MNRAS 000, 1–16 (2025) Preprint 30 September 2025 Compiled using MNRAS LATEX style file v3.2

Technique-agnostic exoplanet demography for the *Roman* era – I. Testing a retrieval framework using simulated *Kepler*-like transit datasets

A. Priyadarshi[1] and E. Kerins[1]★
[1]*Jodrell Bank Centre for Astrophysics, University of Manchester, Oxford Road, Manchester, M13 9PL, UK*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The Nancy Grace Roman Space Telescope (*Roman*) will unveil for the first time the full architecture of planetary systems across Galactic distances through the discovery of up to 200,000 cool and hot exoplanets using microlensing and transit detection methods. *Roman*'s huge exoplanet haul, and Galactic reach, will require new methods to leverage the full exoplanet demographic content of the combined microlensing and transit samples, given the different sensitivity bias of the techniques to planet and host properties and Galactic location. We present a framework for *technique-agnostic exoplanet demography* (TAED) that can allow large, multi-technique exoplanet samples to be combined for demographic studies. Our TAED forward modelling and retrieval framework uses parameterised model exoplanet demographic distributions to embed planetary systems within a stellar population synthesis model of the Galaxy, enabling internally consistent forecasts to be made for all detection methods that are based on spatio-kinematic system properties. In this paper, as a first test of the TAED framework, we apply it to simulated transit datasets based on the *Kepler* Data Release 25 to assess parameter recovery accuracy and method scalability for a single large homogeneous dataset. We find that optimisation using differential evolution provides a computationally scalable framework that gives a good balance between computational efficiency and accuracy of parameter recovery.

Key words: planets and satellites: detection – Galaxy: stellar content – methods: data analysis

1 INTRODUCTION

Whilst there are now more than 6,000 confirmed exoplanets, these planets still reflect a very limited region of the exoplanet discovery space. The vast majority of known systems lie within 1 kpc of the Sun and therefore occupy only a small fraction of the Galactic stellar volume. This volume contains an even smaller fraction of all Galactic stars, and these stars are a biased subset of all Galactic stars in terms of age, chemistry and kinematics (e.g. Gaudi et al. 2021). The deep connection between planets and their host stars means that biases in the host star sample likely translate into biases in the characteristics of the exoplanet sample, with respect to the Galactic population as a whole. Such bias can lead to an incomplete and skewed measure of exoplanet occurrence, or a misrepresentation of the prevalent architecture and characteristics of planetary systems across the Galaxy.

The potential for bias in the stellar sample is compounded further by the use of different detection techniques that have different sensitivities to planet and host properties. The transit and radial velocity methods have so far provided us with the bulk of confirmed planetary systems, though they are samples that favour hotter (shorter period) and larger (more massive) exoplanets orbiting seismically quiet main sequence stars. Other survey techniques that are contributing significantly to current statistics include gravitational microlensing and direct imaging surveys, and it is expected that large samples of astrometrically-detected planets will soon be delivered from reductions of *Gaia* observations (Perryman et al. 2014). These techniques allow us to explore the occurrence of cooler (longer period) exoplanets, with microlensing probing cool planets over wide mass and distance scales, and direct imaging and astrometry typically targeting mostly nearby and larger planetary mass scales.

A picture of exoplanet demographics that is reflective of the Galactic planetary population as a whole is needed to provide the strongest constraints on the physics of planet formation. We require exoplanet surveys over larger Galactic volumes that contain a representative mix of stars. We need to ensure that these representative volumes can be reached by multiple detection methods that can span the full range of exoplanet architecture, including large and small planets, and hot and cold planets. Lastly, we need methods that can tie together the full set of exoplanet demographic information provided by these different techniques, in a manner that takes full account of their variation in sensitivity to planetary orbit and bulk properties, host characteristics, and Galactic location.

The Nancy Grace Roman Space Telescope (hereafter *Roman*), scheduled for launch in late 2026, will take an enormous stride in improving our understanding of exoplanet demography over Galactic distance scales. The core science aim of *Roman*'s Galactic Bulge Time Domain Survey (GBTDS) is to detect and measure the masses of cool exoplanets using the gravitational microlensing method, including free-floating planets unbound from host stars (Penny et al. 2019; Johnson et al. 2020). Simulations indicate that *Roman* will be capable of finding around 1,500 planets using microlensing, provid-

★ Corresponding author: eamonn.kerins@manchester.ac.uk

ing a cool planet sample that is comparable in size to that of the *Kepler* transit sample.

The same survey is also predicted to find 60,000 - 200,000 transiting planets (Montet et al. 2017; Wilson et al. 2023) over Galactic distance scales. The GBTDS will therefore be the first exoplanet survey sensitive to essentially the full range of planetary orbits across Galactic distances for planets of Earth size and larger.

The remaining hurdle is to employ demographic modelling methods that can be used on large exoplanet samples obtained with different detection methods probing different host star populations. These methods will need to be able to account for correlations not just between planet and host properties, but also between host properties and Galactic location. In this paper, we refer to demographic modelling that is agnostic to spatio-kinematic based detection methods (e.g. transit, radial velocity, microlensing and astrometry) as *technique agnostic exoplanet demography* (hereafter TAED). A TAED approach requires the integration of models of Galactic structure, stellar populations and exoplanet demography. It's an approach capable of taking full advantage of the *Roman* exoplanet samples for demographic studies.

Several studies have attempted to model exoplanet demographics, incorporating the effect of detection bias. For instance, Hsu et al. (2018) use a hierarchical Bayesian modelling framework applied to their `SysSim` forward model applied to *Kepler*. This is an example of a retrieval framework that is tailored to a specific survey dataset. `EPOS` (Mulders et al. 2018) is another model which aims to constrain the properties of exoplanet populations, like occurrence rates and orbital architectures, using *Kepler* data. Similar to the `SysSim` model, `EPOS` samples the planets from a distribution over grids of radius and period. Both of these models could be adapted and applied to other surveys and detection techniques. But, both lack a Galactic model to account for variations in host properties with Galactic location. This makes it difficult to use such models to combine results from methods with very different distance sensitivity, such as transits and microlensing.

In this paper, we develop and present a working and scalable TAED-based method for the retrieval of exoplanet demographic parameters from large samples of exoplanets obtained potentially over a range of Galactic distances. In this paper, we restrict the application of our TAED framework only to relatively local samples of transiting planets simulated from a realistic *Kepler*-like survey with known detection efficiency, in order to establish the efficacy and feasibility of the framework. Application to other methods (or multiple methods), and to samples that simulate *Roman*'s specific detection characteristics, will be the focus of future papers.

The present paper is structured as follows. In Section 2 we introduce our TAED approach and simulate a stellar sample for a *Kepler*-like transit survey. The simulated transit sample itself is discussed in Section 3. The initial set of toy planetary models and the likelihood function used for retrieval are discussed in Section 4. The ability of the TAED framework to successfully retrieve accurate demographic information is examined in Section 5, where we explore different retrieval schemes and show that those based on differential evolution appear promising. We recover optimal parameters for simple planet demographic models applied to the true *Kepler* dataset in Section 6, highlighting how the TAED framework can measure demographic model parameters, but also how it shows up deficiencies in the flexibility of simple demographic models. We summarise our findings in Section 7.

2 TECHNIQUE-AGNOSTIC EXOPLANET DEMOGRAPHY

Using samples of planets for demographic studies that have been obtained through different detection methods requires the ability to account for differences in selection bias between methods. For methods like transit and microlensing, these differences are substantial.

The majority of transit events to date are located within 1 kpc from us and involve planets on relatively short orbital periods. Transit samples are also biased in favour of larger planets, which yield deeper transits. Additionally, there is a strong correlation between host type and distance from the observer, with transits involving lower-mass hosts tending to be close by, as these hosts are intrinsically fainter.

By contrast, as it does not rely on the need to detect the host star, microlensing is sensitive to planetary systems over a large range of distances, and so typically reside several kpc from the Sun. Microlensing planets are observable when their projected separation is comparable to the Einstein radius of their host, which means they are typically on much wider orbits than transiting planets. Unlike transiting planets, microlensing observations yield the planet–host mass ratio and sky-projected separation, rather than the planet–host size ratio and orbital period. Planet samples detected through microlensing are biased towards the most common types of star (low mass M and K dwarfs), whereas transit surveys have tended to target FGK-type hosts that exhibit low-levels of intrinsic variability.

Demographic studies to date that have aimed to synthesise transit and microlensing samples have usually done so by restricting their scope to sub-samples that have properties in common, such as transit and microlensing samples that involve hosts of a similar type. Whilst this approach is a valid way to proceed, the relatively small intersection in common properties between microlensing and transit samples means that many valid planets must be excluded from the joint analysis. This, in turn, strongly restricts the scope of demographic information that can be recovered.

In order to fully exploit exoplanet samples from different detection techniques, we must consider not just the demographics of planets, but also the demographics of their hosts. By considering both jointly, we can develop a technique agnostic exoplanet demographic (TAED) framework that can provide consistent forecasts for planet yields across different surveys, including surveys employing different detection techniques. In this way, a retrieval framework can then use the union, rather than the intersection, of the joint sample to recover the underlying planet demography.

The TAED approach shares some similarity to the Exoplanet Population Observation Simulator (EPOS, Mulders et al. 2018) in that it employs a parameterised forward model to describe the underlying planet population. The main difference between TAED and EPOS is that in the TAED approach, we are layering our planet model on top of a model for the Galactic stellar population (i.e. the host stars). Without this additional layer, it is not possible to make meaningful joint forecasts for methods like transits and microlensing that have very different sensitivity to host type and Galactic location.

2.1 Host star model

We use the Besançon Galactic Model (Robin et al. 2003; Marshall et al. 2006; Robin et al. 2012, hereafter BGM) to describe the host star population. The BGM uses input models for the spatial distribution, masses and kinematics of stars as a function of age, together with stellar atmosphere models, to predict the observable properties of a synthetic catalogue of stars that would be seen along any given direction by magnitude and/or kinematics-limited surveys. The BGM incorporates a self-consistent 3D dust map so that stellar ap-

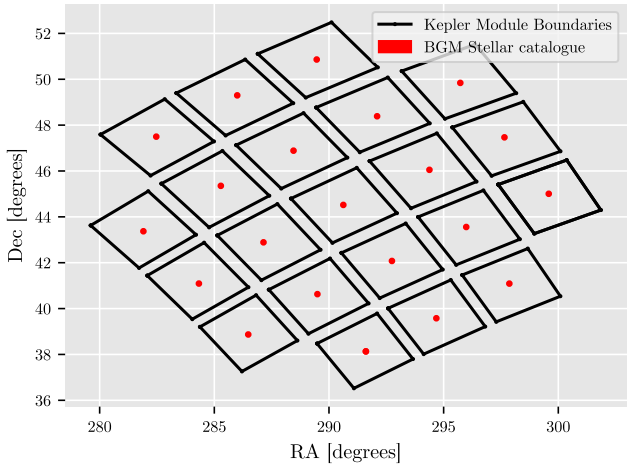

Figure 1. The *Kepler* focal plane array showing the arrangement of the 21 detector modules (black outlines). The red dots show the locations used for our BGM synthetic stellar catalogues.

parent magnitudes can be generated in a range of standard optical and infrared passbands.

For our test of the TAED framework, we simulate a *Kepler*-like transit-observables by generating BGM stellar catalogues towards the *Kepler* field location[1]. We simulate stars within 0.1 square degree pencil beams centred around each of the 21 *Kepler* detector array modules as shown in Figure 1. This spatial sampling enables realistic simulation of host star populations across the *Kepler* field, accounting for Galactic structure and stellar diversity.

Whilst the expected *Roman* transit sample is expected to dwarf that from *Kepler*, *Kepler* nonetheless provides a useful testbed for simulating transit samples of significant size. Additionally, the *Kepler* completeness is well characterised and is publicly available through KeplerPORTs (Burke et al. 2015). This means we can simulate a transit sample using a realistic efficiency framework that we may expect to share some qualitative similarity with the framework that will be ultimately developed for *Roman*, even though, quantitatively, the *Roman* and *Kepler* detection efficiencies will be very different.

The output from BGM comprises catalogues of artificial stars with distance, magnitudes and colours in several Johnson-Cousins passbands, stellar effective temperature, surface gravity, mass, and radius. Colour-magnitude relations from Jordi et al. (2006) and Brown et al. (2011) are used to transform from Johnson-Cousins bands to *Kepler* K_p magnitudes. Our final shortlist of BGM stars was obtained by keeping only BGM stars that were deemed close analogues of *Kepler* stars. This selection is detailed in Section 3.

2.2 Exoplanet model

Having generated a catalogue of host stars, we now populate planets around them. Planet's physical characteristics are assigned using parameterised distribution functions. We determine the number of planets around each star using a weighted planet occurrence rate, which depends on the stellar mass and type, as described later in this section. Subsequently, these planets are assigned physical parameters required to compute their detectability for a given method. In this paper, we are testing the approach using only simulated transit samples, and so we provide details on the generation of observables relevant to transit detection.

Motivated by Pascucci et al. (2018), we adopt the planet–host mass ratio as the primary and universal descriptor of planet bulk characteristics that, together with host mass, controls planet occurrence.

Accordingly, we assume a two-stage broken power law for the number N of planets with planet-star mass ratio, q, of the form

$$\frac{\mathrm{d}N}{\mathrm{d}\log q} \propto \left(\frac{q}{q_{\rm br}}\right)^n \text{ where } n = \begin{cases} n_1 & (q_1 \le q \le q_{\rm br}) \\ n_2 & (q_{\rm br} < q \le q_2) \end{cases}, \tag{1}$$

where n is a power-law exponent.

