A numerical inversion of m sin i exoplanet distribution: the sub-Saturn desert is more depleted than observed and hint of a Uranus mass gap

Jean-Loup Bertaux, Anastasia Ivanova

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **512,** 5552–5571 (2022) https://doi.org/10.1093/mnras/stac777
Advance Access publication 2022 March 22

A numerical inversion of $m \sin i$ exoplanet distribution: the sub-Saturn desert is more depleted than observed and hint of a Uranus mass gap

Jean-Loup Bertaux [1]★ and Anastasiia Ivanova[1,2]

[1] *LATMOS/IPSL, Université de Versailles Saint-Quentin, 11, Boulevard d'Alembert, F-78280 Guyancourt, France*
[2] *Space Research Institute (IKI), Russian Academy of Science, Moscow 117997, Russia*

Accepted 2022 March 16. Received 2022 March 16; in original form 2021 September 20

ABSTRACT

The detection of an exoplanet orbiting another star with the radial velocity (RV) method allows to determine only a minimum mass of the planet, $m \sin i$, m being the true mass and i the angle of inclination of the planet orbital polar axis with the line of sight. Given an observed discretized distribution of $m \sin i$ apparent masses $f_0(m \sin i)$, we have designed a simple algorithm to find a unique true mass distribution $f(m)$ that would reproduce exactly the observed distribution $f_0(m \sin i)$. The method is based on a particular geometrical representation of exoplanets. It calls for the use of spheres and cylinders, and is somewhat similar (though different) to the Abel inversion, widely used in atmospheric physics. We have applied this algorithm to the latest sample of RV discovered planets containing 909 planets. We confirm the existence of a sub-Saturn desert (at least for periods < 100 d), most depleted in the mass Srange in the range 0.1–0.2 $M_{\rm jup}$ ($\sim$32–64 $M_\oplus$), detected in the raw $m \sin i$ distribution, and amplified in the inverted $f(m)$ true mass distribution by a factor $\sim$1.7. We argue that this result is robust, and would remain even if other biases of the RV surveys would be included. Differences with a recent model of population synthesis are discussed. Focusing on lighter planets, we found a likely statistically significant gap of planets in the observed $m \sin i$ distribution in the narrow range of 13.7–15.2 $M_\oplus$ containing Uranus.

Key words: methods: data analysis – methods: numerical – techniques: radial velocities – planets and satellites: formation – planetary systems.

1 INTRODUCTION

The use of radial velocity (RV) of stars has proven very successful at the indirect detection of planets orbiting other stars, as fostered by Connes (1985) with his proposal to use a high resolution, cross-dispersing spectrometer design, CCD 2D detectors, and high stability. Mayor & Queloz (1995) discovered the first exoplanet around a solar type star with this method. Then the HARPS spectrometer was built with the spectrometer under vacuum and accurate thermal control, as recommended by Connes, yielding an accuracy better than 1 m s^{-1}. The Kepler's third law may be used to compute the radius of the orbit from the observed orbital period, knowing the mass of the star. Unfortunately the mass m of the exoplanet cannot be retrieved: only the product $m \sin i$ is derived from the amplitude of the RV wobble, where i is the inclination of the polar axis of the orbit on the line of sight (LOS) from the observer to the star (e.g. Lovis & Fischer 2010).

However, when a reasonable number of exoplanets are detected, giving an observed distribution of $m \sin i$ it is possible to retrieve the distribution function of planetary masses $f(m)$ that will give the observed distribution $f_0(m \sin i)$. One has to make the assumption that the orientations of orbital polar axis are isotropically distributed in space, and independent of the distribution $f(m)$.

Such an exercise was done for instance in Jorissen, Mayor & Udry (2001) and Lopez & Jenkins (2012). There is also the work of Chandrasekhar & Münch (1950), which addressed a different problem: the inversion of the distribution of $V_{\rm rot} \sin i$, where $V_{\rm rot}$ is the rotational velocity of the star at equator, and i is the inclination of the star rotation axis on the LOS to the star. The $V_{\rm rot} \sin i$ may be determined from the wavelength width of one particular stellar spectral line, because of the variable Doppler shift over the disc of the star. Formally, the two problems are identical. Chandrasekhar & Münch (1950) showed that with some change of variables, the inversion of a $V \sin i$ distribution can be transformed into the inversion of an Abel's integral, for which an analytical formulation is known. However, as noted in their paper, 'while equation (14) represents the formal solution of the problem, it is not of much practical use, since it requires differentiation and it is known that the differentiation of an observed frequency function can lead to results which are misleading unless the observations are of high precision'.

The authors then prefer to determine only some moments of the distribution, testing for analytical forms of the V distribution.

More recently, the numerical inversion of Abel's integral has made substantial progresses in the field of atmospheric physics, leading to the famous 'onion peeling' vertical inversion method (See Appendix B2). Assuming a spherical distribution of the density of a given atmospheric constituent (e.g. Bertaux et al. 2010, ozone measured in the star occultation geometry), and a series of column densities recorded at various altitudes, it is possible to retrieve the

★ E-mail: jean-loup.bertaux@latmos.ipsl.fr

vertical distribution of the density of ozone $= f(z)$. See for instance Quémerais et al. (2006), and variants with Tikhonov regularization.

Stimulated by this analogy between exoplanets and atmospheric physics, we have developed a geometrical representation of exoplanets which may be treated in a fashion similar (but not identical) to the atmospheric vertical inversion, to get the actual distribution of planetary masses m.

The methods developed by Jorissen et al. (2001) and Lopez & Jenkins (2012) are rather mathematically sophisticated. The new method that we present here should not give much different results; however, being based on a simple geometrical representation, it is likely simpler to understand, and easier to implement since it is discretized. It is based on the solution of a triangular linear system of equations, which may be resolved step by step (equation by equation) or with the inversion of a triangular matrix.

In Section 2 we describe the principles of our inversion algorithm, while the exact formulation with spheres and cylinders is described in Appendix A. In Section 3 we apply our algorithm to an ensemble of 909 RV planets with three different mass grids, focusing on various parts of the mass histogram. We discuss in Section 4 some gaps identified in the radius/mass observed distribution, in relation to the composition of the solid core, and the existence (or not) of an extended atmosphere. The inverted RV mass distribution is compared in Section 5 with a recent model of population synthesis. Some conclusions are sketched in Section 6, while Appendix B describes the classical onion-peeling algorithm used in atmospheric science, allowing to appreciate the similarities and differences with our algorithm implying spheres and cylinders. In Appendix C we describe several simulations with a forward and inverse calculation, on single peak, double peak, and power-law distributions of true mass, with random assignment of a sin i value.

2 GEOMETRICAL REPRESENTATION OF EXOPLANETS POPULATION AND ASSOCIATED FORWARD FORMULATIONS

2.1 Geometrical representation of exoplanets population

On Fig. 1 is shown the geometry of observation of one exoplanet, where the angle i is the inclination of the polar axis of the orbit on the LOS from the observer to the star, axis X. The angle i is $0 \leq i \leq 180^\circ$, if we select the direction of axis which gives a positive rotation about the axis. We represent the exoplanet as the extremity of a vector representing the polar axis of the orbit coming out from the centre of the star, with a vector length equal to the true mass of the planet. The exoplanet (red dot) is therefore located along the angular momentum direction of the orbit at a length 'm'. This red dot point is projected parallel to observer X-axis on axis Z, at the length $p = m \sin i$ (grey dot).

Now we re-assemble all detected exoplanets in a common three-axis reference system by translating all stars at a common origin S_t, with a rotation aligning all LOS along the X-axis (Fig. 2). All XZ planes of Fig. 1 containing one exoplanet will come to various angles ϕ (unknown) with the common axis S_tZ on Fig. 2. The orientation about the X-axis of each exoplanet is not known, but we may assume that their orbital axis are isotropically distributed. All exoplanets which have the same value of m will lie on a sphere of radius m. All exoplanets which have the same value of $m \sin i$ will lie on a cylinder with axis X and radius $p = m \sin i$. Therefore, since the exoplanets (red points) are distributed isotropically, at a distance m from the origin centre S_t the projected distribution in the plane YZ will have a symmetry of revolution (it will not depend on angle ϕ).

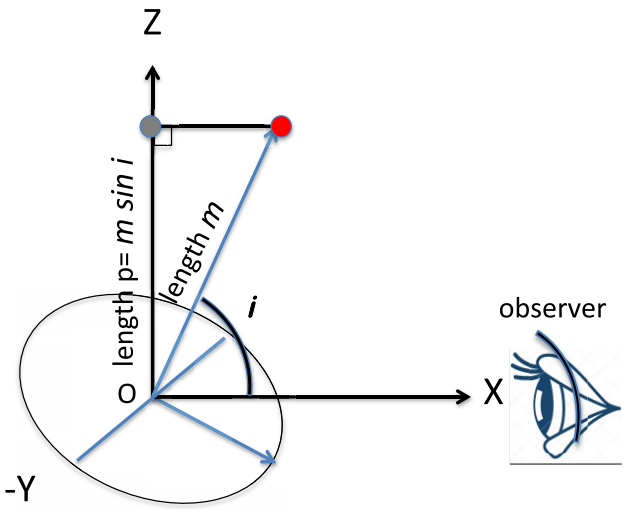

Figure 1. Geometry of observation for one single exoplanet (red point). The star is at the centre O, the LOS being the OX axis. The exoplanet is represented by a point (red), extremity of a vector of length equal to the mass m of the exoplanet, with an inclination angle i to the observer. The apparent mass $p = m \sin i$ is the length of the projection of the exoplanet point on the plane YZ. The XZ plane contains the exoplanet by convention.

2.2 The case when all planets have the same mass

From now on we use the word planet for short of exoplanets. We consider first the case where all planets (in number N) have exactly the same mass m. In our geometric representation, they all lie on the sphere of radius m. We assume that their orbital axis is oriented randomly therefore isotropically. We wish to compute what will be the observed distribution $n(p)$ of apparent masses $p = m \sin i$.

We may state that $p \leq m$; $m \cos i \; di \; = dp$ and the number of planets observed between p and $p + dp$ is defined by $n(p)dp$. The total number of observed planets must be equal to N,

$$\int_0^m n(p)\,\mathrm{d}p = N. \tag{1}$$

In our geometrical representation described above, the surface density S_d of planets on the sphere of radius m is uniform and equal to

$$S_d = \frac{N}{4\,\pi\,m^2}. \tag{2}$$

The area S_{di} of the fraction of sphere which is contained between angles i and $i + di$ (a spherical corona), and projects between p and $p + dp$ on the plane YZ (perpendicular to the observer, passing through the centre of the sphere of radius m), is (Fig. 3)

$$S_{di} = 4\pi p\, m\, \mathrm{d}i. \tag{3}$$

Here we have 2 x $2\pi = 4\pi$ to take into account the two spherical coronas of both hemisphere, one in front of the observer, the other on the other side, projecting on the same plane YZ, as the circular corona contained between p and $p + dp$. The number of planets $n(p)dp$ projecting on this circular corona $(p, p + dp)$ is the product of the area S_{di} and the surface density of planets

$$n(p)\,\mathrm{d}p = S_{\mathrm{di}}\,S_d = \frac{p\,m\,\mathrm{d}i}{m^2} N = N \frac{p}{m\sqrt{m^2 - p^2}}\,\mathrm{d}p \tag{4}$$

$$n(p) = N \frac{p}{m\sqrt{m^2 - p^2}}. \tag{5}$$

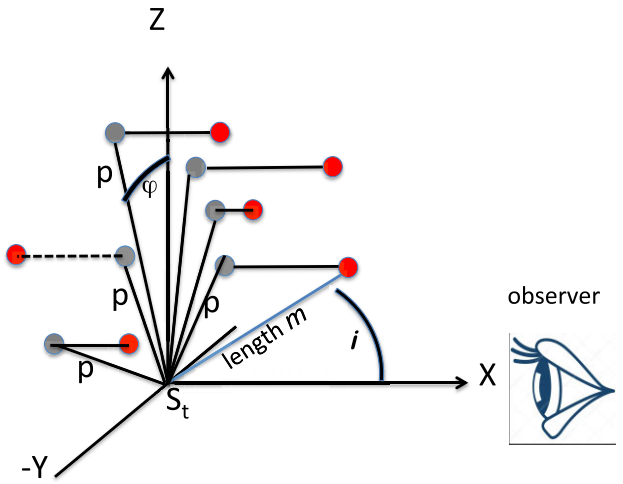

Figure 2. All observed exoplanets and their system of observation are translated and rotated on a common reference system, with all stars being at the centre S_t, the LOS being the OX axis. Their plane XZ of Fig. 1 is now at an angle ϕ (unknown) to axis Z. The exoplanet is represented by a point (red), extremity of a vector of length equal to the mass m of the exoplanet with an inclination angle i to the observer. The apparent mass $p = m \sin i$ is the length of the projection of the exoplanet point on the plane YZ perpendicular to the LOS.

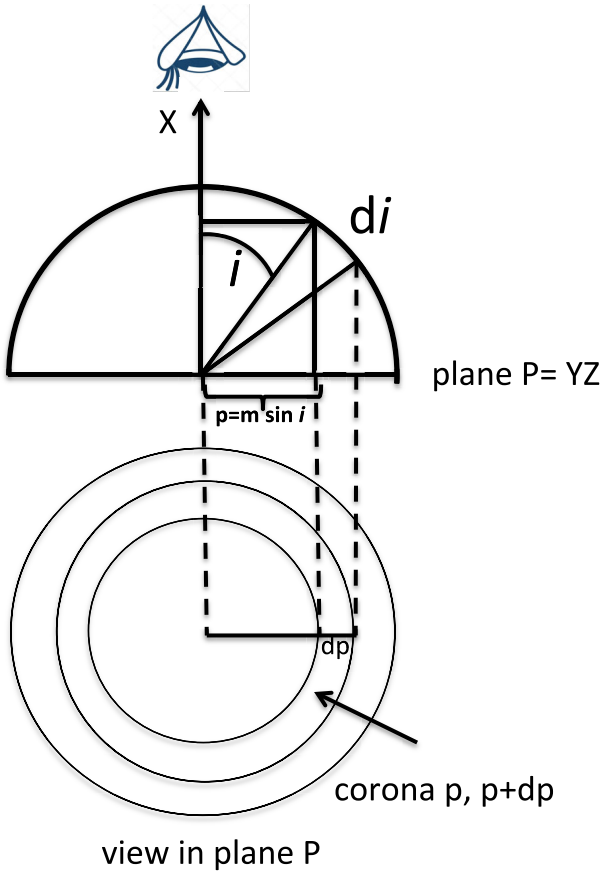

Figure 3. Geometry for the case of iso-mass planets. They lie on a sphere of radius m. Here only one hemisphere is represented. Note the change of orientation for the LOS w.r.t. Figs 1 and 2. All planets with angle inclination between i and $i + di$ are viewed projected in a corona between p and $p + dp$, as shown in the lower part of the figure representing the plane perpendicular to the LOS.

It can be verified that the integral of $n(p)$ from 0 to m is indeed N. This expression becomes infinite when p is approaching m, but the integral is finite. It may be computed that the average value of $p = m \sin i$ for this population $n(p)$ is $m\pi/4$.

Let us now consider a population of planets having all the same mass m = 1 (the unit is not important here). In our geometrical representation of exoplanets (Fig. 3), they are all located on the upper hemisphere,[1] with a constant surface density.

Let us define the cumulative probability function $P_{\rm cum}(x)$ that the value of the variable $\sin i$ be < x,

$$Prob(\sin i < x) = P_{\rm cum}(x). \quad (6)$$

The probability density function $P(x')$ that the variable $x' = \sin i$ be between x and $x + dx$ is therefore such that

$$P(x')dx = P_{\rm cum}(x + dx) - P_{\rm cum}(x). \quad (7)$$

Therefore, the probability density function $P(x')$ is just the derivative of the cumulative probability function (from now on, we use the classical term, the cumulative distribution function, CDF). In general, $P_{\rm cum}(-\infty) = 0$, $P_{\rm cum}(+\infty) = 1$. In the present case, the angle i is $0 \le i \le 90°$, $\sin i$ is defined between 0 and 1, $P_{\rm cum}(0) = 0$ and $P_{\rm cum}(1) = 1$.

