

# Robust inference of sensor-spatially correlated systematic noise in Kepler lightcurves\*

JAMILA S. TAAKI <sup>1</sup>, ATHOL J. KEMBALL <sup>2</sup> AND FARZAD KAMALABADI<sup>1</sup>

<sup>1</sup>*Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign  
306 N. Wright St. MC 702, Urbana, IL 61801-2918*

<sup>2</sup>*Department of Astronomy, University of Illinois at Urbana-Champaign  
1002 W. Green Street, Urbana, IL 61801-3074*

(Received June 1, 2019; Revised January 10, 2019; Accepted January 10, 2023)

Submitted to AJ

## ABSTRACT

Lightcurves produced by wide field transit surveys such as CoRoT, Kepler and TESS are affected by sensor-wide systematic noise, of a priori unknown form. The signal strength of systematic noise typically far exceeds that of, and conceals, astrophysical signals of interest. Thus, methods to remove correlated instrumental noise are necessary to produce accurate photometric data. In this work we develop a method for robust inference of systematic noise affecting lightcurves. Low rank models are well suited to data-driven inference of systematics, however without constraints, estimates are poorly conditioned and thus susceptible to overfitting of astrophysical signals. We use a low rank model with a total variation based prior which encourages correlation among systematic estimates for targets which are spatially close on the sensor. This constraint is introduced foundationally, such that the full systematics model is solved for based on the spatial correlation prior. Herein we fit this model using first order methods. Our model is investigated and evaluated on Kepler lightcurves with simulated injected astrophysical signals. Our method is found to be more robust to overfitting as compared to standard low-rank inference via PCA.

## 1. INTRODUCTION

Transit surveys have contributed significantly to exoplanet science over the past two decades, uncovering exoplanet population statistics at current instrumental completeness limits (Batalha 2014; Bryson et al. 2020). These wide-field surveys include a large number of candidate systems and require a residual photometric precision adequate to detect exoplanet transits (Deeg & Alonso 2018). Space-based surveys, including CoRoT (Auvergne et al. 2009), Kepler (Borucki et al. 2010), and TESS (Ricker et al. 2014) have led to the discovery of over 5100<sup>1</sup> exoplanets to date. Low residual photometric noise requires the optimization of both the instrument design and the associated algorithms for instrumental noise removal and exoplanet transit detection. In this paper we describe an algorithm for addressing spatially-correlated systematic noise across a wide-field imaging sensor used for exoplanet transit detection.

If an exoplanet transit induces a fractional flux density variation  $\Delta F$  in observations, with a differential photometric precision  $\sigma_p$  (after processing) then the signal-to-noise ratio is  $S/N \sim \frac{\Delta F}{\sigma_p}$  (Deeg & Alonso 2018). For example, an Earth-Sun transit requires  $\sigma_p \sim 80$  ppm over the transit duration  $T \sim$  hours (Caldwell et al. 2010). The Kepler telescope achieved  $\sigma_p \sim 30$  ppm over  $T = 6.5$ h for stars with Kepler magnitude  $K_p \sim 12$  (Koch et al. 2010; Gilliland et al. 2011; Christiansen et al. 2012). TESS achieved  $\sigma_p \sim 230$  ppm over  $T \sim 1$  hr for star with TESS magnitude  $T_p \sim 10$  which is sufficient to detect super-Earths around bright stars (Fausnaugh 2018). For reference, the Hubble Space Telescope Space Telescope Imaging Spectrograph (HST/STIS) can attain  $\sigma_p \sim 120$  ppm over  $T = 45$  min (Demory et al. 2015) and non wide-field ground-based telescopes can reach a precision sufficient to detect large Jovian

Corresponding author: Jamila Taaki  
xiazinya@gmail.com

\* Released on June, 10th, 2019

<sup>1</sup> <https://exoplanetarchive.ipac.caltech.edu/>

exoplanets (Tregloan-Reed & Southworth 2013; Stefansson et al. 2017).

The instrumental response will vary over a range of timescales and across various spatial scales at this level of sensitivity (McLean 2008) producing systematic noise effects. Algorithmic innovation to suppress the net residual  $\sigma_p$  arising from unmodeled systematic noise effects therefore is necessary to detect weak exoplanet signals. These effects have diverse physical origins (Jenkins 2017; Tenenbaum & Jenkins 2018) including inter- and intra-pixel sensitivity variation (Toyoizumi & Ashley 2005; Hedges et al. 2021; Smith et al. 2016), a spatially-variant optical PSF and pixel response function (PRF) (Bryson et al. 2010), pointing jitter, thermal events, differential aberration, spacecraft guidance and downlink control events, CCD detector pattern, read, photon, quantization noise (McLean 2008; Gilliland et al. 2011; Caldwell et al. 2010), and saturation errors (Van Cleve & Caldwell 2016; Vanderspek 2018) amongst others. The first correction for noise effects is typically applied in early robust calibration pipelines, Kepler SAP lightcurves are derived using optimized simple aperture photometry (SAP), however residual spatially- and time-variant systematics are present.

The residual systematic noise across the sensor (or regions of the sensor) can effectively be regarded as having shared instrumental origins and are therefore appropriately modeled using low-rank linear models. A low-rank linear basis is either based on time-series of a parameter known to govern noise effects, or derived from a set of contemporaneous observations. The low-rank linear basis is fit against a target time-series, to obtain relative coefficient weightings scaling each basis term, the sum of which is subtracted from the target time-series to remove systematic effects (Kinemuchi et al. 2012). However, the target time-series contains unknown astrophysical variability which although may be statistically uncorrelated with systematics, may be correlated with basis terms and overfit, whether because of i) incidental correlation between systematics and astrophysics (Smith et al. 2018) or ii) the basis terms contain astrophysical content. Further ancillary systematic assumptions can therefore improve performance.

Petigura & Marcy (2012) show an example of spatially-varying residual systematic noise in Kepler data, Moreno et al. (2021) further characterize spatial systematics in Kepler/K2 data. Detrending methods may incorporate spatial structure in the systematic noise models in several forms. In the simplest form, the spatial variation constraint is imposed implicitly by detrending targets across a discrete sensor region as a whole (e.g. a Kepler CCD output channel (Luger et al. 2018; Lund et al. 2021; Roberts et al. 2013; Stumpe et al. 2012; Smith et al. 2012; Stumpe et al. 2013). Greater sensitivity to spatial variation has been incorporated by restricting basis terms by proximity to the target lightcurve, (Luger et al. 2018; Lund et al. 2021), or directly introducing a parametric dependence of fitted systematics by sensor position used in PDC-MAP Smith et al. (2012), MS-MAP Stumpe et al. (2013), and SARS Ofir et al. (2010) methods. We review various detrending methods to explain the context of these methods that incorporate spatial structure, applied to pixel data and lightcurves.

Pixel-level detrending generally describes systematics models based on pixel data, however may be applied to detrend either pixel-data or lightcurves. These methods correct extreme pointing error in K2 data, a more fine-grained characterization of which can be obtained from pixel data than aperture reduced photometry. In Deming et al. (2015) Pixel Level Decorrelation (PLD) pixel time-series from within the target aperture is used to detrend the target lightcurve, Luger et al. (2016); Luger et al. (2018) use PLD, in addition to pixel time-series from nearby targets on the sensor to correct a lightcurve. The Causal Pixel Model Wang et al. (2016); Hattori et al. (2022) corrects pixels within a target aperture prior to aperture photometry, using distant pixel data from external targets. The systematic pointing-error over time can be directly estimated using pixel data as in Huang et al. (2015); Aigrain et al. (2015); Lund et al. (2015); Vanderburg & Johnson (2014); Crossfield et al. (2015), in Huang et al. (2015) the pointing is estimated using nearby targets on the same module.

Herein our method infers systematics for a set of lightcurves collectively. The Trend-Filtering-Algorithm Kovács et al. (2005) uses a uniform selection of external lightcurves to individually detrend a target lightcurve. Lightcurves contains astrophysical content, therefore instead of directly using lightcurves, a derived basis representative of the net systematic content is often obtained and fit individually to lightcurves. The SARS algorithm Ofir et al. (2010) based on Sysrem Tamuz et al. (2005) used for detrending CoRoT light curves, introduces a parametric mathematical model for systematics, which incorporates scaling by external magnitude or spatial position. A least-squares residual between

their systematics model, and observations, is minimized to obtain fitted basis effect terms and various coefficient values.

The truncated singular value decomposition (SVD) or a principal component analysis (PCA) methods are optimal maximum-likelihood estimates of a low-rank linear structure in white noise (due to [Eckart & Young \(1936\)](#)); equivalently minimizing the least-squares residual between the estimated low-rank model and the lightcurves from which the model is derived, among all possible equivalent rank models. A PCA/SVD decomposition always exists and can be exactly computed. To this end, PCA/SVD are often used to obtain a systematics basis ([Tamuz et al. 2005](#); [Thatte et al. 2010](#); [Stumpe et al. 2012](#); [Foreman-Mackey et al. 2015](#); [Taaki et al. 2020](#)).

However, as with any least-squares based estimator, PCA/SVD are susceptible to overfitting of astrophysical signals ([Candès et al. 2011](#)) because astrophysical signals are not appropriately modeled as white noise [Pont et al. \(2006\)](#). To minimize the presence of astrophysical signals in derived basis vectors, the set of lightcurves used to construct the set may be heavily filtered; in [Petigura & Marcy \(2012\)](#) a set of lightcurves which are highly correlated are used to estimate a refined basis using PCA. In the Kepler Presearch Data Conditioning - Maximum-a-posteriori (PDC-MAP) [Stumpe et al. \(2012\)](#); [Smith et al. \(2012\)](#) cotrending basis vectors (CBV) are constructed based on SVD, where the set of lightcurves it is applied on is iteratively refined with an entropy metric.

Furthermore in PDC-MAP Bayesian priors are used to fit CBV's to reduce overfitting due to incidental correlation between an astrophysical signal and the CBV's. The Bayesian priors are empirical and formed over the weights as a function of sensor position and target brightness. Following the development of PDC-MAP, Kepler data was reprocessed using the Multiscale - Maximum-a-posteriori method (MS-MAP) [Stumpe et al. \(2013\)](#). An obtained SVD basis set is orthogonal and therefore not necessarily representative of separate noise effects. This can make conditioning weights difficult, since the weights may not be physically interpretable or show a clear variable dependence. This issue was identified and corrected in MS-MAP by bandsplitting systematics into different timescales, i.e. short timescale thermal unsettling events or longer term quarter to quarter variability. Astrophysically Robust Correction (ARC) [Roberts et al. \(2013\)](#); [Aigrain et al. \(2017\)](#) is closely related to PDC-MAP, with a Bayesian low-rank systematics model, however additionally ARC models a stellar signal in their lightcurve model.

In this paper we consider an exploratory method for systematics inference applied to lightcurves. We propose a low-rank linear systematics model with a least-squares penalty between lightcurves and systematics, alongside a prior on the systematics model, which promotes correlation of systematics between nearest neighbours on the sensor. We iteratively solve for the complete systematics model under the spatially correlated prior, jointly for all lightcurves.

Our form of spatial correlation prior is based on the total variation functional ([Rudin et al. 1992](#)) applied to normalized coefficient weight vectors. The total variation functional places an  $L1$  penalty on the the difference of spatially neighbouring weight vectors, for all such pairs on the sensor, herein we allow the penalty norm to vary between  $[1, 2]$ . This allows flexibility to control the degree to which the prior is Laplacian or Gaussian distributed. While certain systematic effects may be smoothly varying throughout a sensor, there can be sharp discontinuities in the magnitudes of effects on edges of modules. In this case the  $L1$  penalty is more suitable, obtaining solutions with uniform surfaces and a few sudden discontinuities. When the  $L2$  penalty is used, minimizing the prior is exactly equivalent to maximizing the normalized correlation of neighbouring target weights and more smoothly varying solutions are obtained. See the monograph by [Vogel \(2002\)](#). Further weightings are used to tune the strength of the prior for each lightcurve, allowing the strength to be reduced if a region of the sensor (i.e. certain modules) exhibit only weak systematic correlations. The prior formulated in this way, allows for automated inference of the spatial dependence.

To solve for the systematics, a combined penalty consisting of the least-squares residual on the low rank linear model subtracted from lightcurves, and the spatial smoothness prior on the model coefficient weight vectors is minimized. We use variable elimination ([Golub & Pereyra 1973](#)) to reformulate our optimization problem as a function of the model weights, introducing stability to the minimization ([Golub & Pereyra 2003](#); [Shearer & Gilbert 2013](#)). An approximate closed form gradient of the objective function exists which we minimize via gradient descent. In PDC-MAP and related detrending approaches, a basis set is estimated and fit in separate stages, whereas in our approach and methods such

as ARC and SARS, model components are jointly and iteratively solved for.

Our method is investigated in the context of the Kepler sensor. Systematics exhibit magnitude dependency, as observed in Kepler lightcurves (Smith et al. 2012) as well as CoRoT lightcurves (Mazeh et al. 2009; Ofir et al. 2010). This dependency is likely due to the pixels of similar brightness targets experiencing saturation simultaneously. Kepler operated in the magnitude band  $K_p = 9$  to 15 (Koch et al. 2010). In this work, we do not include a prior for magnitude dependency and select lightcurves from a fixed magnitude band  $K_p \in 12.5$  to 13.5, to limit magnitude based variations in systematics.

The prior favours solutions where systematics are spatially correlated, and allows the astrophysical component of lightcurves to be ‘ignored’ in the estimation. We apply our method to Kepler simple-aperture-photometry (SAP) long cadence (30-minute) lightcurves. We perform simulations where we inject a small subset of lightcurves with simulated astrophysical signals and compare the effect of overfitting between our method and PCA. Additionally we simulate lightcurves with spatially correlated systematics and simulated astrophysical signals, our method is demonstrated on these lightcurves alongside PCA. These experiments demonstrate the utility of our method in cases where overfitting occurs, however we do not show general optimality of our method. Performance is evaluated using standard detrending metrics (6 hr combined photometric differential precision, goodness-of-fit). Alongside these metrics, we evaluate the correlation between injected components and estimates. We find that our method achieves lower levels of overfitting of injected astrophysical signals as compared to PCA/SVD. In particular simulated astrophysical signals on longer timescales (e.g. sinusoidal variability) were more likely to be overfit by least-square methods, and corrected by our method.

The paper is organized as follows. Section 2 introduces the light curve signal model, our method for spatial systematics inference, an analysis of spatial correlation in Kepler lightcurves and describes two experimental tests to analyse the performance of our method. In Section we report the results from our experimental tests with Kepler lightcurves. Section 4 is a discussion of the results, and conclusions are presented in Section 5.

## 2. METHODS

The Kepler and K2 missions collected (30-minute) and (2-minute) pixel data within a predefined aperture for targets<sup>2</sup>. TESS obtained (2-minute) and (20-second) target pixel files, as well as (30-minute) full frame images. Lightcurves are constructed by summing pixels within an aperture from the target pixel files. Lightcurves produced from Kepler/K2/TESS target pixel files are termed Simple Aperture Photometry (SAP). Systematic estimation methods described herein are applied to (30-minute) SAP lightcurves produced by the Kepler pipeline. For an overview of pipeline data-products see Jenkins (2017); Tenenbaum & Jenkins (2018).

### 2.1. Lightcurve Decomposition

A collection of lightcurves  $\{\mathbf{y}_i : i \in I\}$  are obtained on a sensor, each lightcurve  $\mathbf{y}_i$  is length  $N$  time-series and  $I$  is the index set of obtained lightcurves. Our lightcurve model approximates processed observations as the additive combination of a systematics term  $\mathbf{l}_i$  and a statistical noise term  $\mathbf{n}_i$ .

$$\mathbf{y}_i = \mathbf{l}_i + \mathbf{n}_i \quad (1)$$

The above model is compactly described in matrix form  $\mathbf{Y} = \mathbf{L} + \mathbf{N}$ , where the collection of lightcurves  $\{\mathbf{y}_i\}$  form the columns of matrix  $\mathbf{Y} \in \mathbb{R}^{N \times I}$  and systematics and  $\mathbf{L} \in \mathbb{R}^{N \times I}$  and statistical noise  $\mathbf{N} \in \mathbb{R}^{N \times I}$  of each lightcurve respectively.

#### 2.1.1. Noise Model

The predominant components of lightcurves are systematics, astrophysical signals and statistical noise. Statistical noise originates due to a mixture of instrumental error and astrophysical variability. Typical sources of instrumental error include those due to pixel statistics (i.e. readout error, thermal noise and Poisson noise on the source and

<sup>2</sup> Data is termed target pixel files

sky count), these are reduced but not completely removed by aperture photometry. For the Kepler/TESS sensor, stellar variability is the most significant source of statistical noise (Gilliland et al. 2011). Stellar targets may exhibit short-timescale (minutes) stochastic variability, as well as longer term oscillations (Conroy et al. 2018). Additionally the noise level of variability ranges greatly, approximately half of Kepler targets are more variable than the sun (McQuillan et al. 2012).

Herein, subject to pre-processing, non-systematic noise  $\mathbf{n}_i$  is modelled as white Gaussian noise  $\mathbf{n}_i \sim \mathcal{N}(0, \sigma_i^2)$ . A white noise model is commonly used to approximate non-systematic lightcurve components in inference of systematic noise modes, and is an implicit assumption in least-squares minimization (Smith et al. 2012; Stumpe et al. 2012, 2013; Aigrain et al. 2017). A white noise model is a reasonable assumption for additive noise components (stellar and instrumental) due to the central limit theorem (Grinstead & Snell 2012). The noise level  $\sigma_i$  is apriori unknown, however reasonably estimated from filtered and coarsely detrended lightcurves. For mathematical ease, hereonafter lightcurves are normalized by  $\sigma_i$  so that the noise matrix is distributed as  $\mathbf{N} \sim \mathcal{N}(0, \mathbf{1})$ .

There are a number of non-statistical sources present in lightcurves of significant amplitude, i.e. oscillatory stellar variability, transit signals and outlier instrumental errors. A-priori there does not exist a simple model that can capture the variety of non-statistical sources. Wherever possible, lightcurves are filtered to remove clear occurrences of non-systematic behaviour, as discussed below in Section 2.3. However it is clearly not possible to remove all such unknown signals. This issue directly motivates our method, wherein systematics are constrained using prior knowledge to avoid overfitting unmodelled signal components.

### 2.1.2. Systematics Model

*Low rank model:*—The origin of noise specific to transit instruments is generally well understood; for an instrument level description of noise sources see (Auvergne et al. (2009); Van Cleve & Caldwell (2016); Vanderspek (2018)). The affect of Instrument noise sources on lightcurves is complex and has no simple generating model, due to common physical origin systematic error is highly correlated between lightcurves and is therefore well approximated from inference on lightcurve data, using a low rank linear model of the form:

$$\mathbf{l}_i[n] = \sum_{k=1}^K c_k^i \mathbf{v}_k[n] \quad (2)$$

Where ( $K \ll N$ ) describes the rank of the noise model,  $\{\mathbf{v}_k : k \in K\}$  are a set basis vectors shared by lightcurves, and each  $c_k^i$  is a coefficient weighting of  $\mathbf{v}_k$  for lightcurve  $i$ . We denote the vector of coefficients for a lightcurve as  $\mathbf{c}_i = [c_1^i, c_2^i, \dots, c_K^i]^T$ , together these form the set  $\{\mathbf{c}_i : i \in I\}$ . A compact representation for the collection of systematics is given by:

$$\mathbf{L} = \mathbf{V}\mathbf{C} \quad (3)$$

where the columns of  $\mathbf{L} \in \mathbb{R}^{N \times I}$  are the systematics in each lightcurve  $\{\mathbf{l}_i\}$ , the columns of  $\mathbf{V} \in \mathbb{R}^{N \times K}$  are the basis vectors  $\{\mathbf{v}_k\}$  and the columns of  $\mathbf{C} \in \mathbb{R}^{K \times I}$  are the coefficients for each lightcurve  $\{\mathbf{c}_i\}$ .

$$\begin{bmatrix} | & | & & | \\ \mathbf{l}_1 & \mathbf{l}_2 & \dots & \mathbf{l}_I \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_K \\ | & | & & | \end{bmatrix} \begin{bmatrix} | & | & & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_I \\ | & | & & | \end{bmatrix} \quad (4)$$

Additionally we define the column normalized coefficient matrix  $\bar{\mathbf{C}}$  where the columns are formed of terms  $\bar{\mathbf{c}}_i = \frac{\mathbf{c}_i}{\|\mathbf{c}_i\|_2}$ :

$$\bar{\mathbf{C}} = \begin{bmatrix} | & | & & | \\ \frac{\mathbf{c}_1}{\|\mathbf{c}_1\|_2} & \frac{\mathbf{c}_2}{\|\mathbf{c}_2\|_2} & \dots & \frac{\mathbf{c}_I}{\|\mathbf{c}_I\|_2} \\ | & | & & | \end{bmatrix} \quad (5)$$

The rank of  $\mathbf{L}$  does not necessarily represent the number of independent noise sources, there can be a greater number of noise sources than the rank. Since  $\text{rank}(\mathbf{V}\mathbf{C}) \leq \min\{\text{rank}(\mathbf{V}), \text{rank}(\mathbf{C})\}$ , it is possible to have a more expansive basis set  $\mathbf{V}$  representing many individual noise effects, but for the relative weightings  $\mathbf{C}$  between lightcurves to have a degenerate and therefore low-rank structure and consequently for  $\mathbf{L}$  to be low rank.



