Universität
Zürich^{UZH}

Geographisches Institut

Interiors of Gas Giant Exoplanets: A Statistical Population Study

ESS 511 Master's Thesis

Author: Lorenzo Peerani, 23-739-592

Supervised by: Dr. Ravit Helled (ravit.helled@uzh.ch), Dr. Saburo Howard (saburo.howard@uzh.ch), Dr. Maarten Eppinga

Faculty representative: Prof. Dr. Maria J. Santos

02.08.2025

Interiors of Gas Giant Exoplanets: A Statistical Population Study

University of Zurich^UZH

Master's Thesis in Earth System Science - ESS511

Author:
Lorenzo Peerani
23-739-592

Supervised by:
Dr. Ravit Helled (ravit.helled@uzh.ch)
Dr. Saburo Howard (saburo.howard@uzh.ch)
(Department of Astrophysics)
Dr. Maarten Eppinga
(Department of Geography)

Faculty Representative:
Dr. Maria Santos
(Department of Geography)

August 2025

Abstract

We investigate how the assumed internal structures shape the inferred heavy-element fraction (Z) inside of gas giants exoplanets. Interiors provide critical clues about how planets formed and evolved. They also encode the internal dynamics that govern planetary behaviour across the universe. Using the evolutionary model CEPAM, coupled to a root–finding retrieval of Z, we compare compact core–envelope (CE), dilute–core (DC/DCA/DCA3), and fully mixed (FM) structural scenarios. Statistical robustness is assessed with Bayesian correlation/covariance analyses and weighted least squares. We find that structural choice redistributes metals and changes the total Z needed to match observed radii. Differences in Z due to structural variations were recorded at a greater magnitude in the low-mass regime. The opacity set by atmospheric composition counter-plays and exceeds density redistribution in inflating radii. This is shown by the overprediction of Z from DCA (and even greater in FM) when compared to CE. Different assumed thermal structures have been inspected too. Results from superadiabatic models stabilised by a compositional gradient shows greater heat retention. Such effect is dependent on the extent of the dilution m_{dilute}. Greater internal temperatures inflate the radius and shift Z estimates to higher degrees. Correlations in M_z/M, Z/M, and $(Z/Z_\star)/M$ have been found statistically significant, while their predictive power is weak, cautioning against simple power-law extrapolations. The found relationships are evidence of possible major core accretion formation dynamics. Future work should enlarge the sample, especially at low masses, couple interiors to measured atmospheric abundances (rather than a fixed Z_{atm}), and allow m_{dilute} to emerge from physics and time-dependent mixing. Looking ahead, future space missions and observational data can constrain atmospheric inventories to restrict the impact of current parameter degeneracies.

“We peer down into the core of our planet, and the blazing interior of our star…
We uncover hidden chapters in the record of our origins.”

— Carl Sagan, *Pale Blue Dot*, 1994

Contents

1 Introduction **7**
- 1.1 Planet formation . . . 7
- 1.2 Evolutionary trends . . . 8
- 1.3 Jupiter and Saturn . . . 9
- 1.4 R-M Relationship . . . 10
- 1.5 Project aims . . . 11

2 Theory **12**
- 2.1 How to build a planetary model . . . 12
  - 2.1.1 Fundamental Equations . . . 12
  - 2.1.2 Equations of State . . . 12
- 2.2 Energy balance and opacity . . . 13
- 2.3 Heat transport . . . 14
- 2.4 Model parameters and assumptions . . . 14

3 Methodology **16**
- 3.1 Sample data . . . 16
- 3.2 Structural Parameters . . . 16
- 3.3 Super adiabatic heat transport . . . 19
- 3.4 Root-find retrieval . . . 19
- 3.5 Data analysis . . . 21

4 Results **22**
- 4.1 Retrieval comparison . . . 22
- 4.2 Bulk metallicity . . . 23
  - 4.2.1 CE-FM . . . 23
  - 4.2.2 DC . . . 25
  - 4.2.3 Models with Z_{atm} . . . 25
- 4.3 Thermal transport . . . 28
- 4.4 Relationships and trends . . . 29
  - 4.4.1 M_z/M Relationship . . . 29
  - 4.4.2 Z/M Relationship . . . 30
  - 4.4.3 $Z_\star$ normalisation . . . 32
- 4.5 Statistics . . . 33
  - 4.5.1 Correlation validity . . . 33
  - 4.5.2 Best-fit analysis: Covariance . . . 36
  - 4.5.3 Best-fit analysis: WLS . . . 37
  - 4.5.4 Break-up points . . . 38

5 Discussion **40**
- 5.1 Planetary trade-offs . . . 40
  - 5.1.1 Density vs. Opacity in controlling Z estimation . . . 40
  - 5.1.2 Density vs. Super Adiabaticity in controlling Z estimation . . . 41
- 5.2 Planetary relations . . . 42
  - 5.2.1 Formation-driven structure . . . 43
- 5.3 Atmospheric metallicity . . . 43
- 5.4 Where is the 'fuzziness' starting? . . . 43

6 Conclusion **45**

7 Acknowledgements **46**

8 Appendix **51**
8.1 Planetary data and Z outputs 52
8.2 Grid Testing and Validation 55
8.3 Convergence evaluation 58
8.4 Single variable testing 59
8.5 MCMC corner plots posterior evaluation 64
8.6 CEA 66

List of Figures

1 Core accretion phases (from Helled and Morbidelli (2021)) . . . 7
2 CE and FM curves for different heavy element fractions Z and EoS(adapted from Howard et al. (2025)) . . . 9
3 M-R Relationship, source: Müller et al. (2024) . . . 10
4 Main internal structures tested . . . 17
5 Metallicity curves at each mass coordinates for the different models . . . 18
6 Effect of different masses to the Z retrieval . . . 20
7 FM and CE data comparison with Howard et al., 2025 . . . 22
8 Model comparison between FM-CE for HD-332231b . . . 23
9 FM-CE-DC distribution comparison for 21 exoplanets . . . 24
10 Model-resolved R-M relationship between DC-CE-FM models . . . 26
11 Model comparison including DCA and DCA3 . . . 27
12 Thermal profile of K2-287b with different f values tested . . . 28
13 M_z/M_{Pl} relationship . . . 30
14 ΔZ between models . . . 31
15 Z/M_{pl} relationship . . . 31
16 Z/M_{st} versus M_{pl} relationship . . . 32
17 Bayesian retrieval Kendall's tau corr. coeff. and p-value distributions for Z/M . . . 34
18 Iterative correlation change with planet removal . . . 35
19 ρ distribution for model covariance for all models . . . 37
A.1 M test grid validation . . . 56
A.2 Age test grid validation . . . 57
A.3 M test grid validation . . . 57
A.4 Variable testing of Z_{dilute} . . . 59
A.5 Variable testing of m_{dilute} . . . 60
A.6 Variable testing of δ_m . . . 60
A.7 Variable testing of Z_{atm} . . . 61
A.8 Structure testing on Z_{atm} and Z_{dilute} for a fixed Z=0.5 . . . 61
A.9 Structure testing on single cooling track . . . 62
A.10 Dilution testing with a core on single cooling tracks . . . 62
A.11 Corner plots for structural models CE and FM across parameters: M_z–M, Z–M, and Z/Z_*–M. . . . 64
A.12 Corner plots for structural models DCA and DCA3 across parameters: M_z–M, Z–M, and Z/Z_*–M. . . . 65
A.13 Model comparison between CE-CEA-DCA . . . 66

List of Tables

1 Kendall's τ and p-values across different models and relations . . . 33
A.1 List of planetary parameters . . . 52
A.2 List of retrieved Z (CE-FM-DC0.25-DC0.5-DC0.75). . . . 53
A.3 List of retrieved Z (DCA-DCA3-CEA-SMA). . . . 54
A.4 Magnitude of different assumed f factors. . . . 55

Glossary

List of symbols and abbreviations

Symbol	Description
M	Planetary mass
M_z	Heavy-element mass in the planet
Z	Heavy-element fraction, M_z/M
$Z_\odot$	Solar metallicity
$M_\oplus$	Earth-mass unit
M_J	Jupiter-mass unit
$Z_\star$	Stellar metallicity ([Fe/H])
Z_atm	Atmospheric metallicity
Z_env	Envelope metallicity (outer layer)
m_dilute	Mass coordinate of steepest composition gradient
δ_m	Steepness of composition gradient
Z_dilute	Central Z in dilute-core models
α	Power-law slope parameter
β	Power-law intercept parameter
CE	Core+Envelope structure
FM	Fully mixed (homogeneous) structure
DC	Dilute core structure
DCA	Dilute core with $Z_\mathrm{atm} = Z_\mathrm{env} = Z_\star$
DCA3	Dilute core with $Z_\mathrm{atm} = 3\,Z_\star$
CEA	Core+metal-rich envelope ($Z_\mathrm{env} = Z_\star$)
SMA	Dilute core with solar atmospheric metallicity ($Z_\mathrm{atm} = Z_\odot$)
f	Scaling factor for double-diffusive convection contribution

1 Introduction

Research in planetary science focuses on characterizing diverse planetary environments to unravel their formation histories. Differentiation, internal structure, and chemical composition serve as each planet's fingerprint, offering insights into the formation dynamics and heat-transport mechanisms that govern planetary interiors and evolution. The internal distribution of heavy elements (i.e. all species heavier than hydrogen and helium) critically shapes a giant planet's cooling rate, affecting its estimated radius and radial contraction. By examining internal structures, we can infer the physical conditions of planets, diving into possible protoplanetary disks dynamics that shaped such internal structures. Moreover, solving internal structures of planets serves as a playground for probing pressure and temperature regimes beyond the reach of current laboratory experiments and instrumentation.

Gas giants play a pivotal role in formation theories. Their immense masses and radii generate extreme temperature and pressure conditions, directly reflecting the early accretion processes that determine planetary size and composition. Understanding the interiors of gas giants addresses fundamental questions: Why do planets reach their observed sizes, and how does composition influence their growth in the formation phase, and their evolution?

1.1 Planet formation

The prevailing theory is core accretion. Micron-sized dust grains coagulate into kilometer-scale planetesimals (or cm–m "pebbles"), which collide and merge to form solid cores of order 5–10 $M_\oplus$. Once a core exceeds this threshold, envelope cooling becomes efficient and the planet enters a runaway gas-accretion phase, rapidly drawing H/He from the nebula. In some cases, massive cores continue to accrete heavy elements, increasing their metal fraction (Pollack et al., 1996; Dominik and Tielens, 1997; Helled and Morbidelli, 2021).

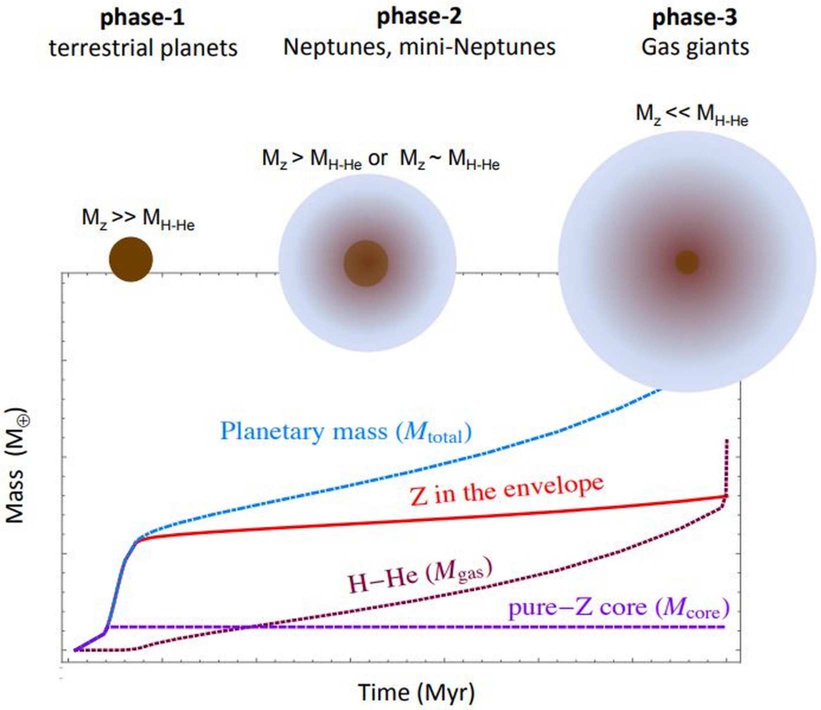

Figure 1: Schematic evolution of total mass (M_{total}), envelope heavy element fraction (Z), H–He mass (M_{gas}), and core mass (M_{core}) changes during the different phases of core accretion (adapted from Helled and Morbidelli (2021)).

Protoplanets that never trigger runaway gas accretion are believed then to remain smaller in size, explaining rocky planets/ gas giants compositional disparity related to their planetary size.

Figure 1 schematically illustrates the three commonly recognized stages in core accretion, where planetary accretion can stop in either one of the three illustrated phases.

At the start, the embryo's mass is dominated by heavy elements (Z), shown as phase 1. As hydrogen and helium accretion proceeds (phase 2), the protoplanet becomes massive enough to also attract gravitationally heavy elements, incorporating them into the envelope, the outermost layer of a gas giant planet (phase 3 . While the amount of heavy elements increases in the envelope, the core mass is believed to be invariant. Thus, the internal structure is shaped by the undergone stages of planetary accretion (in addition on affecting its mass), and these conditions are what interior modelers aim to unveil. The final planet's final characteristics after planetary accretion and formation, are believed to depend on key factors such as orbital distance (which sets the available solid inventory) and formation timing (which controls how much gas remains in the disk when runaway begins). Not all gas giants, however, undergo substantial post-runaway metal enrichment, and the settings determining accretion would result in the resulting planetary radius and mass. Planetary cases could start accreting late, resulting in not undergoing phase 3, stopping at phase 2 if not at phase 1.
An alternative formation pathway mechanism theorised is disk instability; if a massive protoplanetary disk becomes gravitationally unstable and cools rapidly, it may fragment into self-gravitating gas clumps that collapse directly into a gas-dominated objects, bypassing the need for a prior solid core as an initial requirement for planets to form (Cuzzi et al., 2001; Johansen et al., 2007). Although subsequent mixing and differentiation can blur the initial structural differences, comparing population-level predictions from both mechanisms can reveal distinct signatures between the two theories, and putting some light on caveats surrounding both formation mechanisms (Morbidelli and Raymond, 2016).
Since gas giants are the most massive kind of planets (due to their early formation, and thus much more time to accrete and enter phase 3), their relevance for formation studies is central.

1.2 Evolutionary trends

Previous studies have inspected trends to better understand planetary interiors and formation, using evolution models, in order to infer planetary internal structures. By generating evolutionary tracks, which describe how a planet cools and contracts over time, they could estimate the heavy-element content required to match the observed radius and age. With either adding or removing the modeled fraction of heavy elements inside of a planet, the predicted radius by the model can actively get lower or increase, and thus fit the predicted radius with the radius value given from the observation of the inspected planet (using mass and temperature measurements of said planet). Such fitting of planetary cooling to the observed radius and age can be degenerate, as many combinations of different variables like chemical composition, and the internal distribution of chemicals, can shift drastically the result. Efforts done by Thorngren et al. (2016), Müller and Helled (2023) and Howard et al. (2025) aimed at quantifying such heavy element masses in planetary interiors through assumptions on compositions and structure of the planetary interior. Thorngren et al. (2016) looked at population analysis and found a relationship between planetary mass and the amount of heavy elements stored within. Such relation shows an increase in heavy element mass M_z with higher planetary masses M. Similarly, Müller and Helled (2023) aimed at caracterising hot giant exoplanets, and the uncertainties related to evolution model studies. Such uncertanties are also being showed in the work by Howard et al. (2025), where inspected variable changes have been showed from an evolutionary model perspective. Their study shows how different structural assumptions and equations of state (that will be later explained in Section 2) affect the resulting radii from evolution runs. Some of the results found by Howard et al. can be seen in 2.
The figure shows the resulting cooling tracks from the evolutionary model, and how these change according to different structures. A decrease in the expected radius can be seen with an increase in

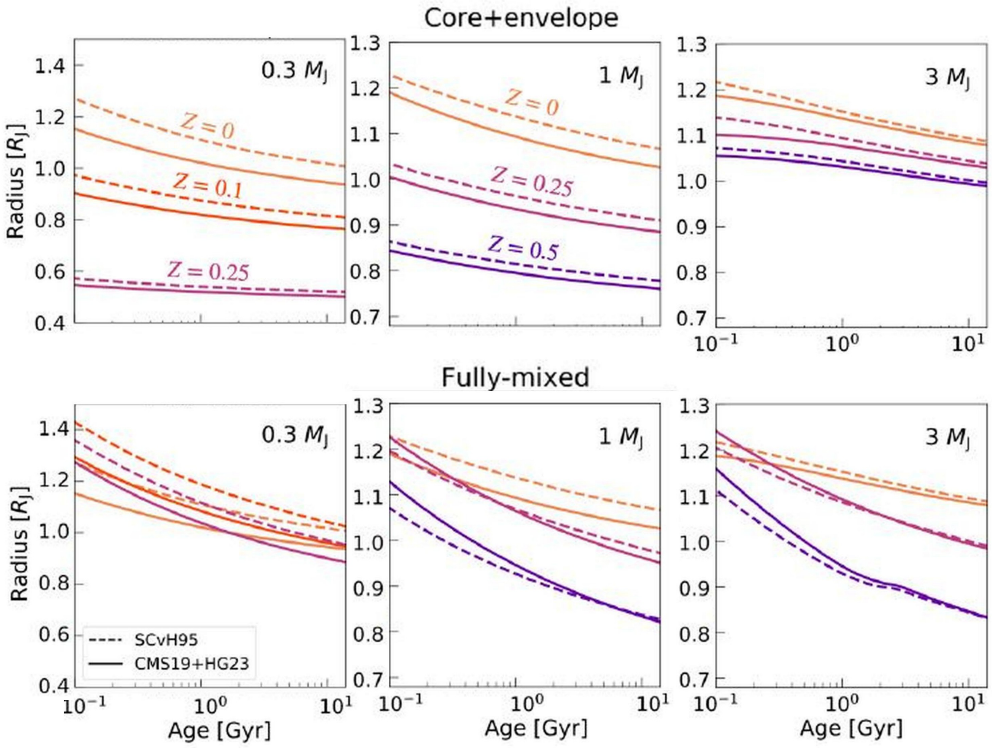

Figure 2: Comparison of FM and CE models for different planetary masses and Z. Comparison with two different EoS (SCvH95 and CMS19+HG23) has been investigated (adapted from Howard et al. (2025)).

a heavy element fraction Z. Using heavy elements as a planetary chemical building block increases the density due to higher mean molecular weight compared to hydrogen and helium. This enhances a greater gravitational contraction of the planet, resulting in a lower radius estimated by the models. Other than composition, structure plays an important role, and it is central to what we want to inspect for explaining formation theories. The tested structures used by Howard et al. (2025) for planetary interiors were two scenarios that have been named Core+Envelope (CE) and Fully Mixed (FM). The first one appears to be as a well defined core made of heavy elements with a hard boundary going into the envelope, the outer layer (with the atmosphere being the outermost fraction of it). A Fully Mixed structure on the other hand is a homogeneous distribution of heavy elements throughout the planet. A change in Z changes to respect of the mass of the planet, (for different planetary masses assumed, respectively 0.3, 1, and 3 Jupiter masses), and for the structure being used. Where in CE the shift is quasi linear (an increase in Z predictably enhances the contraction of the radius), more degeneracies appear for the FM case. In the last study made by Howard (at the time when this was written), the same relationship previously found by Thorngren was further confirmed, even with the change of the structural assumption.

1.3 Jupiter and Saturn

While exoplanetary studies can open new doors to interior modeling, more insightful and precise estimates are provided by studying the gas giants in our Solar system, Jupiter and Saturn. In this case, better observational constraints on the atmospheric chemical composition were possible thanks to the JUNO and Cassini missions carried by NASA for the investigation of the two planets. While JUNO probed the outermost atmospheric composition of the planet, linking such data with gravity data and ring seismology (for Saturn), interior modeling efforts were able to give exhaustive results on the internal structure of the two gas giants. Interior modelers were then able to link the density changes from gravity data analyses to assess the internal composition and structure

(Wahl et al., 2017; Howard et al., 2023). Results indicate that neither Jupiter nor Saturn is best described by a sharp, well-defined core enveloped by a uniform H/He layer, neither by a fully homogeneous composition. Instead, their deep interiors are consistent with a Dilute Core (or “fuzzy core”), in which heavy elements dominate the very center but gradually taper off in concentration with increasing radius, producing a smooth compositional gradient rather than an abrupt boundary. From this knowledge, a more accurate representation of an interior structures could be carried by the Dilute Core structure, where from a metal-rick interior dilutes out into an atmosphere depleted in heavy elements.

Based on these discoveries, efforts in testing such a structure has been made by interior modelers for better representing planetary interiors compared to the initial standard models of Core+Envelope and Fully Mixed. Works like the one carried by Bloot et al. (2023) did test the Dilute structure retrieval from water atmospheric abundances on exoplanets. More complex structures like the introduction of a defined core followed by a dilution gradient was recently tested by van Dijk and Miguel (2025) for four cases of exoplanets. While a rise in use of this structure has been seen in the recent years, there is currently no large statistical population analysis that has been carried for an exhaustive range of planets (similarly to what has been done by Thorngren et al. 2016) with the structural assumption of a composition gradient.

1.4 R-M Relationship

The mutual information between the state of the art of the gas giants, and with the help of observational data, exoplanetary research can provide a statistical framework that would not be possible through just the analysis of Jupiter and Saturn. With a current dataset of nearly 6,000 confirmed exoplanets, approximately 1,000 of which qualify as gas giants according to the NASA Exoplanet Archive. and incorporate observations from key missions and catalogs, including JWST, TESS, Kepler, HATNet, WASP, and the Henry Draper catalog, data that has been used in this study for compiling and retrieving results through planetary evolution models.

Since planetary formation theories aim at explaining the formation of all planets (from rocky, terrestrial planets, to gas giants), efforts in planet characterization over the full mass range has been done on the discovered planets. For a statistical population analysis, efforts like Müller et al. (2020); Müller and Helled (2023); Müller et al. (2024) aimed at seeking hidden exoplanetary relationships, such as the Radius-Mass interplay. Such inspection, over a sample size of 688 exoplanets of known observed radii and masses, can be seen in Figure 3 developed by Müller et al. (2024).

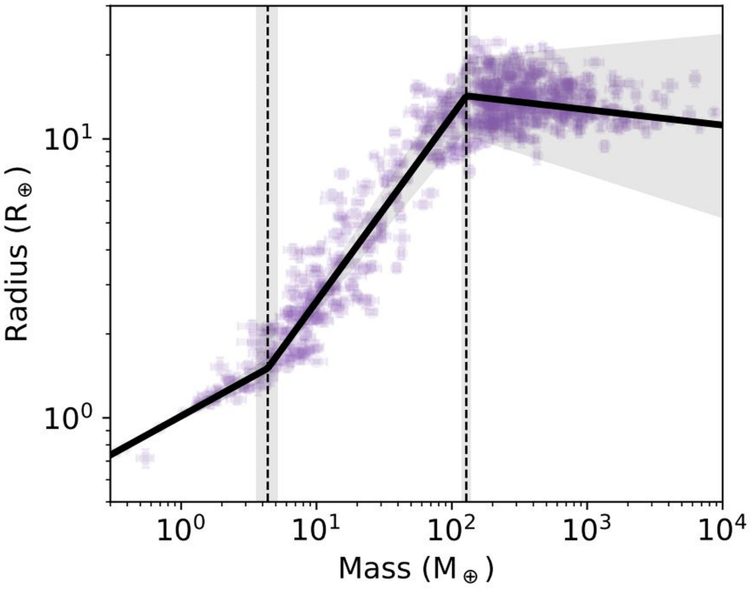

Figure 3: Planetary Mass-Radius relationship with break-up point analysis showing three main distinguishable trends.

The plot shows the distribution of different planetary classes, from rock-dominated planets to an intermediate class symbolising gaseous accretion. At greater masses (127+/-7 earth masses, or $\approx$ 0.4 jupiter masses) the transition towards a greater uncertainty region, show another change in the R-M trend caused by their massive self-gravity, that causes them to contract more over time. Moreover, as planetary mass M increases, the internal pressure rises. At extremely-high pressure environments, hydrogen transitions to a metallic state (because it decopules itself from molecular H_2 into ionised H^- with free electrons). This comformation of free electrons allows them to be closely 'squeezed' together. Such layer, called metallic hydrogen layer, is believed to be present in Jupiter and Saturn too, but at higher masses the radial extent of it will be larger. As this electron degeneracy becomes dominant, the planet will not increase in radii with additional mass (as shown by a gentle negatively-sloped line in Figure 3). Our project, aiming at inspecting in the gas giant mass range realm, will encompasses the fast increase in radii in the middle mass range, and the most massive planetary cases in the degenerate hydrogen domain. Such trend will be considered when discussing our results.

1.5 Project aims

The aim of this study is to investigate exoplanet interiors and quantify heavy element abundances in their held structures. This will be assessed by applying a set of different structural internal compositions to investigate how relations can vary with the change in the assumed structure. The current state of the art on Jupiter and Saturn will be applied to generate the most realistic interpretations. Although Dilute Core models have been mainly explored and applied for Jupiter (e.g. see Howard et al. 2023), they remain largely untested in other gas giants, and the consequences of such compositional gradients have yet to be rigorously quantified on a full population. Evolutionary models will be run by testing a Dilute Core structure (compositional gradient of heavy elements 'diluting out' towards the outer layers). The dilute core structure will be compared with that of the CE and FM end-member models, and with that of the literature. This comprehensive sample enables us to probe the relationship between formation pathways and interior structure across a diverse population of gas giants.
The project's research questions are the following:

- How does the assumed internal structure change the estimated heavy element abundance in planetary interiors?
- How can these estimates inform us on possible planetary formation mechanisms? Is there any relationship with the planetary mass?
- How is the inferred internal structure changing heat transport in the planet?

2 Theory

2.1 How to build a planetary model

2.1.1 Fundamental Equations

Before diving into the scope of this study, it is essential explaining how to study interior models, and how to build a planetary model in general. Planetary models are composed of fundamental equations, equations of state (EoS), and chemical/structural assumptions of the planet. Fundamental equations are partial differential equations that physically describe mass, energy and evolution of a planet. These set of equations are the following:

1. Hydrostatic equilibrium equation:

$$\frac{\partial P}{\partial M} = -\frac{GM}{4\pi R^4} \tag{1}$$

   Its main describers are the pressure gradient where it relates the change of pressure (P) with mass (M), where G is the gravitational constant, R is the radial distance from the centre , assuming perfect sphericity and no rotation in density redistribution or thermal asymmetries. Rotation and magnetic field can be neglected, as in most cases do not play an important role (Guillot, 2005)

2. Energy transport:

$$\frac{\partial T}{\partial M} = \left(\frac{\partial P}{\partial M}\right)\frac{T}{P}\nabla \tag{2}$$

   Makes T, P and M dependent quantities, and is referred to the thermal structure inside the planet.

3. Mass conservation:

$$\frac{\partial R}{\partial M} = \frac{1}{4\pi R^2 \rho} \tag{3}$$

   The radius gradient (mass continuity equation) subsequently describes how the radius changes with the mass as a factor of density (ρ), ensuring proper mass distribution across the planet.

