

REPUBLIC: A variability-preserving systematic-correction algorithm for *PLATO*'s multi-camera light curves

Oscar Barragán,^{1★} Suzanne Aigrain¹ and James McCormac^{2,3}

¹Sub-department of Astrophysics, Department of Physics, University of Oxford, Oxford OX1 3RH, UK

²Physics Department, University of Warwick, Coventry CV4 7AL, UK

³Centre for Exoplanets and Habitability, University of Warwick, Coventry CV4 7AL, UK

Accepted 2024 April 2. Received 2024 March 27; in original form 2023 November 9

ABSTRACT

Space-based photometry missions produce exquisite light curves that contain a wealth of stellar variability on a wide range of time-scales. Light curves also typically contain significant instrumental systematics – spurious, non-astrophysical trends that are common, in varying degrees, to many light curves. Empirical systematics-correction approaches using the information in the light curves themselves have been very successful, but tend to suppress astrophysical signals, particularly on longer time-scales. Unlike its predecessors, the *PLANetary Transits and Oscillations of stars* (*PLATO*) mission will use multiple cameras to monitor the same stars. We present *REPUBLIC*, a novel systematics-correction algorithm which exploits this multi-camera configuration to correct systematics that differ between cameras, while preserving the component of each star's signal that is common to all cameras, regardless of time-scale. Through simulations with astrophysical signals (star spots and planetary transits), *Kepler*-like errors, and white noise, we demonstrate *REPUBLIC*'s ability to preserve long-term astrophysical signals usually lost in standard correction techniques. We also explore *REPUBLIC*'s performance with different number of cameras and systematic properties. We conclude that *REPUBLIC* should be considered a potential complement to existing strategies for systematic correction in multi-camera surveys, with its utility contingent upon further validation and adaptation to the specific characteristics of the *PLATO* mission data.

Key words: Data Methods – Numerical Methods – Algorithms – Techniques: Photometric.

1. INTRODUCTION

The transit method has enabled the detection of thousands of exoplanets so far (see e.g. NASA exoplanet archive,¹ Akeson et al. 2013), and is the most successful exoplanet detection method so far in terms of sheer numbers, thanks in particular to the *Kepler*, *K2*, and *TESS* space missions (Borucki et al. 2010; Howell et al. 2014; Ricker et al. 2015). Typically, transit surveys monitor thousands, if not tens of thousands, of stars simultaneously over a wide Field of View (FOV).

Twicken et al. (2010), Stumpe et al. (2012), and Smith et al. (2012) ensure that a key component of the data processing chain for these surveys is the removal of common-mode systematic effects. These are signals of instrumental (or, for ground-based surveys, telluric) origin that contaminate the light curves of many of the stars to a varying degree, after correction of other known instrumental effects. Methods to identify and correct these effects were first developed in the context of ground-based transit surveys, and typically consist of modelling each light curve as a linear combination of the other light curves observed on the same detector (e.g. Sys-Rem, Tamuz, Mazeh & Zucker 2005), or of external variables such as seeing and airmass. The Sys-Rem method was then refined for *Kepler* to ensure

that the extracted systematic trends, usually referred to as Cotrending Basis Vectors (CBVs), are not dominated by the intrinsic variability of the host stars, and to minimize the suppression of these intrinsic variabilities in the corrected light curves (Twicken et al. 2010; Smith et al. 2012; Stumpe et al. 2012). The resulting methodology, known as 'PDC-MAP', represents the current state of the art. It was used for *TESS* with minimal modifications, and forms part of the standard pipeline currently being developed for the *PLANetary Transits and Oscillations of stars* (*PLATO*) mission (Rauer et al. 2014).

In contrast to previous space-based transit surveys, *PLATO* uses multiple, small-aperture telescopes with partially overlapping FOV, arranged such that a given star is monitored simultaneously by 6, 12, 18, or 24 cameras, depending on its location in the FOV. On the ground, the Next Generation Transit Survey (NGTS, Wheatley et al. 2018) also uses multiple telescopes located at the same observatory site, which can be used to monitor the same FOV simultaneously (e.g. Smith et al. 2020). Both *PLATO* and NGTS produce per-camera light curves for each star. These are systematics-corrected on a per-camera basis (using variants of the Sys-Rem and PDC-MAP algorithms for NGTS and *PLATO*, respectively) and the light curves from the different cameras are then merged to produce a single light curve with higher signal-to-noise ratio for each star.

However, as mentioned above, the per-camera systematics removal step tends to suppress some of the intrinsic stellar variability, particularly on time-scales approaching the duration of the set of observations being processed. In this paper, we present an alternative

