Chapter 2

Transit signal reconstruction

This chapter focuses on improving the planet parameters by improving the accuracy of the transit signal.The motivation for a new stellar variability filter is presented in Section 2.1. The Iterative Reconstruction Filter, the new post-detection stellar variability filter developed during this PhD thesis, is presented in Section 2.4, and tested over simulated data. The performance of the filter is then discussed in Section 2.5.

The work presented in this chapter was published in Alapini & Aigrain (2009).

2.1   Motivation

As smaller and lower-mass planets become increasingly detectable, thanks to space-based transit searches and improvements in ground-based radial velocity instruments, the uncertainties arising from the transit and radial velocity fits are expected to become more important. A specific problem arises when the transits become comparable in depth with the amplitude of the intrinsic brightness fluctuations of the host star. The amplitude of these variations can be several orders of magnitude greater than the transit signal, particularly for terrestrial planets and/or active stars, and they can occur on timescales significantly shorter than the orbital period of the planet (Fig. 2.1, black curve). Stellar variability can thus hinder the detection of planetary transits (Aigrain et al., 2004). A number of 'pre-detection' filters have been developed to tackle this problem.

Pre-detection filters aim to remove stellar variability in light curves to improve the detectability of transits, without any prior knowledge of the transit signal except for the fact that stellar variability typically occurs on longer time scales (hours to days) than the transit signal (minutes to hours). All of the techniques tested in the first CoRoT blind test (Moutou et al., 2005), which range from simple Fourier-domain low-pass filters to slightly more sophisticated implementations involving simultaneous fitting of hundreds of low-frequency sinusoids, or time-domain nonlinear iterative filtering (Aigrain & Irwin, 2004), exploit this difference. These filters proved effective in removing stellar variability

to facilitate the detection of transits but, as pointed out in Moutou et al. (2005) and Bonomo & Lanza (2008), they deform the shape of the transits.

The performance of several of these filters in terms of transit detection was evaluated in the context of first CoRoT blind test, a hare-and-hounds exercise involving 1000 simulated CoRoT light curves containing various transit-like signals, stellar variability and instrumental noise. This test showed that the most successful filters recover a detection threshold close to that obtained in the presence of instrumental noise only, except for a few cases involving the most active and rapidly rotating stars simulated.

However, these filters also have the property of modifying the shape of the transit signal (Moutou et al., 2005; Bonomo & Lanza, 2008), and would destroy any signal at the period of the transit occuring on longer timescales than a few hours.

After introducing, in Sect. 2.2, the simulated data set used for test purposes throughout this chapter, in Sect. 2.3 the effect on the transit signal of a benchmark pre-detection filter, the nonlinear iterative filter of Aigrain & Irwin (2004), is quantified. Described in Sect. 2.4 is the iterative reconstruction filter designed in this thesis, evaluating its effect on the transit signal. The IRF is a post-detection filter that uses the knowledge of the transit period to reconstruct signals at that period while filtering out signals at other timescales. The impact of these two filters on the accuracy of planet parameter measurements are compared in Sect 2.5, and the main results are summarised in Sect 2.6.

2.2   Data set

2.2.1   BT2 light curves

The starting dataset used in this study is a sample of 236 simulated CoRoT light curves taken from the second CoRoT blind test (hereafter BT2; Moutou et al. 2007), which was carried out to compare methods for discriminating between planetary transits and grazing or diluted stellar eclipses. Twenty six (26) of these light curves have planetary transits and 210 have eclipsing binary signals. We selected for this study of transit deformation only the BT2 light curves with planetary transits.

The production of the light curves followed roughly the same steps as that for the first CoRoT blind test (BT1), described in detail in Moutou et al. (2005), incorporating transits simulated with the Universal Transit Modeler (UTM[1], Deeg 2009), instrumental noise simulated using the CoRoT instrument model (Auvergne et al., 2003), and stellar variability curve simulated using a combination of the methods of Lanza et al. (2004) and Aigrain et al. (2004). The stellar variability modelled in the BT2 light curve is pessimistically strong both in terms of amplitude and times scale. The CoRoT data show that most stars are not quite so variable (Aigrain et al., 2009). An updated version of the CoRoT instrument model was used in the BT2, incorporating more realistic satellite jitter

---

[1]See `http://www.iac.es/galeria/hdeeg/`.

and enabling the production of 3-colour light curves, though the 3 bandpasses were summed in the present study to construct a 'white' light curve. The two approaches, used in the BT1 to model stellar variability, were merged in the BT2 using the scaled spot model of Lanza et al. (2004) to simulate rotational modulation of active regions and the stochastic model of Aigrain et al. (2004) to simulate granulation. The simulated transits correspond to planet radii ranging from 0.2 to $1.1\,R_{\mathrm{Jup}}$, orbital periods from 2.6 to 11.0 d, and impact parameters from 0.25 to 0.88.

As in the BT1, the flux in each aperture was modelled as arising from two stars, only one of which contained a transit-like signal. This is to reflect the fact that there is almost always one or more background star in the CoRoT aperture. This has the effect of diluting the transit signal, and to account for it we subtract from each BT2 light curve a constant corresponding to the fraction of the median flux contributed by the star which is not eclipsed (see Tab. 2.1 for contaminant fluxes (percentages of total flux) corrected from each BT2 light curve studied).

An example of a light curve with transit from the BT2 is shown in Fig. 2.1. The full set of light curves is shown in Fig. 2.8.

Table 2.1: Table of flux percentages coming from a contaminant star, for each of the BT2 light curve studied. Each light curve was corrected from the contaminant flux, before deriving transit and planet parameters. The fraction of flux coming from a contaminant star in each colour channel (CoRoT red, green and blue) was given in the parameter file used to build the BT2 light curves. For each light curve, the total contaminant flux was computed as the median of the sum of the contaminant fluxes in each colour channel, normalised by the median of the total flux.

BT2 LC n^o	contaminant flux (%)	BT2 LC n^o	contaminant flux (%)
105	0.2	177	0.6
110	0.1	186	0.3
126	2.2	192	0.8
131	90.6	193	13.1
133	0.2	196	0.9
135	0.1	200	3.3
145	2.3	208	1.8
152	0.3	220	1.9
154	1.9	223	77.4
162	0.1	225	0.6
165	91.1	233	0.6
169	0.5	236	1.4

2.2.2 Reference light curve sample

As the data is simulated, each component of the signal is known and can be studied individually. Thus two sets of reference light curves were constructed, using only the transit signal (no noise, no stellar variability) and the transit signal with instrumental noise only (no variability). We use the first set to evaluate the reference values of
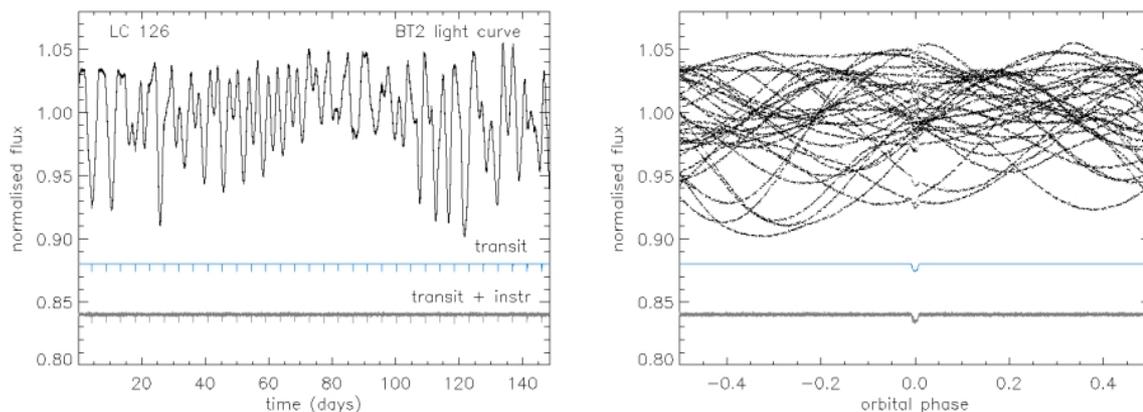
Figure 2.1: BT2 light curve (black), in which transit signal of a Saturn-like planet orbiting a particularly active Sun-like host star with an orbital period of 4.576 d. Transit signal only (blue) and with instrumental noise (grey) plotted underneath for comparison. Left: unfolded light curves. Right: phase-folded versions.

the parameters derived from transit fits. These could have simply been deduced from the input parameters given to the transit modelling software UTM when simulating the light curve. However, there can be differences between those and the parameters recovered from the transit fit due to the fitting process, rather than to the noise, and we wish to keep those effects, which are not specifically of interest here, separate from the effects of the stellar and instrumental noise. The second set was used to provide a benchmark for how well one can measure the parameters of interest in the presence of instrumental (white) noise, i.e. if the stellar variability was removed perfectly. These reference sets are shown in blue and grey respectively in Fig. 2.1.

After visual analysis of our two reference sets of light curves, we discarded two of the 26 light curves, where the transits were so small as to be undetectable even in the light curves with no stellar variability, as such cases would not realistically reach the post-detection stage.

2.3 Quantifying transit deformation with the Non-linear Iterative Filter

In this section, we quantify the impact of the deformation caused by the nonlinear iterative filter (NIF) of Aigrain & Irwin (2004) on the derived planet parameters. The NIF performance as a pre-detection filter was recently compared to a range of other published methods (Bonomo & Lanza, 2008), and it emerged as the method of choice among those compared, which makes it a suitable benchmark for the present work.

