Identification of a 4.3 billion year old asteroid family and planetesimal population in the Inner Main Belt

S. Ferrone, M. Delbo, C. Avdellidou, R. Melikyan, A. Morbidelli, K. Walsh, R. Deienno

HAL Id: hal-04191828

https://hal.sorbonne-universite.fr/hal-04191828v1

Submitted on 26 Oct 2023

Astronomy
&
Astrophysics

Identification of a 4.3 billion year old asteroid family and planetesimal population in the Inner Main Belt

S. Ferrone[1,2] , M. Delbo[1] , C. Avdellidou[1] , R. Melikyan[1,3,5] , A. Morbidelli[1] , K. Walsh[4] , and R. Deienno[4]

[1] Observatoire de la Côte d'Azur, CNRS-Lagrange, Université Côte d'Azur, CS 34229, 06304 Nice Cedex 4, France
[2] LESIA & GEPI, Université Paris Cité, Observatoire de Paris, Université PSL, Sorbonne Université, CNRS, 92190 MEUDON, France
 e-mail: `salvatore.ferrone@obspm.fr`
[3] Lunar and Planetary Laboratory, University of Arizona,1629 E University Blvd, Tucson, AZ 85721, USA
[4] Southwest Research Institute, 1050 Walnut St. Suite 300, Boulder, CO, 80302, USA
[5] Department of Physics and Astronomy, Ithaca college, 953 Danby Road, Ithaca, NY 14850, USA

ABSTRACT

Context. Understanding the conditions that lead to the formation of planetesimals – the building blocks of planets – and their initial size distribution is a central problem of modern planetology. While most of these original planetesimals were accreted onto the terrestrial planets and the cores of the giant planets, some were also stranded in the main belt, where 4.5 Gyr of collisional evolution broke most of them into families of collisional asteroid fragments. However, some planetesimals survived, and are still hidden amongst asteroid fragments in the main belt.

Aims. We make use of astronomical data to identify these leftover planetesimals amongst all other asteroids. Our search is based on separating planetesimal survivors from families of asteroids generated by collisions. Namely, we aim to identify and "clean" the main belt of collisional family members: by doing so, we would be left with the surviving members of the original planetesimals.

Methods. We focus here on the inner portion of the main belt for asteroids with intermediate to high albedo. It is known that current asteroid family catalogs are not suitable for the aforementioned cleaning; they are conservative and only one-quarter of the known asteroids are associated with the approximately 120 distinct asteroid families. We therefore developed methods to inclusively link asteroids to known collisional families in order to better capture their extent. Namely, we apply a hierarchical clustering method (HCM) on asteroids filtered according to the V-shape of the Yarkovsky drift of each family in order to reassess family membership (V-shape-constrained HCM). The identified families were removed and the remaining background population was searched for previously undetected collisional families.

Results. We succeed in using our V-shape-constrained HCM to link family "halos" to their cores. After removing these reassessed families from the asteroid population, our V-shape search reveals a previously unknown collisional family of S-type asteroids in the inner main belt with an age of 4.3 ± 1.7 Gyr and a significance level of 3.4σ. When this ancient collisional family is removed, 34 planetesimals are identified and their size–frequency distribution is presented.

Conclusions. The asteroid belt has two components: planetesimals and collisional fragments. The cumulative size–frequency distribution of planetesimals has a steep power-law index for bodies larger than 100 km in diameter and a much smaller power-law index for planetesimals smaller than 100 km.

Key words. minor planets, asteroids: general – planets and satellites: dynamical evolution and stability

1. Introduction

Understanding the formation of planets and small bodies is a central problem of planetary science. It is recognized that the first stage of this process is the accretion of the so-called planetesimals from the solids in protoplanetary disks (e.g., Birnstiel et al. 2016, and references therein). Studies of the formation of planetesimals have made tremendous progress in the last few decades, and the consensus is that these objects formed from clumps of solid particles that reached sufficient densities to become self-gravitating and to contract to form asteroid-sized bodies (Johansen et al. 2007), a process also favored by disk turbulence (Johansen et al. 2007; Cuzzi et al. 2008) and the Kelvin-Helmholtz instability (Johansen et al. 2007). However, most studies approach the problem of planetesimal formation theoretically, as only limited constraints are available on the original planetesimal size and composition distribution in our Solar System (e.g., Klahr et al. 2022, and references therein).

In order to obtain further such constraints, significant effort has been made to investigate the population of the minor bodies of our Solar System (Bottke et al. 2005; Morbidelli et al. 2009; Delbo et al. 2017; Tsirvoulis et al. 2018). Indeed, asteroids and all the small bodies, such as comets, trans-Neptunian objects, and irregular satellites, are what is left of the original planetesimal disk from the planet-formation era. However, not all asteroids are survivors from primordial times: a large number of these are collisional fragments of the original planetesimals (Delbo et al. 2017; Tsirvoulis et al. 2018; Dermott et al. 2018). Although these fragments still carry the original composition of their progenitors, their sizes, spin vector orientations, and shapes, they do not provide direct information about the accretion processes that led to the formation of planetesimals and, consequently, of the planets.

With this in mind, different studies have attempted to separate the original asteroids that accreted as planetesimals in the protoplanetary disk from the collisional fragments
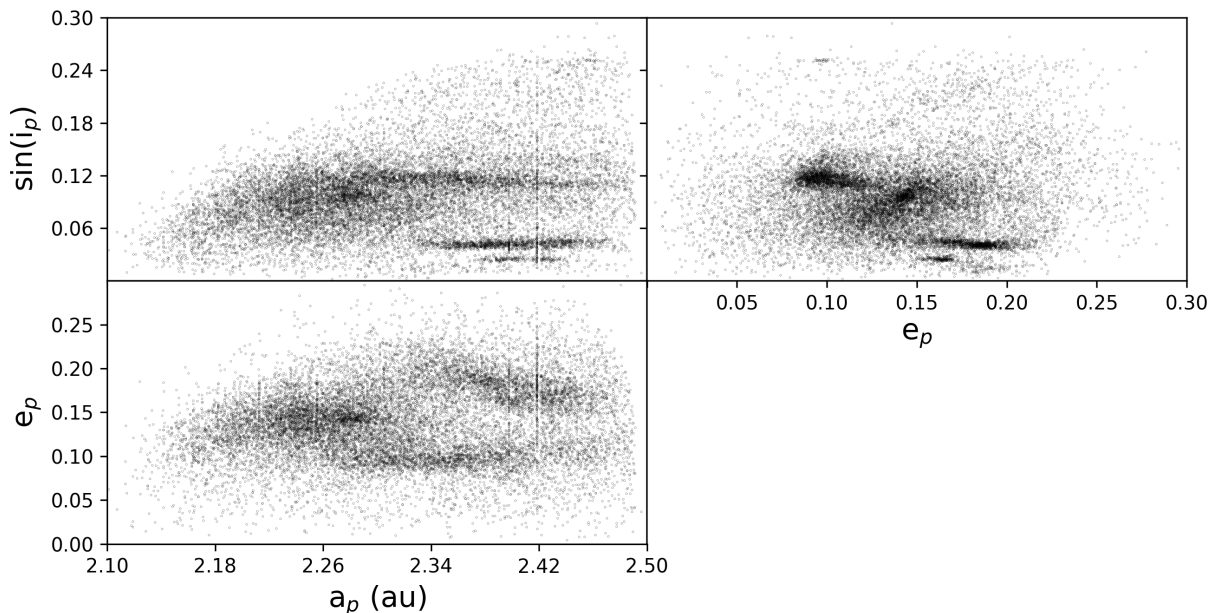
Fig. 1. Proper orbital elements of the inner main asteroid belt. Only bodies with p_V greater than 0.123 are shown. The y-axis is shared for the top two panels, while the x-axis is shared for the left two panels. Clusters of objects correspond to some of the most preeminent asteroid families.

(Delbo et al. 2017, 2019; Tsirvoulis et al. 2018; Bourdelle-de Micas et al. 2022). These works are based on the concept of identifying all families of fragment asteroids that were produced in the main belt after the early formation phases of our Solar System and "cleaning" them from the main belt in order to highlight the survivors of the original asteroids, that is, the planetesimals.

The question remains as to how we can best distinguish asteroid fragments from planetesimals. In a break up process, fragments are launched into space at moderate velocities (some m s^{-1}), which is how, in the main belt, fragments have similar orbital elements to those of their parent body (e.g., Vokrouhlický et al. 2006, and references therein). Thus, fragments become new asteroids themselves, clustered in orbital space, as shown for example in Fig. 1. Additionally, asteroids inside these families have, in general, similar physical properties, such as geometric visible albedo (p_V), (Masiero et al. 2015), color, and spectrum (Parker et al. 2008; Galluccio et al. 2023). These clusters of fragments are the so-called asteroid families and are typically identified using the yierarchical clustering method (HCM). This method looks for clusters of asteroids in the orbital element space of semi-major axis, eccentricity, and inclination (a_p, e_p, i_p; Zappala et al. 1990; Nesvornỳ et al. 2015). The resulting current catalogs of asteroid family membership contain more than 140 000 asteroids belonging to 119 separate families (Nesvornỳ et al. 2015).

However, there are several reasons why current asteroid family catalogs are not suitable for the aforementioned cleaning of the main belt. Firstly, the HCM asteroid families are in general conservative. To demonstrate this, Fig. 2 shows the so-called background asteroid population after removing all family members identified by Nesvornỳ et al. (2015). Immediately, some structures related to families are still visible, with the most obvious being the "halos" in the e_p, sin(i_p) plane. Family halos have also been identified by other works around several families (Parker et al. 2008; Brož & Morbidelli 2013). As asteroids within the halos tend to have the same colors as the family that they surround (Parker et al. 2008), it appears logical to infer that they are constituted by asteroid fragments from the family parent body.

The fact that they had not been linked to the core of the family by the HCM is a result of the fact that HCM-driven asteroid linking is kept conservative. This conservatism is maintained in order to have good separation between adjacent families and to limit family contamination from the so-called interlopers (false positives). This is desirable when family membership information is to be used by authors interested in modeling the family dynamics or studying their composition. However, by keeping the HCM clustering conservative, many family members are missed (e.g., Parker et al. 2008; Brož & Morbidelli 2013). The second reason is that the HCM family catalogs are likely missing families with ages roughly older than 2 Gyr (Brož et al. 2013; Spoto et al. 2015; Bolin et al. 2017). This deficit can be seen by comparing the number of identified families as a function of their age. Although a roughly constant asteroid break-up rate is expected over the last ∼3 Gyr, numerical models of Solar System evolution from Bottke et al. (2005) suggest an even higher original collision rate that decreased in time because of dynamical depletion. Such results raise doubt as to the fidelity of the low number of detected ancient families. One may hypothesize that this deficit of known old families is due to the efficiency of the HCM, which decreases with increasing age of the collisional family to identify.

This latter effect is a result of the fact that families disperse over time. A nongravitational effect – the Yarkovsky effect – slowly changes the orbital semi-major axis a_p of asteroids at a rate $\mathrm{d}a_p/\mathrm{d}t$ proportional to $1/D$ (Vokrouhlicky et al. 2015). Asteroids in prograde rotation have $\mathrm{d}a_p/\mathrm{d}t > 0$ and migrate towards larger heliocentric distances, whereas those in retrograde rotation with $\mathrm{d}a_p/\mathrm{d}t < 0$ migrate towards the Sun. This creates correlations of points in the ($a_p, 1/D$) plane called V-shapes (as they resemble the letter "V"), whose slope (K) indicates family age (Vokrouhlický et al. 2006; Spoto et al. 2015). From modeled or measured $\mathrm{d}a_p/\mathrm{d}t$-values, family age can be determined with some uncertainty (Vokrouhlický et al. 2006; Spoto et al. 2015). Asteroids, as they move, encounter orbital resonances with the planets, which change their orbital e_p and i_p, but not their a_p. Thus, families become harder to identify as clusters of points in (a_p, e_p, i_p) space as they age because they
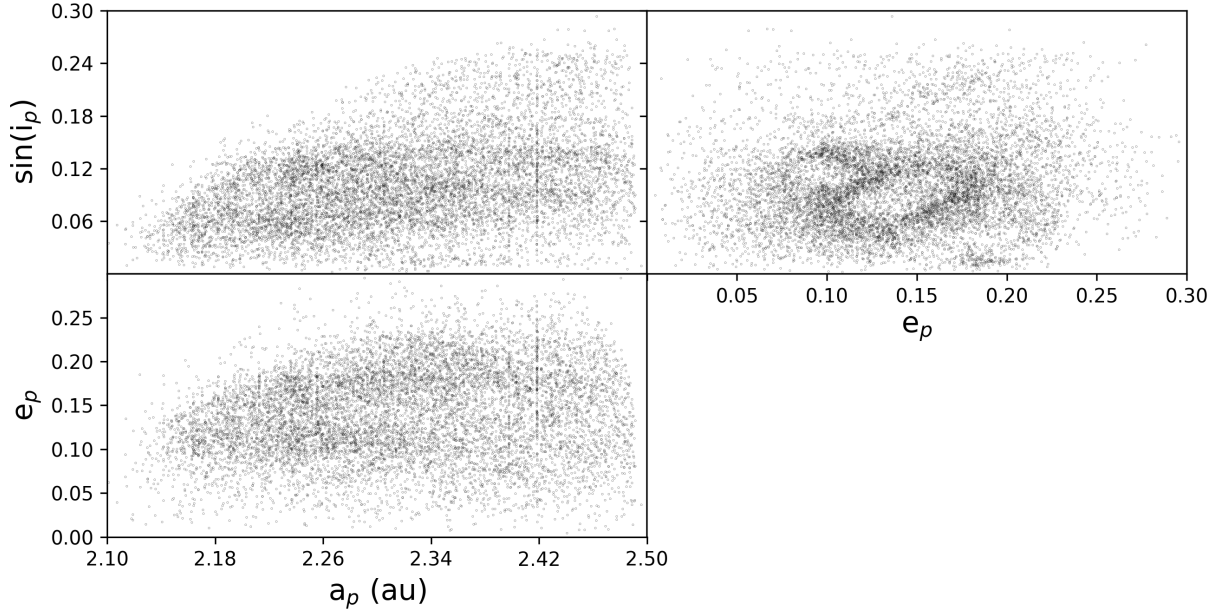
Fig. 2. Same as Fig. 1 but after removing the families identified by Nesvornỳ et al. (2015). The halos are best seen in the $\sin(i_p)$ versus e_p space.

are increasingly dispersed (Delbo et al. 2015) and overlap with each other.

