

# Demographics of close-in *TESS* exoplanets orbiting FGK main-sequence stars

Kaiming Cui <sup>1,2</sup>★ David J. Armstrong <sup>1,2</sup> Andreas Hadjigeorghiou <sup>1,2</sup> Marina Lafarga <sup>1,2</sup>  
Vedad Kunovac <sup>1,2</sup> Lauren Doyle <sup>1,2</sup> Luis Agustín Nieto<sup>3,4</sup> and Rodrigo F. Díaz<sup>3,5</sup>

<sup>1</sup>Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

<sup>2</sup>Centre for Exoplanets and Habitability, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

<sup>3</sup>Instituto de Ciencias Físicas (CONICET / ECyT-UNSAM), Campus Miguelete, 25 de Mayo y Francia, (1650) Buenos Aires, Argentina

<sup>4</sup>Gerencia de Tecnología de la Información y de las Comunicaciones (GTIC), Subgerencia Vinculación y Desarrollo de Nuevas Tecnologías de la Información, DTE-CNEA. Centro Atómico Constituyentes, Av. Gral. Paz 1499, (1650) Buenos Aires, Argentina

<sup>5</sup>Instituto Tecnológico de Buenos Aires (ITBA), Iguazú 341, Buenos Aires CABA C1437, Argentina

Accepted 2025 December 23. Received 2025 December 3; in original form 2025 September 26

## ABSTRACT

Understanding the demographics of close-in planets is crucial for insights into exoplanet formation and evolution. We present a detailed analysis of occurrence rates for close-in (0.5–16 d) planets with radii between 2 and 20  $R_{\oplus}$  around FGK main-sequence stars. Our study uses a comprehensive sample from four years of *TESS* Science Processing Operations Center full-frame image data cross-matched with *Gaia*, analysed through our rigorous detection, vetting, and validation pipeline. Using high-confidence planet candidates, we apply a hierarchical Bayesian model to determine occurrence rates in the two-dimensional orbital period-radius plane. Our results are presented using 10-by-10 bins across the period-radius parameter space, offering unprecedented resolution and statistical precision. We find an overall occurrence rate of  $9.4^{+0.7}_{-0.6}$  per cent. When using identical binning, our occurrence rate posteriors distributions align with *Kepler*'s but have a magnitude smaller uncertainties on average. For hot Jupiters, we estimate the overall occurrence rate of  $0.39^{+0.03}_{-0.02}$  per cent. This value is consistent with the previous *Kepler* FGK-type result within  $1\sigma$ . We find an overall occurrence rate of Neptunian desert planets of  $0.08 \pm 0.01$  per cent, to our knowledge the first such determination. Additionally, in a volume-limited *Gaia* subsample within 100 pc in the same parameter region, we measure an overall planet occurrence rate of  $15.4^{+1.6}_{-1.5}$  per cent and a hot Jupiter occurrence rate of  $0.42^{+0.16}_{-0.12}$  per cent. Our results establish an improved foundation for constraining theoretical models of exoplanet populations.

**Key words:** methods: data analysis – methods: statistical – planets and satellites: detection – planets and satellites: fundamental parameters – planetary systems.

## 1 INTRODUCTION

Over the past three decades, the number of confirmed exoplanets has grown from a handful to several thousand, enabling the transition from individual detections to statistical studies of planetary populations. The statistical characterization of these populations, often referred to as exoplanet demographics, provides crucial constraints on models of planet formation and evolution (e.g. C. Mordasini, Y. Alibert & W. Benz 2009; G. D. Mulders et al. 2019; J. Drażkowska et al. 2023). Large, uniform transit and radial velocity (RV) surveys have revealed striking structures in the period–radius and period–mass planes, such as the ‘hot Jupiter (HJ)’ population (e.g. A. Cumming et al. 2008; J. T. Wright et al. 2012), the ‘Neptunian desert’ (e.g. Gy. M. Szabó & L. L. Kiss 2011; C. Beaugé & D. Nesvorný 2012; T. Mazeh, T. Holczer &

S. Faigler 2016), and the ‘radius valley’ (e.g. B. J. Fulton et al. 2017; V. Van Eylen et al. 2018). Each of these features encodes the imprint of processes such as disc migration, atmospheric escape, and core accretion efficiency (e.g. E. D. Lopez & J. J. Fortney 2013; S. Jin et al. 2014; J. E. Owen & Y. Wu 2017; R. I. Dawson & J. A. Johnson 2018; J. E. Owen & D. Lai 2018), and mapping them with increasing precision remains a central goal of exoplanetary science.

The NASA *Kepler* mission (W. J. Borucki et al. 2010) offers the first opportunity for statistically robust measurements of planetary occurrence rates across a wide range of orbital periods and planetary radii (e.g. A. W. Howard et al. 2012; F. Fressin et al. 2013; C. D. Dressing & D. Charbonneau 2015; E. A. Petigura et al. 2018; D. C. Hsu et al. 2019; S. Bryson et al. 2020; D. C. Hsu, E. B. Ford & R. Terrien 2020; M. Kunimoto & J. M. Matthews 2020; A. Dattilo, N. M. Batalha & S. Bryson 2023). Its unprecedented photometric precision reveals not only the abundance of small planets but also the fine structures in their distribution (e.g. B.

\* E-mail: [Kaiming.Cui@warwick.ac.uk](mailto:Kaiming.Cui@warwick.ac.uk)

J. Fulton et al. 2017; A. Dattilo & N. M. Batalha 2024). However, *Kepler's* focus on a single, narrow field restricted its stellar sample to relatively faint targets, making follow-up and confirmation challenging.

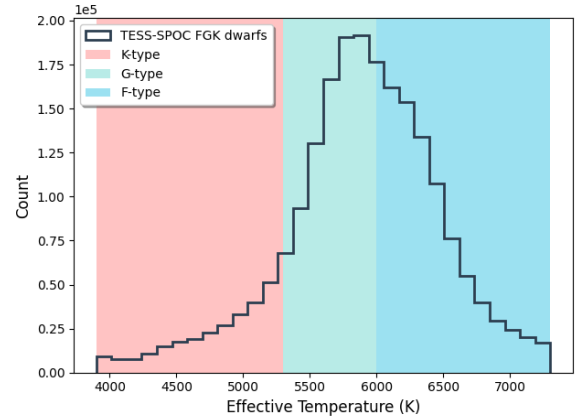
NASA's *Transiting Exoplanet Survey Satellite (TESS)* (G. R. Ricker et al. 2014) enables the extension of the demographic revolution to bright, nearby stars across the entire sky. *TESS* observes the sky in a series of overlapping 27 d sectors, with extended coverage in its continuous viewing zones, enabling sensitive searches for short-period ( $P \lesssim 10\text{--}20$  d) transiting planets (e.g. N. M. Guerrero 2021). Its all-sky coverage, emphasis on bright stars, and systematic observing strategy make it particularly powerful for constraining the occurrence rates of close-in planets. For FGK main-sequence stars, the brightness of *TESS* targets allows precise stellar characterization (e.g. K. G. Stassun et al. 2018, 2019) and facilitates RV follow-up (e.g. C. X. Huang et al. 2018), yielding accurate planetary masses and densities.

The *TESS* mission has already yielded several demographic studies of short-period planets (e.g. R. B. Fernandes et al. 2022; E. M. Bryant, D. Bayliss & V. Van Eylen 2023; K. Ment & D. Charbonneau 2023; E. J. Melton et al. 2024; S. Vach et al. 2024). Most such studies have focused on young stars or M dwarfs, or have analysed early subsets of the *TESS* data set. Some have revealed hints of differences between the *TESS* and *Kepler* planet populations, likely due to differences in target selection, magnitude limits, and observing baselines. A comprehensive, statistically robust census of close-in planets around full-sky FGK main-sequence stars in the *TESS* era is still lacking. Such a study can test the universality of features such as refine the boundaries of the Neptunian desert, improve the statistical precision of hot Jupiter occurrence rates, and provide a uniform, bright-star sample optimized for follow-up.

In this work, we present a systematic demographic analysis of close-in ( $0.5 < P < 16$  d) *TESS* planets orbiting FGK main-sequence stars. We describe our sample selection in Section 2 and our completeness estimation in Section 3, and outline our occurrence-rate methodology in Section 3.5. We then present occurrence rates for our stellar sample (Section 4.2), a *Gaia* 100 pc volume-limited sample (Section 4.3), for hot Jupiters, and Neptunian desert (Section 4.4). Next, we discuss the impact of different models, thresholds and RUWE values on our occurrence rate results (Sections 5.1 and 5.2) and compare our findings with those from *Kepler* (Section 5.3). Finally, we offer future prospects in Section 5.4.

## 2 STELLAR SAMPLE

This study aims to determine the planet occurrence rate from the full-sky *TESS* observations of FGK-type main-sequence stars. To achieve this, we use a main-sequence sample from L. Doyle et al. (2024), which cross-matches objects from the *TESS*-Science Processing Operations Center (SPOC) full-frame image (FFI) releases (D. A. Caldwell et al. 2020) with data from *Gaia* Data Release 2 (A. G. A. Brown et al. 2018) and Data Release 3 (R. L. Smart et al. 2021). The *TESS*-SPOC observations are from the first four years (Sectors 1–55), which cover most of the fields in the Northern and Southern hemispheres and the ecliptic region. The *TESS*-SPOC select sources mainly by FFI field stars with  $T_{\text{mag}} < 13.5$ , but also include some fainter high value objects. Thus we require that the stars in our sample must be brighter than  $G_{\text{mag}} 14$ , have a `parallax_over_error > 5`, and surface gravity



**Figure 1.** Effective temperature distributions of our *TESS*-SPOC FGK main-sequence stars. Shaded regions represent the temperature boundaries for different FGK spectral types.

$\log g > 3.5$  to ensure that they are main-sequence stars. Full details of the sample construction and its properties are available in L. Doyle et al. (2024) and Lafarga et al. (submitted).

From the above sample, we further select FGK-type stars using the *TESS* Input Catalogue v8.2 (TIC; K. G. Stassun et al. 2019; M. Paegert et al. 2021), defined by an effective-temperature range of 3900–7300 K (M. J. Pecaut & E. E. Mamajek 2013; M. Kunitomo & S. Bryson 2020), yielding a stellar sample of 2097 359 targets for our detection pipeline. Although not severe, the *TESS*-SPOC target selection is heterogeneous to some extent. Given the sample's large size and our additional cuts, the resulting targets can be approximated as a uniform, magnitude-limited sample of FGK main-sequence stars. Our planet detection and occurrence rate estimation are established on this sample, and its effective temperature distribution is shown in Fig. 1. We download the FFI light curves of our sample from the Mikulski Archive for Space Telescopes.<sup>1</sup> The pre-search data conditioning simple aperture photometry is used for our planet detection.

## 3 METHODS

The detected planets are characterized by their intrinsic occurrence rate and discovery completeness. To accurately estimate the occurrence rate, it is necessary to assess the efficiency, or completeness, of the selection process. Completeness maps should faithfully represent the entire sequence from observation to validation. We achieve this via simulation, applying the same pipeline to both simulated and observed stellar samples. We will briefly introduce the key procedures and our approach to modelling completeness estimates.

Once we understand the completeness of each stage, calculating the occurrence rate typically involves working backwards. In this study, we implement two algorithms: one for qualitative demonstration and another for quantitative analysis. They can validate each other, offering a comprehensive view with different resolutions and robustnesses.

<sup>1</sup><https://archive.stsci.edu/hlsp/tess-spoc>

### 3.1 Simulated sample

The entire process and details of our simulation are described in A. Hadjigeorgiou et al. (2025). Here, we will introduce the key ideas and choices.

Our simulated sample is built by constructing planetary or binary systems on a target stellar sample. The target stellar sample should consist of single stars and have stellar properties that are identically distributed with respect to those of the observed sample. To do so, we also begin with the stellar sample from L. Doyle et al. (2024, as described in Section 2). Differently from the stellar sample for detection and occurrence rate estimation, we exclude *TESS* Objects of Interest (TOIs), community candidates, *Kepler* Objects of Interest, and K2 candidates listed in ExoFOP<sup>2</sup>, as well as stars with *Gaia* RUWE greater than 1.05, ensuring that the remaining targets are likely single stars without known planetary signals. Note that as part of our false positive simulations, higher RUWE stars are in effect ‘created’ from this sample, so our results are not limited to low RUWE stars.

Given the scope of our study, we restrict their effective temperature range, based on TIC data, from 3000 to 10 000 K. This range is broader than that of our current sample used for occurrence rate calculations, as we are reserving it for future plans. However, for our occurrence rate calculations and completeness map estimations in this work, we limit both our simulated and observed stellar samples to the same Teff range as our stellar sample (i.e. FGK stars). This ensures that the simulated sample is directly comparable to the stellar sample used in our analysis. In the end, about 1.2 million *TESS*-SPOC FFI stars are used as the target sample for our simulation. This number is significantly lower than the stellar sample in Section 2, primarily due to the stringent RUWE criterion we applied excluding almost half of the stars. We will discuss the impact of RUWE on our occurrence rate results in Section 5.2.

