

https://dspace.org/

Robust Detrending of Spatially Correlated Systematics in Kepler Light Curves Using Lowrank Methods

Jamila S. Taaki; Athol Kemball; Farzad Kamalabadi

IOP PUBLISHING LTD
https://hdl.handle.net/10539/40677
WOS

Downloaded from a DSpace repository. https://dspace.org/

The Astronomical Journal, 167:60 (29pp), 2024 February https://doi.org/10.3847/1538-3881/ad1110
© 2024. The Author(s). Published by the American Astronomical Society.
OPEN ACCESS

Robust Detrending of Spatially Correlated Systematics in Kepler Light Curves Using Low-rank Methods

Jamila S. Taaki[1], Athol J. Kemball[2,3], and Farzad Kamalabadi[1]
[1] Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign 306 N. Wright Street MC 702, Urbana, IL 61801-2918, USA; xiaziyna@gmail.com
[2] Department of Astronomy, University of Illinois at Urbana-Champaign 1002 W. Green Street, Urbana, IL 61801-3074, USA
[3] School of Physics, University of the Witwatersrand, P.O. Box Wits, Johannesburg, South Africa
Received 2023 July 22; revised 2023 November 13; accepted 2023 November 28; published 2024 January 15

Abstract

Light curves produced by wide-field exoplanet transit surveys such as CoRoT, Kepler, and the Transiting Exoplanet Survey Satellite are affected by sensor-wide systematic noise, which is correlated both spatiotemporally and with other instrumental parameters such as the photometric magnitude. Robust and effective systematics mitigation is necessary to achieve the level of photometric accuracy required to detect exoplanet transits and to faithfully recover other forms of intrinsic astrophysical variability. We demonstrate the feasibility of a new exploratory algorithm to remove spatially correlated systematic noise and detrend light curves obtained from wide-field transit surveys. This spatial systematics algorithm is data-driven and fits a low-rank linear model for the systematics conditioned on a total-variation spatial constraint. The total-variation constraint models spatial systematic structure across the sensor on a foundational level. The fit is performed using gradient descent applied to, a variable reduced least-squares penalty and a modified form of total-variation prior; both the systematics basis vectors and their weighting coefficients are iteratively varied. The algorithm was numerically evaluated against a reference principal component analysis, using both signal injection on a selected Kepler dataset, as well as full simulations within the same Kepler coordinate framework. We develop our algorithm to reduce the overfitting of astrophysical variability over longer signal timescales (days) while performing comparably relative to the reference method for exoplanet transit timescales. The algorithm performance and application are assessed, and future development is outlined.

Unified Astronomy Thesaurus concepts: Exoplanet detection methods (489); Astronomy data analysis (1858); Wide-field telescopes (1800); Surveys (1671); Transit photometry (1709)

1. Introduction

Transit surveys have contributed significantly to exoplanet science over the past two decades, uncovering exoplanet population statistics at current survey completeness limits (Batalha 2014; Bryson et al. 2020). Space-based surveys, including CoRoT (Auvergne et al. 2009), Kepler (Borucki et al. 2010), and the Transiting Exoplanet Survey Satellite (TESS; Ricker et al. 2014) have led to the discovery of over 5400[4] exoplanets to date. These wide-field surveys include a large number of candidate systems and require a photometric precision after standard calibration and processing (residual photometric precision) adequate to detect exoplanet transits (Deeg & Alonso 2018). Achieving low residual photometric noise requires the optimization of both the instrument design and the associated algorithms for instrumental noise removal. In this paper, we describe a method 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 $\triangle F$ in observations, with a differential photometric precision σ_p (after processing) then the signal-to-noise ratio is $S/N \sim \frac{\triangle F}{\sigma_p}$ (Deeg & Alonso 2018). For example, an Earth–Sun transit requires $\sigma_p \sim 80$ parts-per-million (ppm) over the transit duration $T \sim$ hours (Caldwell et al. 2010). The Kepler telescope achieved $\sigma_p \sim 30$ ppm over $T = 6.5$hr 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 a star with TESS magnitude $Tp \sim 10$, which is sufficient to detect super-Earths around bright stars (Fausnaugh 2018). For reference, the Hubble Space Telescopes Space Telescope Imaging Spectrograph (HST/STIS) can attain $\sigma_p \sim 120$ ppm over $T = 45$ minutes (Demory et al. 2015), and non-wide-field ground-based telescopes can reach a precision sufficient to detect large Jovian exoplanets (Tregloan-Reed & Southworth 2013; Stefansson et al. 2017).

The instrumental response will vary over a range of timescales and across various spatial scales producing systematic noise effects, as expected from general principles (McLean 2008). The primary data products for the Kepler and TESS telescopes include light curves derived using optimized simple aperture photometry (SAP) or image data providing pixel-level light curves (Jenkins 2017; Tenenbaum & Jenkins 2018). Algorithmic innovations to suppress the residual differential photometric precision σ_p, arising from unmodeled systematic noise effects, are therefore critical to detect weak exoplanet signals. These systematic effects have diverse physical origins (Jenkins 2017; Tenenbaum & Jenkins 2018). A single pixel exhibits a nonuniform sensitivity across its surface (Toyozumi & Ashley 2005; Hedges et al. 2021). In addition, there are variations in pixel-to-pixel sensitivity and instrumental pixel response (Van Cleve & Caldwell 2016) including CCD pattern, read, photon, and

[4] https://exoplanetarchive.ipac.caltech.edu/

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

quantization noise (McLean 2008; Caldwell et al. 2010; Gilliland et al. 2011); instrumental CCD terms are expected to vary by CCD output channel or module. A net pixel response function (PRF; Bryson et al. 2010) captures the spatially variant response across the detector including the spatial variation of the optical PSF and instrumental detector sensitivity, among related factors (Jenkins 2017). The spatial response of the detector may, however, include unmodeled error due to pointing jitter, focus changes, uncorrected differential velocity aberration terms, and their interaction (Van Cleve & Caldwell 2016). Thermal effects as well as discrete spacecraft guidance and downlink control events, including momentum dumps, may produce temporal and spatial variation in the systematics, some abrupt (Van Cleve & Caldwell 2016; Vanderspek 2018). Further, bright astrophysical sources may produce pixel saturation or broadening of the PRF (Van Cleve & Caldwell 2016; Vanderspek 2018).

The first correction for noise effects is typically applied in early robust calibration pipelines (Jenkins 2017; Tenenbaum & Jenkins 2018). However residual spatially variant and time-variant systematics unavoidably remain in the calibrated data products given the complex instrumental response at the current level of photometric accuracy. We denote these residual systematics contributions to each light curve i at discrete time sample n as $\boldsymbol{l}_i[n]$ (Section 2.2) where the data may be either pixel or SAP light curves. The systematics $\boldsymbol{l}_i[n]$ has the same discrete time sampling as the measured light curve. The residual systematics vary spatially over the sensor, reflected in equivalent notation $\boldsymbol{l}_{x,y}[n]$, where light curve i maps to sensor position (x, y). Petigura & Marcy (2012) show an example of spatially varying residual systematic noise in Kepler data by characterizing correlations across the sensor. Moreno et al. (2021) similarly characterize spatial systematics correlation in Kepler/K2 data. An a priori analytic instrumental model for $\boldsymbol{l}_{x,y}[n]$ is intractable in practice, and data-driven approaches, informed by physically realistic instrumental assumptions, provide the most effective approach to mitigate the residual systematic noise in the target light curves. There is a rich history in the literature on this subject, which we review briefly only to place our method in context.

A class of methods (hereinafter *external parameter decorrelation* methods; Bakos et al. 2007) model the functional form of the residual systematics $\boldsymbol{l}_{x,y}[n]$ as correlates of other explanatory variables such as pointing error estimates or ancillary engineering data (PDC-LS, Twicken et al. 2010). Charbonneau et al. (2005) demonstrated Spitzer[5] aperture photometry correction using decorrelation with pointing errors estimated using target centroiding. There has been substantial development in this latter area, especially inspired by the pointing challenges of the Kepler K2 mission and pixel-level detrending, including the work described by Vanderburg & Johnson (2014), Huang et al. (2015), Aigrain et al. (2015), Lund et al. (2015) and Crossfield et al. (2015). External parameter decorrelation methods implicitly include spatiotemporal variability in the residual systematics via the encoding of their functional form in the proxy variables.

A second class of methods (hereinafter *cotrending* methods) makes the foundational assumption that a set of measured light curves, or their computed vector basis, comprise an efficient basis of regressors over which to expand the unknown functional form of $\boldsymbol{l}_{x,y}[n]$ as a low-rank linear model. This implicitly assumes that the residual systematics are temporally and spatially correlated across the sensor; this is similarly a physically reasonable assumption. The trend-filtering algorithm (TFA; Kovács et al. 2005) directly uses a uniform selection of external light curves as regressors in a least-squares fit to detrend an individual target light curve. The cotrending approach is analogous to the joint coupled least-squares solution for the product of stellar extinction coefficients and airmasses as posed in a global solution for wide-field multiobject photometry (Sysrem; Tamuz et al. 2005); this solution is equivalent to a principal component analysis (PCA) of target observations. A refinement including additive external parameters is provided by SARS (Ofir et al. 2010).

Cotrending methods have seen significant algorithmic refinement. The PDC-MAP algorithm (Smith et al. 2012; Stumpe et al. 2012) constructs a set of cotrending basis vectors (CBV) using singular value decomposition (SVD) on a template set of light curves selected for their high degree of correlation (therefore capturing foundational instrumental trends) and quiescence. Each CBV is fit against the target light curve using Bayesian maximum a posteriori (MAP) methods to obtain relative coefficient weightings for each basis term, the sum of which is then subtracted from the target time series to remove systematic effects. An empirical Bayesian prior is used to constrain overfitting, constructed on the variation of coefficient value over stellar magnitude and spatial position on the sensor. The CBV are orthogonal by mathematical construction and, therefore, do not map directly to constituent instrumental effects; this dilutes the variable dependence in the empirical coefficient prior. Instrumental effects are often separated by characteristic timescale and this mapping can be improved (MS-MAP; Stumpe et al. 2014) by deriving separate CBV for different timescale wavelet subbands; this provides better temporal separation in the spatiotemporal dependence of $\boldsymbol{l}_{x,y}[n]$. A net composite systematic correction is applied as the multiscale combination across each subband. The Astrophysically Robust Correction algorithm (ARC; Roberts et al. 2013; Aigrain et al. 2017) is related conceptually to PDC-MAP but uses automatic relevance coefficient priors to maximize model evidence for instrumental trends.

Additional cotrending methods include the Causal Pixel Model (CPM; Wang et al. 2016), which uses a template (or training) set of regressor pixel-data light curves that are separated in time from the target light-curve sample being corrected for systematics. The training set is spatially separated from the target but the target light curve itself is included to include autoregressive information. Both measures suppress residual variability except on short transit timescales. A successor method is described by Hattori et al. (2022). The method described by Foreman-mackey et al. (2015) uses a large training set to derive an expanded CBV but uses a joint estimation of systematics and the transit signal in order to better constrain the problem. A related Bayesian formulation is described by Taaki et al. (2020). The pixel-level decorrelation algorithm (PLD; Deming et al. 2015) was developed to implicitly correct photometric decorrelation with pointing errors in Spitzer data. PLD uses pixel light-curve regressors within a summed aperture to model SAP light curves. The regressors encode pointing errors without the need for explicit centroiding as in external parameter decorrelation approaches. The PLD method was extended by Luger et al. (2016) for application to Kepler K2 data by including higher-order terms in the SAP flux dependence on pointing error, by modeling astrophysical variability using a

[5] https://www.spitzer.caltech.edu/

Gaussian process, and by constructing a PCA basis from the regressors to better constrain the model. The method was further improved (Luger et al. 2018) by including L2 coefficient regularization, which was also used in CPM, and by incorporating PLD vectors from nearby stars (nPLD) to improve sensitivity to pointing error for faint stars.

The cotrending methods described above address challenges in estimating intrinsic astrophysical variability. Foundationally it is difficult to perfectly separate systematics and astrophysical variability a priori in a light-curve regressor or derived basis set. Therefore the linear systematics model may include and absorb true astrophysical variability (overfitting) or inject spurious variability into detrended light curves. There may also be an incidental correlation between a relatively clean systematics basis and astrophysical variability (Smith et al. 2018). Overfitting can be reduced by judicious use of prior constraints that represent a physically realistic instrumental response; these constraints may also significantly improve the condition of the problem.

In the cotrending methods described above, spatial structure is incorporated into the systematic noise model using several different approaches. In the simplest form, the spatial variation constraint is imposed implicitly by cotrending targets across a discrete sensor region as a whole (e.g., a CCD output channel; Smith et al. 2012; Roberts et al. 2013; Stumpe et al. 2014; Luger et al. 2018; Lund et al. 2021). Spatial variation has also been incorporated by restricting regressors by proximity to the target light curve (Luger et al. 2018; Lund et al. 2021), or by directly introducing a parametric dependence of fitted systematics on sensor position in an empirical prior (Smith et al. 2012; Stumpe et al. 2014).

In this paper, we present an exploratory algorithm to solve for a cotrending low-rank linear systematics model while incorporating a spatial constraint across the sensor at a foundational level. We refer to this as the *spatial systematics method* in what follows. The spatial constraint is of generalized total variation (TV) from Rudin et al. (1992). This TV constraint on basis vector coefficients promotes the correlation of adjacent neighbors on the sensor and is weighted spatially across the sensor while permitting discontinuities as found in wide-field imaging sensors at module or channel edges. As such, the spatial constraint has a realistic physical motivation and therefore a strong potential to reduce overfitting. The fit the overall SAP light curves is performed by minimizing the sum of the least-squares residual between the light curves and the linear systematics model, and, the total-variation spatial constraint. We use variable elimination (Golub & Pereyra 1973) to reformulate the optimization problem as a function of the weighted coefficients only, introducing stability to the minimization (Golub & Pereyra 2003; Shearer & Gilbert 2013). An approximate closed-form gradient of the objective function is derived and minimized via gradient descent. In this iterative solution both the weighted coefficients and the basis vectors, which functionally are defined by the coefficients, are varied. The algorithm as implemented is available as a public Python package on the GitHub repository[6] or via PyPI under the package name *spatial-detrend*.[7]

We numerically evaluate the performance of our method in the context of the Kepler sensor and against PCA as a reference method. We use both injected signals in real long-cadence (LC) Kepler data and fully simulated data in this evaluation. As noted above (Smith et al. 2012, PDC-MAP) Kepler systematics depend on stellar magnitude. This is also found for CoRoT light curves (Mazeh et al. 2009; Ofir et al. 2010). This effect likely arises due to pixel saturation effects. Kepler operated in the magnitude band Kp $\in$ 9 to 15 (Koch et al. 2010). In this work for simplicity we do not include a prior on stellar magnitude dependence but instead select light curves from a narrow fixed magnitude band Kp $\in$ 12 to 13 within which magnitude-dependent systematics such as saturation are not expected to substantially vary (Smith et al. 2012); this therefore allows for the spatial variability systematics to be isolated. Our numerical evaluation demonstrates that the spatial systematics algorithm has reduced overfitting of astrophysical variability, particularly on day-long timescales, compared to the reference PCA method.

The paper is organized as follows. Section 2 introduces the light-curve signal model, our method for spatial systematics inference, and an analysis of spatial correlation across the Kepler sensor, and describes our experimental tests to numerically evaluate the algorithm. In Section 3 we report the results of our numerical evaluation. These results are discussed in Section 4 and conclusions are presented in Section 5. A table of commonly used symbols is provided in Appendix A for reference.

2. Methods

As described above, the data products for wide-field exoplanet transit surveys may comprise: (i) pixel-level image data either within target apertures or for full frames; and, (ii) SAP target light curves (Jenkins 2017; Tenenbaum & Jenkins 2018). Here we use Kepler SAP target light curves with a sampling cadence of 30 minutes produced by the Kepler science data processing pipeline but not including post-processing by the Kepler Pre-Search Data Conditioning (PDC) module (Twicken et al. 2010; Smith et al. 2012; Stumpe et al. 2012); the latter light curves with full Kepler post-processing are distinguished as PDC-SAP light curves. We present the details of the spatial systematics method that is the subject of this paper.

2.1. Matrix Notation

We define common notation here that is used throughout the paper. The ith row of matrix $\boldsymbol{M}$ is denoted as $[\boldsymbol{M}]_i$ and the jth column as $[\boldsymbol{M}]_{\cdot j}$. The matrix element corresponding to the ith row and the jth column is denoted as $[\boldsymbol{M}]_{i,j}$ or $M_{i,j}$. The transpose of a matrix $\boldsymbol{M}$ is denoted as $\boldsymbol{M}^T$. An identity matrix of size $N \times N$ is denoted as 1_N. The L_p norm of a vector $\boldsymbol{x}$ of length I is defined as: $\|\boldsymbol{x}\|_p = (\sum_{i=1}^{I} |x_i|^p)^{1/p}$. If $\boldsymbol{M} \in \mathbb{R}^{I \times J}$, we use $\|\boldsymbol{M}\|_{p,q}$ to denote the L_p norm applied to each column $[\boldsymbol{M}]_{\cdot j}$, followed by the L_q norm applied to this vector, such that $\|\boldsymbol{M}\|_{p,q} = (\sum_{j=1}^{J} (\sum_{i=1}^{I} |M_{i,j}|^p)^{q/p})^{1/q}$. The Frobenius norm is defined as $\|\boldsymbol{M}\|_F = \|\boldsymbol{M}\|_{2,2}^1$, such that $\|\boldsymbol{M}\|_F^2 = \sum_{j=1}^{J} \sum_{i=1}^{I} |M_{i,j}|^2$ (Golub & Van Loan 2013).

2.2. Light-curve Decomposition

A collection of target light curves $\{\boldsymbol{y}_i : i \in I\}$ are obtained on a sensor, each light curve $\boldsymbol{y}_i$ (over target index set I) is a length-N time series. Each light curve is represented as the sum of a

[6] https://github.com/xiaziyna/spatial-detrend
[7] https://pypi.org/project/spatial-detrend

systematics term $\boldsymbol{l}_i$ and a statistical noise term $\boldsymbol{n}_i$:

$$\boldsymbol{y}_i = \boldsymbol{l}_i + \boldsymbol{n}_i. \tag{1}$$

In matrix form the data model takes the form $\boldsymbol{Y} = \boldsymbol{L} + \boldsymbol{N}$, where $\boldsymbol{y}_i$, $\boldsymbol{l}_i$, and $\boldsymbol{n}_i$ form the columns of matrices $\boldsymbol{Y} \in \mathbb{R}^{N\times I}$, $\boldsymbol{L} \in \mathbb{R}^{N\times I}$, and $\boldsymbol{N} \in \mathbb{R}^{N\times I}$ representing the light curves, systematic noise, and statistical noise, respectively. The goal of this work is to form an accurate estimate of the systematic noise $\boldsymbol{l}_i$.

Sources of systematic errors are described in Section 1; see also Auvergne et al. (2009), Van Cleve & Caldwell (2016), and Vanderspek (2018). Statistical noise originates due to a mixture of instrumental error and astrophysical variability. We approximate statistical noise $\boldsymbol{n}_i$ as white Gaussian noise $\boldsymbol{n}_i \sim \mathcal{N}(0, \sigma_i^2)$ as additive statistical noise sources may reasonably approach $\mathcal{N}(0, \sigma_i^2)$ under the central limit theorem (Grinstead & Snell 2012). This is a common assumption in this domain and implicit in least-squares minimization approaches (Smith et al. 2012; Stumpe et al. 2012, 2014; Aigrain et al. 2017). The noise level σ_i is unknown a priori, but may reasonably be estimated from filtered and coarsely detrended light curves. In what follows, light curves are normalized by an estimate of σ_i so that $\boldsymbol{n}_i \sim \mathcal{N}(0, 1)$ for mathematical convenience.

Light curves may include intermittent sources of systematic noise that produce outliers and are difficult to model. Their influence may be minimized by prior constraints (a further motivation for our approach here), but data-driven outlier filtering is generally necessary in this domain (Section 2.5).

2.3. Systematics Model

Low-rank model: We adopt a cotrending basis model:

$$\boldsymbol{l}_i[n] = \sum_{k\,=\,1}^{K} c_i^k \boldsymbol{v}_k[n], \tag{2}$$

where $(K \ll N)$ describes the rank of the systematic noise model, $\{\boldsymbol{v}_k : k \in K\}$ are a set of basis vectors shared by light curves, and each c_i^k is a coefficient weighting of $\boldsymbol{v}_k$ for light curve i. The coefficient vector for light curve i is $\boldsymbol{c}_i = [c_i^1, c_i^2, \ldots c_i^k]^T$. In matrix form:

$$\boldsymbol{L} = \boldsymbol{V}\boldsymbol{C}, \tag{3}$$

where $\boldsymbol{L} \in \mathbb{R}^{N\times I}$ is defined above, the columns of $\boldsymbol{V} \in \mathbb{R}^{N\times K}$ are the basis vectors $\{\boldsymbol{v}_k : k \in K\}$, and the columns of $\boldsymbol{C} \in \mathbb{R}^{K\,\times\,I}$ are the coefficients $\{\boldsymbol{c}_i : i \in I\}$. Graphically,

$$\begin{bmatrix} | & | & & | \\ \boldsymbol{l}_1 & \boldsymbol{l}_2 & \ldots & \boldsymbol{l}_I \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_K \\ | & | & & | \end{bmatrix} \begin{bmatrix} | & | & & | \\ \boldsymbol{c}_1 & \boldsymbol{c}_2 & \ldots & \boldsymbol{c}_I \\ | & | & & | \end{bmatrix}. \tag{4}$$

We define a column-normalized coefficient matrix $\bar{\boldsymbol{C}}$ of the form $\bar{\boldsymbol{c}}_i = \frac{\boldsymbol{c}_i}{\|\,\boldsymbol{c}_i\,\|_2}$:

$$\bar{\boldsymbol{C}} = \begin{bmatrix} | & | & & | \\ \frac{\boldsymbol{c}_1}{\|\,\boldsymbol{c}_1\,\|_2} & \frac{\boldsymbol{c}_2}{\|\,\boldsymbol{c}_2\,\|_2} & \ldots & \frac{\boldsymbol{c}_I}{\|\,\boldsymbol{c}_I\,\|_2} \\ | & | & & | \end{bmatrix}. \tag{5}$$

The rank of $\boldsymbol{L}$ may be less than the number of independent systematic noise sources. As $\mathrm{rank}(\boldsymbol{V}\boldsymbol{C}) \leqslant \min\{\mathrm{rank}(\boldsymbol{V}), \mathrm{rank}(\boldsymbol{C})\}$, it is possible to have a more expansive basis set $\boldsymbol{V}$ representing many individual noise effects, but for the relative weightings $\boldsymbol{C}$ between light curves to have a degenerate and therefore low-rank structure and consequently for $\boldsymbol{L}$ to be low rank.

SVD/PCA systematics estimation: SVD and PCA are integral to many cotrending approaches (Tamuz et al. 2005; Thatte et al. 2010; Petigura & Marcy 2012; Smith et al. 2012; Stumpe et al. 2012) and are summarized in Appendix B. Under a white noise model for $\boldsymbol{N}$ the maximum likelihood (ML) estimate of a matrix $\boldsymbol{L}$ (not necessarily low-rank) given light curves $\boldsymbol{Y} = \boldsymbol{N} + \boldsymbol{L}$ is equivalent to minimizing the least-squares residual $f(\boldsymbol{V}, \boldsymbol{C}) = ||\boldsymbol{Y} - \boldsymbol{L}||_F^2$:

$$\arg\max_{\boldsymbol{L}} p(\boldsymbol{Y}|\boldsymbol{L}) \equiv \arg\min_{\boldsymbol{L}} ||\boldsymbol{Y} - \boldsymbol{L}||_F^2, \tag{6}$$

where $p(\boldsymbol{Y}|\boldsymbol{L})$ is the likelihood of $\boldsymbol{Y}$ given $\boldsymbol{L}$ (Srebro 2004).

For the data model in Equation (1), without any further constraints, a rank-K optimal least-squares solution of $\boldsymbol{L}$ can be found by rank-thresholding the SVD or PCA decomposition of $\boldsymbol{Y}$ per the Eckart–Young–Mirsky theorem (Appendix B) and in this case takes the form: $\boldsymbol{L} = \boldsymbol{V}_K\boldsymbol{C}_K$. In what follows we use the SVD/PCA method primarily as a comparative method (Section 2.5).

Spatial systematics model: As described in Section 1, the residual systematics are spatially correlated across the sensor. In addition, the white noise model for $\boldsymbol{N}$ is an idealization and the systematics cannot then be perfectly separated by rank thresholding alone. In our spatial systematics algorithm, we retain a low-rank systematics data model $\boldsymbol{L}$ while recognizing the incomplete model for $\boldsymbol{N}$. We introduce a spatial side constraint of total-variation measure (Appendix C) on the normalized coefficient matrix $\bar{\boldsymbol{C}}$ to facilitate a more robust separation of the unknown astrophysical signals. The total-variation constraint admits a small number of bounded discontinuities within generally smooth functions (Rudin et al. 1992; Vogel 2002) and promotes correlation between neighboring light-curve systematics. This choice of spatial constraint is experimentally motivated in Section 2.5. In this formulation, the estimate for $\boldsymbol{L} = \boldsymbol{V}\boldsymbol{C}$ is obtained by minimizing an objective function comprising the least-square residual $f(\boldsymbol{V}, \boldsymbol{C}) = ||\boldsymbol{Y} - \boldsymbol{V}\boldsymbol{C}||_F^2$ and the total-variation penalty constraint $g(\boldsymbol{C}) = \|\boldsymbol{D}_{\boldsymbol{W}}\bar{\boldsymbol{C}}\|_{2,p}^p$ with $p \in [1, 2]$ (Appendix C):

$$\begin{aligned} &\arg\min_{\boldsymbol{V},\boldsymbol{C}:\,\mathrm{rank}(\boldsymbol{V}\boldsymbol{C})\leqslant K} \{f(\boldsymbol{V}, \boldsymbol{C}) + g(\boldsymbol{C})\} \\ &= \arg\min_{\boldsymbol{V},\boldsymbol{C}:\,\mathrm{rank}(\boldsymbol{V}\boldsymbol{C})\leqslant K} \{||\boldsymbol{Y} - \boldsymbol{V}\boldsymbol{C}||_F^2 + \|\boldsymbol{D}_{\boldsymbol{W}}\bar{\boldsymbol{C}}\|_{2,p}^p\}. \end{aligned} \tag{7}$$

Here $\boldsymbol{D}_{\boldsymbol{W}} = \boldsymbol{W}\boldsymbol{D}$ where $\boldsymbol{D}$ is a difference operator, as defined in Appendix C, and $\boldsymbol{W}$ is a diagonal matrix of weights $w^{x,y}$ between [0, 1] to model nonuniform spatial correlation across the sensor. We assume each light curve i has a pixel position $i \rightarrow (x,\ y) \in (X,\ Y)$, where $X, Y \in \mathbb{Z}^+$. Further, we assume uniform spatial cells of size $(\Delta x, \Delta y)$ across the sensor and that there is a one-to-one mapping between light curve i and spatial cell.

