ÖAW IWF

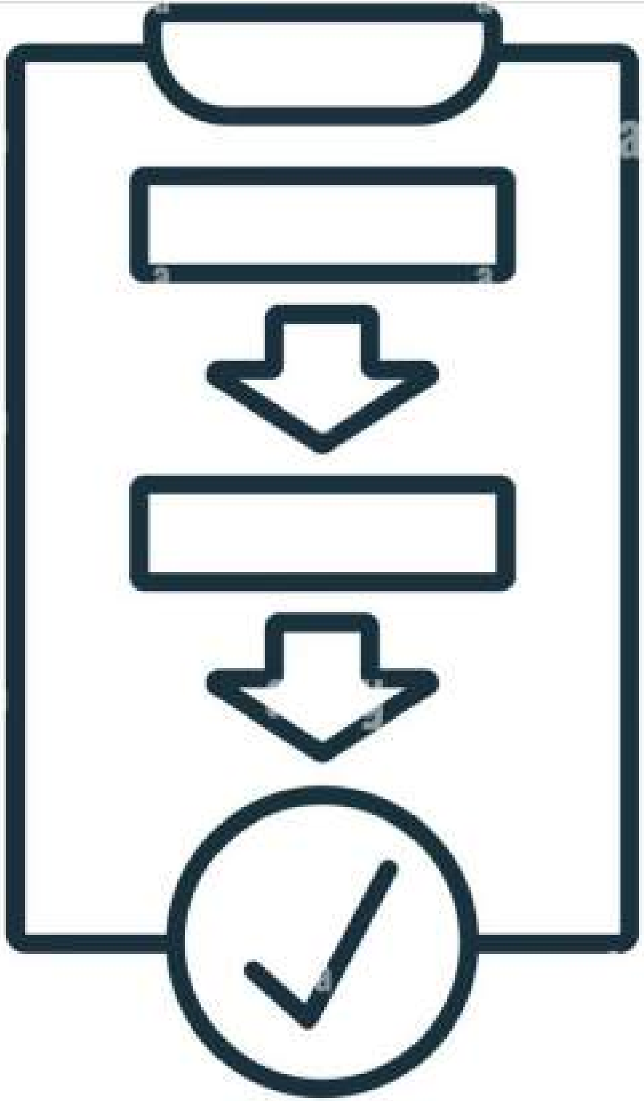
Cleaning RV time series from stellar activity: SN-fit and *bp*-method

Andrea Bonfanti
Austrian Academy of Sciences
Graz, Austria

Seminar @ OATO · 15-11-2022

ÖAW IWF

Outline



1. General introduction
   - Exoplanet detection techniques
   - Radial velocity (RV) method
2. Spectra and CCF extraction
   - CCF-related activity indicators
3. A novel approach for removing stellar activity

Skew Normal (SN) fit + breakpoint method (bp)

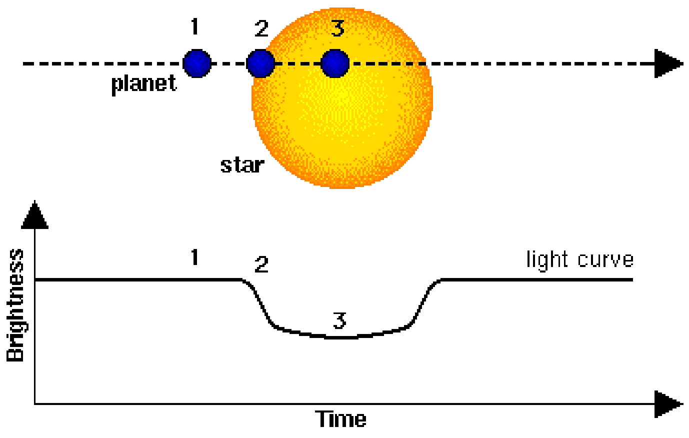
- Discussion
- Simulation tests
- Application on real data

ÖAW IWF

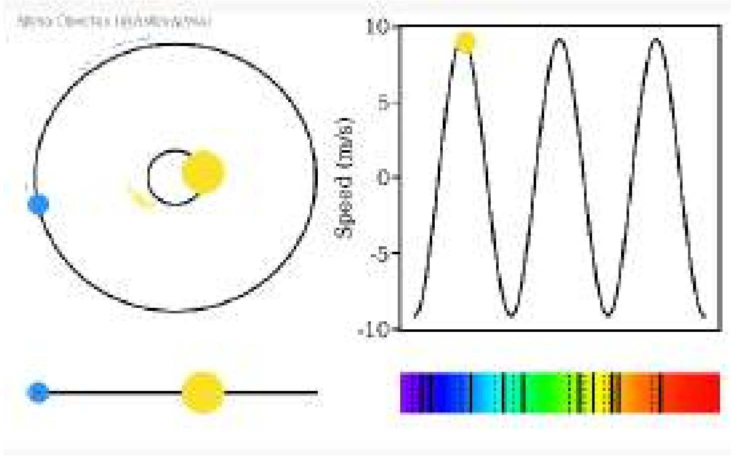
Exoplanet detection methods

Exoplanets are mainly detected through indirect techniques

- Transit method



- Radial velocity method



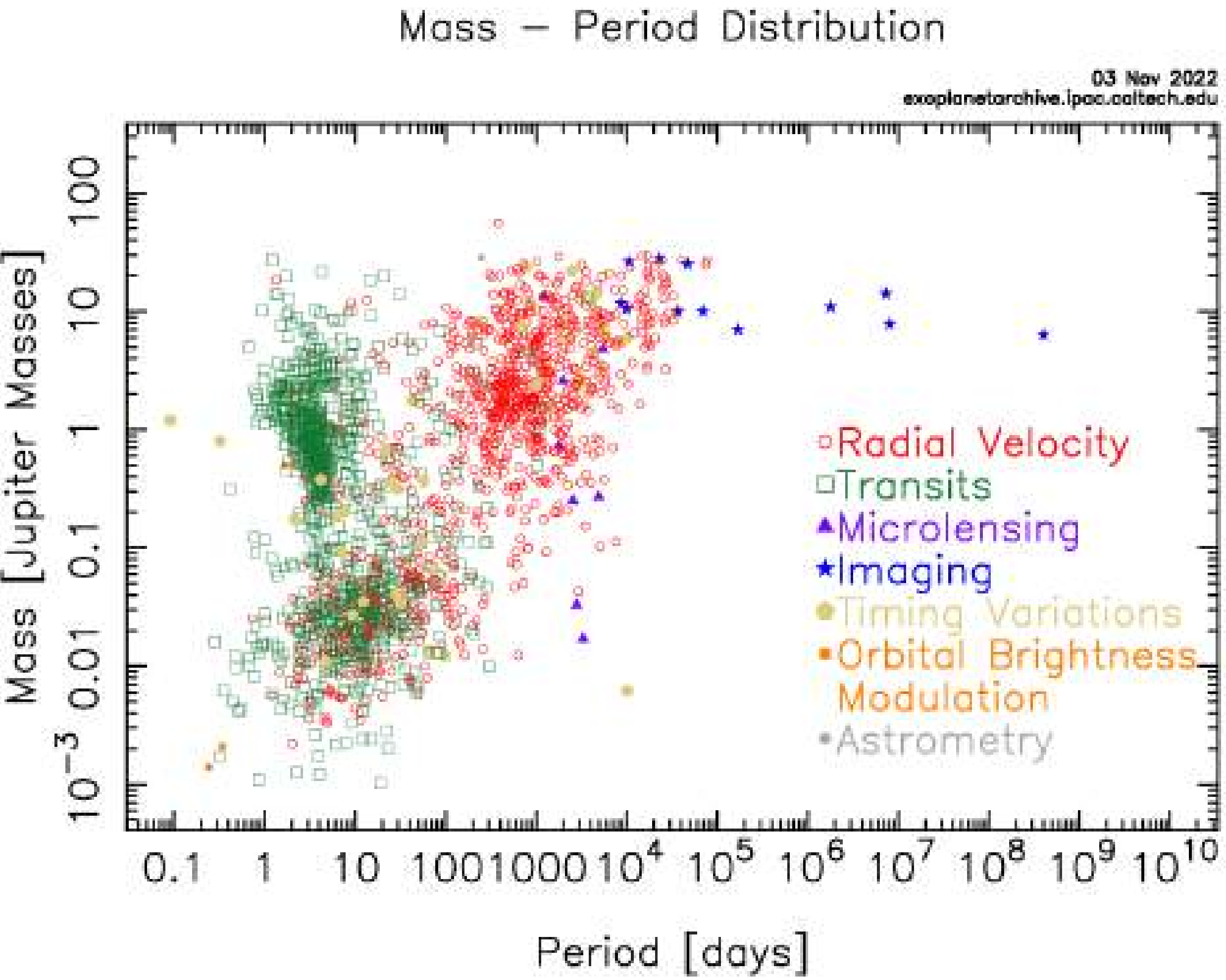
Knowing stellar parameters is essential to determine absolute properties of exoplanets

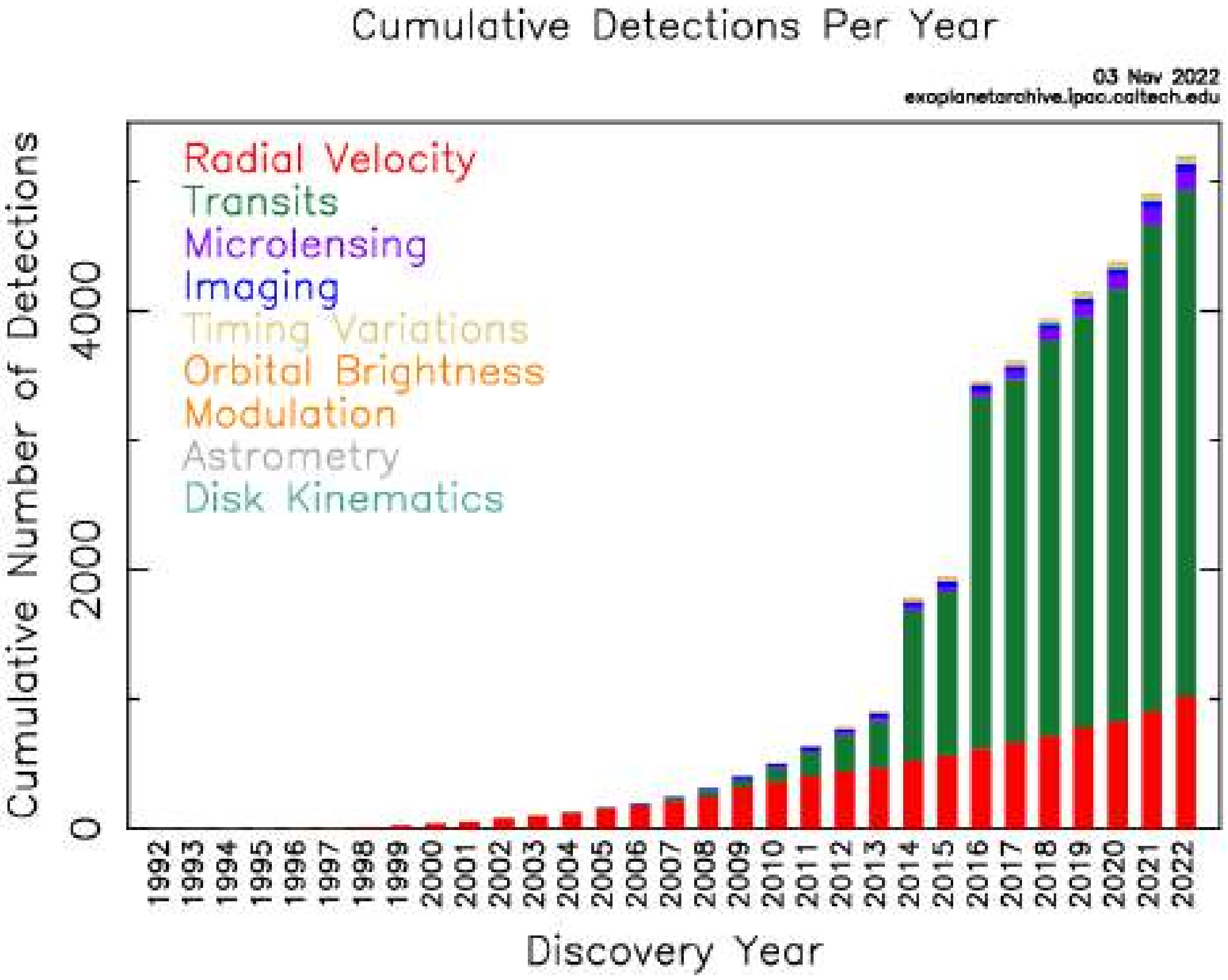
- $R_p \propto R_\star$
- $M_p \propto M_\star^{2/3}$

This sinergy allows us to determine ρ_p, hence unveiling the interior structure of exoplanets

