

# A radius valley between migrated steam worlds and evaporated rocky cores

Received: 20 January 2023

Accepted: 14 December 2023

Published online: 09 February 2024

 Check for updates

Remo Burn<sup>1</sup>✉, Christoph Mordasini<sup>2</sup>✉, Lokesh Mishra<sup>2,3,5</sup>,  
Jonas Haldemann<sup>2</sup>, Julia Venturini<sup>3</sup>, Alexandre Emsenhuber<sup>4,6</sup> &  
Thomas Henning<sup>1</sup>

The radius valley (or gap) in the observed distribution of exoplanet radii, which separates smaller super-Earths from larger sub-Neptunes, is a key feature that theoretical models must explain. Conventionally, it is interpreted as the result of the loss of primordial hydrogen and helium (H/He) envelopes atop rocky cores. However, planet formation models predict that water-rich planets migrate from cold regions outside the snowline towards the star. Assuming water to be in the form of solid ice in their interior, many of these planets would be located in the radius gap contradicting observations. Here we use an advanced coupled formation and evolution model that describes the planets' growth and evolution starting from solid, moon-sized bodies in the protoplanetary disk to mature Gyr-old planetary systems. Employing new equations of state and interior structure models to treat water as vapour mixed with H/He, we naturally reproduce the valley at the observed location. The model results demonstrate that the observed radius valley can be interpreted as the separation of less massive, rocky super-Earths formed in situ from more massive, ex situ, water-rich sub-Neptunes. Furthermore, the occurrence drop at larger radii, the so-called radius cliff, is matched by planets with water-dominated envelopes. Our statistical approach shows that the synthetic distribution of radii quantitatively agrees with observations for close-in planets, but only if low-mass planets initially containing H/He lose their atmosphere due to photoevaporation, which populates the super-Earth peak with evaporated rocky cores. Therefore, we provide a hybrid theoretical explanation of the radius gap and cliff caused by both planet formation (orbital migration) as well as evolution (atmospheric escape).

The driving process leading to the underabundance of planets with radii  $R \sim 1.7 R_{\oplus}$  found in the data from the Kepler spacecraft<sup>1–3</sup> could be related to escape, due to stellar X-ray and ultraviolet irradiation<sup>4–6</sup> or core-powered mass loss<sup>7–9</sup>, of hydrogen and helium (H/He) on top of a core consisting of silicates and iron. In the following, we

call this mixture of solid materials rocky. With this limited compositional inventory, the observed sharp drop-off of the sub-Neptune occurrence at radii greater than  $\sim 3 R_{\oplus}$ —the radius cliff—likely requires an additional mechanism, such as H<sub>2</sub> sequestration in the magma ocean<sup>10</sup>.

<sup>1</sup>Max-Planck-Institut für Astronomie, Heidelberg, Germany. <sup>2</sup>Physikalisches Institut, Universität Bern, Bern, Switzerland. <sup>3</sup>Observatoire de Genève, Versoix, Switzerland. <sup>4</sup>Universitäts-Sternwarte München, Ludwig-Maximilians-Universität München, Munich, Germany. <sup>5</sup>Present address: IBM Research, Rüschlikon, Switzerland. <sup>6</sup>Present address: Physikalisches Institut, Universität Bern, Bern, Switzerland. ✉e-mail: [burn@mpia.de](mailto:burn@mpia.de); [christoph.mordasini@space.unibe.ch](mailto:christoph.mordasini@space.unibe.ch)

Another possible explanation for the radius distribution is that the sub-Neptunes are water-rich, with water mass fractions of several tens of percent<sup>11–13</sup>. Such large water contents are a consistent prediction of planet formation models that include the effect of planet–disk interactions, leading to migration of ice-rich planets from outside the water condensation line towards the star<sup>14,15</sup>. In this scenario, the common assumption that the gaseous envelopes are dominated by H/He is no longer required. This hypothesis has recently gained supporting observational evidence based on planets around M stars<sup>16</sup>, whose bulk density distribution can be well reproduced with silicates and iron for super-Earths and about equal fractions of rock and water for sub-Neptunes. The tentative evidence for this scenario lies in the small scatter of sub-Neptune densities<sup>16</sup>, which might not be expected for rocky cores with a H/He envelope. However, this can also be matched in the classical picture if the initial H/He fraction of the planets exhibits little scatter at fixed planetary mass due to formation and boil-off processes<sup>17</sup>.

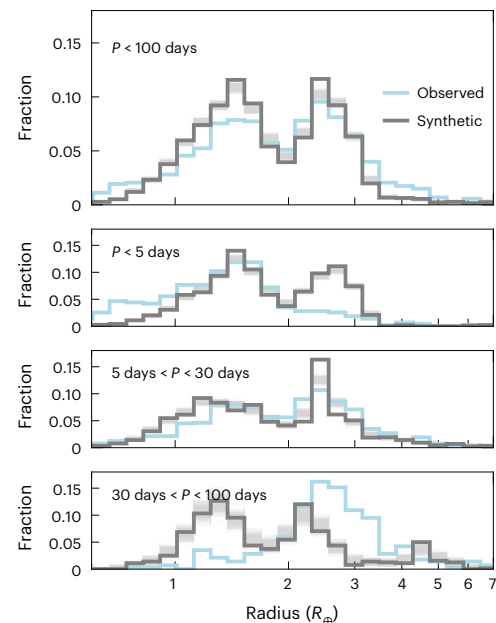
For more massive stars, the smaller observational signal often impedes precise planetary mass determination. In that case, a clear picture of density does not emerge with present-day data, and the different theoretical models cannot easily be falsified<sup>18</sup> unless the planetary masses are constrained using theoretical arguments, such as the output of a planet formation model<sup>15,19,20</sup>.

In the case of water-dominated outer layers of the planet, the phase of water determines the precise radius. For close-in planets, water forms a hot—to a large degree supercritical—vaporized hydrosphere, here called the steam envelope<sup>12</sup>, which increases the radius compared to condensed, solid, high-pressure ice<sup>21</sup>. Due to model limitations, water was not consistently included in this lower-density phase in the first works coupling a planetary mass distribution from planet formation to photoevaporation models<sup>5,19</sup> and/or comparing the observed valley locus with the theoretically predicted one<sup>19,22</sup>.

This prior work excluded that the super-Earths contain substantial amounts of water and favoured water-poor sub-Neptune compositions because planets containing solid ice were located in the radius gap. Later, a combined formation and evolution model with the correct water phases demonstrated the emergence of the radius valley as a separator between dry and wet planets formed within, respectively beyond, the ice line<sup>15</sup>. In that work, the main driver of the dichotomy is varying pebble accretion efficiency based on pebble composition, which produces smaller rocky and larger icy cores. Recently, another study<sup>20</sup> reached the same conclusion using a similar model, incorporating also  $N$ -body interactions between planets.

Here, we use a coupled formation and evolution model to synthesize a population of planets that can be statistically compared to the Kepler mission by applying its detection bias<sup>23</sup>. During the evolution stage, the planets are evolved individually by calculating their interior structure, which captures effects of cooling and contraction. In addition, for the nominal model, we assume that water mixes with H/He at the high temperatures above the runaway greenhouse limit relevant for comparison to the Kepler data<sup>24,25</sup>. For water, we use a new equation of state<sup>26</sup>, which covers all possible phases. Under these assumptions, water is also present in the upper layers of the gaseous envelope or can even be the only volatile constituent. Thus, we weight by mass the loss of H/He (following ref. 27) and water<sup>28</sup> while keeping the envelope water fraction (or metallicity) constant—an approximation most accurate in the limit of efficient hydrodynamic mass loss<sup>29</sup>.

The formation stage is taken from our global planet formation modelling<sup>30,31</sup> growing 1,000 systems of planets. They were generated in a statistically robust way by choosing initial conditions randomly sampled from observations of disks<sup>32</sup>. This enables us to extract the key mechanisms shaping the radius valley and statistically quantify their ability to reproduce observations. In particular, orbital migration and  $N$ -body interactions between growing planets are included. Motivated by prior  $N$ -body simulations<sup>33</sup>, which show that some late dynamical events can influence the inner planetary systems, we extended the



**Fig. 1 | Radius histograms of observed and synthetic planets.** The light blue line shows the observed distribution without correction for the observational bias<sup>2</sup> and the grey line the synthetic one using the updated, nominal evolution model with the bias of the Kepler survey applied. Opaque lines show 500 random realizations of the synthetic planets with 5% error in radii. We restrict the sample to planets with certain orbital periods ( $P$ ), as indicated in the top-left corner of each panel. Histogram bin counts are normalized by the total number of planets in the different samples.

