Signals embedded in the radial velocity noise

Mikko Tuomi
University of Hertfordshire, Centre for Astrophysics Research
Email: mikko.tuomi@utu.fi
Presentation, 13.11.2012

The Bayes' rule

$$\pi(\theta|m) = \frac{l(m|\theta)\pi(\theta)}{P(m)}, P(m) > 0 \tag{1}$$

- $\pi(\theta|m)$ is the posterior density of the parameter given the measurements.
- $\pi(\theta)$ is the prior density, the information on θ before the measurement m was made.
- $l(m|\theta)$ is the likelihood function of the measurement − the statistical model.
- $P(m)$ is constant, usually called the marginal integral, that makes sure that the posterior is a proper probability density.

$$P(m) = \int_{\theta \in \Theta} l(m|\theta)\pi(\theta) d\theta. \tag{2}$$

The Bayes' rule

The Bayes' rule works the same way regardless of the number of measurements (or sets of measurements) available. For independent measurements $m_i, i = 1, ..., N$,

$$\pi(\theta|m_1, ..., m_N) = \frac{l(m_1, ..., m_N|\theta)\pi(\theta)}{P(m_1, ..., m_N)} = \frac{\pi(\theta)\prod_{i=1}^{N} l_i(m_i|\theta)}{P(m_1, ..., m_N)} \quad (3)$$

The best part about the above equation is that:

- The likelihood model(s) can be anything that can be expressed mathematically.
- The measurements can be anything (from different sources: RV, transit, astrometry, etc.).
- No assumptions are required about the nature of the probability densities of model parameters θ.
- Also, if N is large (etc.), the prior $\pi(\theta)$ can also be pretty much anything.

Priors

Priors are the only subjective part of Bayesian analyses – the rest consists of mindless repetitive tasks (i.e. computing).

- Prior probability densities (prior models): something has to be assumed always, a flat prior is still a prior (one that all frequentists' using likelihoods assume).

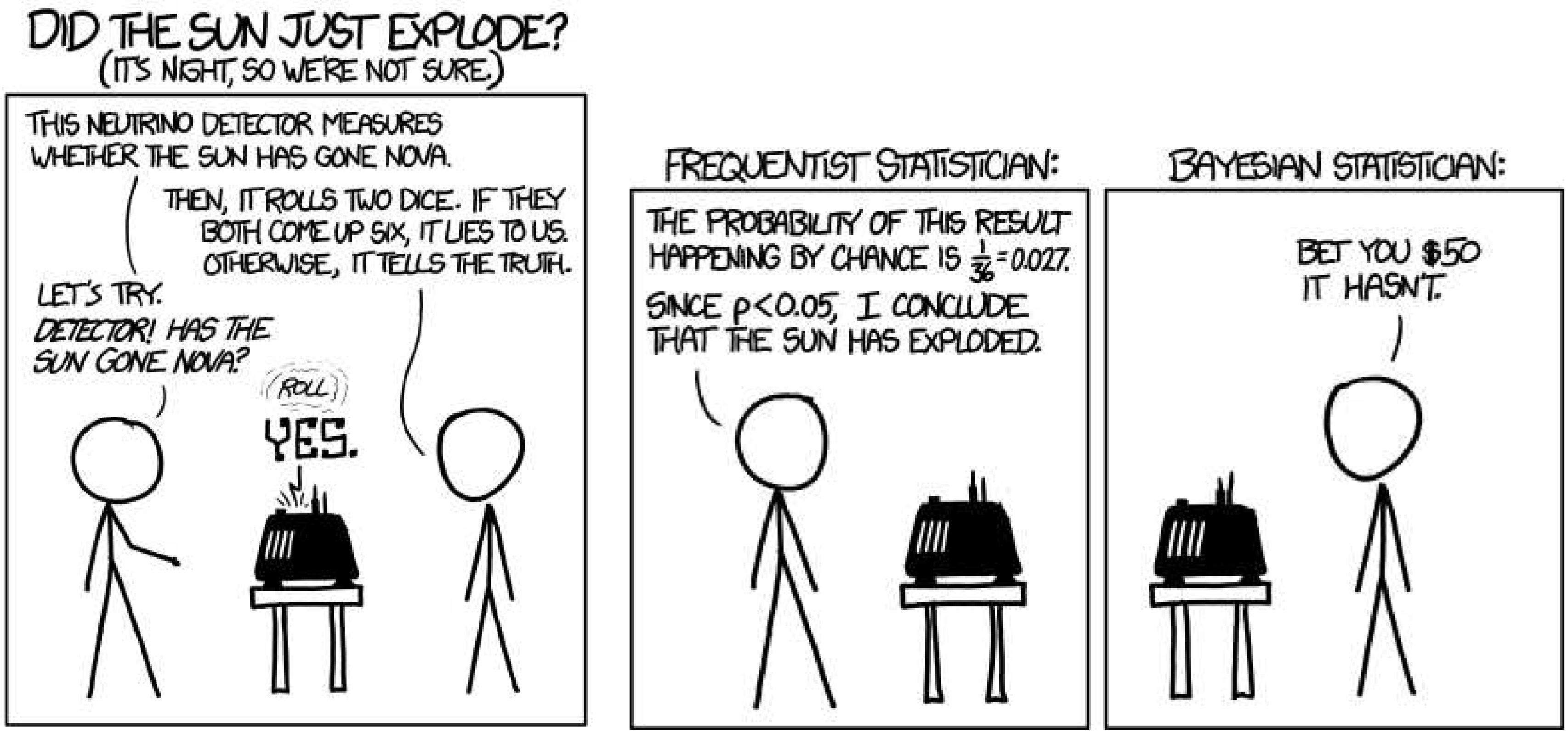
- Fixing parameters (e.g. fixing $e = 0$) corresponds to a delta-function prior.

