Bayesian model comparison in exoplanet data analysis

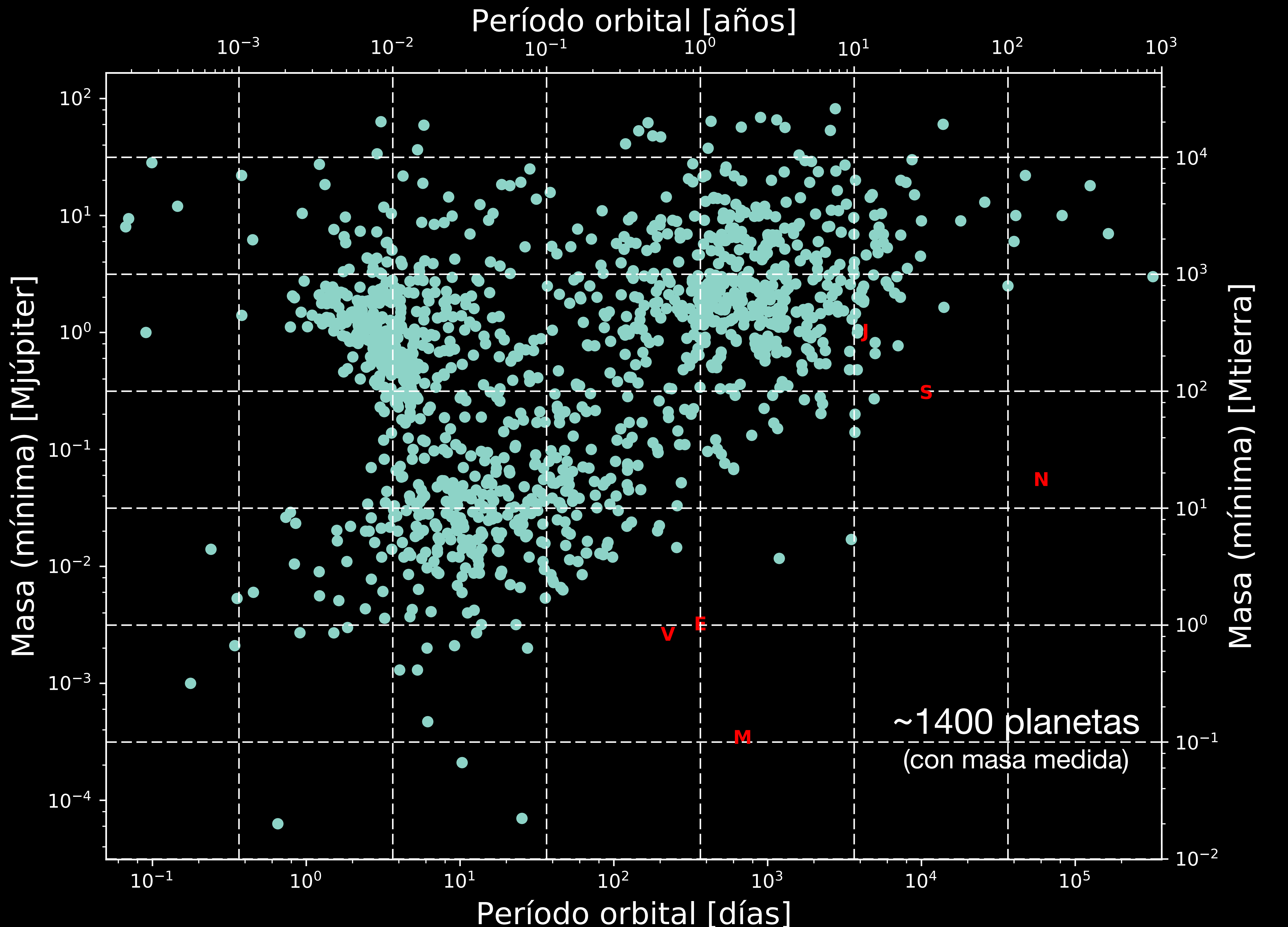
Rodrigo Díaz

rodrigo@iafe.uba.ar

Departamento de Física, FCEN, UBA, July 6th, 2017

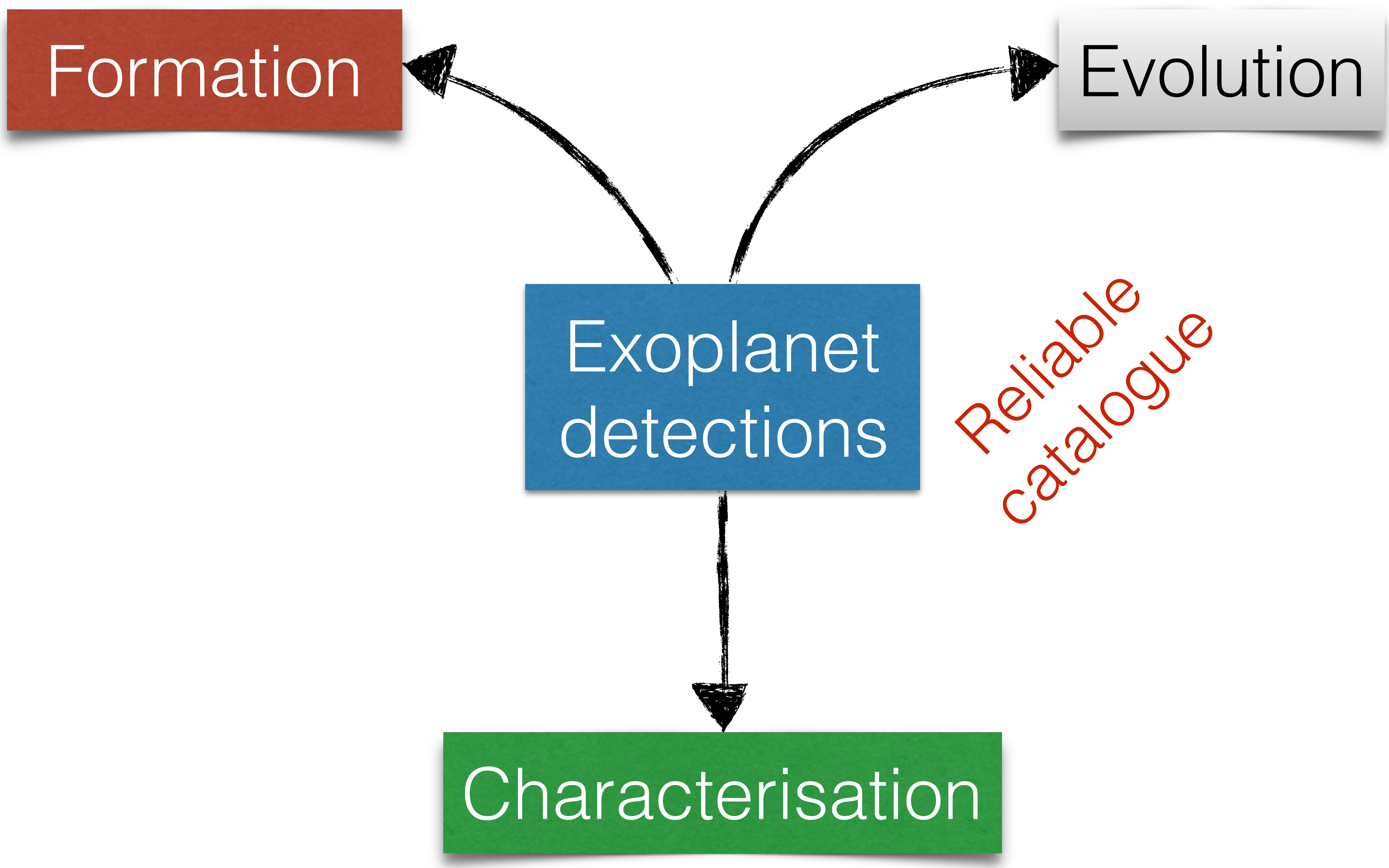
Extrasolar Planets

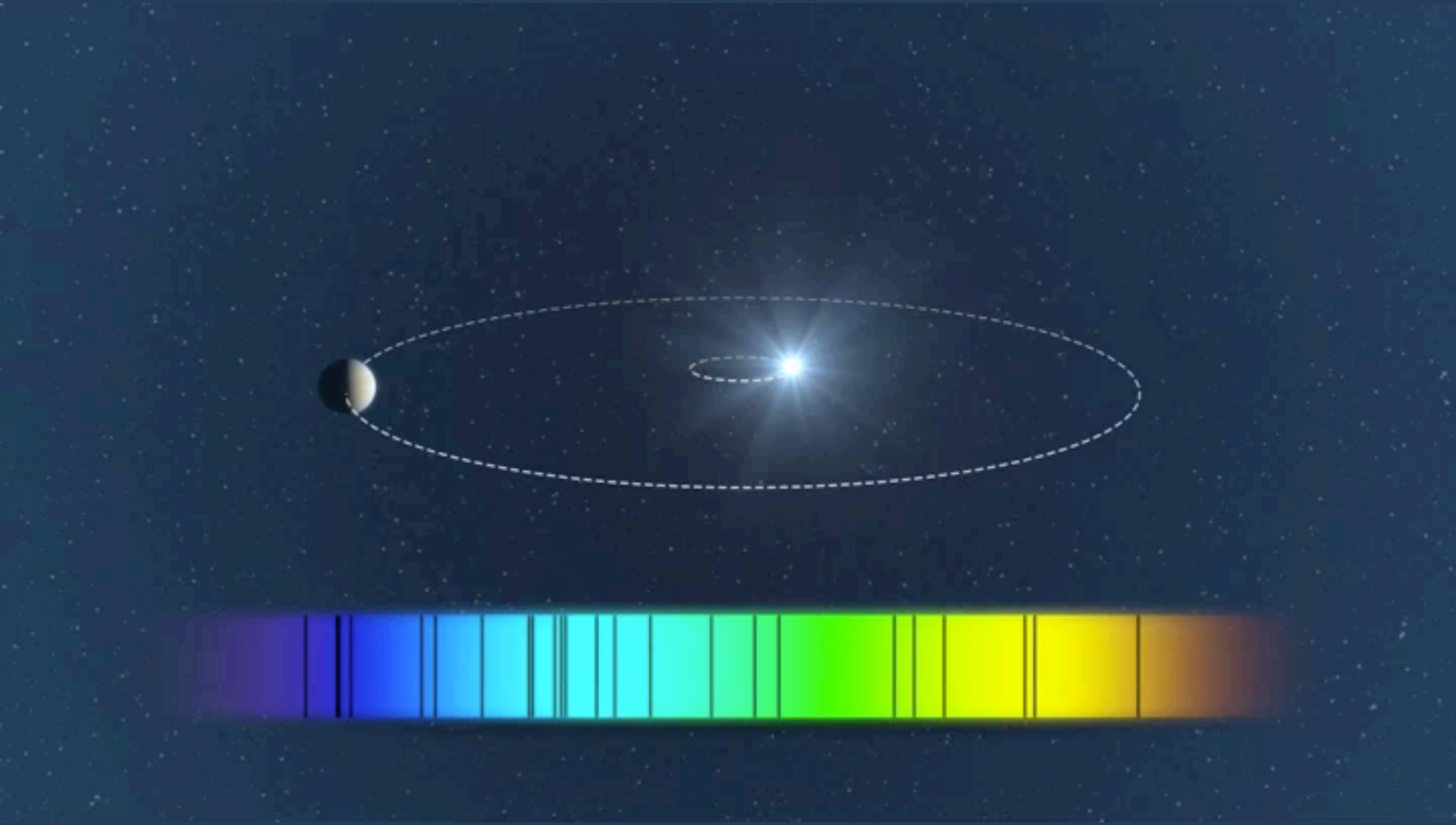
- Modelos de formación y evolución de planetas.
- Estructura interna de planetas gigantes y terrestres.
- Composición atmosférica. Dinámica atmosférica (exoclima).
- Mejor comprensión de la actividad estelar.
- Teoría de disipación por mareas.
- Censo galáctico.
- ¿Estamos solos?



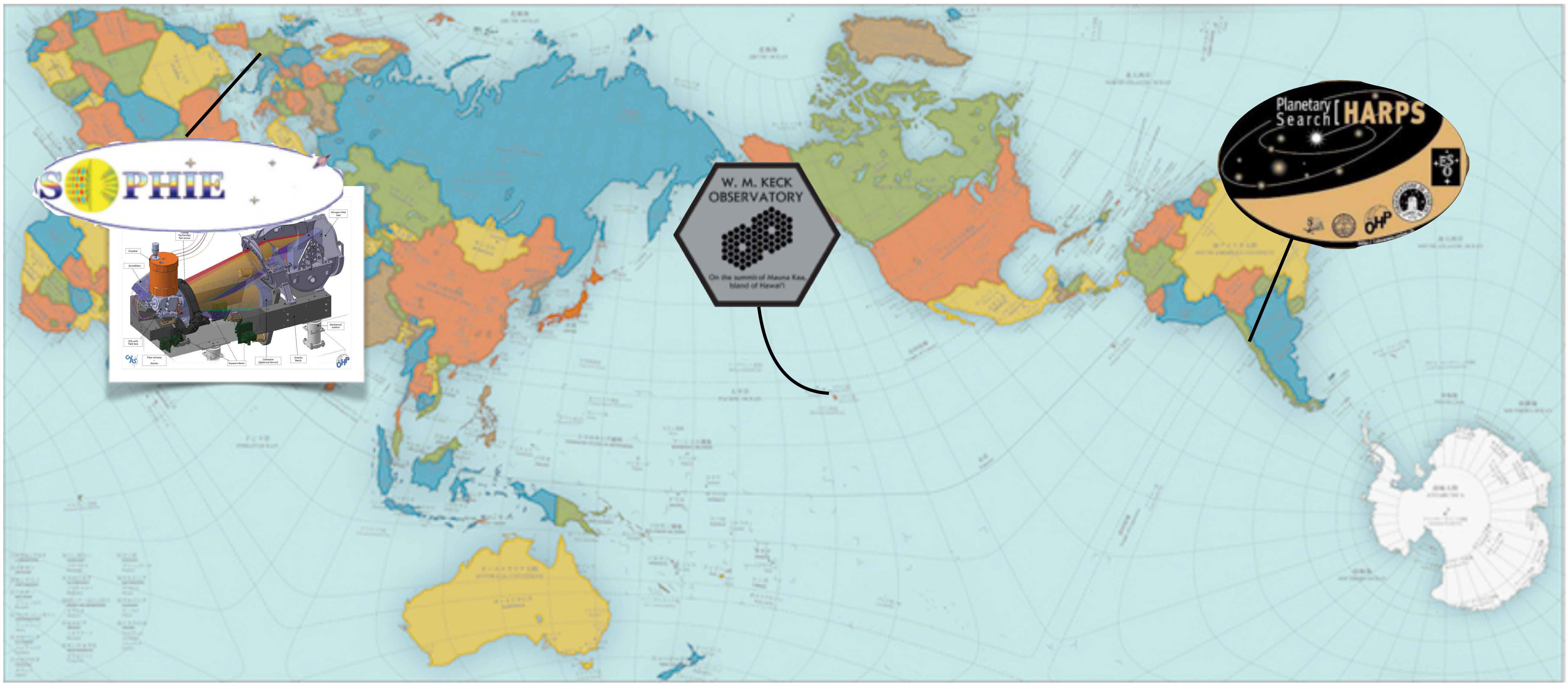
Data from exoplanets.eu (2016-09-14)