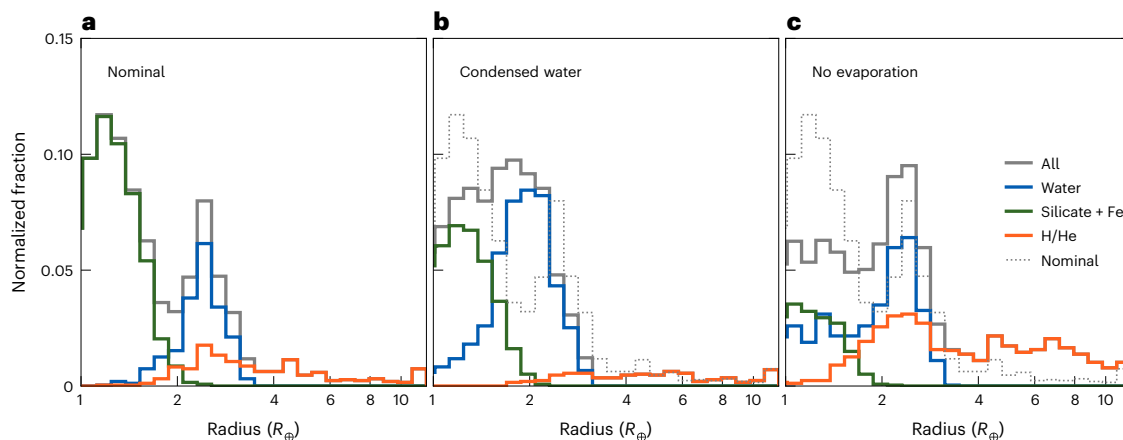
formation phase with  $N$ -body integration and planetesimal accretion to 100 Myr, compared to 20 Myr in the original work<sup>31</sup>. We note that planetesimal accretion essentially terminates after the gaseous disk has dissipated at around 3 Myr due to the missing gas drag damping the planetesimals' eccentricities and inclinations. The formation and evolution processes, detection bias and their implementation are detailed in the Methods section.

## Results

### Radius distribution

We contrast our theoretical radius distribution with applied observational bias to observations<sup>2</sup> in Fig. 1. For the overall distribution, we find an excellent match for the locus of the sub-Neptune peak, the radius valley and the super-Earth peak. The synthetic radius cliff is only marginally steeper, while the relative number of super-Earths in the model is higher than observed. Statistical quantification of the differences (Extended Data Fig. 1) reveals that the radius distribution of super-Earths within 30 days is well matched. At larger distances, we synthesize and theoretically 'detect' more rocky planets than observed. The depth of the radius valley is marginally deeper in the synthetic population even after accounting for typical measurement uncertainties<sup>3</sup>. At the innermost and larger orbital periods, differences between theoretical and observed distributions offer insights for model improvement, as discussed in the following sections. We achieved a natural match with the observed distribution of planetary radii by consistently including more realistic physics, especially various phases of water, compared to previous works<sup>19</sup>. The result is obtained from following the self-consistent formation and evolution of initially 100  $0.01 M_{\oplus}$  planets per disk. We did not adjust any parameters in the underlying formation phase or the evolutionary model.

Figure 2 displays radius distributions of different modelling runs without observational bias applied, categorized by bulk composition. Figure 2a shows that the valley separates smaller, dry planets from



**Fig. 2 | Radius histograms of synthetic planets with orbital periods shorter than 100 days for different model assumptions without any bias applied.**

**a–c**, Full distribution is shown in grey, and coloured histograms show different compositional subsets (pure rocky in green, water-rich without any H/He in blue, with some H/He in orange): nominal model (**a**), water assumed to be condensed

out in a solid ice layer resistant to evaporation below H/He (**b**), nominal treatment of water but excluding evaporation (**c**). The dotted histograms show the full distribution from the nominal simulations. As in Fig. 1, bin counts are normalized by the total number of planets in the different samples.

lower-density, vapour-enveloped, wet planets with little overlap. The planets whose envelopes contain some H/He make up only 13–16% (depending weakly on age) of the sub-Neptunes. If water-rich planets are excluded, we do not reproduce the observed radius distribution (see also Extended Data Fig. 1). Furthermore, most of the planetary envelopes of sub-Neptunes with some H/He (orange), contain less than 10% of H/He by mass.

The radius cliff located at  $\sim 3 R_{\oplus}$  is in our simulations a second compositional transition from planets containing  $\geq 90\%$  water to those that contain several tens of percent of H/He in mass. It is accompanied by a corresponding drop in the occurrence of planets at  $\sim 20 M_{\oplus}$ . Thus it is a feature that emerges due to enhanced gas accretion onto more massive planets.

We further varied model assumptions to understand how robust the results are. First, we found them to be insensitive to any bloating mechanism (Methods). Second, under the assumption that water is buried below the H/He envelope and is forced to remain in condensed form, we do not reproduce the radius valley (Fig. 2b). Instead, water-rich cores populate the radius range where the observed valley is located. This is in agreement with previous works<sup>19,22</sup> and shows that the cause of recovering a radius valley with water-rich planets can be attributed to the (correct) phase of water and its distribution within the envelope.

As a third variation, Fig. 2c shows the case excluding atmospheric mass loss while keeping the water mixed into the H/He envelope. There, we get a distribution that is not in agreement with the presence of a radius valley. Instead, we obtain many (low-mass) large planets with a rocky core and a H/He envelope and also a distribution of water-rich planets that smoothly extend to low radii. In reality, both of these kinds of planets would be unable to retain most of their volatile envelopes. For the same reason, too few rocky planets exist. This highlights the need for atmospheric escape shaping the distribution of planetary radii even for water-rich compositions. It is necessary to populate the super-Earth peak with rocky planets by stripping their H/He envelopes. We conclude that the valley is a hybrid consequence of both formation (migration leading to the sub-Neptune peak) and evolution (evaporation leading to the super-Earth peak).

### Dependency on orbital period

In addition to radii, orbital periods of exoplanets can be determined precisely. Figure 3 shows the period–radius distribution of observed<sup>2</sup> (Fig. 3a) and modelled (Fig. 3b,c) exoplanets. Qualitatively, the observed distribution matches the synthetic distribution with applied

observational bias. We apply the bias of the full Kepler survey without taking into account that not all planets are included in the California Kepler Survey catalogue<sup>2</sup>. Therefore, more numerous points appear in the middle panel. In both the observed and the synthetic distribution, the radius valley can be made out at comparable locations.

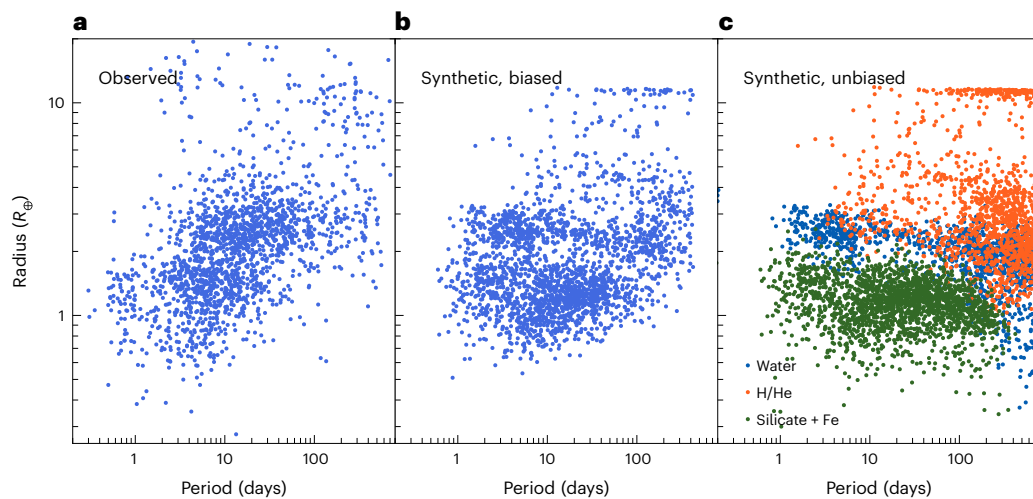
Figure 3c shows the synthetic data without observational bias applied. Additionally, the composition is colour-coded. Rocky Earths and super-Earths (green) are found at small radii below the valley. This distribution is truncated at 300 days, where the equilibrium temperature is close to 300 K.

The planets with water but without H/He (blue) fill the parameter space at radii greater than the rocky planets above the valley and at larger distances. Planets with H/He (orange) populate the even larger radii and are also more common at lower radii further from the star.

As revealed by the selected formation tracks in Extended Data Fig. 2, this pattern is shaped by migration and collisions during the formation stage as well as photoevaporation. Rocky planets generally form at short distances inside the water snowline. They grow first by planetesimal accretion and then (and most importantly) by giant impacts with other rocky protoplanets. However, their growth is limited by the number of building blocks inside the ice line. Due to their small mass, little orbital migration occurs inside the ice line.

