

Testing the recovery of stellar rotation signals from *Kepler* light curves using a blind hare-and-hounds exercise

S. Aigrain,^{1★} J. Llama,^{2★} T. Ceillier,³ M. L. das Chagas,⁴ J. R. A. Davenport,⁵
R. A. García,³ K. L. Hay,^{6,7} A. F. Lanza,⁸ A. McQuillan,⁹ T. Mazeh,⁹
J. R. de Medeiros,⁴ M. B. Nielsen¹⁰ and T. Reinhold¹⁰

¹Department of Physics, University of Oxford, Oxford OX1 3RH, UK

²Lowell Observatory, 1400 West Mars Hill Rd, Flagstaff, AZ 86001, USA

³CEA/DSM, F-91191 Gif-sur-Yvette, France

⁴Departamento de Física Teórica e Experimental, Universidade Federal do Rio Grande do Norte, 59078-970 Natal, Brazil

⁵Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195, USA

⁶Blackett Laboratory, Imperial College London, London SW7 2AZ, UK

⁷SUPA, School of Physics and Astronomy, North Haugh, St Andrews, Fife KY16 9SS, UK

⁸INAF/Catania Observatory, I-95123 Catania, Italy

⁹School of Physics and Astronomy, Tel Aviv University, 69978 Tel Aviv, Israel

¹⁰Institut für Astrophysik, Universität Göttingen, D-37077 Göttingen, Germany

Accepted 2015 April 13. Received 2015 March 26; in original form 2014 November 9

ABSTRACT

We present the results of a blind exercise to test the recoverability of stellar rotation and differential rotation in *Kepler* light curves. The simulated light curves lasted 1000 d and included activity cycles, Sun-like butterfly patterns, differential rotation and spot evolution. The range of rotation periods, activity levels and spot lifetime were chosen to be representative of the *Kepler* data of solar-like stars. Of the 1000 simulated light curves, 770 were injected into actual quiescent *Kepler* light curves to simulate *Kepler* noise. The test also included five 1000-d segments of the Sun's total irradiance variations at different points in the Sun's activity cycle. Five teams took part in the blind exercise, plus two teams who participated after the content of the light curves had been released. The methods used included Lomb–Scargle periodograms and variants thereof, autocorrelation function and wavelet-based analyses, plus spot modelling to search for differential rotation. The results show that the ‘overall’ period is well recovered for stars exhibiting low and moderate activity levels. Most teams reported values within 10 per cent of the true value in 70 per cent of the cases. There was, however, little correlation between the reported and simulated values of the differential rotation shear, suggesting that differential rotation studies based on full-disc light curves alone need to be treated with caution, at least for solar-type stars. The simulated light curves and associated parameters are available online for the community to test their own methods.

Key words: methods: data analysis – techniques: photometric – surveys – stars: rotation – starspots.

1 INTRODUCTION

The rotational modulation of magnetically active regions on the surface of stars produces quasi-periodic variations in their disc-integrated apparent brightness, which have been used for decades to measure rotation periods for young, active stars. The exquisite photometric quality and baseline of space-based telescopes such as *Kepler*, *Convection Rotation and Planetary Transits (CoRoT)*

and *Microvariability and Oscillations of Stars (MOST)* have made it possible to do this for tens of thousands of moderately active field stars, many of which display sub-millimagnitude (sub-mmag) variations that would have been undetectable from the ground. The resulting, extensive rotation period catalogues represent an exciting opportunity to test and refine our understanding of stellar angular momentum evolution, and to develop efficient methods for estimating the ages of field stars based on their rotation rate (gyrochronology). Furthermore, for active stars, the presence of multiple and/or evolving periodicities in the light curves can be used to study phenomena such as differential rotation and active region evolution,

* E-mail: suzanne.aigrain@astro.ox.ac.uk (SA); joe.llama@lowell.edu (JL)

which are important tracers of the dynamo mechanisms driving magnetic field generation in those stars.

To make full use of this new information, however, requires a detailed understanding of the reliability and completeness of measurements of mean and differential rotation rates derived from these full-disc space-based light curves. This paper reports on our attempts to address that problem through a hare-and-hounds exercise, where simulated starspot signals were injected into *Kepler* light curves of otherwise quiet stars, and several teams independently attempted to measure rotation and differential rotation rate.

1.1 Field star rotation studies in the *Kepler* era

While ground-based time-domain photometric surveys have yielded numerous rotation period measurements for stars in young clusters (see e.g. Irwin & Bouvier 2009; Bouvier et al. 2014, and citations therein), the limited precision and time sampling achievable from the ground are typically insufficient to detect rotational modulation in older, less active, slowly rotating field stars. For many years, the Mount Wilson program (Wilson 1978; Vaughan et al. 1981; Baliunas, Sokoloff & Soon 1996), which monitored emission in the cores of the Ca II H&K lines, for large numbers of Sun-like stars over a 20 yr period, with a typical time sampling of one point every few days for each star, was the main source of rotation periods measurements for field stars.

This changed with the advent of space-based photometric monitoring platforms such as *MOST*, *CoRoT* and *Kepler*, which have collectively gathered sub-mmag precision photometry over periods ranging from weeks to years, for over two hundred thousand stars. These, and particularly the *Kepler* light curves, which have the longest baseline and highest precision, have enabled a number of large-scale rotation period studies for field main-sequence stars (e.g. Basri et al. 2011; Affer et al. 2012; McQuillan, Aigrain & Mazeh 2013a; McQuillan, Mazeh & Aigrain 2013b, 2014; Nielsen et al. 2013; Reinhold, Reiners & Basri 2013; Garcia et al. 2014). These studies have produced intriguing results, such as the existence of a bimodality in the period distribution for K and M stars, the sharp upper envelope of the period–mass relation and the existence of a small but significant population of rapid rotators ($P < 10$ d). Some of these studies relied on the Lomb–Scargle (LS) periodogram, which has been widely used for rotation period measurements for decades, while others have introduced new methods based for example on the autocorrelation function (ACF) or a wavelet transform of the light curves.

One of the goals of the present paper was to test the sensitivity and reliability of these different methodologies on realistic but simulated *Kepler* data, where the ground truth is known.

1.2 Differential rotation measurements

Going beyond measurements of the mean or overall rotation periods, a number of studies have also looked for, and reported signs of, surface differential rotation in long-baseline, disc-integrated stellar light curves. The presence of multiple, close but distinct peaks in the Fourier transform or periodogram, and the associated ‘beat’ patterns, which are often observed in the light curves of active stars, are often interpreted as signposts of differential rotation.

Efforts to measure differential rotation from light curves can be divided into two broad classes. One class uses the broadening or splitting of peaks in the Fourier transform or periodogram of the chromospheric activity indicator measurements or light curve (see e.g. Baliunas et al. 1985; Donahue & Baliunas 1992; Reinhold

et al. 2013). The other class relies on fitting a spot model to the light curve, where the spots may have different rotation periods (see e.g. Strassmeier & Bopp 1992; Croll et al. 2006; Lanza et al. 2009; Gondoin et al. 2010; Lanza, Das Chagas & De Medeiros 2014). Once again, these differential rotation studies have yielded intriguing results, such as a trend for increased rotational shear ($\delta P/P$) for increased stellar effective temperature.

The two classes of methods used have different advantages and disadvantages. Fourier-domain methods can be applied systematically for large numbers of stars, but the relationship between the measured peak width or separation and physical quantities of interest such as the rotational shear is non-trivial. Spot modelling is computationally intensive, and therefore restricted to smaller samples, but in principle it has a more direct physical interpretation. In practice, however, spot modelling is highly degenerate, even when using only a few spots. This was recently demonstrated by Walkowicz, Basri & Valenti (2013), who identified degeneracies between stellar inclination and spot latitudes, and showed that differential rotation may be missed if the light curve is dominated by one large active region.

Perhaps even more importantly, both types of approach suffer from a further limitation, which was not explored by Walkowicz et al. (2013): differential rotation is very difficult to distinguish from spot or active region evolution, which can induce very similar beat patterns and broadened or split peaks in the periodogram, and at the very least is an additional source of noise when identifying and modelling individual peaks in the periodogram, or individual star spots (Lanza, Rodono & Zappala 1994). The second goal of the present study was to test the reliability of differential rotation results reported in the recent literature for stars exhibiting low and moderate levels of activity, again by using realistic simulations including both differential rotation and spot evolution.

2 THE HARES: SIMULATING THE LIGHT CURVES

This section describes the process by which the light curves used in the blind exercise were simulated. Note that we do not distinguish between ‘active regions’ and ‘spots’ in the simulation: each active region is effectively assumed to consist of a single dark spot whose area (or equivalently, contrast) evolves over time.

2.1 The spot emergence model

The first task in simulating the light curves is to compute the time-dependent distribution of spots on the stellar surface. On the Sun, the rate and latitude of emergence of sunspots varies over an 11-yr cycle, giving rise to what is known as the ‘butterfly pattern’ first discovered by Maunder (1904, 1922) (for a recent overview of the solar cycle, see Hathaway 2010). At the beginning of the cycle, spots emerge at $\theta \pm 35^\circ$. As the cycle progresses, the spots begin to emerge closer to the equator. During that time, the rate at which spots emerge first rises gradually, then decays again. After 11 yr, the cycle repeats. The nature of this cycle is believed to be intrinsically linked to the solar dynamo (Berdyugina 2005).

Very little is known about the distribution and evolution of spots on stars other than the Sun. Doppler imaging, and Zeeman–Doppler imaging have revealed that, on rapidly rotating stars, starspots are not restricted to two distinct latitude bands. Instead, these stars appear to host spots at all latitudes on the stellar disc (Donati, Collier Cameron & Petit 2003). Indeed, observations of the rapid rotator AB Doradus over many years appear to show no pattern to the

Table 1. Range and distributions used for the simulation parameters.

Parameter	Range	Distribution
Activity level A	$0.3\text{--}3 \times \text{solar}$	Log uniform
Activity cycle length C_{len}	1–10 yr	Log uniform
Activity cycle overlap C_{over}	0.1–3 yr	Log uniform
Minimum spot latitude θ_{min}	$0^\circ\text{--}40^\circ$	Uniform
Maximum spot latitude θ_{max}	$\theta_{\text{min}}\text{--}80^\circ$	Uniform in $(\theta_{\text{max}} - \theta_{\text{min}})^{0.3}$
Inclination i	$0^\circ\text{--}90^\circ$	Uniform in $\sin^2 i$
Equatorial rotation period P_{eq}	10–50 d (90 per cent)	Log uniform
	1–10 d (10 per cent)	Log uniform
Rotational shear $\delta\Omega/\Omega_{\text{eq}}$	0.1–1 (2/3)	Log uniform
	0 (1/3)	
Decay time-scale τ	$(1\text{--}10)P_{\text{eq}}$	Log uniform

Table 2. Parameters of the simulated light curves (full version available online in machine-readable form). For a description of the last three columns, see Section 4.1.

N	A^a	C_{len} (yr)	C_{over} (yr)	θ_{min} ($^\circ$)	θ_{max} ($^\circ$)	R^b	$\sin i$	P_{eq} (d)	$\delta\Omega/\Omega_{\text{eq}}$	τ_{dec} (P_{eq})	KID ^c	P_{obs} (d)	$P_{\text{obs,min}}$ (d)	$P_{\text{obs,max}}$ (d)
1	0.39	1.12	0.27	13	76	0	0.19	1.10	0.46	8.2	893468	1.51	1.26	1.77
2	0.45	9.52	0.19	20	63	0	0.99	1.58	0.40	7.3	1163449	1.87	1.75	2.02
3	1.34	1.54	1.53	33	74	0	0.34	17.04	0.00	1.02	1431091	17.04	17.04	17.04

Notes. ^a A is relative to solar.