However, Bolin et al. (2017) and Delbo et al. (2017) developed a method – based on the initial work of Walsh et al. (2013) – that makes use of the V-shape to identify families. Deienno et al. (2021) studied the efficiency of this method as a function of the age of the family, its position in the main belt, and its contrast with respect to the local background. These latter authors concluded that the V-shape family identification method is particularly useful for locating strongly dispersed collisional families (e_p, i_p), that is, those not trivially visible to the HCM. This is because while (e_p, i_p) of family members are dispersed by the effect of secular and mean motion resonances, these resonances do not affect the a_p and therefore the correlation between a_p and $1/D$ should stay visible for billions of years.

The efficacy of the V-shape method has been demonstrated with the identification of a family that could be as old as the Solar System (Delbo et al. 2017) and another that is roughly 3 Gyr old (Delbo et al. 2019). A third family, but with lower statistical significance, has also been found to be as old as our Solar System (Delbo et al. 2019). These families are quite vast in general, leading the authors to infer that most of the asteroids previously assigned to the background are instead family members. By studying the orbital properties of asteroids in the inner portion of the main belt, Dermott et al. (2018) came to the same conclusion; namely that most of the "background" asteroids are instead family members, either belonging to halos of already known families or those yet-to-be discovered.

Previous works identified families using the V-shape methods in the inner main belt for low- and medium-albedo bodies (Delbo et al. 2017, 2019). Here, we extend these works, focusing on those asteroids with intermediate and high albedo, that is, $p_V > 0.12$, of the inner portion of the main belt. This population of asteroids contains several <2 Gyr-old families already identified by means of the HCM (e.g., Nesvornỳ et al. 2015). Hence, a blind V-shape search of this population would mostly result in the detection of these known families. Therefore, here we proceed with a first additional step compared to previous V-shape searches for unknown families; namely, we reassess which

bodies are members of collisional families that are already cataloged (e.g., by Nesvornỳ et al. 2015), and we remove them to study the remaining population. We need to perform this family membership reassessment in order to link halos to the cores of the respective families. Subsequently, we search amongst the asteroids of the remaining population for the V-shapes that represent the oldest and undiscovered asteroid families.

Once a V-shape is identified, as following previous works (Delbo et al. 2017, 2019), we then consider all asteroids outside of these V-shapes to be unassociated to any asteroid family and catalog them as original planetesimals. This is because a fragment asteroid generated by a collision in the main belt is expected to be inside a V-shape. Therefore, asteroids outside of all V-shapes cannot be associated with collisional families and are most likely pristine primordial objects. Additionally, as noted in Delbo et al. (2017, 2019), most of the identified planetesimals are large and at the same time do not reside near other large objects in (a,1/D) space, meaning that there are no potential larger objects from which the candidate bodies may originate. This affirms that these objects are unlikely to be fragments of undiscovered families that remain below a detection limit. Regardless, if some of these objects are not pristine, the collisions that produced them must have happened before the Solar System reached its current configuration and some transport mechanism that brought them to the main belt must have erased their dynamical linkage to other sibling fragments. However, it is difficult to speculate how many bodies may be fragments from ancient collisions, because the collisional rate of small bodies in this epoch of the Solar System is not well constrained. For instance, the Bern III model from Emsenhuber et al. (2021) includes the late-stage planetary formation phase and has a mass resolution of 1% of the mass of Earth, which is still 100 times larger than the most massive asteroid, Ceres (Emsenhuber et al. 2021).

The paper is structured as follows: in Sect. 2, we describe the input datasets; in Sect. 3 we present the methods used for reassessing the known asteroid families and identifying undiscovered ancient asteroid families. In Sect. 4, we present the results from applying these methods to the data. In Sect. 5,

we discuss the implications of our results on planet-formation models and the limitations of our methods of planetesimal identification, and propose future studies.

2. Data

All data required for this work are described below, including: the proper orbital elements, physical information, previous family definitions, and simulated data. To begin, proper orbital elements approximate quasi-integrals of the full N-body equations of motion of asteroids; they are quasi-constant in time and are the basis for family identification. Proper elements can be used for asteroid clustering (see, e.g., Morbidelli 2002; Knežević 2017, for reviews) and also for the V-shape identification (see, e.g., Bolin et al. 2017). The physical information includes p_V, diameter, and bulk density, or p_V, D, and ρ, respectively. As p_V is indicative of physical composition, it is used to filter the dataset by helping to eliminate false positives in the family membership assignment. As we show, the mean bulk density of the asteroids comprising a given family is needed to compute the age of the family. The density contributes to the magnitude of acceleration of the Yarkovsky effect (Vokrouhlický et al. 2006; Spoto et al. 2015; Bolin et al. 2017), which depends on both the mass and volume (Bottke et al. 2006). Next, instead of searching for families with no prior information, we use their definitions from Nesvorný et al. (2015) to aid our search. Lastly, we use a synthetic dataset of the asteroid belt background population to establish a quasi-random noise, which defines the limits of a family while using the HCM (clustering method). In the following subsections, we review the data sources.

2.1. Orbital elements and physical properties

We retrieved the proper orbital elements and physical properties of asteroids from the Minor Planet Physical Properties Catalogue, MP[3]C[1], which is run and maintained by the Observatoire de la Côte d'Azur (Delbo et al. 2022). This database is a compilation of multiple sources. For instance, most diameters are from NEOWISE (Mainzer et al. 2019), WISE (Masiero et al. 2011), AKARI (Usui et al. 2011), IRAS (Tedesco et al. 2002; Ryan & Woodward 2010), and MSX (Ryan & Woodward 2010). Additionally, some diameters have been measured from stellar occultation campaigns (Ďurech et al. 2011) and spatially resolved images using high-spatial resolution adaptive optics at large telescopes (Hanuš et al. 2017). The proper elements included in MP[3]C are taken from the Asteroids Dynamic Site (Knezevic & Milani 2012).

Next, when viewing the distribution of $\log_{10}(p_V)$ of main belt asteroids, we note that there are two peaks, which are indicative of two major compositional classes. To first order, these peaks approximate the C-types and S-types. In this work, we are interested in studying the S-type population and eliminate all C-types of the inner main belt. To do this, we first distribute $\log_{10}(p_V)$ into 165 bins, which is the square root of the number of data points. We then use a χ^2 least squares minimization to fit two Gaussian peaks to the $\log_{10}(p_V)$ distribution as:

$$\mathcal{P}(x) = \mathcal{P}_0(x; A_0, x_0, \sigma_0) + \mathcal{P}_1(x; A_1, x_1, \sigma_1), \quad (1)$$

where $\mathcal{P}$ is the probability density, $\mathcal{P}_0$ is the probability density function of the low-p_V population and $\mathcal{P}_1$ is of the high-p_V population, x is the logarithm of p_V, A is the amplitude, x_0 and x_1 are the means, and σ is the standard deviation. The histogram
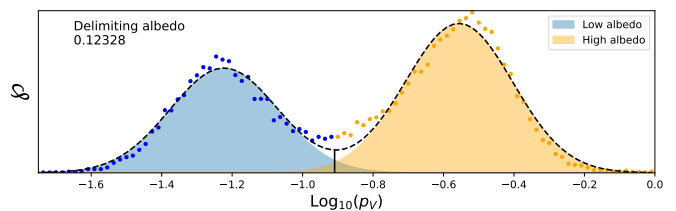
Fig. 3. $\log_{10}$ distribution of p_V for all bodies in the inner main belt. The data points are the bin heights of the histogram, the dotted black line is the best fit of the double-peaked Gaussian histogram. The unimodal components of the low and high p_V are shown in blue and orange, respectively. We impart a delimiting p_V, which is indicated with the vertical black bar.

and fit are shown in Fig. 3. From these, we find the delimiting value of the p_V by finding the ratio of $\mathcal{P}_0/\mathcal{P}_1 = 1$, which is when p_V is 12.328%. This value is consistent with previous findings (Delbo et al. 2017).

We consider this to be a fair delimiting p_V because we have a small degree of contamination and loss. For instance, 1.20% of the area of the low-p_V peak lies above the delimiting p_V and 1.28% of the high-p_V peak lies below it. In this work, we select asteroids whose p_V is higher than 12.328%.

2.2. Synthetic main belt background

We use the synthetic dataset of asteroid diameters and proper orbital elements generated by Deienno et al. (2021), which predicts what the inner main belt would look like if no asteroid collisional families were present. To do this, Deienno et al. (2021) built on the work of Tsirvoulis et al. (2018), who removed all of the families in the "pristine zone" of the main belt between 2.82 and 2.96 au, which contains easily identifiable families. After removing these families, Tsirvoulis et al. (2018) found that the remaining background population has a cumulative size–frequency distribution (SFD) described by $c(D) \propto D^{-q}$, where q is the power-law slope, c is the cumulative counts, and D is the diameter. The value of q was found to be 1.43 and Deienno et al. (2021) assumes this to be the same power law governing the background population of the entire main belt. While generating bodies from this distribution, the authors truncate the diameters at a minimum size of 2 km. This is still valid for our purposes because 70% of the bodies of the inner main belt from the MP[3]C have diameters greater than 2 km and none have diameters of smaller than 1 km. Additionally, the slope of the V does not change with diameter, nor is diameter an input to the HCM.

After their generation, the synthetic orbital elements of Deienno et al. (2021) are randomized and then integrated for 100 Myr. With this, the bodies felt gravitational effects of all Solar System planets from Venus to Neptune and the nongravitational Yarkovsky thermal force. The authors observed that the steady state was reached by 100 Myr and continuing the evolution only depleted the population. Lastly, instead of performing a detailed Hamiltionian analysis to obtain the proper orbital elements, the authors opted to average the Keplerian orbital elements over the last 10 Myr as a proxy[2]. Although different in

---

[1] mp3c.oca.eu

[2] For asteroids with proper eccentricity or inclination larger than the forced eccentricity or inclination, respectively, the time-averaged eccentricity or inclination is close to the proper value. The difference is instead large if the proper eccentricity or inclination is smaller than the respective forced one. Because the forced eccentricities and inclinations are small, the differences between averaged orbital elements and the proper ones is small for most asteroids.
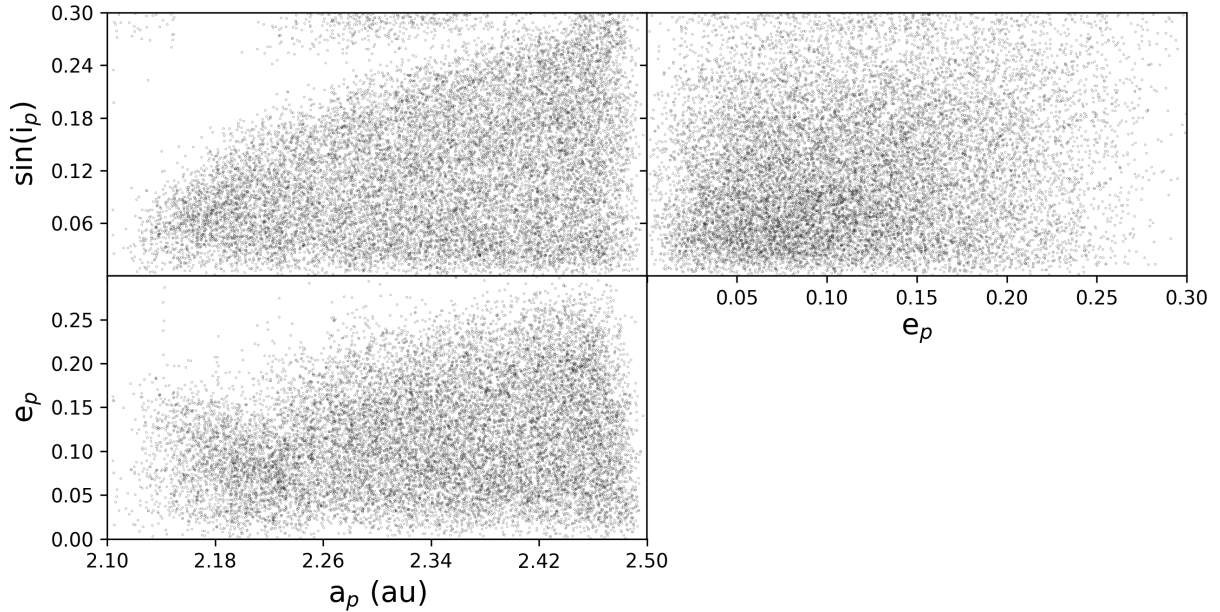
Fig. 4. Axes are the same as in Fig. 1. The data come from the modeled background of the inner main belt from Deienno et al. (2021), which simulates the dynamic evolution of the asteroids of the background population in the main belt.

definition, the simplified time averaging is more than sufficient for our purposes since we are not interested in the precise orbital history of individual simulated asteroid orbits, but the ensemble statistical properties of the main belt. The final result of Deienno et al. (2021) is displayed in Fig. 4.

