*SF2A 2021*
*A. Siebert, K. Baillié, E. Lagadec, N. Lagarde, J. Malzac, J.-B. Marquette, M. N'Diaye, J. Richard, O. Venot (eds)*

# A NEW GENERATION OF STATISTICAL METHODS FOR EXOPLANET DETECTION WITH RADIAL VELOCITY

N. C. Hara[1]

**Abstract.** Radial velocities are essential to estimate the mass of exoplanet, and have the potential to detect Earth-like planets around Sun-like stars. However, radial velocity data are corrupted by instrumental systematics and complex stellar noises, which currently inhibit the detection of Earth analogs. We present new statistical and computational tools for the analysis of time-series, which show greater detection power than previous methods. We then discuss the scientific impact of the improvement of radial velocity precision.

Keywords: Exoplanets, radial velocity, statistics

## 1 Introduction

Since the discovery of the first extra-solar planets around a pulsar in 1992 Wolszczan & Frail (1992), and a Solar type star in 1995 Mayor & Queloz (1995) over four thousands exoplanets have been detected. Most of these detections have been obtained by two techniques: transits, which consist in searching for periodic dimming of the stellar light, indicating that a planet passes between the observer and the star, and radial velocities (or RV), where one searches for periodic variations of the star velocity along the line of sight. As of sept. 6 2021, over seven hundred planets have been discovered thanks to the RV technique.

RVs are essential for several reasons. First, they provide a direct measurement of the planetary projected mass. Combined with the planetary radius, a mass measurement yields the density and surface gravity, which are essential to study internal structures and planetary atmospheres (Batalha et al. 2017). One of the core science cases of the space mission PLATO is to provide a population of exoplanets with known masses, and thus, unlike Kepler (Borucki et al. 2011) will focus on bright stars, and aims at a 10% precision on the mass of planets. Secondly, RVs provide a crucial insight on the architecture of planetary systems. As the orbital period grows, the transit probability decreases. The RVs detection bias is less strong in that parameter space, they were used in particular to study giant planet demographics beyond the ice line Rosenthal et al. (2021); Fulton et al. (2021). Furthermore, transit surveys do not allow the detections of planetary systems with high mutual inclination such as HD 158259 (Hara et al. 2020). Finally, RVs cannot be circumvented for the detection and characterisation of Earth-like planets, as primary detection technique or follow-up of transit candidates. Instruments such as NIRPS and ESPRESSO (Pepe et al. 2021) are conceived with the goal of detecting Earth-like planets in the habitable zone (as a detection technique or to validate transit detections). The atmosphere of these planets will be probed either with transit spectroscopy, or direct imaging for objects within 15 parsecs from the Sun, and this for the best case scenarios of PCS (Kasper et al. 2021), LUVOIR[*], and HabEx[†] (Bouchy et al. 2017).

An Earth-mass planet in the habitable zone of a K dwarf produces a RV signal of the order of tens of centimeters per second, and the last generation spectrograph ESPRESSO demonstrated a stability of 25 cm/s. However, stellar features (granulation, plages and faculae) cause complex, poorly modelled radial velocity signals of the order of at least 50 cm/s (1 m/s for the Sun), which are added to potential instrument systematics (see Hara 2017). As low frequency noises, stellar and instrumental effects are particularly problematic in a parameter space where RV has an excellent potential: small planets at long periods. If not mitigated, these effects inhibits detections of terrestrial planets in the habitable zone of K and G stars, and population studies of planets of Saturn mass and below with periods ($\gtrsim 300$ d)

[1] Observatoire de Genève, Chemin de Pegasi 51, 1290, Versoix, Switzerland
[*] `https://asd.gsfc.nasa.gov/luvoir/resources/`
[†] `https://www.jpl.nasa.gov/habex/documents/`

## 2 Radial velocity data

Radial velocities are acquired as follows: the observer takes spectra of a star at different epochs. If the star has a non-zero velocity in the direction of the line of sight, the spectra will be shifted in frequency thanks to the Doppler effect. From this shift, one extracts the radial velocity. If a planet orbits the star, this one has a reflex motion so that the radial velocity of the star exhibits a periodic pattern at the orbital period. In Fig. 1, we show as an example 190 RV measurements of HD10180 made with the HARPS spectrograph (Lovis et al. 2011), which have a typical nominal uncertainty of 0.6 m/s, but in practice there are other sources of astrophysical and instrumental noise, showing complex structures (e. g. Saar & Donahue 1997; Meunier et al. 2010; Dumusque et al. 2015).

