Quaver User Guide

Krista Lynne Smith

November 2022

Introduction

The Quaver pipeline has been developed to provide a flexible, customizable approach to sources that are likely to be faint and/or have stochastic variability for which an *a priori* model may not exist. It has been designed with active galactic nuclei (AGN) in mind, but is useful for any source that may not be well-suited to pipelines customized for asteroseismology or exoplanet applications for the above reasons.

Software Version Requirements

Python version: 3.9
Lightkurve version: 2.3 (https://docs.lightkurve.org/)
Astropy version: 5.2.1 (https://www.astropy.org/)
Astroquery version: 0.4.6 (https://astroquery.readthedocs.io/en/latest/)

Quaver also requires numpy and matplotlib.

How does it work?

Quaver uses principal component analysis (PCA) to remove spacecraft systematics, variability from contaminating sources, and scattered background light. However, in contrast to simple PCA commands, Quaver utilizes a "hybrid" approach, which analyzes faint background pixels to model additive effects like scattered light separately from bright pixels containing other sources, which contribute multiplicative effects due to both spacecraft systematics and source contamination. The pipeline uses several methods from the `lightkurve` package [1], especially the tasks for regression correcting and building and maintaining design matrices. It also makes use of the TESSCut software, which extracts a small "postage stamp" around the source of interest from the entire TESS Full Frame Image (FFI), which is formidably large otherwise.

[1] https://docs.lightkurve.org/

Before either method is used on a source, the user is allowed to interactively select their extraction aperture, guided by an overlay of the DSS contours of the region in the postage stamp. The user is then allowed to optionally "mask out," or exclude from analysis, any cadences of the light curve for which there are very large systematics.

After these choices are made, Quaver uses one of three methods to reduce the light curve: a simple method using PCA without any hybridization, a hybrid approach in which the background effects are treated with simple background subtraction but are still treated separately from bright pixels and used to correct them, and a "full-hybrid" method that handles both background additive and multiplicative systematics using regression matrices. The user may determine from which method they want to receive the output. As with all methods of systematic removal, the correct choice depends upon the nature of the target source and the science question being addressed. A simple PCA fit very often will over-fit the data, if too many principal components are used, or if the background behavior in any way mimics the behavior of the science target. Pros and cons of each method are discussed in the section below on Detailed Fitting Methods.

There are three outputs of each run of the program: 1) a light curve .dat file containing the time stamp (in TESS Julian Dates, TJD), flux, and flux error, 2) a sector-specific multi-panel image of the corrected light curve, the components used in the regression, and the chosen aperture, saved as a .pdf file, and 3) a fully-stitched light curve of all sectors of the source. These output files are stored in a directory created by Quaver, `quaver_output`, in a sub-directory named after your target.

Tunable Parameters

There some important tunable parameters in Quaver that should be set before you begin reduction. All of them have a default value, and some are more important to consider than others. They can be found at the top of the program code.

- `systematics_correction_method`: The choice of the reduction method for which you want output light curves and figures. The first option, 1, is to perform the full hybrid version of the fit. The second option, 2, is a simple regression matrix fit using principal components. Both methods are described further in the next section.

- `tpf_width_height`: The dimensions of the TESSCut cutout you wish to download. Larger postage stamps will require more processing time, and are more likely to include sources with strong variability that will affect your source's reduction. However, a postage stamp that is too small will not allow for sufficient sampling of the background or lower-order effects in brighter pixels.

- `additive_pca_num, multiplicative_pca_num, pca_only_num`: The number of principle components you wish to use in each of the methods. The hybrid method is set to a default of 3 principle components each to capture the additive and multiplicative effects. The Simple PCA method uses 3 total components. You may adjust these as you like; be warned, however, that using too many components significantly increases the chances that you will overfit the data. Take a look at some of the examples below before choosing. You can always re-run the pipeline with a different number of PCAs if you see questionable effects in the additive or multiplicative components.

