

# Preparation of light curves for rotation and activity analysis

Antonino F. Lanza

*WP 123 000*

*INAF-Catania Observatory, Italy*

[antonino.lanza@inaf.it](mailto:antonino.lanza@inaf.it)

I am indebted to several Colleagues for a better understanding of several instrument and data issues  
(in alphabetical order):

S. Aigrain, F. Baudin, K. Belkacem, D. Brown, W. Chaplin, R. Garcia, A. Moya, R. Samadi, B. Stelzer.

# PLATO Data Products (a short summary)

- **Level-0 (L0):** Validated light curves of selected targets for all individual telescopes; *no instrument corrections except than those already applied on-board*;
- **Level-1 (L1):** Calibrated light curves for each target, *corrected for instrumental effects* (temperature sensitivity, jitter, differential aberration, pixel sensitivity dropout, PSF variations, and so on). The L1 light curves of each target are averaged over the normal telescopes and their uncertainty is provided;
- **Level-2 (L2) for WP 123000:** Data Products 4 (rotation and activity) derived from L1 light curves.

# Preparation of light curves for WP 123 analysis

- WP 123 is interested in a *wide range of periodicities* from minutes (convection and flares) to years (activity cycles);
- This poses *formidable problems* in the preservation of the true variability of the observed stars;
- The range of variability timescales required for asteroseismic or transit analyses is much more restricted;
- Looking at previous space experiments (e.g., CoRoT, Kepler), different approaches have been proposed:
  - Try to preserve all the periodicities, independently of their timescale (very challenging); this may come at the cost of leaving residual instrumental effects in the time series;
  - Optimize transit detection not caring about periodicities on timescale longer than 10-20 days (e.g., Kepler PDC ms-MAP pipeline; see plots by Gilliland et al. (2015) below); this does not affect the detection of transits and p-modes, but it is problem for us in WP123.

# What we want to remove before WP 123 analysis

We identify three main kinds of variations to be removed *after correcting instrumental effects in LO light curves* (e.g., sudden sensitivity dropouts in some pixels, cosmic ray features, jitter effects, relativistic aberration, outliers, etc.):

- Discontinuities produced by astrophysical phenomena (transits and flares); [*flare catalogue to be kept as an ADP; see B. Stelzer's presentation in Session 7*];
- Residual instrumental effects.
  - ❖ Instrumental effects can be classified into two broad categories:
    - a) Those correlated with some measured spacecraft or focal plane parameters (attitude, temperature, etc.); these can be removed by implementing, e.g., the EPD approach (Bakos et al. 2010) that fits an ensemble of basis functions built from the time series of the parameters; => **Warning:** consider the unwanted possibility of fitting some component of the true astrophysical variability;
    - b) Those that are not correlated with the above parameters (*residual instrumental effects*) and appear as trends common to ensembles of stars with similar characteristics (focal plane position, magnitude, colour, amplitude of variability). [see Grolleau et al. 2018, PLATO-LESIA-PDC-RP-023, Sect. 3.2.5];
- Data gaps.

# Detection of discontinuities and correction (I)

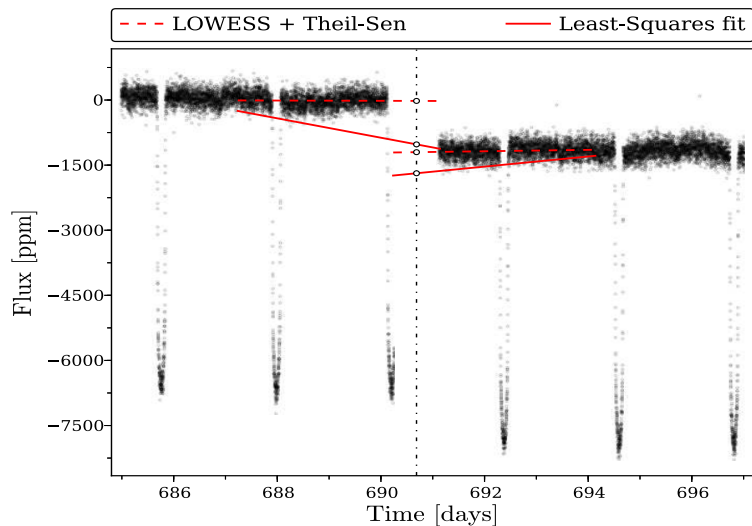
- In Corot, (Ollivier et al. 2016, CoRoT Legacy Book p. 41) an algorithm was devised based on the comparison of the mean normalized flux before and after each point  $i$ :

$$\Delta(i) = \frac{\overline{\text{flux}}_{\text{after } i} - \overline{\text{flux}}_{\text{before } i}}{\sqrt{\sigma_{\text{before } i}^2 + \sigma_{\text{after } i}^2}}$$

$$\Delta(i) > 7.2 \sigma_0$$

Where:

$$\sigma_0 = \text{stddev}(\Delta)$$



Treatment of jumps in the KASOC filter:

Using the BIC, we compare different models (no correction, constant offset, linear trends) to describe a segment centred on the jump (typically of 6 days length);

Jump correction through estimation of the amplitude from a robust linear fitting to segments before and after the jump (Handberg & Lund 2014).

