FliPer: A global measure of power density to estimate surface gravities of main-sequence solar-like stars and red giants

L. Bugnet[1,2], R. A. García[1,2], G. R. Davies[3,4], S. Mathur[5,6,7], E. Corsaro[8], O. J. Hall[3,4], and B. M. Rendle[3,4]

[1] IRFU, CEA, Université Paris-Saclay, 91191 Gif-sur-Yvette, France
e-mail: lisa.bugnet@cea.fr
[2] AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité, 91191 Gif-sur-Yvette, France
[3] School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
[4] Stellar Astrophysics Centre, Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, 8000 Aarhus C, Denmark
[5] Instituto de Astrofísica de Canarias, 38200 La Laguna, Tenerife, Spain
[6] Universidad de La Laguna, Dpto. de Astrofísica, 38205 La Laguna, Tenerife, Spain
[7] Space Science Institute, 4750 Walnut Street Suite 205, Boulder, CO 80301, USA
[8] INAF – Osservatorio Astrofisico di Catania, Via S. Sofia 78, 95123 Catania, Italy

ABSTRACT

Asteroseismology provides global stellar parameters such as masses, radii, or surface gravities using mean global seismic parameters and effective temperature for thousands of low-mass stars ($0.8\,M_\odot < M < 3\,M_\odot$). This methodology has been successfully applied to stars in which acoustic modes excited by turbulent convection are measured. Other methods such as the Flicker technique can also be used to determine stellar surface gravities, but only works for $\log g$ above 2.5 dex. In this work, we present a new metric called FliPer (Flicker in spectral power density, in opposition to the standard Flicker measurement which is computed in the time domain); it is able to extend the range for which reliable surface gravities can be obtained ($0.1 < \log g < 4.6$ dex) without performing any seismic analysis for stars brighter than $Kp < 14$. FliPer takes into account the average variability of a star measured in the power density spectrum in a given range of frequencies. However, FliPer values calculated on several ranges of frequency are required to better characterize a star. Using a large set of asteroseismic targets it is possible to calibrate the behavior of surface gravity with FliPer through machine learning. This calibration made with a random forest regressor covers a wide range of surface gravities from main-sequence stars to subgiants and red giants, with very small uncertainties from 0.04 to 0.1 dex. FliPer values can be inserted in automatic global seismic pipelines to either give an estimation of the stellar surface gravity or to assess the quality of the seismic results by detecting any outliers in the obtained $\nu_{\max}$ values. FliPer also constrains the surface gravities of main-sequence dwarfs using only long-cadence data for which the Nyquist frequency is too low to measure the acoustic-mode properties.

Key words. asteroseismology – methods: data analysis – stars: oscillations

1. Introduction

The precise knowledge of stellar parameters is crucial for a very broad range of fields in astrophysics. It helps us understand stellar evolution; this knowledge also provides important information needed for planetary searches and for studying the chemical and dynamical evolution of our Galaxy. In the last decade, the NASA mission *Kepler* (Borucki et al. 2013) continuously collected very high-quality photometric data for almost 200 000 stars (Mathur et al. 2017) over ∼4 years. These observations not only revolutionized the search for exoplanets, but also opened a window into stellar physics. Asteroseismology proved to be a very powerful tool able to better characterize the stars in terms of mass, radii, and age (Metcalfe et al. 2010; Mathur et al. 2012; Silva Aguirre et al. 2017; Serenelli et al. 2017), and also in terms of their rotation and magnetic activity (McQuillan et al. 2014; García et al. 2014a; Davies et al. 2015; Ceillier et al. 2017; Kiefer et al. 2017). However, stellar oscillations have not been detected in all red giant (RG) stars (∼16 000 reported out of the ∼24 000 in the latest *Kepler* star-properties catalog Mathur et al. 2017) or in all the main-sequence (MS)

solar-like stars. Around 135 000 MS dwarfs have only been observed in long cadence (LC) with a sampling time of 29.4 min, by *Kepler* preventing any direct asteroseismic analyses because their acoustic-mode frequencies are well above the Nyquist frequency and can only be seismically studied with short-cadence (SC) data with a sampling time of 58.85 s (Chaplin et al. 2011b).

To circumvent this, new techniques are being developed to extract precise surface gravities ($\log g$) directly from the photometric data. This is the case of the Flicker method, i.e., the measurement of the brightness variations on timescales shorter than 8 h (Bastien et al. 2013, 2016), the variance in the flux (Hekker et al. 2012), and the granulation (Mathur et al. 2011; Kallinger et al. 2014), and the analysis of the timescales of convective-driven brightness variations (Kallinger et al. 2016). However all these techniques have limitations. Flicker is restricted by construction to stars with $4500 < T_{\rm eff} < 7150$ K and $2.5 < \log g < 4.6$ dex, preventing the study of high-luminosity red-giant branch (RGB) and asymptotic red-giant branch (AGB) stars. To obtain the granulation properties it is necessary to fit a complicated model including different scales of convection with many free parameters (for more details see the

discussions in Mathur et al. 2011; Kallinger et al. 2014; Corsaro et al. 2017). The final method requires that the oscillation signal be temporally resolved, which prevents extending the analysis to MS dwarfs only observed in LC data. It has also been shown that instead of using classic seismic methods, it is possible to apply machine learning algorithms directly on the data. For instance, Hon et al. (2018a) apply a convolutional neural network on spectra to classify stars. This method gives good results for about 99% of their sample of RGs, including some stars that were not already characterized with seismic pipelines. A random forest regression model (Breiman 2001) applied directly on the photometric light curves of variable stellar sources can also estimate their surface gravity with a 0.42 dex uncertainty (Miller et al. 2014).

We present here a new metric called FliPer (Flicker in power, in opposition to the standard Flicker measurement which is computed in the time domain), which links the variability of a star to its surface gravity in a wider range than Flicker, starting at a $\log g \sim 0.1$ and similar effective temperatures ($4500 < T_{\text{eff}} < 7150$ K) covering solar-like pulsating stars. We are limited in the $0.1 < \log g < 4.6$ dex range of surface gravity because of the lack of information we have on extreme surface gravity solar-like stars. There is no intrinsic limits of applicability to the FliPer calculation. We decide to combine powerful methods: we include FliPer values from different lower frequency boundaries into a supervised machine learning random forest algorithm in order to get even more accurate results on the surface gravity estimation. This way, we obtain information about the impact of the lower frequency boundaries and the effective temperature on the estimation of surface gravity.

