MNRAS 000, 1–9 (2018) Preprint 17 November 2018 Compiled using MNRAS LATEX style file v3.0

Gaussian Process modelling of granulation and oscillations in red-giant stars

Luís F. Pereira[1,2]★, Tiago L. Campante[1,2], Margarida S. Cunha[1,2], João P. Faria[1], Nuno C. Santos[1,2], Susana C. C. Barros[1,2], Olivier Demangeon[1], and James S. Kuszlewicz[3]
[1]*Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, PT4150-762 Porto, Portugal*
[2]*Departamento de Física e Astronomia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, PT4169-007 Porto, Portugal*
[3]*Max Planck Institute for Solar System Research, D-37077 Göttingen, Germany*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The analysis of photometric time series in the context of transiting planet surveys suffers from the presence of stellar signals, often dubbed "stellar noise". These signals, caused by stellar oscillations and granulation, can usually be disregarded for main-sequence stars, as the stellar contributions average out when phase-folding the light curve. For evolved stars, however, the amplitudes of such signals are larger and the timescales similar to the transit duration of short-period planets, requiring that they be modelled alongside the transit. With the promise of TESS delivering on the order of $\sim 10^5$ light curves for stars along the red-giant branch, there is a need for a method capable of describing the "stellar noise" while simultaneously searching for an exoplanet transit. In this work, a Gaussian Process regression framework is used to model stellar light curves and the method validated by applying it to TESS-like artificial data. Furthermore, the method is used to characterize the stellar oscillations and granulation of a sample of well-studied *Kepler* low-luminosity red-giant branch stars. The parameters determined are compared to equivalent ones obtained by modelling the power spectrum of the light curve. Results show that the method presented is capable of describing the stellar signals in the time domain and can also return an accurate and precise measurement of $\nu_{\rm max}$, i.e., the frequency of maximum oscillation amplitude.

Key words: asteroseismology – methods: data analysis – planets and satellites: detection – stars: oscillations – techniques: photometric

1 INTRODUCTION

Red-giant stars show significant stellar signals on timescales ranging from hours to weeks (e.g., Mathur et al. 2011; Kallinger et al. 2014, 2016; North et al. 2017). In particular, granulation is responsible for introducing correlated noise in the time series photometry of red giants, with a characteristic timescale similar to the transit duration of a planet in a close orbit. Such "stellar noise" should therefore be taken into account in the modelling of planetary transits if biases on the fitted transit parameters are to be avoided (Carter & Winn 2009; Barclay et al. 2015). Moreover, solar-like oscillations characterized by large mode amplitudes are also present in the time series photometry of red giants and ultimately require similar treatment (Grunblatt et al. 2016, 2017).

With NASA's TESS mission (Ricker et al. 2015; Huang et al. 2018; Vanderspek et al. 2018) expected to deliver on the order of $\sim 10^5$ light curves (from full-frame images) for stars along the red-giant branch (Sullivan et al. 2015; Campante et al. 2016), a need becomes apparent for a method capable of describing such stellar signals while simultaneously searching for an exoplanet transit. Since stellar granulation is a stochastic phenomenon, there is no functional form capable of describing it in the time domain. Nonetheless, there are tools that, while not leading to a parametric model, are capable of defining a model based on a set of properties of the data. One such method consists in the use of Gaussian Process (GP) regression (Rasmussen & Williams 2006). This method has been increasingly adopted in a variety of scenarios, including previous modelling of light curves (e.g., Barclay et al. 2015; Grunblatt et al. 2016, 2017, and references therein) and radial-velocity time series (Brewer & Stello 2009; Faria et al. 2016).

When employing this framework to the modelling of stellar signals in photometric time series it is common, however, to pay little attention to the physical meaning of the resulting model parameters. In the aforementioned works, the main focus has been the convergence of the adopted model whilst the values obtained are rarely analyzed to assess their physical validity with respect to the type of star being observed. A main goal of this work is therefore to evaluate whether it is possible to recover physically meaningful

★ Email: Filipe.Pereira@astro.up.pt

parameters when using GP regression to model stellar signals in the time domain.

The paper starts with an introduction to the theoretical framework of GP regression in Sect. 2. In order to assess the validity and accuracy of the method, the GP regression is first applied to artificial time series similar to those expected from the TESS mission (Sect. 3). The parameters obtained in such fits are then compared to the true values injected in the simulated light curves. Subsequently, the method is tested with real *Kepler* data for low-luminosity red-giant branch (LLRGB) stars (Sect. 4). Here, the aim is to compare the parameters obtained using GP regression to the equivalent parameters obtained when performing a standard fit to the power spectrum of the time series. Conclusions are drawn in Sect. 5.

2 GAUSSIAN PROCESS REGRESSION

2.1 Theoretical framework

Gaussian Processes are non-parametric models capable of describing correlated stochastic signals. They describe each point as a correlated random variable with a mean value and a variance, where any finite collection of those variables has a multivariate Gaussian distribution. The measure of similarity, or correlation, between pairs of points in the signal with respect to the distance between them in time is given by the covariance function or kernel of the GP. The covariance matrix, $\mathbf{K}$, is defined as:

$$K_{ij} = \sigma_i^2 \delta_{ij} + k_{\boldsymbol{\alpha}}(\tau_{ij}), \tag{1}$$

where for each pair of points i and j, σ_i is the uncertainty of the observation, δ_{ij} is the Kronecker delta, and $k_{\boldsymbol{\alpha}}(\tau_{ij})$ is the kernel, with $\tau_{ij} = |t_i - t_j|$ being the absolute distance in time between points i and j. As indicated by the notation, the kernel depends on a set of parameters, $\boldsymbol{\alpha}$.

GP regression consists in the selection of a kernel that describes the correlation between data points and then finding the set of parameters that best represent the observed data. The merit of the fit can be determined using the log-likelihood function

$$\ln\mathcal{L}(\mathbf{r}) = -\frac{1}{2}\mathbf{r}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{r} - \frac{1}{2}\ln|\mathbf{K}| - \frac{n}{2}\ln(2\pi), \tag{2}$$

where $\mathbf{r}$ are the residuals after removing the mean model from the data (this mean model may represent any parametric model used to characterize the data, e.g., a transit model) and n is the number of data points.

When using GP models, the most important decision to be made concerns the selection of an appropriate kernel. A number of aspects have to be taken into account when performing such a selection. Standard computation of Eq. (2) has complexity of order $O(n^3)$, meaning that computation time increases very rapidly with the number of observations. Efforts have been made to reduce this complexity (Ambikasaran et al. 2015; Foreman-Mackey et al. 2017) and hence computational time. However, these efforts inevitably carry with them a number of shortcomings, especially with respect to the functional form of the kernels permitted.