2.3.1 Definition of the NIF

The NIF has been extensively used at a pre-detection / transit search level. Here we briefly describe the main steps of the NIF, we refer the reader to Aigrain & Irwin (2004) for further details.

The NIF separates stellar variability from the transit signal in the time domain, using an iterative procedure with the following steps:

1. apply a short base-line (here we use 7 data-points, $\sim$1 hour) moving median filter to smooth out the white noise and reduce the sharpness of any high-frequency features in the data;

2. apply a longer base-line moving median filter (here we use 24 hours, NIF trade of point to remove stellar variability while keeping away from the transit time domain in these light curves) to the output of the step (1), followed by a shorter base-line (here we use 2 data-points, $\sim$17 minutes) boxcar filter (moving average);

3. subtract the output of step (2) from that of step (1) and evaluate the scatter of the residuals as $\sigma = 1.48 \times \mathrm{MAD}$; [2]

4. flag all outliers differing by more than $n\sigma$ from the continuum;

5. return to step (2) and repeat the process, interpolating over any flagged data points before estimating the continuum and excluding them when estimating the scatter of the residuals, until convergence is reached (typically less than 3 iterations);

6. subtract the final continuum from the original light curve.

As the procedure converges, more and more of the in-transit points become flagged at step (4), so that the effect of the transits on the final continuum estimate is minimal. However, the choice of long base-line for the moving median filter in step (2) and of n in step (4) must reflect a trade-off between appropriately following the stellar variations and incorporating too much of the transit signal when evaluating the continuum. This trade-off results in some of the transit signal been unavoidably filtered along with the variability. For the value of n in step (4), one would normally use $n = 3$ to flag more in-transit points. In the case of the BT2, some light curves contain very strong and rapid variability. Thus, using a low n would clip not only in-transit points but also out-of-transit points where the variability is too rapid to be well modelled by the continuum estimate (e.g. Fig. 2.2 left, green curve compared to black one). Hence, we used a large n

---

[2]The MAD is the median of the absolute deviation from the median of the points, in other words it is the median of the absolute value of the residuals from the median. 1.48*MAD is the equivalent of the standard deviation when using the median of the points rather than the mean. In this thesis, the use of the median and σ=1.48*MAD is preferred to the mean and the standard deviation, as the first combination is more resilient to outliers in the data due to the way it is calculated.

(150) in this study, which effectively means no points are clipped and convergence occurs at the first iteration.

2.3.2   NIF quantitative impact on transit parameters

We applied the NIF to our sample of 24 BT2 light curves. The post-NIF light curves are shown in green on Figs. 2.2 and 2.8. Clear variability residuals are visible in the unfolded post-NIF curves, corresponding to sections of the light curve where the variability is too rapid to be filtered adequately. The phase-folded light curves also show that the shape of the transits is affected by the filter. In practical terms, the transit appears both shorter and shallower than before filtering.
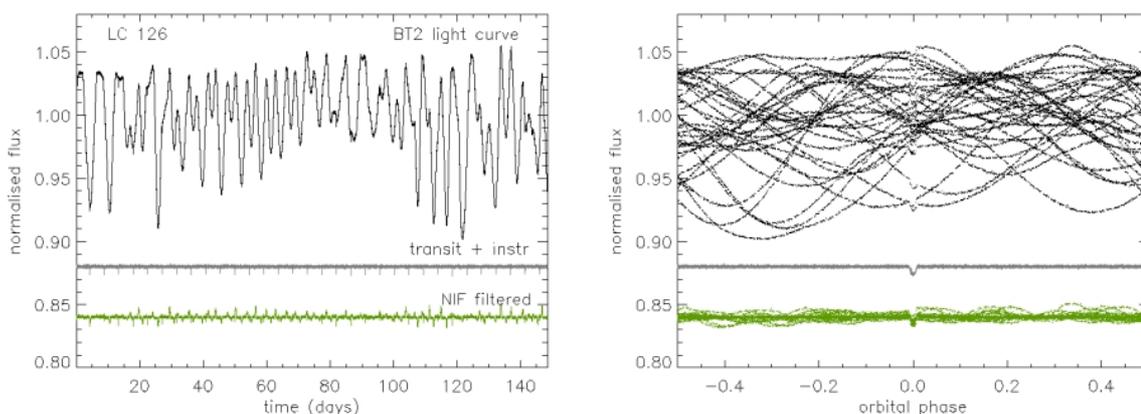

Figure 2.2: As in Figure 2.1 but the NIF-filtered light curve is now shown in green. The right panel shows that the NIF filtering reduces well the amplitude of the variability (green compared to black). The left panel shows that the NIF difficultly filters out fast stellar variability (spikes in green curve).

We then folded all light curves at the period of the injected transits and performed least-squares fits of trapezoidal models to the results to estimate the basic transit parameters: depth δ, internal and external duration d_i and d_e (respectively excluding and including ingress and egress), and the phase ϕ. The light curves were normalised such that the out-of-eclipse level is always 1. The same folding and trapeze fitting procedure was applied to the two reference sets described in Section 2.2.2.

In 4 of the BT2 light curves (Fig. 2.9), the stellar variability was so strong that, after applying the NIF, the phase-folded transits were barely detectable, and meaningful fits to these transits impossible. These 4 light curves were excluded from the comparison sample between the reference and filtered versions of the light curves.

We list the measured values of the transit parameters (δ, d_i, d_e) of direct relevance to the determination of planet parameters for all 20 light curves in Fig. 2.8 (transit parameters in Tab. 2.2). We also show, in Fig. 2.3, cumulative histograms of the relative error $\sigma(\theta) = |\theta - \theta_0|/\theta_0$, where θ is the parameter of interest and the subscript 0 refers to the
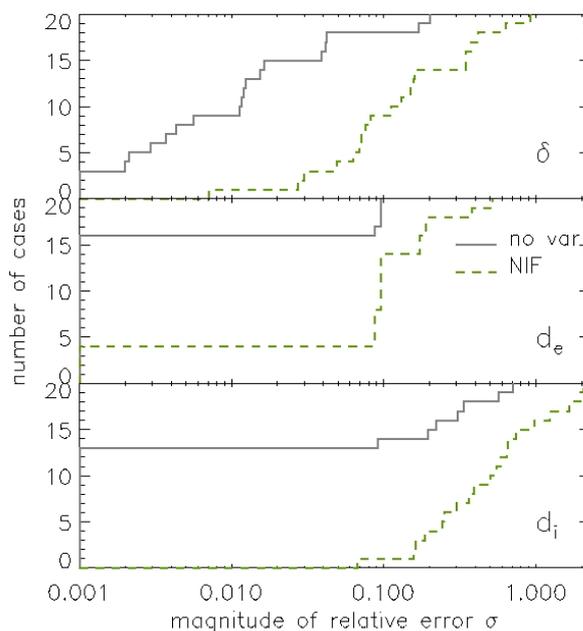
Figure 2.3: Cumulative histograms of the relative error σ (see text for exact definition) on the transit parameters measured from trapezoidal fits to the light curves with no variability (grey), and to NIF-filtered light curves (green). Upper panel: transit depth δ; middle panel: external duration d_e (total transit duration); lower panel: internal duration d_i (duration in full-transit). $\sigma > 1$ when a parameter is mis-estimated by more than its true value.

value measured from the reference light curve with transits only (no noise), contrasting the NIF case (green dashed line) to the case with no variability (black solid line). The median relative errors obtained with the NIF over our sample are $\sigma_{\mathrm{NIF}}(\delta) = 12\%$, $\sigma_{\mathrm{NIF}}(d_e) = 10\%$ and $\sigma_{\mathrm{NIF}}(d_i) = 52\%$, indicating that the planet parameters would be seriously affected if derived from NIF-filtered light curves. We note that the internal duration d_i tends to be systematically underestimated even for the reference set of light curves with no stellar variability. This bias is due to the white noise in the data smoothening the edges of the transit and making it appear more grazing, i.e with smaller orbital inclination so a shorter transit internal duration d_i.

We therefore set out to develop a new post-detection filter: an alternative algorithm, hereafter referred to as 'reconstruction filter', designed to remove variability at other periods than that of the transit and preserve the transit signal, once the transit period has been determined.

2.4 A new stellar variability filter: the Iterative Reconstruction Filter

In an attempt to avoid the undesirable effects of the NIF on the transit shape, an iterative reconstruction filter (IRF, Alapini & Aigrain 2009) was develop to remove the stellar variability post transit detection whilst altering the transit signal as little as possible using the knowledge on the planet orbital period.

2.4.1 Definition of the IRF

The IRF is an iterative approximation of the full signal at the period of the transit. It uses the NIF to simultaneously estimate the continuum variation (i.e., stellar variability). Let $\{Y(i)\}$ (where $i = 1, \ldots, N$, N being the number of data points in the light curve) represent the observed light curve (which is assumed to be normalised), $\{A(i)\}$ the detrended light curve and $\{F(i)\}$ the signal to be filtered out. We give the main steps of the IRF below:

1. Select an initial estimate for $\{F(i)\}$. $\{F(i)\} = 1$ is adopted as the initial estimate of the stellar variability, instead of a closer estimate using a median filter for instance. This is to avoid removing some transit signal before its first evaluation by the IRF, as this signal would never have been evaluated by the IRF as part of the transit signal and the IRF will not know that it has to be recovered.