The construction of planetary systems around the target stars is based on orbital simulations and a light curve generator. The planets’ orbital parameters are sampled from pre-determined distributions. For the orbital period and planetary radius, we have two different distributions. The first distribution is based on planetary occurrence rates, sampling orbital periods and planet radii according to their relative occurrence rates from *Kepler* results (D. C. Hsu et al. 2019). The second distribution is a log-uniform distribution for both period and radius. Both distributions cover orbital periods from 0.5 to 16 d and planet radii from 1 to 16  $R_{\oplus}$ . We call the sample with parameters drawn based on literature occurrence rates the non-uniform sample, and the sample with uniformly drawn parameters the uniform sample.

For the other parameters, inclinations are drawn from an isotropic distribution,  $p(i) \propto \sin i$  for  $i \in [0, \pi/2]$ . The argument of periastron is drawn from a uniform distribution,  $\mathcal{U}(0, 2\pi)$ . Eccentricities follow the distribution from M. Moe & R. D. Stefano (2017), with  $p_e \propto e^\eta$  and  $\eta = -0.3$ , truncated at  $e_{\max}(P)$ , a function of orbital period defined in equation (3) of M. Moe & R. D. Stefano (2017). They are set to zero for periods shorter than 2 d. Albedos are drawn uniformly over  $[0, 1]$ . Planet masses are assigned using the empirical mass–radius relation of S. Müller et al. (2024). These distributions are applied identically to both the uniform and non-uniform samples. We also simulate an-

other independent uniform planetary sample with an increased radius range up to 20  $R_{\oplus}$ ; this is reserved only for completeness estimation.

We also simulate samples for eight different astrophysical false positive (FP) scenarios:

- (1) hierarchical transiting planet (HTP; a planet transits an unresolved stellar companion of the target star)
- (2) nearby transiting planet (NTP; a diluting transit system not associated with the target star)
- (3) background transiting planet (BTP; a background transit system blend with the target star)
- (4) eclipsing binary (EB; a secondary star eclipses the target star)
- (5) background eclipsing binary (BEB; a background EB system blend with the target star)
- (6) nearby eclipsing binary (a diluting EB system not associated with the target star)
- (7) hierarchical EB system (HEB; a wide orbit EB system bound to the target star)
- (8) nearby hierarchical EB system (a diluting HEB system not associated with the target star).

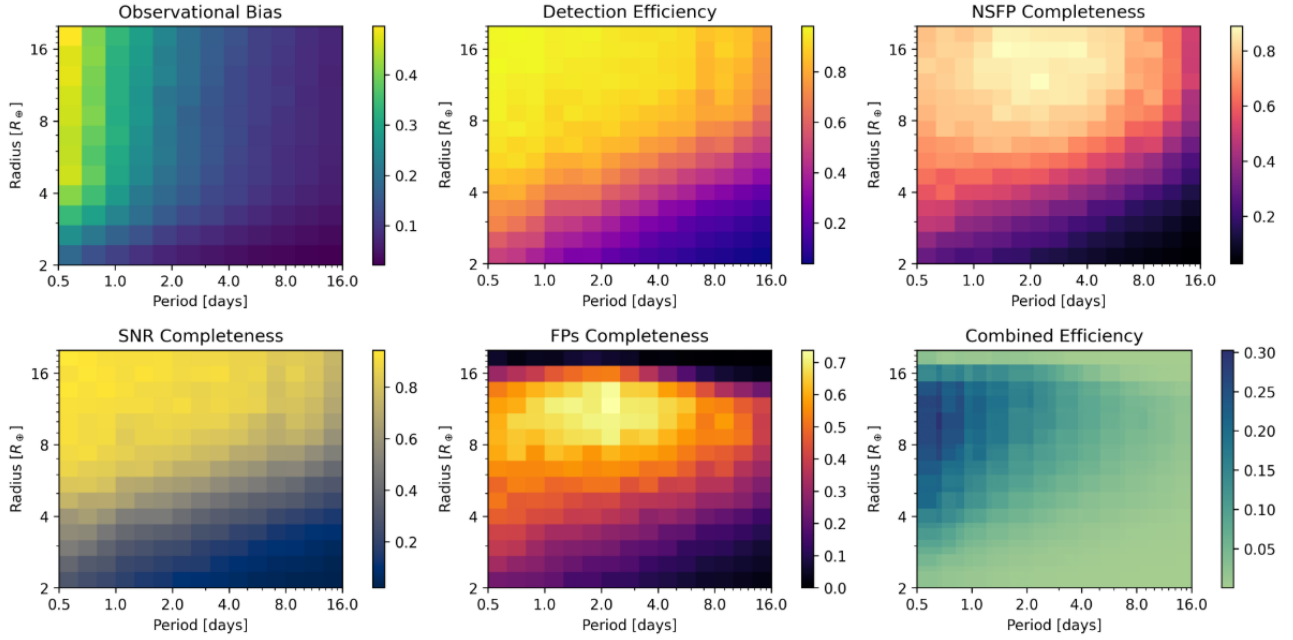
Full details of the FP parameter configuration are provided in A. Hadjigeorgiou et al. (2025). We generate light curves for all simulated systems using JKTEBOP (B. Nelson & W. D. Davis 1972; P. B. Etzel 1981; D. M. Popper & P. B. Etzel 1981; J. Southworth, H. Bruntt & D. L. Buzasi 2007). Consequently, not all simulated systems produce a light curve; this is primarily due to non-transiting orbital configurations, though unrealistic stellar and planetary properties can also cause failures. We also exclude systems whose generated light-curve depth is below 300 ppm to improve efficiency. We omit the BTP scenario in our study because our simulations show an extremely low-success rate (fewer than 500 per million), indicating that such cases are rare and can be safely ignored in our validation. Similar to the planetary simulation, the planetary systems in our FPs (HTP, NTP, BTP) are also simulated with orbital parameters drawn from uniform distributions and from literature.

The simulations are performed using our modified PASTIS code<sup>3</sup> (R. F. Díaz et al. 2014; A. Santerne et al. 2015). We simulate at least 30 000 objects with successfully generated light curves for each scenario, and most scenarios involve more than 100 000 objects. The final simulated light curves are created by injecting the generated transiting or eclipsing light curves into the corresponding target stars’ *TESS*-SPOC light curves.

Unlike some previous work (e.g. S. Giacalone et al. 2020), we do not include period aliases (e.g. twice or half the orbital period) directly in our simulations; however, they appear in our training samples in certain scenarios. We discuss them in Section 3.4. Additionally, we have a non-simulated false positive (NSFP) scenario. It represents any stellar variability or noise features not well modelled in other simulated scenarios. Unlike the other scenarios, the NSFP sample is generated from a randomly selected sample of box least square (BLS) peaks on randomly chosen *TESS*-SPOC light curves after removing TOIs.

<sup>2</sup><https://exofop.ipac.caltech.edu/tess/>

<sup>3</sup><https://github.com/ckm3/pastis-dev>



**Figure 2.** Observational bias, detection efficiency, NSFP completeness, signal-to-noise ratio (SNR) completeness, FP completeness, and combined efficiency of transiting planets in our stellar sample as functions of orbital period and planet radius. Each subplot title describes its process; higher values indicate more planets retained.

### 3.2 Observational bias

Not all planetary systems are observable via the transit method. We refer to the observability of transiting planets in our target sample as an observational bias or observational completeness.

With our reserved uniform planet sample, we can directly examine the observational bias of the target stars in orbital period and planet radius space. We first limit the effective temperature of the target stars to FGK spectral types. For each bin in period and radius space, we estimate the yield rate by counting the number of successfully generated transiting planets over the total number simulated. Due to the photometric precision of *TESS* and the reliability requirements for our occurrence rates, we restrict our observational-bias calculations to radii  $\geq 2 R_{\oplus}$ . For consistency, our other completeness calculations use the same parameter-space ranges.

The upper-left subplot of Fig. 2 shows the observational bias map for our target stars. Higher completeness values are concentrated at shorter orbital periods and larger planetary radii, primarily reflecting the geometric transit probability in our stellar sample. We also require a minimum candidate transit depth of 300 ppm to ensure that candidate signals are sufficiently deep to be robustly detectable in our data.

### 3.3 Detection efficiency

Detection efficiency quantifies the completeness of transiting planets that are successfully detected. We first describe our detection pipeline.

We begin by detrending the light curves using the Savitzky–Golay filter (A. Savitzky & M. J. E. Golay 1964), applying a third-degree moving polynomial with a four-day window length. Then,

we apply a GPU-based BLS algorithm using CUVARBASE<sup>4</sup> to the entire light curve to identify the five most significant peaks for each stellar target. We further select candidates with a signal detection efficiency (G. Kovács, S. Zucker & T. Mazeh 2002) greater than 7 and a multiple event statistic (MES) greater than 0.8. Our MES algorithm is a modified implementation of J. M. Jenkins et al. (2016); see A. Hadjigeorgiou et al. (2025) for details. Selected candidates are further excluded if their depth varies by more than 50 per cent from the average across sectors or if the depth in any sector is below 200 ppm. Note this is a separate sector-by-sector cut, as opposed to our 300 ppm average depth limit in our simulation.

Analogous to our treatment of observational bias, we compute the detection efficiency in the orbital period–planet radius plane. In each bin, we calculate the fraction of correctly recovered planets (period difference within 2 per cent, epoch difference within 0.5 d, and detected-to-injected transit-depth ratio and its reciprocal must be less than 3) relative to the number injected. The top-middle subplot of Fig. 2 shows the detection efficiency estimated from our simulated sample. Generally, shorter periods and larger radii yield higher detection efficiency, consistent with previous signal-to-noise ratio (SNR)-based estimates (e.g. G. D. Mulders et al. 2018).

### 3.4 Vetting and validation completeness

Although many planetary candidates pass the thresholds after the BLS process, a significant number of false positives remain. To reliably identify true exoplanets and rule out signals caused by instrumental effects or astrophysical phenomena, further vetting and validation are usually required (e.g. T. D. Morton et al.

<sup>4</sup><https://github.com/johnh2o2/cuvarbase>

2016; S. E. Thompson et al. 2018; S. Giacalone et al. 2020; D. J. Armstrong, J. Gamper & T. Damoulas 2021; H. Valizadegan et al. 2022). In this work, vetting and validation are carried out by our RAVEN pipeline<sup>5</sup> (A. Hadjigeorgiou & D. J. Armstrong 2024; A. Hadjigeorgiou et al. 2025), which is designed for TESS exoplanet candidates. Here, we provide only a brief overview of the pipeline; for more details, please refer to A. Hadjigeorgiou et al. (2025).

RAVEN uses two different feature-based machine learning classifiers for its vetting and validation, specifically a Gradient Boosted Decision Tree (GBDT) based on the XGBoost (T. Chen & C. Guestrin 2016) and a Gaussian process (GP) classifier (C. E. Rasmussen & C. K. I. Williams 2008). The classifiers are binary and trained to distinguish true planetary signals against each astrophysical FP scenario and the NSFP. Due to the uniqueness of the latter, the classification result is only used for vetting the candidates. All classifiers are trained on our simulated dataset, which is split into training, validation, and test sets in an 80 per cent, 10 per cent, and 10 per cent ratio, respectively. The features used by the classifiers generally fall into five categories: stellar properties (e.g. BP-RP, distance), eclipse parameters (e.g. transit depth, duration, and ratio of semimajor axis and planetary radius), eclipse significance (e.g. SNR, MES, Single Event Statistic), nearby metrics (target fraction and nearby fraction, used only for nearby scenarios), and Self-Organising Map (T. Kohonen 1982; D. J. Armstrong, D. Pollacco & A. Santerne 2017) features. A complete list of features is available in A. Hadjigeorgiou et al. (2025).

The feature data may contain orbital period aliases. We exclude these aliases from our planetary scenarios because aliased planets are not considered true detections. However, we include the aliases in our EB-based scenarios, as they represent a significant subset of common FPs. Therefore, since our model lacks a training set for aliased planets, our final planet candidates require additional alias removal (see Section 4.1).

Given the presence of seven different FP scenarios, we train separate classifiers for each scenario to distinguish FPs from planets. Including NSFPs, we thus have eight binary classifiers, with each classifier utilizing two machine learning models, resulting in a total of 16 models. To ensure reliable statistical results, we also balance the sample sizes so that, for each FP and planet pair, the training set contains an equal number of FPs and planets. We use the average score of the GBDT and GP models as the combined score. The classification precision on our test set exceeds 0.99 across all scenarios, while recalls are above 0.85 in all cases except for NTP (0.78) and HTP (0.80).

We also evaluate the calibration performance, which generally follows the perfect calibration line (see fig. 13 in A. Hadjigeorgiou et al. 2025). Therefore, we interpret the model-predicted scores as probabilities for each scenario. Then, we compute posterior probabilities by combining these scenario probabilities with priors for each candidate under Bayes' theorem. The prior for a candidate includes the occurrence rate, detection rate, and positional probability (see A. Hadjigeorgiou & D. J. Armstrong 2024). Thus, vetting and validation here can be achieved by setting appropriate posterior probability thresholds to rule out FPs and NSFPs.

