The imprint of global magma oceans on exoplanet demographics

Martin Schlecker[1] and al.

[1]*Department of Astronomy/Steward Observatory, The University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA*

ABSTRACT

. . . magma oceans . . .

Here, we assess the ability of space and ground-based telescopes to test this hypothesis using Bioverse, a simulation framework that leverages contextual information from the overall planet population. . . .

1. INTRODUCTION

...just dropping material and references for now...

> papers: (, introduce runaway GH) (Turbet et al. 2019, statistical detection of GH transition), (Luger & Barnes 2015, water loss, O2 buildup around M dwarfs), (???, Solubility of H in magma oceans) Hamano et al. (2013, 2015); Barth et al. (2021); Downey et al. (2022).

For observations of rocky exoplanets, the currently best-probed regime is that of warm, close-in planets. These bodies experience climatic conditions that are similar to the environment of the inner Solar System bodies at early stages of their formation. Studies of the geophysical state and evolution of hot exoplanets can thus contribute to our understanding of the early formation stages of Earth and other habitable worlds.

... Venus and Earth, while having accreted from the same mass reservoir and despite their similar bulk properties, evolved into planets with very different environmental conditions on their surfaces. At the formation of the Solar System, both planets underwent a giant impact phase [CITE!] that melted their mantles, leading to a magma ocean stage. Due to these similar early formation phases, it was commonly assumed that the divergence of Venus and Earth – in particular Venus's water loss – occurred late in their evolution (Elkins-Tanton 2013). Hamano et al. (2013) showed that such a late desiccation process is not needed to explain the stark differences: for a planet rich in volatiles and receiving high enough radiation levels from its host star, a steam atmosphere limits the outgoing radiation from the (molten) planetary surface. This runaway greenhouse state prevents the planet's rapid cooling and can extend the magma ocean stage to hundreds of Myr, enough to remove the entire water reservoir from a rocky planet by hydrodynamic escape.

> important for motivating the existence of the pattern: "the solidification timescale can become comparable to the main-sequence lifetime of the star (Hamano et al. 2013, 2015)" (Lichtenberg et al. 2022)

Under the right conditions, different orbital distances alone can thus decide about a planet's water content and ultimately its habitability. If received flux makes the difference

between our own habitable planet and dry, dead Venus, there is good reason to believe that such patterns exist in exoplanet systems as well.

...long-term goal/overarching objective: derive the geophysical history of rocky extrasolar planets. More concrete: Constrain the limits of runaway greenhouse transitions, and thus the inner edges of the habitable zone. Also: How close is Earth from this runaway greenhouse irradiation limit? ...

> introduce magma oceans and their influence on planetary radii (Dorn & Lichtenberg 2021).

...current/future observations of planets that are currently in the runaway greenhouse phase may constrain properties of their planetary mantles... make connection between interior and atmospheres. ... Planetary radius changes due to the combined effect of steam atmospheres and water dissolved in the molten mantle of planets within the runaway greenhouse threshold is expected to cause a discontinuity in the radius distribution of small (DEFINITION) exoplanets.

2. METHODS

> Caveats: see Turbet et al. (2019) Sect. 3, Bower et al. (2019)

We determined – for different survey configurations – the confidence level with which the runaway greenhouse threshold can be detected in simulated exoplanet observations. For this purpose we expanded the Bioverse framework (Bixel & Apai 2020, 2021)[1] to generate synthetic stellar and planetary samples into which we injected the runaway greenhouse signal, simulated observations of this sample, and computed Bayesian evidences in favor of the runaway greenhouse hypothesis (see diagram in Fig. 1).

2.1. Synthetic star and planet sample

The first step in our analysis is to generate a sample of stars that host synthetic planets.

2.1.1. Stellar sample from Gaia DR3

Different sample sizes are realized by varying the maximum distance of stars to the solar system $d_{\max}$.

> KEVIN: describe stellar population

...

2.1.2. Stellar luminosity tracks

Planetary systems are hosted by stars of a wide range of ages, and stellar luminosities evolve with time. Since the emergence of a runaway greenhouse phase on a planet is highly dependent on the level of radiation it receives, and thus on the luminosity of the host star, we assigned age-dependent luminosities to our synthetic stars. Stellar ages are notoriously poor-constrained (CITE); we thus drew random ages from a

---

[1] Bioverse is actively maintained and documented open source software written in Python. Its latest version and documentation can be found at https://github.com/danielapai/bioverse.
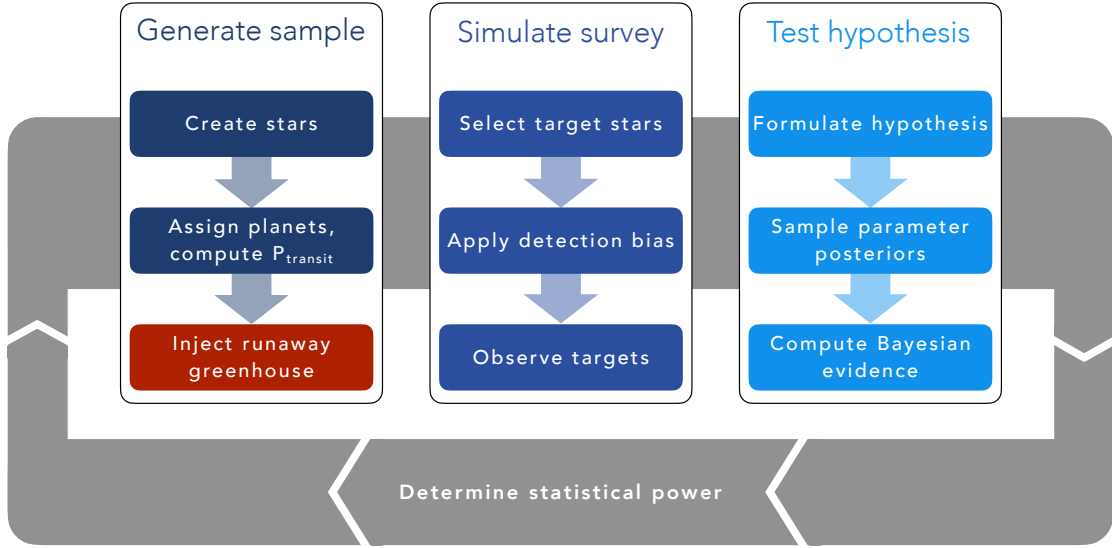
Figure 1. Workflow of our hypothesis testing with Bioverse. In the first block, we generate a sample of stars and populate them with planets based on *Kepler* demographics. A fraction of them are then assigned a runaway greenhouse climate based on the model described in Sect. 2.1.6. The second block simulates an exoplanet survey, whereby selection effects and detection biases are introduced. Finally, the third block deals with testing the runaway greenhouse hypothesis based on data from the survey simulation. By iterating through these steps, we compute the statistical power of testing the hypothesis given the assumed survey design.

uniform distribution from $0\,\mathrm{Gyr}$ to $10\,\mathrm{Gyr}$. We then assigned each star a luminosity from the mass-dependent evolutionary models in Baraffe et al. (1998). Figure 2 shows the corresponding luminosity evolution as a function of stellar mass and age.

2.1.3. *Synthetic planet sample*

Next, we assigned to the stellar sample planetary systems with frequencies, orbital parameters, and bulk properties derived from the *Kepler* mission. We adopted the planetary occurrence rates in orbital period and radius recently derived in (Bergsten et al. 2022). Following (Youdin 2011), their inferred occurrence rate density can be expressed in the form

$$\frac{\mathrm{d}^2 n}{\mathrm{d}R\,\partial P} = F_0 C_n g(R, P), \tag{1}$$

where F_0 ...

GALEN: Briefly describe underlying planet occurrence rates of synthetic population, its Mstar dependency, any caveats or assumptions.

