An algorithm for correcting CoRoT raw light curves[★]

D. Mislis^{1,2}, J. H. M. M. Schmitt¹, L. Carone³, E. W. Guenther⁴, and M. Pätzold³

¹ Hamburger Sternwarte, Gojenbergsweg 112, 21029 Hamburg, Germany
 e-mail: mdimitri@hs.uni-hamburg.de

² Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK

³ Rheinisches Institut für Umweltforschung, Abteilung Planetenforschung, an der Universität Köln, Aachener Str. 209, 50931 Köln, Germany

⁴ Thüringer Landessternwarte Tautenburg, Sternwarte 5, 07778 Tautenburg, Germany

Received 1 March 2010 / Accepted 13 July 2010

ABSTRACT

We introduce the CoRoT detrend algorithm (CDA) for detrending CoRoT stellar light curves. The algorithm CDA has the capability to remove random jumps and systematic trends encountered in typical CoRoT data in a fully automatic fashion. Since enormous jumps in flux can destroy the information content of a light curve, such an algorithm is essential. From a study of 1030 light curves in the CoRoT IRa01 field, we developed three simple assumptions upon which CDA is based. We describe the algorithm analytically and provide some examples of how it works. We demonstrate the functionality of the algorithm in the cases of CoRoT0102702789, CoRoT0102874481, CoRoT0102741994, and CoRoT0102729260. Using CDA in the specific case of CoRoT0102729260, we detect a candidate exoplanet around the host star of spectral type G5, which remains undetected in the raw light curve, and estimate the planetary parameters to be $R_p = 6.27 R_E$ and $P = 1.6986$ days.

Key words. methods: data analysis – surveys – techniques: photometric – planets and satellites: detection

1. Introduction

The CoRoT satellite was successfully launched in 2006. Onboard CoRoT, there is a small 27 cm telescope feeding two science channels to study astroseismology and transits, respectively (Baglin et al. 2000). The CoRoT has a field of view (FOV) of $\sim 2.7^\circ \times 3.05^\circ$. In its first observed field (IRa01 – $\alpha = 6^h 46^m 53^s$ & $\delta = -00^\circ 12' 00''$), CoRoT observed continuously for 60 days, producing uninterrupted light curves for the first time. The data for the IRa01 field have been public since December 2008 and the astronomical community has access to these data. Unfortunately, the CoRoT light curves are affected by a variety of instrumental problems that severely hamper the data interpretation. To overcome these difficulties, we have developed the CoRoT detrend algorithm (CDA). In this paper, the algorithm is presented and we demonstrate its functionality on some typical CoRoT data sets.

2. CoRoT light curves: the problems

The CoRoT data files contain multi-color light curves that were produced by inserting a low-resolution dispersing prism into the telescope beam. This set-up is designed to provide simultaneous light curves in the red (R), green (G), and blue (B) bands, although, these bands do not correspond to true photometric filters and, the bands may indeed differ from star to star. In this paper, we study the multi-color data, but also consider the total (white) flux obtained by summing up the individual light curves through $W = R + G + B$.

Figure 1 are typical CoRoT light curves from IRa01. The first panel of Fig. 1 shows a typical exponential jump very similar to a flare star. A trend is also evident. In the second light curve, there appears to be a box-shape jump, and in the third and fourth light curves one can discern features similar to those in the first and second light curves, except that the jumps are downwards. We note that the downward jump in the third light curve is very similar to a transit event, thus making the detection of true transits difficult. Combinations of all the above features appear to produce a rather typical CoRoT light curve. There are, two basic instrumental problems with all CoRoT light curves. First, there is a long-term trend, forcing a secular decrease in the light curve intensity over the full observing period of 60 days. The strengths of the trends in different sources may be different; the physical cause of these trends is not well understood. The second and even more serious problem is the instrumental jumps in the light curves. The term “jump” refers to a sudden variation in intensity without any obvious reason. Many of these jumps resemble stellar flares, although, the vast majority are clearly instrumental nature. Their physical explanation could be cosmic radiation and the time evolution of bright pixels (Pinheiro da Silva et al. 2008). These jumps are a random phenomenon and affect each filter differently. An inspection of hundreds of CoRoT light curves similar to those presented in Fig. 1 allows us to classify the observed shapes of jumps into five groups:

- sudden intensity increase and exponential decrease (Fig. 1 – panel a);
- sudden intensity increase and decreases (box shape, Fig. 1 – panel b);
- sudden intensity decrease and exponential increase afterwards (Fig. 1 – panel c);

[★] The code is only available in electronic form at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via <http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/522/A86>

Fig. 1. Jumps and trends in CoRoT light curves. CoRoT01027 –21492 a), –24482 b), –40879 c), –49307 d), –27431 e).