2.3. Usage of the known families

We use the definitions of the high-p_V inner main belt families from Nesvorný et al. (2015) as a starting point. These families are Baptistina, Flora, Lucienne, Massalia, Nysa, and Vesta. This is a simpler picture of the families in this region because we are not considering the complex subfamily structures. For instance, there is evidence of a subfamily within Nysa whose largest object is (135) Hertha (Dykhuis & Greenberg 2015). Given that we are searching to remove all known families, this problem does not concern us, because removing Nysa will also remove Hertha.

The parameters describing the V-shape are taken from Nesvorný et al. (2015). More precisely, we take the C_0 -values, which can be related to the slope, K, of the V-shape by $K = \sqrt{p_V}/(1329 C_0)$. Additionally, when we apply the HCM, we search for the families one by one, instead of searching for multiple clusters by applying the HCM to the entire dataset, as has been done in previous works (Masiero et al. 2013). While searching for one cluster at a time, the HCM must be given a central body from which to start its search. The central bodies of the families are also taken from Nesvorný et al. (2015) and are typically the largest asteroid in a given family. Given this, we proceed to our own family membership assessment in order to be more inclusive, following an approach similar to that of Tsirvoulis et al. (2018), but with important differences which are explained in the methods section that follows.

3. Methods

Given the problem that the current asteroid family membership lists – found with HCM – are conservative, as detailed in Sect. 1, our first procedure is to reassess the extent of known families

using a new method that attempts to merge their cores with their halos. This consists in constraining the HCM using the V-shapes of the families and a statistically robust method to stop HCM clustering. Once we have removed the reassessed known families, we hunt for missing ones – which are likely older than ~2.5 Gyr – by searching for their Yarkovsky effect signature (i.e., their V-shape; Bolin et al. 2017). As we demonstrate below, this search is successful.

This experiment is procedural, which is to say that results from one step are the initial conditions for the next. In order to estimate uncertainties, at each step, we explore all possible values for all free parameters with a Monte Carlo approach and restrict the range of each free parameter to a valid domain. This allows us to know the sensitivity of our results with respect to the free parameters. Here, we layout the assumptions and actions taken in each step to quantify and limit our uncertainty. First, the underlying assumptions are as follows, to first order:

1. The asteroid belt is composed of collisional families, which are amongst the planetesimals that comprise the "background" population.
2. Members of a collisional family cannot drift farther in the semi-major axis than allowed by the maximum Yarkovsky drift rate; if asteroids encounter zones of dynamical instability during this process, then these bodies are likely removed from the main belt.
3. The SFD of planetesimals in the pristine zone of the main belt (as estimated by Tsirvoulis et al. 2018) is the same as the SFD of the inner main belt.
4. The p_V of an asteroid is indicative of composition, and asteroids of different composition do not originate from the same parent body.

A significant consequence of these assumptions is that all bodies outside all V-shapes are planetesimals. The last three assumptions clearly have limitations, but we consider them valid for the scope of this work. For instance, collisional families containing members of distinct composition have been reported (Oszkiewicz et al. 2015; Fornasier et al. 2016); but in general, members have uniform composition within each family

(Nesvorný et al. 2015). With these assumptions in mind, we can outline our methods, showcasing where and how we treat the uncertainties and free parameters. To specify, we define uncertainties as quantities that have associated errors bars, namely the diameter, D, and its 1σ uncertainty, δD. The free parameters are subjective parameters that take a fixed value per calculation, affect the result, and whose best value cannot be known beforehand. The two free parameters considered here are the absolute number density of the synthetic background N_d, and the width of the lobes used in the V-shape detection, a_w (see Bolin et al. 2017, Fig. 3). Finally, when searching for the original SFD, we consider two different models to correct for the depletion of bodies as a function of diameter via dynamical and collisional loss.

3.1. The V-shape-constrained hierarchical clustering method

Here, we present a new technique for rejecting more false positives than was possible before and for determining the quasi-random noise, which is an inherent problem when using the HCM (Zappala et al. 1990). We briefly describe the input and output datasets and the variables involved in this step of the experiment. We then describe our method in detail, while in parallel we describe how clustering was done in previous works.

1. Inputs:
   - $\{a_p, e_p, i_p, D\}_{\mathrm{Real}}$. These are proper elements and diameters for all high-p_V asteroids in the inner main belt within the MP^{3}C catalog.
   - $\{a_p, e_p, i_p, D\}_{\mathrm{Syn}}$. These are proper elements and diameters of the synthetic dataset from Deienno et al. (2021).
2. Free parameters considered here:
   - $[N_d]$, which is the absolute value of the number density of the synthetic background.
3. Uncertainties considered here:
   - $\{\delta a_p, \delta e_p, \delta i_p\}_{\mathrm{Syn}}$. These are the proper elements of the synthetic background while applying the clustering algorithm.
4. Outputs:
   - $\{a_p, e_p, i_p, D\}_{\mathrm{FR}}$. This is the "Family Removed" dataset, which is a subset of $\{a_p, e_p, i_p, D\}_{\mathrm{Real}}$.

To build from the start, the classical HCM works by first placing each body in its own cluster and then subsequently merging clusters as a given distance parameter is increased (Zappala et al. 1990). In the case of asteroid families, this distance metric is a velocity, v, which can be understood as the Δv required to change the orbit of one asteroid to that of another. The velocity that is used to define the size of the family is known as the velocity cut off (v_c, Zappala et al. 1990; Nesvorný et al. 2015). Deciding the value of v_c has known difficulties. For instance, at small v_c, a family cluster clearly contains too few members. On the other hand, as v_c is increased, the HCM will include more family members but at the same time will include bodies that belong to other families or the background. Indeed, if v_c is large enough, the entire dataset would be grouped into one family, mostly including false positives into the real family.

The ideal situation would be to increase the v_c as much as possible without accepting any false positives. One excellent way to do this is to reject all bodies that cannot belong to the family a priori, and restrict the HCM to search for a cluster over a smaller subset of data that only includes plausible family members.

The first criterion used in this study to reject false positives is the p_V. To reject false positives even further, we introduce a new technique, which is to constrain the HCM to cluster asteroids that are only inside the V-shape of a particular family. That is to say, we subsample the real dataset, $\{a_p, e_p, i_p, D\}_{\mathrm{Real}}$, based on

the semi-major axis and the diameter. Qualitatively, all bodies outside of the target family's V-shape are rejected. Quantitatively, the V-shape of a family is determined by the bodies that experience maximum Yarkovsly drift, the diameters of such bodies given with the relation: $1/D = K|a_p - a_c|$, where a_c and K are the vertex and slope parameters of a family's V-shape, and a_p and D are of an asteroid undergoing maximum drift rate. Then, asteroids (denoted with index j) are rejected if $1/D_j < K|a_{p,i} - a_c|$ and accepted if $1/D_j > K|a_{p,i} - a_c|$. For each family, we do not redetermine a_c and K but simply take them from the compiled list from Nesvorný et al. (2015) and use them as fixed values. We find that this is a viable method because the V-shape parameters are in principle independent of the number of family members added (Milani et al. 2014) meaning that, the V-shape should not change by reassessing family membership.

Most bodies that are outside of the V-shape should be unrelated to the family, because the Yarkovsky drift is not able to carry them that far; we say most, because some phenomena, such as impulse-like events including close encounters with massive asteroids (and planets in rare cases), can push asteroids outside of the Yarkovsky V-shape – not to mention that measurement uncertainties may also cause a small percentage of members to be counted as being outside of the V-shape. These asteroids that experienced impulse-like events had their orbits dramatically changed over short timescales, which broke their dynamical link to the family. This makes it impossible to differentiate them from background asteroids and relate them to a specific family on the basis of dynamics alone. Fortunately, it has been shown that the effect of these processes is small (Delbo et al. 2017).

We note that, within a V-shape, all values of e_p and i_p are accepted; therefore, while viewing the subset of data in orbital element space for a particular family, a_p will be restricted but e_p and i_p will span their entire respective domains. With this subset, we apply the HCM to the central body of each family and study the growth of the number of its members as a function of v_c.

The second novelty of the present work is the method we use to determine the most optimal v_c. We extend the idea of a "quasi-random noise", which was introduced by Zappala et al. (1990). We apply the HCM at sequentially increasing values of v_c to both a real and a reference dataset in parallel. The reference dataset is a random distribution of points that should have no clusters. When the HCM begins forming clusters on the reference dataset, the v_c has been determined. This is because the algorithm is clustering objects from statistical chance, and not because a cluster exists there.

Previously, Zappala et al. (1990) generated the reference dataset by shuffling the real data in such a way as to dissolve the clusters while preserving large-scale variations in number density. Though ground-breaking at the time, we are now able to improve on this by taking advantage of a synthetic dataset generated by Deienno et al. (2021). To determine the quasi-random noise, we subsample the synthetic main belt using the V-shape of each family in the same way as the real dataset, and then add the central body. The HCM then searches for a cluster about the central body. The v_c is determined when two bodies are clustered in the synthetic data set. This entire process is demonstrated in Fig. 5.

By introducing the synthetic dataset, in addition to inserting a modeling assumption, as listed in Sect. 3, we introduce a free-parameter that will impact the results. That is, although the spatial variation of the synthetic dataset is well modeled, the absolute value of the count density, N_d, is arbitrary. We therefore perform a sensitivity analysis on N_d, spanning a range from physically too large to excessively small. First, we start with
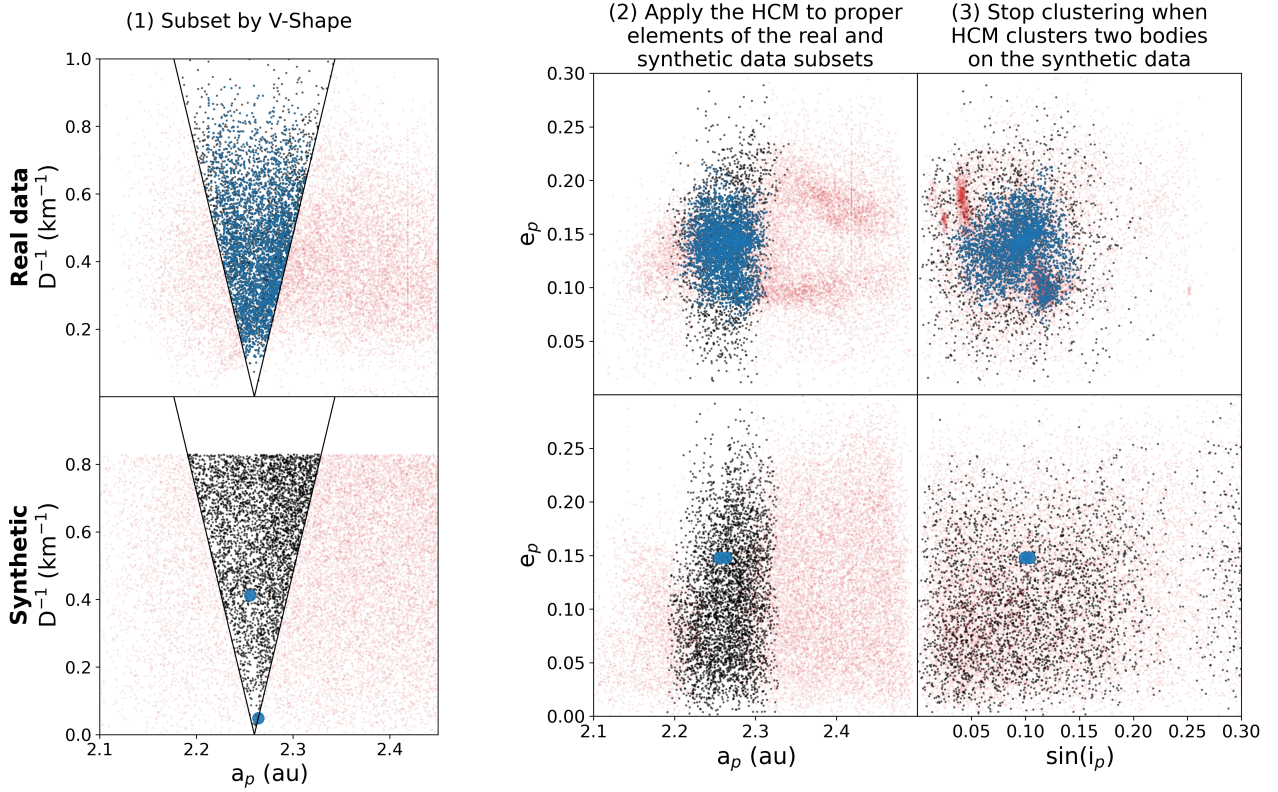
Fig. 5. Demonstration of the V-shape-constrained HCM. The top row depicts the real data while the bottom shows the synthetic background of Deienno et al. (2021). The red bodies are a priori rejected since they are exterior to the family's V-shape. The blue bodies are those clustered by the HCM. In the bottom panels, the two bodies are enlarged for clarity and only two bodies are clustered because forming a cluster on the synthetic dataset determines the v_c (clustering threshold) for the real data. Thus, the blue bodies in the top plot correspond to family members. The black bodies are those who were not a priori rejected but also not clustered at the given v_c. The plots on the right show the same dataset as the left column but plotted in proper element space, which is the data space to which the HCM is applied.