2.2 Algorithm

In this work, the implementation of GP regression chosen was the Python package celerite[1] (Foreman-Mackey et al. 2017). This package applies some restrictions to the functional form of the kernels in order to achieve a computational complexity of order $O(n)$, an acceptable level if one desires to perform fast computations on long time series, like the ones provided by TESS.

The celerite implementation requires that the kernel $k_{\boldsymbol{\alpha}}$ be a mixture of exponential functions:

$$k_{\boldsymbol{\alpha}}(\tau_{ij}) = \sum_{m=1}^{M} a_m \exp(-c_m\, \tau_{ij}), \tag{3}$$

where $\boldsymbol{\alpha} = \{a_m,\ c_m\}$. By introducing complex parameters, $a_m \rightarrow a_m \pm i\ b_m$ and $c_m \rightarrow c_m \pm i\ d_m$, and rewriting the exponentials in Eq. (3) as sums of sines and cosines, the equation can be written as a mixture of quasi-periodic oscillators:

$$\begin{aligned} k_{\boldsymbol{\alpha}}(\tau_{ij}) = \sum_{m=1}^{M} [&a_m \exp(-c_m\, \tau_{ij}) \cos(d_m\, \tau_{ij}) \\ &+ b_m \exp(-c_m\, \tau_{ij}) \sin(d_m\, \tau_{ij})\ ], \end{aligned} \tag{4}$$

this time with $\boldsymbol{\alpha} = \{a_m,\ b_m,\ c_m,\ d_m\}$.

To provide physical insight into the previous equation, consider the power spectral density (PSD) of a stochastically-driven, damped harmonic oscillator:

$$S(\omega) = \sqrt{\frac{2}{\pi}} \frac{S_0 \omega_0^4}{(\omega^2 - \omega_0^2)^2 + \omega_0^2\omega^2/Q^2}, \tag{5}$$

where ω is an angular frequency, ω_0 is the frequency of the undamped oscillator, Q is the oscillator's quality factor, and S_0 is proportional to the power of the spectrum at $\omega = \omega_0$, i.e.,

$$S(\omega_0) = \sqrt{\frac{2}{\pi}}\, S_0\, Q^2. \tag{6}$$

Following Foreman-Mackey et al. (2017), the PSD in Eq. (5) can be used to describe the power spectrum of Eq. (4), and a relation can thus be written between $\boldsymbol{\alpha}$ and the parameters S_0, Q and ω_0. Doing so, the kernel in Eq. (4) translates to

$$k(\tau;\ S_0,\ Q,\ \omega_0) = S_0\, \omega_0\, Q\, \mathrm{e}^{\frac{\omega_0\tau}{2Q}} \times \begin{cases} \cosh(\eta\omega_0\tau) + \frac{1}{2\eta Q} \sinh(\eta\ \omega_0\ \tau), & 0 < Q < 1/2, \\ 2(1 + \omega_0\tau), & Q = 1/2, \\ \cos(\eta\omega_0\tau) + \frac{1}{2\eta Q} \sin(\eta\ \omega_0\ \tau), & 1/2 < Q, \end{cases} \tag{7}$$

where $\eta = |1 - (4Q^2)^{-1}|^{1/2}$.

The work of Foreman-Mackey et al. (2017) remarks upon some limits of physical interest regarding the power spectrum in Eq. (5) which translate also to the kernel in Eq. (7). When considering the limit $Q = 1/\sqrt{2}$, the kernel simplifies to

$$k(\tau) = S_0\, \omega_0\, \mathrm{e}^{-\frac{1}{\sqrt{2}}\omega_0\tau} \cos\left(\frac{\omega_0\tau}{\sqrt{2}} - \frac{\pi}{4}\right), \tag{8}$$

with the corresponding PSD becoming

$$S(\omega) = \sqrt{\frac{2}{\pi}} \frac{S_0}{(\omega/\omega_0)^4 + 1}. \tag{9}$$

Equation (9) has the same functional form as the equation commonly used to model the granulation in a power spectrum analysis (e.g., Kallinger et al. 2014). Therefore, the kernel in Eq. (8) can be used to capture the granulation signal in the time domain. The other interesting limit of Eq. (7) is that, for values of $Q > 1$, the PSD of the kernel can be used as an approximation to the Gaussian-like shape of the oscillation bump (or power excess) found in the power spectrum.

[1] https://github.com/dfm/celerite

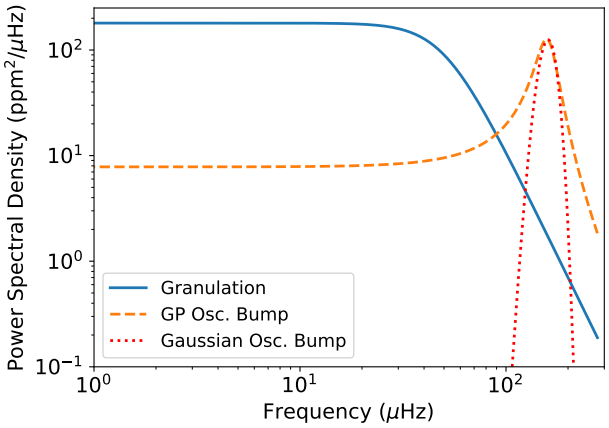

Figure 1. Solid (blue) curve depicts the power spectral density of a granulation profile (Eq. 9). Dashed (orange) and dotted (red) curves are the power spectral densities of the functions used to capture the signal from the oscillation bump in a GP and power spectrum analyses, respectively.

2.3 Implementation

The two models adopted in this work are a sum of the kernels described above. The first model, Model 1, contains a single granulation kernel (Eq. 8), an oscillation bump kernel (Eq. 7; $2 < Q < 15$, where the limits were defined empirically for this work), as well as a shot noise kernel (which is simply a constant value added in quadrature to the diagonal of the covariance matrix). The second model, Model 2, results from adding a second granulation kernel to Model 1. This stems from the fact that, as Kallinger et al. (2014) observed, at least two granulation components should be adopted when estimating the stellar background signal for red-giant stars, owing to granulation being relevant at different timescales. Corsaro et al. (2015) added a third component to model long-term variations (on the order of 1 cycle per 11 days). Since typical TESS time series are considered in this work (i.e., with a duration of 27.4 days), such long-term variations would not be detectable and hence a third component is discarded.

