

Comparative blind test of five planetary transit detection algorithms on realistic synthetic light curves

C. Moutou¹, F. Pont^{1,5}, P. Barge¹, S. Aigrain², M. Auvergne³, D. Blouin¹, R. Cautain¹, A. R. Erikson⁶, V. Guis¹, P. Guterman^{1,7}, M. Irwin², A. F. Lanza⁴, D. Queloz⁵, H. Rauer⁶, H. Voss⁶, and S. Zucker^{5,*}

¹ LAM, Traverse du Siphon, BP 8, Les Trois Lucs, 13376 Marseille Cedex 12, France
e-mail: Claire.Moutou@oamp.fr

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

³ OPM, Place J. Janssen, 92195 Meudon Cedex, France

⁴ INAF - Osservatorio Astrofisico di Catania, via S. Sofia, 78, 95123 Catania, Italy

⁵ Observatoire de Genève, 51 Chemin des Maillettes, 1290 Sauverny, Switzerland

⁶ DLR Institute of Planetary Research, Rutherfordstr. 2, 12489 Berlin, Germany

⁷ Gemplus Card International, La Ciotat, France

Received 8 November 2004 / Accepted 16 February 2005

Abstract. Because photometric surveys of exoplanet transits are very promising sources of future discoveries, many algorithms are being developed to detect transit signals in stellar light curves. This paper compares such algorithms for the next generation of space-based transit detection surveys like CoRoT, Kepler, and Eddington. Five independent analyses of a thousand synthetic light curves are presented. The light curves were produced with an end-to-end instrument simulator and include stellar microvariability and a varied sample of stellar and planetary transits diluted within a much larger set of light curves. The results show that different algorithms perform quite differently, with varying degrees of success in detecting real transits and avoiding false positives. We also find that the detection algorithm alone does not make all the difference, as the way the light curves are filtered and detrended beforehand also has a strong impact on the detection limit and on the false alarm rate. The microvariability of sun-like stars is a limiting factor only in extreme cases, when the fluctuation amplitudes are large and the star is faint. In the majority of cases it does not prevent detection of planetary transits. The most sensitive analysis is performed with periodic box-shaped detection filters. False positives are method-dependent, which should allow reduction of their detection rate in real surveys. Background eclipsing binaries are wrongly identified as planetary transits in most cases, a result which confirms that contamination by background stars is the main limiting factor. With parameters simulating the CoRoT mission, our detection test indicates that the smallest detectable planet radius is on the order of 2 Earth radii for a 10-day orbital period planet around a K0 dwarf.

Key words. planetary systems – methods: data analysis – techniques: photometric – methods: observational

1. Introduction

Transit searches have recently shown their potential in discovering planetary candidates. The ground-based OGLE project, for instance (Udalski et al. 2002a,b, 2003, 2004), detected 177 planetary transit candidates, among which so far 5 are confirmed as short-period planets (Konacki et al. 2003; Bouchy et al. 2004; Pont et al. 2004; Konacki et al. 2005, submitted). Space-based transit searches are expected to be much more efficient, because of i) their continuous time sampling over long periods; and ii) the stabler photometric signal. At precisions of a few mmag, the main limitation comes from residual systematics due to the instrument and from intrinsic stellar variability. These are the problems that the transit detection

algorithms should face in future space missions for long-term planet searches: CoRoT (Baglin 2003), Kepler (Borucki et al. 2004), and Eddington (Favata 2004).

Several transit detection algorithms were proposed in the recent literature: Bayesian algorithms (Doyle et al. 2000; Defaÿ et al. 2001; Aigrain & Favata 2002), matched filters (Jenkins et al. 1996), box-shaped transit finder (Aigrain & Irwin 2004), and the Box-fitting Least Squares (BLS) method (Kovács et al. 2002). A theoretical comparison of these methods was proposed (Tingley 2003), which concluded that “no detector is clearly superior for all transit signal energies”, but an optimized BLS algorithm still performs slightly better for shallower transits. Here, we adopt a more empirical approach to make the comparison by using as a testbench a set of synthetic light curves with detailed simulations of the instrumental noise and astrophysical sources of variability. The test of these five

* *Present adress:* Faculty of Physics, Weizman Institute of Science, Rehovot 76100, Israel.

different transit detection techniques was blind, as the five different detection teams had no prior knowledge of their content.

This comparison of detection algorithms is likely to be relevant for all transit-search programmes, both from the ground and from space, although it was focussed here on CoRoT, to be launched in 2006, as the first space mission largely dedicated to transit searches. The CoRoT characteristics are given in Boisnard & Auvergne (2004), and its planet detection capability is estimated in Bordé et al. (2003). This ability is empirically addressed in this paper.

The goals of this blind detection simulation are the following:

- to independently apply several light curve analysis methods to the same simulated light curves, removing the possible “subjective” elements, such as possible biases when the same person both simulates the transit and detects it;
- to compare their ability to detect faint transits, avoiding false positives (hereafter defined as the noise features from instrumental or stellar micro-variability origin, accidentally picked up as a transit signature);
- to estimate the impact of star micro-variability for transit searches;
- to test the ability to distinguish between a planetary transit and an eclipsing binary from the light curve alone.

Applied to CoRoT, this exercise will help in deriving an estimate for the detection limits of this instrument and its limiting factors, as well as defining the strategy for light curve analysis and the required follow-up.

Section 2 presents the light curve building procedure; Sect. 3 then describes the five light curve analysis methods, and Sect. 4 discusses the results and reaches conclusions.

2. Generating simulated light curves

The synthetic light curves were built by combining several components: the instrumental model, stellar micro-variability, and in some cases, an additional event, such as a planetary transit, an eclipsing binary or a variable star.

2.1. Instrumental model

An instrument model (Auvergne et al. 2003) has been designed for CoRoT in order to evaluate the instrument detection capabilities and test the onboard and ground-based software. We use the output of this model as the basis of our synthetic light curve construction. Let us recall that the CoRoT onboard software will perform photometry on a pre-determined list of stars (12 000 per pointing) every 8 min during 150 days, by summing all the signal within pre-defined aperture covering between 100 and 60 pixels depending on the magnitude. Environmental perturbations, such as light scattered by the Earth, radiation flux, Attitude Control System jitter and temperature variations, are computed by specialised models. The outputs are light curves at the focal plane level, proton fluxes with a 10 mm CCD shielding, satellite angular depointing and temperature curves for the most sensitive sub-systems. Monochromatic PSFs are then provided using an optical model of the telescope, and used

to compute white PSFs, taking into account the optical transmission, CCD quantum efficiency, and target flux for main sequence stars in the effective temperature range 3500 to 9000 K. The appropriate photometric aperture is computed, depending on the star position, magnitude and colour (Llebaria et al. 2003).

We build 25 basic light curves based on stars scanning 5 mag from 12 to 16 and 5 temperatures from 4500 K to 6750 K, all located at the same CCD position. They contain the following realistic noise contributions:

- (i) photon noise (Poisson statistics);
- (ii) flat-field noise, with a 1% non-uniformity;
- (iii) read-out noise of 10 electrons/pixel/read-out;
- (iv) no jitter amplitude, as it is negligible in the CoRoT broad bandpass;
- (v) zodiacal light, the unique source of sky background in space, a uniform offset of 12 electrons/pixel/s over the CCD remaining constant along the orbit. It is corrected by subtraction and the resulting additional photon noise is kept;
- (vi) proton impacts. The exposures corresponding to the crossing of the South-Atlantic (SAA) anomaly are not usable and the final data thus contain a large number of quasi-periodic gaps (typical duration of 30 min each interval of 1.7 h) that should be handled by the detection algorithms (Fig. 1);
- (vii) earth-scattered light, which is not uniform over the CCD and varies along the orbit, almost following the orbital period. We insert a scattered light contribution with a realistic maximum value of 1 electron/pixel/s. As it will be corrected in the processed CoRoT light curves to a certain level, we subsequently remove the scattered light contribution to first order, leaving a random <50% residual. The correction applied may lead to a positive or a negative residual signal, corresponding respectively to an overestimation or underestimation of the actual scattered light level (Fig. 1). This allows us (i) to test the robustness of the detection algorithms, especially against a negative (i.e. when it is over-corrected), quasi-periodic signal; and (ii) to create 999 light curves with varying scattered-light noise amplitudes, produced from a parent set of 25 instrumental curves. Note that scattered light is the dominant systematic signal in the CoRoT instrumental noise and the only instrumental systematics included in the simulation; this is the reason we deliberately took a conservative value for its level of correction.

