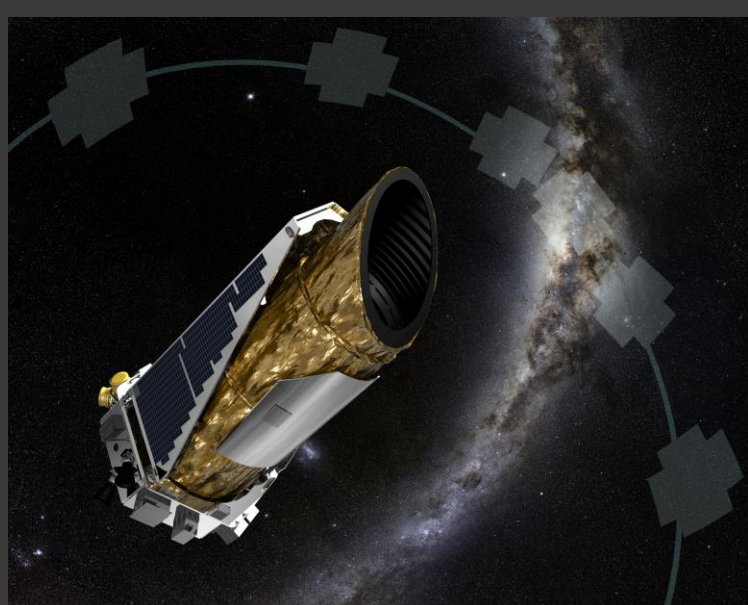# Detection and Removal of Periodic Noise of *Kepler/K2* Photometry with Principal Component Analysis

Riley Clarke, Federica Bianco, and John Gizis
Department of Physics and Astronomy, University of Delaware

Abstract

We present a novel method for detrending systematic noise from time series data using Principal Component Analysis (PCA) in Fast Fourier Transforms (FFT). This method is demonstrated on time series data obtained from Campaign 4 of the Kepler K2 mission, as well as two additional objects of interest. Unlike previous detrending techniques that utilize PCA, this method performs the detrending in Fourier space rather than temporal space. The advantage of performing the analysis in frequency space is that the technique is sensitive purely to the periodicity of the unwanted signal and not to its morphological characteristics. This method could improve measurements of low signal-to-noise photometric features by reducing systematics. We also discuss challenges and limitations associated with this technique.
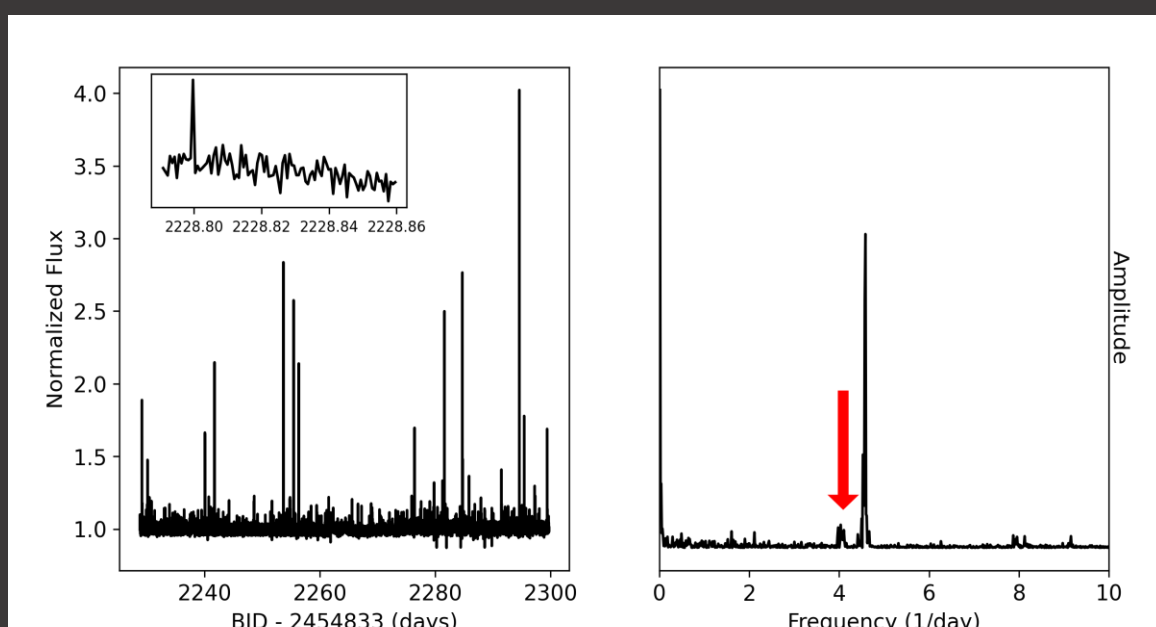
Kepler & Periodic Signals