Study of stellar signals in the literature is commonly performed in the frequency domain by means of power-spectrum fitting. To be able to compare the parameters S_0, Q and ω_0 present in the kernels with the characteristic quantities found for stellar signals in power-spectrum fitting, a correspondence needs to be established between the functions used to capture the signals in both domains.

Concerning granulation, Eq. (9) can be related to the one presented in (Kallinger et al. 2014):

$$S(\nu) = \frac{2\sqrt{2}}{\pi} \frac{a_{\rm gran}^2 / b_{\rm gran}}{(\nu/b_{\rm gran})^4 + 1}, \tag{10}$$

where $a_{\rm gran}$ and $b_{\rm gran}$ are the characteristic amplitude and frequency of the granulation signal, respectively. Note that the above PSD varies in frequency, ν. For the oscillation bump, which is commonly modelled as a Gaussian in a power spectrum analysis, Eq. (5) can be related to (Kallinger et al. 2014)

$$S(\nu) = P_g \exp\left(\frac{-(\nu - \nu_{\rm max})^2}{2\sigma^2}\right), \tag{11}$$

where P_g is the height of the oscillation bump, $\nu_{\rm max}$ is the frequency of maximum oscillation amplitude, and σ is the width of the bump. Again, note that this PSD varies in frequency.

Parameter	Lower Bound	Upper Bound
$a_{\rm gran,1}$ (ppm)	10	400
$b_{\rm gran,1}$ (μHz)	10	200
$P_{\rm g}$ (ppm)	10	1500
Q	2	15
$\nu_{\rm max}$ (μHz)	100	220
Shot Noise (ppm)	0	300

Table 1. Lower and upper bounds chosen for the uniform distributions used as priors for parameters in Model 1.

An illustrative example of the power spectra of both kernels described above is shown in Fig. 1. It shows the shape of the granulation PSD (Eq. 9), common to both the GP and power spectrum analyses, while highlighting the differences in the functions used to capture the signal from the oscillation bump, namely, the low-frequency tail and the slow decrease in power at high frequencies in the GP model.

In order to make a meaningful comparison between the parameters in Eqs. (9) and (5) and those in Eqs. (10) and (11), respectively, all equations need to be consistently normalized. Appendix A details the steps taken to determine the exact correspondence between the parameters in both approaches. After applying a constant normalization factor to both Eqs. (10) and (11), the parameters returned by the two methods can be compared:

$$\begin{aligned} a_{\rm gran} &= \sqrt{\sqrt{2}\, S_{0,\rm gran}\, \omega_{0,\rm gran}}\,, \\ b_{\rm gran} &= \frac{\omega_{0,\rm gran}}{2\pi}\,, \\ P_g &= 4\, S_{0,\rm bump}\, Q_{\rm bump}^2\,, \\ \nu_{\rm max} &= \frac{\omega_{0,\rm bump}}{2\pi}\,. \end{aligned} \tag{12}$$

The subscripts "gran" and "bump" emphasize that, while the parameter names S_0, Q and ω_0 are the same for both kernels, these are separate kernels in the model, and thus have different values.

2.4 Optimization

To not only determine the most appropriate value for each parameter in a model but also study its uncertainties, the EMCEE[2] PYTHON package is used. This package provides Bayesian parameter estimation given a prior for each of the parameters in the model.

In this work, only uniform priors have been considered. Tables 1 and 2 present the lower and upper bounds chosen for the prior distributions of all parameters in Models 1 and 2, respectively.

3 APPLICATION TO TESS-LIKE ARTIFICIAL DATA

The method is first tested with TESS-like artificial time series for a set of 20 LLRGB stars (with effective temperature $4800 < T_{\rm eff} < 5500$ K, frequency of maximum oscillation amplitude $105 < \nu_{\rm max} < 185$ μHz, and apparent magnitude $V < 11$). Generation of the artificial light curves is performed originally in

[2] https://github.com/dfm/emcee

Parameter	Lower Bound	Upper Bound
$a_{\mathrm{gran},1}$ (ppm)	10	400
$b_{\mathrm{gran},1}$ (μHz)	10	60
$a_{\mathrm{gran},2}$ (ppm)	10	400
$b_{\mathrm{gran},2}$ (μHz)	70	200
P_{g} (ppm)	10	1500
Q	2	15
ν_{max} (μHz)	100	220
Shot Noise (ppm)	0	300

Table 2. Lower and upper bounds chosen for the uniform distributions used as priors for parameters in Model 2.

the frequency domain by using scaling relations, after which an inverse Fourier transform is applied and the 30-min cadence of TESS full-frame images considered. A photometric noise model for TESS (Sullivan et al. 2015; Campante et al. 2016) is used in order to predict the rms noise for a given exposure time. A systematic noise term of 20 ppm $\mathrm{hr}^{1/2}$ was included in this calculation. To model the granulation power spectral density, a scaled version (to predict TESS granulation amplitudes) of model F of Kallinger et al. (2014) was adopted, which contains two granulation (or Harvey-like) components. No aliased granulation power was considered. Individual radial, (mixed) dipole and quadrupole modes were also modelled (Kuszlewicz et al., submitted).

For each of the 20 simulated stars, 10 independent 27.4-day time series were generated. The values and corresponding uncertainties computed for each of the model parameters take into account the analysis of these 10 independent time series so as to reduce the possibility of systematic errors. Specifically, each parameter was estimated as the median of the values obtained for the 10 independent time series. Its uncertainty was defined as the sum in quadrature of the uncertainty associated with the median value and the standard deviation of the values determined for the 10 independent time series.

The method is applied twice to all data sets, making use of each of the models, i.e., Models 1 and 2. The parameters derived from the GP regression may then be compared to the input parameters used when generating the light curves. Figures 2 and 3 show such a comparison for the parameters of Models 1 and 2, respectively. The parameters represented in Fig. 2 for Model 1 are the amplitude and characteristic frequency of the single granulation component in the model, $a_{\mathrm{gran},1}$ and $b_{\mathrm{gran},1}$, respectively, the frequency of maximum oscillation, ν_{max}, and the shot noise. Figure 3 depicts the same parameters with the addition of the amplitude and characteristic frequency of the second granulation component, $a_{\mathrm{gran},2}$ and $b_{\mathrm{gran},2}$, respectively. Both the oscillator's quality factor, Q, and the power at $\nu = \nu_{\mathrm{max}}$, P_{g}, are not represented as the artificial light curves result from simulating individual oscillation modes and hence there are no *bona fide* input values to compare with.