2. Compute a corrected time-series $\hat{Y}(i) \equiv Y(i)/F(i)$.

3. Estimate $\{\hat{A}(i)\}$ by folding $\{\hat{Y}(i)\}$ at the transit period and boxcar averaging it in intervals of a fixed duration in phase units (binning is used to reduce high frequency noise). For the BT2 light curves, a duration of 0.09% of the phase was found to be suitable (this value was selected by trial and error, longer duration implying lower noise in the estimate of $\{\hat{A}(i)\}$ but more distortion of the transit signal).

4. Unfold $\{\hat{A}(i)\}$ to obtain $\{A(i)\}$. Compute a new estimate of $\{F(i)\}$ by applying the NIF (described in Section 2.3.1) to $\{Y(i)/A(i)\}$. The baseline for the median filter used in the NIF at this step can be chosen on a case-by-case basis, and can be significantly shorter than in the pre-detection case, because it is applied to a light curve from which most of the transit signal has been removed. In the present study, we adopt a baseline of 12 hours, the rest of the NIF parameters being the same as in Section 2.3.1.

5. Return to step (2) with the new estimate of $\{F(i)\}$, and iterate until the condition $|\mathcal{D}_{j-1} - \mathcal{D}_j| < 10^{-4}$ is satisfied for two consecutive iterations, where j is the iteration

number (initialisation at $j = 0$), and

$$\mathcal{D}_j = \frac{\sum_{i=1}^{N} \left[Y(i)/A_j(i) - F_j(i)\right]^2}{N - 1}.$$

In the case of the BT2 light curves, the convergence was reached after 4 iterations (i.e D_j was calculated up to $j = 6$).

The final detrended light curve is given by $\{Y(i)/F(i)\}$, where $\{F(i)\}$ is the last (presumably best) estimate of the stellar variability. The steps of the IRF are sketched out in Figure 2.4.

$$\{F_0(i)\} = 1$$

1st iter. j=1

j_5

Start: $\{F_{j-1}(i)\}$ ← $\{F_j(i)\}$

NIF continuum

$\{Y(i)/F_{j-1}(i)\}$ j_1

$\{Y(i)\}$

j_4

phase-fold,

j_2

j_3 $\{Y(i)/A_j(i)\}$

$\{A_j(i)\}$

smoothe, unfold

$$D_j = \frac{\sum_i \left[Y(i)/A_j(i) - F_j(i) \right]^2}{N-1}$$

for j > 3

$$\left| D_{j-1} - D_j \right| \quad \left| D_{j-2} - D_{j-1} \right| \quad \left| D_{j-3} - D_{j-2} \right|$$

If all < 10^{-4}

If not, iterate
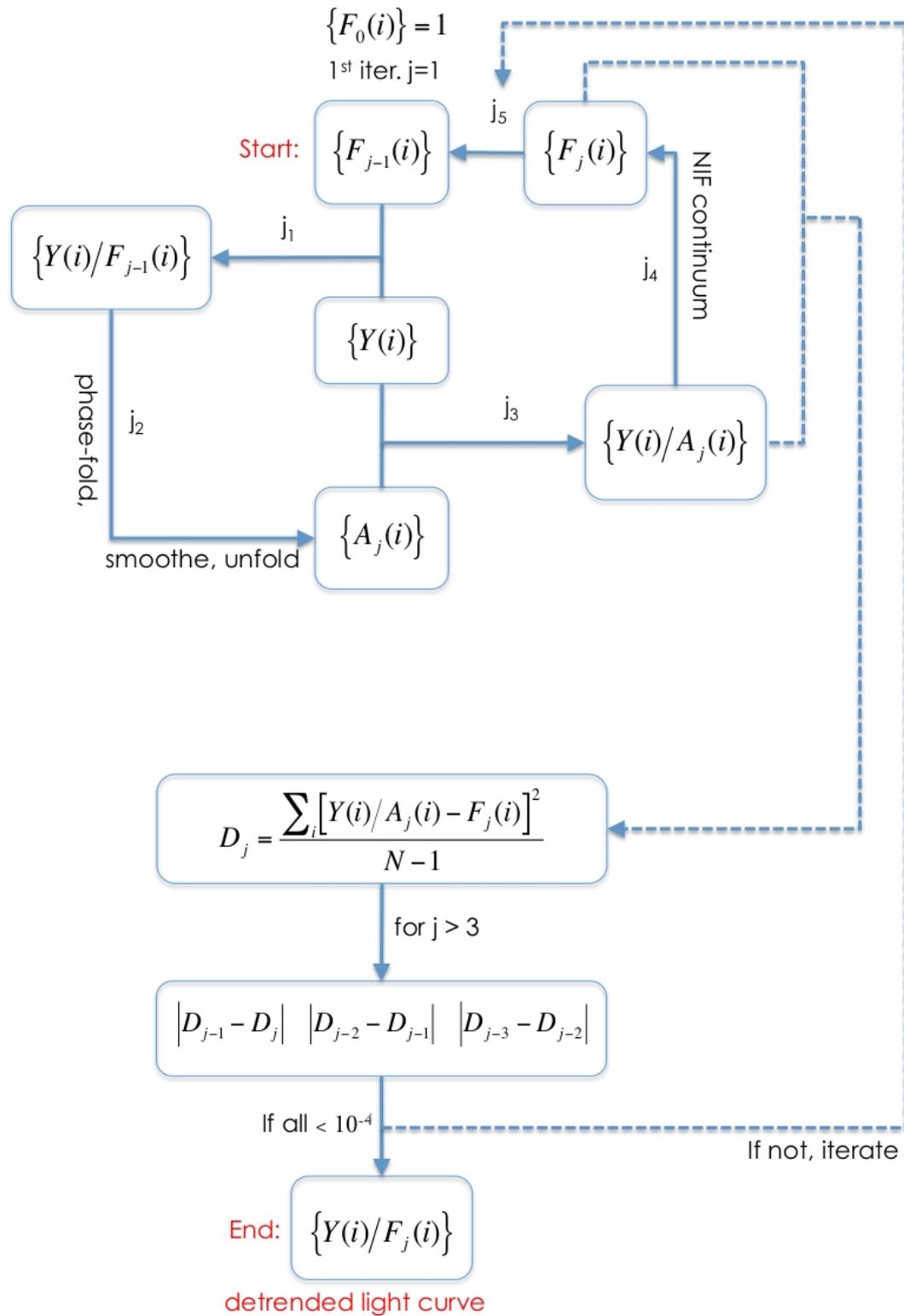
End: $\{Y(i)/F_j(i)\}$

detrended light curve

Figure 2.4: Flow chart of the IRF. $\{Y(i)\}$ represents the observed light curve, $\{A(i)\}$ the detrended light curve and $\{F(i)\}$ the signal to be filtered out. i is the data points index (1 to N) and j is the iteration index.

2.4.2   Comparison with the Trend Filtering Algorithm (TFA)

This algorithm is in some ways analogous to the TFA (Kovács et al., 2005) in post-detection mode. For clarity, we briefly list the main similarities and differences between the two algorithms.

- The TFA is designed to remove systematic trends which are common to large numbers of light curves in the transit surveys, rather than stellar variability which is individual to each object. Both algorithms work by decomposing each light curve into three components: the signal of interest $\{A(i)\}$ (the transits), the signal to be filtered out $\{F(i)\}$ (the systematics in the case of the TFA and the stellar variability in the case of the IRF), and the residuals. In the TFA, the signal to filter out (systematics) is modeled as a linear combination of a number of template light curves selected from the survey sample. In the IRF, the signal to filter out (stellar variability) is taken as the continuum of the light curve estimated with the NIF. In this analogy, the NIF would be equivalent to TFA in pre-transit-detection mode. When used in reconstruction mode (post-detection), both methods make use of the knowledge of the transit period to iteratively improve the evaluation of the transit signal and of the signal to be filtered out (which is assumed not to be periodic).

- Whereas $\{F(i)\}$ and $\{A(i)\}$ are treated additively in the TFA, they are treated multiplicatively here since the signal to be filtered out is intrinsic to the star, and the planet blocks out a certain fraction of the flux emitted by the star. This results in a different initialisation of $\{F(i)\}$. In Kovács et al. (2005), the first estimate of $\{F(i)\}$ is obtained from the pre-detection implementation of the TFA. In the IRF, it would be counter-productive to use the NIF-filtered light curve as the initial estimate of $\{F(i)\}$, since we have shown that the NIF affects the transit signal we are trying to reconstruct (see Section 2.3.2), so the initial estimate of $\{F(i)\}$ is taken to be constant at 1.

- Finally, the IRF treats high frequency effects by smoothing the phase-folded signal, while the TFA treats them by filtering out common outlier values.

2.4.3   Performance of the IRF on the BT2 transits

The IRF was applied to the 24 BT2 light curves described in Section 2.2, with the filtering parameters described in Section 2.4.1. The red curves in Fig. 2.5 and Figs. 2.8, show the light curves after applying the IRF.