ÖAW IWF

Exoplanet discovery by method

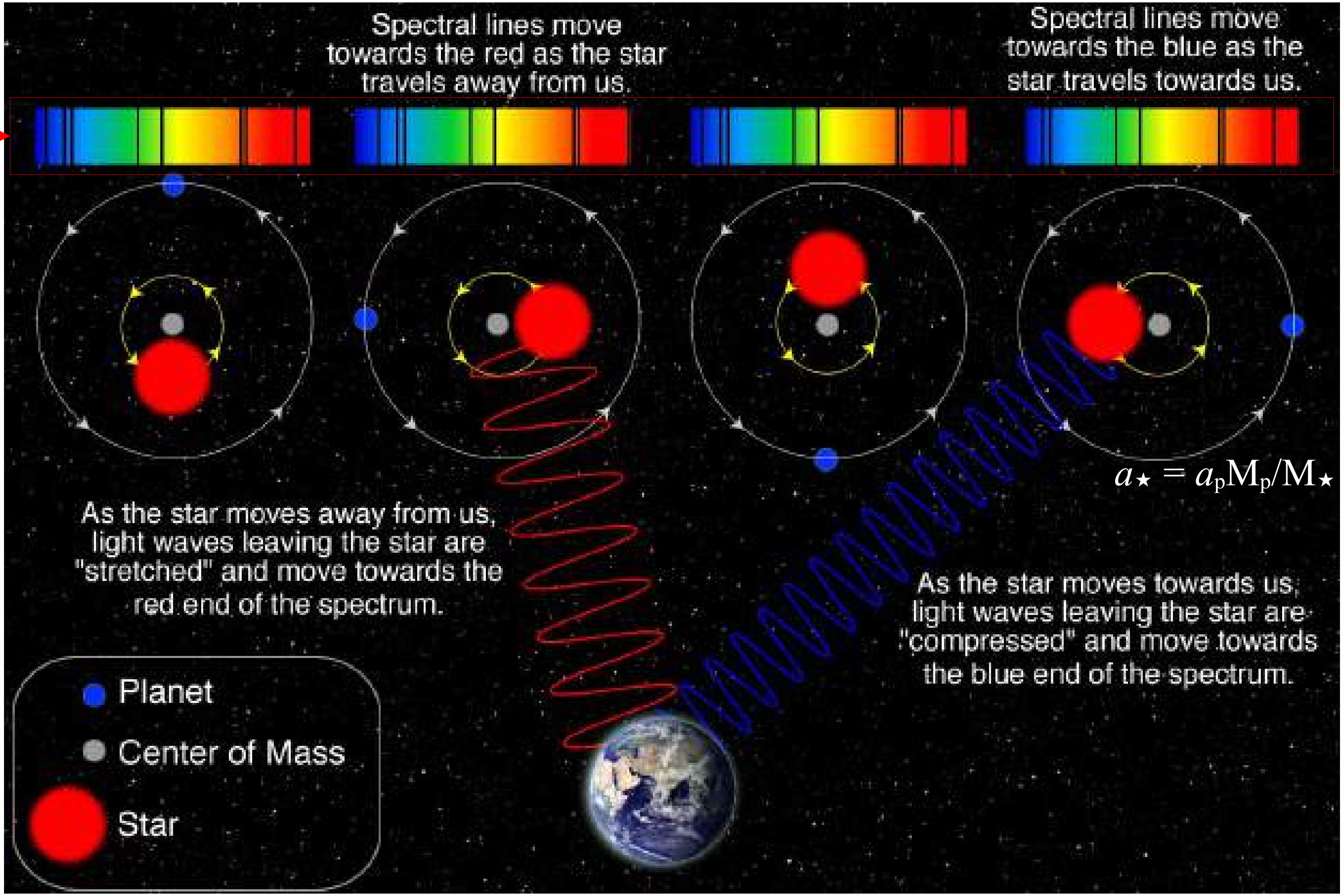




Let’s focus on the RV technique!

RV technique

Periodic shift of absorption lines in stellar spectra



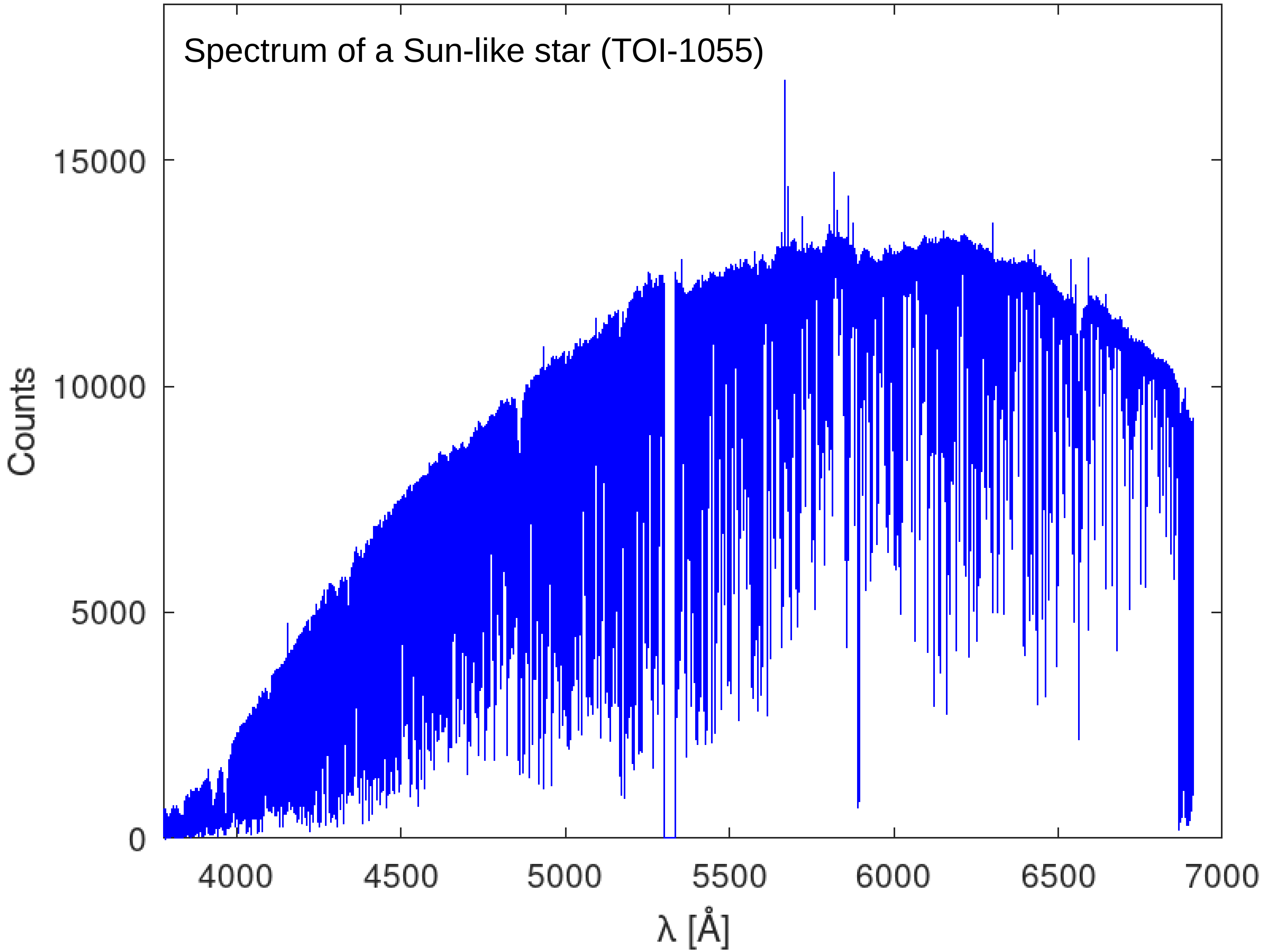
- The star and the planet orbit around their common barycenter
- The reflex motion of the star induced by the gravitational perturbation of the planet may be detected through the…

  ... Doppler effect

Extracting radial velocity of stars from spectra...

Information on the instantaneous Doppler shift is contained in the many thousands of absorption lines present in the stellar spectra

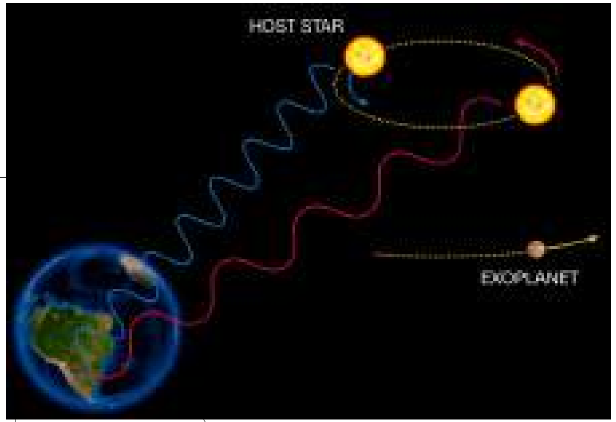
By measuring the centroid of stellar absorption line the radial velocity of the stellar reflex motion at different epochs is inferred from Doppler effect



ÖAW IWF

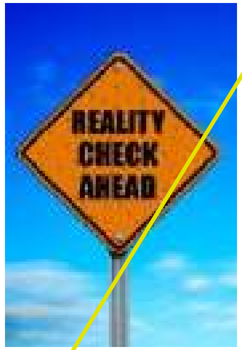
Activity-induced Doppler signals

Stellar activity in the form of *oscillations*, *granulations*, *spots* or *faculae* also produces Doppler signals that can induce spurious RVs variations



- The Doppler reflex motion induced by a planet shifts the entire stellar spectrum without changing the lines' shape
- Stellar activity modifies the shape of spectral lines ⇒ creates a spurious shift of the stellar spectrum

It should be easy to disentangle these two effects



Stellar activity can induce signals up to a few hundreds m·s^{-1}, which

- may completely hide the Keplerian signal of exoplanets
- Induce changes in the shape of the spectral lines that are challenging to be detected

To measure such tiny variations, a common approach is to average the information of all the lines in the spectrum

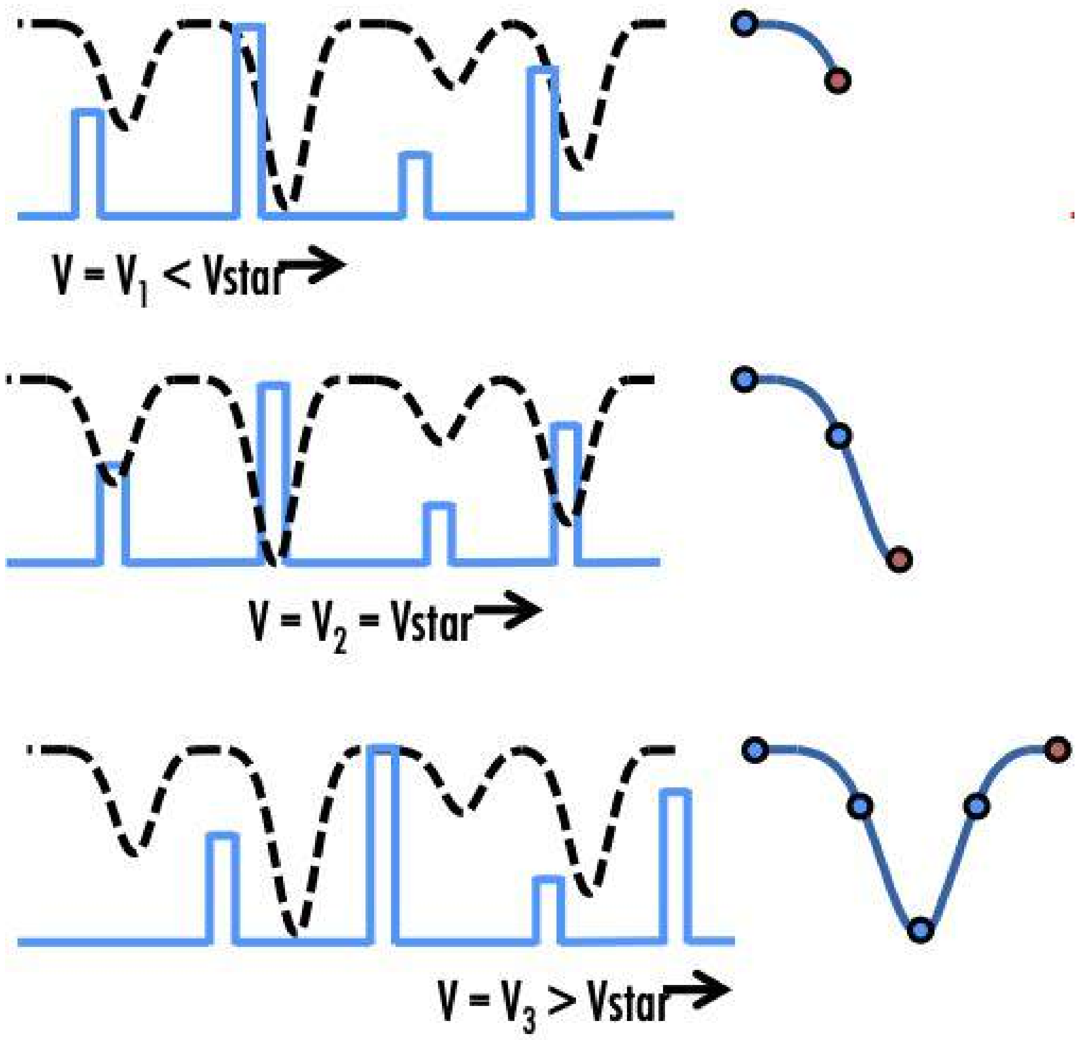
CCFs

Cross-correlation spectroscopy

- The instantaneous Doppler shift is contained in the many thousands of spectral lines
- Averaging the info contained in the spectral lines is done by cross-correlating the spectrum *S* with a synthetic stellar template (the mask *M*)
- This operation gives the **cross-correlation function → CCF**

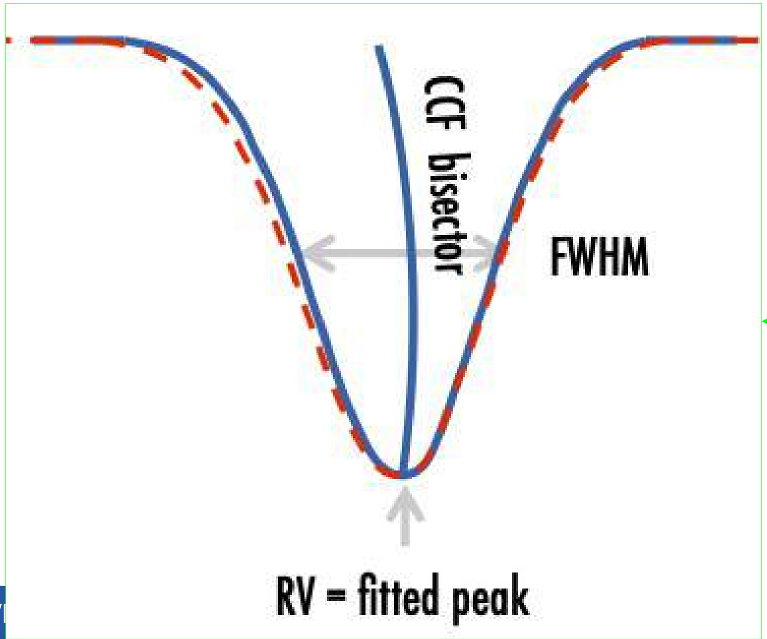
$$C(\epsilon) = \int_{-\infty}^{+\infty} S(v)\, M(v-\epsilon)\, dv$$