2. Observations, data selection, and preparation

In this work, LC data (Gilliland et al. 2010) obtained by NASA's *Kepler* main mission are used. The light curves have been corrected and the different quarters concatenated following García et al. (2011). Two high-pass filters have been used with cutoff frequencies corresponding to 20 and 80 days. To minimize the effects of the gaps in the observations (García et al. 2014b) the missing observations have been interpolated using inpainting techniques (Pires et al. 2015). The power spectrum density is then computed for each star (calibrated as a single-sided spectrum). Data are corrected for apodization following Chaplin et al. (2011a).

We selected ~15 000 RG stars among the ones in Mathur et al. (2017) showing stellar pulsations and characterized using the A2Z asteroseismic pipeline (Mathur et al. 2010). These stars have $0.1 < \log g < 3.4$ dex and 3285 K $< T_{\text{eff}} < 7411$ K. In addition, 254 MS stars with 4951 K $< T_{\text{eff}} < 6881$ K are used to extend the study towards a higher surface gravity range, reaching 4.5 dex. These stars have *Kepler* magnitudes brighter than 14 ($Kp < 14$).

It is important to note that the values of ν_{max} computed by A2Z do not show any systematic biases at a level of ~1% when compared to other seismic pipelines as shown by Pinsonneault et al. (2018).

3. The new metric: FliPer

The complete power spectrum contains contributions from the stellar variability at all timescales, such as oscillation modes, surface granulation, and rotation. We define FliPer as

$$F_p = \overline{\text{PSD}} - P_n, \tag{1}$$

where $\overline{\text{PSD}}$ represents the averaged value of the power spectrum density from a given frequency (see Sect. 3.1) to the Nyquist frequency and P_n is the photon noise. This noise can be calculated by taking the average value of the PSD over a range of frequencies close to the Nyquist frequency, but this method leads to biased estimation of FliPer for stars that oscillate with a frequency close to the Nyquist frequency, as explained in detail by Bugnet et al. (2017). Then the photon noise is computed following the empirical expression obtained by Jenkins et al. (2010).

The value of FliPer is dominated by a combination of the granulation and the oscillation modes, which both depend on the evolutionary stage of the star. The more evolved the star, the larger their oscillation and granulation amplitudes (e.g., Mosser et al. 2012; García & Stello 2015), while the frequency of maximum power ν_{max} decreases (e.g., Bedding 2014).

It is important to note that the signature of strong rotation (and its harmonics) would bias FliPer. This does not have a large impact in the case of RGs because a very small fraction of them show signatures of the rotation in the PSD, as shown by Ceillier et al. (2017), but needs to be studied in detail for MS solar-like stars (see Sect. 3.3).

3.1. Computing FliPer from data

The observational frequency range used to compute $\overline{\text{PSD}}$ is limited at high frequency by the Nyquist frequency. For most stars (those observed in LC) we cannot get information above ~283 μHz. Therefore, we selected a first set of calibrator stars including a RG pulsating at a frequency lower than 300 μHz and for which asteroseismic parameters were available. A second set of known seismic MS dwarfs was used to study FliPer with LC data only.

The low-frequency limit of $\overline{\text{PSD}}$ is given by the cutoff frequency used in the calibration of the data. For most of the stars, a 20-day high-pass filter light curve is used. The associated cutoff frequency of the signal is 0.58 μHz. We thus establish a low-frequency limit for the analysis at 0.7 μHz. As MS stars can rotate with a period shorter than 20 days, FliPer is computed with a low-frequency limit at 7 μHz (i.e., ~1.6 days) avoiding most of the pollution induced by rotation signals (see Sect. 3.3). For stars showing rotation harmonics at higher frequencies, the low-frequency boundary should be taken even higher (e.g., 20 μHz) to avoid any additional impact on FliPer from the peaks associated with rotation. Finally, a small number of RGs in our sample are either high-luminosity RGB or AGB stars ($\log g < 1.2$ dex) pulsating at frequencies smaller than the 20-day cutoff frequency of the calibrated data. For these stars, an 80-day filter is used in the calibration process. It allows us to properly measure the stellar signal down to 0.2 μHz (which is the limit frequency utilized in this analysis) and to include oscillation-mode power into the FliPer value.