*SVD and PCA:*—A brief overview of the Singular Value Decomposition (SVD) and Principle Component Analysis (PCA) is provided as they are frequently referenced throughout and motivate the form of model used in our work. For a complete reference see the monographs (Chapter 14.5 [Hastie et al. \(2009\)](#), [Golub & Van Loan \(2013\)](#)). Every matrix  $\mathbf{Y} \in \mathbb{R}^{n \times m}$  admits an SVD:  $\mathbf{Y} = \mathbf{V}\Sigma\mathbf{U}^T$  with unitary matrices  $\mathbf{V} \in \mathbb{R}^{n \times n} : \mathbf{V}^T\mathbf{V} = \mathbf{I}_n$  and  $\mathbf{U} \in \mathbb{R}^{m \times m} : \mathbf{U}^T\mathbf{U} = \mathbf{I}_m$  and a diagonal matrix  $\Sigma \in \mathbb{R}^{n \times m}$  of ordered non-negative values. The columns of  $\mathbf{V}$  and  $\mathbf{U}$  are referred to as the left and right singular vectors respectively, and the values of  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_{\min(m,n)})$  are the singular values. The number of non-zero singular values determines the rank of  $\mathbf{Y}$  which we denote as  $r$ . The SVD may be equivalently defined with only the singular vectors corresponding to non-zero singular values, in which cases  $\mathbf{V} \in \mathbb{R}^{n \times r} : \mathbf{V}^T\mathbf{V} = \mathbf{I}_r$ ,  $\mathbf{U} \in \mathbb{R}^{m \times r} : \mathbf{U}^T\mathbf{U} = \mathbf{I}_r$  and  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$ , this is termed compact SVD and is the form used in this work. There are a number of efficient algorithms to compute an SVD (see [Golub & Van Loan \(2013\)](#)) making it extremely practical for analysis.

The Eckart–Young–Mirsky theorem ([Eckart & Young 1936](#)) states that the optimal rank  $K$  matrix approximation of  $\mathbf{Y}$  which minimizes the least-squares residual is given by setting all but the top  $K$  singular values of the SVD of  $\mathbf{Y}$  to zero.

$$\arg \min_{\mathbf{L}: \text{rank}(\mathbf{L}) \leq K} \|\mathbf{Y} - \mathbf{L}\|_F^2 = \mathbf{V}\Sigma_K\mathbf{U}^T \quad (6)$$

With  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_K)$ . A compact SVD can be defined by restricting matrices  $\mathbf{V}$  and  $\mathbf{U}$  to the leading  $K$  vectors corresponding to the  $K$  singular values, we denote this restriction as  $\mathbf{V}_K$  and  $\mathbf{U}_K$  respectively.

PCA is the solution to the same objective as in Equation 6 with a specific decomposition over the matrix  $\mathbf{L}$  as  $\mathbf{V}\mathbf{C}$  with  $\mathbf{V} \in \mathbb{R}^{n \times K}$ ,  $\mathbf{C} \in \mathbb{R}^{K \times m}$  and  $\mathbf{V}^T\mathbf{V} = \mathbf{I}_K$ . The PCA solution can be obtained from the compact SVD solution with the same  $\mathbf{V}$  and  $\mathbf{C} = \Sigma_K\mathbf{U}^T$ .

Any  $\mathbf{V}' = \mathbf{V}\mathbf{U}^T$  and  $\mathbf{C}' = \mathbf{U}\mathbf{C}$  is also an equivalent low rank solution  $\mathbf{L} = \mathbf{V}'\mathbf{C}' = \mathbf{V}\mathbf{C}$ , where  $\mathbf{U} \in \mathbb{R}^{K \times K}$  is any unitary matrix such that  $\mathbf{U}^T\mathbf{U} = \mathbf{I}_K$ .

*SVD/PCA for systematics:*—Under a white noise model for  $\mathbf{N}$  the maximum likelihood estimate of a matrix  $\mathbf{L}$  (not necessarily low-rank) given lightcurves  $\mathbf{Y} = \mathbf{N} + \mathbf{L}$  is equivalent to minimizing the least-squares residual between  $\mathbf{Y}$  and  $\mathbf{L}$  (see Section 6.6 of the Appendix).

$$\arg \max_{\mathbf{L}} p(\mathbf{Y}|\mathbf{L}) \equiv \arg \min_{\mathbf{L}} \|\mathbf{Y} - \mathbf{L}\|_F^2 \quad (7)$$

For the data-model in Equation 1 where  $\mathbf{L}$  is modelled as low rank, a rank  $K$  optimal least-squares solution is found by rank-thresholding the SVD or PCA decomposition of  $\mathbf{Y}$  due to the Eckart–Young–Mirsky theorem, such that the maximum likelihood systematics are found as  $\mathbf{L} = \mathbf{V}_K\mathbf{C}_K$ .

*Spatial systematics model:*—As described in Section 2.1.1 and the introduction, the white noise model for  $\mathbf{N}$  is not realistic, as lightcurves may contain outlier noise due to astrophysical signals or non-white instrumental noise and for which no simple analytic model exists. As  $\mathbf{N}$  deviates from the white noise model, systematics cannot be perfectly separated by rank thresholding alone. This issue motivates us to place further constraints on the form of low rank systematics  $\mathbf{L} = \mathbf{V}\mathbf{C}$  based on prior knowledge.

The coefficient matrix  $\mathbf{C}$  is conditioned by a probabilistic prior which encourages neighbouring systematics coefficients on the sensor to be correlated, hereonafter we refer to this as the spatial prior. The reason the spatial prior is placed on  $\mathbf{C}$  instead of  $\mathbf{L}$  is to achieve convenient expressions for optimizing the final objective described in Section 2.2. In a realistic setting we expect some amount of discontinuity/variation in the systematic coefficients, since the true functional dependence is unknown and edge effects/outliers may be introduced by saturated lightcurves or module/output edges (See Section 2.3 for further information). To incorporate the presence of such effects, we use a functional prior common in image processing, where images are often described as functions of bounded variation ([Rudin et al. \(1992\)](#); [Vogel \(2002\)](#)). A particular measure of the functions variation, known as total variation is used as the negative prior. The total variation measure applied to  $\mathbf{C}$  penalizes the  $L_p$  norm of the difference of normalized systematic coefficients  $\bar{c}_i$  for neighbouring lightcurves. To be precise, we consider the pixel position of each lightcurve

$i \rightarrow (x, y) \in (X, Y)$  and sum total variation terms, over the nearest neighbours in  $x$  and  $y$  for each lightcurve. This prior allows a degree of discontinuity in systematics, the degree to which bounded variation is tolerated is controlled via the parameter  $p \in [1, 2]$ . In our work  $p$  may take on any value between  $[1, 2]$ , depending on the degree of smoothness of systematic variations over a sensor.

The objective is compactly stated as:

$$\arg \min_{\mathbf{V}, \mathbf{C} : \text{rank}(\mathbf{VC}) \leq K} \|\mathbf{Y} - \mathbf{VC}\|_F^2 + \|\mathbf{D}\bar{\mathbf{C}}\|_{2,p}^p \quad (8)$$

This objective can also be shown as equivalent to MAP estimation under a Bayesian formulation, where the first term corresponds to the white noise model for  $\mathbf{N}$  described in Section 2.1.1 and the second term is the total variation prior on  $\mathbf{C}$ . See Appendix 6.6 for a complete derivation of the equivalent probabilistic objective. The difference operator  $\mathbf{D}$  applied to  $\bar{\mathbf{C}}$  computes for each  $k \in K$  and pixel  $(x, y)$ , the difference of neighbouring coefficients  $[\bar{c}_k^{x,y} - \bar{c}_k^{x,y+1}, \bar{c}_k^{x,y} - \bar{c}_k^{x+1,y}] \in \mathbb{R}^2$  as the columns of  $\mathbf{D}\bar{\mathbf{C}} \in \mathbb{R}^{2 \times K(X-1)(Y-1)}$ . In expanded form the prior term is  $\|\mathbf{D}\bar{\mathbf{C}}\|_{2,p}^p = \sum_{k \in K} \sum_{(x,y)} (|\bar{c}_k^{x,y} - \bar{c}_k^{x,y+1}|^2 + |\bar{c}_k^{x,y} - \bar{c}_k^{x+1,y}|^2)^{\frac{p}{2}}$ .

Besides systematic discontinuities or outliers described above which the prior is designed to tolerate, the degree of spatial correlation may be non-uniform throughout the sensor. The origin of non-uniformity may be a dominance of thermal/saturation effects or smearing along one direction (i.e. bleed columns). At a coarse spatial level differences of coefficient vectors in  $\mathbf{D}\bar{\mathbf{C}}$  are weighted by a fixed weight matrix  $\mathbf{W}$ , a diagonal matrix of the weightings between  $[0, 1]$ . For ease of representation, we will write the product of the weight matrix and the difference operator as a single operator:  $\mathbf{D}\mathbf{W} = \mathbf{W}\mathbf{D}$ . The objective with this weight matrix incorporated is then:

$$\arg \min_{\mathbf{V}, \mathbf{C} : \text{rank}(\mathbf{VC}) \leq K} \|\mathbf{Y} - \mathbf{VC}\|_F^2 + \|\mathbf{D}\mathbf{W}\bar{\mathbf{C}}\|_{2,p}^p \quad (9)$$

Basis vectors  $\mathbf{V}$  obtained with SVD and PCA are constrained to be orthogonal, however orthogonality is not a necessary constraint as a property of a systematics model, nor to achieve a minimal optimization cost. Our model does not restrict  $\mathbf{V}$  to be orthogonal.

It can be readily seen why it is necessary to place the prior on the normalized coefficients  $\hat{\mathbf{C}}$ . As if instead the prior was of the form  $\|\mathbf{D}\mathbf{W}\mathbf{C}\|_{2,p}^p$  any solution  $\mathbf{C}$  can be replaced with a solution  $\mathbf{C}' = \alpha\mathbf{C}$  for  $\alpha < 1$  that produces a lower value of the prior. Taking  $\mathbf{V}' = \frac{\mathbf{V}}{\alpha}$  so that there is no change to the least-squares penalty  $\|\mathbf{Y} - \mathbf{VC}\| = \|\mathbf{Y} - \mathbf{V}'\mathbf{C}'\|$ , a trivial minima can be found simply as  $\alpha \rightarrow 0$ .

*The spatial prior with  $p = 2$ :*—When  $p = 2$ , optimizing the spatial prior is exactly equivalent to maximizing the normalized correlation of neighbouring systematic coefficient vectors under the low rank model. We describe this special case (without the inclusion of the weightings) as it provides intuition for the more general form. For ease of notation, denote the set of all pairs  $(i, j)$  of neighbouring lightcurves as  $n(I)$ , the following objective maximizes the correlation of all pairs of coefficients in  $n(I)$ :

$$\arg \max_{\mathbf{V}, \mathbf{C} : \text{rank}(\mathbf{VC}) \leq K} \sum_{(i,j) \in n(I)} \bar{\mathbf{c}}_i^T \bar{\mathbf{c}}_j \quad (10)$$

Since  $\|\bar{\mathbf{c}}_j\|_2 - \|\bar{\mathbf{c}}_i\|_2 = 1$  and  $\|\bar{\mathbf{c}}_j - \bar{\mathbf{c}}_i\|_2^2 = \|\bar{\mathbf{c}}_j\|_2^2 + \|\bar{\mathbf{c}}_i\|_2^2 - 2\bar{\mathbf{c}}_j^T \bar{\mathbf{c}}_i = 2 - 2\bar{\mathbf{c}}_j^T \bar{\mathbf{c}}_i$  an equivalent formulation is given by:

$$\arg \min_{\mathbf{V}, \mathbf{C} : \text{rank}(\mathbf{VC}) \leq K} \sum_{(i,j) \in n(I)} \|\bar{\mathbf{c}}_i - \bar{\mathbf{c}}_j\|_2^2 \quad (11)$$

The summation can be replaced by a linear operator which computes the difference of columns of  $\bar{\mathbf{C}}$ :

$$\arg \min_{\mathbf{V}, \mathbf{C} : \text{rank}(\mathbf{VC}) \leq K} \|\mathbf{D}\bar{\mathbf{C}}\|_F^2 \quad (12)$$

*Correlation of systematics for  $p=2$* —If  $p = 2$  the spatial prior equivalently maximizes the normalized correlation of neighbouring coefficient vectors. It's useful to understand how the correlation of the normalized coefficients  $\bar{\mathbf{c}}_i$  is proportional to the correlation of the systematics themselves  $\mathbf{l}_i = \mathbf{V}\mathbf{c}_i$ . In this section we describe this relation as it provides intuition for the properties of the spatial systematics model.

Define normalized systematics as  $\bar{\mathbf{l}}_i = \frac{\mathbf{V}\mathbf{c}_i}{\|\mathbf{V}\mathbf{c}_i\|_2}$ , the correlation of systematics  $\mathbf{l}_i$  and  $\mathbf{l}_j$  is the inner product of the normalized systematics  $\bar{\mathbf{l}}_i^T \bar{\mathbf{l}}_j$ . Likewise the correlation between coefficients  $\mathbf{c}_i$  and  $\mathbf{c}_j$  is given by  $\bar{\mathbf{c}}_i^T \bar{\mathbf{c}}_j$ . The deviation of  $\bar{\mathbf{l}}_i^T \bar{\mathbf{l}}_j$  from  $\bar{\mathbf{c}}_i^T \bar{\mathbf{c}}_j$  can be bounded and understood in terms of  $\mathbf{V}$ .

In the simplest case, if  $\mathbf{V}$  is orthonormal then  $\|\mathbf{V}\mathbf{c}\| = \|\mathbf{c}\|$ , using this property the correlations are equivalent:

$$\|\bar{\mathbf{l}}_i - \bar{\mathbf{l}}_j\|_2^2 = \left\| \frac{\mathbf{V}\mathbf{c}_i}{\|\mathbf{V}\mathbf{c}_i\|_2} - \frac{\mathbf{V}\mathbf{c}_j}{\|\mathbf{V}\mathbf{c}_j\|_2} \right\|_2^2 = \left\| \mathbf{V} \left( \frac{\mathbf{c}_i}{\|\mathbf{c}_i\|} - \frac{\mathbf{c}_j}{\|\mathbf{c}_j\|} \right) \right\|_2^2 = \|\bar{\mathbf{c}}_i - \bar{\mathbf{c}}_j\|_2^2 \quad (13)$$

$$\implies \bar{\mathbf{l}}_i^T \bar{\mathbf{l}}_j = \bar{\mathbf{c}}_i^T \bar{\mathbf{c}}_j \quad (14)$$

However  $\mathbf{V}$  is not strictly orthonormal. In the general case, the deviation of the systematics  $\bar{\mathbf{l}}_i^T \bar{\mathbf{l}}_j$  from the coefficient  $\bar{\mathbf{c}}_i^T \bar{\mathbf{c}}_j$  can be understood through the singular values of  $\mathbf{V}$ . Define  $\mathbf{V}$  to be rank  $K$  with singular values of descending magnitude  $\sigma_1 \geq \sigma_2 \dots \geq \sigma_K \geq 0$  and left singular vectors  $\mathbf{u}_1, \mathbf{u}_2 \dots \mathbf{u}_K$ . Without loss of generality assume  $\sigma_1 = 1$  and  $\|\mathbf{c}_i\| = \|\mathbf{c}_j\| = \|\bar{\mathbf{c}}_i\| = \|\bar{\mathbf{c}}_j\| = 1$ . Define the angle  $\phi$  between normalized coefficient vectors as  $\cos(\phi) = \bar{\mathbf{c}}_i^T \bar{\mathbf{c}}_j$ . The maximal change of the systematics correlation  $\bar{\mathbf{l}}_i^T \bar{\mathbf{l}}_j$  from the coefficient correlation  $\bar{\mathbf{c}}_i^T \bar{\mathbf{c}}_j$  occurs when  $\mathbf{c}_i$  and  $\mathbf{c}_j$  lie in the span of  $\mathbf{u}_1$  and  $\mathbf{u}_K$ . The worst case deviation is given as:

$$\frac{\sigma_K^2 - \tan^2(\phi)}{\tan^2(\phi) + \sigma_K^2} \leq \bar{\mathbf{l}}_i^T \bar{\mathbf{l}}_j \leq \frac{1 - \sigma_K^2 \tan^2(\phi)}{1 + \sigma_K^2 \tan^2(\phi)} \quad (15)$$

As the correlation  $\bar{\mathbf{c}}_i^T \bar{\mathbf{c}}_j \rightarrow 1 \implies \phi \rightarrow 0$ , the lower and upper bounds smoothly converge to 1, so that  $\bar{\mathbf{l}}_i^T \bar{\mathbf{l}}_j \rightarrow 1$ , this implies the possible deviation of systematics correlation is lower when the coefficient correlation is high. Furthermore the worst case deviation of the correlation depends on the ratio of the smallest and largest singular values of  $\mathbf{V}$ , when  $\sigma_K = \sigma_1$  there is no deviation. In Figure 2.1.2.0 for a hypothetical value  $\sigma_K = .6$  the systematics correlation bounds are shown as a function of  $\phi$  and as a function of the coefficient correlation respectively. See Appendix Section 6.3 for a sketch of the derivation of this relation.

## 2.2. Fitting systematics

The objective in Equation 20 minimizes the penalty over variables  $\mathbf{V} \in \mathbb{R}^{N \times K}, \mathbf{C} \in \mathbb{R}^{K \times I}$ , however the method to minimize the penalty has not yet been described. In this section we describe our choice of minimization method. For simplicity, we will decompose Equation 20 into the least squares penalty  $f(\mathbf{V}, \mathbf{C}) = \|\mathbf{Y} - \mathbf{V}\mathbf{C}\|_F^2$  which depends on both variables  $\mathbf{V}, \mathbf{C}$ , and the normalized correlation constraint  $g(\mathbf{C}) = \|\mathbf{D}\mathbf{w}\bar{\mathbf{C}}\|_{2,p}^p$  which only depends on  $\mathbf{C}$ . The least-squares penalty  $f(\mathbf{V}, \mathbf{C})$  is not separable with respect to  $\mathbf{V}, \mathbf{C}$ . The variables  $\mathbf{V}, \mathbf{C}$  can be minimized by alternately updating either  $\mathbf{V}$  or  $\mathbf{C}$  while the other is held fixed.

$$\mathbf{V}' = \arg \min_{\mathbf{V}} f(\mathbf{V}, \mathbf{C}') \quad (16)$$

$$\mathbf{C}' = \arg \min_{\mathbf{C}} f(\mathbf{V}', \mathbf{C}) + g(\mathbf{C}) \quad (17)$$

The minimum of the least-squares term  $f(\mathbf{V}, \mathbf{C})$  with respect to one variable has an analytic solution, however the total variation prior  $g(\mathbf{C})$  does not. There are many minimization methods for total variation penalties, see Vogel (2002), which can be used to compute an approximate minimum  $\mathbf{C}'$ . We choose not to use the alternating updates minimization approach, and instead eliminate variable dependence on  $\mathbf{V}$  in the least squares penalty, so that the objective may be minimized over  $\mathbf{C}$  alone. For a particular  $\mathbf{C}$  the minimizing value  $\mathbf{V}$  of the least-squares penalty is denoted as  $h(\mathbf{C})$ .

