Current Natural Sciences

OPTIMIZATION

Alain BILLIONNET

Designing Protected Area Networks

A Mathematical Programming Approach

Science Press

EDP sciences

Current Natural Sciences

OPTIMIZATION

Designing Protected Area Networks

A Mathematical Programming Approach

Alain BILLIONNET

There is a broad consensus in considering that the loss of biodiversity is accelerating which is due, for example, to the destruction of habitats, overexploitation of wild species and climate change. Many countries have pledged at various international conferences to take swift measures to halt this loss of biodiversity. Among these measures, the creation of protected areas – which also contribute to food and water security, the fight against climate change and people' health and well-being – plays a decisive role, although it is not sufficient on its own.

In this book, we review classic and original problems associated with the optimal design of a network of protected areas, focusing on the modelling and practical solution of these problems.

We show how to approach these optimisation problems within a unified framework, that of mathematical programming, a branch of mathematics that focuses on finding good solutions to a problem from a huge number of possible solutions. We describe efficient and often innovative modellings of these problems.

Several strategies are also proposed to take into account the inevitable uncertainty concerning the ecological benefits that can be expected from protected areas. These strategies are based on the classical notions of probability and robustness.

This book aims to help all those, from students to decision-makers, who are confronted with the establishment of a network of protected areas to identify the most effective solutions, taking into account ecological objectives, various constraints and limited resources.

In order to facilitate the reading of this book, most of the problems addressed and the approaches proposed to solve them are illustrated by fully processed examples, and an appendix presents in some detail the basic mathematical concepts related to its content.

ISBN : 978-2-7598-2521-9

9 782759 825219

edp sciences

www.edpsciences.org

Current Natural Sciences

Alain BILLIONNET

Designing Protected Area Networks

A Mathematical Programming Approach

edp sciences

Alain Billionnet is Professor emeritus at the École Nationale Supérieure d'Informatique pour l'Industrie et l'Entreprise (ENSIIE), researcher at the Centre d'études et de recherche en informatique et communication (Cédric) of the Conservatoire national des arts et métiers (Cnam).

Financial support for the publication of this book was provided by the Combinatorial Optimisation team of the Cédric laboratory. This support is gratefully acknowledged.

Printed in France

EDP Sciences – ISBN(print): 978-2-7598-2521-9 – ISBN(ebook): 978-2-7598-2537-0
DOI: 10.1051/978-2-7598-2521-9

To Ulysse, Valentin, Albane, and Agathe

Introduction

Biodiversity, a contraction of the words biological and diversity, represents the diversity of living organisms (animals, plants, fungi, and bacteria) and their living environments (aquatic, terrestrial, underground, aerial). The diversity of living organisms refers to the diversity of species and the diversity of genes within each species. The species richness of a given place, *i.e.*, the number of species present in that place, is a widely used measure to quantify the biodiversity of that place. This measure is easy to interpret, but inventorying the species present at a given location can be a difficult and costly exercise. Some species such as higher plants and vertebrates are easy to observe but this is not the case, for example, with fungi and bacteria. The number of individuals of each species is also an interesting indicator. It can also be difficult to estimate. Genetic diversity essentially corresponds to the diversity of alleles – versions of a gene – within individuals of the same species. This diversity, which results from mutations and reproduction between individuals, allows species to adapt to changes in their environment. Observing this diversity requires sophisticated and relatively expensive techniques. The notion of biodiversity also includes the interactions that exist between living organisms and also the interactions between these organisms and their living environments. Today's biodiversity is the result of a slow evolution of the living world, spread over billions of years and affecting the entire planet. There is a broad consensus that the preservation of biodiversity is currently a major issue. Biodiversity provides irreplaceable and essential goods for our survival, such as food, oxygen, medicines, and raw materials. In addition, species such as insects, bats, and birds pollinate plants. Finally, natural environments contribute to natural water purification, flood prevention, landscape structuring, and the quality of our living environment.

There is also a broad consensus to consider that biodiversity loss is accelerating and that the five major causes of biodiversity loss are habitat destruction (*e.g.*, urbanization, deforestation, wetland drying), biological invasions (*e.g.*, Japanese knotweed, coypu), pollution (*e.g.*, release of a large number of toxic substances into

DOI: 10.1051/978-2-7598-2521-9.c991

the environment and wide distribution of these substances), overexploitation of species (*e.g.*, African and Asian rhinos, bluefin tuna, ebony) and climate change, including its rate. For example, according to the World Wildlife Fund (WWF), global vertebrate populations declined by almost 70% between 1970 and 2016. Stopping the loss of biodiversity is, therefore, one of the major challenges facing the international community today. Many countries are committed to take early action to halt biodiversity loss.

Protected areas play a decisive role in maintaining biodiversity because they make it possible to directly target the protection of elements at high risk of extinction. Thus, at the 10th Conference of the Parties in Nagoya (COP 10), the signatory countries adopted a Strategic Plan for the period 2011–2020 with 20 key objectives to improve biodiversity conservation. Target 11 sets the global coverage of protected areas to be at least 17% of terrestrial and inland water areas and at least 10% of marine areas. These protected areas may require the restoration of degraded habitats such as reforestation, reintroduction of species, control of invasive species, and restoration of wetlands. They can be created on a regional, national or even international scale and be linked in networks in a physical or organizational way. They already occupy a significant fraction of the Earth's surface and generally aim to preserve several aspects of biodiversity simultaneously. They can also protect species still unknown to scientists. Some species are very sensitive to human presence and many activities can be prohibited within protected areas, such as habitat transformation, hunting, fishing, tourism, and sports. The term protected area is now very often used. However, there are many other terms used to designate these regulated areas for nature protection: nature park, nature reserve, protected zone, conservation area, protected site, etc. The International Union for Conservation of Nature (IUCN) identifies six categories of protected areas, terrestrial and marine, according to their management objectives and defines a protected area as "a clearly defined, recognized, dedicated and managed geographical area, by any effective legal or other means, to ensure the long-term conservation of nature and its associated ecosystem services and cultural values". Ideally, each threatened species or, more broadly, each threatened ecosystem should benefit from an area whose protection ensures its future. In some regions, protected areas may be the only remaining natural areas. As a result, they can support species that are not found elsewhere. To simplify the presentations we will mainly focus in this book on the protection of species, but all the developments could be applied to the protection of other aspects of biodiversity.

Several indicators can be used to measure the effects generated by the creation of protected areas, such as the number of protected species, their degree of vulnerability, the population size of each protected species, the genetic or phylogenetic diversity of protected species, or combinations of these indicators. This measure of the effects of protection may also include the ecosystem services it provides, such as food, water, cultural values, health products, and recreational areas. However, these aspects can be difficult to assess. The protection of natural areas is also an effective strategy for mitigating climate change. Protected areas must be large enough and suitable for the protection of the targeted protection, but at the same time must not be too detrimental to the needs and habitats of the populations living in or near these areas. Given these human pressures and the direct costs associated with protected areas, a trade-off will often have to be made between the ideal size of protected areas and the size ultimately chosen. The delimitation of protected areas often helps to avoid excessive habitat fragmentation. Non-contiguous protected areas can be organized in a network for global management. They can also be more or less linked by biophysical connections such as biological corridors. Finally, an assessment of the effects of protection must be carried out regularly to ensure that it is effective, *i.e.*, that the objectives of maintaining biodiversity are being met. Indeed, the objective of some protected areas may not be achieved due, for example, to illegal behaviour or climate change. Good management of these areas is, therefore, extremely important. There are other species' protection strategies such as, for example, the control of invasive species or captive breeding followed by reintroduction into the wild. The latter strategy may be necessary in an emergency situation. Of course, the development of protected areas, although extremely effective in conserving biodiversity, is not, on its own, sufficient to ensure such conservation. Thus, using land-use and biodiversity models, researchers have recently shown that an approach combining important land protection measures and a transformation of the food system would make it possible to redress the curve of biodiversity loss by 2050.

We are interested here in the choice of natural sites to be protected with the main objective of protecting biodiversity – representation and persistence – but this biodiversity protection can be combined with other objectives (*e.g.*, preservation of drinking water, cultural heritage, and creation of a recreational, research or educational area, flood prevention). As the resources available for this protection are obviously limited, it is important to use them as effectively as possible. It is recognized that protected areas have saved important species and natural environments. However, the erosion of global biodiversity continues at a rapid rate. This is why the creation of new protected areas as well as the optimal choice of them is important. The objectives, many and varied, must be well defined, the possible actions must be identified and the impact of these actions must be assessed. For example, a good knowledge of the geographical distribution of endangered animal and plant species is fundamental. A large number of studies on the selection of sites whose protection is relevant to biodiversity conservation have been published in the operational research and biological conservation literature.

In this book, we provide an overview of classical but also original problems related to the "optimal" design of a network of protected areas, focusing on the modelling approach and finally on their resolution. By "design of a network of protected areas" we mean the process of choosing, within a territory, portions of territories to be protected, *i.e.*, managed with the explicit aim of contributing to the protection of certain species and ecosystems associated with these territories. These territories and portions of territories can be very different in size. Many problems are considered and described in detail in this book – some of them have already been the subject of occasional publications on my part – but this overview is far from exhaustive, as there are so many questions inherent in the optimal design of a protected area network. Numerous references are presented. They concern both the field of optimisation in general and the field of protected area design.

We show how to approach these optimisation problems in a unified framework, that of mathematical programming. Within this framework, we propose efficient and often innovative modellings. Mathematical programming (linear, quadratic, fractional, and convex, in real or integer variables, by objectives) is a branch of mathematics that focuses on finding the "best" solution to a problem, among a very large number of possible solutions. It generally consists in studying and solving a problem expressed as the search for the optimum of a function of n variables. This function – called objective function or economic function – enables the quality of a solution to be measured in relation to the pursued objectives, the variables being subject to linear or non-linear constraints expressed by equalities and inequalities. The objectives may be technical, ecological, sociological, economic or a combination of them. Mathematical programming is, therefore, a very general framework for addressing optimization problems that arise in many fields. Research in this domain of mathematics has been stimulated for many years by the possibility of using more and more powerful solvers such as, for example, IBM-ILOG-CPLEX, FICO-XPRESS or GUROBI. Their impressive performance is based on theoretical and algorithmic results, the effective implementation of these results and the spectacular increase in computer computing speed. It is thus currently possible to solve mathematical programs with thousands of variables and constraints and even much more in the case of linear programming. One of the important advantages of mathematical programming – compared to other approaches for dealing with optimization problems – is its flexibility. It is very easy to modify the objective function and constraints, if this is necessary to take into account, for example, variations in the objectives or characteristics that the desired protected areas must satisfy. In this book, we study many optimization problems associated with the design of a protected area network and show how to formulate them in the framework of mathematical programming. We will see that all kinds of complex objectives and constraints can be easily taken into account. The considerable interest of this approach lies in the fact that, when a problem is formulated in this way, the computer implementation of its resolution is particularly simple using a modelling language coupled with a solver, and powerful languages of this type as well as extremely efficient solvers – mentioned above – are available. The efficient computer implementation of an algorithm specially designed for a particular problem is generally much more difficult. The mathematical programming approach is, therefore, particularly appropriate to help a decision-maker to quickly consider a project to design a network of protected areas. We have just mentioned the technical advantages of mathematical programming to address the problems associated with the design of protected zones. Another advantage of this approach is that in order to be tackled in this way, the problems must be analysed rigorously, since the objectives, constraints and data must be precisely defined. This will often provide an opportunity to clarify certain points. Finally, and this last aspect is extremely important, the solutions proposed are impartial and transparent. However, the fact that a problem can be formulated as a mathematical program does not imply that it can be solved in a reasonable time. Furthermore, the decision-makers and protected area managers must be closely involved in the construction of the models. Note that graph theory is also widely used in this book, mainly as a modelling tool used prior to a mathematical programming formulation. Graph theory is a rich branch of discrete mathematics that studies networks of points connected by lines called arcs or edges.

Many publications in the biological conservation literature address these optimization problems related to the delimitation of protected areas, but they often propose to deal with them by approximate methods, specific heuristics or metaheuristics. These latter are generic heuristics that must be adapted to each problem. These approximate methods are relatively easy to implement and may require less computation time than that required to solve a mathematical program, but they can provide solutions whose value is quite far from the value of the optimal solution. Moreover, it is not generally known whether the value of the solution provided is close or not to the value of the optimal solution. More recently, some problems related to the creation of protected areas that are "optimal" in terms of biodiversity protection have been addressed within the framework of constraint programming.

An important aspect to be taken into account in the design of a network of protected areas is the uncertainty regarding the effects of these nature protection policies. Indeed, a large number of uncertainties exist in the medium and long term about the factors influencing biodiversity. Some are due to human activities such as agriculture, urbanization or climate change, at least in part, others are simply due to errors in measurements and forecasts. There are many approaches to try to account for this uncertainty. It can be conventionally translated into probabilities – difficult to define. These probabilities concern specific events affecting biodiversity and likely to occur in the future given the protection policies adopted. For example, it can be estimated that the probability that a certain species will have disappeared from a certain site in 10 years is 0.9 if no particular action is taken for the protection of this species in this site. This uncertainty can also be taken into account in other ways. For example, it can be assumed that several scenarios – coherent sets of assumptions – are possible and the forecasts used to construct the models will depend on the scenario. For example, it can be considered that the sites whose protection would allow the survival of a given species over a 50-year period are different depending on the scenario considered. Both scenarios and probabilities can also be taken into account simultaneously. For example, it can be considered that the survival probability of a species in a given area and over a certain time horizon depends on the scenario. Another way of taking uncertainty into account is simply to consider that the different values of measurements or forecasts, in the medium or long term, are subject to uncertainties or errors. For example, the population size of a given species in a given area may be estimated to be between 1,000 and 1,500 units after 10 years, or the survival probability of a given species in a given area over a 50-year period may be estimated to be between 0.8 and 0.9. Of course, this type of uncertainty can be combined with considering several scenarios.

Let us now present a little more precisely the general framework of this book. We consider a set of species, animal or plant, or other aspects of biodiversity, that are in risk of disappearing. To simplify the presentation throughout this book, reference will almost always be made to a set of threatened species, but all the proposed approaches would easily adapt to other threatened aspects of nature and biodiversity, such as valuable habitats or ecological processes. It should be noted that according to the World Wildlife Fund (WWF) many common species are also experiencing a significant decline that should at least be slowed down. A certain horizon (*e.g.*, 10 years, 50 years or 100 years) and a set of zones – also called sites or parcels or areas – where these species live are considered. These zones can be very different in nature (*e.g.*, natural zones of ecological, faunistic and floristic interest, zones of the Conservatoire du Littoral, rivers, wetlands). The protection of these different zones can have a very different but complementary impact on biodiversity protection. At the beginning of the horizon considered, it may be decided to protect certain zones in order to provide some protection to the species considered and present in these zones, and thus increase their chance of survival. These decisions may eventually be called into question throughout the horizon under consideration if this is still possible. Protection measures are appropriate to the conservation objectives sought and vary from one zone to another. Thus, certain activities may be authorized in one protected zone and strictly prohibited in another (*e.g.*, destruction of embankments or hedges, construction, hunting, fishing, certain agricultural activities, public circulation, gathering). One way to protect a zone is to include it in a nature reserve. Protecting a zone has a cost. This cost takes into account, for example, the acquisition of the zone and its management over time. It may also reflect some costs that are more complex to assess such as social costs. It is also considered, as mentioned above, that the decisions taken require consideration, as far as possible, of the various uncertainties. To protect a given species or a given set of species, different measures to protect the zones can be adopted. In general, the more important these measures are, the greater the chances of survival of the species concerned – their survival probabilities – are. Thus, with any subset of protected zones is associated an assessment of the value of protecting these zones. For example, it can be simply considered that there are only two possible decisions for a zone, to protect it or not during the period considered, and that its protection automatically ensures the survival – survival probability equal to 1 – of the species present in that zone at the beginning of the period. Thus, in this case and for figure I.1a, the protection of the zones z_2, z_5, z_{16}, and z_{18} ensures the survival of the species s_3, s_4, s_6, s_7, s_9, and s_{11}.

Let us now look at the survival probability of the species. First of all, let us consider the case where only one scenario is envisaged. By definition, the survival probability of a given species throughout the period considered depends on the protection measures decided in favour of that species at the beginning of the period. In one of the extreme cases, where this probability takes the value 0, the species certainly disappears and in the other extreme case, where this probability takes the value 1, it certainly survives. Let us again take the example of figure I.1a and assume that the survival probability of the species present in a zone at the beginning of the period considered is equal to 0 if the zone is not protected and 0.5 if the zone is protected. Let us also assume that the interest associated with the protection of zones is measured by the mathematical expectation of the number of species that will survive, in all zones, protected or not, until the end of the period. Thus, the protection of the zones z_2, z_4, and z_{11} provides an interest equal to 2.5 while the protection of the zones z_{10}, z_{19}, and z_{20} provides an interest only equal to 2.375. For these calculations, it is assumed that all the probabilities are independent. In a

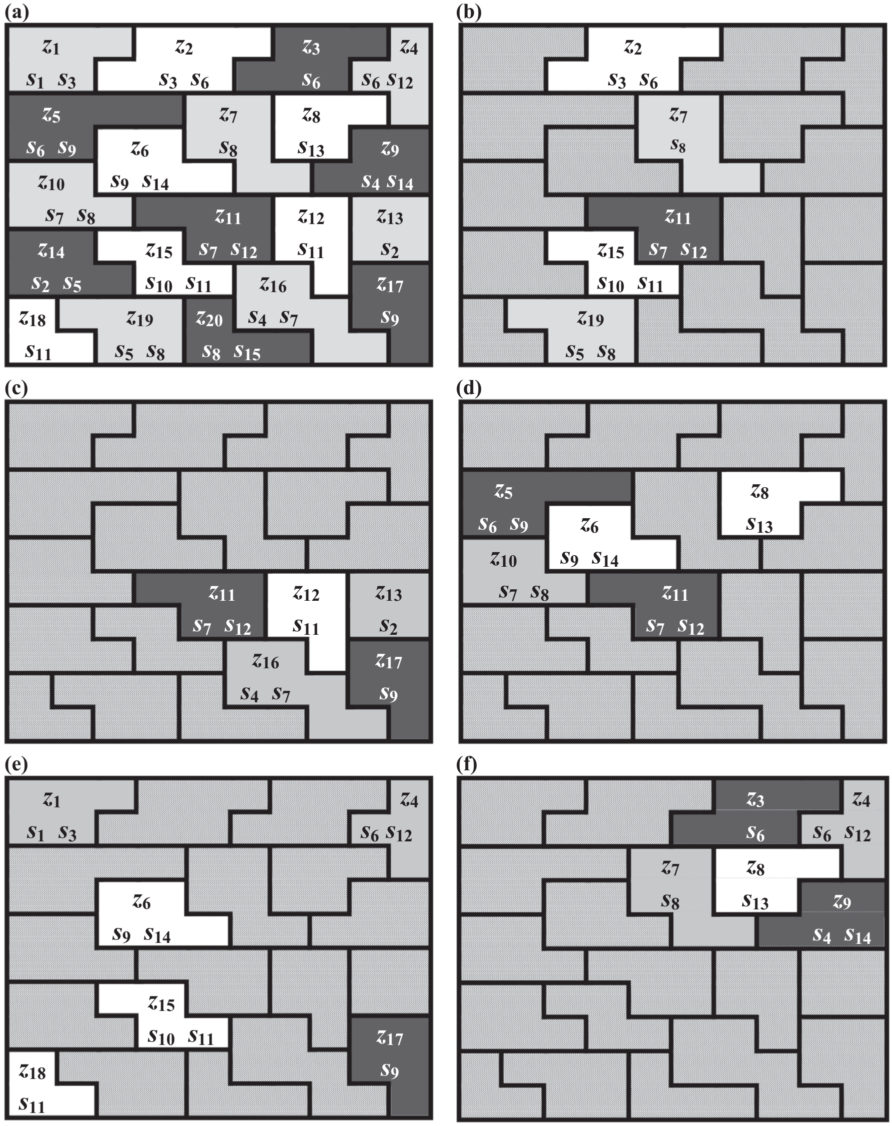


FIG. I.1 – (a) A hypothetical set of 20 candidate zones for protection and the list of species living in each of these zones, among the 15 species considered. The cost of protecting the white zones is equal to one unit, the cost of protecting the light grey zones is equal to two units and the cost of protecting the dark grey zones is equal to four units. (b) Protection of five zones forming a one-piece but not very compact reserve which protects, at least in some way, the 8 species s_3, s_5, s_6, s_7, s_8, s_{10}, s_{11}, and s_{12}. (c) Protection of five zones forming a single, compact reserve. (d) Protection of five zones forming two reserves in one piece and relatively close to each other. (e) Protection of six zones forming a highly fragmented reserve. (f) Five zones are protected but only z_8 belongs to the central part of the reserve, the other four zones share a common border with unprotected zones and thus form a buffer part of the reserve.

general way, these probabilities are obviously difficult to establish. One way of taking into account the uncertainty that inevitably affects these probabilities is to consider, for example, that they belong to a certain interval.

Consider now the case where several scenarios are possible. By definition, the survival probability of a given species throughout the period considered depends as before on the protection measures decided in favour of that species at the beginning of the period but also on the scenario that is envisaged. As in the case of a single scenario, this probability can take any value between 0 and 1, including the values 0 or 1, or it may not be known with certainty, in which case only the interval to which it belongs is known. Similarly to the case of a single scenario, with any subset of protected zones is associated an assessment of the interest provided by the protection of these zones – in terms of biodiversity protection – but, in the case of several scenarios, this interest depends on the scenario under consideration.

Below are some examples of constraints that may be imposed on a set of zones that are being considered for protection. For example, we can impose purely spatial constraints on this set of zones, which we call, for the sake of simplicity, "reserve". These constraints may concern the shape of the reserve, its connectivity, *i.e.*, the contiguity of the different zones composing it, its compactness and its degree of fragmentation measured by different indicators, its edge length, *i.e.*, the length of the transition zones between two different habitats, etc. It should be noted that biodiversity and habitat quality within these transitional areas, the edges, can be negatively affected (alterations at the microclimate level, interactions between species such as predation and competition, development of invasive species). Therefore, efforts will generally be made to limit the "edge effect" as much as possible. However, these areas are sometimes favourable to certain interesting species.

Let us return to the example in figure I.1. Figure I.1b shows a set of 5 protected zones, in one piece but relatively non-compact. On the contrary, figure I.1c shows a set of 5 protected zones, in one piece and compact. Figure I.1d shows a set of 5 zones, relatively compact but made up of two groups of zones of a one-piece each. Figure I.1e shows a highly fragmented reserve of 6 zones.

Once we are able to define the interest associated with the protection of any subset of zones, for any possible scenario, several problems naturally arise. A first type of problem is to determine the optimal set of zones to protect, given limited resources and constraints on the selected zones. In the case of a single scenario, an optimal set of zones is a set of zones of maximal interest. In the case of several scenarios, an optimal set of zones is more difficult to define. This could be, for example, a set of maximal interest in the worst-case scenario, *i.e.*, in the scenario that is the most unfavourable to the set of selected zones. We can also search for a set of zones whose interest, regardless of the scenario that occurs, is not too far from the interest of the set of maximal interest for that scenario. This allows for the identification of a feasible set of zones within the available budget and minimizing the maximal relative difference, the maximal "regret", over all the scenarios, between the interest provided by the protection of this set of zones for the scenario under consideration and the maximal interest that could have been achieved if it had been known that this scenario would occur. A second type of problem is to determine the feasible set of zones of minimal cost that must be protected to achieve a certain interest. In the case of a single scenario, this amounts to determining a set of zones, with a minimal cost and whose protection interest is greater than or equal to a certain value. In the case of several scenarios, an approach may be developed to identify a set of zones, with minimal cost and protection interest that is greater than or equal to a certain value for all the scenarios considered. This value may depend or not on the scenario.

We now give some examples of measures of the interest associated with the protection of a subset of zones – called a reserve for simplicity's sake – with regard to biodiversity. This interest can be assessed by the following measures, or a combination of them: the number or mathematical expectation of the number of species protected by the reserve; the diversity or mathematical expectation of the diversity of species protected by the reserve, measured in different ways (*e.g.*, phylogenetic diversity, Simpson diversity index); the size of the populations of the species protected by the reserve; the amount of carbon sequestered and/or captured by the reserve over time. If more than one scenario is considered, all these measures may be scenario-dependent.

Some examples are also given below of conditions that must be met with regard to the zones of the reserve in order to increase the biodiversity protection in this reserve and over the period considered: the reserve must contain, at the beginning of the period, a total number of species of a given set greater than or equal to a certain threshold value; the zones of the reserve must be sufficiently close to each other or even contiguous; the reserve must have a central part and a buffer part (for example, a zone can be considered to belong to the central part of the reserve if it is "completely surrounded" by other zones of the reserve, see figure I.1f); the reserve may have several contiguous "sub-reserves" but these must be linked by a network of biological corridors; in order to guard against natural risks that may occur and destroy certain zones of the reserve (*e.g.*, storm, fire, and flooding) species must be protected by several zones. Again, these conditions may depend on the scenario.

In everything we have just seen, protection strategies consist, for a given zone, in protecting it or not. The result is a set of protected zones and finally a more or less strong protection, possibly non-existent, of the species or ecosystems concerned. The relationship between "protected zone" and "chances of survival of a species" can be quite complex. A generalization of all this consists in considering that there are, for each zone, several levels of protection and not just one. For example, for a given zone, the survival probability in that zone of a given species is 0.5 if the zone is not protected, 0.8 if a certain level of protection is provided for that zone, and 0.9 if another – higher – level of protection is provided.

In summary, an important objective of this book is to help those who have to make decisions regarding the establishment of a network of protected areas to do so in an "optimal" way, *i.e.*, in the best possible way with regard to the protection of some biodiversity aspects while taking into account various constraints. Specifically, this means selecting the zones to be protected from a set of candidate zones and determining the level of protection to be applied to these zones. These decisions, aiming at the best possible protection of biodiversity, must take into account the criteria chosen to assess biodiversity, the information available in relation to these criteria, limited resources, random factors and spatial constraints of varying complexity. Thus we hope to have shown in this book the interest of optimisation models in designing a network of protected areas. We also hope that the reader will not be too put off by the mathematical formalism that is needed for the presentation of mathematical programs. We hope that the very numerous examples will facilitate his/her reading.

The study of these optimization problems involves several steps: problem definition and modelling, formulation by a mathematical program, possible reformulation by a mathematical program that can be solved effectively, *i.e.*, within a reasonable computation time, pre-treatments, *i.e.*, study of the structure of the problem in order to reduce the number of variables and/or constraints, and possible improvement of the chosen formulation. There are generally several ways to model an optimization problem using a mathematical program and an important question is, therefore, to find the "right" model, *i.e.*, the one that solves the problem in a reasonable computing time while not being too difficult to interpret. These different steps are illustrated in many examples. In order to lighten the presentation and allow the reader to follow the different steps, these examples are hypothetical but generally described in detail. However, the optimisation models that are presented, even if they sometimes include simplifying hypotheses so as to not lose perspective on the proposed approach, can be applied to real-world problems. All the mathematical programs associated with these examples have been modelled using the AMPL language and resolved by CPLEX, a solver based on the most efficient algorithms available today. The experiments have been carried out on a PC with an Intel Core Duo 2 GHz processor and mainly using the solver CPLEX version 12.6. The results obtained and the study and interpretation of the solutions are presented. It is often interesting to examine several solutions of a given problem: all the optimal solutions as well as some solutions close to them. Indeed, this may allow certain criteria that are difficult to formalize to be taken into account. Performance indications such as computation times are also provided for large instances.

The whole approach described above can be an effective decision-making tool for the actors involved in biodiversity conservation policies based on the creation of protected zones. This tool does not replace the actors but can be used to recommend behaviour by clearly highlighting the consequences of the various possible decisions in relation to the objectives of these actors. It should be noted that a protected zone is envisaged on the basis of ecological objectives and criteria, but that its actual establishment depends on a number of other factors, including stakeholder-dependent economic and political ones. The significant gap between theoretical studies and practical implementations is often mentioned in the conservation literature. This gap can certainly be narrowed by establishing closer collaboration between "theorists" and "practitioners" during all the stages of a protected zone network design project.

The reader will not find in this book a study of the specific problem in which he/she is interested because the possible optimization problems, in connection with the creation of protected zones, are extremely numerous and varied. On the other hand, he/she will generally find a similar problem from which he/she can draw inspiration, thanks to the flexibility of mathematical programming, to approach his/her own. Above all, he/she will be able to find a general approach, applicable in many contexts, to address, through mathematical programming, the formulation and resolution of a specific protected zone design problem. It should be noted that there are generic tools such as C-Plan, Marxan and SITES to address these issues. These tools, often based on heuristic methods, have the advantage of being fairly general, but the disadvantage of this generality is that they may not be easily adapted to a specific context. Certainly, in many cases, specific tools need to be developed and we hope that this book will help in the design of such tools.

Plan of the Book

Each chapter deals with a particular aspect involved in the selection of a set of zones to be protected, among a set of candidate zones, and aimed at preserving biodiversity as much as possible. As already mentioned, we mainly deal with species protection in this book, but all the developments presented could be applied to other components of biodiversity. Each chapter first of all presents the interest of the aspect considered with regard to biodiversity protection and then proposes, within the unified framework of mathematical programming, models, formulations and solutions to optimization problems naturally linked to this aspect. Most of these problems are illustrated by detailed examples and numerous computational experiments to evaluate the effectiveness of the proposed approaches are presented. As mentioned, each chapter deals with a specific aspect related to the choice of a set of zones to be protected, but the concrete choice of these zones will generally have to combine several of these aspects.

Chapter 1 deals with the basic problem associated with the optimal choice of zones to be protected as well as some variants of this problem. A use of the AMPL modelling language, coupled with the CPLEX solver, is also presented.

The basic problem can be expressed as follows: what is the set of zones to be protected, among a set of candidate zones, in order to preserve biodiversity "as best as possible"? In this basic problem, we assume that, for each species considered, we know either all the zones whose protection individually ensures the protection of this species, for example its survival, or the minimal population size of this species which must be present in the reserve, *i.e.*, in the set of protected zones, for this species to be considered as protected. Protecting a zone has a cost and protecting biodiversity as best as possible can have several meanings. For example, one can seek to protect as many species as possible within an available budget or seek to protect, at a minimum, a number of species through a minimal cost reserve. A dynamic version of this basic problem, in which zones are progressively protected over time, taking into account a budget constraint related to each period under consideration, is also presented and discussed in this chapter. These elementary problems of zone selection are NP-difficult. In other words, it is conjectured that there is no polynomial-time algorithm to solve them. An algorithm is said to be polynomial in time if the number of elementary operations required to perform it can be expressed as a polynomial depending on the size of the data. However, many optimization problems related to the design of a network of protected areas, although NP-difficult, can be solved efficiently, especially through mathematical programming.

Chapters 2, 3, 4, and 5 deal with the spatial aspects of a set of protected zones.

The spatial configuration of a nature reserve – a set of protected zones – is an essential factor for the survival of the species that live there. Fragmentation, connectivity, compactness or edge length are three important and interdependent aspects of this configuration. Fragmentation is associated with the dispersion of the zones that make up the reserve (chapter 2). This dispersion often results from the fragmentation of space due to artificial phenomena such as the presence of urbanized areas, intensive agricultural areas or transport infrastructures. In contrast, in a connected reserve, all the zones are contiguous and species can circulate easily throughout the whole reserve (chapter 3). The compactness of a reserve corresponds to the distance separating the zones from each other (chapter 4) and this distance can be measured in different ways. The edge of a reserve consists of the transition zones between the reserve and the surrounding matrix (chapter 5). Urban and agricultural development as well as logging can make it difficult to build relatively compact and low-fragmented reserves. Fragmentation, combined with lack of compactness, prevents species from moving around the reserve as they should and could in a compact and non-fragmented reserve, contributing to a loss of biodiversity. Of course species are affected differently by the fragmentation and compactness of their habitat. It should be noted that the ease of movement of species within a reserve is not always without its disadvantages, as it can increase the risk of disease transmission or facilitate the proliferation of invasive species. Chapter 4 also addresses the problem of selecting a set of zones by taking into account both the connectivity and compactness criteria.

Chapter 6 deals with biological – or wildlife – corridors. These allow species to move through more or less fragmented landscapes.

Landscape fragmentation, mainly due to urbanization, agriculture and forestry, is an important cause of biodiversity loss as it prevents species from moving as they should. One of the options commonly used to establish – or restore – some connectivity between different habitat areas is the establishment of corridors. This connectivity within a landscape is considered an essential element for biodiversity conservation. Several aspects of optimal corridor design are presented in this chapter, including the restoration of an existing corridor network in order to increase its permeability.

Chapters 7, 8, and 9 deal with the choice of a set of zones to be protected in view of the inevitable uncertainties affecting the protection effects. Several ways of taking these uncertainties into account are presented.

In all the previous chapters it is assumed that the effects of protection – or not – of the different zones are perfectly known. In chapters 7, 8, and 9, we introduce the integration of a certain uncertainty in these effects. A first way to reflect uncertainty is to assume that protecting a zone ensures the survival of a given set of species in that zone with a certain probability – difficult to establish – for each of those species (chapters 7 and 9). A second way of translating uncertainty about the effects of zone protection is to consider, as before, that the protection of a zone enables the survival of certain species with a certain probability, but it is now assumed that these probabilities can be affected by errors (chapter 7). Finally, a third way to translate uncertainty about the effects of zone protection is to consider that several scenarios are possible (chapters 8 and 9). A scenario is a set of consistent hypotheses on the evolution of the direct or indirect factors that may affect the survival of the considered species. It is hypothesised that it is possible to assess the impact of this evolution. The effects of protecting a zone then depend on the scenario that occurs.

Chapter 10 concerns the choice of zones to be protected in order to maximize the phylogenetic diversity of the impacted species. This measure takes into account both the evolutionary history of the species under consideration and their kinship relationships. The information necessary to implement this approach may be relatively difficult to obtain.

Many authors suggest that the effectiveness of protected areas could be significantly enhanced by taking into account criteria other than species richness or abundance when assessing a set of species from a biodiversity perspective. An interesting measure which is increasingly being used in the field of conservation is phylogenetic diversity. It is based on the concept of the phylogenetic tree associated with the set of species considered and reflects the evolutionary history of these species and their kinship relationships. There are different ways to define phylogenetic diversity. We consider here that the phylogenetic diversity of a set of species is equal to the sum of the branch lengths of the phylogenetic tree associated with this set. Several ways of taking into account the inevitable uncertainty affecting the phylogenetic tree associated with a set of species – tree structure and branch length – are also proposed.

Chapter 11 deals with the selection of zones to be protected, based on different measures of the diversity of a set of species that have not been considered in previous chapters.

In the first part of this chapter, we examine the choice of the zones to be protected by measuring the diversity of the species thus protected by indicators other than species richness or phylogenetic diversity. We measure this diversity in three different ways: the first takes into account the dissimilarity or distance between 2 species which can be represented, for example, by the genetic distance calculated from the differences between DNA sequences; in the other two cases, we are interested in the diversity of protected species as measured by two classical indices, the Simpson's index and the Shannon–Wiener index. These two indices take into account both species richness and abundance of each species. In the second part of this chapter, we focus on the set of individuals, of a given species, concerned by the choice of zones to be protected and we measure the diversity of this set by its average kinship.

Chapter 12 takes into account an increasingly important issue to incorporate into the design of protected zones, namely climate change. Indeed, a substantial number of species can lose valuable habitat in a set of protected zones if the climate changes. We are also interested in the choice of zones to be protected in order to mitigate climate change.

Climate change appears to be an important emerging issue to be taken into account in the development of protected zones. Most of the issues discussed in the previous chapters can be re-examined in the context of climate change. Thus, the approach is illustrated by taking as a starting point some basic problems, some of which having already been discussed in previous chapters. In the context of optimal choice of zones to be protected, climate change can be taken into account in different ways: some zones are likely to protect certain species at certain times but this is no longer the case in later periods and conversely, some zones, at certain times, do not allow for the protection of certain species but will allow it in later periods; the population size of the different species considered in each zone changes over time and it is assumed that this change is known; the area of habitat favourable to a given species in a given candidate zone changes over time and it is assumed that this change is known. We also examine cases where there is uncertainty in predicting the impact of climate change, using a probabilistic approach and also a scenario-based approach. We are also interested in a dynamic choice of the zones to be protected: some zones, acquired at certain times to be protected, may be ceded in subsequent periods. Finally, in this chapter we examine a two-criterion problem consisting in selecting a reserve whose interest is assessed by the number of species it allows to protect but also by the quantity of carbon – one of the main greenhouse gases – it allows to capture and stock. Protected zones can, for example, limit the loss of forests, which is considered an important cause of climate change since forests contain the largest terrestrial carbon stock.

The appendix presents basic concepts concerning mathematical programming, graph theory and Markov chains, in relation to the content of this book, as well as references to further explore these topics. These concepts are illustrated by many examples.

References and Further Reading

Aubertin C., Rodary E., Eds (2008) Aires protégées, espaces durables ? IRD Editions.

Ball I.R., Possingham H.P., Watts M.E. (2009) Marxan and relatives: software for spatial conservation prioritization, *Spatial conservation prioritization* (A. Moilanen, K.A. Wilson, H.P. Possingham, Eds), Oxford University Press.

Barnes M.D., Craigie I.D., Dudley N., Hockings M. (2016) Understanding local-scale drivers of biodiversity outcomes in terrestrial protected areas, *Ann. N.Y. Acad. Sci.* **1399**, 42.

C. Berge (1970) *Graphes et Hypergraphes*. Dunod.

Beyer H.L., Dujardin Y., Watts M.E., Possingham H.P. (2016) Solving conservation planning problems with integer linear programming, *Ecol. Model.* **328**, 14.

Billionnet A. (2007) *Optimisation discrète*. Dunod.

Billionnet A. (2013) Mathematical optimization ideas for biodiversity conservation, *Eur. J. Oper. Res.* **231**, 514.

Bonn A., Gaston K.J. (2005) Capturing biodiversity: selecting priority areas for conservation using different criteria, *Biodiver. Conserv.* **14**, 1083.

Butchart S.H.M. *et al.* (2015) Shortfalls and solutions for meeting national and global conservation area targets, *Conserv. Lett.* **8**, 329.

Cabeza M., Moilanen A. (2001) Design of reserve networks and the persistence of biodiversity, *Trends Ecol. Evol.* **16**, 242.

Chape S., Harrison J., Spalding M., Lysenko I. (2005) Measuring the extent and effectiveness of protected areas as an indicator for meeting global biodiversity targets, *Philosophical Trans. Royal Soc. B* **360**, 443.

Chen D.S., Batson R.G., Dang Y. (2010) *Applied integer programming: modeling and solution*, John Wiley & Sons.

Chvatal V. (1983) *Linear Programming*, W.H. Freeman & Company.

Cocks K.D., Baird I.A. (1989) Using mathematical programming to address the multiple reserve selection problem: an example from the Eyre Peninsula, South Australia, *Biol. Conserv.* **49**, 113.

Costello C., Polasky S. (2004) Dynamic reserve site selection, *Res. Energy Econ.* **26**, 157.

De Werra D., Liebling T.M., Hêche J.-F. (2003) *Recherche Opérationnelle pour Ingénieurs I*, Presses polytechniques et universitaires romandes.

Eklund J., Cabeza M. (2017) Quality of governance and effectiveness of protected areas: crucial concepts for conservation planning, *Ann. N.Y. Acad. Sci.* **1399**, 27.

Faith D.P., Walker P.A. (1996) How do indicator groups provide information about the relative biodiversity of different sets of areas? On hotspots, complementarity and pattern-based approaches, *Biodiver. Lett.* **3**, 18.

Fischer D.T., Church R.L. (2005) The SITES reserve selection system: a critical review, *Environ. Model. Assess.* **10**, 215.

Garey M.R., Johnson D.S. (1979) *Computers and intractability: a guide to the theory of NP-completeness*, W. H. Freeman & Co, New York.

Geldmann J., Barnes M., Coad L., Craigie I.D., Hockings M., Burgess N.D. (2013) Effectiveness of terrestrial protected areas in reducing habitat loss and population declines, *Biol. Conserv.* **161**, 230.

Joppa L.N., Baillie J.E.M., Robinson J.G. (2016) *Protected areas: are they safeguarding biodiversity?* Wiley.

Justeau-Allaire D., Birnbaum P., Lorca X. (2018) Unifying reserve design strategies with graph theory and constraint programming, *Principles and practice of constraint programming* (J. Hooker, eds), CP 2018, Lecture notes in computer science, Springer, Cham, vol. 11008.

Justeau-Allaire D., Vismara P., Birnbaum P., Xavier Lorca X. (2019) Systematic conservation planning for sustainable land-use policies: a constrained partitioning approach to reserve selection and design, *Proc. Twenty-Eighth Int. Joint Conf. Artificial Intell. (IJCAI-19)*, 5902.

Justus J., Sarkar S. (2002) The principle of complementarity in the design of reserve networks to conserve biodiversity: a preliminary history, *J. Biosci.* **27**, 421.

Kingsland S.E. (2002) Designing nature reserves: adapting ecology to real-world problems, *Endeavour* **26**, 9.

Kingsland S.E. (2002) Creating a science of nature reserve design: perspectives from history, Environ. Model. Assess. **7**, 61.

Leclère D., Obersteiner M., Barrett M., *et al.* (2020) Bending the curve of terrestrial biodiversity needs an integrated strategy, *Nature* **585**, 551.

Lecomte J., Sarrazin F., Primack R.B. (2012) *Biologie de la conservation*, Dunod. (A Primer of Conservation Biology, in English).

Lindenmayer D. *et al.* (2008) A checklist for ecological management of landscapes for conservation, *Ecol. Lett.* **11**, 78.

Margules C.R., Pressey R.L. (2000) Systematic conservation planning, *Nature* **405**, 243.

Margules C.R., Sarkar S. (2007) *Systematic conservation planning*, Cambridge University Press.

Ministère de la Transition écologique et solidaire (2018) Biodiversité et paysages, https://www.ecologique-solidaire.gouv.fr/politiques/biodiversite-et-paysages.

Millspaugh J., Thompson F., Eds (2008) *Models for planning wildlife conservation in large landscape*, Elsevier.

Minoux M. (2008) *Programmation mathématique – Théorie et algorithmes*, Lavoisier.

Moffett A., Sarkar S. (2006) Incorporating multiple criteria into the design of conservation area networks: a minireview with recommendations, *Divers. Distrib.* **12**, 125.

Moilanen A. (2004–2014) Zonation, Spatial conservation planning methods and software, User Manual. Conservation Biology Informatics Group, University of Helsinki.

Moilanen A., Kujala H., Leathwick J.R. (2009) The zonation framework and software for conservation prioritization, *Spatial conservation prioritization* (A. Moilanen, K.A. Wilson, H.P. Possingham), Oxford University Press.

Naro-Maciel E., Sterling E.J., Rao M. (2008) Protected areas and biodiversity conservation I: reserve planning and design, *Lessons Conserv.* **2**, 19, Published by Network of Conservation Educators and Practitioners, Center for Biodiversity and Conservation, American Museum of Natural History.

Pressey R.L., Watts M.E., Barrett T.W., Ridges M.J. (2009) The C-Plan conservation planning system: origins, applications, and possible futures, *Spatial conservation prioritization* (A. Moilanen, K.A. Wilson, H.P. Possingham, Eds), Oxford University Press.

Parrish J.D., Braun D.P., Unnasch R.S. (2003) Are we conserving what we say we are? Measuring ecological integrity within protected areas, *BioScience* **53**, 851.

Pollock L., Thuiller W., Jetz W. (2017) Large conservation gains possible for global biodiversity facets, *Nature* **546**, 141.

Possingham H.P., Ball I.R., Andelman S. (2000) Mathematical methods for identifying representative reserve networks, *Quantitative methods for conservation biology* (S. Ferson, M. Burgman, Eds), Springer-Verlag, pp. 291–305.

Pressey R.L., Cowling R.M. (2001) Reserve selection algorithms and the real word, *Conserv. Biol.* **15**, 275.

Pressey R.L., Johnson I.R., Wilson P.D. (1994) Shades of irreplaceability: towards a measure of the contribution of sites to a reservation goal, *Biodiver. Conserv.* **3**, 242.

Pressey R.L., Possingham H.P., Margules C.R. (1996) Optimality in reserve selection algorithms: when does it matter and how much? *Biol. Conserv.* **76**, 259.

Pressey R.L., Cabeza M., Watts M.E., Cowling R.M., Wilson K.A. (2007) Conservation planning in a changing world, *Trends in Ecol. Evol.* **22**, 583.

Ramade F. (2008) *Dictionnaire encyclopédique des sciences de la nature et de la biodiversité*, Dunod.

Rao M., Naro-Maciel E., Sterling E.J. (2009) Protected areas and biodiversity conservation II: management and effectiveness, Center for Biodiversity and Conservation of the American Museum of Natural History.

Rodrigues A.S., Gaston K.J. (2002) Optimization in reserve selection procedures - Why not? *Biol. Conserv.* **107**, 123.

Rodrigues A.S. *et al.* (2004) Effectiveness of the global protected area network in representing species diversity, *Nature* **428**, 640.

Roy B. (1969–1970) *Algèbre moderne et théorie des graphes orientées vers les sciences économiques et sociales*, Tome 1: *Notions et résultats fondamentaux*, Tome 2: *Applications et problèmes spécifiques*. Dunod.

Salkin H.M., Mathur K. (1989) *Foundations of integer programming*, North-Holland.

Sarkar S., Pressey R.L., Faith D.P., Margules C.R., Fuller T., Stoms D.M., Moffett A., Wilson K.A., Williams K.J., Williams P.H., Andelman S. (2006) Biodiversity conservation planning tools: present status and challenges for the future, *Annual Rev. Environ. Res.* **31**, 123.

Stolton S., Dudley N., Eds (2010) Arguments for protected areas: multiple benefits for conservation and use, *Earthscan*.

Tilman D., Clark M., Williams D.R., Kimmel K., Polasky S., Packer C. (2017) Future threats to biodiversity and pathways to their prevention, *Nature* **546**, 73.

Vanderbei R.J. (1998) *Linear programming: foundations and extensions*, Kluwer Academic Publishers.

Vanderkam R.P.D., Wiersma Y.F., King D.J. (2007) Heuristic algorithms vs. linear programs for designing efficient conservation reserve networks: evaluation of solution optimality and processing time, *Biol. Conserv.* **137**, 349.

Wang Y., Önal H., Fang W. (2018) How large spatially-explicit optimal reserve design models can we solve now? An exploration of current models' computational efficiency, Nat. *Conserv.* **27**, 17.

Watts M.E., Stewart R.R., Martin T.G., Klein C.J., Carwardine J., Possingham H.P. (2017) Systematic conservation planning with Marxan, *Learning landscape ecology* (S. Gergel, M. Turner, Eds). Springer.

Wiersma Y.F., Sleep D.J.H. (2016) A review of applications of the six-step method of systematic conservation planning, *Forestry Chronicle* **92**, 322.

Wolsey L. (1998) *Integer Programming*, Wiley-Interscience.

Contents

CHAPTER 5

CHAPTER 6

CHAPTER 7

CHAPTER 10

CHAPTER 11

CHAPTER 12

Chapter 1

Basic Problem and Variants

1.1 Introduction

We are interested in a group of species, animal or plant, which, for various reasons, are threatened. They may thus disappear in the more or less near future. For example, the IUCN (International Union for Conservation of Nature) Red List provides information on whether or not a given species is threatened. This list classifies threatened species into three categories, according to their level of extinction risk: "Vulnerable", "Endangered", and "Critically Endangered". This classification is made taking into account various factors such as the population size of the species in question, the rate of decline of this population, the loss and/or fragmentation of its habitat or its genetic erosion. It is also possible to look at a set of focal species, threatened or not, as their protection automatically leads to the protection of many other species. It should be noted that many species, although common, are also in decline and this should be reversed. We are also interested in a set of geographical zones, spread over a territory, that we can decide whether or not to protect, from a given moment on, in order to ensure a certain protection for the species in question and thus increase their chance of survival. The terms "sites", "parcels", "patches", "tasks", "areas", and "islets" are also used to designate these parts of territory. In this book, we will essentially use the term "zone" which, because of its generality, is appropriate in many contexts. The focus here is on species protection, but all of the following could easily be adapted to other threatened aspects of biodiversity such as valuable habitats.

In sections 1.2–1.5 of this chapter, it is considered that there is only one level of protection for the zones. In other words, a zone is protected or not. Decisions on protective actions to be taken at the beginning of the time horizon (*e.g.*, 10 years, 50 years or 100 years) are made at the beginning of this horizon, at which time the candidate zones are in a certain state. Protecting a zone has a cost. It differs from one zone to another and may include monetary, ecological and social aspects.

We can look at the consequences of these decisions at the end of this horizon. For example, it can be assumed that a given species in a protected zone survives at the

DOI: 10.1051/978-2-7598-2521-9.c001

end of the considered time horizon and that this is not the case if it does not occur in a protected zone. The relevance of these hypotheses presupposes that a large amount of information is available, such as the life history and dynamics of the species studied and the size of their population. More simply, it is possible to assess the impact of the protection of a set of zones by the number of species concerned by this protection, without prejudging as precisely the future of these species. It is then only supposed that the chances of survival are greater in protected zones than in unprotected ones. Section 1.6 addresses a significantly different problem for the reason that different protection actions can be considered for each zone. The level of protection of the species present in a zone depends on these actions.

We denote by $S = \{s_1, s_2,\ldots, s_n\}$ the set of species of interest and $Z = \{z_1, z_2,\ldots, z_n\}$ the set of zones that are candidates for protection. To simplify the presentation, a set of protected zones, $R \subseteq Z$, is called a "reserve". For any reserve $R \subseteq Z$, we are interested in the number of species that are protected – at least in a certain way – because of the protection of the zones of R. It is therefore the criterion of species richness that is used here. Thus, this number, which may be difficult to estimate, may represent the number of species that will survive at the end of the chosen time horizon if it is decided to protect the zones of R or, less precisely, the number of species concerned by this protection. We are interested in the overall effect of the protection of the zones of R, *i.e.*, the species richness of these zones considered as a whole – complementarity principle. The cost associated with protecting zone z_i is denoted by c_i. As mentioned above, it can cover several aspects: monetary costs – or possibly gains – (*e.g.*, leasing or acquisition of the zone, potential restoration of the zone, removal of invasive species, zone management, compensation to third parties, income from nature tourism), ecological costs or gains (*e.g.*, habitat quality and ease of movement of the considered species through the zone, involuntary protection of invasive species) and also social costs or gains, which are often difficult to assess (*e.g.*, reduction in possible uses of the zone by the public, access road closures, welfare gains for certain social groups, cultural gains). This cost can also, more simply, represent the area of the zone. Generally, the protection cost of a set of zones, $R \subseteq Z$, is equal to the sum of the protection costs of each of the zones in that set; it is denoted by $C(R)$. The term $\underline{S}$ refers to the set of indices of the species considered and the term $\underline{Z}$ refers to the set of indices of the zones that can be selected for protection. We have thus $\underline{S} = \{1, 2,\ldots, m\}$ and $\underline{Z} = \{1, 2,\ldots, n\}$. It is considered here that any subset of Z can be a priori protected except when a limited budget must be taken into account, since in this case the total cost of protecting the selected zones must not exceed the available budget. It is assumed that the population size of each species in each zone is known. The population size of species s_k in zone z_i is denoted by n_{ik}.

Two different situations are considered, in which a given species, s_k, is protected by a reserve, R. In the first, the protection of an adequate zone is sufficient to protect s_k. In the second, s_k is protected by R if its total population size in R is greater than or equal to a certain threshold value. The number of species protected by a reserve, R, is thus calculated in two different ways. In the first, the result of which is denoted by $\mathrm{Nb}_1(R)$, it is assumed that all the zones whose protection ensures the protection of the species (*e.g.*, its survival) are known for all the species, *i.e.*, for all k of $\underline{S}$. This set is denoted by Z_k and the corresponding set of indices is denoted by $\underline{Z}_k$. In other words, for species s_k to be protected, it is necessary and sufficient that at least one of the zones of Z_k be protected. For example, it is considered here that the protection of a zone makes it possible to protect all the species present in that zone provided that their population sizes in that zone are greater than or equal to a certain threshold value. We note v_{ik} the threshold value associated with species s_k in zone z_i. In other words, $Z_k = \{z_i \in Z : n_{ik} \geq v_{ik}\}$ (see example 1.1 below). In the second way of calculating the number of species protected by a reserve, R, the result of which is denoted by $\mathrm{Nb}_2(R)$, a reserve is considered to protect species s_k, $k = 1, 2,\ldots, m$, if and only if the total population size of that species in the reserve is greater than or equal to a certain threshold value, denoted by θ_k (see example 1.1 below). It should be noted that data on the size of the different populations may be difficult to obtain. The models considered in this chapter are basic models. They can be considered as a starting point to help a decision-maker in thinking about a relevant set of zones to be protected. The fact that solutions are determined, as we will see, by solving a relatively simple mathematical program, facilitates the task. These models can then be extended to take into account different additional aspects. Here again, the mathematical programming approach makes it easy to take these additional aspects into account. We will see many examples of this approach in the rest of this book.

Example 1.1. Consider the instance described in figure 1.1. Suppose that zones z_1, z_2, and z_3 are protected – $R = \{z_1, z_2, z_3\}$ – and that v_{ik} is equal to 4 for any couple (i, k). We obtain $\mathrm{Nb}_1(R) = 4$. Indeed, if the protection of a zone makes it possible to protect the species that are present in that zone provided that their population size is greater than or equal to 4 units, species s_1, s_3, s_6, and s_{11} are protected by reserve $R = \{z_1, z_2, z_3\}$. If we look at the measure $\mathrm{Nb}_2(R)$, for the same reserve, we obtain, assuming that to be protected a species must be present on the reserve with a population whose total size is greater than or equal to 10 – $\theta_k = 10$ for all k –, $\mathrm{Nb}_2(R) = 2$. In this case, only species s_3 and s_6 are protected.

1.2 Protection by a Reserve of All the Considered Species

1.2.1 The Protection of Each Zone Ensures the Protection of a Given Set of Species; the Number of Species Protected by a Reserve, R, is then Denoted by Nb₁(R)

The first question that can be addressed is: what is the set of zones to be protected, at minimal cost, to protect all the species? This problem, which can be stated concisely as the minimization problem $\min_{R \subseteq Z, \mathrm{Nb}_1(R)=m} C(R)$, can be formulated as a linear program in Boolean variables by associating to each zone z_i, $i = 1,\ldots, n$, a Boolean variable x_i, *i.e.*, a variable that can only take the values 0 or 1 (see appendix at the end of this book). By convention, this decision variable takes the value 1 if and

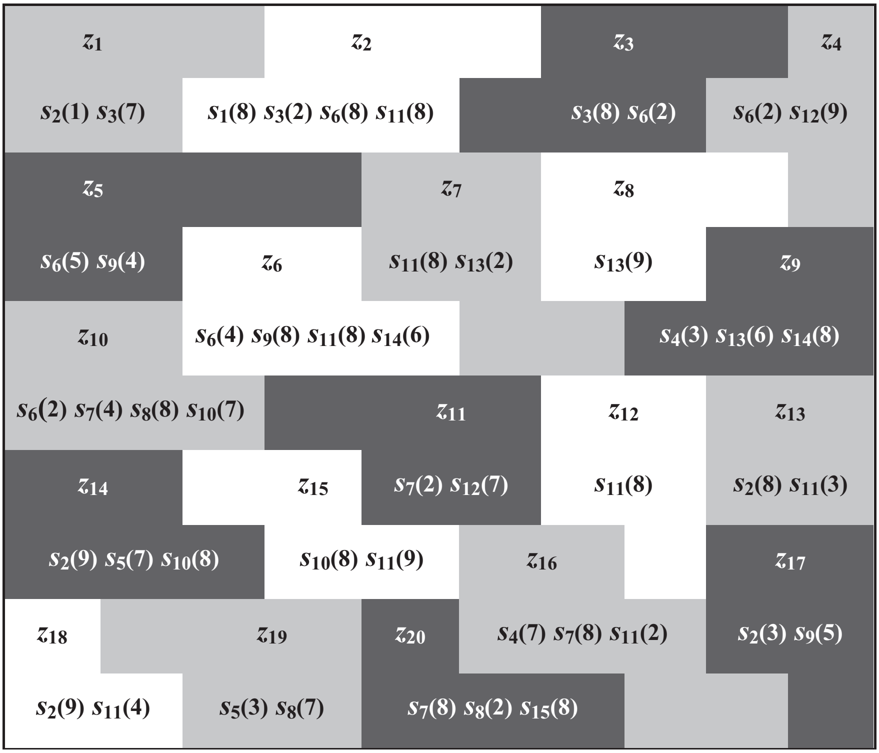


FIG. 1.1 – Twenty zones, z_1, z_2,..., z_{20}, are candidates for protection and fifteen species, s_1, s_2, ..., s_{15}, living in these zones, are concerned. For each zone, the species present and their population size – in brackets – are indicated. The cost of protecting the white zones is equal to 1 unit, the cost of protecting the light grey zones is equal to 2 units and the cost of protecting the dark grey zones is equal to 4 units. For example, species s_6, s_9, s_{11}, and s_{14} are present in zone z_6, their population size is equal to 4, 8, 8, and 6 units, respectively, and the cost of protecting this zone is equal to 1 unit.

only if zone z_i is selected for protection. Program $P_{1.1}$ corresponds to the determination of a reserve of minimal cost allowing all the species to be protected. Program $P_{1.1}$ can admit several optimal solutions, *i.e.*, there may be several reserves allowing all the species to be protected at the lowest cost. In this case, the examination of all the optimal solutions and their evaluation using additional criteria may be necessary to determine the reserve that will finally be selected.

$$
\mathrm{P}_{1.1} : \begin{cases} \min & \sum_{i \in \underline{Z}} c_i x_i \\ \text{s.t.} & \left| \begin{array}{lll} \sum_{i \in \underline{Z}_k} x_i \geq 1 & k \in \underline{S} & (1.1.1) \\ x_i \in \{0,1\} & i \in \underline{Z} & (1.1.2) \end{array} \right. \end{cases}
$$

The objective function of $P_{1.1}$ expresses the total cost associated with protecting the selected zones. Indeed, the cost associated with zone z_i is equal to $c_i x_i$. If we decide to protect zone z_i, which corresponds to $x_i = 1$, then this cost is equal to c_i; if we decide not to protect zone z_i, which corresponds to $x_i = 0$, then this cost is equal to 0. Constraints 1.1.1 express that, for any species s_k, at least one zone of Z_k must be selected for protection. Indeed, at least one zone of Z_k is selected to be protected if and only if at least one of variables x_i – corresponding to zone z_i of Z_k – takes the value 1. Remember that the set Z_k is defined as follows: $Z_k = \{z_i \in Z : n_{ik} \geq v_{ik}\}$. Constraints 1.1.2 specify the Boolean nature of the variables x_i. The problem associated with $P_{1.1}$ is known, in operational research, as the set-covering problem (see appendix at the end of the book).

Example 1.2. Take again the instance described in figure 1.1, with $v_{ik} = 4$ for each couple (i, k). The cheapest strategy for protecting all the species is provided by the resolution of program $P_{1.1}$ – more precisely by the version corresponding to this example – and consists in protecting the 9 zones z_1, z_2, z_4, z_6, z_8, z_{10}, z_{14}, z_{16}, and z_{20}; it costs 19 units. Are there other reserves that cost 19 and protect all the species? This question can be answered simply by looking for a solution that satisfies constraints 1.1.1 and 1.1.2 as well as the 2 additional constraints $\sum_{i \in \underline{Z}} c_i x_i = 19$ and $x_1 + x_2 + x_4 + x_6 + x_8 + x_{10} + x_{14} + x_{16} + x_{20} \leq 8$. This new set of constraints allows for a feasible reserve – of cost 19 – consisting of zones z_1, z_2, z_4, z_6, z_8, z_{14}, z_{16}, z_{19}, and z_{20}.

A variant of this first problem is to consider that, in order to be protected, species s_k must be present – with a sufficient population size – not in at least one protected zone, but in at least β_k protected zones. Indeed, an effective way to guard against random events that could affect a zone (*e.g.*, storm, fire, pollution) and thus eliminate the species present in that zone is to protect several zones for each species. This increases the chances of survival of this species (replication principle). Figure 1.1 shows that, if $\beta_k = 2$ for any k, then the protected zones in the first solution of example 1.2 only protect species s_6, s_7, s_{10}, and s_{11}. It may be noted that, in this example, it is not possible to protect a set of zones in such a way that each species is present – with a sufficient population size – in at least 2 zones of the set. This problem can be formulated as a linear program in 0–1 variables by replacing in $P_{1.1}$ constraints 1.1.1, $\sum_{i \in \underline{Z}_k} x_i \geq 1,\ k \in \underline{S}$, by the constraints $\sum_{i \in \underline{Z}_k} x_i \geq \beta_k,\ k \in \underline{S}$. Indeed, these latter constraints require that, among the variables x_i – corresponding to zone z_i of Z_k – at least β_k of these variables take the value 1.

Other economic functions representing the cost of a reserve may be taken into account. For example, the candidate zones for protection can be considered as a set of q clusters, $\mathrm{Cl} = \{\mathrm{Cl}_1, \mathrm{Cl}_2, \ldots, \mathrm{Cl}_q\}$. More precisely, the q clusters form a partition of the set of zones, Z. Thus, each zone belongs to one and only one cluster and every cluster includes at least one zone. Let us denote by $\underline{\mathrm{Cl}}$ the set of cluster indices. In this case, the cost of protecting a zone consists of two costs: a cost associated specifically with the zone (*e.g.*, acquisition, restoration) and a cost associated with the cluster. The cost associated with cluster Cl_j, which we denote by d_j, is to be supported as soon as one of its zones is selected for protection. On the other hand, if several zones of the same cluster are selected for protection, the cost associated with the cluster is to be supported only once. This cost corresponds, for example, to the delivery of human and material resources to the cluster. The problem of protecting all the species at the lowest cost can then be formulated as the linear program in Boolean variables $P_{1.2}$. To do this, we associate, as before, a Boolean variable x_i to each zone z_i. In addition, with each cluster Cl_j is associated a Boolean variable, u_j, which, by convention, is equal to 1 if and only if at least one zone of cluster Cl_j is selected to be protected.

$$\mathrm{P}_{1.2}: \begin{cases} \min \sum_{i\in \underline{Z}} c_i x_i + \sum_{j\in \underline{\mathrm{Cl}}} d_j u_j \\ \text{s.t.} \left| \begin{array}{llll} \sum_{i\in \underline{Z}_k} x_i \geq 1 & k \in \underline{S} & (1.2.1) & \mid \; x_i \in \{0,1\} \quad i \in \underline{Z} \quad (1.2.3) \\ u_j \geq x_i & (j,i) \in \underline{\mathrm{Cl}} \times \underline{Z} : z_i \in \mathrm{Cl}_j & (1.2.2) & \mid \; u_j \in \mathbb{R} \quad j \in \underline{\mathrm{Cl}} \quad (1.2.4) \end{array} \right. \end{cases}$$

The first part of the economic function represents the cost associated with protecting the zones selected for protection (see $P_{1.1}$) and the second part represents the cost associated with the clusters concerned by this protection, *i.e.*, clusters in which at least one of the zones is selected for protection. Constraints 1.2.1 express that all the species must be protected (see $P_{1.1}$). Because of constraints 1.2.2 and the fact that we are seeking to minimize the costs, the real variable u_j takes the value 0 at the optimum of $P_{1.2}$ if none of the zones of Cl_j is selected for protection and the value 1 in the opposite case. Constraints 1.2.3 specify the Boolean nature of the variables x_i. Note that it is not necessary to further constrain the real variables u_j, $j \in \underline{\mathrm{Cl}}$. Indeed, because of the fact that we are seeking to minimize the quantity $\sum_{j\in\underline{\mathrm{Cl}}} d_j u_j$ and taking into account constraints 1.2.2, the variable u_j takes, at the optimum of $P_{1.2}$, either the value 0 or the value 1.

1.2.2 A Species is Protected by a Reserve, R, if its Total Population Size in R Exceeds a Certain Value; the Number of Species Protected by R is then Denoted by $Nb_2(R)$

In this case, the basic problem, which consists in selecting a set of zones, of minimal cost, whose protection ensures the protection of all the species, corresponds to the minimization problem $\min_{R\subseteq Z, \mathrm{Nb}_2(R)=m} C(R)$ and can be formulated as the linear program in Boolean variables $P_{1.3}$.

$$\mathrm{P}_{1.3}: \begin{cases} \min \sum_{i\in \underline{Z}} c_i x_i \\ \text{s.t.} \left| \begin{array}{lll} \sum_{i\in \underline{Z}} n_{ik} x_i \geq \theta_k & k \in \underline{S} & (1.3.1) \\ x_i \in \{0,1\} & i \in \underline{Z} & (1.3.2) \end{array} \right. \end{cases}$$

As in the previous models, the reserve retained is formed by zone z_i such that $x_i = 1$. The economic function expresses the cost of the reserve (see $\mathrm{P}_{1.1}$). Constraints 1.3.1 express that the total population size of species s_k in the reserve, $\sum_{i\in \underline{Z}} n_{ik}x_i$, must be greater than or equal to the minimal value required for the survival of this species, θ_k, and this for any k of $\underline{S}$.

Example 1.3. Let us take the instance described in figure 1.1 and set θ_k to 7 for any k of $\underline{S}$. The least costly strategy for protecting all the species, when the number of species protected by a reserve R is assessed by $\mathrm{Nb}_2(R)$, is provided by the solution of $\mathrm{P}_{1.3}$. This strategy consists of protecting the 10 zones z_1, z_2, z_4, z_6, z_8, z_9, z_{10}, z_{14}, z_{16}, and z_{20}, and costs 23 units.

1.3 Protection by a Reserve of a Maximal Number of Species of a Given Set Under a Budgetary Constraint

A second basic problem is to determine the zones to be protected, taking into account an available budget, in order to protect, at least in a certain way, the greatest possible number of species. This problem, which consists in maximizing the species richness of the selected reserve, can be expressed in the form of the maximization problem $\max_{R\subseteq Z, C(R)\leq B} \mathrm{Nb}_1(R)$ or $\max_{R\subseteq Z, C(R)\leq B} \mathrm{Nb}_2(R)$, depending on the method of calculating the number of species protected by reserve R. B is the available budget.

1.3.1 The Number of Species Protected by a Reserve, R, is Assessed by Nb₁(R)

The problem can be formulated as a linear program with Boolean variables. As in the previous programs, a Boolean decision variable, x_i, is associated with each zone z_i. With each species s_k is also associated a "working" Boolean variable, y_k, which, by convention, takes the value 1 if and only if at least one of the zones selected to be protected protects species s_k. Thus, when the number of species protected by a reserve, R, is evaluated by $\mathrm{Nb}_1(R)$, the problem considered can be formulated as the mathematical program $\mathrm{P}_{1.4}$.

$$\mathrm{P}_{1.4}: \begin{cases} \max \sum\limits_{k\in \underline{S}} y_k \\ \text{s.t.} \left| \begin{array}{llll} y_k \leq \sum\limits_{i\in \underline{Z}_k} x_i & k\in \underline{S} & (1.4.1) & \mid \quad x_i\in\{0,1\} \quad i\in \underline{Z} \quad (1.4.3) \\ \sum\limits_{i\in \underline{Z}} c_i x_i \leq B & & (1.4.2) & \mid \quad y_k\in\{0,1\} \quad k\in \underline{S} \quad (1.4.4) \end{array} \right. \end{cases}$$

The objective of $\mathrm{P}_{1.4}$ is to maximize the expression $\sum_{k\in \underline{S}} y_k$, *i.e.*, the number of protected species. Indeed, according to constraints 1.4.1 and considering that we are seeking to maximize the quantity $\sum_{k\in \underline{S}} y_k$, variable y_k, which is a Boolean variable, necessarily takes the value 0 if $\sum_{i\in \underline{Z}_k} x_i = 0$, *i.e.*, if no zone of Z_k is selected, and the value 1, at the optimum of $\mathrm{P}_{1.4}$, if $\sum_{i\in \underline{Z}_k} x_i \geq 1$, *i.e.*, if at least one zone of Z_k is selected. Variable y_k therefore takes, as it should, at the optimum of $P_{1.4}$, the value 1 if and only if the zones selected for protection allow to protect species s_k. Note that constraints 1.4.4 could be replaced by constraints $y_k \leq 1,\ k \in \underline{S}$. The quantity $\sum_{i \in \underline{Z}} c_i x_i$ expresses the cost associated with the reserve and constraint 1.4.2, therefore, expresses the budgetary constraint. Note that if one wishes to obtain, among the optimal solutions of $P_{1.4}$, a lowest cost solution, one way is to solve program $P_{1.4}$ with the modified economic function, $\sum_{k \in \underline{S}} y_k - \varepsilon \sum_{i \in \underline{Z}} c_i x_i$, where ε is a sufficiently small constant. This technique can be applied in many cases when two criteria are considered, one in the economic function – here, the number of species – and the other in a constraint – here the cost.

Example 1.4. Let us take the instance described in figure 1.1 with $v_{ik} = 4$ for each couple (i, k) and assume that we have a budget of 8 units. The optimal use of this budget is provided by the resolution of $P_{1.4}$. It consists of protecting the zones z_1, z_2, z_6, z_8, z_{10}, and z_{18}, which protects 11 species, all the species except s_4, s_5, s_{12}, and s_{15}. The totality of the available budget is used.

Here again, it can be considered that the chances of survival of each species s_k are only really increased if β_k zones that contribute to this increase are protected. This problem can be modelled by a linear program in Boolean variables by replacing in $P_{1.4}$ constraints 1.4.1, $y_k \leq \sum_{i \in \underline{Z}_k} x_i$, $k \in \underline{S}$, by constraints $\beta_k y_k \leq \sum_{i \in \underline{Z}_k} x_i,\ k \in \underline{S}$. Thus, if the number of selected zones in the set Z_k, $\sum_{i \in \underline{Z}_k} x_i$, is less than β_k, the Boolean variable y_k can only take the value 0. Otherwise, and because of the fact that we are seeking to maximize $\sum_{k \in \underline{S}} y_k$, variable y_k takes the value 1 at the optimum. Note that, in this case, constraints 1.4.4 cannot be replaced by constraints $y_k \leq 1,\ k \in \underline{S}$. It can also be considered, as in section 1.2.1, that the zones are divided into q clusters. The problem of protecting a maximal number of species under a budgetary constraint can then be formulated as program $P_{1.5}$.

$$
P_{1.5}: \begin{cases} \max \sum_{k \in \underline{S}} y_k \\ \text{s.t.} \left| \begin{array}{lll} y_k \leq \sum_{i \in \underline{Z}_k} x_i & k \in \underline{S} & (1.5.1) \\ u_j \geq x_i & (j, i) \in \underline{\mathrm{Cl}} \times \underline{Z} : z_i \in \mathrm{Cl}_j & (1.5.2) \\ \sum_{i \in \underline{Z}} c_i x_i + \sum_{j \in \underline{\mathrm{Cl}}} d_j u_j \leq B & & (1.5.3) \\ x_i \in \{0, 1\} & i \in \underline{Z} & (1.5.4) \\ y_k \leq 1 & k \in \underline{S} & (1.5.5) \\ u_j \in \mathbb{R} & j \in \underline{\mathrm{Cl}} & (1.5.6) \end{array} \right. \end{cases}
$$

Due to constraints 1.5.2, the real variable u_j must take a value greater than or equal to 0 if none of the zones of Cl_j is selected for protection and a value greater than or equal to 1 in the opposite case. Considering all constraints of $P_{1.5}$, this implies that variable u_j can take the value 0 if none of the zones of Cl_j are selected for protection and the value 1 in the opposite case. Constraint 1.5.3 expresses that the cost associated with the reserve (see $P_{1.2}$) must not exceed the available budget, B. If one wishes to obtain, among the optimal solutions of $P_{1.5}$, a minimal cost solution, one way to do so is to solve program $P_{1.5}$ with the modified economic function, $\sum_{k\in\underline{S}} y_k - \varepsilon(\sum_{i\in\underline{Z}} c_i x_i + \sum_{j\in\underline{Cl}} d_j u_j)$, where ε is a sufficiently small constant.

Example 1.5. Let us consider the 20 zones in figure 1.1, with $v_{ik} = 4$ for each pair (i, k), and assume that these 20 zones are divided into 5 clusters, Cl_1, Cl_2, Cl_3, Cl_4, and Cl_5, as follows: $Cl_1 = \{z_1, z_2, z_5, z_6, z_7, z_{10}\}$, $Cl_2 = \{z_3, z_4, z_8, z_9\}$, $Cl_3 = \{z_{11}, z_{16}, z_{20}\}$, $Cl_4 = \{z_{14}, z_{15}, z_{18}, z_{19}\}$, and $Cl_5 = \{z_{12}, z_{13}, z_{17}\}$. Suppose, moreover, that the cost associated with each cluster is equal to 2 units. The optimal strategy to protect a maximal number of species with an available budget of 11 units is provided by the resolution of $P_{1.5}$. This strategy consists of protecting the 5 zones z_1, z_2, z_6, z_{10}, and z_{18}. These zones, distributed over the 2 clusters Cl_1 and Cl_4, make it possible to protect the 10 species s_1, s_2, s_3, s_6, s_7, s_8, s_9, s_{10}, s_{11}, and s_{14}. We present below, for illustration purposes, a way to solve this example using the AMPL modelling language and the CPLEX solver. Three files, named respectively "Example-1.5.mod", "Example-1.5.dat" and "Example-1.5.run", are used. The first corresponds to the translation of program $P_{1.5}$ into the AMPL language, the second describes the data in this example that are not already defined in "Example-1.5.mod", *i.e.*, c_i, n_{ik}, and a_{ij}, and the third is to start the resolution by CPLEX and display the solution obtained. The Boolean parameter a_{ij} describes the composition of each cluster: $a_{ij} = 1$ if and only if zone z_i belongs to cluster Cl_j.

1.3.2 The Number of Species Protected by a Reserve, R, is Assessed by Nb₂(R)

In this case, the basic problem of selecting a set of zones with a cost less than or equal to B and whose protection ensures the protection of a maximal number of species can be formulated as the linear program in Boolean variables $P_{1.6}$.

$$P_{1.6}: \begin{cases} \max \sum_{k\in\underline{S}} y_k \\ \text{s.t.} \left| \begin{array}{lll} \theta_k y_k \le \sum_{i\in\underline{Z}} n_{ik} x_i \quad k\in\underline{S} & (1.6.1) & \mid \quad x_i \in \{0,1\} \quad i\in\underline{Z} \quad (1.6.3) \\ \sum_{i\in\underline{Z}} c_i x_i \le B & (1.6.2) & \mid \quad y_k \in \{0,1\} \quad k\in\underline{S} \quad (1.6.4) \end{array} \right. \end{cases}$$

Let us examine constraints 1.6.1. There are two possibilities. Either $\sum_{i\in\underline{Z}} n_{ik} x_i < \theta_k$ and then the Boolean variable y_k can only take the value 0, or $\sum_{i\in\underline{Z}} n_{ik} x_i \ge \theta_k$ and then variable y_k takes the value 1 at the optimum of $P_{1.6}$ since we seek to maximize the expression $\sum_{k\in\underline{S}} y_k$. These constraints, therefore, reflect the fact that species s_k is protected if and only if the total population size of this species in the reserve is greater than or equal to θ_k. If one wishes to obtain, among the optimal solutions of $P_{1.6}$, a least-cost solution, one way to do this is to solve

Example-1.5.mod

```
#------Data-------------------------------------------------------------------------------------
param c{i in 1..20};
param a{i in 1..20, j in 1..5} default 0;
param n{i in 1..20, k in 1..15} default 0;
param d{j in 1..5}:=2;
param nu{i in 1..20, k in 1..15}:=4;
param B:=11;
#------Variables--------------------------------------------------------------------------------
var x{i in 1..20} binary;
var y{k in 1..15} <=1;
var u{j in 1..5} <= 1;
#------Model------------------------------------------------------------------------------------
maximize f: sum{k in 1..15} y[k];
subject to
C1{k in 1..15}: y[k]<=sum{i in 1..20 : n[i,k]>=nu[i,k]} x[i];              # (1.5.1)
C2{j in 1..5, i in 1..20 : a[i,j]=1}:    u[j]>=x[i];                        # (1.5.2)
C3: sum{i in 1..20} c[i]*x[i] + sum{j in 1..5} d[j]*u[j]<=B;               # (1.5.3)
#------Fin--------------------------------------------------------------------------------------
```

Example-1.5.dat

```
data;
#-----------------------------------------------------------------------------------------------
param c:=
1 2, 2 1, 3 4, 4 2, 5 4, 6 1, 7 2, 8 1, 9 4, 10 2, 11 4,
12 1, 13 2, 14 4, 15 1, 16 2, 17 4, 18 1, 19 2, 20 4;
#-----------------------------------------------------------------------------------------------
param a:=
1 1 1, 2 1 1, 3 2 1, 4 2 1, 5 1 1, 6 1 1, 7 1 1, 8 2 1, 9 2 1, 10 1 1, 11 3 1,
12 5 1, 13 5 1, 14 4 1, 15 4 1, 16 3 1, 17 5 1, 18 4 1, 19 4 1, 20 3 1;
#-----------------------------------------------------------------------------------------------
param n:=
1 2 1, 1 3 7, 2 1 8, 2 3 2, 2 6 8, 2 11 8, 3 3 8, 3 6 2, 4 6 2, 4 12 9, 5 6 5, 5 9 4, 6 6 4,
6 9 8, 6 11 8, 6 14 6, 7 11 8, 7 13 2, 8 13 9, 9 4 3, 9 13 6, 9 14 8, 10 6 2, 10 7 4, 10 8 8,
10 10 7, 11 7 2, 11 12 7, 12 11 8, 13 2 8, 13 11 3, 14 2 9, 14 5 7, 14 10 8, 15 10 8, 15 11 9,
16 4 7, 16 7 8, 16 11 2, 17 2 3, 17 9 5, 18 2 9, 18 11 4, 19 5 3, 19 8 7, 20 7 8, 20 8 2, 20 15 8;
#-----------------------------------------------------------------------------------------------
```

Example-1.5.run

```
reset;
option solver cplex1260;
model Example-1.5.mod;
model Example-1.5.dat;
solve;
```

```
print '**********************************************************';
print '            Solution Example 1.5:';
print '**********************************************************';
print ' Total cost of protection:', sum{i in 1..20} c[i]*x[i] + sum{j in 1..5} d[j]*u[j];
print '-------------------------------------------------------------';
print ' Indices of the zones to be protected:',{i in 1..20 : x[i]=1}i;
print '-------------------------------------------------------------';
print ' Number of protected species:', f;
print '-------------------------------------------------------------';
print ' Indices of protected species:', {k in 1..15 : y[k]=1}k;
print '**********************************************************';
```

To solve the considered example, it is sufficient to type "model Example-1.5.run;" in the console ampl: **ampl: model Example-1.5.run;** which starts the resolution using the CPLEX 12.6.0.0 solver and then displays the resulting solution:

```
CPLEX 12.6.0.0: optimal integer solution; objective 10
20 MIP simplex iterations
0 branch-and-bound nodes
*******************************************
            Solution Example 1.5:
*******************************************
Total cost of protection: 11
-------------------------------------------------------------
Indices of the zones to be protected: 1 2 6 10 18
-------------------------------------------------------------
Number of protected species: 10
-------------------------------------------------------------
Indices of protected species: 1 2 3 6 7 8 9 10 11 14
*******************************************
```

program $P_{1.6}$ with the modified economic function, $\sum_{k \in \underline{S}} y_k - \varepsilon \sum_{i \in \underline{Z}} c_i x_i$, where ε is a sufficiently small constant.

Example 1.6. Let us take again the instance described in figure 1.1, set θ_k to 7 for every k of $\underline{S}$ and assume that we have a budget of 8 units. An optimal use of this budget, when the number of species protected by a reserve, R, is evaluated by $\text{Nb}_2(R)$, is provided by the resolution of $P_{1.6}$. It consists in protecting the 6 zones z_2, z_6, z_8, z_{10}, z_{16}, and z_{18}, which allows the protection of 10 species, all the species except s_3, s_5, s_{12}, s_{14}, and s_{15}.

1.3.3 Remarks on the Problems Addressed in Sections 1.3.1 and 1.3.2

In all the problems addressed in sections 1.3.1 and 1.3.2, it is possible to give a different importance to the protection of each species by replacing in the corresponding mathematical programs the economic function $\sum_{k \in \underline{S}} y_k$ with the economic function $\sum_{k \in \underline{S}} w_k y_k$ where w_k represents the weight assigned to the species s_k. These weights reflect the relative importance of the different species considered. It should also be noted that, for all these problems, a decision-maker may be interested in knowing their optimal solution for different values of the available budget, B. In this way, he/she can easily assess the marginal effect of an additional investment. This can be done by solving the corresponding mathematical programs with different values of B. It is also possible to look, almost equivalently, at the minimal budget needed to achieve a certain level of species protection. Let us consider, for example, the case where the number of species protected by a reserve, R, is assessed by $\mathrm{Nb}_1(R)$. To know, in this case, the budget necessary to protect, at least, Ns species for all possible values of Ns, it is sufficient to solve program $\mathrm{P}_{1.7}$ by varying Ns from 1 to m.

$$\mathrm{P}_{1.7}: \begin{cases} \min \sum\limits_{i \in \underline{Z}} c_i x_i \\ \text{s.t.} \left| \begin{array}{llll} y_k \leq \sum\limits_{i \in \underline{Z}_k} x_i & k \in \underline{S} & (1.7.1) & \mid \quad x_i \in \{0,1\} \quad i \in \underline{Z} \quad (1.7.3) \\ \sum\limits_{k \in \underline{S}} y_k \geq \mathrm{Ns} & & (1.7.2) & \mid \quad y_k \in \{0,1\} \quad k \in \underline{S} \quad (1.7.4) \end{array} \right. \end{cases}$$

Constraint 1.7.2 states that the number of species protected by the reserve must be greater than or equal to Ns. It should be noted that the number of species actually protected may be greater than $\sum_{k \in \underline{S}} y_k$. It is in fact equal to the cardinal of the set $\{k : \sum_{i \in \underline{Z}_k} x_i \geq 1\}$. If we wish to solve the problem under consideration while maximizing the number of protected species, we can deduct from the economic function of $\mathrm{P}_{1.7}$ the quantity $\varepsilon \sum_{k \in \underline{S}} y_k$ where ε is a sufficiently small constant.

Example 1.7. Let us again take the instance described in figure 1.1 assuming that the number of species protected by a reserve, R, is assessed by $\mathrm{Nb}_1(R)$ and that $v_{ik} = 4$ for each pair (i,k). Table 1.1 gives the optimal solution of $\mathrm{P}_{1.7}$ – after subtracting $\varepsilon \sum_{k \in \underline{S}} y_k$ to the economic function – for all possible values of Ns. Figure 1.2 shows the curve illustrating the minimal cost of a reserve as a function of the number of species to be protected.

1.4 Gradual Establishment of a Reserve Over Time to Protect a Maximal Number of Species of a Given Set, with a Time-dependent Budget Constraint

As previously $Z = \{z_1, z_2, \ldots, z_n\}$ designates the set of candidate zones but now the protection of the zones of Z is done gradually over a time horizon, T, composed of r periods (r years for example), $T_1, T_2, \ldots, T_r$, in order to spread the costs. However, all the protection decisions are taken at the beginning of the horizon considered. In addition, any zone protected from a certain period remains protected for all the subsequent periods in the time horizon considered. Let $\underline{T} = \{1, 2, \ldots, r\}$. The set of

TAB. 1.1 – Resolution of program $P_{1.7}$ for the instance described in figure 1.1. Presentation of the best strategy to adopt and its cost, taking into account the number of species to be protected.

Minimal number of species to be protected (Ns)	Set of zones to be protected, of minimal cost	Cost	Number of species that are actually protected	Protected species
1	z_6	1	4	s_6 s_9 s_{11} s_{14}
2	z_6	1	4	s_6 s_9 s_{11} s_{14}
3	z_6	1	4	s_6 s_9 s_{11} s_{14}
4	z_6	1	4	s_6 s_9 s_{11} s_{14}
5	z_2 z_6	2	5	s_1 s_6 s_9 s_{11} s_{14}
6	z_6 z_{10}	3	7	s_6 s_7 s_8 s_9 s_{10} s_{11} s_{14}
7	z_6 z_{10}	3	7	s_6 s_7 s_8 s_9 s_{10} s_{11} s_{14}
8	z_6 z_{10} z_{18}	4	8	s_2 s_6 s_7 s_8 s_9 s_{10} s_{11} s_{14}
9	z_6 z_8 z_{10} z_{18}	5	9	s_2 s_6 s_7 s_8 s_9 s_{10} s_{11} s_{13} s_{14}
10	z_2 z_6 z_8 z_{10} z_{18}	6	10	s_1 s_2 s_6 s_7 s_8 s_9 s_{10} s_{11} s_{13} s_{14}
11	z_2 z_6 z_8 z_{10} z_{16} z_{18}	8	11	s_1 s_2 s_4 s_6 s_7 s_8 s_9 s_{10} s_{11} s_{13} s_{14}
12	z_1 z_2 z_6 z_8 z_{10} z_{16} z_{18}	10	12	s_1 s_2 s_3 s_4 s_6 s_7 s_8 s_9 s_{10} s_{11} s_{13} s_{14}
13	z_1 z_2 z_4 z_6 z_8 z_{10} z_{16} z_{18}	12	13	s_1 s_2 s_3 s_4 s_6 s_7 s_8 s_9 s_{10} s_{11} s_{12} s_{13} s_{14}
14	z_1 z_2 z_4 z_6 z_8 z_{10} z_{14} z_{16}	15	14	s_1 s_2 s_3 s_4 s_5 s_6 s_7 s_8 s_9 s_{10} s_{11} s_{12} s_{13} s_{14}
15	z_1 z_2 z_4 z_6 z_8 z_{10} z_{14} z_{16} z_{20}	19	15	s_1 s_2 s_3 s_4 s_5 s_6 s_7 s_8 s_9 s_{10} s_{11} s_{12} s_{13} s_{14} s_{15}

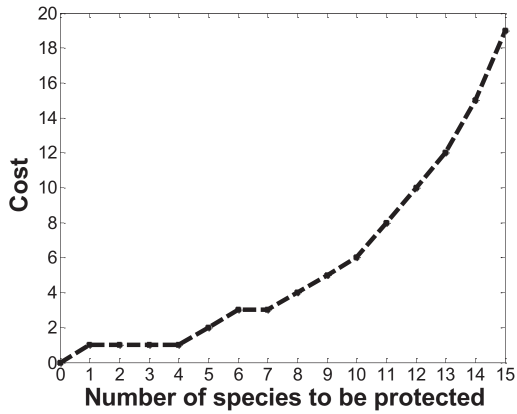


Fig. 1.2 – Curve associated with table 1.1: cost of the cheapest strategy according to the number of species to be protected.

species likely to be protected by a zone depends on the time at which that zone is protected. Indeed, the possible evolution of the environment between two periods may change the role of the different zones in protecting species. For example, a zone protected since the period T_t allows the protection of a certain set of species, but if this zone is protected only from the period $T_{t+\tau}$ it no longer allows the protection of all the species of this set. Typically, for each pair $(z_i,\ T_t)$ we know the set of species that are protected until the end of the time horizon if we protect z_i from the beginning of the period T_t. We denote by Z_{kt} the set of zones which, if they are protected from the beginning of the period T_t, protect species s_k. We denote by $\underline{Z}_{kt}$ the set of corresponding indices. The objective is to determine the zones to be protected from the beginning of each period and under a period-specific budgetary constraint so that a maximal number of species are protected at the end of the r periods. The cost of protecting the zones can vary over time. Thus the cost related to the decision to protect zone z_i at the beginning of the period T_t is denoted by c_{it}, $i \in \underline{Z}$, $t \in \underline{T}$, and this cost must be borne at the beginning of the period T_t. We know the available budget, B_t, at the beginning of the period T_t. In the simple model we consider, it is assumed that all the data are known at the beginning of the horizon T and do not change during this horizon. The problem can be formulated as a mathematical program. To do this, with each zone z_i and each period T_t is associated a Boolean variable, x_{it}, which, by convention, is equal to 1 if and only if we decide to protect zone z_i from the beginning of period T_t. As in the previous models, with each species s_k is associated a Boolean variable, y_k, which is equal to 1 if and only if species s_k is protected. This results in program $\mathrm{P}_{1.8}$.

$$\mathrm{P}_{1.8}: \begin{cases} \max \sum_{k \in \underline{S}} y_k \\ \text{s.t.} \left| \begin{array}{llll} y_k \leq \sum_{t \in \underline{T}} \sum_{i \in \underline{Z}_{kt}} x_{it} & k \in \underline{S} & (1.8.1) & | \quad x_{it} \in \{0,1\} \quad i \in \underline{Z}, t \in \underline{T} \quad (1.8.4) \\ \sum_{t \in \underline{T}} x_{it} \leq 1 & i \in \underline{Z} & (1.8.2) & | \quad y_k \in \{0,1\} \quad k \in \underline{S} \quad (1.8.5) \\ \sum_{i \in \underline{Z}} c_{it} x_{it} \leq B_t & t \in \underline{T} & (1.8.3) & | \end{array} \right. \end{cases}$$

The economic function of $\mathrm{P}_{1.8}$ expresses the number of protected species. Because of constraints 1.8.1 and the fact that we are seeking to maximize the expression $\sum_{k \in \underline{S}} y_k$, variable y_k takes the value 1, at the optimum of $\mathrm{P}_{1.8}$, if and only if at least one of variables x_{it}, $t \in \underline{T}$, $i \in \underline{Z}_{kt}$, is equal to 1, in other words, if and only if there is at least one period T_t, at the beginning of which at least one zone of Z_{kt} is protected. Constraints 1.8.2 express that any zone can only be protected from a single period of the horizon. Constraints 1.8.3 correspond to period-specific budgetary constraints. They express that the budget allocated to the protection of the zones at the beginning of each period T_t should not exceed the available budget, B_t. In this model, the resources not used in the period T_t are lost. If this does not correspond to reality, constraints 1.8.3 can be replaced by the set of constraints $\mathrm{C}_{1.1}$. In this case, the resources available but not used in the period T_t can be used from period T_{t+1}.

$$\mathrm{C}_{1.1}: \begin{cases} \sum_{i \in \underline{Z}} c_{i1} x_{i1} + \delta_1 = B_1 & \\ \sum_{i \in \underline{Z}} c_{it} x_{it} + \delta_t = B_t + \delta_{t-1} & t \in \underline{T}, t \geq 2 \\ \delta_t \geq 0 & t \in \underline{T} \end{cases}$$

Variable δ_t, $t \in \underline{T}$, corresponds to the quantity of unused resources at the beginning of period T_t. The first constraint expresses, on the one hand, that the expenses at the beginning of period T_1 must not exceed the available budget, B_1, and, on the other hand, that variable δ_1 is equal to the amount of unused resources, *i.e.*, the quantity $B_1 - \sum_{i \in \underline{Z}} c_{i1} x_{i1}$. The following set of constraints expresses, on the one hand, that the expenses at the beginning of period T_t, $t \in \underline{T}, t \geq 2$, must not exceed the budget available at the beginning of this period, *i.e.*, $B_t + \delta_{t\text{-}1}$ and, on the other hand, that variable δ_t is equal to the amount of resources not used at the beginning of period T_t, *i.e.*, the quantity $B_t + \delta_{t-1} - \sum_{i \in \underline{Z}} c_{it} x_{it}$. In other words, variable δ_t corresponds to the resources not yet used up to period T_t, including period T_t, *i.e.*, the quantity $B_1 + B_2 + \cdots + B_t$ minus the quantity $\sum_{i \in \underline{Z}} c_{i1} x_{i1} + \sum_{i \in \underline{Z}} c_{i2} x_{i2} + \cdots + \sum_{i \in \underline{Z}} c_{it} x_{it}$.

1.5 Reserve Necessarily Including Certain Zones

All the problems discussed in this chapter consist in selecting an "optimal" set of zones to be protected. It may be that, for different reasons, some of the candidate zones must be selected.

1.5.1 Selection of a Reserve Taking into Account Already Protected Zones

It may be that in the problems studied in sections 1.2 and 1.3, some of the zones that could form the reserve are already protected zones. They have acquired this status in the past and still have it when the time comes to establish a new optimal reserve. To take this constraint into account, it is sufficient to solve the mathematical program corresponding to the problem under study by setting variables x_i to 1 for zone z_i whose protection is mandatory. One way of doing this is to add to this program constraint $x_i = 1$ for all the indices i concerned. Remember that in all the mathematical programs considered in these two sections, the Boolean variable x_i takes the value 1 if and only if zone z_i is selected to form the reserve.

Example 1.8. Let us look again at the instance described in figure 1.1 and consider the problem of determining a reserve, R, which respects a budgetary constraint and maximizes $\text{Nb}_1(R)$. Suppose, as in example 1.4, that $v_{ik} = 4$ for any couple (i, k) and that we have a budget of 8 units. If the protection of zone z_5 is mandatory, an optimal use of this budget consists in protecting the 4 zones z_2, z_5, z_6, and z_{10}, which allows the 8 species s_1, s_6, s_7, s_8, s_9, s_{10}, s_{11}, and s_{14} to be protected. If the protection of zone z_{11} is mandatory, an optimal use of this budget consists in protecting the 4 zones z_2, z_6, z_{10}, and z_{11}, which allows the 9 species s_1, s_6, s_7, s_8, s_9, s_{10}, s_{11}, s_{12}, and s_{14} to be protected. Remember that if there are no zones whose protection is mandatory, an optimal use of a budget of 8 units is to protect the 6 zones z_1, z_2, z_6, z_8, z_{10}, and z_{18}, which allows all species to be protected except s_4, s_5, s_{12}, and s_{15} (see example 1.4).

1.5.2 Gradual Establishment of a Reserve Over Time: Review of Decisions Taken at the Beginning of the Considered Horizon

Let us return to the constitution of a reserve discussed in section 1.4. A disadvantage of the considered model is that the decisions are made, definitively, at the beginning of the time horizon even if some zones are effectively protected only from a certain period of this horizon. We discuss below the possibility of revising these decisions over time, taking into account possible changes in projected costs, budgets and capacities of zones to protect species. Suppose that in the solution to the problem in section 1.4, the list of zones to be protected at each period is defined by $x_{it} = \tilde{x}_{it}$, $i \in \underline{Z}$, $t \in \underline{T}$ ($\tilde{x}_{it}$ is a constant which is 0 or 1). Let us also assume that we have arrived at the beginning of the period T_j and that the forecasts for the coming periods are reviewed, including the available budget. Thus Z_{kt} becomes $\hat{Z}_{kt}$, $k \in \underline{S}, t \in \underline{T}, t \geq j$, B_t becomes $\hat{B}_t$, $t \in \underline{T}, t \geq j$, and c_{it} becomes $\hat{c}_{it}$, $i \in \underline{Z} - \underline{R}_j$, $t \in \underline{T}, t \geq j$, where R_j is the reserve already constituted and $\underline{R}_j$ the set of corresponding indices. The decisions that had been taken at the beginning of the horizon considered for the periods $T_j, T_{j+1}, \ldots, T_r$ can be abandoned and new optimal decisions can be sought, taking into account not only the new forecasts but also the reserve already constituted and the species that it allows to be protected. We are, therefore, in the case of establishing a reserve which must necessarily include certain zones. One could also consider abandoning some zones, but this is not considered here (see chapter 12, section 12.3.3.2). One way to formulate the problem is to slightly modify program $\mathrm{P}_{1.8}$: Constraints 1.8.3 are now to be taken into account only for $t \geq j$ and constraints 1.9.4 which stipulate that some zones have already been selected must be added. It is also necessary to set $\hat{Z}_{kt} = Z_{kt},\ k \in \underline{S}, t \in \underline{T}, t \leq j-1$ to take into account the species already protected by R_j. This gives program $\mathrm{P}_{1.9}$.

$$
\mathrm{P}_{1.9}: \begin{cases} \max \sum_{k \in \underline{S}} y_k & \\ \text{s.t.} \left| \begin{array}{lll} y_k \leq \sum_{t \in \underline{T}} \sum_{i \in \underline{\hat{Z}}_{kt}} x_{it} & k \in \underline{S} & (1.9.1) \\ \sum_{t \in \underline{T}} x_{it} \leq 1 & i \in \underline{Z} & (1.9.2) \\ \sum_{i \in \underline{Z}} \hat{c}_{it} x_{it} \leq \hat{B}_t & t \in \underline{T}, t \geq j & (1.9.3) \\ x_{it} = \tilde{x}_{it} & i \in \underline{Z}, t \in \underline{T}, t \leq j-1 & (1.9.4) \\ x_{it} \in \{0,1\} & i \in \underline{Z}, t \in \underline{T} & (1.9.5) \\ y_k \in \{0,1\} & k \in \underline{S} & (1.9.6) \end{array} \right. \end{cases}
$$

1.6 Case Where Several Conservation Actions are Conceivable in Each Zone

1.6.1 The Problem

This section considers a slightly different case from the ones studied in the previous sections insofar as several different protection actions are possible for each zone. Thus, for each candidate zone, a decision can be made to protect it or not to protect it, but if it is decided to protect it, several protection actions can be considered. We present below an example of reserve selection relevant to this issue and developed drawing on the references (Cattarino *et al.*, 2015; Salgado-Rojas *et al.*, 2020). In this example it is considered that a species present in a given zone is exposed to different threats and that, if the protection of that zone is decided, different actions can be taken to remove all or part of these threats. An optimal reserve is a reserve that maximizes an "overall ecological benefit" for the species considered within an available budget. This benefit takes into account both protected zones and actions taken in these zones to eliminate certain threats.

Let $S = \{s_1, s_2, \ldots, s_m\}$ be the set of species, animal or plant, in which we are interested and $Z = \{z_1, z_2, \ldots, z_n\}$ be the set of zones that we can decide whether or not to protect. S_i refers to the set of species present in zone z_i. In addition, there are a number of threats, $M = \{\mu_1, \mu_2, \ldots, \mu_g\}$, affecting these species. We denote by M_{ik} ($\subseteq M$) the set of threats affecting species s_k in zone z_i and M_i the set of threats to be considered in zone z_i. We have thus $M_i = \cup_{k:\, s_k \in S_i} M_{ik}$. The protection of zone z_i costs c_i and the elimination of threat μ_j in zone z_i costs d_{ij}. As we have said, the protection strategy has two levels. It is defined by the set of zones that it has decided to protect and, for each of these zones, by the set of threats that it has decided to eliminate. For a species s_k living in zone z_i of the reserve, we consider that the degree of protection of s_k in z_i is equal to the ratio between the number of eliminated threats weighing on s_k in z_i and the total number of threats weighing on s_k in z_i. The degree of protection of a species s_k present in a protected zone z_i where it is not threatened is equal to 1 for that zone. The degree of protection of a species s_k in an unprotected zone z_i is equal to 0. We denote by w_{ik} the square of the degree of protection of species s_k in zone z_i. A we have seen, the value of this variable results from the strategy adopted for zone z_i: not protecting it or protecting it and eliminating a number of threats. The problem is to determine the optimal strategy given the available budget. The value of a strategy is measured by the sum of the squares of the degrees of protection, w_{ik}, for all pairs $(z_i,\ s_k)$ where z_i is a candidate zone and s_k is a species present in that zone. We denote, respectively, by $\underline{S}$, $\underline{S}_i$, $\underline{Z}$, $\underline{M}$, $\underline{M}_i$, and $\underline{M}_{ik}$ the set of indices of the sets S, S_i, Z, M, M_i, and M_{ik}.

1.6.2 Mathematical Programming Formulation

We use the Boolean variables x_i, $i \in \underline{Z}$, which take the value 1 if and only if zone z_i is selected to be part of the reserve and the Boolean variables y_{ij}, $i \in \underline{Z}, j \in \underline{M}_i$, which take the value 1 if and only if we decide to eliminate the threat μ_j from zone z_i. The problem considered can be formulated as program $\mathrm{P}_{1.10}$.

$$
\mathrm{P}_{1.10}: \begin{cases}
\max \sum\limits_{i \in \underline{Z}} \sum\limits_{k \in S_i} w_{ik} \\
\text{s.t.} \left|\begin{array}{lll}
\sum\limits_{i \in \underline{Z}} c_i x_i + \sum\limits_{i \in \underline{Z},\, j \in \underline{M}_i} d_{ij} y_{ij} \le B & & (1.10.1) \\
w_{ik} \le \Big(\sum\limits_{j \in \underline{M}_{ik}} y_{ij} / |M_{ik}|\Big)^2 & i \in \underline{Z}, k \in \underline{S}_i, |M_{ik}| > 0 & (1.10.2) \\
w_{ik} \le x_i & i \in \underline{Z}, k \in \underline{S}_i & (1.10.3) \\
x_i \in \{0,1\} & i \in \underline{Z} & (1.10.4) \\
y_{ij} \in \{0,1\} & i \in \underline{Z}, j \in \underline{M}_i & (1.10.5) \\
w_{ik} \in \mathbb{R} & i \in \underline{Z}, k \in \underline{S}_i & (1.10.6)
\end{array}\right.
\end{cases}
$$

Since variable w_{ik} represents the square of the degree of protection of species s_k in zone z_i, the economic function represents the sum, for all pairs $(z_i,\ s_k)$ where s_k is a species present in zone z_i, of the square of the degree of protection of species s_k in zone z_i. If zone z_i is not selected – $x_i = 0$ – then, due to constraints 1.10.3, $w_{ik} = 0$ for all the species living in this zone. If zone z_i is selected – $x_i = 1$ – then two cases are possible: (1) species s_k is not threatened in this zone – $|M_{ik}| = 0$ – and $w_{ik} = 1$ because of constraints 1.10.3 and the economic function to be maximized, (2) species s_k is threatened in this zone – $|M_{ik}| > 0$ – and because of constraints 1.10.2 and the economic function to be maximized $w_{ik} = \left(\sum_{j\in \underline{M}_{ik}} y_{ij}/|M_{ik}|\right)^2$. The economic function, therefore, expresses well the sum of the squares of the degrees of protection, w_{ik}, for all pairs $(z_i,\ s_k)$ where z_i is a protected zone and s_k, a species present in this zone. Constraint 1.10.1 is the budget constraint and constraints 1.10.4 and 1.10.5 specify the Boolean nature of variables x_i and y_{ij}.

1.6.3 Example

Consider the instance described in figure 1.3 (20 zones and 15 species). The optimal protection strategies are given in table 1.2 when the available budget is 25 units.

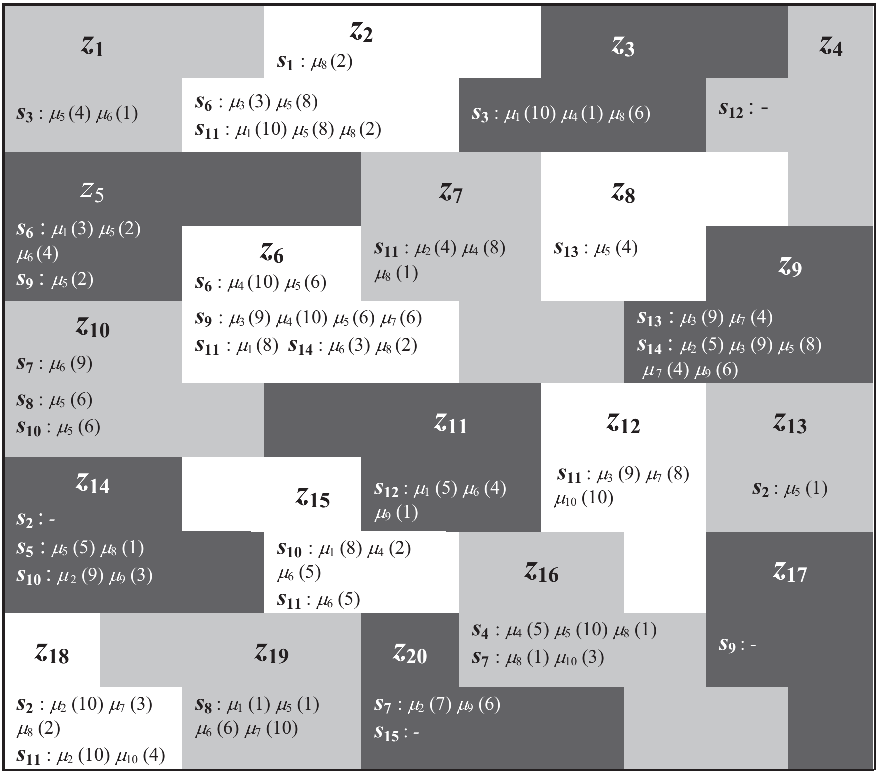


FIG. 1.3 – Twenty zones, z_1, z_2,…, z_{20}, are candidates for protection and fifteen species, s_1, s_2, …, s_{15}, living in these zones are concerned. For each zone, the species present and the threats associated, with their removal costs in brackets are indicated. The cost of protecting the white zones is equal to 1 unit, the cost of protecting the light grey zones is equal to 2 units and the cost of protecting the dark grey zones is equal to 4 units. For example, species s_7 and s_{15} are present in zone z_{20}, threats μ_2 and μ_9 affect species s_7 in this zone and there are no threats to species s_{15}. The cost of protecting this zone is equal to 4 units and the cost of removing threats μ_2 and μ_9 in this zone is equal to 7 and 6 units, respectively.

Tab. 1.2 – Optimal protection strategies for the instance described in figure 1.3 when the available budget is 25 units.

Protected zone	Protection cost of the zone	Species present in the zone	Threats associated to the couple (zone, species)	Threats removed	Total cost of removal threats in the zone	Square of the degree of protection of the couple (zone, species)
z_2	1	s_1	μ_8	μ_8	2	1
		s_6	μ_3 μ_5	–		0
		s_{11}	μ_1 μ_5 μ_8	μ_8		0.11
z_4	2	s_{12}	–	–	0	1
z_{10}	2	s_7	μ_6	–	6	0
		s_8	μ_5	μ_5		1
		s_{10}	μ_5	μ_5		1
z_{13}	2	s_2	μ_5	μ_5	1	1
z_{14}	4	s_2	–	–	1	1
		s_5	μ_5 μ_8	μ_8		0.25
		s_{10}	μ_2 μ_9	–		0
z_{17}	4	s_9	–	–	0	1
Total	15				10	7.36

References and Further Reading

Arponen A., Heikkinen R.K., Thomas C.D., Moilanen A. (2005) The value of biodiversity in reserve selection: representation, species weighting, and benefit functions, *Conserv. Bio.* **19**, 2009.

Arthur J.L., Hachey M., Sahr K., Huso M., Kiester A.R. (1997) Finding all optimal solutions to the reserve site selection problem: formulation and computational analysis, *Environ. Ecol. Stat.* **4**, 153.

Cabeza M., Moilanen A. (2001) Design of reserve networks and the persistence of biodiversity, *Trends Ecol. Evol.* **16**, 242.

Camm J.D., Polasky S., Solow A.R., Csuti A.R. (1996) A note on optimal algorithms for reserve site selection, *Biol. Conserv.* **78**, 353.

Cattarino L., Hermoso V., Carwardine J., Kennard M.J., Linke S. (2015) Multi-action planning for threat management: a novel approach for the spatial prioritization of conservation actions, *PLoS ONE* **10**, e0128027.

Church R.L., Stoms D.M., Davis F.W. (1996) Reserve selection as a maximal covering location problem, *Biol. Conserv.* **76**, 105.

Clemens M.A., ReVelle C.S., Williams J.C. (1999) Reserve design for species preservation, *Eur. J. Oper. Res.* **112**, 273.

Cocks K.D., Baird I.A. (1989) Using mathematical programming to address the multiple reserve selection problem: an example from the Eyre Peninsula, South Australia, *Biol. Conserv.* **49**, 113.

Csuti B., Polasky S., Williams P., Pressey R., Camm J., Kershaw M., Kiester R., Downs B., Hamilton R., Huso M., Sahr K. (1997) A comparison of reserve selection algorithms using data on terrestrial vertebrates in Oregon, *Biol. Conserv.* **80**, 83.

Fiorella K., Cameron A., Sechrest W., Winfree R., Kremen C. (2010) Methodological considerations in reserve system selection: a case study of Malagasy lemurs, *Biol. Conserv.* **143**, 963.

Fischer D.T., Church R.L. (2005) The SITES reserve selection system: a critical review, *Environ. Model. Assess.* **10**, 215.

Haight R.G, Travis L.E. (2008) Reserve design to maximize species persistence, *Environ. Model. Assess.* **13**, 243.

Hamaide B., ReVelle C.S., Malcolm S.A. (2006) Biological reserves, rare species and the trade-off between species abundance and species diversity, *Ecol. Econ.* **56**, 570.

Hamaide B., Williams J.C., ReVelle C.S. (2009) Cost-efficient reserve site selection favoring persistence of threatened and endangered species, *Geog. Anal.* **41**, 66.

Hamaide B., Sheerin J. (2011) Species protection from current reserves: economic and biological considerations, spatial issues and policy evaluation, *Ecol. Econ.* **70**, 667.

Hamaide B., Albers H.J., Busby G. (2014) Backup coverage models in nature reserve site selection with spatial spread risk heterogeneity, *Socio-Econ. Planning Sci.* **48**, 158.

Jafari N., Nuse B.L., Moore C.T., Dilkina B., Hepinstall-Cymerman J. (2017) Achieving full connectivity of sites in the multiperiod reserve network design problem, *Comp. Oper. Res.* **81**, 119.

Juutinen A., Mönkkönen M. (2007) Alternative targets and economics efficiency of selecting protected areas for biodiversity conservation in boreal forest, *Environ. Res. Econ.* **37**, 713.

Kingsland S.E. (2002) Creating a science of nature reserve design: perspectives from history, *Environ. Model. Assess.* **7**, 61.

Knight A.T., Cowling R.M., Rouget M., Balmford A., Lombard A.T., Campbell B.M. (2008) Knowing but not doing: selecting priority conservation areas and the research-implementation gap, *Conserv. Biol.* **22**, 610.

Malcolm S.A., ReVelle C.R. (2005) Models for preserving species diversity with backup coverage, *Environ. Model. Assess.* **10**, 99.

Margules C.R., Usher M.B. (1981) Criteria used in assessing wildlife conservation potential: a review, *Biol. Conserv.* **21**, 79.

Margules C.R., Nicholls A., Pressey R. (1988) Selecting networks of reserves to maximise biological diversity, *Biol. Conserv.* **43**, 63.

Memtsas D.P. (2003) Multiobjective programming methods in the reserve selection problem, *Eur. J. Oper. Res.* **150**, 640.

Moilanen A., Wilson K.A., Possingham H.P., Eds (2009) *Spatial conservation prioritization.* Oxford University Press.

Moore J.L., Folkmann M., Balmford A., Brooks T., Burgess N., Rahbek C., Williams P.H., Krarup J. (2003) Heuristic and optimal solutions for set-covering problems in conservation biology, *Ecography* **26**, 595.

Önal H. (2003) First-best, second-best, and heuristic solutions in conservation reserve site selection, *Biol. Conserv.* **115**, 55.

Polasky S., Camm J.D., Garber-Yonts B. (2001) Selecting biological reserves cost-effectively: an application to terrestrial vertebrate conservation in Oregon, *Land Econ* **77**, 68.

Possingham H.P., Day J., Goldfinch M., Salzborn F. (1993) The mathematics of designing a network of protected areas for conservation, *12th Australian Operations Research Conference* (D. Sutton, F. Cousins, C. Pierce, Eds). University of Adelaïde, pp. 536–545.

Possingham H.P., Ball I., Andelman S. (2000) Mathematical methods for identifying representative reserve networks, *Quantitative methods for conservation biology* (S. Ferson, M. Burgman, Eds), Springer-Verlag, pp. 291–305.

Pressey R.L., Possingham H.P., Margules C.R. (1996) Optimality in reserve selection algorithms: when does it matter and how much? *Biol. Conserv.* **76**, 259.

Pressey R.L., Possingham H.P., Day J.R. (1997) Effectiveness of alternative heuristic algorithms for identifying indicative minimum requirements for conservation reserves, *Biol. Conserv.* **80**, 207.

Pressey R.L., Possingham H.P., Logan V.S., Day J.R., Williams P.H. (1999) Effects of data characteristics on the results ofreserve selection algorithms, *J. Biogeography* **26**, 179.

Pressey R.L., Cowling R.M. (2001) Reserve selection algorithms and the real world, *Conserv. Biol.* **15**, 275.

ReVelle C.S., Williams J.C., Boland J.J. (2002) Counterpart models in facility location science and reserve selection science, *Environ. Model. Assess.* **7**, 71.

Rodrigues A.S., Cerdeira J.O., Gaston K.J. (2000) Flexibility, efficiency, and accountability: adaptative reserve selection algorithms to more complex conservation problems, *Ecography* **23**, 565.

Rodrigues A.S.L., Gaston K.J. (2002) Optimisation in reserve selection procedures – why not? *Biol. Conserv.* **107**, 123.

Rosing K.E., ReVelle C.S., Williams J.C. (2002) Maximizing species representation under limited resources: a new and efficient heuristic, *Environ. Model. Assess.* **7**, 91.

Salgado-Rojas J., Álvarez-Miranda E., Hermoso V., Garcia-Gonzalo J., Weintraub A. (2020) A mixed integer programming approach for multi-action planning for threat management, *Ecol. Model.* **418**, 108901.

Snyder S.A., Haight R.G. (2016) Application of the maximal covering location problem to habitat reserve site selection: a review, *Int. Reg. Sci. Rev.* **39**, 28.

Strange N., Thorsen B.J., Bladt J. (2006) Optimal reserve selection in a dynamic world, *Biol. Conserv.* **131**, 33.

Tanskanen A. (2006) On optimization in reserve selection. Thesis, University of Jyväskylä, p. 74.

Underhill L.G. (1994) Optimal and suboptimal reserve selection algorithms, *Biol. Conserv.* **70**, 85.

Vanderkam R.P.D., Wiersma Y.F., King D.J. (2007) Heuristic algorithms vs. linear programs for designing efficient conservation reserve networks: evaluation of solution optimality and processing time, *Biol. Conserv.* **137**, 349.

Van Teeffelen A.J.A, Moilanen A. (2008) Where and how to manage: optimal selection of conservation actions for multiple species, *Biodivers. Info.* **5**, 1.

Williams J.C., ReVelle C.S., Levin S.A. (1997) Applying mathematical programming to reserve selection, *Environ. Model. Assess.* **2**, 167.

Williams J.C., ReVelle C.S. (1998) Reserve assemblage of critical areas: a zero-one programming approach, *Eur. J. Oper. Res.* **104**, 497.

Chapter 2

Fragmentation

2.1 Introduction

The spatial configuration of a nature reserve plays an important role in the survival of the species that live there. In this chapter, we are interested in the fragmentation of a reserve, *i.e.*, the dispersion of the patches – or zones – that compose it, in relation to each other (see figure 2.1). This phenomenon, which is often associated with the decrease in the area of various patches, is considered to be one of the main causes of biodiversity loss. The fragmentation of a reserve is indeed one of the main factors preventing species from moving around the reserve as they should and could in a non-fragmented one. This habitat fragmentation, therefore, significantly increases the extinction risk of many species. It can be natural but more often results from a fragmentation of the space due to artificial phenomena such as the presence of urbanized zones, intensive agricultural zones or transport infrastructures. It should be noted that species are affected differently by habitat fragmentation. A reserve may appear to be very fragmented for some species, those that will have great difficulty moving from one patch to another, and not very fragmented for others, those that, despite some distance between patches, will still be able to travel most of these patches due, for example, to their ability to fly or cross obstacles such as roads or zones treated with pesticides. Fragmentation is also a handicap in terms of species' adaptation to climate change. It should be noted, however, that the ease of movement of species within a reserve is not always without its drawbacks: it can increase the risk of disease transmission between wildlife species in the reserve and also the transmission of these diseases to domestic species. It can also facilitate the proliferation of invasive species, a phenomenon currently considered to be one of the major causes of biodiversity loss. There has been much debate about the desirable size of protected zones: is it more interesting to have a single large protected zone or several small ones with the same total size – SLOSS: Single Large Or Several Small. This debate focuses mainly on ecological aspects, but it is worth noting that the

DOI: 10.1051/978-2-7598-2521-9.c002

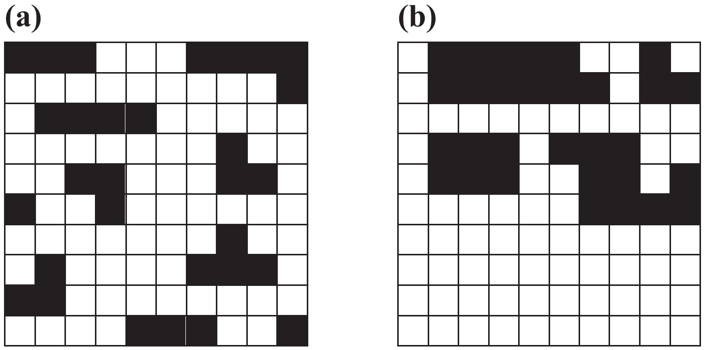


Fig. 2.1 – A hypothetical landscape represented by a grid of square and identical cells. Two reserves – in black – with a total area of 30 units. (a) A highly fragmented reserve. (b) A less fragmented reserve.

management of a fragmented set of zones is generally more difficult and costly than the management of a non-fragmented set.

Given a set of zones spread over a territory and such that any two zones have no common parts, many indicators of fragmentation can be associated with this set. We will examine, for example, the following indicators: the Mean Nearest Neighbour Distance (MNND), the Mean Shape Index (MSI), and the Mean Proximity Index (MPI).

The problem related to the notion of fragmentation, which naturally arises in the presence of a set of zones – without common parts – that can be protected, consists in selecting, among these zones and under certain constraints, a subset of zones to be protected that is optimal with regard to these indicators or that respects some of values of them.

2.2 The Indicators MNND (Mean Nearest Neighbour Distance), MSI (Mean Shape Index) and MPI (Mean Proximity Index)

First, let us look at the MNND indicator associated with a reserve, R, *i.e.*, a subset of zones of $Z = \{z_1, z_2, \dots, z_n\}$. Let us denote by d_{ij} the distance between zones z_i and z_j. Here, it is the straight line distance between the two zones. More precisely, d_{ij} is defined as the shortest distance that can be found between a point in zone z_i and a point in zone z_j. The distance between two zones could very well be defined differently, taking into account, for example, the difficulty for the species under consideration to move from one zone to another. One could thus take into account the obstacles to be overcome or the inhospitable nature of the areas to be crossed, *i.e.*, the surrounding matrix and not only the distance to be covered. For each zone z_i of

R, we are interested in the distance between this zone and its nearest neighbour belonging to R. The index corresponding to this nearest neighbour is equal to $\min_{j\in\underline{R},j\neq i} d_{ij}$ where $\underline{R}$ designates the set of indices of the zones of R. The MNND indicator associated with a reserve, R, can therefore be formulated as follows:

$$\mathrm{MNND}(R) = \frac{1}{|R|}\sum_{i\in\underline{R}} \min_{j\in\underline{R},j\neq i} d_{ij}.$$

The indicator MNND applied to reserve R concerns all the zones of R and is equal to the average of the distances between each zone of R and the zone closest to it. The dimension of MNND is a length. If the zones closest to each zone are further away, then MNND increases and the "inter-zone" movements of the different species concerned become more difficult. Low values of $\mathrm{MNND}(R)$ correspond to a larger grouping of zones of R. We assume, for the definition of $\mathrm{MNND}(R)$, that there are at least two zones in reserve R.

Let us now look at the indicator MSI. It reflects a relationship between the perimeter of a zone and its area. More precisely, for each zone of the set R considered, we use the ratio between the perimeter of this zone and the square root of its area, all this multiplied by the coefficient 0.25. The value of the indicator MSI associated with a reserve, R, is then equal to the average of these values over all the zones of R. By noting, respectively, l_i and a_i the perimeter and the area of zone z_i, the indicator MSI associated with R is written

$$\mathrm{MSI}(R) = \frac{1}{|R|}\sum_{i\in\underline{R}} \frac{0.25\, l_i}{\sqrt{a_i}}.$$

For example, the value of this indicator is 0.89 for a circular zone, 1 for a square zone and 1.74 for a rectangular zone ten times longer than wide. MSI is dimensionless and minimal when all the zones have regular contours – circles. MSI increases with the irregularity of the contours of the zones.

Let us now look at the indicator MPI. Although the indicator MNND is useful for assessing the isolation of zones, considering only the zone closest to a given zone may not adequately represent the ecological neighbourhood of the zone under consideration. To remedy this weakness, we can consider the mean proximity index, MPI. This index takes into account both the proximity and the area of zones whose distance to a given zone is less than or equal to a certain value, d. The contribution of each zone to this index is calculated by summing, over all the zones within a given radius, the area of the zone divided by the square of the distance from the zone under consideration. The value of the index associated with a subset, R, of Z is then equal to the average of the values obtained for each zone of R. We obtain

$$\mathrm{MPI}\,(R, d) = \frac{1}{|R|}\sum_{i\in\underline{R}}\ \sum_{j\in I_i(R,d)} \frac{a_j}{d_{ij}^2},$$

where $I_i(R, d) = \{j \in \underline{R} : j \neq i, d_{ij} \leq d\}$. The contribution of a zone of R that does not have neighbouring zones – belonging to R – located at a distance less than or equal to the threshold distance, d, is equal to 0. $\mathrm{MPI}(R, d)$ is dimensionless and

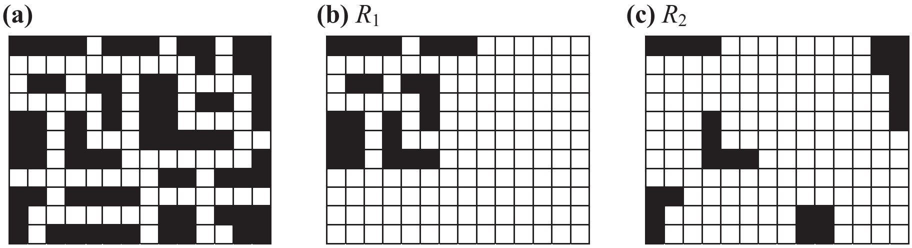


FIG. 2.2 – (a) The hypothetical landscape is represented by a grid of square and identical cells; 17 zones (in black) are candidates for protection. Two examples of reserves with the same area, R_1 and R_2, built from these zones: (b) MNND(R_1) = 1, MSI(R_1) = 1.18, MPI (R_1,2) = 11.5; (c) MNND(R_2) = 2.73, MSI(R_2) = 1.23, MPI(R_2,2) = 0.9.

increases with the size and proximity of the surrounding zones. This indicator measures the relative isolation of zones within a landscape.

Figure 2.2 illustrates the calculation of the 3 indicators MNND, MSI, and MPI on a small instance with 17 candidate zones.

2.3 Reserve Minimizing the Indicator MNND

With regard to the indicator MNND, the basic problem is to select, under certain constraints, a subset of zones that minimizes this indicator. Consider, for example, the problem of selecting, under a budgetary constraint, a subset of zones, $R \subseteq Z$, which allows to protect, at a minimum, a certain number, Ns, of species and which minimizes MNND. The set of species considered is $S = \{s_1, s_2, \ldots, s_m\}$. Let us situate ourselves in the case where the number of species protected by a reserve, R, is estimated by the quantity $\mathrm{Nb}_1(R)$ (see chapter 1, section 1.1). Recall that, in the calculation of $\mathrm{Nb}_1(R)$, it is assumed that the protection of a zone allows all the species present in that zone to be protected, provided that their population size is greater than or equal to a certain threshold value. We note Z_k the set of zones whose protection results in the protection of species s_k and $\underline{Z}_k$ the corresponding set of indices. We assume that we know the set Z_k for all $k \in \underline{S} = \{1, 2, \ldots, m\}$. Let us adopt the following notations: $\underline{Z} = \{1, \ldots, n\}$, $I_i = \{j \in \underline{Z} : j \neq i\}$ for all $i \in \underline{Z}$ and, for all vector x of $\{0,1\}^n$, $I(x) = \{i \in \underline{Z} : x_i = 1\}$, and $I_i(x) = \{j \in \underline{Z} : j \neq i,\ x_j = 1\}$ for all $i \in \underline{Z}$. Note that if x is the characteristic vector of reserve R ($x_i = 1 \Leftrightarrow z_i \in R$) then $I(x) = \underline{R}$ and $I_i(x) = \underline{R} - \{i\}$. The problem considered can be formulated as the fractional mathematical program in Boolean variables $\mathrm{P}_{2.1}$ (see appendix at the end of the book).

$$
\mathrm{P}_{2.1}: \left\{
\begin{array}{ll}
\min & \sum\limits_{i \in I(x)} \min\limits_{j \in I_i(x)} d_{ij} \Big/ \sum\limits_{i \in \underline{Z}} x_i \\
\text{s.t.} & \left|
\begin{array}{llcll}
\sum\limits_{i \in \underline{Z}} c_i x_i \leq B & (2.1.1) & | & x_i \in \{0,1\} \quad i \in \underline{Z} & (2.1.4) \\
\sum\limits_{k \in \underline{S}} y_k \geq \mathrm{Ns} & (2.1.2) & | & y_k \in \{0,1\} \quad k \in \underline{S} & (2.1.5) \\
y_k \leq \sum\limits_{i \in \underline{Z}_k} x_i \quad k \in \underline{S} & (2.1.3) & | & &
\end{array}
\right.
\end{array}
\right.
$$

This program consists in determining the values of variables x_i and y_k that respect constraints 2.1.1–2.1.5 and that minimize an economic function expressed as a fraction whose denominator is a linear function. We will see how to also express the numerator of this fraction by a linear function in order to finally obtain an economic function expressed as the ratio of two linear functions. Lemma 2.1 below shows how to express, for any vector x of $\{0,1\}^n$, the value of the expression $\sum_{i\in I(x)} \min_{j\in I_i(x)} d_{ij}$ as the optimal value of an integer linear program including the decision variables x_i, $i \in \{1,\ldots,n\}$, and the additional Boolean "working" variables t_{ij}, $(i,j) \in \underline{Z}^2, i \neq j$. By definition, variable t_{ij} is equal to 1 if and only if, on the one hand, zones z_i and z_j are selected and, on the other hand, zone z_j is, among the selected zones, the one closest to z_i.

Lemma 2.1. For all vector x of $\{0,1\}^n$,

$$\sum_{i\in I(x)} \min_{j\in I_i(x)} d_{ij} = \min\Bigg\{ \sum_{\substack{(i,j)\,\in\,\underline{Z}^2\\ i\neq j}} d_{ij}t_{ij} : t \in \{0,1\}^{n\times n},$$
$$\sum_{j\in I_i} t_{ij} = x_i,\ t_{ij} \leq x_j\ ((i,j)\in \underline{Z}^2, i\neq j)\Bigg\}.$$

Proof.

$$\begin{aligned}
\sum_{i\in I(x)} \min_{j\in I_i(x)} d_{ij} &= \sum_{i\in \underline{Z}} x_i \min_{j\in I_i(x)} d_{ij}\\
&= \sum_{i\in \underline{Z}} x_i \min\Bigg\{ \sum_{j\in I_i} d_{ij}t_{ij} : t\in\{0,1\}^{n\times n},\\
&\qquad \sum_{j\in I_i} t_{ij} = 1,\ t_{ij}\leq x_j\ (j\in I_i)\Bigg\}\\
&= \sum_{i\in \underline{Z}} \min\Bigg\{ \sum_{j\in I_i} d_{ij}t_{ij} : t\in\{0,1\}^{n\times n},\\
&\qquad \sum_{j\in I_i} t_{ij} = x_i,\ t_{ij}\leq x_j\ (j\in I_i)\Bigg\}\\
&= \min\Bigg\{ \sum_{(i,j)\in \underline{Z}^2, i\neq j} d_{ij}t_{ij} : t\in\{0,1\}^{n\times n},\\
&\qquad \sum_{j\in I_i} t_{ij} = x_i\ (i\in \underline{Z}),\ t_{ij}\leq x_j\ ((i,j)\in \underline{Z}^2, i\neq j)\Bigg\}.
\end{aligned}$$

Lemma 2.1 allows program $\mathrm{P}_{2.1}$ to be rewritten as program $\mathrm{P}_{2.2}$.

$$
\mathrm{P}_{2.2}\ :\ \left\{
\begin{array}{lll}
\min & \displaystyle\sum_{(i,j)\in\underline{Z}^2,\ i\neq j} d_{ij}t_{ij} \Big/ \sum_{i\in\underline{Z}} x_i & \\
\text{s.t.} & \left|
\begin{array}{lll}
\displaystyle\sum_{j\,:\,(i,j)\in\underline{Z}^2,\ i\neq j} t_{ij} = x_i & i\in\underline{Z} & (2.2.1)\\
t_{ij}\leq x_j & (i,j)\in Z^2, i\neq j & (2.2.2)\\
\displaystyle\sum_{i\in\underline{Z}} c_i x_i \leq B & & (2.2.3)\\
\displaystyle\sum_{k\in\underline{S}} y_k \geq \mathrm{Ns} & & (2.2.4)\\
y_k \leq \displaystyle\sum_{i\in\underline{Z}_k} x_i & k\in\underline{S} & (2.2.5)\\
x_i\in\{0,1\} & i\in\underline{Z} & (2.2.6)\\
y_k\in\{0,1\} & k\in\underline{S} & (2.2.7)\\
t_{ij}\in\{0,1\} & (i,j)\in Z^2, i\neq j & (2.2.8)
\end{array}\right.
\end{array}\right.
$$

Program $\mathrm{P}_{2.2}$ consists of minimizing the ratio of two linear functions whose variables are subject to linear constraints. This problem can be solved using the algorithms of fractional programming, for example the Dinkelbach algorithm (see appendix at the end of the book). In this case, the auxiliary problem associated with the – combinatorial – fractional program $\mathrm{P}_{2.2}$ consists in minimizing the linear function, parameterized by the scalar λ, $\sum_{(i,j)\in Z^2,\, i\neq j} d_{ij}t_{ij} - \lambda\sum_{i\in\underline{Z}} x_i$, under the same constraints as those of program $\mathrm{P}_{2.2}$.This auxiliary problem is a linear program in Boolean variables.

2.4 Examples of Reserves Minimizing the Indicator MNND

Consider a set of 20 rectangular zones spread over a 15 km square territory (figure 2.3). The total area of these 20 zones is 79 km^2 and the value of the indicator MNND for these 20 zones is 1.05 km.

We are interested in 10 species and, for each of the zones, we know all the species that live there in sufficient numbers to ensure that the protection of the zone will lead to the protection of this set of species. We also know the cost associated with protecting each zone. This information is summarized in table 2.1. We are looking for a subset of zones, R, which minimizes MNND(R), which protects, at a minimum, a fixed number of species, Ns, and whose cost is less than or equal to the available budget, B. The results obtained by solving program $\mathrm{P}_{2.2}$ are presented in table 2.2 for different values of B and Ns.

