# II.2

# Description of processes and corrections from observation to delivery

M. Ollivier[1,2], A. Deru[2], S. Chaintreuil[2], A. Ferrigno[2], A. Baglin[2], J.-M. Almenara[3], M. Auvergne[2], S. Barros[4], F. Baudin[1], P. Boumier[1], P.-Y. Chabaud[3], H. Deeg[5], P. Guterman[3,6], L. Jorda[3], R. Samadi[2], and T. Tuna[2,7]

[1] Institut d'Astrophysique Spatiale, UMR 8617, CNRS/Univ. Paris-Sud, Université Paris-Saclay, Bât 121, Université Paris-Sud, 91405 Orsay cedex, France

[2] LESIA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, UPMC Univ. Paris 06, Univ. Paris Diderot, Sorbonne Paris Cité, 5 place Jules Janssen, 92195 Meudon, France

[3] Laboratoire d'Astrophysique de Marseille, UMR 7326, 13388, Marseille, France

[4] Centro de Astrofisica da Universidade do Porto, Rua das Estrelas, 4150-762 Porto, Portugal

[5] Instituto de Astrofisica de Canarias, Dept. de Investigacion, C/ Via Lactea s/n, 38200 La Laguna, Tenerife, Spain

[6] Division Technique INSU, BP 330, 83507, La Seyne cedex, France

[7] now at the Centre Européen de Recherche et d'Enseignement des Géosciences de l'Environnement, Aix en Provence, France

## 1. Introduction

The CoRoT observatory, launched the 27th of December 2006 stopped its observations the 2nd of November 2012 after the failure of the second readout detection chain. After 6 months of investigations, the scientific mission was officially ended the 20th of June 2013. The desorbitation procedure of the satellite has been initiated in March 2013, bringing the satellite from its initial orbit at 896 km to a polar orbit at an altitude of 650 km. CoRoT was completely passivated the 17th of June 2013 and should finally fall down on the Earth in about 35 yr.

During its observation phase, CoRoT recorded the photometric light curve of about 150 distinct bright stars (with a visible magnitude between 5.8 and 8) using the bright stars channel (also known as the seismology channel) and more than 160 000 faint stars (with a visible magnitude between 10.5 and 16) using the faint stars channel (also known as the exoplanet channel). Even if the instrument principle and design has been extensively detailed in the literature (Boisnard & Auvergne 2006), the aboard processing, the final data processing pipeline and the successive correction routines have never been extensively described. That is the aim of this paper. We thus present the global data generation, the correction philosophy and the associated algorithms. The detail of corrections applied to the different extends at each version of the data is described in Chaintreuil et al. (2015) in this book.

## 2. The CoRoT data generation and correction philosophy

The CoRoT camera is made of 4 E2V CCDs:

- two CCDs dedicated to the observation of a few bright stars, with a main scientific goal of asteroseismology (this channel was usually defined as the Asteroseismology channel). The image on the detector is slightly defocused to avoid the pixels saturation. The integration sampling rate is 1 readout per second for 5 bright stars per CCD. The acquisition is driven by the instrument clock.
- two CCDs dedicated to the observation of faint stars, with a main scientific goal of planets hunting (Exoplanet channel). About 6000 stars are simultaneously observed per CCD and their image is slightly dispersed on the detector using a double prism in the optical path. This leads to "chromatic" products where pixels are added in order to obtain 3 "colored" fluxes. Contrary to the bright stars channel, the acquisition sampling rate is 1 readout per 32 s. The acquisition at 32 s is driven by the platform clock. Data at 512-s sampling rate are obtained by stacking aboard 16 measurements at the 32-s sampling rate (see next section). Observation always begins at 512-s sampling rate; it is then updated during the observation run thanks to an algorithm that analyses

directly the data when they are acquired and detects transit candidates, the sampling rate is adjusted (alarm mode: Quentin et al. 2006; Surace et al. 2008; Chabaud et al. 2011; Bonomo et al. 2012). The sampling rate of each target can thus be changed during the observation run. That is why several targets (e.g. CoRoT-2b, Alonso et al. 2008) exhibits various sampling rates on a same light curve.

The CCDs are driven by 2 separate chains of electronics for redondancy. Each chain processes 1 CCD of each program. The CCDs are read half by half, and the gain and offset are adapted to each readout process.

The number of oversampled stars in the faint stars channel is limited by the global telemetry budget. At the beginning of the mission, where telemetry had to deal with the four CCDs, the number of oversampled stars was limited to 500 stars per CCD (over about 6000 observed stars). After the failure of one of the two detection chains in March 2009, the telemetry budget could strongly be relaxed and the number of oversampled targets was increased to reach 2000 stars on the remaining CCD.

In the same time, "imagettes" (small parts of the CCD) were acquired on both channels. For the bright stars channel, the star windows are $35 \times 35$ pixels large and correspond to the observed stars (5 per CCD). For the faint stars channel, 20 windows of $15 \times 10$ pixels per CCD sampled at 32 s could be acquired and downloaded. At the beginning of the project, they were considered to monitor in real time the evolution of the instrument and the quality of the image (engineering use), in order to anticipate a mismatch between the stellar images and the available photometric masks. After the failure of the first detection chain, the number of imagettes on the remaining faint stars CCD (E2) was increased to 40 and the imagettes were used for science. For instance, imagettes were put on slightly saturated stars ($V$ magnitude slightly higher than 10), in order to optimize the photometry measurement and get good S/N ratio light curves.

The CoRoT final data are the result of a complex process that happens at different levels:

- on board: some corrections are directly performed during the observation phase, and before the data transfer to the Earth;
- on the ground: thanks to two successive pipelines.

During the observation phase, the CoRoT photometric data are packed and transmitted to the ground from the satellite through telemetry packets. They are collected at CNES after reception at Kiruna, Wien, Aussagel, and Natal facilities. The telemetry packets are then gathered to generate the first level (0-level) of data named N0. The N0 data can thus be considered as the "raw data" and should be processed before distribution to the scientists.

The global correction philosophy, adopted by the project from the beginning, is to consider and correct only effects that can be identified, quantified and modeled because they can be associated to a well defined effect (e.g.: background, cross talk, temperature jumps...). It means that no systematics removal algorithm is used in the process (at least not before the end of the complete process), and no correction is applied if the origin, nature or amplitude of the effect to be corrected is not identified.

The raw data then enter two successive pipelines of corrections:

- the level 0 to level 1 pipeline (also named pipeline N0->N1): in this pipeline, corrections are made on the fly when data are available and do not require parameters that should be computed using the whole data set. It means that N1 data are ready more or less at the end of the observation run, and do not include any correction made with a further knowledge of the instrument behavior or ageing;
- the level 1 to level 2 pipeline (also named pipeline N1->N2): corrections are applied at the end of the run, and use quantities or parameters evaluated with the data of the complete run. In some cases, corrections were applied only at the end of the mission, thanks to a better understanding of the instrument, its response, its ageing...

In this paper, we will first describe the corrections made aboard, and then on the ground without distinguishing systematically whether they are performed on level 0 or level 1 data. The data contained in the legacy archive are level 2 data. They are available at CDS[1] at IAS[2], and associated mirror site at NASA and in Spain. Level 1 data however are available upon request either on the IAS archive or at CNES but require a deep understanding of the instrument and the observation conditions to be scientifically usable.

# 3. Onboard processes and corrections

During observation phases, several processes are run aboard. They gather a bunch of routines driving key-actions:

- acquisition of the astrophysical signal and other quantities (background, offsets...);
- performing basic signal processing elements: subtraction of biases, stacking of observations, sum of pixels contents, evaluation of observation quality...

Each channel (bright stars and faint stars) has its own set of corrections. In this section, we describe these processes. A more detailed description of each process including the complete algorithms can be found in the software User Requirements Documents written by Plasson (2009).

## 3.1. Bright stars channel

In the bright stars channel, fluxes are computed every second.

The Astero Scientific Processing Service (ASPS) is the service run on the bright stars channel during the observation runs. It is successively made of:

*The Astero Offset window Processing (AOP)*
Each offset window is used to determine the electronic mean level of the associated half-CCD updated at each exposure. The software computes for each offset window:

- the median value of each offset window;
- the number of bright and impacted pixels using a threshold applied to the deviation to the median offset value;

---

- the arithmetic mean of non impacted and non-bright pixels;
- the variance on non-impacted and non-bright pixels;
- the electronic offset value on each half CCD.