The *Kepler* Space Telescope (pictured left) was designed to detect periodic astrophysical signals such as:

- Exoplanet transits
- Photospheric variability
- Eclipsing binaries

but the Kepler catalogue includes periodic systematic noise due to scheduled rolling motion every 6 hours. Spectral analysis tools like Fast Fourier Transforms (FFT) can help analyze not only astrophysical periodicity, but systematics as well.
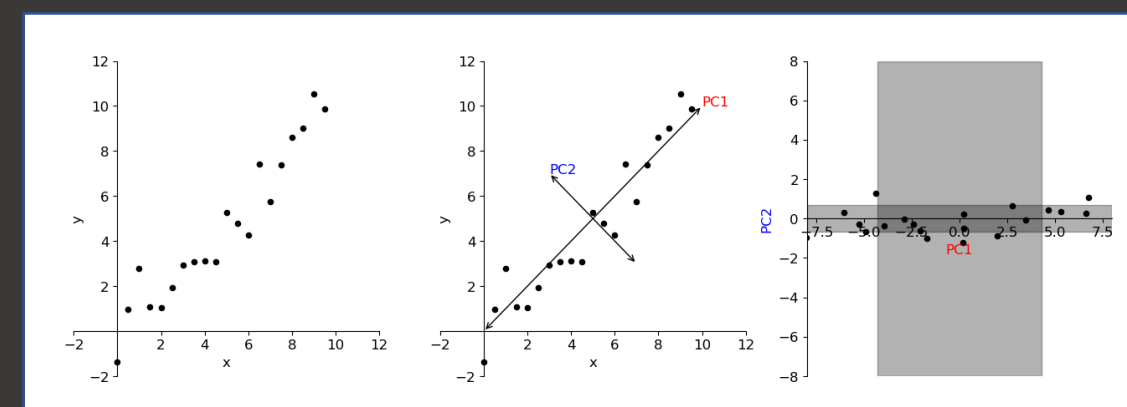

Left: Sample *Kepler* lightcurve of an M7.5 brown dwarf and the conjugate square of its Discrete Fourier Transform (DFT), i.e. the power spectrum (PS). The high-amplitude spike is an astrophysical signal caused by photospheric variability. The smaller spike, indicted by the red arrow, is the systematic *Kepler* roll frequency.

Principal Component Analysis

Principal Component Analysis (PCA), is a dimensionality reduction technique that transforms the data into a new orthonormal coordinate space, the basis vectors of which are the principal components (PCs). The associated eigenvalues of each PC indicate what fraction of a data point, in our case a vector of flux points, is projected along that axis.