- Flat priors on different parameterisations, e.g. using parameters (e, ω) vs. $(e \sin \omega, e \cos \omega)$, result in different results.

- Prior probabilities for models do not have to be equal.

- The collection of candidate models is also selected *a priori* – comparison of these models might be indistinguishable from comparison of different prior models.

Priors


DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)
THIS NEUTRINO DETECTOR MEASURES WHETHER THE SUN HAS GONE NOVA.
THEN, IT ROLLS TWO DICE. IF THEY BOTH COME UP SIX, IT LIES TO US. OTHERWISE, IT TELLS THE TRUTH.
LET'S TRY.
DETECTOR! HAS THE SUN GONE NOVA?
ROLL
YES.
FREQUENTIST STATISTICIAN:
THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS 1/36 = 0.027.
SINCE p<0.05, I CONCLUDE THAT THE SUN HAS EXPLODED.
BAYESIAN STATISTICIAN:
BET YOU $50 IT HASN'T.

Bayes' rule and dynamical information

In case of detections of exoplanet systems, there is a very useful additional source of information – the Newtonian (or post-Newtonian if necessary) mechanics.

Because we cannot expect to detect an unstable planetary system, we can say that the prior probability of detecting something unstable is zero (at least negligible). Hence:

$$\pi(\theta|m,\mathcal{S}) = \frac{l(m,\mathcal{S}|\theta)\pi(\theta)}{P(m,\mathcal{S})} = \frac{l(m|\mathcal{S},\theta)l(\mathcal{S}|\theta)\pi(\theta)}{P(m,\mathcal{S})} \tag{4}$$

Because Newton's laws do not depend on what we measured but what we can measure depends on them. We call $\mathcal{S}$ the "dynamical information".

But what is the likelihood function of dynamical information, $l(\mathcal{S}|\theta)$?

Bayes' rule and dynamical information

The approximated Lagrange stability criterion for two subsequent planets (Barnes & Greenberg, 2006) is defined as

$$\alpha^{-3}\left(\mu_1 - \frac{\mu_2}{\delta^2}\right)\left(\mu_1\gamma_1 + \mu_2\gamma_2\delta\right)^2 > 1 + \mu_1\mu_2\left(\frac{3}{\alpha}\right)^{4/3}, \qquad (5)$$

where $\mu_i = m_i M^{-1}$, $\alpha = \mu_1 + \mu_2$, $\gamma_i = \sqrt{1 - e_i^2}$, $\delta = \sqrt{a_2/a_1}$, $M = m_\star + m_1 + m_2$, e_i is the eccentricity, a_i is the semimajor axis, m_i is the planetary mass, and $m_\star$ is stellar mass.

We simply set $l(\mathcal{S}|\theta) = c$ if the criterion is satisfied and $l(\mathcal{S}|\theta) = 0$ otherwise (we call the set of stable orbits in the parameter space $B \subset \Theta$).

Alternatively, we could use a simpler form that only prevents orbital crossings of the planets. Note that the stellar mass is one of the parameters.

The above does not take e.g. resonances into account, and is only a rough approximation. Can we do better?

Posterior sampling with dynamics

A posterior sample from a MCMC analysis: $\exists \theta_i \sim \pi(\theta|m), i = 1, ..., K$.

Each θ_i as an initial state of orbital integration: K chains of N vectors with $\theta_i^j, j = 1, ..., N$, and $\theta_i^j = \theta_i(t_j)$ and t_j is some moment between $t_0 = 0$ and the duration of the integrations $t_N = T$.

Hence, we can approximate the posterior probability of finding $\theta \in I_l \subset \Theta$ of dynamical information and data for each n-interval I_l as

$$\begin{aligned} P(\theta \in I_l|\mathcal{S}, d) &\approx \frac{1}{K}\sum_{i=1}^{K} P(\theta \in I_l|\mathcal{S}, \theta_i) \\ &\approx \frac{1}{KN}\sum_{i=1}^{K}\sum_{j=1}^{N} \mathbf{1}_l(\theta_i^j)\mathbf{1}(\theta_1^j), \end{aligned} \tag{6}$$

where

$$\mathbf{1}_l(\theta) = \begin{cases} 1 \text{ if } \theta \in I_l \\ 0 \text{ otherwise} \end{cases} \quad \text{and } \mathbf{1}(\theta) = \begin{cases} 1 \text{ if } \theta \in B \\ 0 \text{ otherwise} \end{cases}$$

How to detect a signal from RVs?

There are k periodic signals in the radial velocities if

- $P(\mathcal{M}_k|m) > \alpha P(\mathcal{M}_{k-1}|m)$ for a selected threshold $\alpha > 1$.
- If the radial velocity amplitudes of all signals are statistically distinguishable from zero, i.e. their BCSs (*) do not overlap with zero for a selected threshold $\delta \in [0, 1]$.
- All periodicities get well constrained from above and below.
- The planetary system corresponding to the solution (parameters k and θ) is stable.

$$\text{(*) Set } \mathcal{D}_\delta = \left\{ \mathcal{D}_\delta \subset \Theta : \int_{\theta \in \mathcal{D}_\delta} \pi(\theta|m) = \delta, \pi(\theta|m)|_{\theta \in \partial\mathcal{D}_\delta} = c \right\}. \quad (7)$$

Modelling radial velocity noise

- Usually RV data is binned (somehow) by calculating the average of few velocities within an hour or so.
- Binning will always result in loss of information (because the transformation called “binning” is not a bijective mapping).
- Instead, model the noise as realistically as possible.
- Possibility to have the “binning” procedure as a part of the statistical model, which enables comparisons of different procedures.