* E-mail: oscaribv@gmail.com

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

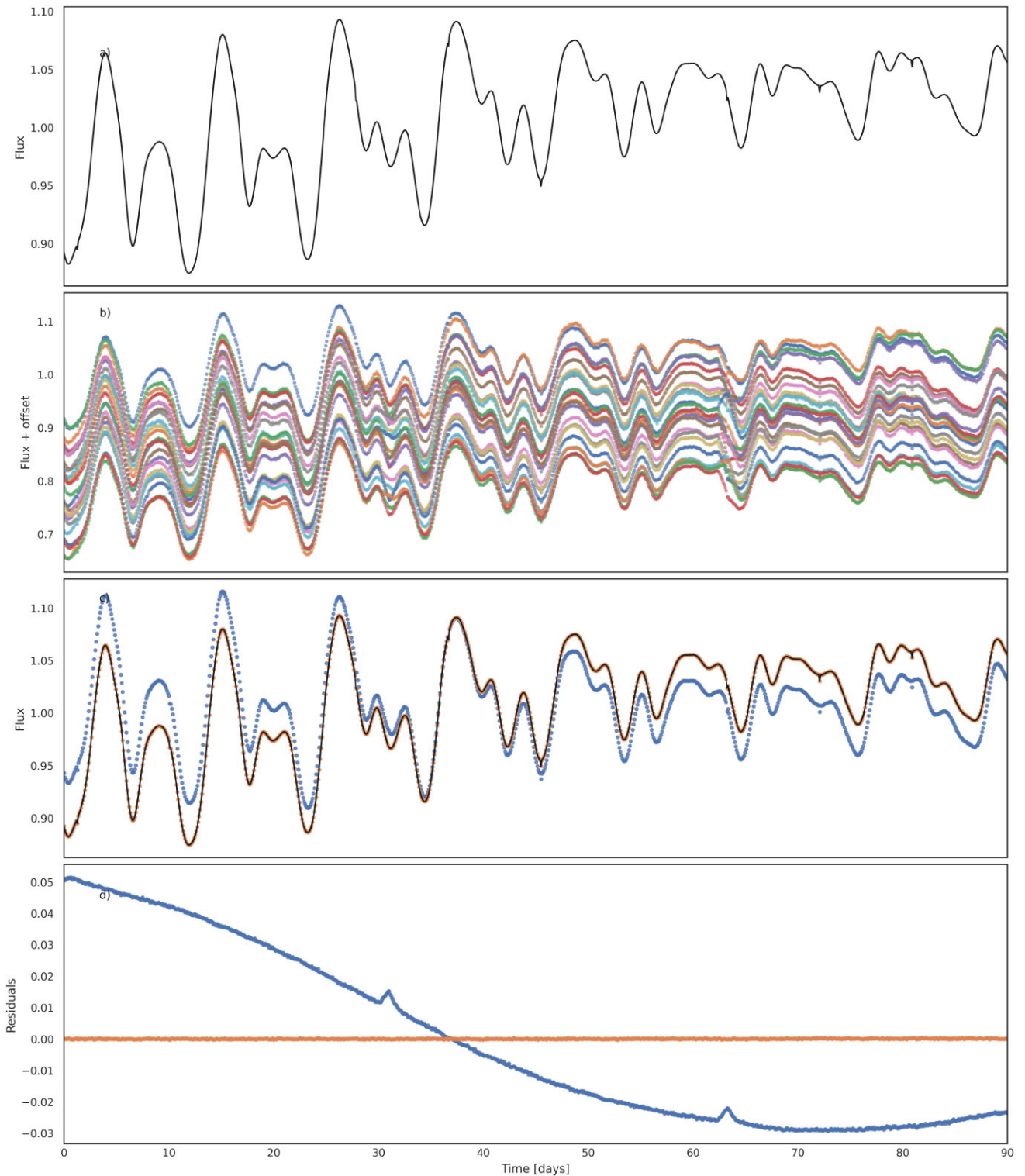


Figure 1. Example application of REPUBLIC to a simulated multi-camera light curve. Panel a shows the synthetic astrophysical signal. Panel b shows the light curves for the 24 simulated cameras. They contain systematics and white noise as described in the text, in addition to the astrophysical signal and an offset to help visibility. Panel c shows the recovered astrophysical signal using the REPUBLIC (orange circles) and PDC-LS-like (blue circles) algorithms. We also show the true signal (solid black line) for comparison. Panel d shows the difference between the merged, systematics-corrected light curve, and the injected astrophysical signal, using REPUBLIC (orange circles) and PDC-LS (blue circles).

systematics-correction method, dubbed REPUBLIC, which uses the multiple light curves for a given star across the different cameras, rather than the multiple light curves of all the stars on a given camera. Unlike existing methods, REPUBLIC explicitly models the

stellar signal, which is the same in all the light curves, alongside the systematics, which are specific to each camera. In REPUBLIC, all cameras get a ‘vote’ in determining the ‘true’ astrophysical signal. This allows for an efficient separation of any astrophysical signal

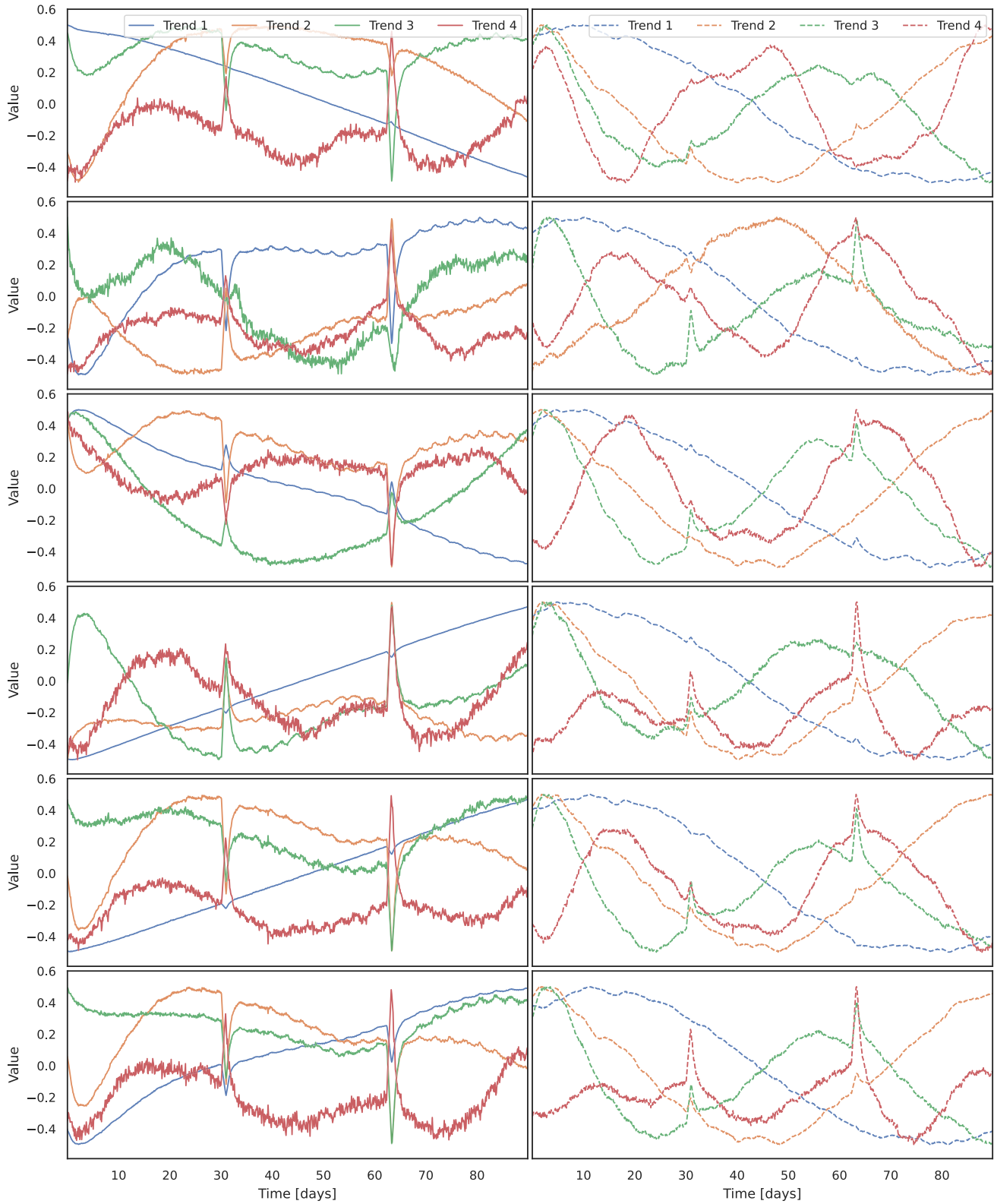


Figure 2. Left panel: Systematic trends added to our simulated light curves. These trends correspond to *Kepler* CBVs from its tenth quarter. Each panel (from top to bottom) corresponds to a different camera, while each colour represents a different trend. Right panel: Recovered systematic trends using PCA decomposition. Panels and colours represent the same cameras and trends as in the left panel.