At larger distances to the star, more solids are available for accretion, given the assumed radial power-law slope of  $-1.5$  for the planetesimal disk surface density. Also, the condensed icy material outside the water–ice line can be accreted as solid. This leads to promoted planetary growth up to several Earth masses, where migration is more efficient<sup>34</sup>. Thus, planets migrate to the inner region, where they collide with smaller rocky planets, especially when the gas disk dissipates and gas-induced eccentricity damping ceases. Collisions with other bodies—or, at the closest distances, the radiation flux of the host star—can lead to the expansion and Roche lobe overflow of the gaseous envelope and thus remove the H/He content. However, due to limitations in our model, this is not the case for the heavier water. Only after switching to the evolution stage do we assume that the constituents mix perfectly. These processes therefore give rise to a population of planets with pure water envelopes. More massive planets or those located at larger distances share a similar formation history but are not stripped completely of H/He (orange planets and tracks in Fig. 3 and Extended Data Fig. 2).

During the evolution stage, planets cool and are subject to photoevaporation, which reduces their size. The most frequent evolutionary pathway is the loss of a (pure) H/He atmosphere from a rocky core.



**Fig. 3 | Transit radii as a function of orbital period.** **a**, Observational data from the California Kepler Survey<sup>2</sup>. **b**, Synthetic data with applied observational bias using KOBE (Methods). **c**, Unbiased synthetic distribution colour-coded by composition (as in Fig. 2).

A mixed or water-dominated envelope can also be lost completely, resulting in a bare, rocky core. This happened for 17% of the (eventually) rocky super-Earths in the biased, synthetic population, which had at least 10% water by mass after the formation stage. This outcome occurs for the lightest migrating planets.

Although there is an overlap in mass between volatile-rich and rocky planets (Extended Data Fig. 3), the overall formation pathway occurs along the following lines: rocky cores are lower-mass planets that formed almost in situ by a giant impact stage, while volatile-rich planets are more massive and for that reason migrated substantially to their present-day location. Low-mass, volatile-rich planets at large distances, in contrast, do not migrate towards the host star (which would fill the valley), because type I migration is slower for lower planet mass<sup>34</sup>.

## Discussion

### Mass distribution and mass–radius diagram

A fundamental property of planets with regards to their formation is their mass. We obtained the mass distribution, shown in Extended Data Fig. 3, by modelling planet growth from the embryo stage onward via solid and gas accretion followed by long-term evolution. This forward modelling approach differs from (and allows for contrasting insights compared to) using an educated guess for the planetary masses<sup>35</sup> or retrieving mass distributions by solving the inverse problem and starting with radii found by the Kepler satellite<sup>18,35</sup>. We obtain a trend in mass from low-mass, rocky planets over more massive, migrated, steamy water worlds to the H/He-rich (sub)giants. However, the mass distributions of each kind of planet are broad and overlap, which results in a unimodal total mass distribution.

The characteristic masses are a few Earth masses for rocky planets and about ten Earth masses for water worlds. These mass scales can be understood analytically by recalling that most of the rocky planets went through a giant-impact phase after a destabilization of the systems. The appropriate mass scale then depends on the number of solid building blocks in a dynamically enhanced feeding zone during the giant-impact phase<sup>36</sup>. In contrast, for the migrated, water-rich sub-Neptunes, the equality of migration against growth or, if present, the saturation of corotation torques sets the mass scale<sup>36</sup>.

This implies that the mass of the rocky planets depends on the solid accretion mechanism, while the mass of migrated steam worlds is less sensitive to it. Independently from our work, a study<sup>15</sup> using a global model with pebble accretion instead of planetesimal accretion resulted in a mass distribution with more distinct separation of rocky and water-rich planets—that is, a bimodal distribution—in mass. An

overall mass distribution similar to ours was retrieved in a Bayesian framework<sup>18</sup> by fitting the radius distribution of Kepler planets, as shown in Extended Data Fig. 3 (dashed). This favours the unimodal distribution, although a low probability of planets containing water was inferred by the authors. We attribute it to their assumption that water is in the condensed phase at all temperatures.

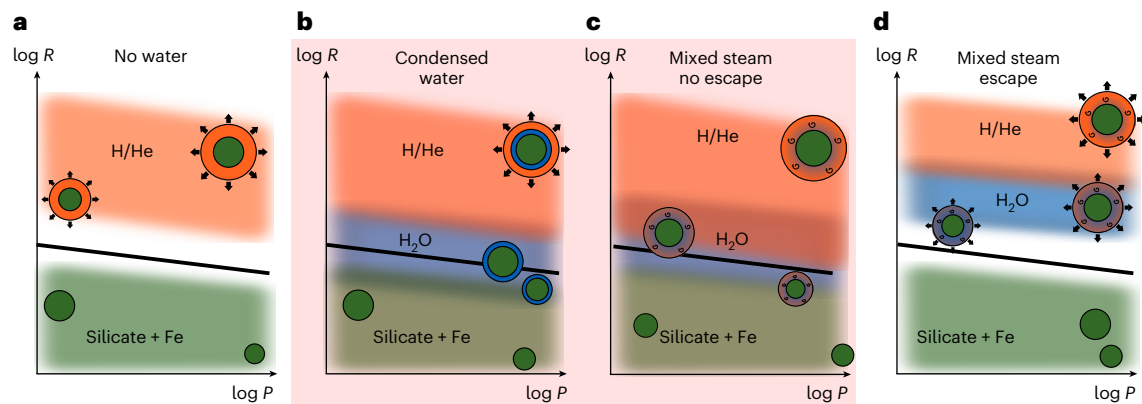
In Extended Data Fig. 4, we also show the synthetic mass–radius diagram for the nominal model. We find that our model leads to planets covering the area occupied by precisely characterized, observed planets around Solar-type stars. A tentative overdensity of both observed and simulated planets is located close to the lowest-density planets without H/He (uppermost blue points in the diagram). This would be in agreement with recent observational findings<sup>16</sup> for M stars but is for Solar-type stars without statistical significance. So far, no unbiased statistical sample of characterized planets with precise mass and radius measurements exists for a more detailed comparison.

### Potential reasons for discrepancy of close-in, wet planets

While the synthetic and observed radii match well, we find a difference in the orbital period distribution of the close-in sub-Neptunes. In Fig. 3, by comparing the observed (Fig. 3a) with the synthetic, biased (Fig. 3b) population, it becomes apparent that a group of sub-Neptunes is theoretically predicted at orbital periods of 1–5 days but is absent in the observed sample. Instead, a similar number of planets are missing at longer orbital periods. It is expected<sup>37</sup> that even at high irradiation, steam envelopes can be kept for sufficient core masses. However, with the exception of 55 Cancri e<sup>38</sup>, there are no observed planets with low bulk densities in this regime. Thus, the very close-in sub-Neptune population that we obtain is disfavoured by observations.

A possible explanation is that their migration might halt at greater orbital distances than the model predicts. Currently, the inner edge of the gas disks, placed at corotation radii of observed young T Tauri stars<sup>39</sup>, acts as migration trap. Since sub-Neptunes form early in the gaseous disks (to be able to migrate substantially), and given a sufficient source of disk turbulence in the inner disk, viscous heating would be efficient. Therefore, the location of thermal ionization—that is, the inner edge of the dead zone to the magnetorotational instability<sup>40</sup>—can lie further from the star. This introduces a migration trap<sup>41,42</sup> and can cause the distribution of migrated planets to peak at larger orbits. Future formation models should include the physical transition in the strength of magnetorotational turbulence together with tracking the heavy-element content in the gaseous phase. In addition to a potential resolution to the discrepancies, this will enable a more





**Fig. 4 | Schematic visualization of the four considered scenarios.** The location of the observed radius valley is represented as black line, and the planetary population of pure silicate and iron bodies (green), water-rich planets (blue) and H/He gas giants (orange) are shown as blurred rectangles. Example interior structure schematics are shown with indicators for mixing and escape. **a,d**, We identified in this work that the radius valley is shaped by either a mass-loss

mechanism without any water on the planets (**a**) or photoevaporation combined with water in supercritical steam envelopes (**d**), which is in better agreement with the predictions of formation theory, in particular that planets should be undergoing orbital migration. **b,c**, Neither the scenario with condensed water (**b**) nor that without photoevaporation (**c**) reproduces the observed valley, and are thus marked in red.

detailed discussion of the radius valley's slope and time dependency in the steam-world scenario for sub-Neptunes.

To summarize, by coupling a global, end-to-end planet formation model with orbital migration and *N*-body interactions to different evolution pathways, we identify two scenarios (Fig. 4a,d) that lead to a distribution of planetary radii consistent with the observed location of the radius valley<sup>1</sup>. A key aspect is the state of water, which in this work is in the correct phase—that is, typically supercritical vapour—which we assume to be mixed with H/He.