2.2. Stellar micro-variability

Stellar micro-variability curves are taken from two independent models. These effects are independent of the instrument and are usually thought to be among the main limitations of transit detection. Considering two types of micro-variability curves, there are 55 different light curves. To build the final light curves the micro-variability curves are all scaled by a random factor between 0.5 and 2, to account for the dispersion in the variability level observed in real stars. They are also phase shifted by

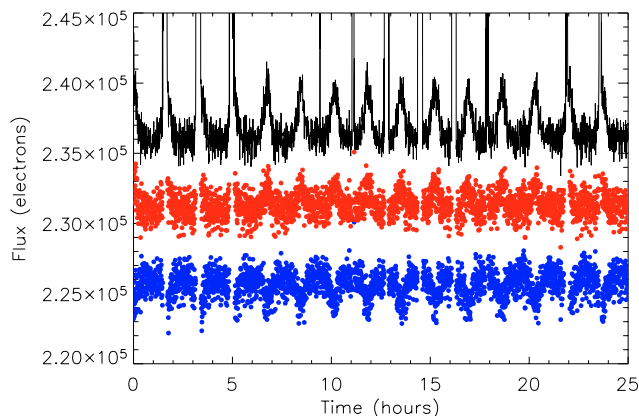


Fig. 1. Example of an instrumental light curve before (*top*) and after the partial correction of scattered light (once underestimated (*middle*), and once overestimated (*bottom*)). The sharp peaks in the upper plot are due to the SAA crossing; they become gaps in the output light curves.

an arbitrary value, and rebinned in their time sampling by a random factor between 1.0 and 1.2, to avoid excessive similarity between the final light curves.

2.2.1. A scaled solar-like model for stellar variability

Lanza et al. (2003) model the variation of the Total Solar Irradiance (TSI) by considering a simple stellar-like model based on the rotational modulation of the visibility of three active regions plus a uniformly distributed background component that accounts for the surface features affecting the mean level of the solar flux. Each active region consists of faculae and sunspots in a fixed area ratio and with brightness contrasts that are functions of the limb angle. A time interval of 153 days ranging from 1st July to 1st December 2000 is selected as representative of the variability of the TSI close to the maximum of the 11-yr cycle. The model is applied to successive subintervals each of 14 days, separated by 7 days from each other, to obtain the coordinates and the areas of the three model active regions plus the uniform background term.

In order to simulate optical light curves for main-sequence stars with faster rotation than the Sun and with a higher activity level, the rotation period and the areas of the three model active regions are varied: the areas of the three active regions, as well as the uniform background term, are multiplied by a factor $f = A(P, Sp)/A_{\odot}$, where $A(P, Sp)$ is the average amplitude of the optical light curves of a star of rotation period P and spectral class Sp derived from Messina et al. (2003), and where $A_{\odot} = 2.2 \times 10^{-3}$ mag is the maximum amplitude of the solar optical variability.

For stars with a rotation period longer than 12 days, there is no information on the amplitude of the rotational modulation in the optical passband (except for the Sun), so that f is assumed to be in the range 1.5 to 6 for a spectral type varying from F5V to K5V. The coordinates of the three active regions are those of the solar model active regions, and the inclination of the stellar rotation axis with respect to the line of sight is fixed at 90° . To reduce the impact of the small discontinuities occurring every

7.0 days at the passage from a fit to the next, the model parameters are linearly interpolated in time between successive best fits. The brightness contrast coefficients and their center-to-limb variations are the solar ones.

The ratio of the area of the faculae to that of the sunspots in an active region is estimated by extrapolating the relationship given by Chapman et al. (1997) to larger sunspot areas. The resulting facular contribution is found to be negligible for stars with a rotation period shorter than 20 days and spectral type later than G8. The variability on time scales significantly shorter than the rotation period is modelled by scaling the residuals of the best fits to the solar TSI variations, which are due to the evolution of the solar active regions on time scales shorter than 4–5 days (Lanza et al. 2003, 2004). In order to increase the amplitude of the short-term stellar variability to make the planetary transit search more challenging, the residual solar variability is multiplied by a factor $3f$ and linearly interpolated to get an even time sampling of 8 minutes. Finally, Poisson random fluctuations with a relative standard deviation of $[3 \times (fA_{\odot})^2]^{-0.5} = 3.8 \times 10^{-3} f^{-1}$ are added to simulate short-term variations due to microflaring or convection on time scales of several minutes.

In addition to the original TSI light curve, 9 light curves were produced with this method, with spectral types F5, G0, and G8 and rotation periods 3, 10, and 20 days. The amplitude of micro-variability ranges from 0.1 to 4%. The stellar optical time series so obtained are dominated by the rotational modulation except for rotation periods longer than 15–20 days for which the active region evolution prevails on the rotational modulation signal. A few small discontinuities are present, due to the passage from a 14-d fit to the successive one, but they never exceed 5% of the amplitude of the rotational modulation, even in the case of the most active stars.

2.2.2. Light curves from SIMLC

SIMLC is a tool to simulate stellar micro-variability for stars with spectral types F5 to K5 and ages later than 625 Myr. It works by computing an artificial power spectrum, starting from a fit to solar data and scaling it using empirical scaling laws. The power spectrum is then sampled as appropriate, given the time sampling and light curve duration required, coupled with a random phase array and reverse Fourier-transformed to the time domain. More details can be found in Aigrain et al. (2004), so only a brief summary is given here.

Following Andersen et al. (1994), the power spectrum of the Sun's total irradiance variations up to $\sim 600 \mu\text{Hz}$ (as observed with the PMO6 radiometer, which is part of the VIRGO experiment on SOHO) is modelled as a sum of three broken power laws, each characterised by an amplitude, characteristic timescale, and slope. There are 3 components with timescales of 10 days, 4 days, and 10 min. The powerlaw slopes are 3.8, 1.8, and 2.0. All these values are those measured for the Sun. Note that because the slope of the first powerlaw is quite steep it falls off quickly for timescales larger than 10 days, while the second powerlaw, which is quite shallow, is still the dominant component at $100 \mu\text{Hz}$ (timescales of a few hours, typical of

transits). The amplitude of the lowest frequency, or “active regions”, component is correlated with simultaneous measurements of the CaII K-line index indicator of chromospheric activity. Higher frequency components, which have much smaller amplitude, are thought to be related, respectively, to super- or meso- granulation and to a superposition of granulation, oscillations, and photon noise.

Empirically derived scaling laws can be used to scale the amplitude and timescale of each power law to what might be expected for other stars. Currently this can be done only for the dominant low-frequency component, using chromospheric activity as a proxy. Observational constraints are currently insufficient to derive scaling laws for the other components, including the second component that corresponds to the timescales characteristic of planetary transits, so those are thus left as they have been measured in the Sun. Upcoming data, in particular from the MOST (Micro-variability and Oscillations of STars) satellite (Walker et al. 2003), are expected to provide constraints on this component in the near future.

A set of 45 light curves lasting 150 days, with 8 min sampling, were generated for the present exercise. They correspond to a grid of stars of spectral type F5, F8, G0, G2, G5, G8, K0, K2, and K5, and ages 0.625, 1, 2, 3, and 4.5 Gyr. The amplitude of the dominant, “active regions” component of the variation scales with convection zone thickness (which is larger in later spectral types) and the inverse of the rotation period (which is larger in older stars), while the characteristic timescale scales roughly with the rotation period. As a result, at 0.625 Gyr the most variable stars are F-stars, while at 4.5 Gyr they are K-stars. The amplitude of micro-variability ranges from 0.01 to 0.1%, a level much lower than those obtained with the method described in Sect. 2.2.1. This is thought to be due to the more coherent nature of micro-variability in active stars, which SIMLC currently cannot reproduce.

2.3. Transits

Twenty planet transits were simulated. For a thousand light curves, this represents about an order of magnitude more transit events than expected in real samples (Bordé et al. 2003). It is important that light curves without transit vastly outnumber those with transits in the simulation, so that the detection thresholds have to be set realistically high. The characteristics of the inserted transits are not chosen with the goal of reproducing planet statistics, because those are mostly unknown in the range where CoRoT will discover planets; the idea is instead to test limitations and to explore the borders of detectability. The objectives are then (1) to sample a variety of system cases; and (2) to investigate the detection limit by including a large number of small planets in light curves with a varying noise level. The characteristics of the transits are summarized in Table 1. The planet size spans the range from 1.6 Earth radius (R_E) to 1.3 Jupiter radius (R_J). One system with two planets is inserted. The period domain is 4 to 90 days. Target stars with the planetary transits are chosen at “directed random”, with the aim of exploring the regions near the limit of detectability. For instance, the largest planets are inserted in the

light curve of faint and/or active stars. The largest planets are also the ones with the lower number of transits (the hot Jupiter configurations, as easy cases for space transit searches, are not emphasized here).

The transit light curves are simulated with the aid of the Universal Transit Modeler (Deeg 1999). Limb darkening of stars are estimated from recent calculations from ATLAS9 models and the CoRoT bandpasses (Barban priv. comm., see method in Barban et al. 2003), considering both a linear limb-darkening law and a classical mixing-length theory.