# Kepler algorithm to find discontinuities

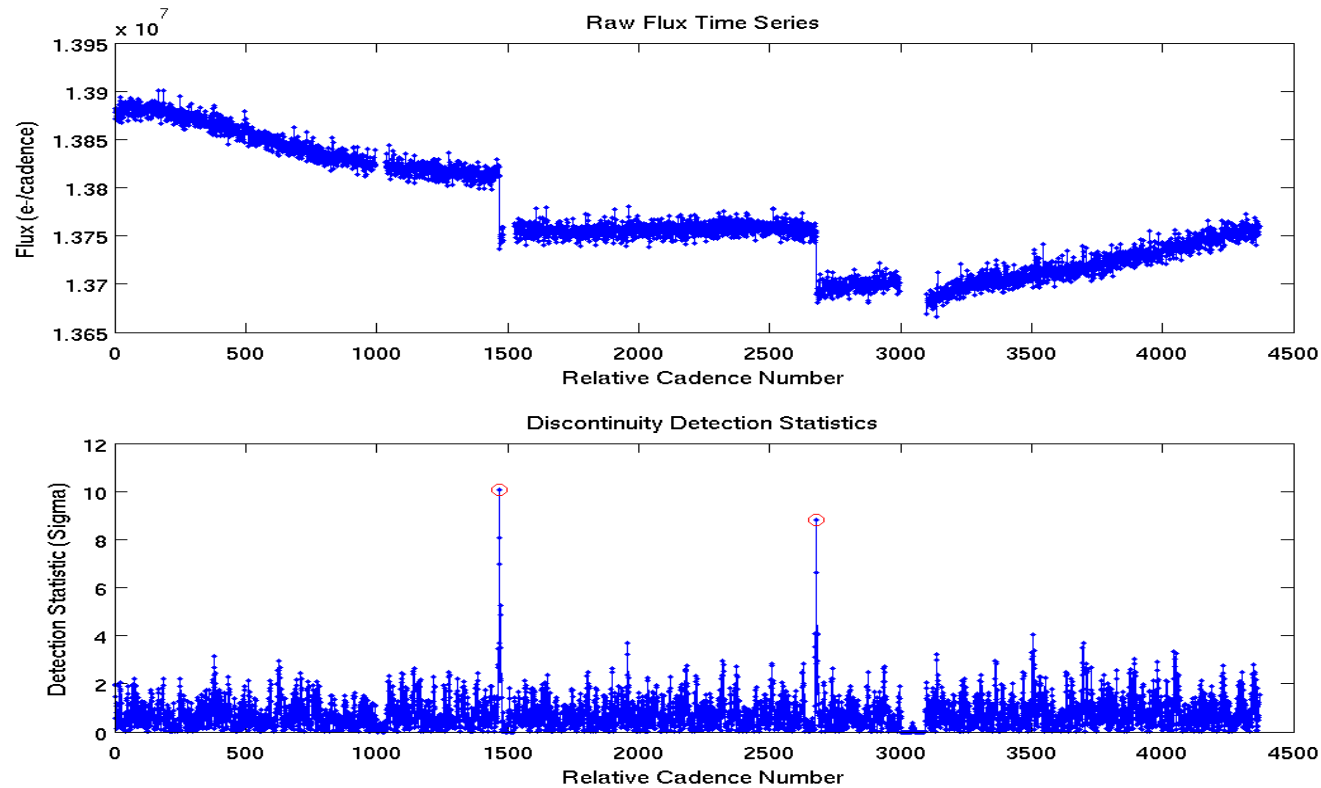
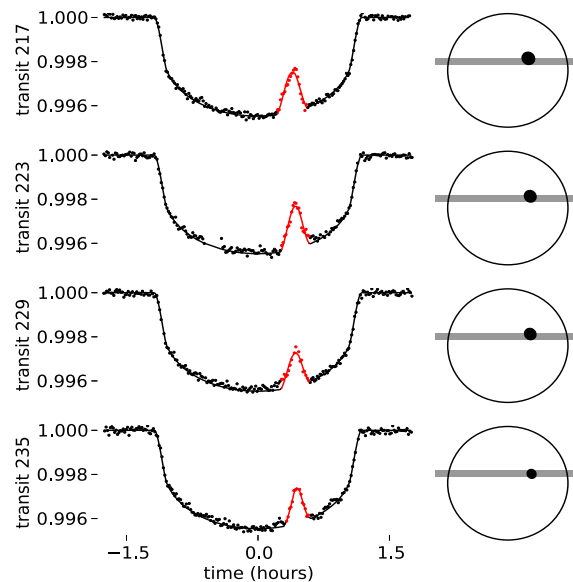


Figure 3: A raw flux time series with two random discontinuities is shown in the upper panel. The detection statistics formulated by correlating the second derivative of the time series with the discontinuity template are plotted in the lower panel. The detection statistics are circled for the cadence indices returned by the detector.