Defining β as the ratio of the sum of masses of all planets around a host star to the mass of the star, we have

$$\beta = \langle N_p \rangle \frac{\langle M_{\rm p} \rangle}{M_\star} = \langle N_p \rangle \langle q \rangle, \tag{2}$$

where $\langle \ldots \rangle$ denotes averaging within a given host system and N_p is the number of planets per star. From Equation (1), the average mass ratio is

$$\langle q \rangle = \frac{\frac{1}{n_1+1} q_{\rm br}^{1-n_1}(q_{\rm br}^{n_1+1} - q_1^{n_1+1}) + \frac{1}{n_2+1} q_{\rm br}^{1-n_2}(q_2^{n_2+1} - q_{\rm br}^{n_2+1})}{\frac{1}{n_1} q_{\rm br}^{1-n_1}(q_{\rm br}^{n_1} - q_1^{n_1}) + \frac{1}{n_2} q_{\rm br}^{1-n_2}(q_2^{n_2} - q_{\rm br}^{n_2})}. \tag{3}$$

By fixing a universal value for β, we compute $\langle N_p \rangle = \beta / \langle q \rangle$ and then, for each star, select the number of planets per star, N_p, to be a Poisson deviate about $\langle N_p \rangle$. Whilst the model may predict multiple planets exist around each host, we do not model multi-planet transit detections. The role N_p plays is in determining the overall yield of single-planet transits. The larger the planet multiplicity, the larger the overall expected yield will be of single-planet transit lightcurves.

Like q, the planet–host separation, a, is allowed to follow a two-stage power-law distribution of the form

$$\frac{\mathrm{d}N}{\mathrm{d}\log a} \propto \left(\frac{a}{a_{\rm br}}\right)^y \text{ where } y = \begin{cases} y_1 & (a_1 \le a \le a_{\rm br}) \\ y_2 & (a_{\rm br} < a \le a_2). \end{cases} \tag{4}$$

For adopted values of β, n, $q_{\rm max}$, $q_{\rm min}$, and $q_{\rm br}$ we can use Equations (3) and (2) to determine N_p for a given system and then use Equations (1) and (4) to randomly generate N_p pairs of q and a. q is converted to planet mass via $M_p = qM_\star$ and then to radius via the mass-radius-temperature relations in Edmondson et al. (2023). Planet temperature is computed using the BGM output values for stellar effective temperature, together with a, to determine equilibrium temperature for a fixed Bond Albedo of 0.3. Finally, the period of the exoplanet, P, is calculated directly from a and $M_\star$ via Kepler's 3rd law.

We make a number of strong assumptions for the initial tests presented in this paper. Firstly, we assume that planets lie on circular orbits. Whilst eccentricity is not readily observable in transit profiles, eccentric orbits will affect the transit probability. The main effect of ignoring eccentric orbits on demographic modelling of transits would be a potentially biased recovery of β, which governs planet multiplicity. We also do not model the effect of mutual inclination and instead assume all planets to be coplanar within a given system. Mutual inclination increases the probability of observing at least one transit around a given system, so ignoring this can also bias recovery of β.

In principle, it is straightforward to add both eccentricity and mutual inclination distributions into the TAED framework. Ultimately, as the number of fit parameters increases, the recovery precision will

[1] https://keplergo.github.io/KeplerScienceWebsite

be governed by the size and quality of the dataset, and the extent to which the data provides complementary sensitivity to parameters.

This is where the addition of a microlensing exoplanet sample may be helpful. In the case of planets detected through microlensing, the distribution of eccentricities will imprint directly upon the distribution of projected planet–host separation. A joint demographic analysis of transit and microlensing samples may well therefore be able to recover separately the planet multiplicity and eccentricity distribution. The inclusion of mutual inclination is likely to be important for observed transit yield but unimportant for microlensing. This complementary sensitivity will therefore play a crucial role within a TAED-based analysis.

As we are focusing in this paper only on transiting planets, we will leave to future papers investigations of how joint transit and microlensing samples may be able to constrain eccentricity and mutual inclination distributions.

3 SIMULATING A *KEPLER*-LIKE TRANSIT SAMPLE

The *Roman* transit sample is expected to be 1.5-2 orders of magnitude larger than that obtained by *Kepler*. It will comprise transits observed over Galactic distances, not just within 1 kpc of the Sun. Nonetheless, the *Kepler* exoplanet sample is the largest to date obtained by a single observatory, from the ground or space, via any detection method. Additionally, it has a very well characterised detection efficiency (Burke & Catanzarite 2017). It therefore provides the best available analogue for testing a TAED framework applicable to *Roman*.

To simulate a sample indicative of one observed by a mission like *Kepler*, we need to assert an equivalence between a BGM simulated host star and a real star within the *Kepler* input catalogue. This matching was achieved by assigning a similarity score

$$S = \sqrt{\left(\frac{\Delta T_\star}{T_\star}\right)^2 + \left(\frac{\Delta R_\star}{R_\star}\right)^2 + \left(\frac{\Delta K_p}{K_p}\right)^2 + \left(\frac{\Delta M_\star}{M_\star}\right)^2}, \tag{5}$$

where small S is better and $S = 0$ would indicate a perfect match. The numerators in Equation (5) refer to the difference in the value of the parameter of the star from the BGM sample to that from the *Kepler* input catalogue (i.e $\Delta T_\star = T_{\star,\mathrm{BGM}} - T_{\star,\mathrm{Kepler}}$, and so on). Denominators refer to the parameters from the *Kepler* catalogue. The parameters $R_\star$, $T_\star$, K_P, and $M_\star$ were selected as they are the stellar parameters used by KeplerPORTs (Burke & Catanzarite 2017) to determine detection efficiency.

We ensure that matched stars are located within the same *Kepler* detector module, and we discard BGM stars that have $S > 0.15$. Not all stars that *Kepler* is able to detect are used within KeplerPORTs. So the number of matching BGM stars does not bear direct correspondence to the number within the KeplerPORTs catalogue. Because of this, our simulated yields are rescaled to match what they would be if the simulated and real samples were of the same size.

At this point, we have a TAED transit simulation framework to compute observable transit yields for a given set of planet model parameters. But what we ultimately wish to do is to perform model retrieval – i.e. be able to recover the planet model parameters that provide the best match to an observed dataset. This is what we now turn to.

4 PLANETARY MODEL RETRIEVAL

To test the sensitivity and efficacy of a TAED retrieval framework, we perform injection tests where we use the planet-host model in Section 2 with a known set of model parameters to generate a synthetic observed dataset. We then perform a retrieval on the synthetic dataset to test the recovery of the parameters.

For our initial tests, we use a simplified set of planetary models with single slope power laws in q and a. So we set $n = n_1 = n_2$ in Equation (1) and $y = y_1 = y_2$ in Equation (4). But even these simplified injection sets provide 7 parameters to retrieve: the minimum and maximum of planet-host mass ratio q (q_1, q_2), along with the q distribution slope n; the minimum and maximum for host separation a (a_1, a_2), along with the a distribution slope y; and β which is the average mass per host in planets, normalised to the host mass. We defer to Section 6 to test recovery for the full 11-parameter models that employ broken power law distributions for q and a.

We select 4 different injected models to represent different ground truths. Set 1 is chosen to span a relatively narrow range of q, but with a wide and flat distribution in a. Set 2 spans a broad range in q with a modest slope favouring larger q. It has a more restricted span in a than Set 1. Set 3 has the same span in q as Set 2 but with a larger slope. It also has a very restricted range in a. Set 4 has the same q model as Set 3 but has the broadest range of a of all of the models. In each case, a value of β is chosen so that the predicted average total number of planets per host is comparable to the number of Solar System planets.

For each model set, we ran the simulation to provide an "observed" yield in the limits of the perfect detection efficiency. Given differences in the models, this means the underlying number of generated planets differs markedly between the sets. We list our injected model parameters in Table 1, with the corresponding distributions of q and a shown in Figure 2.

The simulations are used to generate a two-dimensional histogram of planet periods and radii for the injected parameter vector. This represents our synthetic "observed" dataset. For each trial model, we obtain an equivalent 2D histogram, representing the model prediction. The likelihood of a model is determined by the probability, Pr, that the 2D histograms of the observation and model match, via:

$$\mathcal{L} = \prod_{i,j}^{\text{bins where } N_{ij}^{\mathrm{obs}}>0} \Pr(N_{ij}^{\mathrm{obs}} \mid N_{ij}^{\mathrm{mod}}), \tag{6}$$

where N_{ij}^{obs} gives the number of planets in the i, j period-radius bin for the synthetic observed dataset, and N_{ij}^{mod} gives the same for the model prediction.

When attempting to retrieve results from the actual *Kepler* data, normally the model prediction would be multiplied through by the KeplerPORTs detection efficiencies to determine N^{mod} in Equation (6), so that N^{mod} could be compared directly to the observed sample represented by N^{obs}. In this case, for the correct model, we would expect N^{obs} to be Poisson distributed about N^{mod} within each period-radius bin. However, this approach requires efficiency correction to every likelihood evaluation, which is computationally very expensive.

An alternative approach is to instead apply the KeplerPORTs efficiencies to the observed sample to determine $N_{ij}^{\mathrm{obs},\epsilon} = \sum_k 1/\epsilon_{ijk}$, where the KeplerPORTs efficiency, ϵ_{ijk}, is extracted for every observed transit belonging to period-radius bin ij. Here, index k spans all transit properties over which the detection efficiency is tabulated, other than planet period and radius. $N_{ij}^{\mathrm{obs},\epsilon}$ then provides an estimate of the yield expected for the survey were it to have had perfect detection efficiency, allowing it to be compared directly to the *uncorrected* model prediction, N_{ij}^{mod}. In this case, we only need to apply efficiency correction once, to the observed sample, and not to each

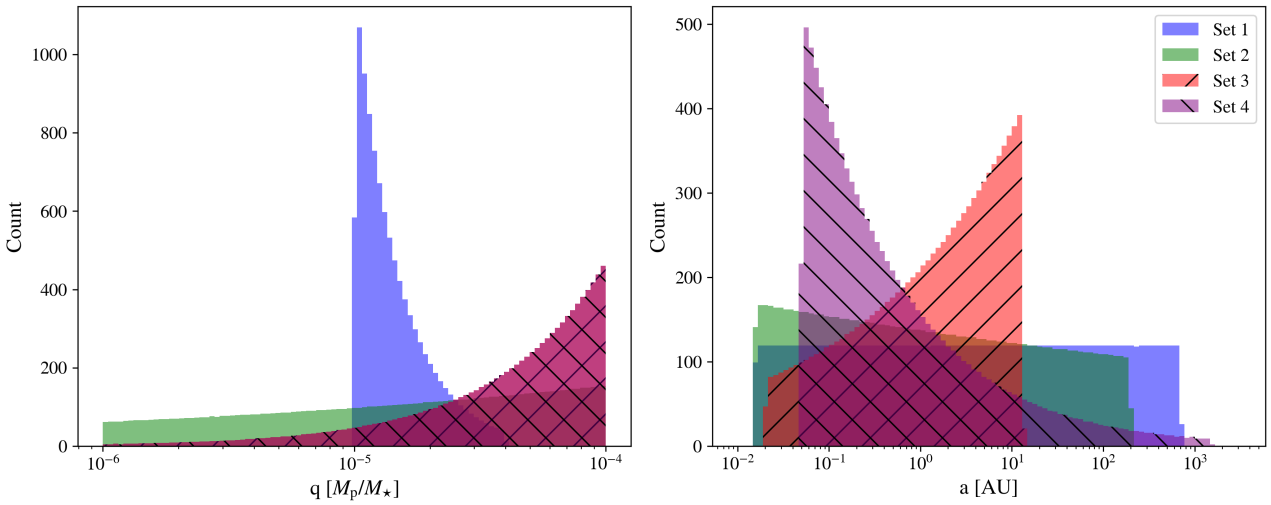

Figure 2. The distribution of planet-to-star mass ratio and semi-major axis in the different injected sets listed in Table 1. Note that, in the left panel, the q distributions for Sets 3 and 4 are identical and so overlay on top of each other. These distributions define the synthetic planetary populations used to test the retrieval framework. The parameter ranges are chosen to span a wide variety of physical scenarios, including compact and extended planetary systems, as well as narrow- and broad-range of mass-ratios of planets. This ensures that the retrieval framework is tested across a broad and realistic spectrum of exoplanetary architectures.

model likelihood evaluation. This is the approach we take, though it should be stressed that it has some limitations.

Firstly, unlike $N^{\rm obs}$, $N^{{\rm obs},\epsilon}$ is non-integer and therefore cannot be Poisson distributed about $N^{\rm mod}$. Instead, the effect of ϵ is to amplify Poisson fluctuations in $N^{\rm obs}$ by a factor $1/\epsilon$. Since the Poisson distribution mean and variance both scale linearly with $N_{\rm mod}$, a $1/\epsilon$ noise amplification in the mean and variance of $N^{\rm obs}$ should not significantly bias the retrieval, such that the distribution of $N^{{\rm obs},\epsilon}$ can be considered *quasi*-Poissonian about an uncorrected $N^{\rm mod}$. Nonetheless, in regions of low detection efficiency (high noise amplification) $N^{{\rm obs},\epsilon}$ becomes numerically unstable. This means that we should restrict likelihood evaluations to regions of planet-host parameter space where ϵ is comfortably above zero. This is straightforward to handle in the retrieval algorithm, meaning our simplified and much faster retrieval approach should still yield a reliable recovery for a survey that has good efficiency over at least some region of the discovery space.

