Kepler Data Analysis: Non-Gaussian Noise and Fourier Gaussian Process Analysis of Stellar Variability

Jakob Robnik and Uroš Seljak

1 University of Ljubljana, Faculty of Mathematics and Physics, Jadranska 19, 1000 Ljubljana, Slovenia
2 Department of Astronomy and Department of Physics, University of California, Berkeley, CA 94720, USA
3 Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 93720, USA

Received 2019 October 2; revised 2020 March 26; accepted 2020 March 27; published 2020 April 22

Abstract

We develop a statistical analysis model of Kepler stellar flux data in the presence of planet transits, non-Gaussian noise, and stellar variability. We first develop a model for the Kepler noise probability distribution in the presence of outliers, which make the noise probability distribution non-Gaussian. We develop a signal likelihood analysis based on this probability distribution, in which we model the signal as a sum of the star variability and planetary transits. We argue that these components need to be modeled together if optimal signal is to be extracted from the data. For the stellar variability model we develop an optimal Gaussian process analysis using a Fourier-based Wiener filter approach, where the power spectrum is non-parametric and learned from the data. We develop high dimensional optimization of the objective function, where we jointly optimize all the model parameters, including thousands of star variability modes, and planet transit parameters. We apply the method to Kepler-90 data and show that it gives a better match to the stellar variability than the existing methods, and robustly handles noise outliers. As a consequence, the planet radii have a higher value than what the existing methods give, including splines and celerite.

Unified Astronomy Thesaurus concepts: Exoplanet detection methods (489); Non-Gaussianity (1116); Astronomy data modeling (1859); Exoplanet astronomy (486); Astronomy data analysis (1858)

1. Introduction

The Kepler Space Telescope operated on its primary mission in the years 2009–2013, aiming to photometrically detect exoplanets using transits. It measured flux from 200,000 stars and detected 18,000 potential planets of which 2000 are confirmed as of now, spanning the range of masses down to approximately Earth-sized objects (Muirhead et al. 2012).

A common approach to the planet detection is to search through possible periods of planets, folding signal phase-wise, and seeking high signal-to-noise events as described in Tenenbaum et al. (2013). One of the challenges is to distinguish the true planets from the false positives. For example, a proposed planet could be an eclipsing binary star, a single or multiple noise event resembling a planet transit, a fluctuation of host star’s brightness, an event in an offset star, which is a star in the same field of view but has no physical contact with a given star (Bryson et al. 2017), etc. A traditional approach is to perform a series of tests, each designed to target a specific group of false positives, and eliminate them if a candidate does not pass these individual tests. Transits are checked for uniformity of transit depths, consistency, possible correlation with other known planets in the given system, and for the shape of transit by calculating a metric distance (locality preserving projection metric) from the known planet shapes (Thompson et al. 2018).

A first step in the analysis is to have a good probabilistic model of all the components that contribute to the observed data. In this paper we present such analysis, where we statistically analyze several different components constituting the measured flux. Our goal is to develop a probability distribution of the data, which can then be used to assess the probability of a given observed transit-like shape to be a planet. Here we will build a model describing the incident flux and analyze its components. We first build a model of the noise probability distribution based on the analysis of stars where there are no planets (Section 2). It enables us to rigorously analyze outliers and it eliminates the need to use robust statistics such as outlier rejection. This has a significant impact on the statistical significance of the proposed planets. We proceed by modeling stellar variability and planet transits in Section 3. We argue it is crucial to analyze these components simultaneously. A common practice in the literature is to fit the spline with different knot spacings, and iteratively remove the outliers, while searching for the lowest Bayesian information criterion in order to prevent overfitting (Vanderburg & Johnson 2014). The spline is then removed from the data and the planets are fitted separately. This procedure is suboptimal because it eliminates the signal with no knowledge about its origin, potentially removing part of the planet signal as well. We propose an algorithm for a joint analysis in Section 3.3, where both star variability and the planets are fitted together. We model star variability with the Fourier-based Gaussian process (GP). Other GP-based analysis procedures have been developed in the recent years, such as Farr et al. (2018), K2 Systematics Correction (Aigrain et al. 2016), Everest (Luger et al. 2016), and celerite (Foreman-Mackey et al. 2017). We demonstrate the advantages of our analysis compared to the naive spline fit and the sophisticated as well as scalable celerite in Section 4. We apply different approaches to a signal from Kepler-90, a star known to host eight planets, which is the largest known planetary system, together with our own.