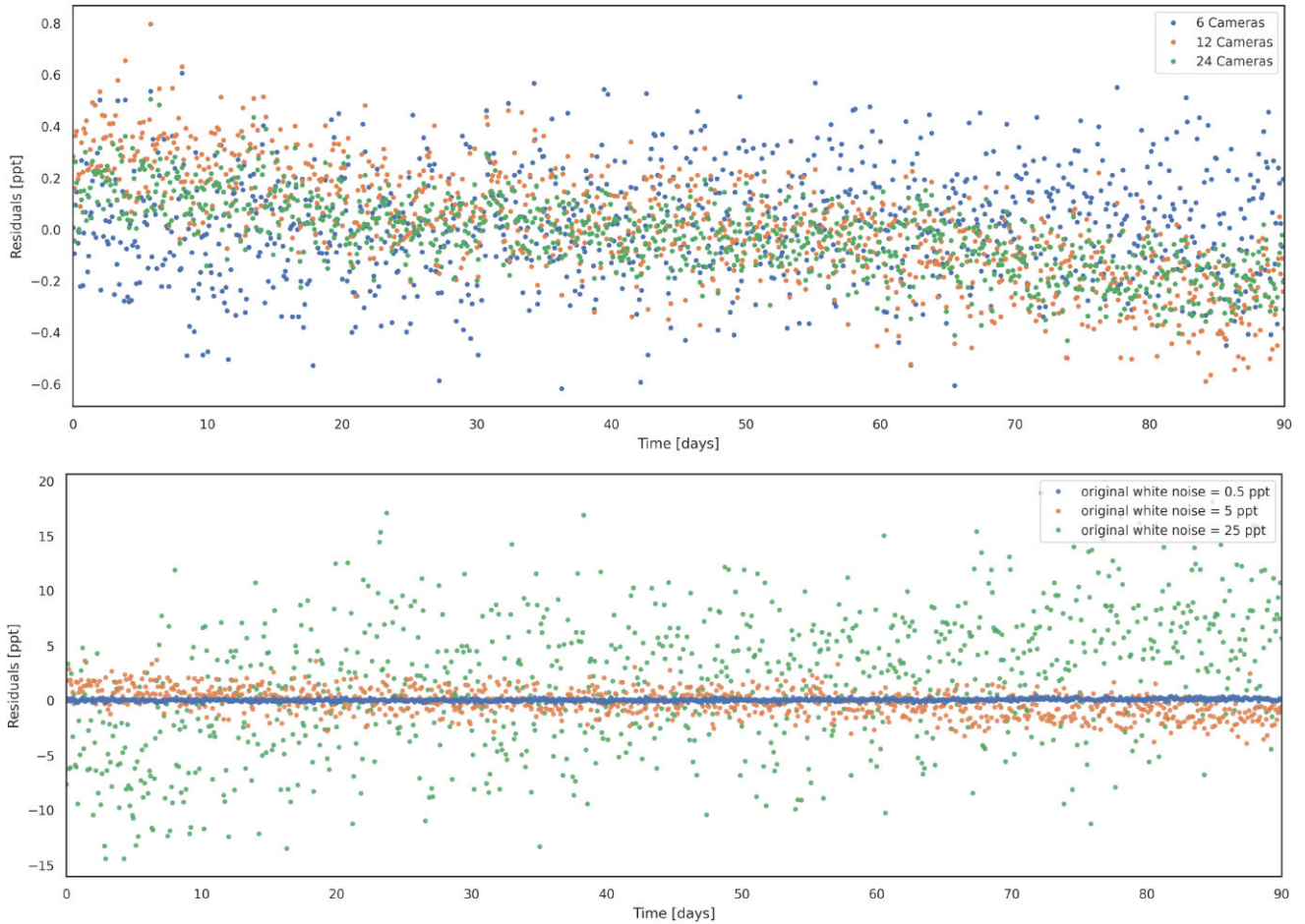


Figure 3. Dependence of the performance on the number of cameras (top panel) and the white noise level per camera (bottom panel). In both panels, we show the difference between the REPUBLIC-corrected, merged light curve and the true astrophysical signal, leaving behind white noise and residual systematic trends. The white noise level was fixed at 0.5 ppt in the top panel, and the number of cameras at 24 in the bottom one. Note that this example corresponds to the same light curve presented in Fig. 1.

from the systematics, provided that the latter differ sufficiently from camera to camera.

2. THE REPUBLIC ALGORITHM

2.1 The model

Let us assume we have I targets, observed by J cameras, with K observations per camera. The observed (background-subtracted) flux for target i on camera j in frame k can be written as

$$F_{ijk} = A_{ik} + B_{ijk} + E_{ijk}, \quad (1)$$

where A_{ik} represents the intrinsic flux of target i in frame k , B_{ijk} the systematic noise for target i in frame k on camera j , and E_{ijk} the corresponding random noise, which we assume to be white and Gaussian with known standard deviation σ_{ijk} . We note that A_{ik} does not depend on the camera j as we expect to observe the same astrophysical signal in all cameras.

Equation (1) assumes that the systematics are additive rather than multiplicative, and the contamination and aperture losses are the same in all cameras. In fact, the relative contribution of nearby contaminant stars to the measured flux for the target does vary from camera to camera. While this is not explicitly accounted for in our formalism, the time-dependent component of the contamination and aperture

losses, caused by the drift of the stars' positions over the camera pixel, can be accounted for by the systematics component of the model, which we detail below.

The systematics are modelled as a linear combination of individual 'trends', which are specific to each camera but common to all targets on that camera (similar to the PDC-MAP approach, Smith et al. 2012). Specifically, we write the systematic component for each star in each camera as

$$B_{ijk} = \sum_{n=1}^{N_j} W_{ijn} T_{jkn}, \quad (2)$$

where T_{jkn} is the value of the n th systematic trend for camera j in frame k , N_j is the number of systematic trends per camera j , and W_{ijn} is the coefficient linking that trend to target i .

2.2 Trend identification for each camera

To implement the model described in the previous section, the first step consists in identifying the systematic trends T_{jkn} for each camera. In this paper, we follow the PDC-MAP approach to extracting CBVs (Smith et al. 2012; Stumpe et al. 2012), which consists in selecting the most mutually correlated light curves on a given camera, then using the first few principal components of the resulting matrix as

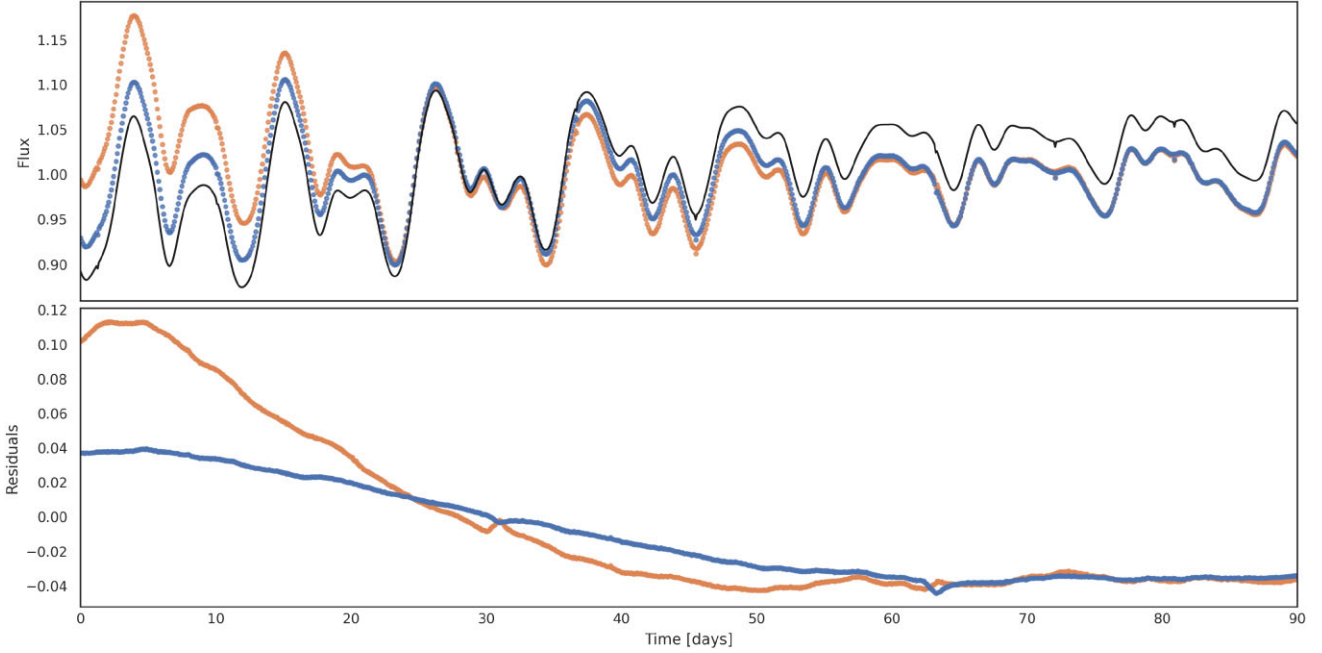


Figure 4. Impact of imperfectly known trends on REPUBLIC’s performance. The top panel shows the merged, REPUBLIC-corrected light curve with the PCA-extracted trends in orange, compared with the true astrophysical signal in black. We also show the performance of the PDC-LS-like algorithm using these imperfect trends in blue. The bottom panel shows the corresponding residuals for each case.

the trends for that camera. The number of CBVs per camera is determined by a user-specified threshold in the fraction of the total variance ‘explained’ by the trends. We note, that one could also use external variables such as airmass and seeing (for ground-based observations), or satellite attitude data, instead of (or in addition to) the CBVs.

