SYSTEMATICS-INSENSITIVE PERIODIC SIGNAL SEARCH WITH K2

RUTH ANGUS1,2, DANIEL FOREMAN-MACKEY3, and JOHN A. JOHNSON2
1 Subdepartment of Astrophysics, University of Oxford, OX1 3RH, UK; ruthangus@gmail.com
2 Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA, USA
3 Center for Cosmology and Particle Physics, New York University, NY, USA

Received 2015 May 22; accepted 2015 December 29; published 2016 February 11

ABSTRACT

From pulsating stars to transiting exoplanets, the search for periodic signals in K2 data, Kepler’s two-wheeled extension, is relevant to a long list of scientific goals. Systematics affecting K2 light curves due to the decreased spacecraft pointing precision inhibit the easy extraction of periodic signals from the data. We here develop a method for producing periodograms of K2 light curves that are insensitive to pointing-induced systematics; the Systematics-insensitive Periodogram (SIP). Traditional sine-fitting periodograms use a generative model to find the frequency of a sinusoid that best describes the data. We extend this principle by including systematic trends, based on a set of “eigen light curves,” following Foreman-Mackey et al., in our generative model as well as a sum of sine and cosine functions over a grid of frequencies. Using this method we are able to produce periodograms with vastly reduced systematic features. The quality of the resulting periodograms are such that we can recover acoustic oscillations in giant stars and measure stellar rotation periods without the need for any detrending. The algorithm is also applicable to the detection of other periodic phenomena such as variable stars, eclipsing binaries and short-period exoplanet candidates. The SIP code is available at https://github.com/RuthAngus/SIPK2.

Key words: asteroseismology – methods: data analysis – methods: statistical – stars: rotation – techniques: photometric

1. INTRODUCTION

The excellent precision achieved by the original Kepler mission relied on extremely precise pointing, for which three reaction wheels were required. After the failure of one of these wheels, the Kepler team devised a new pointing scheme in which the spacecraft is stabilized by the solar wind for ecliptic plane viewing zones (Howell et al. 2014). In this configuration the spacecraft is able to maintain an unstable equilibrium, with the two functioning reaction wheels controlling pitch and yaw while the spacecraft slowly rolls about the boresight. The spacecraft fires its thrusters once every ~6 hr (Vanderburg & Johnson 2014; hereafter VJ14) to correct for this slow drift and, as stars move across pixels with different sensitivities, their flux varies. The extraction of high-precision photometry from K2 target pixel files, despite the reduced pointing precision, is a requirement for many fields of research and several methods for the extraction and detrending of K2 light curves have already been developed. For example, VJ14 and Crossfield et al. (2015) use simple aperture photometry and correct the light curve of each star individually and Aigrain et al. (2015) use a Gaussian process to model the nonlinear dependence of stellar flux on the roll angle of the telescope.

While these methods successfully remove most systematic trends and produce light curves suitable for exoplanet search and some stellar variability studies, residual systematics can still affect the light curves on timescales relevant to asteroseismology and stellar rotation. In particular, the ~6 hr thruster firing signal may still appear with high power in the periodograms of these detrended light curves (see Figure 1). A detrending method for K2 light curves, specifically intended for the asteroseismic analysis of giant stars, has been developed by Lund et al. (2015), in which the systematics due to roll are corrected, again on a star-by-star basis and any remaining periodic signals at 47 μHz (6 hr period) or its harmonics are removed by prewhitening. The method developed here, the Systematics-insensitive Periodogram (SIP), produces periodograms of K2 light curves without the need for detrending or prewhitening.

1.1. Asteroseismology

As well as providing data that have led to the discovery thousands of exoplanets, the original Kepler mission revolutionized many fields of stellar astronomy, particularly asteroseismology. Fundamental stellar parameters—in some cases, extremely precise ones—can be calculated for Kepler asteroseismic stars from the power spectra of their light curves. Although Sun-like stars oscillate at high frequencies and require short-cadence observations, pulsations of giant stars lie below the Nyquist frequency set by the 28.5 minute sampling rate of long-cadence Kepler data: 283 μHz. Asteroseismic analysis of data from the original Kepler mission is traditionally conducted upon detrended light curves. For short-cadence Kepler data, this detrending method is described in García et al. (2011). Due to the precise pointing of the original Kepler mission, systematics present in these light curves, caused by temperature fluctuations and minor pointing shifts, are relatively low amplitude.

