Radboud Repository

Radboud University Nijmegen

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/72096

Please be advised that this information was generated on 2017-12-06 and may be subject to change.

Astronomy
&
Astrophysics

Correcting for background changes in CoRoT exoplanet data[*]

R. Drummond[1], V. Lapeyrere[2], M. Auvergne[2], B. Vandenbussche[1], C. Aerts[1,3], R. Samadi[2], and J. E. S. Costa[4]

[1] Instituut voor Sterrenkunde, Katholieke Universiteit Leuven, Celestijnenlaan 200D, 3001, Belgium
e-mail: rachel@ster.kuleuven.be
[2] Observatoire de Paris, LESIA, CNRS UMR 8109, 92195 Meudon, France
[3] Department of Astrophysics, IMAPP, Radboud University Nijmegen, PO Box 9010, 6500 GL, Nijmegen, The Netherlands
[4] Instituto de Física, UFRGS, 91501-970 Porto Alegre, RS, Brazil

ABSTRACT

Context. The CoRoT satellite is a highly accurate photometer with 2 channels respectively optimised for asteroseismology and exoplanet finding. The design includes an effective straylight rejection system, however residual straylight reaches the detectors.
Aims. We test four different background models in order to apply the best possible background correction to the 12 000 stars observed over the 2 exoplanet CCDs. We identify the best correction method for two types of data reduction – the validation and production data. We also describe a bright pixel correction method and compare background correction quality before and after this correction.
Methods. We used jackknifing – a particular example of bootstrapping – to increase our statistical analysis of a small dataset, comparing the background model with a background datapoint. This enabled us to quantify the background correction quality, which would be impossible when applying the correction to star data.
Results. Our examination of the in-orbit data from two CoRoT runs shows that they give very different results. The commissioning run had far from optimal background window placement. Both runs demonstrate that the closest window correction is very sensitive to the bright pixel problem. Using three windows increases the chance of including a bright pixel impacted window and does not increase performance. Both median and polynomial fit methods give a good correction in most cases, but the median is overall most efficient.
Conclusions. We have shown that a median of all available background windows is the correction method most resistant to bright pixels. It also gives a good background correction for data post-bright-pixel correction. This method has been implemented in the CoRoT pipeline for both validation and production data.

Key words. instrumentation: photometers – space vehicles: instruments – methods: statistical – methods: data analysis – techniques: image processing

1. Introduction

The CoRoT space telescope is a French-led European mission, with significant international participation: ESA, Austria, Belgium, Germany all contribute to the payload, and Spain and Brazil contribute to the ground segment. CoRoT uses the CNES PROTEUS small-satellite bus, and it was placed in a polar orbit by a Soyuz/ST launch in December 2006. The CoRoT mission has two scientific programmes, both requiring long uninterrupted observations with very high photometric accuracy. They work simultaneously on adjacent regions of the sky. For more information on the scientific goals of the mission and its current status, we refer the reader to the CoRoT websites[1].

The asteroseismology channel is sampled every second, covering 10 stars with a magnitude range of 5.5 to 9. The exoplanet channel is sampled either at 32 s or averaged over 512 s, but has some 12 000 targets, of m_v 12−16. To remain within a tight telemetry budget, all stars are assigned photometric masks, and the total light within these masks is downloaded for the exposure time (de Oliveira Fialho & Auvergne 2006; Llebaria & Guterman 2006). On the exoplanet channel, these masks cover between 35 and 120 pixels each (average 60). Brighter stars are sampled in three colours thanks to a dispersal prism placed in front of the CCD (Barge et al. 2006). CoRoT has set a goal of extremely high-precision photometry. We need to be able to detect flux variations at a level of 1 ppm in five days on the asteroseismology channel and 700 ppm over one hour on the exoplanet channel (Baglin et al. 2006). It is therefore very important that the flux does not vary significantly for reasons other than real changes in the stars under observation. There are many sources of variation other than the stars. The major ones are tabulated in Table 1. Some of these can be corrected for, using other downloaded data (such as CCD offset windows). The CoRoT correction pipeline contains many corrections for data received from the satellite. Amongst these are the electronic offset, gain, electromagnetic interference, outliers, background, and jitter corrections. This paper examines the quality of the background correction. A paper outlining the overall performance of CoRoT in flight will be published by Auvergne et al. (in preparation).

2. Straylight and background correction

The major sources of straylight are diffuse galactic background, zodiacal light, and the straylight from the Earth. Other minor contributors are moonlight, the parasite light due to diffusion and the reflection of the stars' light inside the instrument. The

---

[1] http://smsc.cnes.fr/CoRoT/andhttp://corot.oamp.fr

Table 1. The main noise sources and their relative importance for each of the science channels.

Asteroseismology			
Source	Level (e$^-$)	ppm (1 s)	ppm (5 days)
Star	5.49×10^6		
Photon noise	2.34×10^3	426	0.65
Read noise	187	34	5×10^{-2}
Dark current	6	1.1	1.67×10^{-3}
PRNU	191	34.8	5×10^{-2}
Bright pixel	10	1.8	2.7×10^{-3}
Background	233	42.5	2.2×10^{-2}
Exoplanetology			
Source	Level (e$^-$)	ppm (32 s)	ppm (1 h)
Star	4.31×10^4		
Photon noise	207	4820	454
Read noise	86.6	202	19
Dark current	15.5	360	34
PRNU	4.8	111	10
Bright pixel	320	7425	700
Background	173	4020	379

Note 1. PRNU refers to the non-uniform response of CCD pixels. We assume a star of magnitude 6 ($T = 5750$ K) for asteroseismology, covering a mask of 350 px. For the exoplanet channel, we consider a star of magnitude 15.5 ($T = 4500$ K) covering 75 px. The stellar signal and photon noise levels come from Lapeyrere (2006), where values are tabulated for a wider range of star magnitudes. The other noise sources were calculated from summaries of the noise levels described by Boisnard & Auvergne (2004), Lapeyrere (2006) and Drummond (2007).

amount of straylight that manages to reach the focal block depends strongly on the angle between the line of sight of the satellite and the limb of the Earth, as well as the angle to the Sun. CoRoT's orbit also means that it passes regularly through the South Atlantic Anomaly (SAA – an area where the Van Allen radiation belts drop to lower altitudes - resulting in higher intensity radiation). Energetic particles that reach the detectors can impart their energy to the CCD, resulting in a temporary increase in flux seen in one or more pixels. The Earth straylight and the SAA are periodic elements in the background, so they present a larger problem. As the satellite passes over the sunlit side of the Earth, there will clearly be an increase in straylight. The size of these peaks and their duration in time depends on the orientation of the orbit, and this varies throughout the observations (Boisnard & Auvergne 2006; Samadi et al. 2006). The background level is not uniform across the focal plane. This was expected, despite the 1×10^{-12} straylight attenuation of the baffle (Plesseria et al. 2004). The average background level seen in the first CoRoT runs is about 12.5 e$^-$/px/s. This varies around an orbit with an amplitude of about 0.1 e$^-$/px. The noise level included in Table 1 is the shot noise on the average level, as this is the major source of noise from the background light.

The choice of background correction method for the asteroseismology channel was the closest background window. This was a simple choice for this science channel, as only 10 stars are observed during a run, and therefore 10 background windows are placed on the CCDs, each one related to one star. There is far more choice of position, and this window is close enough to be considered representative of the background levels that the star will receive. However, on the exoplanet channel, some 6000 stars are observed in each run, making the CCD mask assignments complicated. The longer exposure time of 32 s also means that far fainter stars are observed and that fainter objects still begin to affect the background measurements. The telemetry budget (1.5 Gbit/day) is not the only limit on how many background windows we can assign. We must be sure that windows are not placed on faint stellar objects, faults in the CCD or too near a star, due to the jitter problem (Drummond et al. 2006).

Four hundred background windows (10×10 px) are assigned positions using the first full-frame images from each run. About 100 are at the 32 s sampling and the remaining 300 at the 512 s sampling. These windows are placed at positions where there are minimal faint stars i.e. areas of minimum convolution with a constant kernel. This process uses a convolution technique on a median taken across several pixels, in order to avoid clusters of "bright pixels" that could be a star. The convolution is carried out on smaller segments of the CCDs to ensure a homogeneous sampling of any background structure present. The accuracy that is needed of the background correction is determined by the signal that CoRoT aims to detect. The CoRoT design specifications set a detection limit of a 0.07% (700 ppm) change in flux, which equates to a drop of just 20 electrons within the average star mask. This is comparable to a planet of twice the diameter of the Earth around a star of magnitude $15^{\mathrm{m}}5$. The duration of a transit depends on the size of the orbit, but is about an hour. Thus, any background correction applied must have a stability better than 20 e$^-$ within a star mask over one hour. We focus on the exoplanet channel during our discussions, since the 400 windows allow more background modelling options.