^bIf R is 0, spots follow a butterfly pattern, otherwise they are randomly distributed in latitude.

^cA KID of 0 means that the noise-free light curve was used, while the entry ‘Sun’ in this column means that the light curve is based on a segment of the Sun’s observed total irradiance variations (see Section 2.4 for details).

distribution of star spots. Rather, the observations show magnetic activity at all latitudes in all the observations (Collier Cameron, Donati & Semel 2002; Jeffers, Donati & Collier Cameron 2007).

The latitude distribution of starspots and its evolution over the star’s activity cycle has a direct influence on the disc-integrated brightness fluctuations, and hence on the extent to which information about rotation and differential rotation can be recovered from the light curves. We therefore took care to incorporate a wide variety of butterfly-like patterns in the simulated light curves, as well as some cases with no explicit relationship between spot latitude and the phase of the activity cycle. Our starting point was the code developed by Llama et al. (2012), itself derived from the work of Mackay et al. (2004). This code, initially designed to simulate the emergence of magnetic flux on the surface of the Sun, allows for the creation of butterfly patterns of varying activity levels, cycle lengths and distribution patterns. It was already used by Llama et al. (2012) to simulate starspot distribution patterns in the context of *Kepler* light curves, specifically light curves containing planetary transits. Their goal was to investigate whether stellar butterfly patterns could be revealed by looking for phase changes in the appearance of bumps in transit light curves of stars hosting a planet in a misaligned orbit.

The names, ranges and distributions of the input parameters of the starspot generator are listed in the first five rows of Table 1. They are as follows: the star’s overall activity level A , relative to that of the Sun; the duration of the activity cycle, C_{len} , in years; the amount of overlap between consecutive activity cycles C_{over} , in years and the minimum and maximum spot latitudes, θ_{min} and θ_{max} . For each simulated star, the model generates a list of spot emergence times, t_k , longitudes ϕ_k and latitudes $\theta_{k,0}$, as well as peak magnetic flux density $B_k^{(\text{max})}$ (in Gauss) at the centre of each spot. The activity rate, A , controls the overall rate of starspot and magnetic flux emergence relative to the Sun: a value of $A = 1$ leads to a total number of spots and flux emergence, integrated over a 27-d period, similar to what is observed on the Sun. The individual parameter values used for

each simulation are listed in Table 2, and the top panel of Fig. 1 shows three examples of the resulting butterfly patterns.

In most cases, we simulate a butterfly pattern by allowing the latitude at which the spots emerge to decrease monotonically from θ_{max} to θ_{min} over the duration of the cycle in two latitude bands (one in the Northern hemisphere and one in the Southern hemisphere), as illustrated in the top panel of Fig. 1. At a given time t in the simulation, the activity cycle phase is given by $\Phi = t \bmod C_{\text{len}}$, where $\Phi \in [0, 1]$. The spot emergence latitude as a function of phase is then given by $\theta(\Phi) = \theta_{\text{max}} - (\theta_{\text{max}} - \theta_{\text{min}})\Phi$. We also assume a latitude scatter in the spot emergence to match the observed solar butterfly pattern (Hathaway 2010). The spots are assumed to emerge at random longitudes, and the number of spots per unit time is defined to be inversely proportional to the square of the spot size (Schrijver & Harvey 1994). For 20 per cent of the cases, however, the spots were simulated to emerge randomly between θ_{max} and θ_{min} . This is denoted by a 1 (rather than a 0) in the column R in Table 2.

At the start of each simulation there are no spots on the stellar surface. To allow the simulation to reach a realistic ‘steady state’, the butterfly patterns were simulated starting N days before the start of the actual simulated light curves, where N was a uniform random number between 200 and 400 d.

2.2 Simulating the photometric signal

After using the model described in Section 2.1 to generate starspot emergence and distribution patterns, the next step is to simulate the photometric signal of the spots on the rotating stellar surface. The names, ranges and distributions of the parameters controlling the light-curve simulations are listed in the remaining rows of Table 1, and their individual values for each simulation in the remaining columns of Table 2, while the remaining panels of Fig. 1 illustrate the process. The parameter ranges and distributions were chosen

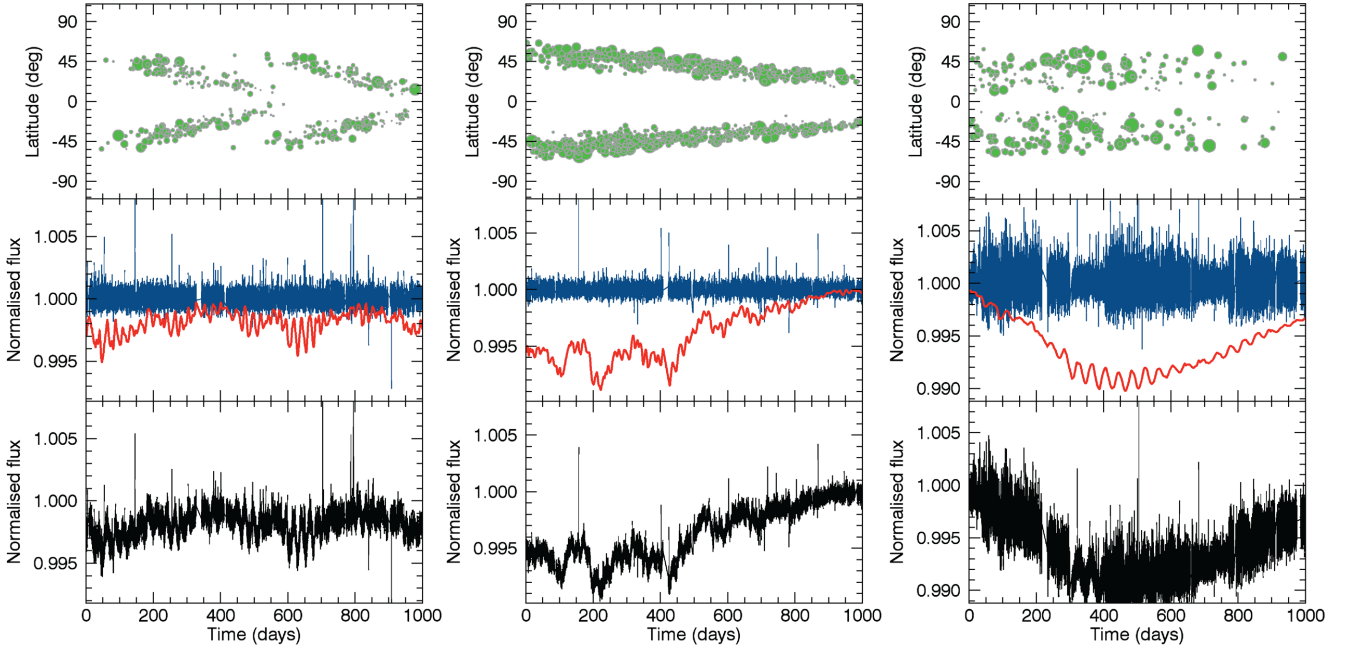


Figure 1. Example simulated light curves (left: no. 9, $P_{\text{eq}} = 23.5$ d, $\tau = 5.2 P_{\text{eq}}$, $\Delta\Omega/\Omega_{\text{eq}} = 0.33$; middle: no. 31, $P_{\text{eq}} = 20.8$ d, $\tau = 1.1 P_{\text{eq}}$, $\Delta\Omega/\Omega_{\text{eq}} = 0.35$; right: no. 12, $P_{\text{eq}} = 18.5$ d, $\tau = 1.6 P_{\text{eq}}$, $\Delta\Omega/\Omega_{\text{eq}} = 0.56$). The top panel shows the emergence times and latitudes of the simulated active regions; the symbol size scales as the peak magnetic flux density (Llama et al. 2012). In the middle panel, the red dots show the simulated, noise-free photometric signal, the blue points the *Kepler* PDC-MAP light curve the signal was injected into. The bottom panel shows the final, noisy light curve included in the blind exercise sample. The first two light curves have similar equatorial rotation periods and differential rotation shears, but the second has more rapid spot evolution. The third has both rapid spot evolution and very strong differential rotation, which appears as a random rather than butterfly-like spot distribution.

to be as realistic as possible while also covering (and hopefully slightly exceeding) the range of detectable signals.

Each star was randomly assigned an inclination i (defined as the angle between the stellar equator and the line of sight) and an equatorial rotation period P_{eq} . The inclinations were drawn from a uniform distribution in $\sin^2 i$. This means that the simulated stars have a tendency to be seen closer to equator, though all orientations are possible. For 90 per cent of the light curves, P_{eq} was drawn from a log uniform distribution between 10 and 50 d, while for the remainder it was drawn from a log uniform distribution between 1 and 10 d. This reproduces approximately the distribution of periods measured for *Kepler* F, G and K main-sequence stars by Reinhold et al. (2013) and McQuillan et al. (2014).

We then implemented Sun-like differential rotation by setting the rotation rate of each spot to

$$\Omega(\theta_k) = \Omega_{\text{eq}} \left[1 - \frac{\delta\Omega}{\Omega_{\text{eq}}} \sin^2(\theta_k) \right], \quad (1)$$

where $\Omega_{\text{eq}} = 2\pi/P_{\text{eq}}$. Two-thirds of the light curves were given fairly strong differential rotation ($\delta\Omega/\Omega$ drawn from a log uniform distribution between 0.1 and 1) while the remaining third had no differential rotation.

We simulate the photometric signature of each spot using the very simple analytical model of Aigrain, Pont & Zucker (2012). The instantaneous relative flux drop caused by a single spot emerging at latitude θ_k and longitude $\phi_k^{(0)}$ is given by

$$\delta F_k(t) = f_k(t) \text{MAX}[\cos \beta_k(t), 0], \quad (2)$$

where $\beta_k(t)$ is the angle between the spot normal and the line of sight:

$$\cos \beta_k(t) = \cos \phi_k(t) \cos \theta_k \cos i + \sin \theta_k \sin i, \quad (3)$$

$\phi_k(t) = 2\pi(t - t_k)/P(\theta_k) + \phi_k^{(0)}$ and $f_k(t)$ is the flux drop that the spot would cause if located at the centre of the stellar disc.

In physical terms, f_k depends on the spot size, shape and contrast relative to the ‘clear’ photosphere. However, in the present study, we set f_k directly, using the following expression to allow for spot evolution:

$$f_k(t) = \begin{cases} f_k^{(\text{max})} \exp[(t - t_k)/(2\tau_{\text{em}})] & \text{for } t < t_k, \\ f_k^{(\text{max})} \exp[(t_k - t)/(2\tau)] & \text{for } t \geq t_k, \end{cases} \quad (4)$$

where

$$f_k^{(\text{max})} \equiv 3 \times 10^{-4} A B_k^{(\text{max})} / \langle B_k^{(\text{max})} \rangle \quad (5)$$

(the constant of proportionality was chosen to give rise to approximately Sun-like levels of variability for $A = 1$), τ is the spot lifetime, which we assume to be the same for all spots on a given star, and τ_{em} was set to either $\tau/5$ or 2 d, whichever was the longer. This two-sided behaviour was chosen to mimic the rapid growth and slower decay of spots observed on the Sun, while avoiding discontinuities in the light curves. A wide range of spot and/or active region lifetimes have been observed on the Sun and on other stars, ranging from a few days for the shortest lived solar active regions to years for particularly active stars (Bradshaw & Hartigan 2014). After some experimentation, we found that drawing τ from a log uniform distribution between 1 and $10 P_{\text{eq}}$ gave rise to light-curve morphologies broadly similar to those observed by *CoRoT* and *Kepler*.