As shown in Fig. 2.5, the IRF preserves any signal at the period of the transit. If the stellar variability contains power at this period, it is also preserved, inducing a flux gradient around the transit which must be removed before fitting the transits. This correction setting the out-of-transit level constant at 1, was done by fitting a 2nd order polynomial

fit – the lowest-order found to give satisfactory results – to the data about the phase-folded transit. The data used for the polynomial fit are two segments, each lasting 0.1 in phase, and offset by 0.15 in phase from the center of the transit on either side. This is a significant improvement over the common practice of performing a local polynomial fit to the vicinity of each transit, since the latter option has many more free parameters (one set of free polynomial parameters per transit, rather than one for the entire light curve). Fig 2.5 right panel gives an example of polynomial fit of the continuum about the transit (black segment superimposed to red curve) and of the resulting re-normalised transit (orange).
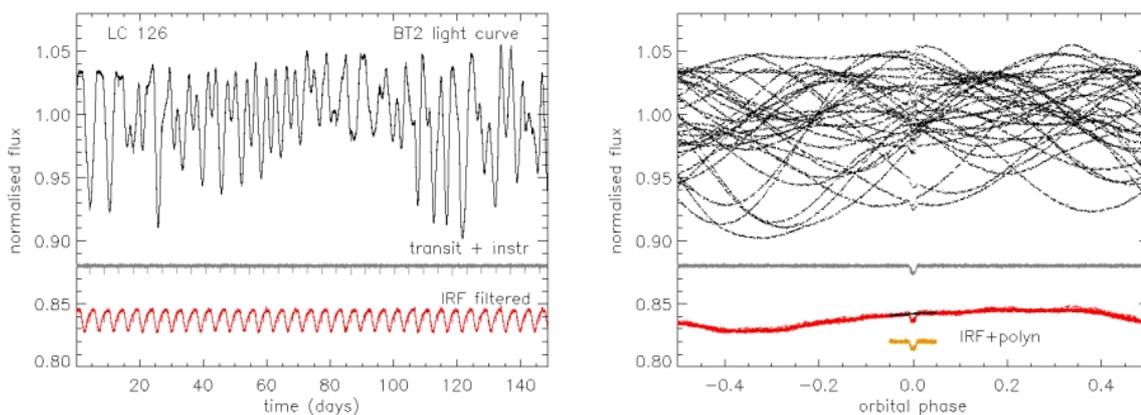

Figure 2.5: IRF-filtered light curve is red. Black and grey same as Fig. 2.1, plotted for comparison; the black curve is the starting point, the grey curve is the level of clean filtering we want to reach. These graphs show that the IRF conserves all variations at the period of the transit. The right panel shows that the IRF recovers the transit signal better than the NIF in Fig. 2.2. The IRF-filtered transit can be corrected from the non-constant local continuum by dividing it by a 2nd order polynomial fit about the transit (black line superimposed to the phase-folded IRF-filtered transit signal). The locally re-normalised transit is shown in orange.

The transit parameters were then estimated from a trapezoidal fit to the resulting phase-folded transit, in the same way as described in Section 2.3.1 for the NIF case. The results are listed in Tab. 2.2 and shown as the red dash-dot curves in Fig. 2.6. For the 20 BT2 transit light curves which were also used to evaluate the performance of the NIF, the IRF gives median relative errors of $\sigma_{\mathrm{IRF}}(\delta) = 3\%$, $\sigma_{\mathrm{IRF}}(d_{\mathrm{e}}) < 10^{-4}\%$ and $\sigma_{\mathrm{IRF}}(d_{\mathrm{i}}) = 42\%$, representing a significant improvement over the NIF case. Additionally after applying the IRF, in 2 of the 4 cases which are not included in the comparison sample as the transits were barely detectable after applying the NIF (Fig. 2.9), the transits are now clearly detectable and yield meaningful fits. In the two other light curves, the IRF-filtering gives a light curve closer to the reference version than the NIF, but the transits – already hidden in the instrumental noise in the reference set – stay barely detectable even in the IRF-filtered version.

Looking at Fig. 2.6, we see that while a relative error on the transit depth in excess of
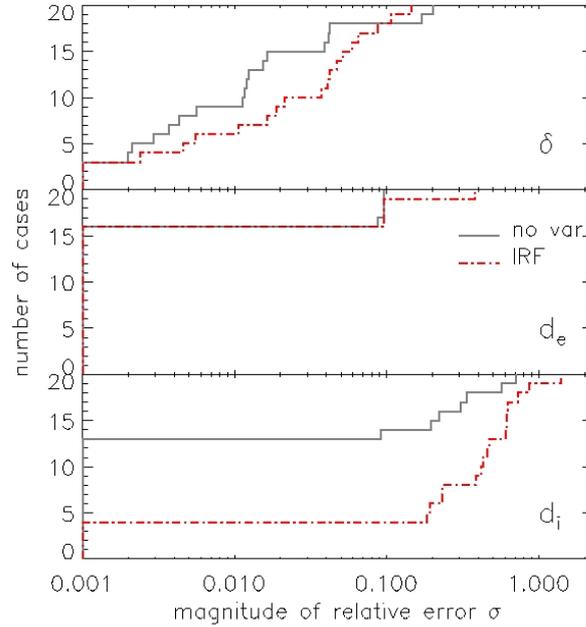
Figure 2.6: Same legend as Fig. 2.3 but for the IRF.

10% (essentially precluding any meaningful constraints on the planet structure) occurs in 60% of the cases studied with the NIF, it occurs in only 5% of the cases with the IRF. Similarly, the NIF yielded $\sigma(\delta) < 3\%$ (potentially allowing discrimination between different kinds of evolutionary models as well as a reliable basic structure determination) in only 15% of the cases, but the IRF did so in 50% of the cases.

It is also clear that the external transit duration is recovered near-optimally in the light curves treated with the IRF, with $\sigma(d_e) < 0.1\%$ in 80% of the cases and $\sigma(d_e) < 10\%$ in 95% of the cases, compared to a significantly decreased performance with the NIF. However, although the IRF also systematically improves the determination of the internal transit duration compared to the NIF, this improvement is much less significant, and the relative errors remain large (more than 10% for 80% of the cases studied). This implies that the IRF would probably not significantly increase the number of cases where both internal (2nd to 3rd contact) and external (1st to 4th contact) duration can be determined precisely enough to break the degeneracy between system scale and inclination, and thus to constrain the stellar density in a model-independent fashion.

2.5 Discussion on the IRF performance

2.5.1 Star-planet parameters

Although the basic trapezoidal fits performed in the previous two sections provide a quick estimate of the degree of deformation of the transit signal due to the variabil-

ity filtering process, one would in practice perform a full transit fit based on a physical model of the star-planet system. Mandel & Agol (2002) provided an analytical formulation which has become very widely used for such purposes, and was also used to generate the transits injected in the BT2 light curves.

We used the quadratic limb darkening prescription of Mandel & Agol (2002) to fit transit models to the 20 BT2 transit light curves where the transits were clearly detectable with both filters. We also performed these fits on both reference sets described in Section 2.2.2, as well as on the BT2 light curves themselves after applying the NIF on one hand, and the IRF – followed by a polynomial fit to the region around the transit (as described in Section 2.4.3) – on the other hand. The best transit fits were derived using MPFIT, an IDL implementation of the Levenberg-Marquardt algorithm[3]. The parameters of the model used are the transit epoch T_0, the period P, the system scale $a/R_{\rm s}$ (where a is the semi-major axis), the star-to-planet radius ratio $R_{\rm p}/R_{\rm s}$, the orbital inclination i (or impact parameter $b \equiv a \cos i/R_{\rm s}$), and the quadratic limb-darkening coefficients u_a and u_b. In this study, we fixed the period and limb-darkening coefficients at the values used to build the light curves[4]. The initial epoch was taken directly from the trapezoidal fits. The initial value for $a/R_{\rm s}$ was derived from the period using Kepler's $3^{\rm rd}$ law, assuming $R_{\rm s} = R_\odot$ and $M_{\rm s} = M_\odot$. In order to ensure convergence in both grazing and full transits we selected, after some trial and error, an initial inclination corresponding to an impact parameter $b = 0.7$. We assumed zero eccentricity in all cases (all the transit light curves in our sample were simulated for circular orbits).

The results of the transit fits are listed in Table 2.3, while the fits themselves are shown in Figures 2.8 and 2.9. They are also compared in cumulative histogram form in Fig. 2.7. Instead of the relative error σ, we show the absolute error $\xi = |\theta - \theta_0| \equiv \sigma \times \theta_0$ with respect to the no noise case (subscript 0), for θ the key planet parameters $R_{\rm p}/R_{\rm s}$, $a/R_{\rm s}$ and b.

The IRF provides an overall improvement over the NIF in all three parameters, reducing the median of $\xi(R_{\rm p}/R_{\rm s})$ from 0.007 to 0.003, $\xi(a/R_{\rm s})$ from 1.7 to 1.0, and $\xi(b)$ from 0.07 to 0.04 for b. For comparison, the corresponding median values for the case with no variability are 0.003, 1.4 and 0.07 respectively. However, the situation is not as defined as when viewed in terms of transit parameters: there are a few cases where the NIF gives a better match with the parameters obtained from the noise-free light curves, and even cases where the largest error occurs in the light curves containing instrumental noise only. In an attempt to understand the reason for this, we examined all the light curves one by one (Figures 2.8 and 2.9). The light curves separate fairly

---

[3]MPFIT is kindly provided by C. Markwart on http://cow.physics.wisc.edu/∼craigm/idl/fitting.html

[4]Visual examination of the phase-folded light curves revealed that the folding was not perfect even in the no noise case, suggesting that the period values used may have been slightly inaccurate. We attempted to refine the periods but did not succeed. It seems that the observation dates in the light curve files themselves, rather than the periods, suffer from a small rounding error. It is not possible to remedy this problem without re-generating the entire light curve set, but it is not expected to affect the results strongly, and any effect would be common to all versions of a given light curve.

naturally into three broad classes:

1. cases where the IRF performed better than the NIF (transit shape and derived planet parameters closer to the shape and parameter obtained in the absence of stellar variability): light curves 126, 162, 169, 196, 200, and 223. These are cases where the original light curves contain large amplitude, short timescale stellar variability (active and rapidly rotating stars).