3.2. First surface gravity estimator

For stars with solar-like oscillations, seismic surface gravities are directly obtained from the frequency of maximum oscillation power ν_{max} computed with the A2Z pipeline (Mathur et al. 2010) and effective temperatures from the *Kepler* DR25 catalog (Mathur et al. 2017). Knowing seismic surface gravities with their uncertainties allows us to study the behavior of FliPer with the evolutionary state of the stars using only LC light curves even for MS stars. It is important to note that for MS stars the seismic
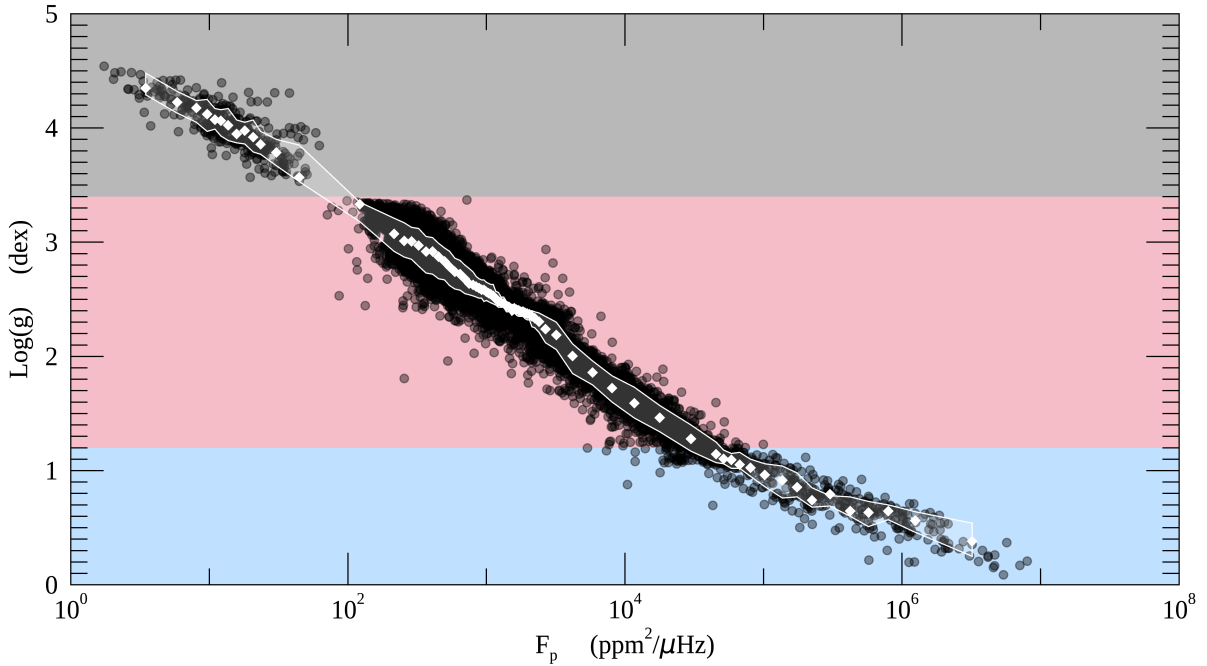
Fig. 1. Seismic log *g* vs FliPer for ~15 000 stars observed with *Kepler* long-cadence observational mode (black dots). The gray, red, and blue shaded regions correspond to MS and subgiant stars, RG, and high-luminosity RG stars or AGB, respectively. White diamonds represent the weighted mean value of log *g* (see Eq. (2)) computed with 30 MS stars and 300 RGs each. The white area delimits the location where 68% of the stars in our sample are around the mean value. The white boundary represents the equivalent of a 1σ uncertainty (standard deviation) on the surface gravity obtained from FliPer.

log *g* has been seismically inferred using SC data although FliPer has been computed using LC data.

In Fig. 1 the seismic log *g* is represented as a function of FliPer (black dots). Three different areas have been identified depending on the evolutionary state of the star: MS and subgiant stars (gray shaded region), RG (red), and high-luminosity RG stars from the branch and from the asymptotic branch (blue). For each of these star categories FliPer was computed with a different low-frequency limit of 7μHz (avoiding in most cases the region of possible pollution by rotation signatures present on data filtered with a high-pass filter at 20μHz), 0.7μHz (20-day filter), and 0.2μHz (80-day filter), respectively. The color scheme is the same in the captions.

In order to characterize the relationship between FliPer and log *g* represented in Fig. 1, we calculate an average value of log *g* for each bin of *n* stars (*n* = 300 for RG and *n* = 30 for MS and subgiant stars) as

$$\overline{\log g} = \frac{\sum_{i=1}^{n} \frac{1}{\delta \log g_i} \times \log g_i}{\sum_{i=1}^{n} \frac{1}{\delta \log g_i}}, \qquad (2)$$

where $\log g_i$ represents the value of surface gravity for each star and $\delta \log g_i$ the corresponding uncertainty.

These values are represented by the white diamonds, and are located at the averaged value of FliPer over each bin. To define the 1σ uncertainties, we compute the area containing 68% of the stars of the sample (marked by a white contour region in Fig. 1). Mean values and their corresponding ±1σ uncertainties are given in Table A.1. By using these mean values it is possible to estimate the stellar surface gravity directly from the FliPer estimator. The uncertainties obtained on log *g* extend from 0.05 to 0.2 dex, depending on the evolutionary state of the star. Because of the calculation of the mean values, the boundaries in log *g* are reduced to the range 4.35–0.38 dex, as shown in Table A.1.

3.3. Disentangling main-sequence stars from red giants

As defined, FliPer is mostly dominated by a combination of the power coming from granulation and oscillation modes (when the latter are below the Nyquist frequency). The limitation in the use of the calibrated values from Table A.1 to directly estimate surface gravity of stars appears when the spectrum shows a specific behavior that strongly modifies the mean value of the power density. For instance, in stars showing large excess of power (e.g., due to spikes at thrusters frequency in K2 data or to pollution from a background binary), the value of FliPer is biased towards high power density (Bugnet et al. 2017). On the contrary, in stars with a low signal-to-noise ratio the value of FliPer is biased towards lower values because most of the spectrum is dominated by the instrumental noise. As a consequence, FliPer is higher than expected for fast rotating MS stars due to the rotation peaks and their harmonics, which can be particularly high for young MS dwarfs. For these stars, the log *g* inferred from Table A.1 could be such that it corresponds to a RG star and not to a MS star, even if we calculate FliPer with the 7μHz frequency limit.

