

# How Internal Structure Shapes the Metallicity of Giant Exoplanets

Lorenzo Peerani<sup>1</sup>, Saburo Howard<sup>1</sup>, and Ravit Helled<sup>1</sup>

Institut für Astrophysik, Universität Zürich, Winterthurerstr. 190, CH8057 Zurich, Switzerland,  
e-mail: lorenzo.peerani@uzh.ch

Received Month date, year; accepted Month date, year

## ABSTRACT

**Context.** The composition and internal structure of gas giant exoplanets encode key information about their formation and subsequent evolution.

**Aims.** We investigate how different interior structure assumptions affect the inferred bulk metallicity and its correlation with planetary mass.

**Methods.** For a sample of 44 giant exoplanets ( $0.12\text{--}5.98 M_J$ ), we compute evolutionary models with CEPAM and retrieve their bulk metallicities under three structural hypotheses: Core+Envelope (CE), Dilute Core (DC), and Fully Mixed (FM).

**Results.** Across all structures, we recover a significant positive correlation between total heavy-element mass ( $M_Z$ ) and planetary mass ( $M$ ), and a negative correlation between metallicity ( $Z$ ) and  $M$  (also for  $Z/Z_\star$  vs.  $M$ ). DC structures yield metallicities comparable to CE models, regardless of the assumed gradient extent. Increasing atmospheric metallicity raises the inferred bulk metallicity, as enhanced opacities slow planetary cooling. Non-adiabatic DC models can further increase the retrieved metallicity by up to 35%. Sensitivity analyses show that the mass–metallicity anti-correlation is primarily driven by low-mass, metal-rich planets ( $M < 0.2 M_J$ ), while massive planets ( $\gtrsim 1 M_J$ ) exhibit unexpectedly high metallicities ( $Z \sim 0.1\text{--}0.3$ ).

**Conclusions.** Improved constraints on convective mixing, combined with upcoming precise measurements of planetary masses, radii, and atmospheric compositions from missions such as *PLATO* and *Ariel*, will enable more robust inferences of interior structures and formation pathways for gas giant planets.

**Key words.** planets and satellites: gaseous planets – planets and satellites: interiors – planets and satellites: composition –

## 1. Introduction

Inferring a planet’s bulk composition and internal structure provides a powerful diagnostic of its formation pathway (Helled &odenheimer 2011; Mordasini et al. 2016; Turrini et al. 2021). Gas giants represent an ideal case study in this context: as the largest planets, they must have formed early, when ample gas and solids were still available for accretion. Their size and rapid growth mean that they preserve valuable information about the composition of the protoplanetary disk (Pollack et al. 1996; Helled & Morbidelli 2021; Mordasini et al. 2016). Hence, constraining their composition is key to probing the conditions under which planets form.

The observational landscape now includes over 6,000 confirmed exoplanets (NASA Exoplanet Archive 2025), with a large, well-characterised subset of giants having measured masses and radii from space and ground-based surveys (e.g. *Kepler*, *TESS*, *WASP*, *HAT/HATSouth*). The large dataset allows not only inferences of heavy-element content in individual planets, but also statistical analyses across populations, enabling the identification of universal trends constrained by bulk properties.

Interior models aim to reproduce the observed bulk properties by assuming specific distributions of heavy elements. These models often explore two simple end-member configurations: a Core+Envelope structure, where a distinct heavy-element core lies beneath a H/He envelope, and a Fully Mixed structure, where the heavy elements are homogeneously distributed throughout the planet (Thorngren et al. 2016; Howard et al. 2025b). However, gravity-field measurements and ring seismology for Jupiter and Saturn suggest more complex interiors. Interior models tai-

lored to match these data suggest the interior of the two gas giants have compositional gradient of heavy elements or ‘dilute cores’ (Wahl et al. 2017; Debras & Chabrier 2019; Mankovich & Fuller 2021; Miguel et al. 2022; Militzer et al. 2022; Howard et al. 2023; Ziv et al. 2024). Nevertheless, efforts in applying such a dilute core structure to exoplanetary studies is still embryonic. Recent retrieval studies have begun testing such structural prescriptions in individual exoplanets. For instance, Bloor et al. (2023) applied homogeneous and inhomogeneous models within a Bayesian framework to a limited sample, highlighting strong degeneracies that prevent quantitative discrimination between the two. Similarly, van Dijk & Miguel (2025) extended this approach by incorporating hot Jupiters, showing that current observations allow only qualitative assessments for a few well-characterized cases. Motivated by these findings, we extend interior modeling beyond the traditional end-member cases to include composition gradients and dilute-core structures that capture gradual heavy-element distributions.

In this work, we assess how the inferred bulk metallicity ( $Z$ ) depends on the assumed interior structure for a sample of gas giant exoplanets. We further examine how different structural assumptions affect the resulting population-level trends. This study is important for the interpretation of data from JWST and upcoming missions such as *PLATO* and *Ariel*, which will provide precise measurements of planetary radii, ages, and atmospheric compositions, enabling tighter constraints on exoplanet interiors and formation histories. (Rauer et al. 2014; Tinetti et al. 2018).

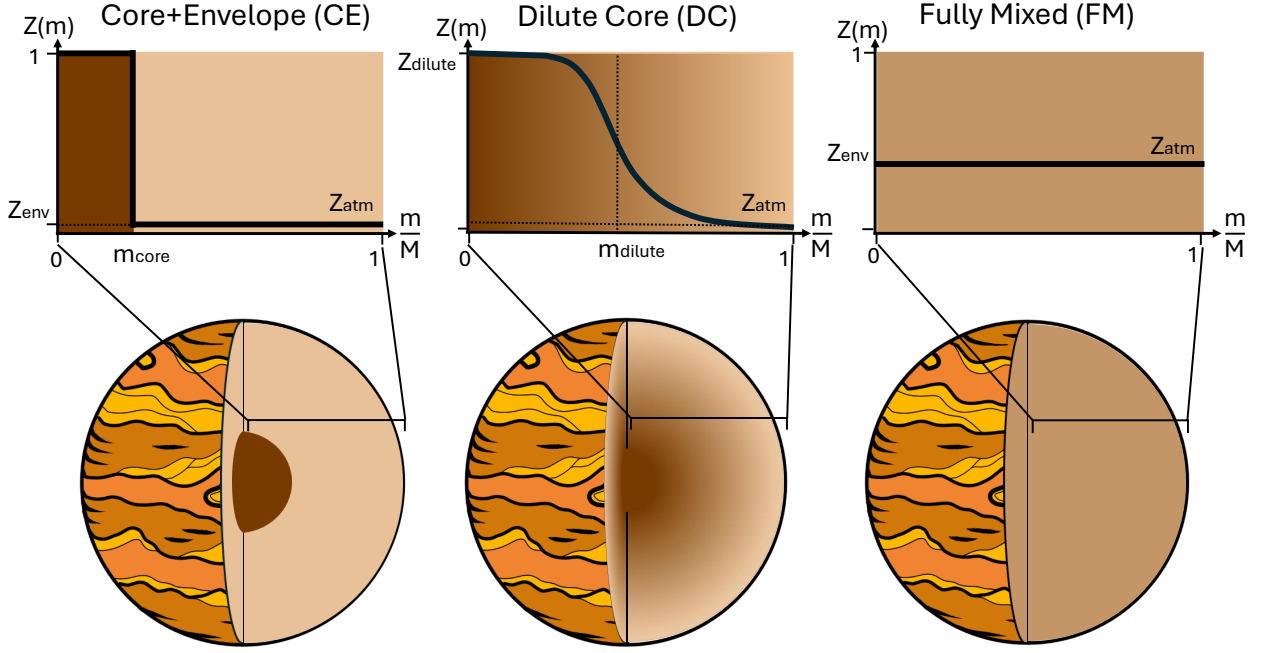


Fig. 1: Three main internal structures tested in this study, with their respective heavy elements structure profiles: Core+Envelope (CE): all the heavy elements situated in a well defined core, Dilute Core, a gradient applied to the planet’s core, and Fully Mixed: an homogeneous envelope throughout the planet.

## 2. Methods

### 2.1. Model parameters

We used CEPAM (Guillot & Morel 1995; Howard et al. 2024) to compute planetary evolution. CEPAM numerically solves the 1D planetary structure and thermal-evolution equations (mass conservation, hydrostatic equilibrium, energy transport, and energy conservation) coupled to equations of state, that set density and entropy as functions of P, T, and composition. The sample size consists of 44 exoplanets taken from the Otegi et al. (2020); Parc et al. (2024) with masses between 0.12 and 5.98  $M_J$ . As in Howard et al. (2025b), we considered exoplanets with radius, mass, age, and  $T_{eq}$  measurements and with relative measurement uncertainties smaller than 5 and 10 % for radius and mass, respectively. Exoplanets with equilibrium temperatures  $T_{eq} > 1000$  K were excluded from this study, since additional energy sources can significantly inflate their radii (Guillot et al. 2006; Thorngren & Fortney 2018). Such inflation mechanisms are not considered, and only planets for which our models reproduce the observed radius at the reported age were included.

The ice-to-rock ratio is kept constant for all the tested structures, and assumed to be a 1:1 ratio. For hydrogen and helium, the ratio between the two elements is assumed to be protosolar ( $Y/(X + Y) = 0.27$ ) (Asplund et al. 2021).

As shown by Howard et al. (2025b), the equation of states (EoS) used are important, and they affect the results inferred by the evolution models. The EoSs used in this study are the SESAME water and dry sand for ices and rocks (Lyon & Johnson 1992), and the CMS19+HG23 EoS for H-He (Chabrier et al. 2019; Howard & Guillot 2023).

The atmosphere in the model is treated as a non-gray, fully radiative atmosphere. Opacities are treated as two bands (visible and IR), and the mean opacities have been tabulated accordingly (Freedman et al. 2008, 2014; Parmentier et al. 2015). Opacity

enhancement caused by heavy elements follows the methodology described in Valencia et al. (2013). For simplicity, we do not account for compositional mixing of heavy elements (the structure does not evolve over time). For all structures, the P-T profile is assumed to be adiabatic in the planet’s interior.