Considering Fig. 3, and a sphere of radius 1 (instead of m), an isotropic distribution of polar axis inclination with angle i (from 0 to 90°) with the looking direction X means an equal probability over the whole hemisphere of radius 1. All directions with an angle smaller than a given value i are contained in the cone of axis X and semi-angle i. The probability that the angle of inclination be < i is therefore the ratio of the solid angle Ω sub-tended by the cone to the hemisphere which subtends 2π steradian. We know that $\Omega = 2\pi(1 - \cos i)sr$, and the probability that the angle of inclination be < i is then

$$P(\text{angle} < i) = 1 - \cos i. \quad (8)$$

It is also the probability that $\sin i < x$, for a given value of x. Therefore we get the expression of the cumulative probability of sin i under the form:

$$P_{cum}(\sin i < x) = 1 - \cos i = 1 - \sqrt{1 - \sin^2 i} = 1 - \sqrt{1 - x^2} \quad . \quad (9)$$

Calling $y = P_{cum}(\sin i < x) = 1 - \sqrt{1 - x^2}$, the cumulative probability function of $\sin i$ is represented on Fig. 4 (axes x and y are different from X and Y of Fig. 3). Actually, it is a quarter of a circle of equation $(y - 1)^2 + x^2 = 1$, which centre is at point x = 0, y = 1.

Drawing one single value of $\sin i$ at random (given the probability density function $P(\sin i)$) is done simply by drawing a random number between 0 and 1 (say 0.14 on Fig. 4), and finding the value of $x = \sin i$ for which $P_{\rm cum}(\sin i) = 0.14$. Indeed, we see on Fig. 4 that the probability of a value of $\sin i$ between x and $x + \Delta x$ is $\Delta P_{\rm cum}$, linked to Δx by

$$\frac{\Delta P_{\rm cum}}{\Delta x} = \frac{{\rm d}P_{\rm cum}}{{\rm d}x} = P(x). \quad (10)$$

Fig. 4 shows that, considering a population of planets of the same mass m, half of them (upper part of graph) will have an inclination i larger than 60° and will produce in RV observations a signal (semi-amplitude in m s^{-1}) corresponding to $p = m \sin i$ between $\frac{\sqrt{3}}{2} \sim 0.86$

[1] angles i above 90° have been put back between 0 and 90° by using $180° - i$ instead of i

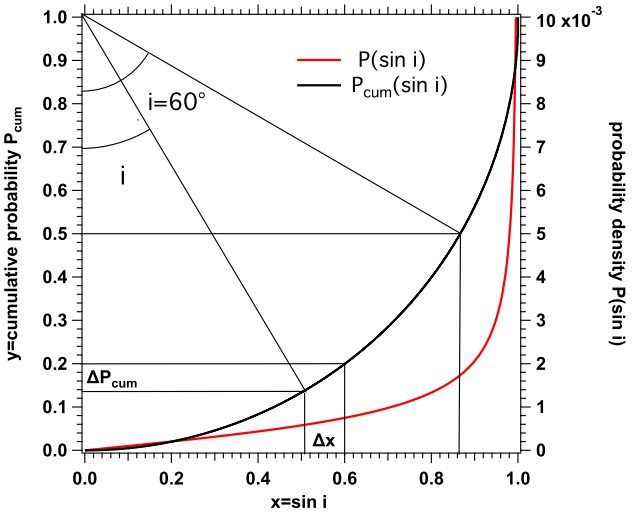

Figure 4. The cumulative probability $y = P_{\rm cum}(\sin i)$ is represented as a function of $x = \sin i$ (black, left scale). It is a quarter of a circle, centred on x = 0, y = 1. The derivative of this function is the probability function $P(\sin i)$ is the solid red curve (right scale), which goes to ∞ for $\sin i = 1$. See text for explanation of Δx and $\Delta P_{\rm cum}$. The angle of inclination i is represented by a line from the centre of the circle. Drawing a line with $i = 60°$ intersects the circle at $P_{\rm cum} = 0.5$, showing that half of the planets have a $p = m \sin i$ between $\frac{\sqrt{3}}{2} \sim 0.86m$ and the true mass m.

and 1 m (therefore, not too far from the true mass m), and the other half a signal corresponding to p smaller than 0.86 m (from 0 to 0.86 m).

2.3 The general case of a distribution of planets

Let us assume that there is a distribution of planets $N(m)dm$, where $N(m)$ is the number of planets with true masses between m and $m + dm$. According to equation (5) above, the observed distribution $f_0(p)$ will be the sum of contributions to p over all masses m, from the minimum value p up to a maximum mass *max*.

$$f_0(p) = \int_p^{\max} n(p)\ {\rm d}m = \int_p^{\max} N(m)\,{\rm d}m \frac{p}{m\sqrt{m^2 - p^2}} \tag{11}$$

$$f_0(p) = p \int_p^{\max} N(m) \frac{{\rm d}m}{m\sqrt{m^2 - p^2}}. \tag{12}$$

This is exactly the same equation as equation (9) in Chandrasekhar & Münch (1950) arrived at to describe the problem of the distribution of $V_{\rm rot}\sin i$, in which $V_{\rm rot}$ is the rotation velocity of the star and angle i is again the inclination angle to the observer.

However, the mathematical demonstration of Chandrasekhar & Münch (1950) applied to the problem of $V_{\rm rot}\sin i$ is not so simple to follow, while our demonstration is based on simple geometrical considerations and we think that it is easier to follow.

As said in Section 1, Chandrasekhar & Münch (1950) showed that with some change of variables, equation (7) may be transformed into the inversion of an Abel's integral, for which an analytical formulation is known. Because of the sensitivity to the derivative, the analytical solution is unstable, as said in the Introduction. Instead, in modern atmospheric physics, the famous 'onion-peeling' numerical method is preferred, most of time with some regularization scheme (e.g. Quémerais et al. 2006, Tikhonov regularization).

Therefore, we are presenting a numerical scheme to resolve the inversion of $m \sin i$ distribution, which is inspired by (but not identical to) the atmospheric onion-peeling method. It is based on our new geometrical 3D representation of the population of planets as a 'gas', in which there is a spherical symmetry of the distribution of planets (isotropic distribution of orbital polar axis). It calls for the computation of volumes which are the intersections of spherical shells and cylindrical shells in the 3D space of exoplanet masses. Instead of considering a continuous distribution of true masses m, the problem is discretized by assuming that the 'number density' of planets $\rho(m_n)$ is constant between two consecutive spherical shells of radius m and $m + \Delta m$ (but the mass may vary within the limits of the mass bin).

Let us assume that we have a distribution $N(m)$, $N(m)$ being the number of planets in a mass-bin Δm. In our geometric representation, the density $\rho(m_n)$ of planets at distance m_n from the centre is linked to the number $N(m_n)$ by

$$\rho(m_n) = \frac{N(m_n)}{4\pi m_n^2 \Delta m} \tag{13}$$

and its integral (fully accurate) form

$$\rho(m_n) = \frac{N(m_n)}{\frac{4\pi}{3}\left(m_n^3 - m_{n+1}^3\right)}. \tag{13bis}$$

In the classical onion-peeling method applied to planetary atmospheres, a LOS crosses a number of spherical shells within which the density is assumed to be constant (Appendix B).The path-length of the LOS through each spherical shell is computed by geometry and multiplied by the local density to get the integrated line (or column) density, which is the observed quantity in the onion-peeling atmospheric scheme. In our problem, it is a whole cylinder of radius p which may be considered as a 'cylinder of sight', the equivalent of a LOS is the onion-peeling scheme. However, since we have discretized the problem with constant density of planets between two spheres of radius m_n and m_{n+1}, the equivalent of a LOS for our problem is the volume between two cylinders of radius $p = m_n$ and $p' = m_{n+1}$, which contains all observed planets with $m \sin i$ between m_n and m_{n+1}. Their total is constituted from the contributions of the various spherical shells with a radius $> m_n$, which are the product of the volumes common to the cylinder and the various spheres, by the density $\rho(m_n)$ in each spherical shell of constant density. Their formulation is established in Appendix A. Below, we give first a simple example of the inversion scheme with three spheres and cylinders, and show the more general formulation.

2.4 The scheme of inversion with spheres and cylinders

2.4.1 A simple example

Let $p = m \sin i$, the 'projected' mass (or apparent mass) of a planet with true mass m, and determined by the RV method to have an apparent mass p. We illustrate our method of inversion (or retrieval) with a simple case in which we consider only three values of masses from 3 to 1 (in arbitrary units), with densities ρ_0 (between R_0 and R_1), ρ_1 (between R_1 and R_2), ρ_2 (between R_2 and 0) in our geometrical representation. The radius of spheres are $R_0 = 3$, $R_1 = 2$, $R_2 = 1$, $R_3 = 0$ (Fig. 5). The largest radius is the first (as is usual in onion-peeling method where computation starts from outside and progresses inwards), decreasing down to zero for the last value $R_3 = 0$. There is one more value of the radius than the number of mass bins. The series of cylinder radii are taken as identical to sphere radii: $C_j = R_j$.

Fig. 5 is a cross-section of the volumes delimited by cylinders and spheres, obtained by revolution about axis X of Figs 1 and 5. The

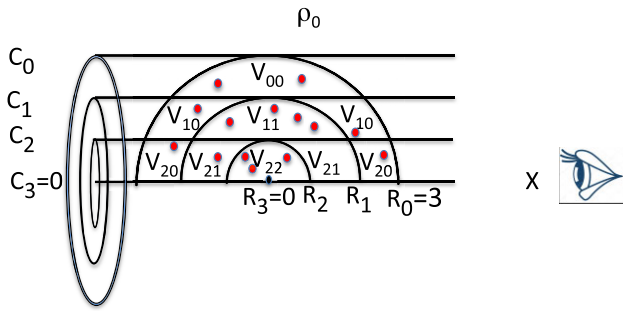

Figure 5. Cross-section of spheres with radii $R_0 = 3$, $R_1 = 2$, $R_2 = 1$, $R_3 = 0$, and cylinders of sight of equal radii $C_0 = R_0$, $C_1 = R_1$, $C_2 = R_2$, $C_3 = 0$. The various volumes contained in both a cylindrical shell and a spherical shell (indicated by V_{ik}) are obtained by a revolution about X-axis.

volumes V_{ik} are identified on Fig. 5 by the first index i, designating the cylindrical shell between C_i and C_{i+1}, and by the second index k, designating the spherical shell contained between R_k and R_{k+1}.

We assume that there is no planet of mass larger than the largest mass, R_0. Considering the largest mass bin from $R_1 < R_0$ to R_0, the number of observed planets is by definition $N(p_0 = m_0 \sin i)$, and these planets belong both to the spherical shell $(R_0; R_1)$, and to the cylindrical shell $(C_0; C_1)$. Let V_{00} be the volume common to both shells (spherical and cylindrical), we may write

$$N(p_0) = \rho_0 V_{00} \tag{14}$$

and similarly, with the definition of the various volumes V_{ik} indicated on Fig. 5

$$\begin{aligned} N(p_0) &= \rho_0 V_{00} \\ N(p_1) &= \rho_0 V_{10} + \rho_1 V_{11} \\ N(p_2) &= \rho_0 V_{20} + \rho_1 V_{21} + \rho_2 V_{22}. \end{aligned} \tag{15}$$

Therefore, we have a linear system of three equations with three unknowns, ρ_0, ρ_1, ρ_2, the number densities of planets of masses $2 < m_0 \leq 3$, $1 < m_1 \leq 2$, $0 < m_2 \leq 1$. The true number $N_{\rm true}$ of planets in each mass interval interval is the product of the number densities ρ_i of planets per unit volume (in the planets space) by the volume of the spherical shells containing all planets with a true mass in this interval (equation 13 bis)

$$\begin{aligned} N_{true}(m_0 = 3) &= \frac{4\pi}{3}\left(R_0^3 - R_1^3\right)\rho_0 \\ N_{true}(m_1 = 2) &= \frac{4\pi}{3}\left(R_1^3 - R_2^3\right)\rho_1 \\ N_{true}(m_2 = 1) &= \frac{4\pi}{3}\left(R_2^3 - R_3^3\right)\rho_2. \end{aligned} \tag{16}$$

It should be noted, as indicated on Fig. 5, that a volume *V*ik is in fact composed of two identical sub volumes, one being before the plane of projection (passing through the centre of all spheres), the other being after the plane of projection. Also, note that all planets within a bin mass need not strictly to be of the same mass. They just need to be with a mass between the defined mass bin limits, and be with their polar axis isotropically distributed, whatever is their mass.

2.4.2 *Generalization with the volume matrix*

The linear system of equations (15) may be solved by extracting ρ_0 from the first equation of the system and reporting into the second equation, and so on and so forth, in a 'onion-peeling' scheme. It may also be put under a matrix form, and solved by the inversion of the triangular matrix M_V:

$$[N] = M_V[\rho] \tag{17}$$

$$[\rho] = M_V^{-1}[N], \tag{18}$$

where [N] and $[\rho]$ are column vectors containing respectively the number of observed $m \sin i$ masses in each 'cylindrical shell of sight' and the 'densities' of planets in corresponding true mass bins.

We may call matrix M_V the volume matrix. The computation of the elements of the matrix, which are the volumes, common to spherical and cylindrical shells is detailed in Appendix A.

Once the vector of densities $[\rho]$ is found by the resolution of the linear system 15, the true number of planets contained in a particular mass bin is retrieved by a formula of the type 13bis, more accurate than the differential form 13:

$$N_{true}(m_k) = \frac{4\pi}{3}\left(R_k^3 - R_{k+1}^3\right)\rho_k. \tag{19}$$

2.4.3 *The computation of the errors (or uncertainties)*

The uncertainties on the resulting retrieved histogram of true masses are basically produced by statistical uncertainties on the number N_j of planets in each bin j of the initial $p = m \sin i$ histogram, which is simply (Poisson law) $\sqrt{N_j}$. Below we describe how these initial uncertainties in the observations may be propagated to the final number of planets in a true mass bin.

Once the limits of mass bins have been selected, the volume matrix M_V can be computed without any error. Each element of the matrix is the volume of a ring as defined above. Therefore, the inverse matrix M_V^{-1} may be also computed, with no error.

Equation (18) relates the number density ρ_i to the observed number of planets in each bin of cylinders (apparent masses) N_j.

Calling a_{ij} an element of matrix M_V^{-1}, we have the relation

$$\rho_i = \sum_j a_{ij}\, N_j. \tag{20}$$

The number of planets N_j are integer numbers, and should follow a statistic Poisson law, in which the variance of N_j is N_j, and the error on N_j is $\sqrt{N_j}$. To propagate these errors into the errors on ρ_i, we apply the formulation of the variance of a linear combination of random variables X and Y,

$$Var(aX + bY) = a^2\, Var(X) + b^2\, Var(Y) \tag{21}$$

$$Var(\rho_i) = \sum_j a_{ij}^2 Var(N_j) = \sum_j a_{ij}^2 N_j. \tag{22}$$

The number of planets with true mass $N_{\rm true}(m_i)$ within the spherical shell between R_i and R_{i+1} is related to the density ρ_i by equation (19). Therefore, the variance of $N_{\rm true}(m_i)$ is

$$Var\left(N_{true}\left(m_i\right)\right) = \left(\frac{4}{3}\pi\left(R_i^3 - R_{i+1}^3\right)\right)^2 Var\left(\rho_i\right) \tag{23}$$

$$Var\left(N_{true}\left(m_i\right)\right) = \left(\frac{4}{3}\pi\left(R_i^3 - R_{i+1}^3\right)\right)^2 \sum_j a_{ij}^2 N_j \tag{24}$$

and the errors on $N_{\rm true}(m_i)$ are simply the square roots of the variances.