2.4. Eclipsing binaries and large-amplitude variable stars

Simulations (Brown 2003) and the results of the OGLE planetary transit follow-up (Bouchy et al. 2004b; Pont et al. 2004b) indicate that for a given transit signal depth, the contamination by grazing and background eclipsing binaries (EB) will be at least as numerous as the planet transits themselves, or could even largely outweigh the true planet events. To simulate this contamination, we inserted ten low-depth stellar eclipse signals among the light curves. There are grazing binaries (6 events), background binaries (4 events), and one hierarchical triple stellar system. Finally, we inserted five background variable stars: a low-amplitude delta Scuti, a classical Cepheid, a β Cephei, the semi-regular variable Z UMa, and the irregular Z Cam. The background variables and background eclipsing binaries refer to fainter objects included in the same aperture, 3 to 7 mag fainter than the main target. In the case of grazing eclipses, the binary star is the main target itself, i.e. in the magnitude range 12–16 for CoRoT.

The characteristics of these light curves are summarized in Table 2. Again, the characteristics of the systems are chosen to cover most possible combinations rather than to reproduce the expected characteristics of real samples. Our eclipsing binary transits include curves with anti-transit signals, with sine and double-sine modulations outside the transits due to the ellipsoidal deformation of the primary under the gravitational influence of the secondary, V-shaped eclipses (grazing) and U-shaped eclipses (central eclipse in a background contaminant system). For grazing eclipsing binaries, the algorithms of Mandel & Agol (2002) and Wichmann (1998) are used. The Universal Transit Modeler (Deeg 1999) is used for background eclipsing binaries and the triple star. The variable star light curves are taken from the literature and from the archives of the AAVSO (American Association of Variable Star Observers).

2.5. Crowding

Another consequence of background stars is to contribute to the flux variations measured in the aperture placed on the primary target. To simulate this effect, we systematically added the contribution of one background star to the primary light curve, which is characterised by a light curve constructed with the same procedure as for the main target and a magnitude difference with a distribution probability $\sim 2^{\Delta m}$ in the range 0–6 mag (thus including stars up to 22th magnitude). For the second star,

Table 1. The characteristics of the transits that were inserted in the light curves: the star radius R (in solar radius units), the stellar limb darkening coefficient (LD), the planet radius r , the orbital period in days, the system inclination in degrees, the semi-major axis a , the star magnitude, the final standard deviation of the light curve in percents, and some comments. The detection flag shows a series of + and – signs, corresponding to each team, respectively, from 1 to 5; + means a positive detection (for Team 1 in position 1, etc.), – means that the event is missed.

ID	R (R_{\odot})	LD	r (R_{\odot})	Period (days)	Inc (deg)	a (R_{\odot})	m	Std dev. %	Comment	Detection flag
34	0.92	0.6	0.025	5.52	88.8	12.77	13	0.17		+++++
85	1.1	0.4	0.099	26.4	88.9	37.876	15	0.75		+++++
168	0.92	0.5	0.13	11.5	87.4	20.827	15	0.71		--+++
207	0.92	0.5	0.11	88.4	90.0	79.89	16	1.42		+++++
317	1.1	0.6	0.02	33.8	89.5	44.66	12	0.09		-----
326	0.85	0.6	0.017	6.8	89.9	13.9	14	0.40		-----
390	0.92	0.6	0.022	8.0	89.	16.35	12	0.07		+++++
460	1.1	0.3	0.076	32.9	89.52	23.49	15	0.82		+++++
474	0.92	0.6	0.028	11.34	89.	20.63	13	0.18		+++++
533	0.92	0.7	0.095	6.4	90.0	7.89	16	1.54		+++++
537	0.85	0.6	0.015	2.78	89.9	7.68	12	0.09		--+-+
575	0.85	0.6	0.019	15.9	90.0	24.57	14	0.40		-----
613	1.1	0.6	0.026	4.8	89.4	12.16	14	0.29		+--++
618	1.3	0.6	0.023	8.48	89.	19.55	12	0.09		-----
624	1.1	0.6	0.029	6.7	89.8	15.18	14	0.29		+--++
681	1.1	0.6	0.023	19.8	89.6	31.27	13	0.20		-----
715	1.3	0.3	0.098	10.1	86.4	21.96	15	0.75	Planet 1	-----
			0.07	63.8	89.7	75.0			Planet 2	-----
835	1.1	0.4	0.084	42.6	89.3	52.10	15	0.74		+++++
915	1.5	0.25	0.13	58.32	89.9	70.0	15	0.74	Planet	-+ +--
		0.3	1.1	2.9	86.0	11.4			Binary	+++++
917	0.85	0.6	0.028	30.4	89.7	37.8	13	0.18		+++++

another stellar micro-variability curve is used. Thus, each final simulated light curve consists in the addition of two different contributors.

2.6. The final set of 999 simulated light curves

The sample of 999 light curves was composed from a combination of the individual elements described so far, as it is developed in this section. The parent lightcurves are the 25 instrumental curves from a grid of 5 mag and 5 color temperatures, with a level of scattered-light residual noise different in all lightcurves. The magnitude of each target was drawn from a probability distribution: $p(m) \sim 2^m$, approximating an isotropic distribution near the Galactic plane, between 12 and 16 mag. The distribution of color temperatures was selected to roughly match a spectral type distribution that is realistic for magnitude-limited, transit-search fields near the Galactic plane; from an analysis of the stellar population in future CoRoT fields, there are 40% of F dwarfs, 40% of G dwarfs, and 20% of K dwarfs (Moutou et al., in prep.). Finally, the micro-variability fluctuations were inserted; from the 55 parent light curves, all final micro-variability contributions are unique due to the applied amplitude and temporal extension factors (Sect. 2.2), and due to the injection of a fainter stellar light curve (Sect. 2.5). The micro-variability light curve was also selected to match the color temperature (or spectral type) of the

instrumental light curve. In total, 964 light curves do not have any transit or EB/variable star signal.

The temporal sampling of the final light curve is 8 min, with a duration of 150 days, as for CoRoT long observing runs. A complete light curve contains 25 056 data points.

The package of 999 light curves (identified with ID 1 to 999 in the following) were supplied to the detection teams with information neither on their content nor on the way they were calculated, whether the number of hidden planets or the nature of injected noise sources. In the real case with CoRoT light curves, some data will be known beforehand, such as the star magnitude, spectral type, luminosity class, contamination by neighbours, and pipeline processing parameters. This knowledge is not fundamental for transit detection but will obviously help in the identification of the detected events.

3. Blind search for transit events

In this section, we describe the five methods used for detrending the light curves and detecting the transits. Their elements span a wide range of complexity from fairly basic to very evolved. They also differ by their previous use: one team started from scratch with no experience in transit detection, two teams use algorithms that they developed for ground-based transit surveys (BEST and OGLE), and two teams are working on algorithms for space-based transit searches.

Table 2. Table of contaminating events introduced into the light curves: magnitude, event type (“BEB” stands for background eclipsing binaries, “GrB” stands for grazing binaries), period and relative flux (contribution of the background star to the total flux), and the standard deviation of the final light curve. Detection flag: detection and correct identification (+), wrong identification (*i*), no detection (–), for each team from 1 to 5. References: UTM (Deeg, 1999, UTM), Nightfall (Wichmann, 1998, W98), (Mandel & Agol, 2002, MA), AAVSO (American Association of Variable Star Observers), Andreasen (1988) (A88).

ID	mv	Event type	Period (days)	Relative flux	Std dev. %	Reference	Detection flag
31	14	BEB	24.7	0.03	0.38	UTM	+++++
131	14	δ Cepheid	5.86	–	0.40	A88	----+
249	16	triple star	3.9	–	1.03	UTM	+++++
259	15	GrB	1.4132	–	0.82	W98	+----+
271	15	Z Cam	–	–	0.82	AAVSO	-----
384	15	β cephei	0.2835	0.001	0.81	AAVSO	+---++
386	15	GrB	17.1	–	0.66	UTM	iiiiii
486	15	BEB	2.4128	0.01	0.66	UTM	--i-+
518	15	GrB	78.3	–	0.82	MA	-----
553	15	δ scuti	0.07342	0.003	0.66	AAVSO	--++++
599	15	GrB	1.874	–	0.75	W98	+--+i+
650	14	semi regular	–	–	0.37	AAVSO	-----
809	15	GrB	3.2	–	0.75	W98	i-+++
919	16	GrB	13.2	–	1.02	UTM	i++ii
937	15	BEB	8.452	0.01	0.81	UTM	iiiiii
985	15	BEB	5.19	0.01	0.71	UTM	iiiiii

3.1. Team 1: Correlation with a sliding transit template

The first algorithm is based on correlation of the light curve with a single sliding template but without prior detrending. Systematic noise on short timescales is removed from the correlation function, then candidates with a high signal in the correlation function are examined individually by eye to pick up the final detections.