As an example of how well GP regression allows for the characterization of the stellar signal in the time domain, Fig. 4 shows a blowup of the fit performed to one of the artificial time series. Figure 5 shows the PSD of that same GP regression output compared to the PSD of the light curve. Both figures show the results obtained when fitting Model 2 to the data.

In what follows, a given parameter is considered to have been accurately determined if the null offset (red dashed line) is within the 1σ interval (black dashed lines) associated with the median of the data points (or bias; black solid line). Looking at the results obtained when employing Model 1 (Fig. 2), the amplitude, $a_{\mathrm{gran},1}$, and characteristic frequency, $b_{\mathrm{gran},1}$, of the single granulation component in the model are not correctly retrieved, showing a bias of about 17% relative to the input values in the lower-frequency granulation component in the data. This is not too surprising, since the model being considered is incomplete: both granulation parameters are being overestimated in an attempt to capture the power in the two granulation components present in the data. It should be noted that the low-frequency tail of the oscillation bump profile (see Sect. 2.3 and Fig. 1) contributes to somewhat attenuating the offset found for these two parameters. Nevertheless, the estimation of ν_{max} is robust, with this parameter being accurately (0.76% bias) and precisely (3.08% scatter) recovered. Finally, the shot noise level is recovered to within 8% of the input value. The slight, overall underestimation of the shot noise level is to be expected because of the non-negligible contribution of the oscillation bump profile at high frequencies (see Fig. 1). Concerning the outlying artificial star, it has the highest value of ν_{max} amongst the stars in the sample. Upon inspection of its PSD, it becomes clear that the proximity of the oscillations to the Nyquist frequency ($\nu_{\mathrm{Nyq}} \approx 283$ μHz for the 30-min cadence of the simulated light curves) prevents the shot noise level from being robustly determined.

Concerning Model 2 (Fig. 3), the introduction of a second granulation component in the model leads to an overall improvement in the fit to the granulation signal. The first (lower-frequency) granulation component is now accurately recovered, with the amplitude and characteristic frequency within 4.47% (6.30% scatter) and 1.49% (7.87% scatter) of the input values, respectively. The amplitude of the second (lower-power) granulation component is also accurately recovered to within 9.66% (12.39% scatter). The same cannot be said of its characteristic frequency, whose bias amounts to about 16% (9.70% scatter), hinting at a correlation with $\log g$. At lower values of surface gravity for the artificial stars where both granulation components and the oscillations are closer together (in frequency), the offset is higher between the real value of the characteristic frequency of the second granulation component and its value determined by the method. As the $\log g$ increases the offset becomes lower reaching values close to zero at the stars with higher surface gravity in the sample. Finally, the estimation of ν_{max} continues to be robust (1.13% bias and 3.30% scatter) and the results for the shot noise level are similar to those obtained when using Model 1.

4 APPLICATION TO *KEPLER* LLRGB STARS

As a second test, the method is applied to the same sample of 19 *Kepler* LLRGB stars considered in Corsaro et al. (2015). In order to mimic the typical amount of data expected from TESS and to account for systematics that might be present in the light curve, 10 non-overlapping subsets of 27.4 days of observations were considered from the full *Kepler* light curve for each star.

The values and corresponding uncertainties computed for each of the model parameters again take into account the analysis of these 10 independent subsets, as described in the previous section. The only exception concerns the shot noise component, whose values presented here are upper bounds containing 99% of the posterior probability. This choice was motivated by the shape of the posterior probability distributions obtained for this model parameter, which typically exhibited uniformly distributed probabilities up to a certain value, above which they would abruptly fall off.

Besides performing a GP regression in the time domain, a standard fit to the power spectrum was also conducted making use

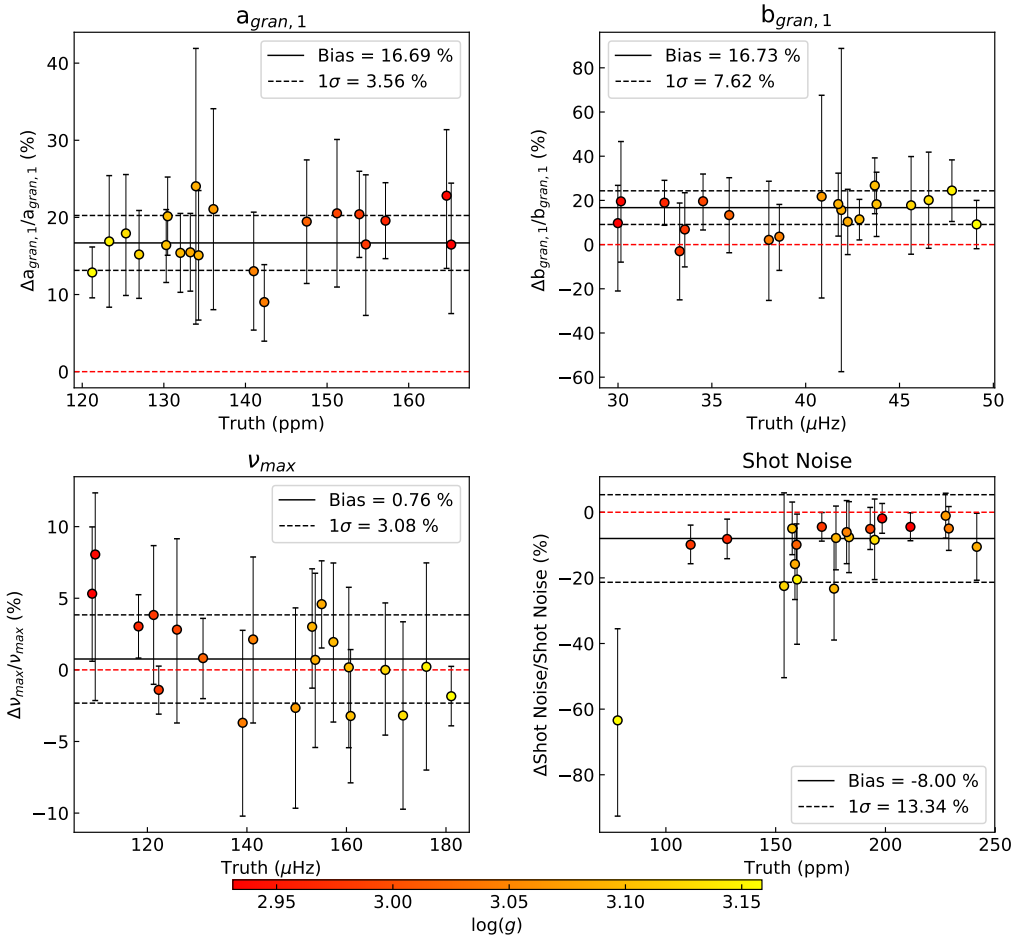

Figure 2. Comparison of the parameters in the fit of Model 1 to the TESS-like artificial data with the input used to generate those data. Data points represent the relative deviation with respect to the input value, with error bars corresponding to the uncertainties returned by the GP regression method. Black solid and dashed lines represent the median and standard deviation of the data points, respectively, with their numerical values shown in the inset. The red dashed line denotes a null offset. Data points are colour-coded according to a star's surface gravity, $\log g$.

of a model equivalent to the GP kernel (see Appendix A). The latter analysis was performed using the DIAMONDS code[3] (Corsaro & De Ridder 2014), which fits the power spectrum and determines the model parameters within a Bayesian framework.