3. The bright pixel problem

The commissioning run for CoRoT (henceforth IRa01) took place from 8 February to 3 April 2007. During examination of the full dataset, we noticed that all windows suffered impacts during the crossing of the SAA. Scientific analysis of the data will avoid using results that are flagged as SAA crossing moments because of these impacts. However, not every impact clears immediately out of the CCD structure. Some of them cause longer lasting damage, due to defects generated in the silicon lattice. The damaged pixels generate an excess of dark current and are therefore called bright pixels. The dark current generally decreases when/if the defects switch to a more stable configuration, but this can take several hours or days or be a permanent problem. These impacts can therefore cause a higher than average value of a background or star window and, in the case of the background windows, affect the correction accuracy significantly. Background over-correction on stars due to bright pixels in the background windows can mimic a planetary transit and create difficulties for the detection algorithms. The background correction methods fitting all available windows were envisaged as being resistant to a few windows with these bright pixel defects, but it was still interesting to look at methods to detect and correct for these bright pixel errors.

The CoRoT IRa01 data show different behaviours of these bright pixels (Pinheiro da Silva et al. 2008). For some bright pixels, the succession of decreasing steps, occurring with a probability that decreases with the time since impact, leads to an exponential decay-like shape. A good approximation of this decay can be achieved by fitting an exponential function with a time constant between a few hours to a few days. But other pixels have more complicated behaviour, with a succession of steps up and down to different levels, characteristic of a random telegraph signal (RTS). However for all bright pixels, we can consider that between two steps the dark current of the pixel is stable and the signal is useful for the background evaluation. This is why we consider a method for modelling all these steps, in order to correct them without changing the evolution of the signal between

two discontinuities. The goal is to obtain a clean signal with no trends or discontinuities any more.

A method developed for gamma-ray burst detection by Scargle (1998) was used to characterise intensity variation in a time-domain signal. The principle is to compute the most probable segmentation of the signal into time intervals during which the electron rate has no statistically significant variations. This is done by comparing the probabilities that the signal follows two different models. The first one, P1, is the probability that the signal is constant during the entire time interval. And the second, P2(t), is the probability that the signal has one value before the instant t and another one after. Probability P2(t) is computed for all possible values of t in the time interval. The ratio of these two different probabilities, P2(T)/P1, has a maximum value for the most probable moment (t) of a segmentation. If this ratio is greater than a threshold, the segment is cut into two different time intervals. Each new segment is also decomposed until no more segmentation is found. Since this method is based on Bayesian statistics, the resulting time intervals are called Bayesian blocks.

Before applying this algorithm, all the variations due to orbital or long-term perturbations and background fluctuations must be subtracted, to be sure that all the variations corrected by this method come from the bright pixels' evolution. This is done using an evaluation of the background at the position of the target window using windows in the neighbourhood. In other words, a local background fitting is performed to allow detection of the bright pixel that, once corrected, allows for more accurate background fitting. The method used for now is a principal component analysis (Murtagh 1993) of the ten neighbouring background windows. The first component is used as the local value of the background.

When all background variations are subtracted, the signal is a succession of segments with fluctuations around a constant value. No more orbital variation or trends are visible. The signal now can be decomposed into Bayesian blocks. Figure 1 shows an example modelling of a bright pixel evolution. We can see that only significant steps in the signal are modelled, while small fluctuations are preserved. The bright-pixel evolution model can then be subtracted from the original background to obtain a corrected background curve shown in Fig. 2. It turned out that the method by Scargle (1998) performs well in solving CoRoT's bright pixel problem. This has therefore been carried out on N1 background data from IRa01 and also 40 days from the first long run (21 May to 30 June 2007, henceforth LRc01). The following sections include an analysis of how the bright pixel correction affects the background correction.

4. The CoRoT data we studied

We present here mainly the results for the first long run (21 May to 30 June 2007, LRc01). For the IRa01 (commissioning run), the automatic background window placement convolution algorithm was run using 4 segments per CCD. This resulted in a non-optimal window placement. The LRc01 window placement used 49 segments per CCD, which resulted in a more uniform window placement. Since the LRc01 results are more representative of future runs, we focus on the results from this run. Figure 3 shows the window placement on full-frame images including the background. The full-frame images have been corrected for CCD offset and gain, high-energy particle impacts and electro-magnetic interference. We have also cut out the stars. We then ran a sliding box median over the images to show overall trends and gradients more easily and to smooth out the jumps caused by the star
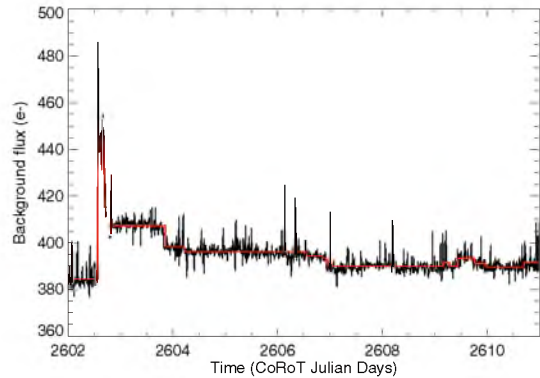
Fig. 1. Background window variations due to a bright pixel. Orbital components and general background variations are not shown on this curve. The red line is the result of the Bayesian block decomposition.
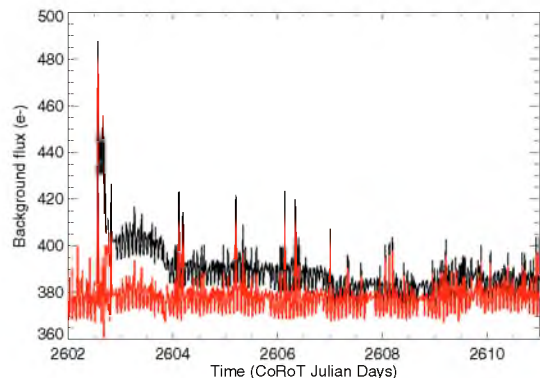
Fig. 2. Background window during a bright pixel creation. The red curve is the same background after correction with the Bayesian block model.

removal. The window positions are overplotted with 3 different symbols to show the window flux distribution. The symbols are larger than the true window size (10×10 px). We also include Fig. 4 to show the distribution of 512 s and 32 s windows without the full-frame image. A surface plot of the flux seen in these windows for one exposure from the LRc01 is shown in Fig. 5.

5. Background correction methodology

We examine here which background correction methods are best, both with and without the bright pixel correction. The reason for this is twofold. Firstly, the exact bright pixel correction method has not yet been fixed. Secondly, examining the improvement in background data accuracy allows us to judge the added value of the bright pixel correction in a way that is impossible on real star data due to the intrinsic science signal. The background correction methods described here, apart from the median method, were proposed and implemented by Costa and described in Samadi et al. (2006). What we present here is both the evaluation of these methods and the new median and distance-weighted average methods, using real data.