- `lowest_dss_contour`: The lowest flux contour, in multiples of the maximum flux in the image, for the overlay of the DSS image on the aperture selection plot. If a very bright star is in the field, for example, the default value of 0.4 may be too high for you to see the DSS contour of your faint target source. Conversely, in a very crowded field of only faint sources, too many contours may be visible corresponding to noise in the DSS image.

- `sys_threshold`: The threshold for determining whether "major" systematics remain in the light curve. The default value, 0.2, is arbitrary but captures the majority of spike-type artifacts and large thermal ramps at the beginnings and ends of the TESS orbits. Making this more stringent will cause Quaver to ask you more often whether you want to mask any regions; and vice versa.

- `max_masked_regions`: The maximum number of masked regions allowed. This is the number of times that Quaver will ask you to mask out regions of the light curve that it believes are subject to major systematics (i.e., that exceed the threshold mentioned in the previous item). In practice, there are rarely more than 3 such regions, and over-masking does remove data from your final result.

- `plot_index`: The index of the cadence used to display the aperture selection panel. Each sector has approximately ~ 1200 cadences. Those near the beginning, middle, and end of a sector are generally bad choices for this, as these may be during periods of thermal renormalization or when systematics are highly prevalent. The default is 200, but if you see that the aperture selection image seems to be blank or otherwise not sensible, it is possible that this cadence corresponds to a bad timestamp, so it is a good idea to try a few others before giving up on an object. Sometimes, an object is listed as in the TESS FOV when it is actually not quite on the detector, and in this case, nothing can be done. If this is the case, you will see no change in the aperture selection panel at different cadence indices.

- `bf_threshold`: The definition of the bright-faint threshold, in multiples of sigma times the median of the flux across the entire TPF, that divides the definition of bright and faint pixels in the calculation of principal components. Faint pixels are used to determine the additive scattered light background effects, while bright pixels are used to determine higher-order systematics (after being corrected for the background). The default is 1.5σ, but the correct choice will depend on whether there are any very bright (or only many, very faint) sources in your TPF.

Details of the Fitting Methods

Each of the three options for fitting methods are described in this section. The workflow of the program for each method is shown graphically in Figure 1.

All three methods retain the aperture selection and cadence-masking that make Quaver so flexible, but the user should determine which method is most appropriate for their science case. All three methods do a similar job of reproducing high-frequency power spectral slopes in Seyfert galaxies, but the treatment of the low-frequency behavior and the robustness against some high-frequency multiplicative systematics is different from method to method.

Simple PCA

The PCA method simply calculates the variability principal components for *all* pixels in the target pixel file (TPF) that are not in your chosen aperture. The number of principal components used is set by the tunable parameter `pca_only_num`.

These principal components are stored in a design matrix, and the `lightkurve` class `RegressionCorrector` is used to linearly regress the light curve extracted from the custom aperture against the principal components.

This is the simplest version of reduction. It has several drawbacks: if any of the targets in the field have variability that mimics your source, this method is extremely likely to over-fit your source. It is also less likely to capture subtle (typically electronic) systematics that can affect very faint targets, and can introduce spurious periodicities.

Simple Hybrid Method

Quaver's hybrid methods model the dominant additive effects, such as the scattered light from the Moon and the Sun, that affect TESS light curves, separately from the subtler multiplicative effects due to electronics, some types of contamination from other varying sources, or other systematics. Both methods impose a 1-pixel buffer zone around your chosen aperture that is excluded from all fitting, to minimize the chance that leakage of real variability from your source sneaks into the fit vectors and is removed.

Both methods utilize the `lightkurve` class, `RegressionCorrector`, which removes trends using linear regression against a matrix of the principal components describing non-source variability.

