Chapter 6

The COROT blind experiment

6.1 Introduction

This experiment, which took place between January and May 2004, was organised by a network of research groups across Europe who are involved in the preparation of the Franco-European transit search mission COROT, in particular its exo-planet search program.

At the 5th COROT week in December 2003, the need to test the capabilities of various detrending and transit detection algorithms on realistic light curves emerged. The satellite and instrument design phases were complete, and the instrument model was at a sufficiently advanced stage to provide realistic simulated light curves at least in white-light. An experiment, carried out in a concerted way between the different groups involved in the preparation of the data analysis, would minimise duplication of work while highlighting any areas where difficulties arose, and test some of the details of the observing strategy. It was therefore agreed to carry out such an experiment within the COROT Exo-planet Working Group (EWG), the results of which were to be announced at the 6th COROT Week in May 2004. The light curves from this experiment would then be made available to the wider transit search community. This was foreseen to be the first in a series of such experiments, subsequent generations drawing on the lessons from the first.

A collaborative publication in which all the participants will describe their contributions is currently in preparation (Moutou et al. 2004a). The purpose and organisation of the experiment and the generation of the light curves (Sections 6.1.1 to 6.2), whose descriptions are mainly drawn from the draft of this publication, are summarised here in order to set the stage for the transit search I performed in collaboration with M. Irwin (Section 6.3.1). A brief description of the methods used by the other participants is also given, though I have only limited information on the algorithms some of the groups used at this stage. The lessons learnt from the exercise are

discussed in Section 6.4, and future prospects in Section 6.5

6.1.1 Purpose of the blind experiment

Rather than a competition between the groups involved, the experiment was designed to identify the algorithm(s) most suited to the COROT data, and to estimate preliminary detectability limits in the presence of all the noise sources currently foreseen. It was designed along a 'hare and hound' scenario: light curves containing noise and stellar variability were built by a single 'game master' from components provided by various contributors, and transits and mimics of stellar origin were inserted. The final set of 1001 light curves – their contents known only to the game master – was then made available to the members of the EWG, who attempted to detrend the light curves and detect the transits by the methods of their choosing.

From my point of view, this was a golden opportunity to put the tools developed in Chapters 2 and 4 to the test and to compare their performance to that of others. It was also an opportunity to apply the simulator described in Chapter 3, which was designed exactly for this type of purpose, for the benefit of the wider exo-planet community.

6.1.2 Participants

The light curves included instrumental and photon noise simulated by D. Blouin from the Laboratoire d'Astrophysique de Marseille (LAM), stellar micro-variability simulated independently by A. Lanza from the Osservatorio Astrofisico di Catania and myself, planetary and stellar eclipses and variable star light curves simulated or collected from the literature by C. Moutou and F. Pont (LAM). The different components were combined to form light curves by C. Moutou, who coordinated the experiment.

Although any members of the EWG were invited to take part in the detection process, only five groups were able to send their results to C. Moutou by the deadline of May 1st: a team from the Institute of Planetary Research in Berlin-Aldershof, led by H. Rauer, two independent teams from LAM (P. Guterman and V. Guis), one team from Geneva Observatory led by S. Zucker and the IoA team composed of M. Irwin and myself.

6.2 The simulated light curves

6.2.1 General characteristics

The simulated light curves correspond to a COROT long run (lasting 150 days with 8.5 min sampling), and to the characteristics of the exo-planet field (see Section 1.2.3.2 for details of the COROT design and the distinctions between asteroseismology and exo-planet program). All light curves were simulated as though they had arisen from the same observation run, i.e. they have a common set of observation times and common time-dependent systematics.

Only light curves in white light were generated, as neither the COROT instrument model nor the stellar micro-variability simulating tools were at the time capable of producing coloured light curves. Work is now underway to upgrade the instrument model, and Lanza et al. (2004) have developed a technique for including colour information in stellar light curves. I am also planning to work on this question in the near future.

6.2.2 Instrumental and photon noise

The instrument model works by taking as input a number of simulated PSFs and a list of star positions and magnitudes, generating images by addition of the appropriate noise sources, computing appropriate masks, performing aperture photometry within each mask (applying the appropriate noise sources again), and then correcting for the known noise sources as well as we expect to be able to correct them in the real data.

25 PSFs were used, corresponding to stellar effective temperatures between 4500 and 6750 K and magnitudes between 12 and 16. Photon noise was added following Poisson statistics. A flat-field non-uniformity of $\sim 1\%$ was introduced. The read-out noise is $10\,e^{-}\text{pixel}^{-1}$ and the jitter noise is negligible in white light. Zodiacal light is uniform over the CCD and constant along the orbit (with a value of $12\,e^{-}\text{pixel}^{-1}\text{s}^{-1}$), and is thus fully corrected for, resulting only in additional photon noise. On the other hand, the Earth scattered light is more difficult to correct, because it varies over the CCD, as well as along the orbit. This noise source was thus corrected only to first order, in a delibrate attempt to simulate a worst-case scenario. A scattered light variation component with a maximum amplitude of $1\,e^{-}\text{pixel}^{-1}\text{s}^{-1}$ was simulated and added to each light curve. A correction was then computed by multiplying the scattered light component by a factor between 0.9 (underestimation) and 1.1 (overestimation), and applied to the light curve, leaving residuals of $\pm 10\%$. As illustrated by Figure 6.1, residual scattered light is the dominant signal in

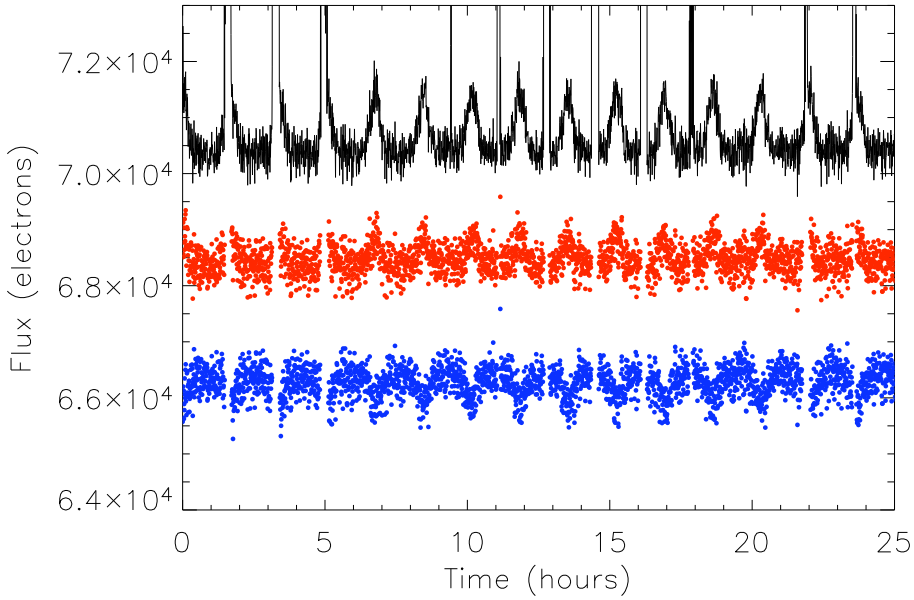

Figure 6.1: Example of an instrumental light curve before (top) and after the partial correction of the scattered light variations, which are either underestimated (middle) or overestimated (bottom). The sharp peaks in the upper panel are due to the South Atlantic Anomaly crossing, and correspond to gaps in the lower two panels. Plot provided by C. Moutou. .

the COROT instrumental light curves.

The large numbers of impacts by high energy particles as COROT passes through the South Atlantic Anomaly (SAA) makes the exposures taken during the SAA crossings unusable. This leads to semi-periodic gaps in all light curves lasting $\sim$ 10 min, occurring $\sim$ 10 times per day. The gaps were not exactly synchronous in all the simulated light curves because an upper flux threshold was applied to each light curve individually to determine where the gaps would be.

6.2.3 Stellar micro-variability

Two sets of light curves containing stellar micro-variability were simulated independently.

One set was produced by myself, using the model described in Chapter 3. It contained 45 light curves covering a grid of 9 spectral types (F5, F8, G0, G2, G5, G8, K0, K2 & K5) and 5 ages (0.625, 1.0, 2.0, 3.0 & 4.5 Gyr). The light curves are illustrated in Figure 6.2.

The second set was produced by A. Lanza and collaborators from Catania Observatory. They have developed a method to reproduce the Total Solar Irradiance (TSI) variations, as observed by SoHO/VIRGO, based on the differential rotation of 3

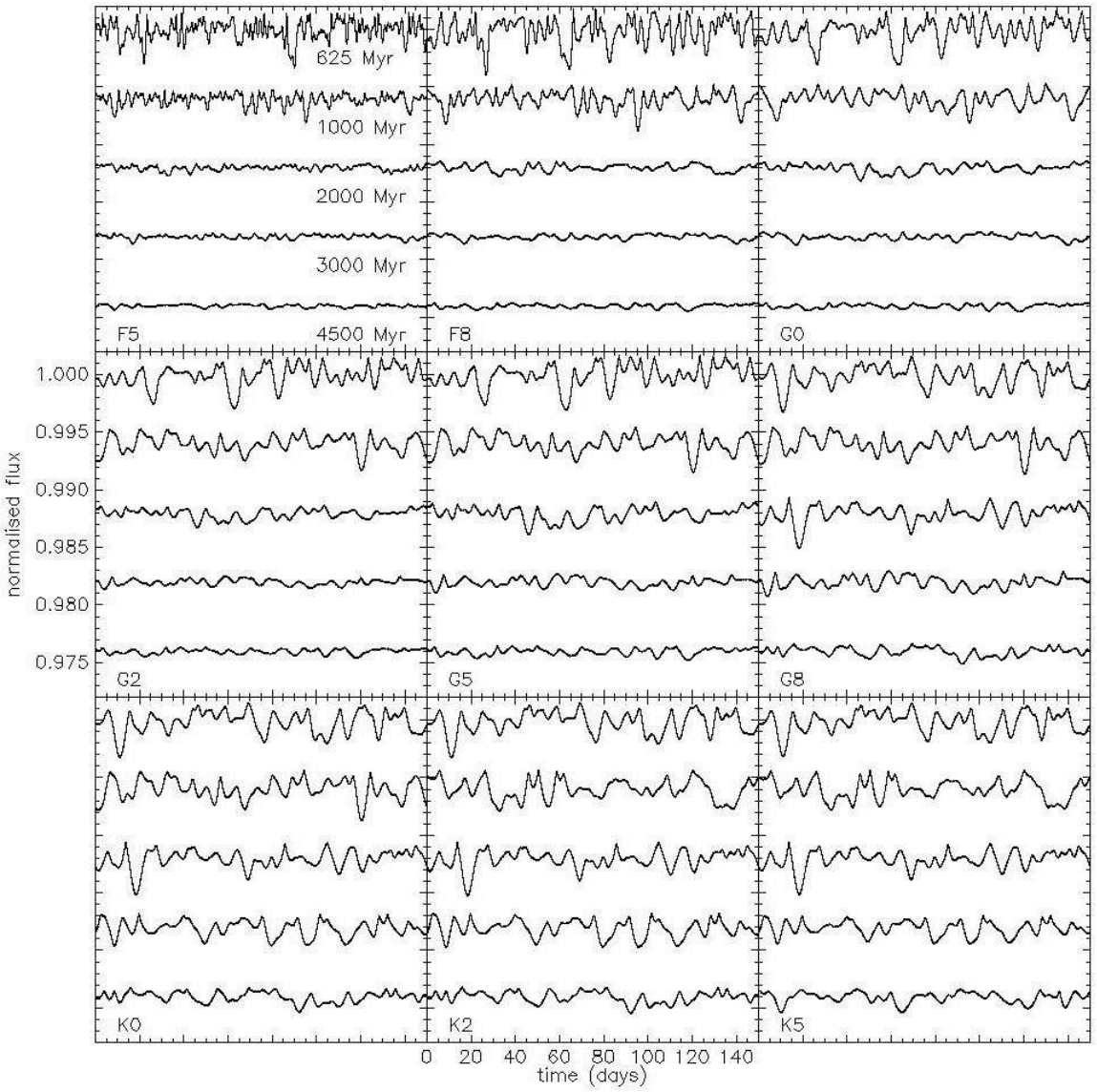

Figure 6.2: The 45 stellar micro-variability light curves simulated for the COROT blind experiment with the model described in Chapter 3.

active regions, composed of faculae and spots, whose sizes and positions on the star's surface are allowed to vary, but whose contrast factor is fixed (Lanza et al. 2003). Fits are made to 14 d portions of the TSI light curve spaced by 7 d, and linear interpolation between successive best fits is used to minimise discontinuities between fits. To produce light curves for stars other than the Sun, Lanza et al. apply a scaling factor f to the area of the active regions, and the rotation period is modified to reproduce that of the star being modelled. The factor f is deduced from an empirical scaling law established by Messina et al. (2001, 2003), relating the amplitude of

$V-$band variations to the active region area coverage. For stars with rotation periods longer than 12 d, for which no optical variability information is available, f was taken to be between 1.5 and 6 for F5V to K5V stars respectively. In order to account for shorter timescale variations, which cannot be explained by rotational modulation, they also inserted an additional component in their light curves, calculated by scaling the residuals from their fit to the solar light curves by a factor $3f$. This higher scaling factor was chosen arbitrarily to mimic a worst-case scenario. Ten light curves were thus generated, with spectral types G5, G0 and G8 and rotation periods 3, 10 and 20 d.

Some light curves in the resulting set (totalling 55) were then duplicated to reproduce the expected distribution of spectral types (40% F dwarfs, 40% G dwarfs, 20% K dwarfs). In order to obtain a sufficient number of non-identical stellar micro-variability light curves, C. Moutou then applied an arbitrary amplitude scaling factor between 0.5 and 2.0, and a time scaling factor between 1.0 and 1.2, as well as arbitrary time shifts. The contributions by A. Lanza and myself were treated in the same way, each light curve being assigned a stellar micro-variability component from one set or the other at random.

6.2.4 Transit signals

Twenty planetary transits were simulated by C. Moutou using the Universal Transit Modeller (UTM) of Deeg (1999), using limb-darkening coefficients calculated from ATLAS9 models and the COROT bandpasses (Barban, priv. comm.).