However, this is not the case for K2 light curves: the precision over a 6 hr timescale is estimated to be four times worse in K2 data (Howell et al. 2014), therefore new approaches to the treatment of systematics are necessary. Figure 1 demonstrates the need for careful systematics treatment of K2 photometry for asteroseismology. The top panel shows a Lomb–Scargle4 (LS) periodogram of the raw, simple aperture photometry5 of EPIC 201183188, a pulsating

---

4 All LS periodograms produced in this project were produced using the gatspy Python module: https://github.com/astroML/gatspy/tree/master/gatspy/periodic
5 The method used to extract this photometry is described in Section 2.

---


giant star. The large peaks at \(\sim 47 \mu\text{Hz}\) and its harmonics are caused by the regular spacecraft thruster fires of the spacecraft. The bottom panel shows the LS periodogram of this light curve, after it has been detrended using the method of VJ14. The large peaks are still present in the detrended light curve. The dominant source of these peaks is the removal of the outlying data points that appear in the K2 light curves every 6 hr, caused by the spacecraft thruster fires. The rapid motion of the spacecraft results in stellar flux being smeared out over the detector and the circular apertures do not adequately encompass the resulting point-spread function. It is the removal of these data points that produces the 47 \(\mu\text{Hz}\) peak in the VJ14 periodogram. While the noise source at 47 \(\mu\text{Hz}\) does not interfere with the detection of high-signal-to-noise transit events for periods greater than \(\sim 1\) day (Vanderburg et al. 2015), it does hamper the detection of smaller signals, particularly on time scales comparable to that of thruster fires. These peaks lie in an important region of parameter space for giant star asteroseismology and could affect the stellar parameters measured for thousands of giants if not dealt with appropriately.

### 1.2. Stellar Rotation

Stellar rotation studies have hugely benefited from the era of high-precision space photometry. Active regions on the surface of rotating stars produce periodic variations in flux and stellar rotation periods can therefore be measured from Kepler light curves. Stellar rotation is a field of active interest as the rotation period of star can be used to infer its age via gyrochronology (Skumanich 1972; Barnes 2007; Epstein & Pinsonneault 2014; Angus et al. 2015), is thought to be tied to the stellar magnetic dynamo, and could even reveal dynamical interactions with companion stars or planets (e.g., Béky et al. 2014; Poppenhaeger & Wolk 2014). Current methods for measuring rotation periods from Kepler light curves include periodogram (e.g., Reinhold & Reiners 2013), AutoCorrelation Function (ACF) (McQuillan et al. 2013a) and wavelet (e.g., García et al. 2014) analysis, or some combination thereof. Stellar variability is not typically sinusoidal, therefore sine-fitting periodograms are not perfectly suited to measuring rotation periods (McQuillan et al. 2013a). For this reason, the ACF method is often favored over the periodogram method. However, because autocorrelation is performed directly on detrended light curves, and cannot be written down as a generative model, it is not possible to use autocorrelation techniques on untreated K2 data.

A quasi-periodic Gaussian process is a much better effective model for stellar variability than a sinusoid; however, we choose to focus on the more generally applicable (and computationally tractable) sine-wave periodogram, leaving the Gaussian process periodogram for future consideration.

In this paper we focus on the examples of asteroseismology and stellar rotation; however, many other fields of astronomy utilize periodic information in K2 light curves. These include studies of eclipsing binaries, variable stars, exoplanets, white dwarfs and even active galactic nuclei. The development of tools for extracting periodic information from K2 data is essential if it is to be as revolutionary in time-domain astronomy as the original Kepler mission was.