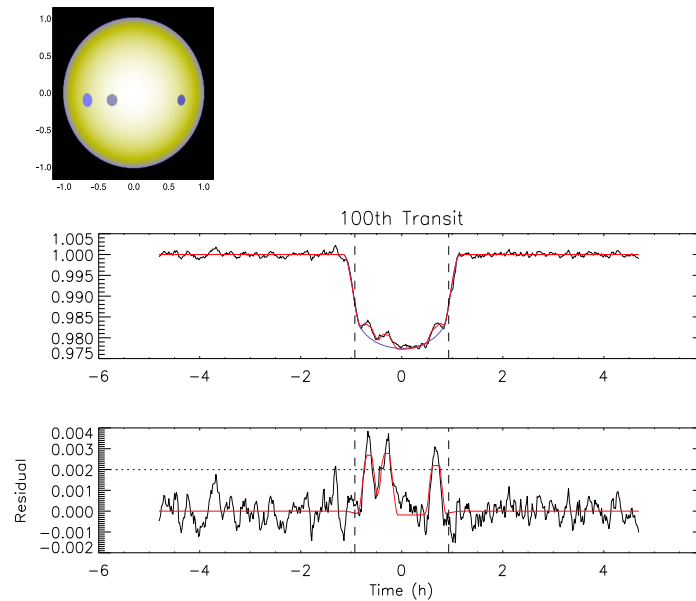
After Twicken et al. (2010). Derivatives of the photometric time series are computed numerically with Savitzky-Golay filtering.

# Detection of discontinuities and corrections (II)

- The detection of planetary transits and eclipsing binaries will be performed by specific modules of the pipeline outside of WP 123;
- We shall use the provided transit models to correct points inside the transits [an *interface* should be established];
- **Warning:** systematic residuals are present in the case of active stars, due to the occultation of starspots. These will be modelled by WP 123 600 (coordinated by A. Valio) [*Catalogue of occulted spots to be kept as an ADP*].



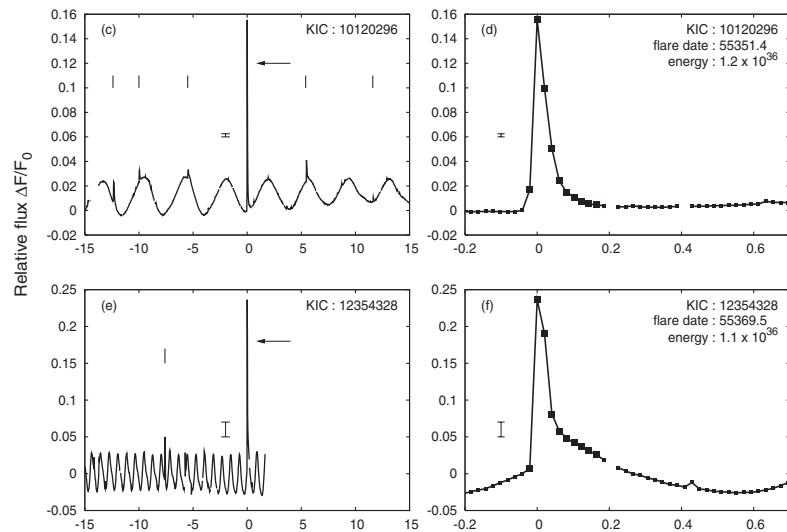
HAT-P-11 (Beky et al. 2014)



Kepler-17 (Valio et al. 2017)

# Detection of discontinuities and corrections (III)

- Stellar flares (see talk by Beate Stelzer in Session 7 on Thursday afternoon);
- Can be fitted with an appropriate model and removed [a WG was recently established to this purpose];
- Important for young F and G stars and for young and old K and M stars on the main sequence.

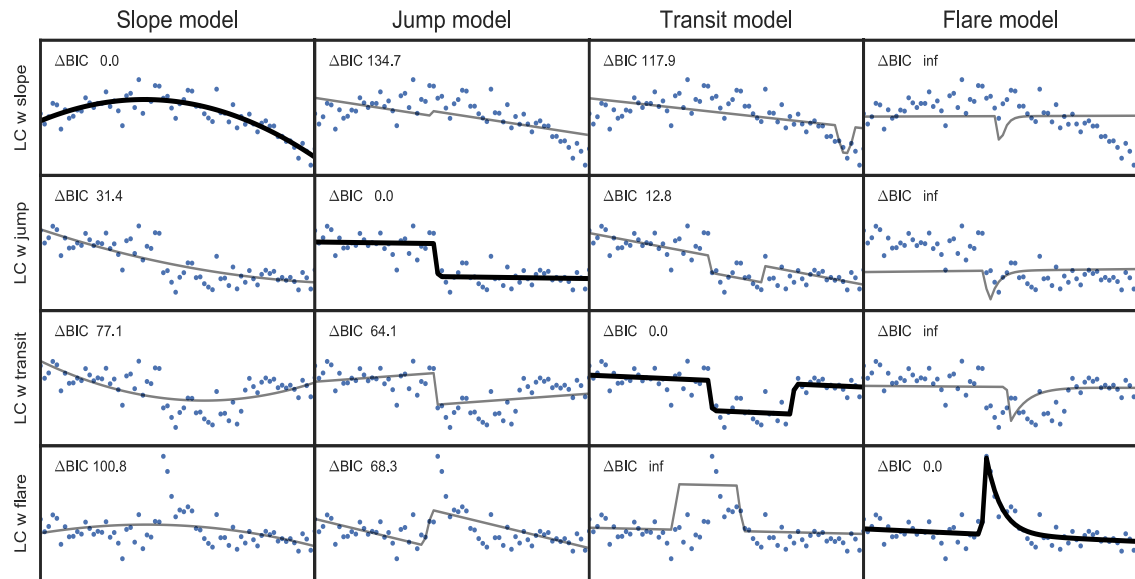
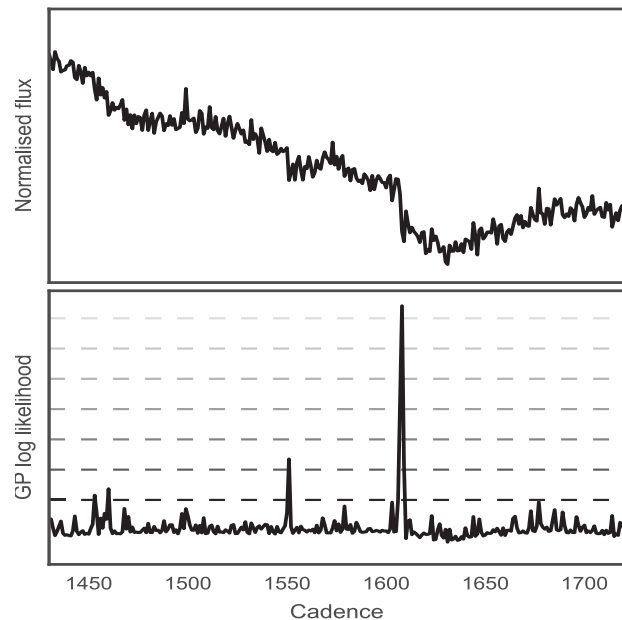