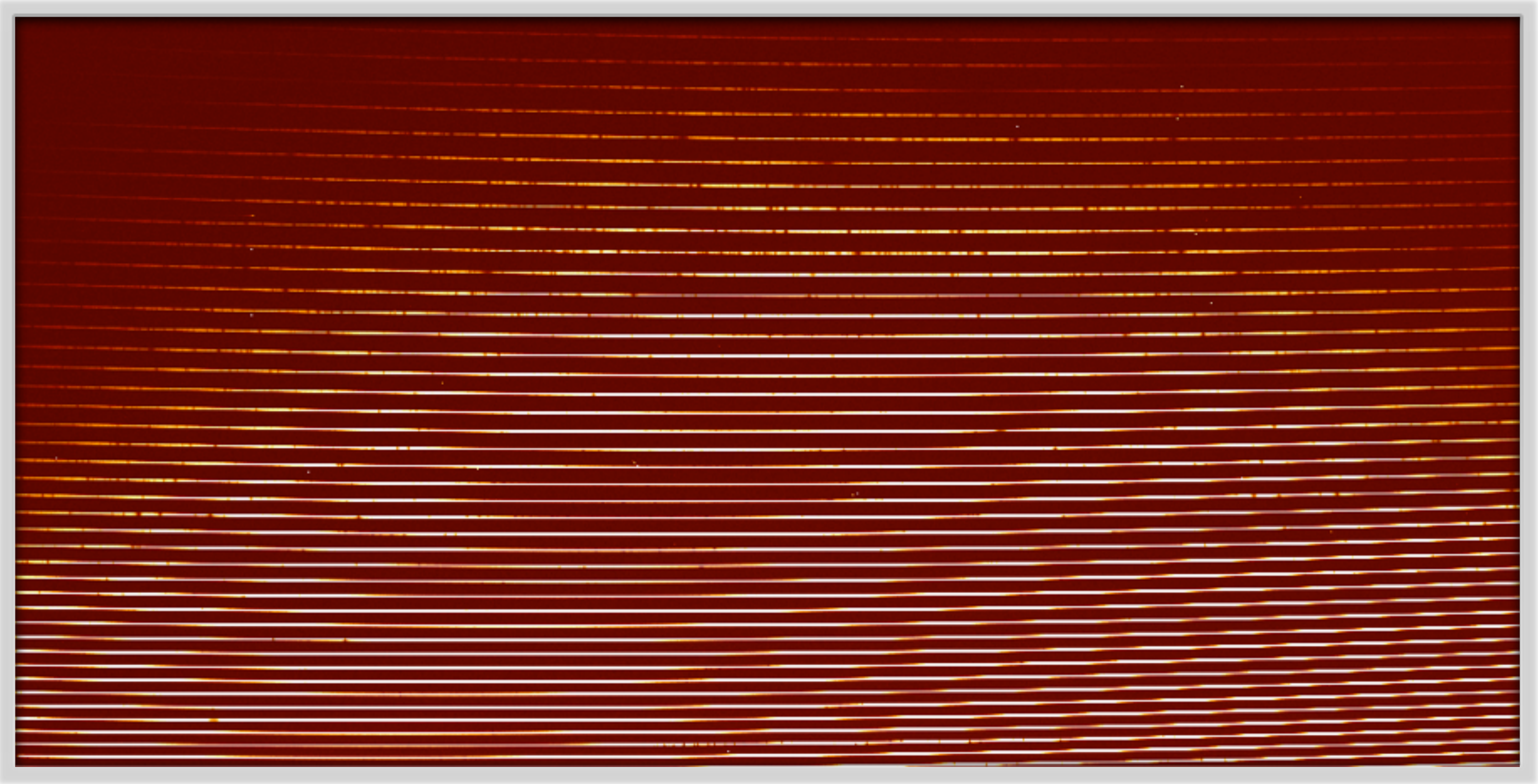
Exoplanets

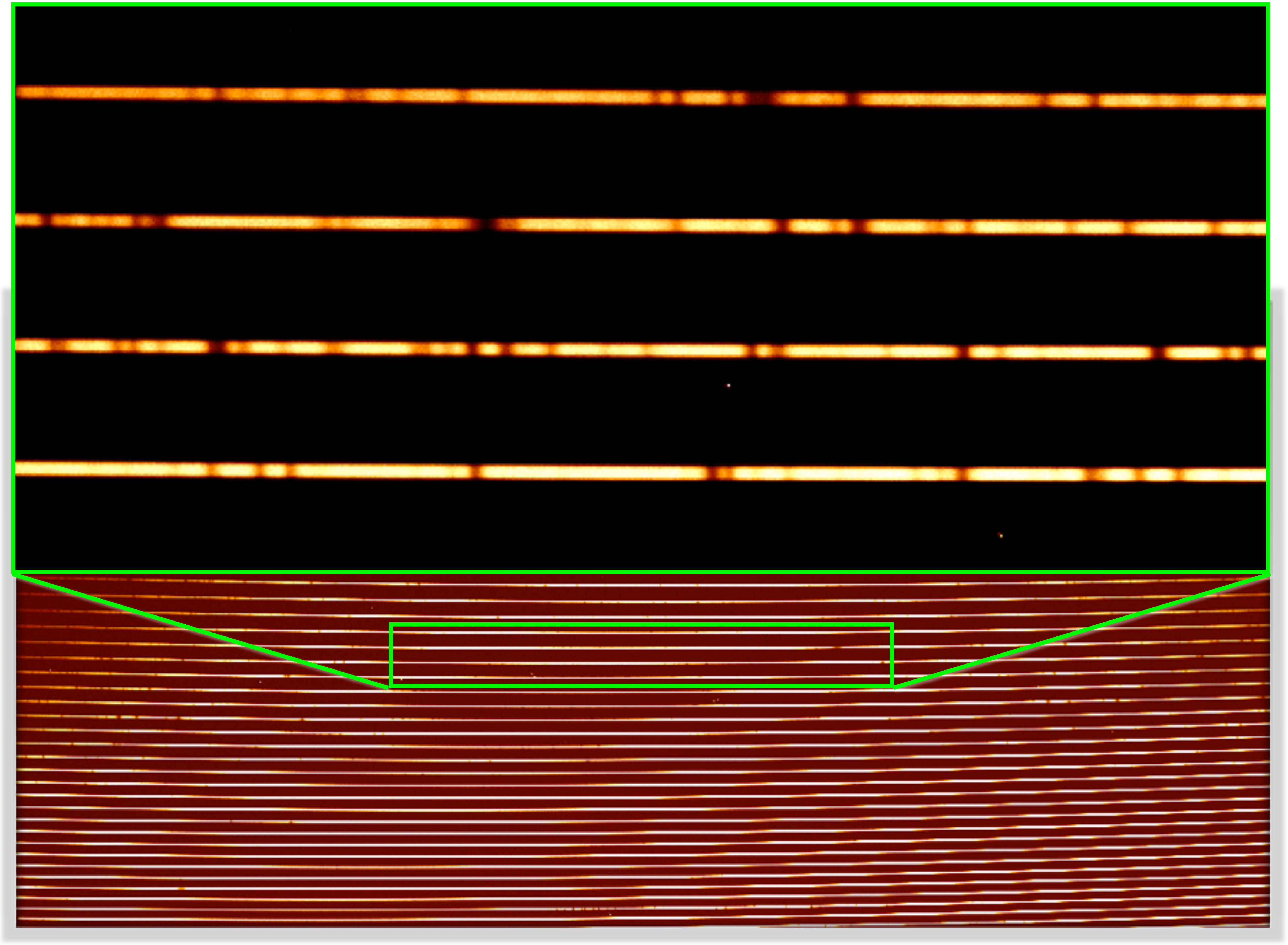

Formation
Evolution
Exoplanet detections
Reliable catalogue
Characterisation



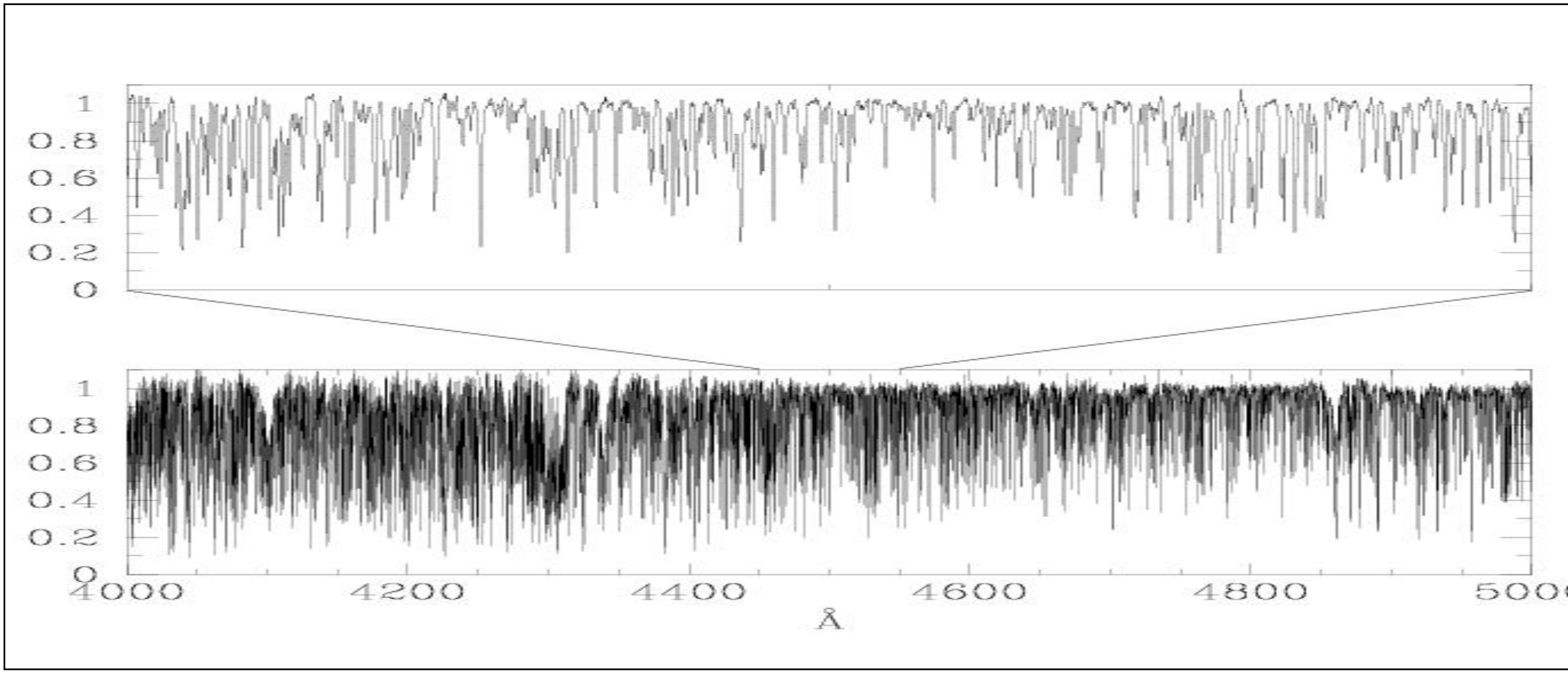
High-precision radial velocity surveys



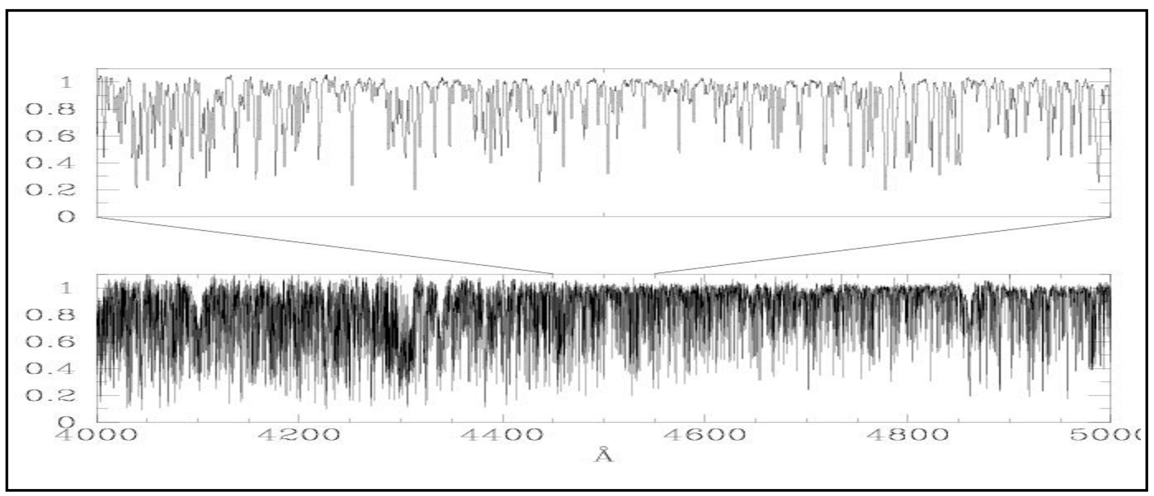


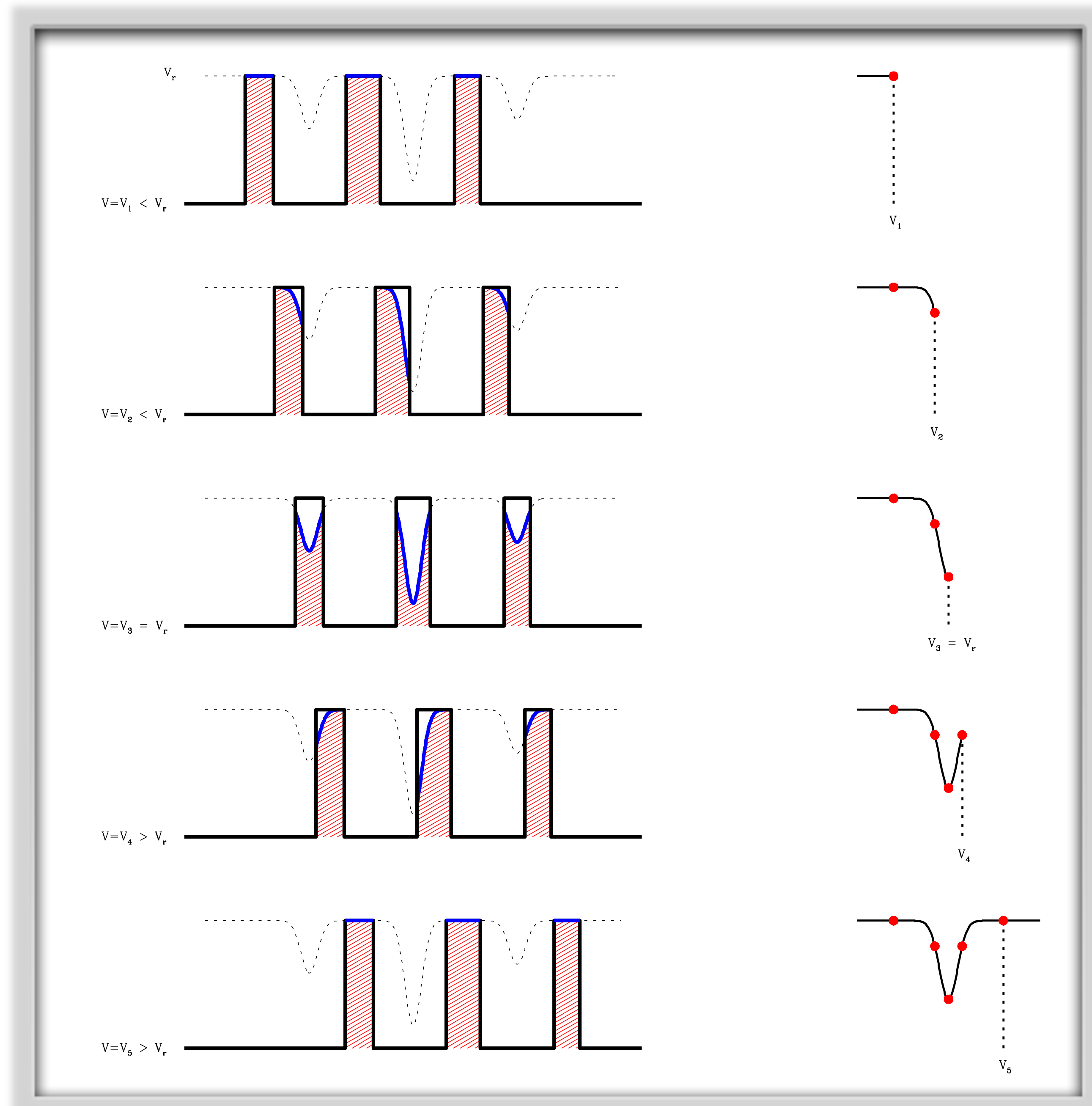


Cross correlation

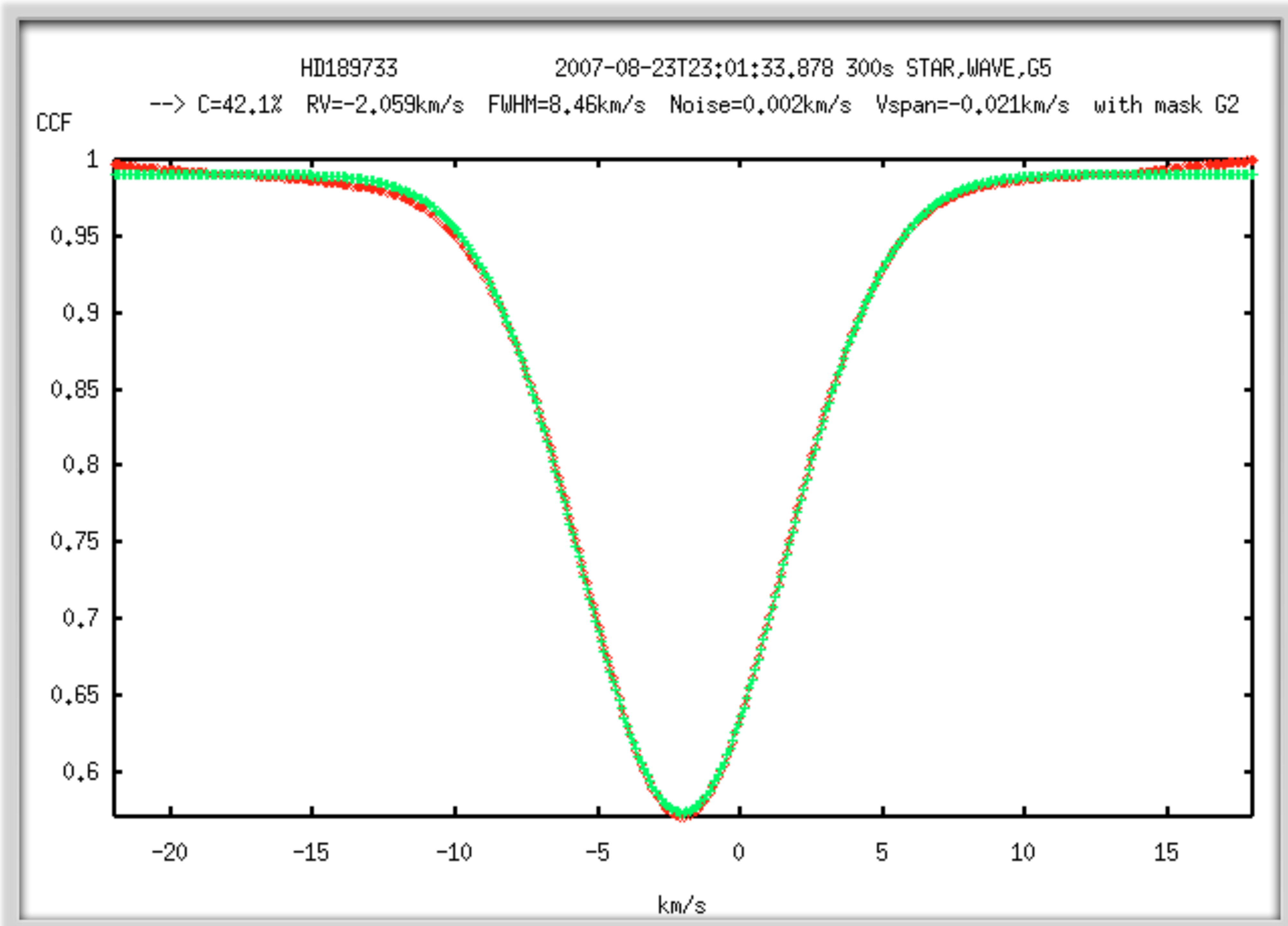


Cross correlation

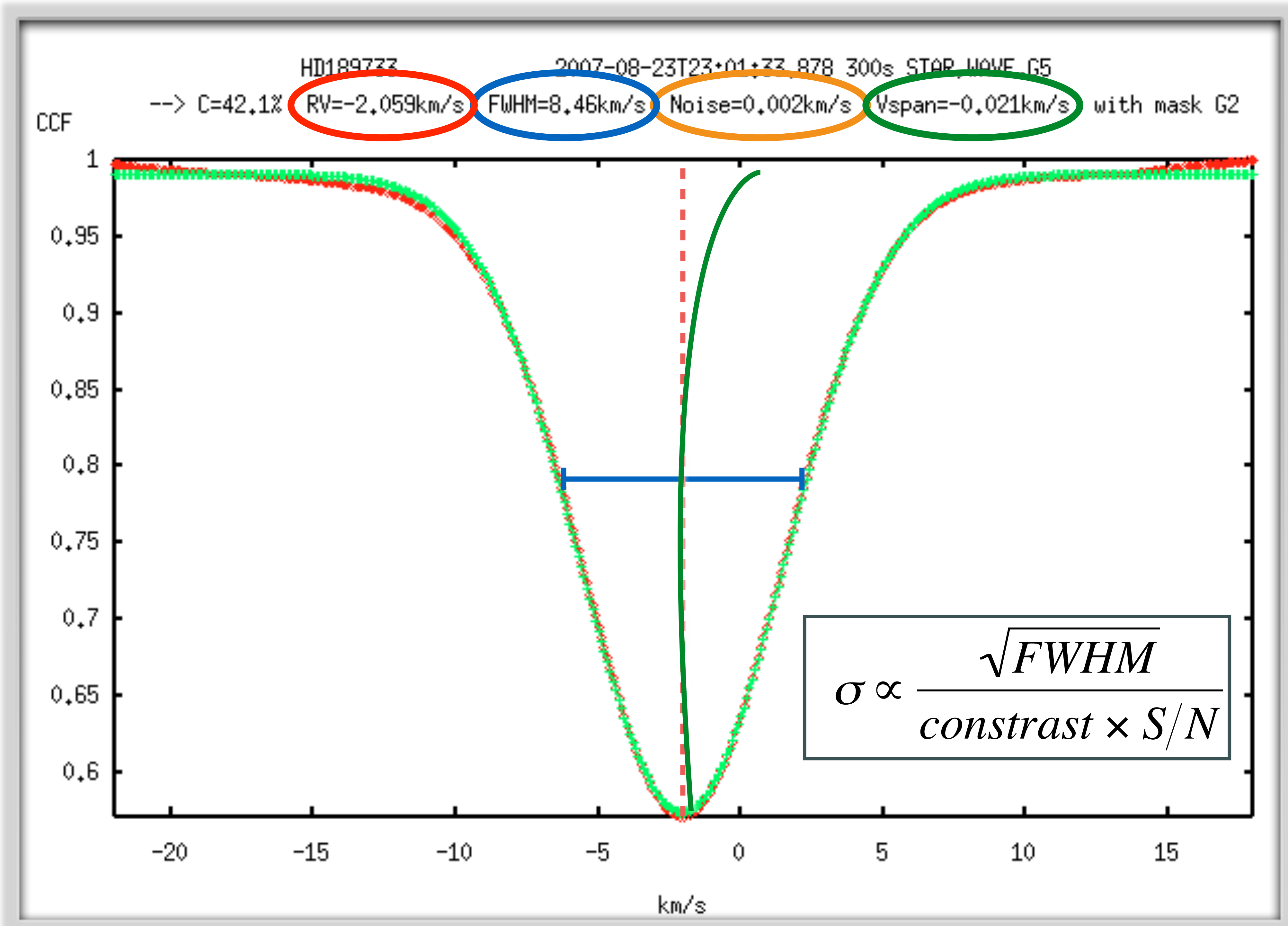




Cross correlation function

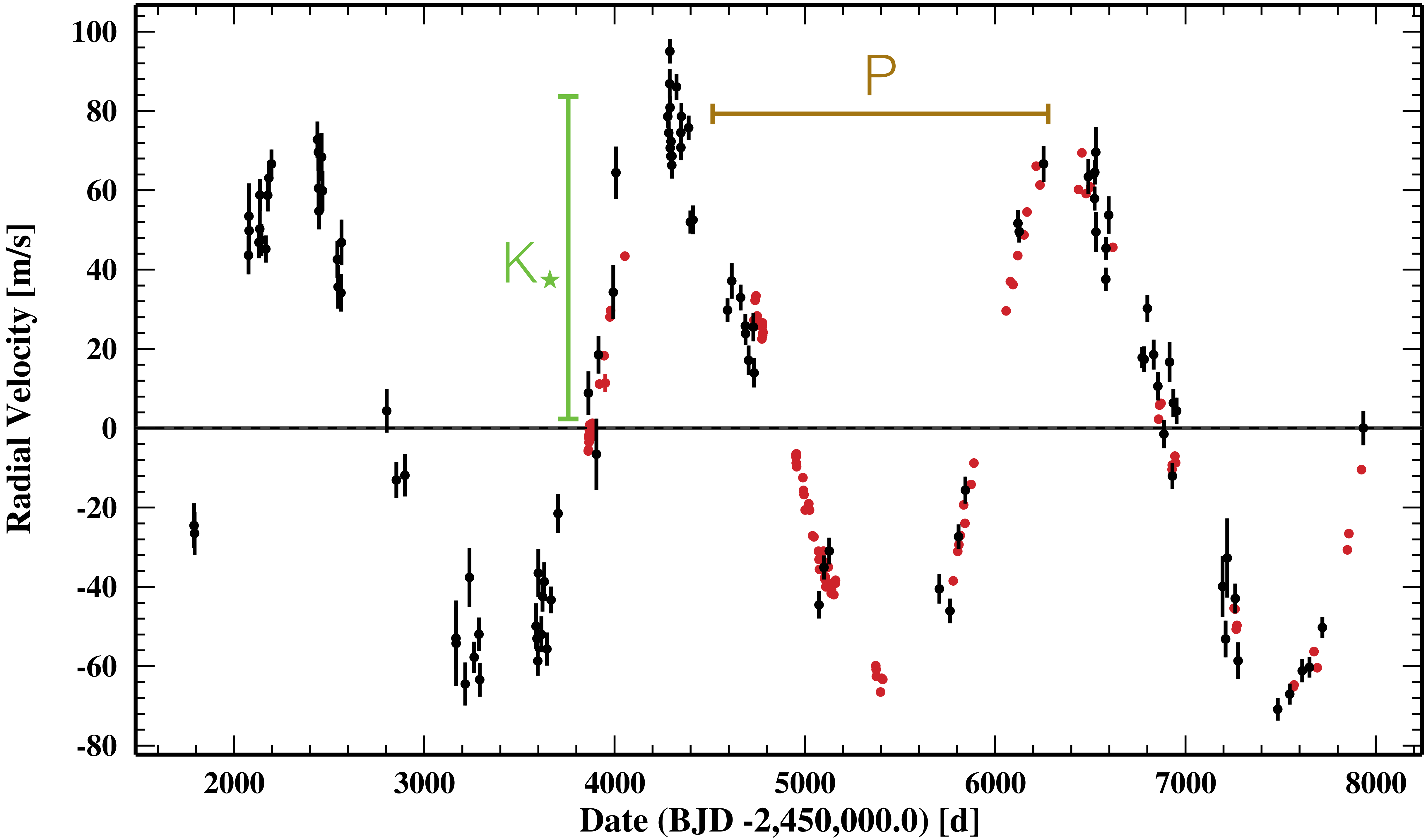


Cross correlation function



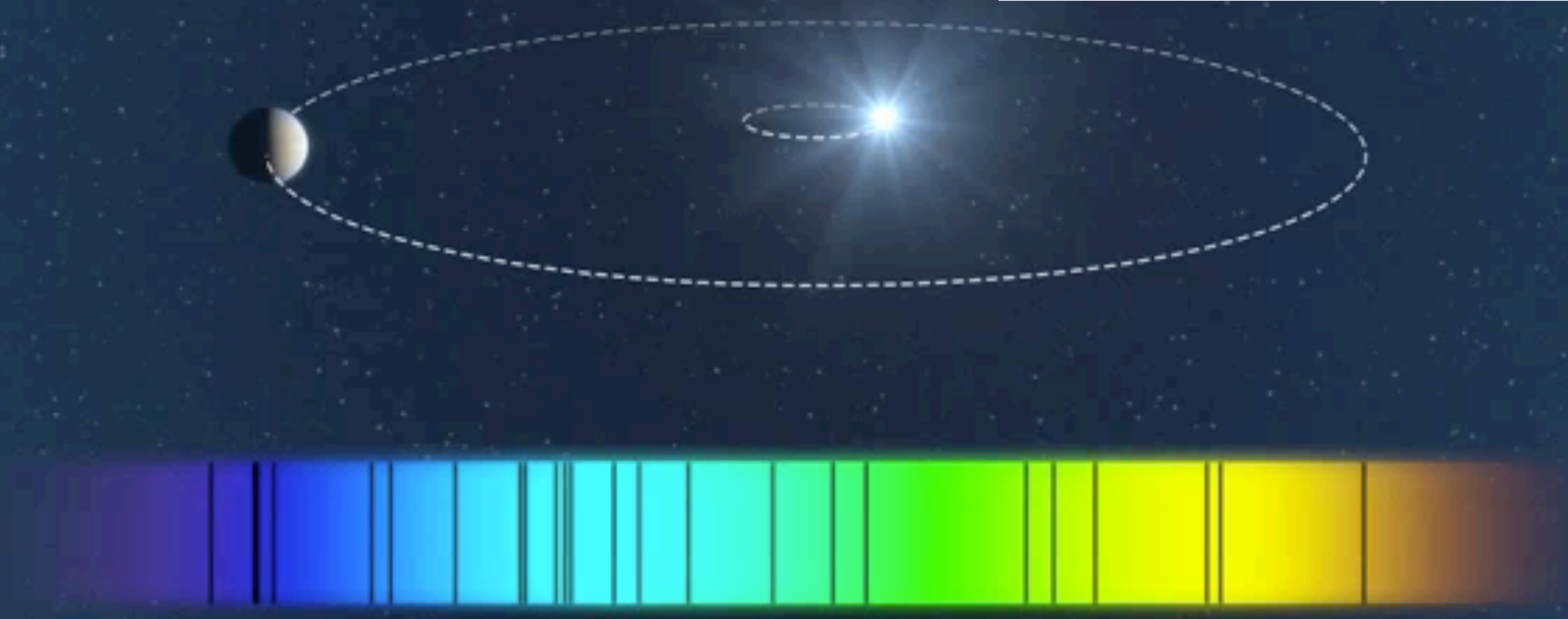
$$\sigma \propto \frac{\sqrt{FWHM}}{constrast \times S/N}$$

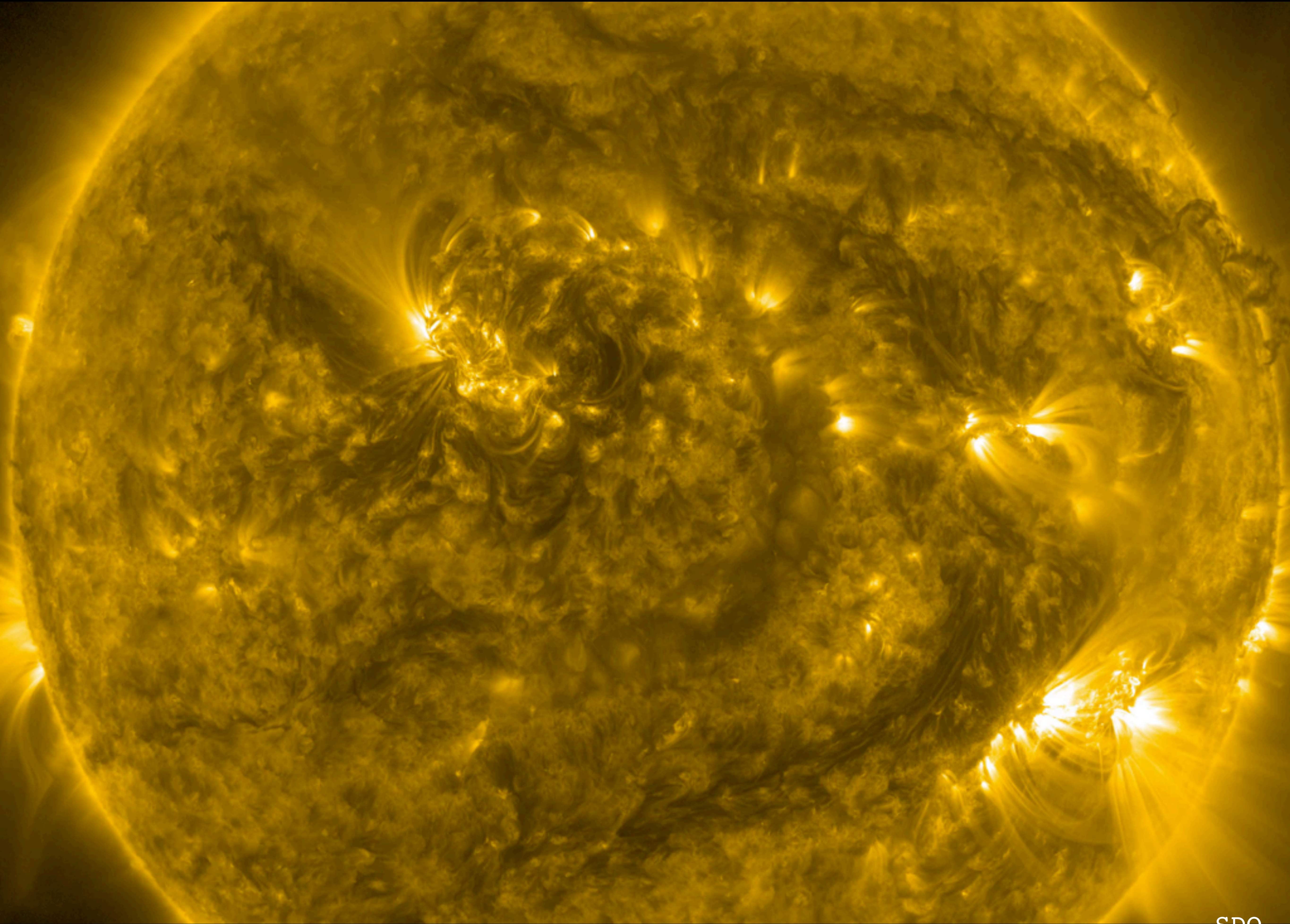
Radial velocity time series