2.5 The case of a power law mass distribution

Several authors (e.g. Ananyeva et al. 2020) have compared observed $m \sin i$ distributions to true mass distributions coming from planets

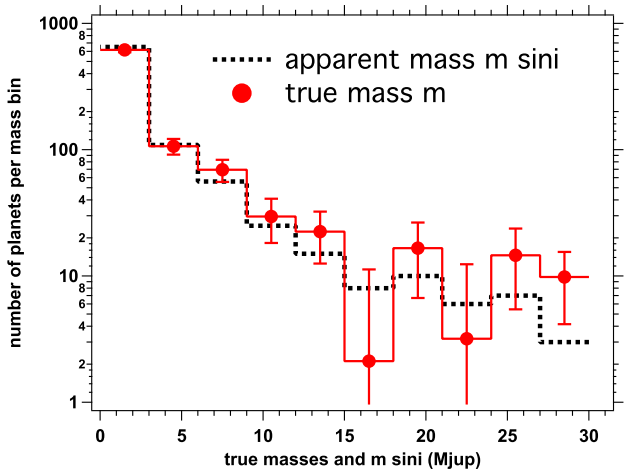

Figure 6. Histogram of observed apparent masses $m \sin i$ (black dashed line) and retrieved true masses (red solid line and red filled circles with error bars). The bin mass size is 3 $M_{\rm jup}$. The log scale is chosen to emphasize the small numbers of large planets. The irregularities in the bins 15–18 and 21–24 $M_{\rm jup}$ (with small numbers of planets) of the apparent mass distribution are amplified by the inversion process. For masses below $\leq 12 M_{\rm jup}$, the values of original and inverted mass distributions are similar.

formation models. They usually examine the apparent mass distributions in terms of dN/dm, (here, m is the apparent mass) and are using power-law fits of the form $\mathrm{d}N/\mathrm{d}m \propto m^{-k}$. And they generally take for granted, without demonstration, that the true mass distribution of the observed planets is having the same power-law index $-k$. This question has not been addressed, neither by Chandrasekhar & Münch (1950) nor by Jorissen et al. (2001) or any other authors to the best of our knowledge. With our geometrical representation of exoplanets, it can be demonstrated (see below) that it is indeed the case: both apparent masses and true masses have the same power law index.

First, we demonstrate (at end of Appendix B) that the Abel transform $F(y)$ (equation B1) of a power-law function $f(r) = r^{-k}$ is also a power law with a higher exponent, $F(y) = Constant \times y^{-k+1}$.

Then we examine equation (12) when we put a true mass distribution of the form $\mathrm{d}N/\mathrm{d}m = N(m) = m^{-j}$. It comes

$$f_0(p) = p \int_p^{max} m^{-j-2} \frac{m}{\sqrt{m^2 - p^2}} \, \mathrm{d}m. \qquad (25)$$

We recognize the integral as the Abel transform of the power law m^{-j-2}. Therefore, it will yields a power law p^{-j-1}. And since it is multiplied by p in equation (25) to yield the resulting apparent mass distribution $f_0(p)$ will be $\propto p^{-j}$. The conservation of the power-law index is also true when a distribution dN/d($logm$) is considered, because $\mathrm{d}N/\mathrm{d}(logm) = m^{-j+1}$ when $\mathrm{d}N/\mathrm{d}m = m^{-j}$.

In Appendix C, we describe simulations (forward modeling and our inversion scheme), with a single-peak mass distribution, a double peak distribution, and a simulation attached to a modification of the observations presented in Fig. 6 (0–30 $M_{\rm jup}$) by the addition of one extra-bin with high masses.

All in all, these simulations, together with the estimate of the error bars, are cross-validating our inversion scheme, and we proceed with confidence to three examples of histograms, differing by the limits of the mass bins, but otherwise using our general sample of 909 planets discovered by the RV technique and screened as explained below.

3 APPLICATIONS OF THE INVERSION METHOD TO 2021 DISTRIBUTION OF $m \sin i$

In this section we apply our inversion method to the catalogue of observed exoplanets with the radial velocity method (RV planets), for which $m \sin i$ was determined. The data were extracted in 2021 August from the Encyclopedia of exoplanets, and contained 960 exoplanets at that time. However, we had to exclude from this ensemble some exoplanets for which the $m \sin i$ was determined by the RV method, but which had been discovered first by the transit method, and later on followed in RV in order to determine the mass (even if the planet of a multiplanet system was discovered by RV method). Indeed, in such a case we know that the inclination i is near 90°, and keeping them would distort the distribution of inclination angles that our method assumes to be random. This is the case of all Kepler planets, as well as other surveys like WASP, KELT, Corot, HAT, and TOI (Tess Objects of Interest). On the contrary, some transiting exoplanets were kept, because their existence was first determined from a RV survey and later on it was found that they were transiting. Rejecting them would also distort unduly the statistics. After such a filtering the resulting sample contained a total of 909 exoplanets orbiting around 651 stars.

One caveat is in order here. In addition to the $m \sin i$ effect, there are several factors that can bias an observed apparent mass distribution from the really true mass distribution: small masses are below the detection threshold, long period planets missed by the limited time span of a survey, different stellar spectral types may have different planet mass distribution etc. In the present paper, we address a method, with some applications to the whole sample of available RV discoveries, and ignoring all possible other biases. It is clear that the method could be applied to samples of various stellar spectral types, or could be used in conjunction with some methods that were designed to estimate other biases. This is well beyond the scope of this paper.

In our method, the bins of masses may be selected *ad libitum*. The matrix of volumes will be different for each version of the bin grid, and we will show several examples with different bin grids, focusing on various aspects of the mass distribution.

3.1 Linear histogram up to 30 $M_{\rm jup}$

We first built a histogram of 10 mass bins, with a size bin of three Jupiter masses ($M_{\rm jup}$) up to an upper boundary of $m \sin i = 30\ M_{\rm jup}$ which defines our outer boundary for 'radii of spheres'. There are 891 planets with $m \sin i$, up to 30 $M_{\rm jup}$. On Fig. 6 is represented the histogram of apparent masses $m \sin i$, (black dashed line) as well as the inverted distribution of 'true' masses m (red solid line and red circles with their error bars) which is giving the observed distribution of $m \sin i$ when randomly observed over inclination angles. It is seen that the number of planets observed in the $m \sin i$ bins 15–18 and 21–24 $M_{\rm jup}$ show a slight deficit w.r.t. their neighbouring bins, irregularities likely due to statistical fluctuations of low integer numbers (~6–10). These slight deficits are largely amplified after inversion into the 'true' mass distribution, but not to the point to get negative in this case (there is no mathematical constraint on the sign of the result of the inversion). This is a well-known effect of the inversion of Abel's integral. The sum of the number of planets in the true mass distribution (which are all floating point, decimal numbers) is also 891: no loss of planets in the inversion process. The shape of the distributions, both the $m \sin i$, and the true mass distribution, are quite similar on this coarse histogram. The error bars are also

displayed and reflect the errors (statistical fluctuations) on the input histogram of apparent masses.

3.1.1 Conservation of planets in the inversion process

In the frame of our inversion scheme, we assume that there are no planet with a true mass beyond a certain upper limit, which is common to both true masses and observed apparent masses. Therefore, with this limiting assumption, the planets within the cylinder of maximum radius p_{max} are assumed to have also a true mass $\leq p_{max}$. This is why the number of planets is conserved in our inversion process. It is true that a planet having a real mass $m > p_{max}$ (out the sphere of maximum radius) but an apparent mass $p < p_{max}$ lies within the observed cylinder of radius. The inversion algorithm will put it in a true mass bin smaller than p_{max} therefore at a 'true' mass smaller than the real mass, most likely in the bins of largest masses. The planet is not lost, but put at the upper end of the mass interval (actually the first bin).

We have studied (as described in Appendix C1) the effect of adding an extra bin of large masses to the exponential histogram of Section 3.2. The effect is insignificant on the number of planets $<12.89\ M_{jup}$, which remain the same, with or without the extra bin. See Appendix C1 and Table C1.

Classically, the mass limit between exoplanets and Brown Dwarfs (BD) is set at 13 M_{jup}, above which internal conditions allow for Deuterium fusion, while the upper limit of BD is $\sim$80 M_{jup}. Therefore, eliminating RV discovered objects (BD) with $m \sin i > 30 M_{jup}$, or keeping them, which do not affect the retrieved true mass up to 12.89 M_{jup} as shown above, has no influence on the realm of exoplanets ($m < 13 M_{jup}$). However, this exercise shows also that it is important to keep the RV discovered objects up to $\sim 30 M_{jup}$, well within the regime of BD, in order to keep unaffected the regime of exoplanets.

3.2 Exponential histogram from 1 $M_\oplus$ to 26 M_{jup}.

With the choice of the linear 10 mass bins of 0–30 M_{jup} (Fig. 6), we see that the number of planets is strongly decreasing with their mass (observed or inverted), but we do not learn anything on the details of the mass distribution of planets below $\sim$3 M_{jup}. To learn more about the details of this mass regime, a choice of a log scale for the mass bins is preferable, as described below.

Here we are starting from the same set of 909 exoplanets as in Section 3.1, but the bin limits on $m \sin i$ (and m) are different. The bin containing the smallest planet is between 1 and 2 Earth Mass = 1 $M_\oplus$ = 0.00314 M_{jup}. The next bin is from 2–4 $M_\oplus$, and the limits are doubled for each successive bin up to the largest mass at 25.76 M_{jup}. In total, 885 planets are contained in all the 13 bins so defined. This distribution of apparent masses ($m \sin i$) is plotted on Fig. 7 as a black dashed line. We have inverted this $m \sin i$ distribution with our method, and the resulting true mass distribution is represented by the red line on Fig. 7, with associated error bars. It is a unique 'true' mass distribution of exoplanet that would result in the observed distribution of $m \sin i$, for a random sample of inclination angles.

Let us first forget the black lozenges and compare the histogram of the observed $m \sin i$ (black dashed line on Fig. 7) with the histogram of the retrieved true masses (red histogram and red circles with error bars). Both histograms show two distinct peaks. Though the two peaks are at the same position, the retrieved histogram of true masses is systematically shifted to higher masses w.r.t the observed $m \sin i$ histogram, for all the four slopes of the histograms. The shift

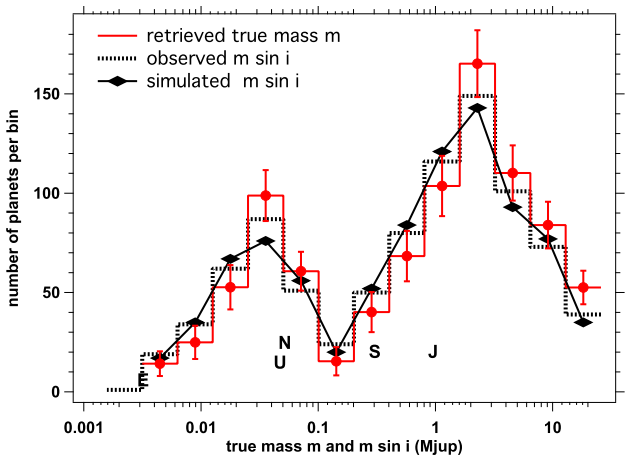

Figure 7. Dashed black line: Distribution of $m \sin i$ of 885 planets detected by the radial velocity method, from 1 $M_\oplus$ = 0.00314 M_{jup} up to 25.7 M_{jup}. Solid red line and red filled circles: distribution of true masses obtained by inversion with our new method with error bars Letters E, U, N, S, J are representing the masses of Earth, Uranus (14.54 $M_\oplus$), Neptune (17.2 $M_\oplus$), Saturn (95 $M_\oplus$), and Jupiter (317.8 $M_\oplus$), respectively. The lowest bin in the central valley has limits 0.1–0.2 M_{jup} ($\sim$32–64 $M_\oplus$). This sub-Saturn desert is even more depleted after inversion of the $m \sin i$ distribution. The black lozenges are the result of a simulation in which a random value of $\sin i$ was assigned to an ensemble of 885 planets distributed in mass according to our retrieved histogram of true masses. Then their histogram is plotted as black lozenges: they have indeed similar values as the original observed values of apparent masses.

is about the same in this log plot of the mass. This reflects the fact that adding sin i to the true mass decreases the apparent mass w.r.t. the true mass, and we see the average effect on the histograms (factor $\pi/4$).

Both histogram distributions (apparent and true masses) are similar, showing two conspicuous peaks separated by a deep valley, the so-called 'sub-Saturn desert' (Arriagada et al. 2013). The low mass peak (left) in the bin 0.025–0.05 M_{jup} is certainly an artefact due to depletion of lightest planets, below the accuracy threshold (in m s^{-1}) of the RV detection method. The high mass peak (bin 1.66–3.22 M_{jup}) is genuine, and the valley is deepest in the bin 0.1 to 0.2 M_{jup} ($\sim$32–64 $M_\oplus$).

The black lozenges are the result of a simulation, performed following a suggestion of the referee. From the retrieved true mass histogram, we have built a mass distribution of 885 planets with a linear variation of the mass inside the bin between the limits of the bin. Then, we have multiplied each mass by a random series of $\sin i$, and built the histogram of this new, simulated, $m \sin i$ distribution (black lozenges). When compared to the observed histogram of original observations, we see that the lozenges are not far from the original observed $m \sin i$ histogram (black dashed line). Small differences may be assigned to the random values of $\sin i$.

In addition, simulations were made on a distribution with two narrow mass peaks, though separated by a mere factor 1.36 (Appendix C). The inversion algorithm is quite efficient to retrieve distinctly the two peaks.

Another similar exercise was performed, with results displayed on Fig. 8. An artificial distribution was built from the retrieved histogram, with the values within the five bins between the two peaks being replaced with a linear variation between the peaks, in order to fill the gap between the two peaks. As above, we have built a distribution of 1264 planets with linear variation of the mass within

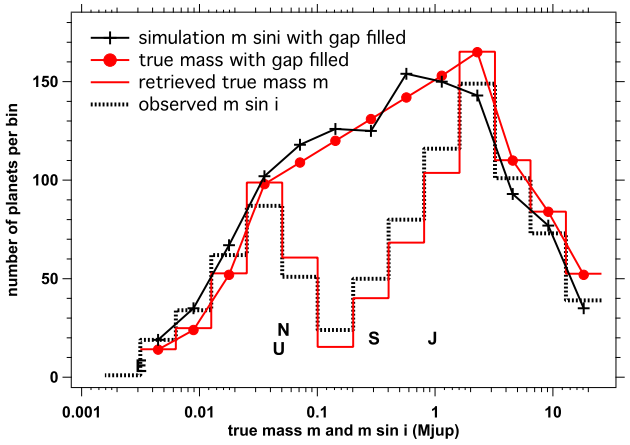

Figure 8. The two histograms are the same as on Fig. 7. The red circles represent a dummy true mass distribution, modified from the retrieved histogram by filling the bins in the valley (linear variation assumed between the two peaks), containing 1264 planets. Then a random value of sin i was assigned to all planets, The black + is the histogram of of the new m sin i. It shows no valley.

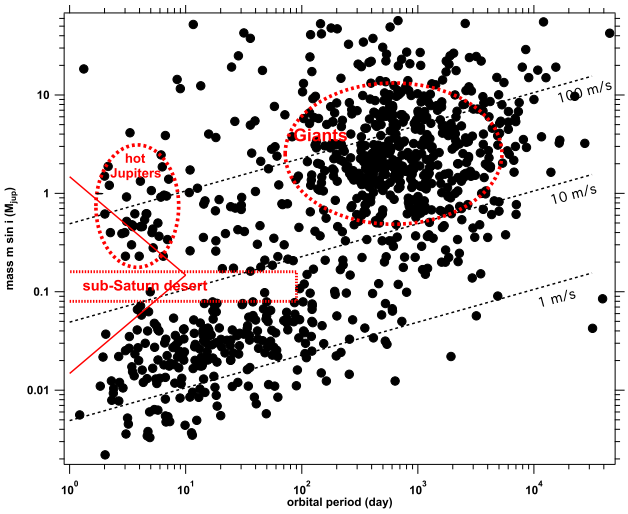

Figure 9. Each black point represents one of the 909 exoplanets that were examined extracted form archives, with period and m sin i as coordinates. The rectangular red dashed line is a contour of a region particularly void of planets, w.r.t. the vicinity. It corresponds to the sub-Saturn desert and seems to affect mostly planets with periods <100 d. The two solid lines correspond to the desert limits of Mazeh, Holczer & Faigler (2016) (see text in Section 4.1). A hot Jupiters cluster and a Giants cluster are delimited with a red dashed oval. Three dashed lines at 1,10, 100 m s^{-1} indicate what is the reflex motion of a star of 1 solar mass, as a function of the period for a circular orbit. This gives an idea of what could be potential biases of RV surveys.

the limits of each bin and multiplied by a random distribution of sin i, multiplied by a random series of sin i, and built the histogram of this new, simulated, m sin i distribution (black + of Fig. 8).