Examples of two Kepler superflare stars (Shibayama et al. 2013)



# Detection of discontinuities and correction (iv)

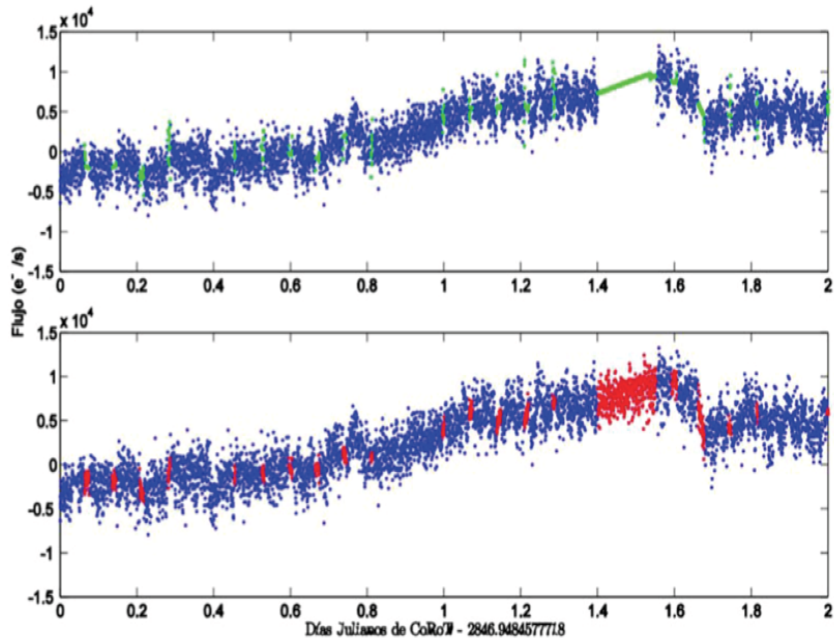
- The proposed standard methods to detect discontinuities (see above) work well in white noise and provided that there are few data gaps;
- Aigrain et al. (2017) proposed an alternative method based on:
  - a) Gaussian Process (GP) regression with a change-point built into the covariance matrix; the discontinuity detection is based on a moving-window likelihood-ratio test between with- and without-discontinuity GP models.



BIC-based model selection for discontinuities (Aigrain et al. 2017)