2.1.4. *Eccentricities, orbital inclinations, and planet masses*

Eccentric orbits alter the probability of a planet to transit (e.g., Barnes 2007). The distribution of eccentricities e of exoplanets has been found to resemble a Beta function (Kipping 2013), which we chose to draw synthetic eccentricities from. Fol-
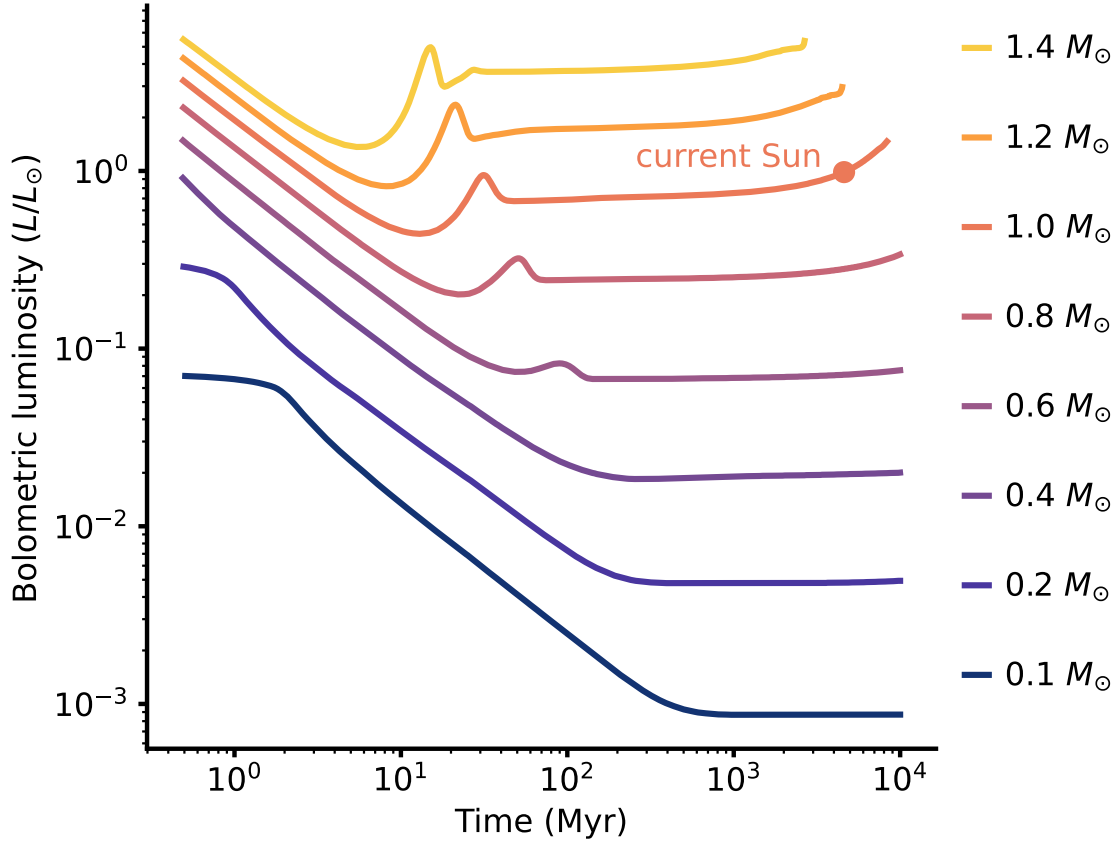
Figure 2. Bolometric luminosity tracks of stars with different masses, computed from stellar evolution models in Baraffe et al. (1998). Low-mass stars, which make up the majority of stars in the solar neighborhood, undergo an extended early phase of several magnitudes higher luminosity before entering a lifetime of relative faintness.

lowing Kipping (2013), we used a Beta distribution with parameters $a = 0.867$ and $b = 3.03$, and truncated the distribution at $e = 0.8$.

Another parameter with an even higher impact on transit probabilities is the orbital inclination toward the observer i. Assuming isotropic alignments of orbits, we assigned each planet an inclination drawn from a distribution uniform in $\cos(i)$.

To assign masses to our planets, we use the semi-empirical mass-radius relationship derived in Zeng et al. (2016). The default relation our planets follow is one assuming a pure $MgSiO_3$ composition.

> JUSTIFY this choice (TIM?)

> run a test: do we expect that the choice of M-R relation impacts the results? Perhaps have an appendix section.

This represents the baseline bulk density before any climate-related effects are applied.

2.1.5. *Transit probability*

We model the occurrence of transits by assuming isotropic orientations of planetary orbits, calculating the impact parameters $b = \cos(i)/R_\star$ and considering only observable

planets with $|b| < 1$. For these cases we calculate the transit depth

$$\delta = \left(\frac{R_\mathrm{P}}{R_\star}\right)^2,\tag{2}$$

which is relevant for the detection probability of the respective planet (see Sect. 2.2.1). Excluding all non-transiting planets diminishes the sample to $999\,\%$ of its original size.

2.1.6. *Planet radius changes due to runaway greenhouse climates*

"Methods Section" for the geophysical models. Describe specific parametrization, dependency on bulk planet/orbit params

Main assumptions:
- Assume that every planet has *some* water. Enough for runaway GH under the right conditions. Baseline water content: 20%
- Limitation to small planets without significant atmospheric envelopes.
- we ignore water loss by H_2O photolysis in the upper atmosphere and subsequent H escape, which would eventually cool even planets within the runaway greenhouse regime (Lichtenberg et al. 2022).
- we ignore tidal heating, which could extend the magma ocean phase of close-in planets (and change their orbits via tidal orbital evolution)(Jackson et al., 2008; Barnes et al., 2013 (?)).

...

ARNAUD: revise definition of S following our discussion on 2022-12-22. Introduce net instellation and clarify what is meant by it.

The power from the host star per unit area at the position of a planet or instellation S in units of Earth's insolation is given by

$$\frac{S}{S_\oplus} = \left(\frac{L_\star}{L_\odot}\right)\left(\frac{au}{a}\right)^2.\tag{3}$$

Here, we assume global redistribution of incoming flux and a fixed albedo of 0.3, comparable to Earth's [CITE]. The absolute value of $S_\oplus$, by which all synthetic planets are scaled, is then $238\,\mathrm{W\,m^{-2}}$.

Central to our procedure is to explore the detectability of population-level trends with future exoplanet surveys, and here the trend in question is the proposed radius discontinuity. To enable a quantitative assessment of this detectability, we injected the signal into the simulated planet sample before observing it with simulated surveys. While we search for the runaway greenhouse pattern in demographic quantities such as average planet radii, the injected changes happen on the planetary level: We changed each planet's radius based on its individual set of properties and the associated predictions from steam atmosphere and water incorporation models. Relevant properties are a planet's mass M, its instellation S, and its bulk water inventory expressed as a water mass fraction x_{H_2O}. We consider three different cases:

A. No radius change. This case serves as our null hypothesis, where planetary radii are not affected by the amount of irradiation the planets receive. It also applies outside the runaway greenhouse limit, where we assume a solid planetary surface and no steam atmosphere. A planet that lies *within* the runaway greenhouse limit but that is already completely desiccated will have the same radius.

Mention if this case is considered in the model or not (as of now, no).

B. Radius inflation due to a steam atmosphere. We assigned planets within the runaway greenhouse limit steam atmospheres, which increase atmospheric scale heights significantly.

    C. As B, but including radius decrease of inflated planets due to water in the melt. Due to greenhouse forcing, planetary surfaces below steam atmospheres are thought to be molten. Part of the water will then partition into the melt, removing it from the steam atmosphere, which in turn decreases the radius inflation. This effect is generally small compared to the radius inflation from a steam atmosphere but resulting radius changes typically reach measurable values.

In case A, a planet simply retains the radius it was assigned based on exoplanet occurrence rates (see Sect. 2.1.3). There is no additional dependency on orbital distance or installation.

Case B assumes an inflated radius due to a steam atmosphere for all planets receiving a dayside-averaged installation exceeding a threshold of $S_{\mathrm{thresh}} = 280\,\mathrm{W\,m^{-2}}$. This value was found to be a limit for the flux a planet can emit in a runaway greenhouse situation (Goldblatt et al. 2013; Leconte et al. 2013). To quantify the radius change, we applied the mass-radius relationships derived by Turbet et al. (2020) using a 1D inverse radiative-convective model (Turbet et al. 2019). For the rocky interiors, their calculations rely on the same mass-radius relations from (Zeng et al. 2016) that we applied for case A-planets. For each planet above the installation threshold, we assigned the predicted radius for the given water mass fraction and planet mass. We note that an additional radius change due to potentially molten interiors (Bower et al. 2019) is not taken into account here.

Case C contains case B, but also includes a radius decrease due to water partitioning in the melt. The magnitude of this effect is again dependent on the planet mass and on the water mass fraction. We follow Dorn & Lichtenberg (2021) and use their computed radius deviations between a wet magma ocean and a solid mantle for a tropopause pressure $P_{\mathrm{iso}} = 0.1\,\mathrm{bar}$. We then add the (in almost all cases negative) radius deviations to the planet radii computed for case B.

Figure 3 shows the different mass-radius relations of the three cases for a fiducial water mass fraction of $x_{H_2O} = 0.005$. In general, steam atmospheres (case B) cause a significant radius increase, which is slightly reduced when water incorporation in the melt is allowed (case C).