*The Astero Background Local Processing (ABLP)*
This processing is made to determine the mean and variance values of the local sky background. This value is computed for each window and updated at each exposure. The computation is based on a moving average and variance of the non-impacted pixels of the window, of the last 30 exposures. The software :

- subtracts the electronic offset to all the pixels;
- computes the median pixel intensity;
- computes the number of impacted and brilliant pixels;
- computes the arithmetic mean of non-impacted and non-bright pixels;
- computes the variance of non-impacted and non-bright pixels;
- computes the value of the sky background per pixel.

*The Astero Star window Main Processing (ASMP)*
This processing was made to perform noise corrections and a temporal photometry of each observed star using a mask method. In the case of a SAA crossing, the processing could be modified to detect and correct additional noises due to the proton impacts. As proton impacts did not disrupt the processing aboard, the parameters were set so that this mechanism was not functional. For each stellar window, the software:

- subtracts the electronic offsets to all the pixels;
- subtracts the sky background value to all the pixels;
- computes the barycenter (photocenter) of the windows;
- computes the total intensity (photometric process);
- writes the computed parameters in the corresponding product.

*The Astero Star window Accumulation Processing (ASAP)*
This processing was designed to record small parts of images called "imagettes" accumulated aboard. This processing was performed only on several selected astero windows and only if the previous ASMP processing did not detect errors during noise correction.

In conclusion, one should bear in mind that, in the case of bright stars, offsets windows as well as background windows are computed on-board and subtracted on-board to the flux of the stars.

## 3.2. Faint stars channel

Observations last 32 s; the ordinary behaviour is to accumulate 16 observations, leading to a 512-s sampling. Some chosen stars are observed at 32 s, in a so-called oversampling rate.

The Exo Scientific Processing Service (ESPS) is the service run on the faint stars channel during the observation runs. It is made of:

*Exo Monochromatic window Processing (EMP)*
This process computes the raw flux of all the exoplanet monochromatic windows. The software :

- sums the intensity of each image in the case of 32-s sampling rate. The result is the final intensity;

- performs data accumulation (16 times the previous process) in the case of data sampled at 512 s;
- computes the mean and standard deviation of each of the 16 images, in the case of data sampled at 512 s;
- computes the number of bright pixels.

*Exo Chromatic window Processing (ECP)*
This process computes the raw flux in each chromatic zone of all the chromatic windows. The software :

- computes the exo spectrum by adding the flux in each column of the mask;
- computes the blue, green and red fluxes by summing the flux of spectral elements in the specific chromatic zones defined by the chromatic mask limits, as defined during the observation setup phase, at a 32-s sampling rate;
- performs data accumulation (16 times the previous process) to get data sampled at 512 s;
- computes the mean and standard deviation of each chromatic flux of the 16 integrations, in the case of data sampled at 512 s.

Chromatic processing applies to the brightest stars (magnitude from 10 to 14) whereas monochromatic processing applies to fainter stars. Monochromatic processing also applies to offset and background windows; all types of windows are undifferentiated on-board. They are processed differently only on ground where offsets and backgrounds are subtracted from the light-curves of the stars.

*Exo Imagette window Processing (EIP)*
This processing is performed on each imagette window to transmit all the pixels of each exposure. Let us remind than an imagette is a $10 \times 15$ pixels window, acquired at a sampling rate of 1 exposure per 32 s.

# 4. Ground-based processes and corrections

In this section, we only describe the final version of the pipelines and associated algorithms and data. For a detailed list of the process versions, see Chaintreuil et al. (2015) in this book.

In this paper, we decided to detail only the effects and algorithms that have not been described extensively previously. This paper should thus be associated to other ones: Auvergne et al. (2009) that was published after 2 yr in orbit and Lapeyrere (2006) for the specific question of the detectors.

## 4.1. Chronology of ground based corrections

### 4.1.1. Bright stars channel data

After several corrections made aboard (cf. previous section), several corrections are applied to the bright stars channel data, thanks to the two ground-based pipelines generating level 1 and 2 data.

The successive global or specific processes leading to level 1 and 2 data and their various extends from level 0 data are described on Fig. II.2.1.
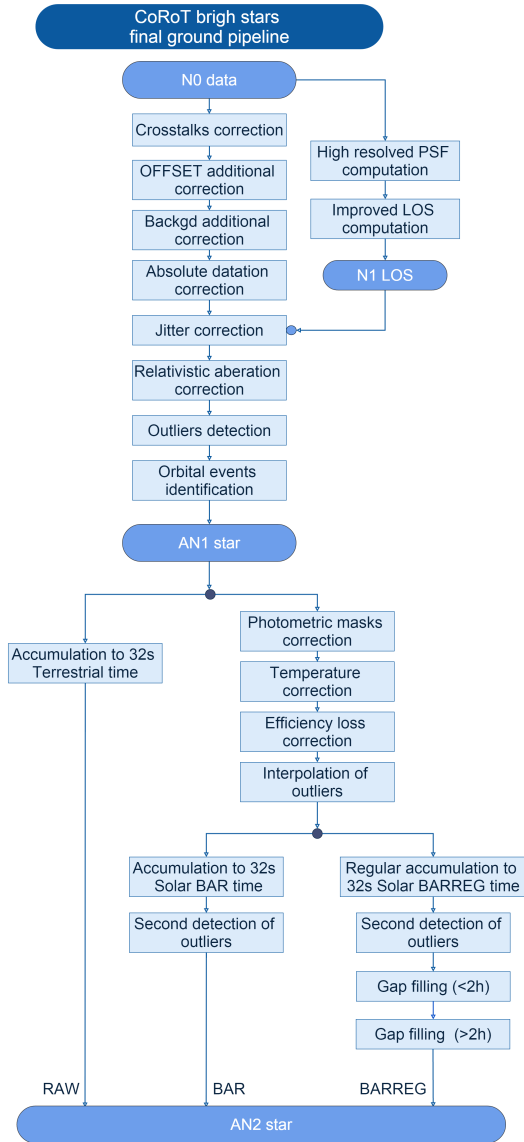
**Fig. II.2.1.** Synoptic view of the Level 1 and 2 data generation from N0 data, on the bright stars channel. AN1 star data are intermediary products corrected only using on-the-fly algorithms and do not include corrections that are made globally considering the whole run data.

## 4.1.2. Faint stars channel data

Contrary to the bright stars, no correction are performed aboard for the faint stars. The data corrections are made only from the ground and are detailed on Fig. II.2.2.

In the next subsections, we detail the main correction algorithms.

## 4.1.3. Faint stars channel imagettes

The faint stars channel imagettes are acquired and transmitted as they are to the ground where they are processed to get the light curves. The detail of the imagette pipeline is given on Fig. II.2.3. Note that the colored light curves once created are processed by the same pipeline as described in previous section (except for what concerns the creation of white photometry and the correction of jumps).
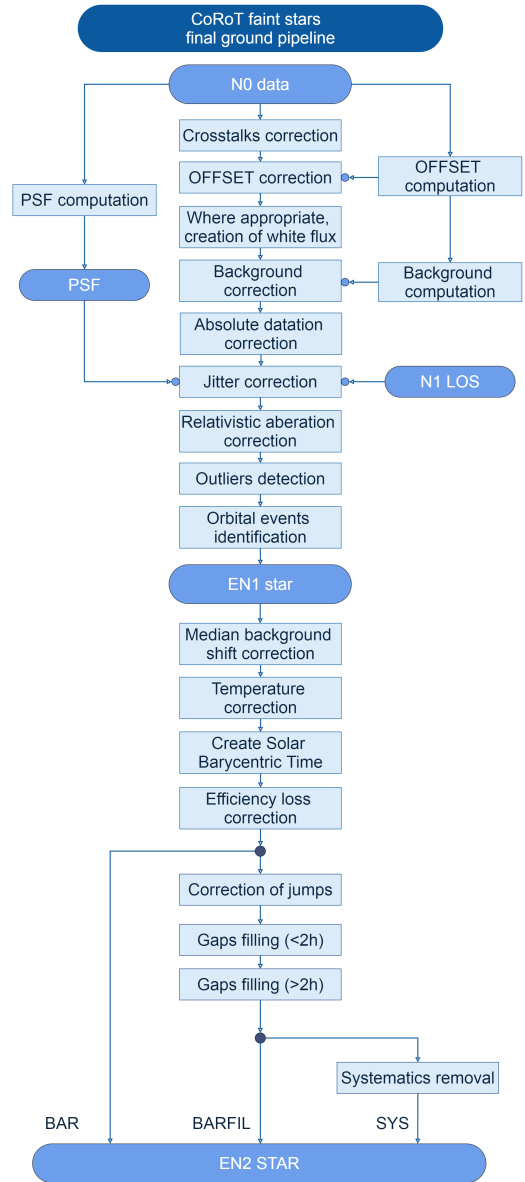


**Fig. II.2.2.** Synoptic view of the Level 1 and 2 data generation from N0 data, on the faint stars channel. EN1 star data are intermediary products corrected only using on-the-fly algorithms and do not include corrections that are made globally considering the whole run data.

