A&A, 696, A94 (2025)
https://doi.org/10.1051/0004-6361/202452434
© The Authors 2025

Astronomy & Astrophysics

Earth-like planet predictor: A machine learning approach

Jeanne Davoult[1,2,3,⋆], Romain Eltschinger[2,3,4], and Yann Alibert[4]

[1] Institute of Planetary Research, German Aerospace Center (DLR), Rutherfordstrasse 2, 12489 Berlin, Germany
[2] NCCR PlanetS, Universität Bern, Gesellschaftstrasse 6, 3012 Bern, Switzerland
[3] Space research & Planetary Sciences (WP), Universität Bern, Gesellschaftsstrasse 6, 3012 Bern, Switzerland
[4] Center for Space and Habitability (CSH), Universität Bern, Gesellschaftstrasse 6, 3012 Bern, Switzerland

Received 30 September 2024 / Accepted 28 February 2025

ABSTRACT

Context. Searching for planets analogous to Earth in terms of mass and equilibrium temperature is currently the first step in the quest for habitable conditions outside our Solar System and, ultimately, the search for life in the universe. Future missions such as PLAnetary Transits and Oscillations of stars or Large Interferometer For Exoplanets will begin to detect and characterise these small, cold planets, dedicating significant observation time to them.

Aims. The aim of this work is to predict which stars are most likely to host an Earth-like planet (ELP) to avoid blind searches, minimises detection times, and thus maximises the number of detections.

Methods. Using a previous study on correlations between the presence of an ELP and the properties of its system, we trained a Random Forest to recognise and classify systems as 'hosting an ELP' or 'not hosting an ELP'. The Random Forest was trained and tested on populations of synthetic planetary systems derived from the Bern model, and then applied to real observed systems.

Results. The tests conducted on the machine learning (ML) model yield precision scores of up to 0.99, indicating that 99% of the systems identified by the model as having ELPs possess at least one. Among the few real observed systems that have been tested, eight have been selected as having a high probability of hosting an ELP, and a quick study of the stability of these systems confirms that the presence of an Earth-like planet within them would leave them stable.

Conclusions. The excellent results obtained from the tests conducted on the ML model demonstrate its ability to recognise the typical architectures of systems with or without ELPs within populations derived from the Bern model. If we assume that the Bern model adequately describes the architecture of real systems, then such a tool can prove indispensable in the search for Earth-like planets. A similar approach could be applied to other planetary system formation models to validate those predictions.

Key words. methods: data analysis – methods: statistical – planets and satellites: detection – planets and satellites: general – planets and satellites: terrestrial planets

1. Introduction

Detecting planets as small and cold as Earth is a major technical challenge in exoplanet research for the coming decades. The upcoming PLAnetary Transits and Oscillations of stars mission (PLATO; Rauer et al. 2014) and the concept of mission Large Interferometer For Exoplanets (LIFE; Kammerer & Quanz 2018; Quanz et al. 2022) will be dedicated to this task, but their long periods (potentially 1 year or more) consume significant observation time. Although various studies on planet demographics suggest that small terrestrial planets with short periods are very common around main sequence stars (e.g. Mayor et al. 2011; Tuomi et al. 2019; Kunimoto & Matthews 2020), the abundance of terrestrial planets with longer periods in the habitable zone of their star is more uncertain (e.g. Hsu et al. 2019; Bryson et al. 2021). Understanding and anticipating where Earth-like planets (ELPs in the rest of the paper) form first, and thus targeting observations to avoid blind searches, minimizes the average observation time for detecting an ELP and maximizes the number of detections. Studies conducted on the architecture and correlations in multi-planet systems over the years (e.g. Lissauer et al. 2011; Millholland et al. 2017; Weiss et al. 2018; Gilbert & Fabrycky 2020; Mishra et al. 2023; Emsenhuber et al. 2023; Davoult et al. 2024 among others) have highlighted correlations between the properties of planets in the same system. For example, correlations have been discovered between the presence of an inner terrestrial planet and the presence of an outer giant planet (e.g. Zhu & Wu 2018; Zhu 2024; Bryan & Lee 2024), but it exists an anti-correlation between the presence of a hot Jupiter and the 'peas-in-a-pod' formation (Weiss et al. 2018; Latham et al. 2011; Steffen et al. 2012). Thus, the architecture of systems, representing the arrangement of planets in a system, is not the result of chance but of simultaneous formation within the same system. In other words, the planets in the same system bear the imprint of each other's formation. Therefore, detected planets could provide insights into undetected planets within the same system.

Attempts to predict yet-undetected exoplanets based on detected exoplanets' properties have emerged in recent years. For example, Bovaird & Lineweaver (2013), Bovaird et al. (2015), Lara et al. (2020) and Mousavi-Sadr et al. (2021) have attempted to use a logarithmic relationship between planetary periods, akin to the Titius-Bode law, to predict missing planets within systems. Similarly, Dietrich & Apai (2020) and Sandford et al. (2021) utilised statistical data from already-detected planetary populations to forecast future observations. However, all these previous studies relied on data from observed exoplanet populations. Here, we propose using synthetic planetary systems from the Bern model – systems in which all planets are known – avoiding observational bias.

⋆ Corresponding author; jeanne.davoult@dlr.de

Open Access article, published by EDP Sciences, under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
This article is published in open access under the Subscribe to Open model. Subscribe to A&A to support open access publication.

In a previous study (Davoult et al. (2024), D24 in the rest of the paper), we have established correlations between the presence of an ELP in the temperate zone of its star and other properties of its system, including the architecture of the planetary system as described in the paper, and the mass, radius, and period of the innermost detectable planet (IDP in the rest of the paper) of that system. In this study, we present the results of algorithms using a machine learning (ML) model capable of learning the differences in properties between systems hosting an ELP and those not hosting an ELP, in order to predict whether a given system hosts an ELP or not.

The use of ML models requires very large datasets, which makes it impossible to use only data from observed systems. In addition to the small number of known planetary systems to date (just under 5000 in July 2024), there is the problem of partial knowledge of these systems. ELPs, being small and relatively cold planets, are difficult to detect using the most efficient detection methods (i.e., transits and radial velocities). Indeed, only 24 systems with at least one ELP are known (following the definition of Sect. 2.2), representing 0.5% of all systems observed to date. Herefore, using those data in a ML-based approach is impossible.

To address these two major problems, this study utilises populations of several thousand synthetic planetary systems generated from the Bern model. Studies have examined the outputs of this model and compared them to observed systems (e.g. Mulders et al. 2019; Schlecker et al. 2021a; Burn et al. 2021; Mishra et al. 2021, 2023; Davoult et al. 2024; Emsenhuber et al. 2025), revealing that these synthetic systems possess similar system-level characteristics as observed systems – such as similar architectures (Mishra et al. 2023; Davoult et al. 2024), recurring patterns in Peas-in-a-Pod (Mishra et al. 2021), correlations between outer giants and inner earths (Schlecker et al. 2021a), etc. These comparisons lead us to believe that synthetic systems generated from the Bern model serve as reasonable training data for ML models. Additionally, a study (Schlecker et al. 2021b) using a data-driven approach was also successfully conducted with the synthetic planetary system populations from the Bern model, aiming to predict the types of planets in a system based on the initial conditions of the protoplanetary disk and planetary embryos.

Section 2 briefly describes the Bern model and the populations used. Section 3 outlines the various ML models, observational biases, and system features used. In Section 4, we describe the results obtained for the different models, and we discuss and conclude in Section 5.

2. Synthetic population of planetary systems

2.1. The Bern model and synthetic populations

The planetary system formation and evolution model used in this study is the Generation III of the Bern model, described in detail in Emsenhuber et al. (2021a). This global model utilises the population synthesis method, as explained in detail in Mordasini (2018), and is based on the core accretion paradigm (Pollack et al. 1996). The planetary formation is modelled over 20 Myr, during which 20 planetary embryos embedded in a disk of gas and planetesimals accrete material to form planets, migrate, and dynamically interact, leading to ejections, giant impacts, or resonance traps. At the end of this formation phase, the model tracks the planets' thermodynamical evolution (consisting mainly of cooling and contraction) for 10 Gyr. During this evolution phase, atmospheric escape and tidal migration are also monitored. For more details on the parameterisation of the protoplanetary disk and the various physical processes involved in the formation and evolution of planets, refer to Emsenhuber et al. (2021a,b).