In this way, we provide theoretical support for the scenario of formation or, more specifically, gas-driven orbital migration, shaping the distribution of mostly water-rich, steam-envelope planets populating the sub-Neptune peak (Fig. 4d). Only at larger planetary radii do H/He become the dominant gaseous constituents. This is in agreement with an earlier study<sup>15</sup>, albeit the formation mechanism assumed in that study was pebble accretion. The bimodal mass distribution (below  $20 M_{\oplus}$ ) shaped by pebble accretion and its isolation mass is not required to match observed radii. Instead, a unimodal mass distribution can reproduce the observations.

Assuming our model's mass distribution is approximately correct, we can further falsify a condensed-out ice layer scenario (Fig. 4b). In any case, at the observed planets' equilibrium temperatures, water is not in the solid ice form<sup>21,25</sup>.

At the same time, we also give evidence for the necessity of an evolutionary mass-loss mechanism. Atmospheric escape is necessary to populate the super-Earth peak with evaporated rocky cores. Without evaporation, rocky planets with small H/He atmospheres that form during the gas disk stage inside the ice line would lead to low-mass planets with large radii, resulting in a radius distribution inconsistent with observations (Fig. 4c).

Our results are not distinctly sensitive to modifying the photoevaporation model or the presence of a bloating mechanism. With both orbital migration and atmospheric escape causing the radius valley, one can speak of a hybrid origin of the radius valley caused by both formation and as evolution. However, we were limited by the availability of detailed and fast models of photoevaporation of high-metallicity envelopes to further explore the effects of chemical fractionation due to mass loss or molecular cooling, which need to be developed in the future to study the sub-Neptune population in further detail<sup>29,43</sup>.

Finally, based on pure planet evolution calculations, we cannot rule out the classical picture of photoevaporation of H/He envelopes on top of exclusively rocky cores<sup>6</sup> as the only mechanism shaping the radius

distribution. However, many different planet formation models consistently predict the migration of  $\sim 10 M_{\oplus}$  water-rich planets close to the host star, a prediction already made, for example, in the first generation of population syntheses<sup>14</sup>. We show here that the resulting population of close-in, water-rich planets can agree with the observed valley. If the prediction of the presence of water-rich sub-Neptunes should turn out to be incorrect, it would call for a revision of fundamental aspects of formation theory. These aspects could, for example, be protoplanetary disk structures leading to less orbital migration<sup>44,45</sup> or a high efficiency of volatile loss during planet assembly<sup>46</sup>.

Today, observational constraints on the bulk composition of sub-Neptunes are still inconclusive<sup>47–49</sup>, but with the help of the James Webb Space Telescope and the future ARIEL mission, we expect to find more evidence for either the water-rich or the H/He-dominated composition and advocate the investigation of sub-Neptune atmospheres to resolve the mysteries of the radius valley.

## Methods

### Formation model

The formation part of the Bern model gathers the evolution of a viscously accreting gas disk, the dynamical state of the solids in the disk, the concurrent accretion of solids and gas by the protoplanets, planet–disk interactions leading to gas-driven migration and dynamical interactions between the protoplanets. Both the gas and solid disks are described by one-dimensional, axisymmetric, vertically integrated profiles. The model solves the viscous evolution equation of the gas disk<sup>50</sup>, with additional sink terms for the accretion by the protoplanets and external photoevaporation. The standard  $\alpha$  parametrization is used to compute the viscosity, while a radiative balance is used to determine the vertical structure<sup>51</sup>. Solids are assumed to be in planetesimals whose dynamical state (eccentricity and inclination) is evolved due to damping by the gas disk, self-stirring and stirring by the protoplanets<sup>52</sup>.

Planet formation follows the core accretion paradigm with the concurrent accretion of planetesimals and gas. At time zero, 100 embryos of  $0.01 M_{\oplus}$  are randomly placed in the disk with a uniform probability in the logarithm of the distance between the inner edge and 40 AU. Planetesimal accretion is assumed to occur in the oligarchic regime<sup>52</sup>, and the planetesimal radius is set to 300 m. The model solves the one-dimensional, spherically symmetric internal structure equations<sup>53</sup> to obtain either the mass or the radius of the envelope. During the early stages, gas accretion is limited by the cooling of the planet<sup>54</sup> and the envelope mass is retrieved from the structure. Cooling efficiency

improves as the planet grows, leading to a point where the gas accretion rate exceeds the supply from the disk, which is set according to the Bondi rate. Once this point is reached, the gas accretion is known and the internal structure equations are used to track the contraction of the envelope<sup>55</sup>, yielding the planet radius and luminosity.

The model prescribes gas-driven migration<sup>34</sup> plus a reduction of the corotation torques due to planet eccentricity and inclination<sup>56</sup>. Gas-induced eccentricity and inclination damping are also included<sup>57</sup>. Dynamical planet–planet interactions are tracked using the mercury N-body code<sup>58</sup>.

### Improved evolution model

The formation model is coupled to an evolution model that evolves the planets individually over Gyr timescales, which is crucial to obtain realistic planetary radii. The evolution model starts at the time when the disk has dissipated. When initializing the evolution model with the planets after the formation stage, we use the luminosity the planet had at the time of disk dispersal but the mass from the time after the complete formation stage. In a test simulation, we checked the impact of this simplification, which we had to make due to the limitations of the formation model. In agreement with a recent study<sup>59</sup>, we found negligible differences in the final planet properties compared to starting the evolution model after 100 Myr. This is due to evolution tracks converging to the same state after Gyr timescales.

In the following, we describe in detail the changes in the evolution model compared to the previous version, which was described in full detail in ref. 30. The contraction and cooling of the interior structure as well as the tidal evolution remained unchanged.

The first major change concerns the treatment of water in the interior structure model. Previously, we assumed that it is always in condensed form and resides in a pure water/ice shell between the inner iron–silicate core and a possible outer gas envelope that consists of pure H/He. In the New Generation Planetary Population Synthesis (NGPPS) series<sup>30,31</sup>, this approximation was required, as our interior structure model was not updated for mixed H/He + H<sub>2</sub>O envelopes. Instead, we now allow for all phases of water—including vapour and supercritical—to exist. The orbital distance at which the valley is observed lies closer to the star than the habitable zone, meaning the temperatures at the top of the atmospheres exceed the threshold where water could condense out. Additionally, a strong runaway greenhouse occurs in vapour atmospheres at these distances<sup>21,25</sup>. Adding the fact that H<sub>2</sub>O is also miscible in H/He on a molecular level<sup>24,25</sup>, this implies that water does not form a separate condensed layer but is mixed into a H/He + H<sub>2</sub>O vapour envelope (or a pure water vapour envelope if the planet does not contain H/He). One of the latest updates to our interior structure and planet evolution model was to include such mixed compositions<sup>60</sup>. Specifically, we mix water described with the AQUA equation of state (EoS, see also the description below)<sup>26</sup> with H/He (ref. 61). Our treatment of the water and H/He mix implies that the fraction of heavy elements  $Z$  is radially constant in the envelope. We further note that this averaged  $Z$  is used to calculate molecular opacities as a function of temperature but assuming Solar-like elemental abundances and equilibrium chemistry<sup>62</sup>.

Second, and related to the previous step, we improved the prescription for X-ray and extreme ultraviolet (XEUUV)-driven atmospheric escape of the planetary envelopes. XEUUV radiation drives an energy-limited (EL) mass loss rate<sup>5</sup> of

$$\dot{M}_{\text{esc,EL}} = \epsilon \frac{\pi F_{\text{XEUUV}} R_{\tau=2/3} R_{\text{base}}^2}{GM_{\text{tot}} K(\xi)}, \quad (1)$$

where  $F_{\text{XEUUV}}$  is the flux received in both X-ray (dominating early) and EUV wavelengths,  $R_{\text{base}}$  is the radius of the base of the ionization layer,  $R_{\tau=2/3}$  is the radius of the optical depth  $\tau = 2/3$  layer,  $G$  is the gravitational constant,  $M_{\text{tot}}$  is the mass of the planet,  $K(\xi) = 1 - \frac{3}{2\xi} - \frac{1}{2\xi^3}$ ,  $\xi = R_{\text{Roche}}/R_{\tau=2/3}$