2. cases where the NIF performance was already satisfactory, and the IRF gives results similar to the NIF: light curves 145, 152, 186, 193, 208, 225, and 233.

3. cases where, while the transit reconstructed with the IRF appears closer to the original than the transit in the NIF-treated curve, the fitted parameters are not significantly improved or worsened: light curves 131, 133, 135, 154, 177, 192, 220. These are typically low signal-to-instrumental noise transits, where it becomes difficult to break the degeneracy between impact parameter and system scale. The radius ratio is typically less affected, except in the highest impact parameter cases (grazing transits).

Thus, we can see that where the limiting factor was stellar variability, the IRF is very successful in improving the errors on the planet parameters. As might be expected, the improvement is minor or non-existent where the limiting factor was the signal-to-white noise or the grazing nature of the transits.
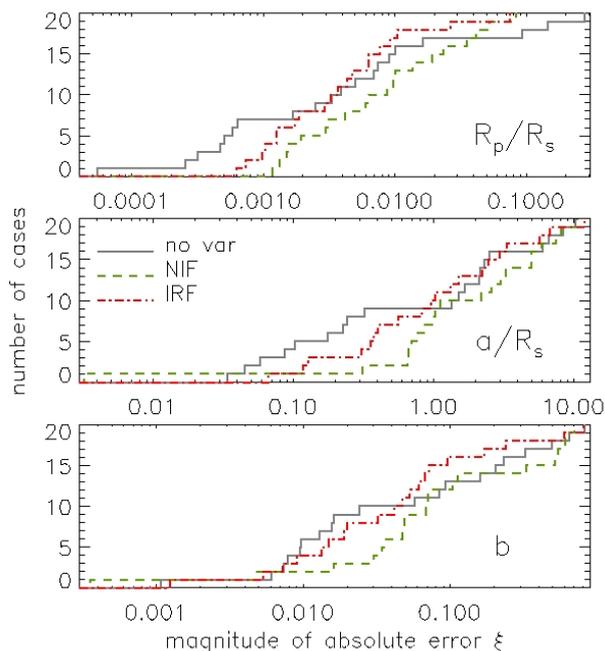
Figure 2.7: Cumulative histograms of the absolute error ξ on the planet parameters derived from the light curves with no variability (grey), NIF-filtered (green) and IRF-filtered (red). The planet to star radius ratio $R_\mathrm{p}/R_\star$ is plotted in the upper panel, the planet orbit to stellar radius ratio $a/R_\star$ in the middle panel, and the impact parameter b in the lower panel.

2.5.2 Application to orbital signal reconstruction

The fact that the IRF preserves any signal at the period of the transit has positive consequences: it implies that potentially interesting signals, such as secondary eclipses, reflected light variations, or thermal emission variations, are preserved. The IRF therefore presents itself as an interesting tool to detect these signals. However as the remaining variations at the period of the transit after IRF-filtering can also be due to stellar variability signal at the planet's orbital frequency, any detection of planet phase variations will need to be analysed carefully. The residual stellar variability at the orbital period of the planet is a worse problem for the detection of the phase curve than for the detection of the secondary eclipse as the latter happens on a shorter timescale.

As the BT2 light curves were not built with any of these orbital signals, the study of the IRF performance in detecting planet orbital signals will have to be done on another sample of light curves.

2.5.3 Potential application to transit detection

Another potential application of the IRF would be at the detection stage. Among the 24 light curves of our sample, there were 2 where the transit signal was larger than the instrumental noise but where the residual stellar variability after NIF-filtering was too

strong to perform any kind of meaningful fit. Naturally, these events were not detected in the NIF-filtered light curves during the original blind test for which the light curves were generated. There are two more cases which we did include in our 20-strong comparison sample, as their transits after NIF-filtering could still be fitted, but for which transits were not detected in the original exercise: light curves 192 and 200. After applying the IRF, two of these 4 cases became detectable[5] (light curves 165 and 200), the other 2 cases remained undetectable due to the level of instrumental noise. Using the IRF as part of the detection process might therefore enable the detection of transits which would otherwise be missed around particularly active stars. However, since the IRF would have to be run at each trial period, and is relatively computationally intensive, this would require a very large amount of CPU time unless the algorithm can be significantly optimised. However, as radial velocity measurements are also affected by stellar activity (which induces radial velocity jitter and line bisector variations at the rotation period of the star), the new photometry detections will be difficult to follow-up in radial velocity, so it is not clear at this stage if the above CPU investment would be justified.

2.6   Conclusion

In the absence of a prior knowledge of the planet's orbital period, the transit and the stellar signal cannot be separated effectively if they overlap too much in the frequency domain. Because of this, commonly used pre-detection stellar variability filters, such as the NIF, alter the transit signal, causing systematic errors in the resulting star and planet parameters. We have quantified this effect using 20 CoRoT BT2 simulated light curves including transits, instrumental noise and stellar variability. We found that the effect on the transit signal can be very significant, leading to errors on the star-planet radius ratio of up to 50%.

  We thus developed the IRF to take advantage of the strictly periodic nature of planetary transits (in the absence of additional bodies in the system) to isolate the transit signal more effectively, following a method similar to the TFA algorithm previously developed for the reconstruction of transits in the presence of systematics. The IRF requires accurate knowledge of the transit period. We evaluated the performance of the IRF relative to the NIF and the no variability light curves by comparing a) the transit parameters from trapezoidal fits, b) the star-planet parameters from analytic transit fits, and c) the light curves themselves by visual examination. The results can be summarised as follows: the transits reconstructed with the IRF are systematically closer to the no variability case than the NIF-processed transits, and the improvement in the transit depth and duration can be very significant particularly in cases with large am-

---

[5]The detectability of the events was evaluated using the transit search algorithm of Aigrain & Irwin (2004), which was used in both CoRoT blind tests.

plitude and high frequency stellar variability. However, the full transit fits are affected by other factors including instrumental noise and the well known degeneracy between system scale and impact parameter, which dominate the final parameter estimates in approximately one third of the cases in our sample, or about half of the cases where the IRF provided a visual improvement over the NIF. The IRF will be most useful when applied to light curves which are strongly affected by stellar variability. The improvement in the planet parameters is likely to be better seen when the signal-to-noise ratio – for other noise than the stellar variability – of the light curve is high, as the noise allows degenerate solutions to the transit fit and thus keeps us from measuring the real impact of the IRF.

The IRF preserves any signal at the period of the transit, which implies that potentially interesting signals, such as secondary eclipses, reflected light variations, or thermal emission variations, are preserved.

Any power in the stellar variability signal at the frequency corresponding to the planet's orbital period is also preserved by the IRF. If required, this remaining variability can be removed locally using polynomial fits about the desired phase, but it is likely to limit the extent to which the IRF can be used to recover signals associated with the planet which vary continuously in phase.

2.7 Appendix

2.7.1 Best-fit parameters to BT2 transits

Table 2.2: Transit parameters (transit depth δ, total transit duration d_e, and internal transit duration d_i) derived from trapezoidal fits to the light curves with transit signal only ('no noise'), transit signal and instrumental noise only ('no stvar'), the BT2 light curves filtered using the pre-detection nonlinear iterative filter ('NIF'), and the same light curves filtered using post-detection iterative reconstruction filter ('IRF').

LC	period (days)	δ no noise	no stvar	NIF	IRF	d_e/P no noise	no stvar	NIF	IRF	d_i/P no noise	no stvar	NIF	IRF
126	4.576	0.00501	0.00495	0.00326	0.00504	0.0153	0.0153	0.0138	0.0153	0.0064	0.0064	0.0084	0.0079
131	6.880	0.00477	0.00469	0.00437	0.00448	0.0134	0.0121	0.0108	0.0121	0.0056	0.0017	0.0098	0.0098
133	8.128	0.00168	0.00161	0.00155	0.00160	0.0058	0.0058	0.0047	0.0057	0.0016	0.0010	0.0026	0.0015
135	3.733	0.00155	0.00148	0.00144	0.00152	0.0147	0.0147	0.0147	0.0147	0.0062	0.0062	0.0073	0.0090
145	5.557	0.00938	0.00949	0.00931	0.00923	0.0167	0.0153	0.0167	0.0167	0.0054	0.0064	0.0067	0.0086
152	7.360	0.00185	0.00185	0.00158	0.00176	0.0115	0.0125	0.0104	0.0115	0.0060	0.0065	0.0090	0.0071
154	10.987	0.00056	0.00065	0.00054	0.00061	0.0172	0.0172	0.0155	0.0237	0.0088	0.0038	0.0067	0.0051
162	4.171	0.00933	0.00922	0.00585	0.00894	0.0167	0.0167	0.0138	0.0167	0.0037	0.0037	0.0074	0.0070
169	5.195	0.00772	0.00770	0.00504	0.00769	0.0209	0.0209	0.0191	0.0209	0.0066	0.0066	0.0109	0.0107
177	7.339	0.00271	0.00267	0.00252	0.00260	0.0209	0.0209	0.0191	0.0209	0.0066	0.0046	0.0090	0.0107
186	4.373	0.00683	0.00679	0.00649	0.00690	0.0209	0.0209	0.0191	0.0209	0.0087	0.0087	0.0134	0.0127
192	3.915	0.00085	0.00102	0.00071	0.00076	0.0086	0.0094	0.0078	0.0078	0.0029	0.0023	0.0078	0.0070
193	6.763	0.00749	0.00747	0.00847	0.00858	0.0167	0.0167	0.0167	0.0167	0.0070	0.0070	0.0065	0.0086
196	4.608	0.01378	0.01384	0.00509	0.01288	0.0248	0.0248	0.0201	0.0248	0.0127	0.0127	0.0201	0.0175
200	5.995	0.00317	0.00313	0.00185	0.00311	0.0095	0.0095	0.0059	0.0086	0.0023	0.0023	0.0052	0.0038
208	4.064	0.00313	0.00301	0.00278	0.00302	0.0267	0.0267	0.0242	0.0267	0.0136	0.0136	0.0158	0.0162
220	7.253	0.00215	0.00212	0.00181	0.00216	0.0230	0.0250	0.0210	0.0230	0.0050	0.0030	0.0148	0.0140
223	5.237	0.00771	0.00736	0.00065	0.00761	0.0184	0.0184	0.0088	0.0200	0.0059	0.0059	0.0082	0.0102
225	2.613	0.01061	0.01053	0.01032	0.00998	0.0344	0.0344	0.0311	0.0344	0.0073	0.0073	0.0224	0.0174
233	3.083	0.00461	0.00460	0.00431	0.00459	0.0153	0.0153	0.0153	0.0153	0.0035	0.0035	0.0040	0.0049