In a population synthesis, some parameters are fixed while others vary. In the populations of planetary systems used in this study, the fixed general parameters of the systems include the mass of the central star (1, 0.5, or 0.2 $M_\odot$), the number of planetary embryos (20), the gas viscosity ($\alpha = 2 \times 10^{-3}$), the distribution of gas and planetesimals in the protoplanetary disk (Veras & Armitage 2004), the size of the planetesimals (radius = 300 m), and their density (rocky 3.2 g cm^{-3}, icy 1g cm^{-3}). The rest of the initial conditions are randomly drawn according to a probability distribution constrained by observations, which allows for diversity in the resulting synthetic planetary systems. The variable parameters include the initial mass of the gas disk, M_g (Beckwith & Sargent 1996), the external photo-evaporation rate $M_{\rm wind}$ (Haisch et al. 2001), the dust-to-gas ratio, $f_{\rm D/G}=M_s/M_g$ (where M_s is the mass of the solid disk) (Murray et al. 2001; Santos et al. 2003), the inner edge of the gas disc, $R_{\rm in}$, and the initial location of the embryos.

The three populations of synthetic systems used in this study differ only in the mass of the central star. This single difference directly influences the mass of the protoplanetary disk and thus the amount of material available for planet formation. As a result, the three populations exhibit different occurrences and properties for the same type of planet, highlighting the importance of studying various types of stars.

The three populations used are:

- G-pop: 24 365 systems around solar mass stars
- earlyM-pop: 14 559 systems around 0.5 solar mass stars
- lateM-pop: 14 958 systems around 0.2 solar mass stars.

For a detailed analysis of the different types of planets and their occurrences in the above populations, refer to D24.

2.2. Earth-like Planet

This study aims to predict which systems host an Earth-like planet or not. The ELP category refers to a small terrestrial planet with a mass ranging from 0.5 to 3 $M_\oplus$, orbiting the temperate zone of its star. The mass range was chosen in accordance with the work of Kopparapu et al. (2018) and Burn et al. (2021). The temperate zone, defined in Davoult et al. (2024), is defined much broader as the habitable zone and extends in terms of equilibrium temperature ($T_{\rm eq}$) from 160 to 510 K, calculated as follows:

$$T_{\rm eq}[{\rm K}] = 279 \cdot a[{\rm AU}]^{-1/2} \cdot L_\star[{\rm L}_\odot]^{1/4}, \quad (1)$$

where a is the semi-major axis of the planet and $L_\star$ is the luminosity of the star. This correspond to a zone between 0.39 and 3.9 AU around a G-type star, between 0.25 and 2.52 around a early-M type star and between 0.15 and 1.48 around a late-M type star. By extending the target zone, we increase the number of systems with an ELP, and we reduce the imbalance in terms of proportion in the data, which is beneficial for ML models.

As seen in D24, the occurrence of a certain type of planet varies depending on the type of star it orbits. Thus, in our three populations, we find 60% of systems with an ELP around solar-mass stars, 74% around stars of 0.5 $M_\odot$, and 40% around stars of 0.2 $M_\odot$.

2.3. Correlations between ELP and the properties of their systems

In D24, we investigate correlations between planets in the synthetic planetary systems from the Bern model and their

architecture to define a typical profile of a system hosting an ELP. Our conclusions highlight a correlation between the presence of an ELP, the architecture of its system, and the properties of the innermost detectable planet (IDP). Indeed, Earth-like planets tend to form in systems mainly composed of low-mass planets (M < 20 $M_\oplus$). In systems with more massive planets, the properties of the IDP, such as mass, radius, and period, can be indicators of ELP presence. A small, low-mass IDP suggests in-situ formation in a low-mass disk, while a giant IDP suggest a massive disk and/or planetary migration, unfavorable for a stable Earth-like planet in the habitable zone. The IDP's period indicates the positions of other planets: a close-in IDP suggests inward planet grouping, leaving the HZ empty, while slightly longer periods (>tens of days) indicate outward grouping, increasing HZ planet probability. Thus, ELP presence correlates with the system's architecture, and IDP's mass, radius, and period.

Table 7 of D24 summarises the conditional probabilities of ELP's presence in systems according to the mass of the central star, the observed system architecture, and the properties of the IDP, providing an overview of the combinations most favourable for ELP formation in a system. The present paper uses part of their results to develop a predictor incorporating a ML model that the community can use to predict whether a system is likely to host an ELP or not based on its observable properties. We relied on the work presented in D24 to define the observable properties used in this prediction.

3. Method

3.1. Machine learning classifier (MLC)

In ML methodologies, algorithms typically perform two main tasks: classification and regression. This problem is a case of classification, aiming to classify a system into the categories of 'hosts an ELP' or 'does not host an ELP'. ML models are trained to recognise data falling into one category or another using a dataset of thousands of data points. Once trained, they can predict, on an independent dataset, which class an instance falls into.

There are many classifier tools available, with the most common being decision trees, support vector machines (SVMs), or Random Forests, among others. Random Forests fall into the category of 'ensemble' learning methods. They consist of multiple sub-classifiers, with each 'local' classifier trained on a subset of data (which is not the entire training set). Then, all local classifiers are queried to classify an element. The 'global' classifier (which includes all local classifiers) decides based on the majority of vote: if a majority of local classifiers vote to classify an element into the 'True' category, then the final response of the global classifier is 'True'. In this case, the category 'hosts an ELP' is True, and the category 'does not host an ELP' is False.

We aim to predict whether a system hosts an ELP or not to target observations to avoid wasting observation time. Therefore, we want to ensure that the positive responses given by our algorithm can be trusted, meaning it produces very few 'false positives'. To ensure this, we want to maximise the precision score (PS), which measures the ratio of 'true positives' to all elements labelled as 'positive' (true positives and false positives). The precision score is the ability of the classifier not to label a negative sample as positive:

$$\mathrm{PS} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}. \tag{2}$$

When the precision score increases, the recall score (RC) decreases. The recall score is the ability of the classifier to find all the positive samples and is computed as follows:

$$\mathrm{RC} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}. \tag{3}$$

In other words, the more we focus on elements most likely to be labelled 'True', the more we miss true positives in the batch. This is not necessarily a major issue because we do not particularly want to maximise the RC. The False Negative rate is characterised by the RC. The lower the RC, the higher the number of false negatives. Given the time required to detect an Earth-like planet, we chose to focus on maximising the PS rather than the RC. It seems more important to ensure the method's reliability by concentrating on the most robust systems. Indeed, there exist many potential targets for a limiting telescope time. In the opposite situation (plenty of telescope time, but few targets), we would like to optimise the recall score (minimising the number of false negatives). The issue arises when the classifier fails to find any positives in our efforts to maximize true positives.

The significant advantage of ensemble algorithms is adjusting the voting threshold ourselves. Instead of declaring the threshold at 50% of local classifiers as the threshold for the global classifier to decide, we can adjust this threshold. In the rest of the study, we examine several thresholds: 70, 80, and 90%. We define 'voting rate' as the proportion of local classifiers that have classified an instance as True.

The Random Forest used in this study is made up of 500 decision trees, allowing to reduce the variance through ensemble learning, while keeping the training time reasonable. Each Decision Tree is trained on a minimum sample of 100 instances in order to increase the diversity between the trees, while allowing the generalisation of the classification. Trees trained on fewer instances have a tendency to learn details and overfit. Finally, the maximum depth of each Tree is limited to five in order to limit the complexity of the model, forcing it to capture only the most important relationships in the data. Tests conducted with a higher maximum depth (no limitation) did not change the results very much, proving the stability of the results.

3.2. Observational bias

To use the observable properties of synthetic planetary systems from the Bern model, we apply an observational bias to retain only the planets that could be theoretically observed. This bias involves a radial velocity (RV) semi-amplitude threshold on the star. Planets with an RV semi-amplitude above this threshold are considered detected, while those below are considered undetected. The detected planets form the new planetary system from which we extract the characteristics used by the ML model. The RV semi-amplitude that a planet induces on its star is calculated as follows:

$$K_{\mathrm{RV}}[\mathrm{ms}^{-1}] = 0.6395 \cdot P[\mathrm{days}]^{1/3} \cdot M_p[\mathrm{M}_\oplus] \cdot M_\star[\mathrm{M}_\odot]^{-2/3}, \tag{4}$$

where P is the period of the planet, M_p its mass and $M_\star$ the mass of the star.