2.3 Least-squares solution for each star

Once the trends are known, it is straightforward to solve for the stellar signal A_{ik} and the systematic trend weights W_{jn} for a given star i using the fluxes measured for that star on all cameras. From this point, we drop the subscript i in derivations, since we are considering one star at a time.

Assuming that the white noise standard deviation is known (which is a reasonable assumption for high-precision photometry), maximizing the model likelihood with respect to the parameters is equivalent to minimizing the total χ^2 . The total χ^2 for a given target is given by

$$\chi^2 = \sum_j \sum_k \sigma_{jk}^{-2} \left[F_{jk} - A_k - \sum_n W_{jn} T_{jkn} \right]^2. \quad (3)$$

Expanding the square:

$$\begin{aligned} \chi^2 = & \sum_j \sum_k \sigma_{jk}^{-2} \left[F_{jk}^2 + A_k^2 - 2F_{jk}A_k \right. \\ & + \sum_{n=1}^{N_j} W_{jn} T_{jkn} \sum_{m=1}^N W_{jm} T_{jkm} \\ & \left. + 2(A_k - F_{jk}) \sum_{n=1}^{N_j} W_{jn} T_{jkn} \right]. \quad (4) \end{aligned}$$

To solve for the parameters, we evaluate the derivatives of equation (4) with respect to A_k and W_{jn} and set them to zero:

$$0 = \frac{\partial \chi^2}{\partial A_k} = 2A_k \sum_j \frac{1}{\sigma_{jk}^2} - 2 \sum_j \frac{F_{jk}}{\sigma_{jk}^2} + 2 \sum_j \sum_n \frac{W_{jn} T_{jkn}}{\sigma_{jk}^2} \quad (5)$$

and

$$\begin{aligned} 0 = \frac{\partial \chi^2}{\partial W_{jn}} = & 2W_{jn} \sum_k \frac{T_{jkn}^2}{\sigma_{jk}^2} + 2 \sum_{m \neq n} W_{jm} \sum_k \frac{T_{jkn} T_{jkm}}{\sigma_{jk}^2} \\ & + 2 \sum_k \frac{A_k T_{jkn}}{\sigma_{jk}^2} - 2 \sum_k \frac{F_{jk} T_{jkn}}{\sigma_{jk}^2}. \quad (6) \end{aligned}$$

2.4 Matrix formulation

We now re-arrange equations (5) and (6), and combine them into a single matrix equation of the form

$$X\mathbf{p} = \mathbf{y}, \quad (7)$$

where \mathbf{p} is a vector containing all the unknowns quantities, i.e. the free parameters of the model, namely the A_k ’s and W_{jn} ’s, while the matrix X and the vector \mathbf{y} depend only on the known values F_{jk} , σ_{jk} , and the B_{jkn} . Specifically, the vector \mathbf{p} is given by

$$\mathbf{p} = (\mathbf{a}, \mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_j, \dots, \mathbf{w}_J)^T, \quad (8)$$

where

$$\mathbf{a} = (A_1, A_2, \dots, A_k, \dots, A_K)^T \quad (9)$$

and

$$\mathbf{w}_J = (W_{j1}, W_{j2}, \dots, W_{jn}, \dots, W_{jN_j})^T. \quad (10)$$

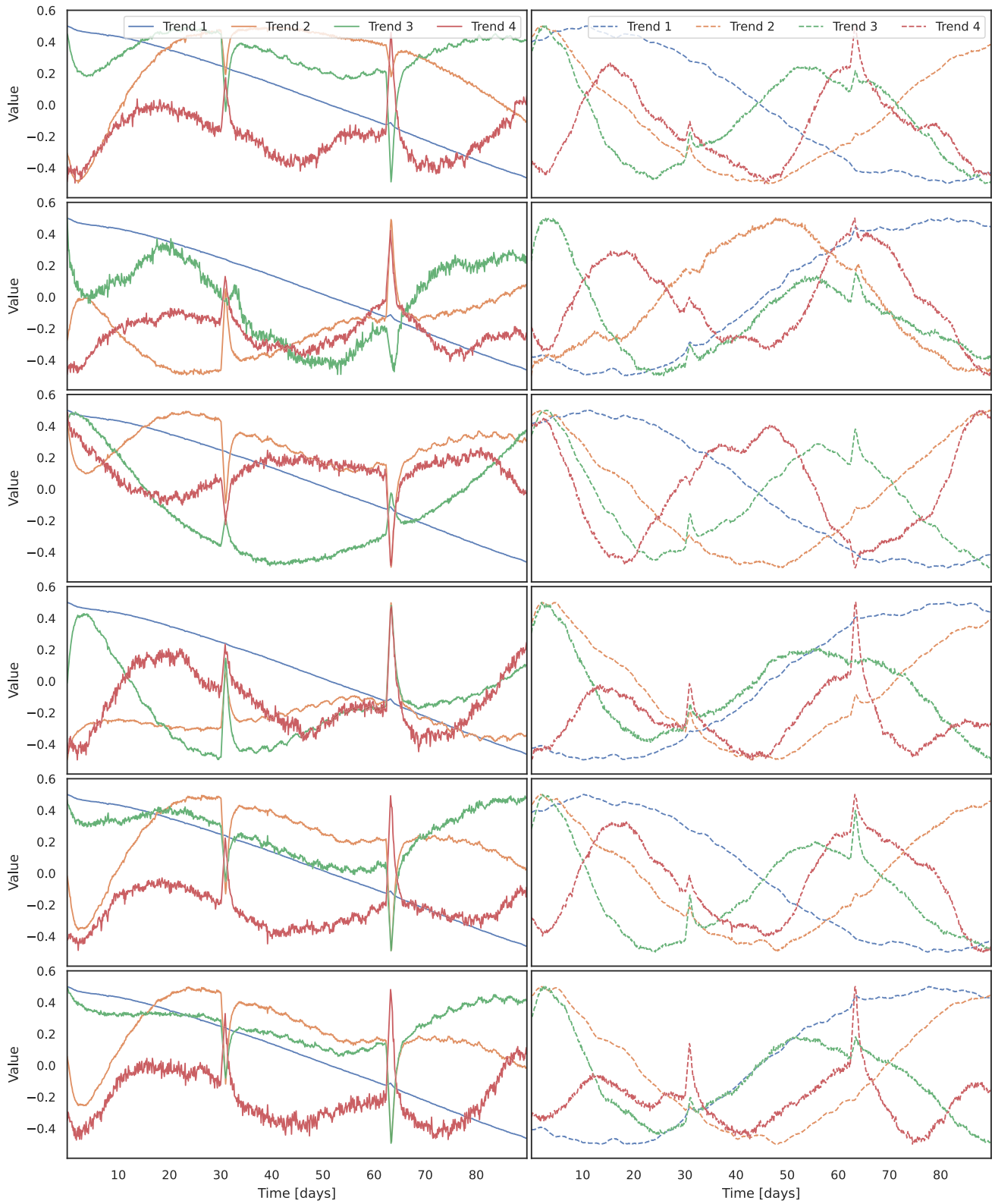


Figure 5. Plots are the same as in Fig. 2, for the identical trend case. In the left panel, one of the trends (blue) is identical for all the cameras.