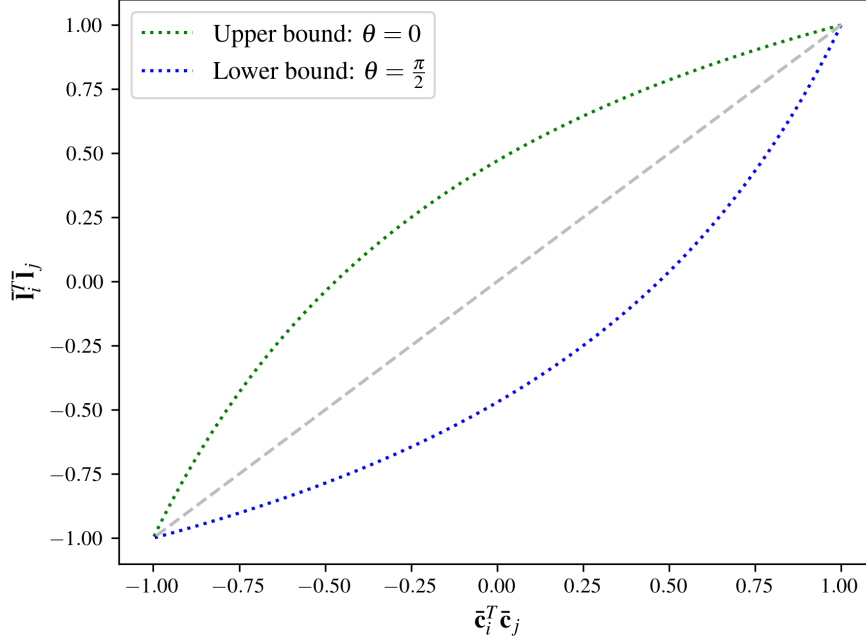
$$h(\mathbf{C}) = \arg \min_{\mathbf{V}} f(\mathbf{V}, \mathbf{C}) \quad (18)$$

For the least-squares penalty the minimizing  $\mathbf{V}$  for a particular  $\mathbf{C}$  can be exactly obtained, this function  $h(\mathbf{C})$  can be inserted into the original objective:

$$\min_{\mathbf{V}, \mathbf{C}} f(\mathbf{V}, \mathbf{C}) + g(\mathbf{C}) \rightarrow \min_{\mathbf{C}} f(h(\mathbf{C}), \mathbf{C}) + g(\mathbf{C}) \quad (19)$$



Example deviation of systematics correlation  $\bar{\mathbf{I}}_i^T \bar{\mathbf{I}}_j$  from coefficient correlation  $\bar{\mathbf{c}}_i^T \bar{\mathbf{c}}_j$ ,  $\sigma_1 : \sigma_k = .6$



**Figure 1.** Deviation of systematics correlation  $\bar{\mathbf{I}}_i^T \bar{\mathbf{I}}_j = \frac{\mathbf{c}_i^T \mathbf{V}^T \mathbf{V} \mathbf{c}_j}{\|\mathbf{V} \mathbf{c}_i\| \|\mathbf{V} \mathbf{c}_j\|}$  from coefficient correlation  $\bar{\mathbf{c}}_i^T \bar{\mathbf{c}}_j$  for  $\sigma_1 : \sigma_k = .6$ . The upper and lower bounds are symmetric, and converge as the coefficient correlation  $\bar{\mathbf{c}}_i^T \bar{\mathbf{c}}_j$  increases.

Using the theory described in (Golub & Pereyra (2003)) and see Appendix 6.4, the objective reduces to a function of a single variable  $\mathbf{c}$  as:

$$\arg \min_{\mathbf{C} : \text{rank}(\mathbf{C}) \leq K} \|(I - \mathbf{C}^T \mathbf{C}^\dagger) \mathbf{Y}^T\|_F^2 + \|\mathbf{D} \mathbf{w} \bar{\mathbf{C}}\|_{2,p}^p \quad (20)$$

There are not results to prove either alternating updates or variable elimination is preferable, the derivations in the appendix can be used to obtain alternating update steps if preferred.

The derived objective is not differentiable due to the  $L_p$  norm in the correlation constraint  $g(\mathbf{C})$  which is discontinuous for entries of  $\mathbf{D} \mathbf{w} \bar{\mathbf{C}}$  which are 0, when  $p = 1$ . Proximal methods are standard for obtaining minima of objective functions with non differentiable penalties such as  $L1$  norms, see the monograph by (Parikh & Boyd (2013)) for further information, in particular the Split-Bregman method (Goldstein & Osher (2009)) is particularly well suited to total variation regularized problems. Such techniques involve multiple optimization steps and tunable optimization parameters, however the goal of this work is to construct a simple method which works with little tuning. A preferred approach is to replace the non-differentiable  $L_p$  norm with a close approximation that largely retains the same robustness properties. We follow the method described in Vogel (2002) where the  $L_p$  norm is replaced with a continuously differentiable Huber functional. This approximation adds a small error  $\delta$  to the individual norm terms so they are bounded from 0 and the approximate norm behaves quadratically as  $x_i \rightarrow 0$ . In practise in the limit as the magnitude of a discontinuity in  $\mathbf{c}$  goes to 0, discontinuities are not preserved and more likely to be smoothed. The details of the approximate form and derived gradients are described in the Appendix 6.5.

We minimize the reduced objective via gradient descent, the general algorithm applied to Equation is described in Algorithm 1.

**Algorithm 1:** Variable Projection Gradient Descent for Systematics

---

Initialize  $\mathbf{C}$  for fixed  $K$ , fix weight matrix  $\mathbf{W}$ . The step-size  $\alpha^i$  may be fixed or vary at each step.

**while**  $|\mathbf{C}_{i+1} - \mathbf{C}_i| \geq \epsilon$  **do**  
  1)  $\nabla f(\mathbf{C}^i) = \nabla_{VP}(\mathbf{C}^i) + \mathbf{W}\nabla_{TV}(\mathbf{C}^i)$   
  2)  $\mathbf{C}^{i+1} = \mathbf{C}^i - \alpha^i \nabla f(\mathbf{C}^i)$   
**end**  
 $\mathbf{L} = \mathbf{V}_{|\mathbf{C}}\mathbf{C}$  with  $\mathbf{V}_{|\mathbf{C}} = \mathbf{Y}\mathbf{C}^\dagger$ .

---

A discussion of how to choose parameters  $p, \alpha, W, \epsilon$  and model order  $K$  is given in Section 2.3.3.

### 2.3. Experimental analysis

The systematic inference method is described algorithmically in Section 2.2. To evaluate the appropriateness of this method we apply it to Kepler lightcurves, as well as simulated lightcurves and use the standard low-rank estimation model PCA as a comparative measure. This section describes implementation details and experimental setup, as well as validation metrics. First, we briefly investigate the degree of spatial correlation present in lightcurves and necessary pre-processing steps.

#### 2.3.1. Lightcurve pre-processing

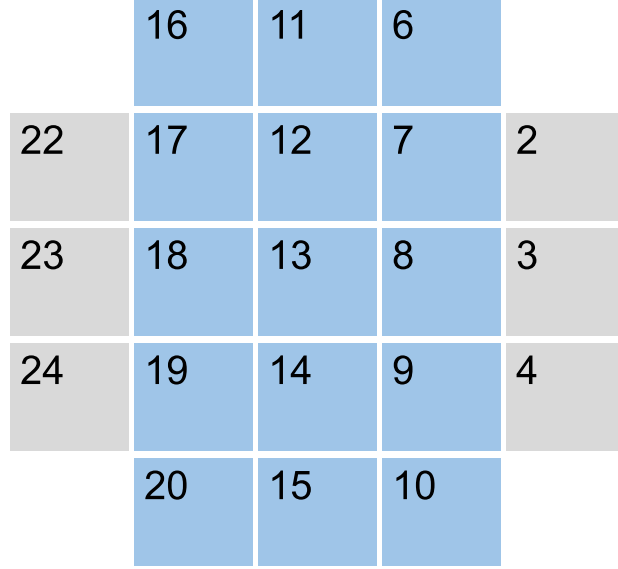
We investigate spatial correlation in Kepler SAP long-cadence lightcurves, with a 29.4 minute integration time (Jenkins et al. 2010) selected from the MAST database<sup>3</sup>. Every 4 quarters, approximately the same set of stars is overlaid on the sensor (Aigrain et al. 2017). We apply the described method to lightcurves from quarters (Q hereinafter) (6, 10 and 14). Since the stellar sample has a large degree of overlap between these quarters, we do not expect stellar variability to significantly change quarter to quarter thus providing a metric of the stability of the detrending method. Stellar targets were selected within the range  $Kp \in [12.5, 13, 5]$  and  $\text{CDPP } 12[hr] \leq 40$ , this target selection criteria was informed by Petigura & Marcy (2012). For quarters (6, 10 and 14) respectively there are (6179, 6286 and 6049) lightcurves in this magnitude range.

The Kepler sensor is broken up into 24 modules (each containing 4 output channels or 2 CCD's), the layout of the modules is shown in Figure 2.3.1 with each square representing an individual module. The readout origin for each output channel starts at one of the corners of the module, hence pixel position for each output channel must be remapped to a global coordinate system. Each module is approximately 2200 x 2200 pixels in size. For a complete overview of the Kepler sensor see the Kepler Instrument Handbook Van Cleve & Caldwell (2016). The locations of the sample targets as a function of their position on the sensor is shown in Figure 2.3.1, for ease of processing we focus only on the region shown in light blue and discard the targets on modules (2, 3, 4, 22, 23, 24).

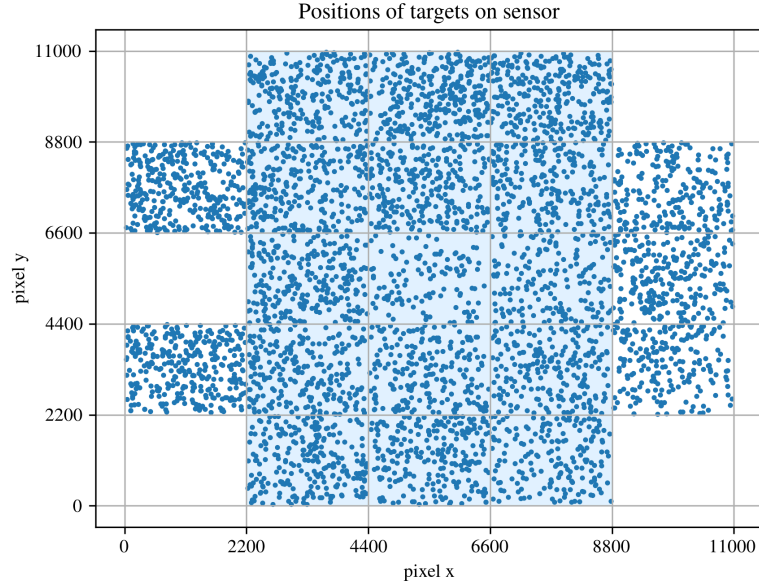
The method of systematics inference is applicable to a uniform grid layout, however targets are non-uniformly distributed on the sensor. We discretize the sensor into (220x220) pixel sized cells and select one lightcurve per cell. The lightcurve is selected from the filtered set of targets as the lightcurve closest to the center of the defined cell. If there is no lightcurve in that cell we select a lightcurve that falls within a neighbouring cell (choosing a lightcurve that has not already been used and is itself closest to the current cell) - here a single cell was filled in this manner. As such the region of the sensor analysed is of dimension (30 x 50) cells, with one lightcurve per cell. Generalizing this method to non-neighbouring pixel correlations, or non-uniform layouts is left for further work.

All missing/erroneous data values were gap-filled with a simple linear interpolation between the start and the end of the gap. Each lightcurve is linearly detrended via a line fit, Figure 2.3.1 shows the effects of linear detrending. It can be seen that lightcurves with opposite slopes are almost identical once the slope is removed, the linear component of a lightcurve does not seem to be clearly correlated with short term systematics. The linear trend of a lightcurve may be systematic or otherwise astrophysical however since co-located lightcurves with highly correlated short term systematics, have anti-correlated slopes we advise that if a linear trend is incorporated in the model the spatial correlation weighting for this term be set to 0 or handled separately as we do here. The basis vectors otherwise

<sup>3</sup> <https://archive.stsci.edu/>

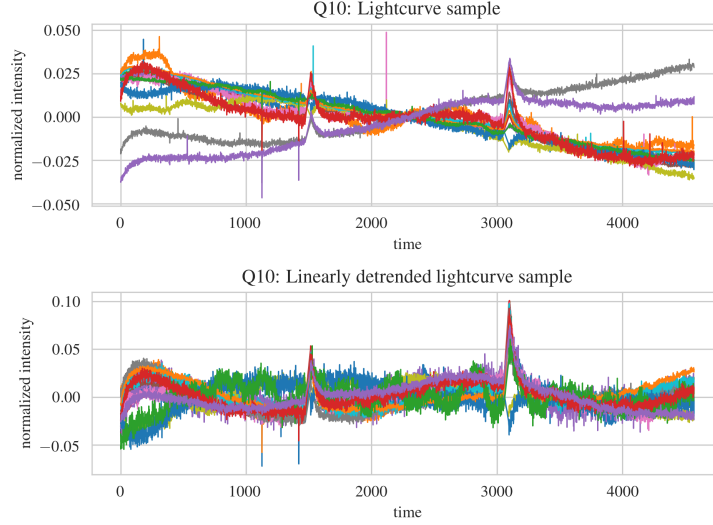


**Figure 2.** Kepler module layout labeled by module number, excluded modules are shown in light grey.



**Figure 3.** Positions of all targets on the sensor for Q (10) are shown, axes lines indicate edges of modules. The colored region indicates the output modules used in this study.

may include mixtures of uncorrelated long-term slopes and short term features lightcurves, such vectors may overfit lightcurves thus injecting systematics which are not present in the lightcurve. In the MS-MAP detrending method due to (Stumpe et al. 2013), a more detailed observation of this effect motivates multiscale detrending; systematic effects occurring on separate timescales are approximately independent. The linear component may indeed be due to astrophysical variability and can be added back in as a post-processing step.



**Figure 4.** The first figure shows a sample of 15 lightcurves from Q (10), the second figure shows the effect of removing the linear component of each lightcurve and non-correlated outlier filtering. Primarily the systematics shown are due to the quarterly roll and earth point recoveries. These shorter term effects, which may be multiplicative, are more easily identified by removing a long term linear component which seems to be more variable between lightcurves.

Lightcurves are median normalized<sup>4</sup> and filtered to include the intersection of lightcurves which are among the 90% of lightcurves with lowest absolute first order difference variance and the 90% of lightcurves with lowest variance. This leaves a sample of approximately 6000 lightcurves for a given quarter. Finally, as a coarse means to remove astrophysical outliers, lightcurves are roughly detrended via PCA and the point to point scatter is  $3\sigma$  thresholded. A further 4 outlier cadences which were identified by eye are masked. PCA is performed using the implementation in the Python module sklearn<sup>5</sup>

### 2.3.2. Implementation

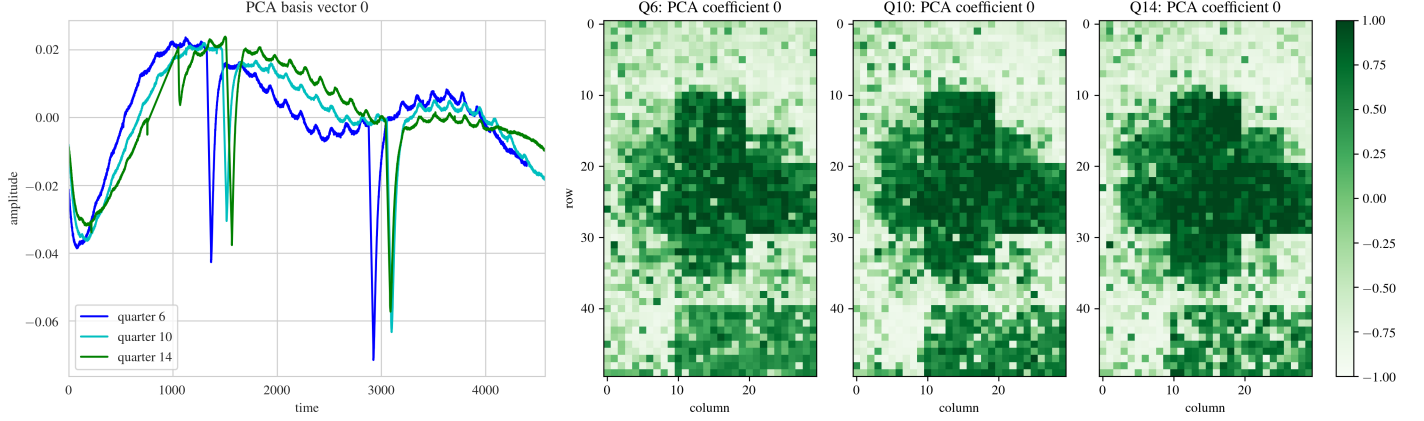
The pixel response function model used to construct lightcurves is characterized per output channel, and systematics are correlated within an output channel/module, generally motivating the choice of many algorithms to apply detrending methods on discrete regions independently. Furthermore, systematics may be spatially correlated, within a discrete sensor region or across them (Moreno et al. 2021; Petigura & Marcy 2012; Lund et al. 2021). We perform a brief exploratory analysis into the structure of spatial correlation by applying PCA to the selected lightcurves from (Q6, Q10 and Q14). In particular we investigate i) the presence of spatial correlation, ii) If spatial correlation is stronger within rows or columns, iii) the degree of spatial correlation per module.

The leading PCA basis vector for each quarter is shown in Figure 2.3.2 alongside the coefficient values for each cell (lightcurve) mapped to the location of the cells on the sensor. The leading systematics basis vector is informative of general systematics as it is the strongest term. The leading systematics basis vector is very similar between the quarters, as is the magnitude of this effect on the sensor. This indicates that this term represents a cyclic systematic effect. Within a module, the coefficient values are highly correlated as can be seen from the blocking of fitted coefficient values by module. This pattern is persistent over time and among different targets selected per cell, indicating it is not astrophysical in origin or coincidental.

If we wished to simply use within module correlation as a prior, we could construct a difference matrix which computes all pairwise differences within a module. However neighboring correlation is empirically stronger than within module correlation alone. Defining the correlation between  $\mathbf{x}$  and  $\mathbf{y}$  as  $corr(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}$ , Figure 2.3.2 shows the mean pairwise correlation computed for all pairs of lightcurves within a module, for each module. Indexing the set of all

<sup>4</sup>  $\hat{x} = \frac{x}{med(x)} - 1$

<sup>5</sup> <https://scikit-learn.org/stable/>



**Figure 5.** Left: The leading PCA basis vector for quarters (6, 10 and 14). Right: leading PCA component plotted as a function of sensor position for quarters (6, 10 and 14). Put mean value of each module on figure and module number in corner.

lightcurves within a module by  $M$ .

$$corr_M = \frac{1}{2|M| - 2} \sum_{m \in M} \sum_{n \in M \setminus \{m\}} corr(\mathbf{y}_m, \mathbf{y}_n) \quad (21)$$

Alongside this, the difference between mean module correlation  $corr_M$  and mean neighbour pairwise correlation (including x and y neighbours) for all lightcurves within a module  $corr_{dX,dY}$ , per module is shown in purple. Denoting the cell positions as  $X$  and  $Y$  within a module, the mean neighboring correlation within a module is calculated as:

$$corr_{dX,dY} = \frac{1}{2XY} \sum_{x \in X} \sum_{y \in Y} corr(\mathbf{y}_{x,y}, \mathbf{y}_{x,y+1}) + corr(\mathbf{y}_{x,y}, \mathbf{y}_{x+1,y}) \quad (22)$$

Generally systematics are highly correlated within a module (Smith et al. 2012), and for certain modules systematics undergo very little variability. As can be seen, the neighbour to neighbour correlations are generally higher, referring to Figure 2.3.2 in modules where the leading systematic coefficient transitions gradually e.g. (7, 8, 9, 14, 17), there is a larger degree of neighboring correlation as compared to within module correlation. Generally it can be seen that systematics undergo 'changes' roughly at module edges leading to anti-correlations in certain places.

The choice of where to use the spatial correlation model and to what degree can be fine-tuned by the choice of weight matrix  $\mathbf{W}$ . In Figure 2.3.2 we marginalize the neighbouring correlations along each row and column of the sensor, the edges have a higher degree of spatial correlation than the central rows and columns.

### 2.3.3. Model Parameters

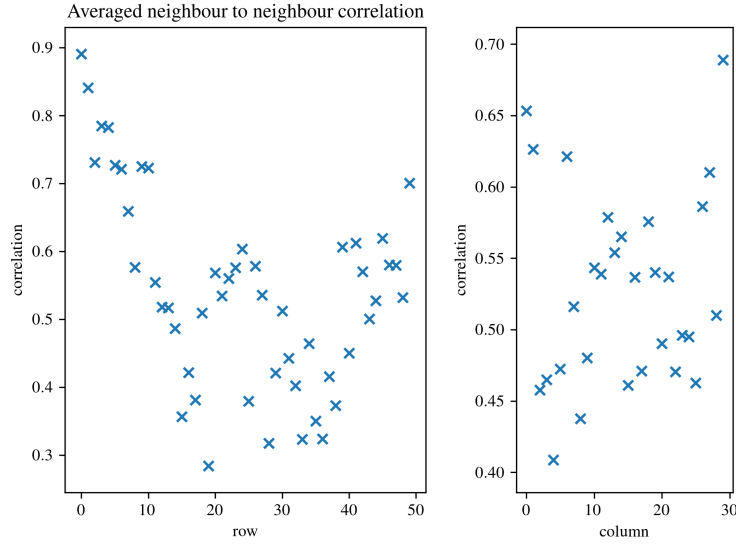
There are several free parameters to be chosen in our method, both in the choice of prior and the descent method. Briefly these are, the choice of initialization  $\mathbf{C}_0$ , the model rank  $K$ , the choice of prior norm  $p$ , the gradient step size  $\alpha$ , the prior weighting matrix  $\mathbf{W}$  as well as the algorithm stopping condition. The results and performance heavily depend on the chosen model parameters, and were generally empirically chosen based on experimental results for randomly generated data.