Detecting the transits: The light curves are correlated with a sliding template to compute a correlation function $C(t)$. The template is a transit shape based on the algorithm of Mandel & Agol (2002). The use of a unique transit template is sufficient and makes the method much simpler; the optimum template has a transit duration of ~ 8 h and is bordered by two flat segments of ~ 14 h. Previous filtering of the long-term variations is not crucial in this case, because the template covers only a small part of the light curve at a time. Figure 2 shows the resulting correlation functions for a few cases. In this method, no periodicity is assumed in the transit signal, and the period is estimated a posteriori.

One advantage of the correlation method is that it is not affected by gaps in time coverage of the data. Missing epochs simply make no contribution to the correlation function, which avoids those problems caused by any interpolation of the data in the gaps.

Detrending the light curves: As explained above, no detrending was done on the long-term variations. Correlation curves show a common pattern of perturbation on short time scales, associated with instrumental effects like temperature changes (“breathing”), scattered light or pointing jitter. We assume that

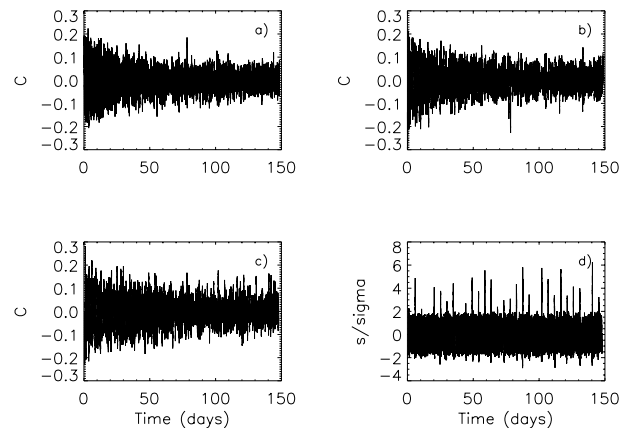


Fig. 2. **A)** and **B)** are two correlation functions (“detection curves” DC) showing systematic noise. Artefacts are sometimes obvious (synchronous spikes and similar envelope) or can be hidden, with a known or unknown origin. **C)** and **D)** show DC613 before then after detrending (note the very different y -axis scales).

this instrumental noise introduces a common noise in all correlation functions, except for a scale factor. We model this by $\vec{C}_i = \vec{s}_i + \lambda_i \cdot \vec{p}$ where \vec{C}_i is the temporal correlation curve, \vec{s}_i , the unknown noise-free correlation curve, and p (with $\|\vec{p}\| = 1$ by convention) the unknown instrumental perturbation common to all objects, weighted by the unknown λ_i . It appears that the average of λ is close to zero, so that p cannot be simply estimated by averaging the curves. To retrieve p we apply the following sequence:

1. choose an initial guess for p from a light curve strongly dominated by p ;
2. estimate a first-guess λ_i by projecting \vec{C}_i on \vec{p} ;

3. the mean of C/λ over all objects is our refined p , giving the refined λ_i .

Actually, we found that the instrumental noise pattern was not common to all stars, but could be classified into a number of families. We therefore applied the above procedure to determine different p for each empirically determined family.

Detrended correlation functions exhibiting a strong signal (i.e. about 5% of the light curve sample) are then examined by eye, selecting the candidates with strictly periodic signals and folding accordingly each light curve to point out autosimilarity of the shape.

Discussion: It turns out that the “families” of objects used to remove the noise in the correlation function often correspond to sets of light curves based on the same parent noise curve. Therefore, with this method, the removal of the systematic noise is probably more efficient on simulated data than it would be in reality.

Correlation with a sliding transit template is among the simplest possible methods for transit detection, short of direct examination of all light curves by eye, and the results of this algorithm on our synthetic sample can be used as a reference point of comparison for the performances of the other algorithms.

3.2. Team 2: Box search with lowpass filtering and broken-line detrending

The algorithm searches for box-shaped signals in normalized, filtered, variability fitted, and unfolded light curves. It was designed to detect single, as well as periodic, transit events.

Detrending the light curves: In a first step all the light curves are normalized, neglecting all the epochs without flux value. These epochs, covering a maximal time span of 43 min caused by crossings of the SAA, are short compared with the transit durations of a minimal 2 h. Therefore the missing epochs are linearly interpolated, without risk of introducing false transit events. A Fourier analysis is carried out, giving a dominant systematic periodic signal at a period of $P \sim 1.13$ h – the orbital period of the satellite (residuals of the scattered light contribution). A standard lowpass-filter is used to eliminate this signal and other high frequency signals. The cut-off frequency is varied between 0.059 day^{-1} and 0.177 day^{-1} . The shape of the transit signals is moderately deformed by this kind of filtering, but for the purpose of a detection tool the influence of this side-effect is negligible. Another side-effect of the lowpass-filtering is that an additional modulation of the light curves occurs at the beginning and the end of the data. Therefore the data of the first and last 10 days are excluded from the transit search. The stellar variability is fitted locally. The light curves are separated into sub-sections and a linear least-squares fit to the data is performed for every sub-section. The size of the sub-sections is varied in a range between 0.5 to 3 days to ensure that no transit-like signals are significantly altered. In a subsequent step the fit is subtracted from the data (see Fig. 3).

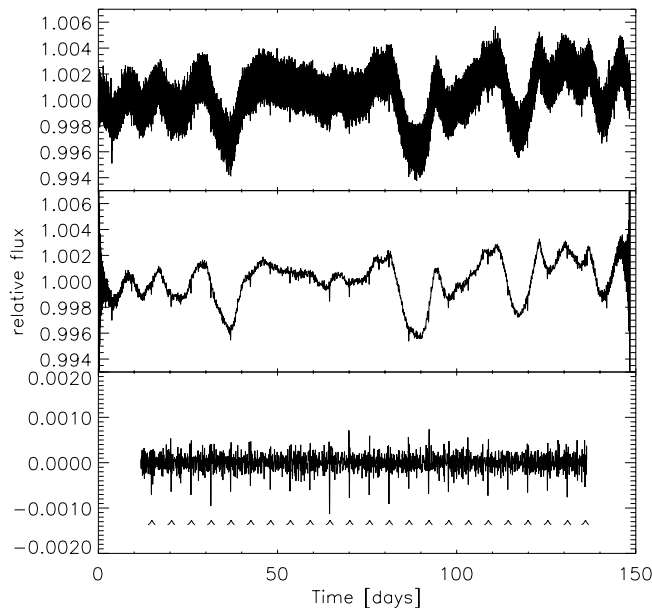


Fig. 3. The different steps of the light curve analysis of Team 2 for light curve ID34. First the data gaps (not visible at this scale) are interpolated and the light curve is normalized (*top*), and a lowpass-filter is applied to remove high frequency signals (*middle*). Finally the stellar variability is modeled and a search for period signals performed (*bottom*). The periodic signal found is marked in the figure.

Detecting the transits: The standard deviation of the normalized, lowpass-filtered and variability-fitted light curves is calculated. Subsequently a box search for transit-like events is carried out. All data points deviating from the average signal by 3σ are identified and the neighbouring deviating points combined into a single detection. A maximal and a minimal signal length are defined, corresponding to transit lengths between 1 to 30 h. Mean epochs of the signals are determined, spurious detections are excluded (this mostly concerns an instrumental artefact that is identified in all light curves), and all remaining single detections are listed for further inspections. Thereafter the epochs of the potential events found are automatically searched for periodicity: time differences between all detected events with approximately the same detection level are estimated and retained when a single time difference or multiples of it have occurred several times within a given error margin. For possible cases a detailed investigation of the potential transit events is performed where the depths and duration of the events are determined. The corresponding light curves are manually inspected for secondary eclipses and gravitationally induced modulations caused by high-mass secondaries.

Finally all light curves with detected events are classified as either possible transit-like or other events.

Discussion: The algorithm is based on a search routine for single transit events developed for the ground-based transit search BEST (Rauer et al. 2004). The adapted version can detect both single and periodic transit-like events. It was also adapted to be able to detrend the microvariability that is not an issue for ground-based wide angle searches. In general, to discriminate between real and false transit signals, more

information about the host star is helpful, especially for non-periodic single transit events.

3.3. Team 3: Box least-squares with 200-harmonics filtering

In this method, the light curves are detrended by fitting 200+5 harmonics, then transits are detected with a box-fitting on the phase-folded signal.