A comparison of the output obtained with both approaches is shown in Figs. 6 and 7 for Models 1 and 2, respectively. The parameters depicted are the same as in Figs. 2 and 3. A blowup of the output of the GP regression when applied to one of the *Kepler* LLRGB stars in the sample is shown in Fig. 8. Figure 9 shows the PSD of that same GP regression output compared to the PSD of the light curve. Both figures show the results obtained when fitting Model 2 to the data.

Contrary to the previous section, where the method's accuracy was assessed with artificial time series, the method is now tested on real *Kepler* data for a sample of well-studied LLRGB stars. By doing this, the parameters derived through GP regression can be compared to the equivalent parameters obtained when performing a standard fit to the power spectrum. This test thus allows for a comparison with the more traditional methodology used in studies of stellar light curves.

[3] https://fys.kuleuven.be/ster/Software/Diamonds

Looking at the results obtained when considering Model 1 (Fig. 6), the parameters describing the single granulation component in the model, $a_{\rm gran,1}$ and $b_{\rm gran,1}$, are underestimated, with relative biases of about -5% and -9%, respectively. This underestimation is likely a consequence of the presence of the low-frequency tail of the oscillation bump profile (see Fig. 1). Concerning $\nu_{\rm max}$, a small relative bias of 2.05% (1.08% scatter) is found between methods. Finally, results for the shot noise level show a relative bias of about -31% (15.57% scatter) between the two methods. It should be borne in mind that shot noise levels for the *Kepler* stars in the sample are relatively low (*Kepler*'s effective collecting area is larger than that of the individual TESS cameras by a factor of $\sim 10^2$), which, coupled with differences in the models adopted in either method (see Fig. 1), results in large relative differences. Inspection of the absolute value of this same bias reveals differences no greater than 12 ppm.

With respect to Model 2 (Fig. 7), excellent agreement is seen between the parameters describing the higher-power granulation component, i.e., $a_{\rm gran,1}$ and $b_{\rm gran,1}$, with relative differences of -0.18% and -3.56%, respectively. The low-frequency tail of the oscillation bump profile does not appear to be affecting this component. It does, however, impact the parameters describing the second (lower-power) granulation component, $a_{\rm gran,2}$ and $b_{\rm gran,2}$, with the GP method showing systematically lower values, with relative bi-

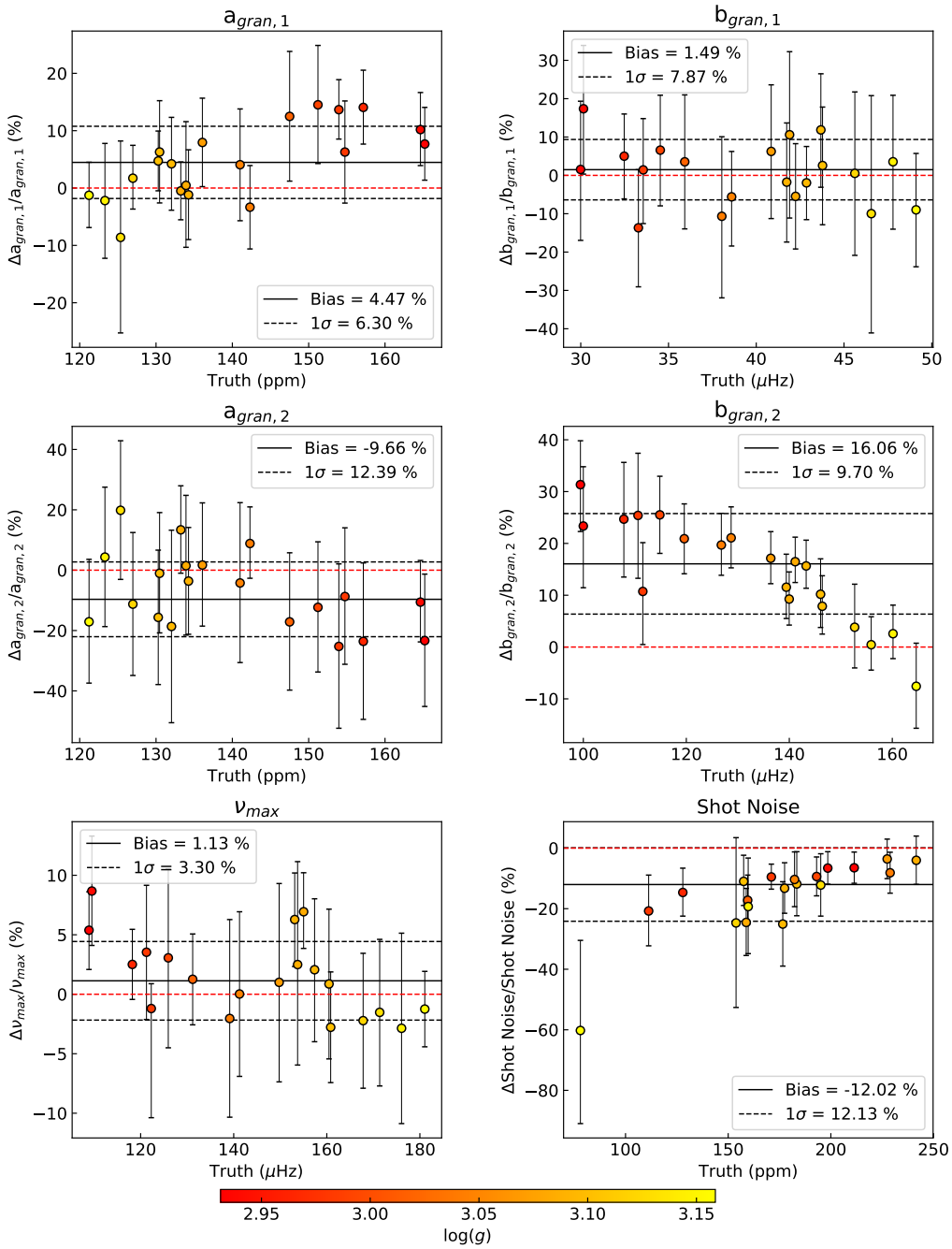

Figure 3. Comparison of the parameters in the fit of Model 2 to the TESS-like artificial data with the input used to generate those data. Data points represent the relative deviation with respect to the input value, with error bars corresponding to the uncertainties returned by the GP regression method. Black solid and dashed lines represent the median and standard deviation of the data points, respectively, with their numerical values shown in the inset. The red dashed line denotes a null offset. Data points are colour-coded according to a star's surface gravity, $\log g$.