### 2.2. Tested interior structures

We considered different interior structures for the planetary dataset. The three main tested structures are Core+Envelope, Fully Mixed and Dilute Core. They can be described as follows:

- **Core+Envelope (CE):** this structure assumes a well-defined core consisting of heavy elements, with the envelope being composed of H-He. As the metallicity in the core  $Z_{core}$  is equal to 1, we use the core mass  $M_{core}$  as a free parameter to infer bulk metallicities. For the basic Core+Envelope scenario, no heavy elements are inputted in the envelope and the atmosphere ( $Z_{env} = Z_{atm} = 0$ ), where  $Z_{env}$  is the metallicity in the envelope and  $Z_{atm}$  in the atmosphere).
- **Core+Envelope with stellar atmosphere (CEA):** To assess the contribution of metals mixed into the envelope, we allow heavy elements in the upper planetary layers  $Z_{env} = Z_{atm} > 0$ .
- **Fully Mixed (FM):** this structure assumes homogeneous composition throughout the planetary interior. No core is present. The bulk metallicity is equal to  $Z_{env} = Z_{atm}$ .
- **Dilute Core (DC):** this structure represents a gradual distribution of heavy elements. Similarly to what done in Miguel (2022); Bloot et al. (2023); Howard et al. (2023), the structure is being described with the following function:

$$Z(m) = Z_{env} + \frac{Z_{dilute} - Z_{env}}{2} \left[ 1 - \operatorname{erf} \left( \frac{m - m_{dilute}}{\delta m} \right) \right], \quad (1)$$

where  $Z_{dilute}$  is the maximum heavy–element fraction in the dilute core, the parameter  $\delta m$  controls the steepness of the

gradient, and  $m_{\text{dilute}}$  is the mass coordinate at which the gradient is steepest (see Fig. 1).  $\delta m$  has been chosen to be fixed at 0.125 (as variations in this parameter are not believed to be substantially affecting planetary  $Z$  estimate (Bloom et al. 2023)). For the basic Dilute Core structure,  $Z_{\text{env}} = Z_{\text{atm}} = 0$ .

- **Dilute Cores with Stellar and (Super)stellar atmospheres (DCA and DCA3):** similar to DC, but now we allow atmospheric metallicity. The atmospheric abundances tested are matching the host star’s ( $Z_{\star}$ ) metal abundance (so  $Z_{\text{atm}} = Z_{\star}$ ). Similar to Jupiter’s atmospheric enrichment (Helled & Stevenson 2024), a more enriched case with  $Z_{\text{atm}} = 3 \times Z_{\star}$  has also been tested (named DCA3). To note that in these cases  $m_{\text{dilute}}$  will be fixed to a value of 0.5.

For each structure consistent with the observational constraints from each planet, we report  $M_Z$  (heavy-element mass) and  $Z$  (bulk metallicity), related by  $Z = M_Z/M$ ,  $M$  being the planetary mass.

### 2.3. Root-find $Z$ retrieval

In this section, we describe the statistical tool designed to retrieve the possible ranges of  $Z$  that explain each planet’s radius for every structural assumption. The retrieval of planetary bulk metallicities  $Z$  was designed as an inversion of the evolution model: instead of using  $M$  and  $Z$  to predict  $(R, \text{Age})$ , we seek the metallicity that reproduces the observed  $(M, R, \text{Age})$  within their uncertainties. Our approach incorporates observational uncertainties by treating the mass and radius as normally distributed around their observed values ( $M \sim \mathcal{N}(M_{\text{obs}}, \sigma_M)$ ,  $R \sim \mathcal{N}(R_{\text{obs}}, \sigma_R)$ ), while the age is assumed to follow a uniform distribution ( $\text{Age} \sim \mathcal{U}(\text{Age}_{\text{min}}, \text{Age}_{\text{max}})$ ). The method proceeds in three steps:

- **Step 1 – Evolution runs:** We first generated a set of cooling tracks with CEPAM across a broad grid of input parameters. Metallicities were sampled in the range  $0 \leq Z \leq 1$  with steps of 0.025, while masses were varied around the observed mean value, including the central value,  $\pm 1\sigma$ ,  $\pm 2\sigma$ , and  $\pm 3\sigma$ . This ensemble of runs ensures that the parameter space captures both the observational uncertainties in mass and the possible degeneracies between metallicity and mass in shaping planetary radii. A dense coverage of  $Z$  values is particularly important for detecting any degeneracies in the  $M$ – $Z$ – $R$  relation.
- **Step 2 – Grid interpolation:** The model outputs from CEPAM (describable as a Radius/Age relation for different structures and chemical abundances) were interpolated to build a continuous parameter grid. While linear interpolation is generally sufficient, we employed cubic spline interpolation (via `scipy.interpolate.interp1d`) since it better resolves curved behaviours and possible degeneracies in the cooling tracks. The interpolated grid was refined to a resolution of  $\Delta Z = 0.0025$ , corresponding to 400 metallicity steps<sup>1</sup>. This step provides smooth predictions of planetary radii across the three-dimensional space  $(M, Z, \text{Age})$ , which are later required for the inversion procedure.
- **Step 3 – Root-finding and sampling:** To invert the grid and

estimate  $Z$ , we repeatedly sampled observational values of mass, radius, and age. For each planet, we drew  $n = 5000$  tuples  $(M, R, \text{Age})$ . For each tuple, Brent’s method was applied to solve

$$f(Z) = R_{\text{pred}}(Z, M, \text{Age}) - R_{\text{obs}} = 0, \quad (2)$$

where  $R_{\text{pred}}$  is the interpolated model prediction and  $R_{\text{obs}}$  is the sampled observational value. This procedure yields the metallicity  $Z$  that reproduces the observed radius. Repeating this process 5000 times provides a posterior distribution of  $Z$ , capturing both measurement uncertainties and intrinsic model degeneracies. A similar approach can be found in Müller & Helled (2023).

The root-finding retrieval provides the posterior distributions of metallicity  $Z$  for each assumed interior structure and for every exoplanet analysed. These distributions were smoothed using a Kernel Density Estimate (KDE), and the corresponding mean and standard deviation of  $Z$  were extracted and subsequently used to investigate planetary trends.

## 3. Results

### 3.1. Structural properties

Figure 2 shows the retrieved curves for CE, FM, and three DC structures that assume a different extent of the dilute core ( $m_{\text{dilute}} = 0.25, 0.5$  or  $0.75$ ). Although the analysis is performed for the full sample, we first focus on three planets with representative masses: K2-287b ( $0.315 M_J$ ), TOI-1811b ( $0.972 M_J$ ), and TOI-5153b ( $3.26 M_J$ ). In all cases, CE yields a lower bulk metallicity than FM. We verified this offset by reproducing the approach of Howard et al. (2025b); the results are consistent with our pipeline (see Appendix B). The physical driver is that, in FM, heavy elements mixed into the envelope and atmosphere increase the mean molecular weight and, critically, the Rosseland mean opacity in the radiative zone. The higher opacity slows cooling, and keeps radii larger at fixed age; matching the observed radius then requires a larger heavy-element mass than in the CE case. Howard et al. (2025b) likewise found FM evolution tracks to be systematically warmer, especially at early ages. The opacity effect is the main contributor in FM structures for higher expected bulk metallicities.

When examining the DC curves at different  $m_{\text{dilute}}$  values (Fig. 2), we find that DC models generally underestimate  $Z$  compared to CE and FM, and their solutions more closely resemble CE than FM. No systematic distinction or trend is evident across the three different  $m_{\text{dilute}}$  values tested, as the model-to-model differences (and their intrinsic uncertainties) remain smaller than the observational errors. Since these DC structures exclude atmospheric metallicity ( $Z_{\text{atm}} = 0$ ), the observed effects are driven solely by the internal distribution of heavy elements. Introducing a dilution gradient, rather than a sharp core boundary, redistributes heavy material into outer layers, increasing local density and pressure. This accelerates planetary contraction over time, meaning that a smaller heavy-element mass is required to reproduce the observed radius and age. Nonetheless, such an effect appears to be of lower magnitude compared to the FM-CE estimates difference.

After testing different  $m_{\text{dilute}}$  structures, we now examine dilute core models with atmospheric metallicities set to one and three times the stellar value (DCA and DCA3, respectively). Results for the three representative planets are shown in Fig. 3,

<sup>1</sup> The choice of step sizes reflects a balance between accuracy and computational time. Tests on two representative planets with different  $M$ – $Z$  step combinations showed that denser sampling in metallicity is more critical than in mass. Such steps are assumed to be precise enough for the whole dataset. Future data with better observational constraints, or simply applying such a grid to Jupiter and Saturn, might be too coarse and unreliable, and it would require adjustments.

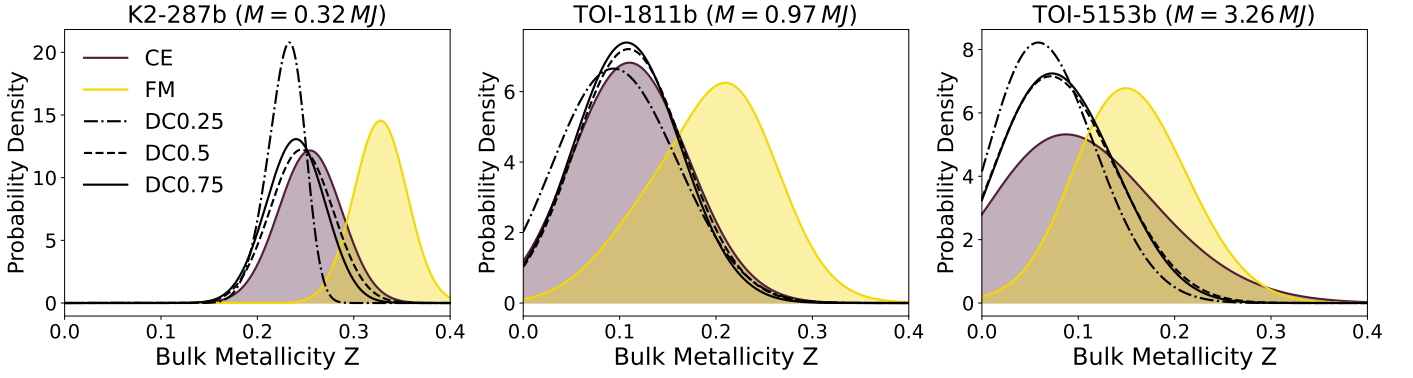


Fig. 2: Retrieved curves for CE, FM, and DC0.25-0.5-0.75 structures across three different planetary cases for different masses. CE curves are shown in purple, FM in yellow, and the DC curves are in black, and can be distinguished through the different line style.

while the full dataset, including mean values and standard deviations for all structures, is reported in Appendix A. Since  $Z_{\text{atm}}$  (and thus the opacity effect) is the varying parameter, DCA and DCA3 directly probe its effect on  $Z$  estimates. As Fig. 3 illustrates, increasing atmospheric metallicity systematically shifts the inferred  $Z$  upward. As a result, in the majority of cases DCA expects higher heavy elements in planetary interiors, followed by DCA3 and FM. In the three cases shown, both DCA and DCA3 yield higher  $Z$  than CE. For some planets (e.g., HAT-P-12b, TOI-2010b, TOI-2180b), however, DCA results converge with DC models, suggesting that opacity effects are insufficient to alter the estimates. Overall, Fig. 3 confirms that opacity is the primary driver of the retrieved heavy-element abundance, whereas density redistribution alone produces comparatively minor shifts. The CEA (bright-red) curves closely track the DCA estimates (especially for TOI-1811b and TOI-5153b in the figure), indicating that both models recover similar  $Z$ . Because they use identical opacities and differ only in interior structure (distinct core in CEA vs. composition gradient in DCA), this agreement reinforces that opacity is the dominant lever over density redistribution.