When compared to the observed histogram of original observations, we see the following features:

(i) the whole new histogram (+) is globally displaced to the left [multiplication by sin i (average value $= \pi/4$)].
(ii) between the two peaks, the gap is filled.
(iii) at the peaks and outside the peaks, the (+) are not far from the original values of the observed histogram.

This exercise demonstrates that if the true mass distribution had no valley between the two peaks, the corresponding m sin i would not show either a valley.

Coming back to Fig. 7, we see that the overall shape of the retrieved histogram is more peaked than the original histogram: the sub-Saturn desert is more pronounced.

These three features (two peaks and one valley in between) are therefore conserved in the inversion process, and are even enhanced through inversion. This is somewhat expected, since for a given bin of observed m sin i, there are contributions of several populations of m, from all $m > m$ sin i. This has the effect of smoothing the original true mass m distribution, and the inverse of smoothing is expected to enhance the differential features.

The contrast between the trough ('desert') and the two peaks has indeed increased in the inversion process. The ratio of number of planets in the high mass peak bin (1.66–3.22 $M_{\rm jup}$) to the bin trough (0.1–0.2 $M_{\rm jup}$) increases from 6.21 in the observed m sin i distribution to 10.7 in the true mass distribution, a factor 1.72 of increase. Similarly, the ratio of number of planets in the low mass peak bin (0.025 to 0.05 $M_{\rm jup}$) to the bin trough (0.1–0.2 $M_{\rm jup}$) increases from 3.6 in the observed m sin i distribution to 6.4 in the true mass distribution, a factor 1.78 of increase. For planets discovered by the RV method, the 'sub-Saturn desert' is more depleted in the true mass distribution than in the m sin i measured distribution.

The void of planets is roughly between the masses of Neptune and Saturn. There is no planet either in our Solar system in this interval of masses. On Fig. 9 is represented in a m sin i/period diagram the population of 909 exoplanets that we have selected, being discovered by observations with the RV method. We have delimited a rectangular contour (dashed red line) whose inside is conspicuously depleted in planets. For periods larger than 100 d, the depletion is less visible, or absent. We are also delimiting by dashed ovals two regions of locally higher concentration of planets, a hot Jupiters cluster and a Giants cluster. We discuss further the mass distribution in Section 4.1.

3.3 Composite histogram from 1 to 230 $M_\oplus$

Quite recently (in the last 4 yr) there has been an interesting development in the statistics of the radius of transiting planets, with the discovery of a valley between super Earths and Neptune type planets, e.g. van Eylen et al. (2018), and Fulton & Petigura (2018).

In particular, Petigura (2020) was able to update the radii of transiting planets in the Kepler surveys, with three improvements: the use of the Gaia results which allows a better estimate of the occulted star radius, eliminating inaccurate determinations when the transit is on the edge of the star (impact parameter must be <0.8 R_s), and performing simulations of the photometric noise, keeping the posterior median. Fig. S1 (Supplementary material) is the net result of Petigura (2020) investigation on planet radii distribution, showing a deep valley (or gap) between two peaks at ∼1.4 and ∼2.5 $R_\oplus$, assigned respectively to Earths or super-earths (probably without an extended thick atmosphere), and to Neptune size planets with a solid core and a thick atmosphere.

Stimulated by this important new finding, we wished to investigate if there were in the mass distribution of planets, a counterpart of this valley (gap) identified in the radius distribution. The fact that RV planets yield the m sin i apparent mass is detrimental to this kind of investigation, because features in the true mass distribution might be smoothed out by the distribution of the inclination angle i.

Nevertheless, we attempted to look carefully to the m sin i distribution. On Fig. 10 is represented the m sin i of our sample

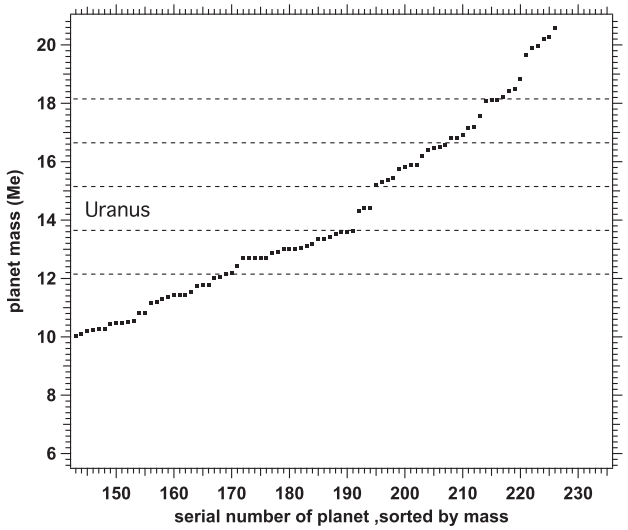

Figure 10. The set of 909 RV planets are sorted by increasing apparent mass ($m \sin i$), and the mass is plotted as a function of the serial number of the planet after sorting. This is a detail of the sorted diagram, suggesting a lack of observed planets in a particular range of masses which contains Uranus. Horizontal dashed lines are the limits of some bins used in the inversion of the $m \sin i$ distribution.

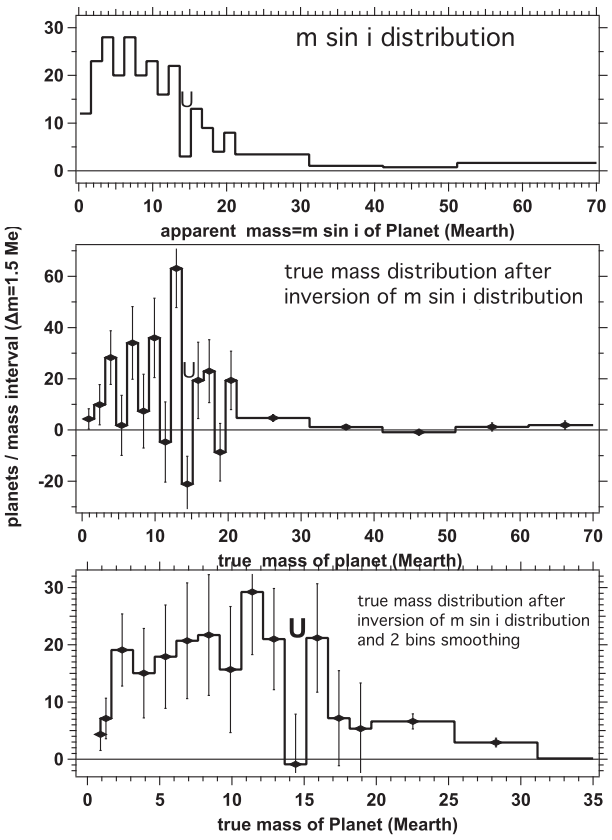

Figure 11. Top panel: observed distribution of apparent masses ($m \sin i$) of 909 planets discovered with RV methods, with bins of equal width = 1.5 $M_\oplus$ up to mass 21.5 $M_\oplus$, followed by 10 $M_\oplus$ bin sizes up to 211.5 $M_\oplus$, with a total of 378 planets. The bin marked U (for Uranus) contains only three planets, between two adjacent bins containing 22 and 13 planets. The middle panel contains the result of the inversion yielding the true mass distribution with 1σ error bars, showing several negative numbers. The bottom panel is obtained by a sliding two bin average of the middle panel numbers and corresponding error bars. The lowest number (below 30 $M_\oplus$) contains −0.87 planets, confirming the low number of planets in this true mass bin.

of 909 planets, sorted by increasing $m \sin i$. The slope of the curve is d($m \sin i$)/dN, (N number of planets) therefore a low slope indicates a large number density of planets per *mass*sin i bin, dN/d($m \sin i$), and a large slope indicates on the contrary a low number density of planets. We were struck by the conspicuous behaviour of the mass sorted curve around planets number 190–200: there are only three planets between planet n° 191 with mass (sin i) 13.64 $M_\oplus$ and planet n° 195 with mass (sin i) 15.197. We have therefore constructed a histogram of $m \sin i$ with adjusted positions of bin limits. We started the histogram at 0.15 $M_\oplus$, with a bin size of 1.5 $M_\oplus$, up to mass 21.5 $M_\oplus$. Then we used 10 $M_\oplus$ bin sizes up to 211.5 $M_\oplus$, with a total of 378 planets up to this limit. The limits of some of the 1.5 $M_\oplus$ bins are indicated by horizontal dashed lines on Fig. 10. The bin which has only three planets contains Uranus at 14.54 $M_\oplus$. The histogram of $m \sin i$ planets is represented on Fig. 10 (top panel), showing indeed the Uranus bin with only three planets, while the bin before contains 22 planets and the bin after contains 13 planets. On the figure, the number of planets for the larger bins have been multiplied by 1.5/10, in order to homogenize the values of dN/d($m \sin i$) across the various bins therefore yielding a 'number density' of planets per 1.5 $M_\oplus$ interval. Of course, the original numbers in all bins were conserved to perform the inversion. The same homogenization was done for the other distributions shown on Fig. 11.

The raw result of the inversion is somewhat noisy with some non-physical negative numbers (middle panel of Fig. 11 with 1σ error bars), and we performed a simple smoothing, by averaging together two successive bins. The abscissae had to be adjusted accordingly by a shift of 0.5 bin size. We also homogenized the distribution by scaling down the content of larger bins, in order to get, whatever is the size of the bin, a number of planets per 1.5 $M_\oplus$ bin. The resulting 'true mass' histogram is represented on Fig. 11, bottom, with 1σ error bars. Two successive 10 $M_\oplus$ bins covering the range 32–52 $M_\oplus$ have a very low number density of planets, respectively 0.15 and 0.18 planets per 1.5 $M_\oplus$ interval (only the beginning of the bin is seen on this panel). They correspond to the sub-Saturn desert, confirming our analysis of Section 3.2, with even a lower number density dN/dm of planets of ∼0.1 planet per 1 $M_\oplus$ interval, instead of ∼0.5 planet per 1 $M_\oplus$ interval as indicated in Fig. 7

The number of planets in the bin corresponding to the mass of Uranus has decreased from 3 in the raw data histogram of $m \sin i$, to a slightly negative value of −0.87 planets. The truth is certainly positive, but it indicates indeed a very low number density dN/dm in this narrow mass range. Could the low number of three planets in the Uranus bin of the $m \sin i$ histogram just due to random noise among low integer numbers? This bin is between two bins containing, respectively, 13 and 22 planets: if we interpolate the neighbour trends to the bin with 3, we might expect, instead of 3, 17.5 (interpolating between 13 and 22). Alternately, we could also in addition consider the bin before containing 19 planets, yielding an estimate for the Uranus bin 15.4 planets. Then we compute the probability to get an integer number less than 4 (0,1,2, or 3), with a Poisson law with parameter λ of 15.4, or 17.5. We use the CDF and find a probability of 1.5 10^{-4} for a Poisson parameter $\lambda = 15.4$, and 2.7 10^{-5} for $\lambda = 17.5$. Therefore we conclude that this low number of three planets detected in the Uranus mass bin is likely not due to random fluctuations. To the best of our knowledge, this is the first time that a deficit of planets is identified

in this narrow mass window, containing Uranus, as a puzzling question.

This gap is in fact quite narrow, $\Delta m/m = 0.1$; much narrower than the valley, or gap, in the radii distribution as found by Petigura (2020; Fig. S4, Supporting material) from ∼1.7 to 2.2 $R_\oplus$ or $\Delta R/R = 0.4$. With a constant density of planets, we would expect on the contrary $m \propto R^3$ and $\Delta m/m = 3\Delta R/R$. But the density of planets is far from being constant, because the core may be formed of ice, rocks, or iron, and the atmosphere may extend way off, increasing the opaque radius accordingly. In the next section we discuss further the questions of sub-Saturn desert, radius valley, and Uranus gap in the context of a diagram radius versus mass of observed exoplanets.

4 DISCUSSION ABOUT THE RADIUS/MASS DISTRIBUTION OF OBSERVED PLANETS

4.1 The mass distributions and the sub-Saturn desert

When examining a population of exoplanet (either observed or modeled) in a two-parameters representation (mass/period, or mass/semi-axis), a region fully empty of planets may be qualified as a 'desert', especially if it is surrounded by populated regions. We may extend somewhat this definition to include a region significantly less populated than the vicinity, without requiring it to be a region totally void of planets. We may also define a 'cluster' of planets, as a region substantially more populated than the surroundings. Actually, between two distinct clusters, there will be a desert. If a cut through a two-parameters distribution is made by fixing one parameter, the resulting 1D curve will show a dip at the place where the cut goes through a desert, which can also take the form of a gap between two peaks.

The first mention of a 'planetary desert' in the extra-solar planets distribution can be found in the paper of Ida & Lin (2004) who modeled the planet formation process with a core-accretion paradigm. They examined giant planets growth and found that since planets masses grow rapidly from 10 to 100 $M_\oplus$, the gas giant planets rarely form with asymptotic masses in this intermediate range. Some versions of their model predict a paucity of extra-solar planets with mass in the range 10–100 $M_\oplus$ and semimajor axis less than 3 AU. Both the mass lower and upper boundaries of the desert depends on the semi-axis a. Actually, they argue that the location of these two boundaries are dictated by the conditions of formation and evolution, (i.e. the core mass that can initiate the onset of rapid gas accretion, the truncation mechanism of gas accretion, the region where type II migration can be halted, etc). The location of the boundaries of the desert changes with model parameters (Ida & Lin 2004), but their lower boundary is always around 4–8 $M_\oplus$, for periods ≤50 d. This prediction was contradicted by Howard et al. (2010), reporting on the RV survey of 235 nearby stars. They found nine planets plus four candidates and an estimated 15 missed planets (with a de-biasing exercise) in the range 3–30 $M_\oplus$ and periods ≤50 d. So, Howard et al. (2010) concluded that 'the models need substantial revision'. We are comparing on Fig. S1 (Supplementary Material) the histogram of our sample of 216 planets with periods <50 d to Howard et al. (2010) histogram: they are consistent within error bars up to ∼100 $M_\oplus$.

To the best of our knowledge, the first paper mentioning a gap in the observed $m \sin i$ distribution is Mayor et al. (2011), based on the 155 planets found at that time (in 102 planetary systems), from which more than 2/3 were discovered from HARPS and CORALIE surveys. In the Section 1 they say: 'A bi-modal mass distribution is observed for low mass planets and giants'. In the caption of their Fig. 10 representing the $m \sin i$ histogram of these 155 planets, it is said: 'We also remark a gap in the histogram between planets with masses above and below ∼30 $M_\oplus$'. The authors noticed that there is a sharp decrease of the distribution between a few Earth masses and ∼40 $M_\oplus$, followed by an increase. The minimum in-between is claimed by the authors to correspond to the minimum predicted by the planet population synthesis modeling produced by Mordasini et al. (2009). Therefore, this minimum is at much larger masses than the 'desert' predicted by Ida & Lin (2004). In fact, the position of this minimum coincides with the position of the minimum of the observed sub-Saturn desert shown on Fig. 8. We note that this Mayor et al. (2011) was never published, is only available on arXiv, but was cited 609 times (2021 August).