Quaver Workflow

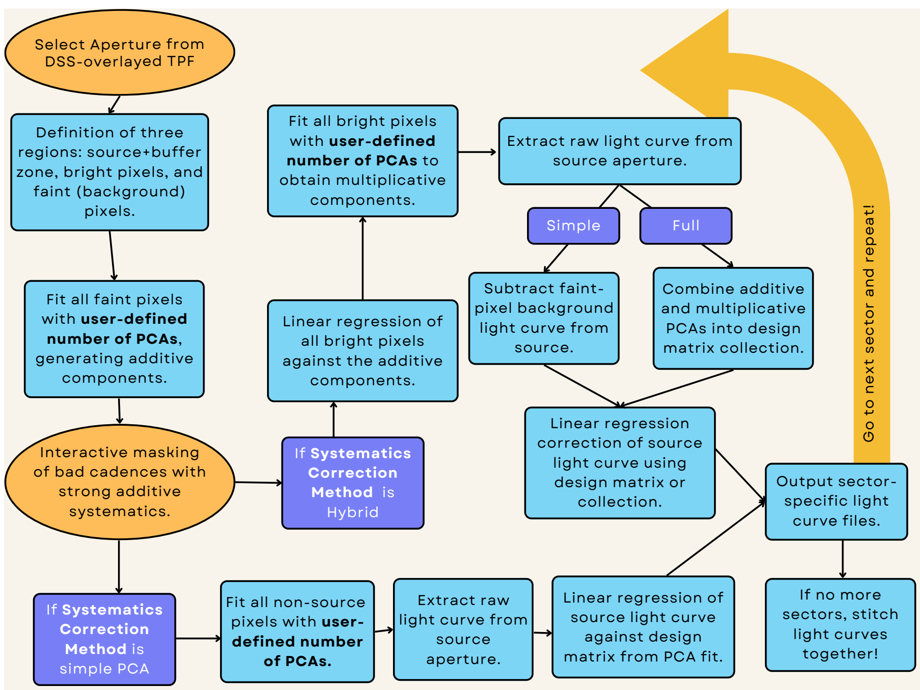


Figure 1: Flowchart depicting the Quaver method. Round, orange elements are interactive steps. Text in bold indicates an option chosen by the user.

Quaver first imposes a flux threshold of 1.5σ above the median, and categorizes all pixels below this threshold as faint background pixels, and all pixels above as bright pixels potentially containing a source. It creates a design matrix describing the background (additive) systematics using PCA of all the faint pixels, with the number of components given as a tunable parameter by the user (see previous section).

With the additive design matrix in hand, Quaver linearly regresses all of the bright source pixel light curves against this matrix, correcting them for the additive effects.

Quaver then computes the *multiplicative* effects using PCA on the additively-corrected bright pixels, again using the number of components specified by the user. The multiplicative correction matrix has now been constructed, and is ready to be used on the source light curve.

Next, Quaver extracts the raw light curve from the user-defined aperture. To handle the background light, it computes a generic background light curve from the same faint pixels used to define additive effects, scales this background light curve to an appropriate level to the number of pixels in the extraction aperture, and performs simple background subtraction. The source light curve, now background-subtracted, is regression-corrected for multiplicative effects using the design matrix derived from the additively-corrected bright pixels.

This method provides robust correction of insidious multiplicative effects that can be much less prominent than the scattered light backgrounds, while avoiding over-fitting of long-term variability by regression against additive trends, as evidenced by very good agreement between Quaver-extracted TESS light curves and simultaneous ground-based data. It is more thorough than the simple PCA method, and produces light curves with variability amplitudes in excellent agreement with simultaneous (but more sparsely-sampled) ground-based light curves. However, because the additive background is handled only with simple subtraction, it is not as aggressive at removing long-term systematics as the full hybrid method, described next.

Full Hybrid Method

This method is the same as the Simple Hybrid method up to the point where the bright pixels are corrected by the faint systematics. In the previous method, the background effects are removed by simple scaled subtraction, before the multiplicative effects derived from the bright pixels are corrected using their design matrix.

In this method, the same additive design matrix used to correct the bright pixels is *combined* with the multiplicative design matrix, to create a `DesignMatrixCollection` object. The `RegressionCorrector` class then uses this more complex object as the linear regressor, upon the source pixels.