In the absence of additional heat sources, opacity controls the degree of radius inflation in our models. Consequently, the magnitude of the opacity effect is planet-specific and cannot be captured by a single global scaling (reflecting planetary variations between transport, boundary conditions, and composition degeneracies). Furthermore, fixed atmospheric constraints such

as a three times  $Z_{\star}$  is ruling out possible structural scenarios for low- $Z$ , high- $M$  planetary systems, where there are cases where DCA3 and FM models give similar results (like in the case of TOI-5153b in Fig.3), or even in some cases the DCA3 estimates exceed the FM ones (like TOI-4127b or WASP-162b cases that can be found in the Appendix A). An atmospheric enrichment of  $3 Z_{\star}$  therefore creates a bias by the model in pushing metallicities of low estimated  $Z$  up. Tightening atmospheric constraints (e.g., on metallicity and thermal structure) therefore becomes crucial for breaking high- $M$  degeneracies.

### 3.2. Mass-Metallicity Relation

After analysing the overall behaviour of metallicity estimates across different structural models, we next examined the full dataset of 44 planets for possible population-level trends in the inferred composition. In particular, we investigated the heavy-element mass ( $M_Z$ ) and the bulk metallicity ( $Z$ ), and how they scale with planetary mass ( $M$ ). To assess these relations, we performed a Bayesian regression. The parameters of the power-law  $M_Z [M_{\oplus}] = \beta M [M_J]^{\alpha}$  are estimated using a Markov chain Monte Carlo (MCMC, Python package emcee (Foreman-Mackey et al. 2013), following Müller & Helled (2023) and Howard et al. (2025b)). Posterior distributions of the slope ( $\alpha$ ) and intercept ( $\beta$ ) were sampled from 5,000 chains, taking into account  $1\sigma$  uncertainties in both axes. Medians and standard de-

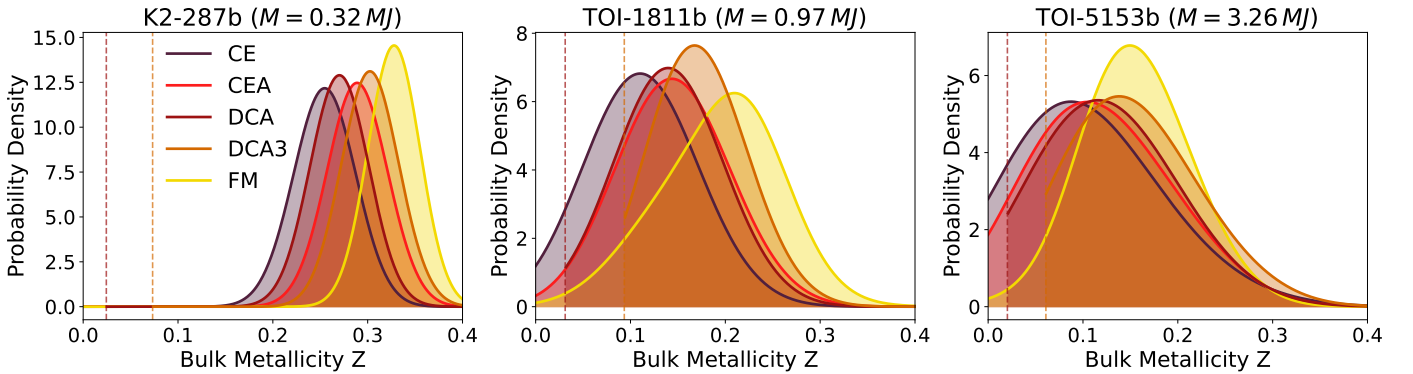


Fig. 3: Retrieved curves for CE, FM, Core+Envelope + Stellar atmospheric composition (CEA), Dilute Core with 1 and  $3 Z_{\star}$  (DCA and DCA3), in vinaccia and orange colour respectively. Dashed lines (in wine colour and orange colour) show the atmospheric abundance assumed for DCA and DCA3 models respectively. Such lines will act as a lower threshold for these models as the assumed  $R_{\text{atm}}$  will be kept fixed for each planet case.

viations have been derived from these posteriors are shown in Fig. 4 for the four structural models (FM, DCA3, DCA, CE). All models reveal a positive sloping between  $M_Z$  and  $M$ , with similar slopes but offsets in intercept depending on the assumed structure. FM systematically predicts the highest  $M_Z$ , while CE predicts the lowest (with  $\beta \approx 36$ ). This scaling is broadly consistent with the relation reported by Thorngren et al. (2016), Howard et al. (2025b), and Chachan et al. (2025). Our results suggest a steeper dependence, with DCA3 yielding  $\alpha > 0.7 \pm 0.07$ , only marginally inconsistent with the Thorngren’s relation of  $M_Z \propto \sqrt{M}$  when uncertainties are considered.

For this reason, an inspection of the reliability of such curves has been done by testing the correlation between  $\alpha$  and  $\beta$ . The covariance analysis shows a weak but positive coupling ( $\rho \sim 0.2$  for all the models), indicating that slope and intercept are not fully independent, and that small shifts in one parameter do affect the other. We then also evaluated the fits with a weighted least-squares (WLS)  $\chi^2$  test. In all cases,  $\chi^2 \gg 1$ , meaning the power-law relations do not formally explain the data within the quoted errors, and reliance of such curves should be treated with caution (see Section D for more information about the WLS goodness-of-fit approach).

Since the weighted least-squares test revealed the limited reliability of the retrieved power-law fits, it is essential to verify whether the apparent slopes of these relations genuinely reflect an underlying correlation within the data. We further tested the correlations of the dataset for each structural assumption using the non-parametric Kendall’s  $\tau$  rank test. To account for observational uncertainties, instead of computing the correlation on the data’s mean values, we computed a bootstrap resampling approach was employed: 5,000 synthetic datasets were generated by drawing from the reported error distributions, and  $\tau$  was calculated for each realization. The median  $\tau$  values and associated

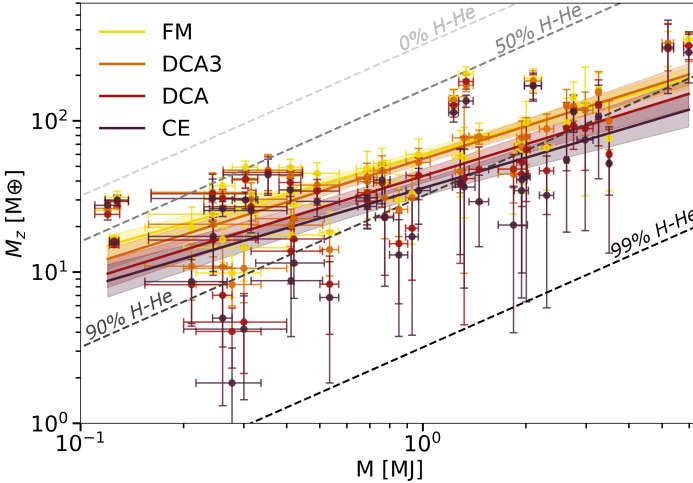


Fig. 4: Relationship between the retrieved heavy element mass  $M_Z$  and planetary mass  $M$ , shown for the four main structures tested (CE, FM, DCA, and DCA3). Shown as dashed lines in gray gradient are the different curves showing how the trend looks for fixed planetary compositions (0%, 50%, 90%, and 99% gas composition respectively). Bayesian fits have been shown using a power law function  $M_Z = \beta M^\alpha$ . Median values of  $\beta$  are respectively 59.57 for FM, 56.02 for DCA3, 43.11 for DCA, and 35.9 for CE. Values for  $\alpha$  are 0.653 for FM, 0.718 for DCA3, 0.700 for DCA, and 0.667 for CE. Posterior uncertainties  $\sigma_\alpha$  and  $\sigma_\beta$  are respectively 0.068 and 4.5 for FM, 0.074 and 4.3 for DCA3, 0.098 and 4.4 for DCA, and 0.113 and 4.5 for CE.

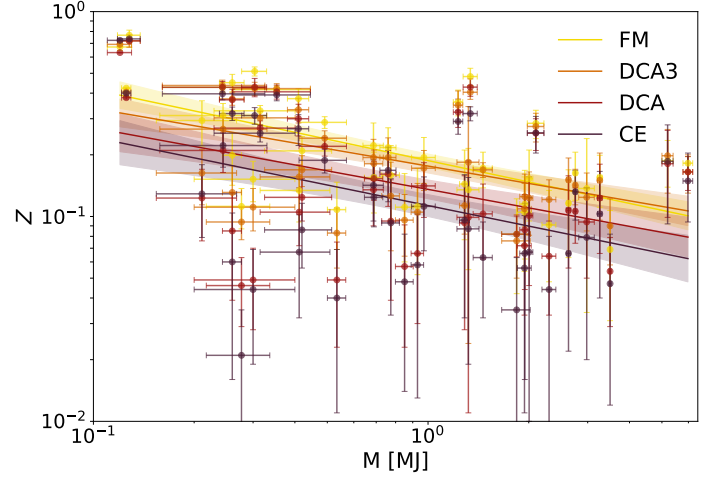


Fig. 5: Relationship between the retrieved heavy element mass  $Z$  and planetary mass  $M$ , shown for the four main structures tested (CE, FM, DCA, and DCA3). Bayesian fits have been shown using a power law function  $Z = \beta M^\alpha$ . Median values of  $\beta$  are respectively 0.187 for FM, 0.176 for DCA3, 0.136 for DCA, and 0.113 for CE. Values for  $\alpha$  are -0.347 for FM, -0.282 for DCA3, -0.3 for DCA, and -0.333 for CE. Posterior uncertainty  $\sigma_\alpha$  and  $\sigma_\beta$  are respectively 0.069 and 0.014 for FM, 0.073 and 0.014 for DCA3, 0.097 and 0.015 for DCA, and 0.115 and 0.014 for CE.

p-values are reported in Table 1. For the  $M_Z$ – $M$  relation, nearly all runs returned p-values well below the adopted significance threshold of  $p = 0.05$ , confirming the robustness of the positive correlation.

| Model | (a) $M_Z$ vs. $M$ |                        | (b) $Z$ vs. $M$ |                       |
|-------|-------------------|------------------------|-----------------|-----------------------|
|       | $\tau$            | p-value                | $\tau$          | p-value               |
| CE    | 0.514             | $1.20 \times 10^{-6}$  | -0.273          | $1.00 \times 10^{-2}$ |
| DCA   | 0.612             | $7.59 \times 10^{-9}$  | -0.245          | $2.07 \times 10^{-2}$ |
| DCA3  | 0.660             | $4.44 \times 10^{-10}$ | -0.314          | $3.05 \times 10^{-3}$ |
| FM    | 0.660             | $4.44 \times 10^{-10}$ | -0.397          | $1.79 \times 10^{-4}$ |

Table 1: Kendall’s  $\tau$  coefficients and associated p-values for four structural models (CE, DCA, DCA3, FM) across the two tested relationships.