The structure of this paper is as follows. In Section 2 we develop the noise probability distribution model. In Section 3 we develop the flux probability distribution model, accounting for the stellar variability and the planet signal. In Section 4 we apply these components to analyze planet signals in an example of Kepler-90. This is followed by conclusions in Section 5.
In this section we build a model of the noise probability distribution that will enable us to rigorously analyze planet detections and make use of robust statistics, such as outlier removal, unnecessary. We use the Kepler data\(^4\) processed through the pre-search data conditioning module (Jenkins et al. 2017), which eliminates systematic instrumental errors like features of the 90 day rotation of a telescope, temperature changes of the aperture, etc. We use Pre-search Data Conditioning Simple Aperture Photometry (PDCSAP) flux, where long-term trends have also been eliminated. For this data reduction the noise is observed to be uncorrelated, i.e., flat in the frequency domain (Figure 2).

In the absence of planets the signal \(y(t)\) is composed of stellar variability, denoted as \(s(t)\), and noise, \(n(t)\),

\[
y(t) = s(t) + n(t).
\]

To determine the noise distribution we also need to determine the stellar variability, \(s(t)\). We will describe a method to do so in the next section: here we assume we have this available, which means we need to iterate on the noise and on the stellar variability. Assuming we have converged on \(s(t)\), in Figure 1 we show the probability distribution of the noise, \(n(t)\). The central core of the data is normally distributed, except for the outliers in the tails. A single light curve may not have enough outliers to analyze them statistically, so we show the results from a single star after planet removal, as well as combining flux from 76 stars, choosing the stars that are believed to have no planets, which enables us to study the negative outliers as well.

Flux from the different stars has a different amplitude and so also a different level of noise: we rescale it in order to treat it as a realization of the same noise distribution. We choose the normalization in such a way that the contribution of the Gaussian distribution to the noise has equal variance (\(\sigma^2\)). To compute \(\sigma\) we first identify the central, Gaussian part of the distribution. A measurement will be defined to belong to the central part of a distribution if its deviation from the average of a central part is less than \(1\sigma\): \(|n - \mu| < \sigma\). This choice is arbitrary, but it suffices to pick any interval where the distribution is still Gaussian and not influenced by the outliers. \(\mu\) and \(\sigma\) are initially unknown and determined by an iterative procedure. An initial guess is \(\mu_0 = E[n]\) and \(\sigma_0^2 = \text{Var}[n]\), where \(n\) are all the noise measurements from the star. In the next step only measurements that are less than \(1\sigma\) away from the average are taken into account. In this way the impact of the outliers is eliminated. A new \(\sigma\) is computed in such a way to

---

\(^4\) https://exoplanetarchive.ipac.caltech.edu/bulk_data_download/
match \( \text{Var}[n_i; n_i \in n, |n_i - \mu| < \sigma] = \sigma^2 \int_{-\infty}^{\mu} N(\lambda^2) \, d\lambda \). We analyze the combined fluxes to extract the probability density. As shown in Figure 1, the result is a combination of the Gaussian distribution in the core and a shallower slope governing the outliers.

We form the noise probability density function as a sum of two terms: the dominant Gaussian contribution, accounting for the majority of the noise events, and a function modeling the outliers. The density function for the outliers is a power law that stretches to high variance, but becomes irrelevant for small values when compared to the Gaussian contribution. It is also evident that the negative and positive outliers are asymmetric: there are considerably more positive outliers and they also have different slopes. We model this with a non-central Student’s \( t \)-distribution, which can cover asymmetric non-Gaussian probability distributions. The noise probability function is then given by a mixture model

\[
p(n) = (1 - a) \, N(\mu, \sigma) + a \, \text{NCT} \left(\frac{n - b}{c}, \nu, \psi\right),
\]

where NCT is a non-central \( t \)-distribution with parameters \( \nu \) and \( \psi \). Here \( b \) and \( c \) are used to rescale \( n \), while parameter \( a \) is a measure of the contribution of the outliers compared to the Gaussian core and is on the order of 1%. The power-law decay of the Student’s distribution for large values dominates the Gaussian distribution, while small \( a \) make it negligible for small \( n \).

Optimal parameters of the Gaussian and non-Gaussian components of analytic probability density function and their relative weight \( a \) can be extracted by minimizing the negative log-likelihood of the star flux data with no planets. The resulting noise model is shown in Figure 1, together with the data. We see that we obtain a satisfactory fit to the data.