Finally, the total observed flux for the star is simply

$$F = F_0 \left(1 - \sum_{k=1}^K F_k \right), \quad (6)$$

where F_0 is the flux that would be observed in the absence of any spots. We simulated 1000 light curves in this manner, each lasting

1000 d with a cadence of 0.5 h (approximately equal to the *Kepler* long-cadence interval). It is worthy of note that in our simulations the absolute continuum level is known and is used to normalize the entire light curve. In reality, the absolute continuum in *Kepler* light curves is unknown. Since our simulated light curves have been consistently normalized, short stellar cycles could potentially be confused with rotation periods which may not be detected in real *Kepler* photometry. Three examples of the resulting light curves are shown in the middle panel of Fig. 1.

The simplified formalism we have adopted treats spots as point like, ignoring the effects of spot shape, differential projection effects within the spot area and overlapping spots. Limb darkening is also ignored. These approximations could be problematic if one were trying to infer individual spot properties but not for the present study, where the main quantity of interest is the rotation period P . Another important advantage over a more physically realistic but more complex approach is computational speed: it takes only a few seconds to generate each simulated light curve. Finally, we note that our light curves are principally representative of stars exhibiting solar-like activity. From the full *Kepler* data set it is clear that a significant fraction of stars exhibit long-lived active regions that remain for many stellar rotations. Our simulated light curves do not reproduce such persistent spot patterns and so the findings reported in this exercise should not be extrapolated to such stars.

2.3 Injection into PDC-MAP light curves

In order to reproduce the noise properties of the *Kepler* data, the signals as described in Section 2.2 were injected into actual *Kepler* light curves of quiet main-sequence stars. These were randomly selected from a sample of 7788 dwarf F, G, K and M stars in which McQuillan et al. (2014) found no evidence of any periodic variability. Three such examples are shown in the middle panel of Fig. 1 (blue line), with the resulting, combined light curve shown in the bottom panel. Of the 1000 simulated light curves, noise was added in this manner to the first 770. The remaining 230 were kept noise-free, in order to enable us to distinguish between the effects of noise and instrumental systematics, and more fundamental limitations of rotation period recovery in stellar light curves.

The approach adopted here does not reproduce the impact of the *Kepler* systematics removal process on stellar signals. The PDC-MAP pipeline used to remove systematics in *Kepler* data does affect astrophysical signals, albeit at a moderate level for time-scales significantly shorter than a *Kepler* quarter. The PDC-MAP retains signals with periods less than 3 d, whilst periods longer than 20 d are likely removed by the pipeline or distributed to other frequencies (Thompson, Christiansen & Jenkins 2013). Ideally, we would inject the simulated signals into the *Kepler* light curves before running them through the PDC-MAP pipeline, but this was not an option since the pipeline is not publicly available.

2.4 Solar test cases

To test the ability of the participating teams to recover the Sun's rotation period and differential rotation signal, we also included in the blind exercise set five light curves based on the composite total solar irradiance (TSI) time series (Fröhlich & Lean 1998; Fröhlich 2000, 2006), taken during the last solar cycle, and thus based on data from the Variability of Solar Irradiance and Gravity Oscillations (VIRGO) instrument on the *Solar and Heliospheric Observatory* (SOHO) spacecraft. The data were downloaded from the PMOD/WRC website (<http://www.pmodwrc.ch/>), and rebinned

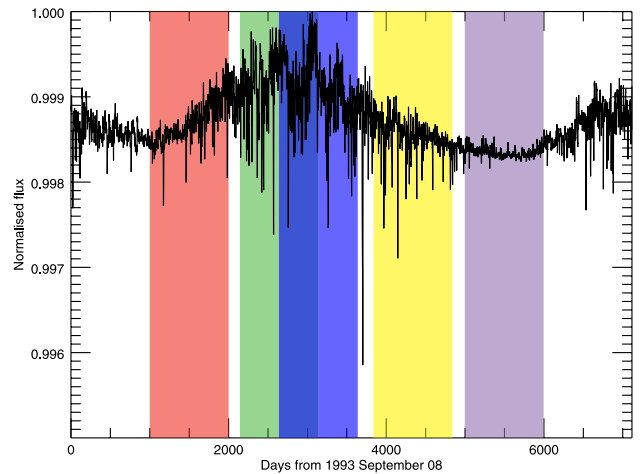


Figure 2. Segments of total solar irradiance variations used to generate the last five light curves included in the set.

to the same cadence as the simulated light curves described in Section 2.2. We then selected five 1000-d long segments spanning the rising phase, maximum, decaying phase and subsequent minimum of the last complete solar activity cycle, as illustrated in Fig. 2.

3 THE HOUNDS: MEASURING ROTATION AND DIFFERENTIAL ROTATION

3.1 The Göttingen/Reinhold team

This team analysed the light curves using the LS periodogram of the full (1000-d) light curves, in a standard pre-whitening approach as described in Reinhold et al. (2013). This approach consists of locating the most significant peak in the LS periodogram, subtracting the best-fitting sinusoid with that period from the data, computing the LS periodogram of the residuals and repeating the process five times. Once the five periods have been identified, a global sine fit is performed according to

$$y_{\text{fit}} = \sum_{k=1}^5 a_k \sin\left(\frac{2\pi}{P_k} t - \phi_k\right) + c, \quad (7)$$

where c is a global offset, and P_k , a_k and ϕ_k are the period, the amplitude and phases of the k th component. The best parameters were found using χ^2 -minimization, resulting in five refined period values for each light curve. The procedure is illustrated for an example light curve in Fig. 3.

The LS periodograms were computed over the period range 0.5–100 d, and the first period, P_1 , was selected within this range. To minimize the number of cases where only the first harmonic of the true rotation period was detected, P_1 was compared to the remaining periods P_k by computing $|2P_1 - P_k|/2P_1$. If this quantity was <0.05 , the highest peak period P_k satisfying this relation, which is most likely to be the true rotation period, was chosen as the main reported period.

The presence of a second period P_2 adjacent to P_1 is indicative of either differential rotation, spot evolution or a combination of both effects. If there was a period P_2 satisfying the relation

$$0.01 \leq |P_1 - P_2|/P_1 \leq 0.30, \quad (8)$$

this period was also reported, in order to test whether the interval between P_1 and P_2 could be used as a measure of differential

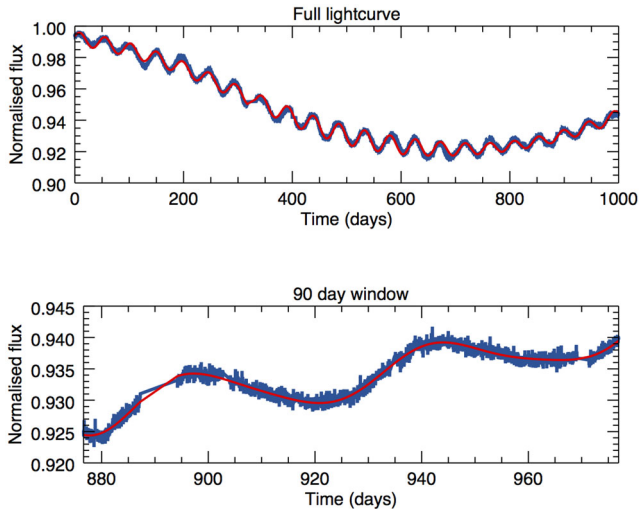


Figure 3. Example multicomponent sinusoidal fits obtained by the Göttingen/Reinhold team for light-curve 9 (see text for details). The top panel shows the results for the full light curve, and the bottom panel for a randomly selected 90-d segment. The data are shown in blue and the fits in red.

rotation. If any P_3 , P_4 and P_5 also satisfied this criterion, the largest and smallest of these were reported as P_{\min} and P_{\max} .

The analysis described above was also applied to a randomly selected 90-d window from each light curve, this time searching for periods up to 45 d to account for the much shorter time span. After visual inspection of the light curves, periodograms and periods returned by the two approaches (full light curve versus 90-d segment), the periods extracted from the full light-curve analysis were deemed to be more reliable in most cases (>90 per cent). The only exceptions were cases where the main period derived from the analysis of the 90-d window was less than 10 d: the analysis of the full light curve appeared less sensitive to short periods. Thus, the final reported periods P_1 (and P_2 if measurable) were chosen from the analysis of the full light curve, except for these short-period cases, where the periods derived from the 90-d segments were reported.

The method was applied automatically to all light curves, and a value was reported for all cases where the main period was <100 d, leading to a total of 840 detections, with at least two periods reported in 545 cases. This detection rate is larger than in Reinhold et al. (2013), where it was ~ 60 per cent. The difference is probably due to the longer period cut-off used here (100 instead of 45 d), which enables the detection of longer periods but also, potentially, spurious signals caused by instrumental effects or long-term spot evolution. To quantify this, 100 light curves were inspected visually, in order to evaluate the expected false-positive rate. About 20 of the visually inspected light curves had detected periods without clear counterpart in the light curve, leading to a predicted false-positive rate of approximately 20 per cent.

3.2 The Göttingen/Nielsen team

This team also used the LS periodogram, but worked with shorter segments of data, equivalent to individual *Kepler* quarters (90 d). The analysis method was based on Nielsen et al. (2013) and was identical to that study in all respects, apart from the number of quarters analysed. Each 1000-d time series was divided into 10 artificial quarters of roughly 90 d. The LS periodogram was computed for

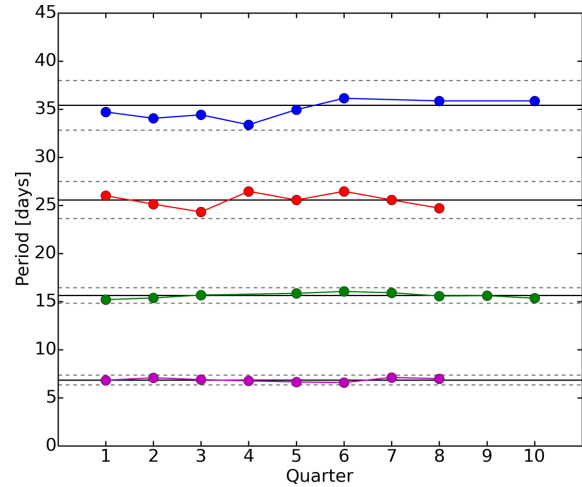


Figure 4. Illustration of the selection criteria used by the Göttingen/Nielsen team. The periods measured in each quarter for four stars are shown in different colours, while black horizontal lines denote the corresponding median values (solid) and ± 3 times the median absolute deviation (dashed). Only periods lying between the dashed lines were retained.

each quarter using the method described in Frandsen et al. (1995). For each quarter, the highest peak was recorded, provided that it was at least higher than four times the white noise level. This level was computed based on the root mean square (rms) of the time series, as described in Kjeldsen & Bedding (1995). The LS periodogram was computed over the range 0.5–90 d, but cases where the period of the highest peak was >45 d were discarded, in an effort to minimize spurious detections from uncorrected instrumental variability or other sources of red noise.

For each star, the median absolute deviation of the periods measured from each quarter was then computed. Any star exhibiting a median absolute deviation above 1 d was discarded. This ensures that only stars with long lived variability are accepted, thus increasing the likelihood that the reported periods are real. This procedure is illustrated in Fig. 4. The remaining stars were required to show at least seven periods within three median absolute deviations of the median. We note that, if differential rotation is present, this would select only light curves where the signal was dominated by spots located at a specific latitude (or rotating at a specific rate) over most of the time span of the data. For stars that satisfy the above criteria, the median of the detected periods was reported as the final period, and the median absolute deviation as an estimate of the uncertainty on that period. This resulted in 158 detections, a somewhat larger detection rate than in Nielsen et al. (2013), but the latter searched the entire *Kepler* sample, which included early-type, giant and old stars, all of which are not expected to lead to period detections.