the ratio of the planet's Roche limit to its radius and  $\epsilon$  is an escape efficiency factor.  $R_{\text{base}}$  is located in the planetary structure by equating the partial pressure to the pressure where an optical depth of one is reached for ultraviolet photons<sup>63</sup>. If this critical pressure lies exterior to the resolved envelope structure, we extrapolate using the scale height determined at 1 bar. As improvements to the NGPPS escape model<sup>19,30</sup>, the time evolution of  $F_{\text{XEUUV}}$  is updated to a tabulated model based on recent observational data<sup>64</sup>. Furthermore, we calculate escape rates for both water and H/He separately. For water, we use equation (1) with a time ( $t$ )-variable  $\epsilon(F_{\text{XEUUV}}(t))$  calculated such that the mass-loss values from a one-dimensional, chemical–hydrodynamic model applied to a pure water vapour atmosphere of an Earth-mass planet in the habitable zone<sup>28</sup> are reproduced. For the escape of H/He, we use extended tables based on a hydrodynamic model accounting for various heating (including XEUUV) and cooling mechanisms<sup>27</sup>. The two rates from water and H/He escape are summed by weight:  $\dot{M}_{\text{esc,total}} = \dot{M}_{\text{esc,H}_2\text{O}}Z + \dot{M}_{\text{esc,H/He}}(1-Z)$ , where  $Z$  is the mass fraction of H<sub>2</sub>O in the envelope. When removing mass from the envelope, we assume perfect mixing and leave  $Z$  unchanged. This choice is motivated by the fractionation<sup>29</sup> found for the pure water case<sup>28</sup> in the limit of large escape rates supported by a recent work including mixed H–H<sub>2</sub>O (ref. 43). We note, however, that this simple model cannot account for effects due to the interplay of species, such as molecular cooling by H<sub>2</sub>O (ref. 43,65), which can reduce the escape rate of H/He, and that at larger planetary masses and lower metallicity, larger escape of H over H<sub>2</sub>O is expected<sup>29</sup>.

The third and final relevant variation is in the mechanism for so-called bloating, where we use an observationally derived, empirical relation<sup>66</sup> in the nominal model. Bloating is the term used to describe an empirically found increase of planetary radii of mostly hot Jupiters over the expected theoretical values. To reproduce observations, an additional heat source in the deeper interior of the planetary structure is required. We model this as an additional luminosity added to the energy equation at the core–envelope boundary. In addition to the empirical model<sup>66</sup>, we introduce and explore variations to the bloating prescription below.

### AQUA equation of state

The AQUA equation of state is a collection of seven individual equations of state of H<sub>2</sub>O, which together cover a large domain in pressure and temperature useful to model planetary interiors<sup>26</sup>. H<sub>2</sub>O is a very peculiar molecule that has a large number of distinct solid phases and a complex phase diagram. Given its importance in industrial applications, H<sub>2</sub>O is well described by the EoS at low pressures: that is,  $P < 1$  GPa. A single EoS that incorporates the many phases of H<sub>2</sub>O and covers the necessary large range in pressure and temperature has not yet been published. All commonly used H<sub>2</sub>O EoSs that cover a large range in pressure and temperature make considerable simplifications in terms of the number of treated phases and the location of the phase transitions. AQUA thus combines seven state-of-the-art EoSs into a single tabulated EoS that covers a domain from 0.1 Pa to 400 TPa in pressure and 150 K to 10<sup>5</sup> K in temperature. References for each individual EoS can be found in the paper introducing AQUA<sup>26</sup>. Each EoS is used in a distinct region in  $P$ – $T$  space, and the included phases are (1) ice-Ih; (2) ice-II, ice-III, ice-V and ice-VI; (3) ice-VII, ice-VII\* and ice-X; (4) low-temperature gas, low-pressure liquid and low-pressure supercritical fluid; (5) higher-pressure supercritical fluid; (6) high-temperature gas, including ionization and dissociation of H<sub>2</sub>O; and (7) supercritical fluid at very high pressures, including superionic phases.

The locations of the seven regions from individual EoSs are shown in Supplementary Fig. 1. Since there are region boundaries that do not follow a physical phase transition, AQUA provides interpolated values to assure a smooth transition in all provided state parameters. The state parameters that AQUA provides for a given pressure and temperature are density  $\rho$ , adiabatic gradient  $\Delta_{\text{Ad}} = \left(\frac{\partial \ln T}{\partial \ln P}\right)_S$ , entropy  $s$ , internal energy

$u$ , bulk speed of sound  $w$ , mean molecular weight  $\mu$ , ionization fraction  $x_{\text{ion}}$ , dissociation fraction  $x_{\text{d}}$  and a phase identifier to identify the corresponding phase.

## KOBE

Kepler Observes Bern Exoplanets (KOBE)<sup>23</sup> is a program that simulates transit surveys of exoplanets. KOBE is publicly available ([github.com/exomishra/kobe](https://github.com/exomishra/kobe)). If Kepler (or TESS, PLATO and so on) was to, hypothetically, observe a synthetic population of planets then KOBE identifies those synthetic planets that would have been detected by such a survey. As a result, KOBE allows theoretical models of planet formation (such as the Bern model used here) to be compared with transit observations.

KOBE is organized in three sequential modules. The first module, KOBE-Shadow, finds transiting planets from a synthetic population of planets given their orbital elements. A planet can transit only when its orbit is aligned with respect to a hypothetical observer's line of sight. KOBE-Transits, the next module, calculates transit parameters. It applies the detection biases coming from physical limitations; large planets in tight orbits around a quiet star are strongly favoured. For comparison to Kepler, transiting planets that transit at least three times and have a signal-to-noise ratio  $\geq 7.1$  are potentially detectable. KOBE-Vetter, the final module, applies the completeness and reliability of the Kepler pipeline by emulating Kepler's Robovetter<sup>67</sup>. Transiting planets that are identified as planetary candidates by KOBE-Vetter make up the KOBE-biased catalogue that we use for comparison to the Kepler-based California Kepler Survey results<sup>2</sup>. The planets in the KOBE catalogue are comparable to the exoplanet population discovered by Kepler.

## Impact of the evaporation model

Here, we investigate the effect of several key model assumptions on the final planet properties. We use the same planetary population after the formation stage but change the used description for the effects of photoevaporation and bloating of the planetary envelopes.

As a first test, we see that removing the effect of photoevaporation results in a distribution of radii without a valley. Photoevaporation is a necessary process to populate the envelope-stripped, super-Earth peak (Fig. 2), with planets located close to or within the observed radius valley otherwise. Furthermore, without photoevaporation, planets in the super-Earth regime would commonly contain large amounts of water, which is not realistic at these planetary masses and distances to the star<sup>37</sup>.

To model photoevaporation, we nominally used the tabulated evaporation rates based on recent detailed radiation–hydrodynamic models for H/He-dominated<sup>27</sup> and water-dominated<sup>28</sup> envelopes. To contrast them to a more simple model, we show in Supplementary Fig. 2b the resulting distribution of planetary radii using energy- and radiation-recombination-limited escape rates<sup>519</sup> and further, in Supplementary Fig. 2c, a variant thereof that accounts for an increase in mean molecular weight as a function of the envelope metallicity  $Z$  when calculating the base layer of the evaporative flow and a scaling of the efficiency factor  $\epsilon$  with  $Z$  in equation (1) taken from ref. 37. While the former gives results similar to the nominal evaporation model, the  $Z$ -dependent version results in a less pronounced super-Earth peak as fewer planets lost their primordial envelope.

For reference, we also show the resulting planetary masses in Supplementary Fig. 3 for the same model variations. A shifted overall peak at higher masses is found in mass space when removing the effect of photoevaporation (Supplementary Fig. 3f) but no significant differences result based on the different evaporation model.

## Impact of bloating

Since bloating of the envelope leads to an increase of the planetary radius and therefore a larger surface area irradiated by the star, positive feedback between bloating and photoevaporation of the planets is expected. This warrants exploring an alternative model based on the physical potential temperature advection mechanism<sup>68</sup> instead

of the nominal, empirically derived bloating model used in the main analysis<sup>66</sup>. We further investigate the effects on the population of synthetic planets in the absence of bloating. The latter exploration is interesting if the so-far-unknown bloating mechanism is not acting on lower planetary masses in the super-Earth and sub-Neptune regimes and would instead apply only to giant planets.

## Bloating model using a fit to the potential temperature advection model

To include potential temperature advection as a bloating mechanism<sup>68</sup>, we add a luminosity term at the core–mantle boundary in the interior structure calculation. We start with a literature fit of the temperature at the layer where a pressure of 100 bar is reached. The corresponding temperature is a function of the stellar irradiation flux  $F$  (ref. 68). We further subtract 200 K, motivated by the results from three-dimensional simulations<sup>69</sup>. We then calculate the entropy and, using our interior structure models, the required bloating luminosity  $L_{\text{bloat}}$  to match this entropy level.

To take into account the mass fraction of heavy molecules (that is, water) in the envelope  $Z$ , we use a  $Z$ -dependent fit of the form

$$L_{\text{bloat}}(F, Z) = a(Z) + b(Z)\log_{10}F + c(Z)\log_{10}^2F, \quad (2)$$

where  $F$  is the irradiation flux. To obtain results for continuous values of  $Z$ , we interpolate linearly between the fits, and we further found a linear dependency of the bloating luminosity on planetary mass to obtain the desired entropy. The resulting parameters for equation (2) are listed in Supplementary Table 1.