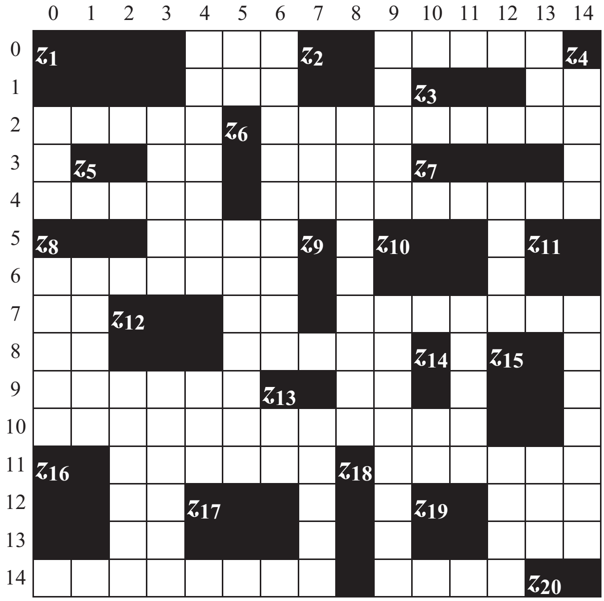


FIG. 2.3 – A set of 20 rectangular zones, z_1, z_2,..., z_{20}, distributed over a 15 km square territory represented by a grid of 15 × 15 identical square cells whose area is equal to 1 km^2. The total area of these 20 zones is 79 km^2 and the value of the indicator MNND for these 20 zones is 1.05 km.

TAB. 2.1 – Cost associated with protecting each zone of figure 2.3 and list of the species living in each of these zones in sufficient numbers.

Zone	Cost	Species living in the zone	Zone	Cost	Species living in the zone
z_1	2	s_5	z_{11}	3	s_4
z_2	2	s_5	z_{12}	3	s_9 s_{10}
z_3	1	s_6	z_{13}	4	s_4
z_4	5	s_7	z_{14}	3	s_4
z_5	1	s_1 s_8	z_{15}	4	s_1 s_4
z_6	2	s_2	z_{16}	3	s_{10}
z_7	2	s_9	z_{17}	3	s_1 s_2
z_8	4	s_1 s_3	z_{18}	2	s_4
z_9	5	s_1	z_{19}	4	s_5
z_{10}	5	s_1 s_4	z_{20}	1	s_8

TAB. 2.2 – Results corresponding to the minimization of the indicator MNND for the instance described in figure 2.3 and table 2.1, for different values of the minimal number of species to be protected, Ns, and the available budget, B.

Number of species to be protected (Ns)	B	Protected species	Budget used	Zones selected	Protected area in % of the initial area	MNND(R) (km)
5	5	s_1 s_6 s_8 s_9 s_{10}	5	z_3 z_5 z_{12}	13.92	4.36
5	8	s_1 s_3 s_8 s_9 s_{10}	8	z_5 z_8 z_{12}	13.92	1.00
7	9	s_1 s_2 s_5 s_6 s_8 s_9 s_{10}	9	z_2 z_3 z_5 z_6 z_{12}	22.78	1.40
7	12	s_1 s_2 s_3 s_5 s_6 s_8 s_9	12	z_2 z_3 z_5 z_6 z_7 z_8	24.05	1.00
8	10	–	–	–	–	–
8	12	s_1 s_2 s_4 s_5 s_6 s_8 s_9 s_{10}	12	z_2 z_3 z_5 z_6 z_{11} z_{12}	27.85	1.67
10	20	all	20	z_1 z_3 z_4 z_5 z_6 z_8 z_{12} z_{18}	37.97	1.33
10	21	all	21	z_1 z_3 z_4 z_5 z_8 z_{12} z_{17} z_{18}	41.77	1.00

– No solution.

2.5 Reserve Minimizing the Indicator MNND with a Constraint on the Indicator MSI

Several optimization problems can be considered with regard to the indicator MSI. For example, we consider the following problem: determine, under a budgetary constraint, a subset of zones that can protect at least a certain number of species, Ns, whose MSI value is less than or equal to a given value, $\mathrm{MSI}_{\max}$, and which minimizes the value of the indicator MNND. As in sections 2.3 and 2.4, the number of species protected by a reserve, R, is estimated by $\mathrm{Nb}_1(R)$. This optimization problem can be formulated as the fractional combinatorial program $\mathrm{P}_{2.2}$ to which is added the linear constraint $0.25 \sum_{i\in\underline{Z}} (l_i/\sqrt{a_i})\, x_i \leq \mathrm{MSI}_{\max} \times \sum_{i\in\underline{Z}} x_i$. The obtained program can be solved, like $\mathrm{P}_{2.2}$, by the Dinkelbach algorithm. The auxiliary program associated with the fractional program obtained consists in minimizing the parameterized linear function $\sum_{(i,j)\in Z^2,\, i\neq j}\ d_{ij} y_{ij} - \lambda \sum_{i\in\underline{Z}} x_i$ under the set of constraints of $\mathrm{P}_{2.2}$ plus the constraint on the maximal MSI value.

2.6 Examples of Reserves Minimizing the Indicator MNND with a Constraint on the Indicator MSI

Let us take the same instance as described in section 2.4 and look for a subset of zones, R, of minimal fragmentation, *i.e.,* minimizing $\mathrm{MNND}(R)$, which protects, at a minimum, a fixed number of species, Ns, and whose associated MSI indicator value is less than or equal to a given value, $\mathrm{MSI}_{\max}$. The results obtained are presented in table 2.3.

2.7 Reserve Maximizing the Indicator MPI

Many optimization problems can arise in connection with this indicator. Consider, for example, the following problem: determine, under a budgetary constraint, a set, R, of zones to be protected in order to protect at least Ns species, while maximizing $\mathrm{MPI}(R,\, d)$. As in the previous sections, the number of species protected by reserve R is estimated by $\mathrm{Nb}_1(R)$. To formulate this problem, simply replace the objective of $\mathrm{P}_{2.1}$ by the function $\left(\sum_{i\in I(x)}\ \sum_{j\in I_i(x,d)} \left(a_j/d_{ij}^2\right)\right)\Big/\sum_{i\in\underline{Z}} x_i$ to be maximized where, for any i of $\underline{Z}$ and any x of $\{0,1\}^n$, $I_i(x,\,d) = \{j \in \underline{Z} : j \neq i, x_j = 1, d_{ij} \leq d\}$. We will see how to reformulate the program obtained as a fractional combinatorial program consisting in maximizing the ratio of two linear functions under linear constraints.

Lemma 2.2. Program $\mathrm{P}_{2.3}$ is equivalent to the fractional linear program $\mathrm{P}_{2.4}$.

$$\mathrm{P}_{2.3} : \begin{cases} \max\ \left(\sum_{i\in I(x)}\ \sum_{j\in I_i(x,d)} \left(a_j/d_{ij}^2\right)\right)\Big/\sum_{i\in\underline{Z}} x_i \\ \text{s.t.}|\ \ x_i \in \{0,1\} \quad i \in \underline{Z} \quad (2.3.1) \end{cases}$$

TAB. 2.3 – Results associated with the instance described in figure 2.3 and table 2.1: Minimization of the indicator MNND for different values of the minimal number of species to be protected, Ns, and the available budget, B, with a maximal value of the indicator MSI, MSI_{max}.

Number of species to be protected (Ns)	MSI_{max}	B	Protected species	Selected zones	Budget used	Protected area in % of the initial area	MNND(R) (km)	MSI(R)
5	1.02	8	$s_1\ s_2\ s_5\ s_9\ s_{10}$	$z_2\ z_{12}\ z_{17}$	8	20.25	3.80	1.01
	1.02	12	$s_1\ s_4\ s_5\ s_9\ s_{10}$	$z_{12}\ z_{15}\ z_{19}$	12	20.25	2.61	1.01
	1.50	5	$s_1\ s_6\ s_8\ s_9\ s_{10}$	$z_3\ z_5\ z_{12}$	5	13.92	4.36	1.08
	1.50	8	$s_1\ s_2\ s_5\ s_6\ s_8\ s_9$	$z_1\ z_3\ z_5\ z_6\ z_7$	8	25.32	1.00	1.14
8	1.02	12	–	–	–	–	–	–
	1.02	17	$s_1\ s_2\ s_4\ s_5\ s_7\ s_8\ s_9\ s_{10}$	$z_2\ z_4\ z_5\ z_{11}\ z_{12}\ z_{17}$	17	29.11	3.52	1.02
	1.50	12	$s_1\ s_2\ s_4\ s_5\ s_6\ s_8\ s_9\ s_{10}$	$z_2\ z_3\ z_5\ z_6\ z_{11}\ z_{12}$	12	27.85	1.67	1.07
	1.50	15	$s_1\ s_2\ s_3\ s_5\ s_6\ s_8\ s_9\ s_{10}$	$z_2\ z_3\ z_5\ z_6\ z_8\ z_{12}$	15	26.58	1.00	1.09

– No solution.

$$\mathrm{P}_{2.4}:\left\{\begin{array}{ll}\max & \sum_{i\in\underline{Z}} v_i \Big/ \sum_{i\in\underline{Z}} x_i \\ \text{s.t.} & \left|\begin{array}{llll|llll} v_i \le \sum\limits_{j\in I_i(d)} \dfrac{a_j}{d_{ij}^2} x_j & i\in\underline{Z} & (2.4.1) & & v_i \ge 0 & i\in\underline{Z} & (2.4.3) \\ v_i \le M_i x_i & i\in\underline{Z} & (2.4.2) & & x_i\in\{0,1\} & i\in\underline{Z} & (2.4.4)\end{array}\right.\end{array}\right.$$

where M_i is a constant greater than or equal to the value of the expression $\sum_{j\in I_i(d)} (a_j/d_{ij}^2)x_j$ in an optimal solution of $\mathrm{P}_{2.3}$. We can take, for example, $M_i = \sum_{j\in I_i(d)} (a_j/d_{ij}^2)$. By examining successively the two possible values of x_i, it can easily be verified that constraints 2.4.1 and 2.4.2 imply $v_i = x_i \sum_{j\in I_i(d)} \frac{a_j}{d_{ij}^2} x_j$, at the optimum of $\mathrm{P}_{2.4}$. The objective of $\mathrm{P}_{2.4}$ is therefore equivalent to maximizing the expression $\left(\sum_{i\in\underline{Z}} x_i \sum_{j\in I_i(d)} \frac{a_j}{d_{ij}^2} x_j\right) \Big/ \sum_{i\in\underline{Z}} x_i$. This last expression, to be maximized, is a rewriting of the economic function of $\mathrm{P}_{2.3}$, since it is easy to verify that $\sum_{i\in\underline{Z}} x_i \sum_{j\in I_i(d)} \frac{a_j}{d_{ij}^2} x_j = \sum_{i\in I(x)} \sum_{j\in I_i(x,d)} \frac{a_j}{d_{ij}^2}$. $\mathrm{P}_{2.4}$ is therefore equivalent to $\mathrm{P}_{2.3}$.

Finally, the problem considered – determining, taking into account an available budget, B, a set of zones, R, to be protected in order to protect at least Ns species, while maximizing MPI(R, d) – can be formulated as the fractional mathematical program $\mathrm{P}_{2.5}$.

$$\mathrm{P}_{2.5}:\left\{\begin{array}{ll}\max & \sum_{i\in\underline{Z}} v_i \Big/ \sum_{i\in\underline{Z}} x_i \\ \text{s.t.} & \left|\begin{array}{llll|llll} v_i \le \sum\limits_{j\in I_i(d)} \dfrac{a_j}{d_{ij}^2} x_j & i\in\underline{Z} & (2.5.1) & & y_k \le \sum\limits_{i\in\underline{Z}_k} x_i & k\in\underline{S} & (2.5.5) \\ v_i \le M_i x_i & i\in\underline{Z} & (2.5.2) & & v_i \ge 0 & i\in\underline{Z} & (2.5.6) \\ \sum\limits_{i\in\underline{Z}} c_i x_i \le B & & (2.5.3) & & x_i\in\{0,1\} & i\in\underline{Z} & (2.5.7) \\ \sum\limits_{k\in\underline{S}} y_k \ge \mathrm{Ns} & & (2.5.4) & & y_k\in\{0,1\} & k\in\underline{S} & (2.5.8)\end{array}\right.\end{array}\right.$$

The auxiliary problem associated with $\mathrm{P}_{2.5}$ is to maximize the parameterized linear function $\sum_{i\in\underline{Z}} v_i - \lambda \sum_{i\in\underline{Z}} x_i$ under the same constraints as those of $\mathrm{P}_{2.5}$.

Example 2.1. Consider the instance described in figure 2.3 and table 2.1 and the problem of maximizing the indicator MPI for different values of the threshold distance, d, minimal number of species to be protected, Ns, and available budget, B. The results obtained, by solving program $\mathrm{P}_{2.5}$, are presented in table 2.4.

TAB. 2.4 – Results concerning the maximization of the indicator MPI for the instance described in figure 2.3 and table 2.1, for different values of Ns, d, and B.

Number of species to be protected (Ns)	d (km)	B	Protected species	Selected zones	Used budget	Protected area in % of the initial area	MPI(R, d) (km)
5	4	6	s_1 s_2 s_5 s_6 s_8	z_1 z_3 z_5 z_6	6	20.25	5.66
	4	10	s_1 s_2 s_5 s_6 s_8 s_9	z_1 z_2 z_3 z_5 z_6 z_7	10	30.38	8.23
	6	6	s_1 s_2 s_5 s_6 s_8	z_1 z_3 z_5 z_6	6	20.25	5.73
	6	10	s_1 s_2 s_5 s_6 s_8	z_1 z_2 z_3 z_5 z_6 z_7	10	30.38	8.34
	6	12	s_1 s_3 s_8 s_9 s_{10}	z_1 z_5 z_6 z_8 z_{12}	12	27.85	8.53
8	4	10	–	–	–	–	–
	4	12	s_1 s_2 s_4 s_5 s_6 s_8 s_9 s_{10}	z_1 z_3 z_5 z_6 z_{11} z_{12}	12	32.91	4.42
	6	10	–	–	–	–	–
	6	12	s_1 s_2 s_4 s_5 s_6 s_8 s_9 s_{10}	z_1 z_3 z_5 z_6 z_{11} z_{12}	12	32.91	4.57
	6	15	s_1 s_2 s_3 s_5 s_6 s_8 s_9 s_{10}	z_1 z_2 z_3 z_5 z_6 z_8 z_{12}	15	36.71	8.52

– No solution.

References and Further Reading

Bajalinov E.B. (2003) *Linear-fractional programming theory, methods, applications and software.* Springer-Science+Business Media, B.V.

Battisti C. (2003) Habitat fragmentation, fauna and ecological network planning: toward a theoretical conceptual framework, *Ital. J. Zool.* **70**, 241.

Bennett A.F., Saunders D.A. (2010) Habitat fragmentation and landscape change, *Conservation biology for all* (N.S. Sodhi, P.R. Ehrlich, Eds). Oxford University Press, Ch. 5, pp. 88–106.

Billionnet A. (2010) Optimal selection of forest patches using integer and fractional programming, *Oper. Res.: Int. J.* **10**, 1.

Collinge S.K. (2000) Effects of grassland fragmentation on insect species loss, colonization, and movement patterns, *Ecology* **81**, 2211.

Dinkelbach W. (1967) On nonlinear fractional programming, *Manage. Sci.* **13**, 492.

Fahrig L. (2003) Effects of habitat fragmentation on biodiversity, *Annu. Rev. Ecol. Evol. Syst.* **34**, 487.

Fischer D.T., Church R.L. (2003) Clustering and compactness in reserve site selection: an extension of the biodiversity management area selection model, *Forest Sci.* **49**, 555.

Fischer J., Lindenmayer D.B. (2007) Landscape modification and habitat fragmentation: a synthesis, *Global Ecol. Biogeo.* **16**, 265.

Fletcher R.J. Jr. (2005) Multiple edge effects and their implications in fragmented landscapes, *J. Animal Ecol.* **74**, 342.

Hanski I., Ovaskainen O. (2000) The metapopulation capacity of a fragmented landscape, *Nature* **404**, 755.

Hargis C.D., Bissonette J.A., David J.L. (1998) The behaviour of landscape metrics commonly used in the study of habitat fragmentation, *Landscape Ecol.* **13**, 167.

Hargis C.D., Bissonette J.A., Turner D.L. (1999) The influence of forests fragmentation and landscape pattern on American martens, *J. Appl. Ecol.* **36**, 167.

Harrison S., Bruna E. (1999) Habitat fragmentation and large-scale conservation: what do we know for sure? *Ecography* **22**, 225.

Henle K., Davies K.F., Kleyer M., Margules C., Settele J. (2004) Predictors of species sensitivity to fragmentation, *Biodiver. Conserv.* **13**, 207.

Hilty J.A., Lidicker W.Z. Jr., Merenlender A.M. (2006) *Corridor ecology, the science and practice of linking landscapes for biodiversity conservation*. Island Press.

Lindenmayer D.B., Cunningham R.B., Pope M.L., Donnelly C.F. (1999) A large-scale "experiment" to examine the effects of landscape context and habitat fragmentation on mammals *Biol. Conserv.* **88**, 387.

Lindenmayer D.B., Fischer J. (2006) *Habitat fragmentation and landscape change: an ecological and conservation synthesis*. Island Press.

Marks B.J., McGarigal K. (1994) Fragstats: spatial pattern analysis program for quantifying landscape structure. Technical report, Forest Science Department, Oregon State University.

Moilanen A., Wilson K.A., Possingham H.P., Eds (2009) *Spatial conservation prioritization.* Oxford University Press.

Noss R.F. (1987) Corridors in real landscapes: a reply to Simberloff and Cox, *Conserv. Biol.* **1**, 159.

Ohman K., Lamas T. (2005) Reducing forest fragmentation in long-term forest planning by using the shape index, *Forest Ecol. Manage.* **212**, 346.

Önal H., Briers R.A. (2002) Incorporating spatial criteria in optimum reserve network selection, *Proc. Royal Soc. London B* **269**, 2437.

Radzik T. (1998) Fractional combinatorial optimization, *Handbook of combinatorial optimization* (D.-Z. Du, P.M. Pardalos, Eds), Springer, vol. 1, pp. 429–478.

Rebelo A.G., Siegfried W.R. (1992) Where should nature reserves be located in the Cape Floristic region, South Africa? Models for the spatial configuration of a reserve network aimed at maximizing the protection of floral diversity, *Conserv. Biol.* **6**, 243.

Ricketts T.H. (2001) The matrix matters: effective isolation in fragmented landscapes, *Am. Nat.* **158**, 87.

Saunders D.A., Hobbs R.J., Margules C.R. (1991) Biological consequences of ecosystem fragmentation: a review, *Conserv. Biol.* **5**, 18.

Schaible S. (1995) Fractional programming, *Handbook of global optimization* (R. Horst, P. Pardalos, Eds). Kluwer Academic Publishers, pp. 495–608.

Shafer C.L. (2001) Inter-reserve distance, *Biol. Conserv.* **100**, 215.

Tischendorf L., Fahrig L. (2000) On the usage and measurement of landscape connectivity, *Oikos* **90**, 7.

Vemema H.D. (2005) Forest structure optimization using evolutionary programming and landscape ecology metrics, *Euro. J. Oper. Res.* **164**, 423.

Walters J.R. (1998) The ecological basis of avian sensitivity to habitat fragmentation, *Avian conservation: research and management* (J. Marzluff, R. Sallabanks, Eds), Island Press.

Williams J.C., ReVelle C.S., Levin S.A. (2005) Spatial attributes and reserve design models: a review, *Environ. Model. Assess.* **10**, 163.

Chapter 3

Connectivity

3.1 Introduction

In this chapter, we focus on the selection of an optimal set of zones to form a connected reserve, *i.e.,* a one-piece reserve. In this type of reserve, species can move between all zones of the reserve without leaving it (figure 3.1). Many publications present the advantages and disadvantages of such reserves. It should be noted that this single notion of connectivity – also called connexity or contiguity – does not allow the shape or contour of the selected reserve to be controlled. Connectivity properties are also of interest, particularly in large reserves, to help some species in their adaptation to climate change. As in the previous chapters, several variants of the problem of selecting optimal reserves, linked to the meaning given to the adjective "optimal", can arise with this connectivity constraint. For example, one can seek to protect all the species or a given number of species, at a minimum, at the lowest cost, or to protect a maximal number of species taking into account an available budget (see chapter 1). As before, we denote by $Z = \{z_1, z_2,\dots, z_n\}$ the set of candidate zones, $\underline{Z} = \{1, 2,\dots, n\}$ the set of corresponding indices, $S = \{s_1, s_2,\dots, s_m\}$ the set of species concerned, and $\underline{S} = \{1, 2,\dots, m\}$ the set of corresponding indices. For the presentation of the different approaches that can be used to address this issue of connectivity – which is difficult – we retain the problem of determining a least-cost reserve, R, that allows all species to be protected. In addition, a species is considered as protected by reserve R if at least one of the zones of R allows this species to be protected, and for each species we know the list of zones allowing to protect it. By noting $\mathrm{Fc}(Z)$ the family of subsets of zones of Z forming a connected reserve and $C(R)$ the cost of a reserve, R, this problem can be formulated as the minimization problem $\min_{R\in \mathrm{Fc}(Z),\mathrm{Nb}_1(R)=m} C(R)$ where $\mathrm{Nb}_1(R)$ refers to the number of species protected by reserve R (chapter 1, section 1.1). It should be recalled that, for the calculation of $\mathrm{Nb}_1(R)$, it is considered that the protection of a zone, z_i, allows all the species present in this zone to be protected provided that their population size in this zone is greater than or equal to a certain threshold value.

DOI: 10.1051/978-2-7598-2521-9.c003

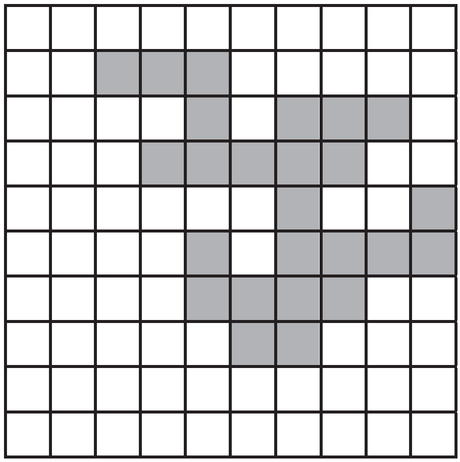


FIG. 3.1 – The set of candidate zones are represented by a grid of 20 × 20 square and identical zones. Two zones are considered adjacent if they share a common side. The 25 grey zones form a connected reserve.

We denote by n_{ik} the population size of the species s_k in zone z_i and v_{ik} the threshold value for species s_k in this zone. For each species s_k we therefore know the set Z_k of the zones whose protection allows this species to be protected: $Z_k = \{z_i \in Z : n_{ik} \geq v_{ik}\}$. We denote by $\underline{Z}_k$ the set of indices of the elements of Z_k. It is also necessary to define the notion of adjacency between two zones: for each pair of zones, it must be decided whether they can be considered as adjacent or not. For example, the length of their common border can be taken into account. The notion of adjacency may vary from one species to another. Indeed, for a given species, this notion simply reflects the possibility of being able to move from one zone to another without having to face a potentially inhospitable environment.

Example 3.1. Figure 3.1 presents a connected reserve.

The search for an optimal connected reserve can be formulated in many ways within the framework of mathematical programming. Some of these formulations are presented below.

3.2 Protection by a Connected Reserve of All the Species Considered, at the Lowest Cost: Graph Formulation

As mentioned above, we address the problem of selecting a set of zones that form a connected reserve, at minimal cost and that protect all the species considered. With the set of candidate zones is associated a graph, $G = (\underline{Z}, U)$, where the set of vertices, $\underline{Z} = \{1,\ldots, n\}$, corresponds to the set of indices of the zones of $Z = \{z_1, z_2, \ldots, z_n\}$, and where the set of arcs, U, includes an arc going from vertex i to vertex j if

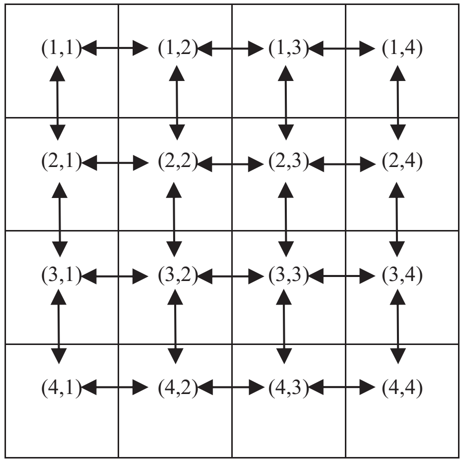


FIG. 3.2 – Graph G associated with a set of 16 candidate zones represented by a grid of 4×4 square and identical zones. Two zones are considered adjacent if they share a common side. Each vertex of the graph corresponds to a zone and vice versa. A vertex is identified by a pair (i, j) where i is the row index and j is the column index. The double arrow between the two vertices (i, j) and (k, l) represents the arc from (i, j) to (k, l) and the arc from (k, l) to (i, j).

and only if zones z_i and z_j are adjacent. The graph G thus defined is, therefore, a symmetric graph (figure 3.2). The problem can then be formulated as follows: find a subset $\underline{\hat{Z}}$ of $\underline{Z}$ of minimal cost and such that:

(i) Each species is protected by at least one zone associated with a vertex of $\underline{\hat{Z}}$;
(ii) The sub-graph induced by $\underline{\hat{Z}}$ is connected, *i.e.,* for each pair of vertices (s, t) of $\underline{\hat{Z}}$, there is a path of G, from s to t, which only uses vertices of $\underline{\hat{Z}}$ (see appendix at the end of the book).

Note that, to simplify the presentation of the examples in this chapter, the set of candidate zones is represented by a grid of nr $\times$ nc square and identical zones. Each zone of this grid is identified by the couple (i, j) where i is its row index and j, its column index. Apart from the examples, the candidate zones are represented by the set $Z = \{z_1, z_2, \ldots, z_n\}$.

Example 3.2. Figure 3.2 shows the graph associated with a set of 16 candidate zones.

3.3 Approach Based on the Search for a Set of Zones Inducing an Arborescence

The property (ii) of the sub-graph searched for in section 3.2 can also be stated as follows: the sub-graph of G induced by the vertices of $\underline{\hat{Z}}$ contains an arborescence $\mathcal{A}$, *i.e.*, a graph that satisfies the following 3 properties (figure 3.3):

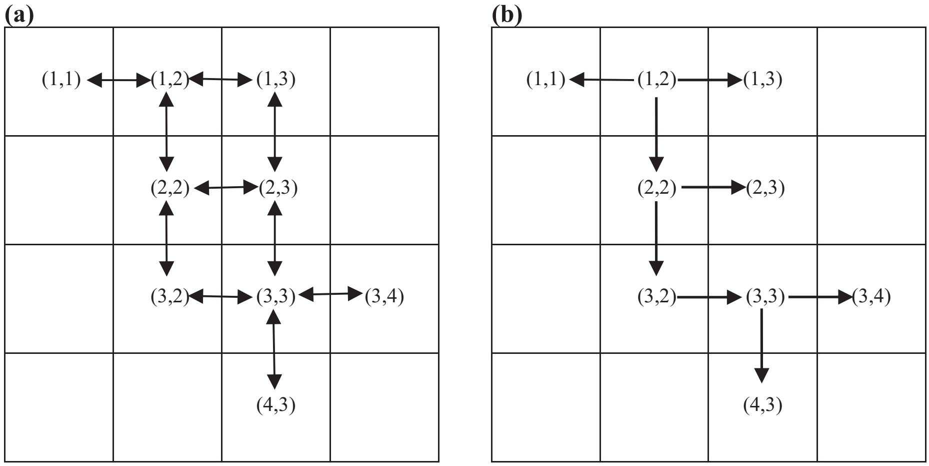


FIG. 3.3 – (a) $G_{\underline{\hat{Z}}} = (\underline{\hat{Z}}, U_{\underline{\hat{Z}}})$, the sub-graph of graph G of figure 3.2 induced by the set of vertices $\underline{\hat{Z}} = \{(1,1),(1,2),(1,3),(2,2),(2,3),(3,2),(3,3),(3,4),(4,3)\}$. (b) $\mathcal{A}$, a partial sub-graph of $G_{\underline{\hat{Z}}}$ which is an arborescence; the vertices (1, 2) is the root of this arborescence.

(i) Each vertex of $\mathcal{A}$ has at most one predecessor;
(ii) $\mathcal{A}$ includes $\underline{\hat{Z}} - 1$ arcs;
(iii) $\mathcal{A}$ does not contain circuits.

Example 3.3. Figure 3.3 shows an example of a connected sub-graph – from the graph G in figure 3.2 – and an associated arborescence.

3.3.1 Case Where No Zone is Mandatory

First of all, we are dealing with the case where none of the candidate zones must be necessarily retained in the reserve. We use the Boolean variables x_i which, by convention, take the value 1 if and only if vertex i – associated with zone z_i – is selected and the Boolean variables y_{ij} which by convention take the value 1 if and only if the arc (i, j), *i.e.*, the arc from vertex i to vertex j, is retained to form the arborescence. We also use the non-negative variables t_i which represent a value assigned to each vertex of the graph. By requiring these values to respect some constraints, we are sure to retain a set of arcs that does not form a circuit. This technique is based on a classic formulation of the travelling salesman's problem by a mixed-integer linear program with a polynomial number of constraints. We thus obtain a formulation of the problem by program $P_{3.1}$.

$$
\mathrm{P}_{3.1}: \begin{cases} \min \sum_{i \in \underline{Z}} c_i x_i \\ \text{s.t.} \left| \begin{array}{lll} \sum_{i \in \underline{Z}_k} x_i \geq 1 & k \in \underline{S} & (3.1.1) \\ y_{ij} \leq x_i & (i,j) \in U & (3.1.2) \\ \sum_{j \in \mathrm{Adj}_i^-} y_{ji} \leq x_i & i \in \underline{Z} & (3.1.3) \\ \sum_{(i,j) \in U} y_{ij} = \sum_{i \in \underline{Z}} x_i - 1 & & (3.1.4) \\ t_j \geq t_i + 1 - M(1 - y_{ij}) & (i,j) \in U & (3.1.5) \\ t_i \geq 0 & i \in \underline{Z} & (3.1.6) \\ x_i \in \{0,1\} & i \in \underline{Z} & (3.1.7) \\ y_{ij} \in \{0,1\} & (i,j) \in U & (3.1.8) \end{array} \right. \end{cases}
$$

Remember that U is the set of arcs of the graph associated with the candidate zones. For all $i \in \underline{Z}$, Adj_i^- refers to the set of vertices predecessors of vertex i. In other words, $\mathrm{Adj}_i^- = \{j \in \underline{Z} : (j,i) \in U\}$. M is a sufficiently large constant (*e.g.*, a value greater than or equal to the number of zones in an optimal reserve). The economic function expresses the cost of the zones selected to form the reserve. Constraints 3.1.1 express the fact that each species must be protected by at least one zone of the reserve. Given a subset of vertices, $\hat{\underline{Z}}$, and x its characteristic vector, a vector y of $\mathbb{R}^{|U|}$ defines an arborescence on the sub-graph induced by $\hat{\underline{Z}}$ if and only if constraints 3.1.2–3.1.8 are satisfied. Constraints 3.1.2 impose that, if vertex i is not selected, then none of the selected arcs should have this vertex as their initial end. If vertex i is selected, the corresponding constraint is inactive. Constraints 3.1.3 express that, in the case where vertex i is not selected, no arc with i as its terminal end can be retained. In the case where vertex i is selected, the corresponding constraint expresses that at most one arc with i as its terminal end can be retained. Constraint 3.1.4 expresses that the total number of retained arcs is equal to the number of retained vertices, less 1. Constraints 3.1.5, where M is a sufficiently large constant, eliminate the possibility that the retained arcs form a circuit. These constraints are similar to those used to eliminate the sub-tours in a classic formulation of the travelling salesman's problem by a mathematical program with a polynomial number of constraints. A positive or zero value t_i is assigned to each vertex i of the graph. If the arc (i, j) is retained – $y_{ij} = 1$ – then t_j must be greater than or equal to $t_i + 1$. Thus, all the selected arcs cannot form a circuit. If the arc (i, j) is not retained – $y_{ij} = 0$ – then the corresponding constraint 3.1.5 is always satisfied provided that the values t_i are less than or equal to M−1.

The problem can also be formulated as a slightly different mixed-integer linear program using the Boolean variables u_i which are equal to 1 if and only if the vertex i is chosen as the root. This gives program $\mathrm{P}_{3.2}$.

$$
\mathrm{P}_{3.2} : \begin{cases} \min \sum_{i \in \underline{Z}} c_i x_i \\ \text{s.t.} \left| \begin{array}{lll} \sum_{i \in \underline{Z}_k} x_i \geq 1 & k \in \underline{S} & (3.2.1) \\ y_{ij} \leq x_i & (i,j) \in U & (3.2.2) \\ \sum_{i \in \underline{Z}} u_i = 1 & & (3.2.3) \\ \sum_{j \in \mathrm{Adj}_i^-} y_{ji} = x_i - u_i & i \in \underline{Z} & (3.2.4) \\ t_j \geq t_i + 1 - M(1 - y_{ij}) & (i,j) \in U & (3.2.5) \\ t_i \geq 0 & i \in \underline{Z} & (3.2.6) \\ x_i \in \{0,1\},\ u_i \in \{0,1\} & i \in \underline{Z} & (3.2.7) \\ y_{ij} \in \{0,1\} & (i,j) \in U & (3.2.8) \end{array} \right. \end{cases} .
$$

Program $\mathrm{P}_{3.2}$ is obtained by replacing constraints 3.1.3 and 3.1.4 in $\mathrm{P}_{3.1}$ by constraints 3.2.3 and 3.2.4. Constraint 3.2.3 requires to choose the root in one and only one vertex of $\underline{Z}$. Constraints 3.2.4 express that any retained vertex i must be the terminal end of one and only one arc unless this vertex has been chosen as root – $x_i - u_i = 0$ – in which case it must not be the terminal end of any arc. Note that, according to constraints 3.2.4 and since the quantity $\sum_{j \in \mathrm{Adj}_i^-} y_{ji}$ is always positive or null, u_i can take the value 1 only if $x_i = 1$.

3.3.2 Case Where at Least One Zone is Mandatory

If the problem data are such that at least one vertex is mandatory – for example, because of constraints $\sum_{i \in \underline{Z}_k} x_i \geq 1,\ k \in \underline{S}$ – this vertex can be chosen as the root of the arborescence sought without loss of generality, and program $\mathrm{P}_{3.3}$ below, where r refers to this vertex, solves the problem.

$$
\mathrm{P}_{3.3}: \begin{cases}
\min \displaystyle\sum_{i \in \underline{Z}} c_i x_i & & \\
\text{s.t.} \left|
\begin{array}{lll}
\displaystyle\sum_{i \in \underline{Z}_k} x_i \geq 1 & k \in \underline{S} & (3.3.1) \\
y_{ij} \leq x_i & (i,j) \in U & (3.3.2) \\
\displaystyle\sum_{j \in \mathrm{Adj}_i^-} y_{ji} = x_i & i \in \underline{Z}, i \neq r & (3.3.3) \\
\displaystyle\sum_{i \in \mathrm{Adj}_r^-} y_{ir} = 0 & & (3.3.4) \\
t_j \geq t_i + 1 - M(1 - y_{ij}) & (i,j) \in U & (3.3.5) \\
x_i \in \{0,1\} & i \in \underline{Z} & (3.3.6) \\
t_i \geq 0 & i \in \underline{Z} & (3.3.7) \\
y_{ij} \in \{0,1\} & (i,j) \in U & (3.3.8)
\end{array}
\right.
\end{cases}
$$

Program $\mathrm{P}_{3.3}$ is obtained from program $\mathrm{P}_{3.2}$ by replacing constraints 3.2.3 and 3.2.4 with constraints 3.3.3 and 3.3.4. Constraints 3.3.3 express that, for all the selected vertices except the root, one and only one arc must have this vertex as its terminal end. It also expresses that no selected arc should have an unselected vertex as its terminal end. Constraint 3.3.4 expresses that no arcs should arrive on the vertex chosen as root. All other constraints are identical to those of program $\mathrm{P}_{3.2}$.

3.4 Approximate Solution When the Set of Candidate Zones is Represented by a Grid

In this section, we present an integer linear program to solve the problem in an approximate way in the case where the set of candidate zones for protection is represented by a grid. We denote by nr the number of rows in this grid and nc the number of its columns. This grid, therefore, includes nr × nc square and identical zones. The advantage of this approach is that the solution is obtained much faster than with the programs in section 3.3. On the other hand, the solution obtained, although often optimal, is not always optimal. This program differs from the previous ones in the way in which the prohibition of circuits is formulated. It is known that the technique used in programs $\mathrm{P}_{3.2}$ and $\mathrm{P}_{3.3}$ may be relatively ineffective since the computation time required to solve these programs may be prohibitive, even for medium-sized problems. We will use another technique. First of all, let us introduce the classic constraints prohibiting circuits of length 2, then specific constraints for the graph in question – associated with a grid – to prohibit circuits of greater length. These constraints are based on the following idea: in order to ensure that the

$$\begin{array}{ccccccc} & & (i-1, j) & \qquad & (k-1, l) & & \\ & & \downarrow & & \downarrow & & \\ (i, j-1) & \leftarrow & (i, j) & & (k, l) & \rightarrow & (k, l+1) \end{array}$$

FIG. 3.4 – These two types of paths of length 2 are prohibited by constraints $C_{3.1.2}$ and $C_{3.1.3}$.

solution does not contain circuits, it is sufficient to prohibit, in addition to circuits of length 2, paths formed by two arcs and of type $\{(i{-}1, j), (i, j), (i, j{-}1)\}$ or $\{(k{-}1, l), (k, l), (k, l+1)\}$ (figure 3.4). If the first type of path is prohibited then the solution cannot have circuits of length greater than or equal to 2 and "clockwise rotating"; if the second type of path is prohibited then the solution cannot have circuits of length greater than or equal to 2 and "anti-clockwise rotating" (figure 3.5). We can now formulate the program to solve the problem in an approximate way. To do this, it is sufficient to replace, in program $P_{3.3}$, constraints 3.3.5 and 3.3.7 by the 3 families of constraints of set $C_{3.1}$ after having adapted this program to the case of a grid – x_i becomes x_{ij} and y_{ij} becomes y_{ijkl}. The resulting program is called P_{approx}.

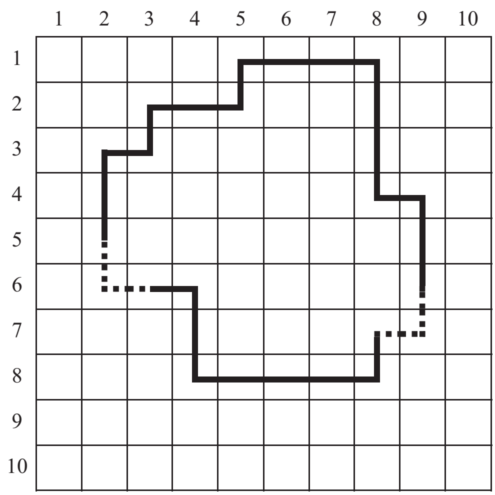


FIG. 3.5 – A cycle on a grid of 10 × 10 square and identical zones. By following this cycle in a clockwise direction, one necessarily encounters a path of length 2 of type $\{(i{-}1, j), (i, j), (i, j{-}1)\}$, for example the path $\{(6,9),(7,9),(7,8)\}$; by following this cycle in an anti-clockwise direction, one necessarily encounters a path of length 2 of type $\{(k{-}1, l), (k, l), (k, l+1)\}$, for example the path $\{(5,2),(6,2),(6,3)\}$.

$$\mathrm{C}_{3.1}: \begin{cases} y_{ijkl} + y_{klij} \leq 1 & ((i,j),(k,l)) \in U, k > i \text{ or } l > j \quad (\mathrm{C}_{3.1.1}) \\ y_{i-1,j,i,j} + y_{i,j,i,j-1} \leq 1 & i = 2,\ldots, nr\,; j = 2,\ldots, nc \quad (\mathrm{C}_{3.1.2}) \\ y_{k-1,l,k,l} + y_{k,l,k,l+1} \leq 1 & k = 2,\ldots, nr\,; l = 1,\ldots, nc-1 \quad (\mathrm{C}_{3.1.3}) \end{cases}$$

Constraints $\mathrm{C}_{3.1.1}$ prohibit circuits of length 2. Constraints $\mathrm{C}_{3.1.2}$ and $\mathrm{C}_{3.1.3}$, by prohibiting certain paths of length 2, prohibit circuits of length greater than 2. Variables t_i used in $\mathrm{P}_{3.2}$ and $\mathrm{P}_{3.3}$ are now useless. Remember that this approach can lead to a sub-optimal solution. Indeed, prohibiting certain paths – using the constraint set $\mathrm{C}_{3.1}$ – can prevent the consideration of arborescences that would be associated with an optimal solution.

3.5 Simple Flow Approach

This section presents a different formulation of the problem of selecting an optimal reserve than that in section 3.3. This formulation is based on the notion of flow in a graph (see appendix at the end of the book). As before, the graph $G = (\underline{Z}, U)$ is associated with the set of candidate zones, $Z = \{z_1, z_2,\ldots, z_n\}$. The set of vertices of the graph, $\underline{Z} = \{1, 2,\ldots, n\}$, corresponds to the zones and there is an arc from vertex i to vertex j if and only if zones z_i and z_j are adjacent. As we have seen in section 3.3, the problem can be formulated as follows: find a subset, $\hat{\underline{Z}} \subseteq \underline{Z}$, of minimal cost, to protect all the species, and such that the sub-graph of G induced by the vertices associated with $\hat{\underline{Z}}$ contains an arborescence. There are several ways to ensure that this sub-graph contains an arborescence. We consider here that it must admit a path from the root to all the other vertices, which allows us to formulate the problem as follows:

(i) Each species is protected by at least one zone associated with a vertex of $\hat{\underline{Z}}$.
(ii) A vertex of $\hat{\underline{Z}}$ is chosen as the root – or source.
(iii) In the sub-graph of G induced by the vertices associated with $\hat{\underline{Z}}$ there is a path from the root to all the other vertices.

In terms of flow in a graph, property (iii) can be expressed as follows: as soon as a vertex, different from the root, is selected to constitute the reserve, it must receive at least one unit of flow emitted by the root and routed along a path passing only through vertices of $\hat{\underline{Z}}$. This ensures that there is indeed a path from the root to all the other selected vertices. In this formulation, we use the Boolean variable x_i, $i \in \underline{Z}$, which is equal to 1 if and only if vertex i is selected – zone z_i is selected – and the non-negative variable ϕ_{ij}, $(i, j) \in U$, which represents the flow circulating on the arc (i, j).

3.5.1 Case Where No Zone is Mandatory

When no zone is mandatory, it is impossible to choose a priori a vertex – a zone – as a root. We therefore use the Boolean variable u_i, $i \in \underline{Z}$, which, by convention, takes the value 1 if and only if vertex i – zone z_i – is chosen as a root. The problem considered can then be formulated as the mixed-integer linear program $\mathrm{P}_{3.4}$.

$$
\mathrm{P}_{3.4} : \begin{cases} \min \sum_{i \in \underline{Z}} c_i x_i & \\ \text{s.t.} \left| \begin{array}{lll} \sum_{i \in \underline{Z}_k} x_i \geq 1 & k \in \underline{S} & (3.4.1) \\ \sum_{i \in \underline{Z}} u_i = 1 & & (3.4.2) \\ \sum_{j \in \mathrm{Adj}_i^+} \phi_{ij} \leq M x_i & i \in \underline{Z} & (3.4.3) \\ \sum_{j \in \mathrm{Adj}_i^-} \phi_{ji} - \sum_{j \in \mathrm{Adj}_i^+} \phi_{ij} \geq x_i - M u_i & i \in \underline{Z} & (3.4.4) \\ u_i \in \{0,1\},\ x_i \in \{0,1\} & i \in \underline{Z} & (3.4.5) \\ \phi_{ij} \geq 0 & (i,j) \in U & (3.4.6) \end{array} \right. \end{cases}
$$

Adj_i^+ refers to the set of successors to vertex i. The economic function of $\mathrm{P}_{3.4}$ expresses the total cost of the selected zones. Constraints 3.4.1 express that each species must be protected. Constraint 3.4.2 expresses that one and only one vertex should be chosen as a root. Constraints 3.4.3 express that, if zone z_i is not retained, then the total amount of flow starting from the vertex associated with z_i must be zero. On the other hand, if this zone is selected, then the amount of flow starting from the vertex associated with this zone is not limited if constant M is set to a sufficiently large value. This quantity M is, therefore, an upper bound of the value of the flow starting from the root; it can be taken, for example, as equal to the number of candidate zones. Now let us look at constraints 3.4.4 and, first of all, in case zone z_i is not chosen as a root – $u_i = 0$. The corresponding constraint becomes $\sum_{j \in \mathrm{Adj}_i^-} \phi_{ji} - \sum_{j \in \mathrm{Adj}_i^+} \phi_{ij} \geq x_i$ and two cases are then possible: (i) zone z_i is not retained, then according to constraints 3.4.3 and 3.4.6 $\sum_{j \in \mathrm{Adj}_i^+} \phi_{ij} = 0$, and in this case the corresponding constraint 3.4.4 is always satisfied since $x_i = 0$; (ii) zone z_i is retained, the corresponding constraint then becomes $\sum_{j \in \mathrm{Adj}_i^-} \phi_{ji} - \sum_{j \in \mathrm{Adj}_i^+} \phi_{ij} \geq 1$ and it expresses the fact that the sum of the incoming flows on i must be at least equal to the sum of the outcoming flows from i plus the unit of flow absorbed by i. Now let us look at the case where vertex i – corresponding to zone z_i – is chosen as a root. The corresponding constraint becomes $\sum_{j \in \mathrm{Adj}_i^-} \phi_{ji} \geq \sum_{j \in \mathrm{Adj}_i^+} \phi_{ij} + x_i - M$. The second member of this constraint is always negative or zero, whether x_i is equal to 0 or 1, and the constraint is, therefore, always satisfied. Note that in all feasible solutions, the root is well chosen among the selected zones.

Let us now show precisely that a solution of $\mathrm{P}_{3.4}$ provides an optimal connected reserve. (1) Let (x, ϕ, u) be a feasible solution of $\mathrm{P}_{3.4}$. Let us show that the sub-graph of G, G', generated by the vertices i such that $x_i = 1$ is connected. Suppose that it is not. Let C be a connected component of this sub-graph not containing vertex i such that $u_i = 1$. For each vertex i of this component, we have $\sum_{j\in \mathrm{Adj}_i^-} \phi_{ji} - \sum_{j\in \mathrm{Adj}_i^+} \phi_{ij} \geq 1$. We deduce $\sum_{i\in C} (\sum_{j\in \mathrm{Adj}_i^-} \phi_{ji} - \sum_{j\in \mathrm{Adj}_i^+} \phi_{ij}) \geq |C|$ what leads to $\sum_{(i,j)\in U, i\notin C, j\in C} \phi_{ij} - \sum_{(i,j)\in U, i\in C, j\notin C} \phi_{ij} \geq |C|$. This last inequality cannot be verified since, according to constraint 3.4.3, $\sum_{(i,j)\in U, i\notin C, j\in C} \phi_{ij} = 0$. Indeed, any initial end, i, of the arcs entering C – and belonging to G' – verifies $x_i = 0$ because if it were not the case, C would not be a connected component. (2) Let G' be a connected sub-graph of G. It contains an arborescence, A. Let r be the root of this arborescence. We can verify that (x, ϕ, u) defined as follows is a feasible solution of $\mathrm{P}_{3.4}$: $u_i = 1$ $(i \in \underline{Z}, i = r)$, $u_i = 0$ $(i \in \underline{Z}, i \neq r)$, $x_i = 1$ $(i \in \underline{Z}, i \in G')$, $x_i = 0$ $(i \in \underline{Z}, i \notin G')$, ϕ_{ij} = number of vertices of A that can be reached by a path of A whose first arc is (i, j), $((i,j) \in A)$, and $\phi_{ij} = 0$ $((i,j) \in U, (i,j) \notin A)$.

3.5.2 Case Where at Least One Zone is Mandatory

If at least one of the candidate zones is mandatory, it can be chosen as a root without loss of generality. Knowing the root allows the problem to be formulated as program $\mathrm{P}_{3.5}$, which is a simplification of program $\mathrm{P}_{3.4}$. Constraint 3.4.2 becomes useless and constraints 3.4.4 becomes, $\sum_{j\in \mathrm{Adj}_i^-} \phi_{ji} - \sum_{j\in \mathrm{Adj}_i^+} \phi_{ij} \geq x_i$, $i \in \underline{Z}, i \neq$ root.

$$
\mathrm{P}_{3.5}: \begin{cases} \min \sum\limits_{i\in \underline{Z}} c_i x_i & \\ \text{s.t.} \left| \begin{array}{lll} \sum\limits_{i\in \underline{Z}_k} x_i \geq 1 & k \in \underline{S} & (3.5.1) \\ \sum\limits_{j\in \mathrm{Adj}_i^+} \phi_{ij} \leq M x_i & i \in \underline{Z} & (3.5.2) \\ \sum\limits_{j\in \mathrm{Adj}_i^-} \phi_{ji} - \sum\limits_{j\in \mathrm{Adj}_i^+} \phi_{ij} \geq x_i & i \in \underline{Z}, i \neq \text{root} & (3.5.3) \\ x_i \in \{0,1\} & i \in \underline{Z} & (3.5.4) \\ \phi_{ij} \geq 0 & (i,j) \in U & (3.5.5) \end{array} \right. \end{cases}
$$

3.6 Multi-Flow Approach

Here we place ourselves in the case where at least one zone, z_r, is mandatory and we consider the corresponding vertex, r, as a source. With each vertex i of $\underline{Z}$, different from r, is associated a product type and all the vertices i of $\underline{Z}$, different from r, constitute, if retained, sinks for a unit of flow of the product i which has to be transported from source vertex r to vertex i. For each arc (k, l) of U and each vertex i different from the source, we introduce a Boolean variable y_{ikl} which takes the value 1 if and only if the arc (k, l) transports a unit of flow of product i – from vertex k to vertex l. If vertex i is selected, then it becomes a sink for the flow of product i. The problem can be posed as follows: select a subset of vertices, $\hat{\underline{Z}}$, including the source

r and such that the routing of one unit of flow of type i, from the source to vertex i, is possible by only passing through vertices of $\underline{\hat{Z}}$, and this for all i of $\underline{\hat{Z}}$ different from r. This gives program $\mathrm{P}_{3.6}$.

$$
\mathrm{P}_{3.6}: \begin{cases} \min \sum_{i \in \underline{Z}} c_i x_i & & \\ \text{s.t.} \left| \begin{array}{lll} \sum_{i \in \underline{Z}_k} x_i \geq 1 & k \in \underline{S} & (3.6.1) \\ \sum_{l \in \mathrm{Adj}_r^-} y_{ilr} = 0 & i \in \underline{Z}, i \neq r & (3.6.2) \\ \sum_{l \in \mathrm{Adj}_i^-} y_{ili} = x_i & i \in \underline{Z}, i \neq r & (3.6.3) \\ \sum_{l \in \mathrm{Adj}_i^+} y_{iil} = 0 & i \in \underline{Z}, i \neq r & (3.6.4) \\ \sum_{l \in \mathrm{Adj}_k^+} y_{ikl} = \sum_{l \in \mathrm{Adj}_k^-} y_{ilk} & k \in \underline{Z}, k \neq r, i \in \underline{Z}, i \neq r, k & (3.6.5) \\ y_{ikl} \leq x_k;\, y_{ikl} \leq x_l & i \in \underline{Z}, i \neq r, (k,l) \in U & (3.6.6) \\ y_{ikl} \geq 0 & i \in \underline{Z}, i \neq r, (k,l) \in U & (3.6.7) \\ x_i \in \{0,1\} & i \in \underline{Z} & (3.6.8) \end{array} \right. \end{cases}
$$

Constraints 3.6.1 require that all the species considered be protected. Constraints 3.6.2 require that the flow of product i entering the source is zero for any i. Constraints 3.6.3 and 3.6.4 require that any selected vertex i, other than the source, be a sink for product i. Constraints 3.6.5 require that the flow of product i be conserved at each vertex k different from sink i and source r. Constraints 3.6.6 require that, for any product i, the flow of this product circulating on each arc must be zero, if one of the two ends of this arc is not selected, and less than or equal to 1 if the two ends are selected. Constraints 3.6.7 impose to variables y_{ikl} to be non-negative. Given constraints 3.6.6 and 3.6.7, these variables can, therefore, take values between 0 and 1. In fact, it can be shown that they take either the value 0 or the value 1 at the optimum, in accordance with their definition. Finally, constraints 3.6.8 specify the Boolean nature of variables x_i. A similar approach to that of the previous section could be followed to rigorously prove that a solution of $\mathrm{P}_{3.6}$ provides an optimal connected reserve.

3.7 Constraint Generation Approach

We again state the problem as follows: determine a subset of vertices, $\underline{\hat{Z}}$, such that: (1) the protection of the zones associated with $\underline{\hat{Z}}$ allows all the species considered to be protected, (2) the sub-graph of G induced by the vertices of $\underline{\hat{Z}}$ contains an arborescence $\mathcal{A}$. Here, the arborescence $\mathcal{A}$ is defined as follows:

(i) Each vertex of $\mathcal{A}$, except the source – or root – has one and only one predecessor;
(ii) $\mathcal{A}$ does not contain circuits.

We also assume that at least one zone is mandatory, which allows the root to be fixed. The formulation we present here is derived from the classic formulation of the travelling salesman's problem by integer linear programming. Like the programs in the previous sections, this program uses the Boolean variables x_i, $i \in \underline{Z}$, which take the value 1 if and only if vertex i is retained and the Boolean variables y_{ij} which take the value 1 if and only if the arc (i, j) is retained to build the arborescence. This results in program $\mathrm{P}_{3.7}$.

$$
\mathrm{P}_{3.7}: \begin{cases} \min \sum_{i \in \underline{Z}} c_i x_i & \\ \text{s.t.} \left| \begin{array}{lll} \sum_{i \in \underline{Z}_k} x_i \geq 1 & k \in \underline{S} & (3.7.1) \\ \sum_{j \in \mathrm{Adj}_i^-} y_{ji} = x_i & i \in \underline{Z}, i \neq \text{root} & (3.7.2) \\ \sum_{j \in \mathrm{Adj}_i^+} y_{ij} \leq d_i^+ x_i & i \in \underline{Z} & (3.7.3) \\ \sum_{(i,j) \in V^2 \cap U} y_{ij} \leq |V| - 1 & \forall V \subseteq \underline{Z} & (3.7.4) \\ x_i \in \{0,1\} & i \in \underline{Z} & (3.7.5) \\ y_{ij} \in \{0,1\} & (i,j) \in U & (3.7.6) \end{array} \right. \end{cases}
$$

According to constraints 3.7.2 and 3.7.6, if vertex i is not retained, then no arcs should arrive on this vertex. On the other hand, if this vertex is retained and is different from the root, then only one arc should arrive on this vertex. According to constraints 3.7.3 and 3.7.6, if vertex i is not retained, then no arc should start from this vertex. On the other hand, if this vertex is retained, then some arcs can start from this vertex and the corresponding constraint is inactive since their number is limited to the outdegree of this vertex, d_i^+. Finally, constraints 3.7.4 prohibit setting variables y_{ij} to values such that the selected arcs form a circuit.

The difficulty of $\mathrm{P}_{3.7}$ lies in the very large number of constraints 3.7.4. One method is to initially consider only a "small" subset of these constraints and solve the resulting program. If the solution of this program is an arborescence, it is an optimal solution. If, on the contrary, it is not an arborescence – the obtained graph contains circuits – then we add, for each circuit that is present, the corresponding constraint of type 3.7.4. The process is iterated until an arborescence is obtained. The process can be initialized by considering only constraints 3.7.4 corresponding to the prohibition of circuits of length 2.

3.8 Computational Experiments

In order to test the different formulations proposed in the previous sections and to solve the problem by directly using a solver of integer linear programs, we considered 6 different instances, I_1, I_2,..., I_6, generated as follows: the hypothetical candidate zones are represented by a grid of 10×10 identical square zones and 100 different species are concerned. The presence of each species in each zone is randomly determined, uniformly, with a probability equal to 0.06. With this choice of probabilities, some species appear in only one zone for the first 5 instances. These zones are therefore mandatory. The fact that some zones are mandatory generally facilitates the resolution of the problem. Indeed, this reduces its combinatorial aspect and also allows the use of a formulation that takes this information into account (see previous sections). In the sixth instance, no zone is mandatory. We consider here that the costs associated with each zone are identical. Therefore, in each formulation, the economic function $\sum_{i \in \underline{Z}} c_i x_i$ is replaced by the function $\sum_{i \in \underline{Z}} x_i$. The computation results are presented in table 3.1. Each line in the table corresponds to an instance. For each instance, the number of mandatory zones and the value of an optimal solution, *i.e.,* the minimal number of zones forming a connected reserve and allowing all the species to be protected, are given. The first 5 instances were resolved by the 7 formulations presented above, *i.e.,* by the 7 programs $\mathrm{P}_{3.1}$, $\mathrm{P}_{3.2}$, $\mathrm{P}_{3.3}$, $\mathrm{P}_{\text{approx}}$, $\mathrm{P}_{3.4}$, $\mathrm{P}_{3.5}$, and $\mathrm{P}_{3.6}$. The sixth instance was resolved by programs $\mathrm{P}_{3.1}$, $\mathrm{P}_{3.2}$, $\mathrm{P}_{\text{approx}}$, and $\mathrm{P}_{3.4}$ which do not require knowledge of at least one mandatory zone. For each program, table 3.1 presents the value of the continuous relaxation (relax), the total CPU time required to solve the program, expressed in seconds (cpu), and the number of nodes developed in the search tree (nodes). In the case of program $\mathrm{P}_{\text{approx}}$, table 3.1 also presents the value of the solution obtained (value) which may differ from the value of an optimal solution since this program only provides an approximate solution to the problem.

Table 3.1 shows that, when we take into account the fact that at least one zone is mandatory (programs $\mathrm{P}_{3.3}$, $\mathrm{P}_{3.5}$, and $\mathrm{P}_{3.6}$), the fastest exact resolution, on average, is performed by program $\mathrm{P}_{3.3}$. We also note that $\mathrm{P}_{3.4}$ and $\mathrm{P}_{3.6}$ do not allow all the instances considered to be solved in one hour of computation. With respect to the exact resolution, regardless of the fact that some zones are mandatory (programs $\mathrm{P}_{3.1}$, $\mathrm{P}_{3.2}$, and $\mathrm{P}_{3.4}$), the formulation $\mathrm{P}_{3.2}$ is the fastest, on average. With respect to the approximate resolution, $\mathrm{P}_{\text{approx}}$ provides a very fast resolution – 20 s on average – and the obtained solution is optimal for 4 instances out of 6. In case it is not optimal – instances I_3 and I_6, bolded values – the difference is only one unit. Table 3.1 also shows that computation times are highly dependent on the instance. Not surprisingly, instances with many mandatory zones are generally easier to resolve, but it is also observed that instance I_5, which has only one mandatory zone, is resolved quickly – compared to the other instances. The values of the continuous relaxations are comparable for the 7 formulations. The relative difference between the value of the continuous relaxation and the optimal value is relatively large – about 26% on average – but varies little from one instance to another. Note that for the approximate resolution – program $\mathrm{P}_{\text{approx}}$ – even if some zones are mandatory, the root of the searched arborescence is not fixed because this could prevent, in some cases, to find the optimal solution. Indeed, let us consider the sub-graph induced by a connected set of zones. This graph contains a set of arborescences, E, and a set of arborescences, $E' \subset E$, when one of the zones is chosen as a root. It is possible that all the arborescences of E' contain paths of two arcs of type $\{(i-1,j),(i,j),(i,j-1)\}$ or

TAB. 3.1 – Determination of a connected reserve, at a lower cost, to protect the 100 species present in a set of 100 zones represented by a grid of 10×10 square and identical zones. Resolution of the problem by 7 different formulations and for 6 instances. The rarity of certain species implies that certain zones must be necessarily protected in the case of the first 5 instances.

Instance	Number of mandatory zones	Optimal value	$P_{3.1}$			$P_{3.2}$		
			Relax	CPU	Nodes	Relax	CPU	Nodes
I_1	2	26	17.6	1,879	1,885,010	17.6	1,418	1,257,615
I_2	6	28	22.2	34	41,402	22.2	22	28,386
I_3	3	26	19.8	98	145,239	19.8	34	36,497
I_4	2	27	18.8	2,443	2,803,734	18.8	1,600	955,498
I_5	1	24	19.7	36	43,535	19.7	25	29,428
I_6	0	26	18.5	2,634	2,551,301	18.5	1,455	1,136,704
Average				1,187	1,245,037		759	574,021

Instance	$P_{3.3}$			P_{approx}			
	Relax	CPU	Nodes	Relax	CPU	Nodes	Value
I_1	18.1	176	198,916	17.1	28	32,910	26
I_2	22.4	44	56,637	22.2	1	381	28
I_3	20.0	32	26,406	20.0	9	3,983	**27**
I_4	19.2	279	334,209	18.9	34	39,680	27
I_5	20.2	12	12,386	19.7	3	1,484	24
I_6				18.5	47	49,749	**27**
Average		109	125,711		20	21,365	

Instance	$P_{3.4}$			$P_{3.5}$			$P_{3.6}$		
	Relax	CPU	Nodes	Relax	CPU	Nodes	Relax	CPU	Nodes
I_1	17.1	*(26/25)	*704,410	17.1	903	367,950	17.1	*(26/23)	*5,420
I_2	21.7	181	42,241	21.9	4	1,392	21.9	1,599	3,221
I_3	19.3	55	16,564	19.3	33	10,571	19.4	*(32/25)	*6,133
I_4	18.2	3,102	620,832	18.2	662	206,956	18.3	*(29/24)	4,307
I_5	19.5	52	13,967	19.6	40	15,691	19.6	1,413	2,989
I_6	18.1	*(27/25)	*579,915						
Average					328	120,512			

*The optimal solution could not be obtained in one hour of computation. In this case, the number of nodes indicated is the number of nodes developed during this computation time. In the column "cpu", the value of the best solution found followed by the corresponding lower bound is indicated in parentheses. For example, program $P_{3.6}$ did not resolve instance I_1 in one hour of computing. At the end of this computation time, 5,420 nodes have been developed in the search tree, the best solution found has a value of 26 and we are sure that the optimal solution has a value of at least 23.

$\{(k-1,l),(k,l),(k,l+1)\}$ while some arborescences of E do not contain these types of path. Consider instance I_2 of table 3.1. In this instance, the 6 zones z_{27}, z_{37}, z_{38}, z_{43}, z_{66}, and $z_{10,6}$ are mandatory. Figure 3.6a shows the solution obtained for this instance with $P_{3.2}$. Zone z_{98} was selected during the resolution of $P_{3.2}$ as the root of the searched arborescence. Figure 3.6b shows the solution obtained with $P_{3.3}$, for the same instance, by first setting zone $z_{10,6}$ as the root of the sought arborescence.

Consider instance I_3 of table 3.1. In this instance, 3 zones are mandatory: z_{18}, z_{21}, and z_{31}. Figure 3.7 shows the solution obtained for this instance with P_{approx}. Zone z_{95} was chosen during the resolution of P_{approx} as the root of the searched arborescence. The value of this approximate solution is equal to 27 while the value of the optimal solution is equal to 26.

Limits of the method. To find out the limitations of the method we tested programs $P_{3.3}$ and P_{approx} on instances of size 15 × 15 with 100 species. To generate these instances, the probability of each species being present in each zone is now 0.025. The results obtained, by limiting the computation time to one hour, are presented in table 3.2.

The optimal solutions of $P_{3.3}$ and P_{approx} could not be obtained in one hour of computation for any of the 5 instances except in one case: P_{approx} solved instance IV in 1,920 s of computation. Table 3.2 presents, for each instance, the value of the best feasible solution (Value) and the best lower bound obtained by the solver within one hour of computation (Lower bound). For example, for instance I, the best connected reserve found by $P_{3.3}$ – to protect all the species – has 45 zones and the optimal connected reserve has at least 38 zones. The best connected reserve found by P_{approx} has 43 zones and the best connected reserve that could be obtained by resolving

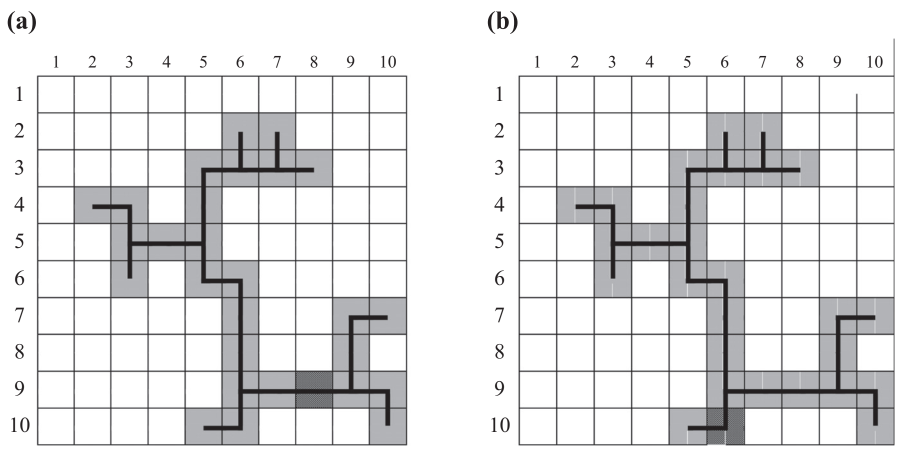


FIG. 3.6 – (a) Optimal solution of instance I_2 of table 3.1 provided by $P_{3.2}$. The resolution of $P_{3.2}$ fixed zone z_{98} as the root of the arborescence. (b) Optimal solution of instance I_2 of table 3.1 provided by the resolution of $P_{3.3}$. Among the 6 mandatory zones, $z_{10,6}$ has been prefixed as the root of the arborescence. The two optimal reserves are identical.

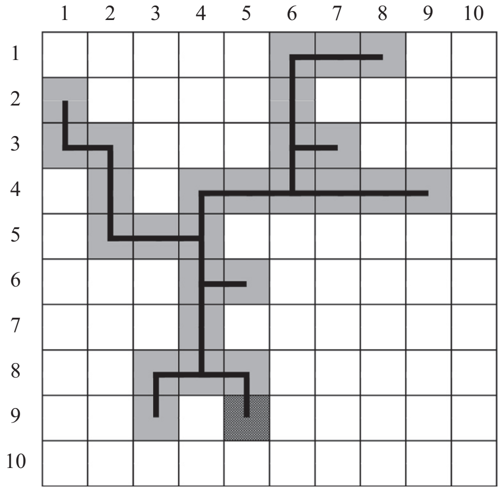


FIG. 3.7 – Approximate solution of instance I_3 of table 3.1 provided by P_{approx}. In this instance, zones z_{18}, z_{21}, and z_{31} are mandatory. The resolution of P_{approx} set zone z_{95} as the root of the arborescence.

TAB. 3.2 – Determination of a connected reserve, at a lower cost, to protect the 100 species present in a set of 225 zones represented by a grid of 15 × 15 square and identical zones. The exact solution of the problem is searched by program $P_{3.3}$ and the approximate one by program P_{approx}. The rarity of some species implies that certain zones must be necessarily protected.

Instance	# mandatory zones	$P_{3.3}$		P_{approx}	
		Lower bound	Value	Lower bound	Value
I	1	38	45	41	43
II	4	40	54	44	47
III	2	37	52	41	44
IV	1	38	47	43	43 (1,920 s)
V	2	38	60	43	44

P_{approx} has at least 41 zones. Table 3.2 also shows that, for the 5 instances considered and within one hour of computation, the best reserve is obtained by solving program P_{approx}.

References and Further Reading

Billionnet A. (2012) Designing an optimal connected nature reserve, *Appl. Math. Modell.* **36**, 2213.
Billionnet A. (2016) Designing connected and compact nature reserves, *Environ. Model. Assess.* **21**, 211.

Bondy J.A., Murty U.S.R. (1976) *Graph theory with applications.* Elsevier.

Briers R.A. (2002) Incorporating connectivity into reserve selection procedures, *Biol. Conserv.* **103**, 77.

Bunn A.G., Urban D.L., Keitt T.H. (2000) Landscape connectivity: a conservation application of graph theory, *J. Environ. Manage.* **59**, 265.

Carvajal R., Constantino M., Goycoolea M., Vielma J.P., Weintraub A. (2013) Imposing connectivity constraints in forest planning models, *Oper. Res.* **61**, 824.

Cerdeira J.O., Gaston K.J., Pinto L.S. (2005) Connectivity in priority area selection for conservation, *Environ. Model. Assess.* **10**, 183.

Cerdeira J.O., Pinto L.S. (2005) Requiring connectivity in the set covering problem, *J. Combin. Optim.* **9**, 35.

Cerdeira J.O., Pinto L.S., Cabeza M., Gaston K.J. (2010) Species specific connectivity in reserve-network design using graphs, *Biol. Conserv.* **143**, 408.

Conrad J.M., Gomes C.P., van Hoeve W.J., Sabharwal A., Suter J.F. (2012) Wildlife corridors as a connected subgraph problem, *J. Environ. Econ. Manage.* **63**, 1.

Crooks K.R., Sanjayan M., Eds (2008) *Connectivity conservation.* Cambridge University Press.

Dilkina B., Gomes C.P. (2010) Solving connected subgraph problems in wildlife conservation, CPAIOR-10: International Conference on Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems. Bologna, Italy.

Fahrig L., Merriam H.G. (1985) Habitat patch connectivity and population survival, *Ecology* **66**, 1762.

Fischer J., Lindenmayer D.B. (2007) Landscape modification and habitat fragmentation: a synthesis, *Global Ecol. Biogeo.* **16**, 265.

Garfinkel R.S. (1985) Motivation and modeling, *The traveling salesman problem* (E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan, D.B. Shmoys, Eds). John Wiley & Sons, Ch. 2.

Haight R.G., Snyder S.A. (2009) Integer programming methods for reserve selection and design, *Spatial conservation prioritization* (A. Moilanen, K.A. Wilson, H.P. Possingham, Eds) Oxford University Press.

Hanski I. (1999) Habitat connectivity, habitat continuity, and metapouplations in dynamic landscapes, *Oikos* **77**, 209.

Hilty J.A., Lidicker W.Z. Jr., Merenlender A.M. (2006) *Corridor ecology, the science and practice of linking landscapes for biodiversity conservation.* Island Press.

Jafari N., Hearne J. (2013) A new method to solve the fully connected reserve network design problem, *Eur. J. Oper. Res.* **231**, 202.

Jafari N., Nuse B.L., Moore C.T., Dilkina B., Hepinstall-Cymerman J. (2017) Achieving full connectivity of sites in the multiperiod reserve network design problem, *Comp. Oper. Res.* **81**, 119.

McDonnell M., Possingham H., Ball I., Cousins E. (2002) Mathematical methods for spatially cohesive reserve design, *Environ. Model. Assess.* **7**, 107.

Moilanen A., Nieminen M. (2002) Simple connectivity measures in spatial ecology, *Ecology* **83**, 1131.

Moilanen A., Wilson K.A., Possingham H.P., Eds. (2009) *Spatial conservation prioritization.* Oxford University Press.

Önal H., Briers R.A. (2006) Optimal selection of a connected reserve network, *Oper. Res.* **54**, 379.

Shirabe T. (2005) A model of contiguity for spatial unit allocation, *Geographical Analy.* **37**, 2.

Shirabe T. (2009) Districting modeling with exact contiguity constraints, *Environ. Plan. B: Plan. Des.* **36**, 1053.

Taylor P.D., Fahrig L., Henein K., Merriam G. (1993) Connectivity is a vital element of landscape structure, *Oikos* **68**, 571.

Tischendorf L., Fahrig L. (2000) How should we measure landscape connectivity? *Landscape Ecol.* **15**, 633.

Urban D., Keitt T. (2001) Landscape connectivity: a graph theoretic perspective, *Ecology* **82**, 1205.

Wang Y., Önal H. (2011) Designing connected nature reserve networks using a graph theory approach, *Acta Ecologica Sinica* **31**, 225.

Wang Y., Önal H. (2013) Designing a connected nature reserve using a network flow theory approach, *Acta Ecologica Sinica* **33**, 253.
Williams J.C. (2002) A zero-one programming model for contiguous land acquisition, *Geogr. Anal.* **34**, 330.
Williams J.C., ReVelle C.S., Levin S.A. (2005) Spatial attributes and reserve design models: a review, *Environ. Model. Assess.* **10**, 163.
Worboys G.L., Francis W.L., Lockwood M. (2010) *Connectivity conservation management: a global guide* Earthscan.

Chapter 4

Compactness

4.1 Introduction

As we have seen in the previous chapters, the spatial configuration of a nature reserve is a determining factor for the survival of the species that live there. Chapter 2 deals with fragmentation and chapter 3 with connectivity – or contiguity. In this chapter, we discuss another spatial aspect of a reserve, compactness. This aspect, which can be assessed in several ways and is not completely distinct from the notion of fragmentation, takes into account the distance separating the different zones. The smaller these distances, the easier it is for species to move within the reserve. It can therefore be said that the more compact a reserve is, the more effective the means devoted to its protection. On the other hand, a compact reserve is generally easier to manage than a non-compact reserve.

We will also look at the length of the reserve boundaries (which we also call the reserve perimeter), *i.e.*, the length of the transition zones between the reserve and the surrounding matrix, this criterion being, in a way, related to compactness (figure 4.1).

4.2 Compactness Measures of a Reserve

The compactness of a reserve, *i.e.*, a set of zones selected for some protection, can be assessed in many ways. For example, the diameter of the reserve, *i.e.*, the maximal distance between two points of the reserve, can be considered. The minimization of this criterion leads to the selection of a set of zones with an external contour that is close to a circle. The minimal distance between two zones can also provide some information on compactness. Another measure of the reserve compactness is its total perimeter. Minimizing this latter criterion allows to obtain groups of zones whose shape is close to a square or a circle, but the distance between the groups is not controlled. The compactness of a reserve can also be measured by the sum of the

DOI: 10.1051/978-2-7598-2521-9.c004

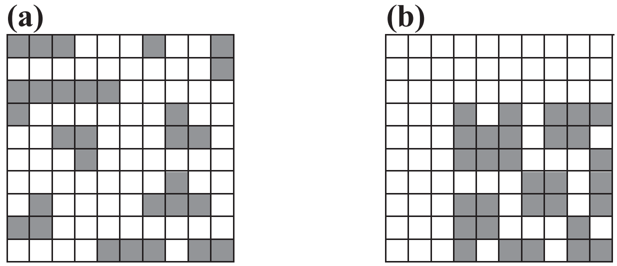


FIG. 4.1 – Two reserves defined on a set of 100 candidate zones represented by a grid of 10×10 square and identical zones whose length of the sides is equal to one unit. The area of these 2 reserves – made up of grey zones – is equal to 30 units. (a) A very fragmented and uncompact reserve with a perimeter of 80 units. (b) A fragmented but relatively compact reserve with a perimeter of 64 units.

distances between all the pairs of selected zones or by its total perimeter divided by its total area. In the latter case, the aim is to minimise the value of the corresponding ratio. Minimizing this ratio also has the effect of reducing the edge effect, which is generally considered as unfavourable to biodiversity protection.

Let $S = \{s_1, s_2, \ldots, s_m\}$ be the set of concerned species, $Z = \{z_1, z_2, \ldots, z_n\}$ be the set of zones that can be protected, $\underline{S}$ be the set of indices of the species of S and $\underline{Z}$ be the set of indices of the zones of Z. Let us specify some supplementary data – and corresponding notations – that we use in this chapter: l_i, the perimeter of zone z_i, a_i, the area of zone z_i, l_{ij}, the length of the border common to zones z_i and z_j, and d_{ij}, the distance "as the crow flies" between zones z_i and z_j. We denote by $\text{Comp}(R)$ the compactness of a reserve, R, and we examine several measures of this compactness.

It is assumed here that we know, among the zones of Z, those whose protection leads to the protection of species s_k (*e.g.*, its survival), and this for all species, *i.e.*, for all $k \in \underline{S} = \{1, 2, \ldots, m\}$. This subset of Z is denoted by Z_k and the corresponding set of indices is denoted by $\underline{Z_k}$. Thus, the protection of species s_k is ensured if and only if at least one of the zones of Z_k is protected and the number of protected species is noted $\text{Nb}_1(R)$ (chapter 1, section 1.1). For example, it is considered here that the protection of a zone allows all the species present in that zone to be protected, provided that their population size is greater than or equal to a certain threshold value. This value is denoted by v_{ik} for zone z_i and species s_k. In other words, $Z_k = \{z_i \in Z : n_{ik} \geq v_{ik}\}$ where n_{ik} refers to the population size of species s_k in zone z_i.

4.3 Some Problems of Selecting Compact Reserves and their Mathematical Programming Formulation

As with the other spatial criteria we examined, several zone selection problems may arise with the objective of obtaining a compact set of zones. We note R the set of selected zones, *i.e.*, the reserve obtained, and $\underline{R}$ the set of indices of the zones forming reserve R. Some of these problems are discussed below, by way of example. Let us recall that we denote by $\mathrm{Comp}(R)$ the value of the compactness criterion associated with reserve R. As we have seen, this criterion can correspond to different aspects of the compactness of a reserve – and can, therefore, be calculated in several different ways. Table 4.1 summarizes the compactness criteria considered in this chapter and that we seek to minimize.

To simplify the presentation, we assume that the set of candidate zones are represented by a grid with nr rows and nc columns. The zones are then designated by z_{ij} where i is the row index and j is the column index. It should be noted that everything presented in the rest of this chapter can be applied directly to any other set of candidate zones.

Take again the reserves in figure 4.1 to illustrate these 3 criteria. First of all, the distance between two candidate zones – represented by two squares whose length is equal to one unit – is defined by the distance, in a straight line, separating the centre of these two zones. Other definitions of the distance between two zones could be considered. Let us compare the compactness of the two reserves in figure 4.1 using the 3 criteria in table 4.1. Using criterion No. 1, the compactness of the reserve in figure 4.1a is equal to the distance between the centres of zones $z_{1,1}$ and $z_{10,10}$, *i.e.*, 12.73 ($\sqrt{162}$) while the compactness of the reserve in figure 4.1b is equal to the distance between zones $z_{4,4}$ and $z_{10,10}$, *i.e.*, 8.49 ($\sqrt{72}$). Using criterion No. 2, the compactness of the reserve in figure 4.1a is equal to 2.67 since its total perimeter is equal to 80 and its total area to 30, while the compactness of the reserve in figure 4.1b is equal to 2.13 since its total perimeter is equal to 64 and its total area to 30. Finally, using criterion No. 3, the compactness of the reserve in figure 4.1a is equal to 2,514.79 while the compactness of the reserve in figure 4.1b is equal to 1,716.43.

Different problems of determining an optimal compact reserve can be considered. For each of these problems, the 3 compactness criteria that we have just defined can be taken into account. Table 4.2 presents the 3 problems we will consider. Recall that $\mathrm{Nb}_1(R)$ represents the number of species protected by reserve R and that species s_k is considered to be protected by reserve R if at least one of the zones of Z_k belongs to R where, for any k of $\underline{S}$, Z_k is a known subset of Z (see section 4.2).

TAB. 4.1 – Three compactness criteria for a reserve, R.

Criterion number	Statement	Formulation
1	Diameter of the reserve, *i.e.*, maximal distance between two zones of the reserve.	$\max\{d_{ij} : (i,j) \in \underline{R}^2, i<j\}$
2	Total perimeter of the reserve, divided by the total area of the reserve.	$\left(\sum_{i\in\underline{R}} l_i - 2\sum_{(i,j)\in\underline{R}^2,\, i<j} l_{ij}\right) \Big/ \sum_{i\in\underline{R}} a_i$
3	Sum of the distances between all pairs of zones in the reserve.	$\sum_{(i,j)\in\underline{R}^2,\, i<j} d_{ij}$

TAB. 4.2 – Three reserve selection problems with a compactness objective.

Problem no.	Problem statement	Formulation
I	Selection of a reserve, R, of minimal cost, allowing to protect, at a minimum, a given number of species, Ns, and such that the compactness indicator, Comp(R), is lower than or equal to a given value, ρ.	$\begin{cases} \min \quad C(R) \\ \text{s.t.} \left\| \begin{array}{l} R \subseteq Z \\ \mathrm{Nb}_1(R) \geq \mathrm{Ns} \\ \mathrm{Comp}(R) \leq \rho \end{array} \right. \end{cases}$
II	Selection of a reserve, R, of cost less than or equal to a given value, B, allowing the greatest possible number of species to be protected, and whose value of the compactness indicator, Comp(R), is less than or equal to a given value, ρ.	$\begin{cases} \max \quad \mathrm{Nb}_1(R) \\ \text{s.t.} \left\| \begin{array}{l} R \subseteq Z \\ C(R) \leq B \\ \mathrm{Comp}(R) \leq \rho \end{array} \right. \end{cases}$
III	Selection of a reserve, R, of cost less than or equal to a given value B, allowing to protect, at a minimum, a given number of species, Ns, and minimizing the value of the compactness indicator, Comp(R).	$\begin{cases} \min \quad \mathrm{Comp}(R) \\ \text{s.t.} \left\| \begin{array}{l} R \subseteq Z \\ C(R) \leq B \\ \mathrm{Nb}_1(R) \geq \mathrm{Ns} \end{array} \right. \end{cases}$

Also remember that $C(R)$ refers to the cost of reserve R: $C(R) = \sum_{i \in \underline{R}} c_i$ where c_i is the cost associated with zone z_i.

4.3.1 Problem I: Protection, at the Lowest Cost, of at Least Ns Species of a Given Set, with a Compactness Constraint

This problem can be formulated as the mathematical program $\mathrm{P}_{4.1}$ in which ρ designates the value that the compactness indicator of the selected reserve must not exceed. As in all the programs we have studied, the Boolean variable x_i is equal to 1 if and only if zone z_i is selected to form the reserve.

$$\mathrm{P}_{4.1}: \begin{cases} \min \quad \sum\limits_{i \in \underline{Z}} c_i x_i \\ \text{s.t.} \left| \begin{array}{llll|llll} y_k \leq \sum\limits_{i \in \underline{Z}_k} x_i & k \in \underline{S} & (4.1.1) & & x_i \in \{0,1\} & i \in \underline{Z} & (4.1.4) \\ \sum\limits_{k \in \underline{S}} y_k \geq \mathrm{Ns} & & (4.1.2) & & y_k \in \{0,1\} & k \in \underline{S} & (4.1.5) \\ \mathrm{Comp}(R) \leq \rho & & (4.1.3) & & & & \end{array} \right. \end{cases}$$

The economic function to be minimized represents the cost of the reserve. According to constraints 4.1.1, variable y_k can take the value 1 if and only if at least one zone in Z_k is protected. Constraint 4.1.2 expresses that the number of protected species must be greater than or equal to Ns. Constraint 4.1.3 imposes a compactness index less than or equal to ρ. Note that if we seek to protect all the species – Ns $= m$ – we can replace constraints 4.1.1 and 4.1.2 by the single family of constraints $\sum_{i\in\underline{Z}_k} x_i \geq 1,\ \ k \in \underline{S}$. If we want to obtain, among the optimal solutions of $\mathrm{P}_{4.1}$, a solution that maximizes the number of protected species, we only need to subtract from the economic function to be minimized the quantity $\varepsilon \sum_{k\in\underline{S}} y_k$ where ε is a sufficiently small constant. Similarly, if one wants to obtain, among the optimal solutions of $\mathrm{P}_{4.1}$, a solution that minimizes the value of the compactness criterion, it is sufficient to add to the economic function to be minimized the quantity $\varepsilon\mathrm{Comp}(R)$ where ε is a sufficiently small constant. Let us now study constraint 4.1.3 according to the criterion retained to measure compactness. Recall that $R = \{z_i : i = 1,..., n;\ x_i = 1\}$.

Criterion No. 1. The compactness of a reserve is measured by the diameter of the reserve, *i.e.*, by the maximal distance between two zones of the reserve (see appendix at the end of the book). In this case, $\mathrm{P}_{4.1}$ solves the problem by replacing the – generic – constraint 4.1.3 with one of the specific constraint sets $\mathrm{C}_{4.1}$ or $\mathrm{C}_{4.2}$:

$$\mathrm{C}_{4.1} : x_i + x_j \leq 1 \quad (i, j) \in \underline{Z}^2,\ i<j,\ d_{ij} > \rho;$$

$$\mathrm{C}_{4.2} : x_i + \sum_{j\in\underline{Z},\, j\, >\, i,\, d_{ij}\, >\, \rho} x_j \leq 1 + M(1 - x_i) \quad i \in \underline{Z}.$$

Constraints $\mathrm{C}_{4.1}$ express that if the distance between any two zones, z_i and z_j, is greater than ρ then these two zones cannot both be part of the reserve. In other words, in this case, variables x_i and x_j cannot simultaneously take the value 1. According to constraints $\mathrm{C}_{4.2}$, if zone z_i is selected – $x_i = 1$ – then none of the zones located at a distance greater than ρ from z_i can belong to the reserve. In case zone z_i is not retained – $x_i = 0$ – the corresponding constraint is inactive provided that the constant M is chosen large enough.

Criterion No. 2. Let us now consider the case where the compactness of a reserve, R, is measured by the total perimeter of the reserve divided by its total area: $\mathrm{Comp}(R) = \left(\sum_{i\in\underline{R}} l_i - 2\sum_{(i,j)\in\underline{R}^2, i<j}\ l_{ij}\right) \Big/ \sum_{i\in\underline{R}}\ a_i$. In this case, the problem considered can be solved by program $\mathrm{P}_{4.1}$ by replacing the – generic – constraint 4.1.3 by the specific constraint $\mathrm{C}_{4.3}$:

$$\mathrm{C}_{4.3} : \frac{\sum_{i\in\underline{Z}} l_i x_i - 2\sum_{(i,j)\in\underline{Z}^2, i<j}\ l_{ij} x_i x_j}{\sum_{i\in\underline{Z}}\ a_i x_i} \leq \rho.$$

The perimeter of the reserve is calculated by summing the perimeters of all the zones that constitute the reserve, $\sum_{i\in\underline{Z}} l_i x_i$, and by subtracting twice the sum of the lengths of the borders common to each pair of zones of the reserve, $2\sum_{(i,j)\in\underline{Z}^2,\, i<j} l_{ij}x_i x_j$. Note that, in the latter expression, many terms l_{ij} are equal to 0. The quantity $\sum_{i\in\underline{Z}} a_i x_i$ represents the sum of the areas of each zone constituting the reserve, *i.e.*, the total area of the reserve. Constraint $C_{4.3}$ is equivalent to constraint $C_{4.4}$ in which the first member is quadratic and the second member is linear:

$$\mathrm{C}_{4.4} : \sum_{i\in\underline{Z}} l_i x_i - 2 \sum_{(i,j)\in\underline{Z}^2, i<j} l_{ij} x_i x_j \leq \rho \sum_{i\in\underline{Z}} a_i x_i.$$

It is possible to replace $C_{4.4}$ with equivalent linear constraints (see appendix at the end of the book). To do this, each product $x_i x_j$ is replaced in $C_{4.4}$ by variable u_{ij} and 2 families of linearization constraints are added to force variables u_{ij} to be equal to the products $x_i x_j$, at the optimum of the obtained program. Finally, the problem can be solved by program $P_{4.1}$ by replacing the – generic – constraint 4.1.3 by the set of specific constraints $C_{4.5}$:

$$\mathrm{C}_{4.5} : \begin{cases} \sum_{i\in\underline{Z}} l_i x_i - 2 \sum_{(i,j)\in\underline{Z}^2, i<j} l_{ij} u_{ij} \leq \rho \sum_{i\in\underline{Z}} a_i x_i \\ u_{ij} \leq x_i \quad (i,j) \in \underline{Z}^2, i<j,\ l_{ij} > 0 \\ u_{ij} \leq x_j \quad (i,j) \in \underline{Z}^2, i<j,\ l_{ij} > 0 \end{cases}.$$

Note that if the compactness criterion is the perimeter of the reserve and not the perimeter-to-area ratio, it is sufficient to replace the first constraint of $C_{4.5}$ by the constraint $\sum_{i\in\underline{Z}} l_i x_i - 2\sum_{(i,j)\in\underline{Z}^2,i<j} l_{ij}u_{ij} \leq \rho$ where ρ now refers to the maximal allowed perimeter.

Criterion No. 3. Let us now consider the case where the compactness of a reserve is measured by the sum of the distances between all the pairs of zones of the reserve: $\mathrm{Comp}(R) = \sum_{(i,j)\in\underline{R}^2,\, i<j} d_{ij}$. In this case, the problem considered can be solved by program $P_{4.1}$ by replacing the – generic – constraint 4.1.3 by the specific constraint $C_{4.6}$:

$$\mathrm{C}_{4.6} : \sum_{(i,j)\in\underline{Z}^2, i<j} d_{ij} x_i x_j \leq \rho.$$

Constraint $C_{4.6}$ is quadratic since it involves the products $x_i x_j$. As in the case of $C_{4.3}$, it is possible to replace this constraint by a set of linear constraints (see appendix at the end of the book). To do this, each product $x_i x_j$ is replaced in $C_{4.6}$ by variable u_{ij} and 2 sets of linearization constraints are added to force variables u_{ij} to be equal to products $x_i x_j$ – at the optimum of the obtained program. Finally, the problem can be solved by program $P_{4.1}$ by replacing constraint 4.1.3 by the set of constraints $C_{4.7}$:

$$\mathrm{C}_{4.7}: \begin{cases} \sum\limits_{(i,j)\,\in\,\underline{Z}^2,\ i<j} d_{ij} u_{ij} \leq \rho \\ 1 - x_i - x_j + u_{ij} \geq 0 & (i,j) \in \underline{Z}^2, i<j \\ u_{ij} \geq 0 & (i,j) \in \underline{Z}^2, i<j \end{cases}.$$

4.3.2 Problem II: Protection, Under a Budgetary Constraint, of the Largest Possible Number of Species of a Given Set, with a Compactness Constraint

This problem can be formulated as the mathematical program $\mathrm{P}_{4.2}$.

$$\mathrm{P}_{4.2}: \begin{cases} \max \quad \sum\limits_{k \in \underline{S}} y_k \\ \text{s.t.} \left| \begin{array}{llll|lll} \sum\limits_{i \in \underline{Z}} c_i x_i \leq B & & (4.2.1) & & x_i \in \{0,1\} & i \in \underline{Z} & (4.2.4) \\ y_k \leq \sum\limits_{i \in \underline{Z}_k} x_i & k \in \underline{S} & (4.2.2) & & y_k \in \{0,1\} & k \in \underline{S} & (4.2.5) \\ \mathrm{Comp}(R) \leq \rho & & (4.2.3) & & & & \end{array} \right. \end{cases}$$

The economic function, to be maximized, expresses the number of protected species. Constraint 4.2.1 reflects the budgetary constraint: the cost associated with the reserve retained must not exceed the budget, B. For the meaning of the other constraints (4.2.2–4.2.5), the reader may refer to program $\mathrm{P}_{4.1}$. To solve the problem with the 3 compactness criteria considered, it is sufficient to replace in $\mathrm{P}_{4.2}$ constraint 4.2.3 by the appropriate constraints: $\mathrm{C}_{4.1}$ or $\mathrm{C}_{4.2}$ for criterion No. 1, $\mathrm{C}_{4.5}$ for criterion No. 2, and $\mathrm{C}_{4.7}$ for criterion No. 3 (see section 4.3.1). As in program $\mathrm{P}_{4.1}$, zone z_i belongs to reserve R if and only if $x_i = 1$.

4.3.3 Problem III: Protection, Under a Budgetary Constraint, of at Least Ns Species of a Given Set, with Optimal Compactness

This problem can be formulated as the mathematical program $\mathrm{P}_{4.3}$.

$$\mathrm{P}_{4.3}: \begin{cases} \min \quad \mathrm{Comp}(R) \\ \text{s.t.} \left| \begin{array}{llll|lll} y_k \leq \sum\limits_{i \in \underline{Z}_k} x_i & k \in \underline{S} & (4.3.1) & & x_i \in \{0,1\} & i \in \underline{Z} & (4.3.4) \\ \sum\limits_{i \in \underline{Z}} c_i x_i \leq B & & (4.3.2) & & y_k \in \{0,1\} & k \in \underline{S} & (4.3.5) \\ \sum\limits_{k \in \underline{S}} y_k \geq \mathrm{Ns} & & (4.3.3) & & & & \end{array} \right. \end{cases}$$

The economic function, to be minimized, expresses the compactness of the reserve. As in programs $\mathrm{P}_{4.1}$ and $\mathrm{P}_{4.2}$, zone z_i belongs to reserve R if and only if

$x_i = 1$. The meaning of the constraints in $P_{4.3}$ is presented in the two previous sections. Let us now look at how to solve the problem with the 3 compactness criteria considered.

Criterion No. 1. The compactness of a reserve is measured by the diameter of the reserve, *i.e.*, the maximal distance between two zones of the reserve. To solve the problem, variable α is introduced, the – generic – economic function of $P_{4.3}$ is replaced by α and the set of constraints $C_{4.8}$ is added:

$$C_{4.8} : \alpha \geq d_{ij} x_i x_j \qquad (i,j) \in \underline{Z}^2,\ i<j.$$

These constraints express that if zones z_i and z_j are retained – $x_i x_j = 1$ – then the value of variable α must be greater than or equal to the distance between these two zones. This results in a program with a linear economic function but with some quadratic constraints. These constraints can be replaced by the equivalent set of linear constraints $C_{4.9}$:

$$C_{4.9} : \begin{cases} \alpha \geq d_{ij}(-1 + x_i + x_j) & (i,j) \in \underline{Z}^2,\ i<j \\ \alpha \geq 0 \end{cases}.$$

Criterion No. 2. The compactness of reserve R is measured by the ratio of the total perimeter of the reserve divided by its total area. In this case, the problem considered can be solved by program $P_{4.3}$ by replacing the – generic – economic function of this program with the expression $\left(\sum_{i\in\underline{Z}} l_i x_i - 2\sum_{(i,j)\in\underline{Z}^2, i<j} l_{ij} x_i x_j\right) \Big/ \sum_{i\in\underline{Z}} a_i x_i$. The numerator of this expression is quadratic and the denominator is linear (see appendix at the end of the book). This expression can be transformed into a ratio of two linear functions by replacing each product $x_i x_j$ with variable u_{ij} and adding the linear constraints $u_{ij} \leq x_i$ and $u_{ij} \leq x_j$. The problem can thus be reformulated as program $P_{4.4}$.

$$P_{4.4} : \begin{cases} \min \ \left(\sum_{i\in\underline{Z}} l_i x_i - 2\sum_{(i,j)\in\underline{Z}^2,\, i<j} l_{ij} u_{ij}\right) \Big/ \sum_{i\in\underline{Z}} a_i x_i \\ \text{s.t.} \left| \begin{array}{lll} \sum\limits_{i\in\underline{Z}} c_i x_i \leq B & & (4.4.1) \\ y_k \leq \sum\limits_{i\in\underline{Z}_k} x_i & k \in \underline{S} & (4.4.2) \\ \sum\limits_{k\in\underline{S}} y_k \geq \mathrm{Ns} & & (4.4.3) \\ u_{ij} \leq x_i;\ u_{ij} \leq x_j & (i,j) \in \underline{Z}^2, i<j, l_{ij} > 0 & (4.4.4) \\ x_i \in \{0,1\} & i \in \underline{Z} & (4.4.5) \\ y_k \in \{0,1\} & k \in \underline{S} & (4.4.6) \end{array} \right. \end{cases}$$

We can therefore use the algorithms of fractional programming to solve $\mathrm{P}_{4.4}$, for example the Dinkelbach algorithm (see appendix at the end of the book). The core of this algorithm is to solve the auxiliary problem $\mathrm{P}_{4.5}(\lambda)$ which is a linear program in 0–1 variables.

$$\mathrm{P}_{4.5}(\lambda) : \begin{cases} \min \sum\limits_{i \in \underline{Z}} l_i x_i - 2 \sum\limits_{(i,j) \in \underline{Z}^2,\, i<j} l_{ij} u_{ij} - \lambda \sum\limits_{i \in \underline{Z}} a_i x_i \\ \text{s.t.} |\ (4.4.1) - (4.4.6) \end{cases}$$

Criterion No. 3. Finally, consider the case where the compactness of a reserve is measured by the sum of the distances between all the pairs of zones of the reserve: $\mathrm{Comp}(R) = \sum_{(i,j)\in \underline{R}^2,\, i<j} d_{ij}$. In this case, the problem considered can be solved by program $\mathrm{P}_{4.3}$ by replacing its – generic – economic function with the expression $\sum_{(i,j)\in \underline{Z}^2,\, i<j} d_{ij} x_i x_j$. We then obtain a mathematical program whose economic function is quadratic and whose constraints are linear (see appendix at the end of the book). One way to solve the resulting program is to linearize the economic function, and there are many techniques to do so. A simple technique that we have already presented consists of replacing products $x_i x_j$ by variables u_{ij} and adding the set of linear constraints $\mathrm{C}_{4.10}$ that force variable u_{ij} to be equal to product $x_i x_j$:

$$\mathrm{C}_{4.10} : \begin{cases} 1 - x_i - x_j + u_{ij} \geq 0 & (i,j) \in \underline{Z}^2, i<j \\ u_{ij} \geq 0 & (i,j) \in \underline{Z}^2, i<j \end{cases}.$$

Another technique consists in rewriting the economic function as $(1/2)\sum_{i\in\underline{Z}} x_i \sum_{j\in\underline{Z}} d_{ij} x_j$ then replacing, for any i of $\underline{Z}$, the expression $x_i \sum_{j\in\underline{Z}} d_{ij} x_j$ by the real, non-negative variable t_i. The new economic function – to be minimized – is therefore written $(1/2)\sum_{i\in\underline{Z}} t_i$. Then we must add the set of linear constraints $\mathrm{C}_{4.11}$ which force variable t_i to be equal, at the optimum, to the expression $x_i \sum_{j\in\underline{Z}} d_{ij} x_j$:

$$\mathrm{C}_{4.11} : \begin{cases} t_i \geq \sum_{j\in\underline{Z}} d_{ij} x_j - M(1 - x_i) \quad i \in \underline{Z} \\ t_i \geq 0 \quad i \in \underline{Z} \end{cases}.$$

If variable x_i is equal to 1 then, because of the first family of constraints of $\mathrm{C}_{4.11}$ and the fact that we seek to minimize the expression $\sum_{i\in\underline{Z}} t_i$, variable t_i takes the value $\sum_{j\in\underline{Z}} d_{ij} x_j$. On the contrary, if variable x_i is equal to 0, then the set of constraints $\mathrm{C}_{4.11}$ and the fact that we seek to minimize the expression $\sum_{i\in\underline{Z}} t_i$ force variable t_i to take the value 0. M denotes a sufficiently large constant.

4.4 Computational Experiments

A hypothetical set of candidate zones represented by a grid of 10×10 square and identical zones is considered. The cost associated with each zone of the grid is randomly drawn, in a uniform way, from the set $\{5, 6, \ldots, 10\}$ and is shown in

	1	2	3	4	5	6	7	8	9	10
1	10	7	7	9	7	6	10	5	9	6
2	9	6	8	10	6	9	7	7	10	9
3	7	7	8	8	8	5	5	9	8	7
4	5	10	8	9	8	4	6	9	7	4
5	7	5	8	7	5	8	5	10	9	9
6	7	10	4	10	9	6	6	4	7	5
7	7	5	5	10	5	9	5	10	7	7
8	9	6	7	9	6	8	8	6	6	5
9	8	9	8	10	6	6	5	9	5	4
10	7	8	5	5	7	9	4	7	8	7

FIG. 4.2 – Cost associated with each of the 100 candidate zones.

figure 4.2. The available budget is 150 units and 100 species are considered. The presence of a species in a given zone – with a sufficient abundance to be protected if the corresponding zone is protected – is randomly drawn, with a probability equal to 0.1. Figure 4.3 shows, for each candidate zone, the list of the species that are protected if the zone is itself protected. It should be noted that, in this example, 9 of the 100 species considered cannot be protected. These are species s_7, s_{13}, s_{16}, s_{23}, s_{33}, s_{61}, s_{73}, s_{83}, and s_{90}.

Table 4.3 presents the results obtained for Problem I of table 4.2 with the compactness criteria No. 1 and No. 2 of table 4.1 (see section 4.3.1). All instances were resolved in less than one second of computation. Some of the reserves obtained are shown in figure 4.4.

Table 4.4 presents the results obtained for Problem II of table 4.2 with the compactness criteria No. 1 and No. 2 of table 4.1 and when the available budget, B, is equal to 150 (see section 4.3.2). All the instances considered were resolved in less than one second of computation. Some of the reserves obtained are shown in figure 4.5.

Table 4.5 presents the results obtained for Problem III of table 4.2 with the compactness criteria No. 1 and No. 2 of table 4.1 and an available budget, B, equal to 150 (see section 4.3.3). All the instances considered were resolved in less than one second of computation. Some of the obtained reserves are presented in figure 4.6.

4.5 Compactness Measure Specific to a Connected Reserve: Protection of a Maximal Number of Species of a Given Set by a Connected and Compact Reserve, under a Budgetary Constraint

In this section, we focus on reserves that must, on the one hand, be connected and, on the other hand, meet a compactness criterion. In a connected reserve, the species can move through the whole reserve without leaving it (see chapter 3). With regard to compactness, we consider here a different measure from those studied in the

	1	2	3	4	5	6	7	8	9	10
1	15 75	4 11 18 69 98	3 5 85 89	-	14 31	47 80	34 85	1 22 38	35 37 77 88	27 87
2	4 79	28 92	31 42 65	35 37	6 25 68 84	19 32 63 99	11 25 40	29 53 85 86	34 51 99	95
3	42	17 63	-	20	41 99 100	21 44	50	-	2	-
4	8 40	-	-	67	63 95	6 26 79	9 44 54	-	8 40 54 58 69	43
5	98	35 66	36	46 72 77 87	25 76	18 32 53	88 96	53	69	26 44
6	52	67	28 82	-	44	27 37 41 88	19 66	84	5 59 64 88	57
7	71 85	43 48	76 78	10 37	-	46 59	14 62	10 12 27 59 91	15 25 59 60 70	87
8	69 85	41 45 80 86 91	46	48 49	30 39 94	46 76	8 14 97	3 70	-	19
9	63 99	12 24 39 64 68	2 5 14 25 88	98	26 94	99	76	18	8 50 72 81	75
10	3 47 71 74 97	81	17 58	48 93	55	19	12 60 85	53	18	56

FIG. 4.3 – For each of the 100 candidate zones, list of the indices of the species protected due to the protection of the zone. For example, the protection of zone z_{67} leads to the protection of species s_{19} and s_{66}.

previous sections and which can only be applied to connected reserves, in contrast to the 3 measures presented in table 4.1. To measure the compactness of such reserves, we define the distance between two zones z_i and z_j as the shortest distance to travel from zone z_i to zone z_j without leaving the reserve. As announced, this definition of the distance between two zones of a reserve implies that this reserve is connected, in contrast to the definition of the distance between two zones used in the compactness measures No. 1 and No. 3 presented in table 4.1. The zones outside the connected and compact reserves in which we are interested should also form a set of connected zones. In other words, it must be possible to cross all the areas outside the reserve without crossing the reserve in order to protect it from external disturbances. The problem considered is to select a set of zones included in the set of candidate zones, $Z = \{z_1, z_2, \ldots, z_n\}$, to constitute a connected reserve that maximizes the number of species protected by this reserve while respecting a compactness criterion and budgetary constraint. The compactness criterion used here is described in detail

TAB. 4.3 – Results obtained for Problem I of table 4.2 with the compactness criteria No. 1 and No. 2 of table 4.1, and for the instance described in figures 4.2 and 4.3. We consider 2 different values of the minimal number of species to be protected, Ns, 30 and 60. For each pair (Ns, No. of the compactness criterion) considered, we study 2 values of the compactness criterion, ρ, that should not be exceeded.

Problem I							
Ns	No. of the compactness criterion considered	ρ	Number of selected zones	Cost of the selected reserve	Number of protected species	Actual value of the criterion	Associated figure
30	1	4	9	62	30	3.6	4.4a
		7	8	48	30	6.1	4.4b
	2	0.9	20	125	31	0.9	–
		1.5	10	68	30	1.5	–
60	1	4	–	–	–	–	–
		7	26	170	60	6.7	–
	2	0.9	32	219	60	0.9	4.4c
		1.5	24	159	60	1.5	4.4d

– No feasible solution.

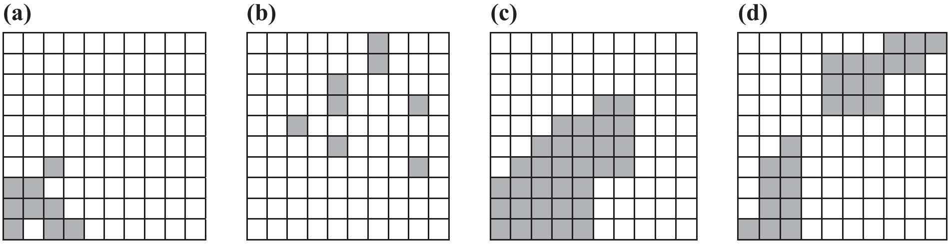


FIG. 4.4 – Obtained reserves for 4 instances of Problem I (see table 4.3).

below. It can be assumed that the non-selected zones, *i.e.*, those located outside the reserve, will be used for urban, industrial or agricultural development. We are interested in a set, $S = \{s_1, s_2, \ldots, s_m\}$, of rare or threatened species present in these zones. For each zone z_i we know the list of the species present in this zone and, for each species, its population size. The population size of species s_k in zone z_i is denoted by n_{ik} and reserve, R, is considered to protect species s_k if and only if the total population size of species s_k in this reserve is greater than or equal to a certain threshold value, θ_k. Thus, the interest in protecting reserve, R, is measured by the quantity $\mathrm{Nb}_2(R)$ which expresses the number of species whose total population size in the reserve is greater than or equal to the threshold value (chapter 1, section 1.1). It is assumed that the movements of the species under consideration are only

TAB. 4.4 – Results obtained for Problem II of table 4.2 with the compactness criteria No. 1 and No. 2 of table 4.1, for the instance described in figures 4.2 and 4.3, and when the value of the available budget, B, is equal to 150. Two values of ρ, the compactness criterion value that should not be exceeded are studied for each considered compactness criterion.

Problem II						
No. of the compactness criterion considered	ρ	Number of selected zones	Actual cost of the selected reserve	Number of protected species	Actual criterion value	Associated figure
1	4	14	101	35	3.6	4.5a
	7	23	149	57	6.7	4.5b
2	0.9	20	142	41	0.9	–
	1.5	24	150	57	1.5	–

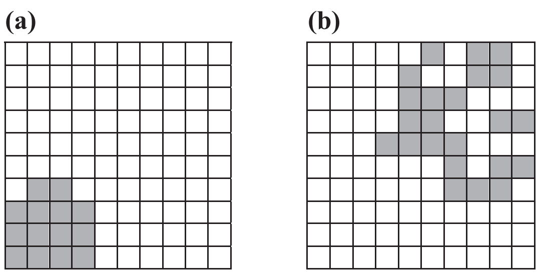


FIG. 4.5 – Obtained reserves for 2 instances of Problem II (see table 4.4).

TAB. 4.5 – Results obtained for Problem III of table 4.2 with the compactness criteria No. 1 and No. 2 of table 4.1, for the instance described in figures 4.2 and 4.3, and when the value of the available budget, B, is equal to 150. We consider 2 different values, 30 and 60, of the minimal number of species to be protected, Ns.

Problem III						
Ns	No. of the compactness criterion considered	Number of selected zones	Actual cost of the selected reserve	Number of protected species	Value of the criterion	Associated figure
30	1	11	82	30	3.2	4.6a
	2	23	148	31	0.9	4.6c
60	1	25	150	60	7.3	4.6b
	2	22	150	60	1.7	4.6d

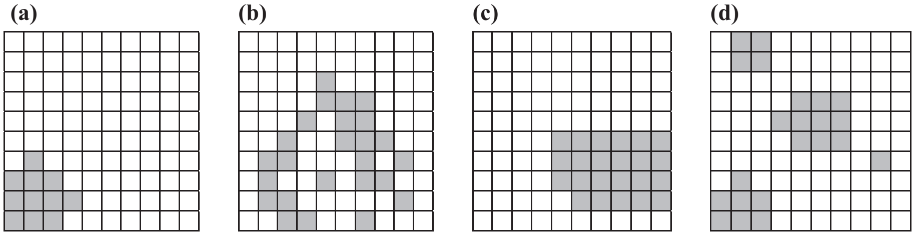


FIG. 4.6 – Obtained reserves for 4 instances of Problem III (see table 4.5).

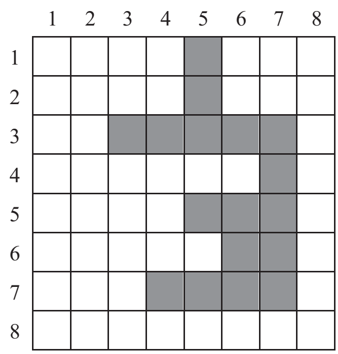


FIG. 4.7 – The candidate zones form a grid of dimension 8×8. Each zone is a square whose side length is equal to one unit. Among the 64 candidate zones, 17 are selected and form a connected reserve.

possible within the reserve – the outside of the reserve being a too hostile environment – and these movements are made from a zone of the reserve to an adjacent zone of this reserve. This notion of adjacency is considered to be the same for all species considered. To facilitate the presentation of the examples, all the candidate zones form a grid and two zones are considered adjacent if they share a common side. It is then assumed, for measuring the length of a route, that the movements are made gradually from the centre of a zone to the centre of an adjacent zone. We are looking for a connected and compact reserve. In such a reserve, thanks to connectivity, the species can circulate throughout the reserve without leaving it (figure 4.7) and, thanks to compactness, the distance they have to travel within the reserve to get from one zone to another is not too long. Let us now look at the precise definition of compactness that we have chosen here.

Compactness. The compactness indicators usually use the Euclidean distance between zones. This is the case for the criteria No. 1 and No. 3 of section 4.3, and also for a measure of the compactness of a reserve equal to the radius of the smallest circle containing the whole reserve (*e.g.*, all the centres of each zone). On the reserve in figure 4.8, we see that this radius is equal to $\sqrt{8}$ and the centre of the corresponding circle is located in the centre of the zone located at the intersection of row 5 and column 5. This structural measure may not be relevant from a functional point of view. Indeed, two zones may be relatively close as the crow flies but distant when trying to travel from one to the other without leaving the reserve. This is the case for the two light grey zones of the reserve shown in figure 4.8. The distance to be covered to join these two zones is equal to 2 units – assuming that the distance between two adjacent zones is equal to one unit – but the associated route leaves the reserve. On the other hand, the minimal distance to be covered to join these two zones without leaving the reserve is equal to 14 units. We will therefore use a more realistic measure of compactness than the radius of the smallest circle containing the whole reserve. Denote by R the reserve, *i.e.*, the set of zones that constitute it. Let $d_{ij}(R)$ be the minimal distance that the species must cover to get from zone z_i to zone z_j without leaving the reserve. To each zone z_i of the reserve R, its eccentricity is associated. It is denoted by $\mathrm{ecc}(z_i,R)$ and defined as follows: $\mathrm{ecc}(z_i, R) = \max\{d_{ij}(R) : z_j \in R\}$. The eccentricity of zone z_i in reserve R is therefore equal to the distance – defined above – between zone z_i and the zone which is furthest from z_i. Finally, we define the compactness of a reserve by the minimal value of this quantity, *i.e.*, $\mathrm{Comp}(R) = \min\{\mathrm{ecc}(z_i, R) : z_i \in R\}$. This measure of the compactness of R is also called the radius of R. With this definition, the compactness of the reserve shown in figure 4.8 is equal to 7 – the eccentricity of the zone located at the intersection of row 5 and column 6.

Connectivity of the zones outside the reserve. When a set of zones is selected to form a reserve, it may be desirable, in order to minimize disturbance of the reserve, to be able to move through all the zones not belonging to the reserve without crossing it. Indeed, some species and/or their habitats can be very sensitive to human presence. This is the case, for example, for plant species that are damaged by trampling (*e.g.*, plants to stabilize dunes), animals whose normal behaviour is easily

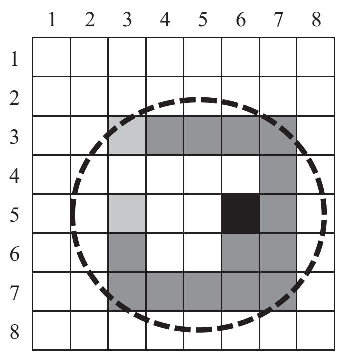


FIG. 4.8 – The candidate zones form a grid of dimension 8 × 8. Each zone is a square whose side length is equal to one unit. Among the 64 candidate zones, 17 are selected and form a connected reserve. The structural distance between the 2 zones z_{33} and z_{53} is equal to 2 units while the functional distance between these two zones is equal to 14 units.

disturbed, or species that are particularly affected by introduced diseases or invasive species. The zones that do not belong to the reserve are made up of unselected candidate zones to which one or more zones representing the territory located outside the candidate zones are added. It is therefore necessary that this set of zones that do not belong to the reserve be connected. It is assumed here that off-reserve movements, such as movements within the reserve, can only be made by gradually moving from one zone to an adjacent one. Consider figure 4.9 where the candidate zones are represented by a grid of 64 square and identical zones. The selected reserve contains 18 grey zones. Non-reserve zones are those zones of the grid that are not selected, to which a zone representing the outside of the grid is added. Zones of the grid that touch the outside of the grid are considered to be adjacent to the zone representing the outside of the grid. The reserve shown in this figure is not an admissible reserve because it is impossible, for example, to join the two zones marked with a "x" without crossing the reserve.

On the other hand, the reserve shown in figure 4.10 is admissible since it is possible to move to all the zones outside the reserve – including the zone outside the grid – without crossing the reserve.

Some definitions of graph theory (see appendix at the end of the book). Let $G = (V, E)$ be a connected graph where $V = \{v_1, v_2, \ldots, v_n\}$ is the set of vertices and E is the set of edges. Denote by d_{ij} the length of the minimal length chain connecting vertices v_i and v_j. The eccentricity of vertex v_i, $\mathrm{ecc}(v_i)$, is equal to the quantity $\max_{j:v_j \in V} d_{ij}$. The centre of G is, by definition, a vertex of minimal eccentricity and its eccentricity is called the radius of the graph. In other words, the radius is the smallest possible value that satisfies the following property: the distance between a vertex – to be determined – and any other vertex is less than or equal to this value. A connected graph has one or more centres. The graph in figure 4.11 includes 3 centres, v_2, v_5, and v_8.

A tree is a connected graph without cycles. To work on a tree, it can be interesting to particularize one of its vertices to make it a root. For tree $\mathcal{A}$ of root r, the father of vertex v is the vertex adjacent to v and belonging to the chain connecting

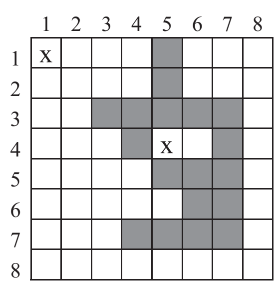


FIG. 4.9 – The 64 candidate zones form a grid of size 8 × 8. 18 zones are selected to form a connected reserve. This reserve is not admissible because the zones that are not affected to the reserve are not all connected.

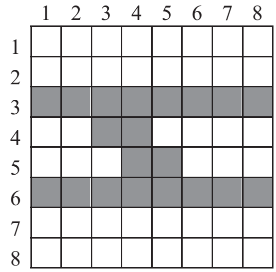


FIG. 4.10 – The 64 candidate zones form a grid of size 8 × 8. 20 zones are selected to form a connected reserve. This reserve is admissible because the zones that are not affected to the reserve are all connected – possibly through the zone outside the grid.

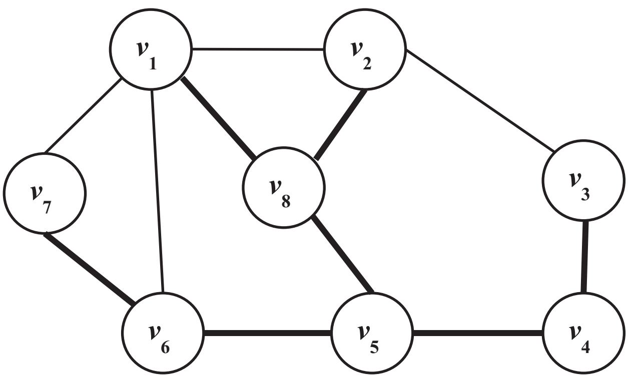


FIG. 4.11 – A connected graph, of radius 2, whose centres are v_2, v_5, and v_8. The edges drawn in bold define a spanning tree of height 2 if v_5 is selected as the root of this tree.

the root to v. Root r is the only vertex of $\mathcal{A}$ without a father. The sons of a vertex v are the vertices adjacent to v that are not the father of v. A leaf of $\mathcal{A}$ is a vertex without sons and its degree is therefore equal to 1. The height of $\mathcal{A}$, that we denote by $h(\mathcal{A})$, is the length of the chain of maximal length that connects the root to a leaf. If all the edges of $\mathcal{A}$ are transformed into arcs oriented from the chosen root, r, towards the leaves, we obtain an arborescence. A spanning tree of $G = (V, E)$, is a tree whose all edges belong to G, and which connects – covers or spans – all the vertices of G. A spanning tree of G is therefore a tree, $\mathcal{A} = (V, E_{\mathcal{A}})$, such that $E_{\mathcal{A}} \subseteq E$. An induced sub-graph of G is a sub-graph of G defined by a subset of vertices of V. More precisely, H is an induced sub-graph of G if, for any pair of vertices $\{v_i, v_j\}$ of H, v_i is connected to v_j by an edge of H if and only if v_i is connected to v_j by an edge of G.

The above definitions allow us to state Property 4.1 below that we use in the formulation of the problem.

Property 4.1. The radius of a connected graph G is less than or equal to ρ if and only if G admits a spanning tree with a height less than or equal to ρ.

Proof. If G admits a spanning tree of height less than or equal to ρ, then the eccentricity of its root, in G, is less than or equal to ρ, and the radius of G is, therefore, itself less than or equal to ρ. Conversely, if the radius of G is less than or equal to ρ, the graph composed of the shortest chains connecting a centre of G to all the other vertices of G is, by definition, a spanning tree with a height less than or equal to ρ.

Expression of the problem in terms of graphs. Let us associate to the set of candidate zones $Z=\{z_1, z_2,\ldots, z_n\}$ a non-oriented graph whose vertices are $\underline{Z} = \{1, 2, \ldots, n\}$ and such that there is an edge between vertex i and vertex j if two zones z_i and z_j are adjacent. Defining a reserve whose compactness is less than or equal to ρ consists in selecting a subset of zones, *i.e.*, a subset of vertices of the graph associated with the candidate zones, such that the sub-graph induced by this subset is connected and with a radius less than or equal to ρ. With each vertex of the graph – candidate zone – is associated a cost and the cost of a sub-graph is equal, by definition, to the sum of the costs of its vertices.

The problem can then be formulated as follows: determine a subset of vertices, with a cost less than or equal to the available budget, B, which induces a connected sub-graph of radius less than or equal to ρ and which allows the greatest possible number of species to be protected. Using property 4.1 above, the problem can be reformulated as follows: determine a subset of vertices, of cost less than or equal to the available budget, B, which admits a spanning tree with a height less than or equal to ρ, and which allows the greatest possible number of species to be protected. The consideration of the connectivity constraint for zones not belonging to the reserve is considered later in section 4.5.2.

4.5.1 Mathematical Programming Formulation

The Boolean variables t_{ih}, $i \in \underline{Z}, h = 1,\ldots, \rho+1$, are used, which take the value 1 if and only if zone z_i is selected and assigned to level h of the searched spanning tree. Level 1 corresponds to the root of the tree and the vertices of level $h+1$ are connected to the root by a chain of length h. Thus, $t_{ih} = 1$ implies that there is a chain, of length less than or equal to $h-1$, from z_i to the root and passing only through the selected vertices. We also use the Boolean variables y_k, $k \in \underline{S}$, which take the value 1 if and only if species s_k is protected by the selected reserve. In other words, and taking into account the conditions for a species to be protected, $y_k = 1$ if and only if the total population size of species s_k present in the zones selected to form the reserve is greater than or equal to the threshold value, θ_k. To simplify the presentation, we also use the working Boolean variables x_i which can be simply expressed as a function of variables t_{ih} and which take the value 1 if and only if zone

z_i is affected to the reserve. We can now formulate the problem as the linear program in Boolean variables $P_{4.6}$.

$$
P_{4.6}: \begin{cases}
\max \sum_{k \in \underline{S}} w_k y_k & \\
\text{s.t.} \left| \begin{array}{lll}
x_i = \sum_{h=1}^{\rho+1} t_{ih} & i \in \underline{Z} & (4.6.1) \\
\theta_k y_k \leq \sum_{i \in \underline{Z}} n_{ik} x_i & k \in \underline{S} & (4.6.2) \\
\sum_{i \in \underline{Z}} t_{i1} = 1 & & (4.6.3) \\
t_{ih} \leq \sum_{j \in \mathrm{Adj}_i} t_{j,h-1} & i \in \underline{Z}, h = 2, \ldots, \rho+1 & (4.6.4) \\
\sum_{i \in \underline{Z}} c_i x_i \leq B & & (4.6.5) \\
x_i \in \{0,1\} & i \in \underline{Z} & (4.6.6) \\
t_{ih} \in \{0,1\} & i \in \underline{Z},\ h = 1, \ldots, \rho+1 & (4.6.7) \\
y_k \in \{0,1\} & k \in \underline{S} & (4.6.8)
\end{array} \right.
\end{cases}
$$

The economic function measures the weighted number of protected species – w_k refers to the weight associated with species s_k. Indeed, due to constraints 4.6.2, the Boolean variable y_k is necessarily equal to 0 if the total population size of species s_k in the reserve is lower than the threshold value, θ_k. Otherwise, it takes the value 1 – at the optimum of $P_{4.6}$ – since the aim is to maximize the value of the economic function. Constraints 4.6.1 express variables x_i as a function of variables t_{ih}: $x_i = 1$ if and only if zone z_i is assigned to one (and only one) of the $\rho + 1$ levels of the tree. Constraint 4.6.3 corresponds to the choice of the root vertex: the root must be chosen in one and only one vertex. According to constraints 4.6.4, if zone z_i is selected and assigned to level h of the tree, then at least one of its adjacent zones must be selected and assigned to level h–1. Adj_i refers to all the indices of the zones adjacent to zone z_i. Constraint 4.6.5 expresses the budgetary constraint. Finally, constraints 4.6.6–4.6.8 specify the Boolean nature of all variables in program $P_{4.6}$.

4.5.2 Connectivity of Zones Outside the Reserve

To determine an optimal reserve that takes this constraint into account, we can proceed as follows: (1) solve $P_{4.6}$ without taking it into account, (2) if the constraint is satisfied by the solution obtained, then this solution is the optimal solution to the problem, (3) if not, solve $P_{4.6}$ again, but with an additional constraint to prohibit the obtained configuration. The process is repeated until an admissible reserve is obtained. Computational experiments have shown that, at least under our experimental conditions, an admissible reserve is obtained directly – *i.e.*, by solving $P_{4.6}$

once – in more than one case out of three and that if this is not the case, a few iterations are sufficient to obtain an admissible solution. Let us look more precisely at a way of implementing point (3). Given a reserve, a set of "isolated" zones is defined as a set of candidate zones, not belonging to the reserve, in one piece – connected – that cannot be reached from outside the grid without crossing the reserve and that is maximal in the inclusion sense. A set of "isolating" zones is associated with any set of isolated zones as follows: any reserve containing all the zones of the set of isolating zones is admissible only if it contains all the zones of the set of isolated zones. In fact, we associate to any subset of isolated zones a set of isolating zones, minimal in the inclusion sense. If Z^a denotes a set of isolated zones and Z^b, the set of isolating zones associated with Z^a, then any admissible reserve must satisfy constraint $\mathrm{C}_{4.11}$. It should be noted that the reserve that had been obtained is no longer admissible.

$$\mathrm{C}_{4.11}: \sum_{i\,:\,z_i \in Z^a} x_i \geq |Z^a| \left(1 - \sum_{i:\,z_i \in Z^b} (1 - x_i)\right).$$

Indeed, if all the zones of the set of isolating zones are selected, constraint $\mathrm{C}_{4.11}$ becomes $\sum_{i\,:\,z_i \in Z^a} x_i \geq |Z^a|$ and imposes that all the zones of the set of isolated zones are selected in the reserve. On the contrary, if $q\ (\geq 1)$ zones of the set of isolating zones are not selected, this constraint becomes $\sum_{i\,:\,z_i \in Z^a} x_i \geq |Z^a|\,(1 - q)$ and is then inactive – always satisfied.

In summary, if the solution obtained by $\mathrm{P}_{4.6}$ corresponds to a reserve with at least one set of isolated zones, it is necessary to add constraint $\mathrm{C}_{4.11}$ to the program and solve it again. The process must then be iterated – keeping the constraints already added – until a reserve is obtained without a set of isolated zones. Let us again take the example of figure 4.9 and assume that the resolution of $\mathrm{P}_{4.6}$ results in the reserve of this figure. It is then necessary to add to $\mathrm{P}_{4.6}$ the constraint $x_{45} + x_{46} \geq 2\,(1 + x_{35} + x_{36} + x_{44} + x_{47} + x_{55} + x_{56} - 6)$.

4.5.3 Computational Experiments

In order to test the effectiveness of the approach, we considered different instances of the problem and solved them with the mathematical program $\mathrm{P}_{4.6}$. We considered hypothetical instances constructed from a set of zones forming a grid of 20×20 identical square zones whose length of sides is equal to one unit, and 100 hypothetical species, which are divided into 3 groups whose weight in the economic function is equal to 10, 5, and 1, respectively.

- Group I (species numbered 1 to 20): These species are rare; they are present in only 10% of the candidate zones and their presence is randomly selected.
- Group II (species numbered 21 to 50): These species are relatively rare; they are present in only 20% of the candidate zones and their presence is randomly selected.

- Group III (species numbered 51 to 100): These species are relatively common; they are present in 30% of the candidate zones and their presence is randomly selected.

For each species present in a zone, its population size in that zone is chosen at random according to the uniform law, between 5 and 10 units. In order to simplify the presentation, the minimal size of the total population necessary for the survival of each of the 100 species considered is set to 25. The distance between two adjacent zones is equal to the distance between their centres, *i.e.*, one unit. The costs associated with each zone are randomly selected, according to the uniform law, between 5 and 20 units. With regard to compactness, 5 values of ρ are considered, 4, 5, 6, 7, and 8, and for each of these values, 3 values of the available budget, B, are considered, 150, 300, and 450. We also randomly select 3 zones that must necessarily belong to the reserve and, on the contrary, 20 zones that cannot be included in it. With these values, the reserves obtained allow for the protection of 0–100 species and the value of the economic function – the weighted number of protected species – varies from 0 to 400. The computation results are presented in table 4.6. All the instances considered could be solved. When $\rho = 3$, there are no admissible reserves, regardless of the value of B considered, and the resolution of $P_{4.6}$ for these 3 values of B requires less than one second of computation. As expected, the resolution of $P_{4.6}$ is very fast for small radius values, ρ, and slower for large values. Indeed, the number of Boolean variables t_{ijh} – associated with zone z_{ij} and level h – increases rapidly with the value of ρ; in our experiments, this number is equal to $400(\rho + 1)$. Several hours of computation are required to solve the problem when $\rho = 8$ and $B = 300$. It can also be seen that, for any fixed value of the radius, the CPU time increases with the budget up to a certain value and then decreases. The resolution of program $P_{4.6}$ provides, for 9 instances out of 15, a reserve with at least one enclave – zones outside the reserve and isolated. The results presented in table 4.6 show that, for these 9 instances, only a few iterations are sufficient to determine a reserve without an enclave. They also show that taking into account the connectivity constraint for the zones outside the reserve deteriorates only slightly the value of the economic function. Figure 4.12a shows the reserve obtained by solving $P_{4.6}$ with $\rho = 6$ and $B = 300$ without taking into account the connectivity constraint for the zones outside the reserve. We see that zone z_{66} is isolated. Figure 4.12b shows the optimal reserve without an enclave. Respecting the "no enclave" constraint does not significantly penalize the value of the solution: it decreases by only 5 units out of 310 – less than 2%. On the other hand, the structure of the reserve has been profoundly modified. A table such as table 4.6 can help a decision-maker to choose the level of compactness of the envisaged reserve. In this example, if he/she can only use a budget of 150 units, he/she can afford to look for a very compact reserve. Indeed, in this case, the compactness constraint does not influence the optimal value of the weighted number of protected species. In contrast, if he/she has a larger budget, such as 300 units, he/she must deal with a compromise between the compactness and the weighted number of protected species.

TAB. 4.6 – Results obtained by solving program $P_{4.6}$ for hypothetical instances, constructed from a grid 20×20 with 100 hypothetical species, when $\theta_k = 25$ for all $k \in \underline{S}$.

ρ	B	Number of protected species in each group (I, II, III)	Economic function value	CPU time (s)	Presence of enclaves in the obtained reserve	Number of iterations to obtain a reserve without enclaves	Final value of the economic function	Additional CPU time (s)
4	150	1, 11, 37	102	1	No	–	–	–
	300	7, 25, 50	245	1	Yes	4	235	3
	450	8, 30, 50	280	1	Yes	2	270	1
5	150	1, 11, 37	102	3	No	–	–	–
	300	10, 28, 50	290	11	Yes	1	290	4
	450	16, 30, 50	360	1	Yes	3	350	3
6	150	1, 11, 37	102	8	No	–	–	–
	300	11, 30, 50	310	174	Yes	2	305	292
	450	20, 30, 50	400	3	Yes	5	390	28
7	150	1, 11, 37	102	24	No	–	–	–
	300	13, 28, 50	320	1,046	Yes	1	320	881
	450	20, 30, 50	400	3	Yes	2	400	7
8	150	1, 11, 37	102	207	No	–	–	–
	300	13, 30, 50	330	10,267	Yes	3	325	27,292
	450	20, 30, 50	400	5	No	–	–	–

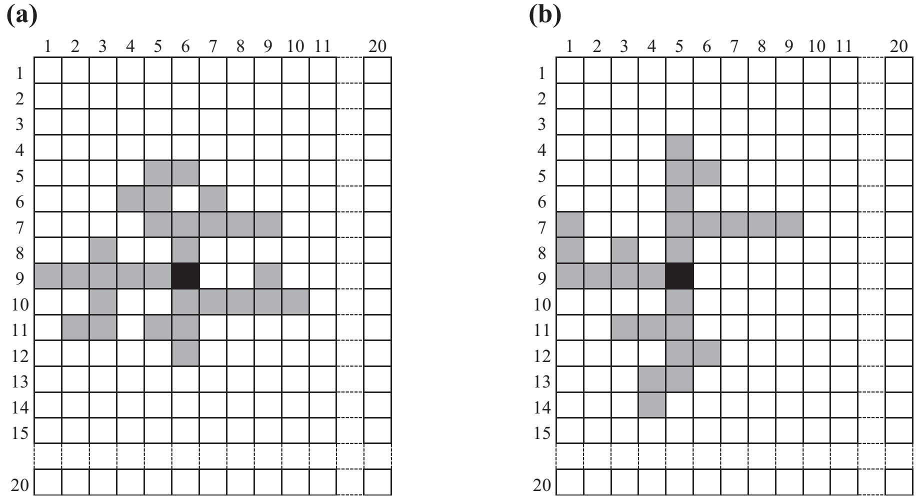


FIG. 4.12 – Two optimal connected reserves, with a radius of 6 or less and a cost of 300 or less. (a) The centre of the reserve is zone z_{96}. 11 species in Group I and all the species in Groups II and III are protected. The value of the solution is 310 but zone z_{66}, which does not belong to the reserve, is isolated by zones z_{56}, z_{67}, z_{76}, and z_{65}. (b) The centre of the reserve is zone z_{95}. 11 species in Group I and all the species in Groups II and III, except one species in Group II, are protected. The value of the solution is 305 and there is no enclave in this reserve.