An attempt was also made to use the measured periods to estimate the differential rotation, if present. (This part of the analysis method was not included in previously published work.) For the stars satisfying the selection criteria described above, a linear trend was fit to the measured periods versus quarter number. However, in almost all cases, no significant trend was found: the linear fit did not significantly reduce the scatter in the measured periods. Therefore, only the median period and uncertainty were reported.

3.3 The Tel Aviv team

This team analysed the light curves using the ACF method first introduced by McQuillan et al. (2013a). The ACF is essentially a measure of the degree of self-similarity of the light curve at different time intervals, or lags. Periodic or quasi-periodic signals with period P in the light curve give rise to an oscillating ACF with regularly spaced peaks at lags $P, 2P, \dots$, modulated by a decaying envelope resulting from correlated noise, spot evolution and/or differential rotation.

The light-curve pre-processing and implementation of the ACF method used in this exercise were identical to that applied to the full *Kepler* sample, as described in McQuillan et al. (2014), except that (a) a pre-processing step was implemented to remove very long term variations, and (b) no attempt was made to select reliable periods, manually or automatically.

The simulated light curves contain very long term variations caused by changes in overall spot coverage. These were not present in the *Kepler* light curves analysed in McQuillan et al. (2014), because the PDC-MAP processing removes them. It was necessary to remove them here, as they would otherwise disrupt the identification of the rotation period from the ACF. The light curve was smoothed using a non-linear filter (Aigrain & Irwin 2004), consisting of a median filter followed by a boxcar filter, with 5000 and 2500-point windows, respectively. The resulting smoothed light curve was subtracted from the original, before computing the ACF. The reported period and associated uncertainty were then measured by identifying up to four regularly spaced peaks in the ACF, and fitting a straight line to peak lag versus peak number. Fig. 5 shows the ACF and period range obtained for light-curve 9.

In previous studies using the ACF, reliable periods were selected either visually (McQuillan et al. 2013a,b) or automatically using a criterion based on the relative height of the first ACF peak, stellar effective temperature and detected period (McQuillan et al. 2014). The latter could not be applied here, since effective temperatures were not available. As selection based on visual inspection would have run the risk of introducing human bias into the results of the exercise, no selection was done and a period was reported for every light curve.

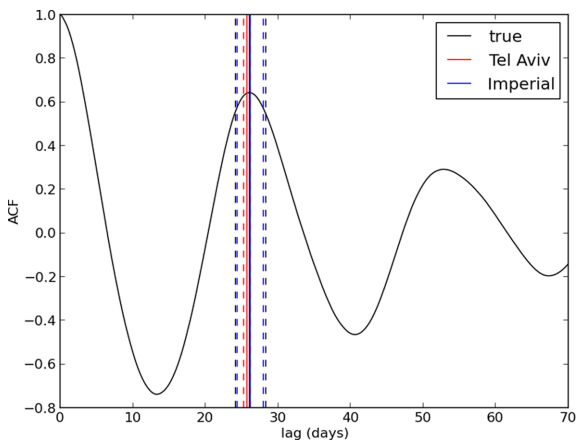


Figure 5. ACF of light-curve 9, as computed by the Imperial team, with the period ranges reported by the Tel Aviv and Imperial teams shown by the vertical red and blue lines, respectively. The vertical black lines show the ‘true’ period range, computed as described in Section 4.1.

3.4 The Imperial team

This team also used the ACF, but with a slightly different implementation. In particular the pre-processing of the light curves was as follows. First, a Savitzky–Golay filter (Savitzky & Golay 1964) with third-order polynomials over 3-month sections of the light curve was applied to smooth long term variability whilst preserving the underlying signal. The light curves were then normalized by dividing them by their median and subtracting unity. Any gaps were linearly interpolated across and Gaussian noise added. Finally, a median filter of width 11 time steps (equivalent to 5.5 h) was applied, followed by a boxcar filter, also of width 11 time steps, to smooth out the high-frequency variability.

The ACF of the resulting light curves was then computed. Peaks in the ACF were deemed significant if higher than 0.15. If no peaks above this value were present, no period was reported for that light curve. In most cases, the first peak was taken to correspond to the rotation period. However, when the second peak was higher than both the first and the third peaks, the second peak was selected as the primary peak. Any primary peaks corresponding to a rotation period of less than 2 d were excluded, as these are particularly strongly affected by correlated noise in the light curve, and their locations are thus uncertain. Primary peaks beyond 50 d were also excluded, as signals longer than this would be removed by the PDC-MAP pipeline in *Kepler* data, and any remaining signal beyond 50 d would typically be due to instrumental effects. This resulted in detections for 970 of the 1005 simulated light curves.

Once the primary peak was identified, up to four subsequent peaks within 10 per cent of integer multiples of the primary one were also identified. A rough estimate of the uncertainty on the lag of each peak was estimated by measuring the full width at half-maximum (FWHM) of each of the selected peaks, the uncertainty on the lag of each peak was then taken as $\text{FWHM}/(2\sqrt{2\ln 2})$. A straight line was then fit to peak lag versus peak number, and the slope of the fit and the resulting uncertainty were adopted as the period and period error, respectively. Fig. 5 shows the ACF and period range obtained for light-curve 9.

3.5 The Natal team

This team used a combination of periodogram, ACF and spot modelling to analyse the light curves, following the procedure described in Lanza et al. (2014). As this process involves a number of manual steps, and the spot modelling is computationally expensive, only the last 500 light curves (from no. 506 onwards) were analysed by this team.

First, the light curves were detrended following the method adopted by De Medeiros et al. (2013). This involves identifying discontinuities using a simple algorithm, and removing long-term trends by fitting a third-order polynomial to each quarter, similar to the recipe of Basri et al. (2011). When applied to raw *Kepler* data, this procedure gives results that are fairly similar to those obtained by subtracting a linear combination of the PDC-MAP cotrending basis vectors from the raw data.

Next, a generalized LS periodogram (Zechmeister & Kürster 2009) was computed for each light curve. The highest peak of the periodogram was used as a preliminary estimate of the star’s rotation period (with no additional selection criteria), as an input to the spot modelling.

The next step used the ACF to identify the light curves with the most stable rotational signal, which are the best candidates to estimate the mean rotation period as well as differential rotation. Working with a small sample of active stars observed by *CoRoT*, *MOST* and *Kepler*, Lanza et al. (2014) were able to measure what they interpreted as a differential rotation signal when the relative height of the first peak in the ACF was around 0.6–0.7. The present hare-and-hounds exercise represents a unique opportunity to test this conclusion. Thus, the spot modelling was also applied to cases where the first ACF peak was lower than this value (down to 0.5). However, owing to time limitations, it was only applied to 303 of the possible 500 light curves.

The remaining 303 light curves were fit using a two-spot model, as described in detail in Lanza et al. (2014). To allow for spot evolution, the time series were cut into segments of equal duration T , during which spots can be considered to remain stable. Each segment is then fit with a two-spot model. The angular rotation rate of the two spots was initially set to the value given by the periodogram analysis mentioned above, but was allowed to vary independently within ± 15 per cent of this value. The inclination of the star was initially fixed to 60° , as it is very degenerate. The spot model was fit using `MPFIT`, an IDL implementation of the Levenberg–Marquart algorithm for non-linear least squares, which allows for some or all of the parameters to be constrained within certain bounds. As this is a local optimizer, the choice of initial parameters can strongly affect the results, and not all segments result in acceptable fits, even for short T . Starting with T equal to the light-curve duration, shorter and shorter intervals were considered (by gradually increasing the total integer number of intervals) until the fit became acceptable (see below).

The segments were then fit again with a two-spot model without differential rotation, i.e. forcing the two spots to share the same rotation period. The initial period of both spots was set to that of the first spot as derived from the differential rotation fit (and all the other parameters' initial values were also taken from the differential rotation fit). For each segment, the relative improvement in the fit between the differential and solid rotation models was computed using the Bayesian information criterion (BIC). Any segment where $\Delta\text{BIC} > 2$ in favour of the differential rotation model were visually examined, and the team manually selected the segment that showed the clearest evidence for differential rotation, using it to report the final minimum and maximum periods. For each light curve, this visual examination was done for a range of segment durations T , and the acceptability of the fit was evaluated at the same time. Typical values for the final segment duration are of the order of 1.5–2 rotation periods.

As there is a strong degeneracy between stellar inclination and several of the spot parameters (mainly latitude and size), the inclination was not explicitly varied in the fits. Instead, it was initially fixed at 60° . If no acceptable fit was found with this inclination, progressively lower values were tried (in steps of 5°), until an acceptable fit was obtained.

Finally, for the two cases deemed to show the strongest evidence for differential rotation (light curves 506 and 513), a Markov chain Monte Carlo (MCMC) analysis was used to obtain posterior distribution over the spot parameters. Fig. 6 shows the results of applying a two spot model to light curve 506.

3.6 Non-blind teams

Two more teams took part in the exercise after the content of the simulations had been made public. These teams did analyse the light

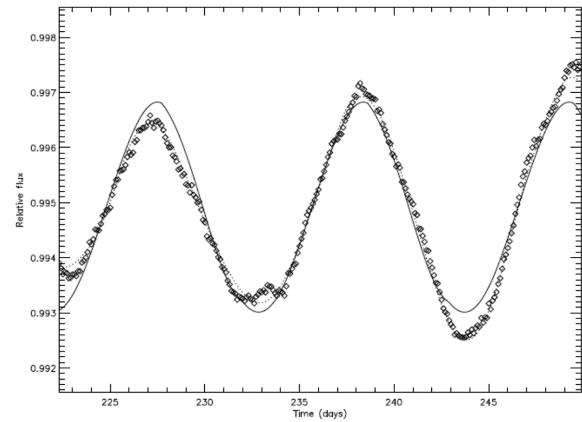


Figure 6. Illustration of the two-spot modelling carried out by the Natal team. This shows, for light-curve 506, the segment showing the strongest evidence for differential rotation. The data are shown by the open lozenges, the best fit obtained with differential rotation by the dotted line and the best fit obtained with rigid rotation by the solid line. The minimum and maximum periods reported by the Natal team for this light curve were 10.77 and 11.62 d, respectively, while the ‘true observable’ minimum and maximum periods, as defined in Section 4.1, were 10.62 and 11.39 d.

curves blindly, in the sense that they did not check the content of individual light curves. Nonetheless, they had access to considerably more information about the range and distribution of the simulated parameters, and also had benefited from the prior experience of the other teams, whose results were discussed before these two additional teams took part.

3.6.1 The CEA team

This team used the ACF, an analysis based on a wavelet decomposition of the light curves and the product of the two. The three results thus obtained are then compared to ascertain that the detected period is not method dependent. This analysis is very similar to the one used by Garcia et al. (2014), and identical to Cellier et al. (in preparation), and is illustrated in Fig. 7 for one example light curve.

The pre-processing of the light curves used in this exercise is based on a simplified version of the tools developed to treat the *Kepler* simple aperture photometry light curves (García et al. 2011). In particular, the light curves were smoothed by convolving them with a 5000-point (~ 100 d) triangular kernel. As the signals were injected into the PDC-MAP rather than raw light curves, no jump correction or quarter-by-quarter detrending was applied. The light curves were then rebinned by a factor of 4 to speed up the analysis.