The weighted difference operator $\boldsymbol{D}_{\boldsymbol{W}}$ applied to $\bar{\boldsymbol{C}}$ computes for each $k \in K$ and cell $(x,\ y)$, the weighted difference of neighboring coefficients $[\bar{c}_{x,y}^k - \bar{c}_{x,y+1}^k, \bar{c}_{x,y}^k - \bar{c}_{x+1,y}^k] \in \mathbb{R}^2$ as the columns of $\boldsymbol{D}_{\boldsymbol{W}}\bar{\boldsymbol{C}} \in \mathbb{R}^{2\,\times\,K\cdot X\cdot Y}$ and where $\bar{c}_{x,y}^k = \bar{c}_{i\,\rightarrow(x,y)}^k$. This is depicted graphically in Figure 1. In expanded form $\|\boldsymbol{D}_{\boldsymbol{W}}\bar{\boldsymbol{C}}\|_{2,p}^p = \sum_{k\in K}\sum_{(x,y)} w^{x,y}(|\bar{c}_{x,y}^k - \bar{c}_{x,y+1}^k|^2 + |\bar{c}_{x,y}^k - \bar{c}_{x+1,y}^k|^2)^{\frac{p}{2}}$.

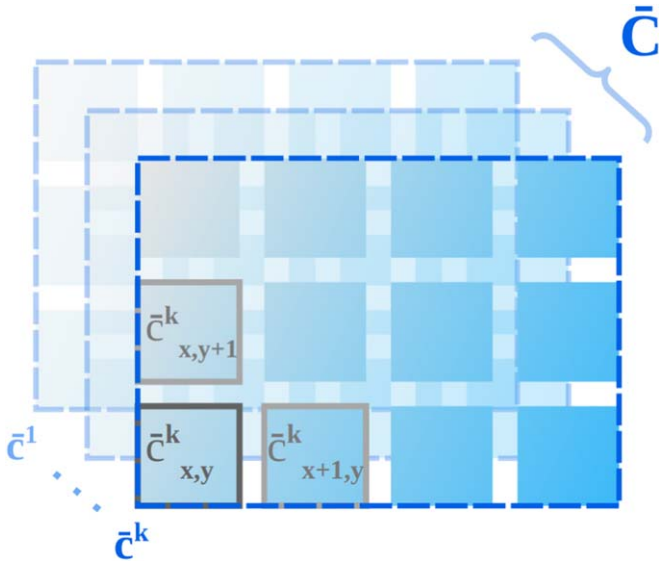


Figure 1. Visual representation of the coefficient matrix $\bar{\boldsymbol{C}}$. Each image layer $\bar{\boldsymbol{C}}^k$ represents the coefficient weights for a basis term k, ($[\bar{\boldsymbol{C}}]_k$) organized by 2D sensor cell position. The mapping of light curve i to cell position $i \rightarrow (x, y)$ is implicit in this notation, as described in the main text.

When $p = 2$ minimizing the total-variation spatial constraint is equivalent to maximizing the correlation between neighboring coefficient vectors; this is shown in Appendix C.1.

Placing the spatial prior on the normalized coefficients $\bar{\boldsymbol{C}}$ instead of the full systematics model $\boldsymbol{L} = \boldsymbol{VC}$ has mathematical advantages as described in Section 2.4 and is shown in Appendix E to promote correlation between neighboring systematics estimates by proxy.

The prior is calculated using normalized coefficients $\bar{\boldsymbol{C}}$ to avoid the objective being minimized to a trivial minimum. If instead the prior was of the form $\|\boldsymbol{D_W C}\|^p_{2,p}$ any solution $\boldsymbol{C}$ can be replaced with a solution $\boldsymbol{C}' = \alpha \boldsymbol{C}$ for $\alpha < 1$ that produces a lower value of the spatial prior $\|\boldsymbol{D_W C}'\|^p_{2,p} = \alpha \|\boldsymbol{D_W C}\|^p_{2,p}$. Taking $\boldsymbol{V}' = \frac{\boldsymbol{V}}{\alpha}$, there is no change to the least-squares penalty $\|\boldsymbol{Y} - \boldsymbol{VC}\| = \|\boldsymbol{Y} - \boldsymbol{V}'\boldsymbol{C}'\|$; therefore a trivial minimum can be found as $\alpha \rightarrow 0$.

Basis vectors 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 $\boldsymbol{V}$ or $\boldsymbol{C}$ to be orthogonal.

2.4. Systematics Inference

The systematics $\boldsymbol{L} = \boldsymbol{VC}$ are inferred by minimizing the objective function defined in Equation (7). This function is defined over variables $\boldsymbol{V} \in \mathbb{R}^{N \times K}$, $\boldsymbol{C} \in \mathbb{R}^{K \times I}$, and comprises the least-squares residual $f(\boldsymbol{V}, \boldsymbol{C}) = \|\boldsymbol{Y} - \boldsymbol{VC}\|^2_F$, which depends on both variables $\boldsymbol{V}$, $\boldsymbol{C}$, and the total-variation penalty constraint $g(\boldsymbol{C}) = \|\boldsymbol{D_W}\bar{\boldsymbol{C}}\|^p_{2,p}$, which only depends on $\boldsymbol{C}$. The least-squares residual $f(\boldsymbol{V}, \boldsymbol{C})$ is not separable over $\boldsymbol{V}$, $\boldsymbol{C}$.

Coupled least-squares problems of this type can be solved by iterative alternating solutions for $\boldsymbol{V}$ and $\boldsymbol{C}$ (Wold 1973; Tamuz et al. 2005) or by using variable elimination to remove dependence on one of the variables (Golub & Pereyra 2003). There is no evidence that either is preferable and we adopt the latter variable-projection method here.

For fixed $\boldsymbol{V}$ or fixed $\boldsymbol{C}$, the least-squares residual $f(\boldsymbol{V}, \boldsymbol{C})$ has an analytic solution for the complementary free variable. Therefore we can eliminate variable dependence on $\boldsymbol{V}$ in the least-squares residual so that the objective can be minimized over $\boldsymbol{C}$ alone. For a value of $\boldsymbol{C}$, the minimizing value $\boldsymbol{V}$ of the least-squares residual is denoted as a function $h(\boldsymbol{C})$:

$$h(\boldsymbol{C}) = \arg\min_{\boldsymbol{V}} f(\boldsymbol{V}, \boldsymbol{C}). \tag{8}$$

This function $h(\boldsymbol{C})$ can be inserted as $\boldsymbol{V}$ into the least-squares residual so that the original objective depends only on $\boldsymbol{C}$:

$$\min_{\boldsymbol{V},\boldsymbol{C}} \{f(\boldsymbol{V}, \boldsymbol{C}) + g(\boldsymbol{C})\} \rightarrow \min_{\boldsymbol{C}} \{f(h(\boldsymbol{C}), \boldsymbol{C}) + g(\boldsymbol{C})\}. \tag{9}$$

Using the theory described in Golub & Pereyra (2003), as elaborated in Appendix F.1, the variable-reduced objective takes the form:

$$\arg\min_{\boldsymbol{C}:\ \mathrm{rank}(\boldsymbol{C}) \leqslant K} \{\|(\mathbb{1}_I - \boldsymbol{C}^T\boldsymbol{C}^\dagger)\boldsymbol{Y}^T\|^2_F + \|\boldsymbol{D_W}\bar{\boldsymbol{C}}\|^p_{2,p}\}. \tag{10}$$

The minimum of the derived objective does not have an analytic solution but can be obtained numerically using gradient-descent optimization methods. However, the L_p norm in the total-variation constraint $g(\boldsymbol{C})$ is not directly differentiable as it is discontinuous for zero-valued entries of $\boldsymbol{D_W}\bar{\boldsymbol{C}}$ when $p = 1$. Proximal methods are standard for obtaining the minima of objective functions with nondifferentiable penalties such as L_1 norms, as described in Parikh & Boyd (2013). In particular, the Split–Bregman method (Goldstein & Osher 2009) is well suited to total-variation-regularized problems. A preferred approach is to replace the nondifferentiable L_p norm with a continuously differentiable Huber loss function as an approximation (Vogel 2002). Our general gradient-descent algorithm to minimize the variable-reduced objective is described in Algorithm 1. We refer to the variable-reduced objective function in Equation (10) as $w(\cdot)$, the gradient of $w(\cdot)$ with respect to $\boldsymbol{C}$ is denoted $\nabla w(\cdot)$. The gradient of the variable-reduced least-squares term $f'(\boldsymbol{C}) = f(h(\boldsymbol{C}), \boldsymbol{C}) = \|(\mathbb{1}_I - \boldsymbol{C}^T\boldsymbol{C}^\dagger)\boldsymbol{Y}^T\|^2_F$ in the objective is denoted $\nabla f'(\cdot)$ and the gradient of the total-variation constraint term $g(\boldsymbol{C}) = \|\boldsymbol{D_W}\bar{\boldsymbol{C}}\|^p_{2,p}$ as $\nabla g(\cdot)$. These gradients are used to iteratively update $\boldsymbol{C}^t$ at each timestep t. We note that this variable-projection algorithm also updates $\boldsymbol{V}$ implicitly, denoted $\boldsymbol{V}_{|\boldsymbol{C}}$. The details of the Huber loss function and the derived gradients are described in Appendix F.2.

Algorithm 1. Variable-projection Gradient Descent for the Spatial Systematics Algorithm

Initialize $\boldsymbol{C}$ for rank K and weight matrix $\boldsymbol{W}$. The step size α^t may be fixed or vary at each step t.
while $|w(\boldsymbol{C}^{t+1}) - w(\boldsymbol{C}^t)| \geqslant \epsilon$ **do**
1) $\nabla w(\boldsymbol{C}^t) = \nabla f'(\boldsymbol{C}^t) + \nabla g(\boldsymbol{C}^t)$
2) $\boldsymbol{C}^{t+1} = \boldsymbol{C}^t - \alpha^t \nabla w(\boldsymbol{C}^t)$
end
$\boldsymbol{L} = \boldsymbol{V}_{|\boldsymbol{C}}\boldsymbol{C}$ with $\boldsymbol{V}_{|\boldsymbol{C}} = \boldsymbol{YC}^T(\boldsymbol{CC}^T)^{-1}$.

2.5. Algorithm Implementation and Evaluation

In this section, we describe further details concerning the implementation and evaluation of the spatial systematics algorithm presented in this paper. As noted above, this method has been developed to explore new algorithmic approaches to exoplanet transit detection. Our prime focus is to demonstrate

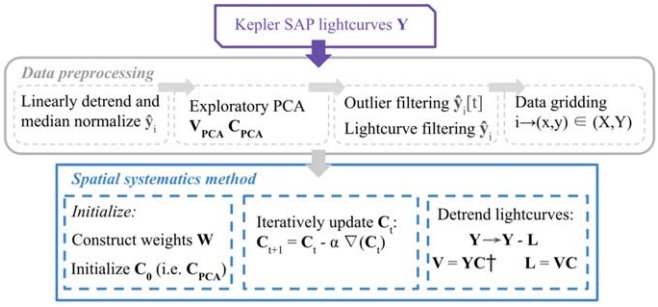


Figure 2. Data analysis framework for numerical evaluation of the spatial systematics algorithm. The data flow includes input data (purple box), preprocessing (gray box; Section 2.5.1), and application of the spatial systematics algorithm (blue box; Sections 2.4 and 2.5).

the feasibility of the algorithm and to provide an initial assessment of its performance. Accordingly, this work is not designed to support or claim the optimality of this method over existing specialized detrending approaches.

The spatial systematics algorithm was evaluated numerically on a selected subset of Kepler test data using both injection tests with time-variable signals (Section 2.5.4) and fully simulated light curves in the same Kepler coordinates (Section 2.5.5). Further, a high-level comparison between the spatial systematics algorithm and several standard detrending methods was performed over the same Kepler test data (Section 2.5.6). The Kepler test data are described in Section 2.5.1.

The data analysis framework used for the numerical evaluation tests is depicted in Figure 2. The free parameters to be chosen, shown in our core Algorithm 1 include the initial value of $\boldsymbol{C}_0$, the model rank K, the total-variation norm parameter p, the gradient step size α, and the spatial weighting matrix $\boldsymbol{W}$. The analysis framework and initial values are discussed in further detail below.

The data analysis was performed in Python and using a single CPU core. The computational time complexity of the spatial algorithm acting on a collection of light curves $I = X \times Y$ is quadratic in $|I|$ and linear in the number of gradient step iterations, and the model rank K. The leading order time complexity of a gradient step in Algorithm 1 defined in Appendix F.2 is $O(|I|^2 K)$. As for difference matrices $\boldsymbol{D}_x \in \mathbb{R}^{Y(X-1)\times YX}$, $\boldsymbol{D}_y \in \mathbb{R}^{(Y-1)X\times YX}$, per gradient step, a matrix multiplication of $\boldsymbol{D}_x^T \boldsymbol{D}_x \in \mathbb{R}^{I\times I}$ and $\boldsymbol{D}_y^T \boldsymbol{D}_y \in \mathbb{R}^{I\times I}$ with $\boldsymbol{C}^T \in \mathbb{R}^{I\times K}$ is performed. The difference matrices $\boldsymbol{D}_x$ and $\boldsymbol{D}_y$ are sparse in form and for either matrix approximately $\frac{2}{|I|}$ fraction of elements are nonzero. In our implementation, matrices $\boldsymbol{D}_x$ and $\boldsymbol{D}_y$ are generated and stored in compressed-sparse-row (CSR) format with Scipy sparse.

2.5.1. Kepler Test Data

The test data were selected from the Kepler long-cadence SAP light curves; these have an integration time of 29.4 minutes (Jenkins et al. 2010). The mapping of a stellar target to approximate sensor position repeats every four Kepler quarters (Aigrain et al. 2017) and accordingly to investigate the consistency of spatial trends over several quarters, we selected quarters (Q hereinafter) Q6, Q10, and Q14 separated at this cadence. Quarter Q6 was chosen to match Petigura & Marcy (2012) and allow for intercomparison with their prior work on spatial correlation; the decision to include successor quarters (Q10, Q14) rather than preceding quarters (Q2) was also made due to calibration issues during early Kepler operations (Van Cleve 2010). Stellar targets were selected within a range of Kepler magnitude $Kp \in [12, 13]$ and with a combined differential photometric precision over 12 hr $\mathrm{CDPP}_{12h} \leqslant 40$ (Christiansen et al. 2012),[8] resulting in 6179, 6286, and 6049 light curves for Q6, Q10, and Q14, respectively. These target selection criteria were informed by Petigura & Marcy (2012); however, we do not believe that our numerical algorithm evaluation is highly sensitive to these exact parameter choices. The narrow range in photometric magnitude was used to isolate the known dependence of systematics on magnitude (Stumpe et al. 2014) and to allow us a focus on spatial systematics in our algorithm.

Preprocessing: As noted in Section 2.2 the assumption of Gaussian statistical noise $\boldsymbol{N}$ is an idealization and preprocessing is necessary to address several data conditioning issues, as described here. First, any missing data values, and four visually identified outlier cadences, were substituted using linear interpolation between the preceding and following data samples spanning each gap. Median normalization was then applied to each light curve as $\hat{\boldsymbol{y}} = \boldsymbol{y}/\mathrm{med}(\boldsymbol{y}) - 1$, where med $(\boldsymbol{y})$ is the median of $\boldsymbol{y}$, and the subscript on light-curve index i (Equation (1)) is dropped for clarity. Here, each preprocessing step is sequential with generic input $\boldsymbol{y}$ and output $\hat{\boldsymbol{y}}$. We then removed a first-order linear trend from each light curve as systematics are known to depend on timescale (Stumpe et al. 2014). This is evident in Figure 3, where there is no clear correlation between short- and long-timescale effects. Removing the first-order trend allows us a focus on spatial systematics in our algorithm. While it is possible to apply our algorithm to each timescale bandwidth separately in a decoupled manner that generalization is beyond the scope of the current paper. After linear detrending, the first-order difference variance $\sigma_z^2 = \mathrm{var}(\{z\colon y_n - y_{n-1}\ \forall n\})$ and variance $\sigma^2 = \mathrm{var}(\boldsymbol{y})$ were computed for each light curve, where $n \in N$ here denotes the time sample index in an individual light curve. The light curves were then filtered to retain only those that are in the lowest 90% in both σ_z^2 and σ^2, ranked separately. In a second filtering step, light curves are removed if they do not satisfy a minimum correlation requirement. Each light curve $\boldsymbol{y}_i$ must have a correlation coefficient ρ greater than 0.6 with at least 10 other light curves $\boldsymbol{y}_j\colon j \neq i$, where the correlation coefficient is calculated as $\rho(\boldsymbol{y}_i, \boldsymbol{y}_j) = \frac{\boldsymbol{y}_i^T \boldsymbol{y}_j}{\|\boldsymbol{y}_i\| \|\boldsymbol{y}_j\|}$. To identify outlier samples the light curves were then coarsely detrended using exploratory PCA[9] (Appendix B) and all samples exceeding a 3σ point-to-point scatter were then flagged. This exploratory PCA detrending was not retained however; it was only used to flag and linearly interpolate outlier samples in the input data at this point in the pre-preprocessing. These initial preprocessing and filtering steps resulted in selecting 4749/6179 (76%), 4850/6286 (77%), and 4741/6049 (78%) of light curves for Q6, Q10, and Q14, respectively.

The final step in preprocessing concerned mapping the data to a regular spatial grid on the sensor. The Kepler sensor comprises 25 modules, of which 21 contain two CCDs across four output channels (Van Cleve & Caldwell 2016). Each CCD is 2200×1024 pixels in size and each module contains a total

[8] Light curve data were obtained from the MAST archive:archive.stsci.edu/kepler/data_search/search.php.
[9] https://scikit-learn.org/stable/

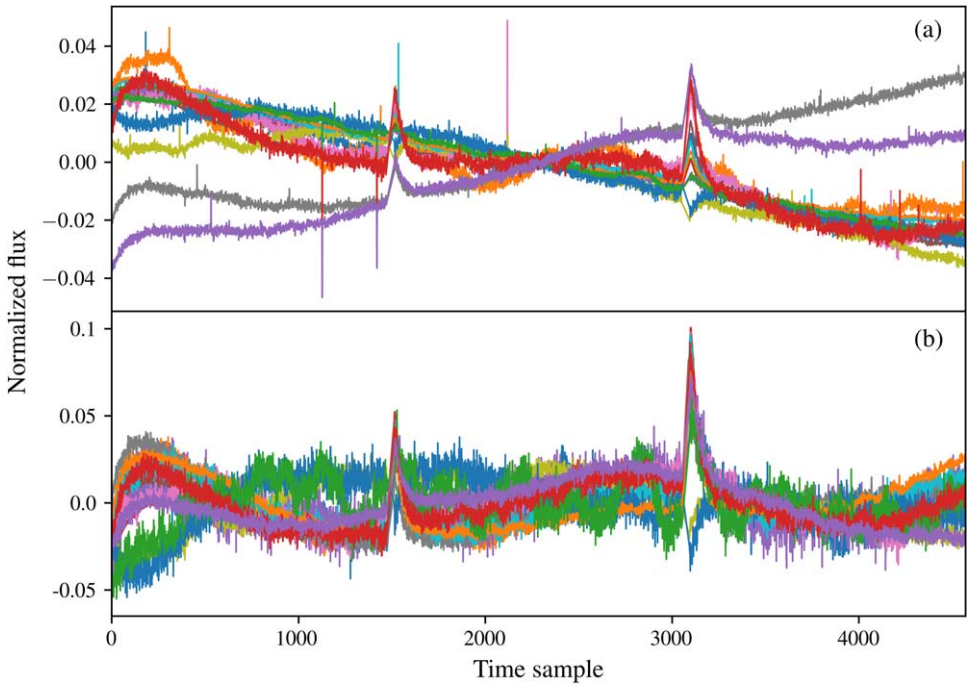


Figure 3. Representative sample of 15 light curves from Q10 shown before (a) and after (b) linear detrending and outlier removal. The x-axis index is in the unit of long-cadence time samples ($\Delta t = 29.4$ minutes). The dominant systematics visible are due to the quarterly roll and earth point recoveries (Stumpe et al. 2012).

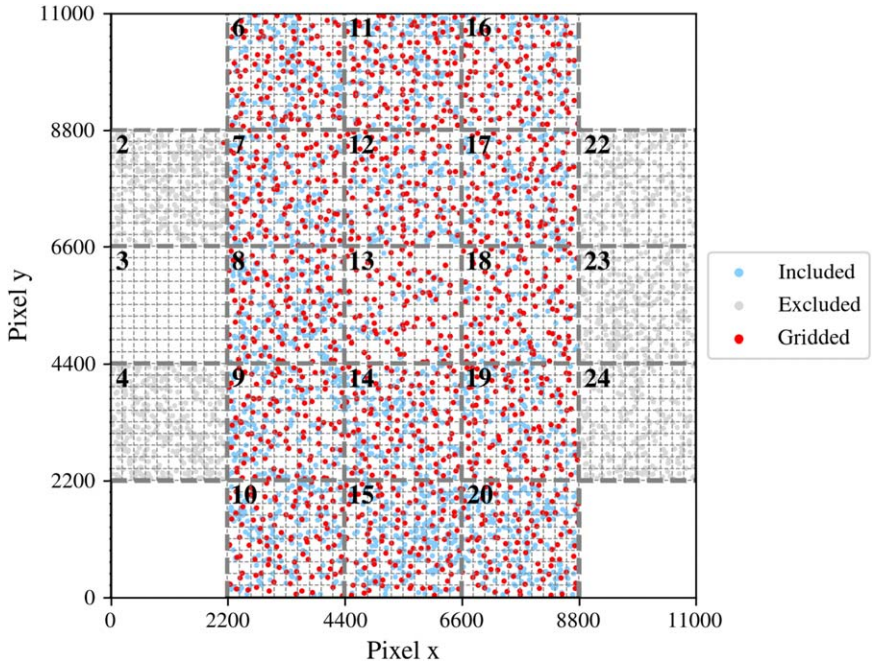


Figure 4. Positions of all (4850/6286) preprocessed Kepler test data light curves for Q10 plotted on the Kepler sensor in global pixel coordinates (11,000×11,000). Kepler modules are demarcated by bold dashed lines and labeled by module number. The (30 × 50) spatial grid is drawn in dashed lines with individual cells of size (220 × 220) pixels. All light curves with coordinates included in the rectangular gridded region are shown in blue. The subset of light curves that were gridded, and therefore processed by the spatial systematics algorithm, are shown in red at their original positions on the CCD. These light curves were either gridded to the (30 × 50) spatial cell in which they fall or to a nearby empty spatial cell, as described in the main text. After gridding, no spatial cell contained multiple light curves and there were no empty spatial cells. Light curves on modules excluded from the data grid are shown in gray. There are no targets in module 3 due to module failure.

of 2200 × 2048 CCD pixels (Van Cleve & Caldwell 2016). The pixel position of each target light curve on a CCD was mapped to a rescaled square global pixel coordinate grid across the sensor of size 11,000 × 11,000 for ease of analysis. In this process each CCD row coordinate was rescaled by a factor 1100/1024 and gaps between CCDs were removed; this resulted in convenient module coordinates of size 2200 × 2200. The total-variation constraint depends only on light-curve adjacency, so this rescaling has no impact on our algorithm.

The positions of the preprocessed test data light curves on the Kepler sensor are shown in blue in Figure 4. This Figure uses global pixel coordinates and is further labeled by Kepler module number. The difference operators $\boldsymbol{D}_x$ and $\boldsymbol{D}_y$ are easier to implement on a rectangular data grid and light curves on

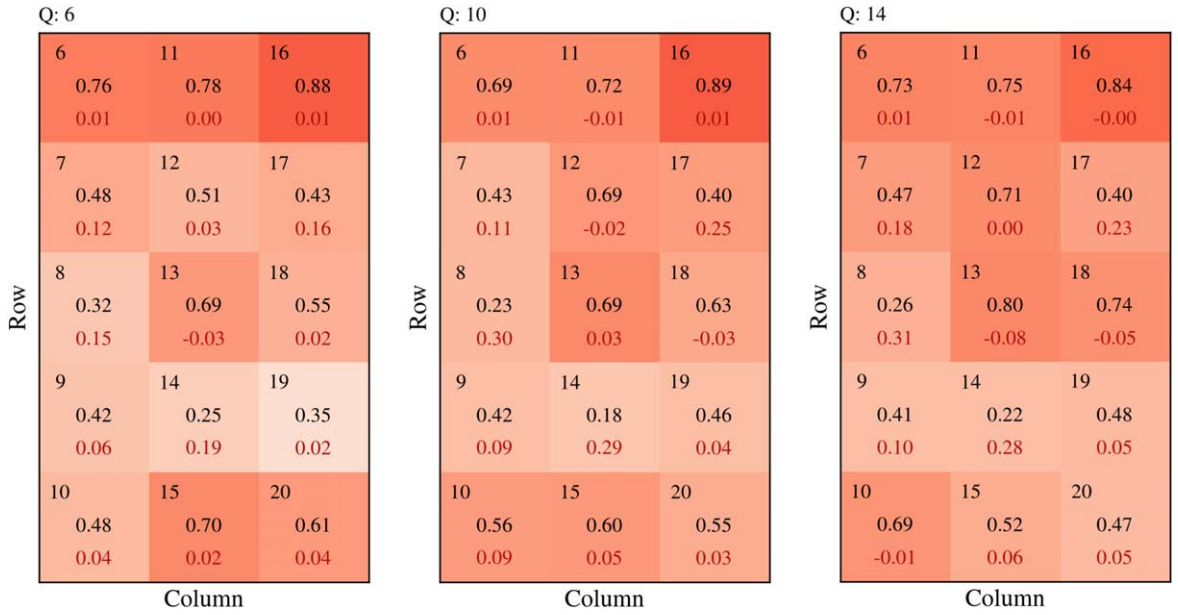


Figure 5. Spatial correlation structure within modules for Q6, Q10, and Q14 across the (30 × 50) gridded region of the Kepler sensor. The module number is shown in the top left corner of each module. The mean pairwise correlation $\bar{\rho}_M$ within each module is tabulated in black. The difference $\bar{\rho}_{M_{\rm XY}} - \bar{\rho}_M$ is tabulated in red, where $\bar{\rho}_{M_{\rm XY}}$ is the mean pairwise correlation of neighboring light curves within each module. The shaded color intensity is proportional to $\bar{\rho}_{M_{\rm XY}}$. In most modules, the correlation between neighboring light curves is greater than the mean pairwise correlation between all pairs of light curves on the module.

modules (2, 3, 4, 22, 23, 24) were discarded, with light curves on a total of 15 modules retained. However, a rectangular grid is not an algorithmic requirement. As described in Section 2.3 our implementation of the total-variation constraint requires a one-to-one mapping between light curve i and individual cells in a regular spatial grid. We adopted a spatial cell size of 220 × 220 pixels for this gridding, comprising (30 × 50) cells across the remaining modules with (10 × 10) cells per module; this grid is shown in Figure 4. The light curve mapped to each cell was selected randomly from the preprocessed light curves falling within each cell. Where a cell contained no light curves, the light curve closest to the center of the cell was selected; this occurred for 12% of 30 × 50 targets. The 1500 gridded light curves are shown in red in Figure 4.

