Universidad Nacional de San Martín

Machine learning for exoplanet detection using the radial velocity method

Nicolas Waehner[1], Luis Agustín Nieto[2, 3], Rodrigo F. Díaz[3]

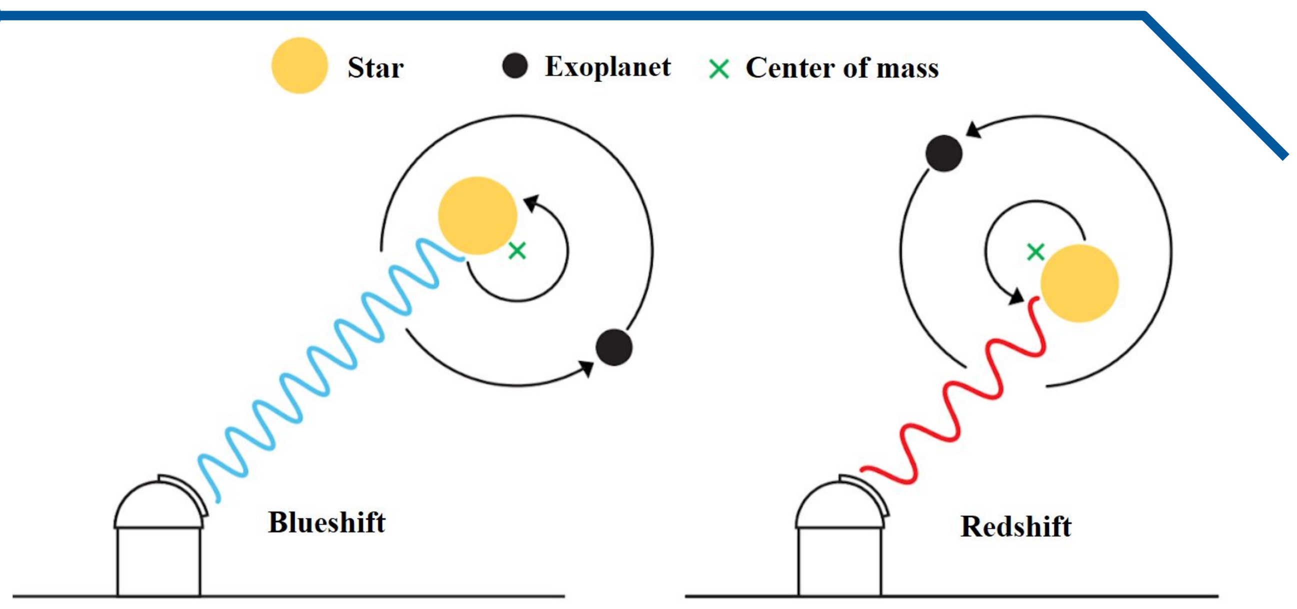
[1] **University of Buenos Aires (UBA)**, Faculty of Exact and Natural Sciences, Department of Physics, Buenos Aires, Argentina.
[2] **Information and Communication Technology Management (GTIC)**, DTE, National Atomic Energy Commission (CNEA), Constituyentes Atomic Center, San Martín, Buenos Aires, Argentina.
[3] **International Center for Advanced Studies (ICAS)** and **ICIFI (CONICET)**, School of Science and Technology (ECyT), National University of San Martín (UNSAM), San Martín, Buenos Aires, Argentina.

The radial velocity (RV) method has proven to be one of the most successful and promising techniques for detecting planets through the motions they induce on their host stars. While instrumental improvements have enabled the measurement of increasingly smaller velocity variations, stellar activity and irregular sampling can make the detection of planetary signals more difficult and lead to false positives. For this reason, machine learning techniques have recently begun to be explored to address this challenge.

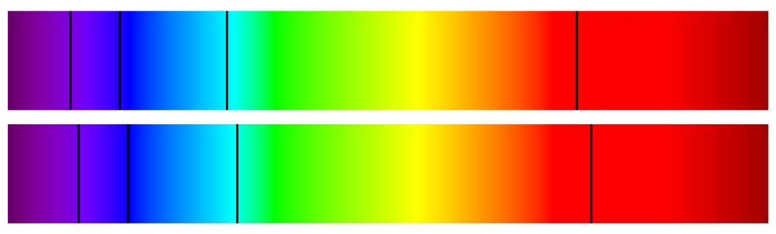
In this work, we develop a convolutional neural network with an attention layer to detect planetary signals in Sun-like stars, using simulated RV measurements generated over observation calendars representative of exoplanet searches, following the study by Nieto & Díaz (2023) [1]. The network achieves 54% fewer false positives than the traditional null-hypothesis-based approach, without increasing the number of false negatives. This improvement is mainly concentrated in low amplitude signals, associated with low mass planets. In addition, the attention layer weights were analyzed to identify which regions of the input the model prioritizes during classification, revealing a correlation between these weights and the network's predictions.

