Astronomy & Astrophysics manuscript no. Kepler30AcceptedVersion
March 31, 2021

On the multiscale behaviour of stellar activity and rotation of the planet host Kepler-30

D. B. de Freitas[1,2] ⋆, A. F. Lanza[1], F. O. da Silva Gomes[2], and M. L. Das Chagas[3]

[1] INAF-Osservatorio Astrofisico di Catania, Via S. Sofia 78, I-95123, Catania, Italy
[2] Departamento de Física, Universidade Federal do Ceará, Caixa Postal 6030, Campus do Pici, 60455-900 Fortaleza, Ceará, Brazil
[3] Faculdade de Física â Instituto de Ciências Exatas, Universidade Federal do Sul e Sudeste do Pará, Marabá, PA 68505-080, Brazil

ABSTRACT

Context. The Kepler-30 system consists of a G dwarf star with a rotation period of ∼16 days and three planets orbiting almost coplanar with periods ranging from 29 to 143 days. Kepler-30 is a unique target to study stellar activity and rotation in a young solar-like star accompanied by a compact planetary system.
Aims. We use about 4 years of high-precision photometry collected by the Kepler mission to investigate the fluctuations caused by photospheric convection, stellar rotation, and starspot evolution as a function of the timescale. Our main goal is to apply methods for the analysis of timeseries to find the timescales of the phenomena that affect the light variations. We correlate those timescales with periodicities in the star as well as in the planetary system.
Methods. We model the flux rotational modulation induced by active regions using spot modelling and apply the Multifractal Detrending Moving Average algorithm (MFDMA) in standard and multiscale versions for analysing the behaviour of variability and light fluctuations that can be associated with stellar convection and the evolution of magnetic fields on timescales ranging from less than 1 day up to about 35 days. The light fluctuations produced by stellar activity can be described by the multifractal Hurst index that provides a measure of their persistence.
Results. The spot modeling indicates a lower limit to the relative surface differential rotation of $\Delta\Omega/\Omega \sim 0.02 \pm 0.01$ and suggests a short-term cyclic variation in the starspot area with a period of ∼ 34 days, virtually close to the synodic period of 35.2 days of the planet Kepler-30b. By subtracting the two timeseries of the SAP and PDC Kepler pipelines, we reduce the rotational modulation and find a 23.1-day period close to the synodic period of Kepler-30c. This period also appears in the multifractal analysis as a crossover of the fluctuation functions associated with the characteristic evolutionary timescales of the active regions in Kepler-30 as confirmed by spot modelling. These procedures and methods may be greatly useful for analysing current TESS and future PLATO data.

Key words. stars: activity – stars: rotation – techniques: photometric – stars: individual: Kepler-30 (KOI-806)

1. Introduction

Stellar rotation is a fundamental physical parameter in stellar astrophysics and plays an important role in the formation and evolution of stars (Kraft 1967; Skumanich 1972; Kawaler 1988). Rotation controls stellar magnetism, mixing in the stellar interior, and tidal interactions in close binary systems. Moreover, the relationship between rotation and magnetic activity has important implications for the detectability and characterization of planets orbiting solar-like stars (Maxted 2016; Collier Cameron 2018). Such stars comprise the vast majority of the Kepler exoplanet targets and also allow us to investigate stellar variability due to magnetic activity. In particular, starspots and active regions in the stellar photospheres modulate the stellar flux on the rotation period (Walkowicz & Basri 2013).

The Kepler space mission found several multi-planet systems, among which that orbiting Kepler-30 (KOI-806), that is a star with nearly solar mass and radius, an effective temperature of 5500 ± 55 K, and a rotation period of ∼16.0 days as measured by means of the Lomb–Scargle periodogram (Sanchis-Ojeda et al. 2012; Lanza, Das Chagas & De Medeiros 2014). These properties make it a very interesting young solar analogue, thus motivating our choice of investigating its rotation and activity behaviour. Moreover, it is orbited by three planets named Kepler-30b, Kepler-30c and Kepler-30d with masses of 9.2, 536 (∼ 1.7 Jupiter masses) and 23.7 Earth masses, radii of 3.75, 11.98 and 8.79 Earth radii (Panichi et al. 2018), and orbital periods of 29.3, 60.3 and 143.3 days (cf. Sanchis-Ojeda et al. 2012), respectively. Based on the values of radius and mass, Panichi et al. (2018) estimated that Kepler-30b is a Neptune-like planet rather than a super-Earth, Kepler-30c is a Jovian planet, while Kepler-30d is classified as a Neptune-mass planet, with bulk densities given by ∼0.96 g·cm^{-3}, ∼1.71 g·cm^{-3} and ∼0.19 g·cm^{-3}, respectively. In this system, the orbits of the planets are on the same plane almost perpendicular to the stellar spin, a behaviour similar to our solar system. As a consequence, we observe Kepler-30 almost equator-on. This provides a fundamental constraint to reduce the degeneracies in the spot modelling of the star, that is, an additional motivation for applying our approach (see Section 4). By using the spot modeling, Bonomo & Lanza (2012) and Lanza et al. (2019) performed an analysis for the star Kepler-17, a solar-like star younger than the Sun hosting an hot Jupiter that occults starspots during transits. In both studies, the authors analysed activity and rotation of this star using starspots as tracers, comparing the longitudes of the spots mapped from the out-of-transit photometry with those of the spots occulted during transits to validate their spot modelling method. The conclusions of these studies are that spot modelling would allow us: i) to derive active longitudes where spots potentially form; ii) to estimate the

⋆ Email: danielbrito@fisica.ufc.br

lifetime of active regions; iii) to determine a lower limit for the stellar differential rotation, and; iv) to evaluate short-term and long-term activity cycles depending on the length of the available timeseries.

Inspired by the peculiar properties of the Kepler-30 system and the availability of almost four years of Kepler high-precision photometry, we decided to invest our efforts in an unprecedented analysis of this system. In particular, the present work carries on for the first time a joint analysis exploting spot modeling and multifractal methods, both widely used in the astrophysical context, but never confronted with each other.

Over the last years, several methods have been applied for an appreciation of the (multi)fractal structure of time series in astrophysical systems (e.g., de Freitas et al. 2016, 2017; Belete et al. 2018; de Franciscis et al. 2019; de Freitas et al. 2019a,b). The estimation of local fluctuations and long-term dependency of astrophysical time series is a problem that has been recently studied to understand the effect of long-memory processes due to the stellar rotation (de Freitas et al. 2019a,b). As discussed by de Freitas et al. (2013a), the global Hurst exponent (Hurst 1951) is a powerful (multi)fractal indicator for analysing CoRoT and Kepler time series, which is mainly used to elucidate the persistence due to rotational modulation in the series themselves. Recently a novel technique proposed by Gu & Zhou (2010), known as the multifractal detrended moving average (MFDMA) method, has been applied for the multifractal characterization of Kepler timeseries (de Freitas et al. 2017, 2019a,b). The MFDMA method filters out the local trends of nonstationary series by subtracting the local means. In addition, the MFDMA method investigates the local fluctuations of timeseries using an a priori (fixed) timescale. According to Wang, Shang & Cui (2014), fixing an a priori scaling range may lead to a crossover falling within the scaling range by mistake, and therefore the results could be biased. A method to avoid mistakes due to fixed scaling ranges is to measure the multifractal properties considering the entire timescales. To this end, a new approach named the multiscale MFDMA method (MFDMAτ) will be introduced in this paper (see Section 6).