Howard (2013) wrote an overview paper on exoplanets, adding new detections of planets with the transit method. He discussed again the mass distribution predicted by models: '. . . these models produce 'deserts' of reduced planet occurrence precisely where Doppler and transit surveys detect a great abundance of planets'. Looking at his fig. 2, it is clear that, when he is talking of 'desert', Howard (2013) is referring to the mass range below 30 $M_\oplus$, like in his previous 2010 paper ('from 5–30 Earth masses'), and not above 30 $M_\oplus$. Therefore, this 'populated desert' has nothing to do with our sub-Saturn desert, most conspicuous in the range ∼32–64 $M_\oplus$. With our new analysis of RV surveys, we see also plenty of planets below 30 $M_\oplus$, and therefore there is no contradiction here between Howard et al. (2010) and our work: we find no desert below 30 $M_\oplus$, the $m \sin i$ distribution is increasing with lower masses, until a peak is reached (due to RV sensitivity limits). Howard (2013) claimed nothing about a desert, present or absent, in the range of masses 30–1000 $M_\oplus$, having only six RV planets (period <50 d) spread over three mass bins, with two planets per bin.

Who used first the denomination sub-Saturn desert? Arriagada et al. (2013) published the discovery of two planets around a K7 star, one of which with a minimum mass ($m \sin i$) of 53 $M_\oplus$ in the sub-Saturn desert mass range. It seems that this paper contains the first denomination of the sub-Saturn desert, both in the title, and in the abstract, but nowhere else in the body of the paper.

In retrospect, there are some papers showing scatter plots of exoplanets where a close examination, and integration 'by eye' mass/period, suggests strongly the existence of this sub-Saturn desert. For instance, Benítez-Llambay, Masset & Beaugé (2011) produced a period/ mass scatter plot (their fig. 1) where the gap is obvious by eye but did not pointed out the gap. Howard (2013) showed a mass/radius scatter plot of transiting planets (his fig. 3). Integrating by eye over the radius, a deficit of planets in the range 30–60 $M_\oplus$ is observed. We have produced a new histogram with all points of his fig. 3 where the deficit is obvious in the sub-Saturn desert (see fig. S2, Supporting material, on line).

Mazeh et al. (2016) have examined the mass/period distribution of exoplanets an ensemble of RV and transit discovered planets); focusing their attention on short-period Neptunian exo-planets. They tried to find mathematical relationships describing the lower boundary and the upper boundary, in the spirit of Ida and Lin (2004): 'The derived shape of the desert, which might extend up to periods of 5–10 d, could shed some light on the formation and evolution of close-in planets'. They found that the desert boundaries have the shape of two straight lines (represented as red solid lines on Fig. 9 forming a triangle (in a logM–logPeriod plot) converging at a period of 10 d. This is somewhat in contrast with our RV sample displayed on Fig. 9, where a rectangular zone (in red) materialize the upper and lower boundaries of the sub-Saturn desert up to periods of 100 d. Is this difference linked to the fact that our sample contains only RV planets? This cannot be excluded. We note, however, that

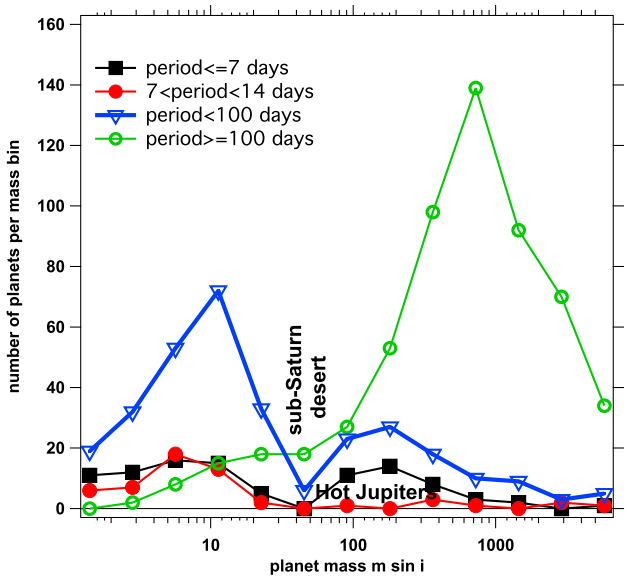

Figure 12. The planets displayed on Fig. 9 have been integrated over various slices of orbital periods to yield the corresponding $m \sin i$ distributions: period $\leq$7 d (black squares), 7 $\leq$ period $\leq$ 14 d (red filled circles), period $<$ 100 d (blue triangles), and period $\geq$100 d (green open circles). The position of hot Jupiters and sub-Saturn desert are indicated. The desert is most conspicuous on period $<$ 100 d blue curve. Other curves obtained for 100 $\leq$ period $\leq$ 1000 and period $\geq$1000 d (not shown here for clarity) may be seen in Supplementary material (on line). See text for comments.

Yakovlev et al. (2022) have shown that, if transiting planets whose mass has been determined, not by the RV method, but more indirectly (and with less accuracy) by the Transit Time Variation method, are excluded from the sample of transit planets, then most of transit planets disappear from the rectangular sub-Saturn area, at least up to ∼50 d. It is interesting, in this context, to examine the mass distribution of our RV sample, when restricted to various ranges of periods, as displayed on Fig. 12 (1–7 d, 7–14 d, 14–100 d, >100 d).

The mass distribution (actually, $m \sin i$) for the short periods 1–7 d (black squares) displays two peaks separated by the sub-Saturn desert (0 planet in 32–64 $M_\oplus$ bin): the first peak of light planets on the left, and a second peak of heavy planets on the right which corresponds to the cluster of hot Jupiters, which is conspicuous on Fig. 9. The 7–14 d distribution (red circles) has a similar light planets peak, but the hot Jupiters are absent. The 14–100 d distribution has a strong light planet peak (certainly biased for very light planets), with again some heavier planets, and a gap at the sub-Saturn desert. The distribution of planets with periods ≥100 d has a strong peak in the mass range 512–1024 $M_\oplus$ (1.6–3.2 M_{jup}), with much smaller numbers of planets below ∼100 $M_\oplus$. In particular, the two bins 16–32 $M_\oplus$ and 32–64 $M_\oplus$ both contain 18 planets. It is likely that these numbers are biased by the sensitivity of the RV survey. But it is clear that the bias factor is larger for the 16–32 $M_\oplus$ bin than for the 32–64 $M_\oplus$ bin (mass effect on RV surveys). After de-biasing, the 16–32 $M_\oplus$ bin will contain more planets than for the 32–64 $M_\oplus$ bin. Actually, the same argument holds also for all the other curves, and therefore holds for the sum of all distributions. This sum is the black dashed line of Fig. 7, well showing the sub-Saturn desert in the bin 32–64 $M_\oplus$. We note that our de-biasing technique is based on the multiplication of the vector N representing the $m \sin i$ distribution by a matrix M_V^{-1}, which operates as a linear operator. Therefore, the inversion of a sum of distributions is equal to the sum of the inversions of all distributions.

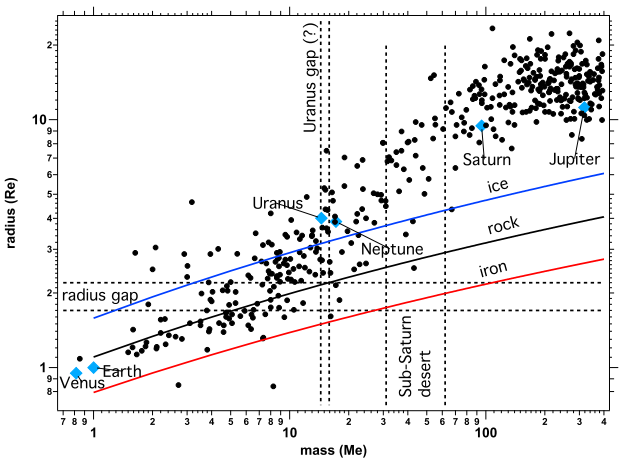

Figure 13. Mass and radius of known 570 exoplanets, distributed along a band through the diagram. The radius was determined form transit observation, and the mass from the RV method. Solid lines are the relations between mass and radius, according to models of Fortney et al. (2007a,b), for pure ice, rock, and iron. Three domains are delimited by vertical or horizontal lines: the sub-Saturn desert, the radius gap as defined by Petigura (2020), and the putative Uranus gap. See text for further discussion.

We note that, in a recent re-analysis of the Mayor et al. (2011) sample of 155 RV planets, Bennett, Ranc & Fernandes (2021) claimed the absence of a sub-Saturn desert in this historical sample, when splitting the planet mass distributions in several bins of periods. However, in their second bin of periods 30 $<$ period $<$ 240 d, they mix together 30–100 and 100–240 d, which will blur the sub-Saturn gap, because the sub-Saturn desert is seen mainly up to 100 d (Fig. 12). In addition, in the $m \sin i$ histogram of fig. 10 of Mayor et al. (2011), the bin 30–48 $M_\oplus$ contains 0 planet, while bracketing bins contain five and four planets. The Poisson probability to get 0 planet if the true average were 4.5 is 1.1 per cent therefore a very low probability of occurrence, pointing to a likely true value significantly smaller in this bin than in the bracketing bins: a gap. At any rate, our sample of 885 planets is now certainly more statistically significant, and supersedes the earlier 155 planets sample.

In conclusion of this sub-section, it can be claimed that the sub-Saturn desert is real. It is pretty obvious and indisputable, on the $m \sin i$ distribution of 885 observed RV exoplanets plotted in Fig. 7, and more visible with the planets with periods shorter than ∼100 d (Fig. 12). We have verified that this 'desert' is certainly not an artefact due to the uncertainty on the true mass m (only $m \sin i$ is measured). On the contrary, we could show that this desert is even amplified in the true mass distribution, obtained with inversion with our numerical procedure, by a significant factor ∼1.7. Actually, comparisons of observations to synthetic models should preferably be done with the true mass distribution, after the proposed inversion, rather than with the raw $m \sin i$ distribution. The general question of de-biasing the observations from some other observational biases is not addressed here. What might be argued is that the inversion should be performed *after* other de-biasing processes, which should take into account the change of sensitivity of the RV surveys around the threshold (in m s^{-1}) and period (connected to the time extent of the RV survey).

4.2 The radius /mass distribution

On Fig. 13 is represented a scatter plot of planets as a function of their mass and their radius. We keep only the transiting planets which transit allows to determine the radius, and which true mass

was determined by RV measurements ($\sin i \sim 1$). We used in this case the NASA archive of discovered transiting planets with RV mass determination. We discarded the planets where the error on the mass is NaN, which indeed result in many outliers lying outside the main stream of the remaining 551 planets, plus 19 planets first discovered by RV measurements which were found later to be transiting (yielding a total of 570 planets). We also draw lines relating the radius to the mass of a (mainly solid) planet for pure ice, rock or iron, from the formulas of Fortney, Marley & Barnes (2007a,b). Also displayed on Fig. 13 are the regions depleted in planets, delimited by horizontal or vertical black dashed lines: the sub-Saturn desert, the radius gap (as determined by Petigura 2020), and the Uranus gap as suggested by our study in Section 3.3.

The existing planets are grouped along a domain delimited by a band crossing the radius/mass diagram in a diagonal direction, the band of existing planets.

(i) the intersection of the mass Sub-Saturn desert and the radius gap occurs in a region where there is no planet at all: all planets in this mass region have a radius showing a composition much lighter than pure iron, most of them being lighter than pure ice, containing therefore a significantly massive atmosphere of hydrogen and helium, and this intersection has no particular meaning. Similarly, the intersection of the radius gap and the Uranus gap has probably no special meaning.

(ii) the sub-Saturn desert: it cuts the domain of existing planets mostly above the pure ice core line, in a radius regime above $\sim 5R_{\oplus}$, indicating that these planets are embedded in a large and significantly massive atmosphere. The classical explanation of this gap is that the rate of accretion from the disc is very rapid during the formation process for this kind of mass, and the probability to stop the accretion process by lack of gas in the disc is small.

(iii) the radius gap: the inner edge of the radius gap at 1.8 $R_{\oplus}$, as found by Petigura (2020), intersects the band of observed planets approximately where the core of the planet is made of pure rock. Below this radius, the core is made of a mixture of rock and iron. Above this radius, the core contains some ice, and apparently is accreting also a significant atmosphere which increases the radius, which will propel these planets beyond the outer edge of the radius gap. Note that, with planets shown in Fig. 13, the radius gap is much less obvious than the one resulting from the work of Petigura (2020), reproduced on Fig. S4 (Supp.mat). This is because the planets plotted in Fig. 13 come from the NASA archive with the originally quoted radius, while Petigura (2020) has made some screening out and some corrections to the actually published values of the radius of exoplanets.

(iv) the Uranus gap (in mass) intersects the band of observed exoplanets approximately at the place of pure ice core. We could speculate that the inner edge of the gap is a limit to the mass of the core of this population of planets, and if they acquire some atmosphere, then they acquire a significant mass of atmosphere, of at least 10 per cent of the total mass ($\Delta m/m = 0.1$). Actually, the model of initial mass function of Mordasini et al. (2009) displays a trough in the range 5–10 $M_{\oplus}$ (Fig. S5 of Supporting material). The putative Uranus gap might be an evolved remnant of this trough.

It could be argued that the very existence of Uranus in the Solar system, which mass puts it right in our tentative Uranus gap, is a strong argument against the reality of such a Uranus gap. However, Uranus is very special in the Solar system, with a spin axis inclined at 98° from the ecliptic polar axis, almost in the ecliptic plane. Safronov (1966) suggested that this large inclination of the spin axis was the result of an early collision with another massive planet.

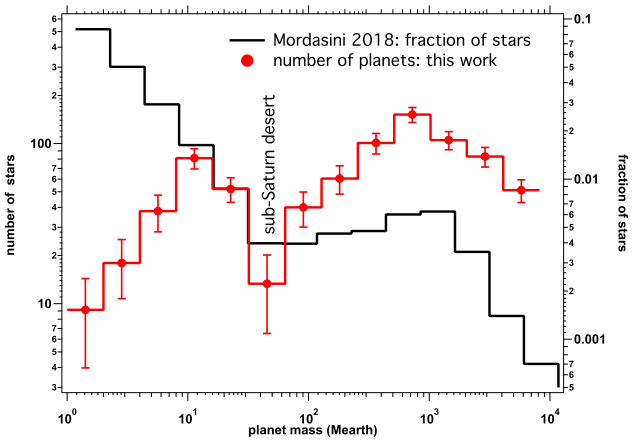

Figure 14. Comparison of the RV star/planet sample (red, left scale) with the synthesis population model of Mordasini (2018; black, left scale). The two scales have been adjusted to be at the same level in the mass bin 16–32 $M_{\oplus}$. The model curve is the fraction of stars which have at least one planet in the mass bin. The red data curve is obtained by our $m \sin i$ inversion scheme applied to the histogram of the number of stars from the original RV sample which have at least one planet in the mass bin. Differences are discussed in the text. The sub-Saturn desert is in the mass bin 32–64 $M_{\oplus}$.

Morbidelli et al. (2012) investigated this mechanism in order to explain the observed system of satellites of Uranus. Kegerreis et al. (2018) performed hydrodynamics simulations to investigate in detail the results of a giant impact on the young Uranus. They found that a number of observed features (inclination of spin axis, lack of heat flow from the interior, asymmetric magnetic field) are best explained by a collision with a 2 $M_{\oplus}$ mass impactor, which mass will be added to the mass of the pre-impact Uranus. Therefore, we may imagine that the pre-impact Uranus had a mass slightly below the lower edge of the putative Uranus gap, and jumped into the Uranus mass gap with the collision of a 2 $M_{\oplus}$ mass impactor.

5 COMPARISON OF THE MASS DISTRIBUTION WITH A PLANETARY POPULATION SYNTHESIS MODEL

In principle, comparisons of observations to synthetic models should preferably be done with the true mass distribution, after some de-biasing process accounting for RV survey biases, followed by the presently proposed inversion, rather than with the raw $m \sin i$ distribution. This is far beyond the scope of this paper. However, inasmuch as our inversion scheme is somewhat approaching the true mass distribution more than than the raw $m \sin i$ distribution, and inasmuch as we anticipate that the mass de-biasing will mainly increase more the number of lighter planets, it was felt useful to compare our retrieved mass distribution results (after $m \sin i$ inversion) on RV discovered planets with the recent work of Mordasini (2018) containing results of population synthesis model. On Fig. 14 is represented the results contained in his fig. 10 on the mass distribution: black histogram, right scale. Actually, it is the fraction of stars which contains at least one planet in the considered mass bin. It is characterized (going up in mass) by a fast decreasing distribution, a break point at 30 $M_{\oplus}$, a gentle increase up to two 10^3 $M_{\oplus}$, and another fast decrease.