Spatial structure on Kepler *sensor:* The pairwise neighbor correlation of the gridded light-curve sample across the sensor was measured to inform our choice of algorithm parameter $\boldsymbol{W}$ (Equation 7). The correlation between vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ is denoted $\rho(\boldsymbol{p}, \boldsymbol{q}) = \frac{\boldsymbol{p}^T \boldsymbol{q}}{\|\boldsymbol{p}\| \|\boldsymbol{q}\|}$. We denote the set of gridded cell positions within each module as $x \in X_M$ and $y \in Y_M$, respectively. The set of gridded light curves within a module is then the Cartesian set product $M = X_M \times Y_M$. The mean pairwise correlation $\bar{\rho}_M$ between all gridded light curves $(\boldsymbol{y}_m, \boldsymbol{y}_n)$ in a module was computed as:

$$\bar{\rho}_M = \frac{1}{|M|^2 - |M|} \sum_{(m,n) \in M \times M,\, (m \neq n)} \rho(\boldsymbol{y}_m, \boldsymbol{y}_n). \tag{11}$$

The mean pairwise neighboring correlation within a module $\bar{\rho}_{M_{\rm XY}}$ was calculated as:

$$\bar{\rho}_{M_{\rm XY}} = \frac{1}{2X_M Y_M - 2} \times \sum_{(x,y) \in (X_M - 1) \times (Y_M - 1)} \rho(\boldsymbol{y}_{x,y}, \boldsymbol{y}_{x,y+1}) + \rho(\boldsymbol{y}_{x,y}, \boldsymbol{y}_{x+1,y}). \tag{12}$$

The spatial correlation structure across the gridded (30 × 50) region of the Kepler sensor is shown in Figure 5. The mean pairwise correlation $\bar{\rho}_M$ for each module is tabulated in black in this Figure, and the difference $\bar{\rho}_{M_{\rm XY}} - \bar{\rho}_M$ is shown in red. The mean pairwise neighbor correlation $\bar{\rho}_{M_{\rm XY}}$ is generally higher than the mean pairwise correlation $\bar{\rho}_M$ across the module as a whole. The mean pairwise correlation for all gridded light curves over all modules was computed to be $\bar{\rho} = 0.26$ and is notably lower than the per-module correlations. Moreno et al. (2021) describe how time-dependent systematics, which lags with the radius, may plausibly give rise to spatial correlations as observed in Figure 5.

We further explored the spatial structure of the Kepler sensor by performing an exploratory PCA decomposition of the gridded light curves for Q6, Q10, and Q14. The resulting leading PCA basis vector $\boldsymbol{v}_1$ for each quarter is shown in Figure 6 along with a color map of the leading coefficient value c_i^1 in each cell. We note the implicit mapping of light-curve index i to cell position $i \rightarrow (x, y) \in (X, Y)$ defined in Section 2.3. The leading systematics basis vector is informative of general systematics as it is the strongest term. As shown in Figure 6 the leading coefficient values are highly spatially correlated, with a blocked structure per module, with anti-correlations present between modules. This effect is explained by Petigura & Marcy (2012) as due to PSF variation from the momentum cycle. The overall spatial structure is also persistent across these quarters.

As discussed in Section 1 and in Moreno et al. (2021), Petigura & Marcy (2012), and Lund et al. (2021), the systematics exhibit spatial dependence that is key to our choice of a total-variation constraint.

2.5.2. Algorithm Initial Parameter Values

As described above, there are several free parameters to be chosen in our method, including the choice of initialization $\boldsymbol{C}^0$,

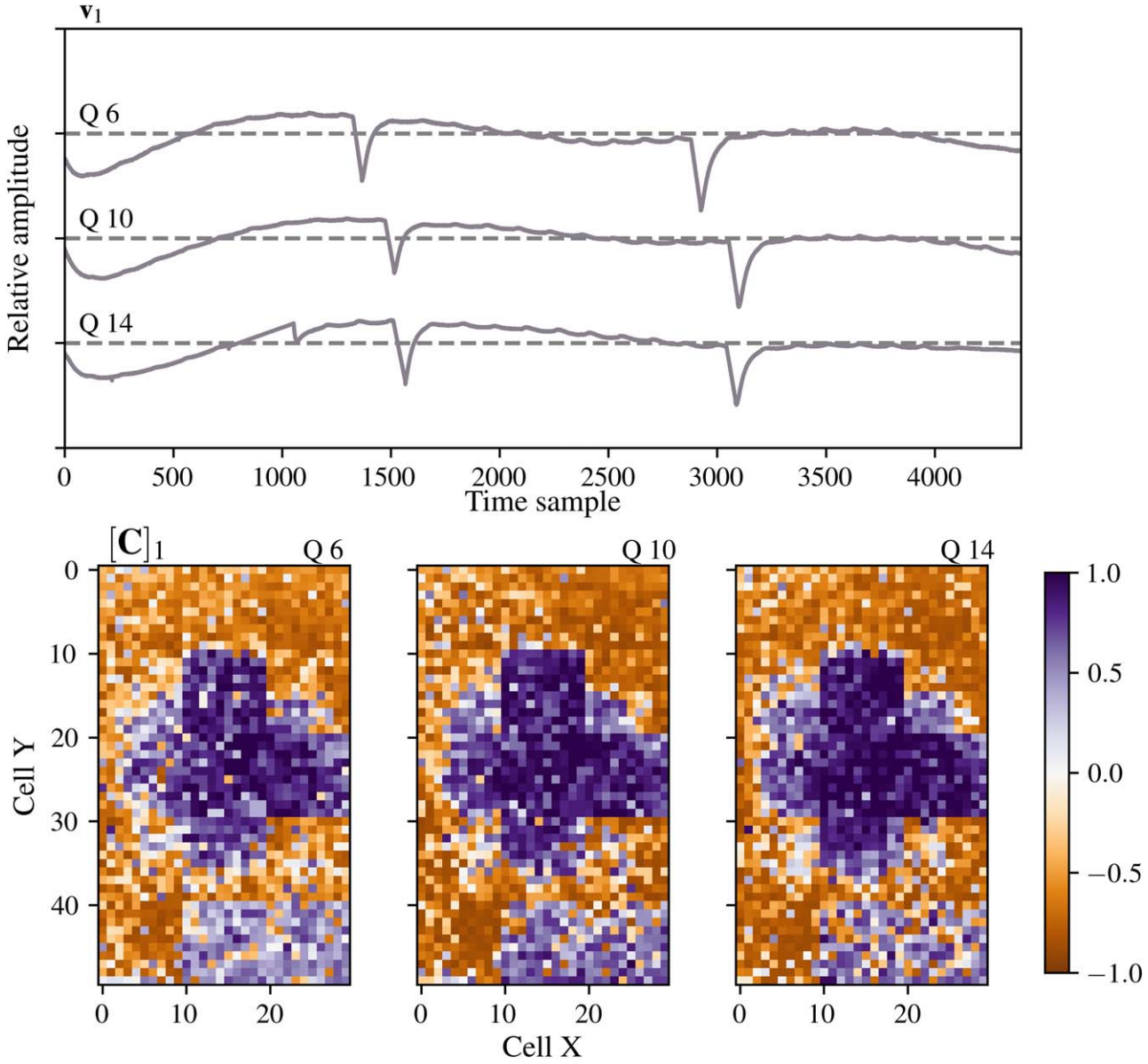


Figure 6. Leading PCA basis vector $\boldsymbol{v}_1$ for quarters Q6, Q10, and Q14 (top) and an associated color map of the leading PCA coefficient $c^1_{i\leftarrow(x,y)}$ at each spatial cell position (bottom). The spatial structure of the leading coefficient term is persistent between quarters.

the model rank K, the choice of prior norm p, the gradient step size α, and the prior weighting matrix $\boldsymbol{W}$.

The weighting matrix $\boldsymbol{W}$ was formed as neighboring pairwise correlations of gridded light curves $\boldsymbol{y}_i$ scaled by a factor 0.2 as $w^{x,y} = \frac{0.2}{2}[\rho(\boldsymbol{y}_{x,y}, \boldsymbol{y}_{x+1,y}) + \rho(\boldsymbol{y}_{x,y}, \boldsymbol{y}_{x,y+1})]$. A comprehensive parameter optimization was not performed; however, heuristic evaluation for these data (Q6, Q10, and Q14) showed slightly improved convergence for this choice of $\boldsymbol{W}$ compared to uniform weighting.

The exact model rank is unknown and for these Kepler test data, a model rank of $K = 20$ was chosen based on the singular value distribution of the light-curve sample. This singular value distribution was consistent with that shown in Smith et al. (2012) where a Kepler CBV rank of $K = 16$ was adopted.

The initial $\boldsymbol{C}^0$ was obtained from PCA applied to $\boldsymbol{Y}$. The choice of p in the total-variation constraint determines the degree of smoothness in the spatial correlation and the tolerance of discontinuities (Appendix C). In Appendix D the choice of p is interpreted in a Bayesian framework, where for $p = 1$ the spatial constraint is equivalent to a Laplacian prior on the difference of coefficients, and for $p = 2$ the spatial constraint is equivalent to a Gaussian on the difference of coefficients. Referring to Figure 6, PCA fitted coefficients representative of the underlying systematic structure show a mostly spatially uniform structure with discontinuities at module edges. We adopted a value $p = 1.1$, chosen empirically as the observed coefficients are closer to Laplacian in form.

Gradient step size α^t at iteration t was set using a backtracing line search (Armijo 1966) starting from $\alpha^t = 0.1$. This was found to allow for the minimization to progress adequately during initial iterations but with fine-tuning in later iterations, thereby saving computation time. Our stopping condition was based on the difference in our cost function in Equation 10 between successive iterates falling to a negligible level $\|w(\boldsymbol{C}^{t+1}) - w(\boldsymbol{C}^t)\| \leqslant 10^{-5}$.

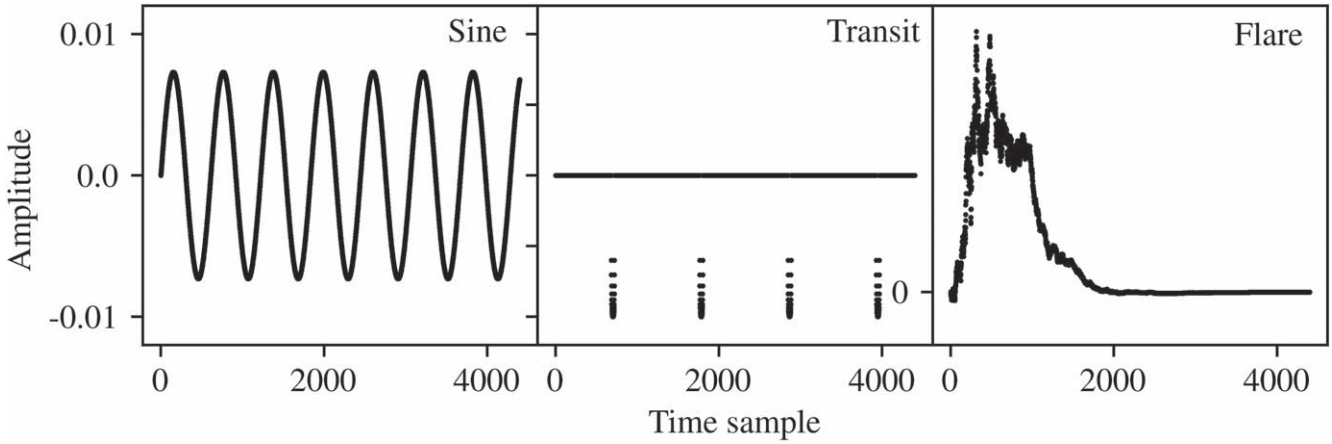


Figure 7. Example simulated sine $\boldsymbol{a}_s$, transit $\boldsymbol{a}_t$, and flare $\boldsymbol{a}_f$ signals as defined in Section 2.5.3. The inset panel (blue outline) shows the tapered transit profile at higher temporal resolution.

2.5.3. Simulated Astrophysical Signals

We describe here the functional form of the simulated transient and variable signals used in the injection tests and full simulations described below. We denote an individual simulated astrophysical signal in light curve form as vector $\boldsymbol{a}$ defined on $n \in \{1,.., N\}$ and relative to normalized light curve $\boldsymbol{y}$: $\|\boldsymbol{y}\| = 1$. The functional form of the simulated signals was chosen to be either a sinusoid ($\boldsymbol{a}_s$), a periodic exoplanet transit ($\boldsymbol{a}_t$), or a transient flare ($\boldsymbol{a}_f$). Figure 7 shows a simulated example signal of each type. These cover a representative range of timescales. The simulated signals were randomly generated relative to light curve $\boldsymbol{y}$ as follows:

1. Sine wave: $\boldsymbol{a}_s[n] = A\sin(2\pi\beta n)$. The amplitude A was drawn from the uniform distribution $U(0.005\|\boldsymbol{y}\|, 0.015\|\boldsymbol{y}\|)$. The angular frequency β was drawn from $U\left(\frac{4}{N}, \frac{8}{N}\right)$.
2. Simulated exoplanet transit signal: $\boldsymbol{a}_t[n]$. A transit signal is simulated as a periodic repeating transit profile $b[n]$, with randomized parameters: orbital period P, epoch n_0, duration d, and transit depth δ.

$$\boldsymbol{a}_t[n] = \begin{cases} \delta \cdot b[(n - n_0) \bmod P], & \text{if } (n - n_0) \bmod P \leqslant d \\ 0, & \text{otherwise} \end{cases}. \tag{13}$$

The transit profile $b[n]$ is an empirical tapered symmetric profile, defined on the first half of the transit $n \in [1, \frac{d}{2}]$ as: $b[n] = 0.4\exp(-0.3n) - 1$ and on the second half $n \in [\frac{d}{2} + 1, d]$ as $b[n] = 0.4\exp(-0.3(d - n)) - 1$. The depth of this profile generally ranges from -0.6 at transit ingress and egress to -1 at the midpoint. Transits were simulated to allow for at least three transits to occur in the light curve. The period P was drawn from the uniform distribution $U(96, 360)$, which corresponds to a range of 4–15 days. The transit epoch n_0 is drawn from the uniform distribution $U(0, P)$. The duration d of the transit profile was drawn from the uniform distribution $U(8, 24)$, which corresponds to a range of 4–12 hr. The depth δ is selected between $U(0.002\|\boldsymbol{y}\|, 0.016\|\boldsymbol{y}\|)$.
3. Simulated flare: $\boldsymbol{a}_f[n]$. The time sample index n_p of the flare peak was drawn from the uniform distribution $U(1, N)$. The flare profile is simulated as a rising exponential up to n_p followed by a slower decaying exponential. The overall signal is multiplied by a random walk stochastic process $x[n]$, generated as $x[n+1] = x[n] + w[n]$: $w[n] = \mathcal{N}(0, 1)$, $x[0] = 0$. The simulated flare is defined on the interval $n \in [1, n_p]$ as $\boldsymbol{a}_f[n] = x[n]\exp\left(-\frac{20\cdot(n_p - n)}{N}\right)$ and on the interval $n \in [n_p, N]$ as $\boldsymbol{a}_f[n] = x[n]\exp\left(-\frac{10\cdot(n - n_p)}{N}\right)$.

The simulated signals vary in strength but are generally not weaker than the sample variance of the first-order difference across each light curve (Section 2.5.1), and not greater than the magnitude of systematics. Each injected signal has on average 30% of the energy of the systematics term $\frac{\|\boldsymbol{a}_s\|}{\|\boldsymbol{y}\|} = 0.3$.

2.5.4. Experiment A: Injected Transient and Variable Signals

In this experiment (A) we numerically evaluated the performance of the spatial systematics algorithm in the recovery of simulated signals (Section 2.5.3) injected into the gridded Kepler test data (Figure 2), here selected for quarter Q10. We denote the preprocessed gridded Kepler test data here as matrix $\boldsymbol{Y}$ (Section 2.3). We simulated nine astrophysical signals, with random parameters drawn as described in Section 2.5.3, divided equally into three types: sinusoid, flare, and exoplanet transit. Each of the nine individual signals was randomly assigned to a separate randomly selected light curve in the test data in a one-to-one mapping. A priori the Kepler test data contain unknown astrophysical variability. Therefore choosing a small number of injected signals matches this realistic incidence and avoids significantly biasing the sample. We represent the injected signals as matrix $\boldsymbol{A}$, with the same shape as $\boldsymbol{Y}$, and where $\boldsymbol{A}$ is null except for nine randomly selected columns containing the simulated injection signals. The post-injection data are denoted: $\boldsymbol{Y} \rightarrow \boldsymbol{Y} + \boldsymbol{A}$.

A least-squares PCA solution was obtained using the data model $\boldsymbol{Y} = \boldsymbol{V}_{\mathrm{PCA}}\boldsymbol{C}_{\mathrm{PCA}} + \boldsymbol{N}$ (Section 2.3), yielding detrended light curves using PCA: $\boldsymbol{Y}'_{\mathrm{PCA}} = \boldsymbol{Y} - \boldsymbol{V}_{\mathrm{PCA}}\boldsymbol{C}_{\mathrm{PCA}}$. The post-injection Kepler test data $\boldsymbol{Y}$ were then detrended using the spatial systematics algorithm (Algorithm 1) using $\boldsymbol{C}_{\mathrm{PCA}}$ to initialize $\boldsymbol{C}$ and yielding detrended light curves $\boldsymbol{Y}'$. For both methods, we adopt a model rank $K = 20$. This model rank is empirically sufficient to represent the systematics and is smaller

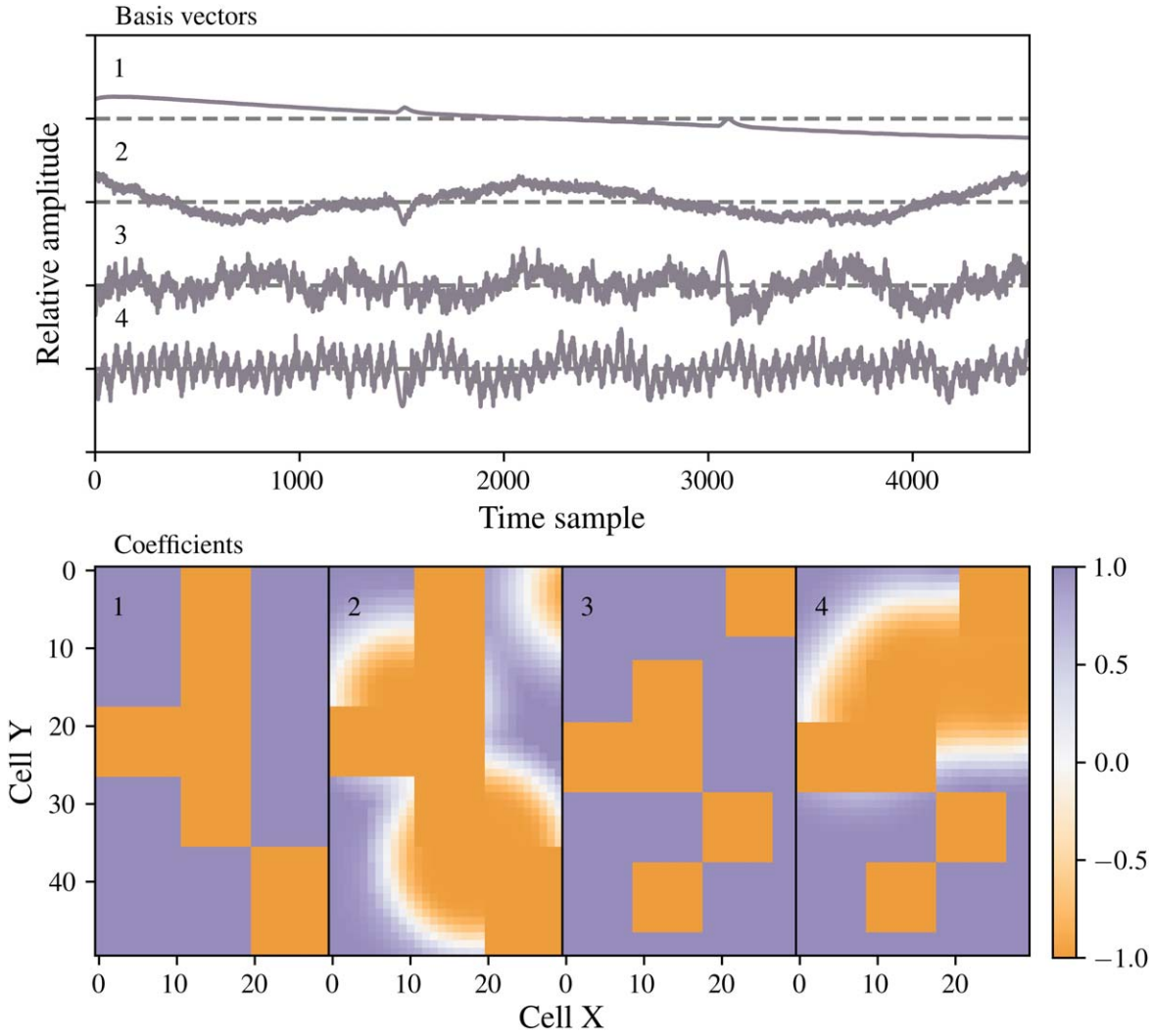


Figure 8. Basis vectors $\boldsymbol{V}$ (upper) and coefficients $\boldsymbol{C}$ (lower) used for the full simulation of light curves in experiment B using a rank $K = 4$ model. The basis vectors were selected at approximately equal spacing as index sets (1, 5, 11, 16) from the 16 Kepler CBV published for Q10 and channel 30 (Stumpe et al. 2012). The coefficient spatial dependence was simulated as a mixture of random smooth features and sharp discontinuities. Both components are arbitrarily chosen and the discontinuities are not constrained to match the Kepler module gridding.

than both the number of light curves 1500 and the length $N \sim 4000$ of each light curve.

This experiment was performed over an ensemble of ten instances of the random parameters defining the injected signals (Section 2.5.3) and random light-curve assignment from $\boldsymbol{Y}$. The performance of both the spatial systematics and PCA algorithm was evaluated in terms of the estimated level of residual systematics after detrending and in terms of their recovery of injected signals, as described in further detail below.

2.5.5. Experiment B: Fully Simulated Light Curves

In this experiment (B) we fully simulated the light curves using only the coordinate framework of the Kepler test data. No simple generative model exists to fully simulate light curves due to the complex origin of residual systematics across the sensor, as discussed in Section 1. Accordingly, we adopted a simple low-rank linear model for the systematics $\boldsymbol{L} = \boldsymbol{VC}$ with rank $K = 4$. The basis vectors $\boldsymbol{V}$ were selected at approximately equal spacing as index set (1, 5, 11, 16) from the 16 Kepler cotrending basis vectors (CBV) published for Q10 and channel 30 (Stumpe et al. 2012). The coefficients $\boldsymbol{C}$ were simulated with a random mix of discontinuities and smooth features, broadly comparable to the spatial structure found for the Kepler sensor test data but not constrained to be an exact match, including module gridding (Figure 6). We chose this spatial structure to be representative but distinct from the Kepler test data. The basis vectors $\boldsymbol{V}$ and coefficients $\boldsymbol{C}$ used in the full simulations are shown in Figure 8. Median normalization was applied to each simulated systematic term as $\hat{\boldsymbol{l}}_i = \boldsymbol{l}_i / \mathrm{med}(\boldsymbol{l}_i) - 1$, where $\mathrm{med}(\boldsymbol{l}_i)$ is the median of $\boldsymbol{l}_i$ for $i \in \boldsymbol{I}$. We note that a PCA decomposition of the simulated systematics will not be identical. The simulated basis vectors and coefficients are rank $K = 4$; however, the simulated systematics are rank $K_s = 5$. An extra basis term is introduced in a small number of light curves due to the median normalization of the simulated systematics. The corresponding basis vector is a constant offset. We simulated 1500 light curves in the 30×50 cell region, chosen to mirror the Kepler sensor region and discretization used in Section 2.5.1. A total of 300 astrophysical signals were simulated and are represented as matrix $\boldsymbol{A}$ as introduced in Section 2.5.4. Equal proportions of

sinusoid, flare, and exoplanet transit signals were simulated with random signal parameters drawn as described in Section 2.5.3. Each simulated signal was injected into a single randomly selected light curve in a one-to-one mapping. The light curves $\boldsymbol{Y}$ (Section 2.3) were simulated as $\boldsymbol{Y} = \boldsymbol{A} + \boldsymbol{VC} + \boldsymbol{N}$, where $\boldsymbol{N}$ is iid Gaussian noise $[\boldsymbol{N}]_{n,i} \sim \mathcal{N}(0, \sigma^2)$, where $\sigma^2 = \frac{1}{4\,|\,I\,|}\sum_{i\in I}\mathrm{var}(\boldsymbol{l}_i)$. The simulations were performed over one run as no ensemble was necessary given the sufficient sample size of simulated and injected astrophysical signals.

As in Experiment A (Section 2.5.4), both the PCA method and the spatial systematics algorithm were used to detrend the data $\boldsymbol{Y}$ using model rank $K = 8$, yielding $\boldsymbol{Y}'_{\mathrm{PCA}}$ and $\boldsymbol{Y}'$, respectively; as before the PCA coefficients were used to initialize the spatial systematics algorithm (Algorithm 1). This experiment deliberately simulates a high level of overfitting as the fitted model rank exceeds the rank of the simulated systematics. The performance of the algorithms was evaluated in terms of the accuracy with which the known systematics and astrophysical signals were recovered and the level of residual noise in the detrended light curves, as described in further detail below.