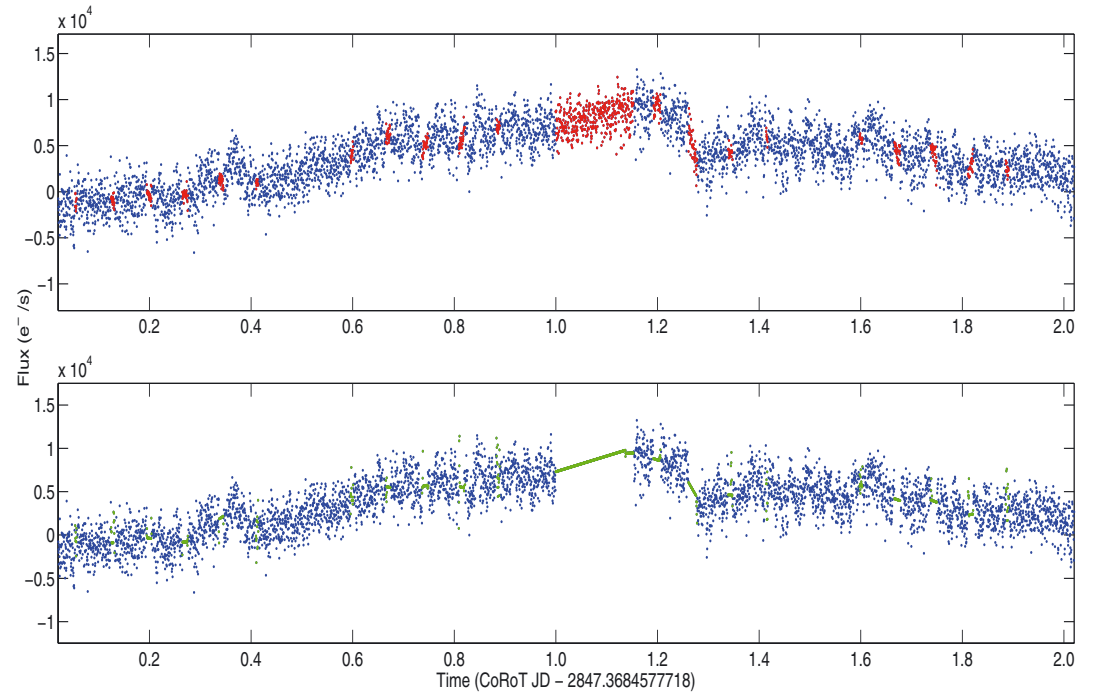
# Gap filling

- Data gaps must be filled for the computation of the power spectrum or the autocorrelation function;
- Short and long gaps are usually treated differently (see Ollivier et al. 2016 for CoRoT or Twicken et al. 2010 for Kepler);
- Some algorithms (e.g., Lomb-Scargle, spot models) can work also in the presence of gaps, but show systematic effects due to them, especially when the gaps are long;
- Therefore, gaps filling algorithms are important;
- Several methods have been proposed, for example:
  - a) Inpainting by Pires et al. (2015), based on the assumption that the light curves consists of a discrete set of generalized Fourier components;
  - b) Interpolation by auto-regressive and moving averages (e.g., MIARMA, Pascual-Granado et al. 2015).
- For our purposes, it will be very useful to analyse light curves filled with different methods to see which methods may produce systematic effects;
- A WG coordinated by S. Aigrain has been created to deal with light curve stitching.



**Fig. II.2.11.** Example of gap filling, using the Inpainting method (Pires et al. 2015).

After Ollivier et al. (2016), showing an example of the Inpainting method by Pires et al. (2015)



**Fig.6.** Comparison between ARMA interpolation (*upper panel* in red) and linear interpolation (*lower panel* in green) for the light curve of HD 49933 observed by the CoRoT satellite. Valid data are depicted in blue (both panels).

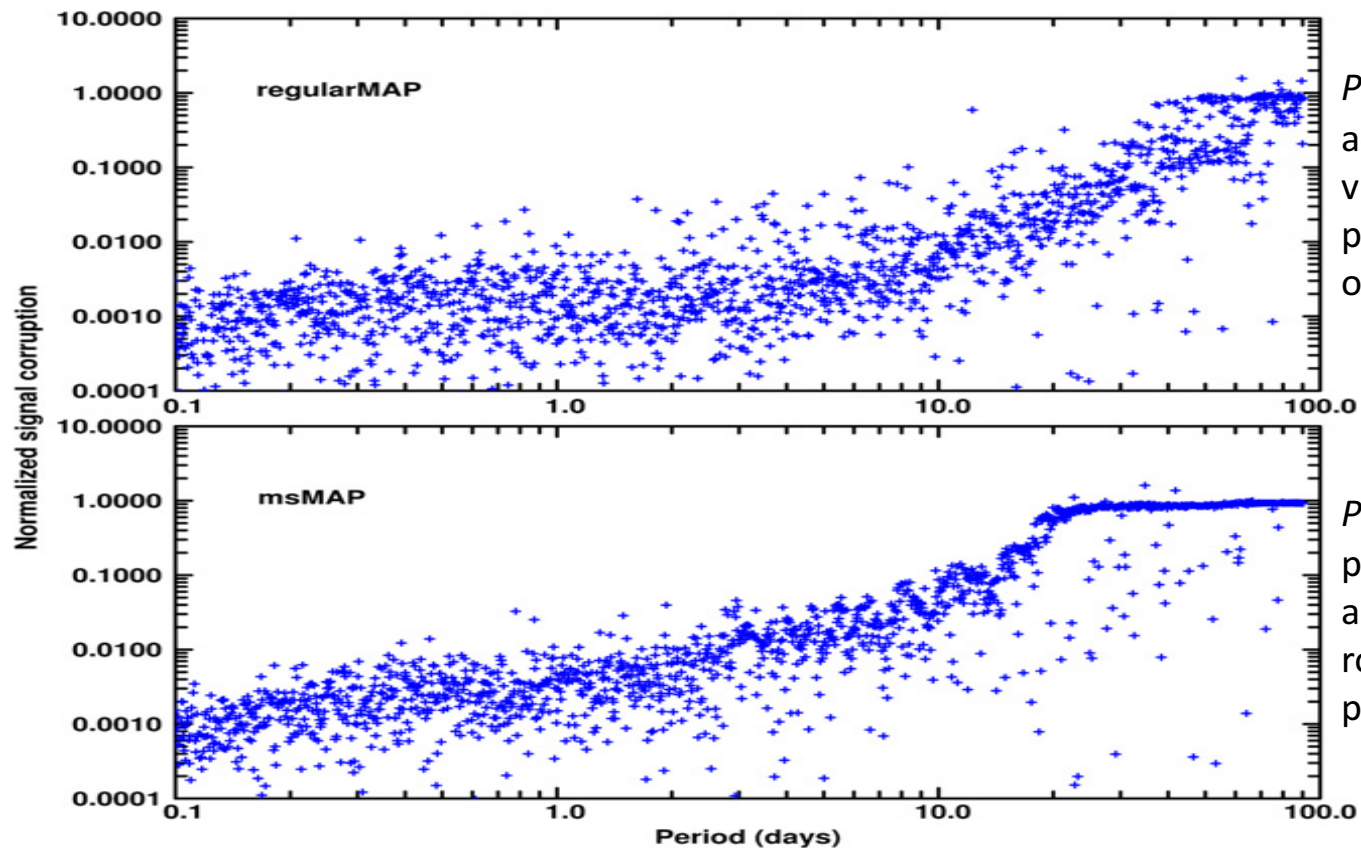
Pascual-Granado et al. (2015) showing gaps filling by their ARMA method