To avoid this problem and to disentangle any MS stars from RGs, we need an additional parameter that takes into account the power due to rotation. The simplest solution is to combine different FliPer values, including some at higher frequencies than the 7μHz limit. For each star in our sample, we then calculate FliPer with several low-frequency limits (e.g., $F_{\mathrm{P}_{0.2}}$ from 0.2, $F_{\mathrm{P}_{0.7}}$ from 0.7, F_{P_7} from 7, $F_{\mathrm{P}_{20}}$ from 20, and $F_{\mathrm{P}_{50}}$ from 50μHz). For MS stars with small rotation signatures the value of FliPer is almost the same for all the low-frequency boundaries (see panel a in Fig. 2). However, when rotation peaks are present, there is a large difference between the FliPer parameters, depending on the frequency of the rotational peaks (see Fig. 2). This is the case for the stars KIC 8298090 and KIC 5357446 represented in panels b and c. In panel b all the rotational components are below the 0.7μHz
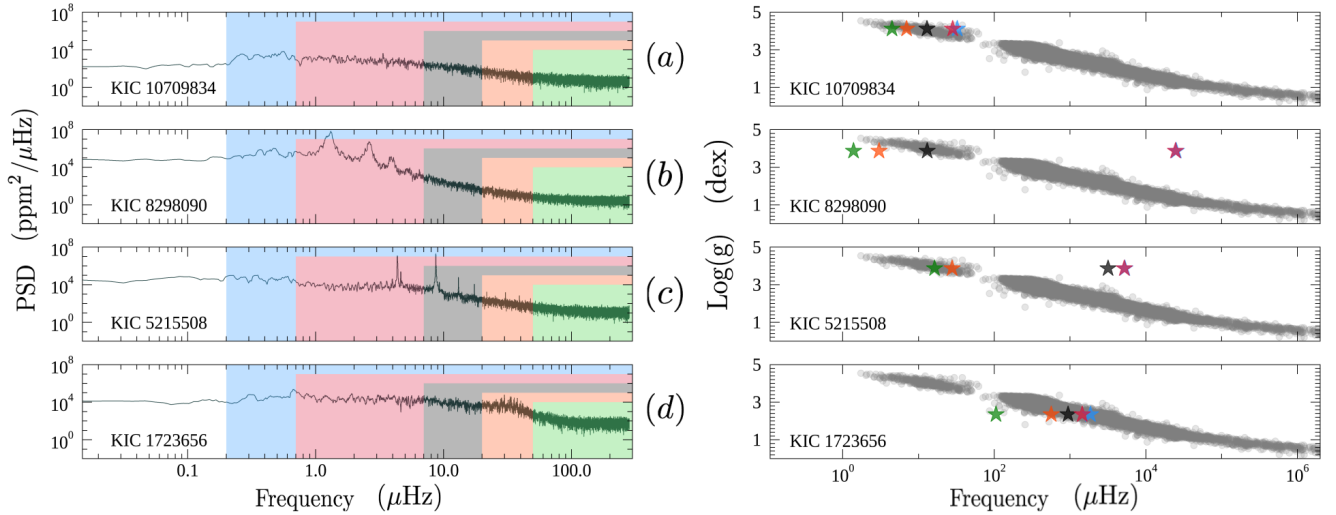
Fig. 2. Impact of the lower frequency limit in the FliPer calculation on the estimation of surface gravity for different types of star. *Left panels*: power density spectra of four *Kepler* targets. Colored areas (blue, red, black, orange, and green) represent the different ranges of frequency used for FliPer calculation (respectively from 0.2, 0.7, 7, 20, and 50 μHz to the Nyquist frequency). The color scheme is the same in the captions. *Right panels*: all studied *Kepler* stars from Fig. 1 are represented in gray in the log *g* V.S FliPer diagram. Colored stars (blue, red, black, orange, and green) show the positions in the diagram of the four stars from the left panels with their color corresponding to the low-frequency boundaries used to compute the FliPer value. *Panel a* represents a MS star without any visible rotation component, *panel b* a MS star showing rotation, *panel c* a high-frequency rotating MS star, and *panel d* a RG star.

boundary, meaning that parameters $F_{p_{20}}$ and $F_{p_{50}}$ were not necessary to classify this star as a MS star. However, in panel c the rotation peaks reach higher frequencies; in order to estimate the surface gravity of this star the two new high-frequency parameters are needed. Panel d shows a RG star for comparison. In the regime of RG stars, all the FliPer values are very similar, except the lowest ones coming from the calculation with the high-pass filter that does not include the range of frequency of oscillation modes. By comparing the values of FliPer computed with different low-frequency limits, it is then possible to disentangle MS stars with a high rotation signature from RG stars. This can be done in a star-by-star analysis (see Fig. 2). However, it is possible to automatize this procedure, as is explained in the following section.

4. Seismic independent surface gravity prediction from 0.1 to 4.5 dex

The direct estimation of surface gravity from Table A.1 gives good results only when the evolutionary state of the star is already known, and when the spectrum does not show a specific behavior that strongly modifies the mean value of the power density (e.g., when the PSD is polluted by spikes of a background binary or a classical pulsator). The reason is that we only use one value of FliPer computed from one lower frequency limit. Estimating surface gravities of unclassified or complex stars requires a different use of the FliPer method.

4.1. Using machine learning

As explained above, combining different FliPer values is a powerful way to detect MS stars showing high rotation signal among RGs. It also means that by using different high-pass filters in the calculation of FliPer we are sensitive to different physical signatures in the PSD. Combining them in the study thus improves the characterization of the star, and we intend to use this wisely to predict surface gravities. To do so, we train a random forest

regressor algorithm (Breiman 2001) on a random subsample representing 80% of our set of stars. The random forest method is based on the aggregation of a large number of decision trees (see Appendix B for a description of the method) that has already been proved to be useful in asteroseimology (e.g., Miller et al. 2014). The trees are constructed from a training data set and internally validated to give a prediction based on the predictor for future observations. The random forest method not only allows the use of a large number of parameters but also estimates their individual impact on the regression. The parameters used to estimate surface gravities are $F_{p_{0.2}}$, $F_{p_{0.7}}$, F_{p_7}, $F_{p_{20}}$, $F_{p_{50}}$, T_{eff}, and Kp. They represent the values of FliPer calculated from a low-frequency limit (0.2, 0.7, 7, 20, 50 μHz), the effective temperature, and the *Kepler* magnitude of the star.

4.2. Building training and test sets of parameters for the random forest algorithm

We intend to take into account the uncertainties on the parameters during the testing of the algorithm to estimate the intrinsic bias and/or uncertainties of our methodology. The uncertainties on effective temperature δT_{eff} are taken directly from the Mathur et al. (2017) catalog. The error on the surface gravity $\delta \log g$ comes from the uncertainty on ν_{max} from the seismic analysis of the stars. We can estimate the uncertainty due to the photon noise in the spectra (following a chi-squared distribution with two degrees of freedom) impacting the determination of FliPer by considering negligible the uncertainty made on the photon noise. Hence, the uncertainty on FliPer can be explicitly written as

$$\delta F_p = \sqrt{\overline{\delta \text{PSD}}^2} = \frac{\delta P_{\text{tot}}}{N_{\text{bin}}} = \frac{\sqrt{\sum_i \delta P_i^2}}{N_{\text{bin}}}, \qquad (3)$$