We then examined the relation between  $Z$  and  $M$ . In contrast to the  $M_Z$ – $M$  trend, the  $Z$ – $M$  relation shows a clear negative correlation. The ranking of models remains similar: FM predicts the highest average  $Z$ , followed by DCA3, DCA, and CE. Compared to the  $M_Z$ – $M$  relation, the data are more scattered, particularly at lower planetary masses. Kendall’s  $\tau$  again indicates statistically significant negative correlations.

(Table 1), although the correlation strength is weaker. A portion of the reduced significance likely reflects mathematical coupling from dividing by the  $x$ -axis variable ( $Z = M_Z/M$ ), which weakens the apparent correlation and inflates p-values. Median p-values are still generally below 0.05, but with a modest fraction of realizations exceeding the threshold ( $\sim 2\%$  for FM and DCA3, and  $\sim 10$ – $11\%$  for DCA and CE). Despite this, the correlations remain statistically compelling across all models.

Such a relationship has also been tested with the normalisation to the host star ( $Z/Z_\star$  vs  $M$ ). The results show similar trends (both in best fit and correlation tests) to those obtained for  $Z/M$ .

Such an investigation would have been essential if we were artificially inputting anti-correlation given from the star’s metallicity (especially since a model assumption is the stellar composition in the planet’s atmospheric structure). Although recent studies (Buchhave et al. 2014; Müller & Helled 2025) found trends between the planetary heavy element mass and the chemical composition of their host stars, the dataset used in this study does not show any correlation between  $Z_\star$  and  $M$ .

## 4. Discussion

### 4.1. Non-adiabatic structure

In our models with a dilute core (DC, DCA and DCA3), we assumed an adiabatic temperature gradient. However, the presence

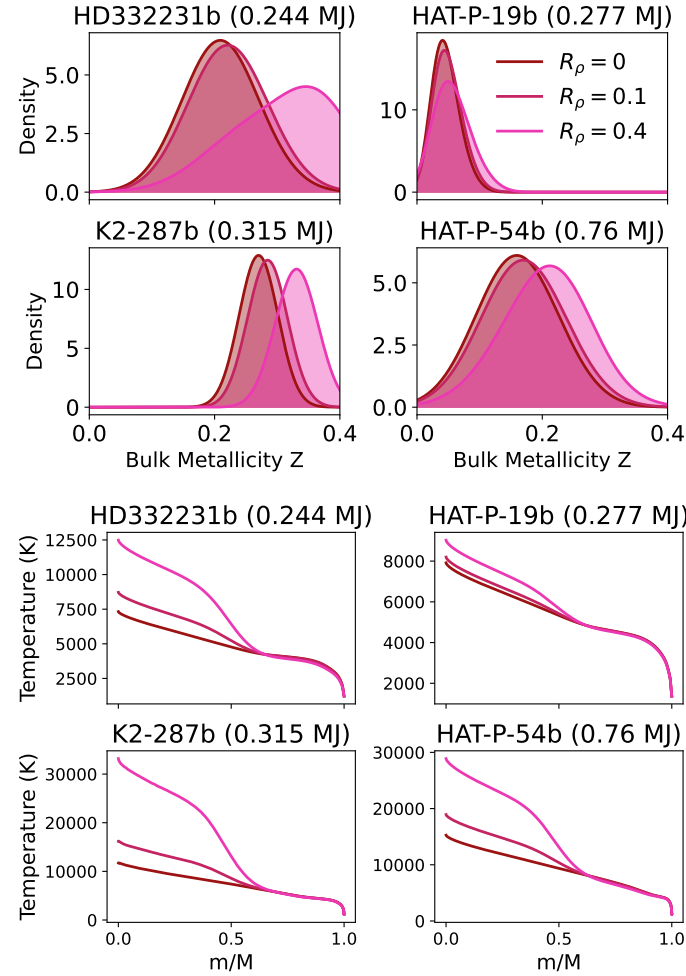


Fig. 6: Top: Four panels using the retrieved method explained in Section 2.3. Retrieved metallicity curves for a DCA model with the assumption of adiabaticity are compared with superadiabatic cases where the  $R_\rho$  scaling factor is either 0.1 or 0.4. This is shown for four different planetary cases. Bottom: Corresponding temperature profiles for the same four planets, using the median metallicity of the DCA model at a fixed age of 4 Gyr. The different heat transport scenarios show a linear increase of heat retention (and thus deviation from adiabaticity) according to the  $R_\rho$  value defined in Eq. 3. The x-axis shows the bulk metallicity  $Z$ .

of a composition gradient implies that the temperature gradient could differ from adiabatic. To assess the effect of a non-adiabatic temperature gradient on the inferred bulk metallicity, we parameterize the deviation from adiabaticity (likely caused by double-diffusive convection (Debras et al. 2021; Leconte & Chabrier 2012)) as in Howard et al. (2025a), using:

$$\nabla_T = \nabla_{\text{ad}} + R_\rho B, \quad (3)$$

where  $B = \frac{\chi_\rho}{\chi_T} \left( \frac{d \ln \rho}{d \ln Z} \right)_{P,T}$ ,  $\nabla_Z, \chi_\rho = \left( \frac{d \ln P}{d \ln \rho} \right)_{T,Z}$ ,  $\chi_T = \left( \frac{d \ln P}{d \ln T} \right)_{\rho,Z}$  and  $\nabla_Z = \frac{d \ln Z}{d \ln P}$ .

For a smaller selection of planets, we calculated evolution models using  $R_\rho = 0.1$  and  $0.4$  and compared it to the adiabatic case ( $R_\rho = 0$ ). The inferred bulk metallicities  $Z$  are shown in Fig. 6. We find that  $R_\rho = 0.1$  and  $R_\rho = 0.4$  lead to increased values of  $Z$  by up to 7 % and 35 %, respectively. The super-adiabatic temperature gradients obtained with  $R_\rho = 0.1$  and  $R_\rho = 0.4$  yield hotter interiors compared to the adiabatic case (see bottom panel of Fig. 6) and naturally lead to an increased amount of heavy-element to match the planets’ radii.

However, a clear limitation of our models is that convective mixing is not included. Such a process could affect the thermal evolution and the distribution of heavy elements. Knierim & Helled (2025) showed for a Jupiter-mass planet that neglecting mixing could overestimate the radius by  $0.08 R_J$ . The values of  $R_\rho$  we chose were arbitrary and it would be necessary to assess whether dilute cores with such a temperature structure can remain stable against convection. Future studies should further assess the effect of non-adiabaticity on the mass–metallicity relation and the thermal evolution of gas giants.

### 4.2. Connection with formation

**Extent of the dilute core.** For some high-metallicity cases ( $Z \gtrsim 0.5$ ; e.g., TOI-4010d and WASP-196b), no solutions were found for compact gradients with  $m_{\text{dilute}} = 0.25$  or  $0.5$ : the interior could not accommodate the required heavy-element fraction unless the gradient was shifted outward to larger  $m_{\text{dilute}}$ . This rules out highly compact structures and implies a dilute core that extends farther in mass and radius.

In more massive planets, a wide non-adiabatic dilute core could substantially increase the heat retained in the deep interior, altering the radius evolution. Constraining  $m_{\text{dilute}}$  is therefore crucial for massive planets. A well-confined dilute core may have little impact on inflation (the thermal profile remains close to adiabatic), whereas large  $m_{\text{dilute}}$  values can stabilize significantly more heat. In non-adiabatic cases, the steepness of the gradient may become important too.

**Mass-metallicity relation.** Our results, obtained across different structural configurations, remain consistent with previous studies (Thorngren et al. 2016; Teske et al. 2019; Howard et al. 2025b; Chachan et al. 2025). We recover a clear positive correlation between the total heavy-element mass ( $M_Z$ ) and planetary mass ( $M$ ), and a weaker yet significant negative correlation between metallicity ( $Z$ ) and  $M$ , also when normalised by ( $Z_\star$ ).

The overall  $Z$ – $M$  trend indicates a gradual decrease in metallicity with increasing mass; however, the behaviour departs from a sharp decrease of estimated metallicities at low masses (between  $0.1$  and  $0.2 M_J$ ). In the mass range of  $\sim 0.2$ – $0.6 M_J$ , we observe a drop in metallicity, consistent with the curvature captured by the fitted power law, but even more evident when expressed as an inverse linear relation, as proposed by Chachan

et al. (2025). In this regime, the data exhibit substantial scatter and a near-plateau where further increases in mass do not lead to a corresponding decline in metal content, again reminiscent of the behaviour reported by Chachan et al. (2025) above  $\sim 0.28 M_J$ . This is also explaining why the rank correlation test expressed lower correlation coefficients and higher  $p$ -values. At the high-mass end, we find hints of renewed enrichment, opposite to the expectations from classical runaway gas accretion, which would predict a dilution of heavy elements as the envelope grows.

This apparent enrichment, together with the significant dispersion across the mass–metallicity plane, likely reflects the diversity of formation pathways and interior structures among giant planets. The presence of high- $Z$ , low- $M$  objects (potentially with compositions more akin to Neptune-like planets) suggests that distinct formation channels may operate in this transition regime. On the other side, more massive planets ( $M > 4 M_J$ ) appear to be metal-rich. Even if discerning any conclusion from two planets is premature to say the least, this may suggest a secondary accretion of heavy elements through late planetesimal capture or giant impacts after disk dispersal, leading to enhanced metal enrichment in their deep envelopes (Mordasini et al. 2016; Helled & Stevenson 2017; Venturini & Helled 2020).

