Open Access Online Journal on Astronomy and Astrophysics

Acta Astrophysica Taurica

www.astrophysicatauricum.org

Acta Astrophys. Tau. 7(1), 1–3 (2026) EDN: SVNIOT

Correction for observational selection effects when analyzing the statistics of exoplanets discovered with the radial velocity technique

V.I. Ananyeva, A.V. Tavrov, O.I. Korablev, A.A. Salamashkina

Space Research Institute (IKI) of the Russian Academy of Sciences, Profsoyuznaya 84/32, Moscow 117997, Russia
e-mail: a-lada@yandex.ru

Received 2 October 2023

ABSTRACT

We propose an improved method for correcting observational selection effects when studying the statistics of exoplanets discovered with the radial velocity technique. We verify the proposed method through the numerical simulation of planets whose mass distribution follows a power law with an exponent of −1, −1.5, −2, −2.5, −3. The Lomb – Scargle periodograms are used to evaluate the detectability of planets. The proposed algorithm taking into account the number of radial velocity measurements is shown to successfully reconstruct the minimum mass distribution of exoplanets.

Key words: exoplanets, radial velocity technique, observational selection

1 Introduction

The radial velocity (RV) technique is one of the most effective techniques for detecting exoplanets, which also enables the reliable mass determination of transiting planets. As of early September 2023, over one thousand exoplanets have been discovered using the RV technique. This number is enough to carry out a statistical analysis of their distributions, in particular by masses and orbital periods. However, the distributions extracted directly from catalogs (e.g., the NASA Exoplanet Archive, the Encyclopaedia of Extrasolar Planets) are significantly affected by observational selection. The planets listed in these catalogs were detected using spectrographs with varying levels of instrumental uncertainty, different numbers of radial velocity measurements, observational programs of different durations, and in stars with different levels of activity. All this makes the existing catalogs substantially inhomogeneous.

To correct for this inhomogeneity, a detectability window regularization algorithm was proposed by Ivanova et al. (2021), which for brevity we will hereafter refer to as the “detectability window” method. The method is based on a matrix whose elements represent the detectability of a planet with a given orbital period and minimum mass (P, m) using a combined set of observational programs. To account for observational selection effects, each known exoplanet is considered with a statistical weight inversely proportional to its detectability.

Early publications on exoplanet statistics did not take into account the inhomogeneity of observational data. Thus, Butler et al. (2006) constructed a minimum mass distribution of the 167 exoplanets known at that time and approximated it by a power law, $\mathrm{d}N/\mathrm{d}m \propto m^{-1.1}$, without taking into account differences between observational programs. Marcy et al. (2005) attempted to resolve this problem by considering only planets discovered at the Lick Observatory and the W.M. Keck Observatory using spectrographs with an identical single-measurement instrumental uncertainty of 3 m/s, thereby selecting 104 planets out of the 152 known at that time. Marcy et al. (2005) found that the distribution follows a power law, $\mathrm{d}N/\mathrm{d}m \propto m^{-1}$.

Analyzing the mass distribution of planets orbiting 166 solar-type stars observed at the Keck Observatory with the HIRES spectrograph, Howard et al. (2010) introduced a completeness function $\mathrm{C}(P, m)$ defined as a fraction of stars for which the presence of a companion planet with a given period and minimum mass can be confidently excluded. Howard et al. (2010) demonstrated that the mass distribution of planets with periods shorter than 50 days can be approximated by a power law, $\mathrm{d}N/\mathrm{d}\log(m) \propto m^{-0.48^{+0.12}_{-0.14}}$, which corresponds to $\mathrm{d}N/\mathrm{d}m \propto m^{-0.48^{+0.12}_{-0.14}}$.

Tuomi et al. (2019) analyzed 23 473 individual radial velocity measurements of 426 M dwarfs acquired by the HARPS, HIRES, PFS, UVES, and other spectrographs. To account for the varying duration and sensitivity of the observational programs, the detection probability function $p_i(\Delta m, \Delta P)$ was introduced for each observed star, taking discrete values of 0 or 1 (1 – if the derived data assume the existence of a planet within the mass and orbital period $(\Delta m, \Delta P)$, and 0 – otherwise). The general planet detection probability function $f_p(\Delta m, \Delta P)$ was obtained by summing all $p_i(\Delta m, \Delta P)$ according to the number of observed stars (N = 426, Tuomi et al., 2019):

$$f_p(\Delta m, \Delta P) = 1/N \sum_{i=1}^{N} p_i(\Delta m, \Delta P). \quad (1)$$

The mass and orbital period ranges $(\Delta m, \Delta P)$ were represented by an 8×8 grid, with orbital periods spanning the interval $P = 1 \div 10^4$ days and minimum masses spanning $m = 1 \div 10^3$ Earth masses.

Tuomi et al. (2019) did not aim to investigate the mass distribution of planets but attempted to determine the occurrence rates of planets around M dwarfs. Nevertheless, the method proposed by Tuomi et al. (2019), after certain modifications, was adopted as the basis for studying the mass distribution of RV detected planets around stars of all spectral types.

