LETTER TO THE EDITOR

The nature of the radius valley

Hints from formation and evolution models

Julia Venturini¹, Octavio M. Guilera^{2,4,5}, Jonas Haldemann³, María P. Ronco^{4,5}, and Christoph Mordasini³

¹ International Space Science Institute, Hallerstrasse 6, 3012 Bern, Switzerland
e-mail: julia.venturini@issibern.ch

² Instituto de Astrofísica de La Plata, CCT La Plata-CONICET-UNLP, Paseo del Bosque s/n (1900), La Plata, Argentina

³ Department of Space Research & Planetary Sciences, University of Bern, Gesellschaftsstrasse 6, 3012 Bern, Switzerland

⁴ Instituto de Astrofísica, Pontificia Universidad Católica de Chile, Santiago, Chile

⁵ Millennium Nucleus of Planetary Formation (NPF), Valparaíso, Chile

Received 10 August 2020 / Accepted 4 October 2020

ABSTRACT

The existence of a radius valley in the *Kepler* size distribution stands as one of the most important observational constraints to understand the origin and composition of exoplanets with radii between those of Earth and Neptune. In this work we provide insights into the existence of the radius valley, first from a pure formation point of view and then from a combined formation-evolution model. We run global planet formation simulations including the evolution of dust by coagulation, drift, and fragmentation, and the evolution of the gaseous disc by viscous accretion and photoevaporation. A planet grows from a moon-mass embryo by either silicate or icy pebble accretion, depending on its position with respect to the water ice line. We include gas accretion, type I–II migration, and photoevaporation driven mass-loss after formation. We perform an extensive parameter study evaluating a wide range of disc properties and initial locations of the embryo. We find that due to the change in dust properties at the water ice line, rocky cores form typically with $\sim 3 M_{\oplus}$ and have a maximum mass of $\sim 5 M_{\oplus}$, while icy cores peak at $\sim 10 M_{\oplus}$, with masses lower than $5 M_{\oplus}$ being scarce. When neglecting the gaseous envelope, the formed rocky and icy cores account naturally for the two peaks of the *Kepler* size distribution. The presence of massive envelopes yields planets more massive than $\sim 10 M_{\oplus}$ with radii above $4 R_{\oplus}$. While the first peak of the *Kepler* size distribution is undoubtedly populated by bare rocky cores, as shown extensively in the past, the second peak can host half-rock–half-water planets with thin or non-existent H-He atmospheres, as suggested by a few previous studies. Some additional mechanisms inhibiting gas accretion or promoting envelope mass-loss should operate at short orbital periods to explain the presence of ~ 10 – $40 M_{\oplus}$ planets falling in the second peak of the size distribution.

Key words. planets and satellites: formation – planets and satellites: atmospheres – planets and satellites: composition

1. Introduction

The California-*Kepler* Survey revealed that exoplanets within a 100-day orbital period present a bimodal size distribution, with peaks at ~ 1.3 and $\sim 2.4 R_{\oplus}$ (Fulton et al. 2017). More recent analysis of better characterised sub-samples showed the peaks at ~ 1.5 and $\sim 2.7 R_{\oplus}$, and the valley or gap at ~ 1.9 – $2 R_{\oplus}$ (Van Eylen et al. 2018; Martinez et al. 2019; Petigura 2020).

The valley can be explained by atmospheric mass-loss mechanisms, such as photoevaporation (e.g. Owen & Wu 2017; Jin & Mordasini 2018) or core-powered mass-loss (e.g. Ginzburg et al. 2018; Gupta & Schlichting 2019). Both models are able to reproduce the correct position of the valley only if the naked cores resulting from the mass-loss are rocky in composition. This has led to the interpretation that most *Kepler* planets with radii between Earth and Neptune accreted only dry condensates, and were therefore formed within the water ice line (Owen & Wu 2017; Gupta & Schlichting 2019).

From a formation point of view, it is hard to envision scenarios where planets with masses below $20 M_{\oplus}$ are devoid of water. Accretion beyond the ice line is usually prominent, and type I migration tends to move planets in the mass range ~ 1 – $20 M_{\oplus}$ inwards in a very effective way (e.g. Tanaka et al. 2002). Hence,

a pure dry core composition for most short-period exoplanets is not really expected from formation models (Raymond et al. 2018; Bitsch et al. 2019; Brügger et al. 2020). A possible way out is to invoke migration traps due to the existence of dead zones in the disc (Alessi et al. 2020). However, even if the super-Earths produced by those models are dry, they cannot account for the *Kepler* radius bimodality.

Recent studies, based on mass-radius relations, suggest instead that only the first peak of the radius distribution corresponds to rocky planets, while the second are water-rich objects (Zeng et al. 2019). The problem with associating the second peak with water-rich planets is that it cannot explain why such planets do not fill the valley. Cores containing 50% rock–50% ice by mass would fall in the radius valley if they had a mass of ~ 3 – $6 M_{\oplus}$ (Sotin et al. 2007; Zeng et al. 2019; Haldemann et al. 2020; Owen & Wu 2017; Gupta & Schlichting 2019). Zeng et al. (2019) showed that the *Kepler* size distribution can be matched if the icy planets are assumed to follow the mass distribution suggested by RV measurements, which encompasses masses in the range ~ 6 – $15 M_{\oplus}$, with a peak at $\sim 9 M_{\oplus}$. However, no explanation for the origin of this mass distribution is offered.

In an accompanying paper (Venturini et al. 2020, hereafter Paper I) we show that when pebble accretion is computed

self-consistently from dust growth and evolution models, pure rocky planets are typically less massive than $5 M_{\oplus}$. In that work we also show that the change in dust properties at the ice line affects dramatically the growth mode of planets, which was originally proposed by Morbidelli et al. (2015) to explain the dichotomy of gaseous versus terrestrial planets in the Solar System. In this letter we show that a bimodality in core mass and composition from birth naturally renders a radius valley at $\sim 1.5\text{--}2 R_{\oplus}$. We also discuss the effect of gaseous envelopes and their photoevaporation on the *Kepler* size bimodality.