3. Flux Variability Model

To develop the flux variability model we assume two components, one for the stellar variability and another for the planet transits. In this work we ignore additional possible components, such as eclipsing binaries. We will then proceed in the next section to perform a joint likelihood analysis of the two, allowing for a combination of planets and stellar variability, given the non-Gaussian noise distribution encoded by the likelihood.

3.1. Stellar Variability

We approach stellar variability as an non-parametric GP whose kernel hyperprior parameters we determine from the data. The GP assumes that the data are Gaussian distributed around the mean, with correlations between the flux values. These correlations help determine the stellar flux as a function of time, and it is crucial that the correlations are properly described. Typically in a GP one uses a kernel description of the correlation function, with a simple form such as a Gaussian or Mattern kernel, and a few parameters only to describe the kernel. If the correlation structure is complex such a description is not sufficient. Stellar variability models can be very complex, exhibiting phenomena such as quasi-periodic oscillations on several scales (Basri et al. 2010). We expect the process to be stationary, and for this reason we will use Fourier basis expansion, in which case the Gaussian hyperprior is the power spectrum of the flux. Because the Kepler data are sampled uniformly, sampling theorem guarantees that the Fourier expansion is complete as long as the power vanishes at Nyquist frequency. As a consequence the kernel parameters can be modeled essentially non-parametrically, by evaluating the power spectrum as a function of frequency in terms of bandpowers, where several nearby frequencies are averaged over, such that the resulting power spectrum estimate has a sufficiently small error to prevent overfitting. Once this is determined, the flux reconstruction itself is minimal variance (and hence optimal), which is known as the Wiener filter. Our model differs from other GP approaches such as celerite (Foreman-Mackey et al. 2017), in that our kernel optimization is non-parametric and thus close to optimal. We will show below that this has important implications for the planet amplitude determination. Our method is also very fast, since it is based on fast Fourier transforms (FFT), which scale as \( O(N \ln N) \), where \( N \) is the number of time stamps. We apply a small amount of zero padding at the ends to avoid issues with periodic boundary conditions with the FFTs, treating them as masked data, similar to how we treat the gaps in the data. Zero padding and gaps lead to a power suppression that can be corrected for with simulations, as we describe in more detail below.

We work with stellar variability in the frequency domain. Fourier components of stellar flux are \( \xi(\nu) = \mathcal{F}^{-1}(s(t)) = u(\nu) + iv(\nu) \), satisfying \( s(-\nu) = s(\nu) = u(\nu) - iv(\nu) \), as we have real valued signal in the time domain and complex Fourier modes. The goal of this optimization is to find optimal \( s(\nu) \) by minimizing the negative log-likelihood function. We have as a hyperprior the power spectrum \( P(\nu) = \langle |s(\nu)|^2 \rangle \), which needs to be determined first. We will assume that the stellar variability is a GP, which we will verify by the final result of the analysis.

3.2. Planet Transits

We will model planetary transits with two parameters that determine the limb-darkening profile of a star \( (u_1, u_2) \), which will be considered as known in the process, and four parameters of interest that are properties of the planet: period \( T \), phase \( \phi \), depth of transit \( A \), and time of transit \( \tau \). Our approach is not novel, many implementations for computing transit curves are similar (Parviainen 2015), and we use a minimalistic model with only the necessary parameters. The shape of the transit is primarily determined by the stellar flux density \( j \), which is the intensity per surface area emitted in the line of sight. Limb darkening makes it a function of angle \( \alpha \), which is spanned by the observer, center of the star, and the point on the surface of the star emitting the flux. This is approximated by the second-order polynomial in \( \cos \alpha \), as proposed in Kipping (2013),

\[
\frac{j(\cos \alpha; u_1, u_2)}{j(1; u_1, u_2)} = 1 - u_1 (1 - \cos \alpha) - u_2 (1 - \cos \alpha)^2.
\]

We integrate over the planet’s shadow to get the total flux reduction. More sophisticated models, accounting for terms beyond quadratic exist (Mandel & Agol 2002), but we found that this model has sufficiently small residuals when applied. We assume constant velocity of the planet during the time of transit and neglect orbital eccentricity. Orbital inclination is
taken into account by leaving the time of transit as a parameter, and not fixing it to \( \tau = R_\star \sqrt{4T/\pi G M_\star} \), as would be expected for a perfectly aligned transit using Kepler’s law. The radius of a planet \( (r) \) is not an independent parameter, but can be calculated from \( A \) as

\[
A = \frac{j(1; u_1, u_2) \pi (\frac{r}{R_\star})^2}{2\pi j(\cos \alpha; u_1, u_2) \cos \alpha \, d \cos \alpha} = \frac{(r/R_\star)^2}{1 - u_1/3 - u_2/6}, \tag{4}
\]