Our paper is structured as follows. Observations and technical details on the Kepler mission are presented in Sect. 2. In Sect. 3, we process and prepare the Kepler-30 timeseries as obtained by the pre-search data conditioning (PDC) and simple aperture photometry (SAP) pipelines. In particular, we investigate the impact of the PDC pipeline over stellar variability when compared to the SAP data reduction, giving an overview of the characteristics of the PDC and SAP data using quality flags and discussing the difference between these pipelines. In the next sections 4, 5 and 6, we present our methods: spot modelling, classical MFDMA method, and a new approach named as multiscale MFDMA method (MFDMAτ), respectively. In Sect. 7, we apply our set of methods to the Kepler-30 data and present the results of spot modelling first and then introduce the results of the application of the multifractal analysis. In the last section 8, our conclusions and final remarks are summarised.

2. Observations

From 2009 to 2013, the Kepler mission performed 17 observational runs, a.k.a. quarters, of ∼90 days each, composed of long cadence (data sampling every ∼ 30 min; Jenkins et al. 2010a) and short cadence (sampling every ∼1 min) observations (Van Cleve et al. 2010; Thompson et al. 2013). The mission public archive provides two types of data called Simple Aperture Photometry (SAP) and Pre-search Data Conditioning (PDC) time series. The SAP flux is the sum of all calibrated background-subtracted flux values of the pixels belonging to the target aperture and for which a basic calibration has been performed, while the PDC flux is further corrected to remove instrumental and systematic effects by fitting a linear combination of the so-called Co-trending Basic Vectors (CBV) (Van Cleve & Caldwell 2009; Kinemuchi et al. 2012). These vectors describe the common trends observed in the 50% less variable and more mutually correlated targets falling within each CCD of the focal plane (Smith et al. 2012; Stumpe et al. 2014).

Considering our interest in the out-of-transit variations on timescales comparable with the stellar rotation, we shall made use of the long-cadence data of Kepler-30 publicly available at the MAST archive[1]. They covers four years and the mean relative accuracy of each data point is ∼260 ppm.

3. Kepler-30 preparation and pre-processing

The PDC pipeline was designed to maximize the chance of detecting planetary transits and not to study stellar intrinsic variability on timescales much longer than the transit duration. Therefore, the removal of the instrumental and systematic trends by fitting a linear combination of CBVs often results in overfitting and removal of the intrinsic target variability. The effect is particularly pronounced for the brighter and more intrinsically variable targets and leads to a remarkable attenuation of the components of the variability with timescales longer than 15-20 days (Stumpe et al. 2014; Gilliland et al. 2015). This is particularly relevant for our analysis. On the other hand, the SAP timeseries, although presenting some residual instrumental effects and long-term systematic trends, is not affected by this problem. Therefore, we shall investigate the variability of Kepler-30 by analysing both the SAP and PDC timeseries.

Because of the flux offsets between adjacent quarters and the long-term instrumental trends within each quarter (cf. Jenkins et al. 2010b), the time series corresponding to different quarters were separately treated as follows:

Firstly, in each quarter residual outliers were removed as described in Sect. 2 of De Medeiros et al. (2013). This procedure resulted in less than 0.2% of data points being flagged as outliers and discarded. Furthermore, each Kepler timestamp has a quality flag which alerts of systematic or instrumental effects that may affect the quality of the photometric measurement. We discarded all the datapoints having the quality flag SAP_QUALITY different from zero because they could potentially be affected by such systematics (Kinemuchi et al. 2012). In this way, the percentage of data points removed was less than 2%.

After applying this procedure, we checked its impact by calculating the cross-correlation coefficient for zero lag (i.e., when the timeseries are not shifted) of the series before and after removing the outliers and datapoints with non-zero quality flag. To this end, we use the MATLAB function `crosscorr`[2]. The values of the coefficients found were 0.92 for the PDC data and 0.94 for the SAP one. The gaps in our timeseries are short enough as to have a negligible impact on our analysis. Therefore, we decided to retain them and to avoid gap filling because artifacts could arise from the necessary interpolation processes.

Secondly, a third-order polynomial was fitted to the data to detrend the time series, separately for each quarter, as suggested by de Freitas et al. (2019b). The planetary transits had a very

[1] `http://archive.stsci.edu/kepler`

[2] See `https://in.mathworks.com/help/econ/crosscorr.html`, for more details.

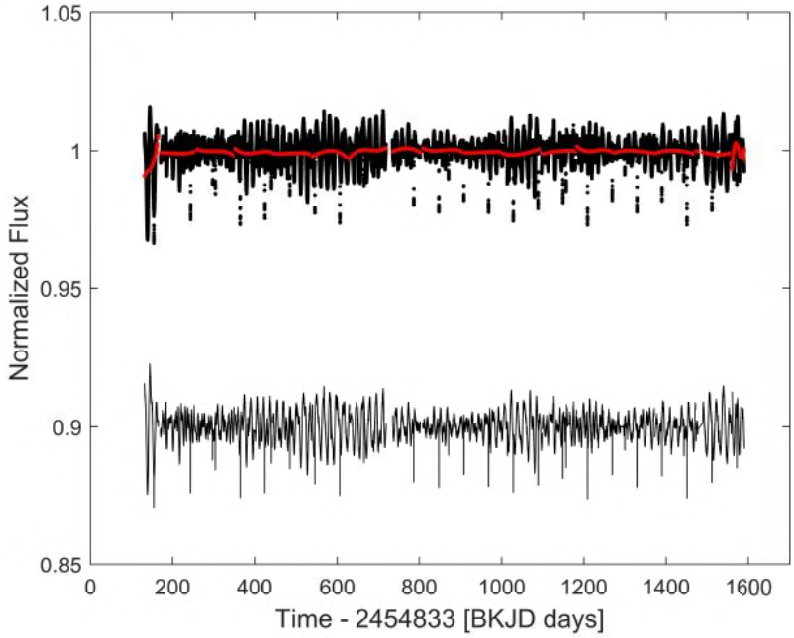

Fig. 1. Kepler-30 PDC time series. Top subfigure: PDC time series (black dots) and polynomial fits of order 3. Bottom subfigure: final PDC time series adjusted by polynomial functions. The signatures of three planets are maintained.

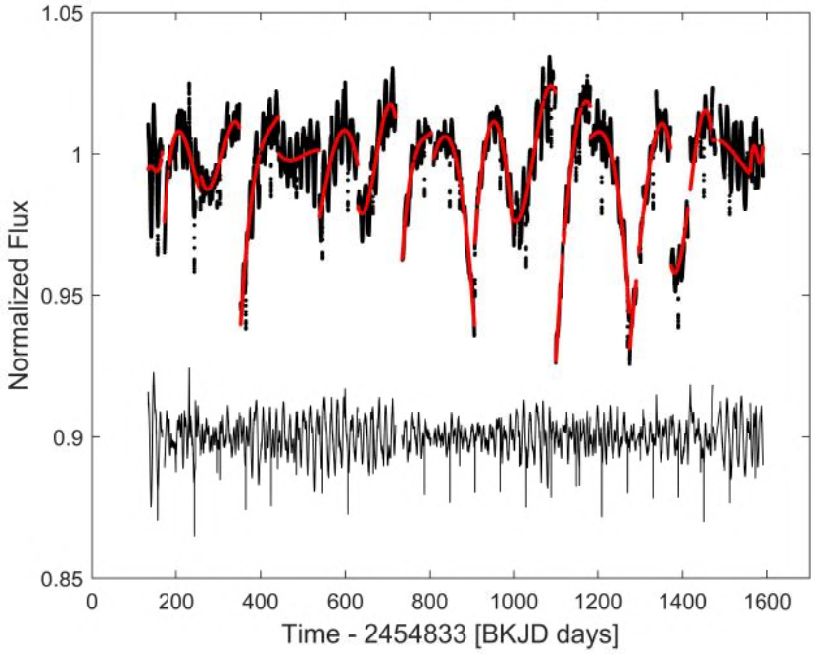

Fig. 2. Idem Figure 1 for SAP time series.

small impact on the polynomial fits given the small planetary radii and their long orbital periods. Our interest is in investigating the out-of-transit fluctuations on the timescale of stellar rotation or longer, so the deformation of the transit profiles is not an issue also in view of their removal in the next step. Figures 1 and 2 show the result of the two steps described above.