Our simulation setup is such that all planets receiving an installation $S < S_{\mathrm{thresh}}$ follow relation A, and a fraction f_{rgh} of the planets with $S > S_{\mathrm{thresh}}$ follow relation C. The choice $f_{\mathrm{rgh}} < 1$ comprises all reasons why a planet has no runaway greenhouse climate despite high installation, for instance the absence of an atmosphere or of volatiles that could form a steam atmosphere, or prior complete desiccation. This parametrization introduces a discontinuity of average planet radii and bulk densities as a function of stellar irradiation. In the following, we test if and under what conditions this trend is large enough to be detected with high significance.

2.2. *Survey simulations*
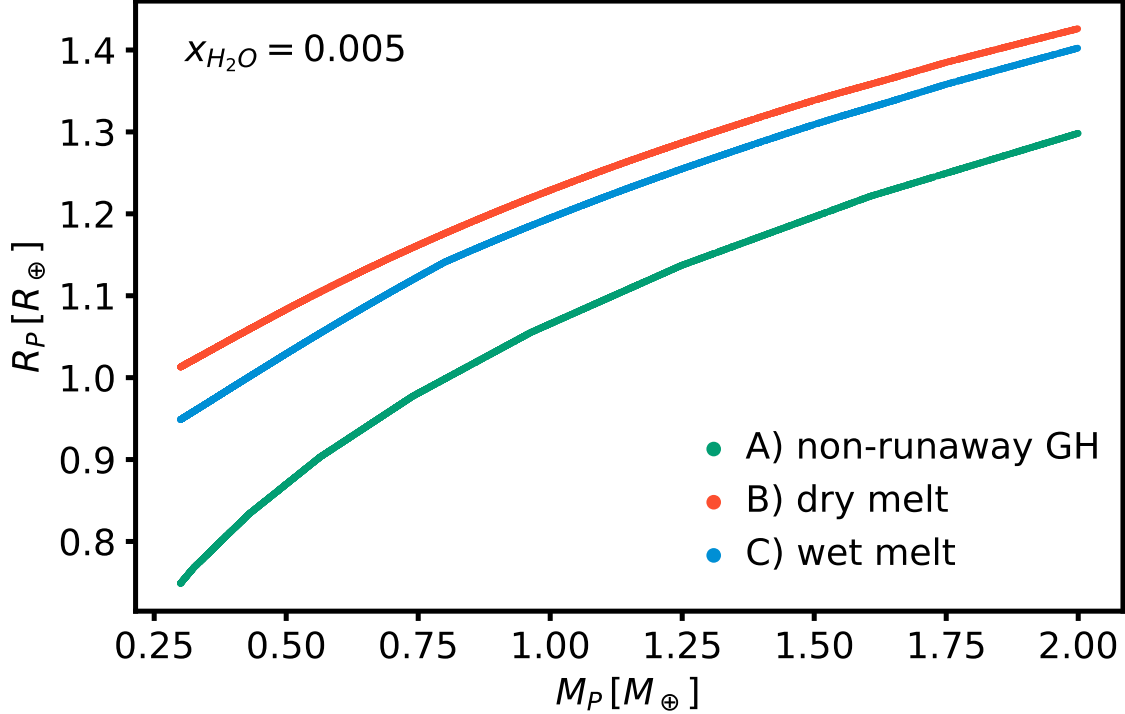
Figure 3. Mass-radius relationships for a water mass fraction $x_{H_2O} = 0.005$ and different planet states: A) fully solid planets without magma oceans. B) planets with steam atmospheres. C) As B), but including the effect of water incorporation in the melt.

The survey module of Bioverse converts the synthetic planet sample into a set of uncertainty-laden measurements on a subset of that sample. This task includes selection of the targets, application of detection biases, and conducting simulated measurements. All of the above are specific to the particular survey.

2.2.1. *Detection bias and sample selection*

Not all transiting planets are detectable with the same likelihood and detection biases have an impact on the demographic measurements we are interested in. A detailed characterization of the detection biases of individual missions would not be justifiable given the uncertainties of the theoretical predictions. Instead, we derived generic observing limits that reflect the limitations of state-of-the-art transit surveys.

The runaway greenhouse transition is expected to occur at orbits sufficiently close to let mission duration not be a limiting factor: S_{thresh} occurs close to Earth's orbital period for G-type stars and on significantly shorter periods for the more prevalent M dwarf hosts (Goldblatt & Watson 2012). However, for a transit signal to be detected in a transit survey, it must reach a sufficient signal-to-noise ratio (S/N). This S/N is sensitive on the photometric precision that can be achieved (e.g., Burke et al. 2015; Hardegree-Ullman et al. 2019). In its in-orbit commissioning phase, the *CHEOPS* mission demonstrated a photometric precision of 75 ppm for faint ($V \approx 12$) stars (Benz et al. 2021). In comparison, *PLATO*'s expected precision for similar stars

was estimated to be $\sim$50 ppm (ESA 2017, Matuszewski et al., in prep.). To represent these photometric limits, we chose a minimum transit depth of 75 ppm as a detection limit and consider only measurements of planets exceeding this threshold. We further exclude target stars with *Gaia* magnitudes $M_{\mathrm{G}} > 11$.

The runaway greenhouse effect becomes obsolete where no atmosphere can be maintained due to proximity to the host star and resulting atmospheric erosion. We thus excluded planets with extremely high instellation that does not allow long-term preservation of an atmosphere and clear our sample from all planets with an instellation $S > 2000\,\mathrm{W\,m^{-2}}$.

use criterion of Zahnle & Caitling 2017 instead? $0.8S^{0.25} < R$

We further consider only rocky planets with masses below $2\,\mathrm{M_{\oplus}}$.

2.3. *Hypothesis tests*

We now turn to assessing the simulated surveys in terms of their ability to detect the runaway greenhouse transition and their constraining power on parameters of this trend. To do this, we rely on a Bayesian hypothesis testing approach where we quantify the evidence of a hypothesis over another based on the (simulated) data. For our specific problem, this implies comparing evidences for a demographic imprint of the runaway greenhouse effect to its absence.

2.3.1. *Definition of the runaway greenhouse hypothesis*

As a null hypothesis, we consider the case where the planetary radius distribution is independent of the instellation,

$$H_0(\theta, S) = \theta, \tag{4}$$

where θ is the set of parameters defining the radius distribution. We further define an alternative hypothesis that describes radius changes due to steam atmospheres and water in the melt. As motivated above, this hypothesis takes the form of a step function in instellation S, where the step occurs at the outer edge of the runaway greenhouse region. Our main observable shall be the average planet radius in the planet population inside and outside this threshold. The runaway greenhouse hypothesis is then defined as

$$H_{\mathrm{rgh}}(\theta, S) = \begin{cases} H_0, & S \leq S_{\mathrm{thresh}} \\ \langle R_{\mathrm{P}} \rangle (f_{\mathrm{rgh}}, \Delta R_{\mathrm{stm}}, \Delta R_{\mathrm{wtr}}), & S > S_{\mathrm{thresh}}. \end{cases} \tag{5}$$

Here, f_{rgh} is the fraction of planets experiencing a runaway greenhouse effect. ΔR_{stm} and ΔR_{wtr} are predicted radius changes from the steam atmosphere and water incorporation models, respectively. They are assumed to act additively on the planet radii and thus on their average $\langle R_{\mathrm{P}} \rangle$.

The only free parameter of the null hypothesis, which assumes the average radius to be independent of instellation, is the predicted mean radius $\langle R_{\mathrm{P}} \rangle$. The functional

form of the runaway greenhouse hypothesis is more complex: Besides the mean radius of planets outside the threshold $\langle R_P \rangle_{\text{out}}$, which is a nuisance parameter necessary to define the hypothesis, it relies on the threshold instellation for the "step" S_{thresh}, the planetary water mass fraction x_{H_2O} encoding the amount of radius change, and the fraction of planets above the threshold instellation that are in a runaway greenhouse climate f_{rgh}.

The measured radii $R_{P,i}$ cannot be directly used for the hypothesis tests as they include intrinsic scatter that is not caused by measurement errors. H_{rgh} and H_0 should thus be tested against a statistical estimator that represents the population mean. To avoid binning and the artificial patterns it may introduce, we chose to test our hypotheses against a simple moving average SMA along the instellation axis with a window of size 25 centered around each measurement. We further computed the uncertainty of this moving average by propagating the individual measurement errors and applying a rolling standard error of the mean.