2. Methodology in brief

Our physical model is the same as in Paper I, except that planets are always allowed to migrate. We recall it here briefly. An embryo grows from lunar mass by pebble and gas accretion, embedded in an α -disc that undergoes X-ray photoevaporation from the central star. The adopted α -values are 10^{-3} and 10^{-4} . The pebble surface density is computed self-consistently from dust coagulation, fragmentation, drift, and ice sublimation at the water ice line (Barnes et al. 2011; Drążkowska et al. 2016; Guilera et al. 2020). We consider the growth of one embryo per disc, which accretes either rocky or icy pebbles, depending on its position with respect to the water ice line. The fragmentation threshold velocity of icy pebbles is taken as $v_{\text{th}} = 10 \text{ m s}^{-1}$ and $v_{\text{th}} = 1 \text{ m s}^{-1}$ for rocky ones (see Paper I and Appendix A for a discussion about this choice). In this work ‘rocky’ means Earth-like composition (i.e. 1/3 iron and 2/3 silicates by mass).

Gas accretion is computed in the attached and the detached phases. To reduce computational time in the attached phase, the interior structure of the planets is calculated using the method presented in Alibert & Venturini (2019), which uses deep neural networks, trained on pre-computed structure models. Before the core reaches the pebble isolation mass (when $M_{\text{core}} = 0.9 M_{\text{iso}}$), we switch to solving the internal structure equations to capture the increase of gas accretion resulting from the halt of pebble accretion (see Paper I). Type I migration prescriptions account for the possibility of outwards migration due to corotation (Jiménez & Masset 2017) and thermal torques (Masset 2017). Planets switch to type II migration once a partial gap opens in the disc (Crida et al. 2006). We perform a total of 665 planet formation simulations, spanning a wide range of initial conditions and disc properties, as detailed in Appendix B.

Once the disc dissipates, the final planetary radius is computed after 5 Gyr of cooling and photoevaporation by solving the internal structure equations and checking where the semi-grey atmosphere becomes optically thick (see details in Paper I). We employ two photoevaporation models. In Model A the water is assumed in the form of ice, mixed with the rocks in the planetary core. A H-He envelope lies on top and undergoes mass loss. In Model B the water is assumed in the form of vapour and uniformly mixed with the H-He, forming a H-He-H₂O envelope, with all its compounds affected by the mass loss (see details in Sect. 2.1.2 of Mordasini 2020). For the cases where we neglect the presence of the gaseous envelope, the planetary radius is computed following Zeng et al. (2019, see Methods), who provide a power-law mass–radius relation determined by the mass of rocks (assumed Earth-like in composition) and water.

3. Results

The water ice line splits a protoplanetary disc into two distinct growth environments. This is because fragmentation renders silicate pebbles considerably smaller than icy ones (see

Fig. 1. *Top panel:* formation tracks corresponding to disc 1 (see Appendix B), $Z_0 = 0.0144$, and $\alpha = 10^{-4}$. The white and green circles indicate the times 0.012 Myr, 0.25 Myr, and 2 Myr. The 0.25 Myr circle of the core starting its formation just inside the ice line is below the 2 Myr circle. M_{iso} is reached in each simulation when M_{core} stops growing. The core growing in a vertical line grows so fast that it practically does not migrate before reaching M_{iso} . *Bottom panel:* evolution of the Stokes number at the planet location for the seven cases shown in the top panel. The labels indicate the initial semi-major axis. The grey circles show the time when planets enter the region $r < r_{\text{ice}}$.

Paper I), resulting in an increase in Stokes number at the water ice line (Morbidelli et al. 2015). In Fig. 1a we illustrate this effect, showing the growth tracks of seven planetary embryos that form in the same disc (one at a time). Three embryos start their growth within the ice line and four beyond. The colour-bar indicates the ice mass fraction of the core. The planets that start forming beyond the ice line always remain water-rich ($f_{\text{ice}} \approx 0.5$) because they grow quickly and attain the pebble isolation mass beyond r_{ice} (also found by Brügger et al. 2020; Lambrechts & Johansen 2014). We note that all the cores that start beyond the ice line and reach $a \lesssim 0.43 \text{ au}$ (or $P < 100$ days for a Sun-mass star) are considerably more massive than the ones forming inside it. This is due to the two-order-of-magnitude jump in Stokes number (Fig. 1b) and to the large Stokes number that enhances the pebble accretion rate (Ormel & Klahr 2010; Lambrechts & Johansen 2012). In addition, the pebble isolation mass is higher at longer orbital periods (Lambrechts et al. 2014; Bitsch et al. 2018; Ataiee et al. 2018), which renders the icy cores effectively more massive than the rocky ones (Fig. 1a).

Figure 2 shows the core and envelope mass after formation of all the simulated planets that finish with $P \leq 100$ days.

Fig. 2. Core mass vs envelope mass after formation, for all the cases with final orbital period within 100 days.

Again, the colour-bar indicates the water mass fraction of the core. The effect of the ice line in the core growth is noticeable: icy cores (blue) tend to be more massive than rocky ones (red). This is more clear when we plot a histogram of the core masses (Fig. 3a). We note that the distribution of rocky core masses ($f_{\text{ice}} = 0$, red bars) is quite narrow, with a peak at $\sim 3 M_{\oplus}$ and maximum core mass of $\sim 5 M_{\oplus}$, in agreement with Paper I. On the contrary, the distribution of icy cores is more spread out, with $1 \lesssim M_{\text{core}} \lesssim 36 M_{\oplus}$. However, the peak occurs clearly for higher core masses ($\sim 10 M_{\oplus}$) compared to the rocky case. The median for those planets occurs at $M_{\text{core}} = 10.9 M_{\oplus}$, and only 25% of the icy cores have $M_{\text{core}} < 8.1 M_{\oplus}$. Hence, the effect of the change in composition with the corresponding transition in the Stokes number at the water ice line is inherited in the overall population. Figure 3b shows the histogram of the core radii for the same cases as the left panel. Interestingly, the two peaks of the *Kepler* size distribution are very well reproduced, with a clear paucity of core radii at $R_{\text{valley}} \approx 1.6\text{--}2 R_{\oplus}$.