where \( R_\star \) is the radius of a star. It impacts the flux profile through the integral over the planet’s shadow and in the edge of the transit where only a part of the planet’s shadow covers a star. Both effects are small when the radius is sufficiently small and will be computed from the initial guess on the planet’s radius, and then held fixed during the optimization. If this assumption is not valid, as for example for binary stars, optimization should be iterated with respect to the radius. For our purposes it is valid and enables us to prepare the shape of transit in advance: we represent it with a spline interpolation \( S(t; r, u_1, u_2) \). We take into account the finite time lapse between the measurements \( \Delta t \), so that what we measure is in fact an average value of the flux during this time lapse. Thus the shape of transit for \( t \in [0, T) \) is given by

\[
U(t, T, \phi, A, \tau) = \frac{A}{\Delta t} \int_{t_\Delta t}^{t + \Delta t} S(t' - \phi \tau; r, u_1, u_2) \, dt', \tag{5}
\]

which in the absence of time transit variations (TTV) repeats periodically with a period \( T \).

### 3.3. Joint Analysis

In our model the observed data \( y(t) \) is composed of three terms: stellar variability, planet transits, and noise, so that the noise is

\[
n(t) = y(t) - s(t) - \sum_i U(t, T_i, \phi_i, A_i, \tau_i). \tag{6}
\]

where \( U \) is the shape of the transits defined in Equation (5), \( T_i, \phi_i, A_i, \tau_i \) are parameters of the \( i \)th planet, \( s(t) \) is the stellar flux, and \( n(t) \) is a realization of the noise distributed according to Equation (2). In previous sections we developed models for each contribution. In order to extract the parameters of planet transits correctly both stellar variability and planets should be analyzed simultaneously.

Posterior of the model parameters is the product of the data likelihood \( p(n) \) and the prior. The former is given by the noise probability distribution of Equation (2), while the latter is given by the Gaussian probability distribution of \( p(s(\nu)) \), with variance given by the power spectrum \( P(\nu) \); we assume a flat non-informative prior for the planet terms. We minimize the loss function, \( \mathcal{L} \), which is the negative log posterior against \( s(\nu) \) and \( T_i, \phi_i, A_i, \tau_i \), iterating on the power spectrum \( P(\nu) \),

\[
\mathcal{L} = -2 \sum_i \ln p(n(t)) + \sum_\nu \ln \frac{[s(\nu)]^2}{P(\nu)} + \ln p(\nu), \tag{7}
\]

where \( n(t) \) is given by Equation (6) and \( p(n(t)) \) by Equation (2). The first sum is over all the measured time stream data: if there is no data in certain time bands we simply omit it from the sum.

Minimization of \( \mathcal{L} \) is performed in the high dimensional space of the planet parameters and Fourier modes. It is aided by the availability of gradients for the FFT modes, so despite many components (typically thousands of modes), their optimization is not expensive, since the model is linear in terms of their amplitude, so the problem is convex and the second-order Newton method obtains an exact solution with a single step. On the other hand, planet parameters are only a few, but they are nonlinear, except transit depth. A better approach is thus to perform the fast convex optimization with respect to Fourier components at fixed planets parameters, and with respect to planets parameters while holding the Fourier components fixed. We first fix the planetary transit parameters to the current best guess and minimize the loss function \( \mathcal{L} \), given in Equation (7), with respect to the stellar variability parameters \( s(\nu) \). Then we fix the stellar variability parameters, and find a better solution of the planetary transit parameters. We iterate this procedure for convergence. The loss function is minimized with a gradient-based algorithm, L-BFGS-B, implemented in SciPy (Virtanen et al. 2020). The final results are optimal parameters of the planet’s transits and the stellar variability parameters. The hyperprior on the spectral power of the stellar variability is assumed to be given in this optimization, but is in practice unknown, and also requires an iterative procedure. We address this issue in the following section.