Table 1. Statistical analysis of 1030 CoRoT light curves from IRa01.

<i>R</i> filter	<i>G</i> filter	<i>B</i> filter	Total
38.14%	14.4%	15.1%	67.6%

Notes. Jumps appear in more than 50% of all light curves in all filters; in 0.82% of all light curves, jumps in all filters occur at the same time.

- sudden intensity decrease and increase (negative box shape, Fig. 1 – panel d);
- all of the combinations above (Fig. 1 – panel e).

A statistical analysis of IRa01 field (visual inspection) shows that only a small minority (Table 1) of all jumps are powerful enough to simultaneously appear in each colour. Most of the light curves are affected not only by one single jump, but by many jumps occurring in the different filters at different times. In Table 1, we show the results of a statistical study of the appearance and the shapes of jumps using data for IRa01. The three first columns of Table 1 show the number of light curves affected by jumps in the respective filter filter and the fourth column indicates the total size of the effect.

3. The CDA algorithm

3.1. General features

It is quite difficult to describe all the features perturbing a CoRoT light curve with a given function, since there are many different shapes of jumps with many different functional forms. Furthermore, the problem is complex because we do not know which light curve features are real signals (real transits, real flares etc.) and which are instrumental effects. The algorithm is based on three assumptions. We first assume that trends appear in almost all light curves and both flux increases and decreases can occur. The trends are not periodic and we assume them to be a long-term phenomenon (Aigrain et al. 2009). Our second assumption is based on the statistical analysis results for the data. The study of 1030 light curves from IRa01 field shows that only 0.82% of them are affected by a jump in all three filters at the same time. In these cases, the jump is very large and affects all bands with the same temporal pattern, although, in most cases the jumps affect only one band at any given time (Fig. 2). We therefore ignore the cases in which jumps occur simultaneously in all three bands. Our third assumption is that real transits must appear in all three filters, while, of course, the intensity and transit depth can vary from filter to filter. In summary, for the CDA we assume that

- long-term trends appear in all CoRoT light curves;
- jumps are random phenomena appearing in different filters at different times;
- the real signals from transits appear in all three bands.

We emphasize that CDA works only for events (such transits) that appear in two or more bands. Hence CDA does not work for stellar flares, since most stellar flares do not show any flux enhancements in the red and green band, but only in the blue band. In these circumstances, CDA will destroy real signals, unless the flare is powerful enough to appear in all bands.

Fig. 2. CoRoT0102729260. Three filter light curves (*R* a), *G* b), *B* c)) from a data set. The jumps in the red light curve do not appear in the other filters and vice versa.

3.2. The algorithm

To remove instrumental feature, CDA uses simultaneously all the colour light curves. The basic idea of CDA is to use the cleanest filter band as a proxy for the whole light curve. The raw data files of each CoRoT light curve have a quality flag (CoRoT files – Col. 4), indicating the quality of each data point (Mazeh et al. 2009). We first remove all the “bad points” (points flagged by CoRoT to have high noise); we note that these “bad points” are the same for all the filters of each star. In this paper, we use light curves from which all “bad points” have been removed (as in Fig. 1). As noted in our first assumption, trends are a long-term phenomenon. A 3rd degree polynomial is fit to the entire light curve to remove the trend in each filter per star. Because each

Fig. 3. Standard deviation versus number of blocks.

CoRoT light curve typically has thousands of data points, the polynomial does not fit short-term variations and real short-term events such as transits. We thus write

$$\text{Flux} = a + b \cdot \text{JD} + c \cdot \text{JD}^2 + d \cdot \text{JD}^3, \quad (1)$$

where JD is the Julian date (normalized to range $-1 \leq \text{JD} \leq 1$) and a , b , c , and d are the fit parameters for the third degree polynomial. At the end of this procedure, we have a detrended light curve per filter for each star.

After this step, CDA proceeds in removing the jumps. To identify the cleanest light curve for a reliable jump removal, we create “sub-light curves”, which have a typical duration of a day. Thus, for the IRa01 field we create 60 “sublight” curves, called simply light curves in the following. These 60 blocks were selected after we checked various combinations. If the number of blocks is too small, the probability of including a jump in the “sublight” curve increases. Figure 3 shows the optimal block number versus standard deviation.