5.1. Closest background window

The simplest correction method examined uses the background window closest to the star in question. If the star is an oversampled one (32 s exposure) and the background window is averaged over 512 s, then the background correction can be interpolated to the correct time. In its simplest form, one background window
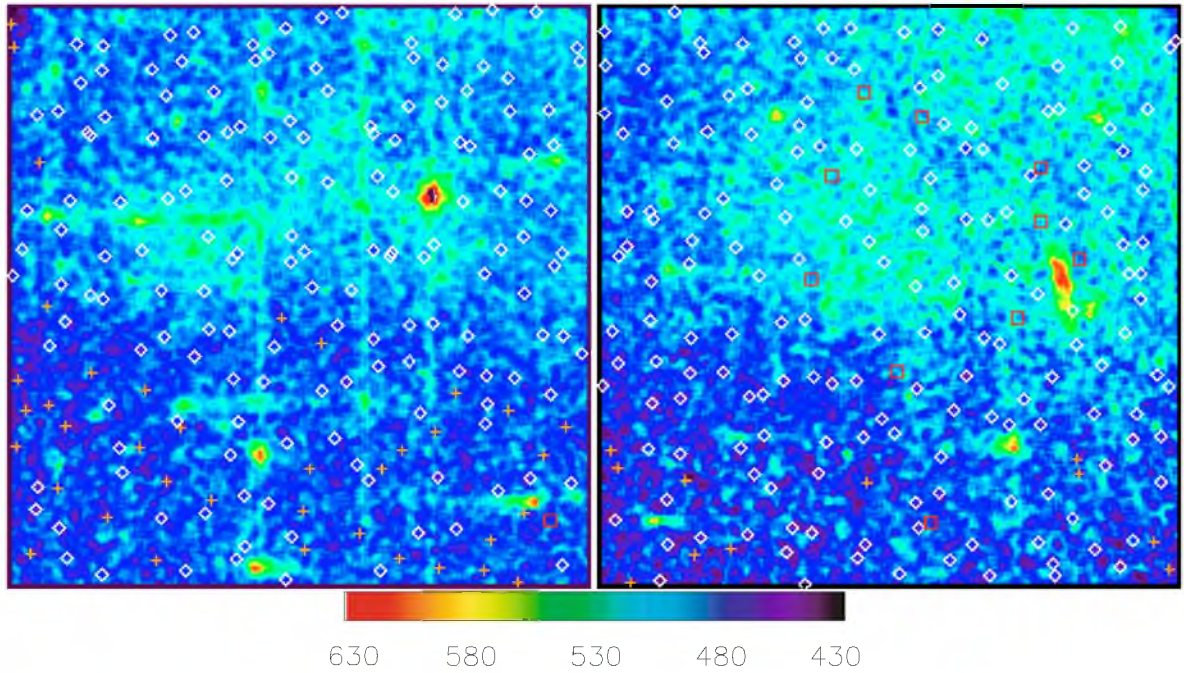
Fig. 3. *Left*: CCD 1, *right*: CCD 2. Full-frame images from the beginning of the LRc01. The level is shown in the colour table, in electrons. Overplotted are the background window positions, filtered by their average flux level over one orbit. The orange crosses represent windows of flux < 453 e⁻, the white diamonds of 453 e⁻ < flux < 530 e⁻ and the red squares of flux > 530 e⁻.
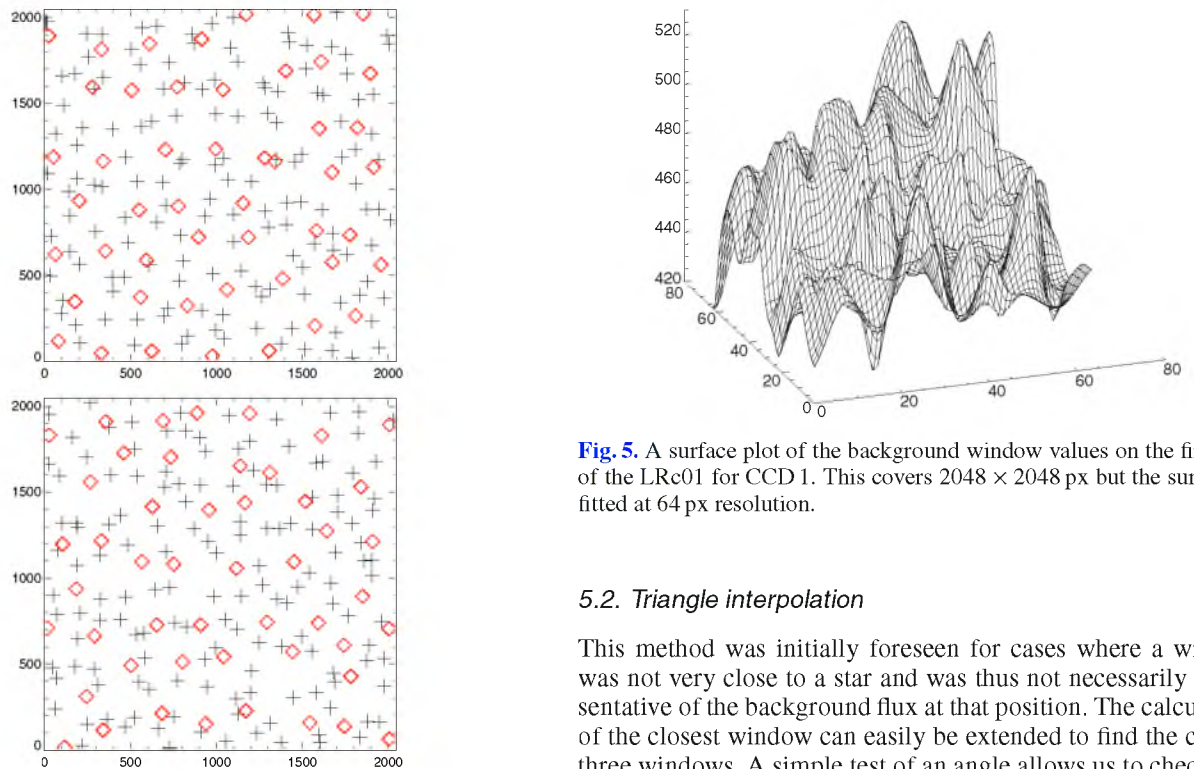
Fig. 4. The distribution of the background windows for the LRc01. The 512 s windows are marked with plus signs, the 32 s with diamonds. *Top*: CCD 1, *bottom*: CCD 2.

Fig. 5. A surface plot of the background window values on the first day of the LRc01 for CCD 1. This covers 2048 × 2048 px but the surface is fitted at 64 px resolution.

5.2. Triangle interpolation

This method was initially foreseen for cases where a window was not very close to a star and was thus not necessarily representative of the background flux at that position. The calculation of the closest window can easily be extended to find the closest three windows. A simple test of an angle allows us to check that a large enough surface of the CCD is covered with these three positions, and we extrapolate the background flux at the star position from the three background window positions and flux values. Ideally, this method is less vulnerable to one window with a bright pixel error.

5.3. Median

In view of the bright pixel problem, which affects several pixels on the CCD at any one time, a median method was included in
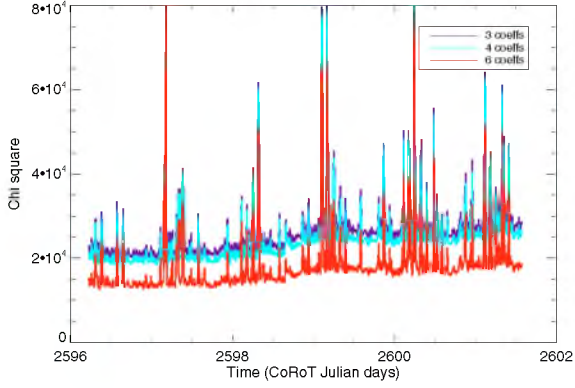
becomes the default for a star and is used constantly for the correction. A complexity could be to allow flagging of a background window with a bright pixel problem, to exclude it from use as a background correction until the excess flux has dropped away.

Fig. 6. Six days of data demonstrate the difference in the χ^2 values according to how many coefficients are used in the polynomial fit.

Fig. 7. Histogram of the differences between a window and the fit at its position using three to six coefficients. This shows a similar standard deviation for all three models, however there are more worse-case instances for the six-coefficient fit.

Fig. 8. Comparing a three-coefficient fit across both CCDS with a fit on only one CCD. Here we show the comparison for CCD 2.

the methods to be tested. A median is resistant to a few outlying background windows, but of course it can only represent a flat background. We therefore take the median of all 200 windows and consider this to be the background value across the whole CCD for that exposure.

5.4. Polynomial fitting

The natural step after using three windows to model background flux at a given point is to use all of the data available to us. This offers several extra options for the corrections. We can use all 400 data points across both CCDs for a fit covering 4096×2048 px every eight minutes. We may also consider modelling the background at a higher time resolution, by using the ~100 32-s windows to fit a polynomial 16 times more often. This would give us a better idea of orbital variations, but the spatial resolution is clearly lower when using fewer windows.

An initial look at the full frame images (Fig. 3) and their structure suggested that a polynomial fit of the following form would be appropriate:

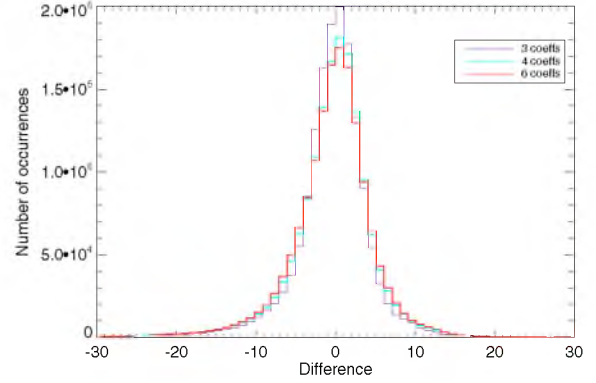$$\text{Flux} = a_0 + a_1 x + a_2 y + a_3 xy + a_4 x^2 + a_5 y^2. \qquad (1)$$

This represents a parabolic-shaped background, but we can imagine a flat gradient occurring, in which case using only a_0 to a_2 or a_3 is more appropriate. We performed tests on all of these options, in order to see if adding another degree of complexity to the fit decreased the χ^2 value of the result significantly. We perform a non-weighted multiple linear regression to the background windows and their coordinates, as this allows easy modification of the number of coefficients one wants to fit and proved robust enough (in comparison to the single value decomposition method also tested). The time factor was also part of this choice, as the regression fits six coefficients to 50 days of data (512 s sampling) in 10 s, compared to 8 min for the single value decomposition method.