Taking $N_{ij}^{{\rm obs},\epsilon}$ to be quasi-Poisson distributed about $N_{ij}^{\rm mod}$ for every period-radius bin, i, j, our retrieval aims to find the model parameter vector that maximises

$$\log \mathcal{L} = \sum_{i,j}^{N_{ij}^{{\rm obs},\epsilon}>0} N_{ij}^{{\rm obs},\epsilon} \log(N_{ij}^{\rm mod}/N_{ij}^{{\rm obs},\epsilon}) + N_{ij}^{{\rm obs},\epsilon} - N_{ij}^{\rm mod}, \qquad (7)$$

where Stirling's approximation is used for $\log(N^{{\rm obs},\epsilon}!)$. Note that, whist $N^{{\rm obs},\epsilon}!$ is undefined for non-integer $N^{{\rm obs},\epsilon}$, Stirling's approximation is defined for non-integers.

5 TESTING RETRIEVAL ALGORITHMS

With an eye on the very large planet samples that are expected to be delivered by *Roman*, we test four different retrieval algorithms for their efficiency and accuracy. These include: nested sampling; random uniform sampling; a 2-stage machine learning implementation; and differential evolution (Storn & Price 1997).

5.1 Nested sampling

For nested sampling, we employed the widely used Dynesty algorithm (Speagle 2020), though we also did some testing with Ultranest (Buchner 2021), which was found to be consistent with but slower than Dynesty. The main issue we encountered with Dynesty was that after ~1,000,000 calls, the remaining evidence typically converged to a value much larger than the default convergence threshold. We attribute this to the fact that both the model and observed samples are derived from discrete simulation data, meaning that both are intrinsically noisy, which hampers the retrieval of the optimal model. This behaviour was confirmed through testing the retrieval of simpler toy model distributions. The results from nested sampling are presented in Table 1, and the posterior distribution for Set 1 (broad-a, peaked q) is shown in Figure 3. The results demonstrate the method's ability to recover input demographics, though with varying precision across parameters. In all the instances, the injected values lie within or close to the 1-sigma error interval. Some parameters (e.g., n and a_2) show broader distributions, suggesting degeneracies or lower sensitivity in the dataset to those parameters. In Set 1, the possible q range is quite narrow, spanning from 1×10^{-5} to 4×10^{-5}. This limited dynamic range can reduce the sensitivity of the slope parameter n, as the model lacks sufficient leverage to distinguish between subtle variations. Moreover, the transit probability scales inversely with a, so as a_2 increases, there are more planets with a lower transit likelihood. This leads to a loss of sensitivity in the model's ability to constrain parameters associated with distant orbits. This is evident in the broader posterior distribution and reduced density in the scatter plots involving a_2. The limits q_1 and q_2 exhibit some negative correlation with n, with Pearson's coefficients of -0.47 and −0.39 respectively. When n is increased, for a fixed β, it increases the proportion of high-q planets.

5.2 Random uniform sampling

Uniform random sampling represents a rather crude and inefficient retrieval method, compared to the other three approaches we test.

Table 1. Injected (IN) and recovered parameters using four different retrieval methods applied to a 7-parameter planetary demographic model. The methods used are: Dynesty nested sampling (NS); uniform random sampling (UR), accelerated uniform random sampling through a two-stage machine learning model (ML); and differential evolution based retrieval (DE). The last row shows the limits of corresponding priors. The range of β was chosen to yield between 1 and 15 planets per star on average, across the full range of other parameters. Parameters n and y were sampled from linear priors, whilst other parameters were sampled from log-uniform space. The timings are reported while using 60 CPU cores on an AMD Ryzen Threadripper PRO 3995WX, except for NS, where we found that using a single core was faster than multiple cores for our retrieval. For ML retrievals, we mention the timings to train the model (using 60 cores) and for likelihood prediction (using a single core).

Set		n	q_1	q_2	β	a_1	a_2	y	Runtime (s)
1	IN	-2.50	1.00×10^{-5}	4.00×10^{-5}	2.00×10^{-4}	0.01	700.00	0.00	
	UR	$-2.15^{+2.02}_{-0.69}$	$(9.00^{+3.76}_{-8.45}) \times 10^{-6}$	$(2.73^{+379.00}_{-2.26}) \times 10^{-4}$	$(1.64^{+6.30}_{-1.00}) \times 10^{-4}$	$(1.54^{+0.18}_{-0.91}) \times 10^{-2}$	$97.60^{+627.12}_{-78.10}$	$-0.04^{+0.32}_{-0.30}$	35,279
	ML	$-0.21^{+0.80}_{-1.68}$	$(9.37^{+821.31}_{-8.11}) \times 10^{-8}$	$(8.99^{+703.41}_{-8.43}) \times 10^{-4}$	$(1.93^{+0.05}_{-1.77}) \times 10^{-3}$	$(8.89^{+5.33}_{-3.73}) \times 10^{-3}$	$11.63^{+240.38}_{-1.15}$	$0.53^{+0.16}_{-0.74}$	19+37
	NS	$-1.47^{+1.72}_{-1.13}$	$(6.63^{+5.10}_{-6.49}) \times 10^{-6}$	$0.15^{+0.44}_{-0.15}$	$(1.75^{+6.54}_{-1.39}) \times 10^{-4}$	$0.01^{+0.04}_{-0.01}$	$130.70^{+664.11}_{-109.62}$	$-0.17^{+0.58}_{-0.22}$	243,696
	DE	$-2.67^{+1.16}_{-0.25}$	$(1.01^{+0.10}_{-0.54}) \times 10^{-5}$	$(3.18^{+59.39}_{-2.90}) \times 10^{-3}$	$(1.43^{+1.87}_{-0.93}) \times 10^{-4}$	$(1.49^{+0.58}_{-0.39}) \times 10^{-2}$	$11.91^{+245.61}_{-1.32}$	$0.02^{+0.24}_{-0.15}$	4,079
2	IN	0.20	1.00×10^{-6}	1.00×10^{-4}	2.50×10^{-4}	0.01	200.00	-0.05	
	UR	$0.06^{+0.62}_{-0.57}$	$(7.82^{+19.49}_{-7.58}) \times 10^{-7}$	$(1.53^{+13.71}_{-0.86}) \times 10^{-4}$	$(2.53^{+7.89}_{-1.61}) \times 10^{-4}$	$(1.52^{+0.18}_{-0.89}) \times 10^{-2}$	$461.25^{+774.05}_{-431.97}$	$-0.09^{+0.31}_{-0.27}$	35,426
	ML	$-0.01^{+0.62}_{-0.48}$	$(2.84^{+16.64}_{-2.71}) \times 10^{-7}$	$(2.19^{+17.71}_{-1.49}) \times 10^{-4}$	$(1.46^{+0.37}_{-1.23}) \times 10^{-3}$	$(1.21^{+0.34}_{-0.63}) \times 10^{-2}$	$10.06^{+215.10}_{-0.05}$	$0.40^{+0.13}_{-0.60}$	19+37
	NS	$0.12^{+0.48}_{-1.11}$	$(7.31^{+28.03}_{-7.04}) \times 10^{-7}$	$(1.10^{+19.72}_{-0.47}) \times 10^{-4}$	$(2.78^{+8.08}_{-2.26}) \times 10^{-4}$	$0.02^{+0.06}_{-0.01}$	$25.50^{+406.69}_{-12.28}$	$-0.05^{+0.38}_{-0.28}$	360,040
	DE	$0.12^{+0.21}_{-0.62}$	$(1.07^{+1.11}_{-0.72}) \times 10^{-6}$	$(1.30^{+4.15}_{-0.36}) \times 10^{-4}$	$(3.33^{+4.00}_{-1.73}) \times 10^{-4}$	$(1.49^{+0.83}_{-0.40}) \times 10^{-2}$	$180.63^{+483.95}_{-144.17}$	$-0.02^{+0.23}_{-0.14}$	3,454
3	IN	1.00	1.00×10^{-6}	1.00×10^{-4}	5.00×10^{-4}	0.02	13.00	0.25	
	UR	$0.96^{+0.51}_{-1.40}$	$(6.13^{+4467.55}_{-1.55}) \times 10^{-9}$	$(9.90^{+83.35}_{-3.69}) \times 10^{-5}$	$(5.01^{+7.96}_{-3.64}) \times 10^{-4}$	$(2.01^{+0.36}_{-1.33}) \times 10^{-2}$	$26.71^{+405.49}_{-13.28}$	$0.19^{+0.23}_{-0.46}$	35,378
	ML	$0.54^{+0.72}_{-1.12}$	$(2.84^{+576.87}_{-2.10}) \times 10^{-8}$	$(1.68^{+14.99}_{-0.92}) \times 10^{-4}$	$(1.92^{+0.05}_{-1.68}) \times 10^{-3}$	$(7.00^{+9.91}_{-2.22}) \times 10^{-3}$	$11.27^{+268.80}_{-0.89}$	$0.57^{+0.15}_{-0.78}$	19+37
	NS	$0.90^{+0.39}_{-1.35}$	$(1.16^{+6.60}_{-1.13}) \times 10^{-6}$	$(1.21^{+10.42}_{-0.37}) \times 10^{-4}$	$(1.27^{+0.46}_{-1.05}) \times 10^{-3}$	$(5.66^{+13.35}_{-1.19}) \times 10^{-3}$	$12.34^{+245.85}_{-1.66}$	$0.52^{+0.47}_{-0.57}$	303,082
	DE	$0.96^{+0.21}_{-0.71}$	$(5.30^{+22.91}_{-5.04}) \times 10^{-7}$	$(1.11^{+2.18}_{-0.25}) \times 10^{-4}$	$(7.32^{+4.83}_{-5.05}) \times 10^{-4}$	$0.02^{+0.02}_{-0.01}$	$22.48^{+142.05}_{-8.61}$	$0.27^{+0.18}_{-0.22}$	3,294
4	IN	1.00	1.00×10^{-6}	1.00×10^{-4}	5.00×10^{-4}	0.05	1500.00	-0.40	
	UR	$1.18^{+0.38}_{-1.60}$	$(5.28^{+694.43}_{-4.36}) \times 10^{-8}$	$(1.03^{+5.72}_{-0.38}) \times 10^{-4}$	$(5.53^{+8.14}_{-3.13}) \times 10^{-4}$	$0.05^{+0.01}_{-0.04}$	$45.63^{+489.97}_{-30.07}$	$-0.44^{+0.46}_{-0.04}$	35,666
	ML	$0.03^{+0.98}_{-1.13}$	$(3.06^{+557.59}_{-2.31}) \times 10^{-8}$	$(2.93^{+40.51}_{-2.06}) \times 10^{-4}$	$(1.89^{+0.07}_{-1.43}) \times 10^{-3}$	$0.03^{+0.02}_{-0.02}$	$10.10^{+263.09}_{-0.07}$	$0.54^{+0.14}_{-0.80}$	19+37
	NS	$1.07^{+0.25}_{-0.79}$	$(5.32^{+158.02}_{-4.43}) \times 10^{-8}$	$(9.67^{+10.12}_{-1.86}) \times 10^{-5}$	$(5.13^{+5.54}_{-3.60}) \times 10^{-4}$	$(4.99^{+0.31}_{-3.37}) \times 10^{-2}$	$1306.09^{+430.15}_{-1264.71}$	$-0.36^{+0.37}_{-0.09}$	446,420
	DE	$0.97^{+0.22}_{-0.71}$	$(8.86^{+25.98}_{-8.63}) \times 10^{-7}$	$(1.03^{+1.68}_{-0.20}) \times 10^{-4}$	$(5.37^{+2.69}_{-3.11}) \times 10^{-4}$	$0.05^{+0.01}_{-0.02}$	$68.47^{+428.94}_{-44.94}$	$-0.35^{+0.33}_{-0.08}$	2,883
Priors		$[-3.2, 1.8]$	$[4 \times 10^{-9}, 2 \times 10^{-5}]$	$[3 \times 10^{-5}, 1.23]$	$[7 \times 10^{-6}, 2 \times 10^{-3}]$	[0.004, 0.5]	[10, 2000]	$[-0.5, 1.5]$	