However, big cores tend to accrete large amounts of gas, as Fig. 2 shows. What does the size distribution look like when the gaseous envelopes are not neglected and atmospheric mass-loss is accounted for? We show this in Fig. 4. Solid lines indicate size distributions accounting for photoevaporation, while dashed grey lines show what the distributions would look like in the absence of it, to assess the precise effect of this mass-loss mechanism. We note that the appearance of the first peak at $R_P \approx 1.3 R_{\oplus}$ is a consequence of photoevaporation. The left panels correspond to evaporation model A, where only the loss of H-He is considered. The right panels correspond to evaporation model B, where the water is assumed to be homogeneously mixed with the primordial H-He envelope and can also be removed. We note in this figure that the second peak (of originally icy cores) gets considerably wiped out compared to Fig. 3b. Most cores of $10 M_{\oplus}$ have envelopes of equal mass just after formation (Fig. 2), and evaporation cannot remove much gas for such massive cores. Then part of the second peak moves to $R_P \approx 8 R_{\oplus}$. Planets concentrated at this radius correspond to discs of low viscosity ($\alpha = 10^{-4}$). This can be noted by comparing the solid black and blue dotted lines in the upper histograms of Fig. 4. Such low viscosity is necessary to form rocky planets (see Paper I), but creates an overdensity of icy and gas-rich planets at $R_P \approx 8 R_{\oplus}$. This could suggest a viscosity transition at the water ice line, although α is expected to decrease with radial distance (Kretke & Lin 2007). Alternatively, an efficient

envelope-loss or gas-accretion-inhibitor process might operate, which renders the planets as nude cores, as Fig. 3b suggests. Despite the reduction in the number of planets at the second peak in Fig. 4 compared to Fig. 3, it is interesting to note that the paucity of planets at $R_P \sim 1.6\text{--}2 R_{\oplus}$, compared to $R_P \sim 1\text{--}1.6 R_{\oplus}$, remains for both evaporation models and that for model B a valley and small second peak appear at the position reported by Fulton et al. (2017) (Fig. 4, lower right panel).

Next, we analyse the resulting planet mass. We plot the three cases described above (bare cores after formation, and evaporation models A and B) in a mass–radius diagram in Fig. 5. The bare cores are shown colour-coded according to the core water mass fraction. Planets run under model A are depicted as magenta triangles and as green diamonds under model B. The grey dots represent real exoplanets from the NASA Exoplanet Archive. It is interesting to note that the three models overlap with existing exoplanets, and actually bracket the observed population fairly well. We note that evaporation model A can strip out H-He envelopes completely for $M_{\text{core}} \lesssim 8 M_{\oplus}$. Larger cores retain sufficient H-He to be kicked out of the second peak. Evaporation model B retains more planets in the second peak, but leaves all planets having $R_P < 4 R_{\oplus}$ with $M_P < 6 M_{\oplus}$. We discuss the implications of this in Sect. 5.

4. Composition of super-Earths/sub-Neptunes

While the composition of first-peak exoplanets is undoubtedly rocky (Owen & Wu 2017; Jin & Mordasini 2018; Gupta & Schlichting 2019, and this work) planets with radius in the second peak have an intrinsic degenerate composition, with rocky planets with thin H-He atmospheres yielding the same radius as icy-dominated objects (e.g. Dorn et al. 2017; Zeng et al. 2019). Atmospheric mass-loss models tend to suggest that second-peak planets correspond to the first type. What do our combined formation and evolution models show? They do not form rocky planets with masses above $\sim 5 M_{\oplus}$, and Model A strips the envelopes of these planets completely for cases with orbital periods concentrated at 10 days¹. At larger orbital periods some H-He can survive (see Appendix C and Paper I). Since water is not removed in Model A, the few planets falling in the valley and second peak of that case are bare ice-rich cores (Fig. 5).

To understand the composition of second-peak planets coming from Model B, we plot in Fig. 6, the bulk content of water and rocks, and the planet’s H-He mass fraction (f_{HHe}) just after formation (left panel) and after atmospheric mass-loss by evaporation (right panel). The only quantity that remains invariable between the two panels is the mass of rocks. The border colour of the circles distinguishes between cases that end up in the first (yellow) or second peak (black). Let us analyse first the case after evolution. First-peak planets are basically devoid of water and H-He. Regarding the second peak, most planets have water in similar amounts than rocks. These planets are not completely depleted of H-He, and have f_{HHe} spanning 0.2% and 10%². Nevertheless, a few second-peak objects are basically dry and have also a H-He mass fraction below 10%, as found by pure evaporation models.

¹ Due to our choice of disc inner edge, most of the short-period planets that we form finish with $a = 0.1$ au (or $P \approx 10$ days). We discuss this choice in Appendix C.

² This also explains why the second small peak shown in the bottom panels of Fig. 4 occurs at a bit larger radius for model B compared to A.

Fig. 3. Histogram of core masses (*left*) and core radii (*right*) of the full population with $P \leq 100$ days, just after formation. Red: $f_{\text{ice}} < 5\%$, green: $5\% \leq f_{\text{ice}} < 45\%$, blue: $f_{\text{ice}} \geq 45\%$. Black: all together. The vertical lines indicate the position of the peaks as reported by [Fulton et al. \(2017\)](#).

Fig. 4. Radius histogram of the synthetic planets with $P \leq 100$ days, after computing the cooling during 5 Gyr with mass-loss driven by evaporation (solid lines). The dashed grey lines show the overall distribution when evaporation is neglected. *Top panels:* all the populations. *Lower panels:* zoom in on radius between that of Earth and Neptune. *Left panels:* model A (evaporation of H-He envelopes). *Right panels:* model B (evaporation of H-He-H₂O envelopes). Red, blue, and green indicate different initial water core fractions as in Fig. 3, and black lines the overall distributions. The blue dotted line in the upper panels shows water-rich planets born in discs of $\alpha = 10^{-3}$ (the remaining cases correspond to $\alpha = 10^{-4}$).

It is also interesting to know if first and second-peak planets accreted from inside or outside the ice line. The left panel of Fig. 6 shows the same quantities as the right panel, but just after formation, before mass-loss takes place. The border colours still indicate the posterior belonging to the first and second peak. We note that in this case where the semi-major axis is typically $a \approx 0.1$ au (see Appendix C), all second-peak objects were born water-rich (also clear from the lower right panel of

Fig. 4); that is, they migrated from beyond the ice line. Interestingly, even though most first-peak planets were born dry (i.e. within the ice line), a few also started with water that was then lost. This means that bare rocky cores could also originate beyond the ice line and lose all their volatile content (H, He, and water) due to stellar irradiation. The number of first-peak objects with this origin should decrease with increasing orbital period.

Fig. 5. Mass-radius of all the planets with final orbital period within 100 days. Filled circles with colour (indicating the water mass fraction of the core after formation) correspond to the mass-radius of the cores of the planets (i.e. the envelope is neglected). The radius is calculated following Zeng et al. (2019) for this case. Magenta triangles show the results of evaporation of H-He after formation. Green diamonds show the same, but assuming mass-loss of H, He, and H₂O. Grey small circles are true exoplanets with orbital periods of less than 100 days, planet radius below 12 $R_{\oplus}$, error on radius of less than 20%, and error on mass of less than 75% (taken from the NASA Exoplanet Archive, July 14, 2020). Yellow shaded areas highlight the two-modes of the *Kepler* size distribution, with darker tones towards the peaks. The gap is delimited by the grey horizontal lines for $1.82 \leq R_p \leq 1.96$, following Martinez et al. (2019).