4. Conservation of energy:

$$\frac{\partial L}{\partial M} = \varepsilon - T\frac{\partial S}{\partial t} \tag{4}$$

   The last equation showed is the luminosity gradient, essential for this study in developing an evolutionary model (as this gives information about temperature changes over time). Its descriptors are luminosity L, ε is an additional energy generation rate per unit mass input (set as 0 in this study), and S the specific entropy (in this case it's change over time).

2.1.2 Equations of State

Equations of state (EoS) are needed to solve these sets of equations and link them with their chemical properties. EoS are equations or tables that describe the relation between the thermodynamical quantities. Equations of state are calculated in high-pressure laboratories, with instrumentation like the diamond-anvil-cell and shock-compression experiments, that directly probe material's properties to pressures up to a few hundred gigapascals. Beyond those limits, ab-initio methods like density functional theory (DFT) and quantum Monte Carlo, extend the EOS into the multi-megabar regime by simulating atomic interactions, ensuring P–T relations deep inside giant-planet interiors where experiments cannot yet reach. Pressure and temperature are shaped by density, and will reflect the composition's chemical and physical properties. This is highly dependent on the composition assumed. Generally, in interior modelling, chemical composition can be divided into light and heavy elements. Light elements comprise of:

- Hydrogen, helium, and their interaction: the main component of gas giants in the majority of cases. This comprises of the phisical properties of hydrogen, helium, and their interaction at different P-T conditions. It also describes electron degeneracies due to metallic hydrogen.

Heavy elements, all the elements and molecules with mean molecular weight greater than hydrogen and helium, are subdivided in two broad assumed cathegories:

- Ices: This term includes information about water (not necessarily in solid state, depending on temperature T and pressure P), as well as volatile materials such as ammonia, methane, and carbon dyoxide.
- Rocks: These are refractory materials such as silicates and metals (iron, quartz -SiO_2- and magnesium silicates -$MgSiO_3$-). Together with the ices, they constitute the heavy elements modelled in the planet Helled and Howard (2024).

When a mixture of these compositions has been added together, the linear mixing rule has been applied. In other words, the properties of each component (entropy, mean molecular weight, density...) are simply being summed and weighted by their mass fraction at each mass coordinate in the planet's interior McMahon et al. (2012).

2.2 Energy balance and opacity

The energy balance of a planet can always be simplified by the amount of energy entering the system (throught stellar irradiation) and the energy lost from the planet's interior.

$$L = L_\odot + L_{\text{int}} \tag{5}$$

where L is the planet's total emergent luminosity, arising from the sum of two contributions. The term $L_\odot$ denotes the equilibrium luminosity (the power reradiated after absorbing stellar flux) while L_{int} represents the intrinsic cooling luminosity generated by the planet's internal heat. Both L and $L_\odot$ follow the Stefan–Boltzmann law under black-body approximation,

$$L = 4\pi\,\sigma\,R^2\,T_{\text{eff}}^4, \tag{6}$$

$$L_\odot = 4\pi\,\sigma\,R^2\,T_{\text{eq}}^4. \tag{7}$$

Energy in and energy out of the system are also controlled by the opacity effect, that through the action of absorption of photons in the atomic/molecular/grain medium, converts radiant energy into internal energy. Scattering of photons slows down the transport of energy through redirecting to a different vectorial direction. Both phenomena affect the energy in from the stellar irradiation, and the energy out from internal heat dissipation. The total (or extinction) opacity, can then be simply explained by the addition of both phenomena.

$$\kappa_{\text{ext}} = \kappa_{\text{abs}} + \kappa_{\text{sca}} \tag{8}$$

In order to compute so for the whole radius coordinates moving in the planet's interior, the integration of the changing opacity must be done. The formula is the following:

$$\tau(r) \;=\; \int_r^\infty \kappa_{\text{ext}}(r')\,\rho(r')\,\mathrm{d}r' \tag{9}$$

where $\tau(r)$ is the optical depth at radius r . This radial integral definition is essential for connecting the opacity κ_{ext} and density structure $\rho(r)$ to the planet's radiative transfer: it tells you how "thick" the material is to photons as you move outward in radius. Opacity data is given through opacity tables that again inform the physical properties of each material at different P-T conditions. A Rosseland mean radiative opacity has been computed for the whole thermal band (Freedman et al.,

2008; Valencia et al., 2013).
While such opacity calculations is being done for the envelope, in the atmosphere is treated as a fully radiative layer cutting off at 10 bar (common atmosphere-envelope cut-off transition) using a non-gray atmosphere that accounts for the visible (light coming in) band, and the thermal (internal) one. For more information please see (Parmentier et al., 2015). The method found from Valencia et al. (2013) is the one being used for the opacity enhancement in the outermost layer.

2.3 Heat transport

Heat transport is controlled by the efficiency in dissipating energy outwards. Radiative heat transfer has been mentioned for the atmosphere, however at greater pressures convection dominates heat transport. It is all dependant on which one is the most efficient in carrying more heat outwards at each layer. Whether ratiative or convective transfer is the preferential 'carrier' is defined by the following relationship given by the Schwarzschild criterion:

$$\nabla_T = \begin{cases} \nabla_{\mathrm{rad}}, & \text{if } \nabla_{\mathrm{rad}} < \nabla_{\mathrm{ad}}, \\ \nabla_{\mathrm{ad}}, & \text{if } \nabla_{\mathrm{rad}} > \nabla_{\mathrm{ad}}, \end{cases} \tag{10}$$

where $\nabla_T \equiv \frac{d\ln T}{d\ln P}$, and it is the local temperature gradient at each mass location. ∇_{rad} denotes the radiative temperature gradient, and the ∇_{ad} being the adiabatic/convective gradient. Adiabatic behaviour is defined as a parcel of gas (or material) that either sinks or rises (buoyancy) without any heat exchange with the surroundings. The convective regime follows the adiabatic behaviour since convection quickly adjusts the local temperature gradient to the adiabatic one. Adiabatic internal structures are widely used as internal heat transport approximations. Under the Schwarzschild criterion, a layer is convectively unstable if $\nabla_{\mathrm{rad}} > \nabla_{\mathrm{ad}}$ under the assumption of constant composition. This will not be the case in internal structures such as the one aimed to be tested in this study, where a composition gradient can play a stabilizing role under such statement. The Ledoux criterion, that accounts for a stabilising composition gradient, states that for a mean molecular weight contribution $\nabla_\mu = \dfrac{d\ln\mu}{d\ln P}$, the resulting relationship will read as follows:

$$\nabla_{\mathrm{rad}} > \nabla_{\mathrm{ad}} + \frac{\phi}{\delta}\nabla_\mu \tag{11}$$

where

$$\delta = -\Big(\frac{\partial\ln\rho}{\partial\ln T}\Big)_{P,\mu}, \qquad \phi = \Big(\frac{\partial\ln\rho}{\partial\ln\mu}\Big)_{P,T}.$$

It can happen that the two criterions are discordant one to the other. If the Schwarzschild test predicts convection instability (meaning convection-driven), but Ledoux predicts stability (inhibition of convection) it is believed a semi-convective regime, such as double diffusive convection (small overturning cells that still carry heat to the above layers), takes place in this dichotomy between the two criteria (Debras and Chabrier, 2019). In such case, deviations from an adiabatic interior will be shown.

2.4 Model parameters and assumptions

Mass, radius, equilibrium temperature (T_{eq}), and age jointly constrain a planet's bulk metallicity (Z). The mass M fixes gravity and pressures via the structure equations; the planetary radius R responds to the mixture of low-density H/He and high-density heavy elements set by the EoS. T_{eq} supplies the outer boundary condition (setting atmospheric entropy and thus the interior adiabat), while the age selects the point along the cooling track (planets shrinking in size over time). Matching the observed (Age, R) (evolution model output) at the known T_{eq} and mass (input variables) with evolution models (solving the fundamental formulas with the use of appropriate EoS) yields Z; omitting any one observable leaves additional degeneracies in the inferred heavy-element content.

These four variables will then be needed for the exoplanets inspected.
In running evolution models, multiple assumptions have to be made.

- Assuming an initial planetary radius at the zeroth time step is needed. The initial radius chosen needs to be higher than the current observed radius because of the higher entropy during gravitational collapse in early-stage planetary formation.
- No rotation nor magnetic field have been accounted
- No mixing has been considered. While the model is dynamic, the structure remains static and constant for consistent interpretations.
- Unlike solid planets, viscosity can be neglected in gas giants
- An initial entropy has to be assumed. Based on the current knowledge on planet formation, the model recreates a high entropy starting points (also called 'hot start') of $S_{\mathrm{init}} \sim 11\,k_{\mathrm{B}}$ per baryon Guillot (2005).

All these assumptions are believed to be valid for the mass range of exoplanets that have been inspected. For mixing dynamics, further inspections not covered by this study might be needed for assessing the stability of a structure over time.

3 Methodology

3.1 Sample data

A first set of 21 exoplanet has been taken for this investigation. The initial mass range inspected is from 0.211 to 3.5 Jupiter masses. The observational parameters have been taken from Howard et al. (2025). The filtering parameters used for the selection were: Planets with the required observational parameters to feed in the model. As previously mentioned, for each planet, its mass, radius, equilibrium temperature, and age range (taken from the star, and assumed to be similar to the planet's since planetary formation is relatively quick compared to their lifecycle) data must be present. The chosen equilibrium temperature range is between 350 and 1000 K. This temperature range is well above the equilibrium temperature of Jupiter and Saturn (respectively 122 and 95 K) because of the grater amount of gas giant exoplanets discovered closer to their host stars. Such convenience for 'spotting' gas giants transiting the star inherently creates a sampling bias in the population (Fressin et al., 2013). Proximity to the star also drives poorly understood "inflation" mechanisms (e.g., tidal dissipation, ohmic heating, enhanced atmospheric circulation) that can deposit significant extra heat into the planet's envelope, inflating its radius beyond what pure H/He structure models predict. Because these additional heating and circulation processes remain poorly understood, and lie outside the scope of our current project, we exclude all planets with $T_{\text{eq}} > 1000\,\text{K}$ from our sample. We considered planets with accurate measurements for mass and radius to focus on planets that we can accurately infer their internal structure. We considered planets with a relative error in the radius below 3%, and below 10% for mass estimates. A further loosening of this filtering criterion has been made to allow more planet data for the statistical analysis. This led to the use of 43 planet data with a radius error smaller than 5% instead of the initial 21. The mass range has then also beeen increased at a later stage to 0.12 Mj for the lowest planetary mass, and 5.98 Mj for the biggest. This was done for increasing the sample size during statistical analysis.

3.2 Structural Parameters

CEPAM *(Code d'Évolution Modulaire et Adaptatif)* has been used to create planetary evolution tracks (Guillot and Morel, 1995). Different internal structures will be tested, and the main types can be seen in Figure 4. For each structure, the heavy element-driven structure and its extent will be adjusted accordingly to how much heavy element mass/ fraction will be inserted. To mention that such heavy element mass will be referred as M_Z and the fraction of heavy elements in planetary interiors, also known as the bulk metallicity Z, is the resulting $Z = \frac{M_Z}{M_{\text{pl}}}$, calculated by dividing the heavy element mass by the total planetary mass.
The tested planetary interiors can be subdivided in:

- **Core+Envelope structure (CE):** in this structure, the planetary interior can be subdivided in two main layers: the core (the innermost part of the planet), and the envelope (outermost layer also comprising of the atmosphere above 10 bars). Consequently, the metallicity in those layers will be called Z_{core} and Z_{env}. In all cases, the metallicity in the envelope will be the same as the one in the atmosphere, so $Z_{\text{env}} = Z_{\text{atm}}$. Z_{core} will always be equal to 1 (completely made of heavy elements), so to vary its extent, the mass of the core will be changed. So, in a CE case with no heavy elements in the envelope, Z can be calculated as $\frac{M_{\text{core}}}{M_{\text{pl}}}$, where M_{pl} is the total planetary mass. The CE structure can be seen in Figure 4, as well in the Z distribution in Figure 5.

- **Fully Mixed structure (FM):** contrarily to the CE model, FM does not have a core, and the deep interior Z_{deep} is homogeneous throughout and it is equal to Z_{atm}. Calculations of Z is then simply $Z = Z_{\text{deep}} = Z_{\text{atm}}$.This structure can be seen at the second place from the right in Figure 4, and shown as a constant Z, and visualised by the horizontal line in the composition distribution in Figure 5 (in yellow). The rest of the planetary composition will be filled with a H-He mixture.

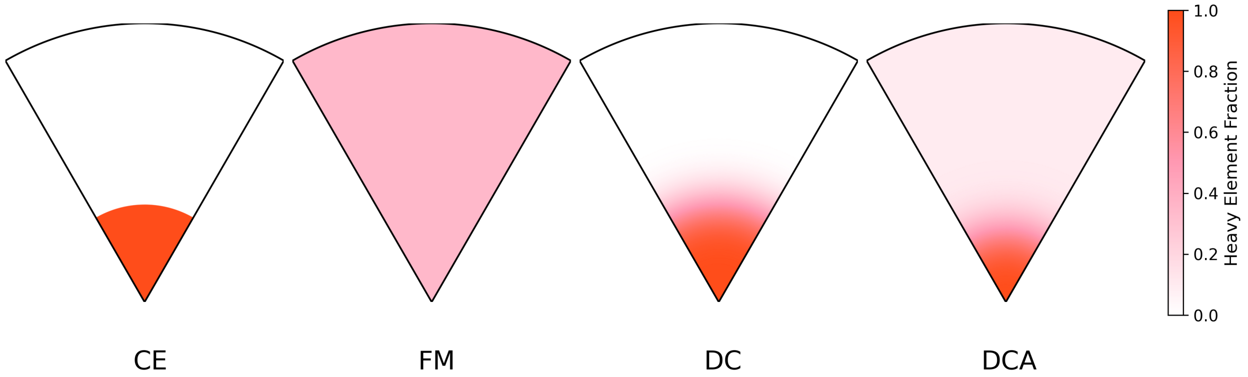

Figure 4: Main planetary interior structures tested in this study: Core–Envelope (CE): $Z_{\mathrm{core}} \neq Z_{\mathrm{env}}$, where $Z_{\mathrm{env}} = 0$; Fully Mixed (FM): $Z_{\mathrm{core}} = Z_{\mathrm{env}}$; Dilute Core (DC): $Z_{\mathrm{core}} \to 0$ outward, $Z_{\mathrm{env}} = 0$; Dilute Core + Atmosphere (DCA): $Z_{\mathrm{core}} \to Z_{\mathrm{env}}$, where $Z_{\mathrm{env}} > 0$ and $Z_{\mathrm{env}} < Z_{\mathrm{core}}$.

- **Dilute Core structure (DC):** this will be the main structure to be inspected. To describe a composition gradient diluting towards the outside of the planet, an error function curve has been implemented in CEPAM to compute the different Z at each mass coordinate. A smooth step sigmoidal transition is what it has been chosen to represent such dilution in Z, and it is being represented in Figure 5. The following equation describes such curve:

  $$Z(m) = Z_{\mathrm{atm}} + \frac{Z_{\mathrm{dilute}} - Z_{\mathrm{atm}}}{2}\left[1 - \mathrm{erf}\left(\frac{m - m_{\mathrm{dilute}}}{\delta m}\right)\right] \tag{12}$$

  where Z_{atm} the outer (atmospheric) heavy–element mass fraction, Z_{dilute} the deep Z at the centre of the planet, erf is the Gauss error function to smooth the dilution curve. δm sets the steepness of the dilution curve, and m_{dilute} is the mass location where the function is at its steepest, or in other words the middle point of the dilution at the fraction coordinates $Z = \frac{1}{2}(Z_{\mathrm{dilute}} + Z_{\mathrm{atm}})$. These parameter can be seen in Figure 5 in the black dashed line. Z_{dilute} has been chosen to be the variable that will be brought up or down to adjust Z. So, for each model tested, the other parameters will be fixed. While δm has been chosen to be fixed at 0.125 as done in previous literature such as Bloot et al. 2023), the effects of changing m_{dilute} and Z_{atm} will be inspected. To start, DC models have been ran for no atmospheric metallicity ($Z_{\mathrm{atm}} = 0$) and with varying m_{dilute} to only study this extent of this variable when isolated. The m_{dilute} values tested are 0.25, 0.5, and 0.75. The resulting models will be referred in this study as DC0.25, DC0.5, and DC0.75. To calculate the total metallicity of such structures the integral of the function must be done. A simplified calculation of $Z_{\mathrm{dilute}}/m_{\mathrm{dilute}}$ can be made since $Z_{\mathrm{atm}} = 0$ and δm is symmetrical.

- **Dilute Core structure with Stellar Atmospheric metallicity (DCA):** After testing for different m_{dilute} scenarios, such value will be fixed at m/M = 0.5 (as shown in Figure 5). The question of which Z_{atm} has been chosen is though a different hypothesis. First of all, we tested a Dilute Core structure with the atmosphere matching the stellar metallicity $\mathrm{Z}_{\star}$, of which we have observational data. $\mathrm{Z}_{\star}$ information of the host star for each exoplanet can be found in Table 5 in the appendix. assuming Z_{atm}=$\mathrm{Z}_{\star}$ then a quick Z calculation can be done with the following formula:$(Z_{\mathrm{dilute}}+Z_{\mathrm{atm}})/m_{\mathrm{dilute}}$.
  In order to get the star's metallicity $Z_{\star}$, some additional steps need to be done. We first take the solar metal mass fraction and assume iron ratios are a good proxy of heavy element abundance in stars ($Fe/H \approx Z/H$). While this is not really the case, it is a well adopted method for approximation (also used in previous literature such as in Thorngren et al. (2016), Teske et al. (2019), and Howard et al. (2025)). To find $Z_{\star}$, the following formula must be applied:

  $$Z_{*} \approx Z_{\odot}\, 10^{[\mathrm{Fe/H}]}, \tag{12}$$

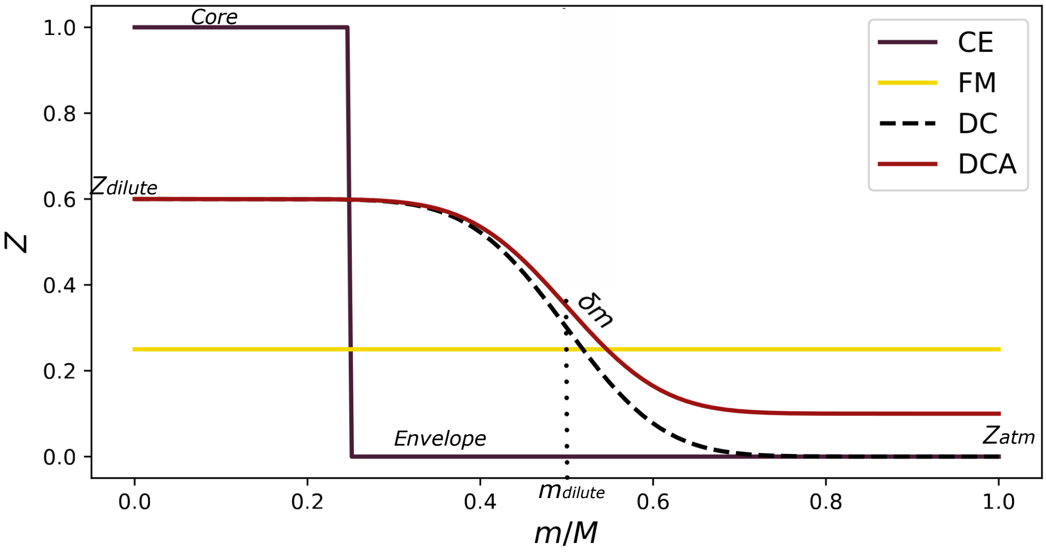

Figure 5: Metallicity curves at each mass coordinates for the different models: CE, FM, DC, and DCA.

where $Z_\odot$ is the Sun's metallicity.

- **Dilute Core structure with 3 times the Stellar Atmospheric metallicity (DCA3):** A tested variation of the Z_{atm} composition has been made with our current knowledge that we have on the atmospheric compositions of Jupiter (both from measurements and models, see Helled and Howard,2024). Jupiter results in having 3 times the metallic abundance of the Sun in its atmosphere. For such reason, a DCA3 structure will be tested with Z_{atm} = 3*Z$_{st}$ to see how this may vary outcomes of Z and M_z.
- **Dilute Core structure with Atmospheric metallicity = Solar (SMA):** what if instead of using a varying metallicity (according to the planet's host star) a fixed one is being used? SMA structure will be inspect the change in Z for a fixed Z_{atm}=Z_{Sun}=0.0142. Such value has been taken from Asplund et al. (2009).
- **Core+Envelope with Stellar metallicity in the envelope (CEA):** does the structure really matter, or by adding Z in the outermost layer we get similar results to DCA? Such hypothesis will be tested with sais structure.

We cannot possibly know the chemical composition of heavy elements inside of planetary interiors. Thus, a 1:1 ratio of rocks and ices has been assumed for this study. Recent modelling studies (van Dijk and Miguel, 2025) tried more complex structures like a core only composed of rocks, and the envelope made of ices, to mimic a possible differentiation of material made by the difference in mean molecular weight through planetary mixing. For the sake of keeping interpretations clear on the extent of the dilute curve variables, composition has been kept constant throughout.
As shown in Howard et al. (2025), the type of EoS used does have an impact in the retrieval. Since the importance of EoS has been inspected in the above-mentioned paper, here we are using CMS19 from Chabrier and Debras (2021) for hydrogen and helium, SESAME water for ices, and SESAME dry sand for rocks. These will be kept constant for all exoplanets and all the models inspected.
The used abundances of hydrogen and helium in our models is assumed to be matching the solar abundance (Y/X= 0.27, where Y is the helium mass fraction and X is the hydrogen fraction). The ratio between the two chemical elements adaptively (in the case of the dilute core, for core envelope it has been manually changed when adding material to the envelope) remains the same when heavy metal composition Z has been added, such as that X+Y+Z=1.

3.3 Super adiabatic heat transport

Deviations from the assumption of a fully adiabatic interior will be made. Referring back to the theory, we will apply the Ledoux criterion for the DCA models, in order to account for the contribution of an increased mean molecular weight caused by the gradient in heavy elements. This would cause deviations from adiabatic behaviour (becoming super adiabatic). In order to account for double-diffusive convection (that overturns and counteracts the stabilising effect of the compositional gradient), we have applied a scaling factor to the criterion. The simplified formula can be represented as the following:

$$\nabla_{\text{rad}} > \nabla_{\text{ad}} + f\,\nabla_{\mu} \tag{13}$$

where the scaling factor f will simply be an arbitrary value between 0 and 1, where 1 means no double-diffusive convection, and 0 means no mean molecular weight contribution (the Schwarzschild criterion, following adiabatic behaviour). Since a completely efficient compositional stabilisation (f = 1) is unlikely to happen according to Debras and Chabrier (2019), different f values will be tested. Interior structures tested with such scaling factor are denoted as 'DCAf' models. This inspection is aiming at quantifying the amount of additional heat that can be stored with a fixed compositionally diluted structure (and quantify its deviation from an adiabatic profile), and to also inspect any deviations in the estimated Z.

3.4 Root-find retrieval

This section will describe the process developed for the retrieval of Z distributions. Our goal is to infer each planet's bulk metallicity Z in a way that reflects the actual observational uncertainties, and that accounts for such range. While estimating a lower and upper bound of radius and age can be done in evolution model outputs, this will not inheritently reflect the real data distribution of the most likely Z in planetary interiors. The aim was then to create a statistical method that retrieves a distribution of all possible metallicities and that quantifies better the behaviour of such possibilities, and accounts for possible degeneracies. Such effort is to reduce as possible the error generated by the retrieval method, and to improve the estimate (moving side-by-side with improvements in observational constrains due to technological advancement).

The retrieval can be divided into three main steps:

- **Step 1- Multiple evolution runs:** CEPAM will be automatically running for multiple Mass and Z combinations to retrieve the respective cooling tracks outcomes (R and Age information). The Z steps chosen are going from 0 to 1, with steps of 0.025. Mass values will instead will represent the mean mass from the observed data of each planet, its +/- 1 and +/- 2 σ ranges. Not accounting for the mass range, and relying on a fixed value, will shift the outcome (as shown in Figure 6. It is important to keep the metallicity steps (as much as computational time permits) as low as possible in order to explain any possible degeneracies that might happen.

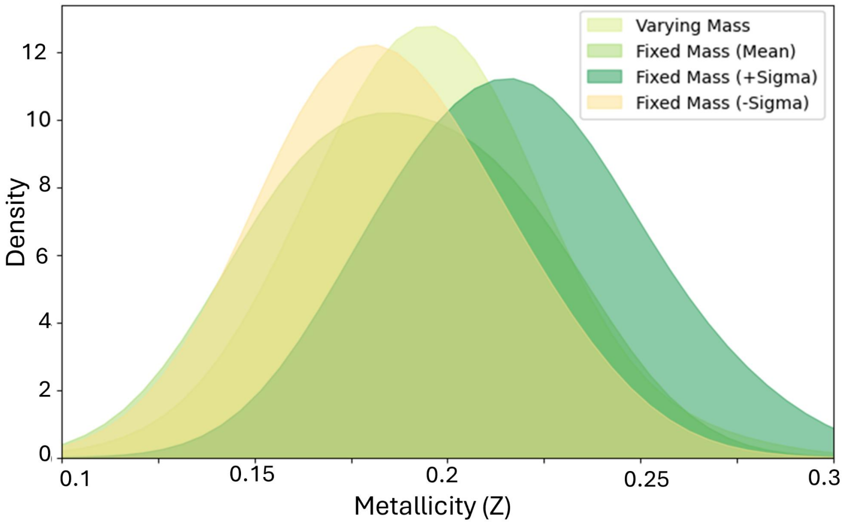

Figure 6: Comparison between the used grid with varying mass, and the behavioural shift when such varying mass is not applies, and when a fixed mass (either the mean or the mean value +/- 1 σ) shift. The case is shown for the metallicity of HATS-75b for the CE structure.

- **Step 2- Grid creation:** From the multiple runs saved from step 1, now using cubic spline (linear in the majority of cases should be sufficient, but the cubic spline method captures degeneracies better) interpolation between the retrieved R/Age lines, but also in the other two dimentions of Z and M. This has been done by using `interp1d` from `scipy.interpolate`, `method='cubic'`. A grid comprising of Z values of steps every 0.0025 (from 0 to 1, with a total of 400 steps) have been interpolated, together with the gaining information for +/- 1 σ of the mean mass value (creating 9 steps for the used mass range). Mass range has been extrapolated to be ranging in the +/- 3 σ range. Such range explains 99.7% of the data, and even if results' error are being slightly under represented (error bars resulting slightly smaller than supposed to be when considering the full mass range), we believe such approximation does not greatly affect the output. Further tests and results for assessing the right grid size for a good balance between precision and computational time and energy, can be found in the Appendix section (Section 8.2).
- **Step 3- Root-find:** Once the grid for the specific planet investigated has been created, we can proceed to the last step for metallicity estimate retrieval. The Brent's method (`scipy.optimize.brentq`) will be applied (Brent 2013; for additional information on the computational side see Virtanen et al. 2020) . Bulk metallicity retrieval has been done for each planet. One retrieval consists of drawing a single tuple (M, R, Age) from the observational uncertainty distributions and then running the evolution + root–finding procedure to obtain the unique metallicity Z that reproduces that tuple. We repeat this $n = 5000$ times, producing 5000 independent $(M, R, \mathrm{Age}) \to Z$ solutions and thus an empirical posterior for Z.