## Inconclusive evidence between bloating models

In Supplementary Fig. 2, we show resulting distributions of planetary radii for the nominal, empirical bloating model<sup>66</sup> (Supplementary Fig. 2a), for the model with radius inflation caused by the potential temperature advection mechanism<sup>68</sup> (Supplementary Fig. 2d) and without any bloating mechanism (Supplementary Fig. 2e). If included, we assume that the bloating mechanism is acting on all planetary masses, and we see a small impact of the bloating mechanism on the sub-Neptune population (blue histogram). It is expected that bloating by ohmic dissipation is efficient in sub-Neptunes<sup>70</sup>. Without a bloating mechanism, the valley would be more populated by smaller water-rich planets, therefore making it less deep and visible. However, the effect is too small to be constrained by the current observational data. Both compared bloating models give very similar results. Overall, we find that the presence of a radius valley is robust against uncertain extrapolations of the radius-inflation mechanism to lower planetary masses and different compositions.

## Data availability

Raw data used in this study are accessible at <https://dace.unige.ch> under identifier NG76. Derived data supporting the findings of this study are publicly available at <https://doi.org/10.5281/zenodo.7646318>. Supplementary data are available upon reasonable request.

## Code availability

The code for the statistical analysis and to generate all figures of the manuscript is available at <https://doi.org/10.5281/zenodo.7646318>. The modified planetary evolution code will be published in open access in the future and is available on reasonable request.

## References

1. Fulton, B. J. et al. The California- Kepler Survey. III. A gap in the radius distribution of small planets. *Astron. J.* **154**, 109 (2017).
2. Fulton, B. J. & Petigura, E. A. The California-Kepler Survey. VII. Precise planet radii leveraging Gaia DR2 reveal the stellar mass dependence of the planet radius gap. *Astron. J.* **156**, 264 (2018).



3. Ho, C. S. K. & Van Eylen, V. A deep radius valley revealed by Kepler short cadence observations. *Mon. Not. R. Astron. Soc.* **519**, 4056–4073 (2023).
4. Lopez, E. D. & Fortney, J. J. The role of core mass in controlling evaporation: the Kepler radius distribution and the Kepler-36 density dichotomy. *Astrophys. J.* **776**, 2 (2013).
5. Jin, S. et al. Planetary population synthesis coupled with atmospheric escape: a statistical view of evaporation. *Astrophys. J.* **795**, 65 (2014).
6. Owen, J. E. Atmospheric escape and the evolution of close-in exoplanets. *Annu. Rev. Earth Planet. Sci.* **47**, 67–90 (2019).
7. Ginzburg, S., Schlichting, H. E. & Sari, R. Super-Earth atmospheres: self-consistent gas accretion and retention. *Astrophys. J.* **825**, 29 (2016).
8. Gupta, A. & Schlichting, H. E. Sculpting the valley in the radius distribution of small exoplanets as a by-product of planet formation: the core-powered mass-loss mechanism. *Mon. Not. R. Astron. Soc.* **487**, 24–33 (2019).
9. Rogers, J. G., Gupta, A., Owen, J. E. & Schlichting, H. E. Photoevaporation versus core-powered mass-loss: model comparison with the 3D radius gap. *Mon. Not. R. Astron. Soc.* **508**, 5886–5902 (2021).
10. Kite, E. S., Fegley Jr, B., Schaefer, L. & Ford, E. B. Superabundance of exoplanet sub-Neptunes explained by fugacity crisis. *Astrophys. J.* **887**, L33 (2019).
11. Zeng, L. et al. Growth model interpretation of planet size distribution. *Proc. Natl Acad. Sci. USA* **116**, 9723–9728 (2019).
12. Mousis, O. et al. Irradiated ocean planets bridge super-Earth and sub-Neptune populations. *Astrophys. J.* **896**, L22 (2020).
13. Aguichine, A., Mousis, O., Deleuil, M. & Marcq, E. Mass-radius relationships for irradiated ocean planets. *Astrophys. J.* **914**, 84 (2021).
14. Mordasini, C., Alibert, Y., Benz, W. & Naef, D. Extrasolar planet population synthesis II. Statistical comparison with observations. *Astron. Astrophys.* **501**, 1161–1184 (2009).
15. Venturini, J., Guilera, O. M., Haldemann, J., Ronco, M. P. & Mordasini, C. The nature of the radius valley. *Astron. Astrophys.* **643**, L1 (2020).
16. Luque, R. & Pallé, E. Density, not radius, separates rocky and water-rich small planets orbiting M dwarf stars. *Science* **377**, 1211–1214 (2022).
17. Rogers, J. G., Schlichting, H. E. & Owen, J. E. Conclusive evidence for a population of water worlds around M dwarfs remains elusive. *Astrophys. J.* **947**, L19 (2023).
18. Rogers, J. G. & Owen, J. E. Unveiling the planet population at birth. *Mon. Not. R. Astron. Soc.* **503**, 1526–1542 (2021).
19. Jin, S. & Mordasini, C. Compositional imprints in density–distance–time: a rocky composition for close-in low-mass exoplanets from the location of the valley of evaporation. *Astrophys. J.* **853**, 163 (2018).
20. Izidoro, A. et al. The exoplanet radius valley from gas-driven planet migration and breaking of resonant chains. *Astrophys. J.* **939**, L19 (2022).
21. Turbet, M. et al. Revised mass-radius relationships for water-rich rocky planets more irradiated than the runaway greenhouse limit. *Astron. Astrophys.* **638**, A41 (2020).
22. Owen, J. E. & Wu, Y. The evaporation valley in the Kepler planets. *Astrophys. J.* **847**, 29 (2017).
23. Mishra, L. et al. The New Generation Planetary Population Synthesis (NGPPS) VI. Introducing KOBE: Kepler observes Bern exoplanets—theoretical perspectives on the architecture of planetary systems: peas in a pod. *Astron. Astrophys.* **656**, A74 (2021).
24. Vazan, A., Sari, R. & Kessel, R. A new perspective on the interiors of ice-rich planets: ice-rock mixture instead of ice on top of rock. *Astrophys. J.* **926**, 150 (2022).
25. Pierrehumbert, R. T. The runaway greenhouse on sub-Neptune waterworlds. *Astrophys. J.* **944**, 20 (2023).
26. Haldemann, J., Alibert, Y., Mordasini, C. & Benz, W. AQUA: a collection of H<sub>2</sub>O equations of state for planetary models. *Astron. Astrophys.* **643**, A105 (2020).
27. Kubyskhina, D. I. & Fossati, L. Extending a grid of hydrodynamic planetary upper atmosphere models. *Res. Notes AAS* **5**, 74 (2021).
28. Johnstone, C. P. Hydrodynamic escape of water vapor atmospheres near very active stars. *Astrophys. J.* **890**, 79 (2020).
29. Lammer, H. et al. Loss and fractionation of noble gas isotopes and moderately volatile elements from planetary embryos and early Venus, Earth and Mars. *Space Sci. Rev.* **216**, 74 (2020).
30. Emsenhuber, A. et al. The New Generation Planetary Population Synthesis (NGPPS) I. Bern global model of planet formation and evolution, model tests, and emerging planetary systems. *Astron. Astrophys.* **656**, A69 (2021).
31. Emsenhuber, A. et al. The New Generation Planetary Population Synthesis (NGPPS) II. Planetary population of solar-like stars and overview of statistical results. *Astron. Astrophys.* **656**, A70 (2021).
32. Tychoniec, Ł. et al. The VLA Nascent Disk and Multiplicity Survey of Perseus Protostars (VANDAM). IV. Free–free emission from protostars: links to infrared properties, outflow tracers, and protostellar disk masses. *Astrophys. J. Suppl. Ser.* **238**, 19 (2018).
33. Izidoro, A. et al. Breaking the chains: hot super-Earth systems from migration and disruption of compact resonant chains. *Mon. Not. R. Astron. Soc.* **470**, 1750–1770 (2017).
34. Paardekooper, S.-J., Baruteau, C. & Kley, W. A torque formula for non-isothermal Type I planetary migration—II. Effects of diffusion. *Mon. Not. R. Astron. Soc.* **410**, 293–303 (2011).
35. Gupta, A. & Schlichting, H. E. Signatures of the core-powered mass-loss mechanism in the exoplanet population: dependence on stellar properties and observational predictions. *Mon. Not. R. Astron. Soc.* **493**, 792–806 (2020).
36. Emsenhuber, A., Mordasini, C. & Burn, R. Planetary population synthesis and the emergence of four classes of planetary system architectures. *Eur. Phys. J. Plus* **138**, 181 (2023).
37. Lopez, E. D. Born dry in the photoevaporation desert: Kepler’s ultra-short-period planets formed water-poor. *Mon. Not. R. Astron. Soc.* **472**, 245–253 (2017).
38. Fischer, D. A. et al. Five planets orbiting 55 Cancri. *Astrophys. J.* **675**, 790–801 (2008).
39. Venuti, L. et al. CSI 2264: investigating rotation and its connection with disk accretion in the young open cluster NGC 2264. *Astron. Astrophys.* **599**, A23 (2017).
40. Flock, M., Fromang, S., Turner, N. J. & Benisty, M. Radiation hydrodynamics models of the inner rim in protoplanetary disks. *Astrophys. J.* **827**, 144 (2016).
41. Mohanty, S., Jankovic, M. R., Tan, J. C. & Owen, J. E. Inside-out planet formation. V. Structure of the inner disk as implied by the MRI. *Astrophys. J.* **861**, 144 (2018).
42. Ataiee, S. & Kley, W. Pushing planets into an inner cavity by a resonant chain. *Astron. Astrophys.* **648**, A69 (2021).
43. Yoshida, T., Terada, N., Ikoma, M. & Kuramoto, K. Less effective hydrodynamic escape of H<sub>2</sub>–H<sub>2</sub>O atmospheres on terrestrial planets orbiting pre-main-sequence M dwarfs. *Astrophys. J.* **934**, 137 (2022).
44. Coleman, G. A. L. & Nelson, R. P. Giant planet formation in radially structured protoplanetary discs. *Mon. Not. R. Astron. Soc.* **460**, 2779–2795 (2016).
45. Ogiwara, M., Kokubo, E., Suzuki, T. K. & Morbidelli, A. Formation of close-in super-Earths in evolving protoplanetary disks due to disk winds. *Astron. Astrophys.* **615**, A63 (2018).
46. Lichtenberg, T. et al. A water budget dichotomy of rocky protoplanets from 26Al-heating. *Nat. Astron.* **3**, 307–313 (2019).
47. Kasper, D. et al. Nondetection of helium in the upper atmospheres of three sub-Neptune exoplanets. *Astron. J.* **160**, 258 (2020).