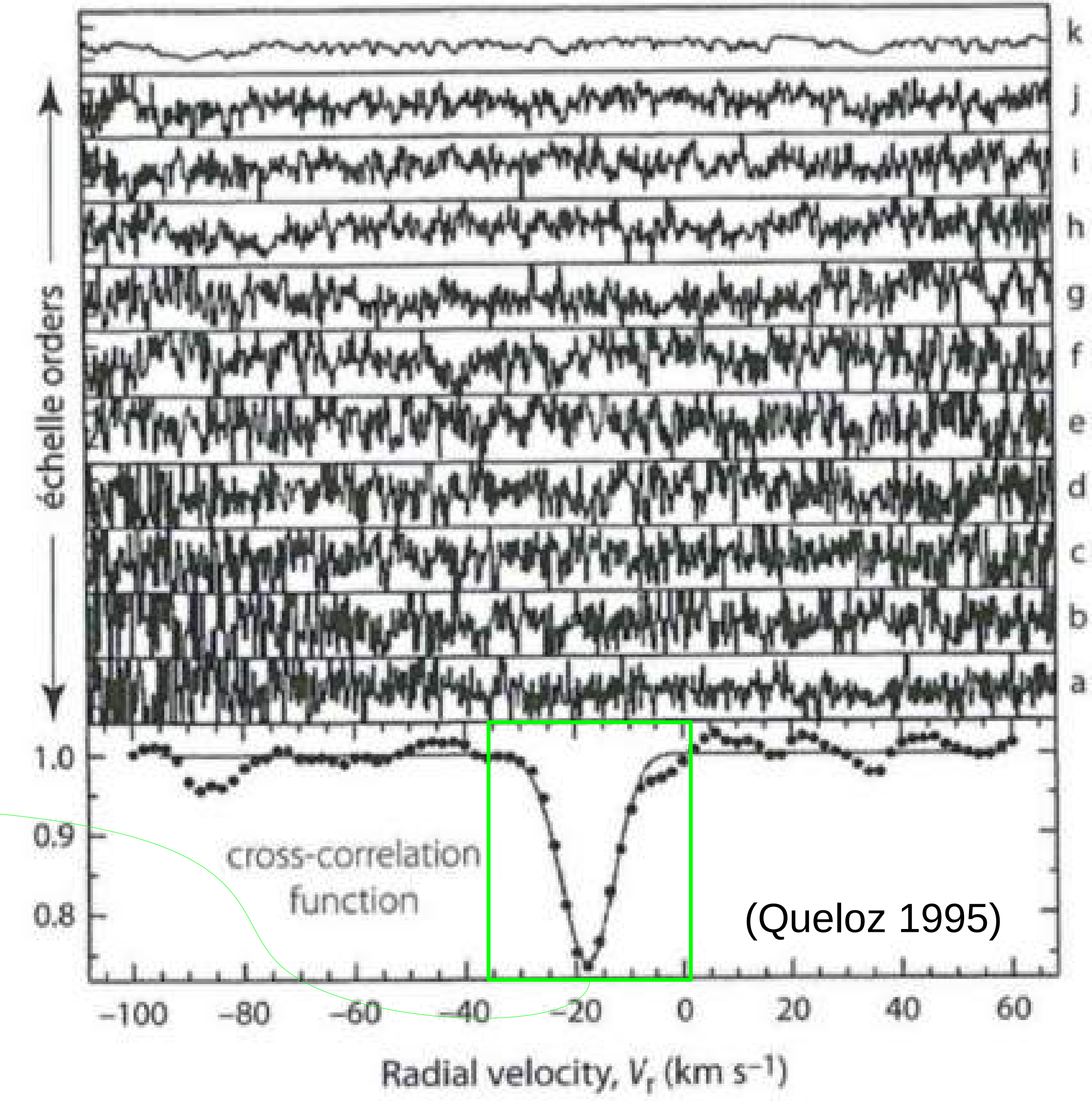
The spectrum S and the mask M are both expressed in the velocity space v and ϵ is the shift



CCF extraction

- Typical pipelines operate on each echelle order of the spectrograph to produce individual CCFs
- CCFs are usually fit with Gaussians to produce order-by-order RV measurements that can be used to scrutinize chromatic behaviors
- The order-by-order CCFs are then summed, and the sum is usually fit by a Gaussian as well to measure a single RV value for that observation





(Queloz 1995)

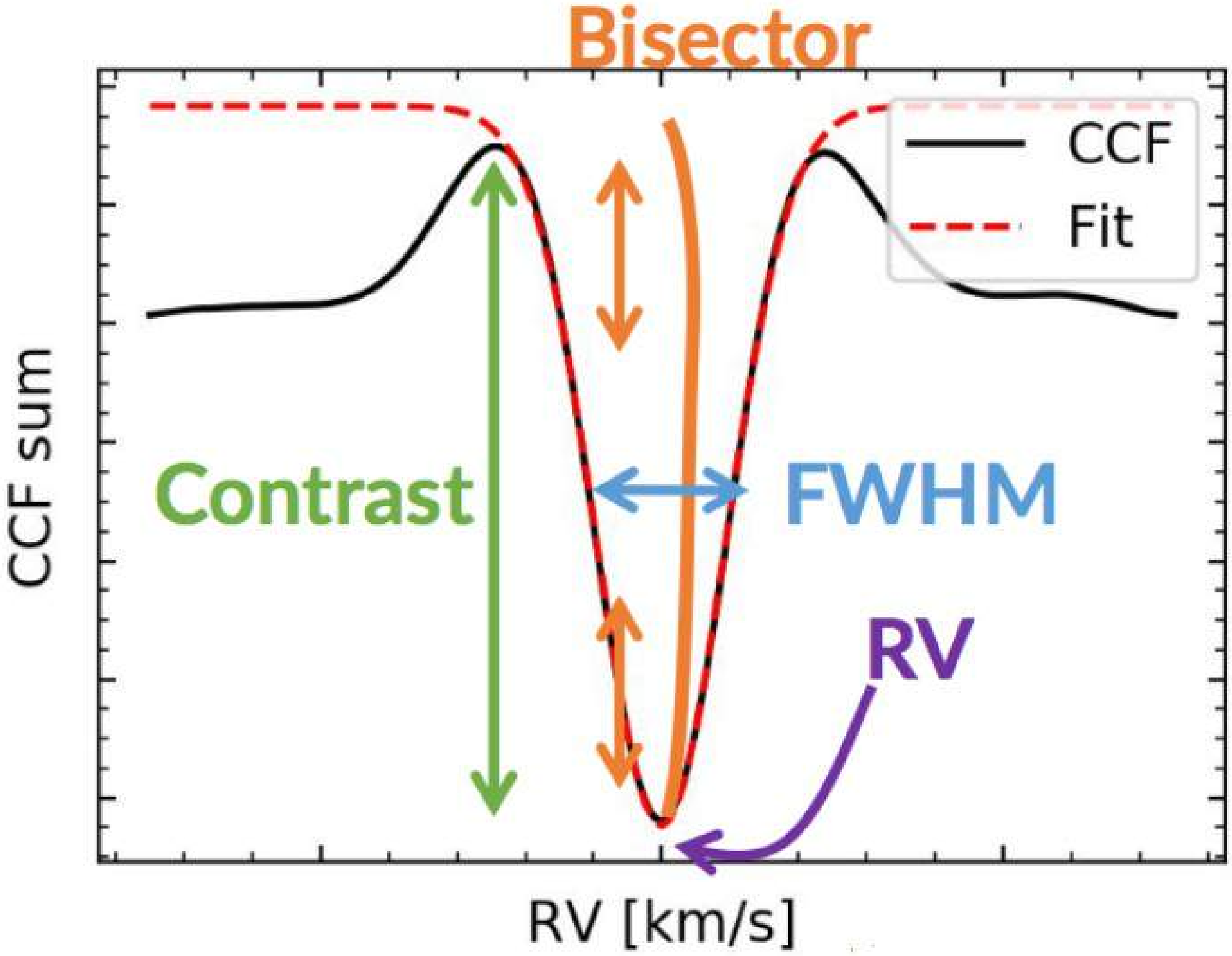
ÖAW IWF

Activity indicators from the CCF

If a spurious RV signal is induced by activity, it is essential to find out activity indicators which are expected to correlate with RV(t) in order to mitigate the stellar activity contribution

Usual approach:

- Gaussian fit onto the CCF (Normal fit: N-fit) to infer:
  - RV
  - FWHM
  - A → Contrast
- At a following stage the asymmetry of the CCF is quantified, e.g. through the Bisector Velocity Span → BIS



Credits: M. Lafarga

ÖAW IWF

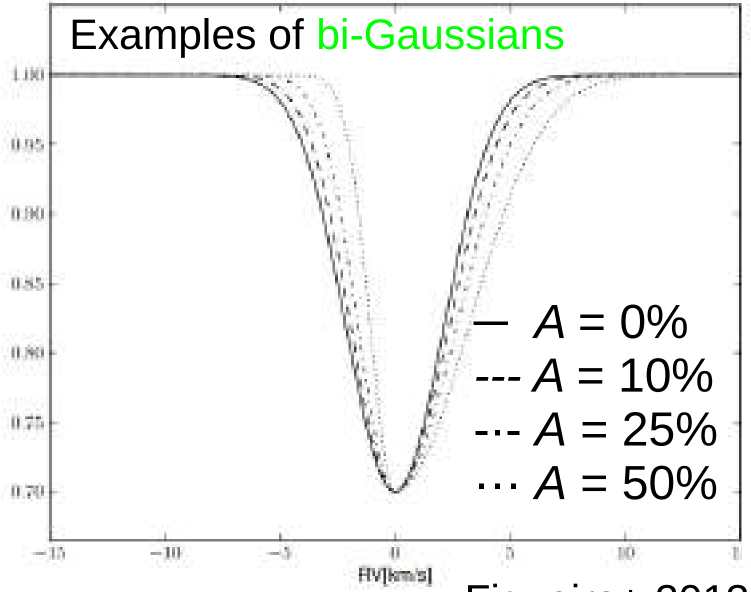
Evaluating the CCF asimmetry as in the literature

- BIS SPAN = V_t - V_b (Queloz et al. 2001)
- V_{span}: difference between the RV measured by fitting a Normal density to the upper and lower parts of the CCF (Boisse et al. 2011)
- BIS$^+$
- BIS$^-$
- V_{asy}

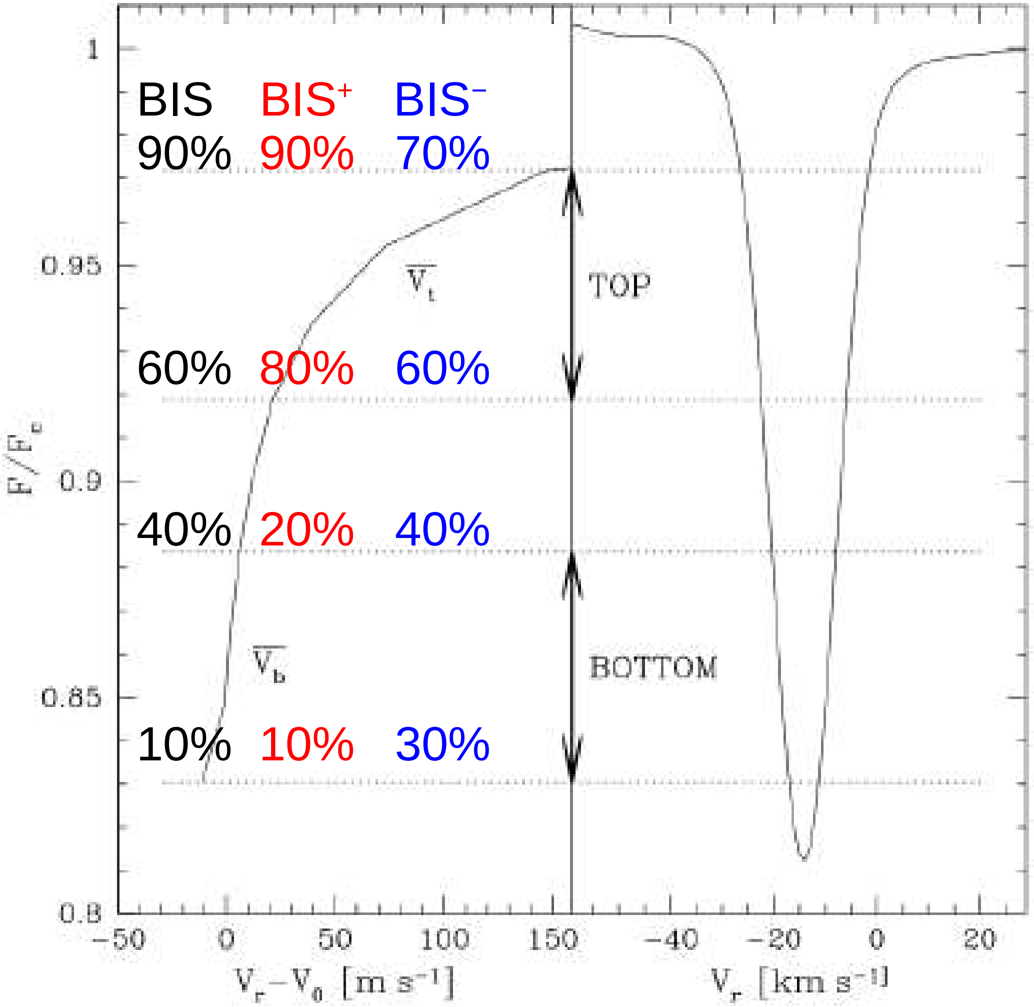
  RV difference as inferred from the info content in the red and blue wings of spectral lines

(BIS$^+$, BIS$^-$, V_{asy}: Figueira at al. 2013)

- $\Delta V = RV_{bG} - RV_G$ difference determined by comparing a Gaussian with a bi-Gaussian fitting (Nardetto+ 2006, Figueira+ 2013)



Figueira+ 2013



Queloz+ 2001

ÖAW IWF

Skew Normal fit to catch the CCF asymmetry

Almost all the methods use a Normal density fitted to the CCF and the asymmetry is estimated using a separate approach.