In the third step, the `WOTAN` package proposed by Hippke et al. (2019) was used to detrend the out-of-transit part of the signal in preparation for the transit removal in the fourth step (see below) as can be seen in the top panels of Figs. 3 and 4 (red lines). To this end, we use the iterative biweight method with a window of 0.75 days as proposed by Hippke et al. (2019). The detrended flux for both pipelines can be seen in the middle panels of Figs. 3 and 4.

In the fourth step, planetary transits were removed according to the parameters given in Table 1 of Sanchis-Ojeda et al. (2012). We also performed a visual inspection and checked that the obtained out-of-transit series has no evident transit signal left

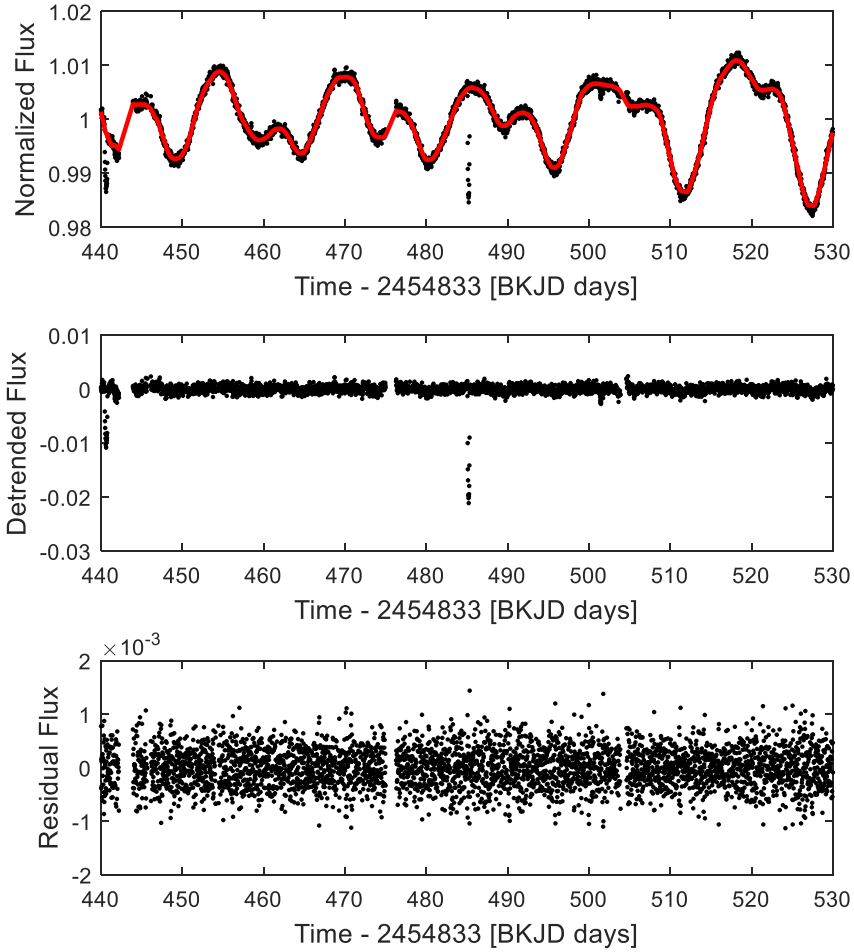

Fig. 3. Detrending of Kepler-30 with the iterative biweight filter using window size of 0.75 days (red line), which allows for a detection of the transits. Normalized flux (top), detrended flux (middle) and residual flux (bottom), both only shown for a segment of the total time series. Black dots indicate the PDC pipeline of Kepler-30 after the first three steps of our procedure. Detrended flux is a combination of noise background and planet transits, indicating the trending of the time series potentially due to rotational modulation of the star.

after the application of our procedure. The gaps introduced by discarding the transits do not exceed 9 hours, i.e., the duration of the transit of the most distant planet Kepler-30d. These gaps shall have a negligible effect on our spot modelling because the rotation period of the star is ∼16 days. In the same way as mentioned above, we did not perform any interpolation process to fill in gaps, since the number of data removed is negligible. Examples of the final results of our procedure can be seen in the bottom panels of Figs. 3 and 4.

Lastly, the final time series is obtained by combining the 17 quarters and consists of four years of data. From each quarter series, the median value was subtracted and then all the residual quarter series were stitched together. The final timeseries are shown in the upper panels of Figs. 5 and 6, where the median of the errors of the single photometric measurements is 2×10^{-4} in relative flux units. The number of datapoints is $N = 60007$ and $N = 60018$ for the PDC and SAP timeseries, respectively.

3.1. Lomb–Scargle periodograms and residual flux between SAP and PDC pipelines

The final PDC and SAP time series and their Lomb–Scargle periodograms (cf. Lomb 1976; Scargle 1982) with the significance levels are illustrated in Fig. 5. For both timeseries, there are two main peaks at 7.8 and 16.8 days. Nevertheless, for SAP time series, a lower peak at 46.6 days also appears. As mentioned above, the SAP pipeline preserves other variabilities that are attenuated or removed by the PDC pipeline. Consequently, the Residual Time Series (hereafter RTS) (i.e., the difference) between the SAP and PDC time series can reveal the origin of that peak (see upper panel of Fig. 6).

In Fig. 6 (bottom panel), we plot the Lomb–Scargle periodogram of the difference time series (RTS). Only a few points were lost during the process of subtracting the time series. The Lomb–Scargle periodogram of Fig. 6 clearly shows that the period of 46.6 days, as well as the periods of 23.1 and 33.9 days, are preserved by the SAP pipeline. The Lomb–Scargle technique has a clear disadvantage, because it is limited to analysing in the frequency space and therefore is not capable of resolving the time-

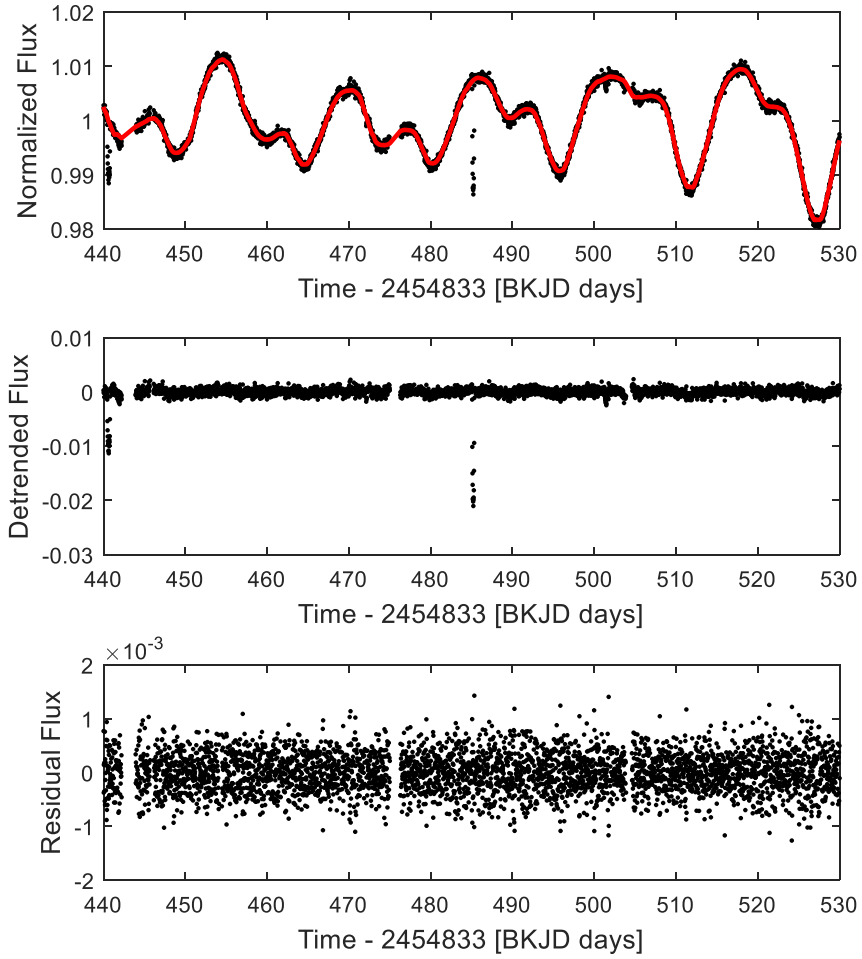

Fig. 4. Idem Figure 3 for SAP pipeline.