$$K_\star \approx \left(\frac{2\pi G}{P}\right)^{1/3} \frac{1}{\sqrt{1-e^2}} \frac{m_2 \sin i}{m_1^{2/3}}$$

Planeta	Período	Amplitud VR
Júpiter	3 d	140 m/s
Júpiter	11.9 yr	12 m/s
Neptuno	3 d	7.6 m/s
Tierra	3 d	44 cm/s
Tierra	1 yr	9 cm/s

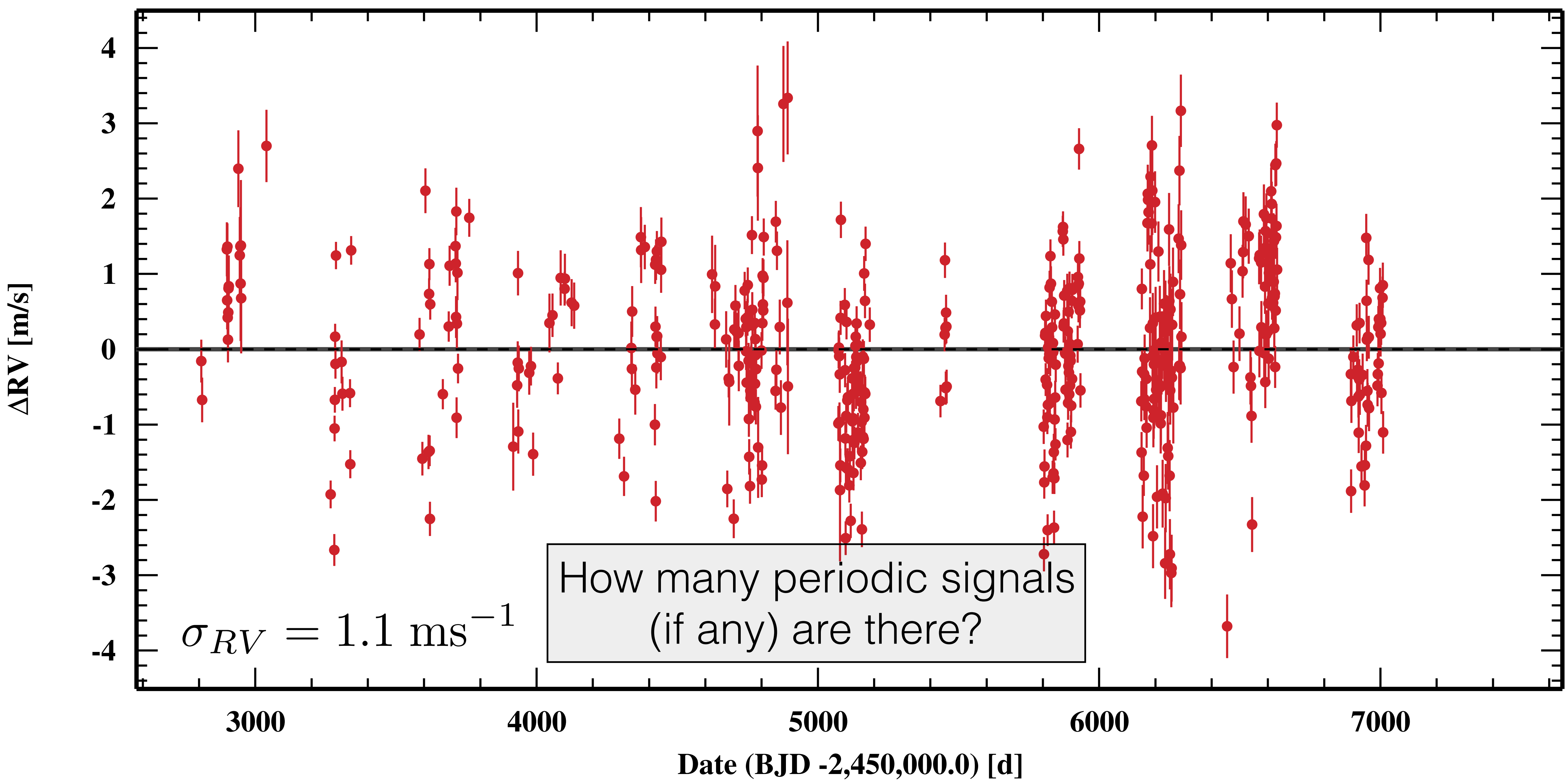




SDO

State-of-the-art radial velocities

HARPS dataset on Tau Ceti: m/s-precision over 11.5 years



Traditional approach: the GLS periodogram

Generalised Lomb Scargle (GLS): Sort of Fourier Transform for unevenly sampled data.

$$p(\omega) = \frac{\chi_0^2 - \chi_\omega^2}{\chi_0^2}$$

χ^2 for the weighted mean

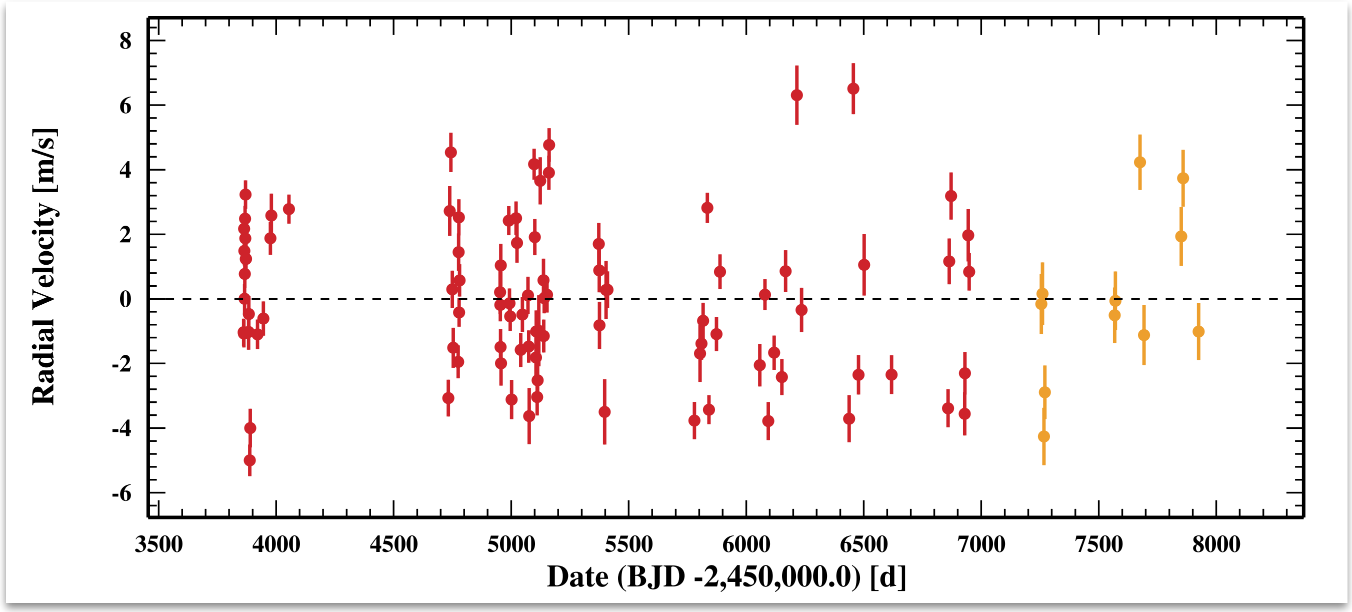
χ^2 for a linear sinusoidal model

$$A\sin(\omega t) + B\cos(\omega t) + C$$

Traditional approach: the GLS periodogram

Generalised Lomb Scargle (GLS): Sort of Fourier Transform for unevenly sampled data.