48. Zhang, M. et al. Detection of atmospheric escape from four young mini-Neptunes. *Astron. J.* **165**, 62 (2023).
49. Cherubim, C. et al. TOI-1695 b: a water world orbiting an early-M dwarf in the planet radius valley. *Astron. J.* **165**, 167 (2023).
50. Lynden-Bell, D. & Pringle, J. E. The evolution of viscous discs and the origin of the nebular variables. *Mon. Not. R. Astron. Soc.* **168**, 603–637 (1974).
51. Nakamoto, T. & Nakagawa, Y. Formation, early evolution, and gravitational stability of protoplanetary disks. *Astrophys. J.* **421**, 640–650 (1994).
52. Fortier, A., Alibert, Y., Carron, F., Benz, W. & Dittkrist, K.-M. Planet formation models: the interplay with the planetesimal disc. *Astron. Astrophys.* **549**, A44 (2013).
53. Bodenheimer, P. & Pollack, J. B. Calculations of the accretion and evolution of giant planets: the effects of solid cores. *Icarus* **67**, 391–408 (1986).
54. Lee, E. J. & Chiang, E. To cool is to accrete: analytic scalings for nebular accretion of planetary atmospheres. *Astrophys. J.* **811**, 41 (2015).
55. Bodenheimer, P., Hubickyj, O. & Lissauer, J. J. Models of the in situ formation of detected extrasolar giant planets. *Icarus* **143**, 2–14 (2000).
56. Coleman, G. A. L. & Nelson, R. P. On the formation of planetary systems via oligarchic growth in thermally evolving viscous discs. *Mon. Not. R. Astron. Soc.* **445**, 479–499 (2014).
57. Cresswell, P. & Nelson, R. P. Three-dimensional simulations of multiple protoplanets embedded in a protostellar disc. *Astron. Astrophys.* **482**, 677–690 (2008).
58. Chambers, J. E. A hybrid symplectic integrator that permits close encounters between massive bodies. *Mon. Not. R. Astron. Soc.* **304**, 793–799 (1999).
59. Affolter, L., Mordasini, C., Oza, A. V., Kubyshkina, D. & Fossati, L. Planetary evolution with atmospheric photoevaporation II: fitting the slope of the radius valley by combining boil-off and XUV-driven escape. *Astron. Astrophys.* **676**, A119 (2023).
60. Linder, E. F. et al. Evolutionary models of cold and low-mass planets: cooling curves, magnitudes, and detectability. *Astron. Astrophys.* **623**, A85 (2019).
61. Chabrier, G., Mazevet, S. & Soubiran, F. A new equation of state for dense hydrogen-helium mixtures. *Astrophys. J.* **872**, 51 (2019).
62. Freedman, R. S. et al. Gaseous mean opacities for giant planet and ultracool dwarf atmospheres over a range of metallicities and temperatures. *Astrophys. J. Suppl. Ser.* **214**, 25 (2014).
63. Murray-Clay, R. A., Chiang, E. I. & Murray, N. Atmospheric escape from hot Jupiters. *Astrophys. J.* **693**, 23–42 (2009).
64. McDonald, G. D., Kreidberg, L. & Lopez, E. The sub-Neptune desert and its dependence on stellar type: controlled by lifetime X-ray irradiation. *Astrophys. J.* **876**, 22 (2019).
65. Koskinen, T. T. et al. Mass loss by atmospheric escape from extremely close-in planets. *Astrophys. J.* **929**, 52 (2022).
66. Sarkis, P., Mordasini, C., Henning, T., Marleau, G. D. & Mollière, P. Evidence of three mechanisms explaining the radius anomaly of hot Jupiters. *Astron. Astrophys.* **645**, A79 (2021).
67. Thompson, S. E. et al. Planetary candidates observed by Kepler. VIII. A fully automated catalog with measured completeness and reliability based on data release 25. *Astrophys. J. Suppl. Ser.* **235**, 38 (2018).
68. Tremblin, P. et al. Advection of potential temperature in the atmosphere of irradiated exoplanets: a robust mechanism to explain radius inflation. *Astrophys. J.* **841**, 30 (2017).
69. Sainsbury-Martinez, F. et al. Idealised simulations of the deep atmosphere of hot Jupiters. Deep, hot adiabats as a robust solution to the radius inflation problem. *Astron. Astrophys.* **632**, A114 (2019).
70. Pu, B. & Valencia, D. Ohmic dissipation in mini-Neptunes. *Astrophys. J.* **846**, 47 (2017).

## Acknowledgements

We thank L. Kreidberg (Max-Planck-Institut für Astronomie) for fruitful discussion. We acknowledge support from the the European Research Council under the European Union's Horizon 2020 research and innovation programme under grant no. 832428-Origins (R.B. and T.H.); from DFG under Germany's Excellence Strategy EXC 2181/1-390900948, Exploratory project EP 8.4 (the Heidelberg STRUCTURES Excellence Cluster, R.B.); and from the Swiss National Science Foundation under grant nos. 200021\_204847 'PlanetsInTime' (C.M. and A.E.) and PZ00P2\_208945 (J.V.). Parts of this work were carried out within the framework of the NCCR PlanetS supported by the Swiss National Science Foundation under grant nos. 51NF40\_182901 and 51NF40\_205606. Calculations were performed on the Horus cluster at the University of Bern.

## Author contributions

R.B. analysed the synthetic data and wrote the manuscript with contributions from all other authors. C.M., L.M., R.B. and A.E. ran the numerical experiments. C.M. and A.E. are the lead developers of the numerical code. L.M. developed the code for comparing theory with observations. Most of L.M.'s work was done at the University of Bern and the University of Geneva. J.H., J.V. and C.M. integrated the updated EoS. C.M. conceived the original idea. T.H. and C.M. supervised the project.

## Funding

Open access funding provided by Max Planck Society.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41550-023-02183-7>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41550-023-02183-7>.