#### 3.4. Power Spectrum Estimation

We would like the kernel hyperparameters to be the power spectrum \( P(\nu) \), and we would like it to be as parameter-free as possible, while at the same time being determined with a sufficiently small error to prevent overfitting, so that the subsequent GP/Wiener filter analysis is reliable. Stellar variability among the different stars is very diverse, so using other stars is unlikely to be useful to determine the hyperprior, and for this reason we will determine it only from the data of the given star itself. The idea is that if we find the correct hyperprior, then stellar variability parameters extracted from the data with the joint analysis in Section 3.3 will be the same as if we used this hyperprior to simulate stellar variability. We will thus construct a iteration scheme which will match the two. The initial guess for the hyperprior power spectrum can be any function that resembles a power spectrum of a star: we use a flat hyperprior at low frequencies and a power-law decay at high frequencies. In the \( n \)th step of the iteration the joint fit gets us optimal Fourier modes of the stellar variability \( s^{(n)}(\nu) \) given our current best guess of the hyperprior \( P^{(n)} \). We simulate the data of the stellar variability and noise under the assumption that \( P^{(n)} \) is the correct hyperprior. We choose random phases with given power \( P^{(n)} \), transform it into the time domain, add noise drawn from the noise probability distribution model found in Section 2, add zeros when Kepler did not measure the flux and in the zero padding region at the edges, Fourier transform back into the frequency domain, and compute the power spectrum. We repeat this random realization of \( P_{\text{sim}} \) to compute the expectation \( \langle P_{\text{sim}} \rangle \). We compare this simulated power spectrum with the power spectrum of the stellar variability \( [s^{(n)}] \) found from the joint fit assuming hyperprior \( P^{(n)} \). If the two agree we have converged. If not we can correct
There is a 25% offset between the hyperprior and drives the Wiener filter. A sign of convergence is that the power spectrum drawn from the simulation with the given hyperprior results in a spectrum that is not important for reconstruction, as the effect of noise, and we eliminate it by a low-pass filter to zero. We see that the measured power spectrum contains low frequency power caused by the stellar variability that exceeds the noise power.

4. Application to Planet Detections

We apply it first to simulations, to show how it performs in situations where we know the answer, and then to the real data, using the example of known planets of Kepler 90.

We also compare our method to other existing methods that are currently used in the literature. Here we choose two popular methods, one based on the spline fitting and second based on the GP in the time domain.

4.1. Spline Fitting with Bayesian Information Criterion

A classical method, frequently used in the literature, is fitting a spline to eliminate the stellar variation, and then separately fitting the planets. To prevent overfitting, the spline method is repeatedly applied with different knot spacings to obtain the lowest Bayesian information criterion coefficient (Vanderburg & Johnson 2014). The method is relatively simple, but risks eliminating the flux that contains information about the planets. This can lead to a reduction in the amplitude of the planet and lower signal-to-noise ratio (S/N) of the spline fit. As we argue above, for our Fourier GP fit method we first determine the optimal power across the frequency range, with no need to use heuristics such as Bayesian information criterion (Weakliem 1999).

4.2. Celerite—GP

GP provides a more elaborate alternative as compared to the naive spline fit. GP in the time domain is not easily applicable to the large data sets as it scales as $O(N^3)$, where $N$ is number of time stamps. Celerite (Foreman-Mackey et al. 2017) is a specialization limited to the specific form of the kernel functions, but with the benefit of scaling as $O(NJ^2)$, where $J$
is the number of GP components. This is an advantage that makes it possible to apply it to the time series of interest.

We construct a celerite type kernel for a GP which will fit the stellar variability and planet’s transits parameters. Mean of the GP is given by the planets’ flux. Following Foreman-Mackey et al. (2017) we model the kernel function or autocorrelation \( k(\tau) \) with the stationary noise term and three stellar rotation terms:

\[
\begin{align*}
k(\tau) &= \sigma^2 \delta_{\tau_0} + \sum_{i=1}^{J} a_i \cos \left( \frac{2\pi \tau}{P_i} \right) e^{-\tau/L_i}.
\end{align*}
\]

(9)

In the celerite paper only one stellar rotation term was proposed, but in our case the power spectrum is dominated by three peaks as can be seen from Figure 2, so we choose \( J = 3 \). Adding two additional such terms gives a much better resemblance with the true power spectrum and reduces the bias in the planet amplitudes. This GP thus has \( 10 + 4 \cdot \# \) planets parameters, which determine the log-likelihood function. Optimal values of the parameters are then found by minimizing it with a nonlinear optimizer. A limitation of this method is in restricting the kernel function to a very specific form. The power spectrum is a Fourier transform of the correlation function \( k(\tau) \)

\[
P(\omega) = \sigma^2 + \sum_{i=1}^{3} \frac{a_i L_i}{1 + L_i^2 (\omega - \frac{2\pi \tau}{P_i})^2},
\]

(10)

falls off quadratically with frequency at high frequencies, which may not hold for the real star.