The first tests were to determine the order of the polynomial needed and whether the CCDs should be fitted as one surface or separately. The quality of fit was judged by the non-normalised chi-squared value
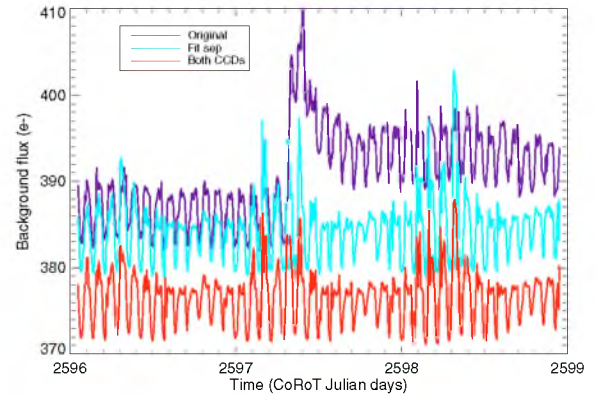
$$\chi^2 = \sum (F(x, y)_t - Z_t)^2, \qquad (2)$$

where $F(x, y)$ is the fit of the background values at positions x and y of the windows at time t, and Z is the background value of a window left out of the fit.

These tests showed that fitting six coefficients gave a lower χ^2 value than three (flux = $a_0 + a_1 x + a_2 y$) or four coefficients (Fig. 6). To quantify a few results more precisely, at time 2596.28 the relative χ^2 values are: three coeffs: 24 075, four coeffs: 22 371, six coeffs: 18 930. However, the histogram showed more worse cases for the six-coefficient model than for the three-coefficient one (Fig. 7). Differences of more than 5 e$^-$ are above the shot noise, as we will see later. The three and four coefficient fits have almost exactly the same model result, indicating that the extra $a_3 xy$ term is not necessary. We chose to focus on the three-coefficient fit, which also makes the fit faster, an important consideration for the pipeline implementation, since finding the closest window is a very fast process. The values shown here are for the IRa01. The same analysis was carried out for the LRc01, but again showed no significant improvement with an increase in the number of coefficients used for the fit. In the case of a similar histogram width, the simplest fit was taken to be the best option. Thus, in the results section, we present only results for the three-coefficient fit. We did not consider a one-coefficient fit (a mean) as we felt that the median method was more robust, considering the bright pixel impact problem.

The option of fitting both CCDs simultaneously (Fig. 8) was rejected fairly quickly. The two CCDs have a very small gap between them (less than 4 mm), so one would expect a uniform distribution of any background that penetrates the baffling, as can be seen in the full-frame images in Fig. 3. However, the 512 s exposures on each channel are not summed over the same 512 s. There are two electronic chains servicing the four scientific CCDs, and they are assigned to one asteroseismology CCD and one exoplanet CCD each. Both exoplanet CCDs are read out at the same moment every 32 s, but due to a telecommand

shift, CCD 2 starts reading out and summing exposures later than CCD 1. As a result there is a shift between data for the two CCDs. This is 128 s for the IRa01 and 96 s for the LRc01. We see a clear orbital variation in the background data and cannot neglect a phase shift by trying to fit both CCDs with one polynomial. Preliminary results supported this theory, as the χ^2 value for the fit over both CCDs was systematically higher than the fit per CCD.

5.5. DWA

Reegen et al. (2006) describes a method in between the median and closest window idea. For a pixel at position l, m at time n, we can model the background value using other background windows at various other positions (but the same time n) using

$$F_\lambda(x, y) \equiv \frac{\sum_{x,y} R_{lm} I_{lmn}}{\sum_{x,y} R_{lm}},$$
$$R_{lm} \equiv \left[(x - l)^2 + (y - m)^2 \right]^{\frac{-\lambda}{2}}, \tag{3}$$

where λ is a positive integer, representing a smoothing coefficient. This results in a correction where the closest window is the most important in the predicted background flux, but all windows are used, according to their relative distance to the point at which we wish to determine the levels. This is equivalent to a Whittaker filter. One may expect this method to be sensitive to bright pixel impacts pre-correction, in the same way as the closest window/triangle interpolation methods, but more reliable post-correction.

6. Comparison of correction methods using jackknifing

Jackknifing is a particular example of bootstrapping (Efron & Tibshirani 1994), whereby we increase our statistical analysis of a small dataset (N values) by running analyses on many datasets of $N - 1$ values. In the context of the background windows, this means that, to analyse the quality of each method of correction, we remove 1 of the 200 background windows per CCD from consideration in the method and then compare the result of the method with the value of the Nth datapoint. This is especially clear in a polynomial fit, where a fit of 199 windows does not suffer a large setback in quality, and the fit value at the position of the 200th window can easily be compared to the true background datapoint we have at that position. One can run this analysis up to 200 times and thus gauge the variance of the fit quality. This can also be done with the closest and triangle interpolation methods by taking the 200th point as a test value and using the correction code to find the appropriate correction. For the tests at 512 s, we can loop ~200 times, whereas the studies on the 32 s sampling is only performed ~50 times (depending on the number of windows). However, even 50 times the number of datapoints significantly improves our statistics on the best correction method.

For each method that passed preliminary inspection (χ^2 values and histogram of differences check), we carried out jackknifing tests. To compare the quality of the corrections, we examine the distribution of the differences between the predicted value at the jackknife position and the real data points. This does not, generally, centre on zero, which means combining data from many windows does not allow us to see the stability of the correction due to the offset included in the plot. Thus, for each jackknifed window time series, we take the mean and look at

the differences around this mean ($z - \overline{z}$). We use the variance of the distribution as a first measure of the quality of the correction, but also consider the shape of the distribution: does a particular method perform well overall, but poorly for a particular set of windows? This gives us the accuracy of the correction at either a 32 s or 512 s sampling. For easier comparison with the exoplanet channel requirements that are specified per hour (minimum detectable transit duration), we use a mean smoothing over the number of exposures that make up one hour (112 or 8) in order to see the deviation of differences over one hour.

CoRoT passes through the SAA on most orbits. Since science will not be performed during the SAA crossing, we remove these datapoints from our analysis using the orbital events log from CoRoT. The statistics we show are therefore outside the SAA. This reduces the number of points for our statistics, so jackknifing 50−200 times is very beneficial. The accuracy of the background correction we can attain depends on the average level of background in the image. For the data we have studied, this is about $400 \, e^-/\text{px}/32$ s. This has a shot noise of $20 \, e^-/\text{px}/32$ s, and the background windows cover 100 pixels on the exoplanet channel. If we assume that we are also photon-noise limited in the background windows, then the accuracy with which we can measure a background window value is $2 \, e^-/32$ s. All of the values we discuss in this section are per pixel per 32 s results. However, the 512 s sampling has an average of 16 values, so its photon noise should be reduced by a factor four. This means the measurement value of one window is accurate to within $20 \, e^-/\text{px}/32$ s and $5 \, e^-/\text{px}/512$ s.

7. Results

Jackknifing tests were performed on all of the options presented in Sect. 5 on a dataset from each run. We focus on the performances of the five best methods on the LRc01. These were the closest window, triangle interpolation, a median fit, and a three-coefficient fit of the form $a_0 + a_1 x + a_2 y$. Each of these were performed on the two CCDs separately. We examined both pre-bright pixel correction and post-correction data at 512 s sampling. We also examine the difference in accuracy if we perform these fits at 32 s sampling rather than 512 s. For this we used only bright-pixel corrected data.

7.1. Closest window and triangle interpolation

The jackknife analysis for the closest window and triangle interpolation uses all ~200 windows per CCD, summed to 512 s if necessary. For each loop, one of these windows is extracted as the test and its closest background window (or three closest) are found. The difference between the two background values is then examined. To plot the histogram for all of the windows together, we take the mean of the difference for each jackknife window, to centre each of the windows' differences around zero. If we run a sliding median along the difference, covering eight exposures (8×8 min ~1 h), then we get the histograms in Figs. 9 and 10.