`specify error propagation`

As our procedure involves random sampling of the model parameters θ, we need to define the probability of obtaining a dataset given the model parameters, i.e., a likelihood function. We assumed here that the individual moving averages SMA_i are measured with a normally distributed uncertainty σ_{SMA_i} and adopted a normal distribution

$$\mathcal{L}(SMA \mid \boldsymbol{\theta}) = \prod_i^N \frac{1}{\sqrt{2\pi\sigma_{SMA_i}^2}} \exp\left(-\frac{(SMA_i - H(\boldsymbol{\theta}, S_i))^2}{2\sigma_{SMA_i}^2}\right). \tag{6}$$

Here, $H(\boldsymbol{\theta}, S_i)$ corresponds to the functional form of the runaway greenhouse or null hypothesis.

2.3.2. *Bayesian model comparison*

We are now ready to assess the relative plausibility of H_{rgh} and H_0 given the synthetic data we have generated, assigning equal a priori probabilities to these models. This is done by comparing the Bayesian evidence $\mathcal{Z}$ of the models, which we estimate with the nested sampling (Skilling 2004) algorithm *dynesty* (Speagle 2020). Our criterion to reject the null hypothesis is

$$\Delta \ln \mathcal{Z} = \ln \mathcal{Z}_{\text{rgh}} - \ln \mathcal{Z}_0 > 3. \tag{7}$$

The chosen threshold roughly corresponds to the $p = 0.05$ threshold for statistical significance popular in frequentist approaches.

`CITATION`

A sensible choice of priors is central for evidence estimation via nested sampling. As the parameters we are interested in recovering, S_{thresh} and x_{H_2O}, are poorly constrained by previous data, we used relatively uninformative priors to sample the entire physically plausible parameter space. For S_{thresh}, we chose a log-uniform prior

in $[10, 1000]\,\mathrm{W\,m^{-2}}$. Similarly, we sample x_{H_2O} from a log-uniform distribution to imply scale-invariant ignorance. Its boundaries $[10^{-5}, 0.1]$ are motivated by the water mass fractions covered by the geophysical models (Sect. 2.1.6). Finally, we adopt a broad, uniform prior for $\langle R_P \rangle_{\mathrm{out}}$ bound by $[0.1, 15]\,\mathrm{R_\oplus}$. We then ran the nested sampler to compute the evidence and sample the posterior distributions.

3. RESULTS

3.1. *Statistical signature of the runaway greenhouse threshold*

To characterize the population-level imprint of individual radius changes, we generated a generic planet population with an injected runaway greenhouse effect assuming a water fraction $x_{\mathrm{H_2O}} = 0.005$. Figure 4 shows the resulting planetary radii. When ordered in orbital period space, the different planet types overlap, diluting the demographic imprint. With net instellation as an independent variable, planets above and below the runaway greenhouse threshold separate. We also show the predictions of observable average planet radii from the statistical hypotheses defined above. Within the runaway greenhouse limit, an average radius change of $15\,\%$ occurs.

3.2. *Testability of the runaway greenhouse hypothesis*

Figure 5 shows a prototypical detection of the runaway greenhouse threshold. For the sake of clarity, we chose an optimistic scenario where the sample is large ($N = 281$) and the measurement uncertainties are small ($\sigma_R = 2\,\%, \sigma_S = 5\,\%$). We assumed that the fraction of those planets irradiated stronger than S_{thresh} that have runaway greenhouse climates is 0.8, and we chose a water mass fraction of 0.005 for each planet. In this case, the pattern was detected with high significance ($\Delta \ln \mathcal{Z} \approx 100$).

With such a strong signal, we can attempt an inference of the parameters defining the injected effect. Figure 6 shows the posterior distributions of S_{thresh}, x_{H_2O}, and f_{rgh} as determined by the nested sampler. The threshold instellation can be accurately constrained, $S_{\mathrm{thresh}} = 277^{+70}_{-219}\,\mathrm{W\,m^{-2}}$. Both a higher water mass fraction and a higher fraction of planets in a runaway greenhouse state lead to larger average radii, thus these parameters are strongly correlated and cannot be well constrained without independent measurements.

Given an optimistic survey design, what influences the probability of correctly rejecting a false null hypothesis? Figure 7 explores the statistical power achieved in the above scenario for different combinations of the poorly constrained parameters x_{H_2O} and f_{rgh}. The higher the water content and thus the radius change, and the higher the fraction of planets in runaway greenhouse climates, the greater the prospect of a statistically sound detection. f_{rgh} dominates this trend: It enters linearly into the average planet radius, whereas the contribution of x_{H_2O} - as predicted by the geophysical models - is sublinear with a power-law exponent of ~ 0.3. As long as f_{rgh} is larger than ~ 0.1, a sample size of 281 is sufficient for a $50\,\%$ detection rate even for water ratios as low as $x_{H_2O} = 0.001$.
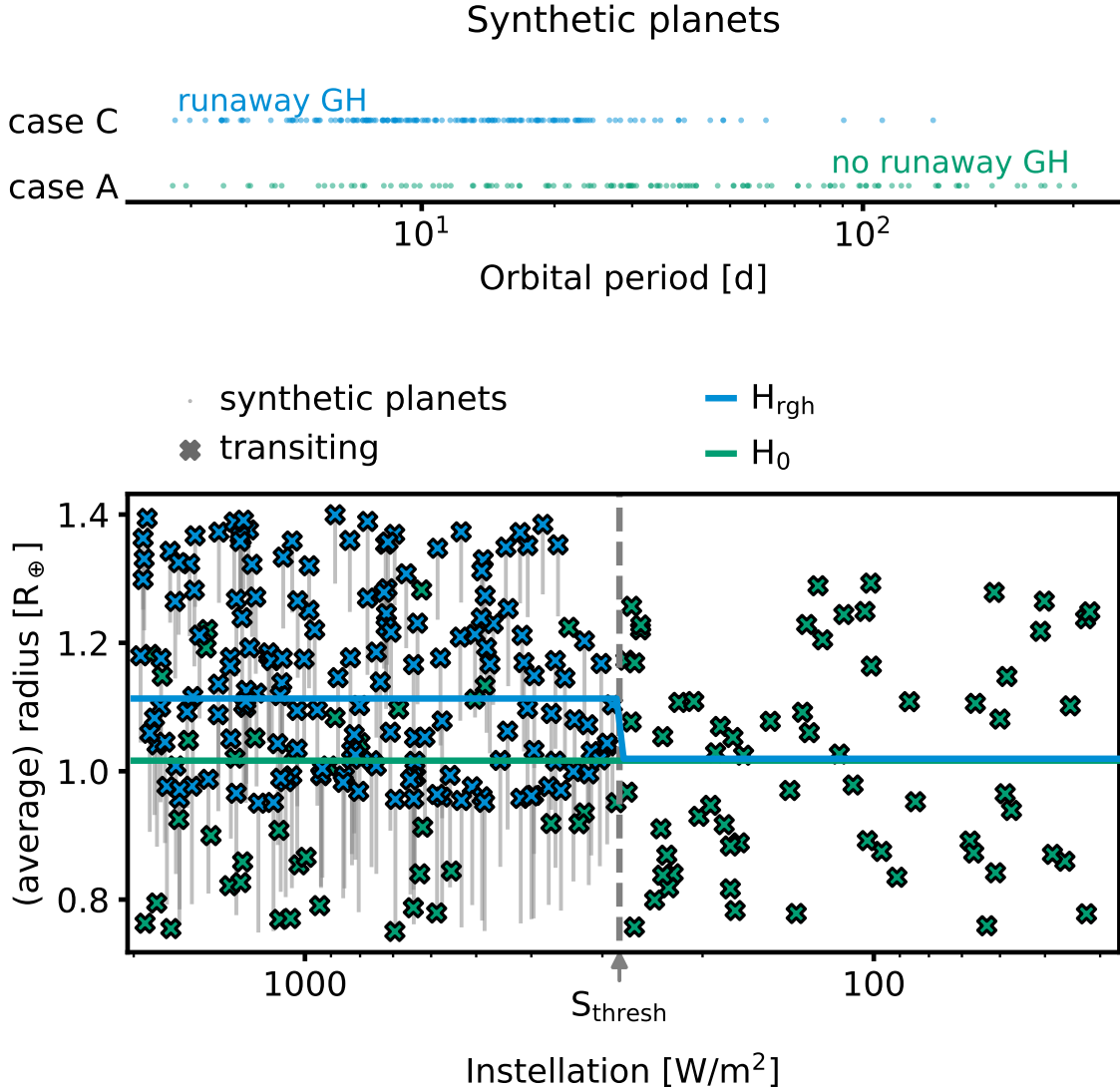
Figure 4. Synthetic planets above and below the runaway greenhouse threshold. *Top*: Planet state as a function of orbital period. Planets with and without a runaway greenhouse climate mix and are not distinguishable in period space. *Bottom*: Radii of synthetic planets with injected radius deviation as a function of instellation. Only the planets marked as transiting are observable. Above the runaway greenhouse threshold $S_{\text{thresh}} = 280\,\mathrm{W\,m^{-2}}$, the radii of a fraction of planets are inflated by the amount indicated with gray lines. The sharp boundary at S_{thresh} causes a discontinuity in the average planet radius (blue). This runaway greenhouse hypothesis can be tested against the null hypothesis H_0 (green), where average radii are independent of instellation.