In the case where several planet orbit the star, if the gravitational interaction between the planets can be neglected on the observation timespan – which is true most of the time – the RV signals of the different planets add up. Searching for planets in radial velocity essentially comes down to a classical time series analysis problem: finding parametric periodic signals in an unevenly sampled timeseries, embedded in complex noises.

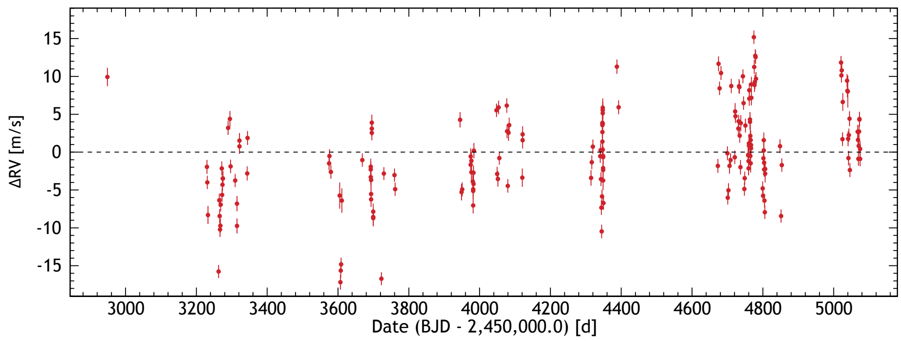


**Fig. 1.** 190 HARPS radial velocity measurements of HD10180 (Lovis et al. 2011).

## 3 New statistical tools

### 3.1 Detection criterion

Exoplanet d etections are typically claimed based on the computations of a statistical significance metric, which often are either the “Bayes factor” (e. g. Gregory 2007; Tuomi 2011; Díaz et al. 2016) or a $p-$value associated to a periodogram (e. g. Baluev 2008, 2009; Zechmeister & Kürster 2009). If the significance is greater than a certain thresold, a detection is claimed. In Hara et al. (2021b), we define a new metric called the true inclusion probability (TIP) (Hara et al. 2021b). The TIP is defined as the probability to have a planet whose frequency lies in a certain frequency interval. We then define the probability of absence of a planet as the False inclusion probability (FIP) = 1 - TIP.

In Fig. 2, we show a comparison of the different detection criteria. We generate a thousand mock RV datasets and, for different metrics, compute the number of false detections and missed detections as the detection threshold varies. The $x$ axis of Fig. 2 shows the number of false detections, and the $y$ axis shows the sum of false and missed detections. The FIP (in purple) shows a smaller number of missed detections for a given number of false detections, and it can be mathematically proven that, provided an accurate model of the data, the FIP outperforms other detection criteria. However, stellar and instrumental noises models might be inaccurate. To further test the reliability of detections, one can test whether the candidate planetary signals have a constant phase, amplitude and frequency – as they should exhibit – or display significant variations, indicating a non planetary origin of the signal (Hara et al. 2021a).

### 3.2 Numerical methods

The classical tool to model stellar activity is Gaussian processes. Manipulating them requires heavy computations, in particular matrix inversions. For some forms of the noise models, the computations can be brought from $O(N^3)$ to $O(N)$, which is the principle behind the CELERITE package (Foreman-Mackey et al. 2017).