To assess the robustness of the inferred anti-correlation, we examined how the correlation significance changes when progressively removing planets from the mass extremes. We quantified this sensitivity by recalculating the Kendall  $\tau$  correlation and corresponding  $p$ -values after each removal (Fig. B.1). Excluding the lowest-mass planets rapidly diminishes the correlation strength for the CE and DCA models (after only two removals), and after six removals only the FM model retains marginal significance. Conversely, removing the most massive planets has a far smaller impact. This behaviour indicates that the observed  $Z$ – $M$  anti-correlation is predominantly shaped by the high- $Z$ , low- $M$  population.

These findings also raise a classification question. Planets with  $Z \gtrsim 0.5$  and low masses likely reside near the transition between Neptunian and Jovian regimes, where electron degeneracy pressure no longer dominates the  $M$ – $R$  relation (Müller et al. 2024, and references therein). Treating these as a distinct subpopulation may therefore be warranted. Although the current sample size precludes a robust identification of potential breakpoints via multi-linear or orthogonal-distance regressions,

as performed by Müller et al. (2024), future datasets will allow this.

## 5. Conclusions

We quantified bulk metallicities for transiting gas giants under multiple interior structures using CEPAM with a root-finding retrieval. Our key conclusions can be summarised as follows:

- We confirm that the inferred bulk metallicity for Fully-Mixed (FM) structures is larger than for Core+Envelope (CE) structures, due to the opacity effect that prevails on the density redistribution effect (Guillot 2005; Müller et al. 2020; Howard et al. 2025b). This, in turn, leads to Dilute Core (DC) structures that yield similar bulk metallicities than CE, independently of the extent of the composition gradient.
- When increasing the atmospheric metallicity to the dilute core structure by one (DCA) or three times stellar (DCA3), the inferred bulk metallicity gets closer to the one yielded by FM. This shows how atmospheric composition measurements are crucial to help constrain bulk composition.
- When considering non-adiabatic temperature gradients in DC structures, we find that the inferred bulk metallicity can be increased by up to 35% for a used value of  $R_p = 0.4$ . However, future studies should explore a larger parameter space and assess the long-term stability of composition gradients.
- Using a sample of 44 giant exoplanets, we confirm that heavy element mass increases with planetary mass. The slope of the  $M_Z$ – $M$  relationship does not change with the assumed structure, but only the intercept. This confirms that giant planets continue to accrete heavy elements as they grow.
- The planetary bulk metallicity decreases with planetary mass. Planets with  $M < 0.2 M_J$  have metallicities between 0.4 and 0.8. These low  $M$  – high  $Z$  planets are required to hold the negative correlation between  $Z$  and  $M$ . Planets with  $0.2 < M_p < 4 M_J$  have a strong scatter in metallicities, which can range from less than 0.01 to 0.5, showing how diverse the composition of giant exoplanets can be. For high planetary masses ( $M > 4 M_J$ ), although we have only 2 planets in our sample, the metallicities are still relatively high (between 0.1 and 0.3). This suggests that giant planets keep accreting heavy elements during runaway gas accretion or post-formation.

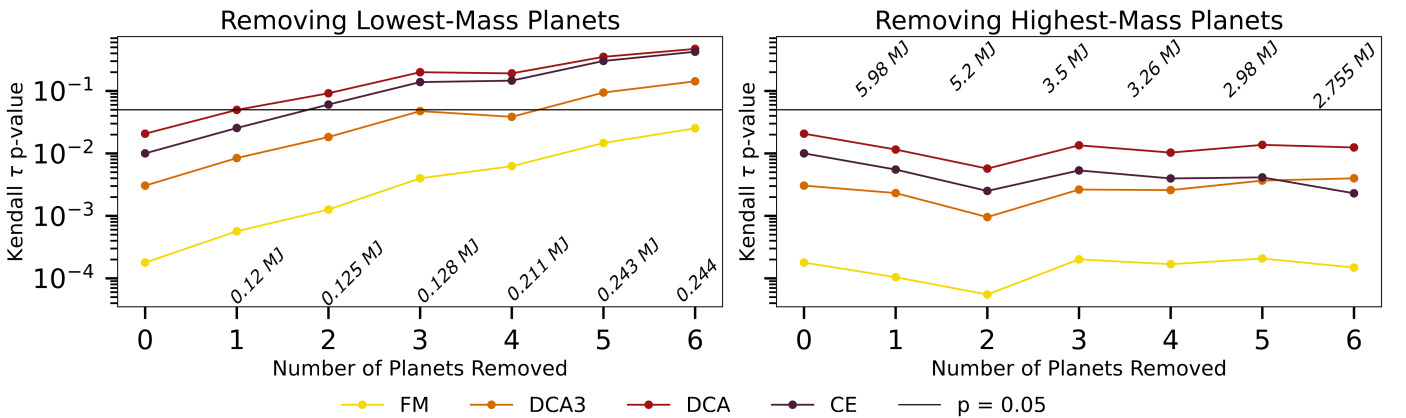


Fig. 7: Panels showing iterative removal of either smallest mass planets (left) or bigger mass ones (right), and how the correlation coefficient  $p$ -values changes as a consequence of that. Colour coded are the four different models. The horizontal line shows the  $p=0.05$  acceptance threshold. Planetary mass removed each time is noted on the x-axis.

Future studies that account for additional physical processes (such as convective mixing), and also observe atmospheric compositions will improve characterisation of planetary interiors. Forthcoming measurements from *PLATO* and *Ariel* (expected to be launched in 2026 and 2029, respectively) will sharpen bulk metallicity estimates by providing tighter constraints on mass, radius, age, and atmospheric composition. This will allow a better understanding of the interior of giant planets and of their formation pathways.

*Acknowledgements.* We thank Stefano Wirth, Simon Müller, Henrik Knierim, and Shay Zucker for the insightful discussions.

## References

- Asplund, M., Amarsi, A. M., & Grevesse, N. 2021, *A&A*, 653, A141
- Bloot, S., Miguel, Y., Bazot, M., & Howard, S. 2023, *Monthly Notices of the Royal Astronomical Society*, 523, 6282
- Buchhave, L. A., Bizzarro, M., Latham, D. W., et al. 2014, *Nature*, 509, 593
- Chabrier, G., Mazevet, S., & Soubiran, F. 2019, *ApJ*, 872, 51
- Chachan, Y., Fortney, J. J., Ohno, K., Thorngren, D., & Murray-Clay, R. 2025, arXiv e-prints [2509.20428], arXiv:2409.12345 [astro-ph.EP]
- Debras, F. & Chabrier, G. 2019, *The Astrophysical Journal*, 872, 100
- Debras, F., Chabrier, G., & Stevenson, D. J. 2021, *The Astrophysical Journal Letters*, 913, L21
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *Publications of the Astronomical Society of the Pacific*, 125, 306
- Fortney, J. J., Mordasini, C., Nettelmann, N., et al. 2013, *The Astrophysical Journal*, 775, 80
- Freedman, R. S., Lustig-Yaeger, J., Fortney, J. J., et al. 2014, *The Astrophysical Journal Supplement Series*, 214, 25
- Freedman, R. S., Marley, M. S., & Lodders, K. 2008, *The Astrophysical Journal Supplement Series*, 174, 504
- Guillot, T. 2005, *Annu. Rev. Earth Planet. Sci.*, 33, 493
- Guillot, T. & Morel, P. 1995, *A&AS*, 109, 109
- Guillot, T., Santos, N. C., Pont, F., et al. 2006, *A&A*, 453, L21
- Helled, R. & Bodenheimer, P. 2011, *Icarus*, 211, 939
- Helled, R. & Morbidelli, A. 2021, in *ExoFrontiers: Big questions in exoplanetary science* (IOP Publishing Bristol, UK), 12–1
- Helled, R. & Stevenson, D. 2017, *ApJ*, 840, L4
- Helled, R. & Stevenson, D. J. 2024, *AGU Advances*, 5, e2024AV001171
- Howard, S. & Guillot, T. 2023, *A&A*, 672, L1
- Howard, S., Guillot, T., Bazot, M., et al. 2023, *Astronomy & Astrophysics*, 672, A33
- Howard, S., Helled, R., Bergermann, A., & Redmer, R. 2025a, arXiv e-prints, arXiv:2507.06288
- Howard, S., Helled, R., & Müller, S. 2025b, *A&A*, 693, L7
- Howard, S., Müller, S., & Helled, R. 2024, *Astronomy & Astrophysics*, 689, A15
- Knierim, H. & Helled, R. 2025, *A&A*, 698, L1
- Leconte, J. & Chabrier, G. 2012, *Astronomy & Astrophysics*, 540, A20
- Lyon, S. & Johnson, J. 1992, Los Alamos National Laboratory
- Mankovich, C. R. & Fuller, J. 2021, *Nature Astronomy*, 5, 1103
- Miguel, Y. 2022, in *Bulletin of the American Astronomical Society*, Vol. 54, 202.01
- Miguel, Y., Bazot, M., Guillot, T., et al. 2022, *Astronomy & Astrophysics*, 662, A18
- Militzer, B., Hubbard, W. B., Wahl, S., et al. 2022, *The planetary science journal*, 3, 185
- Mordasini, C., van Boekel, R., Mollière, P., Henning, T., & Benneke, B. 2016, *The Astrophysical Journal*, 832, 41
- Müller, S., Baron, J., Helled, R., Bouchy, F., & Parc, L. 2024, *Astronomy & Astrophysics*, 686, A296
- Müller, S., Ben-Yami, M., & Helled, R. 2020, *ApJ*, 903, 147
- Müller, S. & Helled, R. 2023, *Astronomy & Astrophysics*, 669, A24
- Müller, S. & Helled, R. 2025, *Astronomy & Astrophysics*, 693, L4
- NASA Exoplanet Archive. 2025, NASA Exoplanet Archive, <https://exoplanetarchive.ipac.caltech.edu/>, accessed: 24 September 2025
- Otegi, J. F., Dorn, C., Helled, R., et al. 2020, *A&A*, 640, A135
- Parc, L., Bouchy, F., Venturini, J., Dorn, C., & Helled, R. 2024, *A&A*, 688, A59
- Parmentier, V., Guillot, T., Fortney, J. J., & Marley, M. S. 2015, *Astronomy & Astrophysics*, 574, A35
- Pollack, J. B., Hubickyj, O., Bodenheimer, P., et al. 1996, *icarus*, 124, 62
- Rauer, H., Catala, C., Aerts, C., et al. 2014, *Experimental Astronomy*, 38, 249
- Teske, J. K., Thorngren, D., Fortney, J. J., Hinkel, N., & Brewer, J. M. 2019, *The Astronomical Journal*, 158, 239
- Thorngren, D. P. & Fortney, J. J. 2018, *The Astronomical Journal*, 155, 214
- Thorngren, D. P., Fortney, J. J., Murray-Clay, R. A., & Lopez, E. D. 2016, *The Astrophysical Journal*, 831, 64
- Tinetti, G., Drossart, P., Eccleston, P., et al. 2018, *Experimental Astronomy*, 46, 135
- Turrini, D., Schisano, E., Fonte, S., et al. 2021, *The Astrophysical Journal*, 909, 40
- Valencia, D., Guillot, T., Parmentier, V., & Freedman, R. S. 2013, *The Astrophysical Journal*, 775, 10
- van Dijk, E. & Miguel, Y. 2025, *Monthly Notices of the Royal Astronomical Society*, 540, 1544
- Venturini, J. & Helled, R. 2020, *A&A*, 634, A31
- Wahl, S. M., Hubbard, W. B., Militzer, B., et al. 2017, *Geophysical Research Letters*, 44, 4649
- Ziv, M., Galanti, E., Howard, S., Guillot, T., & Kaspi, Y. 2024, *Astronomy & Astrophysics*, 692, A251