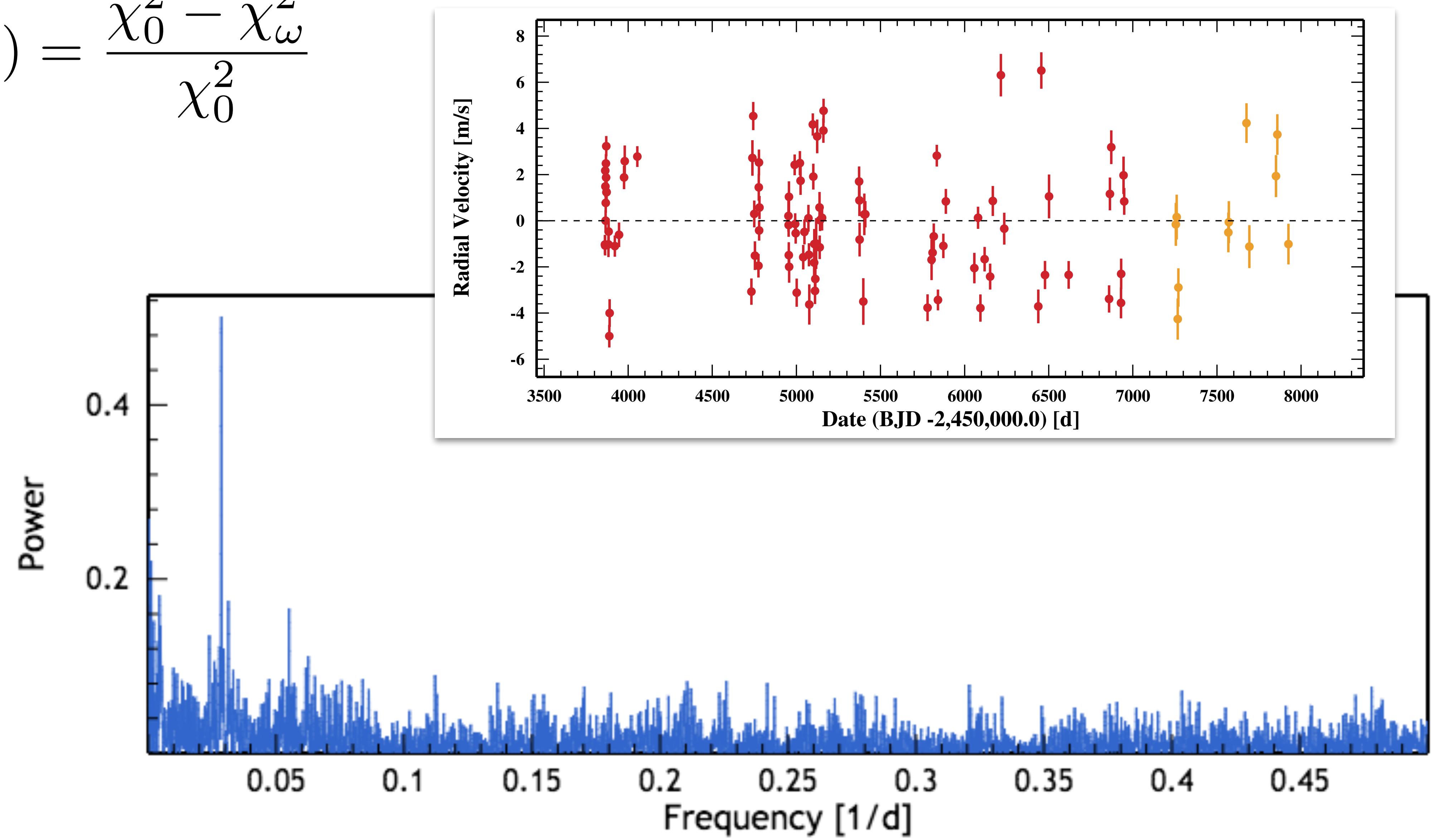
$$p(\omega) = \frac{\chi_0^2 - \chi_\omega^2}{\chi_0^2}$$



Traditional approach: the GLS periodogram

Generalised Lomb Scargle (GLS): Sort of Fourier Transform for unevenly sampled data.

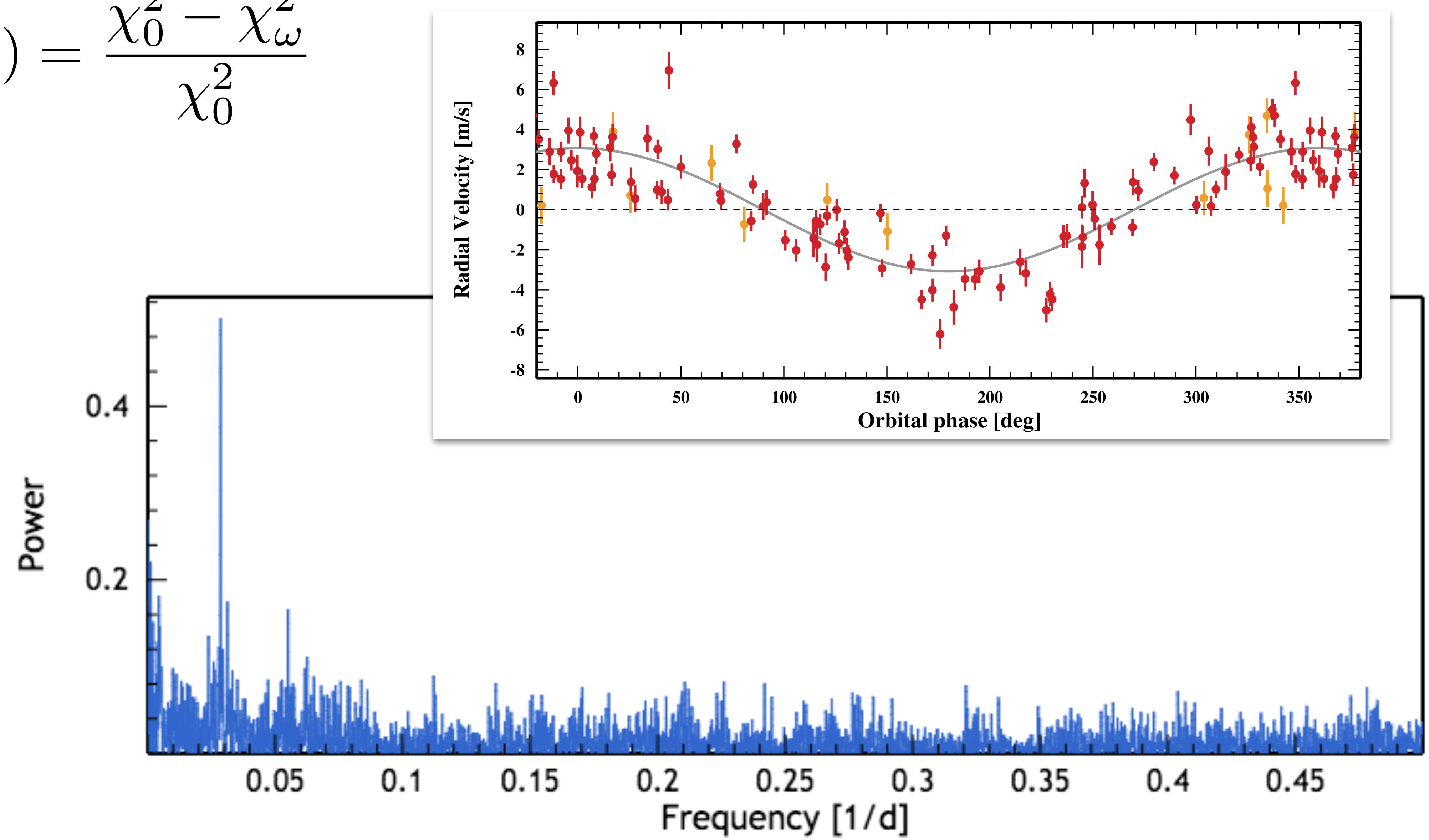
$$p(\omega) = \frac{\chi_0^2 - \chi_\omega^2}{\chi_0^2}$$



Traditional approach: the GLS periodogram

Generalised Lomb Scargle (GLS): Sort of Fourier Transform for unevenly sampled data.

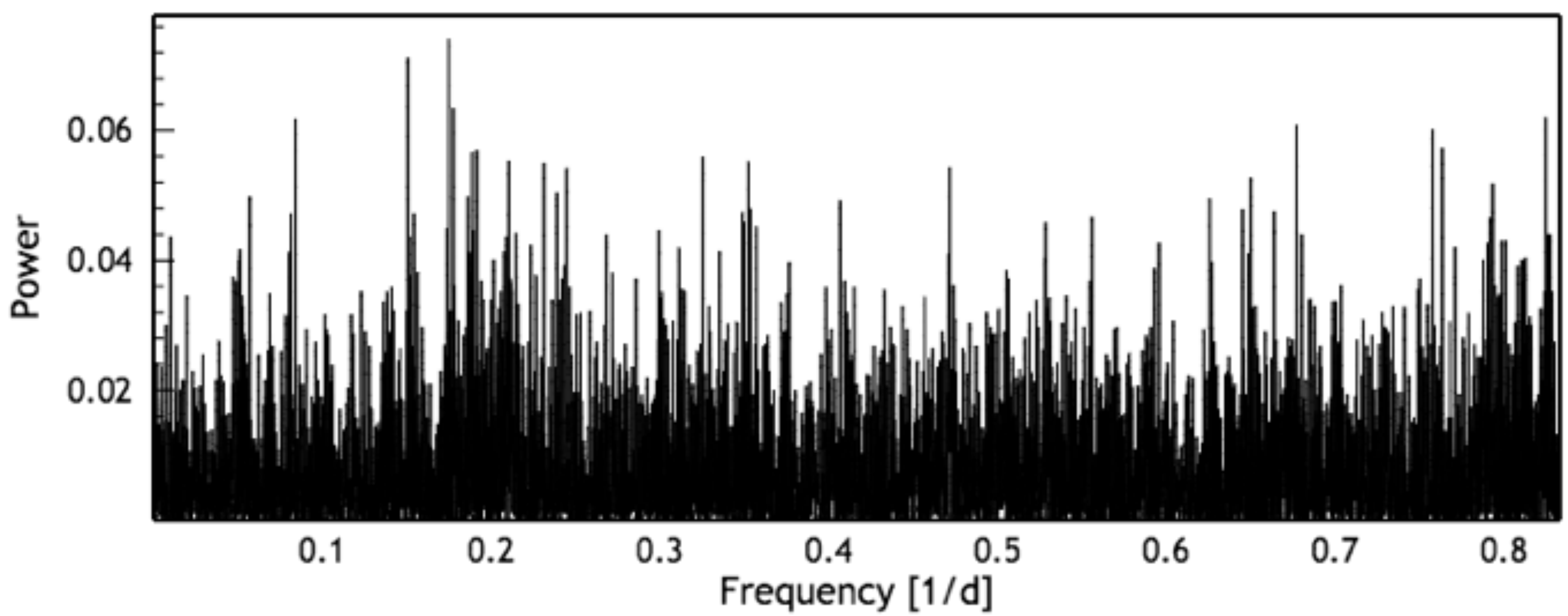
$$p(\omega) = \frac{\chi_0^2 - \chi_\omega^2}{\chi_0^2}$$



GLS periodogram: peak significance

Generalised Lomb Scargle (GLS): Sort of Fourier Transform for unevenly sampled data.

$$p(\omega) = \frac{\chi_0^2 - \chi_\omega^2}{\chi_0^2}$$



Widely used statistic for hypothesis test is $\max_{\omega} [p(\omega)] = \max_{\omega} \left[\frac{\chi_0^2 - \chi_\omega^2}{\chi_0^2} \right]$

GLS periodogram: peak significance

$$p(\omega) = \frac{\chi_0^2 - \chi_\omega^2}{\chi_0^2}$$

If errors are Gaussian, Lomb (1976) demonstrated that $p(\omega)$ is distributed as χ^2 with 2 degrees of freedom.

$$\Pr\left[(p(\omega) > z)\right] = \exp(-z)$$

If N independent frequencies are tested, the probability that they are all below z is

$$\left[1 - \exp(-z)\right]^N$$

Then, the probability that at least one peak is above z (called False Alarm Probability) is

$$\mathrm{FAP} = 1 - \left[1 - \exp(-z)\right]^N$$

GLS periodogram: peak significance

Many (good) reasons _not_ to trust the Gaussian assumption:

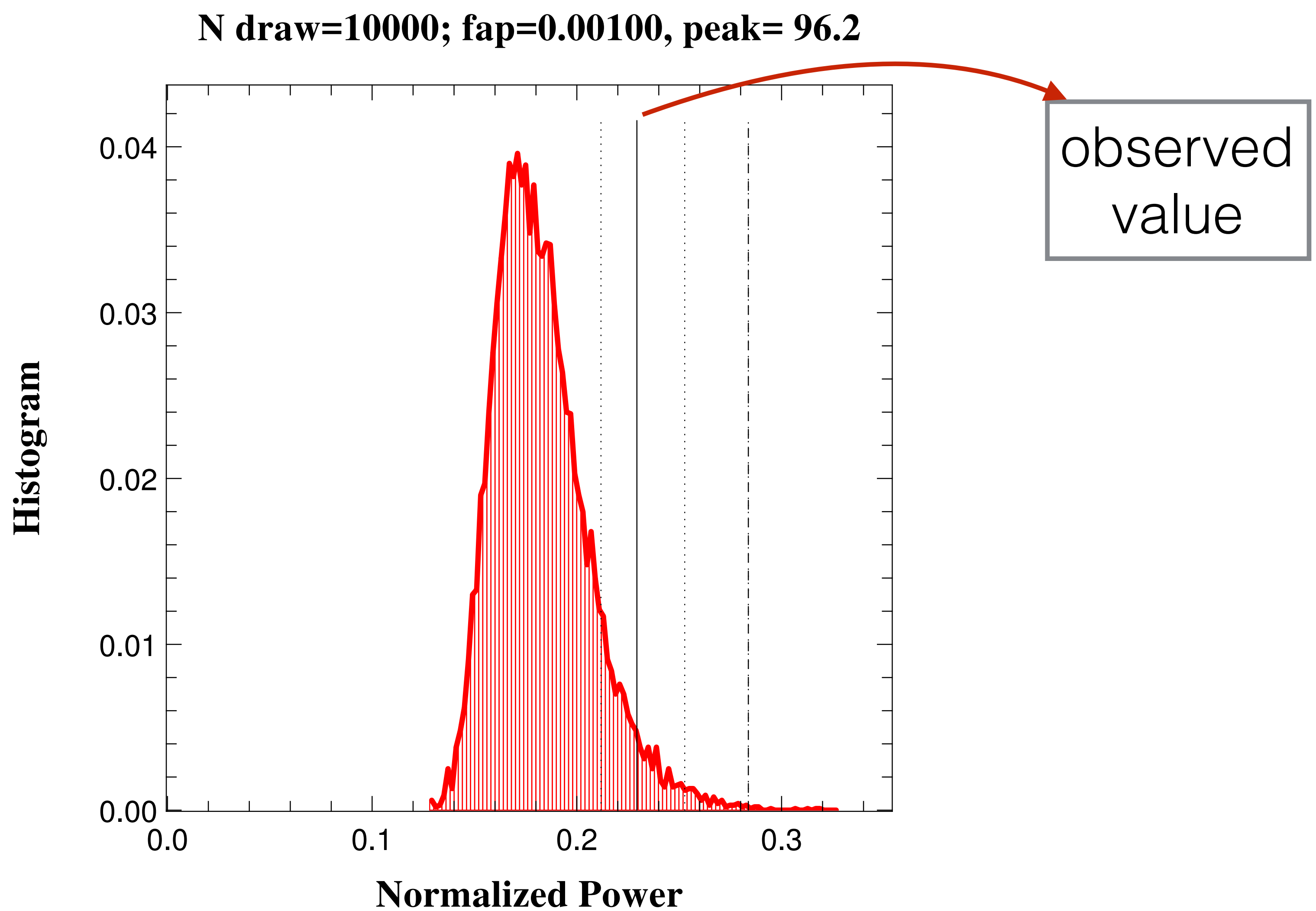
- the data contains a signal.
- errors are under- or overestimated.
- intrinsic variability is present.

Monte Carlo simulations to estimate distribution of the test statistic.

$$\max_{\omega} \left[p(\omega) \right] = \max_{\omega} \left[\frac{\chi_0^2 - \chi_\omega^2}{\chi_0^2} \right] = z$$

GLS periodogram: peak significance

Empirical distribution of test statistics.