The first thing we notice in these figures and in Table 2 is the very clear improvement in results once the data have been corrected for bright pixels. The rms deviation of the histogram decreases significantly in all cases. The second thing we notice is that the triangle interpolation is not better than the closest window. The poor results from using three windows before bright pixel correction is to be expected, since we increase the chances of using an impacted background window by selecting three windows. The poor results after bright pixel correction point to
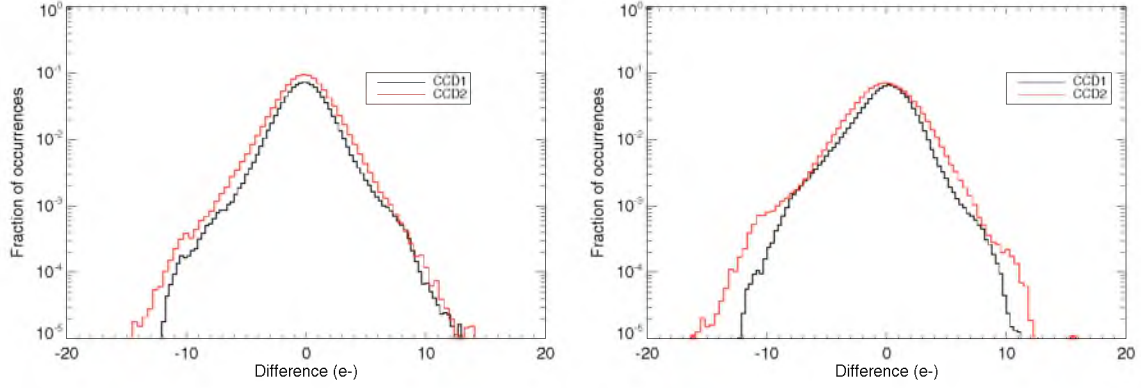
Fig. 9. *On the left* – the histogram of the differences for the closest window correction for the LRc01 using all windows. *On the right* – the same using only 32 s windows. Both use data corrected for bright pixels and smoothed over one hour.
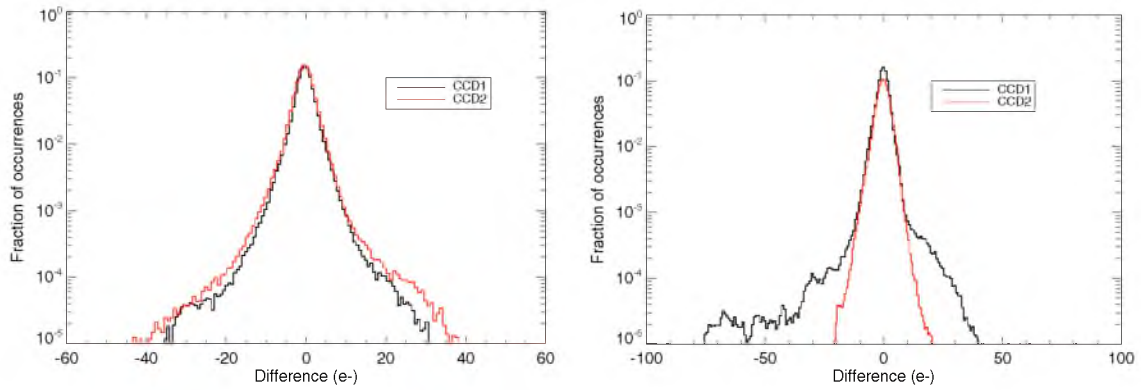
Fig. 10. *On the left* – the histogram of the differences for the triangle interpolation correction for the LRc01 using all windows. *On the right* – the same using only 32 s windows. Both use data corrected for bright pixels and smoothed over one hour. The abscissae ranges are not the same on these two plots due to extreme values that needed to be included.

Table 2. The rms deviation σ (e$^-$/px) of the histogram of differences between the estimated background flux and the real background flux for all of the background correction methods (LRc01).

	Before bright pixel correction		After	
Closest window	$\sigma_{512\,s}$	$\sigma_{1\,h}$	$\sigma_{512\,s}$	$\sigma_{1\,h}$
CCD 1	15.14	13.29	3.81	2.05
CCD 2	8.32	7.65	3.91	2.21
Triangle interpolation				
CCD 1	14.94	12.99	5.53	3.00
CCD 2	12.73	11.84	5.63	3.22
Median	$\sigma_{512\,s}$	$\sigma_{1\,h}$	$\sigma_{512\,s}$	$\sigma_{1\,h}$
CCD 1	10.67	9.36	2.91	1.77
CCD 2	6.05	5.58	3.05	2.00
3-coefficient fit				
CCD 1	10.52	9.23	2.74	1.55
CCD 2	5.96	5.51	2.89	1.82
DWA				
CCD 1	19.18	16.83	1.66	1.13
CCD 2	5.34	5.18	1.93	1.47

poor selection of three representative background windows or to a background that is not locally flat.

7.2. Median background value per CCD

The advantages of using a median value of 200 background windows are clear. We can avoid any effect of bright pixels, unless half the windows are affected: it is very robust, can be calculated

quickly, and the result allows easy examination of the overall change in background flux over time. Of course, if the background contains a great deal of structure, then this simple solution will not give a good fit. The jackknife method was used as previously, even though leaving out one window has almost no effect on the median value. In fact, the results gave only two different values over the whole jackknife sample for one CCD per exposure, differing by e.g. 0.15 e$^-$/px/512 s. To see how representative this median value was of each of the windows, we again plotted a histogram of the differences (Fig. 11).

Figure 11 and Table 2 show that the median method performs better than the closest window method. This is most noticeable in the pre-bright pixel correction results, which means that this method would be better for the validation data than the closest window method initially foreseen for the pipeline. A good quality fit from a median value over time implies that the background variation is flat. We can see this in the histogram widths, but also in other values that we do not plot here. An examination of the offset of each window from the median (the mean value that we use to centre the histogram) shows that 60% of the offsets are less than 20 e$^-$/px away from the median value ($\sim$470 e$^-$ for the LRc01). The most extreme value is a 60 e$^-$ offset. Further statistics on the histogram show that 98% of the differences are less than 10 e$^-$/px. This demonstrates that the offset does not vary much. The science on the exoplanet channel is not as sensitive to a constant background correction of 20 e$^-$ too little or too much, but it is sensitive to a variation in this correction, since this can mimic a transit event. The median correction could be iterated to permit post-processing correction of an offset, if it is detected
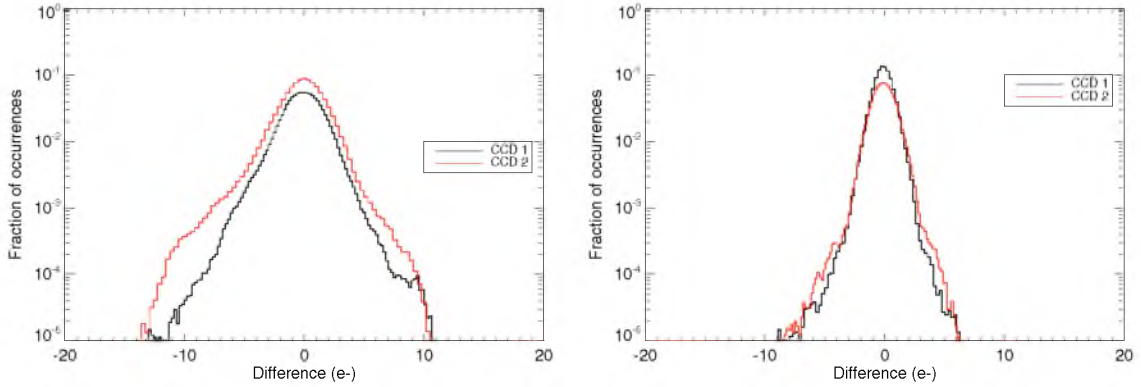
Fig. 11. Histograms for the median method, for LRc01 data corrected for bright pixels and smoothed over one hour. *On the left* we use all the windows, *on the right* only the 32 s.
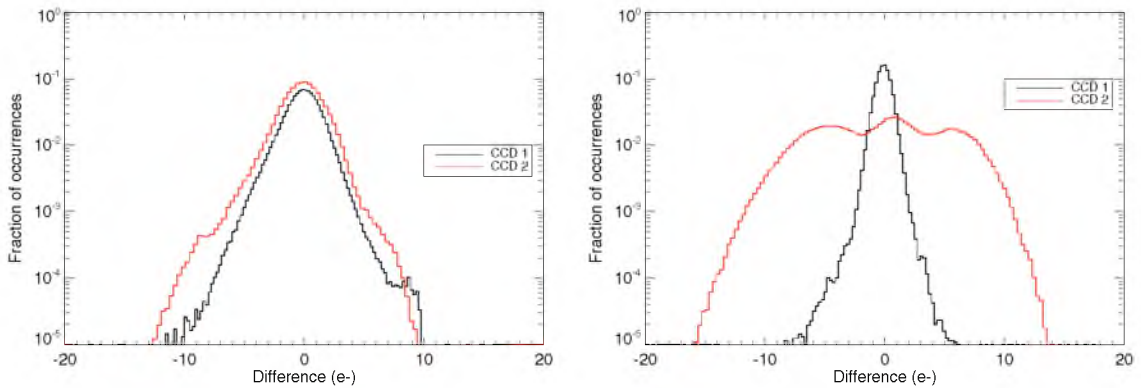
Fig. 12. Histograms for the three-coefficient fit, using LRc01 data corrected for bright pixels and smoothed over one hour. *On the left* we use all the windows, and *on the right* the 32 s only.

over a long period of time; however, the raw method still shows an improvement over the closest window correction.