In general increasing the model order  $K$  can always lower the objective cost, however the benefit of more complete systematics modelling is offset by the inclusion of non-systematic signals. This form of model mismatch is explored in Experiment 2.3.4.0 wherein the fitted model order is explicitly chosen to be larger than the ground truth systematics. Typically, when the exact rank is unknown, model order is empirically chosen based on the singular value distribution of the data. An SVD can be performed as an investigative procedure, singular values  $\Sigma$  represent the significance of corresponding basis vectors and can be used to choose the rank  $K$ , suitable to capture the net systematics. In Figure 2.3.3 sample singular values for the set of 1500 lightcurves from all Kepler modules (described in Section 2.3.1) are shown and can be seen to rapidly fall off. Theoretically  $\mathbf{N}$  is full rank with singular values  $\mathbf{1}$ , so  $\Sigma$  will not reach zero.

Within-module mean pairwise correlation

column	6	11	16
	0.66	0.67	0.88
	0.02	-0.02	0.01
	7	12	17
	0.42	0.50	0.51
	0.05	-0.01	0.19
	8	13	18
	0.41	0.58	0.44
	0.21	0.06	-0.03
	9	14	19
	0.46	0.38	0.38
	0.07	0.23	0.03
	10	15	20
	0.63	0.61	0.54
	0.08	0.04	0.02
	row		

**Figure 6.** Module number is shown in gray. Black: The mean pairwise correlation of all lightcurves within a module. The mean pairwise neighbour correlation of all lightcurves within a module, as a difference from black is shown in purple. A positive value represents an increase. The overall mean pairwise correlation is 0.26.

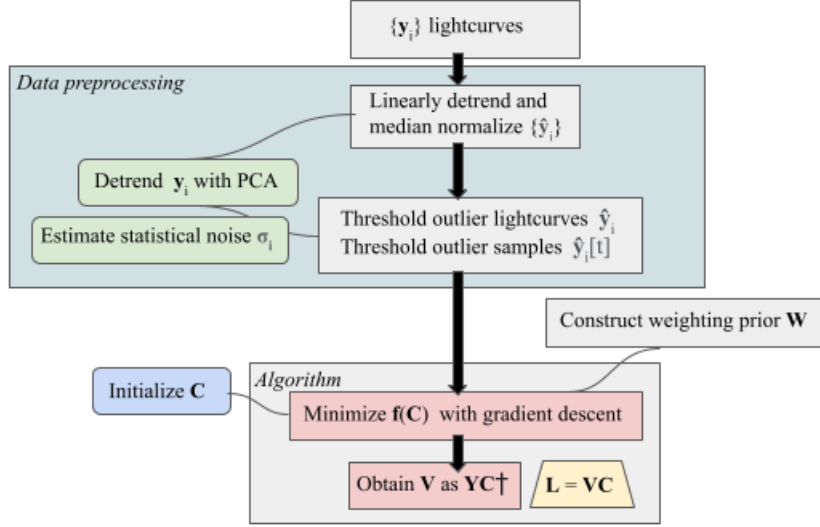


**Figure 7.** Q10 neighbour to neighbour correlation of lightcurves averaged for all columns and rows of the sensor. For comparison the average pairwise correlation between all lightcurves is 0.26, this suggests lightcurves are more likely to be spatially correlated than correlated with a random lightcurve (in this sample). These values are computed including negative correlations at module edges.

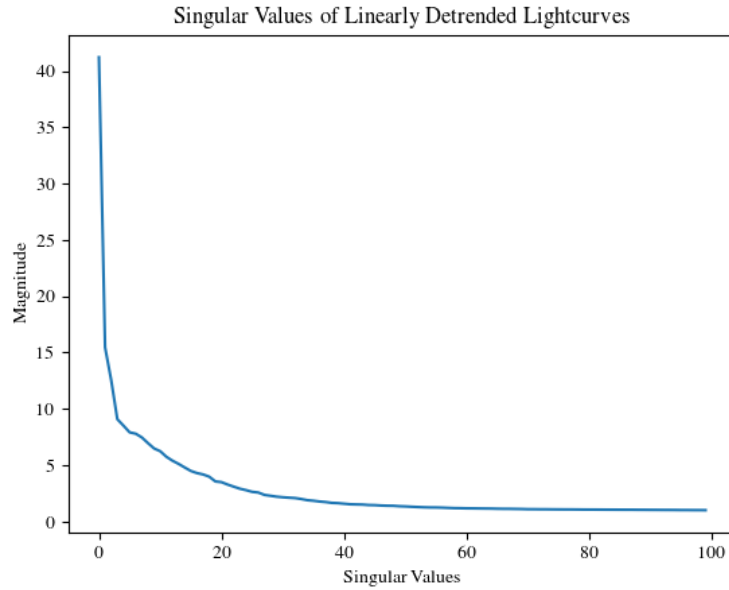
We initialized the algorithm with  $\mathbf{C}_0$  obtained from PCA applied to  $\mathbf{Y}$ , this is because it is likely that the PCA solution is close to the true systematics model and can be improved by the model iterations. The choice of  $p$  determines the degree of smoothness of correlations, and the tolerance of discontinuities and is discussed in Appendix Section 6.2, in Figure 2.3.3 the effect of choices of  $p$  on resulting coefficient solutions is shown. For the simulations described we used  $p = 1.1$ . The weighting matrix  $\mathbf{W}$  was formed as pairwise correlations of lightcurves  $\mathbf{y}_i$  and scaled by a factor .1, and gradient step size  $\eta = 0.04$ . Our stopping condition was based on the difference between successive iterates falling to a negligible level.

#### 2.3.4. Simulation



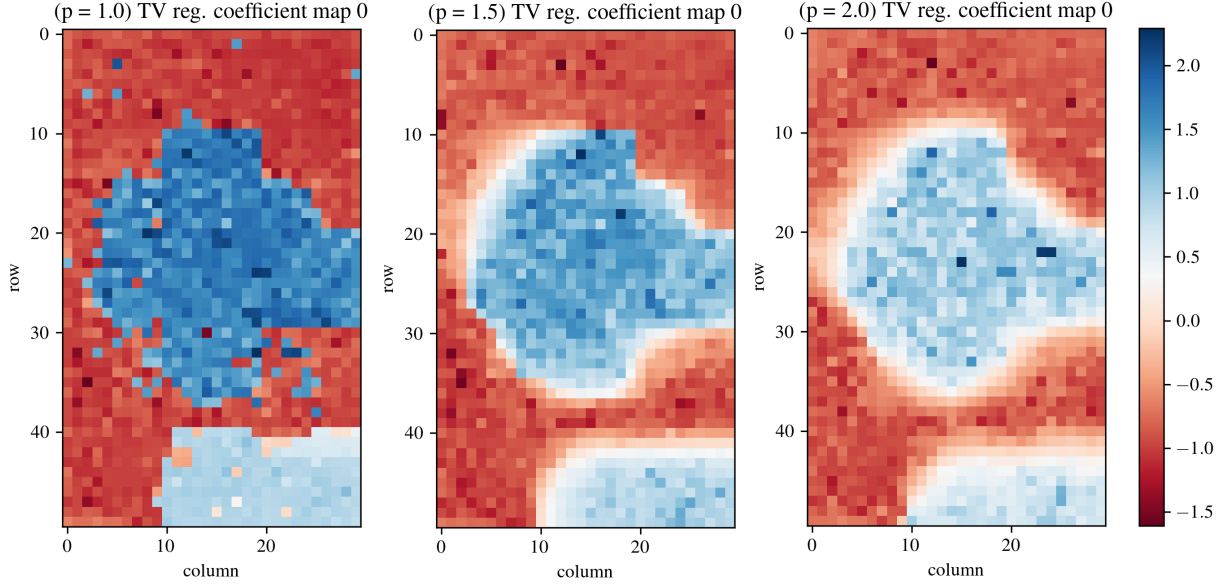


**Figure 8.** Dataflow of the algorithm.

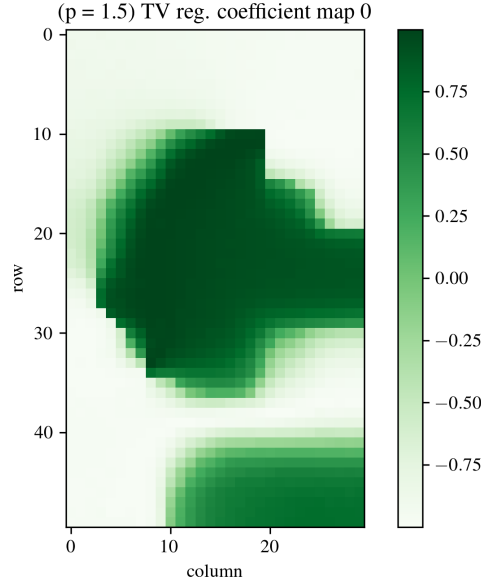


**Figure 9.** Singular values of sample lightcurves, the lightcurves have been minimally filtered (a linear component has been removed from each of them). Approximately a sample of 6000 lightcurves of length 4000. Only the top 100 singular values are shown. The lightcurves contain a small number of high magnitude components attributable to systematic effects.

As described in the introduction, there are many approaches to handling non-Gaussian lightcurve components in low rank systematics estimation. These can be approximately categorized by methods for constructing a robust basis, and those for robust fitting of a basis to individual lightcurves. In the former category, the set of lightcurves used to construct a basis may be filtered to remove 'outlier' lightcurves. A basis may be further refined by post-processing to remove features which are astrophysical, otherwise a basis may overfit and remove such features present in a lightcurve or inject these features into lightcurves where they are not present. A basis consisting only of systematic features can still be overfit to astrophysical features (Smith et al. 2018) since it is unlikely that a basis is completely orthogonal to all possible astrophysical signals, thus the latter approach of constrained fitting is necessary.



**Figure 10.** The effect of varying  $p$  on inferred systematics.



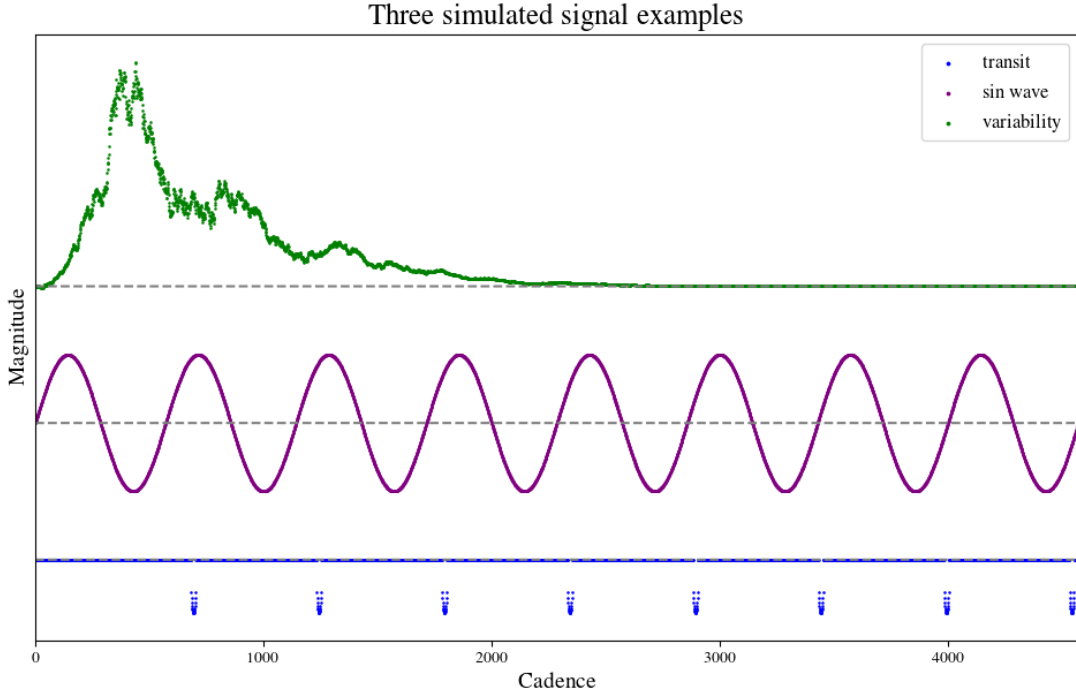
**Figure 11.** Normalized coefficients

The goal of our systematics detrending method which is described in Section 2.2 is to provide greater robustness to overfitting of astrophysical variability in lightcurves. Apriori there exists no absolute ground truth for systematics in Kepler data, to this end we use two complementary experimental simulations described in Section 2.3.4.0 where our method is compared to standard low-rank techniques of PCA and least-squares and the robustness is evaluated with simulated astrophysical signals. We evaluate the experimental performance using a number of metrics described in Section 3.1. The experiments are not intended to show general optimality of our method over the unconstrained least-squares solution, rather they are focused on cases where the latter approach may be susceptible to the overfitting

problems described above. Nor do we intend to claim optimality over specialized detrending approaches such as MS-MAP.

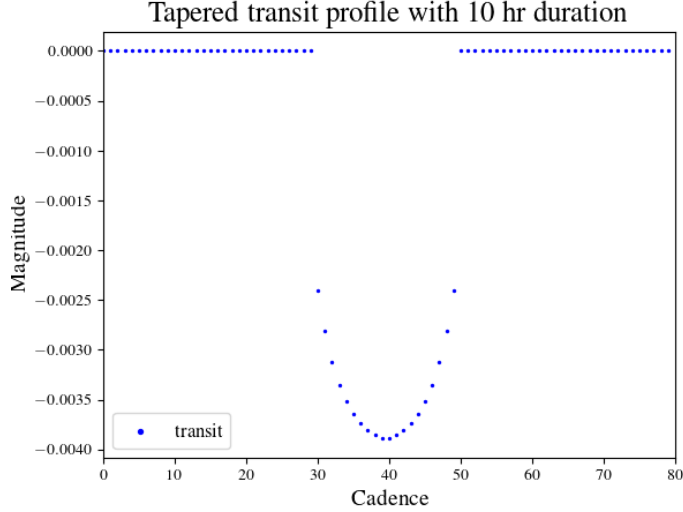
*Simulated Astrophysical Signals*—Three forms of signals are simulated and injected into lightcurves; sinusoidal signals, simulated flare events and transit signals. These forms were selected as they are often the subject of study within a lightcurve, and because they exhibit features on a range of short and long timescales. We denote an individual simulated astrophysical signal as  $\mathbf{a}_s$  where  $s \in \text{sin}, t, \text{flr}$ . The simulated signals are uniquely generated based on randomly selected parameter values. The simulated signals vary in strength but are generally not weaker than the p2p scatter and not greater than the magnitude of systematics, they have a mean norm, as compared to lightcurves of unity norm, of  $\frac{\|\mathbf{a}_s\|}{\|\mathbf{y}\|} = 0.3$ . In Figure 12 three sample simulated signals are shown.

- A sin wave, denoted as *sin*. The amplitude is uniformly selected between  $U(0.005, 0.015)$  of the lightcurve norm  $\|\mathbf{y}\|$ , with angular frequency  $U(\frac{4}{N}, \frac{8}{N})$  long cadence samples, where  $N$  is the number of samples in a lightcurve.
- Simulated flare signals denoted as *flr*. Randomly generated at a random location, profile generated as two exponentials of different taper, multiplied by a random walk.
- Simulated transit signals denoted as *t*. Transits are simulated to allow at minimum 5 transits to occur in the lightcurve, therefore the period is uniformly selected between  $U(1, 14)\text{days}$ . The duration is uniformly selected between  $U(4, 12)\text{hr}$ . The depth is selected between  $U(0.002, 0.016)$  relative to a unity-norm lightcurve. A tapered profile is used to approximate a limb-darkened transit profile, an example of a tapered profile for a transit with a 10 hr duration is shown in Figure 21.



**Figure 12.** Example of three simulated signals.

*Experiment 1*—The degree of overfitting that can occur depends on the form of systematics and astrophysical signals present, therefore in the first experiment we use real lightcurves with real but unknown systematics to investigate overfitting. The lightcurves are selected from two neighbouring modules (16, 17) and are broken into a  $(10 \times 20)$  cell



**Figure 13.** Tapered transit profile.

region. This region is shown, relative to the Kepler sensor region for the leading fitted PCA coefficient in Figure 15. Two modules were selected to provide a sample of lightcurves which can reasonably be studied individually, however the choice of was arbitrary.

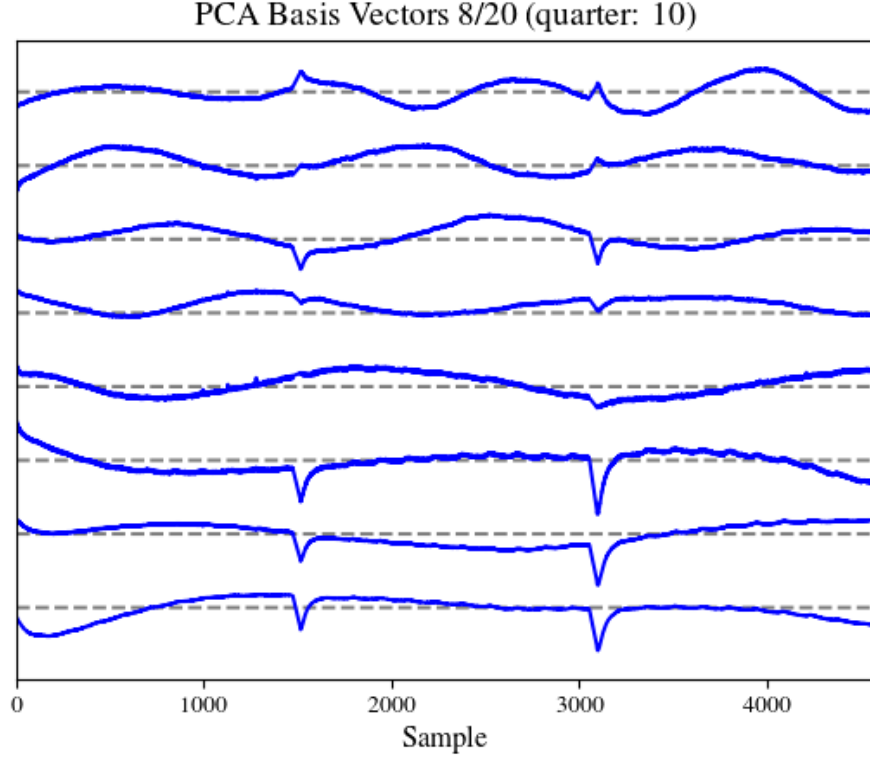
Let  $\mathbf{Y}$  be the set of real lightcurves containing unknown systematics as well as unknown astrophysical variability. We look at how well we can recover simulated and injected astrophysical signals denoted by  $\mathbf{A}$ . A description of simulated astrophysical signals is given in Section 2.3.4.0. By simulating astrophysical signals we can exactly quantify how well we recover these signals. We simulate 3 forms of astrophysical signal (flare event, sinusoidal and transit), each of which is injected into a randomly selected lightcurve. The total number of injected lightcurves is small since the lightcurves already contain unknown astrophysical variability and therefore we do not wish to significantly bias the degree of astrophysical variability realistically present. Lightcurves which have not been injected are used to evaluate our algorithm performance by ensuring that lightcurves have not been corrupted by simulated astrophysical signals in other lightcurves as a consequence of our detrending method. See section 3.1 for a full description of the metrics used to evaluated performance.

We compare the performance of our method to PCA, however we form our PCA basis denoted by  $\mathbf{V}_{PCA}$  and shown in Figure 14, before injecting astrophysical signals so that we can be sure that the basis does not include the injected astrophysical signals. This allows us to assess whether a basis which explicitly does not contain astrophysical signals can overfit them. After injecting the simulated astrophysical signals  $\mathbf{A}$ , least-square fits of the basis to lightcurves are formed denoted by  $\mathbf{C}_{LS}$  and the lightcurves are detrended with these estimates, denoted as  $\mathbf{Y}_{PCA}$ . Our method is initialized with  $\mathbf{C}_{LS}$  and produces fitted coefficients  $\mathbf{C}_{ALG}$  and detrended lightcurves  $\mathbf{Y}_{ALG}$ . Both our method and the least-squares PCA comparison we use a model order  $K = 20$ , empirically sufficient for representing systematics, and smaller than both the number of lightcurves 200 and long-cadence samples in a lightcurve  $\sim 4000$ .