Their analysis then consisted of three parts. First, the Morlet wavelet decomposition of the light curve is computed, yielding a time-frequency spectrum. This decomposition is then projected on to the period axis to obtain the global wavelet power spectrum (GWPS), which is similar to a Fourier power spectrum but with degraded resolution. As shown by Mathur et al. (2010), the GPWS is less susceptible to detecting harmonics of the true rotation period compared to a standard periodogram. Moreover, the time-frequency spectrum can be used to check if the periodic signal is present throughout the light curve, or is caused by a localized data artefact. The main peaks of the GWPS are then fitted with Gaussian functions and the period corresponding to the highest peak is stored as $P_{\text{rot,GPWS}}$. The half width at half-maximum of the fitted Gaussian is taken as the uncertainty on this period, this also takes into account any possible contribution from differential rotation.

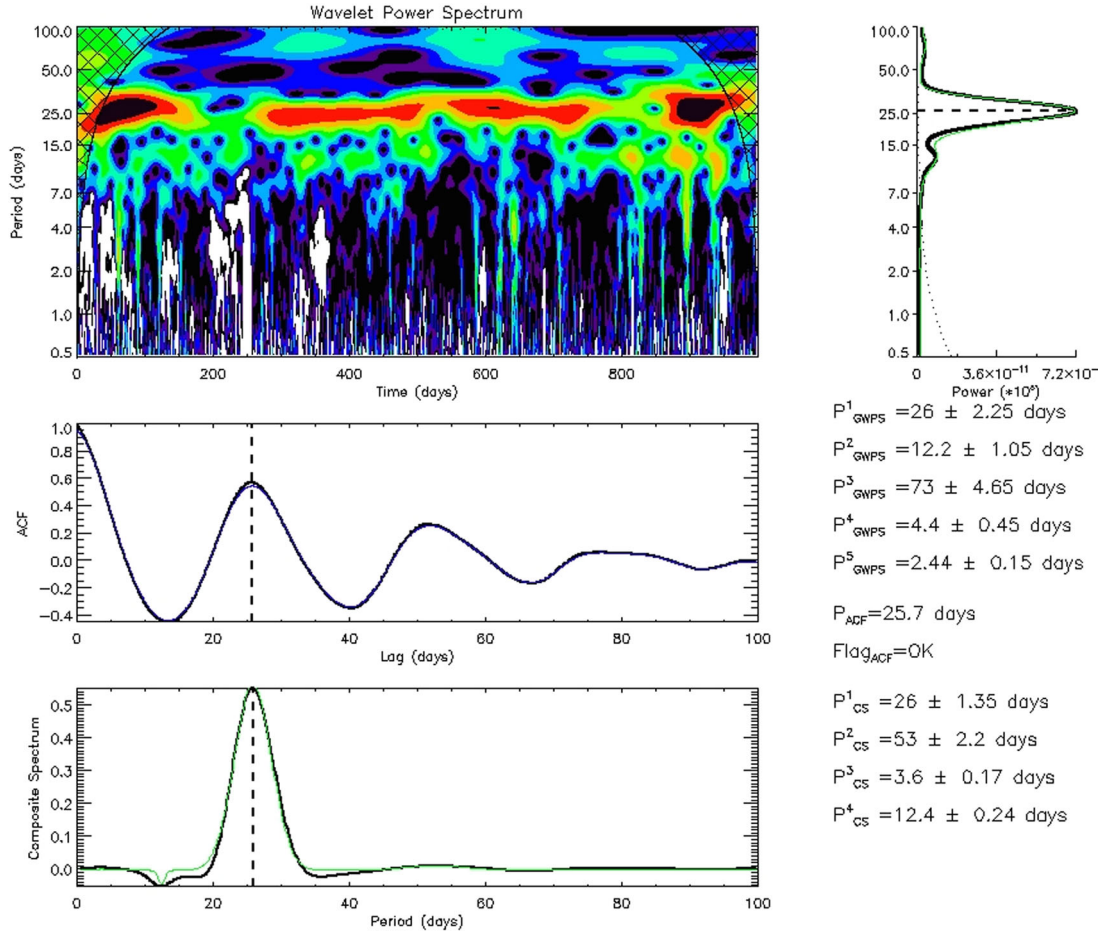


Figure 7. Example graphical output produced by the CEA team for light-curve 9. The top left-hand panel shows wavelet transform of the detrended light curve. The top right-hand panel shows this transform collapsed on to the period axis, while the middle and bottom panel show the light-curve ACF and the composite spectrum obtained by multiplying the wavelet spectrum with the ACF. In the last three panels, the thick black line shows the data, thin green lines show a fit to this composed of multiple Gaussians and the thin purple line shows the smoothed version of the ACF used to identify the ACF peaks. The various periods identified from the different spectra are reported in the bottom right-hand corner of the figure.

Secondly, the ACF of the light curve is computed and smoothed. The smoothing length is determined by the strongest peak in the periodogram of the ACF. Then the period corresponding to the first ACF peak is stored as $P_{\text{tot,ACF}}$. Following McQuillan et al. (2013a), the team checked for repeated peaks in the ACF and regular patterns in the light curves to ensure that the detection is genuine. As it is difficult to calculate uncertainties from the ACF, they did not give an error on this period. The primary goal of computing the ACF period was to validate the GPWS period detection.

Finally, a composite periodogram was obtained by multiplying the ACF with the GWPS. This is done to boost the peaks that are present in both the ACF and the GWPS and to reduce the amplitude of the peaks that only appear in only one of the two methodologies. They call the result the composite spectrum (CS). As for the GWPS, the main peaks of this CS are fitted with Gaussian functions and the period corresponding to the highest peak is returned as $P_{\text{tot,CS}}$, while the uncertainty is given by the half width at half-maximum.

For each light curve, the three periods, $P_{\text{tot,GWPS}}$, $P_{\text{tot,ACF}}$ and $P_{\text{tot,CS}}$ were then compared. If the periods were consistent to within 10 per cent of each other, the detection was considered confirmed. The final rotation period was then taken to be $P_{\text{tot,GWPS}}$ with the associated uncertainty. This last, automatic validation step is the main difference between the present analysis and the approach used

in Garcia et al. (2014), which used a much more time-consuming, systematic visual check.

3.6.2 The Seattle team

The Seattle team used a similar approach to the period finding as the Göttingen/Reinhold team, based on the normalized LS Periodogram. Each light curve was first detrended to remove the longer time-scale variations and residual *Kepler* systematics. A boxcar smoothing algorithm with a width of 3000 data points (~ 82 d) was applied to the light curve and the result subtracted from the original, before computing the LS periodogram. The LS periodogram was not oversampled in this analysis.

Since a smoothing kernel of 82 d was applied, only periods shorter than this were considered significant. Manually examining the light curves led to a conservative upper limit of 75 d and a lower limit of 0.25 d for the range of allowed periods. Rather than selecting the highest amplitude peak in the LS power (P_x) as the most likely period, the Seattle team used a modified criterion. The highest amplitude peak of the quantity $R_x = P_x \sqrt{\nu}$ was selected, where P_x was the normalized LS power as a function of frequency, and ν the corresponding frequency. This had the effect of requiring peaks at longer periods (shorter frequencies) to have larger amplitudes. The

standard deviation over all frequencies of R_x was also computed, and peaks were considered ‘significant’ and thus reported for this exercise if they were above this standard deviation.

Only a single peak was measured in this modified LS analysis, and no checks for alias periods were performed. After discussions with the participating (blind and non-blind) teams, and inspection of the simulation input parameters, the Seattle team did not directly measure a second period from the 1000-d light curves. Instead, the spread of the LS power about the peak was estimated, which was observable as multiple peaks near the primary period, or power distributed over a range of periods with no distinct subpeaks. The LS spread was estimated by first boxcar smoothing the R_x curve with a kernel of five frequency bins, and then least-squares fitting the smoothed R_x with a Gaussian function. The standard deviation of the Gaussian, σ_{R_x} , was reported, as it does correlate with the range of periods simulated via differential rotation for some light curves. However, non-zero LS spread was also observed for objects having no differential rotation but with several starspots emerging at different longitudes (or equivalently rotation phase) throughout the 1000-d light curve.

4 RESULTS

In this section, we evaluate how well the different teams were able to measure the ‘overall’ rotation period, as well as the amount of differential rotation, in each light curve. First, we must define the ‘ground truth’ to compare the different team’s results to.

4.1 Observable periods

For each star, the simulations define an equatorial and a polar rotation period, plus a minimum and maximum spot latitude. Coupled with the differential rotation prescription, this corresponds to a minimum and maximum rotation period that could be present in the data. However, these values cannot necessarily be compared directly to the periods measured from the light curves. If the activity cycle duration is significantly longer than the light-curve duration, the simulated spots may cover only part of the allowed latitude range. Furthermore, the signal will tend to be dominated by the largest active region, and may not be representative of the periods of spot population as a whole.

To enable a more adequate comparison between the simulated and measured periods, we proceed as follows. For each light curve, each spot was assigned a weight proportional to its peak effective area (see Section 2.2). The ‘overall’, minimum and maximum periods were then defined as the median, 10th and 90th percentiles of the weighted periods. These values are reported in the last three columns of Table 2.

4.2 Recovery of overall rotation period

Depending on the team and on the individual light curves, in some cases a single period estimate was reported along with an uncertainty, in others two or more period estimates and in others still a period range. Where a team reported more than one value without explicitly identifying which was the dominant one, we used the mid-point of the range of reported periods as the ‘detected’ period. Where a single uncertainty value was reported rather than a range, we defined the detected range as the detected period $\pm 1\sigma$.

A period detection was considered ‘good’ if the true and detected period were within 10 per cent of each other, ‘ok’ if there was any overlap between the simulated and detected period ranges and ‘bad’

otherwise. Figs 8 and 9 show, for the blind and non-blind teams, respectively, for each team, the simulated versus reported periods, for the noisy and noise-free light curves. Table 3 reports, for all the teams, the fraction of light curves for which a period was reported, and the fraction of light curves for which the reported period was ‘good’ or ‘ok’ according to the definition above.

Of the five teams who participated in a genuinely blind fashion, three (Göttingen/Reinhold, Göttingen/Nielsen and Natal) identified the ‘main’ period as the highest peak in the LS periodogram, or variants thereof, but with different inputs (length of light-curve segment considered, range of periods searched) and selection criteria (coherence across multiple quarters, visual examination). As a result, the fraction of the light curves for which these teams report a detection varies from ~ 15 per cent (Göttingen/Nielsen) to ~ 80 per cent (Göttingen/Reinhold). As one might expect, increasing the number of detections comes at the expense of reliability: between the most and the least conservative of these (Göttingen/Nielsen and Göttingen/Reinhold), the percentage of ‘good’ detection drops from 100 to ~ 75 . The remaining two blind teams (Tel Aviv and Imperial) used the ACF to measure the ‘overall’ period. Both report even higher detection fractions (>95 per cent), while retaining fairly good reliability (just over 70 per cent of the detections were ‘good’ and close to 90 per cent were ‘ok’).

Overall, these results are very encouraging: for the range of periods and amplitudes simulated, overall rotation periods can be recovered with high completeness, and high reliability, using either the periodogram or the ACF. If one is interested in a fully automated method that will give results for almost all light curves while retaining high reliability, an ACF-based method appears the most suitable. A periodogram-based approach can give even better results when the rotation signal is strong, at the expense of some completeness.

We now turn to the results of the non-blind teams. The Seattle team falls somewhere in between the two Göttingen teams in terms of completeness and reliability, reporting results for ~ 70 per cent of the light curves, of which ~ 90 per cent were ‘good’. This is essentially a consequence of slightly different selection criteria. The CEA team reported periods for almost 80 per cent of all light curves, of which close to 90 per cent were ‘good’. These represent the best combination of completeness and reliability among all the teams. Bearing in mind that this was not a completely blind analysis, it nonetheless suggests that a combination of wavelet and ACF analysis, together with dedicated pre-processing, can lead to even better results.