In Section 2 we outline the method behind the SIP. In Section 3 we apply the SIP to real K2 light curves, using some giant asteroseismic pulsators and rotating stars as test cases and then provide the results of some simple tests which show exactly how “insensitive” the SIP is to systematic features. Finally, we demonstrate the SIP’s usefulness regarding other periodically varying objects in this section, before presenting our conclusions in Section 4.
2. METHOD

The method implemented in this paper is an extension of the planet-search algorithm developed by Foreman-Mackey et al. (2015) (hereafter FM15). All targets observed by Kepler move on the CCD in the same way, therefore the systematics affecting each individual star’s light curve have shared properties. The FM15 method uses this fact by decomposing the light curves into a set of “eigen light curves” (ELCs) using Principle Component Analysis (PCA), which can be used to model any individual star’s light curve with very little loss of information. This process is similar to the method used to produce PDC-MAP data for the original Kepler mission (Smith et al. 2012; Stumpe et al. 2012). The resulting ELCs from campaign 1 can be used to model any campaign 1 K2 light curve (campaign 0 ELCs for campaign 0, etc.) and, specifically, can model the data in combination with an arbitrary physical model.

In order to construct sets of ELCs for campaigns 0 and 1, FM15 downloaded the target pixel files for all stars in these two fields. The position of each star was predicted using the World Coordinate System and 10 circular apertures placed around the star with radii varying from 1 to 5 pixels in steps of 0.5 pixels. Following the procedure of VJ14, the aperture producing the light curve with the lowest CDDP within a 6 hr window (Christiansen et al. 2012) was selected. PCA was then performed on the full set of targets in order to produce ELCs.

FM15 used 150 of these ELCs, plus a transit model, in order to search for exoplanet candidates without the need for a “detrending” step. The likelihood of the data, conditioned on the ELC-plus-transit model, was calculated over a fine grid of periods and transit depths, resulting in the detection of 36 new exoplanet candidates. We use a very similar technique to find periodic signals in K2 data. The primary difference is that we use a sinusoid rather than a transit model. This model is linear, therefore the likelihood function conditioned on a specific frequency can be calculated and the systematics model marginalized over analytically.

Following the notation in FM15, our model for the kth star can be written

\[ f_k = A w_k + \text{noise}, \]

where \( f_k \) is the vector of \( N \) flux values,

\[ f_k = (f_{k,1}, f_{k,2}, ..., f_{k,N})^T \]

at times \( t_k = (t_1, t_2, t_3, ..., t_N)^T \).

\( A \) is the design matrix:

\[
A = \begin{pmatrix}
1 & \sin(2\pi n_1) & \cos(2\pi n_1) \\
\sin(2\pi n_2) & \cos(2\pi n_2) & \vdots \\
\sin(2\pi n_N) & \cos(2\pi n_N)
\end{pmatrix}
\]

(4)

where the \( x_{ij} \) are the ELCs, with \( i \) denoting the ELC number and \( j \) the time index. The design matrix contains the basis functions of the linear model. The basis functions for the systematic features in the light curves are the ELC values at each time index, the sine and cosine terms are the basis functions of the sinusoidal signal of interest, and the “1’s” describe a linear offset. Any K2 light curve can be reproduced as a linear combination of these basis functions. We are interested in the last two elements of the weight vector: the coefficients of the sinusoidal signal. The maximum likelihood solution for the weight vector, \( w \) is

\[ w_* = (A^T A)^{-1} A^T f_k. \]

Under this linear model with Gaussian uncertainties, the marginalized likelihood for the periodic amplitude is a two-dimensional Gaussian with mean given by the last two elements (\( \alpha \)) of \( w^* \) and covariance given by the bottom right two-by-two block (\( S_* \)) of \( (A^T C^{-1} A)^{-1} \), where the \( C \) matrix contains observational uncertainties on the diagonal. These uncertainties are estimated as \( 1.48 \times \) the median absolute deviation, following Aigrain et al. (2015). Therefore, the signal-to-noise ratio, \( S/N \), of the amplitude measurement is \( S_*^2 / S \). The \( S/N \) operation takes into account the goodness of fit, i.e., if the amplitude of the sinusoid at a given frequency is not well constrained, it is penalized. The SIP algorithm scales linearly with the number of frequencies evaluated and with the cube of the number of basis functions used. As an example of the typical computation time, calculating a single SIP for a K2 campaign 1 light curve, over a grid of 1000 frequencies with 150 basis vectors, takes 2–3 s on a 2.7 GHz CPU.