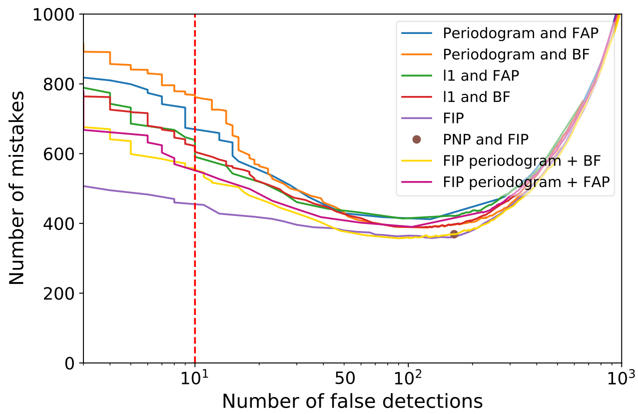


**Fig. 2.** Number of mistakes as a function of the number of false detections (log scale) for the different detection methods. The data corresponds to a simulation of 1000 datasets, where a random number of planets (0,1 or 2) are injected with, white and correlated noise. Mistakes are defined as the sum of missed and false detections. See Hara et al. (2021b) for details.

In Delisle et al. (2020), the CELERITE model is generalised to a wider class of noise models, which can in particular model simultaneously instrumental and stellar effects.

### 3.3 *Extracting velocities*

The previous techniques concern the time series of radial velocity. In those, the information contained in the spectrum is reduced to the RV as well as a few activity indicators. While planets affect identically all the spectral lines as a pure Doppler shift, activity and instrumental effects might have different effects depending on the wavelength. To exploit fully this property, one has to analyse the time series of spectra, which is the approach taken in Cretignier et al. (2021).

As we can see in Fig. 3, the map of spectra residuals after YARARA (bottom) is significantly cleaner than before its application (upper panel). Different recipes in YARARA allows a strong mitigation of the impact of cosmics, telluric and micro-telluric effects (e.g. Cunha et al. (2014)), ghosts (Dumusque et al. 2021) and HARPS detector stitching (Dumusque et al. 2015).

## 4 Conclusion

By combining the new processing of the spectra with the new numerical tools and detection criteria, we are able to significantly improve the capabilities of the radial velocity method to detect terrestrial planets in the habitable zone. To test our methods, we applied them to a hundred bright stars of the HARPS spectroscopic data, showing a nominal precision of 0.7 - 1 m/s. We are able to retrieve with high confidence three new candidate signals with minimum masses below 2 $M_{\oplus}$ and periods greater than 30 days, while there was only one known so far ($\tau$ ceti h, Feng et al. 2017). This analysis is still in progress. We have seen in section 1 that RVs are essential for mass measurements, probing outer regions and for the detection of planets in the habitable zone. It will play for instance a major role in the follow-up of the PLATO photometric mission. With the improvements made and further refinements, the use of RV for demographics studies will gain in precision and give access to Neptunes beyond the ice line., and eventually the detection and characterisation of terrestrial planets in the habitable zone.

This work has been carried out within the framework of the NCCR PlanetS supported by the Swiss National Science Foundation

## References

Baluev, R. V. 2008, MNRAS, 385, 1279
Baluev, R. V. 2009, MNRAS, 393, 969
Batalha, N. E., Kempton, E. M. R., & Mbarek, R. 2017, ApJ, 836, L5