Modelling radial velocity noise

An effective analogue of "binning" is e.g. a noise model with moving average (MA) component. This statistical model can be written as

$$m_i = f_k(t_i) + \epsilon_i + \sum_{j=1}^{p} \phi_j \big[m_{i-j} - f_k(t_{i-j}) \big], \tag{8}$$

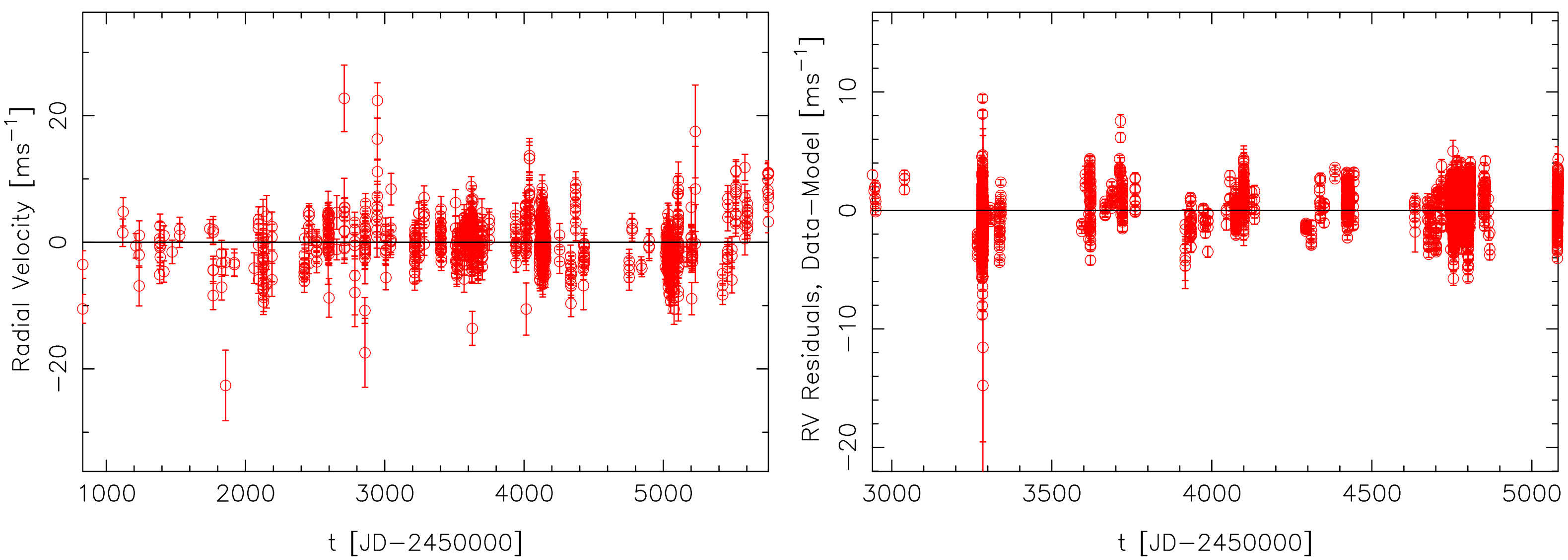
where measurement m_i at epoch t_i is modelled using the function f_k and some convenient white noise component ϵ_i.

The analyses of HARPS radial velocities indicate, that this noise model is much better than pure white noise – and information is not lost if the MA coefficients ϕ_i are selected conveniently (or even better, free parameters).

τ Ceti velocities

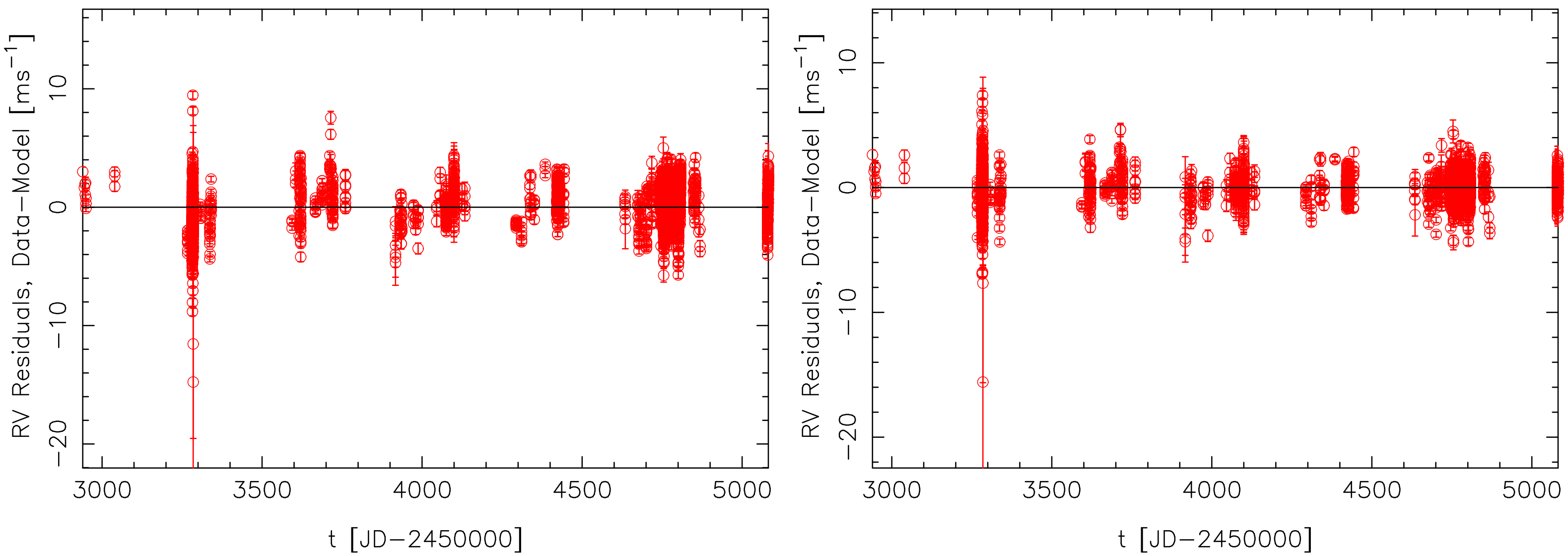
Example: τ Ceti, whose radial velocity curve is considered "flat" despite $\sim$ 4400 HARPS high-precision velocities, $\sim$ 800 AAPS radial velocities, and $\sim$ 1000 HIRES precision velocities from the 10m Keck telescope (though, the star is likely too bright for Keck anyway).