3. APPLICATION TO REAL LIGHT CURVES

An example LS periodogram of the raw K2 photometry for giant star, EPIC 201183188 is shown in Figure 1. Peaks appearing at 47 \( \mu \)Hz and its harmonics are produced by the regular \( \sim 6 \) hr thruster fires that repoint the spacecraft. These peaks are also present in periodograms of the VJ14 detrended light curves. The presence of systematic signals at these timescales are problematic for asteroseismic analysis since they lie in a region of frequency space that is often populated by giant asteroseismic modes. It is possible to remove these signals by “prewhitening” the data, i.e., subtracting a sinusoid of that frequency from the data; however, this process will artificially suppress all signals, both systematic and astrophysical, at that frequency. The SIP method eliminates the necessity for any such procedure. The bottom panel of Figure 1 shows the SIP for the same star, demonstrating the ability of the SIP method to produce periodograms that are free from thruster firing signals.

In order to search for high signal-to-noise asteroseismic modes in the giant star candidates of GO1059, we searched for a power excess in the SIPs using the method of Huber et al. (2009): autocorrelation functions were calculated for sections of the SIP in order to search for regions of increased correlation and locate the frequency of maximum power. The increased correlation arises from the even frequency spacing of acoustic modes, and the frequency of maximum correlation at the location of the power excess corresponds to the large frequency separation, \( \Delta \nu \). Figures 2(a)–(e) show example power spectra of six targets for which we detect pulsations using this method.

The top panel of Figure 3 shows the VJ14-detrended light curve of an active, rotating star, EPIC 201133037, with a linear trend subtracted off. The brightness fluctuations clearly visible...
in the light curve of this target are produced by cool active regions on the stellar surface, which reduce the stellar flux periodically. The rotation period of this star is therefore around 20 days. The middle and bottom panels show an ACF and LS periodogram of the detrended light curve. The top panel of Figure 4 shows the raw light curve of the same target in gray, with the conditioned light curve in black. This conditioned light curve was produced by removing the best fitting systematic

Figure 2. SIPS of six long cadence K2 giants with asteroseismic oscillations. These were selected from the guest observing program, GO1059 and identified using the method of Huber et al. (2009).
trends, described by a certain combination of the ELCs, at the best fitting period of the sinusoid. The bottom panel shows the SIP. Each of these three methods measures a rotation period of around 20 days for this target. This example demonstrates the ability of the SIP to recover rotation periods that agree with those measured from detrended light curves by autocorrelation.

We also include an example that demonstrates the ability of the SIP to outperform a periodogram of detrended data. Figure 5 shows the light curve, ACF and LS periodogram of another rotating star, EPIC 201142043 and Figure 6 shows its SIP. This star shows lower amplitude variability than the previous example and the careful treatment of systematics is much more important. Whereas the ACF method is able to measure a
rotation of ∼3 days for this star, the LS periodogram of the detrended light curve incorrectly measures a period of 59 days. Although there is a small peak at the rotation period of the star, it is not the dominant periodic signal. The SIP method is, by definition, insensitive to these long-term systematics and is able to measure a period of ∼2 days. This example further demonstrates the fact that long-term systematic trends caused by slow pointing variations are often not removed by conventional detrending methods. The 59 day signal is almost certainly a systematic trend produced by spacecraft pointing variations.

Figure 5. Top: light curve of EPIC 201142043, detrended using the method of V14. Middle: autocorrelation function of the detrended light curve. The autocorrelation function method measures a rotation period of 3 days for this star. Bottom: the LS periodogram of the detrended light curve. The highest peak in the periodogram is located at 59 days and is likely to be a systematic trend produced by spacecraft pointing variations.