This is the most aggressive method available in Quaver. It is well-suited to rapidly-varying sources, in which the science case is mainly concerned with high-frequency phenomena or properties. It does the most thorough job of handling all types of systematics, but is prone to over-fitting low-frequency or long-term effects (although it does not always over-fit).

Steps for Using Quaver

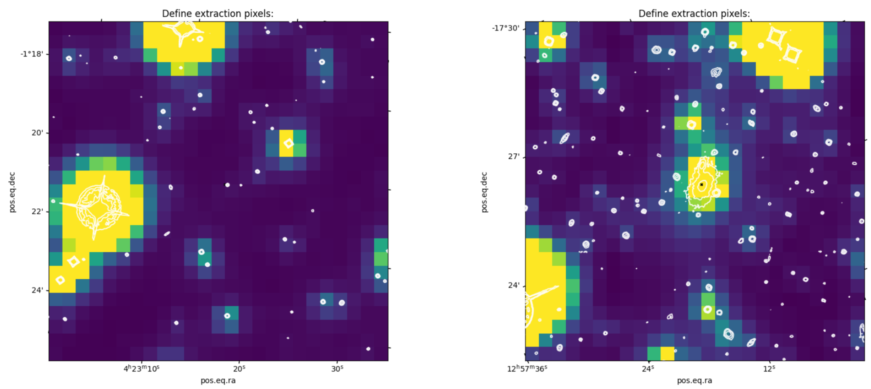
1. By default, Quaver will ask you for the common name of your source. Here you can type something like "3C 120" or "TYC871-819-1". If you prefer to use coordinates, you may hit enter to skip this and prompt the query for coordinates. Enter the right ascension and declination in decimal degrees, separated by a comma without a space. If you use a common name to search, this name will automatically be used to title your output files. If you use coordinates to search, Quaver will ask you for the desired name of your object for the output files.

2. Once your source has been resolved, Quaver will display a list of Sectors in which your object was observed by TESS. If your object has never been observed by TESS, you will be told. Following the list is a short guide of which Sectors belong to which Cycle. Quaver asks you to choose which Cycle you wish to work on. If your observation has multiple Sectors in the chosen Cycle, you will be able to work through all of them in a single Quaver run. For different Cycles, you will need to re-run Quaver; this is good practice in any case, as observations in different Cycles are very far apart in time.

3. After you have entered your Cycle, Quaver will display the extraction aperture selection panel. This tool is helpful for both faint sources, which may be hard to see without the guidance of the overlaid Digital Sky Survey image, bright extended sources with complex morphologies, or especially crowded fields. Examples of aperture selection panels for some of these cases are shown in Figure 2. The black "x" denotes the center of the TESS postage stamp, which your source should be very near. Select your aperture by clicking on the desired pixels. A green check mark will appear on selected pixels. If you wish to deselect a pixel, a red "x" will appear on the pixel, indicating it has been removed from the list. When you are finished constructing your extraction aperture, close the window.

   Note: sometimes, lightkurve believes a source was monitored by TESS when in fact, it was off-silicon or very close to the detector edge. If you encounter a sector map which resembles static, or a solid color, this is most likely the case. Simply close the aperture selection window without making any selections, and the sector will be skipped without disrupting the process.

4. Quaver will now begin extracting and correcting your object's light curve. Often, a light curve will experience strong spacecraft systematics in small time regions, which appear as spikes or large "swoops" in the additive components. If Quaver detects any of these, it will prompt you to decide



(a) Aperture selection panel showing the TESSCut postage stamp and overlaid white contours from the DSS, for a faint and fairly isolated source.

(b) Aperture selection panel showing the TESSCut postage stamp and overlaid white contours from the DSS, for an extended source in a crowded field.