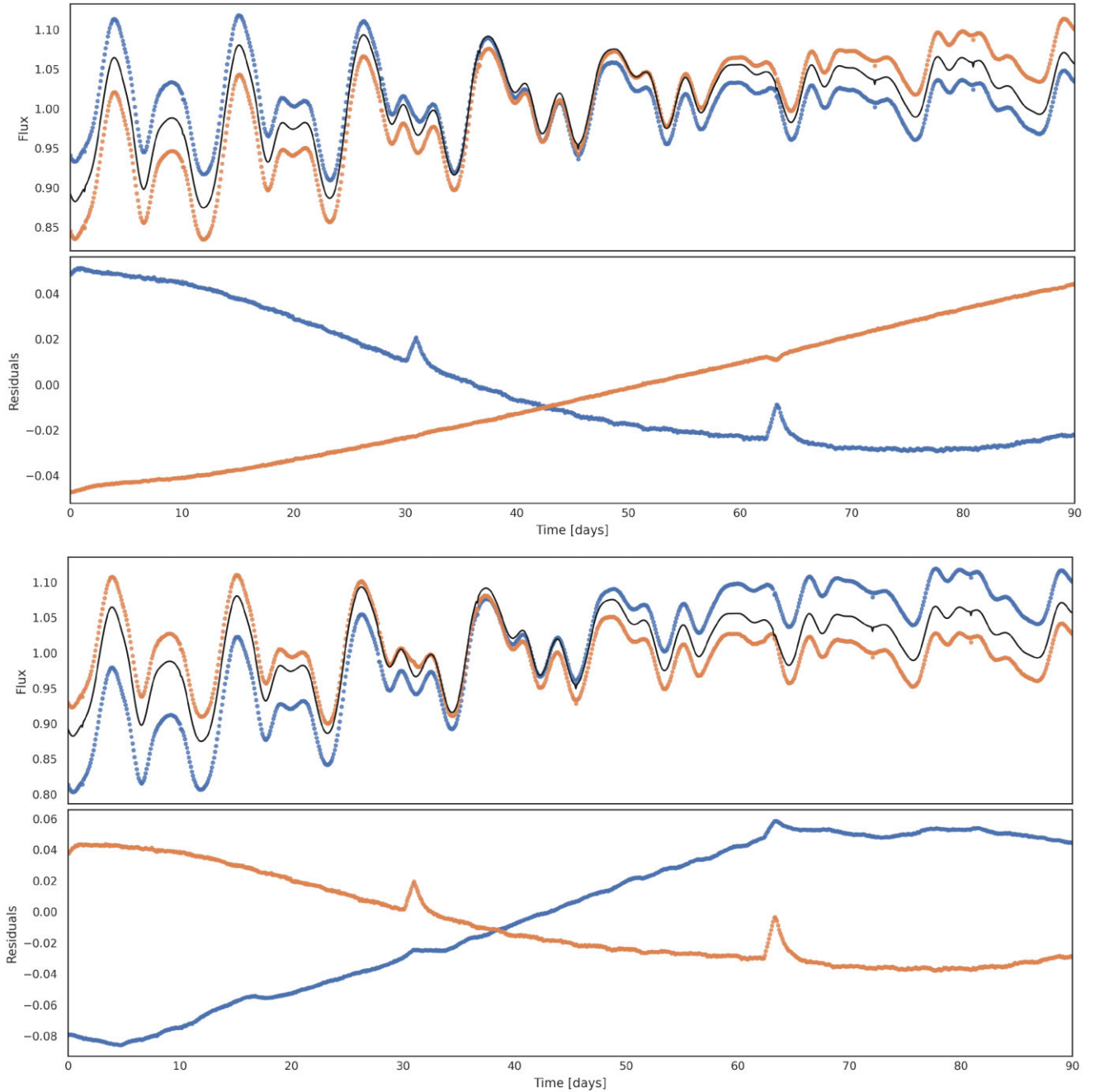


Figure 6. REPUBLIC (orange) and PDC-LS-like (blue) performance in the case of similar trends in the CBVs. Top panel shows the result of perfectly known trends while bottom panel shows the results with the PCA-extracted trends. The black line shows the true astrophysical signal. We show the corresponding residuals for each case.

The matrix X is given by

$$X = \begin{bmatrix} C & D_1 & D_2 & \dots & D_J \\ D_1^T & E_1 & 0 & \dots & 0 \\ D_2^T & 0 & E_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ D_J^T & 0 & 0 & \dots & E_J \end{bmatrix}, \quad (11)$$

$$C = \text{diag} \left\{ \sum_j \sigma_{jk}^{-2}, k = 1 \rightarrow K \right\}, \quad (12)$$

$$D_j = \begin{bmatrix} \frac{T_{j11}}{\sigma_{j1}^2} & \frac{T_{j12}}{\sigma_{j1}^2} & \dots & \frac{T_{j1N_j}}{\sigma_{j1}^2} \\ \frac{T_{j21}}{\sigma_{j2}^2} & \frac{T_{j22}}{\sigma_{j2}^2} & \dots & \frac{T_{j2N_j}}{\sigma_{j2}^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{T_{jK1}}{\sigma_{jK}^2} & \frac{T_{jK2}}{\sigma_{jK}^2} & \dots & \frac{T_{jKN_j}}{\sigma_{jK}^2} \end{bmatrix}, \quad (13)$$

where

and

$$E_j = \begin{bmatrix} \sum_k \frac{T_{jk1}^2}{\sigma_{jk}^2} & \sum_k \frac{T_{jk1}T_{jk2}}{\sigma_{jk}^2} & \cdots & \sum_k \frac{T_{jk1}T_{jkN_j}}{\sigma_{jk}^2} \\ \sum_k \frac{T_{jk1}T_{jk2}}{\sigma_{jk}^2} & \sum_k \frac{T_{jk2}^2}{\sigma_{jk}^2} & \cdots & \sum_k \frac{T_{jk2}T_{jkN_j}}{\sigma_{jk}^2} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_k \frac{T_{jk1}T_{jkN_j}}{\sigma_{jk}^2} & \sum_k \frac{T_{jk2}T_{jkN_j}}{\sigma_{jk}^2} & \cdots & \sum_k \frac{T_{jkN_j}^2}{\sigma_{jk}^2} \end{bmatrix}. \quad (14)$$

Finally, the vector \mathbf{y} is given by

$$\mathbf{y} = (\mathbf{q}, \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_j, \dots, \mathbf{r}_J)^T, \quad (15)$$

where

$$\mathbf{q} = \left(\sum_j \frac{F_{j1}}{\sigma_{j1}^2}, \sum_j \frac{F_{j2}}{\sigma_{j2}^2}, \dots, \sum_j \frac{F_{jK}}{\sigma_{jK}^2} \right)^T \quad (16)$$

is the variance-weighted average of the observed flux across all cameras, and

$$\mathbf{r}_j = \left(\sum_k \frac{F_{jk}T_{jk1}}{\sigma_{jk}^2}, \sum_k \frac{F_{jk}T_{jk2}}{\sigma_{jk}^2}, \dots, \sum_k \frac{F_{jk}T_{jkN}}{\sigma_{jk}^2} \right)^T. \quad (17)$$

2.5 Inverting the matrix X

We can now rewrite equation (11) in order to simplify the solution of the linear problem. X can be rewritten as a 2×2 block matrix of the form

$$X = \begin{bmatrix} C & D \\ D^T & E \end{bmatrix}, \quad (18)$$

where C , D , and E are computed from equations (12), (13), and (14), respectively. The inverse of X can then be computed as (see e.g. Lu & Shiou 2002)

$$X^{-1} = \begin{bmatrix} C^{-1} + C^{-1} D S^{-1} D^T C^{-1} & -C^{-1} D S^{-1} \\ -S^{-1} D^T C^{-1} & S^{-1} \end{bmatrix}, \quad (19)$$

where

$$S = E - D^T C^{-1} D, \quad (20)$$

is the Schur complement of the block C . Thus, to evaluate X^{-1} we only need to invert the matrices C and S . As C is diagonal, its inverse is given by

$$(C^{-1})_k = (C_k)^{-1} = \left(\sum_j \sigma_{jk}^{-2} \right)^{-1}. \quad (21)$$

Furthermore, S is Hermitian and positive-definite, so its inverse can be computed efficiently using Cholesky decomposition. We note that the size of matrix E is independent of the number of observations, while the size of matrix D scales quadratically with the number of observations, but it never needs to be inverted. Therefore, the computational cost of REPUBLIC scales linearly with the number of observations, and cubically with the total number of systematic trends across all cameras, while its memory usage scales cubically with the number of observations, and linear with the number of systematic trends.