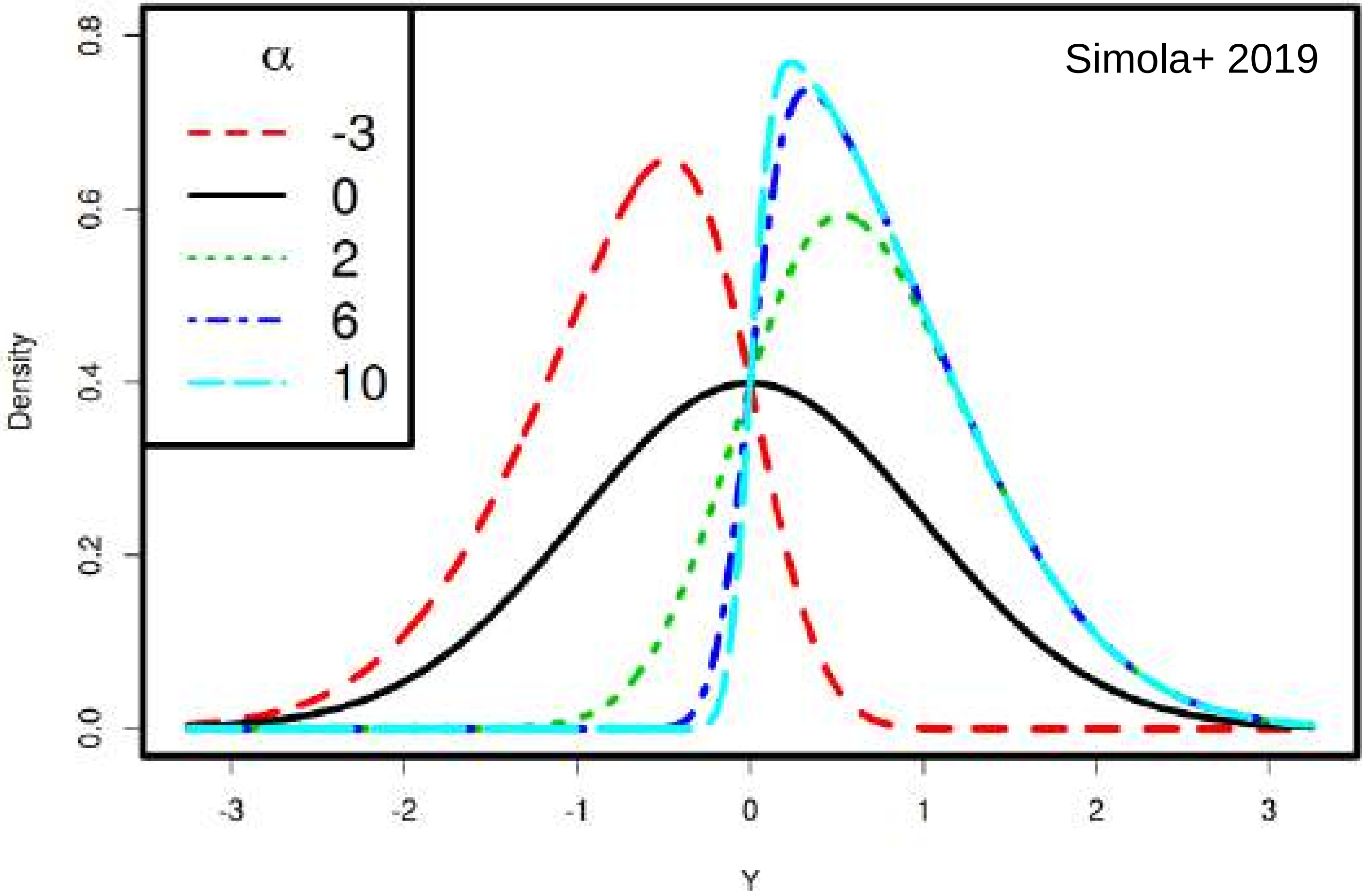
What we propose is...

Skew Normal (SN) density to estimate with a single fit of the CCF the RV, FWHM, and asymmetry of the CCF, as this function includes a Skewness parameter

$$\mathrm{SN}(y;\xi,\omega,\alpha) = \frac{2}{\omega}\phi\left(\frac{y-\xi}{\omega}\right)\Phi\left(\frac{\alpha(y-\xi)}{\omega}\right)$$

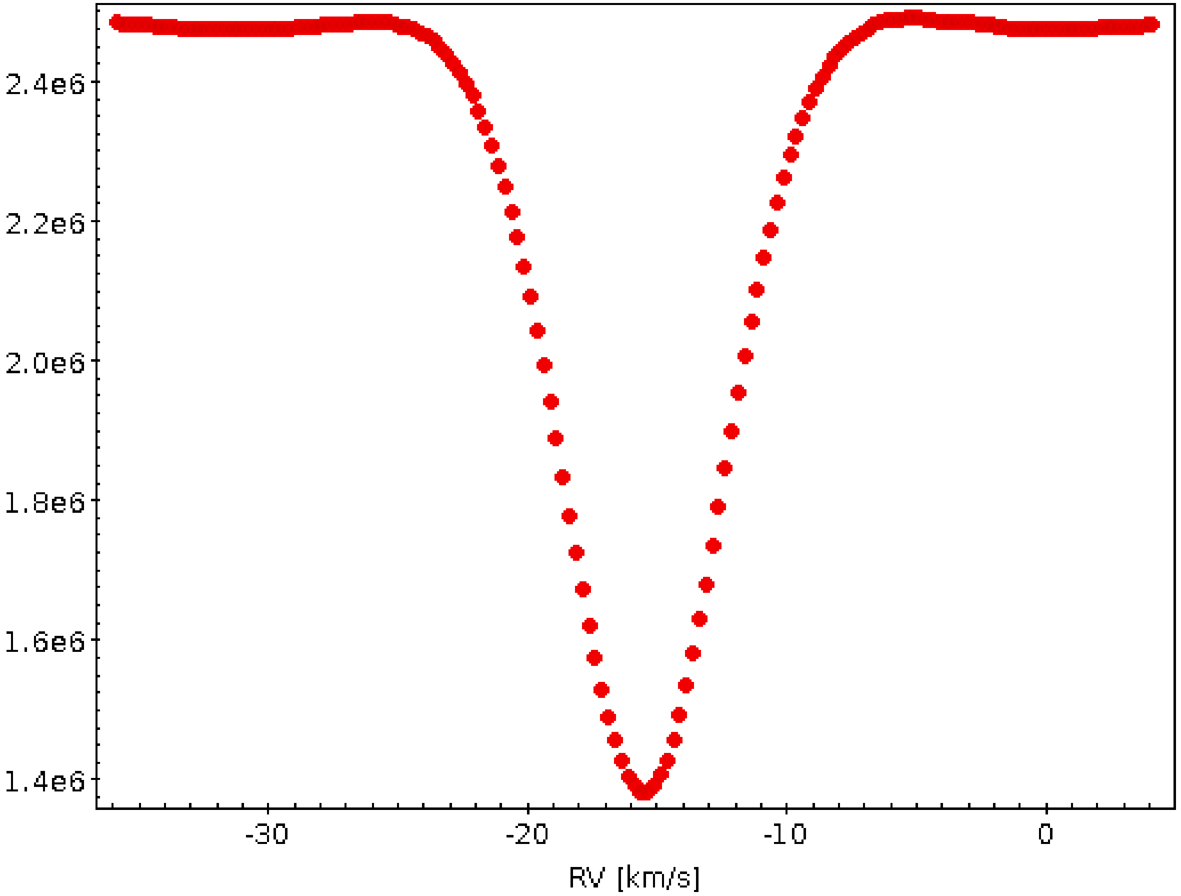
$\phi()$: Gaussian density function
$\Phi()$: Gaussian cumulative distribution function (CDF)

Location parameter	ξ	⟷	μ mean
Scale parameter	ω		σ rms
Asymmetry parameter	α		γ skewness



SN-fit parameters from the CCF

Example of a CCF of a Sun-like star (G2 mask)

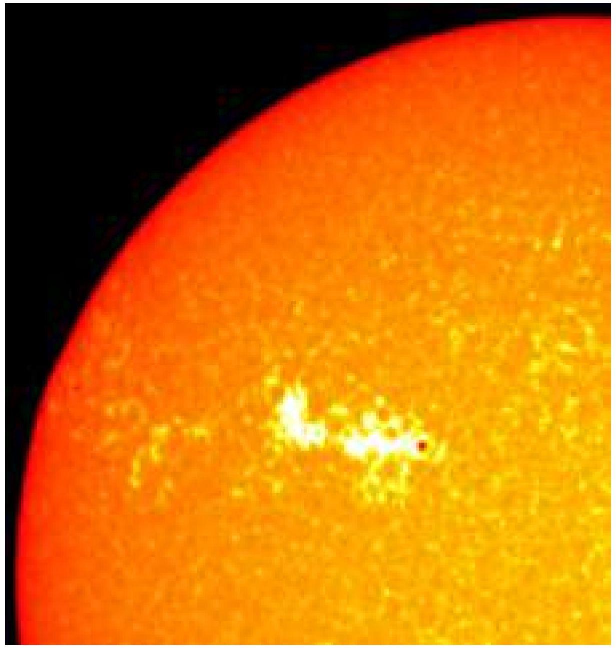


- Stellar activity alters asymmetry and width of the CCF
- Skew-normal [SN] fit onto the CCF following Simola+ 2019, so to retrieve
  - RV
  - $FWHM_{SN}$
  - A → contrast
  - γ → skewness

(t, RV) build up the RV time series

(t, $FWHM_{SN}$, A, γ) is the vector against which perform any data detrending

SN-fit performances on synthetic stars



Synthetic Sun-like star

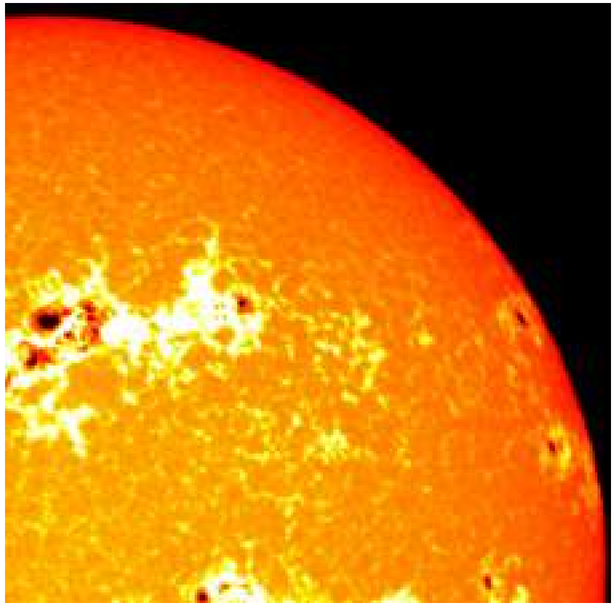
SIMULATION 1

SIMULATION 2

With a facula

With a spot

Generated with SOAP 2.0 (Dumusque et al. 2014)



SOAP computes the CCFs for each desired phase of stellar rotation

RVs are extracted by both fitting

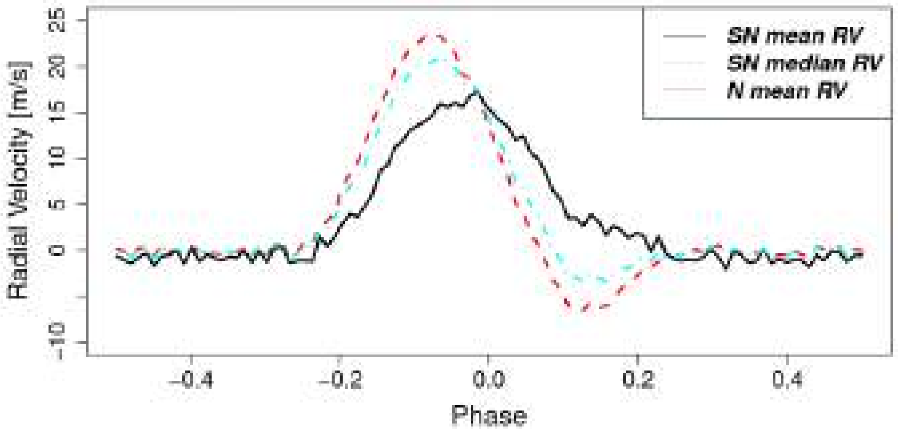
1) a Gaussian (N fit)
2) a Skew Normal (SN fit)

onto the CCFs ⇒ Two different RV timeseries are avaible to compare the performances of the two fits