No planets reported from AAPS (left; Wittenmyer et al., 2006) or HARPS (right; Pepe et al., 2011).



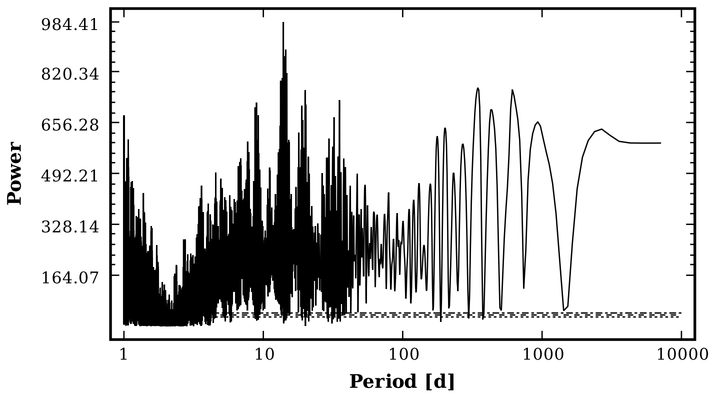
τ Ceti velocities

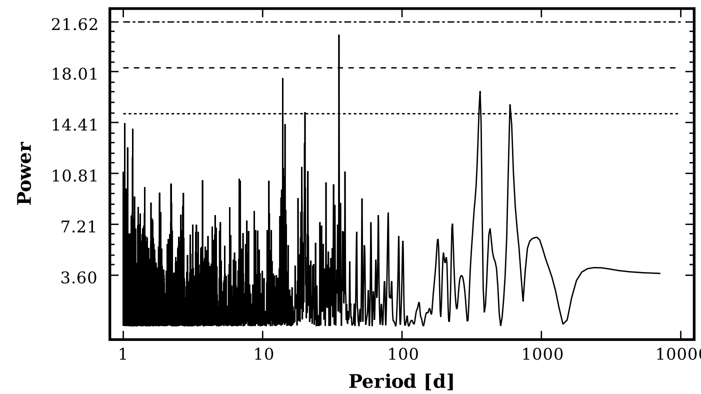
The HARPS velocities before (left) and after (right) removing correlations.



Extremely significant (!) improvement in the statistical model.

τ Ceti velocities



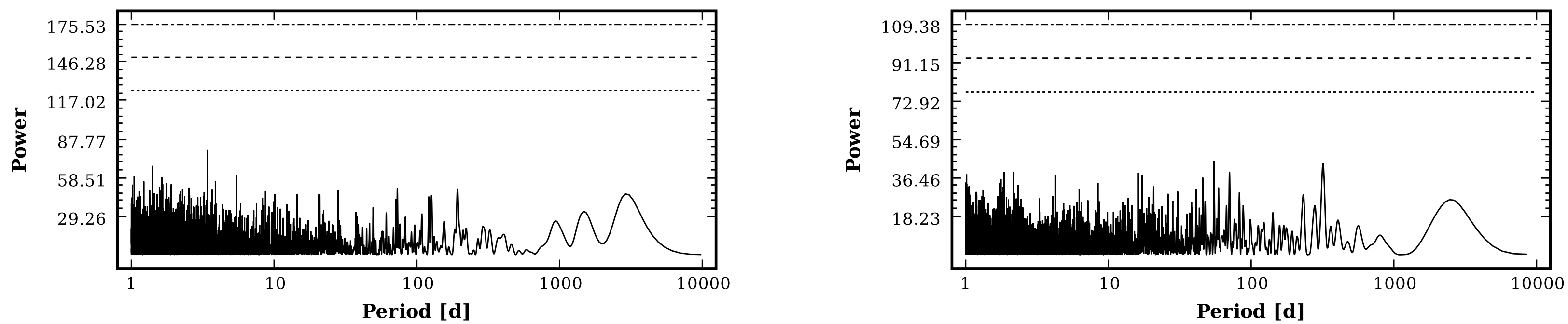


The Lomb-Scargle periodograms of “raw” CCF velocities of τ Ceti (left) and after removing intra-night correlations (right).

The correlation time-scale is 1.19 [0.98, 1.40] h^{-1} and adding the MA components in the model decreases the excess white noise component (the stellar “jitter”) from 1.60 [1.51, 1.69] to 1.06 [1.02, 1.11] ms^{-1}.

What are the periodogram powers (exceeding 10% FAP) at 14, 35, 300, and 600 days?