Due to the different planetary signal injections (Section 3.1), we have two simulated planetary samples: one uniform and one

non-uniform. Accordingly, we train two models per classifier for each planetary FP – one on the uniform sample and one on the non-uniform sample. As a result, we obtain two posterior probabilities for each object, depending on whether the uniform or non-uniform model is used. The differing orbital period and radius distributions influence each model to learn the respective intradistribution of its sample. Comparing these two models provides a test of the robustness of our occurrence rate estimation.

In this work, we set the thresholds for the NSFP and FP classifiers to 0.9 (where 0 corresponds to 100 per cent false positives and 1 to 100 per cent planets) for both non-uniform and uniform models. We selected this value because model calibration is more reliable for scores above 0.9 (A. Hadjigeorgiou et al. 2025). Additionally, excessively high thresholds would result in fewer candidates, causing completeness estimation to fail due to numerous zero detections, as well as increased Poisson noise in our occurrence rate calculation.

Similar to the detection efficiency, we estimate our NSFP and FPs completeness using the simulated planet sample. Specifically, in each bin, we compute the fraction of correctly classified planets relative to the number injected. For FPs, a planet candidate is counted as correctly classified only if all seven models have posterior probabilities greater than 0.9. The top-right and bottom-middle subplots of Fig. 2 illustrate the completeness of the NSFP and FP with non-uniform models. We observe lower completeness in the FP for regions with larger radii and longer periods, where candidates can resemble EBs. In areas with smaller radii and shorter periods, the model becomes confused by scenarios such as BEB, NTP, and other nearby cases. Similarly, Fig. 3 shows the completeness of the NSFP and FP classifiers trained on the uniform sample. Because the uniformly simulated planet distribution only reflects observational bias, the FP and NSFP classifiers learn this distribution, resulting in higher completeness in regions where planets are more abundant.

Additionally, we introduce an SNR cut to further eliminate weak signals, particularly those resulting from misclassified NSFPs. Our SNR is estimated from the trapezoidal fit results, and we apply a threshold of 10. The bottom-left subplot of Fig. 2 shows the resulting completeness map after our injection-and-recovery tests on the simulated planets. The SNR completeness shows the expected smoothly increasing completeness towards shorter period and larger radius.

Different thresholds result in different model outputs and corresponding completeness maps, which in turn affect the occurrence rates. We discuss the impact of varying thresholds further in Section 5.1.

### 3.5 Occurrence rate estimation

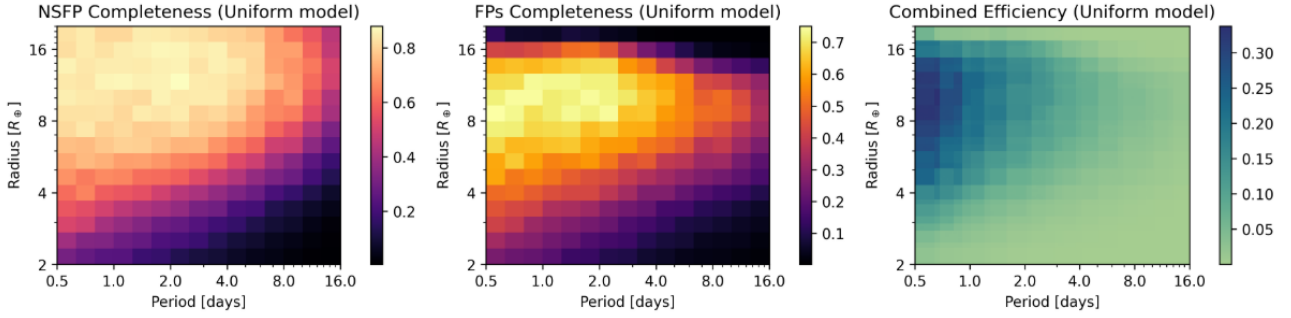
We apply two commonly used methods to estimate the occurrence rate: the inverse detection efficiency method (IDEM; e.g. A. W. Howard et al. 2012; A. Dattilo et al. 2023) and a hierarchical Bayesian model of the Poisson process (e.g. D. Foreman-Mackey, D. W. Hogg & T. D. Morton 2014).

#### 3.5.1 Inverse detection efficiency method

The occurrence rate of IDEM in a specific bin, denoted as  $\Gamma$ , can be generalized using the following equation:

$$\Gamma = \frac{N_p \langle r \rangle}{N_* \langle q \rangle}, \quad (1)$$

<sup>5</sup><https://github.com/ahadjigeorgiou/RAVEN-Pipeline>



**Figure 3.** NSFP completeness, FPs completeness, and combined efficiency as a function of the orbital period and planet radius. The NSFP and FPs models are trained on our uniformly injected sample.

where  $N_p$  represents the number of planets within a specific bin,  $N_*$  denotes the number of host stars,  $\langle r \rangle$  is the average reliability for that bin, and  $\langle q \rangle$  is the average completeness for that bin. In most cases, the  $\langle r \rangle$  is one as they are vetted or confirmed as planets by different observations. However, IDEM is not well-established in probability theory and can lead to biased results because it does not adequately account for uncertainties in transit parameters. Also, when detection efficiency is low or detections are few or absent, it leads to an underestimation and significant fluctuations in the occurrence rate (e.g. D. Foreman-Mackey et al. 2014; D. C. Hsu et al. 2018; W. Zhu & S. Dong 2021).

We demonstrate the use of IDEM with a kernel density function (KDE) to qualitatively measure planet occurrence, known as detection efficiency weighted KDE (wKDE). This method has been applied in various studies (e.g. T. D. Morton & J. Swift 2014; E. A. Petigura et al. 2022; A. Dattilo et al. 2023) to provide a smooth distribution without binning artifacts. We compute the wKDE by applying a KDE with weights  $w = r/q$ . Our reliability uses the RAVEN probability (the minimum of FP probabilities and NSFP probability). Our completeness is modelled using Beta distributions with the Jeffreys prior, based on our injection and recovery tests ( $\alpha = 1/2 + n_{\text{rec}}$ ,  $\beta = 1/2 + n_{\text{inj}} - n_{\text{rec}}$ ; H. Jeffreys 1946). The bottom right subplots of Figs 2 and 3 display the combined efficiency for the non-uniform and uniform models. This is the product of observational bias and the completeness of our pipeline, where completeness is measured as the number of simulated planets that pass through the whole pipeline divided by the total number of injected planets.

### 3.5.2 Hierarchical Bayesian method

To reliably quantify the occurrence rate, we use the hierarchical Bayesian methodology to model the underlying Poisson process (e.g. S. Tabachnik & S. Tremaine 2002; A. N. Youdin 2011; D. Foreman-Mackey et al. 2014; B. J. Fulton et al. 2021). Compared to IDEM, this approach provides robust uncertainty estimations and is more reliable for cases with low detections and low efficiencies. This is particularly important for quantifying the Neptunian desert region. In this work, we slightly modify the original form to include the uncertainties of completenesses, and the reliabilities of our candidates. The log-likelihood can be written as

$$\ln \mathcal{L}(\theta) = - \int Q(\omega) \Gamma(\omega | \theta) d\omega + \sum_{k=1}^K r_k \cdot \ln \left( \frac{1}{N_k} \sum_{n=1}^{N_k} \frac{Q(\omega_k^{(n)}) \Gamma(\omega_k^{(n)} | \theta)}{p(\omega_k^{(n)})} \right). \quad (2)$$

Here,  $\theta$  are the parameters for the occurrence rate model.  $\omega$  represents physical parameters (orbital period  $P$  and planet radius  $R_p$  in this study). Thus  $\Gamma(\omega)$  is the occurrence rate density, defined as  $\Gamma = d^2 N_p / (d \ln P d \ln R_p)$ .  $Q(\omega)$  is the completeness map, while  $r_k$  denotes the reliability of the  $k$ -th planet.  $N_k$  is the number of posterior samples of the  $k$ -th planet's physical parameters. The term  $\omega_k^{(n)}$  refers to the  $n$ -th sample for the same  $k$ -th planet. Additionally,  $p(\omega_k^{(n)})$  indicates each parameter's prior probability distribution for every companion's posteriors. Here,  $r_k$  serves as a Bernoulli weight on the log-likelihood. When  $r_k = 0$ , the contribution from the  $k$ -th planet is omitted; when  $r_k = 1$ , the log-likelihood reduces to its standard form (D. Foreman-Mackey et al. 2014).

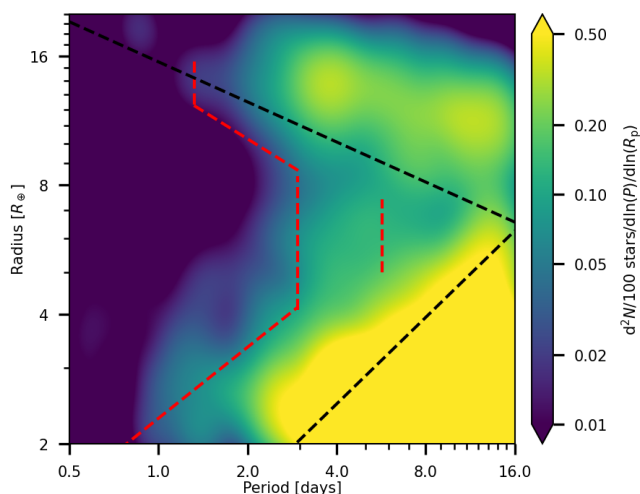
To account for uncertainties in completeness, we also sample the completeness maps from Beta distributions:  $Q(\omega) \sim \text{Beta}(\alpha, \beta)$ , as described in Section 3.5.1. We assign the RAVEN probability of each candidate as  $r_k$ , consistent with IDEM. The posteriors of the planets' physical parameters are fitted and sampled using JULIET (N. Espinoza, D. Kossakowski & R. Brahm 2019), with normal distributions as priors  $p(\omega_k^{(n)})$ . These normal distributions use the fitted BLS results as the means and set the standard deviations to 0.1 for both  $P$  and  $R_p$ . Detailed light curve processing and modelling fitting is available in Section 4.1.

We then perform Markov Chain Monte Carlo (MCMC) sampling of our likelihood function using EMCEE (D. Foreman-Mackey et al. 2013). The MCMC is run for a number of steps exceeding 50 times the autocorrelation time to ensure convergence. Running can be slow due to the large number of posterior samples of physical parameters and completeness samples, so we limit the number of posterior sample to 1000. Also, MCMC sampling may struggle to converge with more than 50 bins. Therefore, we divide the period–radius space into multiple patches if more than 50 bins are required, run them separately, and then combine them together. The division is only applied along the orbital period axis because the uncertainties in period are significantly low and can be ignored, so the covariance of occurrence rates is not affected.

## 4 RESULTS

### 4.1 Planet candidates

From our TESS-SPOC FGK main-sequence stellar sample, we first run our detection pipeline for all of them as described in Section 3.3. Then, using the RAVEN pipeline (Section 3.4), we identify 1361 planet candidates from our stellar sample, with periods and



**Figure 4.** wKDE in period and radius space, with the colour scale normalized logarithmically for better visibility and higher dynamic range. Two Neptunian desert boundaries and a potential ‘ridge’ region estimated from *Kepler* are shown with dashed black lines (T. Mazeh et al. 2016) and dashed red lines (A. Castro-González et al. 2024). The bandwidth of KDE is 0.3, which is manually tuned for clearly showing the general features like hot Jupiters and desert boundaries. The range of our colour scale is limited to match that of Fig. 5 to have consistent visibility.

radii ranging from 0.5 to 16 d and 2 to 20  $R_{\oplus}$ , respectively. However, the sample is not perfectly clean, particularly for candidates exhibiting orbital period aliases.

As discussed in Section 3.4, our models are not designed to handle period aliases for planets; an incorrect period would introduce biases into our occurrence estimates. We therefore manually inspect the data by folding the light curve at different aliases of the detected period, excluding only candidates that display clear period aliasing. This yields a final sample of 1301 planet candidates. The number can vary depending on the models and thresholds applied, but the resulting occurrence rates should remain consistent. This consistency arises because our occurrence rate model explicitly accounts for detection completeness and candidate reliability, both of which change alongside different models and thresholds. We further assess the statistical robustness of our occurrence rates in Section 5.1 by applying various models and thresholds.

As introduced in Section 3.5.2, we account for the parameter uncertainties of our planet candidates to achieve a robust occurrence rate estimation. This is accomplished using JULIET (N. Espinoza et al. 2019), which incorporates GP light-curve modelling, transit modelling, and nested sampling. Detailed parameter settings and prior distributions are provided in Lafarga et al. (submitted). Both the nested sampling posteriors and priors are stored for our occurrence rate calculations.

#### 4.2 TESS planetary occurrence rates

Given the final set of planet candidates, together with our combined completeness maps and reliabilities, we first estimate the occurrence rate distribution via the wKDE method (Section 3.5.1). Fig. 4 shows the wKDE result. We can clearly see the differences in occurrence rate density across various regions of the parameter space. Known features such as the Neptunian desert, together with planet populations like HJs and the abun-

dance of subNeptunes, are all distinctly recovered. Although the wKDE provides a quick and intuitive view of the occurrence-rate distribution, we still need a more statistically robust model to quantify occurrence rates and compare them with results from the literature.