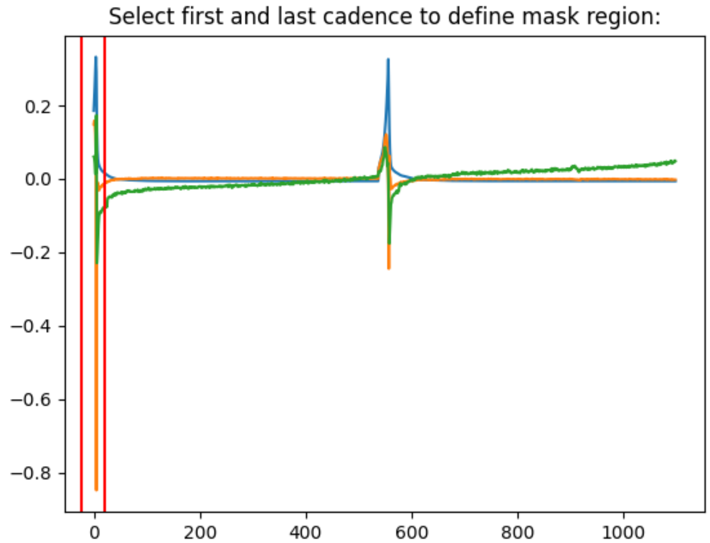
Figure 2

whether you wish to manually mask out these regions. Choose Yes or No (Y or N).

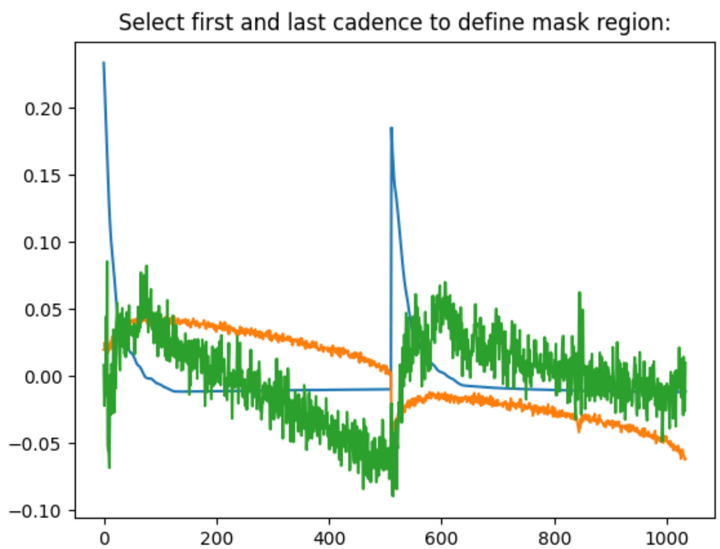
5. If you choose to mask out the affected cadences, the interactive cadence masking window will appear. To mask a region, click once to the left of the region you wish to mask, and once to the right. Red bars will appear indicating the bounds of your selected region. Only choose one region first, then close the window. Quaver will re-fit the additive components, and if significant systematics remain, will display another window to allow you to mask another region. Proceed as before. This will continue until no further systematics above the default threshold remain, or until the maximum allowed number of masked regions is reached. Both of these parameters may be tuned in the Tunable Parameters module at the top of the program. At any point, you may choose to not mask any regions in the window that you see, and close the window without choosing anything.

   Obviously, it is desirable to mask out as few cadences as possible, to preserve the majority of the light curve for analysis. Generally, the only cadences that need to be masked are those displaying very high-amplitude, narrow "spikey" systematics that persist for very few cadences, like both of the outlying regions shown in Figure 3a. Gradual or "swoopy" systematics, like those pictured in Figure 3b, are well-handled by both the hybrid and simple methods, and generally can be left alone.

6. Once you have closed the final cadence masking window, Quaver will finish correcting your light curve. It will display a panel showing the corrected



(a) Sharp, high-amplitude systematics that are likely to persist in the final light curve and should be masked before proceeding with the reduction.



(b) Systematics with gradual or smooth components that are handled well by the mitigation in both reduction methods, and can usually safely be left alone.