# Correction of residual instrumental effects

Several algorithms have been proposed to detect trends common to ensemble of similar stars and correct them:

- ❑ Trend filtering algorithms (e.g., TFA, Kovacs et al. 2005; SysRem, Tamuz et al. 2005) looking for trends common to groups of stars;
  - ❑ Algorithms adopting a simple model (linear) dependence of the systematic effects on some parameters, e.g, position on the focal plane (e.g., Guterman et al. 2016, in CoRoT Legacy Book).
- The main limitation is that they tend to fit part of the intrinsic source variability, especially in the case of the long-term trends for which the amount of information is limited (often less than one complete cycle of the modulation is covered by the timeseries);
  - Adding a priori information to limit “aggressive correction” by chi-square minimization can mitigate this problem (see Kepler PDC MAP approach in my talk at PW8; or the ARC2 pipeline by Aigrain et al. 2017);
  - This increases the complexity of the algorithm and therefore the computational cost;
  - Be prepared to make tests of the suppression of long-term periodicities by PLATO pipeline because no algorithm is perfect.

# Suppression of a periodic signal in Kepler PDC

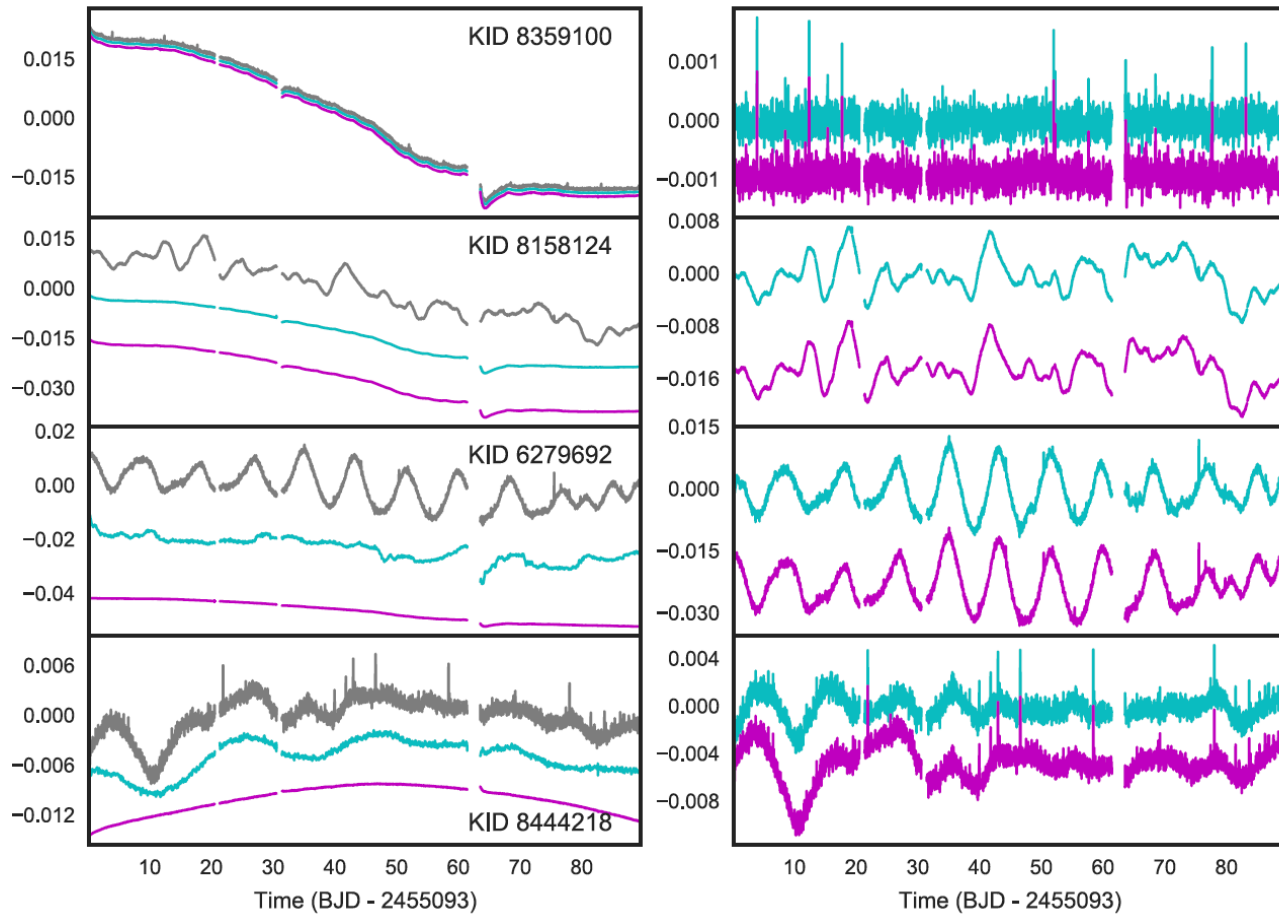


*PDC MAP pipeline:* not optimized according to the timescale of the variation: non-ideal transit search performance, but better preservation of long-term variability.