In order to compare our results to this model prediction, we had to screen out from our sample all stars which have two or more planets in the same mass bin, keeping only one couple star/planet per mass

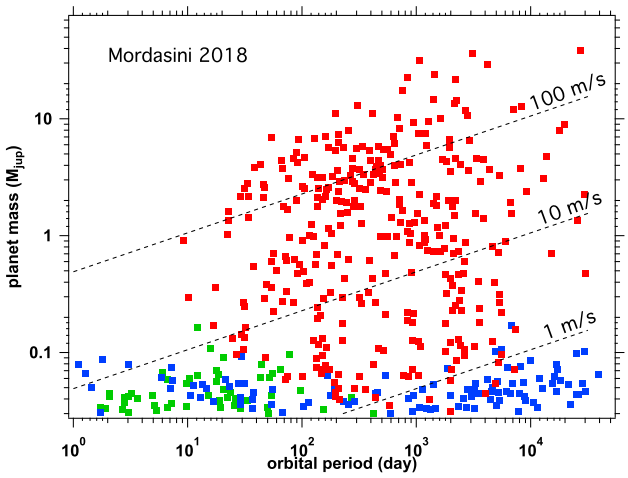

Figure 15. Result of one simulation of formation and evolution of planets (modeled distribution presented in fig. 8 of Mordasini 2018). Blue symbols are planets that have (partially) accreted volatile material (ices) outside of the iceline(s), while green symbols have only accreted refractory solids. The red points are giant planets with a mass of the gaseous envelope larger than the mass of the solid core. Only planets with mass >0.03 $M_{\rm jup}$ are kept from the original sample. The orbital period was computed from the semi-axis. The mass unit is $M_{\rm jup}$. The detection limits for RV surveys at 1, 10, 100 m s^{-1} are indicated.

bin. Doing so, our sample of planets was reduced to 800 planets. As we had done before with the original sample (Fig. 7), we have inverted the distribution of $m \sin i$ with our method, and we plotted the result also on Fig. 14, histogram and points with their error bars of the 'true' mass distribution. The mass bin grid is the same as before, but expressed in units of $M_\oplus$. The bin limits of the synthetic population are similar (but not fully identical) as our grid. We have adjusted the left scale (number of planets in the bin, only one couple star/planet per mass bin) to the right scale: a normalization in the mass bin from 16 to 32 $M_\oplus$.

We see substantial differences between the model and the observation of the mass distribution:

(1) For low mass planets, the observations show a moderate deficit w.r.t. the model in the bin 8–16 $M_\oplus$, strongly increasing with smaller masses. This deficit is of course the result of the lower sensitivity of RV methods to lower masses. Basically, the RV sensitivity threshold is of the order of 1 m s^{-1}.

(2) The bin 32–64 $M_\oplus$ contains much less stars/planets than the model: the drop of number of stars/planets is more severe in the observations than in the model, when progressing in mass from the normalization bin 16–32 $M_\oplus$. This is the so-called sub-Saturn desert.

(3) For all masses above ∼64 $M_\oplus$ there are many more observed planets than predicted by the model.

While the first difference is clearly due to an artefact of the RV method, it is our contention that the two other features (sub-Saturn desert and strong peak of giant planets) cannot be due to an artefact of the RV method. We have not been able to imagine a mechanism by which the RV method would be biased non-monotonously as a function of planet mass, or as a function of period.

The data-model comparison can be refined by looking at the two-parameters period-mass diagram (Fig. 9), or a-m (a is the semimajor axis) diagram of Mordasini (2018)). In order to ease the comparison, the original population model (fig. 8 of Mordasini 2018) has been redrawn (Fig. 15). The distribution has been truncated by keeping planets with mass ≥ 0.03 $M_{\rm jup}$; the period was computed from the semi-axis and third Kepler law around a star with 1 solar mass (which is assumed in the population model). The unit of mass has been taken as $M_{\rm jup}$ to be consistent with Fig. 9.

There are three main differences that can be readily identified 'by eye' between data and model distributions in the mass/period diagrams, by comparing Figs 9 and 15,

(1) the hot Jupiters cluster, conspicuous in the data, is absent in the model.

(2) the giant planet cluster is much less obvious in the model than in the data.

(3) the sub-Saturn desert identified in the data (in particular for periods ≤ 100 d) is weaker in the model, but not absent (better seen on Fig. 14).

We are aware of some different conditions between data and model which might affect the comparison. The RV sample is represented with the apparent mass $m \sin i$, while the model gives the true mass. There is also a bias factor of the RV method, favouring detections of larger masses and shorter periods, affecting the data but not the model.

In spite of these different conditions and biases, we estimate that the discrepancies between data and model most likely cannot be accounted on these differences/biases and are robust conclusions than can be derived from the comparison. For instance, the cluster of giant planets is waning for shorter periods at constant mass while the RV sensitivity is increasing with shorter periods; the cluster of hot Jupiters is waning for larger masses at constant period while the RV sensitivity is increasing with larger masses; the sub-Saturn desert cannot be produced, at constant period, by a mass bias. And we have shown that accounting for the $m \sin i$ distribution through our method of inversion actually increase the depth of the trough.

As a result, it is now easier to interpret the data-model discrepancies in the planet mass distribution of Fig. 14. The sub-Saturn desert is deeper in the observation, because of two reasons: (i) the number of observed planets are scarcer than in the model in the high-mass tail of the distribution of light planets (32–64 $M_\oplus$), at least for periods <100 d; (ii) for periods >100 d, in spite of the fact that the observed giant planets are more numerous than in the model, they have a strongly peaked mass distribution at 1.6–3.2 $M_{\rm jup}$ which falls off rapidly with smaller mass, that would otherwise fill somewhat the sub-Saturn desert. In that sense, it is also somewhat legitimate to describe this fast fall-off as a sub-Jupiter desert.

However, we consider that the question of the prolongation of the sub-Saturn desert for periods longer than 100–1000 d is still open. It is related to the mass distribution above super Earths. Sumi et al. (2010) analysed the sample of two super Earths, three cold Neptunes, one at ∼90 $M_\oplus$, and four Jupiters discovered at long periods by the micro-lensing technique. With a sophisticated de-biasing exercise, they estimated that existing cold Neptunes are three times more abundant than Jupiters. This estimate is at variance with the observed statistics of RV planets. We have plotted the mass distribution of RV planets with periods > 1000 d on Fig. S3 of the Supporting material (on line), which shows that there are 38 planets discovered in the range 0.8–1.6 $M_{\rm jup}$, only three in the range 16–32 $M_\oplus$, and one in the range 8–16 $M_\oplus$. For a fair comparison these numbers should be de-biased from known RV biases and inverted from $m \sin i$, which is beyond the scope of the present paper. Therefore we just mention that the micro-lensing estimate (based on a very small sample) is in strong disagreement with the mass distribution model of Mordasini (2018) (Fig. 14).

6 CONCLUSIONS

Based on a particular geometrical representation of exoplanet masses, we have designed a new numerical method (an algorithm) which allows to find the distribution $f(m)$ of true planet masses that will fit exactly an observed distribution $f_0(m \sin i)$ of apparent masses $m \sin i$ determined from the RV method. Of course, this is only a statistical method, not allowing to determine the true mass of each observed planets.

On the methodological side, Lucy (1974) has investigated the problem of $V_{\rm rot} \sin i$ (the rotation of stars) which is formally identical to exoplanet problem. He first identified a class of statistical methods, which consists of discretizing the solution of the integral equation (12) resulting in a linear system of equations. Clearly our method belongs to this class of methods, with some peculiarities: we have introduced an intermediate variable $\rho(m')$, the 'number density' of planets with $m \leq m' \leq m + dm$.

Lucy (1974) preferred an iterative method that he successfully applied to a histogram of 251 stellar rotations. Though the proposed iterative method is attractive and would deserve to be tested, it is much beyond the scope of our present work. We also note a big difference between the planet problem and the stellar rotation case which might render the iterative solution less attractive. The difference is that the dynamic range of the masses of exoplanets is huge in a linear histogram (our Fig. 6), while the stellar rotation distribution is much more compact (a factor of 4 between most and least populated bins, Lucy's fig. 5). This is certainly linked to the great difference of physical processes involved: continuous slowing of the rotation of the stars on one hand, agglomeration/accretion of dispersed grains/planetesimals and gas on the other hand. One issue with the iterative method is to decide, subjectively, when to stop the iterations: going too far will take noise for real signal. Also, Lucy's critic of the statistical methods seem a little outdated by 2021: 'Their wide application is, however, restricted by their demands on computer time and programming skill'. In the atmospheric community, this method is now classical with the onion-peeling technique, and our algorithm is directly inspired from it.

Our algorithm which is fully versatile in terms of mass bin sizes has been applied to three different grids of masses and a RV sample of 909 planets. The main result of our study is to show that the so-called sub-Saturn desert is confirmed at least for periods shorter than 100 d, with a depletion of planets in the true mass range 0.1–0.2 $M_{\rm jup}$ ($\sim$32 to 64 $M_\oplus$). This is not due to an artefact that would be induced by the fact that we measure only the apparent masses instead of the true masses. On the contrary, the depletion of planets in this mass range is even larger in the true mass distribution than in the apparent mass distribution. Taking as a reference the peak of large masses in the bin 1.66–3.22 $M_{\rm jup}$ (for all periods), the ratio of number of planets in the high mass peak bin to the bin trough (0.1−0.2 $M_{\rm jup}$) increases from 6.21 in the observed $m \sin i$ distribution to 10.7 in the true mass distribution, a factor 1.72 of increase. This is a robust result that most likely will not change much with more sensitive RV instruments or longer surveys. Indeed we have argued that the mass-dependent bias factor is larger for lighter planets, and accounting for it will increase the depth of the sub-Saturn at bin 32–64 $M_\oplus$ w.r.t. the 16–32 $M_\oplus$ bin (Fig. 12). We note that the sub-Saturn desert is more obvious for planets with periods below $\sim$100 d than for longer periods (Fig. 12. It still exists, though shallow, for periods between 100 and 1000 d; Fig. S3 of Supporting information).

We have also applied the algorithm to planets with apparent masses lighter than $\sim$200 $M_\oplus$. We found a gap of planets in the observed $m \sin i$ distribution in the narrow range of 13.7–15.2 $M_\oplus$ which contains Uranus (14.5 $M_\oplus$) and seems to be statistically significant. This Uranus gap contains only three planets in the observed histogram of apparent masses, while adjacent bins contain 13 and 22 planets. After inversion to get the true $f(m)$ mass distribution, the same bin is fully depleted of planets. While the actual existence of such a gap is still putative, it is argued that the very presence of Uranus in this gap could be due to a gain of $\sim$2 $M_\oplus$ during an collision, which is generally invoked to explain the 98° inclination of Uranus spin axis.

There are other observational biases than the $m \sin i$ which is addressed with our algorithm, e.g. the sensitivity limit of RV measurements, and the limited time span of observations, losing planets with long periods. In spite of this limitation, we compared our inverted mass distribution with the synthetic population model of Mordasini (2018)). We found two main differences: the observed sub-Saturn desert is deeper than in the model, and the observed peak of giant planets is larger and more peaked than in the model. We think that this result is robust, and would be confirmed by a more complete analysis, accounting for mass biases, $m \sin i$ bias, and period biases. We note that this $m \sin i$ inversion algorithm should be used in conjunction with other de-biasing methods, in order to better constrain the synthetic population models of planet formation and migration.

Archimed (287–212 BC) lived in Syracuse, Sicily. He worked a lot on spheres and cylinders, in order to find the volume of a sphere. He expressed the wish that his tomb should be engraved with a figure showing a sphere and a cylinder. Our algorithm using spheres and cylinders could be called by his name in recognition of his passionated work.

ACKNOWLEDGEMENTS

This work was partially supported by A. Ivanova Vernadski Scholarship for PhD students, sponsored by the French Government and the Russian Ministry of Science and Higher Education. We wish to thank Karim Gizatullin for the checking of the mathematical method while he was in IKI, and Jean-François Mariscal for the representation of the volumes 'bracelets' and 'napkins', resulting from the intersection of spheres and cylinders. We wish to thank the anonymous referee for very useful suggestions yielding significants improvements in the manuscript (twice). This research has made use of data obtained from the portal exoplanet.eu of The Extrasolar Planets Encyclopaedia, and from the NASA Exoplanet Archive via https://exoplanetarchive.ipac.caltech.edu.

DATA AVAILABILITY

The primary source of RV planets comes from portal exoplanet.eu of The Extrasolar Planets Encyclopaedia. We have made some screening out for planets which were discovered in transit and later on measured with the RV method. The exact list of planets used in the various applications of our algorithm may be obtained from the authors upon request. The mass and radius of the 570 planets were acquired from the NASA Exoplanet Archive via https://exoplanetarchive.ipac.caltech.edu. The exact list of planets used in Figs 12 and 13 may be obtained from the authors upon request.

REFERENCES

Ananyeva V., Ivanova A., Venkstern A., Shashkova I., Yudaev A., Tavrov A., Korablev O., Bertaux J.-L., 2020, Icarus, 346, 113773

Arriagada P., Anglada-Escudé G., Butler R. P., Crane J. D., Shectman S. A., Thompson I., Wende S., Minniti D., 2013, ApJ, 771, 42
Benítez-Llambay P., Masset F., Beaugé C., 2011, A&A, 528, A2
Bennett D. P., Ranc C., Fernandes R. B., 2021, AJ, 162, 243
Bertaux J.-L. et al., 2010, ACP, 10, 12091
Chandrasekhar S., Münch G., 1950, ApJ, 111, 142
Connes P., 1985, Astrophys. Space Sci., 110, 211
Fortney J. J., Marley M. S., Barnes J. W., 2007a, ApJ, 659, 1661
Fortney J. J., Marley M. S., Barnes J. W., 2007b, ApJ, 668, 1267
Fulton B. J., Petigura E. A., 2018, AJ, 156, 264
Howard A., 2013, Science, 340, 572
Howard A. W. et al., 2010, Science, 330, 653
Ida S., Lin D., 2004, ApJ, 604, 388
Jorissen A., Mayor M., Udry S., 2001, A&A, 379, 992
Kegerreis J. A. et al., 2018, ApJ, 861, 52
Lopez S., Jenkins J. S., 2012, ApJ, 756, 177
Lovis C., Fischer D., 2010, in Seager S., Tucson A. Z., eds, Exoplanets. University of Arizona Press, p. 526
Lucy L., 1974, AJ, 79, 745
Mayor M., Queloz D., 1995, Nature, 378, 355
Mayor M. et al., 2011, arXiv Astro-ph, 1
Mazeh T., Holczer T., Faigler S., 2016, A&A, 589, A75
Morbidelli A., Tsiganis K., Batygin K., Crida A., Gomes R., 2012, Icarus, 219, 737
Mordasini C., 2018, in Deeg H. J., Belmonte J. A., eds, Handbook of Exoplanets. Springer International Publishing, AG, New-york, p. 143
Mordasini C., Alibert Y., Benz W., Naef D., 2009, A&A, 501, 1161
Petigura E. A., 2020, AJ, 160, 89
Quémerais E., Bertaux J.-L., Korablev O., Dimarellis E., Cot C., Sandel B. R., Fussen D., 2006, J.Geophys.Res., 111, E09S04
Safronov V. S., 1966, AJ, 9, 987
Sumi T. et al., 2010, ApJ, 710, 1641
van Eylen V., Agentoft C., Lundkvist M. S., Kjeldsen H., Owen J. E., Fulton B. J., Petigura E., Snellen I., 2018, MNRAS, 479, 4786
Wikipedia, Abel Transform, 2021, https://en.wikipedia.org/wiki/Abel_transform
Yakovlev O. Y., Ananyeva V. I., Ivanova A. E., Tavrov A. V., 2022, MNRAS, 509, L17

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Fig. S1. Comparison of two apparent mass histograms of RV detected exoplanets (m sin i) with orbital periods < 50 d.
Fig. S2. (Top): Radius and mass of 175 planets as shown on fig. 3 of Howard (2013).
Fig. S3. This figure is the same as fig. 12 of the main text, on which are added the two m sin i distributions for the period ranges from 14 to 100 d (red open circles) and period>1000 d (open circles, magenta).
Fig. S4. Distribution of planet radii observed in transit, showing the radius gap around 1.8–2.0 $R_\oplus$.
Fig. S5. (Fig. 3 of Mordasini et al. 2009b). Quote : 'Planetary initial mass function, corresponding to the moment in time when the gaseous protoplanetary disc disappears.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: COMPUTATION OF THE VOLUME MATRIX

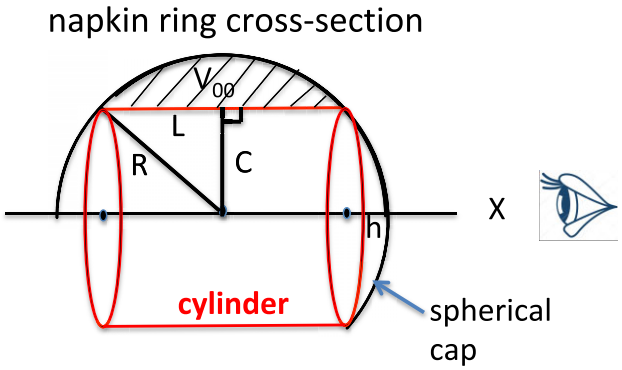

Figure A1. A napkin ring is defined as the volume of a sphere (radius R) which is outside of a cylinder (radius $C < R$) of infinite length. It is generated by revolution of the hatched area about the x-axis. It is equal to the volume of the sphere minus the red portion of the cylinder and two spherical caps, of height h and one on each side of the cylinder.