The experimental steps are summarized as follows:

1. Construct a set of systematic basis vectors  $\mathbf{V}_{PCA}$  from applying PCA to  $\mathbf{Y}_{KEP}$ . The inferred PCA basis vectors are shown in Figure 14.
2. Inject  $M$  lightcurves each with simulated astrophysical signals  $\mathbf{A}$ :  $\mathbf{Y}_{KEP} \rightarrow \mathbf{A} + \mathbf{Y}_{KEP}$ .
3. Least-squares fit the basis  $\mathbf{V}_{PCA}$  which explicitly does not contain  $\mathbf{A}$  to the injected  $\mathbf{Y}_{KEP}$  to form  $\mathbf{C}_{LS}$ . Calculate detrended lightcurves as  $\mathbf{Y}_{LS} = \mathbf{Y} - \mathbf{V}_{PCA}\mathbf{C}_{LS}$ .
4. Initialize Algorithm 1 with  $\mathbf{C}_{LS}$  and obtain solution  $\mathbf{C}_{ALG}$  and detrended lightcurves  $\mathbf{Y}_{ALG}$ .

*Experiment 2*—In Experiment 1 the true lightcurve decomposition is unknown, in Experiment 2 we simulate lightcurves as additive combinations of simulated astrophysical signals, simulated systematics and white noise, enabling us to

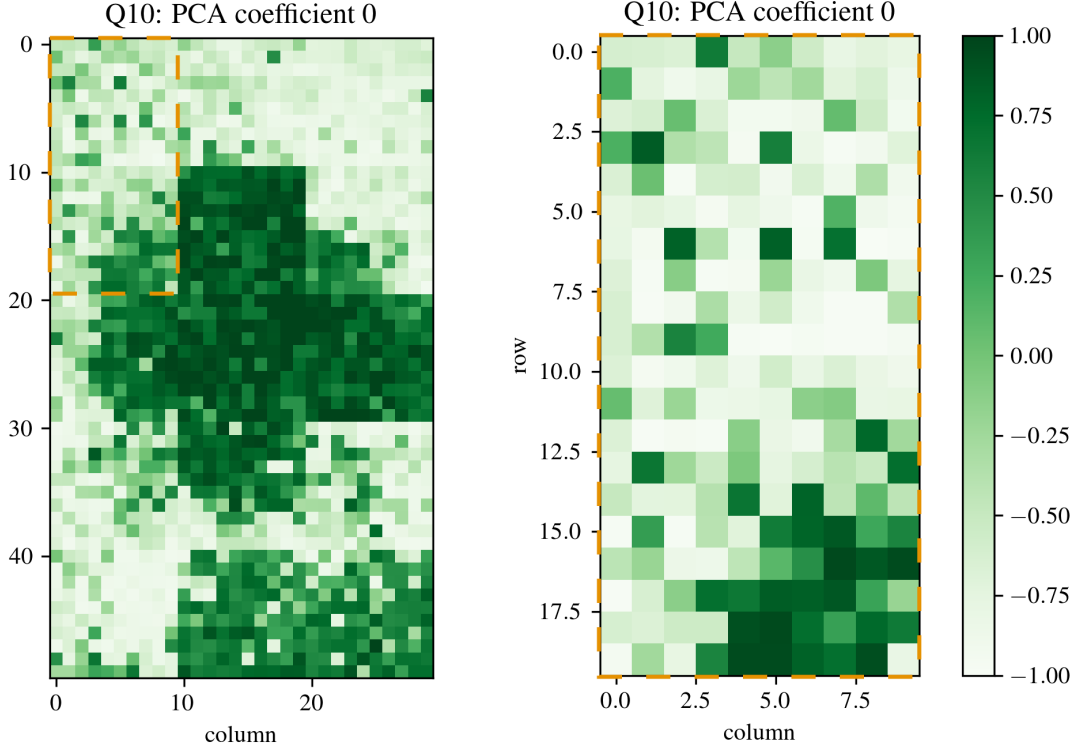


**Figure 14.** Experiment 1: The first 8/20 PCA basis vectors obtained from lightcurves. The basis was obtained prior to injecting simulated astrophysical signals.

study the algorithm performance with ground truth. Systematic effects originate due to unmeasured changes in the sensor performance over time, and across a variety of physical sensor scales. Due to the complex origin of systematics a generative model does not apriori exist. As such in this experiment, we simulate systematics using a low rank linear model of a simple form. Broadly this experiment is informative of the performance of our algorithm for systematics where a correlated low rank model is applicable however it is not informative on the true form of systematics.

Denote simulated lightcurves by  $\mathbf{Y}$ , we simulate 1500 lightcurves in a  $30 \times 50$  cell region, chosen to mirror the Kepler sensor region and discretization used in Section 2.3.1. Systematics are simulated with rank  $K = 4$ , with basis vectors  $\mathbf{V}$  and coefficients  $\mathbf{C}$ . Let  $\mathbf{V}$  consist of Kepler CBV's, selected from among 16 available for Q 10 channel 30. In Figure 16 the selected basis vectors are shown. As discussed in the data pre-processing Section 2.3.1, in certain places systematics are spatially correlated within a module with sharp discontinuities at module edges, or at others spatially correlated across modules. Spatially smooth coefficients  $\mathbf{C}$  are approximately simulated with these properties, and are shown in Figure 16. Simulated lightcurves  $\mathbf{Y}$  are given by  $\mathbf{VC} + \mathbf{N} + \mathbf{A}$ , where  $\mathbf{N}$  is additive Gaussian noise with noise level  $\frac{\sigma(\mathbf{VC})}{4}$  and  $\mathbf{A}$  are simulated astrophysical signals. A total of 300 astrophysical signals are simulated, each injected into a single lightcurve. We simulate (3) forms of astrophysical signals (flare event, sinusoidal and transit) in equal proportion, the form of the simulated astrophysical signals is described Section 2.3.4.0. As compared to experiment 1, only simulated astrophysical variability is present in lightcurves.

A low rank decomposition of the simulated lightcurves  $\mathbf{Y}$  is obtained with PCA, with basis  $\mathbf{V}_{PCA}$  and coefficients  $\mathbf{C}_{PCA}$ , detrended lightcurves are denoted by  $\mathbf{Y}_{PCA}$ . The performance of PCA detrending is compared to our method, where detrended lightcurves are denoted by  $\mathbf{Y}_{ALG}$ . In this experiment we deliberately simulate a high level of overfitting by allowing the fitted model rank to be higher than the simulated systematics as  $K' = 8$ . Since PCA, and our method are studied comparatively, we study i) the relative degree of overfitting occuring for each type of simulated astrophysical



**Figure 15.** The leading PCA coefficient term for each cell of the Kepler sensor is shown, on the left for the Kepler sensor modules defined in Section 2.3.1 and on the right for modules 16, 17, which are the subject of Experiment 1.

signal and ii) our algorithmic performance when over-fitting is definitively possible. The algorithmic performance is examined via noise metrics described in Section 3.1. The experiment steps are summarized as follows:

1. Perform PCA on the simulated lightcurves  $\mathbf{Y}$  to obtain  $\mathbf{V}_{PCA}$  and  $\mathbf{C}_{PCA}$  and detrended lightcurves  $\mathbf{Y}_{PCA}$ .
2. Initialize Algorithm 1 with  $\mathbf{C}_{PCA}$  and obtain solution  $\mathbf{C}_{ALG}$  and detrended lightcurves  $\mathbf{Y}_{ALG}$ .

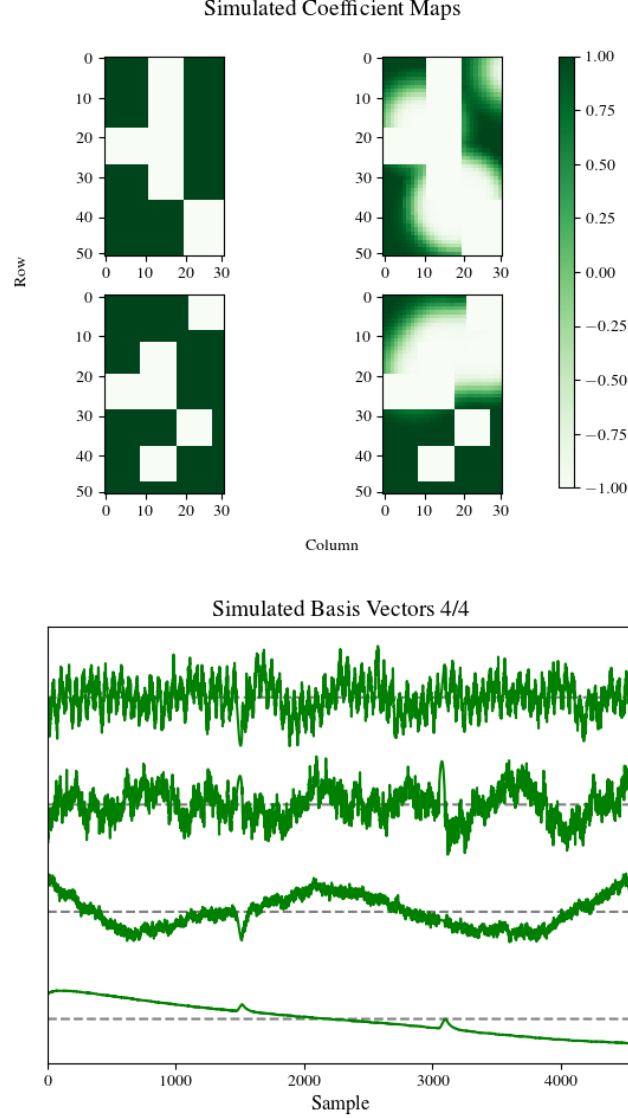
### 3. RESULTS

#### 3.1. Performance metrics

To evaluate detrending one must discriminate between astrophysical variability and residual systematics, consequently evaluation metrics are somewhat subjective [Stumpe et al. \(2012\)](#). The combined differential photometric precision (CDPP) over 6 hrs is used in the Kepler pipeline and serves as an estimate of the level of white noise on a transit timescale. This measures a combination of residual systematics after detrending, shot noise, measurement noise and stellar variability. As described in ([Aigrain et al. \(2016, 2017\)](#)) a full calculation of the CDPP is complex, instead we follow the procedure described in [Gilliland et al. \(2011\)](#) to compute an approximate measure of the CDPP. A 10 day Savitsky-Golay filter is used to remove low-frequency variability from the detrended lightcurves, which are then binned with a 6 hour window. The approximate CDPP is computed as the standard deviation of this binned time series. We use this metric to ensure comparable noise levels between both methods.

To evaluate the degree to which systematics are left or introduced into lightcurves, we compute a goodness-of-fit correlation metric described in [Stumpe et al. \(2012\)](#). This metric is calculated as the average cubed pairwise correlation between all pairs of lightcurves  $\sum_{i \neq j} |c(\mathbf{y}'_i, \mathbf{y}'_j)|^3$ . The absolute correlation is cubed as it is minimal at 0 only if lightcurves are completely flat, which runs counter to the goal of preserving astrophysical variability and hence some level of correlation is expected. Hence by cubing the pairwise absolute correlation, significant correlations contribute more towards the calculated metric relative to small incidental correlations.





**Figure 16.** Experiment 2 simulated systematic noise with dimensionality  $K = 4$  (a): Simulated coefficients; (b): CBV basis vectors used to simulate systematics.

For lightcurves where an astrophysical signal has been injected, post detrending the preservation of the signal is measured by the correlation of the injected signal with the corresponding detrended lightcurve denoted by  $y'$ , the correlation is denoted by  $c(\mathbf{a}_s, \mathbf{y}'_s)$  where  $s \in \text{sin}, t, \text{flr}$ . For either experiment, the reported values are averaged by form of astrophysical signal.

In experiment 1 since we inject a single simulated signal of each form, we also report the absolute correlation of the injected signal with all lightcurves where no astrophysical signal was injected, denoted as  $c(\mathbf{a}_s, \mathbf{y}'_s)$ . Here the absolute correlation is used as the simulated signal may corrupt estimated systematics with a negative scaling. In experiment 2, in addition to ground truth astrophysical signals, the systematics are also known. We report the correlation of the estimated systematics denoted by  $\mathbf{l}'$  (where injected astrophysical signals are subtracted from detrended lightcurves) with the ground truth systematics denoted by  $\mathbf{l}$ . For lightcurves with an astrophysical signal present the averaged correlation is denoted as  $c(\mathbf{l}_s, \mathbf{l}'_s)$  and for all lightcurves  $c(\mathbf{l}, \mathbf{l}')$ .

### 3.2. Experiment 1

Experiment 1 described in Section 2.3.4.0 is evaluated using metrics defined and described in Section 3.1. In Table 1 metrics are averaged over 10 instances of simulated astrophysical signals. Additionally in Table 2, a single instance

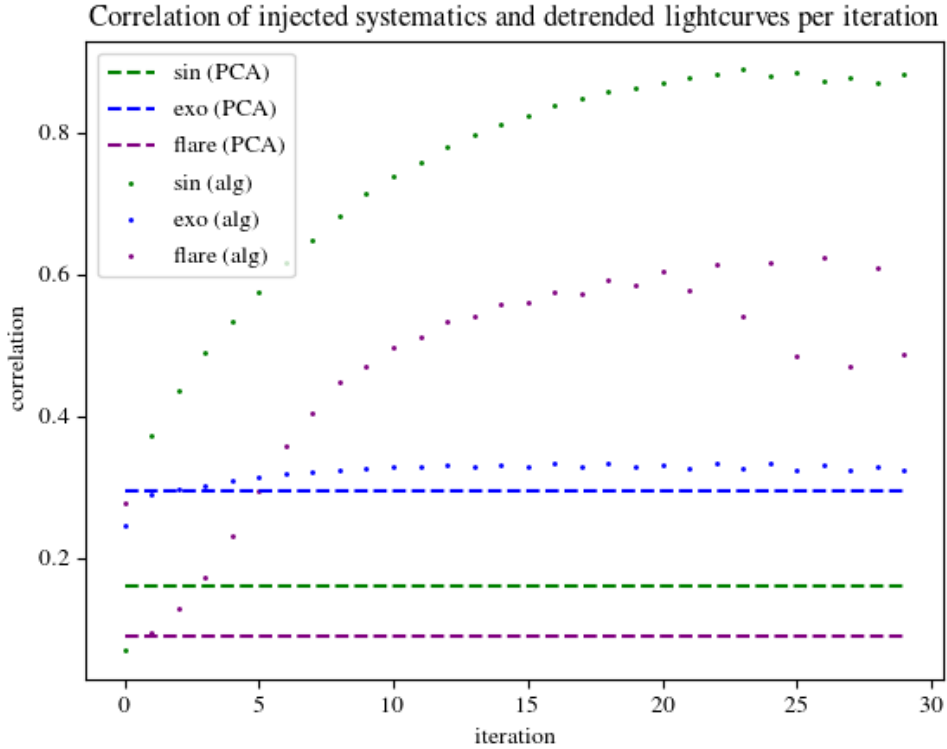
**Table 1.** Experiment 1 (10 randomizations)

	$c(\mathbf{a}_{sin}, \mathbf{y}'_{sin})$	$ c(\mathbf{a}_{sin}, \mathbf{y}'_{sin}) $	$c(\mathbf{a}_t, \mathbf{y}'_t)$	$ c(\mathbf{a}_t, \mathbf{y}'_t) $	$c(\mathbf{a}_{flr}, \mathbf{y}'_{flr})$	$ c(\mathbf{a}_{flr}, \mathbf{y}'_{flr}) $	g-o-f	CDPP 6 hr
PCA	0.097	0.013	0.627	0.034	0.291	0.015	0.031	0.024
Alg	0.385	0.084	0.549	0.04	0.458	0.096	0.033	0.034

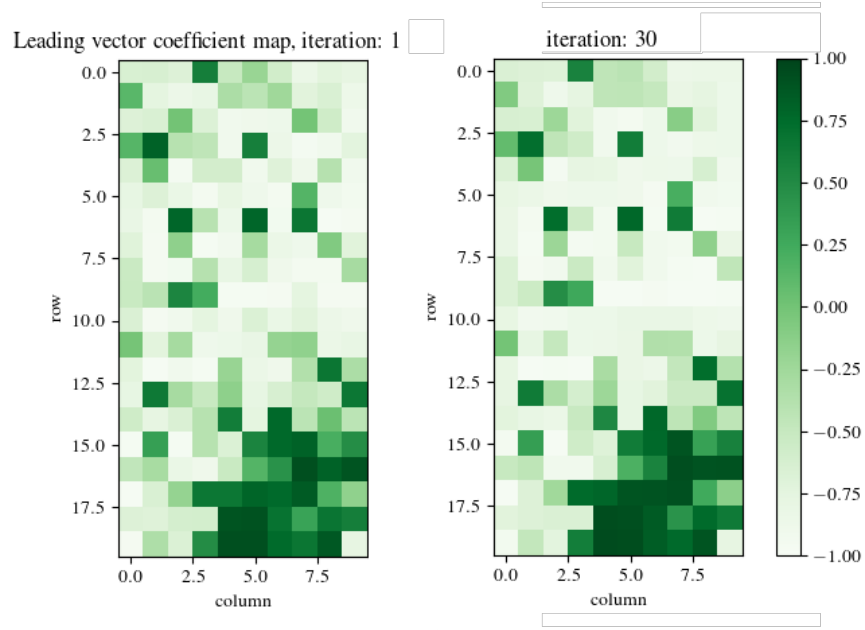
**Table 2.** Experiment 1 results (1 randomization)

	$c(\mathbf{a}_{sin}, \mathbf{y}'_{sin})$	$ c(\mathbf{a}_{sin}, \mathbf{y}'_{sin}) $	$c(\mathbf{a}_t, \mathbf{y}'_t)$	$ c(\mathbf{a}_t, \mathbf{y}'_t) $	$c(\mathbf{a}_{flr}, \mathbf{y}'_{flr})$	$ c(\mathbf{a}_{flr}, \mathbf{y}'_{flr}) $	g-o-f	CDPP 6 hr
PCA	0.16	0.02	0.29	0.027	0.08	0.01	0.001	0.032
Alg	0.88	0.08	0.32	0.032	0.6	0.12	0.002	0.034

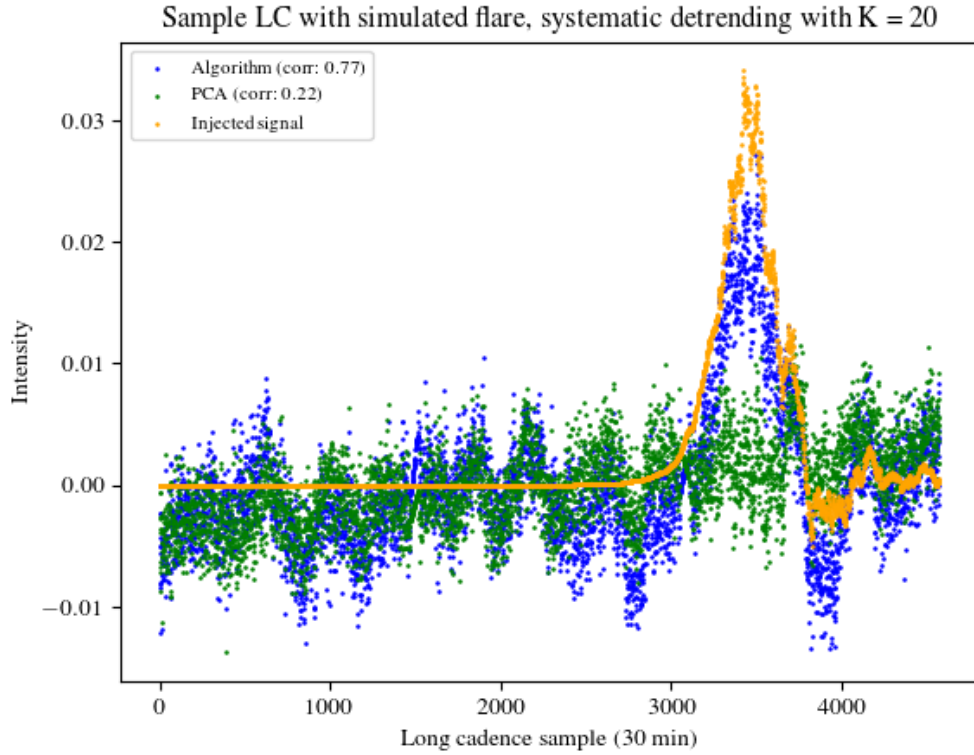
of the experiment with figures is given as an example. In Figure 17 the correlation of the detrended lightcurves and the injected astrophysical signals defined above as  $c(\mathbf{a}_s, \mathbf{y}'_s)$  is shown per iteration for our method, alongside the PCA values. Generally convergence occurred after a small number of steps  $\sim 30$ . In Figures 20, 21 and 19 detrended lightcurves are shown from both our method and PCA alongside the injected astrophysical signals for each of the forms. In Figure 18 the coefficient map of the leading basis term is shown for the first and last iteration.