References and Further Reading

Alagador D. (2011) Quantitative methods in spatial conservation planning. Thesis, Instituto Superior de Agronomia, Universidade Técnica de Lisboa.

Billionnet A. (2012) Designing an optimal connected nature reserve, *Appl. Math. Model.* **36**, 2213.

Billionnet A. (2016) Designing connected and compact nature reserves, *Environ. Model. Assess.* **21**, 211.

Briers R.A. (2002) Incorporating connectivity into reserve selection procedures, *Biol. Conserv.* **103**, 77.

Carvajal R., Constantino M., Goycoolea M., Vielma J.P., Weintraub A. (2013) Imposing connectivity constraints in forest planning models, *Oper. Res.* **61**, 824.

Cerdeira J.O., Gaston K.J., Pinto L.S. (2005) Connectivity in priority area selection for conservation, *Environ. Model. Assess.* **10**, 183.

Cerdeira J.O., Pinto L.S., Cabeza M., Gaston K.J. (2010) Species specific connectivity in reserve-network design using graphs, *Biol. Conserv.* **143**, 408.

Conrad J.M., Gomes C.P., van Hoeve W.J., Sabharwal A., Suter J.F. (2012) Wildlife corridors as a connected subgraph problem, *J. Environ. Econ. Manage.* **63**, 1.

Fischer D.T., Church R.L. (2003) Clustering and compactness in reserve site selection: an extension of the biodiversity management area selection model, *For. Sci.* **49**, 555.

Fischer D.T., Church R.L. (2005) The SITES reserve selection system: a critical review, *Environ. Model. Assess.* **10**, 215.

Groeneveld R.A. (2010) Species-specific spatial characteristics in reserve site selection, *Ecol. Econ.* **69**, 2307.

Haight R.G., Snyder S.A. (2009) Integer programming methods for reserve selection and design, *Spatial conservation prioritization* (A. Moilanen, K.A. Wilson, H.P. Possingham, Eds). Oxford University Press.

Helzer C.J., Jelinski D.E. (1999) The relative importance of patch area and perimeter-area ratio to grassland breeding birds, *Ecol. Appl.* **9**, 1448.

Jafari N., Hearne J. (2013) A new method to solve the fully connected reserve network design problem, Eur. *J. Oper. Res.* **231**, 202.

Lindenmayer D., *et al.* (2008) A checklist for ecological management of landscapes for conservation, *Ecol. Lett.* **11**, 78.

Marianov V., ReVelle C., Snyder S. (2008) Selecting compact habitat reserves for species with differential habitat size needs, *Comput. Oper. Res.* **35**, 475.

McDonnell M., Possingham H., Ball I., Cousins E. (2002) Mathematical methods for spatially cohesive reserve design, *Environ. Model. Assess.* **7**, 107.

Moilanen A. (2005) Methods for reserve selection: interior point search, *Biol. conserv.* **124**, 485.

Moilanen A., Wilson K.A., Possingham H.P., Eds. (2009) *Spatial conservation prioritization.* Oxford University Press, Newyork.

Nalle D.J., Arthur J.L., Montgomery C.A., Sessions J. (2002) Economic and spatial impacts of an existing reserve network on future augmentation, *Environ. Model. Assess.* **7**, 99.

Nalle D.J., Arthur J.L., Sessions J. (2002) Designing compact and contiguous reserve networks with a hybrid heuristic algorithm, *For. Sci.* **48**, 59.

Önal H., Briers R.A. (2002) Incorporating spatial criteria in optimum reserve network selection, *Proc. R. Soc. London* **B 269**, 2437.

Önal H., Briers R.A. (2005) Designing a conservation reserve network with minimal fragmentation: a linear integer programming approach, *Environ. Model. Assess.* **10**, 193.

Önal H., Briers R.A., (2006) Optimal selection of a connected reserve network, *Oper. Res.* **54**, 379.

Urban D., Keitt T. (2001) Landscape connectivity: a graph theoretic perspective, *Ecol.* **82**, 1205.

Vogiatzis C., Veremyev A., Pasiliao E.L., Pardalos P.M. (2014) An integer programming approach for finding the most and the least central cliques, *Optim. Lett.* **9**, 615.

Walteros J.L., Vogiatzis C., Pasiliao E.L., Pardalos P.M. (2014) Integer programming models for the multidimensional assignment problem with star costs, *Eur. J. Oper. Res.* **235**, 553.

Wang Y., Önal H. (2013) Designing a conected nature reserve using a network flow theory approach, *Acta Ecologica Sinica* **33**, 253.

Wang Y., Fang Q., Dissanayake S.T.M., Önal H. (2020) Optimizing conservation planning for multiple cohabiting species, *PLoS ONE* **15**, e0234968.

Williams J.C., ReVelle C.S., Levin S.A. (2005) Spatial attributes and reserve design models: a review, *Environ. Model. Assess.* **10**, 163.

Williams J.C. (2008) Optimal reserve site selection with distance requirements, *Comput. Oper. Res.* **35**, 488.

Chapter 5

Other Spatial Aspects

5.1 Introduction

Chapters 2, 3 and 4 cover the most common spatial aspects involved in reserve design: fragmentation, connectivity, and compactness. In this chapter, we are interested in other spatial aspects involved in the design and management of a reserve. In section 5.2, we distinguish, among the zones of a reserve, those that can be considered as belonging to a central part and those that can be considered as belonging to a buffer part. We also associate two types of species with this kind of reserve: those that can live either in the central part or in the buffer part and those that can only live in the central part. In sections 5.3 and 5.4, we are interested in protecting species living in forests benefiting from a degree of management by also distinguishing several types of species: those that live mainly in forest parcels where the wood has been harvested – cut – those that live mainly on the edge, *i.e.*, at the border between a cut and an uncut parcel, and those that live in an uncut parcel network. Finally, in section 5.5, we look at the classic problem of selecting zones from a set of candidate zones in order to constitute a reserve, but here the definition of the candidate zones implies that not all of these zones are necessarily two-by-two disjoined. The protection of one zone can, therefore, automatically lead to the protection of part of another zone.

5.2 Reserve with Central Part and Buffer Part

In this section, we consider reserves formed by two parts: a "central" part and a "non-central" part. A zone of the reserve is said to be in the central part if it is "far enough" from the outside of the reserve. This central part is thus protected from the negative effects of activities outside the reserve. It can also simply, by being far enough from the outside of the reserve, ensures a certain climate in this part (heat, sunshine, humidity). The central part, which can be considered as the core of the

DOI: 10.1051/978-2-7598-2521-9.c005

reserve, therefore has no common border with the outside of the reserve. To be ecologically effective, taking into account the conservation objectives, the non-central part, called the buffer part, must be large enough. The distance that must be maintained between the part considered as central and the outside of the reserve depends on (1) the species that are to be protected in that central part, (2) the nature of the activities that take place outside the reserve, and (3) the ability of the buffer part to protect the central part from the effects of these activities. For example, the topography of the reserve can be taken into account. The buffer part completely surrounds the central part. The biodiversity protection within the buffer part itself may be relatively limited. On the other hand, this buffer part, which provides additional protection for the zones of the central part, may be fundamental for the protection of biodiversity in the central part. It should be noted that certain activities incompatible with the protection of biodiversity in the central part may be authorized in the buffer part.

Example 5.1. Let us consider a set of candidate zones represented by a grid of 15×15 square and identical zones whose side lengths are equal to one unit. Figure 5.1 shows 2 reserves, defined on this grid, with a central part and a buffer part. In both cases, according to its definition, the central part never touches the outside of the reserve but, in case (b), the buffer part is larger than in case (a). Indeed, in case (a) the smallest distance separating a point of the reserve from a point outside it is equal to one unit. In case (b), this distance is equal to two units. The area of the reserve allocated to the buffer part depends, as we have said, on the desired level of protection for the central part but also on the size and shape of the reserve. Proportionally, the area of the buffer part is larger for a small and/or non-compact reserve than for a large and/or compact reserve. Figure 5.2 illustrates that for two relatively compact reserves, the buffer part is proportionally smaller in a large reserve than in a small one. In case (a), the area of the buffer part represents about 50% of the reserve while in case (b), it represents 60%.

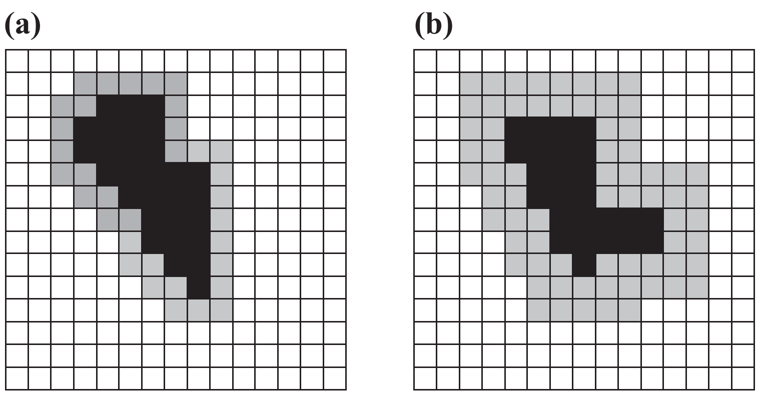


FIG. 5.1 – Two examples of reserves including a central part and a buffer part. The zones in the central part are shown in black and those in the buffer part are shown in grey. In case (b), the buffer part is larger than in case (a).

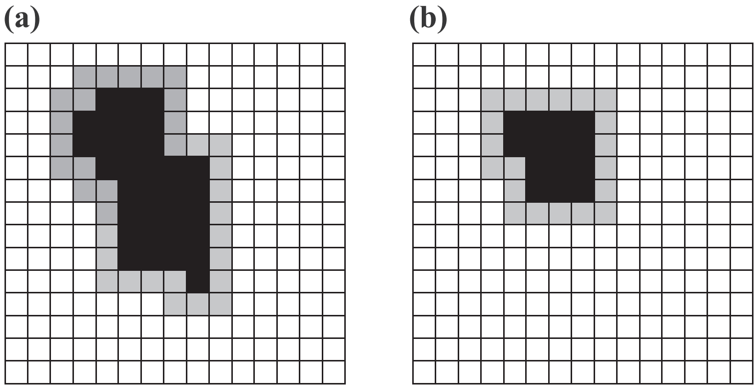


FIG. 5.2 – Two examples of reserves of different sizes including a central part and a buffer part. The zones in the central part are shown in black and those in the buffer part are shown in grey. Proportionally, the buffer part in case (b) is larger than in case (a).

From a species protection perspective, the zones in the central part of the reserve can be considered to be used to protect threatened species and the buffer part can be considered to only provide additional protection to the zones in the central part. It can also be considered that the characteristics of the buffer part are favourable to certain species and that these species are, therefore, protected if they live in the buffer part. These species can also generally be considered as protected if they live in the central part. In addition, as mentioned above, certain activities may be authorized in the buffer part, such as forest harvesting, environmentally friendly farming, and recreational activities.

5.2.1 Minimal Cost Protection of All the Considered Species

We examine here the selection of reserves with central parts and buffer parts. As in the previous chapters, we consider a set of candidate zones, $Z = \{z_1, z_2, \ldots, z_n\}$, and we denote by $\underline{Z}$ the set of corresponding indices. To simplify the presentation, a zone of the reserve is considered to be in the central part if it is completely surrounded by other zones of the reserve. It should be noted that the model studied in the following could very easily be adapted to different and/or more elaborate definitions of the central part. We also consider a set of species to be protected, $S = \{s_1, s_2, \ldots, s_m\}$. This set is divided into two groups, S_1 and S_2. To be protected, a species of the group S_1 must occur in at least one protected zone belonging to the central part and a species of the group S_2 must occur in at least one protected zone belonging either to the central part or to the buffer part. We denote by $\underline{S}_1$ and $\underline{S}_2$ the set of indices corresponding to the sets of species S_1 and S_2, respectively. As mentioned above, a selected zone is considered to be in the central part of the reserve if all the surrounding zones have also been selected, either in the central part or in the buffer part. Several problems may arise with regard to the protection of the species under consideration. The problem here is to determine a subset of zones, of minimal cost, that can protect all the species. For each zone, we know the list of the species present in that zone. If this zone is protected and is located in the central part, it is considered to ensure the protection of all the species of this list; if this zone is protected and is located in the buffer part, it is considered to only ensure the protection of the species of the list belonging to the group S_2. We denote by Z_k the set of zones hosting species s_k, and $\underline{Z}_k$, the set of corresponding indices.

5.2.2 Mathematical Programming Formulation

Consider the Boolean variable t_i, $i \in \underline{Z}$, which is equal to 1 if and only if zone z_i is selected and belongs to the central part of the reserve, and the Boolean variable x_i, which is equal to 1 if and only if zone z_i is selected and belongs either to the buffer part or to the central part. Thus, variable x_i is equal to 1 if and only if zone z_i is selected to form the reserve. It should be noted that the reserve may be made up of several separate "sub-reserves". In this case, there will be several central parts and several buffer parts. Let $\underline{ZC} \subseteq \underline{Z}$ be the set of indices of the candidate zones for the central parts and $\underline{I}_i \subseteq \underline{Z}$ the set formed by the index i and the set of indices of the neighbouring zones of zone z_i, and which must be selected, either in a central part or in a buffer part, to give zone z_i the status of a zone belonging to a central part. The problem can be formulated as the linear program in Boolean variables $\mathrm{P}_{5.1}$.

$$\mathrm{P}_{5.1}: \begin{cases} \min \sum_{i \in \underline{Z}} c_i x_i & \\ \text{s.t.} \left| \begin{array}{lll|lll} \sum_{i \in \underline{Z}_k} t_i \geq 1 & k \in \underline{S}_1 & (5.1.1) & t_i \in \{0,1\} & i \in \underline{ZC} & (5.1.4) \\ \sum_{i \in \underline{Z}_k} x_i \geq 1 & k \in \underline{S}_2 & (5.1.2) & x_i \in \{0,1\} & i \in \underline{Z} & (5.1.5) \\ t_i \leq x_j & i \in \underline{ZC}, j \in \underline{I}_i & (5.1.3) & & & \end{array} \right. \end{cases}$$

The economic function expresses the total cost of the reserve. Constraints 5.1.1 express that, for each species s_k of group S_1, at least one of the zones hosting that species must be, on the one hand, retained in the reserve and, on the other hand, located in the central part of that reserve. Constraints 5.1.2 express that, for each species s_k of group S_2, at least one of the zones hosting that species must be retained in the reserve, *i.e.*, located either in the central part or in the buffer part. Constraints 5.1.3 express that if zone z_i is retained to constitute the central part of the reserve then all the surrounding zones, *i.e.*, all zones z_j, $j \in \underline{I}_i$, must also be retained in the reserve, either in the central part or in the buffer part. Constraints 5.1.4 and 5.1.5 specify the Boolean nature of variables t_i and x_i.

5.2.3 Example

Consider a set of 100 candidate zones for protection represented by a grid of 10×10 square and identical zones (figure 5.3). It is considered here that a retained zone belongs to the central part of the reserve if and only if the 8 surrounding zones are also part of the reserve. This example concerns 10 species and figure 5.3 shows the names of the species that are hosted by each of the zones, and also the cost of protecting these zones.

The group of species S_1, *i.e.*, those species which, in order to be protected, require to be present in at least one zone of the central part of the reserve, consists of the 3 species s_1, s_2, and s_3; the group S_2 consists of the other 7 species, s_4, s_5, s_6, s_7, s_8, s_9, and s_{10}. Recall that the problem considered is to determine a subset of zones, of minimal cost, that allows all the species to be protected.

The optimal reserve is shown in figure 5.4. Its cost is 113 units; it is made up of two parts, not connected, comprising a total of 24 zones, 3 of which are located in a central part. For example, zone z_{99} is in the central part because the 8 surrounding zones, z_{88}, z_{89}, $z_{8,10}$, z_{98}, $z_{9,10}$, $z_{10,8}$, $z_{10,9}$, and $z_{10,10}$ are part of the reserve.

	1	2	3	4	5	6	7	8	9	10
1	s_8 **6**	**6**	s_3 **6**	s_2 **4**	s_3 **4**	**4**	**4**	**8**	**8**	**8**
2	s_2 **6**	**6**	**6**	s_5 **4**	s_1 **4**	**4**	**4**	s_2 s_5 **8**	s_8 s_{10} **8**	s_3 **8**
3	**6**	s_8 **6**	**6**	**4**	**4**	s_1 **4**	s_{10} **4**	**8**	**8**	**8**
4	s_5 **5**	s_1 **5**	s_9 **5**	s_3 s_7 **3**	s_9 **3**	**3**	s_9 **3**	s_8 **7**	**7**	**7**
5	**5**	**5**	**5**	**3**	**3**	**3**	s_6 s_{10} **3**	**7**	**7**	s_3 s_4 **7**
6	s_3 **5**	**5**	s_5 **5**	**3**	**3**	**3**	s_6 **3**	s_2 **7**	**7**	s_4 **7**
7	**5**	**5**	**5**	**3**	s_8 **3**	**3**	**3**	s_8 **7**	s_2 **7**	**7**
8	s_2 **4**	**4**	**4**	**6**	s_1 **6**	**6**	s_9 s_{10} **6**	**5**	s_2 **5**	s_6 s_9 **5**
9	**4**	**4**	**4**	**6**	s_3 s_7 **6**	**6**	s_{10} **6**	s_9 **5**	s_3 **5**	**5**
10	**4**	**4**	**4**	s_8 **6**	**6**	**6**	**6**	s_2 s_3 **5**	**5**	**5**

FIG. 5.3 – A set of 100 candidate zones for protection and, for each zone, the species that are present and the cost associated with the protection of the zone. For example, the zone at the intersection of row 8 and column 7 contains species s_9 and s_{10}, and the cost of its protection is 6.

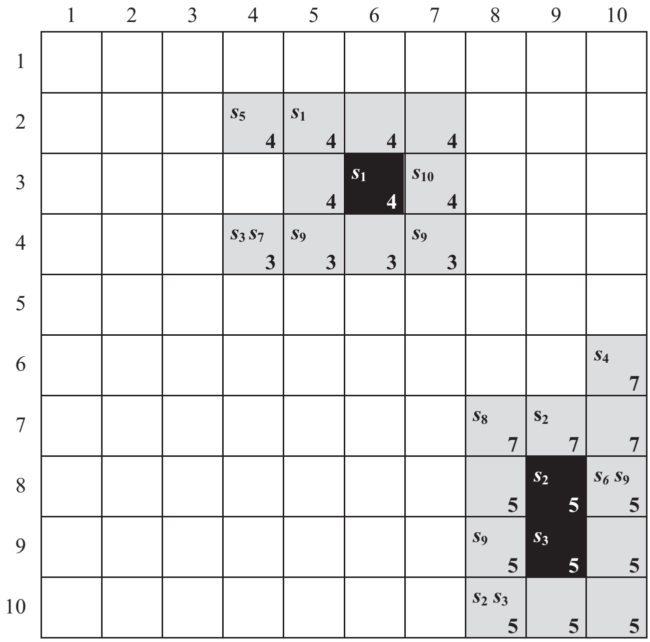


FIG. 5.4 – Optimal reserve that allows all the species to be protected for the instance described in figure 5.3. The 3 zones located in the central part of the reserve are represented in black and allow species s_1, s_2, and s_3 to be protected.

5.3 Edge Effect in Forest Exploitation

Sustainable forest management has economic, environmental and human well-being aspects. Since 1992, this concept has been clarified by international conferences and France has included it in the 2001 Forest Policy Act: "Sustainable forest management guarantees their biological diversity, productivity, regeneration capacity, vitality and ability to satisfy, now and in the future, relevant economic, ecological and social functions, at the local, national and international levels, without causing damage to other ecosystems". There are many publications on the subject involving the notion of optimization. Below, we study, as an example, a sustainable forest management problem that takes into account the impact of edges in the protection of certain species and is presented by Hof and Bevers (1998). This problem is about how to exploit the forest, more precisely how to harvest it, in order to protect two species as efficiently as possible, knowing that the harvested zones constitute a habitat favourable for the former and that boundaries between harvested and non-harvested zones – the edges – constitute a favourable habitat for the latter. Other criteria could easily be added to determine an optimal forest harvesting strategy such as, for example, income from harvested timber.

5.3.1 Optimal Protection of Two Species

We consider a set, Z, of forest zones – or parcels – that are square and identical, represented by a grid of nr rows and nc columns and two species, s_1 and s_2. Denote by z_{ij} the zone at the intersection of row i and column j, l the side length of the zones and $\underline{Z}$, the set of index pairs associated with the zones, *i.e.*, $\underline{Z} = \{1,..., \text{nr}\} \times \{1,..., \text{nc}\}$. The habitat of species s_1 is mainly in cut zones and the habitat of species s_2 is mainly in the edges between cut and uncut zones. For example, the goshawk population likes this edge habitat, in the vicinity of which there are open zones where it can hunt small mammals living in the same habitat. To simplify the presentation, it is considered that all the zones represented by the grid are initially uncut and that the zone outside the grid is a cut zone. The total expected population size of species s_1 in each cut (resp. uncut) zone z_{ij} is equal to n_{ij} (resp. 0). The total expected population size of species s_2 is equal to gL where g refers to the expected population size of species s_2 for each kilometre of edge and L to the total edge length taking into account the cuts made. The problem is to determine the zones to be cut and the zones to be left as they are in order to maximize the weighted sum of the total population sizes of species s_1 and s_2. The weighting reflects the different importance given to the two species. The weight w_1 is assigned to the population of species s_1 and weight w_2 to the population of species s_2.

5.3.2 Mathematical Programming Formulation

First, let us give the formulation proposed by Hof and Bevers (1998). These authors associate to each zone z_{ij} the Boolean variable x_{ij} which is equal to 1 if and only if the zone is not cut. They also associate to each zone z_{ij} the additional positive or zero variable d_{ij}, which represents the number of sides of this zone – from 0 to 4 – that do not form part of the edge when this zone is uncut; in the case where zone z_{ij} is cut, variable d_{ij} is equal to 0. Finally, these authors formulate the problem as the mixed-integer linear program $\text{P}_{5.2}$.

$$\text{P}_{5.2}: \begin{cases} \max \;\; w_1 \sum\limits_{(i,j)\,\in\,\underline{Z}} n_{ij}(1 - x_{ij}) + w_2 gl \sum\limits_{(i,j)\,\in\,\underline{Z}} (4x_{ij} - d_{ij}) \\ \text{s.t.} \left| \begin{array}{lll} d_{ij} \geq \sum\limits_{(k,l)\in \text{Adj}_{ij}} x_{kl} - \left|\text{Adj}_{ij}\right|(1 - x_{ij}) & (i,j) \in \underline{Z} & (5.2.1) \\ d_{ij} \geq 0 & (i,j) \in \underline{Z} & (5.2.2) \\ x_{ij} \in \{0,1\} & (i,j) \in \underline{Z} & (5.2.3) \end{array} \right. \end{cases}$$

In program $\text{P}_{5.2}$, Adj_{ij} refers to the set of couples (k, l) such that zone z_{kl} is adjacent to zone z_{ij}. Remember that w_1 and w_2 are the weighting coefficients and l is the side length of each parcel. The first part of the economic function expresses the total weighted population size of species s_1. Indeed, the total population size of this species in zone z_{ij} is equal to n_{ij} if the parcel z_{ij} is cut – $x_{ij} = 0$ – and to zero if the parcel is not cut – $x_{ij} = 1$. The second part of the economic function expresses the weighted total population size of species s_2 since the total length of the edge can be calculated by summing, on all uncut zones, the zone's contribution to this edge. We can verify that with the definition of variable d_{ij}, the contribution of the uncut zone z_{ij} to the length of the edge is equal to $4x_{ij}-d_{ij}$. The total length of the edge is therefore equal to $l\sum_{(i,j)\,\in\,\underline{Z}}(4x_{ij}-d_{ij})$ and the total population size of species s_2 is therefore equal to this last quantity multiplied by g. Let us now examine the behaviour of the positive or zero variable d_{ij} in relation to the Boolean variable x_{ij}. Because of the economic function to be maximized, variable d_{ij} takes, at the optimum of $\mathrm{P}_{5.2}$, the smallest possible value, *i.e.*, because of constraints 5.2.1 and 5.2.2, the value $\max\left\{\sum_{(k,l)\in\mathrm{Adj}_{ij}} x_{kl} - \left|\mathrm{Adj}_{ij}\right|(1-x_{ij}),\ 0\right\}$. If zone z_{ij} is not cut – $x_{ij}=1$ – variable d_{ij} is equal to $\sum_{(k,l)\in\mathrm{Adj}_{ij}} x_{kl}$, *i.e.*, the number of uncut zones adjacent to z_{ij}. If zone z_{ij} is cut – $x_{ij}=0$ – $d_{ij}=\max\left\{\sum_{(k,l)\in\mathrm{Adj}_{ij}} x_{kl} - \left|\mathrm{Adj}_{ij}\right|, 0\right\}$, which implies $d_{ij}=0$. This means that, if zone z_{ij} is cut, its contribution to the edge length is equal to 0. Finally, the quantity $4x_{ij}-d_{ij}$ is well equal to the number of sides of zone z_{ij} that are part of the edge when this zone is uncut and to 0, when this zone is cut.

We propose below an alternative formulation of the problem, based on the following observation: an edge separating two zones, z_{ij} and z_{kl}, is to be taken into account in the calculation of the edge length if and only if zone z_{ij} is cut while zone z_{kl} is not or if it is the opposite. In order not to count the same edge several times, only the following two adjacent zones are considered for any zone z_{ij}: the one located "to the right" of z_{ij} and the one located "under" z_{ij}. As in the previous formulation, with each zone z_{ij} is associated the Boolean variable x_{ij} which is equal to 1 if and only if the zone is uncut. The problem can then be formulated as the non-linear program in Boolean variables $\mathrm{P}_{5.3}$.

$$\mathrm{P}_{5.3}:\begin{cases} \max\ w_1 \sum\limits_{(i,j)\,\in\,\underline{Z}} n_{ij}(1-x_{ij}) \\ \quad + w_2 gl\left(\sum\limits_{(i,j,k,l)\,\in\,P}(x_{ij}+x_{kl}-2x_{ij}x_{kl}) + \sum\limits_{(i,j)\,\in\,Q} x_{ij} + \sum\limits_{(i,j)\,\in\,U} x_{ij}\right) \\ \text{s.t.}|\ x_{ij}\in\{0,1\} \quad (i,j)\,\in\,\underline{Z} \quad (5.3.1) \end{cases}$$

where $P=\{(i,j,k,l)\in\underline{Z}^2 : (k,l)=(i+1,j) \text{ or } (k,l)=(i,j+1)\}$, $Q=\{(i,j)\in(M\times N) : i=1 \text{ or } i=n \text{ or } j=1 \text{ or } j=m\}$, and $U=\{(1,1),(1,m),(n,1),(n,m)\}$.

The first part of the economic function is identical to that of $\mathrm{P}_{5.2}$. Let us look at the second part. Consider two zones, z_{ij} and z_{kl}, z_{kl} being adjacent to z_{ij} and located to the right or below it. Let us check that the quantity is indeed equal to the number of sides to be taken into account – 0 or 1 – in the edge possibly generated by the adjacency of zones z_{ij} and z_{kl}. This is indeed the case since if these 2 zones are not cut, this quantity is equal to $1+1-2=0$, if these two zones are cut, it is equal to $0+0-0=0$ and, finally, if one of the zones is cut and the other not, it is equal to $1+0-0=1$ or $0+1-0=1$. The quantity $\sum_{(i,j,k,l)\in P}(x_{ij}+x_{kl}-2x_{ij}x_{kl})$ is therefore well equal to the number of sides belonging to the edge and coming from the adjacency of all the pairs of zones. To count the total number of sides belonging to the edge, it is still necessary to take into account the uncut zones that are adjacent to the outside of the grid. Since the zone outside the grid is considered a cut zone, it is easy to verify that the number of sides belonging to these zones and forming part of the edge is equal to $\sum_{(i,j)\in Q} x_{ij} + \sum_{(i,j)\in U} x_{ij}$. It must be taken into account that if the zones z_{ij}, $(i, j) \in U$, are not cut, two of their sides belong to the edge between these zones and the outside of the grid.

The advantage of this formulation is that the matrix of constraints associated with its classic linearization is totally unimodular (TU), which is not the case with the matrix of constraints associated with program $P_{5.2}$ (see appendix at the end of the book). The classic linearization of $P_{5.3}$ consists in replacing the products $x_{ij}x_{kl}$ by variables y_{ijkl} and adding the linear constraints $1 - x_{ij} - x_{kl} + y_{ijkl} \geq 0$ and $y_{ijkl} \geq 0$ to force the equality $y_{ijkl} = x_{ij}x_{kl}$ at the optimum (see appendix at the end of the book). We show that the constraint matrix of this linearization is TU, based on the fact that the vertex-edge incidence matrix of a bipartite graph is TU (see, for example, Nemhauser & Wolsey, 1988). This formulation, therefore, allows large-sized instances of the problem to be solved without difficulty. Let us examine this second formulation of the problem; it is written as the mixed-integer linear program $P_{5.4}$.

$$P_{5.4} : \begin{cases} \max w_1 \sum\limits_{(i,j)\,\in\,\underline{Z}} n_{ij}(1 - x_{ij}) \\ \qquad + w_2 gl \left(\sum\limits_{(i,j,k,l)\in P} (x_{ij} + x_{kl} - 2y_{ijkl}) + \sum\limits_{(i,j)\in Q} x_{ij} + \sum\limits_{(i,j)\in U} x_{ij} \right) \\ \text{s.t.} \left| \begin{array}{lll} 1 - x_{ij} - x_{kl} + y_{ijkl} \geq 0 & (i, j, k, l) \in P & (5.4.1) \\ y_{ijkl} \geq 0 & (i, j, k, l) \in P & (5.4.2) \\ x_{ij} \in \{0, 1\} & (i, j) \in \underline{Z} & (5.4.3) \end{array} \right. \end{cases}$$

Let us study the matrix A associated with the set of constraints $C_{5.1}$ below and derived from constraints 5.4.1 and 5.4.3.

$$C_{5.1} : \begin{cases} x_{ij} + x_{kl} - y_{ijkl} \leq 1 & (i, j, k, l) \in P \\ x_{ij} \leq 1 & (i, j) \in \underline{Z} \end{cases}$$

The matrix A is composed of the three sub-matrices, B, C, and D: $A = \left(\frac{B \mid C}{D}\right)$ (figure 5.5). Let $G = (\underline{Z}, E)$ be the graph defined as follows: with each zone z_{ij} of Z is associated a vertex, and two vertices, (i, j) and (k, l), are connected by an edge if and only if the two zones z_{ij} and z_{kl} have a common side. This graph is a grid and, therefore, a bipartite graph (figure 5.6). Matrix B is the transposed matrix of the vertex-edge incidence matrix of the graph; it is therefore TU. Each column in matrix C has a single non-zero element that is equal to -1. Using calculation of matrix determinants (expansion by cofactors), it can be shown that the determinant of any square sub-matrix of (B, C) belongs to $\{-1, 0, 1\}$. (B, C) is therefore TU. Similarly, each row of D has a single non-zero element that is equal to 1; the matrix $\left(\frac{B \mid C}{D}\right)$ is therefore TU. Recall that the minor, M_{ij}, of a square matrix M is the determinant of the matrix obtained by eliminating the ith row and the jth column of M. The cofactor, C_{ij}, of the matrix M is defined by $C_{ij} = (-1)^{i+j}M_{ij}$.

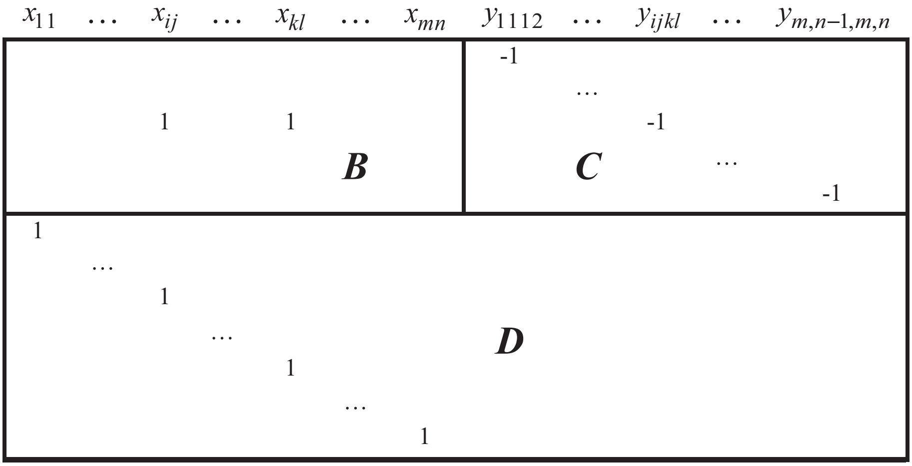


FIG. 5.5 – The non-zero terms of the matrix, A, *i.e.*, the matrix corresponding to the set of inequalities $C_{5.1}$: $x_{ij} + x_{kl} - y_{ijkl} \leq 1$, $(i, j, k, l) \in P$, and $x_{ij} \leq 1$, $(i, j) \in \underline{Z}$.

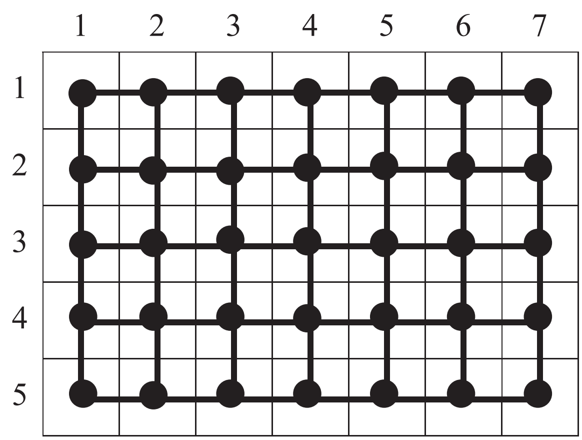


FIG. 5.6 – Graph associated with a grid with 5 rows and 7 columns. The edges are represented by bold lines and the vertices by black circles.

In conclusion, since (1) matrix A associated with the constraints of $P_{5.4}$ is TU and (2) the vector of the second members of the constraint set $C_{5.1}$ is an integer vector, the problem considered can be formulated as the linear program in real variables $P_{5.5}$ which corresponds to program $P_{5.4}$ in which the integrality constraint has been relaxed.

$$
\mathrm{P}_{5.5}: \begin{cases} \max w_1 \sum\limits_{(i,j)\,\in\,\underline{Z}} n_{ij}(1-x_{ij}) \\ \qquad + w_2 gl \left(\sum\limits_{(i,j,k,l)\,\in\,P} (x_{ij}+x_{kl}-2y_{ijkl}) + \sum\limits_{(i,j)\,\in\,Q} x_{ij} + \sum\limits_{(i,j)\,\in\,U} x_{ij} \right) \\ \text{s.t.} \left| \begin{array}{lll} 1-x_{ij}-x_{kl}+y_{ijkl} \geq 0 & (i,j,k,l) \in P & (5.5.1) \\ y_{ijkl} \geq 0 & (i,j,k,l) \in P & (5.5.2) \\ 0 \leq x_{ij} \leq 1 & (i,j) \in \underline{Z} & (5.5.3) \end{array} \right. \end{cases}
$$

5.3.3 Examples

Example A. Consider the instance represented by a grid of 5×5 square and identical zones (figure 5.7a). The values n_{ij} are indicated in each zone of the grid. The side length of each zone is equal to 3 units, the weights associated with species s_1 and s_2 are equal to 2 and 1, respectively, and the coefficient g is equal to 1.26157. The solution is given in figure 5.7b in which the uncut zones are shown in grey. In this solution, 5 zones are uncut, the number of species s_1 is equal to 191, the number of species s_2 is equal to 60.56, the number of sides belonging to the edge is equal to 16 and the value of the economic function is equal to 442.56.

Example B. Consider a second instance represented by a grid of 10×10 identical square zones and presented in figure 5.8. The values n_{ij} are indicated in each zone of the grid. The side length of each zone is equal to 3 units, the weights associated with species s_1 and s_2 are respectively equal to 1 and 5, and the coefficient g is equal to 1.26157.

The optimal solution for this instance is given in figure 5.9a in which the uncut zones are shown in grey. In this solution, 21 zones are uncut, the number of species s_1 is

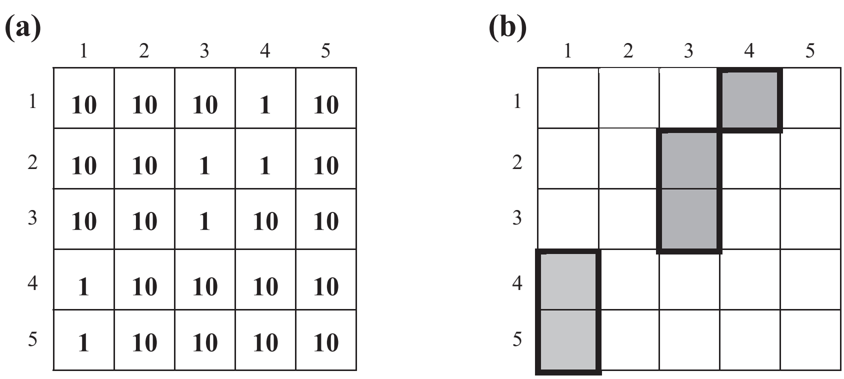


Fig. 5.7 – A hypothetical forest massif represented by a grid of 5×5 square and identical zones. (a) The values n_{ij} are given in each zone. (b) Cut – white – and uncut – grey – zones in an optimal solution.

	1	2	3	4	5	6	7	8	9	10
1	**84**	**68**	**97**	**98**	**64**	**89**	**82**	**71**	**74**	**76**
2	**87**	**83**	**98**	**75**	**60**	**90**	**78**	**67**	**92**	**94**
3	**84**	**68**	**70**	**81**	**67**	**61**	**73**	**92**	**86**	**90**
4	**79**	**62**	**86**	**79**	**73**	**84**	**76**	**98**	**84**	**90**
5	**62**	**72**	**66**	**72**	**92**	**80**	**71**	**91**	**87**	**70**
6	**85**	**77**	**63**	**93**	**90**	**94**	**76**	**81**	**99**	**98**
7	**76**	**63**	**66**	**84**	**94**	**93**	**72**	**92**	**79**	**65**
8	**76**	**63**	**92**	**69**	**60**	**88**	**79**	**93**	**66**	**73**
9	**92**	**82**	**77**	**72**	**77**	**81**	**89**	**95**	**80**	**80**
10	**88**	**89**	**83**	**86**	**69**	**78**	**91**	**64**	**94**	**92**

FIG. 5.8 – A hypothetical forest massif represented by a grid of 10×10 square and identical zones. The values n_{ij} are given in each zone.

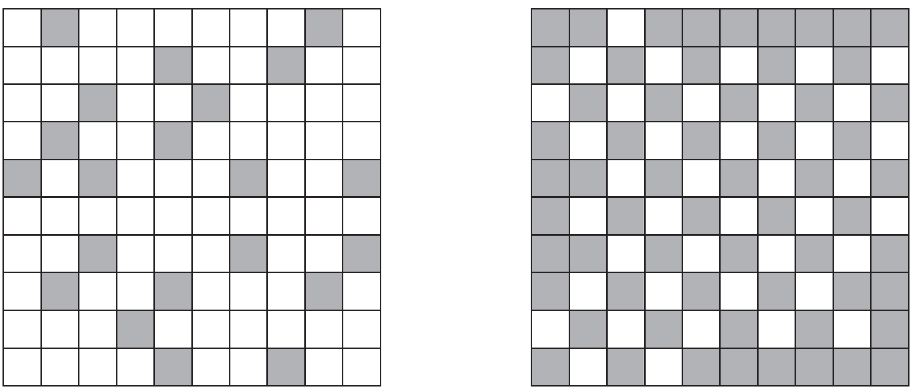


FIG. 5.9 – (a) Cut zones – in white – and uncut zones – in grey – in an optimal solution of the instance described in figure 5.8; (b) Cut zones – in white – and uncut zones – in grey – in an optimal solution of the instance described in figure 5.8 in the case where the number of uncut zones must be greater than or equal to 60.

6,630, the number of species s_2 is 317.92, the number of sides belonging to the edge is 84 and the value of the economic function is 8,219.58.

Example C. Now consider the same instance as in Example B above but with the following additional constraint: the number of uncut zones must be greater than or equal to 60. Adding this constraint causes the loss of the TU property of the constraint matrix associated with this variant of the initial problem. To solve it, it is therefore necessary to solve the mathematical program $P_{5.4}$ to which is added the constraint $\sum_{(i,j) \in \underline{Z}} x_{ij} \geq 60$. The optimal solution of this instance is given by figure 5.9b in which the uncut zones are represented in grey. In this solution, 60 zones are uncut, the number of species s_1 is 3,272, the number of species s_2 is 696.39, the number of sides belonging to the edge is 184 and the value of the economic function is 6,753.93.

5.4 Connectivity Properties in Forest Exploitation

5.4.1 Optimal Protection of Two Species

The problem studied in this section is similar to the one studied in section 5.3, in that it consists in defining the exploitation of a forest region in order to protect certain species present in that region. We consider a set, Z, of square and identical forest zones represented by a grid of nr rows and nc columns and two species, s_1 and s_2, living in these zones. Denote by z_{ij} the zone at the intersection of row i and column j, and $\underline{Z}$ the set of index pairs associated with the zones, *i.e.*, $\underline{Z} = \{1,..., \mathrm{nr}\} \times \{1,..., \mathrm{nc}\}$. The problem is to determine the zones to be cut and the zones to be left as they are in order to maximize the weighted sum of the population sizes of species s_1 and s_2. The weight w_1 is assigned to the population size of species s_1 and weight w_2 to the population size of species s_2. The total expected population size of species s_1 in each cut (resp. uncut) zone z_{ij} is equal to n_{ij} (resp. 0). The calculation of the population size of species s_2 differs from that of section 5.3. The habitat of this species is composed of uncut zones but in each zone, its population size depends on the connection of this zone with the other uncut zones, more precisely on the probability that this zone is connected to at least one other uncut zone. Several studies aiming to optimize landscape configuration take into account this type of dependence between zones. Hof and Bevers (1998) propose a simple linear approximation of the population size of species s_2 for a particular case of the connection probabilities. We propose here a general method, which can be used with any set of connection probabilities, to estimate with great accuracy the population size of this species. As in section 5.3, with each zone z_{ij} is associated a Boolean variable, x_{ij}, which is equal to 1 if and only if the zone is uncut. The population size of species s_1 in the set of considered zones is then equal to $\sum_{(i,j)\in\underline{Z}} n_{ij}(1 - x_{ij})$. The population size of species s_2 is more difficult to estimate. As mentioned above, its habitat consists of uncut zones. The population size of this species is equal to $\sum_{(i,j)\in\underline{Z}} \pi_{ij}\mathrm{PR}_{ij}x_{ij}$ where PR_{ij} $(0 \le \mathrm{PR}_{ij} \le 1)$ refers to the connectivity of zone z_{ij} with other uncut zones, and π_{ij} is the population size of species s_2 in zone z_{ij} when $\mathrm{PR}_{ij} = 1$ and $x_{ij} = 1$. Two zones, z_{ij} and z_{kl}, are considered to be connected with a certain probability, denoted by pr_{ijkl} $(0 \le \mathrm{pr}_{ijkl} < 1)$. It is further assumed that all these probabilities are independent. The connectivity of zone z_{ij} is measured by the probability that this zone is connected to the other uncut zones and we assume that this probability is equal to the probability that this zone is connected to at least one other uncut zone. The connectivity of zone z_{ij}, PR_{ij}, is therefore equal to $1 - \prod_{(k,l)\in\underline{Z}} \left(1 - \mathrm{pr}_{ijkl}x_{kl}\right)$ with $\mathrm{pr}_{ijij} = 0$. It is further assumed that species s_2 can only survive in the set of considered zones if at least TH of these zones are not cut.

5.4.2 Illustration of the Problem

Let us consider a forest region represented by a grid of 5×5 square and identical zones (figure 5.10). In each zone, the values of π_{ij} and n_{ij} are specified, π_{ij} being placed above n_{ij}. A feasible solution is shown in this figure in which the grey zones are the uncut zones ($x_{ij} = 1$).

Given zone z_{ij}, suppose that $\text{pr}_{ijkl} = 0.5$ if zone z_{kl} belongs to the set of zones that immediately surround zone z_{ij}, that $\text{pr}_{ijkl} = 0.15$ if zone z_{kl} belongs to the set of zones that surround the previous set of zones, and that $\text{pr}_{ijkl} = 0$ for the other zones z_{kl}. Figure 5.11 shows the values of pr_{44kl}. The values of PR_{ij} for the uncut zones in figure 5.10 are presented in figure 5.12.

For the solution presented in figure 5.10, the population size of species s_1, $\sum_{(i,j)\in \underline{Z}} n_{ij}(1 - x_{ij})$, is equal to 815 and that of species s_2, $\sum_{(i,j)\in \underline{Z}} \pi_{ij}\text{PR}_{ij}x_{ij}$, is equal to 66.66.

	1	2	3	4	5
1	5.0 60	5.0 64	4.5 66	3.5 70	2.0 75
2	3.5 70	4.5 74	5.0 78	4.5 82	3.5 87
3	2.0 82	3.5 86	4.5 90	5.0 94	4.5 100
4	2.0 87	3.5 90	4.5 93	5.0 96	4.5 100
5	2.0 90	3.5 92	4.5 95	5.0 98	5.0 100

FIG. 5.10 – A hypothetical forest region consisting of 25 zones where the values of π_{ij} and n_{ij}, π_{ij} above n_{ij}, are specified. A solution with 9 cut zones and 16 uncut zones.

	1	2	3	4	5
1	0	0	0	0	0
2	0	0.15	0.15	0.15	0.15
3	0	0.15	0.5	0.5	0.5
4	0	0.15	0.5	0	0.5
5	0	0.15	0.5	0.5	0.5

FIG. 5.11 – Values of pr_{44kl} for all $(k, l) \in \{1,..., 5\}^2$.

	1	2	3	4	5
1	0.92	0.98	0.99	0.97	-
2	0.94	0.99	1	0.99	0.94
3	-	-	1	0.99	-
4	0.62	-	0.99	0.98	-
5	-	-	0.92	0.91	-

FIG. 5.12 – Values of PR_{ij} associated with the solution of figure 5.10.

5.4.3 Mathematical Programming Formulation

Using variable σ_1 (resp. σ_2) to represent the population size of species s_1 (resp. s_2) and the Boolean variable b to express the constraint on the number of zones that must be uncut so that species s_2 can survive, Hof and Bevers (1998) propose to formulate the problem as the mixed-integer non-linear program $\mathrm{P}_{5.6}$.

$$
\mathrm{P}_{5.6}: \begin{cases} \max \quad w_1\sigma_1 + w_2\sigma_2 \\ \text{s.t.} \left| \begin{array}{lll} \sigma_1 = \sum_{(i,j)\in \underline{Z}} n_{ij}(1-x_{ij}) & & (5.6.1) \\ \sigma_2 \leq \sum_{(i,j)\in \underline{Z}} \pi_{ij}\mathrm{PR}_{ij}x_{ij} & & (5.6.2) \\ \mathrm{PR}_{ij} = 1 - \prod_{(k,l)\in \underline{Z}} (1-\mathrm{pr}_{ijkl}x_{kl}) & (i,j)\in \underline{Z} & (5.6.3) \\ \sigma_2 \leq \gamma\, b & & (5.6.4) \\ b \leq \frac{1}{\mathrm{TH}} \sum_{(i,j)\in \underline{Z}} x_{ij} & & (5.6.5) \\ 0 \leq \mathrm{PR}_{ij} \leq 1,\, x_{ij} \in \{0,1\} & (i,j)\in \underline{Z} & (5.6.6) \\ \sigma_1 \geq 0, \sigma_2 \geq 0, b \in \{0,1\} & & (5.6.7) \end{array} \right. \end{cases}
$$

γ is a constant that must be greater than or equal to the maximal value that variable σ_2 can take. We can set, for example, $\gamma = \sum_{(i,j)\in \underline{Z}} \pi_{ij}$. Thus, if $b = 0$, constraint 5.6.4 forces variable σ_2 to take the value 0 and if $b = 1$ this constraint is inactive. Because of constraint 5.6.5, the Boolean variable b takes the value 0 if $\sum_{(i,j)\in \underline{Z}} x_{ij} < \mathrm{TH}$ and the value 1 – at the optimum – in the opposite case. The economic function and all the constraints are linear, except constraints 5.6.2 and 5.6.3. Hof and Bevers (1998) solve $\mathrm{P}_{5.6}$ in an approximate way using a simple linear approximation of the population size of species s_2 for a particular case of the connection probabilities. We propose below a general method for solving $\mathrm{P}_{5.6}$, also in an approximate way, but valid whatever the definition of the connection probabilities. In addition, the population size of species s_2 is evaluated with a great accuracy.

Using the same technique as in section 7.5 of chapter 7, an approximate solution of program $P_{5.6}$ and an upper bound of its optimal value can be obtained by solving a mixed-integer linear program. To do this, first rewrite $P_{5.6}$ as $P_{5.7}$ by replacing, in $P_{5.6}$, the product of variables $PR_{ij}x_{ij}$ with variable e_{ij}. Because of the objective function to be maximized, constraints 5.7.3 and 5.7.4 imply $e_{ij} = PR_{ij}x_{ij}$ at the optimum.

$$
P_{5.7}: \begin{cases} \max \quad w_1\sigma_1 + w_2\sigma_2 \\ \text{s.t.} \left| \begin{array}{lll} \sigma_1 = \sum_{(i,j)\in \underline{Z}} n_{ij}(1-x_{ij}) & & (5.7.1) \\ \sigma_2 \le \sum_{(i,j)\in \underline{Z}} \pi_{ij} e_{ij} & & (5.7.2) \\ 1 - e_{ij} \ge \prod_{(k,l)\in \underline{Z}} (1 - \mathrm{pr}_{ijkl} x_{kl}) & (i,j)\in \underline{Z} & (5.7.3) \\ e_{ij} \le x_{ij} & (i,j)\in \underline{Z} & (5.7.4) \\ \sigma_2 \le \gamma\, b & & (5.7.5) \\ b \le \frac{1}{\mathrm{TH}} \sum_{(i,j)\in \underline{Z}} x_{ij} & & (5.7.6) \\ 0 \le e_{ij} \le 1, x_{ij} \in \{0,1\} & (i,j)\in \underline{Z} & (5.7.7) \\ \sigma_1 \ge 0, \sigma_2 \ge 0, b \in \{0,1\} & & (5.7.8) \end{array} \right. \end{cases}
$$

Using the properties of the logarithmic function and taking into account that variables x_{kl}, $(k,l) \in \underline{Z}$, are Boolean, $\log\left[\prod_{(k,l)\in \underline{Z}} (1 - \mathrm{pr}_{ijkl} x_{kl})\right] = \sum_{(k,l)\in \underline{Z}} \log(1 - \mathrm{pr}_{ijkl}) x_{kl}$, and $P_{5.7}$ is equivalent to $P_{5.8}$.

$$
P_{5.8}: \begin{cases} \max \quad w_1\sigma_1 + w_2\sigma_2 \\ \text{s.t.} \left| \begin{array}{lll} \sigma_1 = \sum_{(i,j)\in \underline{Z}} n_{ij}(1-x_{ij}) & & (5.8.1) \\ \sigma_2 \le \sum_{(i,j)\in \underline{Z}} \pi_{ij} e_{ij} & & (5.8.2) \\ \log(1 - e_{ij}) \ge \sum_{(k,l)\in \underline{Z}} \log(1 - \mathrm{pr}_{ijkl}) x_{kl} & (i,j)\in \underline{Z} & (5.8.3) \\ e_{ij} \le x_{ij} & (i,j)\in \underline{Z} & (5.8.4) \\ \sigma_2 \le \gamma\, b & & (5.8.5) \\ b \le \frac{1}{\mathrm{TH}} \sum_{(i,j)\in \underline{Z}} x_{ij} & & (5.8.6) \\ 0 \le e_{ij} \le 1, x_{ij} \in \{0,1\} & (i,j)\in \underline{Z} & (5.8.7) \\ \sigma_1 \ge 0, \sigma_2 \ge 0, b \in \{0,1\} & & (5.8.8) \end{array} \right. \end{cases}
$$

$P_{5.8}$ is not yet a linear program – in mixed-integer variables – because of the expression $\log(1 - e_{ij})$ that appears in constraints 5.8.3. Using the same technique as in section 7.5 of chapter 7, a relaxation of program $P_{5.8}$ is obtained by replacing constraints 5.8.3 by the constraints $\frac{1}{u_v}(1-e_{ij}) + \log u_v - 1 \geq \sum_{(k,l)\in\underline{Z}} \log(1-\mathrm{pr}_{ijkl})x_{kl}$, $(i,j) \in \underline{Z}$, $v = 1,\ldots, q$, where u is a vector of $\mathbb{R}^q$ such that $0 < u_1 < u_2 < \cdots < u_q = 1$. This results in program $\mathrm{P}_{5.9}$.

$$
\mathrm{P}_{5.9}: \begin{cases} \max \quad w_1\sigma_1 + w_2\sigma_2 \\ \text{s.t.} \left| \begin{array}{lll} \sigma_1 = \sum\limits_{(i,j)\in\underline{Z}} n_{ij}(1-x_{ij}) & & (5.9.1) \\ \sigma_2 \leq \sum\limits_{(i,j)\in\underline{Z}} \pi_{ij} e_{ij} & & (5.9.2) \\ \frac{1}{u_v}(1-e_{ij}) + \log u_v - 1 & & \\ \quad \geq \sum\limits_{(k,l)\in\underline{Z}} \log(1-\mathrm{pr}_{ijkl})x_{kl} & (i,j)\in\underline{Z}, v = 1,\ldots,q & (5.9.3) \\ e_{ij} \leq x_{ij} & (i,j)\in\underline{Z} & (5.9.4) \\ \sigma_2 \leq \gamma\, b & & (5.9.5) \\ b \leq \frac{1}{\mathrm{TH}} \sum\limits_{(i,j)\in\underline{Z}} x_{ij} & & (5.9.6) \\ 0 \leq e_{ij} \leq 1, x_{ij} \in \{0,1\} & (i,j)\in\underline{Z} & (5.9.7) \\ \sigma_1 \geq 0, \sigma_2 \geq 0, b \in \{0,1\} & & (5.9.8) \end{array} \right. \end{cases}
$$

An optimal solution of $\mathrm{P}_{5.9}$, $(\tilde{\sigma}, \tilde{e}, \tilde{x}, \tilde{b})$, gives a feasible solution to the problem considered, *i.e.*, of $\mathrm{P}_{5.6}$. The actual value of this solution is $w_1\tilde{\sigma}_1 + w_2 \sum_{(i,j)\in\underline{Z}} \pi_{ij}\, \tilde{x}_{ij} \Big(1 - \prod_{(k,l)\in\underline{Z}} (1-\mathrm{pr}_{ijkl}\, \tilde{x}_{kl})\Big)$, and $w_1\tilde{\sigma}_1 + w_2\tilde{\sigma}_2$ is an upper bound of the optimal value of $\mathrm{P}_{5.6}$. To obtain a good approximate solution of $\mathrm{P}_{5.6}$, q must be large enough but, the larger q is, the greater the number of constraints 5.9.3 is.

5.4.4 Example

Let us take again the example of section 5.4.2 and set $w_1 = 1$, $w_2 = 20$, $\mathrm{TH} = 8$, and $u_v = u_1^{(q-v)/(q-1)}$ with $q = 10$ and $u_1 = 0.01$. This results in $u = (0.01, 0.02,$ $0.03,\ 0.05,\ 0.08,\ 0.13,\ 0.22,\ 0.36,\ 0.60,\ 1.00)$. The value of the optimal solution of $\mathrm{P}_{5.9}$ is 2,238.17; it is defined by $x_{ij} = 1$ if and only if $(i,j) \in \{(1,1),(1,2),(1,3),(1,4),(2,1),(2,2),(2,3),(2,4),(3,3),(3,4)\}$. These values of x_{ij} provide an – approximate – solution to the problem, with a value of 2,237.8. The relative error made in retaining this feasible solution rather than an optimal solution is, therefore, at most equal to $1.7\ 10^{-4}$. Note that, in this approximate solution, the true value of the population size of species s_2 is 43.3401 while the value of variable σ_2 is 43.3587.

5.5 Optimal Reserve in the Case of Non-Disjoint Candidate Zones

We have always considered that the zones that are candidates for protection to form a reserve and thus protect certain species – or certain ecosystems – are all disjoint. In this section, we examine the case where this is not the case. Indeed, in some situations, the definition of the candidate zones is made in such a way that the selection of one zone can automatically lead to the selection of a part of another zone.

5.5.1 Optimal Protection of the Considered Species by a Reserve of Limited Area

Let $S = \{s_1, s_2, \ldots, s_m\}$ be the set of threatened species of interest and $Z = \{z_1, z_2, \ldots, z_n\}$ be the set of candidate zones for protection, that is, the set of zones that can be decided to be protected or not. As mentioned in the introduction to this book, a set of species is being considered to simplify the presentation. Other aspects of nature and biodiversity could be added, but this would not significantly change the proposed approaches to addressing the questions posed later in this section. The protection of a zone brings some protection for species living in that zone. For each zone z_i, we know all the species that live there and we denote by Z_k, the set of zones in which species s_k lives. We denote by $\underline{Z}_k$, the set of corresponding indices. The difference with the models studied so far and which will have to take into account is that the n zones $z_1, z_2, \ldots, z_n$ are not necessarily all disjoint (see figure 5.13). For example, the candidate zones may be very different in nature and the data available – and it is not possible to have more precise data based, for example, on a redefinition of the zones – may indicate that species s_k lives in zone z_i and that species s_l lives in zone z_j that is not disjoined from z_i. This property significantly modifies the models associated with the selection of optimal reserves, particularly with regard to the level of protection of species and the area of the reserves selected. The level of protection of a species can be defined in several ways, taking into account the zones selected for protection. It is considered here that the level of protection of a species, s_k, depends only on the number of zones of Z_k that are included in the reserve and that this level of protection is proportional to this number. There is, however, a small difficulty because some zones of Z_k can be included only partially in the reserve since the n zones $z_1, z_2, \ldots, z_n$ are not necessarily all disjoint. To take this phenomenon into account, to each species s_k a protection coefficient is assigned which is equal to the sum of the fractions of areas of each zone of Z_k which is protected. This coefficient is, therefore, equal to $\sum_{i \in \underline{Z}_k} \alpha_i / a_i$ where α_i is equal to the area of zone z_i that is actually protected and a_i, to the total area of zone z_i. Remember that 2 situations can occur with regard to the protection of zone z_i: (1) it is decided to protect z_i and then the whole z_i area is protected, (2) it is decided not to protect z_i, but part of the z_i area may nevertheless be protected because of the decision to protect some zones of Z that have a non-empty intersection with z_i. Note that the whole zone z_i can be protected in this way. This is the case, for example, when zone z_i

is completely included in a zone that it is decided to protect. In summary, the value of protecting a set of zones, R, is assessed here by a weighted species richness criterion that is equal to the quantity $\sum_{k \in \underline{S}} \sum_{i \in \underline{Z}_k} \alpha_i / a_i$ and that we call the "weighted number of protected species".

5.5.2 Illustration of the Problem

The calculation of the interest in protecting a set of zones, R, is illustrated in figure 5.13 and table 5.1. As mentioned, the decisions to be made are whether to select an entire zone for protection or not, but this may result in some fractions of unselected zones still being protected. Let us examine the solution which consists in selecting the grey zones, z_1, z_3, z_4, z_7, z_8, z_{10}, and z_{12}, and, therefore, not selecting zones z_2, z_5, z_6, z_9, z_{11}, z_{13}, z_{14}, and z_{15}. The choice of the selected zones implies that a fraction of zones z_2, z_6, z_9, z_{11}, and z_{13} are still protected. For this solution and taking into account the species likely to be protected by each zone, the weighted number of protected species is equal to 30.10 (see table 5.1 for details of the calculation). The total protected area is equal to 94 units.

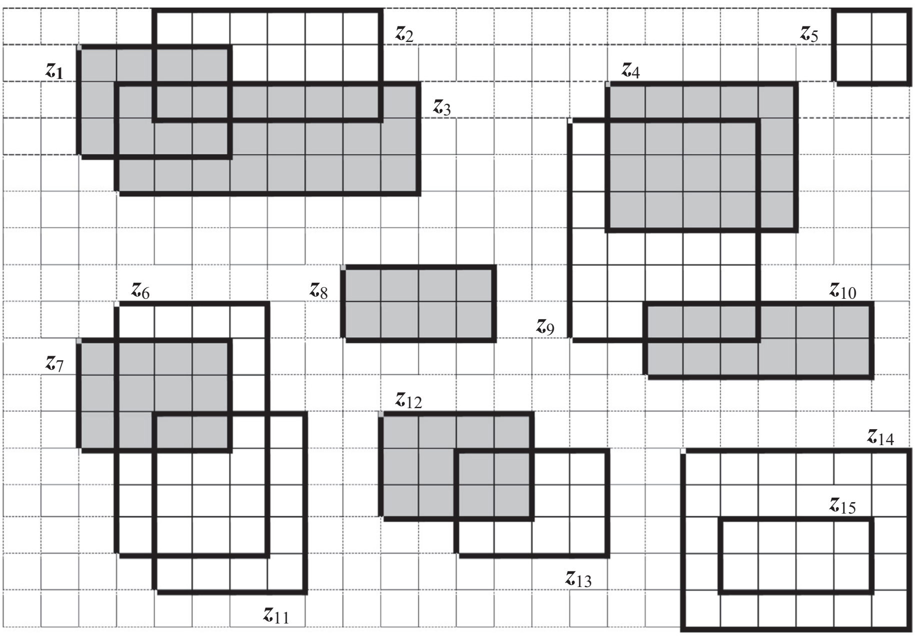


FIG. 5.13 – The region under consideration is represented by a grid of 17 rows and 24 columns. Each cell in this grid is a square whose side length is equal to one unit. The candidate zones are squares or rectangles made up of a subset of these cells, all in one piece. It is assumed that each of the zones z_1, z_5, z_7, z_{10}, z_{12}, and z_{15} is able to protect the 4 species s_1, s_2, s_3, and s_4, that each of the zones z_2, z_4, z_8, z_{11}, and z_{13} is able to protect the 3 species s_5, s_6, and s_7, and that each of the zones z_3, z_6, z_9, and z_{14} is able to protect the 3 species s_8, s_9, and s_{10}.

Tab. 5.1 – Detail of the calculation of the weighted number of protected species for the example described in figure 5.13.

Species	Weighting	Species	Weighting
s_1	4	s_6	$8/18 + 1 + 1 + 4/12 + 2/20 = 2.88$
s_2	4	s_7	$8/18 + 1 + 1 + 4/12 + 2/20 = 2.88$
s_3	4	s_8	$1 + 9/28 + 15/30 = 1.82$
s_4	4	s_8	$1 + 9/28 + 15/30 = 1.82$
s_5	$8/18 + 1 + 1$ $+ 4/12 + 2/20 = 2.88$	s_{10}	$1 + 9/28 + 15/30 = 1.82$

Several reserve selection problems involving the weighted number of protected species may arise. The problem we are considering here is to determine the zones to be protected in order to maximize the weighted number of protected species while respecting an upper limit, $A_{\max}$, on the total protected area.

5.5.3 A First Mathematical Programming Formulation

We use the Boolean variable x_i which is equal to 1 if and only if we decide to protect zone z_i. To simplify the presentation, we limit ourselves to the case where the non-empty intersections of zones concern at most 3 zones but we could easily generalize the approach in case more than 3 zones can have a non-empty intersection. We pose:

$$a_{ij} = \begin{cases} \text{area of the intersection of the zones } z_i \text{ and } z_j, \text{ if } i<j \\ 0 \ \text{ otherwise} \end{cases}$$

$$a_{ijk} = \begin{cases} \text{area of the intersection of the zones } z_i, z_j, \text{ and } z_k, \text{ if } i<j<k \\ 0 \ \text{ otherwise} \end{cases}$$

Note that a_{ij} can also be equal to 0 for some values of i and j such as $i < j$ and that a_{ijk} can also be equal to 0 for some values of i, j, and k such as $i < j < k$. The real variable, positive or zero, α_i, which represents the protected area of zone z_i, is used when deciding not to select this zone for protection. Indeed, as we have seen above, if it is decided not to protect zone z_i, a part of this zone may possibly be protected because of the protection of other zones. For example, in the case of figure 5.13, the protection of zones z_1 and z_3 induces a partial protection of zone z_2 and $\alpha_2/a_2 = 8/18$. In general, the area of zone z_i that is protected as a result of the protection of other zones is equal to $\sum_{j\,:\,(i,j)\in\underline{Z}^2} (a_{ij} + a_{ji})\, x_j - \sum_{j,k\,:\,(i,j,k)\in\underline{Z}^3} (a_{ijk} + a_{jik} + a_{jki})\, x_j x_k$. The problem can be formulated as the mixed-integer non-linear program $\mathrm{P}_{5.10}$.

$$
\mathrm{P}_{5.10}: \begin{cases} \max \sum\limits_{k \in \underline{S}} \sum\limits_{i \in \underline{Z}_k} \left(x_i + \frac{\alpha_i}{a_i} \right) & \\ \text{s.t.} \left| \begin{array}{lll} \sum\limits_{i \in \underline{Z}} a_i x_i - \sum\limits_{(i,j) \in \underline{Z}^2} a_{ij}\, x_i x_j + \sum\limits_{(i,j,k) \in \underline{Z}^3} a_{ijk}\, x_i x_j x_k \le A_{\max} & & (5.10.1) \\ \alpha_i \le \sum\limits_{j:(i,j) \in \underline{Z}^2} (a_{ij} + a_{ji})\, x_j - \sum\limits_{j,k:(i,j,k) \in \underline{Z}^3} (a_{ijk} + a_{jik} + a_{jki})\, x_j x_k & i \in \underline{Z} & (5.10.2) \\ \alpha_i \le a_i(1 - x_i) & i \in \underline{Z} & (5.10.3) \\ x_i \in \{0, 1\}, \alpha_i \ge 0 & i \in \underline{Z} & (5.10.4) \end{array} \right. \end{cases}
$$

Because of the economic function to be maximized, variable α_i takes, at the optimum of $\mathrm{P}_{5.10}$, the largest possible value. Due to constraints 5.10.2 and 5.10.3, variable α_i therefore takes the value 0 if zone z_i is protected, *i.e.*, if $x_i = 1$, and the value $\sum_{(i,j) \in \underline{Z}^2} (a_{ij} + a_{ji})\, x_j - \sum_{(i,j,k) \in \underline{Z}^3} (a_{ijk} + a_{jik} + a_{jki})\, x_j x_k$ if z_i is not protected, *i.e.*, if $x_i = 0$. As we have seen, this last value is equal to the area of z_i that is protected because of the protection of other zones whose intersection with z_i is not empty. The economic function, to be maximized, represents the weighted number of protected species. For each species s_k, the weighting is equal to the sum of the fractions of protected areas in the zones that protect that species, *i.e.*, the zones of Z_k. Either zone $z_i \in Z_k$ is selected – $x_i = 1$ and $\alpha_i = 0$ – and, in this case, the contribution of this zone to the economic function is equal to 1, or z_i is not selected – $x_i = 0$ and $\alpha_i \ge 0$ – and, in this case, the contribution of this zone is equal to α_i / a_i. Constraint 5.10.1 expresses that the total protected area must be less than or equal to A_{max}. Constraints 5.10.1 and 5.10.2 of program $\mathrm{P}_{5.10}$ are not linear. They can be linearized and thus a mixed-integer linear program is obtained. To do this, we replace each product $x_i x_j$ by variable y_{ij}, each product $x_i x_j x_k$ by variable v_{ijk}, and we add the set of constraints $\mathrm{C}_{5.2}$ below. The first 4 families of constraints correspond to the linearization of the products $x_i x_j$ and the next 5 families correspond to the linearization of the products $x_i x_j x_k$.

$$
\mathrm{C}_{5.2}: \left\{ \begin{array}{l} y_{ij} \le x_i \\ y_{ij} \le x_j \\ 1 - x_i - x_j + y_{ij} \ge 0 \\ y_{ij} \ge 0 \end{array} \right\} i < j, z_i \cap z_j \ne \emptyset, \qquad \begin{array}{l} v_{ijk} \le x_i \\ v_{ijk} \le x_j \\ v_{ijk} \le x_j \\ v_{ijk} \ge x_i + x_j + x_k - 2 \\ v_{ijk} \ge 0 \end{array} \right\} i < j < k, z_i \cap z_j \cap z_k \ne \emptyset
$$

The first 4 families of constraints concern only certain couples (i, j) and the following 5, only certain triplets (i, j, k). Indeed, in program $\mathrm{P}_{5.10}$, the products $x_i x_j$ appear with a non-zero coefficient if $i < j$ and $z_i \cap z_j \ne \emptyset$, and the products $x_i x_j x_k$ appear with a non-zero coefficient if $i < j < k$ and $z_i \cap z_j \cap z_k \ne \emptyset$.

5.5.4 Example

Let us take again the example described in figure 5.13 and use the linearization of program $\mathrm{P}_{5.10}$ to determine the best set of zones to select if the total area is limited to 94 units, *i.e.*, to the area of the non-optimal solution presented in figure 5.13.

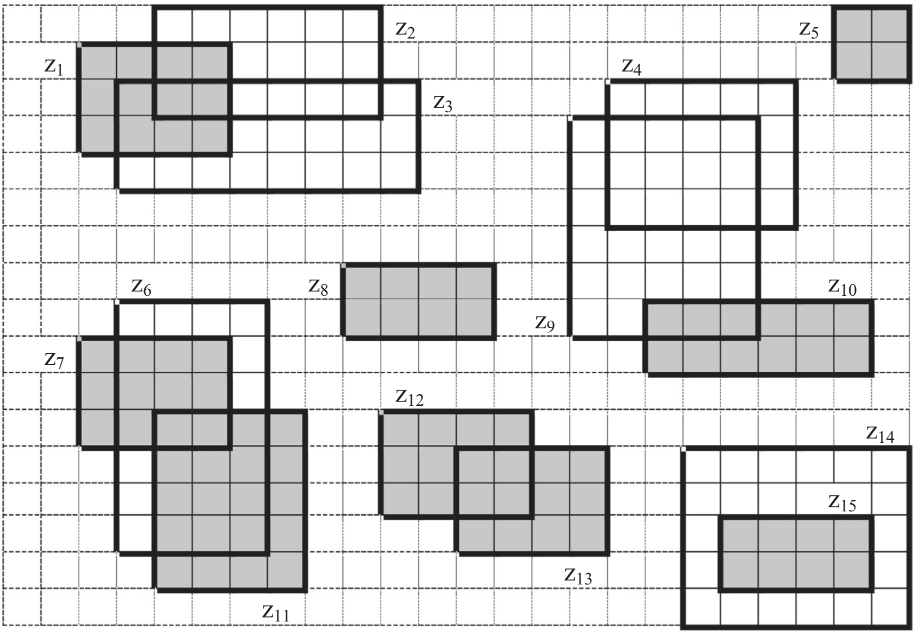


FIG. 5.14 – Each of the zones z_1, z_5, z_7, z_{10}, z_{12}, and z_{15} can participate in the protection of the 4 species s_1, s_2, s_3, and s_4, each of the zones z_2, z_4, z_8, z_{13}, and z_{14} can participate in the protection of the 3 species s_5, s_6, and s_7, and each of the zones z_3, z_6, z_9, and z_{11} can participate in the protection of the 3 species s_8, s_9, and s_{10}. The optimal solution is to decide to select the grey zones z_1, z_5, z_7, z_8, z_{10}, z_{11}, z_{12}, z_{13}, and z_{15}, and, therefore, not to select zones z_2, z_3, z_4, z_6, z_9, and z_{14}. However, the choice of the zones selected for protection implies that a fraction of zones z_2, z_3, z_6, z_9, and z_{14} is also protected. For this solution, the weighted number of protected species is 37.55 and the total area of protected zones is 94.

The zones to be selected are z_1, z_5, z_7, z_8, z_{10}, z_{11}, z_{12}, z_{13}, and z_{15} (see figure 5.14). The total area of these zones is equal to 94 units and the weighted number of protected species is equal to 37.55. This represents an improvement of about 25% over the solution shown in figure 5.13.

5.5.5 A Second Mathematical Programming Formulation

As before, the candidate zones for protection are considered to belong to a region represented by a grid of nr rows and nc columns. We put $M = \{1,\ldots, \text{nr}\}$ and $N = \{1,\ldots, \text{nc}\}$. All the cells in this grid are identical squares whose side length is equal to one unit. Each candidate zone is made up of a set of cells in the grid, all in one piece. The number of cells in zone z_i is denoted by n_i. Each cell is described by a pair composed of its row index and column index. In the example in figure 5.13 there are 17 rows and 24 columns and zone z_7 contains the 12 cells (10, 3), (10, 4), (10, 5), (10, 6), (11, 3), (11, 4), (11, 5), (11, 6), (12, 3), (12, 4), (12, 5), and (12, 6). In this new formulation of the problem, we use, as in the previous formulation, the Boolean variable x_i which is equal to 1 if and only if we decide to select zone z_i and we also use the Boolean variable t_{rc} which is equal to 1 if and only if the cell $(r,\ c)$ is selected (taking into account the decisions made regarding the zones to be selected). Note that it is not necessary to define variables t_{rc} on all the grid cells representing the region in question; it is sufficient to define them on all the cells belonging to at least one candidate zone. We denote by RC all the pairs $(r,\, c) \in M \times N$ such that the cell $(r,\ c)$ belongs to at least one zone. So, $\mathrm{RC} = \{(r,\, c) \in M \times N : \exists\, z_i \in Z \text{ such that } (r,\, c) \in z_i\}$. The notation "$(r,\, c) \in z_i$" means that the cell $(r,\, c)$ – r is the row index and c is the column index of this cell – is included in the zone z_i. Note also that in this formulation, variable α_i used in the previous formulation is no longer necessary. The linear program in Boolean variables $\mathrm{P}_{5.11}$ solves the problem.

$$
\mathrm{P}_{5.11}: \begin{cases}
\max \sum\limits_{k\in\underline{S}} \sum\limits_{i\in\underline{Z}_k} \sum\limits_{(r,c)\,\in\, z_i} t_{rc}/a_i & \\
\text{s.t.} \left| \begin{array}{llll}
\sum\limits_{(r,c)\,\in\,\mathrm{RC}} t_{rc} \le A_{\max} & & (5.11.1) & \mid \quad x_i \in \{0,1\} \quad i \in \underline{Z} \qquad (5.11.4) \\
\sum\limits_{(r,c)\,\in\, z_i} t_{rc} \ge n_i x_i & i \in \underline{Z} & (5.11.2) & \mid \quad t_{rc} \in \{0,1\} \quad (r,\, c) \in \mathrm{RC} \quad (5.11.5) \\
t_{rc} \le \sum\limits_{i\in\underline{Z}\,:\,(r,c)\,\in\, z_i} x_i & (r,\, c) \in \mathrm{RC} & (5.11.3) & \mid
\end{array} \right.
\end{cases}
$$

In the expression of the economic function of $\mathrm{P}_{5.11}$, the quantity $\sum_{(r,c)\in z_i} (t_{rc}/a_i)$ represents the proportion of the area of zone z_i that is protected. The economic function – to be maximized – therefore represents the weighted number of protected species. Constraint 5.11.1 expresses the area constraint since the total protected area is equal to $\sum_{(r,c)\in\mathrm{RC}} t_{rc}$. Constraints 5.11.2 express the fact that if it is decided to select zone z_i – $x_i = 1$ – then all the cells in this zone are protected. In other words, if $x_i = 1$, then $t_{rc} = 1$ for all the cells $(r,\ c)$ of z_i. According to constraints 5.11.3, a cell is selected if at least one of the zones containing it is selected. Constraints 5.11.4 and 5.11.5 specify the Boolean nature of variables x_i and t_{rc}.

5.5.6 Computational Experiments

The formulation of the problem by program $\mathrm{P}_{5.11}$ is much easier than by program $\mathrm{P}_{5.10}$ – and its linearization. Indeed, the formulation $\mathrm{P}_{5.10}$ requires the list of the zones, the area of each zone, but also the list of all the intersections of zones, 2 to 2, 3 to 3, etc. Formulation $\mathrm{P}_{5.11}$ only requires the list of zones and, for each zone, the list of the cells that compose it. In both formulations, it is also necessary to know, of course, the list of the species living in each zone. Table 5.2 gives some computational results with the formulation $\mathrm{P}_{5.11}$. The zones are rectangles distributed in a grid, as in figure 5.13. The coordinates, in the grid, of the cell located at the top left of each rectangle are drawn at random. The lengths of each side of the rectangles are random integers drawn uniformly between 1 and 50. The number of species is set at 200 and the presence of a given species in a given zone is also randomly selected with a certain probability.

TAB. 5.2 – Resolution of $P_{5.11}$: Some computational results on large-sized instances.

Dimension of the grid (nr × nc)	Total number of candidate zones (n)	Total area of candidate zones	Probability of occurrence of species s_k in the zone z_i	Maximal area of protection ($A_{\max}$)	Solution value	Number of zones selected	Total area	CPU time (s)	Number of nodes in the search tree
500 × 500	1,000	212,909	0.01	4,000	304.4	88	3,999	228	0
500 × 500	1,000	212,909	0.1	4,000	2,585.4	104	4,000	236	222
1,000 × 1,000	1,000	479,858	0.01	4,000	275.4	89	3,998	253	143
1,000 × 1,000	1,000	479,858	0.1	4,000	2,283.9	105	3,998	236	0
1,000 × 1,000	2,000	712,883	0.1	4,000	3,234.4	138	4,000	973	0
1,000 × 1,000	2,000	712,883	0.1	8,000	4,723.1	200	8,000	1,147	0
1,000 × 1,000	2,000	712,883	0.01	8,000	560.3	170	7,999	862	0

The results presented in table 5.2 show that large-sized instances of the problems can be solved relatively quickly. The longest instance to resolve requires about 19 min of CPU time. It can also be seen that the number of nodes developed in the search tree by the solver is low and often even zero. This is partly due to the fact that the value of the optimal solution of the continuous relaxation of program $P_{5.11}$, which is obtained by replacing $x_i \in \{0,1\}$ and $t_{rc} \in \{0,1\}$ by $0 \leq x_i \leq 1$ and $0 \leq t_{rc} \leq 1$, respectively, is not far from the value of the optimal solution of $P_{5.11}$.

References and Further Reading

Batisse M. (1982) The biosphere reserve: a tool for environmental conservation and management, *Environ. Conserv.* **9**, 101.

Batisse M. (1990) Development and implementation of the biosphere reserve concept and its applicability to coastal regions, *Environ. Conserv.* **17**, 111.

Billionnet A. (2010) Solving a cut problem in bipartite graphs by linear programming: application to a forest management problem, *Appl. Math. Model.* **34**, 1042.

Billionnet A. (2011) Spatial optimization of wildlife populations with probabilistic habitat connections, *For. Sci.* **57**, 336.

Clemens M.A., ReVelle C.S., Williams J.C. (1999) Reserve design for species preservation, *Euro. J. Oper. Res.* **112**, 273.

Ebregt A., De Greve P. (2000) *Buffer zones and their management - policy and best practices for terrestrial ecosystems in developing countries*. National Reference Centre for Nature Management, International Agricultural Centre, Wageningen, the Netherlands, p. 64.

Fletcher R.J. Jr. (2005) Multiple edge effects and their implications in fragmented landscapes, *J. Anim. Ecol.* **74**, 342.

Hamaide B., Williams J.C., ReVelle C.S. (2009) Cost-efficient reserve site selection favoring persistence of threatened and endangered species, *Geogr. Analysis* **41**, 66.

Hof J., Bevers M. (1998) *Spatial optimization for managed ecosystems*. Columbia University Press.

Holzkämper A., Lausch A., Seppelt R. (2006) Optimizing landscape configuration to enhance habitat suitability for species with contrasting habitat requirements, *Ecol. Model.* **198**, 277.

Laurance W.F., Yensen E. (1991) Predicting the impacts of edge effects in fragmented habitats, *Biol. Conserv.* **55**, 77.

Lindenmayer D., Franklin J., Eds (2002) *Conserving forest biodiversity - a comprehensive multiscaled approach*. Island Press.

Martino D. (2001) Buffer zones around protected areas: a brief literature review, *UCLA, Electro. Green J.* **1**, 20.

Millspaugh J.J., Thompson F.R., Eds. (2008) *Models for planning wildlife conservation in large landscape*. Elsevier.

Nemhauser G., Wolsey L.A. (1988) *Combinatorial optimization*. Wiley-Interscience.

Schonewald-Cox C., Bayless J. (1986) The boundary model: a geographical analysis of design and conservation of nature reserves, *Biol. Conserv.* **38**, 305.

Williams J.C., ReVelle C.S. (1996) A 0-1 programming approach to delineating protected reserves, *Environ. Plan. B: Plan. Des.* **23**, 607.

Williams J.C., ReVelle C.S. (1998) Reserve assemblage of critical areas: a zero-one programming approach, *Euro. J. Oper. Res.* **104**, 497.

Woodroffe R., Ginsberg J.R. (1998) Edge effects and the extinction of populations inside protected areas, *Science* **280**, 2126.

Chapter 6

Biological Corridors

6.1 Introduction

As we have pointed out in previous chapters, landscape fragmentation is an important cause of biodiversity loss. This fragmentation is mainly due to urbanization, agriculture and forestry. It prevents species from moving as they should because they would have to cross often inhospitable zones. These zones may, for example, lack food resources or may host many predators. The viability of the species concerned by this fragmentation then depends strongly on how the fragments can be connected. This connectivity between habitat zones within a landscape has become an essential element for biodiversity conservation. One of the options commonly used to establish – or restore – this connectivity is the establishment of corridors. Thus, the "trame verte et bleue" is a key measure of the Grenelle Environnement (set of political meetings organised in France in 2007 concerning actions to be undertaken in favour of the environment and in particular biodiversity) aimed at halting the decline of biodiversity through the preservation and restoration of ecological continuities or biological corridors. In the biological conservation literature, corridors have multiple definitions – and functions. They are natural spaces, generally linear, *i.e.*, longer than wide, allowing species to move through a fragmented set of zones that are natural habitats for them. These are therefore routes used by species to move, reproduce, flee, migrate, etc. They are highly dependent on the species of interest. They do not necessarily imply the notion of contiguous spaces. In other words, some routes can be easily used by some species – and thus be considered as corridors for these species – but not by others. For example, it will be difficult for some species to overcome obstacles such as transport infrastructure or zones treated with pesticides, which will not be the case for species capable of flying. However, it should be noted that the latter may face hunting when they move from one protected site to another. Another example is lighting, which can be a real obstacle, but only for certain nocturnal species (black corridors). The fact that species can move between the different zones without too many difficulties is an

DOI: 10.1051/978-2-7598-2521-9.c006

essential element for their survival. Indeed, these corridors allow, for example, the increase in population sizes, the resettlement of certain species in certain zones, the maintenance of genetic diversity, the access to different habitats and the increase in places for food. Corridors can also be used as a refuge for species when their usual habitat zones are threatened. In addition, some authors have also highlighted the value of corridors in the context of climate change, since it will force many species to migrate in order to conserve favourable habitats. The creation – or restoration – of these corridors is, therefore, one of the major strategies for protecting species threatened by habitat fragmentation. These corridors must themselves be zones favourable to the life of the species concerned, to enable them to feed, rest and protect themselves from predators during their movements. They can be of very different form and nature. Some studies clearly distinguish between habitat and travel functions in the characteristics of a corridor. Corridors or fractions of corridors can exist naturally. This is the case, for example, for agricultural hedges, riversides or old railway lines. They can be implemented through the protection of certain zones of the landscape. They may also include completely artificial elements such as wildlife crossings built above or below transport infrastructure. To fulfil their functions, these corridors must be made up of zones that benefit from some protection. It should be noted that the fact that biodiversity reservoirs are linked by a network of corridors may have certain disadvantages. Indeed, this network facilitates the movement between the reservoirs and is also an entry point to these reservoirs. It can therefore facilitate the spread of diseases, parasites, invasive species and predators from one reservoir to another, but also facilitate their introduction into the reservoirs. In addition, these corridors, which are often very long, are more difficult to control than reservoirs, which are generally more compact zones. This control concerns the threats we have just mentioned, but also, for example, hunting, poaching, and tourism. Also because of the ease of movement provided by the corridors, wildlife species that are present in reservoirs can become pests in other habitats such as agricultural or livestock zones. These species can also transmit diseases to non-wild species such as livestock and vice versa. It should also be noted that efforts to maintain the effectiveness of a corridor network consume significant human and financial resources. These could possibly be better used to protect other habitat zones, for example zones where the ratio expressing the area of the zone, divided by the length of its edge, is more important (see chapter 4). Finally, if the corridors are not well designed, they can present a high risk of mortality for the species that use them and thus contribute to their extinction. This mortality risk can come from predators encountered during the use of these corridors or from accidents occurring in crossing dangerous zones such as roads. It should be noted that the movements of certain species in certain corridors can take several years and even several generations. The reader can consult the many references cited at the end of this chapter for an in-depth discussion on corridor design and evaluation, a careful examination of the balance between ecological benefits and economic costs associated with maintaining or implementing corridors, and a presentation of the software available to assist in the design of these corridors. In this chapter, we present two optimization problems that we believe are representative of the design of a new corridor network or the restoration of an existing corridor network.

6.2 Least Cost Design of Corridor Networks

6.2.1 The Problem

We are interested in a landscape with a set of well-identified biodiversity reservoirs $BR_1, BR_2,..., BR_N$. These reservoirs are protected zones that provide habitat for a given set of species. To simplify the presentation, it is assumed here that any route that can be considered as a corridor for one of the species concerned can also be considered as a corridor for all the species concerned. Each reservoir is in one piece, meaning that all the species considered can fully traverse it without leaving it. In other words, these reservoirs can be considered as connected reserves (see chapter 3). Outside these reservoirs, the landscape has two types of zones to consider: zones already protected and providing habitat favourable to the species under consideration, and unprotected zones that can become protected zones and provide habitat favourable to the species under consideration.

Both types of zones can, therefore, contribute to the constitution of corridors. The second type corresponds either to completely new zones – from the protection point of view – or to old zones to be restored. A cost is associated with this second type of zones (figure 6.1). This cost can cover many aspects: monetary costs (rental or acquisition, possible restoration, management of the zones), ecological costs (travel facilities for species through the zone, mortality risk, distance travelled) and also social costs (negative or positive social impact generated by the selection of the zone to constitute a corridor). The consideration of monetary costs is obviously a key issue since financial resources are of course limited. In the following, we consider that the cost associated with the first type of zones is zero, but it would be very easy to consider a non-zero cost corresponding, for example, to the management of the zone.

The aim is to determine type 2 zones to be protected in order to connect, possibly using type 1 zones, all the biodiversity reservoirs. Two reservoirs are said to be connected if the species can move from one to the other only through either type 1 zones, type 2 zones that have been decided to be protected, or through one of the reservoirs. The selected type 2 zones, possibly with the addition of type 1 zones, form a network of corridors that link all the reservoirs. The problem we are studying here is to build this network of corridors at the lowest cost (figure 6.2). To simplify the presentation of the general problem of developing a network of corridors linking a set of biodiversity reservoirs, it is considered that the landscape is represented by a grid of nr $\times$ nc square and identical zones. Each zone of this landscape is denoted by z_{ij} where i denotes its row index and j, its column index. This landscape includes N biodiversity reservoirs, $BR_1, BR_2,..., BR_N$, each reservoir being formed by a connected subset of zones. These reservoirs are disjoint. It should be noted that the method we are going to propose would easily adapt to any other set of zones and reservoirs. As mentioned above, some of the zones that do not belong to the reservoirs are already protected and can provide habitat favourable to the species under consideration, while others can be protected and possibly restored to also provide habitat favourable to the species under consideration. The cost of protecting

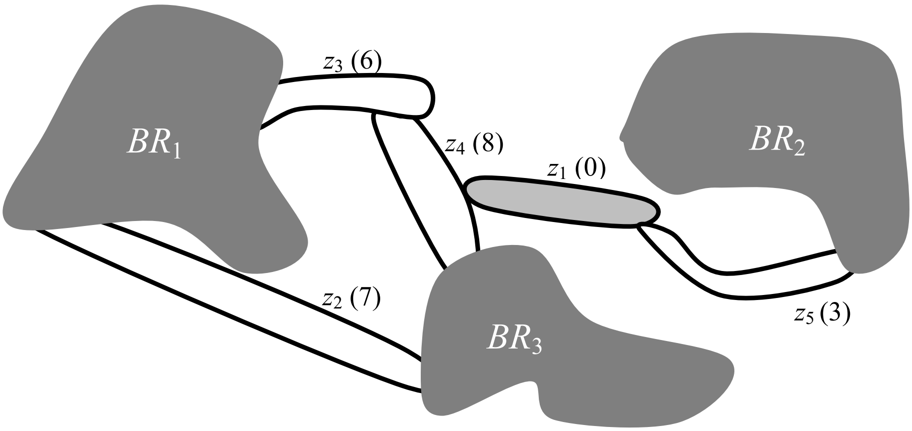


FIG. 6.1 – A hypothetical landscape with 3 biodiversity reservoirs, BR_1, BR_2, and BR_3, an already protected zone, z_1, which provides habitat favourable to the species under consideration, and 4 zones, z_2, z_3, z_4, and z_5, which can be protected and thus provide habitat favourable to the species under consideration, after possible restoration, with the associated costs indicated in brackets.

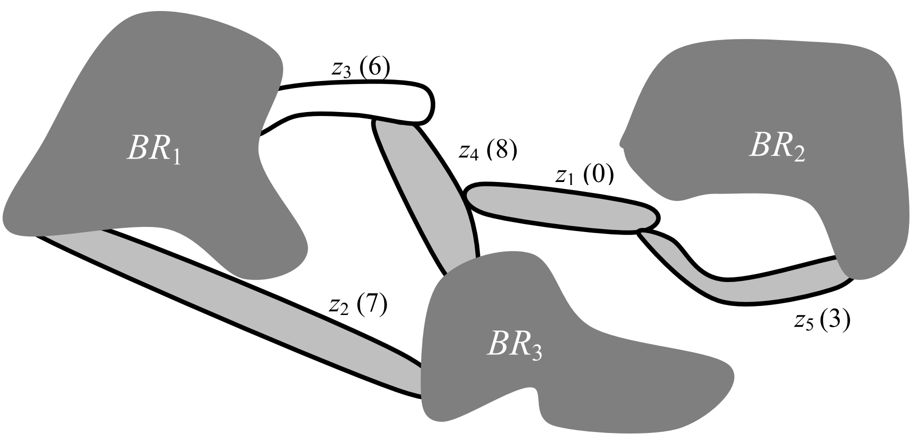


FIG. 6.2 – Zones z_1, z_2, z_4, and z_5 were selected. The corresponding cost is 18. The corridor connecting BR_1 to BR_2 consists of z_2, BR_3, z_4, z_1, and z_5, the corridor connecting BR_1 to BR_3 consists of the single zone z_2, and the corridor connecting BR_2 to BR_3 consists of z_5, z_1, and z_4.

zone z_{ij} is denoted by c_{ij}. In the case where z_{ij} is an already protected zone – not part of a reservoir – and provides habitat favourable to the species under consideration, this cost is zero. There are also zones in the considered landscape that, for different reasons, cannot be protected and, therefore, cannot contribute to the development of corridors (figure 6.3).

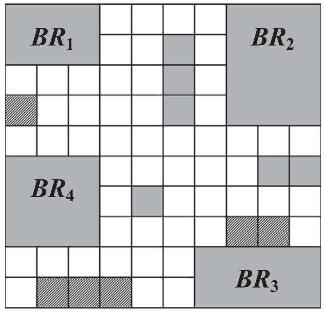


Fig. 6.3 – A hypothetical landscape represented by a grid of 10×10 square and identical zones. It includes 4 biodiversity reservoirs, BR_1, BR_2, BR_3, and BR_4, and 6 already protected zones that can contribute to the development of a corridor for the species under consideration, z_{26}, z_{36}, z_{46}, z_{69}, $z_{6,10}$, and z_{75}. The cost of protecting each zone – not yet protected – is equal to one unit; the cost associated with already protected zones is equal to 0; and finally, zones z_{41}, z_{88}, z_{89}, $z_{10,2}$, $z_{10,3}$, and $z_{10,4}$ cannot contribute to the development of a corridor.

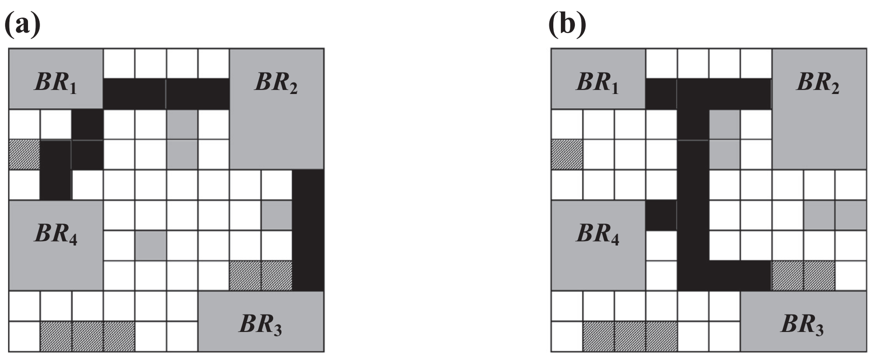


Fig. 6.4 – (a) The 12 black zones form a network of corridors linking the 4 biodiversity reservoirs BR_1, BR_2, BR_3, and BR_4. The associated cost is equal to 10 since, among these 12 zones, 2 were already protected. The length of the corridor connecting BR_3 and BR_4 is equal to 18. (b) The 13 black zones form a network of corridors linking the 4 reservoirs. The associated cost is 11 since, among these 13 zones, 2 were already protected. The length of the corridor connecting BR_3 and BR_4 is equal to 6.

The problem is to determine the zones to be protected, and possibly restored, in order to connect all the reservoirs at the lowest cost. Two reservoirs BR_i and BR_j are considered to be connected if it is possible, for the species under consideration, to move from BR_i to BR_j only through protected zones or zones belonging to a reservoir, and gradually moving from one zone to an adjacent one. Two zones are considered as adjacent if they share a common side. Figure 6.4 shows two different corridor networks to connect the 4 biodiversity reservoirs in figure 6.3.

The problem of connecting biodiversity reservoirs through a network of corridors has similarities to the problem of designing a connected reserve discussed in chapter 3. In both cases, the ultimate goal is to obtain a connected set of zones, *i.e.*, a set of zones in which species can move without leaving it. In chapter 3, the set of zones to be protected is determined in such a way as to ensure the best possible survival of certain species, taking into account protection costs. In this chapter, the set of zones to be protected is chosen in such a way as to link a set of already protected zones, at the lowest cost, and possibly taking into account certain constraints. Expressed in terms of graphs, both problems consist in determining, in a given graph, a subset of vertices inducing a connected sub-graph that checks certain constraints and takes into account certain costs.

6.2.2 Graph Optimization Formulation

Let us now look at how to state the problem as a graph optimization problem (see appendix at the end of the book). Let us associate to the grid of the nr $\times$ nc zones a graph, $G = (\underline{Z}, U)$, where the set of vertices, $\underline{Z}$, corresponds to the pairs of indices associated with a zone and where $((i, j), (k, l))$ is an arc of U if and only if zones z_{ij} and z_{kl} are adjacent – share a common side. For each biodiversity reservoir BR_k, let us choose one of its zones to represent it and denote by $z_{i(k),j(k)}$ this zone. The problem can be formulated as follows: determine a partial sub-graph of $G = (\underline{Z}, U)$, $G' = (\hat{\underline{Z}}, A)$, checking the following properties: all the vertices associated with a zone representing a reservoir belong to $\hat{\underline{Z}}$ and, for all $r \in \{1,..., N-1\}$, there is in this graph a path from the vertex associated with the zone representing reservoir BR_r to the vertex associated with the zone representing reservoir BR_N. This problem is similar to the Steiner tree problem which, in a general way, can be expressed as follows: given a graph whose edges are assigned with a weight, and a subset S of vertices of this graph, find a subset of edges of minimal weight that induces a connected sub-graph containing all the vertices of S (see appendix at the end of the book).

6.2.3 Mathematical Programming Formulation

We give below a flow type formulation of this problem (see appendix at the end of the book). Let ϕ_{ijkl} be the Boolean variable which is equal to 1 if and only if at least one of the $N-1$ paths, from the vertex representing reservoir BR_r, $r = 1,\ldots, N-1$, to the vertex representing reservoir BR_N, follows the arc $((i, j), (k, l))$ and let μ^r_{ijkl} be the Boolean variable which is equal to 1 if and only if, among these paths, the one from BR_r to BR_N follows the arc $((i, j), (k, l))$. Denote by Adj_{ij} the set of index pairs associated with the zones adjacent to zone z_{ij} and define constant d^r_{ij}, $(i, j) \in \underline{Z}$, $r = 1,..., N-1$, as follows: $d^r_{ij} = 1$ if zone z_{ij} represents reservoir BR_r, $d^r_{ij} = -1$ if zone z_{ij} represents reservoir BR_N, and $d^r_{ij} = 0$ in all the other cases. The problem can then be formulated as the 0–1 linear program $\mathrm{P}_{6.1}$.

$$\mathrm{P}_{6.1}: \begin{cases} \min \displaystyle\sum_{(i,j)\,\in\,\underline{Z}} c_{ij} \sum_{(k,l)\,\in\,\mathrm{Adj}_{ij}} \phi_{ijkl} \\ \text{s.t.} \left| \begin{array}{lll} \phi_{ijkl} \geq \mu^r_{ijkl} & ((i,j),(k,l)) \in U, r = 1,\ldots, N-1 & (6.1.1) \\ \displaystyle\sum_{(k,l)\in\,\mathrm{Adj}_{ij}} \mu^r_{ijkl} - \sum_{(k,l)\in\,\mathrm{Adj}_{ij}} \mu^r_{klij} = d^r_{ij} & (i,j) \in \underline{Z}, r = 1,\ldots, N-1 & (6.1.2) \\ \phi_{ijkl} \in \{0,1\} & ((i,j),(k,l)) \in U & (6.1.3) \\ \mu^r_{ijkl} \in \{0,1\} & ((i,j),(k,l)) \in U, r = 1,\ldots, N-1 & (6.1.4) \end{array} \right. \end{cases}$$

In this formulation $c_{ij} = 0$ if zone z_{ij} is already protected or belongs to a reservoir. We give below some indications to justify the formulation $\mathrm{P}_{6.1}$. First of all, it can be shown that the optimal solution to the problem has a tree structure. More precisely, the selected arcs, *i.e.*, the arcs $((i, j), (k, l))$ such that $\phi_{ijkl} = 1$ and the corresponding vertices satisfy the following property: any vertex (i, j) selected and different from the vertex $(i(N), j(N))$ is the initial end of one and only one selected arc. Thus $\sum_{(k,l)\,\in\,\mathrm{Adj}_{ij}} \phi_{ijkl} = 1$ for any vertex (i, j) selected and different from $(i(N), j(N))$, and $\sum_{(k,l)\,\in\,\mathrm{Adj}_{ij}} \phi_{ijkl} = 0$ for any vertex (i, j) not selected or when $(i, j) = (i(N), j(N))$. We deduce from this that the vertex (i, j), different from $(i(N), j(N))$, is selected if and only if $\sum_{(k,l)\,\in\,\mathrm{Adj}_{ij}} \phi_{ijkl} = 1$. It should also be noted that, for all $r \in 1,..., N-1$, any vertex (i, j) is the initial end of at most one arc of the path from the vertex representing reservoir BR_r to the vertex representing reservoir BR_N, and also the terminal end of at most one arc of the same path. Thus, for any vertex (i, j), $\sum_{(k,l)\in\,\mathrm{Adj}_{ij}} \mu^r_{ijkl} \leq 1$ and $\sum_{(k,l)\in Adj_{ij}} \mu^r_{klij} \leq 1$. Constraints 6.1.1 force variable ϕ_{ijkl} to take the value 1 if at least one of variables μ^r_{ijkl}, $r = 1,..., N-1$, takes the value 1. Consider constraints 6.1.2 for the 3 types of zones. If zone z_{ij} represents reservoir BR_N – $d^r_{ij} = -1$ for all $r \in 1,..., N-1$ – these constraints express, taking into account the above remarks, that for all $r \in 1,..., N-1$, $\sum_{(k,l)\in Adj_{ij}} \mu^r_{klij} = 1$ and $\sum_{(k,l)\in\,\mathrm{Adj}_{ij}} \mu^r_{ijkl} = 0$. If zone z_{ij} represents reservoir BR_r, – $d^r_{ij} = 1$ – these constraints express that for all $(i,j) \in \underline{Z}$ and for all $r \in 1,..., N-1$ such that z_{ij} represents reservoir BR_r, $\sum_{(k,l)\in A_{ij}} \mu^r_{klij} = 0$ and $\sum_{(k,l)\in A_{ij}} \mu^r_{ijkl} = 1$. Finally, if zone z_{ij} does not represent any of reservoirs – $d^r_{ij} = 0$ for all $r \in 1,..., N-1$ – these constraints express that, for all $(i,j) \in \underline{Z}$ and for all $r \in 1,..., N-1$, if (i, j) is the terminal end of an arc of the path from the vertex representing reservoir BR_r to the vertex representing reservoir BR_N, then (i, j) is also the initial end of an arc of the same path.

This type of formulation has been used to define a network of corridors suitable for grizzly bear movement in the northern Rocky Mountains of the United States. A disadvantage of this formulation is that the lengths of the corridors connecting two reservoirs cannot be controlled in the searched solution except for the pairs of reservoirs (BR_r, BR_N), $r = 1,..., N-1$. In this case, it is sufficient to add the constraint $\sum_{(i,j,k,l)\in I_{r,N}} \mu^r_{ijkl} \leq L^{rN}_{\max}$, where $I_{r,N} = \{((i,j),(k,l)) \in U, z_{ij} \notin BR_r \cup BR_N\}$ and $L^{rN}_{\max}$ indicates the maximal authorised length for the corridor connecting BR_r to BR_N. In other words, $I_{r,N}$ refers to the set of arcs for which the zone associated with their initial end does not belong to either reservoir BR_r or reservoir BR_N.

We propose below a slightly different formulation of the corridor design problem that does not have this disadvantage. We keep variables μ^r_{ijkl} with their same meaning and replace variables ϕ_{ijkl} by variables x_{ij} that are equal to 1 if and only if at least one of the N−1 paths from, BR_r, $r = 1,\ldots, N-1$, to BR_N passes through the vertex (i, j). The result is program $\mathrm{P}_{6.2}$, which has fewer variables and fewer constraints than $\mathrm{P}_{6.1}$, and allows a limit to be imposed on the length of the corridor connecting any two reservoirs.

$$
\mathrm{P}_{6.2}: \begin{cases} \min \sum_{(i,j)\in \underline{Z}} c_{ij}x_{ij} \\ \text{s.t.} \left| \begin{array}{lll} x_{ij} \geq \sum_{(k,l)\in \mathrm{Adj}_{ij}} \mu^r_{ijkl} & (i,j)\in \underline{Z}, r=1,\ldots,N-1 & (6.2.1) \\ \sum_{(k,l)\in \mathrm{Adj}_{ij}} \mu^r_{ijkl} - \sum_{(k,l)\in \mathrm{Adj}_{ij}} \mu^r_{klij} = d^r_{ij} & (i,j)\in \underline{Z}, r=1,\ldots,N-1 & (6.2.2) \\ x_{ij}\in\{0,1\} & (i,j)\in \underline{Z} & (6.2.3) \\ \mu^r_{ijkl}\in\{0,1\} & ((i,j),(k,l))\in U, r=1,\ldots,N-1 & (6.2.4) \end{array} \right. \end{cases}
$$

Constraints 6.2.1 express that, if at least one of the N−1 paths from the vertex representing reservoir BR_r to the vertex representing reservoir BR_N passes through an arc of initial end (i, j), then zone z_{ij} is retained. Constraints 6.2.2 are identical to constraints 6.1.2.

As in the previous formulation, a constraint can be introduced limiting the length of the corridor connecting BR_r and BR_N. To limit to $L^{st}_{\max}$ the length of the corridor connecting any two reservoirs, BR_s and BR_t, a new Boolean variable ψ^{st}_{ijkl} is defined, which is equal to 1 if and only the path from the vertex representing BR_s to the vertex representing BR_t follows the arc $((i, j), (k, l))$ and the set of constraints $\mathrm{C}_{6.1}$ is added where $\delta^{st}_{ij} = 1$ if z_{ij} represents BR_s, $\delta^{st}_{ij} = -1$ if z_{ij} represents BR_t, and $\delta^{st}_{ij} = 0$ in the other cases.

$$
\mathrm{C}_{6.1}: \begin{cases} \sum_{((i,j),(k,l))\in U, z_{ij}\notin BR_s \cup BR_t} \psi^{st}_{ijkl} \leq L^{st}_{\max} & \\ x_{ij} \geq \sum_{(k,l)\in \mathrm{Adj}_{ij}} \psi^{st}_{ijkl} & (i,j)\in \underline{Z} \\ \sum_{(k,l)\in \mathrm{Adj}_{ij}} \psi^{st}_{ijkl} - \sum_{(k,l)\in \mathrm{Adj}_{ij}} \psi^{st}_{klij} = \delta^{st}_{ij} & (i,j)\in \underline{Z} \end{cases}
$$

6.2.4 Example

The hypothetical landscape studied is represented by a grid of 20 × 20 square and identical zones and includes 7 biodiversity reservoirs, BR_1, $BR_2,\ldots, BR_7$. Among the zones that do not belong to the reservoirs, some are already protected and provide habitat favourable to the species under consideration, others can be protected and possibly restored to also provide habitat favourable to the species under

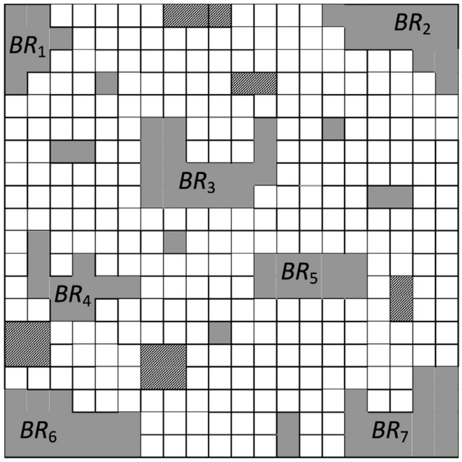


FIG. 6.5 – A hypothetical landscape represented by a grid of 20×20 square and identical zones. It includes 7 biodiversity reservoirs, BR_1, BR_2, BR_3 BR_4, BR_5, BR_6, and BR_7, and 11 zones already protected and providing habitat favourable to the species concerned, z_{45}, $z_{6,15}$, z_{73}, z_{74}, $z_{9,17}$, $z_{9,18}$, $z_{11,8}$, $z_{15,10}$, $z_{19,13}$, and $z_{20,13}$. The cost associated with the not already protected zones is equal to one unit and the cost associated with the already protected zones is equal to 0. Finally, zones z_{18}, z_{19}, $z_{1,10}$, $z_{4,11}$, $z_{4,12}$, $z_{13,18}$, $z_{14,18}$, $z_{15,1}$, $z_{15,2}$, $z_{16,1}$, $z_{16,2}$, $z_{16,7}$, $z_{16,8}$, $z_{17,7}$, and $z_{17,8}$ cannot be protected.

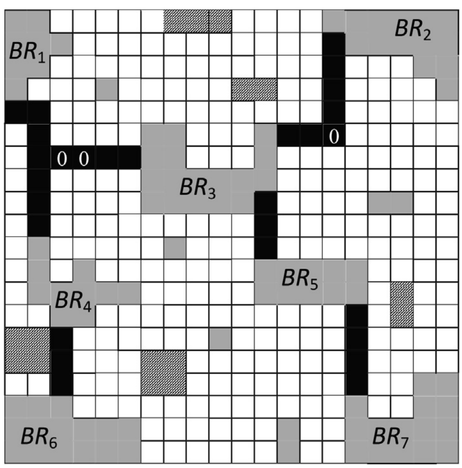


FIG. 6.6 – Optimal solution associated with the instance of figure 6.5. The corridors are shown in black. The protection and possible restoration of a total of 28 zones allows all the reservoirs to be connected. The cost associated with each black zone is equal to 1 except for the zones that were already protected. For these zones, the cost is 0 and is shown in the figure. The total cost of the corridor network is equal to 25 units. The length of the corridor connecting BR_4 and BR_5 is equal to 18.

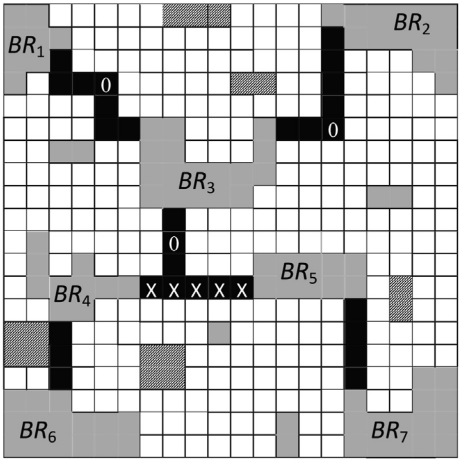


FIG. 6.7 – Optimal solution associated with the instance described in figure 6.5 when the length of the corridor connecting reservoirs BR_4 and BR_5 is required to be less than or equal to 9. The cost of this solution is equal to 26 units. The zones constituting the corridor linking BR_4 and BR_5 are marked with a cross. The length of this corridor is equal to 5.

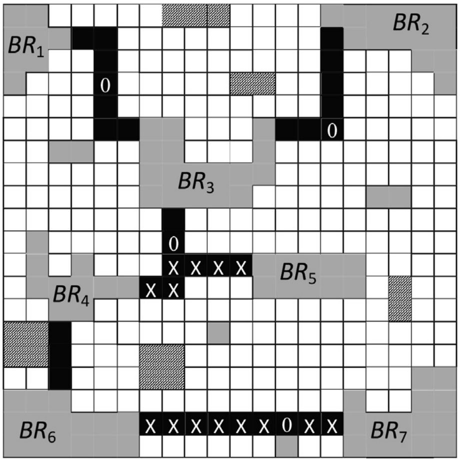


FIG. 6.8 – Optimal solution associated with the instance described in figure 6.5 when the length of the corridor connecting reservoirs BR_4 and BR_5 is required to be less than or equal to 9, and the length of the corridor connecting reservoirs BR_6 and BR_7 is required to be less than or equal to 11. The cost of this solution is equal to 30 units. The zones constituting the corridors connecting BR_4 to BR_5 and BR_6 to BR_7 are marked with a cross. The corresponding lengths are 6 and 9, respectively. On the other hand, the length of the corridor connecting reservoirs BR_5 and BR_7 increases significantly, compared to the previous solution, from 4 to 28.

consideration. In addition, some zones cannot be protected and will, therefore, not be able to contribute to the constitution of corridors (figure 6.5).

Figure 6.6 shows the low-cost corridor network linking the 7 reservoirs. Figure 6.7 shows the least-cost network when the length of the corridor connecting BR_4 and BR_5 is limited to 9 and figure 6.8 presents the least-cost network when, in addition, the length of the corridor connecting BR_6 and BR_7 is limited to 11.

6.3 Optimizing the Permeability of an Existing Corridor Network Under a Budgetary Constraint

6.3.1 The Problem

This problem consists in improving and/or restoring an existing network of corridors with the best cost-effectiveness ratio. In other words, there is a certain budget available to carry out developments to improve the permeability of the network and the aim is to carry out these developments in such a way as to increase this permeability as much as possible, while respecting the financial constraint. Some authors have considered this type of problem but have proposed – approximate – resolutions based on simulation methods. We present here an – exact – resolution based on mixed-integer linear programming. We consider a network of corridors and a set of species all having the same behaviour in this network. The network is represented by a graph, $G = (\underline{BR}, C)$, where $\underline{BR}$ is the set of indices associated with the set of the N biodiversity reservoirs, BR_1, BR_2,…, BR_N, corresponding to habitats favourable to the species under consideration, and where C is the set of arcs. For any couple, $(i, j) \in \{1,..., N\}^2,\ i \neq j$, (i, j) is an arc of the graph if there is a corridor between BR_i and BR_j. Note that $G = (\underline{BR}, C)$ is a symmetric graph. For various reasons – road and rail infrastructure, urbanization, agriculture, etc. – the condition of these corridors is more or less deteriorated. The problem is to restore this network of corridors as efficiently as possible, *i.e.*, to optimize its permeability, under a budgetary constraint. This permeability is measured by the mathematical expectation of the distance travelled in the network by the species under consideration. It is assumed that when an animal is in reservoir BR_i, it randomly and equiprobably chooses one of the corridors leading to this reservoir – and thus also leaving this reservoir. It thus chooses the corridor $[BR_i, BR_j]$ with the probability $1/\mathrm{d}_i$ where d_i indicates the degree of the vertex associated with reservoir BR_i, and then tries, eventually, to use this corridor. A certain probability is associated with this possibility. If it decides to use the corridor, it is assumed that it succeeds in reaching the other end, *i.e.*, reservoir BR_j, also with a certain probability and that it does not succeed in reaching it, being killed beforehand, with the complementary probability. Restoring a corridor $[BR_i, BR_j]$ increases the last two probabilities – trying to follow the corridor and succeeding in its course. The more resources are devoted to restoration, the higher the values of these probabilities. For a given corridor, several levels of investment are possible. The set of these levels, for the corridor connecting reservoirs BR_i and BR_j, is designated by $H_{ij} = \{0, 1, ..., h_{ij}\}$ with $h_{ij} = h_{ji}$. It is assumed that for each corridor, the possible values of the different probabilities mentioned above and the associated costs are known. The level 0 investment consists in doing nothing – the corridor remains in its current state – and costs 0. Denote by $r^1_{ijh}, (i, j) \in C, h \in H_{ij}$, the probability for an animal, located in BR_i, to try to use the corridor $[BR_i, BR_j]$ if level h investment is made in this corridor and $r^2_{ijh}, (i, j) \in C, h \in H_{ij}$, the probability, having chosen to use the corridor, to reach BR_j. These probabilities are not necessarily symmetric. Thus, probability r^1_{ijh} may be different from probability r^1_{jih} and probability r^2_{ijh} may be different from probability r^2_{jih}.

6.3.2 Associated Markov Chain

With the corridor network is associated a Markov chain (see appendix at the end of the book) whose set of states is made up of N transient states corresponding to the N reservoirs and a $(N + 1)$th, absorbing, state corresponding to the death of the animal. These states are denoted by 1, 2,…, N, $N + 1$. We denote by pr_{ij}, $i = 1, ..., N + 1, j = 1, ..., N + 1$, the transition probability from state i to state j. The probability $\mathrm{pr}_{N+1,N+1}$ is equal to 1 and, for all $j \in \{1, ..., N\}$, the probability $\mathrm{pr}_{N+1,j}$ is equal to 0. The probability pr_{ij}, $i = 1, ..., N, j = 1, ..., N$, corresponds to the probability that an animal present in reservoir BR_i at time t is present in reservoir BR_j at time $t + 1$. Note that the probability pr_{ii}, $i = 1,…, N$, is to be considered. It corresponds to the fact that an animal, present in reservoir BR_i at time t, can give up using one of the corridors leaving BR_i and thus be again in BR_i at time $t + 1$. The probability $\mathrm{pr}_{i,N+1}$, $i = 1, ..., N$, corresponds to the probability that an animal in reservoir BR_i at time t is dead at time $t + 1$. It is assumed that at the initial moment there is an animal in each of the N transient states, *i.e.*, in each of the reservoirs. The duration of a transition will depend on the context of the study and in particular on the type of corridor networks and the type of species considered.

Let us consider the transition probability matrix, $\Pi = \begin{pmatrix} Z & D \\ 0 & 1 \end{pmatrix}$, Z corresponding to the transition probabilities between transient states and D, to the transition probabilities from transient states to the absorbing state. Let us denote by $\mathcal{N}$ the $N \times N$ – matrix whose general term, n_{ij}, represents the expected number of passages through transient state j for an animal starting from transient state i, before being absorbed, *i.e.*, before being in the state $N + 1$. According to Markov's chain theory, $\mathcal{N} = (I - Z)^{-1}$ where I denotes the $N \times N$ identity matrix. Let $w_i = \sum_{j=1}^{N} n_{ji}$, $i = 1, ..., N$. The quantity w_i thus represents the expected total number of passages through state i, before being absorbed. We deduce that the expected total number of routes in the corridor $[BR_i, BR_j]$, from BR_i to BR_j, is equal to $w_i \mathrm{pr}_{ij}$. We can show that the only solution of the system of equations $w_i - \sum_{j=1}^{N} w_j \mathrm{pr}_{ji} = 1$, $i = 1, ..., N$, in which the quantities w_i, $i = 1,…, N$, are the unknowns, checks $w_i = \sum_{j=1}^{N} n_{ji}$.

6.3.3 Mathematical Programming Formulation

The problem of choosing the investments to be made in each corridor in order to maximize the expected value of the total distance travelled by N animals, one animal being initially located in each of the N reservoirs, can therefore be formulated as the mathematical program $\mathrm{P}_{6.3}$.

$$\mathrm{P}_{6.3}: \begin{cases} \max \displaystyle\sum_{(i,j)\in C,\, i<j} l_{ij}(w_i \mathrm{pr}_{ij} + w_j \mathrm{pr}_{ji}) & \\ \text{s.t.} \left| \begin{array}{lll} w_i - \displaystyle\sum_{j=1}^{N} w_j \mathrm{pr}_{ji} = 1 & i = 1,\ldots, N & (6.3.1) \\ \Pi \in \tilde{\Pi} & & (6.3.2) \end{array} \right. \end{cases}$$

where $C = \{(i,j) : [BR_i, BR_j] \text{ is a corridor}\}$, l_{ij} is the length of the corridor $[BR_i, BR_j]$, w_i is a real variable that represents the expression $\sum_{j=1}^{N} n_{ji}$ and $\tilde{\Pi}$ is a set of stochastic matrices, of dimension $(N+1) \times (N+1)$, of general term pr_{ij} and admissible for the problem. It should be recalled that the set of possible investment levels in the corridor $[BR_i, BR_j]$ is $H_{ij} = \{1, 2,\ldots, h_{ij}\}$ with $h_{ij} = h_{ji}$. Let x_{ijh}, $(i,j) \in C$, $h \in H_{ij}$, be the Boolean variable which is equal to 1 if and only if the level h investment is made in the corridor $[BR_i, BR_j]$ and c_{ijh}, be the cost of this investment. This cost is defined for $i < j$. Remember that for all $(i,j) \in C, i<j$, $c_{ij0} = 0$. We put $x_{ijh} = x_{jih}$. Let en_{ij}, $(i,j) \in C$, be the positive or zero variable that represents the expected total number of routes in the corridor $[BR_i, BR_j]$, from BR_i to BR_j, *i.e.*, the quantity $w_i \mathrm{pr}_{ij}$. The problem considered can then be formulated as program $\mathrm{P}_{6.4}$.

$$\mathrm{P}_{6.4}: \begin{cases} \max \displaystyle\sum_{(i,j)\in C, i<j} l_{ij}(\mathrm{en}_{ij} + \mathrm{en}_{ji}) & \\ \text{s.t.} \left| \begin{array}{lll} \displaystyle\sum_{(i,j)\in C, i<j, h\in H_{ij}} c_{ijh} x_{ijh} \le B & & (6.4.1) \\ \displaystyle\sum_{h\in H_{ij}} x_{ijh} = 1 & (i,j) \in C, i<j & (6.4.2) \\ \mathrm{en}_{ij} = \frac{w_i}{d_i} \displaystyle\sum_{h\in H_{ij}} r^1_{ijh}\, r^2_{ijh}\, x_{ijh} & (i,j) \in C & (6.4.3) \\ \frac{w_i}{d_i} \displaystyle\sum_{j:(i,j)\in C, h\in H_{ij}} r^1_{ijh}\, x_{ijh} & & \\ \quad = 1 + \displaystyle\sum_{j:\,(i,j)\in C} \mathrm{en}_{ji} & i = 1,\ldots, N & (6.4.4) \\ x_{ijh} = x_{jih} & (i,j) \in C, i<j, h \in H_{ij} & (6.4.5) \\ x_{ijh} \in \{0,1\} & (i,j) \in C, h \in H_{ij} & (6.4.6) \\ w_i \ge 0 & i = 1,\ldots, N & (6.4.7) \\ \mathrm{en}_{ij} \ge 0 & (i,j) \in C & (6.4.8) \end{array} \right. \end{cases}$$

The economic function of $\mathrm{P}_{6.4}$ expresses the expected total distance travelled in the corridors. Constraint 6.4.1 expresses the financial constraint. Constraints 6.4.2

and 6.4.5 express that, for each corridor, only one level of investment must be selected. Constraints 6.4.3 express the expected number of routes in the corridor $[BR_i,\ BR_j]$, from BR_i to BR_j. Constraints 6.4.4 reflect constraints 6.3.1. Indeed, these last constraints can be written as $w_i = 1 + \sum_{j=1,\dots,N,j\neq i} w_j \mathrm{pr}_{ji} + w_i \mathrm{pr}_{ii}$ or alternatively $w_i(1-\mathrm{pr}_{ii}) = 1 + \sum_{j=1,\dots,N,j\neq i} \mathrm{en}_{ji}$. Let us express probability pr_{ii}, *i.e.*, the probability, for an animal present at time t in reservoir BR_i, of being present again in this reservoir at time $t+1$. This occurs when the animal chooses any corridor leaving from BR_i and renounces trying to travel that corridor. Remember that an animal present in BR_i chooses the corridor $[BR_i,\ BR_j]$ with probability $1/d_i$. Moreover, when it has chosen the corridor $[BR_i,\ BR_j]$, it tries to use it with probability r^1_{ijh} if level h investment has been made in this corridor. So we have $\mathrm{pr}_{ii} = \sum_{j:(i,j)\in C} \left(1 - \sum_{h\in H_{ij}} r^1_{ijh} x_{ijh}\right)/d_i$ and constraints 6.3.1 can therefore be written $w_i\left(1 - \sum_{j:\,(i,j)\in C} \left(1 - \sum_{h\in H_{ij}} r^1_{ijh} x_{ijh}\right)/d_i\right) = 1 + \sum_{j=1,\dots,N,j\neq i} \mathrm{en}_{ji}$ or alternatively $\frac{w_i}{d_i} \sum\limits_{j:(i,j)\in C, h\in H_{ij}} r^1_{ijh}\, x_{ijh} = 1 + \sum_{j=1,\dots,N,j\neq i} \mathrm{en}_{ji}$. $\mathrm{P}_{6.4}$ can be transformed into a mixed-integer linear program by linearizing the quadratic expressions $w_i x_{ijh}$, which are products of the real, non-negative variable w_i by the Boolean variable x_{ijh} (see appendix at the end of this book). To do this, we replace each product $w_i x_{ijh}$ with variable v_{ijh} and add the set of linear constraints $\mathrm{C}_{6.2}$ below to force v_{ijh} to be equal to $w_i x_{ijh}$, $(i,j) \in C, h \in H_{ij}$.

$$
\mathrm{C}_{6.2} : \begin{cases} v_{ijh} \leq \mathrm{UB}_i\, x_{ijh} & (i,j) \in C, h \in H_{ij} \\ \sum\limits_{h\in H_{ij}} v_{ijh} = w_i & (i,j) \in C \\ v_{ijh} \geq 0 & (i,j) \in C, h \in H_{ij} \end{cases}
$$

UB_i is a constant greater than or equal to the optimal value of w_i in program $\mathrm{P}_{6.4}$. By examining successively the two possible values of x_{ijh}, while taking into account constraints 6.4.2 and 6.4.5, we see that $v_{ijh} = w_i x_{ijh}$ if and only if the constraints of $\mathrm{C}_{6.2}$ are satisfied. Finally, the problem can be solved by program $\mathrm{P}_{6.5}$.

$$
\mathrm{P}_{6.5} : \begin{cases} \max \sum\limits_{(i,j)\in C,\, i<j} l_{ij}\left(\mathrm{en}_{ij} + \mathrm{en}_{ji}\right) \\ \text{s.t.} \left| \begin{array}{lll} (6.4.1), (6.4.2), (6.4.5), (6.4.6), (6.4.7), (6.4.8) & & \\ \mathrm{en}_{ij} = \frac{1}{d_i} \sum\limits_{h\in H_{ij}} r^1_{ijh}\, r^2_{ijh}\, v_{ijh} & (i,j) \in C & (6.5.1) \\ \frac{1}{d_i} \sum\limits_{\substack{j:(i,j)\in C \\ h\in H_{ij}}} r^1_{ijh}\, v_{ijh} = 1 + \sum\limits_{j:(i,j)\in C} \mathrm{en}_{ji} & i = 1,\dots, N & (6.5.2) \\ v_{ijh} \leq \mathrm{UB}_i\, x_{ijh} & (i,j) \in C, h \in H_{ij} & (6.5.3) \\ \sum\limits_{h\in H_{ij}} v_{ijh} = w_i & (i,j) \in C & (6.5.4) \\ v_{ijh} \geq 0 & (i,j) \in C, h \in H_{ij} & (6.5.5) \end{array} \right. \end{cases}
$$

6.3.4 Example 1

Consider the example described in figure 6.9 and table 6.1.

In this example, we assume that there are 4 possible types of restoration for each corridor to reduce the barrier effect – reflected by the probabilities r^1_{ijh} and r^1_{jih} – and mortality risk – reflected by the probabilities r^2_{ijh} and r^2_{jih}. Table 6.1 gives, for each corridor $[BR_i, BR_j]$, its length, l_{ij}, the probabilities r^1_{ijh}, r^1_{jih}, r^2_{ijh}, and r^2_{jih}, for $h = 0$, 1, …, 4, and the associated costs, c_{ijh}, for $h = 0$, 1,…, 4. By definition, c_{ij0} is equal to 0. In this example, the effects of the possible restorations for each corridor are not symmetric, neither with regard to the barrier effect since r^1_{ijh} may be different from r^1_{jih}, nor with regard to the mortality risk since r^2_{ijh} may be different from r^2_{jih}. Note that, in this example, r^1_{ijh}, r^1_{jih}, r^2_{ijh}, r^2_{jih}, and c_{ijh} are increasing as a function of h.

The computational experiments were conducted with different values of the available budget, B. The results are presented in table 6.2. Remember that, in this example, the effects of corridor restoration are not symmetric, neither in terms of barrier effect nor in terms of mortality risk. In order to obtain, among the equivalent solutions of $\mathrm{P}_{6.5}$, a minimal cost solution, we subtract from the objective function the quantity $\varepsilon \sum_{(i,j)\in C, i<j, h\in H_{ij}} c_{ijh}x_{ijh}$, where ε is a sufficiently small constant.

If no restoration is carried out in the corridors, the expected total distance travelled is 60 km, and if the best possible restoration is carried out in view of the pursued objective – which requires a budget of 133 units – this expected distance becomes equal to 213 km. The results in table 6.2 show that, in some cases, it is not

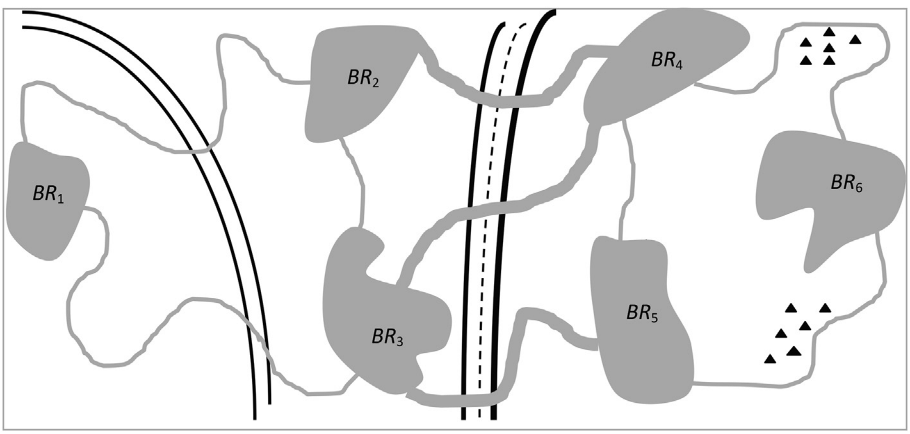


FIG. 6.9 – A hypothetical network of corridors associated with 6 biodiversity reservoirs. The corridors $[BR_1, BR_2]$, $[BR_1, BR_3]$, $[BR_4, BR_6]$, and $[BR_5, BR_6]$ are long and narrow. Dwellings are located near the corridors $[BR_4, BR_6]$ and $[BR_5, BR_6]$. The corridors $[BR_2, BR_3]$ and $[BR_4, BR_5]$ are short and narrow. The corridors $[BR_2, BR_4]$, $[BR_3, BR_4]$, and $[BR_3, BR_5]$ are relatively short and wide. These last 3 corridors are crossed by a main road and the corridors $[BR_1, BR_2]$ and $[BR_1, BR_3]$ are crossed by a small road.

TAB. 6.1 – Each cell in columns 2 to 6 shows, for a corridor $[BR_i, BR_j]$ of the network in figure 6.9 and for a given value of the investment level $h \in \{0, 1, 2, 3, 4\}$, the probabilities r^1_{ijh}, r^1_{jih}, r^2_{ijh}, r^2_{jih}, and the associated costs, c_{ijh}, in this order. For example, in the cell located at the intersection of row [3, 5] – associated with the corridor $[BR_3, BR_5]$ – and column $h = 3$, $r^1_{353} = 0.7$, $r^1_{533} = 0.7$, $r^2_{353} = 0.9$, $r^2_{533} = 0.8$, and $c_{353} = 12$. The last column of the table shows the length of each corridor, l_{ij}, in kilometres. The total cost associated with the maximal investments that can be made in each corridor is equal to 150.