## Appendix A: Sample of Exoplanets and reported results

Table A.1: List of exoplanets used in the study. Planetary parameters are taken from Howard et al. (2025b) and are present in the publicly available planet database on [dace.unige.ch](https://dace.unige.ch). Planet names followed by an asterisk (\*) are the planets with  $\sigma_R < 3\%$ , while the whole dataset is confined by  $\sigma_R < 5\%$ .

| Planet        | M ( $M_J$ )       | R ( $R_J$ )       | Age (Myr)   | $Z_{st}$          | T (K) | $Z_{CE}$          | $Z_{FM}$          |
|---------------|-------------------|-------------------|-------------|-------------------|-------|-------------------|-------------------|
| TOI-4010d     | $0.12 \pm 0.01$   | $0.551 \pm 0.013$ | 3000–9200   | $0.036 \pm 0.07$  | 650   | $0.724 \pm 0.011$ | $0.672 \pm 0.026$ |
| HATS-72b      | $0.125 \pm 0.004$ | $0.722 \pm 0.003$ | 11710–12410 | $0.019 \pm 0.014$ | 739   | $0.4 \pm 0.008$   | $0.423 \pm 0.004$ |
| WASP-156b     | $0.128 \pm 0.01$  | $0.51 \pm 0.02$   | 2400–10400  | $0.027 \pm 0.12$  | 970   | $0.736 \pm 0.01$  | $0.767 \pm 0.045$ |
| HAT-P-12b*    | $0.211 \pm 0.057$ | $0.959 \pm 0.025$ | 500–4500    | $0.008 \pm 0.05$  | 963   | $0.129 \pm 0.05$  | $0.294 \pm 0.073$ |
| TOI-3629b*    | $0.243 \pm 0.082$ | $0.74 \pm 0.014$  | 5000–13800  | $0.055 \pm 0.093$ | 711   | $0.397 \pm 0.026$ | $0.427 \pm 0.025$ |
| HD332231b*    | $0.244 \pm 0.086$ | $0.867 \pm 0.027$ | 2400–6800   | $0.017 \pm 0.059$ | 876   | $0.222 \pm 0.048$ | $0.311 \pm 0.045$ |
| K2-329b       | $0.26 \pm 0.022$  | $0.774 \pm 0.026$ | 500–4000    | $0.019 \pm 0.068$ | 650   | $0.318 \pm 0.024$ | $0.45 \pm 0.044$  |
| WASP-69b      | $0.26 \pm 0.017$  | $1.057 \pm 0.047$ | 500–3000    | $0.021 \pm 0.077$ | 963   | $0.06 \pm 0.044$  | $0.199 \pm 0.057$ |
| HAT-P-19b*    | $0.277 \pm 0.06$  | $1.008 \pm 0.014$ | 3200–11200  | $0.023 \pm 0.07$  | 981   | $0.021 \pm 0.014$ | $0.112 \pm 0.025$ |
| TOI-4406b*    | $0.3 \pm 0.1$     | $1.0 \pm 0.02$    | 2200–3600   | $0.019 \pm 0.05$  | 904   | $0.044 \pm 0.025$ | $0.152 \pm 0.034$ |
| TOI-1268b     | $0.303 \pm 0.026$ | $0.812 \pm 0.054$ | 110–380     | $0.035 \pm 0.06$  | 919   | $0.311 \pm 0.029$ | $0.511 \pm 0.026$ |
| K2-287b*      | $0.315 \pm 0.086$ | $0.847 \pm 0.013$ | 3700–5700   | $0.024 \pm 0.04$  | 804   | $0.255 \pm 0.024$ | $0.328 \pm 0.02$  |
| HATS-49b*     | $0.353 \pm 0.092$ | $0.765 \pm 0.013$ | 8500–11900  | $0.025 \pm 0.053$ | 835   | $0.392 \pm 0.025$ | $0.416 \pm 0.018$ |
| WASP-132b     | $0.41 \pm 0.03$   | $0.87 \pm 0.03$   | 500–2500    | $0.026 \pm 0.13$  | 763   | $0.268 \pm 0.046$ | $0.376 \pm 0.044$ |
| TOI-5344b***  | $0.412 \pm 0.097$ | $0.946 \pm 0.021$ | 5100–13800  | $0.041 \pm 0.088$ | 689   | $0.067 \pm 0.035$ | $0.134 \pm 0.04$  |
| TOI-201b**    | $0.42 \pm 0.095$  | $1.008 \pm 0.014$ | 380–1330    | $0.027 \pm 0.036$ | 759   | $0.086 \pm 0.03$  | $0.209 \pm 0.037$ |
| HATS-75b*     | $0.491 \pm 0.079$ | $0.884 \pm 0.013$ | 10600–13800 | $0.051 \pm 0.4$   | 772   | $0.188 \pm 0.025$ | $0.288 \pm 0.019$ |
| HAT-P-17b**   | $0.534 \pm 0.034$ | $1.01 \pm 0.029$  | 4500–11100  | $0.015 \pm 0.08$  | 792   | $0.04 \pm 0.029$  | $0.108 \pm 0.039$ |
| WASP-84b*     | $0.687 \pm 0.048$ | $0.976 \pm 0.025$ | 500–3700    | $0.019 \pm 0.12$  | 833   | $0.142 \pm 0.047$ | $0.222 \pm 0.064$ |
| TOI-3714b*    | $0.689 \pm 0.044$ | $0.944 \pm 0.02$  | 6000–13800  | $0.038 \pm 0.086$ | 764   | $0.124 \pm 0.035$ | $0.155 \pm 0.051$ |
| HAT-P-54b*    | $0.76 \pm 0.042$  | $0.944 \pm 0.028$ | 1800–8200   | $0.011 \pm 0.08$  | 818   | $0.168 \pm 0.05$  | $0.217 \pm 0.054$ |
| K2-290c       | $0.774 \pm 0.047$ | $1.006 \pm 0.05$  | 3200–5600   | $0.013 \pm 0.1$   | 676   | $0.093 \pm 0.06$  | $0.159 \pm 0.063$ |
| TOI-1478b     | $0.851 \pm 0.052$ | $1.06 \pm 0.04$   | 5200–12200  | $0.018 \pm 0.069$ | 918   | $0.048 \pm 0.034$ | $0.111 \pm 0.051$ |
| K2-140b       | $0.93 \pm 0.04$   | $1.21 \pm 0.09$   | 5300–13200  | $0.019 \pm 0.1$   | 962   | $0.058 \pm 0.045$ | $0.112 \pm 0.06$  |
| TOI-1811b*    | $0.972 \pm 0.079$ | $0.994 \pm 0.025$ | 1900–10800  | $0.031 \pm 77.0$  | 962   | $0.112 \pm 0.044$ | $0.194 \pm 0.049$ |
| WASP-130b     | $1.23 \pm 0.04$   | $0.89 \pm 0.03$   | 400–7900    | $0.028 \pm 0.1$   | 833   | $0.291 \pm 0.039$ | $0.362 \pm 0.052$ |
| TOI-2010b*    | $1.286 \pm 0.044$ | $1.054 \pm 0.027$ | 600–4200    | $0.023 \pm 0.055$ | 400   | $0.096 \pm 0.064$ | $0.144 \pm 0.067$ |
| TOI-5542b     | $1.32 \pm 0.1$    | $1.009 \pm 0.036$ | 7200–12900  | $0.009 \pm 0.08$  | 441   | $0.087 \pm 0.068$ | $0.135 \pm 0.08$  |
| HATS-17b      | $1.338 \pm 0.065$ | $0.777 \pm 0.056$ | 800–3400    | $0.031 \pm 0.03$  | 814   | $0.318 \pm 0.025$ | $0.482 \pm 0.047$ |
| HATS-74Ab*    | $1.46 \pm 0.096$  | $1.032 \pm 0.021$ | 5900–13800  | $0.05 \pm 0.027$  | 895   | $0.063 \pm 0.031$ | $0.172 \pm 0.034$ |
| Kepler-117c   | $1.84 \pm 0.18$   | $1.101 \pm 0.035$ | 3900–6700   | $0.014 \pm 0.1$   | 704   | $0.035 \pm 0.028$ | $0.082 \pm 0.04$  |
| TIC237913194b | $1.942 \pm 0.092$ | $1.117 \pm 0.051$ | 4000–7400   | $0.021 \pm 0.05$  | 974   | $0.056 \pm 0.045$ | $0.105 \pm 0.051$ |
| HAT-P-15b     | $1.946 \pm 0.066$ | $1.072 \pm 0.043$ | 5200–9300   | $0.026 \pm 0.08$  | 904   | $0.066 \pm 0.05$  | $0.115 \pm 0.053$ |
| TOI-4515b     | $2.005 \pm 0.052$ | $1.086 \pm 0.039$ | 1000–1400   | $0.017 \pm 0.03$  | 705   | $0.067 \pm 0.057$ | $0.153 \pm 0.058$ |
| K2-114b       | $2.1 \pm 0.12$    | $0.932 \pm 0.031$ | 2700–11700  | $0.04 \pm 0.037$  | 701   | $0.256 \pm 0.052$ | $0.285 \pm 0.045$ |
| TOI-4127b     | $2.3 \pm 0.11$    | $1.096 \pm 0.036$ | 2700–6900   | $0.021 \pm 0.12$  | 605   | $0.044 \pm 0.036$ | $0.091 \pm 0.043$ |
| HATS-76b*     | $2.629 \pm 0.034$ | $1.079 \pm 0.031$ | 600–13300   | $0.032 \pm 0.057$ | 940   | $0.066 \pm 0.044$ | $0.116 \pm 0.053$ |
| TOI-2180b*    | $2.755 \pm 0.03$  | $1.01 \pm 0.021$  | 6800–9600   | $0.028 \pm 0.057$ | 348   | $0.132 \pm 0.036$ | $0.163 \pm 0.043$ |
| NGTS-20b      | $2.98 \pm 0.16$   | $1.07 \pm 0.04$   | 1400–6800   | $0.022 \pm 0.08$  | 688   | $0.079 \pm 0.059$ | $0.137 \pm 0.103$ |
| TOI-5153b*    | $3.26 \pm 0.05$   | $1.06 \pm 0.04$   | 4400–6400   | $0.02 \pm 0.08$   | 906   | $0.103 \pm 0.063$ | $0.155 \pm 0.044$ |
| TOI-2589b*    | $3.5 \pm 0.029$   | $1.08 \pm 0.03$   | 9000–13000  | $0.02 \pm 0.04$   | 582   | $0.047 \pm 0.035$ | $0.069 \pm 0.038$ |
| WASP-162b     | $5.2 \pm 0.2$     | $1.0 \pm 0.05$    | 10620–13800 | $0.029 \pm 0.13$  | 910   | $0.186 \pm 0.094$ | $0.187 \pm 0.048$ |
| TOI-2338b     | $5.98 \pm 0.2$    | $1.0 \pm 0.02$    | 5000–9000   | $0.026 \pm 0.04$  | 799   | $0.149 \pm 0.055$ | $0.182 \pm 0.014$ |