Detrending the light curves: The scattered light periodic effect is verified as having the same period in all light curves, though its shape is varying among them. Therefore it seems plausible to describe it (in each light curve independently) as a sum of a small number of harmonics (5) of a fundamental frequency – $f_S = 14.18 \text{ day}^{-1}$:

$$F_S(t_n) = \sum_{k=1}^5 a_k \cos(2\pi k f_S t_n) + \sum_{k=1}^5 b_k \sin(2\pi k f_S t_n).$$

Separately, the long term stellar variability is also modelled as a sum of harmonics. The fundamental frequency used this time is $f_L = \frac{1}{2T}$, where T is the whole duration of the light curve (about 150 days). The number of harmonics to consider for the long-term variability model N_L is fixed to 200. Thus, the highest frequency in this model corresponds to a period of 1.5 days. We expect the energy in a planetary transit signal to be mainly concentrated in higher frequencies, because of the relatively sharp ingress and egress. Therefore the harmonics we fit should include only a negligible fraction of the energy in the transit signal:

$$F_L(t_n) = \sum_{k=1}^{N_L} c_k \cos(2\pi k f_L t_n) + \sum_{k=1}^{N_L} d_k \sin(2\pi k f_L t_n).$$

Also including the average level, the model is parameterized by 411 parameters, estimated by a least-square fit. Naïvely, that would involve solving a system of 411 linear equations (known as the normal equations) with the same number of unknowns. This may be prohibitively time-consuming; therefore, we consider only the times t_n for which a valid measurement existed for all light curves, i.e., times which are guaranteed to lie outside the SAA. This amounts to about two thirds of the original sampling times. Using only those points which are common to all light curves allows us to calculate the SVD (Singular Value Decomposition) pseudoinverse of the normal equation matrix (Press et al. 1993) and then use the *same* matrix to solve for the 411 coefficients in each light curve separately. As it turns out, the price we pay by using only part of the points is negligible, because of the very good time coverage. After the fit, the derived coefficients are used to model and remove the scattered light and stellar variability from the complete set of points of the light curve.

Detecting the transits: We apply the Box-fitting Least-Squares (BLS) algorithm, presented in Kovács et al. (2002), on the detrended data. We use a logarithmic sampling of the frequency space, with 2000 frequencies between 0.01 and 3 day^{-1} .

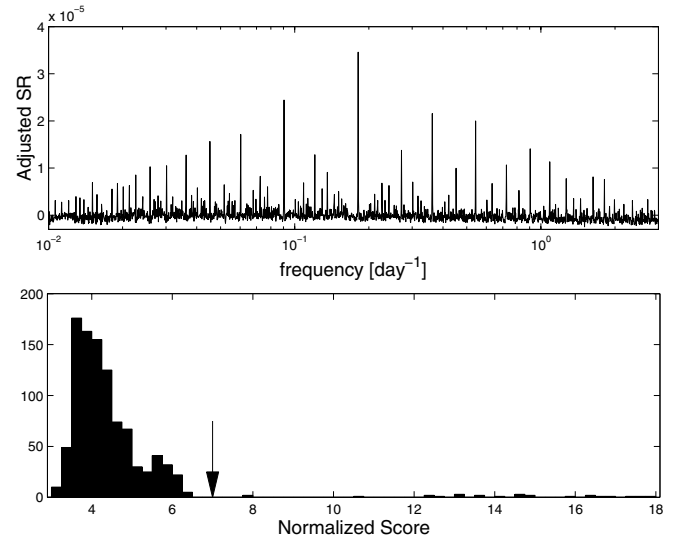


Fig. 4. *Top:* The adjusted SR function for light curve 34, which shows the typical peaks of a transit signal. *Bottom:* The distribution of the normalized SR of the 999 light curves. The arrow points to the adopted detection threshold of 7.0 (Team 3).

The maximum allowed transit width is proportional to the cubic root of the period, as is suitable for Keplerian orbits, and 5 phase bins are allowed in a maximum width transit. Eventually, a simple function ($a + \frac{b}{f}$, where f is the frequency) is fitted and subtracted from the SR (Signal Residue) function (Kovács et al. 2002) to adjust it for the varying number of configurations tested in each frequency (Fig. 4, top). Figure 4 (bottom) displays the distribution of the normalized relative heights of the SR peak for all 999 light curves. One can clearly see a bell-shape distribution, where all samples all lie below a value of 7.0. Thus, we fix 7.0 as the detection threshold, tagging all the scores above it as detections.

Discussion The main attractive feature of the harmonic-fitting procedure is that it does not require any interpolation of the measurements onto a regular grid. Such interpolation would have introduced interpolation noise with some periodic nature, due to the SAA gaps, and probably introduced more false alarms.

Removing harmonics with periods as short as 1.5 days may modify the shape of the transit signal a little, but it does not affect the detection capability. Final characterization of the transit signal is done by fitting a simplified transit model. We used linear ingress and egress, and a 'flat-bottom' transit. Fitting it together with the harmonics proved quite easy, using the SVD pseudoinverse method, and the derived transit signal is not modulated by the harmonics.

3.4. Team 4: Matched filter with image-processing detrending

In this method, the signal is denoised with the ‘‘Gauging Filter’’ (GF), and the detection is performed with a standard matched-filter associated to a three-criteria decision process (Guis & Barge 2004).

Detrending the light curves: In order to remove the low frequency modulations from the signal, the GF is applied to the light curves following a procedure described in Guis & Barge (2004). In this method the time plot drawing of a light curve is considered as a 2D-image which splits into two parts: one corresponds to the area below the time plot, the other to the area above. Then, a denoised signal S_F is defined as the mean value of the two curves resulting from the processing of these two images. The result may still contain residual components at various frequencies. These residuals are removed thanks to a second-order filtering at various scales and Fourier transforms.

The GF detrending procedure is the following: (i) the light curve is successively under-sampled and expanded with a linear recursive interpolation method over the data gaps in order to keep the total size of the light curve unchanged; (ii) the resulting light curve is smoothed out with a 4-width smoothing filter (widths are 2^n with $n = 6, 7, 8, 9$), producing smoothed light curves with different low frequency ranges; (iii) the final light curve is chosen as the optimum of the four filtered light curves. The final choice is made in Fourier space looking at the local minima of the energy contained in the four light curves and selecting the one within the lowest frequency range (i.e. the furthest from the transit frequencies).

For a given light curve, the best fit of the low frequency modulations obtained with our detrending method is denoted C_{LF} . In most light curves, the low frequency modulations are quite weak and using under-sampled light curves with loose smoothing is sufficient.

Detecting the transits: The detection method is based on the classical matched filter aimed at detecting a single feature in a noisy signal (Defaÿ 2001). It is composed of three main steps: (i) subtraction of C_{LF} from the light curve; (ii) convolution of the detrended light curve with a reference filter (based on a model of planetary transit) resulting in a convolution curve M ; (iii) identification of local convolution maxima in M , which directly provides possible positions for the transit-like features. The convolution maxima are selected in two different ways: one is a correlation in Fourier space with a library of periodic signals; the other is a sorting of the convolution peaks and selection according to criteria based on statistical parameters like the variance or the entropy of the peak distribution (Guis & Barge 2004). In summary, our detection method combines the standard matched filter, which is well-suited to localising isolated features, and the Fourier correlation, which permits finding periodicities more efficiently.

With the above method a total of 25 light curves were found to contain transit-like features (Tables 1, 2): 19 are identified thanks to Fourier correlation; 13 (resp. 2) correspond to single (resp. bi-) periodic features present all along the light curve, and 4 have the characteristics of an eclipsing binary. Bi-periodic events are characterized by two non-commensurable periods. Selection by peak sorting allowed the identification of the 6 other detections, with a lower confidence level but also some secondary features.

Discussion: Detection is made using pairs of peaks chosen among the set of selected peaks. The period corresponding to this pair is then checked against the positions of other selected peaks. The larger the number of the pairs or the shorter the periodicities, the higher the confidence level. In some cases (IDs 983, 985), the noise level is so strong, probably due to star micro-variability, that detections become less reliable. Indeed, the matched filter is very sensitive to strong discontinuities in the signal. Further, it can be noted that selection by peak sorting can bring out some potentially interesting cases, as for example signals with rough periodicity on parts of a light curve. However, such cases were removed from our list of possible events because their periodicity was not firmly established.

3.5. Team 5: Box maximum-likelihood with iterative 1-D filtering

Detrending the light curves: Residual scattered light variations, whose period is determined by sine-fitting over the range 0.065 to 0.075 day, are removed by phase-folding each light curve at the best-fit period, smoothing it using a 1-D filter (median, then boxcar filter, with respective widths of 511 and 11 data points), and finally subtracting the smoothed light curve from the original. Other “glitches” common to all light curves are removed by scaling each light curve to unit median, computing the median of all scaled light curves, and subtracting a scaled version of this “common component” from each light curve. Bad data points (large scatter in “common component” light curve) are also flagged at this stage.