In our experiments with \( J = 3 \) celerite was slower when compared to our Fourier GP fit, but the detailed timings can depend on the choice of the nonlinear optimization in celerite. In contrast, Fourier GP does not have nonlinear optimization: the power spectrum is determined non-parametrically, and the Wiener filter is a linear optimization problem that can be solved with a single Newton–Raphson step.

The celerite kernel is based on an impulse approximation of a stellar pulsation process, which has its advantage in that it is physics motivated. However, there are stellar variation processes that are not well described with this model, and for these our non-parametric Fourier power spectrum description is a better choice. Thus we expect that Fourier GP will be particularly useful for complex power spectra that require many components for accurate modeling.

4.3. Comparison on Simulations

We first test our method on simulated flux, comparing our method to the spline fit and celerite. A performance comparison of these methods is shown in Figure 3. We simulate the stellar variability (random phases of Fourier components with power resembling the power spectrum of Kepler 90), white noise, and a planet transit with parameters typical of small inner planets (10 day period and 4 hr time transit). We vary the amplitude \( A \) of the planet’s transit depth, which is directly proportional to the \( S/N \) of the planet. We compare the methods in their ability...
to reproduce the simulated amplitude. As shown in Figure 3, our joint fit with Fourier GP results in a more reliable estimate of amplitude, typically within 2% of correct values, except for very low S/N, where in most cases it rises to a 5% deviation at S/N = 6. In contrast, the corresponding spline and celerite fits are systematically too low, with the negative bias up to 30%.

4.4. Comparison on Kepler-90 Data

We will now demonstrate our method on a concrete example of the Kepler-90 star. We chose this star because it is known to host seven or eight planets and has received a lot of attention in the literature. We use light curves that were preprocessed by the Kepler pipeline. This includes bias voltage correction, calibrated pixels, removed artifacts like cosmic rays, and identified pixels of target stars (Jenkins et al. 2017). We use long cadence light curves, which have a 1460 day long signal (with gaps), with 29.4 minute spacing. We normalize the flux in different quarters as described in Section 2, subtract the average, and add zeros with infinite variance in places where there is no data so quarters can be concatenated in one signal with evenly spaced measurements, unit variance, and zero average. Starting from the published planet parameters we do a joint fit of planets and stellar variability.

In Figure 4 we compare the methods in time and in frequency domain. The top panel shows that planets dominate the spectral density at high frequencies, a consequence of strong planetary signatures in this system. We remove the planets using the Fourier GP fit, and measure the power spectrum of the resulting stellar component fit. We see that our method assigns less power at high frequencies than the celerite, which explains why the celerite method underestimates the amplitude of the planet transits, since it assigns some of the planet signal to the stellar variation. The spline method has a comparable power spectrum to our method, but in general is a worse fit to the stellar variability. Overall, our joint Fourier GP method is better at fitting planets and stellar variability than these alternatives.

The results for seven known planets and the proposed eighth planet Kepler-90 are given in Table 1. In the first two columns, a spline is repeatedly fitted to find optimal Bayesian information criterion, eliminated, and then the planets are fitted. An absence of an individual planet in the fit gives a higher $\chi^2$, and we report the square root of the difference. If we take in account the non-Gaussian contribution of the noise we report the corresponding value of the square root of $-2\Delta \ln p$. The third column shows $[-2\Delta \ln p]^{1/2}$ for the joint analysis with the Fourier GP and the planets fitted together. The fourth column is from the NASA Exoplanet Archive (Kepler Science Operations Pipeline 2016). It is a systematic search through many solar systems and does not take into account out-of-phase transits of Kepler-90g and Kepler-90 h caused by TTVs, which results in a lower S/N for these planets, compared to the spline fit using Gaussian noise. We can see that a joint fit often gives lower S/N when compared to the spline fit, despite the fact that we predict higher S/N on simulated planets. We argue this is a consequence of our improved fit of stellar variability, and that our model is better at modeling the star as a source of false positives. Kepler-90 g and Kepler-90 h (the two biggest planets) have lower $-2\Delta \ln p$ than $\Delta \chi^2$. Their large amplitudes in $\chi^2$ are likely to be noise fluctuations caused by outliers.
Table 1