3.3. *Detecting the runaway greenhouse transition with current and planned exoplanet missions*

Introduce survey simulation(s). PLATO, (Ariel, LIFE(reach out to team; Hendrik Linz produced a LIFE Technology Overview document)) ?, Nautilus
What are threshold missions that are able to detect this signal? =>X meter working for Y years, or X' meter working for Y' years => conclusion could be that there is interesting science to do with intermediate size telescope.

The planetary properties in question can only be probed with a finite precision that is specific to each exoplanet mission.
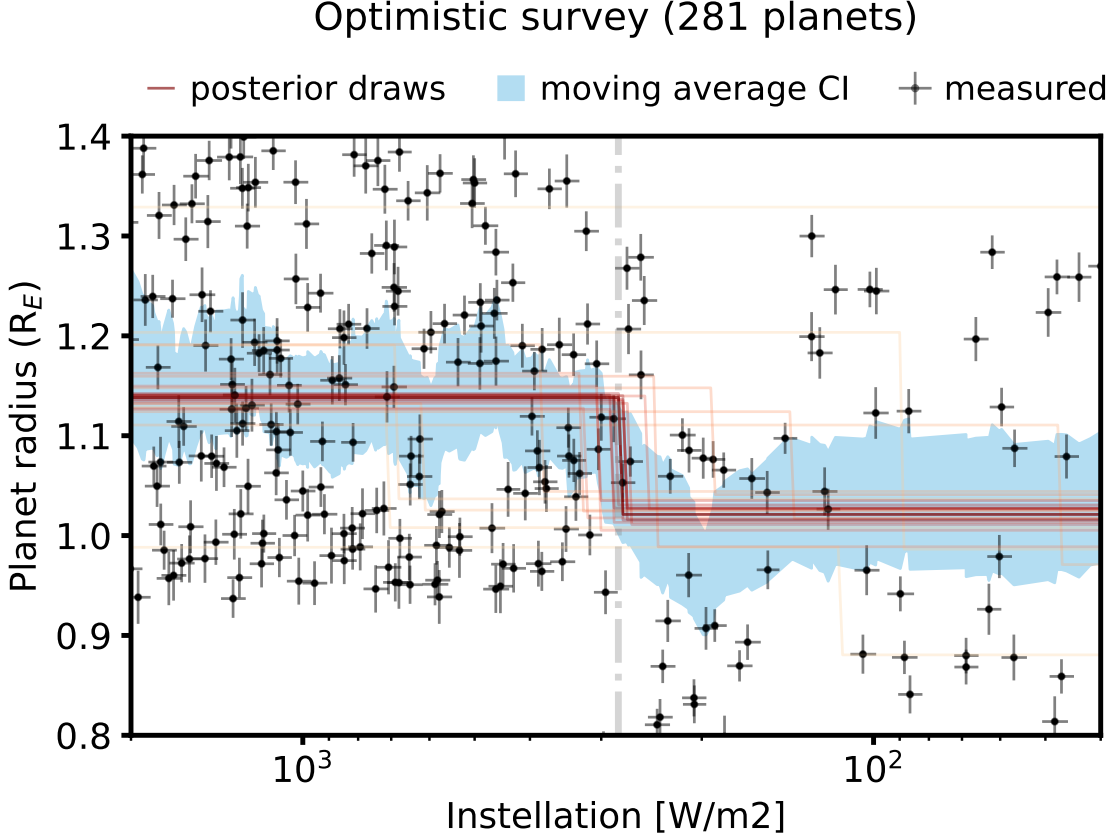
Figure 5. Detection of the runaway greenhouse threshold in the optimistic case. From simulated radius and instellation measurements of a large ($N = 281$) survey, we compute the moving average (blue confidence intervals) and fit the runaway greenhouse hypothesis to it (Eqn. 5, random draws from the posterior in red). The pattern is detected with high significance.

PLATO (PLAnetary Transits and Oscillation of stars) is an ESA mission designed to characterize terrestrial planets in the habitable zones of Sun-like stars, a goal to be achieved via long-term high-precision photometric monitoring of a sample of bright stars (Rauer et al. 2016). The *PLATO* team has released an estimate on the number of exoplanets that will be characterized in the course of the main survey mission (Matuszewski, priv. comm.). In the FGK sample of the mission, a minimum of 500 Earth-sized (0.8-$1.25\,R_\oplus$) planets are expected to be detected. *PLATO*'s definition study report (ESA 2017) further states precision requirements for planet radii ($3\,\%$), planet masses through radial velocity follow-up ($10\,\%$), and stellar masses, radii, and ages ($10\,\%$). We base our assessment of *PLATO*'s diagnostic power on these estimates.

For a volume-limited sample of 500 planets, considering all host star spectral types, we find that *PLATO* is able to obtain a significant detection if the fraction of runaway greenhouse planets is larger than ~ 0.1 (see Fig. 8). The minimum needed fraction rises to 0.5 for $N = 100$. For much smaller samples, even an optimistically strong signal can no longer be detected.
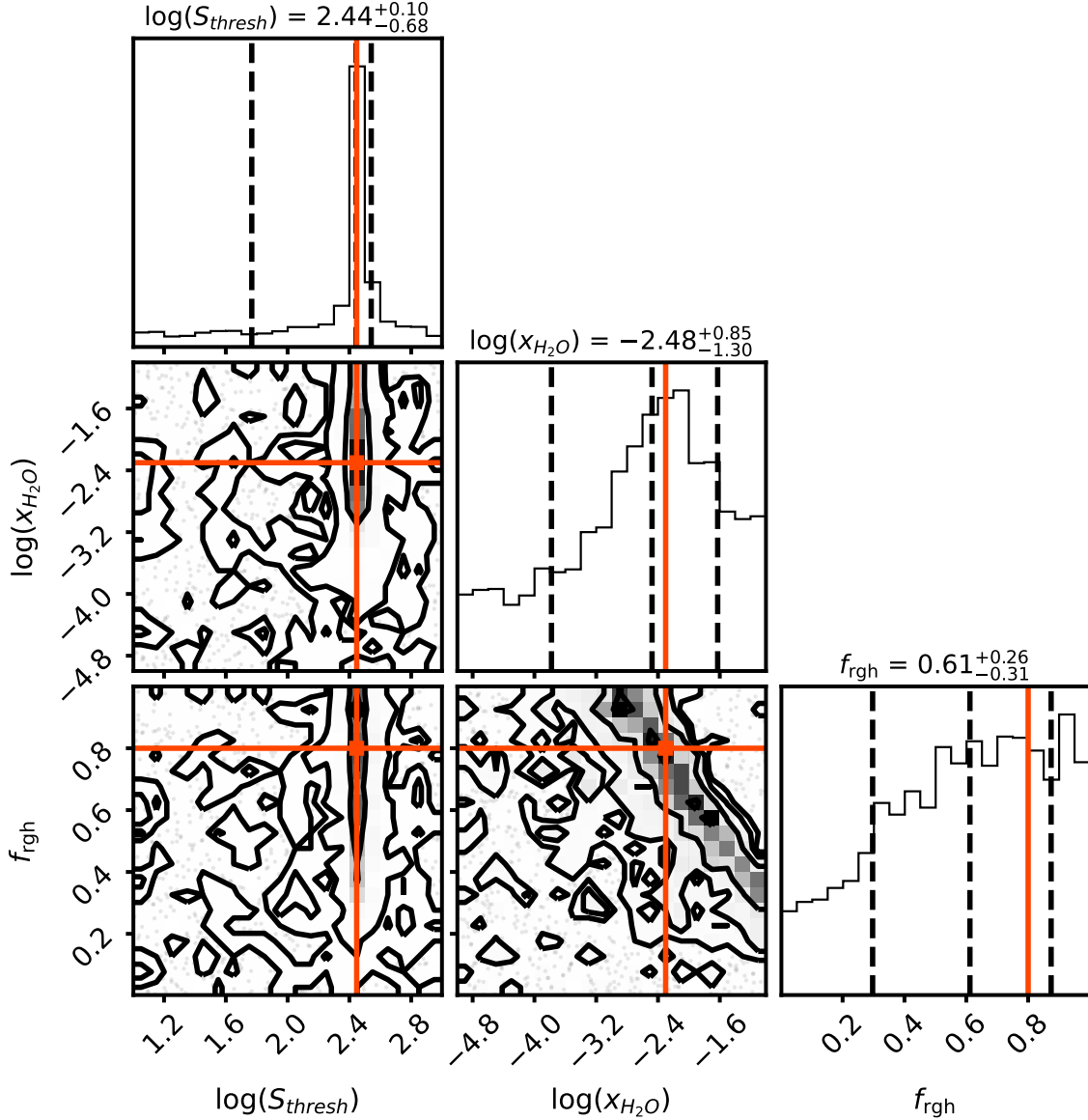
Figure 6. Retrieved posterior distribution of key parameters in the optimistic scenario. The density maps in each panel show relationships between and marginalized distributions of the model parameters as they could be retrieved with a high-precision transit survey and a sample of 281 planets. True values of the parameters for the injected effect are shown in orange. The threshold instellation can be well constrained; the predominant water fraction and the fraction of planets with runaway greenhouse climates are degenerate.