The detection threshold is set to exclude ELPs from this study. Ignoring systems with detected ELPs is reasonable, given that only 24 systems (0.5%) among nearly 4900 observed (as of July 2024) are known to host a planet following our definition (see Sect. 2.2). The RV threshold for detectability varies between populations due to two factors: the limits of the temperate zone

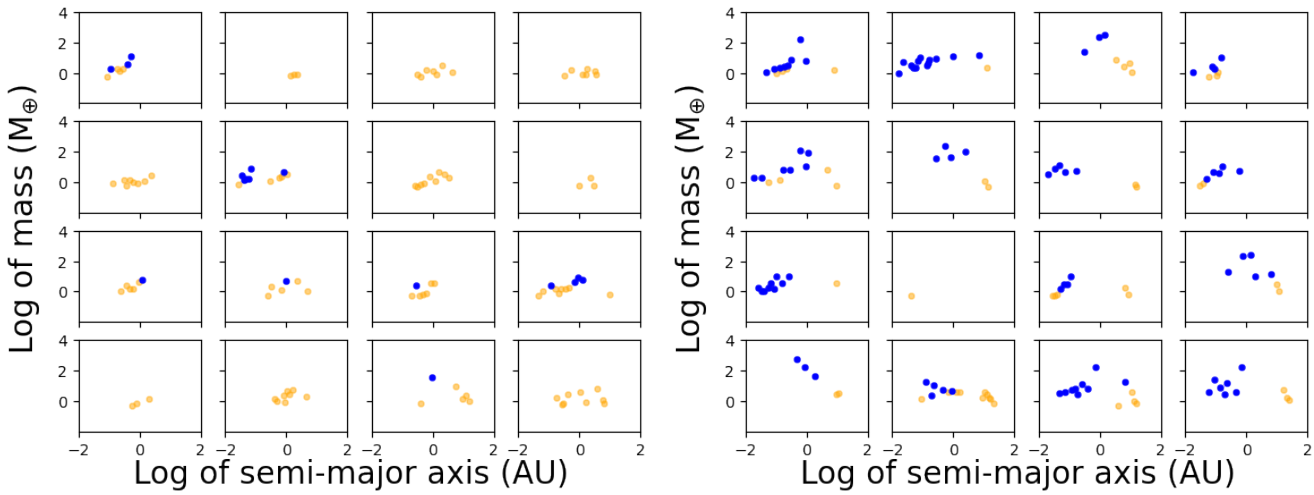

Fig. 1. Representation of 16 systems with ELP (left) and 16 systems without ELP (right) in a semi-major axis – planetary mass diagram (in log scale for both axes). Blue dots represent 'detectable' planets and yellow dots 'undetectable' planets.

Table 1. RV semi-amplitude thresholds retained for each population.

Central star's mass	Threshold retained
1 $M_\odot$ (G-pop)	0.43 ms^{-1}
0.5 $M_\odot$ (earlyM-pop)	0.76 ms^{-1}
0.2 $M_\odot$ (lateM-pop)	1.55 ms^{-1}

vary depending the population ($T_{\rm eff} \propto L_\star^{1/4}$) and for a given planetary mass and period, the RV semi-amplitude signal varies as $M_\star^{-2/3}$. The values used are presented in Table 1.

Although this observation bias is too simple to be considered accurate, D24 have shown that it can reproduce the proportions in architectures observed in multiplanet systems, which is sufficient for this study. An analysis of the impact of this bias on the synthetic populations of planetary systems used here is available in D24.

3.3. Features of interest

When using a ML model, it projects the dataset into an N-dimensional space, where N is the number of dimensions of the dataset. In our case, N represents the amount of information about each system provided to the algorithm for learning. Each of the three populations contains between ~15 000 and ~25 000 systems. After removing empty systems (systems with no planets) and systems with no visible planets (systems with planets but that cannot be classified in our architecture classes), only about 5000 to 20 000 instances remain in each population. While the size of this dataset allows us to conduct this study, it remains limited. If N is too large, the data may become lost in a high-dimensional space, making the task challenging for the model and increasing the risk of overfitting. Therefore, it is important to describe each instance – each system in this case – with a reasonable number of features to mitigate the risk of overfitting. The challenge lies in selecting the right features, the most useful ones that provide the most information. Given that the aim of this project is observational, the information provided to the ML model must be easily observable quantities. We present two strategies: the first strategy utilises the findings of D24, while the second strategy involves defining the features based on a manual analysis.

3.3.1. Observables derived from D24

In D24 we present a study of correlations between the presence of an ELP in a system and observable quantities of those systems. The conclusions link the presence of an ELP with a system's 'biased' architecture, as well as the mass, radius, and period of the innermost detectable planet (IDP). The biased architecture of a system refers to the architecture of a system considering only the detectable planets in that system. The method used to calculate the observational bias is the same in this article as in D24, ensuring a similar approach.

In D24, we also introduce a method for classifying each system into a different architecture class using Principal Component Analysis (PCA) applied in the mass – semi-major axis plane of the visible planets in the system, along with the mass of the most massive visible planet in the system and the number of visible planets. Thus we define five classes:

- Low-mass: systems with at least two visible planets in which all planets are less massive than 20 $M_\oplus$
- Anti-Ordered: systems with at least two visible planets, with at least one planet more massive than 20 $M_\oplus$ and a general tendency for the planetary masses to decrease with the distance to star increasing.
- Ordered: systems with at least two visible planets, with at least one planet more massive than 20 $M_\oplus$ and a general tendency for the planetary masses to increase with the distance to the star increasing.
- Mixed: systems with at least two visible planet and a planet more massive than 20 $M_\oplus$, and a large variability in the planetary masses, inducing no special tendency.
- n = 1: systems with only one visible planet

These four descriptive features make up the first set: the architecture of the visible system, and the IDP's mass, radius, and period.

3.3.2. Manual feature selection

Looking at the systems generated from the Bern model, it is evident that systems with ELPs are very similar to each other, whereas, conversely, they are very different from systems without ELPs. Figure 1 depicts two types of systems: on the left are sixteen systems with at least one ELP randomly selected from

the Sun-like stars population, and on the right are sixteen systems without ELPs randomly selected from the same population. The blue dots represent planets that have passed the detection threshold, while the yellow dots represent 'non-detectable' planets. Systems without ELPs (left) are all very similar to each other. They consist mostly of small planets, with relatively few detectable planets, and few planets more massive than Neptune. Additionally, compact, clustered systems are observed around one AU. In contrast, systems without ELPs, on the right, display more detectable planets, including more massive planets. It is common to find a giant or at least a sub-giant planet in these systems. The systems are more spread out in terms of semi-major axis range, but we can still find clusters of small terrestrial planets, which are shifted inward, very close to the star, at a fraction of an AU. These visible differences allow us to easily classify a system as 'host an ELP' or 'does not host an ELP'. To use these features in an ML model, we need to quantify them, describing each system with a limited number of features.

Our choice of features, which we believe best capture the differences observed visually, is as follows:

- number of visible planets
- number of giant planets ($M_p > 100\ M_\oplus$)
- IDP's mass,

to which we add the star's mass, known to be correlated with the type of planets present in the system. Indeed, as studied in Section 2.2, the proportion of systems with ELPs is not the same in the three populations because the central star's mass plays a role in planetary formation. These five features make up our second set.

3.4. Train and test dataset

For this study, we utilise three different populations of synthetic planetary systems. Initially, we train our algorithm on a 'training set', where each system is labelled as 'True' (host an ELP) or 'False' (does not host an ELP). Thus, the algorithm learns to recognise which systems host an ELP and which are ELP-free. This training set comprises the majority (80%) of our synthetic systems. Once the algorithm is trained, we can test it on a 'test set', which is an unlabeled dataset on which the model makes predictions to analyse its responses and determine its precision and recall scores. The test set consists of the remaining 20% of the dataset to ensure that the systems on which we test the algorithm are not the same as those on which it was trained, which would bias the results.

It is also important to ensure that the different proportions are respected in both datasets. For example, if the test set comprises 80% of systems with an ELP while the overall proportion in the population is 40%, the test is biased.

To ensure a consistent training and test set, we divided the systems with an ELP from the systems without an ELP (¬ELP) in each distinct population, resulting in six subgroups (1 $M_\odot$/ELP, 1 $M_\odot$/¬ELP, 0.5 $M_\odot$/ELP, 0.5 $M_\odot$/¬ELP, 0.2 $M_\odot$/ELP, and 0.2 $M_\odot$/¬ELP, where ELP means the systems with at least one ELP and ¬ELP means the systems without any ELPs). Then, 80% of each subgroup constitutes the training set, and the remaining 20% forms the test set. When creating training and test sets with the three mixed populations, we ensure that the proportion of each population remains the same. Thus, we choose the population with the fewest systems and adjust the other populations to match this number. This way, we ensure that we have the same proportion of systems from each population (1, 0.5, and 0.2 $M_\odot$) in both datasets. On the other hand, we do not scale the number of systems with and without ELP. The proportion of systems with ELP in each population is a feature in itself that the model must account for.