We assume that there are three full light curves for a given star in each band with N points per light curve which we denote by $F_{R,i}$, $F_{G,i}$, and $F_{B,i}$ with $i = 1, N$ the individual data values in the red, green, and blue filters, respectively. We then divide each color light curve into 60 sub-light curves (one sub-light curve per day for IRa01 – 60 days). For each sub-light curve, we calculate the mean value MR , MG , and MB and normalize each sub-light curve by its mean value. We then compute new, normalized sub-light curves NF to be

$$NF_{R,G,B,i} = \frac{F_{R,G,B,i}}{M_{R,G,B}}, \quad (2)$$

for each filter band, and it is clear that all of these light curves have a mean of unity. This normalization is necessary otherwise the entire process would be dominated by the light curve of the strongest signal, which is usually the red light curve. As a side effect, CDA normalizes the depth of a possible transit in all filters using Eq. (2), so when the algorithm continues with its next steps, all transit events in each filter will have the same depth, thus CDA does not destroy real signals from the transits.

The normalized light curves now have the same mean, but their dispersions, differ. Our next goal is to identify the instrumental scatter caused, for example by jumps in each light curve and differentiate this instrumental scatter from statistical noise. To achieve this, CDA extracts five random packages of twenty adjacent points each from all colour bands and calculates the

standard deviation of each package per filter; the result should represent a good estimate of the correct light curve value at that time. If we use many packages, the probability of including jumps increases. The correct combination of packages and points is a function of the duration of the jumps, which is a random value, thus there is no ideal combination. We define as the mean standard deviation (MSD) the mean value of these five packages for each filter

$$MSD_{R,G,B} = \frac{1}{5} \sum_{j=1}^5 \frac{1}{20} \sqrt{\sum_{i=k_j}^{k_j+20} (NF_{R,G,B,i} - Mean_{\min})^2}, \quad (3)$$

where k_j denotes 5 different random data points of the light curve and $Mean_{\min}$ is the mean value of the flux of each package. In general, each filter has a different MSD value, which is compared with the standard deviation of each filter TSD defined to be

$$TSD_{R,G,B} = \frac{1}{N} \sqrt{\sum_{i=1}^N (NF_{R,G,B,i} - Mean_{\min})^2}. \quad (4)$$

Finally, the relative standard deviation of each filter RSD is computed and defined to be

$$RSD_{R,G,B} = \frac{TSD_{R,G,B}}{MSD_{R,G,B}}. \quad (5)$$

At the end of this process, we have three normalized light curves $NF_{R,i}$, $NF_{G,i}$, and $NF_{B,i}$, and three values for the relative standard deviation RSD_R , RSD_G , and RSD_B for each filter light curve, respectively. The algorithm CDA compares these three numbers and refers to the light curve with the minimum RSD as the base and the light curve with the maximum RSD as the target. To help illustrate the procedure, we continue with an example. We assume that the base is the blue light curve ($NF_{B,i}$) and the target is the red ($NF_{R,i}$) light curve. Using both the base and the target, CDA calculates a new mean light curve (AF_i); in our example, CDA computes

$$AF_i = \frac{1}{2}(NF_{R,i} + NF_{B,i}), \quad (6)$$

and then refers to AF_i as the light curve with the maximum RSD (in this example, it defines AF_i to be $NF_{R,i}$). According to assumptions 2 and 3, in the AF_i light curve any possible real signal remains but all the fake (jump) signals tend to be reduced, because jumps appear only at specific times in each filter. As a final result, we have in our example a red light curve detrended and two others (green and blue) untouched. If we try to run the algorithm again, we notice that the new values of RSD have changed because one light curve has changed. This means that every time we run the previous step of the algorithm, CDA removes a part of a fake signal (Fig. 3).

When these loops end, we renormalize the final light curve of the red channel to the raw mean value

$$NFR_{\text{final}} = NR \cdot NF_{R,i}, \quad (7)$$

the procedure is completed, and NFR_{final} is the final sub-light curve. The final step is to place all the 60 sub-light curves together. This produces the final light curve with which we are ready to search for exoplanets (Fig. 5). We employ of course many procedure loops, but if we use too many, CDA begins to destroy the light curve because it is obvious that after a certain number of loops a “saturation” is reached in the procedure. To avoid this effect, we do not use the same loop number for each

Table 2. CoRoT01270289.

Loop No.	RSD_R	RSD_G	RSD_B
#1	5.0485	0.9497	1.0658
#2	1.8632	0.9497	1.0658
#3	1.0665	0.9497	1.0658
#4	0.9688	0.9497	1.0658
#5	0.9688	0.9497	0.9868

Notes. This shows how RSD varies in each loop. In the first four loops, red filter is the target and green the base. By loop five, this situation has changed: blue is the target now and green is the base. These values refers to the RSD values of the full light curve of each filter.

light curve. We calculate the standard deviation in each light curve after each loop and CDA stops when the standard deviation begins to increases.