Long term (stellar) variations are then removed using an iterative clipped non-linear filter (Aigrain & Irwin 2004). First the light curve is pre-filtered with a combined median/boxcar filter (duration 7, 3 samples) to remove short duration glitches and to minimise the removal of signal from transit-like features. A “continuum” is then computed from this pre-filtered curve by iteratively applying a similar median/boxcar filter (duration $2d$, d samples, where d is the trial transit duration), flagging points where the difference between continuum and original is $>3\sigma$, and recomputing the continuum without the flagged points up to 5 times. The σ is robustly re-computed at each iteration from the median of the absolute deviations of the difference signal. The final clipped continuum is subtracted from the original signal and the median level restored to give the filtered (white-noise-like) light curve (see example in Fig. 5).

Detecting the transits: The box-shaped transit finding algorithm of Aigrain & Irwin (2004) is applied to the filtered light curves. This algorithm, based on likelihood maximisation of a box-shaped, periodic transit model, maximises the transit signal-to-noise ratio $S = \sqrt{N_{tr}} \times \Delta F / \sigma$, where N_{tr} is the number of in-transit points, ΔF the transit depth (which is the mean deviation from the median of the in-transit points), and σ the robustly estimated scatter. The parameters are the transit duration, period and epoch. Note that the optimal transit depth is fully determined by the light curve and is thus not a free parameter. The maximum multiple and single transit statistics

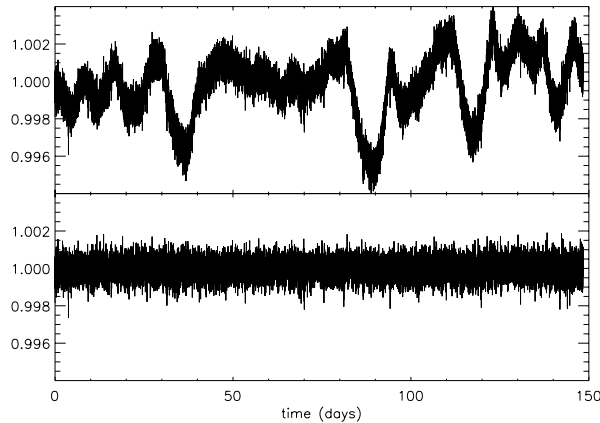


Fig. 5. Light curve 34 before (*top*) and after (*bottom*) iterative non-linear filtering with a trial duration of 3.3 h (Team 5). The Y-axis represents a relative flux.

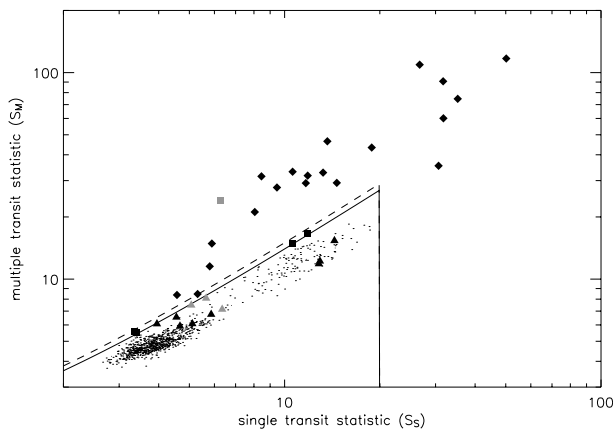


Fig. 6. Candidate selection in the multiple (S_M) versus single (S_S) transit detection statistic plane (trial duration 3.3 h). Solid line: detection threshold. Dashed line: low-confidence threshold. Diamonds: correct detections. Squares: false detections (grey: excluded at the light curve examination stage). Triangles: missed detections (grey: detected using another trial duration). (Team 5)

(S_M and S_S respectively) are then saved and plotted for all light curves (see Fig. 6).

Light curves without events form a clump at low S_S and S_M , while those containing significant residual stellar variations form a tail at high S_S , with $S_M \propto S_S$. A threshold of the form $S_M \geq a + b \times S_S$ was therefore used to pick out periodic events, with $a = 1$ (a makes the threshold more stringent at low S_S 's) and $b = 1.3$. All events below a similar line with $b = 1.4$ are marked as low-confidence events. All light curves with $S_S \geq 20$ are also included in the candidate lists as potentially containing single deep transits.

The long-term variation filtering and transit search are run for trial transit durations of 3.3, 6.7 and 13.3 h, yielding 3 initial lists of 30, 74, and 167 candidates respectively. After examining the corresponding light curves by eye to remove obviously spurious candidates, the final (merged) list contains 31 candidates, of which 6 are low-confidence detections ($S_M \leq 1.0 + 1.4 \times S_S$), 5 are identified as eclipsing binaries due to visible secondary eclipses, 1 as a triple star system and 1 as showing only sinusoidal variations (no transits).

The actual duration of transit candidates is estimated as the full-width at half-minimum of the transits in the phase-folded, filtered light curve. If that differed from the trial duration, the filtering is re-run using the measured duration to obtain a better transit depth measurement; period and epoch were deduced from the transit search itself. Along with the transit search, a search for periodic variations with $0.5 \leq P \leq 4$ days is run by sine-fitting, providing improved period estimates for the stellar variables identified by the transit search, and one additional detection of sinusoidal variation.

4. Comparison and analysis

Analysis of the results is performed in two steps: first, an objective comparison of the individual results and second, comments by each subteam on its own performance.

4.1. “Blind” analysis

Tables 1 and 2 give details of the detection ability of each team for each transit and other contaminating events. From direct comparison of the individual results we observe that:

- Nine transits were detected by all teams (ID numbers 34, 85, 207, 390, 460, 474, 533, 835, 917), and are clearly validated by 5 independent detections. The measured parameters are very similar, except for the transit duration, whose estimation probably depends on the measurement protocol. Periods are always estimated with a $<0.1\%$ dispersion around the actual value. A Jupiter-type planet with two transits is detected around an $m_V = 16$ G dwarf star. The smallest planet detected by all teams (ID 390) has a radius of $2.4 R_E$, a period of 8 days, and it orbits a bright G-type dwarf star ($m = 12$).
- Seven transit events were not detected by any team (ID numbers 317, 326, 575, 618, 681, 715). One (ID 715) corresponds to a giant planet in grazing eclipse on a strongly variable, faint and large-radius star. The other non-detected transits correspond to small planets (1.8 to $2.5 R_E$) and possibly illustrate the detection limit expected for CoRoT (Fig. 7 and next section).
- Five transits were detected by only some of the teams; ID 168, 537, 613, and 624 were detected by 2 to 4 teams. They correspond to star-planet systems with more than 10 transit events in the total light curve duration. Finally, ID 915, detected by two teams, had a “trick”: it is a 58-day period planet around a binary star; only two teams saw the planet itself, whereas the binary was easily detected by all teams.
- Nine false positives were announced. Teams 2, 4, and 5 detected 1, 3, and 5 false events, respectively, while Teams 1 and 3 did not detect any. It never happens that a false event is detected by two independent teams on the same light curve, which probably represents the most remarkable result of this study. This result is very positive as it clearly proves that false alarms are method-dependent. It is probable that using such independent multiple analyses will

considerably reduce the false-alarm rate with real data as well. Again, this result is not specific to CoRoT.

- Six eclipsing binaries and variable stars were detected by all teams but sometimes wrongly identified as planetary transits, when they are in reality background or grazing eclipsing binaries (ID 31, 249, 386, 919, 937, 985). Note that such cases should be identified by spectroscopic or/and photometric ground-based follow-up.
- Three of the contaminating events are not detected. ID 271 and 650 are non-periodic variables that thus do not affect the transit search. ID 518 is the only eclipsing binary that is never detected, but it has only two shallow transits over the 150-day period (grazing binary with an M-dwarf companion).

Quantitative comparisons of the computing time required do not prove major discrepancies between the teams; moreover, it was not always the priority of the detection teams to minimize the analysis time, so that a crude comparison is not realistic at this level. Despite this, no analysis method requires a computing time which is incompatible with the data flow expected from space transit searches. Also, none of the methods described in this paper is highly sensitive to the short and frequent gaps in the data due to the SAA. Finally, the strong residual scattered-light noise never limits the detection, whatever method is used, even when an over-correction of the scattered light led to a periodic, negative signal, more easily confused with transit signatures.

The results show that the simple correlation method proposed by Team 1 is already an effective detection tool (22 detected events over 38 inserted). It also appears that Teams 3 and 5 have detected significantly more transit events than the other three teams (26 detected events). Team 3, moreover, had no false positive, compared to five false positives for Team 5. Team 5 could have included less false positives with a higher threshold (see Fig. 6), but the method of Team 3 has the additional advantage of a very natural way of setting the threshold (Fig. 4, bottom). This points towards a greater robustness of the method used by Team 3. It confirms that the BLS algorithm is more sensitive to faint transits, a result which also shows up in the theoretical comparison performed by Tingley (2003) or in the recent re-analysis of the OGLE data (Udalski et al. 2003). The better results of Team 3 could also be due to a more efficient detrending technique.