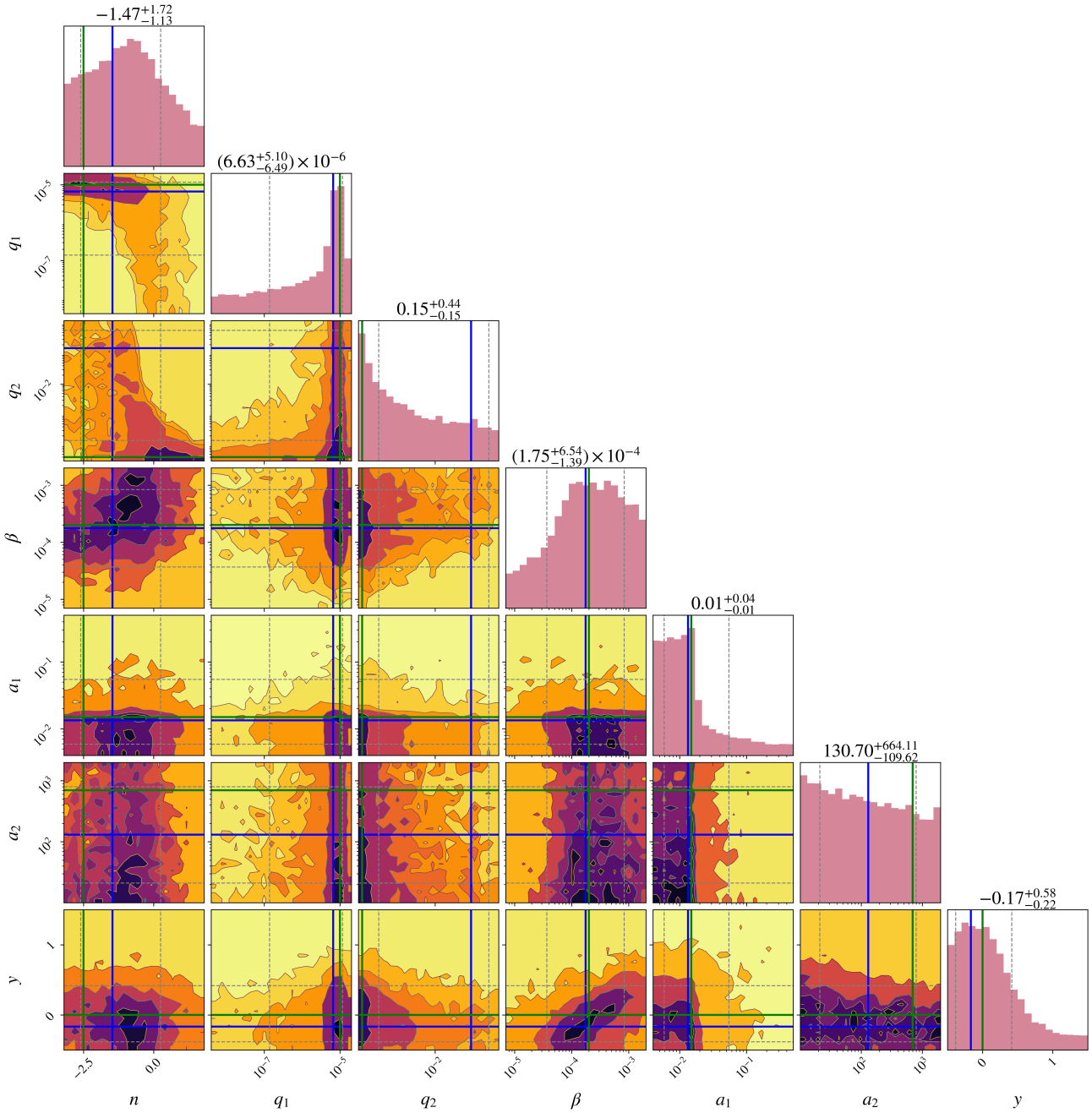

Figure 3. Cornerplot showing 1-D marginals (diagonals) and 2-D joint posteriors (off-diagonals) of samples of parameters while recovering parameters for Set 1 (broad a, peaked q) using Dynesty nested sampling (NS). The multi-dimensional best-fit value is overplotted in blue colour, while the injected parameters are overplotted in green colour. The dashed lines show the upper and lower errors on the best-fit value.

However, it forms the basis for one of the other methods we employ, namely the two-stage machine learning approach. For this reason, its inclusion provides a good benchmark for the speed and accuracy of the other methods.

We generated 10,000,000 parameter vectors sampled randomly from the uniform distribution of priors shown in Table 1. Then we calculated the corresponding log-likelihood values. As the samples are distributed uniformly, the density of points in the posterior distribution is also uniform, so they must be weighted to obtain a meaningful posterior distribution. We calculate the weights, w, of each sample i as

$$w_i \propto \exp\left(\alpha\left(\log\mathcal{L}_i - \max(\log\mathcal{L})\right)\right). \quad (8)$$

We applied a scale factor $\alpha = 10^{-4}$ to suppress the dynamic range in $\mathcal{L}$ and mitigate against numerical underflow. We present the results from these tests in Table 1, and show the posterior plots for Set 1 in Figure 4. Similar to the retrieval with Dynesty in Section 5.1, we see that the injected values are typically within a 1-sigma interval of the retrieved values, and the posterior distribution is analogous to that found from nested sampling. It presents a simple and robust solution for benchmarking, but it is not scalable.

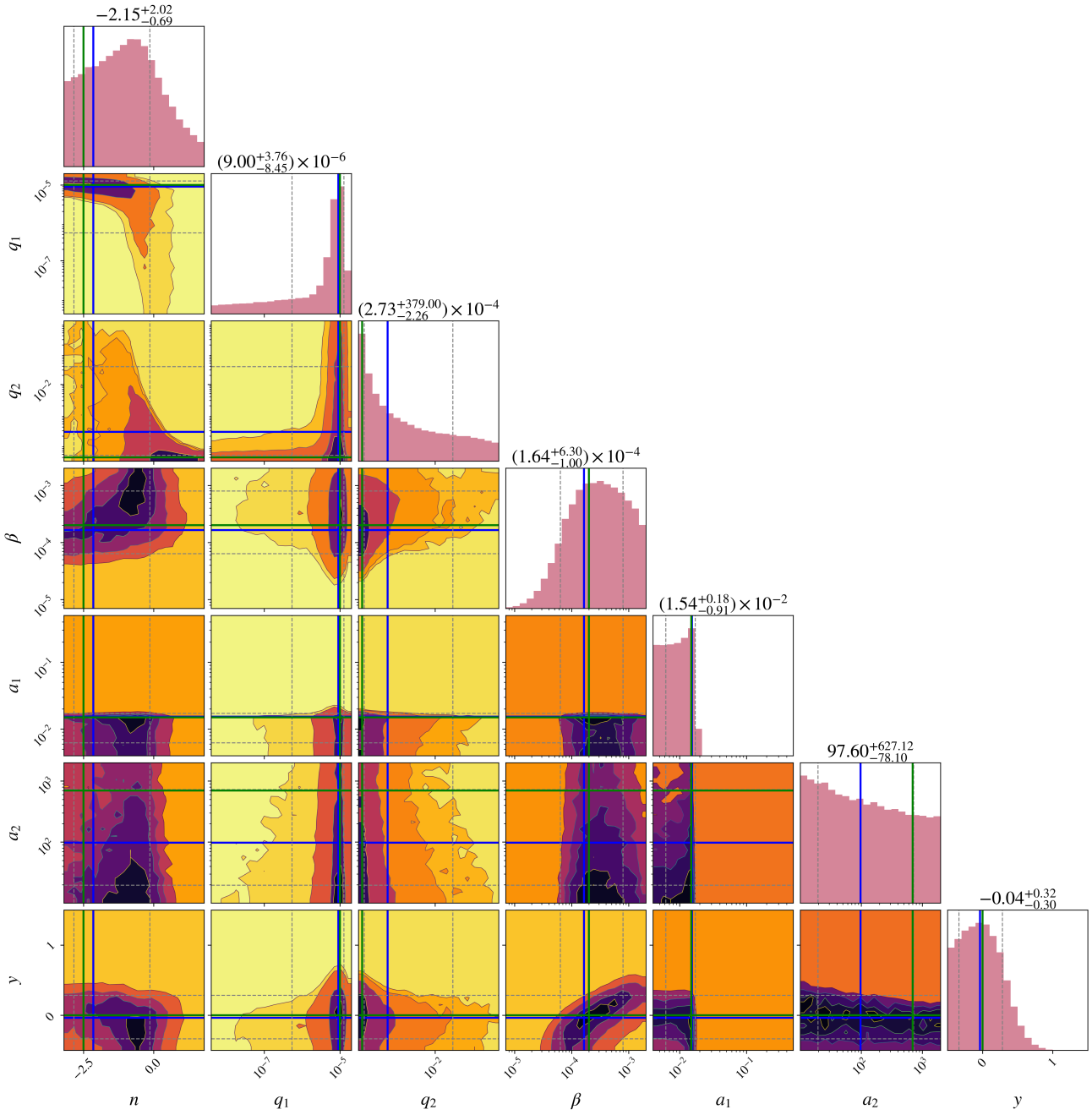

Figure 4. Cornerplot showing 1-D marginals (diagonals) and 2-D joint posteriors (off-diagonals) of samples of parameters while recovering parameters for Set 1 using the random uniform sampling method (UR). The multi-dimensional best-fit value is overplotted in blue colour, while the injected parameters are overplotted in green colour. The dashed grey lines show the upper-error and the lower-error on the best-fit value. The samples were uniformly distributed over the displayed parameter space, hence the density corresponds to the weights calculated in Section 5.2.

5.3 Two-stage machine learning

In order to accelerate our likelihood evaluations, we prepared a 2-stage machine learning model to predict the likelihoods of parameter vectors. For a test dataset of exoplanets, we calculated the likelihoods corresponding to 250,000 random uniform samples for parameter vectors, with diminishing returns found for larger samples. Additional features were included in each sample through second-order combinations of all parameters in the samples. Around 46% of the samples resulted in no exoplanets and so are penalised in the likelihood calculation. As these samples are not highly localised in parameter space, and given their potential to bias the prediction model, we create a Random Forest Classifier as a first stage to identify such samples. The second stage involves a Light Gradient Boost regression model (Zhang et al. 2017) to predict the likelihood of samples. For this stage, the likelihoods were standardised using a Z-score to avoid feature dominance simply through having a larger numerical range. We show the results from our prediction model in Figure 5. On the test dataset, the classifier model had an F1 score of 0.98, and the R^2 value of the regressor model was 0.99. The R^2 of the combined 2-stage solution was 0.94. Using this 2-stage machine

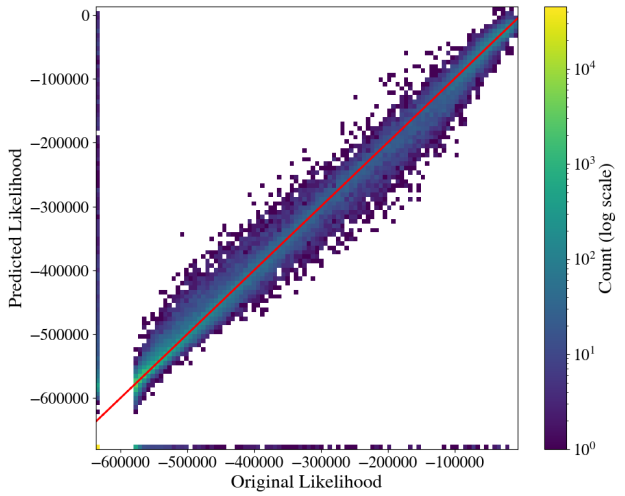

Figure 5. Figure showing the comparison of predicted likelihood using the 2-stage machine learning model for Set 1. The column of points on the left-most edge (false positives) of the original likelihood axis and a row of points on the bottom-most edge (false negatives) of the predicted likelihood axis correspond to 2.2% of all data points. These points originate from misclassification in the first stage of the machine learning model.

learning model, we proceeded to generate log-likelihood predictions for 10,000,000 random samples. The weights of the sampled points are assigned using Equation (8). The results from these analyses are shown in Table 1, and the posterior plot for Set 1 is presented in Figure 6. The cornerplot shows that this method is capable of replicating key features of the posterior distribution seen from nested sampling retrieval, including the joint distributions between n and q_1, and n and q_2. However, when comparing with the retrievals from NS and UR, the marginalised distribution for β shows a preference for higher values. This method is significantly faster than all other methods investigated, but presents a risk of producing unphysical high-likelihood islands.

5.4 Differential evolution to retrieve parameters

The final retrieval method we consider is differential evolution (DE) (Storn & Price 1997; Virtanen et al. 2020). DE is a stochastic method to find the global minimum of a function. It is designed to be efficient even when both the data and models are discrete, and is easy to highly parallelise. DE starts with an initial population of candidate parameter vectors and, for every candidate, creates a new trial candidate vector based on mixed and mutated combinations of existing vectors. A trial candidate vector succeeds its predecessor if it has a higher likelihood. The algorithm continues with a new generation of trial vectors and halts only when

$$\sigma(E) \leq A_{\rm tol} + R_{\rm tol} \cdot |\mu(E)|, \tag{9}$$

where $E = -\mathcal{L}$ are the "energies" of the samples from the current generation of candidate parameter vectors, and $\sigma(E)$ and $\mu(E)$ are the standard deviation and mean of E. $R_{\rm tol}$ and $A_{\rm tol}$ are relative and absolute tolerances, both of which are set to their default values of 0.01 and 0, respectively. The idea is that through mixing and mutation of parameter vectors, each generation of trial candidate samples will have a tendency to migrate towards an increasingly localised parameter region centred around the model of highest likelihood. Equation (9) uses a measure of the tightness of this localisation as the basis for the halting condition.

Given the smooth forms of our planet's demographic distributions, we choose the Sobol sampling method (Sobol' 1967) to seed the initial generation of parameters. Sobol sampling is efficient at maximising the coverage of the parameter space, whilst avoiding clustered sampling encountered in random sampling, which would only decrease the convergence efficiency of the DE method. We set the initial population to be 8192.

To generate a candidate sample in each iteration, we used the `randtobest1bin` method (Qiang & Mitchell 2014), which has been shown to perform better than other methods across diverse problem types and distributions (Jeyakumar & Shanmugavelayutham 2011). For each candidate parameter array, $\mathbf{b}$, in the parent population, this strategy generates an intermediate candidate parameter array as

$$\mathbf{b}' = \mathbf{x}_{r_0} + F \cdot \left(\mathbf{x}_0 - \mathbf{x}_{r_0} + \mathbf{x}_{r_1} - \mathbf{x}_{r_2}\right), \tag{10}$$

where $\mathbf{x}_{r_0}$, $\mathbf{x}_{r_1}$, and $\mathbf{x}_{r_2}$ are 3 random candidate parameter arrays drawn from the parent population, and $\mathbf{x}_0$ is the parameter array with the best likelihood from the current generation of candidate vectors. F is the mutation parameter, which is sampled randomly for each generation from the interval $[0.5, 1)$. The parameter array for a new trial candidate, $\mathbf{b}_t$, is assigned by randomly selecting its elements from among those used for $\mathbf{b}'$ or $\mathbf{b}$. It is ensured that at least one parameter is selected from $\mathbf{b}'$. This is implemented by randomly choosing a parameter on each instance and then using the corresponding value from $\mathbf{b}'$ for the trial candidate. $\mathbf{b}$ is discarded in favour of $\mathbf{b}_t$ if $\mathbf{b}_t$ has a higher likelihood. In this way, a new successor generation of candidate parameter vectors is created.