**Correspondence and requests for materials** should be addressed to Remo Burn or Christoph Mordasini.

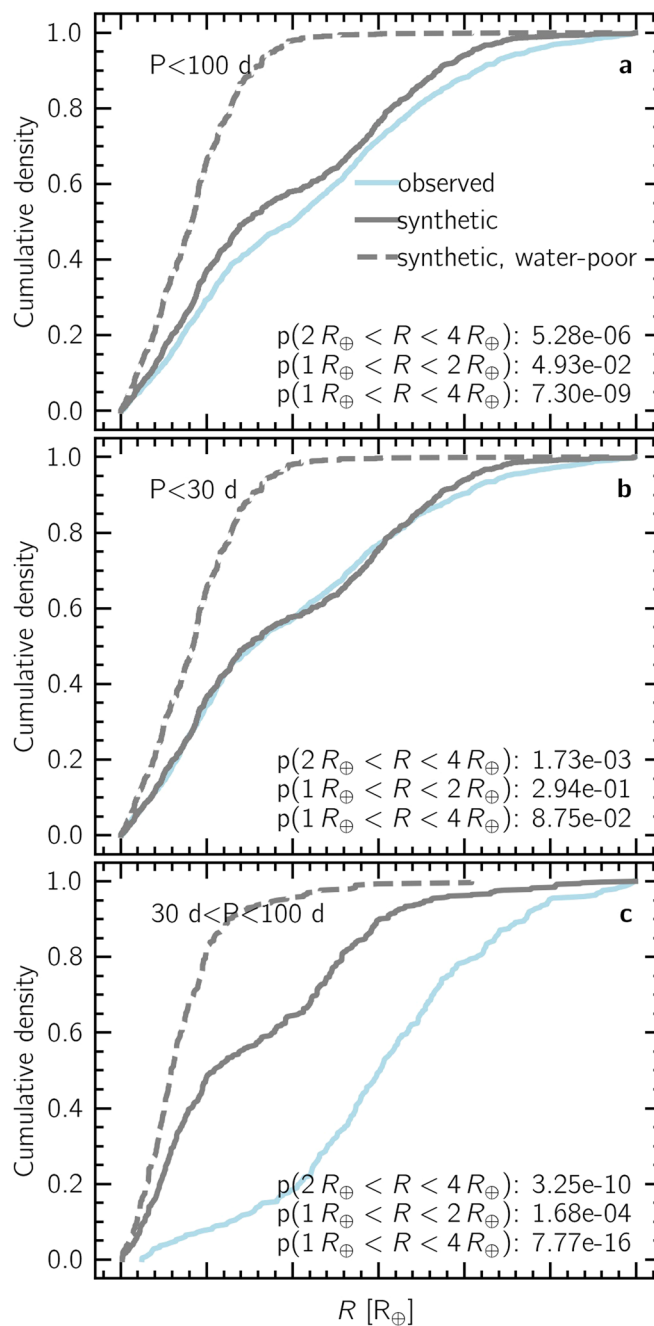
**Peer review information** *Nature Astronomy* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

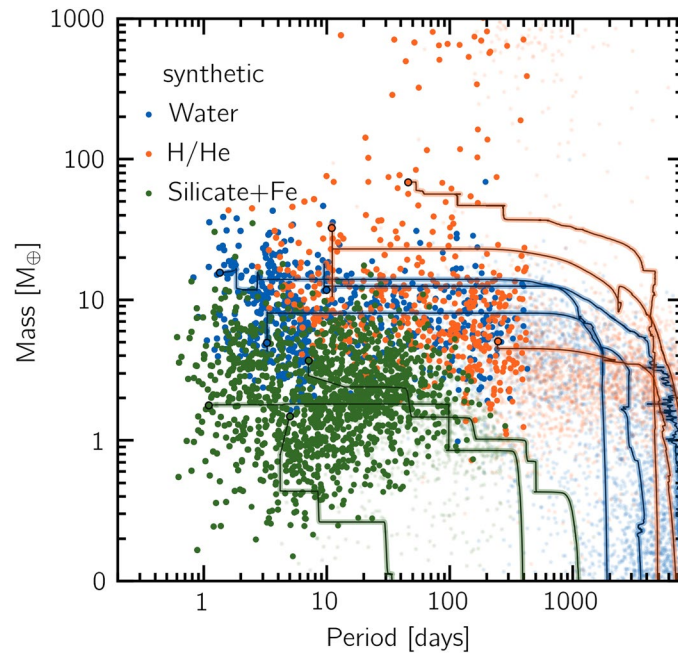
**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024



**Extended Data Fig. 1 | Cumulative radius distributions of observed and nominal, synthetic planets ( $R < 4 R_{\oplus}$ ) in different orbital regions. a–c, Distributions at different orbital period ranges marked in the top left. Two-sample, Kolmogorov-Smirnov (KS) test p-values are listed in the bottom right. The null hypothesis that the observed and the synthetic distributions are the same can be rejected for the region outside of 30 day orbital periods and for**

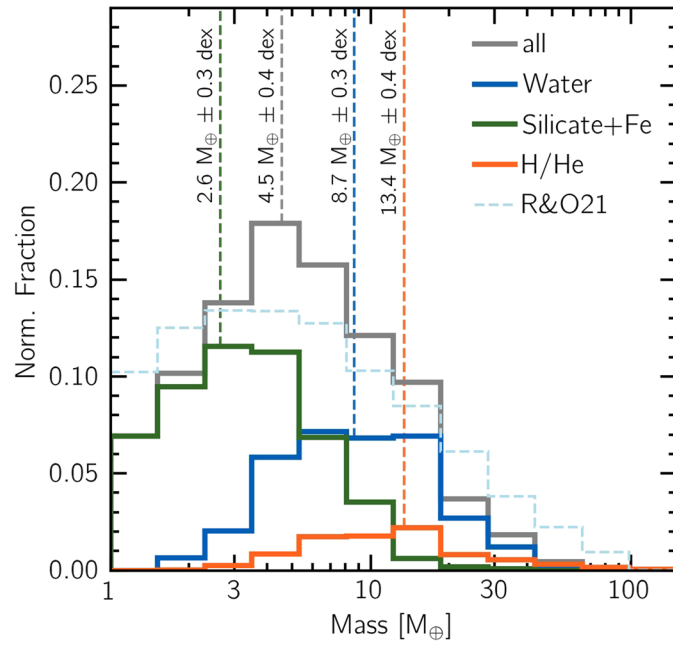
all sub-Neptunes but not for close-in super-Earths (p-values above 0.05). This agreement leads to a p-value above 0.05 for the whole radius range if only planets on short orbits are considered. In clear disagreement with observations is the synthetic sub-sample of planets with less than 50% water mass content in the envelope (dashed).



**Extended Data Fig. 2 | Orbital period against masses of synthetic planets after 5 Gyr of evolution.** Planets observable and non-observable with the Kepler space telescope are shown (KOBÉ bias applied) as opaque or transparent points respectively. The points are colored as in Fig. 3 and we show in addition the

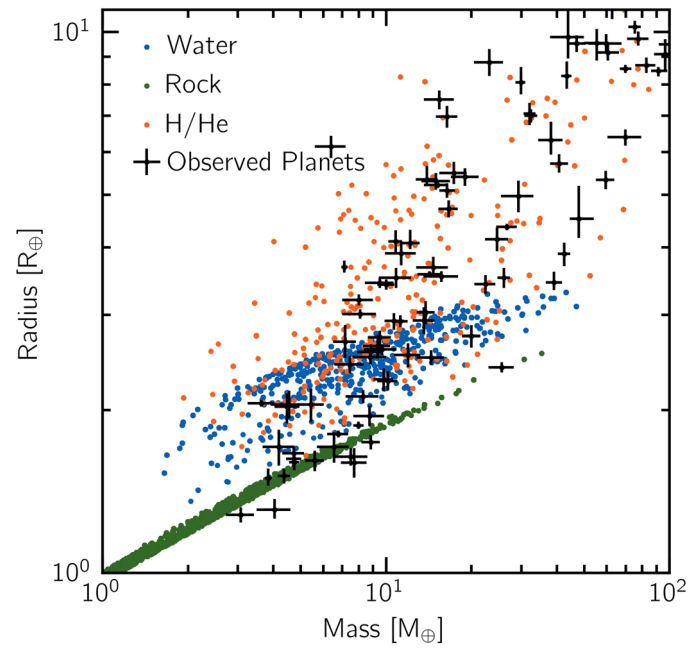
formation tracks of a selection of planets from the different categories as lines. Sudden jumps in mass are caused by giant impacts. Since there is no effect of the lower density material on the planetary mass, the radius valley is not visible in mass space.





**Extended Data Fig. 3 | Mass histogram of the nominal simulation with applied Kepler bias.** Lines show different groups: all planets (gray), rocky planets (only iron core and silicate mantle, green), planets with water but no H/He (blue), and planets with H/He (orange). Dashed lines mark logarithmic mean values of each

category and are labeled with mean and standard deviations. The distribution of core masses retrieved by Rogers and Owen<sup>18</sup> (R&O21) in a pure photoevaporation scenario is shown for comparison.



**Extended Data Fig. 4 | Total planetary masses against planetary radii of observed and synthetic planets.** The observational data was taken from the NASA Exoplanet Archive (composite data, accessed on Dec 9, 2022) filtering

for relative mass, stellar mass, and radius standard errors smaller than 20% and stellar masses ranging from  $0.75$  to  $1.25 M_{\odot}$ . Errorbars show the standard error. We caution that this sample of planets does not come from a homogeneous survey.