Table 1. v_c at which the HCM clustering is stopped for each family for different N_d of the artificial background in units of m s^{-1}.

| N_d (frac.)	1	3/4	2/3	1/2	1/3	1/4	1/5	1/10	1/100
Baptistina	150	150	150	150	165	165	175	200	540
Euterpe	80	80	80	80	115	225	160	280	540
Flora	200	200	200	200	225	225	225	265	547
Lucienne	95	95	95	95	110	180	180	252	667
Massalia	135	135	135	135	140	185	195	215	497
Nysa	140	140	140	150	150	190	195	215	450
Vesta	145	145	145	145	180	205	205	245	517

Notes. These are the median values found from applying the HCM to 100 realizations of the artificial background as shown in Fig. A.1. The columns correspond to uniformly decreasing the N_d of the artificial background to fractions of the original total number of bodies, 18 191, as explained in Sect. 3.1.

18 191 bodies, which is the number of bodies provided from Deienno et al. (2021) and conveniently greater than the size of the real dataset, which has 16 200 bodies. We choose a lower limit of $N_d/100$, which only contains 182 bodies. This intuitively appears to be too few, which we later show to be the case. Within this range, we explore several densities as a fraction of the original number of bodies, as listed in Table 1.

Next, to create each subset, we randomly removed bodies via their index until we lowered the original number density to the desired fraction, ensuring that we do not alter the relative density variations of the proper elements. Yet, this random subsampling adds a stochastic uncertainty as we apply the HCM. That is to

say, in one random subsampling, an asteroid could be placed in close proximity to the central body, providing a smaller v_c, yet in another case the closer bodies could be removed, leading to larger values of v_c. To combat this, for each density we perform 100 random subsamplings and apply the HCM to each of those. Then, we examine the distributions of v_c for each N_d, which is presented in Fig. A.1. One thing to notice is that for all densities greater than 1/10, the distributions are well behaved, whereas when the background number density is 1/100 there is a seemingly uniformly random selection of v_c, indicating that this N_d is too low. Given this, we then only consider densities between 1 and 1/10 for the remainder of the study. At each selected density,
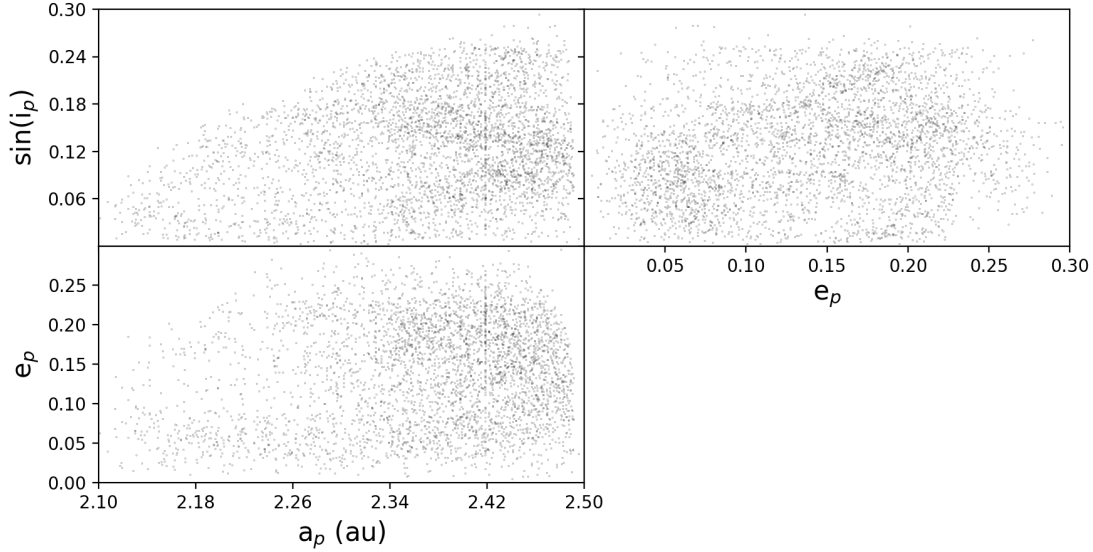
Fig. 6. Inner main belt after removing all of the asteroid families identified in Table 1 given our membership reassessment.

we choose the median value of the distributions of the v_c-values as the final velocity cut-off for each family.

Next, we apply the HCM – constrained by the V-shape on the real data using the v_c obtained at the previous step – to each family of Table 1. After performing the reassessment for each family, we removed all objects identified as collisional fragments. As an output from this step, we have eight family-removed datasets: $\{a_p, e_p, i_p, D\}_{FR}$, one for each N_d considered. By doing so, for the next steps where we search for undetected families, we can also explore the sensitivity of our final results to the free parameter N_d.

3.2. V-shape search for ancient families

After reassessing and removing the known families, we are left with a "background" in which there are no obvious signs of clusters of families in proper element space (as can be seen in Fig. 6). In an effort to discover previously obscured families, we search for traces of them using the V-shape detection method (Bolin et al. 2017; Delbo et al. 2017). By continuing from the family-removed background from the previous step, we apply a modified V-shape detection scheme that considers the following:

1. Inputs:
   - $\{a_p, e_p, i_p, D\}_{FR}$, which is the family removed background.
2. Free-parameters:
   - $[N_d]$, the absolute number density of the synthetic background.
   - $[a_w]$, the width of the lobes in the V-shape search method.
3. Uncertainties:
   - $\{\delta D\}_{FR}$, the uncertainties in the asteroid diameters.
4. Outputs:
   - $(a_c \pm \delta a_c, K \pm \delta K)$, V-shape parameters of the detected families.
   - $t_{age} \pm \delta t_{age}$, the age of the detected families.
   - X-σ, the probability that the ancient family detection is not a statistical coincidence.
   - $\{a_p, e_p, i_p, D\}_{AF}$, members of the V-shape detected families.
   - $\{a_p, e_p, i_p, D\}_P$, planetesimals.

In summary, the V-shape detection method works by detecting a change in density in the $(a, 1/D)$ plane while transitioning from a region occupied by the family to a region unoccupied by the family. In other words, the entire space is scanned with a series of test vertices and slopes of the V-shape, which creates a two-dimensional grid known as the score-map. There are a few different ways of performing this scan, though we use the a_w method, which was developed by Bolin et al. (2017) and Delbo et al. (2017) and further studied by Deienno et al. (2021).

In this method, two lobes are created interior and exterior to the best-fit line (see Fig. 3 of Bolin et al. 2017, or Fig. 7 in this work for a visualization of the lobes). These two lobes are bounded by the best-fit line and other V-shape lines that have the same slope and that are merely shifted above or below by user-defined distance in semi-major axis, a_w. The interior lobe is therefore between the line $y = \mathrm{K}(|a - a_c| + a_w)$ and the best-fit line, whereas the exterior lobe is between $y = \mathrm{K}(|a - a_c| - a_w)$ and the best-fit line. The y for the best-fit line ($y = \mathrm{K}|a - a_c|$) has units of $1/D$ and would be an asteroid that is experiencing maximum Yarkovsky drift.

To measure the change in number density, we count the number of bodies within the interior and exterior lobes (we note that this is a number density in $(a, 1/D)$ space and not the number density, N_d, in proper element space). The ratio found between the number of interior bodies squared, N_{int}^2, and the number of exterior bodies, N_{ext}, is used as the score. A grid of test values is created that spans a range of both a_c and K. A score is calculated at each coordinate. Local maxima of this score map are candidate detections of family V-shapes.

In this study, we are interested in detecting the oldest families in the inner main belt that are likely older than about two billion years (which are thought to be beyond the effectiveness of the HCM; see Brož et al. 2013; Spoto et al. 2015; Bolin et al. 2017), and therefore we search for slopes smaller than 1.5 km^{-1} au^{-1}. Given the age estimation from Nesvorný et al. (2015), we calculate the age of an asteroid family as:

$$t_{age} \approx 1\,\mathrm{Gyr} \times \left(\frac{1}{5}\right)^{1/2} \left(\frac{1}{1329\,\mathrm{km}}\right) \left(\frac{1}{10^{-4}\,\mathrm{au}}\right)$$
$$\times \left(\frac{a_c}{2.5\,\mathrm{au}}\right)^2 \left(\frac{\rho}{2.5\,\mathrm{g\,cm^{-3}}}\right) \left(\frac{1}{K}\right), \tag{2}$$

where ρ is the mean bulk density of all asteroids in the family. A family with a slope of 1.5 km^{-1} au^{-1} and a mean bulk
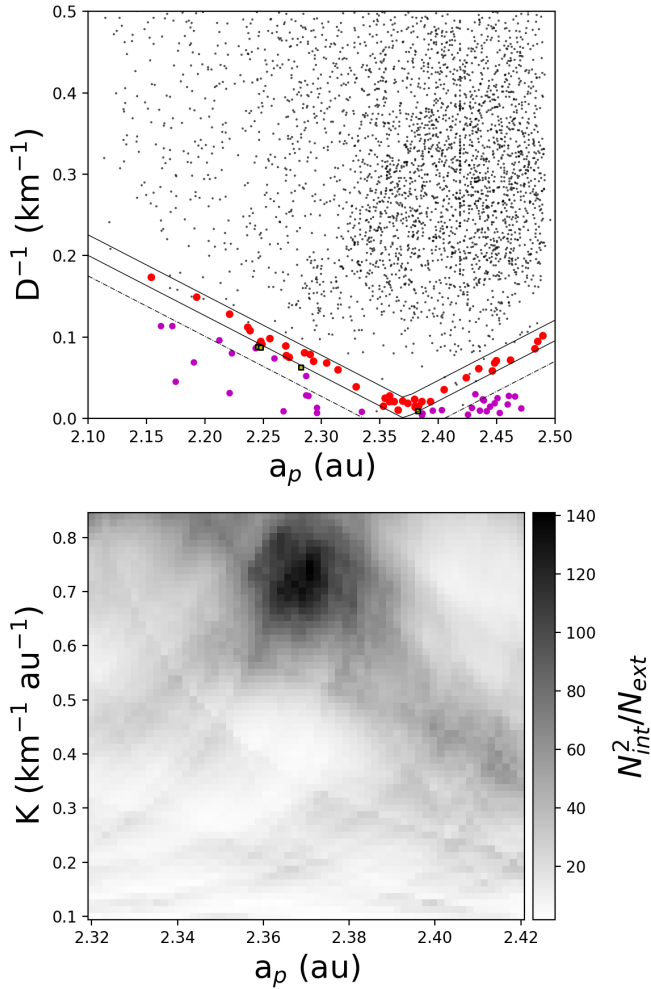
Fig. 7. Detection of an ancient family through the Yarkovsky V-shape. Top: the V-shape down in the a_p, $1/D$ plane. The inner and outer lobes of the V-shape are shown; these are requisite of the detection method, as described in the main text. As a proxy for membership, those within the interior lobes are considered to be family members and are shown as red dots. The purple dots are planetesimals, gray dots are uncategorized, and squares could be either planetesimals or ancient family members. Bottom: score map consisting of a stacked average of 700 Monte Carlo iterations in detecting the ancient family. The maxima gives the coordinates of the V-shapes vertex and slope parameters, a_c and K.

density of 2.5 g cm^{-3} placed in the inner main belt would have an age of 1.9 Gyr. Therefore, covering all slopes smaller than 1.5^{-1} au^{-1} ensures that we scan for all families that are likely beyond the HCM detection limit (Bolin et al. 2017). Additionally, we use Eq. (2) for computing the age of detected families and accordingly find its error as:

$$\sigma^2_{t_{age}} \approx \left(\frac{\partial t_{age}}{\partial a_c}\right)^2 \sigma^2_{a_c} + \left(\frac{\partial t_{age}}{\partial K}\right)^2 \sigma^2_K + \left(\frac{\partial t_{age}}{\partial \rho}\right)^2 \sigma^2_\rho + 2\frac{\partial t_{age}}{\partial a_c}\frac{\partial t_{age}}{\partial K}\sigma^2_{a_p K},$$
(3)

where σ represents the standard deviation of the corresponding variable and $\sigma^2_{a_c k}$ is the covariance between the slope and vertex.

As we show in Sect. 4, our V-shape scan results in a detection of one of the most ancient families known. We want to gain some insight into the membership of the ancient family, which is unfortunately beyond the HCM detection limit. To bypass this, we consider bodies within the interior lobe as being family members. These clearly comprise a fraction of the entire family, that

is, the few members that have near-maximum Yarkovsky drift, and there should be many more family members interior to the V-shape. Unfortunately, from the data we consider here alone, we are unable to discern them from other bodies that are in the V-shape, notably: planetesimals or missed bodies from younger clustering families, known or unknown. In fact, this proxy of using the interior lobe to probe membership is the reason we chose the a_w method. Previous works also considered the dK method, which, instead of defining the lobes by two other lines that have the same slope with a slight vertical offset, defines the lobes with two lines that have slightly different slopes but share the same vertex. As the lobes come to a point at the vertex, bodies with small $1/D$ may not be interior to the lobe, and may therefore be excluded from the membership list.