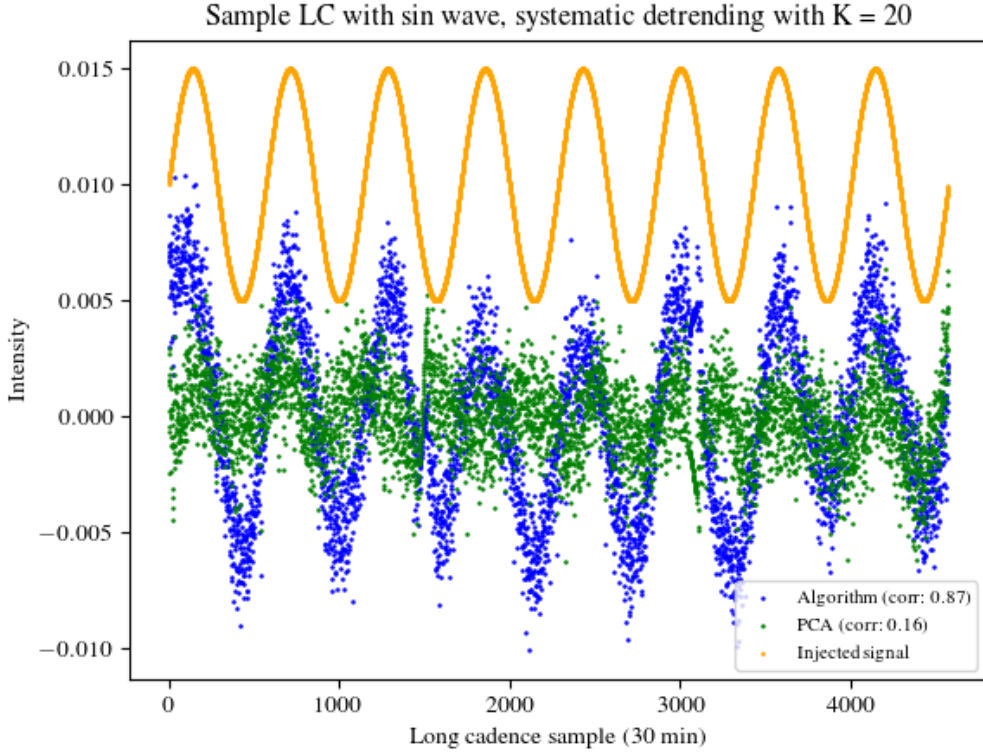
**Figure 17.** Convergence of the correlation between detrended lightcurve and injected astrophysical signal, per iteration. Corresponding data found in Table 2.



**Figure 18.** The coefficient map produced by our method is shown for the first and last iteration. It can be seen that in regions where there is low contrast between values some smoothing has occurred, while outlier values have been generally preserved relative to their neighbours.



**Figure 19.** Experiment 1: An example of a detrended lightcurve with an injected shock signal. Visually our method (shown in blue) has preserved the shock (shown in orange), as compared with the PCA detrended lightcurve (shown in green).



**Figure 20.** An example of a detrended lightcurve with an injected sin wave.

**Table 3.** Experiment 2 results

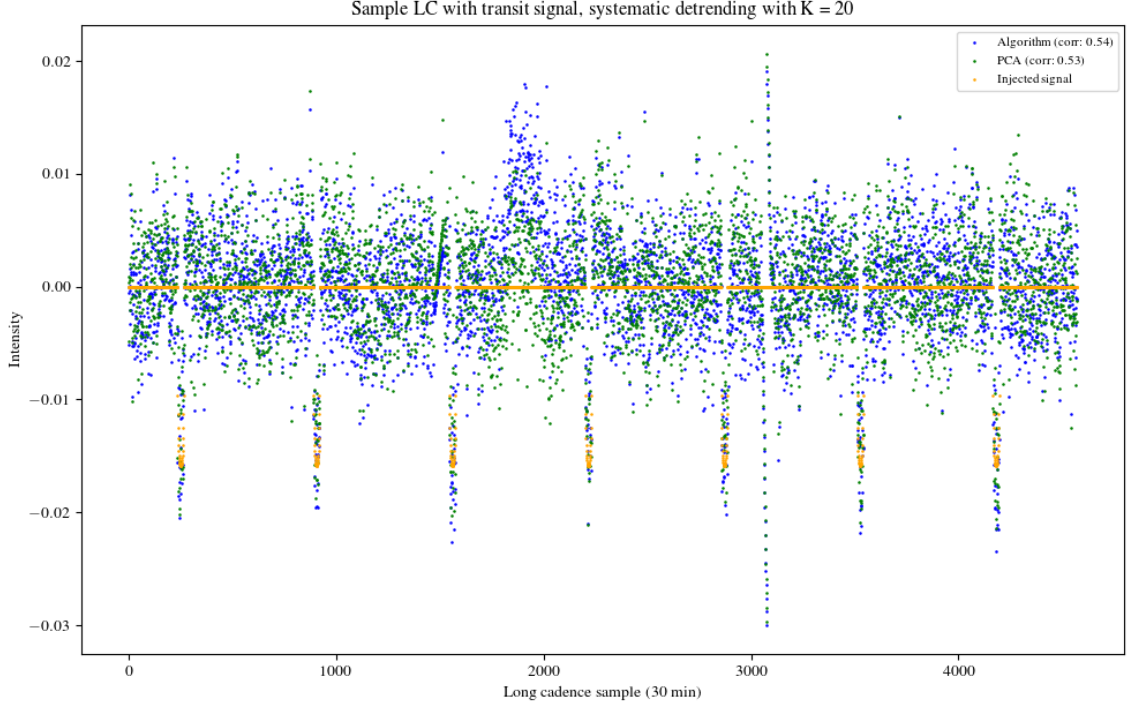
	$c(\mathbf{l}, \mathbf{l}'_s)$	$c(\mathbf{l}, \mathbf{l})$	$c(\mathbf{a}_s, \mathbf{y}'_s)$	Goodness-of-fit	CDPP 6 hr
PCA	0.966	0.994	0.397	0.000153	0.394
Alg	0.983	0.995	0.476	0.000168	0.4

### 3.3. Experiment 2

Experiment 2 is evaluated using metrics defined and described in Section 3.1. In Table 3 metrics are provided for a single randomization since there are a large number (300) of injected and simulated astrophysical signals. In Figure 3.3 the fitted PCA basis vectors are shown, and exhibit features on a range of timescales. In Figure 25 the averaged correlation of the detrended lightcurves and the injected astrophysical signals  $c(\mathbf{a}_s, \mathbf{y}'_s)$  is shown per iteration for our method, alongside the PCA values. Generally convergence occurred after a small number of steps  $\sim 30$  as in experiment 1. In Figure 26 the leading term coefficient map for the first and last iteration is shown.

## 4. DISCUSSION

Results for Experiment 1 are shown in Table 1 and for Experiment 2 shown in Table 3. In Experiment 1 and considering  $c(\mathbf{a}_s, \mathbf{y}'_s)$ , our method as compared to PCA produced detrended lightcurves which were more strongly correlated with injected astrophysical signals with a ratio of 1.37 : 1 (averaged over 10 iterations). In Experiment 2 we similarly confirmed found that our method compared to PCA  $c(\mathbf{a}_s, \mathbf{y}'_s)$  with a ratio of 3.39 : 1. This metric should not be used as a measure of absolute performance however, due to the presence of other unknown lightcurve components and the somewhat unrealistic nature of simulated astrophysical components. Herein this metric is primarily used for the purposes of comparison. Note that the correlation of injected astrophysical signal and detrended lightcurve is not



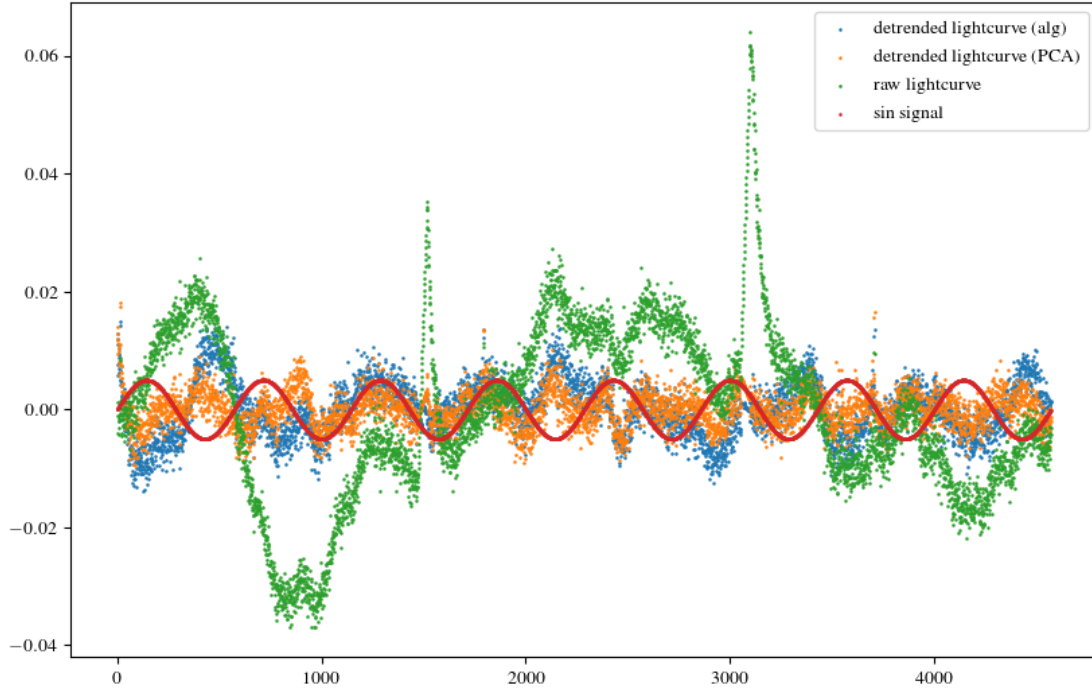
**Figure 21.** An example of a detrended lightcurve with an injected transit signal.

expected to be 1 even with perfect recovery, due to the magnitude of other signals present in lightcurves.

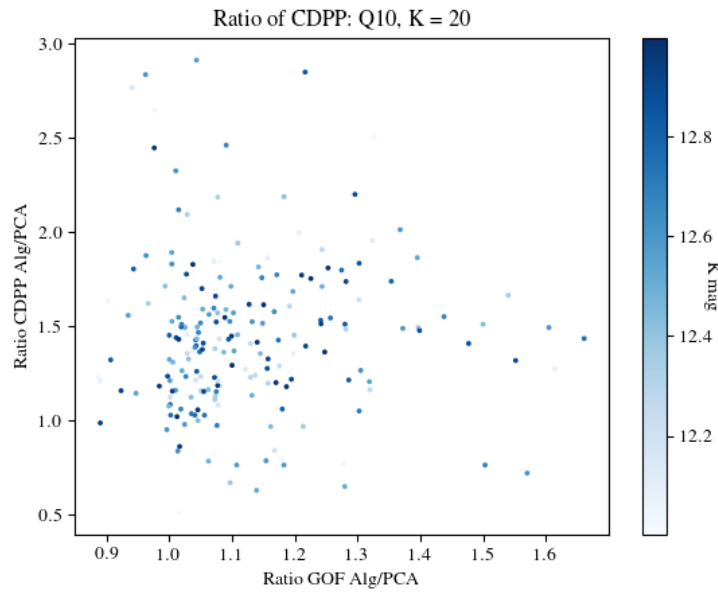
Broken down by signal form, sin and flare type signals were both more susceptible to overfitting by PCA detrending, shown by low values of  $c(\mathbf{a}_{sin}, \mathbf{y}'_{sin})$   $c(\mathbf{a}_{flr}, \mathbf{y}'_{flr})$ . Comparatively transit signals were the least overfit by PCA, with a higher value of  $c(\mathbf{a}_t, \mathbf{y}'_t)$  than for our method. Simulated transit signals have a short duration compared to the timescale of systematics, and therefore are less linearly dependent on an inferred systematics basis. Comparatively our method reduced the level of overfitting for sin and flare signals, with a ratio of  $c(\mathbf{a}_{sin}, \mathbf{y}'_{sin})$  in Experiment 1 between our method and PCA of 4 : 1 and  $c(\mathbf{a}_{flr}, \mathbf{y}'_{flr})$  of 1.7 : 1. However for transit signals in Experiment 1 our method performed comparatively worse, the ratio of  $c(\mathbf{a}_t, \mathbf{y}'_t)$  given by 0.88 : 1. A caveat of this metric is that if a greater amount of the other unknown signal components are preserved, which may include unknown astrophysical variability, the correlation between the detrended lightcurve and the injected signal will be proportionally smaller. Furthermore since PCA is guaranteed to maximally flatten a lightcurve, our method can only increase overall lightcurve variability, thus slightly lowering the value of this metric even if astrophysical signals have been preserved.

In Experiment 1, both our method and PCA produce low values of correlation of injected astrophysical signals with non-injected lightcurves  $|c(\mathbf{a}_s, \mathbf{y}'_s)|$ , however our method produces higher values compared with PCA. As described above, increased lightcurve variability of our method compared to PCA may seem to cause worse correlation metrics as variability may be incidentally correlated with injected signals. However, it may also indicate that our method has corrupted non-injected lightcurves, with injected astrophysical signals. We visually inspect the most highly correlated lightcurve (from the non-injected set) with the injected sin signal, shown in Figure 22, the raw lightcurve is shown in green alongside the sin signal with which the lightcurve is correlated, in orange the lightcurve detrended via PCA is shown and with our method in blue. In this example the detrended lightcurve produced by our method is 3.5 times more correlated with the sin signal than the PCA detrended lightcurve. However, where the lightcurve is strongly correlated with the sin signal (i.e. between cadences 1000-2000) can be seen to be features in the original lightcurve (prior to signal injection) and therefore the high value of  $|c(\mathbf{a}_{sin}, \mathbf{y}'_{sin})|$  may in this case be caused by incidental

Most strongly correlated detrended lightcurve (among non-injected lightcurves) with injected sin wave, correlation ratio (ALG/PCA): 3.51

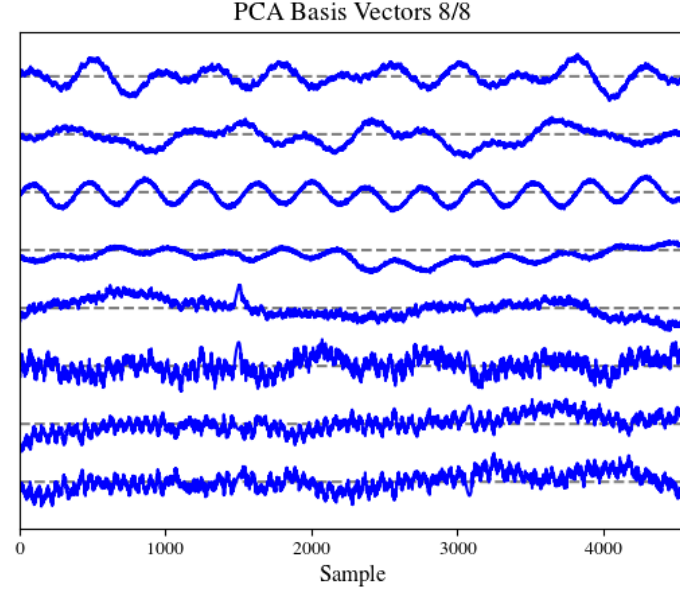


**Figure 22.** Most strongly correlated detrended lightcurve, from among non-injected lightcurves, with injected sin wave.

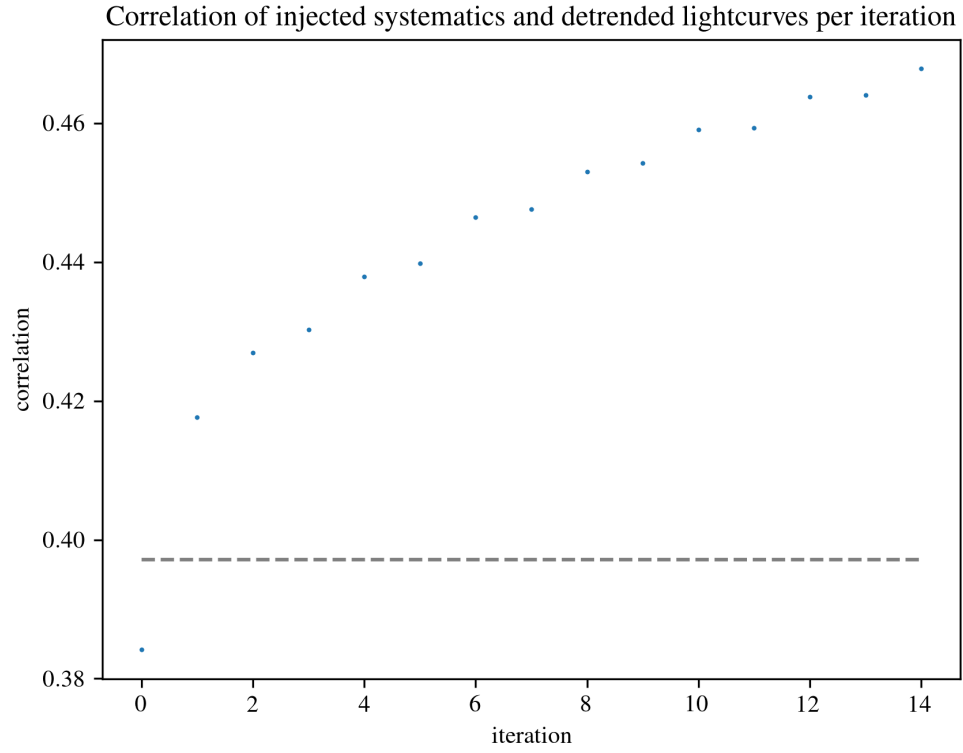


**Figure 23.** Experiment 1: Ratio of CDPP to GOF metrics of our method to PCA. Our method has higher values of CDPP and GOF for most targets, indicating a greater level of scatter in lightcurves detrended via our method.

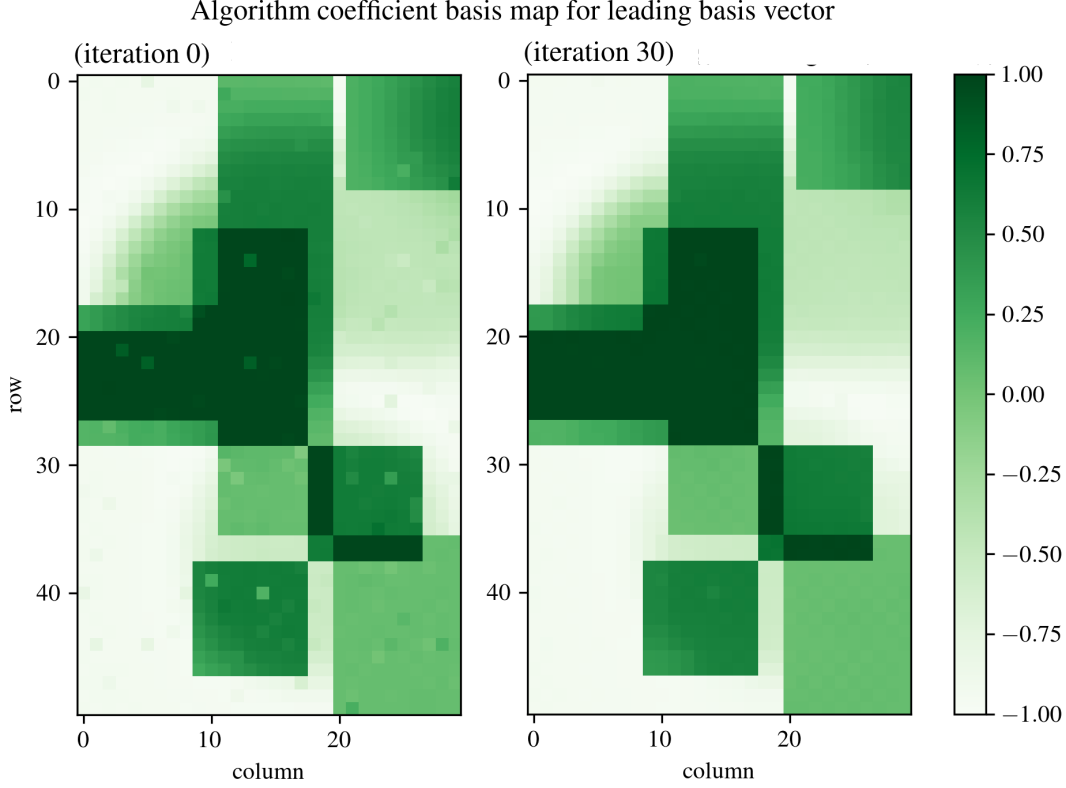




**Figure 24.** PCA fitted basis vectors with  $K' = 8$



**Figure 25.** Experiment 2 iteration



**Figure 26.** Experiment 2: Our method is initialized with the PCA fitted coefficients shown on the left at iteration 0. Each cell represents a fitted value, however the PCA values show speckling, which indicates a cell has been overfit to an injected astrophysical signal, as the coefficients should be spatially smooth. After 30 iterations of our method the obtained coefficients are spatially smooth, thereby overfitting of coefficients to injected signals is reduced.

correlation with unknown lightcurve variability.