τ Ceti velocities

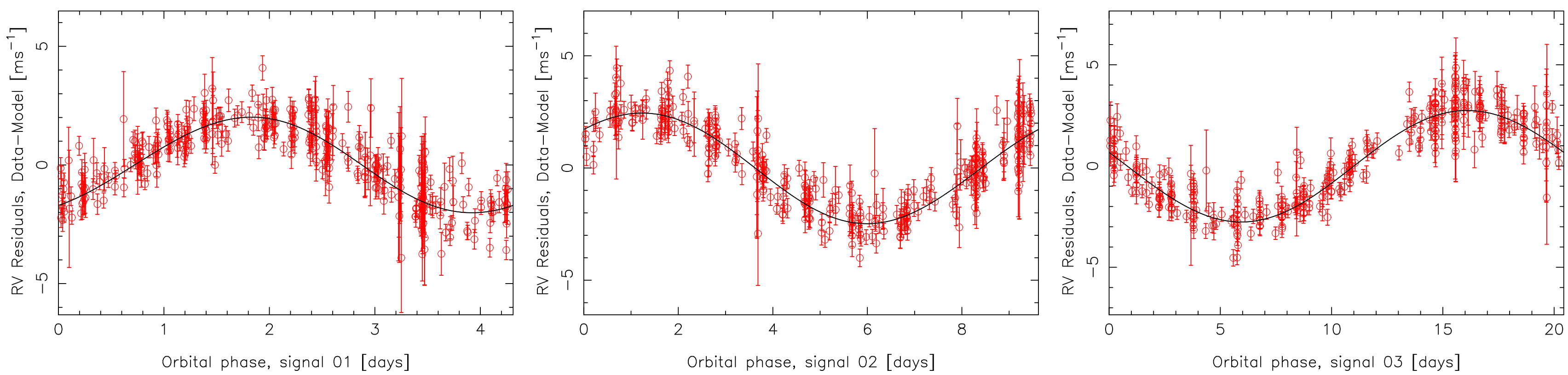


The Lomb-Scargle periodograms of τ Ceti after removing correlations for AAPS (left) and HIRES (right) data.

Very "flat" spectra but the results of Baluev (2012, submitted) and Tuomi & Jenkins (2012, submitted) indicate that models taking into account noise correlations in a time-scale of $\sim$ 10 days can adapt to the signals and decrease their significances in residual periodograms (at least they did for HARPS data of GJ 581).

The peculiar case of HD 40307

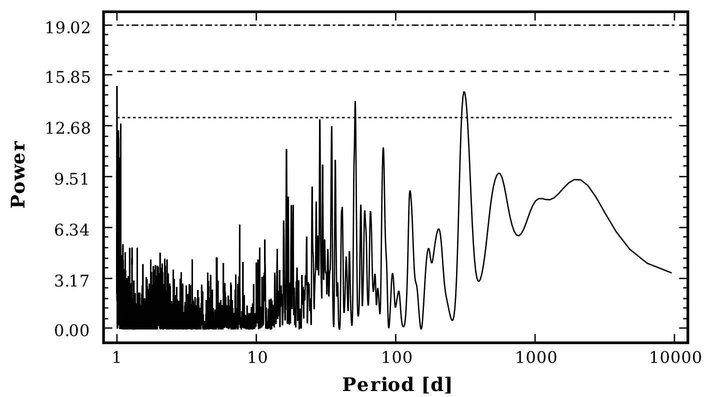
HD 40307 is known to be a system of three super-Earths with orbital periods of 4.31, 9.62, and 20.44 days and minimum masses of 4.3, 7.0, and 10.5 $M_\oplus$, respectively (Mayor et al. 2009).

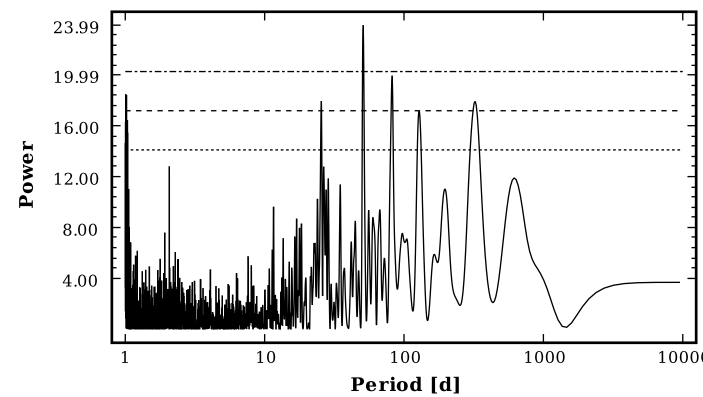


But this is a result received by binning the HARPS radial velocities...

The peculiar case of HD 40307

The residual periodograms of three-Keplerian model.