We now have all the ingredients to compute

$$\mathbf{p} = X^{-1} \mathbf{y}, \quad (22)$$

and thus extract simultaneously the best-fitting astrophysical signal \mathbf{a} and the weights of the systematics model in each camera, \mathbf{w}_j . We can

also construct a systematics-corrected light curve in each camera:

$$\tilde{f}_j = f - \mathbf{w}_j^T \mathbf{T}_j. \quad (23)$$

3. TESTS

As a proof of concept, we created multi-camera, *PLATO*-like synthetic light curves containing astrophysical signals, systematic trends, and white noise. We then applied the algorithm described in Section 2 and compared the results with a naive (least-squares) implementation of the standard PDC algorithm, which corrects the light curves from each camera separately.

There are two main tools currently available to simulate large numbers of *PLATO* light curves. *PlatoSim* (Marcos-Arenal et al. 2014) is a full, end-to-end simulator, which starts by generating full images including astrophysical signals and all known instrumental effects, then mimics the pixel-level correction and light curve extraction and pre-processing implemented in the *PLATO* level 0 and level 1 pipelines. However, at the time of writing, the behaviour of the instrument remains limited, only a few of the cameras have been fully tested, and the level 1 pipeline is not yet fully implemented in *PlatoSim*. The only multi-camera, multi-quarter data set produced with *PlatoSim* (Janssen et al. 2023) was generated using pessimistic assumptions concerning some of the key instrumental noise sources, resulting in strong systematics that are common to all cameras, which makes it unsuitable to test REPUBLIC. Alternatively, the *PLATO* Solar-like Light curve Simulator (Samadi et al. 2019) is a lighter weight tool that uses semi-analytic approximations to generate systematics and noise time series to which the stellar signals are added separately. It produces outputs that are entirely uncorrelated between cameras, which would constitute an unrealistically favourable test for REPUBLIC. Therefore, we opted to produce our own, simpler simulations, where we control every aspect of the process, but note that REPUBLIC should be tested on more realistic simulated data once the behaviour of the satellite and instrument are better understood, and the pipelines have been finalized.

3.1 Light curve generation

3.1.1 Astrophysical signals

We used *citlalicue* (Barragán et al. 2022) to create synthetic stellar light curves containing transit signals and stellar activity signals. Granulation and stellar oscillations were not included in our simulations, as we wished to focus on astrophysical signals with time-scales similar to both the dominant systematics and the duration of the simulated observations, but we note that the REPUBLIC model is entirely agnostic regarding the nature and time-scales of the astrophysical signal.

Each light curve consists of 1000 equally spaced observations over 90 d (corresponding to one *PLATO* quarter). Note that this results in an interval between consecutive observations of just over 2 h, significantly longer than the cadence of real *PLATO* observations (which will vary between 2.5 and 600 s, depending on the stellar sample). This longer cadence is sufficient for a proof-of-concept, and results in much smaller data volumes. However, we note that REPUBLIC is agnostic of the sampling of the data, its performance will be the same even in the 2.5 s cadence. While we do not report on these tests here for the reasons outlined above, we have run REPUBLIC on up to eight quarters of simulated *PLATO* data at 25 s cadence, without encountering any CPU or memory usage problems.

To mimic the stellar activity signals, we use random samples from a Gaussian process (GP) with a quasi-periodic kernel:

$$K(t_i, t_j) = A^2 \exp \left[-\frac{(t_i - t_j)^2}{2\lambda_e^2} - \frac{\sin^2[\pi(t_i - t_j)/P_{\text{GP}}]}{2\lambda_p^2} \right] \quad (24)$$

where A is the amplitude (or standard deviation) of the GP; P_{GP} is its characteristic period; λ_p is the characteristic length-scale of the periodic component (the harmonic complexity Γ is related to λ_p via $\Gamma = 1/2\lambda_p^2$); and λ_e is the characteristic evolution time-scale. For each light curve, these hyperparameters were drawn randomly from uniform distributions in the following intervals: $A \in [1 \times 10^{-5}, 5 \times 10^{-3}]$ ppm; $\lambda_e \in [10, 1000]$ d; $\lambda_p \in [0.1, 2]$; and $P_{\text{GP}} \in [3, 30]$ d.

We also include one planetary transit signal per light curve. The parameters of the injected planetary signals were also drawn from uniform distributions with the following intervals: transit epoch $T_0 \in [0, 5]$ d; orbital period $P \in [0, 10]$ d; impact parameter $b \in [0, 1]$; scaled semi-major axis, $a/R_\star \in [1.5, 50]$; and scaled planet radius $r_p/R_\star \in [0.01, 0.1]$. We assumed all orbits to be circular, and did not include limb-darkening. An example of one of the synthetic stellar signals is shown in panel a of Fig. 1.

3.1.2 Systematic trends and white noise

We then produced multiple copies of each stellar signal, to which we added different simulated systematic effects to represent the observations taken by the different cameras. In the case of *PLATO*, the 24 cameras are arranged in four groups of six cameras each. Each camera group observes the same FOV and timestamps, with a new integration starting every 25 s. On the other hand, the FOV of the different camera groups overlap only partially, and their timestamps are staggered relative to each other (by 6.25 s) to allow for more efficient use of the readout electronics and processing power. In this paper, we ignore the fact that the timestamps are not simultaneous, but we note that the light curves from different camera groups would need to be interpolated onto the same grid of time-stamps before applying REPUBLIC.

To make our simulated systematics trends as realistic as possible, we used the LIGHTKURVE package (Lightkurve Collaboration 2018) to download CBVs extracted from the *Kepler* data by the PDC-MAP pipeline. We used the CBVs obtained for the tenth quarter of the mission, but we emphasize that this selection is arbitrary and these tests work with any other *Kepler* quarter. For each camera, we selected one of the 80 *Kepler* readout channels at random and used the first four CBVs from that channel as our systematic trends. We then assigned a random weight, between -50 and 50 ppt, to each CBV in each camera. In order to account for the difference in time-sampling between the *Kepler* data and our simulations, we interpolated the original CBVs to the timestamps of the synthetic light curves. In Fig. 2, we present an illustration of the four distinct systematics observed in the *Kepler* data set as they are included in six of our simulated *PLATO*-like light curves. Notably, these signals exhibit discernible trends and manifest abrupt discontinuities approximately every 30 d. These discontinuities correspond to the interruptions in *Kepler* observations to enable data to be transmitted back to Earth.

Finally, we added Gaussian white noise to the data, maintaining a standard deviation of 500 ppm for each individual exposure across all cameras. Panel b of Fig. 1 depicts a set of 24 synthetic multi-camera light curves corresponding to a single target, including the aforementioned white noise.