In Experiment 2, the correlation of estimated systematics and simulated ground truth systematics  $c(\mathbf{l}, \mathbf{l}'_s)$  in Table 3 is marginally greater for our method as compared to PCA. In particular, this effect is due to lightcurves where an astrophysical signal was injected, for which the PCA estimated systematics were overfit to the injected astrophysical signals. In Figure 26 the leading term coefficient map for the first and last iteration is shown. In the first iteration, our method is initialized with the PCA solution, where an astrophysical signal was injected cells have been overfit as can be seen by speckled values. In the last iteration these cells have been 'corrected' so that regions are spatially smooth as in the simulated coefficient maps in Figure 2.3.4.0. Furthermore our method produces higher correlation between injected astrophysical signals and detrended lightcurves  $c(\mathbf{a}_s, \mathbf{y}'_s)$  as was also demonstrated for Experiment 1.

In Experiment 1 and 2, the goodness-of-fit metric is marginally larger for our method which indicates that the detrended lightcurves have a higher level of residual correlation. However in Experiment 2 estimated systematics were shown to be more accurate, thus it is possible that this slightly increased goodness-of-fit metric is due to increased lightcurve variability. Our method also has a higher level of 6-hr CDPP suggesting more scatter due to our method. Figure 21 shows a detrended lightcurve from Experiment 1, in certain instances our method left in large residual scatter at Earthpoint downlink times. In Figure 23 the ratio between our method and PCA of the CDPP values is plotted against the ratio of the goodness-of-fit values, where each target is plotted with its Kepler magnitude for targets from Experiment 1. In general it can be seen that for most targets our method produces larger CDPP/goodness-of-fit values, however the scatter is concentrated close to 1 (with less scatter in goodness), no strong pattern associated with Kepler magnitude is observed.

We investigated and presented evidence that lightcurves exhibit spatially correlated systematics and furthermore, in certain regions neighbouring lightcurves are more meaningful predictors of systematics in a lightcurve. We present a method to estimate systematics where spatial correlation is used as a constraint. Our systematics model is low rank and iteratively refined constrained by a spatial-correlation prior. Spatially proximal pixels on the sensor exhibit highly correlated systematics, hence our spatial correlation prior is based on the total variation functional [Rudin et al. \(1992\)](#) applied to the normalized weight vectors, mapped by 2D sensor location. This model is agnostic to the presence of astrophysical signals but demonstrates robustness to overfitting of unknown astrophysical components. Experiment 1 studies algorithmic performance for real lightcurves, Experiment 2 on the otherhand studies performance for simulated systematics under the assumed noise model. Overall experiment 1 and experiment 2 obtained complementary results, demonstrating through metrics and visually non-negligible improvements in overfit signals as compared with least-square methods.

Care was taken to fairly examine both methods, and use realistic systematics, or real lightcurves. However we do not claim general applicability of our method. As described within, the results and performance are highly dependent on a number of tunable parameters (model order, convergence criterion, sensor discretization, weighting matrix, step size), as well as the form of simulated astrophysical signal. For example herein the simulated transit signals have a small duration ( $\sim 20$  lc samples for a 10hr transit) as compared to the number of samples over which systematics are modeled ( $\sim 4000$  samples) with a ratio (1:200), in data where the duration of transits is relatively longer i.e. comparing a 10hr transit in TESS data to a single sector of data a ratio of (1:72) there may be a higher level of overfitting.

There are several drawbacks to our method which will need to be addressed in future work. Namely the uniform sensor layout over which the spatial correlation model is built. Additionally if lightcurves are not sufficiently spaced apart on the sensor, their point-spread-functions may overlap, leading to astrophysical leakage. This will present an issue for our method as astrophysical content will be spatially correlated and can corrupt the systematics model, although if only two lightcurves contain the occurring signal, this issue can be handled by the weighting/L1 prior. In our work the distance between lightcurves, due to the sensor discretization, is much larger than the PSF width.

As described in Section 2.3.1, a refined set of lightcurves (within a fixed magnitude band) with low variability was used to solve for the systematics model as systematics are somewhat magnitude dependent. More generally our approach of incorporating a prior into a low-rank systematics model may be expanded to include other priors, for example a magnitude dependency. Future work will incorporate ancillary priors of this sort, herein we have demonstrated low-rank inference from first principles focusing on a single prior.

We hypothesize that this model can prevent overfitting, the key assertion being that if nearby lightcurves are more informative of a particular lightcurve than uniform selection, then our model can utilize this information. Furthermore by introducing the constraint foundationally, the full systematics model (basis and coefficients) may iteratively refine a basis and coefficients, better suited to the spatial prior. We do not show general optimality over methods such as PDC-MAP where a basis is fit conditioned on a Bayesian prior. However the described method of spatial-systematics inference does not require the explicit construction of a basis or prior, and iteratively refines systematic components as compared to a single Bayesian estimate.

## 5. CONCLUSION

We developed a method for systematics inference in lightcurves based on a low-rank linear model which is conditioned by a spatial prior at a foundational level. This prior encourages reliance of systematic estimates for a lightcurve on neighbouring lightcurves, to mitigate overfitting of astrophysical signals in contrast to singularly relying on minimizing a least-squares residual for the lightcurve under consideration. The spatial prior calculates a measure of total variation of normalized coefficient vectors organized by target (x, y) sensor location, and includes tunable parameters which control the discontinuity and magnitude of the spatial smoothness. A variable-reduced objective function is formulated which is based on the closed form solution of basis vector terms in alternating minimization and depends only on the systematic coefficients. The low-rank systematics are fitted to lightcurves by minimizing the variable-reduced objective with gradient descent using derived gradient steps in Appendix 6.5. The performance of our systematics inference

method was evaluated in two forms of experiment i) using real lightcurves with simulated astrophysical signals and ii) using simulated lightcurves with simulated systematics and astrophysical signals as well as white noise. The principle conclusions of the paper are:

1. The spatially correlated systematics model demonstrated reduced overfitting of simulated astrophysical signals as compared to a standard least-squares low rank systematics model. In Experiment 1 our method compared to PCA produced detrended lightcurves which were more strongly correlated with injected astrophysical signals with a ratio of 1.37 : 1 (averaged over 10 iterations). In Experiment 2 we similarly confirmed found that our method compared to PCA with a ratio of 3.39 : 1.
2. The main reduction in overfitting was found for slow-varying astrophysical signals, comparable performance of both methods was found for transit signals. We posit that this is because there is a greater degree of inherent correlation between systematics as found in Kepler lightcurves, and the form of simulated slow-varying astrophysical signals.
3. In lightcurves where spatially-correlated systematics were simulated and of known form, our model more accurately recovers such systematics than a standard least-squares low rank systematics model. In real lightcurves our model can benefit from the inclusion of other known parametric dependencies (such as Kepler magnitude).

## 6. APPENDIX

### 6.1. Notation

We define some commonly used notation here. The  $i$ 'th row of matrix  $\mathbf{M}$  is denoted as  $[\mathbf{M}]_i$ , the  $j$ 'th row of matrix  $\mathbf{M}$  is denoted as  $[\mathbf{M}]_{.,j}$ . The entry corresponding to the  $i$ 'th row and  $j$ 'th column is denoted as  $[\mathbf{M}]_{i,j}$ . The transpose of a matrix  $\mathbf{M}$  is denoted as  $\mathbf{M}^T$ . An identity of size  $N$  is denoted as  $\mathbf{1}_N$ .

### 6.2. Functions of Bounded Total variation

For an overview of total variation topics see the monograph by [Vogel \(2002\)](#) and Chapter 3.6 [Karl \(2005\)](#). In this section we provide a general description of functions of bounded total variation, which the spatial prior is based on.

A function of bounded variation is defined as being integrable,  $f \in L^1$  and of bounded total variation  $TV(f) \leq \infty$ . We restrict the normalized coefficient vectors as a function of sensor position to belong to this class and use the associated total variation measure as a spatial smoothness prior. Specifying the total variation measure for a function  $f(x, y)$ , with partial derivative vector  $\nabla f = \left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right)$ :

$$TV(f) = \int \int \|\nabla f\|_2 dx dy \quad (23)$$

A discretized approximation of the total variation measure for  $f$  sampled at points  $X \times Y$  is given by:

$$TV_{discrete}(f) = \sum_{i \in X} \sum_{j \in Y} \|(D_x f)_{i,j}, (D_y f)_{i,j}\|_2 \quad (24)$$

Where  $(D_x f)_{i,j} = f_{i,j} - f_{i-1,j}$  and  $(D_y f)_{i,j} = f_{i,j} - f_{i,j-1}$ . If matrix  $Df$  has  $vec(D_x f)$  as the first row and  $vec(D_y f)$  the second row (where  $vec$  denotes vectorization), then we equivalently obtain:

$$TV_{discrete}(f) = \|Df\|_{2,1}^1 \quad (25)$$

A function of bounded total variation may have discontinuities, and when solving for a minimizing  $f$  of the TV measure, the solutions will be mostly spatially uniform with sharp changes in the function value. However, a generalization of this measure is given by  $\|Df\|_{2,p}^p : p \in [1, 2]$ . Taking  $p = 2$  the measure will obtain solutions  $f$  which are gradually spatially varying. See Figure 2.3.3 for an example where the different values of  $p$  obtain solutions with different smoothness properties.

### 6.3. Transformed Correlation

We sketch the proof for the bound in Equation 15, see <sup>6</sup> for details of the trigonometric calculation. Without loss of generality, assume  $\mathbf{c}_i$  and  $\mathbf{c}_j$  are unit vectors. A rank  $K$  SVD of the basis vectors  $\mathbf{V}$  always exists, denote the left singular vectors as  $\mathbf{U}$  and singular values  $\Sigma$ . Writing the systematics correlation using the SVD of  $\mathbf{V}$ :

$$\frac{\mathbf{c}_i^T \mathbf{V}^T \mathbf{V} \mathbf{c}_j}{\|\mathbf{V} \mathbf{c}_i\| \|\mathbf{V} \mathbf{c}_j\|} = \frac{\mathbf{c}_i^T \mathbf{U}^T \Sigma^2 \mathbf{U} \mathbf{c}_j}{\|\Sigma \mathbf{U} \mathbf{c}_i\| \|\Sigma \mathbf{U} \mathbf{c}_j\|} \quad (26)$$

Without loss of generality, assume vectors  $\mathbf{c}_i, \mathbf{c}_j$  are unit norm  $\|\mathbf{c}_i\| = \|\mathbf{c}_j\| = 1$ . The vectors  $\mathbf{c}_i, \mathbf{c}_j$  lie in a plane in  $K$  dimensional space and are at an angle  $\phi : \cos(\phi) = \mathbf{c}_i^T \mathbf{c}_j$ . We can see from the SVD decomposition that  $\mathbf{V} \mathbf{c}_i$  projects  $\mathbf{c}_i$  onto orthonormal vectors  $\{\mathbf{u}_k : k \in K\}$  and scales the result by the corresponding  $\{\sigma_k : k \in K\}$ . Geometrically the maximum deviation occurs when the coefficients  $\mathbf{c}_i$  and  $\mathbf{c}_j$  lie in the span of  $\mathbf{u}_1$  and  $\mathbf{u}_k$ , where the maximal scaling of one dimension occurs relative to the other i.e.  $\mathbf{c}_i = \alpha \mathbf{u}_1 + \beta \mathbf{u}_k$  and  $\mathbf{V} \mathbf{c}_i = \sigma_1 \alpha \mathbf{u}_1 + \sigma_k \beta \mathbf{u}_k$ . Since this case is just defined by the span of  $\mathbf{u}_1, \mathbf{u}_k$  and the two vectors  $\mathbf{c}_i$  and  $\mathbf{c}_j$ , the worst case orientation of  $\mathbf{c}_i$  and  $\mathbf{c}_j$  to produce the maximum deviation of the angle between  $\bar{\mathbf{l}}_i$  and  $\bar{\mathbf{l}}_j$  relative to this plane can be found irrespective of the rank  $K$ . Define the angle between  $\mathbf{u}_1$  and the vector which bisects  $\mathbf{c}_i$  and  $\mathbf{c}_j$  as  $\theta$ , this represents the relative rotation of the coefficient vectors in the plane spanned by  $\mathbf{u}_1$  and  $\mathbf{u}_k$ . The maximal expansion of the correlation angle  $\phi$  occurs when  $\theta = 0$  (i.e. the coefficient vectors are bisected by  $\mathbf{u}_1$ ) and the maximal compression of the correlation angle  $\phi$  occurs when  $\theta = \frac{\pi}{2}$  (i.e. the coefficient vectors are bisected by  $\mathbf{u}_k$ ). Using trigonometry, these cases give the upper and lower bound for the systematics correlation in Equation 26 as a function of the angle between  $\mathbf{c}_i$  and  $\mathbf{c}_j$  i.e.  $\cos(\phi) = \mathbf{c}_i^T \mathbf{c}_j$ .

### 6.4. Objective function via Variable Projection

In this section we use variable projection to eliminate variable dependence of least-squares penalty on  $\mathbf{V}$ , so that it is only a function of  $\mathbf{C}$ . For an overview of variable projection applied to separable non-linear least-squares (least-squares can be seen as a special case), see (Golub & Pereyra (2003)). For simplicity, we work with the transposed version of the terms inside the least-squares penalty. Denoting the least-squares penalty as  $f(\mathbf{V}, \mathbf{C})$  as in Section 2.2:

$$f(\mathbf{V}, \mathbf{C}) = \|\mathbf{Y} - \mathbf{V} \mathbf{C}\|_F^2 = \|\mathbf{Y}^T - \mathbf{C}^T \mathbf{V}^T\|_F^2 \quad (27)$$

Denote the minimizing  $\mathbf{V}^T$  of this penalty for a particular  $\mathbf{C}$  as  $h(\mathbf{C})^T = \arg \min_{\mathbf{V}^T} \|\mathbf{Y} - \mathbf{C}^T \mathbf{V}^T\|_F^2$ . The minimum  $\mathbf{V}$  is given by the minimum-norm least-squares solution, which is known to be computed as the pseudoinverse of  $\mathbf{C}^T$  applied to  $\mathbf{Y}^T$  (see the monograph by Golub & Van Loan (2013)). Denoting the pseudoinverse of  $\mathbf{C}^T$  as  $\mathbf{C}^\dagger = (\mathbf{C} \mathbf{C}^T)^{-1} \mathbf{C}$  we obtain  $h(\mathbf{C})^T = \mathbf{C}^\dagger \mathbf{Y}^T$ . Rewriting the first term using the conditional solution:

$$f(h(\mathbf{C}), \mathbf{C}) = \|\mathbf{Y}^T - \mathbf{C}^T h(\mathbf{C})^T\|_F^2 \quad (28)$$

$$= \|\mathbf{Y}^T - \mathbf{C}^T \mathbf{C}^\dagger \mathbf{Y}^T\|_F^2 \quad (29)$$

$$= \|(\mathbf{1}_I - \mathbf{C}^T \mathbf{C}^\dagger) \mathbf{Y}^T\|_F^2 \quad (30)$$

Noting that  $\mathbf{C}^T \mathbf{C}^\dagger = \mathcal{P}_{R(\mathbf{C}^T)}$  is the projection onto the range space of  $\mathbf{C}^T$ . Furthermore,  $\mathbf{1}_I$  is an  $I \times I$  identity matrix, so  $(\mathbf{1}_I - \mathcal{P}_{R(\mathbf{C}^T)}) = \mathcal{P}_{R(\mathbf{C}^T)^\perp}^\perp$  is the projection onto the orthogonal complement to the range space of  $\mathbf{C}^T$ . The variable reduced least-squares penalty is equivalently stated as:

$$f(h(\mathbf{C}), \mathbf{C}) = \|\mathcal{P}_{R(\mathbf{C}^T)^\perp}^\perp \mathbf{Y}^T\|_F^2 \quad (31)$$

### 6.5. Gradient Descent

We use gradient descent to minimize  $\|(\mathbf{1}_I - \mathbf{C}^T \mathbf{C}^\dagger) \mathbf{Y}^T\|_F^2 + \|\mathbf{D} \mathbf{w} \tilde{\mathbf{C}}^T\|_{2,p}^p$  the following function with respect to  $\mathbf{C}^T \in \mathbb{R}^{I \times K}$ . The gradient  $\nabla(\mathbf{C}^T) \in \mathbb{R}^{I \times K}$  is the matrix of partial derivatives. We denote the gradient of the variable projection least-squares term as  $\nabla_{VP}(\mathbf{C}^T)$  and the gradient of the total-variation prior as  $\nabla_{TV}(\mathbf{C}^T)$  such that the overall gradient is:

$$\nabla(\mathbf{C}^T) = \nabla_{VP}(\mathbf{C}^T) + \nabla_{TV}(\mathbf{C}^T) \quad (32)$$

Note  $\nabla(\mathbf{C})^T = \nabla(\mathbf{C}^T)$ .

<sup>6</sup> <https://stats.stackexchange.com/questions/205099/how-does-cosine-similarity-change-after-a-linear-transformation>

### 6.5.1. Variable projection least-squares gradient

For clarity in computing  $\nabla_{VP}(\mathbf{C}_i^T)$ , expand the variable projection least squares penalty as  $\|P_{R(\mathbf{C}^T)}^\perp \mathbf{Y}^T\|_F^2 = \sum_N \|P_{R(\mathbf{C}^T)}^\perp [\mathbf{Y}]_n\|^2$  where  $[\mathbf{Y}]_n$  is the  $n^{th}$  row of  $\mathbf{Y} \in \mathbb{R}^{N \times I}$ . The gradient of this term is obtained by the chain rule, and using the symmetry of the projection operator, the full proof is given in (Harville (2008)). The gradient  $\nabla_{VP}(\mathbf{C}^T) \in \mathbb{R}^{I \times K}$  of the objective is:

$$\nabla_{VP}(\mathbf{C}^T) = -P_{R(\mathbf{C}^T)}^\perp \left( \sum_n [\mathbf{Y}]_n [\mathbf{Y}]_n^T \right) (\mathbf{C}^\dagger)^T \quad (33)$$

This can be compactly expressed as:

$$\nabla_{VP}(\mathbf{C}^T) = -P_{R(\mathbf{C}^T)}^\perp \mathbf{Y}^T \mathbf{Y} (\mathbf{C}^\dagger)^T \quad (34)$$

### 6.5.2. Total variation gradient

The derivative of  $\|\mathbf{D}_w \bar{\mathbf{C}}^T\|_{2,p}^p$  can be computed by application of the chain rule. For brevity we refer to  $\mathbf{D}_w$  as simply  $\mathbf{D}$ .

$$\frac{d}{d\bar{\mathbf{C}}^T} \|\mathbf{D} \bar{\mathbf{C}}^T\|_{2,p}^p = \frac{d}{d\bar{\mathbf{C}}^T} \|\mathbf{D} \bar{\mathbf{C}}^T\|_{2,p}^p \cdot \frac{d\bar{\mathbf{C}}^T}{d\bar{\mathbf{C}}^T} \quad (35)$$

The matrix  $\mathbf{D} \bar{\mathbf{C}}^T \in \mathbb{R}^{2 \times K(X-1)(Y-1)}$  has columns for each coefficient  $k \in K$  and each pixel  $j \rightarrow (x, y) \in (X-1) \times (Y-1)$ , formed as  $[[\mathbf{D}_x \bar{\mathbf{C}}^T]_{j,k}, [\mathbf{D}_y \bar{\mathbf{C}}^T]_{j,k}] = [\bar{c}_k^{x,y} - \bar{c}_k^{x+1,y}, \bar{c}_k^{x,y} - \bar{c}_k^{x,y+1}]$  where  $\mathbf{D}_x$  and  $\mathbf{D}_y$  are difference matrices in  $x$  and  $y$  respectively.

The total variation penalty computes the  $L_p$  norm over the  $L_2$  norm of each column of  $\mathbf{D} \bar{\mathbf{C}}^T$ , given by  $\|\mathbf{D} \bar{\mathbf{C}}^T\|_{2,p}^p = \sum_n \|[\mathbf{D} \bar{\mathbf{C}}^T]_{\cdot,n}\|_2^p = \sum_k \sum_j \|[\mathbf{D}_x \bar{\mathbf{C}}^T]_{j,k}, [\mathbf{D}_y \bar{\mathbf{C}}^T]_{j,k}\|_2^p$ . Each  $n$  maps to some  $j, k$ , but the ordering does not matter.