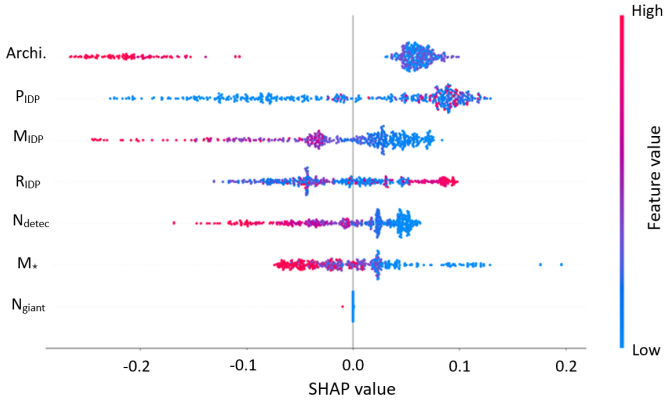

Fig. 2. Bee swarm plot of the seven features considered. The x-axis represents the SHAP value of the feature for each instance, and the y-axis represents the seven features considered ranked from the most important (top) to the least (bottom). The colour of the dots represents the value of the feature itself, red being high values and blue being low values.

Once trained on the training set, we test the algorithm on the test set and calculate its performance using the different scores. We then apply it to a list of observed planetary systems to predict whether a system is likely to host an ELP or not. This likelihood is characterised by the algorithm's voting rate.

4. Results

To optimise the classification model, we first conduct several tests on the training data and the systems' descriptive features, described in the following paragraphs. Once the best strategy is identified, we use the model trained on a sample of 1567 observed systems to predict the presence of an ELP.

4.1. Features analysis

As discussed in Sections 3.3.1 and 3.3.2, we have selected seven potentially useful descriptive features for this study. As mentioned in Section 3.3, we need to identify the ones that provide the most information about the presence of an Earth-like planet to maximise the performance of the Random Forest model.

To select the most useful features, we conduct a feature analysis. We apply the Shapley value concept to assess the features' importance of all the features described in Sect. 3.3. Originating from cooperative game theory, Shapley values are frequently used in machine learning to analyse the importance of features. They represent each feature's contribution to the model's prediction by evaluating all possible feature combinations and measuring the impact of adding or removing each feature.

Fig. 2 presents a bee swarm plot where each point represents the SHAP (SHapley Additive exPlanations) (Lundberg & Lee 2017) value of a specific feature for an individual instance in the dataset. This visualisation shows how each feature's contribution affects the model's prediction for that instance. The y-axis lists the features from most influential (top) to least influential (bottom), while the x-axis shows the SHAP value of each feature for each dataset instance. Negative SHAP values indicate a stronger contribution to the decision 'without ELP', while positive values indicate a stronger contribution to the decision 'with ELP'. Additionally, the colour of the points represents the feature value itself, with higher values in red and lower values in blue. In

this diagram, we have removed systems with no visible planets to facilitate readability. Indeed, when no planets are visible in a system, the IDP's mass, radius, and period are set to −1000 to indicate the absence of values for these features. This procedure results in the final bee swarm plot being polluted by very low values, making it difficult to interpret the values of the different features.

From Fig. 2, we observe that architecture emerges as the most important feature, with lower values indicating a greater likelihood of containing an ELP. We assigned values from 1 to 5 to the architectures (n = 1: 1, Low-mass: 2, Anti-Ordered: 3, Ordered: 4, Mixed: 5). A low value for the architecture indicates either n = 1 or Low-mass, which are the dominant classes hosting an ELP.

The period and mass of the innermost detectable planet (IDP) also play a significant role. Systems where the IDP has a greater distance from the star are more likely to be classified as 'with ELP', consistent with the findings of D24. Conversely, systems with a less massive IDP are also more likely to be classified as 'with ELP', further supporting D24's results.

The impact of the IDP's radius is nuanced, as observed in D24: for a given architecture, either a larger or smaller IDP radius can be more favourable for detecting an ELP. This makes it a more difficult characteristic to use, because there is no clear cut.

The number of visible planets, although lower ranked, also provides valuable information: the more visible planets there are, the less likely the system is to host an ELP, which confirms the observations discussed in Section 3.3.2.

However, the influence of the central star is not consistent with the first analysis, which showed that systems around stars of 0.2 $M_\odot$ had proportionately fewer ELPs (only 40% of systems, compared with 75 and 60% for 0.5 and 1 $M_\odot$ respectively). In this representation, we have removed systems without planets larger than 0.5 $M_\oplus$ (see Appendix A of D24). However, the vast majority of these empty systems are systems without any ELP, which reverses the proportion of systems with ELP if they are not considered. In this representation, we therefore have systems with a low-mass star (blue represents stars of 0.2 $M_\odot$) classified as having an ELP, systems with a star of 1 $M_\odot$ (red) classified as having no ELP, and systems with a star of 0.5 $M_\odot$, being hard to classify because they have almost equal numbers of systems with and without ELP.

Finally, the number of giant planets in the system provides limited information. Specifically, as the number of giant planets increases, the model tends to classify the system as ELP-free. However, in the absence of giant planets, the model has difficulty making a clear decision. This mirrors the data observed: systems with giant planets are much less likely to have an ELP, while those without giant planets may or may not host an ELP.

Based on Fig. 2, it appears that the most important features are the architecture, and the mass and period of the innermost detectable planet (IDP).

To compare the performance of the Random Forest Classifier based on the descriptive features used, we conducted four tests, each time changing the descriptive features. The first test includes all features, the second includes only manually selected features (Sect. 3.3.2), the third includes features derived from D24 (Sect. 3.3.1), and the fourth includes the top features selected from Fig. 2. Those four tests are resumed in Table 2.

Table 3 displays the Random Forest results for voting rate thresholds of 50% (default), 70, 80, and 90%. For each test and each threshold, the table shows the confusion matrix and the precision score (PS). As a reminder, a confusion matrix is constructed as $\begin{pmatrix} TN & FP \\ FN & TP \end{pmatrix}$, with TN and FN representing True Negatives and False Negatives, and TP and FP representing True Positives and False Positives. The confusion matrix is beneficial in unbalanced datasets like this one. It allows us to assess not only whether the model correctly classifies the instances but also its performance (very few false positives and true positives indicate that the model struggles to understand what a system with an ELP looks like).

Table 2. List of planetary systems features used in each test.

N° test	Features included
Test n°1	All
Test n°2	$N_{\rm detec}$, $N_{\rm giant}$, $M_{\rm IDP}$, $M_\star$
Test n°3	$R_{\rm IDP}$, $M_{\rm IDP}$, $P_{\rm IDP}$, architecture, $M_\star$
Test n°4	$M_{\rm IDP}$, $P_{\rm IDP}$, architecture, $M_\star$

Unsurprisingly, for the default threshold of 50%, the results are fair but not excellent. True positive answers account for just above 80% of all positive answers. As the threshold increases, the precision score improves, indicating the model's ability to recognise patterns that distinguish systems with ELPs. Increasing the precision score also means increasing the TP/FP ratio. However, in these cases, we also notice an increase in false negatives (FN), indicating that the model becomes more conservative, missing more positives in its effort to reduce false positives. From a threshold of 80%, all three tests show precision scores above 0.9, indicating that true positives account for 90% of the model's positive predictions, demonstrating its excellent capability.

Although the four tests show similar overall performance, a closer look into the confusion matrix reveals that Test n°1 is less effective than the other three tests. Specifically, above 90%, Test n°1 shows fewer TP and FP, resulting in fewer overall positive answers. While the ratio between TP and FP remains similar, the lower number of total positive answers indicates that it recognises systems with an ELP less effectively. Tests n°2 exhibits a slightly lower PS for all thresholds above 80%. Test n°3 and Test n°4 show same precision score for thresholds above 80% but Test n°4 exhibit slightly fewer False Negative and more True Positive, this reflects its ability to recognise a system with an ELP more effectively. For this reason, Test N°4 is used in the remainder of the study.

4.2. Population analysis

Now that we have determined the features to use, we need to decide on which populations of synthetic systems the model should be trained to achieve optimal performance. Several strategies are considered:

- Mass-Specific Training: to predict the outcome of a system, we use a model trained exclusively on systems with similar central star's mass. The star's mass is not a feature provided to the model but is considered when choosing which training data to use.
- Global Population Training: we train the model on a combined population that includes systems with central stars of different masses, regardless the central star's mass. Here, the star's mass is an input to the model so that it can

Table 3. Performance results of the model trained on four different tests. The Test n°1 uses all the features.