3.2 Results in the fiducial case

Our first set of simulations used different CBVs as the systematic trends in each camera, so that the resulting systematics were not strongly mutually correlated. Furthermore, we initially assumed that these trends were known perfectly, by setting the T_{jkn} in the REPUBLIC model equal to the CBVs used to generate the light curves. We then fit for the systematic weights W_{ijn} alongside the astrophysical signal A_{ik} as described in Sections 2.3–2.5. For comparison, we also perform a least-squares fit for the systematic weights in each camera separately, ignoring the astrophysical signal, before averaging the resulting per-camera light curves. Hereafter, we refer to this approach as PDC-LS, after the least-squares version of the PDC algorithm, which it approximates.

Panel c of Fig. 1 shows an example of the signals recovered by both algorithms. In general, both algorithms result in per-camera light curves with typical point-to-point scatter (measured from the first difference of the light curve, which preserves only short-term signals) of ~ 500 ppm, which averages down to ~ 115 ppm, for 24 cameras. This is the expected behaviour (following $\approx \sigma/\sqrt{J}$, where J is the number of cameras) and it shows that both algorithms treat the white noise component of the signal in the same way.

Short-term scatter aside, the injected astrophysical signals, such as star-spot modulation and/or transits, are visible in the corrected light curves produced by both methods, but the long-term component of the stellar signal is suppressed in the PDC-LS version, whereas it is preserved in the REPUBLIC version. To emphasize this, panel d of Fig. 1 shows the difference between the corrected, merged light curves and the true astrophysical signal. Only white noise is present in the REPUBLIC version, whereas the PDC-LS-like method has ‘injected’ a spurious long-term signal into the corrected light curve. It is also evident that the PDC-LS method exhibits minor fluctuations at 30 and 60 d, attributable to the *Kepler* trends incorporated into the light curves. These perturbations are clearly observed in Fig. 2 (they will also appear in the residuals of other figures in the paper). We only show an example in this paper, but we note that this behaviour was seen consistently across the entire set of 1000 simulated stars.

3.3 Sensitivity to number of cameras and white noise

The next step was to test how REPUBLIC performs with a smaller number of cameras. Besides the 24 cameras case described before, we also tested for 6 and 12 detectors. While stars located near the centre of the *PLATO* FOV will be observed by all 24 cameras, other parts of the FOV will be observed by only 18, 12, or 6 cameras. We found that the nominal white noise of 500 ppm averages down to ~ 220 and ~ 150 ppm for 6 and 12 cameras, respectively. This behaviour is as expected, similar to the case with 24 cameras.

The top panel of Fig. 3 shows the residuals (REPUBLIC-corrected light curve minus true astrophysical signal) for the 6, 12, and 24 camera cases for a randomly selected example light curve. We can see visually the improvement in white noise when increasing the number of cameras. In some cases, we also observed that residual systematic trends start to appear in the REPUBLIC-corrected light curves as the number of cameras decreases, although this is partly masked by the larger white noise. This behaviour is entirely in line with expectations.

We also explore the behaviour of the REPUBLIC correction when light curves contain higher levels of white noise. We reprocessed the light curves created in Section 3.1 to add white noise of 5 and 25 ppt (significantly larger than the nominal 0.5 ppt that we used before). We applied a REPUBLIC correction to each set of light curves and

we found that the noise goes down to 1.1 and 5.8 ppt for the light curves with 5 and 25 ppt noise levels, respectively. This behaviour is nominal, scaling approximately as σ/\sqrt{J} , where $J=24$. An example of this test is shown in the bottom panel of Fig. 3. Despite the increase in residual dispersion, increasing the white noise level does not affect REPUBLIC’s ability to correct the systematic trends (provided, of course, the trends themselves are known – we address that point in the next section).

We have shown then that in the ideal scenario where the systematics across all cameras are distinct, the REPUBLIC algorithm demonstrates robust performance, unaffected by either the number of cameras or variations in white noise levels. Consequently, even with a reduced number of cameras or in the presence of increased white noise, REPUBLIC’s efficacy in systematic correction remains consistently high, showcasing its potential for application in diverse observational conditions.

3.4 Breaking assumptions

Until this point, we have tested the REPUBLIC in idealized simulations, demonstrating that the algorithm functions as expected. However, several of the assumptions we have made cannot be guaranteed in practice. In particular, we assumed that the systematic trends were perfectly known and mutually uncorrelated. In a real-world scenario, the dominant trends are unknown and are typically extracted from the ensemble of light curves itself (see e.g. Smith et al. 2012). Furthermore, at least some sources of error affect the different cameras in the same way, leading to systematics that can be strongly correlated across the different cameras. In this subsection, we explore how these real-world constraints affect the performance of the REPUBLIC algorithm.

3.4.1 Unknown trends

Instead of assuming the trends were known a priori, we attempted to extract them from the sample of 1000 light curves in 24 cameras created in Section 3.1. For each camera, we perform a Principal Component Analysis (PCA) decomposition of the set of 1000 light curves (LCs), and extract the first four Principal Components to be used as estimated trends. Fig. 2 shows the recovered trends for six of the simulated cameras. As we can see, the PCA-inferred trends provide a good, but imperfect approximation to the true (*Kepler* CBV-derived) trends.

We then used these PCA-inferred trends as the T_{jkn} in the REPUBLIC algorithm to correct the light curves. Fig. 4 shows the recovered astrophysical signal using the PCA-inferred trends for a given light curve, compared with the ideal case where the trends are known. The recovered astrophysical signal retains the bulk of the intrinsic stellar and planetary signals, but the imperfect knowledge of the underlying systematic trends results in significant residual long-term trends that are not of astrophysical origin. We also performed a PDC-LS approach to correct the light curves using the PCA-inferred trends. Fig. 4 shows the recovered astrophysical signal in this case. We can see that, as in the REPUBLIC case, the imperfect trends leave a residual signal in the recovered astrophysical signal. These results highlight the importance of the trend extraction in the systematics-correction process.

It is worth noting that the light curve sample used in this test is not entirely realistic. For example, all the light curves share the same white noise level and the same parent parameter distributions for the stellar and planetary signals. Furthermore, the method employed here

to extract the trends is very simplistic compared with the procedure used in the PDC-MAP pipeline (Stumpe et al. 2012), which is also foreseen to be used for *PLATO*. The PDC-MAP trend extraction also uses PCA, but works with a sub-set of the light curves pre-selected to be mutually correlated. This pre-selection ensures that the extracted CBVs are not excessively affected by the variability of individual stars. In our sample, all the simulated light curves contain planetary signals and significant stellar variability, making the identification of a ‘low variability’ or systematic-dominated subset impossible. A more sophisticated trend extraction approach applied to a more realistic light curve sample would likely result in a better set of estimated trends, so the results in this section should be seen as a worst-case scenario.

3.4.2 Similar trends across multiple cameras

We then proceed to relax another key assumption, namely that the systematic trends are substantially different between one camera and the next. This could happen in a real-world scenario if the dominant sources of systematics affect multiple cameras at the same time, e.g. due to changes in airmass or seeing (for ground-based observations from the same site), telescope motion (for multiple cameras on the same mount), or differential kinematic aberration (for multiple cameras sharing the same FOV). We note that from a theoretical point of view, repetition of the exact same trend across multiple cameras renders the least-squares problem ill-posed and this could cause the REPUBLIC algorithm to fail. To take this scenario to its extreme, we modified the procedure described in Section 3.1.2 to simulate the trends, forcing one of the four trends to have the same behaviour by all cameras. To do so we took one of the *Kepler* CBVs and we assign as the first trend injected to each light curve. The only variation is that for each of them we added a slightly white noise of 50 ppm. This scenario represents a stringent test case, where the trends are nearly identical, simulating conditions that are, in practice, less likely due to the natural variations expected in real observations where shared trends would exhibit more significant differences. The left panel of Fig. 5 shows the trends used for this analysis for six of the simulated cameras. We can see that Trend 1 (blue line) has the same behaviour for all the cameras. We then created another set of 1000 light curves using this new set of trends and applied the REPUBLIC correction as we did in Section 3.2. The top panel of Fig. 6 shows the REPUBLIC and PDC-LS corrections to one of the light curves. We can see that both algorithms provide a light curve that resembles the real one. The general shape of the light curve is maintained, but both algorithms leave a residual trend. However, we note that REPUBLIC does better at correcting the bumps at ~ 30 and ~ 60 d.