The choice of transit signals was aimed at sampling a variety of cases and testing detectability limits by including a large number of shallow transits in light curves of varying noise level. The planet radii used were in the range $1.6\,R_\oplus \leq R_{\rm pl} \leq 1.3\,R_{\rm J}$, and the periods in the range $4.8 \leq P \leq 88.4$ d. One double planet system was included. The process by which the transit signals were inserted in the light curve was not entirely random, in that care was taken to insert the deepest transits into the light curves of faint/active stars.

This part of the experiment was not designed to be realistic – 2% of the light curves contained planetary transits, which is approximately an order of magnitude larger than what is expected (Bordé et al. 2003), and the parameters of the planets do not reproduce the observed distributions. Instead, the number was chosen to allow variety while keeping the size of the dataset small for this first generation exercise, and the parameters to test the algorithms in limiting cases which were expected to be problematic.

6.2.5 Stellar mimics and variables

Ten low depth stellar eclipses were simulated. Six were grazing binaries, simulated using the algorithm of Mandel & Agol (2002) and the 'Nightfall' software of Wichman (1998). Four were diluted background binaries, simulated with UTM. In addition, one triple eclipsing system was simulated, also with UTM. Again, the parameters were chosen to explore a variety of cases rather than follow observed distributions. The background eclipsing binaries were 3 to 5 mag fainter than the principal target in the mask. As well as the V- (grazing) or U-shaped (annular) primary eclipses, the eclipsing binary light curves include secondary eclipses and, when applicable, out-of-eclipse modulation due to ellipsoidal deformation of the primary.

In addition, the light curves of stellar variables were taken from the literature and the archive of the AAVSO (American Association of Variable Star Observers) and inserted. These included a low amplitude δ-Scuti, a classical Cepheid, a β Cephei, the semi-regular variable Z Uma and the irregular Z Cam.

6.2.6 Background pollution

The relatively large size of the COROT masks imply that a number of faint background stars are expected to fall in the mask of most COROT targets. The variability of these background stars will be superimposed with that of the main target and may significantly contribute to the overall result. To simulate this effect, the contribution of one background star – computed in the same way as the principal target's, but with a different stellar variability component – was added to each light curve. The background star's magnitude followed a $2^{\Delta m}$ distribution with $0 \leq \Delta m \leq 6$, thus including stars up to 22^{nd} magnitude (see note regarding magnitude distribution in the next Subsection).

6.2.7 The final set of 1001 light curves

Of the final 1001 light curves, 964 have no stellar or planetary transit signal (apart from the stellar micro-variability component, which all light curves have). The temperature of the simulated stars follows a uniform distribution between 4000 and 6000 K, approximating a magnitude-limited (rather than volume limited) sample. The magnitudes were drawn from a 2^m distribution, approximating an isotropic spatial distribution. This corresponds to the distribution measured from observations of the COROT fields obtained in the context of the ground-based preparatory observations program, and is in agreement with the star counts expected from the Besançon model of the Milky Way (Robin et al. 2003, 2004). All light curves have a temporal sampling rate of once per 8.5 min, except two which have 32 s sampling (this 'oversampling' mode will be

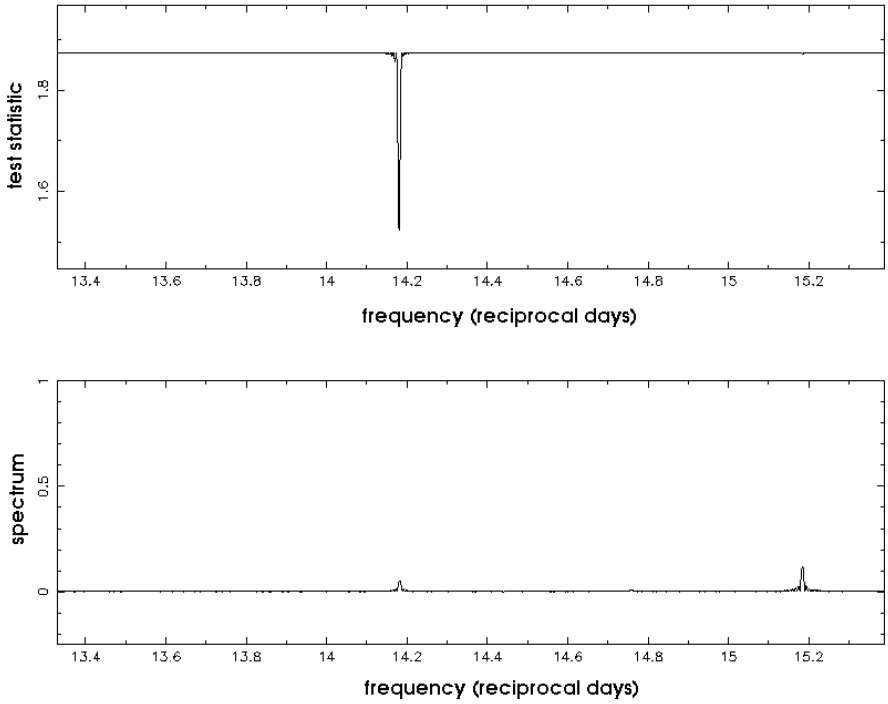

Figure 6.3: Sine-curve fitting to identify the period of the scattered light variations. Top panel: test statistic versus trial frequency. The test statistic is the reduced χ^2 of the residuals from the best fitting-sine curve at each frequency. The best-fit frequency is given by the minimum value. Bottom panel: window function amplitude spectrum.

possible for a small number of COROT light curves). The light curves that do contain planetary transits and stellar events are summarised in Tables 6.3 and 6.4.

These light curves, in units of photon counts, were supplied to the detection teams with no information about the way in which they were generated. When real COROT light curves come to be analysed, some information regarding magnitude, spectral type, luminosity class and contamination by neighbours *will* be available. However, this information, while useful at the transit candidate characterisation stage, is not fundamental at the transit detection stage. The light curves as received by the detection teams are referred to hereafter as level 0 light curves.

6.3 The transit search

The detection teams were asked to return a list of canditate transits with their parameters, as well as a list of other events identified in the course of the transit search. The results were to be sent by email to C. Moutou by May 1st 2004.

6.3.1 Method used by the IoA team (team 5)

The method proceeds along sucessive steps outlined in Figure 6.4 and detailed below.

6.3.1.1 Removal of residual scattered light variations

Upon receipt of the level 0 light curves, the first noticeable feature were the short period variations due to residual Earth-scattered light, which were positive or negative depending on the light curves, and the SAA gaps.

The exact period of the residual scattered light variations was determined from least-squares fitting of sine curves in the period range 0.065 to 0.075 d (see Figure 6.3). In the spirit of total ignorance of the light curve contents, this period determination was carried out individually for each light curve, though as expected the best fit period was always found to be very close to the satellite orbital period of 0.0705 d, the variations being within the errors on the fitted period. In each case, the light curve was then phase folded at the best-fit period. The phase-folded light curve was then smoothed, using a box-car filter with a width of 501 data points, and the smoothed version was subtracted from the phase-folded light curve, before remapping to the original time base and adding the original median level. The resulting light curves are referred to hereafter as level 1 light curves. An example of the scattered light removal procedure is illustrated in Figure 6.5. The effect of the removal on a section of the same light curve is also shown in the top and middle panels of Figure 6.7.

6.3.1.2 Removal of other trends and 'glitches' common to all light curves

During the scattered light removal stage, a number of 'glitches' and trends common to all light curves became apparent. This is consistent with the fact that all light curves were simulated as though they shared the same set of observation times. For example, two of the scattered light variation cycles appeared to be 'missing' approximately 78 d after the start of the level 0 light curves. The subtraction of the smoothed phase-folded light curve therefore induced two artificial 'bumps' or 'dips' at that particular location in the level 1 light curves.

In order to remove all systematic trends common to all light curves, a Principal Component Analysis (PCA) approach, based on a variant of Gram-Schmidt orthogonalisation, was initially considered. However, it involved the manipulation of 1000 by 25 056 element arrays (each light curve contains 25 056 data points), which would be computationally rather expensive. Additionally, the mean of all the light curves appears to contain most of the systematic trends and glitches which needed to be removed. The following, simplified procedure was therefore adopted:

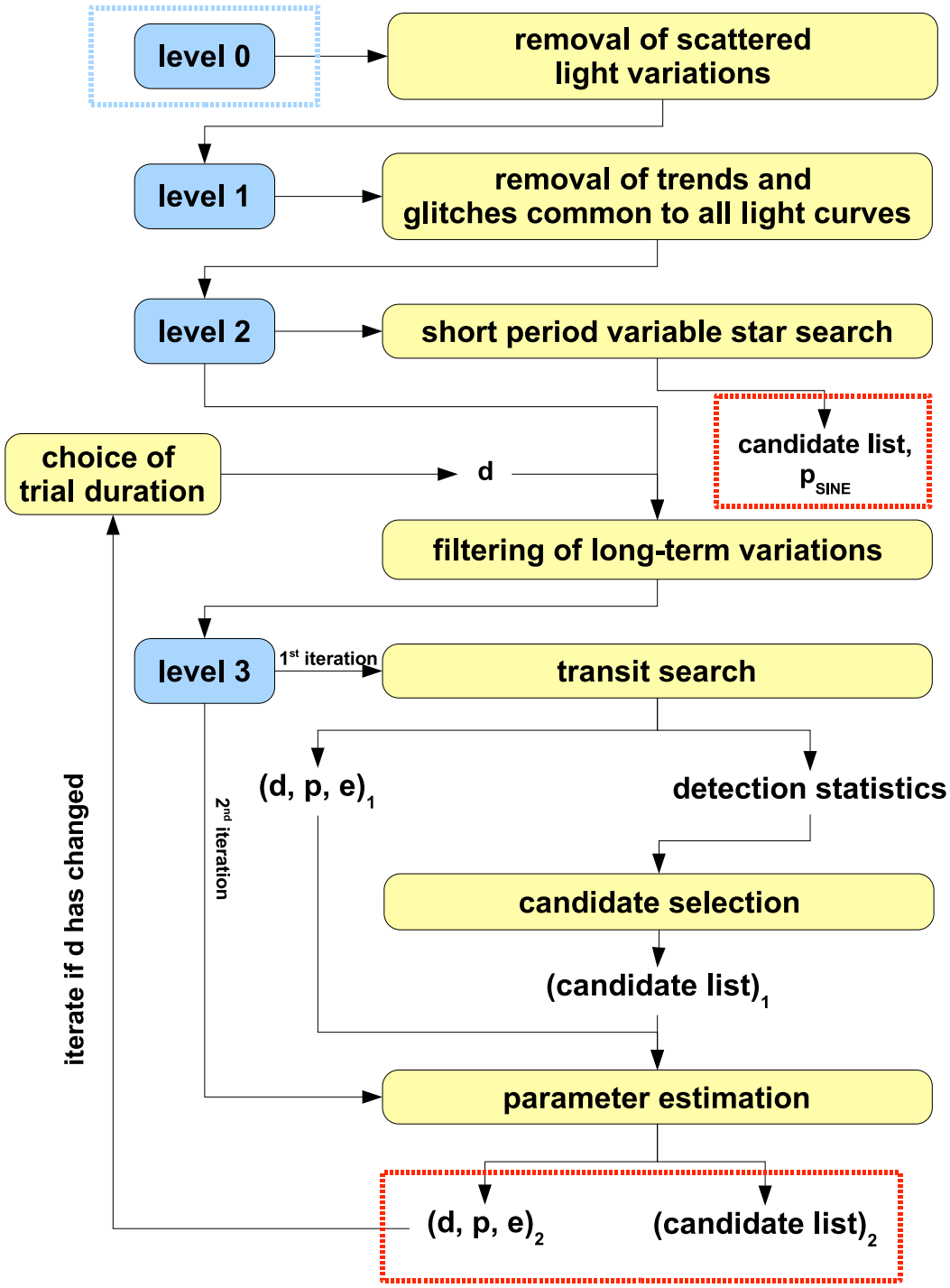

Figure 6.4: Main steps of the method employed by the IoA team, starting from the light curves as received from C. Moutou (blue dashed box) and ending with the short period variable and planetary transit candidate lists (red dashed boxes).

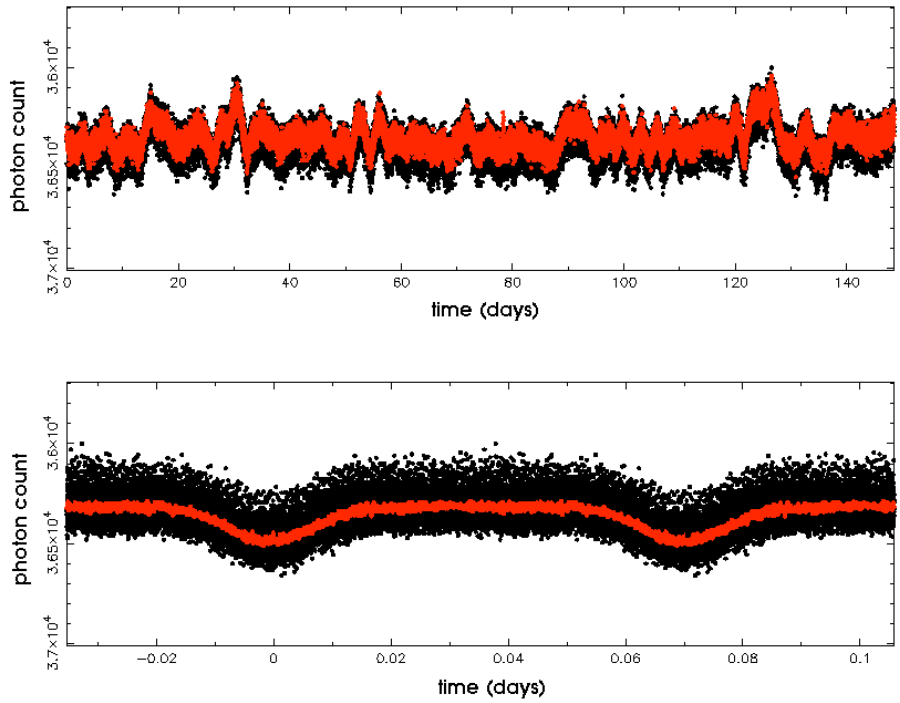

Figure 6.5: Removal of the scattered light variations. Top panel: Input (level 0, black) and output (level 1, red) light curves. Bottom panel: Light curve phase-folded at the best-fit period (black) and smoothed (red). The level 1 light curve (red curve in top panel) is obtained by subtracting the red from the black curve in the bottom panel and remapping to the original time base.