Important fact: in the case of faint stars channel data, whatever they are extracted from (classical photometry or imagettes), several extends are created:

- BAR: data processed and time given in barycentric time;
- BARFILL: same as before but the "jumps" are corrected using a two step correction algorithm (see Sect. 4.2.14) and the gaps due to data loss (SAA crossing or telemetry loss) are filled using an Inpainting algorithm (see Sect. 4.2.15). These data are particularly useful when a continuous time serie is required for analysis;
- SYS: it is BARFILL data plus a systematics removal algorithm (see Sect. 4.2.16). These data are particularly suited for research programs that require very clean and continuous data.
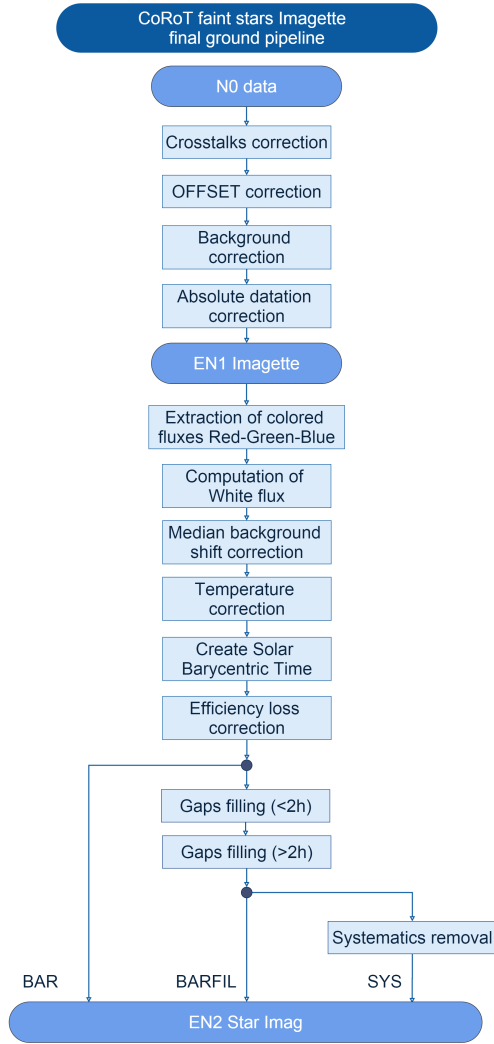
Fig. II.2.3. Synoptic veiw of the imagette pipeline.

An exemple of the same light curve but in its different extends BAR, BARFILL and SYS is given on Fig. II.2.4

## 4.2. Correction algorithms

### 4.2.1. Elimination of the aliasing (crosstalk)

This process concerns both channels data. Crosstalk results from the fact that bright stars and faint stars CCDs are not read and operated the same way aboard:

- bright stars CCDs are windowed around the observed targets (5 stars per CCD) and the 5 stellar and background windows are read once every second. The process is driven by the instrument clock;
- faint stars CCDs are read once every 32 s. The process is driven by the clock of the service module.

Fortunaltely, both processes are synchronized by the service module clock, so that the perturbation patterns on both types of CCD are stable in time, each perturbation corresponding to a specific command (one of the 15 low level commands such as pixel digitisation, line transfer to the register... ). Because of the stability of the process, the amplitude of the pattern associated to one specific low-level
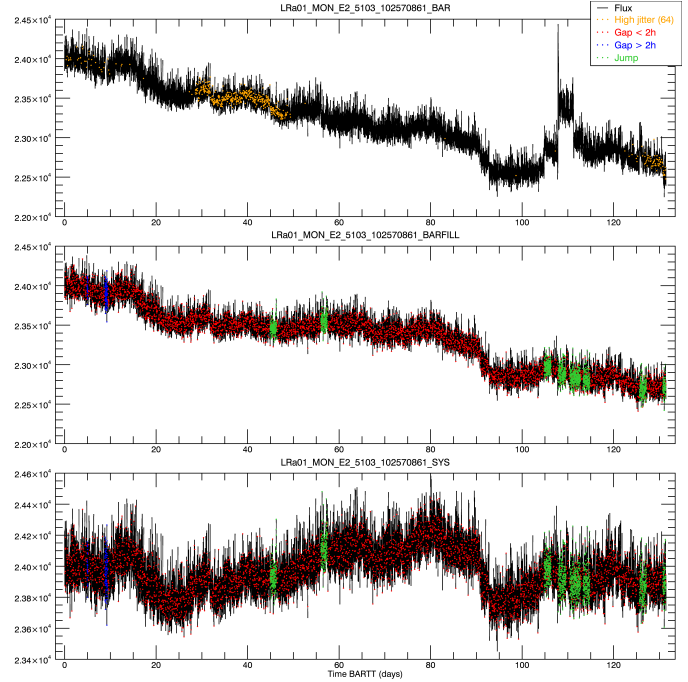


Fig. II.2.4. Example of data of the faint stars channel in its 3 extends BAR, BARFILL and SYS.

command can be calibrated and a final global crosstalk pattern can be computed for each CCD. A complete description of crosstalk components and images of associated patterns can be found in Auvergne et al. (2009).

The elimination of the aliasing is thus a simple process where the crosstalk pattern effet is computed for each photometric mask and its global effect is subtracted from the corresponding light curve.

In very few cases, when the star is positioned in the lower rows of the CDD, read at first, the calibrated correction introduces a periodical artefact at a period of 32 s. This is quite penalizing, thus, in this case, an ad-hoc correction is applied.

### 4.2.2. Elimination of offset and background contribution

This process concerns only the faint stars channel data. For the bright stars channel data, the process is performed aboard (see previous section).

The background estimation all over the CCD is given by 196 windows ($10 \times 10$ pixels), defined during the observation setup phase and distributed all over the CCDs in dark starless zones; 147 windows are sampled at 512 s and 49 others at 32s, per CCD as shown in Fig. II.2.6. The background correction is done in 3 steps:

- removal of the background median value, the method is extensively described in Drummond (2007), and Drummond et al. (2008);
- correction of the Y-dependency, due to the ageing of the detectors and the increase of the dark current with time;
- correction of the differences between the 32-s and 512-s median background estimations, due to a bias in the positioning of the background windows at different sampling rates.
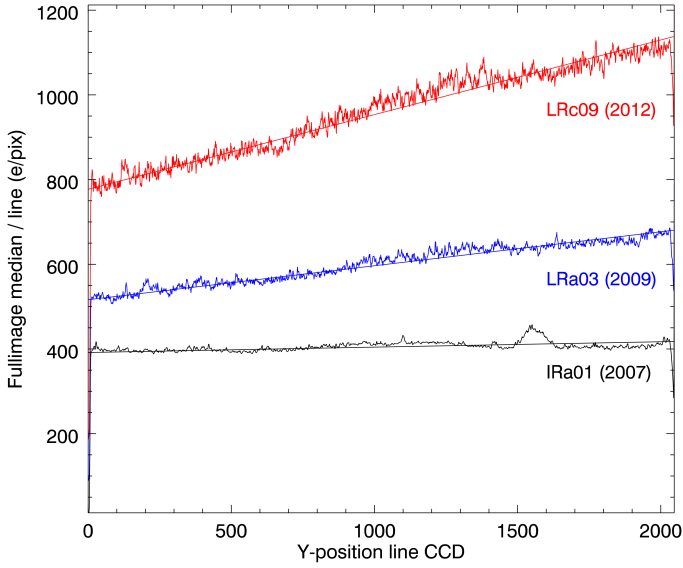
45

**Fig. II.2.5.** Evolution of the background due to the dark current as a function of time.



196 windows (49 for 32s and 147 for 512s)

**Fig. II.2.6.** Position of the background windows at 512-s sampling rate (blue) and 32-s sampling rate (red).

With the ageing of the telescope, phenomena that were minimized at the beginning of the mission are now very influential, like the value of the dark-current. The median method used in a first time (step 1 of the process described above), is not adapted any longer to properly evaluate the level of the background. Assuming that the faint stars CCD readout process duration is 23 s, the increase of the dark current causes a dependency between the background level and the position on the CCD, especially with the Y axis, along the readout direction. This dependency is modelled at the beginning of each run by a first order model, and used to adjust the background correction as following.

$$\text{Flux}_{\text{corrected}}\,(t, y_{\text{CCD}})) = Flux(t) - BG_{\text{median}} - (a*y_{\text{CCD}} + b) \tag{1}$$

where $y_{\text{CCD}}$ is the y position of the star on the CCD, and $\{a, b\}$ the parameters of the dark current model. The parameters $\{a, b\}$ are a function of the run, and are evaluated thanks to the field image acquired at the beginning of each run. Figure II.2.5 shows clearly that the dark current increases with time.