Figure 6. Top: the raw light curve of EPIC 201142043 is shown in gray and the conditioned light curve is shown in black. The conditioned light curve is produced by removing the trends that best describe the data, at the best fitting frequency. Bottom: an SIP, produced by modeling the data as a linear combination of the top 150 ELCs plus a sine and cosine function at a range of frequencies, measuring a rotation period of 2 days. The SIP is, by definition, insensitive to the long-timescale systematics that dominate the LS periodogram of the detrended data, shown in Figure 5.
periodograms from detrended data. However, it has been shown that the ACF method often performs better than periodogram methods in general for measuring stellar rotation periods (McQuillan et al. 2013a, 2013b; Mazeh et al. 2015). For stars with relatively high-amplitude variability, for which perfect removal of systematic trends is less important, performing the ACF method on detrended data is likely to produce similar results to the SIP method. The SIP method is ideally suited to low-amplitude cases, where systematic trends could drastically influence rotation period measurements. While the SIP method may outperform ACF in the low-amplitude cases, any “marginal” rotation period measurements calculated using either method should be treated with caution unless a representative uncertainty is provided. In general neither ACF nor periodogram methods are equipped to provide such uncertainties. In practice, we recommend using both the SIP and ACF methods, in combination with a by-eye check, to measure rotation periods for K2 stars.

3.1. Tests and Discussion

In order to demonstrate the consistent ability of the SIP method to remove the signal at 47μHz, corresponding to the periodic ~6 hr thruster firings, we computed SIPs for 4923 targets from the GO1049 proposal: “Galactic archaeology on a grand scale” (PI: D. Stello). For each target, an SIP of its raw photometry and a LS periodogram of its VJ14 light curve was calculated for frequencies between 40 and 54 μHz. Both the height and frequency of the highest peak in the SIP and the highest peak in the LS periodogram were recorded. A histogram of the frequencies of the highest peaks in the SIPs of all 4923 targets is shown in the top panel of Figure 8. The bottom panel shows the histograms of peak heights within the correspondingly colored ranges indicated in the top panel. This figure shows that while there is a greater number of maximum peaks around 47 μHz, the S/Ns of these peaks are comparable to those found just above and just below this frequency. Figure 9 shows the equivalent results for the VJ14 light curves. There is a significant number of large peaks at ~47 μHz in the LS periodograms of the detrended light curves; the highest peak in the LS periodograms was almost always located at ~47 μHz. Furthermore, the distribution of peak power within the range 46.5–48 μHz is skewed toward higher powers, i.e., a substantial fraction of the peaks at ~47 μHz have a large power. The SIP method is able to consistently remove the 47 μHz signal which is present in almost every VJ14 light curve.

We performed an injection and recovery experiment in order to test the detection efficiency of the SIP algorithm. A total of 4000 sinusoids with frequencies ranging from 10 to 270 μHz and amplitudes ranging from 1 to 100 ppm were injected into the raw K2 light curve of target star EPIC 201121245, a relatively non-variable giant with low-amplitude acoustic oscillations. In order to recover the injected signals we calculated a SIP of the original light curve, subtracted this from the SIP of the injected light curve, and searched for excess power in the residuals. This allowed us to perform injection and recovery tests on this target star without being affected by the star’s own intrinsic variability. We then measured the position of the highest peak in the resulting residual SIP and recorded the successful detections, defined as those that lay within 1 μHz of the injected value. SIPs were computed over a grid of frequencies ranging from 10 to 270 μHz, with a spacing of 0.1 μHz. The resulting detection efficiency map is shown in Figure 7. This figure demonstrates that the SIP can recover the frequencies of signals with amplitudes less than 10 ppm.

Photometric variability in dwarf stars on timescales less than 8 hr, often known as flacker, has been linked to surface gravity (Bastien et al. 2013; Kipping et al. 2014). The metrics used to quantify photometric variability include finding the range in intensity, counting the number of zero crossings and calculating the root mean square (rms) of the light curve. Although these features are related to signal processing, they are operations performed on detrended light curves, not inferred from periodograms. However, it may be possible to derive a property of the periodogram that scales with the density or surface gravity of a star, for example, the mean excess power at frequencies near those relevant to granulation timescales. The SIP method presented here would be useful for such a technique.

4. CONCLUSIONS

We demonstrate that modeling campaign 1 K2 photometry as a linear combination of 150 PCA components plus a sinusoid can produce periodograms that are almost completely free from instrumental systematic signals, without the need for detrending. We find that the 47 μHz signal, generated by the spacecraft thruster fires, is not present in the vast majority of SIPs for more than 4000 targets selected from the K2 guest observer program, GO1059, “Galactic archaeology on a grand scale” (PI: D. Stello). The SIP method is highly successful for campaign 1 targets where the large number of stars, observed for a baseline of 80 days, ensures that most of the systematics are captured in the ELCs and we anticipate that it will be equally effective for the up-and-coming campaigns.