ÖAW IWF

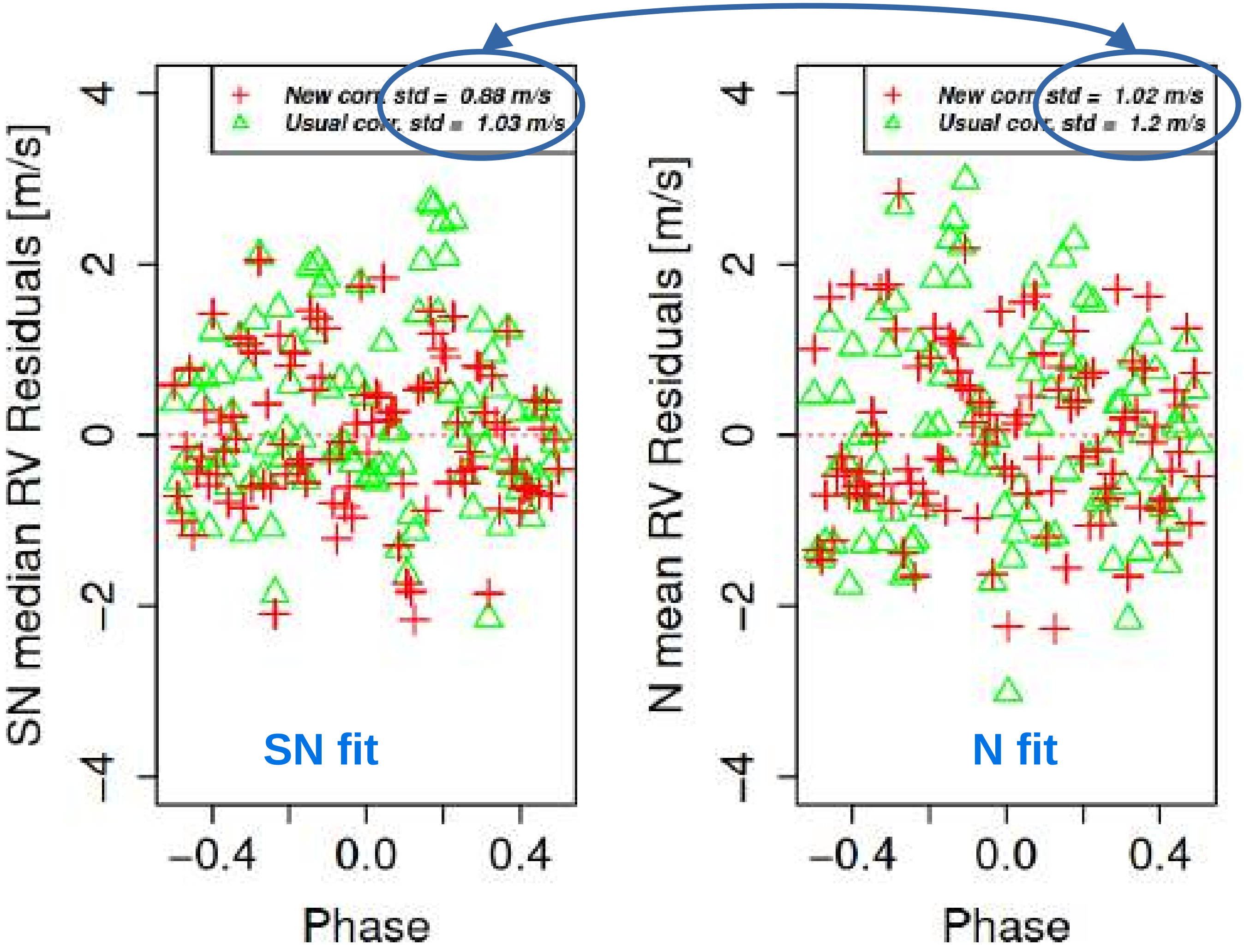
Facula removal

The facula (face-on at phase=0) induce a spurious RV signal varying as a function of the rotational phase



Residuals

SN fit implies lower rms values in the residuals



The two different RV time series are detrended using 1st order polynomials containing

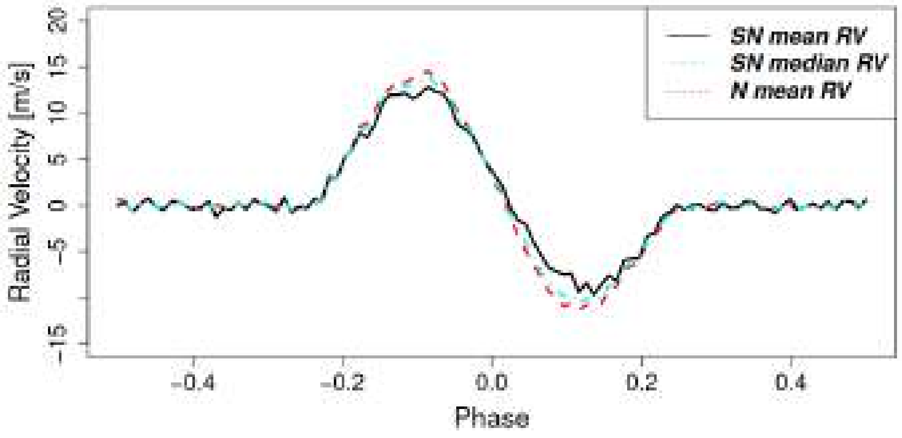
- SN hyperparameters
- ($FWHM_{SN}$; γ) ↔ (FWHM; BIS)

Plots from Simola+ 2019

ÖAW IWF

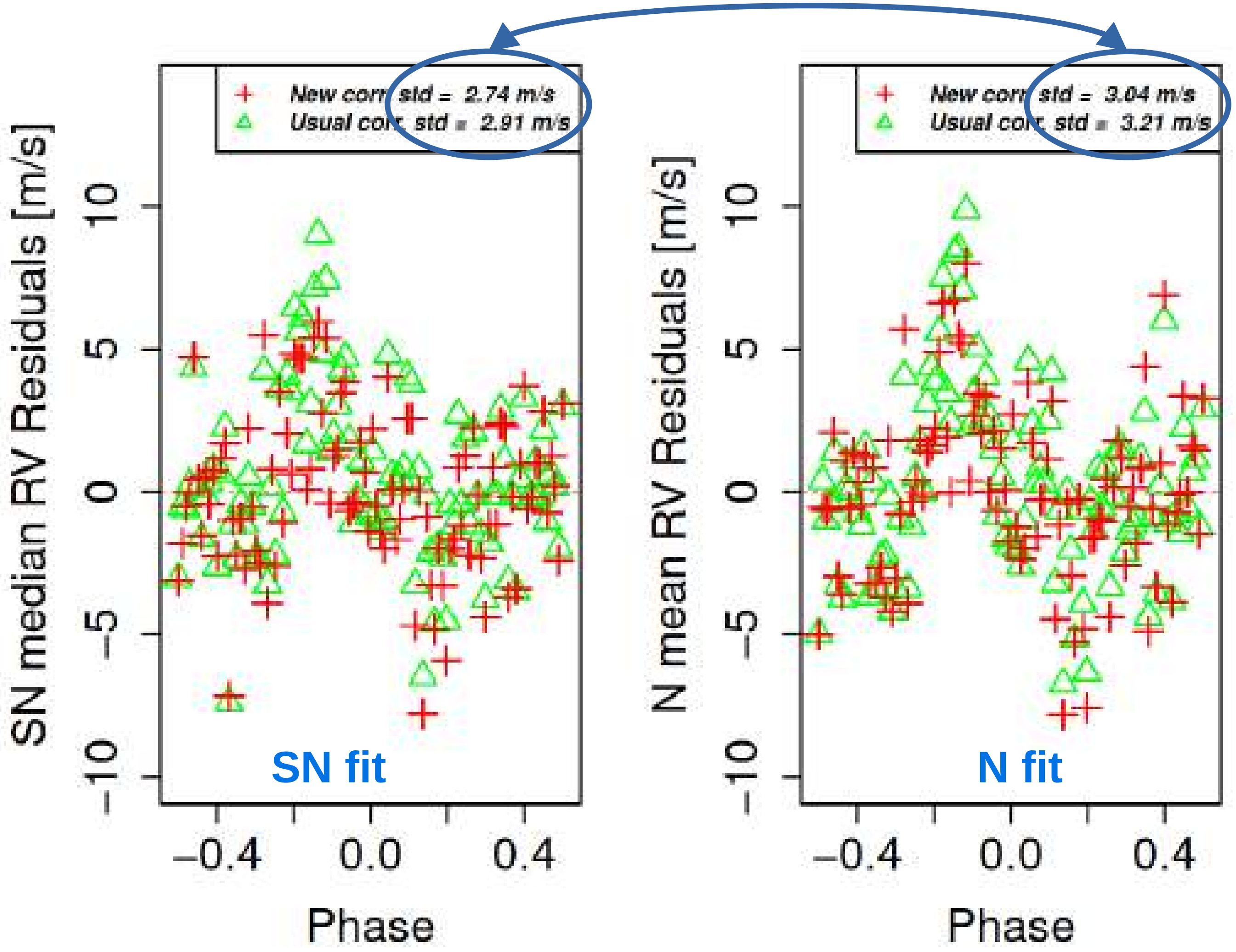
Spot removal

The spot (face-on at phase=0) induce a spurious RV signal varying as a function of the rotational phase



Residuals

SN fit implies lower rms values in the residuals



The two different RV time series are detrended using 1st order polynomials containing

- SN hyperparameters
- (FWHM$_{SN}$; γ) ↔ (FWHM; BIS)

Plots from Simola+ 2019

ÖAW IWF

Activity indicator performances on α Cen B

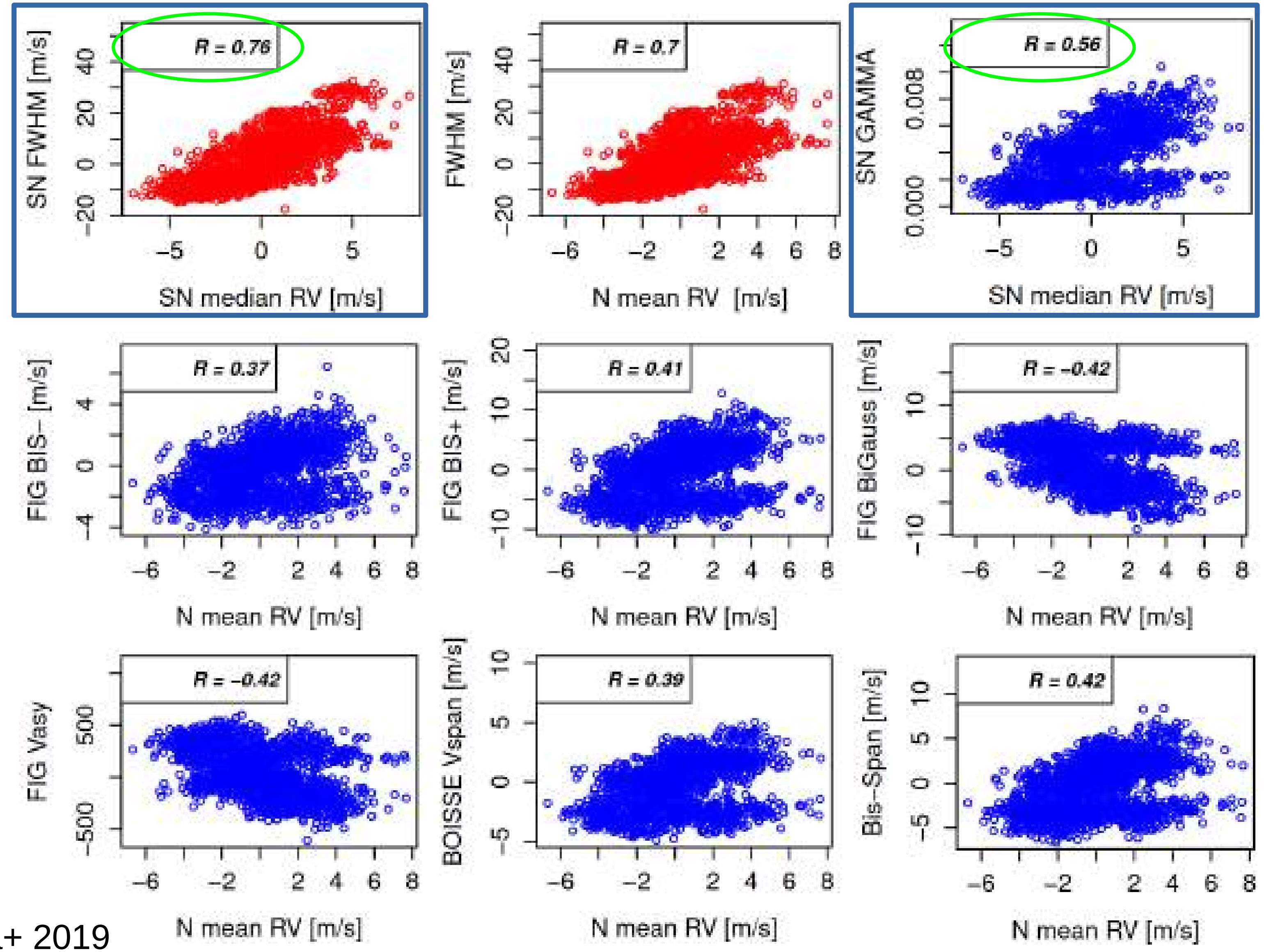
α Cen B does not likely host any planet (Rajpaul+ 2015, Simola+ 2022)

Good target to evaluate how the activity indicators correlate with the RV signal (which is expected to be due to stellar activity only)

Parameters derived from the SN-fit are generally more sensitive to activity

Strongest correlation with RV (highest R)

Helpful when characterising stellar activity signals



Plots from Simola+ 2019

ÖAW IWF

Precision of the SN-fit-based parameters

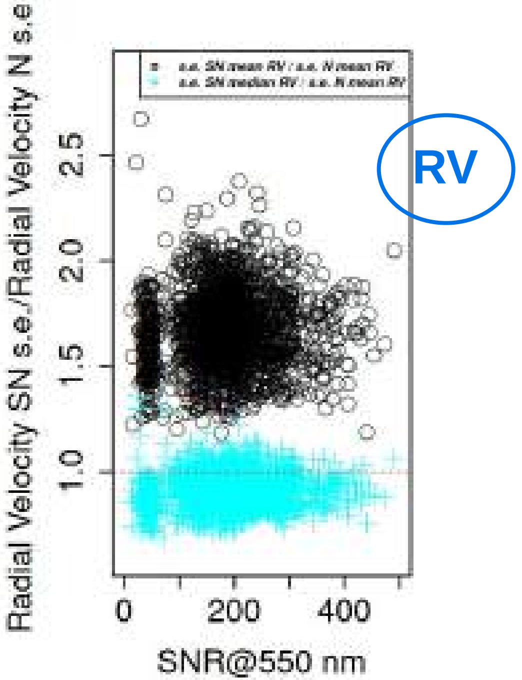
Estimates are based on a simulation study applied to real HARPS measurements