Fig. 6. Bulk water versus rock content after formation (*left*) and after evolution (*right*) for Model B. The colour-bar indicates the planet H-He mass fraction at each corresponding epoch. Yellow-line circles represent cases that end up in the first peak ($1 < R_p \leq 1.7 R_{\oplus}$) and black-line circles cases that finish in the second peak ($1.7 < R_p < 4 R_{\oplus}$).

5. Discussion

We found that for the radius valley to exist, it is not mandatory that all planets are dry, as pure evolution models suggest. From a planet formation perspective, many of the existing super-Earths/sub-Neptunes are expected to form beyond the water ice line, as shown in this work and many others (e.g. Alibert et al. 2013; Bitsch et al. 2019; Schlecker et al. 2020). Our results indicate that second-peak planets can be often half water and half rock with ~ 0.01 – 10% H-He by mass (Fig. 6). Some second-peak exoplanets present water signatures in their spectra (Kreidberg et al. 2020; Benneke et al. 2019). In addition, planets in the first peak could actually have lost all their H-He and water, and remain as bare rocky cores. Thus, planets starting their formation beyond the ice line can end up as purely rocky as well. Our study suggests that interpreting the origin of super-Earths/sub-Neptunes can be more cumbersome than previously thought.

When analysing the final mass-radius in Fig. 5, we note that the results of our models encompass the short-period exoplanet

population. When combining formation and evaporation models, it seems difficult to obtain planets with mass of ~ 10 – $40 M_{\oplus}$ and radius below that of Neptune. Nevertheless, such objects could be bare cores of half water–half rock if some missing mechanism could inhibit gas accretion or remove the gas after the formation. A process proposed to hinder the entire build-up of the envelope at short orbital periods is known as atmospheric recycling (Ormel et al. 2015), although more recent works adopting non-isothermal discs report that the process only abates gas accretion (Kurokawa & Tanigawa 2018; Lambrechts & Lega 2017; Cimerman et al. 2017). More work on the topic is needed to elucidate the importance of this mechanism. Another possibility is the accretion of planetesimals in addition to pebbles (Alibert et al. 2018; Venturini & Helled 2020). In such a hybrid scenario the heat released by planetesimals delays the accretion of gas once pebble accretion stops at isolation mass (Guilera et al. 2020).

Finally, we have neglected the effect of collisions, which can also remove gas, especially once the disc dissipates. We estimate the magnitude of collisions on the envelope loss in Appendix D.

When one giant impact (per planet) takes place after disc dispersal, we find that the mass–radius of the observed exoplanets is much better reproduced (Fig. D.1) and that the *Kepler* size bimodality is fairly well recovered (Fig. D.2a). Too many collisions would promote compositional mixing (Raymond et al. 2018), smearing out the radius valley (Schlecker et al. 2020; Van Eylen et al. 2018).

6. Conclusions

By studying pebble-based planet formation we found that the change in dust properties at the water ice line combined with the increase in the pebble isolation mass with orbital distance yields two distinct populations of planetary cores, one rocky peaked at $\sim 3 M_{\oplus}$ with all masses below $\sim 5 M_{\oplus}$, and another icy, more spread, and peaked at $\sim 10 M_{\oplus}$. Remarkably, when neglecting the presence of the gaseous envelopes, such mass-bimodality accounts naturally for the bimodal size distribution of the *Kepler* exoplanets.

When considering the formed planets with their envelopes, by computing the photoevaporation of the accreted atmospheres, we corroborate that this process can by itself render the correct radius gap. Nevertheless, contrary to pure evaporation studies, we find that the gap typically separates dry from wet planets. Future atmospheric characterisation with JWST and ARIEL will be crucial to learn how water-rich and water-poor second-peak exoplanets are, and will provide precious constraints for planet formation and evolution models.

By considering extreme-case scenarios with and without gaseous envelope, we find that the exoplanet population is fairly well bracketed by these end-members (Fig. 5). This suggests, on the one hand, that reality might be in between, and on the other, that a much more effective gas-accretion-inhibitor and/or gas-loss mechanism might be operating to explain planets with masses in the range $\sim 10\text{--}40 M_{\oplus}$ and falling on the second peak. The combination of different processes such as hybrid pebble-planetesimal accretion, collisions, photoevaporation, and core-powered mass-loss into a single framework might be an important venue to bridge the gap between theory and observations.

Acknowledgements. We thank the anonymous referee for valuable criticism. J. V. and O. M. G. thank the ISSI Team “Ice giants: formation, evolution and link to exoplanets” for fruitful discussions. O. M. G. thanks ISSI Bern for their support and hospitality during a monthly stay. J. H. acknowledges the Swiss National Science Foundation (SNSF) for supporting research through the SNSF grant 200020_19203. This work has been carried out in part within the framework of the NCCR PlanetS supported by the Swiss National Science Foundation. O. M. G. is partially support by PICT 2018-0934 and PICT 2016-0053 from ANPCyT, Argentina. O. M. G. and M. P. R. acknowledge financial support from the Iniciativa Científica Milenio (ICM) via the Núcleo Milenio de Formación Planetaria Grant. M. P. R. acknowledges financial support provided by FONDECYT Grant 3190336.

References

Alessi, M., Inglis, J., & Pudritz, R. E. 2020, *MNRAS*, **497**, 4814
 Alibert, Y., & Venturini, J. 2019, *A&A*, **626**, A21
 Alibert, Y., Carron, F., Fortier, A., et al. 2013, *A&A*, **558**, A109
 Alibert, Y., Venturini, J., Helled, R., et al. 2018, *Nat. Astron.*, **2**, 873
 Andrews, S. M., Wilner, D. J., Hughes, A. M., Qi, C., & Dullemond, C. P. 2010, *ApJ*, **723**, 1241
 Ataiee, S., Baruteau, C., Alibert, Y., & Benz, W. 2018, *A&A*, **615**, A110
 Aumatell, G., & Wurm, G. 2014, *MNRAS*, **437**, 690
 Bell, K. R., & Lin, D. N. C. 1994, *ApJ*, **427**, 987
 Benneke, B., Wong, I., Piaulet, C., et al. 2019, *ApJ*, **887**, L14