  We assume (motivated by the measurement methods) Gaussian errors for mass and radius and a uniform (top–hat) age range $M \sim \mathcal{N}(M_{\mathrm{obs}}, \sigma_M)$, $R \sim \mathcal{N}(R_{\mathrm{obs}}, \sigma_R)$, $\mathrm{Age} \sim \mathcal{U}(\mathrm{Age}_{\mathrm{min}}, \mathrm{Age}_{\mathrm{max}})$.

  Each sampled tuple is independent. The root finder (Brent's method) searches the pre-computed model grid (or calls the model interpolator) to return the Z that matches the sampled (M, R, Age). This can be explained by the following formula:

$$f(Z) = R_{\mathrm{pred}}(Z_{\mathrm{pred}}, M_{\mathrm{pred}}, \mathrm{Age}_{\mathrm{pred}}) - R_{\mathrm{obs}} = 0, \quad (14)$$

where R_{pred} denotes the predicted values stored in the grid, and R_{obs} refers to the observed values generated with their respective distributions. While the formula aims at a result of exact zero, we allow for a three decimal places margin. For the root-finding to successfully work, the predicted radius should lie within a range such that:

$$f(a) \cdot f(b) < 0,$$

where a and b represent values within the range. This condition ensures that the equation is valid and the function changes signs, i.e., one side is positive and the other negative. If $f(a) > 0$ and $f(b) < 0$, it means the predicted radius at $Z = a$ is larger than R_{obs}, and the predicted radius at $Z = b$ is smaller than R_{obs}, so there must be a metallicity Z_{root} in between where $R_{\text{pred}} = R_{\text{obs}}$. The chosen range for the metallicity Z is from 0 to 1. While this range could be narrowed, using 0 and 1 guarantees that all possible physical values (from 0% metal content to a planet made entirely of heavy elements) are considered. The 5000 draws can capture not only the propagated observational uncertainties but also any asymmetries in the inferred Z. Possibilities of unsuccessful roots in the method can be due to retrieved values outside the 0-1 range, or for tuple combinations that resulted in no convergence. Further evaluation on the percentage of successfull convergences is further discussed in the Appendix, together with a comparison with previous Z estimation methods.

At the end of this step, density curves of the Z distributions have been plotted and shown as histograms using kernel density estimates (KDE) using the Python function `sns.kdeplot`.

3.5 Data analysis

We first plot M_z and Z versus planetary mass using the mean and $\pm 1\sigma$ values. We use a Markov Chain Monte Carlo (MCMC) sampler to explore parameter space and to fit a model representative of the data (Foreman-Mackey et al., 2013). Several chains wander through possible parameter values, preferring those that fit the data better. After an initial "burn-in" period (early steps we discard while the chains settle), the later steps act like random samples from the true solution. The $10'000$ 'random' draws will still follow a normal distribution in both axes, where the data point represents the mean value, and the error bar the 1σ range. This has been achieved with the `emcee` function on Python. The posterior distribution of the draws are shown in corner plots, and can they can be seen in Figure A.11 and Figure A.12 in the Appendix (Section 8.5), also showing the uncertanty of slope and intercept of the best-fit lines. Such study has been done separately for each structural assumption, and the CE, FM, DCA, and DCA3 models are the chosen structures that underwent such analysis.

To quantify any correlation between the plotted variables, we computed Kendall's rank correlation coefficient τ (as done by Müller and Helled 2025; Thorngren et al. 2016; Teske et al. 2019) for each Monte Carlo realization, and for their average point. Kendall's $\tau \in [-1, 1]$ measures the strength and direction of a monotonic relationship based on the difference between the number of concordant and discordant data pairs. Like all correlation test, $\tau > 0$ indicates positive, and $\tau < 0$ negative correlation. $\tau \approx 0$ implies no detected monotonic trend. For each realization we also record the two–sided p-value: the probability, under the null hypothesis of no monotonic association, of obtaining a τ at least as large as observed. We will use an acceptance level of p=0.05, a standard limit adopted in data analysis. Correlation tests will aim to give statistical validity to patterns and trends in exoplanetary characterisation that will be covered in the following sections.

4 Results

4.1 Retrieval comparison

First of all, to check that the retrieved values are consistent with the ones found by Howard et al. (2025) using the same dataset, $Z_\mathrm{p}/Z_\star$ (planetary bulk metallicity normalised by the star's metallicity) over planetary mass has been plotted. The initial 21 samples have been compared with the same planet samples where Z has been calculated with a different method (as $Z_\star$ values used are the same). Both the results from CE and FM structures are being inspected (shown in Figure 7). In the panels on the second row of Figure 7, the normalised difference between the retrieve results, and the ones from Howard, has been calculated. The normalised difference shown in the lower panels is being calculated with the following formula:

$$\sigma = \frac{\text{Howard} - \text{CE/FM}}{\sqrt{\sigma^2_{\text{Howard}} + \sigma^2_{\text{CE/FM}}}} \tag{15}$$

where σ_{Howard} and $\sigma_{\text{CE/FM}}$ are respectively the +/- 1 σ ranges for respectively Howard's results and the ones retrieved with the root-find method (for both CE and FM).

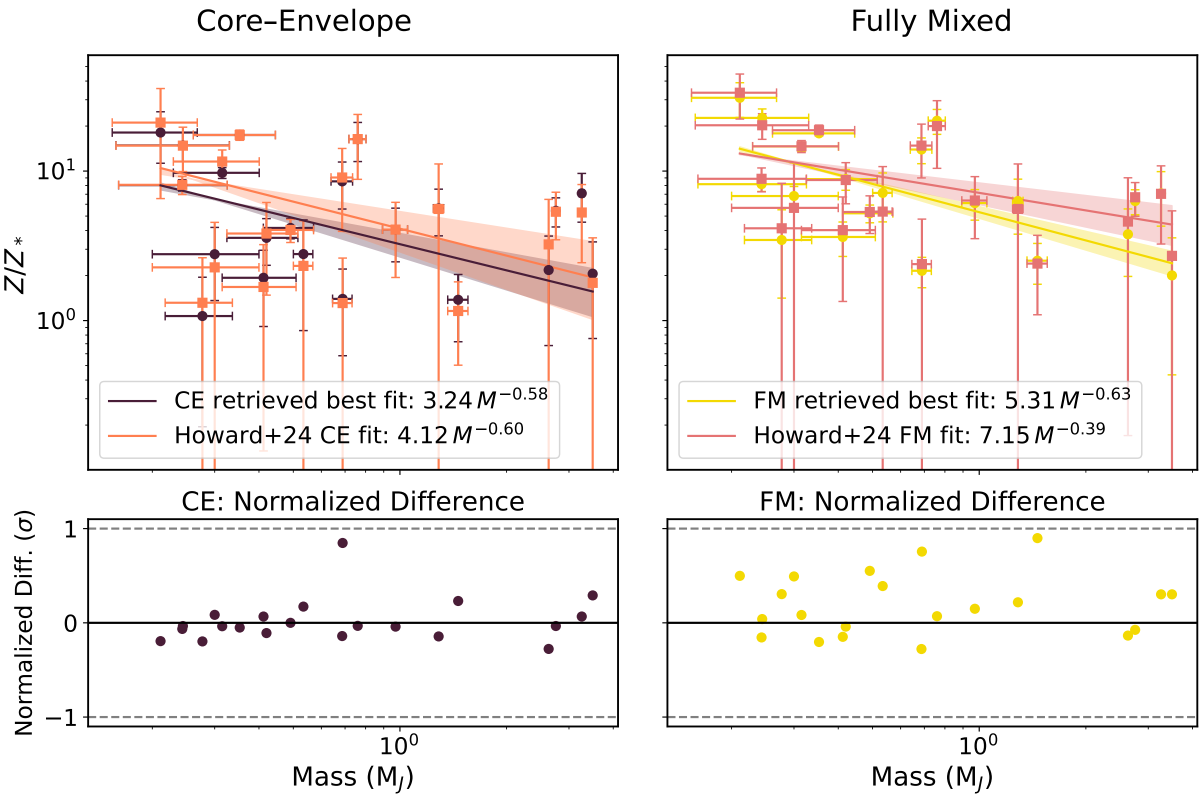

Figure 7: Comparison between the retrieved results for FM models made by this study with the ones retrieved by Howard et al. (2025), plotted as normalised Z by the stellar metallicity $Z_\star$, over planetary mass M.

The upper plots in Figure 7 show a reduction in the error bars generated in the retrieved results. This is especially seen in the lower $Z_\mathrm{p}/Z_\star$ values where the log-scale is 'stretching out' more the lower Z error bars. In terms of trends and best-fit line comparison, there are slight deviations between the two lines. The model used to explain the trend is a power law function, that becomes a straight line when plotted in log-log scale. For the Core-Envelope, the slope of the two function is the same (being around -0.6), while there is a small underestimation by our retrieved model that shifts the intercept down by approximately one digit. Conversely for the Fully Mixed comparison, The slope of the trend augmented, with the starting point intercept sitting above Howard's prediction, and ending underpredicting at higher M. In general, there is an under prediction of the retrieved results compared to Howard et al. (2025). In order to evaluate whether the deviation of the planetary retrieved values of Z is outside the normalised difference of +/- 1 σ, the difference between the two models has been plotted. All values are sitting within this range, meaning that the retrieved values do not substantially deviate from the ones from Howard's (accounting for their error ranges). FM shows more scatter that CE, that could be caused by the difference in dealing with degeneracies between the two approaches. Furthermore, there is no seen trend in the lower panels in CE and FM, showing no systematic error on the mass range. Since all values sits within the σ range, we decided that the results retrieved by this study were convincing enough to make us continue with such methodology.

4.2 Bulk metallicity

4.2.1 CE-FM

Now that we evaluated our approach and we proved its self-consistency, we can now inspect CE and FM results, to then continue with the other tested structures where comparison for the structures tested cannot be made. Starting easy with the comparison between Core+Envelope and Fully-Mixed, we can see the resulting Z distributions from the root-find method, and plotted using the KDE smoothing function. The two curves that can be seen in Figure 8 for planet HD332231b, where it show the results for these two structures.

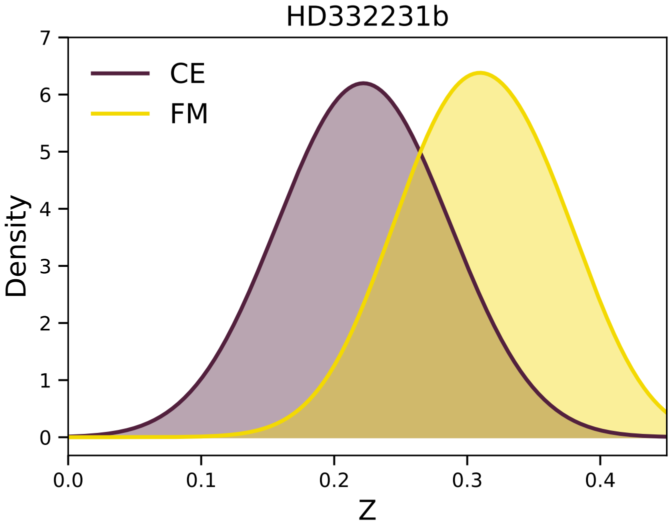

Figure 8: Retrieved heavy element fraction curve distributions for HD-332231b. Fully Mixed is shown in yellow, while Core+Envelope in dark purple.

As shown in Figure 8, here is a difference between the retrieved values, with FM expecting a higher fraction of heavy elements (with the average of FM being around 31% of the total planet, and CE expecting approximately 22%). While the density of these values can be informative on the error ranges of these curves, such shifts could also be subject of the difference in successful converges (especially for curves close to zero where no Z can go below 0. Such issue is being mentioned in the appendix, and for accurate mean and σ values of the curves please consult Table A.2 and Table A.3 in Section 8.1.

While planet HD-332231b expresses a difference between the two models, to draw accurate conclusions about systematic differences between the models, the full data range must be inspected. For this, the initial 21 planets (and their relative distributions) have been displayed for comparison

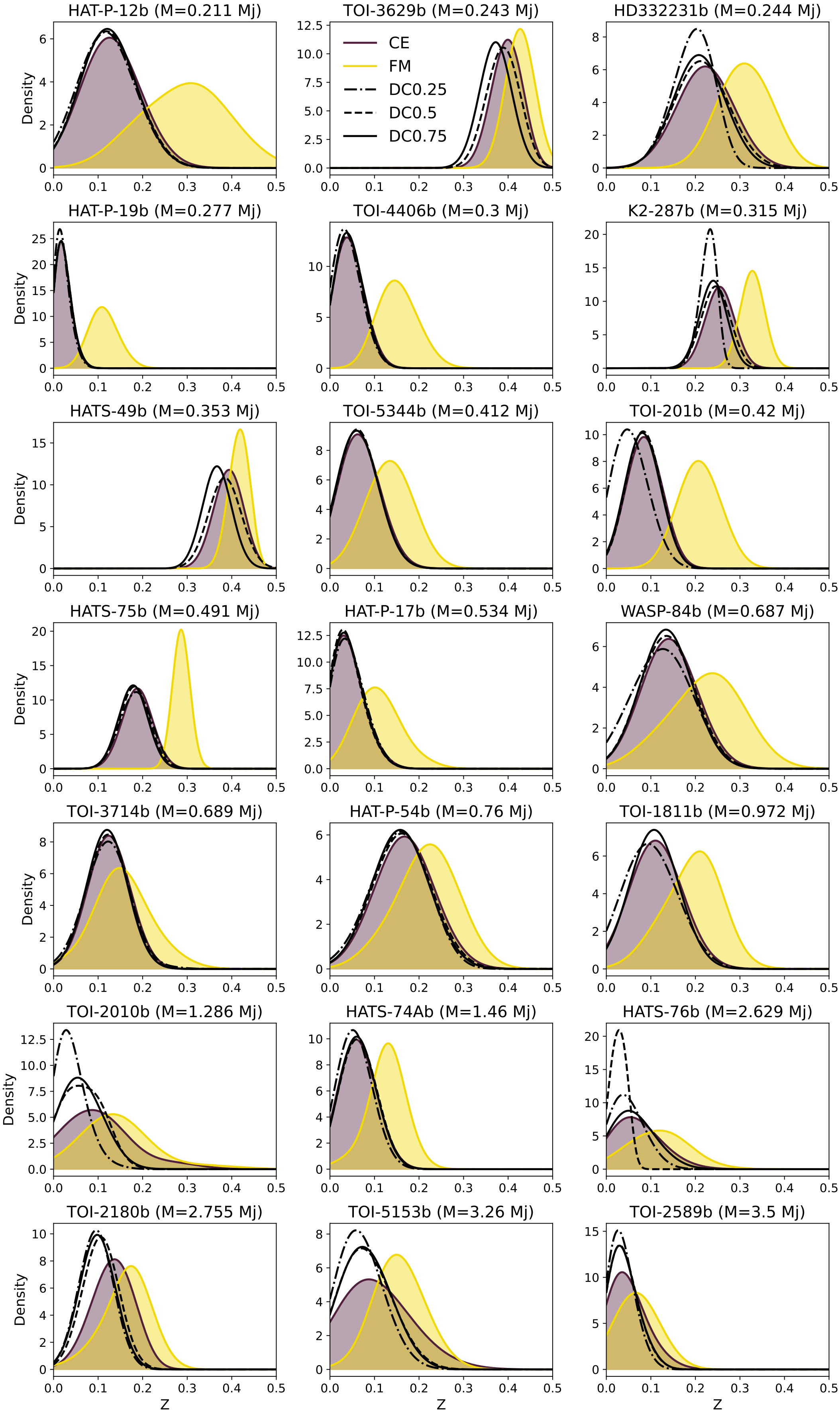

Figure 9: Retrieved heavy element fraction curve distributions for the initial 21 planets. The models compared in this picture are the CE (in yellow), FM (in purple) and DC0.25, 0.5, and 0.75 (different mdilutes, respectively in dotted-dashed, dashed, and continuous lines).

]

sorted by their masses (smaller to bigger). Figure 9 shows the distributions for FM and CE, together with the dilute core with changing m_{dilute} values. For all the original 21 planets, the estimated Z from the assumed FM structure are expected to be higher values compared to what estimated with a CE structure. In a nutshell, for the FM case, the greater difference in estimates is mainly controlled by the opacity increased by the atmospheric metallicity. A more insightful explanation of this mechanism can be found in Section 5.1.1.
It is also shown how the size and spread of distributions, together with the expected Z, varies a lot between different planetary examples. The spread will be mainly controlled by the observational error. FM curves appear to be overall slightly wider than the CE (with lower probability density), and this might indicate a greater error range given by this structure's estimates.

4.2.2 DC

Other than FM and CE models, in Figure 9 are also plotted the retrieved distribution curves for DC models with no atmospheric metallicity. The different extents of the variable m_{dilute} for tested values of 0.25, 0.5, and 0.75 of the mass extent, are respectively shown in dotted-dashed, dashed and continuous black line. We can see that for all the three scenarios, the estimated Z does not significantly differ from CE. All three curves match quite well with what can be seen for the Core-Envelope estimates. While in some planetary examples (like TOI-5344b, HATS-75b, HAT-P-17b, and TOI-3714b) the three DC models exactly match with the CE distribution (both in mean value, wideness, and probability density), in other cases such structures underestimate Z abundances with respect of CE. The most obvious example of this is shown in planets like TOI-2010b, TOI-2180b, and TOI-5153b. To note that planet TOI-3629b does not have a solution for DC0.25, since the expected Z is around 0.4, and a structure such as the one of DC0.25 cannot hold more heavy elements in its interior without increasing the m_{dilute} value (at maximum possible Z_{dilute} of 1). Overall, DC models underpredict on average the planetary metallicity. The distribution of heavy elements is what differentiates CE and DC models, as there is no opacity increase due to Z_{atm}=0 in both structures. Differentiation between these three models is difficult, as the observational error is too large for a reliable assessment. Once again, more details on the reasoning on predictive Z shift can be found in the following Discussion Section.

4.2.3 Models with Z_{atm}

In this section we will display the results for the other tested structures. Since Z_{atm} is believed to be controlling Z estimates, it is important to further inspect and take into account this effect when running dilute core structure. For this reason, we will delve into the changes applied to Z caused by varying Z_{atm}.
Before running evolution models for each planet, an additional inspection has been made. We first wanted to see any structural behaviours for fixed parameters (so not tailored for each planet). By running the evolution model for fixed synthetic cases, it could provide predictory information on the general behaviour that can then be seen in the following planetary Z distributions. CEPAM has ran and retrieved $R_{(Age)}$ values for different masses, ranging for different ages and equilibrium temperatures. Ages tried where 1 and 5 Gyr, and T_{eq} of 500 and 1000 K. By running the model for different masses (respectively 0.1, 0.3, 0.5, 0.8, 1, 3, 6, and 10), we were able to create structural curves that predict the radius-mass relationship of gas giants exoplanets over a 0.1-10 Mj mass range. Such plot can be seen in Figure 10 (only showing the output for Age=1Gyr and T_{eq}=1000K). The tested structures were all fixed to a standards M_z of 30 $M_\oplus$ (and thus not varying for adapting to the planet). Plotted in grey are the planetary data from the extended dataset found in Howard et al. (2025).

As shown in Figure 10, the different tested structures do play a role in the estimated radius. The structures follows the planetary data. All structures capture the electron degeneracies happening at higher masses, where the resolved radius does not increase as mass increases (as previously explained in Section 1.4, and in Figure 3). Planets that seem to be not represented by the curves is

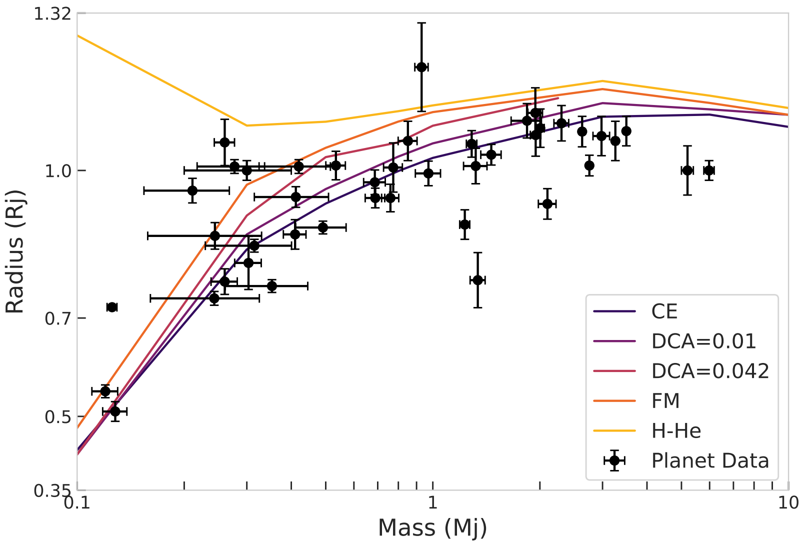

Figure 10: Radius–mass relationship predicted by CEPAM. Planetary radii were computed at fixed equilibrium temperature $T_{\mathrm{eq}} = 1000\,\mathrm{K}$ and age $t = 1\,\mathrm{Gyr}$ for four internal structures and compositions: (1) a core–envelope model with heavy-element mass in the core $z_{\mathrm{core}} = 30\,M_{\oplus}$ and in the envelope $z_{\mathrm{env}} = 0$; (2) a dilute-core model with 1% atmospheric metallicity; (3) a dilute-core model with three-times–solar metallicity (Jupiter-like); and (4) a fully mixed distribution with a $30\,M_{\oplus}$ fixed at $30\,M_{\oplus}$. Models were evaluated at planet masses $M_p = 0.1$, 0.3, 0.5, 0.8, 1, 3, 6, and $10\,M_{\mathrm{J}}$.

just because of the fixed $30\,M_{\oplus}$ of M_Z tested. Planets that seems to 'sit' on top of the H-He curve are not inflated, as all the planets inspected have a Z solution when individually tailored with the respective planetary data. The different curves show the radius results for respectively: A planet fully composed of hydrogen and helium (H-He), a Core-Envelope structure CE, a Fully Mixed one (FM), a Dilute Core structure with a 1% atmospheric metallicity (DCA=0.01), and a Dilute Core with 3 times the Solar metallicity (likewise Jupiter, denominated DCA=0.042 as Z_{atm}=0.042). DC models have been run with an assumed m_{dilute} of 0.5. Radius estimates explained by the curves follow a pattern in this succession from lower to highest: CE, DCA=0.01, DCA=0.042, FM, and H-He predicting the highest radius (even at low masses). All curves are showing a similar direction through planetary masses, apart from the great difference between H-He and the other mothels at low masses. At low M, differences between CE, DCA0.01, and DCA0.042 are hard to interpret, as the majority of the mass is composed by heavy elements. On the rest of the mass range, even if the M_z is kept fixed, the resulting higher expected radii of DCA=0.042 compared to DCA=0.01 shows some dependency on Z_{atm}. While the plot only shows the example for T=1000K and Age=1Gyr, same patterns can be seen in the other combinations of these two parameters. This plot not only shows how CEPAM is able to replicate R-M relationships in gas giant exoplanets, but it also informs us on what are we expected to see for the following DCA models.

After this preliminary inspection, we now can see the actual estimated Z for DCA models uning Z_{atm} as the planet's stellar metallicity, together with DCA3. Figure 11 is showing the different Z distributions for four models: FM and CE (previously discussed) with the addition of DCA and DCA3 retrieved curves (shown in orange and vinaccia colour respectively). The vertical dashed lines show the lowest possible metallicity that the DCA and DCA3 models can allow (colours matching the model curve distributions). This is because a constant stellar (or 3 times stellar) is forced by the structure, so below such metallicity limit the structure cannont go (hence the hard break on the curves). Since a inverse composition gradient could be unstable over time, the lower limit will be kept as Z_{atm}=$Z_{\star}$ for DCA, and $Z_{\mathrm{atm}} = 3 \cdot Z_{\star}$. The results prove what previously inspected in Figure10. In the majority of the planets tested (here shown for the original 21 planets), DCA and DCA3 sit in between the two structural end-members CE and FM. In planets like K2-287b, TOI-

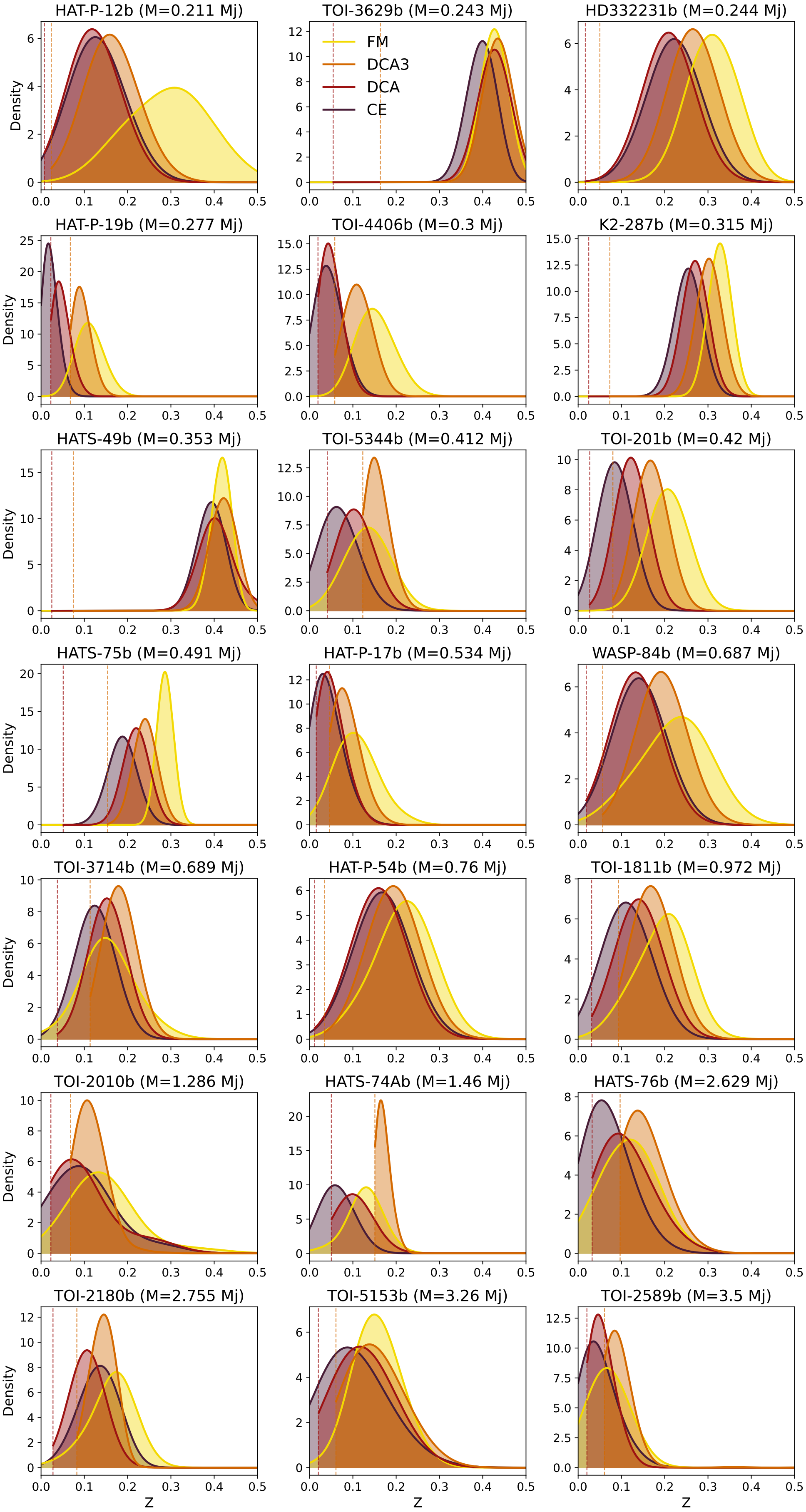

Figure 11: Retrieved heavy element fraction curve distributions for the initial 21 planets. The models compared are FM, CE, DCA, and DCA3.