4.2.1 Nature of the ‘bad’ detections

For all teams except the CEA one, ‘bad’ detections, particularly at the longer end of the period range searched, occur at approximately the same rate among both noisy and noise-free light curves. This implies that these bad detections are not predominantly caused by residual systematics in the *Kepler* PDC-MAP light curves.

Almost all of the cases where one or more team reported a ‘bad’ period that was significantly longer than the ‘true’ value have short spot lifetimes. These are cases where the detected signal is not due to rotational modulation, but to the evolution of the star’s overall spot coverage. Although this evolution is not in fact periodic, it can give rise to some apparently quasi-periodic signal in the light curves.

The lower rate of this kind of bad detection among the CEA results may be due to the light-curve pre-processing approach used

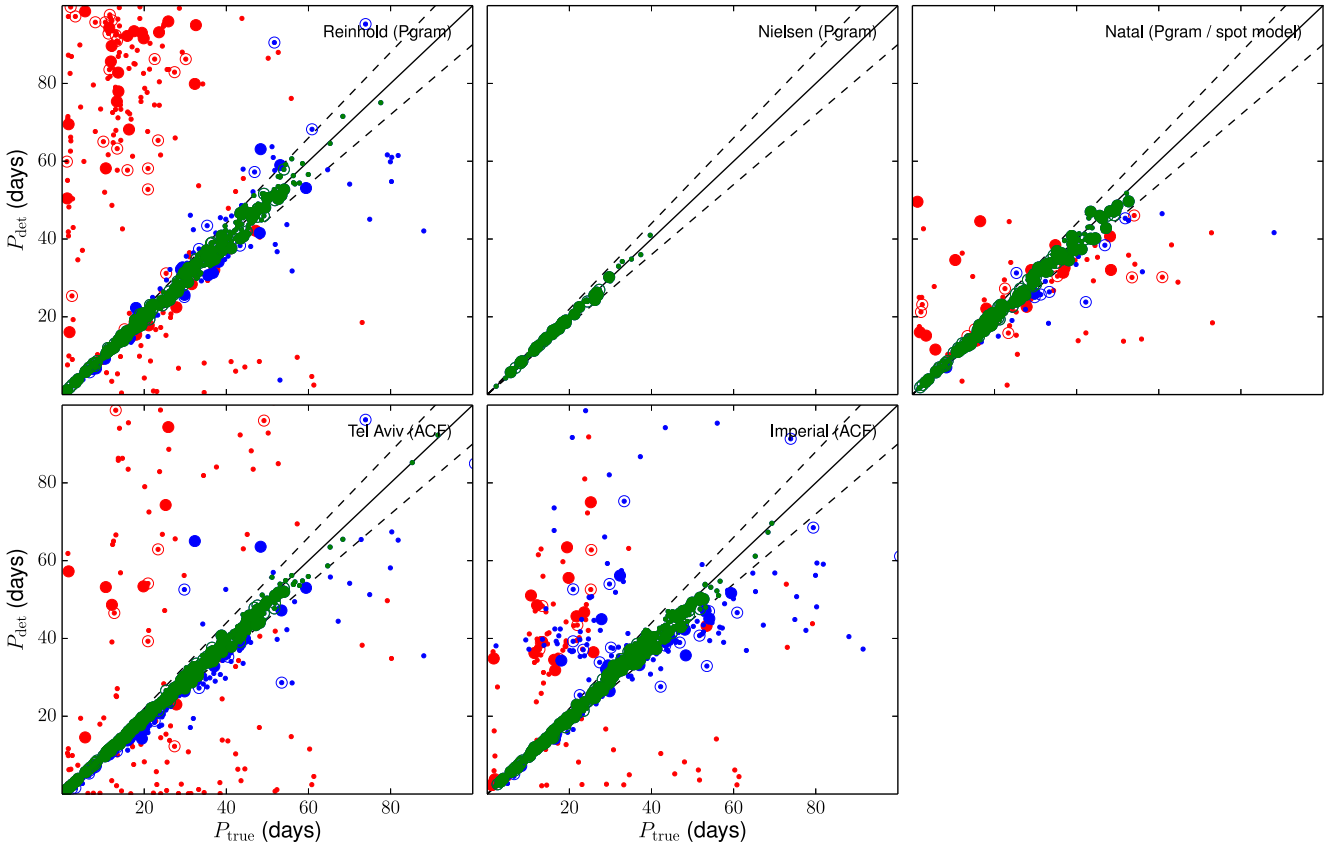


Figure 8. Detected versus simulated periods for each team. Green, blue and red symbols correspond to ‘good’, ‘ok’ and ‘bad’ detections, respectively (see text for details). The larger symbols correspond to the noise-free light curves. The solid and dashed lines mark the one-to-one and ± 10 per cent correspondence between simulated and detected periods. The unfilled symbols correspond to stars with a cycle length less than 3 yr, which may be confused with a long rotation period in the analysis performed by the teams.

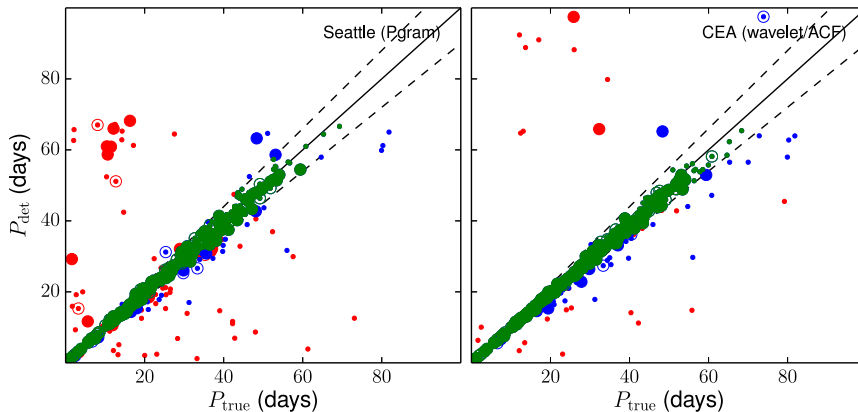


Figure 9. Detected versus simulated periods for the two teams who analysed the light curves in non-blind mode. Lines and symbols are the same as in Fig. 8.

by that team. On the issue of pre-processing, it is important to note that the present exercise did not test the full impact of the PDC-MAP processing on stellar rotation signals, since the signals were injected into the processed light curves. We note again, that in our simulated light curves the entire light curve is normalized to the absolute continuum which is not known for real *Kepler* light curves. This introduces an additional issue for the teams, who may have confused the stellar cycle with stellar rotation. In reality this would not be an issue when dealing with real *Kepler* data because the normalization between quarters is unknown. The teams that either restricted their analysis to shorter periods, or those teams that fit out the longer

variations (in effect treating our light curves more like real *Kepler* data) performed better than those teams who did not. The unfilled circles in Figs 8 and 9 show those stars with a stellar cycle length less than 3 yr. A considerable number of the detections for these stars overestimated the rotation period of the star, suggesting that the short cycle length can indeed be confused with the rotation period of the star.

All teams also reported a smaller number of spurious detections that were significantly shorter than the true values. Interestingly, these occur only in the noisy light curve. A more detailed inspection of these cases revealed that the detected signal was present, albeit

Table 3. Overall rotation period detection results. For the noisy and noise-free simulated light curves, the three columns show the percentage of cases for which each team reported a value, and the percentage of reported values that were ‘good’ and ‘ok’, respectively (see text for details). For the solar cases, we report numbers rather than percentages.

Method	Noisy			Noise-free			Solar	
	Per cent det	Per cent good	Per cent ok	Per cent det	Per cent good	Per cent ok	No. det	No. ok
Blind teams								
Reinhold (periodogram)	82	67	76	87	65	77	4	2
Nielsen (periodogram)	16	100	100	15	100	100	0	0
Natal (periodogram/spot model)	27	69	76	72	80	85	5	2
Tel Aviv (ACF)	100	68	80	100	75	90	5	4
Imperial (ACF)	95	71	87	98	73	88	5	4
Non-blind teams								
Seattle (periodogram)	68	81	88	75	84	90	0	0
CEA (wavelets/ACF)	78	88	95	82	92	99	2	2

at a low level, in the ‘quiet’ light curves into which the simulated signals were injected. Thus, these detections may not be entirely spurious, though it is not possible to determine whether the detected signal in those cases was genuinely due to rotational modulation of starspots.

4.2.2 Solar cases

For the solar cases, a detection was considered ‘ok’ if the reported period range overlapped with the interval 25–30 d. This interval covers the expected solar synodic rotation periods for latitudes 0° – 45° , the approximate latitude range within which sunspots tend to be located.

Two of the teams using the periodogram recovered the solar period in two of the five solar cases, both the teams using the ACF did so in four of the five solar cases, and the CEA team recovered it in two of the five cases. This suggests, once more, that the ACF method may perform slightly better than the periodogram for Sun-like stars which are neither fast rotators nor very active. Additionally, we note that, on average, the measured period decreases as one progresses in the solar activity cycle, which is an encouraging sign for the detectability of differential rotation for *Kepler* stars having shorter activity cycles. Previous studies of the TSI have shown that the true rotation period of the Sun was detectable only during phases of minimum activity during the 11 yr solar cycle when the modulation was dominated by faculae (Lanza et al. 2003; Lanza, Rodonò & Pagano 2004). A re-analysis of the TSI using an autocorrelation approach was carried out in Lanza & Shkolnik (2014) where they reached similar conclusions.

4.3 Recovery of differential rotation

Not all teams explicitly searched for differential rotation signals. For those that did, Fig. 10 shows a comparison of the simulated versus measured differential rotation signal, quantified by the rotational shear α , which we define as $\alpha \equiv (P_{\max} - P_{\min})/P$, following Reinhold et al. (2013), where P is the ‘overall’ or mean rotation period. To compute the simulated value, α_{true} , shown in the figure, we use the ‘observable’ period range defined in Section 4.1, but the results are similar when using the absolute maximum and minimum periods defined by the spot latitude range and the value of $\Delta\Omega$ used in the simulations.

The three teams whose results are shown in Fig. 10 used very different methods: identification of multiple individual peaks in the periodogram, spot modelling and measurement of the spread of the

power in the LS periodogram about the strongest peak. Strikingly, none of these methods enables an unambiguous detection of the differential rotation signal. All teams reported non-zero shear in numerous cases where the simulated shear was zero, and vice versa. There are, of course, cases where each team reported a non-zero shear and the simulations did include differential rotation. Although there is a lot of scatter, there is some tendency for those to cluster around the one-to-one line, at least for the Göttingen/Reinhold and Natal teams, particularly when considering only ‘good’ period detections (green points), but the correlation is weak.

Rapid spot evolution can lead to light-curve morphologies that closely resemble those caused by differential rotation. In an attempt to disentangle between the effects of spot evolution and differential rotation, we plotted in the bottom panel of Fig. 10 only the cases with relatively long spot lifetimes $\tau \geq 5P$. There are fewer ‘bad’ or ‘ok’ period detections (red and blue points) in that subsample, but the picture in terms of recovery of differential rotation is not substantially different. It may be that differential rotation signals can only be identified reliably in cases where τ is even larger, but the present blind exercise did not include enough such cases to test that hypothesis.