Birnstiel, T., Ormel, C. W., & Dullemond, C. P. 2011, *A&A*, **525**, A11
 Bitsch, B., Johansen, A., Lambrechts, M., & Morbidelli, A. 2015a, *A&A*, **575**, A28
 Bitsch, B., Lambrechts, M., & Johansen, A. 2015b, *A&A*, **582**, A112
 Bitsch, B., Morbidelli, A., Johansen, A., et al. 2018, *A&A*, **612**, A30
 Bitsch, B., Raymond, S. N., & Izidoro, A. 2019, *A&A*, **624**, A109
 Brügger, N., Burn, R., Coleman, G. A. L., Alibert, Y., & Benz, W. 2020, *A&A*, **640**, A21
 Choukroun, M., Altwegg, K., Kürt, E., et al. 2020, *Space Sci. Rev.*, **216**, 44
 Cimerman, N. P., Kuiper, R., & Ormel, C. W. 2017, *MNRAS*, **471**, 4662
 Cossou, C., Raymond, S. N., Hersant, F., & Pierens, A. 2014, *A&A*, **569**, A56
 Crida, A., Morbidelli, A., & Masset, F. 2006, *Icarus*, **181**, 587
 Dorn, C., Venturini, J., Khan, A., et al. 2017, *A&A*, **597**, A37
 Drażkowska, J., & Alibert, Y. 2017, *A&A*, **608**, A92
 Drażkowska, J., Alibert, Y., & Moore, B. 2016, *A&A*, **594**, A105
 Flock, M., Turner, N. J., Mulders, G. D., et al. 2019, *A&A*, **630**, A147
 Fulton, B. J., Petigura, E. A., Howard, A. W., et al. 2017, *AJ*, **154**, 109
 García, A. J. L., & González, J.-F. 2020, *MNRAS*, **493**, 1788
 Ginzburg, S., Schlichting, H. E., & Sari, R. 2018, *MNRAS*, **476**, 759
 Guilera, O. M., Sándor, Z., Ronco, M. P., Venturini, J., & Miller Bertolami, M. 2020, *A&A*, **642**, A140
 Gundlach, B., & Blum, J. 2015, *ApJ*, **798**, 34
 Gundlach, B., Kilias, S., Beitz, E., & Blum, J. 2011, *Icarus*, **214**, 717
 Gupta, A., & Schlichting, H. E. 2019, *MNRAS*, **487**, 24
 Haack, D., Otto, K. A., Gundlach, B., et al. 2020, *A&A*, in press, <https://doi.org/10.1051/0004-6361/202037763>
 Haldemann, J., Alibert, Y., Mordasini, C., & Benz, W. 2020, *A&A*, in press, <https://doi.org/10.1051/0004-6361/202038367>
 Inamdar, N. K., & Schlichting, H. E. 2015, *MNRAS*, **448**, 1751
 Izidoro, A., Ogihara, M., Raymond, S. N., et al. 2017, *MNRAS*, **470**, 1750
 Jiménez, M. A., & Masset, F. S. 2017, *MNRAS*, **471**, 4917
 Jin, S., & Mordasini, C. 2018, *ApJ*, **853**, 163
 Kreidberg, L., Mollière, P., Crossfield, I. J. M., et al. 2020, *AAS J.*, submitted [arXiv:2006.07444]
 Kretke, K. A., & Lin, D. N. C. 2007, *ApJ*, **664**, L55
 Krijt, S., Ormel, C. W., Dominik, C., & Tielens, A. G. G. M. 2016, *A&A*, **586**, A20
 Kurokawa, H., & Tanigawa, T. 2018, *MNRAS*, **479**, 635
 Lambrechts, M., & Johansen, A. 2012, *A&A*, **544**, A32
 Lambrechts, M., & Johansen, A. 2014, *A&A*, **572**, A107
 Lambrechts, M., & Lega, E. 2017, *A&A*, **606**, A146
 Lambrechts, M., Johansen, A., & Morbidelli, A. 2014, *A&A*, **572**, A35
 Liu, B., & Ormel, C. W. 2017, *A&A*, **606**, A66
 Lodders, K., Palme, H., & Gail, H. P. 2009, *Landolt Börnstein*, **4B**, 712
 Martinez, C. F., Cunha, K., Ghezzi, L., & Smith, V. V. 2019, *ApJ*, **875**, 29
 Masset, F. S. 2017, *MNRAS*, **472**, A204
 Morbidelli, A., Lambrechts, M., Jacobson, S., & Bitsch, B. 2015, *Icarus*, **258**, 418
 Mordasini, C. 2020, *A&A*, **638**, A52
 Mulders, G. D., Pascucci, I., Apai, D., & Ciesla, F. J. 2018, *AJ*, **156**, 24
 Musiolik, G., & Wurm, G. 2019, *ApJ*, **873**, 58
 Ogihara, M., & Hori, Y. 2020, *ApJ*, **892**, 124
 Ogihara, M., Kokubo, E., Suzuki, T. K., & Morbidelli, A. 2018, *A&A*, **615**, A63
 Oka, A., Nakamoto, T., & Ida, S. 2011, *ApJ*, **738**, 141
 Okuzumi, S., & Tazaki, R. 2019, *ApJ*, **878**, 132
 Ormel, C. W., & Klahr, H. H. 2010, *A&A*, **520**, A43
 Ormel, C. W., Shi, J.-M., & Kuiper, R. 2015, *MNRAS*, **447**, 3512
 Owen, J. E., & Wu, Y. 2017, *ApJ*, **847**, 29
 Petigura, E. A. 2020, *AJ*, **160**, 89
 Petigura, E. A., Marcy, G. W., Winn, J. N., et al. 2018, *AJ*, **155**, 89
 Raymond, S. N., Boulet, T., Izidoro, A., Esteves, L., & Bitsch, B. 2018, *MNRAS*, **479**, L81
 Ronco, M. P., Guilera, O. M., & de Elía, G. C. 2017, *MNRAS*, **471**, 2753
 Savvidou, S., Bitsch, B., & Lambrechts, M. 2020, *A&A*, **640**, A63
 Schlecker, M., Mordasini, C., Emsenhuber, A., et al. 2020, *A&A*, in press, <https://doi.org/10.1051/0004-6361/202038554>
 Schoonenberg, D., Liu, B., Ormel, C. W., & Dorn, C. 2019, *A&A*, **627**, A149
 Sotin, C., Grasset, O., & Mocquet, A. 2007, *Icarus*, **191**, 337
 Tanaka, H., Takeuchi, T., & Ward, W. R. 2002, *ApJ*, **565**, 1257
 Tatsuuma, M., Kataoka, A., & Tanaka, H. 2019, *ApJ*, **874**, 159
 Van Eylen, V., Agentoft, C., Lundkvist, M. S., et al. 2018, *MNRAS*, **479**, 4786
 Venturini, J., & Helled, R. 2020, *A&A*, **634**, A31
 Venturini, J., Guilera, O. M., Ronco, M. P., & Mordasini, C. 2020, *A&A*, submitted [arXiv:2008.05497]
 Zeng, L., Jacobsen, S. B., Sasselov, D. D., et al. 2019, *Proc. Nat. Acad. Sci.*, **116**, 9723