4.2. Derived detection limits of CoRoT

Figure 7 shows the three types of results (5 detections, 1 to 4 detections, 0 detection) against the main parameters that affect detection sensitivity: transit depth d and number of transits n in the light curve. The non-detected events are all situated below the empirical detection curve $d \simeq 2 \cdot 10^{-3} n^{-1/2}$, except one which corresponds to a difficult case described earlier (ID 715). The detection capability of CoRoT derived from this blind test analysis (where r is the planet radius and R the star radius) are:

- $n = 50$ (period < 3 days): $r > 0.017R$ is detected.
- $n = 15$ (period < 10 days): $r > 0.023R$ is detected.
- $n = 3$ (period < 50 days): $r > 0.034R$ is detected.

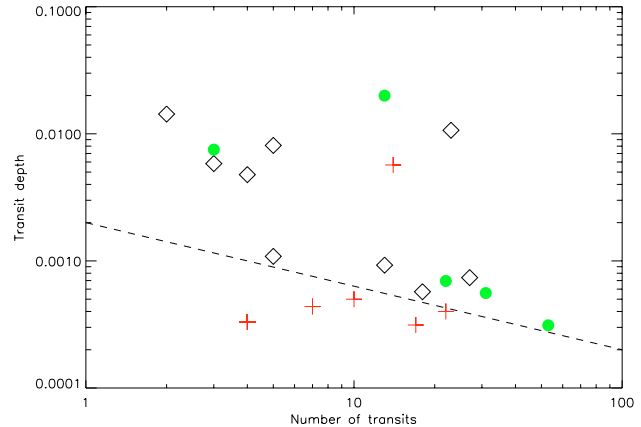


Fig. 7. Depth of the transits versus number of transits. Plus signs show the non-detected events, diamonds show the events detected by five groups independently, and filled circles correspond to 1 to 4 detection occurrences. The dashed line thus shows the border of the simulated CoRoT detection limit (proportional to $n^{-1/2}$). The only plus sign above the detection line is a grazing planet on a faint fluctuating star.

Table 3. Minimum planet radius for F0V, G0V, K0V, and M0V stars in unit of Earth radius, corresponding to the empirical detection curve estimated by the blind test, which possibly overestimates the minimal radius of the detected planets at the longest periods. The star radii are from Allen (2000), i.e. 1.5, 1.1, 0.85, and 0.6 solar radius, respectively.

Period	F0V	G0V	K0V	M0V
50-day	5.6	4.0	3.2	2.2
10-day	3.75	2.75	2.1	1.5
3-day	2.8	2.0	1.6	1.1

This “law” may overestimate the minimum detected size when the number of transits is small. It also does not account for the detrending of systematics, which may have an unpredictable impact on the detection.

Table 3 gives the corresponding values of the minimal detected planet size for four types of parent stars, F0V, G0V, K0V, and M0V.

4.3. Lessons learned per team

4.3.1. Team 1

Non-detections: The residual pollution by stellar micro-variability may explain some non-detections. Light curve ID 915, where the transit is mixed with a fast eclipsing binary, was missed because such a possibility was not considered. It would, however, not have been detected, since the small event was embedded in a secondary detection peak. It shows one of the detection limits of the method.

False detections: None, due to the low sensitivity limit of the method and to visual elimination steps.

Prospects for further improvements: The periodicity of the transit signal could be used in the detection. The removal of the instrumental noise could be improved with a tool such as

Principal-Component Analysis. Filtering the long-period variations would also be useful.

4.3.2. Team 2

Non-detections: For most of the non-detections some individual events were detected on a very low confidence level, but most signals were below the detection limit of our routine. To detect these events a search in folded light curves is necessary to improve the S/N ratio of the signals. For ID 168, several transit-shaped events were detected with medium confidence level, but many resulted from the variability of the stars, thus confusing the detection algorithm searching for a periodicity. Consequently the treatment of the variability of the stars and the robustness of the periodicity search has to be improved.

False detections: Only one false detection was made by the team. In light curve ID 213, simulating a faint star, a false transit event was found. This detection had the lowest confidence level of all our detections (3σ).

Prospects for further improvements: A first step would be to search for transits in folded light curves to be able to detect fainter transits in noisier environments. We also plan to test a Fourier analysis and remove frequencies that can be identified as instrumental noise. The deformation of the transit events can be prevented that way. Additionally, the light curves of variable stars have to be analysed more carefully to reduce periodic variations that can confuse the detection algorithm.

4.3.3. Team 3

Non-detections; False detections: We did not detect six of the simulated transits, and we had no false detections. Examining the non-detections reveals that the correct peak appears in the SR for two of them – ID numbers 317 and 575. In order to detect light curve 575 we would have to lower the detection threshold to 5.6, which would have resulted in 86 false detections! The score of light curve 317 was 3.9, which would mean an unrealistic lowering of the threshold. Most of the non-detected (ID numbers 317, 326, 575, 618, and 681) transits corresponded to planets with the smallest radii – less than $0.023 R_{\odot}$. Light curve 715 was affected by the presence of two different periods and escaped detection. Visual inspection of some of the non-detected transits also suggests that maybe some additional variability still exists after removal of the harmonics, but this effect has not been quantified yet.

Prospects for further improvements: The detrending process may benefit significantly from new procedures recently developed for systematic-effect removal (Tamuz et al. 2005; Kovács et al. 2005). This procedure may remove a significant part of the stellar variability, but also some systematic effects that were not modelled in this exercise.

The detection stage may benefit from the correction to the BLS algorithm proposed by Tingley (2003). In theory, the corrected BLS should be somewhat more powerful in

distinguishing between a transit signal and random noise, thus improving detection ability. Another improvement in application of the BLS may be related to a better sampling of the frequency space, fine tuning of the algorithm parameters (maximum transit width, bin width, etc.), or better adjustment of the SR function. Finally, one could also make a 2-D search that looks at both the “SDE” and “DDE” parameters of Kovács et al. (2002) to check whether this allows some gain in the detection capability.

4.3.4. Team 4

Non-Detections: A posteriori analyses show that the algorithm cannot detect a planet with radius less than $0.02 R_{\odot}$ or when the noise (likely stellar noise) is so strong that the denoising algorithm starts modifying the transit itself.

False detections: Among the detected signals, three of them turned out to be false detection (IDs 701, 703, 983). The case of ID 983 corresponds to a discontinuity of the light curve produced by the stellar variability simulation (Sect. 2.2.1). In the other cases (IDs 701, 703), transit features were erroneously identified with the peak sorting method due to a random and unlucky location of the peaks in the convolution curve M . This kind of false detection should, however, not be specific to our algorithm. Our best results were obtained when the matched filter was associated with a peak selection by Fourier correlation. No false alarm is found in this case, while selection by peak sorting can lead to a number of false alarms due to ambiguities with noise artefacts. Finally, the number of false positives does not change with the detection threshold, which is automatically optimized from an estimate of the noise in the input signal. The threshold thus strongly depends on the quality of the detrending process.

Prospects for further improvement: The method developed in the present exercise can certainly be improved for higher noise level. A new filter based on image processing is presently being tested to improve the detection capacity. It is developed on the same grounds as the detrending tool presented in Sect. 3.4.

Another issue is the actual robustness of the algorithms to periodicity changes, due for example to binarity, secondary planets, or residual instrumental drift. This question has not been addressed in the present exercise since transit signals were assumed strictly periodic.

4.3.5. Team 5

Non detections; false detections: Three noteworthy points emerge from Fig. 6. First, the tail of small dots with high S_S and $S_M \simeq S_S$ (i.e. no clear evidence of a periodic signal) represents light curves with residual non-periodic variations. In some cases, these were on too short a timescale to be fully filtered. However, some of the stellar light curves built with the method of Sect. 2.2.1 contained transit-like features, that are probably artefacts. Second, all the false alarms that escaped removal at the light curve examination

stage were low-confidence events. Had the higher threshold of $S_M \geq 1.0 + 1.4S_S$ been used, there would have been no false alarms. This is the result of a conscious decision to include low-confidence detections, in order to pinpoint the detectability limit. Third, there were 11 missed detections for the 3.3 h trial duration. Of those, 3 were detected at other trial durations and one was a planet orbiting a binary causing non-periodic events, which we didn't tailor our search to. By lowering the threshold, we could have detected ID 575 and ID 317 at the cost of 1 and 17 additional false alarms, respectively. Of the others, 3 produced S_M 's close to our algorithm's white-noise limit of ~ 6 (Aigrain & Irwin 2004) and are thus probably beyond the detectability limit of CoRoT. The other 3 were highly variable light curves.

Prospects for further improvements: Future improvements will include refinement of the detrending stages, of the choice of threshold through Monte Carlo simulations, and of the post-detection transit characterisation.