- The median count level and scatter σ of each level 1 light curve were computed. Note that a robust scatter estimate, $\sigma = 1.48 \times$ MAD, where the MAD is the median of the absolute deviations from the median, was used throughout the present analysis. This estimate approximates the standard deviation in the case of white Gaussian noise, but is less sensitive to outliers (Hoaglin et al. 1983).
- Each level 1 light curve was median subtracted and scaled to unit scatter.
- A 'scaled common average' light curve was then computed by taking the median of all 1001 median subtracted, scaled light curves at each time point (NB: light curves 1000 and 1001 were simply rebinned to 8.5 min sampling and treated throughout the analysis identically to the other light curves), and scaling the result to have unit variance. An array of scatter values for each time point – the 'scaled common error' array – was also computed. Excerpts from the scaled common average and error arrays are shown in Figure 6.6.
- The scaled common average light curve was then 'fitted' to each level 1 light curve by finding the coefficient k which minimises $\sum_{i=1}^{N}(d_i - k\,m_i)^2$, where N is

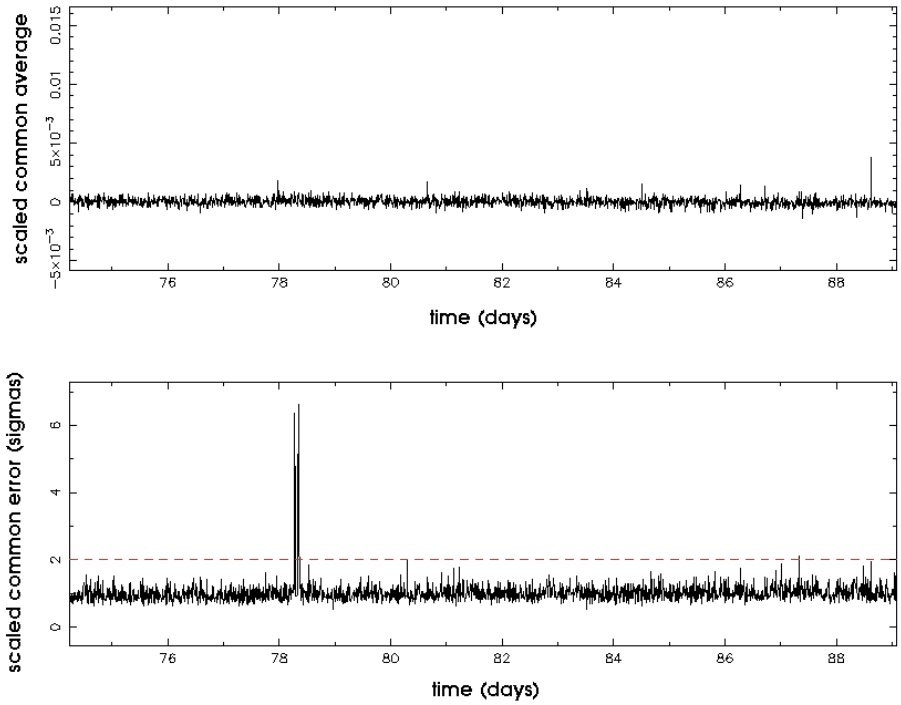

Figure 6.6: Excerpts from the scaled common average light curve (top) and the scaled common error array (bottom) in the time range 76 to 88 days. The 2 σ cut-off (red dashed line in bottom panel) is used to exclude common 'glitches' from all light curves.

the number of data points and d_i and m_i are the i^{th} points in the level 1 and scaled common average light curves respectively. If m_i is normalised such that $\sum_{i=1}^{N} m_i^2 = 1$ then this simplifies to $k = \sum_{i=1}^{N} d_i\, m_i$. The scaled common average light curve, multiplied by the factor k, was then subtracted from the level 1 light curve.

- The scaled common error array was also used to exclude common 'glitches': a 2σ cut-off was applied and time points corresponding to values of the scaled common error array above the cut-off were excluded from all light curves.
- The resulting light curves, free of common systematic trends (to first order at least) and glitches, are referred to hereafter as level 2 light curves.

Excerpts from an example of level 0, level 1 and level 2 light curve are shown in Figure 6.7, illustrating the removal of both residual scattered light variations and glitches.

6.3.1.3 Identification of short-period variable stars

At that stage, the least-squares sine-curve fitting program was run on all level 2 light curves with a range of trial periods p of 0.1 to 5 d (this range was designed to probe

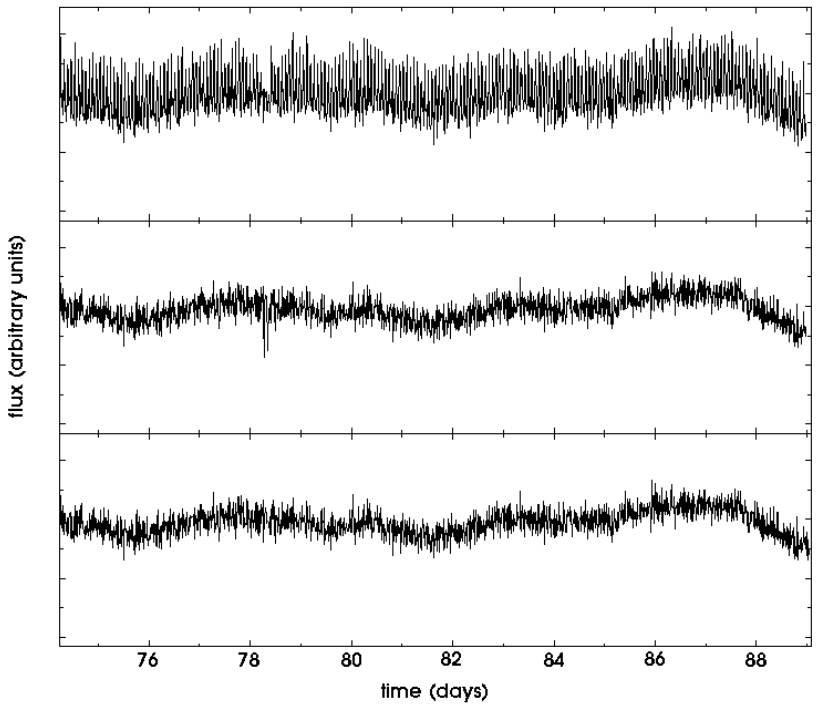

Figure 6.7: Excerpts from an example of level 0 (top), level 1 (middle) and level 2 (bottom) light curves in the time range 76 to 88 days. Note the residual scattered light variations in the level 0 light curve and the 'glitches' at ∼79 d in the level 1 light curve.

periods shorter than those explored through the transit search). This process took approximately 12 hours on a 2 GHz Dual-AMD Linux PC. In seven cases, the subtraction of the best-fit sine component led to a significant reduction in the reduced χ^2 of the light curve. These are summarised in Table 6.1. For these seven cases, the search was iterated to find additional periodicities with an extended period search range.

6.3.1.4 Variability filtering and transit search

The iterative non-linear filter described in Chapter 4 was applied to each light curve prior to running the box-shaped transit finder described in Chapter 2. An alternative in terms of filtering would have been to use the optimal filter (after estimating the power spectrum by least-squares fitting of sine-curves), an option also described in Chapter 4. However, the good results obtained in tests of the iterative non-linear filter and its low computational requirements made it the filter of choice, at least for the initial detection stage. The filtered light curves are referred to as level 3 light curves.

We found that pre-smoothing, using a median filter with a width of 7 data points followed by a linear filter with a width of 3 data points, before running the iterative

Table 6.1: Short-period variables identified by sine-curve fitting.

ID	p (days)	Comment
249	3.903327	
	44.647	a
259	0.706374	
	0.710492	a
384	1.781245	
553	0.461317	
599	0.937934	b
809	1.601433	c
915	1.451470	c

Comments: (a) second period found after subtracting primary periodic component, (b) eclipsing binary with orbital period p (secondary eclipses visible in phase-folded light curve), (c) eclipsing binary with orbital period $2p$ (marked difference between odd- and even-numbered eclipses).

non-linear filter to construct the continuum to be subtracted from the data[1], improved the results of the filtering, minimising the amount of transit signal removed by the filter for a given trial duration d. The filtering process is illustrated in Figure 6.8.

Three trial transit durations were used: 25, 50 and 100 time steps, corresponding to 3.6, 7.1 and 14.2 hrs respectively. The last value is above the range of durations expected for most planetary transits, but was used as a check, as previous experience had shown that false alarms due to residual stellar micro-variability tend to lead to detections at the longest trial transit duration.

Performing the transit search on the 1001 light curves at each transit duration took $\sim$ 1 hr on a 2 GHz Dual-AMD Linux PC. For each light curve and trial duration, the following information was recorded: the best epoch e_S and detection statistic S_S for a single transit (maximum from the top panel of Figure 6.9), and the best period p, epoch e_M and detection statistic S_M for multiple transits (maxima from the middle and bottom panels respectively of Figure 6.9).

If the light curves had contained only transits and white Gaussian noise, a simple threshold in S_M (which is equal the signal-to-noise ratio of the combined transits) would have led to a given confidence level. Provided the filtering process was successful in all light curves, a simple threshold should also have been sufficient.

However, plots of S_M versus S_S for all light curves at a given transit duration (see e.g. Figure 6.10) show a large number of points with high S_S and $S_M \simeq S_S$. These correspond to light curves with a single transit-like event. Examination of these light curves showed no convincing transit candidates. It is also interesting to note that the tail increases in number and extends towards higher S_S and S_M for longer trial durations, and that most points in the tail corresponds to cases where the scatter

[1] See details of the iterative non-linear filter in Chapter 4 for more detailed explanation.

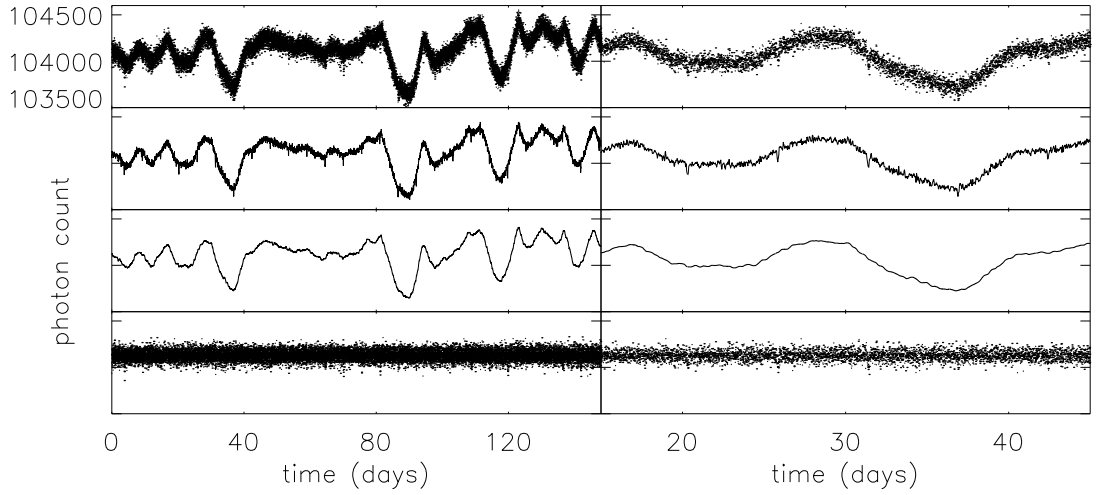

Figure 6.8: Example of iterative non-linear filtering with a trial duration of 3.5 hrs for light curve ID 34. Left column: entire light curve. Right column: detail. Top row: level 2 light curve. 2nd row: pre-smoothed light curve. 3rd row: continuum. Bottom row: level 3 (filtered) light curve. This light curve contained transits with a depth of 0.07% and a period of 5.52 d. They are clearly visible in the 2nd row, and are present in the 4th, though buried in the (now close to white) noise.

in the level 3 light curve was significantly higher than expected from simple photon counting statistics. This suggests that these points are due to incompletely filtered stellar micro-variability. Variations of stellar origin are mostly on timescales longer than a transit, so that shorter trial (hence filter) durations give better results. That was expected, but the large number of light curves still significantly contaminated by residual variability even at the shortest trial duration was not. A possible explanation for this is discussed in Section 6.4.3.1. For the purposes of the present exercise, it was necessary to devise a method to discriminate between real (periodic) transit candidates and these events.