Figure 3

light curve, the components in the design matrices, and the extraction aperture. It will save this image, in a new folder titled using the name of your source. Once you close this panel, if the source has more Sectors in your chosen Cycle, you will be shown the aperture selection panel for the next Sector. If you do not want to extract a given Sector, simply close the aperture selection panel without choosing any pixels.

7. If the object is monitored for more than one sector per chosen cycle, the final step is to stitch those light curves together. At present, Quaver uses the arithmetic mean of the fluxes of the final ten cadences of one sector and the first ten cadences of the next sector to compute an additive offset, then additively scales the sectors to match. In future versions of Quaver, more sophisticated mechanisms involving modeling long-term behavior maybe incorporated, but at present, testing with these methods unquestionably introduces spurious variability.

Example Objects

- PKS 0035-252

  In this section we explore how to reduce systematics in a faint source. Because the source is so faint, one is tempted to use a small aperture to reduce background noise in the final light curve. However, in this case, if one uses the smallest feasible aperture and leaves the default (3) number of principle components to be used in the hybrid method, significant overfitting occurs, as shown in Figure 4.

  Notice that the third (green) additive component is identical to the source variability. This means that some faint trace of the source variability is still present in the faint pixels used to calculate the additive components.

  Figure 5 shows the Quaver output when a larger aperture is used, which fully excludes all source flux from the faint background that is used to compute the additive components. In this case, the hybrid method and the simple PCA method both work well.

  A good match between the two methods, and avoidance of overfitting, is also achieved if the smaller aperture is used, but with fewer (2) principal components in the PCA of the additive background in the hybrid method, as shown in Figure 6.

  Given this example, it may seem that using the simple PCA method is best. However, this is not true in the majority of cases. The use of simple PCAs can very often result in overfitting, especially when the source is in a crowded region or a region containing several other strongly-varying sources, as the next example shows.

- PKS 0420-01

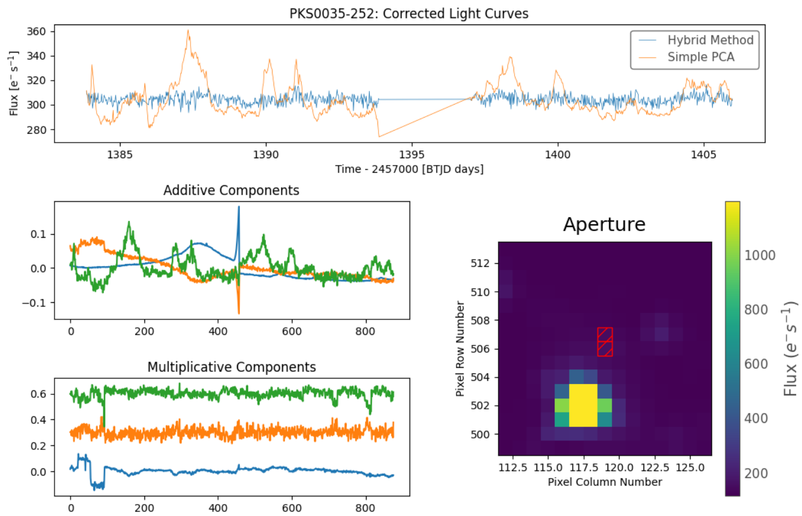


Figure 4: Quaver output when the smallest aperture is used for a very faint source.