- The eigenvalues represent how much of the variance present in the data is explained by each PC.
- The components can be sorted by the amount of data variance explained (i.e. the first component is the one that alone explains the largest fraction of the variance in the data).
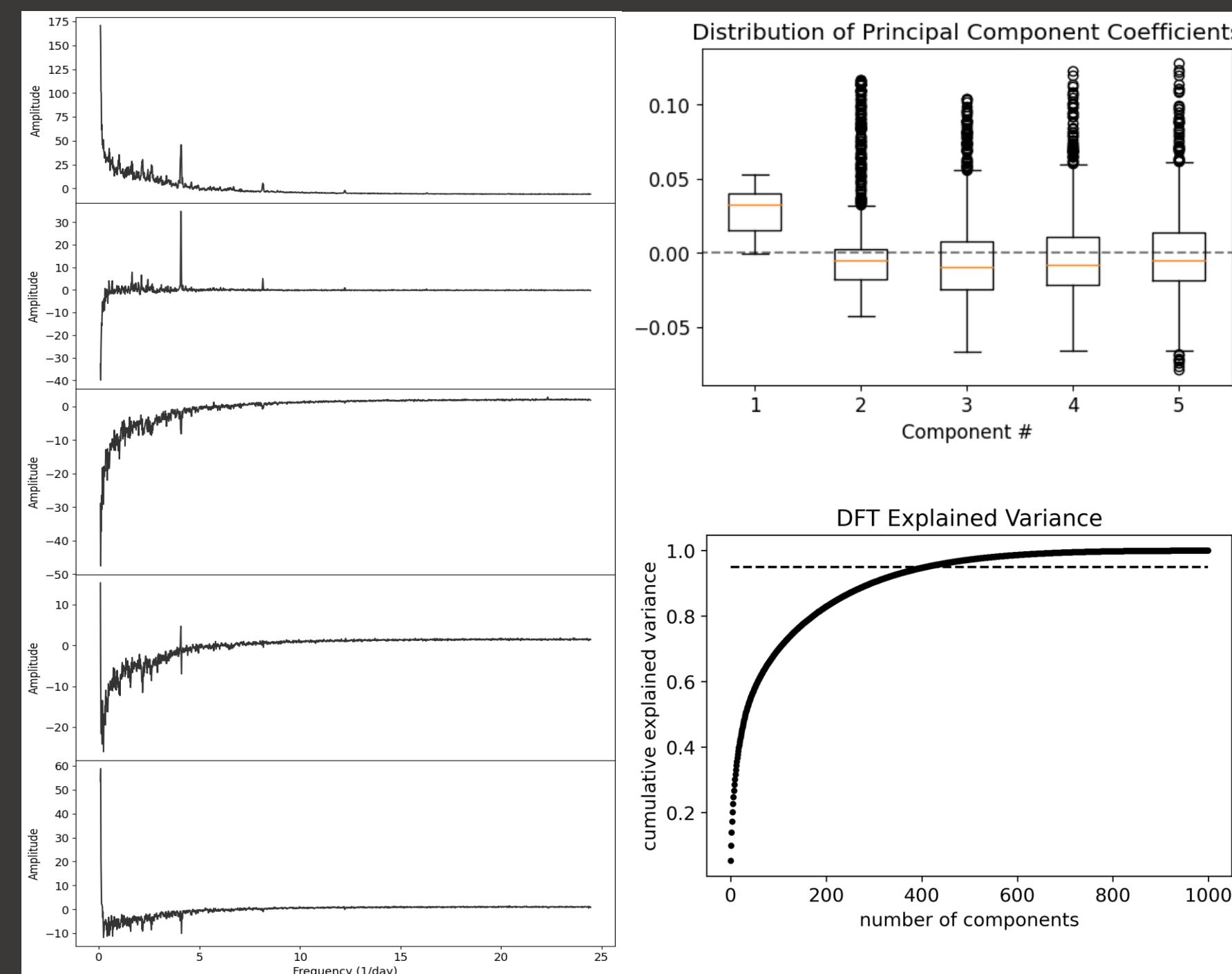

Far left: A synthetic illustrative dataset; Mid left: Principal Components of the data are identified on top of the dataset. Near left: The original data is transformed by PCA. The first component captures the majority of the variance of the dataset. The grey shaded regions indicate the standard deviation along PC1 and PC2 axis (eigenvectors).

To detect and characterize systematic periodic noise in a set of synchronous, evenly sampled lightcurves, the following steps are performed:

- the power spectrum of each lightcurve is obtained by taking the conjugate square of the Fourier Transform (Oliphant 2006),
- a PCA decomposition was generated using the power spectra as inputs,
- high order PCs were inspected.

The first 5 PCs of the power spectra ensemble are shown below in ascending order from top to bottom. The coefficients of the first PC are uniformly positive and non 0 for (almost) all lightcurves, which means that PC-1, with its prominent peak at 4.09 days^{-1}, is required to reconstruct every lightcurve in the dataset, connecting PC-1 to a systematic phenomenon.

Left: First five principal components of the power spectra, shown in ascending order from top to bottom. A sharp peak at a frequency of 4.09 days (5.86 hours) is noticeable in the 1st and 2nd components, and less prominently in all other components shown.
Upper right: Box-and-whisker plots of the coefficients of the PCs of the power spectra of 1000 lightcurves. Each box represents the interquartile range of the distribution of PC coefficients for PC-1 through PC-5, as indicated at the top; the orange line represents the median value, the whiskers the edges of the distribution, and points represent ``outliers''.
Lower right: Cumulative variance of the discrete Fourier transform PCs explained by successive coefficients. The dotted black line is 95% explained variance.

Principal Component Filtering

The systematic periodic signatures in our sample of *Kepler* lightcurves were removed according to Algorithm 1 and some sample results are shown below.