A transit signal repeated $N_{\rm tr}$ times in a noise-free light curve would lead to $N_{\rm tr}$ identical peaks in the distribution of single transit signal as a function of trial epoch, and $S_M = \sqrt{N_{\rm tr}} \times S_S$. However, noise in the light curves and the discrete sampling in trial parameter space lead to variations in the single transit signal-to-noise ratio values at each trial epoch corresponding to a real transit, while S_S is taken to be the maximum of these values. One therefore expects S_M to be slightly smaller than $\sqrt{N_{\rm tr}} \times S_S$. Any light curve for which $S_M \geq a + b \times S_S$ were therefore selected as potentially containing two or more transits, with b = 1.3 (slightly smaller than $\sqrt{2}$). The constant a, taken to be 1.0, was designed to make the threshold more stringent for less significant events. All cases below a similar threshold line with b = 1.4 were flagged as low confidence level candidates. Any light curves with $S_S \geq 20$ were also

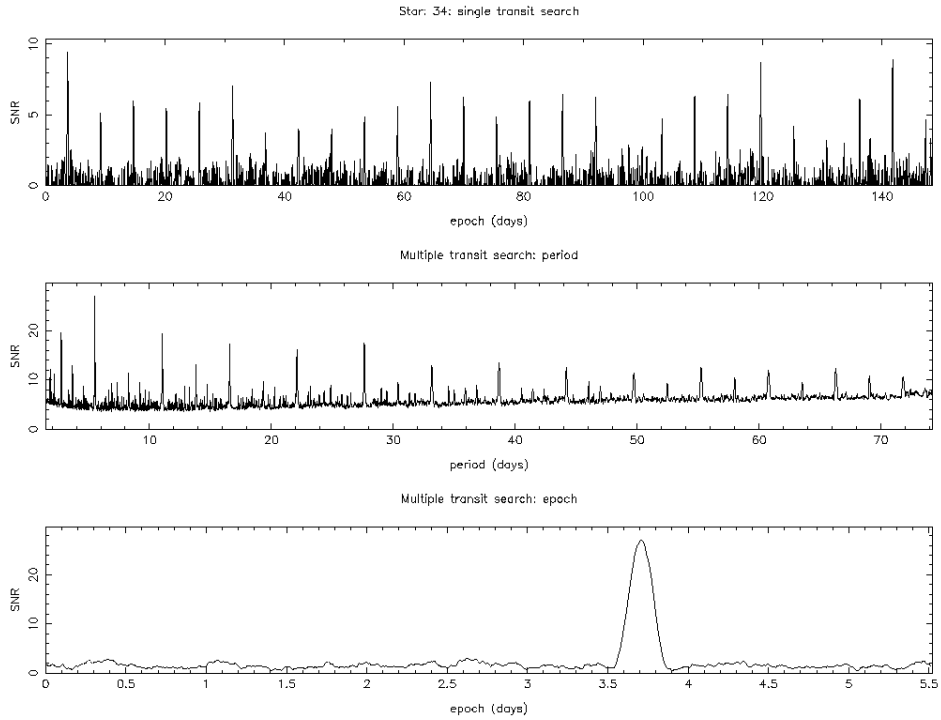

Figure 6.9: Example of the transit detections statistic distributions as a function of epoch e for a single event (top), as a function of period p for a multiple event (middle) and as a function of epoch at the best period (bottom)

selected as potentially containing single but significant transits (e.g. single transit of a long-period Jupiter-like planet). This automatic selection process yielded three initial candidate lists, one for each trial duration.

The level 2 and 3 light curves were then examined by eye in the transit vicinity, and obviously spurious detections eliminated from the list. The three lists were also merged at this stage to create a single candidate list.

6.3.1.5 Basic transit parameter estimation

Once the candidate list was 'weeded out' and merged, the basic transit parameters were estimated in the following way:

- The starting point was the level 3 light curve corresponding to the trial duration closest to the true duration as estimated by eye.
- A more accurate duration was computed by estimating the 'full width at half-minimum' of the transits as follows:
  - phase-folding the light curve at the best trial period;
  - rebinning it into bins lasting approximately one fifth of the current duration

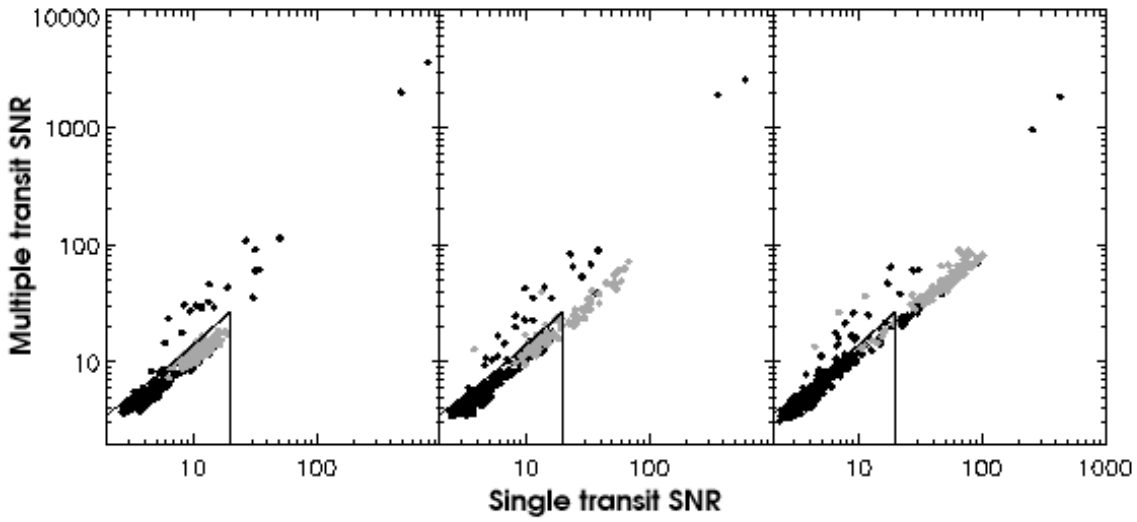

Figure 6.10: Multiple versus single transit detection statistic for trial durations of 3.5, 7.1 and 14.2 hrs (left, centre and right panels respectively). Gray symbols: measured scatter in filtered (level 3) light curves exceeds the expected scatter from photon statistics only.

estimate;

- finding the minimum in the rebinned phase-folded light curve;
- measuring the depth of the transit from that minimum value;
- finding the half-minimum points by moving outward from the minimum one point at a time;
- the new duration estimate is then given by the interval between the half-minimum points,

- A new level 3 light curve was generated by applying the non-linear filter to the level 2 light curve with the newly estimated duration.
- The duration was estimated again, also yielding a transit depth estimate and a refined epoch estimate.

A small number of very shallow transit candidates were eliminated at this last stage as convergence in the duration estimate failed to be reached.

The results, as summarised in Table 6.2, were then send to C. Moutou. The phase-folded level 3 light curves of all transit candidates (corresponding to the last iteration of the transit parameter estimation procedure) are shown an Appendix at the end of this chapter.

Table 6.2: Transit candidates and their parameters. NB: in cases where there was evidence (from visual inspection of the light curve) that the true period was a harmonic of the detected period, the suspected true period is given.

ID	N_{tr}	$\Delta F/F$ (%)	e (days)	p (days)	d (days)	Confidence	Comment
31	6	0.27	7.34	24.724	0.141	high	a
34	27	0.06	3.662	5.528	0.107	high	
85	6	0.9	0.175	26.427	0.175	high	
117	2	1.78	29.751	69.841	0.14	low	b
168	13	0.27	4.803	11.513	0.035	high	
207	2	1.5	41.45	88.486	0.279	high	c, d
249	48	28.06	2.729	3.903	0.105	high	
	3	11.69	36.957	44.657	0.21	high	e
259	52	0.3	0.327	1.414	0.14	high	a, f, g
276	2	1.54	28.564	66.985	0.142	low	h
375	6	0.04	10.024	24.905	0.175	low	i
384	–	–	–	1.781	–	high	f, j
386	9	0.66	10.073	17.117	0.035	high	
390	19	0.05	0.561	8.01	0.108	high	
406	6	0.05	10.025	24.905	0.175	low	i
460	5	0.54	2.718	32.932	0.489	high	
474	13	0.08	6.175	11.351	0.142	high	
483	4	0.03	1.606	48.643	0.176	low	i
486	31	0.08	1.893	4.83	0.07	high	a, k
533	23	1.34	2.887	6.407	0.209	high	
537	53	0.02	1.699	2.783	0.108	high	i, l
599	159	0.23	0.392	0.938	0.028	high	f, m
613	16	0.05	1.077	9.611	0.106	high	
624	22	0.05	6.709	3.238	0.142	high	
809	47	0.25	1.012	3.203	0.049	high	a, f
835	4	0.57	9.277	42.641	0.252	high	
915	51	28.51	2.84	2.903	0.105	high	a, n
917	5	0.11	25.459	30.428	0.213	high	
919	11	1.76	7.753	13.213	0.098	high	
937	18	0.35	4.932	8.459	0.105	high	
985	28	0.16	4.697	5.196	0.105	high	
1001	22	0.04	5.377	6.808	0.066	low	i

Comments: (a) eclipsing binary (secondary eclipses visible in phase-folded light curve), (b) very noisy light curve, (c) true period outside period search range, (d) detected period was half the true period, (e) triple system, second period was found by sine-fitting, (f) evidence of sinusoidal modulation, (g) transits exhibit phase shift with respect to the sinusoidal modulation, (h) light curve highly variable on short timescales, (i) detected transit very shallow, (j) no transits visible (other periodic behaviour), (k) detected period was 1.5 times the true period, (l) large number of transits improves confidence level, (m) detected period was 4 times the true period, (n) detected period was 2 times the true period.

6.3.2 Methods used by the other teams

The methods employed by the other teams for the light curve detrending and transit detection are briefly summarised below. In some cases, the information available at the time of writing is incomplete. For a fuller description see the upcoming article reporting on the exercise (Moutou et al. 2004a, in prep.) or the individual references for each method.

6.3.2.1 Team 1

A low-pass filter with a cutoff frequency between 0.059 and 0.117 d^{-1} was used to remove short-term variations (residual scattered light). This type of filter does not proceed to the edge of a dataset, and the first and last ten days of each light curve were removed before proceeding further. Local linear fits to small (0.5 to 3 d) sections of each light curve where then used to model and remove long-term stellar variations. The transit search was carried out with an algorithm described in Rauer et al. (2004), by which data points further than 3 σ from the light curve average are flagged and contiguous sequences of such data points within a certain range of durations are considered as transit candidates. After removal of obviously spurious candidates, a period search was run on the epochs of the highlighted sequences. Finally, detailed visual examination was used to determine transit parameters and detect signs of stellar binaries (secondary eclipses and out-of-eclipse variations).

6.3.2.2 Team 2

In the method used by team 2, transit detection proceeds before detrending. The transit detection is carried out by comparing, at a given epoch, a short sequence of data to a short reference sequence (model transit) by plotting one versus the other and measuring the width (rms of residuals from a linear fit) of the resulting cloud. The sequence of widths at each trial epoch forms the detection curve DC, on which the detrending is applied. Detrending proceeds by applying a model of the form $DC_{i,j} = s_{i,j} + \lambda_i p_j$, where i is the detection curve number, j is the data point index, s is the signal, p is a perturbing trend common to a subset of all DCs and λ a scaling factor for each DC. Starting with an initial guess for p, the λ_i's are found by projecting each DC onto p. A new guess for p is obtained by taking the group average of DC_i/λ_i over all i's, and the λ's are then recalculated. The DC with the highest scatter is taken as the first guess for p, and the resulting correction applied to those light curves which closely resemble it (high λ's), then the most noisy of the remaining light curves is taken as the next guess, etc... Detrended DCs are visually examined and those showing periodic signals selected as candidates.

6.3.2.3 Team 3

Team three modelled the scattered light variations as a sum of harmonics of the fundamental orbit frequency (5 harmonics included), and the long term stellar variations as a sum of harmonics of the frequency $1/2T$, using 200 harmonics (i.e. down to timescales of 1.5 d). The 411 parameter model (5 high frequency harmonics each with sine & cosine components, 200 low frequency harmonics and a constant level) is fitted to each light curve and subtracted. This is basically identical to the least-squares fitting procedure described in Section 4.3 for power-spectrum estimation of data with gaps, but only the frequencies corresponding to the noise sources to be removed are fitted and the result is subtracted – resulting in the equivalent of a band-pass filter. The fitting process is made less computationally expensive by considering only those time points common to all light curves and using a single SVD pseudo-inverse of the fitting matrix to perform all the fits. The BLS algorithm of Kovács et al. (2002) – which is very similar to our box-shaped transit finder – was then used for transit detection. It is interesting to note (see Section 6.4.3.2), that a fit of the form $a + b/f$ where f is the trial frequency was subtracted from the distribution of their SR statistic with frequency, which is the equivalent in the BLS of the distribution of our S_M statistic with trial period. Light curves yielding a peak SR statistic greater than 7 were taken as candidates.

6.3.2.4 Team 4

Team 4 used a filter inspired from image processing techniques, the Gauging filter (Guis & Barge 2004, in prep.), for detrending. This filter treats the plot of a light curve as two 2-D images: one is black above and white below the light curve, and the other vice versa. Each image is undersampled and smoothed with a 2-D boxcar filter, the results for the two images are averaged and subtracted from the original light curve. Transit detection then proceeds with a standard matched filter, candidates being those light curves with periodic peaks in the distribution of correlation versus trial epoch, or those with one or two peaks deemed significant (according to criteria based on the entropy and variance of the peak distribution, see Guis & Barge 2004 for details).

6.4 Discussion of the results

The results of the exercise are summarised in Tables 6.3, 6.4 and 6.5. The first two of these tables list all the transit and stellar events inserted in the light curves together with relevant parameters and whether they were or weren't detected by

each group. The last table contains the false alarm and missed detection rates achieved by each group for transit and transit-like (eclipsing binary) events, and the results obtained by combining the candidates of all the groups.

Table 6.3: Characteristics of the transits that were inserted in the light curves. LD*: stellar limb darkening coefficient. i: orbital inclination. a: semi-major axis. V: apparent visual magnitude. Detection flag: each sign corresponds, from left to right, to teams 1 to 5. +: the event was detected by the team in question. –: the event was missed. Contents provided by C. Moutou.*

ID	$R_\star$ ($R_\odot$)	LD	$R_{\rm pl}$ ($R_\odot$)	p (days)	i (°)	a ($R_\odot$)	V	*	Detection flag
34	0.92	0.6	0.025	5.52	91.2	12.77	13		+ + + + +
85	1.1	0.4	0.099	26.4	88.9	37.876	15	a	+ + + + +
168	0.92	0.5	0.13	11.5	87.4	20.827	15		– – + + +
207	0.92	0.5	0.11	88.4	90	79.89	16		+ + + + +
317	1.1	0.6	0.02	33.8	90.5	44.66	12		– – – – –
326	0.85	0.6	0.017	6.8	89.9	13.9	14		– – – – –
390	0.92	0.6	0.022	8.0	91	16.35	12		+ + + + +
460	1.1	0.3	0.076	32.9	89.52	23.49	15	b	+ + + + +
474	0.92	0.6	0.028	11.34	91	20.63	13		+ + + + +
533	0.92	0.7	0.095	6.4	90	7.89	16		+ + + + +
537	0.85	0.6	0.015	2.78	90.1	7.68	12		– – + – +
575	0.85	0.6	0.019	15.9	90	24.57	14		– – – – –
613	1.1	0.6	0.026	4.8	89.4	12.16	14		– + + – +
618	1.3	0.6	0.023	8.48	91	19.55	12		– – – – –
624	1.1	0.6	0.029	6.7	90.2	15.18	14		– + + + +
681	1.1	0.6	0.023	19.8	91.4	31.27	13		– – – – –
715	1.3	0.3	0.098	10.1	86.4	21.96	15		– – – – –
			0.07	63.8	89.7	75		c	– – – – –
835	1.1	0.4	0.084	42.6	89.3	52.1	15		+ + + + +
915	1.5	0.25	0.13	58.32	89.9	70	15		+ – + – –
		0.3	1.1	2.9	86	11.4		d	+ + + + +
917	0.85	0.6	0.028	30.4	90.3	37.8	13		+ + + + +
1000	1.1	0.6	0.02	33.8	90.5	44.66	14	e	– – – – –
1001	1.1	0.6	0.02	33.8	90.5	44.66	13	e	– – – – +

* (a) planet + ring, (b) planet + moon, (c) 2 planets, (d) circumbinary planet, (e) 32 s sampling rate.