ases of respectively −6.08% and −16.21%, the latter being significant. Regarding $\nu_{\rm max}$, just like with Model 1, a small relative bias of 0.81% (0.64% scatter) is found between methods. For the shot noise level, results show a less pronounced discrepancy compared to the one seen for Model 1, with the relative bias between methods now being of −13.86% (16.89% scatter) and the absolute difference never exceeding 6 ppm.

5 CONCLUSIONS

Gaussian Processes were employed to model the stellar signals (i.e., granulation and oscillations) of low-luminosity red-giant stars in the time domain. Two models were considered: Model 1 contains a granulation component, an oscillation bump component and a shot noise component; Model 2 results from adding an extra granulation component to Model 1. Both models were applied to TESS simulated data (generated considering the presence of two granulation

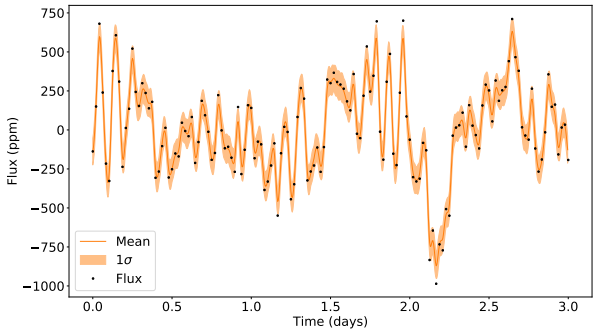

Figure 4. Predictive model (Model 2) output by the GP regression (mean and 1σ interval) when applied to one of the artificial TESS-like time series. The plot is zoomed in on the first ~ 3 days of simulated data to improve visualization.

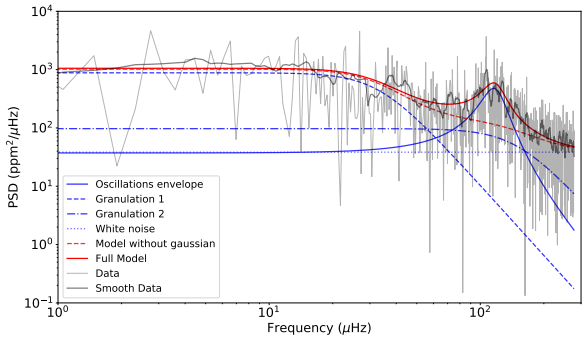

Figure 5. Power spectral density of the same (full) light curve depicted in Fig. 4. The PSD of the light curve is shown in light grey, with a slightly smoothed version overlapped in dark grey. The PSD of the GP regression output (Model 2) is shown as a solid red curve, with individual components shown in different line styles and colors (see legend).

components) to test the validity of the method and its applicability to TESS light curves. Furthermore, the models were also applied to a sample of *Kepler* stars in order to compare this method with the commonly used power-spectrum fitting method.

Following the analysis and discussion of the results presented in Sections 3 and 4, some conclusions can be drawn:

- Application to TESS simulated data showed that GP regression is capable of capturing the stellar background signal and oscillations directly in the time domain. In particular, an accurate determination of $\nu_{\max}$ was made possible, with a small bias of $\approx$1%.
- Comparison with the power-spectrum fitting method using *Kepler* data showed that both methods find the same stellar signals, with any disagreement in specific parameters being attributed to differences in the models adopted to capture the signal of the oscillation bump.
- Provided a physically motivated model is chosen, Gaussian Processes can be used to model stellar signals in the time domain, hence becoming a valid alternative to the commonly used power-spectrum fitting method.

Considering these conclusions, the method described in this work presents a chance to improve on previous methodologies (Barclay et al. 2015; Grunblatt et al. 2017, e.g.)) of modelling the light curves of planetary transits together with the host star's signals by describing a more complete model of the stellar components and having physically motivated parameters.

Moreover, the method also ties in with one of the "Important scientific opportunities for Kepler & K2 Data" (Barentsen et al. 2018), specifically, on performing asteroseismology in the time domain.

ACKNOWLEDGEMENTS

The authors would like to thank Enrico Corsaro for assistance with the implementation of DIAMONDS. F.P. acknowledges support from fellowship PD/BD/135227/2017 funded by FCT (Portugal) and POPH/FSE (EC). The project leading to this publication has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 792848. T.L.C. acknowledges support from grant CIAAUP-12/2018-BPD. S.C.C.B. acknowledges support from FCT through Investigador FCT contract IF/01312/2014/CP1215/CT0004. This work was supported by FCT - Fundação para a Ciência e a Tecnologia through national funds and by FEDER through COMPETE2020 - Programa Operacional Competitividade e Internacionalização by these grants: UID/FIS/04434/2013 & POCI-01-0145-FEDER-007672; PTDC/FIS-AST/30389/2017 & POCI-01-0145-FEDER-030389; PTDC/FIS-AST/32113/2017 & POCI-01-0145-FEDER-032113 and PTDC/FIS-AST/28953/2017 & POCI-01-0145-FEDER-028953.