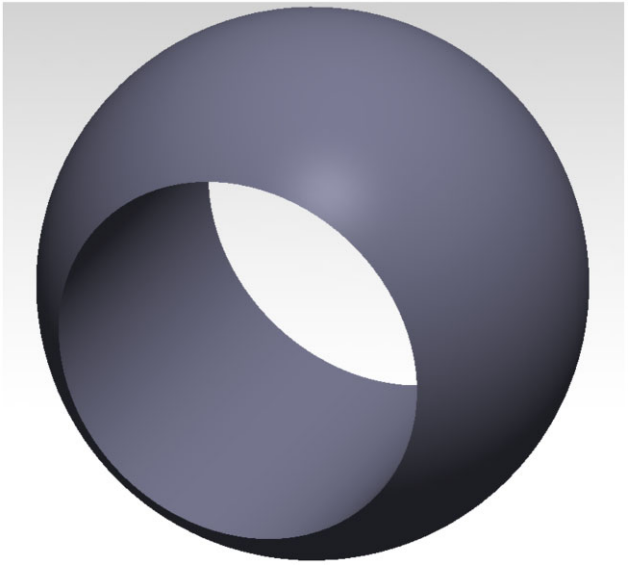

Figure A2. Perspective view of a napkin ring as defined in Fig. A1, the volume of a sphere (radius R) which is outside of a cylinder (radius $C > R$) of infinite length.

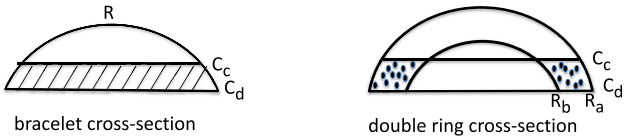

Figure A3. Left: A bracelet is defined as the difference of two napkins with same sphere radius R and between two cylinders C_c and C_d. Right: a (double) ring is defined as the difference of two bracelets.

A1 The grid of spheres and cylinders

We start from an observed distribution of apparent masses $p = m_{\rm true} \sin i$. The inversion problem (or retrieval problem) is discretized by deciding to cut the distribution to a maximum mass $m_{\rm max}$, excluding the observations with $p > m_{\rm max}$. A grid of true masses is chosen, defined by the limits of mass bins as a series of radii of concentric spheres, with $R_0 = m_{\rm max}$, a set of decreasing radii R_k

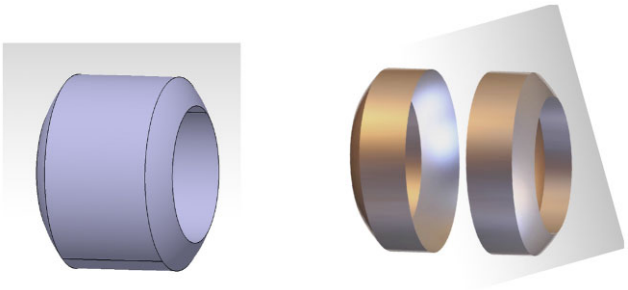

Figure A4. Left: Perspective view of a bracelet, a volume inside a sphere and between two cylinders as defined in Fig. A3. Right: Perspective view of a double ring as defined in Fig. A3. It is a volume between two spheres and two cylinders.

with the last one being $R_{\rm kmax} = 0$ or a minimum value $R_{\rm kmax} = R_{\rm min}$ (here there is a notable difference with the usual atmospheric onion-peeling method, with the grid of radii being of course all much greater than the radius of the Earth). An identical series of radii of cylinders C_i is also defined. The grid of spheres and cylinders radius may be chosen *ad libitum*: equally spaced or not, exponential; or composite. An element of matrix M is therefore the volume which is common to a spherical shell (defined by two radii R_a and R_b) and to a cylindrical shell (defined by two cylinders of infinite length and radii C_c and C_d). Such a volume is a double ring, as sketched on Figs A3 left and A4 left, and may be called $V_{\rm ring}(R_a, R_b, C_c, C_d)$. Its mathematical expression may be found in three steps, as described below.

A2 The volume of a napkin

A napkin volume is defined as the volume of the part of a sphere of radius R which is outside of a cylinder of radius C and infinite length. The name comes from the shape of a volume generated by rotation around X-axis of a portion of circle limited by a chord of length 2L, with $L^2 = R^2 - C^2$ (Fig. A1).

The volume of the sphere which is inside the cylinder is composed of one cylinder, and two spherical caps, one at each end of the cylinder.

Volume of sphere V_R : $V_R = 4/3\pi R^3$.

Volume of cylinder V_C, radius C, length 2L,

$$V_C = 2\pi C^2 L. \tag{A1}$$

Volume of one spherical cap of height h = R − L,

$$V_{cap} = \pi h^2 (R - h/3). \tag{A2}$$

Volume of napkin,

$$V_{\rm nap} = V_R - V_C - 2V_{cap} = 4/3\pi R^3 - 2\pi C^2 L - 2\pi (R-L)^2 (R - (R-L)/3). \tag{A3}$$

If $C = R$, $V_{nap} = 0$. By convention, we state that if $C > R$, $V_{\rm nap} = 0$.

A3 The volume of a bracelet

We define a bracelet as the portion of space inside a sphere (radius R) and comprised between two cylinders of radii C_1 and C_2 ($C_1 > C_2$). We see (Fig. A3) that the volume of a bracelet is the difference of the volumes of two napkins,

$$V_{\rm brac}(R, C_c, C_d) = V_{\rm nap}(R, C_d) - V_{\rm nap}(R, C_c). \tag{A4}$$

Table A1. Elements of the volume matrix for a simple example with three equal mass bins : 1, 2, 3.

Column 1	Column 2	Column 3
46.8321	0	0
26.1838	21.7656	0
6.5711	7.55594	4.18879

A4 The volume of a ring

We define the volume of a ring as the portion of space bounded by two spheres and two cylinders. In fact, it is a double ring of two equal volumes (Figs A3 and A4), one being in front of the projection plane, the other behind the projection plane. The volume of a (double) ring is the difference of the volumes of two bracelets

$$V_{\rm ring}(RA, RB, C_c, C_d) = V_{\rm brac}(RA, C_c, C_d) - V_{\rm brac}(RB, C_c, C_d). \tag{A5}$$

When RA and RB are contiguous in the grid of R_k, like R_k and R_{k+1}, and C_c and C_d are also contiguous (C_i and C_{i+1}), the (double) ring contain all planets which have the same true mass defined by the bin (R_k, R_{k+1}), and have been observed with an apparent mass $m \sin i$ defined by the bin (C_i, C_{i+1}). Therefore, the element V_{ik} on line i and column k of the matrix M_V relating the number densities of true mass planets ρ_i and the distribution of apparent masses N_k (equation 12) is

$$V_{ik} = V_{\rm ring}(R_k, R_{k+1}, C_i, C_{i+1}) \tag{A6}$$

which may be computed form the expressions of $V_{\rm brac}$ (A4) and $V_{\rm nap}$ (A3).

As an example, we give in Table A1 the elements of the matrix of dimension 3 when $R_0 = 3$, $R_1 = 2$, $R_2 = 3$, $R_4 = 0$, and identical values for the cylinders: $C_0 = 3$, $C_1 = 2$, $C_2 = 3$, $C_4 = 0$. Only the three true masses $2 < m_0 \leq 3$, $1 < m_1 \leq 2$, $0 < m_2 \leq 1$ (arbitrary units) are considered.

APPENDIX B: ABEL INVERSION AND THE ONION PEELING ALGORITHM

B1 The Abel transform and its formal inversion

We shortly describe the onion-peeling method, heavily used in space atmospheric physics, in order to retrieve the vertical profile $f(r)$ of the concentration [number of molecules $\rm cm^{-3}$, or density] of one atmospheric constituent from a series of measurements $F(y)$ of the integrated density (or column density, or slant path density, all the same thing, in $\rm cm^{-2}$) acquired by an orbiting space platform. An example is the series of line densities of ozone measured with stellar occultations by GOMOS on board ENVISAT platform as described in Bertaux et al. (2010). It is assumed that the atmosphere has a spherical symmetry, the concentration depending only on altitude z, or radial distance r to the centre of the planet, $r = R_p + z$, where R_p is the radius of the planet. Here, y is the impact parameter of the LOS from a space platform, the distance of the LOS from centre of the planet.

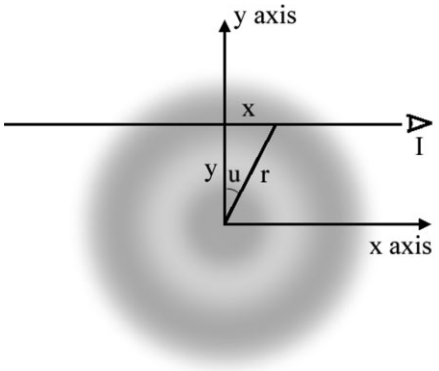

Figure B1. A geometrical interpretation of the Abel transform in two dimensions. An observer (I) looks along a line parallel to the x-axis at distance y above the origin. What the observer sees is the projection (i.e. the integral) of the circularly symmetric function $f(r)$ along the line of sight. The function $f(r)$ is represented in grey in this figure (taken from Wikipedia 2021). The observer is assumed to be located infinitely far from the origin so that the limits of integration are $\pm\infty$. The gaseous coma of a comet is an example of such a natural spherical distribution.

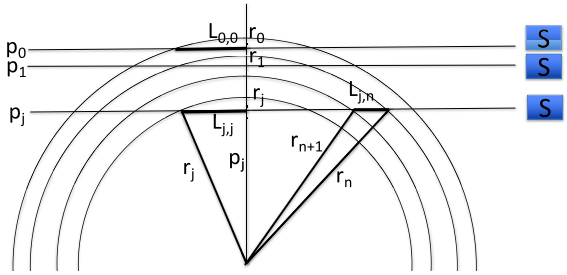

Figure B2. Geometry of LOS above the limb when an integrated density is obtained for each LOS (the case of GOMOS/ENVISAT, ozone, H_2O). The spacecraft S is drifting on its orbit, allowing a series of measurements with various impact parameters p_j of each LOS.

The column density $F(y)$ may be written as

$$F(y) = 2\int_y^\infty \frac{f(r)\, r\, \mathrm{d}r}{\sqrt{r^2 - y^2}}. \tag{B1}$$

This is the mathematical definition of the Abel transform, which may be represented as the sketch of Fig. B1

The formal solution of the inversion of Abel transform does exist,

$$f(r) = -\frac{1}{\pi}\int_r^\infty \frac{\mathrm{d}F}{\mathrm{d}y}\frac{\mathrm{d}y}{\sqrt{y^2 - r^2}}. \tag{B2}$$

In principle, equation (B2) gives the answer to the atmospheric problem. Unfortunately, in practice it does not work so well, because there is a derivative dF/dy, which is usually affected by noise, and makes unpractical the direct use of B2 in most cases, the resulting profile being very noisy, with often negative values of the local density. Therefore, atmospherists have developed since a long time a numerical algorithm : the so-called onion peeling algorithm.

B2 The onion peeling numerical algorithm

B2.1 General

The key, as usual in inversion problems, is to describe the forward model mathematically. Here, the forward model is the expression of what should be observed, $F(y)$, if we have the distribution $f(r)$. The forward model is formula B1 above, which will be discretized. The atmosphere is considered as an onion, with concentric spherical layers. The concentration is assumed to be constant into each layer. It is assumed 0 outside of the most external layer.

Consider a series of m measurements along various LOS at tangent altitudes $z_0, z_1, z_2, \ldots z_j, \ldots z_{m-1}$ going from top to bottom (Fig. B2). It corresponds to impact parameters $p_0, p_1, p_2, \ldots p_j, \ldots p_{m-1}$ with $p_j = z_j + R_p$. We may define the radii of layers as follows (this is partially arbitrary):

$$\begin{aligned} r_0 &= p_0 + (p_0 - p_1)/2 \\ r_1 &= (p_0 + p_1)/2 \\ r_2 &= (p_1 + p_2)/2 \\ &\ldots \\ r_n &= (p_{n-1} + p_n)/2 \\ r_m &= p_{m-1} - \text{something, and } r_m \text{ may be equal to 0, must be} < p_{m-1}. \end{aligned}$$

B2.2 Construction of lengths of segments of LOS through the layers

We have to compute the length of segments of LOS $L_{j,n}$ between two successive layers with the help of Pythagore theorem. The first index j indicates on which LOS is the segment and its impact parameter p_j. The second index n indicates that the segment is included between two spheres with radius r_n and r_{n+1}, $r_n > r_{n+1}$.

$L_{0,0}$ is the half-segment of the cord which encompass the tangent point of the LOS,

$$L_{0,0} = \sqrt{r_0^2 - p_0^2}$$

$$L_{1,0} = \sqrt{r_0^2 - p_1^2} - \sqrt{r_1^2 - p_1^2}$$

in general, the length $L_{j,n}$ of segment (j, n) is

$$L_{j,n} = \sqrt{r_n^2 - p_j^2} - \sqrt{r_{n+1}^2 - p_j^2} \tag{B3}$$

for $n = j$, there is a somewhat slightly different formula because the sphere of radius $n + 1$ is not involved,

$$L_{j,j} = \sqrt{r_j^2 - p_j^2} \tag{B4}$$

B2.3 Step-by-step retrieval of densities in each layer from resolution of a system of linear equations

The integrated densities $F(p)$ are expressed with the sum of products of lengths of cords by the constant density between two spherical shells, along the LOS with impact parameter p. We use indifferently z or $r = R_p + z$ and get a series of linear equations with unknowns $f(z_k)$ and data $F(p_j)$,

$$\begin{aligned} F(p_0) &= 2f(z_0)\, L_{0,0} \\ F(p_1) &= 2f(z_0)\, L_{1,0} + 2f(z_1)\, L_{1,1} \\ &\ldots \\ F(p_j) &= 2\sum_0^{j-1} f(z_k)\, L_{1,k} + 2f(z_j)\, L_{j,j} \end{aligned} \tag{B5}$$

where sum $\sum$ extends from $k = 0$ to $k = j - 1$, and the system contains m equations and m unknown $f(z_k)$.