where δP_i stands for the error made on the power contained in each bin and N_{bin} is the total number of bins in the power density spectrum. The individual δP_i cannot be extracted directly
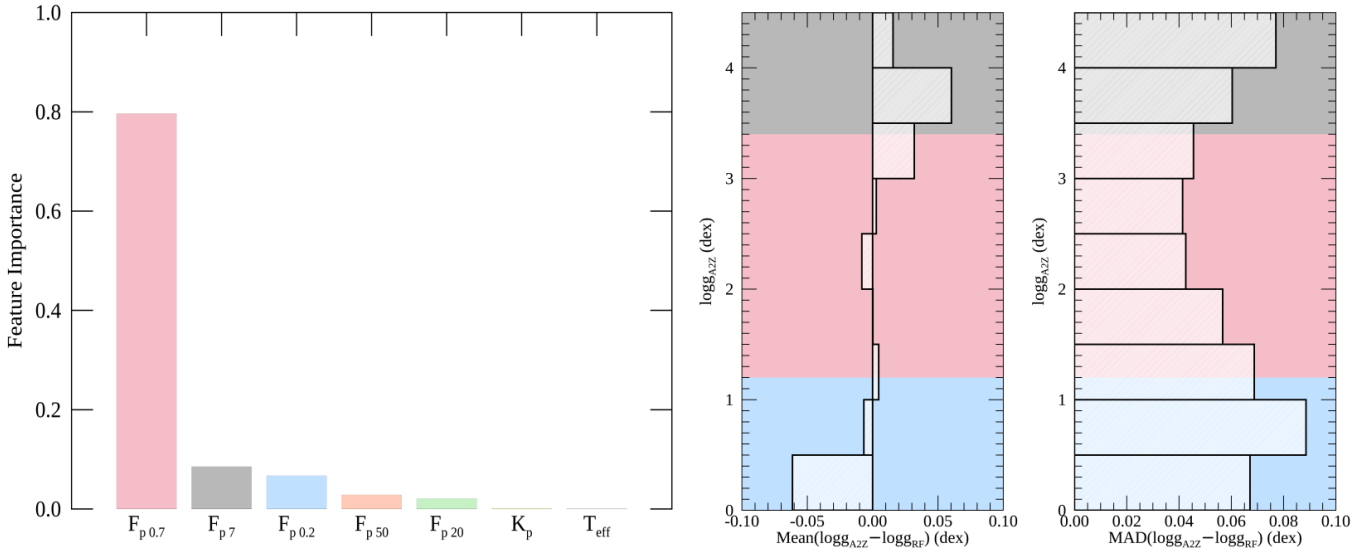
Fig. 3. *Left panel*: importance of the different parameters $F_{p_{0.2}}$, $F_{p_{0.7}}$, F_{p_7}, $F_{p_{20}}$, $F_{p_{50}}$, T_{eff}, and Kp in the training process. The color scheme is the same in the captions. *Middle panel*: histogram of the residuals of the estimated $\log g_{RF}$ values from the references $\log g_{A2Z}$. *Right panel*: histogram of the mean absolute deviation from the expected value MAD($\log g_{A2Z} - \log g_{RF}$).

because the χ_2^2 noise distribution does not have Gaussian errors. We thus use the central limit theorem and we rebin the spectrum by a factor of $n = 50$. The total amount of power in the spectrum is then expressed as

$$P_{tot} = \sum_j P_{n,j}, \tag{4}$$

where $P_{n,j}$ follows a quasi-normal distribution with $2n$ degrees of freedom. This assumes that the signal does not change dramatically over this range of 50 bins, which is consistent with the shape of the spectra in solar-like stars. The uncertainty on the mean from each n bins is then expressed as

$$\delta P_{n,j} = 2 \times \frac{P_{n,j}}{2n} \times \sqrt{n}, \tag{5}$$

leading to a global uncertainty on FliPer values of

$$\delta F_p = \frac{\sqrt{\sum_j \left(2 \times \frac{P_{50,j}}{2n} \times \sqrt{n}\right)^2}}{N_{bin}}. \tag{6}$$

Then we include the effect of these errors on the different parameters on the testing of the algorithm. To do so, we perform a Monte Carlo simulation by generating for each star in our test sample (representing 20% of the total number of stars in our study) 100 artificial sets of parameters from their corresponding normal distributions. With $\mathcal{G}_{0 \leq i \leq 100}$ being 100 random values following the standard distribution, we calculate for each X parameter ($F_{p_{0.2}}$, $F_{p_{0.7}}$, F_{p_7}, $F_{p_{20}}$, $F_{p_{50}}$, T_{eff}, and $\log g$) 100 new values $X_{0 \leq i \leq 100}$ following Eq. (7) below. However, the *Kepler* magnitude of the star remains constant as it has no uncertainties, and it completes each of the 100 new sets of parameters:

$$X_{0 \leq i \leq 100} = X + \delta X \times \mathcal{G}_{0 \leq i \leq 100}. \tag{7}$$

Each new group of 100 sets of parameters is considered in the following study as a representation of a hundred stars to test the algorithm.

4.3. Impacts of parameters on the training

We used 80% of our stars to train the algorithm to estimate the surface gravity. The remaining 20% was used to test the performance of the algorithm by taking into account uncertainties on the different parameters as explained in Sect. 4.2. The impacts of the different parameters on the training process are represented in Fig. 3 A.

A predictable result is that the $F_{p_{0.7}}$ parameter largely dominates the training. It comes from the fact that this is the most suitable parameter to study RG, representing more than 90% of the total number of stars. Other relevant values of the filtering appear to be 7 and 0.2 μHz. Indeed, F_{p_7} plays an important role in the training because of its ability to distinguish MS stars from RGs, and the $F_{p_{0.2}}$ parameter helps in the prediction of surface gravity for high-luminosity stars. The other parameters $F_{p_{20}}$ and $F_{p_{50}}$ have lower impacts on the training, but still help the learning for high rotating MS stars. Impacts of the effective temperature and Kp do not exceed a small fraction. We confirm from Fig. 3 that combining different lower frequency boundaries in the FliPer calculation makes a great difference for the estimation of robust surface gravities.

4.4. Results

To evaluate the performance of the algorithm, the estimate of surface gravity from the test sample is compared to the corresponding A2Z estimation of surface gravity. The mean absolute deviation (MAD) of the random forest surface gravities from reference values is reported in Table 1. This estimator of the deviation is chosen to be robust against outliers to avoid any issue coming from an eventual remaining error in the A2Z estimation of surface gravity.