3.3. Simulations

To verify the functionality of CDA, we simulated CoRoT light curves as shown in Fig. 4. We simulated in particular a light curve in three filters (R , G , B), where jumps and trends appear at different times in each filter; a long-term trend was also included. In these light curves a transit pattern with period $P = 520$ time units and a relative depth $\Delta\text{Flux} = 0.01$ was included. The transits were masked by the high noise. As can be seen in Fig. 4, all jumps were removed and the resulting output light curve shows some regions with higher noise and some others with lower noise, but this does not affect the real signal. By applying transit detection algorithms (e.g. box least squares – BLS Kovács et al. 2002), the included transit pattern can also be detected.

4. Results

To illustrate the algorithm with real light curves, CDA is applied to four CoRoT light curves, i.e., CoRoT0102702789, CoRoT0102874481, CoRoT0102741994, and CoRoT0102729260.

4.1. The case of CoRoT0102702789

In Fig. 5, we show the raw red light curve, which includes a trend and jumps, and the final light curve after applying CDA with 5 loops. The light curve of CoRoT012702789 has one huge jump around JD ~ 2614 and many other smaller jumps. The RSD_R value of the raw light curve is 5.048 and of the final light curve is 0.95. Table 2 shows analytically the values of RSD from the total light curves, in these 10 loops of each filter. The green filter has the minimum RSD value, thus CDA uses it as a base. The red filter on the other hand has the maximum value and we call it the target, but in principal CDA defines different filters as either base or target in each loop. For this reason, in the first four loops the target is the red filter and base the green filter, then target changes to blue and green remains as base etc.; as already mentioned, the red light curve is the most common filter used to search for transits.

The example of CoRoT012702789 shows us how CDA works and how it removes jumps from a distorted light curve. As far as we can tell from our reconstructed light curve, there are no clear flares or transits in the light curve of CoRoT012702789. The critical question at this point is how CDA works if the raw light curve has real events such as transits.

Fig. 4. Simulated data. *R* – **a**), *G* – **b**) and *B* – **c**) color, respectively. Plot **d**) is the final light curve after CDA and the plot **e**) is the phase diagram of the transit after CDA & BLS.

4.2. The case of CoRoT0102874481

An even more extreme case is CoRoT0102874481, the light curve of which is affected by many jumps. The raw (red) light curve of CoRoT0102874481 is shown in Fig. 6. In the raw data, it is very difficult to distinguish real from instrumental events. As demonstrated in Fig. 6, CDA corrects all the jumps except for a real transit around JD ~ 2612 . The standard deviation before and after CDA is 2203.13 and 336.44 ADUs, respectively. Only a small jump from green and blue filters remains at the end of light curve.

Because this transit is the only transit in the light curve, we can determine neither the period nor the nature of the transiting object. Figure 7 shows that CDA does not reduce the depth of the transit, which is ~ 0.036 . According to the CoRoT team¹, the

host star’s spectral type is A0IV. Assuming the typical radius and mass of this star to be $R_s = 4.4R_\odot$ and $M_s = 2.8M_\odot$ and assuming the transiting object to be a true exoplanet, we determine the planet’s radius to be $R_p = 4.28R_J$ by using the relation between radius and transit depth (Seager & Mallén-Ornelas 2003)

$$R_p = R_s \cdot \sqrt{\Delta \text{Flux}}, \quad (8)$$

where R_s is the radius of the star and R_p is the radius of the planet. From Kepler’s 3rd law, the semi-major axis of the orbit is $a > 0.78$ AU, because the period is $P > 60$ days.

4.3. The case of CoRoT0102741994

The source CoRoT0102741994 appears to be a binary system. In this example, our main interest is not to check whether CDA

¹ <http://idoc-corot.ias.u-psud.fr>

Fig. 5. CoRoT012702789 red light curve and CDA results. Raw data **a)**, after 1 **b)**, 3 **c)**, and 5 **d)** loops respectively. All jumps are removed.

can remove the jump but to check how the algorithm preserves the eclipses and the flux of the light curve. Figure 8 shows how the algorithm converts the light curve. The light curve is affected by only a weak jump ($\Delta\text{Flux} \sim 1.25\%$) around JD ~ 2615 . The flux depth of the primary and secondary eclipse is 9% and 7%, respectively. In the top figure, is the light curve of the star before the applying CDA. The two eclipses are obvious, while the bottom figure shows the light curve after application of CDA. The jump is clearly removed completely. The depth of the primary and secondary eclipses are now 9.5% and 6.5%, respectively. As a general result, we can say that CDA does not remove the real signal but corrects the jumps.