201b and HATS-75b, the transition between models is clear and shows the order from the lowest Z predicted to the highest, respectively: CE, DCA, DCA3, FM. However, in many situations this trend cannot be observed; in some situations, like in planets TOI-3714b, HATS-74Ab, HATS-76b, and TOI-2589b, DCA3 models predict higher Z mean values compared to FM. In the special case of HATS-74Ab, what the KDE resolved curve represents in reality is a tail of a normal distribution, recognised as a peak by the KDE. This is because the 3 times $Z_\star$ forced by DCA3 models narrows down the Z possibility range. This could be due to an inappropriately-high tested Z_{atm}, but it also highlights how atmospheric metallicity (if known) can constrain the range of possibilities. Similarly, some DCA models do show underprediction of Z when compared to CE. This can be seen in examples like HAT-P-12b and HD332231b. Such retrieved DCA curves look very similar to the ones resolved without atmospheric metallicity (DC).

CEA models, so Core+Envelope with Z_{atm}=$Z_\star$, were having similar retrieved Z results to the ones given by the DCA that have been tested with the same metallicity value in the atmosphere. A plot showing such results (and comparison between CE, CEA, and DCA) can be found in the Appendix in Figure A.13 (Section 8.6). Similarly, SMA models with solar atmospheric metallicities, are similar in the estimated fractions, with not big notable shifts that would not be accounted by the observational error carried. There is no plot showing the SMA, however their mean Z values and σ can be found in the appendix in Table A.3.

To further inspect the full parameter space created by the dilute core function, additional tests on the TOI-1811b planet case study have been done by isolating one of the variables of the gradient function. This has been done by retrieving a single evolution cooling track, and then compare it with changing structural variables. To note that this method does not capture the degenerate behaviour, and that the results can be generalised for all the planetary dataset, such assessment was only valid on the case for TOI-1811b that has been inspected. Results of the investigation can be found in section 8.4 in the Appendix, and are consistent with what we have investigated by the above-mentioned findings.

4.3 Thermal transport

Varying the superadiabatic gradient with different f values (as shown in Equation 13) shows distinctions caused by retaining more internal temperature. The f values tested in Figure 12 range from 0 (assumption of the thermal gradient following an adiabat), to 0.1, and finally 0.4. No greater values have been tested to avoid unrealistically hot interiors (even though this is case dependent). For computational time reasons, this study has been done for a fewer set of planets (10).

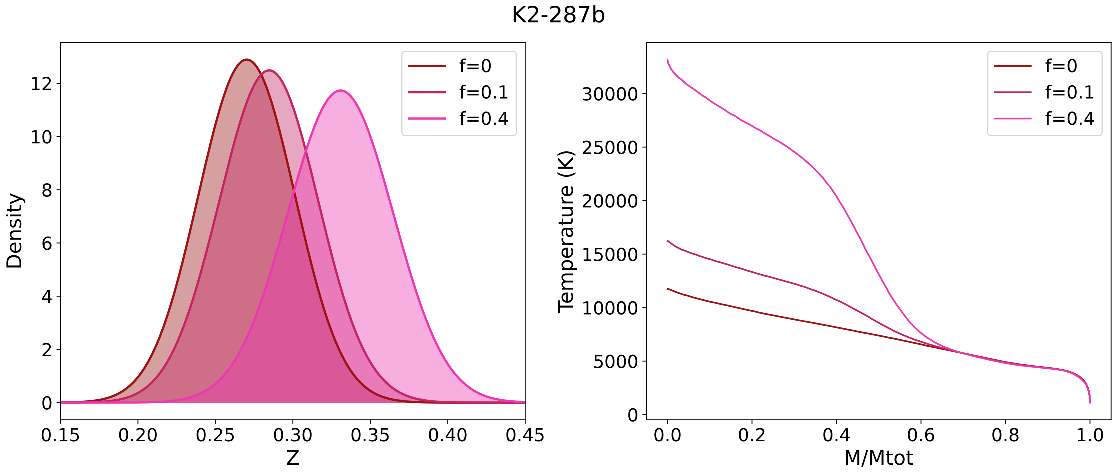

Figure 12: Z distributions and thermal profiles for different assumed internal heat transport, from adiabatic (f=0) to superadiabatic (f=0.1 and 0.4).

They all show the same results, just to different magnitudes. Figure 12 shows the different Z curves with such f factor change (panel on the left), while on the right the internal heat structure (Temperature changing the planetary mass fraction, where 1 is outside, and 0 is the centre) follows adiabatic behaviour for $f = 0$ (no gradient contribution), and the two super adiabatic heat structures shown in orange and red. With higher f the extent of the temperature stabilised in the planetary interior is drastically higher in $f = 0.4$. Such hotter interior allowed to introduce more heavy elements shown by the shift on the left panel of Figure 12. There is a drastic change in temperatures reached by the different models, with the adiabatic estimating approximately $10'000$ K, and the DCA$f = 0.4$ exceeding the $30'000$ K. A greater scaling factors yields larger estimates, as gradient stabilisation is being increased according to the Ledoux criterion. The composition curve steepens out at the m_{dilute}. So, the extent of m_{dilute} also determines how much hotter the interior will be, and how much it deviates from an adiabatic profile. Looking at planet K2-287b, the extent of estimated distributions at $f = 0.4$ is lower compared to its Z estimates with an assumed FM structure (around 32%). This is not the case for planet TOI-3629b that matches the same estimated bulk metallicity, or even by planets like HD332231b and HATS-49b that marginally have higher retrieved estimates for DCAf compared to FM (This can be checked by comparing the values between Tables A.2 and A.4. Moreover, the increase created by the scaling factor seems linear in the plot provided, but it is not in other planetary examples tested.

4.4 Relationships and trends

After retrieving all the bulk metallicity distributions for the main inspected models (FM, CE, DCA, and DCA3), now we can inspect possible trends shown by the resulting mean values and standard errors for M_z and Z. In this section, we will show the different relationships for different compared variables. This will be done over the full set of 43 planets, as the subsequent statistical assessment on the plots will be strengthened by a bigger sample size.

4.4.1 M_z/M Relationship

In Figure 13 we see plotted M_z (in Earth masses) on the y-axes, and the mass of the planet M in the x-axes. For all models, their estimated M_z and relative errors both axes is being implemented. The Bayesian stochastic approach done with the MCMC to retrieve best fit lines and error can also be seen. The fits are a power law, that when placed in a log-log space (since both axes are in log scale) it becomes linear. The error bars shown by the best fits shows the uncertainty on those model predictions, coming from the two fitting parameters α (slope) and β (intercept). The best-fits show an increase in heavy element mass with planetary mass (i.e. more massive planets hold more heavy elements). FM data points are generally estimating higher M_z. This is further confirmed by the FM best-fit line having the highest intercept value of $\beta \approx 60\, M_\oplus$. FM, is followed by DCA3, DCA, and finally by CE, similarly to what can be seen in Figure 11. The slope α of the best fits is relatively steep for all the models, ranging from 0.65 of FM to 0.72 for DCA3, that shows the steepest line.
When compared to the relationship found by Thorngren et al. (2016), results from this study appear to show a steeper relationship (given by a higher correlation and slope of the model) compared to their results. While the $M_Z \propto M_p^{0.61\pm0.08}$ relationship found by them stands and is within the error range for FM and CE models, DCA and DCA3 show steeper slopes that would be underestimated by Thorngren's statement, especially with their simplified relation of $M_Z \propto \sqrt{M_p}$.
Shown in the plot are also the lines showing how would the M_z be for fixed Z values, and how would it change with mass. If all planets would be fully made of heavies (Z=1), the followed behaviour would follow the 0% H-He line, and similarly for 50% and 99% in heavy element composition. The slopes from the best-fits do not follow the same tendency, meaning that composition is not being constant throughout the mass range.

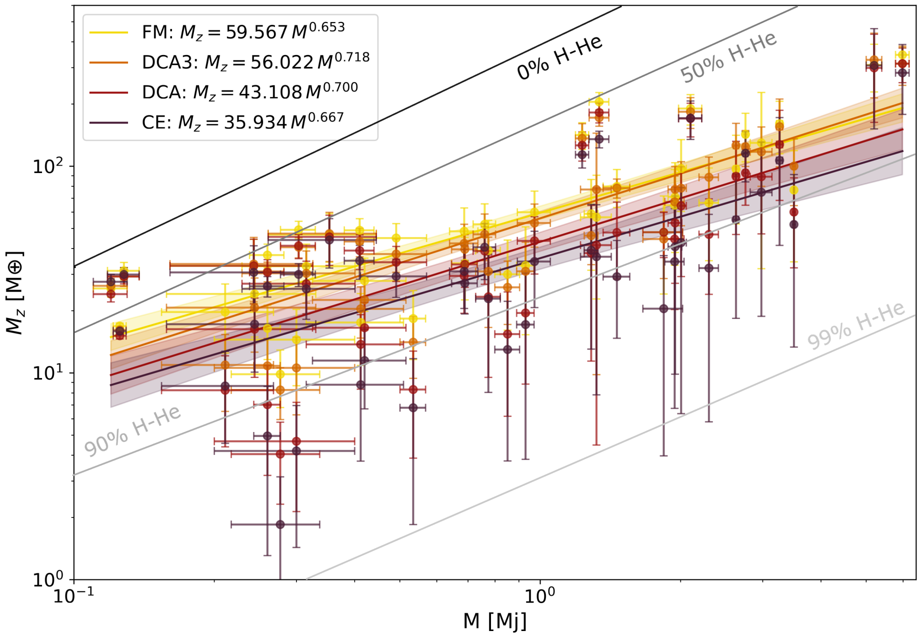

Figure 13: Plotted M_z (heavy element mass, in Earth masses units) versus planetary mass (in Jupiter masses units) for four different structures: FM, CE, DCA, and DCA3.

4.4.2 Z/M Relationship

After looking at Figure 11 and the comparison between CE, FM, DCA, and DCA3, it is worth to inspect the magnitude of the Z deviation caused by model assumptions. We want to inspect if such shifts in metallicity in planetary interiors is related to the inspected mass range of the planet sample. To inspect such shifts, we computed and recorded the respective Z deviations from the FM model (in average the highest estimated Z in the whole sample). The shifts in the metallicity estimates will be plotted over the sample's planetary masses M. Z deviations are named ΔZ and it is showed in Figure 14. In said figure, the mean average points from the distributions of the four different models have been taken. From there, ΔZ has been computed for each model in comparison to the FM reference by simple subtraction, such as the different lines (purple, vinaccia and orange) represent ΔZ respectively of FM-CE, FM-DCA, and FM-DCA3. If ΔZ is positive, it means the FM structure predicts higher metallicity compared to the model compared with. In contrast, negative values would reflect that the compared model is expecting higher Z than FM. Best-fit lines have been shown for the found points, and they have been retrieved with a simple ordinary least of squares. This linear model retrieval, together with just plotting mean values of ΔZ is an approximation that does not account for variance and model uncertainty and it is for the only scope of visualising in simple terms. Nonetheless, a statistical correlation test by running a single Kendall's tau correlation coefficient for each models has been done for inspecting correlation between the two components.

Results show negative correlation, shown by a moderately negative Kendall's τ coefficient. τ ranges from -0.25 for FM-CE, to -0.39 for FM-DCA, and -0.34 for FM-DCA3 (also shown on the figure). This is further visually facilitated by the best-fit lines that carry this trend. CE always expects less Z than FM, with the only exception of the interesting scenario of the lowest-mass planet in the sample set, planet TOI-4010d. So, The differences between FM and CE are reduced at bigger masses, and similarly to DCA and DCA3, where the only difference is an increased overprediction from DCA3 moving to more massive planets (trend going into the negative regime). At low masses, there is a greater variance of ΔZ, but it could be caused by a greater amount of sample data at relatively lower masses. DCA3 records many negative ΔZ, with the FM-DCA3 linear model predicting the slope entering in the negatives at higher masses. This symbolises what previously discussed in

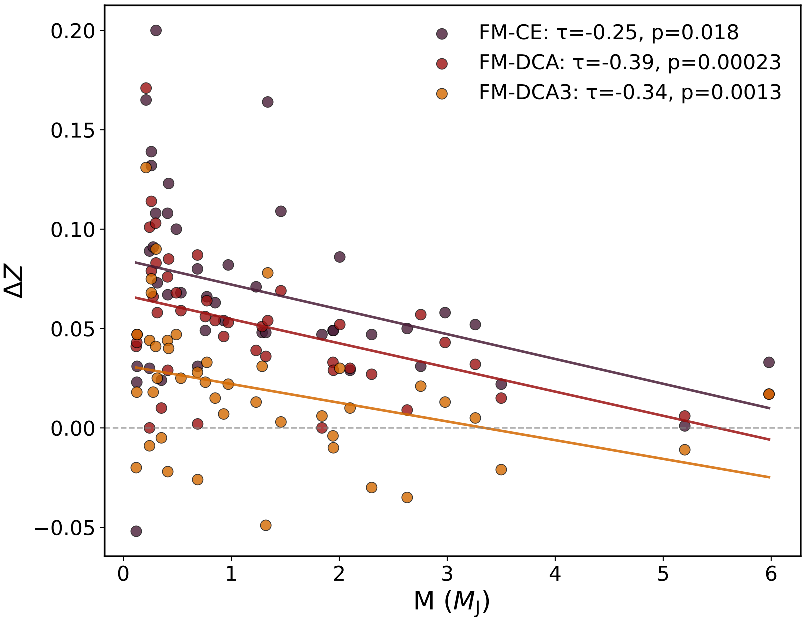

Figure 14: δZ comparison between FM, CE, DCA, and DCA3. ΔZ is the difference between FM and the other models. Kendall's τ also computed for these relationships. The zero line delines no difference between the FM model with the one being compared with.

Figure 11 in forcing high metallicities at high masses, that might not be suited. In such scenario, we need to prove that at greater planetary masses Z is overall reducing in planetary interiors, and thus high Z_{atm} might not be suited.

Similarly to what has been done for Figure 13, we investigated the relationship between two variables, in this case Z and M. Metalliciy has been plotted on the y-axis and planetary mass in the x. The relationship can be seen in Figure 15. Again, all estimated Z mean points and error bars have been plotted and inspected through a power law model.

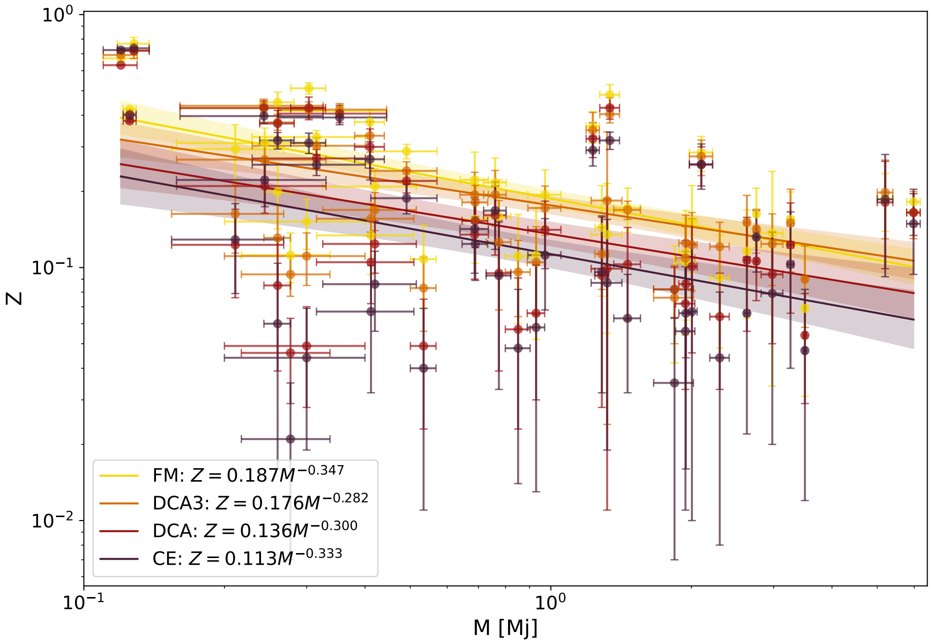

Figure 15: Plot of the heavy-element fraction Z as a function of planetary mass M (in Jupiter masses) for all four models (CE, DCA, DCA, and FM).

As predicted in the M_z/M plot, there is a negative trend shown by all four models when plotting the two quantities Z and M. As previously seen in Figure 13 in the distributions, but also in Figure 13 with the M_z relationship, FM predicted Z overall are higher-end estimates, followed gradually by DCA3, DCA, and finally CE. The best-fit slopes are similar (α ranging from -0.28 of DCA3 to approx. -0.35 of FM). Such negative trend is also found in previous literature like in Thorngren et al. (2016). This behaviour was envisioned in the previous Figure regarding ΔZ, where at higher masses it was expected (especially for DCA3) to estimate higher metallicities than FM, and this is further confirmed in Figure 15 with the DCA3 retrieved line crosses the FM model. As also stated, previously, there seems to be a great variability in Z ranges at low M.

4.4.3 $Z_\star$ normalisation

This variability of planets, but also the trends, could be controlled by the nature of different stellar systems. Metallicity, on a formation perspective, is remnant of dynamics in the propoplanetary disk, but it is ultimately controlled by the composition of the initial gas where the stellar system is being formed. For instance, a planet that formed from a heavy-element rich primordial gas will likely result in higher metallicities independently from its formation process. Furthermore, there is the risk of further inducing a correlation, especially for DCA and DCA3 models, where it has already been forced a parameter of stellar metallicity in their atmospheric composition. For this reason, inspecting the relationship normalised by $Z_\star$ is a must.

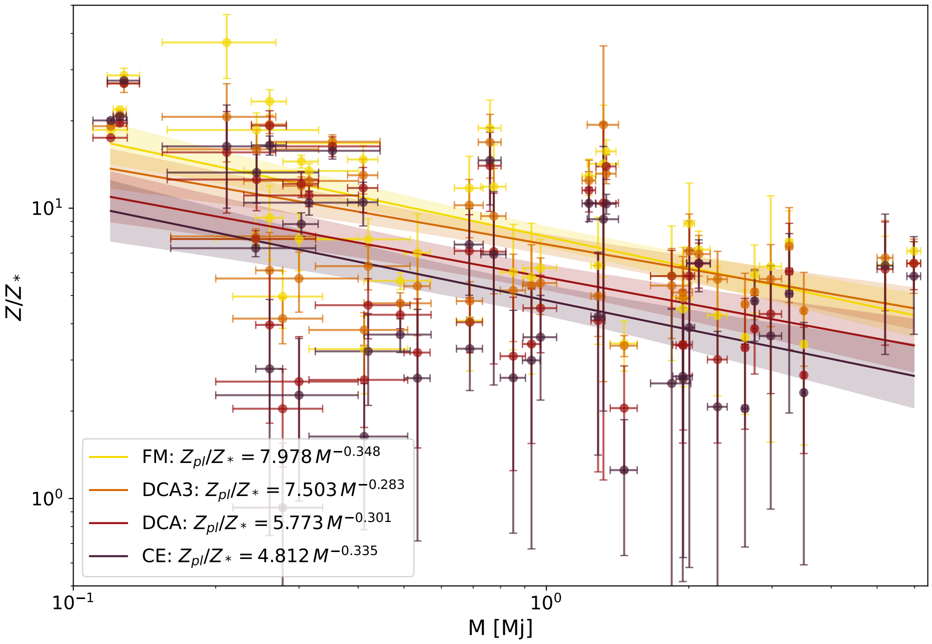

Figure 16: Plot of the heavy-element fraction Z_p normalized by the host star's metallicity $Z_\star$ as a function of planetary mass M_p (in Jupiter masses).

In order to normalise the retrieved Z for their respective $Z_\star$, a simple division between the two values has been done. As previously done, the values have been plotted and fitted over M (shown in Figure 16. The relationship shows similar results to what found for Figure 15, proving consistency in the relationship. Slope parameters α did not really change compared to what seen in the Z relationship. A further analysis has been done in inspecting if there is no correlation between the pure $Z_\star$ and the planetary mass, but no statistically valid result has been found. This is consistent with what found by Teske et al. (2019), where no strenghtening relationship has been found relative to $Z_\star$. Conversely, correlation between M_z residuals (M_z inferred / M_z predicted by the best-fit) and $Z_\star$ shows correlation. While the limited sample size of this study, and in Teske et al. (2019), compared to the one from Müller and Helled (2025) could have played a role, the separation into different star types done in Müller's paper might have simplified the correlation, and this is not something delved in this study.

Going back to our relationship, even with the normalisation, trends remain consistent across different interior models, with a shift in Z by ∼10–20% without altering fundamental relationships. This is consistent with Thorngren et al. (2016); Teske et al. (2019); Howard et al. (2025), where similar best-fit lines have been tested. In Thorngren's, a best-fit line of $(9.7 \pm 1.28)\ M^{(-0.45 \pm 0.09)}$ has been found for a CE structural assumption. while the intercept is higher than the one found in this study, the compensation of α to a steeper line is consistent with what found here. Moreover, the error range of α=-0.45 ± 0.09 is within the found error of this study, with α=-0.33 ± 0.11 (error ranges can be found in Figures A.11 and A.12 in the appendix section showing all the posterior corner plot distributions from the MCMC best fit lines). Retrieved results from Teske et al. (2019) reflect the ones from Thorngren's since the same dataset has been used. Validation with Howard et al. (2025) has already been discussed in section 4.1 and in Figure 7.

Worth mentioning that, while all values show enrichment ($Z/Z_\star$>1), in the case of planet HAT-P-19b shows same enrichment, if not lower (accounting for the error range), for the CE model. The change in the value is being pushed up a lot by the following models, but such preditions are still low compared to its planetary counterparts at around the same mass (0.3 Mj).

4.5 Statistics

4.5.1 Correlation validity

So far we have only talked about the trends carried by the best fits. In pure data analysis terms, before relying on trend lines from linear models, one must first confirm that the underlying data exhibit a genuine correlation. In this section, we will report and comment on the results found from the Kendall's τ correlation tests ran for each previously-explained relationships. For more information on how the Kendall' τ correlation coefficient is being calculated, please refer to Kendall (1938).

In order to show complete transparency on this study, it is important to mention that correlation test validity (given for an acceptance level of p=0.05) was not met initially. A p-value higher than 0.05 denoted failure in finding any true association between the observed variables. This was because of the limited sample size (represented by the initial 21 samples), that has been increased to 43 planets at a later stage. The following results will also be compared to what found in the initial dataset. Filtering criteria are mentioned in section 3.1 to avoid cherry-picking. Kendall's τ for all models for all the three plotted relationships can be found in Table 1 for the final dataset of 43 samples. In all three cases, the correlation matches the previously-decribed trends, positive correlation has been achieved by the M_z plot showing moderatively-high τ values (ranging from 0.51 to 0.66). For Z, a negative correlation can be seen with a weaker correlation given by a negative τ closer to 0 (ranging from -0.4 to -0.245). As previously mentioned, since $Z_\star$ does not strenghten the relationship, resulting τ values are not being substantially affected. In all cases, there is a low chance that such observed correlation is just caused by randomness. The p-values are all under the

	(a) M_z vs M		**(b) Z vs M**		**(c) Z/Z_* vs M**	
Model	**τ**	**p-value**	**τ**	**p-value**	**τ**	**p-value**
CE	0.514	1.20×10^{-6}	−0.273	1.00×10^{-2}	−0.290	6.11×10^{-3}
DCA	0.612	7.59×10^{-9}	−0.245	2.07×10^{-2}	−0.284	7.38×10^{-3}
DCA3	0.660	4.44×10^{-10}	−0.314	3.05×10^{-3}	−0.308	3.62×10^{-3}
FM	0.660	4.44×10^{-10}	−0.397	1.79×10^{-4}	−0.390	2.30×10^{-4}

Table 1: Kendall's τ and p-values across four structural models (CE, DCA, DCA3, FM) for the three different relationships.

acceptance level of 5%, thus we can conclude that there is a true correlation between the data in all three relationships.

The nature of the resulting p is varying between the inspected M_z and Z plots. The M_z/M results in lower p-values for all the four models, and thus for a more reliable conclusion of proving correlation. This is caused by the nature of Z, that has to be constrained to be between 0 and 1 (either no heavy elements of fully made of them). By flattening the signal of the response variable on the y-axis, one also magnifies the scatter of the data, weakening the statistical correlation. Such behaviour is not recorded in Thorngren et al. (2016). However, Thorngren's dataset was encompassing a bigger range, with approximately the same amount of sample points used here (46). We ran a little test of reducing the dataset from the results found of Z/M in Thorngren et al. (2016), to the mass range found in our own dataset (so from 0.12 to 5.98 Mj). We found out an expected increase in the p-value from the correlation test, going from a magnitude of $\times 10^{-5}$, to a p-value of $\times 10^{-3}$. By further reducing the mass range to the initial mass range of the 21 planets used (so 0.211 to 3.5 Mj), we see that the resulting p-value is above 0.05 for Thorngren's, agreeing with what we had with the reduced sample size.

Such case was also present in the initial dataset composed by 21 planets. The limited dataset showed the same pattern , a p-value lower than 2×10^{-4} for M_z/Z relations for all models, with the only difference that for Z/M (and in the normalised scenario too), p-values of the computed correlations was met only by FM (p≈0.01), while the rest ranged from 0.06 to even 0.38. The fact that correlation was met to the absolute (M_z) value instead of the normalised Z (together with the correlation met for FM) led us to think that it was related to sample size, and the results from the extended dataset is the evidence of it.

While the sample size did play a role in bringing statistical validity to the correlation analysis, Kendall's τ is being computed only on the mean values of the data points. While this is a good approximation for data with small error, highly uncertain data points can hide contradictory results. Other than the p-values, correlation coefficients can also be shifted in log-log spaces (like the ones in our study), where large, lower-end errors are being 'stretched' by the non-linear scale.