2.5.6. Experiment C: Comparison with Standard Detrending Methods

We performed a high-level comparison against several standard detrending methods including the Kepler cotrending basis vector method (CBV; Smith et al. 2012; Stumpe et al. 2012), self-flat-fielding (SFF; Vanderburg & Johnson 2014), and pixel-level-decorrelation (PLD; Deming et al. 2015; Luger et al. 2016, 2018), all as implemented in the Lightkurve package (Lightkurve Collaboration et al. 2018). We caution that this is a high-level comparison only and is not intended to assess the optimality of any method. For this comparison, the gridded Kepler targets (Figure 2) selected for Q10 were used. The spatial systematics method was applied to these data with rank $K = 20$ after preprocessing (Section 2.5.1; Figure 2) and with no injection of astrophysical signals. However, for the Lightkurve methods, the data for this target list were downloaded and processed inside that package as is customary.

In general, for all Lightkurve methods, we used the recommended parameters in the Lightkurve tutorial.[10] Specifically for the CBV and SFF detrending methods SAP preprocessing was performed using the *remove_nans* and *remove_outliers* functions. The PLD detrending method uses pixel-level data and no additional preprocessing was applied. The Lightkurve module masks bad quality-flagged cadences and removes ∼100 cadences per light curve for these test data. We did not remove these cadences in our processing using the spatial systematics method (Figure 2). However, in our calculation of performance metrics, we use the intersection of nonmasked cadences among Lightkurve light curves. This excludes a further ∼100 cadences per light curve. For the Lightkurve CBV detrending we used all single-scale CBVs available $K = 16$. For Lightkurve SFF detrending we specified a window of width 401 cadences for preprocessing; this removes long-term variability and flattens each light curve. The Lightkurve SFF detrending was performed using 20 such windows. No other Lightkurve input parameters were specified. The broad relative performance of the spatial systematics algorithm against the comparison of Lightkurve methods was evaluated in terms of the estimated level of residual systematics after detrending (Section 3.1).

[10] https://docs.lightkurve.org/tutorials/index.html#removing-instrumental-noise

3. Results

In this section, we describe the results of experiments A, B, and C.

3.1. Performance Metrics

The metrics used to numerically evaluate algorithm performance fall into the following two broad categories.

Residual Noise in Detrended Light Curves: This section concerns metrics used to assess the level of residual systematic noise in detrended light curves. The combined differential photometric precision (CDPP) over 6 hr is used in the Kepler pipeline and serves as an estimate of the level of white noise remaining after detrending on a transit timescale (Christiansen et al. 2012).

Following Aigrain et al. (2016, 2017) we used the procedure described in Gilliland et al. (2011) to compute an approximate measure of the CDPP in our detrended light curves. A 2D quadratic Savitsky–Golay filter was applied to the detrended light curves to remove low-frequency variability and the data were then averaged over 6 hr bins. The approximate CDPP was then computed as the standard deviation of the binned time series and scaled by a value of 1.168. The median CDPP across the set of detrended light curves is reported.

The second metric for residual systematic noise was adopted from the goodness metric defined by Stumpe et al. (2012) and is denoted here as $G_{y'}$. This metric is formed as the average absolute cubed pairwise correlation between detrended light curves as $G_{y'} = \frac{2}{N(N-1)}\sum_{i\neq j}|\rho(\boldsymbol{y}'_i, \boldsymbol{y}'_j)|^3$. Lower values of $G_{y'}$ suggest a lower residual systematic noise in detrended light curves. The cube downweights correlations, which may be spuriously low if all residual astrophysical variability is removed and higher correlations make a more significant contribution to this metric.

Cross-Correlation against Known Signals and Systematics: In this section, we consider metrics used to measure the degree to which known astrophysical signals or systematics are recovered in the injection tests or full simulations.

The injection matrix $\boldsymbol{A}$ defined above is a null matrix with a subset of columns containing simulated vector light curves that are either sinusoids ($\boldsymbol{a}_s$), flares ($\boldsymbol{a}_f$), or exoplanet transit signals ($\boldsymbol{a}_t$), as defined in Section 2.5.3. Each simulated light curve is mapped to a random light curve in a one-to-one mapping. We denote the mean correlation of all simulated astrophysical signals of type $\boldsymbol{a}_{[m]}, m \in \{s,\ t,\ f\}$ against their associated detrended light curves $\boldsymbol{y}'_{[m]}$ and averaged over all simulation runs as $\bar{\rho}(\boldsymbol{a}_{[m]}, \boldsymbol{y}'_{[m]})$. This is a measure of the degree to which the simulated or injected astrophysical signal has been recovered from the detrended light curve. When averaged across all signal types this is denoted $\bar{\rho}(\boldsymbol{a}, \boldsymbol{y}'_{\boldsymbol{a}})$.

We denote the mean absolute correlation of all simulated astrophysical signals of type $\boldsymbol{a}_{[m]}$ against the set of all detrended light curves where no astrophysical signal was injected $\boldsymbol{y}'_{\boldsymbol{a}^c}$ and averaged over all simulation runs, as $|\bar{\rho}(\boldsymbol{a}_{[m]}, \boldsymbol{y}'_{\boldsymbol{a}^c})|$, $m \in \{s, t, f\}$. The absolute value is used as the simulated signal may be anti-correlated with the estimated systematics. This correlation measures whether a detrending algorithm is corrupting light curves that do *not* contain an injected astrophysical signal ($\boldsymbol{y}'_{\boldsymbol{a}^c}$).

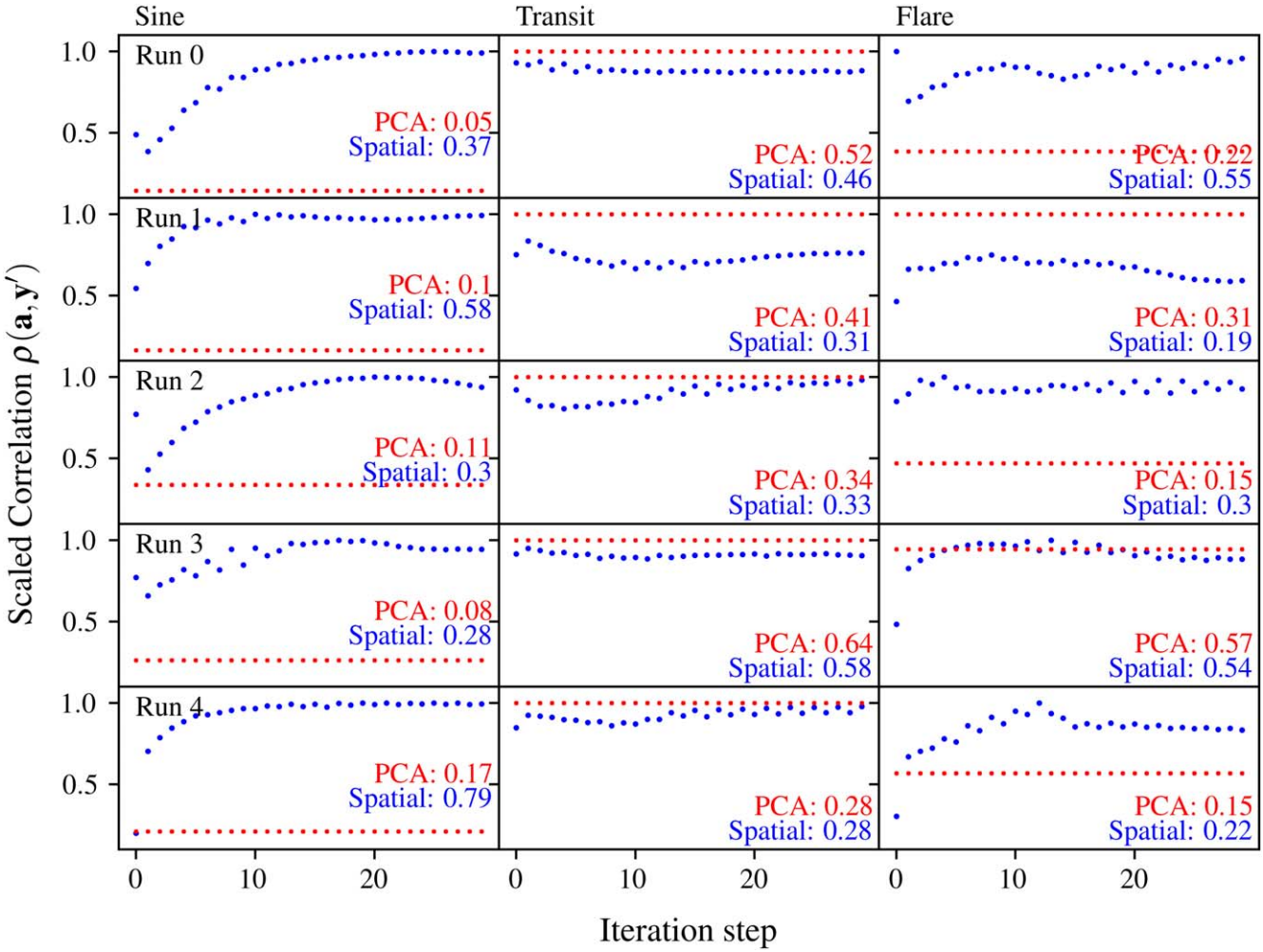


Figure 9. Scaled mean correlation of injected signals and associated detrended light curves $\bar{\rho}(\boldsymbol{a}_{[m]}, \boldsymbol{y}'_{[m]})$ for Experiment A for sine, flare, and transit signal types $m \in \{s, t, f\}$ (across columns) for each of five separate simulation runs (across rows). This mean correlation, defined in the main text, is plotted for the spatial systematics algorithm (blue) and the PCA decomposition method (red). The mean correlation is scaled by the maximum value achieved by either method; this maximum is shown in the legend of each subplot. The PCA algorithm is noniterative and is depicted as a straight line accordingly. The x-axis label is the iteration number of the spatial systematics algorithm. The spatial systematics algorithm is generally convergent and typically has comparable or improved performance relative to PCA but this is not universal for all runs.

The unprocessed light curves before detrending ($\boldsymbol{y}_{\boldsymbol{a}^c}$) may, however, be incidentally correlated with an injected astrophysical signal and therefore we normalize by the mean absolute correlation of the nondetrended light curves $\frac{|\bar{\rho}(\boldsymbol{a}, \boldsymbol{y}'_{\boldsymbol{a}^c})|}{|\bar{\rho}(\boldsymbol{a}, \boldsymbol{y}_{\boldsymbol{a}^c})|}$. Values below unity indicate that the detrending algorithm is not introducing spurious correlation with the injected signals beyond that present incidentally in the nondetrended light curves.

In Experiment B, the constituent systematic light-curve vectors $\boldsymbol{l}_i$, $(i \in I)$ comprising matrix $\boldsymbol{L}$ are also known, in addition to the injected astrophysical signals $\boldsymbol{a}_{[m]}$, $m \in \{s, t, f\}$. We denote the estimated systematics vector for light curve i as $\hat{\boldsymbol{l}}_i$. The mean correlation of the simulated $\boldsymbol{l}_i$ and estimated systematics vectors $\hat{\boldsymbol{l}}_i$ averaged over all light curves is denoted $\bar{\rho}(\boldsymbol{l}, \hat{\boldsymbol{l}})$. If this metric is restricted to light curves where an astrophysical signal $\boldsymbol{a}_{[m]}$, $m \in \{s, f, t\}$ was injected, it is denoted $\bar{\rho}(\boldsymbol{l}_{\boldsymbol{a}}, \hat{\boldsymbol{l}}_{\boldsymbol{a}})$.

3.2. Experiment A

Experiment A, as described in Section 2.5.4, was performed to numerically evaluate the spatial systematics algorithm in terms of its ability to recover transient and variable signals injected into Kepler test data. A standard PCA decomposition method was used as a comparison. The convergence of the spatial systematics algorithm for five representative runs of this experiment is shown in Figure 9. This figure plots the mean correlation $\bar{\rho}(\boldsymbol{a}_{[m]}, \boldsymbol{y}'_{[m]})$ (over the light curve) of the injected astrophysical signal types $m \in \{s, t, f\}$ as a function of the spatial systematics iteration number. This mean correlation is a proxy for the degree of recovery of injected astrophysical signals, as described in Section 3.1. Figure 9 shows the general convergence of the spatial systematics algorithm within ∼30 iterations. The figure shows signal recovery performance comparable to or exceeding the reference PCA method in the majority of cases but not without exception for individual flare and transit runs.

The performance of the spatial systematics algorithm is shown in more detail for a single representative run in Figures 10–12. Figure 10 shows the estimated basis vectors $\boldsymbol{v}_k$ obtained with the spatial systematics and reference PCA algorithms. The estimated basis vectors $\boldsymbol{v}_k$ are generally similar but differences increase at higher-order k. Figure 11 shows the

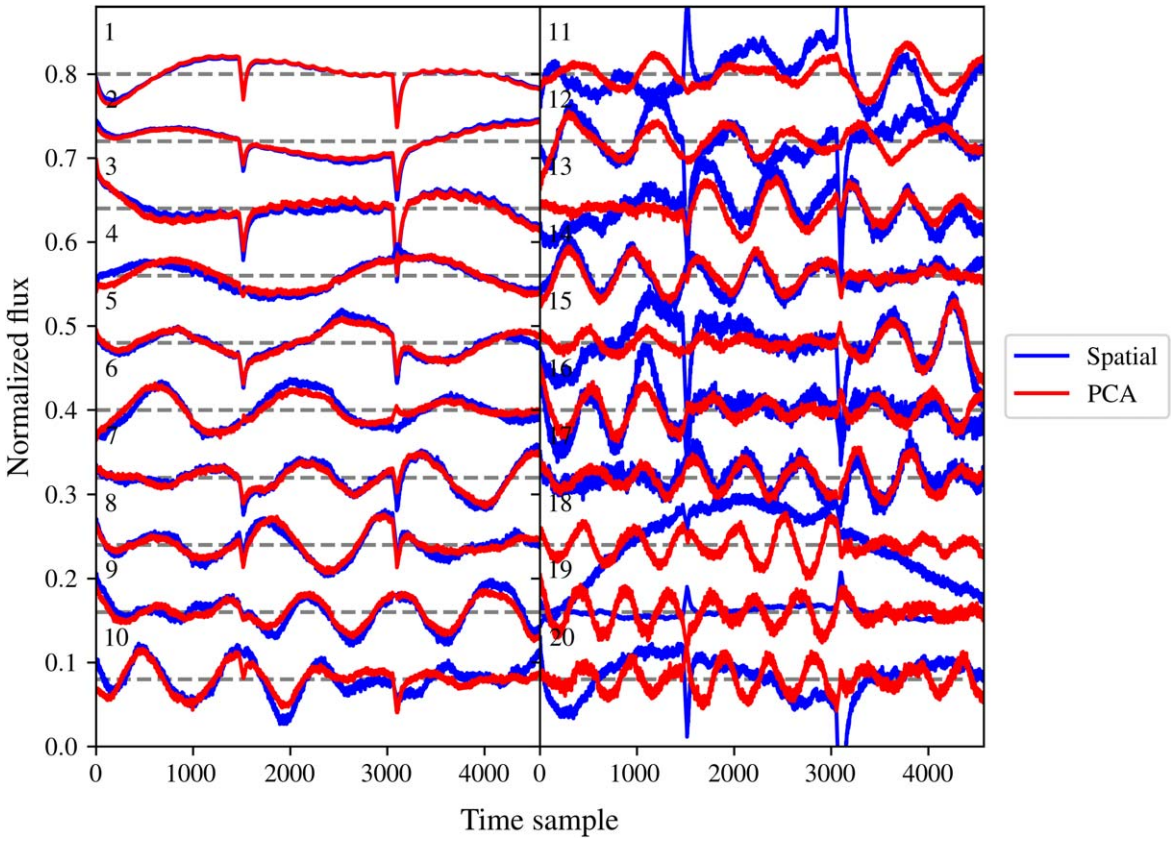


Figure 10. Estimated basis vectors ν_k, $k \in \{1, 20\}(K = 20)$ obtained for run #0, a single representative run of Experiment A, shown here for the spatial systematics algorithm (blue) and the reference PCA method (red). The spatial basis vectors and PCA basis vectors closely overlap for the first two terms.

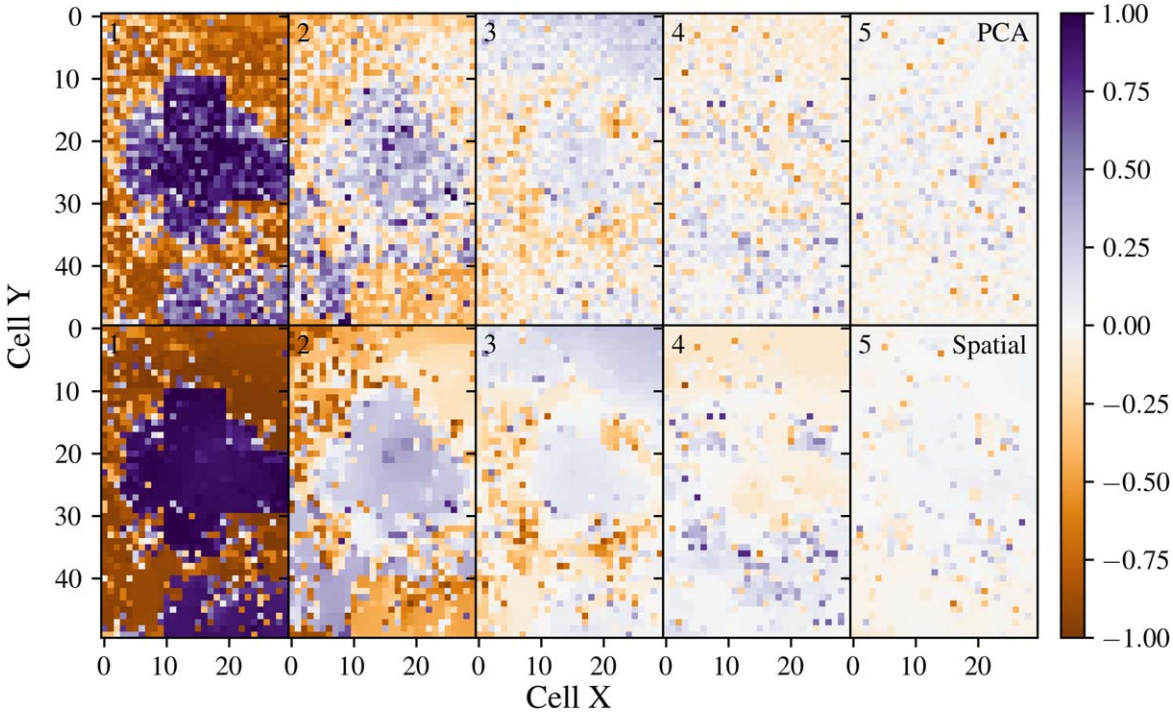


Figure 11. Basis vector coefficients c_i^k, $k \in \{1 ,..., 5\}(K = 20)$ for light curve $i \rightarrow (x, y) \in (X, Y)$ obtained using the spatial systematics (top row) and PCA algorithms (bottom row) for run #0, a representative run of Experiment A. The coefficient index k is shown in the top left of each subfigure. The x- and y-axes are in units of gridded spatial cells.

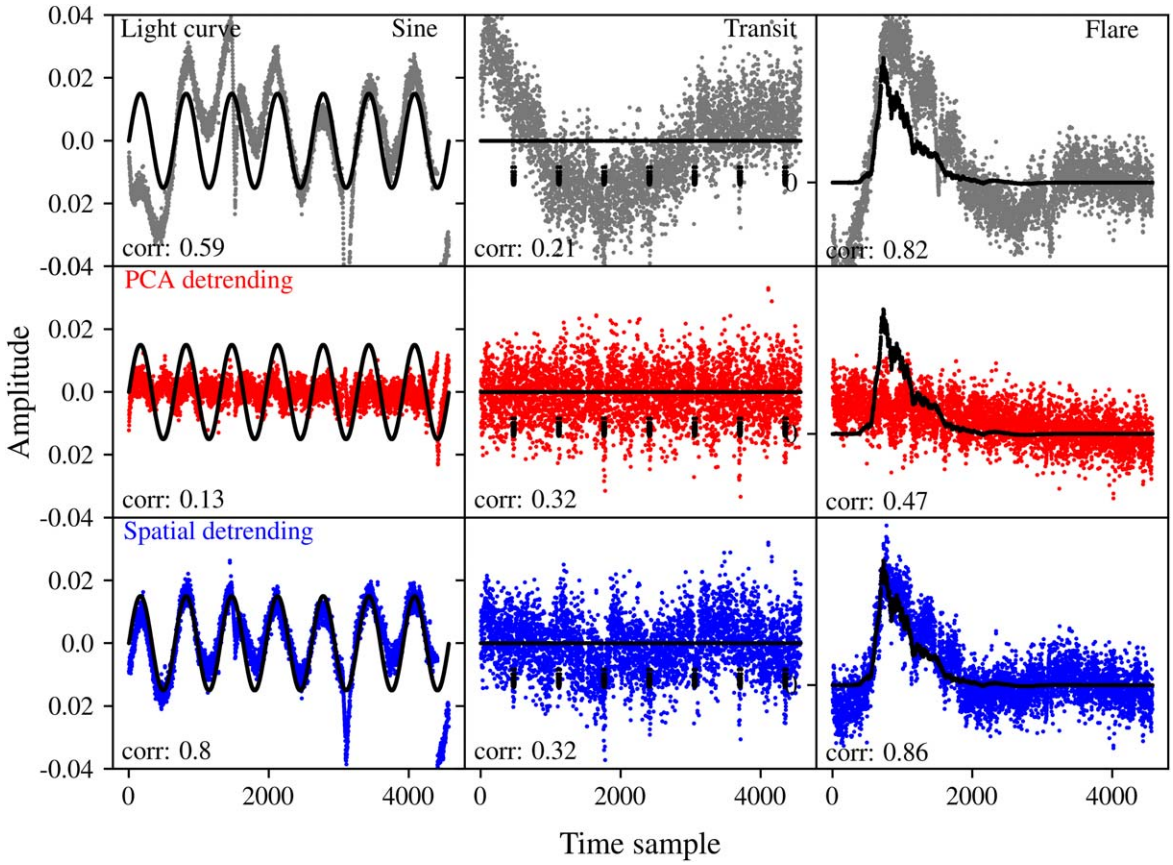


Figure 12. Three sample light curves across injected signal type $\boldsymbol{a}_{[m]}$ $m \in \{s, t, f\}$ (sine, transit, and flare) from run #0, a representative run of Experiment A. The figure shows the preprocessed light curve before detrending $\boldsymbol{y}$ (upper row), the detrended light curve using PCA $\boldsymbol{y}'_{\mathrm{PCA}}$ (middle row), and the detrended light curve using the spatial systematics algorithm $\boldsymbol{y}'$ (lower row). Simulated injected signals are overlaid in black. The x-axis is Kepler's long-cadence time sample index. This sample shows an instance where PCA detrending removed the injected astrophysical sine and flare signals, while spatial detrending preserved those signals.

fitted coefficients c_i^k, $k \in \{1,..,5\}$ ($K = 20$) obtained by the spatial systematics and PCA algorithms across the gridded light curves $i \rightarrow (x,\ y) \in (X,\ Y)$. This figure shows that the total-variation constraint in the spatial systematics algorithm enforces smoothness in the coefficients while preserving discontinuities. In Figure 12, examples of detrended light curves are shown over injected astrophysical signal type. In this example, the spatial systematics algorithm has preserved the injected astrophysical variability to a greater degree than the reference PCA method. However, the spatial systematics algorithm has a higher residual scatter in the detrended light curves than PCA, as is visible near cadence 3000 for the light curve injected with a sine signal.

The summary noise metrics (Section 3.1) for Experiment A for the final iteration of the spatial systematics algorithm and averaged over all runs in the ensemble are listed in Table 1. This table shows that for these simulations the spatial systematics algorithm generally outperforms the reference PCA method with caveats discussed in further detail in Section 4.

3.3. Experiment B

Experiment B (Section 2.5.5) was designed to numerically evaluate the spatial systematics algorithm using fully simulated data generated using only the Kepler test data coordinates. The algorithm was evaluated in terms of its ability to recover the known astrophysical signals and systematics used in simulating the light curves.

The simulated systematics are rank $K_s = 5$ but were estimated with rank $K = 8$ as described above. The estimated and simulated basis vectors $\boldsymbol{v}_k$, $k \in \{1,..,8\}$ for Experiment B are shown in Figure 13. The first four estimated basis vectors are identical between the spatial systematics and PCA algorithms. Higher-order basis vectors deviate between the two methods and are also poorly constrained. However, fitted coefficients for these basis terms ($K > 4$) are small. For the spatial systematics algorithm, these fitted coefficients contribute $\sim$4% in total magnitude ($\sum_{k=5}^{K} \|c_k\| / \sum_{k=1}^{K} \|c_k\|$), while for PCA the contribution is $\sim$11%. Figure 14 shows the true simulated coefficients and those estimated using the PCA and spatial systematics algorithms for the first 5/8 basis terms. The coefficients estimated using PCA have a more speckled spatial substructure than the smoother distribution obtained by the spatial systematics algorithm.

Example individual detrended light curves for sine, transit, and flare simulated signal types are shown in Figure 15. Summary results for Experiment B, in terms of the metrics described in Section 3.1, are provided in Table 2.