- SNR from 10 up to 500
- K-type stars: HD 215152 (K3V), HD 192310 (K2V), CoRoT-7 (K0V)

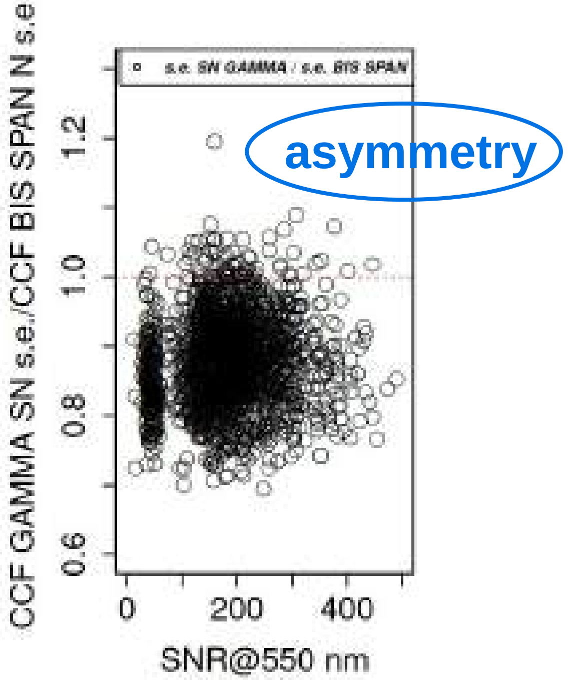
Bootstrap approach

- Perturbation of stellar spectra
- Computation of the CCFs
- Extraction of the CCF parameters by performing
  - a N-fit
  - a SN-fit
- From the bootstrap approach the parameter distributions are available
- rms of the distributions as proxies of the errors

1) SN median RV more precise than N mean RV by 10%



2) Skewness γ more precise than BIS by 15%



Plots from Simola+ 2019

RV detrending

After the SN-fit onto the CCF

- (t, RV) build up the RV time series
- (t, $FWHM_{SN}$, A, γ) is the vector against which perform any data detrending

How to do it effectively?

Let's start from some key points:

- Spurious RV signals are due to stellar activity
- Active regions evolve on the stellar photosphere over time
- Correlations between the RV observations and the activity indicators will correspondingly be anisotropic

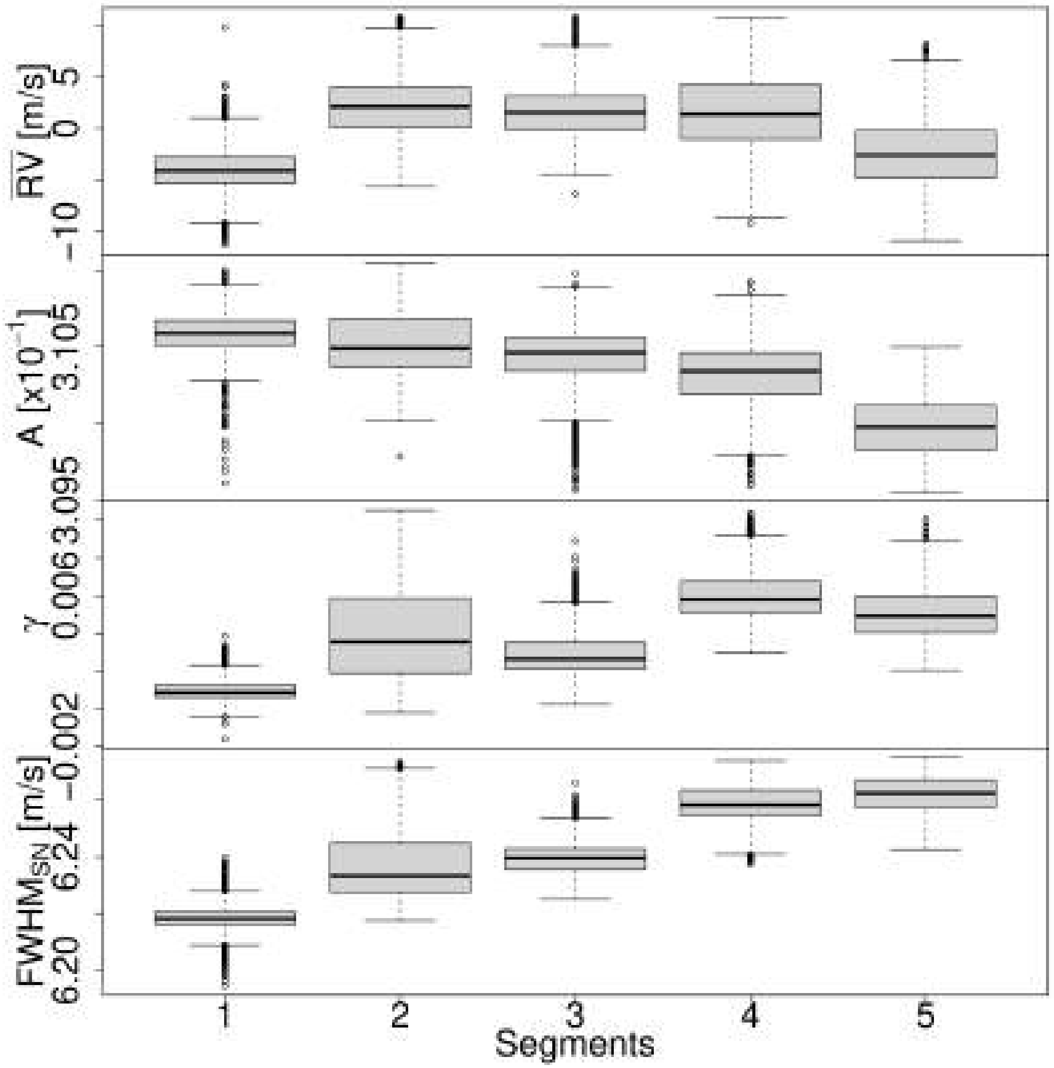
Breakpoint method

If we were able to recognise the RV locations where the correlations between the RV observations and the activity indicators significantly change
$\Rightarrow$ We could better account for variations in RV caused by stellar activity

ÖAW IWF

The breakpoint [bp] method: outline

- The breakpoint [*bp*] algorithm (Simola, Bonfanti, et al. 2022) belongs to the class of Change Point Detection [CPD] methods
- It detects those locations in a time series where correlation changes against parameters of interest ($FWHM_{SN}$, A, γ in our case) are statistically significant
  ⇒ The time series is splitted into segments: the *bp* method is in charge of determining the optimal number of segments and their locations
  ⇒ The detrending is applied to each stationary segment rather than performing an overall [oc] correction to the entire time series
  ⇒ Optimal removal of stellar activity to hopefully detect planetary signals, if any



Application to the **α Cen B time series**

ÖAW IWF

The bp method: the algorithm

- The number D of segments the RV time series shall be divided into is unknown a priori
- The researcher selects
  - a set of D values, D = 1, 2, …, D_{max}
  - a regression model of stellar activity
- For each D value and for each possible partition (i.e. the breakpoint locations), the bp algorithm computes the likelihood defined by the adopted regression model
- The best D with its optimal partition is the solution for which the BIC-penalised likelihood is maximised

A polynomial whose basis vector is (A, $FWHM_{SN}$, γ) in our case

The bp algorithm is implemented either in

R (e.g. in the `strucchange` package)

`python` (e.g. through the `ruptures` package)

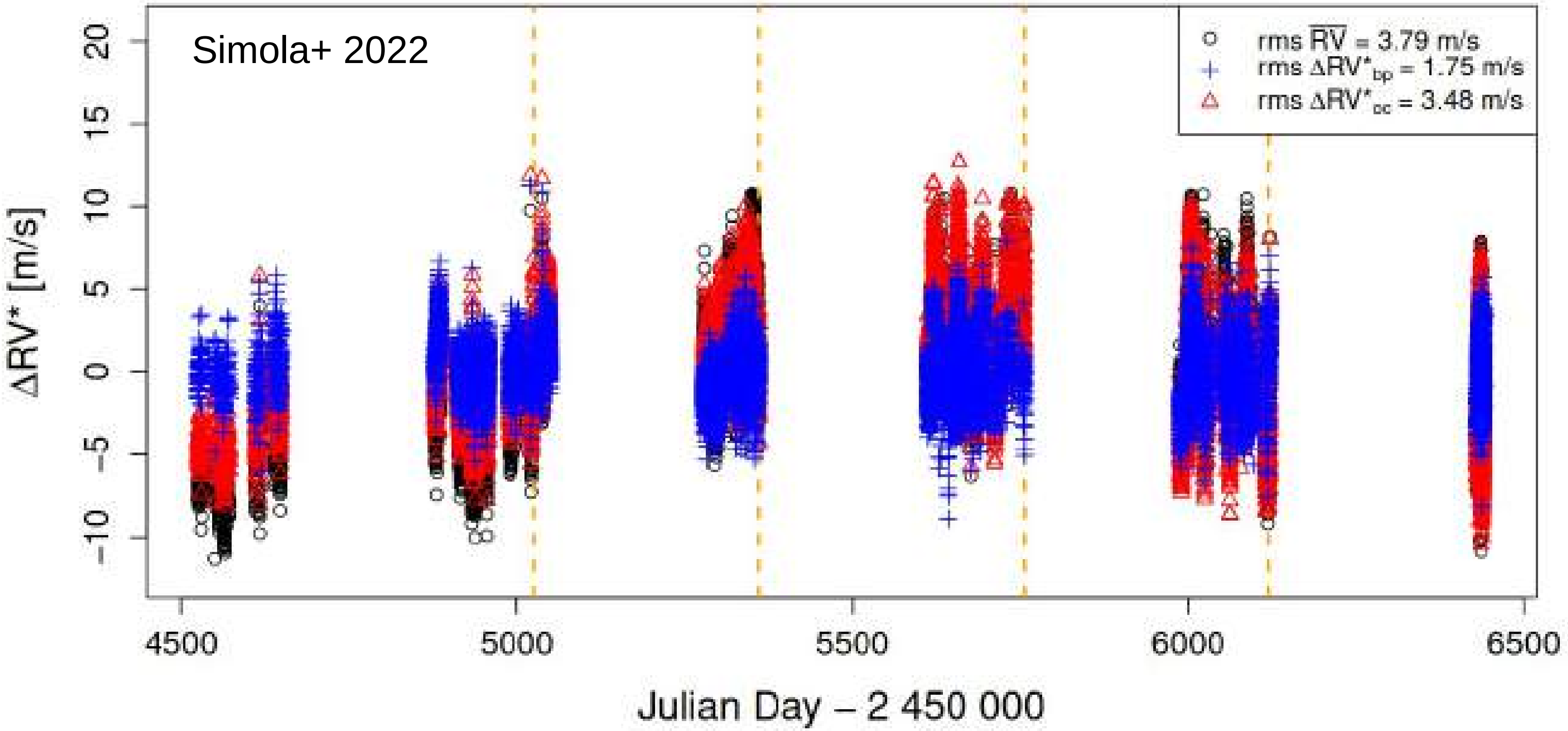
ÖAW IWF

bp vs. oc method: RV time series detrending

Starting points
1) a time series: **α Cen B**
2) a model of stellar activity
$RV^* = \beta_0 + \beta_1 \cdot FWHM_{SN} + \beta_2 \cdot \gamma + \beta_3 \cdot A$ to be applied on
- the entire time series → **oc method**
- each piecewise stationary segment → **bp method**

Residuals of the detrended time series $\Delta RV^* = RV - RV^*$

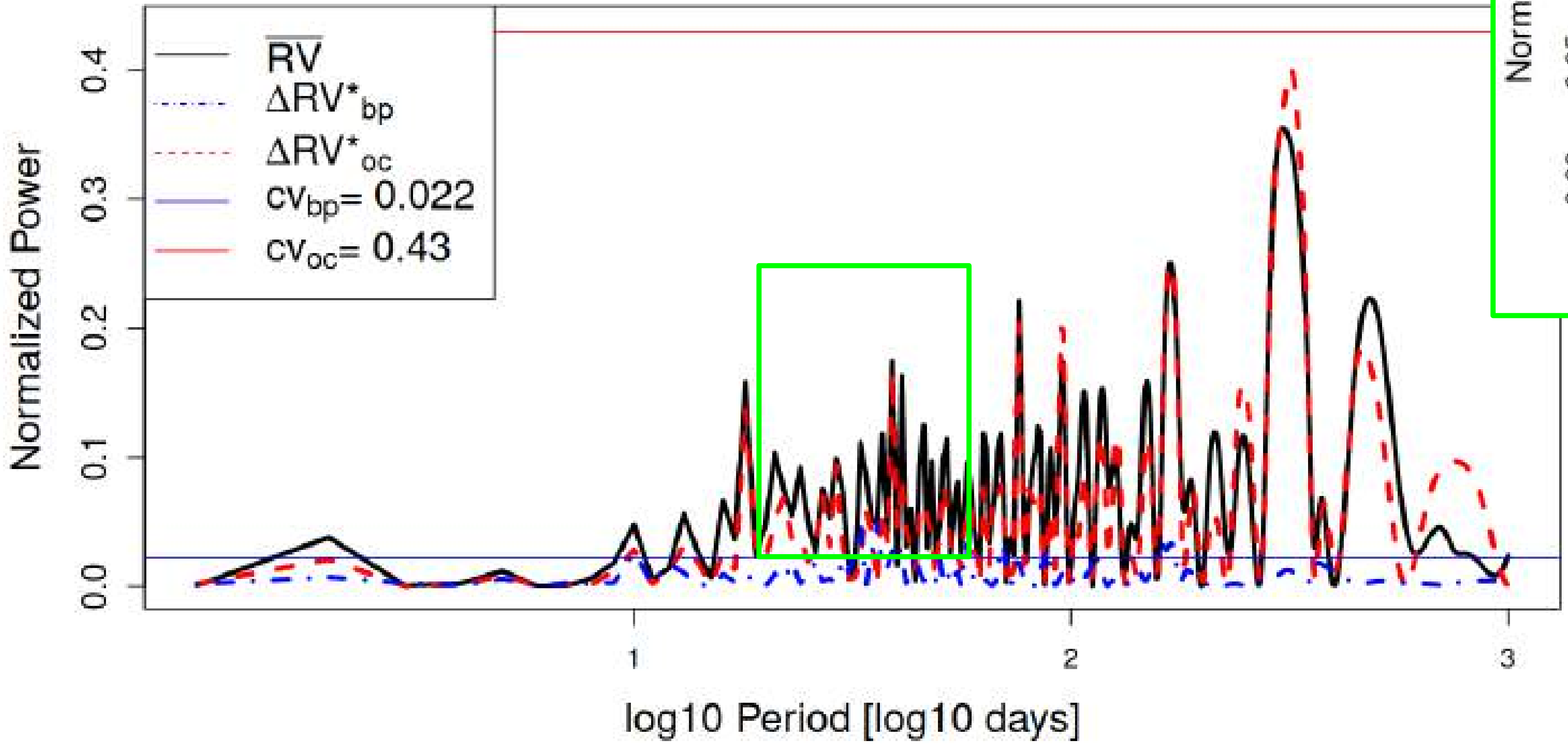
Original RV time series
oc-detrended residuals
bp-detrended residuals