	Test n°1	Test n°2	Test n°3	Test n°4
50%	$\begin{pmatrix} 1035 & 285 \\ 148 & 1451 \end{pmatrix}$ PS=0.84	$\begin{pmatrix} 1027 & 331 \\ 131 & 1430 \end{pmatrix}$ PS=0.81	$\begin{pmatrix} 990 & 308 \\ 172 & 1449 \end{pmatrix}$ PS=0.83	$\begin{pmatrix} 1061 & 287 \\ 144 & 1427 \end{pmatrix}$ PS=0.83
70%	$\begin{pmatrix} 1095 & 225 \\ 208 & 1391 \end{pmatrix}$ PS=0.86	$\begin{pmatrix} 1100 & 258 \\ 215 & 1346 \end{pmatrix}$ PS=0.84	$\begin{pmatrix} 1063 & 285 \\ 226 & 1395 \end{pmatrix}$ PS=0.86	$\begin{pmatrix} 1106 & 242 \\ 194 & 1377 \end{pmatrix}$ PS=0.85
80%	$\begin{pmatrix} 1309 & 11 \\ 859 & 740 \end{pmatrix}$ PS=0.99	$\begin{pmatrix} 1337 & 21 \\ 811 & 750 \end{pmatrix}$ PS=0.97	$\begin{pmatrix} 1281 & 17 \\ 860 & 760 \end{pmatrix}$ PS=0.98	$\begin{pmatrix} 1328 & 20 \\ 788 & 783 \end{pmatrix}$ PS=0.98
90%	$\begin{pmatrix} 1316 & 4 \\ 1423 & 176 \end{pmatrix}$ PS=0.98	$\begin{pmatrix} 1345 & 13 \\ 973 & 588 \end{pmatrix}$ PS=0.98	$\begin{pmatrix} 1290 & 8 \\ 942 & 679 \end{pmatrix}$ PS=0.99	$\begin{pmatrix} 1343 & 5 \\ 861 & 710 \end{pmatrix}$ PS=0.99

Notes. Test n°2 uses the number of detectable planets, number of giant planets, mass of the innermost detectable planet and central star's mass. Test n°3 uses the radius, mass and period of the innermost detectable planet, the architecture of the system and the mass of the central star. Test n°4 corresponds to the most important features: the mass and period of the innermost detectable planet, the architecture of the system and the mass of the central star. The different percentages correspond to different voting rate thresholds of the Random Forest.

Table 4. Different training population used in the tests.

Population name	Synthetic population included
MS-1	1 $M_\odot$
MS-0.5	0.5 $M_\odot$
MS-0.2	0.2 $M_\odot$
Subset	1 $M_\odot$ + 0.5 $M_\odot$
Global	1 $M_\odot$ + 0.5 $M_\odot$ + 0.2 $M_\odot$

differentiate between different types of systems mixed in the overall training population.

– Subset Training: we create subsets of training data. For example, we train the model on populations of systems with 1 and 0.5 $M_\odot$ stars together because they share similarities, particularly in terms of the proportion of systems with an ELP (60% for solar-mass stars and 74% for 0.5 $M_\odot$ stars, compared to only 40% for 0.2 $M_\odot$ stars), and separately on the population with 0.2 $M_\odot$ stars.

We construct several training populations to evaluate which strategy is best. These populations are summarised in Table 4. We then train the Random Forest, and test it for each strategy. The results are shown in Table 5.

A quick glance shows that the last two populations (Subset and Global) yield very similar results. The models trained on the three populations built for the Mass-Specific strategy (MS-1, MS-0.5 and MS-0.2) show however different results: the model trained on MS-1 and MS-0.5 have better results than on MS-0.2. This can be explained both by the fact that the population of 0.2 $M_\odot$ stars is unbalanced negatively (only 40%) of systems host an ELP) but also because there are a lot less systems in this population than in the two others. After correction for empty systems, the MS-0.2 has 4862 systems, MS-0.5 has 10 158 systems and MS-1 has 20 365 systems. Although above a threshold of 90%, all populations yield the same result (PS = 0.99) except for MS-0.2 (PS = 0.94), we chose the mass-specific strategy. This strategy allows for the maximum use of training data and helps avoid overfitting. For the subset training and global training strategies, populations are scaled to have the same number of systems. In other words, systems are randomly removed from the larger populations to match the number of data points in the smallest population.

4.3. Prediction of detection

The developed and trained model can now be used to predict which systems are most likely to host an ELP. We use a sample of 1567 known systems around MKG stars from `exoplanet.eu`[1] (Schneider et al. 2011) in which at least one planet and its mass is known, regardless its detection method.

The dataset is then divided into three subsets: 1025 systems with central star masses between 0.7 and 1.2 $M_\odot$, 342 systems with central star masses between 0.35 and 0.7 $M_\odot$, and 200 systems with central star masses less than 0.35 $M_\odot$. Each subset corresponds to a specific training dataset: 1, 0.5, and 0.2 $M_\odot$, respectively. We apply the same observational bias to each subgroup as described in Section 3.2, according to the mass of the central star. For each system, we extract the corresponding features for Test n°4: the system's architecture with planets that overpass the observational bias, the mass, and the period of the IDP, if one planet remained in the system after applying the bias. We then use the model trained on the populations corresponding to each subset to obtain the voting rate of each planetary system. Among the 1567 total systems in the three subsets, 51 achieved a voting rate of more than 90%. We exclude binary systems because the Bern model produced only single stars and the habitable zone is calculated differently in binaries (Haghighipour 2015), and the 44 remaining systems with their associated voting rates are listed in Table 6.

To evaluate the possibility of a planet's existence in these systems, we use the stability criterion from Fabrycky et al. (2014), which was also employed in Chen et al. (2024). The Hill-stability criterion H is defined as:

$$H = \frac{a_{\rm out} - a_{\rm in}}{R_{\rm H}}, \tag{5}$$

with $a_{\rm in}$ and $a_{\rm out}$ referring to the semi-major axes of the inner and outer planets, respectively, and $R_{\rm H}$ being the mutual Hill radius relevant for dynamical interactions (Fabrycky et al. 2014):

$$R_{\rm H} = \left(\frac{M_{\rm in} + M_{\rm out}}{M_\star}\right)^{1/3} \frac{(a_{\rm in} + a_{\rm out})}{2}, \tag{6}$$

with $M_{\rm in}$ and $M_{\rm out}$ being the masses of the inner and outer planets, respectively, and $M_\star$ being the mass of the star. For a

[1] Available at `https://exoplanet.eu/catalog/`

Table 5. Performance results of the model trained on different populations.

	50%	70%	80%	90%
MS-1	$\begin{pmatrix} 824 & 205 \\ 238 & 2807 \end{pmatrix}$ PS=0.93	$\begin{pmatrix} 945 & 84 \\ 361 & 2684 \end{pmatrix}$ PS=0.97	$\begin{pmatrix} 967 & 62 \\ 406 & 2639 \end{pmatrix}$ PS=0.98	$\begin{pmatrix} 1011 & 18 \\ 544 & 2501 \end{pmatrix}$ PS=0.99
MS-0.5	$\begin{pmatrix} 23 & 466 \\ 28 & 1516 \end{pmatrix}$ PS=0.77	$\begin{pmatrix} 43 & 446 \\ 69 & 1475 \end{pmatrix}$ PS=0.77	$\begin{pmatrix} 482 & 7 \\ 1318 & 226 \end{pmatrix}$ PS=0.97	$\begin{pmatrix} 487 & 2 \\ 1364 & 180 \end{pmatrix}$ PS=0.99
MS-0.2	$\begin{pmatrix} 837 & 3 \\ 91 & 42 \end{pmatrix}$ PS=0.93	$\begin{pmatrix} 838 & 2 \\ 93 & 40 \end{pmatrix}$ PS=0.95	$\begin{pmatrix} 838 & 2 \\ 95 & 38 \end{pmatrix}$ PS=0.95	$\begin{pmatrix} 839 & 1 \\ 117 & 16 \end{pmatrix}$ PS=0.94
Subset	$\begin{pmatrix} 437 & 566 \\ 140 & 2923 \end{pmatrix}$ PS=0.84	$\begin{pmatrix} 502 & 501 \\ 209 & 2854 \end{pmatrix}$ PS=0.85	$\begin{pmatrix} 964 & 39 \\ 1556 & 1507 \end{pmatrix}$ PS=0.98	$\begin{pmatrix} 982 & 21 \\ 1672 & 1391 \end{pmatrix}$ PS=99
Global	$\begin{pmatrix} 1061 & 287 \\ 144 & 1427 \end{pmatrix}$ PS=0.83	$\begin{pmatrix} 1106 & 242 \\ 194 & 1377 \end{pmatrix}$ PS=0.85	$\begin{pmatrix} 1328 & 20 \\ 788 & 783 \end{pmatrix}$ PS=0.98	$\begin{pmatrix} 1341 & 5 \\ 861 & 710 \end{pmatrix}$ PS=99

Notes. MS-1, MS-0.5 and MS-0.2 are used in the Mass-Specific training strategy and correspond to individual populations. Subset corresponds to the population of 0.5 and 1 $M_\odot$ stars combined for the Subset strategy. Global population is the three populations combined for the global strategy.

two-planet system, Chen et al. (2024) defines $H > 7.1$ for the system to be stable. In a system with more than two planets, a more stringent criterion is used:

$$H_{\rm in} + H_{\rm out} > 18, \tag{7}$$

with $H_{\rm in}$ and $H_{\rm out}$ being the Hill-stability criteria for the inner and outer planet pairs.