REFERENCES

Ambikasaran S., Foreman-Mackey D., Greengard L., Hogg D. W., O'Neil M., 2015, IEEE Transactions on Pattern Analysis and Machine Intelligence, 38
Barclay T., Endl M., Huber D., Foreman-Mackey D., Cochran W. D., MacQueen P. J., Rowe J. F., Quintana E. V., 2015, ApJ, 800, 46
Barentsen G., Hedges C., Saunders N., Cody A. M., Gully-Santiago M., Bryson S., Dotson J. L., 2018, preprint, p. arXiv:1810.12554 (arXiv:1810.12554)
Brewer B. J., Stello D., 2009, MNRAS, 395, 2226
Campante T. L., et al., 2016, ApJ, 830, 138
Carter J. A., Winn J. N., 2009, ApJ, 704, 51
Corsaro E., De Ridder J., 2014, A&A, 571, A71
Corsaro E., De Ridder J., García R. A., 2015, A&A, 579, A83
Faria J. P., Haywood R. D., Brewer B. J., Figueira P., Oshagh M., Santerne A., Santos N. C., 2016, A&A, 588, A31
Foreman-Mackey D., Agol E., Ambikasaran S., Angus R., 2017, AJ, 154, 220
Grunblatt S. K., et al., 2016, AJ, 152, 185
Grunblatt S. K., et al., 2017, AJ, 154, 254
Huang C. X., et al., 2018, preprint, (arXiv:1809.05967)
Kallinger T., et al., 2014, A&A, 570, A41
Kallinger T., Hekker S., Garcia R. A., Huber D., Matthews J. M., 2016, Science Advances, 2, 1500654
Mathur S., et al., 2011, ApJ, 741, 119
North T. S. H., et al., 2017, MNRAS, 465, 1308
Rasmussen C., Williams C., 2006, Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, USA
Ricker G. R., et al., 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003
Sullivan P. W., et al., 2015, ApJ, 809, 77

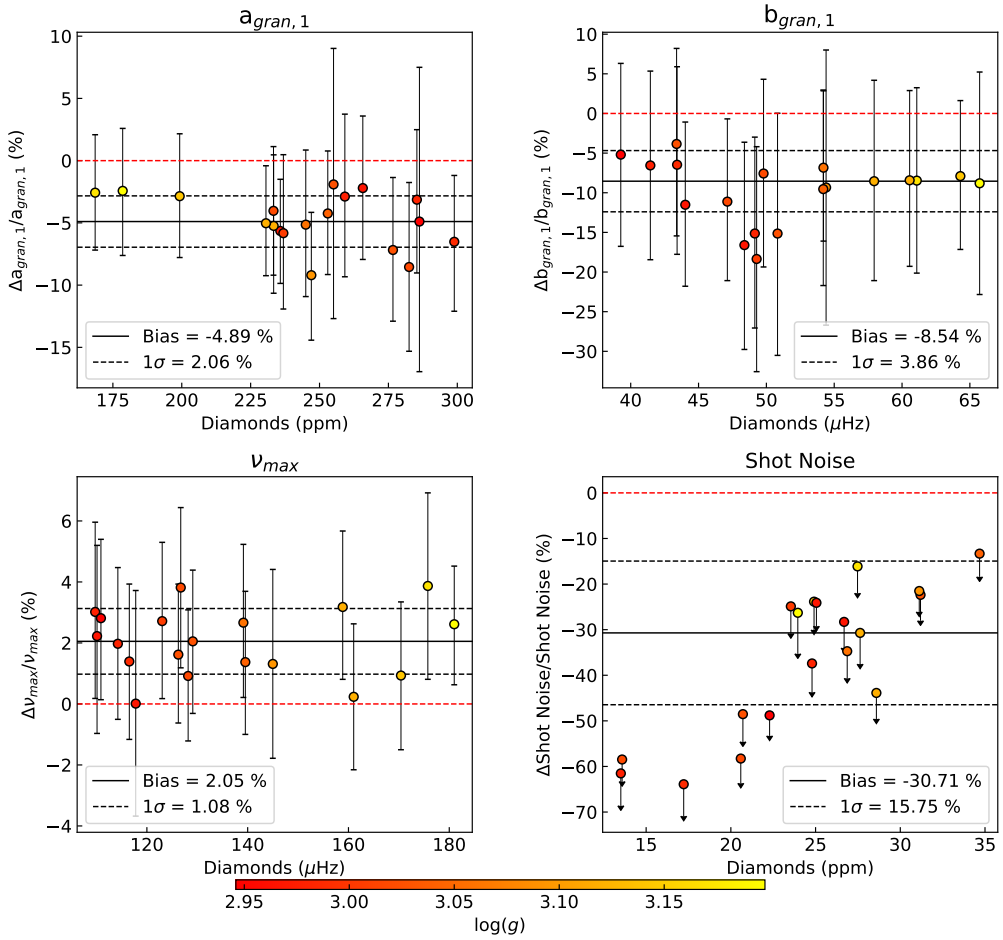

Figure 6. Comparison of the parameters in the fit of Model 1 to the *Kepler* time series data both by means of a GP regression and power-spectrum fitting. Data points represent the relative deviation with respect to the value determined using the PSD-fitting procedure, with error bars corresponding to the sum in quadrature of the uncertainties of both methods. Relative deviations of the shot noise level are computed based on an upper bound returned by the GP regression (therefore explaining the downward arrow). Black solid and dashed lines represent the median and standard deviation of the data points, respectively, with their numerical values shown in the inset. The red dashed line denotes a null offset. Data points are colour-coded according to a star's surface gravity, $\log g$.

Vanderspek R., et al., 2018, preprint, (arXiv:1809.07242)

APPENDIX A: PARSEVAL NORMALIZATION OF THE CELERITE PSD

The PSD in Eq. (9),

$$S(\omega) = \sqrt{\frac{2}{\pi}}\frac{S_0}{(\omega/\omega_0)^4 + 1}, \tag{A1}$$

shares the functional form of the PSD describing the granulation in Kallinger et al. (2014),

$$S(\nu) = \frac{2\sqrt{2}}{\pi}\frac{a^2/b}{(\nu/b)^4 + 1}. \tag{A2}$$

However, unlike Eq. A2, Eq. A1 is not normalized according to Parseval's theorem.

For a Parseval-normalized PSD, the variance of the light curve must equal a^2. In order to normalize Eq. A1, a constant K needs to be found that ensures that the previous condition is met. From Foreman-Mackey et al. (2017), the variance of a light curve described by the kernel in Eq. (8) is

$$k(\tau = 0) = \frac{S_0\omega_0}{\sqrt{2}}, \tag{A3}$$

which gives

$$a^2 = \frac{S_0\omega_0}{\sqrt{2}}. \tag{A4}$$

Moreover, from Foreman-Mackey et al. (2017), ω_0 is expressed as

$$\omega_0 = 2\pi b\,. \tag{A5}$$

Equating Eqs. A1 and A2,

$$K\sqrt{\frac{2}{\pi}}\frac{S_0}{(\omega/\omega_0)^4 + 1} = \frac{2\sqrt{2}}{\pi}\frac{a^2/b}{(\nu/b)^4 + 1}, \tag{A6}$$

and substituting for Eqs. A4 and A5, K becomes

$$K = 2\sqrt{2\pi}. \tag{A7}$$

Having obtained a value for K, Eq. A1 can be normalized according to Parseval's theorem as