The estimation of surface gravity resulting from the machine learning on the test sample has an averaged deviation of ~0.046 dex from our reference values (see Table 1). We also obtain surface gravity deviation and errors from the reference values for different ranges of surface gravity in Table 1. We conclude that for all our stars, the new method gives a very good precision on surface gravity; the Flicker method in the range of $\log g$

Table 1. Summary of statistical results on the test set from Fig. 3.

log g range (dex)	$\overline{\log g_{A2Z} - \log g_{RF}}$ (dex)	MAD (dex)
ALL	-4.5×10^{-4}	0.046
[0−0.5]	−0.061	0.067
[0.5−1]	−0.007	0.089
[1−1.5]	0.005	0.069
[1.5−2]	0.000	0.057
[2−2.5]	−0.008	0.043
[2.5−3]	0.003	0.046
[3−3.5]	0.032	0.041
[3.5−4]	0.060	0.060
[4−4.5]	0.016	0.077

Notes. MAD is the mean absolute deviation.

of 2.5−4.6 dex has typical errors between 0.1 and 0.2 dex. Here, errors are in the range from ~0.04 dex for RGs to ~0.09 dex for high-luminosity stars (see Fig. 3 right panel). Our estimates are on average centered around the log g reference values (see Table 1). There is a small bias for HL and MS stars (see Fig. 3 middle panel) because the estimation of extreme surface gravities is the hardest for the algorithm, which sometimes becomes slightly biased by the presence of many RGs in the sample. This bias, which depends on the evolutionary state of the star, should be taken into account, but it remains smaller than the uncertainties on the original surface gravity values.

Our algorithms are available on GitHub[1] where the functions to calculate FliPer and the random forest algorithm are provided. We also provide the already trained algorithms for the estimation of surface gravities. They can be directly applied to any solar-type star to estimate its surface gravity from 0.1 to 4.6 dex.

5. Discussion and conclusion

In this work we present a new method used to estimate surface gravity of solar-like stars that extracts information from global power in their spectra. The sample of ~15 000 stars is constituted of MS and subgiant stars, stars on the red-giant branch (RGB) and clump stars, and also high-luminosity stars on the asymptotic giant branch (AGB). This way, we study stars with $0.1 < \log g < 4.5$ dex in which mode oscillations are expected to arise from surface convection. Power spectra should then present patterns of granulation power, rotation components, and oscillation-mode power.

FliPer values are calculated by taking the average power density normalized by the photon noise of the star from different lower frequency limits to the Nyquist frequency. Our first method consists of calibrating surface gravity of stars from their FliPer value with a 1σ uncertainty (see Table A.1). We explained how these values can be used directly to give a first estimate of surface gravity; however, it works only on stars that are already characterized. The evolutionary state has to be known or the star must have a weak rotational signature in order to distinguish MS stars from RGs. To give estimations of surface gravities for any star, we introduce a second method. A random forest regressor algorithm is trained to estimate surface gravity on a sample of our stars. We use FliPer values computed with

different frequency ranges, spectroscopic effective temperatures, and seismic surface gravities. This way, stars are better characterized during the process, and no additional information is needed to provide accurate estimations of surface gravity, even for highly rotating MS stars. By testing the algorithm on the rest of our sample, we obtain estimates of surface gravity with a mean absolute deviation of 0.046 dex from seismic log g. The training relies on seismic observations of solar-like stars representing 80% of our sample. However, there is no need for additional seismic measurements to obtain precise estimations of surface gravities on the test set of stars. The uncertainty on our results largely improves upon previous non-seismic estimations of surface gravity. Spectroscopic estimations are known to have 0.1–0.3 dex error bars (Frasca et al. 2016; García Pérez et al. 2016). Recent methods such as Flicker (Bastien et al. 2016) give estimates with errors higher than 0.1 dex, while the study of the granulation timescale (Kallinger et al. 2016) is limited to stars showing a visible oscillation pattern but with better uncertainties, around 0.018 dex. In addition, FliPer is extended to a wider range of surface gravities, reaching log g as small as 0.1 dex with a mean absolute deviation of 0.046 dex comparable to the other RGs.

For MS stars that oscillate at high frequency (above the *Kepler* LC Nyquist frequency), FliPer computed from LC data does not contain mode power, but only granulation-related power (Corsaro et al. 2017) and rotation signals. However, Fig. 1 clearly shows that FliPer values for MS stars are still correlated with surface gravity. This is new evidence of the link between granulation and asteroseismic properties (Mathur et al. 2011; Kallinger et al. 2014), allowing us to estimate ν_{max} or rather surface gravities on LC data for which high-frequency modes are not measured. Thus, proper surface gravities can be precisely inferred for any *Kepler* LC solar-like target, from MS to high-luminosity stars, without using direct seismic analysis.

Lots of studies concern the estimation of seismic parameters of stars with new techniques directly from the properties of the time series or the power spectrum density (Bugnet et al. 2017; Hon et al. 2018b; Pande et al. 2018; Bell et al. 2018). We thus adapt our methodology to estimate ν_{max} instead of the surface gravity based on the same sample of stars (see Appendix C). The results are of course very similar to those on the surface gravity, with uncertainties on ν_{max} about 0.044 dex and a mean distance to references ν_{max} values ($\sim 1.3 \times 10^{-3}$ dex) negligible. Moreover, FliPer has already been included by Bell et al. (2018) as a validation procedure to their seismic results and it is also being implemented as one of the parameters to be used in the classification algorithm that is being developed for the NASA TESS mission (Ricker et al. 2014) using a random forest classifier (Tkachenko et al., in prep.).