The last part of the background correction is done in the N1 to N2 pipeline. The mapping of the 196 background windows on the CDD is biased: the distribution between the 512-s windows and the 32-s ones is not namely homogenous along the Y axis (i.e.: the barycentre of the windows distribution is not the same, see Fig. II.2.6). Since the Y position has an impact on the background level, the median level estimated thanks to the windows at the different sampling rate is not the same, leading to a bias between 32-s and 512-s sampled data. This bias leads to a jump in the photometry when the sampling rate is changed during the observation sequence. By resampling the 32-s windows at 512 s, a global median value is computed on the 196 windows, and then used to compute the mean shift to apply to each part.

$$Shift_{32} = <BGMed_{32}(t) - BGMed_{\text{global}}(t)> \tag{2}$$
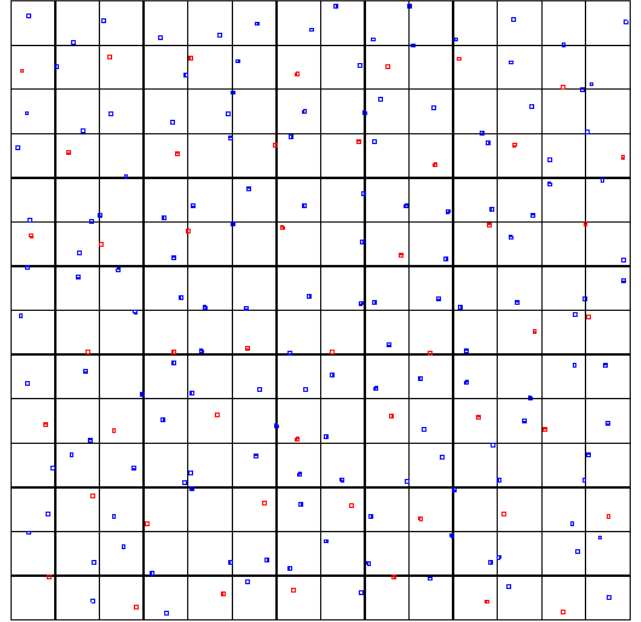
$$Shift_{512} = <BGMed_{512}(t) - BGMed_{\text{global}}(t)> \tag{3}$$

The corrected flux reaches the mean level it would have had if the 196 background windows were used to compute the median background level.

## 4.2.3. Correction of the exposure time

This process concerns both bright stars and faint stars channels data.

As already mentioned above, the CoRoT satellite includes several aboard clocks: among them, one:

– at the service module level (platform clock), giving a top every 32 seconds and used to drive the faint stars channel and synchronize the bright stars channel processes;
– at the instrument level (instrument clock), giving a top at each second, driving the bright stars channel acquisition processes.

Both clocks are driven by oscillating quartzs, but do not have the same accuracy. The platform clock has a stability of about 2 microseconds per second. This stability is a function of the satellite temperature while the instrument clock is much more accurate. Because the 32-s integration cycle for both channels is driven by the platform clock, the faint stars channel exposure times exhibit fluctuations with time and temperature. In addition, because the 1-second integration process in the bright stars channel is driven by the instrument clock and the overall 32-seconds cycle driven by the platform clock, elementary integration time (and thus flux) corrections have to be done. These corrections have been described extensively by Samadi et al. (2007).

On the bright stars channel, the 32 s are in fact 31 s given by the instrument clock plus a duration corresponding to the time between the 31st second top given by the instrument clock and the platform top that synchronizes

the two channels every 32 s, at the accuracy of the platform clock. The correction is done using the aboard GPS time considered as the primary reference. The value of each counter and corresponding GPS time is recorded at each platform top so that it is possible to know the exact duration of the last "second" of bright star channel integration process. The correction applied is thus a simple proportional correction to know the exact duraction of the last second of integration.

On the faint star channel, the GPS time is used to determine the exact duration of the integration and the exact date of the exposure.

At the end of the process of correction at L1/N1 level, the flux is given:

− per second (TT) for the bright stars channel data;
− per 32 seconds (TT) for the faint stars channel data.

### 4.2.4. Absolute datation

This process concerns both bright stars and faint stars channels data.

The way data are absolutely dated changed from what it was at the beginning of the mission. Now (final version of the data), absolute datation is done as following:

− on the bright stars channel: the absolute date is the barycentre of the integration phases;
− in the faint stars channel: the absolute date is the middle of the integration process.

Datation is now done in terrestrial time (TT) which is much more accurate than Universal Time (UT). The conversion from UT to TT has been done using the relation:

$$TT = DATEJD + \frac{32,184s + \text{offset}}{86400} \qquad (4)$$

with offset = 35, 34 or 33.

### 4.2.5. Data resampling

This process concerns only the bright stars channel data.

The data of the bright stars channel are a stack of 32 elementary integrations acquired at a 1-second sampling rate. During the resampling, only the valid acquisitions are taken into account if they are sufficiently numerous. If there are too many invalid exposures, all the 32 acquisitions are combined and the status of the resulting point is set accordingly. The minimum number of valid exposures taken into account varies between the different extends of the data. For the RAW extend, only 7 valid exposures upon 32, are necessary to consider the point as valid. For the 2 other extends, BAR and BARREG, at least 16 valid exposures upon 32 are required to consider the point as valid. If less than 16 (or 7 for the RAW) exposures are valid, then the 32 exposures are cumulated and the points are flagged as invalid.

### 4.2.6. Correction of mask switch jumps

This process concerns only the bright stars channel data.

In observation mode on the bright stars channel, two successive photometric masks are applied to each star. The first one is a rough mask used for the calibration at the beginning of the run. The second one is the specific mask computed and adapted to accurate stellar photometry. However the modification of the masks leads to a discontinuity in the light curve, which needs to be corrected.

By superposing the photometric masks with a time serie of imagettes recorded at the beginning of the run, the mean flux produced by each mask can be measured. The ratio of the 2 fluxes gives the amplitude of the discontinuity, and can be used as a multiplicative factor to correct the light curve of the first sequence. With this method, a table is created, containing a multiplicative factor for each star of the bright stars channel.

In few cases (it happened 2 or 3 times), this method does not give the expected results. In these cases, an empiric method is used to correct the mask discontinuity. The amplitude of the discontinuity is directly measured on the light curve and the first sequence is adjusted to the level of the second sequence. To know which correction was used in the data, a keyword in the primary header of the file has been created: "MASK_COR". If it is equal to 1, the first method using the mask is used, and if it is equal to 2, then the empirical one is used.

### 4.2.7. Correction of jitter effects

The jitter effects are a slight variation of the line of sight of the instrument linked to the intrinsic performance of the plateform pointing system (AOCS). In the case of CoRoT, the AOCS was driven by the instrument itself, in order to increase the stability performance. Nevertheless, pointing residuals reaching 0.3 arcsec (about 1/10th of pixel) at $1\sigma$ per axis could be observed in the raw data. The pointing residuals were frequencially located at even harmonics of the orbital period, and in a large spectral band around $5 \times 10^4$ $\mu$Hz (Auvergne et al. 2009).

Specific methods of correction based on PSF fitting have been developed for the bright stars (Drummond 2007; Drummond et al. 2008) and the faint stars (Fialho et al. 2007) channels. These methods require 5 steps:

− building of reference PSFs on both channels, with a spatial resolution higher than the resolution of the CoRoT images;
− estimation of the global fluctuations of the line of sight, using the housekeeping and the information given by the AOCS and the instrument;
− estimation of the position fluctuations of each star on the field assuming the line of sight variations, using the observation geometrical model;
− evaluation of photometric consequences of the jitter on a well spatially sampled grid, thanks to the reference PSFs, and construction of a jitter surface;
− interpolation of the grid at any position fluctuation given by the AOCS for any star in the field.