Thus, we quantify the occurrence rate within our *TESS* FGK main-sequence stellar sample using the Bayesian method described in Section 3.5.2. Fig. 5 presents the occurrence rates on the orbital period–radius plane, using a 10-by-10 grid with both axes in logarithmic scale. The value in each bin represents the average occurrence as a percentage, that is, the number of planets per 100 stars. Generally, our quantified result shows consistent features with the wKDE result in Fig. 4. The overall occurrence in our parameter space ( $0.5 < P < 20$  d,  $2 < R_p < 20 R_{\oplus}$ ) is  $9.4^{+0.7}_{-0.6}$  per cent. The direct comparison with *Kepler* is conducted in Section 5.3.

This occurrence map reveals several notable features while providing direct quantitative estimates of occurrence rates for regions such as the distinct Neptunian desert, the piled-up HJs, and the abundance of sub-Neptunes. However, within the scope of this paper, we will focus briefly on HJs and the Neptunian desert, comparing our results with previous studies in Section 4.4. Detailed investigations of the those features will be reserved for future work.

The occurrence rate for each bin is provided in the Appendix Table A1. However, we also offer an online interactive plot<sup>6</sup> where users can click and select any region using the tools like box select or lasso select on the side toolbar to instantly obtain the total occurrence rate and associated uncertainties from the title. We recommend that readers use this tool rather than manually calculating the sum and propagating errors. We have also made the posterior distributions of occurrence rate density for each cell available online,<sup>7</sup> facilitating future applications and comparisons.

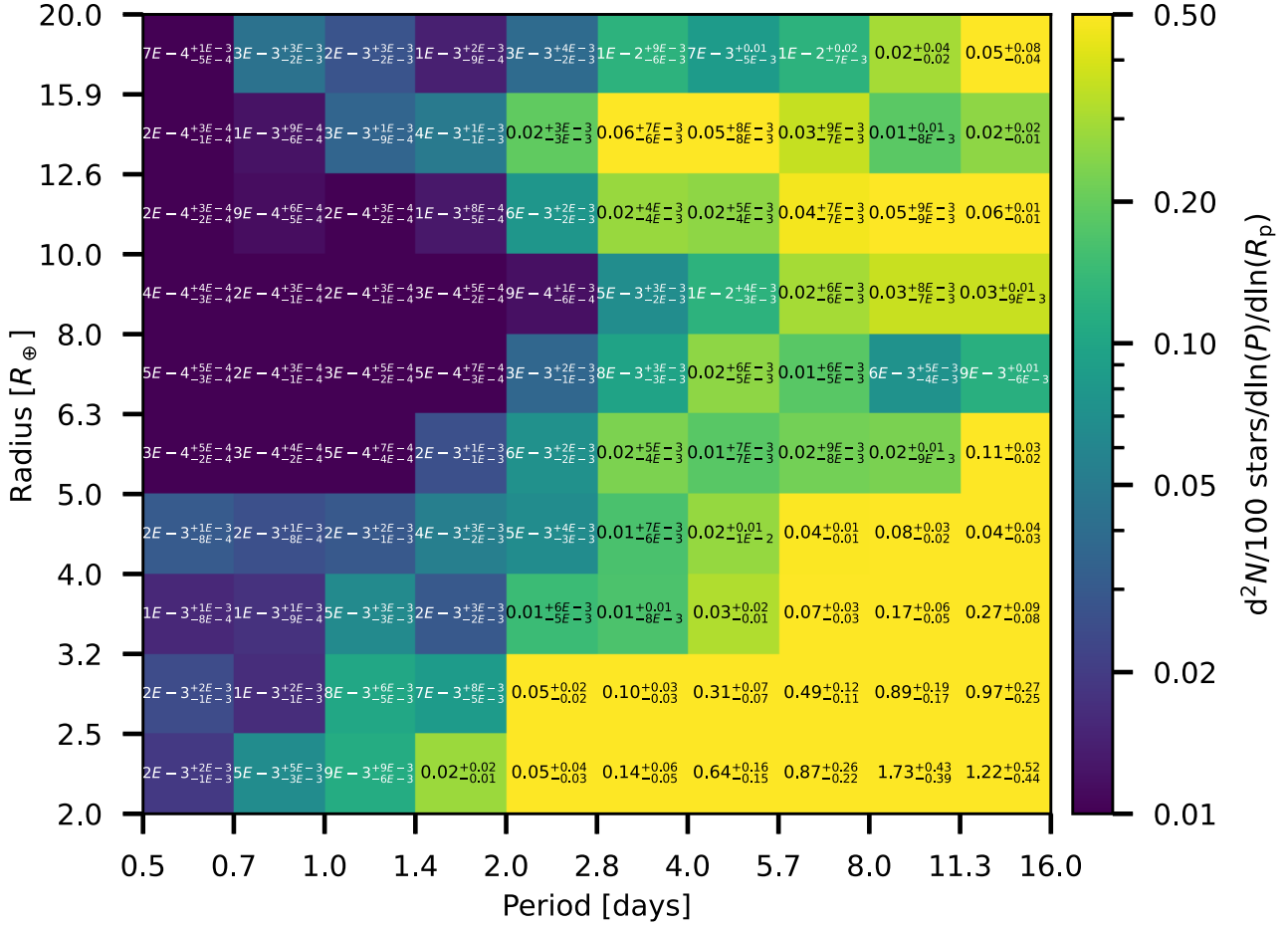
#### 4.3 Gaia 100 pc volume-limited occurrence rate

Enabled by *Gaia*’s precise astrometry results and strategically selecting stars within 100 pc by *TESS*-SPOC, here we tentatively present an analysis of volume-limited planetary occurrence rates for FGK main-sequence stars, which is an important comparison that has rarely been conducted before. Utilizing data from *Gaia* and *TESS*, we achieve this by cross-matching the *Gaia* 100 pc sample with the TIC catalogue. The *Gaia* Nearby Star Catalogue (R. L. Smart et al. 2021) comprises 331 312 objects within 100 pc. We perform a cross-match with TIC using an 8 arcsec radius and selected the closest matches, resulting in 329 921 common objects, a completeness rate of approximately 99.6 per cent.

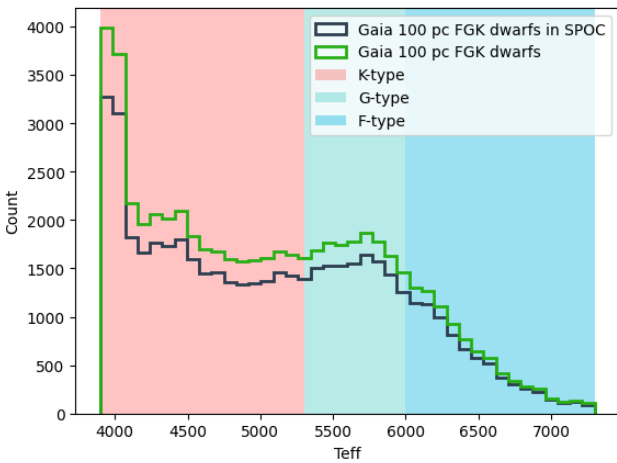
Subsequently, we identified FGK main-sequence stars (with effective temperatures between 3900–7300 K and  $\log g > 3.5$ ) based on TIC parameters, yielding a subset of 56 197 stars. Of these, *TESS*-SPOC observed 48 622 within Sectors 1–55. We also evaluated the completeness of our cross-matched sample across different temperature ranges as illustrated in Fig. 6. Although the number of *TESS*-SPOC cross-matched samples is relatively lower than that of *Gaia*’s FGK dwarfs within the same distance range (100 pc), quantitatively, the completeness for F, G, and K types all remain above 85 per cent.

<sup>6</sup><https://cuikaiming.com/TESS-SPOC-Occurrence-Rate/>

<sup>7</sup>[10.5281/zenodo.17804280](https://doi.org/10.5281/zenodo.17804280)



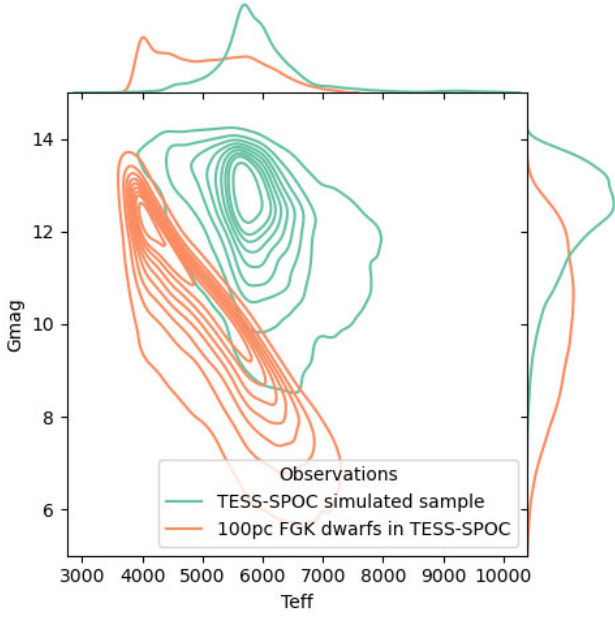
**Figure 5.** Occurrence rates of planet candidates around *TESS*-SPOC FGK main-sequence stars. Each cell displays the average occurrence as a percentage (number of planets per 100 stars), along with the corresponding uncertainties (the difference between the 16th and 84th percentiles and the median). For clarity, text in cells with an occurrence rate below 0.01 per cent is displayed in white. The colour of each cell encodes the occurrence rate density,  $d^2N_p/(d \ln P d \ln R_p)$ , per 100 stars. The logarithmic colour scale is truncated at 0.5 to enhance visibility. We also provide an [online interactive plot](#) that allows users to easily select any region and obtain its total occurrence rate.



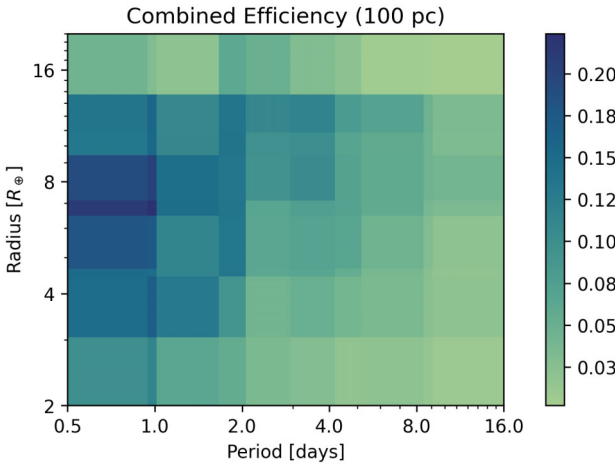
**Figure 6.** Temperature distributions of FGK main-sequence stars within 100 pc from *Gaia* data, compared to those observed by *TESS*-SPOC across the 1–55th sectors. Shaded regions represent the temperature boundaries for different FGK spectral types.

We then apply the same pipeline to the observed volume-limited sample, using the non-uniform FP models, setting the NSFP, FPs, and SNR thresholds to 0.9, 0.9, and 10, respectively. This yields 125 planetary candidates within our parameter space range ( $0.5 < P < 16$  d,  $2 < R_p < 20 R_\oplus$ ). We further fit their planetary parameters and estimate the occurrence rate by incorporating the completeness map and using the Bayesian method as previously described. Notably, because of the significant differences in stellar properties between magnitude-limited and volume-limited samples, all completeness maps must be recalculated. The 100 pc sample, in particular, contains a higher proportion of K dwarfs, and its stars tend to be apparently brighter. Fig. 7 shows effective temperature and *Gaia* Gmag distributions of the observed 100 pc sample and our simulated sample.

To address this, importance sampling is applied to our simulated sample to match the effective temperature and Gmag distributions of the 100 pc sample: the simulated sample is weighted by the ratio between the target (100 pc sample) probability density and our simulated probability density. Effective temperature and Gmag are used because, for main-sequence stars, most other stellar properties, such as stellar radius, light curve noise levels, and crowdings are either closely related to these two parameters or can be assumed to have the same distribution in the simulated



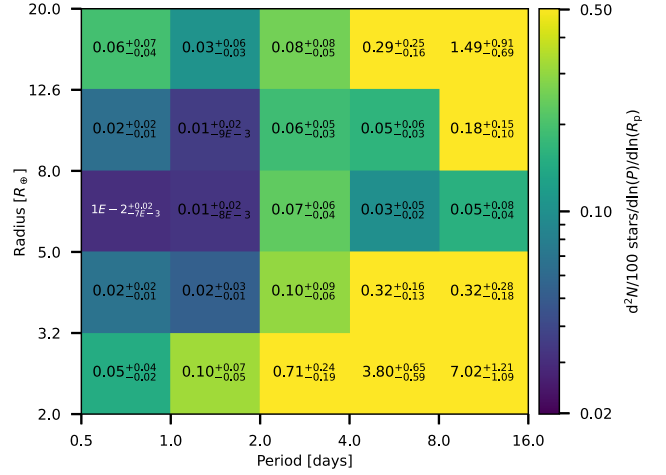
**Figure 7.** 2D KDE contours showing the effective temperature and *Gaia* *G*-band magnitude distributions for our *TESS*-SPOC simulated sample and for *Gaia* 100 pc FGK main-sequence stars observed by *TESS*-SPOC.