$[i, j]$	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 4$	l_{ij}
[1, 2]	0.2/0.2/0.7/0.7/0	0.5/0.6/0.8/0.8/3	0.7/0.8/0.9/0.9/8	0.8/0.8/0.9/0.9/11	0.9/0.9/0.8/0.9/15	10
[1, 3]	0.2/0.2/0.6/0.6/0	0.3/0.4/0.7/0.7/4	0.6/0.5/0.7/0.7/10	0.7/0.7/0.7/0.8/13	0.9/0.8/0.8/0.9/17	10
[2, 3]	0.2/0.2/0.5/0.5/0	0.5/0.6/0.6/0.6/2	0.5/0.7/0.7/0.6/7	0.7/0.8/0.7/0.7/11	0.8/0.9/0.8/0.9/17	2
[2, 4]	0.2/0.2/0.6/0.7/0	0.4/0.6/0.7/0.7/4	0.7/0.7/0.7/0.7/9	0.8/0.8/0.9/0.8/13	0.9/0.9/0.8/0.9/16	5
[3, 4]	0.2/0.2/0.6/0.6/0	0.4/0.4/0.6/0.6/3	0.7/0.7/0.8/0.6/8	0.8/0.8/0.8/0.8/15	0.9/0.9/0.8/0.9/19	6
[3, 5]	0.2/0.3/0.5/0.7/0	0.4/0.5/0.6/0.7/5	0.6/0.7/0.8/0.7/9	0.7/0.7/0.9/0.8/12	0.9/0.9/0.8/0.9/18	5
[4, 5]	0.2/0.2/0.7/0.7/0	0.4/0.4/0.7/0.7/2	0.5/0.5/0.8/0.7/6	0.8/0.8/0.8/0.8/11	0.9/0.9/0.9/0.9/16	2
[4, 6]	0.2/0.3/0.7/0.5/0	0.5/0.5/0.7/0.7/4	0.7/0.7/0.7/0.7/9	0.8/0.8/0.7/0.7/12	0.9/0.9/0.7/0.9/15	7
[5, 6]	0.2/0.2/0.6/0.6/0	0.5/0.5/0.7/0.7/5	0.7/0.8/0.8/0.8/10	0.8/0.8/0.9/0.8/13	0.9/0.9/0.8/0.9/17	7

TAB. 6.2 – Results obtained, by solving program $P_{6.5}$, for the example described in figure 6.9 and table 6.1.

B	Expected total distance travelled (km)	Actual cost	$(en_{ij}+en_{ji})$			CPU time (s)
			Min	Av	Max	
0	60	0	0.86	1.06	1.37	0.0
30	147	29	0.74	2.03	5.54	0.1
60	181	60	0.70	2.65	5.18	0.2
90	203	87	0.64	3.11	4.81	0.2
120	211	117	0.79	3.63	4.29	0.1
150	213	133	3.21	4.94	4.43	0.1

TAB. 6.3 – Detailed results corresponding to the optimal solution of the example described in figure 6.9 and table 6.1 for a budget of 90 units.

	1	2	3	4	5	6
1		11	17			
2			0	0		
3				19	12	
4					0	15
5						13
6						

(a) c_{ijh}, amount of investment in each corridor - a total of 87 units.

	1	2	3	4	5	6
1		0.8/0.9	0.9/0.8			
2	0.8/0.9		0.2/0.5	0.2/0.6		
3	0.8/0.9	0.2/0.5		0.9/0.8	0.7/0.9	
4		0.2/0.7	0.9/0.9		0.2/0.7	0.9/0.7
5			0.7/0.8	0.2/0.7		0.8/0.9
6				0.9/0.9	0.8/0.8	

(b) r^1_{ijh} / r^2_{ijh}, probabilities associated with barrier effect, r^1_{ijh}, and mortality risk, r^2_{ijh}.

	1	2	3	4	5	6	7
1	0.15	0.36	0.36	0	0	0	0.13
2	0.24	0.60	0.03	0.04	0	0	0.09
3	0.18	0.02	0.35	0.18	0.16	0	0.11
4	0	0.03	0.20	0.45	0.03	0.16	0.12
5	0	0	0.19	0.05	0.43	0.24	0.09
6	0	0	0	0.40	0.32	0.15	0.12
7	0	0	0	0	0	0	1

(c) pr_{ij}, transition probabilities.

	1	2	3	4	5	6
1		4.81	4.52			
2			0.64	0.81		
3				4.49	3.58	
4					0.83	4.21
5						4.08
6						

(d) $(en_{ij}+en_{ji})$, expected number of routes along each corridor; average value=3.11.

	1	2	3	4	5	6
1	2.24	2.25	2.01	1.26	0.91	0.49
2	1.49	4.05	1.64	1.23	0.79	0.45
3	1.01	1.26	3.10	1.75	1.37	0.71
4	0.66	1.00	1.81	3.30	1.25	0.96
5	0.62	0.83	1.83	1.80	3.06	1.20
6	0.55	0.79	1.55	2.25	1.75	2.09

(e) n_{ij}, general term of $(I\text{-}Z)^{-1}$, expected number of passages through BR_j for an individual starting from BR_i - before its disappearance.

TAB. 6.4 – Detailed results corresponding to the optimal solution of the example described in figure 6.9 and table 6.1, in the case of a budget of 90 units and when the number of routes along each corridor must be greater than or equal to 1.5.

	1	2	3	4	5	6
1		8	17			
2			2	13		
3				15	12	
4					6	4
5						13
6						

(a) c_{ijh}, amount of investment in each corridor - a total of 90 units.

	1	2	3	4	5	6
1		0.7/0.9	0.9/0.8			
2	0.8/0.9		0.5/0.6	0.8/0.9		
3	0.8/0.9	0.6/0.6		0.8/0.8	0.7/0.9	
4		0.8/0.8	0.8/0.8		0.5/0.8	0.5/0.7
5			0.7/0.8	0.5/0.7		0.8/0.9
6				0.5/0.7	0.8/0.8	

(b) r^1_{ijh} / r^2_{ijh}, probabilities associated with barrier effect, r^1_{ijh}, and mortality risk, r^2_{ijh}.

	1	2	3	4	5	6	7
1	0.20	0.32	0.36	0	0	0	0.12
2	0.24	0.30	0.10	0.24	0	0	0.12
3	0.18	0.09	0.28	0.16	0.16	0	0.14
4	0	0.16	0.16	0.35	0.10	0.09	0.14
5	0	0	0.19	0.12	0.33	0.24	0.12
6	0	0	0	0.18	0.32	0.35	0.16
7	0	0	0	0	0	0	1

(c) pr_{ij}, transition probabilities.

	1	2	3	4	5	6
1		3.46	3.63			
2			1.55	3.24		
3				2.97	2.90	
4					1.85	1.82
5						3.69
6						

(d) $(\mathrm{en}_{ij} + \mathrm{en}_{ji})$, expected number of routes along each corridor; average value=2.79.

	1	2	3	4	5	6
1	2.05	1.42	1.69	1.21	0.80	0.46
2	1.04	2.42	1.39	1.51	0.79	0.50
3	0.88	1.03	2.54	1.36	1.08	0.58
4	0.60	1.04	1.30	2.63	1.06	0.75
5	0.50	0.70	1.29	1.33	2.54	1.12
6	0.41	0.62	0.99	1.36	1.54	2.29

(e) n_{ij}, general term of $(I\text{-}Z)^{-1}$, expected number of passages through BR_j for an individual starting from BR_i - before its disappearance.

worthwhile to use all the financial resources to optimize the permeability of the network. For example, when $B = 90$, the best solution – 203 km – is obtained by investing only 87 units. If the entire budget is required to be used by transforming in program $\mathrm{P}_{6.5}$ the inequality constraint $\sum_{(i,j)\in C, i<j, h\in H_{ij}} c_{ijh}x_{ijh} \leq B$ into the equality constraint $\sum_{(i,j)\in C, i<j, h\in H_{ij}} c_{ijh}x_{ijh} = B$, the best solution obtained corresponds to an expected distance of only 201 km. Table 6.3 gives detailed results when $B = 90$.

We see in table 6.3d that the corridors $[BR_2, BR_3]$, $[BR_2, BR_4]$, and $[BR_4, BR_5]$ are little used, compared to others. The mathematical programming approach allows additional constraints to be easily taken into account. For example, the expected number of routes along each corridor can be required to be greater than or equal to 1.5. To do this, simply add the constraints $\mathrm{en}_{ij} + \mathrm{en}_{ji} \geq 1.5$, $(i, j) \in C$, $i<j$, to program $\mathrm{P}_{6.5}$. In this case, the expected total distance travelled along the corridors becomes 165 km instead of 203 km. The detailed characteristics of this solution are given in table 6.4.

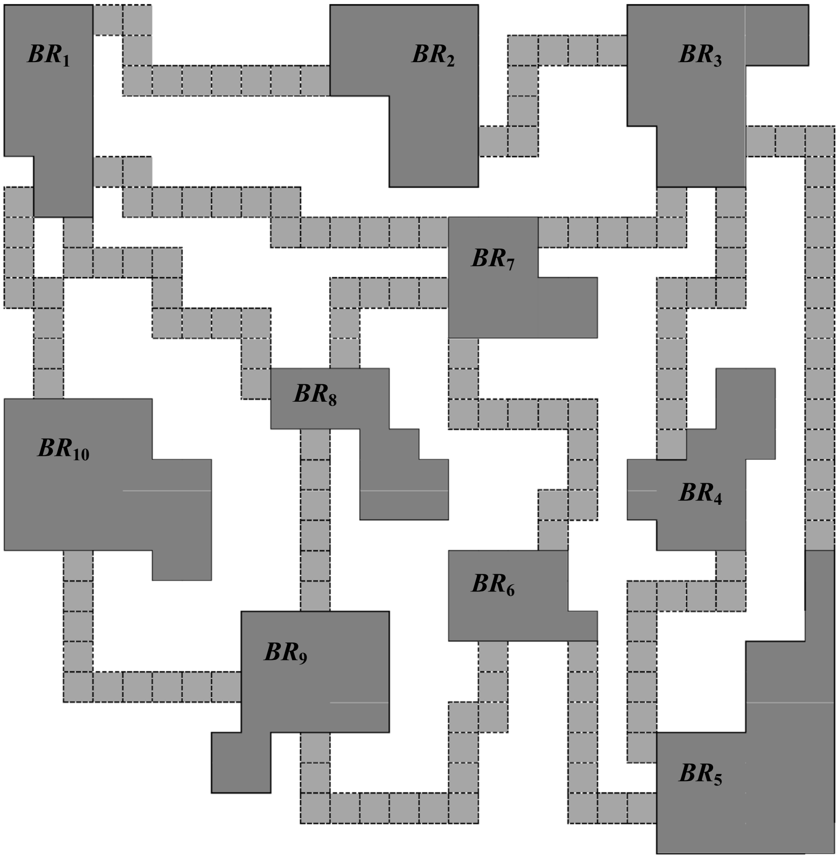


FIG. 6.10 – A hypothetical network with 10 reservoirs and 15 corridors.

6.3.5 *Example 2*

From a theoretical point of view, there is no limit in the size of the instances – number of reservoirs and number of corridors – that can be handled by program $P_{6.5}$. However, for large-sized instances, the computation time required to resolve them can become very important. We tested an instance with 10 reservoirs and 15 corridors and considered that there could be 7 levels of restoration for each of these corridors (figure 6.10). In this example, the landscape is represented by a grid of 28×28 square and identical zones, each side of which measures 500 m, and includes 10 biodiversity reservoirs, BR_1, BR_2,…, BR_{10}. As in the example in section 6.3.4, it is assumed that the effects of corridor restoration are not symmetric – r^1_{ijh} may be different from r^1_{jih} and r^2_{ijh} may be different from r^2_{jih}.

In figure 6.10, the length of the corridor connecting two reservoirs, BR_i and BR_j, is proportional to the number of grid cells that must be traversed along this corridor

TAB. 6.5 – This table shows, for each corridor $[BR_i, BR_j]$ of the network in figure 6.10 and for each possible value of the investment level $h \in \{0, 1, 2, 3, 4, 5, 6, 7\}$, the probabilities r^1_{ijh}, r^1_{jih}, r^2_{ijh}, r^2_{jih} and the associated costs, c_{ijh}, in this order. For example, the cell located at the intersection of row [3, 5] – associated with the corridor $[BR_3, BR_5]$ – and column $h = 6$, $r^1_{356} = 0.81$, $r^1_{536} = 0.81$, $r^2_{356} = 0.86$, $r^2_{536} = 0.86$, and $c_{356} = 31$. The last column of the table shows the length of each corridor, l_{ij}, in kilometres – 16 for the corridor $[BR_3, BR_5]$.

$[i, j]$	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 4$
[1, 2]	0.21/0.23/0.56/0.58/0	0.31/0.33/0.61/0.63/1	0.41/0.43/0.66/0.68/4	0.51/0.53/0.71/0.73/9	0.61/0.63/0.76/0.78/15
[1, 7]	0.22/0.22/0.57/0.57/0	0.32/0.32/0.62/0.62/1	0.42/0.42/0.67/0.67/4	0.52/0.52/0.72/0.72/9	0.62/0.62/0.77/0.77/15
[1, 8]	0.21/0.22/0.56/0.57/0	0.31/0.32/0.61/0.62/1	0.41/0.42/0.66/0.67/5	0.51/0.52/0.71/0.72/11	0.61/0.62/0.76/0.77/18
[1, 10]	0.22/0.23/0.57/0.58/0	0.32/0.33/0.62/0.63/1	0.42/0.43/0.67/0.68/5	0.52/0.53/0.72/0.73/11	0.62/0.63/0.77/0.78/18
[2, 3]	0.24/0.21/0.59/0.56/0	0.34/0.31/0.64/0.61/1	0.44/0.41/0.69/0.66/5	0.54/0.51/0.74/0.71/10	0.64/0.61/0.79/0.76/17
[3, 4]	0.22/0.23/0.57/0.58/0	0.32/0.33/0.62/0.63/1	0.42/0.43/0.67/0.68/5	0.52/0.53/0.72/0.73/10	0.62/0.63/0.77/0.78/17
[3, 5]	0.21/0.21/0.56/0.56/0	0.31/0.31/0.61/0.61/1	0.41/0.41/0.66/0.66/4	0.51/0.51/0.71/0.71/9	0.61/0.61/0.76/0.76/15
[3, 7]	0.21/0.21/0.56/0.56/0	0.31/0.31/0.61/0.61/1	0.41/0.41/0.66/0.66/4	0.51/0.51/0.71/0.71/8	0.61/0.61/0.76/0.76/13
[4, 5]	0.21/0.21/0.56/0.56/0	0.31/0.31/0.61/0.61/1	0.41/0.41/0.66/0.66/5	0.51/0.51/0.71/0.71/10	0.61/0.61/0.76/0.76/17
[5, 6]	0.22/0.23/0.57/0.58/0	0.32/0.33/0.62/0.63/1	0.42/0.43/0.67/0.68/4	0.52/0.53/0.72/0.73/9	0.62/0.63/0.77/0.78/15
[6, 7]	0.22/0.23/0.57/0.58/0	0.32/0.33/0.62/0.63/1	0.42/0.43/0.67/0.68/5	0.52/0.53/0.72/0.73/11	0.62/0.63/0.77/0.78/18
[6, 9]	0.24/0.22/0.59/0.57/0	0.34/0.32/0.64/0.62/1	0.44/0.42/0.69/0.67/4	0.54/0.52/0.74/0.72/9	0.64/0.62/0.79/0.77/15
[7, 8]	0.23/0.22/0.58/0.57/0	0.33/0.32/0.63/0.62/1	0.43/0.42/0.68/0.67/5	0.53/0.52/0.73/0.72/10	0.63/0.62/0.78/0.77/17
[8, 9]	0.21/0.23/0.56/0.58/0	0.31/0.33/0.61/0.63/1	0.41/0.43/0.66/0.68/4	0.51/0.53/0.71/0.73/7	0.61/0.63/0.76/0.78/12
[9, 10]	0.22/0.23/0.57/0.58/0	0.32/0.33/0.62/0.63/1	0.42/0.43/0.67/0.68/4	0.52/0.53/0.72/0.73/8	0.62/0.63/0.77/0.78/13

$[i, j]$	$h = 5$	$h = 6$	$h = 7$	l_{ij}
[1, 2]	0.71/0.73/0.81/0.83/23	0.81/0.83/0.86/0.88/32	0.91/0.93/0.91/0.93/42	10
[1, 7]	0.72/0.72/0.82/0.82/22	0.82/0.82/0.87/0.87/30	0.92/0.92/0.92/0.92/40	14
[1, 8]	0.71/0.72/0.81/0.82/27	0.81/0.82/0.86/0.87/38	0.91/0.92/0.91/0.92/50	12
[1, 10]	0.72/0.73/0.82/0.83/27	0.82/0.83/0.87/0.88/38	0.92/0.93/0.92/0.93/50	8
[2, 3]	0.74/0.71/0.84/0.81/25	0.84/0.81/0.89/0.86/35	0.94/0.91/0.94/0.91/47	8
[3, 4]	0.72/0.73/0.82/0.83/26	0.82/0.83/0.87/0.88/35	0.92/0.93/0.92/0.93/47	11
[3, 5]	0.71/0.71/0.81/0.81/22	0.81/0.81/0.86/0.86/31	0.91/0.91/0.91/0.91/41	16
[3, 7]	0.71/0.71/0.81/0.81/20	0.81/0.81/0.86/0.86/28	0.91/0.91/0.91/0.91/37	6
[4, 5]	0.71/0.71/0.81/0.81/26	0.81/0.81/0.86/0.86/36	0.91/0.91/0.91/0.91/47	10
[5, 6]	0.72/0.73/0.82/0.83/23	0.82/0.83/0.87/0.88/32	0.92/0.93/0.92/0.93/42	8
[6, 7]	0.72/0.73/0.82/0.83/27	0.82/0.83/0.87/0.88/37	0.92/0.93/0.92/0.93/49	12
[6, 9]	0.74/0.72/0.84/0.82/22	0.84/0.82/0.89/0.87/31	0.94/0.92/0.94/0.92/41	14
[7, 8]	0.73/0.72/0.83/0.82/26	0.83/0.82/0.88/0.87/36	0.93/0.92/0.93/0.92/47	6
[8, 9]	0.71/0.73/0.81/0.83/19	0.81/0.83/0.86/0.88/26	0.91/0.93/0.91/0.93/34	6
[9, 10]	0.72/0.73/0.82/0.83/19	0.82/0.83/0.87/0.88/27	0.92/0.93/0.92/0.93/36	10

TAB. 6.6 – Computational results for the example described in figure 6.10 and table 6.5.

B	Expected total distance travelled (km)	Actual cost	$(\text{en}_{ij} + \text{en}_{ji})$ Min	Av	Max	CPU time (s)
0	133	0	0.77	0.87	1.00	0
150	443	147	0.84	2.39	6.62	9
300	688	296	0.86	3.95	7.17	20
450	917	445	0.96	5.49	7.58	2
600	1,065	576	1.00	6.87	7.53	0.1
750	1,111	620	7.16	7.34	7.62	0.03

to get from BR_i to BR_j. For example, the length of the corridor connecting BR_1 to BR_2 is equal to 5 km and the length of the corridor connecting BR_6 to BR_7 is equal to 6 km. All data for this example are summarized in table 6.5.

We see in table 6.6 that the resolution of this instance is very fast for the six B values considered. The case that requires the most computation time is when the financial resources are limited to about 0.5 times the maximum potential investment, $\sum_{(i,j) \in C, i<j} c_{ij}^7 = 650$. We also see in this table that the difference between the extreme values of $(\text{en}_{ij} + \text{en}_{ji})$ is often significant. We solved the problem with $B = 450$ and the additional constraints, $\text{en}_{ij} + \text{en}_{ji} \geq 2$, $(i, j) \in C, i<j$. In this case, the minimal number of routes along a corridor is equal to 2.33 and the maximal number of routes along a corridor is equal to 6.15, but the expected total distance travelled along the corridors is only equal to 765 km. It should be noted that taking this constraint into account significantly increases the computation time, since it increases from 2 to 49 s. We also solved the problem with the constraints $\text{pr}_{ij} \geq 0.1$ – without constraints on the number of routes in each corridor. In this case, the minimal number of routes along a corridor is equal to 2.32, the maximal number of routes along a corridor is equal to 6.09 and the expected total distance travelled along the corridors is equal to 744 km (39 s of computation time). With the constraints $\text{pr}_{ij} \geq 0.15$, $(i, j) \in C$, the minimal number of routes along a corridor is equal to 2.86, the maximal number of routes along a corridor is equal to 5.47 and the expected total distance travelled along the corridors is equal to 676 km (1.12 s of computation time). Finally, there is no feasible solution when $\text{pr}_{ij} \geq 0.2$, $(i, j) \in C$. In this case, with the budgetary constraint corresponding to $B = 450$, it is impossible to make investments in the corridors in such a way that $\text{pr}_{ij} \geq 0.2$ for all the reservoir pairs connected by a corridor. Remember that pr_{ij} is the probability, for an animal leaving reservoir BR_i, of reaching the adjacent reservoir BR_j in one transition. The same applies to any available budget value less than or equal to 569. On the other hand, for any available budget value greater than or equal to 570, there is a feasible solution. For example, for the maximal potential investment – $B = 650$ – the minimal number of routes along a corridor is equal to 7.16, the maximal number of routes along a corridor is equal to 7.62, and the expected total distance travelled along the corridors is equal to 1,111 km (0.02 s of computation time).

References and Further Reading

Álvarez-Miranda E., Ljubić I., Mutzel P. (2013) The rooted maximum node-weight connected subgraph problem, *Integration of AI and OR techniques in constraint programming for combinatorial optimization problems* (C. Gomes, M. Sellmann, Eds). CPAIOR 2013. Lecture notes in computer science, Springer, vol. 7874.

Anderson A.B., Jenkins C.N. (2006) *Applying nature's design: corridors as a strategy for biodiversity conservation*. Columbia University Press.

Beale E.M.L. (1988) *Introduction to optimization*. Wiley.

Beier P., Noss R. (1998) Do habitat corridors provide connectivity? *Conserv. Biol.* **12**, 1241.

Bell S., Apostol D. (2008) *Designing sustainable forest landscapes*. Taylor & Francis, London.

Bergès L., Roche P., Avon C. (2010) Corridors écologiques et conservation de la biodiversité, intérêts et limites pour la mise en place de la Trame verte et bleue, *Sci. Eaux Territoires* **3**, 34.

Billionnet A. (2010) Integer programming for optimizing habitat network permeability, *Manage. Environ. Qual.* **21**, 570.

Billionnet A. (2013) Mathematical optimization ideas for biodiversity conservation, *Eur. J. Oper. Res.* **231**, 514.

Bond M. (2003) *Principles of wildlife corridor design*. Report Center for Biological Diversity, Idyllwild, California.

Carrière S.M., Hervé D., Andriamahefazafy F., Méral P. (2011) Les corridors, passage obligé ? L'exemple malgache (C. Aubertin, E. Rodary, Eds). *Aires protégées, espaces durables ?* IRD Editions.

Challéat S. (2017) *Empreinte lumineuse, trame nocturne et corridors noirs: quelques precisions*. RENOIR research group, Interdisciplinary Night-time Environment Observatory.

Clergeau P., Désiré G. (1999) Biodiversité, paysage et aménagement: du corridor à la zone de connexion biologique, *Mappemonde* **55**, 19.

Closset-Kopp D., Wasof S., Decocq G. (2016) Using process-based indicator species to evaluate ecological corridors in fragmented landscapes, *Biol. Conserv.* **201**, 152.

Conrad J., Gomes C., van Hoeve W.J., Sabharwal A., Suter J. (2012) Wildlife corridors as a connected subgraph problem, *J. Environ. Econ. Manage.* **63**, 1.

Conservation corridor, connectivity conservation specialist group, available at: https://conservationcorridor.org/.

Conservation corridor, Connecting Science to Conservation, Programs and tools. Available at: http://conservationcorridor.org/corridor-toolbox/programs-and-tools.

Cushman S.A., McRae B., Adriaensen F., Beier P., Shirley M., Zeller K. (2013) Biological corridors and connectivity, *Key topics in conservation biology 2* (D.W. Macdonald, K.J. Willis, Eds). Wiley-Blackwell, Hoboken, NJ, pp. 384-404.

Dilkina B., Gomes C.P. (2010) Solving connected subgraph problems in wildlife conservation. CPAIOR-10: International Conference on Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems, Bologna, Italy.

Downes S.J., Handasyde K.A., Elgar M.A. (1997) The use of corridors by mammals in fragmented Australian eucalypt forests, *Conserv. Biol.* **11**, 718.

Finke J., Sonnenschein M. (2008) Simulation and optimization of habitat network permeability, *Inf. Technol. Environ. Eng.* **1**, 12.

Finke J., Strein M., Sonnenschein M. (2008) A simulation framework for modeling anthropogenic disturbances in habitat networks, *Ecol. Inf.* **3**, 26.

Grinstead C.M., Snell J.L. (1997) *Introduction to probability*. American Mathematical Society (2nd edition 2006), Providence, USA.

Gustafsson L., Hansson L. (1997) Corridors as a conservation tool, *Ecol. Bull.* **46**, 182.

Harris L.D., Scheck J. (1991) From implications to applications: the dispersal corridor principle applied to the conservation of biological diversity, *The role of corridors* (D.A. Saunders, R.J. Hobbs, Eds). Surrey Beatty, Chipping Norton, pp. 189-200.

Harrison R. (1992) Toward a theory of inter-refuge corridor design, *Conserv. Biol.* **6**, 293.

Hilty J.A., Lidicker W.Z. Jr., Merenlender A.M. (2006) *Corridor ecology, the science and practice of linking landscapes for biodiversity conservation*. Island Press.

Hobbs R.J. (1992) The role of corridors in conservation: solution or bandwagon? *Trends Ecol. Evol.* **7**, 389.

Holzgang O., Pfister H.P., Heynen D., Blant M., Righetti A., Berthoud G., Marchesi P., Maddalena T., Muri H., Wendelspiess M., Dandliker G., Mollet P., Bornhauser-Sieber U. (2001) Les corridors faunistiques en Suisse. Cahier de l'environnement n° 326, Office fédéral de l'environnement, des forêts et du paysage (OFEFP), Société suisse de Biologie de la Faune (SSBF) & Station ornithologique suisse de Sempach, Bern, p. 120.

Jordan F. (2000) A reliability-theory approach to corridor design, *Ecol. Model.* **28**, 211.

Kemeny J.G., Snell J.L. (1983) *Finite Markov chains*. Springer.

Margules C.R., Nicholls A.O., Pressey R.L. (1988) Selecting networks of reserves to maximize biological diversity, *Biol. Conserv.* **43**, 63.

McEuen A. (1993) The wildlife corridor controversy: a review, *Endanger. Species Update* **10**, 1.

Menard P., Clergeau P. (2001) La notion de zone de connexion biologique, son application en aménagement du territoire, *Mappemonde* **64**, 24.

Ministère de la Transition écologique et solidaire, available at: https://www.ecologique-solidaire.gouv.fr/trame-verte-et-bleue.

Naidoo R., Kilian J.W., Du Preez P., Beytell P., Aschenborn O., Taylor R.D., Stuart-Hill G. (2018) Evaluating the effectiveness of local- and regional-scale wildlife corridors using quantitative metrics of functional connectivity, *Biol. Conserv.* **217**, 96.

Naro-Maciel E., Sterling E.J., Rao M. (2008) Protected areas and biodiversity conservation I: reserve planning and design, *Lessons Conserv.* **2**, 19.

Natural England. (2015) Green Bridges, a literature review. Natural England Commissioned Report (NECR181).

Noss R.F. (1987) Corridors in real landscapes: a reply to Simberloff and Cox, *Conserv. Biol.* **1**, 159.

Önal H., Briers R.A. (2006) Optimal selection of a connected reserve network, *Oper. Res.* **54**, 379.

Rosenberg D.K., Noon B.R., Meslow E.C. (1995) Towards a definition of biological corridor, *Integrating people and wildlife for a sustainable future* (J.A. Bissonette, P.R. Krausman, Eds). International Wildlife Management Congress, Bethesda, Maryland, pp. 436-439.

Salkin H.M., Mathur K. (1989) *Foundations of integer programming*. North-Holland.

Saunders D.A., Hobbs R.J., Eds (1991) *Nature conservation 2: the role of corridors*. Surrey Beatty & Sons.

Sessions J. (1992) Solving for network connections as a Steiner network problem, *For. Sci.* **38**, 203.

Simberloff D., Cox J. (1987) Consequences and costs of conservation corridors, *Conserv. Biol.* **1**, 63.

Simberloff D., Farr J.A., Cox J., Mehlman D.W. (1992) Movement corridors: conservation bargains or poor investments? *Conserv. Biol.* **6**, 493.

St John R. (2016) Explicitly controlling geometric characteristics of corridors in spatial optimization. PhD, University of Washington.

Suter J.F., Conrad J., Gomes C.P., van Hoeve W.J., Sabharwal A. (2008) Optimal corridor design for grizzly bear in the U.S. Northern Rockies. Paper presented at American Agricultural Economics Association Annual Meeting, July 27–29, Orlando, FL.

West S., Cairns R., Schultz L. (2016) What constitutes a successful biodiversity corridor? A Q-study in the Cape Floristic Region, South Africa. *Biol. Conserv.* **198**, 183.

Williams J.C. (1998) Delineating protected wildlife corridors with multi-objective programming, *Environ. Model. Assess.* **3**, 77.

Williams J.C., Snyder S.A. (2005) Restoring habitat corridors in fragmented landscape using optimization and percolation models, *Environ. Model. Assess.* **10**, 239.

Wolsey L.A. (1998) *Integer programming*. Wiley, New York, NY.

Chapter 7

Species Survival Probabilities

7.1 Introduction

As in the previous chapters, we are interested in a set of threatened species, $S = \{s_1, s_2,\ldots, s_m\}$, and a set of zones that we can decide whether or not to protect, $Z = \{z_1, z_2,\ldots, z_n\}$. It is hypothesized that protecting a zone increases the chances of survival, in that zone, of the species that live there and in which we are interested. It is also assumed that the effects generated by the protection of the zones are independent. In other words, the chances of survival of a species in one zone depend only on whether the zone is protected or not; they do not depend on decisions that are made with regard to other zones. Thus, the main characteristic of the different models presented in this chapter lies in the fact that the uncertainty – which has many sources – concerning the survival of species s_k in zone z_i is reflected by a certain probability, and this for all $i \in \underline{Z} = \{1, 2,..., n\}$ and for all $k \in \underline{S} = \{1, 2,..., m\}$. Note that these probabilities are generally difficult to establish since it is particularly difficult, in this field as in many others, to predict the future based on past events. First, it is assumed that the survival probability of species s_k in zone z_i is equal to p_{ik} if zone z_i is not protected and q_{ik} in the opposite case. Note that this model is very general since the values p_{ik} and q_{ik} can be equal to 0 for some couples (i, k). In particular, some protected zones do not contribute to the protection of certain species. This also enables, for example, to consider a zero survival probability for certain species in unprotected zones. In a second step, it is assumed, as before, that the protection of zone z_i ensures the survival of species s_k in this zone with the probability q_{ik} but it is realistically admitted that a certain error may affect this probability. More precisely, it is assumed that the survival probability of species s_k in the protected zone z_i belongs to the interval $[q_{ik} - \delta_{ik},\ q_{ik} + \gamma_{ik}]$, and this for all $i \in \underline{Z}$ and for all $k \in \underline{S}$. On the other hand, to simplify the presentation, it is considered, in this case, that there is no uncertainty about the survival probabilities of the species in unprotected zones and that all these probabilities are equal to 0.

DOI: 10.1051/978-2-7598-2521-9.c007

7.2 Reserve Ensuring a Certain Survival Probability for the Largest Possible Number of Species, of a Given Set, Under a Budgetary Constraint

The problem is to define a reserve, *i.e.*, a set of zones to be protected, whose protection cost is less than or equal to the available budget, denoted by B, and which maximizes the number of species of S whose survival probability in the set of zones considered – protected or not – is greater than or equal to a certain threshold value. We denote by ρ_k the threshold value corresponding to species s_k. As we saw in the introduction, the survival probability of species s_k in zone z_i is denoted by p_{ik} if z_i is not protected and q_{ik} in the opposite case, and it is assumed that these survival probabilities are independent. Let us introduce the Boolean decision variable x_i which takes the value 1 if and only if zone z_i is protected. The extinction probability of species s_k in zone z_i can then be written, as a function of variables x_i, $1 - p_{ik}(1 - x_i) - q_{ik}x_i$. It can be deduced that the probability of disappearance of species s_k from the set of zones considered is equal to $\prod_{i \in \underline{Z}} (1 - p_{ik}(1 - x_i) - q_{ik}x_i)$, and finally that the survival probability of species s_k in these same zones is equal to $1 - \prod_{i \in \underline{Z}} (1 - p_{ik}(1 - x_i) - q_{ik}x_i)$. The problem is, therefore, to determine the zones to be protected, *i.e.,* the values of variables x_i, in order to satisfy, for as many species s_k as possible, $k \in \underline{S}$, the constraint $1 - \prod_{i \in \underline{Z}} (1 - p_{ik}(1 - x_i) - q_{ik}x_i) \geq \rho_k$. Let us also introduce the Boolean variable y_k which takes the value 1 if and only if this last constraint is verified, *i.e.,* if the survival probability of species s_k in the set of candidate zones is greater than or equal to the threshold value, ρ_k. The problem considered can then be formulated as the mathematical program in Boolean variables $\mathrm{P}_{7.1}$.

$$
\mathrm{P}_{7.1}: \begin{cases} \max \sum_{k \in \underline{S}} y_k & \\ \text{s.t.} \left| \begin{array}{lll} \sum_{i \in \underline{Z}} c_i x_i \leq B & & (7.1.1) \\ 1 - \prod_{i \in \underline{Z}} (1 - p_{ik}(1 - x_i) - q_{ik}x_i) \geq \rho_k y_k & k \in \underline{S} & (7.1.2) \\ x_i \in \{0, 1\} & i \in \underline{Z} & (7.1.3) \\ y_k \in \{0, 1\} & k \in \underline{S} & (7.1.4) \end{array} \right. \end{cases}
$$

The economic function of $\mathrm{P}_{7.1}$ expresses the number of species whose survival probability in the set of candidate zones is greater than or equal to the threshold value. This function should be maximized. Constraint 7.1.1 expresses that the total cost of protecting the reserve must be less than or equal to the available budget, B. Constraints 7.1.2 force the Boolean variables y_k, $k \in \underline{S}$, to take the value 0 if the survival probability of species s_k, in the set of candidate zones, is below the threshold value, ρ_k. Otherwise, and because of the expression of the economic function to be maximized, variable y_k takes the value 1 at the optimum of $\mathrm{P}_{7.1}$. Constraints 7.1.3 and 7.1.4 specify the Boolean nature of variables x_i and y_k. The economic function is linear but constraints 7.1.2 are not linear since they involve the products of the n linear functions $1 - p_{ik}(1 - x_i) - q_{ik}x_i$. We will see that these constraints 7.1.2 can be linearized and therefore, finally, the solution to the problem considered can be determined by solving a linear program in Boolean variables. First of all, let us rewrite constraints 7.1.2 as $\prod_{i\in\underline{Z}}(1 - p_{ik}(1 - x_i) - q_{ik}x_i) \leq 1 - \rho_k y_k,\ \ k \in \underline{S}$. To simplify the presentation, it is assumed that p_{ik}, q_{ik}, and ρ_k are strictly less than 1 (a method to take into account probabilities that can take the value 1 is presented in section 7.5.1). Constraints 7.1.2 are equivalent to $\log\Big(\prod_{i\in\underline{Z}}(1 - p_{ik}(1 - x_i) - q_{ik}x_i)\Big) \leq \log(1 - \rho_k y_k)$ or alternatively to $\sum_{i\in\underline{Z}} \log(1 - p_{ik}(1 - x_i) - q_{ik}x_i) \leq \log(1 - \rho_k y_k),\ k \in \underline{S}$. Since x_i and y_k are Boolean variables, $\log(1 - p_{ik}(1 - x_i) - q_{ik}x_i) = x_i \log(1 - q_{ik}) + (1 - x_i)\log(1 - p_{ik})$ and $\log(1 - \rho_k y_k) = y_k \log(1 - \rho_k)$. The non-linear constraints 7.1.2 are, therefore, equivalent to the linear constraint $\sum_{i\in\underline{Z}}[x_i \log(1 - q_{ik}) + (1 - x_i)\log(1 - p_{ik})] \leq y_k \log(1 - \rho_k),\ \ k \in \underline{S}$. Finally, the solution to the problem considered can be determined by solving the linear program in Boolean variables $\mathrm{P}_{7.2}$:

$$
\mathrm{P}_{7.2}:\left\{\begin{array}{ll}
\max \displaystyle\sum_{k\in\underline{S}} y_k & \\
\text{s.t.} & \left|\begin{array}{lll}
\displaystyle\sum_{i\in\underline{Z}} c_i x_i \leq B & & (7.2.1)\\
\displaystyle\sum_{i\in\underline{Z}}[x_i \log(1 - q_{ik}) + (1 - x_i)\log(1 - p_{ik})] & & \\
\qquad \leq y_k \log(1 - \rho_k) & k \in \underline{S} & (7.2.2)\\
x_i \in \{0, 1\} & i \in \underline{Z} & (7.2.3)\\
y_k \in \{0, 1\} & k \in \underline{S} & (7.2.4)
\end{array}\right.
\end{array}\right.
$$

By setting $\alpha^1_{ik} = \log(1 - q_{ik})$, $\alpha^2_{ik} = \log(1 - p_{ik})$ and $\beta_k = \log(1 - \rho_k)$, program $\mathrm{P}_{7.2}$ is rewritten as program $\mathrm{P}_{7.3}$.

$$
\mathrm{P}_{7.3}:\left\{\begin{array}{ll}
\max \displaystyle\sum_{k\in\underline{S}} y_k & \\
\text{s.t.} & \left|\begin{array}{lll|lll}
\displaystyle\sum_{i\in\underline{Z}} c_i x_i \leq B & & (7.3.1) & x_i \in \{0, 1\} & i \in \underline{Z} & (7.3.3)\\
\displaystyle\sum_{i\in\underline{Z}}[\alpha^1_{ik} x_i + \alpha^2_{ik}(1 - x_i)] \leq \beta_k y_k & k \in \underline{S} & (7.3.2) & y_k \in \{0, 1\} & k \in \underline{S} & (7.3.4)
\end{array}\right.
\end{array}\right.
$$

7.3 Least-Cost Reserve Ensuring a Certain Survival Probability for All Species Under Consideration

Consider the following variant of the problem studied in the previous section, which consists in determining a reserve, *i.e.,* a set of zones to be protected, with a minimal cost, and which ensures that all species of S have a survival probability – in the set of candidate zones – greater than or equal to a certain threshold value. The solution to this problem is obtained by solving the linear program in Boolean variables $\mathrm{P}_{7.4}$.

$$\mathrm{P}_{7.4}: \begin{cases} \min \sum\limits_{i \in \underline{Z}} c_i x_i \\ \text{s.t.} \left| \begin{array}{lll} \sum\limits_{i \in \underline{Z}} \left[\alpha^1_{ik} x_i + \alpha^2_{ik}(1 - x_i)\right] \le \beta_k & k \in \underline{S} & (7.4.1) \\ x_i \in \{0,1\} & i \in \underline{Z} & (7.4.2) \end{array} \right. \end{cases}$$

The economic function of $\mathrm{P}_{7.4}$ expresses the cost of protecting the reserve. This function should be minimized. Constraints 7.4.1 express that the survival probability of species s_k, in the set of candidate zones, must be greater than or equal to the threshold value, ρ_k, associated with this species, and this for all $k \in \underline{S}$.

7.4 Study of the Two Previous Problems When the Survival Probabilities of the Species Considered are All Equal to Zero in the Unprotected Zones

7.4.1 Mathematical Programming Formulation

It is assumed here that all the unprotected zones will be assigned to activities incompatible with the protection of the species that live there. This corresponds to the particular cases of the problems studied in the 2 previous sections, obtained by considering that the survival probability of species s_k in zone z_i is equal to 0 if zone z_i is not protected and to q_{ik} in the opposite case. The survival probability of species s_k in the set of candidate zones is then equal to the survival probability of species s_k in the reserve, *i.e.*, expressed as a function of variables x_i, to $1 - \prod_{i \in \underline{Z}} (1 - q_{ik} x_i)$. The particular case corresponding to the problem in section 7.2 is to determine a reserve, whose protection cost is less than or equal to the available budget and which maximises the number of species of S whose survival probability in the reserve is greater than or equal to a certain threshold value. We obtain the mathematical program in Boolean variables $\mathrm{P}_{7.5}$. This program is obtained by replacing in $\mathrm{P}_{7.1}$ constraints 7.1.2 by constraints 7.5.2.

$$\mathrm{P}_{7.5}: \begin{cases} \max \sum\limits_{k \in \underline{S}} y_k \\ \text{s.t.} \left| \begin{array}{lllllll} \sum\limits_{i \in \underline{Z}} c_i x_i \le B & & (7.5.1) & | & x_i \in \{0,1\} & i \in \underline{Z} & (7.5.3) \\ 1 - \prod\limits_{i \in \underline{Z}} (1 - q_{ik} x_i) \ge \rho_k y_k & k \in \underline{S} & (7.5.2) & | & y_k \in \{0,1\} & k \in \underline{S} & (7.5.4) \end{array} \right. \end{cases}$$

Like constraints 7.1.2, constraints 7.5.2 can be linearized and the solution to the problem considered can, therefore, be determined by solving the linear program in Boolean variables $\mathrm{P}_{7.6}$.

$$
\mathrm{P}_{7.6}:\begin{cases}\max \sum_{k\in\underline{S}} y_k & \\ \text{s.t.}\left|\begin{array}{llll} \sum_{i\in\underline{Z}} c_i x_i \le B & & (7.6.1) & \mid\ x_i\in\{0,1\} \quad i\in\underline{Z} \quad (7.6.3)\\ \sum_{i\in\underline{Z}} x_i \log(1-q_{ik}) \le y_k \log(1-\rho_k) & k\in\underline{S} & (7.6.2) & \mid\ y_k\in\{0,1\} \quad k\in\underline{S} \quad (7.6.4)\end{array}\right.\end{cases}
$$

Finally, by putting $\alpha^1_{ik} = \log(1-q_{ik})$ and $\beta_k = \log(1-\rho_k)$, program $\mathrm{P}_{7.6}$ is rewritten as program $\mathrm{P}_{7.7}$.

$$
\mathrm{P}_{7.7}:\begin{cases}\max \sum_{k\in\underline{S}} y_k & \\ \text{s.t.}\left|\begin{array}{llll} \sum_{i\in\underline{Z}} c_i x_i \le B & & (7.7.1) & \mid\ x_i\in\{0,1\} \quad i\in\underline{Z} \quad (7.7.3)\\ \sum_{i\in\underline{Z}} \alpha^1_{ik} x_i \le \beta_k y_k & k\in\underline{S} & (7.7.2) & \mid\ y_k\in\{0,1\} \quad k\in\underline{S} \quad (7.7.4)\end{array}\right.\end{cases}
$$

The problem of section 7.3, in the particular case where all the survival probabilities are zero in unprotected zones, is to determine a minimal cost reserve that ensures that all the species of S have a survival probability – in the set of candidate zones and, therefore, in the reserve – greater than or equal to a certain threshold value. This problem can be solved by the mathematical program in Boolean variables $\mathrm{P}_{7.8}$ obtained by replacing in program $\mathrm{P}_{7.4}$ constraints 7.4.1 by the constraints $\sum_{i\in\underline{Z}} \alpha^1_{ik} x_i \le \beta_k$, $k\in\underline{S}$.

$$
\mathrm{P}_{7.8}:\begin{cases}\min \sum_{i\in\underline{Z}} c_i x_i & \\ \text{s.t.}\left|\begin{array}{lll} \sum_{i\in\underline{Z}} \alpha^1_{ik} x_i \le \beta_k & k\in\underline{S} & (7.8.1)\\ x_i\in\{0,1\} & i\in\underline{Z} & (7.8.2)\end{array}\right.\end{cases}
$$

Remark. The problems considered in section 7.4 can be interpreted in a slightly different way as some authors have done: the presence of a given species in a given zone is defined by a probability. Thus q_{ik} refers to the probability of occurrence of species s_k in zone z_i. These probabilities can be determined using statistical methods such as logistic regression. The two problems considered above then become: (1) determine a reserve, whose cost of protection is less than or equal to the available budget and which maximizes the number of species of S whose probability of occurrence in the reserve is greater than or equal to a certain threshold value, (2) determine a minimal cost reserve which ensures to all the species of S a probability of occurrence in the reserve greater than or equal to a certain threshold value.

7.4.2 Examples

Let us illustrate the previous results on a hypothetical set of candidate zones represented by a grid of 8×8 square and identical zones. As already noted, the set of

	1	2	3	4	5	6	7	8
1	$s_1 : 0.7$ $s_8 : 0.7$ $s_{10} : 0.3$ 9	$s_3 : 0.5$ 2	3	$s_5 : 0.4$ $s_8 : 0.6$ 3	$s_6 : 0.6$ $s_{10} : 0.7$ 3	$s_1 : 0.6$ 3	$s_2 : 0.4$ $s_4 : 0.4$ 9	$s_2 : 0.9$ $s_7 : 0.6$ $s_{10} : 0.8$ 5
2	4	$s_9 : 0.8$ 1	$s_9 : 0.5$ 1	$s_5 : 0.4$ 9	$s_1 : 0.6$ 5	$s_3 : 0.6$ $s_7 : 0.8$ 6	3	4
3	7	8	10	$s_3 : 0.6$ $s_{10} : 0.5$ 6	4	$s_9 : 0.9$ 8	$s_7 : 0.5$ 10	4
4	3	$s_9 : 0.6$ 3	$s_2 : 0.6$ 8	$s_4 : 0.3$ 7	6	$s_9 : 0.7$ 5	9	9
5	1	10	2	$s_4 : 0.9$ 7	$s_3 : 0.8$ $s_7 : 0.5$ 3	$s_7 : 0.5$ $s_8 : 0.8$ 6	$s_5 : 0.4$ 1	6
6	$s_2 : 0.6$ 3	$s_4 : 0.8$ $s_9 : 0.7$ 4	6	2	$s_3 : 0.6$ 10	6	$s_4 : 0.7$ 1	$s_5 : 0.6$ 7
7	$s_2 : 0.4$ 8	2	$s_8 : 0.3$ 5	1	$s_3 : 0.8$ $s_4 : 0.4$ 3	$s_8 : 0.8$ 6	8	$s_6 : 0.5$ 8
8	5	10	$s_1 : 0.5$ $s_3 : 0.7$ 5	$s_2 : 0.5$ $s_4 : 0.5$ $s_6 : 0.4$ 5	$s_3 : 0.6$ 4	$s_2 : 0.7$ 7	$s_5 : 0.7$ 4	$s_6 : 0.3$ 7

FIG. 7.1 – A set of 64 candidate zones for protection represented by a grid of 8×8 square and identical zones. 10 species s_1, s_2,…, s_{10} are concerned. The corresponding survival probabilities, q_{ijk}, and the protection costs are indicated in each zone. Consider, for example, zone z_{56}. Species s_7 and s_8 are concerned. The survival probabilities of these 2 species in this zone, if protected, are 0.5 and 0.8, respectively. The cost of protecting this zone is equal to 6.

candidate zones is represented by a grid in order to simplify the presentation, but all the following could easily be adapted to other sets of candidate zones. In this example, 10 species are concerned. The data are presented in figure 7.1. The zones are designated by z_{ij} where i represents the row index of the zone and j, its column index. On each zone is indicated the list of the species whose survival probability is positive if the zone is protected and the corresponding survival probability, denoted by q_{ijk} for species s_k in zone z_{ij}. This probability corresponds to probability q_{ik} defined at the beginning of this chapter but, here, a candidate zone is defined by the index pair (i, j). In this example, all the survival probabilities in the unprotected zones are zero.

TAB. 7.1 – List of the zones whose protection guarantees for species s_k, $k = 1,\ldots, 10$, a positive survival probability.

s_k	Z_k	s_k	Z_k
s_1	z_{11} z_{16} z_{25} z_{83}	s_6	z_{15} z_{78} z_{84} z_{88}
s_2	z_{17} z_{18} z_{43} z_{61} z_{71} z_{84} z_{86}	s_7	z_{18} z_{26} z_{37} z_{55} z_{56}
s_3	z_{12} z_{26} z_{34} z_{55} z_{65} z_{75} z_{83} z_{85}	s_8	z_{11} z_{14} z_{56} z_{73} z_{76}
s_4	z_{17} z_{44} z_{54} z_{62} z_{67} z_{75} z_{84}	s_9	z_{22} z_{23} z_{36} z_{42} z_{46} z_{62}
s_5	z_{14} z_{24} z_{57} z_{68} z_{87}	s_{10}	z_{11} z_{15} z_{18} z_{34}

The cost associated with protecting each zone is indicated in the lower right corner of the corresponding zone. To facilitate the analysis of this example, we give in table 7.1 the composition of the sets Z_k for all $k \in \underline{S}$, *i.e.*, the sets of candidate zones where the survival probability of species s_k is strictly positive. In other words, $Z_k = \{z_{ij} \in Z : q_{ijk} > 0\}$. We denote by $\underline{Z}_k$ the set of indices of the zones of Z_k.

In this example, the threshold value is considered to be the same for all species considered and so we set $\rho_k = \rho$ for all $k \in \underline{S}$. We will examine both problems defined below.

Problem I. Determine a reserve that respects a certain budget, B, and maximizes the number of species whose survival probability – in the set of candidate zones and, therefore, in the reserve – is greater than or equal to a certain threshold value, ρ. We consider the 4 values of ρ, 0.80, 0.85, 0.90, and 0.95, and the 4 values of B, 20, 40, 60, and 80.

Problem II. Determine a minimal cost reserve that ensures that all the species considered have a survival probability – in the set of candidate zones and therefore in the reserve – greater than or equal to a certain threshold value, ρ. We consider the 4 values of ρ, 0.80, 0.85, 0.90, and 0.95.

The solution to Problem I is obtained by solving program $P_{7.7}$ and that of Problem II, by solving program $P_{7.8}$. The results obtained for Problem I are presented in table 7.2. The optimal reserves for some instances of table 7.2 are presented in figure 7.2. The results obtained for Problem II are presented in table 7.3. The optimal reserves corresponding to the instances in table 7.3 are presented in figure 7.3. The resolution of all these instances – by program $P_{7.7}$ or $P_{7.8}$ – is instantaneous and the number of nodes developed in the search tree is very often zero.

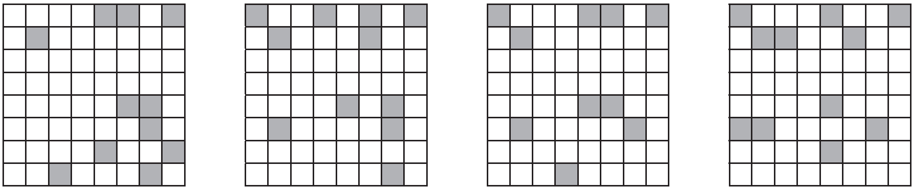
Section 7.4.3 discusses the resolution of large-sized instances.

7.4.3 Computational Experiments on Large-Sized Instances

In these instances, 300 species are concerned and the set of candidate zones is represented by a grid of 20×20 square and identical zones (figure 7.4). The zones are designated by z_{ij} where i represents the row index of the zone and j, its column

TAB. 7.2 – Problem I: Results obtained by solving program $P_{7.7}$ for the example described in figure 7.1, for different threshold values, ρ, and different values of the available budget, B.

B	ρ	Number of species with a survival probability $\geq \rho$	Cost of the reserve	Number of zones in the reserve	Number of nodes in the search tree	Associated figure
20	0.80	7	20	7	0	–
	0.85	5	20	7	0	–
	0.90	5	20	7	0	–
	0.95	3	19	8	0	–
40	0.80	10	40	11	0	7.2a
	0.85	9	40	11	12	7.2b
	0.90	8	40	10	0	7.2c
	0.95	6	39	11	0	7.2d
60	0.80	10	58	14	0	–
	0.85	10	60	15	0	–
	0.90	9	60	15	0	–
	0.95	8	60	17	0	–
80	0.80	10	58	14	0	–
	0.85	10	80	18	0	–
	0.90	10	79	19	0	–
	0.95	9	79	19	0	–



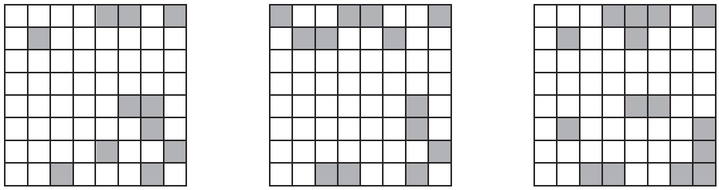
(a) $\rho = 0.80$ **(b)** $\rho = 0.85$ **(c)** $\rho = 0.90$ **(d)** $\rho = 0.95$

FIG. 7.2 – Problem I: Optimal reserves for the instances in table 7.2 corresponding to $B = 40$.

TAB. 7.3 – Problem II: Results obtained by solving program $P_{7.8}$ for the example described in figure 7.1 and for different threshold values.

ρ	Cost of the reserve	Number of zones in the reserve	Number of nodes in the search tree	Associated figure
0.80	40	11	0	7.3.a
0.85	52	13	0	7.3.b
0.90	69	15	0	7.3.c
0.95	–	–	–	–

–: No feasible solution.



a) $\rho = 0.80$ **b)** $\rho = 0.85$ **c)** $\rho = 0.90$

FIG. 7.3 – Problem II: Optimal reserves for the instances in table 7.3.

index. The 300 species considered are divided into 4 groups and, in order to give more or less importance to the different species, a weight is given to each species. All species in the same group have the same weight.

- Group I (species numbered from 1 to 50): This group includes species with a critical extinction risk. The weight of the species in this group is set at 8.
- Group II (species numbered from 51 to 100): This group includes species with a certain extinction risk. The weight of the species in this group is set at 4.
- Group III (species numbered from 101 to 150): This group includes species that are relatively rare but do not currently present an extinction risk. The weight of the species in this group is set at 2.
- Group IV (species numbered from 151 to 300): This group includes relatively common species that do not currently present an extinction risk. The weight of the species in this group is set at 1 (according to the World Wildlife Fund (WWF), many common species are also experiencing a significant decline that should at least be slowed down).

In these experiments, the cost of protecting a zone is generated randomly, in a uniform way, in the set of values $\{1, 2,\ldots, 10\}$. Three values of the available budget, B, are considered: 20, 40, and 60. The probabilities q_{ijk} – the survival probability of species s_k in zone z_{ij} when this zone is protected – are generated at random as follows: for each triplet (i, j, k), a number is generated at random in a uniform way throughout the set $\{1, 2,\ldots, 20\}$. If this number is less than or equal to 18, then $q_{ijk} = 0$ otherwise q_{ijk} is generated at random and uniformly in the set of values $\{0.1, 0.2,\ldots, 0.9\}$. The results obtained for Problem I applied to these instances are presented in table 7.4 for different values of the available budget, B, and the threshold value, ρ. The results obtained for Problem II applied to these instances are presented in table 7.5 for different threshold values, ρ.

We see in table 7.4 that the optimal solutions were obtained in less than 1,800 s of computation for only five instances out of the twelve that are considered. For the other seven instances, the optimal solutions could not be obtained in 1,800 s of computation. For these seven cases, the solutions obtained after 1,800 s of computation are described in the table. In fact, these solutions may be optimal, but even if they are, we do not have any proof of that. Let us examine the case where $B = 40$ and $\rho = 0.85$. The value of the best solution found after 1,800 s of computation is

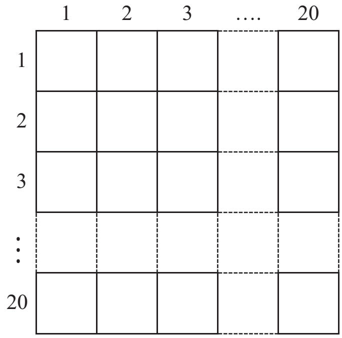


FIG. 7.4 – A set of 400 candidate zones represented by a grid of 20 × 20 square and identical zones.

698 and we are sure that the relative difference, between the value of the optimal solution and the value of this solution, is less than or equal to 0.8%. In the found solution, the number of species whose survival probability in the reserve is greater than or equal to 0.85 is equal to 221: 45 species in Group I, 42 species in Group II, 36 species in Group III, and 98 species in Group IV. This reserve costs 40 units and is composed of 28 zones. In addition, for this instance, 357,119 nodes were developed in the search tree during the 1,800 s of computing.

We see in table 7.5 that the optimal solutions could not be obtained in 1,800 s of computation for the 4 values of ρ considered. For these 4 cases, we describe the solutions obtained after 1,800 s of computation. As with the results in table 7.4, these solutions may sometimes be optimal, but even if they are, we do not have proof of this. Let us look at the case where $\rho = 0.85$. The cost of the best reserve obtained after 1,800 s of computation is equal to 88 and we are sure that the relative difference, between the cost of the optimal reserve and the cost of the obtained reserve, is less than or equal to 2.7%. This reserve is composed of 49 zones. In addition, for this instance, 228,996 nodes were developed in the search tree during the 1,800 s of computation.

7.5 Reserve Maximizing, Under a Budgetary Constraint, the Expected Number of Species of a Given Set that will Survive there

As in the previous sections, the protection of zone z_i ensures the survival of species s_k in this zone with the probability q_{ik} and this for all $i \in \underline{Z}$ and for all $k \in \underline{S}$. As in section 7.4, all the survival probabilities of the species in unprotected zones are considered to be zero. It is therefore assumed that none of the species considered will be able to survive outside the reserve. We consider here that probabilities q_{ik} can be equal to 1, which was not the case in the previous sections. The proposed approach

TAB. 7.4 – Problem I: Results obtained by solving program $P_{7.7}$ for the example described in section 7.4.3 (20×20 candidate zones and 300 species), for different threshold values, ρ, and different values of the available budget, B.

B	ρ	Value of the solution	Number of species with a survival probability $\geq \rho$ and their distribution in each group	Cost of the reserve	Number of zones in the reserve	Number of nodes in the search tree	CPU time (s)
20	0.80	508	151 (36, 27, 24, 64)	20	16	8,652	28
	0.85	443	129 (30, 27, 23, 49)	20	17	31,634	96
	0.90	382	125 (25, 20, 22, 58)	20	19	74,030	280
	0.95	245 (15.8%)	69 (17, 15, 12, 25)	20	17	311,322	1,800
40	0.80	742	245 (47, 41, 45, 112)	40	30	107,106	554
	0.85	698 (0.8%)	221 (45, 42, 36, 98)	40	28	357,119	1,800
	0.90	626 (6.1%)	201 (40, 38, 31, 92)	40	31	334,580	1,800
	0.95	494 (15.7%)	145 (35, 28, 20, 62)	40	30	307,813	1,800
60	0.80	836	287 (50, 50, 49, 138)	60	41	94,303	390
	0.85	807 (1.9%)	267 (50, 48, 46, 123)	60	40	312,308	1,800
	0.90	759 (5.0%)	249 (49, 42, 41, 117)	60	41	282,521	1,800
	0.95	682 (6.5%)	218 (43, 41, 40, 94)	60	43	284,086	1,800

TAB. 7.5 – Problem II: Computation results obtained by solving program $P_{7.8}$ for the example described in section 7.4.3 – 20 × 20 candidate zones and 300 species – for different threshold values, ρ.

ρ	Cost of the reserve	Number of zones in the reserve	Number of nodes in the search tree	CPU time
0.80	80 (5.1%)	46	185,254	1,800
0.85	88 (2.7%)	49	228,996	1,800
0.90	104 (5.0%)	58	168,910	1,800
0.95	131 (4.5%)	65	153,877	1,800

for this case can be easily adapted to other contexts. The aim is to determine a reserve, *i.e.,* a set of zones to be protected, with a cost less than or equal to a certain value, B, in order to maximize the expected number of protected species, *i.e.*, here, the expected number of species that will survive in this reserve. Different importance is given to each species – reflected in a weight assigned to each species – and we consider the expected weighted number of species that will survive in the reserve.

7.5.1 Mathematical Programming Formulation

As we have seen in section 7.2, the expression of the survival probability in the reserve of species s_k, as a function of the Boolean variables x_i, is equal to $1 - \prod_{i \in \underline{Z}} (1 - q_{ik} x_i)$. Remember that the reserve is defined by zones z_i such as $x_i = 1$. We deduce that the expected number of species that will survive in the reserve is equal to $\sum_{k \in \underline{S}} \left[1 - \prod_{i \in \underline{Z}} (1 - q_{ik} x_i)\right]$ and that the expected weighted number of species that will survive in the reserve is equal to $\sum_{k \in \underline{S}} w_k \left[1 - \prod_{i \in \underline{Z}} (1 - q_{ik} x_i)\right]$ where w_k is the weight assigned to species s_k. The problem considered can, therefore, be formulated as the mathematical program in Boolean variables $P_{7.9}$.

$$
P_{7.9} : \begin{cases} \max \sum_{k \in \underline{S}} w_k \left(1 - \prod_{i \in \underline{Z}} (1 - q_{ik} x_i)\right) \\ \text{s.t.} \left| \begin{array}{lll} \sum_{i \in \underline{Z}} c_i x_i \le B & & (7.9.1) \\ x_i \in \{0,1\} & i \in \underline{Z} & (7.9.2) \end{array} \right. \end{cases}
$$

Using the real variable β_k to represent the quantity $\prod_{i \in \underline{Z}} (1 - q_{ik} x_i)$, *i.e.*, the disappearance probability of species s_k from the reserve, program $P_{7.9}$ can be rewritten as program $P_{7.10}$.

$$
P_{7.10} : \begin{cases} \max \sum_{k \in \underline{S}} w_k (1 - \beta_k) \\ \text{s.t.} \left| \begin{array}{llll} \beta_k = \prod_{i \in \underline{Z}} (1 - q_{ik} x_i) \quad k \in \underline{S} & (7.10.1) & \mid \quad x_i \in \{0,1\} \quad i \in \underline{Z} & (7.10.3) \\ \sum_{i \in \underline{Z}} c_i x_i \le B & (7.10.2) & \mid \quad \beta_k \ge 0 \quad k \in \underline{S} & (7.10.4) \end{array} \right. \end{cases}
$$

Using the basic properties of the logarithmic function, program $P_{7.10}$ can be rewritten as program $P_{7.11}$ in which, for all $k \in \underline{S}$, $I_k^{<1} = \{i \in \underline{Z} : 0 < q_{ik} < 1\}$, $I_k^{=1} = \{i \in \underline{Z} : q_{ik} = 1\}$, and α_k is a real variable equal to $\prod_{i \in \underline{Z}, q_{ik} \neq 1} (1 - q_{ik} x_i)$, *i.e.*, $\prod_{i \in I_k^{<1}} (1 - q_{ik} x_i)$.

$$
P_{7.11} : \begin{cases} \max \sum\limits_{k \in \underline{S}} w_k (1 - \beta_k) \\ \text{s.t.} \left| \begin{array}{llll} \log \alpha_k = \sum\limits_{i \in I_k^{<1}} x_i \log(1 - q_{ik}) & k \in \underline{S} & (7.11.1) & | \quad x_i \in \{0,1\} \quad i \in \underline{Z} \quad (7.11.4) \\ \beta_k \geq \alpha_k - \sum\limits_{i \in I_k^{=1}} x_i & k \in \underline{S} & (7.11.2) & | \quad \alpha_k, \beta_k \geq 0 \quad k \in \underline{S} \quad (7.11.5) \\ \sum\limits_{i \in \underline{Z}} c_i x_i \leq B & & (7.11.3) & | \end{array} \right. \end{cases}
$$

By convention $\sum_{i \in I_k^{<1}} x_i \log(1 - q_{ik}) = 0$ if $I_k^{<1} = \emptyset$ and $\sum_{i \in I_k^{=1}} x_i = 0$ if $I_k^{=1} = \emptyset$. First of all, let us observe that variable β_k appears in the economic function with a negative coefficient. Since this variable appears only in constraints 7.11.2 and 7.11.5, it takes, in any optimal solution of $P_{7.11}$, the smallest possible value, *i.e.*, the value $\max\{0, \alpha_k - \sum_{i \in I_k^{=1}} x_i\}$, *i.e.*, the value $\max\left\{0, \prod_{i \in I_k^{<1}} (1 - q_{ik} x_i) - \sum_{i \in I_k^{=1}} x_i\right\}$ since constraints 7.11.1 are equivalent to constraints $\alpha_k = \prod_{i \in I_k^{<1}} (1 - q_{ik} x_i)$. If for at least one index $i \in I_k^{=1}$, $x_i = 1$, then the expression $\prod_{i \in I_k^{<1}} (1 - q_{ik} x_i) - \sum_{i \in I_k^{=1}} x_i$ is negative or zero and variable β_k takes the value 0. On the contrary, if $x_i = 0$ for all $i \in I_k^{=1}$, β_k takes the value $\prod_{i \in I_k^{<1}} (1 - q_{ik} x_i)$. Thus, β_k takes the value $\prod_{i \in \underline{Z}} (1 - q_{ik} x_i)$ in any optimal solution of $P_{7.11}$. In this program, all expressions are linear according to variables α_k, β_k, and x_i, except the left member of constraints 7.11.1 which is equal to $\log \alpha_k$.

Remark. If all survival probabilities are strictly less than 1, program $P_{7.11}$ becomes simpler and is transformed into $P_{7.12}$.

$$
P_{7.12} : \begin{cases} \max \sum\limits_{k \in \underline{S}} w_k (1 - \alpha_k) \\ \text{s.t.} \left| \begin{array}{llll} \log \alpha_k = \sum\limits_{i \in Z} x_i \log(1 - q_{ik}) & k \in \underline{S} & (7.12.1) & | \quad x_i \in \{0,1\} \quad i \in \underline{Z} \quad (7.12.3) \\ \sum\limits_{i \in \underline{Z}} c_i x_i \leq B & & (7.12.2) & | \quad \alpha_k \geq 0 \quad k \in \underline{S} \quad (7.12.4) \end{array} \right. \end{cases}
$$

The right-hand side of constraint 7.12.1 can also be written in this case $\sum_{i \in I_k^{<1}} x_i \log(1 - q_{ik})$. Note also that if constraints 7.11.5 and 7.12.4 specify, suitably for mathematical programming, that variables α_k are positive or zero, these variables will in fact take a strictly positive value in any feasible solution of the corresponding programs.

7.5.2 Problem Relaxation and Determination of an Approximate Solution

The approximate resolution of $P_{7.11}$ that we propose can be interpreted as an approximation of the logarithmic function, which is a concave function, by a piecewise linear function greater than or equal to the logarithmic function at all points (see appendix at the end of this book and figure 7.5). The advantage of this approach is that it provides not only an approximate solution to the problem but also a relaxation of the initial problem and thus an upper bound of the true value of the optimal solution. In other words, the method provides – using an integer linear programming solver – an approximate solution as well as some guarantee on the value of this solution.

Relaxation of $P_{7.11}$: To build a relaxation of $P_{7.11}$, the idea is to replace, within the constraints 7.11.1 and for all $k \in \underline{S}$, the logarithmic function, $\log \alpha_k$, by a piecewise linear function, $f(\alpha_k)$, greater than or equal to $\log \alpha_k$ for all α_k such that $0 < \alpha_k \leq 1$ (figure 7.5).

This substitution leads to program $P_{7.13}$. This is a relaxation of $P_{7.11}$ that includes a piecewise linear function. Indeed, any feasible solution of $P_{7.11}$ is also a feasible solution of $P_{7.13}$ since, if the constraint, $\log \alpha_k = \sum_{i \in I_k^{<1}} \log(1 - q_{ik}) x_i$, $k \in \underline{S}$, is satisfied, then the constraint $f(\alpha_k) \geq \sum_{i \in I_k^{<1}} \log(1 - q_{ik}) x_i$, $k \in \underline{S}$, is also satisfied. In addition, the economic functions of $P_{7.11}$ and $P_{7.13}$ are identical. Note that $P_{7.13}$ can easily be converted into a linear program – in Boolean variables – since function f is concave and appears in the left member of a "greater than or equal" constraint (see appendix at the end of the book).

$$P_{7.13} : \begin{cases} \max \sum_{k \in \underline{S}} w_k (1 - \beta_k) \\ \text{s.t.} \left| \begin{array}{ll} (7.11.2), (7.11.3), (7.11.4), (7.11.5) & \\ f(\alpha_k) \geq \sum_{i \in I_k^{<1}} x_i \log(1 - q_{ik}) & k \in \underline{S} \quad (7.13.1) \end{array} \right. \end{cases}$$

Let $(\overline{x}, \overline{\alpha}, \overline{\beta})$ be an optimal solution of $P_{7.13}$. An approximate solution to the initial problem – a reserve – is given by $\overline{x}$; its value, *i.e.*, the expected weighted number of species protected by this reserve, is equal to $\sum_{k \in \underline{S}} w_k \left[1 - \prod_{i \in \underline{Z}} (1 - q_{ik} \overline{x}_i)\right]$. An upper bound of the true optimal value of the problem considered is given by the optimal value of $P_{7.13}$, *i.e.*, $\sum_{k \in \underline{S}} w_k (1 - \overline{\beta}_k)$. The relative gap between the value of the optimal solution and the value of the approximate solution is, therefore, less than or equal to $\left(\sum_{k \in \underline{S}} w_k (1 - \overline{\beta}_k) - \sum_{k \in \underline{S}} w_k \left[1 - \prod_{i \in \underline{Z}} (1 - q_{ik} \overline{x}_i)\right]\right) \Big/ \sum_{k \in \underline{S}} w_k \left[1 - \prod_{i \in \underline{Z}} (1 - q_{ik} \overline{x}_i)\right]$.

Let us now see how to construct the piecewise linear function f such that $f(\alpha_k)$ is greater than or equal to $\log \alpha_k$ for all α_k such as $0 < \alpha_k \leq 1$.

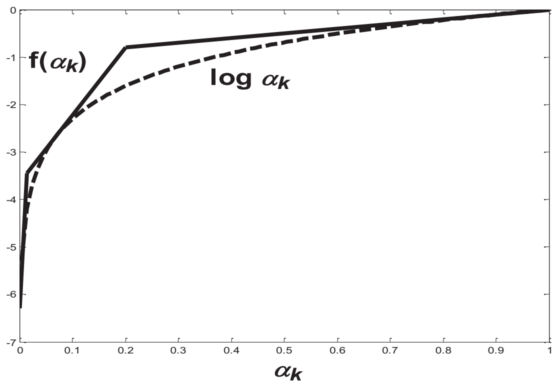


FIG. 7.5 – An upper approximation of $\log \alpha_k$ for $0 < \alpha_k \leq 1$ (dotted curve) by a piecewise linear function, $f(\alpha_k)$ (solid line curve).

Lemma 7.1. For all $a > 0$ and $b > 0$, $\log a + \frac{1}{a}(b - a) \geq \log b$.

The proof is immediate since $1/a$ is the value of the derivative of the function $\log y$ at point a, and the function $\log y$ is concave.

Lemma 7.2. Let u be a vector of $\mathbb{R}^V$ such that $0 < u_1 < u_2 < \cdots < u_V = 1$, and let f be the piecewise linear function composed of the V segments tangent to the curve $\log \alpha_k$ at the V points of abscissa $u_1, u_2,..., u_V$. The following 5 properties are satisfied:

(i) The abscissa of the breaking points of f are $bp_v = \frac{\log u_{v+1} - \log u_v}{1/u_v - 1/u_{v+1}}, v = 1,..., V - 1$.

(ii) For α_k such that $0 < \alpha_k \leq bp_1$, $f(\alpha_k) = \frac{1}{u_1}\alpha_k + \log u_1 - 1$.

(iii) For α_k such that $bp_v \leq \alpha_k \leq bp_{v+1}$, $f(\alpha_k) = \frac{1}{u_{v+1}}\alpha_k + \log u_{v+1} - 1$, $v = 1,..., V - 1$.

(iv) The function $f(\alpha_k)$ is concave.

(v) $f(\alpha_k) \geq \log(\alpha_k)$ for all $\alpha_k \in]0, 1]$.

Proof. The proof of (i) is immediate; it results from elementary calculations.

The proofs of (ii) and (iii) are direct consequences of lemma 7.1.

The successive slopes of the function f are $1/u_1 > 1/u_2 > \cdots > 1/u_V$. The function f is therefore concave.

The proof of (v) is a direct consequence of lemma 7.1.

In summary, the expression $\log \alpha_k$, which appears in $P_{7.11}$, is upper approximated in $P_{7.13}$ by $f(\alpha_k)$ where f is the piecewise linear function defined in lemma 7.2 and illustrated in figure 7.5. Program $P_{7.11}$ can, therefore, be reformulated – in an approximate way – by a mixed-integer linear program by adding a single set of constraints. This gives program $P_{7.14}$.

$$\mathrm{P}_{7.14}: \begin{cases} \max \sum_{k \in \underline{S}} w_k(1-\beta_k) \\ \text{s.t.} \left| \begin{array}{l} (7.11.2), (7.11.3), (7.11.4), (7.11.5) \\ \frac{\alpha_k}{u_v} + \log u_v - 1 \geq \sum_{i \in I_k^{<1}} x_i \log(1-q_{ik}) \quad k \in \underline{S}, v = 1,..., V \quad (7.14.1) \end{array} \right. \end{cases}$$

Another way to proceed is to use a modelling language that automatically performs this conversion. This is the case, for example, with the AMPL language (Fourer *et al.*, 1993). Thus, using the syntax of this language, the function $f(\alpha_k)$ is defined as follows:

< < {v in 1..V-1} bp[v]; {v in 1..V} 1/u[v] > > (fα[k], 1).

The expression between < < and > > describes the piecewise linear function, and is followed by the name of the variable concerned. Here, this variable, which represents $f(\alpha_k)$, is denoted by fα[k]. There are two parts in this expression, the list of the abscissa of the breaking points, bp_v, $v = 1,\ldots, V-1$, corresponding to the changes in the slope of the function, and the list of the slopes, $1/u_v$, $v = 1,\ldots, V$. The two lists are separated by a semicolon. The first slope is the slope before the first breaking point and the last slope is the slope after the last breaking point. For the function to be perfectly defined, it is also necessary to specify at what point it takes the value 0. Here $f(1) = 0$.

An optimal solution of $\mathrm{P}_{7.14}$ provides a feasible solution to the problem, *i.e.*, a set of zones to be protected to form the reserve. By definition, the value of this approximate solution is equal to $\sum_{k\in\underline{S}} w_k[1-\prod_{i\in\underline{Z}}(1-q_{ik}\overline{x}_i)]$ where $\overline{x}_i$ is the value of variable x_i in an optimal solution of $\mathrm{P}_{7.14}$. Since $\mathrm{P}_{7.14}$ is a relaxation of $\mathrm{P}_{7.11}$, the optimal value of $\mathrm{P}_{7.14}$ gives an upper bound of the optimal value of $\mathrm{P}_{7.11}$ and, therefore, an upper bound of the value of the optimal solution of the reserve selection problem considered. Thus, the formulation $\mathrm{P}_{7.14}$ allows us to obtain, with the help of integer linear programming, an approximate solution to our problem but also an upper bound on the gap between the value of this solution and the value of an optimal solution. To obtain a good approximation of $\log \alpha_k$ by the piecewise linear function defined in lemma 7.2, V must be large enough. However, the larger V is, the larger the size of program $\mathrm{P}_{7.14}$ is. The results of the experiments presented in section 7.5.3 show that by choosing the vector u carefully – see the example in section 7.5.3 – we obtain an approximate solution whose value is very close to the value of the optimal solution. Thus, program $\mathrm{P}_{7.14}$ provides a solution to the problem, regardless of the probability values, and provides a guarantee on the quality of the solution obtained.

Remark. Again, the problem considered in this section 7.5 can be interpreted in a slightly different way – as some authors have done – assuming that the presence of species s_k in zone z_i is defined by a probability denoted by q_{ik}. The problem considered above then becomes: determine a reserve that respects a budgetary constraint and maximizes the expected number of species present in this reserve. Thus, in the first interpretation we are interested in the mathematical expectation of the weighted number of species that will survive in the reserve and in the second, in the mathematical expectation of the weighted number of species that are protected, at least in some way, *i.e.*, present in the reserve.

7.5.3 Example

We illustrate the solution to the problem studied in section 7.5, which we recall here: determine a reserve of cost less than or equal to a certain value, B, and which maximizes the expected number of species that will survive in this reserve. A hypothetical set of candidate zones, represented by a grid of 8×8 square and identical zones and described in section 7.4.2, is considered. In this example, 10 species, s_1, s_2,..., s_{10}, are concerned. The description of this example concerns the list of the candidate zones, z_{ij}, $i = 1$,..., 8, $j = 1$,..., 8, the survival probabilities of the species in the different zones if they are protected, q_{ijk} for the survival of species s_k in zone z_{ij}, and finally the cost of protecting zones, c_{ij} for zone z_{ij}. Note that, in this example, all the survival probabilities are strictly less than 1. The weight of species s_k is denoted by w_k and, in this example, $w = (1, 1, 4, 2, 1, 2, 4, 2, 1, 2)$. All survival probabilities in unprotected zones are zero. Remember that Z_k refers to the set of candidate zones in which the survival probability of species s_k is strictly positive in case of protection – $Z_k = \{z_{ij} \in Z : q_{ijk} > 0\}$ – and we denote by $\underline{Z}_k$ the set of index pairs associated with the zones of Z_k. Different values of the available budget, B, are considered. The values obtained for the expected weighted number of species range from 37 to 98% of the largest possible value of the expected weighted number of species which is equal to $\sum_{k=1}^{10} w_k = 20$. The vector u (see section 7.5.2) is chosen as follows: $u_v = u_1^{(V-v)/(V-1)}$ for $v = 1$,..., V with $u_1 = 0.01$ and $V = 20$. Choosing $u_v = u_1^{(V-v)/(V-1)}$, $v = 1$,..., V, produces a regular decrease in the slope of the piecewise linear function f defined in lemma 7.2. since $\frac{1/u_{v+1}}{1/u_v} = \frac{u_v}{u_{v+1}} = \frac{u_1^{(V-v)/(V-1)}}{u_1^{(V-v-1)/(V-1)}} = u_1^{[(V-v)/(V-1)-(V-v-1)/(V-1)]} = u_1^{1/(V-1)}$. The experimental results are presented in table 7.6. For example, let $(\overline{x}, \overline{\alpha}, \overline{\beta})$ be an optimal solution of $\mathrm{P}_{7.14}$ when $B = 50$. In this case, the expected weighted number of species, $\sum_{k\in\underline{S}} w_k \left[1 - \prod_{i\in\underline{Z}} (1 - q_{ik}\overline{x}_i)\right]$, is equal to 18.59, the value of the optimal solution of $\mathrm{P}_{7.14}$, $\sum_{k\in\underline{S}} w_k(1 - \overline{\beta}_k)$, which is an upper bound of the optimal value of the problem considered, is also equal to 18.59, and the associated relative gap is equal to 0.02% – expressing the different results with a precision of two decimal places. Note that the expected weighted number of species is 19.69 if all the zones are selected. The solution corresponding to $B = 50$ is represented by figure 7.6a: 14 zones are selected at a cost equal to the available budget, *i.e.*, 50 units.

The results in table 7.6 show that, in this instance, the approach is very effective in solving the problem under consideration. Indeed, the relative gap is always less than 0.1%. It should also be noted that all the solutions are obtained almost instantaneously. However, the reserves obtained can be very fragmented when there is no compactness constraint. If we introduce a compactness constraint consisting in prohibiting the distance between two zones of the reserve from being more than

TAB. 7.6 – Results concerning the resolution by $P_{7.14}$ of the problem studied in this example. The instance under consideration is described in section 7.4.2 (8×8 zones and 10 species), $w = (1, 1, 4, 2, 1, 2, 4, 2, 1, 2)$, $u_1 = 0.01$, and $V = 20$.

B	Number of selected zones	Budget used	Expected weighted number of species	Upper bound	Relative gap (%)	Associated figure
10	4	10	11.10	11.11	0.08	–
30	11	30	17.04	17.05	0.05	–
50	14	50	18.59	18.59	0.02	7.6a
50*	8	43	13.04	13.05	0.08	7.6b
70	17	70	19.21	19.23	0.05	–
150	29	150	19.67	19.69	0.10	–
200	35	180	19.69	19.70	0.07	–

(a) $B = 50$, no compactness constraints.

(b) $B = 50$, with a compactness constraint.

FIG. 7.6 – An optimal reserve with or without a compactness constraint when the available budget, B, is equal to 50 (see rows "50" and "50*" of table 7.6).

3 units, we obtain, when the budget is equal to 50, the solution described in the 4th row of table 7.6 (50*) and by figure 7.6b. In this case, the compactness constraint decreases the value of the solution by about 30%.

7.5.4 Computational Experiments on Large-Sized Instances

In order to test the effectiveness of the approach, various large-sized artificial instances were tested. In these instances, 300 species are concerned and the set of candidate zones is represented by a grid of 20×20 square and identical zones (figure 7.7). The zones are designated by z_{ij} where i represents the row index of the zone and j its column index. The 300 species considered are divided into 4 groups and a weight is assigned to each species. All species in the same group have the same weight.

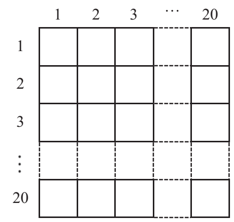


FIG. 7.7 – A set of 400 candidate zones represented by a grid of 20×20 square and identical cells.

- Group I (species numbered from 1 to 50): This group includes species with a critical extinction risk. The weight of the species in this group is set at 8.
- Group II (species numbered from 51 to 100): This group includes species with a certain extinction risk. The weight of the species in this group is set at 4.
- Group III (species numbered from 101 to 150): This group includes species that are relatively rare but do not currently present an extinction risk. The weight of the species in this group is set at 2.
- Group IV (species numbered from 151 to 300): This group includes relatively common species that do not currently present an extinction risk. The weight of the species in this group is set at 1.

In these experiments, the cost of protecting a zone is generated randomly, in a uniform way, within the set of values $\{1, 2,..., 10\}$. Six values of the available budget, B, are considered: 20, 40, 60, 80, 100, and 120. These values of B were chosen in order to obtain an expected weighted number of species ranging from 60 to 100% of the largest possible value of the expected weighted number of species, $\sum_{k=1,300} w_k = 850$. The probabilities q_{ijk} – the survival probability of species s_k in zone z_{ij} if it is protected – are drawn at random as follows: for each triplet (i, j, k), a number is generated at random in a uniform way from the set $\{1, 2,..., 20\}$. If this number is less than or equal to 18, then $q_{ijk} = 0$ otherwise q_{ijk} is randomly drawn uniformly from the set of values $\{0.1, 0.2,..., 0.9\}$. As in the example in section 7.5.3, the vector u is chosen as follows: $u_v = u_1^{(V-v)/(V-1)}$ for $v = 1,..., V$ with $u_1 = 0.01$ and $V = 20$. Note that to obtain even more accurate approximations, the values of u_1 can be chosen according to the value of B. Indeed, when the available budget is large, many survival probabilities of the species in the reserve are close to 1 in an optimal solution, which implies that many extinction probabilities (α_k, see section 7.5.1) are close to 0. To obtain a good approximation of the logarithmic function by the piecewise linear function defined in lemma 7.2, it may therefore be interesting to decrease the value of u_1 when the value of B increases.

The experimental results are presented in table 7.7 for different values of the available budget, B.

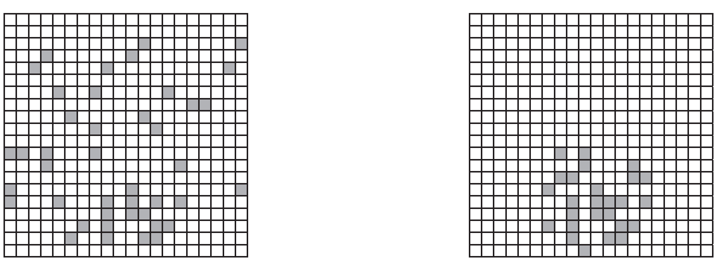
For example, let $(\overline{x}, \overline{\alpha}, \overline{\beta})$ be an optimal solution of $P_{7.14}$ when $B = 60$. In this case, the expected weighted number of protected species, $\sum_{k \in \underline{S}} w_k \Big[1 - \prod_{i \in \underline{Z}}$

TAB. 7.7 – Results regarding the resolution of the problem by $P_{7.14}$ for the example described in this section 7.5.4 (20×20 zones, 300 species, $u_1 = 0.01$, and $V = 20$).

B	Number of selected zones	Budget used	Expected weighted number of species	Upper bound	Relative gap (%)	CPU time (s)	Number of nodes in the search tree	Associated figure
20	19	20	598.59	599.14	0.09	6	641	–
40	31	40	745.54	745.98	0.06	24	4,903	–
60	42	60	802.16	802.62	0.06	59	15,060	7.7a
60*	26	60	686.56	687.08	0.08	45	4,152	7.7b
80	52	80	825.47	826.04	0.07	217	92,935	–
100	59	100	836.41	837.20	0.09	549	407,188	–
120	67	120	842.33	843.10	0.09	538	416,902	–

$(1 - q_{ik}\overline{x}_i)]$, is equal to 802.16, the value of the upper bound, *i.e.*, the value of the optimal solution of $P_{7.14}$, $\sum_{k\in \underline{S}} w_k(1 - \overline{\beta}_k)$, is equal to 802.62, and the associated relative gap, $100\left(\sum_{k\in \underline{S}} w_k(1 - \overline{\beta}_k) - \sum_{k\in \underline{S}} w_k \left[1 - \prod_{i\in \underline{Z}} (1 - q_{ik}\overline{x}_i)\right]\right)$ divided by $\sum_{k\in \underline{S}} w_k \left[1 - \prod_{i\in \underline{Z}} (1 - q_{ik}\overline{x}_i)\right]$, is equal to 0.06%. Also in the case where $B = 60$, the CPU time required to solve the problem by $P_{7.14}$ is equal to 59 s, and 15,060 nodes are developed in the search tree. The solution – without any compactness constraints – is presented in figure 7.8a: 42 zones are selected and the corresponding cost is equal to the value of the budget. Figure 7.8b represents the solution obtained by adding a compactness constraint that imposes a maximal distance of 9 units between two zones.

The results in table 7.7 show that in this example – 300 species and 400 zones – the approach is effective in solving the problem under consideration. Indeed, the average computation time is about 205 s and the relative gap is always less than 0.1%. The larger B is, the longer the computation time required is. It can also be noted that the value of B clearly influences the value of the optimal solution: increasing B from 40 to 100 increases the expected weighted number of species by about 12%. It should also be noted that the solution presented in figure 7.8a is very fragmented since some zones are very far from each other. The diameter of this reserve is equal to approximately 23 units, *i.e.*, the distance between zones $z_{3,20}$ and $z_{16,1}$, if the length of the side of each zone on the grid is equal to one unit and if the distance between two zones is measured by the distance between the centres of these two zones. If we impose the compactness constraint consisting in prohibiting the distance between two zones of the reserve from being greater than 9 units, we obtain, for $B = 60$, the solution described in the 4th row of table 7.7 (60*) and by figure 7.8b. We can see that this compactness constraint decreases the value of the solution by about 15%. For very large problems, it would probably be difficult to obtain the optimal solution within a reasonable computation time. A heuristic approach should then be used. A simple way to do this would be to use a truncated branch and bound procedure, a procedure that is easy to implement by limiting the



(a) Expected weighted number of species: 802.16 (42 zones)

(b) Expected weighted number of species: 686.56 (26 zones)

FIG. 7.8 – Optimal reserves for a budget of 60 units. (a) Without a compactness constraint. (b) With a compactness constraint imposing a maximal distance of 9 units between two zones.

computation time allocated to the solver to address the problem, as we did in chapter 3, section 3.8, to determine “good” connected reserves.

7.6 Consideration of Uncertainties Affecting Species Survival Probabilities

7.6.1 Reserve Ensuring that as Many Species as Possible, of a Given Set, Have a Certain Survival Probability, Under a Budgetary Constraint: A Robust Reserve

As in section 7.2, we consider the problem of selecting a nature reserve, *i.e.*, a set of zones to be protected, which ensures that as many species as possible have at least some survival probability, taking into account a budgetary constraint. An important point in the problem of section 7.2 is that the survival probabilities of each species in each zone are assumed to be perfectly known. Thus, we know the survival probability of species s_k in zone z_i if zone z_i is protected and also if it is not. In reality, there may be errors in determining these probabilities. Ignoring these errors can lead to reserves whose actual interest is quite far from the interest pursued. In this section, we study the case where the values of the survival probabilities of each species in each protected zone are subject to certain errors (see appendix at the end of this book). However, we assume that the number of zones for which these values may be incorrect is limited. We further assume that the survival probabilities of the species in the unprotected zones are all 0. Thus, for each species and each protected zone, we define a set of possible values for the survival probability of the considered species in the considered zone. The problem we are studying is then to determine a reserve that respects a certain budget, B, and ensures that as many species as possible have at least some survival probability regardless of the values taken by the survival probabilities of each species in each zone, in the set of possible values. In other words, the selected reserve guarantees a survival probability greater than or equal to a certain threshold value for a certain number of species, regardless of the errors that have been made – among a set of possible errors – in the evaluation of the probabilities. In addition, there is no reserve – with a cost less than or equal to B – to do this for a higher number of species. We will say that the reserve obtained is an “optimal robust reserve” in the sense that the main property of this reserve – set out above – is independent of any errors that may exist in the data. An optimal robust reserve thus provides some guarantee against data uncertainty, but this guarantee has a cost and we will see in the experiments presented in section 7.6.5.2 that this cost can be high.

7.6.2 Description of Uncertainties

As noted above, it is assumed that there is some uncertainty in estimating the survival probabilities of each species in each zone. Thus, the only certitude is that the value of the survival probability of species s_k in the protected zone z_i belongs to the interval $[q_{ik} - \delta_{ik}, q_{ik} + \gamma_{ik}]$ with $\delta_{ik}, \gamma_{ik} \geq 0$, $\delta_{ik} \leq q_{ik}$, $\gamma_{ik} \geq 0$ and $q_{ik} + \gamma_{ik} \leq 1$, and that, if zone z_i is not protected, the survival probability of any species in this zone is equal to 0. For the sake of simplifying the presentation we assume that q_{ik} is strictly less than 1. The problem considered is to determine an optimal robust reserve, *i.e.*, a reserve that maximizes the number of species whose survival probability in the reserve is greater than or equal to a threshold value – equal to ρ_k for species s_k – regardless of the values taken by the survival probability of species s_k in the protected zone z_i within the interval $[q_{ik} - \delta_{ik}, q_{ik} + \gamma_{ik}]$, $i \in \underline{Z}$, $k \in \underline{S}$. Again, to simplify the presentation we assume that ρ_k is strictly less than 1. A reserve that is optimal for the nominal values q_{ik} of the survival probabilities may not be robust to uncertainty. Note that to determine an optimal robust reserve as we have defined it, we only need to consider that the values of the survival probabilities belong to the restricted intervals $[q_{ik} - \delta_{ik}, q_{ik}]$, $i \in \underline{Z}$, $k \in \underline{S}$.

For example, we can consider that, for each species s_k, a maximal error of $r_k\%$ $(0 \leq r_k \leq 100)$ could have been made on the estimation of the probability q_{ik}, $i \in \underline{Z}$, *i.e.*, $\delta_{ik} = r_k\, q_{ik}/100$. Note that in this case it is assumed that the error depends on the species but not on the zone. Unless other assumptions are made, the optimal robust reserve is obtained by solving program $\mathrm{P}_{7.3}$ in which the survival probabilities are set at their minimal value, $q_{ik} - \delta_{ik}$, for any species s_k and for any protected zone z_i. This results in a reserve that, in a certain way, provides a complete guarantee against uncertainty. Indeed, whatever the value taken by the survival probability of species s_k in zone z_i within the interval $[q_{ik} - \delta_{ik}, q_{ik}]$, $i \in \underline{Z}$, $k \in \underline{S}$, the survival probability of species s_k in the obtained reserve remains greater than or equal to ρ_k if this probability is already greater than or equal to ρ_k when we consider that the survival probability of species s_k in zone z_i is equal to $q_{ik} - \delta_{ik}$. However, retaining this very pessimistic hypothesis may lead to the selection of a very costly reserve. To avoid this pitfall, we consider that it is unlikely that there is an error on all the survival probabilities. In fact, we consider that this is impossible. Thus, we assume that, for a species s_k, the survival probabilities – not zero – in the different zones may differ from their nominal value, q_{ik}, in at most Γ_k zones of Z_k, Z_k designating the subset of zones of Z whose protection generates a strictly positive nominal survival probability of species s_k. We have thus $Z_k = \{z_i \in Z, q_{ik} > 0\}$. For example, Γ_k can be a proportion of the number of zones of Z_k: $\Gamma_k = \lfloor \eta_k\, |Z_k|/100 \rfloor$ where η_k is a constant between 0 and 100 and $\lfloor a \rfloor$ denotes the integer part of a. Thus, setting η_k to 0 means that there is no uncertainty about the different survival probabilities of species s_k and, on the contrary, setting η_k to 100 means that all the survival probabilities of species s_k, other than 0, can take the value $q_{ik} - \delta_{ik}$ instead of their nominal value, q_{ik}. Solving the problem considered in these two extreme cases is like solving a problem without uncertainty with regard to the survival probabilities: if $\Gamma_k = 0$, everything happens as if the survival probability of species s_k in zone z_i were set to q_{ik}, and when $\Gamma_k = |Z_k|$, everything happens as if this probability were set to $q_{ik} - \delta_{ik}$. In intermediate cases $(0 < \eta_k < 100)$, for each species s_k and for a fixed reserve R, the worst-case occurs when, in v_k zones of the reserve, the survival probability takes the value $q_{ik} - \delta_{ik}$ instead of the value q_{ik} with $v_k = \min\{\Gamma_k, |Z_k \cap R|\}$. We will show that these v_k zones correspond to the v_k highest values of the expression $\delta_{ik}/(1 - q_{ik})$ found in the reserve – for species s_k and varying i. Indeed, by using the Boolean variable t_{ik}, $i \in \underline{Z}$, $k \in \underline{S}$, that is equal to 1 if and only if the survival probability of species s_k in the protected zone z_i is equal to $q_{ik} - \delta_{ik}$ instead of q_{ik}, the problem of minimizing the survival probability of species s_k in reserve R, taking into account the uncertainty considered, can be formulated as the maximization of the extinction probability, *i.e.*, the maximization problem:

$$\max_{t_{ik} \in \{0,1\}\ (i \in \underline{Z}_k \cap \underline{R})} \left\{ \prod_{i \in \underline{Z}_k \cap \underline{R}} (1 - q_{ik} + t_{ik}\ \delta_{ik}) : \sum_{i \in \underline{Z}_k \cap \underline{R}} t_{ik} \leq v_k \right\} \qquad \text{(a)}$$

Remember that $\underline{Z}_k$ refers to the set of indices of the zones belonging to Z_k and that $\underline{R}$ refers to the set of indices of the zones belonging to reserve R. We will show that the solution to the maximization problem (a) is obtained by fixing to 1 the v_k variables t_{ik} corresponding to the v_k largest values of the expression $\delta_{ik}/(1 - q_{ik})$, obtained by varying the index i in the set $\underline{Z}_k \cap \underline{R}$. Using the logarithmic function and taking into account that variables t_{ik} can only take the values 0 or 1, the maximization problem (a) is equivalent to the following maximization problem:

$$\max_{t_{ik} \in \{0,1\}\ (i \in \underline{Z}_k \cap \underline{R})} \left\{ \sum_{i \in \underline{Z}_k \cap \underline{R}} \log\ (1 - q_{ik} + t_{ik}\ \delta_{ik}) : \sum_{i \in \underline{Z}_k \cap \underline{R}} t_{ik} \leq v_k \right\}$$

or

$$\max_{t_{ik} \in \{0,1\}\ (i \in \underline{Z}_k \cap \underline{R})} \left\{ \begin{array}{l} \sum_{i \in \underline{Z}_k \cap \underline{R}} \log(1 - q_{ik}) \\ \quad + \sum_{i \in \underline{Z}_k \cap \underline{R}} [\log(1 - q_{ik} + \delta_{ik}) - \log(1 - q_{ik})]\, t_{ik} : \sum_{i \in \underline{Z}_k \cap \underline{R}} t_{ik} \leq v_k \end{array} \right\}$$

which is itself equivalent to

$$\max_{t_{ik} \in \{0,1\}\ (i \in \underline{Z}_k \cap \underline{R})} \left\{ \sum_{i \in \underline{Z}_k \cap \underline{R}} t_{ik} \log\left(1 + \frac{\delta_{ik}}{1 - q_{ik}}\right) : \sum_{i \in \underline{Z}_k \cap \underline{R}} t_{ik} \leq v_k \right\} \qquad \text{(b)}$$

since the constant $\sum_{i \in \underline{Z}_k \cap \underline{R}} \log(1 - q_{ik})$ does not have a role in the maximization problem. An optimal solution of (b) and therefore of (a) is obtained by fixing to 1 the v_k variables t_{ik} corresponding to the largest values of the expression $\delta_{ik}/(1 - q_{ik})$.

7.6.3 Determination of a Robust Reserve by Mathematical Programming

Let us consider a reserve, R, defined by the values of the Boolean variables $\overline{x}_i$, $i \in \underline{Z}$. Zone z_i belongs to the reserve if and only if $\overline{x}_i = 1$. As we have seen previously, the Boolean variable t_{ik} allows us to specify the value taken by the survival probability of species s_k in zone z_i. Variable t_{ik} takes the value 1 if this survival probability is equal to $q_{ik} - \delta_{ik}$. In this case, the extinction probability is equal to $1 - (q_{ik} - \delta_{ik})$.

Variable t_{ik} takes the value 0 if this survival probability is equal to q_{ik}. In this case, the extinction probability is equal to $1 - q_{ik}$. For all $k \in \underline{S}$, let us consider the Boolean variable y_k which takes the value 1 if and only if the extinction probability of species s_k in the reserve considered is less than or equal to the threshold value $1 - \rho_k$, and this regardless of the values taken by the survival probabilities – in the set of possible values. The constraint below forces variable y_k, $k \in \underline{S}$ – which we are seeking to maximize – to take the right value:

$$\max_{t_{ik} \in \{0,1\}\ (i \in \underline{Z}_k)} \left\{ \prod_{i \in \underline{Z}_k} [1 - (q_{ik} - t_{ik}\,\delta_{ik})\,\overline{x}_i] : \sum_{i \in \underline{Z}_k} t_{ik} \le \Gamma_k \right\} \le 1 - \rho_k y_k \qquad \text{(c)}$$

Using the logarithmic function this constraint can also be written:

$$\max_{t_{ik} \in \{0,1\}\ (i \in \underline{Z}_k)} \left\{ \sum_{i \in \underline{Z}_k} \log\,[1 - (q_{ik} - t_{ik}\,\delta_{ik})\,\overline{x}_i] : \sum_{i \in \underline{Z}_k} t_{ik} \le \Gamma_k \right\} \le \log(1 - \rho_k y_k) \qquad \text{(d)}$$

Let us rewrite the objective function to be maximized that appears in constraint (d):

$$\sum_{i \in \underline{Z}_k} \log\,[1 - (q_{ik} - t_{ik}\,\delta_{ik})\overline{x}_i] = \sum_{i \in \underline{Z}_k} \overline{x}_i \log(1 - q_{ik} + t_{ik}\,\delta_{ik}) \qquad (\text{since}\, \overline{x}_i \in \{0,1\})$$

$$= \sum_{i \in \underline{Z}_k} \overline{x}_i \log\,(1 - q_{ik}) + \sum_{i \in \underline{Z}_k} \overline{x}_i t_{ik} [\log(1 - q_{ik} + \delta_{ik}) - \log\,(1 - q_{ik})] \quad (\text{since}\, t_{ik} \in \{0,1\})$$

$$= \sum_{i \in \underline{Z}_k} \overline{x}_i \log\,(1 - q_{ik}) + \sum_{i \in \underline{Z}_k} \overline{x}_i t_{ik} \Delta_{ik} \text{ with } \Delta_{ik} = \log(1 - q_{ik} + \delta_{ik}) - \log(1 - q_{ik}) \qquad \text{(e)}$$

Finally, we can express the constraint allowing the Boolean variable y_k to take the value 1 only if the extinction probability of species s_k, in the reserve defined by $\overline{x}$, is less than or equal to the threshold value, whatever the values taken by the survival probabilities in each zone – in the set of possible values – by the following inequality:

$$\sum_{i \in \underline{Z}_k} \overline{x}_i \log\,(1 - q_{ik}) + \max_{t_{ik} \in \{0,1\}\ (i \in \underline{Z}_k)} \left\{ \sum_{i \in \underline{Z}_k} \overline{x}_i t_{ik} \Delta_{ik} : \sum_{i \in \underline{Z}_k} t_{ik} \le \Gamma_k \right\} \le \log(1 - \rho_k y_k) \qquad \text{(f)}$$

In the maximization problem that appears in constraint (f), the integrality constraints $t_{ik} \in \{0,1\}$, $i \in \underline{Z}_k$, can be relaxed, *i.e.*, replaced by the constraints $0 \le t_{ik} \le 1$, $i \in \underline{Z}_k$. In fact, one solution to this maximization problem – with $t_{ik} \in \{0,1\}$ or with $0 \le t_{ik} \le 1$ – is to set to 1 the ν_k variables t_{ik} corresponding to the ν_k highest values of the product $\overline{x}_i \Delta_{ik}$ with $\nu_k = \min\{\Gamma_k, |\,\underline{Z}_k \cap \{\, i : \overline{x}_i = 1\}|\}$. This maximization problem thus becomes a continuous linear program. As we have just seen, this program admits a finite optimal solution. According to linear programming theory, its dual therefore also admits a finite optimal solution; it is written:

$$\min_{\lambda_k \geq 0,\, \mu_{ik} \geq 0\ (i \in \underline{Z}_k)} \left\{ \sum_{i \in \underline{Z}_k} \mu_{ik} + \Gamma_k \lambda_k : \lambda_k + \mu_{ik} \geq \Delta_{ik} \overline{x}_i \,(i \in \underline{Z}_k) \right\} \qquad \text{(g)}$$

where λ_k is the non-negative dual variable associated with the constraint $\sum_{i \in \underline{Z}_k} t_{ik} \leq \Gamma_k$, and μ_{ik} is the non-negative dual variable associated with the constraint $t_{ik} \leq 1$. Since, by duality, the optimal value of the maximization problem that appears in the constraint (f) is equal to the optimal value of the minimization problem (g), the constraint (f) can be rewritten

$$\sum_{i \in \underline{Z}_k} \overline{x}_i \log(1 - q_{ik})$$
$$+ \min_{\lambda_k \geq 0, \mu_{ik} \geq 0\ (i \in \underline{Z}_k)} \left\{ \sum_{i \in \underline{Z}_k} \mu_{ik} + \Gamma_k \lambda_k : \lambda_k + \mu_{ik} \geq \Delta_{ik} \overline{x}_i \ \ (i \in \underline{Z}_k) \right\} \leq \log(1 - \rho_k y_k).$$

Noting that since y_k is a Boolean variable $\log(1 - \rho_k y_k) = y_k \log(1 - \rho_k)$, we can now formulate the problem of determining an optimal robust reserve as the mixed-integer linear program $\mathrm{P}_{7.15}$. Remember that, here, an optimal robust reserve is a reserve that respects a certain budget and guarantees a given survival probability – in the reserve – to the greatest possible number of species, whatever the values taken by the survival probabilities in the set of possible values.

$$\mathrm{P}_{7.15}: \begin{cases} \max \sum_{k \in \underline{S}} y_k & & \\ \text{s.t.} \left| \begin{array}{lll} \sum_{i \in \underline{Z}} c_i x_i \leq B & & (7.15.1) \\ \sum_{i \in \underline{Z}_k} x_i \log(1 - q_{ik}) + \sum_{i \in \underline{Z}_k} \mu_{ik} & & \\ \quad + \Gamma_k \lambda_k \leq y_k \log(1 - \rho_k) & k \in \underline{S} & (7.15.2) \\ \lambda_k + \mu_{ik} \geq \Delta_{ik} x_i & k \in \underline{S}, i \in \underline{Z}_k & (7.15.3) \\ \mu_{ik} \geq 0 & k \in \underline{S}, i \in \underline{Z}_k & (7.15.4) \\ y_k \in \{0, 1\} & k \in \underline{S} & (7.15.5) \\ \lambda_k \geq 0 & k \in \underline{S} & (7.15.6) \\ x_i \in \{0, 1\} & i \in \underline{Z} & (7.15.7) \end{array} \right. \end{cases}$$

Let $\overline{x}_i$, $i \in \underline{Z}$, be the value of variable x_i in an optimal solution of $\mathrm{P}_{7.15}$. The optimal reserve includes zones z_i such that $\overline{x}_i = 1$, its cost is equal to $\sum_{i \in \underline{Z}} c_i \overline{x}_i$ and the worst case is defined by the worst-case survival probabilities in the v_k zones of the reserve corresponding to the v_k highest values of the expression $\delta_{ik}/(1 - q_{ik})$ with $v_k = \min\{\Gamma_k, |\,\underline{Z}_k \cap \{\, i : \overline{x}_i = 1\}|\}$. In the particular case defined by $\Gamma_k =$

$\lfloor \eta_k |Z_k|/100 \rfloor$, for all k, and $\delta_{ik} = r_k q_{ik}/100$ for all i and for all k, program $\mathrm{P}_{7.15}$ becomes program $\mathrm{P}_{7.16}$.

$$\mathrm{P}_{7.16} : \begin{cases} \max \sum_{k \in \underline{S}} y_k \\ \text{s.t.} \left| \begin{array}{lll} \sum_{i \in \underline{Z}} c_i x_i \le B & & (7.16.1) \\ \sum_{i \in \underline{Z}_k} x_i \log(1 - q_{ik}) + \sum_{i \in \underline{Z}_k} \mu_{ik} + \left\lfloor \frac{\eta_k |Z_k|}{100} \right\rfloor \lambda_k & & \\ \qquad \le y_k \log(1 - \rho_k) & k \in \underline{S} & (7.16.2) \\ \lambda_k + \mu_{ik} \ge \{\log[1 - q_{ik}(1 - \frac{r_k}{100})] - \log(1 - q_{ik})\}\, x_i & k \in \underline{S}, i \in \underline{Z}_k & (7.16.3) \\ \mu_{ik} \ge 0 & k \in \underline{S}, i \in \underline{Z}_k & (7.16.4) \\ \lambda_k \ge 0, y_k \in \{0, 1\} & k \in \underline{S} & (7.16.5) \\ x_i \in \{0, 1\} & i \in \underline{Z} & (7.16.6) \end{array} \right. \end{cases}$$

7.6.4 A Variant of the Previous Problem: A Robust, Least-Cost Reserve that Ensures a Given Survival Probability for All Species Considered

A variant of the problem studied in the previous section is to determine a least-cost reserve ensuring, for each species of S, a survival probability greater than or equal to a certain threshold value, regardless of the values taken by the survival probabilities in each zone, in the set of possible values. This reserve can be qualified as an optimal robust reserve in the sense that there are no other reserves, of lower cost, ensuring that all species have a survival probability greater than or equal to the threshold value, regardless of the values taken by the survival probabilities in each zone, in the set of possible values. This reserve can be determined by the mixed-integer linear program $\mathrm{P}_{7.17}$.

$$\mathrm{P}_{7.17} : \begin{cases} \min \sum_{i \in \underline{Z}} c_i x_i \\ \text{s.t.} \left| \begin{array}{lll} \sum_{i \in \underline{Z}_k} x_i \log(1 - q_{ik}) + \sum_{i \in \underline{Z}_k} \mu_{ik} & & \\ \qquad + \Gamma_k \lambda_k \le \log(1 - \rho_k) & k \in \underline{S} & (7.17.1) \\ \lambda_k + \mu_{ik} \ge \Delta_{ik} x_i & k \in \underline{S}, i \in \underline{Z}_k & (7.17.2) \\ \mu_{ik} \ge 0 & k \in \underline{S}, i \in \underline{Z}_k & (7.17.3) \\ \lambda_k \ge 0 & k \in \underline{S} & (7.17.4) \\ x_i \in \{0, 1\} & i \in \underline{Z} & (7.17.5) \end{array} \right. \end{cases}$$

Let $\overline{x}_i$, $i \in \underline{Z}$, be the value of variable x_i in an optimal solution of $\mathrm{P}_{7.17}$. The optimal reserve includes zones z_i such that $\overline{x}_i = 1$, its cost is equal to $\sum_{i \in \underline{Z}} c_i \overline{x}_i$ and the worst case is defined by the worst-case survival probabilities in the v_k zones of the reserve corresponding to the v_k highest values of the expression $\delta_{ik}/(1-q_{ik})$ with $v_k = \min\{\Gamma_k, |\,\underline{Z}_k \cap \{\, i : \overline{x}_i = 1\}|\}$.

In the particular case where $\Gamma_k = \lfloor \eta_k\, |Z_k|/100 \rfloor$, for all k, and $\delta_{ik} = r_k q_{ik}/100$ for all i and for all k, program $\mathrm{P}_{7.17}$ becomes program $\mathrm{P}_{7.18}$.

$$
\mathrm{P}_{7.18}:\left\{
\begin{array}{lll}
\min \sum\limits_{i\in\underline{Z}} c_i x_i & & \\
\text{s.t.} \left|
\begin{array}{l}
\sum\limits_{i\,\in\,\underline{Z}_k} x_i \log(1-q_{ik}) + \sum\limits_{i\in\underline{Z}_k} \mu_{ik} + \left\lfloor \frac{\eta_k\,|Z_k|}{100} \right\rfloor \lambda_k \\
\qquad \le \log(1-\rho_k) \\
\mu_{ik} + \lambda_k \ge \{\log\left[1 - q_{ik}\left(1 - \frac{r_k}{100}\right)\right] - \log(1-q_{ik})\} x_i \\
\mu_{ik} \ge 0 \\
\lambda_k \ge 0 \\
x_i \in \{0,1\}
\end{array}\right.
&
\begin{array}{l}
\\
k \in \underline{S}, i \in \underline{Z}_k \\
k \in \underline{S}, i \in \underline{Z}_k \\
k \in \underline{S}, i \in \underline{Z}_k \\
k \in \underline{S} \\
i \in \underline{Z}
\end{array}
&
\begin{array}{l}
\\
(7.18.1) \\
(7.18.2) \\
(7.18.3) \\
(7.18.4) \\
(7.18.5)
\end{array}
\end{array}\right.
$$

7.6.5 Computational Experiments

7.6.5.1 Landscape Represented by a 8 × 8 Grid

In this section, we illustrate the results of the previous sections on a set of hypothetical zones represented by a grid of 8×8 square and identical zones. In this example, 10 species are concerned. The data are presented in figure 7.9. The zones are designated by z_{ij} where i represents the row index and j represents the column index of these zones. On each zone is indicated (1) the list of species whose nominal survival probability is positive if the zone is protected and (2) the corresponding nominal survival probability, denoted by q_{ijk} for species s_k in zone z_{ij}. The cost associated with protecting each zone is indicated in the lower right corner of the corresponding zone. All the survival probabilities in unprotected zones are assumed to be zero. To facilitate the analysis of this example, we give in table 7.8 the composition of the sets Z_k, for all $k \in \underline{S}$. We consider the problem discussed in section 7.6.4 and recalled below.

Problem. Determine a least-cost reserve that guarantees a survival probability greater than or equal to 0.8 and then 0.9 for all species considered, which corresponds to $\rho_k = 0.8$ then $\rho_k = 0.9$, for all k, regardless of the values taken by the survival probabilities in the set of possible values.

Definition of the uncertainty affecting the species survival probabilities in each zone. The level of uncertainty, Γ_k, is defined as a percentage, η_k, of the number of zones of Z_k. Remember that Z_k refers to the subset of zone of Z whose protection provides a non-zero nominal survival probability – in the corresponding zone – of species s_k. We therefore have $\Gamma_k = \lfloor \eta_k |Z_k|/100 \rfloor$ where $\lfloor a \rfloor$ refers to the integer part of a. For example, figure 7.9 shows non-zero nominal survival probabilities of species s_3 in zones z_{12}, z_{26}, z_{34}, z_{55}, z_{65}, z_{75}, z_{83}, and z_{85}. So we have $Z_3 = \{z_{12}, z_{26}, z_{34}, z_{55}, z_{65}, z_{75}, z_{83}, z_{85}\}$ and $|Z_3| = 8$. If $\eta_3 = 30$, then the maximal number

	1	2	3	4	5	6	7	8
1	s_1 : 0.7 s_8 : 0.7 s_{10} : 0.3 9	s_3 : 0.5 2	3	s_5 : 0.4 s_8 : 0.6 3	s_6 : 0.6 s_{10} : 0.7 3	s_1 : 0.6 3	s_2 : 0.4 s_4 : 0.4 9	s_2 : 0.9 s_7 : 0.6 s_{10} : 0.8 5
2	4	s_9 : 0.8 1	s_9 : 0.5 1	s_5 : 0.4 9	s_1 : 0.6 5	s_3 : 0.6 s_7 : 0.8 6	3	4
3	7	8	10	s_3 : 0.6 s_{10} : 0.5 6	4	s_9 : 0.9 8	s_7 : 0.5 10	4
4	3	s_9 : 0.6 3	s_2 : 0.6 8	s_4 : 0.3 7	6	s_9 : 0.7 5	9	9
5	1	10	2	s_4 : 0.9 7	s_3 : 0.8 s_7 : 0.5 3	s_7 : 0.5 s_8 : 0.8 6	s_5 : 0.4 1	6
6	s_2 : 0.6 3	s_4 : 0.8 s_9 : 0.7 4	6	2	s_3 : 0.6 10	6	s_4 : 0.7 1	s_5 : 0.6 7
7	s_2 : 0.4 8	2	s_8 : 0.3 5	1	s_3 : 0.8 s_4 : 0.4 3	s_8 : 0.8 6	8	s_6 : 0.5 8
8	5	10	s_1 : 0.5 s_3 : 0.7 5	s_2 : 0.5 s_4 : 0.5 s_6 : 0.4 5	s_3 : 0.6 4	s_2 : 0.7 7	s_5 : 0.7 4	s_6 : 0.3 7

FIG. 7.9 – A set of 64 candidate zones for protection represented by a grid of 8×8 square and identical zones. 10 species s_1, s_2,..., s_{10} are concerned. The corresponding nominal survival probabilities, q_{ik}, and the protection costs are indicated in each zone. Consider, for example, zone z_{56}. Species s_7 and s_8 are concerned. The nominal survival probabilities of these 2 species in this zone, if protected, are 0.5 and 0.8, respectively. The cost of protecting this zone is equal to 6.

of zones where the survival probability of this species may differ from its nominal value is equal to $\Gamma_3 = \lfloor 30 \times 8/100 \rfloor = 2$. In other words, the survival probabilities of s_3 may differ from their nominal value in up to 2 of the 8 zones of Z_3. We therefore admit that, among these 8 probabilities, 2 (at most) may be erroneous but we do not know which ones. Note that since we have to consider the "worst-case", these two probabilities will be equal to $q_{ik} - \delta_{ik}$ instead of q_{ik}. In these experiments we consider that $\eta_k = \eta$, for all k, and we consider 4 different values of η: 0, 20, 30, and 100.

TAB. 7.8 – List of zones whose protection provides species s_k, $k = 1,\ldots,10$, with a positive nominal survival probability.

s_k	Z_k
s_1	z_{11} z_{16} z_{25} z_{83}
s_2	z_{17} z_{18} z_{43} z_{61} z_{71} z_{84} z_{86}
s_3	z_{12} z_{26} z_{34} z_{55} z_{65} z_{75} z_{83} z_{85}
s_4	z_{17} z_{44} z_{54} z_{62} z_{67} z_{75} z_{84}
s_5	z_{14} z_{24} z_{57} z_{68} z_{87}
s_6	z_{15} z_{78} z_{84} z_{88}
s_7	z_{18} z_{26} z_{37} z_{55} z_{56}
s_8	z_{11} z_{14} z_{56} z_{73} z_{76}
s_9	z_{22} z_{23} z_{36} z_{42} z_{46} z_{62}
s_{10}	z_{11} z_{15} z_{18} z_{34}

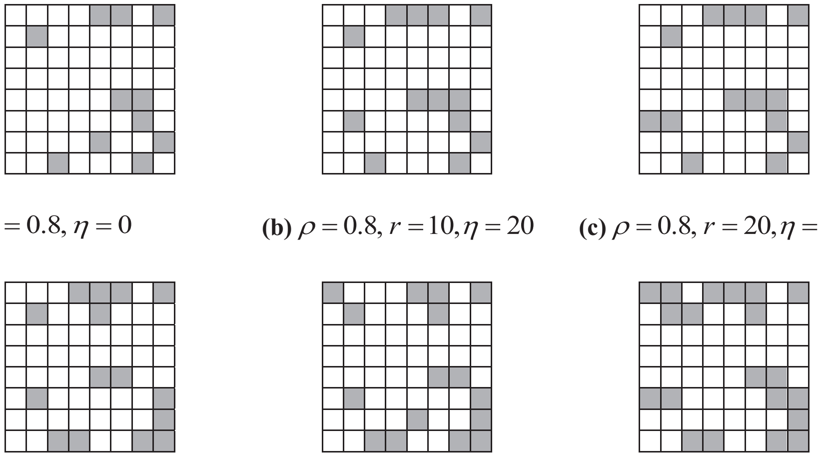
TAB. 7.9 – Results for the example in figure 7.9 when the threshold of survival probability required for each species, ρ, is equal to 0.8 or 0.9, and for different values of η and r.

ρ	r	η	Cost of the optimal robust reserve	Number of zones in the reserve	Associated figure
0.8	10	0	40	11	Figure 7.10a
		20	47	13	Figure 7.10b
		30	51	15	
		100	53	14	
	20	0	40	11	
		20	50	14	Figure 7.10c
		30	53	14	
		100	70	16	
0.9	10	0	69	15	Figure 7.10d
		20	77	16	Figure 7.10e
		30	80	17	
		100	–	–	
	20	0	69	15	
		20	84	20	Figure 7.10f
		30	–	–	
		100	–	–	

–: No feasible solution.

We also assume that $\delta_{ijk} = r_k\, q_{ijk}/100$ with $r_k = r$, for all k, and we consider two values of r: 10 and 20.

The results obtained are presented in table 7.9. Some optimal robust reserves, corresponding to the instances of this table, are presented in figure 7.10.



(a) $\rho = 0.8, \eta = 0$ **(b)** $\rho = 0.8, r = 10, \eta = 20$ **(c)** $\rho = 0.8, r = 20, \eta = 20$

(d) $\rho = 0.9, \eta = 0$ **(e)** $\rho = 0.9, r = 10, \eta = 20$ **(f)** $\rho = 0.9, r = 20, \eta = 20$

FIG. 7.10 – Optimal robust reserves for some instances of table 7.9.

TAB. 7.10 – Details of the results corresponding to figure 7.10b.

Species (s_k)	$\vert Z_k\vert$	$\left\lfloor \frac{\eta\,\vert Z_k\vert}{100} \right\rfloor$	$R^* \cap Z_k$	Zone where the survival probability of species s_k takes its worst value, *i.e.*, $q_{ijk} - r\, q_{ijk}/100$	$P_k(R^*)$	$\tilde{P}_k(R^*)$
s_1	4	0	z_{16} z_{83}	–	0.8	0.8
s_2	7	1	z_{18}	z_{18}: 0.9→ 0.81	0.81	0.9
s_3	8	1	z_{55} z_{83}	z_{55}: 0.8→ 0.72	0.916	0.94
s_4	7	1	z_{62} z_{67}	z_{62}: 0.8→ 0.72	0.916	0.94
s_5	5	1	z_{14} z_{57} z_{87}	z_{87}: 0.7→ 0.63	0.8668	0.892
s_6	4	0	z_{15} z_{78}	–	0.8	0.8
s_7	5	1	z_{18} z_{55} z_{56}	z_{18}: 0.6→ 0.54	0.885	0.9
s_8	5	1	z_{14} z_{56}	z_{56}: 0.8→ 0.72	0.888	0.92
s_9	6	1	z_{22} z_{62}	z_{22}: 0.8→ 0.72	0.916	0.94
s_{10}	4	0	z_{15} z_{18}	–	0.94	0.94

The detailed results corresponding to figure 7.10b are presented in table 7.10. Given the value of η, there can be no errors for species s_1, s_6, and s_{10}. For the other species, errors may only concern one zone. R^* refers to the set of zones selected to form the optimal robust reserve, $P_k(R^*)$, the survival probability of species s_k in that reserve in the worst case and $\tilde{P}_k(R^*)$, the survival probability of species s_k in that reserve, calculated with the nominal survival probabilities in each zone.

Tab. 7.11 – Survival probabilities of the species in the reserve of figure 7.10a. (a) No uncertainty about the survival probabilities, q_{ijk}. (b) Results in the worst case and when the uncertainty is defined by $\eta = 20$ and $r = 20$; probabilities that differ from their nominal values are in bold.

(a)		(b)		
s_k	Species survival probability in the reserve	s_k	$\left\lfloor \frac{\eta\|Z_k\|}{100} \right\rfloor$	Species survival probability in the reserve, in the worst case and when $\eta = 20$ and $r = 20$
s_1	1−(1–0.6)(1–0.5) = 0.8	s_1	0	1−(1–0.6)(1–0.5) = 0.8
s_2	1−(1–0.9) = 0.9	s_2	1	1−(1–**0.72**) = 0.72
s_3	1−(1–0.8)(1–0.7) = 0.94	s_3	1	1−(1–**0.64**)(1–0.7) = 0.892
s_4	1−(1–0.7)(1–0.4) = 0.82	s_4	1	1−(1–**0.56**)(1–0.4) = 0.736
s_5	1−(1–0.4)(1–0.7) = 0.82	s_5	1	1−(1–0.4)(1–**0.56**) = 0.736
s_6	1−(1–0.6)(1–0.5) = 0.8	s_6	0	1−(1–0.6)(1–0.5) = 0.8
s_7	1−(1–0.6)(1–0.5) = 0.8	s_7	1	1−(1–**0.48**)(1–0.5) = 0.74
s_8	1−(1–0.8) = 0.8	s_8	1	1−(1–**0.64**) = 0.64
s_9	1−(1–0.8) = 0.8	s_9	1	1−(1–**0.64**) = 0.64
s_{10}	1−(1–0.7)(1–0.8) = 0.94	s_{10}	0	1−(1–0.7)(1–0.8) = 0.94

Now let us consider the instance where the threshold value chosen for each species s_k is always defined by $\rho = 0.8$ but without taking into account the uncertainty about the species survival probabilities in each zone. The optimal reserve obtained costs 40 units and is shown in figure 7.10a. The survival probabilities of each species in this reserve, calculated from the nominal values of the survival probabilities in each zone, are given in table 7.11a.

Let us always consider the reserve in figure 7.10a but now take into account the uncertainty about the survival probabilities in each zone when $\eta = 20$ and $r = 20$. This means that in at most $\lfloor 20|Z_k|/100 \rfloor$ zones, the true values of the survival probabilities of species s_k are only equal to 80% of their nominal value, q_{ijk}. In this context, the worst case corresponds to the species survival probabilities in the reserve given in table 7.11b. As expected, all these probabilities are less than or equal to those in table 7.11a, but the survival probabilities of species s_2, s_4, s_5, s_7, s_8, and s_9 fall below the fixed threshold value, 0.8. The reserve under consideration is therefore not at all robust. Remember that in this context of uncertainty, the optimal robust solution is given by the reserve in figure 7.10c, the cost of which is equal to 50. Thus, we can say that, for this hypothetical example, protection against uncertainty defined by $\eta = 20$ and $r = 20$, increases the cost of the optimal reserve by 25%.

The problem discussed in this section 7.6.5.1 can be addressed in a slightly different way. We can look, for example, at the maximal error that can be made in estimating the probability q_{ijk} in such a way that there is a reserve that satisfies the required performance – survival probabilities of each species greater than or equal to a certain threshold value. We will see that the cost of the optimal robust reserve associated with this problem increases with uncertainty on q_{ijk} until there is no longer a reserve that meets the required performance.

TAB. 7.12 – Cost of an optimal robust reserve for the landscape in figure 7.9 when $\rho = 0.9$, $\eta = 30$, and for different values of r. For example, if r is between 1 and 6, the optimal robust reserve costs 75 units and has 16 zones when $r \in \{1, 4, 5, 6\}$ and 17 zones when $r \in \{2, 3\}$.

r	Cost of an optimal robust reserve	Number of zones in the reserve for the different values of r
$r = 0$	69	15
$r \in \{1, 2, 3, 4, 5, 6\}$	75	16, 17, 17, 16, 16, 16
$r \in \{7, 8, 9\}$	77	16, 16, 16
$r \in \{10, 11, 12\}$	80	17, 17, 17
$r \geq 13$	No feasible robust reserve	–

Let us look again at the landscape described in figure 7.9, set ρ to 0.9, η to 30, and look for the highest value of r for which there is a robust reserve, *i.e.*, a reserve that ensures that all the species have a survival probability in the reserve that is greater than or equal to ρ in the worst case. This problem can be directly formulated by replacing in program $P_{7.18}$ the economic function with the expression r to be maximized – in this case, r becomes a variable. However, the optimization problem obtained is difficult because the constraints 7.18.2 become non-linear: $\mu_{ik} + \lambda_k \geq \left\{\log\left[1 - q_{ik}\left(1 - \frac{r}{100}\right)\right] - \log(1 - q_{ik})\right\}x_i$. Note that in this case r_k does not depend on k and is therefore denoted r. Another way to determine this limit value of r is to solve $P_{7.18}$, iteratively, by gradually increasing the value of r until there is no longer any feasible reserve. In this case, the survival probability, in any reserve, of at least one species falls below the desired threshold value, in the worst case. Table 7.12 presents the results obtained by this approach: there are robust reserves as long as r is less than or equal to 12 and there are no more such reserves if r is greater than or equal to 13.

Similarly, it may be interesting when ρ and r are fixed to determine the largest value of η for which a robust reserve exists. Here again, the optimization problem is difficult since it consists in solving $P_{7.18}$ after having replaced the economic function by variable η to be maximized. When r and ρ are fixed, all the constraints of $P_{7.18}$ are linear except constraints 7.18.1 which contain the term $\lfloor \eta\,|Z_k|/100 \rfloor \lambda_k$. Here η_k does not depend on k and is therefore denoted η. As before, one way to determine this limit value is to solve $P_{7.18}$ iteratively by gradually increasing the value of η until there is no longer any feasible robust reserve. In this case, the survival probability of at least one species falls below the required threshold, in the worst case, for any reserve. Table 7.13 presents the results obtained by this approach when $\rho = 0.9$ and $r = 20$. It can be seen that, under these conditions, there are robust reserves as long as $\eta \leq 24$ and that there are no more such reserves when $\eta \geq 25$.

7.6.5.2 Large-Sized Instances

In this section we present the results obtained regarding the resolution of $P_{7.18}$ on large-sized instances. The landscapes studied are characterized by the following 4 parameters: the number of zones, the cost of protection of each zone, the number of

TAB. 7.13 – Cost of an optimal robust reserve for the landscape in figure 7.9 when $\rho = 0.9$ and $r = 20$, and for different values of η.

η	Cost of an optimal robust reserve	Number of zones in the reserve
$\eta \in \{0,...,12\}$	69	15,…,15
$\eta \in \{13, 14\}$	70	17, 17
$\eta \in \{15, 16\}$	75	18, 18
$\eta \in \{17, 18, 19\}$	76	19, 19, 19
$\eta \in \{20, 21, 22, 23, 24\}$	84	20, 20, 20, 20, 20
$\eta \geq 25$	No feasible robust reserve	

species concerned and the nominal survival probability of these species in each zone. In these experiments, 50 species are present in the studied landscape and we considered two cases for each of the other 3 parameters, thus obtaining 8 types of landscapes. The landscape is represented by a grid of 15 × 15 identical square zones or a grid of 20 × 20 identical square zones; the costs of protecting the zones are drawn at random, uniformly, between 1 and 10 or between 1 and 20; finally, in a first case, each species appears randomly in 5% of the zones and the nominal survival probabilities in the zones are drawn at random, uniformly, from the set $\{0.5, 0.6, 0.7, 0.8\}$; in a second case, each species appears randomly in 4% of the zones and the nominal survival probabilities in the zones are drawn at random, uniformly, from the set $\{0.7, 0.8\}$. The characteristics of these 8 landscape types are summarized in table 7.14. When a species does not appear in a zone, its survival probability in that zone is zero.

For each of the 8 types of landscape we considered 5 different landscapes by modifying the germ of the random number generator thus obtaining 40 different landscapes.

Now let us see how the values of Γ_k, ρ_k, and δ_{ijk} are defined. As in the experiments in section 7.6.5.1, $\Gamma_k = \lfloor \eta\,|Z_k|/100 \rfloor$, $\rho_k = \rho$, $k \in \underline{S}$, and $\delta_{ijk} = r\,q_{ijk}/100$, $(i, j) \in \underline{Z}$, $k \in \underline{S}$. For each of the 40 landscapes, we studied solutions for the following values of the different parameters: $\rho \in \{0.85, 0.90, 0.95\}$, $\eta \in \{10, 20\}$, and $r \in \{20, 30, 100\}$. For each of the 40 landscapes we also studied the case where there is no uncertainty by solving $\mathrm{P}_{7.18}$ with $\rho \in \{0.85, 090, 0.95\}$ and $\eta = 0$. We thus resolved a total of 840 instances of $\mathrm{P}_{7.18}$.

All the instances were resolved quickly except some instances of size 20 × 20 when the threshold value is equal to 0.95. In this case, the resolution of several instances requires a few hundred seconds, the most difficult instances corresponding to $\eta = 20$ and $r = 20$ or 30. All the instances corresponding to landscapes of size 20 × 20 have a feasible solution. For the instances corresponding to landscapes of size 15 × 15, there is no feasible solution for one of the 5 germs of the random number generator when $\rho = 0.95$ and when each species appears randomly in 4% of the zones with a nominal survival probability drawn at random from the set $\{0.7, 0.8\}$ – landscapes of types 2 and 4. The other instances for which there is no feasible solution are presented in table 7.15.

TAB. 7.14 – Description of the 8 types of landscape considered in the experiments. In all the cases, 50 species are concerned.

Type	Dimension of the grid ($n \times n$)	Set of values in which the costs, c_{ij}, are generated	Probability of presence of each species in each zone	Set of values in which the nominal survival probabilities, q_{ijk}, are generated
1	15×15	$\{1, 2, ..., 10\}$	$5(n \times n)/100$	$\{0.5, 0.6, 0.7, 0.8\}$
2	15×15	$\{1, 2, ..., 10\}$	$4(n \times n)/100$	$\{0.7, 0.8\}$
3	15×15	$\{1, 2, ..., 20\}$	$5(n \times n)/100$	$\{0.5, 0.6, 0.7, 0.8\}$
4	15×15	$\{1, 2, ..., 20\}$	$4(n \times n)/100$	$\{0.7, 0.8\}$
5	20×20	$\{1, 2, ..., 10\}$	$5(n \times n)/100$	$\{0.5, 0.6, 0.7, 0.8\}$
6	20×20	$\{1, 2, ..., 10\}$	$4(n \times n)/100$	$\{0.7, 0.8\}$
7	20×20	$\{1, 2, ..., 20\}$	$5(n \times n)/100$	$\{0.5, 0.6, 0.7, 0.8\}$
8	20×20	$\{1, 2, ..., 20\}$	$4(n \times n)/100$	$\{0.7, 0.8\}$

TAB. 7.15 – Instances without feasible solutions.