*PDC ms-MAP pipeline:* better performance in transit searching, but almost complete suppression of rotation and activity signals for periods longer than 20-25 days.

(Gilliland et al. 2015)

# Some sample light curves: PDC vs. ARC 2



ARC2 is a de-trending pipeline introduced by Aigrain et al. (2017) as an alternative to the PDC pipelines.

*Left column:* Kepler Simple Aperture Photometry (SAP) light curves (in gray), PDC correction (cyan) and ARC2 correction (magenta).

*Right column:* PDC corrected light curves (in cyan) and ARC2 corrected light curves (in magenta).

# Binning of the time series

- Some kind of data analysis can be performed on binned data to reduce the computation burden and the filling in of the gaps (however, a gap changes the SNR of the binned data);
- Algorithms looking for rotational modulation or spot modelling *outside transits* can work with binned time series;
- A binning window of 2-3 hours is usually sufficient to preserve the interesting information.

# Putting everything together (see talk by A. Moya)

## Analysis-ready light curves for stellar physics

Authors: WP128.300 and WP123.000 teams for WP120.000

A. Moya, A. Lanza, W.J. Chaplin, M. Lund, R. García, S. Mathur, G.R. Davies, M. Nielsen, J. Pascual-Granado, P. Gaulme, R. Alonso

### Flow chart version 1

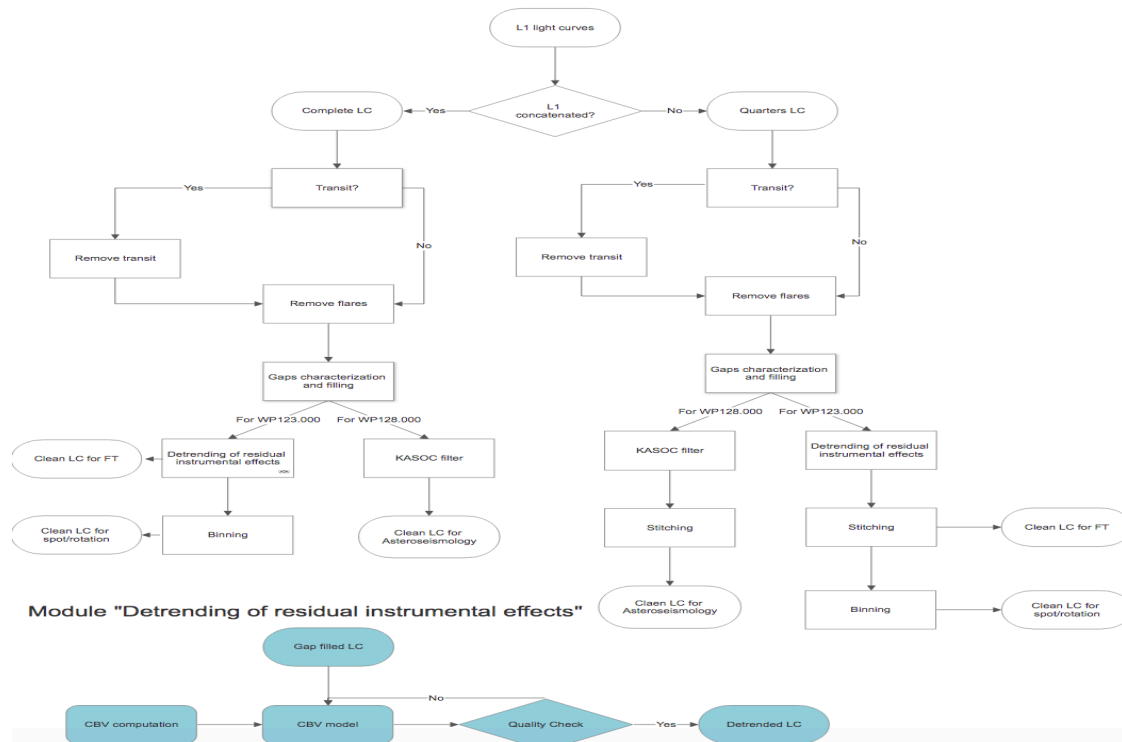


Figure 1: Proposed flow chart for the preparation of light curves for stellar physics analysis.