Left: Results of FFTPCA on the first two PCs of the power spectra ensemble. The full PCs are shown on the left, and a zoomed-in region indicated by the grey dashed box is shown on the right.
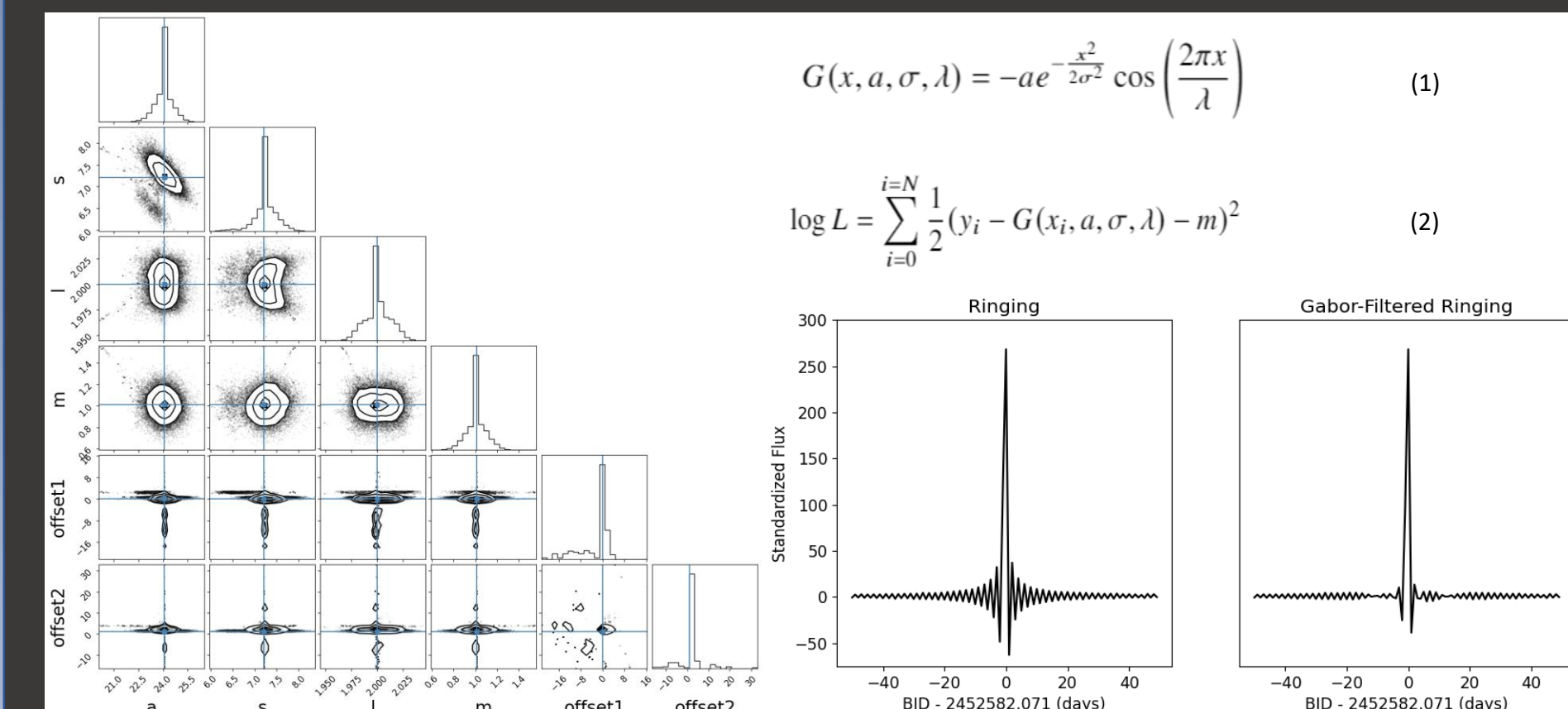
Top Right: Excerpt of a lightcurve with high PC-1 coefficients pre- and post-processing. We note: 1) the signal is largely intact; 2) using only 400 component may have affected the sharpness of signal.

Bottom Right: Power spectra of pre- and post-processed lightcurves shown above

Flares & Ringing Artifacts

Flares are sudden brightening events on a star's photosphere associated with magnetic reconnection events. Discrete Fourier Transforms and bandpass filtering produce oscillatory artifacts in the vicinity of discontinuities, a phenomenon known as "ringing". I designed a method that used a 1D Gabor filter (1), optimized by a Markov Chain Monte Carlo (MCMC) routine (2).

$$G(x, a, \sigma, \lambda) = -ae^{-\frac{x^2}{2\sigma^2}} \cos\left(\frac{2\pi x}{\lambda}\right) \quad (1)$$

$$\log L = \sum_{i=0}^{i=N} \frac{1}{2}(y_i - G(x_i, a, \sigma, \lambda) - m)^2 \quad (2)$$

Left: Corner plot of the posterior distributions of the segmented Gabor filter parameters, where s is the standard deviation of the Gaussian envelope, l is the wavelength of the sinusoid, m is the vertical offset, and ``offset1'' and ``offset2'' are the left and right horizontal offsets, respectively.
Right: Largest flare in EPIC 21032702 before ringing removal beside same flare with ringing reduced using a MCMC-optimized Gabor filter.

References

Cooley, J. W. 1987, Microchimica Acta, 93, 33

Oliphant, T. E. 2006, A guide to NumPy, Vol. 1 (Trelgol Publishing USA)