The method was further evaluated on 159 real signals from stars with at least one confirmed planet and achieved correct classifications in most cases. These data were also used to perform fine-tuning on the network, enhancing its detection capability on real observations. Overall, these results highlight the potential of neural networks as a promising tool for the detection of planetary signals in radial velocity measurements.

Radial velocity method



The **radial velocity** can be obtained from the shift of the absorption lines in the spectrum.



$$\lambda = \lambda_0 \left(1 + \frac{v_r}{c}\right)$$

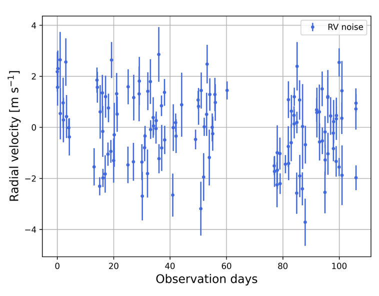
The amplitude K of the oscillation [2] is given by

$$K = \left(\frac{2\pi G}{P}\right)^{1/3} \frac{M_p \sin i}{(M_\star + M_p)^{2/3}} \frac{1}{\sqrt{1-e^2}}$$

Time series

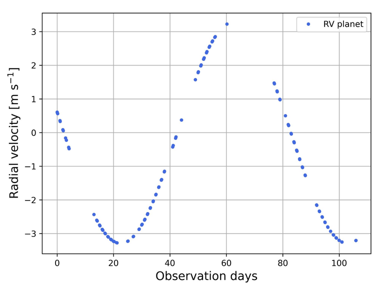
The goal is to evaluate the **detectability of exoplanets** in the presence of **realistic stellar noise** and **observation calendars representative** of exoplanet searches.

To this end, we simulate the **radial velocity** over these calendars, taking into account:



Stellar activity and errors

- Granulation and pulsations [4]
- Rotational modulation
- Photon noise
- Instrument calibration

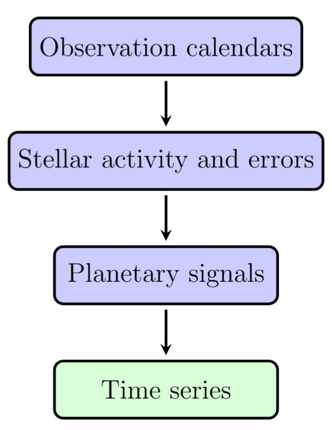


Planetary signals

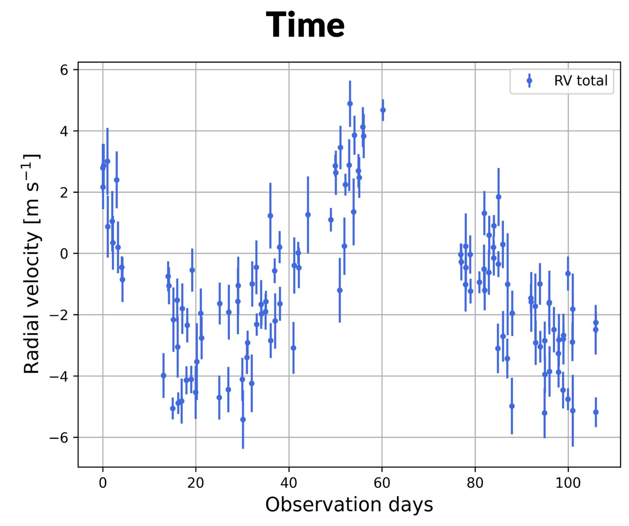
- Circular orbits
- Non-interacting

$$RV(t) = K \sin\left(\frac{2\pi(t-T_0)}{P}\right)$$

$$K \sim \mathcal{U}_{\log}(0.1\,\mathrm{m\,s^{-1}}, 10\,\mathrm{m\,s^{-1}})$$
$$P \sim \mathcal{U}(5\,\mathrm{d}, 200\,\mathrm{d})$$