Table A.2: [List of planetary parameters]List of exoplanets used in the study. Planetary parameters are taken from Howard et al. (2025b) and are present in the publicly available planet database on `dace.unige.ch`. Planet names without the asterisk (\*) are missing data from DC and CEA models, as for them statistical analysis has not been done, and the inspection has been made only for 21 planets with  $\sigma_R < 3\%$ .

| Planet Name   | $Z_{DC0.25}$      | $Z_{DC0.5}$       | $Z_{DC0.75}$      | $Z_{DCA}$         | $Z_{DCA3}$        | $Z_{CEA}$         |
|---------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| TOI-4010d     |                   |                   |                   | $0.631 \pm 0.002$ | $0.692 \pm 0.012$ |                   |
| HATS-72b      |                   |                   |                   | $0.380 \pm 0.006$ | $0.405 \pm 0.005$ |                   |
| WASP-156b     |                   |                   |                   | $0.720 \pm 0.004$ | $0.720 \pm 0.050$ |                   |
| HAT-P-12b*    | $0.119 \pm 0.048$ | $0.124 \pm 0.047$ | $0.124 \pm 0.047$ | $0.123 \pm 0.047$ | $0.163 \pm 0.049$ | $0.129 \pm 0.050$ |
| TOI-3629b*    |                   | $0.389 \pm 0.028$ | $0.372 \pm 0.027$ | $0.427 \pm 0.029$ | $0.436 \pm 0.027$ | $0.453 \pm 0.026$ |
| HD332231b*    | $0.195 \pm 0.035$ | $0.214 \pm 0.045$ | $0.209 \pm 0.043$ | $0.210 \pm 0.046$ | $0.267 \pm 0.045$ | $0.222 \pm 0.049$ |
| K2-329b       |                   |                   |                   | $0.371 \pm 0.047$ | $0.375 \pm 0.026$ |                   |
| WASP-69b      |                   |                   |                   | $0.085 \pm 0.046$ | $0.131 \pm 0.046$ |                   |
| HAT-P-19b*    | $0.019 \pm 0.014$ | $0.021 \pm 0.014$ | $0.021 \pm 0.015$ | $0.046 \pm 0.017$ | $0.094 \pm 0.017$ | $0.048 \pm 0.017$ |
| TOI-4406b*    | $0.038 \pm 0.025$ | $0.043 \pm 0.025$ | $0.043 \pm 0.024$ | $0.049 \pm 0.021$ | $0.111 \pm 0.026$ | $0.051 \pm 0.021$ |
| TOI-1268b     |                   |                   |                   | $0.428 \pm 0.044$ | $0.421 \pm 0.031$ |                   |
| K2-287b*      | $0.229 \pm 0.015$ | $0.246 \pm 0.024$ | $0.240 \pm 0.023$ | $0.270 \pm 0.023$ | $0.303 \pm 0.023$ | $0.289 \pm 0.024$ |
| HATS-49b*     |                   | $0.385 \pm 0.027$ | $0.367 \pm 0.024$ | $0.406 \pm 0.032$ | $0.421 \pm 0.024$ | $0.419 \pm 0.025$ |
| WASP-132b     |                   |                   |                   | $0.300 \pm 0.053$ | $0.332 \pm 0.046$ |                   |
| TOI-5344b*    | $0.064 \pm 0.034$ | $0.065 \pm 0.033$ | $0.064 \pm 0.034$ | $0.105 \pm 0.033$ | $0.156 \pm 0.023$ | $0.108 \pm 0.035$ |
| TOI-201b*     | $0.053 \pm 0.031$ | $0.084 \pm 0.029$ | $0.082 \pm 0.029$ | $0.124 \pm 0.029$ | $0.169 \pm 0.029$ | $0.130 \pm 0.031$ |
| HATS-75b*     | $0.184 \pm 0.026$ | $0.182 \pm 0.025$ | $0.179 \pm 0.024$ | $0.220 \pm 0.023$ | $0.241 \pm 0.021$ | $0.233 \pm 0.026$ |
| HAT-P-17b*    | $0.041 \pm 0.028$ | $0.038 \pm 0.028$ | $0.040 \pm 0.028$ | $0.049 \pm 0.026$ | $0.083 \pm 0.027$ | $0.050 \pm 0.026$ |
| WASP-84b*     | $0.124 \pm 0.051$ | $0.137 \pm 0.046$ | $0.135 \pm 0.044$ | $0.135 \pm 0.045$ | $0.194 \pm 0.044$ | $0.140 \pm 0.048$ |
| TOI-3714b*    | $0.121 \pm 0.038$ | $0.121 \pm 0.035$ | $0.120 \pm 0.035$ | $0.153 \pm 0.033$ | $0.181 \pm 0.030$ | $0.161 \pm 0.036$ |
| HAT-P-54b*    | $0.154 \pm 0.048$ | $0.162 \pm 0.049$ | $0.158 \pm 0.048$ | $0.161 \pm 0.049$ | $0.194 \pm 0.048$ | $0.166 \pm 0.051$ |
| K2-290c       |                   |                   |                   | $0.095 \pm 0.056$ | $0.126 \pm 0.058$ |                   |
| TOI-1478b     |                   |                   |                   | $0.057 \pm 0.034$ | $0.096 \pm 0.032$ |                   |
| K2-140b       |                   |                   |                   | $0.066 \pm 0.036$ | $0.105 \pm 0.040$ |                   |
| TOI-1811b*    | $0.099 \pm 0.046$ | $0.110 \pm 0.042$ | $0.108 \pm 0.041$ | $0.141 \pm 0.042$ | $0.172 \pm 0.038$ | $0.145 \pm 0.044$ |
| WASP-130b     |                   |                   |                   | $0.323 \pm 0.051$ | $0.349 \pm 0.061$ |                   |
| TOI-2010b*    | $0.037 \pm 0.028$ | $0.040 \pm 0.031$ | $0.062 \pm 0.038$ | $0.093 \pm 0.065$ | $0.113 \pm 0.031$ | $0.074 \pm 0.031$ |
| TOI-5542b     |                   |                   |                   | $0.099 \pm 0.088$ | $0.184 \pm 0.160$ |                   |
| HATS-17b      |                   |                   |                   | $0.428 \pm 0.042$ | $0.404 \pm 0.031$ |                   |
| HATS-74Ab*    | $0.056 \pm 0.030$ | $0.064 \pm 0.031$ | $0.063 \pm 0.030$ | $0.103 \pm 0.041$ | $0.169 \pm 0.014$ | $0.106 \pm 0.031$ |
| Kepler-117c   |                   |                   |                   | $0.082 \pm 0.018$ | $0.076 \pm 0.026$ |                   |
| TIC237913194b |                   |                   |                   | $0.072 \pm 0.039$ | $0.109 \pm 0.036$ |                   |
| HAT-P-15b     |                   |                   |                   | $0.086 \pm 0.042$ | $0.125 \pm 0.036$ |                   |
| TOI-4515b     |                   |                   |                   | $0.101 \pm 0.055$ | $0.123 \pm 0.042$ |                   |
| K2-114b       |                   |                   |                   | $0.255 \pm 0.045$ | $0.275 \pm 0.044$ |                   |
| TOI-4127b     |                   |                   |                   | $0.064 \pm 0.031$ | $0.121 \pm 0.030$ |                   |
| HATS-76b*     | $0.046 \pm 0.030$ | $0.028 \pm 0.014$ | $0.060 \pm 0.038$ | $0.107 \pm 0.051$ | $0.151 \pm 0.042$ | $0.097 \pm 0.046$ |
| TOI-2180b*    | $0.096 \pm 0.029$ | $0.105 \pm 0.030$ | $0.098 \pm 0.030$ | $0.106 \pm 0.032$ | $0.142 \pm 0.028$ | $0.142 \pm 0.062$ |
| NGTS-20b      |                   |                   |                   | $0.094 \pm 0.044$ | $0.124 \pm 0.039$ |                   |
| TOI-5153b*    | $0.067 \pm 0.040$ | $0.080 \pm 0.045$ | $0.079 \pm 0.044$ | $0.123 \pm 0.057$ | $0.150 \pm 0.054$ | $0.118 \pm 0.058$ |
| TOI-2589b*    | $0.034 \pm 0.023$ | $0.038 \pm 0.026$ | $0.038 \pm 0.026$ | $0.054 \pm 0.025$ | $0.090 \pm 0.032$ | $0.065 \pm 0.033$ |
| WASP-162b     |                   |                   |                   | $0.181 \pm 0.082$ | $0.198 \pm 0.068$ |                   |
| TOI-2338b     |                   |                   |                   | $0.165 \pm 0.031$ | $0.165 \pm 0.035$ |                   |