Additionally to the retrieval of the curves, we once again relied on the MCMC to sample different tuples of data drawn from their distribution ranges. We then recomputed the Kendall's τ correlation and p-value for each tuble iteration. We evaluated τ for each draw of your model parameters from the MCMC chain to capture the whole Bayesian posterior propagation of uncertainty of the data. The resulting distributions can be found in Figure 17. The figure shows in the first panel on the left the resulting distribution of τ and on the right the one for p-values. A vertical line at p=0.05 to demarcate the acceptance limit above which correlation is not statistically met. To mention that what is shown in Figure 17 are the results only from the relationship Z/M, as the relationship

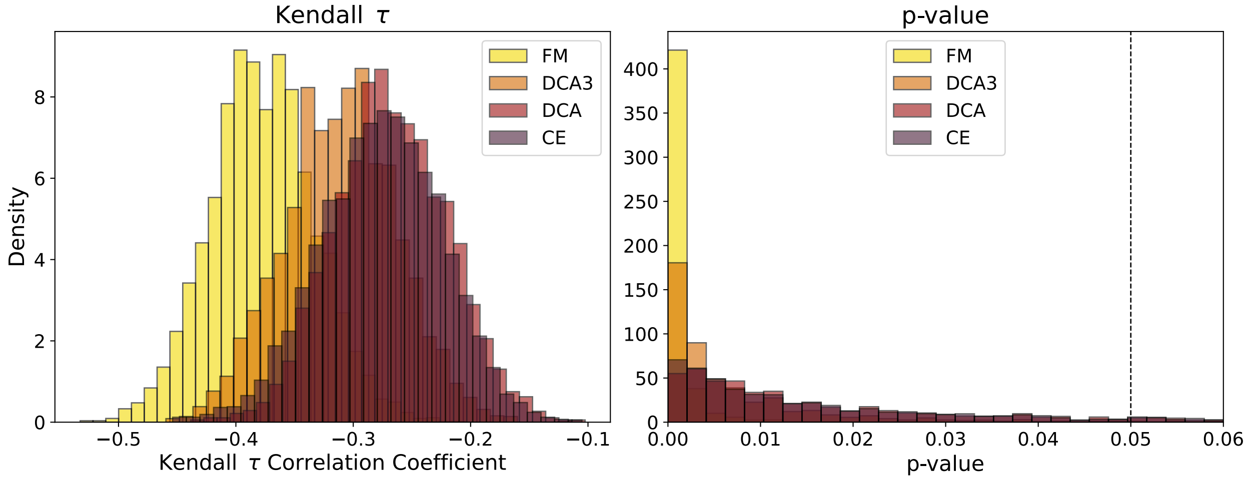

Figure 17: Bayesian retrieval of Kendall's τ (and respective p) distribution for the Z/M relationship.

normalised by $Z_\star$ is essentially the same, and the p-values in M_z/M are well above p=0.05.

The results shown in Figure 17 are consistent with the one computed only for the mean values. In the left panel, the correlation coefficients τ show a Gaussian distribution, with the highest points in the histogram matching the ones from the single computation of τ. FM shows the lowest correlation coefficient, consistent with both the relationship and the previous τ value, while the other models are more or less around the -0.3 value. Such difference of the FM correlation coefficient might be due to its higher Z difference in estimates at low masses compared to the other models (proved in Figure 17), that boosts the correlation (especially for Kendall's method that computes the correlation through ranks).

Talking about the p-value distributions, we can see that for the majority of all the models is situated below the acceptance level of 5%. By noting the percentage of retrieved values below that p=0.05 limit, we see how FM has 99.98% of p-values below the acceptance level, followed by DCA3 with 98.46%, and by DCA and CE with respectively 89% and 88.5%. We see again the recurring follow up of FM, DCA3, DCA, and CE. Thus, according to the internal structure assumed, there is not only a shift in expected Z (either up or down), but also flattening or steepening of the curve on the mass coordinates. Flattening would make correlation rank analysis weaker, and thus showing τ values closer to 0, and higher p-values. Having said that, we believe that even CE, with a recorded 11.5% of the retrieved results above p=0.05, is sufficiently accurate to satisfy the statement that all models show a correlation.

Since Kendall's correlation depends on ranks, so values in y either increasing or decreasing moving on the x-axis, we wanted to do an exploratory analysis on the dependence of the curve's correlation on samples situated on the outermost limits of the x-axis (namely the highest and smallest planetary masses for the Z/M relation plot) and see how the correlation coefficient varies as a result of it. Moreover, we want to quantify until what mass our statistical validity on correlation stands.

We so decided to iteratively remove one planet from the dataset (either from the lowest set of masses or from the highest) and record the change in τ. The results can be seen in Figure 18. On the left panel, lower-mass planets are being removed iteratively, while on the right the bigger planetary masses are being removed. Noted on the plot are the values of the lowest/highest planetary masses being removed (always for the Z/M plot). The different coloured lines show how the p-values from the correlation test that vary with the iterative removal. The horizontal black line denotes when a p-value crosses the p=0.05 threshold.

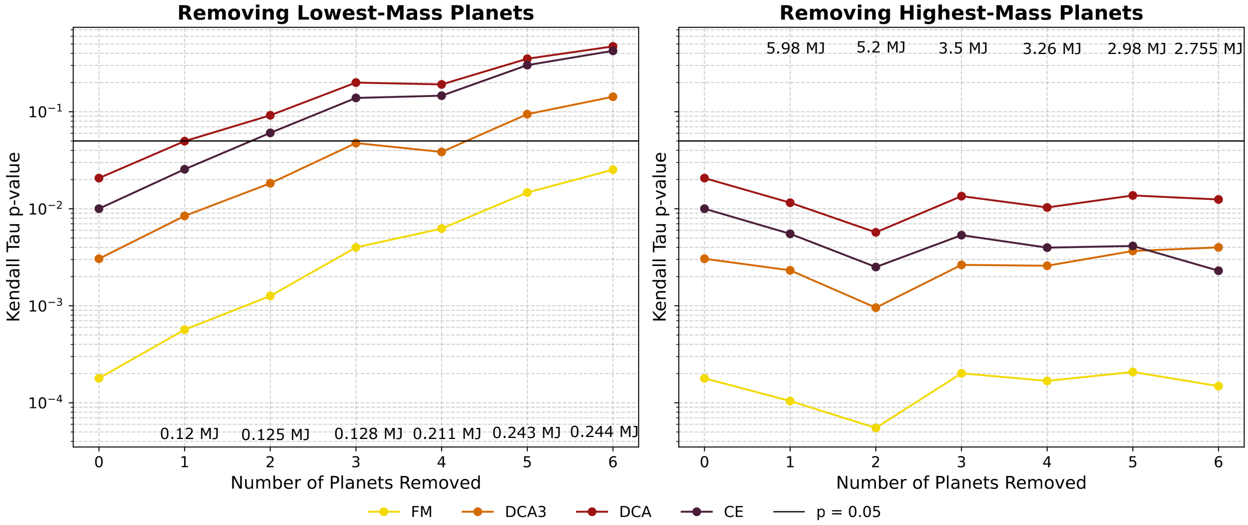

Figure 18: Exploratory analysis on the change in Kendall's τ correlation p-values as planet data (first for lower mass ones on the left, and then iteratively for the largest masses) for the Z/M plot.

As we can see, there is no remarkable change in p with the removal of higher masses. What affect the most statistical reliability are the lower mass planets that, after only three planets removed, the resulting p exceeds the decided acceptance threshold. The plot shows the importance of low-mass gas giants in keeping the given correlation statistically valid with $p<0.05$. This sensitivity is also reflected in the shape of the power-law decay used to describe our data (very steep at low x-axis location, and then flattening out). By removing just a few low-x points you lose most of the trend signal, whereas cutting high-x data barely moves the curve. Therefore, dense sampling of the low-mass end is essential to preserve statistical power and reliably measure the slope.

4.5.2 Best-fit analysis: Covariance

At the end of the previous section we inspected the changes in the correlation, but not the reliability of the best-fit curves. This will be examined in this section. We will continue to use the case study of the Z/M relationship.
To assess the power-law curves, we first computed the covariance of the slope α and intercept β parameters for each of the four models. Covariance inspects dependency between the two describers of the curve, α and β. Correlation on top of the covariance would give a comparable and quantifiable result in the $[-1, 1]$ range. To analyze parameter independence and stability, plotting (or reporting) the normalized correlation coefficient between slope and intercept can be explained with the following formula:

$$\rho = \frac{\mathrm{Cov}(\alpha, \log_{10}\beta)}{\sigma_\alpha\, \sigma_{\log_{10}\beta}}, \tag{16}$$

where ρ is the covariance normalized by their standard deviations. ρ can also be referred as the Pearson correlation coefficient between the two fitted parameters describing the power-law best fit (Pearson, 1896).

- If $|\rho| \ll 0.2$, (ideally 0) fit parameters are essentially independent.
- If $|\rho| \gtrsim 0.5$, they are strongly correlated, complicating separate interpretation (Freedman, 2009).

For the inspected relation of Z/M (but also valid for the power-laws of the other relations), the correlation of the covariance has been done for α and $\log -\beta$, as the formula shows their implications in the power-law function:

$$\log Z = \alpha \log M + \log\beta \tag{17}$$

which implies the following relation that matches the $y = bx^a$ power-law curve.

$$Z = \beta\, M^\alpha \tag{18}$$

The retrieval of such value will follow what has been done in Figure 17 in inspecting the full parameter error space with Bayesian analysis. Curves from the $10'000$ runs will create a distribution that can be interpreted. ρ distributions for all the four (FM, DCA3, DCA, and CE) models for the Z/M relation can be seen in Figure 19.
A shown in the plot, a moderately-weak coupling shown by the ρ value for all models (≥ 0.2) for the heavy element fraction against planetary mass relationship. A value of approximately 0.2 means that the covariance between the slope α and intercept $\log_{10}\beta$ is only 20% of the maximum possible covariance, $\sigma_\alpha\, \sigma_{\log_{10}\beta}$. In other terms, a positive shift to one of the two parameters by 1 sigma, results also in a change in the other parameter by 0.2 of its standard deviation (and this works symmetrically for both the intercept and slope).

This weak but positive coupling implies that α and β are not fully independent: small perturbations in α will nudge $\log_{10}\beta$ upward, and vice versa. Consequently, the fitted power–law curves can be sensitive to the sample size (despite increasing the sample size) and to the intrinsic scatter of planetary properties combined with a relatively narrow mass range. Such sensitivity can undermine the reliability of pure power–law extrapolations. While covariance of intercept and slope does not imply any unreliability of the best-fit, a great covariance value will spike up uncertainty in the predictions. Since in the majority of cases, covariance is not accounted in error propagation, results derived from

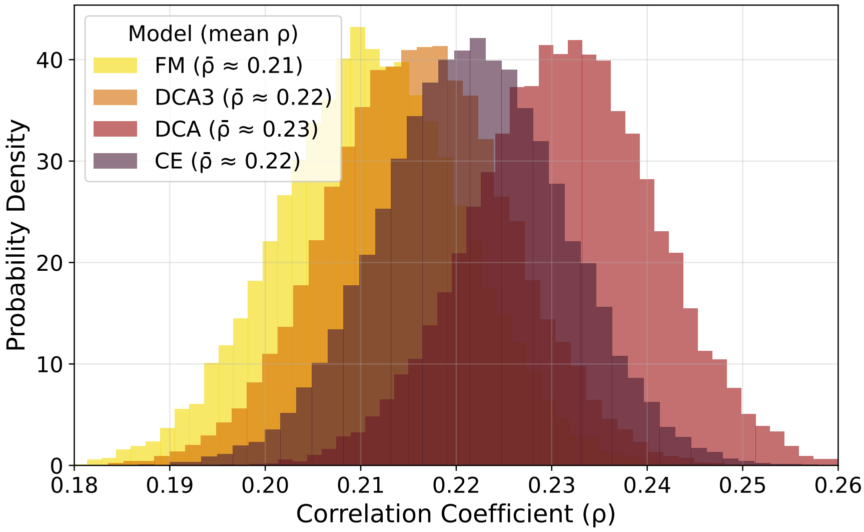

Figure 19: Bayesian analysis of the correlation coefficient ρ between the intercept and the slope (respectively β in log scale and α) for the log–log plot of heavy-element mass over planetary mass (Z/M), accounting for both $x-$ and y–errors. ρ have been calculated for all four models: FM, DCA3, DCA, and CE.

these fits can hold higher uncertainties that just from the data error themselves. In particular, the origin of a positive ρ value can be caused by outliers like HATS-17b and K2-114b, whose Z lie well off the best-fit line in log–log space (see Fig. 11 or Tables A.2-A.3), can disproportionately influence both α and β. Since the best-fit greatly underestimates these values (meaning that they sit way on top of it), a positive ρ to adjust for them is understandable. As a result, even modest shifts in either parameter may lead to substantial changes in the inferred Z trends, and the line (especially in a sample size of 43) becomes dependent on the kind of planets included (especially low-mass ones, as revealed by Fig. 18). Such test has been quickly checked for the retrieved data from other literature such as (Thorngren et al., 2016), that, even when statistical correlation has been met, the sample size of 46 planets (for instance differing from the dataset used in this study) results into the same problem when looking at drawing a linear model in log-log scale, with an even higher covariance reached compared to the one found in this study.

With such analysis we can conclude that the fitting parameters of the best-fit lines are not completely independent, and assuming complete independence without accounting the covariance joint uncertainty can cause overconfidence of the results, and underestimate the error.

4.5.3 Best-fit analysis: WLS

Since such analysis does not imply unreliability of the best-fit lines used, but just independence analysis for error estimation, an additional statistical test has been done. To inspect how well the power-law curves explain the data (with errors), a weighted least-squares (WLS) χ^2 test has been tested. Such test has been preferred over the R squared coefficient of determination because WLS does account for the data's individual uncertainties, and the distribution of residuals, other than only considering the mean points. For high uncertainty data like the one found in this study, we thought WSL was a more suited test.

In a weighted least–squares (WLS) framework, each residual is scaled by its measurement uncertainty, so the goodness-of-fit statistic is the sum of squared, error–weighted residuals:

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

where y_i are the observations, $f(x_i;\theta)$ the model prediction, and σ_i the 1σ uncertainty of each point. This quantity follows a chi-square distribution with $\nu = N - p$ degrees of freedom under the null hypothesis that the model is correct and the errors are Gaussian (Andrae et al., 2010).

For the assessment, we have computed the 'observed' statistics, and compared it to the “critical” value $\chi^2_{(\alpha;\nu)}$ drawn from the chi-square distribution with $\nu = N - p$ degrees of freedom at significance level p. At p=0.05 and $\nu = 41$ (43 samples explained by the two parameters α and β, so 43-2), one finds

$$\chi^2_{(0.05;\,41)} \approx 56.94.$$

The fit is accepted at the 5% level if and only if

$$\chi^2_{\mathrm{obs}} \;\le\; \chi^2_{(\alpha;\nu)}\,,$$

otherwise it is rejected. So if the observed χ^2 retrieved by the fits is lower than the critical value at the acceptance of p=0.05, we accept the null hypothesis of H0:"the model is correctly describing the data and its deviations, with the last one solely caused by measurement error drawn by a Gaussian distribution". Such analysis has been carried for both the HEM/M and Z_{pl}/M relations. Such relation could also be explained as the relationship between $\chi^2_{\mathrm{obs}}/\chi^2_{(\alpha;\nu)}$, where the result is the so called reduced χ^2.

- If $\chi^2_{\mathrm{red}} \approx 1$, the model is consistent with the data within the quoted uncertainties.
- If $\chi^2_{\mathrm{red}} > 1$, the scatter of the residuals is larger than expected, indicating a poor fit or underestimated errors.
- If $\chi^2_{\mathrm{red}} < 1$, the scatter of the residuals is smaller than expected, suggesting overestimated errors or potential overfitting (Bevington and Robinson, 2003).

In all cases, for all models and for both the relationships, the observed χ^2 is way greater than the critical value of 56.94 at the respective acceptance level and degrees of freedom. For the CE structure best fit curve (Z/M), the observed χ^2 is 580.4. The χ^2_{obs}, similarly to the other inspected trends, is the highest for CE, and then followed by DCA, DCA3, and finally FM. We see a value of 465.4 for DCA, 311.7 for DCA3, and the lowest FM with 243.7. This is also the order observed in other assessment done in this paper. Since in all cases χ^2 is above its critical term for the acceptance level, its χ^2_{red} will also be well above 1. The four models have a χ^2_{red} of 14.2, 11.4, 7.6, and 5.94 respectively for CE-DCA-DCA3-FM. We thus need to reject the null hypothesis and accept an alternative hypothesis.

We can then conclude that, even with the correlation is being met for all the three instances, the best fits do not explain the data within the quoted errors. Prediction-wise, this means any extrapolation or forward use of these power laws will carry systematically underestimated uncertainties and should be treated as unreliable unless the model or error budget is revised. Such scatter is not merely noise, but a reflection of the inherent diversity and complexity of planetary systems. Each planet is shaped by unique formation histories and evolutionary paths that ultimately define our estimated Z.

4.5.4 Break-up points

Since part of the problem could be the forcing of a linear power law model in the correlated data, multiple lines with different slopes (and respective break-up points) could explain better the data. Similarly to what Müller et al. (2024) did for the R/M relationship (also shown in Figure 3), we wanted to inspect if a M-R relationship of our data (proved to be somewhat present in Figure 10) could be explained by our Z/M. Similarly, we have applied an orthogonal distance regression (ODR) study to evaluate such option (Muggeo, 2003). This also consisted in taking into account the error bars of the data and so running a stochastic approach of it for $10'000$ tries. While the ODR is increasing the complexity of the model by a lot, a possible observation of the most likely break-up points could be linked with the results found in the radius-mass relationship (found in Müller et al. 2024, and in Chen and Kipping 2017a,b) to see any consistent result. For this reason, the number of breaks tried,and the location of the breaks in the x-axis coordinates, have not been fixed.

Results from the break-up point analysis were not convincing, thus they have not been visually implemented in this work. The ODR ran for the heavy element fraction over planetary mass relationship turned out to be best explained by the initial one-line model in log-log scale. When plotting back in normal scale, the results are different: two break-up points (three lines with different slopes) show the optimal conformity for data explanation. Even though also Müller et al. (2024) found two break-up points for explaining the radius-mass relationship, in this study the break-up points have a huge uncertainty range. While the locations of these points could be somewhat close to the results had in Müller's (0.146 and 0.671 M_J respectively versus the 0.014 and 0.4 M_J in Müller's study), the spread of these values due to the Bayesian method unveiled a +1 sigma value of +2.1 M_J. Such a value, together with the difference in mass range and number of data (with Müller having used a broader planetary range even outside gas giant exoplanets, and a greater sample size). We thus discard this idea as the error range is too wide, and because low mass break-up points were relying on too few data points to be consistent with other findings.

5 Discussion

5.1 Planetary trade-offs

Our retrievals reveal that the inferred heavy element fraction Z of a gas giant hinges on a delicate balance between several trade-offs that can alter heavy element quantities and limit structural estimations. Structural mixing, density redistribution, composition, energy transport, and opacity-driven thermal inflation are only a few of the identified factors affecting interior studies.

5.1.1 Density vs. Opacity in controlling Z estimation

When comparing a classic CE model with DC configurations (e.g., $m_{\mathrm{dilute}} = 0.25, 0.5, 0.75$), a Dilute Core steadily shifts heavy elements outward, reducing central density. A redistribution of heavy elements at higher radial coordinates means that higher densities (given higher mean molecular weight) will be reached at higher mass fractions. The contraction extent in outer layers will be greater, resulting in a reduced expected radius. Thus, it is expected that with a higher m_{dilute}, allowing heavy elements more to the surface, the contraction effect would increase. The evidence of such a statement can be proved by the average trend of underprediction of Z by DC models when compared with CE ones. The accuracy of this impact cannot be unfortunately be stated with clarity within the DC models at different m_{dilute}, where precision of the model and accuracy of the data hides possible scaled effects where DC0.25 predicts the closest to CE, and DC0.75 predicts the least amount of heavy elements in the retrieved Z. Moreover, by fixing m_{dilute} values, Z_{dilute} is the parameter changing, and this also creates degeneracies in the analysis of results, as this actively also reduces or increases internal density redistribution by shifting up and down. The magnitude of this effect can be negligible for the current observational error that we work with.

Yet this structural effect alone does not fully capture the diversity of inferred Z, especially when comparing DCA/DCA3 models with DC ones. Opacity, which governs how heat is trapped in the deep envelope and core, emerges as a predominant lever.

Envelope and atmospheric metallicities both raise mean molecular weight in upper layers, enhancing density. While such case is also what has been inspected in DC0.75 in purely density redistribution terms, there is another phenomenon coming into play that shifts the bulk metallicity up. Radiative opacity enhancement is caused by a change in composition in the uppermost region of the planet. Molecular structures have more ways to interact with photons than simpler structures like H_2 of He. For instance, molecules like water or methane (H_2O or CH_4) have wider possibilities with rotational and vibrational modes that enhance absorption, together with created molecular dipole dynamics (Freedman et al., 2008). In natural systems, condensation and droplets involving H_2O will affect opacity far more than molecular water. This is why atmospheric studies are important to be understood and constrained for metallicity determination. Other than molecular contribution, grain/cloud opacity is also contributing in greater magnituted, but this has not been modelled by CEPAM, as separate efforts are currently working on calibrating and studying atmospheric dynamics.

At higher equilibrium temperatures, partial thermal ionization supplies free electrons, allowing H- and atomic/ionic absorbers to dominate the opacity in the radiative outer atmosphere. This non-linear interplay means that adding metals can simultaneously compact and inflate the planet: higher opacities deepen the radiative gradient toward the adiabatic limit, trapping internal heat and inflating the radius (Valencia et al., 2013; Parmentier et al., 2015; Thorngren et al., 2019).

Consequently, two models with similar density profiles but different opacity prescriptions can return a Z that differs by more than the structural redistribution alone would suggest, hence the degeneracy of such studies. The effect of heavy elements in the outermost layer appears to be stronger that the changes applied to only density variations (DC-DCA differences seen in Figure 11. This can lead, as also concluded by Howard et al. (2025) in situations where a small amount of heavy elements is introduced in the outermost envelope, such strong opacity contribution may estimate higher radii than pure H-He (for FM structures). Composition does not only matter when talking about density changes, but also in the physical properties of each chemical species in contributing to the opacity effect. In the temperature regime inspected (below 1000 K), the opacity-driven effect is dominated by molecular absorbers (H_2O and CH_4 predominantly, thus the 'ices') and by collision-induced absorptions between hydrogen at helium (Freedman et al., 2008, 2014). This is shown by the highest possible estimate retrieved being FM models with heavy elements fully composed of ice, where a reduction due to reduced mean molecular weight, together with this enhancement in opacity, creates a positive feedback for inflated radii. The story might change in colder atmospheres, where water-ice or ammonia clouds can form, and additionally slow down heat escape. While this phenomenon has been seen in Jupiter with NASA's Juno mission (Guillot et al., 2020), this might be happening in planets with low equilibrium temperature like TOI-2180b and TOI-2010b (respectively Teq=348 and 400 K), no big difference between CE and FM models has been seen. Due to the observational bias of detecting exoplanets, and the filtering parameters applied in this study, no exoplanetary analogues such with equilibrium temperature similar to Jupiter and Saturn has been inspected. At T<1000K the situation might be increased by atomic iron and bound-free hydrite ions, but for the dataset used for this study that is not the case.

5.1.2 Density vs. Super Adiabaticity in controlling Z estimation

Introducing superadiabatic gradients under the Ledoux criterion further complicates the picture. Compositional boundaries, such as those created by dilute cores, can suppress large-scale convection, stabilizing layers and allowing enhanced heat retention at the highest steepness created by the chemical structure gradient. Since mean molecular weight gradients play a key role in the Ledoux criterion, composition directly affects the extent of convective inhibition: rocks, with higher molecular weight than ices, provide stronger stabilizing effects for a given abundance gradient. Consequently, cores enriched in rocks are expected to retain more heat than ice-dominated ones, due to a even higher mean molecular weight.

Counterintuitively, steeper internal gradients in heavy-element content (i.e., large differences in Z across the interior) actually enhance thermal retention. This effect becomes more pronounced in high-metallicity planets, where a large ratio of Z_{dilute}/Z_{atm} magnifies the ability of the structure to trap heat. In such cases, the temperature gradient dominates the thermal structure, while density redistribution plays a secondary role, as demonstrated in Figure 12.

However, these effects are likely to be short-lived. Over Giga year timescales, radiative diffusion and double-diffusive mixing are expected to erode the steepness of compositional gradients, diminishing the stabilizing influence and gradually shifting the interior structure toward a more adiabatic state. As mixing proceeds, the once-steep $Z_{(m)}$ curve may flatten or shift in mass coordinate, eventually resembling a fully mixed (FM) configuration (Leconte and Chabrier, 2012b). Although this study does not model that evolution explicitly (for simplicity and due to the uncertainty in compositional mixing timescales), it is important to note that realistic interiors likely evolve dynamically over time.
In this context, high values of the superadiabatic parameter f may still remain Ledoux-stable (i.e., not overturning), but they can support thermal profiles that become excessively steep, causing the planet to retain unrealistically high internal heat. This makes the extreme case of fully inhibited mixing (f=1) unlikely to persist on long evolutionary timescales. Nonetheless, such configurations may still be feasible for young planets, which have not yet undergone significant compositional mixing.

Finally, although some planets show relatively high bulk metallicities under DCA models with f=0.4, the long-term viability of these profiles must be viewed with caution. As the compositional gradient weakens over time, thermal structures will tend to re-equilibrate and approach adiabatic behaviour. While adiabatic models tend to underestimate heat retention, DCA models with fixed intermediate f values may overestimate it, especially if neglecting evolving internal mixing. Future studies should aim to quantify the efficiency and timescales of thermodynamic re-equilibration in steep thermal and compositional gradients to better assess the long-term realism of such interior structures.

5.2 Planetary relations

In this study, a major focus was to inspect the reliability of the planetary relations plotted (namely M_z/M, Z/M, and $(Z/Z_{star})/M$). We succeeded in proving statistical validity for all the tested correlations. While for the current dataset correlation assumptions stand, limited data and mass range might affect the resulting p-value might not be statistically valid anymore.

While correlation has been met, the reliability of best-fit curves is dubious. Not only, with the comparison of our curves with the ones from similar works (Thorngren et al., 2016; Howard et al., 2025), we see how small variability in methodology and dataset used, can shift on both α and β values.

In lower-mass planets ($M \lesssim 0.3\,M_J$), small uncertainties in radius or age translate into disproportionately large swings in Z, as the relative contribution of the core versus envelope mass becomes especially sensitive to model assumptions. This is explained with Figure 14 showing the reduction in ΔZ among the compared models. Such swings in expected Z with internal structures, could undermine variety in planetary characteristics and natures while having similar planetary masses. The presence of low-mass planets with such low estimated metallicities (namely HAT-P-12b, HAT-P-19b, and TOI-4406b), are the evidence of a high dependancy on different planetary formation settings, and ultimately the result of increased p-values in the tested correlations. Even with very distinct results and variability in said planets, expected Z remain low compared to their counterparts in similar mass. While such steep trend is somewhat captured by the negative power-law curve (for the case of the Z/M relationship in Figure 15)

Furthermore, results from the WLS show how the data is poorly explained and represented by the best-fit lines. It is then not recommended to use such models for prediction purposes in formation analysis for instance.

Given the high dependency of such relations on low mass planets, and given the broad categorisation of 'gas giants' that encompasses such a wide range of masses, it is clear to doubt such straight-up trends. In our case, we had lower-mass planets with inferred Z>0.5, and it is debatable on whether planets not predominantly composed of 'gas' (H-He) can be called gas giants, and not be remnants (if not an in between stage) of ice/Neptunian planets, that is the second biggest kind of planets according to the used nomenclature.