Landscape type	ρ	r	η	Number of instances without feasible solutions (out of the 5 considered)
2	0.85	20	100	1
4	0.85	20	100	1
2	0.90	10	100	1
2	0.90	20	100	1
4	0.90	10	100	1
4	0.90	20	100	1
1	0.95	20	100	1
2	0.95	20	100	1
4	0.95	20	100	2

Experiments have shown that the cost of protection against uncertainty is often high. Table 7.17 summarizes the average percentage increase in the cost of an optimal robust reserve – taking into account uncertainty hypotheses – compared to the cost of an optimal reserve when there is no uncertainty, for all the instances of size 20 × 20 – landscapes of types 5, 6, 7, and 8. The average increase is calculated on the 5 instances associated with the 5 values of the generator germ. Consider, for example, the landscapes of type 8 when the threshold value of the survival probability, ρ, is equal to 0.90. The results obtained for $\eta = 30$ and $r = 10$ are given in table 7.16. In this case, the average increase in the cost due to protection against uncertainty is about 55%.

Let us look at table 7.17 when the threshold value of the survival probability is equal to 0.85. We see on this table that, when $r = 10$ and $\eta = 30$ or 100, there is no

TAB. 7.16 – Increase in the costs of an optimal reserve, due to protection against uncertainty, for 5 landscapes of type 8, when $\eta = 30$, $r = 10$, and $\rho = 0.90$.

No. of the instance	Minimal cost when there is no uncertainty	Minimal cost when uncertainty is defined by $\eta = 30$ and $r = 10$	Cost increase (%)
1	119	188	57.98
2	90	147	63.33
3	114	180	57.89
4	124	183	47.58
5	99	148	49.49
Av			55.26

TAB. 7.17 – Cost increases due to uncertainty for the 4 types of landscape of size 20×20 and for 3 values of the parameter ρ: 0.85, 0.90, and 0.95.

		Percentage increase in cost due to uncertainty			
Survival probability to be ensured for each species (ρ)	Definition of the considered uncertainty	Landscape of type 5	Landscape of type 6	Landscape of type 7	Landscape of type 8
0.85	$r = 10, \eta = 30$	+ 21.3	+ 0	+ 20.1	+ 0
	$r = 10, \eta = 100$	+ 21.3	+ 0	+ 20.1	+ 0
	$r = 20, \eta = 30$	+ 51.6	+ 46.5	+ 50.6	+ 55.3
	$r = 20, \eta = 100$	+ 51.6	+ 49.3	+ 50.6	+ 56.5
0.90	$r = 10, \eta = 30$	+ 32.8	+ 46.8	+ 32.1	+ 55.3
	$r = 10, \eta = 100$	+ 32.8	+ 49.3	+ 32.1	+ 56.5
	$r = 20, \eta = 30$	+ 61.1	+ 72.5	+ 63.4	+ 76.4
	$r = 20, \eta = 100$	+ 61.1	+ 72.5	+ 64.5	+ 77.0
0.95	$r = 10, \eta = 30$	+ 28.1	+ 18.2	+ 31.1	+ 17.2
	$r = 10, \eta = 100$	+ 28.1	+ 19.2	+ 31.2	+ 17.4
	$r = 20, \eta = 30$	+ 60.8	+ 57.5	+ 63.0	+ 58.8
	$r = 20, \eta = 100$	+ 62.1	+ 63.4	+ 67.0	+ 66.5

increase in cost for the landscapes of types 6 and 8. However, for the landscapes of types 5 and 7 and for the same values of η and r, the average cost increases by about 20%. The largest average increase, again for this same threshold value of survival probability, occurs for the landscapes of type 8 when $r = 20$ and η=100 and is about 57%. Regardless of the landscape considered, the largest average increase in the table occurs for the instances defined by $\rho = 0.90$, $r = 20$, and η=100. Indeed, in this case, this average increase in cost for the 4 types of landscape is approximately between 61 and 77%.

References and Further Reading

Arponen A., Heikkinen R.K., Thomas C.D., Moilanen A. (2005) The value of biodiversity in reserve selection: representation, species weighting, and benefit functions, *Conserv. Biol.* **19**, 2009.

Arthur J.L., Haight R.G., Montgomery C.A., Polasky S. (2002) Analysis of the threshold and expected coverage approaches to the probabilistic reserve site selection problem, *Environ. Model. Assess.* **7**, 81.

Arthur J.L., Camm J., Haight R.G., Montgomery C.A., Polasky S. (2004) Weighing conservation objectives: maximum expected coverage versus endangered species protection, *Ecol. Appl.* **14**, 1936.

Bertsimas D., Sim M. (2004) The price of robustness, *Oper. Res.* **52**, 35.

Billionnet A. (2011) Solving the probabilistic reserve selection problem, *Ecol. Model.* **222**, 546.

Billionnet A. (2015) Designing robust nature reserves under uncertain survival probabilities, *Environ. Model. Assess.* **20**, 383.

Billionnet A. (2017) *Programmation discrete.* Dunod.

Cabeza M., Araújo M.B., Wilson R.J., Thomas C.D., Cowley M.J.R., Moilanen A. (2004) Combining probabilities of occurrence with spatial reserve design, *J. Appl. Ecol.* **41**, 252.

Camm J.D., Norman S.K., Polasky S., Solow A.R. (2002) Nature reserve site selection to maximize expected species covered, *Oper. Res.* **50**, 946.

Chen D.-S., Batson R.G., Dang Y. (2010) *Applied integer programming, modeling and solution.* John Wiley & Sons.

Cocks K.D., Baird I.A. (1989) Using mathematical programming to address the multiple reserve selection problem: an example from the Eyre Peninsula, South Australia, *Biol. Conserv.* **49**, 113.

Drechsler M. (2005) Probabilistic approaches to scheduling reserve selection, *Biol. Conserv.* **122**, 253.

Fiorella K., Cameron A., Sechrest W., Winfree R., Kremen C. (2010) Methodological considerations in reserve system selection: a case study of Malagasy lemurs, *Biol. Conserv.* **143**, 963.

Fourer R. (1992) A simplex algorithm for piecewise-linear programming III: computational analysis and applications, *Math. Prog.* **53**, 213.

Fourer R., Marsten R.E. (1992) Solving piecewise-linear programs: experiments with a simplex approach, *ORSA J. Comput.* **4**, 16.

Fourer R., Gay D.M., Kernighan B.W. (1993) *AMPL, a modeling language for mathematical programming.* Boyd & Fraser Publishing Company, Danvers, USA.

Gabrel V., Murat C., Thiele A. (2014) Recent advances in robust optimization: an overview, *Eur. J. Oper. Res.* **235**, 471.

Golovin D., Krause A., Gardner B., Converse S.J., Morey S. (2011) Dynamic resource allocation in conservation planning, Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, August 7–11, 2011, San Francisco, California, USA.

Haider Z., Charkhgard H., Kwon C. (2018) A robust optimization approach for solving problems in conservation planning, *Ecol. Model.* **368**, 288.

Haight R.G., ReVelle C.S., Snyder S.A. (2000) An integer optimization approach to a probabilistic reserve site selection problem, *Oper. Res.* **48**, 697.

Haight R.G., Travis L.E. (2008) Reserve design to maximize species persistence, *Environ. Model. Assess.* **13**, 243.

Haight R.G., Snyder S.A. (2009) Integer programming methods for reserve selection and design, *Spatial conservation prioritization* (A. Moilanen, K.A. Wilson, H.P. Possingham, Eds). Oxford University Press.

Margules C.R., Nicholls A.O. (1987) Assessing the conservation value of remnant habitat "islands": mallee patches on the western Eyre Peninsula, South Australia, *Nature conservation: The role of remnants of native vegetation* (D.A. Saunders, G.W. Arnold, A.A. Burbidge, A.J.M. Hop-kins, Eds). Surrey Beatty and Sons, in association with CSIRO and CALM, Sydney, Australia, pp. 89–102.

Margules C.R., Stein J.L. (1989) Patterns in the distributions of species and the selection of nature reserves: an example from Eucalyptus forests in south-eastern New South Wales, *Biol. Conserv.* **50**, 219.

Margules C.R., Nicholls A.O., Usher M.B. (1994) Apparent species turnover, probability of extinction and the selection of nature reserves: a case study of the ingleborough limestone pavements, *Conserv. Biol.* **8**, 398.

Moilanen A., Cabeza M. (2005) Variance and uncertainty in the expected number of occurrences in reserve selection, *Conserv. Biol.* **19**, 1663.

Burgman M., Ben-Haim Y. (2006) Planning for robust reserve networks using uncertainty analysis, *Ecol. Model.* **199**, 115.

Moilanen A., Wintle B.A., Elith J., Burgman M. (2006) Uncertainty analysis for regional-scale reserve selection, *Conserv. Biol.* **20**, 1688.

Moilanen A., Wilson K.A., Possingham H.P., Eds. (2009) *Spatial conservation prioritization.* Oxford University Press.

Nicholls A.O. (1989) How to make biological surveys go further with generalised linear models, *Biol. Conserv.* **50**, 51.

Polasky S., Camm J. D., Solow A.R., Csuti B., White D., Ding R. (2000) Choosing reserve networks with incomplete species information, *Biol. Conserv.* **94**, 1.

Sarkar S., Pappas C., Garson J., Aggarwal A., Cameron S. (2004) Place prioritization for biodiversity conservation using probabilistic surrogate distribution data, *Diver. Distrib.* **10**, 125.

Strange N., Thorsen B.J., Bladt J. (2006) Optimal reserve selection in a dynamic world, *Biol. Conserv.* **131**, 33.

Williams P.H., Araujo M.B. (2000) Using probability of persistence to identify important areas for biodiversity conservation, *Proc. Royal Soc. London* **B 267**, 1959.

Chapter 8

Scenarios

8.1 Introduction

In chapters 1–6, the effect of zones' protection is fully known. In chapter 7, the uncertainty that may exist with regard to the survival of the species, both in protected and unprotected zones, is expressed in terms of probabilities. We also show, in chapter 7, how to take into account, in a certain way, the inevitable uncertainties concerning the values of these probabilities. In this new chapter, we consider another way to take into account the uncertainty about the survival of the species in protected and unprotected zones. For this purpose, we consider that a set of scenarios, $\mathrm{Sc} = \{\mathrm{sc}_1, \mathrm{sc}_2, ..., \mathrm{sc}_p\}$, are possible (see appendix at the end of the book). A scenario is a set of hypotheses on the evolution of factors that can affect the survival of species in protected or unprotected zones. These assumptions may include direct factors such as land use, climate change, pollution, overexploitation or invasive species, and indirect factors such as economic activity, demographic change, and socio-political contexts. Thus, with each set of protected zones is associated a certain protection of the species under consideration and this protection depends on the scenario. We denote by $\underline{\mathrm{Sc}}$ the set of indices of the possible scenarios. As in the previous chapters, $S = \{s_1, s_2, ..., s_m\}$ refers to the set of species, more or less threatened, in which we are interested and $Z = \{z_1, z_2, ..., z_n\}$, the set of zones that we can decide whether or not to protect from a given moment, in order to ensure a certain protection to the species in question and thus increase their chance of survival. $\underline{S}$ and $\underline{Z}$ refer to the set of corresponding indices, respectively. With regard to the survival of the species in protected zones, the following two cases are considered: in the first case, it is assumed that, for any scenario sc_ω, we know the zones whose protection ensures the survival of species s_k, and this for all $k \in \underline{S}$, if scenario sc_ω is realized. This set is denoted by Z_k^ω and the corresponding set of indices is denoted by $\underline{Z}_k^\omega$. In other words, to ensure the survival of species s_k if scenario sc_ω is realized, it is necessary and sufficient that at least one zone of Z_k^ω be protected. As we have generally done in the

DOI: 10.1051/978-2-7598-2521-9.c008

previous chapters, we consider here that there is only one level of protection: a zone is protected or not. More precisely, the protection of zone z_i is considered to protect species s_k in the case of scenario sc_ω realization if the population size of species s_k in this zone is greater than or equal to a certain threshold value, depending on the scenario and denoted by v_{ik}^ω. In other words, $Z_k^\omega = \{z_i \in Z : n_{ik} \geq v_{ik}^\omega\}$ where n_{ik} refers to the population size of species s_k – at the beginning of the horizon considered – in zone z_i. Given a reserve R, we refer to $\mathrm{Nb}_1^\omega(R)$ as the number of species protected by this reserve if scenario sc_ω occurs. In the second case, it is assumed that, for any scenario sc_ω, we know the minimal population size of species s_k that must be present in the entire reserve – at the beginning of the period considered – for this species to be protected if scenario sc_ω occurs, and this for all $k \in \underline{S}$. This minimal population size is denoted by θ_k^ω and $\mathrm{Nb}_2^\omega(R)$ is referred to as the number of species protected by reserve R if scenario sc_ω occurs. This chapter focuses on the determination of optimal robust reserves, *i.e.,* the determination of reserves that allow a certain objective to be "best" achieved, knowing that several scenarios are possible.

Example 8.1. The instance described in figure 8.1 is considered and it is assumed that two scenarios are possible: $\mathrm{Sc} = \{\mathrm{sc}_\omega : \omega = 1, 2\}$. We consider the two ways – described above – of calculating the number of species protected by a reserve, R, when scenario sc_ω occurs: $\mathrm{Nb}_1^\omega(R)$ and $\mathrm{Nb}_2^\omega(R)$. With regard to the calculation of $\mathrm{Nb}_1^\omega(R)$, the values of v_{ik}^ω, $i \in \{1,\ldots,20\}$, $k \in \{1,\ldots,15\}$, $\omega \in \{1, 2\}$, are such that the list of species that will survive in each protected zone and in each of the two scenarios is given in figure 8.2. With regard to the calculation of $\mathrm{Nb}_2^\omega(R)$, the values of θ_k^1 and θ_k^2 are given in table 8.1. For example, if reserve R is composed of the 5 zones z_2, z_3, z_{10}, z_{11}, and z_{16}, we obtain $\mathrm{Nb}_1^1(R) = 6$ since the 6 species s_3, s_4, s_6, s_7, s_8, and s_{12} will survive in the case of scenario sc_1, $\mathrm{Nb}_1^2(R) = 7$ since the 7 species s_1, s_3, s_6, s_9, s_{10}, s_{11}, and s_{12} will survive in the case of scenario sc_2, $\mathrm{Nb}_2^1(R) = 6$ since the 6 species s_3, s_4, s_7, s_{10}, s_{11}, and s_{12} will survive in the case of scenario sc_1, and $\mathrm{Nb}_2^2(R) = 6$ since the 6 species s_1, s_3, s_9, s_{10}, s_{11}, and s_{12} will survive in the case of scenario sc_2.

In the following sections we examine several problems related to the selection of optimal robust reserves. Such reserves provide the best possible protection for the species under consideration, in the presence of several scenarios and taking into account a given robustness criterion.

8.2 Reserve Protecting All Species Considered Regardless of the Scenario that Occurs

A first question that can be raised is the following: what is the set of zones to be protected, at minimal cost, to protect all species considered, regardless of the scenario that occurs. We first examine the case where the interest in protecting a reserve R, if scenario sc_ω is realized, is assessed by $\mathrm{Nb}_1^\omega(R)$ then the case where this interest is assessed by $\mathrm{Nb}_2^\omega(R)$.

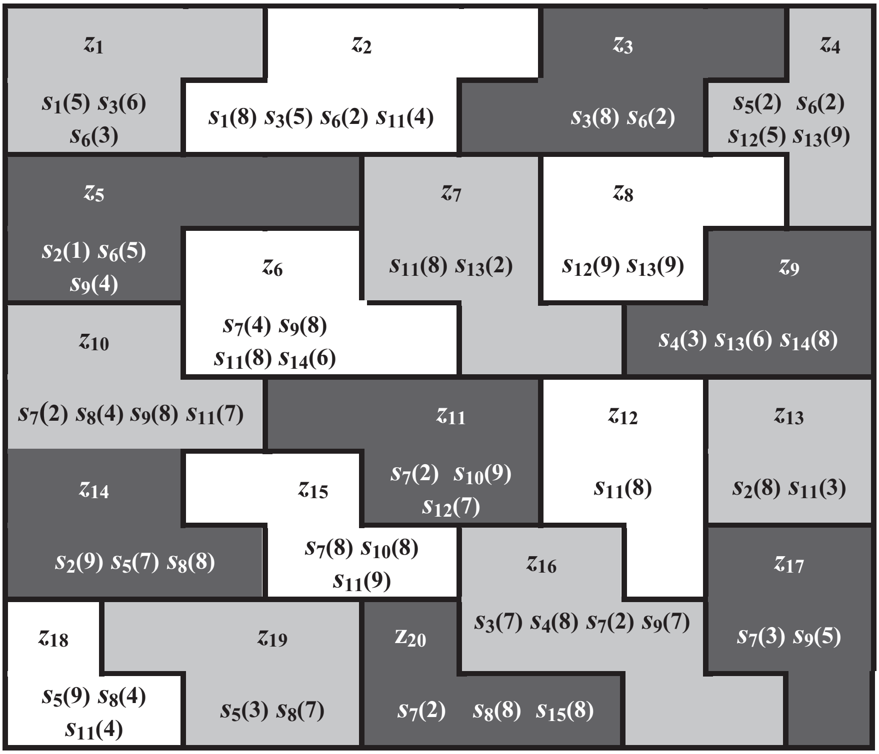


FIG. 8.1 – The 20 zones z_1, z_2,..., z_{20} are candidates for protection and the 15 species s_1, s_2, ..., s_{15} living in these zones are concerned. For each zone, the species present and their population size – in brackets – are indicated. The cost of protecting the white zones is equal to 1 unit, the cost of protecting the light grey zones is equal to 2 units and the cost of protecting the dark grey zones is equal to 4 units. For example, s_7, s_9, s_{11}, and s_{14} are present in zone z_6, their population size is equal to 4, 8, 8, and 6 units, respectively, and the cost of protecting this zone is equal to 1 unit.

8.2.1 Case Where the Number of Species Protected by a Reserve, R, if Scenario sc_ω is Realized, is Assessed by $Nb_1^\omega(R)$; in this Case the Protection of Each Zone Allows to Protect a Given Set of Species Depending on the Scenario

The problem can be formulated as a linear program in Boolean variables by associating to each zone z_i, as in the previous programs, a Boolean variable x_i that takes the value 1 if and only if zone z_i is selected for protection. This results in program $\mathrm{P}_{8.1}$ which is known, in the field of operational research, as the set-covering problem.

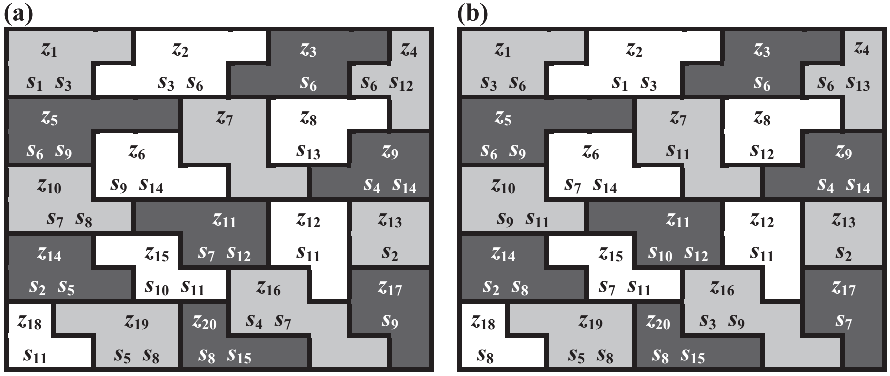


FIG. 8.2 – A region divided into twenty zones. Two scenarios are possible. (a) Species protected by the protection of a zone in the case of scenario sc_1. (b) Species protected by the protection of a zone in the case of scenario sc_2. For example, the protection of zone z_{10} ensures the protection of species s_7 and s_8, if scenario sc_1 occurs, and that of species s_9 and s_{11}, if scenario sc_2 occurs.

TAB. 8.1 – Values of θ_k^1 and θ_k^2. For example, species s_3 will survive in the selected reserve if its total population size in this reserve is greater than or equal to 20, in the case of scenario sc_1, and 17, in the case of scenario sc_2.

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
θ_k^1	9	12	20	8	14	12	6	13	16	7	10	5	17	10	7
θ_k^2	8	16	17	10	18	9	12	12	13	9	9	6	22	12	6

$$\mathrm{P}_{8.1}: \begin{cases} \min \sum_{i \in \underline{Z}} c_i x_i \\ \text{s.t.} \left| \begin{array}{lll} \sum_{i \in \underline{Z}_k^\omega} x_i \geq 1 & k \in \underline{S},\ \omega \in \underline{\mathrm{Sc}} & (8.1.1) \\ x_i \in \{0,1\} & i \in \underline{Z} & (8.1.2) \end{array} \right. \end{cases}$$

The economic function expresses the total cost of protecting the selected zones. Constraints 8.1.1 express that, for any species s_k and scenario sc_ω, at least one zone of Z_k^ω must be selected. It should be noted that wanting to protect all species considered regardless of the scenario that occurs is a very conservative but often unrealistic objective. Indeed, the optimal solution will generally consist in protecting a large number of zones – possibly all of them – to be guarded against the consequences of the different scenarios.

Example 8.2. Let us take again the instance built from figure 8.1 and described by figure 8.2. In this example, the cheapest strategy to protect all species, regardless of the scenario that occurs – among the 2 possible scenarios – is to protect the 12 zones z_1, z_2, z_4, z_6, z_8, z_9, z_{11}, z_{13}, z_{15}, z_{16}, z_{19}, and z_{20}, which costs 26 units (figure 8.3).

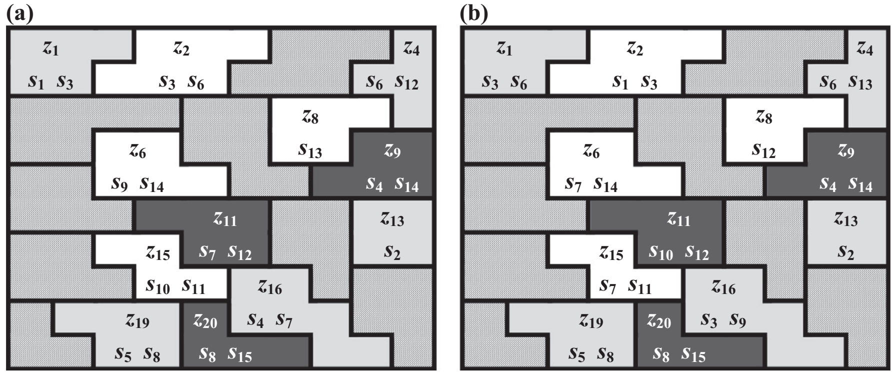


FIG. 8.3 – The least costly solution to protect all species, regardless of the scenario that occurs – among the 2 possible scenarios – is to protect the 12 non-hatched zones: z_1, z_2, z_4, z_6, z_8, z_9, z_{11}, z_{13}, z_{15}, z_{16}, z_{19}, and z_{20}. This protection costs 26 units while the protection of all zones costs 48 units. (a) All species are protected if scenario sc_1 occurs. (b) All species are protected if scenario sc_2 occurs.

As we have already discussed in the case of a single scenario (chapter 1), it can be considered that to be protected species s_k must be protected in at least b_k zones. The formulation of this variant of the problem is obtained by replacing in $\text{P}_{8.1}$ the constraints $\sum_{i \in \underline{Z}_k^\omega} x_i \geq 1,\ k \in \underline{S}, \omega \in \underline{\text{Sc}}$, by the constraints $\sum_{i \in \underline{Z}_k^\omega} x_i \geq b_k$, $k \in \underline{S}, \omega \in \underline{\text{Sc}}$.

8.2.2 Case Where the Number of Species Protected by a Reserve, R, if Scenario sc_ω Occurs, is Assessed by $Nb_2^\omega(R)$; in this Case, a Species is Protected by R if its Total Population Size in R Exceeds a Certain Value Depending on the Scenario

The problem of determining the minimal cost reserve, making it possible to protect all species considered, whatever the scenario that occurs, can be formulated as the linear program in Boolean variables obtained by replacing in $\text{P}_{8.1}$ the constraints $\sum_{i \in \underline{Z}_k^\omega} x_i \geq 1,\ k \in \underline{S}, \omega \in \underline{\text{Sc}}$, by the constraints $\sum_{i \in \underline{Z}} n_{ik} x_i \geq \theta_k^\omega,\ k \in \underline{S}, \omega \in \underline{\text{Sc}}$.

8.3 Reserve Protecting as Many Species – of a Given Set – as Possible Under a Budgetary Constraint and in the Worst-Case Scenario

A second problem that may naturally arise is to determine the zones to be protected, taking into account an available budget, B, in order to protect as many species as possible in the worst-case scenario. The worst-case scenario is related to a set of protected zones. This is the scenario for which the number of protected species is minimal, taking into account the zones selected for protection. This problem, related to species richness, can be written $\max_{R \subseteq Z, C(R) \leq B} [\min_{\omega \in \underline{\mathrm{Sc}}} \mathrm{Nb}_f^{\omega}(R)]$ where $\mathrm{Nb}_f^{\omega}(R)$ refers to the number of protected species – calculated in two different ways depending on the value of f – when the set of zones R is protected and scenario sc_{ω} is realized. $C(R)$ refers to the cost of reserve R. These problems can be formulated as linear programs in Boolean variables. For this purpose, as in all previous programs, with each zone z_i is associated a Boolean decision variable, x_i. With each possible pair (species, scenario) is also associated a "working" Boolean variable, y_k^{ω}, which, by convention, takes the value 1 if and only if the zones selected to be protected allow species s_k to be protected in the event that scenario sc_{ω} is realized.

8.3.1 Case Where the Interest of Protecting a Reserve, R, if Scenario sc_{ω} is Realized, is Assessed by $Nb_1^{\omega}(R)$

In this case, the problem can be formulated as program $\mathrm{P}_{8.2}$.

$$
\mathrm{P}_{8.2}: \begin{cases} \max\ \alpha \\ \text{s.t.} \left| \begin{array}{llll|lll} \alpha \leq \sum\limits_{k \in \underline{S}} y_k^{\omega} & \omega \in \underline{\mathrm{Sc}} & (8.2.1) & & x_i \in \{0,1\} & i \in \underline{Z} & (8.2.4) \\ y_k^{\omega} \leq \sum\limits_{i \in \underline{Z}_k^{\omega}} x_i & k \in \underline{S}, \omega \in \underline{\mathrm{Sc}} & (8.2.2) & & y_k^{\omega} \in \{0,1\} & k \in \underline{S}, \omega \in \underline{\mathrm{Sc}} & (8.2.5) \\ \sum\limits_{i \in \underline{Z}} c_i x_i \leq B & & (8.2.3) & & & & \end{array} \right. \end{cases}
$$

The objective of $\mathrm{P}_{8.2}$ is to maximize variable α. Because of constraints 8.2.1, this variable α takes the value $\min_{\omega \in \underline{\mathrm{Sc}}} \left\{ \sum_{k \in \underline{S}} y_k^{\omega} \right\}$ at the optimum of $\mathrm{P}_{8.2}$ since there is no other constraint on this variable, which corresponds to the number of protected species in the event that the worst-case scenario occurs – for a fixed set of protected zones. According to constraints 8.2.2, variable y_k^{ω}, which is a Boolean variable, takes, at the optimum of $\mathrm{P}_{8.2}$, the value 0 if $\sum_{i \in \underline{Z}_k^{\omega}} x_i = 0$, *i.e.*, if no zone of Z_k^{ω} is selected, and the value 1 if $\sum_{i \in \underline{Z}_k^{\omega}} x_i \geq 1$, *i.e.*, if at least one zone of Z_k^{ω} is selected. Variable y_k^{ω}, therefore, takes the value 1 if and only if the zones selected for protection allow species s_k to be protected, in the event that scenario sc_{ω} occurs. Constraints 8.2.4 and 8.2.5 specify the Boolean nature of all variables.

Example 8.3. Let us take again the instance described in figures 8.1 and 8.2 and assume that the budget available for the protection of the zones is equal to 10 units. By protecting the 7 zones z_2, z_4, z_6, z_8, z_{15}, z_{16}, and z_{19} we are sure that, whatever the scenario, at least 11 species will be protected. Indeed, if scenario sc_1 is realized, the 12 species s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}, s_{11}, s_{12}, s_{13}, and s_{14} will be protected, and if scenario sc_2 is realized, the 11 species s_1, s_3, s_5, s_6, s_7, s_8, s_9, s_{11}, s_{12}, s_{13}, and s_{14} will be protected (figure 8.4).

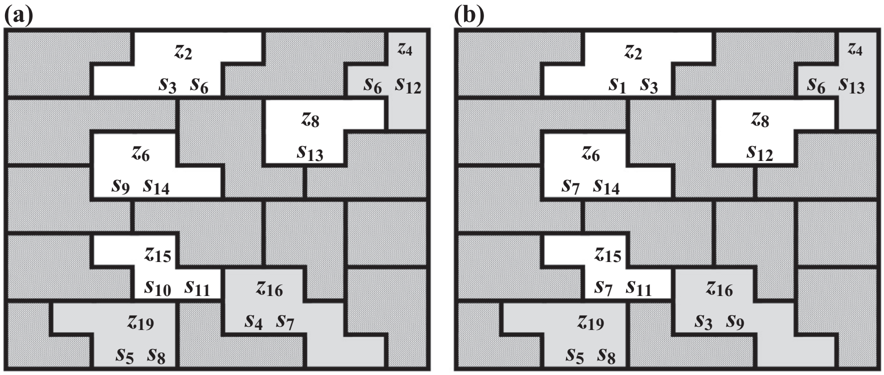


FIG. 8.4 – If the budget available for zone protection is 10 units, the optimal solution for the instance described in figures 8.1 and 8.2 consists in protecting the 7 non-hatched zones z_2, z_4, z_6, z_8, z_{15}, z_{16}, and z_{19}, which costs 10 units. (a) 12 species are protected in scenario sc_1. (b) 11 species are protected in scenario sc_2.

8.3.2 Case Where the Interest of Protecting a Reserve, R, if Scenario sc_ω is Realized, is Assessed by $Nb_2^\omega(R)$

In this case, the problem can be formulated as the mathematical program obtained by replacing in $\mathrm{P}_{8.2}$ the constraints $y_k^\omega \leq \sum_{i \in \underline{Z}_k^\omega} x_i,\ k \in \underline{S}, \omega \in \underline{\mathrm{Sc}}$, by the constraints $\theta_k^\omega y_k^\omega \leq \sum_{i \in \underline{Z}} n_{ik} x_i,\ k \in \underline{S}, \omega \in \underline{\mathrm{Sc}}$.

8.4 Reserve Minimizing the Maximal Relative Regret, for All Scenarios, About the Number of Protected Species, Taking into Account a Budgetary Constraint

Seeking to protect a set of zones in such a way that as many species as possible are protected in the worst-case scenario can have a significant drawback: if one of the scenarios is very "pessimistic" – from the viewpoint of species protection resulting from the protection of zones –, then the optimal solution of $\mathrm{P}_{8.2}$ will essentially take this only scenario into consideration. To overcome this drawback, it is possible to determine the zones to be protected – under a budgetary constraint – in such a way as to minimize the greatest regret, *i.e.,* the greatest relative gap, over all scenarios, between the number of protected species, given the zones selected, and the maximal number of species that could be protected in the scenario considered, by possibly retaining another set of zones. This problem can be written $\min_{R \subseteq Z, C(R) \leq B}$ $\{\max_{\omega \in \underline{\mathrm{Sc}}}[(\mathrm{Nb}_f^\omega(R_f^{*\omega}) - \mathrm{Nb}_f^\omega(R))/\mathrm{Nb}_f^\omega(R_f^{*\omega})]\}$ where the set of zones of maximal interest for scenario sc_ω is designated by $R_f^{*\omega}$ and when the interest of a reserve, R, is assessed by $\mathrm{Nb}_f^\omega(R)$ – f is equal to 1 or 2. To solve this problem, it is first necessary to determine the maximal interest – here, the maximal number of protected species – that can be obtained by protecting a set of zones in the case of scenario sc_ω, for all scenarios.

8.4.1 Case Where the Interest of Protecting a Reserve, R, If Scenario $\mathbf{sc}_\omega$ is Realized, is Assessed by $\boldsymbol{Nb}_1^\omega(\boldsymbol{R})$

In this case, the maximal number of species that can be protected under scenario sc_ω can be calculated by solving the linear program in 0–1 variables $\mathrm{P}_{8.3}(\omega)$.

$$\mathrm{P}_{8.3}(\omega): \begin{cases} \max \quad N_{\max}^\omega = \sum\limits_{k \in \underline{S}} y_k^\omega \\ \text{s.t.} \left| \begin{array}{llll} y_k^\omega \leq \sum\limits_{i \in \underline{Z}_k^\omega} x_i \quad k \in \underline{S} & (8.3_\omega.1) & | \quad x_i \in \{0,1\} \quad i \in \underline{Z} & (8.3_\omega.3) \\ \sum\limits_{i \in \underline{Z}} c_i x_i \leq B & (8.3_\omega.2) & | \quad y_k^\omega \in \{0,1\} \quad k \in \underline{S} & (8.3_\omega.4) \end{array} \right. \end{cases}$$

Because of the economic function to be maximized, $\sum_{k \in \underline{S}} y_k^\omega$, and constraints $8.3_\omega.1$, variable y_k^ω takes the value 1, at the optimum of $\mathrm{P}_{8.3}(\omega)$, if and only if $x_i = 1$ for at least one index i of $\underline{Z}_k^\omega$, *i.e.*, if at least one of the zones that allow species s_k to be protected under scenario sc_ω is selected to be protected. Otherwise, variable y_k^ω can only take the value 0. The value of the economic function at the optimum of $\mathrm{P}_{8.3}(\omega)$ is therefore well equal to the maximal number of species that can be protected if scenario sc_ω is realized, under the budgetary constraint expressed by constraints $8.3_\omega.2$.

Example 8.4. Let us take again the instance built from figure 8.1 and described by figure 8.2, and suppose that the budget available for the protection of the zones is equal to 7 units. If we think that scenario sc_1 will be realized, then the optimal reserve allows 10 species to be protected. It should be noted that if, contrary to the forecasts, scenario sc_2 is realized, then the reserve selected only allows for the protection of 8 species (figure 8.5). If, on the contrary, we think that scenario sc_2 will be realized, then the optimal reserve allows 9 species to be protected. Again, it should be noted that if, contrary to the forecasts, scenario sc_1 is realized, then the reserve selected only allows for the protection of 8 species (figure 8.6).

Once $N_{\max}^\omega$ is determined for all scenarios sc_ω, *i.e.*, for all $\omega \in \underline{\mathrm{Sc}}$, the optimal solution to the problem considered – minimization of the maximal regret – can be calculated by solving the linear program in Boolean variables $\mathrm{P}_{8.4}$.

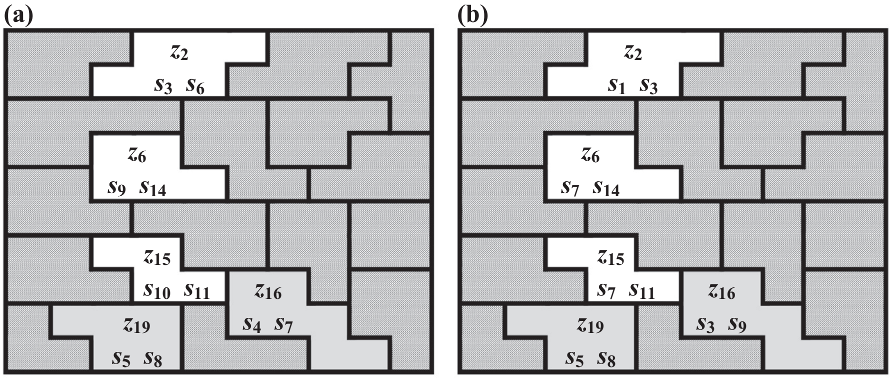


FIG. 8.5 – The budget available for the protection of the zones is equal to 7 units. (a) The optimal solution for scenario sc_1 of the instance described in figures 8.1 and 8.2 is to protect the 5 non-hatched zones z_2, z_6, z_{15}, z_{16}, and z_{19}, which costs 7 units and protects the 10 species s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}, s_{11}, and s_{14}. (b) If scenario sc_2 is realized, the protection of the 5 zones z_2, z_6, z_{15}, z_{16}, and z_{19} protects the 8 species s_1, s_3, s_5, s_7, s_8, s_9, s_{11}, and s_{14}.

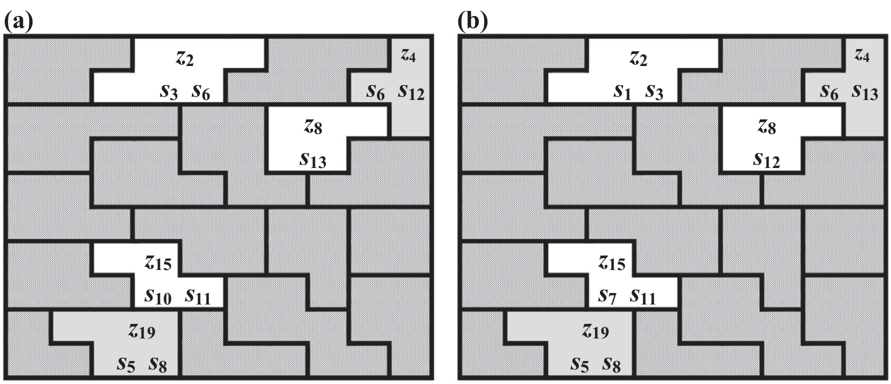


FIG. 8.6 – The budget available for the protection of the zones is equal to 7 units. (a) The optimal solution for scenario sc_2 of the instance described in figures 8.1 and 8.2 is to protect the 5 unhatched zones z_2, z_4, z_8, z_{15}, and z_{19}, which costs 7 units and protects the 9 species s_1, s_3, s_5, s_6, s_7, s_8, s_{11}, s_{12}, and s_{13}. (b) If scenario sc_1 occurs, the protection of the 5 zones z_2, z_4, z_8, z_{15}, and z_{19} protects the 8 species s_3, s_5, s_6, s_8, s_{10}, s_{11}, s_{12}, and s_{13}.

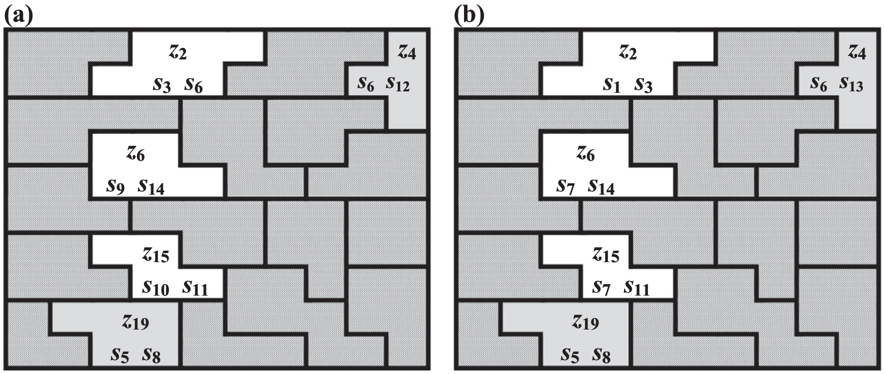


FIG. 8.7 – If the budget available for zone protection is 7 units, the optimal solution of $P_{8.4}$, for the example described in figures 8.1 and 8.2, is to protect zones z_2, z_4, z_6, z_{15}, and z_{19}. (a) The protection of these zones allows the protection of the 9 species s_3, s_5, s_6, s_8, s_9, s_{10}, s_{11}, s_{12}, and s_{14}, if scenario sc_1 is realized. (b) The protection of these zones allows the protection of the 9 species s_1, s_3, s_5, s_6, s_7, s_8, s_{11}, s_{13}, and s_{14} if scenario sc_2 is realized.

$$
\mathrm{P}_{8.4}: \left\{
\begin{array}{lllr}
\min \ \alpha & & & \\
\text{s.t.} & \alpha \geq \left(N^{\omega}_{\max} - \sum_{k\in\underline{S}} y^{\omega}_k\right) \Big/ N^{\omega}_{\max} & \omega \in \underline{\mathrm{Sc}} & (8.4.1) \\
 & y^{\omega}_k \leq \sum\limits_{i\in\underline{Z}^{\omega}_k} x_i & k\in\underline{S}, \omega\in\underline{\mathrm{Sc}} & (8.4.2) \\
 & \sum\limits_{i\in\underline{Z}} c_i x_i \leq B & & (8.4.3) \\
 & x_i \in \{0,1\} & i\in\underline{Z} & (8.4.4) \\
 & y^{\omega}_k \in \{0,1\} & k\in\underline{S}, \omega\in\underline{\mathrm{Sc}} & (8.4.5)
\end{array}
\right.
$$

Since we are seeking to minimize variable α and because of constraints 8.4.1, the Boolean variable y^{ω}_k takes, at the optimum of $P_{8.4}$, the highest possible value. Because of constraints 8.4.2, it therefore takes the value 1 if and only if the zones selected to be protected ensure the protection of species s_k, in the case of scenario sc_{ω}. In other words, y^{ω}_k takes the value 1 if and only if at least one of the zones of Z^{ω}_k is selected. Because of the economic function, α, to be minimized and constraints 8.4.1, variable α takes, at the optimum of $P_{8.4}$, the largest of the values $\left(N^{\omega}_{\max} - \sum_{k\in\underline{S}} y^{\omega}_k\right) \Big/ N^{\omega}_{\max}$, on all scenarios sc_{ω}. The resolution of $P_{8.4}$, therefore, enables the selection of zones whose protection minimizes the maximal relative gap, over all scenarios sc_{ω}, between 1) the number of species that are protected in scenario sc_{ω} given the selected zones – zone z_i is selected if $x_i = 1$ – and 2) the maximal number of species that could have been protected – possibly by protecting another set of zones – in scenario sc_{ω} (figure 8.7).

Example 8.5. Let us take again the instance built from figure 8.1 and described by figure 8.2, and suppose that the budget available for the protection of the zones is equal to 7 units. The reserve associated with the optimal solution of $P_{8.4}$ costs 7 units and allows 9 species to be protected, if scenario sc_1 is realized, and also 9 species, if scenario sc_2 is realized. For scenario sc_1 the value of the expression $\left(N_{\max}^{\omega} - \sum_{k \in \underline{S}} y_k^{\omega}\right) \Big/ N_{\max}^{\omega}$ is equal to $(10-9)/10 = 0.1$ and for scenario sc_2 it is equal to $(9-9)/9 = 0$. The corresponding value of the economic function of $P_{8.4}$, α, is therefore equal, for this example, to $\max\{0.1,\, 0\} = 0.1$. In other words, regardless of which scenario sc_{ω} occurs, the relative gap between the number of species that are protected by protecting the zones corresponding to the optimal solution of $P_{8.4}$ rather than the zones corresponding to the best strategy for scenario sc_{ω} is less than or equal to 10%.

8.4.2 Case Where the Interest of Protecting a Reserve, R, If Scenario sc_{ω} Occurs, is Assessed By $Nb_2^{\omega}(R)$

In this case, the problem of determining the maximal interest, $N_{\max}^{\omega}$, that can be obtained by protecting a set of zones, in the case of scenario sc_{ω}, can be formulated as the program obtained by replacing in $P_{8.3}(\omega)$ the constraints $y_k^{\omega} \le \sum_{i \in \underline{Z}_k^{\omega}} x_i,\ k \in \underline{S}$, by the constraints $\theta_k^{\omega}\, y_k^{\omega} \le \sum_{i \in \underline{Z}} n_{ik} x_i,\ k \in \underline{S}$. Once $N_{\max}^{\omega}$ is determined for all scenarios sc_{ω}, *i.e.,* for all $\omega \in \underline{\mathrm{Sc}}$, one can calculate the optimal solution to the problem under consideration by solving the program obtained by replacing in $P_{8.4}$ the constraints $y_k^{\omega} \le \sum_{i \in \underline{Z}_k^{\omega}} x_i,\ k \in \underline{S}, \omega \in \underline{\mathrm{Sc}}$, by the constraints $\theta_k^{\omega} y_k^{\omega} \le \sum_{i \in \underline{Z}} n_{ik} x_i,\ k \in \underline{S}, \omega \in \underline{\mathrm{Sc}}$.

References and Further Reading

Billionnet A. (2015) Designing robust nature reserves under uncertain survival probabilities, *Environ. Model. Assess.* **20**, 383.

Carvalho S.B., Brito J.C., Crespo E.G., Watts M.E., Possingham H.P. (2011) Conservation planning under climate change: toward accounting for uncertainty in predicted species distributions to increase confidence in conservation investments in space and time, *Biol. Conserv.* **144**, 2020.

Haider Z., Charkhgard H., Kwon C. (2018) A robust optimization approach for solving problems in conservation planning, *Ecol. Model.* **368**, 288.

Kouvelis P., Yu G. (1997) *Robust discrete optimization and its applications*. Kluwer Academic Publishers, Boston.

Moilanen A., Runge M.C., Elith J., Tyre A., Carmel Y., Fegraus E., Wintle B.A., Burgman M., Ben-Haim Y. (2006) Planning for robust reserve networks using uncertainty analysis, *Ecol. Model.* **199**, 115.

Moilanen A., Wintle B.A., Elith J., Burgman M. (2006) Uncertainty analysis for regional-scale reserve selection, *Conserv. Biol.* **20**, 1688.

Snyder S.A., Haight R.G., ReVelle C.S. (2004) A scenario optimization model for dynamic reserve site selection, *Environ. Model. Assess.* **9**, 179.

Chapter 9

Species Survival Probabilities' Scenarios

9.1 Introduction

In this chapter, the uncertainty that may exist with regard to the survival of the species under consideration, taking into account the protection policies of the different candidate zones, is expressed both by the survival probabilities of these species in protected and unprotected zones but also by a set of possible scenarios, $\mathrm{Sc} = \{\mathrm{sc}_1, \mathrm{sc}_2, \ldots, \mathrm{sc}_\mathrm{p}\}$. A scenario corresponds to a set of hypotheses on the evolution of the factors likely to influence the survival probabilities of the species. These assumptions may concern direct factors such as land use, climate change, pollution, overexploitation or invasive species, and indirect factors such as economic activity, demographic change, and socio-political contexts. In this chapter, the survival probabilities of the species are therefore scenario-dependent. As in the previous chapters, the term "species" refers to the set of species of interest, and "candidate zones" refers to the set of candidate zones for protection. Let $S = \{s_1, s_2, \ldots, s_m\}$ be the set of species considered and $\mathrm{Z} = \{z_1, z_2, \ldots, z_n\}$ be the set of candidate zones. Denote by $\underline{S}$, $\underline{Z}$, and $\underline{\mathrm{Sc}}$ the set of indices of the elements of S, Z, and Sc, respectively. Thus, the information concerning the survival of species s_k in the protected zone z_i is provided by the probability p_{ik}^{ω}, $i \in \underline{Z}$, $k \in \underline{S}$, $\omega \in \underline{\mathrm{Sc}}$, assuming that scenario sc_ω is realized. The survival probability of the species s_k in the unprotected zone z_i, $k \in \underline{S}$, $i \in \underline{Z}$, is assumed to be 0 for all scenarios. It is also assumed that all these probabilities are independent. Note that the value of p_{ik}^{ω} may be 0 for some triplets (i, k, ω) even if zone z_i is protected. Indeed, on the one hand, some zones, even if protected, do not contribute to the protection of certain species in any scenario – there are i and k such that $p_{ik}^{\omega} = 0$ for all $\omega \in \underline{\mathrm{Sc}}$ – and, on the other hand, a given zone may contribute to the protection of a certain species in the case of scenario sc_{ω_1} but not contribute to the protection of that species in the case of scenario sc_{ω_2} – there are i, k, ω_1, and ω_2 such as $p_{ik}^{\omega_1} > 0$ and $p_{ik}^{\omega_2} = 0$.

Given a reserve, R, *i.e.*, a subset of zones of Z that one decides to protect, denote by $\mathrm{Int}^{\omega}(R)$ the interest in protecting R in the case of scenario sc_ω. In sections 9.2–9.5

DOI: 10.1051/978-2-7598-2521-9.c009

of this chapter, $\mathrm{Int}^{\omega}(R)$ represent the number of species whose survival probability in reserve R is greater than or equal to a certain threshold value – denoted by ρ_k for species s_k – in the case of scenario sc_{ω}. $\mathrm{Int}^{\omega}(R)$ can therefore be defined as follows: $\mathrm{Int}^{\omega}(R) = \left|\left\{k \in \underline{S} : P_k^{\omega}(R) \geq \rho_k\right\}\right|$ where $P_k^{\omega}(R)$ is the survival probability of species s_k in reserve R, in the case of scenario sc_{ω}. This probability $P_k^{\omega}(R)$ is equal to $1 - \prod_{i \in \underline{R}} \left(1 - p_{ik}^{\omega}\right)$ where $\underline{R}$ denotes the set of indices of the zones belonging to reserve R. In section 9.6 of this chapter, $\mathrm{Int}^{\omega}(R)$ represents the expected number of species that will survive in reserve R if scenario sc_{ω} occurs. Thus, in this case, we have $\mathrm{Int}^{\omega}(R) = \sum_{k \in \underline{S}} P_k^{\omega}(R)$.

In the remainder of this chapter, we focus on the determination of optimal robust reserves by giving several meanings to the term “robust”. In all cases, this qualifier refers to reserves that have a certain level of interest regardless of the scenario that occurs.

9.2 Reserve Ensuring a Certain Survival Probability for the Largest Possible Number of Species, of a Given Set, Under a Budgetary Constraint and Regardless of the Scenario

In this section, let us examine the determination of a reserve, respecting a budgetary constraint and guaranteeing certain objectives, whatever the scenario that occurs. Such a reserve is referred to as “robust”. Recall that for any triplet $(i, k, \omega) \in \underline{Z} \times \underline{S} \times \underline{\mathrm{Sc}}$, the survival probability of species s_k in zone z_i in the case of scenario sc_{ω} is equal to p_{ik}^{ω} if zone z_i is protected and 0 in the opposite case. We consider here that there is only one level of protection: a zone is protected or not. The problem consists in determining a reserve, *i.e.*, a set of zones to be protected, whose protection cost is less than or equal to the available budget, denoted by B, and which satisfies the following property: the number of species of S whose survival probability in the reserve is greater than or equal to a certain threshold value – depending on the species – is maximal in the worst-case scenario. For a given reserve, the worst-case scenario is the one that leads to the lowest number of species whose survival probability – in that reserve – is greater than or equal to the specified threshold value. Denote by ρ_k the threshold value corresponding to species s_k. Using the notation $\mathrm{Int}^{\omega}(R)$ for the interest of reserve R in the case of scenario sc_{ω}, *i.e.*, the number of species whose survival probability is greater than or equal to the set threshold value, this optimization problem can be concisely formulated as follows: $\max_{R \subseteq Z, C(R) \leq B}\left\{\min_{\omega \in \underline{\mathrm{Sc}}} \mathrm{Int}^{\omega}(R)\right\}$. Let us introduce the Boolean decision variable x_i which takes the value 1 if and only if zone z_i is protected. The extinction probability of species s_k in zone z_i can then be written $1 - p_{ik}^{\omega} x_i$, in the case of scenario sc_{ω} and as a function of variable x_i. It is deduced that the probability of disappearance of species s_k from the reserve, in the case of scenario sc_{ω}, is equal to $\prod_{i \in \underline{Z}} \left(1 - p_{ik}^{\omega} x_i\right)$, and finally that the survival probability of species s_k in the reserve, *i.e.*, in the set of protected zones, is equal to $1 - \prod_{i \in \underline{Z}} \left(1 - p_{ik}^{\omega} x_i\right)$. The problem we consider here is to determine the zones to be protected, *i.e.*, the values of variables x_i, in such a way as to ensure, for all scenarios, a survival probability greater than or equal to a certain threshold value, for as many species as possible. In other words, we seek to determine a reserve, R, such that, among the m constraints $1 - \prod_{i \in \underline{Z}} \left(1 - p_{ik}^{\omega} x_i\right) \geq \rho_k$, $k \in \underline{S}$, as many as possible are satisfied for all $\omega \in \underline{\mathrm{Sc}}$. In order to formulate the problem as a mathematical program, let us also introduce the Boolean variable y_k^{ω} which takes the value 1 if and only if the survival probability of species s_k in the reserve is greater than or equal to the threshold value ρ_k, in the case of scenario sc_{ω}. The optimal robust solution can be determined by solving the mathematical program in integer variables $\mathrm{P}_{9.1}$.

$$
\mathrm{P}_{9.1}: \begin{cases} \max\ \alpha & \\ \text{s.t.} \left| \begin{array}{lll} \sum_{i \in \underline{Z}} c_i x_i \leq B & & (9.1.1) \\ 1 - \prod_{i \in \underline{Z}} (1 - p_{ik}^{\omega} x_i) \geq \rho_k y_k^{\omega} & k \in \underline{S}, \omega \in \underline{\mathrm{Sc}} & (9.1.2) \\ \alpha \leq \sum_{k \in \underline{S}} y_k^{\omega} & \omega \in \underline{\mathrm{Sc}} & (9.1.3) \\ x_i \in \{0,1\} & i \in \underline{Z} & (9.1.4) \\ y_k^{\omega} \in \{0,1\} & k \in \underline{S}, \omega \in \underline{\mathrm{Sc}} & (9.1.5) \end{array} \right. \end{cases}
$$

The economic function of $\mathrm{P}_{9.1}$ is variable α to be maximized. Because of constraints 9.1.3, the value of variable α, at the optimum of $\mathrm{P}_{9.1}$ is equal to the number of species with a survival probability in the reserve greater than or equal to the specified threshold value, in the worst-case scenario. Indeed, at the optimum of $\mathrm{P}_{9.1}$, we have $\alpha = \min_{\omega \in \underline{\mathrm{Sc}}} \left\{ \sum_{k \in \underline{S}} y_k^{\omega} \right\}$. Constraint 9.1.1 expresses that the total cost of protecting the reserve must be less than or equal to the available budget, B. Constraints 9.1.2 force the Boolean variables y_k^{ω} to take the value 0 if the survival probability in the reserve of species s_k is less than the threshold value ρ_k, in the case of scenario sc_{ω}. Otherwise, and because of the expression of the economic function to be maximized, the Boolean variables y_k^{ω} take the value 1 at the optimum of $\mathrm{P}_{9.1}$. Constraints 9.1.4 and 9.1.5 specify the Boolean nature of variables x_i and y_k^{ω}. The economic function is linear but constraints 9.1.2 are non-linear since they involve the product of the n linear functions $1 - p_{ik}^{\omega} x_i$. We will see that, as in section 7.2 of chapter 7, these constraints 9.1.2 can be linearized and that therefore, finally, the solution to the problem considered can be determined by solving a linear program in Boolean variables. Let us first rewrite constraints 9.1.2 as $\prod_{i \in \underline{Z}} \left(1 - p_{ik}^{\omega} x_i\right) \leq 1 - \rho_k y_k^{\omega}$. To simplify the presentation, it is assumed that p_{ik}^{ω} and ρ_k are strictly less than 1. These constraints are equivalent to $\log\left(\prod_{i \in \underline{Z}} (1 - p_{ik}^{\omega} x_i)\right) \leq \log(1 - \rho_k y_k^{\omega})$ or $\sum_{i \in \underline{Z}} \log(1 - p_{ik}^{\omega} x_i) \leq \log(1 - \rho_k y_k^{\omega})$. Since variables x_i and y_k^{ω} are Boolean variables, $\log(1 - p_{ik}^{\omega} x_i) = x_i \log(1 - p_{ik}^{\omega})$ and $\log(1 - \rho_k y_k^{\omega}) = y_k^{\omega} \log(1 - \rho_k)$. The nonlinear constraints 9.1.2 are, therefore, equivalent to the linear constraints $\sum_{i\in\underline{Z}} x_i \log(1-p_{ik}^{\omega}) \leq y_k^{\omega}\log(1-\rho_k),\ \ k\in\underline{S}, \omega\in\underline{\mathrm{Sc}}$. Finally, the optimal robust solution to the problem under consideration can be determined by solving the linear program in Boolean variables $\mathrm{P}_{9.2}$.

$$\mathrm{P}_{9.2}:\left\{\begin{array}{ll}\max\ \alpha & \\ \text{s.t.} & \left|\begin{array}{lll|lll}\sum_{i\in\underline{Z}} c_i x_i \leq B & & (9.2.1) & \alpha\leq\sum_{k\in\underline{S}} y_k^{\omega} & \omega\in\underline{\mathrm{Sc}} & (9.2.3)\\ \sum_{i\in\underline{Z}} x_i\log(1-p_{ik}^{\omega}) & & & x_i\in\{0,1\} & i\in\underline{Z} & (9.2.4)\\ \quad\leq y_k^{\omega}\log(1-\rho_k) & k\in\underline{S},\ \omega\in\underline{\mathrm{Sc}} & (9.2.2) & y_k^{\omega}\in\{0,1\} & k\in\underline{S},\ \omega\in\underline{\mathrm{Sc}} & (9.2.5)\end{array}\right.\end{array}\right.$$

By setting $\mu_{ik}^{\omega}=\log(1-p_{ik}^{\omega})$ and $\nu_k=\log(1-\rho_k)$, program $\mathrm{P}_{9.2}$ is rewritten as program $\mathrm{P}_{9.3}$.

$$\mathrm{P}_{9.3}:\left\{\begin{array}{ll}\max\ \alpha & \\ \text{s.t.} & \left|\begin{array}{lll|lll}\sum_{i\in\underline{Z}} c_i x_i \leq B & & (9.3.1) & x_i\in\{0,1\} & i\in\underline{Z} & (9.3.4)\\ \sum_{i\in\underline{Z}} \mu_{ik}^{\omega} x_i \leq \nu_k y_k^{\omega} & k\in\underline{S},\omega\in\underline{\mathrm{Sc}} & (9.3.2) & y_k^{\omega}\in\{0,1\} & k\in\underline{S},\ \omega\in\underline{\mathrm{Sc}} & (9.3.5)\\ \alpha\leq\sum_{k\in\underline{S}} y_k^{\omega} & \omega\in\underline{\mathrm{Sc}} & (9.3.3) & & & \end{array}\right.\end{array}\right.$$

9.3 Least Cost Reserve Ensuring a Certain Survival Probability for All Considered Species, Regardless of the Scenario

The following variant of the problem in the previous section can be considered: determine an optimal robust reserve, that is, here, a set of zones to be protected, of minimal cost, and which ensures that all species of S have a survival probability in the reserve – or equivalently in all candidate zones – greater than or equal to a certain threshold value, in all scenarios. As in the previous section, the survival probability of species s_k in zone z_i, and in the case of scenario sc_{ω}, is equal to p_{ik}^{ω} if zone z_i is protected and 0 if it is not. This optimization problem can be written in a concise way: $\min\{C(R): R\subseteq Z, \mathrm{Int}^{\omega}(R)=m\ (\omega\in\underline{\mathrm{Sc}})\}$ where $\mathrm{Int}^{\omega}(R)$ is the interest in protecting R – here, the number of species whose survival probability is at least equal to the set threshold value – in the case of scenario sc_{ω}. Recall that $P_k^{\omega}(R)$ is the survival probability of species s_k in reserve R in case of scenario sc_{ω}. The problem can also be written as follows: $\min\big\{C(R): R\subseteq Z,\ P_k^{\omega}(R)\geq\rho_k\ (k\in\underline{S},\omega\in\underline{\mathrm{Sc}})\big\}$. The optimal robust solution to the problem can be obtained by solving the linear program in Boolean variables $\mathrm{P}_{9.4}$.

$$\mathrm{P}_{9.4}: \begin{cases} \min \sum\limits_{i\in\underline{Z}} c_i x_i \\ \text{s.t.} \left| \begin{array}{lll} \sum\limits_{i\in\underline{Z}} \mu_{ik}^{\omega} x_i \le \nu_k & k \in \underline{S}, \omega \in \underline{\mathrm{Sc}} & (9.4.1) \\ x_i \in \{0,1\} & i \in \underline{Z} & (9.4.2) \end{array} \right. \end{cases}$$

where the coefficients μ_{ik}^{ω} and ν_k have the same meaning as in $\mathrm{P}_{9.3}$.

9.4 Reserve Subject to a Budgetary Constraint and Minimizing the Maximal Relative Regret, over All Scenarios, About the Number of Species of a Given Set with a Survival Probability Above a Certain Threshold Value

As already pointed out for similar contexts, seeking to protect a set of zones in such a way that as many species as possible have a survival probability greater than or equal to a certain threshold value, in the worst-case scenario, can have a significant drawback: if one of the scenarios is very "pessimistic" then the reserve selected will essentially take into account that single scenario. To overcome this disadvantage, one can seek to determine the zones to be protected – under a budgetary constraint – in such a way as to minimize the greatest relative gap, over all scenarios sc_{ω}, between (1) the number of species with a survival probability greater than or equal to a certain threshold value, taking into account the zones selected, and (2) the maximal number of species that could have a survival probability greater than or equal to the same threshold value in scenario sc_{ω}. This problem of determining an optimal robust reserve can be written $\min_{R\subseteq Z\,:\,C(R)\le B}\left\{\max_{\omega\in\underline{\mathrm{Sc}}}\left[(\mathrm{Int}^{\omega}(R^{*\omega})-\mathrm{Int}^{\omega}(R))/\mathrm{Int}^{\omega}(R^{*\omega})\right]\right\}$ where $R^{*\omega}$ is the set of zones of maximal interest for scenario sc_{ω}. Recall that, in sections 9.2–9.5 of this chapter, the interest of reserve, R, corresponds to the number of species whose survival probability in that reserve is greater than or equal to a pre-set threshold value. To solve this problem, we must first determine the maximal interest that can be obtained – by protecting some zones of Z – in the case of scenario sc_{ω}. The following optimization problem must, therefore, be solved for any scenario sc_{ω}: $\max_{R\subseteq Z\,:\,C(R)\le B} \mathrm{Int}^{\omega}(R)$. This problem can be formulated as the linear program in Boolean variables $\mathrm{P}_{9.5}(\omega)$.

$$\mathrm{P}_{9.5}(\omega): \begin{cases} \max\ \mathrm{Int}_{\max}^{\omega} = \sum\limits_{k\in\underline{S}} y_k^{\omega} \\ \text{s.t.} \left| \begin{array}{llll} \sum\limits_{i\in\underline{Z}} c_i x_i \le B & & (9.5_{\omega}.1) & \mid\ x_i \in \{0,1\} \quad i \in \underline{Z} \quad (9.5_{\omega}.3) \\ \sum\limits_{i\in\underline{Z}} \mu_{ik}^{\omega} x_i \le \nu_k y_k^{\omega} & k \in \underline{S} & (9.5_{\omega}.2) & \mid\ y_k^{\omega} \in \{0,1\} \quad k \in \underline{S} \quad (9.5_{\omega}.4) \end{array} \right. \end{cases}$$

Because of the economic function to be maximized, $\sum_{k\in\underline{S}} y_k^\omega$, and constraints $9.5_\omega.2$, the Boolean variable y_k^ω takes the value 1, at the optimum of $\mathrm{P}_{9.5}(\omega)$, if and only if the survival probability of species s_k in the selected reserve is greater than or equal to ρ_k, in the case of scenario sc_ω. Otherwise, variable y_k^ω can only take the value 0. The value of the economic function, at the optimum of $\mathrm{P}_{9.5}(\omega)$, is therefore equal to the maximal number of species whose survival probability, in the case of scenario sc_ω, is greater than or equal to the pre-set threshold value, taking into account the budgetary constraint expressed by $9.5_\omega.1$. This value is denoted by $\mathrm{Int}^\omega_{\max}$. It corresponds to $\mathrm{Int}^\omega(R^{*\omega})$. Once $\mathrm{Int}^\omega_{\max}$ has been determined for all scenarios, *i.e.*, for all $\omega \in \underline{\mathrm{Sc}}$, the optimal solution to the problem under consideration can be found by solving the mixed-integer linear program $\mathrm{P}_{9.6}$.

$$
\mathrm{P}_{9.6}:\begin{cases}
\min\ \alpha \\
\text{s.t.}\left|\begin{array}{lll}
\sum_{i\in\underline{Z}} c_i x_i \le B & & (9.6.1)\\
\sum_{i\in\underline{Z}} \mu_{ik}^\omega x_i \le \nu_k y_k^\omega & k\in\underline{S},\ \omega\in\underline{\mathrm{Sc}} & (9.6.2)\\
\alpha \ge \left(\mathrm{Int}^\omega_{\max} - \sum_{k\in\underline{S}} y_k^\omega\right)/\mathrm{Int}^\omega_{\max} & \omega\in\underline{\mathrm{Sc}} & (9.6.3)\\
x_i\in\{0,1\} & i\in\underline{Z} & (9.6.4)\\
y_k^\omega\in\{0,1\} & k\in\underline{S},\ \omega\in\underline{\mathrm{Sc}} & (9.6.5)\\
\alpha\ge 0 & & (9.6.6)
\end{array}\right.
\end{cases}
$$

Because of constraints 9.6.2, the Boolean variable y_k can take the value 1 if and only if the zones selected for protection provide species s_k with a survival probability greater than or equal to ρ_k, in the case of scenario sc_ω. Because of the economic function, α, to be minimized and constraints 9.6.3, variable α takes, at the optimum of $\mathrm{P}_{9.6}$, the largest of the values $\left(\mathrm{Int}^\omega_{\max} - \sum_{k\in\underline{S}} y_k^\omega\right)/\mathrm{Int}^\omega_{\max}$ over all scenarios sc_ω. The resolution of $\mathrm{P}_{9.6}$ therefore allows for the selection of a reserve whose protection minimizes the maximal relative gap, over the set of scenarios, between the number of species that have a survival probability in that reserve greater than or equal to the threshold value – zone z_i is selected if $x_i = 1$ – and the maximal number of species that would have a survival probability greater than or equal to the threshold value in a reserve that is optimal for the scenario under consideration.

9.5 Examples

In this section, we illustrate the results of the previous sections on a hypothetical set of candidate zones represented by a grid of 8×8 square and identical zones. In this example, 10 species, s_1, s_2,…, s_{10}, are involved and 2 scenarios, sc_1 and sc_2, are considered. The data are presented in figure 9.1. The zones are designated by z_{ij} where i denotes their row index and j, their column index. For each zone z_{ij}, the non-zero

	1	2	3	4	5	6	7	8
1	s_1 : 0.7, 0.4 s_8 : 0.7, 0.3 s_{10} : 0.3, 0.8 9	s_3 : 0.5, 0.5 2	3	s_5 : 0.4, 0.5 s_8 : 0.6 , 0.3 3	s_6 : 0.6, 0.7 s_{10} : 0.7, 0.3 3	s_1 : 0.6 , 0.4 3	s_2 : 0.4 , 0.6 s_4 : 0.4, 0.5 9	s_2 : 0.9, 0.6 s_7 : 0.6, 0.8 s_{10} : 0.8, 0.4 5
2	4	s_9 : 0.8, 0.6 1	s_9 : 0.5, 0.8 1	s_5 : 0, 0.7 9	s_1 : 0.6, 0.4 5	s_3 : 0.6, 0.8 s_7 : 0.8, 0.5 6	3	4
3	7	8	10	s_3 : 0.6, 0.8 s_{10} : 0.5, 0 6	4	s_9 : 0.9, 0.5 8	s_7 : 0.5, 0.8 10	4
4	3	s_9 : 0.6, 0.8 3	s_2 : 0.6, 0.3 8	s_4 : 0.3, 0.7 7	6	s_9 : 0.7, 0.7 5	9	9
5	1	10	2	s_4 : 0.9, 0.5 7	s_3 : 0.8, 0.4 s_7 : 0.5, 0.8 3	s_7 : 0.5, 0.8 s_8 : 0.8, 0.4 6	s_5 : 0.4, 0.6 1	6
6	s_2 : 0.6, 0.8 3	s_4 : 0.8, 0.4 s_9 : 0.7, 0.8 4	6	2	s_3 : 0.6, 0.3 10	6	s_4 : 0.7, 0.5 1	s_5 : 0.6, 0.8 7
7	s_2 : 0.4, 0.7 8	2	s_8 : 0, 0.6 5	1	s_3 : 0.8, 0.4 s_4 : 0.4, 0.7 3	s_8 : 0.8, 0.8 6	8	s_6 : 0.5, 0.7 8
8	5	10	s_1 : 0.5, 0.7 s_3 : 0.7, 0 5	s_2 : 0.5, 08 s_4 : 0.5, 0.5 s_6 : 0.4, 0.6 5	s_3 : 0.6, 0.4 4	s_2 : 0.7, 0.5 7	s_5 : 0.7, 0.6 4	s_6 : 0, 0.8 7

FIG. 9.1 – A set of 64 candidate zones for protection represented by a grid of 8 × 8 square and identical zones. 10 species are concerned and 2 scenarios are envisaged. The corresponding survival probabilities, p_{ijk}^{ω}, and the costs of protection are indicated in each zone. Consider, for example, zone z_{56}. Species s_7 and s_8 are concerned by this zone. The survival probability of species s_7 in this zone, if protected, is equal to 0.5 for the first scenario and 0.8 for the second one. The survival probability of species s_8 in this zone, if protected, is 0.8 for the first scenario and 0.4 for the second one. The cost of protecting this zone is equal to 6.

survival probabilities of the species in that zone, when protected and for each scenario, are indicated. Here p_{ijk}^{ω} refers to the survival probability of species s_k in the protected zone z_{ij} and in the case of scenario sc_{ω}. Recall that all survival probabilities are zero in unprotected zones. The cost associated with protecting each zone is indicated in the lower right-hand corner of the corresponding zone. In all that follows, we will say that a species is protected by a reserve, R, in the case of scenario sc_{ω}, if the survival probability of that species – in reserve R – is greater than or equal to the set threshold value.

TAB. 9.1 – List of zones whose protection ensures a positive survival probability for species s_k, for at least one of the two scenarios.

Species	Zones	Species	Zones
s_1	z_{11} z_{16} z_{25} z_{83}	s_6	z_{15} z_{78} z_{84} z_{88}
s_2	z_{17} z_{18} z_{43} z_{61} z_{71} z_{84} z_{86}	s_7	z_{18} z_{26} z_{37} z_{55} z_{56}
s_3	z_{12} z_{26} z_{34} z_{55} z_{65} z_{75} z_{83} z_{85}	s_8	z_{11} z_{14} z_{56} z_{73} z_{76}
s_4	z_{17} z_{44} z_{54} z_{62} z_{67} z_{75} z_{84}	s_9	z_{22} z_{23} z_{36} z_{42} z_{46} z_{62}
s_5	z_{14} z_{24} z_{57} z_{68} z_{87}	s_{10}	z_{11} z_{15} z_{18} z_{34}

TAB. 9.2 – Problem I: Optimal robust reserve characteristics for three values of the available budget and for two threshold values.

B	ρ	Number of selected zones	Cost of the reserve	Number of protected species	Species protected in scenario sc_1	Species protected in scenario sc_2	Associated figure
10	0.8	5	9	3	s_3, s_4, s_9	s_2, s_4, s_9	9.2a
	0.9	5	10	2	s_4, s_9	s_4, s_9	–
20	0.8	8	19	5	s_2, s_3, s_4, s_5, s_9, s_{10}	s_2, s_4, s_5, s_7, s_9	9.2b
	0.9	7	19	4	s_2, s_3, s_4, s_9	s_2, s_4, s_7, s_9	–
30	0.8	11	30	7	s_2, s_3, s_4, s_5, s_8, s_9, s_{10}	s_2, s_3, s_4, s_5, s_7, s_8, s_9	–
	0.9	10	30	5	s_2, s_3, s_4, s_7, s_9	s_2, s_3, s_4, s_7, s_9	9.2c

To facilitate the review of this example, table 9.1 lists for all $k \in \underline{S}$ the zones for which the survival probability of species s_k is positive for at least one of the two scenarios.

Let us consider the three following problems, each of which consists of determining an optimal robust reserve:

Problem I. Determine a reserve that respects a certain budget and maximizes the number of species whose survival probability – in the reserve or equivalently in the set of candidate zones – is greater than or equal to 0.8 then 0.9 – $\rho_k = 0.8$ then $\rho_k = 0.9$ for all k – regardless of the scenario.

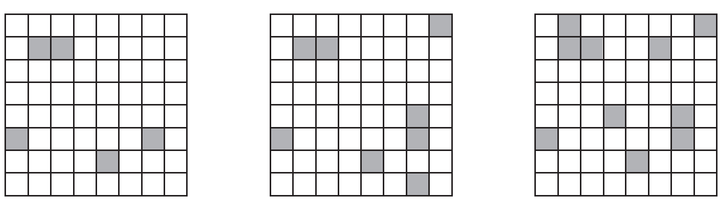
Problem II. Determine a minimal cost reserve that ensures that all species considered have a survival probability – in the reserve or equivalently in the set of candidate zones – greater than or equal to 0.8 then 0.85 then 0.9 – $\rho_k = 0.8$ then $\rho_k = 0.85$ then $\rho_k = 0.9$ for all k – regardless of the scenario.

Problem III. Determine a reserve that respects a certain budget and minimizes the maximal relative regret, over all scenarios, on the number of species whose survival probability – in the reserve or equivalently in the set of candidate zones – is greater than or equal to 0.8 then 0.9 – $\rho_k = 0.8$ then $\rho_k = 0.9$ for all k.

In this example and for these three problems, we refer to ρ the threshold value applicable to all species. Note that the survival probability of a species in the reserve is equal to the survival probability of that species in the set of candidate zones since, by hypothesis, the survival probabilities of the species considered in the unprotected zones are all equal to 0.

The results obtained for Problem I are presented in table 9.2. Some optimal robust reserves corresponding to the instances in this table are shown in figure 9.2. Table 9.3 gives the survival probabilities of each species in the optimal robust reserve, for both scenarios, when the available budget equals 20 and the threshold value equals 0.8.

The results obtained for Problem II are presented in table 9.4. It can be noted that it is not possible to define a reserve to ensure, for all scenarios, a survival probability of at least 0.9 for all species. Some optimal robust reserves corresponding to the instances in table 9.4 are presented in figure 9.3. Table 9.5 gives the survival



(a) $B = 10, \rho = 0.8$ **(b)** $B = 20, \rho = 0.8$ **(c)** $B = 30, \rho = 0.9$

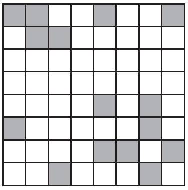
FIG. 9.2 – Problem I: Optimal robust reserves corresponding to certain instances in table 9.2.

TAB. 9.3 – Problem I: Species survival probabilities in the optimal robust reserve, for both scenarios, when $B = 20$ and $\rho_k = 0.8$ for all k.

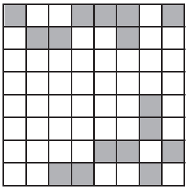
Species	Survival probability in the reserve (scenario sc_1)	Survival probability in the reserve (scenario sc_2)	Species	Survival probability in the reserve (scenario sc_1)	Survival probability in the reserve (scenario sc_2)
s_1	0	0	s_6	0	0
s_2	0.96	0.92	s_7	0.60	0.80
s_3	0.80	0.40	s_8	0	0
s_4	0.82	0.85	s_9	0.90	0.92
s_5	0.82	0.84	s_{10}	0.80	0.40

TAB. 9.4 – Problem II: Optimal robust reserve characteristics for three threshold values.

ρ	Cost of the optimal robust reserve	Associated figure
0.80	55	9.3a
0.85	64	9.3b
0.90	No feasible reserve	–



(a) $\rho = 0.8$



(b) $\rho = 0.85$

FIG. 9.3 – Problem II: Optimal robust reserves for the instances in table 9.4.

TAB. 9.5 – Problem II: Species survival probabilities in the optimal robust reserve for both scenarios, and when $\rho_k = 0.8$ for all k.

Species	Survival probability in the reserve (scenario sc_1)	Survival probability in the reserve (scenario sc_2)	Species	Survival probability in the reserve (scenario sc_1)	Survival probability in the reserve (scenario sc_2)
s_1	0.850	0.820	s_6	0.800	0.910
s_2	0.960	0.920	s_7	0.800	0.960
s_3	0.994	0.820	s_8	0.940	0.860
s_4	0.820	0.850	s_9	0.900	0.920
s_5	0.820	0.840	s_{10}	0.958	0.916

probabilities of each species in the optimal robust reserve, for both scenarios, when the threshold value is equal to 0.8.

The results obtained for Problem III are presented in tables 9.6 and 9.7.

9.6 Reserve Satisfying a Budgetary Constraint and Maximizing the Expected Number of Species, of a Given Set, that will Survive in it

As in the previous sections, $S = \{s_1, s_2,..., s_m\}$ refers to the set of species under consideration, $Z = \{z_1, z_2,..., z_n\}$, the set of candidate zones for protection, and $\mathrm{Sc} = \{\mathrm{sc}_1, \mathrm{sc}_2,..., \mathrm{sc}_p\}$, the set of possible scenarios. Thus, information concerning the survival of species s_k in the protected zone z_i when scenario sc_ω is assumed to occur is provided by the probability p_{ik}^ω, $i \in \underline{Z}$, $k \in \underline{S}$, $\omega \in \underline{\mathrm{Sc}}$. In order to simplify the presentation it is assumed that all these probabilities are strictly less than 1. On the other hand, the survival probabilities of the different species in unprotected zones are all assumed to be zero.

As in chapter 7, section 7.5, the aim is to identify a set of zones to be protected, with a cost less than or equal to a certain value, B, so as to maximize the expected number of species that will survive in that set of zones. The difference with chapter 7, section 7.5, is that now several scenarios are considered. The next two sections 9.6.1 and 9.6.2, consider two slightly different problems.

As before, $P_k^\omega(R)$ refers to the survival probability in reserve R of species s_k in the case of scenario sc_ω. As we saw in the introduction, $P_k^\omega(R) = 1 - \prod_{i\in\underline{Z}}\left(1 - p_{ik}^\omega x_i\right)$. Recall that the reserve is defined by zones z_i such that $x_i = 1$. We deduce that the expected number of species that will survive in the reserve in the case of scenario sc_ω is equal to $\sum_{k\in\underline{S}}\left[1 - \prod_{i\in\underline{Z}}\left(1 - p_{ik}^\omega x_i\right)\right]$. Note that one could give different importance to each species and thus consider the expected weighted number of species that will survive in the reserve, *i.e.*, the quantity $\sum_{k\in\underline{S}} w_k\left[1 - \prod_{i\in\underline{Z}}\left(1 - p_{ik}^\omega x_i\right)\right]$ where w_k is the weight assigned to species s_k.

9.6.1 Reserve Respecting a Budgetary Constraint and Maximizing the Expected Number of Species, of a Given Set, that will Survive in it in the Worst-Case Scenario

In this section, we focus on determining an optimal robust reserve, that is, a reserve that respects a certain budget and maximizes the expected weighted number of species that will survive in this reserve in the worst-case scenario. For a given reserve, the worst-case scenario here is the one for which the expected weighted number of species that will survive in this reserve is the lowest. The problem can, therefore, be formulated as the mixed-integer mathematical program $\mathrm{P}_{9.7}$ in which the Boolean variable x_i takes the value 1 if and only if zone z_i is selected to be part of the reserve.

TAB. 9.6 – Problem III: Optimal reserves for each scenario, for three values of the available budget and for two threshold values.

B	ρ	Scenario	Optimal reserve	Species with a survival probability greater than the threshold value
10	0.8	sc$_1$	z_{18} z_{22} z_{55}	s_2 s_3 s_7 s_9 s_{10}
		sc$_2$	z_{12} z_{23} z_{55} z_{67} z_{75}	s_3 s_4 s_7 s_9
	0.9	sc$_1$	z_{15} z_{18} z_{22} z_{23}	s_2 s_9 s_{10}
		sc$_2$	z_{18} z_{22} z_{23} z_{55}	s_7 s_9
20	0.8	sc$_1$	z_{14} z_{18} z_{22} z_{56} z_{57} z_{67} z_{75}	s_2 s_3 s_4 s_7 s_8 s_9 s_{10}
		sc$_2$	z_{12} z_{14} z_{23} z_{55} z_{57} z_{75} z_{84}	s_2 s_3 s_4 s_5 s_7 s_9
	0.9	sc$_1$	z_{15} z_{18} z_{22} z_{55} z_{62} z_{67} z_{75}	s_2 s_3 s_4 s_9 s_{10}
		sc$_2$	z_{18} z_{22} z_{23} z_{55} z_{67} z_{75} z_{84}	s_2 s_4 s_7 s_9
20	0.8	sc$_1$	z_{16} z_{18} z_{22} z_{56} z_{57} z_{67} z_{75} z_{83} z_{87}	s_1 s_2 s_3 s_4 s_5 s_7 s_8 s_9 s_{10}
		sc$_2$	z_{11} z_{12} z_{14} z_{15} z_{23} z_{55} z_{57} z_{75} z_{84}	s_2 s_3 s_4 s_5 s_6 s_7 s_9 s_{10}
	0.9	sc$_1$	z_{12} z_{14} z_{15} z_{18} z_{22} z_{55} z_{56} z_{62} z_{67}	s_2 s_3 s_4 s_7 s_8 s_9 s_{10}
		sc$_2$	z_{22} z_{23} z_{26} z_{55} z_{61} z_{67} z_{75} z_{84} z_{88}	s_2 s_3 s_4 s_6 s_7 s_9

TAB. 9.7 – Problem III: Characteristics of the optimal robust reserve for three values of the available budget and for two threshold values.

B	ρ	Optimal robust reserve	Cost of the reserve	Protected species (scenario sc$_1$)	Relative regret (scenario sc$_1$)	Protected species (scenario sc$_2$)	Relative regret (scenario sc$_2$)	Maximal relative regret
10	0.8	z_{18} z_{23} z_{67} z_{75}	10	s_2 s_3 s_4 s_{10}	0.2	s_4 s_7 s_9	0.25	0.25
	0.9	z_{22} z_{62} z_{67} z_{75}	9	s_4 s_9	1/3	s_4 s_9	0	1/3
20	0.8	z_{12} z_{18} z_{22} z_{23} z_{55} z_{57} z_{61} z_{67} z_{75}	20	s_2 s_3 s_4 s_7 s_9 s_{10}	1/7	s_2 s_3 s_4 s_7 s_9	1/6	1/6
	0.9	z_{18} z_{22} z_{23} z_{55} z_{57} z_{67} z_{75} z_{84}	20	s_2 s_3 s_4 s_9	0.2	s_2 s_4 s_7 s_9	0	0.2
30	0.8	z_{12} z_{18} z_{22} z_{23} z_{55} z_{57} z_{61} z_{67} z_{75} z_{76} z_{87}	30	s_2 s_3 s_4 s_5 s_7 s_8 s_9 s_{10}	1/9	s_2 s_3 s_4 s_5 s_7 s_8 s_9	0.125	0.125
	0.9	z_{18} z_{22} z_{23} z_{26} z_{34} z_{67} z_{75} z_{84}	28	s_2 s_3 s_4 s_7 s_9 s_{10}	1/7	s_2 s_3 s_4 s_7 s_9	1/6	1/6

$$
\mathrm{P}_{9.7}: \begin{cases} \max\ \alpha & \\ \text{s.t.} \left| \begin{array}{lll} \sum_{i \in \underline{Z}} c_i x_i \leq B & & (9.7.1) \\ \alpha \leq \sum_{k \in \underline{S}} w_k \left(1 - \prod_{i \in \underline{Z}} \left(1 - p_{ik}^{\omega} x_i\right)\right) & \omega \in \underline{\mathrm{Sc}} & (9.7.2) \\ x_i \in \{0,1\} & i \in \underline{Z} & (9.7.3) \end{array} \right. \end{cases}
$$

According to constraints 9.7.2, and since the objective consists in maximizing variable α, this variable takes, at the optimum of $\mathrm{P}_{9.7}$, the smallest of the values $\sum_{k \in \underline{S}} w_k \left(1 - \prod_{i \in \underline{Z}} \left(1 - p_{ik}^{\omega} x_i\right)\right)$ for $\omega \in \underline{\mathrm{Sc}}$, each of these values being equal to the expected weighted number of species that will survive in the selected reserve – formed by zones z_i for which variable x_i is equal to 1 – in the case of scenario sc_{ω}. Using variable μ_k^{ω}, $k \in \underline{S}, \omega \in \underline{\mathrm{Sc}}$, to designate the quantity $1 - \prod_{i \in \underline{Z}} \left(1 - p_{ik}^{\omega} x_i\right)$, program $\mathrm{P}_{9.7}$ can be rewritten as program $\mathrm{P}_{9.8}$.

$$
\mathrm{P}_{9.8}: \begin{cases} \max\ \alpha & \\ \text{s.t.} \left| \begin{array}{lll} \sum_{i \in \underline{Z}} c_i x_i \leq B & & (9.8.1) \\ \alpha \leq \sum_{k \in \underline{S}} w_k \mu_k^{\omega} & \omega \in \underline{\mathrm{Sc}} & (9.8.2) \\ 1 - \mu_k^{\omega} = \prod_{i \in \underline{Z}} \left(1 - p_{ik}^{\omega} x_i\right) & k \in \underline{S}, \omega \in \underline{\mathrm{Sc}} & (9.8.3) \\ x_i \in \{0,1\} & i \in \underline{Z} & (9.8.4) \\ 0 \leq \mu_k^{\omega} \leq 1 & k \in \underline{S}, \omega \in \underline{\mathrm{Sc}} & (9.8.5) \end{array} \right. \end{cases}
$$

Note that, in any feasible solution of $\mathrm{P}_{9.8}$, the value of variable μ_k^{ω} represents the probability $P_k^{\omega}(R)$ where zone z_i belongs to R if and only if $x_i = 1$. Taking the logarithm of the two members of constraints 9.8.3, we obtain the equivalent program $\mathrm{P}_{9.9}$. Recall that it is assumed here that the probabilities p_{ik}^{ω} are all different from 1.

$$
\mathrm{P}_{9.9}: \begin{cases} \max\ \alpha & \\ \text{s.t.} \left| \begin{array}{lll} \sum_{i \in \underline{Z}} c_i x_i \leq B & & (9.9.1) \\ \alpha \leq \sum_{k \in \underline{S}} w_k \mu_k^{\omega} & \omega \in \underline{\mathrm{Sc}} & (9.9.2) \\ \log(1 - \mu_k^{\omega}) = \sum_{i \in \underline{Z}} x_i \log(1 - p_{ik}^{\omega}) & k \in \underline{S}, \omega \in \underline{\mathrm{Sc}} & (9.9.3) \\ x_i \in \{0,1\} & i \in \underline{Z} & (9.9.4) \\ 0 \leq \mu_k^{\omega} \leq 1 & k \in \underline{S}, \omega \in \underline{\mathrm{Sc}} & (9.9.5) \end{array} \right. \end{cases}
$$

The economic function is linear. Constraints 9.9.1 and 9.9.2 are also linear. On the other hand, the left-hand sides of constraints 9.9.3, $\log(1 - \mu_k^{\omega})$, are not linear. We propose below a method, similar to that proposed in section 7.5 of chapter 7, for determining an approximate solution of $P_{9.9}$ with a guarantee on the gap between the value of this solution and the value of the optimal solution. Note that if constraints 9.8.5 and 9.9.5 specify, suitably for mathematical programming, that variables μ_k^ω are less than or equal to 1, these variables will in fact take a value strictly less than 1 in any feasible solution of the corresponding program. The same applies to programs $P_{9.10}$, $P_{9.11}(\omega)$, and $P_{9.12}$.

Let us first consider a relaxation of $P_{9.9}$ (see chapter 7, section 7.5). The values of variables x_i of this relaxation provide a feasible solution of the problem, *i.e.*, a set of zones to be protected, and the optimal value of this relaxation corresponds to an upper bound of the optimal value of the problem, *i.e.*, the best expected weighted number of species that will survive in the worst-case scenario. The relaxation we consider can be interpreted as an upper approximation of the concave function $\log(1-\mu_k^\omega)$ by a concave and piecewise linear function (see Appendix at the end of the book). The relaxation of $P_{9.9}$ is obtained by relaxing constraints 9.9.3. A relaxation of this inequality is obtained by replacing it by the set of linear inequalities $\sum_{i\in\underline{Z}} x_i \log(1-p_{ik}^\omega) \le (1-\mu_k^\omega)/u_v + \log u_v - 1, v=1,\dots,V$, where $u_1, u_2,\dots, u_V$ are constants such that $0 < u_1 < u_2 < \cdots < u_V = 1$. This set of constraints is indeed a relaxation of constraints 9.9.3 since it expresses that the quantity $\sum_{i\in\underline{Z}} x_i \log(1-p_{ik}^\omega)$ is less than or equal to the lower envelope of the V straight lines tangent to the curve $\log(1-\mu_k^\omega)$ at the points of abscissa $u_1, u_2,\dots, u_V$. This relaxation of $P_{9.9}$ is given by $P_{9.10}$. As already noted in section 7.5 of chapter 7, to obtain a good relaxation of $P_{9.9}$, V must be large enough. However, the larger V is, the more constraints have to be taken into account.

$$
P_{9.10}:\left\{\begin{array}{ll}
\max\ \alpha & \\
\text{s.t.} & \left|\begin{array}{lll}
\sum_{i\in\underline{Z}} c_i x_i \le B & & (9.10.1)\\
\alpha \le \sum_{k\in\underline{S}} w_k \mu_k^\omega & \omega\in\underline{\mathrm{Sc}} & (9.10.2)\\
\sum_{i\in\underline{Z}} x_i \log(1-p_{ik}^\omega) & & \\
\qquad \le (1-\mu_k^\omega)/u_v + \log u_v - 1 & k\in\underline{S}, \omega\in\underline{\mathrm{Sc}}, v=1,\dots,V & (9.10.3)\\
x_i\in\{0,1\} & i\in\underline{Z} & (9.10.4)\\
0\le \mu_k^\omega \le 1 & k\in\underline{S}, \omega\in\underline{\mathrm{Sc}} & (9.10.5)
\end{array}\right.
\end{array}\right.
$$

9.6.2 Reserve Respecting a Budgetary Constraint and Minimizing the Maximal Relative Regret, over All Scenarios, About the Expected Number of Species, of a Given Set, that will Survive in it

As already noted in section 8.4 of chapter 8, seeking to protect a set of zones in such a way that the value of that protection is as high as possible in the worst-case scenario can have a significant disadvantage: if one of the scenarios is very "pessimistic" then the set of zones selected will essentially take account of that single scenario. To overcome this disadvantage, we are interested here in determining another type of optimal robust reserve, one that respects a certain budget and that minimizes the maximal relative regret, over all scenarios, about the expected weighted number of species that will survive in this set of zones. First, the optimal reserve – the one that provides the largest expected weighted number of species that will survive in a reserve of cost less than or equal to B – must be determined for each scenario. Let $R^{*\omega}$ be this reserve for scenario sc_ω and $E(R^{*\omega})$ be the corresponding mathematical expectation value. For scenario sc_ω, this problem can be solved – in an approximated way – by program $\mathrm{P}_{9.11}(\omega)$.

$$
\mathrm{P}_{9.11}(\omega):\left\{\begin{array}{l}
\max \displaystyle\sum_{k\in\underline{S}} w_k\mu_k^\omega \\
\text{s.t.}\left|\begin{array}{lll}
\displaystyle\sum_{i\in\underline{Z}} c_i x_i \le B & & (9.11_\omega.1)\\
\displaystyle\sum_{i\in\underline{Z}} x_i \log(1-p_{ik}^\omega) \le (1-\mu_k^\omega)/u_v & & \\
\quad + \log u_v - 1 & k\in\underline{S},\, v=1,\ldots,V & (9.11_\omega.2)\\
x_i\in\{0,1\} & i\in\underline{Z} & (9.11_\omega.3)\\
0\le\mu_k^\omega\le 1 & k\in\underline{S} & (9.11_\omega.4)
\end{array}\right.
\end{array}\right.
$$

The problem of determining an optimal robust reserve can then be solved – in an approximate way – by program $\mathrm{P}_{9.12}$.

$$
\mathrm{P}_{9.12}:\left\{\begin{array}{l}
\min\ \alpha \\
\text{s.t.}\left|\begin{array}{lll}
\displaystyle\sum_{i\in\underline{Z}} c_i x_i \le B & & (9.12.1)\\
\alpha \ge (E(R^{*\omega}) - \displaystyle\sum_{k\in\underline{S}} w_k\mu_k^\omega)/E(R^{*\omega}) & \omega\in\underline{\mathrm{Sc}} & (9.12.2)\\
\displaystyle\sum_{i\in\underline{Z}} x_i \log(1-p_{ik}^\omega) \le (1-\mu_k^\omega)/u_v & & \\
\quad + \log u_v - 1 & k\in\underline{S},\, \omega\in\underline{\mathrm{Sc}},\, v=1,\ldots,V & (9.12.3)\\
x_i\in\{0,1\} & i\in\underline{Z} & (9.12.4)\\
0\le\mu_k^\omega\le 1 & k\in\underline{S},\, \omega\in\underline{\mathrm{Sc}} & (9.12.5)
\end{array}\right.
\end{array}\right.
$$

Example 9.1. Let us take again the instance described in figure 9.1 and assign a weight equal to 1 to each species. Table 9.8 presents the optimal solution of program $\mathrm{P}_{9.11}(\omega)$ and its characteristics, in the case of scenario sc_1, for 3 values of the available budget, 10, 20, and 30. Table 9.9 presents the optimal solution of the same program, in the case of scenario sc_2, for the same values of the available budget. Table 9.10 presents the optimal robust solution, *i.e.*, the one that minimizes the

TAB. 9.8 – Description of the optimal solutions of $P_{9.11}(1)$ (scenario sc_1) for three values of the available budget, B.

B	Selected reserve (R^{*1})	Cost of the reserve	Expected weighted number of protected species ($E(R^{*1})$)	Upper bound (opt. val. of $P_{9.11}(1)$)
10	z_{14} z_{18} z_{22} z_{67}	10	4.80	4.80
20	z_{14} z_{15} z_{16} z_{18} z_{22} z_{55} z_{57} z_{67}	20	7.38	7.38
30	z_{14} z_{15} z_{16} z_{18} z_{22} z_{55} z_{56} z_{57} z_{62} z_{67}	30	8.18	8.18

TAB. 9.9 – Description of the optimal solutions of $P_{9.11}(2)$ (scenario sc_2) for three values of the available budget, B.

B	Selected reserve (R^{*2})	Cost of the reserve	Expected weighted number of protected species ($E(R^{*2})$)	Upper bound (opt. val. of $P_{9.11}(2)$)
10	z_{23} z_{55} z_{57} z_{84}	10	4.50	4.50
20	z_{14} z_{15} z_{22} z_{23} z_{55} z_{57} z_{75} z_{84}	20	6.29	6.29
30	z_{14} z_{15} z_{18} z_{22} z_{23} z_{55} z_{57} z_{75} z_{83} z_{84}	30	7.55	7.55

TAB. 9.10 – Description of the optimal solutions of $P_{9.12}$ (determination of robust reserves) for three values of the available budget, B.

B	Optimal robust reserve	Cost of the reserve	Expected weighted number of protected species in the case of scenario sc_1	Expected weighted number of protected species in the case of scenario sc_2	Maximal regret	Optimal value of $P_{9.12}$
10	z_{15} z_{22} z_{23} z_{55} z_{57} z_{67}	10	4.60	4.22	0.06	0.06
20	z_{14} z_{15} z_{16} z_{18} z_{22} z_{23} z_{67} z_{75}	20	7.16	6.05	0.04	0.04
30	z_{14} z_{15} z_{16} z_{18} z_{22} z_{23} z_{55} z_{57} z_{67} z_{75} z_{84}	29	8.06	7.32	0.03	0.03

maximal relative regret, over all scenarios, regarding the expected weighted number of species that will survive. This optimal robust solution is determined by solving program $P_{9.12}$ in which the quantities $E(R^{*\omega})$, $\omega = 1{,}2$, are set to the values presented in tables 9.8 and 9.9.

References and Further Reading

Arthur J.L., Haight R.G., Montgomery C.A., Polasky S. (2002) Analysis of the threshold and expected coverage approaches to the probabilistic reserve site selection problem, *Environ. Model. Assess.* **7**, 81.

Arthur J.L., Camm J., Haight R.G., Montgomery C.A., Polasky S. (2004) Weighing conservation objectives: maximum expected coverage versus endangered species protection, *Ecol. Appl.* **14**, 1936.

Billionnet A. (2011) Solving the probabilistic reserve selection problem, *Ecol. Model.* **222**, 546.

Camm J.D., Norman S.K., Polasky S., Solow A.R. (2002) Nature reserve site selection to maximize expected species covered, *Oper. Res.* **50**, 946.

Carvalho S.B., Brito J.C., Crespo E.G., Watts M.E., Possingham H.P. (2011) Conservation planning under climate change: toward accounting for uncertainty in predicted species distributions to increase confidence in conservation investments in space and time, *Biol. Conserv.* **144**, 2020.

Drechsler M. (2005) Probabilistic approaches to scheduling reserve selection, *Biol. Conserv.* **122**, 253.

Fourer R. (1992) A simplex algorithm for piecewise-linear programming III: computational analysis and applications, *Math. Program.* **53**, 213.

Fourer R., Marsten R.E. (1992) Solving piecewise-linear programs: experiments with a simplex approach, *ORSA J. Comput.* **4**, 16.

Haider Z., Charkhgard H., Kwon C. (2018) A robust optimization approach for solving problems in conservation planning, *Ecol. Model.* **368**, 288.

Haight R.G., ReVelle C.S., Snyder S.A. (2000) An integer optimization approach to a probabilistic reserve site selection problem, *Oper. Res.* **48**, 697.

Haight R.G., Travis L.E. (2008) Reserve design to maximize species persistence, *Environ. Model. Assess.* **13**, 243.

Haight R.G., Snyder S.A. (2009) Integer programming methods for reserve selection and design, *Spatial conservation prioritization* (A. Moilanen, K.A. Wilson, H.P. Possingham, Eds). Oxford University Press.

Kouvelis P., Yu G. (1997) *Robust discrete optimization and its applications*. Kluwer Academic Publishers, The Netherlands.

Margules C.R., Nicholls A.O., Usher M.B. (1994) Apparent species turnover, probability of extinction and the selection of nature reserves: a case study of the Ingleborough Limestone Pavements, *Conserv. Biol.* **8**, 398.

Moilanen A., Runge M., Elith J., Tyre A., Carmel Y., Fegraus E., Wintle B., Burgman M., Ben-Haim Y. (2006) Planning for robust reserve networks using uncertainty analysis, *Ecol. Model.* **199**, 115.

Moilanen A., Wintle B.A., Elith J., Burgman M. (2006) Uncertainty analysis for regional-scale reserve selection, *Conserv. Biol.* **20**, 1688.

Polasky S., Camm J.D., Solow A.R., Csuti B., White D., Ding R. (2000) Choosing reserve networks with incomplete species information, *Biol. Conserv.* **94**, 1.

Sarkar S., Pappas C., Garson J., Aggarwal A., Cameron S. (2004) Place prioritization for biodiversity conservation using probabilistic surrogate distribution data, *Divers. Distrib.* **10**, 125.

Snyder S.A., Haight R.G., ReVelle C.S. (2004) A scenario optimization model for dynamic reserve site selection, *Environ. Model. Assess.* **9**, 179.

Strange N., Thorsen B.J., Bladt J. (2006) Optimal reserve selection in a dynamic world, *Biol. Conserv.* **131**, 33.

Williams P.H., Araujo M.B. (2000) Using probability of persistence to identify important areas for biodiversity conservation, *Proc. Phys. Soc. London* **B 267**, 1959.

Chapter 10

Phylogenetic Diversity

10.1 Introduction

The rate of biodiversity loss is increasing. If it is not slowed down, this phenomenon will have disastrous consequences in many sectors. Numerous projects are currently being developed to try to remedy this situation. In order to carry out these projects, precise measurements of biodiversity are necessary, especially to be able to focus on the most effective strategies, since the resources are of course limited. The simple measures that are commonly used are the species richness and species abundance of a zone. The former refers to the number of species and the latter to the number of individuals of each species. These criteria have been used extensively in the previous chapters, but many authors consider that protected areas could considerably increase their effectiveness by taking into account other criteria. An interesting measure for assessing a set of species from a biodiversity perspective, and which is increasingly used in the field of conservation, is "phylogenetic diversity" (PD). This is based on the notion of a phylogenetic tree associated with the set of species under consideration and reflects the evolutionary history of these species. Some experts consider the rate of phylogenetic diversity loss to be even greater than the rate of species diversity loss. There are different ways of defining phylogenetic diversity. Here we adopt Faith's definition (1992a): the phylogenetic diversity of a set of species is equal to the sum of the lengths of the branches of the phylogenetic tree associated with that set. This measure is a good reflection of the evolutionary history of the set of species considered and has been much studied in the biological conservation literature.

10.2 Phylogenetic Tree

A phylogenetic tree is a tree, in the sense of graph theory (see appendix at the end of the book). It can be defined, for example, as a connected graph without cycles. It is composed, on the one hand, of internal vertices and, on the other hand, of leaves that

DOI: 10.1051/978-2-7598-2521-9.c010

represent the species – there is a one-to-one correspondence between the leaves of the tree and the species under consideration. Some trees have a root, others do not. Here we are interested in trees with a root. In such trees, there is an implicit orientation of the branches from the root to the leaves. These phylogenetic trees can, therefore, be considered as arborescences, in the sense of graph theory (see appendix at the end of the book), each branch of the tree being in fact an arc with an initial and a terminal end. Thus, the course of a path, from the root to a leaf, follows the arcs of this path from their initial end to their terminal end. Throughout this chapter, we will use the terms phylogenetic "tree" and "branch" or "arc". Each branch of the tree has a value associated with it, called the branch length. This length reflects the accumulation of evolutionary changes that have occurred from the initial end of the branch to its terminal end or, more simply, the elapsed time. In the case where it reflects evolutionary changes, these are related to particular characteristics, morphological or molecular, chosen to construct the tree. In a phylogenetic tree, an internal vertex has all the characteristics common to all its descendants. This internal vertex can be considered as a common ancestor for all its descendants. In summary, the length of a branch provides an overall indicator of the amount of evolution that has taken place between its two ends. Throughout this chapter, a phylogenetic tree will be represented by the quadruplet (V, A, S, λ) where V is the set of vertices, $A = \{a_1, \ldots, a_r\}$, the set of arcs – also called branches –, $S = \{s_1, \ldots, s_m\}$, the set of leaves – the species – and $\lambda = \{\lambda_1, \ldots, \lambda_r\}$, the set of branch lengths. We denote by $\underline{A}$ the set of indices of the arcs and $\underline{S}$ the set of indices of the species. Thus, $\underline{A} = \{1, \ldots, r\}$ and $\underline{S} = \{1, \ldots, m\}$.

A tree is said to be ultrametric if the lengths of all the paths connecting the root to a leaf – a species – are identical. Recall that, by definition, the length of a path is equal to the sum of the lengths of the arcs composing it. For example, a phylogenetic tree in which the length of the branches represents the elapsed time is ultrametric. Figure 10.1 gives two slightly different representations of an ultrametric phylogenetic tree. Figure 10.2 shows a non-ultrametric tree. In all the cases, the length of a branch reflects the extent of evolutionary changes that have occurred between the two ends of the branch.

10.3 Phylogenetic Diversity (PD)

The phylogenetic diversity (PD) of a set of species, S, is equal to the sum of the lengths of the branches of the phylogenetic tree associated with that set. The phylogenetic diversity of a subset of species, $\hat{S} \subseteq S$, is equal to the sum of the lengths of the branches of the smallest sub-tree – of the phylogenetic tree associated with S – which links all the species of this subset as well as the root of the tree. In other words, the phylogenetic diversity of $\hat{S}$ is equal to the sum of the lengths of the branches for which there is at least one path from the terminal end of that branch to one of the species of $\hat{S}$. Figures 10.3 and 10.4 illustrate this notion of phylogenetic diversity.

Let S be a set of species, and S_1 and S_2 two subsets of S. The phylogenetic diversity criterion implies that the set S_1 is more "interesting" than the set S_2 – from the biodiversity point of view – if the phylogenetic diversity of S_1 is greater than that

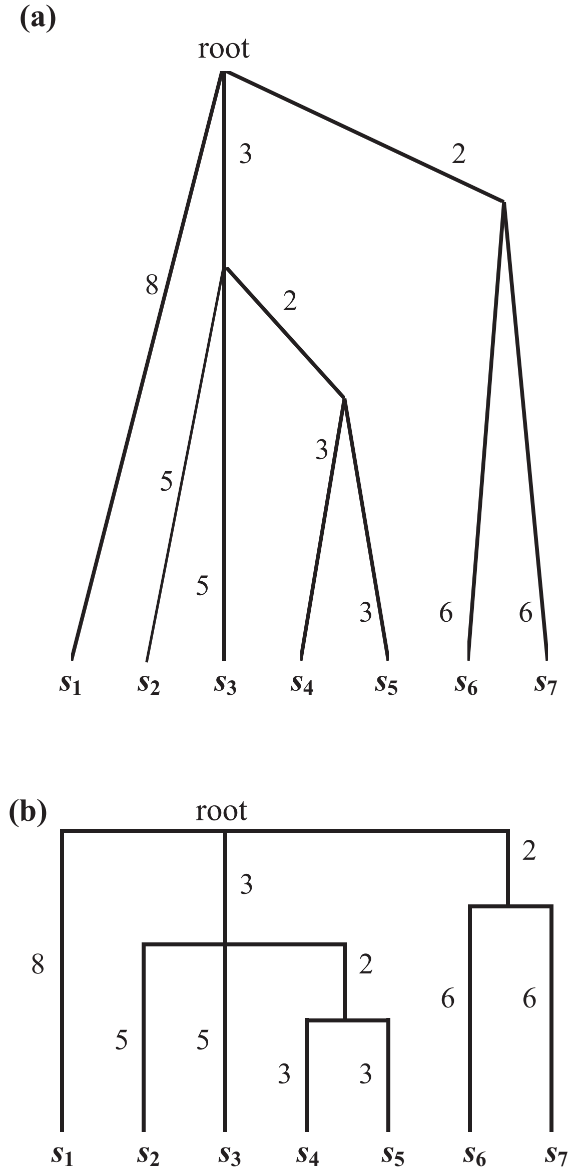


FIG. 10.1 – Two slightly different representations of a hypothetical ultrametric phylogenetic tree with 7 species and 10 branches. The length of each branch is indicated next to the branches. In representation (b) the lengths of the segments representing the branches are proportional to the lengths of the branches.

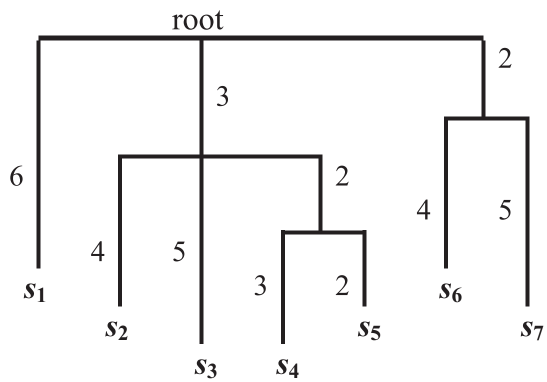


FIG. 10.2 – A hypothetical, non-ultrametric phylogenetic tree with 7 species and 10 branches. The length of each branch is indicated next to the branches. In this figure, the lengths of the segments representing the branches are proportional to the lengths of the branches.

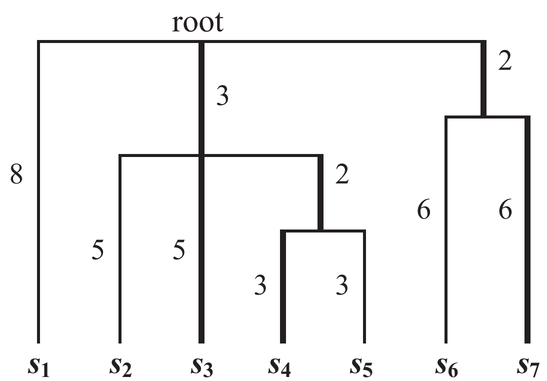


FIG. 10.3 – A hypothetical phylogenetic tree – ultrametric – associated with 7 species, s_1, s_2,…, s_7; it has 10 branches. The length of each branch is indicated next to the branch. The phylogenetic diversity of the complete set of species, $\{s_1,..., s_7\}$, is equal to 43. The phylogenetic diversity of the subset of species $\{s_3, s_4, s_7\}$ is 21 and the 6 branches involved in its calculation are shown in bold.

of S_2. Indeed, in this case, the evolutionary history accumulated by the set of species S_1 is greater than that accumulated by the set of species S_2, and the biodiversity associated with S_1 is then considered to be greater than that associated with S_2. In other words, the disappearance of a species with a long evolutionary history – a species linked to the tree root by a long path – and few living related species, would cause more biodiversity loss than the disappearance of a recently appeared species with living related species. Consider, for example, figure 10.3 and the 7 associated living species. The loss of species s_1 is estimated to be more detrimental to

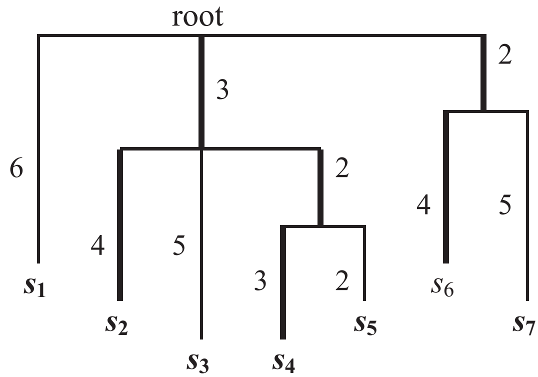


FIG. 10.4 – A hypothetical, non-ultrametric phylogenetic tree, associated with 7 species s_1, s_2,..., s_7; it has 10 branches. The length of each branch is indicated next to the branch. The phylogenetic diversity of the complete set of species, $\{s_1,..., s_7\}$, is equal to 36. The phylogenetic diversity of the subset of species $\{s_2, s_4, s_6\}$ is equal to 18 and the 6 branches involved in its calculation are shown in bold.

biodiversity than the loss of species s_4. Indeed, the disappearance of s_1 results in the loss of 8 evolutionary time units while the disappearance of s_4 results in the loss of only 3 – assuming that s_5 survives.

Note that the problem of determining, knowing the phylogenetic tree associated with a set of species $S = \{s_1, s_2,..., s_m\}$, a subset of species, $\hat{S} \subseteq S$, of given cardinal and maximal PD can be easily solved by a greedy algorithm. This algorithm starts with an empty set, $\hat{S}$, then consists in enriching $\hat{S}$ by progressively adding the species, one after the other. At each step of this algorithm, species s that maximizes the PD of the set of species $\hat{S} \cup \{s\}$ is added to the already obtained set $\hat{S}$ until the set $\hat{S}$ contains the desired number of species. Let us apply this algorithm to the set of species in figure 10.3 to determine a subset of 4 species of maximal PD. This yields, for example, the 4 species s_1, s_2, s_6, and s_7, and the PD of this set is 30. Figure 10.5 shows the phylogenetic tree associated with 13 species of otters (Lutrinae), s_1, s_2, ..., s_{13}, constructed from the data in (Bininda-Edmonds *et al.*, 1999). This tree has 21 branches or arcs, a_1, a_2,..., a_{21}, and 8 internal nodes, i_1, i_2,..., i_8. The lengths of the branches are given next to them, in millions of years. Which are the 5 species of otters – among the 13 considered – of maximal PD? The greedy algorithm provides the set $\{s_1, s_2, s_5, s_8, s_{10}\}$ of PD 46.6.

10.4 Expected Phylogenetic Diversity (ePD)

10.4.1 Definition

Let us consider a phylogenetic tree associated with a set of species $S = \{s_1, s_2,..., s_m\}$ and, to each species s_k of this set, let us associate a survival probability

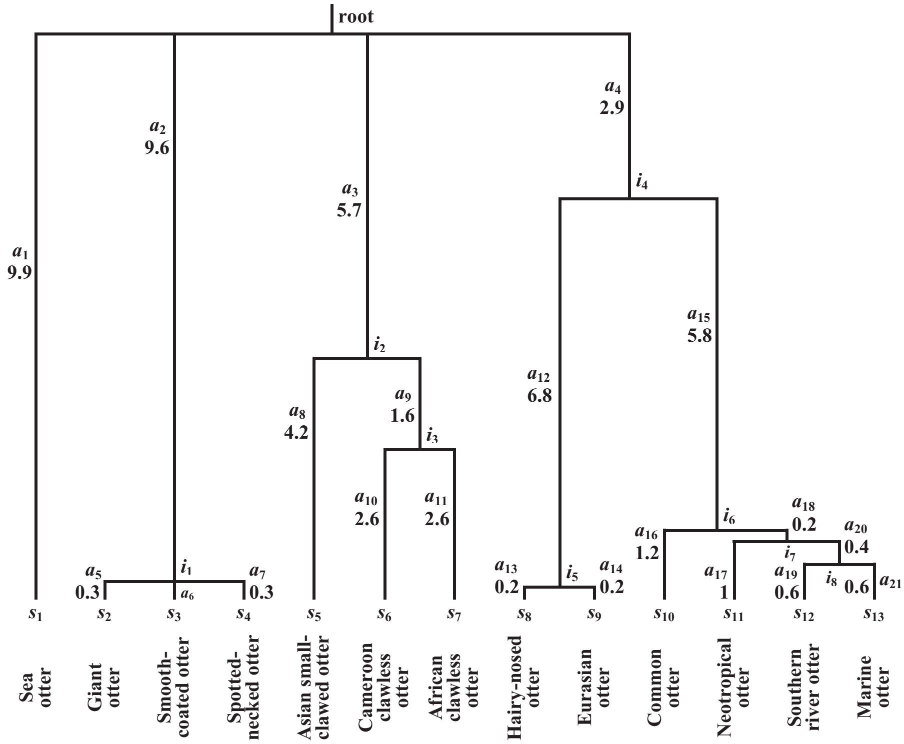


FIG. 10.5 – An ultrametric phylogenetic tree associated with 13 species of otters s_1, s_2,..., s_{13}; it has 21 branches. The length of each branch is indicated next to the branch. The length of the segments representing the branches is proportional to the length of the branches, *i.e.*, the elapsed time. The PD of the set of species s_1, s_2, s_5, s_8, and s_{10} is equal to 46.6. It is a set of 5 species of maximal PD.

denoted by p_k. It is assumed here that all these probabilities are independent of each other. Note that these probabilities are often difficult to estimate. The expected phylogenetic diversity (ePD) of this set of species is, by definition, the sum, on all the branches of the associated phylogenetic tree, of the probabilities that the information represented by this branch is retained, multiplied by the length of the branch. Species s_k is said to be located under branch a_l if and only if there is a path from the terminal end of a_l to species s_k, and the set of species under branch a_l is denoted by F_l. The probability that the information represented by branch a_l is retained, until a certain date, is equal to the probability that at least one of the species of F_l will survive until that date. Indeed, all the species of F_l retain the evolutionary history represented by branch a_l. The probability that the information represented by branch a_l is lost is, therefore, equal to $\prod_{k \in \underline{F}_l} (1 - p_k)$ and the probability that the information represented by branch a_l is retained is, therefore, equal to $1 - \prod_{k \in \underline{F}_l} (1 - p_k)$ where $\underline{F}_l$ denotes the set of indices of the species located under the arc a_l. Finally, the expected phylogenetic diversity of a set of species is equal to $\sum_{l\in\underline{A}} \lambda_l(1 - \prod_{k\in\underline{F}_l}(1 - p_k))$ where A is the set of branches of the phylogenetic tree associated with these species and $\underline{A}$ is the set of corresponding indices. Regarding the probabilities of retaining the information associated with the tree branches, taking into account the survival of the species, we will use, for each branch a_l, the following 3 notations in everything that follows:

σ_l: probability that the information associated with branch a_l is not retained;

$\hat{\sigma}_l$: logarithm of the probability that the information associated with branch a_l is not retained ($\hat{\sigma}_l = \log \sigma_l$);

$\tilde{\sigma}_l$: an approximation of the probability that the information associated with branch a_l is not retained, *i.e.*, of σ_l.

In order to measure the trend of phylogenetic diversity of a set of species to dispersal around its mathematical expectation, one can look at the variance of phylogenetic diversity of this set. Let us denote by $\beta(l)$ the set of indices of branches below branch a_l. Branch $a_{l'}$ is said to be located below branch a_l if and only if there is a path from the terminal end of a_l to the initial end of $a_{l'}$. The variance of phylogenetic diversity of the set of species associated with the tree whose set of branches is A can be written as:

$$\sum_{l\in\underline{A}} \lambda_l^2 \sigma_l(1-\sigma_l) + \sum_{(l,l')\,\in\underline{A}^2:\; l'\in\beta(l)} 2\lambda_l\lambda_{l'}(1-\sigma_{l'})\sigma_l.$$

This expression is established without difficulty using the following classical property concerning the variance of a sum of random variables: if X is a random variable equal to the sum of n random variables X_1, X_2,..., X_n, the variance of X is equal to $\sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} [\mathrm{Esp}(X_iX_j) - \mathrm{Esp}(X_i)\mathrm{Esp}(X_j)]$ where Var (α) designates the variance of the random variable α and $\mathrm{Esp}(\alpha)$, its mathematical expectation. Note that if X_i and X_j are independent, then $\mathrm{Esp}(X_iX_j) = \mathrm{Esp}(X_i)\mathrm{Esp}(X_j)$. In the case of the random variable representing phylogenetic diversity, the random variables associated with branches a_l and $a_{l'}$ are independent if and only if $l \notin \beta(l')$ and $l' \notin \beta(l)$. An example of a detailed calculation of the variance of phylogenetic diversity of a set of species is presented in section 10.6.4.

10.4.2 Example

Let us take again the phylogenetic tree of figure 10.1b and suppose that the 7 species concerned are more or less threatened. Let us associate a survival probability with each of these species (figure 10.6).

To quantify, in this example, the survival probabilities of species, we inspired ourselves from the definition of the 5 categories of threatened species defined by the International Union for the Conservation of Nature (IUCN): Critically Endangered (CR), Endangered (EN), Vulnerable (VU), Near Threatened (NT), Least Concern (LC). Table 10.1 gives, for each category, an interval within which we consider the survival probability may lie. The survival probabilities finally retained for the 7 species are shown in figure 10.6.

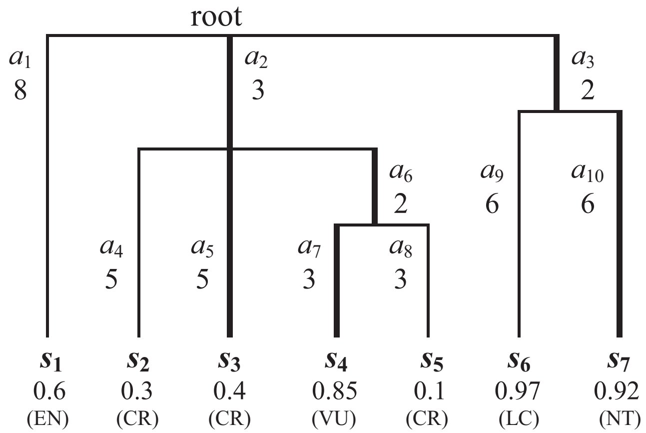


FIG. 10.6 – A hypothetical phylogenetic tree (ultrametric) associated with 7 species $s_1, s_2, \ldots, s_7$; it has 10 branches, $a_1, \ldots, a_{10}$. The length of each branch is indicated next to the branch. For each species, the hypothetical IUCN category to which it belongs (see table 10.1) as well as its survival probability is indicated below each species. The ePD associated with the 7 species $s_1, \ldots, s_7$ is equal to 29.0451 (see table 10.2 for details of the calculation). The associated variance is equal to 37.6194.

TAB. 10.1 – Hypothetical possible values of survival probabilities for each of the 5 categories of threatened species defined by IUCN.

Category	Interval of survival probabilities
CR (Critically Endangered)	[0, 0.50]
EN (Endangered)	[0.50, 0.80]
VU (Vulnerable)	[0.80, 0.90]
NT (Near Threatened)	[0.90, 0.95]
LC (Least Concern)	[0.95, 1]

Many problems arise regarding the selection of protected zones for the protection of species when the criterion of phylogenetic diversity or the criterion of expected phylogenetic diversity is used. Some of these problems are discussed later in the chapter (sections 10.7–10.10). A basic problem concerning the selection of species to be protected in order to maximise the resulting ePD, without consideration of zones to be protected, is presented first in section 10.5. We then present a generalization of this problem in section 10.6.

TAB. 10.2 – Detail of the calculation of the expected phylogenetic diversity of the set of species $\{s_1,\ldots, s_7\}$ in figure 10.6.

Branch	Probability of keeping the information represented by the branch	Branch length	Branch contribution to the value of ePD
a_1	0.6	8	4.8
a_2	1−(1−0.3)(1−0.4)(1−0.85)(1−0.1) = 0.9433	3	2.8299
a_3	1−(1−0.97)(1−0.92) = 0.9976	2	1.9952
a_4	0.3	5	1.5
a_5	0.4	5	2.0
a_6	1−(1−0.85)(1−0.1) = 0.865	2	1.73
a_7	0.85	3	2.55
a_8	0.1	3	0.3
a_9	0.97	6	5.82
a_{10}	0.92	6	5.52

10.5 Noah's Ark Problem

10.5.1 Definition

In this problem, we consider a set of species for which we know the survival probabilities. The carrying out of certain conservation actions can increase these probabilities but these actions have a cost. The problem is to allocate resources – thus increasing the survival probability of certain species – as efficiently as possible. Effectiveness is measured by the expected phylogenetic diversity of the set of species under consideration that is obtained as a result of the conservation actions carried out. Let us consider the phylogenetic tree associated with the set of species $S = \{s_1, s_2,\ldots, s_m\}$ and let us denote by ϕ_k^1 the initial survival probability of species s_k, $k \in \underline{S} = \{1,\ldots, m\}$. As mentioned, it is assumed that certain actions can influence the survival probability of the species. For example, it is possible to increase the survival probability of species s_k, from ϕ_k^1 to a higher value, ϕ_k^2, but this has a cost, denoted by α_k. The problem considered here – known as the "Noah's Ark problem" in the biological conservation literature – consists in choosing the species whose survival probability will be increased in order to maximize the expected PD associated with the set of species under consideration while respecting a budgetary constraint. In some versions of this problem $\phi_k^1 = 0$ for all $k \in \underline{S}$. As mentioned above, survival probabilities are difficult to estimate in general. The same is even more true for estimating these probabilities in view of the protection actions undertaken.

10.5.2 Mathematical Programming Formulation

Let t_k be a Boolean variable which is equal to 1 if and only if we decide to increase the survival probability of species s_k, which costs α_k. The survival probability of species s_k is expressed, as a function of variables t_k, by $t_k\phi_k^2 + (1 - t_k)\phi_k^1$. The extinction probability of species s_k is therefore equal to $1 - t_k\phi_k^2 - (1 - t_k)\phi_k^1$. It is deduced that the probability that the information associated with the arc a_l is kept is equal to $1 - \prod_{k\in\underline{F}_l}(1 - t_k\phi_k^2 - (1 - t_k)\phi_k^1)$, and finally that the expected PD – expressed as a function of variables t_k – is equal to $\sum_{l\in\underline{A}}\lambda_l(1 - \prod_{k\in\underline{F}_l}$ $(1 - t_k\phi_k^2 - (1 - t_k)\phi_k^1)$. The solution of Noah's Ark problem can, therefore, be obtained by solving the mathematical program in Boolean variables $\mathrm{P}_{10.1}$.

$$\mathrm{P}_{10.1}: \begin{cases} \max \sum_{l\in\underline{A}} \lambda_l \left(1 - \prod_{k\in\underline{F}_l} \left(1 - t_k\phi_k^2 - (1 - t_k)\phi_k^1\right)\right) \\ \text{s.t.} \left| \begin{array}{lll} \sum_{k\in\underline{S}} \alpha_k t_k \le B & & (10.1.1) \\ t_k \in \{0,1\} & k \in \underline{S} & (10.1.2) \end{array} \right. \end{cases}$$

Program $\mathrm{P}_{10.1}$ consists of maximizing a non-linear economic function subject to a linear constraint. The economic function expresses the ePD of the set of species considered, taking into account the conservation actions decided on. Constraint 10.1.1 expresses that the total cost of these actions must not exceed the available budget, B. We give below the mixed-integer linear program $\mathrm{P}_{10.2}$ which allows to determine both an approximate solution of Noah's Ark problem and an upper bound of the optimal value of this problem. We comment briefly on program $\mathrm{P}_{10.2}$ but, for a more detailed explanation of the technique used to construct this program, the reader may refer to section 7.5 of chapter 7 and also to sections 10.8.2. and 10.8.3 of this chapter which deals with a closely related problem. The method of section 10.8.3 can indeed easily be extended to solving Noah's Ark problem, *i.e.*, to solving $\mathrm{P}_{10.1}$. The set of pending arcs of the phylogenetic tree under consideration, *i.e.*, the set of arcs whose terminal end represents a species, is designated by A_p. The set of corresponding indices is designated by $\underline{A}_p$. For any pending arc a_l of the tree, $\mathrm{ext}(l)$ designates the index of the species associated with the terminal end of this arc.

$$\mathrm{P}_{10.2}: \begin{cases} \max \sum_{l\in\underline{A}-\underline{A}_p} \lambda_l(1 - \tilde{\sigma}_l) + \sum_{l\in\underline{A}_p,\, k=\mathrm{ext}(l)} \lambda_l\,[\phi_k^2 t_k + \phi_k^1(1 - t_k)] \\ \text{s.t.} \left| \begin{array}{lll} \sum_{k\in\underline{S}} \alpha_k t_k \le B & & (10.2.1) \\ \hat{\sigma}_l \le \frac{\tilde{\sigma}_l}{u_v} + \log u_v - 1 & l \in \underline{A} - \underline{A}_p,\ v = 1,\dots, V & (10.2.2) \\ \hat{\sigma}_l = \sum_{j\in\underline{\mathrm{As}}_l} \hat{\sigma}_j & l \in \underline{A} - \underline{A}_p & (10.2.3) \\ \hat{\sigma}_l = \log(1 - \phi_k^2)t_k + \log(1 - \phi_k^1)(1 - t_k) & l \in \underline{A}_p,\ k = \mathrm{ext}(l) & (10.2.4) \\ t_k \in \{0,1\} & k \in \underline{S} & (10.2.5) \\ \tilde{\sigma}_l \ge 0,\ \hat{\sigma}_l \le 0 & l \in \underline{A} - \underline{A}_p & (10.2.6) \end{array} \right. \end{cases}$$

The variable, real and negative or null, $\hat{\sigma}_l$ represents the logarithm of the probability that the information associated with the arc a_l is lost and the variable, real, positive or null and less than or equal to 1, $\tilde{\sigma}_l$ represents an approximation of this probability. The first part of the economic function expresses an approximate value of the contribution of the non-pending arcs to the ePD and the second part expresses the contribution of the pending arcs. Constraint 10.2.1 expresses the budget constraint. Constraint 10.2.2 makes it possible to obtain, for each non pending arc a_l of the tree, the value of $\tilde{\sigma}_l$ knowing the value of $\hat{\sigma}_l$. The coefficients $u_1, u_2, \ldots, u_V$ are real numbers such that $0 < u_1 < u_2 < \cdots < u_V = 1$. As_l designates the set of arcs whose initial end coincides with the terminal end of the arc a_l and $\underline{\mathrm{As}}_l$ designates the set of corresponding indices. Constraints 10.2.3 express, for each non pending arc a_l of the tree, that the logarithm of the probability that the information associated with this arc is lost ($\hat{\sigma}_l$) is equal to the sum, over all the successor arcs of a_l, of the logarithms of the probabilities that the information associated with these arcs is lost. Constraints 10.2.4 express, for any pending arc in the tree, the logarithm of the probability that the information associated with that arc is lost. Constraints 10.2.5 and 10.2.6 specify the nature of the variables.

10.5.3 Remarks

10.5.3.1 Special Cases Where Noah's Ark Problem can be Solved by a Greedy Algorithm

Some special cases of the problem of selecting a subset of species to be protected, $\hat{S} \subseteq S = \{s_1, s_2, \ldots, s_m\}$, of a given cardinal, in order to maximize the expected phylogenetic diversity of the species of S, can be easily solved by a greedy algorithm. Let us consider two such cases. In both cases, the survival probability of unprotected species is equal to ϕ_k^1, $k \in \underline{S}$. In the first case, the survival probability of protected species is equal to 1 and in the second case, this probability is equal to $1 - \rho(1 - \phi_k^1)$, $k \in \underline{S}$, ρ being a multiplying coefficient independent of k and ranging between 0 and 1. In both cases, the algorithm starts with an empty set, $\hat{S}$, then consists in enriching $\hat{S}$ by progressively adding to it the species, one after the other. At each step of this algorithm, species s that maximizes the ePD of the set of species $\hat{S} \cup \{s\}$ is added to the already obtained set $\hat{S}$ until the set $\hat{S}$ contains the desired number of species.

10.5.3.2 Influence of the Initial Survival Probability Values

Consider the general problem of choosing, from a set of threatened species, S, a subset of species, $\hat{S}$, of a given cardinal, whose protection maximizes the resulting ePD of S. Here we consider that the survival of the protected species is assured, whereas this is not the case for the unprotected species. Thus, the survival probability of all protected species is equal to 1 and that of unprotected species is known and equal to p_k for species s_k. As we have seen in section 10.5.3.1, the problem is easy to solve by a greedy algorithm. This algorithm starts with an empty set, $\hat{S}$, and then consists in enriching $\hat{S}$ by progressively adding species to it, one after the other. At each step of this algorithm, species s that maximizes the PD of the set of species $\hat{S} \cup \{s\}$ is added to the already obtained set $\hat{S}$ until the set $\hat{S}$ contains the desired number of species. The difficulty with this problem is that the extinction probabilities of the different species are difficult to quantify. We found that the choice of the precise value assigned to the extinction probability of each species is not as important as one might think – for the problem under consideration. However, this choice should follow the natural rule: if species s_k is more threatened than species s_j then the extinction probability of species s_k should be larger than the extinction probability of species s_j. For example, the species concerned can be considered to be divided into categories corresponding to more or less threatened species as is done in the IUCN Red List (Critically Endangered, Endangered, Vulnerable, Near Threatened, Least Concern). We found in our experiments that the values of these probabilities have a significant influence on the set of species that are selected for protection but little influence on the resulting ePD. Specifically, if S^{1*} and S^{2*} are the two optimal subsets of species selected for protection corresponding to two different scenarios – two different sets of extinction probabilities – the ePDs of S^{1*} and S^{2*} calculated with the probabilities of scenario sc_1 – or with the probabilities of scenario sc_2 – are not very different. We have randomly generated 10,000 different instances of the problem in the following way:

- A phylogenetic tree, T, is generated. The number of non-leaf nodes of T is randomly and uniformly generated between 50 and 1,000;
- The length of each branch of T is randomly and uniformly generated in a set of possible values;
- Two sets of extinction probabilities (corresponding to two hypothetical scenarios) are randomly and uniformly generated in a set of possible values for the species associated with T;
- The maximum number of species that can be protected is equal to the greatest integral number less than or equal to ρ multiplied by the number of species of T, and ρ is randomly and uniformly generated from the set $\{0.1, 0.2, 0.3, 0.4, 0.5\}$.

