

## PUSHING THE LIMITS OF DE-TRENDING WITH PLD

NICHOLAS SAUNDERS<sup>1,2,3</sup>, RODRIGO LUGER<sup>1,2</sup>

*Draft version November 27, 2017*

### ABSTRACT

We present **sko**, a **python** package to create a forward model of the *Kepler* Space Telescope detector and simulate stellar targets traversing the CCD. Using simulations, we characterize the contribution of CCD sensitivity variation to the noise of *K2* light curves. We also test methods to increase the effectiveness of existing de-trending pipelines for space telescope exoplanet targets, and focus our treatment on stars with bright neighbors or high motion relative to the CCD. With these simulated light curves, we assess the de-trending power of two methods, first: fitting an aperture around the stellar Point Spread Function (PSF) to maximize light received by the target and exclude contribution from neighbors, and second: fitting a mathematical PSF model to each target in the aperture and subtracting contribution by neighbors. The combination of these two methods increases the recovery accuracy of injection tests by (*some amount*). These methods can be applied to the light curves of *K2* targets for existing and future campaigns, and will become particularly useful as *Kepler* begins to run out of fuel. These methods can also be applied to future space telescope missions, such as *TESS*, to improve detection and characterization of exoplanet candidates.

### 1. INTRODUCTION

Despite the failure of two reaction wheels in 2012 and 2013, the *Kepler Space Telescope* has continued to produce valuable data in its new configuration, *K2*, with significantly higher precision than ground based telescopes (Howell et al. 2014). However, due to the unstable pointing caused by the missing reaction wheels, targets have significant motion relative to the quantum sensitivity variation of the telescope detector, creating noise in *K2* light curves (**source**). A number of attempts have been made to isolate and remove the instrumental noise from *K2* data (Armstrong et al. 2015; Lund et al. 2015; Huang et al. 2015; Aigrain et al. 2015, 2016).

Existing de-trending pipelines, notably the K2SFF pipeline, developed by Vanderburg & Johnson (2014), generate Gaussian fits to the stellar PSF to find the centroid position of targets. From this, motion relative to the detector can be backed out and used to de-trend *K2* light curves. However, factors such as light aberration, pixel sensitivity variation, and imperfect PSF fitting can lead to uncertainty in the central pillar of these models: the motion of the star relative to the detector.

The EPIC Variability Extraction and Removal for Exoplanet Science Targets (EVEREST) pipeline, developed by Luger et al. (2016, 2017) utilizes a different method – pixel level decorrelation (PLD), developed by Deming et al. (2015). PLD seeks to remove noise generated by intra-pixel sensitivity variation independent of apparent motion of the star. This method is particularly effective for *K2* light curves despite the magnitude of apparent motion being high, and EVEREST can recover *Kepler*-like accuracy in exoplanet light curves for targets up to  $K_p = 15$ .

There remain cases in which *K2* de-trending pipelines fail to achieve *Kepler*-like accuracy, particularly high

magnitude stars ( $K_p > 15$ ) and targets with bright contaminants near the edge of the aperture (Luger et al. 2017). For PLD, this is problematic because the algorithm uses pixels containing flux from the contaminant source to de-trend the target light curve. Minimizing or removing contaminant flux from pixels included in the PLD model increases the de-trending power of EVEREST. Crowded apertures are not a problem unique to *Kepler/K2*, and future space telescope missions will benefit from having established methods to approach crowded targets.

Further, as the *Kepler Space Telescope* runs out of fuel, its motion due to thruster fires is expected to become less predictable and the magnitude of targets’ motion relative to the detector will increase. With higher motion, targets will traverse more regions of varied pixel sensitivity, contributing more noise to the light curves of *K2* targets. Eventually, the telescope will enter a phase of constant drift, at which point targets will traverse many pixels across the detector and flux pollution from neighbors is likely, as stellar PSFs may overlap over the course of a campaign.

In this paper, we present options for assessing and removing these sources of noise. §2 and §3 explore the issues of pixel sensitivity variation and crowding, respectively. §4 describes our methods to characterize and address these sources of noise: §4.1 focuses on PSF-fitting, §4.2 on Aperture-fitting, and §4.3 on Motion Removal. In §5, we discuss our results, and finally we draw our conclusions in §6.