If we assume that the R-M relationship inspected by Müller et al. (2024) is mainly controlled by a composition shift (a first stage of small M-R made up by fully rocky planets, a second stage of accretion differentiation mimicking ice giants and a transition of Neptunian planets to gas giants, to then end up in a regime where the main composition is H-He, and the relationship is flatten out), the break-up trend, and the one inspected in this study with the bulk metallicity Z, should agree one to the other. The Z/M relationship does flatten out when moving at greater masses, and thus the power-law decay relies mainly on low mass-high metallicity planets (as shown in the mass removal inspection in Figure 18). If so, we take into account only planets above 0.4 M_J (second break-up recognised by Müller), the relationship should have a correlation coefficient of τ close to 0. Due to the limited data being used, we did not find in that range any valid result (a statistically-valid evidence of no correlation, possibly shown by $\tau \approx 0$, and yet with a valid $p - value > 0.05$). Such low p-value has not been reached to give validity to this hypothesis. Further statistical inspections should involve a larger dataset to unveil any tendencies at high masses.

Moreover, break-up point analysis gave poor results in constraining different planetary cases by their masses. The tests were not able to accurately constrain the break-up points found by Müller as further evidence of this planetary transition. The large range in the break-up points implies a dataset too small to be reliable on such assessments (for comparison, Müller et al. (2024) used 688 data points for WLS). If no p-value is being met even at a greater sample size could suggest that structure (and its change through planetary ages), stellar irradiation, additional heating sources and the natural differentiation of chemistry among different stellar cases, can bring additional obstacles that would disrupt such a simple linearity in studying only a two-parameter space, multivariate analysis might be needed.

Nonetheless, it is a topic worth inspecting in the future with a larger dataset (especially as ongoing surveys keep adding thousands of new exoplanets) since broader population studies may reveal cleaner, more robust trends.

5.2.1 Formation-driven structure

Interpreting these relations in a formation context suggests that the timing and efficiency of gas accretion play central roles in establishing a planet's present-day structure. Results from this study agree with what concluded by Miller and Fortney (2011); Thorngren et al. (2016); Venturini et al. (2017). Said tested relationships give us strengthening evidence of previous literature regarding planetary formation. For said relationships, we can conclude that:

- M_z/M shows that more massive planets tend to accrete more heavy elements in their interiors.
- Z/M shows that more massive planets tend to have a smaller heavy element bulk metallicity in their interior.
- Z_{star} normalisation for Z suggest the same trends, as for our dataset there is no seen correlation between Z_{star} and M.

All structural assumptions agree with these statements.

5.3 Atmospheric metallicity

Despite no increased correlation by Z/Z_{star}, it still holds interpretational power. We see how is all cases planets are enrichment in heavy elements compared to their host stars (Z/Z_{star} above 1). The enrichment appears to be decreasing to bigger M, meaning that bigger planets will be of similar composition to their stars. In such scenarios, forcing a Z_{atm}=3Z_{star} might unrealistically high and not physically explained.
The interesting case of planet HATS-74Ab further proves this point (in Figure 11), where the DCA3 structure tested was reducing the range of possible Z. Ruling out possibilities of an adjustable, less enriched atmospheric Z, led to a flattening of the DCA3 trend line in the Z/Z_{star} plotted space.
This reconnects to how gas accretion (or disk instability) can generate planets similar to the primordial gas composition. Once rapid gas accretion is triggered by either one of the two possible thought mechanisms, it pulls in huge amounts of nearly pristine nebular H/He. The heavier the planet, the larger that envelope fraction becomes, so the bulk composition approaches the star's. Push the mass high enough $\approx 13\,M_{\mathrm{J}}$ and it transitions into the brown dwarf regime (essentially showing how massive gas giants are 'failed stars', or 'very successful planets', depending of how one sees it) which naturally shares the parent cloud's composition.
Future work that couples precise atmospheric abundance measurements with interior dilution extents will be crucial to tracing planets back to their formation locations to understand how early disk chemistry imprints on present-day structures.

5.4 Where is the 'fuzziness' starting?

Low-mass giants such as TOI-4010d and HATS-72b, which we found required a higher extent in dilution ($m_{\mathrm{dilute}} \approx 0.75$) since a forced m_{dilute} of 0.5 gave no converged results since it required higher metallicities than Z=0.5 in order to match their observed radii. If the structure informs how much material is being accreted, and thus the planetary mass M , the bulk metallicity shows how the planets had no time in accreting large quantities of gasses during gas accretion phase. A higher extent of gas accretion would automatically lower the Z of the other elements. In classical core-accretion theory, such planets might have not reached a phase 3 of gas+material accretion. A smaller core has less gravitational reach and a shorter window to capture gas before runaway accretion (Pollack et al., 1996). Late formation and accretion in their own protoplanetary disk compared to more massive planets might explain poor development of such planets. For TOI-4010 d and

HATS-72b, it is unlikely their formation mechanism is given by disk instability. However, for cases like HAT-P-19b, it looks like it does not follow the overall trend, estimating lower metallicities than bigger planets, and showing a low normalised $Z/Z_\star$ marginally above 1. A mix of the two formation theories can explain such scatter in the data, together with the natural chemical differentiation normalised by the same assumed chemical ratios in all planets. According to core accretion, more massive planets develop compact cores and accrete voluminous envelopes, a sharp core–envelope boundary might be dilute over time in evolved planetary system, if not even explained by an FM structure (where compositional mixing took over the formation structure via double-diffusive convection (Leconte and Chabrier, 2012a,b; Debras et al., 2021)).

Even if the internal structure of the planet is not yet fully compositionally mixed, the plausibility of high mdilute values could be explained by high Z, low M planets. For large masses would not be feasible unless reasoned by any physical or chemical phenomena that would explain stable compositional gradients at relatively shallow depths. While in the case of this study it was helpful to play with different m_{dilute} values in order to see density distribution changes in a fixed variable space, possible conformations will be feasible in having a central core made fully of heavy elements (and so a Z_{dilute} likely being fixed at 1, and the dilution curve (and so the m_{dilute} variable in the function), will be the one adaptive according to the extent of heavy elements in the planet's interior. This could be tested in future analyses. A fully 'rocky' interior would not only be explained by a core-accreation perspective, footprint of an enlarged planetesimal into the protoplanet, but also explainable by gravity-driven sinking of heavier material (Vazan et al., 2018). Furthermore, for this reason, it is safe to say that the interior will be dominated by silicates and metals more than volatiles like water and carbon species, consistent with models of young giant planets developing "fuzzy" cores (Debras and Chabrier, 2019). Hopefully, this study gathers all the possibilities and informs future directions in planetary interior modelling and chemistry, in order to also increase their intricacies for better realism by still keeping a clear interpretation of results.

6 Conclusion

The aim of this work was to quantify the bulk metallicity (Z) for different gas–giants, and assess how inferred Z changes with the assumed internal structure. We applied the evolutionary model CEPAM with a root–finding retrieval across several structures (CE, DC, DCA/DCA3, FM), and evaluated statistical robustness via Bayesian correlation/covariance analyses and weighted least squares. Although dilute–core concepts are established for Jupiter and Saturn, they remain largely untested across the broader exoplanet population; this study aims at bridging that gap and sets the stage for forthcoming interior modelling studies.

- **Method check:** By comparing our results with Howard et al. (2025), we positively accept our approach as reliable. Both methods concluded with the same results in terms of the relationships identified.
- **Structure matters for Z:** Switching between CE, DC/DCA, and FM redistributes metals and alters the total Z needed to match observed radii. As a consequence, structural and chemical information is crucial.
- **Low-mass giants are vulnerable to assumptions:** For low-mass planets, a higher change in Z in varying structural assumptions is being recorded. An forced 1:1 ratio of ices and rock can additionally contribute to the scatter of a dataset that by nature will show dissimilarities in planetary environments.
- **Opacity vs. density:** The shift seen in the models can be simplified by two main trade-offs; Molecular opacity can both compact and inflate planets, dominating in magnitude over pure density redistribution and creating degeneracies.
- **Heat transport depends on gradients:** Ledoux-stable layers (DCA) retain heat, but large superadiabatic factors are likely transient as double–diffusive mixing erodes gradients over time.
- **Population trends exist but are weak predictors:** M_z/M, Z/M, and $(Z/Z_\star)/M$ correlate, yet WLS reveals poor explanatory power best-fit curves are unreliable for prediction. Although data scatter is adverse to statistics, it does show the varying nature in exoplanets' characteristics and internal compositions.
- **Formation clues:** Massive planets accrete more metals in absolute terms but are less metal-rich fractionally; enrichment above stellar declines with mass. Forcing $Z_{\text{atm}} = 3Z_\star$ is often unjustified. A great scatter in planetary data glances at possible disk instability scenarios.
- **Key degeneracies:** Density redistribution, opacity, and superadiabaticity can mimic each other's effects. Joint interior–atmosphere constraints are essential.

Future improvements in support of this study could be made in increasing the sample size, especially for low-mass giants, to try once again break-up points analysis, and to see whether predictory best-fit curves become more reliable. Coupling interior models to measured atmospheric abundances/opacities (including clouds/grains) instead of fixing Z_{atm} would come with observational studies, together with improvements in atmospheric modelling.

Looking forward, *PLATO* will deliver precise ages and radii, crucial for shrinking Z uncertainties in the most sensitive (low-mass) regime, while *Ariel* will map atmospheric compositions and cloud properties, breaking opacity/structure degeneracies, and helping interior modelers at improving internal structure constrains. Together, these missions will enable self-consistent interior–atmosphere modeling, reveal when and where "fuzzy" cores emerge, and ultimately link accurate structures to their formation pathways with far greater fidelity. In such case, the question will no longer be whether cores are diluted, but solely on how and why they became that way.

7 Acknowledgements

I would like to express my sincere gratitude to the University of Zurich for giving me the opportunity to complete this Master's program and pursue my academic journey.

A big thank you goes to my parents for their unwavering support and for providing me with the opportunity to study and follow an academic path. My father and sister have been a constant source of inspiration in choosing a scientific direction.

I am deeply grateful to the Astrophysics research group for their warm welcome and collegial atmosphere throughout the course of my thesis. Pool matches, 5-a-side, and Connections will be remembered as great moments shared with the group. In particular, I would like to thank Ravit for believing in me and giving me the opportunity to work within her group. A special thanks to Saburo, who has been an exceptional supervisor, always patient, supportive, and generous with his time. Beyond being a mentor, he has also become a great friend, and for that I am truly thankful.

I would also like to acknowledge Juan-Diego and Mike, who initially sparked my passion for planetary sciences and encouraged me to pursue this field academically.

Many thanks to Stefano and Maarten for their support on the statistical side of this work. I am also grateful to Simon for his insights, suggestions, and helpful feedback during the development of this thesis. Additional thanks to Maarten for his assistance in navigating the administrative aspects of the Master's thesis process.

I sincerely thank Dr. Maria Santos that agreed to be the Faculty Member for this thesis, my external evaluator reading this, as well as all other readers of this thesis. It is a privilege to share one's research and passion with the scientific community, and I am deeply appreciative of that opportunity.

Lastly, I would like to extend a general thank you to all the people I met during my time in Zurich. Their presence made this experience even more enriching. The importance of psychological well-being, both academically and personally, cannot be overlooked, and it was ultimately a key part of what made this journey so rewarding.

This research was conducted without any form of external financial support (direct or indirect) and was not influenced by institutional, industrial, or economic interests. It was carried out solely in the spirit of academic inquiry, driven by genuine scientific curiosity and personal dedication.

References

Rene Andrae, Tim Schulze-Hartung, and Peter Melchior. Dos and don'ts of reduced chi-squared, 2010.

Martin Asplund, Nicolas Grevesse, A. Jacques Sauval, and Pat Scott. The Chemical Composition of the Sun. , 47(1):481–522, September 2009. doi: 10.1146/annurev.astro.46.060407.145222.

Philip R. Bevington and D. Keith Robinson. *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, 3 edition, 2003.

S. Bloot, Y. Miguel, M. Bazot, and S. Howard. Exoplanet interior retrievals: core masses and metallicities from atmospheric abundances. , 523(4):6282–6292, August 2023. doi: 10.1093/mnras/stad1873.

Richard P Brent. *Algorithms for minimization without derivatives*. Courier Corporation, 2013.

Gilles Chabrier and Florian Debras. A New Equation of State for Dense Hydrogen-Helium Mixtures. II. Taking into Account Hydrogen-Helium Interactions. , 917(1):4, August 2021. doi: 10.3847/1538-4357/abfc48.

Jingjing Chen and David Kipping. Probabilistic Forecasting of the Masses and Radii of Other Worlds. , 834(1):17, January 2017a. doi: 10.3847/1538-4357/834/1/17.

Jingjing Chen and David Kipping. Forecaster: Mass and radii of planets predictor. Astrophysics Source Code Library, record ascl:1701.007, January 2017b.

Jeffrey N. Cuzzi, Robert C. Hogan, Julie M. Paque, and Anthony R. Dobrovolskis. Size-selective Concentration of Chondrules and Other Small Particles in Protoplanetary Nebula Turbulence. , 546(1):496–508, January 2001. doi: 10.1086/318233.

F. Debras and G. Chabrier. New models of Jupiter in the context of Juno and Galileo. In *AGU Fall Meeting Abstracts*, volume 2019, pages P24B–05, December 2019.

Florian Debras, Gilles Chabrier, and David J. Stevenson. Superadiabaticity in Jupiter and Giant Planet Interiors. , 913(2):L21, June 2021. doi: 10.3847/2041-8213/abfdcc.

C. Dominik and A. G. G. M. Tielens. The Physics of Dust Coagulation and the Structure of Dust Aggregates in Space. , 480(2):647–673, May 1997. doi: 10.1086/303996.

Daniel Foreman-Mackey, David W. Hogg, Dustin Lang, and Jonathan Goodman. emcee: The mcmc hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):306–312, 2013. doi: 10.1086/670067.

Jonathan J. Fortney, Kevin Zahnle, Isabelle Baraffe, Adam Burrows, Sarah E. Dodson-Robinson, Gilles Chabrier, Tristan Guillot, Ravit Helled, Franck Hersant, William B. Hubbard, Jack J. Lissauer, and Mark S. Marley. Planetary Formation and Evolution Revealed with a Saturn Entry Probe: The Importance of Noble Gases. *arXiv e-prints*, art. arXiv:0911.3699, November 2009. doi: 10.48550/arXiv.0911.3699.

David A Freedman. *Statistical models: theory and practice*. cambridge university press, 2009.

Richard S. Freedman, Mark S. Marley, and Katharina Lodders. Line and Mean Opacities for Ultracool Dwarfs and Extrasolar Planets. , 174(2):504–513, February 2008. doi: 10.1086/521793.

Richard S. Freedman, Jacob Lustig-Yaeger, Jonathan J. Fortney, Roxana E. Lupu, Mark S. Marley, and Katharina Lodders. Gaseous Mean Opacities for Giant Planet and Ultracool Dwarf Atmospheres over a Range of Metallicities and Temperatures. , 214(2):25, October 2014. doi: 10.1088/0067-0049/214/2/25.

François Fressin, Guillermo Torres, David Charbonneau, Stephen T. Bryson, Jessie Christiansen, Courtney D. Dressing, Jon M. Jenkins, Lucianne M. Walkowicz, and Natalie M. Batalha. The False Positive Rate of Kepler and the Occurrence of Planets. , 766(2):81, April 2013. doi: 10.1088/0004-637X/766/2/81.

T. Guillot and P. Morel. CEPAM: a code for modeling the interiors of giant planets. , 109:109–123, January 1995.

Tristan Guillot. THE INTERIORS OF GIANT PLANETS: Models and Outstanding Questions. *Annual Review of Earth and Planetary Sciences*, 33:493–530, January 2005. doi: 10.1146/annurev.earth.32.101802.120325.

Tristan Guillot, Cheng Li, Scott Bolton, Shannon Brown, Andrew Ingersoll, Michael Janssen, Steven Levin, Jonathan Lunine, Glenn Orton, Paul Steffes, and David Stevenson. Storms and the Depletion of Ammonia in Jupiter: II. Explaining the Juno Observations. *arXiv e-prints*, art. arXiv:2012.14316, December 2020. doi: 10.48550/arXiv.2012.14316.

Ravit Helled and Saburo Howard. Giant planet interiors and atmospheres. *arXiv e-prints*, art. arXiv:2407.05853, July 2024. doi: 10.48550/arXiv.2407.05853.

Ravit Helled and Alessandro Morbidelli. Planet Formation. In Nikku Madhusudhan, editor, *ExoFrontiers; Big Questions in Exoplanetary Science*, pages 12–1. 2021. doi: 10.1088/2514-3433/abfa8fch12.

S. Howard, T. Guillot, M. Bazot, Y. Miguel, D. J. Stevenson, E. Galanti, Y. Kaspi, W. B. Hubbard, B. Militzer, R. Helled, N. Nettelmann, B. Idini, and S. Bolton. Jupiter's interior from Juno: Equation-of-state uncertainties and dilute core extent. , 672:A33, April 2023. doi: 10.1051/0004-6361/202245625.

S. Howard, R. Helled, and S. Müller. Giant exoplanet composition: The impact of the hydrogen–helium equation of state and interior structure. , 693:L7, January 2025. doi: 10.1051/0004-6361/202452783.

Anders Johansen, Jeffrey S. Oishi, Mordecai-Mark Mac Low, Hubert Klahr, Thomas Henning, and Andrew Youdin. Rapid planetesimal formation in turbulent circumstellar disks. , 448(7157):1022–1025, August 2007. doi: 10.1038/nature06086.

Maurice G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938. doi: 10.2307/2332226.

J. Leconte and G. Chabrier. Layered double diffusive convection: From Earth oceans to giant planet interiors. In S. Boissier, P. de Laverny, N. Nardetto, R. Samadi, D. Valls-Gabaud, and H. Wozniak, editors, *SF2A-2012: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, pages 237–240, December 2012a.

J. Leconte and G. Chabrier. A new vision of giant planet interiors: Impact of double diffusive convection. , 540:A20, April 2012b. doi: 10.1051/0004-6361/201117595.

Jeffrey M. McMahon, Miguel A. Morales, Carlo Pierleoni, and David M. Ceperley. The properties of hydrogen and helium under extreme conditions. *Reviews of Modern Physics*, 84(4):1607–1653, October 2012. doi: 10.1103/RevModPhys.84.1607.

Yamila Miguel and Allona Vazan. Interior and Evolution of the Giant Planets. *Remote Sensing*, 15 (3):681, January 2023. doi: 10.3390/rs15030681.

Neil Miller and Jonathan J. Fortney. The Heavy-element Masses of Extrasolar Giant Planets, Revealed. , 736(2):L29, August 2011. doi: 10.1088/2041-8205/736/2/L29.

Alessandro Morbidelli and Sean N. Raymond. Challenges in planet formation. *Journal of Geophysical Research (Planets)*, 121(10):1962–1980, October 2016. doi: 10.1002/2016JE005088.

V. M. R. Muggeo. Estimating regression models with unknown break-points. *Statistics in Medicine*, 22(19):3055–3071, 2003. doi: 10.1002/sim.1545.

Simon Müller and Ravit Helled. Warm giant exoplanet characterisation: current state, challenges and outlook. *Frontiers in Astronomy and Space Sciences*, 10:1179000, May 2023. doi: 10.3389/fspas.2023.1179000.

Simon Müller and Ravit Helled. The bulk metallicity of giant planets around M stars. , 693:L4, January 2025. doi: 10.1051/0004-6361/202452442.

Simon Müller, Maya Ben-Yami, and Ravit Helled. Theoretical versus Observational Uncertainties: Composition of Giant Exoplanets. , 903(2):147, November 2020. doi: 10.3847/1538-4357/abba19.

Simon Müller, Jana Baron, Ravit Helled, François Bouchy, and Léna Parc. The mass-radius relation of exoplanets revisited. , 686:A296, June 2024. doi: 10.1051/0004-6361/202348690.

Vivien Parmentier, Tristan Guillot, Jonathan J. Fortney, and Mark S. Marley. A non-grey analytical model for irradiated atmospheres. II. Analytical vs. numerical solutions. , 574:A35, February 2015. doi: 10.1051/0004-6361/201323127.

Karl Pearson. Mathematical contributions to the theory of evolution. iii. regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A*, 187:253–318, 1896. doi: 10.1098/rsta.1896.0007.

James B. Pollack, Olenka Hubickyj, Peter Bodenheimer, Jack J. Lissauer, Morris Podolak, and Yuval Greenzweig. Formation of the Giant Planets by Concurrent Accretion of Solids and Gas. , 124(1):62–85, November 1996. doi: 10.1006/icar.1996.0190.

Johanna K. Teske, Daniel Thorngren, Jonathan J. Fortney, Natalie Hinkel, and John M. Brewer. Do Metal-rich Stars Make Metal-rich Planets? New Insights on Giant Planet Formation from Host Star Abundances. , 158(6):239, December 2019. doi: 10.3847/1538-3881/ab4f79.

Daniel Thorngren, Peter Gao, and Jonathan Fortney. New Giant Planet Physics from Statistical Population Models. In *AAS/Division for Extreme Solar Systems Abstracts*, volume 51 of *AAS/Division for Extreme Solar Systems Abstracts*, page 321.06, August 2019.

Daniel P. Thorngren, Jonathan J. Fortney, Ruth A. Murray-Clay, and Eric D. Lopez. The Mass-Metallicity Relation for Giant Planets. , 831(1):64, November 2016. doi: 10.3847/0004-637X/831/1/64.

Diana Valencia, Tristan Guillot, Vivien Parmentier, and Richard S. Freedman. Bulk Composition of GJ 1214b and Other Sub-Neptune Exoplanets. , 775(1):10, September 2013. doi: 10.1088/0004-637X/775/1/10.

E. A. van Dijk and Y. Miguel. Retrieving interior properties of hot Jupiters with Love numbers and atmospheric measurements. , 540(2):1544–1559, June 2025. doi: 10.1093/mnras/staf814.

Allona Vazan, Ravit Helled, and Tristan Guillot. Jupiter's evolution with primordial composition gradients. , 610:L14, February 2018. doi: 10.1051/0004-6361/201732522.

Julia Venturini, Ravit Helled, Yann Alibert, and Willy Benz. The formation of gas-rich planets. In *EGU General Assembly Conference Abstracts*, EGU General Assembly Conference Abstracts, page 15909, April 2017.

Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.

S. M. Wahl, W. B. Hubbard, B. Militzer, T. Guillot, Y. Miguel, N. Movshovitz, Y. Kaspi, R. Helled, D. Reese, E. Galanti, S. Levin, J. E. Connerney, and S. J. Bolton. Comparing Jupiter interior structure models to Juno gravity measurements and the role of a dilute core. , 44(10):4649–4659, May 2017. doi: 10.1002/2017GL073160.

8 Appendix

In this section, all the tables with the observational data and the retrieved Z values retrieved can be found, together with the grid-testing validation, and figures not used in the main text. I hope this could help for future reproducibility and for future statistical studies that might have not been covered in this analysis. Scientific progress continues thanks to the joint work of a community, where each researcher builds on the discoveries of others, critiques and refines hypotheses through peer review, and shares data and methodologies openly.

8.1 Planetary data and Z outputs

Planet	M (M_J)	R (R_J)	Age (Myr)	T (K)	Z_{st} (%)
TOI-4010d*	0.120 ± 0.010	0.551 ± 0.013	3000–9200	650	0.036 ± 0.070
HATS-72b*	0.125 ± 0.004	0.722 ± 0.003	11710–12410	739	0.019 ± 0.014
WASP-156b*	0.128 ± 0.010	0.510 ± 0.020	2400–10400	970	0.027 ± 0.120
HAT-P-12b	0.211 ± 0.057	0.959 ± 0.025	500–4500	963	0.008 ± 0.050
TOI-3629b	0.243 ± 0.082	0.740 ± 0.014	5000–13800	711	0.055 ± 0.093
HD332231b	0.244 ± 0.086	0.867 ± 0.027	2400–6800	876	0.017 ± 0.059
K2-329b*	0.260 ± 0.022	0.774 ± 0.026	500–4000	650	0.019 ± 0.068
WASP-69b*	0.260 ± 0.017	1.057 ± 0.047	500–3000	963	0.021 ± 0.077
HAT-P-19b	0.277 ± 0.060	1.008 ± 0.014	3200–11200	981	0.023 ± 0.070
TOI-4406b	0.300 ± 0.100	1.000 ± 0.020	2200–3600	904	0.019 ± 0.050
TOI-1268b*	0.303 ± 0.026	0.812 ± 0.054	110–380	919	0.035 ± 0.060
K2-287b	0.315 ± 0.086	0.847 ± 0.013	3700–5700	804	0.024 ± 0.040
HATS-49b	0.353 ± 0.092	0.765 ± 0.013	8500–11900	835	0.025 ± 0.053
WASP-132b*	0.410 ± 0.030	0.870 ± 0.030	500–2500	763	0.026 ± 0.130
TOI-5344b	0.412 ± 0.097	0.946 ± 0.021	5100–13800	689	0.041 ± 0.088
TOI-201b	0.420 ± 0.095	1.008 ± 0.014	380–1330	759	0.027 ± 0.036
HATS-75b	0.491 ± 0.079	0.884 ± 0.013	10600–13800	772	0.051 ± 0.400
HAT-P-17b	0.534 ± 0.034	1.010 ± 0.029	4500–11100	792	0.015 ± 0.080
WASP-84b	0.687 ± 0.048	0.976 ± 0.025	500–3700	833	0.019 ± 0.120
TOI-3714b	0.689 ± 0.044	0.944 ± 0.020	6000–13800	764	0.038 ± 0.086
HAT-P-54b	0.760 ± 0.042	0.944 ± 0.028	1800–8200	818	0.011 ± 0.080
K2-290c*	0.774 ± 0.047	1.006 ± 0.050	3200–5600	676	0.013 ± 0.100
TOI-1478b*	0.851 ± 0.052	1.060 ± 0.040	5200–12200	918	0.018 ± 0.069
K2-140b*	0.930 ± 0.040	1.210 ± 0.090	5300–13200	962	0.019 ± 0.100
TOI-1811b	0.972 ± 0.079	0.994 ± 0.025	1900–10800	962	0.031 ± 77.000
WASP-130b*	1.230 ± 0.040	0.890 ± 0.030	400–7900	833	0.028 ± 0.100
TOI-2010b	1.286 ± 0.044	1.054 ± 0.027	600–4200	400	0.023 ± 0.055
TOI-5542b*	1.320 ± 0.100	1.009 ± 0.036	7200–12900	441	0.009 ± 0.080
HATS-17b*	1.338 ± 0.065	0.777 ± 0.056	800–3400	814	0.031 ± 0.030
HATS-74Ab	1.460 ± 0.096	1.032 ± 0.021	5900–13800	895	0.050 ± 0.027
Kepler-117c*	1.840 ± 0.180	1.101 ± 0.035	3900–6700	704	0.014 ± 0.100
TIC237913194b*	1.942 ± 0.092	1.117 ± 0.051	4000–7400	974	0.021 ± 0.050
HAT-P-15b*	1.946 ± 0.066	1.072 ± 0.043	5200–9300	904	0.026 ± 0.080
TOI-4515b*	2.005 ± 0.052	1.086 ± 0.039	1000–1400	705	0.017 ± 0.030
K2-114b*	2.100 ± 0.120	0.932 ± 0.031	2700–11700	701	0.040 ± 0.037
TOI-4127b*	2.300 ± 0.110	1.096 ± 0.036	2700–6900	605	0.021 ± 0.120
HATS-76b	2.629 ± 0.034	1.079 ± 0.031	600–13300	940	0.032 ± 0.057
TOI-2180b	2.755 ± 0.030	1.010 ± 0.021	6800–9600	348	0.028 ± 0.057
NGTS-20b*	2.980 ± 0.160	1.070 ± 0.040	1400–6800	688	0.022 ± 0.080
TOI-5153b	3.260 ± 0.050	1.060 ± 0.040	4400–6400	906	0.020 ± 0.080
TOI-2589b	3.500 ± 0.029	1.080 ± 0.030	9000–13000	582	0.020 ± 0.040
WASP-162b*	5.200 ± 0.200	1.000 ± 0.050	10620–13800	910	0.029 ± 0.130
TOI-2338b*	5.980 ± 0.200	1.000 ± 0.020	5000–9000	799	0.026 ± 0.040

Table A.1: List of exoplanets used in the study. Planetary parameters are taken from Howard et al. (2025) and are present in the publicly available planet database on `dace.unige.ch`. Planet names followed by an asterisk (*) were introduced at a later stage; those without it are the original exoplanets shown in Figures 8 and 10.