Once again, the bright pixel correction improves the accuracy of the background fit significantly. We are now reaching a standard deviation of the histogram of about 2 e⁻/px.

7.3. Three-coefficient fit

The initial assumption was that there would be a complex gradient across the CCDs due to differential scattering of the light that manages to penetrate CoRoT's baffling. This is seen to some extent in the full-frame images (Fig. 3) with generally higher flux in the upper-half of the image but no clear structure to the overall gradient. For both of the runs, the three-coefficient polynomial (slope in two directions) fits the windows with almost as low a χ^2 as the six-coefficient polynomial. This leads us to expect a similar jackknife result from the three-coefficient polynomial to that achieved with the median method.

Figure 12 and Table 2 confirm this theory. The widths of the histograms are very similar to those of the median method, both before and after bright pixel correction. For the LRc01, the better results for both the median and three-coefficient methods before bright pixel correction are seen on CCD 2, whereas this switches to CCD 1 after correction. This implies that CCD 1 had more bright pixel impacts.

7.4. DWA

The distance-weighted average method has the same problem as the closest and triangle interpolation pre-bright-pixel correction.

If the nearest windows experience an impact, then the predicted window flux is far too high. Post bright-pixel correction, however, this gives very good results, as shown in Fig. 13. We tested the jackknifing for various λ values and got the best result for $\lambda = 1$. Using all of the windows, we get a better result with the DWA than for the median method; however, for the 32 s windows, the median method has a lower rms deviation of the histogram. This is no doubt due to the wider spacing of the 32 s windows, meaning that the closest windows can be at some distance and are not necessarily the most representative of the background flux. The lower number of 32 s windows, along with their spacing, means that the median method is more reliable – using a constant value across the entire CCD.

7.5. 32 s sampling results

In the first section of these analyses, we have used all windows summed to 512 s. However, we have many stars sampled at 32 s and they require a 32 s-sampled background correction. For the closest and triangle interpolation analyses, we performed the jackknifing using only the 32 s windows as the jackknife sample. However, we allowed either 32 s or 512 s windows to be found as the closest window and interpolated a correction value as necessary (a linear interpolation with time). For the median and three-coefficient fit, we used only the ~50 windows available at 32 s sampling on each of the CCDs to perform the jackknifing. We chose to present only the bright pixel corrected values, for brevity.

The 32 s sampling results (Table 3) follow the same general pattern as the results from all of the windows, except for the
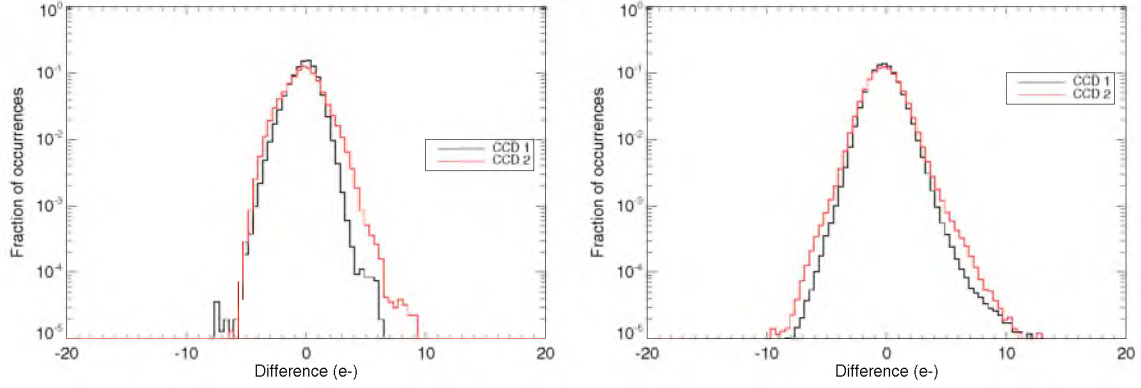
Fig. 13. Histograms for the distance-weighted average method, using LRc01 data corrected for bright pixels and smoothed over one hour. *On the left* we use all the windows, and *on the right* only the 32 s.
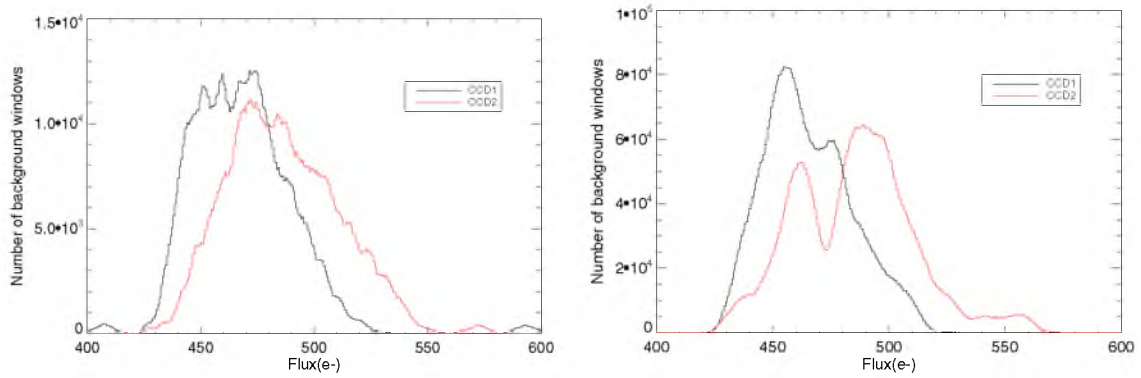
Fig. 14. The flux distribution of only the 512 s windows (*left*) and only the 32 s windows (*right*).

Table 3. The rms deviation σ (e$^-$/px) of the histogram of the differences between the estimated background flux and the real background flux for all of the methods for the 32 s windows, for the LRc01 corrected for bright pixels.

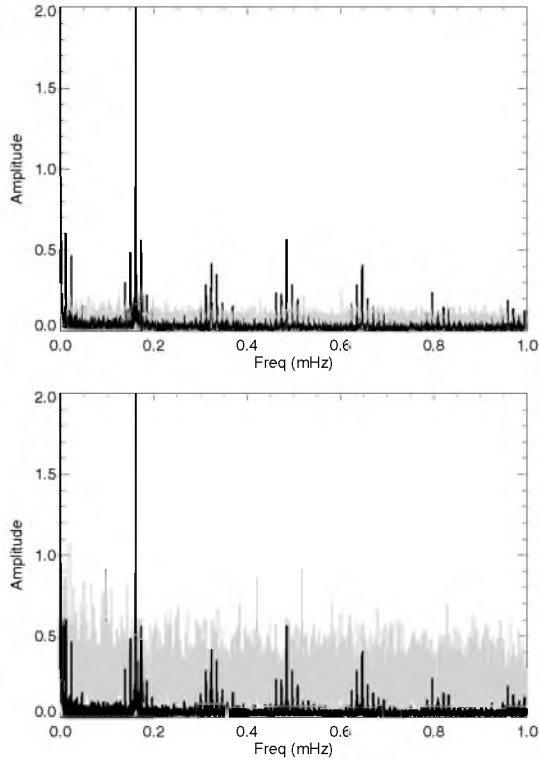| Closest window	$\sigma_{32\ s}$	$\sigma_{1\ h}$
CCD 1	4.55	2.31
CCD 2	4.73	2.61
Triangle interpolation		
CCD 1	7.50	3.89
CCD 2	5.43	2.91
Median		
CCD 1	3.18	0.82
CCD 2	3.27	0.99
3-coefficient fit		
CCD 1	2.90	0.67
CCD 2	8.47	5.00
DWA		
CCD 1	3.04	1.28
CCD 2	3.14	1.45

LRc01 CCD 2 results for the three-coefficient fit. The closest window gives a good correction, after bright pixel correction, but the median and three-coefficient fits are better still, except for CCD 2 of the LRc01. For the LRc01, this is the worst CCD 2 result. We investigated the cause of this, plotting the flux distribution of the 512 s windows compared to those of the 32 s windows for the LRc01 (Fig. 14). The lefthand panel of Fig. 14 shows the distribution of flux in the 512 s windows. CCD 2 has a wider range, but there is no significant other difference between the two CCDs. However, the righthand panel shows the very different 32 s flux distribution. Both CCDs have a dip, but it is more pronounced on CCD 2. The turning point occurs at 473 e$^-$, with the range of fluxes with a lower number of windows encompassing 460–490 e$^-$. We examined which windows belonged to each side of the turning point. There was a slight tendency for the top of the CCD to be in the higher flux section of the histogram (as we saw in the full-frame images), but this was not a clear division. Almost all of the windows in the bottom half of the CCD had fluxes above 473 e$^-$ at some point, 4 of them all of the time. Similarly, windows in the top half of the CCD also had values below 473 e$^-$. Thus, this flux division does not appear to be spatially-dependent, nor is it time-dependent. Further investigation into this strange behaviour, seen only in the 32 s sampled data, forms part of our future projects.