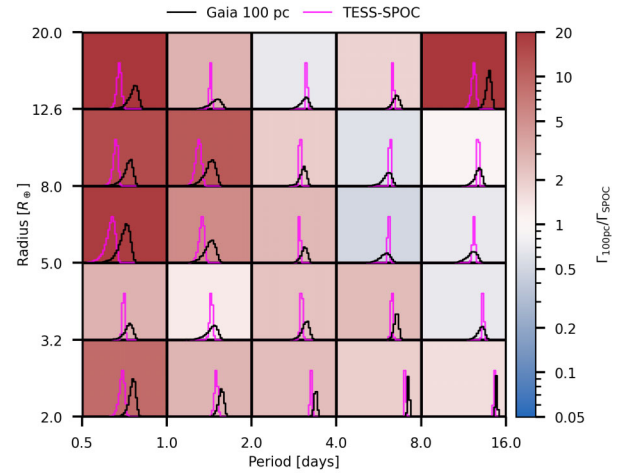
**Figure 8.** Combined completeness of *Gaia* 100 pc FGK main-sequence stars.

and real samples. Note that there are a few faint stars ( $G_{\text{mag}} > 14$ ) in the 100 pc sample that are not included in our *TESS*-SPOC simulated sample, so their contributions to the completeness may not be accurately modelled. Fig. 8 shows the combined completeness map for the 100 pc sample. We reduce the number of bins to mitigate fluctuations caused by a few high-weight points.

Following the same procedure, we estimate the occurrence rate for the 100 pc sample. However, due to importance sampling, the numbers for injection and recovery are not feasible to obtain. Therefore, we use the expected values of completeness directly, without modelling them with Beta distributions. Fig. 9 presents the occurrence rate distributions, and the overall occurrence rate in our parameter space is  $15.4_{-1.5}^{+1.6}$  per cent, which is higher than our *TESS*-SPOC result. In detail, Fig. 10 compares the posteriors



**Figure 9.** Similar to Fig. 5, occurrence rates of *Gaia* 100 pc sample.



**Figure 10.** Posterior distributions of planetary occurrence rates for both the *Gaia* 100 pc sample and our *TESS*-SPOC sample. The black and fuchsia lines represent the *Gaia* 100 pc and *TESS*-SPOC samples, respectively. In each cell, the *x*-axis of occurrence rates ranging from  $1 \times 10^{-6}$  per cent to 50 per cent with logarithmic scale, and the heights of the distributions are scaled according to the highest histogram in each cell. The background diverging colour map indicates the ratio of occurrence rates between the two samples. The colour scale is manually centred at one; the actual minimum ratio is approximately 0.4.

of our magnitude-limited and volume-limited samples within each bin. The colour map in Fig. 10 is scaled so that the white midpoint represents where the *Gaia* 100 pc and *TESS*-SPOC median occurrence rates are equal. On this scale, bluish colours indicate that the *TESS*-SPOC occurrence rate is higher than that of *Gaia* 100 pc, whereas reddish colours indicate that the *Gaia* 100 pc occurrence rate is higher than that of *TESS*-SPOC.

As expected, the *Gaia* 100 pc occurrence rate shows larger uncertainties because of limited sample size. In most bins, the *Gaia* 100 pc sample generally shows higher occurrence rates, although these rates are consistent within  $1\sigma$ . This elevation can be attributed to the sample's larger proportion of K-type stars, reflecting the known temperature dependence of planetary occurrence

rates (e.g. M. Kunimoto & J. M. Matthews 2020). The especially high-occurrence rates in the 100 pc sample for shorter periods ( $< 1$ -2 d) are due to the very small number of detections and the limited size of the stellar sample, which is much smaller than that of the *TESS*-SPOC sample. Because a smaller stellar sample sets a relatively high lower limit on the minimum measurable occurrence rate, even a few detected planets can result in a noticeably higher inferred rate compared to the *TESS*-SPOC sample, which contains far more stars.

#### 4.4 Hot Jupiters and the Neptunian desert

As one of the most prominent features in planet occurrence rate distributions, the occurrence of hot Jupiters has been widely studied. A. W. Howard et al. (2012) found  $0.4 \pm 0.1$  per cent for *Kepler* GK-type stars. Later, F. Fressin et al. (2013) reported  $0.43 \pm 0.05$  per cent for *Kepler* FGK dwarfs, and A. Santerne et al. (2016) obtained  $0.47 \pm 0.08$  per cent for *Kepler* F5–K5 V-type stars. E. A. Petigura et al. (2018) estimated  $0.57^{+0.14}_{-0.12}$  per cent for *Kepler* FGK-type stars. G. Zhou et al. (2019) measured an HJ occurrence rate of  $0.41 \pm 0.10$  per cent for early *TESS* AFG-type stars; subsequently, M. Beleznyay & M. Kunimoto (2022) derived  $0.33 \pm 0.04$  per cent for AFG-type stars in a larger sample. Studies of M dwarfs (e.g. E. M. Bryant et al. 2023; T. Gan et al. 2023) indicate lower occurrence rates toward lower stellar masses. Others have examined HJ occurrence as a function of metallicity, age, and evolutionary stage (e.g. X. Guo et al. 2017; A. Osborn & D. Bayliss 2020; S. Miyazaki & K. Masuda 2023; M. Temmink & I. A. G. Snellen 2023; S. W. Yee & J. N. Winn 2023).

We refine our HJ parameter space to  $0.8 < P < 10$  d and  $8 < R_p < 20 R_{\oplus}$ , identifying 432 HJ candidates in our *TESS*-SPOC sample. This number is broadly consistent with the predictions of S. W. Yee, J. N. Winn & J. D. Hartman (2021). Using the same method, we derive an overall HJ occurrence rate of  $0.39^{+0.03}_{-0.02}$  per cent. This value is fully consistent with previously reported results for the *Kepler* FGK and some *TESS* AFG samples. In our analysis of the *Gaia* 100 pc sample, we also estimate an overall HJ occurrence rate of  $0.42^{+0.16}_{-0.12}$  per cent. This value with its relative larger uncertainty is in agreement with the above estimations within  $1\sigma$ .

Due to the low number of planets in the Neptunian desert, as far as we are aware, no previous work has directly reported its overall occurrence rate. However, we can estimate the Neptunian desert occurrence rate in the *Kepler* data using our posterior distribution derived from D. C. Hsu et al. (2019). Adopting the desert boundary defined by A. Castro-González et al. (2024) and assuming that the occurrence rate within each bin is uniformly distributed on a logarithmic scale, we can calculate the overall Neptunian desert occurrence rate from *Kepler* data. Since much of the desert region remains undetected, we adopt the 95th percentile as a conservative upper limit, which is 0.69 per cent. Within the boundary, our sample has 108 candidates, and applying the same desert boundary to our posterior distributions yields an overall occurrence rate for *TESS*-SPOC of  $0.08 \pm 0.01$  per cent, with a 95th percentile upper limit of 0.10 per cent. Our results provide the first direct constraint on the overall occurrence rate of the Neptunian desert within these boundaries. Our results are consistent with the *Kepler* upper limit, in contrast with the potentially higher desert occurrence rate in *TESS* data reported in the preliminary results of E. J. Melton et al. (2024).

## 5 DISCUSSION

### 5.1 Effect of different models and thresholds

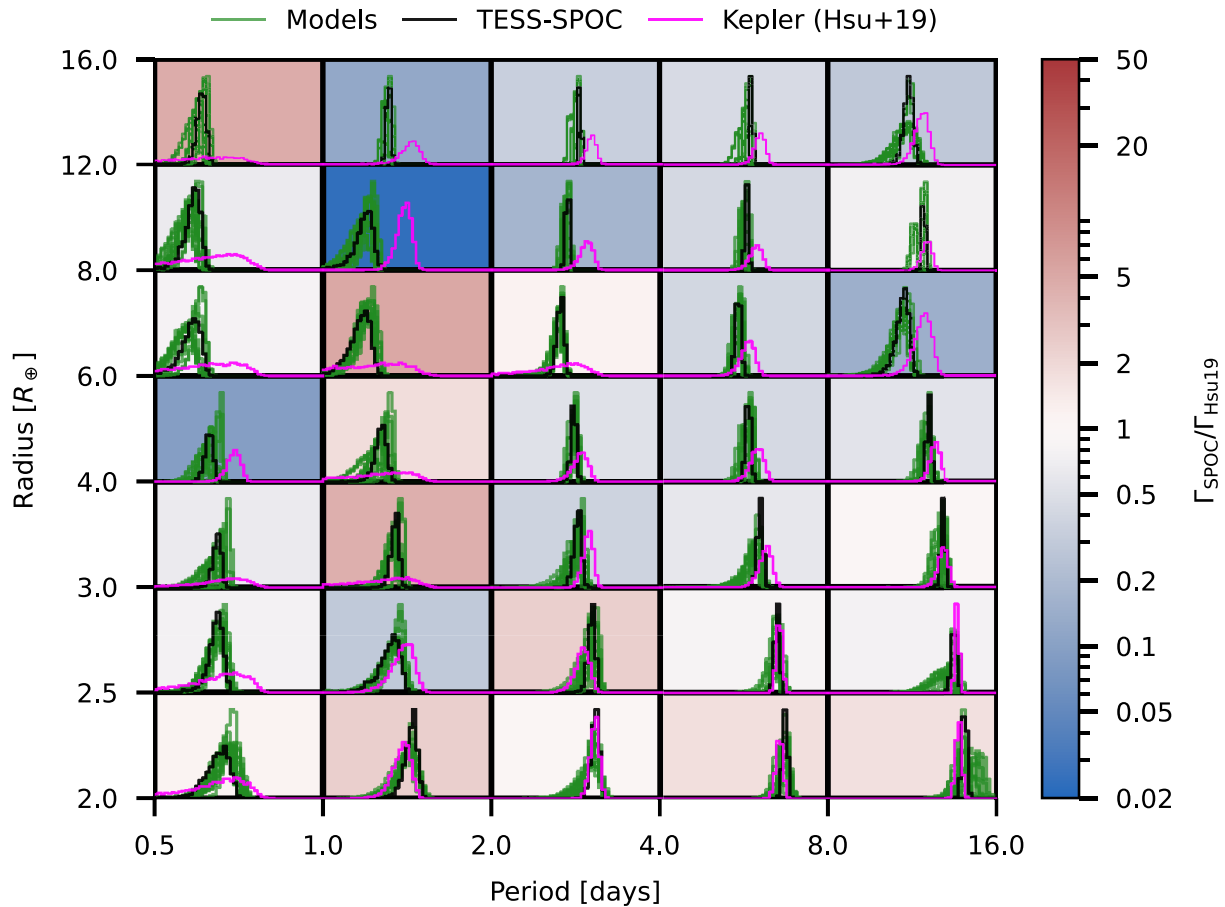
Our occurrence rate calculations rely heavily on our vetting and validation models; they affect both the number of detected planets and the completeness estimates. Nevertheless, the occurrence rate estimation process should yield robust results under the following ideal conditions: (1) The stellar properties in our simulations match those of the observed stellar sample; (2) the simulated sample is processed by the same pipeline as the observed sample, with no data leakage; and (3) the vetting and validation models produce statistically reliable posterior probabilities.

To meet these requirements: (1) We select the simulated sample from the same sources as the observed stellar sample, removing only binaries and known planets; (2) we evaluate completeness with the same pipeline used for detection, maintain a clear training/test split, and never use the observed light curves when training and evaluating models (except for NSFP, which are sampled from observed BLS peaks, but their number is limited); and (3) although it is impossible to know the true posterior probabilities of candidates, except for confirmed planets and FPs, our machine learning models are well calibrated, supporting the reliability of their probabilities. In addition, adopting a relatively high-decision threshold helps ensure that the selected candidates can be treated as confirmed planets.

Based on the discussion above, we test the robustness of our occurrence rate estimates by varying the models and thresholds. Specifically, we explore two validation models (uniform and non-uniform), FP thresholds of 0.90, 0.95, and 0.99, and SNR thresholds of 10, 15, and 20, resulting in 18 distinct combinations. For each of these 18 cases, we estimate the occurrence rate using the same hierarchical Bayesian procedure and obtain their corresponding posterior distributions. For convenience, we use the same bin configuration as D. C. Hsu et al. (2019), since we will be comparing our posteriors with theirs as well (see Section 5.3). Fig. 11 shows the posterior distributions of the occurrence rates for all combinations considered, highlighting both our selected threshold and the results of D. C. Hsu et al. (2019).