What other planned missions are suited to test the runaway greenhouse hypothesis? The ongoing *CHEOPS* mission was designed as a follow-up mission to search for transits of planets discovered with other techniques, in particular with radial velocity measurements (Benz et al. 2021). As such, it will provide precise radius constraints on a sample of small planets; however, only a small number of planets with periods >50 d are being observed. This largely limits *CHEOPS*' coverage to planets within the runaway greenhouse regime, preventing a detection of the transition.
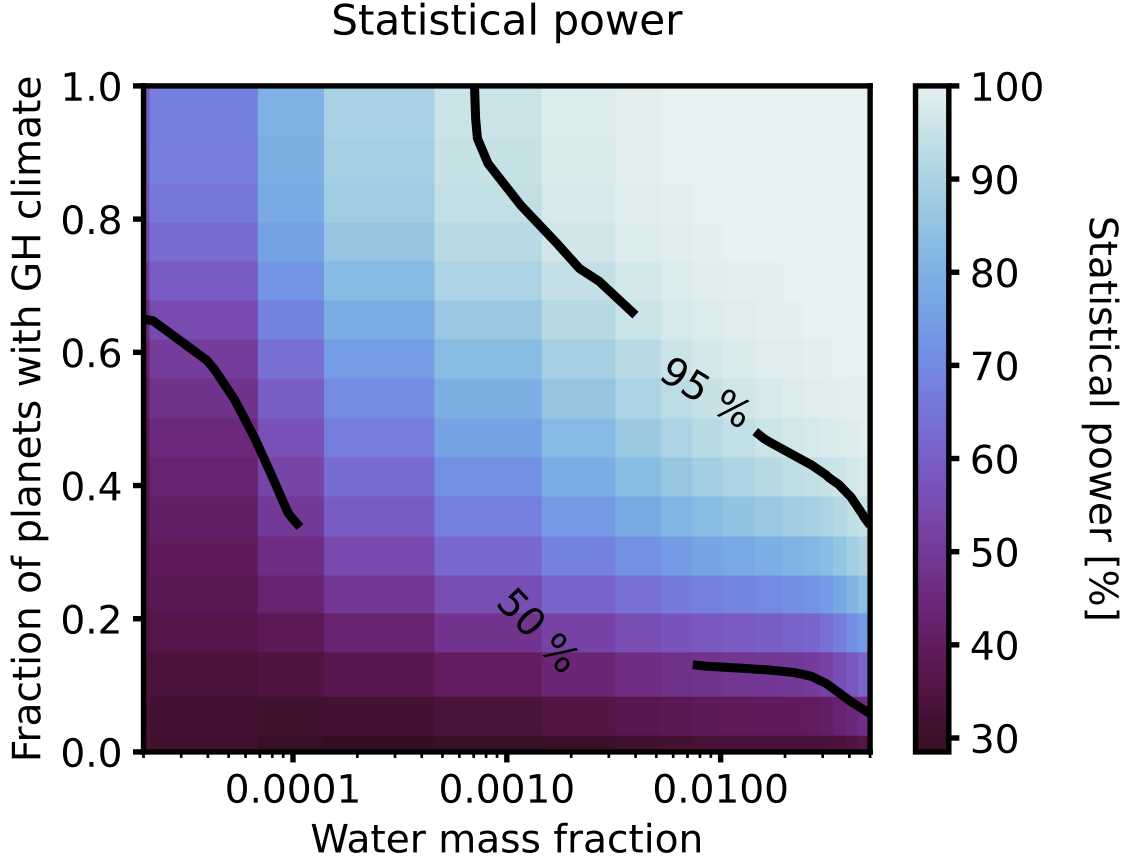
Figure 7. Statistical power of the runaway greenhouse hypothesis test as a function of model parameters. For a sample size $N = 281$, the color code shows the fraction of simulations resulting in a sound detection ($\Delta \ln \mathcal{Z} > 3$) for different combinations of water mass fraction and fraction of greenhouse climate planets. Higher values in either parameter result in a more reliable detection.

As *CHEOPS*, *Ariel* (Puig et al. 2016) will be a follow-up mission that is not designed to provide a large number of new radius measurements. *Ariel*'s primary targets are larger planets in the range of sub-Neptune to Jupiter-like planets. We thus do not expect a significant contribution to exploring the inner edge of the habitable zone for Earth-sized planets.

3.4. *Statistical power of different mission designs*

To explore the impact of mission trade-offs on the detectability of the runaway greenhouse transition, we simulated several transit surveys with different designs and strategies and measured their capability to recover the trend and constrain its parameters. We assessed this capability based on two determinants: the likelihood that the mission is able to detect the injected trend, and the precision with which it can constrain the parameters of that trend.

3.4.1. *Additional planet mass measurements*
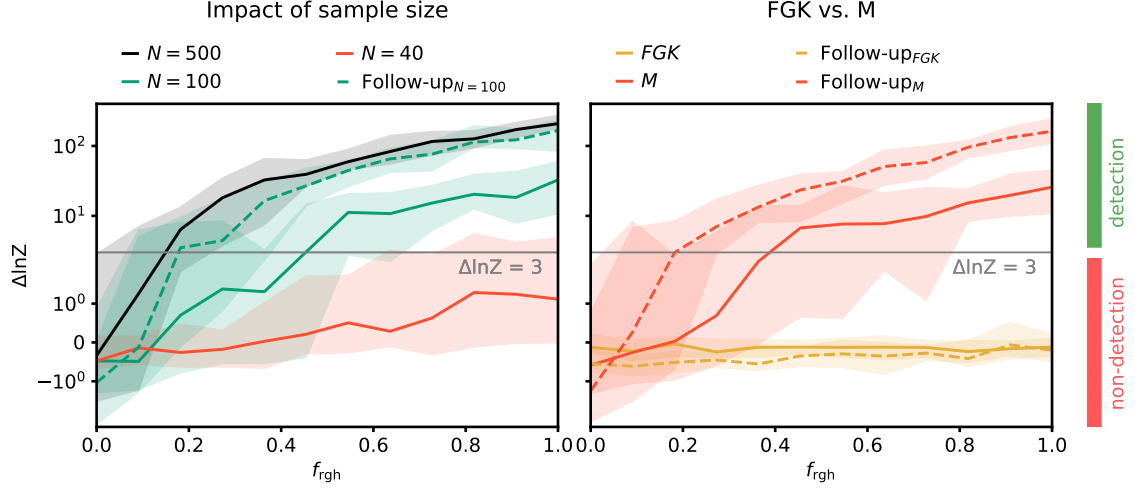
Figure 8. Expected evidences as a function of the fraction of planets harboring runaway greenhouse climates for different versions of the *PLATO* survey. $\Delta \ln Z > 3$ (gray horizontal line) is considered sufficient evidence to reject the null hypothesis. *Left:* For a large planet yield of $N = 500$, even small planet fractions $\gtrsim 0.1$ allow a detection. A much smaller sample of 100 planets is sufficient if their masses are constrained to within 10% (dashed green line). Without such follow-up measurements, sufficient diagnostic power can only be achieved with this sample if $f_{\mathrm{rgh}} \gtrsim 0.5$. Even smaller samples are unlikely to yield a significant detection. *Right:* Evidences when only FGK or only M dwarfs are considered. Only M dwarfs host enough planets on both sides of the threshold instellation to allow a detection of the runaway greenhouse signal.