Time

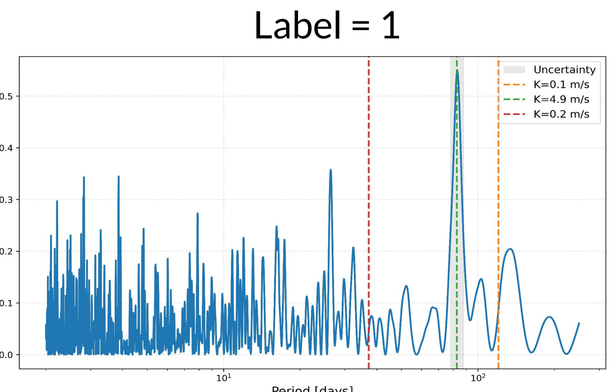


Periodograms and labels

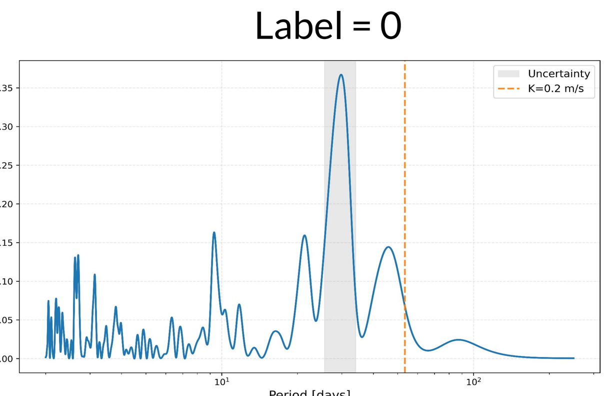
We generate **generalized Lomb-Scargle (GLS) periodograms** from the time series and label them.

$$y(t) = a\cos(\omega t) + b\sin(\omega t) + c \;\Rightarrow\; \chi^2 = \sum_{i=1}^{N} \frac{[y_i - y(t_i)]^2}{\sigma_i^2} \;\Rightarrow\; p(\omega) = \frac{\chi_0^2 - \chi^2(\omega)}{\chi_0^2}$$