4.4. The case of CoRoT0102729260

We find that the light curve of CoRoT0102729260 is a combination of trends and strong and weak jumps. The raw light curve of CoRoT0102729260 does not show any transits. We note that a transit detection algorithm such as BLS does not detect any transit event in this light curve (Fig. 10, top panel). However, after applying CDA to remove all jumps, we again implement BLS on the final light curve and a possible transit appears (Fig. 9, bottom panel).

This transit is only detectable after applying CDA, but not in the raw data. Our analysis of the phased light curve infers period of $P = 1.6986$ days. The photometry by the CoRoT team provides some information about the parameter of the host star, which appears to be a main sequence star (G5V) of apparent brightness $m_V = 14.772$ mags. Assuming the spectral type to be correct, we can estimate the radius of the star to be $R_s \sim 0.91R_\odot$.

Table 3. Physical parameters of CoRoT0102729260.

Color Index	0.752
Star Radius R_s	$0.91R_\odot$
Period	1.6986 days
Planet Radius R_p	$6.27R_E$
Depth (Flux)	0.004

With a transit depth of $\Delta\text{Flux} = 0.004$, we deduce a planetary radius of $R_p = 6.27R_E$ applying Eq. (8). Figure 11 shows the phase-folded light curves. Table 3 also provides additional information about the system.

5. Conclusions

We have introduced and presented a method dubbed CDA that removes instrumental artefacts from CoRoT data and demonstrated its usefulness in some practical applications. We emphasize that the CDA algorithm can be used to prepare CoRoT data for any transit detection but should not be used for transit analysis because it can remove real signal. This is not of course a problem for the detection inasmuch as instrumental jumps affect far more the light curve. From our study of 1030 light curves in the first CoRoT field (IRao01), we found that only very few light curves have no instrumentally caused features and remain as they are, while the vast majority of light curves are appreciably improved. We have presented some examples that show

Fig. 6. CoRoT012874481 – red filter. *Top*: raw light curve. *Bottom*: the same light curve after CDA. Jumps are removed and a clear transit appears. The subframe is a zoom-in plot.

Fig. 7. CoRoT012874481 residuals before minus after. CDA. The signal from the real transit is not reduced by the algorithm.

how the algorithm affects the light curves. Our main conclusion is that instrumental jumps substantially affect the CoRoT light curves, making a transit detection in fainter stars impossible.

To illustrate how the algorithm affects the data of the full sample, we calculated the median absolute deviation (MAD) before and after applying CDA. Figure 12 shows the differences between the two procedures.

We prove our case with the example of CoRoT0102729260, a possible candidate exoplanet that is detected only after applying CDA on the raw data.

Fig. 8. CoRoT012741994 – red filter. *Up*: raw data. We have removed all the “bad points”. The light curve contains one jump around JD ~ 2615 and a trend. *Down*: the same light curve after CDA. The jump intensity is reduced. CDA does not affect the transit depth.

Fig. 9. CoRoT0102729260 – red filter. *Up*: raw data before CDA. *Down*: final light curve after CDA. The algorithm succeeded to remove all the jumps and trends and improve the light curve enough to detect the “concealed” transit.

Fig. 10. CoRoT0102729260 – red filter. *Up*: periodogramm of the raw light curve before CDA without any obvious signal. *Down*: same plot after CDA. A clear periodic signal ($P \sim 1.698$) is detected.

Fig. 11. CoRoT0102729260. *Top*: a phase-folded light curve before CDA. *Bottom*: a phase-folded light curve after CDA.

Fig. 12. Median absolute deviation (MAD) before and after CDA using a 1030 light curve sample.

Acknowledgements. D.M. was supported in the framework of the DFG-funded Research Training Group “Extrasolar Planets and their Host Stars” (DFG 1351/1).

References

- Aigrain, S., Pont, F., Fressin, F., et al. 2009, A&A, 506, 425
- Baglin, A., Vauclair, G., & The COROT Team. 2000, JApA, 21, 319
- Kovács, G., Zucker, S., & Mazeh, T. 2002, A&A, 391, 369
- Mazeh, T., Guterman, P., Aigrain, S., et al. 2009, A&A, 506, 431
- Pinheiro da Silva, L., Rolland, G., Lapeyrere, V., & Auvergne, M. 2008, MNRAS, 384, 1337
- Seager, S., & Mallén-Ornelas, G. 2003, ApJ, 585, 1038