Planet Name	CE	FM	DC0.25	DC0.5	DC0.75
TOI-4010d*	0.724 ± 0.011	0.672 ± 0.026			
HATS-72b*	0.400 ± 0.008	0.423 ± 0.004			
WASP-156b*	0.736 ± 0.010	0.767 ± 0.045			
HAT-P-12b	0.129 ± 0.050	0.294 ± 0.073	0.119 ± 0.048	0.124 ± 0.047	0.124 ± 0.047
TOI-3629b	0.397 ± 0.026	0.427 ± 0.025		0.389 ± 0.028	0.372 ± 0.027
HD332231b	0.222 ± 0.048	0.311 ± 0.045	0.195 ± 0.035	0.214 ± 0.045	0.209 ± 0.043
K2-329b*	0.318 ± 0.024	0.450 ± 0.044			
WASP-69b*	0.060 ± 0.044	0.199 ± 0.057			
HAT-P-19b	0.021 ± 0.014	0.112 ± 0.025	0.019 ± 0.014	0.021 ± 0.014	0.021 ± 0.015
TOI-4406b	0.044 ± 0.025	0.152 ± 0.034	0.038 ± 0.025	0.043 ± 0.025	0.043 ± 0.024
TOI-1268b*	0.311 ± 0.029	0.511 ± 0.026			
K2-287b	0.255 ± 0.024	0.328 ± 0.020	0.229 ± 0.015	0.246 ± 0.024	0.240 ± 0.023
HATS-49b	0.392 ± 0.025	0.416 ± 0.018		0.385 ± 0.027	0.367 ± 0.024
WASP-132b*	0.268 ± 0.046	0.376 ± 0.044			
TOI-5344b	0.067 ± 0.035	0.134 ± 0.040	0.064 ± 0.034	0.065 ± 0.033	0.064 ± 0.034
TOI-201b	0.086 ± 0.030	0.209 ± 0.037	0.053 ± 0.031	0.084 ± 0.029	0.082 ± 0.029
HATS-75b	0.188 ± 0.025	0.288 ± 0.019	0.184 ± 0.026	0.182 ± 0.025	0.179 ± 0.024
HAT-P-17b	0.040 ± 0.029	0.108 ± 0.039	0.041 ± 0.028	0.038 ± 0.028	0.040 ± 0.028
WASP-84b	0.142 ± 0.047	0.222 ± 0.064	0.124 ± 0.051	0.137 ± 0.046	0.135 ± 0.044
TOI-3714b	0.124 ± 0.035	0.155 ± 0.051	0.121 ± 0.038	0.121 ± 0.035	0.120 ± 0.035
HAT-P-54b	0.168 ± 0.050	0.217 ± 0.054	0.154 ± 0.048	0.162 ± 0.049	0.158 ± 0.048
K2-290c*	0.093 ± 0.060	0.159 ± 0.063			
TOI-1478b*	0.048 ± 0.034	0.111 ± 0.051			
K2-140b*	0.058 ± 0.045	0.112 ± 0.060			
TOI-1811b	0.112 ± 0.044	0.194 ± 0.049	0.099 ± 0.046	0.110 ± 0.042	0.108 ± 0.041
WASP-130b*	0.291 ± 0.039	0.362 ± 0.052			
TOI-2010b	0.096 ± 0.064	0.144 ± 0.067	0.037 ± 0.028	0.040 ± 0.031	0.062 ± 0.038
TOI-5542b*	0.087 ± 0.068	0.135 ± 0.080			
HATS-17b*	0.318 ± 0.025	0.482 ± 0.047			
HATS-74Ab	0.063 ± 0.031	0.172 ± 0.034	0.056 ± 0.030	0.064 ± 0.031	0.063 ± 0.030
Kepler-117c*	0.035 ± 0.028	0.082 ± 0.040			
TIC237913194b*	0.056 ± 0.045	0.105 ± 0.051			
HAT-P-15b*	0.066 ± 0.050	0.115 ± 0.053			
TOI-4515b*	0.067 ± 0.057	0.153 ± 0.058			
K2-114b*	0.256 ± 0.052	0.285 ± 0.045			
TOI-4127b*	0.044 ± 0.036	0.091 ± 0.043			
HATS-76b	0.066 ± 0.044	0.116 ± 0.053	0.046 ± 0.030	0.028 ± 0.014	0.060 ± 0.038
TOI-2180b	0.132 ± 0.036	0.163 ± 0.043	0.096 ± 0.029	0.105 ± 0.030	0.098 ± 0.030
NGTS-20b*	0.079 ± 0.059	0.137 ± 0.103			
TOI-5153b	0.103 ± 0.063	0.155 ± 0.044	0.067 ± 0.040	0.080 ± 0.045	0.079 ± 0.044
TOI-2589b	0.047 ± 0.035	0.069 ± 0.038	0.034 ± 0.023	0.038 ± 0.026	0.038 ± 0.026
WASP-162b*	0.186 ± 0.094	0.187 ± 0.048			
TOI-2338b*	0.149 ± 0.055	0.182 ± 0.014			

Table A.2: Retrieved results for FM, CE, and DC models for all the 43 exoplanets. This table contains the results shown in 8 for the original 21 planets with the addition of the remaining 23. DC models have been ran only for the original 21, since no further statistics has ben applied to it

Planet Name	DCA	DCA3	CEA	SMA
TOI-4010d*	0.631 ± 0.002	0.692 ± 0.012		
HATS-72b*	0.380 ± 0.006	0.405 ± 0.005		
WASP-156b*	0.720 ± 0.004	0.720 ± 0.050		
HAT-P-12b	0.123 ± 0.047	0.163 ± 0.049	0.129 ± 0.050	0.138 ± 0.050
TOI-3629b	0.427 ± 0.029	0.436 ± 0.027	0.453 ± 0.026	0.420 ± 0.001
HD332231b	0.210 ± 0.046	0.267 ± 0.045	0.222 ± 0.049	0.144 ± 0.003
K2-329b*	0.371 ± 0.047	0.375 ± 0.026		
WASP-69b*	0.085 ± 0.046	0.131 ± 0.046		
HAT-P-19b	0.046 ± 0.017	0.094 ± 0.017	0.048 ± 0.017	0.046 ± 0.021
TOI-4406b	0.049 ± 0.021	0.111 ± 0.026	0.051 ± 0.021	0.201 ± 0.005
TOI-1268b*	0.428 ± 0.044	0.421 ± 0.031		
K2-287b	0.270 ± 0.023	0.303 ± 0.023	0.289 ± 0.024	0.221 ± 0.040
HATS-49b	0.406 ± 0.032	0.421 ± 0.024	0.419 ± 0.025	0.045 ± 0.049
WASP-132b*	0.300 ± 0.053	0.332 ± 0.046		
TOI-5344b	0.105 ± 0.033	0.156 ± 0.023	0.108 ± 0.035	0.104 ± 0.041
TOI-201b	0.124 ± 0.029	0.169 ± 0.029	0.130 ± 0.031	0.130 ± 0.030
HATS-75b	0.220 ± 0.023	0.241 ± 0.021	0.233 ± 0.026	0.233 ± 0.060
HAT-P-17b	0.049 ± 0.026	0.083 ± 0.027	0.050 ± 0.026	0.088 ± 0.055
WASP-84b	0.135 ± 0.045	0.194 ± 0.044	0.140 ± 0.048	0.177 ± 0.032
TOI-3714b	0.153 ± 0.033	0.181 ± 0.030	0.161 ± 0.036	0.156 ± 0.067
HAT-P-54b	0.161 ± 0.049	0.194 ± 0.048	0.166 ± 0.051	0.141 ± 0.070
K2-290c*	0.095 ± 0.056	0.126 ± 0.058		
TOI-1478b*	0.057 ± 0.034	0.096 ± 0.032		
K2-140b*	0.066 ± 0.036	0.105 ± 0.040		
TOI-1811b	0.141 ± 0.042	0.172 ± 0.038	0.145 ± 0.044	0.144 ± 0.044
WASP-130b*	0.323 ± 0.051	0.349 ± 0.061		
TOI-2010b	0.093 ± 0.065	0.113 ± 0.031	0.074 ± 0.031	0.552 ± 0.049
TOI-5542b*	0.099 ± 0.088	0.184 ± 0.160		
HATS-17b*	0.428 ± 0.042	0.404 ± 0.031		
HATS-74Ab	0.103 ± 0.041	0.169 ± 0.014	0.106 ± 0.031	0.124 ± 0.058
Kepler-117c*	0.082 ± 0.018	0.076 ± 0.026		
TIC237913194b*	0.072 ± 0.039	0.109 ± 0.036		
HAT-P-15b*	0.086 ± 0.042	0.125 ± 0.036		
TOI-4515b*	0.101 ± 0.055	0.123 ± 0.042		
K2-114b*	0.255 ± 0.045	0.275 ± 0.044		
TOI-4127b*	0.064 ± 0.031	0.121 ± 0.030		
HATS-76b	0.107 ± 0.051	0.151 ± 0.042	0.097 ± 0.046	0.127 ± 0.054
TOI-2180b	0.106 ± 0.032	0.142 ± 0.028	0.142 ± 0.062	0.928 ± 0.062
NGTS-20b*	0.094 ± 0.044	0.124 ± 0.039		
TOI-5153b	0.123 ± 0.057	0.150 ± 0.054	0.118 ± 0.058	0.128 ± 0.036
TOI-2589b	0.054 ± 0.025	0.090 ± 0.032	0.065 ± 0.033	0.600 ± 0.066
WASP-162b*	0.181 ± 0.082	0.198 ± 0.068		
TOI-2338b*	0.165 ± 0.031	0.165 ± 0.035		

Table A.3: Retrieved results for FM, CE, DCA, and DCA3 models for all the 43 exoplanets. This table contains the results shown in 10 for the original 21 planets with the addition of the remaining 23. The plot also contains the results from CEA and SMA Solar atmosphere metallicity only done for the original 21 planets, since no further statistics has been applied.

Planet	DCA	DCAf=0.1	DCAf=0.4
HAT-P-12b	0.123 ± 0.047	0.1232 ± 0.0470	
HAT-P-17b	0.049 ± 0.026	0.0500 ± 0.0250	
HAT-P-19b	0.046 ± 0.017	0.0492 ± 0.0178	
HAT-P-54b	0.161 ± 0.049	0.1708 ± 0.0502	0.2054 ± 0.0529
HATS-49b	0.406 ± 0.032	0.4266 ± 0.0444	0.4242 ± 0.0194
HATS-75b	0.220 ± 0.023	0.2201 ± 0.0253	0.2735 ± 0.0287
HD332231b	0.210 ± 0.046	0.2222 ± 0.0470	0.3127 ± 0.0651
K2-287b	0.270 ± 0.023	0.2845 ± 0.0237	0.3306 ± 0.0252
TOI-3629b	0.427 ± 0.029	0.4309 ± 0.0284	0.4752 ± 0.0249
WASP-84b	0.135 ± 0.045	0.1448 ± 0.0464	

Table A.4: Magnitude of different assumed f factors as reported in equation 14. Chemical structures are matching DCA models, and are compared with such model with an assumed adiabatic interior.

8.2 Grid Testing and Validation

In order to choose the right amount of steps for the grid, grid-size testing has been made to ensure a good balance between accuracy and computational efficiency. Combinations of the three variables Z, M, and Age (R is used for the retrieval step explained in the Methodology). In order to carry a grid testing evaluation (to avoid too grids that are too coarse and the results can shift, but also considering root-find execution time and energy). The results from the grid may vary from planet to planet according to their observed uncertainty. For this reason, we have carried this experiment of planet K2-287b that has precise constrains on the variables, and consequently on a younger planet TOI-201b with low error on the age parameter, giving the same results. The results will be shown only for planet K2-287b in the following figures. To test the grid, the first step was to do step 1, and so the retrieval of the several evolution lines from CEPAM. This will be the starting ground from where we start. From this, we ran step 2 for different parameter ranges, and all their possible combinations between them for the three controlled parameters on CEPAM, that is Z, Age, and M. R is the input for the rootfind, and a grid is not being implemented on the value. R is being kept at the maximum precision that CEPAM could retrieve. As unique values of R are found as outputs from cepam, steps in a equally-distance grid has not be applied for this variable. Thus, we will inspect how the retrieved distribution Z changes by playing with the other three parameters. The steps that will vary are the following:

- Z tryed ranging from steps every 0.001, 0.0025, 0.005, 0.01, 0.025, 0.05, and 0.1 (between 0 and 1), giving the total number of steps of 1000, 400, 200, 100, 40, 20, and 10 respectively. Age and M values have been taken fixed at 700 and 9 each.
- Age will be treated differently, as the steps are taken from the full range of 0 to 14 Gyr, that will not necessarely match with the planetary age range. Since CEPAM converges more results at small ages, applying the same distancing might be not ideal for all planets. The inspection of TOI-201b however, gave us the confirmation of the grid size also in the young planetary range. Age steps tested are: 2800, 1400, 700, 140, 28, 14. Since these steps Have been taken throughout the 0-14 Gyr range, we can then say the steps represent successions of every 0.005, 0.01, 0.02, 0.1, 0.5, and 1 Gyr respectively.
- Mass M will be inspected between the tailored observational range of the planet, and thus the steps tested (9, 7, 5, 3, and 1) are the number of steps withing the range [-3σ;+3σ]. The numbers of steps in this range will be: steps of every +/- 0.75σ for the nine steps, 1 σ for steps equal to 7, 1.5 σ for steps=5, for steps=3 the mean value together with the +3σ and -3σ values, while for step 1 only the mean mass will be used, similarly to what shown in Figure 6.

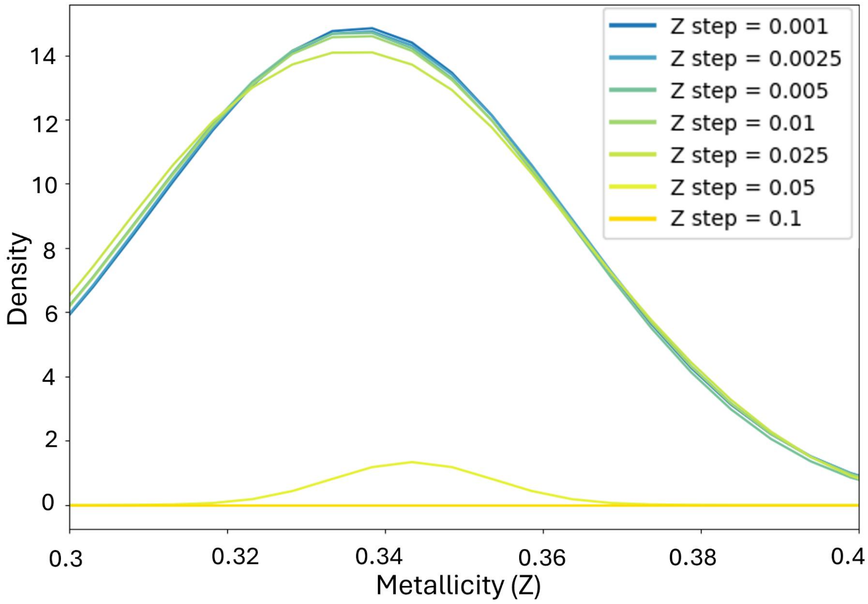

Figure A.1: Different Mass stepts tried (9, 7, 5, 3, 1, with their respective frequency on the mass σ provided in the plot) for a fixed Age and Z grids of 700 and 400 respectively. This has been tested for a fixed Age and M of respectively 700 and 9. Tested on planet K2-287b with an FM model.

Figure A.1 shows the results for the first grid tested at varying Z. To note that the execution time for Z steps $= 0.001$ with the chosen Age and M was long enough to be discarded as an option in processing 43 planets. The figure shows how the estimated curves follow the same trend at least up to a coarse grid of 0.025 (40 total steps case). For Z step $= 0.025$ there is a slight decrease in the retrieved density and a slight shift. When coarsening up to 0.05, even if the convergence still gave an indicative location of the estimates, we decided that the mean, and especially the standard error, was not reliable enough. Z step $= 0.1$ was too coarse to retrieve any result. From these results, we have agreed in taking a Z step of 0.0025 (400 values between 0 and 1 for each unique Age and M combination), as the grid was precise enough staying on the conservative side to avoid case-to-case unreliabilities. This has been adopted as grid for the main methodology, but also kept fixed for the upcoming grid testing validations.

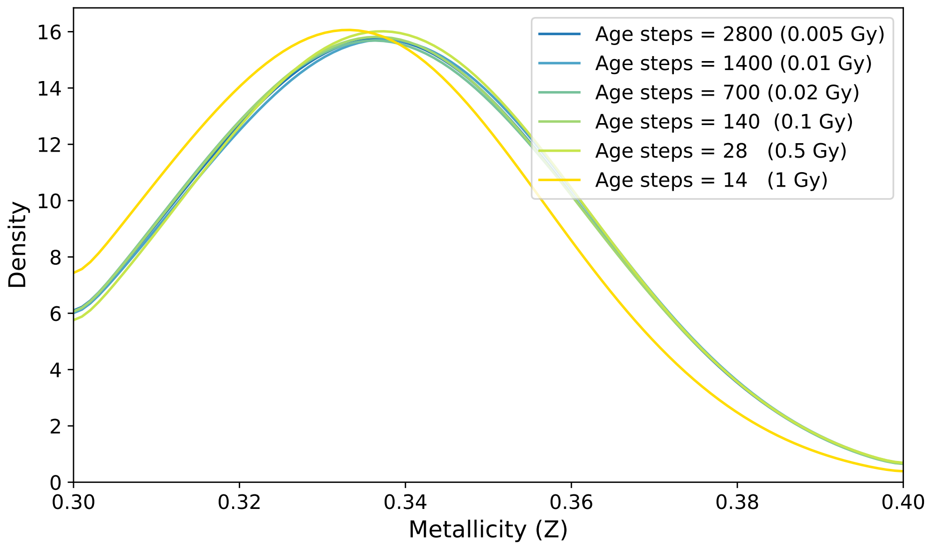

Figure A.2: Different age stepts tried (2800, 1400, 700, 140, 28, and 14) for a fixed M and Z grids of 5 and 400 respectively.Tested on planet K2-287b with an FM model.

The second variable inspected for grid size was *Age*. *Age* steps too high like 2800 were immediately discarded for prolongued computational time wait, as the increase in grid exponentially increases the waiting time for each root, as every unique combination of the three variables *Age*, Z, and M follow an exponential increase. *Age* steps of 14 shows a slight shift in the retrieved output, and even it does not seem to highly affect the output for planet K2-287b, we considered unreliable. We decided on an *Age* step of 700 for this study, as this has been observed to be the best balance between quality and waiting time.

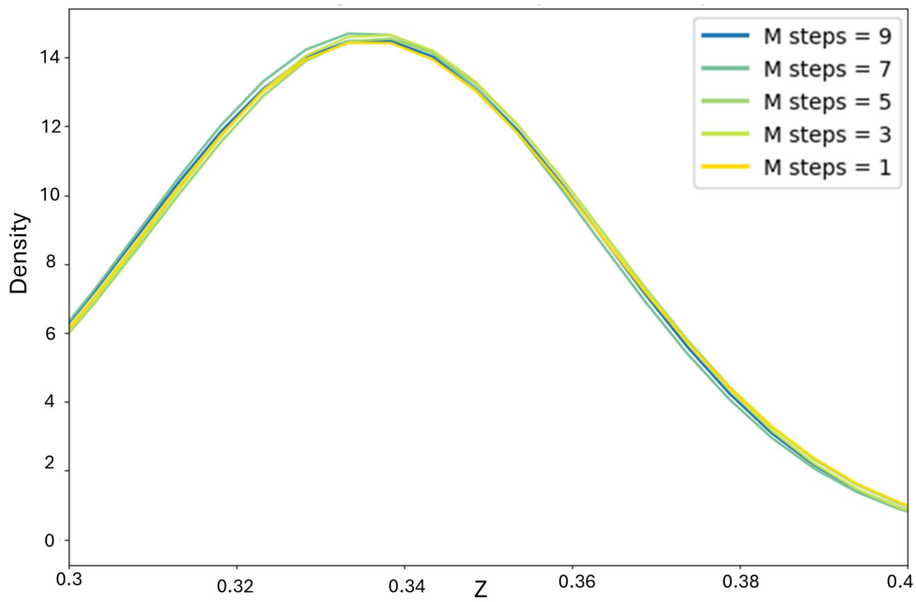

Figure A.3: Different Mass stepts tried (9, 7, 5, 3, 1, with their respective frequency on the mass σ provided in the plot) for a fixed *Age* and Z grids of 700 and 400 respectively. Tested on planet K2-287b with an FM model.

Finally, M has been tested for grid size. In this case, for all masses tested, the result is not varying.

This is because M variability is not playing a major role in Z shift, since the distribution found for using M as a fixed mean value is the same as with considering the full range, even in terms of standard deviation. Since Figure 6 shows the non-linear variability in masses for different planetary cases, we decided to go for the highest possible grid size of 9, since there is not a dramatic increase in computational energy for the size of this variable's grid. Young and a more evolved planets have been inspected to ensure that early stage cooling is still expressed well by the grid size. Obviously these tests have not been executed for the whole sample size, and it is highly dependent on the planet's size, observational accuracy of the planet, and model used. It is then appropriate to say that this grid size is not universally valid, and with future missions and better data constrains, grid size should adapt as a result. We assume that with the similar observational error of our data, the grid size chosen remains valid throughout.

8.3 Convergence evaluation

Previous approaches (Howard et al., 2025; Miguel and Vazan, 2023; Müller et al., 2020; Fortney et al., 2009) aimed at estimating the range of possible metallicities by defining a strict-boundary observational box:

$$R \in [R_{\text{obs}} - \sigma_R, R_{\text{obs}} + \sigma_R] \tag{19}$$

$$\text{Age} \in [\text{Age}_{\text{min}}, \text{Age}_{\text{max}}] \tag{20}$$

with a fixed mass M_{mean}. The range of metallicities is determined by the cooling tracks that touch the extreme points, so the top-right corner $(R_{\text{obs}} + \sigma_R, \text{Age}_{\text{max}})$ and the bottom-left one $(R_{\text{obs}} - \sigma_R, \text{Age}_{\text{min}})$.

This process means that instead of considering only the extreme cases, we sum over all possible values weighted by their probabilities.

Mathematically, the probability distribution of inferred metallicities is:

$$P(Z) = \int_{\text{Age}_{\text{min}}}^{\text{Age}_{\text{max}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(R \mid M, \text{Age}, Z) P(M) P(R) P(\text{Age}) \, dR \, dM \, d\text{Age} \tag{21}$$

where P(M, R, Age) are the probability distributions of each variable. This integral effectively weights metallicity estimates by how likely each mass, radius, and age combination is. While adding another variable such as mass increases the estimated Z error, The integral, dominated by the most probable values ending up in the middle (apart from age), narrows down the inferred metallicity when compared to previous methods of taking the extreme values. The extent of this improvement varies according to the planet inspected, and by the uncertainty of the observed variables.

In situations where the draws of the variables was retrieving a result outside the root-find range, (like negative values or higher than 1, where 0 is a planet completely composed of H-He, and 1 is fully made of heavy elements, ices and rocks) the retrieval would fail. Failure could also be caused when no Z value can be found for the unique set of selected R, M, and Age combination. For this reason, planets with distributions close to zero will give an estimated failure percentage of 10-50%, close to what Thorngren et al. (2016) has estimated with a similar approach. We sustain that even in these cases, with a reduced of succeeded generated samples, the accuracy of the retrieved curves remains sufficiently high.

When inspecting the overall convergence percentage, in the majority of planet cases, the amount of failed convergences was low (around 0.3%). This matches the 3σ range (that explains 99.7% of the data) chosen to be kept in the radius range for the grid size. While extrapolation could have also covered that percentage, such fraction will not impact greatly on the estimated standard deviation (especially where planetary observational data error is substantially smaller than the model-carried error); it has been decided to be not meaningful enough to be considered.

For the cases where convergence success rate was lower than 50% (not caused by out-of-bounds values), this was mainly due to the crashing of the evolution model and the saving of two identical tracks, which disrupted the interpolation step during the grid creation. Re-running CEPAM was enough to solve the problem in the retrieval. To ensure consistency in the method, tests on the grid size have been made to validate the method. A good balance between accuracy, computational cost, and execution time was chosen, resulting in the values discussed in the methodology.

8.4 Single variable testing

In this section, variables from the dilute core structure function will be inspected in effect singularly. Following, some comparison will be made for different structures. Single cooling tracks for a fixed Z (arbitrarily, here we are not trying to match the curves to their observational ranges) and M (using the mean mass). Variable inspection will be made on TOI-1811b, a planet with mass and radius close to the one of Jupiter. Results are being shown in the following figures showing on the left panel the internal structures tested (so how Z changes at different mass fractions), and on the right the resulting cooling tracks generated for comparative purposes. The variable tested are Z_{dilute}, m_{dilute}, δ_m, and Z_{atm}. The results from the inspection of these four variables could be seen in Figures A.4, A.5, A.6, and A.7.