The results of this method are summarised in Table 1, and the comparison of population for Set 1 is shown in Figure 7. The cornerplot for Set 1 is shown in Figure 8. The posterior distributions from DE are more clearly delineated than in other retrievals, but they generally show good recovery of the injected parameters, which is consistent with the width of the posteriors. DE presents the best tradeoff between speed and accuracy of retrievals, which is efficiently parallelisable and scalable.

6 PARAMETER RETRIEVAL USING THE REAL KEPLER DATASET

Having verified the performance of the recovery using simulated data, we now recover parameters from the exoplanet sample observed by *Kepler* itself.

As mentioned in Section (3), comparison of simulated and real *Kepler* star samples required using a rescaling factor to account for the discrepancy between the number of stars in the BGM and the number chosen for the *Kepler* Input Catalogue. *Kepler* observed 150,000 stars, whereas 59,734 stars from the BGM sample qualified as representative of the *Kepler* input catalogue using the similarity factor S in Equation (5) with the criterion $S \leq 0.15$. Our simulated yields are accordingly scaled up by a factor 2.51 to represent the *Kepler* expected yield for a given model. We use the differential evolution (DE) retrieval to obtain our best-fit parameters, as shown in Table 2. The corresponding posterior distributions are shown in Figure 9, and we show the histograms of exoplanet population generated using these values in Figure 10. The recovered planet distribution from the 7-parameter DE model successfully reproduces some key features of the underlying distribution inferred from the *Kepler* population, including the Fulton gap (Fulton et al. 2017) near 2 $R_\oplus$, and the prominent peak in the semi-major axis distribution around $P \sim 10$

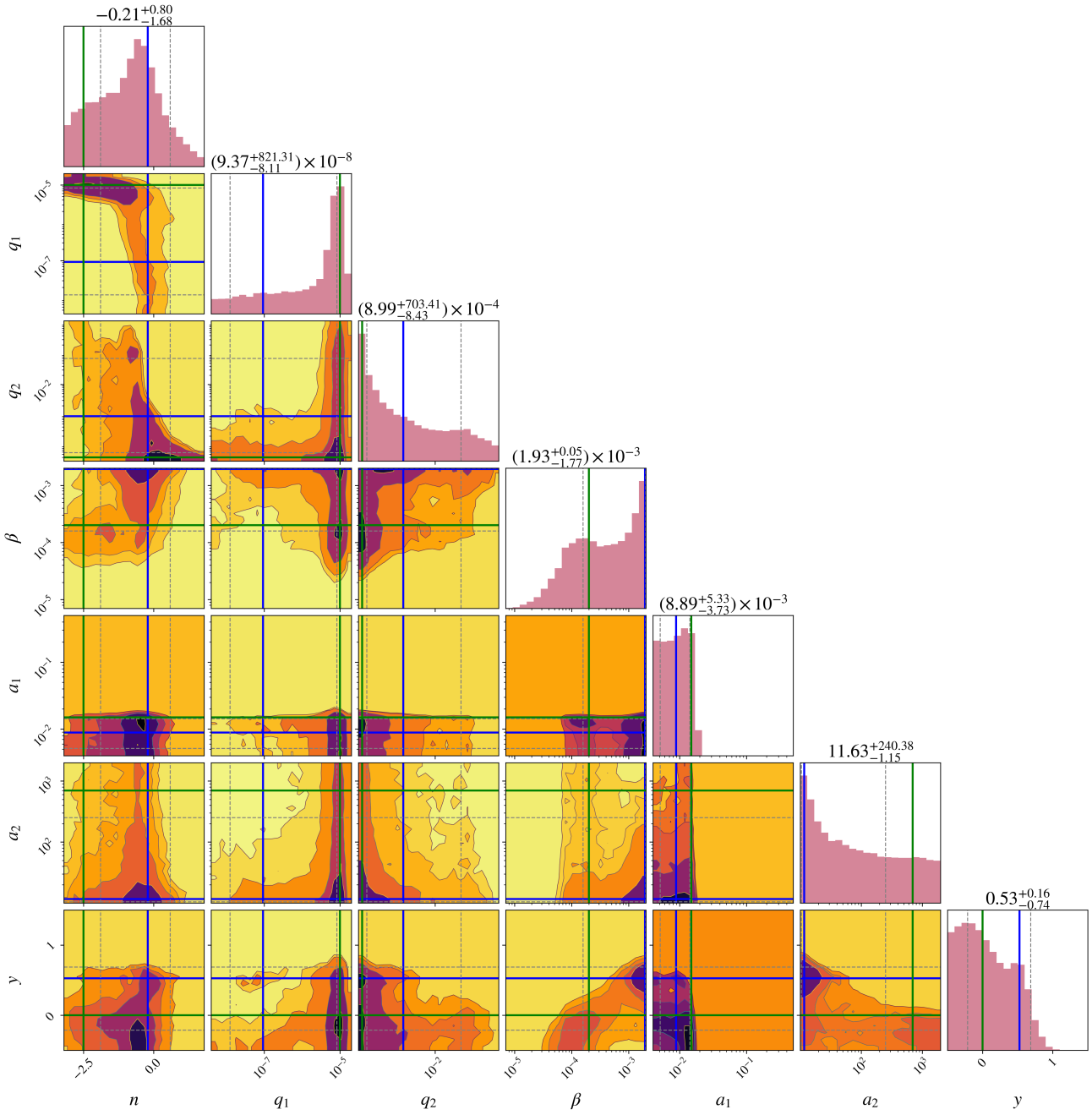

Figure 6. Cornerplot showing 1-D marginals (diagonals) and 2-D joint posteriors (off-diagonals) of samples of parameters while recovering parameters for Set 1, using the 2-stage machine learning model to predict log-likelihoods. The multi-dimensional best-fit value is overplotted in blue colour, while the injected parameters are overplotted in green colour. The dashed grey lines show the upper-error and the lower-error on the best-fit value. The samples were uniformly distributed over the displayed parameter space, hence the density corresponds to the weights calculated in Section 5.2.

days. These agreements indicate that the model captures the dominant demographic trends. However, when the full predicted distribution considering all the bins is examined, the model also generates a significant population of larger planets. Their absence in the observed sample suggests serious deficiencies in this model and also in others that we examine in this paper.

To mitigate the shortcomings in the 7-parameter model, we also consider a 9-parameter model and an 11-parameter model. In the 7-parameter model, both q and a are described by single power-law distributions. While in the 9-parameter case, q is allowed to have a broken power law distribution, while a follows a single power law. Finally, in the 11-parameter model, both q and a have broken power-law distributions. The motivation for adopting a broken power-law in semi-major axis arises from the discrepancy between the model and observations: the recovered population exhibits a relatively flat one-dimensional a-distribution when all bins are considered, whereas the *Kepler* data show a pronounced peak, indicating a possible change in slope across different orbital regimes.

The priors and results for the 9-parameter and 11-parameter models are provided in Table 2, with the corresponding corner plots

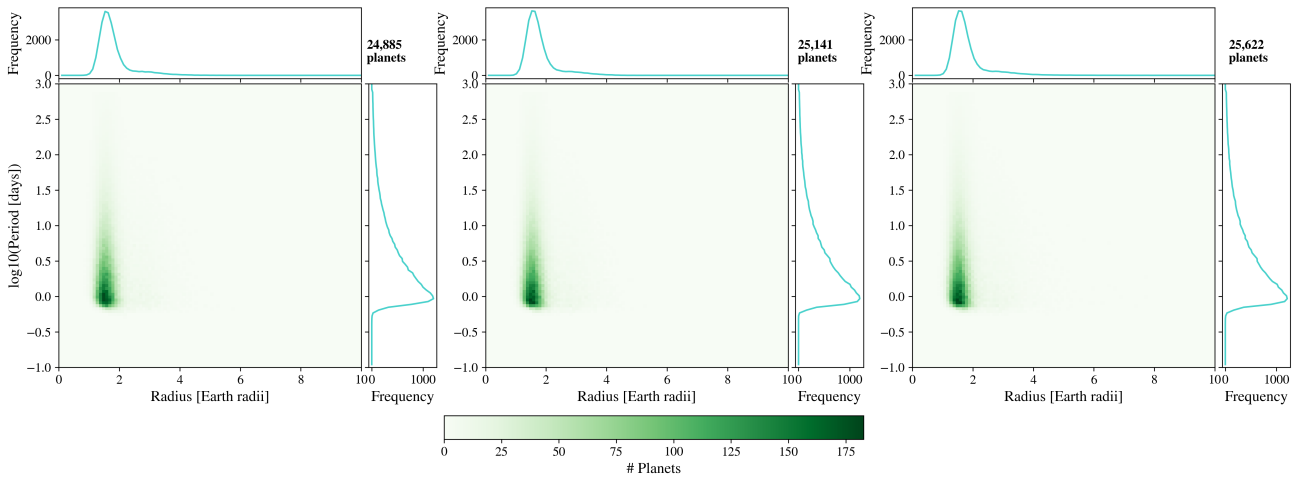

Figure 7. The figure shows the distribution of the injected dataset of exoplanets for Set 1, on the left and the corresponding recovered planets using differential evolution on the middle and right. In the middle panel, we only plot the bins where the injected distribution had non-zero planets. In the rightmost panel, we plot all the bins from the recovered population. The colourmap shows the number of planets on the period-radius axes. The axes are accompanied by 1-D histograms of radius and period separately. In the top right corner of each 2-D histogram, the total number of the total planets is written.

shown in Figures 11 and 12. A comparison of exoplanet population is shown in Figures 13 and 14. The 9-parameter model shows no improvements over the 7-parameter model, while the 11-parameter model shows a better agreement in the comparability of semi-major axis histograms.

We calculate the Bayesian Information Criterion (BIC) (Schwarz 1978; Astropy Collaboration et al. 2022) to compare the 7, 9, and 11 parameter models. There were 1261 bins which had non-zero planets in *Kepler* observation. This value is used as the number of samples for BIC calculation. The BIC came out to be 4830.98, 4734.68, and 4643.19, for 7, 9, and 11 parameter models, respectively. This shows that the 11-parameter model is preferred.

6.1 Modelling limitations

Upon observing the unmasked 2D histograms of predicted *Kepler* exoplanet distribution from the 7,9, and 11 parameter model, we notice that the underlying model predicts a substantial number of planets with $P = 10$ days and $R_p > 2R_\oplus$. Such planets should be easily detected by *Kepler* for many hosts, so this points to a clear deficiency of the model in not accounting for possible correlations between a and q, as may arise due to Hill instability, for example.

For Hill stability, we have (Obertas et al. 2017),

$$a_2 > a_1(1 + 2 \times 3^{1/6} \times (q_1 + q_2)^{1/3}), \tag{11}$$

where the subscripts 1 and 2 denote the inner and outer planet. We consider a simplified single planet proxy of this inequality,

$$a > k \times q_1^z, \tag{12}$$

or,

$$a_{min} = k \times q_1^z, \tag{13}$$

where k and z are parameters to fit, and

$$a \in \text{LogUniform}(a_{\text{min}}, a_2). \tag{14}$$

Using this $a - q$ correlated power-law (CPL) in our 7-parameter model from 5.4, instead of fitting for a_1, we fit for k and z. To keep the model simpler, we consider the semi-major distribution to be log-uniform and remove parameter y. Subsequently, we find the best-fit values for k and z. The BIC for this retrieval was 5789.19. The resulting distributions are shown in Figure 15, and the corresponding best-fit parameters are listed in Table 3. The corresponding posterior distribution is shown in Figure 16, with comparison histogram shown in Figure 17. The distribution of recovered planets using this model shows improved agreement when comparing the 1D histograms, even after considering all the bins. The existence of a few planets with an orbital period ~10 days and $\geq 4\ R_\oplus$ suggests that even this model requires additional physics to suppress larger planets at low periods. The purpose of this paper is to present the TAED framework; we leave the exploration of better-fitting and more complex models to future studies.

7 CONCLUSIONS

The NASA *Roman* Galactic Bulge Time Domain Survey is a five-year high-cadence survey to be undertaken by *Roman* from early 2027. It will hugely expand both the number and locations of known exoplanets. Up to 200,000 exoplanets are expected to be discovered through the transit and microlensing detection techniques, and these will be located towards, in, and beyond the Galactic bulge. *Roman* will access large numbers of hot and cool exoplanets around a representative mix of stellar hosts for the first time. *Roman*'s sample will present a treasure trove for studies of Galactic-wide exoplanet demography and will provide stringent new tests for planet formation models.

With this opportunity comes a new challenge: to develop exoplanet demographic analysis methods that can combine the full set of information available from large microlensing and transit samples, given the differences in properties and Galactic locations of the planets and their hosts within each sample. To address this challenge, we introduce in this paper a *technique-agnostic exoplanet demography* (TAED) framework that is able to make consistent forecasts for multiple detection methods that are based on spatio-kinematic properties. This includes transits, microlensing, radial velocity and astrometric detection methods.

This paper presents the design and first test of the TAED framework, focusing on simulations of a *Kepler*-like transit observable.