Appendix A: Dependence on the fragmentation velocity of grains

The core growth by pebble accretion depends sensitively on the Stokes number (Lambrechts & Johansen 2012, 2014), as we mentioned in Sect. 3. The Stokes number is proportional to the pebbles' mean size, which is affected by the fragmentation velocity of the particles. In this work we adopted the conservative approach of considering a fragmentation threshold velocity of 1 m s^{-1} for silicate grains and 10 m s^{-1} for icy grains (Paper I). Support for these numbers stems from the experimental work of Gundlach et al. (2011), Aumatell & Wurm (2014) and Gundlach & Blum (2015) and was adopted in the study of Drążkowska & Alibert (2017).

However, recent lab experiments seem to challenge this view. Musiolik & Wurm (2019) reported that icy grains of 1.1 mm have similar sticking properties to silicate grains for $T \lesssim 180 \text{ K}$ and $P = 1.5 \text{ mbar}$. In the view of these new experiments, we performed a few test cases to evaluate how our results depend on the adoption of more similar fragmentation threshold velocities between silicate and icy grains. We repeated the simulations of Fig. 1 (where $v_{\text{th}} = 1 \text{ m s}^{-1}$ for $r < r_{\text{ice}}$ and $v_{\text{th}} = 10 \text{ m s}^{-1}$ for $r \geq r_{\text{ice}}$), according to the following cases:

- case 1: $v_{\text{th}} = 1 \text{ m s}^{-1}$ along all the disc;
- case 2: $v_{\text{th}} = 1 \text{ m s}^{-1}$ for $r < r_{\text{ice}}$ and $v_{\text{th}} = 2 \text{ m s}^{-1}$ for $r \geq r_{\text{ice}}$;
- case 3: $v_{\text{th}} = 1 \text{ m s}^{-1}$ for $r < r_{\text{ice}}$ and $v_{\text{th}} = 5 \text{ m s}^{-1}$ for $r \geq r_{\text{ice}}$.

The results of the core growth for these three cases are shown in Fig. A.1. Interestingly, case 1 leads to practically no growth. The reasons for this are twofold. First, the smaller particle sizes translate into a reduced drift velocity, which yields a strong reduction of the pebble flux along the entire disc. This precludes the core growth at all locations. Second, the decrease in the Stokes number resulting from the decrease in the pebble sizes, reduces the pebble accretion rate for the outer planets. This suggests that silicate and icy grains might not have exactly equivalent properties within a protoplanetary disc. For case 2, the growth of the inner planets is very similar to Fig. 1, but the growth of the outer ones is modified. There is still one icy planet that reaches $M_{\text{core}} = 10 M_{\oplus}$. For case 3 the growth of the planets is very similar to the nominal case shown in Fig. 1. This suggests that a reduction in the fragmentation velocity of icy grains by a factor of two would not affect our conclusions.

Regarding the water mass fraction of the resulting cores, we note that the decrease in the pebble accretion rate promoted by a lower v_{th} makes the embryo starting its growth just outside the ice line finish with a more mixed composition in cases 1 and

2. For case 3 the dichotomy in the resulting core ice fraction shown in Fig. 1 is recovered. Some previous works (Bitsch et al. 2019; Schoonenberg et al. 2019) find that planets starting their formation beyond the ice line continue to accrete dry pebbles within it. For this to occur the timescale of core growth has to be longer than the timescale of migration, making it possible for the protoplanet to migrate substantially (crossing the ice line) before reaching the pebble isolation mass. This is the case for the embryo starting its formation just outside the ice line in our case 2. We note that even reducing v_{th} of icy pebbles by a factor of 5 (case 2), the dichotomy between pure rocky and half-rock–half-ice cores is maintained unless the embryo starts to form extremely close to the ice line.

A curious aspect about the formation tracks of Figs. A.1 and 1 is that for case 3 and also for the nominal case (Fig. 1), the planet starting its formation just outside the ice line grows so fast that migration does not have time to modify the trajectory, leading to an in situ formation until the attainment of M_{iso} . The planets starting their formation farther out ($a_{\text{ini}} \geq 4 \text{ au}$) grow slower and the torques have time to act, moving the planets typically farther away at the beginning of the growth due to the thermal torque. The reason why the case with $a_{\text{ini}} = r_{\text{ice},0} + 0.1 \text{ au}$ experiences this extremely fast growth (see green and white dots in Fig. 1a) is the following. The change in v_{th} at the ice line leads to a change in pebble size, and thus to a change in the drift velocities. This provokes a traffic jam in the vicinity of the ice line (at approximately $r_{\text{ice}} \pm 0.5 \text{ au}$), which, at early times, increases the surface density of pebbles at that location (see bottom right panel of Fig. 1, Paper I), leading to a rise in the pebble accretion rate.

To close this section, it is important to mention that the results of Musiolik & Wurm (2019) have been regarded as controversial by some authors. Garcia & Gonzalez (2020) point out that the results of Musiolik & Wurm (2019) disagree with the tensile strength computed numerically by Tatsuuma et al. (2019). Okuzumi & Tazaki (2019) mention that Musiolik's recent experiments are inconsistent with earlier ones performed by Gundlach & Blum (2015), which showed efficient sticking of H_2O grains for temperatures down to 100 K . In addition, missing key aspects such as porosity (Garcia & Gonzalez 2020; Krijt et al. 2016) and the lack of experiments involving mixtures of silicates and ices (Choukroun et al. 2020) might influence the fragmentation velocities. The very recent work of Haack et al. (2020) finds that tensile strengths of mixtures of silicate and ice at 150 K are lower than previously reported. More experimental work is needed to pin down realistic values of dust properties under the environmental conditions of a protoplanetary disc.

Fig. A.1. Repetition of Fig. 1, but considering different v_{th} . Case 1, 2, and 3 of Appendix A are shown in the left, centre, and right panels, respectively. For clarity, the y-axis of case 1 has a very different scale than the other two cases. Planets in case 1 basically do not grow.