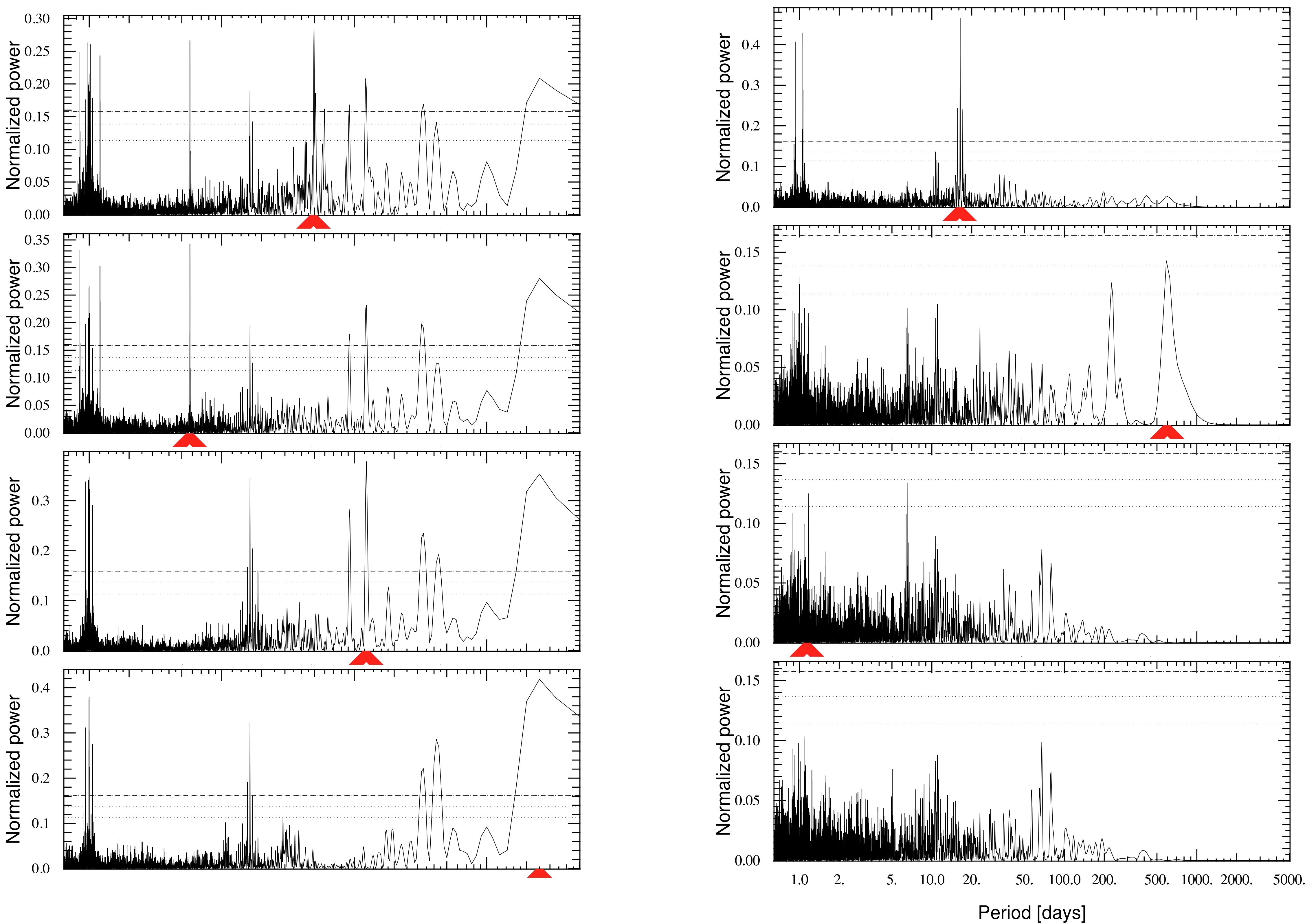
Recent work by Sulis, Mary & Bigot (2017) study different test statistics for exoplanet detection. IEEE TRANSACTIONS ON SIGNAL PROCESSING.

The periodogram analysis

- Perform a frequency analysis of the RV measurements.
    - Many methods exist (Generalized Lomb Scargle, Zechmeister & Kürster (2009); Keplerian periodogram, Baluev 2013; …)
- Remove signals whose amplitude exceed a certain level (issues arise on the definition of this level).
- Continue until no further significant signal is found.
- Perform an analysis including all the frequencies detected.

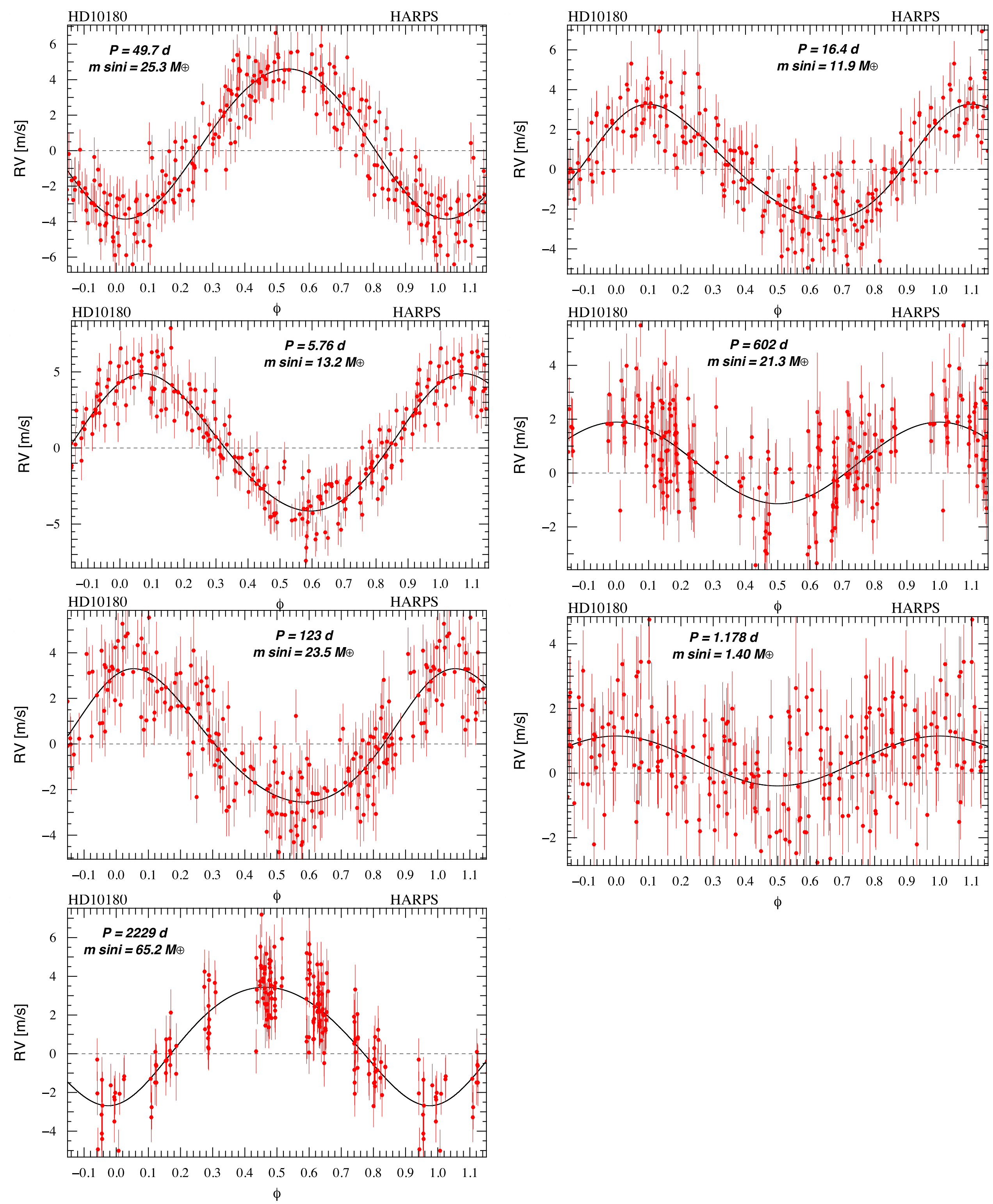
The periodogram analysis

HD10180 (Lovis et al. 2011)



The periodogram analysis

HD10180 (Lovis et al. 2011)



On periodograms

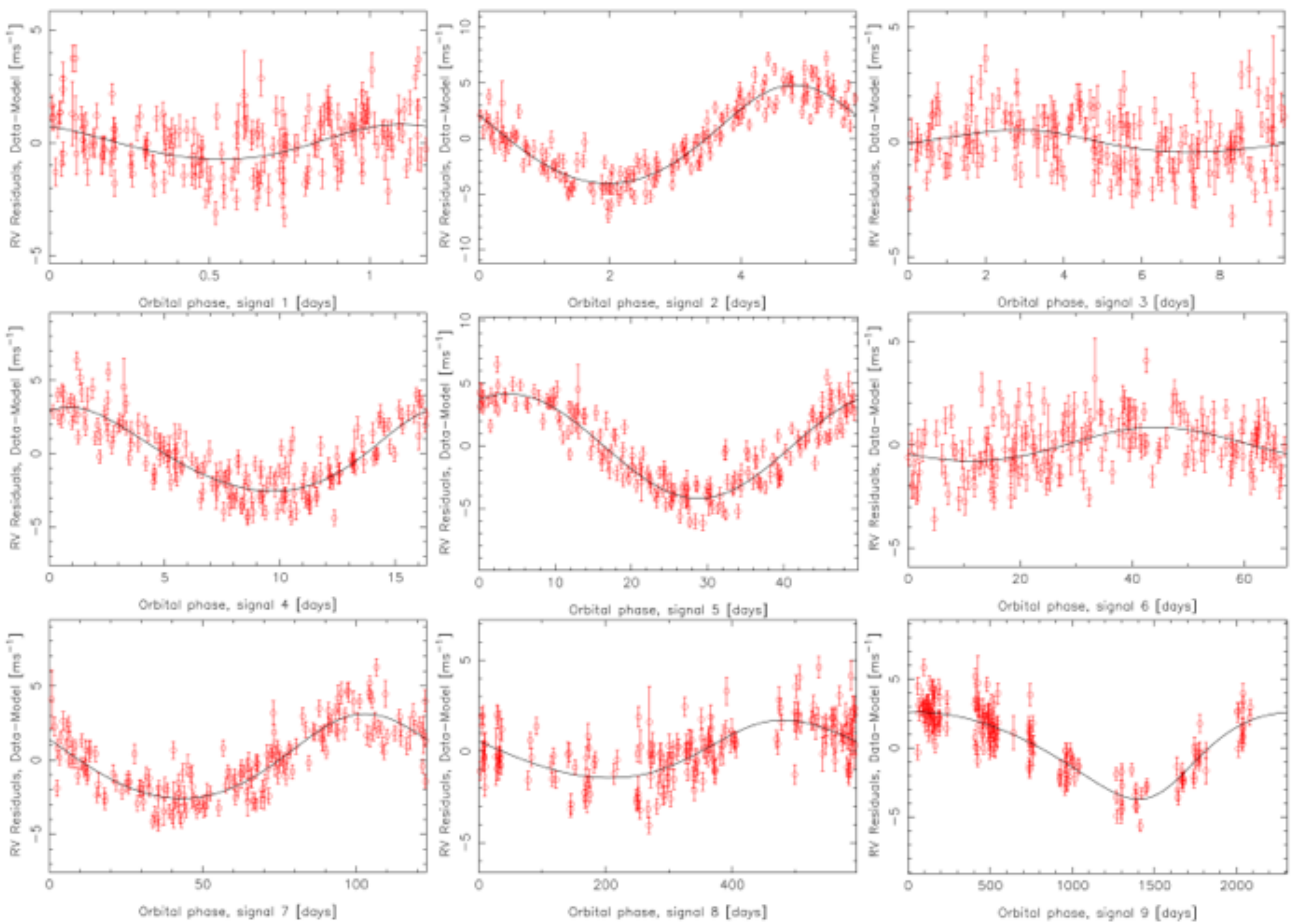
from another more significant flaw. The uncertainties of the signals removed from the data cannot be taken into account, which means that the method assumes the removed signals were known correctly. Obviously this is not the case even with the strongest signals, and even less so for the weaker ones, making the process prone to biases. Therefore, while likely producing reliable results when the signals are clear and their periods can be constrained accurately, this method cannot be expected to provide reliable results for extremely weak signals with large (and unknown) uncertainties. As noted by Lovis et al. (2011), when testing the significance of the 600-day signal, they could not take into account the uncertainties of the parameters of each Keplerian signal, that of the reference velocity, or the uncertainty in estimating the excess noise in the data correctly.

Tuomi (2012)

Periodogram challenged

HD10180

Tuomi (2012): nine signals.



Issues with *p*-values

$$p(D|H_o) \neq p(H_o|D)$$

- There is no reason to believe *p*-values are similar to posterior probabilities of the null hypothesis (KR95).
- Frequentist tests reject the null almost systematically for large samples (KR95).
- Most studies use the conventional limit $p < 0.05$ —> **Large fraction of false positives.**
- Simulations show that ~30% of the cases with $p < 0.05$ the null hypothesis is true (Sellke, Bayarri, Berger, 2001).

KR95: Kass & Raftery (1995)

Issues with *p*-values

- More technical issues:
  - Dependent of experimental design, sampling techniques.
  - Does not comply with the *likelihood principle*.

ESSAY

Why Most Published Research Findings Are False

John P. A. Ioannidis

Published: August 30, 2005 • http://dx.doi.org/10.1371/journal.pmed.0020124

"… a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true…."

BASIC AND APPLIED SOCIAL PSYCHOLOGY, 37:1–2, 2015

ISSN: 0197-3533 print/1532-4834 online
DOI: 10.1080/01973533.2015.1012991

Routledge
Taylor & Francis Group

Editorial

David Trafimow and Michael Marks
New Mexico State University

The *Basic and Applied Social Psychology* (BASP) 2014 Editorial emphasized that the null hypothesis significance testing procedure (NHSTP) is invalid, and thus authors would be not required to perform it (Trafimow, 2014). However, to allow authors a grace period, the Editorial stopped short of actually banning the NHSTP. The purpose of the present Editorial is to announce that the grace period is over. From now on, BASP is banning the NHSTP.

Bayesian alternative

- **Robust logical foundation.**
    - Extension of logic and the works of Cox, Pólya and Jaynes.
- **Answer the question you want to ask.**
- **Bayes factors are applicable equally well to nested and non-nested models.**
- Bayes factors were shown to **favour simpler model even when more complex one was correct** (Atkinson 1978), but Smith & Spiegelhalter (1980) showed that **this only happens when the predictions from the two models are equivalent**.
- **Sensitivity to priors.**
- **Computational difficulty.**

Two basic tasks of statistical inference

Learning process

(parameter estimation)

Decision making

(model comparison)

Learning process

Bayesian **probability** represents a state of

knowledge

$$p(\bar{\theta}|H_i, I) \longrightarrow p(\bar{\theta}|D, H_i, I)$$

Prior Posterior

$\bar{\theta}$: parameter vector

H_i: hypothesis

I: information

D: data

Discrete space

(hypothesis space)

$$p(H_i|I) \longrightarrow p(H_i|I, D)$$

$\bar{\theta}$: parameter vector
H_i: hypothesis
I: information
D: data

Learning process

Enter the **likelihood** `function`

$$p(\bar{\theta}|H_i, I, D) = \frac{p(D|\bar{\theta}, H_i, I)}{p(D|H_i, I)} \cdot p(\bar{\theta}|H_i, I)$$

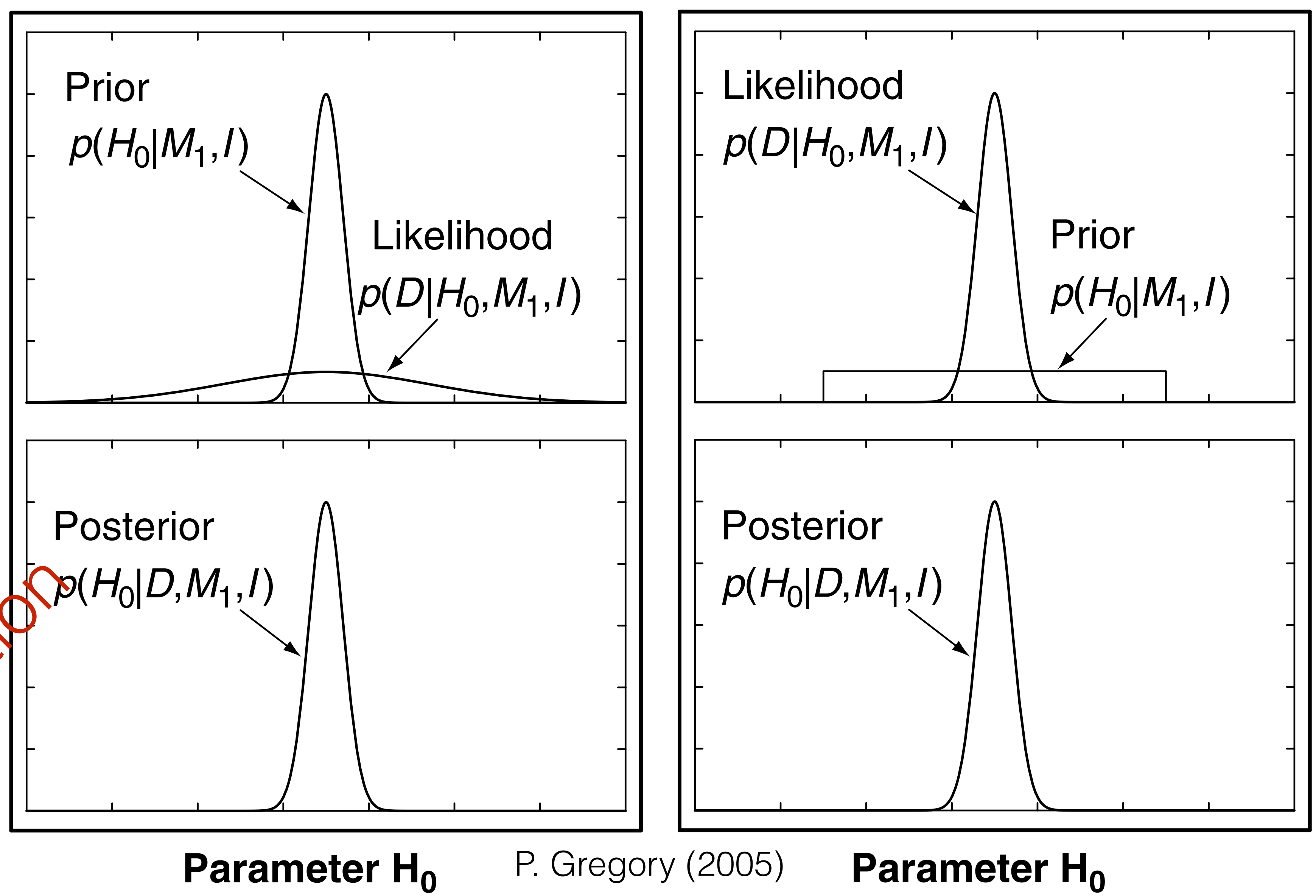
Posterior Prior

$$\boxed{p(\bar{\theta}|D, H_i, I) \propto \mathcal{L}_\theta(H_i) \cdot p(\bar{\theta}|H_i, I)}$$

The proportionality constant has many names: marginal likelihood, global likelihood, model evidence, prior predictive. Hard to compute.

Optimising the learning process

- The likelihood needs to be selective for the learning process to be effective.



Dominated by prior information

P. Gregory (2005)

Dominated by data

Decision making

Bayes' theorem is also the base for model comparison

$$p(H_i|I,D) = \frac{p(D|H_i,I)}{p(D|I)} \cdot p(H_i|I)$$

but now $p(D|H_i,I) = \int_\pi p(D|\bar{\theta},H_i,I)p(\bar{\theta}|H_i,I)\,\mathrm{d}^n\theta$

Computation of the evidence cannot be escaped.