evolution of a given periodicity. In addition, this technique does not allow us to analyse the behaviour of the variability, and of the light fluctuations in particular, over different timescales and amplitudes. To this end, we shall consider now the spot modeling (cf. Sect. 4) and the MFDMA approaches (cf. Sects. 5and 6).

4. Spot modeling of Kepler-30 time series

The small $v \sin i \sim 1.94 \pm 0.25$ km s^{-1} of Kepler-30 (Fabrycky et al. 2012) prevents the application of Doppler Imaging techniques to this star (Strassmeier 2009), therefore, the only possibility of mapping its surface is a spot modeling based on the inversion of its light curve.

Spot modeling is notoriously an ill-posed problem because it tries to extract a bi-dimensional map from a one-dimension time-series. To account for the inherent degeneracies of the solution, we apply a regularization to our inversion. It consists in introducing a priori information to select the unique spot map that maximizes the configurational entropy of the spot pattern for a given level of the chi square of the fit. We choose the configurational entropy to be maximal when the star is unspotted, that is, when the spot map has the least spotted area.

The fundamental assumption of our spot model is that the spot pattern stays stable for a time interval $\Delta t_{\rm f}$ during which the flux received from the star is modulated only by stellar rotation. A moderate random starspot evolution on a timescale significantly shorter than $\Delta t_{\rm f}$ is not a problem because it contributes to the noise level and does not affect the spot modeling thanks to the application of the Maximum Entropy (hereinafter ME) regularization (see Appendix C). However, sizeable changes in the spot pattern on a timescale comparable with $\Delta t_{\rm f}$ can produce systematic errors in the spot map. An optimal selection of $\Delta t_{\rm f}$ is therefore relevant for our modeling and is discussed in Appendix C together with the determination of the other parameters of our model.

A main advantage of modeling a star with transiting planets such as Kepler-30 is the possibility of measuring the obliquity of its spin axis projected on the plane of the sky that allows us to constrain the inclination i of the stellar spin to the line of sight (Sanchis-Ojeda et al. 2012). Given the low projected obliquity, we assume the stellar spin to be normal to the plane of the orbit of the more massive close-by transiting planet Kepler-30c, that is, virtually perpendicular to the line of sight, thus fixing the geometry of the rotating star to be mapped (see Appendix C for the adopted value of the inclination). Note that in the case of a star

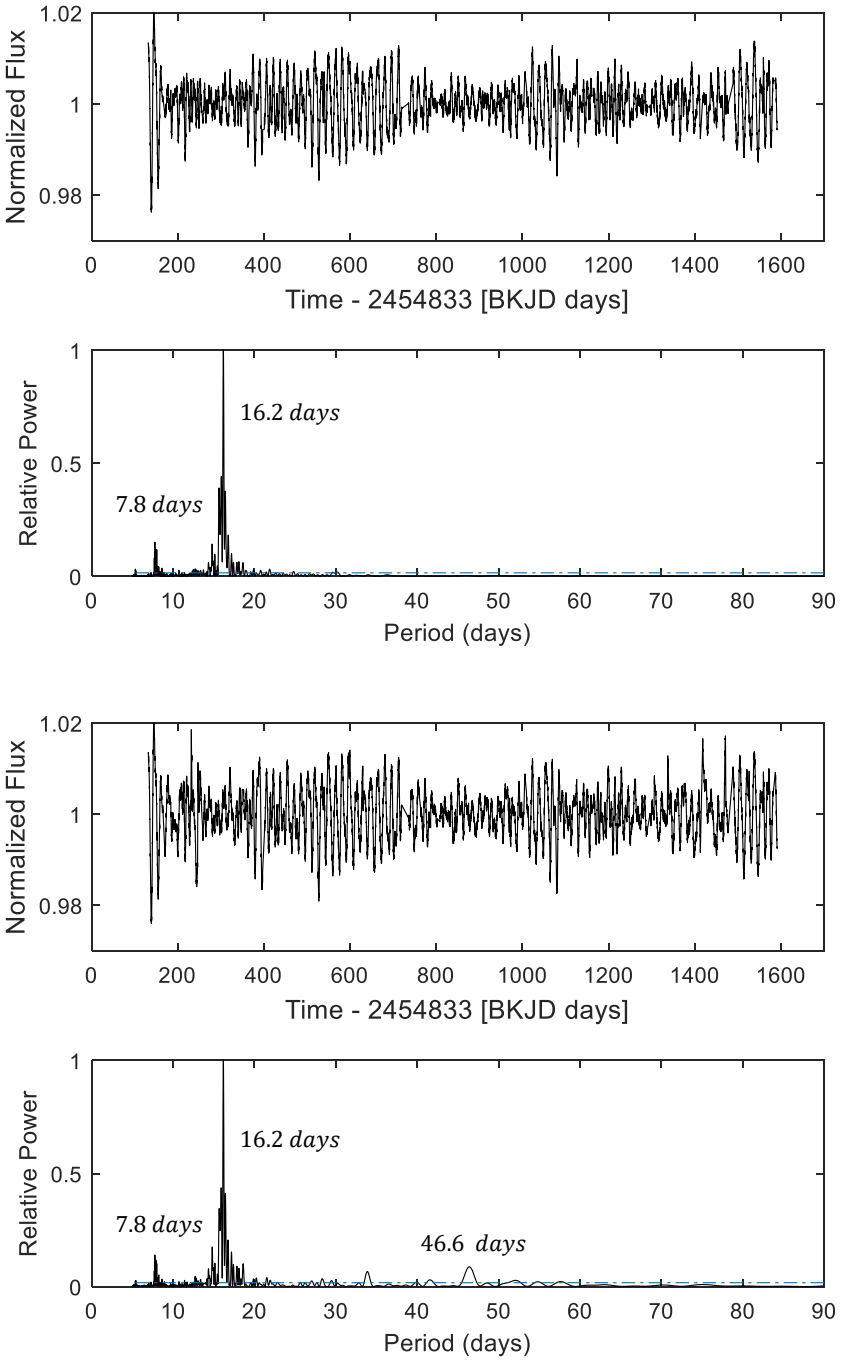

Fig. 5. Final PDC (first panel) and SAP (third panel) time series adjusted according to steps of Section 3. The Lomb-Scargle periodograms are on the bottom side correspondingly. The dashed-dot blue line is the significance threshold given by letting FAP (False Alarm Probability) equals to 10^{-4}. For SAP data, there is a significant peak to 46.6 days also found in Figure 6. As can be seen, the PDC timeseries has a flat periodogram for periods longer than 20 days confirming that the PDC pipeline tends to remove all the long-term trends.

with an unknown inclination, spot modeling is much more uncertain because of the strong degeneracy between the inclination and the latitudes of the spots.