We also checked the results of the other teams, who reported a period range or error bar, even when they did not do so with the specific intention of making a differential rotation measurement, but the results were similar: there was little or no correlation between the simulated shear and the period range or period uncertainty. We also checked for any correlation between simulated shear and a number of ACF properties measured by the Tel Aviv team, which could conceivably be related to differential rotation, such as the width of the ACF peaks, or the gradient in this width for consecutive peaks, but to no avail.

5 SUMMARY AND DISCUSSION

We have carried out a blind exercise to test the recoverability of ‘typical’ rotation and differential rotation signals from *Kepler* light curves. Seven teams analysed up to 1004 light curves containing simulated starspot signals including activity cycles, differential rotation and spot evolution, using a variety of methods including variants on the LS periodogram, ACF, wavelet power spectra and combinations thereof. Five of the teams took part in genuinely blind fashion, while the last two teams analysed the light curves after their content had been made public, though they did not explicitly use this information in their analysis.

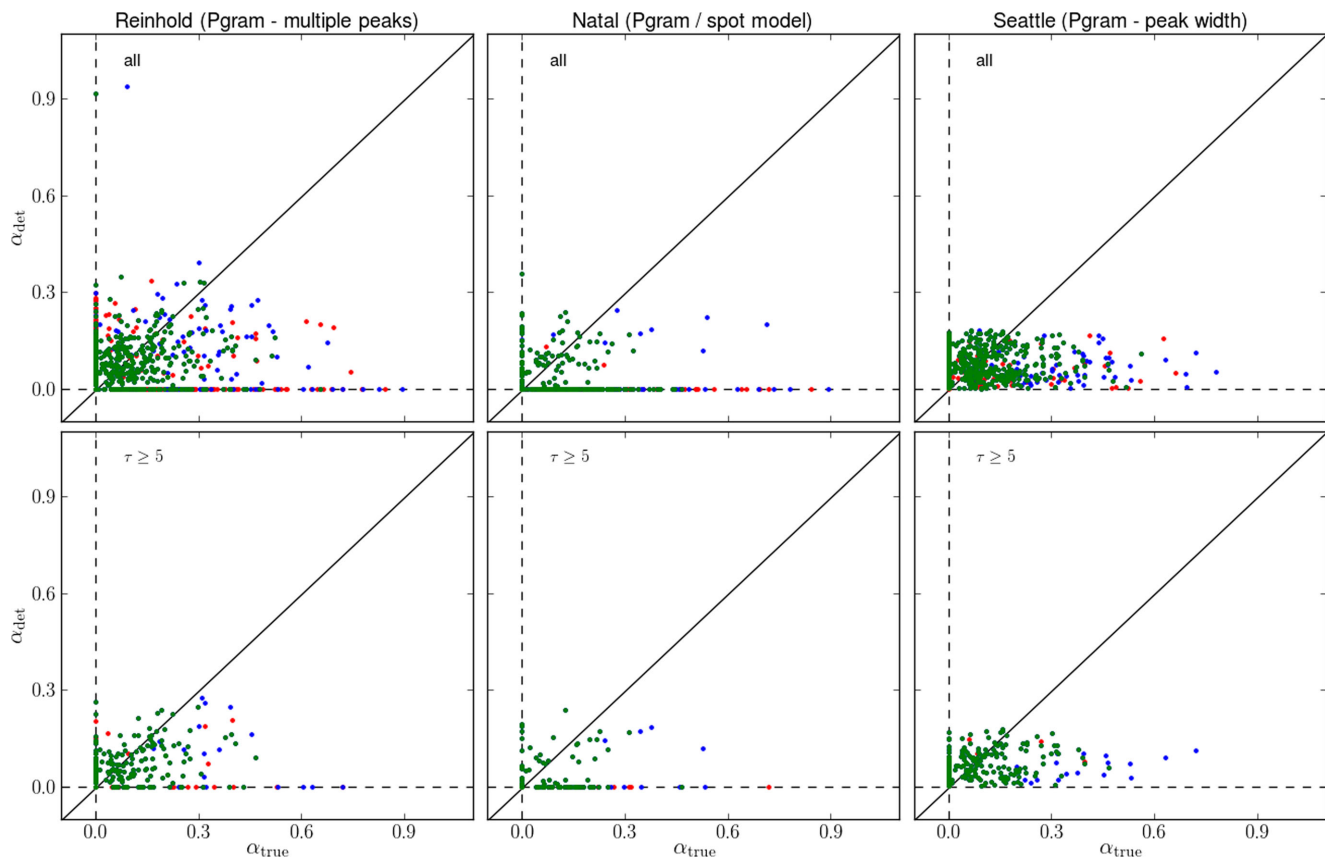


Figure 10. Measured versus simulated differential rotation shear ($\alpha \equiv (P_{\min} - P_{\max})/P$) for the three teams which specifically searched for differential rotation signals. The colour coding of the points is the same as in Fig. 8. The top row shows all the light curves each team reported results for, while the bottom row shows only cases with spot evolution time-scales $\tau > 5P$.

The results of the exercise for the recovery of the ‘overall’ rotation period are positive: the periods reported by the different teams were within 10 per cent of the ‘true’ period between 70 and 100 per cent of the time. There was a clear trade-off between completeness and reliability: one team – Göttingen/Nielsen, using the periodogram with a stringent stability criterion – obtained 100 per cent ‘good’ detections, but reported results for only 27 per cent of the cases, while another – Tel Aviv, using the ACF method with no selection criterion whatsoever – reported results for 100 per cent of the case, but only ~ 70 per cent of these were ‘good’. Arguably, the best combination of completeness and reliability was obtained by the CEA team, using a combination of multiple detrending as well as period-search methods (ACF and wavelets). However, as this team took part after the end of the genuinely blind phase of the exercise, their results cannot be compared directly to those of the other teams.

The results are much more worrying for differential rotation. There seems to be very little correlation between the injected and recovered differential rotation shear. In numerous cases, a non-zero shear was reported by several teams where the simulated shear was zero, and vice versa. The origin of this problem, which persists whatever the method used, is not entirely clear. Spot evolution is likely to be playing a role, as it can mimic differential rotation signals in the light curves and periodograms, but the problem persists even in cases where the simulated spots had relatively long lifetimes (≥ 5 rotation periods). Further tests with fewer, larger, longer lived active regions are needed to see if differential rotations signals can be recovered more reliably in more active stars – which

have typically been the focus of differential rotation studies. In the mean time, the results of the present exercise suggest that differential rotation measurements based on periodogram analysis or spot modelling of the full-disc light curve alone should be treated with considerable caution.

It is worth noting that all of the teams who searched for differential rotation signals did so by looking for multiple periodicities present in the light curves at the same time. An alternative, which may yield more robust results, would be to look for a drift in the mean period over an activity cycle (Scargle, Keil & Worden 2013; Mathur et al. 2014). However, this requires light curves lasting at least a significant fraction of the activity cycle. Furthermore, it implicitly assumes the existence of a Sun-like butterfly pattern, which may or may not be common in other stars.

In the remainder of this section we comment on a few points which were brought up by the exercise and seem worthy of further discussion.

5.1 What do we actually mean by rotation period?

One question which had to be addressed in the course of the exercise, and which is less trivial than it may seem, is: what do we mean by a ‘correct’ rotation period measurement for a star with differential rotation? We chose to address this in a statistical, rather than a strictly physical sense, by considering the distribution of spot rotation periods present in the simulations, weighted by the spot size and comparing that to the reported period(s). This enabled us

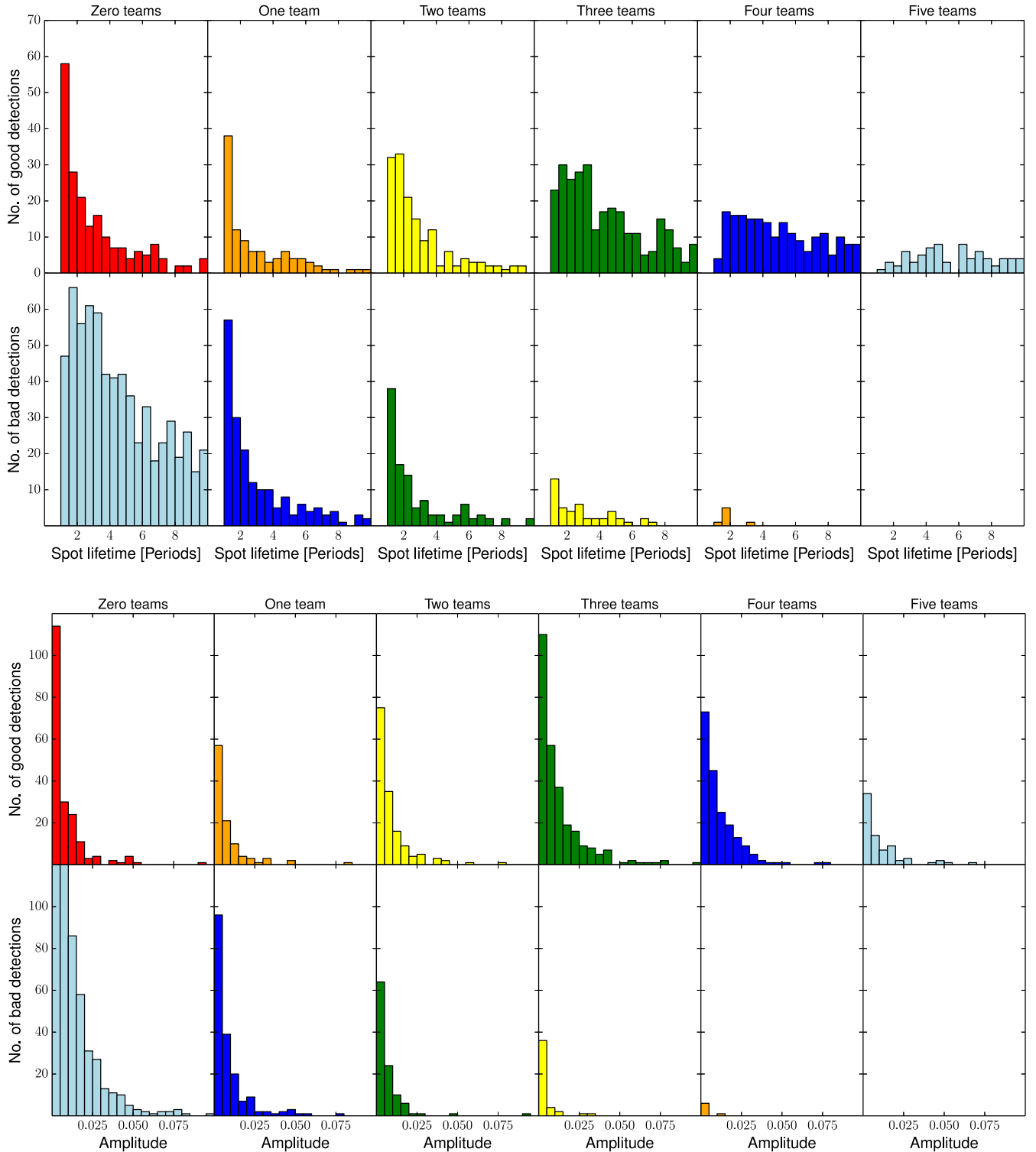


Figure 11. Histograms showing how the value for spot lifetime (top) and amplitude of variability (bottom) gave rise to consistently good or bad detections. For both parameters, the top row of histograms shows the number of teams with good detections and the bottom row the number of bad detections.

to evaluate the effects of noise, spot evolution, inclination angle of the star and the period measurement method used on the results. However, it is important to bear in mind that both the ‘true’ and the reported period ranges are affected by other factors such as the distribution of latitudes at which spots emerge and the specific differential rotation prescription, which would be unknown a priori in a realistic scenario.