Figures 3, 4 and 5 present the systems identified as potential candidates for hosting an ELP in a mass-semi-major axis diagram around G stars, early-M and late-K stars, and late-M stars respectively. The dots indicate the already existing planets: black dots represent visible planets for which we know the mass and are the planets used to assess the voting rates of a system. Grey dots represent planets for which we know the mass but with a RV semi-amplitude lower than the detection bias (they do not contribute to the determination of the architecture). Finally, orange points represent planets for which we only know the radius, and the mass has been derived using the mass-radius relationship from Parc et al. (2024). The latters are only used to evaluate the stability of a system with an additional planet, but they are not used in the model. The green areas outline the regions defining an ELP in terms of mass and equilibrium temperature. The grey areas correspond to regions where the Hill-stability criterion is met and where the presence of an additional planet is possible. The green and grey areas overlap in most of systems identified by our algorithm, indicating the potential for an ELP in these systems. Particularly, for G stars, only the system HIP 41378 does not seem stable with the addition of an ELP. However, if we knew the precise mass of the four planets represented by an orange dot, the model would not have classified it in the category 'with ELP'. For late-K and early-M stars, all of the systems seem stable while adding an ELP. Finally, for late-M stars, only GJ 273 does not seem stable while adding an ELP. These results highlight the effectiveness of our model: 95.5% of the systems identified as likely to host an ELP can theoretically host one.

5. Discussion and conclusion

5.1. Discussion

The model presented in this work presents a few limitations and avenues to improvement that we would like to discuss. Firstly, using this model on a sample of known and observed planetary systems involves assuming that the Bern model accurately replicates observed planetary systems and that the properties correlated with the presence of an Earth-like planet in synthetic systems are the same in real planetary systems. In reality, the synthetic systems modelled with the Bern model only partially resemble actual planetary systems. Other studies (Mulders et al. 2019; Schlecker et al. 2021a; Burn et al. 2021; Mishra et al. 2021, 2023; Burn et al. 2024; Emsenhuber et al. 2025) have shown that populations of planetary systems replicate the basic patterns observed in actual planetary populations. The populations of synthetic systems calculated using the Bern model demonstrate a positive correlation between the occurrence of inner Super-Earths and cold Giants (Schlecker et al. 2021a), albeit weaker than those observed in studies such as Zhu & Wu (2018). Additionally, the model captures trends related to dependencies on stellar metallicity (Schlecker et al. 2021a; Emsenhuber et al. 2025) and stellar mass (Burn et al. 2021), as well as patterns in period ratio distributions (Mulders et al. 2019; Burn et al. 2021; Emsenhuber et al. 2025) and eccentricity distributions (Burn et al. 2021; Emsenhuber et al. 2025). Notable architectural features include similarities and mass or size ordering (Mishra et al. 2021), the 'peas-in-a-pod' structure (Mishra et al. 2021), a bimodal mass function distinguishing sub-Neptunes and Gas Giants (Mulders et al. 2019; Emsenhuber et al. 2025), and a mean observed multiplicity of approximately 1.6 (Emsenhuber et al. 2025). Despite these successes, the model has limitations in reproducing certain observed characteristics of planetary populations. For example, the positive correlation between Super Earths and cold Giants is weaker than observed (Schlecker et al. 2021a). Moreover, the model predicts an overproduction of planets per system – by at least a factor of 1.7 (Mulders et al. 2019; Emsenhuber et al. 2025). Synthetic planets also tend to be closer to their stars than observed (Mulders et al. 2019; Emsenhuber et al. 2025), and the mass distribution does not align precisely with observations (Emsenhuber et al. 2025). Finally, the model produces an excess of planets in or near mean-motion resonances (Mulders et al. 2019; Burn et al. 2021; Emsenhuber et al. 2025), which is inconsistent with the distribution seen in observed systems. In summary, planetary system populations are realistic at the system and architectural scale rather than at the individual planet scale. These results suggest that we can consider synthetic

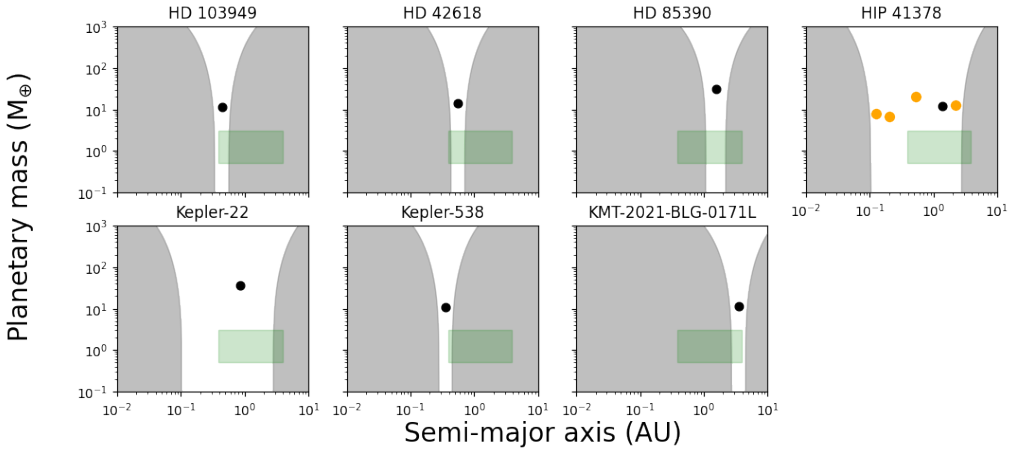

Fig. 3. Systems around G stars with a resulting voting rate above 90%. The green areas represent the definition of an Earth-like planet in the study in terms of equilibrium temperature and mass. The grey areas represent the combinations of mass and semi-major axis for which the Hill-stability criterion is met with the addition of a new planet. The black dots correspond to planets for which we know the mass, and the orange dots correspond to planet for which we only know the radius, and the mass has been derived thanks to the work of Parc et al. (2024).

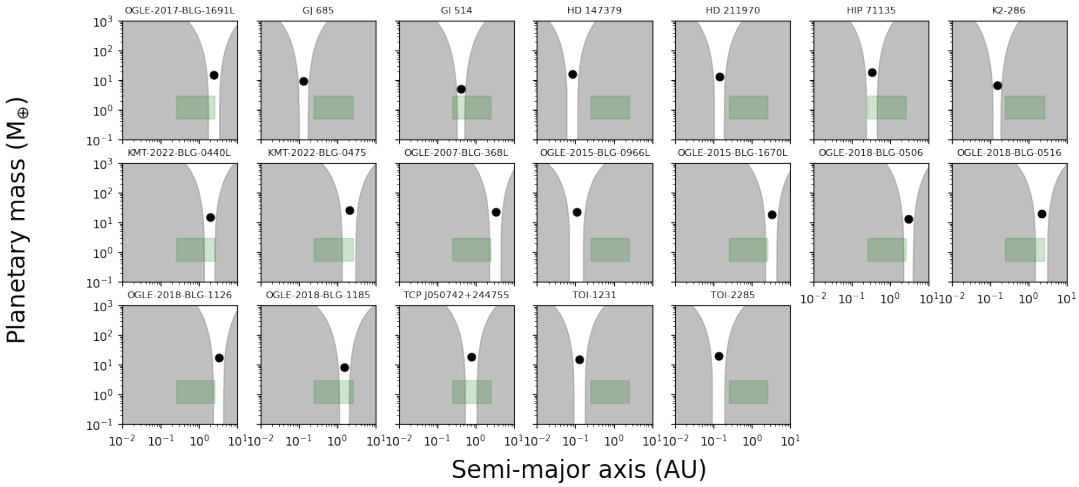

Fig. 4. Systems around early-M and late-K stars with a resulting voting rate above 90%. The green areas represent the definition of an Earth-like planet in the study in terms of equilibrium temperature and mass. The grey areas represent the combinations of mass and semi-major axis for which the Hill-stability criterion is met with the addition of a new planet. The black dots correspond to the planets already known in these systems.