The SIP method is capable of detecting periodicities in K2 data in the region of frequency space relevant to the study of asteroseismic oscillations in giant stars and for any signals with a timescale close to 6 hr. It is also effective at measuring stellar rotation periods and is an improvement upon a simple LS periodogram of detrended data. In practice, the best approach for measuring rotation periods in K2 data is likely to be a combination of the SIP method and the ACF method, where autocorrelation is performed on detrended light curves. The SIP...
code is available for public use and can be found at https://github.com/RuthAngus/SIPK2.

It is a pleasure to thank Dan Huber (Sydney) who provided many excellent comments for this paper and useful astroseismology tips. We would also like to thank David Hogg (NYU) and Suzanne Aigrain (Oxford) for their comments, plus Andrew Vanderburg (Harvard), Ben Montet (Caltech) and Stephanie Douglas (Columbia) for their extremely helpful suggestions and recommendations regarding this
J.A.J. is supported by generous grants from the David and Lucile Packard and Alfred P. Sloan Foundations. The data presented in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST). STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. Support for MAST for non-HST data is provided by the NASA Office of Space Science via grant NNX09AF08G and by other grants and contracts. This paper includes data collected by the Kepler mission. Funding for the Kepler mission is provided by the NASA Science Mission directorate.

REFERENCES
Aigrain, S., Hodgkin, S. T., Irwin, M. J., Lewis, J. R., & Roberts, S. J. 2015, MNRAS, 447, 2880
Angus, R., Aigrain, S., Foreman-Mackey, D., & McQuillan, A. 2015, MNRAS, 450, 1787
Barnes, S. A. 2007, ApJ, 669, 1167
Bastien, F. A., Stassun, K. G., Basri, G., & Pepper, J. 2013, Natur, 500, 427
Béký, B., Holman, M. J., Kipping, D. M., & Noyes, R. W. 2014, ApJ, 788, 1
Christiansen, J. L., Jenkins, J. M., Caldwell, D. A., et al. 2012, PASP, 124, 1279
Crossfield, I. J. M., Petigura, E., Schlieder, J., et al. 2015, arXiv:1501.03798
Epstein, C. R., & Pinsonneault, M. H. 2014, ApJ, 780, 159
Foreman-Mackey, D., Montet, B. T., Hogg, D. W., et al. 2015, ApJ, 806, 215
García, R. A., Ceillier, T., Salabert, D., et al. 2014, A&A, 572, A34
García, R. A., Hekker, S., Stello, D., et al. 2011, MNRAS, 414, L6
Howell, S. B., Sobeck, C., Haas, M., et al. 2014, PASP, 126, 398
Huber, D., Stello, D., Bedding, T. R., et al. 2009, CoAst, 160, 74
Kipping, D. M., Bastien, F. A., Stassun, K. G., et al. 2014, ApJL, 785, L32
Lund, M. N., Handberg, R., Davies, G. R., Chaplin, W. J., & Jones, C. D. 2015, ApJ, 806, 30
Mazeh, T., Perets, H. B., McQuillan, A., & Goldstein, E. S. 2015, ApJ, 801, 3
McQuillan, A., Aigrain, S., & Mazeh, T. 2013a, MNRAS, 432, 1203
McQuillan, A., Mazeh, T., & Aigrain, S. 2013b, ApJL, 775, L11
Poppenhaeger, K., & Wolk, S. J. 2014, A&A, 565, L1
Reinhold, T., & Reiners, A. 2013, A&A, 557, A11
Skumanich, A. 1972, ApJ, 171, 565
Smith, J. C., Stumpe, M. C., Van Cleve, J. E., et al. 2012, PASP, 124, 1000
Stumpe, M. C., Smith, J. C., Van Cleve, J. E., et al. 2012, PASP, 124, 985
Vanderburg, A., & Johnson, J. A. 2014, PASP, 126, 948
Vanderburg, A., Montet, B. T., Johnson, J. A., et al. 2015, ApJ, 800, 59

The Astrophysical Journal, 818:109 (9pp), 2016 February 20

Angus, Foreman-Mackey, & Johnson