Since we observe Kepler-30 virtually equator-on, spot modeling is unable to provide information on the latitude of the starspots because the duration of their visibility along a stellar rotation is independent of their latitude (we neglect the effect of latitudinal differential rotation the value of which is unknown a priori). Therefore, we can only map the distribution of the spotted area vs. the longitude and the time along successive intervals of duration Δt_f. As a consequence, we consider two-dimensional (2D) spot maps only as an intermediate step in our modeling and derive the distribution of the spotted area vs. longitude from them, that is, compute one-dimensional maps of the filling factor by integrating over the latitude in the 2D maps. Such longitude distributions of the filling factor can be directly related to the

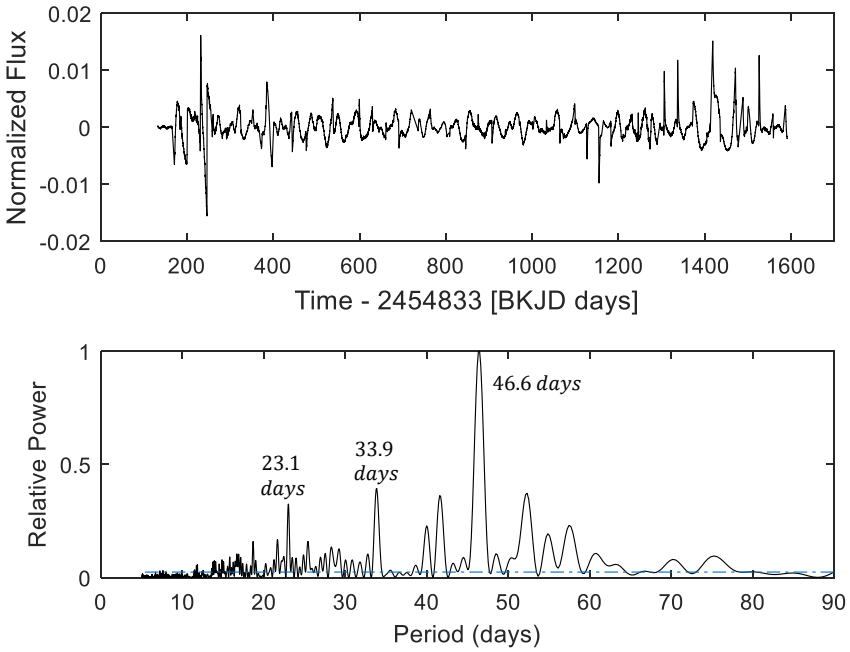

Fig. 6. Difference between the SAP and PDC time series, where the peak of 46.6 days is evidenced. A second and third peaks of 23.1 and 33.9 days are also evidenced.

light modulation under the assumption that the spot contrast is constant. This greatly reduces the degeneracies associated with the non-uniqueness of light curve inversion.

We include the effect of solar-like faculae in our model, but their contribution to the flux modulation is found to be very small and can be neglected for a star as active as Kepler-30 where dark spots dominate (cf. Radick et al. 2018, see also Appendix C). More details of our spot modeling and the impact of the modeling parameters on the spot maps are presented in Appendix C.

Recent works have questioned the validity of spot models assuming a few discrete spots on the stellar surface (e.g., Basri & Shah 2020). Our model is based on a continuous distributions of spots and solar-like faculae on the stellar photosphere and has been extensively tested in the case of the Sun (Lanza, Bonomo & Rodonò 2007) as well as of active stars with transiting planets by comparing the spot maps from light curve inversions with the real distributions of the sunspot groups and the spots detected by occultation during planetary transits, respectively (Silva-Valio & Lanza 2011; Lanza et al. 2019). Thanks to such comparisons, we are confident that our spot modelling is capable of reconstructing the distribution of the spot area vs. the longitude with a resolution of $\approx 60^\circ$ in the case of Kepler-30. In consideration of the systematic differences introduced by the Kepler pipeline between the SAP and the PDC timeseries, we shall apply our modelling to both of them and compare the results to look for artifacts produced by those systematics.

5. Multifractal description of timeseries

5.1. Basic concepts and definitions

The light variations in the time series of Kepler-30 display a wide range of timescales ranging from tens of minutes characteristic of photospheric convection, to the rotational modulation due to photospheric brightness inhomogeneities (starspots), up to months or years characteristic of stellar activity cycles. The rotational modulation signal is quasi-periodic and non-stationary owing to the evolution of starspots on the surface of the star. To characterize the fluctuations on the shortest timescales, mostly produced by stellar convection, we can use methods of multifractal timeseries analysis, that we briefly introduce below. We mainly follow the treatment by Kantelhardt (2015) to whom we refer the reader for a more detailed account because here we limit ourselves to a minimal introduction with the purpose of making this paper accessible also to readers who do not possess previous knowledge of the field.

To characterize the wide frequency spectrum of convective fluctuations, it is useful to look for self-similarity properties in the fluctuations themselves. Specifically, we define $\Delta f(t)$ as the light fluctuation at the time t, measured with respect to some constant mean value. A rescaling of the time by an arbitrary factor a, i.e., $t \rightarrow at$ may require a re-scaling of the fluctuations by a factor a^H, that is, $\Delta f(t) \rightarrow a^H \Delta f(at)$, in order for the rescaled fluctuations to follow the same statistical distribution as the initial ones. A timeseries that obeys this scaling property for an arbitrary factor a is called a self-affine or self-similar timeseries and the exponent H is called the Hurst exponent of the timeseries. Several stochastic fluctuations, including those associated with turbulent convection, satisfy this definition.

The Hurst exponent H characterizes the kind of self-similarity. The simplest example is the timeseries of the mean-square displacement $x^2(t)$ of a point performing a random walk with a normally distributed step size. In this case, the time scaling $t \rightarrow at$ requires a corresponding scaling $x \rightarrow a^{1/2}x$ in order to preserve the statistical distribution of the variable x. In other words, the timeseries $x(t)$ exhibits a self-similar appearance with $H = 0.5$, that is, by expanding the time axis by a factor a and the space axis by a factor a^H the distribution of the statistical fluctuations of $x(t)$ is the same as the original one. Such a scale self-

similarity (or scale invariance) is one of the fundamental properties that define fractals.

Self-affine timeseries are persistent in the sense that a large fluctuation is likely followed by another large fluctuation and a small fluctuation by a small one. The persistence holds on all the timescales for which the self-affine scaling holds. In more complex systems, the persistence may change according to the timescale being stronger on some timescales and weaker on others. For example, the weather timeseries are usually persistent on timescales of hours or days, but the degree of persistence is higher on shorter timescales. Moreover, the characteristic persistence timescale generally changes with the season usually being longer during winter and summer and shorter in autumn and spring.

Considering a stochastic self-affine timeseries $x(t)$ with the variable sampled at equidistant times t_i, with $i = 1, 2, ..., N$, we define the timeseries of the increments $\Delta x_i \equiv x(t_i) - x(t_{i-1})$. In the case of the simple random walk model considered above with $H = 0.5$, the Δx_i are independent of each other, but when $H > 0.5$, a positive Δx_i is likely to be followed by a positive increment (persistence), while, when $H < 0.5$, by a negative one (anti-persistence). The degree of persistence can be quantified by the auto-correlation function in the case of a stationary timeseries, that is, one with constant mean and standard deviation. If we denote the timescale as n, also indicating a number of datapoints in a uniformly sampled timeseries, the autocorrelation is defined as:

$$C(n) = \frac{1}{\sigma^2}\frac{1}{N-n}\sum_{i=1}^{N-n}\Delta x_i \Delta x_{i+n}, \tag{1}$$

where σ^2 is the variance of the timeseries of the increments Δx_i.