Comparison of the S/Ns and Radii of Planets in the Kepler 90 System Using Different Methods

| Planet | Spline $\sqrt{\Delta N^2}$ | Spline $\sqrt{2\Delta \ln \rho}$ | Fourier GP $\sqrt{2\Delta \ln \rho}$ | Official S/N | Spline $r [R_\text{\tiny J}]$ | Fourier GP $r [R_\text{\tiny J}]$ | Official $r [R_\text{\tiny J}]$ |
|--------|-----------------------------|---------------------------------|-----------------------------------|-------------|-----------------------------|-----------------------------|-----------------------------|
| b      | 16.2                        | 16.8                            | 15.5                              | 16.7        | 1.22                        | 1.29 ± 0.04                 | 1.31 ± 0.17                 |
| c      | 14.0                        | 15.5                            | 15.8                              | 16.6        | 1.22                        | 1.36 ± 0.05                 | 1.19 ± 0.14                 |
| i      | 4.9                         | 5.4                             | 6.7                               | unknown     | 0.81                        | 1.02 ± 0.06                 | 1.3 ± 0.2                   |
| d      | 35.1                        | 35.1                            | 29.6                              | 18.9        | 2.63                        | 2.83 ± 0.03                 | 2.9 ± 0.3                   |
| e      | 28.6                        | 28.6                            | 23.5                              | 25.1        | 2.49                        | 2.72 ± 0.04                 | 2.7 ± 0.3                   |
| f      | 24.6                        | 24.6                            | 16.6                              | 30.2        | 2.7                         | 2.69 ± 0.06                 | 2.9 ± 0.5                   |
| g      | 224.1                       | 61.6                            | 113.4                             | 121.8       | 7.72                        | 7.72 ± 0.02                 | 8.1 ± 0.8                   |
| h      | 336.6                       | 49.9                            | 142.6                             | 233.3       | 10.81                       | 10.82 ± 0.02                | 11.3 ± 1.0                  |

Note. Only the statistical error is included in the Fourier GP fit radius, while the official radius error also includes the contribution from the error on the star radius.

Perhaps more importantly, it is evident from Table 1 that the S/N can be dramatically changed if the stars and planets are fitted together and if the true noise probability distribution is used instead of a Gaussian assumption. This is another indication that the outlier contribution to the noise probability distribution is important and that simplification to do a separate fit of planets and stellar variability is not justified. We believe that the Fourier GP fit results in Table 1 give the most reliable estimate of the significance of detection to date.

The fifth and sixth columns show that the spline fitting method generally underestimates the radius of the planet, consistent with the simulation results in Figure 3. The last column is an official estimate on planet’s radius (Rein 2012). One can observe that official errors are significantly larger than those found from our analysis, but this is because we only include the statistical error in our Fourier GP analysis, while the official error also includes the error on the star radius, which is uncertain at about a 10% level, while here we used a fixed value of 1.2 solar radius in our fits.

In terms of the estimated radius we observe discrepancies between the spline and Fourier GP fits, which are consistent with Figure 3. We chose to explicitly investigate it further for Kepler-90 c. In Figure 3 we show results of a synthetic analysis of such a planet, where we inject it into the data at a random phase with a period of 8.7 days and with an amplitude consistent with Kepler-90 c, and then analyze it with the two methods. Results show that the spline analysis underestimates the radius by about 10% (or amplitude by about 20%), comparable to the difference we observe between the spline fit and the Fourier GP fit of actual Kepler-90 c.

In light of these improvements it is worth revisiting the credibility of the reported eighth planet in this system, Kepler-90 i (Shallue & Vanderburg 2018). As one can see from Table 1, the spline fit would assign it to be around 5σ, which is too low for a real detection, since noise realizations can make a signal as high as 6σ due to the large look-elsewhere effect present in planet searches. The Fourier GP analysis increases this to 6.7σ, as we would expect from Figure 3, where the Fourier GP analysis fits are always higher in amplitude and more consistent with the correct value. A more detailed analysis would need to evaluate the Bayesian evidence of this planet, which is beyond the scope of this paper.

5. Conclusions

In this paper we develop methodology for analyzing stationary time-series data, such as stellar flux data, in the presence of non-Gaussian noise and time-correlated stellar variability. After the data have been preprocessed to account for calibrations, one is given a calibrated time series, which may have a non-Gaussian noise distribution. As a first step we develop a likelihood analysis using the true noise probability distribution. Due to the presence of outliers the likelihood has higher probability at high values of flux, which reduces their impact on the fit, and makes the analysis robust without the need to eliminate the outliers by hand. In this paper we define noise as the difference between the observed data and the signal that combines the stellar variability with the planet transits. As a consequence, the noise probability distribution can only be determined as part of the full iterative analysis that also determines the stellar variability and the planet transits at the same time.