Decision making

Model comparison consists in computing the ratio of the posteriors (odds ratio) of two competing hypotheses.

$$\frac{p(H_i|I,D)}{p(H_j|I,D)} = \frac{p(D|H_i,I)}{p(D|H_j,I)} \cdot \frac{p(H_i|I)}{p(H_j|I)}$$

Decision making

Model comparison consists in computing the ratio of the posteriors (odds ratio) of two competing hypotheses.

$$\frac{p(H_i|I,D)}{p(H_j|I,D)} = \boxed{\frac{p(D|H_i,I)}{p(D|H_j,I)}} \cdot \boxed{\frac{p(H_i|I)}{p(H_j|I)}}$$

Bayes factor Prior odds

$2\log_e(B_{10})$	(B_{10})	Evidence against H_0
0 to 2	1 to 3	Not worth more than a bare mention
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
>10	>150	Very strong

Kass & Raftery (1995)

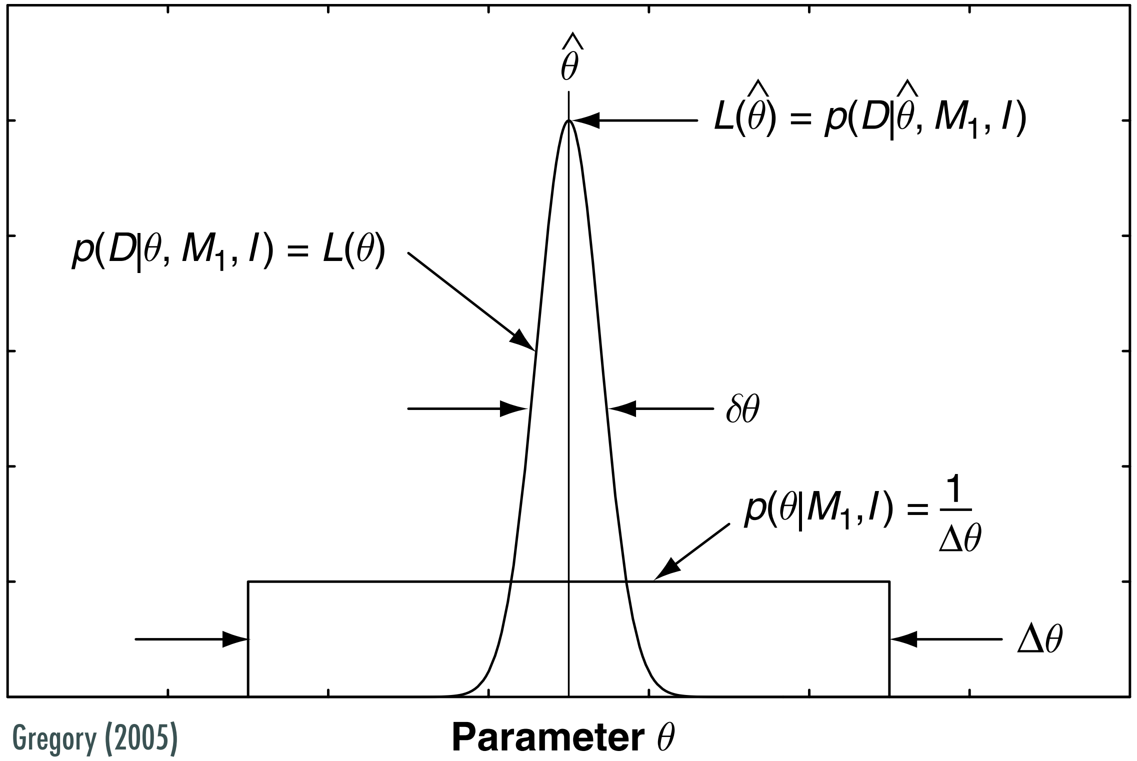
A built-in Occam's razor

The Bayes factor naturally punishes models with more parameters

$$p(D|H_i, I) = \int p(D|\bar{\theta}, H_i, I)p(\bar{\theta}|H_i, I)\, \mathrm{d}\bar{\theta}$$

E.g.: model M_0, without free parameters. Model M_1, with one parameter.

$$M_0 = M_1(\theta = \theta_0)$$



Gregory (2005)

$$p(D|M_0, I) = p(D|\theta_0, M_1, I)$$

$$p(D|M_1, I) = \int p(D|\theta, M_1, I) \cdot p(\theta|M_1, I)\, \mathrm{d}\theta$$
$$= \frac{1}{\Delta\theta} \int p(D|\theta, M_1, I)\, \mathrm{d}\theta$$
$$\approx p(D|\hat{\theta}, M_1, I) \frac{\delta\theta}{\Delta\theta}$$

$$B_{10} \approx \frac{p(D|\hat{\theta}, M_1, I)}{p(D|\theta_0, M_1, I)} \boxed{\frac{\delta\theta}{\Delta\theta}}$$

Occam's factor
One per parameter

Estimating the marginal likelihood

$$p(D|H_i, I) = \int_{\pi} p(D|\bar{\theta}, H_i, I)p(\bar{\theta}|H_i, I)\ \mathrm{d}^n\theta$$

Marginal likelihood is a k-dimensional integral over parameter space.

Accounting for the “size” of parameter space bring many good things to Bayesian statistics, but the integral is intractable…

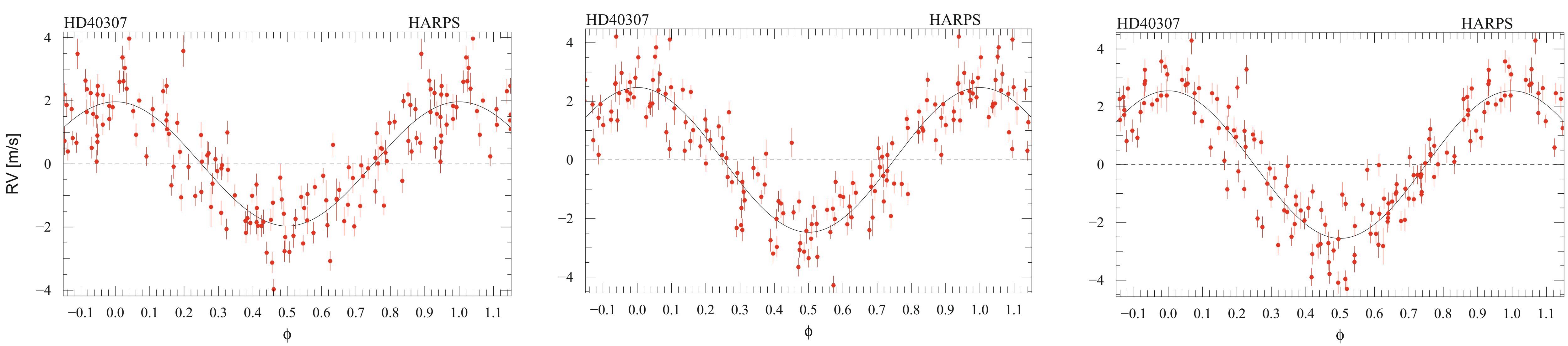
Large number of techniques to estimate the value of the integral:

- Asymptotic estimates (Laplace approximation, BIC).
- Importance sampling.
- Chib & Jeliazkov.
- Nested sampling.

Polemical system: HD40307

Díaz, et al. (2016)

- Polemical system: 3 planets (Mayor+2009); 6 planets (Tuomi+2013)

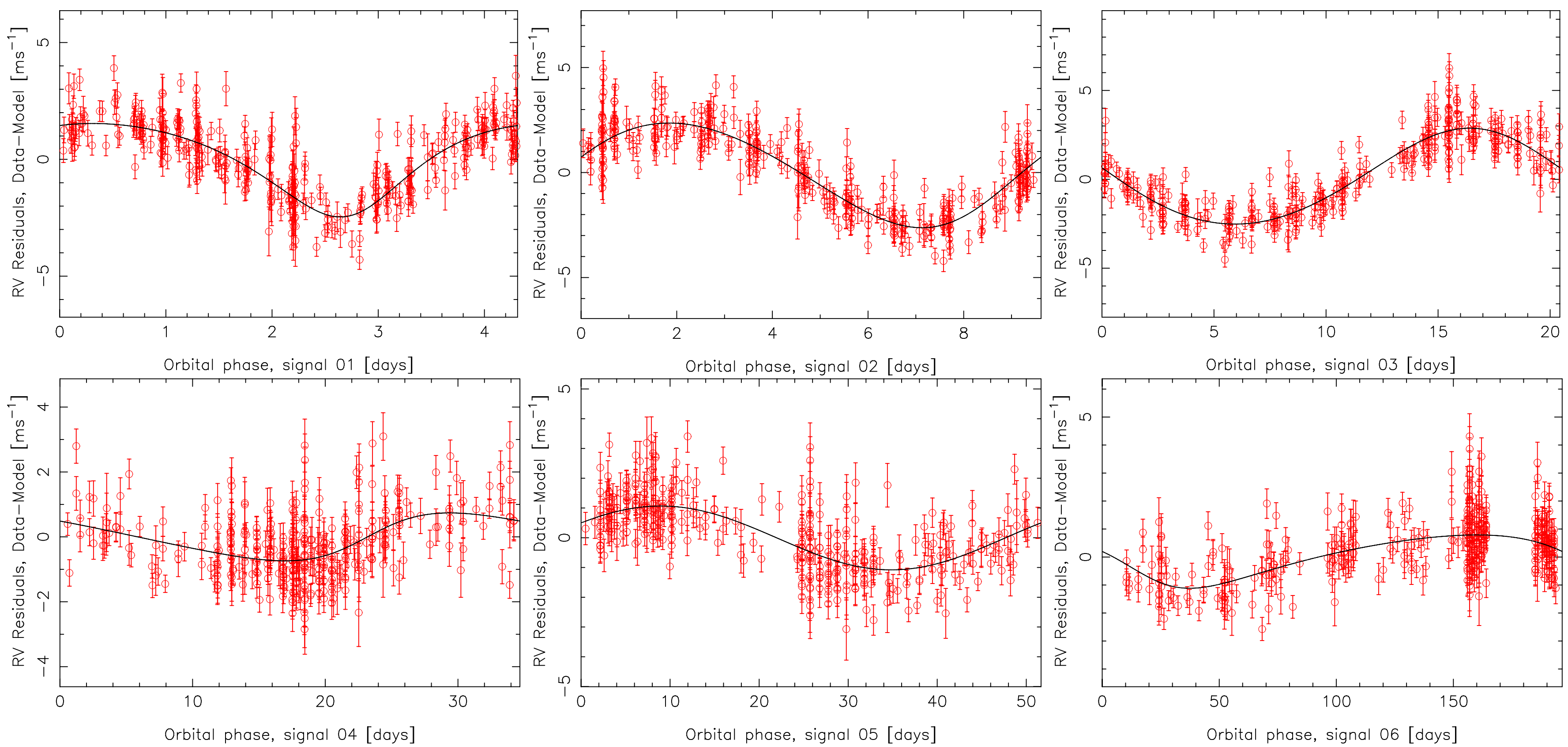


Díaz, et al. (2016)

Polemical system: HD40307

Díaz, et al. (2016)

- Polemical system: 3 planets (Mayor+2009); 6 planets (Tuomi+2013)
- Tuomi et al. announced six planet using roughly the same data, a noise model, and a "Bayesian" approach.
- The sixth planet (g) would be a Super-Earth in the HZ.

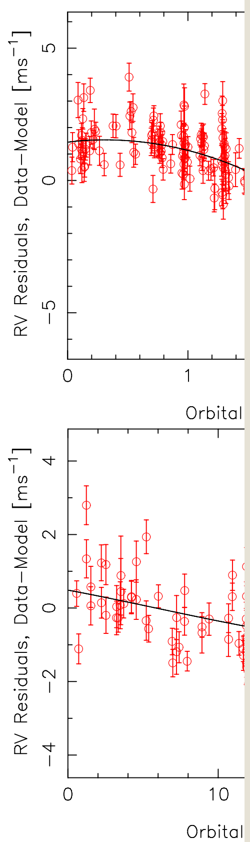


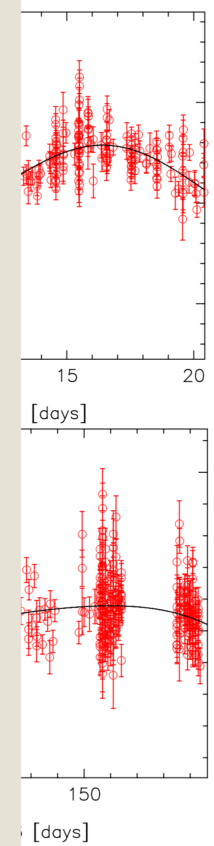
Díaz, et al. (2016)

Polemical sy

Díaz, et al. (2016)

- Polemical syst … (Tuomi+2013)
- Tuomi et al. a … he same data, a noise model, a
- The sixth plan … .





Díaz, et al. (2016)

$$\frac{p(H_{\mathrm{i+1}}|D, I)}{p(H_{\mathrm{i}}|D, I)}$$

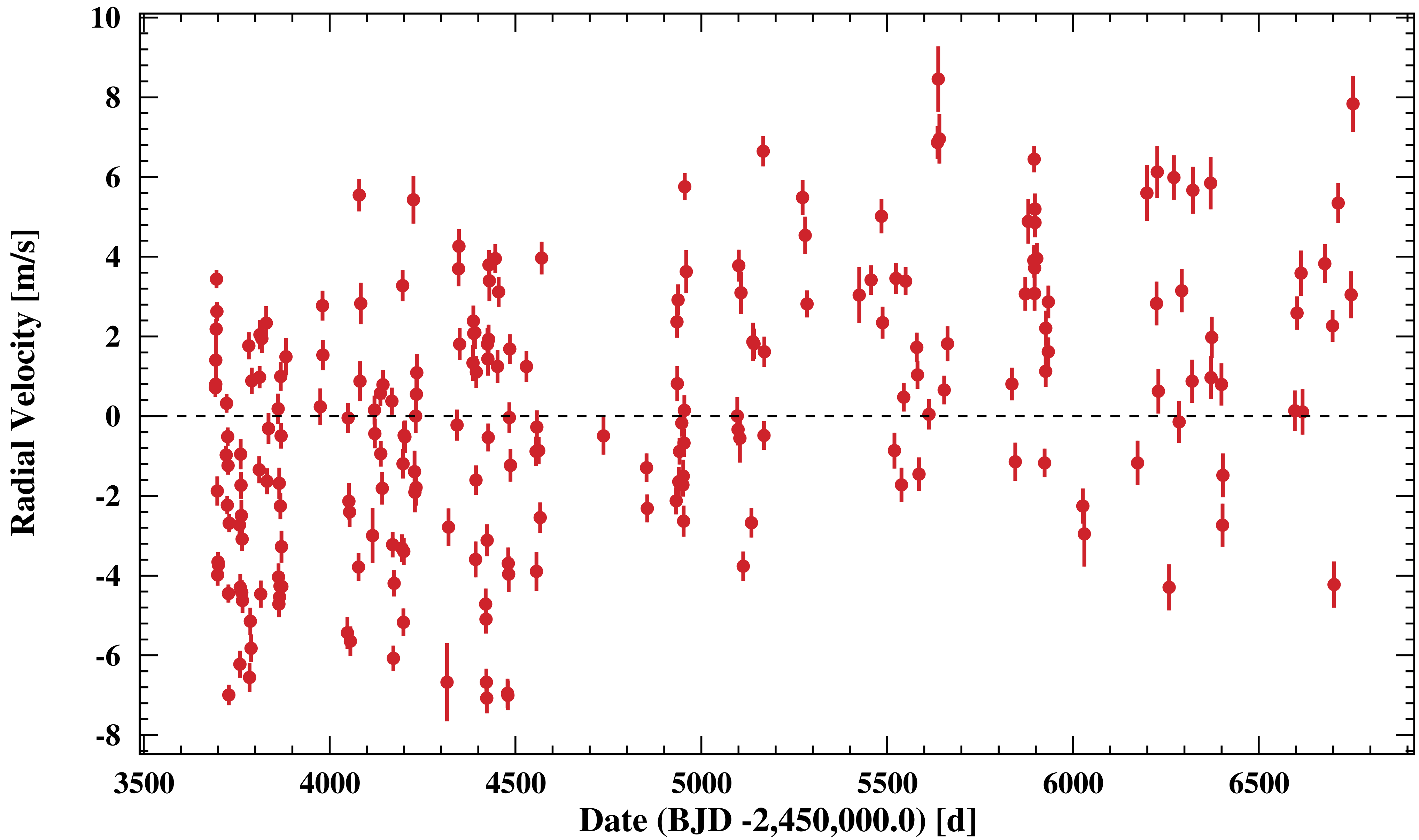
for i = 0, 1, …, N

HD40307

Díaz, et al. (2016a)

Bayesian analysis of new HARPS data.

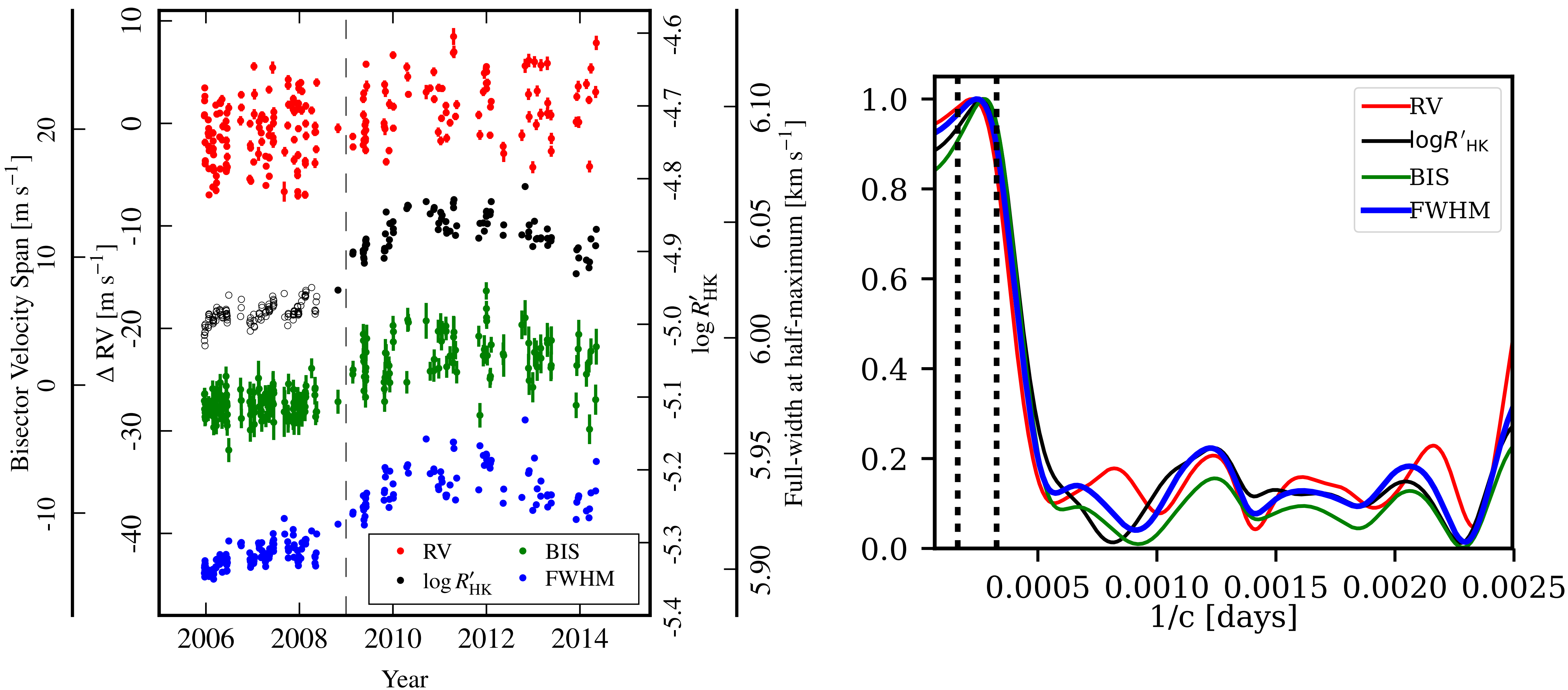
- 10+ years of data: 98 new nightly-averaged points (for a total of 226 measurements).