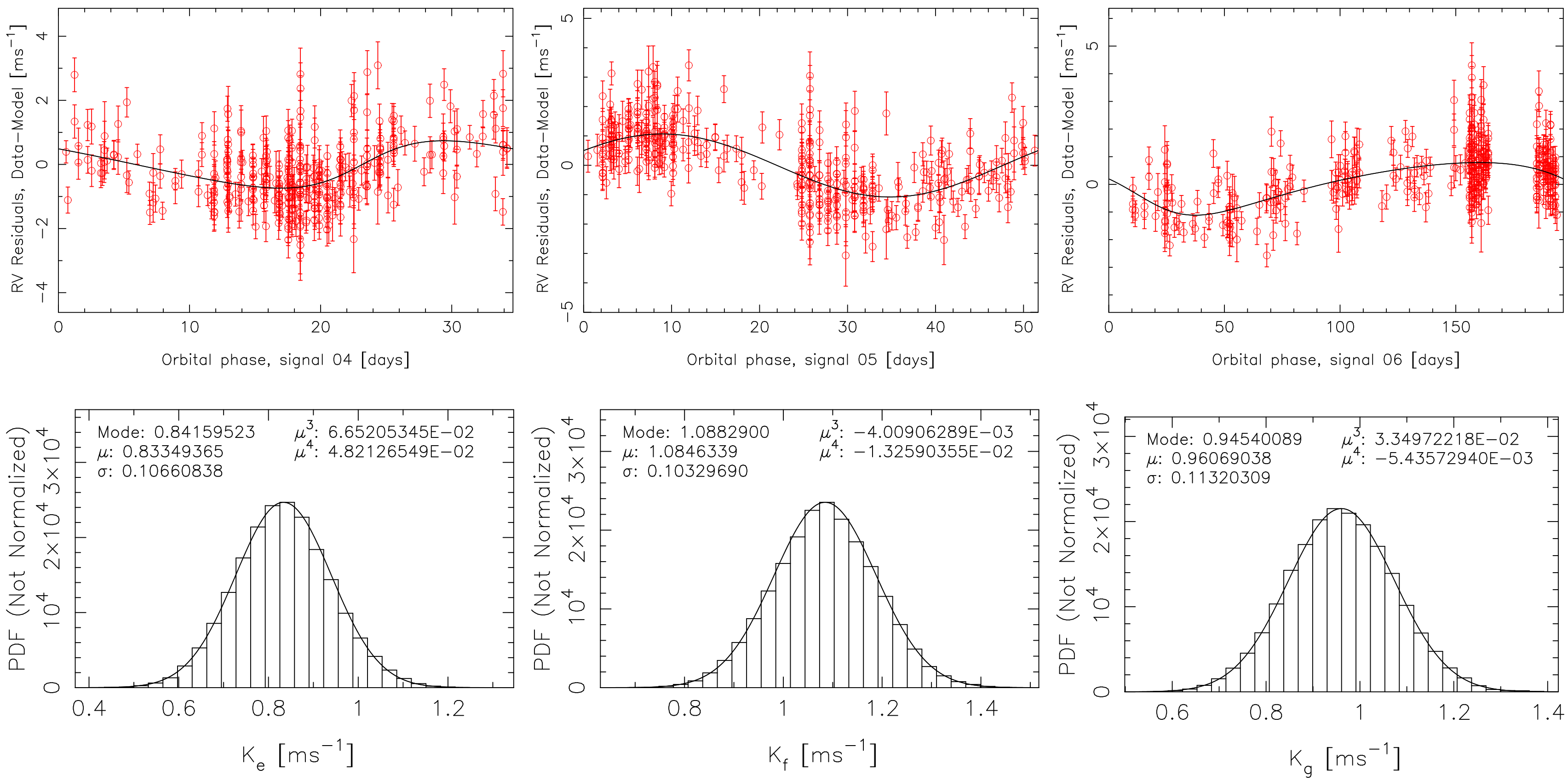


The L-S periodograms of HD 40307 velocities for nightly binned data (left) and all data after removing correlations (right).

The latter is actually obtained by also throwing some data away, i.e. the velocities from the HARPS bluest echelle orders. The 51-day candidate is the dominant power but the activity-induced 300-day periodicity is also visible together with aliases.