Previously, an empirical approach was used to compute the "detectability window" matrix (Ananyeva et al., 2022, 2023). For each planet-hosting star, the noise level was estimated using the mean deviation from the best-fit Keplerian curve $\sigma(O-C)$ and the whole duration of observations T. A hypothetical planet with the orbital period and mass (P, m) was considered detectable if two conditions were simultaneously satisfied: its orbital period P was shorter than twice the total observational time, while the radial velocity semi-amplitude K induced by the planet exceeded the product of $\sigma(O-C)$ and the numerical parameter γ:

$$\left.\begin{array}{c} P \leq 2T \\ K \geq \gamma\,\sigma\,(\mathrm{O}-\mathrm{C}) \end{array}\right\}. \quad (2)$$

The parameter γ was determined empirically based on estimates of the $K/\sigma(O-C)$ ratio for known planets. In total, three values of γ were adopted depending on the planetary mass: $\gamma = 0.75$ for planets with masses below 0.14 Jupiter masses, $\gamma = 1.6$ for planets with masses between 0.14 and 1.7 Jupiter masses, and $\gamma = 2$ for planets with masses between 1.7 and 13 Jupiter masses.

This approach enabled the construction of a composite corrected minimum mass distribution of exoplanets, which is described by a piecewise function consisting of three power laws (Fig. 1).

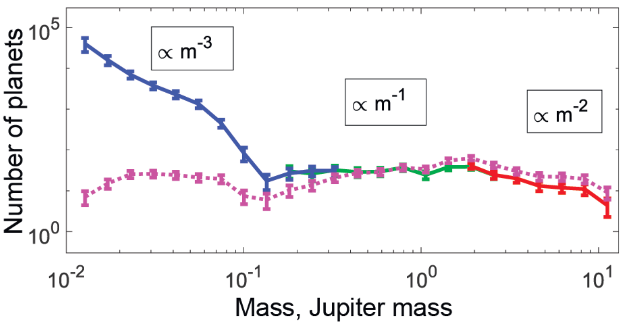

Fig. 1. Composite mass distribution of radial velocity detected planets corrected using the "detectability window" method. The dashed magenta line shows the uncorrected distribution.

The empirical approach did not explicitly account for the different numbers of radial velocity measurements of the host star, which is particularly important when searching for low-mass planets. To incorporate this important quantity, a new approach was developed.

2 Observational selection correction method accounting for the number of radial velocity measurements

The present work aims to study the distribution of low-mass planets; therefore, only planets with orbital periods shorter than 100 days were considered. For the low-mass planets with longer orbital periods, the radial velocity semi-amplitude K becomes too small, causing the corresponding elements of the "detectability window" matrix to become equal to zero, and the correction turns out to be impossible. The mass range from 0.0061 to 0.21 Jupiter masses (1.94–66.6 Earth masses) and the period range from 1 to 100 days were divided into 60 logarithmically equal intervals. A synthetic planet was placed at the center of each grid cell. For each such planet, the radial velocity semi-amplitude K was calculated and an RV signal was simulated assuming a circular orbit, which was then combined with Gaussian noise. A Lomb – Scargle periodogram was subsequently constructed. A planet was regarded as detectable if the corresponding periodogram peak had a significance exceeding 99 % (false alarm probability less than 1 %). Since the peak height may depend on the initial phase of the planet (its position along the orbit), the computation for each synthetic planet was performed for 24 different initial phases (with a 15° step), followed by averaging (Fig. 2).

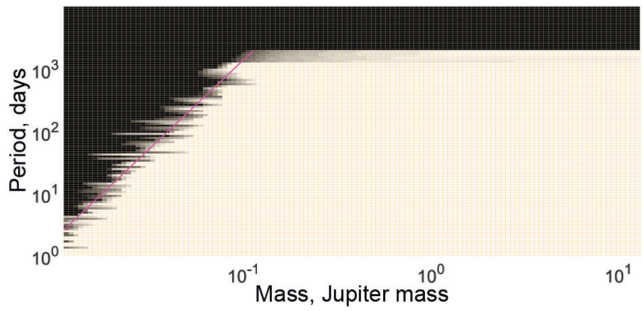

Fig. 2. Example of an individual detectability window for a solar-mass star with $N = 100$ radial velocity measurements, while the total observational time T is 1000 days. The RV signals of synthetic planets were combined with Gaussian noise with a standard deviation of $\sigma = 3$ m/s. The magenta line shows the power-law approximation of the detectability threshold.

The boundary between synthetic planets that can and cannot be detected is well approximated by a power law, $m \sim P^{1/3}$, which corresponds to $K = \text{const}$. It was shown that for each value of the number of radial velocity measurements N, there exists such a γ that for the planets with $K > \gamma \cdot \sigma$ they will be detected, while for those with $K < \gamma \cdot \sigma$ will not, i.e., there exists a characteristic threshold value $\gamma(N)$.