The difference between bright and faint stars channel is only in the spatial resolution of the PSF used for the jitter surface construction. The consequences of correction on the Fourier spectra of photometric light curves is shown in Auvergne et al. (2009). Because of spatial variations of the PSFs over the field of view, the process should be optimized for a specific region of each CCD.
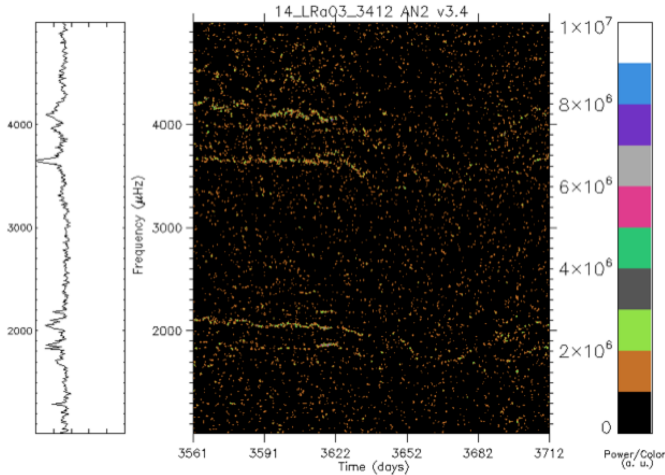
**Fig. II.2.7.** Evolution of the spectrum of HD 43317 with time for which jitter residuals are well visible and identified (power between 2000 and 4000 $\mu$Hz; see the text for complete description) in LRa03.



**Fig. II.2.8.** Spectrum of the photocentre of HD 43317.

The process has been implemented on both channels and runs routinely. The only case where the process is not efficient appears on the bright stars channel, when the photometric mask is too tight compared to the size of the stellar image. In that case, photometric information is lost due to mask side effects.

## 4.2.8. Pointing anomalies and effects on light curves

The effects of the residual jitter of the satellite is corrected on the light curves thanks to an accurate and permanent determination of the line of sight (see previous section). However, a careful examination of some bright stars channel light curves shows clearly some uncorrected flux fluctuations strongly correlated to the fluctuations of the photocentre of the stars that are used for the line of sight determination and the satellite orbital attitude control system. Practically, we observe (cf. Fig. II.2.7):

- the presence of frequency bands around 2 and 4 mHz (2 mHz = 500 s) in the Fourier transform of the stellar flux;
- artifacts are identical on several stars of a same run;
- but all the stars of a same run are not affected;
- the frequency of the artifacts is not constant and the firts band in the spectrum is between 1.5 and 2.5 mHz;
- there is power in the harmonics but it decreases quickly.

After investigations, we came to the conclusion that it was jitter residuals because it was correlated with the barycentre signal (Fig. II.2.8) and that the jitter algorithm itself was not involved in this effect because a large majority of the other observation runs does not exhibit such an effect. After deeper investigations we also came to the conclusion that, in some cases, the size of the window used to estimate the photocentre of the guiding stars (35 × 35 pixels windows) was too small compared to the extend of the PSF in some part of the bright stars CCDs. As a consequence, the barycentre and the variations of the line of sight determined using the guiding stars are slightly biased and
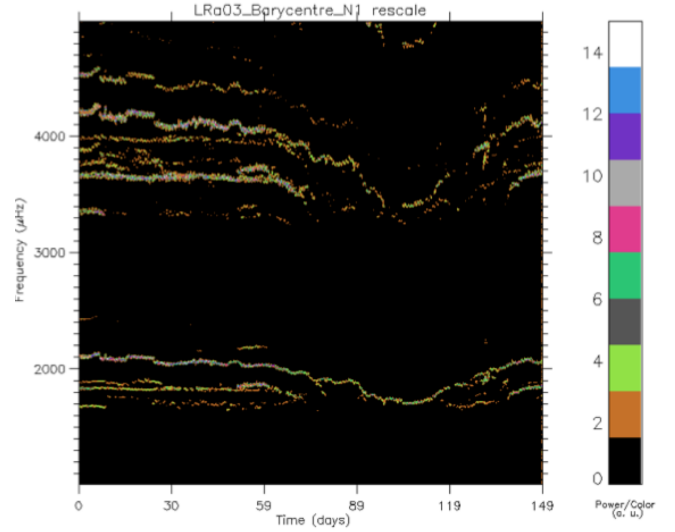
noised. Unfortunately, part of the line of sight information is lost and cannot be extrapolated. We decided to keep the light curves associated with these runs uncorrected and to deliver the position of the photocentre in a special file (AN2_POINTING).

## 4.2.9. Correction of relativistic aberration

This process concerns both bright stars and faint stars channels data.

The relativistic aberration is an apparent change in the direction of an object observed by a moving observatory, and in our case, by CoRoT on its orbit around the Earth, which orbits around the sun. Theoretically speaking let $\theta_s$ be the angle between the direction of the target and the moving direction of CoRoT . Then the apparent deflection angle $\theta_0$ is given by the relation:

$$\cos(\theta_0) = \frac{\cos(\theta_s) - \frac{v}{c}}{1 - \frac{v}{c}\cos\theta_s} \qquad (5)$$

where $v$ is the speed of CoRoT with respect to the observed source and $c$ is the speed of light in vacuum. On the image, the global effect of the relativistic aberration is a field distorsion. The orbital motion of CoRoT around the Earth at about 7.4 km s$^{-1}$ is negligible compared to the motion of the Earth around the Sun at about 30.8 km s$^{-1}$. In the formula above, the speed vector of the observatory is thus the speed vector of the Earth around the Sun. The relativistic aberration is thus maximum more or less in the middle of long runs, when the Earth motion is perpendicular to the line of sight of the instrument. In that case, $\theta_s = \frac{\pi}{2}$ and $\theta_0 = \frac{v}{c} \sim 10^{-4}$ rad $\sim 20$ arcsec. This corresponds to about 8 pixels. Fortunately, the first order of the effect is a global translation of the field and is corrected by the attitude control system of the satellite, thanks to the CoRoT instrument that works as a fringe tracker. Only the differential residuals from one side of the field to the other one have to be corrected. It gives a dilatation (resp contraction) of the field of view, equivalent to a focal length slow variation, i.e.: a radial shift of the stars position with respect to the

line of sight. Assuming a field of view of 1.7 square degrees, this leads to a differential shift (top-left to bottom-right corners) of less than $10^{-5}$ rad, i.e. 2 arcsec, which is less than a pixel (pixel size: 2.3 arcsec).

The correction of the relativistic aberration was not considered in the first versions of the pipeline. At present, the correction of the effect is included in the jitter correction algorithm (see previous subsection), and particularly in the estimation of the line of sight variations. The focus of the telescope, slightly variable with time, is considered in the computation of the observation geometrical model. Thus, the stellar image shift due to the relativistic aberration, which is a function of the position of the stellar image on the field can be estimated, and is added to the global field jitter due to AOCS fluctuations. Relativistic aberration effects are slow compared to jitters variations. The associated variation of the size of the images, both in the faint and in the bright stars channels, and due to the slight defocus, are negligible and neglegted.

## 4.2.10. Correction of the detection chain loss of efficiency

This process concerns both bright stars and faint stars channels data.

During the mission duration, several effects led to a degradation of the detection chain efficiency:

- loss of transmission due to optical ageing;
- increase of the dark current;
- electronic noise increase;
- loss of electron transfer efficiency within the CCD;
- ...

As a consequence, the efficiency of the photons collection decreased and caused a long linear trend, especially visible over a complete long run. The average efficiency loss has been estimated by Auvergne et al. (2009) to about 10%. The value of the slope of the linear trend is different for each star, It is directly linked to the intensity of the star. At the end of the mission, all the available data were used to build a model which links the intensity of a star to the slope of the linear decrease of the efficiency.

$$\text{Loss}_{\text{slope}}(\bar{f}) = \alpha_2 x \bar{f}^2 + \alpha_1 * \bar{f} \qquad (6)$$

Where $\bar{f}$ is the mean flux of the star, and $(\alpha_2, \alpha_1)$ are the two parameters of the model.

In the pipeline, the efficiency loss correction works in two phases:

- the mean flux of the star is used with the model to compute the estimated linear slope of the efficiency loss;
- the light curve is corrected by adding the estimated loss along the time.

$$\text{Flux}_{\text{corr}}(t, \bar{f}) = \text{flux}(t) - \text{Loss}_{\text{slope}}(\bar{f}) * t \qquad (7)$$

The slope used to correct the curve is given in the primary header of the file, in the field "COR_SLOP".