We measure the relative scattering of occurrence rate posterior distributions by calculating the median absolute deviation (MAD) of median occurrence rates over the median uncertainty for posteriors in each cell. Overall, our posterior distributions remain self-consistent across most threshold combinations in most bins, only four bins show the MAD over median uncertainties greater than one. These cases can be found in Fig. 11 by their increased scatter and always occur at the edges of the parameter space. We highlight the most pronounced discrepancies in Appendix B, and we show that the main differences arise from using excessively high FP threshold or uniform models.

In principle, as discussed, the occurrence rate should remain consistent regardless of the FP or SNR threshold and model used. However, at higher thresholds, estimating completeness becomes difficult because the recovered simulated sample size in some bins is very small, sometimes zero, leading to highly uncertain estimates. A similar issue arises for the uniform models: Assuming planets occur uniformly in orbital period and planetary radius space, the simulated uniform sample is dominated by observational bias, yielding very few observable planets at longer periods and smaller radii. Consequently, the simulated sample used to estimate FP completeness can be extremely limited, sometimes



**Figure 11.** Similar to Fig. 10, this figure shows the posterior distributions of planetary occurrence rates for our *TESS*-SPOC sample using various thresholds, alongside *Kepler* results. The fuchsia lines represent the *Kepler* findings from D. C. Hsu et al. (2019), while the green curves correspond to our model’s results across 18 different threshold combinations. The black line denotes our selected model, which uses a non-uniform model with NSFP and FP thresholds of 0.9 and an SNR threshold of 10. In each cell, the  $x$ -axis of occurrence rates ranging from  $1 \times 10^{-5}$  per cent to 100 per cent with logarithmic scale. The background diverging colour map shows the ratio of occurrence rates between our chosen model and the *Kepler* results. The colour map scale is manually centred at one; the actual maximum ratio is approximately five.

nearly zero. Thus it suffers from both too few detections and too few injections. This is evident in Fig. 3, where the bottom-right bin shows zero completeness.

Given these results and analyses, we can confidently state that our occurrence rate estimates are robust.

## 5.2 Effect of RUWE

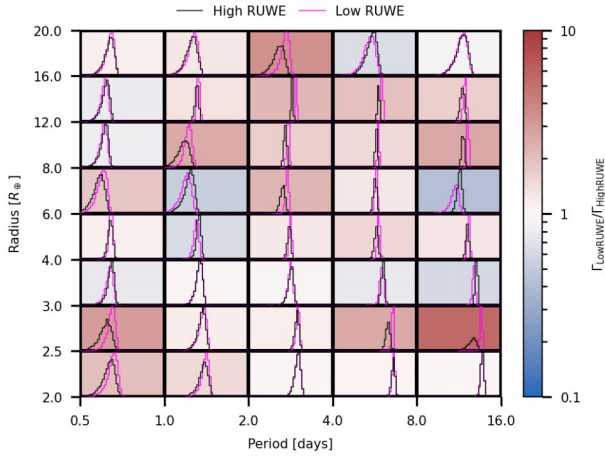
As a commonly used binarity metric, RUWE may indicate different planetary occurrence rates between low and high values. To quantify differences in occurrence rates between low-RUWE ( $<1.05$ ) and high-RUWE ( $>1.05$ ) stars, we divide our stellar sample into these two groups, resulting in 54 per cent low-RUWE stars and 46 per cent high-RUWE stars. In our RAVEN pipeline, RUWE affects our HTP and HEB priors: when RUWE is less than 1.05, we reduce the HTP and HEB priors by preset factors, and when RUWE exceeds 1.4, we increase the HTP and HEB priors by specific factors (detailed in A. Hadjigeorgiou et al. 2025). Thus, we calculate completeness by applying the low-RUWE prior to low-RUWE cases and the high-RUWE prior ( $>1.05$ ) to high-RUWE cases. Next, we separate our high-confidence planet candidates into low RUWE and high RUWE groups and apply

the same occurrence rate estimation methodology to each. This allows us to compare their occurrence rates in detail.

Fig. 12 shows a comparison of planetary occurrence rates between stars with low- and high-RUWE values. Even though all their differences are within one magnitude, the high RUWE sample generally exhibits lower occurrence rates. This trend may be related to the prevalence of unresolved binaries, which can both influence RUWE measurements and impact planet detection. Our results suggest that unresolved binaries may suppress planet formation or reduce the likelihood of detecting planets in these systems. Further investigation into the relationship between RUWE, binarity, and planetary systems will require a deeper understanding of the RUWE parameter, which is expected to improve with the release of *Gaia* DR4.

## 5.3 *TESS* and *Kepler* occurrence rate comparison

There are many occurrence rate estimates based on *Kepler* FGK stars (e.g. D. C. Hsu et al. 2019; M. Kunimoto & J. M. Matthews 2020). Here, we compare our results with those of D. C. Hsu et al. (2019) within the same parameter space range ( $0.5 < P < 16$  d,  $2 < R_p < 16 R_{\oplus}$ ). Since D. C. Hsu et al. (2019) only provides



**Figure 12.** Similar to Fig. 10, this figure shows the posterior distributions of planetary occurrence rates for low-RUWE ( $<1.05$ ) and high-RUWE ( $>1.05$ ) stars. The black lines indicate the high-RUWE sample and the fuchsia lines indicate the low-RUWE sample. The background diverging colour maps the ratio of median occurrence rates between low-RUWE and high-RUWE sample.

occurrence rates with uncertainties at the 16th and 84th percentiles, we fit these uncertainty percentiles with Gamma distributions to directly compare their occurrence rate posteriors with ours. For occurrences where D. C. Hsu et al. (2019) only provides upper limits due to zero detections, we assume the mean of a Gamma distribution is half of the 84th percentile value, which roughly matches their posterior distributions shown in fig. 5 of their paper.

For the overall occurrence rate, D. C. Hsu et al. (2019) gives  $11.4^{+1.2}_{-1.1}$  per cent (this value is estimated from the above posteriors; the direct sum over bins with non-zero detections yields 11.0 per cent. Since we do not have access to their posterior samples, small differences may be present). Our *TESS*-SPOC occurrence rate estimation within  $2 < R_p < 16 R_{\oplus}$  is  $9.2^{+0.7}_{-0.6}$  per cent, consistent within  $1.5\sigma$ . We also adopt the largest common parameter range from M. Kunimoto & J. M. Matthews (2020,  $0.75 < P < 12.5$  d,  $2 < R_p < 16 R_{\oplus}$ ). Compared to their overall median value of 6.7 per cent (not accounting for the upper limit in the desert region), our result of  $7.2 \pm 0.5$  per cent is consistent within  $1\sigma$ .

To have a detailed comparison, we use the same bin configuration as D. C. Hsu et al. (2019) and run our Bayesian model as described in Section 4.2. Fig. 11 compares the posterior distributions of occurrence rate estimates for each bin. The colour scale of this plot is set to be one when the occurrence of *TESS*-SPOC and *Kepler* are equal. Thus the bluish bins show our occurrence rates are lower than *Kepler*, while the reddish colour indicate our *TESS*-SPOC occurrence rate is higher than *Kepler*.

From the Fig. 11, we can see that our occurrence rate posterior distributions are significantly tighter than those of D. C. Hsu et al. (2019) in most bins. On average, the  $1\sigma$  uncertainty ranges of our posteriors (defined as the difference between the 84 per cent and 16 per cent percentiles) are roughly an order of magnitude smaller than those reported by D. C. Hsu et al. (2019). This improvement is particularly evident near the Neptunian desert region. Within the desert, our posteriors provide better constraints, reducing some of the uncertainties present in that region. For example, in the bins with  $0.5 < P < 1.0$  d,  $4.0 < R_p < 6.0 R_{\oplus}$  and

$1.0 < P < 2.0$  d,  $8.0 < R_p < 12.0 R_{\oplus}$ , D. C. Hsu et al. (2019) report anomalously high-occurrence rates compared to their surrounding bins, whereas our results show lower values that are more consistent with those of the adjacent desert regions, resulting in a more well-defined Neptunian desert.

#### 5.4 Future prospects

Given our comprehensive pipeline and occurrence rate results, numerous directions remain to be explored. For instance, future studies can investigate occurrence rates across various stellar types, metallicities, and evolutionary stages, with particular attention to sub-Neptunes and hot Jupiters. Additionally, it will be valuable to precisely quantify the boundaries of the Neptunian desert, as well as determine the locations and significance of the ridge and savanna (V. Bourrier et al. 2023; A. Castro-González et al. 2024), and to compare these findings with theoretical models of planet formation and evolution. We plan to address these topics in future research.

Additionally, extending our pipeline to more *TESS* FFI observations, specifically to sectors beyond sector 55, is a straightforward way to expand the sample size and obtain higher precision.

## 6 CONCLUSION

In this study, we measure the occurrence rates of *TESS* close-in exoplanets orbiting FGK main-sequence stars. Our main results are summarized as below:

(1) We measured the occurrence of close-in exoplanets around FGK main-sequence stars using four years of *TESS*-SPOC observations. Our analysis spans orbital periods of 0.5–16 d and radii of  $2\text{--}20 R_{\oplus}$ , evaluated on a  $10 \times 10$  grid in period–radius space. The results are robust, as confirmed by our tests.

(2) For the magnitude-limited *TESS*-SPOC sample, the overall occurrence rate is  $9.4^{+0.7}_{-0.6}$  per cent, and the distributions are consistent with the *Kepler*-based estimates (D. C. Hsu et al. 2019) across most of the commonly studied parameter space.

(3) The *Gaia* Nearby Star Catalogue, volume-limited subsample within 100 pc exhibits a similar structure in the occurrence map but systematically higher rates with larger uncertainties, yielding an overall occurrence of  $15.4^{+1.6}_{-1.5}$  per cent.

(4) In the HJ regime, we find an occurrence rate of  $0.39^{+0.03}_{-0.02}$  per cent in the full *TESS*-SPOC sample. Our results are consistent with *Kepler* FGK results within  $1\sigma$ . These findings are also consistent with those from the 100 pc sample, which has an HJ occurrence rate density of  $0.42^{+0.16}_{-0.12}$  per cent.

(5) For the Neptunian desert, adopting the boundary from A. Castro-González et al. (2024), we find an overall occurrence rate of  $0.08 \pm 0.01$  per cent, with a 95th percentile upper limit of 0.10 per cent. Our results significantly enhance the precision of occurrence rate estimates made possibly by *TESS*.

(6) Together, these comparisons demonstrate that our systematic detection, validation, and occurrence rate estimation reliably provide stronger constraints on the demographics of close-in exoplanets orbiting FGK main-sequence stars.

*TESS*-SPOC provides a powerful platform for exoplanet statistics. Extending to additional sectors, broader period–radius ranges, and higher-dimensional dependences (e.g. stellar mass and metallicity), alongside continued refinement of our completeness and reliability models, will sharpen the empirical constraints and test theoretical predictions.

## ACKNOWLEDGEMENTS

This paper includes data collected by the *TESS* mission. Funding for the *TESS* mission was provided by the NASA Explorer Programme. Resources supporting this work were provided by the NASA High-End Computing (HEC) Programme through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center for the production of the SPOC data products. This research was funded by the UKRI (Grants ST/X001121/1 and EP/X027562/1). Some of the language has been polished with the assistance of OpenAI API Service (OpenAI, submitted).

## DATA AVAILABILITY

The *TESS*-SPOC data used in this paper can be found in Mikulski Archive for Space Telescopes (D. A. Caldwell et al. ). Our *TESS*-SPOC occurrence rate posterior sample is available (K. Cui 2025).