Next we address the stellar variability using a Fourier space GP (Fourier GP). We adopt non-parametric kernel hyperprior parameters as the Fourier space power spectrum amplitudes at different frequencies, and develop an iterative procedure that determines the power spectrum together with the stellar variability reconstruction. The latter is a linear Wiener filter of the data given the power spectrum. In contrast to existing methods (Vanderburg & Johnson 2014; Foreman-Mackey et al. 2017), ours has the advantage of having more flexibility in terms of the GP kernel, which is essentially non-parametric and can fit even detailed features in the stellar spectrum. It is also very fast since it is based on FFTs, scaling as $O(N \ln N)$, where $N$ is the number of time stamps, and in our tests we found it to be comparable or faster than alternatives. One such alternative is celerite (Foreman-Mackey et al. 2017), which is also a GP, but whose kernel hyperparameters are of a more restricted form and scales as $O(N^2)$, where $J$ is the number of kernel components. As a result of this scaling it slows down considerably when a complex kernel with many components ($J \gg 1$) is needed.

We apply our method to simulated planets injected into simulated and real stellar variability data and show that we recover the signal more accurately than the current standard practices. One of these is a separate spline fit, followed by a planet transit fit, which is a poor fit to the data, and underestimates the planet transit amplitudes. Our Fourier GP method gives essentially unbiased amplitude estimation down to the lowest detectable signals.

We also apply our method to the real data of Kepler-90, and show that it gives considerably different results when compared to the splines with Gaussian and non-Gaussian error distributions, as well as when compared to the official
S/N and radius numbers. This shows that the specifics of the analysis can affect the results. While we only analyzed the effects on the transit amplitude, we expect there will be similar effects on the other planetary parameters, such as period, phase, or TTVs. We argue that our approach is close to optimal and whenever high-precision planetary transit parameters are needed, a joint star variability and planet transit analysis should be performed, together with the proper non-Gaussian noise likelihood analysis. All of these aspects are included in our current version of the code, which is freely available. Other astrophysical sources can be added to the Fourier GP fit, such as eclipsing binaries, which we plan to do in the future.

We thank Stephen Bryson and Dan Foreman-MacKay for useful comments. We acknowledge ASEF for providing funding for J.R. to visit UC Berkeley. This material is based upon work supported by the National Science Foundation under grant Nos. 1814370 and NSF 1839217, and by NASA under grant No. 80NSSC18K1274.

ORCID iDs
Jakob Robnik © https://orcid.org/0000-0003-1929-526X

References
Aigrain, S., Parviainen, H., & Pope, B. 2016, MNRAS, 459, 2408
Basri, G., Walkowicz, L. M., Batalha, N., et al. 2010, ApJL, 713, L155
Bryson, S. T., Batalha, N. M., Colón, K. D., et al. 2017, Kepler Certified False Positive Table, Kepler Science Document, KSCI-19093-003
Farr, W. M., Pope, B. J., Davies, G. R., et al. 2018, ApJL, 865, L20
Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, AJ, 154, 220
Jenkins, J. M., Tenenbaum, P., Seader, S., et al. 2017, Kepler Data Processing Handbook: Transiting Planet Search, Kepler Science Document, KSCI-19081-002
Kepler Science Operations Pipeline 2016, KIC 011442793, https://exoplanetarchive.ipac.caltech.edu/data/KeplerData/011/011442/011442793/tcert/kplr011442793_q1_q17_dr25_obs_tcert.pdf
Kipping, D. M. 2013, MNRAS, 435, 2152
Luger, R., Agol, E., Kruse, E., et al. 2016, AJ, 152, 100
Mandel, K., & Agol, E. 2002, ApJL, 580, L171
Mathur, S., Metcalfe, T., Woitaszek, M., et al. 2012, ApJ, 749, 152
Muirhead, P. S., Johnson, J. A., Apps, K., et al. 2012, ApJ, 747, 144
Parviainen, H. 2015, MNRAS, 450, 3233
Rein, H. 2012, arXiv:1211.7121v2
Shallue, C. J., & Vanderburg, A. 2018, AJ, 155, 94
Tenenbaum, P., Jenkins, J. M., Seader, S., et al. 2013, ApJS, 206, 5
Thompson, S. E., Coughlin, J. L., Hoffman, K., et al. 2018, ApJS, 235, 38
Vanderburg, A., & Johnson, J. A. 2014, PASP, 126, 948
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nat. Methods, 17, 261
Weakliem, D. L. 1999, Sociological Methods & Research, 27, 359

5 https://github.com/JakobRobnik/Kepler-data-analysis