Figure II.2.9 shows an example of transmission loss correction on a bright star from LRa04.
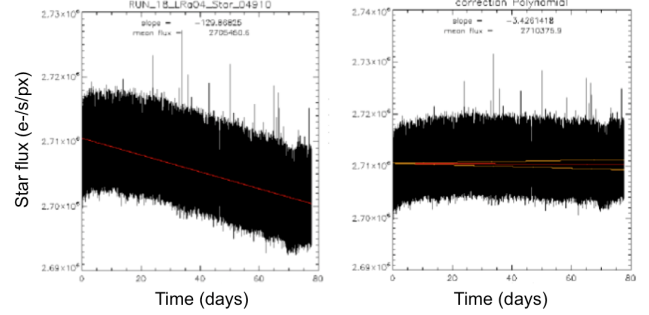


**Fig. II.2.9.** A light curve before (*left*) and after (*right*) correction of the efficiency loss.

## 4.2.11. Identification of orbital events and detection of outliers

CoRoT was placed on a near circular polar orbit, with an apogee at 911 km, a perigee at 888 km, and a period of 6184 seconds. Because it is an inertial orbit, the spacecraft crosses regularly specific zones over the Earth:

- the South Atlantic Anomaly (SAA): it is a wide region of South America and South Atlantic Ocean where the magnetic field is a bit stronger, leading to a decrease of the altitude of the inner part of the Van Halen belts. As a consequence, at a given altitude, the proton impact rate is higher in this region compared to the other ones. In the case of CoRoT, the SAA crossing leads to a dramatic amount of proton impacts on the detector so that the stellar photometric information is partially lost or at least strongly affected. In order to better identify these events, a flag has been created to mark the data acquired during the SAA crossing. CoRoT also helped to draw a more accurate map of this region (see Auvergne et al. 2009);
- the Earth shadow: corresponding more or less to the fly over night regions. The first effect on the photometric light curves is a strong modification of the scattered light, but also, short-term thermal effects at ingress and regress in the shadow of the Earth, modificating for a few seconds the stability of the line of sight. Once again, ingress and regress are flagged because line of sight sudden variations create a strong jitter effect on the photometric light curves (see Pinheiro da Silva et al. 2008).

## 4.2.12. Creation of the white photometry

This process concerns only the chromatic light curves of the faint stars channel. The white photometry is the result of a simple co-addition of the blue, green and red photometry previously computed aboard. This white light curve has at least two advantages:

- for detection: the signal to noise ratios of the white light curves are much higher than those of colored light curves;
- the jitter effects, and particularly due to colors borders are strongly reduced.

## 4.2.13. Correction of thermal jumps

This process concerns both bright stars and faint stars channels data.

The temperature of the CoRoT instrument was not actively servo-controlled. The focal plane of the instrument was connected to a radiator, decreasing its temperature to a value between $-50$ and $-40$ Celsius. This temperature depends on the orientation of the radiator with respect to the Sun. As a consequence, the radiator temperature varies along the Earth orbit, and particularly during a long run. In order to avoid the focal plane temperature varying during an observation run, a stable reference temperature is set by heating the plate where the detectors are mounted and controlling the final temperature. The temperature of the radiator along the orbit and the corresponding temperature of the focal plane are estimated thanks to an abacus assuming the field of view to be observed, and the corresponding orientation of the satellite with respect to the Sun, all along the duration of the run. The reference temperature, applied to the temperature controller is chosen to be slightly higher than the highest temperature given by the abacus, so that the order of the servo control is not changed during the observation and the final temperature of the detector is stable during the run.

The temperature of the CCDs of the satellite was effectively controlled, therefore the time variations of the measured temperature are small enough to avoid requesting a correction along the run. However during the mission, it happened that the order of the temperature control system had to be changed during a run because the controller could not work properly, causing steps in the temperature curves of the detectors. As an example, the temperature order had been changed six times during the LRa01 observation run. Those orders need to be corrected. The purpose of this correction is to match the jumps seen on the temperature curve (HK) and the effects seen on the light curve. Since the exact function, converting a temperature jump into the corresponding jump in the electronic signal, has too many parameters to be correctly computed, the correction uses the optimisation of a numerical model of the jump. The following equations are established on a short window around the temperature jump.

Light curve model $(t) = a(t - b) - c * T2(t) + d$ (8)

with $t$ in a window of $n$ points, and where $T2$ is a temperature curve, normalised as a function of the mean level before the jump:

$$T2(t) = \frac{HK(T)}{\frac{1}{k}\sum_{i=1}^{k} HK(t_i)} - 1 \qquad (9)$$

with $k$ the index of the jump, in the considered window ($k < n$).

Then the four parameters of the light curve model $a$, $b, c, d$ are optimised to fit the light curve, on a window around the date of the temperature order. During the optimisation process, several lengths of windows are tested, from 8 hours to 2 days, in order to avoid fitting the stellar activity of the star. Once the fit is complete, the best model is chosen to generate a correction factor. Actually, only the parameter $c$ of the model is used; it represents the amplitude of the jump. The correction factor is applied from the first point after the change of temperature to the next order, or to the end of the data set, if no other temperature jump is encountered.

$$\text{Corr}_{\text{factor}}(t) = 1 + \left(\frac{HK(t)}{\frac{1}{k}\sum_{i=1}^{k} HK(t_i)} - 1\right) * c \qquad (10)$$

In the header of the N2 data, specific fields give information about the temperature correction results:

- NB_CONSI: gives the number of temperature orders in the run;
- DATE_TPE: gives the date at which the temperature order changes. This field is duplicated 6 times, because it is the maximum number of temperature jumps recorded in the data, but most of the time there is only one;
- COR_TPE: also duplicated 6 times, it gives the status of the temperature correction for the considered jump. This flag can take 3 values:
  - 0: means the correction worked without issues;
  - 1: means the correction worked but with a warning. It may be either a failure during the optimisation process on one of the windows, even though at least one of the other windows worked, or a mismatch between the date of the temperature change and the data, especially if the temperature change happened before the beginning of the faint star sequence;
  - $-1$: means the correction did not work.

## 4.2.14. Correction of non-thermal jumps

This process concerns only the faint stars channel data. It is necessary for aboard computed light curves, not for the light curves obtained on the ground from aboard imagettes; therefore, it is not applied on the later.

The purpose of the jumps correction is to detect and correct the jumps not compensated by the previous algorithm. The CoRoT data, and especially the faint star light curves, show two kinds of jumps:

- the result of impacts of protons on the CCD; they are leading to a sudden increase of the flux in the light curve, followed by a rather exponential decrease. The decrease time is usually unknown and can strongly vary from one jump to another one;
- an abnormal behaviour of the aboard software , leading to an alternation between two states, and between two different reference levels.
  In a first time, the algorithm detects the jumps by comparing the mean values before and after each point of the curve. The size of the moving windows used to compute the mean value depends on the sampling rate of the curve (see Table II.2.1). If the difference with the mean values is greater than $7\sigma$ then a jump is detected.

For each detected jump, a negative exponential model, is fitted on a short window after the jump. The model is used to adjust the light curve as a function of the median level before the jump. The lengths of the windows used to compute the median level or to fit the model are detailed in Table II.2.1.

Jump model$(t) = \mathrm{e}^{a(t-c)} + b$ (with $a \leq 0$) (11)

$$\text{Flux}_{\text{corr}}(t) = \frac{\text{Flux}(t)}{\text{Jump model}(t)} * \text{Median}_{\text{before jump}} \qquad (12)$$

An example of proton impact induced jump correction is given on Fig. II.2.10.

In some cases, the jump detection can identify structures which are not discontinuities, like eclipses of binary stars or the transits of exoplanets. To avoid those events being corrected by the jump detection algorithm, a special
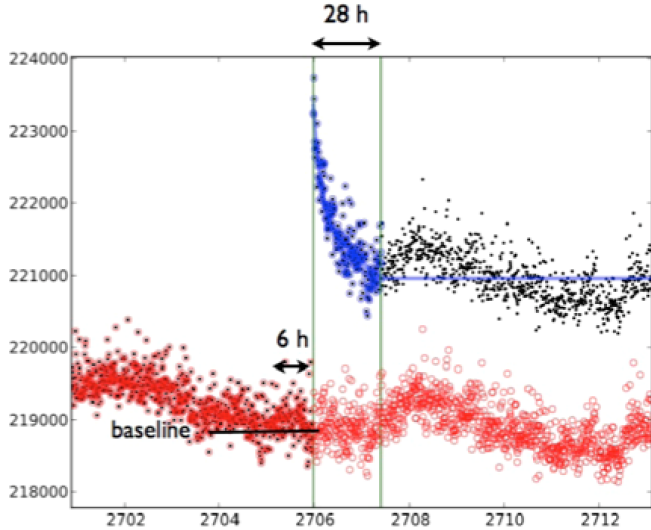
**Fig. II.2.10.** Example of a proton impact-induced.