### 2. SYNTHETIC LIGHT CURVES

#### 2.1. Pixel Sensitivity Variation

Removal of instrumental noise requires a thorough understanding of its source. Stellar motion relative to the pixel sensitivity variation on the *Kepler* CCD causes fluctuation in the amount of light received by the telescope detector. In order to accurately fit a PSF model to multiple targets on the detector, it became necessary to generate a model for the pixel sensitivity variation of the

<sup>1</sup> Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195, USA

<sup>2</sup> Virtual Planetary Laboratory, Seattle, WA 98195, USA

<sup>3</sup> nks1994@uw.edu

CCD.

We did not set out to model the exact sensitivity of the *Kepler* CCD, rather our goal was to simply characterize its contribution to the noise in *K2* light curves. Light curves generated against this model are adequate to serve as well-understood sample targets on which to test de-trending methods. To accurately represent the *Kepler* CCD, we generated a model for the detector that included both inter-pixel sensitivity variation between pixels and intra-pixel sensitivity variation within each pixel. We found that a stochastic distribution of sensitivity with  $\sim 1\%$  variation between pixels and  $\sim 5\%$  variation within pixels, from center to edge.<sup>4</sup>

### 2.2. PSF Model

A stellar PSF was generated with a characteristic two-dimensional Gaussian shape and with covariance between  $x$  and  $y$  dimensions to capture PSF distortion due to incident light aberration on the *Kepler* detector. We define our mathematical model to be

$$F(t) = \sum_{\text{aperture}} \iint_{\text{pixel}} [s(x)s(y)P(x,y)\tau(t)] dx dy, \quad (1)$$

where  $s(x, y)$  is the sensitivity variation function, modeled by a sum of power functions, and  $P(x, y)$  is the PSF of the star, centered at  $(x_0, y_0)$  with amplitude  $A$ .  $\tau(t)$  is a simulated transit function. The pixel sensitivity model is given by

$$s(x)s(y) = \sum_n a_{x,n} x^n a_{y,n} y^n \quad (2)$$

and the PSF model is given by

$$P(x, y) = \sum_m e^{-\frac{(x-x_{0,m})^2}{2\sigma_{x,m}^2} - \frac{(y-y_{0,m})^2}{2\sigma_{y,m}^2}}. \quad (3)$$

Our model for the sensitivity variation was chosen to capture the same noise magnitude as real *K2* targets. For consistency and comparison purposes, we adopt the Combined Differential Photometric Precision, developed by the *Kepler* team (Christiansen et al. 2012), as our noise metric. In Fig. 1, we plot the CDDP of our simulated light curves with no motion against CDDP of original *Kepler* light curves as a function of  $K_p$  Mag. Our synthetic targets follow the trend of photon-limited noise. This is because our benchmark model has no stellar variability, which can add extraneous noise to de-trended light curves.

In Fig. 2, we inject motion from a real *K2* target (EPIC 205998445) to benchmark our model against *K2* data. The noise of our various tests bracket the trend of raw *K2* noise, again following the photon-limited trend due to no variability.

The coefficients  $a_n$  were determined to emulate the magnitude of sensitivity variation on the *Kepler* CCD based on its contribution to the noise in *K2* light curves.

To best capture the characteristics of *K2* data when analyzing the noise contributed by quantum sensitivity

<sup>4</sup> More tests are needed to determine these values with higher certainty, these are just the values we ended up using in our model. Rodrigo - any ideas?

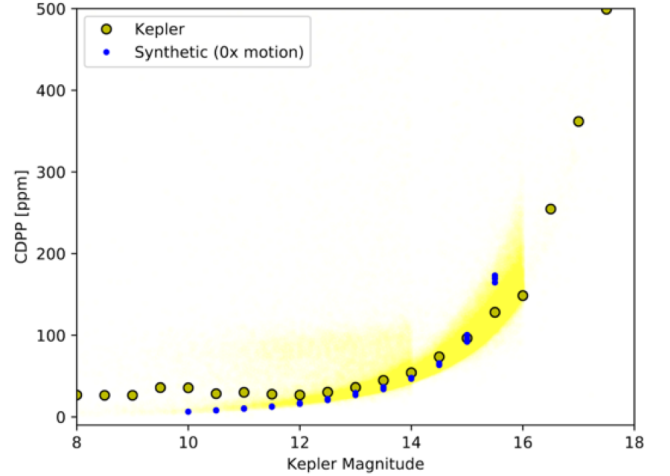


FIG. 1.— CDPP of our simulated light curves with no motion compared to original *Kepler* targets as a function of  $K_p$  Mag.