6.4.1 Relative performances of the various groups

Teams 3 & 5 – the Geneva Observatory team and ourselves – detected significantly more transit events than the other groups. In fact, taking the overlap of the results of teams 3 and 5 – i.e. accepting a detection only if it was made by both teams – gives as good a result as taking the overlap of the result of all the teams (column B in Table 6.5). The transit search algorithms used by both teams were very similar: team 3 used the BLS method of Kovács et al. (2002), which is mathematically identi-

Table 6.4: Characteristics of the stellar events that were inserted in the light curves. BEB*: background eclipsing binary.* GrB*: grazing binary. An* x *in the detection flag column means the event was detected but incorrectly identified as a transit. Contents provided by C. Moutou.*

ID	V	Type	p (days)	$\Delta F/F$	Detection flag
31	14	BEB	24.7	0.003	+ + + + +
131	14	δ Cepheid	5.86	–	– – – + –
249	14	triple star	3.9	–	+ + + + +
259	16	GrB	1.4132	–	– + + + +
271	15	Z Cam	–	–	– – – – –
384	15	β Cephei	0.2835	0.001	– + – + +
386	15	GrB	17.1	–	x x x x x
486	15	BEB	2.4128	0.001	– – x – +
518	15	GrB	78.3	–	– – – – –
553	15	δ Scuti	0.07342	0.003	– – + + +
599	15	GrB	1.874	–	– + + x +
650	14	semi regular	–	–	– – – – –
809	15	GrB	3.2	–	– x + + +
919	16	GrB	13.2	–	+ x + x x
937	15	BEB	8.452	0.001	x x x x x
985	15	BEB	5.19	0.001	x x x x x

Table 6.5: Individual and combined performance of the detection teams. In columns A*,* B *&* C*, which refer to detections made simultaneously by 1, 2 and 3 teams respectively, incorrect identifications of a stellar event count toward true detections if half of the teams that detected the event or more identified it correctly.*

Team	1	2	3	4	5	A	B	C
Planetary transits (23 inserted)								
True detections	10	11	14	11	14	16	15	13
Missed detections	13	12	9	12	9	7	8	10
False alarms	1	0	0	3	5	9	0	0
Stellar events (17 inserted)								
True detections	4	6	8	8	9	10	9	8
Incorrect identifications	3	5	4	5	4	4	4	4
Missed detections	10	6	5	4	4	3	4	5
False alarms	0	0	0	0	0	0	0	0
All periodic transit-like events (34 inserted)								
True detections	16	21	24	21	25	25	24	22
Missed detections	18	13	10	13	9	9	10	12
False alarms	1	0	0	3	5	9	0	0

cal to our box-shaped transit-finder, except that the light curve is phase-folded and rebinned before computing the detection statistic. This suggests that transit search algorithms based on least-squares minimisation with box-shaped transit models are a good choice. The other three teams used a variety of methods, some inspired by image processing techniques, some standard matched filter detectors. All have the common characteristic of being more complex than the methods used by teams 3 and 5. Simplicity therefore appears to be the best guarantee of robustness.

However, team 3 had no false alarms, whereas we had 5. The difference in false alarm rate between teams 3 and 5 may be due to the variability filtering and/or candidate selection methods. The fact that all our false alarms were low confidence candidates lends credibility to the latter hypothesis. The only low confidence candidate that wasn't a false alarm was light curve 1001. In that case the detected period was not the true period (i.e. the detected events were not the transits), so that it can, in a way, be considered to have been a false alarm.

6.4.2 Overall results and implications

A very reassuring result is the absence of overlap between the false alarms due to the various teams. By contrast, there is some overlap on all but one of the true detections – the only true detection made by one team only being light curve 1001, which was one of our low confidence detections. This bodes well for the exploitation of COROT light curves by teams who work in a concerted fashion, but rely on different methods, and suggests that internal exchange of candidate lists is advisable before any results are announced.

On the other hand, one might consider the fact that several of the detected eclipsing binaries were incorrectly identified as transits by most of the teams that detected them rather worrying. A total of 34 'transit-like events' – 23 planetary transits and 11 background or grazing eclipsing binaries – were inserted in the light curves. Of these (counting only detections made by at least two teams simultaneously), 15 were correctly identified as planetary transits, 9 were correctly identified as stellar eclipses, 12 were not detected, and 4 were detected but incorrectly identified as planetary transits by more than half of the teams that detected them.

The fraction of light curves containing stellar binary eclipses was more or less realistic, but the fraction of light curves containing planetary transits was approximately ten times what is expected (from extrapolation of the discovery rates of present RV surveys). If the trends from the present exercise are extrapolated to a more realistic sample, we would expect, for (say) 10 000 light curves, $\sim$ 20 planetary transits and $\sim$ 130 low-depth stellar eclipses to be detectable; leading to $\sim$ 150 transit candidates of which $\sim$ 130 would be stellar but only $\sim$ 90 identified as such,

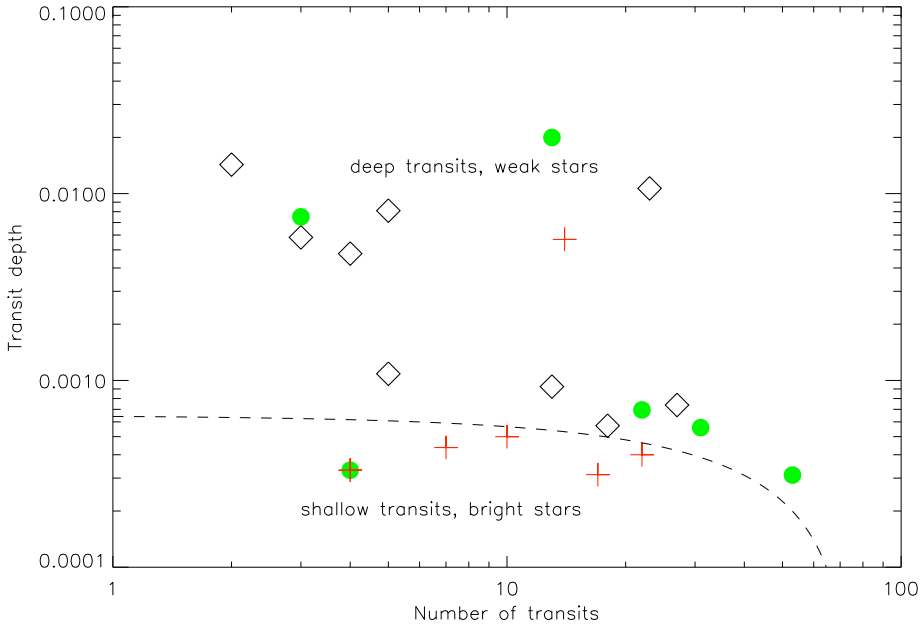

Figure 6.11: Depth $\Delta F/F$ versus number N_{tr} for all inserted planetary transits. Plus signs: non-detected events. Diamonds: events detected by all teams. Circles: events detected by 1 to 4 teams. Dashed line: estimated COROT detectability limit, $\Delta F/F \simeq -8.5 \cdot 10^{-6}\, N_{tr} + 6.5 \cdot 10^{-4}$. The superimposed plus sign and circle corresponds to three events: two undetected (IDs 317 and 1000) and one detected by 1 team only (ID 1001). The plus sign above the line corresponds to a very variable light curve (ID 715). Plot provided by C. Moutou.

leaving $\sim$ 60 planetary transit candidates: a false alarm rate of over 65%! Such a high false alarm rate would not be acceptable in view of the time-consuming and difficult follow-up observations that are needed to weed them out. However, several points of COROT's observation strategy which are specifically aimed at reducing this type of false alarm were not included in the present exercise. It would also have been possible to correctly identify some of the eclipsing binaries from a more detailed examination of their light curves, but most teams chose to focus on the detection process in the limited time available. The false alarms rate due to background or grazing eclipsing binaries would therefore not necessary be as high as this exercise suggests.

C. Moutou used the results of the exercise to estimate detectability limits for COROT, by plotting the transit depth $\Delta F/F$ versus the number of transits N_{tr} (Figure 6.11) and drawing an empirical line, corresponding to a linear relation between p and N_{tr}, that separates events detected by more than one team from those that went undetected or were detected by one team only. The minimum detectable radii are summarised in Table 6.6 as a function of period for different stellar radii (i.e. spectral types).

Table 6.6: Estimated minimum detectable planet radius as a function of period p or number of transits N_{tr}. R_{pl} is expressed in units of the stellar radius $R_\star$, or in units of Earth radii $R_\oplus$ for different stellar radii, corresponding to a given spectral type.

p (days)	N_{tr}	Minimum R_{pl}				
		fractional ($R_\star$)	F0V star ($R_\oplus$)	G0V star ($R_\oplus$)	G2V star ($R_\oplus$)	K0V star ($R_\oplus$)
75	2	0.0252	4.12	3.01	2.83	2.33
50	3	0.0250	4.08	3.00	2.81	2.31
30	5	0.0246	4.03	2.95	2.76	2.28
15	10	0.0238	3.89	2.85	2.67	2.20
10	15	0.0229	3.74	2.74	2.57	2.12
3	50	0.0150	2.45	1.80	1.68	1.39

These limits are slightly more pessimistic than those suggested by the results of the simulations described in Chapter 5 (Section 5.3.3). The latter implied a detection limit between 1.5 and 2.0 $R_\oplus$ for a period of 30 d and a G2V parent star, while the corresponding limiting radius, as given by the line shown on Figure 6.11, is 2.76 $R_\oplus$. This difference can probably be explained by the more complete treatment of the instrumental noise in the present exercise. The method adopted by A. Lanza to simulate short-timescale stellar micro-variability is also more pessimistic than ours, as they effectively apply a scaling factor of 3 while we apply none.

6.4.3 Lessons learnt

6.4.3.1 Simulated light curves

A possible explanation for the residual micro-variability which, in a number of light curves, caused our algorithm to produce high single transit detection statistics (and multiple transit statistics barely higher than the single transit values) can be arrived at by noting that a number of the level 2 light curves contained remarkably similar transit-like variations (i.e. relatively sharp dips lasting a few hours). As shown in Figure 6.12, the resemblance between different light curves (during but also around the dip), together with the lack of repeat events in a given light curve, suggest that these are artifacts of the simulated stellar micro-variability components rather than inserted transit events. The dips are slightly distorted and shifted from one-another – presumably due to the scaling and shifting applied to the stellar micro-variability components to generate a sufficient number of non-identical light curves – but the similarities are striking. It is our opinion that this is due to a problem with the use by

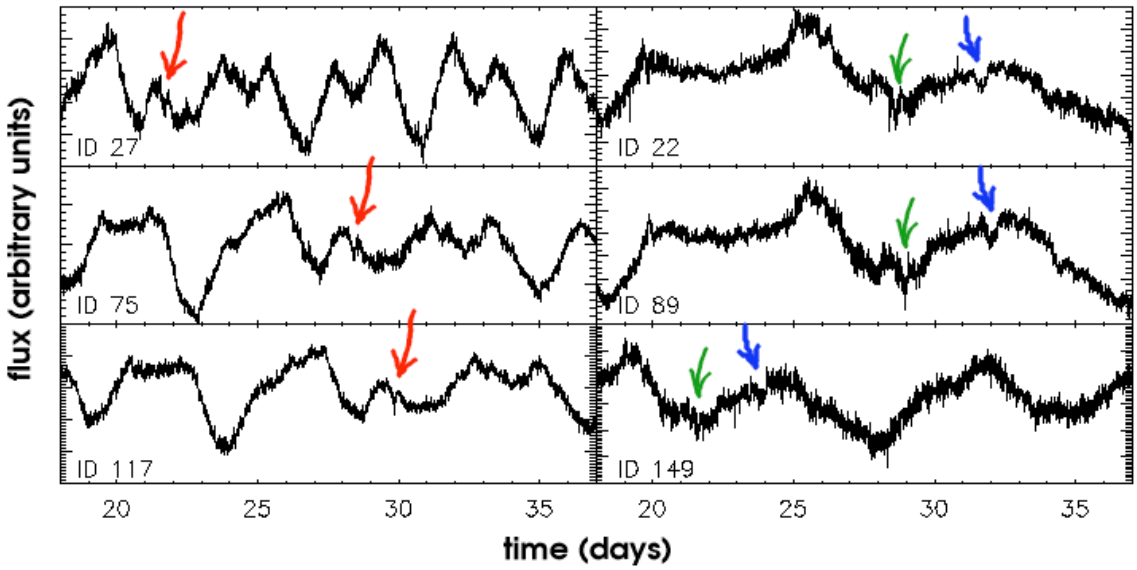

Figure 6.12: Two examples (one per column) of recurrent transit-like features in similar looking light curves. All panels have the same time range. Arrows of the same colour on different panels point to the same transit like dips, which occur at different times due to the shifting and scaling applied to the stellar micro-variability light curves. Almost all the highly variable light curves in the sample closely resemble one of these two examples. Note that ID 117 is one of our false alarms.