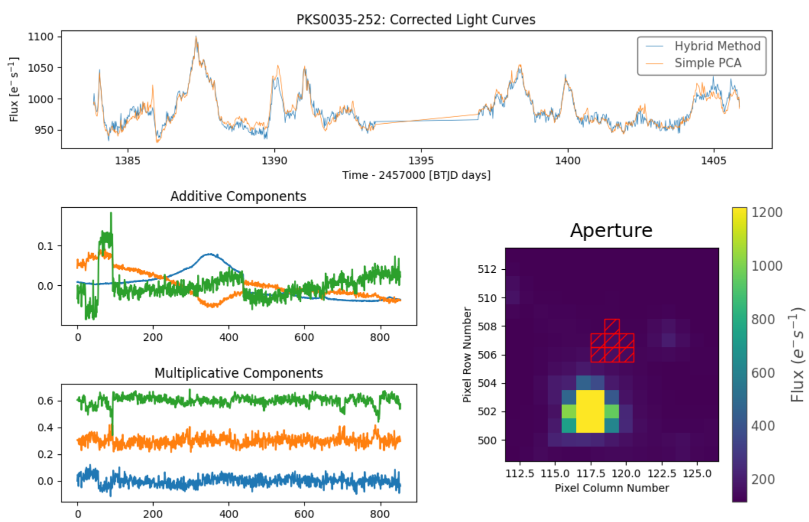


Figure 5: Quaver output when a larger aperture is used for the same source, postage stamp dimensions, and PCA component numbers as in Figure 5.

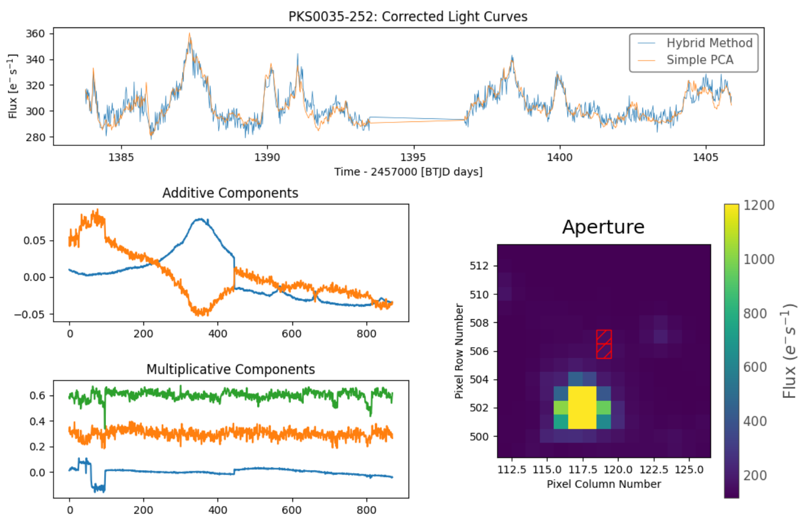


Figure 6: Quaver output when the same small aperture is used for the same source and postage stamp dimensions as Figure 4, but with one fewer principal component allowed for the hybrid additive method.

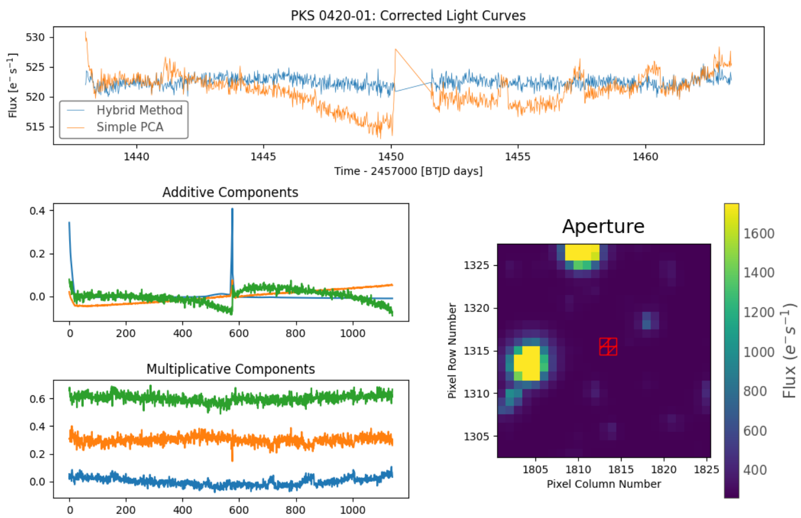


Figure 7: Quaver output for PKS 0420-01 with the default 25×25 pixel aperture.