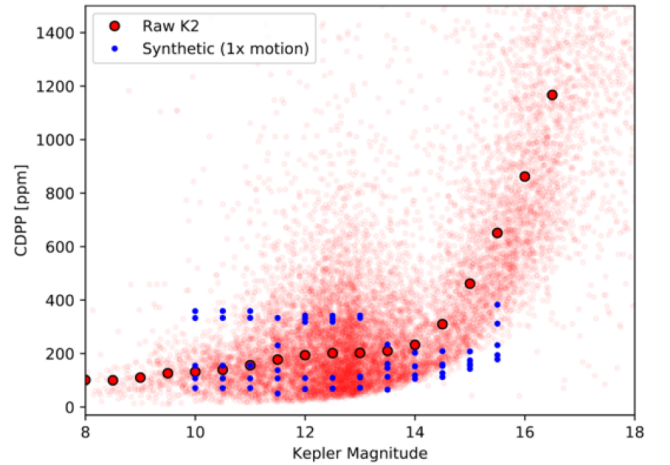


FIG. 2.— CDPP of our simulated light curves with *K2* motion compared to real *K2* targets as a function of  $K_p$  Mag.

variation, our model also included two sources of noise – photon and background noise. (*More here about the sources of noise.*)

This is the same model we fit to the simple aperture photometry (SAP) flux of *K2* targets to subtract contaminant flux (further discussion in §4.1).

### 2.3. Crowding

A significant stumbling block of PLD arises when *K2* targets are observed in crowded fields. In these cases, PLD uses pixels containing flux from neighbors to de-trend the target light curve, and the astrophysical signal or transit signal can be lost.

We define a crowding metric,  $C$ , as a quantitative measure of the significance of neighbor pollution in each pixel.

$$C = 1 - \frac{F_{\text{target}}}{F_{\text{total}}} \quad (4)$$

Applications of PLD have demonstrated that it is not well suited to handle targets with high values of  $C$  in pixels containing flux from the target star. These cases

suffer from overfitting at a higher rate and transit recovery accuracy is compromised significantly. A sample target (EPIC 202072978) and the crowding metric  $C$  for each pixel are shown in Fig. 2. This plot was achieved by simultaneously fitting PSFs to the central star and nearby contaminant, then calculating  $C$  for each pixel from equation (3).

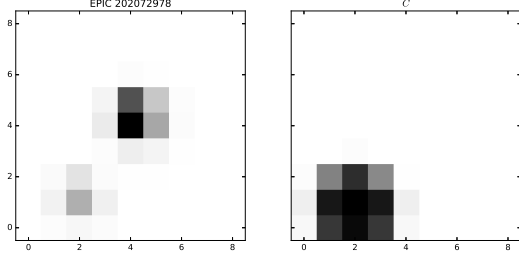


FIG. 3.— A single cadence from target EPIC 202072978, a 12<sup>th</sup> magnitude exoplanet host and its 16<sup>th</sup> magnitude neighbor (left), and the crowding metric  $C$  in each pixel for this cadence, as defined in Eq. 4 (right).

#### 2.4. High Roll

To test how PLD performs on targets with high motion relative to the detector, we inject motion vectors with various coefficients into our synthetic models. A coefficient of 1 corresponds to physical  $K2$  motion (maximum of  $\sim 0.56$  pixels), while a coefficient of 20 results in motion of up to 11.2 pixels.

Motion Coefficient	Physical Motion (pixels)	$\sigma$ (pixels)
1	0.56	0.16
2	1.12	0.32
5	2.8	0.8
10	5.6	1.6
20	11.2	3.2

TABLE 1

MOTION COEFFICIENTS AND CORRESPONDING PIXEL MOTION.

The resulting raw light curve noise can be seen in Fig. 4 compared to raw  $K2$  noise. Our synthetic targets with a motion coefficient of 1 align with the raw  $K2$  trend and increase corresponding to noise coefficient.

### 3. METHODS

The two sources of noise presented above – pixel crowding and detector sensitivity variation – have been obstacles for  $K2$  de-trending pipelines. In the following section, we present methods to address both sources of noise and recover more accurate light curves.

#### 3.1. PLD

##### more detail here

To de-trend our synthetic light curves, we used a variant of the PLD method applied in the EVEREST pipeline. We included a Gaussian Process (GP) to capture and retain the astrophysical signal of the synthetic host star.