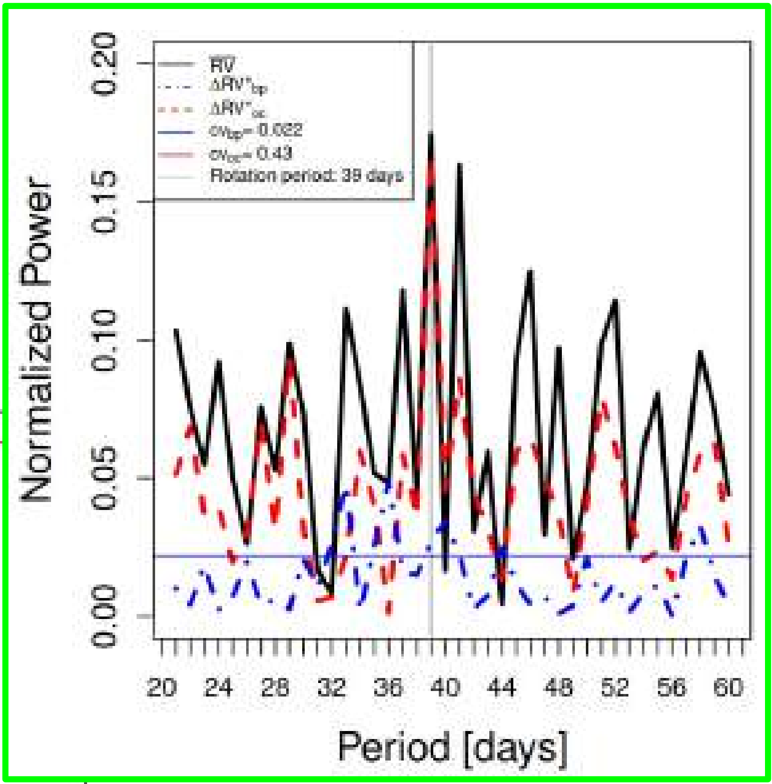


ÖAW IWF

bp vs. oc method: RV time series detrending

Generalised Lomb-Scargle (GLS) Periodogram of α Cen B



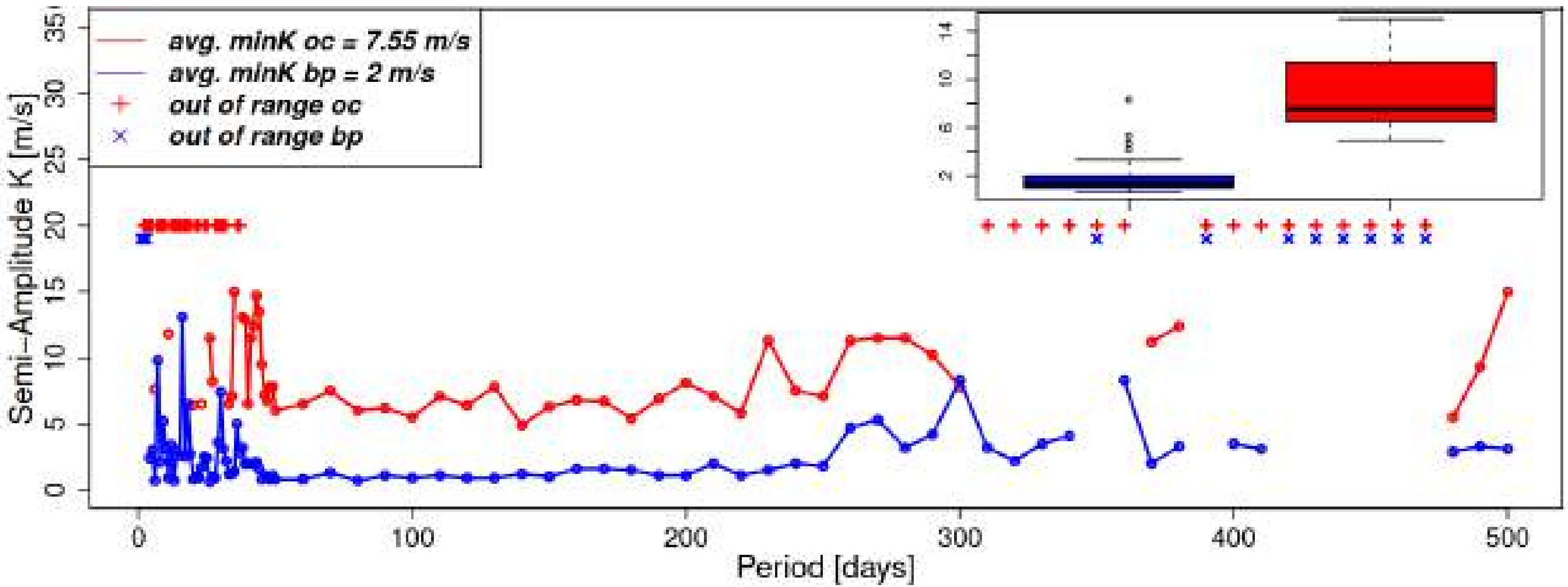


bp vs. oc method: detection of synthetic planets

- The RV time series of α Cen B is essentially made of stellar activity signals (e.g. Rajpaul et al. 2015)
- **We added artificial Keplerian signals** that would be generated by exoplanets in circular orbits **to the α Cen B RV time series**
- We applied both the **bp** and **oc** methods to clean the time series from stellar activity

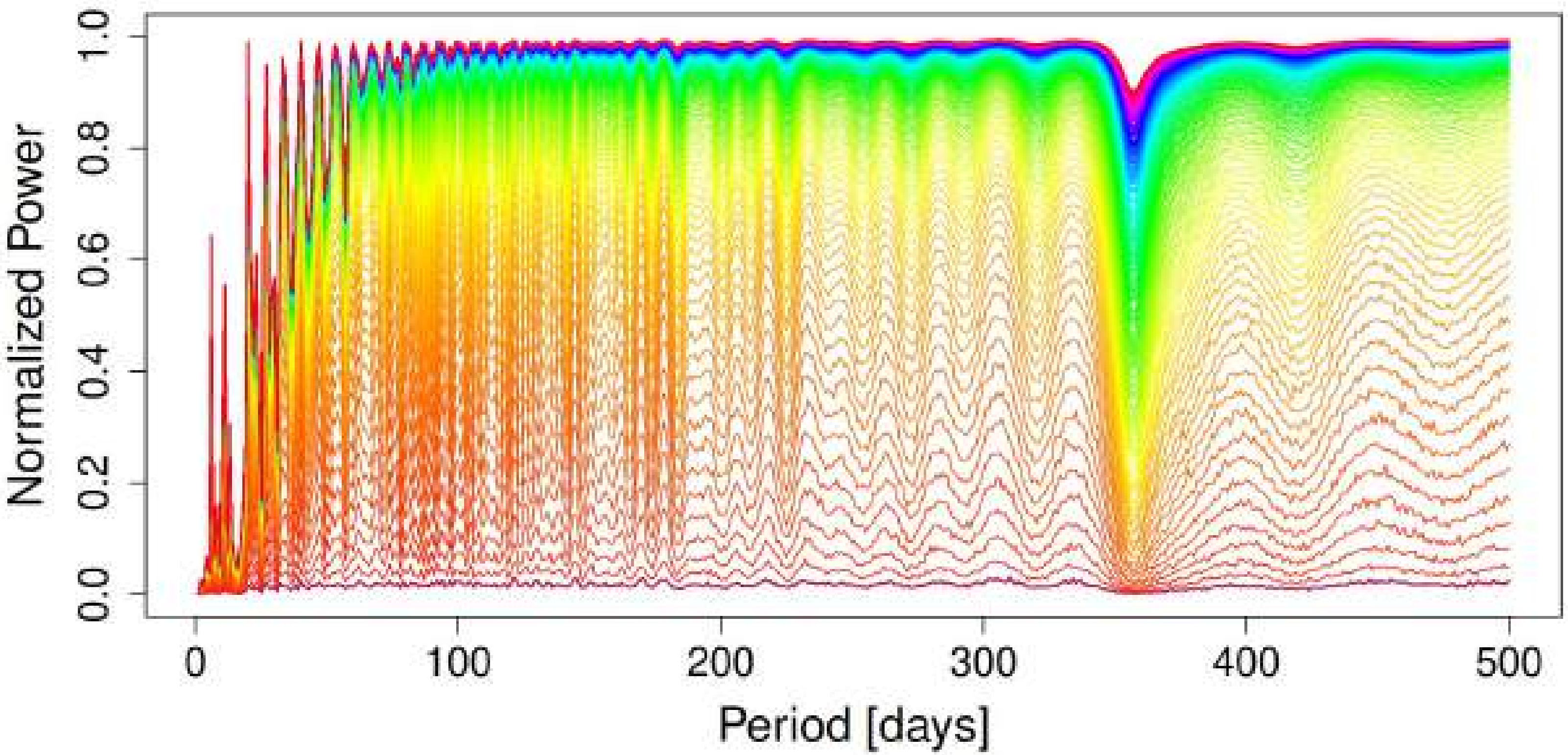
Up to which extent the two methods produce a clean time series where we can retrieve the synthetic signals of the simulated planets?

Minimum semiamplitude K at which a synthetic exoplanet of a given period is recovered from the two different clean time series



bp vs. oc method: detection of synthetic planets

- GLS periodograms of various synthetic planets
- Keplerian signals use the same sampling of the α Cen B time series
- *K* is colour-coded and increase from bottom to top



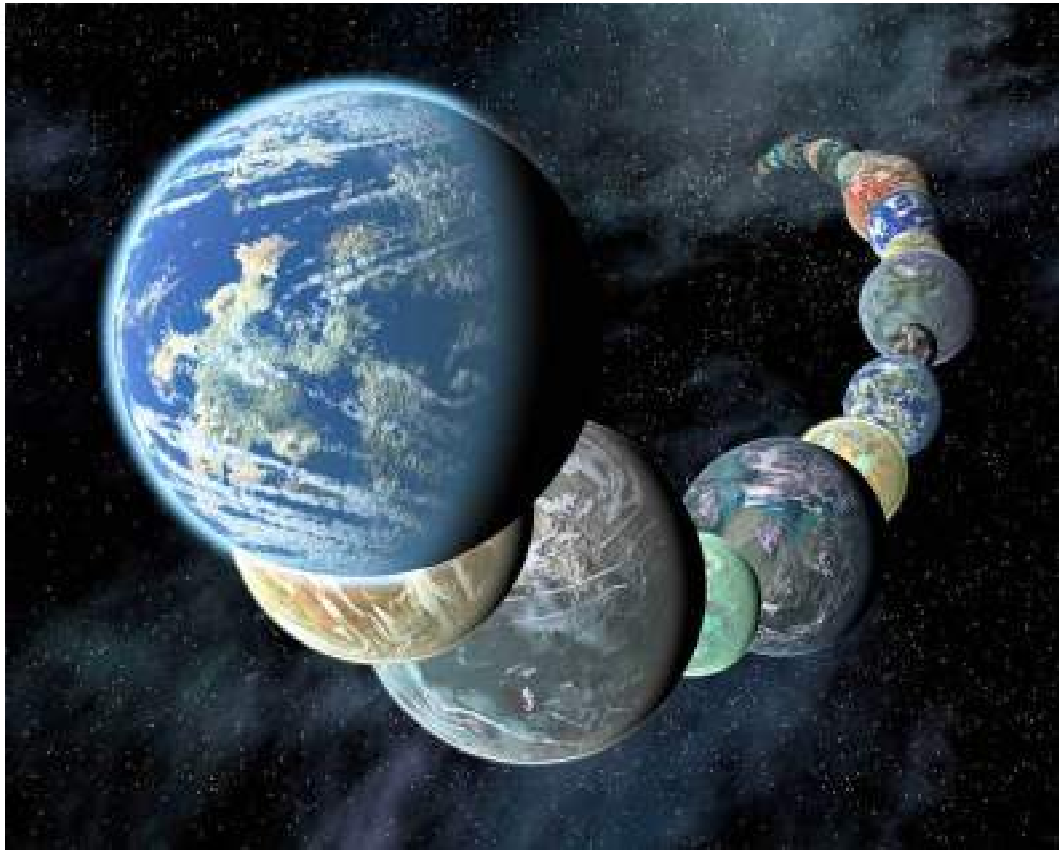
Systematic decrease of the power spectrum at some specific P_{orb} values, which is caused by the data sampling

A clear decrease occurs at P_{orb}~365 d as expected in ground-based observations

bp method: what about real exoplanets?