Table 2.3: Star-planet parameters (planet to star radius ratio $R_p/R_\star$, system scale $a/R_\star$, and impact parameter b) derived from full transit fits. The columns corresponding to the 4 sets of light curves used in the fits are labelled as in Table 2.2.

LC	period (days)	$R_p/R_\star$ no noise	no stvar	NIF	IRF	$a/R_\star$ no noise	no stvar	NIF	IRF	b no noise	no stvar	NIF	IRF
126	4.576	0.0799	0.0800	0.0698	0.0817	12.27	12.09	13.14	11.89	0.862	0.870	0.858	0.872
131	6.880	0.0760	0.1687	0.0779	0.0749	15.64	9.64	10.65	12.30	0.825	1.058	0.893	0.873
133	8.128	0.1389	0.2836	0.0557	0.0642	17.70	15.59	20.99	20.65	1.074	1.233	0.961	0.979
135	3.733	0.0469	0.0369	0.0372	0.0393	10.00	20.34	20.62	16.75	0.916	0.422	0.397	0.676
145	5.557	0.1050	0.1044	0.1080	0.1079	14.35	14.24	13.34	13.38	0.788	0.795	0.822	0.820
152	7.360	0.0481	0.0476	0.0495	0.0475	15.48	15.80	13.29	15.35	0.860	0.836	0.908	0.865
154	10.987	0.0263	0.0313	0.0323	0.0245	11.54	9.24	8.50	23.47	0.829	0.914	0.933	0.023
162	4.171	0.1335	0.1245	0.1102	0.1272	10.91	11.13	11.58	11.21	0.927	0.910	0.911	0.913
169	5.195	0.0918	0.0921	0.0781	0.0925	12.61	12.36	11.84	12.25	0.720	0.736	0.750	0.735
177	7.339	0.0486	0.0555	0.0549	0.0549	17.53	10.92	11.66	11.83	0.150	0.794	0.760	0.749
186	4.373	0.0839	0.0872	0.0869	0.0876	12.71	11.21	11.69	11.69	0.667	0.760	0.738	0.735
192	3.915	0.0504	0.0488	0.0152	0.0240	10.97	12.64	12.09	8.51	0.988	0.978	0.656	0.926
193	6.763	0.0904	0.0929	0.0990	0.0996	15.26	13.90	14.60	14.42	0.723	0.780	0.759	0.764
196	4.608	0.1162	0.1160	0.0970	0.1150	11.87	11.91	14.47	11.98	0.546	0.533	0.000	0.547
200	5.995	0.0860	0.3645	0.0444	0.0755	14.59	12.10	22.72	15.91	0.965	1.290	0.916	0.945
208	4.064	0.0582	0.0588	0.0597	0.0592	9.33	9.30	8.63	8.93	0.710	0.716	0.780	0.763
220	7.253	0.0475	0.0436	0.0433	0.0462	10.71	12.86	15.67	12.93	0.722	0.519	0.019	0.550
223	5.237	0.0976	0.0811	0.0452	0.1008	11.33	19.75	18.84	9.82	0.835	0.017	0.231	0.854
225	2.613	0.1033	0.1028	0.1017	0.0989	8.28	8.22	8.59	8.84	0.624	0.625	0.575	0.552
233	3.083	0.1426	0.1500	0.1414	0.1378	9.29	9.38	9.30	9.23	1.020	1.029	1.020	1.012