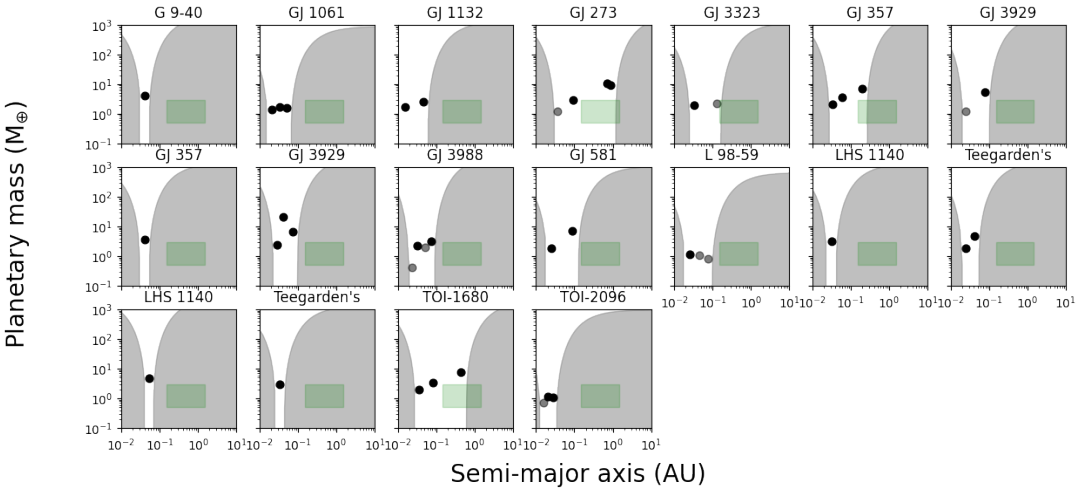

Fig. 5. Systems around late-M stars with a resulting voting rate above 90%. The green areas represent the definition of an Earth-like planet in the study in terms of equilibrium temperature and mass. The grey areas represent the combinations of mass and semi-major axis for which the Hill-stability criterion is met with the addition of a new planet. The dots represent the planets already known in those systems: the black dots for planets with a RV semi-amplitude above the threshold of detection bias (detectable planets) and the grey dots for the planets with a RV semi-amplitude below this threshold. Only the detectable planets count in the calculation of the architecture of the systems.

Table 6. List of 44 systems achieving a voting rate (VR) of over 90%.

System	VR	Reference
G stars		
HD 103949	96%	Feng et al. (2019)
HD 42618	95%	Fulton et al. (2016)
HD 85390	93%	Wittenmyer et al. (2019)
HIP 41378	97%	Vanderburg et al. (2016)
Kepler-22	92%	Borucki et al. (2012)
Kepler-538	96%	Mayo et al. (2019)
KMT-2021-BLG-0171L	94%	Yang et al. (2022)
Late-K and early-M stars		
OGLE-2017-BLG-1691L	97%	Han et al. (2022)
GJ 685	92%	Pinamonti et al. (2019)
Gl 514	95%	Damasso et al. (2022)
HD 147379	96%	Reiners et al. (2018)
HD 211970	92%	Feng et al. (2019)
HIP 71135	95%	Feng et al. (2019)
K2-286	94%	Díez Alonso et al. (2019)
KMT-2022-BLG-0440L	98%	Zhang et al. (2023)
KMT-2022-BLG-0475	92%	Han et al. (2023)
OGLE-2007-BLG-368L	96%	Sumi et al. (2010)
OGLE-2015-BLG-0966L	96%	Street et al. (2016)
OGLE-2015-BLG-1670L	96%	Ranc et al. (2019)
OGLE-2018-BLG-0506	96%	Hwang et al. (2022)
OGLE-2018-BLG-0516	95%	Hwang et al. (2022)
OGLE-2018-BLG-1126	96%	Gould et al. (2022)
OGLE-2018-BLG-1185	96%	Kondo et al. (2021)
TCP J050742+244755	97%	Nucita et al. (2018)
TOI-1231	91%	Burt et al. (2021)
TOI-2285	91%	Fukui et al. (2022)
Late-M stars		
G 9-40	91%	Stefansson et al. (2020)
GJ 1061	98%	Dreizler et al. (2020)
GJ 1132	94%	Bonfils et al. (2018)
GJ 273	92%	Pozuelos et al. (2020)
GJ 3323	95%	Astudillo-Defru et al. (2017b)
GJ 357	98%	Jenkins et al. (2019)
GJ 3929	91%	Kemmer et al. (2022)
GJ 3988	91%	Gorrini et al. (2023)
GJ 581	98%	von Stauffenberg et al. (2024)
L 98-59	98%	Demangeon et al. (2021)
LHS 1140	95%	Dittmann et al. (2017)
Teegarden's	95%	Dreizler et al. (2024)
TOI-1680	91%	Ghachoui et al. (2023)
TOI-2096	98%	Pozuelos et al. (2023)
TOI-2136	91%	Gan et al. (2022)
TOI-237	92%	Waalkes et al. (2021)
Wolf 1061	98%	Astudillo-Defru et al. (2017b)
YZ Cet	98%	Astudillo-Defru et al. (2017a)

planetary systems as proxies for real systems when examining architectures and correlations between planetary properties.

Another limitation of the model is the limited amount of training data: there are only between 5000 and 20 000 instances depending on the populations, due to the time required to generate synthetic systems. This issue could be addressed by the upcoming work of Alibert et al. (2024), which employs a transformer-based generative model to emulate the Bern model and generate millions of synthetic planetary systems in an hour.

Finally, another weakness of the study is the simplistic approach to handling observational bias, which does not account for various factors that may influence planet detection, such as stellar activity, the presence of other planets in the system, observation frequency, and orbital period. Forthcoming works should address this issue.

5.2. Conclusion

In this work, we have developed a model using a Random Forest Classifier to predict which known planetary systems are most likely to host an Earth-like planet. The model was trained on a dataset of synthetic planetary systems from the Bern model to which we applied an observational bias to extract observable properties. We conducted tests to determine the optimal descriptive features of synthetic systems to enhance model performance, finding that the mass, period of the innermost detectable planet (IDP), and system architecture are the three properties that provide the most information about the presence of an Earth-like planet. These findings are consistent with the results of Davoult et al. (2024).

The model demonstrated excellent performance, achieving a precision score of up to 0.99 on the test datasets, which means that 99% of the positive predictions were True Positives. This result proves that the model can accurately identify the properties of systems with and without ELPs within a dataset derived from the Bern model.

Therefore, we used the model to predict the presence of an Earth-like planet in a sample of 1567 observed GKM systems, for which we know at least one planet and the properties necessary for the model to function (the mass and semi-major axis or period of at least one planet and the mass of the central star). The results indicate that 44 systems (listed in Table 6) exhibit architectures suggesting the presence of an Earth-like planet. Further study of the stability state of these systems with the addition of a new planet has shown that 95.5% of those systems would remain stable with the addition of an Earth-like planet.

We caution that the results heavily rely on the Bern model and should be interpreted cautiously. However, we recommend prioritising the study of these systems because both positive and negative outcomes provide conclusive findings. Negative results would indicate that the Bern model is having difficulty reproducing the architecture of the systems, and would be a path for improvement. In the context of predicting exoplanet detection using global models of planetary formation, it is crucial to link observations with theoretical models to rigorously test them closely.

Acknowledgements. This work has been carried out within the framework of the National Centre of Competence in Research PlanetS supported by the Swiss National Science Foundation under grants 51NF40_182901 and 51NF40_205606. The authors acknowledge the financial support of the SNSF.