A. Lanza et al. of scaled residuals to simulate the short-term variability component of their simulated light curves. This procedure is likely to amplify any artifacts in the residuals resulting from imperfections of the rotational modulation fit to the solar irradiance variations, and results in transit-like dips which we believe are not realistic. As these dips last only a few hours, they were not removed by the filtering process, and therefore led to a number of spurious detections.

6.4.3.2 Performance of our filtering and transit detection method

Our team obtained the largest number of correct detections. We also reported 5 'false alarms', although all of these were low confidence detections. The team whose results in terms of sensitivity and robustness can be considered the best is probably team 3, and our results are identical to theirs if the low confidence detections are discounted. It is interesting to discuss the similarities and differences in the methods used by the two teams.

The least squares fitting method used by team 3 is equivalent to a bandpass filter, where the low-frequency continuum and the high-frequency noise are reconstructed by fitting sine curves up to a limiting frequency, then subtracted from the original light curve. In theory, there is no reason to expect this method to be more

capable of discriminating between stellar and transit signals than the iterative non-linear filter we used. The transit search method they used was very similar to ours, so that one wouldn't expect significant differences to arise from that step in the detection process. On the other hand, their candidate selection was based on a single thresold in their SR detection statistic.

Such an approach was not adopted in our case because, as discussed previously, a single threshold applied to the maximum of the distribution of our S_M detection statistic as a function of period gave unsatisfactory results. However, comparison of the two methods brought to our attention the presence of a systematic trend in the distribution of the S_M statistic with period, which is clearly visible in the middle panel of Figure 6.9. This trend is due to the change in the number of independent models tested with period. Among other parameters, the number of trial epochs and the number of in-transit points change with trial period. These effects are less pronounced in the BLS method (though a correction for this effect was applied by Team 3), in which the light curve is phase-folded at the trial period and rebinned into a fixed number of bins. However, that approach gives even statistical weight to each bin, rather than to each data point, which is not optimal, particularly in the presence of irregular sampling or gaps. The number of period cycles in the full light curve also changes with trial period, leading (for strong periodic signals) to discrete steps in the distribution of multiple transit detection statistic with trial period at harmonics of the true period.

It is difficult to derive a theoretical correction for the overall trend observed, because the value plotted for each trial period is itself the result of finding the maximum in the distribution of detection statistics over all trial epochs. Theoretically computing how the trend induced by the varying number of independent tests propagates through this maximum finding process would be prohibitively complex. We therefore experimented with an empirically derived weighting scheme, dividing the value of the detection statistic obtained for each trial period by a factor which scales with the 0.1^{th} power of the trial period (normalised to the middle of the trial period range). This leads to a significant reduction of the trend, and was therefore adopted. This modification alone removes some of the false alarms.

It is also possible to improve our candidate selection by automatically determining the best trial duration, taking into account the ratio R of the scatter measured in the level 3 light curve to that expected from photon statistics (based on the median photon count), and by selecting candidates not only on the basis of the multiple and single transit detection statistics themselves, but also on the significance of the maximum in the distribution of S_M versus trial period. After a process of optimisation by trial and error, where various forms of threshold were experimented

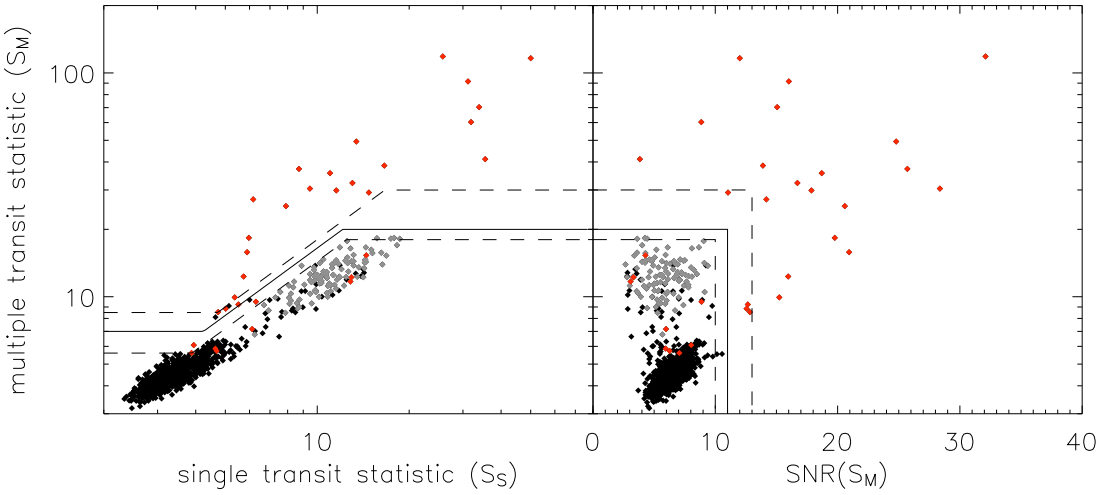

Figure 6.13: Plots of S_M versus S_S (left) and S_M versus SNR(S_M) *(right) after slight modifications to the transit search algorithm and automatic duration selection. Grey: light curves with $R > 1$. Red: Light curves actually containing planetary transits or transit-like stellar events. Solid line: main detection threshold. Dashed lines: upper (high confidence) and lower (low confidence) thresholds.*

with, the following procedure seems to give the best results:

- A single trial transit duration (0.18 d) is used in the filtering process, but three trial durations (0.18, 0.27 and 0.35 d) are used for the transit search. The decision to take this approach was based on the fact that almost all the candidates which were selected on the basis of the two higher trial durations and not the lower in the initial search were either removed at the individual light curve examination stage or were false alarms.
- For each light curve, the optimum trial duration is chosen in the following way:
  - If $R > 1$, which indicates residual stellar variability in the filtered light curve, the best trial duration is taken to be the shortest, which is the one that was used for the filtering, as using a longer filter duration would only lead to more residual stellar variability and hence a higher probability of false detection.
  - If $R < 1$, the best trial duration is taken to be that which leads to the highest ratio of multiple to single transit detection statistic S_M/S_S.
- The candidate selection is then based on the two plots shown in Figure 6.13: S_M versus S_S, but also S_M versus the 'signal-to-noise ratio' of the multiple transit detection statistic, SNR(S_M), that is the ratio of $S_M - \langle S_M \rangle$ to $\sigma(S_M)$, where $\langle S_M \rangle$ is the median of the distribution of S_M with trial period and $\sigma(S_M)$ is the scatter of

this distribution.

- Rather than choosing selection thresholds on the basis of any theoretical consideration, they were chosen on the basis of the distribution of the points on the two plots themselves:
  - The plots were generated for all light curves in a 'blind' fashion, that is without showing which points corresponded to light curves actually containing transits, highlighting only those points for which $R > 1$ in grey.
  - Threshold lines were placed where a clear separation between the bulk of the points and the outliers was visible (solid lines on Figure 6.13). In the S_M versus S_S plot, this results in a line defined by: ($S_M > 7$ AND $S_M > 1.65 \times S_S$) OR $S_M > 20$. In the S_M versus SNR(S_M) plot, the line is defined by: $S_M > 20$ OR SNR(S_M) > 11.
  - Slightly higher and lower threshold lines, designed to respectively exclude and include a few more points in each case, were also chosen (dashed lines on Figure 6.13).
  - Selection then proceeds according to a points system: one point for a point above the lower dashed line, two for a point above the solid line, and three for a point above the upper dashed line in each plot. This results in points scores between 0 and 6, and candidates are light curves with scores of 3 and above. This procedures enables one to probe slightly into the bulk of the points in one of the two plots provided the point corresponding to the same light curve is well separated from the bulk in the other plot.

The above procedure is almost automatic – the user must only choose the position of the threshold lines, but only on the basis of the distribution of the points on the two scatter plots rather than by examining light curves. It leads to zero false alarms and 10 missed detections (counting both planetary transits and transit-like stellar events). The missed detections are the same as before plus ID 460 which was previously detected but is now missed – it is a long duration event.

A number of light curves with scores of 1 or 2 have best durations other than the filter duration used. In those cases, we tried repeating the transit search using the best duration as the filter duration, then applying the same selection procedure. This yields two more detections – IDs 460 and 317 – but also two false alarms – IDs 309 and 995.

Modifications in the transit search and candidate selection procedure therefore allow us to obtain 8 missed detections and two false alarms without having to examine any light curves individually. We cannot objectively say whether the two false alarms would have been excluded at the light curve examination stage as we

already know the content of the light curves and could not reproduce the process in an unbiased way.

6.4.3.3 Parameter estimation and rejection of stellar events

The method used for parameter estimation was very basic. It was also made difficult by the need to make a guess of the transit duration to filter the light curve before the parameters could be measured. In future, methods to filter the light curve without a prior guess of the duration (but given knowledge of the period) will be investigated. The transit parameters will be measured in a more optimal way, e.g. with a matched filter applied to the phase folded-light curve, using the same type of transit modelling software to generate model transit light curves as was used to generate the signals inserted in the light curves for the present exercise.

Can careful consideration of the shape of a suspected planetary transit help reduce the contamination by stellar events? We consider three categories of stellar eclipses that can mimic planetary transits: grazing, high mass ratio, and blended (the light of a third star aligned with the eclipsing system dilutes the eclipses). The first and some of the second may be identified by relating the transit observables (depth $\Delta F/F$, total duration d_1, duration of totality d_2, period p) to the geometry of the system (as illustrated on Figure 1.1).

It is immediately apparent, for example, that the lack of a flat-bottom to the eclipse ($d_2 = 0$) is a tell-tale sign of a grazing event. This is the case for light curve IDs 249, 259, 599, 809 and 915 (which were identified as stellar events due to the presence of secondary eclipses or out-of-eclipse variations), but also of IDs 386 and 919 (which we reported as candidate planetary transits when they were in fact grazing eclipsing binaries). Of course, a planetary transit can also be grazing, but grazing planetary transits are even shallower and rarer than non-grazing ones, which makes it all the more unlikely that one would have been observed, let alone detected.

High mass ratio events can be identified by computing the star's density from the light curve, as discussed by Seager & Mallén-Ornelas (2003). In Chapter 1 we derived the dependence of the transit depth (Equation 1.6), the full transit duration (Equation 1.11) and the duration of totality (Equation 1.12) on the star radius $R_\star$, the planet radius R_p, the period p, the orbital distance a and the inclination i of the system, assuming that the orbit is circular, that the planet is dark, that there is only one star – the eclipsed object – in the system, and that there is a flat bottom to the eclipse – ie the impact parameter $b \leq R_\star$. In this framework, Seager & Mallén-Ornelas (2003) show that, these equations can be used to derive $R_\mathrm{p}/R_\star$, $b/R_\star$ and $a/R_\star$ from the light curve observables ($\Delta F/F$, d_1, d_2 and p) alone, and, using them in combination with Kepler's third law (Equation 1.4), the star density $\rho_\star/\rho_\odot$ (under

the additional assumption that $M_p \ll M_\star$). In theory, we should thus be able to identify eclipsing binaries where the primary is a giant star, without a spectrum or multi-colour information (provided the errors on the observables are low enough, but the high time sampling and photometric precision of COROT should allow this). This method cannot, however, be used to exclude cases where the primary is a main sequence star and the secondary a brown dwarf, as both the density of the eclipsed object and the radius of the transiting object will be consistent with a planet orbiting a main-sequence star.

The unique solution to the transit equations discussed above breaks down in the case of a blend, i.e. if light emitted by third object, spatially coincident or fortuitously aligned, dilutes the observed eclipses. In such cases, any secondary eclipses may be diluted to the point of being undetectable. Other observations than the single bandpass light curve alone are thus needed to identify these events as stellar rather than planetary. This is the case for the remaining two light curves which we incorrectly identified as candidate planetary transits, IDs 937 and 985.

6.5 Future prospects for COROT blind exercises

The present exercise was considered extremely useful by all participants, and a second generation exercise is foreseen for early 2005. Improvements will hopefully include the inclusion of colour information and improvements in the simulation of stellar micro-variability, based among other things on the soon to be released MOST data. It would also be desirable to simulate an entire COROT field, with stellar distributions and crowding derived from galactic models and the results of preparatory ground-based observations. It is foreseen that this second exercise will have a higher profile, groups outside the COROT EWG (e.g. the *Kepler* team) being invited to participate. More emphasis will also be devoted to the use of oversampling and to the determination of transit parameters after detection.

In the mean time the light curves from the present exercise will be made available to the entire transit search community, while the teams which participated will continue to attempt to improve their algorithms now that the content of the light curves is known.

From our point of view, this exercise also highlighted areas where future efforts should be concentrated. There is room for improvement in the variability filtering and transit candidate selection procedures. The pre-processing steps (in particular the removal of systematic trends) and variable star search tools, of which only the very basic lines were laid out here, will also play an important part in maximising the scientific return from COROT and other transit search data, and they need to be

developed further.

Appendix: Phase-folded light curve plots

Figures 6.14 to 6.21 contain plots of the normalised phase folded light curves for all the transit candidates reported by the IoA team. There is one row per transit candidate. In each case the left panel shows the full phase-folded light curve, the right panel the phase folded light curve around the transit. The blue line in the right panel is the binned phase-folded light curve from which the transit parameters are computed. The red line in both panel shows the box-shaped transit model corresponding to the reported parameters. The light curve ID is indicated in the top of the left panel. The period and input epoch (in days), as well as the filtering duration (in number of time steps and in days), are listed at the bottom of the left panel. The measured duration (in number of time steps and in days), epoch (in days) and transit depth (in %) are listed at the bottom of the right panel.

In the cases where sinusoidal variations corresponding to the period of the transits were visible (IDs 259, 384, 553, 599 & 809), the phase folded level 2 light curve is shown in a single panel below the two panels showing the level 3 light curve. The two rows for light curve 249 correspond to the two sets of eclipses with different periods (star 249 turns out to be a triple system). Light curves 384 and 553 showed only sinusoidal variations and no transits, so that only the phase folded level 2 light curve is shown. The plots for star 259 clearly show the shift in phase of the transit relative to the sinusoidal variations. This is probably due to a small error in the detected period.