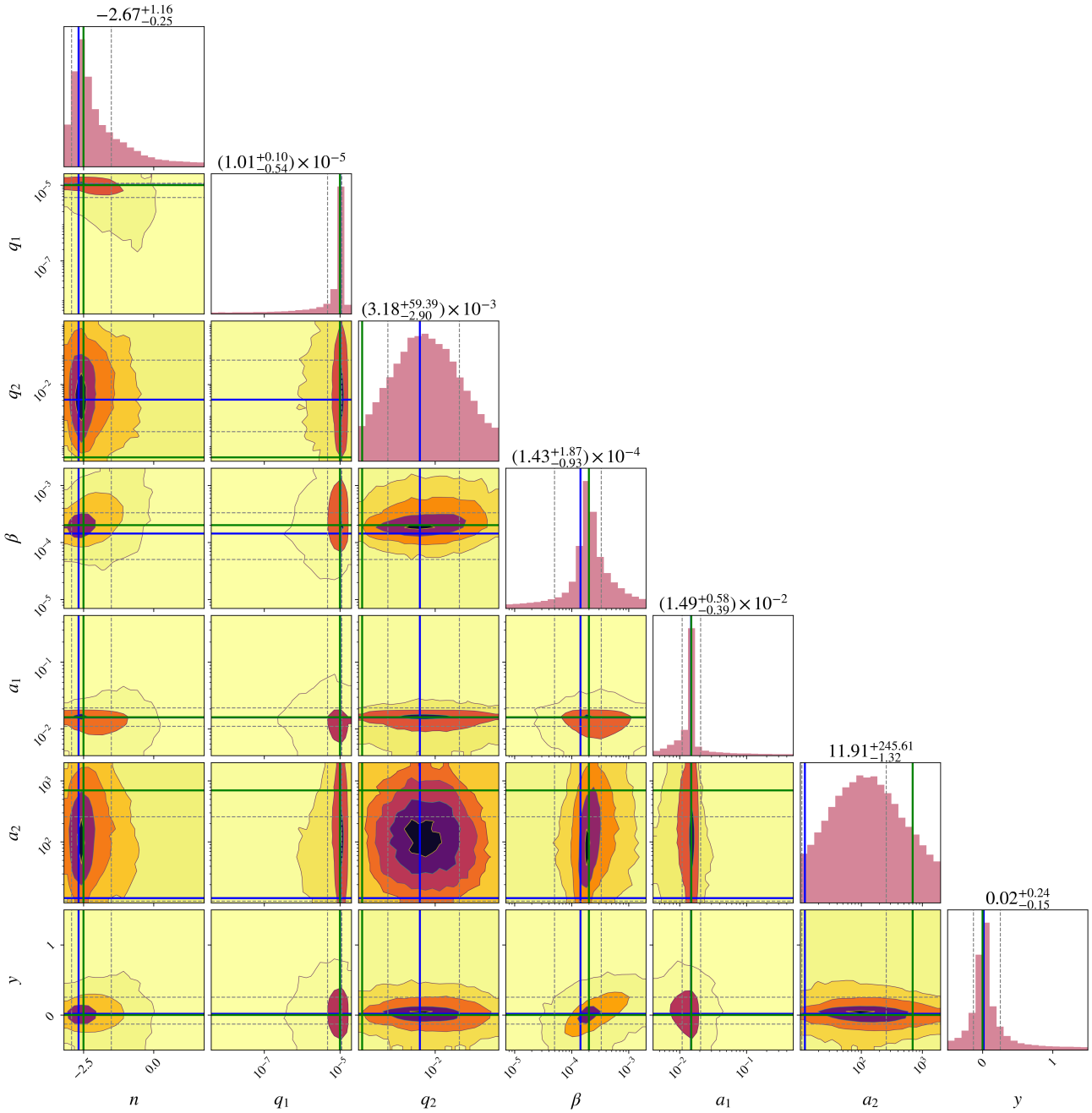

Figure 8. Cornerplot showing 1-D marginals (diagonals) and 2-D joint posteriors (off-diagonals) of samples of parameters while recovering parameters for Set 1 using differential evolution. The multi-dimensional best-fit value is overplotted in blue colour, while the injected parameters are overplotted in green colour. The dashed grey lines show the upper-error and the lower-error on the best-fit value.

We have examined the accuracy and scalability of the recovery of exoplanet demographic parameters from simulated transit datasets where the truth is known. We have validated the TAED framework and have demonstrated a workable retrieval method, based on differential evolution, which can recover the underlying exoplanet population and can scale to large datasets of the kind that *Roman* is expected to deliver. Differential evolution was found to be preferred over several other retrieval methods we investigated, including nested sampling and machine learning, based on its combination of speed and accuracy of recovery.

For our initial test of the TAED framework, we considered several fairly simple toy exoplanet demographic models with 7 to 11 free parameters. In our tests, the TAED framework was successful in retrieving key population features and parameters for our injected distributions. We also applied the TAED retrieval to the *Kepler* DR25 dataset itself, even though the assumed models have known simplicities and limitations. The recovered models successfully reproduced major known demographic features, including the Fulton gap, and the peak in orbital period distribution around 10 days. However, due to their simplicity, the models typically over-predicted exoplanets in regions of parameter space where *Kepler* found few or none. This does not point to a deficiency in the TAED framework itself. Rather,

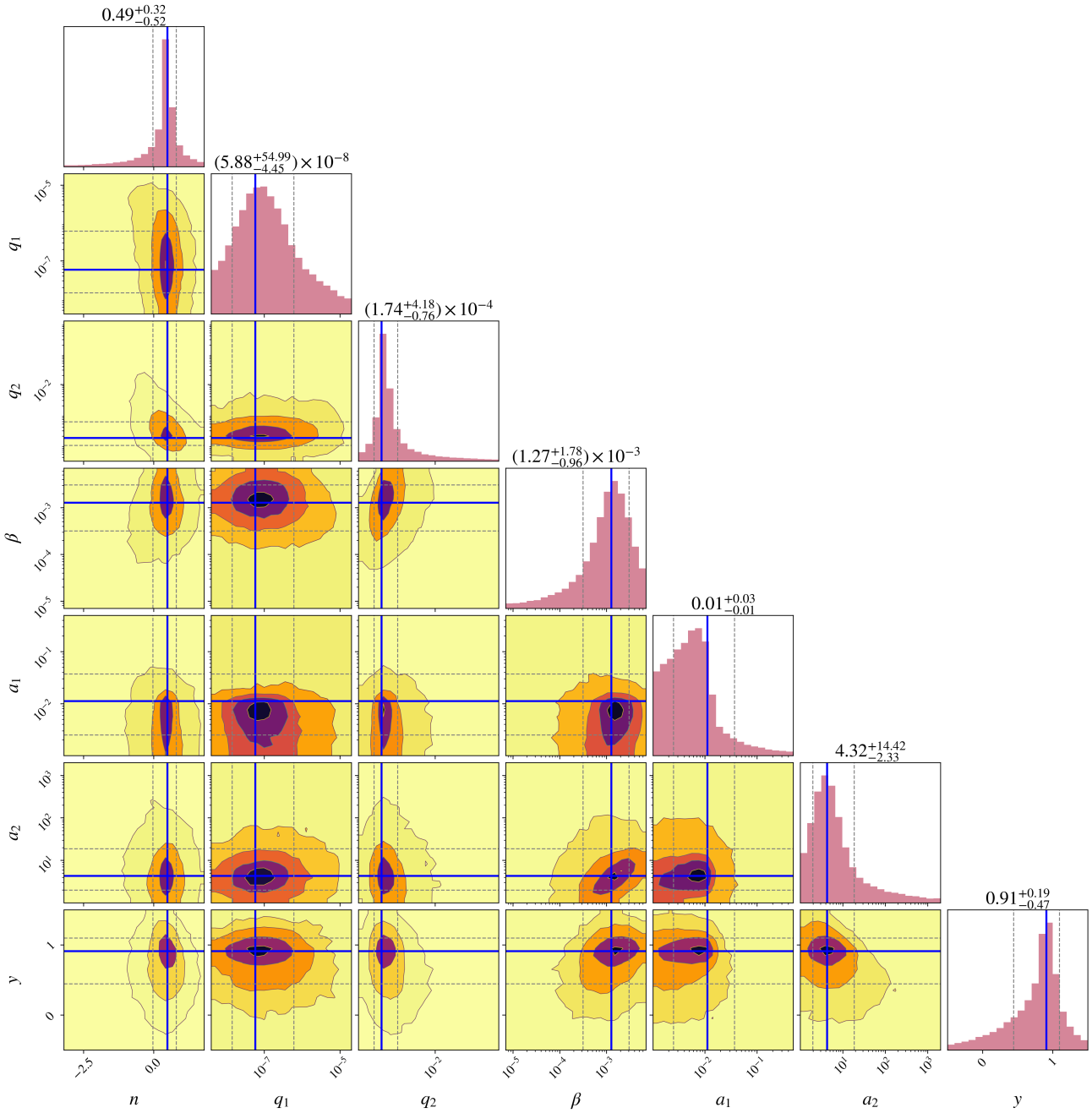

Figure 9. Cornerplot showing 1-D marginals (diagonals) and 2-D joint posteriors (off-diagonals) of samples while fitting for *Kepler* population using differential evolution. The multi-dimensional best-fit value is overplotted in blue colour. The dashed grey lines show the upper-error and the lower-error on the best-fit value.

it highlights how the framework can be used to evolve towards more complex exoplanet models that include additional covariances between planet parameters and between planet and host properties that facilitate a better match to the observations. It was not the aim of this current paper to undertake such a study, though it is clearly a goal for future study.

The logical next step will be to test TAED on simulated microlensing samples, and subsequently on combined fully-scaled transit and microlensing samples, based on the characteristics of the expected *Roman* survey sensitivity. The prospect of deploying TAED on combined samples is particularly exciting. Microlensing and transit methods have different sensitivity to planet size/mass, orbital period/radius, orbit eccentricity and, when considering multi-planet systems, mutual inclination. Using the TAED framework, we expect demographic parameter retrieval from joint microlensing and transit samples to provide rather more powerful constraints on some of these distributions than would arise from studies of the separate samples. This strongly motivates a joint approach to *Roman* microlensing and transit datasets for demographic studies.

Lastly, given that transit and microlensing samples represent hot and cold exoplanets that typically straddle either side of the stellar habitable zone, a joint demographic analysis of the *Roman* transit and microlensing samples likely offers one of the most reliable calibrations of the total Galactic occurrence of habitable zone exoplanets.

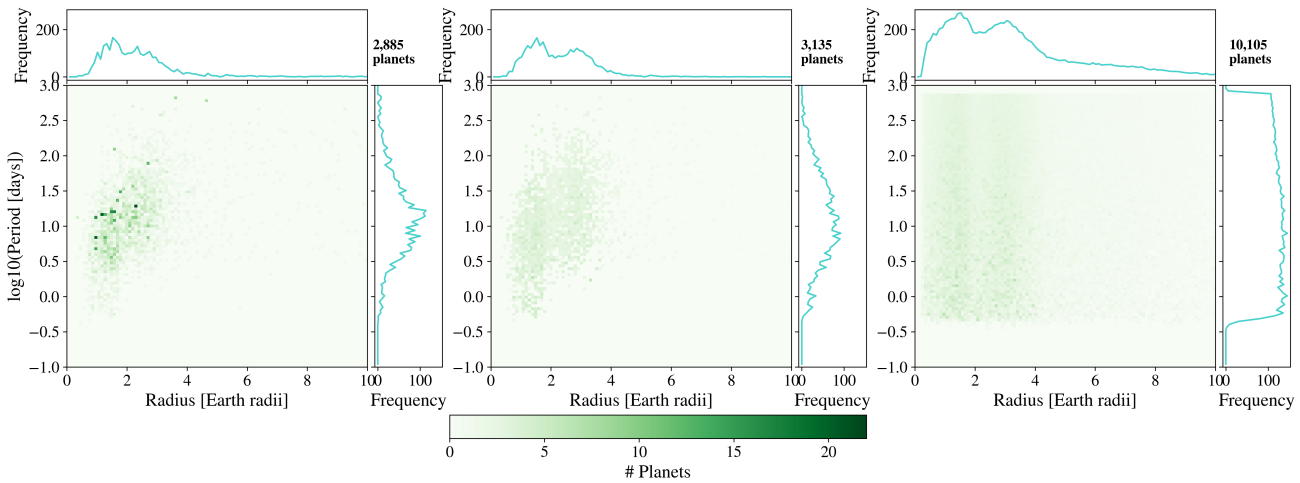

Figure 10. The figure shows the distribution of the projected *Kepler* observations (KeplerPORTs-corrected counts per bin) on the left and the corresponding recovered planets using the 7-parameter model on the middle and right. In the middle panel, we only plot the bins where the projected distribution had non-zero planets. In the rightmost panel, we plot all the bins from the recovered population. The colourmap shows the number of planets on the period-radius axes. The axes are accompanied by 1-D histograms of radius and period separately. In the top right corner of each 2-D histogram, the total number of the total planets is written.

Table 2. The table shows the recovered parameters for *Kepler* observations when using 7-parameter, 9-parameter, and 11-parameter model fitted using differential evolution method. The priors for the parameters are uniform for n_1, n_2, y_1, and y_2, while log-uniform for the others.