KESPRINT

- The bp technique was applied to HARPS time series of planet-hosting-stars also observed by CHEOPS
- Piecewise interpolation implemented within the MCMCI code (Bonfanti & Gillon, 2020) for a comprehensive RV + LC analysis of the system



About the RV time series:

- Polynomial detrending for removal of stellar activity
- Best values for the polynomial orders established according to the BIC minimisation criterion

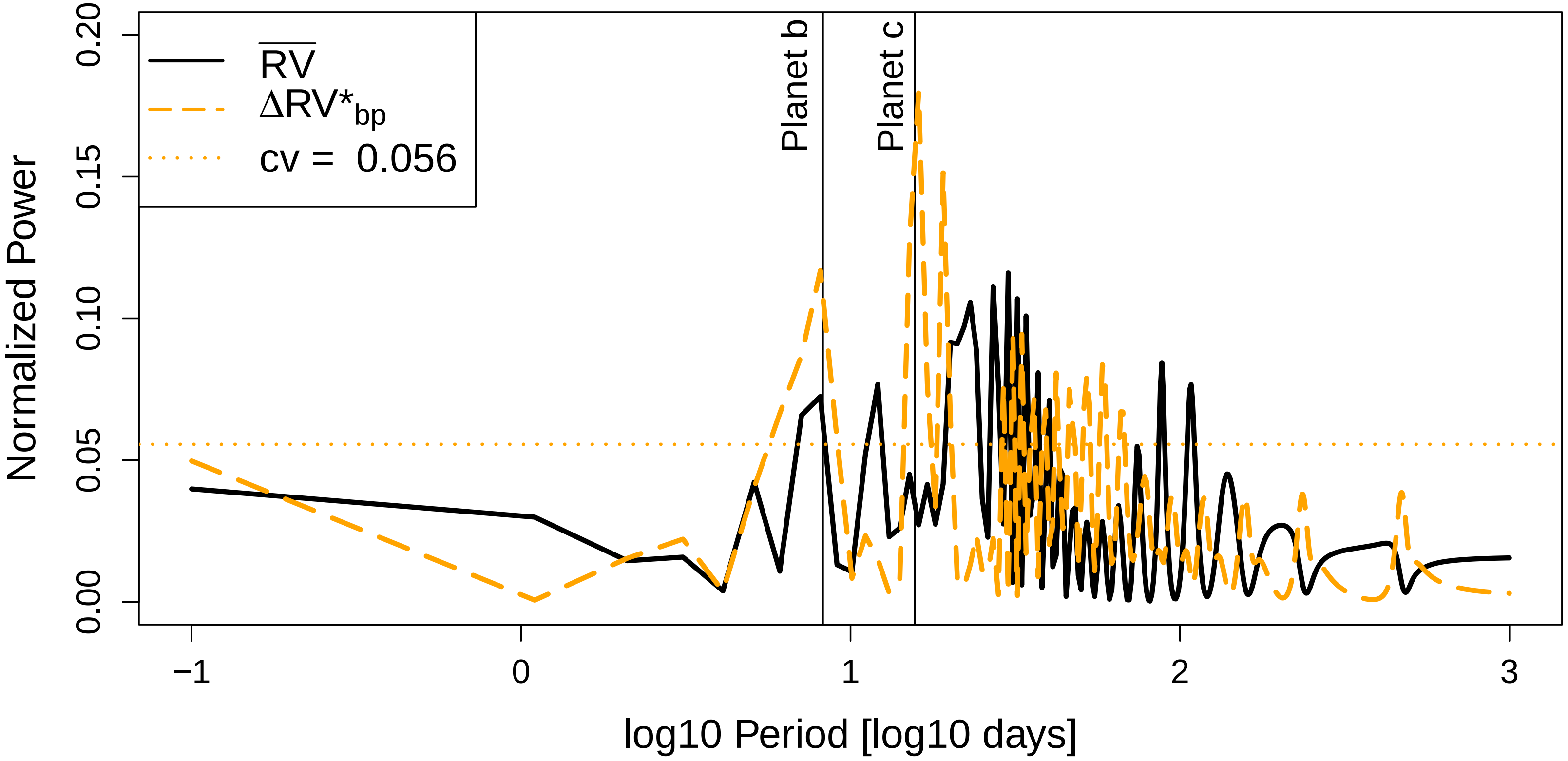
$$RV^*_{activity} = \beta_0 + \sum_k \beta_{t,k} \cdot t^k + \sum_k \beta_{F,k} \cdot FWHM_{SN}^k + \sum_k \beta_{\gamma,k} \cdot \gamma^k + \sum_k \beta_{A,k} \cdot A^k$$

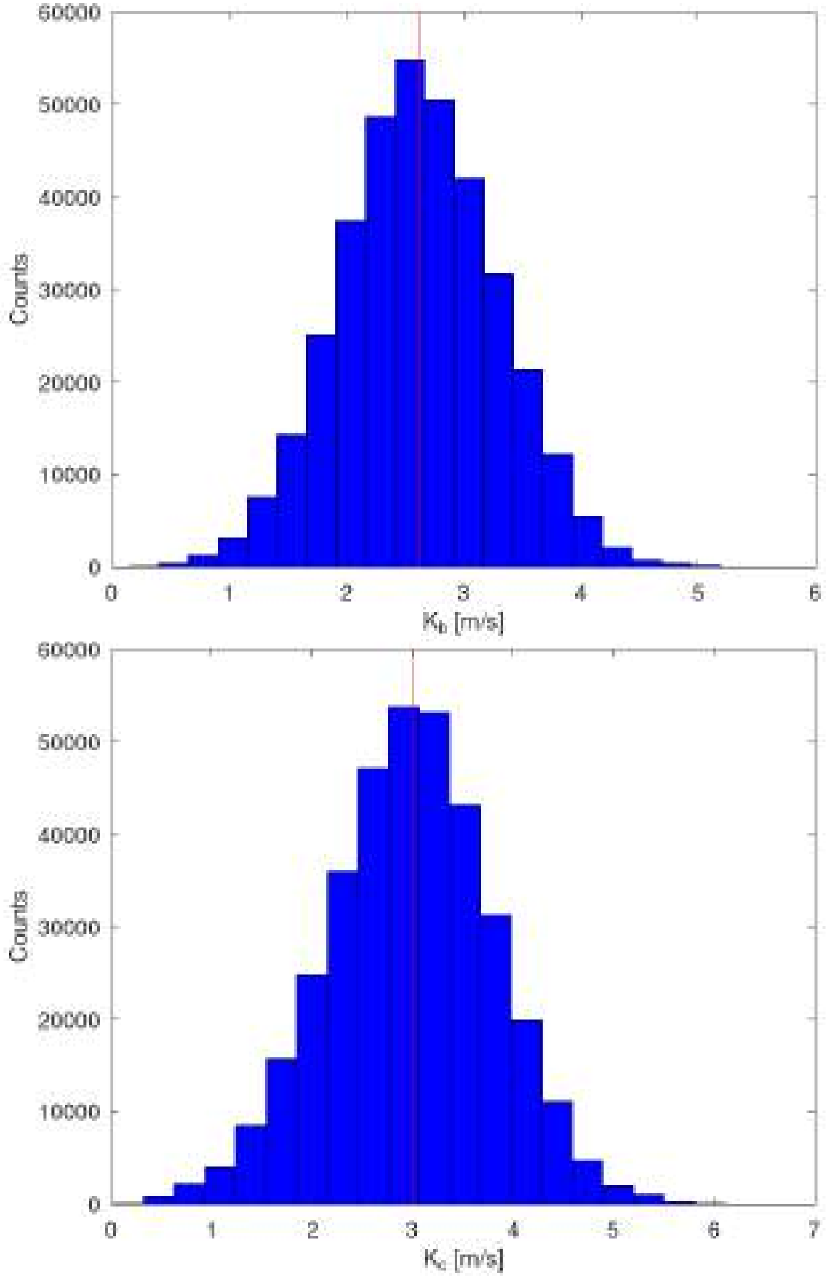
- The detrending is performed simultaneously during the MCMC fit

Breakpoint application to TOI-776

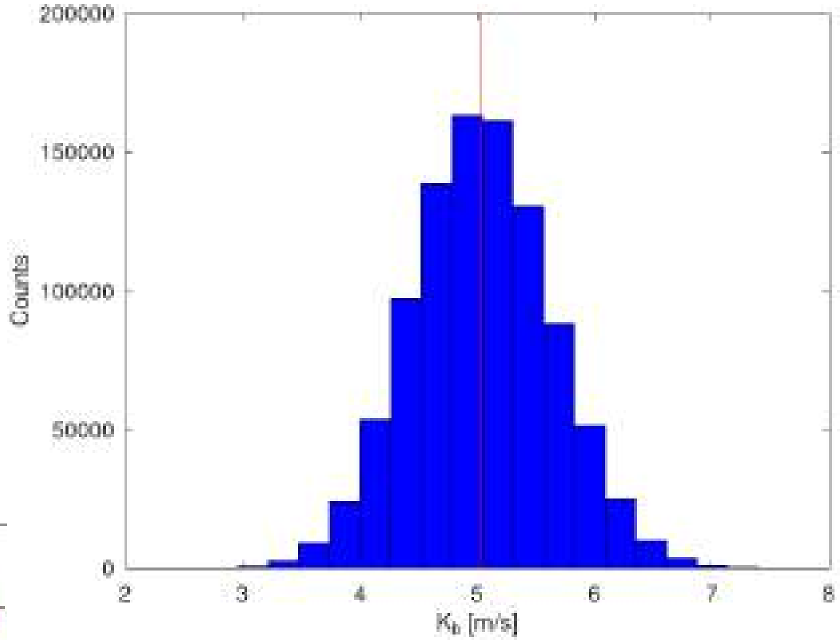
- Two transiting planets with known P_{orb}
- Two HARPS seasons, but…
  … standard RV reduction methods failed in analysing the 2nd season as data were highly contaminated by stellar activity

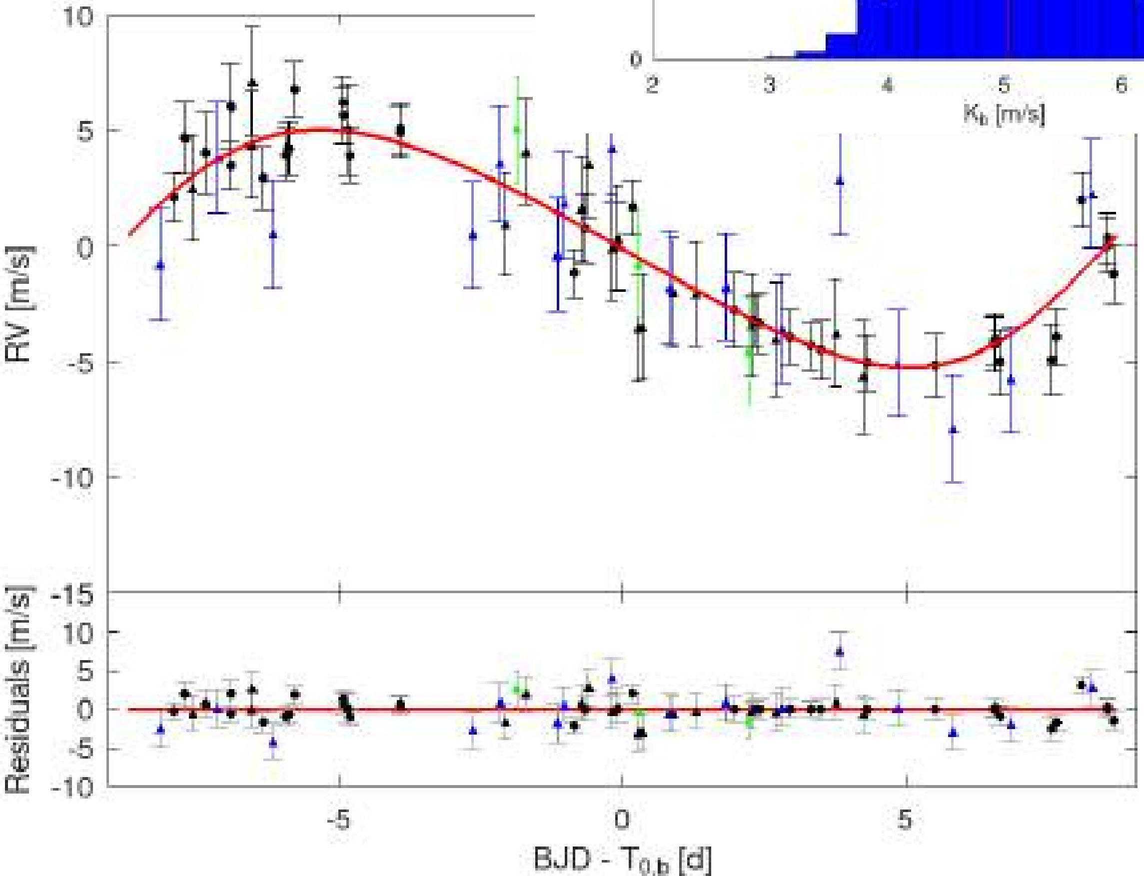
⇒ **Joint LC + RV MCMC analysis using the *bp* method**





Breakpoint application to TOI-1055





- One transiting planet with known P_{orb}
- Rather different estimates of planetary masses in the literature, despite using almost the same input data

Combining:

- More RV HARPS data
- TESS + CHEOPS LCs
- Our *bp* method

We obtain

- The most precise estimate of K and M_p so far
- The lowest jitter term
- ρ_p precise enough for meaningful planetary interior modelling

ÖAW IWF

Many thanks for your attention!!!

Any questions?