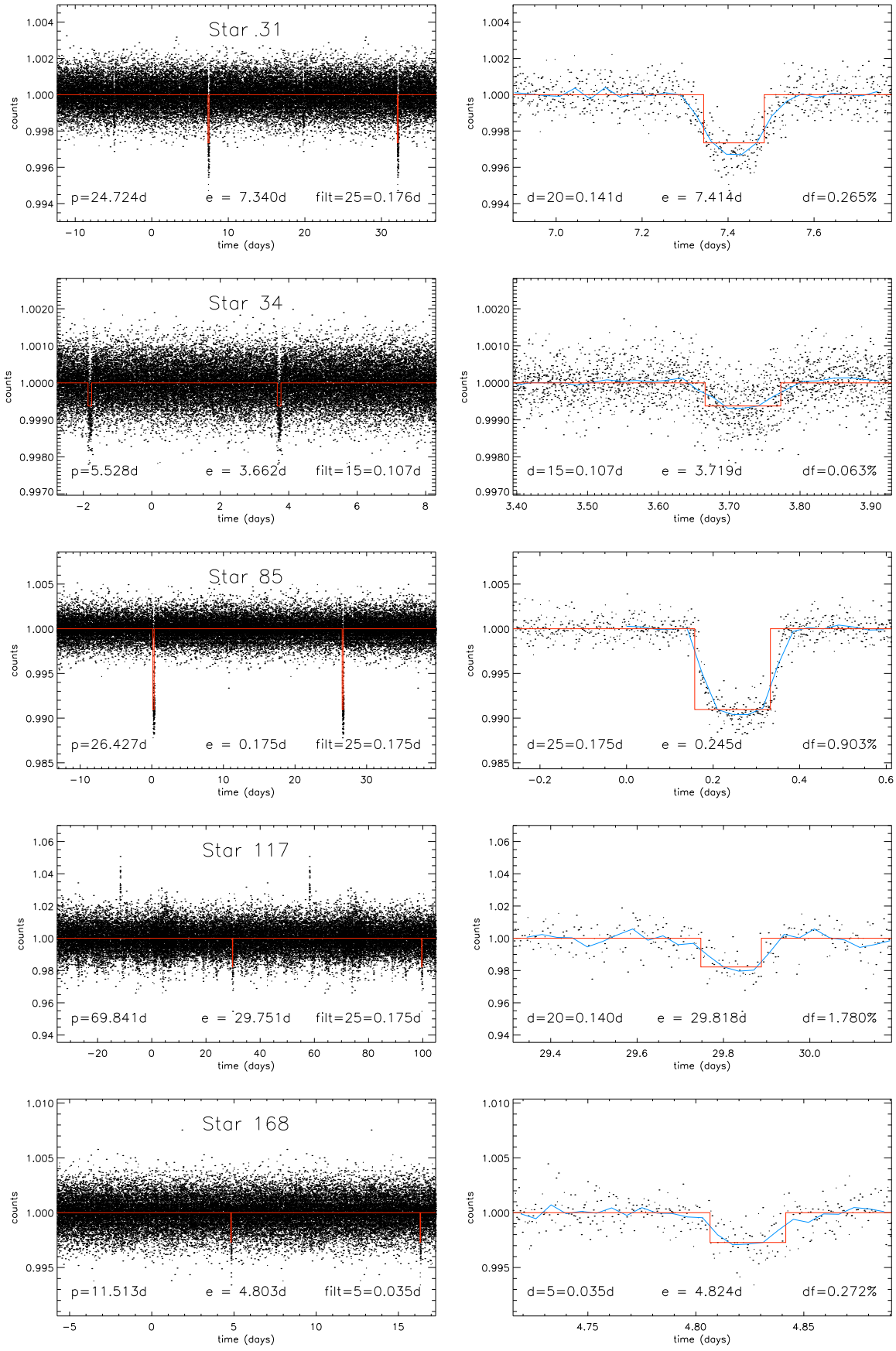

Figure 6.14: Light curves of transit / variable star candidates – page 1.

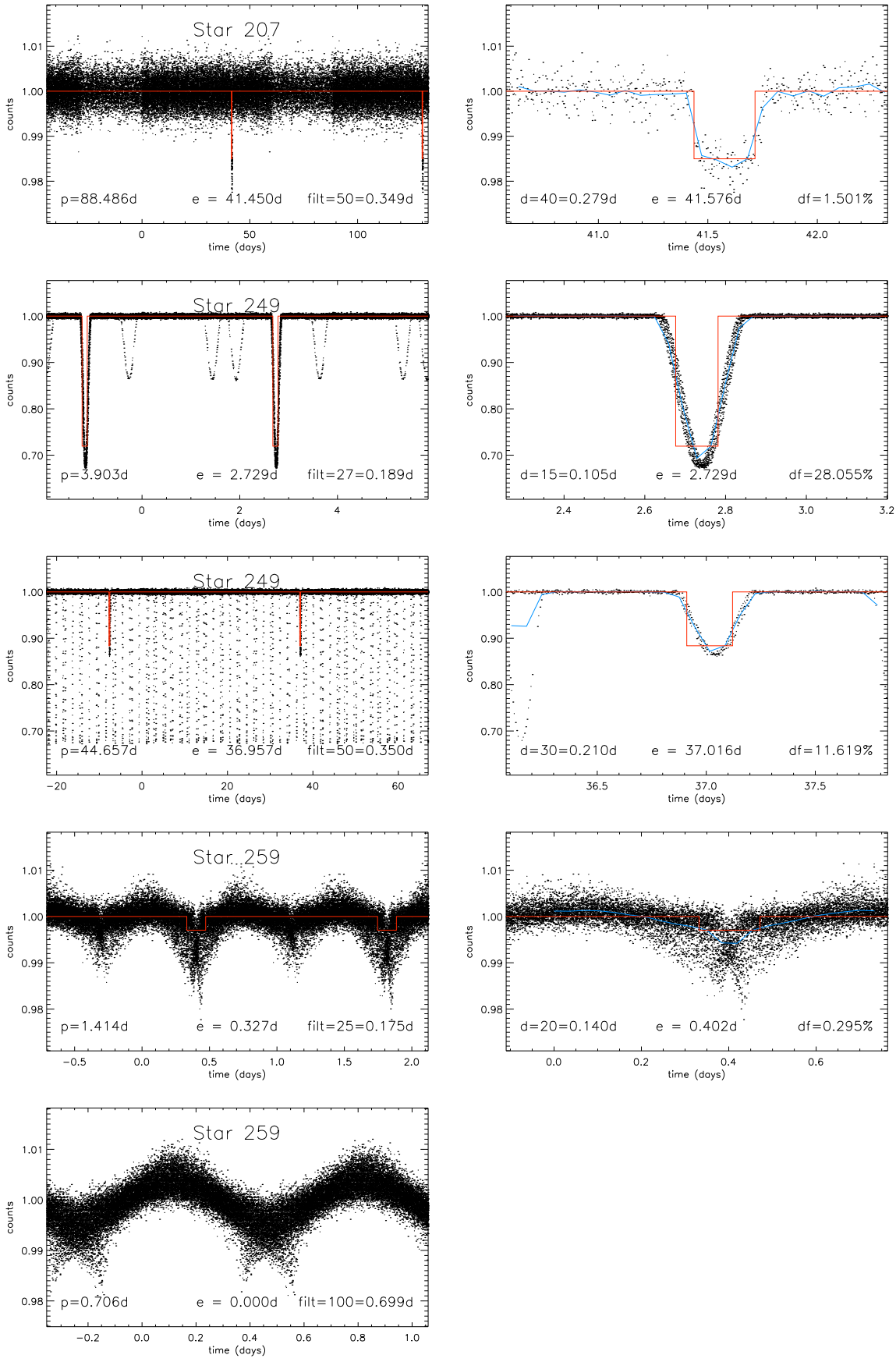

Figure 6.15: Light curves of transit / variable star candidates – page 2.

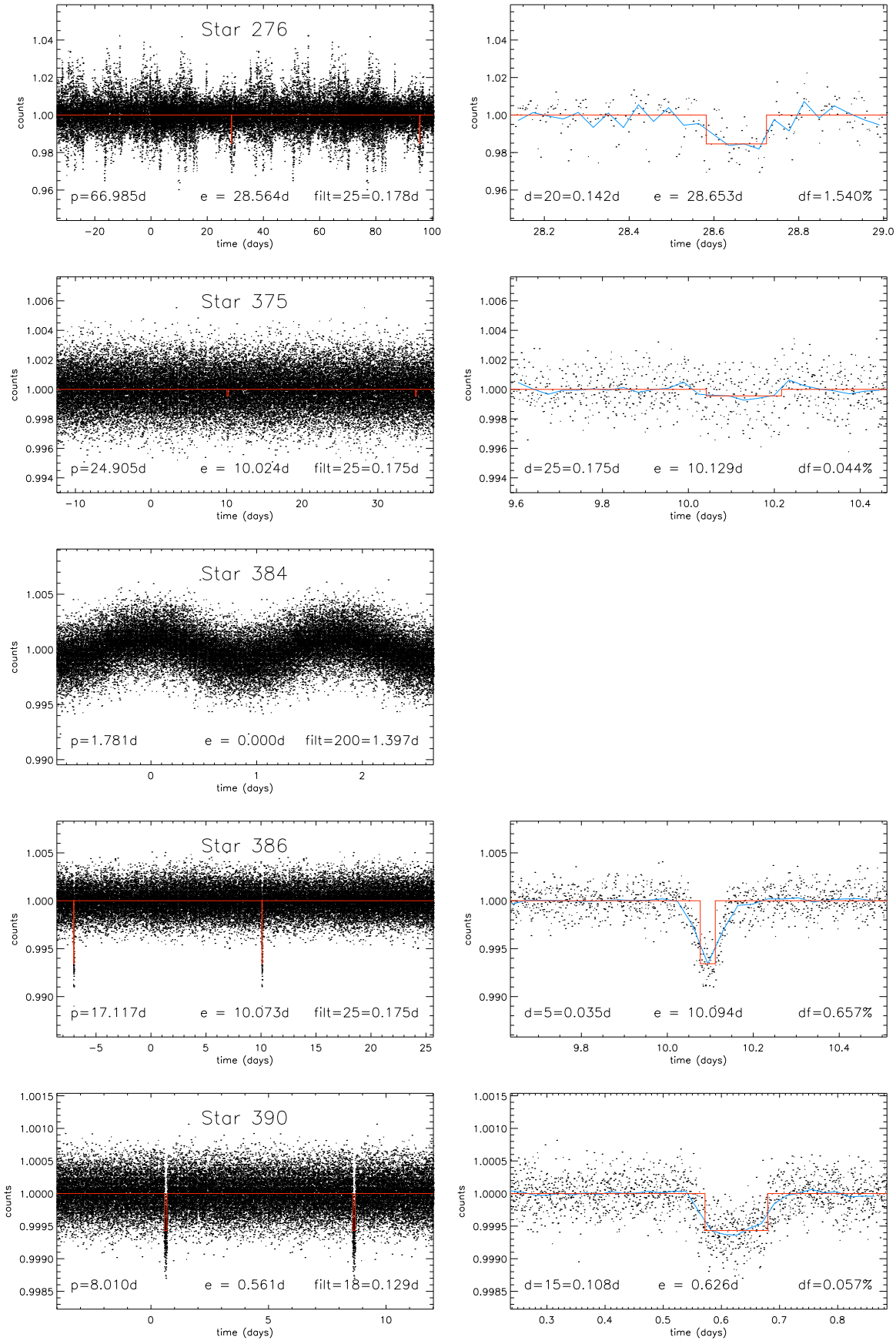

Figure 6.16: Light curves of transit / variable star candidates – page 3.

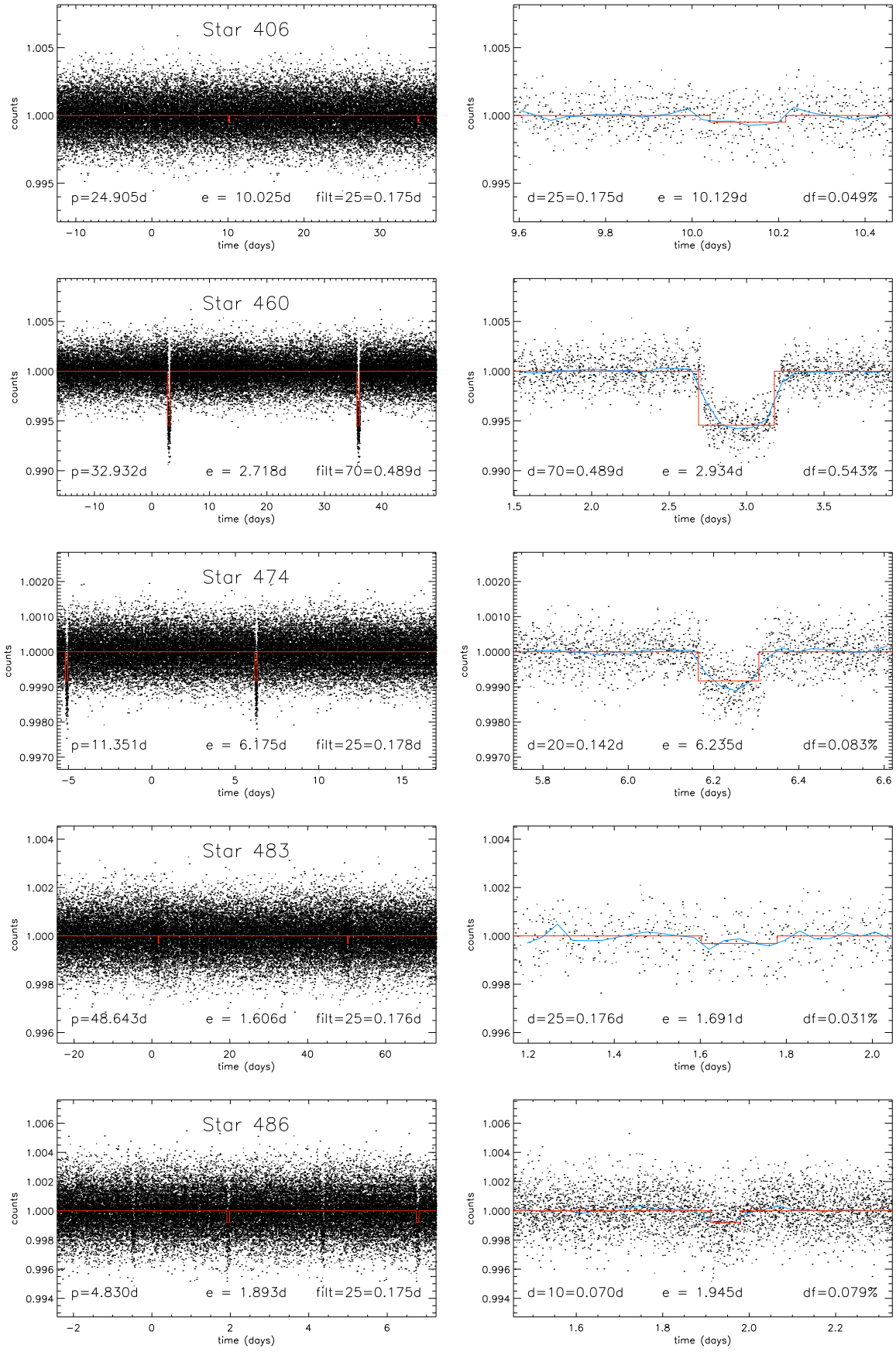

Figure 6.17: Light curves of transit / variable star candidates – page 4.