Parameter	Prior	7-param	9-param	11-param
n_1 (n)	[-3.2, 1.2]	$0.49^{+0.32}_{-0.52}$	$0.66^{+0.24}_{-0.70}$	$0.66^{+0.22}_{-0.55}$
n_2	[-1.2, 1.8]	-	$-0.39^{+0.67}_{-0.35}$	$-0.40^{+0.58}_{-0.36}$
q_1	[4e-09, 2e-05]	$(5.88^{+54.99}_{-4.45}) \times 10^{-8}$	$(9.64^{+213.50}_{-4.07}) \times 10^{-9}$	$(2.00^{+22.00}_{-1.22}) \times 10^{-8}$
q_2	[3e-05, 1.23]	$(1.74^{+4.18}_{-0.76}) \times 10^{-4}$	$(4.12^{+13.88}_{-2.37}) \times 10^{-4}$	$(3.59^{+20.22}_{-3.14}) \times 10^{-3}$
q_{br}	[1e-07, 0.0001]	-	$(4.02^{+2.60}_{-3.55}) \times 10^{-5}$	$(3.46^{+2.62}_{-2.95}) \times 10^{-5}$
β	[7e-06, 0.007]	$(1.27^{+1.78}_{-0.96}) \times 10^{-3}$	$(1.93^{+2.06}_{-1.51}) \times 10^{-3}$	$(3.04^{+1.87}_{-2.38}) \times 10^{-3}$
a_1	[0.001, 0.5]	$0.01^{+0.03}_{-0.01}$	$(3.04^{+9.57}_{-1.45}) \times 10^{-3}$	$(5.64^{+8.48}_{-3.41}) \times 10^{-3}$
a_2	[1, 2000]	$4.32^{+14.42}_{-2.33}$	$4.43^{+8.79}_{-2.43}$	$2.62^{+10.91}_{-1.07}$
y_1 (y)	[-0.5, 2.5]	$0.91^{+0.19}_{-0.47}$	$-0.02^{+0.42}_{-0.20}$	$1.38^{+0.32}_{-0.62}$
y_2	[-0.5, 1.5]	-	-	$0.73^{+0.25}_{-0.53}$
a_{br}	[0.04, 40]	-	-	$0.09^{+0.89}_{-0.03}$
Runtime (s)		5,899	16,305	25,755

Table 3. The table shows the recovered parameters for *Kepler* observations when using CPL model which incorporates correlated a and q power-law relations, following equation 13. The priors for the parameters are uniform for n and z, while log-uniform for the others.

Parameter	Prior	CPL
n	[-3.2, 1.8]	$0.82^{+0.18}_{-0.60}$
q_1	[4e-09, 2e-05]	$(1.82^{+16.21}_{-1.08}) \times 10^{-8}$
q_2	[3e-05, 1.23]	$(1.76^{+1.81}_{-0.55}) \times 10^{-4}$
β	[7e-06, 0.002]	$(4.85^{+4.49}_{-2.98}) \times 10^{-4}$
k	[0.01, 1000]	$6.70^{+76.62}_{-6.32}$
z	[-2, 5]	$0.51^{+0.93}_{-0.40}$
a_2	[10, 2000]	$99.28^{+332.50}_{-71.05}$
Runtime (s)		4,076

This would represent the first true measurement of the third term in the Drake Equation.

ACKNOWLEDGEMENTS

This paper includes data collected by the *Kepler* mission, which is funded by the NASA Discovery Programme. AP is supported by a PhD studentship from the United Kingdom's Science and Technology Facilities Council (STFC). Language refinement in parts of this manuscript was supported by Copilot.

REFERENCES

Astropy Collaboration et al., 2022, ApJ, 935, 167
Brown T. M., Latham D. W., Everett M. E., Esquerdo G. A., 2011, AJ, 142, 112
Buchner J., 2021, The Journal of Open Source Software, 6, 3001

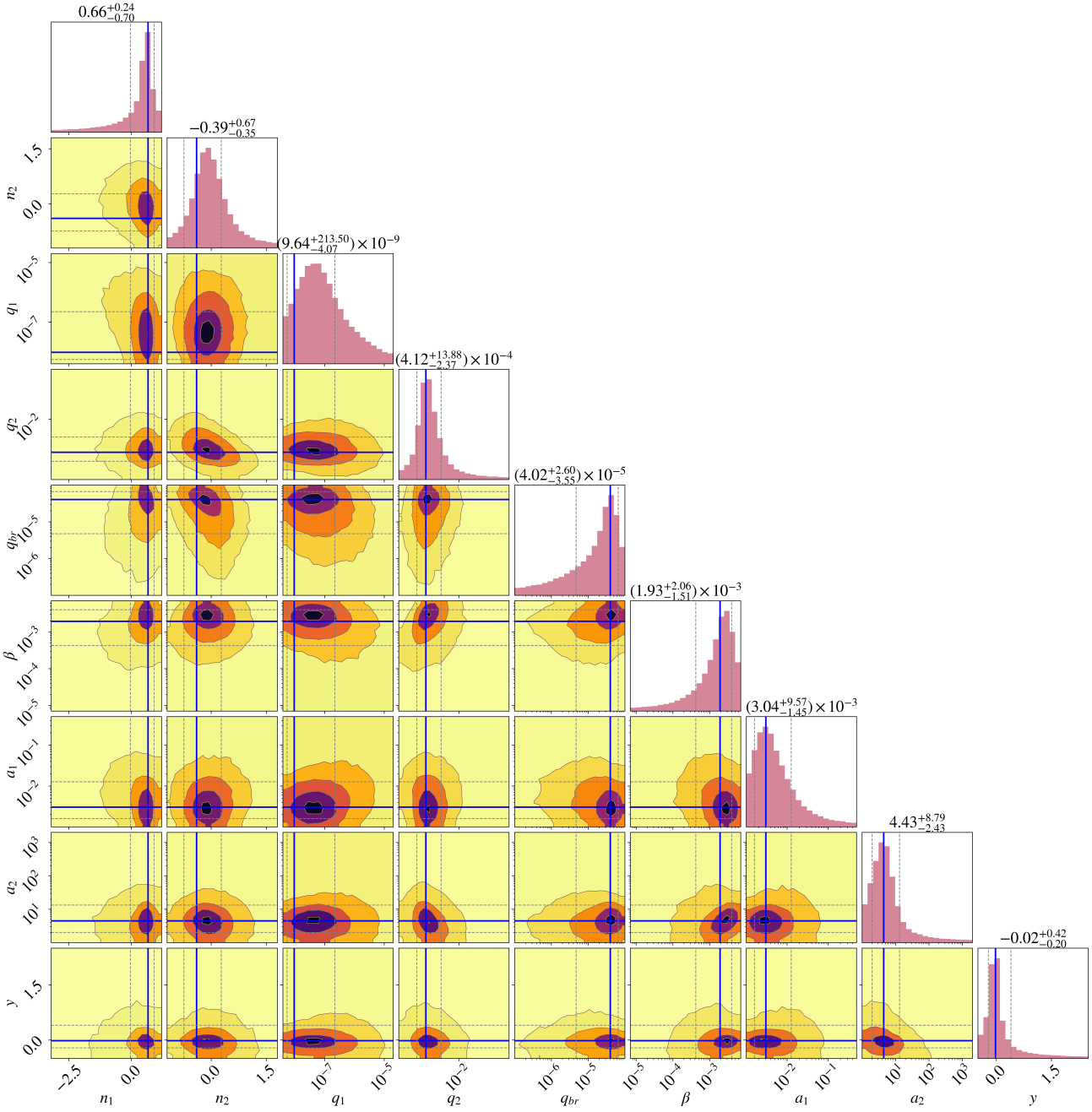

Figure 11. Cornerplot showing 1-D marginals (diagonals) and 2-D joint posteriors (off-diagonals) of samples while fitting for *Kepler* population on 9-parameter model as defined in Section 6, using differential evolution. The multi-dimensional best-fit value is overplotted in blue colour. The dashed grey lines show the upper-error and the lower-error on the best-fit value.

Burke C. J., Catanzarite J., 2017, Planet Detection Metrics: Per-Target Detection Contours for Data Release 25, KSCI-19111-002

Burke C. J., et al., 2015, ApJ, 809, 8

Edmondson K., Norris J., Kerins E., 2023, arXiv e-prints, p. arXiv:2310.16733

Fulton B. J., et al., 2017, The Astronomical Journal, 154, 109

Gaudi B. S., Meyer M., Christiansen J., 2021, in 2514-3433, ExoFrontiers. IOP Publishing, pp 2–1 to 2–21, doi:10.1088/2514-3433/ABFA8FCH2

Hsu D. C., Ford E. B., Ragozzine D., Morehead R. C., 2018, AJ, 155, 205

Jeyakumar G., Shanmugavelayutham C., 2011, International Journal of Artificial Intelligence & Applications, 2, 116–127

Johnson S. A., Penny M., Gaudi B. S., Kerins E., Rattenbury N. J., Robin A. C., Calchi Novati S., Henderson C. B., 2020, The Astronomical Journal, 160, 123

Jordi K., Grebel E. K., Ammon K., 2006, A&A, 460, 339

Marshall D. J., Robin A. C., Reylé C., Schultheis M., Picaud S., 2006, A&A, 453, 635

Montet B. T., Yee J. C., Penny M. T., 2017, Publications of the Astronomical Society of the Pacific, 129, 044401

Mulders G. D., Pascucci I., Apai D., Ciesla F. J., 2018, AJ, 156, 24

Obertas A., Van Laerhoven C., Tamayo D., 2017, Icarus, 293, 52

Pascucci I., Mulders G. D., Gould A., Fernandes R., 2018, ApJ, 856, L28

Penny M., Scott Gaudi B., Kerins E., Rattenbury N., Mao S., Robin A., Calchi Novati S., 2019, Astrophysical Journal, Supplement Series, 241

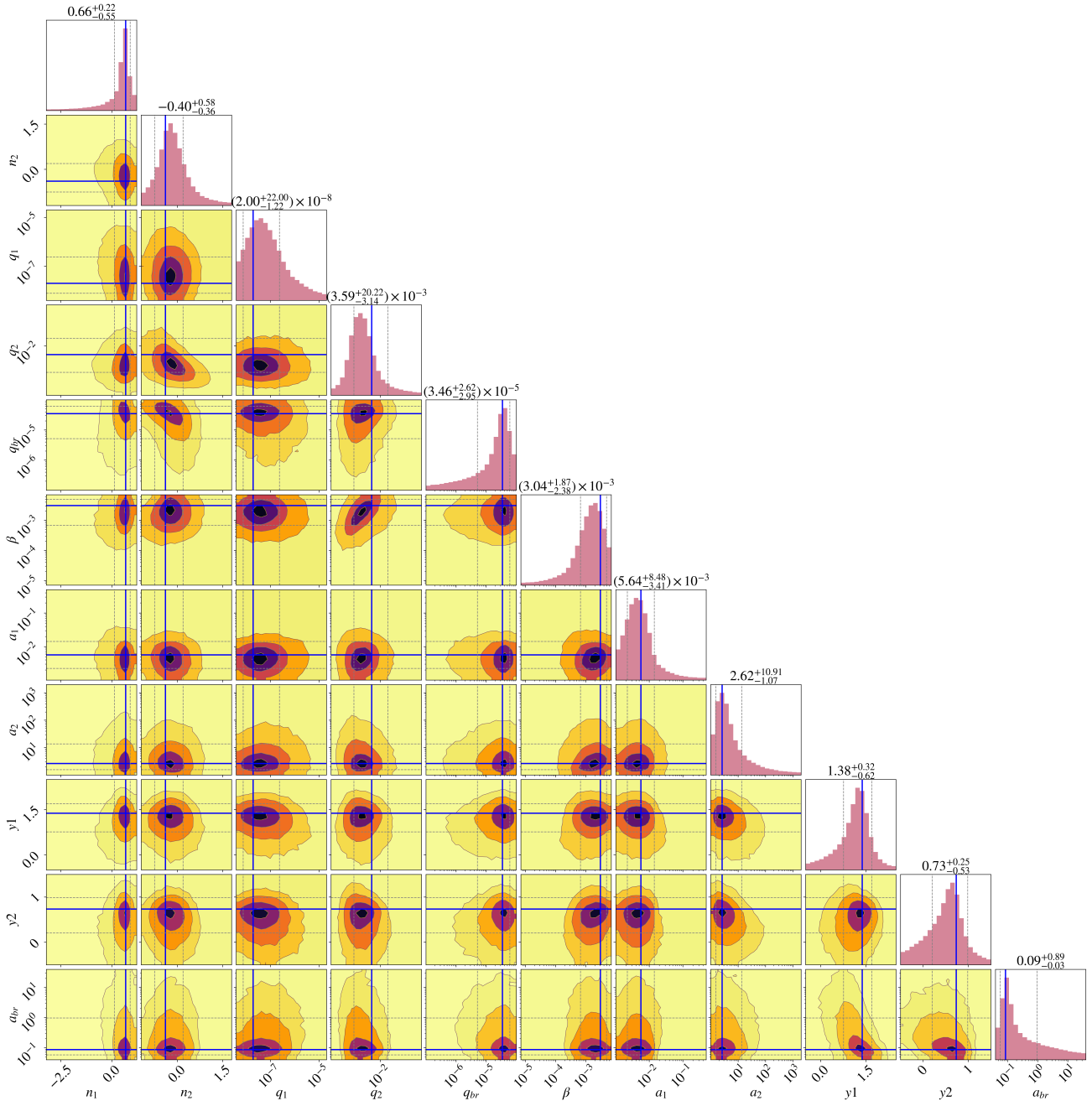

Figure 12. Cornerplot showing 1-D marginals (diagonals) and 2-D joint posteriors (off-diagonals) of samples while fitting for *Kepler* population on 11-parameter model as defined in Section 6, using differential evolution. The multi-dimensional best-fit value is overplotted in blue colour. The dashed grey lines show the upper-error and the lower-error on the best-fit value.