Appendix B: Disc parameters and initial conditions

Table B.1. Observed discs from [Andrews et al. \(2010\)](#) with their parameters and corresponding lifetimes and initial ice line positions.

Disk number	γ	$M_{d,0} [M_{\odot}]$	r_c [au]	$\alpha = 10^{-3}$		$\alpha = 10^{-4}$	
				τ [Myr]	$r_{ice,0}$ [au]	τ [Myr]	$r_{ice,0}$ [au]
1	0.9	0.029	46.0	1.73	2.47	3.54	1.74
2	0.9	0.117	127.0	7.24	2.79	8.84	1.87
3	0.7	0.143	198.0	9.08	1.89	11.07	1.53
4	0.4	0.028	126.0	2.03	1.38	3.16	1.37
5	0.9	0.136	80.0	7.62	3.79	9.63	2.30
6	1.0	0.077	153.0	4.93	2.47	6.47	1.75
7	0.8	0.029	33.0	1.61	2.68	3.65	1.81
8	0.8	0.004	20.0	0.39	1.66	1.75	1.47
9	1.0	0.012	26.0	0.80	2.30	3.25	1.69
10	1.1	0.007	26.0	0.59	2.01	3.00	1.59
11	1.1	0.007	38.0	0.56	1.84	2.87	1.53
12	0.8	0.011	14.0	0.78	2.65	3.51	1.81

We adopt the initial gas surface density profile inferred from the observations of [Andrews et al. \(2010\)](#),

$$\Sigma_g = \Sigma_g^0 \left(\frac{r}{r_c} \right)^{-\gamma} e^{-(r/r_c)^{2-\gamma}}, \quad (\text{B.1})$$

where Σ_g^0 is a normalisation parameter determined by the disc initial mass ($M_{d,0}$), γ is the exponent that represents the surface density gradient, and r_c is the characteristic radius of the disc. All the disc parameters are taken from [Andrews et al. \(2010\)](#) and are shown in Table B.1, with their corresponding lifetime (τ) and initial ice line position ($r_{ice,0}$). For the viscosity we consider $\alpha = 10^{-3}$ and $\alpha = 10^{-4}$. Only the low-alpha case produces pure rocky planets, as found in Paper I.

We run simulations for all the discs with lifetimes between 1 and 12 Myr (19 discs), for which we consider the initial dust-to-gas ratios (Z_0) shown in Table B.2. Such a wide range in dust-to-gas ratios or metallicities spans the metallicities of planet-hosting stars ([Petigura et al. 2018](#)). We launch seven embryos per disc (one embryo at a time), with initial semi-major axes of $a_{ini} = 0.5, 1, r_{ice,0} - 0.1, r_{ice,0} + 0.1, 4, 8$, and 16 au.

All the embryos are inserted at $t = 0$. We checked that changing this initial time to 0.1 Myr (as is customarily done in pebble accretion simulations; e.g. [Lambrechts & Johansen 2014](#); [Bitsch et al. 2015a, 2019](#); [Ogihara et al. 2018](#)) barely modifies the results.

We note that the initial ice line position of all the discs with $\alpha = 10^{-4}$ lies between 1.37 and 2.3 au (see Table B.1). This is comparable to the locations reported by other works, such as [Oka et al. \(2011\)](#) and [Bitsch et al. \(2015b\)](#) for high accretion rates onto the central star, of $\sim 10^{-7} - 10^{-8} M_{\odot} \text{ yr}^{-1}$. It is important to note that our disc model uses the classical opacities of [Bell & Lin \(1994\)](#), suited for micrometre-size grains. Dust coagulation, especially for low disc turbulence, is expected to reduce

Table B.2. Adopted initial dust-to-gas ratio or disc metallicity (Z_0) and the corresponding [Fe/H].

Z_0	[Fe/H]
0.0068	-0.350
0.0099	-0.185
0.0144	-0.025
0.0210	0.138
0.0305	0.300

Notes. [Fe/H] = $\log_{10}(Z_0/Z_{\odot})$, where $Z_{\odot} = 0.0153$ is taken from [Lodders et al. \(2009\)](#).

the grain opacities ([Savvidou et al. 2020](#)), yielding an ice line location closer to the central star. In Paper I and in this work we coupled in a self-consistent manner dust growth and evolution with pebble accretion. Future work should also address the difficult problem of coupling consistently grain growth with the disc's opacities.

Appendix C: Dependence on the disc inner edge

The inner border of the disc determines the minimum semi-major axis that planets can attain by inwards migration. When planets migrate in resonant chains, the innermost planet tends to stop its migration at or near the edge of the protoplanetary disc ([Cossou et al. 2014](#)), although outwards migration can also occur due to the expansion of the inner cavity during disc dispersal ([Liu & Ormel 2017](#)). Since we do not include N -body interactions or the effect of the magnetic cavity in our calculations, most planets tend to end up near the disc inner edge, assumed as $r_{in} = 0.1$ au in our nominal setup (all figures of main text). The final planet's position affects mainly the photoevaporation rate and hence the final mass and thickness of a planet's atmosphere. Our choice of nominal disc inner edge at $r_{in} = 0.1$ au is based on constraints from hydrodynamical simulations ([Flock et al. 2019](#)), and on the observed typical position of the innermost exoplanet in a system ([Mulders et al. 2018](#)). Still, the mean orbital period of second-peak exoplanets is ~ 38 days ([Martinez et al. 2019](#)), which corresponds to $a \approx 0.22$ au for solar-type star. Hence, it is relevant to test how the composition of second-peak planets would change for such orbital periods. We therefore repeat the simulations with $r_{in} = 0.2$ au, together with the histograms of Fig. 4 for Model B. Both histograms (nominal setup and $r_{in} = 0.2$ au) are shown in Fig. C.1. We note that for $r_{in} = 0.2$ au some planets that formed inside the water ice line (and are therefore devoid of water, the red bars) end up in the second peak, meaning that they retain some H-He atmosphere. This is a natural consequence of photoevaporation removing less gas at larger orbital distances. The longer the orbital period, the larger the number of rocky-icy objects that should contribute to the second peak. Future work with population synthesis will be able to quantify this precisely and give quantitative predictions.

Fig. C.1. Same as bottom right panel of Fig. 4, but comparing the nominal case ($r_{\text{in}} = 0.1$ au, *left panel*) with the case where $r_{\text{in}} = 0.2$ au (*right*).