---
[1] https://github.com/lbugnet/FLIPER

References

Bastien, F. A., Stassun, K. G., Basri, G., & Pepper, J. 2013, Nature, 500, 427
Bastien, F. A., Stassun, K. G., Basri, G., & Pepper, J. 2016, ApJ, 818, 43
Bedding, T. R. 2014, Solar-like oscillations: An observational perspective, eds. P. L. Pallé, & C. Esteban, 60
Bell, K. J., Hekker, S., & Kuszlewicz, J. S. 2018, MNRAS, accepted [arXiv:1809.09135]
Borucki, W. J., Agol, E., Fressin, F., et al. 2013, Science, 340, 587
Breiman, L. 2001, Mach. Learn., 45, 5
Bugnet, L., Garcia, R. A., Davies, G. R., Mathur, S., & Corsaro, E. 2017, Proceedings SF2A 2017
Ceillier, T., Tayar, J., Mathur, S., et al. 2017, A&A, 605, A111
Chaplin, W. J., Kjeldsen, H., Bedding, T. R., et al. 2011a, ApJ, 732, 54
Chaplin, W. J., Kjeldsen, H., Christensen-Dalsgaard, J., et al. 2011b, Science, 332, 213
Corsaro, E., Mathur, S., García, R. A., et al. 2017, A&A, 605, A3
Davies, G. R., Chaplin, W. J., Farr, W. M., et al. 2015, MNRAS, 446, 2959
Frasca, A., Molenda-Żakowicz, J., De Cat, P., et al. 2016, A&A, 594, A39
García, R. A., & Stello, D. 2015, Asteroseismology of Red Giant Stars in Extraterrestrial Seismology (Cambridge: Cambridge University Press)
García, R. A., Hekker, S., Stello, D., et al. 2011, MNRAS, 414, L6
García, R. A., Ceillier, T., Salabert, D., et al. 2014a, A&A, 572, A34
García, R. A., Mathur, S., Pires, S., et al. 2014b, A&A, 568, A10
García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., et al. 2016, AJ, 151, 144
Gilliland, R. L., Jenkins, J. M., Borucki, W. J., et al. 2010, ApJ, 713, L160
Hekker, S., Elsworth, Y., Mosser, B., et al. 2012, A&A, 544, A90
Hon, M., Stello, D., & Yu, J. 2018a, MNRAS, 475, 3233
Hon, M., Stello, D., & Zinn, J. C. 2018b, ApJ, 859, 64
Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., et al. 2010, ApJ, 713, L120
Kallinger, T., De Ridder, J., Hekker, S., et al. 2014, A&A, 570, A41
Kallinger, T., Hekker, S., Garcia, R. A., Huber, D., & Matthews, J. M. 2016, Sci. Adv., 2, 1500654
Kiefer, R., Schad, A., Davies, G., & Roth, M. 2017, A&A, 598, A77
Kotsiantis, S. B. 2007, Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies (Amsterdam, The Netherlands: IOS Press), 3
Mathur, S., García, R. A., Régulo, C., et al. 2010, A&A, 511, A46
Mathur, S., Hekker, S., Trampedach, R., et al. 2011, ApJ, 741, 119
Mathur, S., Metcalfe, T. S., Woitaszek, M., et al. 2012, ApJ, 749, 152
Mathur, S., Huber, D., Batalha, N. M., et al. 2017, ApJS, 229, 30
McQuillan, A., Mazeh, T., & Aigrain, S. 2014, ApJS, 211, 24
Metcalfe, T. S., Monteiro, M. J. P. F. G., Thompson, M. J., et al. 2010, ApJ, 723, 1583
Miller, A., Richards, J., Bloom, J. S., & a larger Team 2014, Amer. Astron. Soc. Meet. Abstr., 223, 125.01
Mosser, B., Elsworth, Y., Hekker, S., et al. 2012, A&A, 537, A30
Pande, D., Bedding, T. R., Huber, D., & Kjeldsen, H. 2018, MNRAS, 480, 467
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Learn. Res., 12, 2825
Pinsonneault, M. H., Elsworth, Y. P., Tayar, J., et al. 2018, ApJS, submitted [arXiv:1804.09983]
Pires, S., Mathur, S., García, R. A., et al. 2015, A&A, 574, A18
Ricker, G. R., Winn, J. N., & Vanderspek, R. 2014, SPIE Conf. Ser., 9143, 20
Serenelli, A., Johnson, J., Huber, D., et al. 2017, AJSS, 233, 23
Silva Aguirre, V., Lund, M. N., Antia, H. M., et al. 2017, ApJ, 835, 173

Appendix A: Surface gravity as a function of FliPer

Table A.1. Weighted mean value of log g (diamonds in Fig. 1) with their 1σ uncertainties for each bin of 30 (for MS and HL stars) or 300 (for RGs) stars.

$\log(F_P)$	$\log g$ (dex)	-1σ (dex)	$+1\sigma$ (dex)
0.54	4.35	0.06	0.13
0.77	4.23	0.09	0.09
0.91	4.18	0.12	0.06
0.98	4.12	0.15	0.13
1.04	4.07	0.09	0.10
1.08	4.06	0.14	0.10
1.14	4.02	0.13	0.14
1.20	3.95	0.08	0.12
1.25	3.98	0.12	0.08
1.32	3.92	0.13	0.15
1.37	3.86	0.09	0.11
1.48	3.79	0.12	0.11
1.65	3.57	0.03	0.28
2.09	3.33	0.15	0.02
2.33	3.07	0.15	0.13
2.41	3.01	0.14	0.16
2.46	3.01	0.16	0.13
2.51	2.97	0.18	0.15
2.56	2.92	0.19	0.14
2.61	2.92	0.20	0.13
2.65	2.87	0.19	0.12
2.70	2.83	0.16	0.12
2.73	2.78	0.16	0.13
2.77	2.74	0.15	0.12
2.81	2.72	0.14	0.12
2.84	2.68	0.13	0.12
2.87	2.64	0.10	0.13
2.90	2.63	0.09	0.11
2.93	2.61	0.09	0.09
2.95	2.60	0.09	0.08
2.97	2.58	0.08	0.09
2.98	2.58	0.10	0.09
3.00	2.57	0.08	0.08
3.01	2.55	0.08	0.09
3.03	2.54	0.08	0.08
3.05	2.53	0.07	0.10
3.07	2.50	0.05	0.07
3.09	2.48	0.06	0.08
3.11	2.47	0.04	0.05
3.12	2.46	0.04	0.05
3.14	2.44	0.05	0.04
3.17	2.43	0.04	0.04
3.18	2.42	0.05	0.04
3.20	2.42	0.04	0.03
3.22	2.41	0.05	0.04
3.23	2.40	0.05	0.04
3.25	2.39	0.04	0.03
3.26	2.39	0.04	0.03
3.28	2.38	0.05	0.04
3.29	2.37	0.06	0.04
…	…	…	…

Table A.1. continued.