3.4. Experiment C

Experiment C was designed to provide a high-level comparison of the spatial systematics algorithm against the standard CBV, SFF, and PLD methods (Section 2.5.6). The

Table 1
Experiment A Results

	Sine Recovery $\bar{\rho}(\boldsymbol{a}_s, \boldsymbol{y}'_s)$	Sine Corruption $\lvert\bar{\rho}(\boldsymbol{a}_s, \boldsymbol{y}'_{\boldsymbol{a}^c})\rvert / \lvert\bar{\rho}(\boldsymbol{a}_s, \boldsymbol{y}_{\boldsymbol{a}^c})\rvert$	Transit Recovery $\bar{\rho}(\boldsymbol{a}_t, \boldsymbol{y}'_t)$	Transit Corruption $\lvert\bar{\rho}(\boldsymbol{a}_t, \boldsymbol{y}'_{\boldsymbol{a}^c})\rvert / \lvert\bar{\rho}(\boldsymbol{a}_t, \boldsymbol{y}_{\boldsymbol{a}^c})\rvert$		Flare Recovery $\bar{\rho}(\boldsymbol{a}_f, \boldsymbol{y}'_f)$	Flare Corruption $\lvert\bar{\rho}(\boldsymbol{a}_f, \boldsymbol{y}'_{\boldsymbol{a}^c})\rvert / \lvert\bar{\rho}(\boldsymbol{a}_f, \boldsymbol{y}_{\boldsymbol{a}^c})\rvert$	Signal Recovery $\bar{\rho}(\boldsymbol{a}, \boldsymbol{y}'_{\boldsymbol{a}})$	Residual Systematics	
									$G_{\boldsymbol{y}'}$	CDPP_{6h} (ppm)
PCA	0.14 ± 0.04	0.12 ± 0.02	0.45 ± 0.08	0.60 ± 0.11	PCA	0.35 ± 0.13	0.09 ± 0.03	0.31 ± 0.08	$(6.61 \pm 0.12) \times 10^{-4}$	38.91 ± 0.03
Spatial	0.57 ± 0.11	0.85 ± 0.09	0.44 ± 0.06	0.80 ± 0.12	Spatial	0.64 ± 0.15	0.54 ± 0.09	0.55 ± 0.11	$(1.73 \pm 0.04) \times 10^{-2}$	40.33 ± 0.14
Ratio(PCA/Spatial)	0.25	⋯	1.01	⋯	Ratio(PCA/Spatial)	0.54	⋯	0.56	0.04	0.97

Note. Performance metrics for detrended light curves, as defined in Section 3.1. As described there, *recovery* indicates the degree to which injected signals are preserved correctly in the detrended light curves. *Corruption* indicates the degree to which injected signals incorrectly appear in light curves that contain no injected signals. The uncertainties listed are the standard deviation of the value over the ensemble of ten runs. For reference, the correlation of nondetrended, preprocessed, and noninjected light curves $\boldsymbol{y}_{\boldsymbol{a}^c}$ with simulated astrophysical signals $\boldsymbol{a}_{[m]}$ have values $\lvert\bar{\rho}(\boldsymbol{a}_s, \boldsymbol{y}_{\boldsymbol{a}^c})\rvert = 0.07, \lvert\bar{\rho}(\boldsymbol{a}_t, \boldsymbol{y}_{\boldsymbol{a}^c})\rvert = 0.04, \lvert\bar{\rho}(\boldsymbol{a}_f, \boldsymbol{y}_{\boldsymbol{a}^c})\rvert = 0.18$ for sine, transit, and flare signals, respectively. For nondetrended preprocessed light curves, the goodness metric is $G_{\boldsymbol{y}}$ is 0.26 ± 0.01 and the median CDPP_{6h} is 51.21 ± 0.14 ppm.

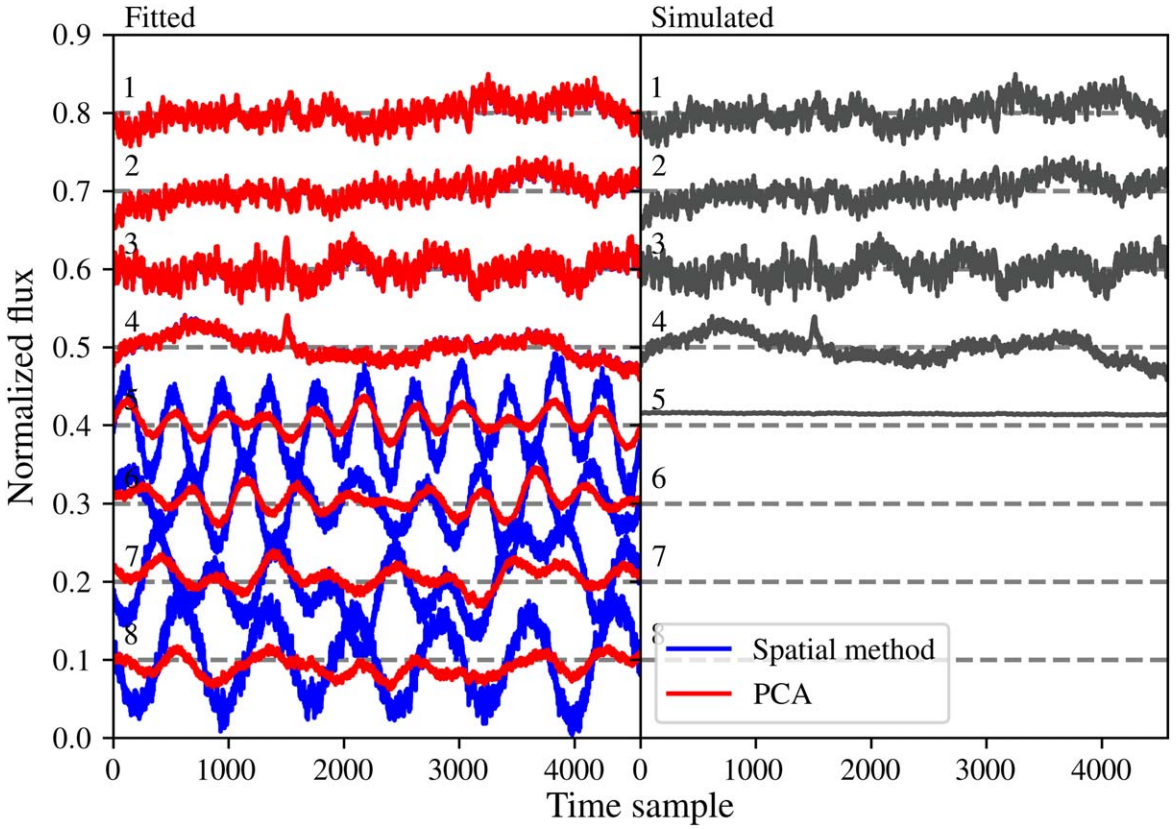


Figure 13. Estimated (left) and known simulated (right) basis vectors $\boldsymbol{v}_k$, $k \in \{1,..,8\}$ for Experiment B. The estimated basis vectors (left) are shown for the spatial systematics algorithm (blue) and the reference PCA method (red). The x-axis is in units of Kepler long-cadence time sample. The estimated spatial and PCA basis vectors (left) for the first four terms overlap one another in these plots.

performance metrics $G_{y'}$ and CDPP_{6h} for the residual noise in the detrended light curves for this experiment are shown in Table 3. A composite plot of CDPP_{6h} versus $G_{y'}$ across all methods in Experiment C is shown in Figure 16. These results show that the CDPP_{6h} metric is broadly comparable across all methods while the $G_{y'}$ metric has a greater dependence on the detrending method. As noted in Section 2.5.6, the data used in this comparison have undergone different preprocessing, and furthermore detrending methods have not been optimized in each specified case. The purpose of this comparison is to provide a high-level reference comparison and not to show the general optimality of any single method.

4. Discussion

The numerical evaluation of the spatial systematics algorithm presented here, using injected signals in Kepler data (Experiment A) and full simulations (Experiment B), shows that in these cases the algorithm matches or outperforms the reference PCA method for astrophysical signal recovery and achieves comparable performance for systematics removal. This is indicated by higher values of $\bar{\rho}(\boldsymbol{a}, \boldsymbol{y}'_{\boldsymbol{a}})$ and $\bar{\rho}(\boldsymbol{l}_{\boldsymbol{a}}, \hat{\boldsymbol{l}}_{\boldsymbol{a}})$, respectively, in Tables 1 and 2.

The improved recovery of injected astrophysical signals by the spatial systematics algorithm is direct evidence that the algorithm is less prone to erroneously absorbing true astrophysical variability in the systematics (overfitting) than the reference PCA method. Astrophysical variability and systematics are not separable a priori and overfitting is thus a foundational concern in the detrending of exoplanet transit light curves (Stumpe et al. 2012; Smith et al. 2018). Overfitting may remove true astrophysical variability from detrended light curves and also introduce spurious variability into other detrended light curves in the sample as these corrupted systematics are applied. A number of detrending approaches infer a basis representative of systematic effects from a correlated set of light curves Petigura & Marcy (2012), Foreman-Mackey et al. (2015) and Kepler PDC-MAP Smith et al. (2012), Stumpe et al. (2012). The set of light curves from which a set of basis vectors is constructed can be robustly filtered to exclude outliers and those with clear astrophysical variability as a means to mitigate overfitting, as in PDC-MAP (Stumpe et al. 2012). A basis consisting only of systematics (which is difficult to realize) can however still be overfitted to astrophysical features (Smith et al. 2018), as it is unlikely that a basis is completely orthogonal to all possible astrophysical signals. Spatial dependence among light curves has been identified by Petigura & Marcy (2012) and Moreno et al. (2021), whereby the latter work explains spatial correlations as time-delayed systematic effects traversing the sensor. The modified total variation prior applies a spatial correlation constraint across the sensor, which is well suited for inference of local effects of this form. Functionally, the spatial systematics model depends on systematic coefficients alone, this model is differentiable, and derived gradients are computationally tractable. The algorithm iteratively refines both the fitted coefficients and the overall systematics model. These algorithm properties help to separate the systematics and astrophysical variability and thereby mitigate overfitting. A position-based prior is used in PDC-MAP to reduce the

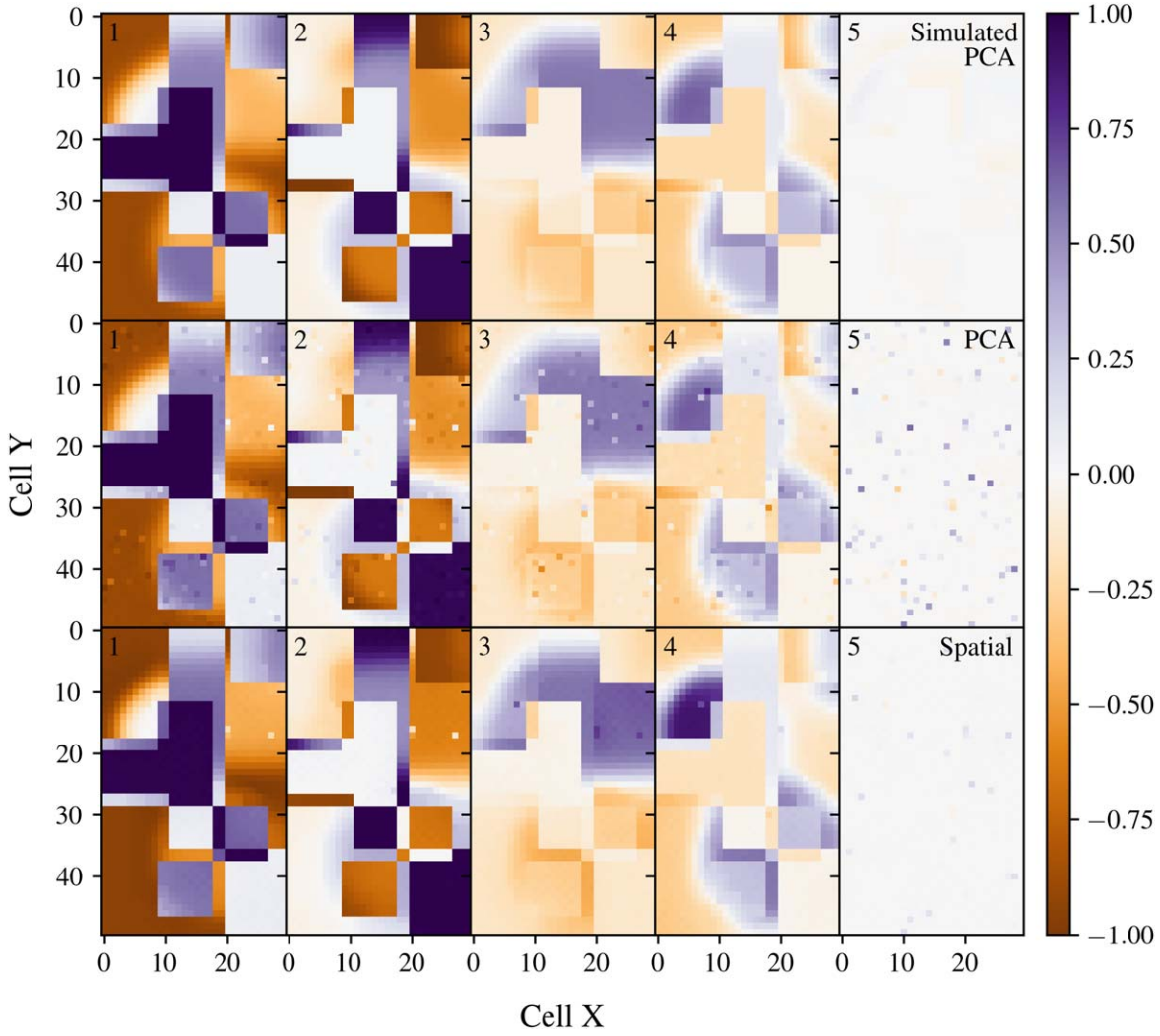


Figure 14. Coefficients c_i^k, $k \in \{1,..,5\}(K = 8)$ for light curve $i \rightarrow (x, y) \in (X, Y)$ for Experiment B. This figure shows the true simulated coefficients (top row), the coefficients obtained using the PCA method (middle row), and the coefficients obtained using the spatial systematics algorithm (bottom row). The coefficient index k is shown in the top left of each subfigure. The x- and y-axes are in units of gridded spatial cells. Coefficients fitted with PCA show speckling, indicative of the overfitting of injected astrophysical signals.

overfitting of cotrending basis vectors to a light curve. The spatial systematics algorithm is novel in that it uses a spatial constraint computed directly from systematic estimates, in total-variation form, and further iteratively refines the overall systematics solution for a collection of light curves. We emphasize the comparative performance to standard PCA and while we perform a high-level comparison to specialized detrending methods (Experiment C), we do not claim optimality for this exploratory algorithm development. The reduced overfitting is most evident in Tables 1 and 2 for sine $\bar{\rho}(\boldsymbol{a}_s, \boldsymbol{y}_s')$ and flare $\bar{\rho}(\boldsymbol{a}_f, \boldsymbol{y}_f')$ signals and comparable for transit signals $\bar{\rho}(\boldsymbol{a}_t, \boldsymbol{y}_t')$. This is also visible in the sample detrended light curves shown in Figures 12 and 15. In summary, the longer-duration astrophysical signals (sine, flare) have reduced overfitting relative to the transit signals, which are of shorter duration. The shorter the duration of an astrophysical signal relative to the timescale of systematics present in Kepler light curves (Figure 10), the less linearly dependent on an inferred systematics basis and therefore the smaller the expected improvement over the reference PCA method. Notably, the spatial systematics algorithm performs well over all astrophysical signal characteristic timescales considered here. The spatial systematics algorithm contains no explicit model for transient or variable astrophysical signals; however, it performs well for the signal types considered here (Section 2.5.3). The consistent relative performance between signal types in Experiments A and B further supports the preceding proposed causal link between the signal and characteristic residual systematic timescales, rather than the exact functional form of the astrophysical signal.

Further evidence for the improved mitigation of overfitting by the spatial systematics algorithm over the reference PCA method is provided by Experiment B. As a full simulation, this experiment includes no unknown astrophysical variability beyond the simulated astrophysical signals. In Figure 14 the true simulated coefficients shown in the top row are spatially smooth but with realistic module discontinuities. In contrast, the coefficients in the second row estimated using the PCA method show speckling in smooth regions. These discrepant coefficient values correspond to cells containing light curves with simulated astrophysical signals that are overfitted by the PCA method. The coefficients estimated by the spatial systematics method shown in the bottom row of this figure

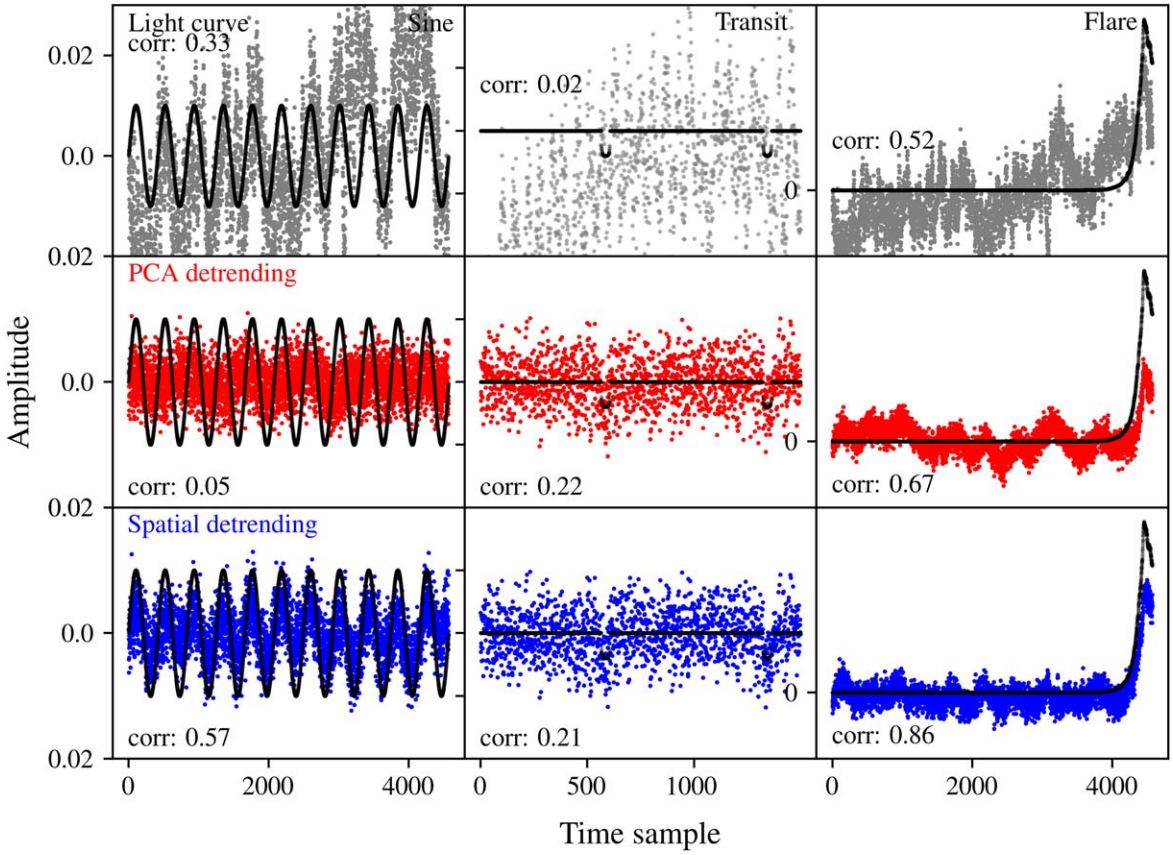


Figure 15. Three sample light curves across injected signal type $\boldsymbol{a}_{[m]}$ $m \in \{s, t, f\}$ (sine, transit, and flare) for Experiment B. The figure shows the simulated light curve before detrending $\boldsymbol{y}$ (upper row), the detrended light curve using PCA $\boldsymbol{y}'_{\mathrm{PCA}}$ (middle row), and the detrended light curve using the spatial systematics algorithm $\boldsymbol{y}'$ (lower row). Simulated injected signals are overlaid in black. The x-axis is Kepler's long-cadence time sample index. This sample shows an instance where PCA detrending moderately removed the injected astrophysical sine and flare signals, while spatial detrending more clearly preserved those signals.

generally have reduced speckling and reflect the true coefficients with greater fidelity. Figure 17 shows the cells where an astrophysical signal was injected into a simulated light curve, alongside fitted coefficients from PCA and the spatial systematics method. This figure supports the argument that PCA has overfitted light curves containing astrophysical variability. However, we note that not all individual cells containing injected astrophysical signals are overfitted by PCA. A light curve and the associated detrending results corresponding to an overfit cell in Figure 17 are shown in Figure 18. The spatial systematics algorithm varies both the coefficients and basis vectors (Figure 13) in the iterative fit. Empirically, in Experiment B the spatial systematics algorithm estimates systematics $\hat{\boldsymbol{l}}$ with a correlation $\bar{\rho}(\boldsymbol{l}, \hat{\boldsymbol{l}})$, against true simulated systematics $\boldsymbol{l}$, which is comparable to the reference PCA method. The correlation $\bar{\rho}(\boldsymbol{l}_a, \hat{\boldsymbol{l}}_a)$ against systematics for light curves containing a simulated astrophysical signal, and therefore a higher risk of overfitting, is slightly higher for the spatial systematics algorithm relative to the reference PCA method. This implies the spatial systematics algorithm shows reduced overfitting relative to the reference PCA method. We believe this is due to the spatial total-variation constraint appropriately estimating spatially correlated systematics.

As a definitional consequence of the reduction in overfitting, the detrended light curves obtained using the spatial systematics algorithm will retain a higher degree of true astrophysical variability relative to those obtained using the reference PCA method in addition to a component for uncorrected systematics, in keeping with all detrending algorithms. We note that PCA is guaranteed to maximally flatten a collection of light curves as the PCA solution is equivalent to a minimal least-squares residual between fitted systematics and light curves (Equation (6)). PCA implicitly assumes a white noise model and is susceptible to overfitting of astrophysical signals (Candès et al. 2011) because astrophysical signals are not appropriately modeled as white noise Pont et al. (2006). In both experiments, the two algorithms obtained low goodness metric values $G_{y'} < 2\%$ indicating largely successful systematics removal in the detrended light curves. The residual systematics metrics $G_{y'}$ and CDPP_{6h} are plotted for run #0 of Experiment A per light curve and over-detrending algorithm in Figures 19 and 20. Figure 19 shows that the distribution of the goodness metric $G_{y'}$ is broadly larger than that for PCA. This metric (Section 3.1) is nonlinear and pairwise, and therefore sensitive to those light curves with poorly modeled systematics (Figure 20). For example, individual light curves may be poorly modeled by the spatial systematics algorithm if their systematics exhibit strong temporal variation insufficiently captured by the fixed spatial model. It is also possible that the spatial systematics algorithm retains a greater degree of astrophysical variability that biases $G_{y'}$ through incidental correlation. We lack sufficient evidence to state either claim strongly here. Future generalizations of the constraint terms are described below as potential future work.

Table 2
Experiment B Results

	Systematics Recovery		Signal Recovery				Residual Systematics	
	$\bar{\rho}(\boldsymbol{l}_a, \hat{\boldsymbol{l}}_a)$	$\bar{\rho}(\boldsymbol{l}, \hat{\boldsymbol{l}})$	$\bar{\rho}(\boldsymbol{a}_s, \boldsymbol{y}_s{}')$	$\bar{\rho}(\boldsymbol{a}_t, \boldsymbol{y}_t{}')$	$\bar{\rho}(\boldsymbol{a}_f, \boldsymbol{y}_f{}')$	$\bar{\rho}(\boldsymbol{a}, \boldsymbol{y}_a')$	$G_{y'}$	CDPP_{6h}
PCA	0.96 ± 0.08	0.99 ± 0.03	0.52 ± 0.30	0.28 ± 0.15	0.39 ± 0.17	0.39 ± 0.24	1.71×10^{-4}	0.39 ± 0.01
Spatial	0.98 ± 0.08	0.99 ± 0.03	0.63 ± 0.22	0.28 ± 0.15	0.48 ± 0.19	0.47 ± 0.24	1.55×10^{-4}	0.40 ± 0.01
Ratio (*PCA*/*Spatial*)	0.98	0.99	0.82	0.98	0.79	0.82	1.10	0.98

Note. Performance metrics for detrended light curves, as defined in Section 3.1 and used in Table 1. As described there, *recovery* indicates the degree to which injected signals are preserved correctly in the detrended light curves. The uncertainties listed are the standard deviation of the value over the ensemble of simulated light curves. No uncertainty is listed for $G_{y'}$ as it is computed over the ensemble of light curves. The goodness metric value computed over the nondetrended simulated light curves G_y is 0.23. The goodness metric value computed between all simulated astrophysical signals G_a is 2.2×10^{-2}.

Table 3
Experiment C Results

	Residual Systematics		
	$G_{y'}$	CDPP_{6h}	
		μ	σ
Spatial	1.13×10^{-2}	38.37	60.17
CBV	8.95×10^{-4}	33.16	97.60
SFF	1.45×10^{-3}	30.78	24.16
PLD	3.92×10^{-1}	36.49	41.44

Note. Residual systematics metrics for detrended light curves produced in Experiment C. The CDPP_{6h} mean over the ensemble of simulated light curves is denoted μ with sample standard deviation σ. No uncertainty is listed for $G_{y'}$ as it is computed over the ensemble of light curves.

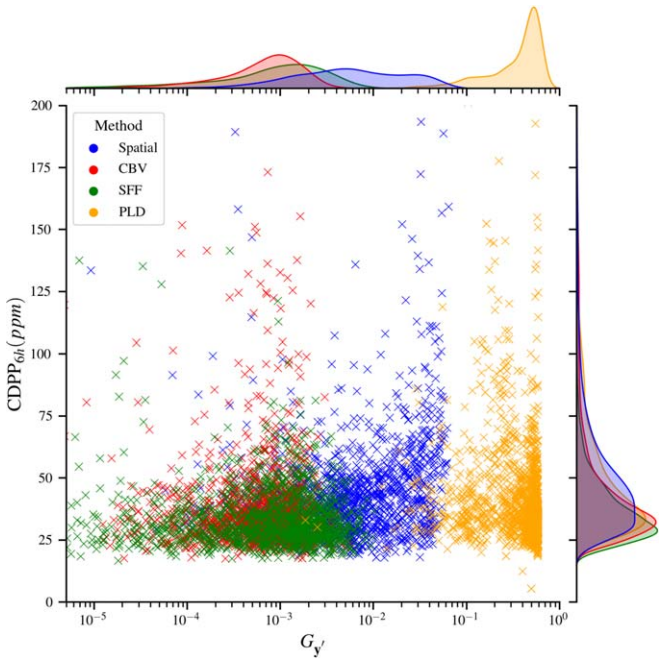


Figure 16. CDPP_{6h} (ppm) vs. goodness metric $G_{y'}$ per target light curve shown for the spatial systematics method and comparison standard detrending methods CBV, SFF, and PLD. Empirical distribution functions for each light-curve type are plotted over $G_{y'}$ (top) and CDPP_{6h} (right). The CDPP_{6h} metric is broadly comparable across methods while the $G_{y'}$ metric is more strongly dependent on the detrending method.

In Experiment B, the goodness metric value $G_{y'}$ was lower for the spatial method than the reference PCA method. For both experiments (Tables 1 and 2) the two algorithms obtained very similar CDPP_{6h} metrics and a reduction from the nondetrended preprocessed light curves, indicating that both methods were equivalently successful at suppressing transit timescale systematics.

In Experiment A the measures of residual systematics $G_{y'}$ and CDPP_{6h} are higher for the spatial systematics method compared to the PCA method. As noted above, these values for the detrended light curves may be elevated primarily due to improved preservation of true astrophysical variability or alternatively due to unmodeled residual systematics. In Figure 12 the example detrended light curve from Experiment A obtained by the spatial systematics algorithm in the first column shows an example of a large residual systematic scatter not present in the PCA detrended light curve; however, the overall variability is dominated in this case by the injected astrophysical signal. We argue, that although the spatial method may retain a greater level of residual systematics, comparably the major component of retained light-curve variability is astrophysical. As discussed above, Experiment B demonstrated a more accurate estimation both of systematics $\hat{\boldsymbol{l}}$ and of astrophysical signals $\boldsymbol{a}$ in light curves containing astrophysical signals. Experiment B, although simulated with a simplified data model, therefore aids the interpretation of metric performance. We note also that the improved preservation of true astrophysical variability in the light curves detrended by the spatial systematics algorithm may reduce the correlations $\bar{\rho}(\boldsymbol{a}_{[m]}, \boldsymbol{y}_{[m]}{}')$, $m \in \{s, f, t\}$ in Tables 1 and 2 discussed above and that these may be underestimated as a result.