In this example, we consider PKS 0420-01, a faint blazar in the vicinity of two very bright sources. In the image below, it appears that the hybrid method may have overfit the source, and the simple method is again showing true variations.

The hybrid method has identified the behaviors that we see in the simple light curve, and removed them. In this case, it turns out the bright star nearby (HD 27779) exhibits exactly the behavior that we see persisting in the simple PCA version of the quasar's light curve. This is evident in Figure 8.

In this case, which is very common, the simple method has failed to identify a major contribution to the (faint!) source's variability from contamination due to a very bright nearby source. The hybrid method, which treats the bright sources and faint background separately, is able to do this successfully. When we reduce the size of the cutout to 14x14 pixels, to remove the bright star from most of the image, we see better agreement between the methods, and can now see in both methods some real variability from our faint source, which the hybrid method had correctly reduced all along. This is demonstrated in Figure 9.

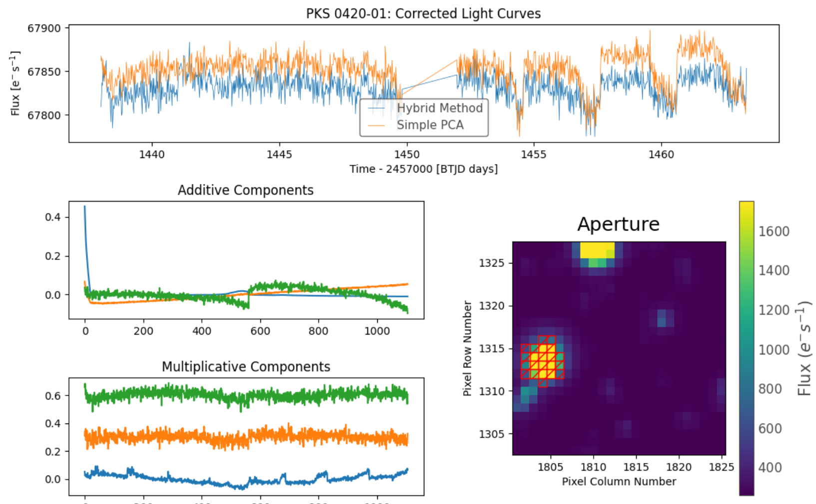


Figure 8: Quaver output when the bright star is selected as the extraction source in the postage stamp containing PKS 0420-01.

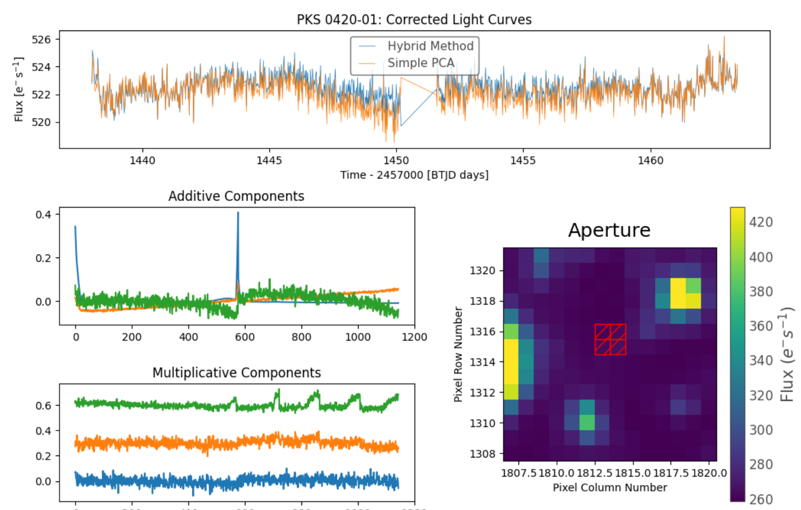


Figure 9: Quaver output for PKS 0420-01 when the postage stamp size is reduced to exclude a nearby source that was contaminating the simple PCA method in the previous figure.