References

Alibert, Y., Davoult, J., & Marques, S. 2024, A&A, submitted
Astudillo-Defru, N., Díaz, R. F., Bonfils, X., et al. 2017a, A&A, 605, L11
Astudillo-Defru, N., Forveille, T., Bonfils, X., et al. 2017b, A&A, 602, A88
Beckwith, S. V. W., & Sargent, A. I. 1996, Nature, 383, 139
Bonfils, X., Almenara, J. M., Cloutier, R., et al. 2018, A&A, 618, A142
Borucki, W. J., Koch, D. G., Batalha, N., et al. 2012, ApJ, 745, 120
Bovaird, T., & Lineweaver, C. H. 2013, MNRAS, 435, 1126
Bovaird, T., Lineweaver, C. H., & Jacobsen, S. K. 2015, MNRAS, 448, 3608
Bryan, M. L., & Lee, E. J. 2024, ApJ, 968, L25

Bryson, S., Kunimoto, M., Kopparapu, R. K., et al. 2021, AJ, 161, 36
Burn, R., Schlecker, M., Mordasini, C., et al. 2021, A&A, 656, A72
Burn, R., Mordasini, C., Mishra, L., et al. 2024, Nat. Astron., 8, 463
Burt, J. A., Dragomir, D., Mollière, P., et al. 2021, AJ, 162, 87
Chen, D.-C., Mordasini, C., Xie, J.-W., Zhou, J.-L., & Emsenhuber, A. 2024, A&A, 687, A25
Damasso, M., Perger, M., Almenara, J. M., et al. 2022, A&A, 666, A187
Davoult, J., Alibert, Y., & Mishra, L. 2024, A&A, 689, A309
Demangeon, O. D. S., Zapatero Osorio, M. R., Alibert, Y., et al. 2021, A&A, 653, A41
Dietrich, J., & Apai, D. 2020, AJ, 160, 107
Díez Alonso, E., González Hernández, J. I., Toledo-Padrón, B., et al. 2019, MNRAS, 489, 5928
Dittmann, J. A., Irwin, J. M., Charbonneau, D., et al. 2017, Nature, 544, 333
Dreizler, S., Jeffers, S. V., Rodríguez, E., et al. 2020, MNRAS, 493, 536
Dreizler, S., Luque, R., Ribas, I., et al. 2024, A&A, 684, A117
Emsenhuber, A., Mordasini, C., Burn, R., et al. 2021a, A&A, 656, A69
Emsenhuber, A., Mordasini, C., Burn, R., et al. 2021b, ApJ, 656, A70
Emsenhuber, A., Mordasini, C., & Burn, R. 2023, Eur. Phys. J. Plus, 138, 181
Emsenhuber, A., Mordasini, C., Mayor, M., et al. 2025, A&A, submitted
Fabrycky, D. C., Lissauer, J. J., Ragozzine, D., et al. 2014, ApJ, 790, 146
Feng, F., Crane, J. D., Xuesong Wang, S., et al. 2019, ApJS, 242, 25
Fukui, A., Kimura, T., Hirano, T., et al. 2022, PASJ, 74, L1
Fulton, B. J., Howard, A. W., Weiss, L. M., et al. 2016, ApJ, 830, 46
Gan, T., Soubkiou, A., Wang, S. X., et al. 2022, MNRAS, 514, 4120
Ghachoui, M., Soubkiou, A., Wells, R. D., et al. 2023, A&A, 677, A31
Gilbert, G. J., & Fabrycky, D. C. 2020, AJ, 159, 281
Gorrini, P., Kemmer, J., Dreizler, S., et al. 2023, A&A, 680, A28
Gould, A., Han, C., Zang, W., et al. 2022, A&A, 664, A13
Haghighipour, N. 2015, in Encyclopedia of Astrobiology, eds. M. Gargaud, W. M. Irvine, R. Amils, I. Cleaves, Henderson James (Jim), D. L. Pinti, J. C. Quintanilla, D. Rouan, T. Spohn, S. Tirard, & M. Viso (Berlin: Springer), 1054
Haisch, Karl E., J., Lada, E. A., & Lada, C. J. 2001, AJ, 553, L153
Han, C., Kim, D., Gould, A., et al. 2022, A&A, 664, A33
Han, C., Lee, C.-U., Bond, I. A., et al. 2023, A&A, 676, A97
Hsu, D. C., Ford, E. B., Ragozzine, D., & Ashby, K. 2019, AJ, 158, 109
Hwang, K.-H., Zang, W., Gould, A., et al. 2022, AJ, 163, 43
Jenkins, J. S., Pozuelos, F. J., Tuomi, M., et al. 2019, MNRAS, 490, 5585
Kammerer, J., & Quanz, S. P. 2018, A&A, 609, A4
Kemmer, J., Dreizler, S., Kossakowski, D., et al. 2022, A&A, 659, A17
Kondo, I., Yee, J. C., Bennett, D. P., et al. 2021, AJ, 162, 77
Kopparapu, R. K., Hébrard, E., Belikov, R., et al. 2018, ApJ, 856, 122
Kunimoto, M., & Matthews, J. M. 2020, AJ, 159, 248
Lara, P., Cordero-Tercero, G., & Allen, C. 2020, PASJ, 72, 24
Latham, D. W., Rowe, J. F., Quinn, S. N., et al. 2011, ApJ, 732, L24
Lissauer, J. J., Ragozzine, D., Fabrycky, D. C., et al. 2011, ApJS, 197, 8
Lundberg, S. M., & Lee, S.-I. 2017, in Advances in Neural Information Processing Systems, eds. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (New York: Curran Associates, Inc.), 30
Mayo, A. W., Rajpaul, V. M., Buchhave, L. A., et al. 2019, AJ, 158, 165
Mayor, M., Marmier, M., Lovis, C., et al. 2011, arXiv e-prints [arXiv:1109.2497]
Millholland, S., Wang, S., & Laughlin, G. 2017, ApJ, 849, L33
Mishra, L., Alibert, Y., Leleu, A., et al. 2021, A&A, 656, A74
Mishra, L., Alibert, Y., Udry, S., & Mordasini, C. 2023, A&A, 670, A68
Mordasini, C. 2018, in Handbook of Exoplanets, eds. H. J. Deeg, & J. A. Belmonte (Berlin: Springer), 143
Mousavi-Sadr, M., Gozaliasl, G., & Jassur, D. M. 2021, PASA, 38, e015
Mulders, G. D., Mordasini, C., Pascucci, I., et al. 2019, ApJ, 887, 157
Murray, N., Chaboyer, B., Arras, P., Hansen, B., & Noyes, R. W. 2001, ApJ, 555, 801
Nucita, A. A., Licchelli, D., De Paolis, F., et al. 2018, MNRAS, 476, 2962
Parc, L., Bouchy, F., Venturini, J., Dorn, C., & Helled, R. 2024, A&A, 688, A59
Pinamonti, M., Sozzetti, A., Giacobbe, P., et al. 2019, A&A, 625, A126
Pollack, J. B., Hubickyj, O., Bodenheimer, P., et al. 1996, Icarus, 124, 62
Pozuelos, F. J., Suárez, J. C., de Elía, G. C., et al. 2020, A&A, 641, A23
Pozuelos, F. J., Timmermans, M., Rackham, B. V., et al. 2023, A&A, 672, A70
Quanz, S. P., Ottiger, M., Fontanet, E., et al. 2022, å, 664, A21
Ranc, C., Bennett, D. P., Hirao, Y., et al. 2019, AJ, 157, 232
Rauer, H., Catala, C., Aerts, C., et al. 2014, Exp. Astron., 38, 249
Reiners, A., Ribas, I., Zechmeister, M., et al. 2018, A&A, 609, L5
Sandford, E., Kipping, D., & Collins, M. 2021, MNRAS, 505, 2224
Santos, N. C., Israelian, G., Mayor, M., Rebolo, R., & Udry, S. 2003, A&A, 398, 363
Schlecker, M., Mordasini, C., Emsenhuber, A., et al. 2021a, A&A, 656, A71
Schlecker, M., Pham, D., Burn, R., et al. 2021b, A&A, 656, A73
Schneider, J., Dedieu, C., Le Sidaner, P., Savalle, R., & Zolotukhin, I. 2011, A&A, 532, A79
Stefansson, G., Cañas, C., Wisniewski, J., et al. 2020, AJ, 159, 100
Steffen, J. H., Ragozzine, D., Fabrycky, D. C., et al. 2012, Proc. Natl. Acad. Sci., 109, 7982
Street, R. A., Udalski, A., Calchi Novati, S., et al. 2016, ApJ, 819, 93
Sumi, T., Bennett, D. P., Bond, I. A., et al. 2010, ApJ, 710, 1641
Tuomi, M., Jones, H. R. A., Butler, R. P., et al. 2019, arXiv e-prints [arXiv:1906.04644]
Vanderburg, A., Becker, J. C., Kristiansen, M. H., et al. 2016, ApJ, 827, L10
Veras, D., & Armitage, P. J. 2004, MNRAS, 347, 613
von Stauffenberg, A., Trifonov, T., Quirrenbach, A., et al. 2024, A&A, 688, A112
Waalkes, W. C., Berta-Thompson, Z. K., Collins, K. A., et al. 2021, AJ, 161, 13
Weiss, L. M., Marcy, G. W., Petigura, E. A., et al. 2018, AJ, 155, 48
Wittenmyer, R. A., Clark, J. T., Zhao, J., et al. 2019, MNRAS, 484, 5859
Yang, H., Zang, W., Gould, A., et al. 2022, MNRAS, 516, 1894
Zhang, J., Zang, W., Jung, Y. K., et al. 2023, MNRAS, 522, 6055
Zhu, W. 2024, Res. Astron. Astrophys., 24, 045013
Zhu, W., & Wu, Y. 2018, AJ, 156, 92