To obtain a detection of such a family, we followed a three-step process, considering the free parameters: N_d and a_w and the uncertainties in the diameters. Again, for each of these we performed a Monte Carlo-style analysis. After these procedures, we identified a family, and we report a confidence level of its detection in Sect. 4.2 and the resultant data products in Appendix B, which include identification of some planetesimals below its V-shape as well as some members of the ancient family.

Step 1: Performing a global search

Deienno et al. (2021) investigated the robustness of a series of techniques for identifying asteroid families, one of which is the a_w technique. The authors find some key results: first, families with a higher contrast to the background are clearly easier to detect. Second, the score map (see Fig. 8 of Deienno et al. 2021 or Fig. 8 here for a visualization) should show a curved X-like shape when a family is detected. Each line of the X corresponds to the right or left sides of the lobes of the V-shape search, and the maximum score should be at the intersection of the two lines of this X. Next, and most importantly, the optimal a_w anti-correlates with the slope of the family; that is, larger values of a_w allow improved detection of more open V-shapes.

However, Deienno et al. (2021) were able to calibrate the a_w method with the simulated data because they knew the correct values of the vertex and slope they were aiming to detect, which is not the case for us. To combat the subjectivity of the choice of a_w, we test a large range from too thin to too thick, and then restrict a_w over a reasonable domain. More precisely, for the global search, we use nine different values of a_w logarithmically spaced from 0.01 au to 0.5 au. The lower limit creates lobes that are too thin, and these therefore only capture a small number of bodies per $N^2_{interior}/N_{exterior}$, and the search is sensitive to stochastic placement of the asteroids in $(a_p, 1/D)$ space. In other words, the sampling using thin lobes is more sensitive to the random placement of bodies than the larger scale variation of where a family begins and ends. The larger limit is 0.5 au, which is the entire width of the inner main belt. As this upper limit is of the same order as the width of the domain of semi-major axis values for our entire dataset, it would not be sensitive to variations within this domain.

To specify the technical details of our procedure, first we begin by defining the scan range, where each score map is a 100×100 linearly spaced grid, where the vertex varies within $2.1 < a_c < 2.5$ au and the slope varies within the range of $0 < K < 1.5$ km^{-1} au^{-1}. We then begin creating base datasets in a Monte Carlo fashion by considering the uncertainties on the diameter, as well as the free parameter N_d. That is, for each of the seven densities ranging between 3/4 and 1/10 in Table 1, we create 20 random realizations of the asteroid diameters based
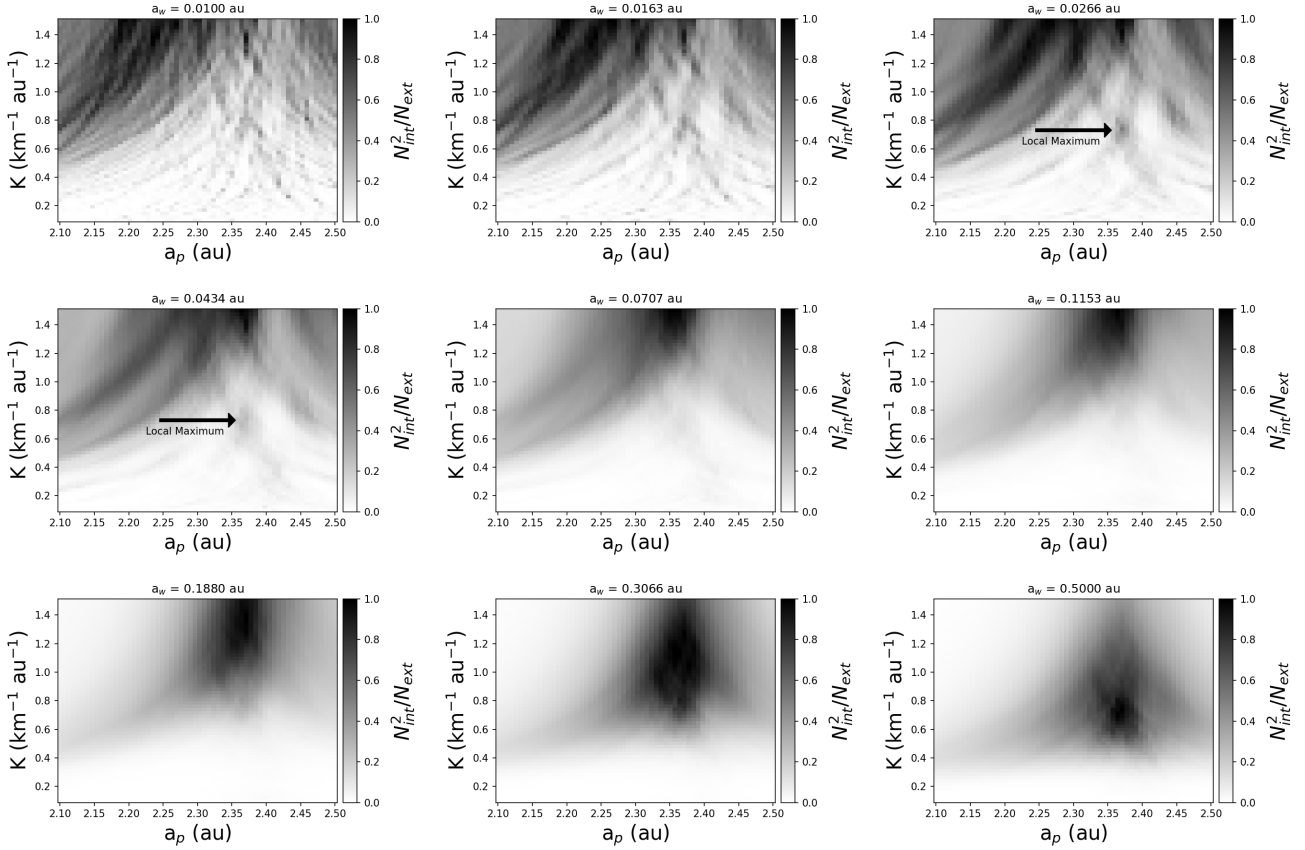
Fig. 8. Each score map is a stacked average of 140 score maps comprising 20 realizations of the diameters for each of the seven choices of the v_c corresponding to N_d. Each score map uses a different lobe width, as indicated by the respective titles. The scores are stretched and normalized such that all values are between zero and one. The signature of the local maximum indicating the ancient family is labeled for two lobe widths.

on Gaussian sampling, where the best value is the mean and the uncertainties are one standard deviation[3]. In all, we create 140 base datasets. Next, we fill in the score maps of each of the 140 base datasets for all nine a_w-values, leading to $9 \times 140 = 1260$ score maps. The 140 datasets are then averaged for each a_w and we are left with nine total score maps to evaluate for our global search.

Step 2: Fine-tuning the detection

With the nine score maps ready, we first evaluate each by eye. If a family is present, we expect that as a_w increases, the a_c of the local maxima remains constant and K decreases. After the candidate family is identified, we then begin fine-tuning the detection by performing a second search over a restricted range of a_c, K, and a_w. We restrict a_w because we anticipate only a small range will converge to the proper detection, as demonstrated by Deienno et al. (2021). We also desire to use the smallest a_w possible. We anticipate that larger a_w will work

best for detecting old families. Nevertheless, we use the widths of the inner lobe as a way to gather family members. In this way, we want the thinnest lobes possible to limit the number of false positives.

In regards to reducing the search range of a_c and K, the motivation is simple. We expect in the global search that a family may be indicated by a local maximum on the score map – which may not necessarily be the global maximum. Therefore, we restrict the search range of a_c and K such that the local maximum of the family is also the global maximum. We opt for this instead of applying another algorithm to sort all local maxima and select the correct coordinates.

Next, after a_w and the search range of the score map are restricted, we perform another search against all 140 base datasets. We consider bodies within the interior lobe to be family members, and those below it to be planetesimals. To establish a confidence of our membership, we track the number of times a body is assigned to either the planetesimals or the ancient family. Bodies that are assigned more often to one class than another are more likely to be true members. Thus, we can choose our membership as a function of statistical significance. In this work, we report bodies at a statistical significance level of 1σ. That is, bodies that are assigned for more than 68% of the total number of iterations.

Step 3: Quantify detection confidence

Finally, we employ a statistical test to establish detection confidence by following the same procedure as Delbo et al. (2019). Briefly, we test the null hypothesis that the $1/D$ distribution has

---

[3] We note that about 10% of the diameters have fractional uncertainties of greater than 1/2, and only a handful are greater than 1. These high fractional uncertainties result in some iterations producing asteroids with negative diameters, which is a nonphysical consequence of approximating the errors as being Gaussian distributed. To avoid this, we take the absolute values. A more thorough treatment could include a bayesian inference of the diameter posterior probabilities given the infrared observations. The limitation arising from this approximation is negligible because the fraction of high-uncertainty asteroids is low, and the diameters of those bodies are small. Small bodies are ultimately not categorized as planetesimals, which are typically large.

no correlation with the distribution of semi-major axis. Here, as in the previous steps, we perform a Monte Carlo approach, considering many base datasets by varying N_d and performing many realizations of the diameters. Effectively, we compare multiple base datasets – which serve as our nominal case – to multiple "shuffled" datasets. To create the shuffled datasets, we first find the distribution of semi-major axis, $\mathcal{D}(a_p)$, and subsequently randomly sample this distribution for each asteroid. Between each nominal and shuffled dataset, we compare the ratio of the number of bodies that are inside of the family's V-shape, $R_{\text{in/out}}$ to those outside. In effect, if the detection is real, $R_{\text{in/out}}$ will be systematically larger in the nominal case than in the shuffled case. Whereas if the detection is a coincidence, then there will be no systematic preference between $R_{\text{in/out}}$ being greater in either the nominal or shuffled cases. To quantify the detection confidence, we perform the following procedure:

- `counts=0`
- **for** N_d in $\{N_d\}$ **do**
  - Evaluate the distribution of the semi-major axes $\mathcal{D}(a_p)$
  - **for** j in 100 **do**
    - Build nominal set $Nom = \{(a_p, D)\}$ by Monte Carlo Gaussian sampling of the diameters: $\{D|\mathcal{N}(D, \delta D)\}$
    - Measure $R_{\text{in/out}}^{\text{nom}}$
    - **for** k in 100 **do**
      - Build shuffled set, Shuff $= \{(a_p, D)\}$, with the same diameters from the nominal set but randomly sampling the semi-major axis: $\{a_p|\mathcal{D}(a_p)\}$
      - Measure $R_{\text{in/out}}^{\text{shuff}}$
      - **if** $R_{\text{in/out}}^{\text{shuff}} \geq R_{\text{in/out}}^{\text{nom}}$ **then**
        - `counts+=1`
      - **end if**
    - **end for**
  - **end for**
- **end for**

At the end of this procedure, we look at `counts`, which is the number of times the shuffled dataset has a larger ratio of the number of bodies inside to outside the V-shape than the nominal case. We then divide this number by the total number of iterations and find the statistical significance.

3.3. Correction of dynamical and collisional loss

By reassessing the known families as well as detecting new ones, we can remove them and obtain a list of planetesimals. At this point, we wish to infer the original SFD during the early days of the Solar System from the SFD we observe today. To do this, we replicate the methods developed in Delbo et al. (2017). Briefly, the authors correct the SFD for the dynamical and collisional loss of asteroids in the inner main belt. To start, a computer program is given a nominal observed SFD of planetesimals. Additionally, the authors appended their list with parent bodies of other known asteroid families. Beyond this, they also corrected for the fragmentation of these bodies. In our case, we append the planetesimals with bodies such as (8) Flora, (27) Euterpe, and (289) Baptistina. In the case of Flora, its current day diameter is estimated to be 130 km, but by adding the mass of its collisional fragments, we can estimate it was a body of 260 km in size. We note that this estimate uses the family membership list from Nesvorný et al. (2015), and not the list found in this work.

Next, the corrections for the dynamical and collisional loss were treated separately. For the collisional loss, a collisional lifetime was determined by Bottke et al. (2005), which is a characteristic time in which a fraction of $1/e$ of the asteroids of a given diameter are destroyed (see Fig. 1 of Marchi et al. 2006, for a visual of the collisional lifetime as a function of diameter). Moving forward, estimation of collisional loss is treated in a Monte Carlo fashion, where the input parameters are the total integration time, T, the integration step time, Δt, the collisional lifetime as a function of diameter, $t_c(D)$, and a list of asteroid diameters $\{D\}$. At each time step, the probability of an asteroid being destroyed is approximated as $\mathcal{P} = \Delta t/t_c$, which is valid in the limit $\Delta t \ll t_c$. In the simulations we consider, the integration step is 10 Myr, while the collisional lifetime of asteroids greater than about 35 km is roughly 4.5 Gyr, or the age of the Solar System. Thus, at each integration step and for each asteroid, a random number is generated between 0 and 1; if it is less than $\mathcal{P}$, the asteroid is destroyed.

We note that this correction for the collisional loss assumes the current state of the Solar System for the entire simulation, and does not take into account the fact that the collisional loss was greater in the earlier days in the Solar System. The collisional loss should decrease as a function of time, as there are less bodies for other bodies to collide into. To compensate for this, Bottke et al. (2005) estimated that prolonging the total integration time by 10 Gyr in the current state of the Solar System is roughly equivalent to integrating for 4.5 Gyr in a Solar System that experienced higher collisional loss in its earlier days. We therefore perform two corrections, one compensating for dynamical loss over 4.5 Gyr and the other for 10 Gyr.