5. Conclusion and prospects

The present paper reports on the first “real size” blind study of a large set of synthetic stellar light curves by five independent expert teams to detect planetary transits. Different transit detection methods were tested, ranging from ad-hoc procedures built from scratch to seasoned algorithms used in existing ground-based surveys (OGLE and BEST) and for future space missions (CoRoT and Kepler). Together, they offer a global view of the most possible approaches to light curve detrending and transit detection algorithms. The results show that specialised algorithms can detect transit signals down to the noise limit. It turns out that rather simple procedures can identify most detectable transit signals, but that the additional effort put into refined detection algorithms is really worthwhile to recover the shallower detectable transits – potentially the most interesting ones.

The results also show that false detections may not be a major difficulty when various detection methods are applied, since no false event was ever detected twice independently in the simulation. Also, one method (harmonic-fitting filtering plus BLS detection) does not suffer from any false detections within the synthetic sample. We note that stellar micro-variability limits the transit detection only when its standard deviation is larger than 0.5% and its main frequency is around 0.1 day^{-1} . In most cases, stellar micro-variability such as simulated here (Sect. 2.2) is not the main limitation, mostly because the fluctuation frequencies are not in the domain of the transit duration, and the amplitude is usually low. This result compares well with the conclusions of Jenkins (2003) and are important in the context of space transit detection missions. Of course, this is true only as long as activity models based on the solar case correctly describe other stars. In the next few years, space astero-seismology missions may provide better constraints on stellar micro-variability on timescales of a few hours.

The present study shows that the detrending method is almost as important for detecting faint transits such as the detection algorithm itself. Precise detrending processes can cancel

almost all the variability and reflected light contamination. On the other hand, artefacts of the detrending can cause spurious transit detections. The relative importance of detrending and detection could be quantified by coupling the detrending and detection phases between the five algorithms, but was not attempted in the present study.

Processing of real data from space will suffer more systematic effects than those introduced in this exercise, due to temperature cycles, pointing jitter, or scattered light gradients along the detector. In ground-based transit surveys, systematics are mainly due to fluctuations caused by Earth's atmosphere. Future work will include a comparative study of gains from the correction of systematics using comparison stars, as recently proposed by Tamuz et al. (2005) and Kovács et al. (2005).

The characterisation of transits (shape, radius ratio, orbital inclination, etc.) requires an entirely different set of analysis tools and no particular insight was obtained about it from the detection simulation – apart from confirming that eclipsing binaries can easily be confused with planetary transits.

Some of the algorithms used here focus on the detection of individual transits, as well as on strictly periodic signals. Detection of not strictly periodic transit signals is an issue that was not considered here. In most realistic cases (two planets, circumbinary planet), the transits will be very nearly periodic and the algorithms for periodic signals will probably be able to detect them. But among the algorithms that are studied here, at least two have reached “maturity” for monochromatic light curves without a priori information. Continuation of this study could consider including more information: e.g. chromatic light curves (CoRoT), colour or spectroscopic information about the target star. It could also include other instrumental contents (Kepler, Eddington) and a refinement of stellar micro-variability in the frequency-amplitude parameter zone where it may mimic transit features.

The 999 light curves produced and a table with the parameters used are available to the community by request to the authors for testing and improving other detection algorithms.

Another by-product of our blind comparison of detecting transits in light curves simulated as CoRoT data is a refined estimate of the detection limitation on this instrument under development: a 3-day $1.1 R_E$ planet around an M0 dwarf star would probably be detected. CoRoT would also detect the transits of a planet like μ Arae d, the 14.5-Earth mass planet with 9.55-day period recently discovered in radial-velocity surveys (Santos et al. 2004), if it is larger than $2.7 R_E$, i.e. with a density up to that of terrestrial planets.

Acknowledgements. We are grateful to the CoRoT PI Annie Baglin and to the whole CoRoT/Exoplanet Working Group for their support and fruitful discussions during this exercise. S.Z. wishes to acknowledge support by the European RTN “The Origin of Planetary Systems” (PLANETS, contract number HPRN-CT-2002-00308) in the form of a fellowship. Finally, we express our acknowledgements to the anonymous referee for his/her detailed reading and many interesting suggestions.

References

- Aigrain, S., & Favata, F. 2002, *A&A*, 395, 625
- Aigrain, S., Favata, F., & Gilmore, G. 2004, *A&A*, 414, 1139
- Aigrain, S., & Irwin, M. 2004, *MNRAS*, 350, 331
- Allen, C. 2000, *Astrophysical Quantities*, ed. A. N. Cox
- Andersen, B. N., Leifsen, T., & Toutain, T. 1994, *Sol. Phys.*, 152, 247
- Andreasen, J. 1988, *A&A*, 196, 159
- Auvergne, M., Boisnard, L., & Buey, J.-T. 2003, *SPIE*, 4853, 170
- Baglin, A. 2003, *Adv. Sp. Res.*, 31, 345
- Barban, C., Goupil, M. J., & Van't Veer-Menneret, C. 2003, *A&A*, 405, 1095
- Boisnard, L., & Auvergne, M. 2004, in 55 th IAF Congress, October 4–7, Vancouver, IAC-04-Q.1.01
- Bordé, P., Rouan, D., & Léger, A. 2003, *A&A*, 405, 1137
- Borucki, W., Koch, D., Boss, A., et al. 2004, in Second Eddington Workshop: Stellar structure and habitable planet finding, ed. F. Favata, S. Aigrain, & A. Wilson, ESA SP-538, 177
- Bouchy, F., Pont, F., & Santos, N. 2004, *A&A*, 421, L13
- Brown, T. 2003, *ApJ*, 593, 125
- Chapman, G. A., Cookson, A., & Dobias, J. 1997, *ApJ*, 482, 541
- Deeg, H. 1999, *Universal Transit Modeler*, www.iac.es/galeria/hdeeg/idl_hans_lib/utm/
- Defaÿ, C. 2001, Ph.D. Thesis
- Defaÿ, C., Deleuil, M., & Barge, P. 2001, *A&A*, 365, 330
- Doyle, L. R., Deeg, H. J., Kozhevnikov, V. P., et al. 2000, *ApJ*, 535, 338
- Favata, F. 2004, in Second Eddington Workshop: Stellar structure and habitable planet finding, ed. F. Favata, S. Aigrain, & A. Wilson, ESA SP-538, 3
- Guis, V., & Barge, P. 2004, *PASP*, accepted
- Jenkins, J., Doyle, L., & Cullers, D. 1996, *Icarus*, 119, 244
- Konacki, M., Torres, G., Jha, S., & Sasselov, D. 2003, *Nature*, 421, 507
- Konacki, M., Torres, G., & Sasselov, D. 2005, *ApJ*, submitted
- Kovács, G., Bakos, G., & Noyes, R. 2005, *MNRAS*, 356, 557
- Kovács, G., Zucker, S., & Mazeh, T. 2002, *A&A*, 391, 369
- Lanza, A. F., Rodonò, M., Pagano, I., Barge, P., & Llebaria, A. 2003, *A&A*, 403, 1135
- Lanza, A. F., Rodonò, M., & Pagano, I. 2004, *A&A*, 425, 707
- Llebaria, A., Guterman, P., & Ollivier, M. 2003, *SPIE*, 5170, 155
- Mandel, K., & Agol, E. 2002, *ApJ*, 580, L171
- Messina, S., Pizzolato, N., Guinan, E., & Rodonò, M. 2003, *A&A*, 410, 671
- Pont, F., Bouchy, F., Queloz, D., et al. 2004, *A&A*, 426, L15
- Press, W., Teukolsky, S., Vetterling, W. T., & Flannery, B. P. 1993, *Numerical Recipes in C*, 2nd ed. (Cambridge Univ. Press)
- Santos, N., Bouchy, F., Mayor, M., Pepe, F., & Queloz, D. 2004, *A&A*, 426, L19
- Tamuz, O., Mazeh, T., & Zucker, S. 2005, *MNRAS*, 356, 1466
- Tingley, B. 2003, *A&A*, 408, L5
- Udalski, A., Paczynski, B., Zebrun, K., et al. 2002a, *Acta Astron.*, 52, 1
- Udalski, A., Zebrun, K., Szymanski, M., et al. 2002b, *Acta Astron.*, 52, 115
- Udalski, A., Pietrzynski, G., Szymanski, M., et al. 2003, *Acta Astron.*, 53, 133
- Udalski, A., Szymanski, M. K., Kubiak, M., et al. 2004, *Acta Astron.*, 54, 313
- Walker, G., Matthews, J., Kuschnig, R., et al. 2003, *PASP*, 115, 1023
- Wichmann, R. 1998, *Nightfall* www.lsw.uni-heidelberg.de/users/rwichman/Nightfall.html