HD40307

Díaz, et al. (2016a)

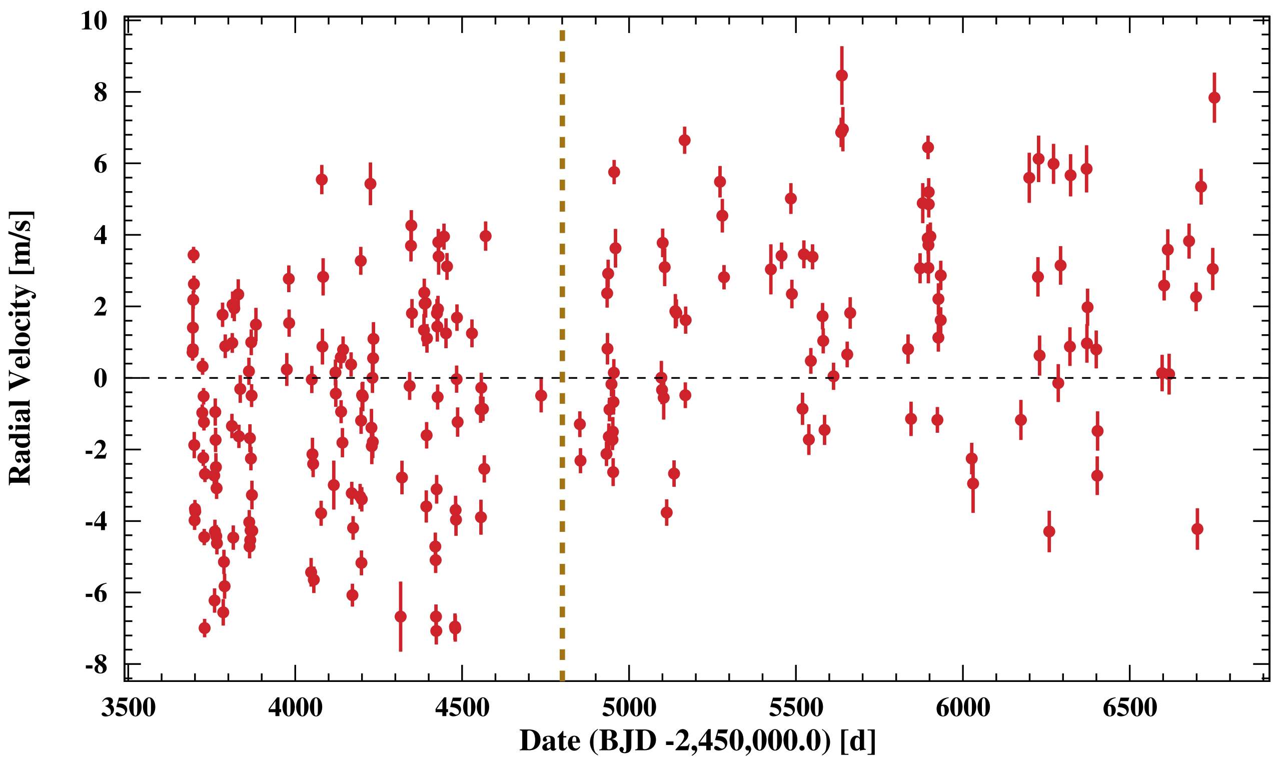
Long-term activity cycle seen in all observables.

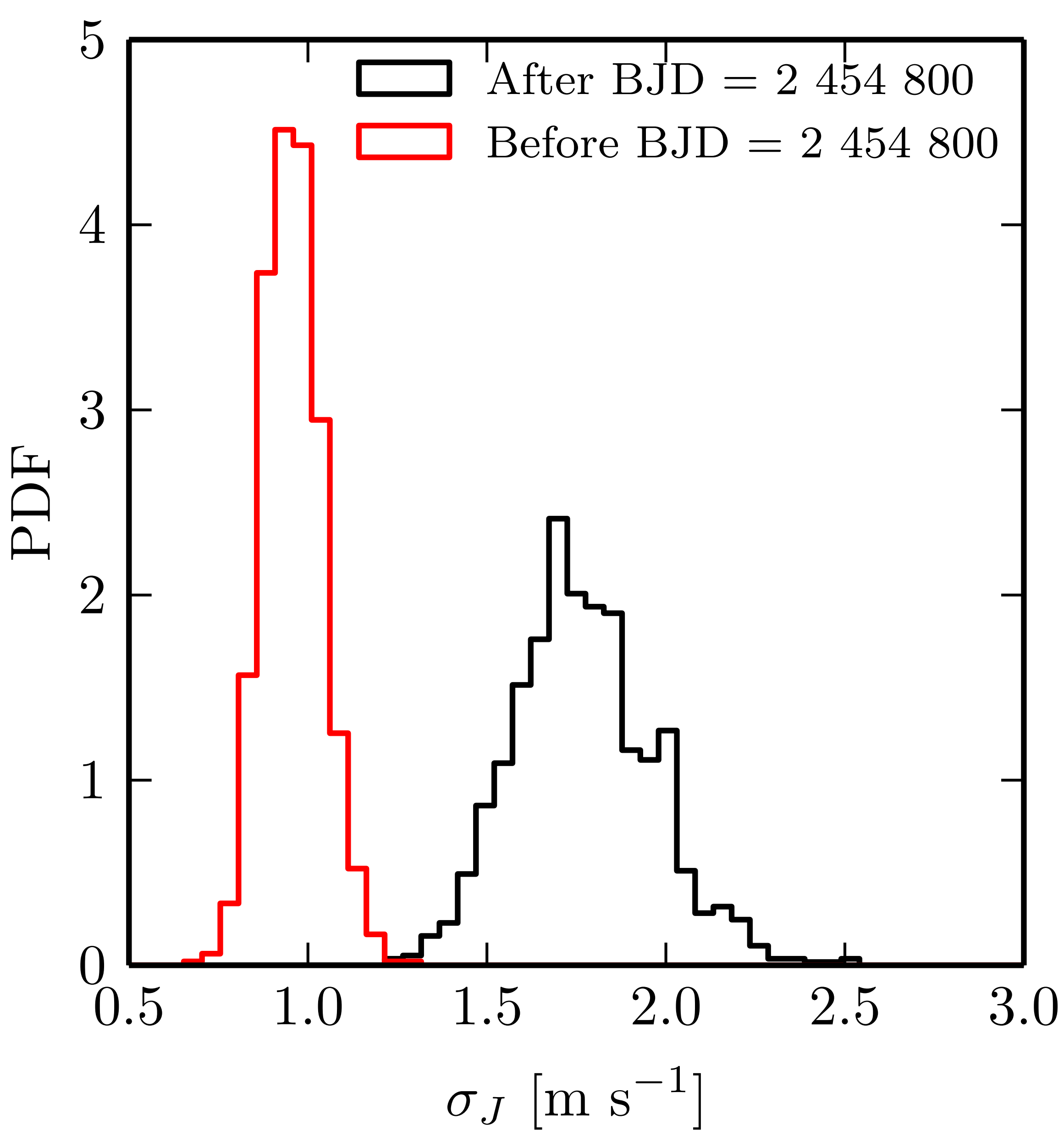


HD40307

Díaz, et al. (2016a)

Stellar noise increases with activity level.





Bayesian analysis of HD40307

Díaz, et al. (2016a)

- Hypotheses: H_n: "there are n planets in the HD40307 system."
- Physical model: n Keplerians. No interactions.
- Long term drift: polynomial fit to R'_{HK} used as prior. Linearity assumed.
- Noise model: Evolving stellar jitter, to consider activity increase.
- Likelihood: normally distributed independent errors.

$$\mathcal{L} = \prod_i \frac{1}{\sqrt{2\pi}\sqrt{\sigma_i^2 + \sigma_{Ji}^2}} \exp\left[-\frac{(v_i - m_i)^2}{2(\sigma_i^2 + \sigma_{Ji}^2)}\right]$$

Díaz, et al. (in prep, 2015)

The model

RV at t_i measured by instrument *k*

Planets and non-“jitter” activity (rotational modulation). Drifts (degree *l*).

Activity cycles modelled as polynomial or Keplerian, using priors based on logR’hk, and a scale factor.

$$a(t_i) = \alpha \cdot P_3^{\log R'_{\rm HK}}(t_i)$$

$$a(t_i) = \alpha' \cdot k_{\log R'_{\rm HK}}(t_i)$$

$$v_i^{(k)} = \delta_k + \sum_{j=1}^{n} k_j(t_i) + p_l(t_i) + a(t_i) + \epsilon_i^{(k)}$$

Independent Gaussian measurements errors assumed.

Number of parameters:
$5j + (1\text{-}3) + 1 + (3 \mid 5) + 2$

Short-term activity “jitter” (including instrumental noise). “Complex physics”; modelled statistically.

$$\epsilon_i \sim \mathrm{N}(0, \sigma_{Ji})$$

$$\sigma_{Ji} \propto \log R'_{HK}$$

i.e., the “jitter” increases with activity.

Parameter posterior PDF sampled using a **MCMC + adaptive PCA** algorithm (details in Díaz +2014. Starting point obtained using Genetic Algorithm.

The prior distribution

Orbital parameters		Prior distribution
Orbital period, P	[days]	$J(1.0, 10^4)$
RV amplitude, K	[$\mathrm{m\,s^{-1}}$]	$U(0.0, 200)$
Eccentricity, e		$B(0.867, 3.03)$
Argument of periastron, ω	[deg]	$U(0, 360)$
Mean longitude at epoch, L_0	[deg]	$U(0, 360)$
Systemic velocity, V_0	[$\mathrm{km\,s^{-1}}$]	$U(28.996, 33.668)$
Velocity drift (long-term activity effect)*		
Scaling constant, α	[$\mathrm{m\,s^{-1}/dex}$]	$U(0, 100)$
Linear	[10^{-4} dex $\mathrm{yr^{-1}}$]	$N(373.8, 8.6)$
Quadratic	[10^{-4} dex $\mathrm{yr^{-2}}$]	$N(46.3, 4.6)$
Cubic	[10^{-4} dex $\mathrm{yr^{-3}}$]	$N(-17.3, 1.0)$
Noise model‡		
Additional noise at $\log(R'_{\mathrm{HK}}) = -5$, $\sigma_{\mathrm{J}}\|_{-5.0}$	[$\mathrm{m\,s^{-1}}$]	$U(0, 50)$
Slope, α_{J}	[$\mathrm{m\,s^{-1}/dex}$]	$U(0, 50)$

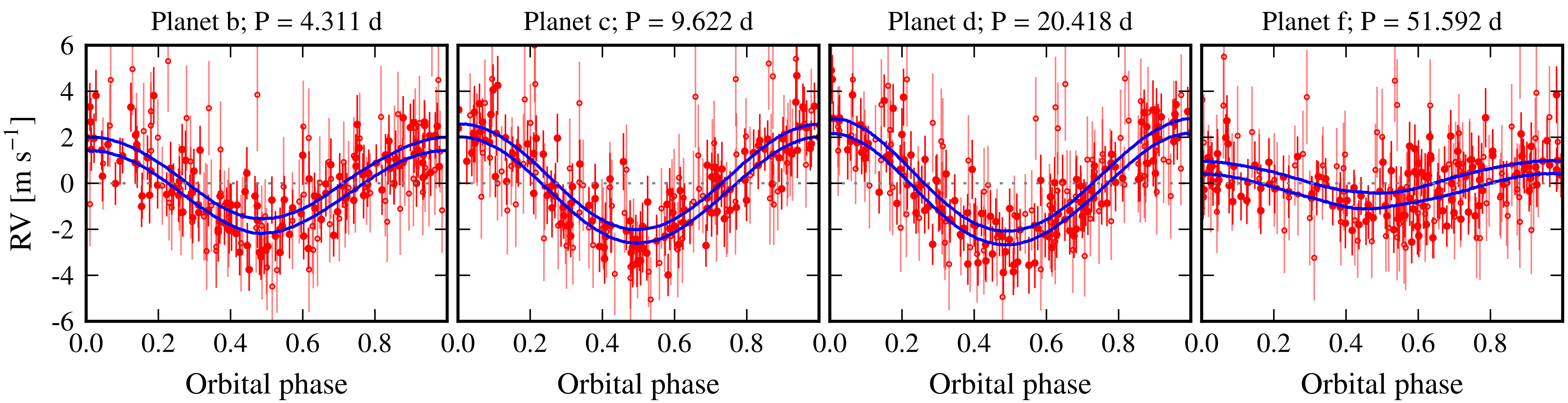
$$\frac{p(H_{\mathrm{i}+1}|D, I)}{p(H_{\mathrm{i}}|D, I)}$$

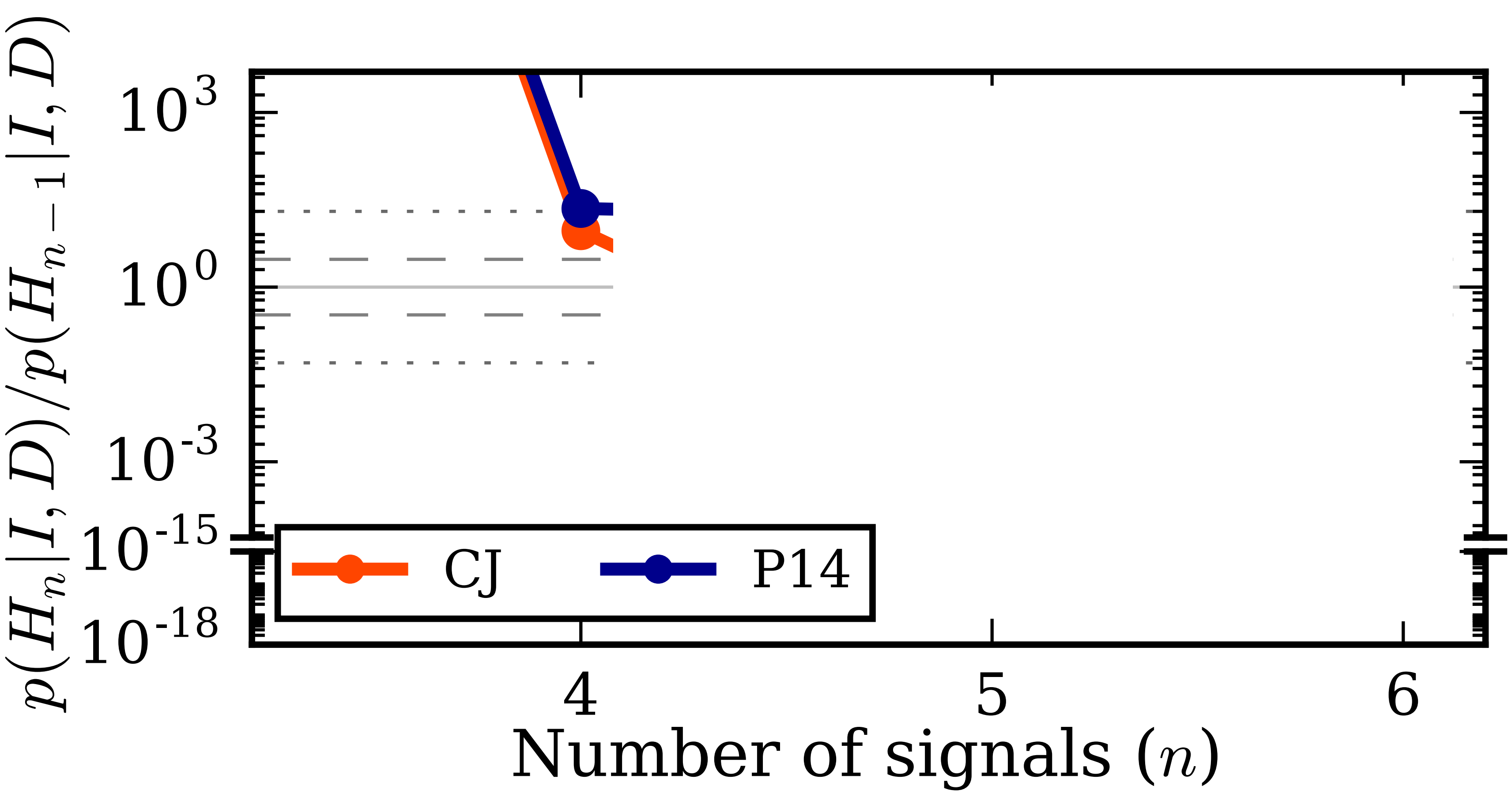
for i = 0, 1, …, N

$2 \log_e(B_{10})$	(B_{10})	Evidence against H_0
0 to 2	1 to 3	Not worth more than a bare mention
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
>10	>150	Very strong

KR95: Kass & Raftery (1995)