Appendix D: Envelope mass-loss by giant impacts

Giant collisions, which may be important at the time of disc dispersal (Ogihara & Hori 2020), are, in addition to photoevaporation, a possible mechanism that could help in removing a planet’s atmosphere. Although we did not consider collisions in our simulations since we formed only one planet per simulation, we can estimate in a simple way the fraction of envelope mass-loss a planet could undergo if we allowed it to collide with another less-massive planet, formed isolated in the same disc. These hypothetical collisions could happen within the first million years of evolution after gas dissipation and before substantial photoevaporation takes place (e.g. Izidoro et al. 2017). The goal is to compute the envelope mass-loss of a planet due to a possible collision plus the subsequent atmospheric loss due to photoevaporation, to explore how the mass-radius of exoplanets (Fig. 5) and the radius distribution (Fig. 4) could be affected.

Following the same procedure as in Ronco et al. (2017, see their Sect. 2.2.3), we compute the core mass of the collision remnant as the sum of the core masses of the target and the impactor. The final gaseous envelope is computed following Inamdar & Schlichting (2015), who calculated the global atmospheric mass-loss fraction for planets with masses in the range of the super-Earths and mini-Neptunes. Although this study does not provide mass-loss fractions for collisions with gas giant planets with extended atmospheres, we use the same results due to lack of works on the subject.

For simplicity and following the results of Ogihara & Hori (2020), who report only one or two giant impacts when accounting for N -body interactions, we allow only one collision per planet, but compute all the possible results of that collision considering that all the less-massive planets in the same disc (with final periods < 100 days), can be the impactor. We compute mean values for the core mass, envelope mass, and core ice fraction for each ‘family of impacts’. The percentage of the envelope mass-loss due to impacts ranges between 11% and 100% with a mean of 55%. If for each family of impacts we consider the most destructive one (the one that generates the maximum envelope mass-loss), the percentage of the envelope mass-loss ranges between 16% and 100%, with a mean value of 72% for this latter

case. Overall, collisions could reduce the mass of the envelope by a factor of ~ 2 .

After computing collisions, we compute the mass-loss due to photoevaporation (only with Model B) for the mean and maximum values of each family of impacts. In Fig. D.1, which is similar to our previous Fig. 5, we compare the planet population affected only by photoevaporation (as in Fig. 5, green diamonds) with the planet population that also suffered a collision after gas dissipation (coloured circles). The colour-bar of the circles represents the final water mass fraction with respect to the total heavy-element content³ after collisions and photoevaporation are calculated. The black and yellow borders of the circles represent those planets with mean and maximum envelope mass-loss after collisions, respectively, followed by envelope mass-loss due to photoevaporation. The grey border circles denote the naked cores of the planets that lost their entire atmosphere either after the collision, or after the collision followed by photoevaporation. For the cores resulting nude after the collision, the radii are computed following Zeng et al. (2019), as in the main text.

The water mass fraction evidences some mixing of material due to collisions, which can be seen in all the figures of this Appendix. Approximately 33% (20%) of the resulting planets of the mean (maximum) collisional model have a final water mass fraction of $0.05 \leq f_{\text{ice},f} < 0.45$, compared to $\sim 4\%$ when collisions were not considered (Sect. 3). Nevertheless, most of the planets with this intermediate $f_{\text{ice},f}$ are still water-rich since they typically have $f_{\text{ice},f} > 0.3$. This occurs because the original half-rock–half-water cores were more massive than the pure dry ones, and hence contribute with a non-negligible amount of water when colliding to pure rocky planets. Moreover, 30% of the resulting planets preserve a pure dry composition (for both collisional models); and 37% (50%) an $f_{\text{ice},f} \approx 0.5$ for the mean (maximum) collisional models. This happens because many collisions occur among cores that originally have the same composition.

There are some remarkable aspects to highlight from Fig. D.1. First, the synthetic planets fill the delimiting M-R trends of the simulated planets of Fig. 5, accounting much better for the mean density diversity of real exoplanets. Second, bare rocky planets can now be as massive as $8 M_{\oplus}$ (compared to

³ It is important to note that due to the simplification of the chemistry in our disc model, ‘water’ and ‘ice’ refer throughout this work to all species with condensation temperatures below 170 K.

$5 M_{\oplus}$ without collisions), which fills better the exoplanets clustering around the Earth-like composition trend, which seems to extend until $M_P \sim 10 M_{\oplus}$. Finally, very energetic impacts are able to produce bare icy cores (grey-bordered blue circles in Fig. D.1), which would explain the existence of a few exoplanets with $M_P \sim 70 M_{\oplus}$ and $R_P \sim 4 R_{\oplus}$. According to our formation-evolution model, these planets should be half rock–half water by mass. In addition, the ability of collisions to produce bare cores would move objects from $R_P \sim 8 R_{\oplus}$ to lower radii. This is better appreciated in Fig. D.2, where we repeat the histograms of

the size distributions from Fig. 4 for the two collisional models. From these histograms it is clear that the model considering the maximum amount of H-He removed by collisions gives the best match with observations: the valley in planet radii occurs at $R_P \approx 1.8\text{--}2.1 R_{\oplus}$ and the peaks at 1.4 and $2.8 R_{\oplus}$, giving a better agreement with the latest estimates (Martinez et al. 2019; Van Eylen et al. 2018) than the pure bare cores of Fig. 3, whose peaks match those of Fulton et al. (2017) fairly well. Overall, the inclusion of a few giant impacts seems to be a crucial process to better reproduce the size and mass of short-period exoplanets.

Fig. D.1. Same as Fig. 5, but showing only model B from that figure (green diamonds), plus the results of hypothetical giant impacts followed by photoevaporation (coloured circles). The colour-bar represents the final water mass fraction. The grey dots represent the observed exoplanet population as in Fig. 5. The yellow-bordered circles represent the planets that underwent the maximum envelope mass-loss due to a collision, and the black-bordered circles represent the mean values of envelope mass-loss for each family of collisions. The grey-bordered circles denote the bare cores that lost their envelopes completely either just after the collision or after photoevaporation.

Fig. D.2. Size distribution of the synthetic planets after undergoing one giant impact and photoevaporation. *Left panel:* maximum mass removed by the collision. *Right panel:* mean mass removed by the collision. Red: $f_{ice,f} < 5\%$, green: $5\% \leq f_{ice,f} < 45\%$, blue: $f_{ice,f} \geq 45\%$, where $f_{ice,f}$ is the final mass fraction of water relative to the total amount of heavy elements. Black lines: overall size distribution. Grey dotted lines: distribution without collisions (as in Fig. 4). Vertical dashed lines: peaks of the *Kepler* size distribution inferred by Fulton et al. (2017) (light violet) and Martinez et al. (2019) (dark violet).