We then tested the behaviour of REPUBLIC on the same set of light curves, but using trends extracted from the light curves themselves using PCA as described in Section 3.4.1. The resulting trends are shown in the right panel Fig. 5. As before, the PCA-recovered trends are qualitatively similar, but not identical, to the injected ones. Importantly, there are now small differences between the first trend (blue curve) in each camera, though the overall behaviour remains the same. We applied the REPUBLIC algorithm to the light curves using these new PCA-extracted CBVs and found that in this case, REPUBLIC works nominally. The results, an example of which is shown in Fig. 6, are similar to the case where the trends are different across all cameras and the trends are extracted using PCA (Section 3.4.1). We note that the identical trends scenario does not affect the PDC-LS-like algorithm, as the PDC-LS correction is performed on a camera-by-camera basis. For completeness, we show

in Fig. 6 the PDC-LS-like correction in the case in which CBVs of different cameras have identical trends.

4. CONCLUSIONS

We have presented REPUBLIC, a new algorithm that can be used to correct instrumental systematics while preserving astrophysical signals in light curves from multi-camera surveys such as *PLATO*. REPUBLIC exploits the fact that the astrophysical signal is the same across all cameras, but the systematics vary from camera to camera, to disentangle the two. Using simulated *Kepler*-like data, we showed that, in idealized conditions, when the systematic trends are known and weakly or uncorrelated, REPUBLIC outperforms a naive correction that ignores the intrinsic variability of the target star.

In particular, REPUBLIC avoids the suppression of signals on time-scales similar to the duration of the observations, which is a hallmark of standard systematic correction methods such as the widely used PDC-MAP method. Conversely, REPUBLIC is designed to preserve signals which are common to multiple cameras, so it will not remove systematics that have the exact same effect on multiple cameras, as those will be incorporated in the ‘stellar’ part of the model. Therefore, REPUBLIC can be seen as a ‘conservative’ alternative to the more ‘aggressive’ PDC-MAP correction. For this reason, REPUBLIC is being implemented in the *PLATO* L0-L1 pipeline (which will produce science-ready light curves) as an alternative to the baseline PDC-MAP algorithm. While the preservation of long-term astrophysical signals is not crucial for applications such as transit detection or asteroseismology, it will be useful for the detection of rotation periods, particularly for slower rotators, as well as the study of longer-term phenomena such as active region evolution and activity cycles.

As with any systematics-correction method, the performance of REPUBLIC depends directly on the accuracy of the trends used to model the systematics. The tests presented in this paper either assumed that the trends were known a priori, or were extracted from the light curves themselves using PCA. A possible extension of the REPUBLIC algorithm would be to solve for the systematic trends at the same time as for the astrophysical signal, over the entire ensemble of light curves at once. This may seem unrealistic, but the mathematical structure and scale of the resulting problem would not be dissimilar to that of the large linear models used in numerous machine learning applications.

The performance of REPUBLIC also depends critically on the degree of co-linearity between the systematic trends affecting the different cameras. If those are too similar, the algorithm becomes numerically unstable and the correction fails. In tests where only one of four trends was shared across all cameras, we found that using PCA to extract the trends from the light curves in each camera alleviates this problem significantly: it introduces subtle differences between the extracted trends in different cameras that break the degeneracy.

DATA AVAILABILITY

The code used to perform the tests and figures in this manuscript is available at https://github.com/oscaribv/REPUBLIC_paper_code.

ACKNOWLEDGEMENTS

OB and SA acknowledge support from the UK Science and Technology Facilities Council (STFC) under grants ST/S000488/1 and ST/R004846/1. This work made use of NUMPY (Harris et al. 2020), MATPLOTLIB (Hunter 2007), SEABORN (Waskom 2021), PANDAS (Pandas Development Team 2020), SKLEARN (Pedregosa et al. 2011), SCIPY (Virtanen et al. 2020), and PYTRANSIT (Parviainen 2015) libraries. This work presents results from the European Space Agency (ESA) space mission *PLATO*. The *PLATO* payload, the *PLATO* Ground Segment, and *PLATO* data processing are joint developments of ESA and the *PLATO* Mission Consortium (PMC). Funding for the PMC is provided at national levels, in particular by countries participating in the *PLATO* Multilateral Agreement (Austria, Belgium, Czech Republic, Denmark, France, Germany, Italy, the Netherlands, Portugal, Spain, Sweden, Switzerland, Norway, and United Kingdom) and institutions from Brazil. Members of the *PLATO* Consortium can be found at <https://platomission.com/>. The ESA *PLATO* mission website is <https://www.cosmos.esa.int/plato>. We thank the teams working for *PLATO* for all their work. OB acknowledges Dr. Baptiste Klein and Dr. Annabella Meech for their supportive words that motivated the author to finally complete this paper.

REFERENCES

- Akeson R. L. et al., 2013, *PASP*, 125, 989
 Barragán O., Aigrain S., Rajpaul V. M., Zicher N., 2022, *MNRAS*, 509, 866
 Borucki W. J. et al., 2010, *ApJ*, 713, L126
 Harris C. R. et al., 2020, *Nature*, 585, 357
 Howell S. B. et al., 2014, *PASP*, 126, 398
 Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
 Janssen N. et al., 2024, *A&A*, 681, A18
 Lightkurve Collaboration, 2018, Astrophysics Source Code Library, record ascl:1812.013
 Lu T.-T., Shiou S.-H., 2002, *Comput. Math. Appl.*, 43, 119
 Marcos-Arenal P. et al., 2014, *A&A*, 566, A92
 Pandas Development Team, 2020, pandas-dev/pandas: Pandas version: 2.2.1. Zenodo, available at: <https://zenodo.org/records/10697587>
 Parviainen H., 2015, *MNRAS*, 450, 3233
 Pedregosa F. et al., 2011, *J. Mach. Learn. Res.*, 12, 2825
 Rauer H. et al., 2014, *Exp. Astron.*, 38, 249
 Ricker G. R. et al., 2015, *J. Astron. Tel. Instr. Syst.*, 1, 014003
 Samadi R. et al., 2019, *A&A*, 624, A117
 Smith J. C. et al., 2012, *PASP*, 124, 1000
 Smith A. M. S. et al., 2020, *Astron. Nachr.*, 341, 273
 Stumpe M. C. et al., 2012, *PASP*, 124, 985
 Tamuz O., Mazeh T., Zucker S., 2005, *MNRAS*, 356, 1466
 Twicken J. D., Chandrasekaran H., Jenkins J. M., Gunter J. P., Girouard F., Klaus T. C., 2010, Proc. SPIE Conf. Ser. Vol. 7740, Software and Cyberinfrastructure for Astronomy. SPIE, Bellingham, p. 77401U
 Virtanen P. et al., 2020, *Nature Methods*, 17, 261
 Waskom M. L., 2021, *J. Open Source Softw.*, 6, 3021
 Wheatley P. J. et al., 2018, *MNRAS*, 475, 4476

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.