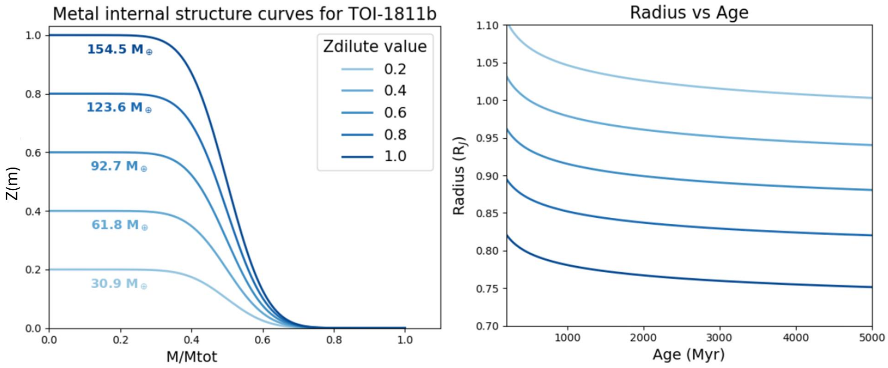

Figure A.4: Different tested Z_{dilute} values on sample planet TOI-1811b

Figure A.4 shows the changes in Z_{dilute} in the internal structure. For each structure, there is a linear increase in the Z_{dilute} value, from 0.2 to 1 at 0.2 distances. Since the other parameters are ifixed, by increasing Z_{dilute}, the fraction and mass of heavy elements increases (one can see the mass values for each structure in Earth masses inside the Figure's left panel. Since the only thing changing in the increase mean molecular weight at the planet's core given by a higher Z, the planet will be more susceptible to contraction due to the denser region formed at the centre. Thus, the planet resists better thermal expansion and inflation mechanisms affecting the radius, at the same pressure and temperature, it occupies less volume. Gravitational self-compression becomes more effective when the mean molecular weight increases, and so it results in a lower expected radius shown by a lower cooling curve in the plot. The cooling is quite linear and very well explainable by the linear increase of Z_{dilute}.

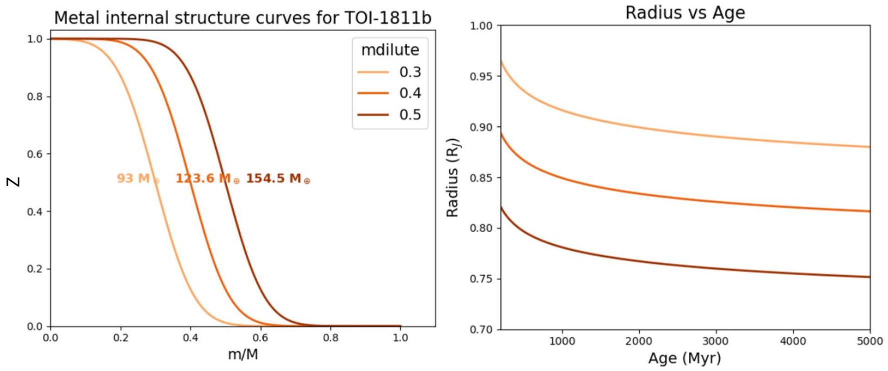

Figure A.5: Different tested m_{dilute} values on sample planet TOI-1811b

Similarly to what shown for Z_{dilute}, Figure A.5 shows how the change in m_{dilute} (with the other parameters remaining fixed) changes the cooling tracks retrieved by CEPAM. By keeping the other variables fixed, in this case we are also having an increase in heavy element mass shown by the numbers in the left panel (colour coded with the curves). Similarly to what seen for Z_{dilute}, the change in the extent of the dilution curve, and its natural increase in Z, results in a lower expected radii calculated by the evolution models. There is again linearity in the transition when changing for m_{dilute} values from 0.3, 0.4, and 0.5. Of course the trend would change when the dilution curve is situated at shallower planetary depth in such a way the curve does not reach Z=0 at the atmospheric layer. However, how CEPAM treats such case, is that it abruptly brings Z to zero (as a vertical gradient), unless Z_{atm} is being actively inputted. No opacity is being changed at the atmospheric interface, even at increased Z at shallow depths due to a flattening of the curve (given by a higher m_{dilute} value).

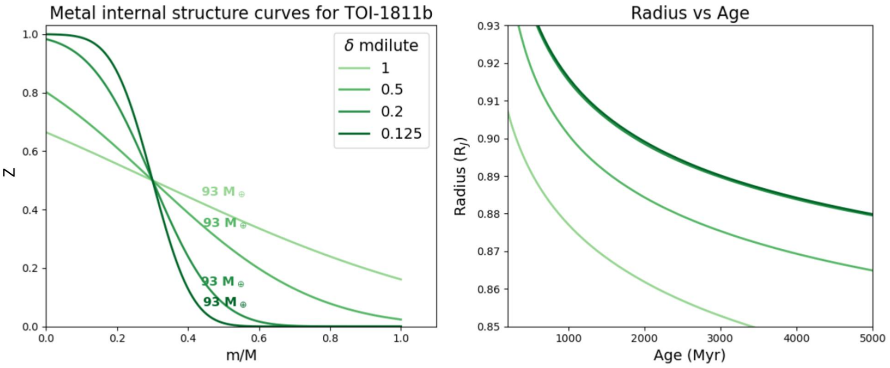

Figure A.6: Different tested δ_m values on sample planet TOI-1811b

The different m_{dilute}-driven changes in the structures can be seen in Figure A.6, where the parameter δ_m is being changed. This parameter changes the slope of the dilution curve, with δ_m=0 would represent a hard boundary similar to the CE model, and δ_m=1 having a linear composition gradient (as shown by the lightest green colour structure on the left panel in Figure A.7. In this case, the simple change on the curve's slope does not affect the fraction Z in a planet's interior. This evaluation then has been done for a fixed M_z=93$M_\oplus$. Results show not much variation in the output when we compare δ_m=0.125 (the actual value being used, and kept fixed, in this study) and δ_m=0.2. While there is a stronger shift in the cooling curves at higher δ_m values. Since there is no active atmospheric metallicity, the results shown are only controlled by density changes.

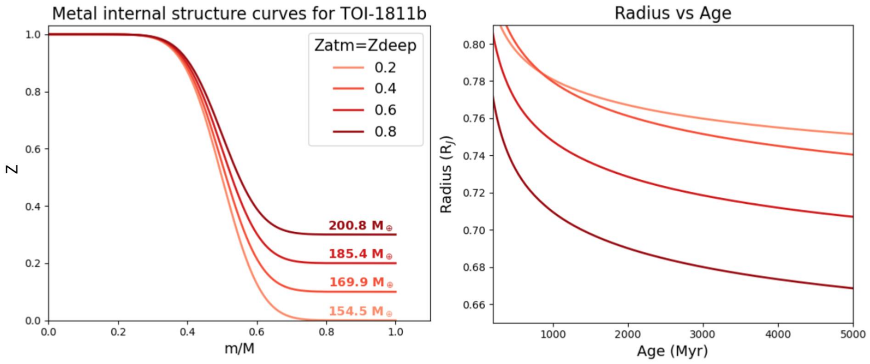

Figure A.7: Different tested Z_{atm} values on sample planet TOI-1811b

To check again if at high Z opacity could inflate planetary radii, a quick check on the magnitude of atmospheric metallicity Z_{atm} should be done. We then tested different values, where Z_{atm}=Z_{deep}=0, 0.1, 0.2, and 0.3. As the other variables are fixed, the fraction Z and the mass of heavy elements M_z inside of the planet will rise. Unexpectedly, Figure A.7 shows a decrease in the cooling track radii by adding more Z_{atm}. We believe this is not related on the opacity effect, but more on the dominance of heavy elements inside of the planet that contracts more than the contribution of keeping the planet warmer by the opacity in near surface.

This is the evidence of the degeneracies around planetary evolution studies, and why such effect in the larger trend when comparing CE with DCA, but it does give an explaination to the rare cases where DCA was in fact underpredicting Z when compared to CE. This shows how at high Z density could overcome any opacity increase over a certain threshold, and so an FM model would not show the highest inflation. This is believed to have affected planet TOI-4010b, the lowest mass planet in the dataset, believed to have high Z, and expected higher in CE rather than FM as shown for the majority of the other cases.

If then the issue is the change of mass more than the atmospheric input, it is worth inspecting for a fixed heavy element fraction Z.

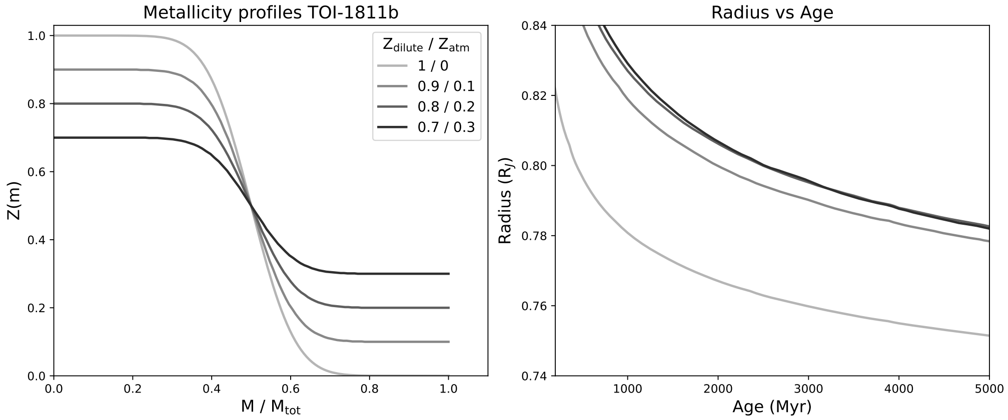

Figure A.8: Different structures tested for a fixed Z=0.5, and different combinations of Z_{atm} and Z_{dilute} Tested on sample planet TOI-1811b

In Figure A.8, we are inspecting the extent of the opacity effect by keeping a constant Z of 0.5. In order to keep Z constant, a reduction in Z_{dilute} or m_{dilute} has to be done to adjust of the compensation in the atmosphere (shown in the left panel). For this reason, we decided to change Z_{dilute} decreasing of 0.1 each time (starting at 1 and ending at 0.7 for this inspection), while similarly the Z_{atm} will be incremented by 0.1 each time for a maximum of 0.3. The resulting curves for the different structures finally show the opacity-driven inflation when heavy elements are not being added. The biggest difference is between the Z_{atm}=0 and Z_{atm}=0.1, where the introduction to the heavy elements in the atmosphere drastically boosts its expected inflation. There is a limit being reached where no more heavy elements being added to the atmosphere would actually inflate more the planet,s radius. This can be seen when comparing Z_{atm}=0.2 and 0.3, where the two expected cooling tracks are similar one to the other. Adding too much Z would then promote contraction and so decrease the expected R. This shows the non linearity, and the degeneracy of finding different suitable structural conditions for planets.

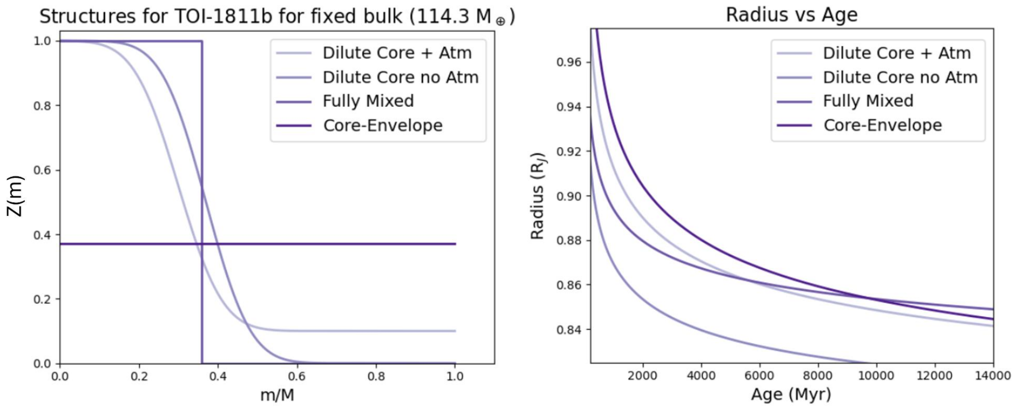

Figure A.9: Different tested structures for a fixed composition constant in all assumptions. Structures tested are FM, CE, DC, and DCA with Z_{atm}=0.1. Tested on sample planet TOI-1811b

Similarly to what done in this study, a quick inspection of the main structures could be done to retrive a single line each. This is being shown in Figure A.9, where CE, FM, DC, and a DC with Z_{atm}=0.1 have been tested. Parameters in all models have been adjusted to match the same Z for all four cases. The cooling tracks show clearly an initial temperature-dominated phase where fast cooling contracts the planet quickly, and then a shallow up in a gravity-controlled setting. R-Age results show underpredicted R notably for the DC model, with the other three being similar overall. The right panel in Figure A.9 shows non linearities between CE, FM, and DCA, and also a difficulty in interpretation of the data. For this reason, the statistical retrieval tool done in this study is essential to untangle challenges in gathering comprehensive results for easier interpretations.

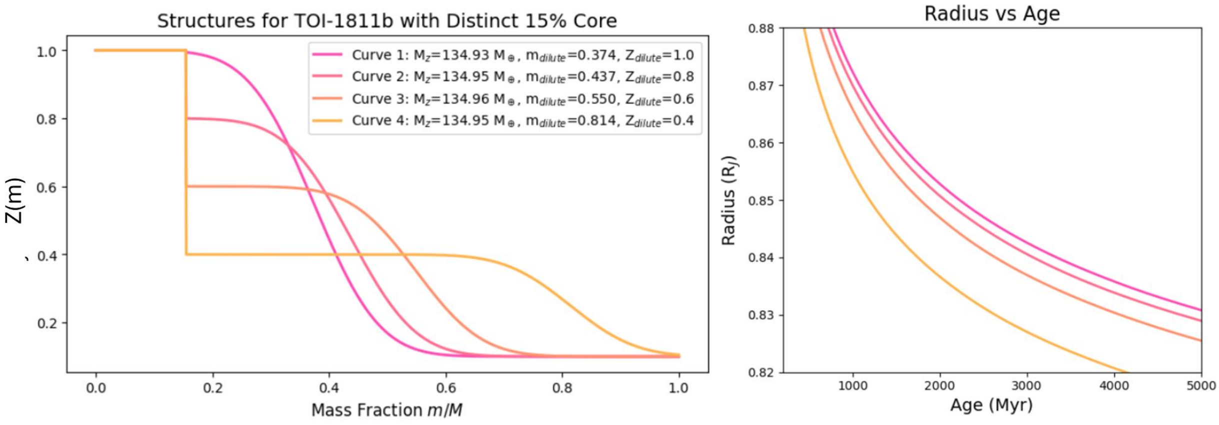

Figure A.10: Different tested dilute structures with an assumed core comprising of 15% of the total mass, tested for different m_{dilute} and m_{dilute} to keep the total planetary Z constant. Tested on sample planet TOI-1811b

Even if not actively tested in this study, the possibility of a distinct core on top of a dilution curve (symbolising chemical mixing over time) is a feasible structure that could be present in planetary interiors, and can be used as an additional constraining variable (van Dijk and Miguel, 2025). This is what has been tested and shown in Figure A.10, where a change in m_{dilute} change to adapt for a decrease in $Z_{dilutec}$ for a fixed M_{core} to maintain the same Z once again. This can still be achieved by adding M_{core} in the formula shown in equation 12. The specifics of each structure can be seen in the legend, followed by the colour coded line. The change of m_{dilute} similarly to what seen in Figure A.5 would suggests linearcdecrease. However, the adjustment made by Z_{dilute} to maintain the same Z changed the behaviour. There is an exponential change in adding denser material at shallower depth (higher m_{dilute}) even with a shallowing up of Z_{dilute}. While in all four structures Z_{atm} is zerothe change of the m_{dilute}/Z_{dilute} is noticeable. Such visible, but relatively-negligible (or at least compared to the current observational error range) changes in the expected radius is limiting an accurate constraining of m_{dilute} and Z_{dilute} variables, as many solutions would fit the same observational range. This once again show the challenge in interior modelling and structure determination in different planetary environments.

8.5 MCMC corner plots posterior evaluation

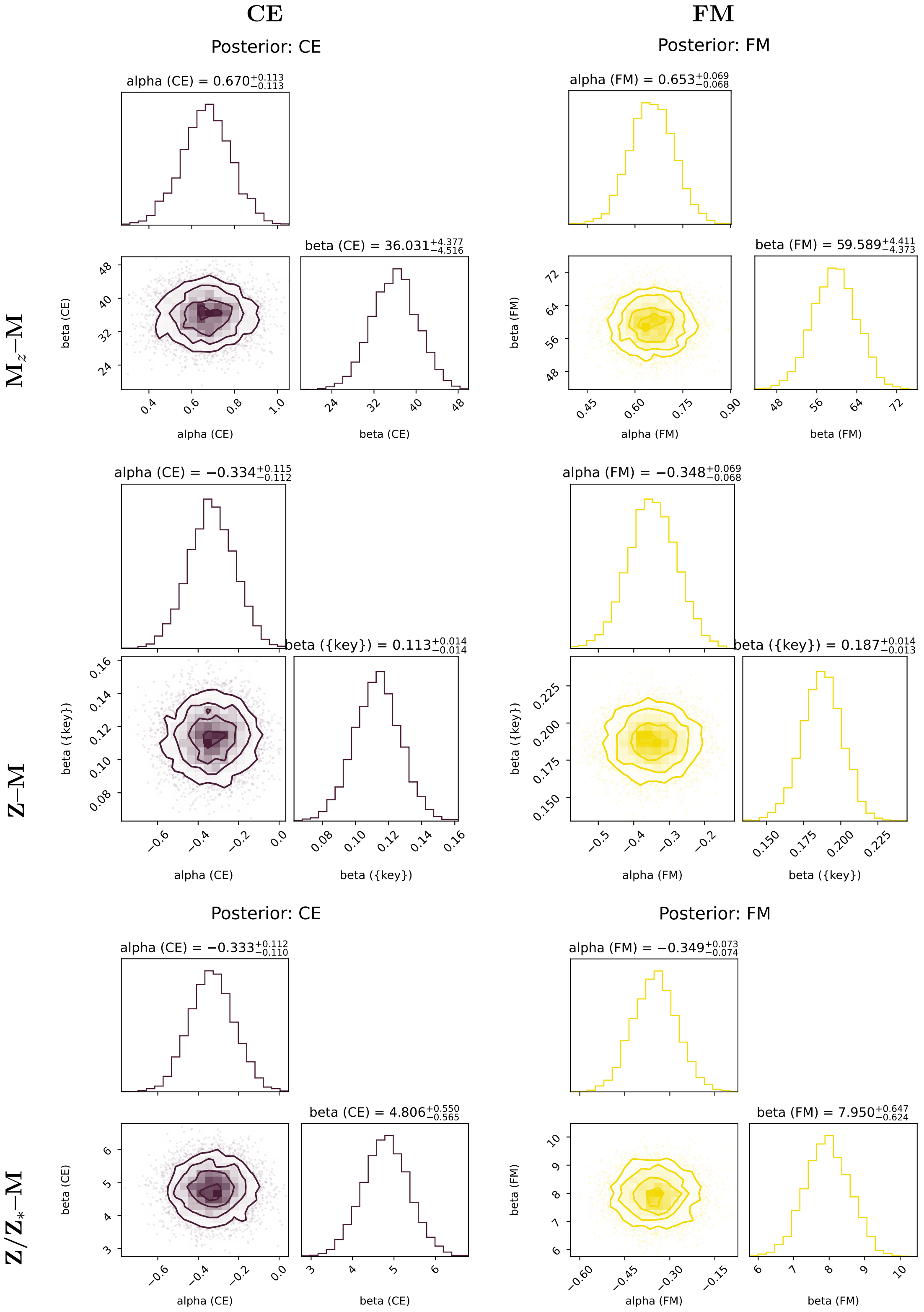

Figure A.11: Corner plots for structural models CE and FM across parameters: M_z–M, Z–M, and Z/Z_*–M.

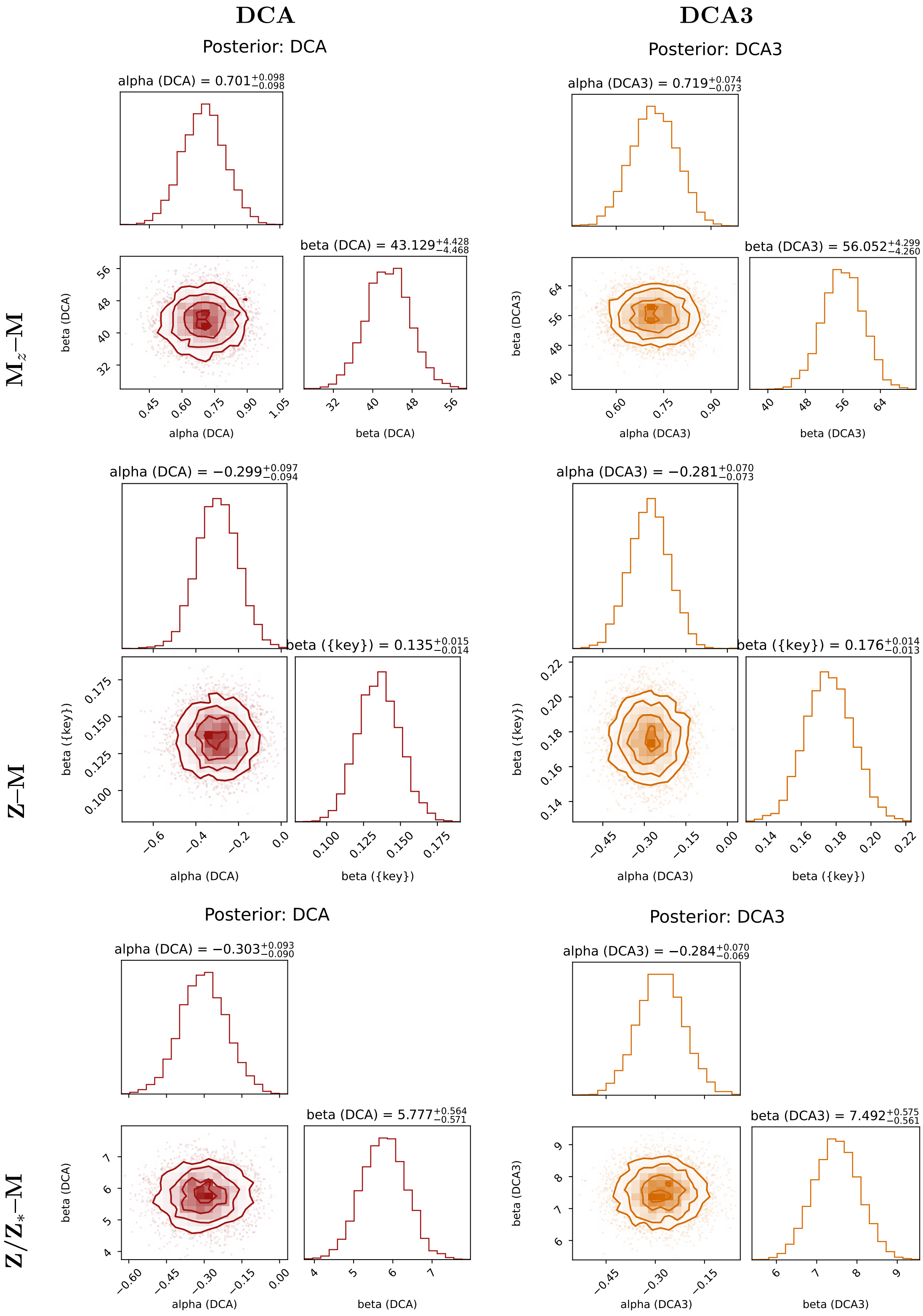

Figure A.12: Corner plots for structural models DCA and DCA3 across parameters: M$_z$–M, Z–M, and Z/Z$_*$–M.

8.6 CEA

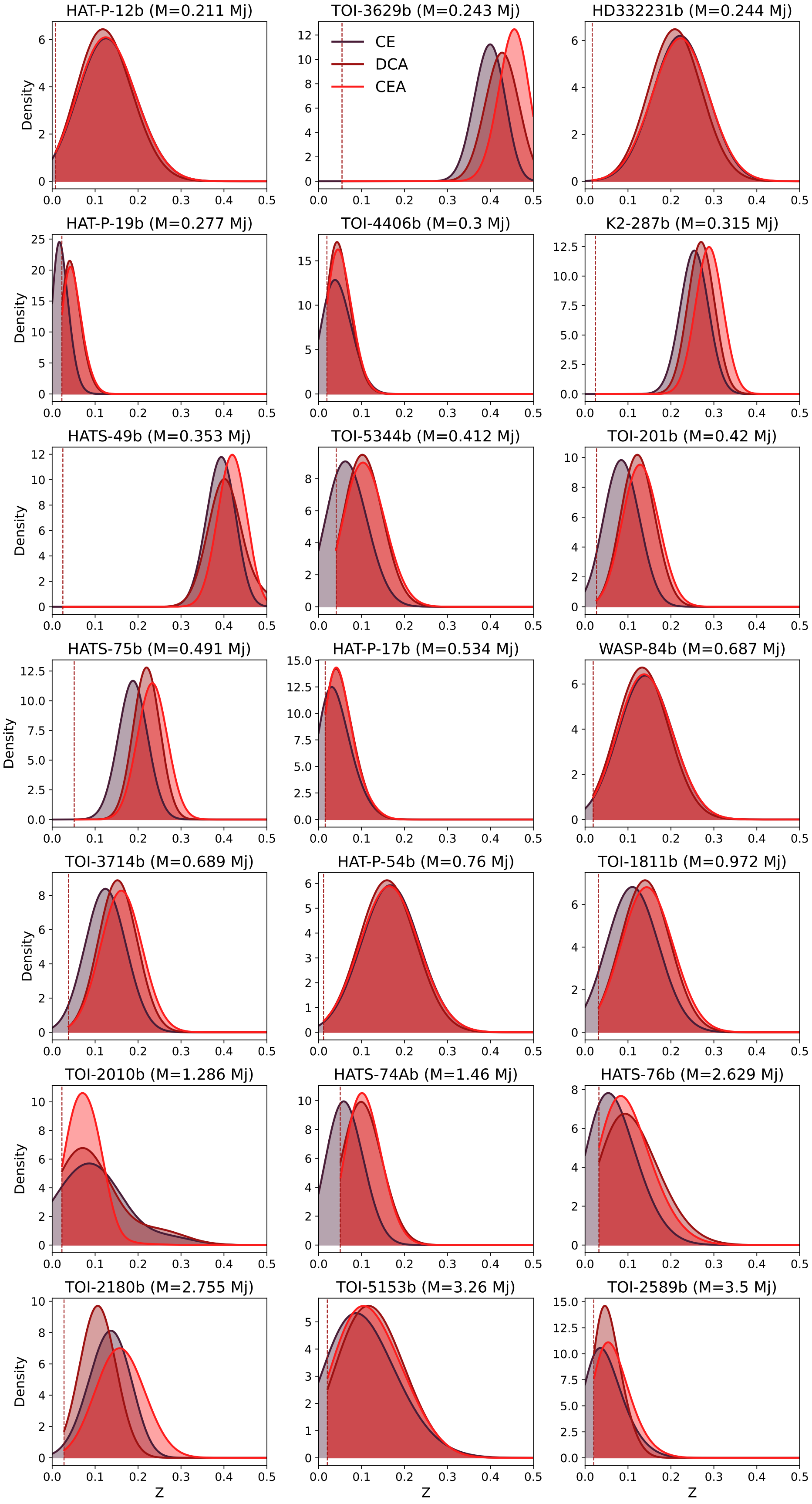

Figure A.13: Retrieved heavy element fraction curve distributions for the initial 21 planets for DCA, CE, and CEA.

Personal Declaration

I hereby declare that the submitted thesis is the result of my own independent work. All external sources are explicitly acknowledged within the thesis.

University of Zürich, 2/8/2025

Lorenzo Peerani