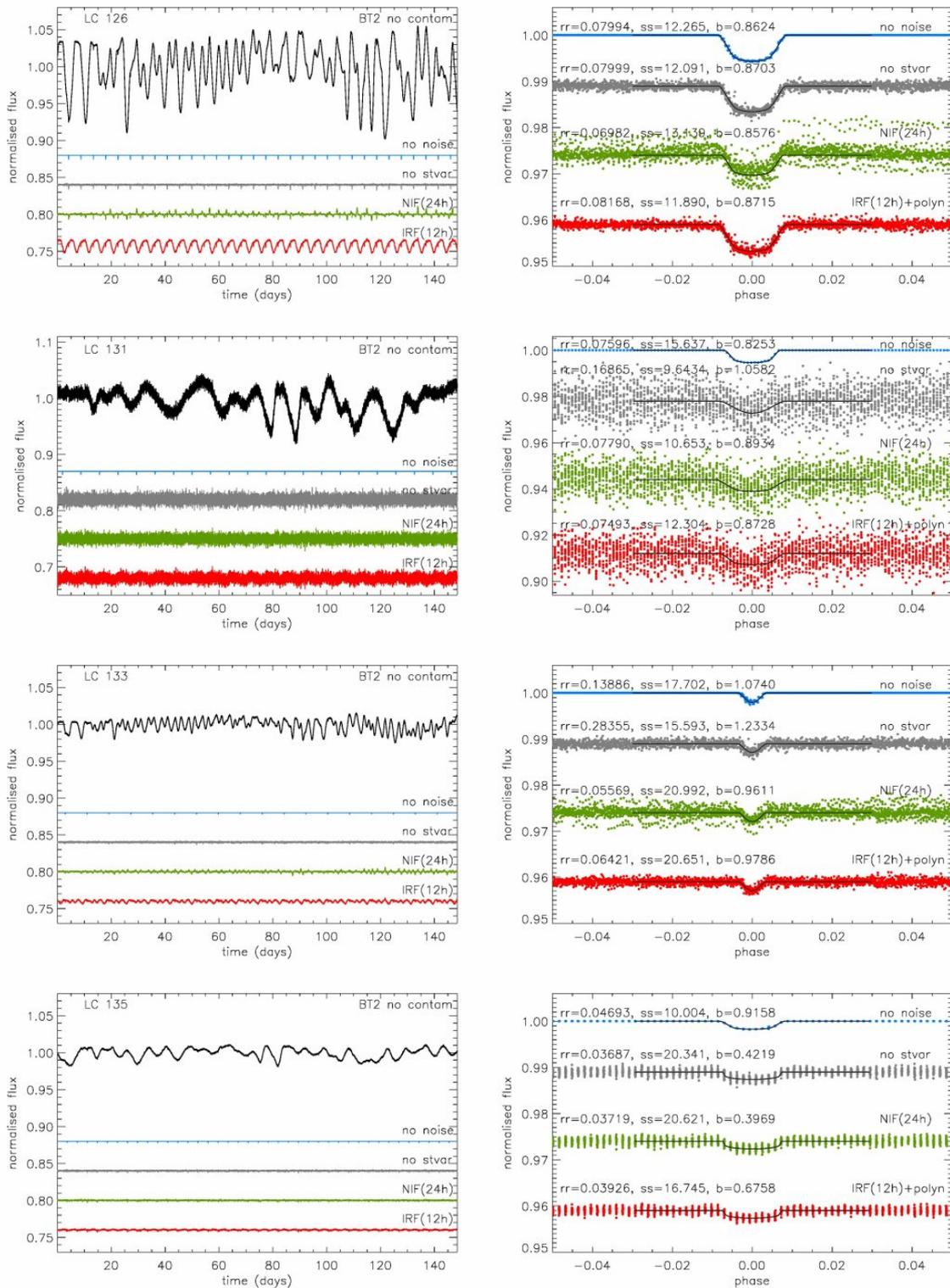### 2.7.2 Full BT2 light curve sample

Figure 2.8: The 20 BT2 light curves in the comparison sample (left:unfolded; right : phase-folded section around the transit). The light curve number is shown on the plots in the left column (original BT2 numbering scheme) and the planet to star radius ratio (rr), system scale (ss), and impact parameter (b) in the right column.
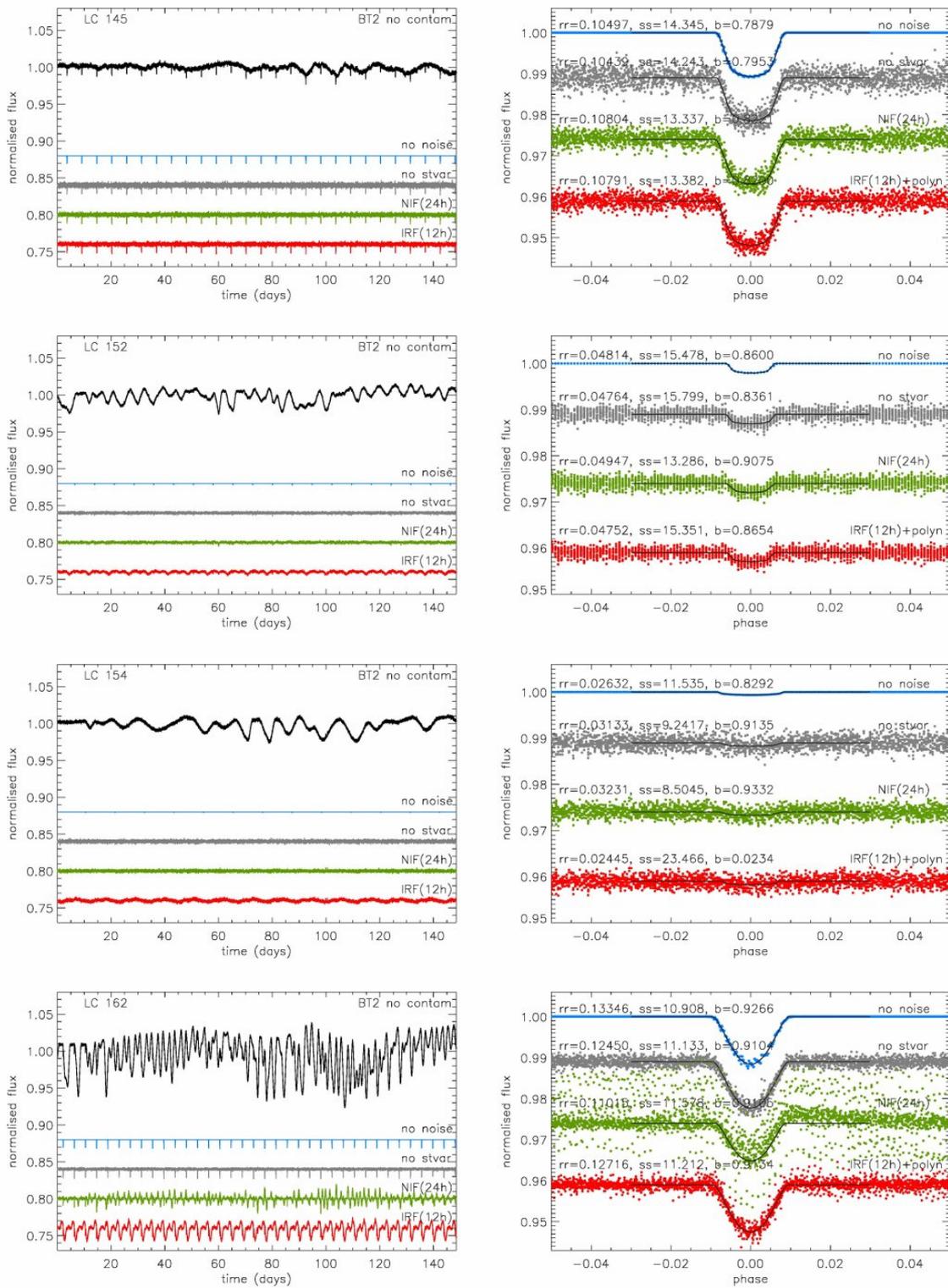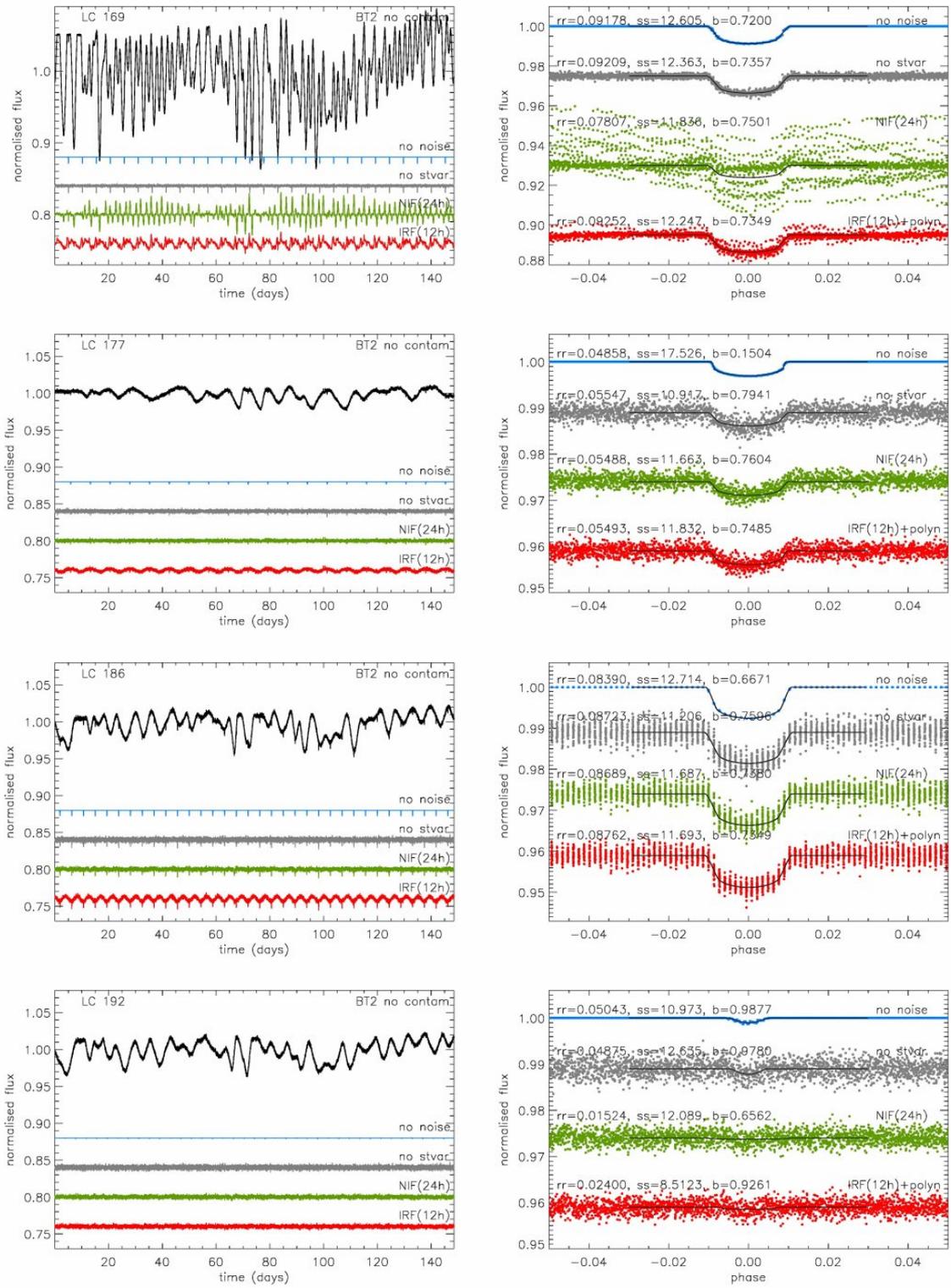
Figure 2.8: continued