To determine the nature of dependence of $\gamma(N)$, calculations of γ were carried out for N = 50, 60, 75, 100, 150,

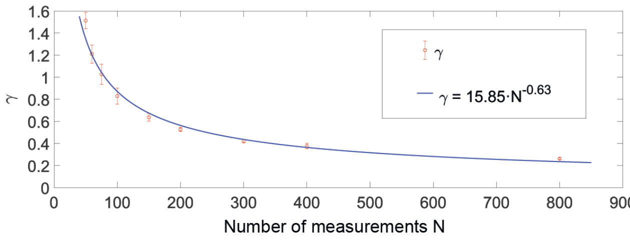

Fig. 3. The $\gamma(N)$ dependence approximated by a power law.

200, 300, 400, and 800 radial velocity measurements, with ten runs for each value of N. The resulting dependence was approximated by a power law, $\gamma(N) = 15.85 \cdot N^{-0.63}$ (Fig. 3).

3 Algorithm verification through numerical experiments

To verify the correct operation of the algorithm, $2 \cdot 10^4$ planets were randomly generated, with masses ranging from 0.0061 to 0.21 Jupiter masses and orbital periods from 1 to 100 days. The planetary mass distribution followed the power law $\sim m^{\alpha}$, where α took values of -1, -1.5, -2, -2.5, and -3, while the period distribution followed a power law with an exponent of -1 ($\sim P^{-1}$).

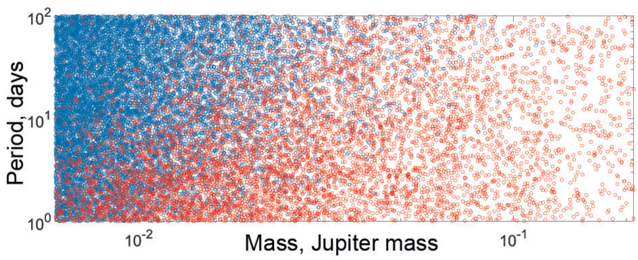

Fig. 4. Randomly generated population of planets on the minimum mass – orbital period diagram. The planets with $K < \gamma \cdot \sigma$ that are most likely undetectable using the Lomb – Scargle periodograms are indicated by blue color; the planets with $K > \gamma \cdot \sigma$ that are most likely detectable are indicated by red color.

Figure 4 shows an example of a generated planet population (in blue) whose mass distribution follows a power law with an exponent of -2.

To evaluate the detectability of these planets, 248 real RV detected planets with masses ranging from 0.0061 to 0.21 Jupiter masses were analyzed, taking into account the number of radial velocity measurements N and the mean deviation from the best-fit Keplerian curve $\sigma(O - C)$. Based on this sample, $2 \cdot 10^4$ sets of observational data with parameters close to the real sample were generated. Subsequently, the detectability of each generated planet was evaluated. A planet was considered detectable if the radial velocity semi-amplitude K it induces exceeded $\gamma(N) \cdot \sigma(O - C)$. The detected planets are shown in red in Fig. 4.

Figure 5 presents the corrected distribution of the planets shown in Fig. 4, obtained by applying the new algorithm to the distribution of "visible" planets.

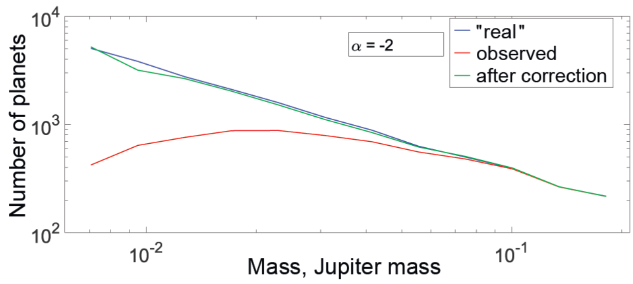

Fig. 5. The original ("real") mass distribution of planets is shown by the blue line. The red line shows the distribution of detectable planets, and the green line shows the distribution corrected using the new algorithm.

4 Conclusions

The proposed numerical algorithm for correcting observational selection, which accounts for the varying number of radial velocity measurements, can be applied to correct for the observational selection effects of real planets detected with the radial velocity technique.

References

Ananyeva V.I., Ivanova A.E., Shashkova I.A., et al., 2022. Astron. Rep., vol. 66, no. 10, pp. 886–917.

Ananyeva V., Ivanova A., Shashkova I., et al., 2023. Atmosphere, vol. 14, no. 2.

Butler R.P., Wright J.T., Marcy G.W., et al., 2006. Astrophys. J., vol. 646, no. 1, p. 505.

Howard A.W., Marcy G.W., Johnson J.A., et al., 2010. Science, vol. 330, no. 6004, pp. 653–655.

Ivanova A.E., Yakovlev O.Y., Ananyeva V.I., et al., 2021. Astron. Lett., vol. 47, no. 1, pp. 43–49.

Marcy G., Butler R.P., Fischer D., et al., 2005. Progress of Theoretical Physics Supplement, vol. 158, pp. 24–42.

Tuomi M., Jones H.R.A., Anglada-Escud'e G., et al., 2019. arXiv: Earth and Planetary Astrophysics.