When the Δx_i are uncorrelated, $C(n) = 0$ for $n > 0$. A short-range correlation can be described by an exponentially decaying $C(n) \propto \exp(-n/n_\mathrm{d})$, where n_d is the decay timescale. For example, this is the case of an autoregressive process defined by the linear recursion $\Delta x_i = c\Delta x_{i+1} + \epsilon_i$, where $c = \exp(-1/n_\mathrm{d})$ and the ϵ_i are normally distributed random deviates.

In the case of long-term correlations, the asymptotic behaviour of $C(n)$, that is, its scaling for large n, can be described by a power law:

$$C(n) \propto n^{-\gamma}, \tag{2}$$

where the exponent $0 < \gamma < 1$ characterizes the long-term correlations. It can be shown that a long-term correlated, i.e., persistent, behaviour of Δx_i leads to a self-affine scaling of the $x(t_i)$ characterized by an Hurst exponent $H = 1 - \gamma/2$. Therefore, the Hurst exponent can be used to characterize the autocorrelation function of the increments of the timeseries.

Ideally, the estimation of the exponent γ can be made by computing the autocorrelation function of the increments or the power spectrum of the original time series $x(t_i)$ (Kantelhardt 2015). In practice, these approaches are doomed to fail because of the noise that dominates the autocorrelation and the power spectrum in the asymptotic regime of large n values to be considered when deriving γ. Therefore, more robust statistical approaches have been developed for such a determination. They are based on the so-called *fluctuation functions* $F_q(n)$ that measure the scaling of the fluctuation amplitude with the timescale n, where q is a parameter defining the moment of the statistical distribution of the $x(t_i)$ measured by the function itself (see Appendix A for examples of fluctuation functions). The fluctuation function based on the standard deviation corresponds to $q = 2$ and plays a special role in the analysis.

The scaling of the fluctuation function with n is characterized by a generalized Hurst exponent or Holder exponent $h(q)$ that depends on the moment of the distribution sampled by $F_q(n)$. To define the fluctuation function $F_q(n)$, we start from the root-mean-square (RMS) fluctuation function $F_\nu(n)$ that is defined by

$$F_\nu(n) = \left\{\frac{1}{n}\sum_{i=1}^{n}\epsilon_\nu^2(i)\right\}^{\frac{1}{2}}, \tag{3}$$

where $\epsilon_\nu(i)$ is the deviation of the i-th datapoint from a suitable moving average introduced to remove the long-term variation in the timeseries (see Appendix A for its precise definition). Starting from the RMS fluctuation function, we define the fluctuation function of order q as

$$F_q(n) = \left\{\frac{1}{N_n}\sum_{\nu=1}^{N_n}F_\nu^q(n)\right\}^{1/q} \quad \text{for } q \neq 0 \tag{4}$$

and, for $q = 0$,

$$\ln\left[F_0(n)\right] = \frac{1}{N_n}\sum_{\nu=1}^{N_n}\ln[F_\nu(n)], \tag{5}$$

where N_n is the number of non-overlapping segments for a given segment of size n in the timeseries, i.e., $N_n \sim N/n$, where N is the total number of datapoints in the series (see Appendix A for the precise definition of N_n). For larger values of n, the fluctuation function scales following a power-law given by

$$F_q(n) \sim n^{h(q)}, \tag{6}$$

that allows us to define the generalized Hurst exponent $h(q)$.

The previously defined Hurst exponent corresponds to the value attained for $q = 2$, that is, $H \equiv h(2)$. By definition, when $h(q)$ is constant, we speak of a fractal (or monofractal) time series, while when h depends on q we have multifractal time series. In the former case, the scaling exponent of the fluctuation self-affine law is independent of the amplitude of the fluctuations, while in the latter case it depends on their amplitude because different values of q corresponds to different moments that sample fluctuations of large or small amplitudes in different ways (cf. equation A.5).

The scaling of the fluctuations with their amplitude can be quantitatively characterized by means of the multifractal spectrum and its characteristic parameters as detailed in Appendix A. Multifractal timeseries can be produced, for example, by simple, non-linear, recurrence laws, such as the so-called logistic law $x_{i+1} = \lambda_\mathrm{ll} x_i(1 - x_i)$, where the coefficient λ_ll falls in a range leading to a chaotic behaviour (e.g., Luo & Han 1992).

5.2. Origins of multifractality and the impact of quasi-periodic modulations in the timeseries

The multifractal character of a stochastic timeseries is manifested by different scaling exponents for small and large fluctuations on the same timescale n. This behaviour can be originated by the simultaneous presence of different long-term correlations and persistences that depend on the amplitude of the fluctuations and/or by their statistical distribution that, even in the absence of long-term correlations, deviates strongly from a Gaussian (normal) distribution. In the latter case, the distribution of the fluctuations usually shows fat tails as, for example, in the case of a Cauchy distribution.

It is possible to discriminate between these two sources of multifractality by considering the so-called surrogates of the original timeseries. A first kind of surrogate is simply produced by randomly shuffling the datapoints over the set of the N indexes i. In this way, we keep the statistical distribution of the datapoints, but remove all short- and long-term correlations producing a new timeseries with $h(q)$ close to 0.5, that is the Hurst exponent of a random time series consisting of uncorrelated points. A second kind of surrogate can be obtained by taking the Fourier transform of the original timeseries and constructing a modified Fourier transform having the same modulus, but a phase randomly extracted in the $[-\pi, \pi]$ interval. The backward transformation of this modified transform into the time domain gives a timeseries the points of which retain most of the correlations present in the original timeseries, but whose distribution is nearly Gaussian, thus effectively eliminating the effect of non-Gaussianity in the original distribution (e.g., Kantelhardt 2015; Wang, Shang & Cui 2014).

By comparing the $h(q)$ function and the multifractal spectrum of the original timeseries with those of the above surrogates, we can extract information on the dominant source of multifractality. Specifically, when the dominant source of multifractality is time correlation, the $h(q)$ function of the randomly shuffled series will be remarkably different from that of the original series, while the $h(q)$ of the phase-randomized series will be similar. On the other hand, when the dominant source of multifractality is a non-Gaussian statistical distribution of the datapoints, the converse will be true.

From a more general point of view, timeseries can result from both deterministic and stochastic processes, the former usually being responsible for long-term trends or oscillations. Those trends should be removed to allow the best determination of the properties of the stochastic fluctuations. This is usually done by fitting the trends with polynomials of different degrees in the calculation of the fluctuation functions (e.g. Kantelhardt 2015) or by subtracting a moving average as in our case (see Appendix A). However, in the case of astrophysical systems, and of magnetically active stars in particular, a clear separation between deterministic long-term trends and short-term stochastic fluctuations is not always possible. For example, the rotational modulation of the flux due to photospheric starspots contains a stochastic component produced by the evolution of the individual starspots often appearing randomly in time and/or longitude (e.g. Lanza et al. 2019, and references therein).

In light of such an impossibility of performing a clear separation between the different components of the photometric timeseries of magnetically active stars, the adopted approach has been that of including the rotational modulation and the evolution of the surface brightness inhomogeneities as additional sources of multifractality. In other words, the multifractal analysis itself has been used as a technique to characterize those processes in addition to standard techniques such as Lomb-Scargle periodogram or autocorrelation to measure stellar rotation periods (e.g., McQuillan et al. 2013) or spot modelling to derive the evolution of the surface features (Lanza 2016). This approach has been extended even to the detection and characterization of strictly periodic phenomena such as planetary transits in the presence of noise (e.g., Agarwal et al. 2017).

In a sample of magnetically-active solar-like stars observed by the CoRoT space telescope, de Freitas et al. (2013a) found a correlation between the Hurst index and their rotation period by assuming that their timeseries were monofractal. In a subsequent work on the same sample, de Freitas et al. (2016) established that their timeseries are indeed multifractal, that is, the correlations produced by magnetic activity and rotation are better described by assuming a generalized Hurst exponent $h(q)$ and adopting the full set of parameters introduced in Appendix A to characterize the multifractal spectrum, in particular $\Delta\alpha$.