In Experiment A, both the spatial systematics algorithm and the reference PCA method produce values $|\bar{\rho}(\boldsymbol{a}_{[m]}, \boldsymbol{y}'_{\boldsymbol{a}^c})| / |\bar{\rho}(\boldsymbol{a}_{[m]}, \boldsymbol{y}_{\boldsymbol{a}^c})| < 1$, for $m \in \{s, t, f\}$. Therefore, as discussed in Section 3.1, this suggests neither algorithm has increased the average incidental correlation between the simulated astrophysical signals and the light curves in which no astrophysical signals were injected (the noninjected subsample). Equivalently, neither method has introduced spurious astrophysical content into this noninjected subsample of light curves. However, the ratio $|\bar{\rho}(\boldsymbol{a}_{[m]}, \boldsymbol{y}'_{\boldsymbol{a}^c})| / |\bar{\rho}(\boldsymbol{a}_{[m]}, \boldsymbol{y}_{\boldsymbol{a}^c})|$ is higher for the spatial systematics algorithm compared to the PCA method (Tables 1 and 2), particularly for sine and flare $m \in \{s, f\}$ astrophysical signals, which have a longer duration. In Figure 21 we plot the light curves in the noninjected subsample for which the spatial systematics algorithm produces

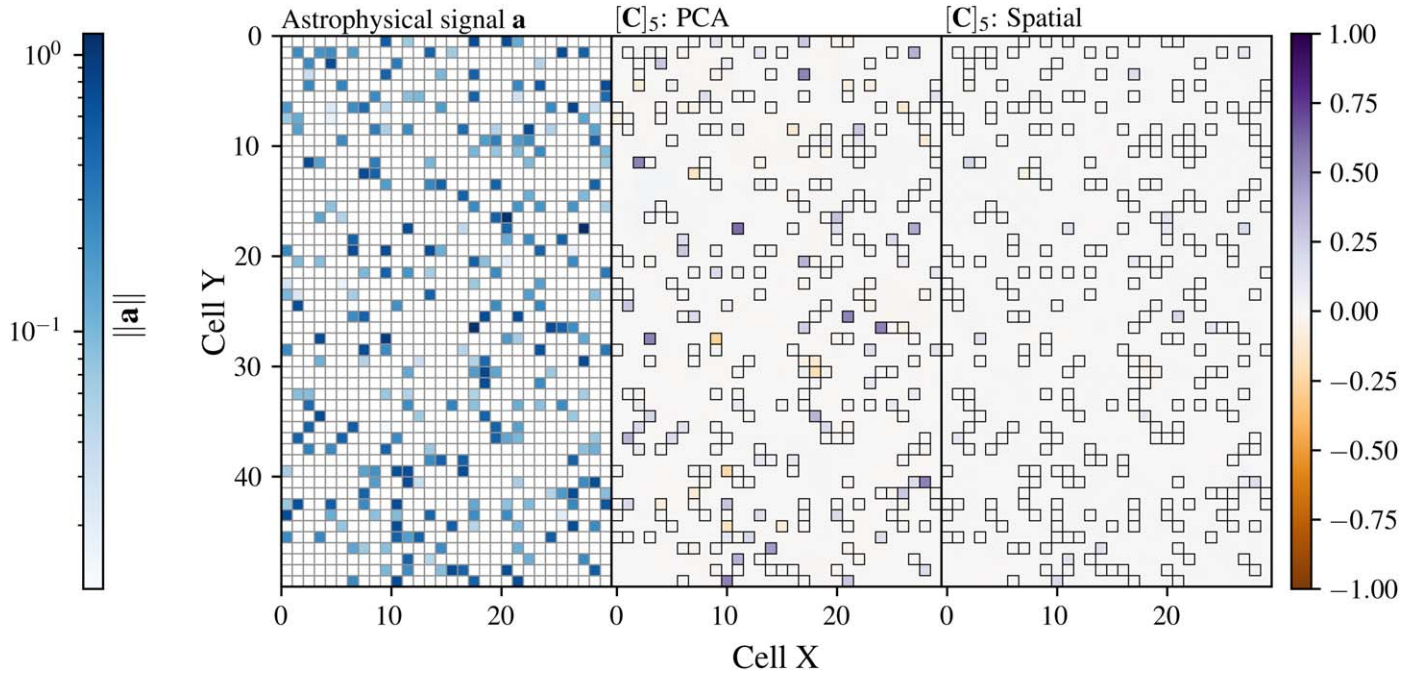


Figure 17. Gridded light-curve cells $i \rightarrow (x, y) \in (X, Y)$ where a simulated astrophysical signal $\boldsymbol{a}_{[m]}$ was injected for Experiment B. The x- and y-axes are in units of gridded spatial cells. The column at left depicts the magnitude of the injected astrophysical signal $\|\boldsymbol{a}_{[m]}\|$ as the color intensity of each cell on a log scale (blue color wedge at left). The middle and right columns depict coefficients c_i^5 for light curve $i \rightarrow (x, y) \in (X, Y)$ at each spatial cell (from Figure 14) for the reference PCA method and the spatial systematics method, respectively. For these columns, the magnitude of the coefficient is shown as color intensity (color wedge at right). A black outline around a cell indicates that the corresponding light curve contained an injected astrophysical signal. The rank 5 coefficient was chosen for clarity of interpretation as this coefficient should be closer to zero. Coefficients fitted with PCA exhibit some overfitting of injected astrophysical signals.

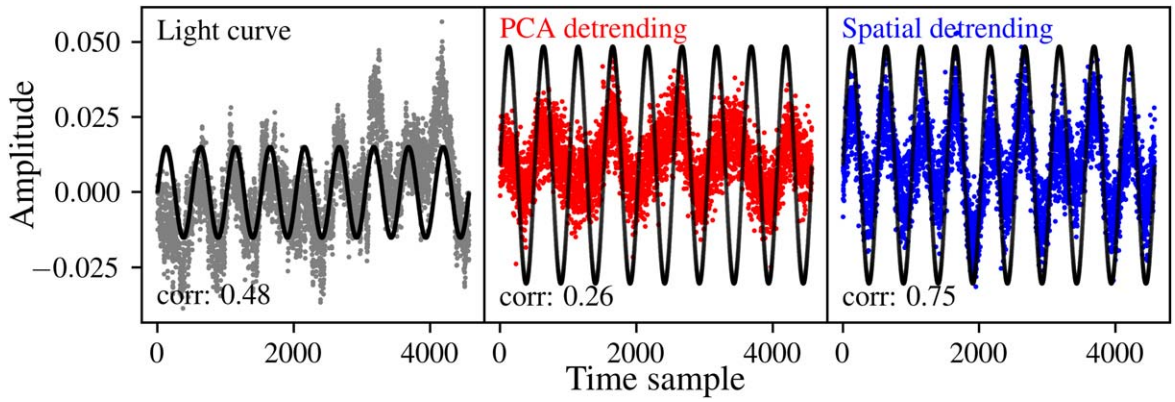


Figure 18. Simulated light curve $\boldsymbol{y}$ before detrending (left), the detrended light curve using PCA $\boldsymbol{y}'_{\mathrm{PCA}}$ (middle), and the detrended light curve using the spatial systematics algorithm $\boldsymbol{y}'$ (right), all at the position of the maximum value of $[\boldsymbol{C}_{\mathrm{PCA}}]_5$ in Figure 17. Simulated injected signals are overlaid in black. The x-axis is Kepler's long-cadence time sample index. This shows an instance where PCA detrending overfits the injected astrophysical sine signal relative to the spatial systematics algorithm.

the highest incidental correlation $\max_{\boldsymbol{y}_{a^c}} |\rho(\boldsymbol{a}_{[m]}, \boldsymbol{y}_{a^c})|$, $m \in \{s, t, f\}$ between the noninjected spatial detrended light curve and any injected astrophysical signal used in the simulation run. This figure shows the worst cases of incidental correlation; in general, the incidental correlation average is low. As described above, the spatial systematics algorithm is more successful at preserving astrophysical variability in detrended light curves and will therefore retain rare incidentally correlated signals, particularly if their timescale matches the simulated astrophysical signal. We believe this explains the higher values of this ratio in the case of sine and flare signal types.

Section 3.4 describes the results of Experiment C, a high-level comparison between the spatial systematics algorithm and the standard detrending methods CBV, SFF, and PLD (Lightkurve Collaboration et al. 2018). As noted above, the default recommended parameters were used for each method as opposed to a customized optimization. In addition, individual methods apply different default preprocessing steps. Accordingly, this high-level comparison cannot address the optimality of different detrending methods.

In this experiment, the CDPP_{6h} residual systematics metric, which is computed per light curve, shows broadly comparable ($\leqslant$8 ppm between mean values) detrending performance between the spatial systematics algorithm and the reference detrending methods CBV, SFF, and PLD (Table 3; Figure 16). The spatial systematics method has a slightly increased mean

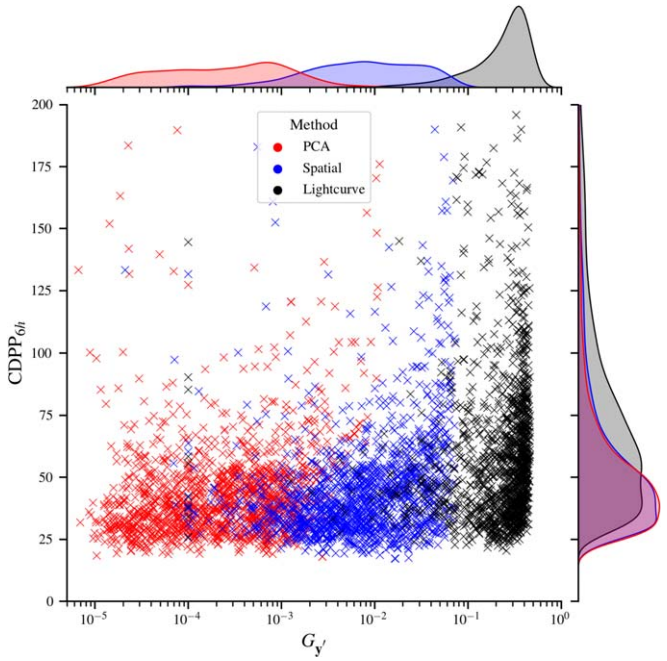


Figure 19. CDPP$_{6h}$ vs. goodness metric $G_{y'}$ per target light curve for a single run #0 of Experiment A. Three light-curve types are shown, including nondetrended preprocessed light curves (black), light curves detrended using PCA (red), and light curves detrended using the spatial systematics algorithm (blue). Empirical distribution functions for each light-curve type are plotted over $G_{y'}$ (top) and CDPP$_{6h}$ (right).

CDPP$_{6h}$ compared to other methods possibly due to instances of residual scatter where the spatial model imperfectly corrects strong temporal systematic variation. The $G_{y'}$ residual systematics metric differs in distribution across the detrending method (Figure 16). By definition (Section 3.1) and as discussed earlier, this metric is nonlinear and pairwise; accordingly, light curves with unmodeled systematics are more heavily weighted. The PLD method has the highest mean value of $G_{y'}$ (Figure 16) but this method retains long-term variability, which is fit by a spline term in the data model, and therefore is expected to have an inherently higher $G_{y'}$ distribution. In contrast, in preprocessing, SFF light curves are flattened using a Savitsky–Golay filter (Lightkurve Collaboration et al. 2018), and SFF is therefore broadly comparable to the CBV method in this metric. The spatial systematics algorithm has a higher mean value of $G_{y'}$ and broader distribution than the CBV and SFF methods (Figure 16). As discussed earlier for Experiments A and B, this may be due to the spatial systematics method retaining a higher level of astrophysical variability or due to the nonlinear contribution to $G_{y'}$ of light curves that do not fit the spatial systematics data model assumptions. Generalizations to make the constraint terms more complete are discussed below as potential future work.

Our current work has several limitations that would benefit from exploration in future work. As described in Section 2.5.2, we chose tunable algorithm parameters, including model rank, convergence criteria, optimization step size, weighting matrix, and initial coefficient values, among other parameters, based on feasibility and empirical evaluation alone. It would be beneficial to sample this algorithm parameter space over a broader range, for completeness. In conjunction, our numerical evaluations relied on an associated set of injected astrophysical signal types (Section 2.5.3), which could be expanded. Generalizing the sensor spatial discretization to nonuniform layouts is left for future work. It has been noted that light curves at small spatial separations may suffer from the blending of astrophysical signals (Kovács et al. 2005; Hattori et al. 2022) requiring future consideration when using the spatial systematics algorithm on densely populated fields. Although some light curves in our data sample have small spatial separation, this is not a major concern in the current work, as the cell separation on the spatial grid used significantly exceeds the PSF width. The sparse population of the field is likely due to the narrow (and bright) magnitude range selected. There is scope to generalize the constraint term to incorporate nonneighboring pixel spatial correlations or weightings of the spatial prior to excluding targets within a pixel separation where astrophysical blending could occur. In addition, the constraint terms may be generalized to include the known magnitude correlation (Smith et al. 2012; Stumpe et al. 2012) and other parametric dependencies (Moreno et al. 2021). As noted earlier, the Kepler test data were selected in a narrow magnitude range to isolate the effect of magnitude dependence in the current work.

5. Conclusions

In this paper, we present an exploratory algorithm for detrending light curves obtained in wide-field exoplanet transit surveys. This spatial systematics algorithm fits a low-rank linear systematics model to a collection of light curves while also including a total-variation spatial constraint across the sensor at a foundational level. The resulting objective function is reduced using variable elimination (Golub & Pereyra 1973), which also stabilizes the solution (Golub & Pereyra 2003; Shearer & Gilbert 2013). An approximate closed-form gradient was developed for minimization by gradient descent; this particular formulation is a modification of total-variation optimization approaches (Vogel 2002). The spatial systematics algorithm was numerically evaluated relative to a reference PCA method using both injection tests with Kepler data (Experiment A) and full simulations including simulated systematics, astrophysical signals, and statistical noise within the same Kepler coordinate framework (Experiment B).

The principal conclusions of the paper are:

1. The spatial systematics algorithm showed reduced overfitting of intrinsic astrophysical variability relative to the PCA method in the two numerical evaluation studies. The reduced overfitting was demonstrated by a comparable or significantly higher correlation between the known astrophysical signals and the detrended light curves in both experiments. In Experiment B the known simulated systematics were comparably or more accurately estimated relative to the PCA method, particularly for light curves containing injected signals. We argue that the reduced overfitting likely arises from the physically realistic total-variation constraint. In addition, the algorithm simultaneously varies both the basis vectors and their coefficient weights. Both factors likely contribute to the improved separation of systematics and astrophysical signals.
2. A marked reduction in overfitting relative to PCA was found for slowly varying astrophysical signals while comparable performance was achieved for shorter exoplanet transit signals. We argue that the longer-

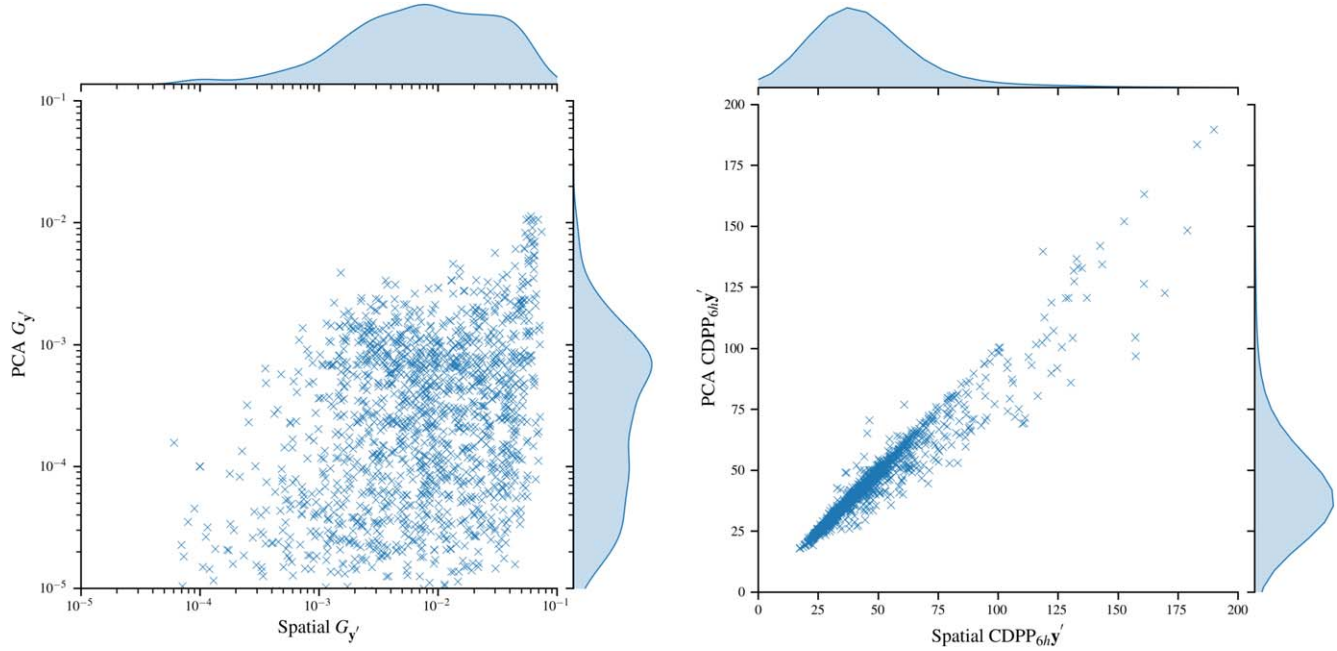


Figure 20. Metrics $G_{y'}$ (left) and CDPP$_{6h}$ (right) for the reference PCA method vs. the spatial systematics method. Both plots are per target light curve for a single run #0 of Experiment A. Empirical distribution functions for each light-curve metric are plotted over the spatial systematics method (horizontal axis) and the PCA method (vertical axis). Although correlated, there is significant scatter in the left plot due to the intrinsic nonlinear and pairwise nature of the metric $G_{y'}$. It is accordingly more sensitive to light curves with poorly modeled systematics or greater retained astrophysical variability.

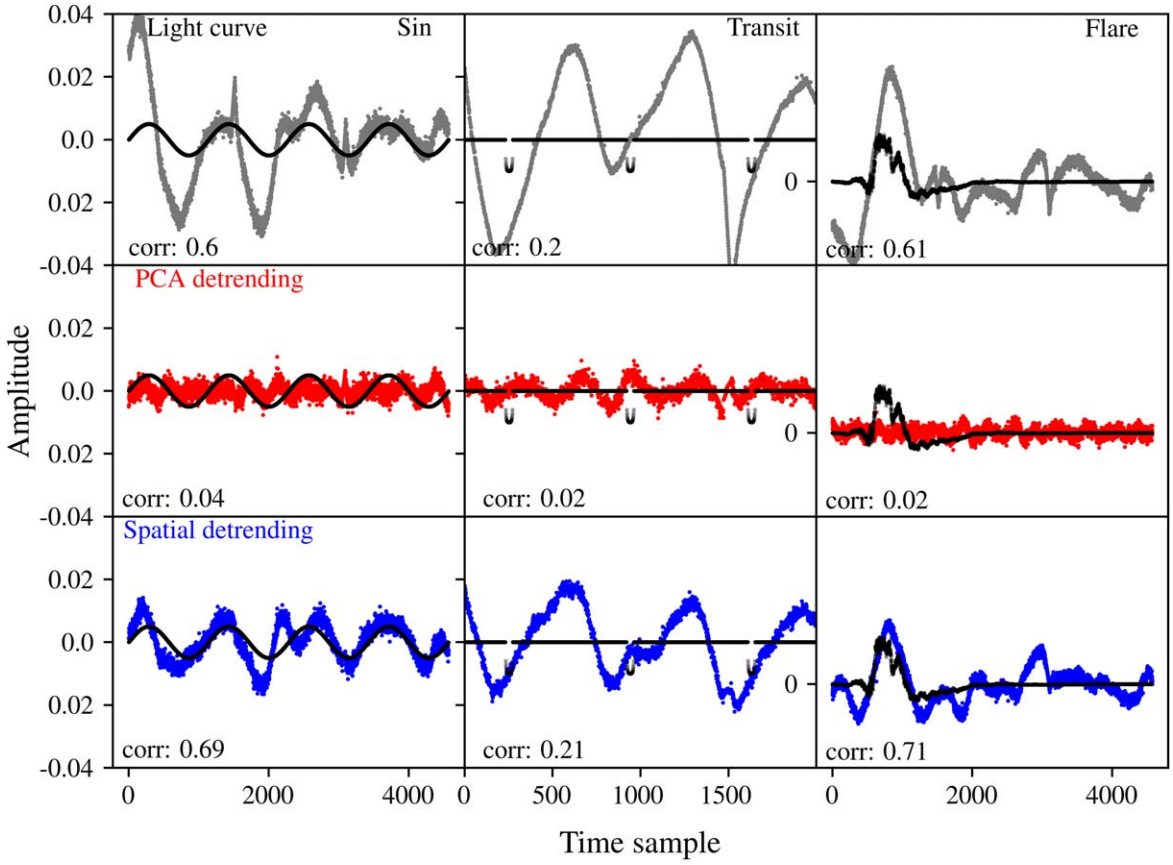


Figure 21. Noninjected light curves $\boldsymbol{y}_{a^c}$ for which the spatial systematics algorithm produces the highest incidental correlation $\max_{\boldsymbol{y}_{a^c}} \rho(\boldsymbol{a}_{[m]}, \boldsymbol{y}_{a^c})$, $m \in \{s, t, f\}$ between the nondetrended light curve $\boldsymbol{y}_{a^c}$ and any injected astrophysical signal $\boldsymbol{a}_{[m]}$ used in the simulation run. The top row shows the nondetrended light curve and the astrophysical signal (injected into another light curve) with which the incidental correlation is highest. The detrended light curves $\boldsymbol{y}'_{a^c}$ are shown as obtained by the reference PCA method (middle row) and the spatial systematics algorithm (bottom row). Simulated injected signals are overlaid in black. In these worst cases, it can be seen that the high level of correlation of the spatial detrended light curve $\boldsymbol{y}'$ with an astrophysical signal $\boldsymbol{a}$ is incidental, as the original light curve $\boldsymbol{y}$ itself is incidentally correlated with $\boldsymbol{a}$. This suggests that high values for the metric of corruption $|\bar{\rho}(\boldsymbol{a}, \boldsymbol{y}'_{a^c})|/|\bar{\rho}(\boldsymbol{a}, \boldsymbol{y}_{a^c})|$ may result incidentally for the spatial systematics algorithm because it retains a higher level of astrophysical variability.

duration astrophysical signals overlap more significantly with the timescale of residual systematics in the Kepler data and therefore allow for the algorithm to achieve a clearer separation of systematics for these signals.

3. In terms of estimates of residual systematics in the detrended light curves, both methods achieved comparable CDPP_{6h} values in Experiment A suggesting largely equivalent success in suppressing transit timescale systematic noise. However, the goodness metric $G_{y'}$ was comparatively larger in this experiment for the spatial systematics algorithm relative to PCA indicative of residual systematics.
4. Neither method increased the incidental correlation between the simulated astrophysical signals and the noninjected subsample of light curves in the numerical evaluations. Equivalently, neither method added spurious astrophysical variability into the detrended light curves.

Acknowledgments

This paper includes data collected by the Kepler mission. Funding for the Kepler mission is provided by the NASA Science Mission Directorate. STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. All of the data presented in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST) at the Space Telescope Science Institute. The specific Kepler observations analyzed include Q6 (STScI 2016a), Q10 (STScI 2016b), and Q14 (STScI 2016c). We thank the anonymous referee of the paper for their suggested revisions; these significantly improved the paper.

Software: This work used the following external packages: Astropy (Astropy Collaboration et al. 2013, 2018, 2022), Scikit-learn (Pedregosa et al. 2011), Matplotlib (Hunter 2007), Seaborn (Waskom 2021), NumPy (Harris et al. 2020), SciPy (Virtanen et al. 2020), and Lightkurve (Lightkurve Collaboration et al. 2018). The algorithm described in this paper is available as a public Python package on GitHub[11] or via PyPI under the package name *spatial-detrend*.[12] A copy of this algorithm concurrent with this study has been deposited to figshare (Taaki 2023).

Appendix A
List of Symbols

A list of important symbols and mathematical notation used in the paper is provided in Table 4.