The  $L_p$  norm is non-differentiable at zero for  $p \in [1, 2)$ , a differentiable approximation of the  $L_p$  norm is used in place known as the Huber functional, which prevents a zero being reached by adding a very small positive value  $\delta$  to every element. A description of the Huber functional and it's use in total variation denoising is given in the monograph by Vogel (2002). We apply the Huber functional to arrive at the modified penalty:

$$\|\mathbf{D} \bar{\mathbf{C}}^T\|_{2,p}^p \approx \sum_n (\|[\mathbf{D} \bar{\mathbf{C}}^T]_{\cdot,n}\|_2^2 + \delta)^{\frac{p}{2}} \quad (36)$$

Where

$$\|[\mathbf{D} \bar{\mathbf{C}}^T]_{\cdot,n}\|_2^2 = [\mathbf{D}_x \bar{\mathbf{C}}^T]_{j,k}^2 + [\mathbf{D}_y \bar{\mathbf{C}}^T]_{j,k}^2 \quad (37)$$

The derivative is decomposable with  $k$ , for simplicity we find the derivative with respect to each column of  $[\bar{\mathbf{C}}^T]_{\cdot,k}$ . First  $\frac{d[\mathbf{D}_x \bar{\mathbf{C}}^T]_{j,k}^2}{d[\bar{\mathbf{C}}^T]_{\cdot,k}} = 2[\mathbf{D}_x \bar{\mathbf{C}}^T]_{j,k} [\mathbf{D}_x]_j^T$ . Using the chain rule

$$\frac{d\|\mathbf{D} \bar{\mathbf{C}}^T\|_p}{d[\bar{\mathbf{C}}^T]_{\cdot,k}} = p \sum_j ([\mathbf{D}_x \bar{\mathbf{C}}^T]_{j,k}^2 + [\mathbf{D}_y \bar{\mathbf{C}}^T]_{j,k}^2)^{\frac{p}{2}-1} ([\mathbf{D}_x \bar{\mathbf{C}}^T]_{j,k} [\mathbf{D}_x]_j^T + \mathbf{D}_y \bar{\mathbf{C}}^T]_{j,k} [\mathbf{D}_y]_j^T) \quad (38)$$

Denote  $\mathbf{L}_k = \text{diag} \left( p([\mathbf{D}_x \bar{\mathbf{C}}^T]_{j,k}^2 + [\mathbf{D}_y \bar{\mathbf{C}}^T]_{j,k}^2)^{\frac{p}{2}-1} \forall j \right)$ . The sum over  $j$  allows the expression to be written compactly as:

$$\frac{d\|\mathbf{D} \bar{\mathbf{C}}^T\|_p}{d[\bar{\mathbf{C}}^T]_{\cdot,k}} = (\mathbf{D}_x^T \mathbf{L}_k \mathbf{D}_x + \mathbf{D}_y^T \mathbf{L}_k \mathbf{D}_y) [\bar{\mathbf{C}}^T]_{\cdot,k} \quad (39)$$

The final term in Equation 35 is the derivative due to the row-wise normalization of  $\bar{\mathbf{C}}^T$ .

$$\frac{d[\bar{\mathbf{C}}^T]_i}{d[\bar{\mathbf{C}}^T]_j} = \begin{cases} (\mathbf{1}_K - [\bar{\mathbf{C}}^T]_i [\hat{\mathbf{C}}^T]_i^T) \frac{1}{\|[\bar{\mathbf{C}}^T]_i\|} & i = j \\ 0 & i \neq j \end{cases} \quad (40)$$



Each row  $i$  of the final derivative is the product of  $\left[\frac{d\|\mathbf{D}\bar{\mathbf{C}}^T\|_p}{d[\bar{\mathbf{C}}^T]_i}\right]_i$  with  $\frac{d[\bar{\mathbf{C}}^T]_i}{d[\bar{\mathbf{C}}^T]_i}$  such that

$$[\nabla_{TV}(\mathbf{C}^T)]_i = [\mathbf{B}_1[\bar{\mathbf{C}}^T]_{,1}, \dots, \mathbf{B}_K[\bar{\mathbf{C}}^T]_{,K}]_i (I_K - [\bar{\mathbf{C}}^T]_i [\hat{\mathbf{C}}^T]_i^T) \frac{1}{\|[\hat{\mathbf{C}}^T]_i\|} \quad (41)$$

Where  $\mathbf{B}_k = \mathbf{D}_x^T \mathbf{L}_k \mathbf{D}_x + \mathbf{D}_y^T \mathbf{L}_k \mathbf{D}_y$ .

For  $p = 2$  an analytic solution can be computed without the use of the Huber functional approximation.

$$[\nabla_{TV}(\mathbf{C}^T)]_i = [(\mathbf{D}_x^T \mathbf{D}_x + \mathbf{D}_y^T \mathbf{D}_y) \bar{\mathbf{C}}^T]_i (\mathbb{1}_K - [\hat{\mathbf{C}}^T]_i [\hat{\mathbf{C}}^T]_i^T) \frac{1}{\|[\hat{\mathbf{C}}^T]_i\|} \quad (42)$$

It can be seen that the effect of the  $L_p$  norm for  $p \in [1, 2)$  is contained in the diagonal weightings  $\mathbf{L}_k$ .

### 6.6. Probabilistic view of the objective

Although the method described was not derived from a Bayesian framework, a probabilistic interpretation of the objective function can be shown. For an overview of Bayesian estimation and low rank models see the monograph [Murphy \(2020\)](#). Here we provide a general description of this equivalence for our objective. In Bayesian MAP estimation, one optimizes the posterior probability of the form:

$$\arg \max_{\mathbf{C}} p(\mathbf{C}|\mathbf{Y}) \propto \arg \max_{\mathbf{C}} p(\mathbf{Y}|\mathbf{C}) p(\mathbf{C}) \quad (43)$$

$$\equiv \arg \min_{\mathbf{C}} \ln(p(\mathbf{Y}|\mathbf{C})) + \ln(p(\mathbf{C})) \quad (44)$$

The signal model in Equation 1,  $\mathbf{Y} = \mathbf{V}\mathbf{C} + \mathbf{N}$ , where  $\mathbf{N}$  is white Gaussian noise. Therefore  $p(\mathbf{Y}|\mathbf{C}, \mathbf{V}) \sim \mathcal{N}(\mathbf{V}\mathbf{C}, \mathbb{1})$ , calculate the conditional basis vectors as  $\mathbf{V}\mathbf{C} = \arg \min_{\mathbf{V}} p(\mathbf{Y}|\mathbf{C})$  to obtain  $p(\mathbf{Y}|\mathbf{C}) \sim \mathcal{N}(\mathbf{V}\mathbf{C}, \mathbb{1})$ . From which  $\ln(p(\mathbf{Y}|\mathbf{C})) = \|\mathbf{Y} - \mathbf{V}\bar{\mathbf{C}}\|_F^2$ .

For the total variation prior used, it is implied that  $\ln(p(\mathbf{C})) \propto \|\mathbf{D}_w \bar{\mathbf{C}}\|_{2,p}^p \implies p(\mathbf{C}) \propto \exp \|\mathbf{D}_w \bar{\mathbf{C}}\|_{2,p}^p$ , in the case where  $p = 2$  this is exactly a Gaussian prior.

## REFERENCES

- Aigrain, S., Hodgkin, S. T., Irwin, M. J., Lewis, J. R., & Roberts, S. J. 2015, MNRAS, 447, 2880, doi: [10.1093/mnras/stu2638](https://doi.org/10.1093/mnras/stu2638)
- Aigrain, S., Parviainen, H., & Pope, B. J. S. 2016, Monthly Notices of the Royal Astronomical Society, 459, 2408, doi: [10.1093/mnras/stw706](https://doi.org/10.1093/mnras/stw706)
- Aigrain, S., Parviainen, H., Roberts, S., Reece, S., & Evans, T. 2017, Monthly Notices of the Royal Astronomical Society, 471, 759, doi: [10.1093/mnras/stx1422](https://doi.org/10.1093/mnras/stx1422)
- Auvergne, M., Bodin, P., Boissard, L., et al. 2009, Astronomy & Astrophysics, 506, 411
- Auvergne, M., Bodin, P., Boissard, L., et al. 2009, A&A, 506, 411, doi: [10.1051/0004-6361/200810860](https://doi.org/10.1051/0004-6361/200810860)
- Batalha, N. M. 2014, Proceedings of the National Academy of Sciences, 111, 12647, doi: [10.1073/pnas.1304196111](https://doi.org/10.1073/pnas.1304196111)
- Borucki, W. J., Koch, D., Basri, G., et al. 2010, Science, 327, 977, doi: [10.1126/science.1185402](https://doi.org/10.1126/science.1185402)
- Bryson, S., Coughlin, J., Batalha, N. M., et al. 2020, The Astronomical Journal, 159, 279
- Bryson, S. T., Tenenbaum, P., Jenkins, J. M., et al. 2010, The Astrophysical Journal, 713, L97, doi: [10.1088/2041-8205/713/2/197](https://doi.org/10.1088/2041-8205/713/2/197)
- Caldwell, D. A., Kolodziejczak, J. J., Cleve, J. E. V., et al. 2010, The Astrophysical Journal, 713, L92, doi: [10.1088/2041-8205/713/2/192](https://doi.org/10.1088/2041-8205/713/2/192)
- Candès, E. J., Li, X., Ma, Y., & Wright, J. 2011, J. ACM, 58, doi: [10.1145/1970392.1970395](https://doi.org/10.1145/1970392.1970395)
- Christiansen, J. L., Jenkins, J. M., Caldwell, D. A., et al. 2012, Publications of the Astronomical Society of the Pacific, 124, 1279, <http://www.jstor.org/stable/10.1086/668847>
- Conroy, C., Strader, J., van Dokkum, P., et al. 2018, The Astrophysical Journal, 864, 111
- Crossfield, I. J. M., Petigura, E., Schlieder, J. E., et al. 2015, The Astrophysical Journal, 804, 10, doi: [10.1088/0004-637x/804/1/10](https://doi.org/10.1088/0004-637x/804/1/10)
- Deeg, H. J., & Alonso, R. 2018, Transit Photometry as an Exoplanet Discovery Method, ed. H. J. Deeg & J. A. Belmonte (Cham: Springer International Publishing), 1–25, doi: [10.1007/978-3-319-30648-3\\_117-1](https://doi.org/10.1007/978-3-319-30648-3_117-1)

- Deming, D., Knutson, H., Kammer, J., et al. 2015, *ApJ*, 805, 132, doi: [10.1088/0004-637X/805/2/132](https://doi.org/10.1088/0004-637X/805/2/132)
- Demory, B.-O., Ehrenreich, D., Queloz, D., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 450, 2043, doi: [10.1093/mnras/stv673](https://doi.org/10.1093/mnras/stv673)
- Eckart, C., & Young, G. 1936, *Psychometrika*, 1, 211
- Fausnaugh, M. 2018, TESS Data Release Notes: Sector 1, DR1, Tech. rep. [https://tasoc.dk/docs/release\\_notes/tess\\_sector\\_01\\_drn01\\_v02.pdf](https://tasoc.dk/docs/release_notes/tess_sector_01_drn01_v02.pdf)
- Foreman-Mackey, D., Montet, B. T., Hogg, D. W., et al. 2015, *The Astrophysical Journal*, 806, 215
- Gilliland, R. L., Chaplin, W. J., Dunham, E. W., et al. 2011, *The Astrophysical Journal Supplement Series*, 197, 6, doi: [10.1088/0067-0049/197/1/6](https://doi.org/10.1088/0067-0049/197/1/6)
- Goldstein, T., & Osher, S. 2009, *SIAM J. Imaging Sciences*, 2, 323, doi: [10.1137/080725891](https://doi.org/10.1137/080725891)
- Golub, G., & Pereyra, V. 2003, *Inverse Problems*, 19, R1, doi: [10.1088/0266-5611/19/2/201](https://doi.org/10.1088/0266-5611/19/2/201)
- Golub, G., & Van Loan, C. 2013, *Matrix Computations*, Johns Hopkins Studies in the Mathematical Sciences (Johns Hopkins University Press). <https://books.google.com/books?id=5U-l8U3P-VUC>
- Golub, G. H., & Pereyra, V. 1973, *SIAM Journal on numerical analysis*, 10, 413
- Grinstead, C. M., & Snell, J. L. 2012, *Introduction to probability* (American Mathematical Soc.)
- Harville, D. 2008, *Matrix Algebra From a Statistician's Perspective* (Springer New York). <https://books.google.com/books?id=kZGBQijgGV8C>
- Hastie, T., Tibshirani, R., & Friedman, J. 2009, *Unsupervised Learning* (New York, NY: Springer New York), 485–585, doi: [10.1007/978-0-387-84858-7\\_14](https://doi.org/10.1007/978-0-387-84858-7_14)
- Hattori, S., Foreman-Mackey, D., Hogg, D. W., et al. 2022, *The Astronomical Journal*, 163, 284, doi: [10.3847/1538-3881/ac625a](https://doi.org/10.3847/1538-3881/ac625a)
- Hedges, C., Luger, R., Dotson, J., Foreman-Mackey, D., & Barentsen, G. 2021
- Huang, X., Soares-Furtado, M., Penev, K., et al. 2015, in *AAS/Division for Extreme Solar Systems Abstracts*, Vol. 47, AAS/Division for Extreme Solar Systems Abstracts, 114.06
- Jenkins, J. M. 2017, *Kepler Data Processing Handbook: Overview of the Science Operations Center*, Kepler Science Document KSCI-19081-002
- Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., et al. 2010, *The Astrophysical Journal Letters*, 713, L87
- Karl, W. C. 2005, in *Handbook of Image and Video Processing* (Elsevier), 183–V
- Kinemuchi, K., Barclay, T., Fanelli, M., et al. 2012, *PASP*, 124, 963, doi: [10.1086/667603](https://doi.org/10.1086/667603)
- Koch, D. G., Borucki, W. J., Basri, G., et al. 2010, *ApJL*, 713, L79, doi: [10.1088/2041-8205/713/2/L79](https://doi.org/10.1088/2041-8205/713/2/L79)
- Kovács, G., Bakos, G., & Noyes, R. W. 2005, *Monthly Notices of the Royal Astronomical Society*, 356, 557, doi: [10.1111/j.1365-2966.2004.08479.x](https://doi.org/10.1111/j.1365-2966.2004.08479.x)
- Luger, R., Agol, E., Kruse, E., et al. 2016, *The Astronomical Journal*, 152, 100, doi: [10.3847/0004-6256/152/4/100](https://doi.org/10.3847/0004-6256/152/4/100)
- Luger, R., Kruse, E., Foreman-Mackey, D., Agol, E., & Saunders, N. 2018, *AJ*, 156, 99, doi: [10.3847/1538-3881/aad230](https://doi.org/10.3847/1538-3881/aad230)
- Lund, M. N., Handberg, R., Davies, G. R., Chaplin, W. J., & Jones, C. D. 2015, *The Astrophysical Journal*, 806, 30, doi: [10.1088/0004-637x/806/1/30](https://doi.org/10.1088/0004-637x/806/1/30)
- Lund, M. N., Handberg, R., Buzasi, D. L., et al. 2021, *The Astrophysical Journal Supplement Series*, 257, 53, doi: [10.3847/1538-4365/ac214a](https://doi.org/10.3847/1538-4365/ac214a)
- Mazeh, T., Guterman, P., Aigrain, S., et al. 2009, *A&A*, 506, 431, doi: [10.1051/0004-6361/200912037](https://doi.org/10.1051/0004-6361/200912037)
- McLean, I. 2008, *Electronic Imaging in Astronomy: Detectors and Instrumentation*, Springer Praxis Books (Springer Berlin Heidelberg). <https://books.google.com/books?id=LXH2UavpcakC>
- McQuillan, A., Aigrain, S., & Roberts, S. 2012, *A&A*, 539, A137, doi: [10.1051/0004-6361/201016148](https://doi.org/10.1051/0004-6361/201016148)
- Moreno, J., Buttry, R., O'Brien, J., et al. 2021, *AJ*, 162, 232, doi: [10.3847/1538-3881/ac205c](https://doi.org/10.3847/1538-3881/ac205c)
- Murphy, K. 2020, *Machine Learning, second edition: A Probabilistic Perspective*, Adaptive Computation and Machine Learning series (MIT Press). <https://books.google.com/books?id=Qs.tDwAAQBAJ>
- Ofir, A., Alonso, R., Bonomo, A. S., et al. 2010, *Monthly Notices of the Royal Astronomical Society: Letters*, 404, L99, doi: [10.1111/j.1745-3933.2010.00843.x](https://doi.org/10.1111/j.1745-3933.2010.00843.x)
- Parikh, N., & Boyd, S. 2013, *Proximal Algorithms*, Foundations and Trends in Optimization (Now Publishers). <https://books.google.com/books?id=DS04ngEACAAJ>
- Petigura, E. A., & Marcy, G. W. 2012, *Publications of the Astronomical Society of the Pacific*, 124, 1073, doi: [10.1086/668291](https://doi.org/10.1086/668291)
- Pont, F., Zucker, S., & Queloz, D. 2006, *MNRAS*, 373, 231, doi: [10.1111/j.1365-2966.2006.11012.x](https://doi.org/10.1111/j.1365-2966.2006.11012.x)
- Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, *Journal of Astronomical Telescopes, Instruments, and Systems*, 1, 014003
- Roberts, S., McQuillan, A., Reece, S., & Aigrain, S. 2013, *Monthly Notices of the Royal Astronomical Society*, 435, 3639, doi: [10.1093/mnras/stt1555](https://doi.org/10.1093/mnras/stt1555)

- Rudin, L. I., Osher, S., & Fatemi, E. 1992, Nonlinear total variation based noise removal algorithms
- Shearer, P., & Gilbert, A. C. 2013, *Inverse Problems*, 29, 045003, doi: [10.1088/0266-5611/29/4/045003](https://doi.org/10.1088/0266-5611/29/4/045003)
- Smith, J. C., Morris, R. L., Jenkins, J. M., et al. 2016, *Publications of the Astronomical Society of the Pacific*, 128, 124501, doi: [10.1088/1538-3873/128/970/124501](https://doi.org/10.1088/1538-3873/128/970/124501)
- Smith, J. C., Stumpe, M. C., Van Cleve, J. E., et al. 2012, *Publications of the Astronomical Society of the Pacific*, 124, 1000
- Smith, K. L., Mushotzky, R. F., Boyd, P. T., et al. 2018, *The Astrophysical Journal*, 857, 141, doi: [10.3847/1538-4357/aab88d](https://doi.org/10.3847/1538-4357/aab88d)
- Stefansson, G., Mahadevan, S., Hebb, L., et al. 2017, *The Astrophysical Journal*, 848, 9, doi: [10.3847/1538-4357/aa88aa](https://doi.org/10.3847/1538-4357/aa88aa)
- Stumpe, M., Smith, J., Catanzarite, J., et al. 2013, *Publications of the Astronomical Society of the Pacific*, 126, doi: [10.1086/674989](https://doi.org/10.1086/674989)
- Stumpe, M. C., Smith, J. C., Van Cleve, J. E., et al. 2012, *Publications of the Astronomical Society of the Pacific*, 124, 985
- Taaki, J. S., Kamalabadi, F., & Kembell, A. J. 2020, *The Astronomical Journal*, 159, 283, doi: [10.3847/1538-3881/ab8e38](https://doi.org/10.3847/1538-3881/ab8e38)
- Tamuz, O., Mazeh, T., & Zucker, S. 2005, *MNRAS*, 356, 1466, doi: [10.1111/j.1365-2966.2004.08585.x](https://doi.org/10.1111/j.1365-2966.2004.08585.x)
- Tenenbaum, P., & Jenkins, J. 2018, *TESS Science Data Products Description Document*, Tech. rep.
- Thatte, A., Deroo, P., & Swain, M. R. 2010, *A&A*, 523, A35, doi: [10.1051/0004-6361/201015148](https://doi.org/10.1051/0004-6361/201015148)
- Toyozumi, H., & Ashley, M. C. 2005, *Publications of the Astronomical Society of Australia*, 22, 257
- Tregloan-Reed, J., & Southworth, J. 2013, *Monthly Notices of the Royal Astronomical Society*, 431, 966, doi: [10.1093/mnras/stt227](https://doi.org/10.1093/mnras/stt227)
- Van Cleve, J. E., & Caldwell, D. A. 2016, *Kepler Instrument Handbook*, Kepler Science Document KSCI-19033-002
- Vanderburg, A., & Johnson, J. 2014, *Publications of the Astronomical Society of the Pacific*, 126, doi: [10.1086/678764](https://doi.org/10.1086/678764)
- Vanderspek, R. 2018, *Tess Instrument Handbook*
- Vogel, C. R. 2002, *Computational Methods for Inverse Problems* (Society for Industrial and Applied Mathematics), doi: [10.1137/1.9780898717570](https://doi.org/10.1137/1.9780898717570)
- Wang, D., Hogg, D. W., Foreman-Mackey, D., & Schölkopf, B. 2016, *PASP*, 128, 094503, doi: [10.1088/1538-3873/128/967/094503](https://doi.org/10.1088/1538-3873/128/967/094503)