The likelihood ( $\mathcal{L}$ ) of our model capturing the true signal of the star is captured closely by a Gaussian distribution given by

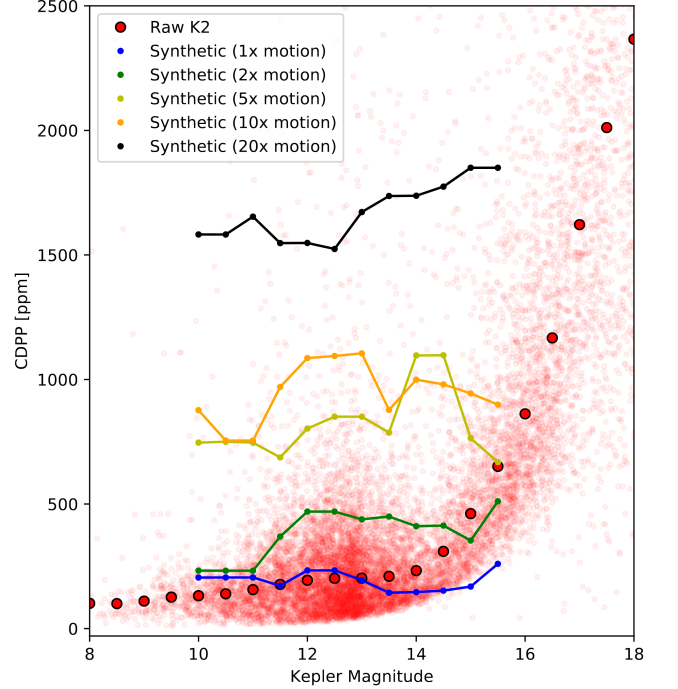


FIG. 4.— CDPP of synthetic targets with 1x to 20x motion versus raw  $K2$  noise as a function of  $K_p$  Mag.

$$\mathcal{L} = \exp \left[ - \left( \frac{\mathbf{y} - \mathbf{X} \cdot \mathbf{w}}{2\mathbf{\Sigma}} \right)^2 \right],$$

where  $\mathbf{y}$  is the raw flux light curve,  $\mathbf{X}$  is the design matrix,  $\mathbf{\Sigma}$  is the covariance matrix between pixels, and  $\mathbf{w}$  is the array of weights. By solving for the derivative of the natural log of this likelihood, we locate the distribution peak and the best values for weights on the pixel flux vectors.

$$\ln \mathcal{L} = -\frac{1}{2} (\mathbf{y} - \mathbf{X} \cdot \mathbf{w})^\top \cdot \mathbf{\Sigma}^{-1} \cdot (\mathbf{y} - \mathbf{X} \cdot \mathbf{w})$$

$$\frac{d \ln \mathcal{L}}{d\mathbf{w}} = 0$$

$$= \frac{d}{d\mathbf{w}} \left[ -\frac{1}{2} (\mathbf{y} - \mathbf{X} \cdot \mathbf{w})^\top \cdot \mathbf{\Sigma}^{-1} \cdot (\mathbf{y} - \mathbf{X} \cdot \mathbf{w}) \right]$$

Now, we identify the ideal weights for our noise model.

$$\mathbf{w} = \left( \mathbf{X}^\top \cdot \mathbf{\Sigma}^{-1} \cdot \mathbf{X} \right)^{-1} \cdot \mathbf{X}^\top \cdot \mathbf{\Sigma}^{-1} \cdot \mathbf{y}$$

The final model can be found by solving

$$\mathbf{m} = \mathbf{X} \cdot \left( \mathbf{X}^\top \cdot \mathbf{\Sigma}^{-1} \cdot \mathbf{X} \right)^{-1} \cdot \mathbf{X}^\top \cdot \mathbf{\Sigma}^{-1} \cdot \mathbf{y}.$$

#### 3.2. PSF-fitting

For some targets, particularly those with exceptionally close neighbors, defining an aperture around the central star can be difficult, because pixels containing flux pollution also contain important stellar signal. Exclusion of

pixels above the prescribed crowding cutoff would render the remaining flux inadequate to thoroughly analyze transiting exoplanet targets.

We present a second method to remove unwanted neighbor signal: fitting each target in the frame with a Gaussian PSF model and subtracting the flux contributed by the neighbor from each pixel. This process involves constructing sums of  $n$  PSFs, where  $n$  is the number of stars within the aperture, and minimizing  $\chi^2$  between the model,  $m$  and flux,  $y$ .  $\chi^2$  is given by

$$\chi^2 = \left( \frac{m - y}{\sigma} \right)^2, \quad (5)$$

Where  $\sigma$  is the flux error. Our model, defined in equation (2), was fit to each cadence of crowded *Kepler* targets.