8. Discussion

Jackknifing allows us to identify the behaviour of the sampled points with respect to one another, allowing for more statistical analysis than is possible with just 200 points every eight minutes. This also shows the robustness of any fit and highlights outlying windows.

We show here the jackknife test for one background window from each run in Fourier space. We cannot conclusively test background correction on a star spectrum, due to the inherent science signal so we show here the initial spectrum of a background window compared to its spectra after the various correction methods have been applied. These are shown for bright pixel corrected data.
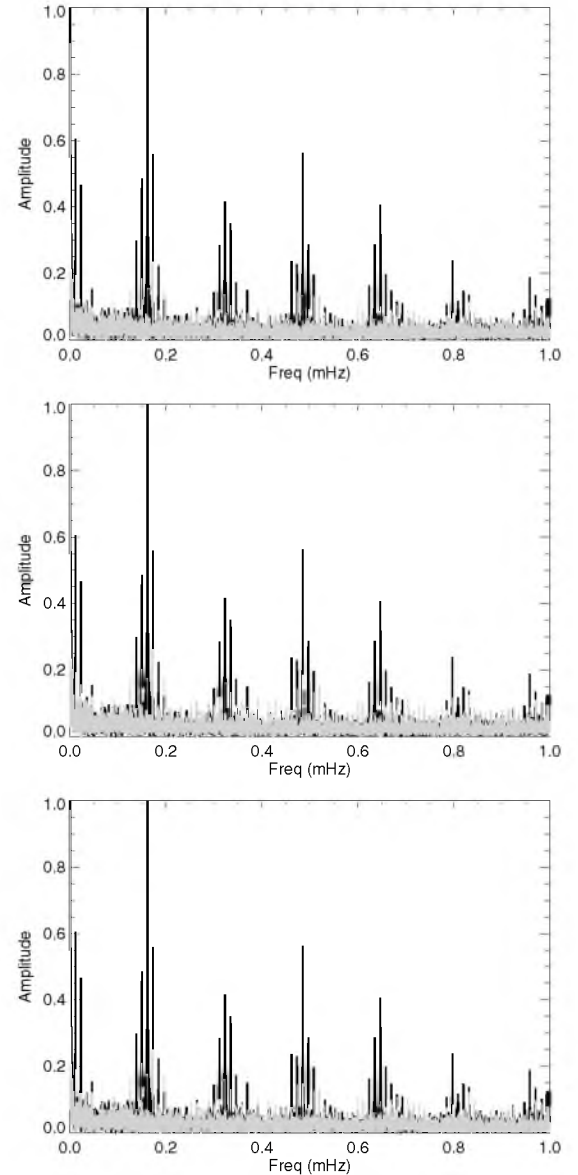
Fig. 15. Spectrum of a background window for the LRc01 before and after correction. In black – the original measured jackknife window. *Top*: in grey – the closest window correction. *Bottom*: in grey – the triangle interpolation background spectrum. The amplitude is in e⁻/px/32 s. A typical mask size is 75 pixels.

The main peak is the orbital frequency peak and its aliases. CoRoT's orbital period is $6180\,\mathrm{s}$, which gives a frequency $f_0 \sim 0.16\,\mathrm{mHz}$. This has the highest amplitude in the pre-correction spectra, but is the most clearly corrected peak for all methods.

Figures 15 and 16 show the spectra for the LRc01. Here, the closest window is clearly not the best option. It can correct for the primary orbital frequency, but does not reduce the overall noise levels. The triangle interpolation also performs badly, increasing the overall noise levels to higher than the harmonics of the orbital frequency. The correction of the three-coefficient fit, distance-weighted average, or median method is far better, removing the orbital frequencies and lowering the general noise. Between these three, the three-coefficient fit performs slightly better, leaving fewer spurious peaks and significantly reducing the orbital peaks. The results are very similar; all these three methods correct for the orbital variations seen in the background levels. These spectra do not show the full amplitude of the peaks, in order to show the small variations more clearly. Before correction on the LRc01, f_0 has an amplitude of $3\,\mathrm{e}^-/\mathrm{px}/32\,\mathrm{s}$.

During our studies we also applied these background corrections to pulsating stars selected from the IRa01. These showed the clearest improvement over long time scales, as the overall straylight levels increase as Earth's orbit around the Sun decreases the angle between CoRoT's line of sight and the Sun's position. Examining these corrections in Fourier space demonstrated that the background light is not a significant factor in the frequencies, and is therefore not a concern for asteroseismology.
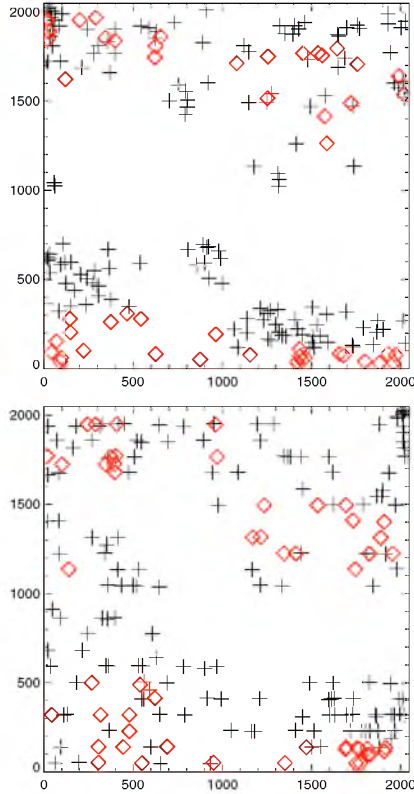

Fig. 16. Spectrum of a background window for the LRc01 before and after correction. In black – the original measured jackknife window. *Top*: in grey – after three-coefficient fit correction. *Middle*: in grey – after median correction. *Bottom*: in grey – after distance-weighted average correction. The amplitude is in e⁻/px/32 s. A typical mask size is 75 pixels.
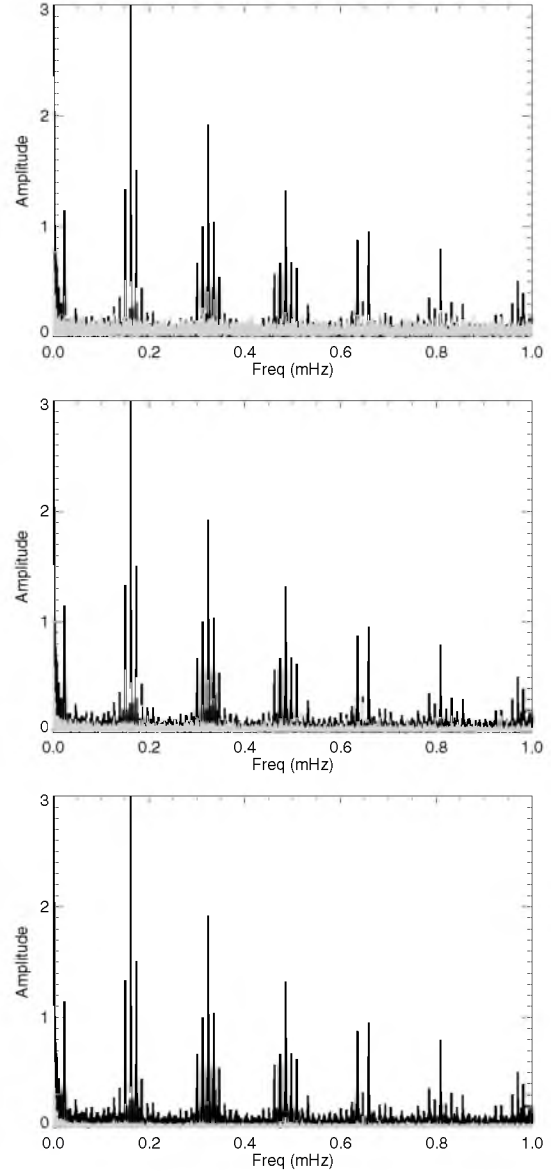
9. IRa01

We have mentioned that the position assignment of the background windows was not optimal for the IRa01. Their placement is shown in Fig. 17. The results of the same jackknifing study highlight the importance of window placement in analysing the quality of the background correction methods.