Later, de Freitas et al. (2017) analysed a sample of 34 M dwarf stars previously investigated by Mathur et al. (2014), to test the behaviour of the Hurst exponent against the magnetic activity indicator $S_{\rm ph}$. The latter index is based on the standard deviation of the flux in the Kepler passband computed over non-overlapping time intervals of k mean rotation periods $P_{\rm rot}$ after removing the contribution of the photon shot noise to the standard deviation. An average index $\langle S_{\rm ph}\rangle_k$ is then defined as the arithmetic mean of all the individual $S_{\rm ph}$ indexes determined from each of the $kP_{\rm rot}$ intervals along the timeseries. Therefore, it includes the contributions of all the sources of light variations from the shortest timescales up to $kP_{\rm rot}$, including the rotational modulation. On the other hand, the index $H = h(2)$ is defined as the scaling exponent of the RMS fluctuation function $F_2(n) \propto F_\nu(n) \sim n^H$ that is evaluated with the fluctuations obtained after subtracting a moving average from the timeseries. This gives a relatively greater weight to the shorter timescales, while reducing the contribution of the rotational modulation and longer timescales. Therefore, with a generally adopted value of $k = 5$ (Mathur et al. 2014), the correlation between the H index and the $\langle S_{\rm ph}\rangle_k$ index is expected to be low because of such a difference in the respectively dominant timescales. This expectation was confirmed by de Freitas et al. (2017) who found a slight anti-correlation between the two indexes with a Pearson coefficient of -0.33 when considering the above mentioned sample of M dwarf stars (see their Figure 6).

On the other hand, even within a limited range in the rotation period (< 15 days), de Freitas et al. (2017) confirmed the strong correlation between H and $P_{\rm rot}$ and found the Hurst exponent to be an indicator of magnetic activity. Recently, de Freitas et al. (2019a,b) extended the field of application to timeseries showing more than one significant rotational periodicity, which can be considered as an indication of stellar differential rotation. Using a large sample of Kepler active stars, de Freitas et al. (2019a) showed that in the relationship between the Hurst exponent and the relative amplitude of the differential rotation $\Delta P/P$ a strong trend of increasing $\Delta P/P$ toward larger H-index is observed. In addition, the study shows that the correlation is stronger for the most active stars in the sample.

6. MFDMA analysis

Recently, de Freitas et al. (2019a) used multifractal analysis to investigate the multi-scale behaviour of a set of ∼ 8000 active stars that were observed by the Kepler mission. de Freitas et al. (2016, 2017, 2019a,b) showed that the multifractal detrending moving average (MFDMA) algorithm, which was developed by Gu & Zhou (2010)[3] and Tang et al. (2015), is a powerful technique that provides valuable information on the fluctuations of a time series. A complete description of the MFDMA algorithm can be found in Appendix A.

Several authors (e.g., Gieratowski, Zebrowski & Baranowski 2012; Wang, Shang & Cui 2014) have suggested that a single scaling exponent ($H = h(2)$) is inadequate to describe the internal dynamics of complex signals, and consequently, classical (multi)fractal methods (e.g., DFA and MFDFA) have been

[3] MATLAB codes for MFDMA analysis can be found in the arXiv version of Gu & Zhou (2010)'s paper: https://arxiv.org/pdf/1005.0877v2.pdf

modified for an estimation of the temporal dependence of the spectrum of scale exponents. These improved methods are used to better quantify the short- and long-range correlations, that is important in our case because multifractality in our timeseries arises mostly because of the correlations (cf. Sect. 4.2 and Table A.1).

Inspired by these studies, we decided to introduce another method to calculate a time-dependent scaling exponent $h(q, \tau)$, where τ is a timescale (see below). We call it the multiscale MFDMA method (MFDMAτ) and it allows us to extend the investigation of stellar variability by including a time dependence in the description of the multiscale fluctuations. MFDMAτ is relatively immune to additive noise and non-stationarity, including the non-stationarity due to occurrence of events of a different dynamics.

To explore the multiscale fluctuations in flux, we introduce the increments $\Delta x(t, \tau) = x(t + \tau) - x(t)$, where τ varies from 29.4 min (Kepler cadence) to ∼60 days. We apply the MFDMA analysis to the increment time series $\Delta x(t, \tau)$, thus obtaining an Hurst exponent $h(q, \tau)$ that is a function also of the timescale τ and in this way introduce the MFDMAτ method. The consideration of the increment timeseries in the MFDMAτ approach strongly reduces the effects of all the stationary modulations having a period equal to τ or to one of its multiples. In other words, this method allows us to isolate the effects of the fluctuations after removing those periodicities. Given the remarkable rotational modulation present in our time series, this method will allow us to study the properties of the fluctuations after removing the modulation itself with an appropriate selection of the timescale τ.

6.1. Optimal delay time τ

We explored different values of τ using a plot of $x(t + \tau)$ vs. $x(t)$ following an approach reminiscent of that used to study multifractal attractors by means of plots of the trajectory of the dynamical system in a bi-dimensional phase space. In Fig. 7 we plot the case corresponding to the optimal value $\tau = 8.7$ days that shows a repetitive, quasi-stationary pattern organized around a fixed point with a behaviour similar to that of an attractor.

According to Henry, Lovell & Camacho (2001), the optimal delay time τ can be determined using the approach of topological properties based on the behaviour of the autocorrelation function. There are various proposals for choosing an optimal delay time using the autocorrelation function. We choose the best value for τ as that corresponding to the maximum of the mean value of the autocorrelation coefficient. In the subplot of Fig. 7, we show that the mean autocorrelation coefficient of the stochastic process $x(t)$, defined as $E[x(t + \tau)x(t)]$ (for all t), is maximum for $\tau = 8.7$ days.

7. Results and Discussions

We present the results of our spot modelling first and then introduce the results of the application of the multifractal analysis to the original photometric timeseries of Kepler-30 as well as to the residuals of the spot models. In this way, we can apply the results of the spot modelling to elucidate those of the multifractal analysis, discussing the characteristic rotational and activity timescales found in the data.

7.1. Best fit of the Kepler-30 light curves

In Fig. 8 we plot the PDC light curve together with the unregularized best fit obtained with our model and the corresponding residuals. The unregularized best fit provides the minimum chi square and effectively reproduces the spot light modulation giving residuals where it has been filtered out.

The distribution of the residuals is shown in Fig. 9 and is almost Gaussian with a mean $\mu_{\rm res}$ very close to zero ($\mu_{\rm res} = -2.18 \times 10^{-7}$) and a standard deviation $\sigma_0 = 3.78 \times 10^{-4}$ in relative flux units, slightly larger than the photon shot noise. This indicates the presence of other processes that contribute to the random noise level, probably associated with surface convection and magnetic fields affecting the stellar flux.

A GLS periodogram of the residuals is shown in Fig. 10. Our spot model strongly reduces any variability with periods longer than about 6 days leaving the variability below 3 − 4 days almost intact. The maximum of the GLS is reached at a period of 2.565 days. The peak is about two times higher than the noise level as indicated by the other nearby peaks, thus it is probably not very significant – usually a peak reaches a significance of 99 percent when its height is at least four times the noise level –, but we shall discuss its possible origin in Sect. 7.3.1.