Table 4
List of Symbols

$[\boldsymbol{M}]_{i,j}$ or $M_{i,j}$	Element at ith row and jth column of matrix $\boldsymbol{M}$
1_N	Identity matrix of size $N \times N$
$\boldsymbol{M}^{\dagger}$	Pseudoinverse matrix operator
$\boldsymbol{M}^{T}$	Matrix transpose operator
$\|\boldsymbol{x}\|_p$	L_p norm of vector $\boldsymbol{x}$
$\|\boldsymbol{M}\|_{p,q}$	Combined $L_{p,q}$ norm for matrix $\boldsymbol{M}$
$\|\boldsymbol{M}\|_F$	Frobenius norm of matrix $\boldsymbol{M}$
$i \in I$	Target light-curve index set
$i \rightarrow (x, y) \in X, Y; X, Y \in \mathbb{Z}^+$	Gridded target position on sensor
$\boldsymbol{y}_i$	Vector light curve for target index i
$\boldsymbol{n}_i$	Vector statistical noise term for light-curve target index i
$\boldsymbol{l}_i$	Vector systematics term for light-curve target index i
$\boldsymbol{Y}$	Matrix of light curves; for vectors $\boldsymbol{y}$
$\boldsymbol{Y}'$	Matrix of detrended light curves; for vectors $\boldsymbol{y}'$
$\boldsymbol{N}$	Matrix of statistical noise
$\boldsymbol{a}_{[m]}$: $m \in \{s, t, f\}$	Simulated astrophysical signal of type: s sine, t transit, f flare
$\boldsymbol{A}$	Matrix of simulated astrophysical signals
K	Rank of systematic noise model
$\boldsymbol{v}_k$	Basis vector $k \in K$ for systematic noise model
$\boldsymbol{L} = \boldsymbol{VC}$	Matrix of low-rank systematic noise ($\boldsymbol{V}$ basis vectors, $\boldsymbol{C}$ coefficient matrix)
$\bar{\boldsymbol{C}}$	Column-normalized coefficient matrix
c_i^k	Coefficient weighting of $\boldsymbol{v}_k$ for light curve $i \in I$
$\boldsymbol{c}_i = [c_i^1, ..., c_i^K]^T$	Coefficient vector for light curve i
$\boldsymbol{W}$	Weight matrix
$\boldsymbol{D}$	Difference operator ($\boldsymbol{D}_{\boldsymbol{W}}$ difference operator with weights $\boldsymbol{W}$)
$\boldsymbol{\alpha}^t$	Step size at iteration t
$f(\boldsymbol{V}, \boldsymbol{C})$	Least-square residual between $\boldsymbol{Y}$ and $\boldsymbol{VC}$
$h(\boldsymbol{C})$	For fixed $\boldsymbol{C}$, minimizing value $\boldsymbol{V}$ of $f(\boldsymbol{V}, \boldsymbol{C})$
$g(\boldsymbol{C})$	Total-variation penalty constraint
$w(.)$	Variable-reduced objective function
$\nabla f'(.)$	Gradient of $f(h(\boldsymbol{C}), \boldsymbol{C})$
$\nabla g(.)$	Gradient of $f(\boldsymbol{C})$
$\boldsymbol{P}_{R(\boldsymbol{C})^T}$	Projection onto the range space of $\boldsymbol{C}^T$
$\rho(\boldsymbol{y}_i, \boldsymbol{y}_j)$	Correlation between vectors $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$
$\bar{\rho}(\boldsymbol{a}, \boldsymbol{y}_{\boldsymbol{a}}')$	Averaged correlation of astrophysical signals $\boldsymbol{a}$ and detrended light curves with an injected signal $\boldsymbol{y}_{\boldsymbol{a}}'$
$\{\boldsymbol{y}_{\boldsymbol{a}^c}\}$	Set of light curves excluding those with a simulated astrophysical signal
$G_{y'}$	Residual detrended systematics: goodness metric
CDPP_{6h}	Residual detrended systematics: 6 hr combined differential photometric precision

[11] Repository: https://github.com/xiaziyna/spatial-detrend.
[12] PyPI: https://pypi.org/project/spatial-detrend.

Appendix B
Principal Component Analysis

Singular value decomposition (SVD) and principal component analysis (PCA) are summarized here, as foundations of the cotrending method. A complete reference for SVD and PCA is provided by Hastie et al. (2009, chap. 14.5) and Golub & Van Loan (2013). Every matrix $\boldsymbol{Y} \in \mathbb{R}^{n\times m}$ admits an SVD: $\boldsymbol{Y} = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^T$ with unitary matrices $\boldsymbol{V} \in \mathbb{R}^{n\times n}$: $\boldsymbol{V}^T\boldsymbol{V} = \mathbb{1}_n$ and $\boldsymbol{U} \in \mathbb{R}^{m\times m}$: $\boldsymbol{U}^T\boldsymbol{U} = \mathbb{1}_m$ and a diagonal matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times m}$ of ordered nonnegative values. The columns of $\boldsymbol{V}$ and $\boldsymbol{U}$ are the left and right singular vectors, respectively, the values $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \ldots \sigma_{\min(m,n)})$ are the singular values, and 1 denotes an identity matrix. The number of nonzero singular values determines the rank r of $\boldsymbol{Y}$. The SVD may be equivalently defined using only nonzero singular values, in which case $\boldsymbol{V} \in \mathbb{R}^{n\times r}$: $\boldsymbol{V}^T\boldsymbol{V} = \mathbb{1}_r$, $\boldsymbol{U} \in \mathbb{R}^{m\times r}$: $\boldsymbol{U}^T\boldsymbol{U} = \mathbb{1}_r$ and $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \ldots \sigma_r)$. This is termed compact SVD and is the form used in this work. There are a number of efficient algorithms to compute an SVD (Golub & Van Loan 2013) making it extremely practical for analysis.

Consider the data model $\boldsymbol{Y} = \boldsymbol{L} + \boldsymbol{N}$, defined in Section 2, where $\boldsymbol{Y} \in \mathbb{R}^{N\times I}$, $\boldsymbol{L} \in \mathbb{R}^{N\times I}$, and $\boldsymbol{N} \in \mathbb{R}^{N\times I}$ represent the light curves, systematic noise, and statistical noise respectively. The Eckart–Young–Mirsky theorem (Eckart & Young 1936) states that the optimal rank-K matrix approximation of $\boldsymbol{Y}$, which minimizes the least-squares residual $||\boldsymbol{Y} - \boldsymbol{L}||_F^2$ (subscript F denoting the Frobenius norm Golub & Van Loan 2013) is obtained by retaining only the top K singular values and associated singular vectors, of the SVD of $\boldsymbol{Y}$; this is known as rank thresholding:

$$\underset{\boldsymbol{L}:\mathrm{rank}(\boldsymbol{L})\leqslant K}{\arg\min}\ ||\boldsymbol{Y} - \boldsymbol{L}||_F^2 = \boldsymbol{V}_K\boldsymbol{\Sigma}_K\boldsymbol{U}_K^T, \tag{B1}$$

with $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_K)$, $\boldsymbol{V}_K$ and $\boldsymbol{U}_K$ denote the associated leading K left and right singular vectors, respectively. PCA is also the solution to the objective function minimization in Equation (B1) with form $\boldsymbol{L} = \boldsymbol{V}_K\boldsymbol{C}_K$, $\boldsymbol{V}_K \in \mathbb{R}^{n\times K}$, $\boldsymbol{C}_K \in \mathbb{R}^{K\times m}$, and $\boldsymbol{V}_K^T\boldsymbol{V}_K = \mathbb{1}_K$. The PCA solution can be obtained from the compact SVD solution, using $\boldsymbol{V}_K$ and setting $\boldsymbol{C}_K = \boldsymbol{\Sigma}_K\boldsymbol{U}_K^T$. Any $\boldsymbol{V}_K{}' = \boldsymbol{V}_K\bar{\boldsymbol{U}}^T$ and $\boldsymbol{C}_K{}' = \bar{\boldsymbol{U}}\boldsymbol{C}_K$ is also an equivalent low-rank solution $\boldsymbol{L} = \boldsymbol{V}_K{}'\boldsymbol{C}_K{}' = \boldsymbol{V}_K\boldsymbol{C}_K$, where $\bar{\boldsymbol{U}} \in \mathbb{R}^{K\times K}$ is any unitary matrix such that $\bar{\boldsymbol{U}}^T\bar{\boldsymbol{U}} = \bar{\boldsymbol{U}}\bar{\boldsymbol{U}}^T = \bar{\boldsymbol{U}}\bar{\boldsymbol{U}}^{-1} = \mathbb{1}_K$.

Appendix C
Total-variation Measure

An overview of total variation is provided in the monograph by Vogel 2002 and in Karl (2005, Chapter 3.6). In the continuous limit, a Lebesque-integrable (Edwards 1994) function $f(x, y) \in L^1$ with vector gradient $\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$ has total variation:

$$TV(f) = \int\int \|\nabla f\|_2\ dxdy, \tag{C1}$$

where $\|.\|_2$ denotes an L_2 vector norm. Functions of the bounded total variation have $TV(f) \leqslant \infty$. The total-variation measure $TV(f)$ can be used as a regularization or penalty constraint on f to enforce spatial uniformity while preserving limited function discontinuities (Vogel 2002).

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

$$TV_d(f) = \sum_{i\in X}\sum_{j\in Y} ||(D_x f)_{i,j}, (D_y f)_{i,j}||_2, \tag{C2}$$

where $(D_x f)_{i,j} = f_{i,j} - f_{i-1,j}$ and $(D_y f)_{i,j} = f_{i,j} - f_{i,j-1}$ are unweighted function differences. If matrix $\boldsymbol{Df}$ is constructed with $\mathrm{vec}(D_x f)^T$ and $\mathrm{vec}(D_y f)^T$ as the first and second rows, respectively (where vec(.) denotes vectorization (Golub & Van Loan 2013), then:

$$TV_d(f) = ||\boldsymbol{Df}||_{2,1}^1. \tag{C3}$$

This may be generalized to: $||\boldsymbol{Df}||_{2,p}^p$: $p \in [1, 2]$, where p controls the degree of spatial uniformity.

C.1. Total-variation Measure for p = 2

We show here that minimizing the generalized total-variation measure $TV_d(f) = ||\boldsymbol{Df}||_{2,p}^p$ for the case $p = 2$ is equivalent to maximizing the correlation ρ_{ij} between neighboring normalized light-curve coefficients $\bar{\boldsymbol{c}}_i$ and $\bar{\boldsymbol{c}}_j$ ($i \in I, j \in I, i, j \in n(I)$), where $n(I)$ is the set of neighboring light-curve tuples. The maximum neighboring coefficient correlation ρ_n takes the form:

$$\rho_n = \underset{\boldsymbol{V},\boldsymbol{C}:\ \mathrm{rank}(\boldsymbol{VC})\leqslant K}{\arg\max} \sum_{(i,j)\in n(I)} \bar{\boldsymbol{c}}_i^T\bar{\boldsymbol{c}}_j. \tag{C4}$$

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

$$\rho_n = \underset{\boldsymbol{V},\boldsymbol{C}:\ \mathrm{rank}(\boldsymbol{VC})\leqslant K}{\arg\min} \sum_{(i,j)\in n(I)} ||\bar{\boldsymbol{c}}_i - \bar{\boldsymbol{c}}_j||_2^2. \tag{C5}$$

The summation can be replaced by a linear difference operator $\boldsymbol{D}$ (here unweighted for simplicity) acting on columns of $\bar{\boldsymbol{C}}$. This difference operator was introduced above in the definition of the generalized total-variation measure. Therefore:

$$\begin{aligned}\rho_n &= \underset{\boldsymbol{V},\boldsymbol{C}:\ \mathrm{rank}(\boldsymbol{VC})\leqslant K}{\arg\min} ||\boldsymbol{D}\bar{\boldsymbol{C}}||_F^2 \\ &= \underset{\boldsymbol{V},\boldsymbol{C}:\ \mathrm{rank}(\boldsymbol{VC})\leqslant K}{\arg\min} TV_d(\bar{\boldsymbol{C}}).\end{aligned} \tag{C6}$$

We note that even if the minimum is not achieved either objective should produce an equivalent result.

Appendix D
Probabilistic View of the Objective

Although the spatial systematics algorithm was not derived within a Bayesian framework, a probabilistic interpretation of the objective function can be shown. An overview of Bayesian estimation and low-rank models is provided in the monograph by Murphy (2020). We adopt the nomenclature of Section 2.3 here. In Bayesian MAP estimation, the posterior probability is maximized:

$$\underset{\boldsymbol{C}}{\arg\max}\, p(\boldsymbol{C}|\boldsymbol{Y}) \propto \underset{\boldsymbol{C}}{\arg\max}\, p(\boldsymbol{Y}|\boldsymbol{C})p(\boldsymbol{C}). \tag{D1}$$

The signal model in Equation (1) is $\boldsymbol{Y} = \boldsymbol{VC} + \boldsymbol{N}$, where $\boldsymbol{N}$ is white Gaussian noise, and therefore $p(\boldsymbol{Y}|\boldsymbol{C}, \boldsymbol{V}) \sim \mathcal{N}(\boldsymbol{VC}, \mathbb{1})$. The conditional basis vectors are $\boldsymbol{V}_{|\boldsymbol{C}} = \underset{\boldsymbol{V}}{\arg\min}\, p(\boldsymbol{Y}, \boldsymbol{V}|\boldsymbol{C})$ and $p(\boldsymbol{Y}|\boldsymbol{C}) \sim \mathcal{N}(\boldsymbol{V}_{|\boldsymbol{C}}\boldsymbol{C}, \mathbb{1})$. Therefore $\ln(p(\boldsymbol{Y}|\boldsymbol{C})) = ||\boldsymbol{Y} - \boldsymbol{V}_{|\boldsymbol{C}}\boldsymbol{C}||_F^2$.

The total-variation prior implies that $\ln(p(\boldsymbol{C})) \propto \|\boldsymbol{D}_W \bar{\boldsymbol{C}}\|_{2,p}^p$, whereby $p(\boldsymbol{C}) \propto \exp \|\boldsymbol{D}_W \bar{\boldsymbol{C}}\|_{2,p}^p$. For $p = 2$ this is equivalent to a Gaussian prior on the product $\boldsymbol{D}_W \bar{\boldsymbol{C}}$.

Appendix E
Relation between Coefficient and Systematics Correlations

We consider here the relation between the correlation $\bar{\boldsymbol{c}}_i^T \bar{\boldsymbol{c}}_j$ of the normalized coefficients $\bar{\boldsymbol{c}}_i \in \mathbb{R}^K$ and the correlation $\bar{\boldsymbol{l}}_i^T \bar{\boldsymbol{l}}_j$ of the normalized systematics $\bar{\boldsymbol{l}}_i = \frac{\boldsymbol{V}\boldsymbol{c}_i}{\|\boldsymbol{V}\boldsymbol{c}_i\|_2^2}$: $\bar{\boldsymbol{l}}_i \in \mathbb{R}^N$.

The maximum deviation of $\bar{\boldsymbol{l}}_i^T \bar{\boldsymbol{l}}_j$ from $\bar{\boldsymbol{c}}_i^T \bar{\boldsymbol{c}}_j$ can be bounded and understood in terms of $\boldsymbol{V} \in \mathbb{R}^{N \times K}$. In the simplest case, if $\boldsymbol{V}$ is orthonormal then $\|\boldsymbol{V}\boldsymbol{c}\| = \|\boldsymbol{c}\|$ (Gentle 2010), and:

$$\begin{aligned} \|\bar{\boldsymbol{l}}_i - \bar{\boldsymbol{l}}_j\|_2^2 &= \left\| \frac{\boldsymbol{V}\boldsymbol{c}_i}{\|\boldsymbol{V}\boldsymbol{c}_i\|_2^2} - \frac{\boldsymbol{V}\boldsymbol{c}_j}{\|\boldsymbol{V}\boldsymbol{c}_j\|_2^2} \right\|_2^2 \\ &= \left\| \boldsymbol{V}\left(\frac{\boldsymbol{c}_i}{\|\boldsymbol{c}_i\|} - \frac{\boldsymbol{c}_j}{\|\boldsymbol{c}_j\|} \right) \right\|_2^2 \\ &= \|\bar{\boldsymbol{c}}_i - \bar{\boldsymbol{c}}_j\|_2^2 \end{aligned} \tag{E1}$$

$$\implies \bar{\boldsymbol{l}}_i^T \bar{\boldsymbol{l}}_j = \bar{\boldsymbol{c}}_i^T \bar{\boldsymbol{c}}_j. \tag{E2}$$

However, $\boldsymbol{V}$ is not restricted to be orthonormal, as discussed in Section 2.3. In the general case, the worst-case deviation between $\bar{\boldsymbol{l}}_i^T \bar{\boldsymbol{l}}_j$ and $\bar{\boldsymbol{c}}_i^T \bar{\boldsymbol{c}}_j$ can be understood in terms of the singular values of $\boldsymbol{V}$; we follow the approach of Huber (2016). Denoting the rank K singular value decomposition as $\boldsymbol{V} = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}^T$ (Appendix B), $\boldsymbol{\Sigma}$ is comprised of singular values of descending magnitude $\sigma_1 \geqslant \sigma_2 ... \geqslant \sigma_K \geqslant 0$ and where, without loss of generality, we adopt $\sigma_1 = 1$. Matrix $\boldsymbol{A}$ contains the left singular vectors $\{\boldsymbol{a}_k : k \in K\}$ and matrix $\boldsymbol{B}$ the right singular vectors $\{\boldsymbol{b}_k : k \in K\}$. The systematics correlation $\bar{\boldsymbol{l}}_i^T \bar{\boldsymbol{l}}_j$ can be expanded as:

$$\bar{\boldsymbol{l}}_i^T \bar{\boldsymbol{l}}_j = \frac{\boldsymbol{c}_i^T \boldsymbol{V}^T \boldsymbol{V}\boldsymbol{c}_j}{\|\boldsymbol{V}\boldsymbol{c}_i\| \|\boldsymbol{V}\boldsymbol{c}_j\|} = \frac{\boldsymbol{c}_i^T \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{A}^T \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}^T \boldsymbol{c}_j}{\|\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}^T \boldsymbol{c}_i\| \|\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}^T \boldsymbol{c}_j\|} \tag{E3}$$

$$= \frac{\boldsymbol{c}_i^T \boldsymbol{B}\boldsymbol{\Sigma}^2 \boldsymbol{B}^T \boldsymbol{c}_j}{\|\boldsymbol{\Sigma}\boldsymbol{B}^T \boldsymbol{c}_i\| \|\boldsymbol{\Sigma}\boldsymbol{B}^T \boldsymbol{c}_j\|}, \tag{E4}$$

as $\boldsymbol{A}$ is orthonormal $\boldsymbol{A}^T\boldsymbol{A} = 1_K$ and therefore $\|\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}^T\boldsymbol{c}\| = \|\boldsymbol{\Sigma}\boldsymbol{B}^T\boldsymbol{c}\|$. By the cosine similarity (Phillips 2021), $\bar{\boldsymbol{c}}_i^T \bar{\boldsymbol{c}}_j = \cos(\phi)$ where ϕ is the angle between the normalized coefficient vectors in K-dimensional space. The correlation $\bar{\boldsymbol{l}}_i^T \bar{\boldsymbol{l}}_j$ is similarly determined by the angle ψ between unit length vectors $\frac{\boldsymbol{\Sigma}\boldsymbol{B}^T\boldsymbol{c}_i}{\|\boldsymbol{\Sigma}\boldsymbol{B}^T\boldsymbol{c}_i\|}$ and $\frac{\boldsymbol{\Sigma}\boldsymbol{B}^T\boldsymbol{c}_j}{\|\boldsymbol{\Sigma}\boldsymbol{B}^T\boldsymbol{c}_j\|}$. The deviation of $\bar{\boldsymbol{l}}_i^T \bar{\boldsymbol{l}}_j$ from $\bar{\boldsymbol{c}}_i^T \bar{\boldsymbol{c}}_j$ is defined as $|\phi - \psi|$.

The transform $\boldsymbol{\Sigma}\boldsymbol{B}^T\boldsymbol{c}_i$ projects $\boldsymbol{c}_i$ along each $\boldsymbol{b}_k$ and scales by σ_k: $\boldsymbol{\Sigma}\boldsymbol{B}^T\boldsymbol{c}_i = [\sigma_1 \boldsymbol{b}_1^T \boldsymbol{c}_i, ..., \sigma_K \boldsymbol{b}_K^T \boldsymbol{c}_i]^T$. Any $\boldsymbol{c}_i \in \mathbb{R}^K$ must lie in the span of orthonormal vectors $\{\boldsymbol{b}_k : k \in K\}$ as $\boldsymbol{b}_k$ are a basis for $\mathbb{R}^K$. For simplicity, assume that $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ lie in the 2D span of two of the basis vectors $\boldsymbol{b}_m$ and $\boldsymbol{b}_n$ in the form $\boldsymbol{c}_i = \alpha_i \boldsymbol{b}_m + \beta_i \boldsymbol{b}_n$ and $\boldsymbol{c}_j = \alpha_j \boldsymbol{b}_m + \beta_j \boldsymbol{b}_n$. Then $\boldsymbol{\Sigma}\boldsymbol{B}^T\boldsymbol{c}_i = \sigma_m \alpha_i \boldsymbol{b}_m + \sigma_n \beta_i \boldsymbol{b}_n$ and $\boldsymbol{\Sigma}\boldsymbol{B}^T\boldsymbol{c}_j = \sigma_m \alpha_j \boldsymbol{b}_m + \sigma_n \beta_j \boldsymbol{b}_n$. Then $\phi = \arctan \beta_j/\alpha_j - \arctan \beta_i/\alpha_i$ and $\psi = \arctan \frac{\sigma_n \beta_j}{\sigma_m \alpha_j} - \arctan \frac{\sigma_n \beta_i}{\sigma_m \alpha_i}$. Therefore the maximum differential rotation $|\phi - \psi|$ occurs when the ratio between singular values $\frac{\sigma_m}{\sigma_n}$ is greatest, i.e., when $m = 1$ and $n = K$. In this case, the transformed vector component along $\boldsymbol{b}_1$ is unchanged (as $\sigma_1 = 1$) while the transformed component along orthogonal vector $\boldsymbol{b}_K$ is scaled by the smallest relative value σ_K; this maximizes rotation. The rotation of a single coefficient vector $\boldsymbol{c}_i$ under the linear transformation $\boldsymbol{\Sigma}\boldsymbol{B}^T$ is illustrated in Figure 22.

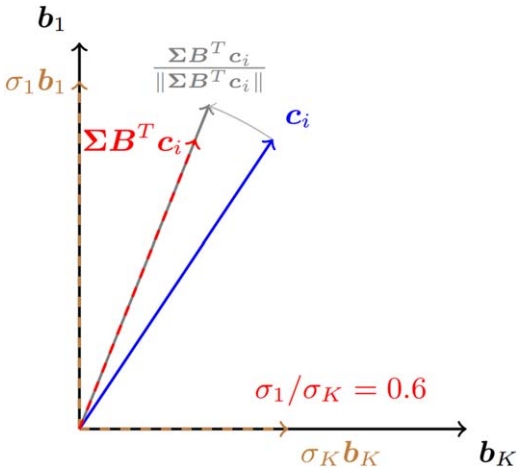


Figure 22. Example coefficient vector $\boldsymbol{c}_i = 3\boldsymbol{b}_1 + 2\boldsymbol{b}_K$ (blue), linearly transformed to vector $\boldsymbol{\Sigma}\boldsymbol{B}^T\boldsymbol{c}_i = 3\boldsymbol{b}_1 + \sigma_K 2\boldsymbol{b}_K$ (red), plotted on the 2D space spanned by $\boldsymbol{b}_1$ and $\boldsymbol{b}_K$ (brown). The normalized transformed vector is also shown (gray). A value $\sigma_K = 0.6$ was adopted for this example (recall $\sigma_1 = 1$).

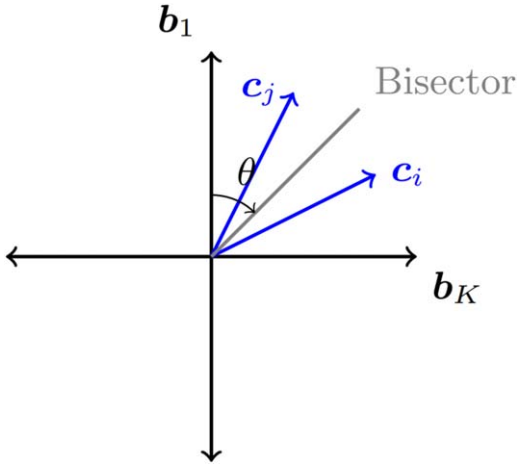


Figure 23. Bisector at angle θ between coefficient vectors $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ in the 2D space spanned by basis vectors $\boldsymbol{b}_1$ and $\boldsymbol{b}_K$.

The maximum differential rotation $|\phi - \psi|$ depends on the relative position of the coefficient vectors $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ before linear transformation. In this 2D case, we denote their bisector angle as θ, as depicted in Figure 23. By geometric inspection and by considering the algebra above for the 2D case, the maximal differential rotation $|\phi - \psi|$ occurs when $\theta = 0$ and the coefficient vectors are bisected by $\boldsymbol{b}_1$. The minimal differential rotation occurs when $\theta = \frac{\pi}{2}$ and the coefficient vectors are bisected by $\boldsymbol{b}_K$. For these values of θ and following the trigonometric approach by Huber (2016), the upper and lower limits for the correlation $\bar{\boldsymbol{l}}_i^T \bar{\boldsymbol{l}}_j$ of the normalized systematics

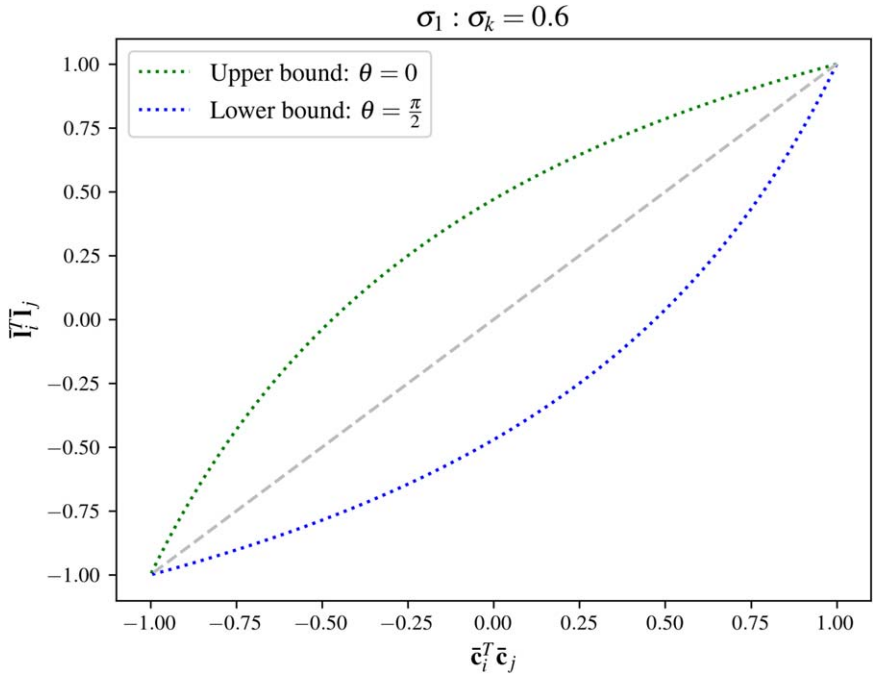


Figure 24. Deviation of the systematics correlation $\bar{\boldsymbol{l}}_i^T \bar{\boldsymbol{l}}_j$, upper bound (green) and lower bound (blue), from the coefficient correlation $\bar{\boldsymbol{c}}_i^T \bar{\boldsymbol{c}}_j$ (gray) for $\sigma_1 : \sigma_K = 0.6$.

vectors can be expressed in the form:

$$\frac{\sigma_K^2 - \tan^2(\phi)}{\tan^2(\phi) + \sigma_K^2} \leqslant \bar{\boldsymbol{l}}_i^T \bar{\boldsymbol{l}}_j \leqslant \frac{1 - \sigma_K^2 \tan^2(\phi)}{1 + \sigma_K^2 \tan^2(\phi)}. \tag{E5}$$

As the correlation $\bar{\boldsymbol{c}}_i^T \bar{\boldsymbol{c}}_j \to 1$ implies $\phi \to 0$, the lower and upper bounds smoothly converge to unity and $\bar{\boldsymbol{l}}_i^T \bar{\boldsymbol{l}}_j \to 1$ in this limit. This implies that the deviation between $\bar{\boldsymbol{l}}_i^T \bar{\boldsymbol{l}}_j$ and $\bar{\boldsymbol{c}}_i^T \bar{\boldsymbol{c}}_j$ is lower when the absolute value of the coefficient correlation is high. Furthermore, the worst-case deviation depends on the ratio of the smallest and largest singular values of $\boldsymbol{V}$; when $\sigma_K = \sigma_1$, $\boldsymbol{V}$ is orthonormal and there is no deviation, also shown above. In Figure 24 the deviation bounds are shown as a function of the coefficient correlation for a hypothetical value $\sigma_K = 0.6$ ($\sigma_1 = 1$).