Perryman M., Hartman J., Bakos G. A., Lindegren L., 2014, The Astrophysical Journal, 797, 14

Qiang J., Mitchell C. E., 2014, IEEE Transactions on Evolutionary Computation

Robin A. C., Reylé C., Derrière S., Picaud S., 2003, A&A, 409, 523

Robin A. C., Marshall D. J., Schultheis M., Reylé C., 2012, \aap, 538, A106

Schwarz G., 1978, Annals of Statistics, 6, 461

Sobol' I. M., 1967, Zh. Vychisl. Mat. Mat. Fiz., 7, 86

Speagle J. S., 2020, MNRAS, 493, 3132

Storn R., Price K., 1997, Journal of Global Optimization, 11, 341

Virtanen P., et al., 2020, Nature Methods, 17, 261

Wilson R. F., et al., 2023, The Astrophysical Journal Supplement Series, 269, 5

Zhang H., Si S., Hsieh C.-J., 2017, GPU-acceleration for Large-scale Tree Boosting (arXiv:1706.08359), https://arxiv.org/abs/1706.08359

This paper has been typeset from a TEX/LATEX file prepared by the author.

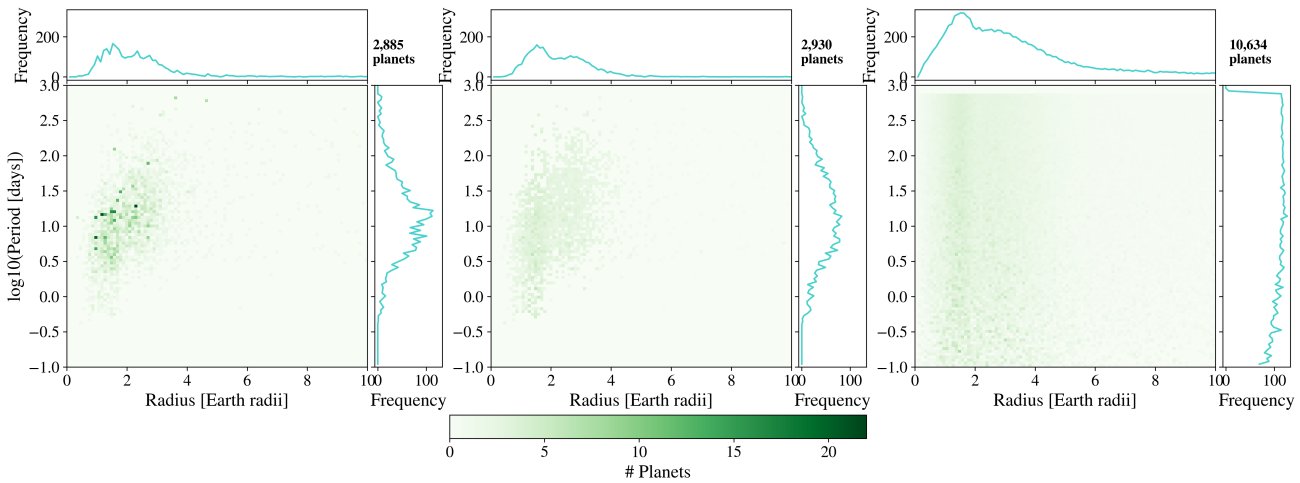

Figure 13. The figure shows the distribution of the projected *Kepler* observations (KeplerPORTs-corrected counts per bin) on the left and the corresponding recovered planets using the 9-parameter model on the middle and right. In the middle panel, we only plot the bins where the projected distribution had non-zero planets. In the rightmost panel, we plot all the bins from the recovered population. The colourmap shows the number of planets on the period-radius axes. The axes are accompanied by 1-D histograms of radius and period separately. In the top right corner of each 2-D histogram, the total number of the total planets is written.

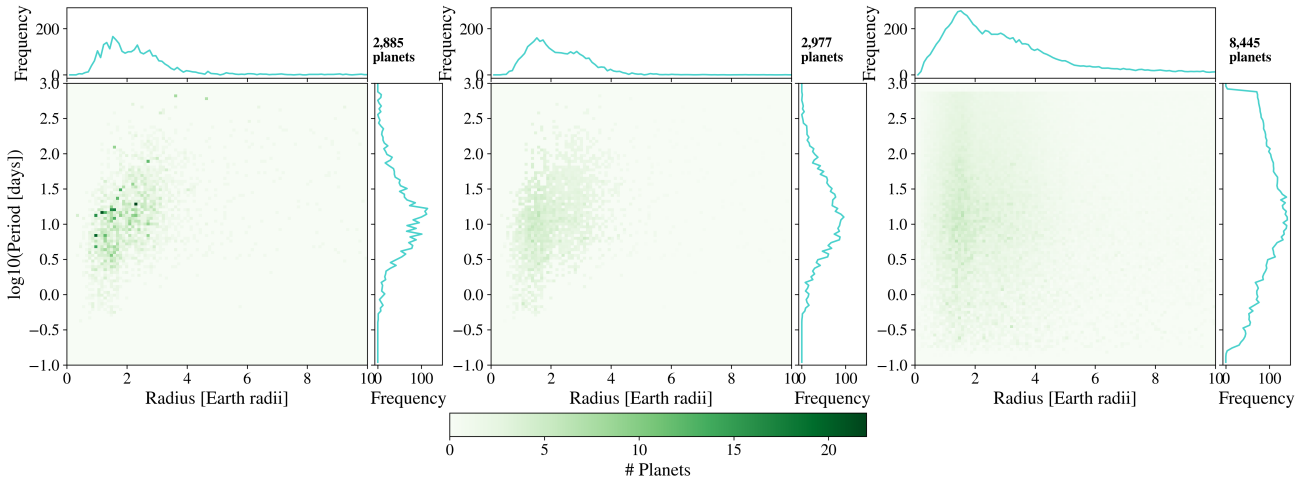

Figure 14. The figure shows the distribution of the projected *Kepler* observations (KeplerPORTs-corrected counts per bin) on the left and the corresponding recovered planets using the 11-parameter model in the middle and right. In the middle panel, we only plot the bins where the projected distribution had non-zero planets. In the rightmost panel, we plot all the bins from the recovered population. The colourmap shows the number of planets on the period-radius axes. The axes are accompanied by 1-D histograms of radius and period separately. In the top right corner of each 2-D histogram, the total number of the total planets is written.

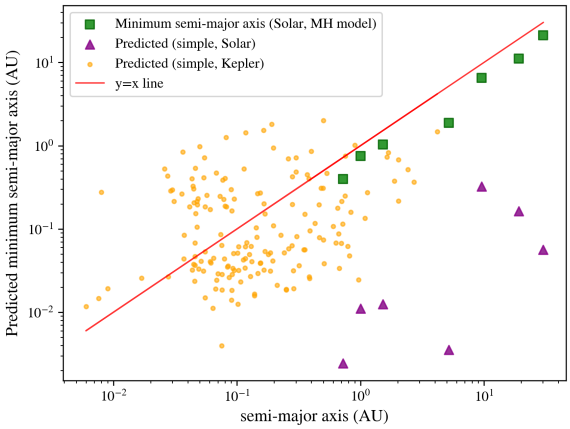

Figure 15. The figure shows the distribution of predicted a_{min} vs observed a for all *Kepler* exoplanets in orange dots, and Solar system planets in purple triangles, calculated using Equation (13). The green squares show the a_{min} calculated using Equation (11) for Solar system planets. The red line shows the $y = x$ line. Ideally, we would want all the orange dots to lie close to, and below the red line.

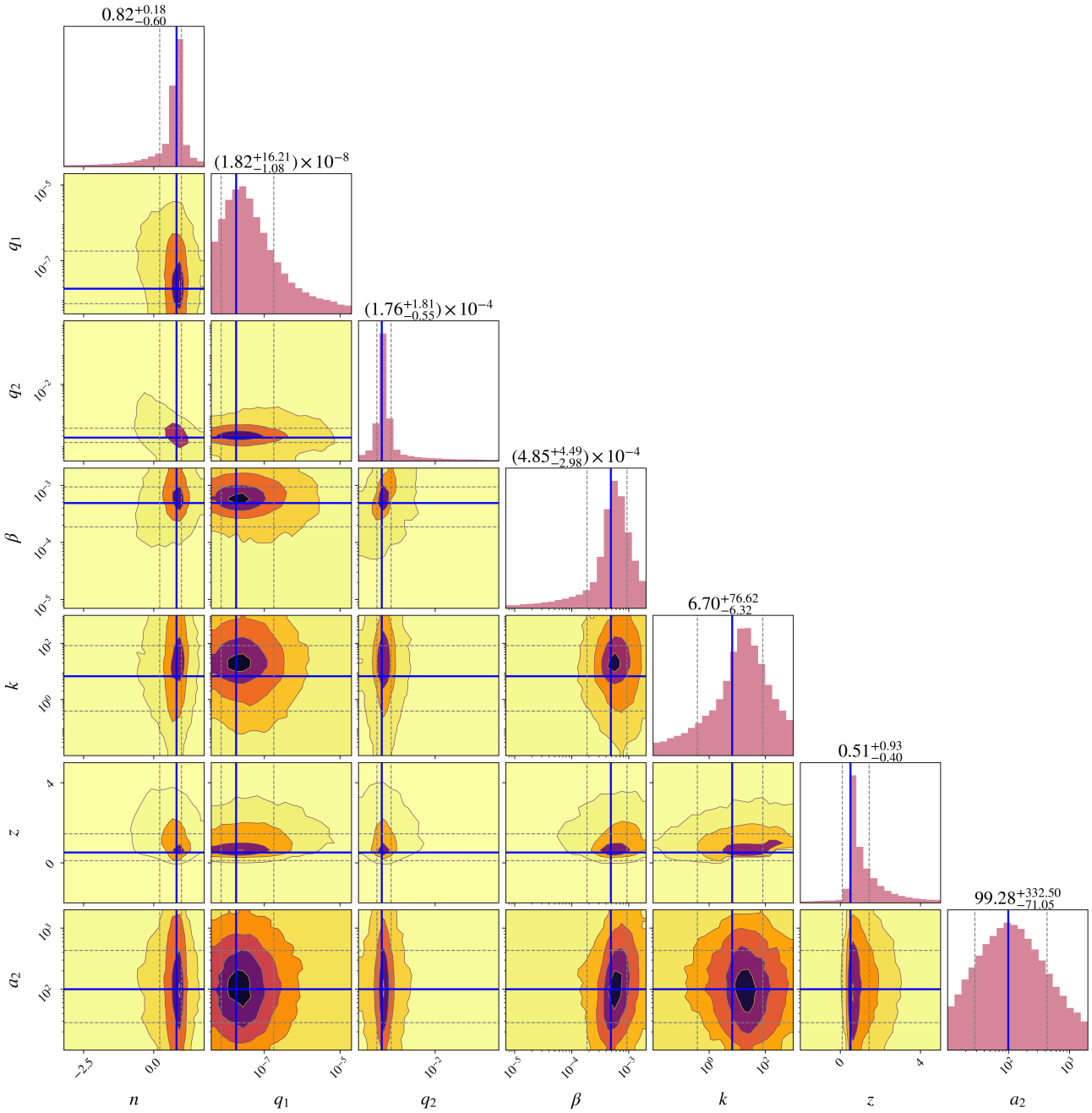

Figure 16. Cornerplot showing 1-D marginals (diagonals) and 2-D joint posteriors (off-diagonals) of samples while fitting for *Kepler* population on 7-parameter CPL model using differential evolution. The multi-dimensional best-fit value is overplotted in blue colour. The dashed grey lines show the upper-error and the lower-error on the best-fit value.

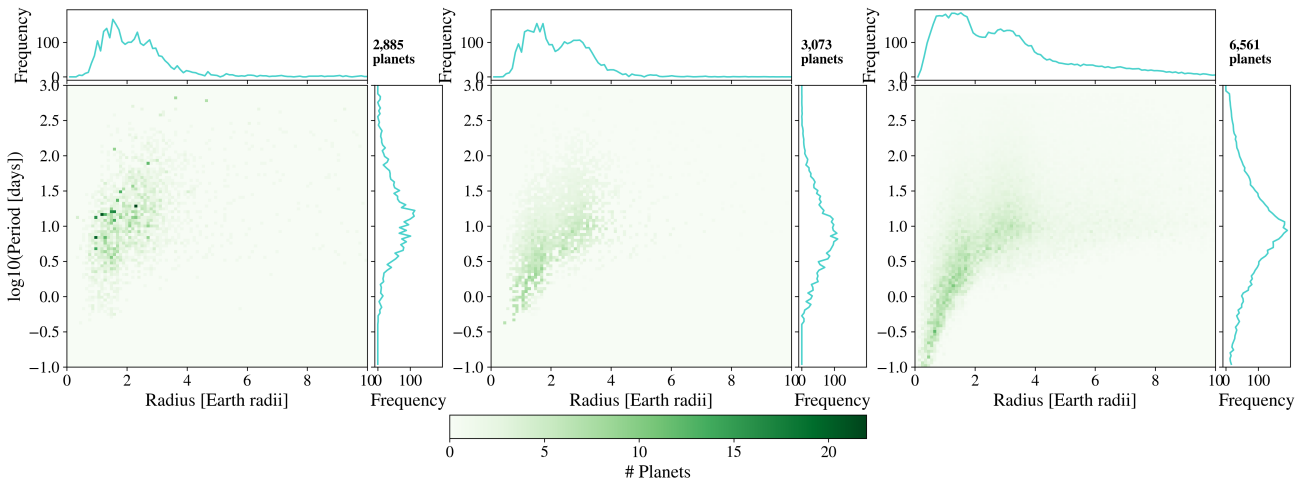

Figure 17. The figure shows the distribution of the projected *Kepler* observations on the left and the corresponding recovered planets using the CPL model (Section 6.1) on the middle and right. In the middle panel, we only plot the bins where the projected distribution had non-zero planets. In the rightmost panel, we plot all the bins from the recovered population. The colourmap shows the number of planets on the period-radius axes. The axes are accompanied by 1-D histograms of radius and period separately. In the top right corner of each 2-D histogram, the total number of the total planets is written.