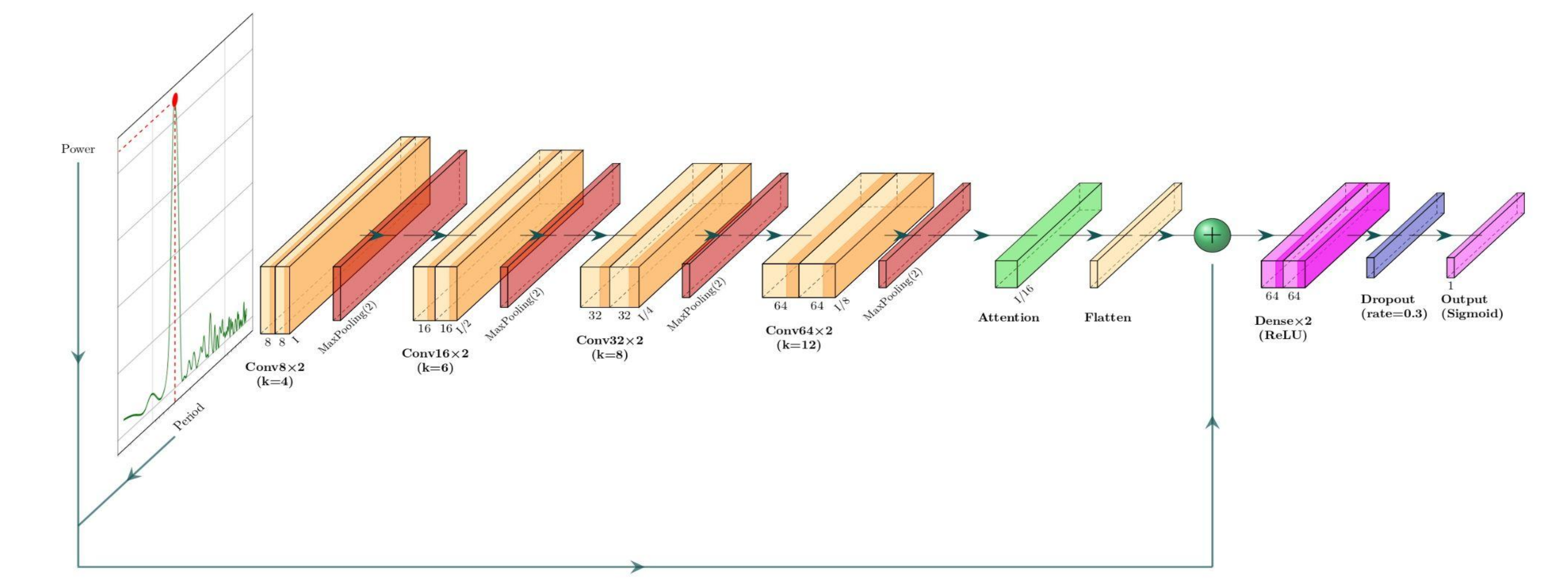
Label = 1



Label = 0

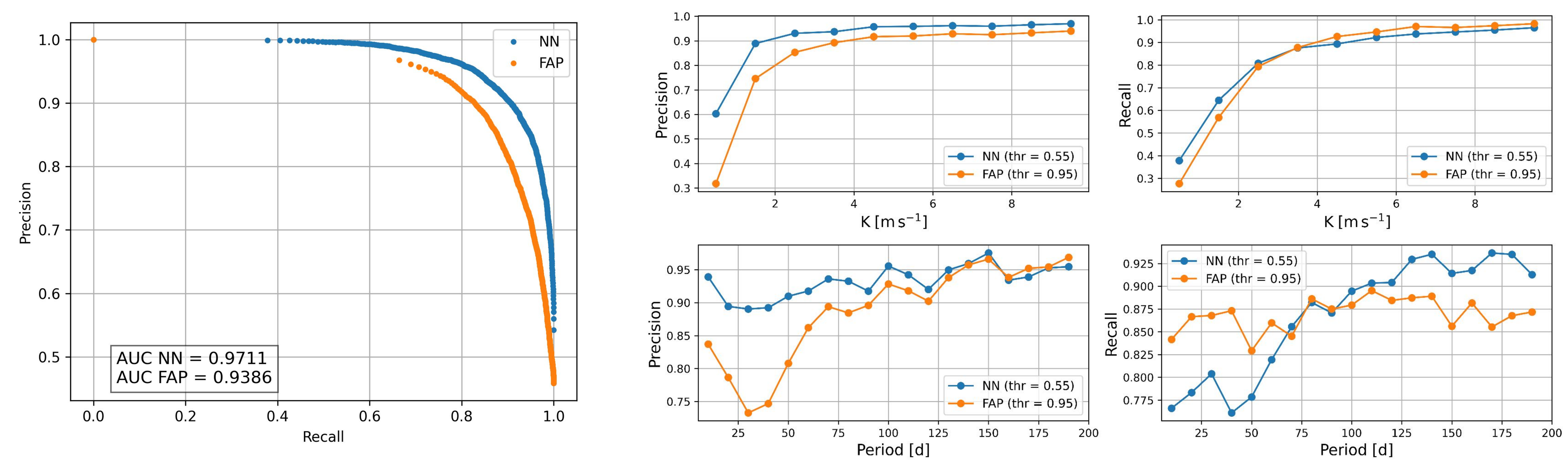


Neural Network

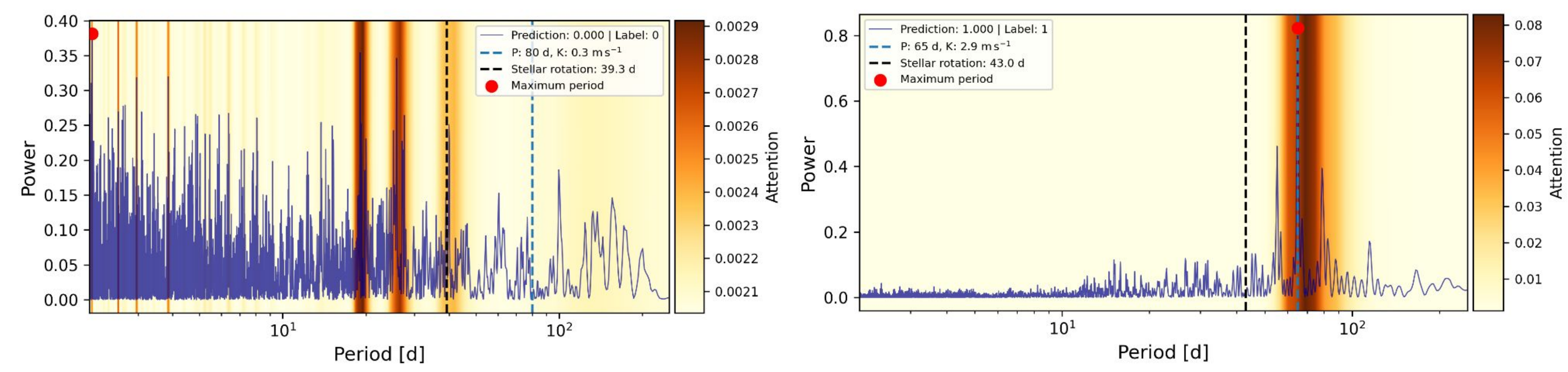


Quality and characteristics of the detections

We compare the results of our neural network with the **false alarm probability (FAP)** to evaluate the **significance** of peaks in periodograms.

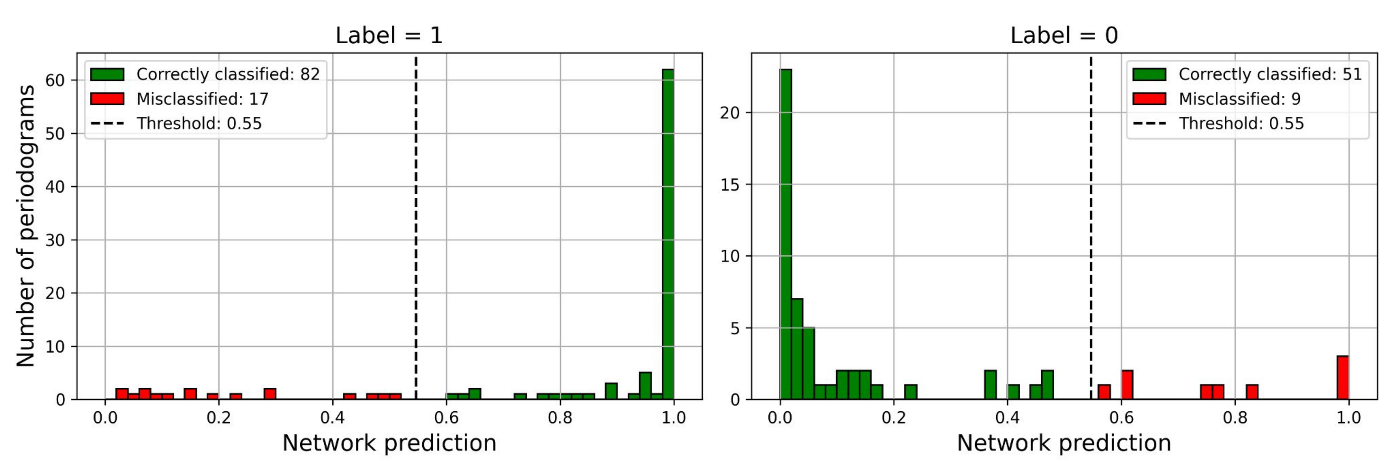


We obtained the **attention weights** and plotted the **average attention** that each point in the periodogram receives from the rest.



Real data

The performance of the neural network was evaluated on a set of **159 periodograms** obtained from **real observational data**.



Conclusions

[✓] The neural network outperforms the traditional statistical method on simulated data, reducing false positives by 54%.

[✓] It is more accurate in detecting low-amplitude planetary signals and close to the stellar rotation period.

[✓] The attention layer focuses its processing on the peaks, and it was found that lower entropy corresponds to higher network prediction.

[✓] The neural network works on real data and can be better adapted to them through fine-tuning techniques.

GitHub & Work

References

[1] Nieto, L. A. & Díaz, R. F. (2023). Exoplannet: A deep learning algorithm to detect planetary signals in RV data. A&A, 677, A48.

[2] Díaz, R. F. (2018). Modelling Light and Velocity Curves of Exoplanet Hosts. In Asteroseismology and Exoplanets: Listening to the Stars and Searching for New Worlds, ASSP, vol. 49, p.199.

[3] Haywood, R. D., et al. (2014). Planets and stellar activity in the CoRoT-7 system. Monthly Notices of the Royal Astronomical Society, 443, 2517–2531

[4] Dumusque, X., Udry, S., Lovis, C., Santos, N. C., & Monteiro, M. J. P. F. G. (2011). Planetary detection limits considering stellar noise. I. Strategies to reduce oscillation and granulation effects. A&A, 525, A140.