Appendix F
Spatial Systematics Optimization Method

In this section we derive gradient-descent steps for minimizing the objective in Equation (7), repeated here for convenience:

$$\underset{\boldsymbol{V},\boldsymbol{C}:\ \mathrm{rank}(\boldsymbol{VC}) \leqslant K}{\arg\min} \left|\left| \boldsymbol{Y} - \boldsymbol{VC} \right|\right|_F^2 + \left|\left| \boldsymbol{D}\bar{\boldsymbol{C}} \right|\right|_{2,p}^p. \tag{F1}$$

As above, the least-squares penalty is denoted as $f(\boldsymbol{V}, \boldsymbol{C}) = ||\boldsymbol{Y} - \boldsymbol{VC}||_F^2$ and the total-variation spatial constraint as $g(\boldsymbol{C}) = ||\boldsymbol{D}\bar{\boldsymbol{C}}||_{2,p}^p$. For clarity of presentation, we use an unweighted difference operator $\boldsymbol{D}$ instead of $\boldsymbol{D}_{\boldsymbol{W}}$ (Equation (7)).

F.1. Variable Projection

In this section we use variable projection to eliminate the dependence of the least-squares penalty on $\boldsymbol{V}$. An overview of variable projection applied to separable nonlinear least-squares problems is provided by Golub & Pereyra (2003). For completeness we note that alternating iterative minimization over $\boldsymbol{V}$ and $\boldsymbol{C}$ could also be used; however, this would not necessarily be superior to the variable projection used here. For simplicity, we work with a transposed version of the least-squares penalty:

$$f(\boldsymbol{V}, \boldsymbol{C}) = \|\boldsymbol{Y} - \boldsymbol{VC}\|_F^2 = \|\boldsymbol{Y}^T - \boldsymbol{C}^T\boldsymbol{V}^T\|_F^2. \tag{F2}$$

We denote as $h(\boldsymbol{C}) = \underset{\boldsymbol{V}}{\arg\min} \|\boldsymbol{Y}^T - \boldsymbol{C}^T\boldsymbol{V}^T\|_F^2$ the value $\boldsymbol{V}$ that minimizes this penalty for a particular $\boldsymbol{C}$. Here, and afterward, $h(\cdot)^T$ denotes the transpose of $h(\cdot)$. The minimum $\boldsymbol{V}^T$ is the minimum-norm least-squares solution, which is the pseudoinverse of $\boldsymbol{C}^T$ applied to $\boldsymbol{Y}^T$ (Golub & Van Loan 2013). Denoting the pseudoinverse of $\boldsymbol{C}^T$ as $\boldsymbol{C}^\dagger = (\boldsymbol{C}\boldsymbol{C}^T)^{-1}\boldsymbol{C}$ we obtain $h(\boldsymbol{C})^T = \boldsymbol{C}^\dagger\boldsymbol{Y}^T$. Rewriting the least-squares term using the conditional solution:

$$f(h(\boldsymbol{C}), \boldsymbol{C}) = ||\boldsymbol{Y}^T - \boldsymbol{C}^T h(\boldsymbol{C})^T||_F^2 \tag{F3}$$

$$= ||\boldsymbol{Y}^T - \boldsymbol{C}^T\boldsymbol{C}^\dagger\boldsymbol{Y}^T||_F^2 \tag{F4}$$

$$= ||(\mathbb{1}_I - \boldsymbol{C}^T\boldsymbol{C}^\dagger)\boldsymbol{Y}^T||_F^2. \tag{F5}$$

As $\boldsymbol{C}^T\boldsymbol{C}^\dagger = \boldsymbol{P}_{R(\boldsymbol{C}^T)}$ is the projection onto the range space of $\boldsymbol{C}^T$ (Gentle 2010), then $(\mathbb{1}_I - \boldsymbol{P}_{R(\boldsymbol{C}^T)}) = \boldsymbol{P}_{R(\boldsymbol{C}^T)}^\perp$ is the projection onto the orthogonal complement to the range space of $\boldsymbol{C}^T$. The variable-reduced least-squares penalty is denoted as $f'(\boldsymbol{C})$:

$$f'(\boldsymbol{C}) = f(h(\boldsymbol{C}), \boldsymbol{C}) = ||\boldsymbol{P}_{R(\boldsymbol{C}^T)}^\perp \boldsymbol{Y}^T||_F^2. \tag{F6}$$

F.2. Gradient Descent

We use gradient descent to minimize $w(\boldsymbol{C}) = ||(\mathbb{1}_I - \boldsymbol{C}^T\boldsymbol{C}^\dagger)\boldsymbol{Y}^T||_F^2 + ||\boldsymbol{D}\bar{\boldsymbol{C}}^T||_{2,p}^p = f'(\boldsymbol{C}) + g(\boldsymbol{C})$ with respect to $\boldsymbol{C}^T \in \mathbb{R}^{I \times K}$. The gradient $\nabla w(\boldsymbol{C}^T) \in \mathbb{R}^{I \times K}$ is the matrix of partial derivatives. We denote the gradient of the variable-projection least-squares term as $\nabla f'(\boldsymbol{C}^T)$ and the gradient of the total-variation constraint as $\nabla g(\boldsymbol{C}^T)$. The total

gradient is:

$$\nabla w(\boldsymbol{C}^T) = \nabla f'(\boldsymbol{C}^T) + \nabla g(\boldsymbol{C}^T) \tag{F7}$$

where $\nabla w(\boldsymbol{C})^T = \nabla w(\boldsymbol{C}^T)$.

F.2.1. Variable-projection Least-squares Gradient

For clarity in computing $\nabla f'(\boldsymbol{C}^T)$, we expand the variable-projection least-squares penalty as $||\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}\boldsymbol{Y}^T||_F^2 = \sum_N ||\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}[\boldsymbol{Y}]_n||_F^2$ where $[\boldsymbol{Y}]_n$ is the n^{th} row of $\boldsymbol{Y} \in \mathbb{R}^{N\times I}$. Noting the Frobenius matrix norm may be expressed as a Frobenius matrix inner product, as $\|\boldsymbol{A}\|_F^2 = \langle \boldsymbol{A}, \boldsymbol{A}\rangle$, where the Frobenius matrix inner product is defined as $\langle \boldsymbol{A}, \boldsymbol{B}\rangle = tr(\boldsymbol{A}^T\boldsymbol{B})$ and tr denotes a trace (Petersen & Pedersen 2012). By application of the chain rule:

$$d||\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}[\boldsymbol{Y}]_n||_F^2 = \langle 2\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}[\boldsymbol{Y}]_n, d\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}[\boldsymbol{Y}]_n\rangle \tag{F8}$$

$$= \langle 2\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}[\boldsymbol{Y}]_n[\boldsymbol{Y}]_n^T, d\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}\rangle. \tag{F9}$$

The derivative of the projection matrix is shown due to Harville (2008; Theorem 15.11.1):

$$d\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)} = -\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}(d\boldsymbol{C}^T)\boldsymbol{C}^{\dagger} - (\boldsymbol{C}^{\dagger})^T(d\boldsymbol{C})\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}. \tag{F10}$$

Substituting this result for $d\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}$ into (F9):

$$d||\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}[\boldsymbol{Y}]_n||_F^2 = -2\langle \boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}[\boldsymbol{Y}]_n[\boldsymbol{Y}]_n^T, \boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)} \times (d\boldsymbol{C}^T)\boldsymbol{C}^{\dagger} + (\boldsymbol{C}^{\dagger})^T(d\boldsymbol{C})\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}\rangle \tag{F11}$$

$$= -2\langle \boldsymbol{P}^{\perp T}_{R(\boldsymbol{C}^T)}\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}[\boldsymbol{Y}]_n[\boldsymbol{Y}]_n^T(\boldsymbol{C}^{\dagger})^T, d\boldsymbol{C}^T\rangle - 2\langle \boldsymbol{C}^{\dagger}\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}[\boldsymbol{Y}]_n[\boldsymbol{Y}]_n^T, (d\boldsymbol{C})\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}\rangle. \tag{F12}$$

Noting that an orthogonal projection matrix $\boldsymbol{P}$ is symmetric $\boldsymbol{P}^T = \boldsymbol{P}$ and idempotent $\boldsymbol{P}\boldsymbol{P} = \boldsymbol{P}$ (Golub & Van Loan 2013), the first term in Equation (F12) may be further simplified. Consider $\boldsymbol{C}^{\dagger}\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)} = (\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}(\boldsymbol{C}^{\dagger})^T)^T$. It can be seen that the range space of $(\boldsymbol{C}^{\dagger})^T$ is the range space of $\boldsymbol{C}^T$; therefore $\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}(\boldsymbol{C}^{\dagger})^T = 0$ and $\boldsymbol{C}^{\dagger}\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)} = 0$, so that the second term in Equation (F12) is dropped:

$$d||\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}[\boldsymbol{Y}]_n||_F^2 = -2\langle \boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}[\boldsymbol{Y}]_n[\boldsymbol{Y}]_n^T(\boldsymbol{C}^{\dagger})^T, d\boldsymbol{C}^T\rangle. \tag{F13}$$

Rearranging and summing together the derivatives as $\sum_N \frac{d}{d\boldsymbol{C}^T}||\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}[\boldsymbol{Y}]_n||_F^2$, the gradient $\nabla f'(\boldsymbol{C}^T) \in \mathbb{R}^{I\times K}$ is:

$$\nabla f'(\boldsymbol{C}^T) = -2\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}\left(\sum_n [\boldsymbol{Y}]_n[\boldsymbol{Y}]_n^T\right)(\boldsymbol{C}^{\dagger})^T. \tag{F14}$$

This can be compactly expressed as:

$$\nabla f'(\boldsymbol{C}^T) = -2\boldsymbol{P}^{\perp}_{R(\boldsymbol{C}^T)}\boldsymbol{Y}^T\boldsymbol{Y}(\boldsymbol{C}^{\dagger})^T. \tag{F15}$$

F.2.2. Total-variation Gradient

The derivative of $||\boldsymbol{D}\bar{\boldsymbol{C}}^T||_{2,p}^p$ can be computed by application of the chain rule:

$$\nabla g(\boldsymbol{C}^T) = \frac{d}{d\boldsymbol{C}^T}||\boldsymbol{D}\bar{\boldsymbol{C}}^T||_{2,p}^p = \frac{d}{d\bar{\boldsymbol{C}}^T}||\boldsymbol{D}\bar{\boldsymbol{C}}^T||_{2,p}^p \cdot \frac{d\bar{\boldsymbol{C}}^T}{d\boldsymbol{C}^T}. \tag{F16}$$

The matrix $\boldsymbol{D}\bar{\boldsymbol{C}}^T \in \mathbb{R}^{2\ \times\ K\cdot X\cdot Y}$ has columns for each coefficient $k \in K$ and each pixel $j \to (x, y) \in X \times Y$, formed as $[[\boldsymbol{D}_x\bar{\boldsymbol{C}}^T]_{j,k}, [\boldsymbol{D}_y\bar{\boldsymbol{C}}^T]_{j,k}] = [\bar{c}^k_{x,y} - \bar{c}^k_{x+1,y}, \bar{c}^k_{x,y} - \bar{c}^k_{x,y+1}]$, where $\boldsymbol{D}_x \in \mathbb{R}^{(X-1)Y\times XY}$ and $\boldsymbol{D}_y \in \mathbb{R}^{X(Y-1)\times XY}$ are difference matrices in x and y, respectively. Note, for the edge pixels in the top row and right-most column, difference values are only included along X or Y, respectively.

The total-variation penalty computes the L_p^p norm over the L_2 norm of each column of $\boldsymbol{D}\bar{\boldsymbol{C}}^T$, given by $\|\boldsymbol{D}\bar{\boldsymbol{C}}^T\|_{2,p}^p = \sum_n||[\boldsymbol{D}\bar{\boldsymbol{C}}^T]_{:,n}||_2^p = \sum_k\sum_j||[[\boldsymbol{D}_x\bar{\boldsymbol{C}}^T]_{j,k}, [\boldsymbol{D}_y\bar{\boldsymbol{C}}^T]_{j,k}]||_2^p$. Here each index n maps to some (j, k), but the ordering does not affect the computation.

The L_p norm is nondifferentiable at zero for $p \in [1, 2)$ and a differentiable approximation of the L_p norm is used instead in the form of the Huber functional (Vogel 2002). The Huber functional avoids zero values by adding a very small positive value δ to every element.

We apply the Huber functional to arrive at the modified spatial constraint:

$$||\boldsymbol{D}\bar{\boldsymbol{C}}^T||_{2,p}^p \approx \sum_n(||[\boldsymbol{D}\bar{\boldsymbol{C}}^T]_{:,n}||_2^2 + \delta)^{\frac{p}{2}}, \tag{F17}$$

where

$$||[\boldsymbol{D}\bar{\boldsymbol{C}}^T]_{:,n}||_2^2 = [\boldsymbol{D}_x\bar{\boldsymbol{C}}^T]^2_{n\to j,k} + [\boldsymbol{D}_y\bar{\boldsymbol{C}}^T]^2_{n\to j,k}. \tag{F18}$$

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

$$\frac{d||\boldsymbol{D}\bar{\boldsymbol{C}}^T||_{2,p}^p}{d[\bar{\boldsymbol{C}}^T]_{:,k}} = p\sum_j([\boldsymbol{D}_x\bar{\boldsymbol{C}}^T]^2_{j,k} + [\boldsymbol{D}_y\bar{\boldsymbol{C}}^T]^2_{j,k} + \delta)^{\frac{p}{2}-1} \times ([\boldsymbol{D}_x\bar{\boldsymbol{C}}^T]_{j,k}[\boldsymbol{D}_x]_j^T + [\boldsymbol{D}_y\bar{\boldsymbol{C}}^T]_{j,k}[\boldsymbol{D}_y]_j^T). \tag{F19}$$

We denote $\boldsymbol{L}_k = \mathrm{diag}(p([\boldsymbol{D}_x\bar{\boldsymbol{C}}^T]^2_{j,k} + [\boldsymbol{D}_y\bar{\boldsymbol{C}}^T]^2_{j,k} + \delta)^{\frac{p}{2}-1}\ \forall j)$. The sum over j allows for the expression to be written compactly as:

$$\frac{d||\boldsymbol{D}\bar{\boldsymbol{C}}^T||_{2,p}^p}{d[\bar{\boldsymbol{C}}^T]_{:,k}} = (\boldsymbol{D}_x^T\boldsymbol{L}_k\boldsymbol{D}_x + \boldsymbol{D}_y^T\boldsymbol{L}_k\boldsymbol{D}_y)[\bar{\boldsymbol{C}}^T]_{:,k}. \tag{F20}$$

The final term in Equation (F16) is the derivative due to the row-wise normalization of $\bar{\boldsymbol{C}}^T$.

$$\frac{d[\bar{\boldsymbol{C}}^T]_i}{d[\boldsymbol{C}^T]_j} = \begin{cases}(\mathbb{1}_K - [\bar{\boldsymbol{C}}^T]_i[\bar{\boldsymbol{C}}^T]_i^T)\dfrac{1}{||[\boldsymbol{C}^T]_i||} & i = j \\ 0 & i \neq j\end{cases}. \tag{F21}$$

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

$$[\nabla g(\boldsymbol{C}^T)]_i = [\boldsymbol{B}_1[\bar{\boldsymbol{C}}^T]_{:,1}, \ldots \boldsymbol{B}_K[\bar{\boldsymbol{C}}^T]_{:,K}]_i \times (\mathbb{1}_K - [\bar{\boldsymbol{C}}^T]_i[\bar{\boldsymbol{C}}^T]_i^T)\frac{1}{||[\boldsymbol{C}^T]_i||}, \tag{F22}$$

where $\boldsymbol{B}_k = \boldsymbol{D}_x^T\boldsymbol{L}_k\boldsymbol{D}_x + \boldsymbol{D}_y^T\boldsymbol{L}_k\boldsymbol{D}_y$.

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

$$[\nabla g(\boldsymbol{C}^T)]_i = [(\boldsymbol{D}_x^T\boldsymbol{D}_x + \boldsymbol{D}_y^T\boldsymbol{D}_y)\bar{\boldsymbol{C}}^T]_i \times (\mathbb{1}_K - [\bar{\boldsymbol{C}}^T]_i[\bar{\boldsymbol{C}}^T]_i^T)\frac{1}{||[\boldsymbol{C}^T]_i||}. \tag{F23}$$

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

ORCID iDs

Jamila S. Taaki https://orcid.org/0000-0001-5475-1975
Athol J. Kemball https://orcid.org/0000-0001-6233-8347

References

Aigrain, S., Hodgkin, S. T., Irwin, M. J., Lewis, J. R., & Roberts, S. J. 2015, MNRAS, 447, 2880
Aigrain, S., Parviainen, H., & Pope, B. J. S. 2016, MNRAS, 459, 2408
Aigrain, S., Parviainen, H., Roberts, S., Reece, S., & Evans, T. 2017, MNRAS, 471, 759
Armijo, L. 1966, Pac. J. Appl. Math., 16, 1
Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, ApJ, 935, 167
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Auvergne, M., Bodin, P., Boisnard, L., et al. 2009, A&A, 506, 411
Bakos, G. Á., Kovács, G., Torres, G., et al. 2007, ApJ, 670, 826
Batalha, N. M. 2014, PNAS, 111, 12647
Borucki, W. J., Koch, D., Basri, G., et al. 2010, Sci, 327, 977
Bryson, S., Coughlin, J., Batalha, N. M., et al. 2020, AJ, 159, 279
Bryson, S. T., Tenenbaum, P., Jenkins, J. M., et al. 2010, ApJL, 713, L97
Caldwell, D. A., Kolodziejczak, J. J., Cleve, J. E. V., et al. 2010, ApJL, 713, L92
Candès, E. J., Li, X., Ma, Y., & Wright, J. 2011, JACM, 58, 1
Charbonneau, D., Allen, L. E., Megeath, S. T., et al. 2005, ApJ, 626, 523
Christiansen, J. L., Jenkins, J. M., Caldwell, D. A., et al. 2012, PASP, 124, 1279
Crossfield, I. J. M., Petigura, E., Schlieder, J. E., et al. 2015, ApJ, 804, 10
Deeg, H. J., & Alonso, R. 2018, Transit Photometry as an Exoplanet Discovery Method (Cham: Springer), 1
Deming, D., Knutson, H., Kammer, J., et al. 2015, ApJ, 805, 132
Demory, B.-O., Ehrenreich, D., Queloz, D., et al. 2015, MNRAS, 450, 2043
Eckart, C., & Young, G. 1936, Psychometrika, 1, 211
Edwards, M. 1994, Advanced Calculus: A Differential Forms Approach (Basel: Birkhäuser)
Fausnaugh, M. 2018, TESS Data Release Notes: Sector 1 DR1, NASA, https://tasoc.dk/docs/release_notes/tess_sector_01_drn01_v02.pdf
Foreman-Mackey, D., Montet, B. T., Hogg, D. W., et al. 2015, ApJ, 806, 215
Gentle, J. E. 2010, Matrix Algebra (Berlin: Springer)
Gilliland, R. L., Chaplin, W. J., Dunham, E. W., et al. 2011, ApJS, 197, 6
Goldstein, T., & Osher, S. 2009, SIIMS, 2, 323
Golub, G., & Pereyra, V. 2003, InvPr, 19, R1
Golub, G., & Van Loan, C. 2013, Matrix Computations, Johns Hopkins Studies in the Mathematical Sciences (Baltimore, Maryland: Johns Hopkins Univ. Press), https://books.google.com/books?id=5U-l8U3P-VUC
Golub, G. H., & Pereyra, V. 1973, SJNA, 10, 413
Grinstead, C. M., & Snell, J. L. 2012, Introduction to Probability (Providence, RI: American Mathematical Society)
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
Harville, D. 2008, Matrix Algebra From a Statistician's Perspective (New York: Springer), https://books.google.com/books?id=kZGBQijgGV8C
Hastie, T., Tibshirani, R., & Friedman, J. 2009, Unsupervised Learning (New York: Springer), 485
Hattori, S., Foreman-Mackey, D., Hogg, D. W., et al. 2022, AJ, 163, 284
Hedges, C., Luger, R., Dotson, J., Foreman-Mackey, D., & Barentsen, G. 2021, AJ, 161, 95
Huang, X., Soares-Furtado, M., Penev, K., et al. 2015, AAS/Division for Extreme Solar Systems Abstracts, 47, 114.06
Huber, W. 2016, How does Cosine Similarity Change after a Linear transformation?, Cross Validated, https://stats.stackexchange.com/q/206083
Hunter, J. D. 2007, CSE, 9, 90
Jenkins, J. M. 2017, Kepler Data Processing Handbook: Overview of the Science Operations Center, Kepler Science Document KSCI-19081-002, NASA
Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., et al. 2010, ApJL, 713, L87
Karl, W. C. 2005, Handbook of Image and Video Processing (Amsterdam: Elsevier), 183
Koch, D. G., Borucki, W. J., Basri, G., et al. 2010, ApJL, 713, L79
Kovács, G., Bakos, G., & Noyes, R. W. 2005, MNRAS, 356, 557
Lightkurve Collaboration, Cardoso, J. V. d. M., Hedges, C., et al., 2018, Lightkurve: Kepler and TESS Time Series Analysis in Python, Astrophysics Source Code Library, ascl:1812.013
Luger, R., Agol, E., Kruse, E., et al. 2016, AJ, 152, 100
Luger, R., Kruse, E., Foreman-Mackey, D., Agol, E., & Saunders, N. 2018, AJ, 156, 99
Lund, M. N., Handberg, R., Buzasi, D. L., et al. 2021, ApJS, 257, 53
Lund, M. N., Handberg, R., Davies, G. R., Chaplin, W. J., & Jones, C. D. 2015, ApJ, 806, 30
Mazeh, T., Guterman, P., Aigrain, S., et al. 2009, A&A, 506, 431
McLean, I. 2008, Electronic Imaging in Astronomy: Detectors and Instrumentation (Berlin: Springer), https://books.google.com/books?id=LXH2UavpcakC
Moreno, J., Buttry, R., O'Brien, J., et al. 2021, AJ, 162, 232
Murphy, K. 2020, Machine Learning (Cambridge, MA: MIT Press), https://books.google.com/books?id=Qs_tDwAAQBAJ
Ofir, A., Alonso, R., Bonomo, A. S., et al. 2010, MNRAS: Letters, 404, L99
Parikh, N., & Boyd, S. 2013, Proximal Algorithms, Foundations and Trends in Optimization (Boston, MA: Now Publishers), https://books.google.com/books?id=DS04ngEACAAJ
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825
Petersen, K. B., & Pedersen, M. S. 2012, The Matrix Cookbook, http://www.math.uwaterloo.ca/~hwolkowi//matrixcookbook.pdf
Petigura, E. A., & Marcy, G. W. 2012, PASP, 124, 1073
Phillips, J. M. 2021, Mathematical Foundations of Data Analysis (Cham: Springer)
Pont, F., Zucker, S., & Queloz, D. 2006, MNRAS, 373, 231
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, JATIS, 1, 014003
Roberts, S., McQuillan, A., Reece, S., & Aigrain, S. 2013, MNRAS, 435, 3639
Rudin, L. I., Osher, S., & Fatemi, E. 1992, Phys. D: Nonlinear Phenom., 60, 259
Shearer, P., & Gilbert, A. C. 2013, InvPr, 29, 045003
Smith, J. C., Stumpe, M. C., Van Cleve, J. E., et al. 2012, PASP, 124, 1000
Smith, K. L., Mushotzky, R. F., Boyd, P. T., et al. 2018, ApJ, 857, 141
Srebro, N. 2004, Thesis, MIT
Stefansson, G., Mahadevan, S., Hebb, L., et al. 2017, ApJ, 848, 9
STScI 2016a, Kepler LC, Q6, STScI/MAST, doi:10.17909/T9388B
STScI 2016b, Kepler LC, Q10, STScI/MAST, doi:10.17909/T9K88P
STScI 2016c, Kepler LC, Q14, STScI/MAST, doi:10.17909/T92C7P
Stumpe, M., Smith, J., Catanzarite, J., et al. 2014, PASP, 126, 100
Stumpe, M. C., Smith, J. C., Van Cleve, J. E., et al. 2012, PASP, 124, 985
Taaki, J., 2023 spatial-detrend, figshare, doi:10.6084/M9.FIGSHARE.24680868
Taaki, J. S., Kamalabadi, F., & Kemball, A. J. 2020, AJ, 159, 283
Tamuz, O., Mazeh, T., & Zucker, S. 2005, MNRAS, 356, 1466
Tenenbaum, P., & Jenkins, J. 2018, TESS Science Data Products Description EXP-TESS-ARC-ICD-0014 Rev D, NASA
Thatte, A., Deroo, P., & Swain, M. R. 2010, A&A, 523, A35
Toyozumi, H., & Ashley, M. C. 2005, PASA, 22, 257
Tregloan-Reed, J., & Southworth, J. 2013, MNRAS, 431, 966
Twicken, J. D., Chandrasekaran, H., Jenkins, J. M., et al. 2010, Proc. SPIE, 7740, 77401U
Van Cleve, J. 2010, Kepler Data Release 3 Notes, Tech. Rep. 3, NASA, https://archive.stsci.edu/missions/kepler/docs/drn/release_notes03/DataRelease_03_20100120.pdf
Van Cleve, J. E., & Caldwell, D. A. 2016, Kepler Instrument Handbook, Kepler Science Document KSCI-19033-002, NASA
Vanderburg, A., & Johnson, J. A. 2014, PASP, 126, 948
Vanderspek, R. 2018, Tess Instrument Handbook
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Vogel, C. R. 2002, Computational Methods for Inverse Problems (Philadelphia, PA: SIAM)
Wang, D., Hogg, D. W., Foreman-Mackey, D., & Schölkopf, B. 2016, PASP, 128, 094503
Waskom, M. L. 2021, JOSS, 6, 3021
Wold, H. O. A. 1973, Nonlinear Iterative Partial Least Squares (NIPALS) Modeling: Some Current Developments (Cambridge, MA: Academic Press), 383