For each of these instances, we calculated the optimal set of species to be protected in scenario sc_ω, $S^{\omega*}, \omega = 1, 2$, and then the two relative gaps gap^1 and gap^2 defined below. For each subset $\hat{S}$ of S and for each scenario sc_ω, we denote by $\text{ePD}^\omega(S, \hat{S})$ the ePD of S generated by the protection of the species of $\hat{S}$ and calculated with the probabilities of scenario sc_ω. We are first interested in the case where a decision is made to protect the species of S^{1*} and we calculate the resulting ePD with the two sets of probabilities, *i.e.*, $\text{ePD}^1(S, S^{1*})$ and $\text{ePD}^2(S, S^{1*})$. We then consider the case where it is decided to protect the species from S^{2*} and calculate the resulting ePD using both sets of probabilities, *i.e.*, $\text{ePD}^1(S, S^{2*})$ and $\text{ePD}^2(S, S^{2*})$. We then compute the two relative gaps gap^1 and gap^2:

$$\text{gap}^1 = [\text{ePD}^1(S, S^{1*}) - \text{ePD}^1(S, S^{2*})]/\text{ePD}^1(S, S^{1*})$$

$$\text{gap}^2 = [\text{ePD}^2(S, S^{2*}) - \text{ePD}^2(S, S^{1*})]/\text{ePD}^2(S, S^{2*}).$$

The first of these gaps provides the relative error if scenario sc_1 occurs when the choice of the species to be protected has been optimized based on scenario sc_2. Conversely, the second relative gap provides the relative error if scenario sc_2 occurs when the choice of the species to be protected has been optimized based on scenario sc_1. We have thus calculated gap^1 and gap^2 for the 10,000 instances considered and the largest gap obtained – among the 20,000 calculated – is equal to 2.6%. Let S^{1*} and S^{2*} be the two optimal subsets associated with the instance corresponding to the largest relative gap. The cardinal of the intersection of these two sets, $\left| S^{1*} \cap \ S^{2*} \right|$, divided by the cardinal of the union of these two sets, $\left| S^{*1} \cup \ S^{*2} \right|$, is equal to 0.3. The two sets S^{1*} and S^{2*} corresponding to the largest relative gap are thus very different since their intersection includes only 30% of the species concerned by S^{1*} or S^{2*}.

10.6 The Generalized Noah's Ark Problem

10.6.1 Definition

In this generalisation, Π_k different conservation policies can be envisaged for each species s_k, $k \in \underline{S}$. The policy π, $\pi \in \{1,\ldots,\Pi_k\}$, applied to species s_k consists in allocating a certain quantity of monetary units, denoted by α_k^π, to the protection of this species, which leads to a certain survival probability of this same species, noted ϕ_k^π. The cost α_k^1 is equal to 0 and the probability ϕ_k^1 therefore corresponds to the case where no action is carried out for the protection of species s_k. Note that, in this model, the amount invested in the protection of a species must belong to a finite and predetermined set of values. It is not possible to invest any amount. The problem is to determine the conservation strategy – the conservation policy to be applied to each species – that maximises the expected phylogenetic diversity of the set of species under consideration, while respecting a budgetary constraint.

10.6.2 Mathematical Programming Formulation

Using the Boolean variables t_k^π that are equal to 1 if and only if the conservation policy π is applied to species s_k, the generalized Noah's Ark problem can be formulated as the mathematical program in Boolean variables $P_{10.3}$.

$$\mathrm{P}_{10.3}:\begin{cases}\max \displaystyle\sum_{l\in \underline{A}} \lambda_l\left(1-\prod_{k\in \underline{F}_l}(1-\sum_{\pi=1}^{\Pi_k}\varphi_k^\pi t_k^\pi)\right) & & \\ \text{s.t.}\left|\begin{array}{lll}\displaystyle\sum_{k\in \underline{S}}\sum_{\pi=1}^{\Pi_k}\alpha_k^\pi t_k^\pi \le B & & (10.3.1)\\ \displaystyle\sum_{\pi=1}^{\Pi_k} t_k^\pi = 1 & k\in \underline{S} & (10.3.2)\\ t_k^\pi\in\{0,1\} & k\in \underline{S},\ \pi=1,\ldots,\Pi_k & (10.3.3)\end{array}\right.\end{cases}$$

The economic function of $\mathrm{P}_{10.3}$ represents the expected phylogenetic diversity, taking into account the chosen conservation strategy. Constraint 10.3.1 expresses the budgetary constraint. Constraint 10.3.2 expresses that one and only one conservation policy must be applied to each species. This program is difficult to resolve because of the non-linearity of the economic function. In the following section, we present a method for obtaining a solution to the problem that is close to the optimal solution.

10.6.3 Resolution

We give below the mixed-integer linear program $\mathrm{P}_{10.4}$ which allows the determination of an approximate solution of the generalized Noah's Ark problem as well as an upper bound of the optimal value of this problem.

$$\mathrm{P}_{10.4}:\begin{cases}\max \displaystyle\sum_{l\in \underline{A}-\underline{A}_p}\lambda_l(1-\tilde{\sigma}_l)+\sum_{l\in \underline{A}_p,\ k=\mathrm{ext}(l)}\lambda_l\sum_{\pi=1}^{\Pi_k}p_k^\pi t_k^\pi & \\ \text{s.t.}\left|\begin{array}{lll}\displaystyle\sum_{k\in \underline{S}}\sum_{\pi=1}^{\Pi_k}\alpha_k^\pi t_k^\pi\le B & & (10.4.1)\\ \displaystyle\sum_{\pi=1}^{\Pi_k}t_k^\pi=1 & k\in\underline{S} & (10.4.2)\\ \frac{\tilde{\sigma}_l}{u_v}+\log u_v-1\ge\hat{\sigma}_l & l\in\underline{A}-\underline{A}_p,\ v=1,\ldots,V & (10.4.3)\\ \hat{\sigma}_l=\displaystyle\sum_{j\in\underline{\mathrm{As}}_l}\hat{\sigma}_j & l\in\underline{A}-\underline{A}_p & (10.4.4)\\ \hat{\sigma}_l=\displaystyle\sum_{\pi=1}^{\Pi_k}\log(1-\varphi_k^\pi)t_k^\pi & l\in\underline{A}_p,\ k=\mathrm{ext}(l) & (10.4.5)\\ t_k^\pi\in\{0,1\} & k\in\underline{S},\ \pi=1,\ldots,\Pi_k & (10.4.6)\\ \tilde{\sigma}_l\ge 0 & l\in\underline{A}-\underline{A}_p & (10.4.7)\\ \hat{\sigma}_l\le 0 & l\in\underline{A}-\underline{A}_p & (10.4.8)\end{array}\right.\end{cases}$$

We quickly comment on this program which has many similarities with program $\mathrm{P}_{10.2}$. As for $\mathrm{P}_{10.2}$, for more detailed explanations, the reader can refer to sections 10.8.2 and 10.8.3 of this chapter which presents an effective method of solving a closely related problem. Indeed, this method can easily be extended to the solution of the generalized Noah's Ark problem, *i.e.*, to the solution of $P_{10.3}$. Variables $\tilde{\sigma}_l$ and $\hat{\sigma}_l$ have the same meaning as in $P_{10.2}$. The first part of the economic function expresses an approximation of the contribution of the non-pending arcs to the ePD and the second part expresses the contribution of the pending arcs. Constraint 10.4.1 expresses the budget constraint. Constraints 10.4.2 express the fact that for each species s_k, one, and only one, of the $\mathbf{\Pi}_k$ possible protection policies must be chosen. Constraints 10.4.3 and 10.4.4 are identical, respectively, to constraints 10.2.2 and 10.2.3 of $P_{10.2}$. Constraints 10.4.5 express, for any pending arc in the tree, the logarithm of the probability that the information associated with that arc will be lost. Constraints 10.4.6, 10.4.7, and 10.4.8 specify the nature of the variables.

Note that in the generalized Noah's Ark problem the survival probabilities as well as the costs can be completely arbitrary. They can be very different from one species to another and no assumptions are needed about the relationship between the survival probability of a species and the amount of resources devoted to its protection. Nor is there any assumption about the structure of the tree or the length of its branches. The only limitation of the model is the ability to define, for each species s_k, a set of conservation policies, the cost of each policy and the associated survival probabilities. This model is therefore very general. Some authors have considered a situation, a priori less realistic, in which the survival probability of species s_k is expressed as a function of any amount of resources, b, devoted to the protection of this species, *i.e.*, by a function $f_k(b)$.

10.6.4 Example

Figure 10.7 shows a hypothetical phylogenetic tree with a root, r, two internal vertices, i_1 and i_2, 6 branches, a_1, a_2, a_3, a_4, a_5, and a_6, and 4 species, s_1, s_2, s_3, and s_4. The length of each branch and, for each species, the possible conservation policies with their associated costs are shown in the figure. For example, 3 conservation policies are possible for species s_2. The first is to do nothing, costs 0 and corresponds to a survival probability of 0.3. The second increases the survival probability from 0.3 to 0.6 and costs 3 units, and the third increases the survival probability to 0.9 and costs 6.2 units. Assume that the total budget available is equal to 16.5 units and consider protection policies 1, 3, 2, and 3 for species s_1, s_2, s_3, and s_4, respectively. The total conservation cost in this case is equal to $0 + 6.2 + 2 + 8 = 16.2$ units and the expected PD is equal to

$$(10.5 \times 0.5) + (4.2 \times (1 - 0.1 \times 0.8 \times 0.1)) + (6.3 \times 0.9) + (4 \times (1 - 0.8 \times 0.1)) + (7 \times 0.2) + (5.1 \times 0.9) = 24.7564.$$

The variance of the PD is 47.4975 (see section 10.4.1). The details of the calculation of this variance are presented in table 10.3.

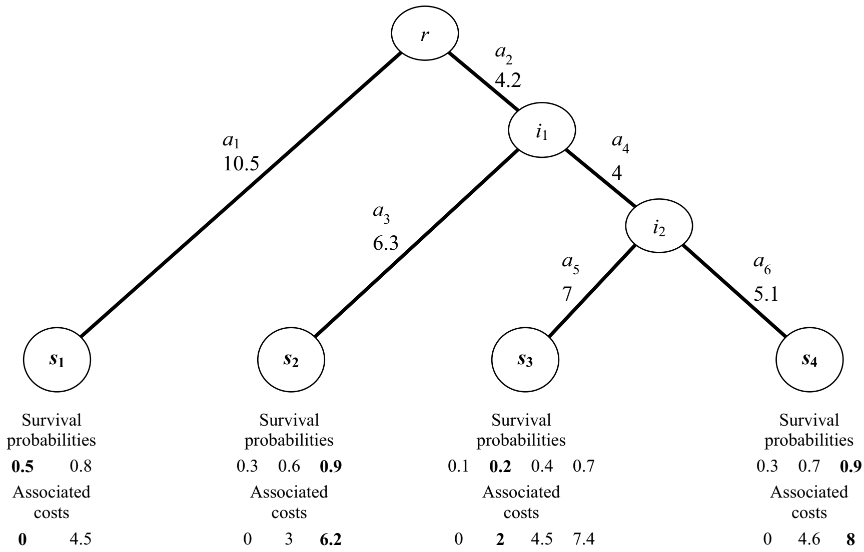


FIG. 10.7 – A hypothetical phylogenetic tree associated with the 4 species s_1, s_2, s_3, and s_4. The branch lengths and the various possible conservation policies – survival probabilities and associated costs – are shown in the figure. For example, 3 conservation policies are possible for species s_2. They cost 0, 3 and 6.2 units and correspond to survival probabilities of 0.3, 0.6, and 0.9, respectively.

TAB. 10.3 – Calculation of the variance of the phylogenetic diversity of the set of species, s_1, s_2, s_3, and s_4 in figure 10.7 when their survival probabilities are 0.5, 0.9, 0.2, and 0.9, respectively.

Branch (a_l)	Probability of keeping the associated information ($1 - \sigma_l$)	Branch length (λ_l)	$\lambda_l^2(1-\sigma_l)\sigma_l$	$(l, l') \in \underline{A}^2$: $l' \in \beta(l)$	$2\lambda_l\lambda_{l'}(1-\sigma_{l'})\sigma_l$
$a1$	0.5	10.5	27.5625	(2, 4)	0.2473
$a2$	0.992	4.2	0.1400	(2, 3)	0.3810
$a3$	0.9	6.3	3.5721	(2, 5)	0.0941
$a4$	0.92	4	1.1776	(2, 6)	0.3084
$a5$	0.2	7	7.8400	(4, 5)	0.8960
$a6$	0.9	5.1	2.3409	(4, 6)	2.9376
Total			42.6331		4.8644

10.7 Reserve Maximizing the PD of the Species of a Given Set Present in It

10.7.1 The Problem

We are interested here in a set of zones susceptible to protection, $Z = \{z_1, z_2, \ldots, z_n\}$, and in a set of species living on these zones, $S = \{s_1, s_2, \ldots, s_m\}$. The problem considered is to select a subset of zones to be protected – a reserve – under a budgetary constraint, so as to maximize the phylogenetic diversity of the species present in these zones. It is thus assumed that the protection of a zone allows for the protection of a given set of species. This can be interpreted in the following way: the protection of a zone ensures the conservation of the species living there and, on the contrary, the species present in an unprotected zone will disappear from that zone. This situation occurs when the unprotected zones are completely "lost" from a conservation point of view, at least for some species, which may be the case, for example, when these zones become urban or agricultural zones. The data for the problem are:

- A set of zones that can be protected.
- For each zone, the list of species that are protected as a result of the protection of that zone.
- The cost of protecting each zone.
- The total budget available.
- The phylogenetic tree of the species concerned – vertices, arcs and arc lengths.

Note that it is possible that the protection of a subset of zones may lead to the protection of all the species under consideration. This is because some zones may host several species and some species may occur in several zones. If, given the budget, it is possible to protect a subset of zones that allows for the protection of all the species considered, this is the optimal solution to the problem. The PD associated with a zone can be defined as the PD of the set of species that live there. Clearly, the PD associated with a set of zones is not the sum of the PD associated with each zone in the set. In other words, if R is a subset of zones, the PD of R is equal to the PD of the union of the sets of species living in each of the zones of R.

10.7.2 Mathematical Programming Formulation

Let x_i be the Boolean variable which is equal to 1 if and only if zone z_i is selected to be protected and y_k be the Boolean variable which is equal to 1 if and only if species s_k is present on one of the protected zones. The first thing to do is to express the phylogenetic diversity of the protected species, *i.e.*, to express this phylogenetic diversity as a function of variables y_k. To do this, we introduce an additional Boolean variable t_l which is equal to 1 if and only if arc a_l contributes to the calculation of the global phylogenetic diversity. This is the case if and only if there is a path from the terminal end of a_l to at least one of the protected species, *i.e.*, to at least one species s_k for which variable y_k takes the value 1. We can therefore express t_l as a function of variables y_k as follows: $t_l = \min\{\sum_{k \in \underline{F}_l} y_k, 1\}$ where $\underline{F}_l$ designates the set of indices of the leaves – the species – of the tree that can be reached by a path starting from the terminal end of a_l. The problem posed can then be solved by the linear program in Boolean variables $\mathrm{P}_{10.5}$.

$$
\mathrm{P}_{10.5}: \begin{cases} \max \displaystyle\sum_{l \in \underline{A}} \lambda_l t_l \\ \text{s.t.} \left| \begin{array}{llll} \displaystyle\sum_{i \in \underline{Z}} c_i x_i \leq B & & (10.5.1) & \mid \quad x_i \in \{0,1\} \quad i \in \underline{Z} \quad (10.5.4) \\ t_l \leq \displaystyle\sum_{k \in \underline{F}_l} y_k & l \in \underline{A} & (10.5.2) & \mid \quad y_k \in \{0,1\} \quad k \in \underline{S} \quad (10.5.5) \\ y_k \leq \displaystyle\sum_{i \in \underline{Z}_k} x_i & k \in \underline{S} & (10.5.3) & \mid \quad t_l \in \{0,1\} \quad l \in \underline{A} \quad (10.5.6) \end{array} \right. \end{cases}
$$

Z_k refers to the set of zones where species s_k lives and we assume, for example, that the protection of at least one zone of Z_k ensures the survival of species s_k. $\underline{Z}_k$ refers to the set of indices of zones of Z_k. The economic function expresses the sum of the lengths of the arcs whose information is retained taking into account the protected species and thus the phylogenetic diversity of these species. Constraint 10.5.1 expresses the budgetary constraint. According to constraint 10.5.2, the Boolean variable t_l can take the value 1 if and only if at least one of the leaves that can be reached from the terminal end of arc a_l corresponds to a protected species. In fact, given the economic function to be maximized, this variable necessarily takes the value 1 – at the optimum – if it can. It therefore reflects the fact that the information associated with arc a_l should be retained if and only if one of the protected species can be reached from that arc. According to constraints 10.5.3 variable y_k can take the value 1 if and only if one of the zones where species s_k lives is protected. Constraints 10.5.4, 10.5.5, and 10.5.6 specify the Boolean nature of the variables.

10.7.3 Example

Consider a set of 20 hypothetical zones represented in figure 10.8 and assume that the phylogenetic tree of the 15 species present in these zones is the one shown in figure 10.9. Suppose also that a budget of 4 units is available. The optimal solution – obtained by solving $\mathrm{P}_{10.5}$ – is to select zones z_2, z_6, and z_{10}, which will protect species s_3, s_6, s_7, s_8, s_9, and s_{14}. The cost of protection is equal to 4 and the phylogenetic diversity obtained is equal to 46. It corresponds to the arcs a_1, a_2, a_4, a_6, a_7, a_8, a_{12}, a_{13}, a_{14}, a_{15}, a_{16}, and a_{21} which are shown in bold in the figure. A non-optimal solution would be, for example, to use the budget of 4 units to protect zones z_2, z_8, z_{12}, and z_{15}, which would allow the protection of species s_3, s_6, s_{10}, s_{11}, and s_{13}, and in this case the phylogenetic diversity of the protected species would only be equal to 36.

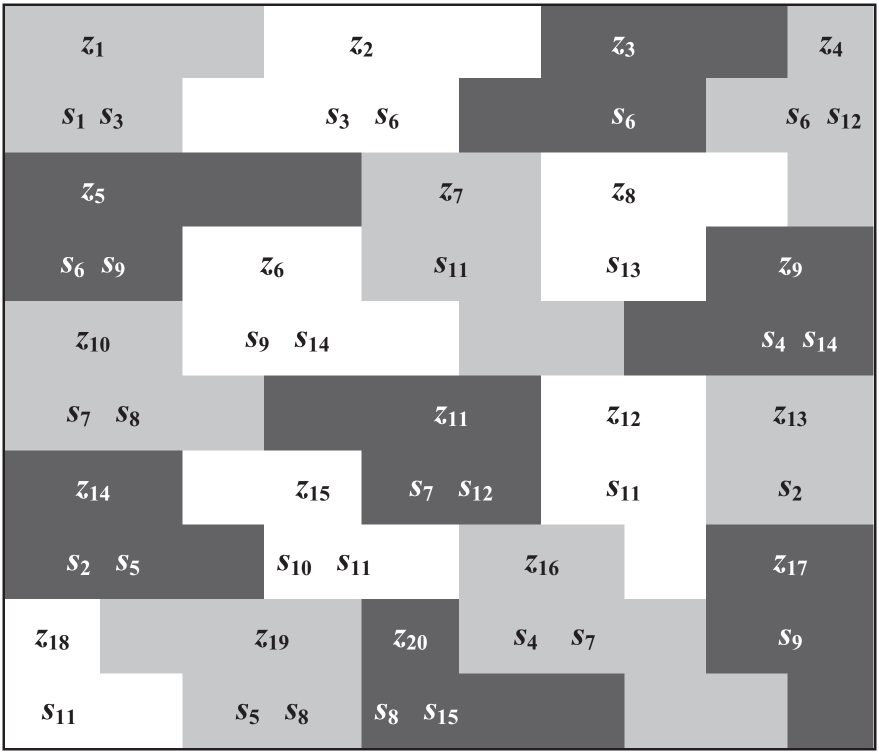


FIG. 10.8 – The 20 zones z_1, z_2,…, z_{20} are candidates for protection and 15 species, s_1, s_2,…, s_{15}, living in these zones are concerned. For each zone, the species present are indicated. The cost of protecting the white zones is equal to 1 unit, the cost of protecting the light grey zones is equal to 2 units and the cost of protecting the dark grey zones is equal to 4 units. For example, species s_9 and s_{14} are present in zone z_6, and the cost of protecting this zone is equal to 1 unit.

10.8 Reserve Maximizing the ePD of the Considered Species

10.8.1 The Problem

Here we are interested in the problem of choosing the zones to be protected – the reserve R – from a set of candidate zones and under a budgetary constraint, so as to maximize the expected phylogenetic diversity of the set of species under consideration $S = \{s_1, s_2,\ldots, s_m\}$ – living in protected or unprotected zones. The phylogenetic tree associated with the set of species in S and the set of candidate zones for protection, $Z = \{z_1, z_2,\ldots, z_n\}$, is considered. Each species is present in one or more zones. The survival probability of species s_k in zone z_i is equal to p_{ik}, if zone z_i is not

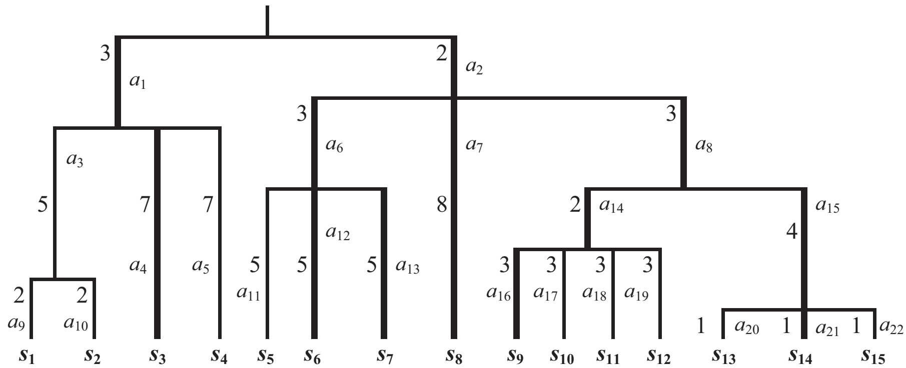


FIG. 10.9 – A hypothetical (ultrametric) phylogenetic tree associated with the 15 species considered in figure 10.8. The lengths of branches a_1, a_2,…, a_{22} are shown next to these branches.

protected, and to q_{ik} if zone z_i is protected. Note that these probabilities can take any value. They are supposed to be independent. For example, the survival probabilities of some species in protected or unprotected zones may be zero. It is also possible that protecting a zone may decrease the survival probability of some species living in that zone. As before, c_i refers to the cost of protecting zone z_i. The problem consists in determining a reserve, R, *i.e.*, a set of zones to be protected, so as to maximize the expected phylogenetic diversity associated with the set of species under consideration, while taking into account the budgetary constraint. The data are therefore:

- A set, Z, of zones that can be protected.
- For each zone $z_i \in Z$:
  - the list of species living in the zone – and in which we are interested;
  - for each species s_k living in that zone, its survival probability, p_{ik}, if the zone is not protected and its survival probability, q_{ik}, if the zone is protected;
  - the protection cost, denoted by c_i.
- The available budget, denoted by B.
- The phylogenetic tree of the species under consideration.

Taking into account the selected reserve, R, the extinction probability of species s_k in the set of zones considered is equal to $\prod_{i \in \underline{Z}: z_i \notin R} (1 - p_{ik}) \times \prod_{i \in \underline{Z}: z_i \in R} (1 - q_{ik})$. The expected phylogenetic diversity of the species concerned can therefore be written as $\sum_{l \in \underline{A}} \lambda_l (1 - \prod_{k \in \underline{F}_l} \prod_{i \in \underline{Z}: z_i \notin R} (1 - p_{ik}) \times \prod_{i \in \underline{Z}: z_i \in R} (1 - q_{ik}))$. Remember that λ_l designates the length of branch a_l and that $\underline{F}_l$ designates the set of indices associated with the species located under the arc a_l.

10.8.2 Mathematical Programming Formulation

As before, we use the Boolean variable x_i which is equal to 1 if and only if zone z_i is protected. The extinction probability of species s_k can be written, as a function of variables x_i, $\prod_{i\in\underline{Z}}(1-p_{ik}(1-x_i)-q_{ik}x_i)$. The expected phylogenetic diversity of the species concerned can then be written, as a function of variables x_i, $\sum_{l\in\underline{A}}\lambda_l(1-\prod_{k\in\underline{F}_l}\prod_{i\in\underline{Z}}((1-p_{ik}(1-x_i)-q_{ik}x_i))$. The problem can thus be formulated as the mathematical program $\mathrm{P}_{10.6}$.

$$\mathrm{P}_{10.6}:\begin{cases}\max \sum\limits_{l\in\underline{A}}\lambda_l\left(1-\prod\limits_{k\in\underline{F}_l}\prod\limits_{i\in\underline{Z}}(1-p_{ik}(1-x_i)-q_{ik}x_i)\right)\\ \text{s.t.}\left|\begin{array}{lll}\sum\limits_{i\in\underline{Z}}c_ix_i\le B & & (10.6.1)\\ x_i\in\{0,1\} & i\in\underline{Z} & (10.6.2)\end{array}\right.\end{cases}$$

The non-linearity of the economic function of program $\mathrm{P}_{10.6}$ makes it difficult to resolve. Let us look at another formulation of the problem. For this, let us associate to each arc – or branch – a_l of the tree a real variable, σ_l, which represents the probability that the information associated with this arc is not conserved – taking into account protected zones and therefore protected species. Let us denote by As_l the set of arcs directly following arc a_l, *i.e.*, the set of arcs whose initial end coincides with the terminal end of arc a_l. Let us denote by $\underline{\mathrm{As}}_l$ the set of corresponding indices. We have the equality $\sigma_l=\prod_{j\in\underline{\mathrm{As}}_l}\sigma_j$. Indeed, the information associated with arc a_l is not kept if and only if the information associated with each arc of As_l is not kept. For each pending arc, a_l, of the tree, we denote by $\mathrm{ext}(l)$ the index of the species corresponding to the terminal end of this arc. Using these variables σ_l and by expressing the probability that the information associated with the arc a_l is not retained as a function of the probabilities that the information associated with the arcs of As_l is not retained, the problem can be formulated as program $\mathrm{P}_{10.7}$.

$$\mathrm{P}_{10.7}:\begin{cases}\max \sum\limits_{l\in\underline{A}}\lambda_l(1-\sigma_l)\\ \text{s.t.}\left|\begin{array}{lll}\sum\limits_{i\in\underline{Z}}c_ix_i\le B & & (10.7.1)\\ \sigma_l=\prod\limits_{j\in\underline{\mathrm{As}}_l}\sigma_j & l\in\underline{A}-\underline{A}_p & (10.7.2)\\ \sigma_l=\prod\limits_{i\in\underline{Z}}(1-p_{ik}(1-x_i)-q_{ik}x_i) & l\in\underline{A}_p,\ k=\mathrm{ext}(l) & (10.7.3)\\ x_i\in\{0,1\} & i\in\underline{Z} & (10.7.4)\\ 0\le\sigma_l\le 1 & l\in\underline{A} & (10.7.5)\end{array}\right.\end{cases}$$

Using the logarithmic function and adding the constraints $\hat{\sigma}_l=\log\sigma_l, l\in\underline{A}$, which require variable $\hat{\sigma}_l$ to be equal to the logarithm of variable σ_l, program $\mathrm{P}_{10.7}$ can be rewritten as program $\mathrm{P}_{10.8}$. In order to simplify the presentation, it is assumed here that the survival probabilities p_{ik} and q_{ik} are all different from 1.

$$
\mathrm{P}_{10.8}: \begin{cases} \max \sum_{l \in \underline{A}} \lambda_l (1-\sigma_l) & \\ \text{s.t.} \left| \begin{array}{lll} \hat{\sigma}_l = \log \sigma_l & l \in \underline{A} & (10.8.1) \\ \sum_{i \in \underline{Z}} c_i x_i \leq B & & (10.8.2) \\ \hat{\sigma}_l = \sum_{j \in \underline{\mathrm{As}}_l} \hat{\sigma}_j & l \in \underline{A} - \underline{A}_p & (10.8.3) \\ \hat{\sigma}_l = \sum_{i \in \underline{Z}} \log[(1-p_{ik}(1-x_i) - q_{ik}x_i)] & l \in \underline{A}_p,\ k = \mathrm{ext}(l) & (10.8.4) \\ x_i \in \{0,1\} & i \in \underline{Z} & (10.8.5) \\ 0 \leq \sigma_l \leq 1,\ \hat{\sigma}_l \leq 0 & l \in \underline{A} & (10.8.6) \end{array} \right. \end{cases}
$$

The second member of constraint 10.8.4, $\sum_{i \in \underline{Z}} \log\left[(1-p_{ik}(1-x_i) - q_{ik}x_i)\right]$, represents the logarithm of the global extinction probability of species s_k, as a function of the zones selected for protection, *i.e.*, as a function of the values of variables x_i. It is easy to verify, by examining the 2 possible values of x_i, that $\sum_{i \in \underline{Z}} \log\left[(1-p_{ik}(1-x_i) - q_{ik}x_i)\right] = \sum_{i \in \underline{Z}} \left[x_i \log(1-q_{ik}) + (1-x_i)\log(1-p_{ik})\right]$. We can therefore finally formulate the problem as program $\mathrm{P}_{10.9}$ in which the economic function is linear and all the constraints, except constraints 10.9.1, are also linear.

$$
\mathrm{P}_{10.9}: \begin{cases} \max \sum_{l \in \underline{A}} \lambda_l (1-\sigma_l) & \\ \text{s.t.} \left| \begin{array}{lll} \hat{\sigma}_l = \log \sigma_l & l \in \underline{A} & (10.9.1) \\ \sum_{i \in \underline{Z}} c_i x_i \leq B & & (10.9.2) \\ \hat{\sigma}_l = \sum_{j \in \underline{\mathrm{As}}_l} \hat{\sigma}_j & l \in \underline{A} - \underline{A}_p & (10.9.3) \\ \hat{\sigma}_l = \sum_{i \in \underline{Z}} [x_i \log(1-q_{ik}) + (1-x_i)\log(1-p_{ik})] & l \in \underline{A}_p,\ k = \mathrm{ext}(l) & (10.9.4) \\ x_i \in \{0,1\} & i \in \underline{Z} & (10.9.5) \\ 0 \leq \sigma_l \leq 1,\ \hat{\sigma}_l \leq 0 & l \in \underline{A} & (10.9.6) \end{array} \right. \end{cases}
$$

Note that if constraints 10.8.6 and 10.9.6 specify, suitably for mathematical programming, that variables σ_l are greater than or equal to 0, these variables will in fact take a value strictly greater than 0 in any feasible solution of the corresponding programs. The same applies to $\mathrm{P}_{10.10}$.

10.8.3 Resolution

We now propose a relaxation of $\mathrm{P}_{10.9}$, similar to the one used in section 10.5 of this chapter and presented, in detail, in section 7.5 of chapter 7. In a feasible solution of this relaxation, the values of variables x_i define a feasible solution to the problem,

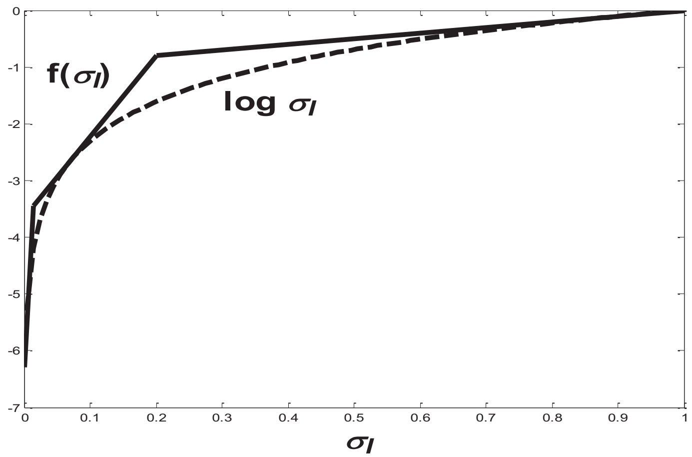


FIG. 10.10 – An upper approximation of the function $\log \sigma_l$ – shown in dashed lines – over the interval $]0, 1]$ by a piecewise linear function $f(\sigma_l)$ – shown in solid lines.

i.e., a feasible set of zones to be protected. The optimal value of this relaxation is an upper bound of the optimal value of the problem, *i.e.*, of the best expected phylogenetic diversity that could be obtained with a reserve of cost less than or equal to B. The relaxation we consider can be interpreted as an upper approximation of the concave function $\log \sigma_l$ by a concave and piecewise linear function (see appendix at the end of the book). This relaxation is obtained by relaxing constraints 10.9.1. Note first of all that, given the economic function to be maximized, we obtain a problem equivalent to $\mathrm{P}_{10.9}$ by replacing the equality constraints 10.9.1 by the inequality constraints $\hat{\sigma}_l \leq \log \sigma_l$, $l \in \underline{A}$. A relaxation of this inequality is obtained by replacing it by the set of V linear inequalities $\hat{\sigma}_l \leq (\sigma_l/u_v) + \log u_v - 1,\ v = 1, \ldots, V$, where $u_1, u_2, \ldots, u_V$ are constants such that $0 < u_1 < u_2 < \cdots < u_V = 1$. To prove that it is indeed a relaxation, it is enough to prove that $\log \sigma_l$ is less than or equal to $(\sigma_l/u_v) + \log u_v - 1$ for all v in $\{1, \ldots, V\}$. This is indeed the case since the latter inequality derives directly from the fact that (1) $1/u_v$ is the expression of the derivative of $\log x$ at the point u_v and (2) the function $\log x$ is concave. Constraints 10.10.1 expresses the fact that the quantity $\hat{\sigma}_l$ is less than or equal to the lower envelope of the V straight lines tangent to the curve $\log \sigma_l$ at the points of abscissa $u_1, u_2, \ldots, u_V$ (figure 10.10). The quantity σ_l is now an approximation of the probability that the information associated with branch a_l is not retained. For better readability, we will note it $\tilde{\sigma}_l$. The relaxation of $\mathrm{P}_{10.9}$ thus obtained is given by $\mathrm{P}_{10.10}$. To obtain a thin relaxation of $\mathrm{P}_{10.9}$, V must be sufficiently large. However, the larger V is, the greater the number of constraints of $\mathrm{P}_{10.10}$.

$$\mathrm{P}_{10.10}: \begin{cases} \max \displaystyle\sum_{l \in \underline{A}} \lambda_l (1 - \tilde{\sigma}_l) \\ \text{s.t.} \left| \begin{array}{lll} \hat{\sigma}_l \leq \frac{\tilde{\sigma}_l}{u_v} + \log u_v - 1 & l \in \underline{A}, v = 1,\ldots, V & (10.10.1) \\ \displaystyle\sum_{i \in \underline{Z}} c_i x_i \leq B & & (10.10.2) \\ \hat{\sigma}_l = \displaystyle\sum_{j \in \underline{\mathrm{As}}_l} \hat{\sigma}_j & l \in \underline{A} - \underline{\mathrm{Ap}} & (10.10.3) \\ \hat{\sigma}_l = \displaystyle\sum_{i \in \underline{Z}} [x_i \log(1 - q_{ik}) & & \\ \qquad + (1 - x_i) \log(1 - p_{ik})] & l \in \underline{\mathrm{Ap}},\ k = \mathrm{ext}(l) & (10.10.4) \\ x_i \in \{0, 1\} & i \in \underline{Z} & (10.10.5) \\ 0 \leq \tilde{\sigma}_l \leq 1,\ \hat{\sigma}_l \leq 0 & l \in \underline{A} & (10.10.6) \end{array} \right. \end{cases}$$

Let $(\overline{x}, \overline{\hat{\sigma}}, \overline{\tilde{\sigma}})$ be an optimal solution of $\mathrm{P}_{10.10}$. An approximate solution to the problem posed is given by $\overline{x}$ – which gives the zones to be protected. The associated expected PD is equal to $\sum_{l \in \underline{A}} \lambda_l (1 - \prod_{k \in \underline{F}_l} \prod_{i \in \underline{Z}} ((1 - p_{ik}(1 - \overline{x}_i) - q_{ik}\overline{x}_i))$ or, equivalently, to $\sum_{l \in \underline{A}} \lambda_l (1 - \exp(\overline{\hat{\sigma}}_l))$ where "exp" denotes the exponential function. An upper bound on the true optimal value of the problem is given by the optimal value of $\mathrm{P}_{10.10}$, *i.e.*, $\sum_{l \in \underline{A}} \lambda_l (1 - \overline{\tilde{\sigma}})$.

10.8.4 Example

The 20 candidate zones, their protection costs and the 15 species considered are described in figure 10.8. The phylogenetic tree associated with these 15 species is the one shown in figure 10.9. The survival probabilities of the species in the unprotected zones are all considered to be zero – $p_{ik} = 0$ for all $i \in \underline{Z}$ and for all $k \in \underline{S}$. As mentioned above, this may correspond to the case where unprotected zones are used for activities such as urbanization or agriculture. To simplify the presentation of this example it is assumed that for a given species, its survival probability is the same in all the protected zones, *i.e.*, the quantity q_{ik} depends only on k. These probabilities are given in table 10.4. Note that in this small example, the most threatened species, s_1, s_2, and s_3, are highly "phylogenetically" related, as is often the case in reality. In other words, as various authors have pointed out, the extinction risks are generally not uniform within a phylogeny. The optimal solutions obtained, for different values of the available budget, are presented in table 10.5. When the available budget is

TAB. 10.4 – Survival probabilities of species s_1, s_2,…, s_{15} in protected zones – independent of zones.

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
q_{ik}	0.2	0.3	0.1	0.4	0.4	0.5	0.8	0.6	0.7	0.8	0.7	0.9	0.6	0.7	0.5

TAB. 10.5 – Optimal solutions for different values of the available budget.

Available budget (% of the total cost of protecting the 20 candidate zones)	Optimal reserve	ePD	Upper bound to the optimal solution
20	z_2 z_4 z_6 z_{15} z_{16} z_{19}	42.13	42.15
50	z_1 z_2 z_4 z_6 z_8 z_9 z_{10} z_{13} z_{14} z_{15} z_{16} z_{19}	56.05	56.05
80	All zones except z_3 z_7 z_{17}	59.91	59.91
100	All zones	60.46	60.46

TAB. 10.6 – The survival probability of each species in the optimal reserve – maximizing the ePD – when the available budget is equal to 50% of the total cost of protection of the 20 zones under consideration.

Species	Zones of the reserve where the survival probability is > 0	Survival probability in the reserve	Species	Zones of the reserve where the survival probability is > 0	Survival probability in the reserve
s_1	z_1	0.20	s_9	z_6	0.70
s_2	z_{13} z_{14}	0.51	s_{10}	z_{15}	0.80
s_3	z_1 z_2	0.19	s_{11}	z_{15}	0.70
s_4	z_9 z_{16}	0.64	s_{12}	z_4	0.90
s_5	z_{14} z_{19}	0.64	s_{13}	z_8	0.60
s_6	z_2 z_4	0.75	s_{14}	z_6 z_9	0.91
s_7	z_{10} z_{16}	0.96	s_{15}	–	0
s_8	z_{10} z_{19}	0.84			

equal to 50% of the total cost of protecting the 20 zones considered, the optimal solution consists of protecting the 12 zones z_1, z_2, z_4, z_6, z_8, z_9, z_{10}, z_{13}, z_{14}, z_{15}, z_{16}, and z_{19}. The detailed calculation of the expected PD in this case is presented in tables 10.6 and 10.7.

10.9 Reserve Maximizing the ePD of the Considered Species, in the case of Uncertain Survival Probabilities

10.9.1 The Problem

The extinction probabilities of species – in protected or non-protected zones – are generally difficult to quantify. One way of taking into account this difficulty is to

TAB. 10.7 – Contribution of each arc of the phylogenetic tree to the maximal ePD that can be obtained when the available budget is equal to 50% of the total cost of protecting the 20 zones under consideration.

Arc of the phylogenetic tree	Arc length	Probability that the information associated with the arc is retained	Contribution of the arc to the ePD
a_1	3	0.885693	2.65708
a_2	2	1	2
a_3	5	0.608	3.04
a_4	7	0.19	1.33
a_5	7	0.64	4.48
a_6	3	0.9964	2.9892
a_7	8	0.84	6.72
a_8	3	0.999935	2.99981
a_9	2	0.2	0.4
a_{10}	2	0.51	1.02
a_{11}	5	0.64	3.2
a_{12}	5	0.75	3.75
a_{13}	5	0.96	4.8
a_{14}	2	0.9982	1.9964
a_{15}	4	0.964	3.856
a_{16}	3	0.7	2.1
a_{17}	3	0.8	2.4
a_{18}	3	0.7	2.1
a_{19}	3	0.9	2.7
a_{20}	1	0.6	0.6
a_{21}	1	0.91	0.91
a_{22}	1	0	0
Total			56.05

consider that several scenarios are possible (see appendix at the end of the book). The set of these scenarios is denoted by $\text{Sc} = \{\text{sc}_1, \text{sc}_2, \ldots, \text{sc}_p\}$, and $\underline{\text{Sc}} = \{1, 2, \ldots, p\}$ is the set of corresponding indices. It is assumed that the survival probabilities of the species, in these different scenarios, are known. We denote by p_{ik}^{ω} the survival probability of species s_k in zone z_i if zone z_i is not protected and if scenario sc_ω occurs, and q_{ik}^{ω} the same probability but in the case where zone z_i is protected. To simplify the presentation, all probabilities p_{ik}^{ω} and q_{ik}^{ω} are assumed to be different from 1. The problem is to determine a robust solution, *i.e.*, a set of zones to be protected so as to obtain a "good" ePD of the species concerned, whatever the scenario (see appendix at the end of the book). The objective that we have retained here is based on the notion of regret and consists in minimizing, in the worst-case scenario, the relative regret associated with this scenario. For a given scenario, the relative regret represents the "shortfall" associated with the choice of the robust solution; it is equal to the relative gap between the ePD of the retained reserve and the ePD of the optimal reserve, both ePDs being calculated in the considered scenario. In other words, the problem is to determine the zones to be protected, taking into account the available resources, in such a way as to minimize the maximum, over all scenarios, of the relative gap between (1) the ePD of the set of species concerned, calculated with the probabilities of the considered scenario taking into account the selected zones, and (2) the maximal ePD of the set of species concerned that could be obtained – by protecting the adequate set of zones – in the considered scenario. The set of zones selected for protection is the optimal robust solution or the optimal robust reserve. Let us define the problem more formally. Let S be the set of species considered and Z the set of candidate zones for protection. Let us denote by $\mathrm{ePD}^{\omega}(S, R)$ the ePD of the species of S obtained in scenario sc_{ω} when the zones of $R \subseteq Z$ are protected. Let us denote by $R^{\omega *}$ the set of zones of Z whose protection maximizes the ePD of the species of S, in the case of scenario sc_{ω}, and by R^* the optimal robust solution. The problem under consideration – the determination of R^* – can then be formulated as follows:

$$\min_{R \subseteq Z} \max_{\omega \in \underline{\mathrm{Sc}}} \{(\mathrm{ePD}^{\omega}(S, R^{\omega *}) - \mathrm{ePD}^{\omega}(S, R))/\mathrm{ePD}^{\omega}(S, R^{\omega *})\}.$$

10.9.2 Mathematical Programming Formulation

Let us first consider the problem of determining the set of zones to be protected in order to maximise the expected phylogenetic diversity of the species concerned, in the case of scenario sc_{ω}. As before, we use the Boolean variable x_i which is equal to 1 if and only if zone z_i is protected. In the case of scenario sc_{ω}, the extinction probability of species s_k can then be written – as a function of variables x_i – $\prod_{i \in \underline{Z}} (1 - p_{ik}^{\omega}(1 - x_i) - q_{ik}^{\omega} x_i)$. The expected phylogenetic diversity of the species concerned can therefore be written – again as a function of variables x_i – $\sum_{l \in \underline{A}} \lambda_l (1 - \prod_{k \in \underline{F}_l} \prod_{i \in \underline{Z}} ((1 - p_{ik}^{\omega}(1 - x_i) - q_{ik}^{\omega} x_i))$. The problem of determining the optimal set of zones to be protected, in the case of scenario sc_{ω}, can therefore be formulated as the mathematical program $\mathrm{P}_{10.11}(\omega)$.

$$\mathrm{P}_{10.11}(\omega) : \begin{cases} \max \displaystyle\sum_{l \in \underline{A}} \lambda_l \left(1 - \prod_{k \in \underline{F}_l} \prod_{i \in \underline{Z}} (1 - p_{ik}^{\omega}(1 - x_i) - q_{ik}^{\omega} x_i) \right) \\ \text{s.t.} \left| \begin{array}{lll} \displaystyle\sum_{i \in \underline{Z}} c_i x_i \le B & & (10.11_{\omega}.1) \\ x_i \in \{0, 1\} & i \in \underline{Z} & (10.11_{\omega}.2) \end{array} \right. \end{cases}$$

Finally, following the same procedure as in sections 10.5 and 10.8, a good approximate solution to the problem under consideration and an upper bound to its optimal value can be obtained by solving program $\mathrm{P}_{10.12}(\omega)$. In this program, variable σ_l^ω represents the probability that the information attached to the arc a_l is not retained in the case of scenario sc_ω, given the protected zones and therefore the protected species in this scenario, variable $\tilde{\sigma}_l^\omega$ represents an approximation of this probability, and $\hat{\sigma}_l^\omega$ is a variable equal to the logarithm of σ_l^ω

$$
\mathrm{P}_{10.12}(\omega):\left\{
\begin{array}{lll}
\max \displaystyle\sum_{l\in\underline{A}} \lambda_l(1-\tilde{\sigma}_l^\omega) & & \\
\text{s.t.} \left|
\begin{array}{l}
\hat{\sigma}_l^\omega \le \frac{\tilde{\sigma}_l^\omega}{u_v} + \log u_v - 1 \\
\displaystyle\sum_{i\in\underline{Z}} c_i x_i \le B \\
\hat{\sigma}_l^\omega = \displaystyle\sum_{j\in\underline{\mathrm{As}}_l} \hat{\sigma}_l^\omega \\
\hat{\sigma}_l^\omega = \displaystyle\sum_{i\in\underline{Z}} x_i \log(1-q_{ik}^\omega) \\
\qquad + \displaystyle\sum_{i\in\underline{Z}} (1-x_i)\log(1-p_{ik}^\omega) \\
x_i \in \{0,1\} \\
0 \le \tilde{\sigma}_l^\omega \le 1,\ \hat{\sigma}_l^\omega \le 0
\end{array}\right.
&
\begin{array}{l}
l\in\underline{A}, v=1,\ldots,V \\
\\
l\in\underline{A}-\underline{A}_p \\
\\
l\in\underline{A}_p,\ k=\mathrm{ext}(l) \\
i\in\underline{Z} \\
l\in\underline{A}
\end{array}
&
\begin{array}{l}
(10.12_\omega.1) \\
(10.12_\omega.2) \\
(10.12_\omega.3) \\
\\
(10.12_\omega.4) \\
(10.12_\omega.5) \\
(10.12_\omega.6)
\end{array}
\end{array}\right.
$$

The searched approximate solution is given by the values of variables x_i in an optimal solution of $\mathrm{P}_{10.12}(\omega)$. The reserve obtained, which we note $R_a^{\omega*}$, is made up of zones z_i such as $x_i = 1$. Its value, $\mathrm{ePD}^\omega(S, R_a^{\omega*})$, is an approximation of $\mathrm{ePD}^\omega(S, R^{\omega*})$. Let us now consider the problem of selecting an optimal robust reserve, R^*. Recall that it can be written:

$$\min_{R\subseteq Z, C(R)\le B}\ \max_{\omega\in\underline{S}} \{(\mathrm{ePD}^\omega(S,R^{\omega*}) - \mathrm{ePD}^\omega(S,R))/\mathrm{ePD}^\omega(S,R^{\omega*})\}.$$

In fact, we will consider the problem.

$$\min_{R\subseteq Z, C(R)\le B}\ \max_{\omega\in\underline{S}} \{(\mathrm{ePD}^\omega(S,R_a^{\omega*}) - \mathrm{ePD}^\omega(S,R))/\mathrm{ePD}^\omega(S,R_a^{\omega*})\}.$$

A "good" approximation of the latter problem is formulated by the mathematical program $\mathrm{P}_{10.13}$.

$$
\mathrm{P}_{10.13}: \begin{cases} \min \alpha & \\ \text{s.t.} \left| \begin{array}{lll} \alpha \geq (\mathrm{eDP}^{\omega}(S, R_a^{\omega *}) - \psi^{\omega})/\mathrm{eDP}^{\omega}(S, R_a^{\omega *}) & \omega \in \underline{\mathrm{Sc}} & (10.13.1) \\ \psi^{\omega} \leq \sum\limits_{l \in \underline{A}} \lambda_l (1 - \tilde{\sigma}_l^{\omega}) & \omega \in \underline{\mathrm{Sc}} & (10.13.2) \\ \hat{\sigma}_l^{\omega} \leq \frac{\tilde{\sigma}_l^{\omega}}{u_v} + \log u_v - 1 & l \in \underline{A}, & \\ & v = 1, \ldots, V, \omega \in \underline{\mathrm{Sc}} & (10.13.3) \\ \sum\limits_{i \in \underline{Z}} c_i x_i \leq B & & (10.13.4) \\ \hat{\sigma}_l^{\omega} = \sum\limits_{j \in \underline{\mathrm{As}}_l} \hat{\sigma}_j^{\omega} & l \in \underline{A} - \underline{A}_p,\ \omega \in \underline{\mathrm{Sc}} & (10.13.5) \\ \hat{\sigma}_l^{\omega} = \sum\limits_{i \in \underline{Z}} [x_i \log(1 - q_{ik}^{\omega}) & & \\ \quad + (1 - x_i) \log(1 - p_{ik}^{\omega})] & l \in \underline{A}_p, & \\ & k = \mathrm{ext}(l), \omega \in \underline{\mathrm{Sc}} & (10.13.6) \\ x_i \in \{0, 1\} & i \in \underline{Z} & (10.13.7) \\ 0 \leq \tilde{\sigma}_l^{\omega} \leq 1,\ \hat{\sigma}_l^{\omega} \leq 0 & l \in \underline{A}, \omega \in \underline{\mathrm{Sc}} & (10.13.8) \\ \psi^{\omega} \geq 0 & \omega \in \underline{\mathrm{Sc}} & (10.13.9) \end{array} \right. \end{cases} .
$$

Variable ψ^{ω} represents an approximation of the ePD of the species concerned by the set of protected zones, $\{z_i \in Z : x_i = 1\}$, in the case of scenario sc_{ω}. The robust reserve is made up of those zones z_i such as $x_i = 1$. We note this reserve R_a^*.

10.9.3 Example

The 20 zones considered, their protection costs and the 15 species present in these zones are described in figure 10.8. The phylogenetic tree associated with the 15 species considered is the one shown in figure 10.9. The available budget is 8 units. The survival probabilities of the species in the unprotected zones are all considered to be zero under all scenarios – $p_{ik}^{\omega} = 0$ for all $i \in \underline{Z}$, for all $k \in \underline{S}$ and for all $\omega \in \underline{\mathrm{Sc}}$. Finally, we consider that for a given species, its survival probability is the same in all the protected zones, *i.e.,* the quantity q_{ik}^{ω} depends only on k and ω. In this example, two scenarios are considered and the corresponding probabilities are given in table 10.8.

Let us first determine reserve $R_a^{\omega *}$ by solving $\mathrm{P}_{10.12}(\omega)$ for each scenario sc_{ω} and the associated ePD. These results are presented in table 10.9. The detailed results of table 10.9 are presented in table 10.10. The robust reserve R_a^* consists of zones z_1, z_2, z_6, z_{16}, and z_{19}, and the survival probabilities of the different species in this reserve are given in table 10.11 for each of the 2 scenarios. The ePD associated with the robust reserve, R_a^*, in each scenario, is deduced from table 10.11 and presented in table 10.12. Finally, the maximal relative regret can be calculated (table 10.13). The

TAB. 10.8 – Extinction probabilities of species s_1, s_2,..., s_{15} according to the 2 scenarios. For all the species, the extinction probabilities are the same in all zones.

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
q^1_{ik}	0.4	0.3	0.5	0.6	0.7	0.8	0.8	0.6	0.9	0.5	0.7	0.2	0.4	0.4	0.5
q^2_{ik}	0.8	0.7	0.6	0.7	0.4	0.3	0.3	0. 3	0.4	0.3	0.6	0.8	0.4	0.7	0.3

TAB. 10.9 – $R_a^{\omega *}$ and its associated ePD for each scenario.

Scenario sc_ω	$R_a^{\omega *}$	$ePD^\omega(S, R_a^{\omega *})$
sc_1	$\{z_2, z_6, z_8, z_{15}, z_{16}, z_{19}\}$	45.98
sc_2	$\{z_1, z_6, z_{15}, z_{16}, z_{19}\}$	39.15

TAB. 10.10 – Details of the results presented in table 10.9.

Species s_k	Scenario sc_1: survival probability of species s_k in R_a^{1*}	Scenario sc_2: survival probability of species s_k in R_a^{2*}
s_1	0	0.8
s_2	0	0
s_3	0.5	0.6
s_4	0.6	0.7
s_5	0.7	0.4
s_6	0.8	0
s_7	0.8	0.3
s_8	0.6	0.3
s_9	0.9	0.4
s_{10}	0.5	0.3
s_{11}	0.7	0.6
s_{12}	0	0
s_{13}	0.4	0
s_{14}	0.4	0.7
s_{15}	0	0

maximal relative gap is equal to 1.01%. In other words, the relative error that can be committed, if the robust reserve, R_a^*, is chosen rather than the optimal reserve for a given scenario, is less than or equal to 1.01% regardless of the scenario that occurs.

Note that, if we choose reserve R_a^{1*} while it is scenario sc_2 that is realized, we obtain an ePD equal to 36.10, whereas the ePD of R_a^{2*} is equal to 39.15 and the ePD

TAB. 10.11 – Survival probabilities of species s_1, s_2,..., s_{15} in the robust reserve, R_a^*, for each of the 2 scenarios.

Species s_k	Scenario sc_1: survival probability of species s_k in the robust reserve R_a^*	Scenario sc_2: survival probability of species s_k in the robust reserve R_a^*
s_1	0.4	0.8
s_2	0	0
s_3	0.75	0.84
s_4	0.6	0.7
s_5	0.7	0.4
s_6	0.8	0.3
s_7	0.8	0.3
s_8	0.6	0.3
s_9	0.9	0.4
s_{10}	0	0
s_{11}	0	0
s_{12}	0	0
s_{13}	0	0
s_{14}	0.4	0.7
s_{15}	0	0

TAB. 10.12 – ePD associated with the robust reserve, R_a^*, in each scenario.

Scenario sc_ω	$\text{ePD}^\omega(S, R^*)$
sc_1	45.65
sc_2	38.76

TAB. 10.13 – Maximal relative regret.

Scenario sc_ω	$\text{ePD}^\omega(S, R_a^*)$	$\text{ePD}^\omega(S, R_a^{\omega*})$	$\dfrac{\text{ePD}^\omega(S, R_a^{\omega*}) - \text{ePD}^\omega(S, R_a^*)}{\text{ePD}^\omega(S, R_a^{\omega*})}$
sc_1	45.65	45.98	0.71%
sc_2	38.76	39.15	1.01%

of the robust reserve R_a^* is equal to 38.76. Conversely, if we choose reserve R_a^{2*} while it is scenario sc_1 that is realized, we obtain an ePD equal to 43.50, whereas the ePD of R_a^{1*} is equal to 45.98 and the ePD of the robust reserve, R_a^*, is equal to 45.65.

10.10 Reserve Maximizing the PD of the Species that are Present or the ePD of the Species Under Consideration, in the Presence of Uncertainties in the Phylogenetic Tree

10.10.1 The Problem

It is recognized that there is generally a lot of uncertainty in the definition of the phylogenetic tree associated with a set of species. These uncertainties may concern the topology of the tree as well as the lengths of its branches. We assume here that the uncertainties are captured by the fact that several trees are plausible for the set of species considered. An important question that arises then is to try to take into account these uncertainties about the tree in the definition of a reserve to preserve biodiversity when biodiversity is measured by phylogenetic diversity or by expected phylogenetic diversity. In the latter case, to simplify the presentation, we assume that there is no uncertainty about the survival probabilities of the species under consideration. We first present different measures of PD that can be used in the presence of uncertainties when these uncertainties are captured by the fact that several phylogenetic trees are plausible. We then show how these measures can be used to select zones, with the aim of maximizing the phylogenetic diversity associated with these zones, and then extend these results to the expected phylogenetic diversity.

10.10.2 Different Measures of PD in the Presence of Uncertainties in the Phylogenetic Tree

We consider a set of species, $S = \{s_1, s_2, \ldots, s_m\}$, and a set of φ plausible phylogenetic trees, $T = \{T_1, T_2, \ldots, T_\varphi\}$, for these species. Let us denote by $\underline{S}$, the set of indices $\{1, \ldots, m\}$ and $\underline{T}$, the set of indices $\{1, 2, \ldots, \varphi\}$. Each tree T_τ of T is represented by the quadruplet $(V^\tau, A^\tau, S, \lambda^\tau)$ where V^τ is the set of vertices, A^τ, the set of arcs, S, the set of leaves – the set of species – and λ^τ, the set of branch lengths. For any subset $\hat{S}$ of S, $\mathrm{PD}_\tau(\hat{S})$ designates the PD of $\hat{S}$ in the tree T_τ. We are then confronted with the problem of evaluating the PD of a group of species taking into account the uncertainties on the phylogenetic tree associated with these species. We propose below several ways to evaluate this PD.

10.10.2.1 Average and Weighted Average Phylogenetic Diversity (aPD and waPD)

One way to take into account the multiplicity of trees associated with a set of species, S, to assess the PD of a group of species $\hat{S} \subseteq S$ is to consider the average PD of $\hat{S}$ across all the trees. This is a very conventional measure. We record it as $\mathrm{aPD}(\hat{S})$. It is expressed as follows:

$$\mathrm{aPD}(\hat{S}) = \frac{1}{\varphi}\sum_{\tau\in\underline{T}} \mathrm{PD}_\tau(\hat{S}).$$

Alternatively, the weighted average of the PDs of $\hat{S}$ across all the trees can be considered, if one wants to give more or less importance to individual trees. We denote this measure by $\mathrm{waPD}(\hat{S})$. It is expressed as follows:

$$\mathrm{waPD}(\hat{S}) = \frac{\sum_{\tau\in\underline{T}} w_\tau \mathrm{PD}_\tau(\hat{S})}{\sum_{\tau\in\underline{T}} w_\tau},$$

where w_τ is the weight associated with the tree T_τ. If $w_\tau = 1$ for all τ then $\mathrm{aPD}(\hat{S}) = \mathrm{waPD}(\hat{S})$. The advantage of these two measures lies in their simplicity, but they have many disadvantages. The measure aPD is in fact the expected PD that would be obtained by assigning the same probability to each tree. Similarly, waPD is the mathematical expectation corresponding to the probability $w_\tau/\sum_{\tau\in\underline{T}} w_\tau$ assigned to the tree T_τ, $\tau \in \underline{T}$. An important and well-known disadvantage of these measures is that they are strongly influenced by the extreme values. Moreover, they allow for compensation between the low and high values. Thus, a group of species with a relatively high aPD can have a very low PD on some trees. It should be noted that if the uncertainty about the phylogenetic tree was only about its branch lengths, one could be interested, for any set of species $\hat{S}$ included in S, in the expected PD of $\hat{S}$, provided that a probability could be associated with each possible length of the different branches.

10.10.2.2 Robust Phylogenetic Diversity (rPD)

In the presence of uncertainty about the phylogenetic tree associated with a group of species, a robust solution may be of interest (see appendix at the end of the book). A robust solution can be defined as any solution that protects the decision-maker from uncertainty in some way. Several such measures have been proposed in the literature. In this section, we focus on a classical measure that we denote by $\mathrm{rPD}(\hat{S})$ for any subset $\hat{S}$ of S. This is a very conservative measure that, for all $\hat{S}\subseteq S$, ensures that the PD of $\hat{S}$, in all the trees considered, is at least equal to $\mathrm{rPD}(\hat{S})$. This measure therefore only takes into account the “worst case scenario”. In practice, it consists of calculating the PD of $\hat{S}$ for each tree and then taking the lowest of the values obtained. It is expressed as follows:

$$\mathrm{rPD}(\hat{S}) = \min_{\tau\in\underline{T}} \mathrm{PD}_\tau(\hat{S}).$$

We will see later that looking for a set of species $\hat{S}\subseteq S$ that maximizes $\mathrm{rPD}(\hat{S})$, under certain constraints, amounts to looking for a set of species $\hat{S}\subseteq S$ that perform relatively well, from a PD point of view, regardless of which tree actually represents the phylogeny associated with $\hat{S}$. It is in this sense that a robust solution protects against uncertainties. This measure, which is interesting regardless of the probabilities associated with each tree – if it is possible to assign such probabilities – may be useful when all the trees are equiprobable. Another measure may also be considered. It consists in establishing a compromise between the pessimistic measure $\mathrm{rPD}(\hat{S})$, which we have just seen, and an optimistic measure that would only take into account the highest PD value of $\hat{S}$ over all trees – Hurwicz's criterion. This measure is therefore a weighted average of the extreme values. For a coefficient of pessimism $\alpha \in [0,1]$, it is written as:

$$\alpha \min_{\tau \in \underline{T}} \mathrm{PD}_\tau(\hat{S}) + (1-\alpha) \max_{\tau \in \underline{T}} \mathrm{PD}_\tau(\hat{S}).$$

10.10.2.3 Ordered Weighted Average of Phylogenetic Diversity (owaPD)

We saw in the previous section a measure associated with a set of species, $\hat{S}$, included in S that took into account the worst situation – associated with the phylogenetic tree providing the lowest value of $\mathrm{PD}_\tau(\hat{S})$ – and also another measure that took into account both the worst and the best situation. We now propose to use an even different measure that somehow takes into account all the situations. The notion of ordered weighted average (owa) was introduced by Yager in 1998 as a tool for aggregating a certain amount of information. In the case at hand, this operator provides a measure of $\hat{S}$ that first takes into account the lowest value of $\mathrm{PD}_\tau(\hat{S})$ then the value immediately following, and so on until the best value is itself taken into account. We denote by $\mathrm{owaPD}(\hat{S})$ this measure. Let $w_1, w_2, \ldots w_\varphi$ be a decreasing list of weights, belonging to the interval [0,1] and whose sum is 1. The calculation of $\mathrm{owaPD}(\hat{S})$ is done as follows: multiply each weight w_k by the kth smallest value of the set $\{\mathrm{PD}_\tau(\hat{S}), \tau \in \underline{T}\}$, and sum all the values so obtained. This can be expressed as follows:

$$\mathrm{owaPD}(\hat{S}) = \sum_{k \in \underline{T}} w_k \mathrm{PD}_{\eta(k)}(\hat{S})$$

where $\eta(k)$ is the index of the tree corresponding to the kth smallest value of the set $\{\mathrm{PD}_\tau(\hat{S}) : \tau \in \underline{T}\}$. The value of $\mathrm{owaPD}(\hat{S})$ lies between the minimal and maximal value of $\mathrm{PD}_\tau(\hat{S})$. Compared to the measure rPD, the measure owaPD is less "conservative" since this measure allows a compromise between several scenarios. Note, however, that owaPD includes rPD as a special case – $w_1 = 1$ and $w_2 = w_3 = \cdots = w_\varphi = 0$. One of the difficulties in using owaPD is the definition of the weights $w_1, w_2, \ldots w_\varphi$. The meaning of these weights is indeed very different from the meaning of the weights that are used in waPD. The latter measure makes an assumption about the importance of each tree – there is no "impartiality". Instead, the weight used in owaPD reflects the importance given to the lowest value, the one immediately after it, and so on, but neither these values nor the tree from which they come are known. The reader can refer to the bibliography of this chapter for a presentation of the owa operators, their interest and their use in different applications.

10.10.2.4 Highest Value of Guaranteed Phylogenetic Diversity for δ Trees ($h_\delta PD$)

This measure applied to a set of species $\hat{S} \subseteq S$ is denoted by $h_\delta \mathrm{PD}(\hat{S})$. If $h_\delta \mathrm{PD}(\hat{S})$ takes value v, it means that, for at least δ trees of T, the quantity $\mathrm{PD}_\tau(\hat{S})$ is greater than or equal to this value v. Since we are interested in the highest possible value, v is therefore equal to the kth lowest value of the set $\{\mathrm{PD}_\tau(\hat{S}) : \tau \in \underline{T}\}$ with $k = (\varphi - \delta + 1)$. An alternative interpretation is that this value does not take into account the $(\varphi - \delta)$ smallest values. The quantity $h_\delta \mathrm{PD}(\hat{S})$ can be expressed as follows:

$$h_\delta \mathrm{PD}(\hat{S}) = \min_{\tau \in I_\delta} \mathrm{PD}_\tau(\hat{S})$$

where I_δ is the set of indices of $\underline{T}$ corresponding to the δ highest values of the set $\{\mathrm{PD}_\tau(\hat{S}), \tau \in \underline{T}\}$. In other words, for δ trees of the set of trees considered, the phylogenetic diversity of $\hat{S}$ is at least equal to $h_\delta \mathrm{PD}(\hat{S})$. This measure corresponds to the special case of owaPD in which the weights $w_1, w_2, \ldots, w_\varphi$ may not be decreasing and are all equal to 0 except $w_{\delta+1}$ which is equal to 1. On the other hand, if $\delta = \varphi$, then $h_\delta \mathrm{PD}(\hat{S}) = \mathrm{rPD}(\hat{S})$. Note that in the case where a probability can be associated with each tree T_τ of T, the probability that the phylogenetic diversity of a set $\hat{S} \subseteq S$ is greater than or equal to $h_\delta \mathrm{PD}(\hat{S})$ is greater than or equal to the sum of the probabilities associated with the trees whose index belongs to I_δ – for example, δ/φ when all the trees are equiprobable.

10.10.2.5 Largest Deviation from Optimal Phylogenetic Diversity (lgapPD)

This measure, which is part of the robust measures (see appendix at the end of the book), is by nature slightly different from the previous ones. It involves the notion of regret. Consider a set of species, $\hat{S}$, included in S and satisfying a set of constraints, C. This measure involves the highest PD that can be associated with a set of species included in S and satisfying C, for each tree considered. To evaluate a set of species $\hat{S}$ included in S – and satisfying C – with this measure, the difference between the PD of $\hat{S}$ and the highest PD value that could be obtained on the same tree, for a set of species included in S and satisfying C – the regret – is calculated for each tree. The largest of these differences – the maximal regret – is then retained. This measure, which we denote by $\mathrm{lgapPD}(\hat{S})$, can be expressed as follows:

$$\mathrm{lgapPD}(\hat{S}) = \max_{\tau \in \underline{T}} \{\mathrm{PD}_\tau^{*C} - \mathrm{PD}_\tau(\hat{S})\}$$

where PD_τ^{*C} is equal to the maximal PD of a set of species included in S and satisfying the constraints C, calculated for the tree T_τ. Thus, for any $\hat{S}$ included in S and satisfying C, the distance between the PD of $\hat{S}$ and the maximal PD of a set of species, included in S and satisfying C, is guaranteed in each tree, to the extent that this distance is less than or equal to $\text{lgapPD}(\hat{S})$.

For all set of species $\hat{S} \subseteq S$ we denote by $\mu\text{PD}(\hat{S})$ the 6 measures we just saw where μ represents a, wa, r, owa, h_δ and lgap.

10.10.2.6 Example

Consider the 4 hypothetical phylogenetic trees in figure 10.11.

These trees were generated to illustrate the measures of phylogenetic diversity presented earlier. They are associated with the set of 6 species $\{s_1, s_2, s_3, s_4, s_5, s_6\}$. Table 10.14 presents the values of $PD_\tau(\hat{S})$ for each tree and also the 6 $\mu\text{PD}(\hat{S})$ values, for $\hat{S} = \{1, 2, 3\}$. In this example, the value of $\text{lgapPD}(\hat{S})$ is calculated assuming that the constraints C simply express the fact that 3 out of 6 species must be selected.

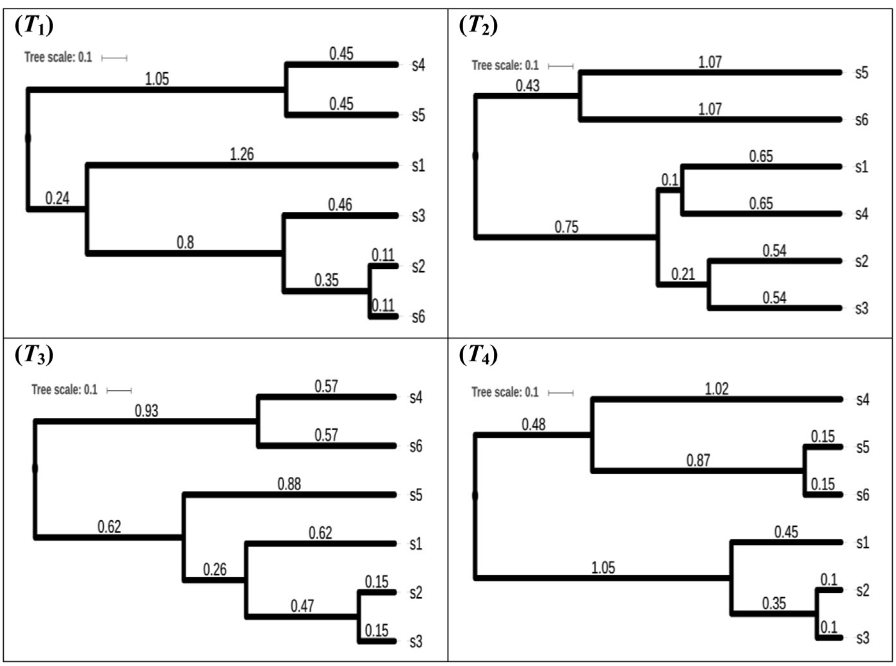


FIG. 10.11 – Four hypothetical phylogenetic trees associated with six species (drawn with iTOL software).

TAB. 10.14 – Values of the PDs of the set {1, 2, 3} for the 4 trees in figure 10.11 and of μPD for μ = a, wa, owa, r, h_2, and lgap.

PD Tree T_1	PD Tree T_2	PD Tree T_3	PD Tree T_4	aPD	waPD $w = (0.5, 0.2, 0.2, 0.1)$	owaPD $w = (0.4, 0.3, 0.2, 0.1)$	rPD	h_2PD	lgapPD
3.22	2.79	2.27	2.05	2.5825	2.827	2.381	2.05	2.79	1.97

10.10.3 Reserve Maximizing the PD of the Species, of a Given Set, Present in It

In this section, we examine the following classical problem, relating to a set of phylogenetic trees, $T = (T_1, T_2,\ldots, T_\varphi)$, associated with a set of m species, $S = \{s_1, s_2,\ldots, s_m\}$, distributed over a set of n geographical zones, $Z = \{z_1, z_2,\ldots, z_n\}$: given a cost c_j associated with each zone z_j, select a subset of zones whose total cost does not exceed a predefined budget, B, and which optimizes μPD$(\hat{S})$ over all the feasible sets of zones where $\hat{S}$ is the set of species present in at least one of the selected zones. We say that species s_k is protected if it is present in at least one of the selected zones. "Optimize" μPD$(\hat{S})$ means "maximize" for μ= a, wa, owa, r, and h_δ, and "minimize" for μ= lgap. For any set of species $\hat{S}$ included in S, it is easy to calculate μPD$(\hat{S})$ for μ = a, wa, owa, r, and h_δ. For μ = lgap, the calculation is a bit more complicated since it requires first of all to determine, for each tree, the maximal PD of a set of species verifying certain constraints. For the problem at hand, the sets of species to be considered are the sets of species present in at least one zone of a set of zones with a total cost less than or equal to B. We are going to propose the formulation of the reserve selection problem considered by integer linear programming.

Let us recall that $T = (T_1, T_2,\ldots, T_\varphi)$ is the set of phylogenetic trees to be considered, that A^τ is the set of arcs of the tree T_τ, and that λ^τ is the set of the branch lengths of this tree. More precisely, $\lambda^\tau = \{\lambda_l^\tau : l \in \underline{A}^\tau\}$ where λ_l^τ is the length of branch a_l in the tree T_τ and $\underline{A}^\tau$ is the set of indices of the arcs of A^τ. F_l^τ is the set of species – leaves – located under the arc a_l in the tree T_τ and $\underline{F}_l^\tau$ is the set of corresponding indices. In other words, the survival of at least one species of F_l^τ preserves the evolutionary history linked to the arc a_l if, however, T_τ is the right phylogenetic tree. We note $\underline{Z} = \{1,\ldots, n\}$. Finally, b_{ik} is a parameter which is equal to 1 if and only if species s_k is present in zone z_i and to 0 otherwise. In order to formulate the problem as a mathematical program we use the following variables:

$x_i \in \{0,1\} (i \in \underline{Z})$: $x_i = 1$ if and only if zone z_i is selected;

$y_k \in \{0,1\}$ $(k \in \underline{S})$: $y_k = 1$ if and only if species s_k is protected. This results from the selection of the zones;

$z_l^\tau \in \{0,1\}$ $(\tau \in \underline{T}, l \in \underline{A}^\tau)$: $z_l^\tau = 1$ if and only if the set of protected species allows the preservation of the evolutionary history linked to the arc a_l in the tree T_τ;

$\alpha \geq 0$: α is a working variable used in the calculation of the minimum of several quantities;

$\beta_\tau \in \{0,1\}$ $(\tau \in \underline{T})$: β_τ is a working variable allowing to take into account, or not, the tree T_τ in the optimization of the h_δPD criterion;

$\mathrm{PD}_\tau \geq 0$ $(\tau \in \underline{T})$: phylogenetic diversity, calculated in the tree T_τ, of the set of protected species, *i.e.*, the set of species $\{s_k : k \in \underline{S}, y_k = 1\}$.

Let us now examine the different constraints that are used to formulate the problem under consideration with the 6 criteria proposed to assess phylogenetic diversity.

(C_0): $z_l^\tau = \max_{k \in \underline{F}_l^\tau} y_k$ $(\tau \in \underline{T}, l \in \underline{A}^\tau)$. The Boolean variable z_l^τ must take the value 1 if and only if at least one of species of F_l^τ is protected. These constraints are not linear, we will replace them by the linear constraints (C1);

(C_1): $z_l^\tau \leq \sum_{k \in \underline{F}_l^\tau} y_k$ $(\tau \in \underline{T}, l \in \underline{A}_\tau)$. In the case where one seeks to maximize z_l^τ, these constraints are equivalent to C_0;

(C_2): $\mathrm{PD}_\tau = \sum\limits_{l \in \underline{A}_\tau} \lambda_l^\tau z_l^\tau$ $(\tau \in \underline{T})$;

(C_3): $y_k \leq \sum_{i \in \underline{Z}} b_{ik} x_i$ $(k \in \underline{S})$. Species s_k can be regarded as protected – $y_k = 1$ – only if at least one of the zones where it occurs is selected;

(C_4): $\sum_{i \in \underline{Z}} c_i x_i \leq B$. The total cost associated with the selected zones must not exceed the available budget, B;

(C_5): $x_i \in \{0,1\}$ $(i \in \underline{Z})$, $y_k \in \{0,1\}$ $(k \in \underline{S})$, $z_l^\tau \in \{0,1\}$ $(\tau \in \underline{T}, l \in \underline{A}^\tau)$, $\mathrm{PD}_\tau \geq 0$ $(\tau \in \underline{T})$.

Once these variables and constraints are specified, it is easy to formulate the problem as a mathematical program.

10.10.3.1 Maximization of waPD, rPD, h_δPD and Minimization of lgapPD

When μ is equal to a, wa, r, h_δ, and lgap, the different programs obtained are presented in table 10.15.

For aPD and waPD the economic function, $\sum_{\tau \in \underline{T}} w_\tau \mathrm{PD}_\tau / \sum_{\tau \in \underline{T}} w_\tau$, expresses waPD for the set of protected species – this set resulting from the set of selected

Tab. 10.15 – Mathematical programs associated with the problem when the phylogenetic diversity is measured by μPD with μ = a, wa, r, h_δ, and lgap.

aPD and waPD:

$$\begin{cases} \max \ \sum_{\tau \in \underline{T}} w_\tau \mathrm{PD}_\tau / \sum_{\tau \in \underline{T}} w_\tau \\ \text{s.t.} \mid \mathrm{C}_1,\mathrm{C}_2,\mathrm{C}_3,\mathrm{C}_4,\mathrm{C}_5 \end{cases}$$

rPD:

$$\begin{cases} \max \ \alpha \\ \text{s.t.} \left| \begin{array}{l} \alpha \leq \mathrm{PD}_\tau \ (\tau \in \underline{T}) \\ \mathrm{C}_1,\mathrm{C}_2,\mathrm{C}_3,\mathrm{C}_4,\mathrm{C}_5 \\ \alpha \geq 0 \end{array} \right. \end{cases}$$

h_δPD:

$$\begin{cases} \max \ \alpha \\ \text{s.t.} \left| \begin{array}{l} \alpha \leq \mathrm{PD}_\tau + H\beta_\tau \quad (\tau \in \underline{T}) \\ \sum\limits_{\tau \in \underline{T}} \beta_\tau = \varphi - \delta \\ \mathrm{C}_1,\mathrm{C}_2,\mathrm{C}_3,\mathrm{C}_4,\mathrm{C}_5 \\ \alpha \geq 0 \\ \beta_\tau \in \{0,1\} \quad (\tau \in \underline{T}) \end{array} \right. \end{cases}$$

lgapPD:

$$\begin{cases} \min \ \alpha \\ \text{s.t.} \left| \begin{array}{l} \alpha \geq \mathrm{PD}_\tau^{*C} - \mathrm{PD}_\tau \quad (\tau \in \underline{T}) \\ \mathrm{C}_1,\mathrm{C}_2,\mathrm{C}_3,\mathrm{C}_4,\mathrm{C}_5 \\ \alpha \geq 0 \end{array} \right. \end{cases}$$

zones – and aPD in the special case where $w_\tau = 1$ for all $\tau \in \underline{T}$. For rPD, according to the constraints $\alpha \leq \mathrm{PD}_\tau$, $\tau \in \underline{T}$, and since α is the economic function to be maximized, α takes, at the optimum, the smallest of the φ values DP_τ, $\tau \in \underline{T}$, *i.e.*, the smallest of the φ PD values – in the φ trees considered – associated with the set of protected species. In the program associated with h_δPD, H is a constant large enough to make the constraints $\alpha \leq \mathrm{PD}_\tau + H\beta_\tau$, $\tau \in \underline{T}$, inactive when $\beta_\tau = 1$. The objective is to maximize variable α. Because of these constraints, the value of α is equal, at the optimum, to the smallest of the quantities $\mathrm{PD}_\tau + H\beta_\tau$, $\tau \in \underline{T}$. The constraint $\sum_{\tau \in \underline{T}} \beta_\tau = \varphi - \delta$ force $\varphi - \delta$ variables β_τ to take the value 1. Since we are seeking to maximize α, variables β_τ that will take the value 1 are those whose index corresponds to the $\varphi - \delta$ smallest values of PD_τ, $\tau \in \underline{T}$. At the optimum, α is thus equal to the value of PD_τ which comes just after. Recall that in the program associated with lgapPD, PD_τ^{*C} designates the maximal PD, calculated in the tree T_τ, of a set of species present in a set of zones whose total cost does not exceed B. According to the constraints $\alpha \geq \mathrm{PD}_\tau^{*C} - \mathrm{PD}_\tau$, $\tau \in \underline{T}$, and since α is the economic function to be minimized, α takes, at the optimum, the largest of the φ values $\mathrm{PD}_\tau^{*C} - \mathrm{PD}_\tau$, $\tau \in \underline{T}$, *i.e.*, the value of the largest deviation – in the φ considered trees – that interests us.

10.10.3.2 Maximization of owaPD

The formulation is somewhat more difficult to establish when the phylogenetic diversity of a set of species, $\hat{S}$, is measured by $\mathrm{owaPD}(\hat{S})$. Let $\gamma_{i\tau}$, $i \in \underline{T}, j \in \underline{T}$, be the Boolean variable which takes the value 1 if and only if the weight w_i is assigned to the PD value of $\hat{S}$ calculated in the tree T_τ. For a set of species, $\hat{S}$, included in S and fixed, $\mathrm{owaPD}(\hat{S})$ is equal to the optimal value of the linear program in Boolean variables $\mathrm{P}_{10.14}(\hat{S})$.

$$
\mathrm{P}_{10.14}(\hat{S}) : \begin{cases} \min \sum\limits_{i \in \underline{T}} w_i \sum\limits_{\tau \in \underline{T}} \mathrm{PD}_\tau(\hat{S})\, \gamma_{i\tau} & \\ \text{s.t.} \left| \begin{array}{lll} \sum\limits_{\tau \in \underline{T}} \gamma_{i\tau} = 1 & i \in \underline{T} & (a) \\ \sum\limits_{i \in \underline{T}} \gamma_{i\tau} = 1 & \tau \in \underline{T} & (b) \\ \gamma_{i\tau} \in \{0, 1\} & i \in \underline{T}, \tau \in \underline{T} & (c) \end{array} \right. \end{cases}
$$

Given the economic function to be minimized, the decreasing weights $w_1, w_2, \ldots, w_\varphi$ will be assigned to the PDs of $\hat{S}$ sorted in increasing order. The economic function, therefore, expresses well, at the optimum of $\mathrm{P}_{10.14}(\hat{S})$, the value of owaPD for the set of protected species. It is known that, in this type of programs, the integrality constraints (c) can be relaxed, *i.e.*, replaced by constraints $0 \leq \gamma_{i\tau} \leq 1$, $i \in \underline{T}, \tau \in \underline{T}$, and finally by the constraints $0 \leq \gamma_{i\tau}$ since the constraints (a) and (b) prevent variables $\gamma_{i\tau}$ from exceeding the value 1. Let us consider the dual program of the program thus obtained by associating respectively to the constraints

(a) and (b) the (dual) real variables ε_i and ν_τ. We obtain the mathematical program $\mathrm{P}_{10.15}(\hat{S})$ whose optimal value is equal to that of $\mathrm{P}_{10.14}(\hat{S})$.

$$\mathrm{P}_{10.15}(\hat{S}) : \begin{cases} \max \sum_{i \in \underline{T}} \varepsilon_i + \sum_{\tau \in \underline{T}} \nu_\tau \\ \text{s.t.} \left| \begin{array}{ll} \varepsilon_i + \nu_\tau \leq w_i \mathrm{PD}_\tau(\hat{S}) & i \in \underline{T}, \tau \in \underline{T} \\ \varepsilon_i \in \mathbb{R} & i \in \underline{T} \\ \nu_\tau \in \mathbb{R} & \tau \in \underline{T} \end{array} \right. \end{cases}$$

To solve the problem we are interested in, it is now sufficient to express, in $\mathrm{P}_{10.15}(\hat{S})$ the quantities $\mathrm{PD}_\tau(\hat{S})$ as a function of variables x_i and y_k. We obtain program $\mathrm{P}_{10.16}$.

$$\mathrm{P}_{10.16} : \begin{cases} \max \sum_{i \in \underline{T}} \varepsilon_i + \sum_{\tau \in \underline{T}} \nu_\tau \\ \text{s.t.} \left| \begin{array}{ll} \varepsilon_i + \nu_\tau \leq w_i \mathrm{PD}_\tau & i \in \underline{T}, \tau \in \underline{T} \\ \varepsilon_i \in \mathbb{R} & i \in \underline{T} \\ \nu_\tau \in \mathbb{R} & \tau \in \underline{T} \\ \mathrm{C}_1, \mathrm{C}_2, \mathrm{C}_3, \mathrm{C}_4, \mathrm{C}_5 & \end{array} \right. \end{cases}$$

10.10.3.3 Examples

Let us look again at the 4 phylogenetic trees in figure 10.11. Suppose that the 4 zones, z_1, z_2, z_3, and z_4, are concerned and that the distribution of the 6 species in these zones is as follows: s_1 and s_2 in z_1, s_3 and s_4 in z_2, s_4 and s_5 in z_3, and s_5 and s_6 in z_4. We are interested in selecting a set of zones with a cost less than or equal to 2, the cost of each zone being equal to 1. This amounts to searching for a set of 2 zones that optimizes μPD. The results are presented in table 10.16 for $\mu = \mathrm{r}$, owa, h_2, and lgap. This table also presents, for comparison purposes, the solution to the problem when each tree is considered separately. In other words, for each tree, we give the set of the 2 zones that maximizes the PD calculated in that tree. This table also presents the worst selection of zones that can be made, based on one of the phylogenetic trees, when the correct tree is another tree.

Let us look at the optimal solution with the rPD criterion. It consists in selecting zones z_1 and z_3, which ensures a PD of at least 4.40 regardless of which tree is taken into account for the calculation. This solution is the best one, in the sense that there is no other set of 2 zones ensuring a PD strictly greater than 4.40 for all the 4 trees. Row 2 of table 10.16 indicates that the optimal selection of 2 zones, based on only one of the trees, may lead to a poor solution if this tree is not the right one. In the worst case, the resulting PD would be equal to 3.44. In this case, the optimal solution for the tree T_1 was chosen, whereas the tree T_2 is the right tree. Choosing the solution that maximizes rPD is therefore a good protection against uncertainty since it ensures, in all the cases, a PD at least equal to 4.40 (about + 28%). Note that, in this small example, the solution that maximizes rPD is the same as the one that minimizes the largest gap from the maximal PD solution in each tree. For this

TAB. 10.16 – Worst possible solutions that can be obtained by making a wrong choice in choosing the tree. Solutions optimizing PD and μPD (μ = r, owa, h_2 and lgap) for the 6 species whose phylogenetic trees T_1, T_2, T_3, and T_4 are shown in figure 10.11. These 6 species are distributed into 4 zones, and 2 zones should be selected.

Set of zones of PD max for each tree and corresponding PD: $\{z_1, z_2\}$: 4.72, $\{z_1, z_4\}$: 4.82, $\{z_1, z_3\}$: 4.50, $\{z_1, z_3\}$: 4.47.
The worst solution is obtained by selecting the set of zones of PD max for tree T_1, $\{z_1, z_2\}$, while the correct tree is tree T_2. Corresponding PD in tree T_2: 3.44.
The largest error is made when selecting the set of zones of PD max for tree T_1, $\{z_1, z_2\}$, when the correct tree is tree T_2. Corresponding error in tree T_2: 1.38.
Set of zones of rPD max and corresponding value: $\{z_1, z_3\}$, 4.40. PD of the set of rPD max, for each tree: 4.71, 4.40, 4.50, 4.47.
Set of zones of owaPD max and corresponding value: $\{z_1, z_3\}$, 4.472. PD of the set of owaPD max, for each tree: 4.71, 4.40, 4.50, 4.47.
Set of zones of h_2PD max and corresponding value: $\{z_1, z_4\}$, 4.50. PD of the set of h_2PD max, for each tree: 4.37, 4.82, 4.50, 3.60.

solution, the largest gap is equal to 0.42, while for the solution that maximizes PD in the tree T_1 – $\{z_1, z_2\}$ – this gap is equal to 1.38.

For a group of species, $\hat{S}$, included in S and needing to verify a set of constraints C, lgapPD($\hat{S}$) measures the largest gap – regret – on all the trees of T, between the PD of $\hat{S}$ and the largest PD value that could be obtained on the same tree for a set of species included in S and satisfying C. One could consider the relative gap instead of the absolute gap. This problem can be solved by replacing the constraints $\alpha \geq \mathrm{PD}_\tau^{*C} - \mathrm{PD}_\tau,\ \tau \in \underline{T}$ in the program associated with lgapPD minimization (table 10.15) with constraints $\alpha \geq (\mathrm{PD}_\tau^{*C} - \mathrm{PD}_\tau)/\mathrm{PD}_\tau^{*C},\ \tau \in \underline{T}$.

10.10.4 Reserve Maximizing the ePD of the Considered Species

As before, we consider a set of threatened species S and a set of zones Z where these species live. The zones of Z may or may not be protected. It is assumed that the survival probabilities of the species considered are known, that they are independent and that they are identical for all the considered phylogenies. The survival probability of species s_k in zone z_i is equal to p_{ik} if zone z_i is not protected and to q_{ik} if it is. To simplify the presentation all these probabilities are assumed to be different from 1. The aim is to identify the zones to be protected – reserve R – from a set of candidate zones and under a budgetary constraint, so as to maximize the expected phylogenetic diversity of the set of species under consideration – living in protected or unprotected zones. Given the uncertainties about the phylogeny of the species under consideration, we will consider the same type of criteria as above – aPD, waPD, owaPD, rPD, h$_\delta$PD, and lgapPD – but this time looking at the expected phylogenetic diversity (ePD) rather than phylogenetic diversity (PD). We denote by aePD, waePD, owaePD, rePD, h$_\delta$ePD, and lgapePD the corresponding criteria. These criteria are no longer associated with a subset of species, $\hat{S}$, of the set of species considered, S. They are associated with the set of species considered, S, and their value depends on the reserve selected since the survival probabilities of the different species of S depend on this reserve. As an example, the determination of a set of zones minimizing lgapePD is presented below in detail. The resulting formulation could easily be extended to other measures associated with ePD.

For a set of constraints C to be satisfied by the targeted reserve R, $\text{lgapePD}(R) = \max_{\tau\in\underline{T}}\{\text{ePD}_\tau^{*C} - \text{ePD}_\tau(R)\}$ where ePD_τ^{*C} is equal to the maximal ePD of S associated with a reserve satisfying C and calculated for the tree T_τ, and $\text{ePD}_\tau(R)$ is the ePD of S associated with R and calculated for the tree T_τ. A robust approach based on the concept of regret is adopted here. It consists of determining a – robust – reserve of cost less than or equal to B that minimizes the maximal gap, over all the possible phylogenies, between (1) the expected phylogenetic diversity associated with the optimal reserve in the phylogeny under consideration, calculated with that phylogeny, and (2) the expected phylogenetic diversity associated with the selected reserve, calculated with that same phylogeny. The problem under consideration can, therefore, be stated as follows: $\min_{R\subseteq Z, C(R)\le B} \text{lgapePD}(R)$ or $\min_{R\subseteq Z, C(R)\le B} \left\{\max_{\tau\in\underline{T}}\{\text{ePD}_\tau^{*C} - \text{ePD}_\tau(R)\}\right\}$. In cases where there are no uncertainties about the phylogeny of the species under consideration, the problem of determining a reserve that maximizes the expected phylogenetic diversity of the species concerned while respecting a budgetary constraint was discussed in section 10.8. As in section 10.8, the data are as follows: the available budget, B, a set, Z, of zones eligible for protection, the cost of protecting each zone, denoted by c_i, the list of species present in each zone – among the set, S, of considered species – with the corresponding survival probabilities. The only difference lies in the phylogenetic tree of the species concerned – vertices, arcs and arc lengths. Here several trees are considered plausible. Let us designate by As_l^τ the set of arcs directly following arc a_l in the phylogenetic tree T_τ, *i.e.*, the set of arcs whose initial end coincides with the terminal end of the arc a_l. Let us designate by $\underline{\text{As}}_l^\tau$ the set of corresponding indices. We have the equality $\sigma_l^\tau = \prod_{j\in\underline{\text{As}}_l^\tau} \sigma_j^\tau$ where σ_l^τ, $l \in \underline{A}^\tau$, is a real variable that represents the probability that the information associated with the arc a_l of the phylogenetic tree T_τ is not retained. Variable $\hat{\sigma}_l^\tau$ designates the logarithm of this probability. For each pending arc a_l of the tree T_τ, we denote by $\text{ext}^\tau(l)$ the index of the terminal end of this arc, *i.e.*, the index of the corresponding species. We denote by A_p^τ the set of pending arcs in tree T_τ and by $\underline{A}_p^\tau$ the set of corresponding indices. For the phylogenetic tree T_τ, the problem of determining the reserve that leads to an ePD of S close to the optimum value can be solved by program $\text{P}_{10.17}(\tau)$, identical to program $\text{P}_{10.10}$ provided that $\tilde{\sigma}_l$ is replaced by $\tilde{\sigma}_l^\tau$ and $\hat{\sigma}_l$ by $\hat{\sigma}_l^\tau$.

$$
\mathrm{P}_{10.17}(\tau):\left\{\begin{array}{ll}
\max & \sum\limits_{l\in\underline{A}^\tau}\lambda_l^\tau(1-\tilde{\sigma}_l^\tau) \\
\text{s.t.} & \left|\begin{array}{lll}
\hat{\sigma}_l^\tau \le \frac{\tilde{\sigma}_l^\tau}{u_v}+\log u_v - 1 & l\in\underline{A}^\tau, v=1,\ldots,V & (10.17_\tau.1)\\
\sum\limits_{i\in\underline{Z}} c_i x_i \le B & & (10.17_\tau.2)\\
\hat{\sigma}_l^\tau = \sum\limits_{j\in\underline{\mathrm{As}}_l^\tau}\hat{\sigma}_j^\tau & l\in\underline{A}^\tau-\underline{A}_p^\tau & (10.17_\tau.3)\\
\hat{\sigma}_l^\tau = \sum\limits_{i\in\underline{Z}} x_i\log(1-q_{ik}) & & \\
\quad+\sum\limits_{i\in\underline{Z}}(1-x_i)\log(1-p_{ik}) & l\in\underline{A}_p^\tau,\ k=\mathrm{ext}^\tau(l) & (10.17_\tau.4)\\
x_i\in\{0,1\} & i\in\underline{Z} & (10.17_\tau.5)\\
0\le\tilde{\sigma}_l^\tau\le 1,\ \hat{\sigma}_l^\tau\le 0 & l\in\underline{A}^\tau & (10.17_\tau.6)
\end{array}\right.
\end{array}\right.
$$

Let us denote by $\mathrm{ePD}_a^{\tau*}$ the ePD associated with the reserve corresponding to the optimal solution of $\mathrm{P}_{10.17}(\tau)$, *i.e.*, a "good" value of the expected phylogenetic diversity that can be obtained – by protecting an adequate set of zones – in the case of the phylogeny T_τ, $\tau\in\underline{T}$. Let us now consider the determination of a robust reserve. As stated above a robust reserve is here a reserve that minimizes the maximal gap, over all possible phylogenies, between (1) the expected phylogenetic diversity associated with the optimal reserve in the phylogeny under consideration, and (2) the expected phylogenetic diversity associated with the selected reserve, calculated with the phylogeny under consideration. An approximate solution to this problem, close to the optimal solution, can be determined by solving the mathematical program $\mathrm{P}_{10.18}$.

$$
\mathrm{P}_{10.18}:\left\{\begin{array}{ll}
\min & \alpha \\
\text{s.t.} & \left|\begin{array}{lll}
\alpha\ge\mathrm{ePD}_a^{\tau*}-\psi^\tau & \tau\in\underline{T} & (10.18.1)\\
\psi^\tau=\sum\limits_{l\in\underline{A}^\tau}\lambda_l^\tau(1-\tilde{\sigma}_l^\tau) & \tau\in\underline{T} & (10.18.2)\\
\hat{\sigma}_l^\tau \le \frac{\tilde{\sigma}_l^\tau}{u_v}+\log u_v - 1 & \tau\in\underline{T}, & \\
 & l\in\underline{A}^\tau, v=1,\ldots,V & (10.18.3)\\
\sum\limits_{i\in\underline{Z}} c_i x_i \le B & & (10.18.4)\\
\hat{\sigma}_l^\tau = \sum\limits_{j\in\underline{\mathrm{As}}_l^\tau}\hat{\sigma}_j^\tau & \tau\in\underline{T}, l\in\underline{A}^\tau-\underline{A}_p^\tau & (10.18.5)\\
\hat{\sigma}_l^\tau = \sum\limits_{i\in\underline{Z}}[x_i\log(1-q_{ik}) & & \\
\quad+(1-x_i)\log(1-p_{ik})] & \tau\in\underline{T}, l\in\underline{A}_p^\tau, & \\
 & k=\mathrm{ext}^\tau(l) & (10.18.6)\\
x_i\in\{0,1\} & i\in\underline{Z} & (10.18.7)\\
0\le\tilde{\sigma}_l^\tau\le 1,\ \hat{\sigma}_l^\tau\le 0 & \tau\in\underline{T}, l\in\underline{A}^\tau & (10.18.8)\\
\psi^\tau\ge 0 & \tau\in\underline{T} & (10.18.9)
\end{array}\right.
\end{array}\right.
$$

The economic function consists in minimizing variable α. Because of constraints 10.18.1, at the optimum of $P_{10.18}$, this variable is equal to the maximal gap – maximal regret – over all the phylogenies between (1) the expected phylogenetic diversity associated with a reserve close to the optimal one in the phylogeny under consideration, $ePD_a^{\tau*}$, and (2) a "good" approximation of the expected phylogenetic diversity associated with the selected reserve and calculated with the phylogeny under consideration, ψ^τ. Indeed, the constraints of $P_{10.18}$ require variable ψ^τ to be equal to an approximation of the ePD, calculated in the phylogeny T_τ, associated with the selected reserve. Recall that the quantity $ePD_a^{\tau*}$, $\tau \in \underline{T}$, is calculated beforehand by solving program $P_{10.17}(\tau)$. As regards the meaning of constraints 10.18.3, 10.18.5, and 10.18.6, the reader may refer again to program $P_{10.10}$ in which similar constraints are used. The only difference is that, in program $P_{10.18}$, these constraints are considered for each phylogeny T_τ. Note that if one were interested in determining a reserve that maximizes rePD, *i.e.*, the smallest value, over all possible phylogenies, of the expected phylogenetic diversity associated with this reserve, a reserve close to this optimal robust reserve could be determined by solving program $P_{10.18}$ in which the objective becomes max α and the constraints $\alpha \geq ePD^{\tau*} - \psi^\tau$, $\tau \in \underline{T}$, are replaced by the constraints $\alpha \leq \psi^\tau$, $\tau \in \underline{T}$.

References and Further Reading

Arponen A., Zupan L. (2016) Representing hotspots of evolutionary history in systematic conservation planning for European mammals, *Biodiversity conservation and phylogenetic systematics, topics in biodiversity and conservation* (R. Pellens, P. Grandcolas, Eds). Springer, 14.

Asmyhr M.G., Linke S., Hose G., Nipperess D.A. (2014) Systematic conservation planning for groundwater ecosystems using phylogenetic diversity, *PLoS One* **9**, e115132.

Billionnet A. (2013) Solution of the generalized Noah's Ark Problem, *Syst. Biol.* **62**, 147.

Billionnet A. (2016) How to take into account uncertainty in species extinction probabilities for phylogenetic conservation prioritization, *Environ. Model. Assess.* **22**, 535.

Billionnet A. (2018a) Quantifying extinction probabilities of endangered species for phylogenetic conservation prioritization may not be as sensitive as might be feared, *Biodivers. Conserv.* **27**, 1189.

Billionnet A. (2018b) Phylogenetic conservation prioritization with uncertainty, *Biodivers. Conserv.* **27**, 3137.

Bininda-Edmonds O.R.P., Gittleman J.L., Purvis A. (1999) Building large trees by combining phylogenetic information: a complete phylogeny of the extant Carnivora (Mammalia), *Biol. Rev.* **74**, 143.

Bordewich M., Rodrigo A., Semple C. (2008) Selecting taxa to save or sequence: desirable criteria and a greedy solution, *Syst. Biol.* **57**, 825.

Bordewich M., Semple C. (2008) Nature reserve selection problem: a tight approximation algorithm, *IEEE/ACM Trans. Comput. Biol. Bioinfo.* **5**, 275.

Bradley S.P., Hax A.C., Magnanti T.L. (1977) *Applied mathematical programming.* Addison-Wesley, Reading, MA, p. 539.

Buerki S., Callmander M.W., Bachman S., Moat J., Labat J.-N., Forest F. (2015) Incorporating evolutionary history into conservation planning in biodiversity hotspots, *Philos. Trans. R. Soc. London, Ser. B* **370**, 20140014.

Calvo T., Mayor G., Mesiar R., Eds (2002) *Aggregation operators: new trends and applications.* Physica-Verlag, Heidelberg.

Chao A., Chiu C.-H., Jost L. (2010) Phylogenetic diversity measures based on Hill numbers, *Philos. Trans. R. Soc. London, Ser. B* **365**, 3599.

Chernomor O., Klaere S., von Haeseler A., Minh B.Q. (2016) Split Diversity: measuring and optimizing biodiversity using phylogenetic split networks, *Biodiversity conservation and phylogenetic systematics, topics in biodiversity and conservation* (R. Pellens, P. Grandcolas, Eds). Springer, 14.

Collen B. (2015) Conservation prioritization in the context of uncertainty, *Anim. Conserv.* **18**, 315.

Donoghue M.J., Ackerly D.D. (1996) Phylogenetic uncertainties and sensitivity analyses in comparative biology,*Philos. Trans. R. Soc. London, Ser. B* **351**, 1241.

Faith D.P. (1992a) Conservation evaluation and phylogenetic diversity, *Biol. Conserv.* **61**, 1.

Faith D.P. (1992b) Systematics and conservation: on predicting the feature diversity of subsets of taxa, *Cladistics* **8**, 361.

Faith D.P. (1994) Phylogenetic pattern and the quantification of organismal biodiversity, *Philos. Trans. R. Soc. London, Ser. B* **345**, 45.

Faith D.P., Carter G., Cassis G., Ferrier S., Wilkie L. (2003) Complementarity, biodiversity viability analysis, and policy-based algorithms for conservation, *Environ. Sci. Policy* **6**, 311.

Faith D.P., Baker A.M. (2006) Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges, *Evol. Bioinfo. Online* **2**, 121.

Faith D.P., Williams K.J. (2006) Yearbook of science and technology, *Phylogenetic diversity and biodiversity conservation.* McGraw-Hill, 233.

Faith D.P. (2008a) Phylogenetic diversity and conservation, *Conservation biology, evolution in action* (S.P. Carroll, C.W. Fox, Eds). Oxford University Press.

Faith D.P. (2008b) Threatened species and the potential loss of phylogenetic diversity: conservation scenarios based on estimated extinction probabilities and phylogenetic risk analysis, *Conserv. Biol.* **22**, 1461.

Faith D.P. (2013) Biodiversity and evolutionary history: useful extensions of the PD phylogenetic diversity assessment framework, *Ann. N.Y. Acad. Sci.* **1289**, 69.

Faith, D.P. (2016) The PD phylogenetic diversity framework: linking evolutionary history to feature diversity for biodiversity conservation, *Biodiversity conservation and phylogenetic systematics, topics in biodiversity and conservation* (R. Pellens, P. Grandcolas, Eds). Springer, 14.

Faller B., Pardi F., Steel M. (2008) Distribution of phylogenetic diversity under random extinction, *J. Theor. Biol.* **251**, 286.

Hartmann K., Steel M. (2006) Maximizing phylogenetic diversity in biodiversity conservation: greedy solutions to the Noah's Ark problem, *Syst. Biol.* **55**, 644.

Hickey G., Carmi P., Maheshwari A., Zeh N. (2008) NAPX: a Polynomial time approximation scheme for the Noah's Ark Problem, *WABI '08 Proceedings of the 8th international workshop on algorithms in bioinformatics*. Springer-Verlag, 76.

Hickey G., Blanchette M., Carmi P., Maheshwari A., Zeh N. (2011) An approximation algorithm for the Noah's Ark problem with random features loss, *IEEE/ACM Trans. Comput. Biol. Bioinfo.* **8**, 551.

Isaac N.J.B., Turvey S.T., Collen B., Waterman C., Baillie J.E.M. (2007) Mammals on the edge: conservation priorities based on threat and phylogeny, *PloS One* **3**, e296.

iTOL, version 4.0.3, Interactive tree of life, An online tool for the display, annotation and management of phylogenetic trees, Available at: https://itol.embl.de/.

Kouvelis P., Yu G. (1997) *Robust discrete optimization and its applications*. Kluwer Academic Publishers.

Lean C., Maclaurin J. (2016) The Value of Phylogenetic diversity, *Biodiversity conservation and phylogenetic systematics, topics in biodiversity and conservation* (R. Pellens, P. Grandcolas, Eds). Springer, 14.

Llamazares B. (2007) Choosing OWA operator weights in the field of social choice, *Info. Sci.* **177**, 4745.

MacLaurin J., Sterelny K. (2008) *What is biodiversity?* The University of Chicago Press.

Mimouni E., Beisner B.E, Pinel-Alloul B. (2016) Phylogenetic diversity and its conservation in the presence of phylogenetic uncertainty: a case study of cladoceran communities in urban waterbodies, *Biodivers. Conserv.* **25**, 2113.

Minh B.Q., Klaere S., Von Haeseler A. (2006) Phylogenetic diversity within seconds, *Syst. Biol.* **55**, 769.

Mooers A., Gascuel O., Stadler T., Li H., Steel M. (2012) Branch lengths on birth-death trees and the expected loss of phylogenetic diversity, *Syst. Biol.* **61**, 195.

Moreno C.E. *et al.* (2017) Measuring biodiversity in the anthropocene: a simple guide to helpful methods, *Biodivers. Conserv.* **26**, 2993.

Moulton V., Semple C., Steel M. (2007) Optimizing phylogenetic diversity under constraints, *J. Theor. Biol.* **246**, 186.

Nemhauser G.L., Wolsey L.A. (1988) *Integer and combinatorial optimization*. Wiley.

Nipperess D.A. (2016) The rarefaction of phylogenetic diversity: formulation, extension and application, *Biodiversity conservation and phylogenetic systematics, topics in biodiversity and conservation* (R. Pellens, P. Grandcolas, Eds). Springer, 14.

Ogryczak W., Sliwinski T. (2003) On solving linear programs with the ordered weighted averaging objective, *Euro. J. Oper. Res.* **148**, 80.

Olender P. (2016) Hybrid models for the OWA optimization, *J. Telecommun. Inf. Technol.* **4**, 22.

Pardi F., Goldman N. (2007) Resource-aware taxon selection for maximizing phylogenetic diversity, *Syst. Biol.* **56**, 431.

Pardi F. (2009) Algorithms on phylogenetic trees. PhD Thesis, University of Cambridge, St Catharine's College.

Pellens R., Grandcolas P., Eds (2016) *Biodiversity conservation and phylogenetic systematics: preserving our evolutionary heritage in an extinction crisis*. Springer.

Perry N. (2010) The ecological importance of species and the Noah's Ark problem, *Ecol. Econ.* **69**, 478.

Pollock L., Thuiller W., Jetz W. (2017) Large conservation gains possible for global biodiversity facets, *Nature* **546**, 141.

Purvis A., Hector A. (2000) Getting the measure of biodiversity, *Nature* **405**, 212.

Rangel T.F., Colwell R.K., Graves G.R., Fučíková K., Rahbek C., Diniz-Filho J.A.F. (2015) Phylogenetic uncertainty revisited: implications for ecological analyses, *Evol.* **69**, 1558.

Rodrigues A.S.L., Gaston K.J. (2002) Maximising phylogenetic diversity in the selection of networks of conservation areas, *Biol. Conserv.* **105**, 103.

Scherson R.A., Faith D.P., Eds (2018) *Phylogenetic diversity, applications and challenges in biodiversity science*. Springer.

Steel M., McKenzie A. (2002) The "shape" of phylogenies under simple random speciation models, *Proceedings of biological evolution and statistical physics conference* (M. Lässig, A. Valleriani, Eds). LNP 585, Springer-Verlag, pp. 162–180.

Steel M. (2005) Phylogenetic diversity and the greedy algorithm, *Syst. Biol.* **54**, 527.

Steel M., Mooers A. (2010) The expected length of pendant and interior edges of a Yule tree, *Appl. Math. Lett.* **23**, 1315.

Swenson N.G. (2009) Phylogenetic resolution and quantifying the phylogenetic diversity and dispersion of communities, *PLoS One* **4**, e4390.

Van Dyke F. (2008) *Conservation biology, foundations, concepts, applications*, Springer, ch. 4.

Vellend M., Cornwell W.K., Magnuson-Ford K., Mooers A.O. (2010) Measuring phylogenetic biodiversity, *Biological diversity: frontiers in measurement and assessment* (A.E. Magurran, B. J. McGill, Eds). Oxford University Press.

Weitzman M.L. (1992) On diversity, *Q. J. Econ.* **107**, 363.

Weitzman M.L. (1998) The Noah's ark problem, *Econometrica* **66**, 1279.

Winter M., Devictor V., Schweiger O. (2013) Phylogenetic diversity and nature conservation: where are we? *Trends Ecol. Evol.* **28**, 199.

Witting L., Loeschcke V. (1995) The optimization of biodiversity conservation, *Biol. Conserv.* **71**, 205.

Yager R.R. (1988) On ordered weighted averaging aggregation operators in multi-criteria decision making, *IEEE Trans. Syst. Man Cybern.* **18**, 183.

Yager R.R., Kacprzyk J., Eds (1997) *The ordered weighted averaging operators: theory and applications*. Kluwer Academic Publishers.

Yule G.U. (1924) A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, *Philos. Trans. R. Soc. London, Ser. B* **213**, 21.

Chapter 11

Specific and Genetic Diversity

11.1 Introduction

In the first part of this chapter, we are interested, as in the previous chapters, in a set of threatened plant or animal species, $S = \{s_1, s_2, \ldots, s_m\}$, living in a set of zones, $Z = \{z_1, z_2, \ldots, z_n\}$, which may be protected from a given moment. The value of protecting a set of zones, R, included in Z is assessed by the diversity of the species that are protected – at least in some way – as a result of the protection of the zones in R, this diversity being calculated in three different ways. In the first, the "dissimilarity" or "distance" between 2 species is taken into account. This can be, for example, the genetic distance calculated from differences between DNA sequences. In this case, the aim is to protect a set of species that maximizes an overall distance. In the other two ways, we look at the species diversity of the protected set of species using two classical indices, Simpson diversity index and Shannon–Wiener diversity index, to measure the overall species diversity of the protected populations. Both indices take into account at the same time species richness and abundance of each species. It is assumed that for each zone, the population size of each of the species concerned by the protection of that zone is known. We denote by n_{ik} the population size of species s_k in zone z_i. Note that these data can be difficult to obtain. We know the set of zones whose protection makes it possible to protect species s_k – for example, to ensure its survival – and this for all the species of S, that is to say for all $k \in \underline{S}$. This set is denoted by Z_k and the corresponding set of indices is denoted by $\underline{Z_k}$. In other words, protecting species s_k requires, and it is sufficient, that at least one zone of Z_k is protected. The reserve formed by zones z_1, z_{12}, and z_{17}, shown in the example in figure 11.1, protect species s_2, s_3, s_9, and s_{11}. The value of protecting the set of zones $\{z_1, z_{12}, z_{17}\}$ is, therefore, assessed by the diversity of the set of species $\{s_2, s_3, s_9, s_{11}\}$.

In the second part of this chapter, we will look at a set of individuals, $I = \{I_1, I_2, \ldots, I_m\}$, of a given species and concerned with the protection of zones of Z. We know, for each zone, the list of the individuals of this set who live there.

DOI: 10.1051/978-2-7598-2521-9.c011

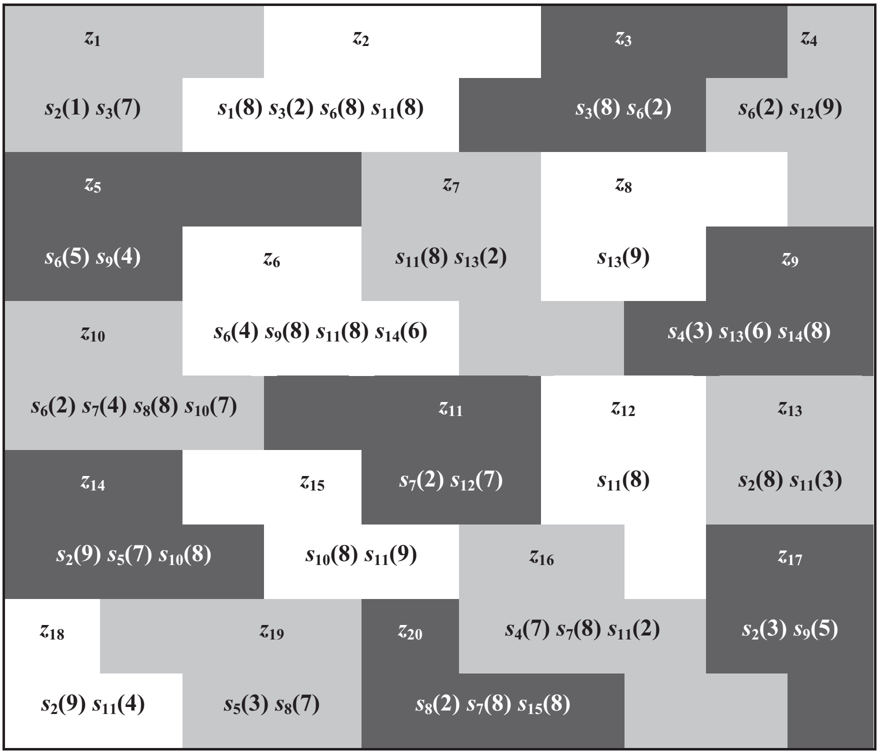


FIG. 11.1 – The 20 zones z_1, z_2,..., z_{20} are candidates for protection and the 15 species s_1, s_2, ..., s_{15} are concerned by the protection of these zones. For each zone, the species concerned by the protection of this zone as well as the size of their population – in brackets – are indicated. The cost of protecting the white zones is equal to 1 unit, the cost of protecting the light grey zones is equal to 2 units and the cost of protecting the dark grey zones is equal to 4 units. For example, species s_6, s_9, s_{11}, and s_{14} are concerned by the protection of zone z_6, their population size is 4, 8, 8, and 6 units, respectively, and the cost of protecting this zone is 1 unit.

We then consider the problem of determining, taking into account a budgetary constraint, a reserve that enables a given number of individuals to be protected while minimizing the average kinship of the protected population. Throughout this chapter, it is considered that there is only one level of protection of zones, *i.e.*, a zone is protected or not. The protection of each zone has a cost and this cost depends on the zone; it is denoted by c_i for zone z_i. The cost of protecting a set of zones $R \subseteq Z$ is equal to the sum of the costs of protecting each of the zones in that set and is denoted by $C(R)$. $\underline{Z}$ is the set of indices for the zones in Z, $\underline{S}$ is the set of indices for the species under consideration, *i.e.*, the species in S, and $\underline{I}$ is the set of indices for the individuals under consideration. Thus we have $\underline{Z} = \{1, 2, \ldots, n\}$ and $\underline{S} = \underline{I} = \{1, 2, \ldots, m\}$.

11.2 Reserve Maximizing the Overall Dissimilarity of the Species – of a Given Set – Present in It

11.2.1 The Problem and Its Mathematical Programming Formulation

The problem is to determine the zones to be protected, taking into account the available budget, B, so as to maximize a certain biological diversity of the subset of protected species. Here, the biological diversity of a set of species is measured by the dissimilarities or distances between species. Let us denote by d_{kl} the distance between species s_k and s_l. This quantity, positive or zero, satisfies $d_{kl} = d_{lk}$ and $d_{kk} = 0$. The overall diversity of a group of species, $\hat{S}$, included in S, is denoted by $D(\hat{S})$ and is equal to the sum of the distances between all the pairs of species of $\hat{S}$:

$$D(\hat{S}) = \sum_{(k,l)\,\in\,\underline{\hat{S}}^2,\; k<l} d_{kl}$$

where $\underline{\hat{S}}$ denotes the set of indices of the species of $\hat{S}$. With each zone z_i of Z is associated the Boolean variable x_i which is equal to 1 if and only if zone z_i is selected for protection. With each species s_k is associated the Boolean variable y_k, which is equal to 1 if and only if species s_k is protected – due to the protection of certain zones of Z. When the interest of a reserve R is evaluated by $D(\hat{S})$, where $\hat{S}$ denotes the species protected by R, this interest is denoted by $D(R)$. The problem can be formulated as $\mathrm{P}_{11.1}$.

$$\mathrm{P}_{11.1}: \begin{cases} \max \displaystyle\sum_{(k,\,l)\,\in\,\underline{S}^2,\; k<\,l} d_{kl}\, y_k\, y_l \\ \text{s.t.} \left| \begin{array}{lll|lll} \displaystyle\sum_{i\in\underline{Z}} c_i x_i \le B & & (11.1.1) & x_i \in \{0,1\} & i \in \underline{Z} & (11.1.3) \\ y_k \le \displaystyle\sum_{i\,\in\,\underline{Z}_k} x_i & k \in \underline{S} & (11.1.2) & y_k \in \{0,1\} & k \in \underline{S} & (11.1.4) \end{array} \right. \end{cases}$$

The mathematical program $\mathrm{P}_{11.1}$ is a non-linear program in 0–1 variables: the economic function to be maximized is a quadratic function since each term involves the product of two variables, $y_k y_l$. Many approaches exist to solve this type of program (see appendix at the end of the book). Classically, to obtain an optimal solution of $\mathrm{P}_{11.1}$ at a lower cost, one can subtract from its economic function the quantity $\varepsilon \sum_{i\in\underline{Z}} c_i x_i$ where ε is a sufficiently small constant.

11.2.2 Example of Application to the Protection of Cranes

Let us consider the 14 species of cranes presented in table 11.1. Many species of cranes are endangered. This is the case, for example, of the Siberian crane classified as Critically Endangered by IUCN. The main threat is due to the draining of

TAB. 11.1 – The 14 species of crane considered.

s_1 South African crane	s_2 Demoiselle crane	s_3 Blue crane	s_4 Wattled crane	s_5 Siberian crane	s_6 Sandhill crane	s_7 Sarus crane
s_8 Australia crane	s_9 White naped crane	s_{10} Eurasian crane	s_{11} Hooded crane	s_{12} Whooping crane	s_{13} Black-necked crane	s_{14} Japanese crane

swamps to produce agricultural land. In this example, we use the "genetic distances" calculated by Krajewski (Krajewski 1989) and presented in table 11.2. These distances represent the differences between DNA sequences associated with the two species. They are determined by DNA hybridization.

We are interested in the problem of determining a reserve, R, that respects a budget constraint and maximizes $D(R)$. Figure 11.2 shows a distribution of these 14 crane species over 20 hypothetical candidate zones for protection. If, in figure 11.2, the species s_k is mentioned in the zone z_i, then the protection of z_i results in protection of s_k. Table 11.3 gives the optimal reserves that can be obtained by solving $P_{11.1}$ with different values of the available budget, B. It can be seen in this table, as one would expect since the criterion to be maximized is the overall distance between protected species, that this overall distance increases with B.

11.3 Reserve Maximizing the Specific Diversity of Species – of a Given Set – Present in It, as Measured by the Simpson and Shannon–Wiener Indices

The biological conservation literature offers a wide range of indices to measure the diversity of a population with individuals of different species. The purpose of these indices is to try to measure the diversity of the population concerned by a single number, the index value. The functional meaning of these indices is often not obvious and slightly different interpretations may appear in the literature. While these indices are relatively difficult to interpret in a very precise way, they can be useful, in the field we are interested in, to compare reserves under certain conditions. The measurement of the diversity of a population – faunistic or floristic – must take into account, in the classical way, on the one hand, the species richness, *i.e.*, the number of species making up this population, and, on the other hand, the relative abundance of each species. The abundance of a species in a population is the total number of individuals of that species present in that population. The relative abundance of a species is equal to the number of individuals of that species divided

TAB. 11.2 – Genetic distances between 14 crane species (Krajewski 1989).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0.00	3.75	3.85	4.10	3.55	3.90	3.70	3.60	3.60	3.55	4.05	3.65	3.55	3.80
2	3.75	0.00	0.50	1.10	1.80	1.35	1.50	1.15	1.05	1.00	1.05	1.25	1.50	1.55
3	3.85	0.50	0.00	1.25	1.90	1.30	1.75	1.00	1.15	1.05	1.20	1.30	1.15	1.75
4	4.10	1.10	1.25	0.00	1.55	1.20	1.50	1.40	1.35	1.10	1.60	1.30	1.25	1.40
5	3.55	1.80	1.90	1.55	0.00	1.45	1.15	1.50	1.60	1.25	1.55	1.65	1.50	1.65
6	3.90	1.35	1.30	1.20	1.45	0.00	1.40	1.20	1.10	1.10	1.45	1.40	1.75	1.55
7	3.70	1.50	1.75	1.50	1.15	1.40	0.00	0.60	0.50	1.15	1.80	1.45	1.50	1.40
8	3.60	1.15	1.00	1.40	1.50	1.20	0.60	0.00	0.65	1.10	1.40	1.50	1.75	1.35
9	3.60	1.05	1.15	1.35	1.60	1.10	0.50	0.65	0.00	1.10	1.15	1.35	1.30	1.05
10	3.55	1.00	1.05	1.10	1.25	1.10	1.15	1.10	1.10	0.00	0.20	0.15	0.60	0.35
11	4.05	1.05	1.20	1.60	1.55	1.45	1.80	1.40	1.15	0.20	0.00	0.35	0.60	0.55
12	3.65	1.25	1.30	1.30	1.65	1.40	1.45	1.50	1.35	0.15	0.35	0.00	0.65	0.65
13	3.55	1.50	1.15	1.25	1.50	1.75	1.50	1.75	1.30	0.60	0.60	0.65	0.00	0.65
14	3.80	1.55	1.75	1.40	1.65	1.55	1.40	1.35	1.05	0.35	0.55	0.65	0.65	0.00

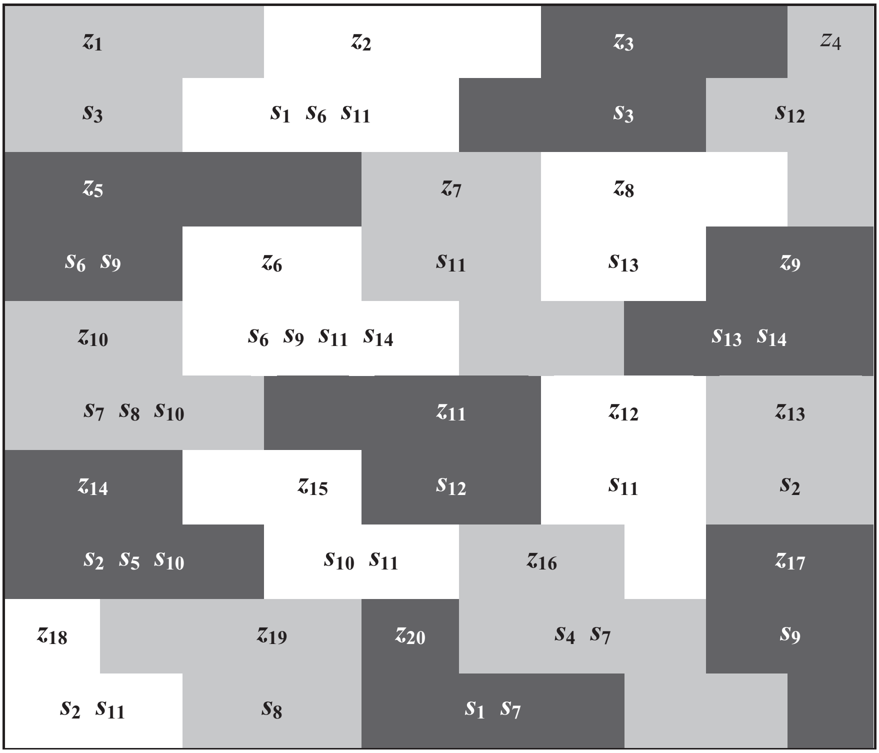


FIG. 11.2 – The 20 hypothetical zones z_1, z_2,..., z_{20} are candidates for protection and the 14 crane species, s_1, s_2,..., s_{14}, concerned by the protection of these zones are indicated. The cost of protecting the white zones is equal to 1 unit, the cost of protecting the light grey zones is equal to 2 units and the cost of protecting the dark grey zones is equal to 4 units.

TAB. 11.3 – Reserves of maximal dissimilarity associated with the data in table 11.2 and figure 11.2, for different values of the available budget, B.

Available budget (B)	Used budget	Zones selected for protection (R)	Protected species	Total dissimilarity ($D(R)$)
4	4	z_2 z_6 z_{10}	s_1 s_6 s_7 s_8 s_9 s_{10} s_{11} s_{14}	48.35
6	6	z_2 z_6 z_8 z_{10} z_{18}	s_1 s_2 s_6 s_7 s_8 s_9 s_{10} s_{11} s_{13} s_{14}	73.95
8	8	z_2 z_6 z_8 z_{10} z_{16} z_{18}	s_1 s_2 s_4 s_6 s_7 s_8 s_9 s_{10} s_{11} s_{13} s_{14}	89.95
14	13	z_1 z_2 z_6 z_8 z_{10} z_{14} z_{16}	All but s_{12}	126.35
15	15	z_1 z_2 z_4 z_6 z_8 z_{10} z_{14} z_{16}	All	143.00

by the total number of individuals. When looking at the diversity of a population protected by a reserve, it is interesting to take into account the species richness of this population as well as the relative abundance of each of the species composing it. Let us consider, for example, two populations composed of 5 species and 35 individuals. In the first, the number of individuals of each species is equal to 9, 8, 12, and 6, respectively, and in the second, it is equal to 5, 20, 4, and 6, respectively. The first population is considered more diverse than the second. It is also considered that the diversity of a reserve with many species and only one dominant species is not really more interesting than the diversity of a reserve with fewer species but in similar abundance. The indices proposed in the literature give more or less importance to these two aspects of diversity. Recall that here we are making the hypothesis that, for any reserve under consideration, the protected population is perfectly known – the list of protected species and the population size of each of these species. We will examine two widely used indices to measure the diversity of a reserve in terms of species richness and relative abundance: the Simpson index and the Shannon–Wiener index. Note that, since we are interested here in identifying reserves with the aim of preserving biodiversity as much as possible, we will need to consider, in addition to the diversity indices just mentioned, the total number of species protected by that reserve.

11.3.1 The Simpson Index

This index was proposed by Simpson in 1949. It measures the probability that two randomly selected individuals in a population of several species do not belong to the same species. It is therefore equal to $1 - \sum_{k \in \underline{S}} f_k^2$ where f_k denotes the frequency of species s_k in the population under consideration. The value of this index starts with 0 – the minimal diversity – and is increasing as the diversity increases, tending towards 1. This index is more sensitive to abundant species than to species richness. Thus, adding rare species to a population hardly changes the value of the index. Consider, for example, a population of 50 individuals divided into 5 species whose respective population sizes are as follows: 7, 12, 10, 9, and 12. The Simpson diversity index associated with this population is equal to 0.79. Let us add to this population a sixth species, comprising 2 individuals. The value of the Simpson diversity index becomes equal to 0.81 (+2.5%). The Simpson diversity index for a given population can be divided by the maximal value that the index can take, given the number of species in that population. The resulting ratio, sometimes referred to as the Simpson evenness index, then reflects the degree of diversity achieved by that population relative to the theoretical maximum. This ratio expresses the dominance of a species, when it tends towards 0, or the fact that the population sizes of the different species are close together, when it tends towards 1. For example, let us consider again the population examined above, composed of 50 individuals distributed among 5 species. The maximal value that the Simpson diversity index can take for a population of 50 individuals distributed among 5 species is equal to $1 - \sum_{k=1,\ldots,5} (1/5)^2 = 0.8$. The value of the ratio in this case is, therefore, equal to 0.99. Recall that n_{ik} refers to the population size of species s_k in zone z_i. The population size of species s_k in a reserve R $(\subseteq Z)$ is therefore equal to $\sum_{i \in \underline{R}} n_{ik}$ and the total population size of the reserve, including all species, is therefore equal to $\sum_{k \in \underline{S}} \sum_{i \in \underline{R}} n_{ik}$. It is deduced that the frequency of species s_k in reserve R, denoted by $f_k(R)$, is equal to $\sum_{i \in \underline{R}} n_{ik} / \sum_{k \in \underline{S}} \sum_{i \in \underline{R}} n_{ik}$. The diversity of the population associated with – protected by – reserve R and measured by the Simpson diversity index, is denoted by DSI(R) and is therefore equal to $\mathrm{DSI}(R) = 1 - \sum_{k\in\underline{S}} [f_k(R)]^2 = 1 - \sum_{k\in\underline{S}} \left(\sum_{i\in\underline{R}} n_{ik} / \sum_{k\in\underline{S}} \sum_{i\in\underline{R}} n_{ik}\right)^2$.

With regard to the selection of protected zones, several problems naturally arise in relation to the Simpson diversity index. We consider the two problems below.

Problem I. Select a reserve with a maximal Simpson diversity index and a budget constraint.

Problem II. Select a minimum cost reserve in order to protect all the species concerned on the one hand, and to maximize the Simpson diversity index on the other hand.

These two problems can be enhanced by adding, for example, a constraint on the minimum total number of individuals to be protected by the reserve.

Mathematical programming formulation of problem I. Let x_i be the Boolean decision variable which is equal to 1 if and only if zone z_i is selected to be part of the reserve and let f_k be the positive or null real variable which represents the frequency of species s_k in the reserve. We get the mixed-integer non-linear program $\mathrm{P}_{11.2}$.

$$\mathrm{P}_{11.2}: \begin{cases} \max \;\; 1 - \sum\limits_{k\in\underline{S}} f_k^2 \\ \text{s.t.} \left| \begin{array}{llll} \sum\limits_{i\in\underline{Z}} n_{ik}x_i = f_k \sum\limits_{i\in\underline{Z}} \sum\limits_{j\in\underline{S}} n_{ij}x_i \quad k\in\underline{S} & (11.2.1) & \mid \quad x_i \in \{0,1\} \quad i \in \underline{Z} & (11.2.4) \\ \sum\limits_{i\in\underline{Z}} \sum\limits_{j\in\underline{S}} n_{ij}x_i \geq 1 & (11.2.2) & \mid \quad 0 \leq f_k \leq 1 \quad k\in\underline{S} & (11.2.5) \\ \sum\limits_{i\in\underline{Z}} c_i x_i \leq B & (11.2.3) & \mid & \end{array} \right. \end{cases}$$

The economic function to be maximized represents the Simpson diversity index associated with the set of species protected by the selected reserve. The quantity $\sum_{i\in\underline{Z}} n_{ik}x_i$ represents the size of the population of species s_k in the reserve and the quantity $\sum_{i\in\underline{Z}} \sum_{j\in\underline{S}} n_{ij}x_i$ represents the size of the total population in the reserve. Constraints 11.2.1 therefore require variable f_k, $k \in \underline{S}$, to be equal to the frequency of species s_k in the reserve, $\sum_{i\in\underline{Z}} n_{ik}x_i / \sum_{i\in\underline{Z}} \sum_{j\in\underline{S}} n_{ij}x_i$. Constraint 11.2.2 requires that the denominator of this ratio be strictly positive, *i.e.*, that the total population size of the reserve is at least equal to one unit. This makes it possible to prohibit the solution of $\mathrm{P}_{11.2}$, of value 1, consisting in not selecting any zone. Constraint 11.2.3 is the budget constraint. Constraints 11.2.4 specify the Boolean nature of variables x_i and constraints 11.2.5 specify that variables f_k belong to the interval [0, 1]. Note that the economic function is quadratic and concave and that the constraints are linear, except constraints 11.2.1 which include the products $f_k x_i$. One way to solve this program is to linearize constraints 11.2.1. The result is a mixed-integer program that consists of maximizing a concave function under linear constraints, which is equivalent to minimizing a convex function under linear constraints. The solution of this program can, therefore, be obtained efficiently using solvers based on classical implicit enumeration algorithms. Indeed, this type of algorithm requires, at each node of the search tree, the resolution of a continuous program which, in this case, is an "easy" problem since it consists in maximizing a concave function under linear constraints (see appendix at the end of the book). We will see that constraints 11.2.1 can be easily linearized.