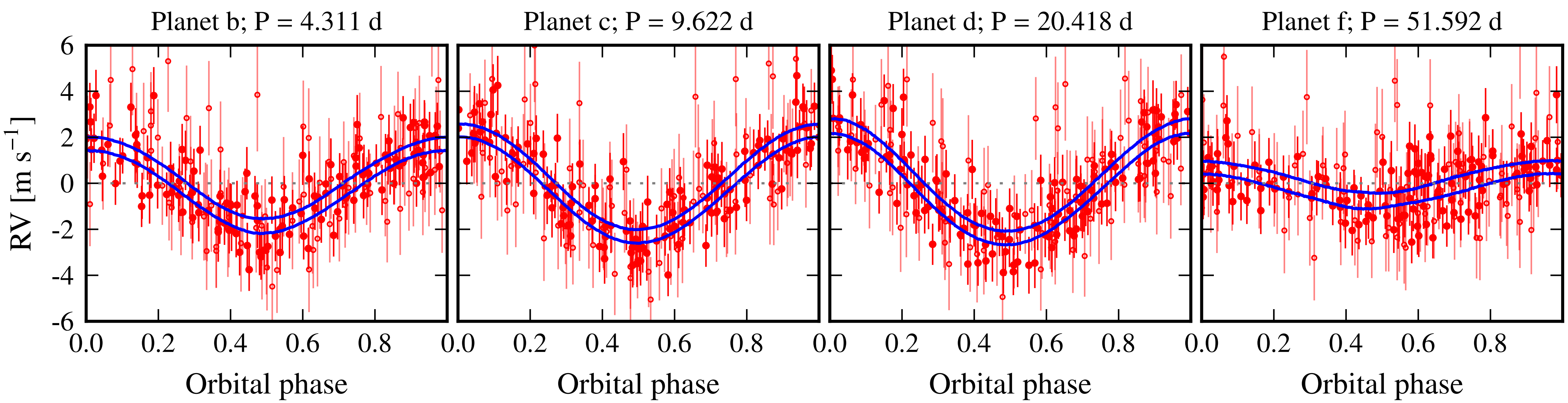
k=4; a four-planet model

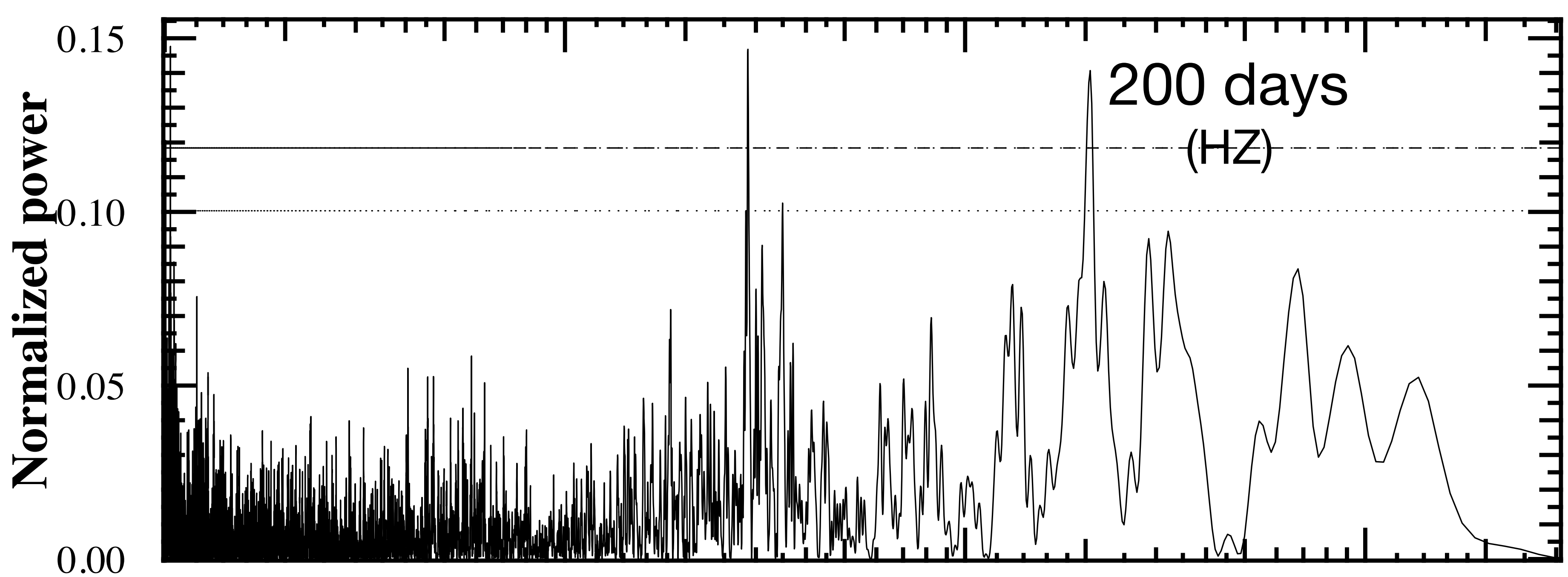




Díaz et al. (2016a)

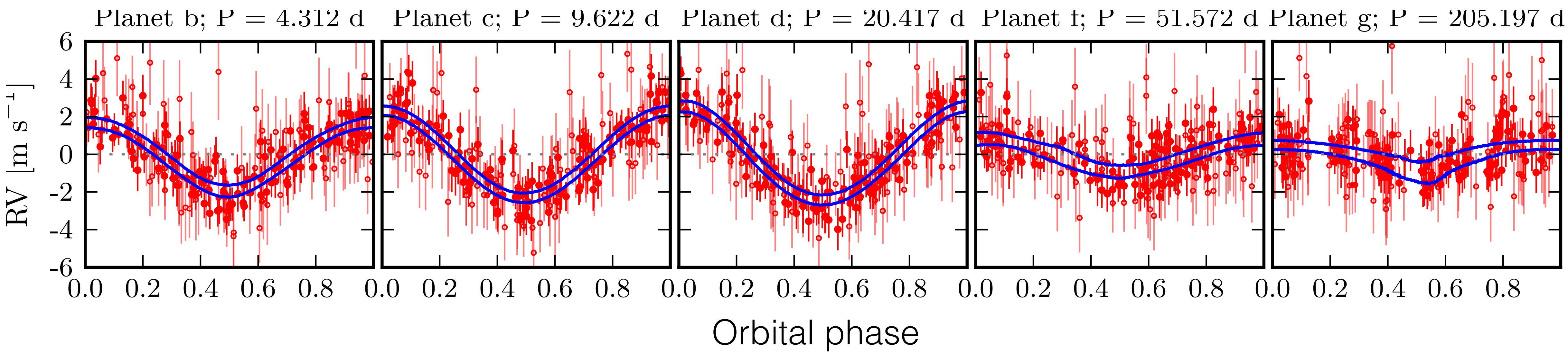
k=4; a four-planet model

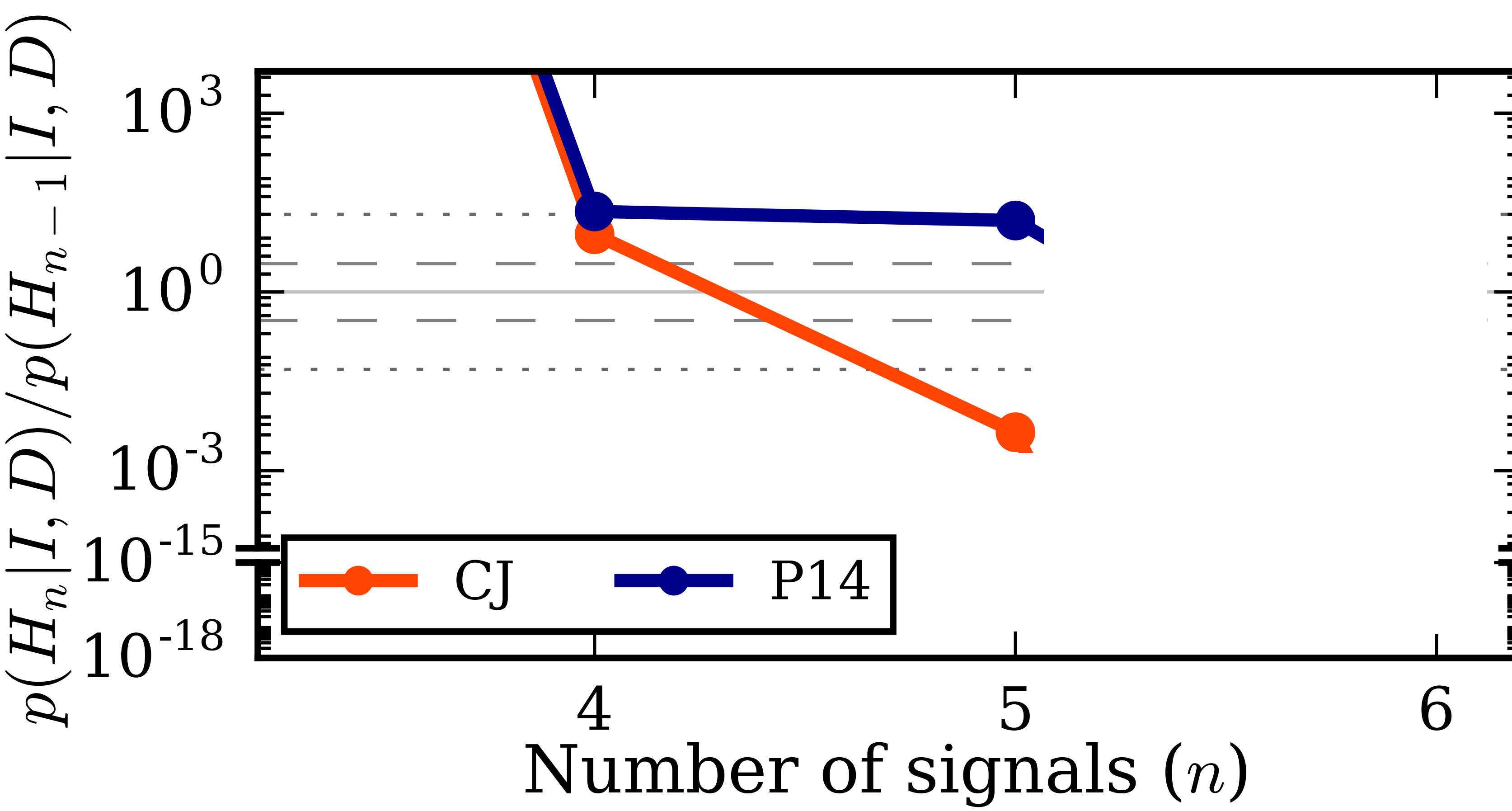




Díaz et al. (2016a)

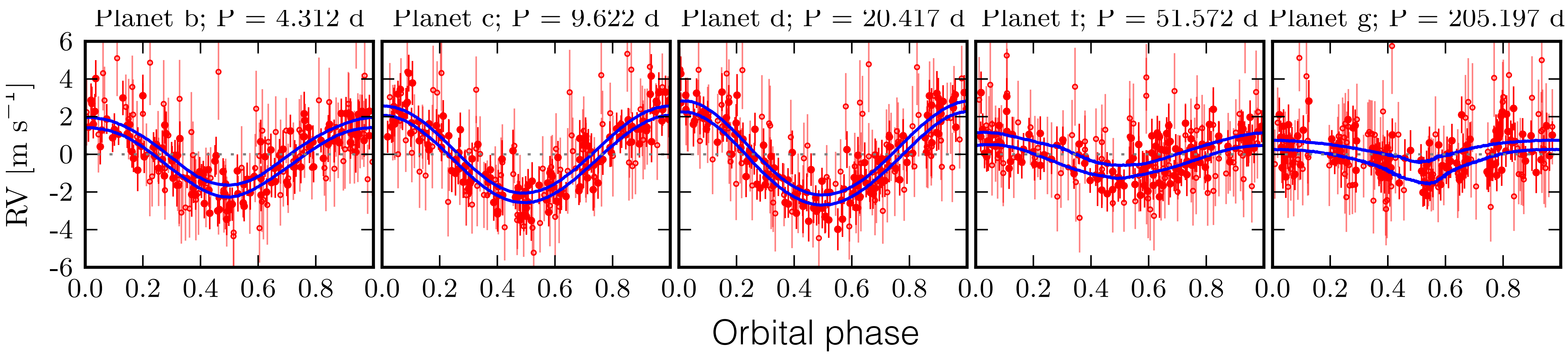
k=5; a five-planet model (one in the HZ)

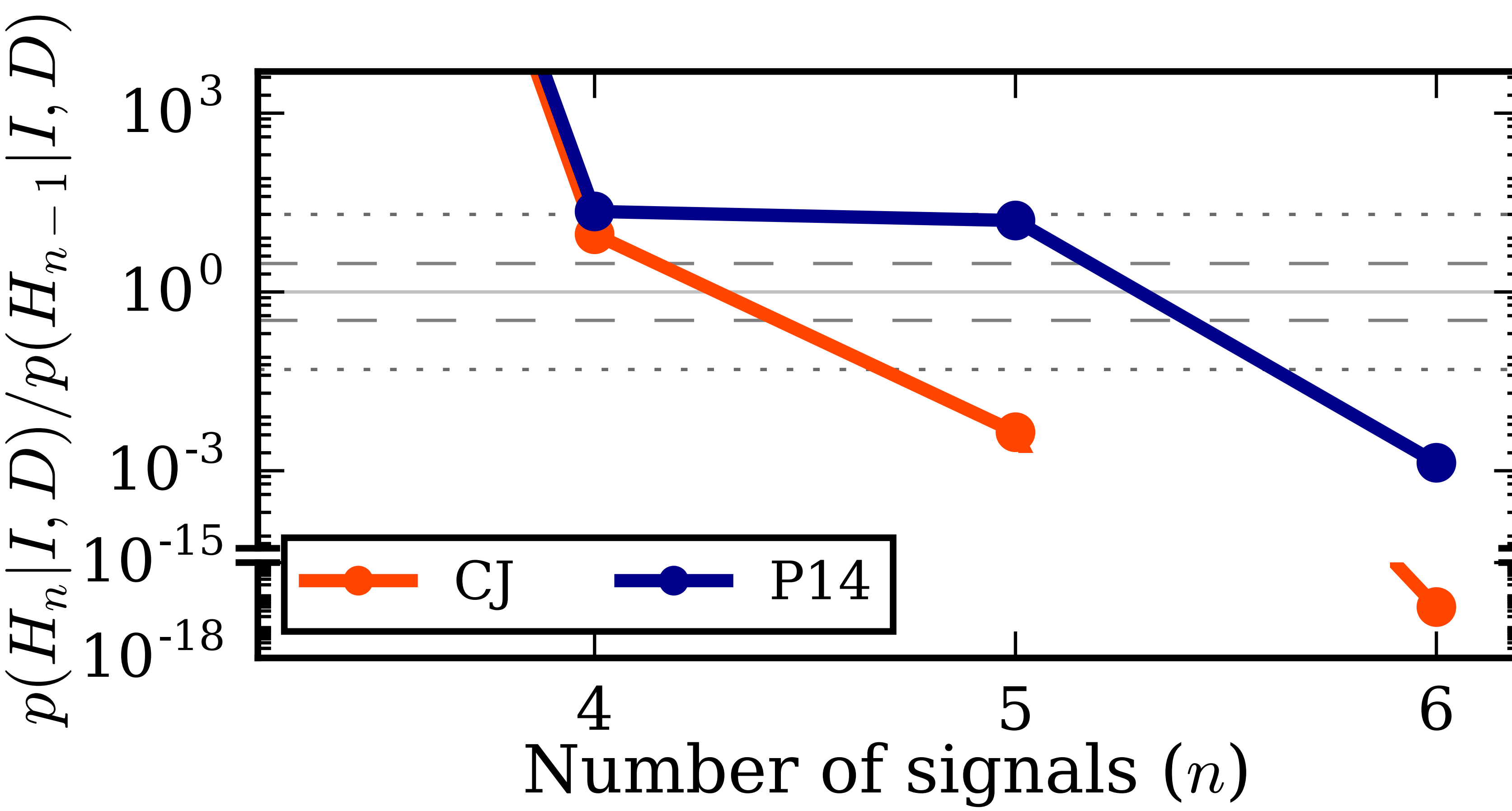




Díaz et al. (2016a)

k=5; a five-planet model (one in the HZ)





Díaz et al. (2016a)

- new statistical and computational methods for signal extraction, signal analysis, and jitter mitigation

THE THIRD WORKSHOP ON

EPRV

EXTREMELY PRECISE RADIAL VELOCITIES

AUG 14-17

STATE COLLEGE, PA, USA

Following the tradition of previous workshops, participants will dig into the "nuts and bolts" of exoplanetary discovery and orbit characterization via Doppler velocimetry, and discuss challenges, lessons learned, and the details of their work, "warts and all".

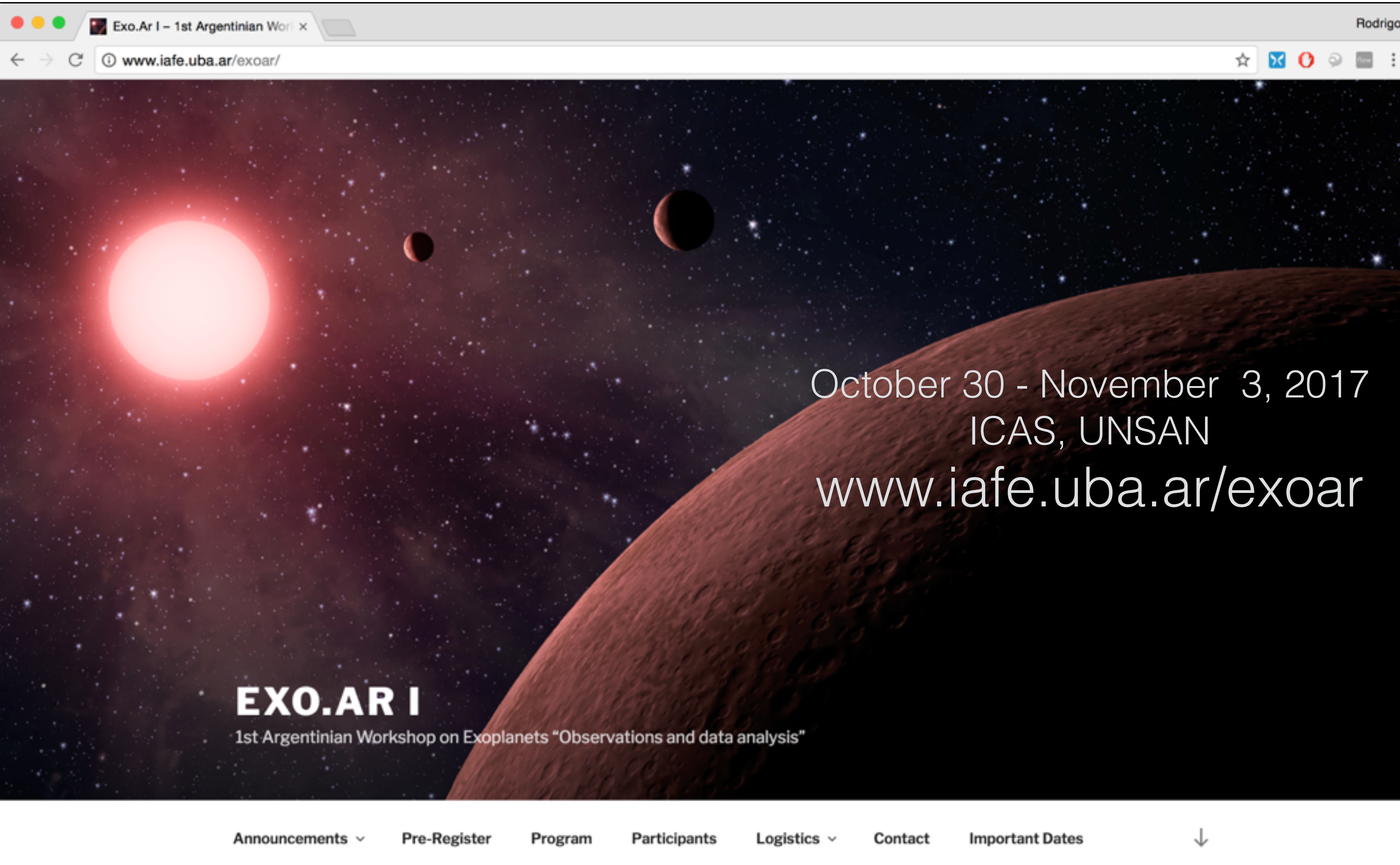
This edition of the workshop will focus on:

- specific hardware challenges
- lessons learned from the newest generation of EPRV instruments
- new statistical and computational methods for signal extraction, signal analysis, and jitter mitigation
- physical models and spectral diagnostics of stellar granulation, activity, and other sources of jitter

BIT.LY/EPRVIII

PennState
Eberly College of Science

Center for Exoplanets
and Habitable Worlds


Exo.Ar I – 1st Argentinian Wor
Rodrigo
www.iafe.uba.ar/exoar/
October 30 - November 3, 2017
ICAS, UNSAN
www.iafe.uba.ar/exoar
EXO.AR I
1st Argentinian Workshop on Exoplanets "Observations and data analysis"
Announcements
Pre-Register
Program
Participants
Logistics
Contact
Important Dates