$\log(F_P)$	$\log g$ (dex)	-1σ (dex)	$+1\sigma$ (dex)
…	…	…	…
3.31	2.37	0.05	0.04
3.33	2.35	0.09	0.05
3.35	2.33	0.09	0.05
3.38	2.30	0.09	0.07
3.43	2.24	0.12	0.11
3.50	2.19	0.13	0.11
3.62	2.00	0.15	0.10
3.77	1.86	0.11	0.10
3.91	1.72	0.12	0.11
4.07	1.59	0.13	0.14
4.25	1.46	0.12	0.11
4.48	1.28	0.11	0.13
4.66	1.14	0.07	0.10
4.71	1.10	0.04	0.08
4.77	1.09	0.07	0.06
4.83	1.05	0.03	0.11
4.91	1.02	0.08	0.08
5.01	0.96	0.10	0.10
5.14	0.92	0.15	0.13
5.24	0.85	0.08	0.13
5.35	0.74	0.06	0.11
5.48	0.79	0.10	0.01
5.63	0.65	0.04	0.12
5.76	0.63	0.12	0.09
5.90	0.65	0.08	0.05
6.10	0.56	0.10	0.08
6.51	0.38	0.13	0.16

Appendix B: Random forest regressor

B.1. Supervised machine learning

A random forest algorithm is a supervised machine learning (ML) method (Kotsiantis 2007). It learns how to predict an output variable ($Y_{\mathrm{predicted}}$) from some training data (X) for which the corresponding result (Y_{known}) is already known. It thus learns a mapping function f from the input(s) to the output:

$$Y_{\mathrm{predicted}} = f(X) \tag{B.1}$$

The algorithm iteratively makes predictions ($Y_{\mathrm{predicted}}$) on the training data (X). These predicted values are corrected to achieve a maximum level of performance by comparing them with the Y_{known} values. We use a supervised ML algorithm for our study because we have input variables X (which are $F_{\mathrm{P_{0.2}}}$, $F_{\mathrm{P_{0.7}}}$, $F_{\mathrm{P_7}}$, $F_{\mathrm{P_{20}}}$, $F_{\mathrm{P_{50}}}$, T_{eff}, and Kp) and an output variable Y_{known} (representing our surface gravity log g).

B.2. Regression trees

The regression tree method is part of the Classification and Regression Trees (CART) introduced by Breiman (2001). A decision tree algorithm constructs a binary tree during the training, with each node representing a split point on a single input variable (X) (a numerical value for regression algorithms, or a class name for classification algorithms). The leaf nodes of the tree contain the output possible predictions ($Y_{\mathrm{predicted}}$), as shown in Fig B.1.
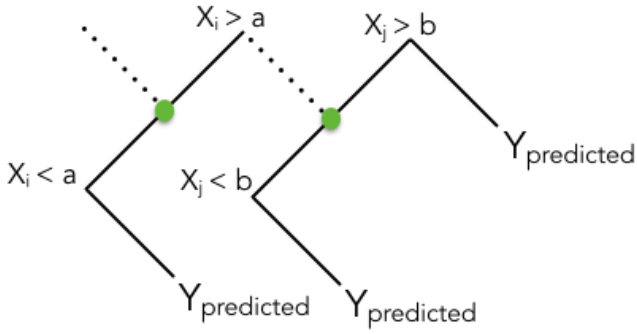
Fig. B.1. Schematic representation of a regression decision tree. At each node (green points) one variable is split at a value such that the cost function (see Eq (B.2)) is minimized.

The tree is built in such way that the cost function is minimized. Equation (B.2) is the cost function used for the regressor, with N_{training} being the number of stars in our training sample:

$$\text{cost} = \sum_{i=1}^{N_{\text{training}}} (Y_{\text{known}} - Y_{\text{predicted}})^2. \tag{B.2}$$

Once the tree is built on the training sample, it is used to evaluate $Y_{\text{predicted}}$ for new X_{new} data.

B.3. Ensemble method random forest regressor

An ensemble method combines the prediction from multiple ML algorithms together. It aims to make even more accurate predictions than any individual model. The random forest regressor is an ensemble method that combines regression trees. It consists in the following:

- Creating many subsamples of the training sample;
- Training a regression tree on each subsample, keeping a low number of variables that can be looked at for each split point. It aims to decrease the correlation between the different trees. For a regression algorithm, the typical number of features that can be searched is $m = \frac{p}{3}$ with p the number of input (X) variables;

- Calculating the average prediction from each model for the new test sample: this averaged value is taken as the estimate for the output variable ($Y_{\text{predicted}}$).

In our work we use the "RandomForestRegressor" function from the "sklearn.ensemble" Python library (Pedregosa et al. 2011) to compute the training on surface gravity.

Appendix C: Automatic estimation of ν_{max}

As a complementary study we also trained our algorithm to estimate the frequency of maximum power ν_{max}. The training is made following Sect. 4 by using ν_{max} instead of $\log g$ as the predicted parameter $Y_{\text{predicted}}$. The training input values are computed as in Sect. 4.2 by combining the ν_{max} values estimated by the A2Z global seismic pipeline for our sample of ~15 000 stars along with their uncertainties.

Results are very similar to the estimation of surface gravity, and are given in Table C.1. The estimation of ν_{max} can be made for any star with solar-like oscillations with $0.1 < \log g < 3.4$ dex, $3285 < T_{\text{eff}} < 7411$ K, $Kp < 14$, and provide a very good prior for any more complex seismic analysis of the star. The complete algorithm for the ν_{max} estimation can be found on the Git repository[1].

Table C.1. Summary of statistical results on ν_{max} on the test set from Fig. 3.

$\log(\nu_{\text{max}})$ (dex)	$\overline{\log(\nu_{\text{max A2Z}}) - \log(\nu_{\text{max RF}})}$ (dex)	MAD (dex)
ALL	-0.3×10^{-3}	0.044
$[-1: -0.5]$	-0.123	0.085
$[-0.5:0]$	-0.011	0.081
$[0:0.5]$	-0.0007	0.075
$[0.5:1]$	-0.004	0.053
$[1:1.5]$	-0.019	0.041
$[1.5:2]$	0.003	0.037
$[2:2.5]$	0.025	0.046
$[2.5:3]$	-0.041	0.061
$[3:3.5]$	0.011	0.082

Notes. MAD is the mean absolute deviation.