Figure 2.8: continued

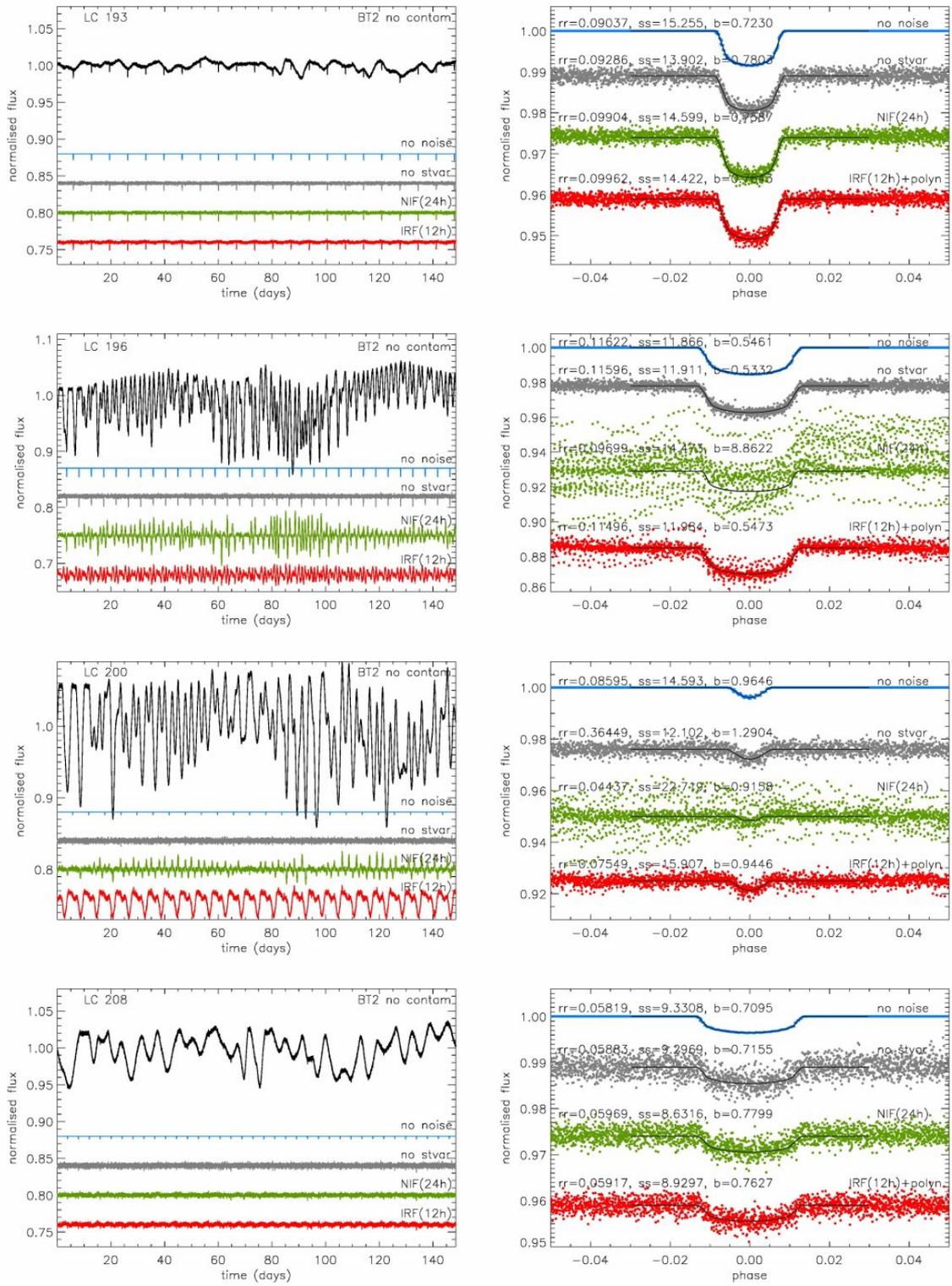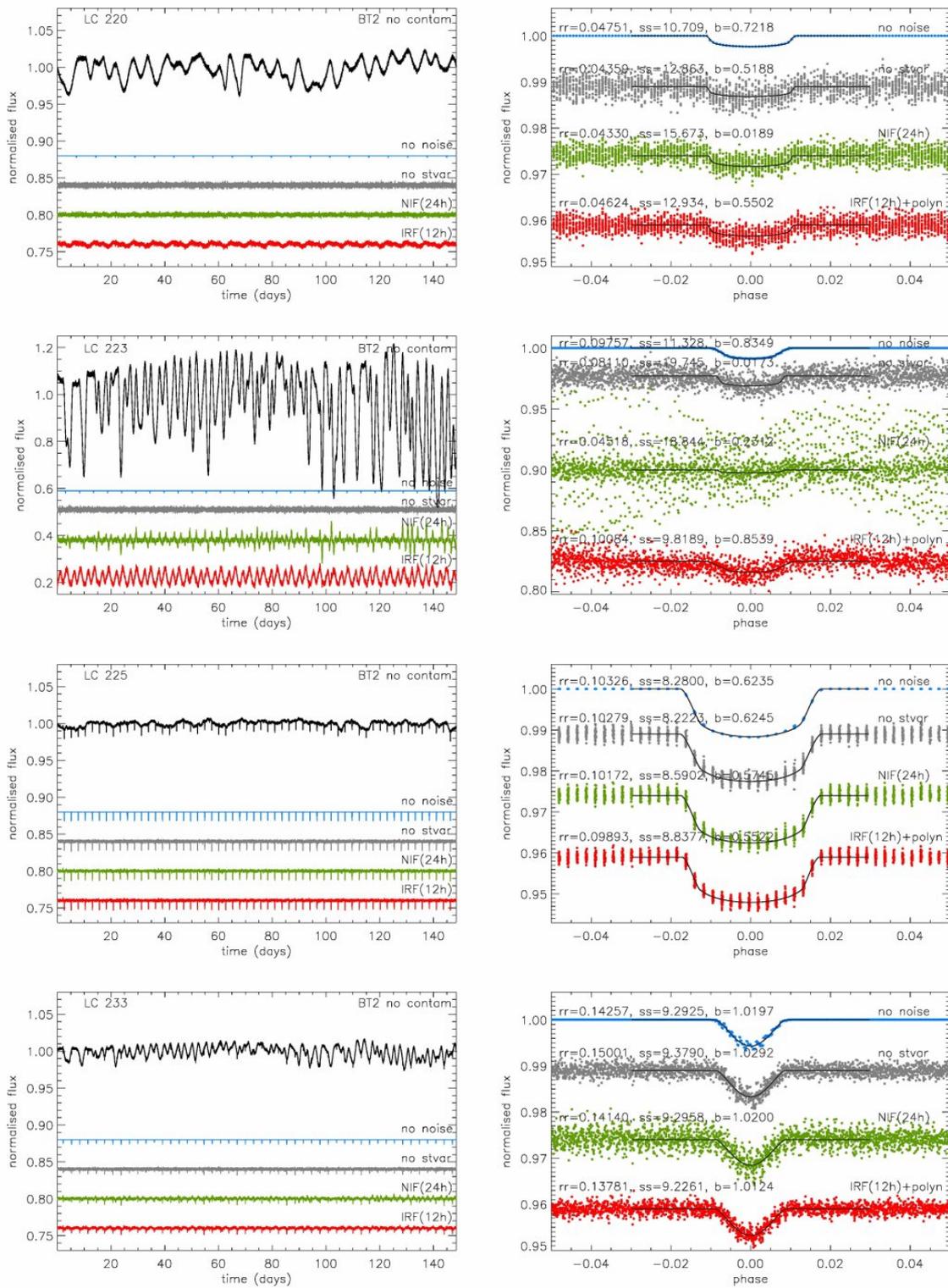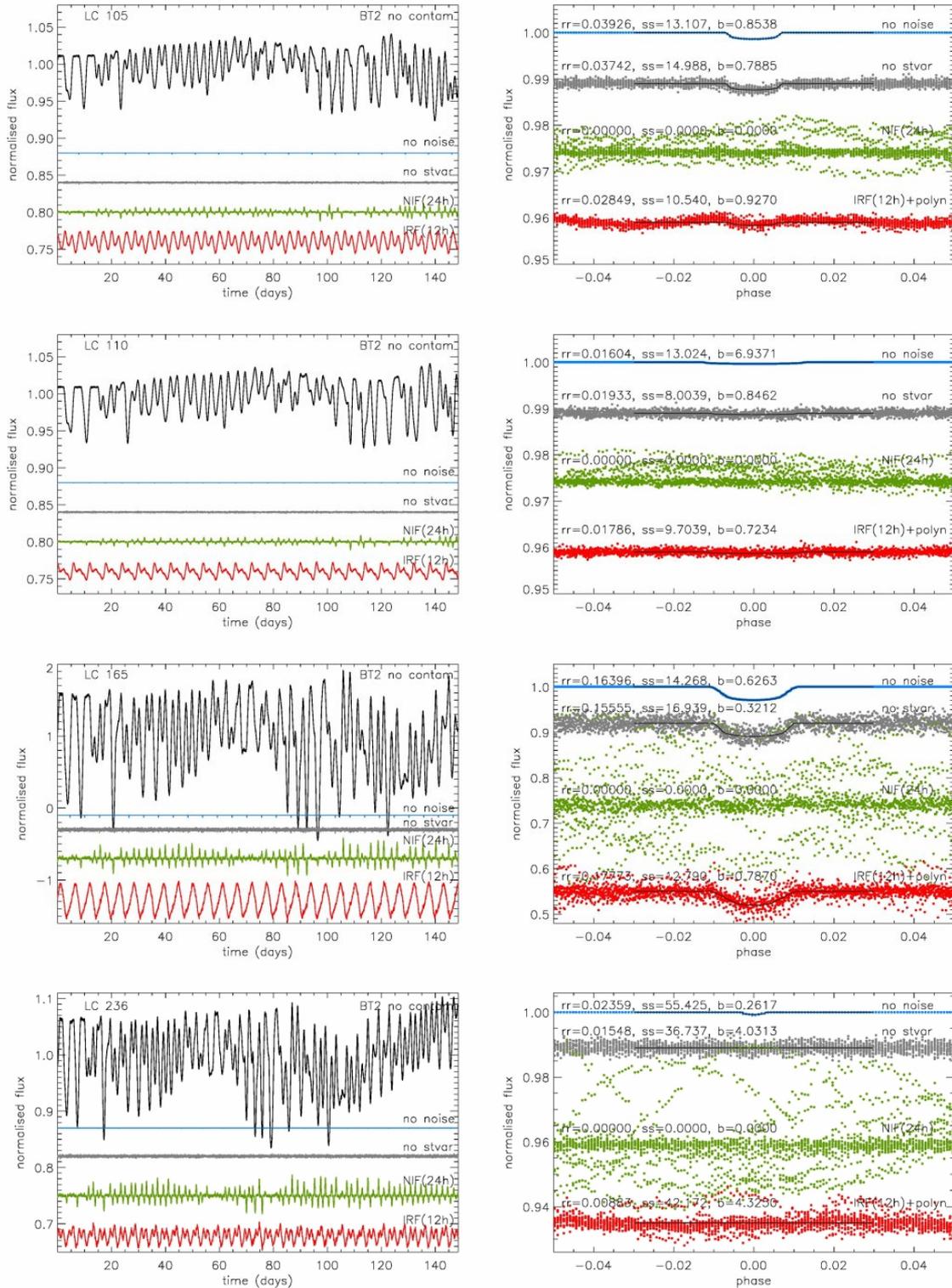Figure 2.8: continued

Figure 2.8: continued

Figure 2.9: The 4 BT2 transit light curves where the transit was undetectable after apply-ing the NIF and miningfull fits to the resulting transits were not possible. Same legend as Fig. 2.8. These transits became boarder-line detectable in the IRF-filtered light curves.