## REFERENCES

- Armstrong D. J., Pollacco D., Santerne A., 2017, *MNRAS*, 465, 2634  
 Armstrong D. J., Gamper J., Damoulas T., 2021, *MNRAS*, 504, 5327  
 Beaugé C., Nesvorný D., 2012, *ApJ*, 763, 12  
 Beleznyay M., Kunimoto M., 2022, *MNRAS*, 516, 75  
 Borucki W. J. et al., 2010, *Science*, 327, 977  
 Bourrier V. et al., 2023, *A&A*, 669, A63  
 Brown A. G. A. et al., 2018, *A&A*, 616, A1  
 Bryant E. M., Bayliss D., Van Eylen V., 2023, *MNRAS*, 521, 3663  
 Bryson S., Coughlin J. L., Kunimoto M., Mullally S. E., 2020, *AJ*, 160, 200  
 Caldwell D. A. et al., 2020, *Res. Notes AAS*, 4, 201  
 Castro-González A., Bourrier V., Lillo-Box J., Delisle J.-B., Armstrong D. J., Barrado D., Correia A. C. M., 2024, *A&A*, 689, A250  
 Chen T., Guestrin C., 2016, Association for Computing Machinery. San Francisco, California, USA, p. 785  
 Cui K., 2025, Sample of Posterior Distributions of Paper ‘Demographics of Close-In TESS Exoplanets Orbiting FGK Main-sequence Stars’. Zenodo. Available at: <https://doi.org/10.5281/zenodo.16785938>  
 Cumming A., Butler R. P., Marcy G. W., Vogt S. S., Wright J. T., Fischer D. A., 2008, *PASP*, 120, 531  
 Dattilo A., Batalha N. M., 2024, *AJ*, 167, 288  
 Dattilo A., Batalha N. M., Bryson S., 2023, *AJ*, 166, 122  
 Dawson R. I., Johnson J. A., 2018, *ARA&A*, 56, 175  
 Díaz R. F., Almenara J. M., Santerne A., Moutou C., Lethuillier A., Deleuil M., 2014, *MNRAS*, 441, 983  
 Doyle L., Armstrong D. J., Bayliss D., Rodel T., Kunovac V., 2024, *MNRAS*, 529, 1802  
 Drażkowska J. et al., 2023, ASP Conf. Ser. Vol. 534, Protostars and Planets VII. Astron. Soc. Pac., San Francisco, p. 717  
 Dressing C. D., Charbonneau D., 2015, *ApJ*, 807, 45  
 Espinoza N., Kossakowski D., Brahm R., 2019, *MNRAS*, 490, 2262  
 Etzel P. B., 1981, in Carling E. B., Kopal Z., eds, *Proceedings of the NATO Advanced Study Institute, Photometric and Spectroscopic Binary Systems*. D. Reidel Publishing Co., Dordrecht, p. 111  
 Fernandes R. B. et al., 2022, *AJ*, 164, 78  
 Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306  
 Foreman-Mackey D., Hogg D. W., Morton T. D., 2014, *ApJ*, 795, 64  
 Fressin F. et al., 2013, *ApJ*, 766, 81  
 Fulton B. J. et al., 2017, *AJ*, 154, 109  
 Fulton B. J. et al., 2021, *ApJS*, 255, 14  
 Gan T. et al., 2023, *AJ*, 165, 17  
 Giacalone S. et al., 2020, *AJ*, 161, 24  
 Guerrero N. M. et al., 2021, *ApJS*, 254, 39  
 Guo X., Johnson J. A., Mann A. W., Kraus A. L., Curtis J. L., Latham D. W., 2017, *ApJ*, 838, 25  
 Hadjigeorgiou A., Armstrong D. J., 2024, *MNRAS*, 527, 4018  
 Hadjigeorgiou A., Armstrong D. J., Cui K., Magro M. L., Nieto L. A., Díaz R. F., Doyle L., Kunovac V., 2025, *preprint (arXiv:2509.17645)*  
 Howard A. W. et al., 2012, *ApJS*, 201, 15  
 Hsu D. C., Ford E. B., Ragozzine D., Morehead R. C., 2018, *AJ*, 155, 205  
 Hsu D. C., Ford E. B., Ragozzine D., Ashby K., 2019, *AJ*, 158, 109  
 Hsu D. C., Ford E. B., Terrien R., 2020, *MNRAS*, 498, 2249  
 Huang C. X. et al., 2018, *ApJ*, 868, L39  
 Jeffreys H., 1946, *Proc. R. Soc.*, 186, 453  
 Jenkins J. M. et al., 2016, in Chiozzi G., Guzman J. C., eds, *Proc. SPIE, Vol. 9913, The TESS Science Processing Operations Center*. SPIE, Edinburgh, UK, p. 99133E  
 Jin S., Mordasini C., Parmentier V., van Boekel R., Henning T., Ji J., 2014, *ApJ*, 795, 65  
 Kohonen T., 1982, *Biol. Cybern.*, 43, 59  
 Kovács G., Zucker S., Mazeh T., 2002, *A&A*, 391, 369  
 Kunimoto M., Bryson S., 2020, *Res. Notes AAS*, 4, 83  
 Kunimoto M., Matthews J. M., 2020, *AJ*, 159, 248  
 Lopez E. D., Fortney J. J., 2013, *ApJ*, 776, 2  
 Mazeh T., Holczer T., Faigler S., 2016, *A&A*, 589, A75  
 Melton E. J., Feigelson E. D., Montalto M., Caceres G. A., Rosenswie A. W., Abelson C. S., 2024, *AJ*, 168, 271  
 Ment K., Charbonneau D., 2023, *AJ*, 165, 265  
 Miyazaki S., Masuda K., 2023, *AJ*, 166, 209  
 Moe M., Stefano R. D., 2017, *ApJS*, 230, 15  
 Mordasini C., Alibert Y., Benz W., 2009, *A&A*, 501, 1139  
 Morton T. D., Swift J., 2014, *ApJ*, 791, 10  
 Morton T. D., Bryson S. T., Coughlin J. L., Rowe J. F., Ravichandran G., Petigura E. A., Haas M. R., Batalha N. M., 2016, *ApJ*, 822, 86  
 Mulders G. D., Pascucci I., Apai D., Ciesla F. J., 2018, *AJ*, 156, 24  
 Mulders G. D., Mordasini C., Pascucci I., Ciesla F. J., Emsenhuber A., Apai D., 2019, *ApJ*, 887, 157  
 Müller S., Baron J., Helled R., Bouchy F., Parc L., 2024, *A&A*, 686, A296  
 Nelson B., Davis W. D., 1972, *ApJ*, 174, 617  
 Osborn A., Bayliss D., 2020, *MNRAS*, 491, 4481  
 Owen J. E., Lai D., 2018, *MNRAS*, 479, 5012  
 Owen J. E., Wu Y., 2017, *ApJ*, 847, 29  
 Paegert M., Stassun K. G., Collins K. A., Pepper J., Torres G., Jenkins J., Twicken J. D., Latham D. W., 2021, *preprint (arXiv:2108.04778)*  
 Pecaú M. J., Mamajek E. E., 2013, *ApJS*, 208, 9  
 Petigura E. A. et al., 2018, *AJ*, 155, 89  
 Petigura E. A. et al., 2022, *AJ*, 163, 179  
 Popper D. M., Etzel P. B., 1981, *AJ*, 86, 102  
 Rasmussen C. E., Williams C. K. I., 2008, *Gaussian Processes for Machine Learning*, 3. print edn. Adaptive Computation and Machine Learning. MIT Press, Cambridge  
 Ricker G. R. et al., 2014, *JATIS*, 1, 014003  
 Santerne A. et al., 2015, *MNRAS*, 451, 2337  
 Santerne A. et al., 2016, *A&A*, 587, A64  
 Savitzky A., Golay M. J. E., 1964, *Anal. Chem.*, 36, 1627  
 Smart R. L. et al., 2021, *A&A*, 649, A6  
 Southworth J., Bruntt H., Buzasi D. L., 2007, *A&A*, 467, 1215  
 Stassun K. G. et al., 2018, *AJ*, 156, 102  
 Stassun K. G. et al., 2019, *AJ*, 158, 138  
 Szabó Gy. M., Kiss L. L., 2011, *AJ*, 727, L44  
 Tabachnik S., Tremaine S., 2002, *MNRAS*, 335, 151  
 Temmink M., Snellen I. A. G., 2023, *A&A*, 670, A26  
 Thompson S. E. et al., 2018, *ApJS*, 235, 38  
 Vach S. et al., 2024, *AJ*, 167, 210  
 Valizadegan H. et al., 2022, *ApJ*, 926, 120  
 Van Eylen V., Agentoft C., Lundkvist M. S., Kjeldsen H., Owen J. E., Fulton B. J., Petigura E., Snellen I., 2018, *MNRAS*, 479, 4786  
 Wright J. T., Marcy G. W., Howard A. W., Johnson J. A., Morton T. D., Fischer D. A., 2012, *ApJ*, 753, 160  
 Yee S. W., Winn J. N., 2023, *ApJ*, 949, L21  
 Yee S. W., Winn J. N., Hartman J. D., 2021, *AJ*, 162, 240  
 Youdin A. N., 2011, *ApJ*, 742, 38  
 Zhou G. et al., 2019, *AJ*, 158, 141  
 Zhu W., Dong S., 2021, *A&AR*, 59, 291

**APPENDIX A: OCCURRENCE RATE TABLES**

Table A1 lists the occurrence rate values and uncertainties of Fig. 5. Table A2 lists the occurrence rate values and uncertainties of Fig. 9.

**Table A1.** Occurrence rates (number of planets per 100 stars) of our 10-by-10 bins of Fig. 5. Values are reported to two significant figures.