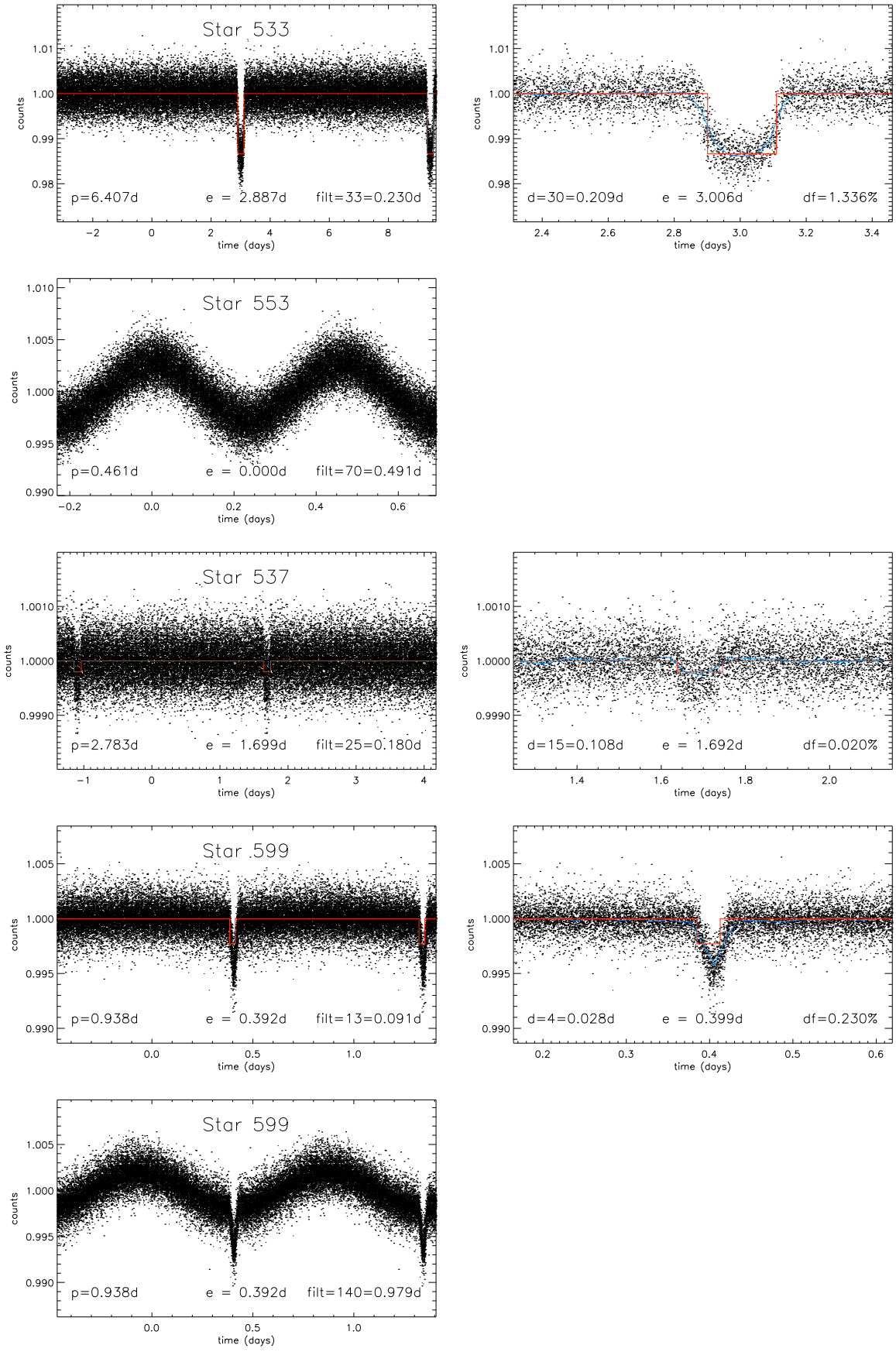

Figure 6.18: Light curves of transit / variable star candidates – page 5.

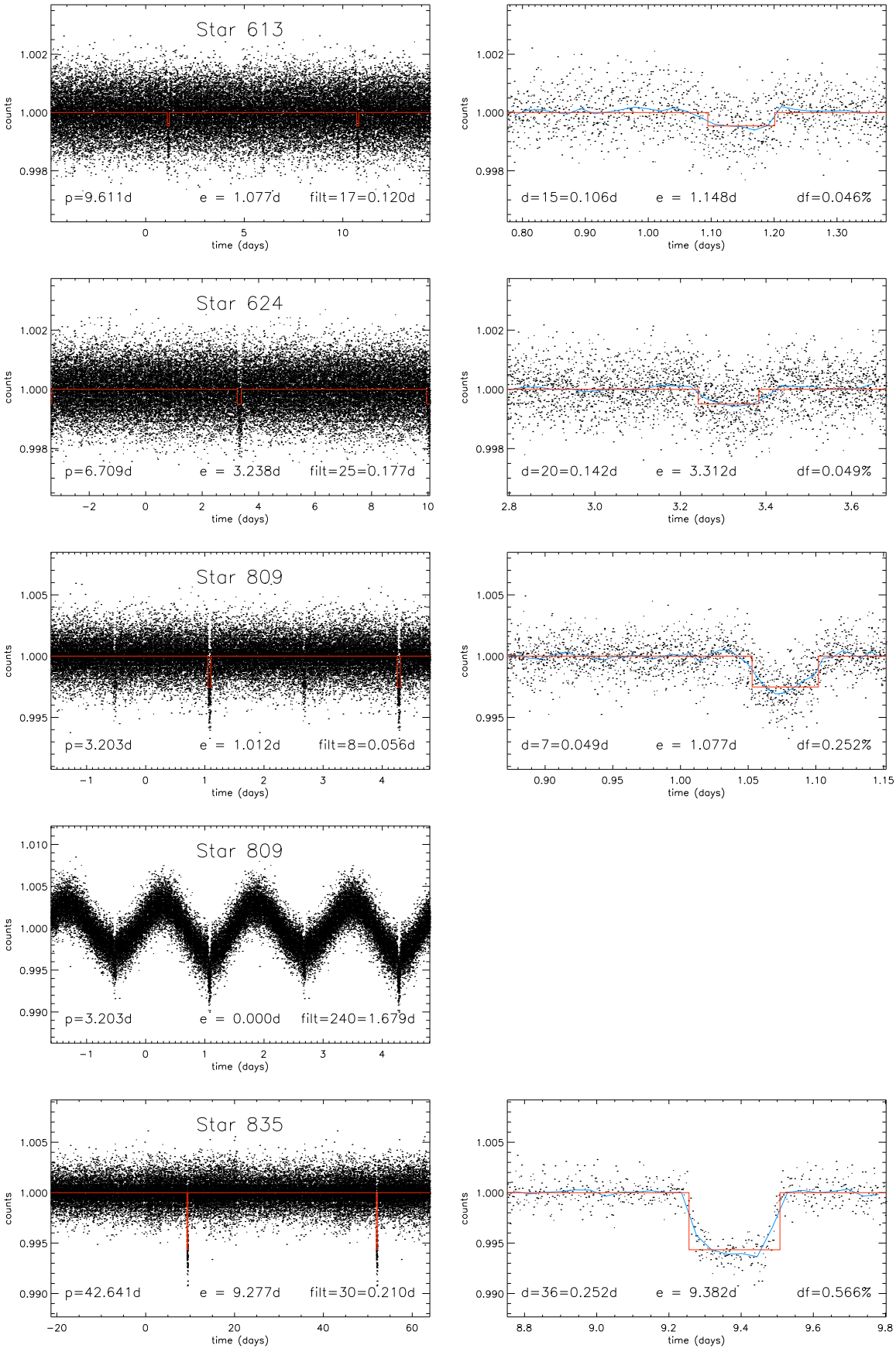

Figure 6.19: Light curves of transit / variable star candidates – page 6.

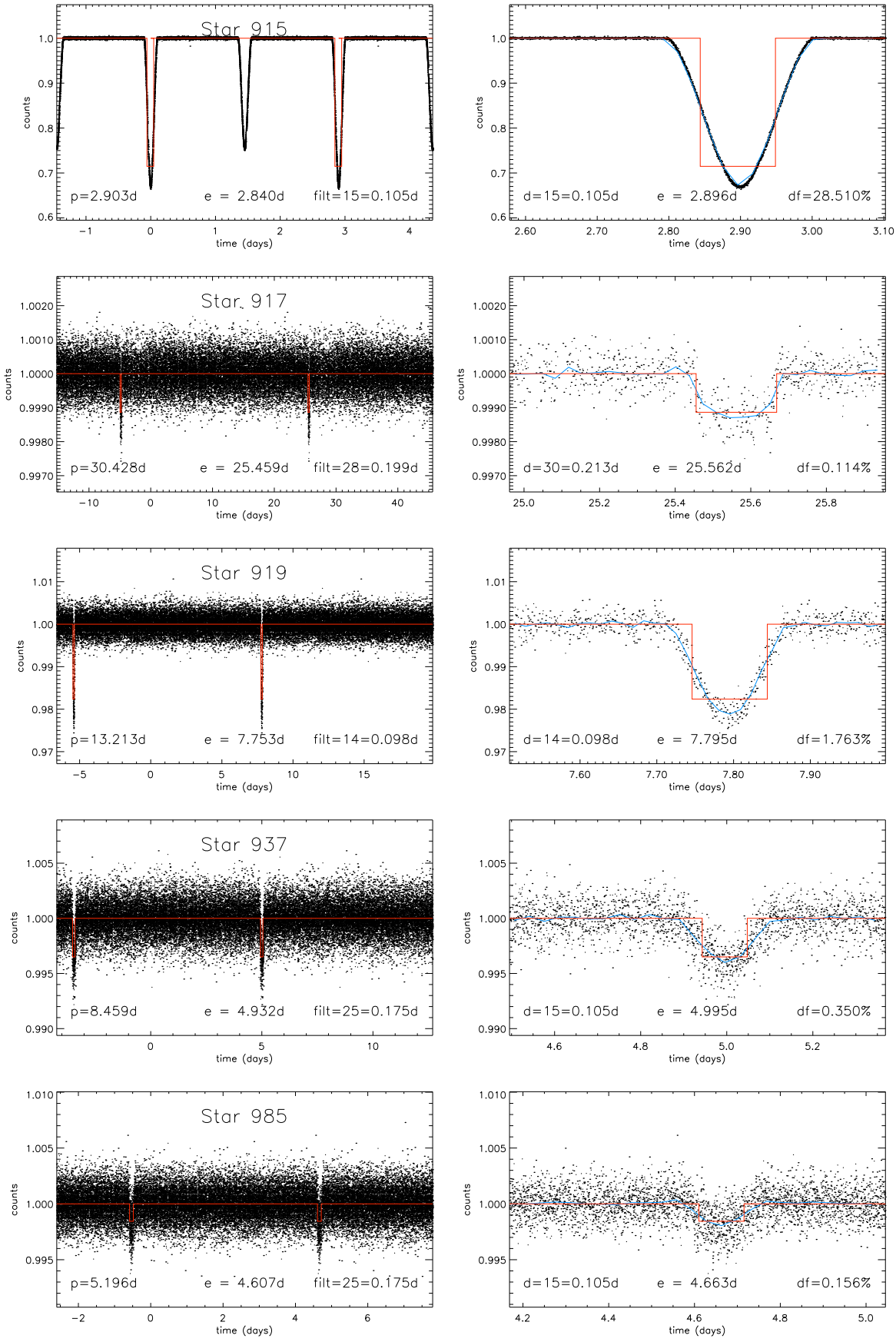

Figure 6.20: Light curves of transit / variable star candidates – page 7.

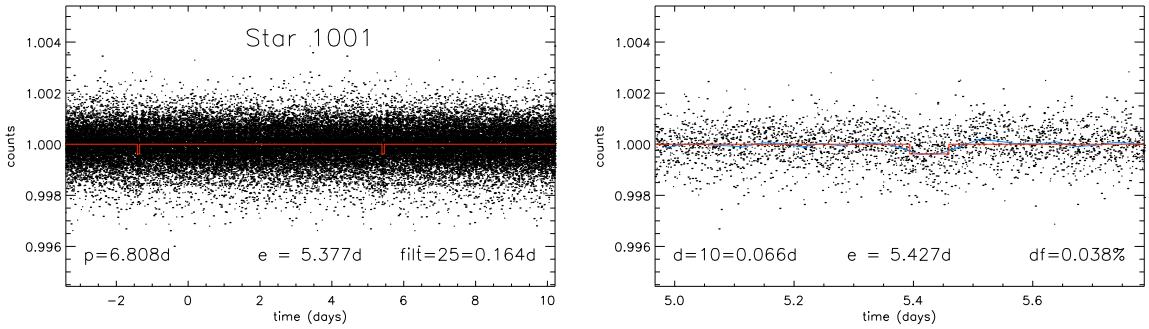

Figure 6.21: Light curves of transit / variable star candidates – page 8.