In total, we simulate the collisional loss of the planetesimals 10 000 times. We are then able to find the average fractional loss as a function of diameter, which is the ratio, $\mathcal{F}_{\text{collision}}(D) = C_{\text{final}}(D)/C_{\text{initial}}(D)$, or the final number of counts of asteroids of size D divided by the initial number of counts.

At the same time, we compensate for the dynamical loss (see the supplemental materials of Delbo et al. 2017): we search for the fractional loss as a function of diameter. To do this, we created a sample of asteroids from the inner main belt, with low inclination, and low eccentricity orbits, particularly those whose perihelion does not cross the orbit of Mars. Then, the simulation is evolved for 4.5 Gyr. The asteroids feel the effects of all the planets from Venus to Uranus, Ceres, and Vesta, and the Yarkovsky effect. The integration time step is 10 days. To compensate for such a high temporal resolution, only about 600 bodies are used for this integration. A body is removed when its semi-major axis falls below 0.5 au or beyond 30 au, or if it comes within the Hill Sphere of a planet. At the end of the simulation, the dynamical fractional loss, $\mathcal{F}_{\text{dynamic}}(D)$, is found using the same method as described above.

Once we have the average collisional and dynamical loss fractions, we then estimate the original SFD simply by multiplying their fractions:

$$C_{\text{original}}(D) = \frac{1}{\mathcal{F}_{\text{dynamic}}} * \frac{1}{\mathcal{F}_{\text{collision}}} C_{\text{observed}}(D). \qquad (4)$$

4. Results

4.1. Synthesis of family removal

An example of the V-shape-constrained HCM method is presented in Fig. 5 for the case of the Baptistina family. This method

was also applied to all the families labeled in Table 1. The figure shows how the V-shape of the family is used to constrain the population of asteroids to which HCM clustering is applied. We note that this restriction applies to a_p but not e_p and i_p. For each of the plots of Fig. 5, the v_c is 150 m s^{-1}. We note that this central body, (298) Baptistina, is near the vertex of the V-shape and that the second body has a much smaller diameter and is therefore not adjacent to (298) Baptistina in this space. However, in proper element space, these bodies are adjacent. As expected, an asteroid family is recovered from the real data, yet not from the synthetic data, which have no clusters.

While viewing Fig. 5, we notice in $\sin(i_p)$ and e_p space that a small satellite cluster located at (0.12, 0.10) is included in the Baptistina family. This is part of the Vesta family and overlaps with the V-shape of Baptistina. Although a piece of the Vesta family is included in our Baptistina family, this is not problematic because our primary objective is to remove all known collisional family members. Specifying the family of a specific body is not a primary objective. Nevertheless, we present a supplementary analysis whereby we obtain a robust family membership using our V-shape-constrained HCM in Appendix A.

We estimate the uncertainty introduced by the free parameter N_d by removing the families for each v_c listed in Table 1. In other words, we remove the families seven different ways for each density listed in Table 1. By doing so, we are left with seven family-removed datasets, $\{a_p, e_p, i_p, D\}_{\text{FR}}$, with a number of bodies of 3866, 3866, 3707, 2735, 1932, 2278, and 1235, for each N_d fraction of 3/4, 2/3, 1/2, 1/3, 1/4, 1/5, and 1/10, respectively. We note that the case $N_d = 1/4$ has less bodies than the case $N_d = 1/5$. This is due to the value of v_c found for Euterpe, which jumps to a large number from $N_d = 1/3$ to $N_d = 1/4$ because of our choice of using the median value and because the distribution of v_c per family per N_d may be under-sampled (see Fig. A.1). We show the conclusions found in the following steps are valid even when considering all of the tested N_d.

One of the family-removed datasets is presented in Fig. 6. This is for the v_c-value found for using either one of the following values for the free parameter $N_d = [1, 3/4, 2/3]$, since the results are equivalent. As expected, there are no obvious signs of other families as clusters in proper-element space. Importantly, we do not see obvious signs of family "halos", which, on the other hand, are clearly visible if we remove families according to a more conservative membership, as in Fig. 2.

4.2. Detection of an ancient family and planetesimal identification

Having removed previously known families, we perform the global search for undetected families. As there are no obvious signs of clusters in the family-removed dataset of Fig. 6, we turn to the V-shape search. We consider nine different a_w-values, taking into account the effects of the free parameter N_d and the uncertainty on asteroid diameters. First, in the score maps shown in Fig. 8, we can see that the map created with $a_w = 0.01$ au is too sporadic and the global maximum is a random "hot pixel" that does not correspond to a large-scale trend indicative of the presence of a family, but is rather characterized by stochastic sampling. Next, considering $a_w = 0.0707$ au and upward, we see a global maximum that remains invariable within the range 2.35–2.40 au. This is the expected behavior for a detection of a family V-shape. The behavior along the K-axis is more complex. As demonstrated by Deienno et al. (2021), we notice the trend that the global maximum on the score map for a family detection decreases in slope as the width, a_w, increases. We
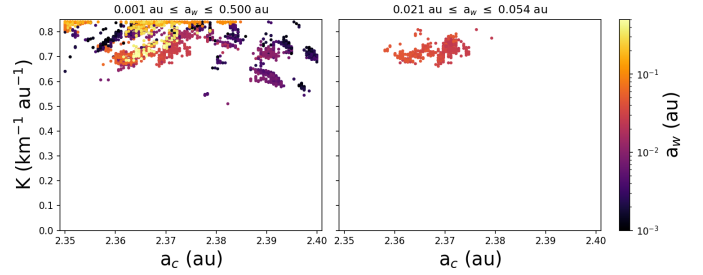
Fig. 9. Demonstration of how the widths of the a_w method were selected. The color indicates the width of the lobes. Each point shows the best found slope and vertex. In the left panel, there are 4200 points (seven values of the v_c, 20 iterations of the diameters uncertainty, and 30 lobe widths.) The right panel is a filtered version of the left, only including lobe widths that converge stably to a solution and includes 700 points.

notice a local maximum at $(a_c, K) = (2.36$ au, 0.7 km^{-1}au^{-1}) that comes into focus and eventually fades away for a_w between 0.0266 and 0.434 au. This local maximum now represents our candidate family.

Next, we aim to fine tune this detection by restricting the range of slopes and vertexes so that the local maximum over this domain is also the global maximum. After the scan range is reduced, we eventually study the behavior of the best a_c and K – which are coordinates of the maxima on the local score maps – in relation to a_w. We reduce a_c to 2.35 and 2.40 au. We reduce the slope range to scan between 0 and 0.84 km^{-1}au^{-1}. A 75 × 75 linearly spaced grid is created within these limits. We then use a higher resolution of 30 values of a_w logarithmically spaced between 0.001 and 0.5 au. Again, in this procedure we take into consideration the dependence of the results on the uncertainty in diameter. Thus, we create 4200 score maps, that is, one for each combination of the seven N_d, the 20 realizations of the diameters, and the 30 values of a_w. Figure 9 is a compilation of these results; contrary to Fig. 8, which shows an average score map per a_w, the left panel of Fig. 9 reports only the maximum found per score map and each marker is color coded according to a_w. Then, the right panel shows the filtered and accepted values of a_w.

To filter the range of a_w, we clip a lower and upper limit of the left panel of Fig. 9 until the resultant score-maxima converge. When a_w is less than 0.021 au, we notice that the values are rather sporadic and maxima span the entire a_c domain. In regard to the upper limit, we see that K correlates too strongly with a_w. For instance, between 0.054 and 0.4 au, the maximum scores are found near to the upper border of the search range. Thus, a_w between 0.021 and 0.054 au is the optimal range because the solutions converge. We note that the largest values of a_w also converge to the same solution. We note that the largest values of a_w also converge to the same solution. However, we opt not to use this for two reasons, the first being that K still correlates too strongly with a_w. This is not the case for the optimal range. Secondly, we use the lobes to determine membership, and large lobes will include more bodies that may be interlopers.

By fine-tuning our detection, we retain five a_w-values tested from 0.021 to 0.056 au. This, in tandem with the 20 realizations of the diameters and the 7 family-removed datasets, leads us to a total number of 700 score maps, which are stacked together and presented in the bottom panel of Fig. 7. In regard to the top panel, we show the nominal V-shape of the ancient family, which is the middle V. The inner and outer Vs are the lines constraining the lobes, using the mean of the filtered a_w values. Next, we

color the planetesimals in purple, the ancient family members in red, and those that could be either in yellow as squares with black edges.

To determine which category a body belongs to, we track the number of times a body falls within the inner lobe and those that are below the determined V-shape for each of the 700 iterations. Those inside the lobe for more than 1σ of all iterations are nominal family members, which represent only a subset of the total family members, namely those experiencing maximum Yarkovsky drift. Those that are below the V-shape are planetesimals. Those bodies that are in yellow belong to either group for more than 68% of the iterations. The rest of the bodies above in gray we now consider to be uncategorized because they could be either members of the ancient family, missed members of the known families, or members of undetected families; they could also be planetesimals, but this is less likely because, as we demonstrate in Sect. 5, their SFD is more consistent with those of families of fragments than with those of the planetesimals. We note that some bodies fall within the interior lobe and are colored black instead of red; this is in light of the presented realization of their diameters in the Monte Carlo sampling; that is, they fell into the lobe for this iteration but not more than 68% of all iterations. In Appendix B, we report a list of nominal family members and planetesimals. Additionally, we show the proper orbital elements in Fig. B.1. Unsurprisingly, the members do not clump but are rather disperse.

With the slope and the vertex of the family's V-shape calibrated, we now follow the procedure identified in step 3 of Sect. 3.2 to establish the detection confidence. Briefly, we find the ratio of bodies interior to the V to those exterior for 700 different iterations of the nominal data and 70 000 different iterations of the shuffled dataset; that is, 100 shufflings of the semi-major axis for each of the realizations of the nominal data. In all, we find 43/70 000 iterations, where $R^{\text{shuf}}_{\text{in/out}} > R^{\text{nom}}_{\text{in/out}}$, leading to a statistical significance of 3.4σ. With the detection confidence established, we return our attention to the right panel of Fig. 9, where we find the mean K and a_c value as well as the cross correlation. We find that the V-shape has a slope of 0.73 ± 0.03 km^{-1} au^{-1}, a_c has a best value of 2.369 ± 0.004 au, and the two variables have a covariance, σ^2_{K,a_p}, of 0.0077 km^{-2}, which is a correlation of 0.55.

4.3. The corrected size–frequency distribution

With the detection of the ancient family established, we can now look at the SFDs of various subsets of the inner main belt, such as: the reassessed known families, the ancient family, the planetesimals, and the remaining uncategorized bodies. This is shown in the top panel of Fig. 10, where the fuzziness of the lines corresponds to the uncertainties from each iteration in the Monte Carlo. Interestingly, the uncategorized bodies have a very similar shape to the known families, indicating that these objects are also collisional fragments. We note that the SFDs of the planetesimals have a much different shape from the SFDs of the known families. Beyond that, if we look at the SFD of the high p_V population of the entire inner main belt, we notice two features that are both humps: the planetesimals corresponding to the first hump of larger bodies on the right, and the known families characterizing the hump for smaller bodies on the left.

The SFD of the ancient family is unlike those of the known families and the planetesimals. However, this is expected, as we know the membership of the ancient family is incomplete, and many are uncategorized. Furthermore, there may be some
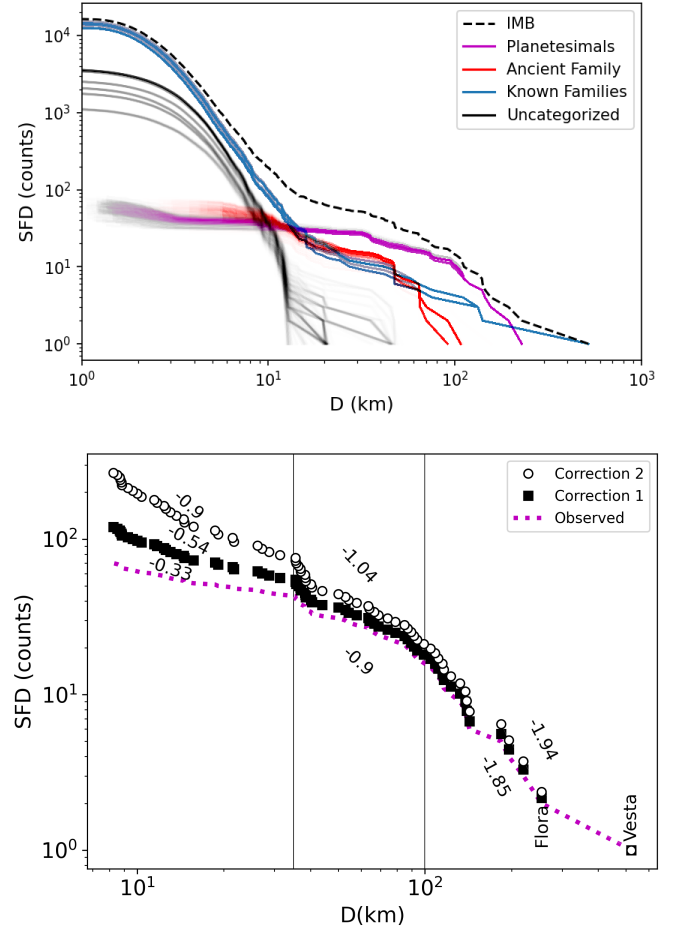
Fig. 10. SFDs of various collections of asteroids. Top: cumulative SFD of five different cases. The high-p_V inner main belt is indicated as IMB, the blue Known Families are the reassessed families listed in Table 1. The uncategorized objects are those that do not belong to the known families, the newly discovered ancient family, or the planetesimals. Multiple curves are drawn according to each iteration in order to reflect the uncertainty. There are 700 highly transparent curves drawn for the Ancient Family, the Planetesimals, and the Uncategorized. There are seven known family curves drawn, and one curve for the inner main belt. Bottom: purple curve labeled "Observed" is a modified version of the Planetesimals from the top panel. This list is a compilation of planetesimals detected with a greater than 1σ confidence, identified planetesimals from previous works, and a few artificial asteroids that should be hidden behind the V-shapes of the ancient families, as explained in the appendix. Next, we provide two different corrections for depletion. Both consider the same dynamical loss while correction 2 accounts for more collision loss than correction 1, as explained in the text. We provide the power-law slope in three different regions: below 35 km, between 35 and 100 km, and above 100 km. We note that the three slopes are reported below 35 km, yet above we only report Correction 1 and 2 for clarity.

planetesimal interlopes within the inner lobes, particularly the large objects that crowd the vertex of the V-shape. Removing only a few large-diameter objects would have a significant effect on the shape of the SFD, particularly in the cumulative count range of $C < 10$.