$$S(\omega) = \frac{4S_0}{(\omega/\omega_0)^4 + 1}\,. \tag{A8}$$

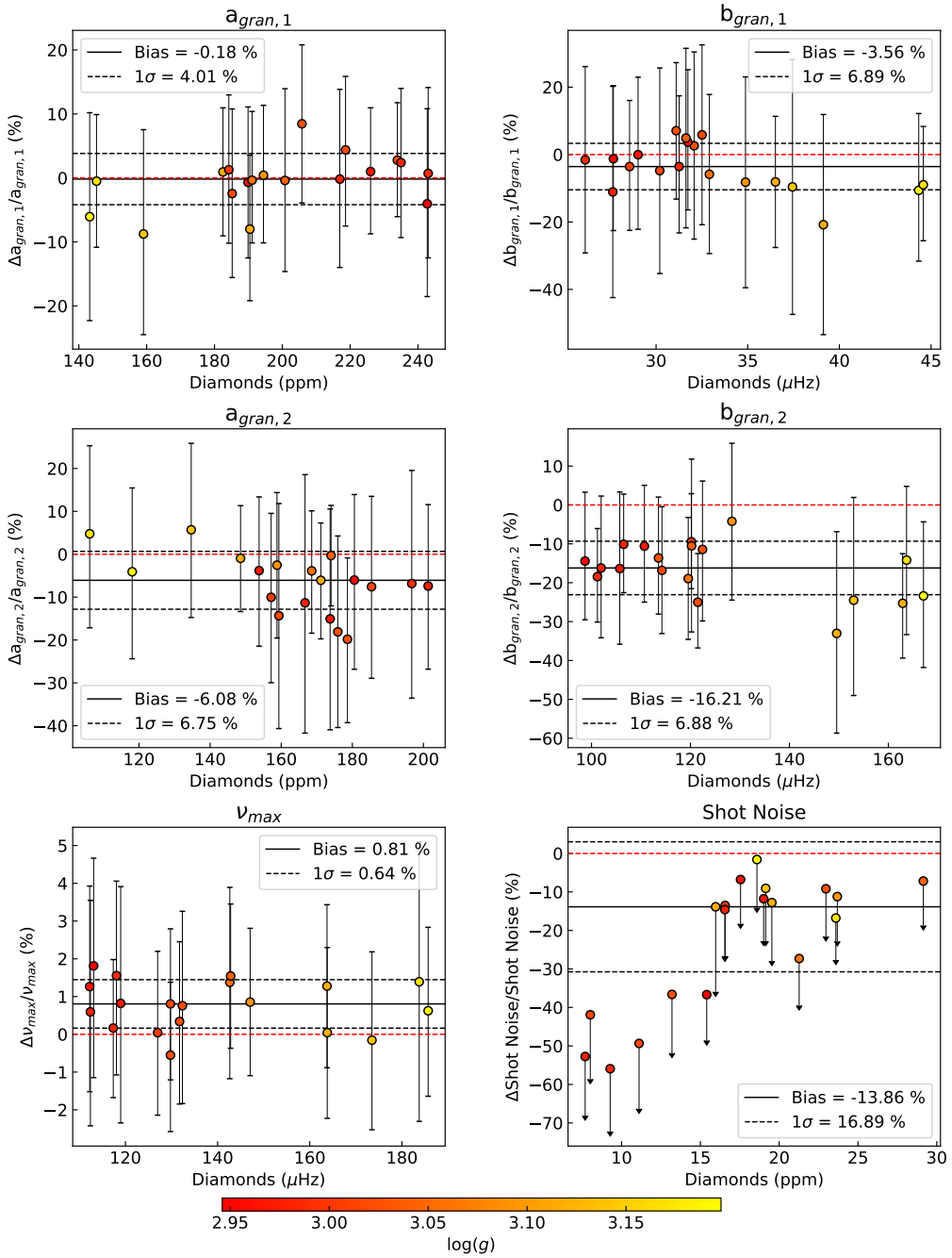

Figure 7. Comparison of the parameters in the fit of Model 2 to the *Kepler* time series data both by means of a GP regression and power-spectrum fitting. Data points represent the relative deviation with respect to the value determined using the PSD-fitting procedure, with error bars corresponding to the sum in quadrature of the uncertainties of both methods. Relative deviations of the shot noise level are computed based on an upper bound returned by the GP regression (therefore explaining the downward arrow). Black solid and dashed lines represent the median and standard deviation of the data points, respectively, with their numerical values shown in the inset. The red dashed line denotes a null offset. Data points are colour-coded according to a star's surface gravity, $\log g$.

This paper has been typeset from a TEX/LATEX file prepared by the author.

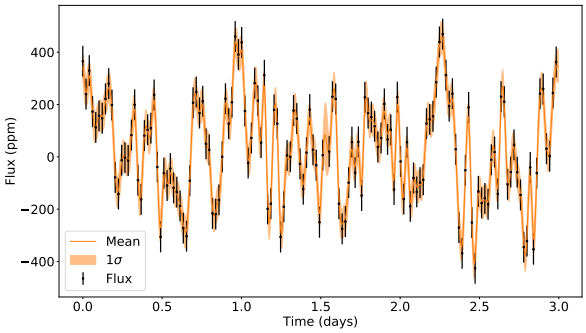

Figure 8. Predictive model (Model 2) output by the GP regression (mean and 1σ interval) when applied to one of the *Kepler* LLRGB stars in the sample. The plot is zoomed in on the first $\sim$ 3 days of observations to improve visualization.

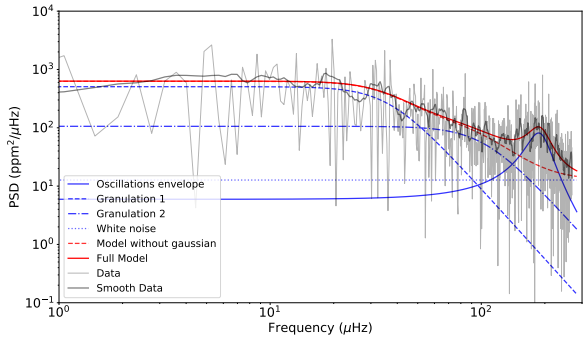

Figure 9. Power spectral density of the same (full) light curve depicted in Fig. 8. The PSD of the light curve is shown in light grey, with a slightly smoothed version overlapped in dark grey. The PSD of the GP regression output (Model 2) is shown as a solid red curve, with individual components shown in different line styles and colors (see legend).