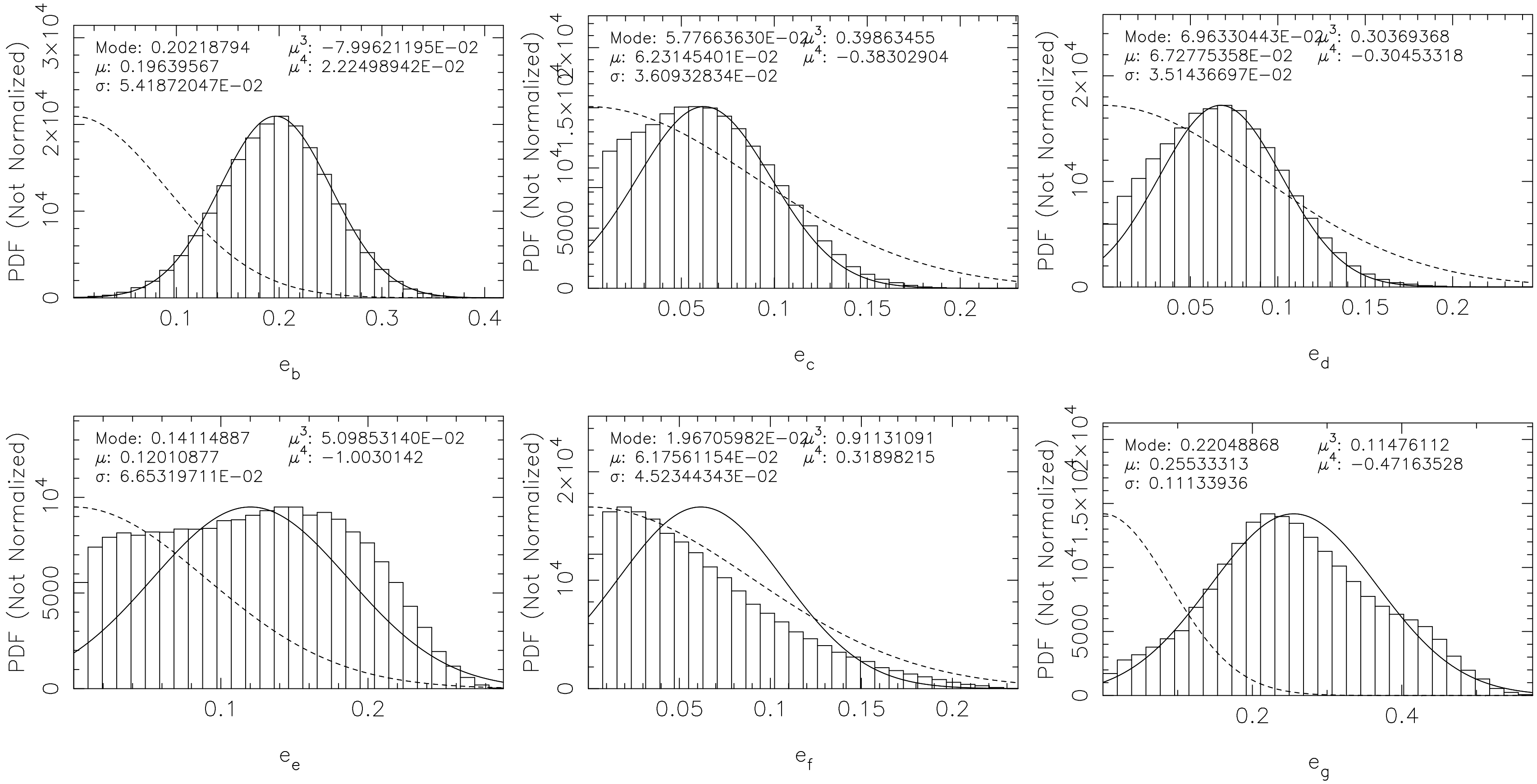
The peculiar case of HD 40307

Three new candidate planets satisfying the detection criteria.



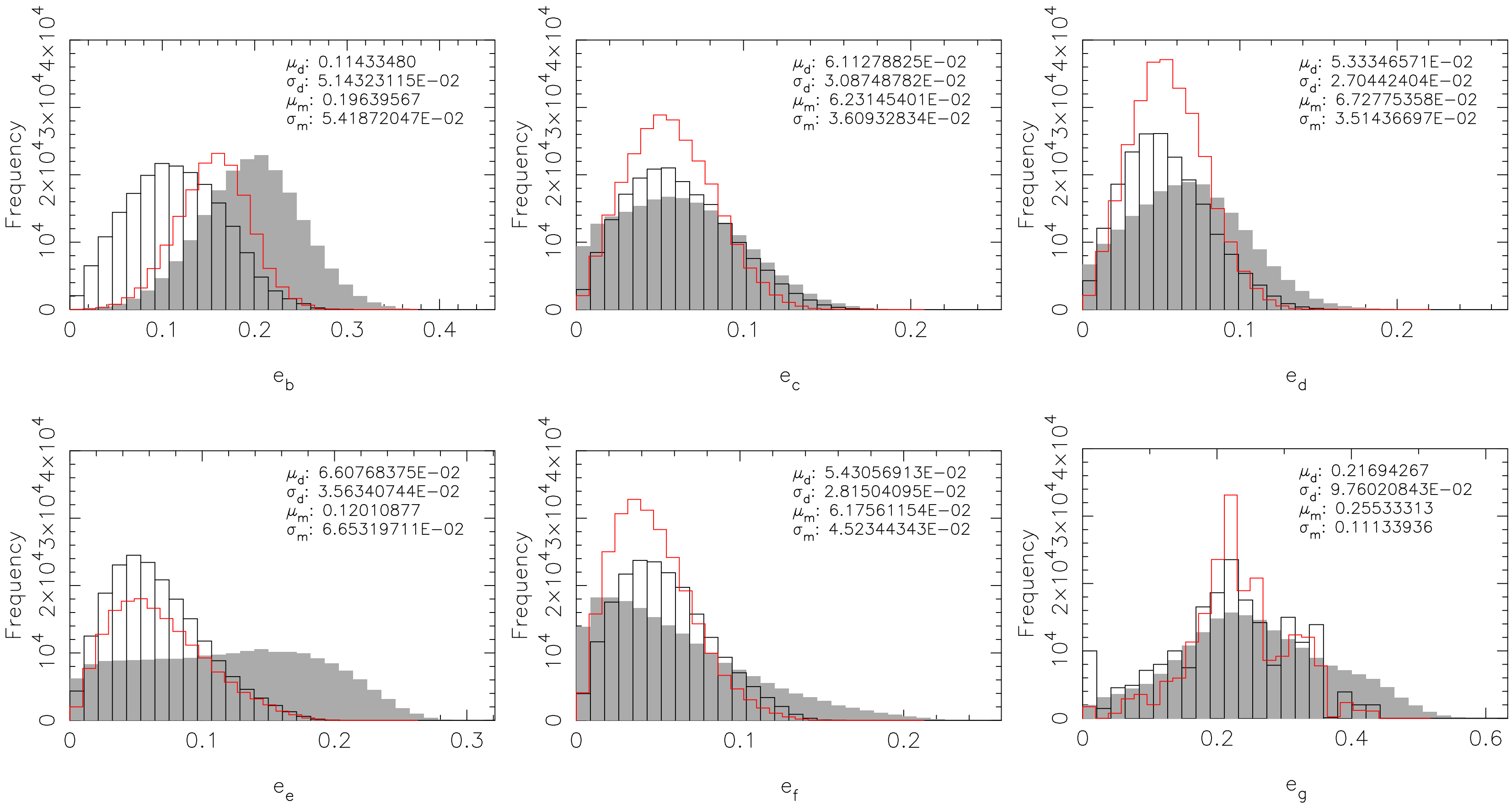
The peculiar case of HD 40307

Comparison of posterior (histogram and solid curve) and prior (dotted curve).



The peculiar case of HD 40307

Additional constraints for eccentricities from dynamical analyses.