To finish, we wish to explore the possible SFD of planetesimals in the inner main belt during the early days of the Solar System. To do this, we take the list of objects identified at the 1σ confidence level from Table B.2 and add the other known planetesimals from Delbo et al. (2017, 2019), which totals 54 bodies.

However, given that the planetesimals were identified through a process of elimination, that is, they were not within any family's V-shape in the a–$1/D$ plane, and that the size and semi-major axis should be uncorrelated for the planetesimals, we can expect more planetesimals to be hidden amongst the ancient family's V-shapes. We therefore compensate for the V-shape obscuration, as explained in Appendix C. Briefly, this compensation is treated in a Monte-Carlo fashion; we often generate between 12 and 25 additional bodies. In general, these bodies have small diameters, which is expected because the magnitude of the obscuration is inversely proportional to the diameter. As a result, the compensation does not change the reported slope for the SFD for bodies larger than 100 km, yet increases the reported slopes by a factor of 1.4 for objects smaller than 30 km and is therefore not negligible.

Next, as described in Sect. 3, we correct for dynamical and collisional loss. Two models are presented, both experience comminution rates based on the current state of the inner main belt. One compensates for this collisional loss over 4.5 Gyr, which would be a lower limit given that the loss rate today is lower than before. The second applies the correction for 10 Gyr, which attempts to balance stronger comminution in the past. These in turn establish a possible range of SFDs we would expect for the original planetesimal population and are shown in the bottom panel of Fig. 10.

5. Conclusions and discussion

Given our new technique of applying the HCM on a family-by-family basis, each time restricted by the Yarkovsky V-shape of the family, we were able to reassess the membership of known asteroid families with high albedo, and the issue of "halos". Subsequently, we removed these known families from the inner main belt and applied a V-shape search method to the remaining population: by doing so we identified one of the most ancient asteroid families, which has the widest open V-shape. From determining the slope and the vertex of the family's V-shape, we have two of the three necessary parameters for an age estimate. Finally, with the most ancient family identified, we are able to isolate planetesimals and present a range of possible SFDs given different corrections for dynamical and collisional loss.

In regards to the family's age, as shown in Eq. (2), its computation depends on the value we adopt for the bulk densities of the member asteroids. Vernazza et al. (2021) were able to constrain the shapes of many asteroids using imaging and spectro-polarimetry. If we take the 12 S-type asteroids reported, we obtain a bulk density of 3.0 ± 0.3 g cm^{-3}, which corresponds to an age estimation of 5 ± 1 Gyr. At the same time, the MP^{3}C database has masses and diameters reported for a handful of members, which are listed in the caption of Table B.1. Assuming the asteroids are spheres, we obtain bulk density measurements of 2.6 ± 1 g cm^{-3}, which results in an age estimate of 4.3 ± 1.7 Gyr. However, it must be taken into account that the age estimation has large error bars. A numerical analysis of the dynamics of this family may be performed – similarly to the work presented in Milani et al. (2017) – to reduce the uncertainty.

In either case, the age of this asteroid family is at the limit of the age of the Solar System. Finding a more precise measurement of the age could add a valuable constraint to Solar System evolution models, especially for what concerns the timing of the dynamical instability that affected the orbits of the giant planets. Moreover, the ancient family that we have found is so dispersed in eccentricity and inclination that it must predate the giant planet instability. The interplay between Yarkovsky effect and resonances cannot produce such a dispersed population; only the giant planet instability can (Brasil et al. 2016); see Tsiganis et al. (2005) for the original concept of the giant planet instability.

With the detection of this ancient family, one low-albedo family from Delbo et al. (2017), and two X-complex families from Delbo et al. (2019), there is now a total of four asteroid families discovered within the inner main belt with ages greater than 2 Gyr. However, there are about 25 other families within the inner main belt whose ages are younger than 2 Gyr. As explained in Bottke et al. (2005), the collisional environment is estimated to have been more violent in the early days of the Solar System and as a result we should expect to observe more ancient asteroid families than young families. However, this is not the case. Although pairing the V-shape with the HCM represents a significant advancement in asteroid family detection, the method still has its limitations and therefore detectable families may exist that are still beyond our current detection capabilities. As shown in Deienno et al. (2021), detection efficiency drops greatly as a family's contrast decreases with respect to the background population.

On the other hand, recent studies have emerged suggesting that the asteroid belt was not necessarily as massive upon formation as it is today, which would be consistent with the finding that there are less ancient families than younger ones. Indeed, Raymond & Izidoro (2017) performed a numerical simulation demonstrating a viable mechanism of populating an originally empty belt that results in the observed S/C-type distribution as a consequence of the formation of terrestrial and giant planets. Other studies indicate that planetesimals formed in ring-like structures at certain distances about their host stars and as such must have been transported to other locations later (Morbidelli et al. 2022; Izidoro et al. 2022). Additionally, Deienno et al. (2022) studied the effects of giant planet instability and radial migration on the asteroid belt and found that a high initial Solar System mass could not reproduce the low-mass belt that we observe today. Future numerical studies of family formation and dissipation in the two paradigms of either a massive or sparse initial asteroid belt are necessary in order to determine which is more likely to reproduce the observed distribution in collisional family ages observed today.

Given dynamical depletion, perhaps only ancient asteroid families with a high initial number density remain detectable today regardless of detection efficiency. However, to know if this depletion can compensate for the observed difference between the number of young and ancient families – of about a factor 5 – requires full-scale dynamical evolution simulations of family generation with collisional models and detection. It is possible that such an experiment will show that dynamical depletion cannot explain the observed discrepancy. In such a case, the estimated collisional rates of the early Solar System may need to be revisited.

The SFDs shown in Fig. 10 have important implications. To contextualize the result within the literature, we must note that the presented SFD considers a new compensation that previous works did not; namely we proportionally increase the count of observed bodies to include those that are potentially obscured by the V-shapes of ancient families. With this in mind and with the discovery of this new ancient family, we have been able to identify more planetesimals than previous works, particularly bodies smaller than 50 km in diameter (Delbo et al. 2017). Notably, the SFD of the low-albedo inner-main-belt planetesimals identified by Delbo et al. (2017) was hindered by lacking bodies smaller than 35 km in diameter. To combat this, the authors

extrapolated a power-law slope for $D < 35$ km. Their predictions, when adjusted for the V-shape obscuration, are consistent with the power law deduced from our analyses, which now contains $D < 35$ km planetesimals.

Second, the SFD of the "uncategorized" objects of Fig. 10 resembles the SFD of the known families, supporting the conclusion that they too are collisional fragments and less likely to be original planetesimals. Before carrying out our analysis, we expected uncategorized bodies to be collisional fragments originating from either the known asteroid families, the newly detected ancient asteroid family, the undetected asteroid families, or the planetesimals. Another a priori expectation was that the SFD of the planetesimals may be incomplete, since some of its members may be amidst the uncategorized objects. However, as the SFD of the uncategorized objects indicates that they are collisional fragments, we infer that the deduced planetesimal SFD is complete or at the very least not missing a major component.

Thirdly, the inferred original SFD of the planetesimals can be used for Solar System evolution models (Klahr et al. 2022). Interestingly, the planetesimal SFD has the steepest power-law index in the regime of $D > 100$ km, a turning point around $D \sim 100$ km, a shallower power index for $35 < D < 100$ km, and an even shallower power index for $D < 35$ km, which indicates that planetesimals mostly formed as large objects. This result is consistent with previous studies of the original asteroid population in the inner main belt (Delbo et al. 2017, 2019), in other regions of the main belt (Tsirvoulis et al. 2018), and with numerical simulations that searched for the original SFD of planetesimals that produced – by collisional objects – an SFD that is consistent with the current SFD of the asteroid belt (Bottke et al. 2005; Morbidelli et al. 2009).

Our results regarding the SFD of the original planetesimals are also consistent with theoretical studies of planetesimal formation by pebble cloud collapse in turbulent protoplanetary disks. In particular, Klahr & Schreiber (2020) predict a relatively narrow differential size distribution of planetesimals with a best effective size of around 100 km; this can be intuitively explained by the fact that smaller pebble clouds – with $D \ll 100$ km – are unlikely to collapse into planetesimals, because their self-gravity is overcome by turbulent diffusion, while and larger pebble clouds – with $D \gg 100$ km – would take longer to accumulate than the time they would need to collapse (see also Klahr et al. 2022, for a review).

Lastly, the list of bodies we identify as planetesimals is robust against the free parameters and uncertainties from the reassessment process of the known families. Potential false positives among the identified primordial objects may be fragments from asteroid families that have disassociated beyond a detectable limit or fragments from families that formed before the giant-planet instability, which may have erased the dynamical linkage between members. However, given that the results are insensitive to the rigour with which the family-removal procedure is applied, we expect this number to be a small contribution. Moreover, the robustness supports that the planetesimals and collisional families are indeed two independent populations. That is, the planetesimals are processed through collisional evolution and create the collisional asteroid families, which is a conclusion also supported by Delbo et al. (2017).

A series of follow-up studies could be carried out to either complement this work, benefit from the newly developed methods, or exploit its results. Firstly, the membership lists provided for either the ancient family or the planetesimals contain some interlopers (false positives). For example, knowing the spin states of the asteroids and obtaining their visible and near-infrared spectra could help to identify these interlopers (e.g., Athanasopoulos et al. 2022, used the spin states to do just this). A plethora of visible and near-infrared spectra already exist (Usui et al. 2011), and with the new *Gaia* Data Release 3, the spin states for many asteroids are now available (Galluccio et al. 2023). Also, with more information about the ancient family, a better age estimate may be obtained, which could add a more robust constraint in planetary formation models.

Next, this method could be replicated for the high- and low-p_V populations of the other sections of the main belt. After a reassessment and removal of other known families, we suspect that more planetesimals and ancient families will be discovered. Again, the discovery of the oldest families in other parts of the belt would provide temporal constraints to planetary formation models. Additionally, it would be interesting to see if the SFDs of the planetesimals in other sections of the belt have different shapes to those observed here. Though not the primary goal of this experiment, reassessment of the known families could be studied in further detail. A spectral analysis could be performed to compare the newly identified members in the "halo" with those in the core. At the same time, another analysis could attempt to disentangle the Baptistina and Flora families, which is not attempted here. Finally, we are hopeful that the SFD of the planetesimals can be used as a constraint in Solar System evolution models.