| Period ranges (d) | Radius ranges ( $R_{\oplus}$ ) | Occurrence rates (per cent) | Period ranges (d) | Radius ranges ( $R_{\oplus}$ ) | Occurrence rates (per cent) |
|-------------------|--------------------------------|-----------------------------|-------------------|--------------------------------|-----------------------------|
| 0.50–0.71         | 2.00–2.52                      | 0.0015[+0.0024, –0.0011]    | 0.50–0.71         | 2.52–3.17                      | 0.0019[+0.0017, –0.0012]    |
| 0.50–0.71         | 3.17–3.99                      | 0.0012[+0.0011, –0.00079]   | 0.50–0.71         | 3.99–5.02                      | 0.0024[+0.0010, –0.00084]   |
| 0.50–0.71         | 5.02–6.32                      | 0.00033[+0.00050, –0.00024] | 0.50–0.71         | 6.32–7.96                      | 0.00047[+0.00048, –0.00031] |
| 0.50–0.71         | 7.96–10.02                     | 0.00043[+0.00041, –0.00027] | 0.50–0.71         | 10.02–12.62                    | 0.00021[+0.00031, –0.00016] |
| 0.50–0.71         | 12.62–15.89                    | 0.00018[+0.00029, –0.00013] | 0.50–0.71         | 15.89–20.00                    | 0.00066[+0.0011, –0.00049]  |
| 0.71–1.00         | 2.00–2.52                      | 0.0053[+0.0049, –0.0033]    | 0.71–1.00         | 2.52–3.17                      | 0.0013[+0.0019, –0.00097]   |
| 0.71–1.00         | 3.17–3.99                      | 0.0014[+0.0015, –0.00092]   | 0.71–1.00         | 3.99–5.02                      | 0.0021[+0.0011, –0.00085]   |
| 0.71–1.00         | 5.02–6.32                      | 0.00025[+0.00043, –0.00019] | 0.71–1.00         | 6.32–7.96                      | 0.00019[+0.00030, –0.00014] |
| 0.71–1.00         | 7.96–10.02                     | 0.00017[+0.00027, –0.00013] | 0.71–1.00         | 10.02–12.62                    | 0.00091[+0.00063, –0.00048] |
| 0.71–1.00         | 12.62–15.89                    | 0.00097[+0.00087, –0.00058] | 0.71–1.00         | 15.89–20.00                    | 0.0035[+0.0030, –0.0021]    |
| 1.00–1.41         | 2.00–2.52                      | 0.0089[+0.0090, –0.0059]    | 1.00–1.41         | 2.52–3.17                      | 0.0082[+0.0062, –0.0049]    |
| 1.00–1.41         | 3.17–3.99                      | 0.0051[+0.0033, –0.0025]    | 1.00–1.41         | 3.99–5.02                      | 0.0023[+0.0015, –0.0012]    |
| 1.00–1.41         | 5.02–6.32                      | 0.00052[+0.00071, –0.00039] | 1.00–1.41         | 6.32–7.96                      | 0.00028[+0.00045, –0.00021] |
| 1.00–1.41         | 7.96–10.02                     | 0.00019[+0.00031, –0.00014] | 1.00–1.41         | 10.02–12.62                    | 0.00021[+0.00033, –0.00015] |
| 1.00–1.41         | 12.62–15.89                    | 0.0028[+0.0011, –0.00085]   | 1.00–1.41         | 15.89–20.00                    | 0.0022[+0.0026, –0.0015]    |
| 1.41–2.00         | 2.00–2.52                      | 0.022[+0.017, –0.012]       | 1.41–2.00         | 2.52–3.17                      | 0.0067[+0.0076, –0.0046]    |
| 1.41–2.00         | 3.17–3.99                      | 0.0023[+0.0026, –0.0016]    | 1.41–2.00         | 3.99–5.02                      | 0.0043[+0.0026, –0.0020]    |
| 1.41–2.00         | 5.02–6.32                      | 0.0021[+0.0014, –0.0010]    | 1.41–2.00         | 6.32–7.96                      | 0.00047[+0.00071, –0.00035] |
| 1.41–2.00         | 7.96–10.02                     | 0.00031[+0.00050, –0.00023] | 1.41–2.00         | 10.02–12.62                    | 0.0010[+0.00076, –0.00053]  |
| 1.41–2.00         | 12.62–15.89                    | 0.0039[+0.0013, –0.0011]    | 1.41–2.00         | 15.89–20.00                    | 0.0011[+0.0018, –0.00085]   |
| 2.00–2.83         | 2.00–2.52                      | 0.049[+0.036, –0.026]       | 2.00–2.83         | 2.52–3.17                      | 0.045[+0.017, –0.015]       |
| 2.00–2.83         | 3.17–3.99                      | 0.011[+0.0065, –0.0052]     | 2.00–2.83         | 3.99–5.02                      | 0.0053[+0.0037, –0.0028]    |
| 2.00–2.83         | 5.02–6.32                      | 0.0057[+0.0024, –0.0020]    | 2.00–2.83         | 6.32–7.96                      | 0.0029[+0.0017, –0.0013]    |
| 2.00–2.83         | 7.96–10.02                     | 0.00094[+0.0010, –0.00063]  | 2.00–2.83         | 10.02–12.62                    | 0.0063[+0.0021, –0.0018]    |
| 2.00–2.83         | 12.62–15.89                    | 0.015[+0.0030, –0.0027]     | 2.00–2.83         | 15.89–20.00                    | 0.0031[+0.0035, –0.0022]    |
| 2.83–4.00         | 2.00–2.52                      | 0.14[+0.063, –0.052]        | 2.83–4.00         | 2.52–3.17                      | 0.10[+0.031, –0.029]        |
| 2.83–4.00         | 3.17–3.99                      | 0.013[+0.011, –0.0080]      | 2.83–4.00         | 3.99–5.02                      | 0.013[+0.0071, –0.0057]     |
| 2.83–4.00         | 5.02–6.32                      | 0.018[+0.0055, –0.0045]     | 2.83–4.00         | 6.32–7.96                      | 0.0077[+0.0031, –0.0025]    |
| 2.83–4.00         | 7.96–10.02                     | 0.0054[+0.0026, –0.0021]    | 2.83–4.00         | 10.02–12.62                    | 0.022[+0.0044, –0.0040]     |
| 2.83–4.00         | 12.62–15.89                    | 0.061[+0.0068, –0.0064]     | 2.83–4.00         | 15.89–20.00                    | 0.0098[+0.0091, –0.0062]    |
| 4.00–5.66         | 2.00–2.52                      | 0.64[+0.16, –0.15]          | 4.00–5.66         | 2.52–3.17                      | 0.31[+0.075, –0.065]        |
| 4.00–5.66         | 3.17–3.99                      | 0.025[+0.017, –0.014]       | 4.00–5.66         | 3.99–5.02                      | 0.021[+0.011, –0.0096]      |
| 4.00–5.66         | 5.02–6.32                      | 0.014[+0.0071, –0.0065]     | 4.00–5.66         | 6.32–7.96                      | 0.020[+0.0056, –0.0051]     |
| 4.00–5.66         | 7.96–10.02                     | 0.0098[+0.0038, –0.0031]    | 4.00–5.66         | 10.02–12.62                    | 0.020[+0.0049, –0.0044]     |
| 4.00–5.66         | 12.62–15.89                    | 0.054[+0.0081, –0.0076]     | 4.00–5.66         | 15.89–20.00                    | 0.0067[+0.010, –0.0050]     |
| 5.66–8.00         | 2.00–2.52                      | 0.87[+0.26, –0.22]          | 5.66–8.00         | 2.52–3.17                      | 0.49[+0.12, –0.11]          |
| 5.66–8.00         | 3.17–3.99                      | 0.071[+0.032, –0.029]       | 5.66–8.00         | 3.99–5.02                      | 0.039[+0.015, –0.013]       |
| 5.66–8.00         | 5.02–6.32                      | 0.017[+0.0089, –0.0075]     | 5.66–8.00         | 6.32–7.96                      | 0.014[+0.0059, –0.0051]     |
| 5.66–8.00         | 7.96–10.02                     | 0.024[+0.0061, –0.0056]     | 5.66–8.00         | 10.02–12.62                    | 0.037[+0.0072, –0.0066]     |
| 5.66–8.00         | 12.62–15.89                    | 0.028[+0.0086, –0.0074]     | 5.66–8.00         | 15.89–20.00                    | 0.0099[+0.016, –0.0073]     |
| 8.00–11.31        | 2.00–2.52                      | 1.7[+0.43, –0.39]           | 8.00–11.31        | 2.52–3.17                      | 0.89[+0.19, –0.17]          |
| 8.00–11.31        | 3.17–3.99                      | 0.17[+0.062, –0.055]        | 8.00–11.31        | 3.99–5.02                      | 0.081[+0.026, –0.023]       |
| 8.00–11.31        | 5.02–6.32                      | 0.018[+0.011, –0.0088]      | 8.00–11.31        | 6.32–7.96                      | 0.0059[+0.0053, –0.0036]    |
| 8.00–11.31        | 7.96–10.02                     | 0.029[+0.0076, –0.0066]     | 8.00–11.31        | 10.02–12.62                    | 0.054[+0.0091, –0.0085]     |
| 8.00–11.31        | 12.62–15.89                    | 0.015[+0.010, –0.0076]      | 8.00–11.31        | 15.89–20.00                    | 0.024[+0.038, –0.018]       |
| 11.31–16.00       | 2.00–2.52                      | 1.2[+0.52, –0.44]           | 11.31–16.00       | 2.52–3.17                      | 0.97[+0.27, –0.25]          |
| 11.31–16.00       | 3.17–3.99                      | 0.27[+0.093, –0.082]        | 11.31–16.00       | 3.99–5.02                      | 0.045[+0.036, –0.027]       |
| 11.31–16.00       | 5.02–6.32                      | 0.11[+0.026, –0.024]        | 11.31–16.00       | 6.32–7.96                      | 0.0095[+0.010, –0.0064]     |
| 11.31–16.00       | 7.96–10.02                     | 0.029[+0.011, –0.0093]      | 11.31–16.00       | 10.02–12.62                    | 0.063[+0.014, –0.012]       |
| 11.31–16.00       | 12.62–15.89                    | 0.021[+0.017, –0.012]       | 11.31–16.00       | 15.89–20.00                    | 0.050[+0.082, –0.038]       |

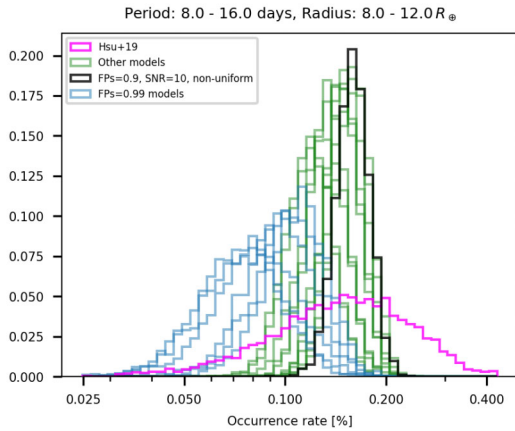
**Table A2.** Occurrence rates (number of planets per 100 stars) of 100 pc sample of Fig. 9. Values are reported to two significant figures.

| Period ranges (d) | Radius ranges ( $R_{\oplus}$ ) | Occurrence rates (per cent) | Period ranges (d) | Radius ranges ( $R_{\oplus}$ ) | Occurrence rates (per cent) |
|-------------------|--------------------------------|-----------------------------|-------------------|--------------------------------|-----------------------------|
| 0.50–1.00         | 2.00–3.17                      | 0.046[+0.037, −0.025]       | 0.50–1.00         | 3.17–5.02                      | 0.022[+0.024, −0.014]       |
| 0.50–1.00         | 5.02–7.96                      | 0.0098[+0.015, −0.0073]     | 0.50–1.00         | 7.96–12.62                     | 0.021[+0.025, −0.014]       |
| 0.50–1.00         | 12.62–20.00                    | 0.061[+0.075, −0.043]       | 1.00–2.00         | 2.00–3.17                      | 0.10[+0.068, −0.048]        |
| 1.00–2.00         | 3.17–5.02                      | 0.017[+0.026, −0.013]       | 1.00–2.00         | 5.02–7.96                      | 0.010[+0.018, −0.0078]      |
| 1.00–2.00         | 7.96–12.62                     | 0.011[+0.019, −0.0086]      | 1.00–2.00         | 12.62–20.00                    | 0.034[+0.055, −0.025]       |
| 2.00–4.00         | 2.00–3.17                      | 0.71[+0.24, −0.19]          | 2.00–4.00         | 3.17–5.02                      | 0.095[+0.088, −0.055]       |
| 2.00–4.00         | 5.02–7.96                      | 0.073[+0.057, −0.039]       | 2.00–4.00         | 7.96–12.62                     | 0.058[+0.046, −0.033]       |
| 2.00–4.00         | 12.62–20.00                    | 0.080[+0.085, −0.054]       | 4.00–8.00         | 2.00–3.17                      | 3.8[+0.65, −0.59]           |
| 4.00–8.00         | 3.17–5.02                      | 0.32[+0.16, −0.13]          | 4.00–8.00         | 5.02–7.96                      | 0.031[+0.051, −0.023]       |
| 4.00–8.00         | 7.96–12.62                     | 0.052[+0.061, −0.035]       | 4.00–8.00         | 12.62–20.00                    | 0.29[+0.25, −0.16]          |
| 8.00–16.00        | 2.00–3.17                      | 7.0[+1.2, −1.1]             | 8.00–16.00        | 3.17–5.02                      | 0.32[+0.28, −0.18]          |
| 8.00–16.00        | 5.02–7.96                      | 0.049[+0.079, −0.037]       | 8.00–16.00        | 7.96–12.62                     | 0.18[+0.15, −0.10]          |
| 8.00–16.00        | 12.62–20.00                    | 1.5[+0.91, −0.69]           |                   |                                |                             |

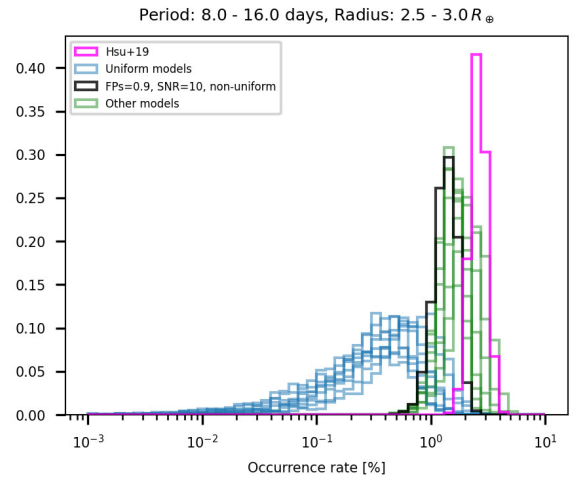
## APPENDIX B: DETAILED POSTERIOR DISTRIBUTION COMPARISONS

Here, selected from Fig. 11, we present the four distributions with the greatest discrepancies when different thresholds are applied.

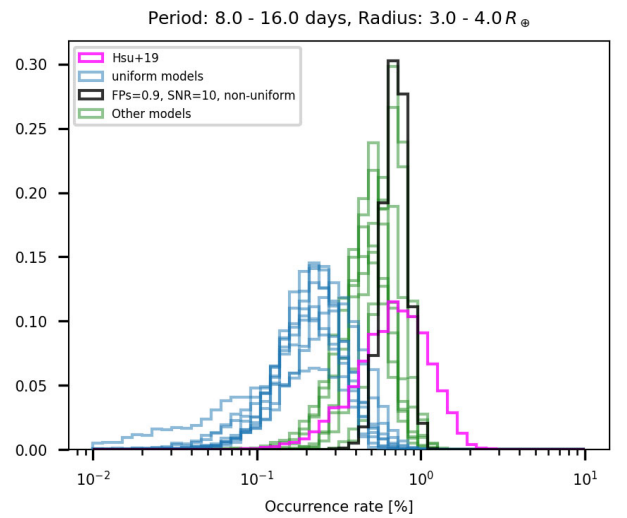
Fig. B1 shows that all models with a 0.99 FP threshold have a significantly lower occurrence rate, while others remain generally consistent, indicating that an excessively high FP threshold eliminates too many candidates, while the completeness estimation fail to account for this. Figs B2, B3, and B4 demonstrate that all models trained on uniformly injected planetary samples show a lower occurrence rate, suggesting these samples are highly deficient in that region and leading to the very limited completeness reliability.



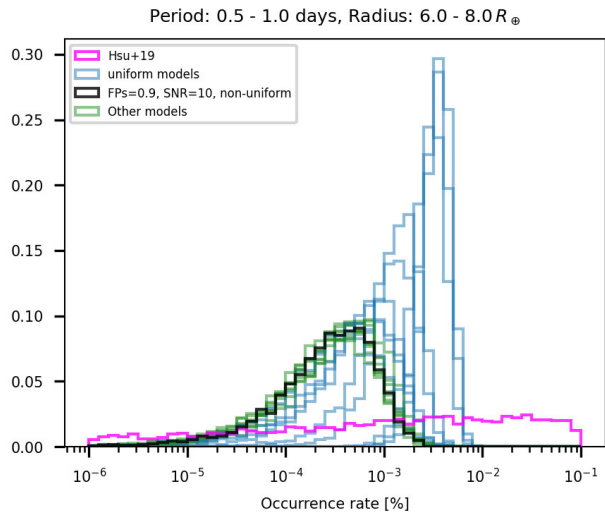
**Figure B1.** Posterior distributions of occurrence rates for various models and the *Kepler* results in a cell with an unusually large discrepancy; the cell’s location is indicated in the title. Different colours denote models with different thresholds, as labelled in the legend.



**Figure B2.** Similar to Fig. B1.



**Figure B3.** Similar to Fig. B1.



**Figure B4.** Similar to Fig. B1.

This paper has been typeset from a  $\text{T}_{\text{E}}\text{X}/\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  file prepared by the author.