Table A.3: List of retrieved adiabatic (DCA) and superadiabatic runs for dilute core structures. Two parametrised  $R_\rho$  values were tested: 0.1 and 0.4. Results for 11 exoplanets are listed below.

| Planet    | $R_\rho = 0$ (DCA) | $R_\rho = 0.1$      | $R_\rho = 0.4$      |
|-----------|--------------------|---------------------|---------------------|
| HAT-P-12b | $0.123 \pm 0.047$  | $0.1232 \pm 0.043$  | $0.189 \pm 0.047$   |
| TOI-3629b | $0.427 \pm 0.029$  | $0.4309 \pm 0.0284$ | $0.4752 \pm 0.0249$ |
| HD332231b | $0.210 \pm 0.046$  | $0.2222 \pm 0.047$  | $0.3127 \pm 0.0651$ |
| HAT-P-19b | $0.046 \pm 0.017$  | $0.0492 \pm 0.022$  | $0.055 \pm 0.022$   |
| K2-287b   | $0.270 \pm 0.023$  | $0.2845 \pm 0.0237$ | $0.3306 \pm 0.0252$ |
| HATS-49b  | $0.406 \pm 0.032$  | $0.4266 \pm 0.0444$ | $0.4342 \pm 0.0194$ |
| HATS-75b  | $0.220 \pm 0.023$  | $0.2201 \pm 0.0253$ | $0.2735 \pm 0.0287$ |
| HAT-P-17b | $0.049 \pm 0.026$  | $0.050 \pm 0.025$   | $0.063 \pm 0.036$   |
| WASP-84b  | $0.135 \pm 0.045$  | $0.1448 \pm 0.0464$ | $0.155 \pm 0.045$   |
| HAT-P-54b | $0.161 \pm 0.049$  | $0.1708 \pm 0.0502$ | $0.2054 \pm 0.0529$ |
| TOI-1811b | $0.141 \pm 0.042$  | $0.142 \pm 0.042$   | $0.177 \pm 0.044$   |

## Appendix B: Result comparison with Howard et al. (2025)

In order to assess the strength and accuracy of the retrieved  $Z$  for the sample used, we took advantage of the fact that the used dataset is the same as the one used by Howard et al. (2025b). We thus took the initial sample of 21 exoplanets with the most precise data on Radius ( $R_{err} < 3\%$ ), and compared it with the same planets processed with the method explained in Howard et al. (2025b). Results are shown in Fig. B.2, where a visual comparison between the trends for the  $Z/Z_\star$  vs  $M$  is being shown. To note how the best fit lines vary from the ones retrieved with the full dataset of 44 exoplanets, showing again the variability and unreliability of such curves for model prediction. Howard also discussed in their paper the shift of the slope  $\alpha$  between the dataset considering only  $R_{err} < 3\%$ , or for the broader range considered of  $R_{err} < 5\%$ . In the bottom panels of Fig. B.2, it is shown how we assessed the difference between the data from this paper compared to the ones from Howard et al. (2025b). While a simple subtraction could provide information about the difference between the two mean values, we also wanted to investigate whether each ranges of values (so accounting for data uncertainty) lie within each other. To do so, a normalisation using the standard deviations of the data has been computed. The normalised difference

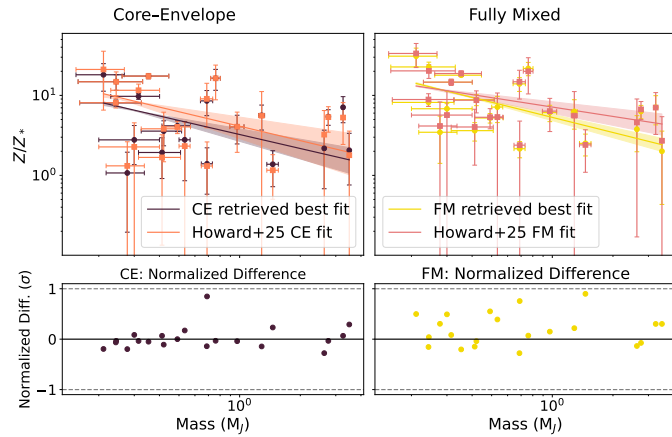


Fig. B.1

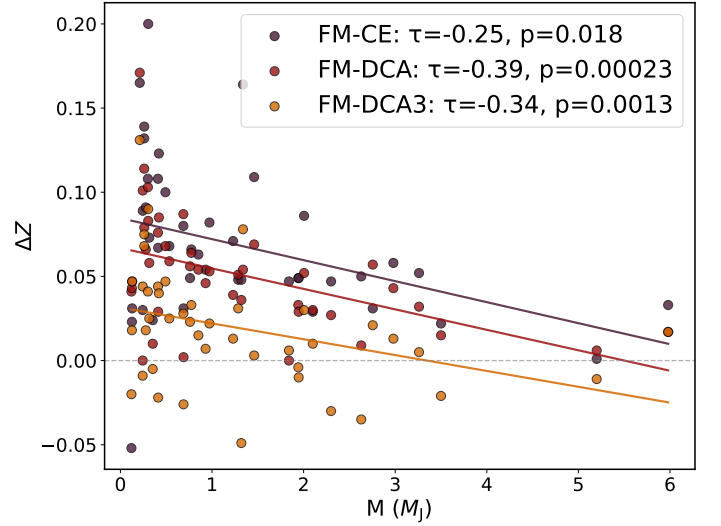


Fig. B.2

can be expressed with the following formula:

$$\sigma = \frac{\text{Howard} - \text{CE/FM}}{\sqrt{\sigma_{\text{Howard}}^2 + \sigma_{\text{CE/FM}}^2}}. \quad (\text{B.1})$$

Results regarding this comparison were considered positive. While there is a slight change in the predictive trends between the two structures CE and FM (slight shift in  $\beta$  for CE, and in  $\alpha$  for FM compared to the ones found in Howard's). The results for the normalised difference ultimately confirmed the similarities between the two retrieved results, and their respective methods; All values lie between a standard deviation of 1 and -1 (or  $|\sigma| < 1$ ), showing statistical consistency for all the 21 cases compared. Furthermore, the normalised difference does not show any systematic error related to the mass (since there is no trend in the normalised difference  $\sigma$ ). We thus conclude by saying that there is no discouraging difference between the two results (as the difference between the two approaches can be explained to be within the data error range and thus not due to randomness). We then accepted the validity of our approach explained in Section 2.3, and proceeded with the inspection of the additional structures.

## Appendix C: Model assumption-driven changes

To isolate the impact of structural assumptions on the inferred metallicity, we quantified model-dependent offsets in  $Z$  across the four models (FM, CE, DCA, DCA3) by forming differences relative to FM, namely  $\Delta Z_{\text{FM-CE}} \equiv Z_{\text{FM}} - Z_{\text{CE}}$ ,  $\Delta Z_{\text{FM-DCA}} \equiv Z_{\text{FM}} - Z_{\text{DCA}}$ , and  $\Delta Z_{\text{FM-DCA3}} \equiv Z_{\text{FM}} - Z_{\text{DCA3}}$ .

Figure B.2 aims at showing such changes in the assumed structure over the full mass regime of the inspected exoplanets. The three lines show the best-fit lines of the data that is computed by the subtraction between the two mean points according to the models compared. No  $\sigma$  in the models has been accounted for, as only the mean points have been inspected. For this reason, no error protraction has been shown on the best-fit lines, that have been kept only for visual reasons. Nonetheless, without counting model errors, the  $\Delta Z$  for all three models being compared to FM show statistical validity as shown by the Kendall's  $\tau$  results (and respective p-values) shown on the Figure's legend. We can

see that, on average, all the models tend to reach 0 (No  $Z$  difference between two structurally different models) in the high- $M$  regime. While the best-fit line for  $\Delta Z_{FM-CE}$  approaches zero, for the difference between the FM-DCA and FM-DCA3 goes to the negative (meaning that fully mixed is not anymore the structural endmember, but DCA and DCA3 overpredicts  $Z$  more than FM). This is caused by the forcing of respectively stellar and three times stellar abundances in their structures. Low- $Z$  scenarios (usually in the high- $M$  region) will be 'ruled out' by forcing an enriched metallicity in the atmosphere. This could suggest that low- $Z$  are unlikely to be superenriched in heavy elements in their atmospheres. This agrees with the theory exposed in Fortney et al. (2013). Further studies must be made to investigate if that is the case, or if degeneracies between mass distribution in massive planets might overshadow the overall magnitude of opacity effect proved to be dominant by this study.

#### Appendix D: Weighted least-squares (WLS) $\chi^2$ test

To assess the reliability of the power-law fits, we performed a weighted least-squares (WLS)  $\chi^2$  test, which accounts for individual data uncertainties rather than relying solely on  $R^2$ . In WLS, each residual is scaled by its measurement error:

$$\chi_{\text{obs}}^2 = \sum_{i=1}^N \left( \frac{y_i - f(x_i; \theta)}{\sigma_i} \right)^2, \quad (\text{D.1})$$

where  $y_i$  are the data,  $f(x_i; \theta)$  the model, and  $\sigma_i$  the  $1\sigma$  uncertainties. Under the null hypothesis that the model describes the data and errors are Gaussian,  $\chi_{\text{obs}}^2$  follows a chi-square distribution with  $\nu = N - p$  degrees of freedom.

For  $N = 43$  and  $p = 2$ ,  $\nu = 41$ , giving  $\chi_{(0.05; 41)}^2 \approx 56.9$ . Fits are accepted at the 5% level if  $\chi_{\text{obs}}^2 \leq \chi_{(0.05; \nu)}^2$ . In all cases,  $\chi_{\text{obs}}^2$  exceeds this threshold, indicating that the fits do not reproduce the data within the quoted errors. For the  $Z$ – $M$  relation, the observed values are:

$$\chi_{\text{obs}}^2 = 580.4 \text{ (CE)}, 465.4 \text{ (DCA)}, 311.7 \text{ (DCA3)}, 243.7 \text{ (FM)},$$

corresponding to reduced  $\chi_{\text{red}}^2$  values of 14.2, 11.4, 7.6, and 5.9, respectively. Since  $\chi_{\text{red}}^2 \gg 1$ , we reject the null hypothesis and conclude that the observed scatter exceeds that expected from measurement errors alone. The diversity of planetary metallicities likely reflects intrinsic astrophysical variability rather than observational noise.