The dependencies of a detection change when additional information beyond planet radii is available for the characterized planet population. Since the planet bulk density $\rho \propto R^{-3}$, the signal is stronger when measured in bulk density instead of radius: With an optimistic choice of geophysical parameters (see above), the average measured radius change is 15% whereas the average density change is -33%. Constraining bulk densities requires planetary mass measurements. We now examine the value of a follow-up campaign to constrain planet masses with respect to the detectability of the runaway greenhouse threshold.

For uncertainties in density measurements, we propagate normally distributed errors of mass measurements assuming $\sigma_{\mathrm{M_P}} = 10\%$, as expected for *PLATO*. As can be seen in Fig. 8, the detectability is enhanced when radius constraints can be augmented with mass measurements: At unchanged sample size, the difference in Bayesian evidence can be up to an order of magnitude larger. Consequently, we achieve a statistically significant detection with smaller samples or lower fractions f_{rgh}. A density-based survey of 100 targets is roughly equivalent to a radius-based survey of 500 targets. At $N = 100$, pure radius measurements require $f_{\mathrm{rgh}} \gtrsim 0.5$ whereas bulk density measurements enable a detection from $f_{\mathrm{rgh}} \gtrsim 0.2$.

3.4.2. *Spectral type of target stars*

Since receiving radiation is sensitive on the spectral type of the host star, the share of planets on either side of S_{thresh} is different for FGK and M dwarfs. We tested

the detectability of the runaway greenhouse signal when only FGK or only M dwarfs are considered (see Fig. 8). The samples are volume and magnitude-limited ($d_{\max} = 55\,\mathrm{pc}$, $M_{\mathrm{G}} < 11$), which makes the M dwarf sample significantly larger with 92 planets compared to 20 planets in the FGK sample. No significant detection is possible in the pure FGK sample, independent of the assumed geophysical parameters. In the M dwarf sample, the evidence threshold is reached around $f_{\mathrm{rgh}} \sim 0.5$, similar to the case above where all spectral types are considered.

3.4.3. *Constraining the threshold instellation*

for different survey modes, compare the constraints on S_{thresh}. Plot posteriors from random draws of simulations.

4. DISCUSSION

Potential discussion points: what is the fraction of *Kepler* planets w/ runaway GH.

4.1. *Statistical imprint of exoplanet climates*

the predicted climate tipping point causes a demographic imprint.
Discuss our ignorance on f_{rgh} and its impact on the imprint.

4.2. *Detectability of the runaway greenhouse threshold*

the demographic imprint is detectable. Under what conditions?
Mention multiplanet systems.

We identified the following drivers of the diagnostic power for detecting the runaway greenhouse transition with a given transit survey:

- Prevalent water inventory: The magnitude of radius change at the runaway greenhouse threshold is highly sensitive on water mass fraction. As a result, the statistical abundance of water in terrestrial planets impacts the strength of the demographic pattern.

- Sample size: Since we are looking for a population-level trend, the statistical significance will increase with a larger sample size.

- Radius measurement precision: The more precise individual planet radii can be determined, the less smeared out the pattern will be. Good *accuracy* is less important, as long as it does not have a systematic error scaling with stellar irradiance.

- Availability and precision of mass measurements: For simple geometric reasons ($\rho \propto R^{-3}$), the expected trend is stronger when measured in bulk density than it is in planet radius space. If transiting planets can be followed up to obtain mass measurements, the statistical significance increases.

Besides these main factors, uncertainties in the measured instellations can influence the result, although they are typically small due to the very precise orbital period measurements available for transiting planets. This can be different for young host

stars when their ages can not be well constrained; in particular the long pre-main sequence phase of M dwarfs shows a large variation in bolometric luminosity (see Fig. 2).

4.2.1. *False positive scenarios*

what other processes could be confused with a magma ocean signal?

Global magma oceans are not the only physical mechanism that may cause a decrease in transit radii for a subset of planets. Other causes of shrinking planet sizes have been put forward, the most widely discussed ones being atmospheric loss due to either photoevaporation through high-energy radiation by the host star (e.g., Owen & Wu 2013; Jin et al. 2014; Mordasini 2020) or due to residual heat from the planet's interior shortly after formation (Ginzburg et al. 2016, 2018; Gupta & Schlichting 2019). Both processes are being traded as potentially sculpting the observed radius bimodality of small, close-in exoplanets from the mission (Fulton et al. 2017; Van Eylen et al. 2018). Like the magma ocean effect discussed here, these processes reduce the radii of some planets, leading to a decrease in average measured planet radii in a specific region of the planetary parameter space. This region is distinct from the one affected by global magma oceans, since...

...elaborate why they cannot be confused

Another false positive contribution may come from potential unknown variations in the occurrence rate gradients in radius-instellation space. They can impact the statistical inference of the transitions, especially if these variations are similar to the expected pattern. Although an abrupt pattern at the expected runaway greenhouse transition seems unlikely, examples of steep occurrence rate density changes exist. An example is the "Neptune desert", a triangular region of low planet occurrence density of close-in planets in period-radius space (Szabó & Kiss 2011; Mazeh et al. 2016; Dreizler et al. 2020). The shape of this region is such that smaller planets become less frequent the closer to the star they are, which to some degree resembles the pattern introduced by the instellation dependency of the magma ocean probability. However, the Neptune desert occurs at smaller orbital periods and its location depends on a planet's size (Szabó & Kiss 2011), which is not expected for the runaway greenhouse transition.

is there overlap/confusion with the radius/density trend in ??

Maps to Sect. 3.4.

Penny et al. (2019) mention that a significant increase in planet yield could be achieved if the telescope's slew speed was increased.

4.2.2. *OPTIONAL...: Atmospheric signatures*

discuss potential atmospheric signatures of magma oceans. e.g.: H/O ratios of sub-Neptunes could be low (because H is in the melt) (Tim's talk)

4.3. *Diagnostic power of upcoming exoplanet missions*

PLATO: needed sample size, precision

The fidelity of a future detection or falsification of a magma ocean signature will depend on the significance with which the null hypothesis can be excluded. As a

posterior of some key parameter and it's ground truth
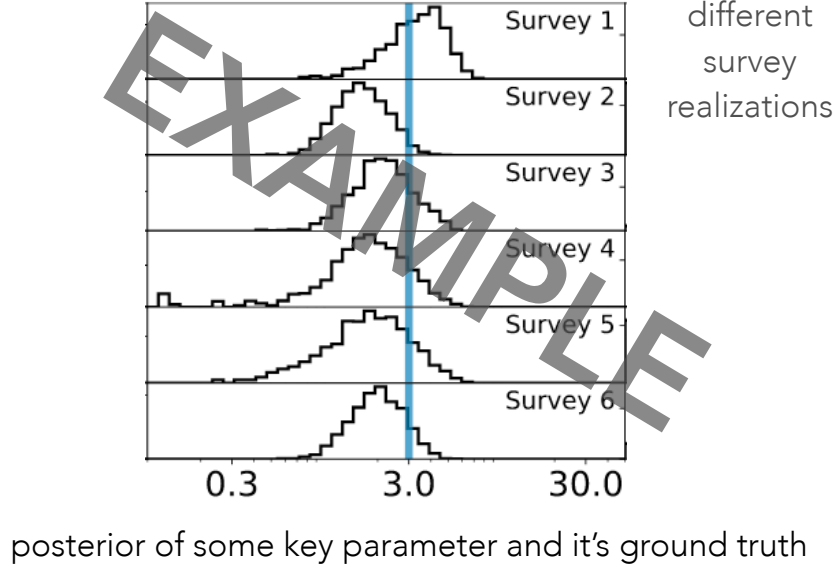
Figure 9. Retrieved posterior distribution of the $< keyparameter >$ for different survey realizations. The orange line corresponds to the true value of the injected signal.

consequence, instrumentation and survey strategy aiming at testing the hypothesis should aim at maximizing the probability of a true positive detection given the existence of the effect. ...

4.4. *Mission design trades*

We now turn to mission design trades that influence the statistical power for testing the runaway greenhouse hypothesis. ... ...FOLLOW-UP...

The advantage of mass measurements for the planet sample can be enhanced with a wise selection of follow-up targets, i.e. favoring those planets residing close to S_{thresh}.

...FGK vs. M...

To date, the majority of planets with radius measurements orbit FGK dwarfs and lie in the runaway greenhouse regime. ...