© A. Moya



# Conclusions and proposals

- Keeping all the astrophysical variability in the L1 light curves from minutes to year timescales is fundamental for WP123, but it is *a formidable task*;
- Exchange of information between different PSM and PDC WPs is crucial to this purpose;
- Presently, there is a coordination between WP 128, WP 123, and WP 111 and related PDC WPs; there is also a WG working on light curve stitching and a WG on stellar flares;
- At some point in the future, a wider WG on the *correction of residual instrumental effects* could be very useful to keep in contact people working on data reduction from L0 to L1 (would they include also this correction ?) and people working on the analysis to produce L2 products;
- Make the full data reduction pipeline (L0 --> L1 --> L2) available for *testing possible overcorrection or distortion* of intrinsic variability; this requires a big effort by different teams, but it is worth doing to fully exploit the scientific potential of PLATO;
- Save information derived during the light curve preparation as ADPs or IDPs (e.g., catalogues of stellar flares, occulted spots).

Thank you for your attention

**Additional material**

# Kepler PDC pipeline

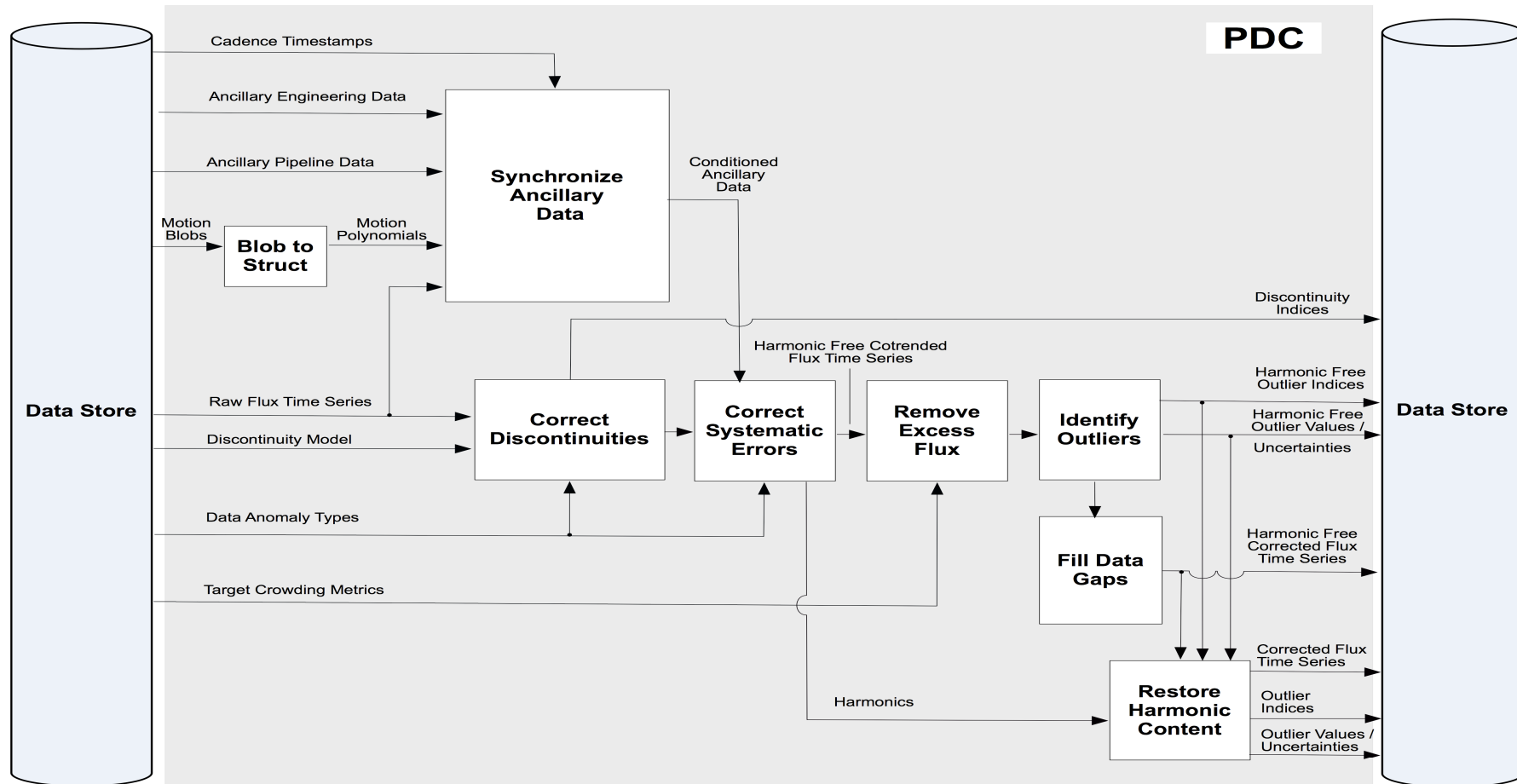


Figure 1: Data flow diagram for the Presearch Data Conditioning (PDC) pipeline module. Inputs are shown at the left and outputs are shown at the right. Inputs are obtained from the Data Store and outputs are written to the Data Store.

Twicken et al. (2010)