5.2 Main factors affecting overall period sensitivity and reliability

Fig. 11 shows the number of teams who reported ‘good’ and ‘bad’ detections, as a function of the (observable) spot lifetime and the peak-to-peak amplitude of the noise-free light curve. For the sake of consistency, only the results of the blind teams are shown, although

the picture does not change appreciably when the results of the non-blind teams are included. As one might expect, both the sensitivity and the reliability of the period search, when evaluated across all the teams, decreases for short-lived spots, smaller amplitudes and longer periods.

5.3 When are unambiguous detections of differential rotation feasible?

While the differential rotation results of the exercise were on the whole disappointing, there were a few cases where individual teams did report a range of periods that closely matched the simulated range – such a case is illustrated for the Natal team in Fig. 6, for example. This naturally begs the question: if differential rotation can be reliably measured under some circumstances, can we identify what those circumstances are? We searched for this by comparing the spot lifetimes, median number of spots present at any one time, overall periods and activity levels for the stars where the reported shear was close to the simulated one, to those where it was not. Unfortunately, we were unable to identify, within the parameter space covered by this exercise, a ‘safe’ region where differential rotation can be reliably measured. Even if future studies were to reveal that differential rotation can be detected reliably for – say – spot lifetimes longer than 10 rotation periods and stars dominated by only 2–3 active regions at any one time, this information would be useful only if coupled with a reliably means of establishing, from the light curve itself, that these conditions apply.

To our knowledge, the only differential rotation studies to explicitly account for spot evolution so far are the work of Fröhlich et al. (2009, 2012) and Frasca et al. (2011), who used the BIC to distinguish between cases with and without differential rotation while allowing for spot evolution. This approach is more sophisticated than most, but unfortunately none of the teams involved in the present exercise used it.

5.4 Limitations of this blind test

The light-curve simulations used in this exercise had a number of limitations, which one might want to address in any future exercise with similar goals.

First, the distribution of simulated periods and amplitudes were chosen to match those found by recent studies based on *Kepler* data. This made sense at the time, but as a result the conclusions of the present exercise apply only within the range of parameter space probed, in particular, periods ranging from 1 to 50 d and activity levels from 0.3 to 3 times solar. In hindsight, it would have been interesting to go to longer periods and smaller amplitude, to enable us to test whether the observed distributions are real, or the result of selection effects. A significant number of light curves in the *Kepler* data show long-lived spot groups that live for many rotations of the star. We do not model such persistent spot groups in our simulated light curves; rather, our investigation was more focused on spot groups with half-lives up to 10 stellar rotations. Simulating light curves of longer lived spots would be interesting since this would negate the confusion between spot evolution and the detection of differential rotation.

Furthermore, the sample did not include any very active stars (the activity rate was limited to three times solar or lower). Because the activity rate controls the number of spots emerging per unit time, while the spot lifetimes were specified as a multiple of the period, this results in particularly small amplitudes for the short-period cases (as can be seen in the right-hand column of Fig. 11). Thus, the

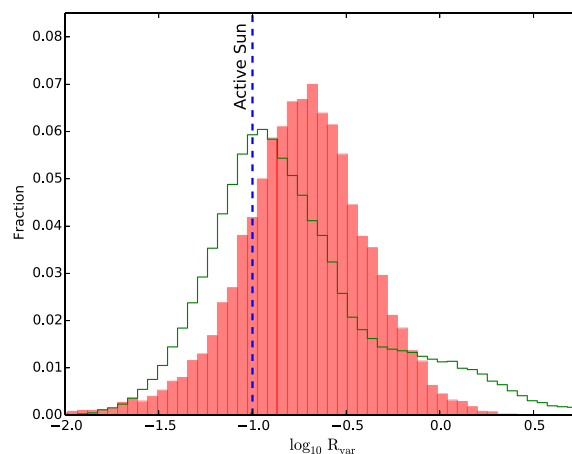


Figure 12. Distribution of the variability in our simulated light curves (red). Overplotted in green is the corresponding distribution for all *Kepler* Q3 stars from Reinhold et al. (2013). Note that both histograms are normalized such that the total area equals one. The dashed blue line shows the variability level of the Sun at solar maximum for reference.

most promising type of stars for differential rotation studies, namely stars with large-amplitude, short-period rotational modulation, were absent from our sample. This makes it difficult to compare our results to those of previous studies attempting to measure differential rotation, which have focused mainly on such stars.

Fig. 12 shows the variability in our simulated light curves derived by splitting each light curve into 90-d windows. We then fit a low-order polynomial to each time window and calculate the range. Overplotted is the analogous distribution from all *Kepler* Q3 stars from Reinhold et al. (2013). The histograms show how our sample does not cover the most active stars in the *Kepler* survey. The results presented here should therefore only be thought of in the context of stars exhibiting low and moderate levels of activity.

ACKNOWLEDGEMENTS

The authors would like to thank the referee, Professor G. Basri, for suggesting several ways to improve our discussion and asking thought-provoking questions that helped us to improve the final paper. The authors are grateful to All Souls College Oxford for supporting and hosting a 3-d workshop during which the present paper was finalized. SA’s contribution to this work was supported by the UK Science and Technology Facilities Council through Consolidated Grant ST/K00106X/1. SA also wishes to thank I. Ribas for pointing out a couple of small errors, which had crept into the expression for f in Aigrain et al. (2012) (these errors were corrected in the present paper). JL acknowledges support through NASA/GALEX grant program under Cooperative Agreement No. NNX12AC19G issued through the Office of Space Science. MLdC acknowledges a CAPES/PNPD fellowship. JRdM and MLdC acknowledge financial support of the INCT INEspaço. TC and RAG want to acknowledge the funding of the CNES grant at the CEA, as well as the ANR (Agence Nationale de la Recherche, France) program IDEE (n ANR-12-BS05-0008) ‘Interaction Des Etoiles et des Exoplanetes’.

REFERENCES

- Affer L., Micela G., Favata F., Flaccomio E., 2012, *MNRAS*, 424, 11
Aigrain S., Irwin M., 2004, *MNRAS*, 350, 331

- Aigrain S., Pont F., Zucker S., 2012, *MNRAS*, 419, 3147
- Baliunas S. L. et al., 1985, *ApJ*, 294, 310
- Baliunas S., Sokoloff D., Soon W., 1996, *ApJ*, 457, L99
- Basri G. et al., 2011, *AJ*, 141, 20
- Berdugina S. V., 2005, *Living Rev. Sol. Phys.*, 2, 8
- Bouvier J., Matt S. P., Mohanty S., Scholz A., Stassun K. G., Zanni C., 2014, in Beuther H., Klessen R., Dullemond K., Henning T., eds, *Protostars and Planets VI*. University of Arizona Press, Tucson, p. 433
- Bradshaw S. J., Hartigan P., 2014, *ApJ*, 795, 79
- Collier Cameron A., Donati J.-F., Semel M., 2002, *MNRAS*, 330, 699
- Croll B. et al., 2006, *ApJ*, 648, 607
- De Medeiros J. R. et al., 2013, *A&A*, 555, A63
- Donahue R. A., Baliunas S. L., 1992, *ApJ*, 393, L63
- Donati J.-F., Collier Cameron A., Petit P., 2003, *MNRAS*, 345, 1187
- Frandsen S., Jones A., Kjeldsen H., Viskum M., Hjorth J., Andersen N. H., Thomsen B., 1995, *A&A*, 301, 123
- Frasca A., Fröhlich H.-E., Bonanno A., Catanzaro G., Biazzo K., Molenda-Żakowicz J., 2011, *A&A*, 532, A81
- Fröhlich C., 2000, *Space Sci. Rev.*, 94, 15
- Fröhlich C., 2006, *Space Sci. Rev.*, 125, 53
- Fröhlich C., Lean J., 1998, *Geophys. Res. Lett.*, 25, 4377
- Fröhlich H.-E., Küker M., Hatzes A. P., Strassmeier K. G., 2009, *A&A*, 506, 263
- Fröhlich H.-E., Frasca A., Catanzaro G., Bonanno A., Corsaro E., Molenda-Żakowicz J., Klutsch A., Montes D., 2012, *A&A*, 543, A146
- García R. A. et al., 2011, *MNRAS*, 414, L6
- García R. A. et al., 2014, *A&A*, 572, 34
- Gondoin P., Fridlund M., Gandolfi D., Günter E., 2010, in Kosovichev A. G., Andrei A. H., Rozelot J.-P., eds, *Proc. IAU Symp. 264, Solar and Stellar Variability: Impact on Earth and Planets*. Cambridge Univ. Press, Cambridge, p. 60
- Hathaway D. H., 2010, *Living Rev. Sol. Phys.*, 7, 1
- Irwin J., Bouvier J., 2009, in Mamajek E. E., Soderblom D. R., Wyse R. F. G., eds, *Proc. IAU Symp. 258, The Ages of Stars*. Cambridge Univ. Press, Cambridge, p. 363
- Jeffers S. V., Donati J.-F., Collier Cameron A., 2007, *MNRAS*, 375, 567
- Kjeldsen H., Bedding T. R., 1995, *A&A*, 293, 87
- Lanza A. F., Shkolnik E. L., 2014, *MNRAS*, 443, 1451
- Lanza A. F., Rodono M., Zappala R. A., 1994, *A&A*, 290, 861
- 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
- Lanza A. F. et al., 2009, *A&A*, 506, 255
- Lanza A. F., Das Chagas M. L., De Medeiros J. R., 2014, *A&A*, 564, A50
- Llama J., Jardine M., Mackay D. H., Fares R., 2012, *MNRAS*, 422, L72
- Mackay D. H., Jardine M., Collier Cameron A., Donati J.-F., Hussain G. A. J., 2004, *MNRAS*, 354, 737
- McQuillan A., Aigrain S., Mazeh T., 2013a, *MNRAS*, 432, 1203
- McQuillan A., Mazeh T., Aigrain S., 2013b, *ApJ*, 775, L11
- McQuillan A., Mazeh T., Aigrain S., 2014, *ApJS*, 211, 24
- Mathur S. et al., 2010, *A&A*, 511, A46
- Mathur S. et al., 2014, *A&A*, 562, A124
- Maunder E. W., 1904, *MNRAS*, 64, 747
- Maunder E. W., 1922, *MNRAS*, 82, 534
- Nielsen M. B., Gizon L., Schunker H., Karoff C., 2013, *A&A*, 557, L10
- Reinhold T., Reiners A., Basri G., 2013, *A&A*, 560, A4
- Savitzky A., Golay M. J., 1964, *Analytical Chem.*, 36, 1627
- Scargle J. D., Keil S. L., Worden S. P., 2013, *ApJ*, 771, 33
- Schrijver C. J., Harvey K. L., 1994, *Sol. Phys.*, 150, 1
- Strassmeier K. G., Bopp B. W., 1992, *A&A*, 259, 183
- Thompson S. E., Christiansen J. L., Jenkins J. M., 2013, *Kepler Data Release 21 Notes (KSCI-19061-001)*. Kepler mission
- Vaughan A. H., Preston G. W., Baliunas S. L., Hartmann L. W., Noyes R. W., Middelkoop F., Mihalas D., 1981, *ApJ*, 250, 276
- Walkowicz L. M., Basri G., Valenti J. A., 2013, *ApJS*, 205, 17
- Wilson O. C., 1978, *ApJ*, 226, 379
- Zechmeister M., Kürster M., 2009, *A&A*, 496, 577

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S2. Parameters of the simulated light curves. (<http://mnras.oxfordjournals.org/lookup/suppl/doi:10.1093/mnras/stv853/-/DC1>).

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.