**Table II.2.1.** Lengths of the windows used to compute the median level or to fit the model as a function of the sampling rate.

|  | 32 s | 512 s |
|---|---|---|
| Nb of points used to compute the mean before and after a jump | 150 points 1.3 hour | 20 points 2.8 hours |
| Nb of points used to fit the model | 200 points 1.77 hour | 200 points 28.4 hours |
| Nb of points used to compute the median level before the jump | 40 points 21.3 min | 40 points 5.7 hours |

treatment is applied to all the binaries and the transit candidates detected by the exoplanet team. Their parameters, ephemerides, period and transit duration, and their associated errors (Deleuil et al., 2015) are used as an input of the function; if a jump is detected in a region of a transit-like event, the jump is not corrected. The safety region around the transit events is determined taking into account the transit duration and its associated uncertainty. In practice, the safety region duration is about a few hours, which corresponds to the longest time window used in the correction algorithm. In some specific cases, the algorithm failed and detected non existing (real) jumps. The correction process introduced extra perturbations. This phenomenum has been identified in about one third of the objects of a specific sample of RR lyrae. This point is considered in Appendix A at the end of the paper.

## 4.2.15. Filling the gaps

This process concerns both bright stars and faint stars channels data.

The Inpainting method developed by Pires et al. (2015) is used on the CoRoT data to fill the gaps in the light curves. The gaps can either result from some missing values during the inflight acquisition, or from the removal of invalid data during the ground-based processes. All the gaps are filled, even those of several days. But to ensure the
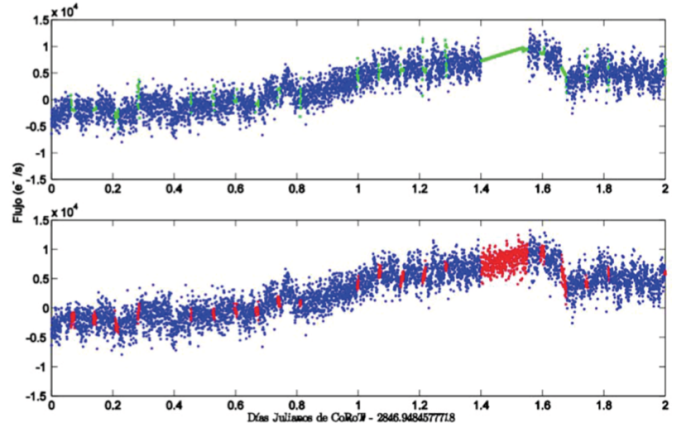


**Fig. II.2.11.** Example of gap filling, using the Inpainting method (Pires et al. 2015).

proper behaviour of the algorithm, process is performed in two steps:

- the short timescale gaps (less than two hours) are first corrected;
- the remaining gaps are filled without limit of duration.

The distinction between the two steps is possible thanks to the status of the new points, each step having a specific status. On the bright stars channel, the BARREG extend is perfectly filled and ready to use in spectral analysis, with a proper 32-s sampling. On the faint star channel, the BARFIL extend is also filled, but it is possible to find some specific gaps in the data. Indeed for the light curves with an oversampling, the gaps between the two parts at different sampling rates, 32s and 512s, cannot be filled. In addition, some gaps do not last an exact number of 32 exposures, and thus cannot be corrected. Those gaps can especially be found at the beginning of the mission. An example of gap filled light curve is given on Fig. II.2.11.

## 4.2.16. Correction of systematics

This process concerns only the faint stars channel data.

Even if the global phylosophy for the data processing consists of correcting only effects that can be clearly identified, modeled or physically quantified, one must admit, that, at the end of the global data processing, some light curve alterations are still visible, and remain uncorrected. Even if one cannot always exactly identify the origin of these residuals – some of them are due to cross correlations between several effects that are corrected individually – we can notice that they are:

- present on each light curve (systematics);
- with an amplitude that cannot always be quantified using a simple parametric description.

At the beginning of the project, it was proposed to apply a "last chance algorithm", a systematic removal algorithm such as the SYSREM algorithm (Tamuz et al. 2005). The method has been optimized and led to a process described by Guterman et al. (2015). The algorithm works on gap filled and jump corrected data, and has been implemented on the last version of the pipeline. A new extend named SYS has been created in that purpose (see Fig. II.2.2).

51

## 4.2.17. Second detection of outliers

This process concerns only bright stars channel data.

A second detection of outliers is performed for the stars of the bright channel during the N1 to N2 pipeline. The parameters of the search, as the size of the moving window or the threshold of the sigma value, are manually tuned for each star. The detected outliers are flagged in the BAR extend and they are replaced by "inpainting" points in the BARREG extend (see the gap filling section).

## 4.2.18. Creation of final flags and file header

This process concerns both bright stars and faint stars channels data.

At the end of the process, each element of the photometric data is flagged using an unsigned long integer (32 bits) named STATUS and giving information of the data, including the orbital events. The final values of STATUS are given in Chaintreuil et al. in this book.

All the information describing the data set and several corrections made within the pipeline are gathered in the file header.

# 5. Other facts about the CoRoT instrument ageing and end of life

## 5.1. Ageing of the instrument and efficiency loss

The space environment leads to ageing of the instrument and at the end of the detection chain, a global efficiency loss. The efficiency loss is difficult to estimate, because:

- no absolute photometry is performed by CoRoT . This means that the instrument has not been absolutely calibrated using photometric standards and is not calibrated in flight;
- a very few number of long-term stable stars have been re-observed and can be used to estimate the long-term efficiency loss;
- when stars have been re-observed, the orientation of the line of sight and the image position on the CCDs were not identical. Because of the variation of the PSF over the focal plane, and the specificity of the masks used for the photometry, it is difficult to compare the absolute photometry from one observation to the other.

However, a first order estimation of the detection chain efficiency loss made on a limited number of stars with comparable observation conditions led to an efficiency loss of the detection chain of about 10% during the whole mission duration (Auvergne et al. 2009). At this level, it is also difficult to assign exactly what is due to:

- a loss of optical transmission;
- a decrease of the quantum efficiency;
- a loss of charge transfer efficency.

## 5.2. Degradation of the images optical quality

The CoRoT telescope structure is made of carbon fibers. One of the key questions raised by the use of this material
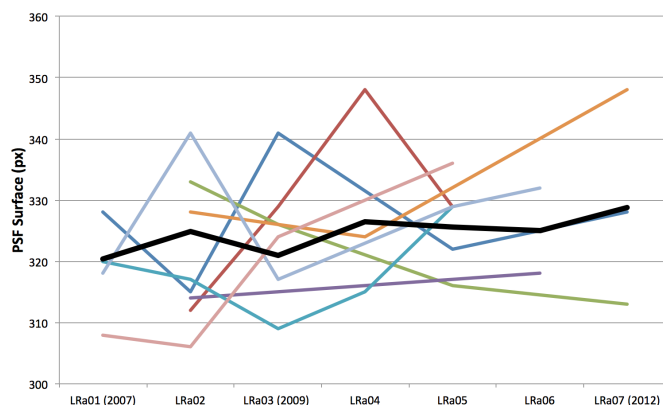


**Fig. II.2.12.** Evolution of the bright star PSF size with time showing no significant change during the mission lifetime (see the text for discussion).

was its behaviour in space and particularly, the evolution of the telescope bars length with time. Carbon fibers are wrapped with an epoxy resin which contains water. In space, and particularly just after the launch, a desorption of the carbon fiber bars was expected, leading to a variation of the telescope effective focusing. The initial variation was anticipated and the telescope was focused on the ground assuming an expected focus variation during the first three months of the mission. After three months in space, the telescope was effectively perfectly focused as expected. The question was to investigate a potential defocus after the three first months of mission. The variation of the position of the stars in the field cannot be used as an indicator of defocus (breathing of the field), because it is corrected simultaneously with the jitter and the relativistic aberration (see Sect. 4.2.9). We proposed thus to monitor the variations of the size of the bright star channel images. These images are slightly defocused to avoid the CCDs saturation. In that way, the size of these images can be an indicator of the focus variations. Because of the optical formula of CoRoT (wide field telescope and dioptric objective), the size and the shape of the PSFs are not constant over the field. Thus, we split the A2 CCD in 9 square regions where we estimated the size of the PSF along the mission lifetime (6 yr). Results are given on Fig. II.2.12. Even if the dispersion is large because the number of images is small, the average size of the PSF (black curve) does not exhibit a significant increase or decrease, tending to show that the focus remained globally unchanged after the initial phase of bars desorption. In the same time, no significant changes have been observed in the images of stars on the faint stars channel, and an update of the photometric mask has not been necessary.