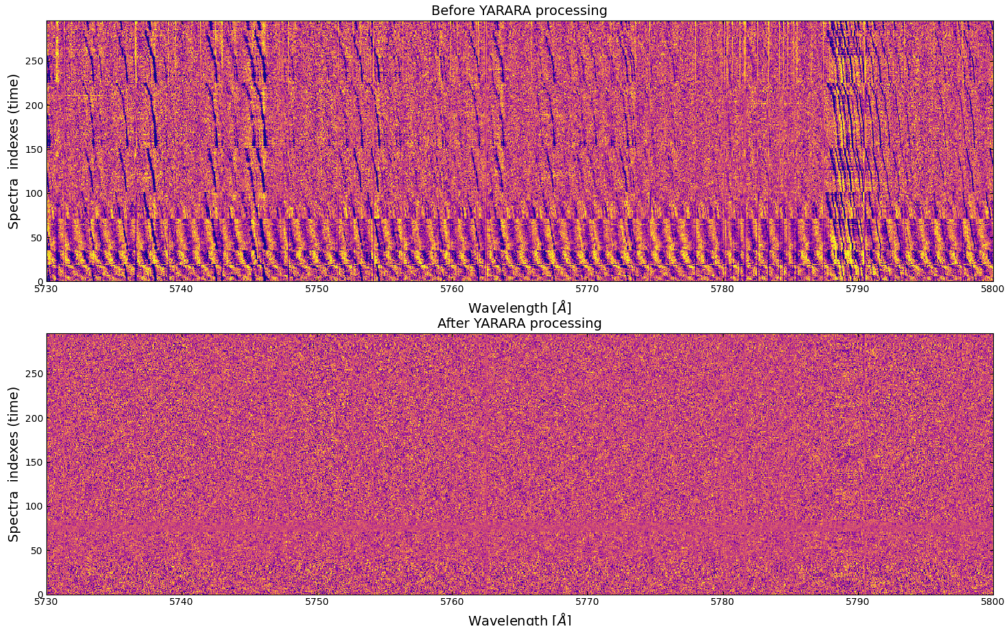


**Fig. 3.** Stacked HARPS spectra represented in color code (1 row is one spectrum for alpha Cen B) **Top:** Spectra before YARARA corrections. Spectra are suffering from interference patterns on two observational seasons (from index 0 to 71). **Bottom:** Residual river diagram after YARARA correction.

Borucki, W. J., Koch, D. G., Basri, G., et al. 2011, The Astrophysical Journal, 736, 19
Bouchy, F., Doyon, R., Artigau, É., et al. 2017, The Messenger, 169, 21
Cretignier, M., Dumusque, X., Hara, N. C., & Pepe, F. 2021, arXiv e-prints, arXiv:2106.07301
Cunha, D., Santos, N. C., Figueira, P., et al. 2014, A&A, 568, A35
Delisle, J. B., Hara, N., & Ségransan, D. 2020, A&A, 638, A95
Díaz, R. F., Ségransan, D., Udry, S., et al. 2016, A&A, 585, A134
Dumusque, X., Cretignier, M., Sosnowska, D., et al. 2021, A&A, 648, A103
Dumusque, X., Pepe, F., Lovis, C., & Latham, D. W. 2015, ApJ, 808, 171
Feng, F., Tuomi, M., Jones, H. R. A., et al. 2017, AJ, 154, 135
Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, AJ, 154, 220
Fulton, B. J., Rosenthal, L. J., Hirsch, L. A., et al. 2021, ApJS, 255, 14
Gregory, P. C. 2007, MNRAS, 381, 1607
Hara, N. C. 2017, PhD thesis, Observatoire de Paris
Hara, N. C., Bouchy, F., Stalport, M., et al. 2020, A&A, 636, L6
Hara, N. C., Delisle, J.-B., Unger, N., & Dumusque, X. 2021a, arXiv e-prints, arXiv:2106.01365
Hara, N. C., Unger, N., Delisle, J.-B., Díaz, R., & Ségransan, D. 2021b, arXiv e-prints, arXiv:2105.06995
Kasper, M., Cerpa Urra, N., Pathak, P., et al. 2021, arXiv e-prints, arXiv:2103.11196
Lovis, C., Ségransan, D., Mayor, M., et al. 2011, A&A, 528, A112
Mayor, M. & Queloz, D. 1995, Nature, 378, 355
Meunier, N., Desort, M., & Lagrange, A.-M. 2010, A&A, 512, A39
Pepe, F., Cristiani, S., Rebolo, R., et al. 2021, A&A, 645, A96
Rosenthal, L. J., Fulton, B. J., Hirsch, L. A., et al. 2021, ApJS, 255, 8
Saar, S. H. & Donahue, R. A. 1997, ApJ, 485, 319
Tuomi, M. 2011, A&A, 528, L5
Wolszczan, A. & Frail, D. A. 1992, Nature, 355, 145
Zechmeister, M. & Kürster, M. 2009, A&A, 496, 577