Linearization of Constraints 11.2.1. Let us introduce the real variables u_{ik} belonging to the interval [0, 1]. A program equivalent to $\mathrm{P}_{11.2}$ is obtained by replacing the products $f_k x_i$ by variables u_{ik} and adding the set of linearization constraints $\mathrm{C}_{11.1}$ to force variable u_{ik} to be equal to the product of variables $f_k x_i$.

$$\mathrm{C}_{11.1}: \begin{cases} u_{ik} \le x_i & i \in \underline{Z}, k \in \underline{S} & | & u_{ik} \ge f_k - (1 - x_i) & i \in \underline{Z}, k \in \underline{S} \\ u_{ik} \le f_k & i \in \underline{Z}, k \in \underline{S} & | & u_{ik} \ge 0 & i \in \underline{Z}, k \in \underline{S} \end{cases}$$

We thus obtain the mixed-integer mathematical program $\mathrm{P}_{11.3}$ whose economic function to be maximized is quadratic and concave, and whose all constraints are linear. As we have said, many solvers allow a direct solution of this type of program.

$$\mathrm{P}_{11.3}: \begin{cases} \max \ 1 - \sum\limits_{k \in \underline{S}} f_k^2 \\ \text{s.t.} \left| \begin{array}{lll} \sum\limits_{i \in \underline{Z}} n_{ik} x_i = \sum\limits_{i \in \underline{Z}} \sum\limits_{j \in \underline{S}} n_{ij} u_{ik} & k \in \underline{S} & (11.3.1) \\ \sum\limits_{i \in \underline{Z}} \sum\limits_{k \in \underline{S}} n_{ik} x_i \ge 1 & & (11.3.2) \\ \sum\limits_{i \in \underline{Z}} c_i x_i \le B & & (11.3.3) \\ u_{ik} \le x_i & i \in \underline{Z}, k \in \underline{S} & (11.3.4) \\ u_{ik} \le f_k & i \in \underline{Z}, k \in \underline{S} & (11.3.5) \\ x_i \in \{0, 1\} & i \in \underline{Z} & (11.3.6) \\ u_{ik} \ge 0 & i \in \underline{Z}, k \in \underline{S} & (11.3.7) \\ 0 \le f_k \le 1 & k \in \underline{S} & (11.3.8) \end{array} \right. \end{cases}$$

Note that, because of the economic function to be maximized, the linearization constraints $u_{ik} \ge f_k - (1 - x_i)$ and $u_{ik} \ge 0$ are unnecessary.

It is easy to modify this program to require that the number of individuals protected by the reserve be greater than or equal to a given value, NI. To do this, simply replace the value 1 by NI in the right-hand side of constraint 11.3.2.

Below is another formulation of Problem I. Let us first note that maximizing the Simpson diversity index means minimizing the quantity $\sum_{k \in \underline{S}} f_k^2$, which can be written $\sum_{k \in \underline{S}} \left(\sum_{i \in \underline{Z}} n_{ik} x_i \Big/ \sum_{i \in \underline{Z}} \sum_{j \in \underline{S}} n_{ij} x_i \right)^2$ or $\sum_{k \in \underline{S}} \left(\sum_{i \in \underline{Z}} n_{ik} x_i \right)^2 \Big/ \left(\sum_{i \in \underline{Z}} \sum_{k \in \underline{S}} n_{ik} x_i \right)^2$. Problem I can thus be formulated as the fractional program in Boolean variables $\mathrm{P}_{11.4}$ (see appendix at the end of the book).

$$
\mathrm{P}_{11.4}:\left\{\begin{array}{ll}
\min & \sum_{k\in\underline{S}}\left(\sum_{i\in\underline{Z}} n_{ik}x_i\right)^2 \Big/ \left(\sum_{i\in\underline{Z}}\sum_{k\in\underline{S}} n_{ik}x_i\right)^2 \\
\text{s.t.} & \left|\begin{array}{lll}
\sum\limits_{i\in\underline{Z}}\sum\limits_{k\in\underline{S}} n_{ik}x_i \geq 1 & & (11.4.1)\\
\sum\limits_{i\in\underline{Z}} c_i x_i \leq B & & (11.4.2)\\
x_i \in \{0,1\} & i\in\underline{Z} & (11.4.3)
\end{array}\right.
\end{array}\right.
$$

The auxiliary program associated with $\mathrm{P}_{11.4}$ consists in minimizing the parameterized economic function $\sum_{k\in\underline{S}}\left(\sum_{i\in\underline{Z}} n_{ik}x_i\right)^2 - \lambda\left(\sum_{i\in\underline{Z}}\sum_{k\in\underline{S}} n_{ik}x_i\right)^2$ where λ is the parameter, under the same constraints as those of $\mathrm{P}_{11.4}$. Program $\mathrm{P}_{11.4}$ is slightly more difficult to implement than program $\mathrm{P}_{11.3}$ since a quadratic problem in variables 0–1 – the auxiliary program – has to be solved iteratively, but it is generally more efficient.

Mathematical programming formulation of Problem II. In this case, all the species must be protected, which is possible over a certain budget that we denote by B_{min}. We are therefore faced with two criteria: the diversity of the group of individuals protected by the selected reserve – measured by the Simpson diversity index – and the cost of this reserve. One way to approach Problem II is to solve $\mathrm{P}_{11.2}$, to which we add the constraints imposing the protection of all the species, $\sum_{i\in\underline{Z}_k} x_i \geq 1,\ k\in\underline{S}$, by gradually increasing the value of the available budget from the value B_{min}. The result is program $\mathrm{P}_{11.5}$. Note that constraint 11.2.2 becomes useless.

$$
\mathrm{P}_{11.5}:\left\{\begin{array}{ll}
\max & 1-\sum\limits_{k\in\underline{S}} f_k^2 \\
\text{s.t.} & \left|\begin{array}{lll}
\sum\limits_{i\in\underline{Z}} n_{ik}x_i = f_k\sum\limits_{i\in\underline{Z}}\sum\limits_{j\in\underline{S}} n_{ij}x_i & k\in\underline{S} & (11.5.1)\\
\sum\limits_{i\in\underline{Z}_k} x_i \geq 1 & k\in\underline{S} & (11.5.2)\\
\sum\limits_{i\in\underline{Z}} c_i x_i \leq B & & (11.5.3)\\
x_i \in \{0,1\} & i\in\underline{Z} & (11.5.4)
\end{array}\right.
\end{array}\right.
$$

Constraints 11.5.1 can be linearized as in program $\mathrm{P}_{11.3}$. The curve representing the maximal value of diversity that can be obtained given the available budget could be plotted. Note that Problem II can also be solved by the fractional mathematical program $\mathrm{P}_{11.4}$ in which constraint 11.4.1 is replaced by the constraints $\sum_{i\in\underline{Z}_k} x_i \geq 1,\ k\in\underline{S}$. Again, it is easy to modify $\mathrm{P}_{11.5}$ to require that the number of individuals protected by the reserve be greater than or equal to a given value, NI. To do this, simply add the constraint $\sum_{i\in\underline{Z}}\sum_{k\in\underline{S}} n_{ik}x_i \geq \mathrm{NI}$.

11.3.2 The Shannon–Wiener Index

The Shannon–Wiener index is commonly used. Like the Simpson index, it takes into account both species richness and relative abundance of each species. For the set of species S = $\{s_1, s_2, \ldots, s_m\}$ it is given by the following formula:

$$SH = -\sum_{k \in \underline{S}} f_k \log_2 f_k$$

where f_k denotes the frequency of species s_k in the population under consideration – number of individuals of species s_k divided by the total population size. It is supposed here that $f_k > 0$ for all $k \in \underline{S}$. This index allows diversity to be expressed by taking into account the number of species concerned and the relative abundance of individuals within each of these species. The value of the index varies from 0 – a single species – to $\log_2 m$ – all the species have the same abundance. Consider, for example, a population of 50 individuals distributed among 10 species and whose respective population sizes are as follows: 2, 3, 2, 4, 5, 20, 2, 3, 4, and 5. The Shannon–Wiener index of this population is equal to 2.8205. If each of the 10 species has 5 individuals, the Shannon–Wiener index would be equal to 3.3219. Note that the Shannon–Wiener index is sensitive to relatively rare species. Let us again take the above example of a population composed of 10 species whose respective population sizes are as follows: 2, 3, 2, 4, 5, 20, 2, 3, 4, and 5. When the 2 individuals of the first species disappear, the Shannon–Wiener index becomes equal to 2.6856 (–5%). As for the Simpson index, we can associate to this index the ratio SH/$\log_2 m$ where $\log_2 m$ represents the maximal value that the Shannon–Wiener index can take. This ratio, between 0 and 1, allows the distribution of individuals within species to be measured, independently of species richness. It reflects the degree of diversity achieved, in relation to the theoretical maximum. In reality, this ratio is commonly around 0.8 or 0.9. In the previous example of a population composed of 10 species, this ratio is equal to 2.8205/$\log_2 10$, *i.e.*, 2.8205/3.3219 or 0.8491. For the population composed of 9 species, it becomes equal to 0.8472. The problem of determining a reserve, R, that maximizes the index associated with the set of species protected by this reserve, DSH(R), can be formulated as program $\mathrm{P}_{11.6}$.

$$\mathrm{P}_{11.6}: \begin{cases} \min \sum_{k \in \underline{S}} f_k \log_2 f_k \\ \text{s.t.} \left| \begin{array}{lll} \sum_{i \in \underline{Z}} n_{ik} x_i = f_k \sum_{i \in \underline{Z}} \sum_{j \in \underline{S}} n_{ij} x_i & k \in \underline{S} & (11.6.1) \\ \sum_{i \in \underline{Z}} \sum_{k \in \underline{S}} n_{ik} x_i \geq 1 & & (11.6.2) \\ \sum_{i \in \underline{Z}} c_i x_i \leq B & & (11.6.3) \\ x_i \in \{0, 1\} & i \in \underline{Z} & (11.6.4) \\ 0 \leq f_k \leq 1 & k \in \underline{S} & (11.6.5) \end{array} \right. \end{cases}$$

Minimizing the economic function of $P_{11.6}$ is equivalent to maximizing the Shannon–Wiener Index. One way to solve $P_{11.6}$ is to approximate $\log_2 f_k$ by a piecewise linear function (see appendix at the end of the book) and then linearize the resulting program using for example the method presented in section 11.3.1 to linearize constraints 11.6.1. Here again, it is easy to modify program $P_{11.6}$ to require that the number of individuals protected by the reserve be greater than or equal to a given value, NI. To do this, simply replace constraint 11.6.2 with the constraint $\sum_{i \in \underline{Z}} \sum_{k \in \underline{S}} n_{ik} x_i \geq \text{NI}$.

11.3.3 Example with the Simpson Index

This example illustrates the selection of a reserve to protect all the species concerned while respecting a budget constraint and maximizing the Simpson diversity index. Consider the instance involving 20 candidate zones and 15 species, described in figure 11.3. Note that in the case of 15 species, the theoretical maximum of the

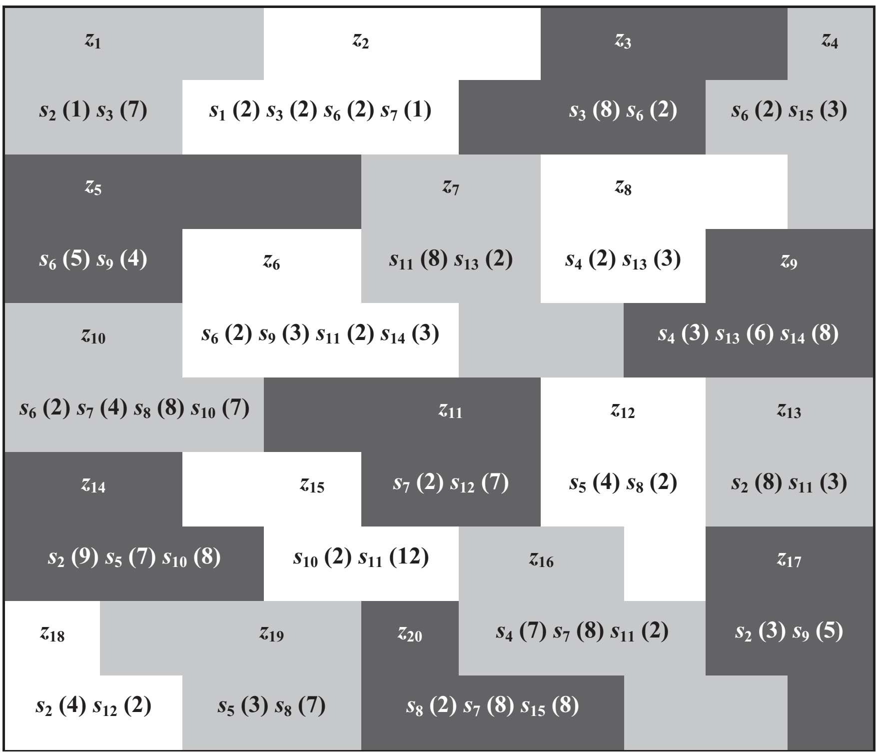


FIG. 11.3 – The 20 zones z_1, z_2,..., z_{20} are candidates for protection and the 15 species s_1, s_2,..., s_{15} are concerned by the protection of these zones. For each zone, the species concerned and the size of their population – in brackets – are indicated. The cost of protecting the white zones is 1 unit, the cost of protecting the light grey zones is 2 units and the cost of protecting the dark grey zones is 4 units. For example, species s_6, s_9, s_{11}, and s_{14} are concerned by the protection of zone z_6, their population size is equal to 2, 3, 2, and 3 units respectively, and the cost of protecting this zone is equal to 1 unit.

Simpson diversity index is equal to $1–15(1/15)^2 = 0.9333$. The problem considered can be formulated as the mathematical program $\mathrm{P}_{11.3}$ in which constraint 11.3.2 is replaced by the constraints $\sum_{i \in \underline{Z}_k} x_i \geq 1$, $k \in \underline{S}$. Note that since the number of protected species is fixed, the maximization of the evenness index associated with the Simpson diversity index is equivalent to the maximization of the Simpson diversity index. Table 11.4 presents the results obtained for different values of the available budget, B. Note that the data in this example are such that the protection of the 20 candidate zones allows for the protection of the 15 species considered. The associated cost is 48, the Simpson diversity index is 0.9214, and the corresponding evenness index is 0.9872. It should also be noted that there is no reserve to protect all the species with a cost less than 8.

11.4 Selecting a Reserve with Species Richness, Abundance, and Cost Constraints

One can search for a reserve with a great diversity without trying to describe this diversity by a single number as in section 11.3 above. One way of doing this is to maximize an economic function involving the – possibly weighted – criteria of species richness and abundance, while respecting constraints on the relative abundance of each species in the reserve and the cost of the reserve. This problem can be formulated as the mixed-integer program $\mathrm{P}_{11.7}$ where variable Ns represents the species richness.

$$
\mathrm{P}_{11.7}: \begin{cases}
\max \quad w_1 \mathrm{Ns} + w_2 \sum_{i \in \underline{Z}} \sum_{k \in \underline{S}} n_{ik} x_i \\
\text{s.t.} \left| \begin{array}{lll}
\mathrm{Ns} = \sum_{k \in \underline{S}} y_k & & (11.7.1) \\
\sum_{i \in \underline{Z}} n_{ik} x_i = f_k \sum_{i \in \underline{Z}} \sum_{j \in \underline{S}} n_{ij} x_i & k \in \underline{S} & (11.7.2) \\
(1/n) \sum_{i \in \underline{Z}_k} x_i \leq y_k \leq \sum_{i \in \underline{Z}_k} x_i & k \in \underline{S} & (11.7.3) \\
f_k \leq (1 + \varepsilon_k^+) \frac{1}{\mathrm{Ns}} + (1 - y_k) & k \in \underline{S} & (11.7.4) \\
f_k \geq (1 - \varepsilon_k^-) \frac{1}{\mathrm{Ns}} - (1 - y_k) & k \in \underline{S} & (11.7.5) \\
\sum_{i \in \underline{Z}} c_i x_i \leq B & & (11.7.6) \\
x_i \in \{0, 1\} & i \in \underline{Z} & (11.7.7) \\
y_k \in \{0, 1\} & k \in \underline{S} & (11.7.8) \\
0 \leq f_k \leq 1 & k \in \underline{S} & (11.7.9) \\
\mathrm{Ns} \in \mathbb{N} & & (11.7.10)
\end{array} \right.
\end{cases}
$$

TAB. 11.4 – Reserves protecting all the species, respecting a budgetary constraint, and with a maximal Simpson diversity index, when the candidate zones, their protection cost and the population size of the different species present in these zones are described in figure 11.3.

Available budget (B)	Used budget	Zones forming the reserve	Population size of each of the 15 species	Simpson diversity index	Evenness index
8	8	z_2 z_4 z_6 z_8 z_{12} z_{15} z_{18}	2 4 2 2 4 6 1 2 3 2 14 2 3 3 3 (53)	0.8843	0.9475
12	11	z_1 z_2 z_4 z_6 z_8 z_{10} z_{12} z_{18}	2 5 9 2 4 8 5 10 3 7 2 2 3 3 3 (68)	0.9109	0.9760
16	16	z_1 z_2 z_4 z_6 z_8 z_{10} z_{11} z_{12} z_{13}	2 9 9 2 4 8 7 10 3 7 5 7 3 3 3 (82)	0.9170	0.9825
20	20	z_1 z_2 z_4 z_6 z_8 z_9 z_{10} z_{11} z_{12} z_{13}	2 9 9 5 4 8 7 10 3 7 5 7 9 11 3 (99)	0.9222	0.9880
24	24	z_1 z_2 z_4 z_7 z_8 z_9 z_{10} z_{11} z_{12} z_{17} z_{18}	2 8 9 5 4 6 7 10 5 7 8 9 11 8 3 (102)	0.9243	0.9903
28	28	z_1 z_2 z_5 z_6 z_8 z_9 z_{11} z_{14} z_{15} z_{19} z_{20}	2 10 9 5 10 9 11 9 7 10 14 7 9 11 8 (131)	0.9270	0.9932
32	29	z_1 z_2 z_5 z_6 z_7 z_8 z_9 z_{11} z_{14} z_{19} z_{20}	2 10 9 5 10 9 11 9 7 8 10 7 11 11 8 (127)	0.9280	0.9943
36	29	z_1 z_2 z_5 z_6 z_7 z_8 z_9 z_{11} z_{14} z_{19} z_{20}	2 10 9 5 10 9 11 9 7 8 10 7 11 11 8 (127)	0.9280	0.9943
40	29	z_1 z_2 z_5 z_6 z_7 z_8 z_9 z_{11} z_{14} z_{19} z_{20}	2 10 9 5 10 9 11 9 7 8 10 7 11 11 8 (127)	0.9280	0.9943
44	29	z_1 z_2 z_5 z_6 z_7 z_8 z_9 z_{11} z_{14} z_{19} z_{20}	2 10 9 5 10 9 11 9 7 8 10 7 11 11 8 (127)	0.9280	0.9943
48	29	z_1 z_2 z_5 z_6 z_7 z_8 z_9 z_{11} z_{14} z_{19} z_{20}	2 10 9 5 10 9 11 9 7 8 10 7 11 11 8 (127)	0.9280	0.9943

Constraint 11.7.1 requires variable Ns to take the value corresponding to the number of protected species. The economic function therefore expresses the weighted sum of the number of species protected by the reserve and the total number of corresponding individuals. The weight assigned to these two quantities is equal to w_1 and w_2, respectively. Constraint 11.7.2 expresses the relative abundance of each species, f_k, $k \in \underline{S}$. Constraints 11.7.3 require the Boolean variable y_k, $k \in \underline{S}$, to be equal to 1 if and only if at least one of the zones hosting species s_k is protected. The coefficients ε_k^+ and ε_k^- are such that: $\varepsilon_k^+ \geq 0$ and $0 \leq \varepsilon_k^- \leq 1$. Constraints 11.7.4 and 11.7.5 express that the relative abundance of each species should not be too far from the ideal value, 1/Ns. Constraints 11.7.7 and 11.7.8 specify the Boolean nature of variables x_i and y_k. Constraint 11.7.9 specifies that variable f_k belongs to the interval [0, 1] and constraint 11.7.10 specifies that variable Ns is an integer variable. Program $\mathrm{P}_{11.7}$ consists of maximizing a linear function whose variables are subject to linear constraints (11.7.1, 11.7.3, and 11.7.6) and also to non-linear constraints (11.7.2, 11.7.4, and 11.7.5). We have already seen in section 11.3.1 how to linearize the quadratic constraints 11.7.2. Let us now examine constraints 11.7.4 and 11.7.5. Introduce variable Ns′ and constraint $\mathrm{Ns}' \times \sum_{j \in \underline{S}} y_j = 1$ that requires variable Ns′ to be equal to 1/Ns. Using variables v_j to represent the products $\mathrm{Ns}' \times y_j$, this last constraint can be written as $\sum_{j \in \underline{S}} v_j = 1$. It only remains to add the set of constraints $\mathrm{C}_{11.2}$ to require variable v_j, $j \in \underline{S}$, to be equal to the product $\mathrm{Ns}' \times y_j$.

$$\mathrm{C}_{11.2} : \left\{ v_j \leq y_j\,,\ v_j \leq \mathrm{Ns}',\ v_j \geq \mathrm{Ns}' - (1 - y_j),\ v_j \geq 0 \ (j \in \underline{S}) \right.$$

Finally, the problem can be formulated as the mixed-integer linear program $\mathrm{P}_{11.8}$.

$$\mathrm{P}_{11.8} : \begin{cases} \max \quad w_1 \mathrm{Ns} + w_2 \sum\limits_{k \in \underline{S}} \sum\limits_{i \in \underline{Z}} n_{ik} x_i \\ \text{s.t.} \left| \begin{array}{lll} (11.7.1), (11.7.3), (11.7.6), (11.7.7), & & \\ (11.7.8), (11.7.9), (11.7.10) & & \\ (\mathrm{C}_{11.1}), (\mathrm{C}_{11.2}) & & \\ \sum\limits_{i \in \underline{Z}} n_{ik} x_i = \sum\limits_{i \in \underline{Z}} \sum\limits_{j \in \underline{S}} n_{ij} u_{ik} & k \in \underline{S} & (11.8.1) \\ \sum\limits_{j \in \underline{S}} v_j = 1 & & (11.8.2) \\ f_k \leq (1 + \varepsilon_k^+) \mathrm{Ns}' + (1 - y_k) & k \in \underline{S} & (11.8.3) \\ f_k \geq (1 - \varepsilon_k^-) \mathrm{Ns}' - (1 - y_k) & k \in \underline{S} & (11.8.4) \end{array} \right. \end{cases}$$

By varying the weighting coefficients w_1 and w_2 as well as the coefficients ε_k^+ and ε_k^-, the resolution of program $\mathrm{P}_{11.7}$ can help a decision-maker to determine the best reserve taking into account the 3 criteria, species richness, abundance and relative abundance. Let us take again the example described in section 11.3.3 and examine, for different values of the available budget, B, the following 4 cases: $w_1 = 1$, $w_2 = 1$, $\varepsilon_k^+ = \varepsilon_k^- = 0.3$ $(k = 1,\ldots,15)$; $w_1 = 10$, $w_2 = 1$, $\varepsilon_k^+ = \varepsilon_k^- = 0.3$ $(k = 1,\ldots,15)$;

$w_1 = 1$, $w_2 = 1$, $\varepsilon_k^+ = \varepsilon_k^- = 0.5$ $(k = 1,\ldots, 15)$; and $w_1 = 10$, $w_2 = 1$, $\varepsilon_k^+ = \varepsilon_k^- = 0.5$ $(k = 1,\ldots, 15)$. The optimal reserves, *i.e.*, those that maximize the weighted sum of species richness and abundance taking into account constraints on relative abundance and budget, are determined by the resolution of programme $\mathrm{P}_{11.8}$ and are presented in tables 11.5 and 11.6. The last column of table 11.5 presents the maximal deviation, *i.e.*, the maximal gap – in absolute value – over all the protected species, between Ns' and the relative abundance of the species, all divided by Ns'. In other words, the maximal deviation is equal to $\max_{k \in \underline{S}\,:\ s_k\,\text{protected}} |(\mathrm{Ns}' - f_k)|/\mathrm{Ns}'$.

11.5 Reserve Minimizing the Average Kinship of Individuals of a Given Species Present in It

11.5.1 Kinship Between Two Individuals

A general and relevant problem in the field of biological conservation is to define, for a certain species, a sub-population of a given population, respecting certain constraints and having "good" genetic diversity. It is recognized that genetic diversity is essential for the survival of species. Several authors have demonstrated that a good measure of genetic diversity in a population is the overall kinship of that population. In 1948, Malécot defined the kinship coefficient between two individuals, I_k and I_l, as the probability that two randomly selected alleles, one on each individual and at any locus, are identical by descent. Two alleles are identical by descent when they are copies of a single allele of a common ancestor. If the pedigree of the population of interest is known, the kinship coefficients between any pair of individuals can be calculated according to simple rules (see below). The general problem considered is thus to extract from a given population a sub-population of minimum overall kinship. If we are interested in a population consisting of the individuals I_1, $I_2,\ldots, I_m$, and if α_{kl} is the kinship coefficient between the individuals I_k and I_l, then the average global kinship of this population is, by definition, equal to $(1/m^2) \sum_{k=1}^{m} \sum_{l=1}^{m} \alpha_{kl}$.

Calculation of kinship coefficients. Let us consider two disjoint generations and calculate the kinship coefficients of the individuals of the generation $g + 1$ from the kinship coefficients of the individuals of generation g. For each individual I_k of the generation $g + 1$ let us denote by I_{k_1} and I_{k_2} its two parents – belonging to the generation g. Recall that, for two individuals I_k and I_l of a generation, we denote by α_{kl} the kinship coefficient between these two individuals. For 2 individuals I_k and I_l of the generation $g + 1$, the kinship coefficient between these two individuals, α_{kl}, is equal to $(\alpha_{k_1 l_1} + \alpha_{k_1 l_2} + \alpha_{k_2 l_1} + \alpha_{k_2 l_2})/4$. The kinship coefficient of an individual I_k with itself, α_{kk}, is equal to $0.5\,(1 + \alpha_{k_1 k_2})$.

Example 11.1. Let us consider an initial – founder – population composed of 10 individuals, 4 males, m_1, m_2, m_3, and m_4, and 6 females, f_1, f_2, f_3, f_4, f_5, and f_6.

Tab. 11.5 – Optimal reserves associated with the data in figure 11.3. These reserves maximize the weighted sum of the species richness and abundance, taking into account constraints on relative abundance and budget, when $w_2 = 1$, $\varepsilon_k^+ = \varepsilon_k^- = 0.3$, $k = 1,\ldots, 15$.

Budget (B)	w_1	Economic function value	Used budget	Zones forming the reserve	Number of protected species	Number of protected individuals	Protected species and their population sizes	Max. deviation
10	1	62	9	z_1 z_{14} z_{15} z_{19}	6	56	$s_2(10)$ $s_3(7)$ $s_5(10)$ $s_8(7)$ $s_{10}(10)$ $s_{11}(12)$	0.2857
	10	116	9	z_1 z_{14} z_{15} z_{19}	6	56	$s_2(10)$ $s_3(7)$ $s_5(10)$ $s_8(7)$ $s_{10}(10)$ $s_{11}(12)$	0.2857
14	1	87	14	z_7 z_9 z_{14} z_{16} z_{19}	9	78	$s_2(9)$ $s_4(10)$ $s_5(10)$ $s_7(8)$ $s_8(7)$ $s_{10}(8)$ $s_{11}(10)$ $s_{13}(8)$ $s_{14}(8)$	0.1923
	10	168	14	z_7 z_9 z_{14} z_{16} z_{19}	9	78	$s_2(9)$ $s_4(10)$ $s_5(10)$ $s_7(8)$ $s_8(7)$ $s_{10}(8)$ $s_{11}(10)$ $s_{13}(8)$ $s_{14}(8)$	0.1923
18	1	98	16	z_5 z_6 z_7 z_8 z_9 z_{10} z_{16}	9	89	$s_4(12)$ $s_6(9)$ $s_7(12)$ $s_8(8)$ $s_9(7)$ $s_{10}(7)$ $s_{11}(12)$ $s_{13}(11)$ $s_{14}(11)$	0.2921
	10	187	18	z_7 z_9 z_{11} z_{14} z_{16} z_{19}	10	87	$s_2(9)$ $s_4(10)$ $s_5(10)$ $s_7(10)$ $s_8(7)$ $s_{10}(8)$ $s_{11}(10)$ $s_{12}(7)$ $s_{13}(8)$ $s_{14}(8)$	0.1954
22	1	118	20	z_8 z_9 z_{11} z_{12} z_{14} z_{15} z_{16} z_{18} z_{19}	10	108	$s_2(13)$ $s_4(12)$ $s_5(14)$ $s_7(10)$ $s_8(9)$ $s_{10}(10)$ $s_{11}(14)$ $s_{12}(9)$ $s_{13}(9)$ $s_{14}(8)$	0.2963
	10	216	22	z_1 z_5 z_6 z_8 z_9 z_{10} z_{13} z_{16} z_{17}	11	106	$s_2(12)$ $s_3(7)$ $s_4(12)$ $s_6(9)$ $s_7(12)$ $s_8(8)$ $s_9(12)$ $s_{10}(7)$ $s_{11}(7)$ $s_{13}(9)$ $s_{14}(11)$	0.2736

Tab. 11.6 – Optimal reserves associated with the data in figure 11.3. These reserves maximize the weighted sum of the species richness and abundance, taking into account constraints on relative abundance and budget, when $w_2 = 1$, $\varepsilon_k^+ = \varepsilon_k^- = 0.5$, $k = 1,\ldots, 15$.

Budget (B)	w_1	Economic function value	Used budget	Zones forming the reserve	Number of protected species	Number of protected individuals	Protected species and their population sizes	Max. deviation
10	1	78	10	z_{12} z_{14} z_{15} z_{16} z_{19}	7	71	$s_2(9)$ $s_4(7)$ $s_5(14)$ $s_7(8)$ $s_8(9)$ $s_{10}(10)$ $s_{11}(14)$	0.3803
	10	154	10	z_4 z_9 z_{10} z_{13}	10	54	$s_2(8)$ $s_4(3)$ $s_6(4)$ $s_7(4)$ $s_8(8)$ $s_{10}(7)$ $s_{11}(3)$ $s_{13}(6)$ $s_{14}(8)$ $s_{15}(3)$	0.4815
14	1	97	14	z_9 z_{12} z_{14} z_{15} z_{16} z_{19}	9	88	$s_2(9)$ $s_4(10)$ $s_5(14)$ $s_7(8)$ $s_8(9)$ $s_{10}(10)$ $s_{11}(14)$ $s_{13}(6)$ $s_{14}(8)$	0.4318
	10	192	13	z_1 z_4 z_6 z_8 z_{12} z_{13} z_{16} z_{19}	12	72	$s_2(9)$ $s_3(7)$ $s_4(9)$ $s_5(7)$ $s_6(4)$ $s_7(8)$ $s_8(9)$ $s_9(3)$ $s_{11}(7)$ $s_{13}(3)$ $s_{14}(3)$ $s_{15}(3)$	0.5000
18	1	117	18	z_1 z_{12} z_{13} z_{14} z_{15} z_{16} z_{19} z_{20}	9	108	$s_2(18)$ $s_3(7)$ $s_4(7)$ $s_5(14)$ $s_7(16)$ $s_8(11)$ $s_{10}(10)$ $s_{11}(17)$ $s_{15}(8)$	0.5000
	10	208	18	z_1 z_5 z_6 z_8 z_9 z_{10} z_{13} z_{16}	11	98	$s_2(9)$ $s_3(7)$ $s_4(12)$ $s_6(9)$ $s_7(12)$ $s_8(8)$ $s_9(7)$ $s_{10}(7)$ $s_{11}(7)$ $s_{13}(9)$ $s_{14}(11)$	0.3469
22	1	144	22	z_5 z_6 z_8 z_9 z_{10} z_{12} z_{14} z_{15} z_{16} z_{19}	11	133	$s_2(9)$ $s_4(12)$ $s_5(14)$ $s_6(9)$ $s_7(12)$ $s_8(17)$ $s_9(7)$ $s_{10}(17)$ $s_{11}(16)$ $s_{13}(9)$ $s_{14}(11)$	0.4211
	10	243	22	z_5 z_6 z_8 z_9 z_{10} z_{12} z_{14} z_{15} z_{16} z_{19}	11	133	$s_2(9)$ $s_4(12)$ $s_5(14)$ $s_6(9)$ $s_7(12)$ $s_8(17)$ $s_9(7)$ $s_{10}(17)$ $s_{11}(16)$ $s_{13}(9)$ $s_{14}(11)$	0.4211

	m_1	m_2	m_3	m_4	f_1	f_2	f_3	f_4	f_5	f_6
m_1	0.5	0	0	0	0	0	0	0	0	0
m_2	0	0.5	0	0	0	0	0	0	0	0
m_3	0	0	0.5	0	0	0	0	0	0	0
m_4	0	0	0	0.5	0	0	0	0	0	0
f_1	0	0	0	0	0.5	0	0	0	0	0
f_2	0	0	0	0	0	0.5	0	0	0	0
f_3	0	0	0	0	0	0	0.5	0	0	0
f_4	0	0	0	0	0	0	0	0.5	0	0
f_5	0	0	0	0	0	0	0	0	0.5	0
f_6	0	0	0	0	0	0	0	0	0	0.5

FIG. 11.4 – Kinship coefficient matrix of the founder population.

TAB. 11.7 – Matings in the initial population and associated offspring.

Mating	Offspring size	Generated individuals
(m_1, f_1)	2	I_1 I_2
(m_1, f_2)	1	I_3
(m_2, f_1)	1	I_4
(m_3, f_3)	1	I_5
(m_3, f_4)	2	I_6 I_7
(m_4, f_5)	1	I_8
(m_4, f_6)	2	I_9 I_{10}

The kinship coefficients of this population are given by the matrix in figure 11.4. The overall kinship of this population is, by definition, equal to $(10 \times 0.5)/100 = 0.05$. Let us now consider a hypothetical population of 10 individuals generated, from the 10 individuals in the initial population, by the matings shown in table 11.7.

The kinship coefficients, α_{kl}, in the generated population are given by the matrix in figure 11.5. The mean kinship coefficient of the generated population is equal to 0.085.

11.5.2 The Problem and its Mathematical Programming Formulation

More specifically, we are concerned here with a set of zones, $Z = \{z_1, z_2, \ldots, z_n\}$, which are likely to be protected and a single species, s, living in these zones and

	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}
I_1	0.5	0.25	0.125	0.125	0	0	0	0	0	0
I_2	0.25	0.5	0.125	0.125	0	0	0	0	0	0
I_3	0.125	0.125	0.5	0	0	0	0	0	0	0
I_4	0.125	0.125	0	0.5	0	0	0	0	0	0
I_5	0	0	0	0	0.5	0.125	0.125	0	0	0
I_6	0	0	0	0	0.125	0.5	0.25	0	0	0
I_7	0	0	0	0	0.125	0.25	0.5	0	0	0
I_8	0	0	0	0	0	0	0	0.5	0.125	0.125
I_9	0	0	0	0	0	0	0	0.125	0.5	0.25
I_{10}	0	0	0	0	0	0	0	0.125	0.25	0.5

FIG. 11.5 – Matrix of the kinship coefficients of the population composed of 10 individuals generated from the founder population as shown in table 11.7.

considered critically endangered. The individuals of this species, I_1, I_2,…, I_m, are distributed over the different zones. The set of these individuals is designated by I and the set of corresponding indices, by $\underline{I}$. The protection of a zone makes it possible to protect all the individuals present on this zone. For $i = 1, 2,\ldots, n$ and $k = 1,\ldots, m$, the presence of the individual I_k on zone z_i is defined by the coefficient a_{ki}. This coefficient is equal to 1 if and only if the individual I_k is present on zone z_i, and to 0 otherwise. We know the kinship coefficient, α_{kl}, associated with each pair of individuals $\{I_k, I_l\}$, including the coefficient α_{kk} for $k = 1, 2,\ldots, m$. The problem considered is to determine the best set of zones to be protected – a reserve – taking into account a budget constraint and the number of individuals one wishes to protect. The value of a reserve, R, is the average kinship coefficient of the population of individuals protected by R, and this coefficient should be minimized. Let us associate to each zone z_i the Boolean variable x_i which is equal to 1 if and only if zone z_i is protected. The individual I_k is protected if and only if the zone where it is present is protected, *i.e.*, if and only if variable x_i such that $a_{ki} = 1$ is equal to 1. Let us associate to each individual I_k a Boolean variable, y_k, which is equal to 1 if and only if the individual I_k is protected – because of the zone protections. By noting n_i the number of individuals present in zone z_i, the number of protected individuals is equal to $\sum_{i\in\underline{Z}} n_i x_i$ where $\underline{Z}$ denotes the set of indices of the elements of Z. This number must be equal to NI and the mean kinship coefficient is equal to $\sum_{(k,l)\in\underline{I}^2} \alpha_{kl} y_k y_l/\mathrm{NI}^2$. The problem considered can thus be formulated as the mathematical program $\mathrm{P}_{11.9}$.

$$\mathrm{P}_{11.9}: \begin{cases} \min \ \frac{1}{\mathrm{NI}^2} \sum_{(k,l) \in \underline{I}^2} \alpha_{kl} y_k y_l \\ \text{s.t.} \left| \begin{array}{llll} \sum_{i \in \underline{Z}} c_i x_i \leq B & (11.9.1) & | & y_k \in \{0,1\} \quad k \in \underline{I} \quad (11.9.4) \\ \sum_{k \in \underline{I}} y_k = \mathrm{NI} & (11.9.2) & | & x_i \in \{0,1\} \quad i \in \underline{Z} \quad (11.9.5) \\ y_k = \sum_{i \in \underline{Z}} a_{ki} x_i \quad k \in \underline{I} & (11.9.3) & | & \end{array} \right. \end{cases}$$

Note that the matrix of the kinship coefficients $(\alpha_{kl})_{(k,l) \in \underline{I}^2}$ is symmetrical and positive semidefinite, resulting in the convexity of the economic function of $\mathrm{P}_{11.9}$. This program, which thus consists in minimizing a convex quadratic function subject to linear constraints, can be directly submitted to a solver that can handle this type of program – whose continuous relaxation is a quadratic and convex program. It can also be linearized (see appendix at the end of the book). Suppose that, among the individuals making up the population under consideration, the individuals $I_k,\ k \in K_1 \subseteq \underline{I}$, are males and the individuals $I_k,\ k \in K_2 \subseteq \underline{I}$, are females (the sets K_1 and K_2 form a partition of $\underline{I}$). Program $\mathrm{P}_{11.9}$ could easily be modified to require, for example, that at least $\rho_1\%$ of the protected individuals be males and at least $\rho_2\%$ of the protected individuals be females. It is sufficient to add to $\mathrm{P}_{11.9}$ the constraint set $\mathrm{C}_{11.3}$.

$$\mathrm{C}_{11.3}: \begin{cases} \sum_{k \in K_1} y_k \geq \rho_1 \mathrm{NI}/100 \\ \sum_{k \in K_2} y_k \geq \rho_2 \mathrm{NI}/100 \end{cases}$$

11.5.3 Example

Let us take again the population resulting from the founder population described in figure 11.4 and composed of 10 individuals, generated according to the information in table 11.7 and whose kinship coefficients are given in figure 11.5. Suppose that in this generated population, the first 5 individuals – indexed from 1 to 5 – are males, designated by M_1, M_2, M_3, M_4, and M_5, and the next 5 – indexed from 6 to 10 – are females, designated by F_1, F_2, F_3, F_4, and F_5. Consider a new population of 10 individuals resulting from the matings described in table 11.8. These 10 individuals are again denoted I_1, I_2,…, I_{10}.

The kinship coefficient matrix of the generated population is given in figure 11.6. The average kinship of this population is equal to 0.10875. Let us now consider a set of 7 zones that can be protected to form a reserve. Figure 11.7 shows, for each of these zones, the individuals of the population I_1, I_2,…, I_{10}, described in figure 11.6, present in these zones. Table 11.9 presents the optimal reserves, *i.e.*, those that minimize the average kinship of the individuals in these reserves, for different values of the available budget, B, and the number of individuals to be protected, NI. These

TAB. 11.8 – Matings in the population described in figure 11.5 and associated offspring.

Mating	Number of offspring	Generated individuals
(M_1, F_1)	1	I_1
(M_2, F_4)	1	I_2
(M_3, F_3)	2	I_3 I_4
(M_3, F_4)	1	I_5
(M_4, F_1)	1	I_6
(M_4, F_2)	2	I_7 I_8
(M_5, F_5)	2	I_9 I_{10}

reserves are obtained by resolving $\text{P}_{11.9}$. Suppose that, among the ten individuals making up the population under consideration, *i.e.*, the population described in figure 11.6, the individuals I_k, $k = 1,\ldots, 4$, are males and the individuals I_k, $k = 5,\ldots, 10$, are females. The following constraint is now imposed: at least 30% of the protected individuals must be males and at least 50% of the protected individuals must be females. The results obtained with these additional constraints are presented in table 11.10.

A variant of this problem is to select a reserve that minimizes the average global kinship of the protected population, but without constraints on the number of individuals to be protected. As before, the selected reserve must respect the available budget and the proportion of males and females. One way to solve this problem is to solve a series of programs $\text{P}_{11.9}$ – with the addition of the constraint set $\text{C}_{11.3}$ – by giving the parameter NI all the possible values and then to choose the best solution – the one that minimizes the average global kinship. This problem can be solved more quickly by program $\text{P}_{11.9}$ – with the addition of the constraint set $\text{C}_{11.3}$ – in which NI is no longer a fixed number but becomes an integer variable. We then

	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}
I_1	0.5	0.0625	0.03125	0.03125	0.03125	0.15625	0.09375	0.09375	0.03125	0.03125
I_2	0.0625	0.5	0.0625	0.0625	0.15625	0.03125	0.03125	0.03125	0.0625	0.0625
I_3	0.03125	0.625	0.5	0.25	0.15625	0	0	0	0.03125	0.03125
I_4	0.03125	0.625	0.25	0.5	0.15625	0	0	0	0.03125	0.03125
I_5	0.03125	0.15625	0.15625	0.15625	0.5	0	0	0	0.0625	0.0625
I_6	0.15625	0.03125	0	0	0	0.5	0.1875	0.1875	0.03125	0.03125
I_7	0.09375	0.03125	0	0	0	0.1875	0.5	0.25	0.03125	0.03125
I_8	0.09375	0.03125	0	0	0	0.1875	0.25	0.5	0.03125	0.03125
I_9	0.03125	0.0625	0.03125	0.03125	0.0625	0.03125	0.03125	0.03125	0.5	0.25
I_{10}	0.03125	0.0625	0.03125	0.03125	0.0625	0.03125	0.03125	0.03125	0.25	0.5

FIG. 11.6 – Kinship coefficients of the ten individuals generated by the population described in figure 11.5, from the matings described in table 11.8.

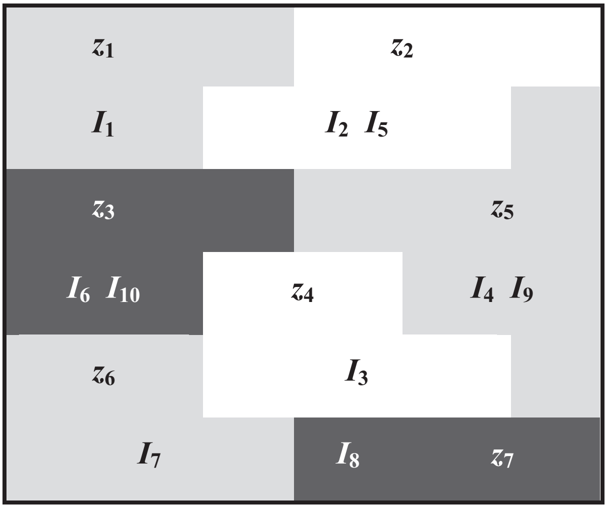


FIG. 11.7 – Seven zones, z_1, z_2,…, z_7, are candidates for protection and ten individuals of the same species, I_1, I_2,…, I_{10}, living in these zones are concerned. For each zone, the individuals present are indicated. The cost of protecting the white zones is equal to 1 unit, the cost of protecting the light grey zones is equal to 2 units and the cost of protecting the dark grey zones is equal to 4 units. For example, the individuals I_6 and I_{10} are present in zone z_3, and the cost of protecting this zone is 4 units.

obtain a fractional program for which the auxiliary problem consists in minimizing the parameterized quadratic function in integer variables $\sum_{(k,l)\in \underline{I}^2} \alpha_{kl} y_k y_l - \lambda\,\mathrm{NI}^2$, subject to constraints 11.9.1–11.9.5 and $C_{11.3}$. The results obtained using the Dinkelbach algorithm (see appendix at the end of the book) are presented in table 11.11.

TAB. 11.9 – Optimal reserves, under a budgetary constraint, associated with figures 11.6 and 11.7: Minimizing average kinship, taking into account the number of individuals to be protected and the available budget.

Available budget (B)	Number of individuals to be protected (NI)	Optimal reserve	Cost of the reserve	Protected individuals	Average kinship
4	5	z_2 z_4 z_5	4	I_2 I_3 I_4 I_5 I_9	0.2275
5	5	z_2 z_5 z_6	5	I_2 I_4 I_5 I_7 I_9	0.1700
6	5	z_2 z_5 z_6	5	I_2 I_4 I_5 I_7 I_9	0.1700
7	5	z_1 z_4 z_5 z_6	7	I_1 I_3 I_4 I_7 I_9	0.1425
8	5	z_1 z_4 z_5 z_6	7	I_1 I_3 I_4 I_7 I_9	0.1425
9	5	z_1 z_4 z_5 z_6	7	I_1 I_3 I_4 I_7 I_9	0.1425

TAB. 11.10 – Optimal reserves, under a budgetary constraint, associated with figures 11.6 and 11.7: Minimization of average kinship, taking into account the number of individuals to be protected, the available budget, and constraints on the number of males and females to be protected.

Available budget (B)	Number of individuals to be protected (NI)	Optimal reserve	Cost of the reserve	Protected individuals	Average kinship
5	5	z_2 z_5 z_6	5	I_2 I_4 I_5 I_7 I_9	0.1700
6	5	z_2 z_5 z_6	5	I_2 I_4 I_5 I_7 I_9	0.1700
7	7	–	–	–	–
8	7	z_2 z_3 z_4 z_5	8	I_2 I_3 I_4 I_5 I_6 I_9 I_{10}	0.1582
9	5	z_1 z_3 z_4 z_6	9	I_1 I_3 I_6 I_7 I_{10}	0.1475

– No feasible solution.

TAB. 11.11 – Optimal reserves, under a budgetary constraint, associated with figures 11.6 and 11.7: Minimization of average kinship, taking into account the proportion of males and females to be protected and the available budget.

Available budget (B)	Optimal reserve	Cost of the reserve	Protected individuals	Average kinship
5	z_4 z_5 z_6	5	I_3 I_4 I_7 I_9	0.1680
6	z_1 z_5 z_6	6	I_1 I_4 I_7 I_9	0.1523
7	z_1 z_2 z_5 z_6	7	I_1 I_2 I_4 I_5 I_7 I_9	0.1458
8	z_2 z_3 z_4 z_6	8	I_2 I_3 I_5 I_6 I_7 I_{10}	0.1458
9	z_3 z_4 z_5 z_6	9	I_3 I_4 I_6 I_7 I_9 I_{10}	0.1354

References and Further Reading

Ablquist J.E., Bledsoe A.H., Sheldon F.H., Sibley C.G. (1987) DNA hybridization and avian systematics, *Auk* **104**, 556.

Allen S.D. (2009) *An integer programming approach to selecting individuals for transfer in pedigreed populations*. North Carolina State University Thesis.

Allen S.D., Fathi Y., Gross K., Mace M. (2010) An optimal and near-optimal strategy to selecting individuals for transfer in captive breeding programs, *Biol. Conserv.* **143**, 2858.

Billionnet A. (2007) *Optimisation discrète, de la modélisation à la résolution par des logiciels de programmation mathématique*. Dunod, Paris.

Billionnet A. (2013) Mathematical optimization ideas for biodiversity conservation, *Eur. J. Oper. Res.* **231**, 514.

Boyce A.J. (1983) Computation of inbreeding and kinship coefficients on extended pedigrees, *J. Heredity* **74**, 400.

Dinkelbach W. (1967) On nonlinear fractional programming, *Manage. Sci.* **13**, 492.

Glover F., Ching-Chung K., Dhir K.S. (1995) A discrete optimization model for preserving biological diversity, *Appl. Math. Model.* **19**, 696.

Hill M.O. (1973) Diversity and evenness: a unifying notation and its consequences, *Ecol.* **54**, 427.

IUCN, The IUCN red list of threatened species, Version 2017–3.

Jacquard A. (1966) Logique du calcul des coefficients d'identité entre deux individus, *Population* **21**, 751.

Keylock C.J. (2005) Simpson diversity and the Shannon-Wiener index as special cases of a generalized entropy, *Oikos* **109**, 203.

Krajewski C. (1989) Phylogenetic relationships among cranes (Gruiformes: Gruidae) based on DNA hybridization, *The Auk* **106**, 603.

Magurran A.E. (1988) *Ecological diversity and its measurement.* Springer-Science+Business Media, B.Y.

Malécot G. (1948) *Les mathématiques de l'hérédité.* Masson, Paris.

Peet R.K. (1974) The measurement of species diversity, *Annu. Rev. Ecol. Syst.* **5**, 285.

Piélou E.C. (1966) The measurement of diversity in different types of biological collections, *J. Theor. Biol.* **13**, 131.

Piélou E.C. (1969) *An introduction to mathematical ecology.* Wiley-Interscience, New York.

Routledge R.D. (1979) Diversity indices: which ones are admissible? *J. Theor. Biol.* **76**, 503.

Schaible S. (1995) Fractional programming, *Handbook of global optimization* (R. Horst, P. Pardalos, Eds). Kluwer Academic Publishers, pp. 495–608.

Shimatani K. (2001) On the measurement of species diversity incorporating species differences, *Oikos* **93**, 135.

Simpson E.H. (1949) Measurement of species diversity, *Nat.* **163**, 688.

Solow A., Polasky S., Broadus J. (1993) On the measurement of biological diversity, *J. Environ. Econ. Manage.* **24**, 60.

Chapter 12

Climate Change

12.1 Introduction

It is widely accepted that human activities are causing an increase in the concentrations of greenhouse gases in the Earth's atmosphere and thus causing an increase in its average temperature. This warming, even modest, by modifying the behaviour of the air masses in the atmosphere, leads to climate change, *i.e.*, changes in average values, measured over long periods and over specific geographical zones, concerning, for example, temperature, precipitation and winds. It is also recognized that climate change is a major threat to biodiversity and that the system of protected zones is a very effective solution to combat this threat. However, climate change is creating major challenges in the design and management of protected zones. By designing reserves without accounting for the effects of climate change, the biodiversity, currently protected by these reserves, may no longer be protected in the near or more distant future. In the previous chapters, the definition of protected zones is largely based on current observations. In these chapters, the general idea behind decisions to protect, or not to protect, a zone is that certain species live in a given zone that is a priori a favourable habitat for them and that protection of that zone, therefore, contributes to the protection of those species. There is no anticipation in this reasoning of possible changes in the quality of the habitat that this zone provides for the species considered. It is therefore quite possible that these species, by not being able to adapt to the effects of climate change in this zone, will disappear completely from this zone. However, some anticipation is present in the previous chapters when assigning survival probabilities to the species or when considering that different scenarios may occur in the future (chapters 7, 8, 9, and 10). It has been observed that the ranges of some species are shifting significantly towards the poles, mountain tops or ocean depths, probably in response to increases in temperature. Some species, such as the pine processionary caterpillar, are able to move quickly to maintain zones of habitat that are favourable to them. Other species, such as trees, are much slower in these movements. It was also found that other species appeared to be unable to adapt to change, in part because of the rate of change. For these species,

DOI: 10.1051/978-2-7598-2521-9.c012

natural selection will take place. Climate change may also result in the proliferation of certain species and thus a profound change in the quality of the habitat of other species. In summary, climate change may cause many species present in a reserve to lose much of the habitat that is currently favourable to them in that reserve. Some of this habitat may disappear or be moved outside the reserve. The definition and also the management of protected areas must therefore take into account the impacts of climate change using, for example, bioclimatic models, even though there is great uncertainty about the effects of these impacts, their significance and when they will occur. The reserves defined must have the capacity to be as resilient as possible to climate change. Note that the properties of contiguity – or connectivity – and compactness of reserves (chapters 3 and 4) and the concepts of fragmentation and biological corridors (chapters 2 and 6) are particularly important in the context of climate change.

Protected zones and the way they are managed also contribute to slowing climate change, in particular by capturing and storing carbon in natural ecosystems. Thus, increasing the size of protected zones and possibly changing their management to sequester more carbon are important actions to combat climate change. Protected zones can, for example, limit the loss of forests, which is considered an important cause of climate change since forests contain the largest terrestrial carbon stock (forests themselves are directly threatened by climate change). Grasslands also contain large reserves of carbon. This aspect should be increasingly taken into account in the choice of zones to be protected. The protection of certain zones may be more effective than the construction of specific infrastructure to combat natural disasters such as floods and storms. Of course, protected zones are not a complete solution and cannot replace efforts to reduce emissions at source. In conclusion, predictions of the effects of climate change are, therefore, becoming important factors to be taken into account in the selection and management of protected zones.

12.2 Three Fundamental Problems of Reserve Selection, Under a Budgetary Constraint, Without Taking Climate Change into Account

All of the issues discussed in the previous chapters can be revisited with climate change in mind. To illustrate this approach, we consider three basic problems, two of which have already been discussed in previous chapters (sections 1.3.1 and 1.3.2 of chapter 1). The issues raised by the consideration of climate change and the approach taken in this chapter would easily extend to other problems concerning the optimal design and management of networks of protected zones. We briefly present these three problems in which climate change is not taken into account. $S = \{s_1, s_2, \ldots, s_m\}$ is the set of species, animal or plant, that we are interested in, $Z = \{z_1, z_2, \ldots, z_n\}$ is the set of zones that we may decide to protect or not, and only one level of protection is possible. The set of protected zones is called a reserve. As already mentioned, to facilitate the presentation we are interested here in a set of species, but the approaches presented here could just as well apply to other aspects of biodiversity. The set of indices for the species in S is denoted by $\underline{S}$ and the set of indices for the zones in Z is denoted by $\underline{Z}$. B is the available budget and c_i, $i \in \underline{Z}$, is the cost of protecting zone z_i. The cost of protecting a set of zones, $R \subseteq Z$, is equal to the sum of the costs of protecting each of the zones in that set. The Boolean variables x_i, $i \in \underline{Z}$, and y_k, $k \in \underline{S}$, are used to formulate these problems as mathematical programs. By definition, $x_i = 1$ if and only if zone z_i is selected for protection – for forming the reserve – and $y_k = 1$ if and only if species s_k is protected by the reserve.

12.2.1 Problem I: Choice of a Reserve Protecting the Greatest Possible Number of Species – of a Given Set – Knowing that the Protection of Each Zone Makes it Possible to Protect a Certain Set of Species

This problem, already discussed in section 1.3.1 of chapter 1, is to determine a set of zones to be protected – a reserve – within an available budget, so as to protect as many species as possible. Here, it is considered that a reserve, R, protects species s_k, from a certain instant, if and only if this species is protected by at least one of the zones of R. For each of species s_k, the list of candidate zones whose protection leads to the protection of s_k is known. We denote by Z_k the set of these zones and by $\underline{Z}_k$ the set of corresponding indices. This problem can be formulated as program $\mathrm{P}_{12.1}$.

$$\mathrm{P}_{12.1}: \begin{cases} \max \displaystyle\sum_{k \in \underline{S}} y_k \\ \text{s.t.} \left| \begin{array}{llll} y_k \le \displaystyle\sum_{i \in \underline{Z}_k} x_i \quad k \in \underline{S} & (12.1.1) & \mid & x_i \in \{0,1\} \quad i \in \underline{Z} \quad (12.1.3) \\ \displaystyle\sum_{i \in \underline{Z}} c_i x_i \le B & (12.1.2) & \mid & y_k \in \{0,1\} \quad k \in \underline{S} \quad (12.1.4) \end{array} \right. \end{cases}$$

12.2.2 Problem II: Selection of a Reserve Protecting as many Species – of a Given Set – as Possible, Knowing that a Species is Protected if its Total Population Size in the Reserve Exceeds a Certain Value

This problem, already discussed in section 1.3.2 of chapter 1 consists of determining a set of zones to be protected, taking into account the available budget, so as to protect the greatest possible number of species. Here, it is considered that a reserve, R, protects species s_k, $k \in \underline{S}$, if and only if the total population size of that species in the reserve is greater than or equal to a threshold value, θ_k. The population size of each species in each of the candidate zones is known and denoted by n_{ik}, $i \in \underline{Z}$, $k \in \underline{S}$. This problem can be formulated as program $\mathrm{P}_{12.2}$.

$$
\mathrm{P}_{12.2}:\left\{\begin{array}{ll}
\max & \sum_{k\in\underline{S}} y_k \\
\text{s.t.} & \left|\begin{array}{llll}
\theta_k y_k \le \sum_{i\in\underline{Z}} n_{ik}x_i & k\in\underline{S} & (12.2.1) & \mid \quad x_i\in\{0,1\} \quad i\in\underline{Z} \quad (12.2.3) \\
\sum_{i\in\underline{Z}} c_i x_i \le B & & (12.2.2) & \mid \quad y_k\in\{0,1\} \quad k\in\underline{S} \quad (12.2.4)
\end{array}\right.
\end{array}\right.
$$

12.2.3 Problem III: Selection of a Reserve that Provides Each Species – of a Given Set – with a Favourable Habitat Area as Close as Possible to a Target Value

This problem consists in determining a set of zones to be protected – a reserve – taking into account an available budget, B, so as to ensure for each of the species under consideration a total area of favourable habitat, included in the reserve, as close as possible to a target value. For each species s_k, the area of habitat in zone z_i that is favourable to it is known; it is denoted by a_{ik}, $i \in \underline{Z}$, $k \in \underline{S}$. The target value for species s_k is denoted by $\min_k$, $k \in \underline{S}$. This problem can be formulated as program $\mathrm{P}_{12.3}$, which uses, in addition to variables x_i, the positive or null variables g_k that express the gap between the total area of habitat in the reserve favourable to species s_k and the target value for this species, $\min_k$, $k \in \underline{S}$. This gap is only taken into account if the total area of habitat in the reserve favourable to species s_k is less than $\min_k$.

$$
\mathrm{P}_{12.3}:\left\{\begin{array}{ll}
\min & \sum_{k\in\underline{S}} \xi_k g_k \\
\text{s.t.} & \left|\begin{array}{llll}
\sum_{i\in\underline{Z}} a_{ik}x_i + g_k \ge \min_k & k\in\underline{S} & (12.3.1) & \mid \quad x_i\in\{0,1\} \quad i\in\underline{Z} \quad (12.3.3) \\
\sum_{i\in\underline{Z}} c_i x_i \le B & & (12.3.2) & \mid \quad g_k \ge 0 \quad k\in\underline{S} \quad (12.3.4)
\end{array}\right.
\end{array}\right.
$$

A “goal programming” approach is used here, which aims to achieve, taking into account a budgetary constraint, some objectives of protection for each species with a penalty when these objectives are not achieved. For each species s_k, this penalty is equal to the product of a positive number, ξ_k, by the difference between the total area of habitat in the reserve favourable to species s_k and the target value, $\min_k$. This approach may be interesting compared to imposing strict targets because, for various reasons, these targets may be unachievable. Note that the economic function of $\mathrm{P}_{12.3}$ expresses the sum of the penalties associated with each species.

12.3 Taking into Account a Certain and Known Climate Evolution in Problems I, II and III

Let us revisit the previous problems in the light of climate change predictions. In this section, we make the – strong – assumption that there is no uncertainty regarding these predictions. The management horizon, T, consists of r periods $T_1,\dots, T_r$ and all the decisions regarding the zones to be protected are made at the beginning of the management horizon. We set $\underline{T} = \{1,\dots, r\}$.

12.3.1 Problem I

With regard to the extension of Problem I, it is assumed that certain zones constitute a favourable habitat for s_k, $k \in \underline{S}$, at certain periods but that this is no longer the case at later periods, even if these zones are protected, because of climate change. Conversely, some zones, at certain periods, do not constitute a favourable habitat for certain species but these zones become a favourable habitat in later periods. We thus assume, in a very general way, that we know Z_{kt}, the set of zones of Z which, if they are protected at the beginning of the horizon considered, constitute a favourable habitat for species s_k during the period T_t. $\underline{Z}_{kt}$ designates the set of indices of these zones. The definition of the sets Z_{kt} requires important prospective studies. The problem that then arises is to determine a set of zones to be protected, at a cost less than or equal to a given value, and optimal with regard to the conservation of the species under consideration. The tricky question is: what is an optimal reserve? The aim here is to determine a reserve that maximizes, within an available budget, the number of species for which, at each period of the time horizon under consideration, at least one zone of the reserve constitutes a habitat that is favourable to them. It can be assumed, for example, that the species living at a certain time in a zone favourable to them will move over time to other zones if that initial zone is no longer favourable to them. In a first step, we do not consider precisely these movement problems, but we could look at connected reserves (chapter 3) or reserves whose different units are linked by biological corridors (chapter 6). Indeed, the connectivity properties of reserves can significantly help certain species to adapt to climate change.

To formulate this problem by mathematical programming, we use the Boolean variable x_i, $i \in \underline{Z}$, which takes the value 1 if and only if zone z_i is protected at the beginning of the horizon considered, and therefore throughout this horizon, and the Boolean variable y_k, $k \in \underline{S}$, which takes the value 1 if and only if species s_k has, at each period of the horizon considered, at least one protected zone which constitutes a favourable habitat for it. It should be remembered that the decisions to protect zones – and the implementation of these protections – are made at the beginning of the horizon, without the possibility of modification. We obtain program $\mathrm{P}_{12.4}$ which is none other than program $\mathrm{P}_{12.1}$ in which constraints 12.1.1 are replaced by constraints $y_k \leq \sum_{i \in \underline{Z}_{kt}} x_i$, $k \in \underline{S}, t \in \underline{T}$. According to these constraints and the economic function to be maximized, variable y_k takes the value 1, at the optimum of the program, if and only if, at all the periods of the horizon, at least one of the zones of R constitutes a favourable habitat for species s_k.

$$
\mathrm{P}_{12.4}: \begin{cases} \max \displaystyle\sum_{k \in \underline{S}} y_k \\ \text{s.t.} \left| \begin{array}{llll} y_k \leq \displaystyle\sum_{i \in \underline{Z}_{kt}} x_i & k \in \underline{S},\ t \in T & (12.4.1) & \mid \quad x_i \in \{0,1\} \quad i \in \underline{Z} \quad (12.4.3) \\ \displaystyle\sum_{i \in \underline{Z}} c_i x_i \leq B & & (12.4.2) & \mid \quad y_k \in \{0,1\} \quad k \in \underline{S} \quad (12.4.4) \end{array} \right. \end{cases}
$$

As mentioned earlier, climate change may force some species to migrate from one zone to another. Let us now look at how to modify the wording of Problem I to take these potential migrations into account. It is assumed, as before, that the set of zones in Z, Z_{kt}, which are favourable habitat for species s_k during the period T_t, are known for all $k \in \underline{S}$ and for all $t \in \underline{T}$. It is now considered that a zone z_i of the reserve protects species s_k in the period T_t if, on the one hand, the habitat in z_i is favourable to species s_k in the period T_t, *i.e.*, if $z_i \in Z_{kt}$, and, on the other hand, if zone z_i or a zone adjacent to z_i already protected species s_k in the period T_{t-1}. It is assumed for all $k \in \underline{S}$, that species s_k is protected by the reserve, during the first period, if at least one of the zones in Z_{k1} belongs to the reserve. Implicit in these hypotheses is the assumption that, to some extent, the species are able to move around the reserve. Therefore, the focus is now on determining a reserve that maximizes, within an available budget, the number of species protected, a species being protected if it is protected at each period of the time horizon considered and, therefore, at the end of the horizon. This new version of Problem I can be formulated as program $\mathrm{P}_{12.5}$ in which the meaning of variable y_k, $k \in \underline{S}$, is the same as in programs $\mathrm{P}_{12.4}$. We also use the Boolean variable α_{ikt}, $i \in \underline{Z}, k \in \underline{S}, t \in \underline{T}$. This variable takes the value 1 if and only if zone z_i protects species s_k during the period T_t, *i.e.*, if and only if the following three conditions are satisfied: (1) zone z_i is selected to be part of the reserve ($x_i = 1$), (2) the habitat of zone z_i is favourable to species s_k at the period T_t ($z_i \in Z_{kt}$), and (3) zone z_i, or a zone adjacent to z_i already protected species s_k at the period T_{t-1}. This third condition is to be satisfied only from the second period of the horizon ($t \geq 2$).

$$
\mathrm{P}_{12.5}: \left\{
\begin{array}{ll}
\max \sum_{k \in \underline{S}} y_k & \\
\text{s.t.} & \left|
\begin{array}{lll}
\alpha_{ikt} \leq x_i & i \in \underline{Z}, k \in \underline{S}, t \in \underline{T} & (12.5.1)\\
\alpha_{ikt} = 0 & k \in \underline{S}, t \in \underline{T}, i \in \underline{Z} - \underline{Z}_{kt} & (12.5.2)\\
\alpha_{ikt} \leq \alpha_{ikt-1} + \sum_{j \in \mathrm{Adj}_i} \alpha_{jkt-1} & i \in \underline{Z}, k \in \underline{S}, t \in \underline{T}, t \geq 2 & (12.5.3)\\
y_k \leq \sum_{i \in \underline{Z}_{kt}} \alpha_{ikt} & k \in \underline{S}, t \in \underline{T} & (12.5.4)\\
\sum_{i \in \underline{Z}} c_i x_i \leq B & & (12.5.5)\\
x_i \in \{0,1\} & i \in \underline{Z} & (12.5.6)\\
y_k \in \{0,1\} & k \in \underline{S} & (12.5.7)\\
\alpha_{ikt} \in \{0,1\} & i \in \underline{Z}, k \in \underline{S}, t \in \underline{T} & (12.5.8)
\end{array}
\right.
\end{array}
\right.
$$

Constraints 12.5.1 require variable α_{ikt}, $i \in \underline{Z}, k \in \underline{S}, t \in \underline{T}$, to take the value 0 if zone z_i is not selected. Constraints 12.5.2 require variable α_{ikt}, $k \in \underline{S}, t \in \underline{T}, i \in \underline{Z} - \underline{Z}_{kt}$, to take the value 0. Indeed, in this case, zone z_i does not constitute a favourable habitat for species s_k during the period T_t. In other cases, variable α_{ikt} may take a priori the value 1 but because of constraint 12.5.3 it can only take the value 1 if at least one of the variables in the set $\{\alpha_{ikt-1}\} \cup \{\alpha_{jkt-1}, j \in \mathrm{Adj}_i\}$ was already taking the value 1. Adj_i designates the set of indices of the zones adjacent to zone z_i. This last constraint thus reflects the fact that zone z_i cannot protect species s_k during period T_t if this zone or a zone adjacent to it did not already protect this species during the previous period, T_{t-1}. Because of the economic function to be maximized and constraints 12.5.4, variable y_k takes the value 1, at the program optimum, if, for each period T_t, at least one of variables α_{ikt}, $i \in \underline{Z}_{kt}$, takes the value 1.

Example 12.1. Let us illustrate this new version of Problem I and its resolution by program $\mathrm{P}_{12.5}$ on a small example with 7 candidate zones, 10 species, and 4 periods. Figure 12.1 describes this example by presenting, for each of the zones, the species for which these zones constitute a favourable habitat and this for each of the 4 periods.

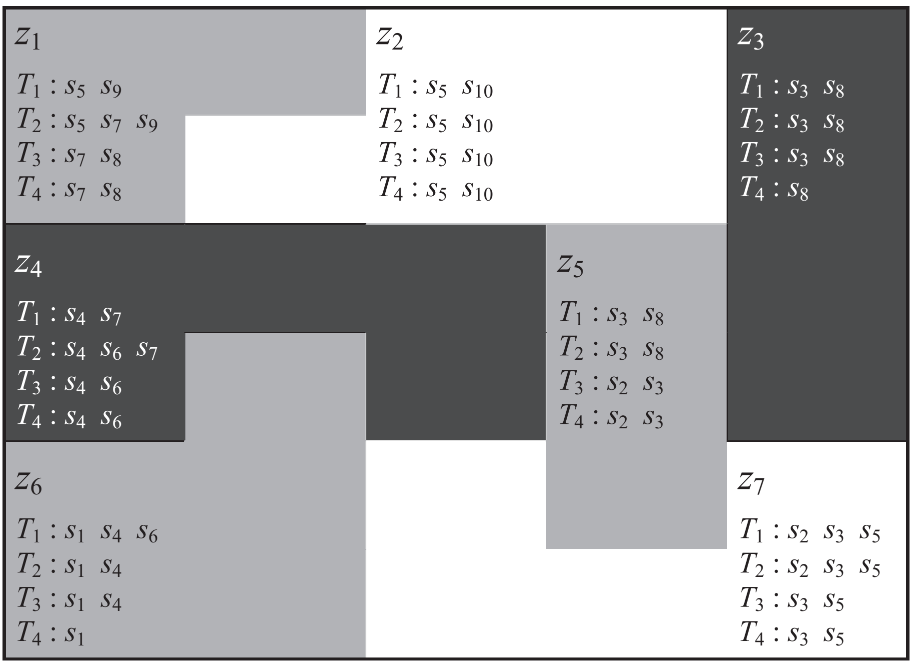


FIG. 12.1 – Seven zones, z_1, z_2,..., z_7, are candidates for protection and 10 species, s_1, s_2,..., s_{10}, living in these zones, in the first period, are concerned. For each zone z_i and each period T_t, the species for which the zone in question constitutes a favourable habitat (if protected), during the period considered are indicated. The cost of protecting the white zones is equal to 1 unit, the cost of protecting the light grey zones is equal to 2 units and the cost of protecting the dark grey zones is equal to 4 units. For example, zone z_5, if protected, provides a favourable habitat for species s_3 and s_8 in the first two periods and for species s_2 and s_3 in the next two periods. The cost of protecting this zone is equal to 2 units.

TAB. 12.1 – Optimal reserves and associated protected species obtained by resolving $P_{12.5}$ in the case of the example shown in figure 12.1, for different values of the available budget, B.

B	Used budget	Reserve	Protected species	B	Used budget	Reserve	Protected species
1	1	z_2	$s_5\ s_{10}$	8	8	$z_1\ z_2\ z_4\ z_6\ z_7$	$s_1\ s_3\ s_4\ s_5\ s_6\ s_7\ s_{10}$
2	2	$z_2\ z_6$	$s_1\ s_5\ s_{10}$	9	8	$z_1\ z_2\ z_4\ z_6\ z_7$	$s_1\ s_3\ s_4\ s_5\ s_6\ s_7\ s_{10}$
3	2	$z_2\ z_6$	$s_1\ s_5\ s_{10}$	10	8	$z_1\ z_2\ z_4\ z_6\ z_7$	$s_1\ s_3\ s_4\ s_5\ s_6\ s_7\ s_{10}$
4	4	$z_2\ z_4\ z_6$	$s_1\ s_4\ s_5\ s_6\ s_{10}$	11	8	$z_1\ z_2\ z_4\ z_6\ z_7$	$s_1\ s_3\ s_4\ s_5\ s_6\ s_7\ s_{10}$
5	4	$z_2\ z_4\ z_6$	$s_1\ s_4\ s_5\ s_6\ s_{10}$	12	12	$z_1\ z_2\ z_3\ z_4\ z_6\ z_7$	$s_1\ s_3\ s_4\ s_5\ s_6\ s_7\ s_8\ s_{10}$
6	6	$z_2\ z_4\ z_6\ z_7$	$s_1\ s_3\ s_4\ s_5\ s_6\ s_{10}$	13	12	$z_1\ z_2\ z_3\ z_4\ z_6\ z_7$	$s_1\ s_3\ s_4\ s_5\ s_6\ s_7\ s_8\ s_{10}$
7	6	$z_2\ z_4\ z_6\ z_7$	$s_1\ s_3\ s_4\ s_5\ s_6\ s_{10}$	14	12	$z_1\ z_2\ z_3\ z_4\ z_6\ z_7$	$s_1\ s_3\ s_4\ s_5\ s_6\ s_7\ s_8\ s_{10}$

The optimal reserves obtained, taking into account the available budget, are presented in table 12.1. It can be seen from this table that even if a budget were available to protect all the 7 zones, only 8 out of the 10 species could be protected. This is due to the two phenomena presented above: (1) a zone constitutes a favourable habitat for a species at a certain period but this is no longer the case at a later period, and (2) a protected zone can protect a given species at a period T_t only if that zone or one of its adjacent zones already protected that species at period T_{t-1}. So, the maximal number of species that can be protected is 8 and the cheapest solution to obtain this protection is to protect all the zones except z_5, which costs 12 units.

Figure 12.2 shows the optimal reserve when the available budget is equal to 7 units, and the species protected by the different zones during the 4 periods. Only

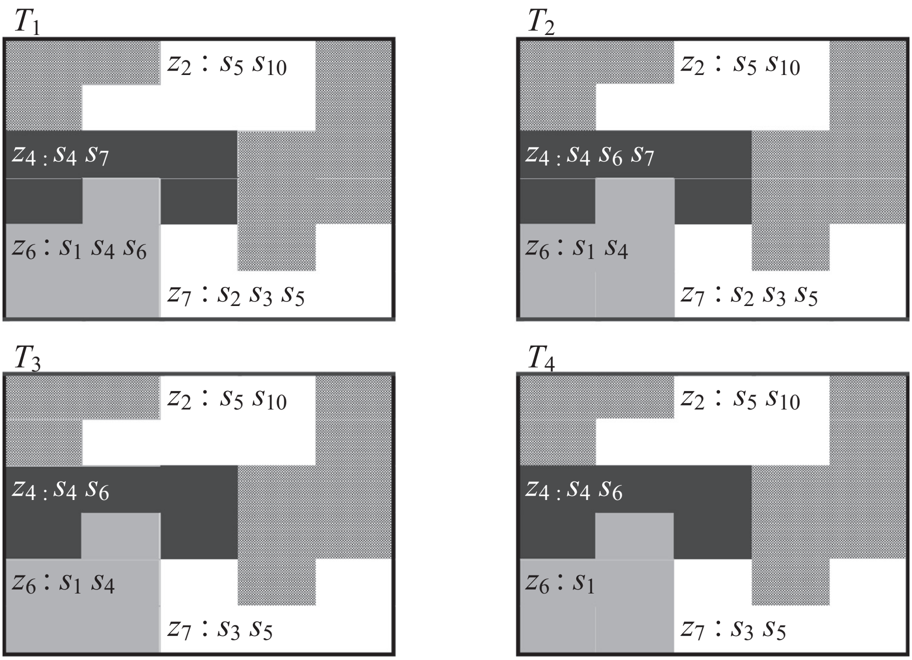


FIG. 12.2 – Optimal solution for a budget of 7 units; species protected by the reserve during the 4 periods of the considered horizon. In period T_1, 8 species are protected; in period T_2, these 8 species are still protected but species s_6 has migrated from zone z_6 to zone z_4; in period T_3, there are only 6 species protected since species s_2 and s_7 are no longer protected, the other species have not migrated; finally in period T_4, there are still 6 species protected although species s_4 is no longer protected by zone z_6 but zone z_4 still protects this species. The 6 species did not have to migrate.

species s_1, s_3, s_4, s_5, s_6, and s_{10} are protected in each period of the horizon considered.

12.3.2 Problem II

With regard to the extension of Problem II, it is assumed, on the one hand, that climate change is causing a change in the population size of the species in each candidate protected zone over time and, on the other hand, that it is possible to estimate this change. Let n_{ikt}, $(i, k, t) \in \underline{Z} \times \underline{S} \times \underline{T}$, be the predicted population size of species s_k, in the protected zone z_i and during the period T_t. Species s_k is assumed to survive in a given reserve, R, in period T_t if its total population size in that reserve is greater than or equal to θ_k. It is assumed here that this threshold value is not time-dependent, but it would be easy to consider the more general case where this value is time-dependent. This value would then be denoted by θ_{kt}. The problem that emerges in the case of Problem II is to determine a set of zones to be protected from the beginning of the considered horizon, with a cost less than or equal to a given value, and optimal with regard to the conservation of the species under consideration. As in Problem I, it is necessary to define what constitutes an optimal reserve. As in Problem I, the aim is to determine a reserve that maximizes, within an available budget, the number of species that survives at the end of the considered horizon. For this problem, the natural assumption is that a species survives at the end of the horizon if it survives at each period of the horizon, *i.e.*, as noted above, if its population size in the reserve, at each period, is greater than or equal to the threshold value.

To formulate this problem by mathematical programming, we use, as for Problem I, the Boolean variable $x_i,\ i \in \underline{Z}$, which takes the value 1 if and only if zone z_i is protected at the beginning of the horizon considered and therefore throughout this horizon, and the Boolean variable $y_k,\ k \in \underline{S}$, which takes the value 1 if and only if species s_k survives until the end of the horizon considered. We obtain program $\mathrm{P}_{12.6}$ which is none other than program $\mathrm{P}_{12.2}$ in which constraints 12.2.1 are replaced by constraints $\theta_k y_k \leq \sum_{i \in \underline{Z}} n_{ikt} x_i,\ k \in \underline{S},\ t \in \underline{T}$. These constraints allow variable y_k to take, at the optimum of program $\mathrm{P}_{12.6}$, the value 1 if and only if, at all the periods of the horizon, the population size of species s_k in the reserve is greater than or equal to the threshold value, θ_k.

$$
\mathrm{P}_{12.6}: \begin{cases} \max \sum_{k \in \underline{S}} y_k \\ \text{s.t.} \left| \begin{array}{llll} \theta_k y_k \leq \sum_{i \in \underline{Z}} n_{ikt} x_i & k \in \underline{S},\ t \in \underline{T} & (12.6.1) & \mid \quad x_i \in \{0,1\} \quad i \in \underline{Z} \quad (12.6.3) \\ \sum_{i \in \underline{Z}} c_i x_i \leq B & & (12.6.2) & \mid \quad y_k \in \{0,1\} \quad k \in \underline{S} \quad (12.6.4) \end{array} \right. \end{cases}
$$

12.3.3 Problem III

12.3.3.1 Static Approach

With regard to Problem III, which has been adapted to take account of climate change, it is assumed that the area of habitat favourable to species s_k in zone z_i is known for all the periods of the time horizon considered if zone z_i is protected from the beginning of the horizon considered. This area is denoted by a_{ikt}, $i \in \underline{Z}, k \in \underline{S}, t \in \underline{T}$. As before, the aim is to define an optimal reserve at the beginning of the horizon under consideration. Thus, each candidate zone is protected, or not, from the beginning of the horizon and during all the periods of this horizon. For this problem, a "goal programming" approach is adopted, *i.e.*, one seeks a reserve that ensures, for each of the species under consideration and at each period of the horizon, a total area of favourable habitat – included in the reserve – as close as possible to a target value. The target value for species s_k is denoted by $\min_k$, $k \in \underline{S}$. To simplify the presentation, it is assumed that it is not time-dependent.

To formulate this problem by mathematical programming, we use as previously the Boolean variable x_i , $i \in \underline{Z}$, which takes the value 1 if and only if zone z_i is protected at the beginning of the horizon considered and therefore throughout this horizon, but also variable g_{kt}, $k \in \underline{S}$, $t \in \underline{T}$, which expresses the gap between the total area of habitat favourable to species s_k in the reserve, at the period T_t, and the target value for this species, $\min_k$ – value independent of the period. This gap is only taken into account if the total area of habitat favourable to species s_k on the reserve at time T_t is less than $\min_k$. In other words, $g_{kt} = \max\{0, (\min_k - \sum_{i \in \underline{Z}} a_{ikt} x_i)\}$. This gives program $\mathrm{P}_{12.7}$, which corresponds to program $\mathrm{P}_{12.3}$ in which the economic function is replaced by $\sum_{k \in \underline{S}, t \in \underline{T}} \xi_k g_{kt}$, and constraints 12.3.1 by constraints $\sum_{i \in \underline{Z}} a_{ikt} x_i + g_{kt} \geq \min_k$, $k \in \underline{S}$, $t \in \underline{T}$. Note that the objective achieved by a given reserve is evaluated globally since it is measured by the sum of the gaps over all species and over all periods.

$$
\mathrm{P}_{12.7}: \left\{
\begin{array}{ll}
\min & \displaystyle\sum_{k \in \underline{S},\, t \in \underline{T}} \xi_k g_{kt} \\
\text{s.t.} & \left|
\begin{array}{lll}
\displaystyle\sum_{i \in \underline{Z}} a_{ikt} x_i + g_{kt} \geq \min_k & k \in \underline{S}, t \in \underline{T} & (12.7.1) \\
\displaystyle\sum_{i \in \underline{Z}} c_i x_i \leq B & & (12.7.2) \\
x_i \in \{0, 1\} & i \in \underline{Z} & (12.7.3) \\
g_{kt} \geq 0 & k \in \underline{S}, t \in \underline{T} & (12.7.4)
\end{array}
\right.
\end{array}
\right.
$$

The economic function expresses the weighted sum, for all the periods T_t and all species s_k, of the gaps between the total area of habitat favourable to species s_k in the reserve in period T_t and the period-independent target value for that species, $\min_k$. It is assumed here that the weighting coefficient, ξ_k, does not depend on t. Since constraint 12.7.4 requires g_{kt} to be non-negative, these gaps are only considered if the total area of habitat favourable to species s_k in the reserve at the time T_t is less than $\min_k$. Constraints 12.7.1 express the value of these gaps at the optimum of the program.

12.3.3.2 Dynamic Approach

Here we examine a "dynamic" variant of the extension of Problem III. The essential difference with the "static" Problem III, presented in the previous section, is that the configuration of the reserve can change over time. However, the decisions are made at the beginning of the horizon (see section 12.3.3.3 for an issue where the decisions may be questioned over time). Thus, at the beginning of each period, a certain budget is available – B_t at the beginning of the period T_t – and a decision can be made to acquire zones for protection but also to cede zones – which would no longer be interesting – in order to increase the budget available at the beginning of this period. It is assumed that the unused budget in a period is lost, but this assumption could easily be changed (see chapter 1, section 1.4). The management horizon, T, is formed as before of r periods $T_1,\ldots,\ T_r$ and $\underline{T}$ designates the set of indices $\{1,\ldots,\ r\}$. As with the static problem, a reserve is searched to ensure that for each species considered and each period of the horizon a total area of favourable habitat – included in the reserve – is as close as possible to a target value. The target value for species s_k, $k \in \underline{S}$, is denoted by $\min_k$. To simplify the presentation it is assumed that it is not period-dependent. It is assumed that the area of habitat favourable to species s_k in zone z_i is known for all the periods of the time horizon considered. This area is denoted by $a_{ikt},\ i \in \underline{Z}, k \in \underline{S}, t \in \underline{T}$; here it does not depend, a priori, on whether zone z_i is protected or not. However, some zones that have not yet been included in the reserve at a given time may be allocated to certain activities that cause them to lose their status as candidate zones for protection. More specifically, as an example, we consider here, that the following two constraints should be taken into account:

- ($C_{12.1}$): A zone of the reserve that has been ceded at a certain period can no longer be acquired at the beginning of a subsequent period to be returned to the reserve.
- ($C_{12.2}$): After a certain period of time, certain zones, which were available for inclusion in the reserve, are no longer available if they have not already been included in the reserve. Let us denote by $T_{t(i)}$, $i \in \underline{Z}$, the period after which it is no longer possible to acquire zone z_i.

We denote by $c_{it},\ i \in \underline{Z},\ t \in \underline{T}$, the cost of acquisition of zone z_i at the beginning of the period T_t in order to protect it, $v_{it},\ i \in \underline{Z},\ t \in \underline{T}$, the cost of cession of zone z_i at the beginning of the period T_t, and B_t the budget available at the beginning of the period T_t, not taking into account the cessions carried out at the beginning of the period T_t. The proceeds of these cessions are considered to be available at the beginning of the period T_t. It should therefore be added to B_t to define the total budget available at the beginning of this period. The composition of the reserve can thus change over time, but all the decisions regarding acquisitions and cessions are made at the beginning of the considered horizon. The following Boolean variables are used to formulate the problem as a mathematical program: $y_{it},\ i \in \underline{Z}, t \in \underline{T}$, which takes the value 1 if and only if zone z_i is acquired at the beginning of the period T_t, and u_{it}, $i \in \underline{Z}$, $t \in \underline{T}$, which takes the value 1 if and only if zone z_i is ceded at the beginning of the period T_t. The Boolean variable x_{it} is also used, which takes the value 1 if and only if zone z_i is part of the reserve at the beginning of the period T_t and thus throughout the period T_t. This variable thus takes the value 1 if and only if zone z_i was acquired at the beginning of one of the periods $T_1,\ldots,\ T_t$ and not ceded at the beginning of one of these same periods. The problem considered can be formulated as the mathematical program $\mathrm{P}_{12.8}$.

$$
\mathrm{P}_{12.8}: \begin{cases}
\min \displaystyle\sum_{k\in\underline{S},\, t\in\underline{T}} \xi_k g_{kt} & & \\
\text{s.t.} \left| \begin{array}{lll}
\displaystyle\sum_{i\in\underline{Z}} a_{ikt}x_{it} + g_{kt} \geq \min_k & k \in \underline{S}, t \in \underline{T} & (12.8.1) \\
\displaystyle\sum_{i\in\underline{Z}} c_{i1}x_{i1} \leq B_1 & & (12.8.2) \\
\displaystyle\sum_{i\in\underline{Z}} c_{it}y_{it} \leq B_t + \sum_{i\in\underline{Z}} v_{it}u_{it} & t \in \underline{T}, t \geq 2 & (12.8.3) \\
x_{it} = x_{it-1} + y_{it} - u_{it} & i \in \underline{Z}, t \in \underline{T}, t \geq 2 & (12.8.4) \\
y_{it} + u_{it} \leq 1 & i \in \underline{Z}, t \in \underline{T}, t \geq 2 & (12.8.5) \\
\displaystyle\sum_{l=t+1,\ldots,r} y_{il} \leq 1 - u_{it} & i \in \underline{Z}, t \in \underline{T}, t \geq 2 & (12.8.6) \\
\displaystyle\sum_{t=t(i),\ldots,r} y_{it} = 0 & i \in \underline{Z} & (12.8.7) \\
x_{it} \in \{0,1\} & i \in Z, t \in \underline{T} & (12.8.8) \\
y_{it} \in \{0,1\}, u_{it} \in \{0,1\} & i \in Z, t \in \underline{T}, t \geq 2 & (12.8.9) \\
g_{kt} \geq 0 & k \in \underline{S}, t \in \underline{T} & (12.8.10)
\end{array} \right.
\end{cases}
$$

The economic function of $\mathrm{P}_{12.8}$ expresses the weighted sum, for all the periods T_t and all species s_k, of the gaps between the total habitat area of the reserve favourable to species s_k, at the period T_t, and the target value for that species, $\min_k$ – independent of the period. These gaps, represented by variables g_{kt}, are only accounted for if the total area of habitat favourable to species s_k in the reserve at time T_t is less than $\min_k$. Constraints 12.8.1 combined with constraints 12.8.10 express the value of these gaps at the program optimum. Constraint 12.8.2 expresses the budget constraint, for the first period, and constraints 12.8.3 express the budget constraint, for all the other periods. Constraint 12.8.4 expresses that zone z_i belongs to the reserve in the period T_t $(x_{it} = 1)$ in the following cases: (1) it already belonged to the reserve at the period T_{t-1} $(x_{it\text{-}1} = 1)$ and was not ceded at the beginning of the period T_t $(u_{it} = 0)$, 2) it did not belong to the reserve at the period T_{t-1} $(x_{it\text{-}1} = 0)$ and was acquired at the beginning of the period T_t $(y_{it} = 1)$. Constraints 12.8.5 express the impossibility of carrying out simultaneously at the beginning of each period the acquisition and the cession of the same zone. Constraints 12.8.6 express that if zone z_i, $i \in \underline{Z}$, has been ceded at the period T_t, $t \in \underline{T}, t \geq 2$, then this zone cannot be

Zone z_1 ($\leq T_2$)

	T_1	T_2	T_3	T_4
s_1	0.7	0.0	0.6	0.5
s_2	0.2	0.5	0.6	0.7
s_3	1.0	0.2	0.8	0.7
s_4	0.1	0.1	0.3	0.6
s_5	0.8	0.0	0.4	0.0
s_6	0.6	0.4	0.0	0.0
s_7	0.2	0.1	0.7	0.3
s_8	0.0	0.0	0.8	0.0

Zone z_2 ($\leq T_1$)

	T_1	T_2	T_3	T_4
s_1	0.3	0.1	0.3	0.7
s_2	0.0	0.7	0.7	0.2
s_3	0.5	0.0	0.6	0.9
s_4	0.0	0.8	0.1	0.3
s_5	0.2	0.5	0.9	0.0
s_6	0.8	0.3	0.7	1.0
s_7	0.8	0.5	0.5	0.8
s_8	0.3	0.4	0.0	0.0

Zone z_3 ($\leq T_4$)

	T_1	T_2	T_3	T_4
s_1	0.2	0.2	0.2	0.8
s_2	0.4	0.7	0.6	0.0
s_3	0.0	0.0	0.6	0.9
s_4	0.0	0.0	0.2	0.6
s_5	0.7	0.2	0.9	0.3
s_6	0.9	0.6	1.0	0.0
s_7	0.3	1.0	0.9	0.1
s_8	0.6	0.3	0.4	0.5

Zone z_4 ($\leq T_3$)

	T_1	T_2	T_3	T_4
s_1	0.4	0.0	0.0	0.9
s_2	0.3	0.5	0.9	1.0
s_3	0.3	0.0	0.5	0.0
s_4	0.6	0.3	0.0	0.2
s_5	0.0	0.7	0.0	0.4
s_6	0.1	0.0	0.0	0.0
s_7	0.0	0.6	0.8	0.0
s_8	0.1	0.1	0.0	0.3

Zone z_5 ($\leq T_2$)

	T_1	T_2	T_3	T_4
s_1	0.1	0.2	0.9	0.0
s_2	0.9	1.0	0.9	0.6
s_3	0.2	0.5	1.0	0.5
s_4	0.1	0.5	0.3	0.8
s_5	0.4	0.6	0.6	0.6
s_6	0.6	0.4	0.7	0.3
s_7	0.3	0.4	1.0	0.9
s_8	0.2	0.7	0.8	0.0

Zone z_6 ($\leq T_1$)

	T_1	T_2	T_3	T_4
s_1	0.2	0.2	0.4	0.1
s_2	0.8	0.3	0.8	0.0
s_3	0.1	0.6	0.9	0.5
s_4	0.0	0.0	0.3	0.7
s_5	0.1	0.1	0.9	0.3
s_6	0.6	0.9	0.3	0.9
s_7	0.8	0.0	0.5	0.0
s_8	0.6	0.8	0.7	0.0

Zone z_7 ($\leq T_1$)

	T_1	T_2	T_3	T_4
s_1	0.6	0.0	0.8	0.3
s_2	1.0	0.2	0.3	0.3
s_3	0.0	0.0	0.6	0.6
s_4	0.4	0.0	0.0	0.1
s_5	0.3	0.1	1.0	0.9
s_6	0.3	0.5	0.7	0.9
s_7	0.2	0.4	0.8	0.0
s_8	1.0	0.9	1.0	0.0

Zone z_8 ($\leq T_4$)

	T_1	T_2	T_3	T_4
s_1	1.0	0.0	0.3	0.1
s_2	0.6	0.3	0.0	0.0
s_3	0.6	0.4	0.8	0.3
s_4	0.9	1.0	0.5	0.9
s_5	1.0	0.8	0.7	0.9
s_6	0.5	0.3	0.8	0.3
s_7	0.5	0.0	0.2	0.8
s_8	0.3	0.0	0.7	0.4

Zone z_9 ($\leq T_2$)

	T_1	T_2	T_3	T_4
s_1	0.0	0.0	0.0	0.2
s_2	0.5	0.4	0.2	0.6
s_3	0.7	1.0	0.8	0.0
s_4	0.0	0.6	0.5	0.9
s_5	0.0	0.0	0.0	0.0
s_6	0.3	0.3	0.7	0
s_7	0.0	0.7	0.2	0.5
s_8	0.8	0.6	0.1	0.5

FIG. 12.3 – Description of a hypothetical instance with 9 zones z_1, z_2,..., z_9 and 8 species s_1, s_2,..., s_8. The area of each zone is equal to one unit. In each zone z_i are indicated: the date until which the zone can be acquired and, for each pair (s_k, T_t), the fraction of the area of this zone that constitutes a favourable habitat for species s_k during the period T_t. For example, zone z_8 may be acquired in the periods T_1, T_2, T_3, or T_4 and 70% of the area of zone z_8 is favourable to species s_5 in the period T_3.

acquired later to be reintegrated into the reserve. Constraints 12.8.7 express that from the period $T_{t(i)}$ onwards it is no longer possible to acquire zone z_i for inclusion in the reserve. Finally, constraints 12.8.8–12.8.10 specify the nature of the variables.

Example 12.2. Consider the instance described in figure 12.3 (9 square and identical zones with an area of one unit, 8 species, 4 periods) and table 12.2.

TAB. 12.2 – Acquisition and cession costs of the 9 zones during the 4 periods.

	c_{1t}	c_{2t}	c_{3t}	c_{4t}	c_{5t}	c_{6t}	c_{7t}	c_{8t}	c_{9t}		v_{1t}	v_{2t}	v_{3t}	v_{4t}	v_{5t}	v_{6t}	v_{7t}	v_{8t}	v_{9t}
$t = 1$	8	4	6	10	7	8	5	8	8	$t = 1$	–	–	–	–	–	–	–	–	–
$t = 2$	5	–	6	6	5	–	–	9	7	$t = 2$	5	8	6	6	5	5	4	9	7
$t = 3$	–	–	8	4	–	–	–	7	–	$t = 3$	10	7	8	4	9	5	6	7	6
$t = 4$	–	–	8	–	–	–	–	4	–	$t = 4$	10	5	8	9	9	7	9	4	8

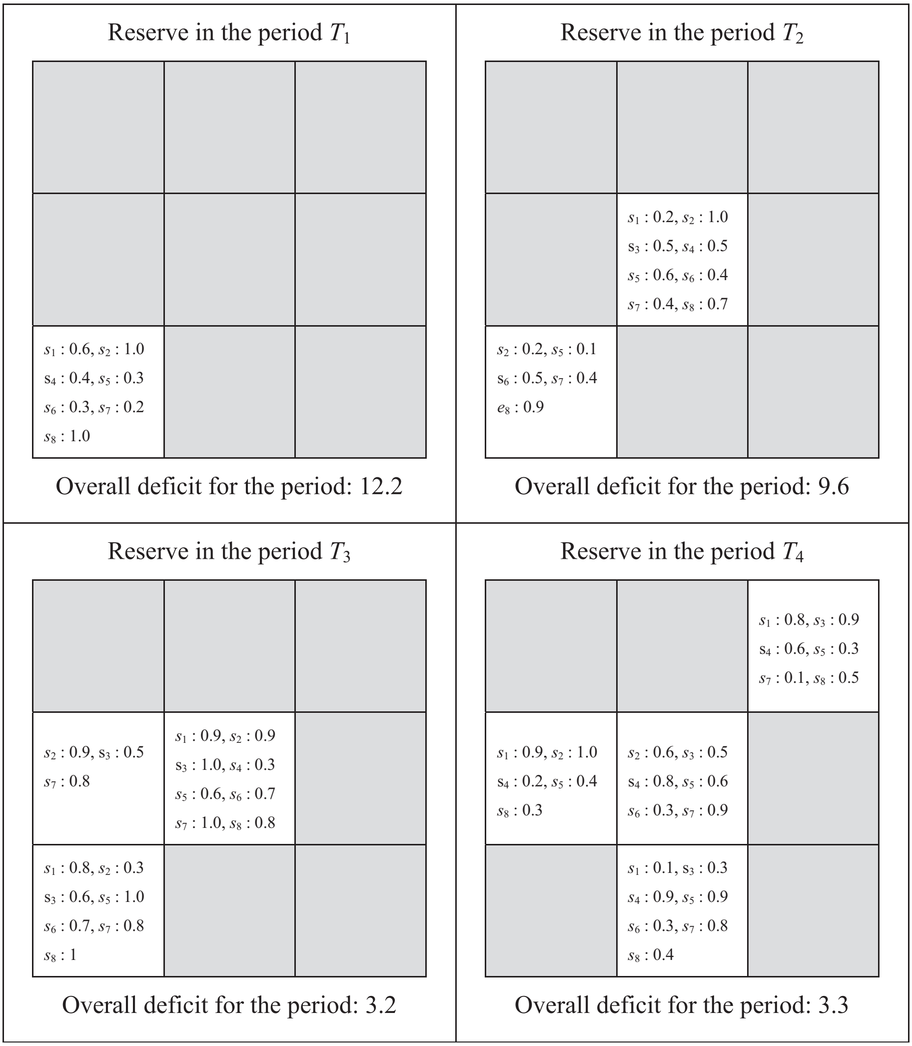


FIG. 12.4 – Optimal solution for the instance described in figure 12.3 and table 12.2. Zone z_7 is acquired at the beginning of the period T_1. At the beginning of the period T_2, zone z_5 is acquired. At the beginning of the period T_3, zone z_4 is acquired. Finally, zones z_3 and z_8 are acquired at the beginning of the period T_4, and zone z_7 is ceded at the beginning of the same period. The overall deficit is equal to 12.2 + 9.6 + 3.2 + 3.3 = 28.3.

Let us solve the problem when the budget available at the beginning of each period, B_t, $t \in \underline{T}$, is equal to 5, the target value for each species, $\min_k$, is equal to 2 units, and $\xi_k = 1$, $k = 1, \ldots, 8$. The optimal solution obtained is described in figure 12.4, which shows the zones belonging to the reserve in the different periods. In each zone and for each period, the fraction of habitat favourable to the species is indicated when this fraction is not zero. The overall deficit associated with each period is also indicated.

12.3.3.3 Adaptive Management: Review at Each Period of Earlier Decisions

Let us return to the problem considered in section 12.3.3.2. Suppose that it has been solved by $\mathrm{P}_{12.8}$, that we have reached the end of the period T_{j-1} and that the forecasts for the following periods, *i.e.*, $T_j, \ldots, T_r$, have changed. Thus, for $i \in \underline{Z}$, $k \in \underline{S}$, $t \in \underline{T}$, $t \geq j$, a_{ikt} has become $\hat{a}_{ikt}$, for $i \in \underline{Z}$, $t \in \underline{T}$, $t \geq j$, c_{it} has become $\hat{c}_{it}$ and v_{it} has become $\hat{v}_{it}$, for $i \in \underline{Z}$, $t(i)$ has become $\hat{t}(i)$ and, finally, for $t \in \underline{T}$, $t \geq j$, B_t has become $\hat{B}_t$. At the end of the period T_{j-1}, the reserve is formed of certain zones. The composition of this reserve is the consequence of the acquisitions and cessions made during the periods $T_1, \ldots, T_{j-1}$. Denote by $\overline{x}_{i,j-1}$ the value taken by variable $x_{i,j-1}$, $i \in \underline{Z}$, in the optimal solution of $\mathrm{P}_{12.8}$ and $\overline{u}_{it}$ the value taken by variable u_{it}, $i \in \underline{Z}, t \in \underline{T}, 2 \leq t \leq j-1$, in the same optimal solution. The composition of the reserve is defined by $\overline{x}_{i,j-1}$, $i \in \underline{Z}$. The zones to be acquired or ceded in subsequent periods, $T_j, \ldots, T_r$, can then be determined, taking into account the updated forecasts, by resolving program $\mathrm{P}_{12.9}$.

$$
\mathrm{P}_{12.9}: \left\{
\begin{array}{lll}
\min \displaystyle\sum_{k \in \underline{S},\, t \in \underline{T},\, t \geq j} \xi_k g_{kt} & & \\
\text{s.t.} \quad \displaystyle\sum_{i \in \underline{Z}} \hat{a}_{ikt} x_{it} + g_{kt} \geq \min_k & k \in \underline{S}, t \in \underline{T}, t \geq j & (12.9.1) \\
\displaystyle\sum_{i \in \underline{Z}} \hat{c}_{it} y_{it} \leq \hat{B}_t + \sum_{i \in \underline{Z}} \hat{v}_{it} u_{it} & t \in \underline{T}, t \geq j & (12.9.2) \\
x_{i,j-1} = \overline{x}_{i,j-1} & i \in \underline{Z} & (12.9.3) \\
x_{it} = x_{it-1} + y_{it} - u_{it} & i \in \underline{Z}, t \in \underline{T}, t \geq j & (12.9.4) \\
y_{it} + u_{it} \leq 1 & i \in \underline{Z}, t \in \underline{T}, t \geq j & (12.9.5) \\
u_{it} = \overline{u}_{it} & i \in \underline{Z}, t \in \underline{T}, t \leq j-1 & (12.9.6) \\
\displaystyle\sum_{l = \max\{j,\, t+1\}, \ldots, r} y_{il} \leq 1 - u_{it} & i \in \underline{Z}, t \in \underline{T}, t \geq 2 & (12.9.7) \\
\displaystyle\sum_{t = \hat{t}(i), \ldots, r} y_{it} = 0 & i \in \underline{Z} & (12.9.8) \\
x_{it} \in \{0, 1\} & i \in \underline{Z}, t \in \underline{T}, t \geq j-1 & (12.9.9) \\
y_{it} \in \{0, 1\} & i \in \underline{Z}, t \in \underline{T}, t \geq j & (12.9.10) \\
u_{it} \in \{0, 1\} & i \in \underline{Z}, t \in \underline{T}, t \geq 2 & (12.9.11) \\
g_{kt} \geq 0 & k \in \underline{S}, t \in \underline{T}, t \geq j & (12.9.12)
\end{array}
\right.
$$

The economic function and constraints of $P_{12.9}$ are similar to the economic function and constraints of $P_{12.8}$. Constraints 12.9.1, 12.9.2, 12.9.4, and 12.9.5 concern only the periods after period T_{j-1}. The composition of the reserve at the end of the period T_{j-1} is defined by constraints 12.9.3. Since $\overline{u}_{it}$ is the value taken by variable u_{it} in the optimal solution of $P_{12.8}$, constraints 12.9.6 and 12.9.7 express that a zone z_i, ceded at period T_t, can no longer be acquired at the beginning of a subsequent period to be reintegrated into the reserve. Constraints 12.9.8 express, for all $i \in \underline{Z}$, that from the period $T_{\hat{t}(i)}$ – subsequent to the period T_j – zone z_i can no longer be included in the reserve if this has not already been done. Constraints 12.9.9–12.9.12 specify the nature of the different variables.

The resolution of program $P_{12.8}$ provides a solution to the problem of section 12.3.3.2, which is to determine the zones that must be acquired for protection and those that must be ceded, at the beginning of each period of the horizon under consideration, in order to ensure for each species a total area of favourable habitat as close as possible to a target value, taking into account the available budget. Decisions are made at the beginning of the horizon, and the relevance of these decisions is highly dependent on the quality of the various forecasts. We have just shown – in this section 12.3.3.3 – how to adapt, at the end of the period T_{j-1}, the optimal solution obtained at the beginning of the horizon under consideration to take into account changes in the forecasts for the periods $T_j, \ldots, T_r$. This process can be repeated at the end of each period, *i.e.*, for $j = 2, \ldots, r$.

12.4 Taking into Account Climate Change, Described by a Set of Scenarios, in Problems I, II and III; Conservative Approach

This section realistically considers that the impacts of climate change are not known for sure and that several hypotheses can be considered. To reflect the uncertainty in the ability of different zones to protect species over a given time horizon, we consider a set of possible scenarios, $\mathrm{Sc} = \{\mathrm{sc}_1, \mathrm{sc}_2, \ldots, \mathrm{sc}_p\}$. A scenario is here a set of assumptions about climate change and its consequences for the survival of the species under consideration in the candidate zones for protection and this for the entire management horizon under consideration. We set $\underline{\mathrm{Sc}} = \{1, 2, \ldots, p\}$. The identification of the different scenarios and the description of their consequences are delicate tasks. We resume Problems I, II and III in this framework.

12.4.1 Problem I

With regard to the extension of Problem I, it is assumed, in a general way, that the ability of the zones to protect certain species – if these zones are protected from the beginning of the considered horizon – depends both on the scenario envisaged and on the period considered. Denote by Z_{kt}^{ω} the set of zones allowing the protection of species s_k during period T_t in the case of scenario sc_ω. In other words, in order to ensure the survival of species s_k during period T_t, if scenario sc_ω occurs, it is necessary and sufficient that at least one of the zones of Z_{kt}^{ω} be protected – from the beginning of the considered horizon. These sets are assumed to be known for any triplet $(k,t,\omega) \in \underline{S} \times \underline{T} \times \underline{\mathrm{Sc}}$. The corresponding set of indices is denoted by $\underline{Z}_{kt}^{\omega}$.

As in the case where climate change and its consequences are assumed to be known with certainty (section 12.3.1), the problem of determining a set of zones to be protected with a cost less than or equal to a given value and "optimal" with regard to the conservation of the species under consideration is considered here. The question then arises: what is an optimal reserve? For this Problem I, several objectives can be considered, as this is done in chapter 8. For example, a very conservative strategy can be adopted by seeking to identify a reserve that maximizes, within an available budget, the number of species that survive at the end of the considered horizon, in the worst-case scenario. The survival of this number of species is then guaranteed regardless of the scenario that occurs.

For Problem I, a species is assumed to survive at the end of the time horizon under consideration and within a given scenario if, at each period of that horizon, at least one of the protected zones is able to protect that species in that scenario. On the other hand, all the protection decisions and their implementation are made at the beginning of the horizon considered.

Mathematical programming formulation. As before, we use the Boolean variable $x_i,\ \ i \in \underline{Z}$, which takes the value 1 if and only if zone z_i is protected at the beginning of the horizon under consideration – and thus throughout this horizon – and the Boolean variable $y_k^{\omega},\ \ k \in \underline{S}, \omega \in \underline{\mathrm{Sc}}$, which takes the value 1 if and only if species s_k survives at the end of the horizon under consideration, when scenario sc_ω occurs. We obtain program $\mathrm{P}_{12.10}$.

$$
\mathrm{P}_{12.10}: \begin{cases} \max \quad \alpha \\ \text{s.t.} \left| \begin{array}{lll} \alpha \leq \sum\limits_{k \in \underline{S}} y_k^{\omega} & \omega \in \underline{\mathrm{Sc}} & (12.10.1) \\ y_k^{\omega} \leq \sum\limits_{i \in \underline{Z}_{kt}^{\omega}} x_i & k \in \underline{S}, \omega \in \underline{\mathrm{Sc}}, t \in \underline{T} & (12.10.2) \\ \sum\limits_{i \in \underline{Z}} c_i x_i \leq B & & (12.10.3) \\ x_i \in \{0,1\} & i \in \underline{Z} & (12.10.4) \\ y_k^{\omega} \in \{0,1\} & k \in \underline{S}, \omega \in \underline{\mathrm{Sc}} & (12.10.5) \\ \alpha \in \mathbb{N} & & (12.10.6) \end{array} \right. \end{cases}
$$

Let us examine constraints 12.10.2. Because of the economic function, α, to be maximized and constraints 12.10.1, variable y_k^{ω} takes, at the optimum of $\mathrm{P}_{12.10}$, the largest possible value. For a given species and for a given scenario, this variable takes the value 1 if and only if, in each period, at least one of the zones of the reserve protects species s_k, and the value 0 if not. We therefore have, at the optimum, $y_k^{\omega} = 1$

if and only if the reserve allows the protection of species s_k in the case of scenario sc_ω. The quantity $\sum_{k \in \underline{S}} y_k^\omega$, which appears in the second members of constraints 12.10.1, thus expresses the number of species protected by the reserve in the case of scenario sc_ω. Because of constraints 12.10.1 and since we are trying to maximize variable α, the value of this variable, at the optimum, is equal to the number of protected species, in the worst-case scenario, *i.e.*, the one corresponding to the smallest number of protected species. Constraint 12.10.3 expresses the budget constraint and constraints 12.10.4–12.10.6 specify the nature of the variables.

12.4.2 Problem II

With regard to the extension of Problem II, it is assumed that the number of species in each protected zone evolves over time and according to the scenario. It is also assumed that this evolution is known, for each scenario, at the beginning of the considered horizon. The population size of species s_k in zone z_i – protected from the beginning of the time horizon – during the period T_t in the case of scenario sc_ω is denoted by n_{ikt}^ω, $(i, k, t, \omega) \in \underline{Z} \times \underline{S} \times \underline{T} \times \underline{\mathrm{Sc}}$. Species s_k is assumed to survive in a given reserve, R, during the period T_t and in the case of scenario sc_ω if its total population size in that reserve, in that period and in that scenario is greater than or equal to a threshold value, θ_k^ω. It is assumed that this threshold value does not depend on t. The aim is to identify a set of zones to be protected, with a cost less than or equal to a given value, and "optimal" with regard to the conservation of the species under consideration. Here again, a very conservative strategy can be adopted by seeking to identify a reserve that maximizes, taking into account an available budget, the number of species that survive at the end of the considered horizon, regardless of the scenario that occurs. A species is assumed to survive at the end of the considered horizon and in the considered scenario if, in each period of that horizon, the size of its population in the reserve, in the considered scenario, is greater than the threshold value.

Mathematical programming formulation. As before, we use the Boolean variable x_i, $i \in \underline{Z}$, which takes the value 1 if and only if zone z_i is protected at the beginning of the considered horizon and thus throughout this horizon and the Boolean variable y_k^ω, $k \in \underline{S}, \omega \in \underline{\mathrm{Sc}}$, which takes the value 1 if and only if species s_k survives at the end of the considered horizon, when scenario sc_ω occurs. To formulate Problem II, it is sufficient to replace in program $\mathrm{P}_{12.10}$ constraints 12.10.2 by constraints $\theta_k^\omega y_k^\omega \le \sum_{i \in \underline{Z}} n_{ikt}^\omega x_i$, $k \in \underline{S}, \omega \in \underline{\mathrm{Sc}}, t \in \underline{T}$. Indeed, according to this constraint, for a given species, s_k, and for a given scenario, sc_ω, variable y_k^ω can only take the value 1 if at each period, the population size of species s_k in the reserve, $\sum_{i \in \underline{Z}} n_{ikt}^\omega x_i$, is at least equal to θ_k^ω. It therefore takes the value 0 otherwise. We thus have, at the optimum, $y_k^\omega = 1$ if and only if the reserve protects species s_k in the case of scenario sc_ω.

12.4.3 Problem III

As regards Problem III adapted to take climate change into account with different scenarios, it is assumed that for all the periods of the considered horizon, T, and in each scenario sc_ω, the habitat area of zone z_i – protected from the beginning of the considered horizon – favourable to species s_k is known. This area is denoted by a^{ω}_{ikt}, $i \in \underline{Z}, k \in \underline{S}, t \in \underline{T}, \omega \in \underline{\mathrm{Sc}}$. As in the previous two sections, the problem is to identify a set of zones to be protected with a cost less than or equal to a given value and "optimal" with regard to the conservation of the species under consideration. Here again, a very conservative strategy can be adopted by seeking to identify a reserve that ensures that for each of the species under consideration, at each period of the horizon, and whatever the scenario that occurs, a total area of favourable habitat – included in the reserve – is as close as possible to a target value. We assume that this target value does not depend on either the time period or the scenario, but it is not mandatory. The target value for species s_k is denoted by $\min_k$, $k \in \underline{S}$.

Mathematical programming formulation. We use as previously the Boolean variable x_i, $i \in \underline{Z}$, which takes the value 1 if and only if zone z_i is protected at the beginning of the horizon considered and thus throughout this horizon and the real and positive or zero variable g^{ω}_{kt}, $k \in \underline{S}$, $t \in \underline{T}$, $\omega \in \underline{\mathrm{Sc}}$ that expresses the gap between the total habitat area of the reserve favourable to species s_k, in scenario sc_ω over the period T_t, and the target value for that species, $\min_k$, a time- and scenario-independent value. This gap is only considered if the total area of habitat in the reserve favourable to species s_k, in scenario sc_ω at time T_t, is less than $\min_k$. In other words, $g^{\omega}_{kt} = \max\{0, (\min_k - \sum_{i\in\underline{Z}} a^{\omega}_{ikt} x_i)\}$. This gives $\mathrm{P}_{12.11}$.

$$
\mathrm{P}_{12.11}: \begin{cases} \min \quad \alpha \\ \text{s.t.} \left| \begin{array}{lll} \alpha \geq \sum\limits_{k\in\underline{S},\, t\in\underline{T}} \xi_k g^{\omega}_{kt} & \omega \in \underline{\mathrm{Sc}} & (12.11.1) \\ \sum\limits_{i\in\underline{Z}} a^{\omega}_{ikt} x_i + g^{\omega}_{kt} \geq \min_k & (k,\omega,t) \in \underline{S} \times \underline{\mathrm{Sc}} \times \underline{T} & (12.11.2) \\ \sum\limits_{i\in\underline{Z}} c_i x_i \leq B & & (12.11.3) \\ x_i \in \{0,1\} & i \in \underline{Z} & (12.11.4) \\ g^{\omega}_{kt} \geq 0 & (k,\omega,t) \in \underline{S} \times \underline{\mathrm{Sc}} \times \underline{T} & (12.11.5) \\ \alpha \geq 0 & & (12.11.6) \end{array} \right. \end{cases}
$$

Let us examine constraints 12.11.2. The quantity $\sum_{i\in\underline{Z}} a^{\omega}_{ikt} x_i$ expresses the total area of the reserve favourable to species s_k at the period T_t in the case of scenario sc_ω. Because of constraint 12.11.1 and the attempt to minimize variable α, variable g^{ω}_{kt} takes, at the optimum, the smallest possible value, *i.e.*, the value 0 if the quantity is greater than or equal to $\min_k$ and the value $(\min_k - \sum_{i\in\underline{Z}} a^{\omega}_{ikt} x_i)$ otherwise. The quantity $\sum_{k\in\underline{S},\, t\in\underline{T}} \xi_k g^{\omega}_{kt}$ thus expresses well, in the case of scenario sc_ω, the weighted sum, for all the species and for all the periods, of the area deficit. Variable α thus takes at the optimum, the value of the overall deficit in the worst-case scenario, *i.e.*, the one that maximizes this deficit. Constraint 12.11.3 is the budgetary constraint. Constraint 12.11.4 requires variable x_i, $i \in \underline{Z}$, to be Boolean, and constraint 12.11.5 specifies that variables g_{kt}^{ω}, $k \in \underline{S}, t \in \underline{T}, \omega \in \underline{\text{Sc}}$, are positive or zero real variables.

12.5 Reserve Minimizing, Under a Budgetary Constraint, the Relative Regret Associated with Problem III in the Worst-Case Scenario of Climate Change

We have just seen how to determine robust solutions for Problems I, II and III. In each of these problems, a value, $\text{Val}^{\omega}(R)$, is associated with the set, R, of selected zones – the reserve – in the case of scenario sc_{ω}, and a robust reserve corresponds to a reserve that takes on the best value in the worst-case scenario. As discussed in other chapters, this objective can have a significant disadvantage: if one of the scenarios is very "pessimistic" then the selection of an optimal reserve will essentially consider that particular scenario. As in chapter 8, other robustness criteria can be considered. For example, one can seek to determine a reserve – under a budget constraint – that minimizes the largest relative gap – over all the scenarios – between the value of the obtained reserve and the value of the optimal reserve in the scenario under consideration. Let us apply this approach to Problem III. In this case $\text{Val}^{\omega}(R)$ represents the overall area deficit – species-weighted – associated with a reserve, R, if scenario sc_{ω} occurs. As we saw in section 12.4.3, $\text{Val}^{\omega}(R) = \sum_{k\in\underline{S},\, t\in\underline{T}} \xi_k g_{kt}^{\omega}$ where $g_{kt}^{\omega} = \max\{0, (\min_k - \sum_{i\in\underline{Z}} a_{ikt}^{\omega} x_i)\}$. The optimization problem considered can be written $\min_{R\subseteq Z, C(R)\leq B}\big\{\max_{\omega\in\underline{\text{Sc}}}[(\text{Val}^{\omega}(R) - \text{Val}^{\omega}(R^{*\omega}))/\text{Val}^{\omega}(R^{*\omega})]\big\}$ where $R^{*\omega}$ is the most interesting reserve for scenario sc_{ω}, *i.e.*, reserve R that minimizes the quantity $\text{Val}^{\omega}(R)$. In order to solve the problem under consideration, we must first calculate $\text{Val}^{\omega}(R^{*\omega})$ for all $\omega \in \underline{\text{Sc}}$. This value, which we denote by $\text{Val}^{*\omega}$ for simplicity, corresponds, for scenario sc_{ω}, to an optimal solution of the mathematical program $\text{P}_{12.12}(\omega)$.

$$
\text{P}_{12.12}(\omega):\left\{\begin{array}{lll}
\min \displaystyle\sum_{k\in\underline{S},\ t\in\underline{T}} \xi_k g_{kt}^{\omega} & & \\
\text{s.t.}\left|\begin{array}{l}
\displaystyle\sum_{i\in\underline{Z}} c_i x_i \leq B \\
\displaystyle\sum_{i\in\underline{Z}} a_{ikt}^{\omega} x_i + g_{kt}^{\omega} \geq \min_k \\
g_{kt}^{\omega} \geq 0 \\
x_i \in \{0,1\}
\end{array}\right. &
\begin{array}{l}
\\
k\in\underline{S}, t\in\underline{T} \\
k\in\underline{S}, t\in\underline{T} \\
i\in\underline{Z}
\end{array} &
\begin{array}{l}
(12.12_{\omega}.1) \\
(12.12_{\omega}.2) \\
(12.12_{\omega}.3) \\
(12.12_{\omega}.4)
\end{array}
\end{array}\right.
$$

Finally, the problem under consideration can be solved by the mathematical program $P_{12.13}$ in which variable Val^{ω} represents the quantity $\mathrm{Val}^{\omega}(R)$ for reserve R retained, *i.e.*, for the reserve formed by zones z_i such that $x_i = 1$.

$$
P_{12.13}: \begin{cases} \min \alpha \\ \text{s.t.} \left| \begin{array}{lll} \sum_{i \in \underline{Z}} c_i x_i \leq B & & (12.13.1) \\ \sum_{i \in \underline{Z}} a_{ikt}^{\omega} x_i + g_{kt}^{\omega} \geq \min_k & (k,\omega,t) \in \underline{S} \times \underline{\mathrm{Sc}} \times \underline{T} & (12.13.2) \\ \mathrm{Val}^{\omega} = \sum_{k \in \underline{S},\ t \in \underline{T}} \xi_k g_{kt}^{\omega} & \omega \in \underline{\mathrm{Sc}} & (12.13.3) \\ \alpha \geq \frac{\mathrm{Val}^{\omega} - \mathrm{Val}^{*\omega}}{\mathrm{Val}^{*\omega}} & \omega \in \underline{\mathrm{Sc}} & (12.13.4) \\ g_{kt}^{\omega} \geq 0 & (k,\omega,t) \in \underline{S} \times \underline{\mathrm{Sc}} \times \underline{T} & (12.13.5) \\ x_i \in \{0,1\} & i \in \underline{Z} & (12.13.6) \\ \mathrm{Val}^{\omega} \geq 0 & \omega \in \underline{\mathrm{Sc}} & (12.13.7) \\ \alpha \geq 0 & & (12.13.8) \end{array} \right. \end{cases}
$$

Constraint 12.13.1 expresses the budgetary constraint. Constraint 12.13.2, associated with constraint 12.13.5, expresses, for each species, for each period and for each scenario, the area deficit associated with the selected reserve, which is intended to be minimized. Constraints 12.13.3 express the global area deficit associated with the selected reserve in each scenario, Val^{ω}, which lightens the writing of constraints 12.13.4. Because of the economic function, α, to be minimized and constraints 12.13.4, variable α takes, at the optimum of $P_{12.13}$, the largest of the values $(\mathrm{Val}^{\omega} - \mathrm{Val}^{*\omega})/\mathrm{Val}^{*\omega}$ over all scenarios sc_{ω}. Finally, constraints 12.13.5–12.13.8 specify the nature of the different variables. The resolution of $P_{12.13}$ therefore allows the selection of zones whose protection minimizes the largest relative gap, over all the scenarios, between the existing global area deficit taking into account the selected zones – zone z_i is selected if $x_i = 1$ – and the minimal global area deficit that could have been obtained in the considered scenario possibly selecting another set of zones.

12.6 Taking into Account Climate Change Described by Several Scenarios Each with a Probability, in Problems I, II and III; Mathematical Expectation Criterion

As in sections 12.4 and 12.5, we consider in this section that there are several possible scenarios for climate change, but it is further assumed that a probability, p^{ω}, can be assigned to the occurrence of each scenario sc_{ω}, $\omega \in \underline{\mathrm{Sc}}$. Problems I, II and

III will be reconsidered in this framework using this time the mathematical expectation criterion and not a robustness criterion as previously.

12.6.1 Problem I

In this new framework, Problem I consists in determining a set of zones to be protected – a reserve – from the beginning of the horizon under consideration that takes into account an available budget and maximizes the expected number of protected species. Recall that we consider here that species s_k is protected in the case of scenario sc_ω if and only if, at each period of the horizon considered, at least one of the zones of the reserve protects s_k, *i.e.*, at least one of the zones of Z^ω_{kt} belongs to the reserve. This problem can be formulated as the mathematical program $\mathrm{P}_{12.14}$.

$$
\mathrm{P}_{12.14}: \begin{cases} \max \quad \sum\limits_{\omega\in\underline{\mathrm{Sc}}} p^\omega \sum\limits_{k\in\underline{S}} y_k^\omega \\ \text{s.t.} \left| \begin{array}{lll} y_k^\omega \le \sum\limits_{i\in\underline{Z}^\omega_{kt}} x_i & k\in\underline{S}, \omega\in\underline{\mathrm{Sc}}, t\in\underline{T} & (12.14.1) \\ \sum\limits_{i\in\underline{Z}} c_i x_i \le B & & (12.14.2) \\ x_i\in\{0,1\} & i\in\underline{Z} & (12.14.3) \\ y_k^\omega \in\{0,1\} & k\in\underline{S}, \omega\in\underline{\mathrm{Sc}} & (12.14.4) \end{array}\right. \end{cases}
$$

Constraints of $\mathrm{P}_{12.14}$ are all already present in $\mathrm{P}_{12.10}$. The expression $\sum_{k\in\underline{S}} y_k^\omega$ corresponds to the number of protected species in the case of scenario sc_ω. The economic function of $\mathrm{P}_{12.14}$ therefore expresses the expected number of protected species, since p^ω is the probability that scenario sc_ω will occur.

12.6.2 Problem II

Like Problem I, Problem II consists in determining a set of zones to be protected – a reserve – at the beginning of the considered horizon that takes into account an available budget and maximizes the expected number of protected species. The only difference is that in Problem II a reserve R is considered as protecting species s_k, $k\in\underline{S}$, if and only if the total population size of that species in the reserve, at each period, is greater than or equal to a threshold value, θ_k. In order to formulate this problem, it is sufficient to replace, in programme $\mathrm{P}_{12.14}$, constraints 12.14.1 by constraints $\theta_k y_k^\omega \le \sum_{i\in\underline{Z}^\omega_{kt}} n^\omega_{ikt} x_i$, $k\in\underline{S}, \omega\in\underline{\mathrm{Sc}}, t\in\underline{T}$. Recall that n^ω_{ikt} is the population size of species s_k in zone z_i – protected from the beginning of the time horizon – during the period T_t and in the case of scenario sc_ω.

12.6.3 Problem III

The problem is to select a set of zones, R, to be protected from the beginning of the considered horizon, with a cost less than or equal to B and such that the expected sum, for all the species and for all the periods, of the species-weighted deficits of the area of R favourable to the species considered in relation to the desired area for the same species is minimal. This optimization problem can be formulated as follows: $\min_{R \subseteq Z,\, C(R) \le B} \sum_{\omega \in \underline{\mathrm{Sc}}} p^{\omega} \sum_{k \in \underline{S},\, t \in \underline{T}} \xi g_{kt}^{\omega}(R)$ where $g_{kt}^{\omega}(R)$ is the above-mentioned deficit for species s_k, at the period T_t and in the case of scenario sc_{ω}. This problem can be formulated as the mathematical program $\mathrm{P}_{12.15}$.

$$\mathrm{P}_{12.15}: \begin{cases} \min & \sum_{\omega \in \underline{\mathrm{Sc}}} p^{\omega} \sum_{k \in \underline{S},\, t \in \underline{T}} \xi_k g_{kt}^{\omega} \\ \text{s.t.} & \left| \begin{array}{lll} \sum_{i \in \underline{Z}} a_{ikt}^{\omega} x_i + g_{kt}^{\omega} \ge \min_k & k \in \underline{S}, \omega \in \underline{\mathrm{Sc}}, t \in \underline{T} & (12.15.1) \\ \sum_{i \in \underline{Z}} c_i x_i \le B & & (12.15.2) \\ x_i \in \{0,1\} & i \in \underline{Z} & (12.15.3) \\ g_{kt}^{\omega} \ge 0 & k \in \underline{S}, \omega \in \underline{\mathrm{Sc}}, t \in \underline{T} & (12.15.4) \end{array} \right. \end{cases}$$

Recall that a_{ikt}^{ω}, $i \in \underline{Z}, k \in \underline{S}, t \in \underline{T}, \omega \in \underline{\mathrm{Sc}}$ is the habitat area of zone z_i – protected from the beginning of the considered horizon – favourable to species s_k during the period T_t and in the case of scenario sc_{ω}. Constraints of $\mathrm{P}_{12.15}$ are all already present in $\mathrm{P}_{12.11}$. Since p^{ω} is the probability that scenario sc_{ω} will occur, the economic function of $\mathrm{P}_{12.15}$ expresses the expected sum, for all the species and for all the periods, of the species-weighted deficits in the area of R favourable to the considered species compared to the desired area for this species.

12.7 Protected Zones and Carbon Sinks

It is widely recognized that protected zones have an important role to play in trying to mitigate climate change. They reduce greenhouse gas emissions by capturing carbon from the atmosphere and protecting the existing carbon stocks. However, the effective management of these zones (*e.g.*, reforestation, forest management) is necessary for them to fulfil their role. For example, degraded forests may contain much less carbon than intact forests. Of course, protected zones are not a complete solution; they are not a substitute for efforts to reduce emissions at source, which are mainly caused by the burning of oil, coal and gas and by deforestation. This section focuses on the definition of protected zones taking into account two aspects simultaneously: (1) species protection and (2) carbon capture and sequestration. Indeed, addressing climate change mitigation must not overshadow the direct protection of biodiversity.

To illustrate this issue simply, let us take up Problem I defined in section 12.2 and briefly recalled here: determine a set of zones to be protected, taking into account an available budget, in order to protect the greatest possible number of species of a given set. Reserve R protects species s_k if and only if that species is present in at least one zone of R and the species present in each of the candidate zones are known. To characterize the quality of a reserve, two additional data will be considered for each protected zone: the amount of carbon stored and the amount of carbon sequestered each year. Some zones (*e.g.*, primary tropical forests, mangroves, peatlands) are more efficient than others with respect to these two quantities. A management horizon of r years is considered. To simplify the presentation, it is assumed that the unprotected zones do not store or sequester carbon. So we have a two-criterion problem because a reserve will be characterized both by the number of species it protects and the amount of carbon it captures and stores. We denote by q_i the amount of carbon stored in zone z_i and by ρ_i the amount of carbon captured and stored by zone z_i each year. We assume here, to simplify the presentation, that these two quantities do not depend on the period, but it would be easy to adapt what follows to the opposite case. Let us recall that Z_k designates the set of zones hosting species s_k and $\underline{Z}_k$, the set of corresponding indices. With each selected zone z_i is associated a cost, c_i, reflecting the acquisition and management of this zone with the aim, on the one hand, of protecting the species and, on the other hand, of capturing and storing carbon. This can be formulated as $\mathrm{P}_{12.16}$.

$$
\mathrm{P}_{12.16}:\begin{cases}
\max\left\{\sum\limits_{k\in\underline{S}} y_k, \sum\limits_{i\in\underline{Z}}(q_i+r\rho_i)x_i\right\}\\
\text{s.t.}\left|\begin{array}{llll}
y_k \le \sum\limits_{i\in\underline{Z}_k} x_i \quad k\in\underline{S} & (12.16.1) & \mid \quad x_i\in\{0,1\} \quad i\in\underline{Z} & (12.16.3)\\
\sum\limits_{i\in\underline{Z}} c_i x_i \le B & (12.16.2) & \mid \quad y_k\in\{0,1\} \quad k\in\underline{S} & (12.16.4)
\end{array}\right.
\end{cases}
$$

This program is identical to program $\mathrm{P}_{12.1}$ except for the objective, which now includes 2 criteria: the number of protected species, $\sum_{k\in\underline{S}} y_k$, and the amount of carbon stored over the management horizon, $\sum_{i\in\underline{Z}}(q_i+r\rho_i)x_i$. One way to deal with the problem is to set the available budget, B, and the number of species to be protected, Ns, and then determine a reserve – if one exists – that maximizes the amount of carbon stored over the management horizon. The goal is to maximize the quantity $\sum_{i\in\underline{Z}}(q_i+r\rho_i)x_i$ under the same constraints 12.16.1–12.16.4 plus constraint $\sum_{k\in\underline{S}} y_k \ge \mathrm{Ns}$. In fact, we will consider the economic function $\sum_{i\in\underline{Z}}(q_i+r\rho_i)x_i+\varepsilon\sum_{k\in\underline{S}} y_k$ where ε is a sufficiently small coefficient. This change in the economic function provides, among the reserves that respect the budget and maximize the amount of carbon they store while protecting at least Ns species, those that maximize the number of species they protect. The solutions obtained thus offer the decision-maker with a given budget several trade-offs between the number of species protected and the amount of carbon stored.

Example 12.3. Consider the instance described in figure 12.5. It includes 20 candidate zones and concerns 15 species.