### 3.3. Aperture-fitting

The simplest solution to the issue of crowded apertures is to only consider pixels that contain the most flux from the desired target and exclude flux received by neighboring stars. The crowding metric in each pixel can be used to determine which pixels should be included to maximize target flux. *(I need to give this a more thorough treatment to determine exactly what value of  $C$  is best to include within an aperture. To do so, I will perform aperture tests on a variety of targets and maximize injected transit recovery accuracy to constrain ideal values of  $C$ .)* In a process we refer to as aperture PLD (aPLD), we define an aperture based on the crowding metric in each pixel to maximize target flux.

Injection tests were performed by removing noise from simulated light curves with aPLD and measuring the transit depth of the de-trended light curve. Comparing this value to the injected depth, we determine the effectiveness of aPLD and constrain useful values of  $C$ . With our injection tests, we find that pixels with  $C > 0.3$  should be excluded to maximize the effectiveness of aPLD.

### 3.4. Motion Removal

$$x_0 = X(c) - v_x(c)$$

$$y_0 = Y(c) - v_y(c)$$

## 4. RESULTS

### 4.1. Motion Tests

When de-trended with second order PLD, our synthetic light curves align with the noise of original *Kepler* observations for motion coefficients of 1x, 2x, and 5x. 10x motion shows increased noise in the de-trended light curve, but still performs better than observations from the ground.

### 4.2. Injection Tests

## 5. FURTHER CONSIDERATIONS

### 5.1. Focus Changes

**TESS stuff, breathing apertures**

### 5.2. Seeing

**APO data**

## 6. CONCLUSION

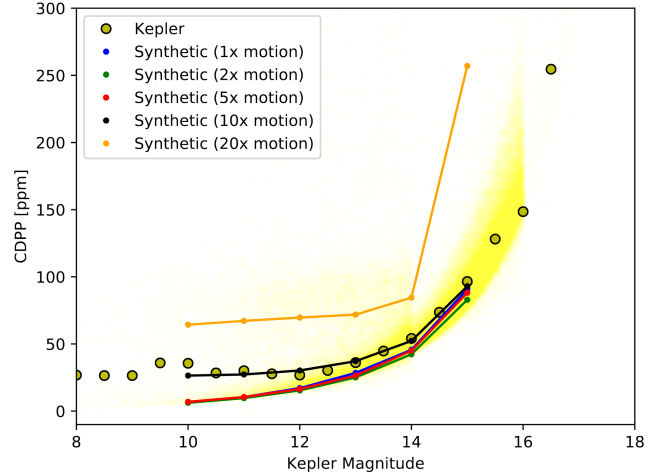


FIG. 5.— De-trended CDPP for synthetic targets with 1x to 20x motion versus original *Kepler* noise as a function of  $K_p$  Mag.

## REFERENCES

- Aigrain, S., Hodgkin, S. T., Irwin, M. J., Lewis, J. R., & Roberts, S. J. 2015, *MNRAS*, 447, 2880
- Aigrain, S., Parviainen, H., & Pope, B. J. S. 2016, *MNRAS*, 459, 2408
- Armstrong, D. J., Kirk, J., Lam, K. W. F., et al. 2015, *A&A*, 579, A19
- Christiansen, J. L., Jenkins, J. M., Caldwell, D. A., et al. 2012, *PASP*, 124, 1279
- Deming, D., Knutson, H., Kammer, J., et al. 2015, *The Astrophysical Journal*, 805, 132.  
<http://stacks.iop.org/0004-637X/805/i=2/a=132>
- Howell, S. B., Sobeck, C., Haas, M., et al. 2014, *PASP*, 126, 398
- Huang, C. X., Penev, K., Hartman, J. D., et al. 2015, *MNRAS*, 454, 4159
- Luger, R., Agol, E., Kruse, E., et al. 2016, *AJ*, 152, 100
- Luger, R., Kruse, E., Foreman-Mackey, D., Agol, E., & Saunders, N. 2017, *ArXiv e-prints*, arXiv:1702.05488
- Lund, M. N., Handberg, R., Davies, G. R., Chaplin, W. J., & Jones, C. D. 2015, *The Astrophysical Journal*, 806, 30.  
<http://stacks.iop.org/0004-637X/806/i=1/a=30>
- Vanderburg, A., & Johnson, J. A. 2014, *PASP*, 126, 948