Figures 18 and 19 show the results for the IRa01 and confirm what we see in Table 4. The closest window correction is seen to be a fairly good correction method, especially after bright-pixel correction. The triangle interpolation performs very badly, which may indicate a non-locally flat background, or simply be indicative of poor window placement. Of course, here we are analysing the correction on a test background window, which is therefore not representative of a star in the middle of the CCD where we have few background windows. The

Fig. 17. The distribution of the background windows for the IRa01, the 512 s windows are marked with plus signs, the 32 s with diamonds. *Top*: CCD 1, *bottom*: CCD 2.

Fig. 18. Spectrum of a background window from the IRa01 before and after correction. In black – the original measured jackknife window. *Top*: in grey – the closest window correction. *Bottom*: in grey – the triangle interpolation background spectrum. The amplitude is in e⁻/px/32 s. A typical mask size is 75 pixels.
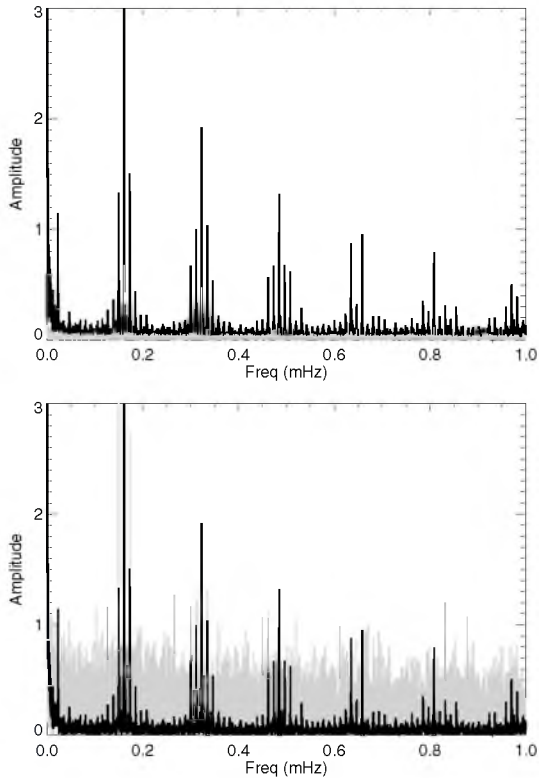
Fig. 19. Spectrum of a background window from the IRa01 before and after correction. In black – the original measured jackknife window. *Top*: in grey – after three-coefficient fit correction. *Middle*: in grey – after median correction. *Bottom*: in grey – after distance-weighted average correction. The amplitude is in e⁻/px/32 s. A typical mask size is 75 pixels.

median and distance-weighted average methods give the best histogram rms deviation, and the spectra also show significant improvement. There are slight remnants at the orbital peaks, but the noise is generally reduced. The triangle interpolation increases the general noise levels in the spectrum and the three-coefficient fit also does not reduce this general level. Both the median and the distance-weighted average methods give a clean spectrum – with the DWA outperforming the median method. The median method was studied earlier and recommended for inclusion in the CoRoT pipeline. This has already been implemented. Extending our study to include the distance-weighted average showed that this is a highly performant method post-bright-pixel correction. Pre-bright pixel correction, however, it is slightly more sensitive to the number of impacts, as seen in Table 2. Since the bright pixel correction can only be performed once a longer time series of data has been obtained, it is useful to

Table 4. The rms deviation σ (e$^-$/px) of the histogram of differences between the estimated background flux and the real background flux for the IRa01.

Before bright pixel correction			After	
Closest window	$\sigma_{512\,s}$	$\sigma_{1\,h}$	$\sigma_{512\,s}$	$\sigma_{1\,h}$
CCD 1	6.98	5.70	4.45	1.92
CCD 2	7.44	6.29	4.38	1.89
Triangle interpolation				
CCD 1	23.19	19.28	13.84	5.68
CCD 2	31.09	24.97	20.15	8.04
Median	$\sigma_{512\,s}$	$\sigma_{1\,h}$	$\sigma_{512\,s}$	$\sigma_{1\,h}$
CCD 1	5.08	4.11	3.48	1.79
CCD 2	5.28	4.38	3.39	1.69
3-coefficient fit				
CCD 1	6.34	5.36	3.92	1.98
CCD 2	5.39	4.49	3.45	1.74
DWA				
CCD 1	4.21	3.93	1.97	1.29
CCD 2	5.15	4.81	2.34	1.49

maintain the median method to correct the light curves for preliminary analysis. It would, however, be useful to implement the distance-weighted average method in the CoRoT pipeline.

10. Conclusions

We have seen that a bright pixel correction is vital to ensuring the accuracy of the N1 background correction of CoRoT data. The method discussed works as long as few enough windows are affected by the bright pixel problem. We used ~10 nearby windows to identify the variations caused by true background variations. If too many windows are affected, this will become more difficult. A solution to this problem is being implemented in the on-board software, to allow summation of the background flux avoiding permanent hot pixels. This software will allow exclusion of pixels within a background window if they are identified as problematic.

Jackknifing has allowed us to examine four possible background correction methods for the bright pixel corrected background data. For the IRa01, where the window placement was not optimal, the median method gave the best results in these studies. We have shown that the median method gives a good correction for the background spectrum. However, we still see residual peaks in the star spectra. This may come from other instrumental effects and will be examined further. For the LRc01, we see the best results for the 512 s bright-pixel corrected background data if we use a three-coefficient fit. The 32 s windows on CCD 2 have a strange flux distribution that means that the three-coefficient fit does not represent the surface well. The 32 s star corrections can be carried out by interpolating the 512 s values, or by using a 32 s median. The median method has now been implemented in the CoRoT pipeline and is being used to correct science data.

We have presented a strategy for analysing which background correction method offers the best correction for each run. These analyses should be carried out on each new run. For future runs we recommend re-running the analysis for the polynomial fit to see how many coefficients best represent the background shape. This should be compared to the full-frame images acquired at the beginning of each run in order to confirm that the window placement represents the background shape. Jackknifing can confirm which the best method is for background correction. Further work could include studies of the background levels as a function of the Sun's position to the line of sight of the baffle, as well as considering and resolving linear correlations between background windows, as in Reegen et al. (2006) for the MOST data reduction. It would be useful to test this method and implement it, now that the baffle's performance turns out to be of such high quality that the main source of background light is the Earth's scattered light.

References

Baglin, A., Auvergne, M., Barge, P., et al. 2006, in The CoRoT Mission: Pre-launch status, ESA-SP 1306, 33
Barge, P., Léger, A., Ollivier, M., et al. 2006, in The CoRoT Mission: Pre-launch status, ESA-SP 1306, 83
Boisnard, L. & Auvergne, M. 2004, in IAC-04-IAF-Q.1.01
Boisnard, L., & Auvergne, M. 2006, in The CoRoT Mission: Pre-launch status, ESA-SP 1306, 19
de Oliveira Fialho, F., & Auvergne, M. 2006, in The CoRoT Mission: Pre-launch status, ESA-SP 1306, 289
Drummond, R., Vandenbussche, B., Aerts, C., De Oliveira Fialho, F., & Auvergne, M. 2006, PASP, 118, 874
Drummond, R. J. 2007, Ph.D. Thesis, Katholieke Universiteit Leuven, Belgium
Efron, B., & Tibshirani, R. J. 1994, An Introduction to the Bootstrap: Monographs on Statistics and Applied Probability, 1st edn. (Chapman & Hall)
Lapeyrere, V. 2006, Ph.D. Thesis, École Doctorale d'Astronomie et d'Astrophysique d'Ile de France
Llebaria, A., & Guterman, P. 2006, in The CoRoT Mission: Pre-launch status, ESA-SP 1306, 289
Murtagh, F. 1993, in Central Activity in Galaxies. From Observational Data to Astrophysical Diagnostics, ed. A. Sandqvist, & T. P. Ray (Berlin: Springer Verlag), Lecture Notes in Physics, 413, 209
Pinheiro da Silva, L., Rolland, G., Lapeyrere, V., & Auvergne, M. 2008, MNRAS, 384, 1337
Plesseria, J. Y., Mazy, E., Defise, J. M., et al. 2004, in Proceedings of the 5th International Conference on Space Optics, ESA SP-554, ed. B. Warmbein, 543
Reegen, P., Kallinger, T., Frast, D., et al. 2006, MNRAS, 367, 1417
Samadi, R., de Oliveira Fialho, F., Costa, J., et al. 2006, in The CoRoT Mission: Pre-launch status, ESA-SP 1306, 317