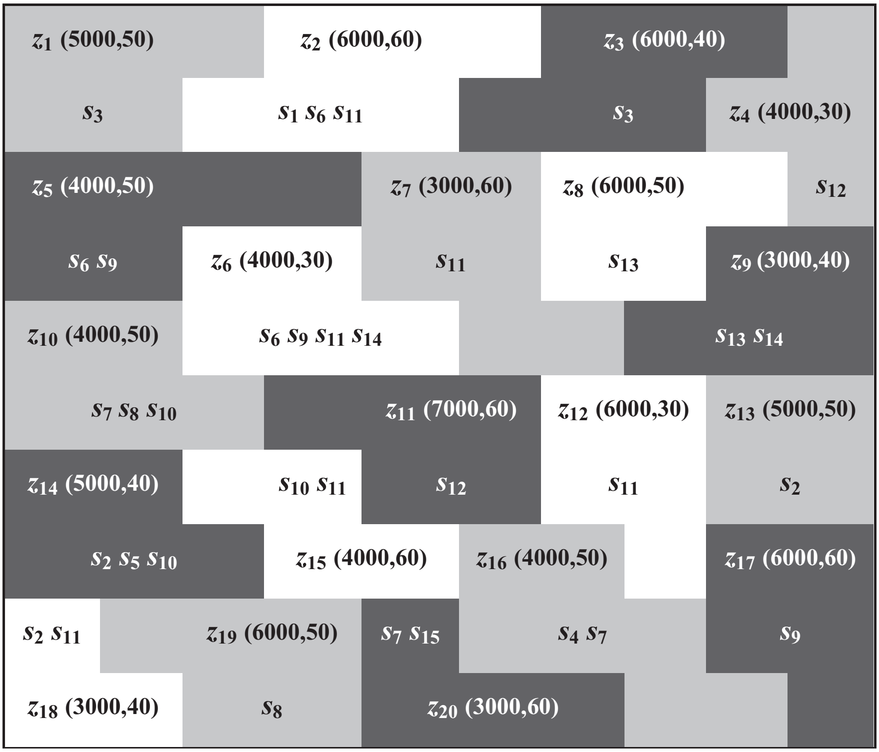


FIG. 12.5 – Twenty zones, z_1, z_2,..., z_{20}, are candidates for protection and fifteen species, s_1, s_2,..., s_{15}, living in these zones are concerned. For each zone, the species present are indicated, as well as the amount of carbon stored followed by the amount of carbon sequestered each year, all in brackets. The cost of protecting the white zones is equal to 1 unit, the cost of protecting the light grey zones is equal to 2 units and the cost of protecting the dark grey zones is equal to 4 units. For example, species s_6, s_9, s_{11}, and s_{14} are present in zone z_6, the amount of carbon stored in this zone is equal to 4,000 tonnes and the amount of carbon sequestered each year is equal to 30 tonnes; the cost of protecting this zone is equal to 1 unit.

Each zone z_i presents the amount of carbon stored, q_i, the amount of carbon sequestered each year, ρ_i, and the protected species, all in the case where this zone is protected. It is assumed here that unprotected zones are not involved in species protection, neither in carbon storage and capture. The problem is to determine a reserve, with a cost less than or equal to the budget, B, that allows for the protection of a number of species at least equal to Ns and that maximizes the amount of carbon stored at the end of the management horizon considered – 20 years in this example. The results are presented in table 12.3 for different values of the parameters B and Ns.

If one has a budget of 4 units and wishes to protect at least 5 species, the optimal reserve costs 4 units, protects 5 species and stores 26,000 tonnes of carbon at the end

Tab. 12.3 – Optimal reserves associated with the instance described in figure 12.5 for different values of the available budget, B, and the minimal number of species to be protected, Ns.

Budget (B)	Minimal number of species to be protected (Ns)	Used budget	Optimal reserve	Number of protected species	Protected species	Amount of carbon
4	5	4	z_2 z_8 z_{12} z_{15}	5	s_1 s_6 s_{10} s_{11} s_{13}	26,000
	10	–	–	–	–	–
	15	–	–	–	–	–
8	5	8	z_2 z_6 z_8 z_{12} z_{15} z_{18} z_{19}	9	s_1 s_2 s_6 s_8 s_9 s_{10} s_{11} s_{13} s_{14}	41,400
	10	8	z_2 z_6 z_8 z_{12} z_{15} z_{16} z_{18}	10	s_1 s_2 s_4 s_6 s_7 s_9 s_{10} s_{11} s_{13} s_{14}	39,400
	15	–	–	–	–	–
12	5	12	z_1 z_2 z_6 z_8 z_{12} z_{13} z_{15} z_{18} z_{19}	10	s_1 s_2 s_3 s_6 s_8 s_9 s_{10} s_{11} s_{13} s_{14}	53,400
	10	12	z_1 z_2 z_6 z_8 z_{12} z_{13} z_{15} z_{18} z_{19}	10	s_1 s_2 s_3 s_6 s_8 s_9 s_{10} s_{11} s_{13} s_{14}	53,400
	15	–	–	–	–	–
16	5	16	z_1 z_2 z_6 z_8 z_{10} z_{12} z_{13} z_{15} z_{16} z_{18} z_{19}	12	s_1 s_2 s_3 s_4 s_6 s_7 s_8 s_9 s_{10} s_{11} s_{13} s_{14}	63,400
	10	16	z_1 z_2 z_6 z_8 z_{10} z_{12} z_{13} z_{15} z_{16} z_{18} z_{19}	12	s_1 s_2 s_3 s_4 s_6 s_7 s_8 s_9 s_{10} s_{11} s_{13} s_{14}	63,400
	15	–	–	–	–	–
20	5	20	z_1 z_2 z_4 z_6 z_7 z_8 z_{10} z_{12} z_{13} z_{15} z_{16} z_{18} z_{19}	13	s_1 s_2 s_3 s_4 s_6 s_7 s_8 s_9 s_{10} s_{11} s_{12} s_{13} s_{14}	72,200
	10	20	z_1 z_2 z_4 z_6 z_7 z_8 z_{10} z_{12} z_{13} z_{15} z_{16} z_{18} z_{19}	13	s_1 s_2 s_3 s_4 s_6 s_7 s_8 s_9 s_{10} s_{11} s_{12} s_{13} s_{14}	72,200
	15	20	z_1 z_2 z_4 z_6 z_8 z_{12} z_{14} z_{16} z_{19} z_{20}	15	s_1 s_2 s_3 s_4 s_5 s_6 s_7 s_8 s_9 s_{10} s_{11} s_{12} s_{13} s_{14} s_{15}	58,000

–: no feasible solution.

of the considered horizon. If one has a budget of 8 units and wishes to protect at least 5 species, the optimal reserve costs 8 units, protects 9 species and stores 41,400 tonnes of carbon at the end of the horizon. Note that, in this case, the number of protected species is greater than the minimal number of species to be protected.

References and Further Reading

Abrahms B., DiPietro D., Graffis A. *et al.* (2017) Managing biodiversity under climate change: challenges, frameworks, and tools for adaptation, *Biodivers. Conserv.* **26**, 2277.

ADEME (2015) Le changement climatique, comprendre ses causes et ses conséquences pour mieux réagir, Guide de l'Agence De l'Environnement et de la Maîtrise de l'Énergie (ADEME).

Alagador D., Cerdeira J.O., Araújo M.B. (2014) Shifting protected areas: scheduling spatial priorities under climate change, *J. Appl. Ecol.* **51**, 703.

Alagador D., Cerdeira J.O., Araújo M.B. (2016) Climate change, species range shifts and dispersal corridors: an evaluation of spatial conservation models, *Methods Ecol. Evol.* **7**, 853.

Albert C.H., Rayfield B., Dumitru M., Gonzalez A. (2017) Applying network theory to prioritize multispecies habitat networks that are robust to climate and land-use change, *Conserv. Biol.* **31**, 1383.

Araújo M.B., Cabeza M., Moilanen A.M., Thuiller W., Hannah L., Williams P.H. (2004) Would climate change drive species out of reserves? an assessment of existing reserve-selection methods, *Global Change Biol.* **10**, 1618.

Araújo M.B. (2009) Climate change and spatial conservation planning, *Spatial conservation prioritization: quantitative methods and computational tools* (A. Moilanen, K. Wilson, H. Possingham, Eds). Oxford University Press, pp. 172–184.

Beaumont L.J., Hughes L., Pitman A.J. (2008) Why is the choice of future climate scenarios for species distribution modelling important? *Ecol. Lett.* **11**, 1135.

Carroll C., Dunk J.R., Moilanen A. (2010) Optimizing resiliency of reserve networks to climate change: multispecies conservation planning in the Pacific Northwest, USA, *Global Change Biol.* **16**, 891.

Carvalho S.B., Brito J.C., Crespo E.J., Possingham H.P. (2010) From climate change predictions to actions-conserving vulnerable animal groups in hotspots at a regional scale, *Global Change Biol.* **16**, 3257.

Carvalho S.B., Brito J.C., Crespo E.G., Watts M.E., Possingham H.P. (2011) Conservation planning under climate change: toward accounting for uncertainty in predicted species distributions to increase confidence in conservation investments in space and time, *Biol. Conserv.* **144**, 2020.

Dissanayake S.T.M., Önal H., Westervelt J.D., Balbach H.E. (2012) Incorporating species relocation in reserve design models: an example from Ft. Benning GA, *Ecol. Model.* **224**, 65.

Dudley N., Stolton S., Belokurov A., Krueger L., Lopoukhine N., MacKinnon K., Sandwith T., Sekhran N., Eds (2009) *Natural Solutions: protected areas helping people cope with climate change*. IUCN-WCPA, The Nature Conservancy, UNDP, Wildlife Conservation Society, The World Bank and WWF, Gland, Switzerland, Washington DC and New York.

Fuller T., Morton D.P., Sarkar S. (2008) Incorporating uncertainty about species' potential distributions under climate change into the selection of conservation areas with a case study from the Arctic Coastal Plain of Alaska, *Biol. Conserv.* **141**, 1547.

Hannah L., Midgley G., Hughes G., Bomhard B. (2005) The view from the Cape: extinction risk, protected areas, and climate change, *BioSci.* **55**, 231.

Hannah L., Midgley G., Andelman S., Araújo M., Hughes G., Martinez-Meyer E., Pearson R., Williams P. (2007) Protected area needs in a changing climate, *Front. Ecol. Environ.* **5**, 131.

Hannah L. (2008) Protected areas and climate change, *Ann. N.Y. Acad. Sci.* **1134**, 201.

Hodgson J.A., Thomas C.D., Wintle B.A., Moilanen A. (2009) Climate change, connectivity and conservation decision making: back to basics, J. *Appl. Ecol.* **46**, 964.

Kouvelis P., Yu G. (1997) *Robust discrete optimization and its applications.* Kluwer Academic Publishers.

Lemieux C.J., Scott D.J. (2005) Climate change, biodiversity conservation and protected area planning in Canada, *The Can. Geographer* **49**, 384.

Nabuurs G.J., Pussinen A., van Brusselen J., Schelhaas M.J. (2007) Future harvesting pressure on European forests, *Eur. J. For. Res.* **126**, 391.

Peters R.L., Darling J.D.S. (1985) The greenhouse effect and nature reserves: global warming would diminish biological diversity by causing extinctions among reserve species, *BioScience* **35**, 707.

Pyke C.R., Fischer D.T. (2005) Selection of bioclimatically representative biological reserve systems under climate change, *Biol. Conserv.* **121**, 429.

Schniederjans M.J. (1995) *Goal programming: methodology and applications.* Kluwer Academic Publishers.

Strange N., Thorsen B.J., Bladt J. (2006) Optimal reserve selection in a dynamic world, *Biol. Conserv.* **131**, 33.

Strange N., Thorsen B.J., Bladt J. (2011) Conservation policies and planning under climate change, *Biol. Conserv.* **144**, 2968.

Téllez-Valdés O., Dávila-Aranda P., Lira-Saade R. (2006) The effects of climate change on the long-term conservation of Fagus grandifolia var. mexicana, an important species of the cloud forest in Eastern Mexico, *Biodivers. Conserv.* **15**, 1095.

Thomas C.D. *et al.* (2004) Extinction risk from climate change, *Nature* **427**, 145.

Thuiller W. (2004) Patterns and uncertainties of species' range shifts under climate change, *Global Change Biol.* **10**, 2020.

Appendix

A.1 Linear Programming

Consider program $P_{A.1}$ below which consists in minimizing the linear function $f(x_1, x_2, \ldots, x_n)$; the variables of this function, x_1, $x_2, \ldots$, x_n, are either Boolean variables, or integer variables or non-negative real variables. These variables are subject to a set of linear constraints and the coefficients c_1, $c_2, \ldots$, c_n, a_{i1}, $a_{i2}, \ldots$, a_{in} ($i = 1$, $\ldots$, m), b_1, $b_2, \ldots$, b_m are arbitrary.

$$
\mathrm{P}_{\mathrm{A.1}}: \begin{cases} \min \quad f(x_1, x_2, \ldots, x_n) = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n \\ \text{s.t.} \left| \begin{array}{lll} a_{i1} x_1 + a_{i2} x_2 + \cdots + a_{in} x_n \le b_i & i = 1, \ldots, p & \text{(A.1.1)} \\ a_{i1} x_1 + a_{i2} x_2 + \cdots + a_{in} x_n = b_i & i = p+1, \ldots, m & \text{(A.1.2)} \\ x_j \in \{0, 1\} & j = 1, \ldots, r & \text{(A.1.3)} \\ x_j \in \mathbb{N} & j = r+1, \ldots, s & \text{(A.1.4)} \\ x_j \ge 0 & j = s+1, \ldots, n & \text{(A.1.5)} \end{array} \right. \end{cases}
$$

$P_{A.1}$ is a mathematical program, called a mixed-integer linear program; it can be written in the condensed form $P_{A.2}$.

$$
\mathrm{P}_{\mathrm{A.2}}: \begin{cases} \min \quad f(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} c_j x_j \\ \text{s.t.} \left| \begin{array}{lll} \sum_{j=1}^{n} a_{ij} x_j \le b_i & i = 1, \ldots, p & \text{(A.2.1)} \\ \sum_{j=1}^{n} a_{ij} x_j = b_i & i = p+1, \ldots, m & \text{(A.2.2)} \\ x_j \in \{0, 1\} & j = 1, \ldots, r & \text{(A.2.3)} \\ x_j \in \mathbb{N} & j = r+1, \ldots, s & \text{(A.2.4)} \\ x_j \ge 0 & j = s+1, \ldots, n & \text{(A.2.5)} \end{array} \right. \end{cases}
$$

Problem $\mathrm{P}_{\mathrm{A.1}}$ – or $\mathrm{P}_{\mathrm{A.2}}$ – consists in determining the values of variables x_1, x_2,…, x_n which respect their specificity, defined by constraints A.2.3–A.2.5, satisfy linear constraints A.2.1 and A.2.2, and minimize the linear economic function $\sum_{j=1}^{n} c_j x_j$. The linear constraints are either inequalities, $\sum_{j=1}^{n} a_{ij}x_j \leq b_i,\ i = 1, 2,\ldots, p,$ or equalities, $\sum_{j=1}^{n} a_{ij}x_j = b_i,\ i = p+1,\ldots, m$. As regards the specificity of the variables in program $\mathrm{P}_{\mathrm{A.1}}$, some of them can only take integer values; this is the case of variables $x_j, j = 1,\ldots, s$. Others can take any non-negative real values; this is the case of variables $x_j,\ j = s+1,\ldots, n$. Among the variables that can only take integer values, some can only take the values 0 or 1; this is the case of variables $x_j\,, j = 1,\ldots, r$. If all the variables of $\mathrm{P}_{\mathrm{A.1}}$ must only take integer values, we have an integer linear program. If all its variables must only take the values 0 or 1, it is a 0–1 linear program or a linear program in Boolean variables. It can always be assumed that all the variables in a mixed-integer linear program are positive or zero. This is because any variable that is not constrained in sign can be expressed as the difference between two non-negative variables. Thus, a real variable can be expressed as the difference between two positive or zero real variables and a variable belonging to the set of integers, as the difference between two variables belonging to the set of natural numbers.

Example A.1. Consider program $\mathrm{P}_{\mathrm{A.3}}$, which consists of minimizing a linear function of the five variables x_1, x_2, x_3, x_4, and x_5, subject to two linear constraints – one equality and one inequality. Variables x_1 and x_2 can only take the values 0 or 1, variable x_3 must take a positive or zero integer value, and both variables x_4 and x_5 must take real, positive or zero values.

$$\mathrm{P}_{\mathrm{A.3}}: \left\{ \begin{array}{ll} \min & f(x_1, x_2, x_3, x_4, x_5) = -x_1 - 3x_2 + 3x_3 - 4x_4 + 7x_5 \\ \text{s.t.} & \left| \begin{array}{ll} 2x_1 - 3x_2 + 3x_3 - 6x_4 - 2x_5 \leq 10 & \text{(A.3.1)} \\ x_1 - 4x_2 + 2x_3 - 5x_4 + 3x_5 = 14 & \text{(A.3.2)} \\ x_1, x_2 \in \{0, 1\} & \text{(A.3.3)} \\ x_3 \in \mathbb{N} & \text{(A.3.4)} \\ x_4, x_5 \geq 0 & \text{(A.3.5)} \end{array} \right. \end{array} \right.$$

Many software packages for solving linear programs are available. Using one of these software packages, the following optimal solution of $\mathrm{P}_{\mathrm{A.3}}$ is obtained: ($x_1 = 1$, $x_2 = 0$, $x_3 = 4$, $x_4 = 0.0714$, $x_5 = 1.7857$). This solution gives the economic function the value 23.2143. A solution that satisfies all the constraints is called a feasible solution. For example, the solution ($x_1 = 0$, $x_2 = 1$, $x_3 = 5$, $x_4 = 0.5$, $x_5 = 3.5$) is feasible but not optimal since it gives the economic function the value 34.5.

Some integer linear programs are easy to solve because any feasible basic solution of their continuous relaxation is an integer solution. A matrix is said to be totally unimodular if the determinants of all its square sub-matrices are 0, 1 or −1. The coefficients of such a matrix can, therefore, only take the values 0, 1 or −1. There are some simple characterizations of totally unimodular matrices. Consider the set of solutions, assumed to be not empty, $\{x \in \mathbb{R}^n : Ax \leq b, x \geq 0\}$. In a general way, if A is totally unimodular and if all the entries of the vector b are integer, then $\{x \in \mathbb{R}^n : Ax \leq b, x \geq 0\}$ is an integral polyhedron, *i.e.*, a polyhedron whose vertices all have integer coordinates. Consider the mathematical program $\mathrm{P}_{\mathrm{A.4}}$ and its continuous relaxation, $\mathrm{P}_{\mathrm{A.5}}$. It is assumed that $\mathrm{P}_{\mathrm{A.4}}$ admits an optimal solution.

$$
\mathrm{P}_{\mathrm{A.4}} : \begin{cases} \min \ \sum_{j=1}^{n} c_j x_j \\ \text{s.t.} \left| \begin{array}{lll} \sum_{j=1}^{n} a_{ij} x_j \leq b_i & i = 1,\ldots, p & \text{(A.4.1)} \\ x_j \in \{0, 1\} & j = 1,\ldots, r & \text{(A.4.2)} \\ x_j \in \mathbb{N} & j = r+1,\ldots, s & \text{(A.4.3)} \end{array} \right. \end{cases}
$$

$$
\mathrm{P}_{\mathrm{A.5}} : \begin{cases} \min \ \sum_{j=1}^{n} c_j x_j \\ \text{s.t.} \left| \begin{array}{lll} \sum_{j=1}^{n} a_{ij} x_j \leq b_i & i = 1,\ldots, p & \text{(A.5.1)} \\ x_j \leq 1 & j = 1,\ldots, r & \text{(A.5.2)} \\ x_j \geq 0 & j = 1,\ldots, s & \text{(A.5.3)} \end{array} \right. \end{cases}
$$

If the matrix associated with constraints A.5.1 and A.5.2 is totally unimodular and if the coefficients b_i, $i = 1,\ldots, p$, are integers, then any basic solution of $\mathrm{P}_{\mathrm{A.5}}$ is integer-valued, and this is the case, in particular, of its optimal solution(s). To solve the integer linear program $\mathrm{P}_{\mathrm{A.4}}$, it is therefore sufficient to solve the continuous linear program $\mathrm{P}_{\mathrm{A.5}}$.

There is an extensive literature on linear programming, a central problem in operational research. A few works, either entirely devoted to linear programming or more general but with parts devoted to linear programming, are mentioned below.

References and Further Reading

Baillargeon G. (1999) *Introduction à la programmation linéaire*. Edition SMG.

Billionnet A. (2017) *Optimisation discrète*. Dunod.

Brezinski C. (2000) *Initiation à la programmation linéaire et à l'algorithme du simplexe*. Ellipses.

Chen D.S., Batson R.G., Dang Y. (2010) *Applied integer programming: modeling and solution*. John Wiley & Sons.

Charon I., Germa A., Hudry O. (1996) *Méthodes d'optimisation combinatoire*. Masson.

Chrétienne P., Grandjean J.C., Pesqueux Y. (1980) *Algorithmes et pratique de la programmation linéaire*. Technip.

Chvatal V. (1983) *Linear Programming*. Freeman.

Dantzig G.B. (1963) *Linear programming and extensions*. Princeton University Press.

De Werra D. (1990) *Éléments de programmation linéaire avec application aux graphes*. Presses Polytechniques Romandes.

Eppen G.D., Gould F.J., Schmidt C.P., Moore J.H., Weatherford L.R. (1998) *Introductory management science*. Prentice Hall.

Faure R., Lemaire B., Picouleau C. (2014) *Précis de recherche opérationnelle*. Dunod.

Guéret C., Prins C., Sevaux M. (2000) *Programmation linéaire avec 65 problèmes d'optimisation modélisés et résolus avec Visual Xpress.* Eyrolles.
Hillier F.S., Lieberman G.J. (2015) *Introduction to operations research.* McGraw-Hill.
Jacquet-Lagrèze E. (1998) *Programmation linéaire, modélisation et mise en œuvre informatique.* P.I.Q poche.
Matousek J., Gärtner B. (2007) *Understanding and using linear programming.* Springer.
Minoux M. (2008) *Programmation mathématique, théorie et algorithmes.* Lavoisier.
Murty K.G. (1983) *Linear programming.* Wiley.
Nering E.D., Tucker A.W. (1993) *Linear programs and related problems.* Academic Press.
Padberg M. (1995) *Linear optimization and extensions.* Springer.
Roos C., Terlaky T., Vial J.-P. (1996) *Linear optimization with interior point methods.* Wiley.
Roseaux (collective of authors) (2003) Exercices et problèmes résolus de recherche opérationnelle (Tome 3), *Programmation linéaire et extensions, problèmes classiques.* Dunod.
Sakarovitch M. (1997) *Optimisation combinatoire.* Hermann.
Salkin H.M., Haas R., Mathur K. (1989) *Foundations of integer programming.* North-Holland.
Schrijver A. (1986) *Theory of linear and integer programming.* Wiley.
Sierksma G. (1996) *Linear and integer programming, theory and practice.* Marcel Dekker.
Simonnard M. (1962) *Programmation linéaire.* Dunod.
Teghem J. (2003) *Programmation linéaire.* Ellipses.
Vajda S. (1981) *Linear programming, algorithms and applications.* Chapman and Hall.
Vanderbei R.J. (1996) *Linear programming: foundations and extensions.* Kluwer Academic Publishers.
Vaserstein L.N. (2003) *Introduction to linear programming*, Prentice Hall.
Williams H.P. (1993) *Model solving in mathematical programming.* John Wiley.

A.2 Quadratic Programming

Linear programming is a powerful tool for formulating and solving a wide variety of optimization problems. However, in some cases, the economic function or constraints associated with the problems of interest do not possess the linearity property. These are referred to as non-linear optimization problems and non-linear mathematical programs. Solving a general non-linear mathematical program is a difficult task. Here we are interested in quadratic programs. Such a program is generally written as $\mathrm{P}_{\mathrm{A.6}}$.

$$
\mathrm{P}_{\mathrm{A.6}}: \begin{cases}
\min \quad q(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} a_j x_j + \sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij} x_i x_j \\
\text{s.t.} \left| \begin{array}{lll}
\sum_{j=1}^{n} e_{kj} x_j + \sum_{i=1}^{n}\sum_{j=1}^{n} c_{kij} x_i x_j \le b_k & k = 1,\ldots,p & \text{(A.6.1)} \\
\sum_{j=1}^{n} e_{kj} x_j + \sum_{i=1}^{n}\sum_{j=1}^{n} c_{kij} x_i x_j = b_k & k = p+1,\ldots,m & \text{(A.6.2)} \\
x_j \in \{0,1\} & j = 1,\ldots,r & \text{(A.6.3)} \\
x_j \in \mathbb{N} & j = r+1,\ldots,s & \text{(A.6.4)} \\
x_j \ge 0 & j = s+1,\ldots,n & \text{(A.6.5)}
\end{array} \right.
\end{cases}
$$

Variables $x_1, x_2, \ldots, x_n$ are the variables of the problem; $a_j\,(j = 1, \ldots, n)$, $q_{ij}\,(i = 1, \ldots, n; j = 1, \ldots, n)$, $e_{kj}\,(k = 1, \ldots, m; j = 1, \ldots, n)$, $c_{kij}\,(k = 1, \ldots, m; i = 1, \ldots, n; j = 1, \ldots, n)$, and $b_k\,(k = 1, \ldots, m)$ are any given coefficients. Constraints A.6.1 are quadratic inequality constraints and constraints A.6.2 are quadratic equality constraints. Solving $P_{A.6}$ is generally difficult, but there are many interesting and much easier special cases. Some of them are discussed below.

A.3 Convex Quadratic Programming

Consider program $P_{A.7}$ that satisfies the following properties: the objective, $q(x) = q(x_1, x_2, \ldots, x_n)$, and the left-hand side of constraint A.7.1, $\sum_{j=1}^{n} e_{kj}x_j + \sum_{i=1}^{n}\sum_{j=1}^{n} c_{kij}x_ix_j$, are convex quadratic functions; the right-hand side of this constraint, b_k, is a positive or zero constant. Program $P_{A.7}$ consists of minimizing a convex function over a convex domain; it is called a convex quadratic program. There are very efficient algorithms to solve it.

$$P_{A.7}: \begin{cases} \min \quad q(x_1, x_2, \ldots, x_n) = \sum\limits_{j=1}^{n} a_jx_j + \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} q_{ij}x_ix_j & \\ \text{s.t.} \left| \begin{array}{lll} \sum\limits_{j=1}^{n} e_{kj}x_j + \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} c_{kij}x_ix_j \le b_k & k = 1, \ldots, m & \text{(A.7.1)} \\ x_j \ge 0 & j = 1, \ldots, n & \text{(A.7.2)} \end{array} \right. \end{cases}$$

Given n real variables, $x_1, x_2, \ldots, x_n$, an expression of the form $q(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_iq_{ij}x_j$ is called a quadratic form. It is written, in matrix form, $q(x) = x^t Q\, x$ where x denotes the vector $(x_1, x_2, \ldots, x_n)$, and Q denotes a symmetric square matrix of dimension $n \times n$. The matrix Q is said to represent the quadratic form $q(x)$. Note that any quadratic form can be represented by one and only one symmetric matrix and that any symmetric matrix represents one and only one quadratic form. The symmetric matrix Q is said to be positive semidefinite if, for all $x \in \mathbb{R}^n$, $x^t Qx \ge 0$. By definition, the quadratic form $q(x) = x^t Qx$ is convex if the symmetric matrix Q is positive semidefinite. There are many characterizations of positive semidefinite matrices.

A special case of $P_{A.7}$, which can be solved very efficiently, consists in minimizing a convex quadratic function subject to linear constraints. It corresponds to program $P_{A.8}$ in which, now, b_k is a coefficient of any sign.

$$P_{A.8}: \begin{cases} \min \quad q(x_1, x_2, \ldots, x_n) = \sum\limits_{j=1}^{n} a_jx_j + \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} q_{ij}x_ix_j & \\ \text{s.t.} \left| \begin{array}{lll} \sum\limits_{j=1}^{n} e_{kj}x_j \le b_k & k = 1, \ldots, m & \text{(A.8.1)} \\ x_j \ge 0 & j = 1, \ldots, n & \text{(A.8.2)} \end{array} \right. \end{cases}$$

Example A.2. Consider the convex quadratic program $P_{A.9}$.

$$P_{A.9}: \begin{cases} \min \quad q(x) = -3x_1 - 2x_2 - 3x_3 - 10x_4 + 4x_5 + 2(x_2-1)^2 + 5(x_3-1)^2 \\ \quad + 2(x_4-1)^2 + (x_5-2)^2 \\ \text{s.t.} \left| \begin{array}{ll} 2x_1 + 3x_2 + 3x_3 + 5x_4 - 2x_5 \le 20 & \text{(A.9.1)} \\ x_1, x_2, x_3, x_4, x_5 \ge 0 & \text{(A.9.2)} \end{array} \right. \end{cases}$$

Many software packages are available to solve $P_{A.9}$. The optimal solution obtained with one of these software packages is: ($x_1 = 5.6$, $x_2 = 0.375$, $x_3 = 0.85$, $x_4 = 1.625$, $x_5 = 1.5$). This solution gives the economic function the value -28.425.

There are many books and articles dealing with convex mathematical programming and, in particular, convex quadratic programming. Some of these publications are mentioned below.

References and Further Reading

Best M.J. (2017) *Quadratic programming with computer programs*. Taylor & Francis.
Fletcher R. (1971) A General quadratic programming algorithm, *IMA J. Appl. Math.* **7**, 76.
Gill P.E., Murray W., Wright M.H. (1986) *Practical optimization*. Elsevier.
Hertog D. den (1994) *Interior point approach to linear, quadratic and convex programming, algorithms and complexity*. Kluwer.
Mitra G. (1976) *Theory and application of mathematical programming*. Academic Press.
Nesterov Y. (2004) *Introductory lectures on convex optimization: basic course*. Springer-Verlag.
Quadri D., Soutil E. (2015) Reformulation and solution approach for non-separable integer quadratic programs, *J. Oper. Res. Soc.* **66**, 1270.
Wolf P. (1959) The simplex method for quadratic programming, *Econometrica* **27**, 382.

A.4 Mixed-Integer Quadratic Programs With Convex Continuous Relaxations

In such programs, the variables are either integer or real. They are generally written as $P_{A.10}$. In this program, the economic function and the left-hand side of constraint A.10.1 are convex quadratic functions, and b_k, $k = 1, 2, \ldots, m$, is a positive or zero constant.

$$P_{A.10}: \begin{cases} \min \quad q(x_1 x_2, \ldots, x_n) = \sum_{j=1}^{n} a_j x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} q_{ij} x_i x_j \\ \text{s.t.} \left| \begin{array}{lll} \sum_{j=1}^{n} e_{kj} x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} c_{kij} x_i x_j \le b_k & k = 1, \ldots, m & \text{(A.10.1)} \\ x_j \in \{0, 1\} & j = 1, \ldots, r & \text{(A.10.2)} \\ x_j \in \mathbb{N} & j = r+1, \ldots, s & \text{(A.10.3)} \\ x_j \ge 0 & j = s+1, \ldots, n & \text{(A.10.4)} \end{array} \right. \end{cases}$$

The continuous relaxation of $P_{A.10}$ is obtained by relaxing the integrality constraints. Specifically, the constraint $x_j \in \{0, 1\}$, $j = 1, \ldots, r$, is replaced by the constraint $0 \leq x_j \leq 1$, $j = 1, \ldots, r$, and the constraint $x_j \in \mathbb{N}$, $j = r+1, \ldots, s$, is replaced by the constraint $x_j \geq 0$, $j = r+1, \ldots, s$. The resulting program is called the continuous relaxation of $\mathrm{P}_{\mathrm{A}.10}$, and it is easy to see that it is a convex quadratic program. There are efficient algorithms for solving mixed-integer quadratic programs with convex continuous relaxation. They are in fact based on implicit enumeration methods that require, at each node of the search tree, the resolution of a continuous relaxation, which in this case can be done efficiently because of convexity properties.

There are also several methods for converting a mixed-integer quadratic program whose continuous relaxation is not convex into a mixed-integer quadratic program whose continuous relaxation is convex. Some examples of these transformations are presented in the following section.

A.5 Quadratic Programming in 0–1 Variables

Quadratic programs in 0–1 variables allow the formulation of a large number of combinatorial optimization problems in various fields. The general form of these programs is given by $\mathrm{P}_{\mathrm{A}.11}$.

$$\mathrm{P}_{\mathrm{A}.11}: \begin{cases} \min \quad q(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} a_j x_j + \sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij} x_i x_j \\ \text{s.t.} \left| \begin{array}{lll} \sum_{j=1}^{n} e_{kj} x_j + \sum_{i=1}^{n}\sum_{j=1}^{n} c_{kij} x_i x_j \leq b_k & k = 1, \ldots, p & \text{(A.11.1)} \\ \sum_{j=1}^{n} e_{kj} x_j + \sum_{i=1}^{n}\sum_{j=1}^{n} c_{kij} x_i x_j = b_k & k = p+1, \ldots, m & \text{(A.11.2)} \\ x_j \in \{0, 1\} & j = 1, \ldots, n & \text{(A.11.3)} \end{array} \right. \end{cases}$$

There are many methods to solve this type of program: linearization methods and convexification methods. The linearization methods consist of transforming $\mathrm{P}_{\mathrm{A}.11}$ into a mixed-integer linear program, using additional variables. The convexification methods consist in transforming $\mathrm{P}_{\mathrm{A}.11}$ into a quadratic problem whose continuous relaxation is convex, possibly using additional variables. Some solvers accept programs $\mathrm{P}_{\mathrm{A}.11}$ directly, and automatically perform a pre-processing – linearization or convexification – that transforms the program into an equivalent one whose continuous relaxation is a linear or a convex quadratic program.

A.5.1 Linearizations

One way to solve $\mathrm{P}_{\mathrm{A}.11}$ is to linearize it and then solve the mixed-integer linear program thus constructed using a mixed-integer linear programming solver. A first linearization method consists in replacing, in the economic function and in the constraints, each product of variables $x_i x_j$ by variable y_{ij}, and in adding to the obtained program constraints which force variable y_{ij} to be equal to product $x_i x_j$. Thus, we obtain program $\mathrm{P}_{\mathrm{A}.12}$ in which IJ designates the set of index pairs $(i, j) \in$

$\{1,\ldots,n\}^2$ such that the product x_ix_j appears either in the economic function or in the constraints of $\mathrm{P}_{\mathrm{A.11}}$.

$$\mathrm{P}_{\mathrm{A.12}}:\begin{cases}\min \quad q(x_1,x_2,\ldots,x_n)=\sum\limits_{j=1}^{n} a_jx_j+\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} q_{ij}y_{ij} \\ \text{s.t.}\left|\begin{array}{lll}\sum\limits_{j=1}^{n} e_{kj}x_j+\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} c_{kij}y_{ij}\le b_k & k=1,\ldots,p & \text{(A.12.1)}\\ \sum\limits_{j=1}^{n} e_{kj}x_j+\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} c_{kij}y_{ij}= b_k & k=p+1,\ldots,m & \text{(A.12.2)}\\ y_{ij}\le x_i;\, y_{ij}\le x_j;\, 1-x_i-x_j+y_{ij}\ge 0 & (i,j)\in \mathrm{IJ} & \text{(A.12.3)}\\ y_{ij}\ge 0 & (i,j)\in \mathrm{IJ} & \text{(A.12.4)}\\ x_j\in\{0,1\} & j=1,\ldots,n & \text{(A.12.5)}\end{array}\right.\end{cases}$$

It is easy to verify, by examining the two possible values of variables x_j, $j=1,\ldots,n$, that any feasible solution of $\mathrm{P}_{\mathrm{A.12}}$ satisfies $y_{ij}=x_ix_j$.

Example A.3. Consider program $\mathrm{P}_{\mathrm{A.13}}$ which consists of minimizing a quadratic function of 5 Boolean variables subject to one linear constraint.

$$\mathrm{P}_{\mathrm{A.13}}:\begin{cases}\min \quad -3x_1-2x_2+3x_3-10x_4+4x_5+2\,x_1x_2\\ -4x_1x_3+5x_2x_3+6x_2x_5-4x_3x_4+6x_3x_5-2x_4x_5\\ \text{s.t.}\left|\begin{array}{ll}2x_1+3x_2+3x_3+5x_4-2x_5\le 20 & \text{(A.13.1)}\\ x_1,x_2,x_3,x_4,x_5\in\{0,1\} & \text{(A.13.2)}\end{array}\right.\end{cases}$$

By applying the linearization shown above, we obtain program $\mathrm{P}_{\mathrm{A.14}}$.

$$\mathrm{P}_{\mathrm{A.14}}:\begin{cases}\min \quad -3x_1-2x_2+3x_3-10x_4+4x_5\\ \qquad +2y_{12}-4y_{13}+5y_{23}+6y_{25}-4y_{34}+6y_{35}-2y_{45}\\ \text{s.t.}\left|\begin{array}{llcll}2x_1+3x_2+3x_3+5x_4-2x_5\le 20 & \text{(A.14.1)} & | & y_{34}\le x_3;\, y_{34}\le x_4 & \text{(A.14.6)}\\ 1-x_1-x_2+y_{12}\ge 0 & \text{(A.14.2)} & | & 1-x_3-x_5+y_{35}\ge 0 & \text{(A.14.7)}\\ y_{13}\le x_1;\, y_{13}\le x_3 & \text{(A.14.3)} & | & y_{45}\le x_4;\, y_{45}\le x_5 & \text{(A.14.8)}\\ 1-x_2-x_3+y_{23}\ge 0 & \text{(A.14.4)} & | & x_1,x_2,x_3,x_4,x_5\in\{0,1\} & \text{(A.14.9)}\\ 1-x_2-x_5+y_{25}\ge 0 & \text{(A.14.5)} & | & y_{12},y_{13},y_{23},y_{25},y_{34},y_{35},y_{45}\ge 0 & \text{(A.14.10)}\end{array}\right.\end{cases}$$

Note that, given the signs of the coefficients of variables y_{ij} in the economic function, some linearization constraints are unnecessary. An optimal solution for $\mathrm{P}_{\mathrm{A.14}}$ is: $(x_1=1,\ x_2=0,\ x_3=1,\ x_4=1,\ x_5=0)$. This solution gives the economic function the value -18.

We now present a second linearization method by applying it to program $\mathrm{P}_{\mathrm{A.15}}$, which consists of minimizing a quadratic economic function whose variables are subject to linear constraints. Here, we assume that the quadratic part of the economic function is written as $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} q_{ij}x_ix_j$. Note that, since $x_i^2=x_i$, we can assume that the economic function does not have any terms x_i^2.

$$
\mathrm{P}_{\mathrm{A.15}}: \begin{cases} \min \quad q(x_1, x_2, \ldots, x_n) = \sum\limits_{j=1}^{n} a_j x_j + \sum\limits_{i=1}^{n-1} \sum\limits_{j=i+1}^{n} q_{ij} x_i x_j \\ \text{s.t.} \left| \begin{array}{lll} \sum\limits_{j=1}^{n} e_{kj} x_j \le b_k & k = 1, \ldots, m & (\mathrm{A.15.1}) \\ x_j \in \{0, 1\} & j = 1, \ldots, n & (\mathrm{A.15.2}) \end{array} \right. \end{cases}
$$

This second linearization consists in rewriting the economic function of $\mathrm{P}_{\mathrm{A.15}}$ by factoring variables x_i in the quadratic part of this function and then replacing, for all $i = 1, \ldots, n-1$, the expression $x_i \sum_{j=i+1}^{n} q_{ij} x_j$ by variable z_i. By proceeding in this way, the economic function is written $\sum_{j=1}^{n} a_j x_j + \sum_{i=1}^{n-1} z_i$. We must then add constraints A.16.2 and A.16.3 to force variable z_i to take, at the optimum of the program obtained, the value of the expression $x_i \sum_{j=i+1}^{n} q_{ij} x_j$. We finally obtain program $\mathrm{P}_{\mathrm{A.16}}$.

$$
\mathrm{P}_{\mathrm{A.16}}: \begin{cases} \min \quad q(x_1, x_2, \ldots, x_n) = \sum\limits_{j=1}^{n} a_j x_j + \sum\limits_{i=1}^{n-1} z_i \\ \text{s.t.} \left| \begin{array}{lll} \sum\limits_{j=1}^{n} e_{kj} x_j \le b_k & k = 1, \ldots, m & (\mathrm{A.16.1}) \\ z_i \ge x_i \sum\limits_{\substack{j = i+1, \ldots, n: \\ q_{ij} < 0}} q_{ij} & i = 1, \ldots, n-1 & (\mathrm{A.16.2}) \\ z_i \ge \sum\limits_{j=i+1, \ldots, n} q_{ij} x_j - (1 - x_i) \sum\limits_{\substack{j = i+1, \ldots, n: \\ q_{ij} > 0}} q_{ij} & i = 1, \ldots, n-1 & (\mathrm{A.16.3}) \\ x_j \in \{0, 1\} & j = 1, \ldots, n & (\mathrm{A.16.4}) \\ z_i \in \mathbb{R} & i = 1, \ldots, n-1 & (\mathrm{A.16.5}) \end{array} \right. \end{cases}
$$

Let us look at constraints A.16.2 and A.16.3. If $x_i = 0$, constraint A.16.2 becomes $z_i \ge 0$ and constraint A.16.3, $z_i \ge \sum_{j=i+1,\ldots,n} q_{ij} x_j - \sum_{j=i+1,\ldots,n:\, q_{ij} > 0} q_{ij}$. The right-hand side of the latter constraint is negative or zero whatever the values taken by variables x_j. Finally, in this case and because we are seeking to minimize variable z_i, this variable takes the value 0 at the optimum of $\mathrm{P}_{\mathrm{A.16}}$. If $x_i = 1$, constraint A.16.2 becomes $z_i \ge \sum_{j=i+1,\ldots,n:\, q_{ij} < 0} q_{ij}$ and constraint A.16.3 becomes $z_i \ge \sum_{j=i+1,\ldots,n} q_{ij} x_j$. Note that $\sum_{j=i+1,\ldots,n} q_{ij} x_j \ge \sum_{j=i+1,\ldots,n:\, q_{ij} < 0} q_{ij}$ whatever the values taken by the Boolean variables x_j. Because we are seeking to minimize z_i, this variable takes, at the optimum of $\mathrm{P}_{\mathrm{A.16}}$, the greater of the 2 values $\sum_{j=i+1,\ldots,n:\, q_{ij} < 0} q_{ij}$, $\sum_{j=i+1,\ldots,n} q_{ij} x_j$, that is to say the value $\sum_{j=i+1,\ldots,n} q_{ij} x_j$. Finally, at the optimum of $\mathrm{P}_{\mathrm{A.16}}$, $z_i = x_i \sum_{j=i+1,\ldots,n} q_{ij} x_j$. Note that the technique just presented for linearizing the economic function could be applied in the same way to linearize quadratic constraints.

Example A.4. Let us go back to $\mathrm{P}_{\mathrm{A.13}}$ and apply this second linearization method to it. The quadratic part of the economic function can be rewritten $x_1(2x_2 - 4x_3) + x_2(5x_3 + 6x_5) + x_3(-4x_4 + 6x_5) + x_4(-2x_5)$. We thus obtain the mixed-integer linear program $\mathrm{P}_{\mathrm{A.17}}$ which is equivalent to program $\mathrm{P}_{\mathrm{A.13}}$.

$$\mathrm{P}_{\mathrm{A}.17}:\begin{cases}\min \quad q(x_1,x_2,\ldots,x_5)=-3x_1-2x_2+3x_3-10x_4+4x_5+z_1+z_2+z_3+z_4\\ \text{s.t.}\left|\begin{array}{ll}2x_1+3x_2+3x_3+5x_4-2x_5\le 20 & \mid \quad z_4\ge -2x_4;\ z_4\ge -2x_5\\ z_1\ge -4x_1;\ z_1\ge 2x_2-4x_3-2(1-x_1) & \mid \quad x_1,x_2,x_3,x_4,x_5\in\{0,1\}\\ z_2\ge 0;\ z_2\ge 5x_3+6x_5-11(1-x_2) & \mid \quad z_1,z_2,z_3,z_4,z_5\in\mathbb{R}\\ z_3\ge -4x_3;\ z_3\ge -4x_4+6x_5-6(1-x_3) & \mid\end{array}\right.\end{cases}$$

The optimal solution for $\mathrm{P}_{\mathrm{A}.17}$ is: $(x_1=1,x_2=0,x_3=1,x_4=1,x_5=0,$ $z_1=-4,z_2=0,z_3=-4,\ z_4=0)$, and this solution gives the economic function the value -18.

A.5.2 Convexifications

There are several methods to transform $\mathrm{P}_{\mathrm{A}.11}$ into an equivalent quadratic program with a convex continuous relaxation. Examples of these methods are given below, in the particular case of minimizing a quadratic function whose variables are subject to linear constraints. Let us therefore consider program $\mathrm{P}_{\mathrm{A}.18}$.

$$\mathrm{P}_{\mathrm{A}.18}:\begin{cases}\min \quad q(x_1,x_2,\ldots,x_n)=\sum\limits_{j=1}^{n}a_jx_j+\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}q_{ij}x_ix_j\\ \text{s.t.}\left|\begin{array}{lll}\sum\limits_{j=1}^{n}e_{kj}x_j\le b_k & k=1,\ldots,m & (\mathrm{A}.18.1)\\ x_j\in\{0,1\} & j=1,\ldots,n & (\mathrm{A}.18.2)\end{array}\right.\end{cases}$$

Let $Q^+=\{(i,j):q_{ij}>0\}$ and $Q^-=\{(i,j):q_{ij}<0\}$. Since variables x_i, $i=1,\ldots,$ n, are Boolean variables, the function $\tilde{q}$ below is equal to the economic function of $\mathrm{P}_{\mathrm{A}.18}$, $q(x_1,x_2,\ldots,x_n)$, for all $x\in\{0,1\}^n$.

$$\begin{aligned}\tilde{q}(x_1,x_2,\ldots,x_n)=\sum_{j=1}^{n}a_jx_j&+\frac{1}{2}\sum_{(i,j)\in Q^+}q_{ij}\big((x_i+x_j)^2-(x_i+x_j)\big)\\&-\frac{1}{2}\sum_{(i,j)\in Q^-}q_{ij}\big((x_i-x_j)^2-(x_i+x_j)\big)\end{aligned}$$

In addition, $\tilde{q}(x_1,x_2,\ldots,x_n)$ is a convex function. Program $\mathrm{P}_{\mathrm{A}.18}$ is equivalent to program $\mathrm{P}_{\mathrm{A}.19}$ and the continuous relaxation of $\mathrm{P}_{\mathrm{A}.19}$ is a convex quadratic program.

$$\mathrm{P}_{\mathrm{A}.19}:\begin{cases}\min \quad \tilde{q}(x_1,x_2,\ldots,x_n)\\ \text{s.t.}\mid (\mathrm{A}.18.1),(\mathrm{A}.18.2)\end{cases}$$

Let us now consider another convexification method. Let us rewrite program $\mathrm{P}_{\mathrm{A}.18}$ using a matrix writing of the quadratic part of the economic function. We obtain program $\mathrm{P}_{\mathrm{A}.20}$ in which the matrix M, of general term m_{ij}, is the symmetric matrix associated with the quadratic form $\sum_{i=1}^{n}\sum_{j=1}^{n}q_{ij}x_ix_j$. We have, for all $i\le j$, $m_{ji}=m_{ij}=\big(q_{ij}+q_{ji}\big)/2$.

$$P_{A.20}:\begin{cases}\min \; q(x_1,x_2,\ldots,x_n)=\sum\limits_{j=1}^{n} a_j x_j \; + x^t\,Mx\\ \text{s.t.}|\;(A.18.1),(A.18.2)\end{cases}$$

One way of reformulating program $P_{A.20}$ – and thus program $P_{A.18}$ – as an equivalent quadratic program whose continuous relaxation is a convex quadratic program, is to add to the function $q(x_1,x_2,\ldots,x_n)$ – assumed to be non-convex – the quantity $\sum_{j=1,\ldots,n}\lambda_{\min}(x_j^2-x_j)$ where $\lambda_{\min}$ denotes the absolute value of the smallest eigenvalue of the square and symmetric matrix, M. This gives the convex quadratic function $\hat{q}(x_1,x_2,\ldots,x_n)=q(x_1,x_2,\ldots,x_n)+\sum_{j=1,\ldots,n}\lambda_{\min}\left(x_j^2-x_j\right)$ which is equal to $q(x_1,x_2,\ldots,x_n)$ for all $x\in\{0,1\}^n$. This method requires a relatively simple pre-processing of program $P_{A.20}$, the calculation of the smallest eigenvalue of the matrix M. Finally, one can thus solve $P_{A.20}$ by solving the quadratic program $P_{A.21}$ whose continuous relaxation is convex.

$$P_{A.21}:\begin{cases}\min \;\; q(x_1,x_2,\ldots,x_n)+\sum\limits_{j=1}^{n}\lambda_{\min}\left(x_j^2-x_j\right)\\ \text{s.t.}|\;(A.18.1),(A.18.2)\end{cases}$$

Example A.5. Let us consider again program $P_{A.13}$ and transform it into a quadratic program whose continuous relaxation is a convex program. To do this, let us add to the economic function the quantity – zero for any feasible solution – $\lambda_{\min}\sum_{j=1}^{n}\left(x_j^2-x_j\right)$ where $\lambda_{\min}$ is equal to the absolute value of the smallest eigenvalue of the matrix associated with the quadratic form $2x_1x_2-4x_1x_3+5x_2x_3+6x_2x_5-4x_3x_4+6x_3x_5-2x_4x_5$, *i.e.*, of the matrix

$$M=\begin{pmatrix}0 & 1 & -2 & 0 & 0\\ 1 & 0 & 2.5 & 0 & 3\\ -2 & 2.5 & 0 & -2 & 3\\ 0 & 0 & -2 & 0 & -1\\ 0 & 3 & 3 & -1 & 0\end{pmatrix}.$$

The smallest eigenvalue of M is equal to -4.19. Program $P_{A.13}$ is therefore equivalent to program $P_{A.22}$ whose continuous relaxation is convex.

$$P_{A.22}:\begin{cases}\min \;\; -3x_1-2x_2+3x_3-10x_4+4x_5+2x_1x_2\\ -4x_1x_3+5x_2x_3+6x_2x_5-4x_3x_4+6x_3x_5-2x_4x_5\\ +4.2\sum\limits_{i=1}^{n}\left(x_i^2-x_i\right)\\ \text{s.t.}\left|\begin{array}{ll}2x_1+3x_2+3x_3+5x_4-2x_5\le 20 & (A.22.1)\\ x_1,x_2,x_3,x_4,x_5\in\{0,1\} & (A.22.2)\end{array}\right.\end{cases}$$

There are other, more elaborate pre-treatments of program $P_{A.18}$ – whose continuous relaxation is non-convex – allowing it to be rewritten as an equivalent quadratic program whose continuous relaxation is a convex quadratic program. These methods, which are based on positive semidefinite programming, also allow the processing of quadratic programs containing simultaneously Boolean variables, integer variables and real variables. There are many publications dealing with these linearization and convexification methods. Some of them are mentioned below.

References and Further Reading

Adams W.P., Sherali H.D. (1990) Linearization strategies for a class of zero–one mixed integer programming problems, *Oper. Res.* **38**, 217.

Adams W.P., Forrester R.J., Glover F.W. (2004) Comparisons and enhancement strategies for linearizing mixed 0–1 quadratic programs, *Discr. Opti.* **1**, 99.

Balas E., Mazzola J.B. (1984a) Nonlinear 0–1 programming: I. Linearization techniques, *Math. Prog.* **30**, 1.

Balas E., Mazzola J.B. (1984b) Nonlinear 0–1 programming: II, Dominance relations and algorithms, *Math. Prog.* **30**, 22.

Billionnet A., Elloumi S. (2007) Using a mixed integer quadratic programming solver for the unconstrained quadratic 0–1 problem, *Math. Prog.* **109**, 55.

Billionnet A., Elloumi S., Plateau M. (2009) Improving the performance of standard solvers for quadratic 0–1 programs by a tight convex reformulation: the QCR method, *Discr. Appl. Math.* **157**, 1185.

Billionnet A., Elloumi S., Lambert A. (2012) Extending the QCR method to the case of general mixed integer programs, *Math. Prog.* **131**, 381.

Billionnet A., Elloumi S., Lambert A. (2016) Exact quadratic convex reformulations of mixed-integer quadratically constrained problems, *Math. Prog.* **158**, 235.

Billionnet A., Elloumi S., Lambert A., Wiegele A. (2017) Using a conic bundle method to accelerate both phases of a quadratic convex reformulation,. *Info. J. Comp.* **29**, 318.

Chardaire P., Sutter A. (1995) A decomposition method for quadratic zero–one programming, *Manage. Sci.* **41**, 704.

Chen D.S., Batson R.G., Dang Y. (2010) *Applied integer programming: modeling and solution.* John Wiley & Sons.

Elloumi S., Faye A., Soutif E. (2000) Decomposition and linearization for 0–1 quadratic programming, *Ann. Oper. Res.* **99**, 79.

Faye A., Roupin F. (2007) Partial Lagrangian relaxation for general quadratic programming, *4OR* **5**, 75.

Fortet R. (1959) L'algèbre de Boole et ses applications en recherche opérationnelle, *Cahiers du Centre d'Etudes de Recherche Opérationnelle* **1**, 5.

Fortet R. (1960) Applications de l'algèbre de Boole en recherche opérationnelle, *Revue Française de Recherche Opérationnelle* **4**, 17.

Glover F., Woolsey E. (1973) Further reduction of zero–one polynomial programming to zero–one linear programming problems, *Oper. Res.* **21**, 156.

Glover F., Woolsey E. (1974) Converting the 0–1 polynomial programming problem to a linear 0–1 program, *Oper. Res.* **22**, 180.

Hammer P.L., Rudeanu S. (1968) *Boolean methods in operations research and related areas.* Springer.

Hammer P.L., Rubin A.A. (1970) Some remarks on quadratic programming with 0–1 variables, *RIRO* **3**, 67.

Rodriguez-Heck E. (2018) Linear and quadratic reformulations of nonlinear optimization problems in binary variables, Thèse de l'Univesité de Liège.

Sherali H.D., Adams W.P. (1986) A tight linearization and an algorithm for 0–1 quadratic programming problems, *Manage. Sci.* **32**, 1274.

A.6 Fractional Programming

The general fractional optimization problem can be written in the form of the mathematical program $\mathrm{P}_{\mathrm{A.23}}$.

$$\mathrm{P}_{\mathrm{A.23}} : \begin{cases} \max & f(x)/g(x) \\ \text{s.t.} & x \in X \end{cases}$$

The set X is a compact, non-empty subset of $\mathbb{R}^n$. The functions $f(x)$ and $g(x)$ are continuous functions with real values defined on the set X. It is assumed here that $g(x) > 0$ for any x belonging to X. There are many methods to solve this problem. We present below one of these methods, the Dinkelbach algorithm.

Let λ be a parameter belonging to the set of real numbers. Let us consider the parametric problem $\mathrm{P}_{\mathrm{A.24}}(\lambda)$ associated with $\mathrm{P}_{\mathrm{A.23}}$.

$$\mathrm{P}_{\mathrm{A.24}}(\lambda) : \begin{cases} \max & f(x) - \lambda\, g(x) \\ \text{s.t.} & x \in X \end{cases}$$

Let us denote by $v(\lambda)$ the optimal value of $\mathrm{P}_{\mathrm{A.24}}(\lambda)$ and by x^*_λ, an optimal solution to this program. We can prove that $v(\lambda) = 0$ if and only if λ is the optimal value of $\mathrm{P}_{\mathrm{A.23}}$ and x^*_λ, an optimal solution to this problem. Thus, we obtain another formulation of program $\mathrm{P}_{\mathrm{A.23}}$:

Find $\lambda \in R$ such that $v(\lambda) = 0$, where $v(\lambda) = \max\{f(x) - \lambda g(x) : x \in X\}$.

From this formulation, we will be able to build algorithms to solve $\mathrm{P}_{\mathrm{A.23}}$, based on classical methods to determine the root of a function – the Newton method. This is the case of the Dinkelbach algorithm presented below.

The Dinkelbach Algorithm

Step 1. $\lambda \leftarrow f(x_0)/g(x_0)$ where x_0 is a point of X.

Step 2. calculate $v(\lambda) = \max\{f(x) - \lambda g(x) : x \in X\}$ and let x_λ be such that $v(\lambda) = f(x_\lambda) - \lambda g(x_\lambda)$.

Step 3. **if** $v(\lambda) \neq 0$ **then** $\lambda \leftarrow f(x_\lambda)/g(x_\lambda)$ and go to 2 **else** x_λ is an optimal solution **endif.**

The difficulty of this algorithm depends on the difficulty of the optimization problem of Step 2.

In the case where the functions $f(x)$ and $g(x)$ are linear or affine and X is a convex polyhedron, $\mathrm{P}_{\mathrm{A.23}}$ is a linear or hyperbolic fractional optimization problem. It is written as $\mathrm{P}_{\mathrm{A.25}}$.

$$\mathrm{P}_{\mathrm{A.25}} : \begin{cases} \max & \left(b_0 + \sum_{j=1}^n b_j x_j\right) \Big/ \left(c_0 + \sum_{j=1}^n c_j x_j\right) \\ \text{s.t.} & \left| \begin{array}{lll} \sum\limits_{j=1}^{n} a_{ij} x_j \le d_i & i = 1,\ldots,m & \text{(A.25.1)} \\ x_j \ge 0 & j = 1,\ldots,n & \text{(A.25.2)} \end{array} \right. \end{cases}$$

where b_0, c_0, b_j $(j = 1,\ldots,n)$, c_j $(j = 1,\ldots,n)$, and $a_{ij} (i = 1,\ldots,m, j = 1,\ldots,n)$ are real coefficients such that, for any feasible solution of $\mathrm{P}_{\mathrm{A.25}}$, $c_0 + \sum_{j=1}^n c_j x_j > 0$. Program

$P_{A.25}$ can be solved by the Dinkelbach algorithm. In this case, the optimization problem of Step 2 consists in solving a continuous linear program. However, by performing the variable changes $y_j = x_j/(c_0 + \sum_{j=1}^{n} c_j x_j)$, $j = 1,\ldots,n$, and $t = 1/(c_0 + \sum_{j=1}^{n} c_j x_j)$, $P_{A.25}$ can be rewritten as the equivalent linear program $P_{A.26}$.

$$
P_{A.26}: \begin{cases} \max \quad b_0 t + \sum_{j=1}^{n} b_j y_j & \\ \text{s.t.} \left| \begin{array}{lll} c_0 t + \sum_{j=1}^{n} c_j y_j = 1 & & (A.26.1) \\ \sum_{j=1}^{n} a_{ij} y_j \le d_i t & i = 1,\ldots,m & (A.26.2) \\ y_j \ge 0 & j = 1,\ldots,n & (A.26.3) \\ t \ge 0 & & (A.26.4) \end{array} \right. \end{cases}
$$

Let us now consider the mixed-integer linear fractional optimization problem $P_{A.27}$.

$$
P_{A.27}: \begin{cases} \max \quad \left(b_0 + \sum_{j=1}^{n} b_j x_j\right) \Big/ \left(c_0 + \sum_{j=1}^{n} c_j x_j\right) & \\ \text{s.t.} \left| \begin{array}{lll} \sum_{j=1}^{n} a_{ij} x_j \le d_i & i = 1,\ldots,m & (A.27.1) \\ x_j \ge 0 & j = 1,\ldots,p & (A.27.2) \\ x_j \in \{0,1\} & j = p+1,\ldots,n & (A.27.3) \end{array} \right. \end{cases}
$$

As before, it is assumed that $c_0 + \sum_{j=1}^{n} c_j x_j > 0$ for any feasible solution. The Dinkelbach algorithm, presented above, can be used to solve $P_{A.27}$. In this case, X is defined by Constraints A.27.1–A.27.3, and Step 2 of the algorithm consists in solving a mixed-integer linear program.

Fractional optimization problems are very diverse. For example, one can look at the ratio of two quadratic functions or at the sum of several ratios. For more information on fractional optimization, the reader can consult the references cited below.

References and Further Reading

Bajalinov E.B. (2003) *Linear-fractional programming: theory, methods, applications and software.* Springer Science + Business Media.

Billionnet A. (2002) Approximate and exact solution methods for the hyperbolic 0–1 knapsack problem, *INFOR: Info. Syst. Oper. Res.* **40**, 97.

Billionnet A., Djebali K. (2006) Résolution d'un problème combinatoire fractionnaire par la programmation linéaire mixte, *RAIRO-Oper. Res.* **40**, 97.

Borrero J.S., Gillen C., Prokopyev O.A. (2016) A simple technique to improve linearized reformulations of fractional (hyperbolic) 0–1 programming problems, *Oper. Res. Lett.* **44**, 479.

Charnes A., Cooper W.W. (1962) Programming with linear fractional functionals, *Naval Res. Logis. Quart.* **9**, 181.

Craven B.D. (1988) *Fractional programming.* Heldermann Verlag.

Dinkelbach W. (1967) On nonlinear fractional programming, *Manage. Sci.* **13**, 492.

Frenk J.B.G., Schaible S. (2005) Fractional programming. *Handbook of generalized convexity and generalized monotonicity* (N. Hadjisavvas, S.Komlósi, S. Schaible, Eds). Nonconvex optimization and its applications, Springer, p. 76.

Freund R.W., Jarre F. (2001) Solving the sum-of-ratios problem by an interior point method, *J. Global Opti.* **19**, 83.

Jeflea A. (2003) A parametric study for solving nonlinear fractional problems, *An. St. Univ. Ovidius Constanta* **11**, 87.

Konno H. (2001) Minimization of the sum of several linear fractional functions, *Generalized convexity and generalized monotonicity* (N. Hadjisavvas, J.E. Martínez-Legaz, J.P. Penot, Eds.) Lecture Notes in Economics and Mathematical Systems. Springer Verlag, Vol. 502, pp. 3–20.

Kuno T., (2002) A branch-and-bound algorithm for maximizing the sum of several linear ratios, *J. Global Opti.* **22**, 155.

Li H. (1994) A Global approach for general fractional programming, *Eur. J. Oper. Res.* **73**, 590.

Nagih A., Plateau G. (1999) Problèmes fractionnaires: tour d'horizon sur les applications et méthodes de resolution, *RAIRO – Oper. Res.* **33**, 383.

Radzik T. (1998) Fractional combinatorial optimization, *Handbook of combinatorial optimization* (D.-Z. Du, P.M. Pardalos). Kluwer Academic Publishers, pp. 429–478.

Rodenas R.G., Lopez M.L., Verastegui D. (1999) Extensions of Dinkelbach's algorithm for solving non-linear fractional programming problems, *Sociedad de Estadistica e Investigacion Operativa* **7**, 33.

Saipe A.L. (1975) Solving a 0–1 hyperbolic program by branch-and-bound, *Naval Res. Logis. Quart.* **22**, 397.

Schaible S., Ibaraki T. (1983) Fractional programming, *Eur. J. Oper. Res.* **12**, 325.

Schaible S. (1995) Fractional programming, *Handbook of global optimization* (R. Horst, P.M. Pardalos). Kluwer Academic Publishers, pp. 495–608.

Schaible S., Shi J. (2003) Fractional programming: the sum-of-ratios case, *Opti. Meth. Soft.* **18**, pp. 219–229.

Stancu-Minasian I.M. (1997) *Fractional programming*. Kluwer Academic Publishers.

Wu T. (1997) A note on a global approach for general 0–1 fractional programming, *Eur. J. Oper. Res.*, 229.

You F., Castro P.M., Grossmann I.E. (2009) Dinkelbach's algorithm as an efficient method to solve a class of MINLP models for large-scale cyclic scheduling problems, *Comp. Chem. Eng.* **33**, 1879.

A.7 Piecewise Linear Functions

In this section, we are interested in mathematical programs involving piecewise linear functions and linear functions in the economic function and/or in the constraints. In fact, in the general case, such programs can be rewritten as mixed-integer linear programs. Note that this notion of piecewise linearity is interesting since any continuous function of one variable can be approximated by a piecewise linear function, the quality of the approximation depending on the size of the segments. Let $f(x)$ be a piecewise linear function defined on the interval $[b_0,\ b_p]$ in the following way:

$$\begin{array}{ll} f(x) = a_1 x + d_1 & b_0 \le x \le b_1 \\ f(x) = a_2 x + d_2 & b_1 \le x \le b_2 \\ f(x) = a_3 x + d_3 & b_2 \le x \le b_3 \\ \dots\dots\dots\dots\dots & \dots\dots\dots \\ f(x) = a_p x + d_p & b_{p-1} \le x \le b_p \end{array}$$

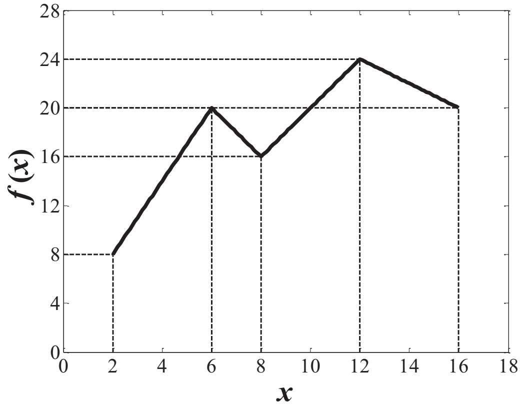


FIG. A.1 – A piecewise linear function, $f(x)$, defined by the 5 points of coordinates (2,8), (6,20), (8,16), (12,24), and (16,20).

The coefficients $b_0, b_1, \ldots, b_p$ are real numbers such that $0 \leq b_0 < b_1 < \cdots < b_p$. The coefficients $a_1, \ldots, a_p$ represent the slope of the different segments. Figure A.1 shows a piecewise linear function, $f(x)$, defined on the interval [2, 16] by the 5 points of coordinates (2, 8), (6, 20), (8, 16), (12, 24), and (16, 20). The 4 corresponding linear – or affine – functions are: $f_1(x) = 3x + 2$, $f_2(x) = -2x + 32$, $f_3(x) = 2x$, and $f_4(x) = -x + 36$.

A first formulation. This formulation allows a piecewise linear function to be expressed as a linear function subject to linear constraints. This formulation uses additional Boolean variables and also additional real variables. Note, first of all, that, for any x between b_i and b_{i+1}, two non-negative reals, λ_i and λ_{i+1}, can be defined, whose sum is 1 and such that $x = \lambda_i b_i + \lambda_{i+1} b_{i+1}$. It is thus deduced that, for any x between b_i and b_{i+1}, the piecewise linear function $f(x)$ can be written $f(x) = \lambda_i f(b_i) + \lambda_{i+1} f(b_{i+1})$ where λ_i and λ_{i+1} satisfy the above properties. Finally, we can therefore write the function $f(x)$ in the form $f(x) = \sum_{i=0}^{p} \lambda_i f(b_i)$ where all λ_i are non-negative real numbers such that $\sum_{i=0}^{p} \lambda_i = 1$, and satisfying the following conditions: if $b_i \leq x \leq b_{i+1}$ then $\lambda_i + \lambda_{i+1} = 1$ and $x = \lambda_i b_i + \lambda_{i+1} b_{i+1}$. We can therefore write $f(x)$ in the form $\sum_{i=0}^{p} \lambda_i f(b_i)$, where variables z_i and λ_i satisfy constraints $C_{A.1}$ below.

$$
C_{A.1}: \begin{cases}
x = \sum_{i=0}^{p} \lambda_i b_i & & \text{(CA.1.1)} & | & \sum_{i=0}^{p} \lambda_i = 1 & & \text{(CA.1.5)} \\
\lambda_0 \leq z_0 & & \text{(CA.1.2)} & | & \sum_{i=0}^{p-1} z_i = 1 & & \text{(CA.1.6)} \\
\lambda_i \leq z_{i-1} + z_i & 1 \leq i < p & \text{(CA.1.3)} & | & \lambda_i \geq 0 & i = 0, 1, \ldots, p & \text{(CA.1.7)} \\
\lambda_p \leq z_{p-1} & & \text{(CA.1.4)} & | & z_i \in \{0, 1\} & i = 0, 1, \ldots, p-1 & \text{(CA.1.8)}
\end{cases}
$$

Indeed, constraints CA.1.6 and CA.1.8 require that one and only one variable z_i be equal to 1. Moreover, if $z_i = 1$, then constraints CA.1.2, CA.1.3, CA.1.4, CA.1.5, and CA.1.7 imply $\lambda_i + \lambda_{i+1} = 1$ and $\lambda_k = 0$, for all k different from i or $i+1$.

Example A.6. Let us apply the approach presented above to the mathematical program $\mathrm{P}_{\mathrm{A.28}}$. In this program, the economic function and one of the constraints are expressed as the sum of a piecewise linear function of variable x, and a linear function of variable x and other variables, t_1, t_2, and t_3. The piecewise linear function $f(x)$ is defined on the interval [1, 7] by the points of coordinates (1, 3), (3, 5), (5, 3), and (7, 5).

$$\mathrm{P}_{\mathrm{A.28}}: \begin{cases} \min\ f(x) + t_1 - 2t_2 + t_3 - 2x \\ \text{s.t.} \left| \begin{array}{llcll} 4.5x + t_1 - t_3 \le 27.5 & (\mathrm{A.28.1}) & | & t_1 \le 6 & (\mathrm{A.28.4}) \\ x + 2f(x) + 2t_1 + 2t_2 \le 24 & (\mathrm{A.28.2}) & | & t_2 \le 4 & (\mathrm{A.28.5}) \\ 1 \le x \le 7 & (\mathrm{A.28.3}) & | & t_1, t_2, t_3 \ge 0 & (\mathrm{A.28.6}) \end{array} \right. \end{cases}$$

The solution of $\mathrm{P}_{\mathrm{A.28}}$ can be obtained by solving the mixed-integer linear program $\mathrm{P}_{\mathrm{A.29}}$ in which variable e represents the value of the piecewise linear function $f(x)$.

$$\mathrm{P}_{\mathrm{A.29}}: \begin{cases} \min\quad e + t_1 - 2t_2 + t_3 - 2x \\ \text{s.t.} \left| \begin{array}{lclclcl} 4.5x + t_1 - t_3 \le 27.5 & | & \lambda_0 \le z_0 & | & z_0 + z_1 + z_2 = 1 & | & t_2 \le 4 \\ x + 2e + 2t_1 + 2t_2 \le 24 & | & \lambda_1 \le z_0 + z_1 & | & \lambda_0 + \lambda_1 + \lambda_2 + \lambda_3 = 1 & | & \lambda_0, \lambda_1, \lambda_2, \lambda_3 \ge 0 \\ e = 3\lambda_0 + 5\lambda_1 + 3\lambda_2 + 5\lambda_3 & | & \lambda_2 \le z_1 + z_2 & | & 1 \le x \le 7 & | & t_1, t_2, t_3 \ge 0 \\ x = \lambda_0 + 3\lambda_1 + 5\lambda_2 + 7\lambda_3 & | & \lambda_3 \le z_2 & | & t_1 \le 6 & | & z_0, z_1, z_2 \in \{0, 1\} \end{array} \right. \end{cases}$$

Note that the values of both variables e and x are entirely defined by the values of variables λ_i, $i = 0, 1, 2, 3$. The optimal solution of $\mathrm{P}_{\mathrm{A.29}}$ is: ($x = 6.1111$, $t_1 = 0$, $t_2 = 4$, $t_3 = 0$); the corresponding values of variables e, λ_i and z_i are: $e = 4.1111$, $\lambda_0 = 0$, $\lambda_1 = 0$, $\lambda_2 = 0.4444$, $\lambda_3 = 0.5556$, $z_0 = 0$, $z_1 = 0$, $z_2 = 1$. The value of the optimal solution is equal to -16.1111.

A second formulation. We present below another way of expressing the piecewise linear function $f(x)$ defined at the beginning of this section. In this formulation, $f(x)$ is expressed as a linear function of real variables, u_i, and Boolean variables, z_i, $i = 1, \ldots, p$, these variables being subject to linear constraints. It is indeed easy to verify that, for any x belonging to the interval $[b_0, b_p]$, $f(x) = \sum_{i=1}^{p} (a_i u_i + d_i z_i)$, provided that variables u_i and z_i satisfy constraints $\mathrm{C}_{\mathrm{A.2}}$ below.

$$\mathrm{C}_{\mathrm{A.2}}: \begin{cases} b_{i-1} z_i \le u_i \le b_i z_i \quad i = 1, \ldots, p & (\mathrm{CA.2.1}) \\ x = \sum_{i=1}^{p} u_i & (\mathrm{CA.2.2}) \\ \sum_{i=1}^{p} z_i = 1 & (\mathrm{CA.2.3}) \\ z_i \in \{0, 1\} \qquad i = 1, \ldots, p & (\mathrm{CA.2.4}) \end{cases}$$

Example A.7. Let us apply this second method to the previous program $P_{A.28}$. Since the piecewise linear function $f(x)$ is defined on the interval [1, 7] by the points of coordinates (1, 3), (3, 5), (5, 3), and (7, 5), we have: $a_1 = 1, a_2 = -1, a_3 = 1$ and $d_1 = 2, d_2 = 8, d_3 = -2$. The equivalent mixed-integer linear program $P_{A.30}$ is obtained in which variable e represents the value of the piecewise linear function $f(x)$.

$$P_{A.30}: \begin{cases} \min \quad e + t_1 - 2t_2 + t_3 - 2x \\ \text{s.t.} \left| \begin{array}{lcl} x + 2e + 2t_1 + 2t_2 \le 24 & | & x = u_1 + u_2 + u_3 \\ 4.5x + t_1 - t_3 \le 27.5 & | & z_1 + z_2 + z_3 = 1 \\ e = u_1 + 2z_1 - u_2 + 8z_2 + u_3 - 2z_3 & | & 0 \le t_1 \le 6;\ 0 \le t_2 \le 4;\ t_3 \ge 0 \\ z_1 \le u_1 \le 3z_1;\ 3z_2 \le u_2 \le 5z_2;\ 5z_3 \le u_3 \le 7z_3 & | & z_1, z_2, z_3 \in \{0,1\} \end{array} \right. \end{cases}$$

Note that the possible values of variables u_i, $i = 1, 2, 3$, are completely defined by the values of variables z_i, $i = 1, 2, 3$, that the value of variable x is completely defined by the values of variables u_i, $i = 1, 2, 3$, and that the value of variable e is completely defined by the values of variables u_i and z_i, $i = 1, 2, 3$. The optimal solution of $P_{A.30}$ is: $(x = 6.1111,\ t_1 = 0,\ t_2 = 4,\ t_3 = 0)$; the corresponding values of variables e, and z_i are: $e = 4.1111,\ z_1 = 0,\ z_2 = 0,\ z_3 = 1$. The value of the optimal solution is -16.1111.

Maximization of a concave piecewise linear function. In the concave case, a piecewise linear function can be expressed as the maximum of a linear function of additional real variables, these variables being subject to linear constraints. In this case, the use of Boolean variables becomes unnecessary. Recall that a function $f(x)$ defined on a domain D is concave if and only if.

$$\forall x_1, x_2 \in D, \forall \lambda \in [0,1] : f(x) = f(\lambda x_1 + (1-\lambda)\, x_2) \ge \lambda f(x_1) + (1-\lambda)\, f(x_2).$$

Figure A.2 shows an example of a concave piecewise linear function.

Let us consider a concave piecewise linear function, $f(x)$, defined on the interval $[b_0, b_p]$ by the points of coordinates $(b_0, y_0), (b_1, y_1), \ldots, (b_p, y_p)$ with $0 \le b_0 < b_1 < \cdots < b_p$. It can be shown that, for all x belonging to the interval $[b_0, b_p]$,

$$f(x) = \max_{u_1,\ldots,u_p} \left\{ y_0 + \sum_{i=1}^{p} \frac{y_i - y_{i-1}}{b_i - b_{i-1}} u_i : x = \sum_{i=1}^{p} u_i, 0 \le u_i \le b_i - b_{i-1}\ (i = 1, \ldots, p) \right\}.$$

Example A.8. Let us apply the previous method to program $P_{A.31}$ which consists in maximizing the sum of two concave piecewise linear functions subject to linear constraints.

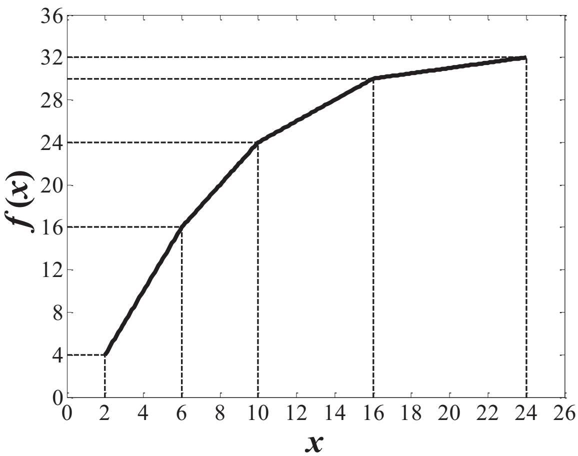


FIG. A.2 – A concave piecewise linear function, $f(x)$, defined by the 5 points of coordinates (2, 4), (6, 16), (10, 24), (16, 30), and (24, 32). The successive slopes of the 4 segments are decreasing and equal to 3, 2, 1, and 0.25, respectively.

$$
\mathrm{P}_{\mathrm{A}.31}:\left\{\begin{array}{l}\max f_1(x_1)+f_2(x_2)\\[2mm] \text{s.t.}\left|\begin{array}{lcl} x_1+2x_2\le 6 & | & 0\le x_1\le 6\\ 3x_1+x_2\ge 10 & | & 0\le x_2\le 8\end{array}\right.\end{array}\right.
$$

The function $f_1(x_1)$ is defined by the points of coordinates (0, 0), (1, 2), (3, 4), and (6, 5), and the function $f_2(x_2)$ is defined by the points of coordinates (0, 0), (2, 4), (4, 6), and (8, 7). The linear program $\mathrm{P}_{\mathrm{A}.31}$ is equivalent to $\mathrm{P}_{\mathrm{A}.32}$.

$$
\mathrm{P}_{\mathrm{A}.32}:\left\{\begin{array}{ll}\max & 2u_{11}+u_{12}+\frac{1}{3}u_{13}+2u_{21}+u_{22}+\frac{1}{4}u_{23}\\[2mm] \text{s.t.} & \left|\begin{array}{lcl} x_1+2x_2\le 6 & | & 0\le u_{11}\le 1;\ 0\le u_{12}\le 2;\ 0\le u_{13}\le 3\\ 3x_1+x_2\ge 10 & | & 0\le u_{21}\le 2;\ 0\le u_{22}\le 2;\ 0\le u_{23}\le 4\\ x_1=u_{11}+u_{12}+u_{13} & | & 0\le x_1\le 6\\ x_2=u_{21}+u_{22}+\alpha_{23} & | & 0\le x_2\le 8\end{array}\right.\end{array}\right.
$$

Note that constraints $0\le x_1\le 6$ and $0\le x_2\le 8$ are useless since the value of variable x_1 (resp. x_2) is completely defined by the values of variables u_{1k} (resp. u_{2k}), $k=1,2,3$. The optimal solution of $\mathrm{P}_{\mathrm{A}.32}$ is $(x_1=3, x_2=1.5)$. Its value is 7. The corresponding values of variables u_{ik} are: $u_{11}=1, u_{12}=2, u_{13}=0$ and $u_{21}=1.5, u_{22}=0, u_{23}=0$.

The concave piecewise linear function $f(x)$, defined on the interval $[b_0,\ b_p]$ by the points of coordinates (b_0, y_0), (b_1, y_1),…, (b_p, y_p) with $0\le b_0<b_1<\cdots<b_p$, can also be expressed as follows:

$$\forall x \in [b_0, b_p], f(x) = \min_{i=1,\dots,p} \{a_i x + d_i\}$$

where, for $i = 1,\dots, p$, $a_i = (y_i - y_{i-1})/(b_i - b_{i-1})$ and $d_i = y_i - a_i b_i$. Applying this property to program $P_{A.31}$ results in the equivalent linear program $P_{A.33}$ in which variable e_1 represents the value of the piecewise linear function $f_1(x)$ and variable e_2 the value of the piecewise linear function $f_2(x)$.

$$P_{A.33}: \begin{cases} \max \quad e_1 + e_2 \\ \text{s.t.} \left| \begin{array}{lcl} e_1 \le 2x_1 & | & e_2 \le 0.25x_2 + 5 \\ e_1 \le x_1 + 1 & | & x_1 + 2x_2 \le 6 \\ e_1 \le (1/3)x_1 + 3 & | & 3x_1 + x_2 \ge 10 \\ e_2 \le 2x_2 & | & 0 \le x_1 \le 6 \\ e_2 \le x_2 + 2 & | & 0 \le x_2 \le 8 \end{array} \right. \end{cases}$$

Note that there is no need to further constrain the – non negative – variables e_1 and e_2. The optimal solution of $P_{A.33}$ is: $(x_1 = 3, x_2 = 1.5, e_1 = 4, e_2 = 3)$; the value of this solution is equal to 7.

For a more complete presentation of the possible processing of piecewise linear functions, the reader can consult the references cited below.

References and Further Reading

Billionnet A. (2017) *Programmation discrète*. Dunod, Paris.

Chen D.-S., Batson R.G., Dang Y. (2010) *Applied integer programming, modeling and solution*. John Wiley & Sons.

Li H.-L., Lu H.-C., Huang C.-H., Hu N.-Z. (2009) A superior representation method for piecewise linear functions, *INFORMS J. Comp.* **21**, 314.

Lin M.-H., Carlsson J.G., Ge D., Shi J., Tsai J.-F. (2013) A review of piecewise linearization methods, *Mathematical problems in engineering*. Hindawi Publishing Corporation.

Vielma J.P., Ahmed S., Nemhauser G. (2010) A note on "A superior representation method for piecewise linear functions", *INFORMS J. Comp.* **22**, 493.

A.8 Robustness in Mathematical Programming

It is often necessary to take into account, in a mathematical program, some uncertainty in the data since this uncertainty can strongly influence the quality and feasibility of the selected solution. The robust approach allows this uncertainty to be taken into account to some extent.

Consider the linear program $P_{A.34}$.

$$P_{A.34}: \begin{cases} \max \quad \sum_{j=1}^{n} c_j x_j \\ \text{s.t.} \left| \begin{array}{lll} \sum_{j=1}^{n} a_{ij} x_j \le b_i & i = 1,\dots, m & (A.34.1) \\ l_j \le x_j \le u_j & j = 1,\dots, n & (A.34.2) \end{array} \right. \end{cases}$$

The coefficients c_j $(j = 1,\ldots, n)$, a_{ij} $(i = 1,\ldots, m, j = 1,\ldots, n)$, b_i $(i = 1,\ldots, m)$, l_j $(j = 1,\ldots, n)$, and u_j $(j = 1,\ldots, n)$ are data, and all the coefficients l_j are positive or zero. For all $i \in \{1,\ldots, m\}$, let J_i be the set of indices j such that the coefficient a_{ij} is uncertain. It is assumed here that, for all $i \in \{1,\ldots, m\}$, each entry a_{ij}, $j \in J_i$, corresponds to a bounded symmetric random variable, $\tilde{a}_{ij}$, $j \in J_i$, which takes its values in the interval $[a_{ij} - \hat{a}_{ij}, a_{ij} + \hat{a}_{ij}]$ where $\hat{a}_{ij}$ is a positive or zero constant. Below we present a robust approach proposed by Bertsimas and Sim (2004).

Maximal protection against uncertainty. In this case, the optimal solution of $\mathrm{P}_{\mathrm{A.34}}$ is the one that maximizes the value of the economic function and is feasible regardless of the values taken by the coefficients a_{ij}, in the set of possible values. The linear program $\mathrm{P}_{\mathrm{A.35}}$ allows this solution to be determined.

$$
\mathrm{P}_{\mathrm{A.35}}: \begin{cases} \max \sum_{j=1}^{n} c_j x_j & \\ \text{s.t.} \left| \begin{array}{lll} \sum_{j=1}^{n} a_{ij} x_j + \sum_{j \in J_i} \hat{a}_{ij} x_j \le b_i & i = 1,\ldots, m & \text{(A.35.1)} \\ l_j \le x_j \le u_j & j = 1,\ldots, n & \text{(A.35.2)} \end{array} \right. \end{cases}
$$

Indeed, any feasible solution of $\mathrm{P}_{\mathrm{A.35}}$ remains feasible for all possible values of the random variables $\tilde{a}_{ij}$, *i.e.*, for all values belonging to the interval $[a_{ij} - \hat{a}_{ij}, a_{ij} + \hat{a}_{ij}]$. The optimal solution of $\mathrm{P}_{\mathrm{A.35}}$ is said to be the optimal robust solution of the problem under consideration.

A less conservative approach. Here, it is considered unlikely that all uncertain coefficients will differ simultaneously from their nominal value. It is thus assumed that Γ_i coefficients a_{ij} can differ from their nominal value – at most from the quantity $\hat{a}_{ij}$. Γ_i is therefore an integer belonging to $[0, |J_i|]$. As before, the optimal solution of $\mathrm{P}_{\mathrm{A.34}}$ is then the one that maximizes the value of the economic function, and which is feasible regardless of the values taken by the coefficients a_{ij}, given the uncertainty assumptions. The search for this solution can be formulated as the mathematical program $\mathrm{P}_{\mathrm{A.36}}$.

$$
\mathrm{P}_{\mathrm{A.36}}: \begin{cases} \max \sum_{j=1}^{n} {}_j x_j & \\ \text{s.t.} \left| \begin{array}{lll} \sum_{j=1}^{n} a_{ij} x_j & & \\ \quad + \max_{S_i \subseteq J_i,\, |S_i| \le \Gamma_i} \left[\sum_{j \in S_i} \hat{a}_{ij} x_j \right] \le b_i & i = 1,\ldots, m & \text{(A.36.1)} \\ l_j \le x_j \le u_j & j = 1,\ldots, n & \text{(A.36.2)} \end{array} \right. \end{cases}
$$

Given a feasible solution of $\mathrm{P}_{\mathrm{A.36}}$, $\tilde{x}$, let us denote by $\beta_i(\tilde{x}, \Gamma_i)$ the quantity $\max_{S_i \subseteq J_i,\, |S_i| \le \Gamma_i} [\sum_{j \in S_i} \hat{a}_{ij} \tilde{x}_j]$. This quantity can be determined by solving the linear program $\mathrm{P}_{\mathrm{A.37}}$.

$$
\mathrm{P}_{\mathrm{A.37}}:\begin{cases}\max \sum_{j\in J_i} \hat{a}_{ij}\tilde{x}_j\alpha_{ij} & \\ \text{s.t.}\left| \begin{array}{lll} \sum_{j\in J_i}\alpha_{ij}\le \Gamma_i & & \text{(A.37.1)}\\ 0\le \alpha_{ij}\le 1 & j\in J_i & \text{(A.37.2)}\end{array}\right.\end{cases}
$$

$\mathrm{P}_{\mathrm{A.37}}$ admits an optimal finite solution, which implies that its dual admits one too (duality theory). Moreover, the values of these two optimal solutions are equal. By associating the dual variable z_i to constraint A.37.1 and the dual variables p_{ij}, $j \in J_i$, to constraints A.37.2, this dual problem is written.

$$
\beta_i(\tilde{x},\Gamma_i)=\begin{cases}\min \sum_{j\in J_i} p_{ij}+\Gamma_i z_i & \\ \text{s.t.}\left| \begin{array}{ll} z_i+p_{ij}\ge \hat{a}_{ij}\tilde{x}_j & j\in J_i\\ p_{ij}\ge 0 & j\in J_i\\ z_i\ge 0 & \end{array}\right.\end{cases}.
$$

From this, it can be deduced that the optimal robust solution to the problem under consideration can be determined by solving program $\mathrm{P}_{\mathrm{A.38}}$.

$$
\mathrm{P}_{\mathrm{A.38}}:\begin{cases}\max \sum_{j=1}^{n} c_j x_j & \\ \text{s.t.}\left| \begin{array}{lll} \sum_{j=1}^{n} a_{ij}x_j + \sum_{j\in J_i} p_{ij}+\Gamma_i z_i\le b_i & i=1,\ldots,m & \text{(A.38.1)}\\ z_i+p_{ij}\ge \hat{a}_{ij}x_j & i=1,\ldots,m; j\in J_i & \text{(A.38.2)}\\ l_j\le x_j\le u_j & j=1,\ldots,n & \text{(A.38.3)}\\ z_i\ge 0 & i=1,\ldots,m & \text{(A.38.4)}\\ p_{ij}\ge 0 & i=1,\ldots,m; j\in J_i & \text{(A.38.5)}\end{array}\right.\end{cases}
$$

Example A.9. Consider the linear program $\mathrm{P}_{\mathrm{A.39}}$.

$$
\mathrm{P}_{\mathrm{A.39}}:\begin{cases}\max\ 6x_1+2x_2+9x_3+10x_4+x_5 & \\ \text{s.t.}\left| \begin{array}{lcll} 4x_1+x_2-5x_3-3x_4-2x_5\le 10 & | & 5x_1-x_2-7x_3+6x_4+7x_5\le 20 & \\ 3x_1+2x_2+9x_3-3x_4-10x_5\le 2 & | & 0\le x_i\le 10 & i=1,\ldots,5\end{array}\right.\end{cases}
$$

The optimal solution of $P_{A.39}$ is $x = (0, 10, 10, 6.25641, 8.92308)$ and its value is 181.4872. Now, suppose that all the coefficients of the constraints are uncertain. The values of the coefficients $\hat{a}_{ij}$ are given by the matrix below.

$$(\hat{a}_{ij}) = \begin{pmatrix} 0.8 & 0.2 & 1.0 & 0.6 & 0.4 \\ 0.6 & 0.4 & 1.8 & 0.6 & 2.0 \\ 1.0 & 0.2 & 1.4 & 1.2 & 1.4 \end{pmatrix}$$

Table A.1 gives the optimal robust solutions and their values for different values of the parameter Γ_i. In this example, it is assumed that this parameter is not dependent on i and we set $\Gamma = \Gamma_i$ for all i.

Table A.1 shows that when the uncertainty is substantial ($\Gamma = 5$), the protection cost against this uncertainty is very high since the value of the economic function decreases, for example, from 91.0816 when $\Gamma = 1$, to 44.0547 when $\Gamma = 5$ (about -52%). Note that in the case where $\Gamma = 5$ the optimal robust solution can be calculated by using the formulation $P_{A.35}$.

Tab. A.1 – Optimal robust solutions associated with program $P_{A.39}$, for different uncertainty domains defined by the parameter Γ.

Γ	Optimal robust solution	Value
1	(0, 0, 9.0204, 0, 9.8980)	91.0816
2	(0, 1.7778, 2.2222, 2.5926, 2.2222)	51.7037
3	(0, 3.3182, 0.7374, 3.0379, 0.6636)	44.3156
4	(0, 0, 0.9701, 3.5323, 0)	44.0547
5	(0, 0, 0.9701, 3.5323, 0)	44.0547

The approach can be extended to mixed-integer linear programs. Consider program $P_{A.40}$ in which some variables are integer while others are real.

$$P_{A.40}: \begin{cases} \max \sum_{j=1}^{n} c_j x_j \\ \text{s.t.} \left| \begin{array}{lll} \sum_{j=1}^{n} a_{ij} x_j \leq b_i & i = 1,\ldots,m & \text{(A.40.1)} \\ x_j \in \mathbb{N} & j = 1,\ldots,p & \text{(A.40.2)} \\ x_j \geq 0 & j = p+1,\ldots,n & \text{(A.40.3)} \end{array} \right. \end{cases}$$

In this case, the optimal robust solution can be determined by solving the mixed-integer linear program $P_{A.41}$.

$$\mathrm{P}_{\mathrm{A.41}}:\begin{cases}\max \sum_{j=1}^{n} c_j x_j \\ \text{s.t.}\left|\begin{array}{lll}\sum_{j=1}^{n} a_{ij}x_j + \sum_{j\in J_i} p_{ij} + \Gamma_i z_i \le b_i & i=1,\ldots,m & \text{(A.41.1)}\\ z_i + p_{ij} \ge \hat{a}_{ij}x_j & i=1,\ldots,m, j\in J_i & \text{(A.41.2)}\\ x_j\in\mathbb{N} & j=1,\ldots,p & \text{(A.41.3)}\\ x_j\ge 0 & j=p+1,\ldots,n & \text{(A.41.4)}\\ z_i \ge 0 & i=1,\ldots,m & \text{(A.41.5)}\\ p_{ij}\ge 0 & i=1,\ldots,m, j\in J_i & \text{(A.41.6)}\end{array}\right.\end{cases}$$

In everything we have just seen, the uncertainty affecting some coefficients is defined by intervals. We now consider the case where the uncertainty is represented by a set of (discrete) scenarios. A scenario is a set of assumptions about the evolution of the factors that may influence the value of the coefficients, and several scenarios are possible. In such an approach, the value of the different coefficients of the mathematical program considered depends on the scenario.

Consider the linear program $\mathrm{P}_{\mathrm{A.42}}$ where the coefficients a_{ij} are uncertain.

$$\mathrm{P}_{\mathrm{A.42}}:\begin{cases}\max \sum_{j=1}^{n} c_j x_j \\ \text{s.t.}\left|\begin{array}{lll}\sum_{j=1}^{n} a_{ij}x_j \le b_i & i=1,\ldots,m & \text{(A.42.1)}\\ x_j\ge 0 & j=1,\ldots,n & \text{(A.42.2)}\end{array}\right.\end{cases}$$

A set of scenarios, $\mathrm{Sc} = \{\mathrm{sc}_1, \mathrm{sc}_2,\ldots, \mathrm{sc}_p\}$, is envisaged and the values of the coefficients a_{ij} depend on the scenario. It is assumed that for each of these p scenarios the values of all the coefficients a_{ij} are known. For $i = 1,\ldots, m$, $j = 1,\ldots, n$, and $\omega = 1,\ldots, p$, a_{ij}^{ω} denotes the value of the coefficient a_{ij} in the case of the scenario sc_{ω}.

The problem of finding a solution to $\mathrm{P}_{\mathrm{A.42}}$ that is feasible for all the scenarios and that is the least costly – an optimal robust solution – can then be formulated as $\mathrm{P}_{\mathrm{A.43}}$.

$$\mathrm{P}_{\mathrm{A.43}}:\begin{cases}\max \sum_{j=1}^{n} c_j x_j \\ \text{s.t.}\left|\begin{array}{lll}\sum_{j=1}^{n} a_{ij}^{\omega}x_j \le b_i & i=1,\ldots,m;\ \omega=1,\ldots,p & \text{(A.43.1)}\\ x_j\ge 0 & j=1,\ldots,n & \text{(A.43.2)}\end{array}\right.\end{cases}$$

We have just shown how to take into account some uncertainty about the coefficients a_{ij} of program $P_{A.42}$. Let us now consider how to take into account uncertainty about the coefficients of the economic function, c_j. Consider the linear program $P_{A.42}$ where the coefficients c_j are uncertain. We consider a set of possible scenarios, $\mathrm{Sc} = \{\mathrm{sc}_1, \mathrm{sc}_2, \ldots, \mathrm{sc}_p\}$, and denote the value of the coefficient c_j in the case of the scenario sc_ω as c_j^ω. Several robustness criteria can be considered (see, for example, Kouvelis and Yu, 1997). We consider here a "max–min" criterion to measure the quality of a solution. In this approach, a solution is better than all others if its worst performance – over all scenarios – is better than the worst performance of all other solutions. We first illustrate this robustness criterion on an optimization problem in graphs (example A.10) and then formulate the search for an optimal robust solution for program $P_{A.42}$ with this robustness criterion.

Example A.10. Let us consider the problem of the path of minimum value in a graph with uncertainty in the arc values, this uncertainty being modelled by a set of possible scenarios. Note that this is a minimisation problem unlike program $P_{A.43}$. Let $G = (X, U)$ be a graph where $X = \{x_1, \ldots, x_n\}$ is the set of vertices and $U = \{a_1, \ldots, a_m\}$, the set of arcs. Let $\mathrm{Sc} = \{\mathrm{sc}_1, \mathrm{sc}_2, \ldots, \mathrm{sc}_p\}$ be the set of possible scenarios. For each scenario $\mathrm{sc}_\omega \in \mathrm{Sc}$, the value of the arc $a_i \in U$ is denoted by c_i^ω. The objective is to determine a path of minimum value, from vertex x_1 to vertex x_n, the value of a path being equal to the sum of the value of its arcs. Here, we use a min–max criterion, *i.e.*, we consider the problem of determining, among all the paths in the graph from x_1 to x_n, the one whose maximal length, over all the scenarios, is minimal. It is obviously possible to consider other criteria to take into account this uncertainty on the arc values. Let us consider an example of the problem with two scenarios.

The considered graph is represented by figure A.3 where each double arrow connecting two vertices x_i and x_j actually corresponds to the 2 symmetric arcs (x_i, x_j) and (x_j, x_i) with identical associated values. For each arc in the graph, the values for

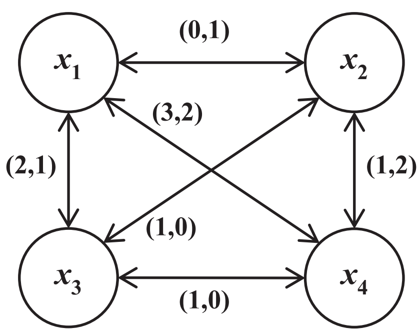


FIG. A.3 – A symmetric graph with 4 vertices. The value of each arc depends on the scenario and is indicated in brackets next to the arc: (value in the scenario sc_1, value in the scenario sc_2). For example, the value of the arc (x_3, x_4) and the arc (x_4, x_3) is equal to 1 for the scenario sc_1 and 0 for the scenario sc_2.

TAB. A.2 – List of all elementary paths, from x_1 to x_4, in the graph of figure A.3, with their values in both scenarios.

Name of the path	Description of the path	Value of the path in the scenario sc_1	Value of the path in the scenario sc_2
π_1	$x_1 \to x_2 \to x_3 \to x_4$	2	1
π_2	$x_1 \to x_2 \to x_4$	1	3
π_3	$x_1 \to x_3 \to x_2 \to x_4$	4	3
π_4	$x_1 \to x_3 \to x_4$	3	1
π_5	$x_1 \to x_4$	3	2

the scenarios sc_1 and sc_2 are given in brackets (arc value for the scenario sc_1, arc value for the scenario sc_2). Let us determine the optimal robust solution, by enumeration. Table A.2 gives, for all elementary paths from x_1 to x_4, their respective values in both scenarios.

We can deduce from table A.2 that, in this example, the optimal path is $\pi_1 = x_1 \to x_2 \to x_3 \to x_4$ and the optimal value is equal to 2.

The program for determining an optimal robust solution of program $\mathrm{P}_{\mathrm{A}.42}$, when the coefficients of the economic function, c_j, are uncertain, is $\mathrm{P}_{\mathrm{A}.44}$.

$$\mathrm{P}_{\mathrm{A}.44} : \begin{cases} \max \quad \alpha \\ \text{s.t.} \left| \begin{array}{lll} \alpha \leq \sum_{j=1}^{n} c_j^{\omega} x_j & \omega = 1,\ldots, p & (\mathrm{A}.44.1) \\ \sum_{j=1}^{n} a_{ij} x_j \leq b_i & i = 1,\ldots, m & (\mathrm{A}.44.2) \\ x_j \geq 0 & j = 1,\ldots, n & (\mathrm{A}.44.3) \end{array} \right. \end{cases}$$

Choosing the best possible solution in the worst-case scenario can have a significant drawback. Indeed, if one of the scenarios is very pessimistic – regarding the value of the economic function coefficients –, then the solution chosen will essentially take into account this single scenario. To overcome this disadvantage, other criteria can be chosen to evaluate a solution of $\mathrm{P}_{\mathrm{A}.42}$ when the coefficients of the economic function are uncertain. For example, we may be interested in the solution that minimizes the largest relative gap or "regret" – over all scenarios – between the value of the selected solution and the value of the optimal solution in the scenario under consideration. To solve this problem, one must first determine the optimal solutions in each scenario, *i.e.*, solve program $\mathrm{P}_{\mathrm{A}.45}(\omega)$ for each scenario, *i.e.*, for $\omega = 1,\ldots, p$.

$$\mathrm{P}_{\mathrm{A}.45}(\omega) : \begin{cases} \max \sum_{j=1}^{n} c_j^{\omega} x_j \\ \text{s.t.} \left| \begin{array}{lll} \sum_{j=1}^{n} a_{ij} x_j \leq b_i & i = 1,\ldots, m & (\mathrm{A}.45_{\omega}.1) \\ x_j \geq 0 & j = 1,\ldots, n & (\mathrm{A}.45_{\omega}.2) \end{array} \right. \end{cases}$$

Let $V^{*\omega}$ be the optimal value of program $\mathrm{P}_{\mathrm{A}.45}(\omega)$ – for the scenario sc_{ω}. The optimal robust solution can be determined by solving program $\mathrm{P}_{\mathrm{A}.46}$.

$$
\mathrm{P}_{\mathrm{A.46}}: \begin{cases} \min\ \alpha \\ \text{s.t.} \left| \begin{array}{lll} \alpha \geq \left(V^{*\omega} - \sum\limits_{j=1}^{n} c_j^{\omega} x_i \right) \Big/ V^{*\omega} & \omega = 1,\ldots,p & \text{(A.46.1)} \\ \sum\limits_{j=1}^{n} a_{ij} x_j \leq b_i & i = 1,\ldots,m & \text{(A.46.2)} \\ x_j \geq 0 & j = 1,\ldots,n & \text{(A.46.3)} \end{array} \right. \end{cases}
$$

Robust optimization is a rapidly growing branch of mathematical optimization that attempts to solve an optimization problem by taking into account as best as possible the various uncertainties that affect it. For a more detailed presentation of the basics of this optimization field, the reader can consult the references cited below.

References and Further Reading

Ben-Tal A., El Ghaoui L., Nemirovski A. (2009) *Robust optimization.* Princeton University Press.

Bertsimas D., Sim M. (2004) Price of robustness, *Oper. Res.* **52**, 35.

Bertsimas D., Brown B., Caramanis C. (2011) Theory and applications of robust optimization, *SIAM Rev.* **53**, 464.

Billionnet A., Costa M-C., Poirion P.-L. (2014) 2-Stage Robust MILP with continuous recourse variables, *Discr. Appl. Math.* **170**, 21.

Gabrel V., Murat C., Thiele A. (2014) Recent advances in robust optimization: an overview, *Eur. J. Oper. Res.* **235**, 471.

Kouvelis P., Yu G. (1997) *Robust discrete optimization and its applications.* Kluwer Academic Publishers.

Roy B. (2007) La robustesse en recherche opérationnelle et aide à la décision: une préoccupation multi facettes, *Annales du LAMSADE* **7**.

A.9 Set-Covering and Set-Partitioning Problems

We consider a set of elements, $E = \{e_1, e_2,\ldots, e_m\}$, and a set of parts of E, $F = \{F_1, F_2,\ldots, F_n\}$. With each element, F_j, of F is associated a cost, c_j. The set covering problem consists in determining a subset of F, of minimal cost, which covers all the elements of E. In other words, the set covering problem consists in determining $X \subseteq F$ such that $\cup_{j\,:\,F_j \in X} F_j = E$, and which minimizes the cost of X, that is the quantity $\sum_{j\,:\,F_j \in X} c_j$. The set X is said to be a cover for E.

One can also look at minimal covers in the inclusion sense. A cover, X, of E is minimal in the inclusion sense if there are no other covers of E strictly included in X.

Mathematical program associated with the set-covering problem. With each element F_j of F, is associated a Boolean variable, x_j, which, by definition, is equal to 1 if and only if F_j is selected to form the minimal cost cover, X, of E, *i.e.*, if $F_j \in X$. Let J_i be the set of indices j belonging to $\{1,\ldots, n\}$ and such that $e_i \in F_j$. The set-covering problem – finding a minimal cost cover – can be formulated as the linear program in Boolean variables $\mathrm{P}_{\mathrm{A.47}}$.

$$\mathrm{P}_{\mathrm{A.47}}: \begin{cases} \min \sum\limits_{j=1}^{n} c_j x_j \\ \text{s.t.} \left| \begin{array}{lll} \sum\limits_{j \in J_i} x_j \geq 1 & i = 1,\ldots,m & \text{(A.47.1)} \\ x_j \in \{0,1\} & j = 1,\ldots,n & \text{(A.47.2)} \end{array} \right. \end{cases}$$

If the inequality constraints in $\mathrm{P}_{\mathrm{A.47}}$ are replaced by equality constraints, the resulting program is associated with what is called a set-partitioning problem. This problem indeed consists in determining a set of elements of F of minimal cost, and which form a partition of E.

Example A.11. Consider the set-covering problem in which $E = \{e_1, e_2, e_3, e_4, e_5\}$, $F = \{F_1, F_2, F_3, F_4\}$ with $F_1 = \{e_1, e_2\}$, $F_2 = \{e_2, e_3, e_5\}$, $F_3 = \{e_2, e_4, e_5\}$, $F_4 = \{e_3, e_4\}$ and $c = \{3, 4, 5, 2\}$. The associated linear program in Boolean variables is $\mathrm{P}_{\mathrm{A.48}}$.

$$\mathrm{P}_{\mathrm{A.48}}: \begin{cases} \min \; 3x_1 + 4x_2 + 5x_3 + 2x_4 \\ \text{s.t.} \left| \begin{array}{lcl} x_1 \geq 1 & | & x_3 + x_4 \geq 1 \\ x_1 + x_2 + x_3 \geq 1 & | & x_2 + x_3 \geq 1 \\ x_2 + x_4 \geq 1 & | & x_1, x_2, x_3, x_4 \in \{0,1\} \end{array} \right. \end{cases}$$

The very particular structure of the programs associated with set-covering and set-partitioning problems make that many simple and effective pre-processing operations are possible. Suppose, for example, that the set of indices, J_r, appearing in a constraint r is contained in the set of indices, J_s, appearing in a constraint s. In the case of the set-covering problem, constraint s can be removed. In the case of the set-partitioning problem, we can set to 0 all variables whose indices belong to J_s without belonging to J_r, and remove the constraint s. Thus, in program $\mathrm{P}_{\mathrm{A.48}}$, the second constraint can be removed. Furthermore, the resolution of the continuous relaxation of $\mathrm{P}_{\mathrm{A.47}}$ often results in an optimal solution in which all the variables take integer values – 0 or 1. The continuous relaxation of program $\mathrm{P}_{\mathrm{A.47}}$ is obtained by replacing in this program the constraints $x_j \in \{0,1\}$, $j = 1,\ldots,n$, by the constraints $0 \leq x_j \leq 1$, $j = 1,\ldots,n$. If this happens, the resolution of $\mathrm{P}_{\mathrm{A.47}}$ is particularly easy since it can be deduced that the optimal solution of the continuous relaxation of $\mathrm{P}_{\mathrm{A.47}}$ – a continuous linear program – is the optimal solution of $\mathrm{P}_{\mathrm{A.47}}$.

Set-covering and partitioning problems are two important issues in operations research with many applications. For more information on the different properties of these problems and how to approach their resolution, the reader can consult the references mentioned below.

References and Further reading

Al-Sultan K.S., Hussain M.F., Nizami J.S. (1996) A genetic algorithm for the set covering problem, *J. Oper. Res. Soc.* **47**, 702.

Balas E., Padberg M.W. (1976) Set partitioning: a survey, *SIAM Rev.* **18**, 710.

Beasley J.E. (1987) An algorithm for set covering problems, *Eur. J. Oper. Res.* **31**, 85.

Beasley J.E., Chu P.C. (1996) A genetic algorithm for the set covering problem, *Eur. J. Oper. Res.* **94**, 392.

Billionnet A. (1978) Transformation du problème de partitionnement en un problème d'ensemble stable de poids maximal, *Revue d'Automatique, d'Informatique et de Recherche Opérationnelle (RAIRO), série verte* **12**, 319.

Billionnet A. (1981) Une nouvelle méthode pour le problème de partitionnement fondée sur une évaluation par excès de la solution, *RAIRO Recherche Opérationnelle* **15**, 139.

Caprara A., Toth P., Fischetti M. (2000) Algorithms for the set covering problem, *Ann. Oper. Res.* **98**, 353.

Chen D.S., Batson R.G., Dang Y. (2010) *Applied integer programming: modeling and solution.* John Wiley & Sons.

Chvatal V. (1979) A greedy heuristic for the set covering problem, *Math. Oper. Res.* **4**, 233.

Galinier P., Hertz A. (2007) Solution techniques for the large set covering problem, *Discr. Appl. Math.* **155**, 312.

Garfinkel R.S., Nemhauser G.L. (1969) The set-partitioning problem: set covering with equality constraints, *Oper. Res.* **17**, 848.

Garfinkel R.S., Nemhauser G.L. (1972) *Integer programming.* John Wiley & Sons.

Vermuganti R.R. (1998) Application of set covering, set packing, and set partitioning models: a survey, *Handbook of combinatorial optimization* (D.Z. Du, P.M. Pardalos, Eds). Springer, pp. 573–746.

A.10 Elements of Graph Theory

An undirected graph, G, is a pair, (X, E), composed of a set of vertices, $X = \{x_1, x_2, \ldots, x_n\}$, and a set of edges, E, each edge connecting two vertices of X called the ends of the edge. In general, two vertices can be connected by more than one edge. If this is the case, we are dealing with a multi-graph. In the rest of this section we consider simple graphs. In such graphs, two vertices are connected by no more than one edge and there is no loop, *i.e.*, an edge whose two ends are identical. Each edge is therefore defined by a pair of distinct vertices, $\{x_i, x_j\}$. The two vertices x_i and x_j are said to be adjacent. The degree of a vertex is equal to the number of edges of which this vertex is one end. An adjacency matrix can be associated with G. It is a $n \times n$-matrix, M, whose general term, m_{ij}, is equal to 1 if and only if vertices x_i and x_j are adjacent. If these two vertices are not adjacent $m_{ij} = 0$.

A directed graph – or oriented graph – G is a pair, (X, A), composed of a set of vertices, $X = \{x_1, x_2, \ldots, x_n\}$, and a set of arcs, A, each arc being defined by a – oriented – pair of vertices, (x_i, x_j). One says that x_i is the initial end of the arc, that x_j is its terminal end, that x_i is a predecessor of x_j, and that x_j is a successor of x_i. An arc whose two ends are identical is called an oriented loop. We are interested here in oriented graphs without loops – oriented – and for which, for any pair of vertices (x_i, x_j), there is at most one arc going from x_i to x_j. The indegree of a vertex is the number of arcs of which this vertex is the terminal end, and the outdegree of a vertex is the number of arcs of which this vertex is the initial end.

Graphs are so named because they can be represented graphically. Each vertex is represented by a point, each edge by a line connecting its ends, *i.e.*, two points, each arc by an arrow from its initial end to its terminal end. A graph can be drawn in several ways: the positions of the points representing the vertices and the shape of the lines or arrows connecting these vertices can vary.

Let $G = (X, U)$ be a directed or undirected graph. An induced sub-graph of G is a graph having for vertices a subset, $\hat{X}$, of the vertices of G, and for arcs/edges only those of G joining the vertices of $\hat{X}$; a partial sub-graph of G is a graph having for vertices a subset, $\hat{X}$, of the vertices of G and for arcs/edges a subset of those of G joining the vertices of $\hat{X}$. In an oriented graph, a path originating at x_i and ending at x_j is defined by a sequence of consecutive arcs, connecting x_i to x_j. If x_i and x_j are identical, then we have a circuit. In the oriented graph in figure A.4, the 3 arcs $(x_2, x_3), (x_3, x_4)$, and (x_4, x_2) define a circuit.

In a graph, oriented or not, a chain connecting x_i to x_j is a sequence of arcs or edges placed end to end, and connecting these two vertices. A chain connecting x_i to x_j is a cycle if x_i and x_j are identical and if the edges of the chain are all distinct. The length of a chain is the number of edges or arcs that compose it, and the length of a path is the number of arcs that compose it. In the undirected graph of figure A.4, the 4 edges $\{x_1, x_2\}, \{x_2, x_5\}, \{x_5, x_4\}$, and $\{x_4, x_6\}$ form a chain connecting vertices x_1 and x_6 and in the directed graph of the same figure, the 4 arcs $(x_2, x_1), (x_2, x_3), (x_3, x_4)$, and (x_6, x_4) form a chain connecting these same two vertices.

A graph – oriented or not – is connected if and only if any pair of vertices is linked by a chain. The distance between two vertices of a connected graph is the length of the chain that links them, with the smallest number of edges – or arcs. The diameter of a connected graph is the greatest distance between two vertices of that graph, among all pairs of vertices. A real value – sometimes called a weight – can be associated with each arc of G. The value of a path/chain is then equal to the sum of the values of the arcs/edges that compose it. A connected component of a graph G is a sub-graph, G_0, of G, which is connected and maximal in the inclusion sense – no other connected sub-graph of G contains G_0.

An oriented graph is strongly connected if and only if for any oriented pair of vertices, (x_i, x_j), there is a path from x_i to x_j. A strongly connected component of an oriented graph, G, is a sub-graph, G_0, of G, which is strongly connected and maximal in the inclusion sense – no other strongly connected sub-graph of G contains G_0. The two graphs in figure A.4 are connected. The sub-graph of the oriented graph in

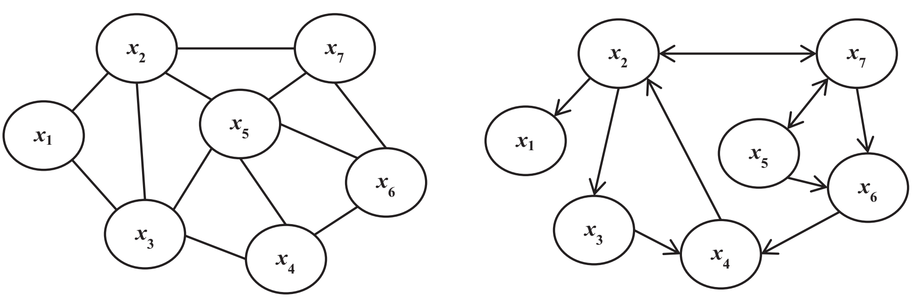


FIG. A.4 – An example of an undirected graph and a directed graph with 7 vertices. Here, a double arrow connecting two vertices x_i and x_j – for example x_2 and x_7 – means that the graph includes an arc from x_i to x_j and an arc from x_j to x_i.

this figure, induced by the vertex set $\{x_2, x_3, x_4, x_5, x_6, x_7\}$, is a strongly connected component of this graph.

A tree is an undirected graph that has no cycle and is connected (figure A.5). A tree can be defined in many ways. For example, G is a tree if G is without cycles and has n−1 edges, or G is a tree if G is connected and has n−1 edges – n denotes the number of vertices of the graph.

If in a tree a particular vertex, r, is chosen and the edges of this tree are oriented so that there is a – unique – path from r to all other vertices, one obtains an arborescence of root r (figure A.6).

We can also define an arborescence as an oriented graph without circuits admitting a particular vertex, r, called root, and such that there is a single path from r to all the other vertices of the graph.

Given a connected undirected graph, G, a spanning tree of G is a partial sub-graph of G whose vertices are those of G, and which is a tree. A classical problem, when values are assigned to the edges of G, is to determine a spanning tree of minimal value, the value of a tree being equal to the sum of the values of its edges. There are efficient algorithms to solve this problem. One can also look at the

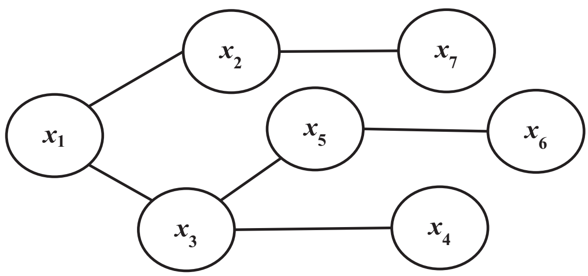


FIG. A.5 – An example of a tree with 7 vertices.

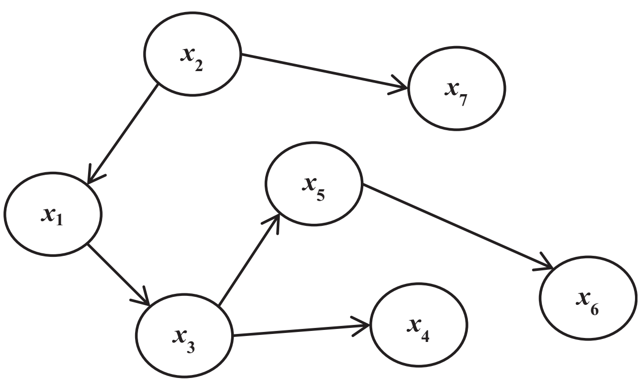


FIG. A.6 – An example of an arborescence constructed from the tree in figure A.5 and whose root is vertex x_2.

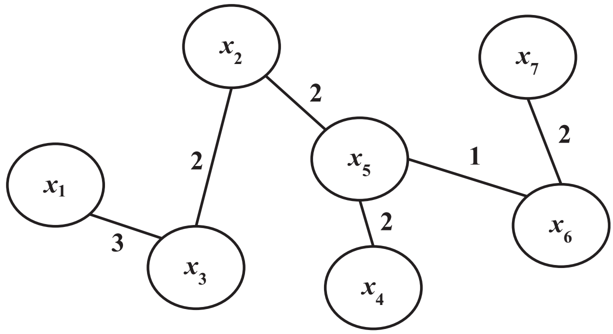


FIG. A.7 – A spanning tree of minimal value (12) associated with the graph in figure A.8.

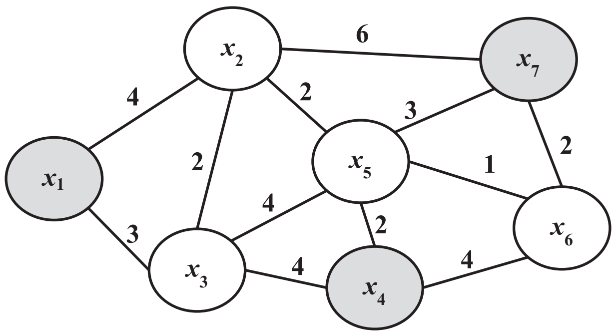


FIG. A.8 – A connected undirected graph with values associated to each edge and a set of mandatory vertices $\hat{X} = \{x_1, x_4, x_7\}$.

spanning tree of minimal value in an oriented graph. The tree in figure A.7 is a spanning tree of minimal value for the graph in figure A.8.

Consider an undirected graph, $G = (X, E)$, and a subset of vertices, $\hat{X}$, included in X. With each edge $\{x_i, x_j\}$ of E, is associated a positive or zero value. The Steiner tree problem consists in determining a partial sub-graph of G that includes all the vertices of $\hat{X}$, which is a tree, and whose value is as small as possible. Recall that the value of a tree is equal to the sum of the values of its edges. This problem is usually difficult. Consider the graph in figure A.8. The values of the edges are shown next to the edges. Suppose that the set $\hat{X}$ consists of vertices x_1, x_4, and x_7 – the required vertices. The Steiner tree of minimal value is given in figure A.9.

A transportation network is defined by an oriented graph, $G = (X, A)$, with two particular vertices, x_1, which is the source of the network and x_n, which is its sink. To simplify the presentation, it is assumed that no arc ends at x_1 and no arc starts at x_n. Each arc in the graph has an associated capacity. It is assumed that there is no useless vertex, *i.e.*, for any vertex x_i of X, there is a path from x_1 to x_n passing through x_i. In a transportation network, G, a flow is the assignment of a

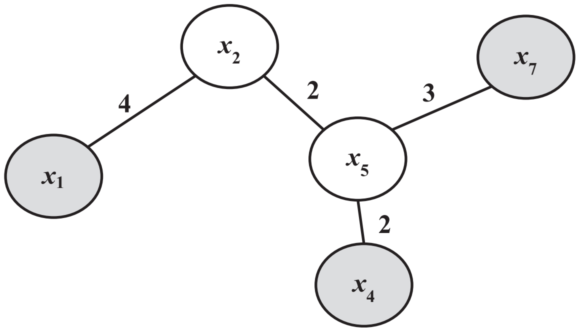


FIG. A.9 – A Steiner tree of minimal value (11) associated with the graph in figure A.8 when the set of mandatory vertices is $\hat{X} = \{x_1, x_4, x_7\}$.

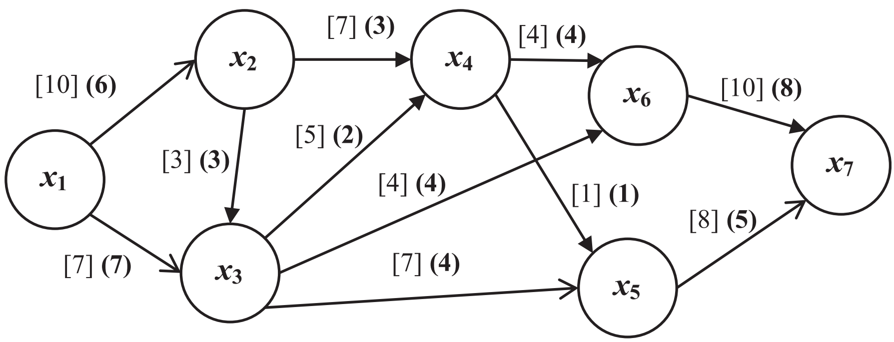


FIG. A.10 – A transportation network whose source is vertex x_1 and sink is vertex x_7, and a flow of value 13 on this network.

non-negative real value to each arc of G – the flow on that arc – which can be interpreted, for example, as a quantity of matter transported on that arc, such that, in each vertex, the sum of the incoming flows – on the arcs of which this vertex is the terminal end – is equal to the sum of the outgoing flows – on the arcs of which this vertex is the initial end. This flow must take into account the capacity assigned to each arc, this capacity reflecting an upper limit of the flow allowed on that arc. The value of the flow is equal to the sum of the flows emanating from x_1 or entering x_n. It is easy to show that these two quantities are equal. A classical problem, for which efficient algorithms exist and which has many applications, consists in determining a flow of maximal value on the considered network. Let us consider the transportation network in figure A.10. The capacities of the arcs are given in square brackets next to the arcs. A flow from x_1 to x_7, of value 13, is shown in figure A.10. The corresponding flow of each arc is noted between brackets next to it.

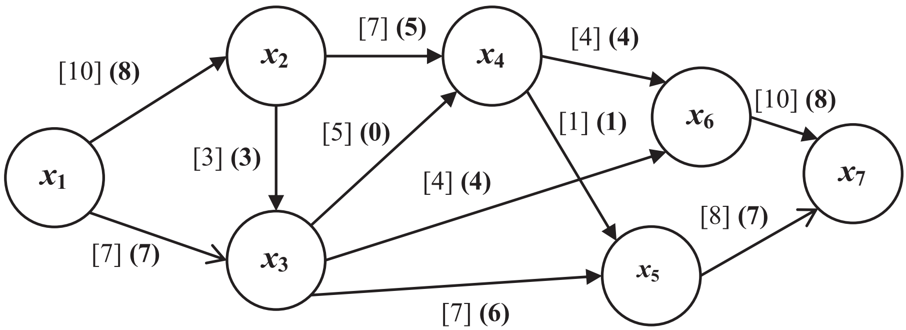


FIG. A.11 – A flow of maximal value (15) on the transportation network in figure A.10.

The flow indicated in figure A.10 is not a maximal flow because there is a flow with a value of 15 from x_1 to x_7 (figure A.11). It can be shown that the value of this new flow is maximal.

Publications concerning graph theory, a branch of mathematics in its own right, are extremely numerous. For more information on the basic notions of this very dynamic discipline, the reader can consult the references cited below.

References and Further Reading

Ahuja R., Magnanti T., Orlin J. (1993) *Network flows: Theory, algorithms, and applications.* Prentice Hall.

Berge C. (1958) *Théorie des graphes et ses applications.* Dunod.

Berge C., Ghouila-Houri A. (1962) *Programmes jeux et réseaux de transport.* Dunod.

Berge C. (1970) *Graphes et Hypergraphes.* Dunod.

Bondy A., Murty U. (2008) Graph theory. Graduate texts in mathematics 244, Springer, Traduction française disponible: https://www-sop.inria.fr/members/Frederic.Havet/Traduction-Bondy-Murty.pdf.

Bretto A., Faisant A., Hennecart F. (2018) *Éléments de théorie des graphes.* Hermes Science Publications.

Promel H.G., Steger A. (2002) *The Steiner tree problem: a tour through graphs, algorithms, and complexity.* Vieweg.

Roseaux (collective of authors) (2005) Exercices et problèmes résolus de recherche opérationnelle (Tome 1), *Graphes, leurs usages, leurs algorithmes.* Dunod.

Roy B. (1969–1970) Algèbre moderne et théorie des graphes orientées vers les sciences économiques et sociales, *Tome 1: notions et résultats fondamentaux, Tome 2: applications et problèmes spécifiques.* Dunod.

Sessions J. (1992) Solving for network connections as a Steiner network problem, *For. Sci.* **38**, 203.

West D. (2000) *Introduction to graph theory.* Prentice Hall.

A.11 Markov Chains

A Markov chain allows to model the dynamic evolution of a random system with N states, S_1, S_2,…, S_N. The system – or the chain – is initially in one of the states, and passes successively from one state to another. Each movement constitutes a step or a transition. If the chain is, at a given instant, in state S_i, then it passes into state

S_j, at the following instant with a probability denoted by p_{ij}, and this probability does not depend on the state in which the chain was previously – nor on the considered instant for a homogeneous chain. The probabilities p_{ij} are called transition probabilities. The process can also remain in state S_i in which it was, and this happens with a probability p_{ii}. If this probability is equal to 1, state S_i is said to be absorbing. The set of these transition probabilities constitutes the transition probability matrix. A graph can be easily associated with this matrix. Figure A.12 presents such a graph for a chain defined on 5 states. Note that, for all $i \in \{1,\ldots, N\}$, $\sum_{j=1}^{N} p_{ij} = 1$.

The starting state of the chain is defined by the initial probability distribution of states S_1, S_2,..., S_N. This is often done by specifying a particular state as the starting state. Markov chains allow a large number of situations to be modelled in a variety of fields. The study of a Markov chain aims at studying the evolution of the system described by this chain. One can be interested, for example, in the probability of being in state S_j at the end of p transitions when the initial state is S_i. One can also seek to determine the probability distribution of the states after a very large (infinite) number of transitions – if this limiting distribution exists. One can also be interested in the probability of entering state S_j for the first time after p transitions, starting from state S_i. Some chains are said to be absorbing. In this case, there is at least one absorbing state and, from any non-absorbing state, an absorbing state can be reached in one or several transitions. For any absorbing Markov chain and for any starting state, the probability of being in an absorbing state after p transitions tends to 1 when p tends to infinity. In such chains, one can look at the probability of ending up in a given absorbing state – if there are several absorbing states – or at the expected number of transitions through the non-absorbing state S_j, starting from the non-absorbing state S_i, before ending up in an absorbing state. One can also look at the number of transitions it will take on average to reach an absorbing state, taking into account the initial state of the chain.

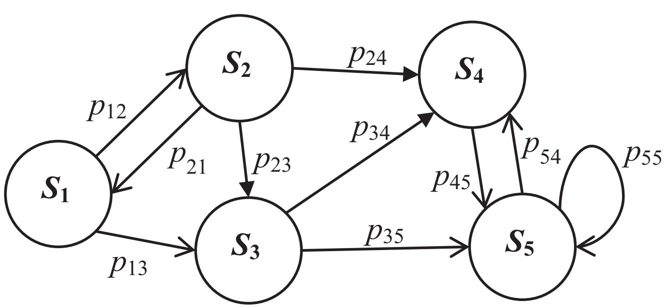


FIG. A.12 – Graph associated with a Markov chain with 5 states S_1, S_2,..., S_5. If, at the time n, the chain is in state S_5, then at the time $n + 1$ it will be in one of the 2 following states: again in state S_5 with the probability p_{55} or in state S_4 with the probability p_{54} ($p_{54} + p_{55} = 1$).

Example A.12. Consider a Markov chain with the 5 states, S_1, S_2, S_3, S_4, and S_5, and whose transition probability matrix is the matrix M below. The value at the intersection of row i and column j is the transition probability from state S_i to state S_j, *i.e.*, the probability p_{ij}. Thus, when the chain is at the instant n in state S_3, it can be at the instant $n+1$ either in state S_1, with the probability 0.4, or in state S_3, with the probability 0.4, or in state S_4, with the probability 0.1, or in state S_5, with the probability 0.1. This chain has 3 transient states, S_1, S_2, and S_3, and 2 absorbing states, S_4 and S_5. A state is transient if the system, being in this state, may not return to this state.

$$M = \begin{pmatrix} 0.4 & 0.5 & 0 & 0 & 0.1 \\ 0 & 0.5 & 0.3 & 0.1 & 0.1 \\ 0.4 & 0 & 0.4 & 0.1 & 0.1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Let us denote by Z the sub-matrix of transition probabilities between transient states and by D the sub-matrix of transition probabilities from a transient state to an absorbing state. In this example,

$$Z = \begin{pmatrix} 0.4 & 0.5 & 0 \\ 0 & 0.5 & 0.3 \\ 0.4 & 0 & 0.4 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} 0 & 0.1 \\ 0.1 & 0.1 \\ 0.1 & 0.1 \end{pmatrix}.$$

Let us denote by π_{ij} the expected number of passages through state S_j – transient – starting from state S_i – transient – before absorption, and by Π the matrix whose general term is π_{ij}. We can show that $\Pi = (I - Z)^{-1}$ where I designates the identity matrix of the same dimension as Z. In our example, Π is equal to the inverse of the matrix

$$I - Z = \begin{pmatrix} 1-0.4 & -0.5 & 0 \\ 0 & 1-0.5 & -0.3 \\ -0.4 & 0 & 1-0.4 \end{pmatrix}, \quad \textit{i.e.}, \quad \Pi = \begin{pmatrix} 2.5 & 2.5 & 1.25 \\ 1 & 3 & 1.5 \\ 5/3 & 5/3 & 2.5 \end{pmatrix}.$$

Thus, starting from state S_2, the system will go on average 3 times through this same state before absorption. Let us now look at the probability, being in state S_i, $i = 1, 2, 3$, of ending up in the absorbing state S_j, $j = 4, 5$. Let a_{ij} be this probability and A be the matrix of general term a_{ij}. We can show that $A = \Pi.D$. In our example,

$$A = \begin{pmatrix} 0.375 & 0.625 \\ 0.45 & 0.55 \\ 1.25/3 & 1.75/3 \end{pmatrix}.$$

Thus, starting from state S_1, the system will end up in the absorbing state S_5 with a probability equal to 0.625.

Markov chain theory has proven to be very effective in modelling and studying many concrete or theoretical random phenomena. For a more detailed presentation of this theory the reader can consult the references cited below.

References and Further Reading

Brémaud P. (2009) *Initiation aux probabilités et aux chaînes de markov*. Springer.
Ching W.-K., Ng M.K. (2006) *Markov chains: models, algorithms and applications*. Springer.
Douc R., Moulines E., Priouret P., Soulier P. (2019) *Markov chains*. Springer.
Gagniuc P.A. (2017) *Markov chains: from theory to implementation and experimentation*. Wiley.
Gordon P. (1964) *Théorie des chaînes de markov finies et ses applications*. Dunod, Paris.
Graham C. (2008) *Chaînes de markov - cours et exercices corrigés*. Dunod.
Grinstead C.M., Snell J.L. (1997) *Introduction to probability*, 2nd edition 2006. American Mathematical Society, Providence.
Kemeny J.G., Snell J.L. (1983) *Finite markov chains*, Springer.
Méléard S. (2016) *Modèles aléatoires en écologie et évolution*. Springer.
Privault N. (2018) *Understanding Markov chains, examples and applications*. Springer.
Roseaux (collective of authors) (2004) Exercices et problèmes résolus de recherche opérationnelle (Tome 2), *Phénomènes aléatoires en recherche opérationnelle*. Dunod, Paris.
Sericola B. (2013) *Chaînes de markov, theorie, algorithmes et applications*. Hermes-Lavoisier.

Index