False positives

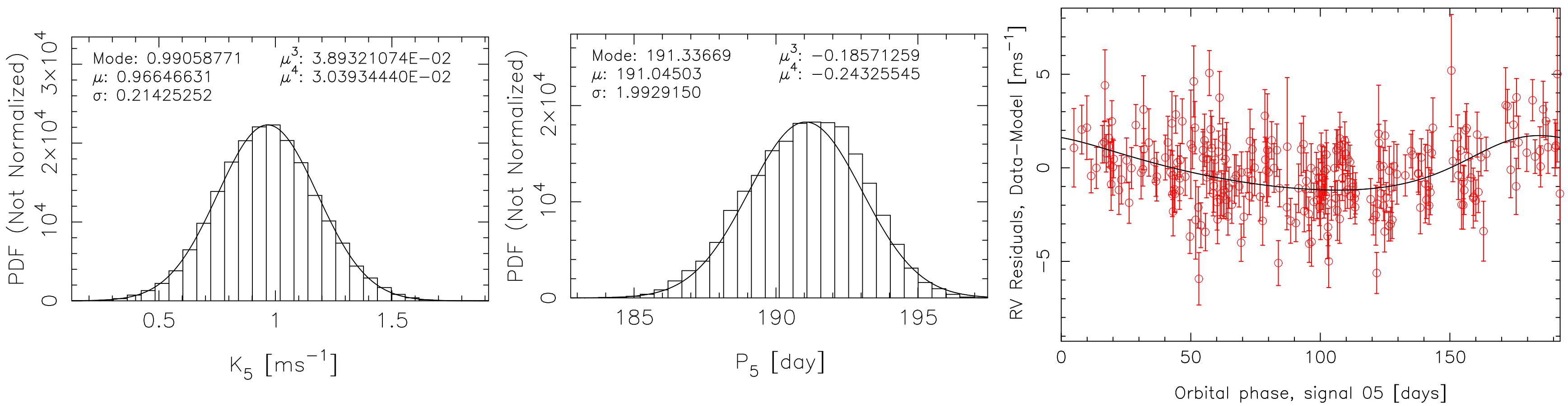
How can one distinguish between false positives and “genuine“ signals?

Reguirements for a ”genuine“ signal:

- Signal present in at least two independent data sets.
- Signal significance increases as a function of amount of data (at least, it does not fall below the detection threshold).
- That the independent data sets are consistent w.r.t. the statistical model (and especially signals within): the *Bayesian model inadequacy criterion* of Tuomi et al. (2011).
- That the noise models of each data set are as ”correct“ as possible.

False positives

Is this a false positive in the HARPS velocities of GJ 581:

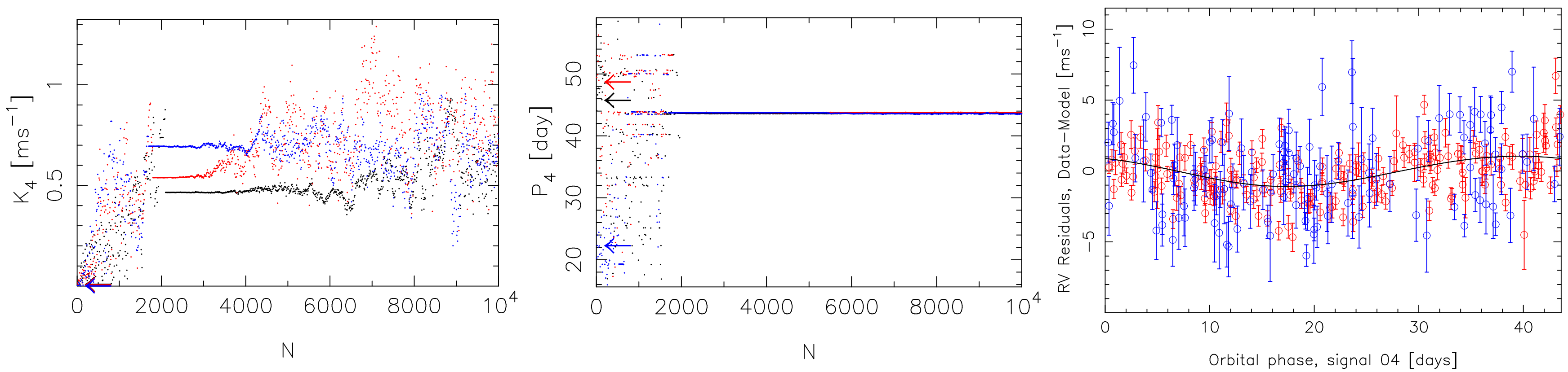


Satisfies all the detection criteria for the Forveille et al. (2011) HARPS data. But not for the combined HARPS and HIRES data set.

What is going on? Poor noise modelling of HARPS, HIRES, or both? A source of noise mimicking periodic behaviour in the HARPS data? Or perhaps biases in the HIRES data hiding the signal?

False positives

Is this a false positive in the combined HARPS and HIRES velocities of GJ 581?



Satisfies the detection criteria for the combined data but only if the noise model consists of pure white noise. Taking into account noise correlations (that does improve the statistical model) makes it impossible to constrain this signal.

A spurious signal caused by insufficient noise modelling? Or a genuine signal that falls below detection thresholds because the noise model is still not "correct"?

Thank You for Your attention

Baluev, R. V., 2012, MNRAS, submitted (astro-ph.EP)

Barnes, R. & Greenberg, R. 2006, ApJ, 647, L163

Forveille, T., Bonfils, X., Delfosse, X., et al., 2011, ApJ, submitted (astro-ph.EP)

Mayor, M., Udry, S., Lovis, C., et al. 2009a, A&A, 493, 639

Pepe, F., Lovis, C., Ségransan, D., et al. 2011, A&A, 534, A58

Tuomi, M. 2012, A&A, 543, A52

Tuomi, M. & Jenkins, J. S. 2012, A&A, submitted (astro-ph.EP)

Tuomi, M., Anglada-Escudé, G., Gerlach, E., et al. 2012, A&A, in press

Tuomi, M., Pinfield, D., & Jones, H. R. A. 2011, A&A, 532, A116

Wittenmyer, R. A., Endl, M., Cochran, W. D., et al. 2006, AJ, 132, 177