The unregularized best fit of the SAP light curve is very similar to that of the PDC light curve except for a few larger residuals probably associated with some small uncorrected instrumental effects. By fitting the residual distribution with a Gaussian, we find a slightly larger mean $\mu_{\rm reg} = -1.737 \times 10^{-6}$, but a standard deviation $\sigma_0 = 3.746 \times 10^{-4}$ in relative flux units virtually identical to that of the PDC residual distribution. Also the periodogram of the residuals is very similar and is not shown here. The residuals of the PDC and SAP light curve fits will be used to study the multifractal properties of the stellar flux variability without the additional complication introduced by the light modulation produced by starspots. Analysing the residuals of the unregularized best fits will allow us to avoid possible systematic effects introduced by the a priori assumptions adopted for the ME regularization that increase the standard deviation and make the mean of the residuals systematically negative (see Appendix C).

7.2. Spot maps and variation of the spotted area

The spot map obtained from the ME regularized best fit of the PDC light curve is plotted in Fig. 11, where we plot the spot filling factor vs. the longitude and the time. The adopted longitude reference frame rotates with the mean rotation period of the star of 16.0 days and the longitude increases in the direction of stellar rotation. Therefore, spots rotating with the mean rotation period of 16 days stay at a constant longitude on this plot. On the other hand, spots rotating faster than the mean rotation period show a longitude that increases vs. the time, while those rotating slower show a longitude decreasing in time. We see that the evolution of the starspots of Kepler-30 is rather fast with sizeable variations of the filling factor over timescales of 20-30 days (the time resolution of our spot modelling is $\Delta t_{\rm f} = 11.963$ days). We introduce a modified Julian date as MJD = HJD −2454900. During the first half of the modelled time series, i.e., up to MJD ∼ 800, spots cluster around two main active longitudes at ∼ 0° and ∼ 200°, while a smaller and intermittent active longitude is present in between them around ∼ 100°. Starting from MJD ∼ 300, the main active longitudes migrate towards greater and lower longitudes, respectively, probably as a consequence of the decay of the spots rotating with the mean rotation period and the formation of new spots that rotate faster and slower, respectively. The intermediate

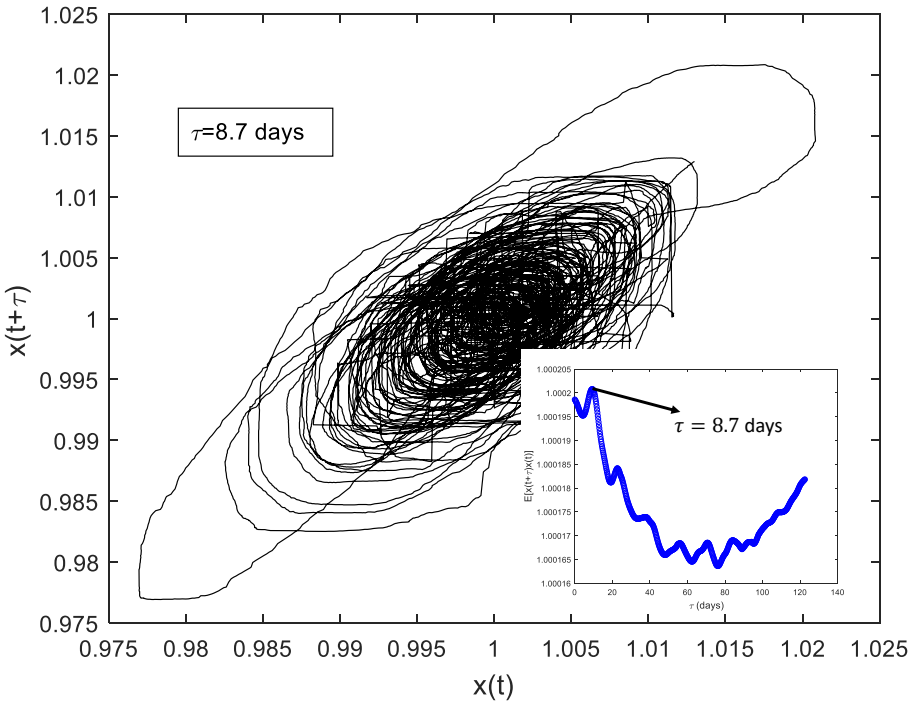

Fig. 7. The time-delayed PDC timeseries $x(t + \tau)$ with the optimal delay $\tau = 8.7$ days vs. the original time series $x(t)$. The subplot shows that $\tau = 8.7$ days corresponds to the maximum value of the autocorrelation function of the timeseries.

active longitude is approached by the main longitude initially at $\sim 200°$ and looses its identity. For MJD $\gtrsim 800$, the active longitudes are less well defined and the pattern is characterized by individual spots with lifetimes ranging from ≈ 30 to ≈ 200 days that rotate at different rates. The spot map obtained with the regularized best fit to the SAP light curve is similar to that displayed in Fig. 11 and is plotted in Appendix C.

The migration of the different spots in longitude can be used to estimate the surface differential rotation, if we assume that the migration is caused by the location of those spots at different latitudes on a differentially rotating star. Given the lack of information on the spot latitudes, we can only derive a lower limit to the surface shear. The estimate should be regarded with great caution because the apparent migration can be the result of the starspot evolution rather than of the differential rotation. In the case of Kepler-30, the evolution of individual starspots occurs on timescales shorter than or comparable with those of the migration of the active longitudes, thus this source of systematic errors is certainly present and casts a further uncertainty on our estimate.

Looking at the spot map in Fig. 11, the best suitable feature to estimate the differential rotation is the active longitude that starts its migration at MJD ~ 400 at a longitude of $\sim 30°$ and reaches a longitude of $\approx 100°$ at MJD ~ 550. Taking into account an uncertainty of at least $\pm 40°$ in the longitude shift between those two dates owing to the intrinsic width and spreading of the active longitude in time, we estimate a relative surface shear of $\Delta\Omega/\Omega \sim 0.020 \pm 0.012$, where Ω is the angular velocity of rotation, the mean value of which corresponds to the mean rotation period of 16 days. A similar estimate based on the spot map obtained from the SAP timeseries (cf. Fig. C.1) gives a comparable result. In any case, we stress that this estimate is particularly uncertain, especially because of the fast spot evolution seen inside the active region with individual spot lifetimes of tens of days, i. e., comparable with the time resolution of our spot modelling. Other active longitudes show a less clearly defined migration pattern and are more affected by the intrinsic evolution of the spots; thus, they are not considered for our estimate.

The total spotted area, obtained by summing up the filling factor over all the longitudes, is plotted in Fig. 12 vs. the time. Only the values obtained from time intervals Δt_f without significant data gaps have been plotted in Fig. 12 to avoid systematic errors associated with the ME regularization that tends to decrease the spotted area in the time intervals containing gaps. To find if the distribution of the datapoints within each interval Δt_f is uniform, we subdivide it into five equal subintervals and count the number q_i of datapoints in each subinterval ($i = 1, .., 5$). If the ratio $(\max\{q_i\} - \min\{q_i\})/\max\{q_i\} \leq 0.2$, the corresponding interval Δt_f is considered as not affected by gaps and its area value is considered in our analysis. In spite of disregarding the area values obtained from intervals with significant gaps, we still see some points with deviations up to ± 10 percent from the mean value in Fig. 12. Such relatively large deviations are due to flux discontinuities at the epochs when successive intervals are stitched together after a repointing of the Kepler telescope or to some small deviations from the convergence criterion adopted to fix the level of regularization (see Appendix C).

A GLS periodogram of the spotted area shows a peak at 33.8 days with a false-alarm probability (FAP) of 0.29 according to the analytical formula by Zechmeister & Kürster (2009). The corresponding best fitting sinusoid is plotted in Fig. 12 with the area values affected by the systematic effects mentioned above clearly deviating from it. Repeating the same analysis with the spot model of the SAP light curve (see Appendix C), not affected by the systematic effects introduced by the PDC pipeline, we find the same periodicity with a period of 33.9 days and a FAP = 0.024, thus confirming the presence of a short-term