One way to solve this system (B5) is to 'peel the onion' from outside to inside. From the first equation of system (B5) we extract the density of the upper layer $f(z_0)$,

$$f(z_0) = F(p_0)/2\, L_{0,0}. \tag{B6}$$

Then the value found for $f(z_0)$ is put in the second equation of system (B5), which allows to extract the value of $f(z_1)$, and successively from one layer to the next below, down to the last LOS and spherical shell.

We may see also the problem of the onion peeling as the resolution of a linear system of m equations with m unknowns, which may be written under a matrix formulation,

$$F = Mf. \tag{B7}$$

M is a triangular matrix of rank m. Its elements are the lengths of segments of LOS between two spherical shells where the density f is assumed constant, and F is the vector of the series of observations of integrated densities.

Of course in order to be resolved, one has to have at least as many values of p (number of measurements) as the number of layers. Otherwise, the number of layers, their exact limits w.r.t the values of p_i can be selected at will. If the number of layers is selected as smaller than the number of measurements, then there is a least-square solution to the problem, which makes use of the transpose matrix M^T and its product by M to get a square matrix.

Different errors in the data (measurements) may be accounted for. In order to avoid vertical oscillations in the retrieval which are not physical, one often use the scheme of Tikhonov regularization (Quémerais et al. 2006).

B3 Change of power-law exponent in the Abel transform

In Section 2.5 of the main text, we expressed the need to demonstrate that, if the true mass distribution is a power law with a certain value of the exponent (or index), then the apparent mass distribution will have show also a power law with the same index. The demonstration requires to explore what happens in the Abel transform of a power law.

So, we apply the Abel transform (equation B1) to the power law $f(r) = r^{-n}$. From Fig. B1 we derive the following relations, with x, y, r, and angle u are defined on Fig. B1.

$$\begin{aligned}
x^2 + y^2 &= r^2 \\
\tan u &= \frac{x}{y} \\
\cos u &= \frac{y}{r} \\
r &= \frac{y}{\cos u} \\
\mathrm{d}r &= y \tan u \frac{1}{\cos u}\mathrm{d}u \\
F(y) &= 2\int_y^\infty \frac{r\, r^{-n}}{\sqrt{(r^2 - y^2)}}\,\mathrm{d}r = 2\int_y^\infty \frac{r^{-n+1}}{\sqrt{(r^2 - y^2)}}\,\mathrm{d}r.
\end{aligned}$$

We replace r by $\frac{y}{\cos u}$ and change the limits of integration with corresponding values of angle u.

$$\begin{aligned}
F(y) &= 2\int_0^{\pi/2} \frac{\left(\frac{y}{\cos u}\right)^{-n+1} y \tan u \frac{1}{\cos u}}{\sqrt{y^2\left(\frac{1}{cos^2 u} - 1\right)}}\,\mathrm{d}u \\
&= 2\int_0^{\pi/2} \frac{\left(\frac{y}{\cos u}\right)^{-n+1} y \tan u \frac{1}{\cos u}}{y \tan u}\,\mathrm{d}u = 2\int_0^{\pi/2} \frac{y^{-n+1}}{\cos^{-n+2} u}\,\mathrm{d}u \\
F(y) &= 2y^{-n+1}\int_0^{\pi/2} \cos^{n-2} u\,\mathrm{d}u.
\end{aligned}$$

Since the integral has a definite value (a constant), the function $F(y)$ is indeed also a power law, with a power law $-n + 1$: the exponent has increased by $+1$ during the Abel transform.

For $n = 2 \Rightarrow F(y) = 2y^{-2+1}\int_0^{\pi/2} \cos^{2-2} u\,\mathrm{d}u = 2y^{-1}(\pi/2 - 0) = = \frac{\pi}{y}$.

This result is well known from cometary scientists. Assuming a constant production rate of a molecule from the nucleus, the cloud of this molecule expands in vacuum at a constant velocity. Therefore, the density in this coma scales as r^{-2} (equaling the flux through a sphere of radius r to the constant production rate), and integrated densities (observed from outside by instruments) scale as y^{-1} (y being the distance of LOS from comet nucleus).

Let us define the constant $W_e = \int_0^{\pi/2} \cos^e u\,\mathrm{d}u$ with $e = n - 2$ used in the expression of $F(y)$. We have already computed for $e = 0, n = 2, W_0 = \frac{\pi}{2}$.

For $e = 1,\ n = 3,\ W_1 = \int_0^{\pi/2} \cos^1 u\,\mathrm{d}u = 1.$

For other values of n and e, we may find a recurrence formula for the constant,

$$\begin{aligned}
W_e &= \int_0^{\pi/2} \cos^e u\,\mathrm{d}u = \int_0^{\pi/2} \cos^{e-1} u \cos u\,du \\
&= (\cos^{e-1} u \sin u)\Big|_{\pi/2}^{0} + +(e-1)\int_0^{\pi/2} \cos^{e-2} u \sin^2 u\,\mathrm{d}u \\
&= (e-1)\int_0^{\pi/2} \cos^{e-2} u\left(1 - \cos^2 u\right)\mathrm{d}u \\
&= (e-1)\left(\int_0^{\pi/2} \cos^{e-2} u\,\mathrm{d}u - \int_0^{\pi/2} \cos^e u\,\mathrm{d}u\right) \\
&= (e-1)(W_{e-2} - W_e) \\
eW_e &= (e-1)W_{e-2}.
\end{aligned}$$

This formula is linking separately the series of odd and even values of e. Therefore, we find that for $e = 2, n = 4, W_2 = \frac{\pi}{4}$ and so on and so forth.

APPENDIX C: SIMULATIONS OF SEVERAL MASS DISTRIBUTIONS

C1 Adding one bin of large masses to the distribution, and effect on the other bins

On Fig. 7 we have studied a histogram with progressive bins limits, with a factor of 2 on the limits of one bin to the next. This exponential histogram (Section 3.2) has 13 (apparent) mass bins, with the first bin (largest apparent masses) with limits between 12.89 and 25.77 $M_{\rm jup}$. We had excluded all 24 RV planets reported with an $m \sin i > 25.77 M_{\rm jup}$, reducing the original sample of 909 planets to 885 planets. What is the influence, on the true mass histogram retrieved by the algorithm, of excluding these high mass planets?

In order to answer, one bin was added between 25.77 and 51.54 $M_{\rm jup}$, containing 17 large mass planets (or rather BD) which are in the list of observed planets, but were excluded from the 13 bins histogram (Table C1, column $m \sin i$ (2)). The same inversion scheme was applied to this 14 bins histogram. The new bin 25.77–51.55 $M_{\rm jup}$ contains 22.9 planets, in the second bin (first bin of 13 bin histogram), the number of planets has decreased from 52.53 to 46.3, and all the other bins contain almost identical values of planet. Therefore, we may conclude that eliminating from our observed sample all

Table C1. Effect of excluding or including one bin with large masses on retrieved histogram. (1) excluding the largest mass bin. (2) including the largest mass bin.

Bin number	Bin lim inf (M_{jup})	Bin lim sup (M_{jup})	Observed planet number $m \sin i$ (1)	Observed planet number $m \sin i$(2)	Retrieved planet number m(1)	Retrieved planet number m(2)
1	25.77	51.55	–	17	–	22.90
2	12.88	25.77	39	39	52.53	46.29
3	6.44	12.88	73	73	84.02	84.43
4	3.22	6.44	101	101	110.19	110.12
5	1.61	3.22	149	149	165.2	165.2
6	0.805	1.61	116	116	103.6	103.6
7	0.402	0.805	80	80	68.3532	68.3533
8	0.201	0.402	50	50	40.16	40.16
9	0.101	0.201	24	24	15.43	15.43
10	0.0503	0.101	51	51	60.77	60.77
11	0.0251	0.0503	87	87	98.85	98.85
12	0.0125	0.0251	62	62	52.71	52.71
13	0.00629	0.0125	34	34	24.91	24.91
14	0.00314	0.00629	19	19	14.24	14.24

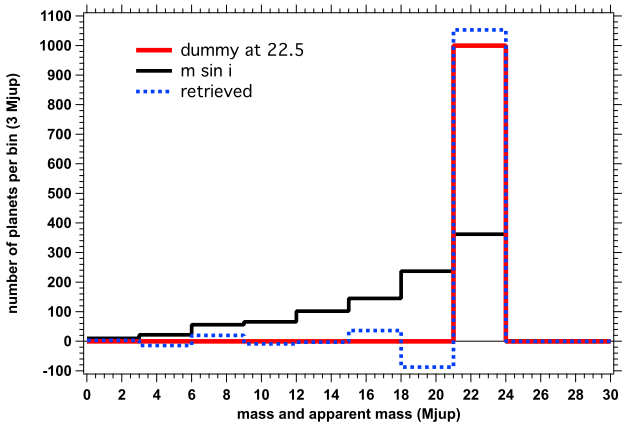

Figure C1. Simulation of the retrieval. A mono-mass ensemble of 1000 planets at 22.5 M_{jup} (red solid histogram, dummy) are assigned a random $\sin i$ value, yielding the $m \sin i$ histogram (black line). This histogram is inverted, yielding the retrieved distribution indicated as the blue dashed line. Both dummy and retrieved histograms mostly overlap.

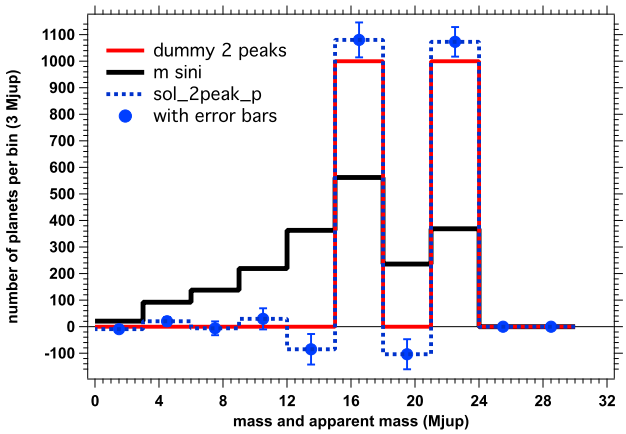

Figure C2. Simulation for a two peaks mass distribution. Red: Histogram of 2000 planets with two values at 16.5 and 22.5 M_{jup}. Black: Histogram of $m \sin i$ with $\sin i$ assigned at random. The retrieved distribution (after inversion with our algorithm) is indicated both by the blue dashed line and blue circles with the error bars. The two peaks are well retrieved. Both dummy and retrieved histograms mostly overlap.

planets above 25.77 M_{jup} does not affect at all the values retrieved below 12.89 M_{jup}. In principle, one definition of an exoplanet is the minimum mass required to get Deuterium burning, which is about 13 M_{jup}. This exercise shows also that it is necessary to keep the observed BD up to ∼26 M_{jup} in order to keep the exoplanet bin 6.44–12-88 M_{jup} unaffected by the algorithm of inversion (edge effect).

C2 Simulation of a single peak mass distribution

We start with a sample of 1000 planets, all having a mass of 22.5 M_{jup}, the centre of bin 21–24 (the unit is not important). Its histogram is represented by the red solid curve on Fig. C1. Each planet was then assigned a random value of $\sin i$ to build an apparent mass distribution, represented by the black dashed line on Fig. C1. Finally, our retrieval algorithm was applied to this simulated $m \sin i$ distribution, yielding the retrieved distribution, indicated by the blue dashed line. The bin 21–24 contains 1053.1 planets, instead of 1000 in the original sample. Below 21, the values in the bins are essentially near 0, with the exception of the bin 18–21 with a negative number of −87.3 planets.

C3 Two peaks simulation

In order to simulate a two peaks mass distribution, another mono-mass distribution of 1000 planets with mass 16.5 M_{jup} was added to the previous single peak distribution (red solid line on Fig. C2). After assignment of a random value of $\sin i$ to each planet, the histogram of the 2000 planets apparent masses is displayed as the black solid line. The retrieved distribution (after inversion with our algorithm) is indicated both by the blue dashed line and blue circles with the error bars. The two peaks are well retrieved. We note a small excess at the two peaks, and a small deficit (negative values) in the two bins just before the two peaks, as in the single peak simulation. This small bias seems systematic. Other bins have an error bar which encompasses zero. The two bins above the bin 21–24 are containing strictly 0 planet, which is an inherent feature of the method.

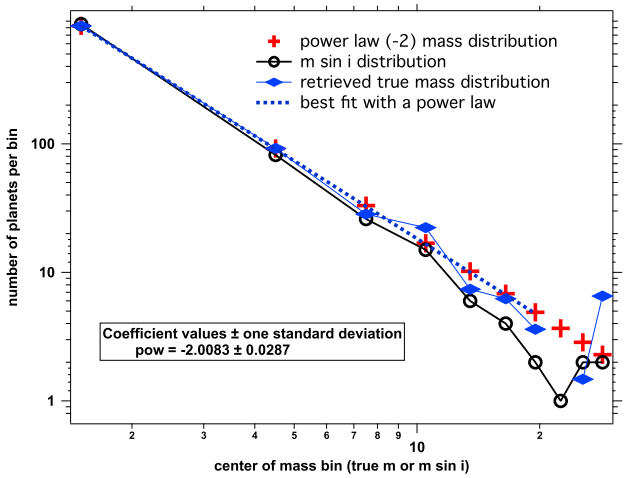

Figure C3. Red crosses: Power law distribution of true masses m, with power index -2. Black circles: Histogram of $m \sin i$, after assignment of a random $\sin i$ value. Blue lozenges: Retrieved true mass histogram after the inversion method applied to black circles distribution. Dashed dark blue line: best-fitting power law to lozenges with index -2.0083 ± 0.0287.

C4 Power law simulation

We have selected for this simulation a -2 power law distribution, with the number of planets per unit mass is $\mathrm{d}N/\mathrm{d}m = Am^{-2}$. We have used the mass bins as for the Fig. 6 inversion, 10 bins of 3 $M_{\rm jup}$ width.

We started from a simulated histogram containing 994 planets, represented by red crosses on Fig. C3. The histogram numbers of planets are not integer values, in order to respect strictly the power law. In each bin, all the stars in the bin are assigned an individual mass by spreading them regularly (linearly) within the limits of the bin. Then a random value of $\sin i$ was also assigned to each planet, by drawing from an isotropic distribution of polar axis. We obtained a dummy $m \sin i$ distribution, which histogram is displayed as black circles on Fig. C3. Note that the black circles are below the red crosses, except for the lightest planets (bin 0 to 3 $M_{\rm jup}$). This can be explained as follows. When all true masses are multiplied by a random value of $\sin i$, the apparent mass $m \sin i$ becomes smaller than m. We may say that the mass has shrinked. The average value of $\sin i$ is $\frac{\pi}{4} = 0.785$, the average shrinking factor. After multiplication by random $\sin i$, some planets are going to move from one bin to the next, containing lower mass planets. Except for the last bin (0–3 $M_{\rm jup}$), which number increases from 827 to 854, the other bins are losing planets, because more planets go to the lighter bin than planets which enter this bin from the heavier bin (we have assumed a negative power law). We note also that the low (integer) numbers for the large mass bins display some strong fluctuations.

Then we have applied our standard inversion scheme to find the 'true' mass distribution with its histogram represented by blue lozenges. One value at bin 21–24 $M_{\rm jup}$ was negative (–0.75) and is not represented on the log–log plot of Fig. C2. Discarding also the two other largest masses bins (24–30 $M_{\rm jup}$) which contain very low number of planets and therefore subject to large statistical fluctuations, we have computed the best-fitting power law of the seven other bins of lightest planets (0–21 $M_{\rm jup}$). We found -2.0083 ± 0.0287, encompassing in its error bar the original -2 value.

Therefore, we may say that our numerical scheme of inversion, operated on a power law of apparent mass distribution, retrieves the correct true mass power law.

This paper has been typeset from a TEX/LATEX file prepared by the author.