Our tests with different spectral types (Sect. 3.4.2) show that the information content of M dwarfs in a sample dominates the hypothesis tests. The reason is, besides their large number in a volume/magnitude-limited sample, that many detected M dwarf planets are located near the threshold instellation. As in the cases above, an M dwarf survey profits from follow-up measurements of planetary masses with an order of magnitude increase in evidence.

4.5. *Implications for planetary habitability*

Magma oceans influence the amount of water available at the surface and atmosphere, which is 1. commonly used as environmental marker to assess habitability, and 2. influences the planet's climate (greenhouse effect).

5. CONCLUSIONS

We have explored the potential of contemporary exoplanet missions to statistically detect a radius discontinuity that is predicted to result from runaway greenhouse phases. ...Bioverse...

The key findings of our work are as follows:

1. The predicted tipping point in planetary climates causes a discontinuity in the radius and density distribution of small exoplanets.

2. The demographic imprint of the runaway greenhouse threshold should be detectable with transit measurements of a large enough sample and sufficient precision.

3. We find that the planned *PLATO* transit survey will provide a sufficient sample to test the runaway greenhouse hypothesis.

4. The diagnostic power of transit missions in testing this hypothesis can be maximized by ...

5. A detection of the feature will constrain the water inventory of rocky exoplanets, providing important context in the assessment of their habitability.

Author contributions:

DATA AVAILABILITY

All data sets and software required to reproduce our results and this article are available through GitHub and Zenodo. The code to reproduce the figures can be accessed via the icon links associated with the respective figure caption.

Software: Bioverse (Bixel & Apai 2021), Astropy (Astropy Collaboration et al. 2018), NumPy (Harris et al. 2020), SciPy (Virtanen et al. 2020), corner.py (Foreman-Mackey 2016), dynesty (Speagle 2020).

REFERENCES

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, The Astronomical Journal, 156, 123, doi: 10.3847/1538-3881/aabc4f

Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 1998, Astronomy and Astrophysics, v.337, p.403-412 (1998), 337, 403

Barnes, J. W. 2007, Publications of the Astronomical Society of the Pacific, 119, 986, doi: 10.1086/522039

Barth, P., Carone, L., Barnes, R., et al. 2021, Astrobiology, 21, 1325, doi: 10.1089/ast.2020.2277

Benz, W., Broeg, C., Fortier, A., et al. 2021, Experimental Astronomy, 51, 109, doi: 10.1007/s10686-020-09679-4

Bergsten, G. J., Pascucci, I., Mulders, G. D., Fernandes, R. B., & Koskinen, T. T. 2022, The Demographics of Kepler's Earths and Super-Earths into the Habitable Zone

Bixel, A., & Apai, D. 2020, The Astrophysical Journal, 896, 131, doi: 10.3847/1538-4357/ab8fad

—. 2021, The Astronomical Journal, 161, 228, doi: 10.3847/1538-3881/abe042

20

Bower, D. J., Kitzmann, D., Wolf, A. S., et al. 2019, Astronomy & Astrophysics, 631, A103, doi: 10.1051/0004-6361/201935710

Burke, C. J., Christiansen, J. L., Mullally, F., et al. 2015, The Astrophysical Journal, 809, 8, doi: 10.1088/0004-637X/809/1/8

Dorn, C., & Lichtenberg, T. 2021, The Astrophysical Journal Letters, 922, L4, doi: 10.3847/2041-8213/ac33af

Downey, B. G., Nimmo, F., & Matsuyama, I. 2022, The Thermal-Orbital Evolution of the Earth-Moon System with a Subsurface Magma Ocean and Fossil Figure

Dreizler, S., I., J., et al. 2020, Astronomy & Astrophysics

Elkins-Tanton, L. T. 2013, Nature, 497, 570, doi: 10.1038/497570a

ESA. 2017, PLATO DEFINITION STUDY REPORT (RED BOOK), Tech. Rep. ESA-SCI(2017)1, European Space Agency

Foreman-Mackey, D. 2016, Journal of Open Source Software, 1, 24, doi: 10.21105/joss.00024

Fulton, B. J., Petigura, E. A., Howard, A. W., et al. 2017, The Astronomical Journal, 154, 109, doi: 10.3847/1538-3881/aa80eb

Ginzburg, S., Schlichting, H. E., & Sari, R. 2016, The Astrophysical Journal, 825, 29, doi: 10.3847/0004-637x/825/1/29

—. 2018, Monthly Notices of the Royal Astronomical Society, 476, 759, doi: 10.1093/mnras/sty290

Goldblatt, C., Robinson, T. D., Zahnle, K. J., & Crisp, D. 2013, Nature Geoscience, 6, 661, doi: 10.1038/ngeo1892

Goldblatt, C., & Watson, A. J. 2012, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 370, 4197, doi: 10.1098/rsta.2012.0004

Gupta, A., & Schlichting, H. E. 2019, Monthly Notices of the Royal Astronomical Society, 487, 24, doi: 10.1093/mnras/stz1230

Hamano, K., Abe, Y., & Genda, H. 2013, Nature, 497, 607, doi: 10.1038/nature12163

Hamano, K., Kawahara, H., Abe, Y., Onishi, M., & Hashimoto, G. L. 2015, The Astrophysical Journal, 806, 216, doi: 10.1088/0004-637X/806/2/216

Hardegree-Ullman, K. K., Cushing, M. C., Muirhead, P. S., & Christiansen, J. L. 2019, The Astronomical Journal, 158, 75, doi: 10.3847/1538-3881/ab21d2

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357, doi: 10.1038/s41586-020-2649-2

Jin, S., Mordasini, C., Parmentier, V., et al. 2014, Astrophysical Journal, 795, doi: 10.1088/0004-637X/795/1/65

Kipping, D. M. 2013, Monthly Notices of the Royal Astronomical Society: Letters, 434, L51, doi: 10.1093/mnrasl/slt075

Leconte, J., Forget, F., Charnay, B., Wordsworth, R., & Pottier, A. 2013, Nature, 504, 268, doi: 10.1038/nature12827

Lichtenberg, T., Schaefer, L. K., Nakajima, M., & Fischer, R. A. 2022, Geophysical Evolution During Rocky Planet Formation

Luger, R., & Barnes, R. 2015, Astrobiology, 15, 119, doi: 10.1089/ast.2014.1231

Mazeh, T., Holczer, T., & Faigler, S. 2016, Astronomy and Astrophysics, 589, doi: 10.1051/0004-6361/201528065

Mordasini, C. 2020, Astronomy and Astrophysics, 638, 1, doi: 10.1051/0004-6361/201935541

Owen, J. E., & Wu, Y. 2013, Astrophysical Journal, 775, 1, doi: 10.1088/0004-637X/775/2/105

Penny, M. T., Gaudi, B. S., Kerins, E., et al. 2019, The Astrophysical Journal Supplement Series, 241, 3, doi: 10.3847/1538-4365/aafb69

Puig, L., Pilbratt, G. L., Heske, A., Sanz, I. E., & Crouzet, P.-E. 2016, in Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave, Vol. 9904 (SPIE), 649–657, doi: 10.1117/12.2230964

Rauer, H., Aerts, C., Cabrera, J., & PLATO Team. 2016, Astronomische Nachrichten, 337, 961, doi: 10.1002/asna.201612408

Skilling, J. 2004, AIP Conference Proceedings, 735, 395, doi: 10.1063/1.1835238

Speagle, J. S. 2020, Monthly Notices of the Royal Astronomical Society, doi: 10.1093/mnras/staa278

Szabó, G. M., & Kiss, L. L. 2011, Astrophysical Journal Letters, 727, 2, doi: 10.1088/2041-8205/727/2/L44

Turbet, M., Bolmont, E., Ehrenreich, D., et al. 2020, Astronomy & Astrophysics, Volume 638, id.A41, <NUMPAGES>10</NUMPAGES> pp., 638, A41, doi: 10.1051/0004-6361/201937151

Turbet, M., Ehrenreich, D., Lovis, C., Bolmont, E., & Fauchez, T. 2019, Astronomy & Astrophysics, 628, A12, doi: 10.1051/0004-6361/201935585

Van Eylen, V., Agentoft, C., Lundkvist, M. S., et al. 2018, Monthly Notices of the Royal Astronomical Society, 479, 4786, doi: 10.1093/mnras/sty1783

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261, doi: 10.1038/s41592-019-0686-2

Youdin, A. N. 2011, Astrophysical Journal, 742, doi: 10.1088/0004-637X/742/1/38

Zeng, L., Sasselov, D. D., & Jacobsen, S. B. 2016, The Astrophysical Journal, 819, 127, doi: 10.3847/0004-637X/819/2/127