## 5.3. Detector ageing

The detector and front-end electronics ageing leads to a global degradation of the performance of the detection chain:

- loss of charge transfer efficiency, as demonstrated on the HST (Goudfrooij et al. 2006);
- decrease of the quantum efficiency;
- increase of the dark current.

The two first effects cannot be investigated using only the data that are available. The last effect has been clearly observed and led to a change in the way the background in the bright and faint stars channels is corrected (see corresponding section in this paper).

## 5.4. Origins of the instrument failure

The instrument failure happened in two steps:

- the 8th of March 2009: loss of the communication with one of the two detection chains leading to the loss of the A1 (bright stars) and E1 (faint stars) CCDs;
- the 2nd of November 2012: loss of the communication with the second detection chain, leading to the end of the scientific mission after further investigation in June 2013.

Even now and after several months of dedicated investigations and tests on the instrument, including a complete switch-off and reboot of all the sub-systems, the instrument remained silent, and the origins of these two failures are not clearly understood. They are certainly due to the ageing of several electronic devices in charge of the power control of the detection chain, preventing from normal reboot of the DPU. Let us remind that the mission was designed for a mission duration of 3 yr, and the final lifetime of the second detection chain reached almost 6 yr.

## 6. Conclusions

The CoRoT legacy data are now delivered to the community. The corrections applied to the raw data correspond to the state-of-the-art knowledge of the instrument behaviour and evolution. Several extends, with different sets of corrections or post-processing are available to maximize their scientific uses. Anyway, our final knowledge is certainly not complete and other effects may be identified in the data in the future, and further investigation may be needed, particularly in the context of CHEOPS and PLATO, for which the same type of detectors will be used. A priori, no new version of the data including new corrections will be created.

## References

Alonso, R., Auvergne, M., Baglin, A., et al. 2008, A&A, 482, L21

Auvergne, M., Bodin, P., Boisnard, L., et al. 2009, A&A, 506, 411

Boisnard, L., & Auvergne, M. 2006, ESA-SP-1306, 19

Bonomo, A. S., Chabaud, P.-Y., Deleuil, M., et al. 2012, A&A, 547, A110

Chabaud, P.-Y., Agneray, F., Meunier, J.-C., et al. 2011, ASP Conf. Proc., 442, 339

Chaintreuil, S., Deru, F., Baudin, F., et al. 2015, this book

Deleuil, M., Moutou, C., Cabrera, J., et al. 2015, this book

Drummond, R., Lapeyrere, V., Auvergne, M., et al. 2008, A&A, 487, 1209

Drummond, R. 2007, Calibration of the CoRoT Space Mission: correction of the photometry for pointing jitter and background changes, Katholieke Universiteit Leuven

Fialho De Oliveira, F., Lapeyrere, V., Auvergne, M., et al. 2007, PASP, 119, 337

Goudfrooij, P., Bohlin, R. C., & Maíz-Appellaníz, J. 2006, PASP, 118, 1455

Guterman, P., Mazeh, T., & Faigler, S. 2015, this book

Lapeyrere, V. 2006, Etalonnage des détecteurs de CoRoT: adaptation aux besoin spécifiques de la mission, PhD Thesis, Université Pierre et Marie Curie (in French)

Pinheiro da Silva, L., Rolland, G., Lapereyre, V., et al. 2008, MNRAS, 384, 1337

Pires, S., Mathur, S., Garcia, R., et al. 2015, A&A, 574, id.A18

Plasson, P. 2009, CoRoT SW user requirement, COROT/DESPA.00.019

Quentin, C. G., Barge, P., Cautain, R., et al. 2006, ESA-SP-1306, 409

Samadi, R., Fialho, F., Costa, J. E. S., et al. 2007, eprint [arXiv:astro-ph/0703354]

Surace, C., Alonso, R., Barge, P., et al. 2008, Proc SPIE, 7019, id. 70193B

Tamuz, O., Mazeh, T., & Zucker, S. 2005, MNRAS, 356, 1466

## Appendix A: Jump detection failure

The jump detection algorithm overcomes the star variability in the light curve by computing a differential vector normalised by the local standard deviation. Thus the impact of the local variability of the star is minimised. For each point of the light curve, the differential vector is computed as following:

$$\Delta(i) = \frac{\overline{flux}_{after\,i} - \overline{flux}_{before\,i}}{\sqrt{\sigma^2_{before\,i} + \sigma^2_{after\,i}}}$$

Then to detect a jump, each point is compared to the global standard deviation of the full length of the differential vector $\sigma_0$.. A jump is detected at the "i" index if:

$$\Delta(i) > 7.2\,\sigma_0$$

Where:

$$\sigma_0 = stddev(\Delta)$$

This method allows considering the global disturbance level of the star. For a strongly perturbed star the $\sigma_0$ value will be higher and the detection of a jump will be more rigorous.

However the method presents two limitations. It works in term of elements and does not consider the time vector. It works only on the filtered data and does not take into account the points recorded during the SAA crossing.

These limitations can lead to false detections especially with highly variable stars as the RR Lyrae stars, because of an unfortunate synchronisation of the star patterns and SAA crossing.

Figure A.1 shows a jump detection failure, the SAA crossing is synchronized with the intensity increase of the star, causing an abnormal increase of the differential value. A SAA crossing longer than usual is sufficient to make the differential vector exceed the detection value (red line).
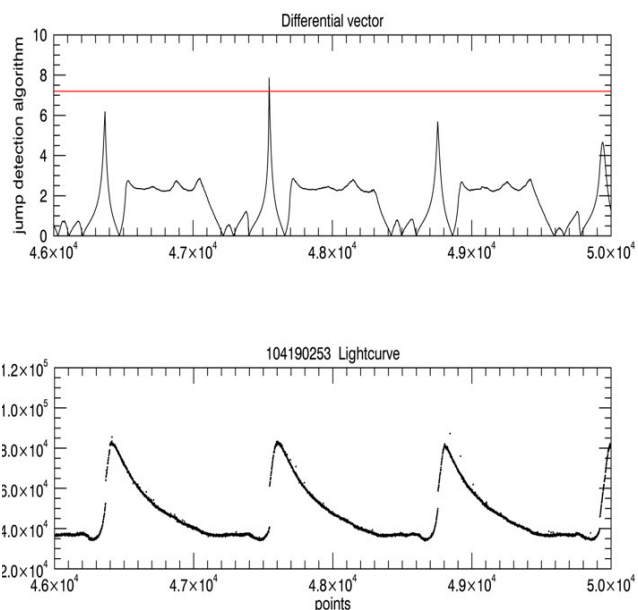
**Fig. A.2.** Jump correction failure showing the variation of the mean level and of the amplitude. *Top*: light curve before correction. *Bottom*: light curve after correction. In abscissa, the time in Julian date.

**Fig. A.1.** Jump detection failure. *Top*: differential vector. *Bottom*: light curve in points. The peaks of the differential vector are synchronized with the gaps caused by filtering the SAA crossing. The red line is the detection limit. In abscissa, points, not time.

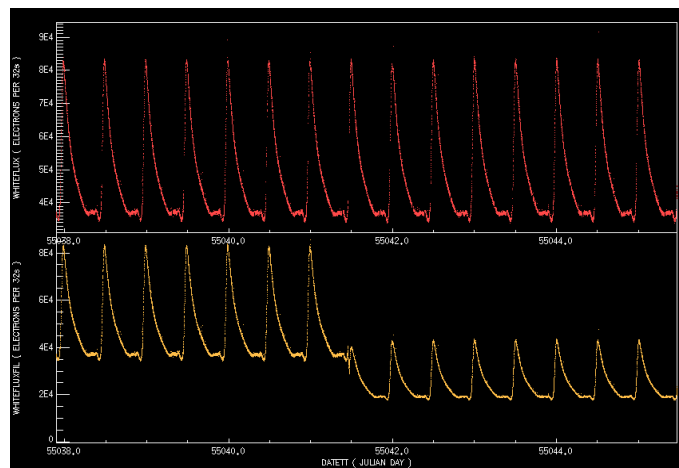The false detection is then corrected by the jump correction causing severe discontinuities in the light curve, see Fig. A.2. It shows also that a reduction of the amplitude can happen.

These correction errors are not frequent but they exist!

When such an anomaly is detected (through a visual inspection), only the BAR extension, available in the N2 final data, has to be used, as this extension does not include the jump correction.