References

Athanasopoulos, D., Hanuš, J., Avdellidou, C., et al. 2022, A&A, 666, A116

Birnstiel, T., Fang, M., & Johansen, A. 2016, Space Sci. Rev., 205, 41

Bolin, B. T., Delbo, M., Morbidelli, A., & Walsh, K. J. 2017, Icarus, 282, 290

Bottke, W. F., Durda, D. D., Nesvorný, D., et al. 2005, Icarus, 175, 111

Bottke, William F., J., Vokrouhlický, D., Rubincam, D. P., & Nesvorný, D. 2006, Ann. Rev. Earth Planet. Sci., 34, 157

Bourdelle-de Micas, J., Fornasier, S., Avdellidou, C., et al. 2022, A&A, 665, A83

Brasil, P., Roig, F., Nesvornỳ, D., et al. 2016, Icarus, 266, 142

Brož, M., & Morbidelli, A. 2013, Icarus, 223, 844

Brož, M., Morbidelli, A., Bottke, W., et al. 2013, A&A, 551, A117

Cuzzi, J. N., Hogan, R. C., & Shariff, K. 2008, ApJ, 687, 1432

Deienno, R., Walsh, K. J., & Delbo, M. 2021, Icarus, 357, 114218

Deienno, R., Izidoro, A., Morbidelli, A., Nesvornỳ, D., & Bottke, W. F. 2022, ApJ, 936, L24

Delbo, M., Mueller, M., Emery, J. P., Rozitis, B., & Capria, M. T. 2015, Asteroid Thermophysical Modeling (Tucson: University of Arizona Press)

Delbo, M., Walsh, K., Bolin, B., Avdellidou, C., & Morbidelli, A. 2017, Science, 357, 1026

Delbo, M., Avdellidou, C., & Morbidelli, A. 2019, A&A, 624, A69

Delbo, M., Avdellidou, C., Bruot, N., & Erard, S. 2022, EPSC Abstracts, 16, EPSC2022-323

Dermott, S. F., Christou, A. A., Li, D., Kehoe, T. J. J., & Robinson, J. M. 2018, Nat. Astron., 2, 549

Ďurech, J., Kaasalainen, M., Herald, D., et al. 2011, Icarus, 214, 652

Dykhuis, M. J., & Greenberg, R. 2015, Icarus, 252, 199

Emsenhuber, A., Mordasini, C., Burn, R., et al. 2021, A&A, 656, A70

Fornasier, S., Lantz, C., Perna, D., et al. 2016, Icarus, 269, 1

Galluccio, L., Delbo, M., De Angeli, F., et al. 2023, A&A, 674, A35

Hanuš, J., Viikinkoski, M., Marchis, F., et al. 2017, A&A, 601, A114

Izidoro, A., Dasgupta, R., Raymond, S. N., et al. 2022, Nat. Astron., 6, 357

Johansen, A., Oishi, J. S., Mac Low, M.-M., et al. 2007, Nature, 448, 1022

Klahr, H., & Schreiber, A. 2020, ApJ, 901, 54

Klahr, H., Delbo, M., & Gerbig, K. 2022, in Vesta and Ceres. Insights from the Dawn Mission for the Origin of the Solar System (NASA), 199

Knežević, Z. 2017, Serbian Astron. J., 194, 1

Knezevic, Z., & Milani, A. 2012, IAU Joint Discussion, P18

Mainzer, A. K., Bauer, J. M., Cutri, R. M., et al. 2019, NASA Planetary Data System

Marchi, S., Paolicchi, P., Lazzarin, M., & Magrin, S. 2006, AJ, 131, 1138

Masiero, J. R., Mainzer, A., Grav, T., et al. 2011, ApJ, 741, 68

Masiero, J. R., Mainzer, A., Bauer, J., et al. 2013, ApJ, 770, 7

Masiero, J. R., DeMeo, F. E., Kasuga, T., & Parker, A. H. 2015, Asteroids IV (Tucson: Unversity of Arizona Press), 32, 3

Milani, A., Cellino, A., Knežević, Z., et al. 2014, Icarus, 239, 46

Milani, A., Knežević, Z., Spoto, F., et al. 2017, Icarus, 288, 240

Morbidelli, A. 2002, Modern Celestial Mechanics: Aspects of Solar System Dynamics (Boca Raton: CRC Press)

Morbidelli, A., Bottke, W. F., Nesvorný, D., & Levison, H. F. 2009, Icarus, 204, 558

Morbidelli, A., Baillie, K., Batygin, K., et al. 2022, Nat. Astron., 6, 72

Nesvornỳ, D., Brož, M., Carruba, V., et al. 2015, Asteroids IV (Tucson: University of Arizona Press), 29, 7

Oszkiewicz, D., Kankiewicz, P., Włodarczyk, I., & Kryszczyńska, A. 2015, A&A, 584, A18

Parker, A., Ivezić, Ž., Jurić, M., et al. 2008, Icarus, 198, 138

Raymond, S. N., & Izidoro, A. 2017, Sci. Adv., 3, e1701138

Ryan, E. L., & Woodward, C. E. 2010, AJ, 140, 933

Spoto, F., Milani, A., & Knežević, Z. 2015, Icarus, 257, 275

Tedesco, E. F., Noah, P. V., Noah, M., & Price, S. D. 2002, AJ, 123, 1056

Tsiganis, K., Gomes, R., Morbidelli, A., & Levison, H. F. 2005, Nature, 435, 459

Tsirvoulis, G., Morbidelli, A., Delbo, M., & Tsiganis, K. 2018, Icarus, 304, 14

Usui, F., Kuroda, D., Müller, T. G., et al. 2011, PASJ, 63, 1117

Vernazza, P., Ferrais, M., Jorda, L., et al. 2021, A&A, 654, A56

Vokrouhlický, D., Brož, M., Bottke, W. F., Nesvorný, D., & Morbidelli, A. 2006, Icarus, 182, 118

Vokrouhlicky, D., Bottke, W. F., Chesley, S. R., Scheeres, D. J., & Statler, T. S. 2015, in Asteroids IV, eds. P. Michel, F. E. DeMeo, & W. F. Bottke (Tucson: University of Arizona Press), 509

Walsh, K. J., Delbó, M., Bottke, W. F., Vokrouhlický, D., & Lauretta, D. S. 2013, Icarus, 225, 283

Zappala, V., Cellino, A., Farinella, P., & Knezevic, Z. 1990, AJ, 100, 2030

Appendix A: Membership of known families

First, while choosing v_c we know the results are going to be sensitive to the number density of particles in the proper element space of the reference synthetic data set. This synthetic data set from Deienno et al. (2021) may be an overestimate. Therefore, we created nine different samplings, beginning from all 18,191

bodies provided in the dataset, and narrowed this value to 182, which is certainly unrealistically few. As explained in section 3.1, for each number density below 1, we randomly remove bodies 100 different ways. Then, for each, we apply the HCM for each central body of the families listed in Table 1. The distributions of the found velocities are reported in Fig. A.1. As expected, the
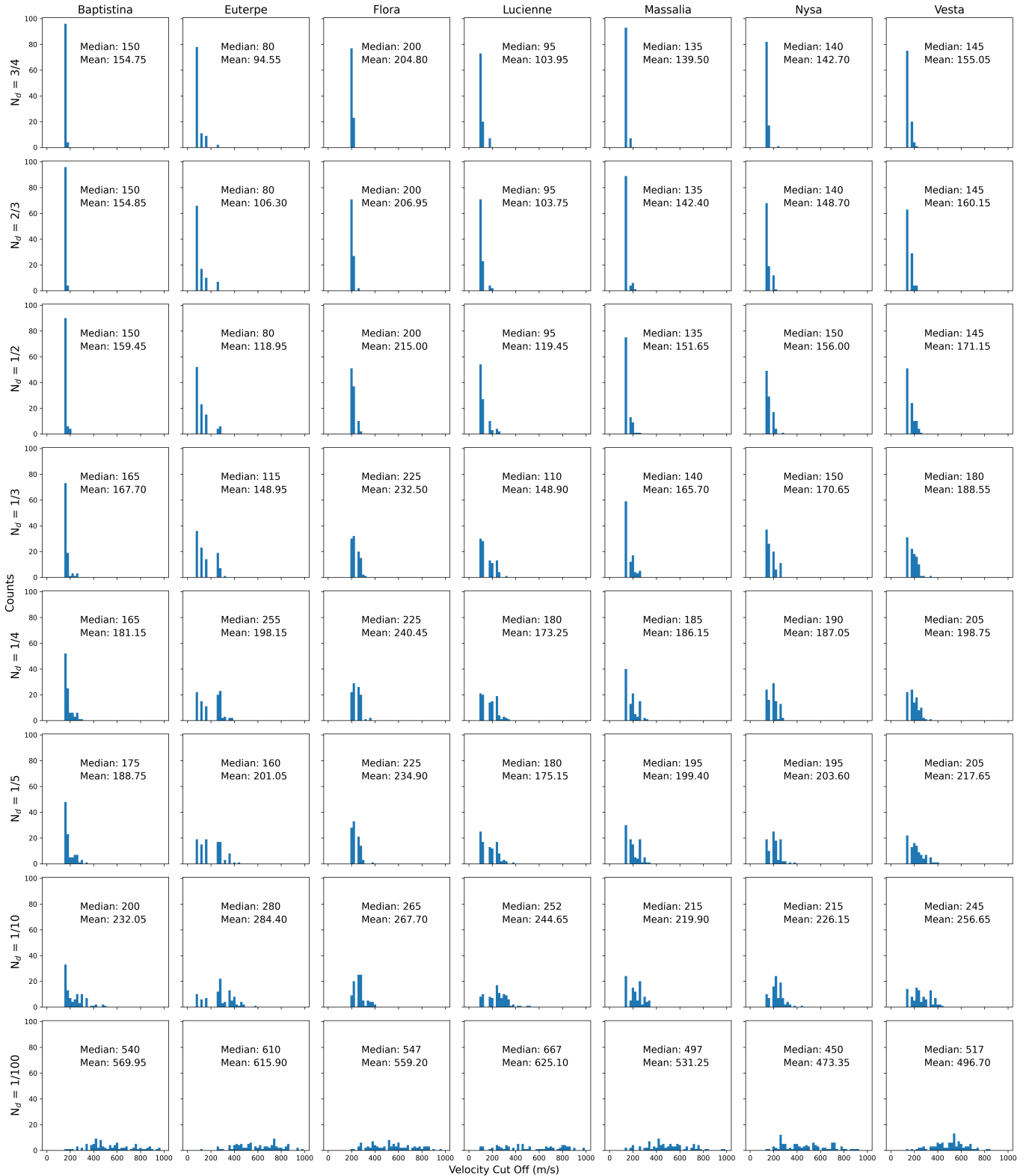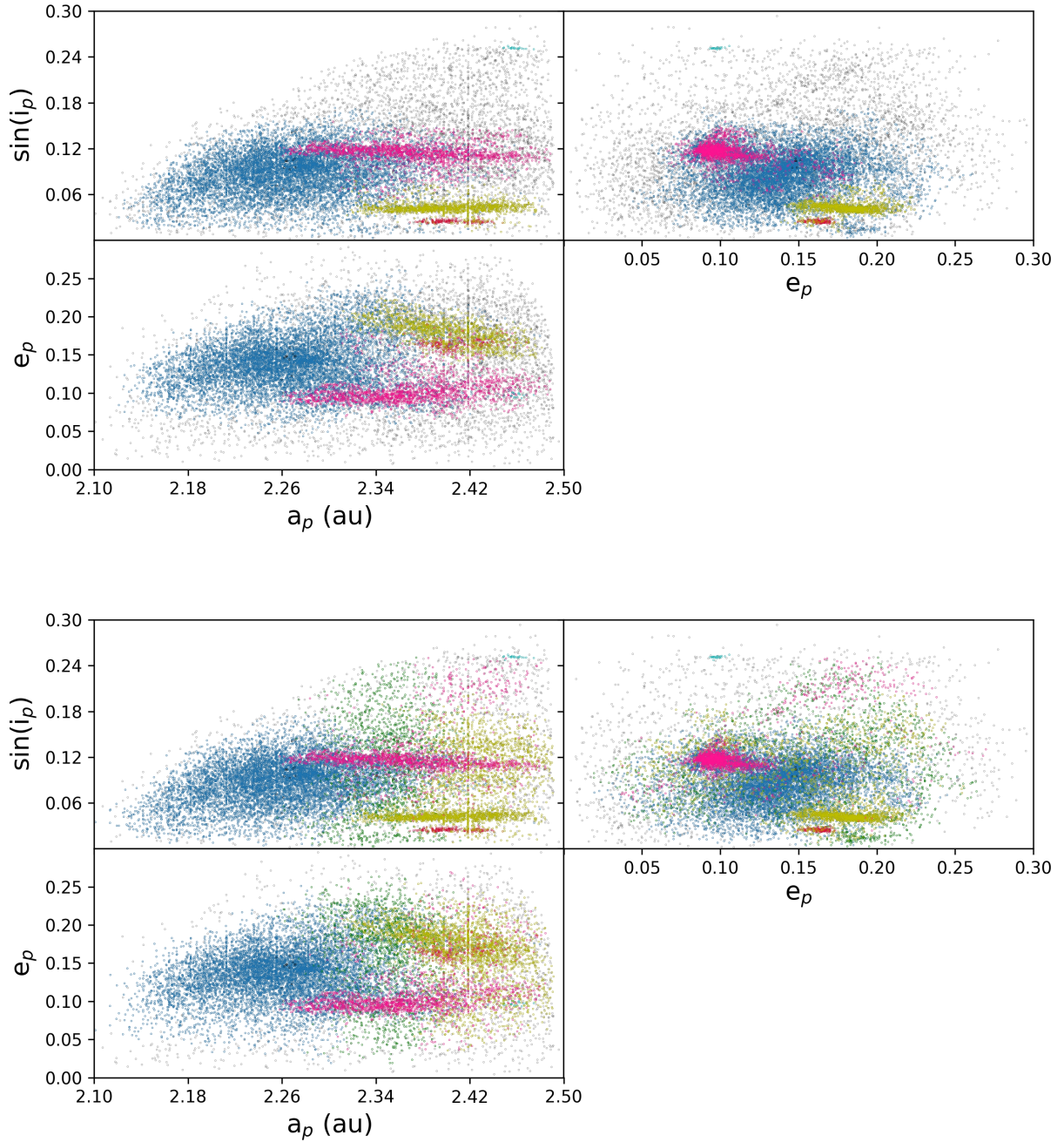

Fig. A.1. Velocity cut-off as a function of the density of the artificial background

Fig. A.2. Membership of each family. The Flora–Baptistina complex is shown in blue, Euterpe in green, Vesta in red, Nysa in yellow, Massalia in purple, and Lucienne in cyan. Background objects are in gray. The top panel uses the v_c values defined using a synthetic background of 3/4 from Table 1, while the bottom panel uses 1/10.

chosen v_c increases as the density decreases. This is always true for the mean and generally true for the median. As seen in the last column, where the density if 1/100, the values are rather spread and no longer consistent with the higher-density background.

The main goal of our experiment is to remove all bodies that are members of collisional families. We were not concerned with providing precise membership. For instance, in our method, a body can be assigned to multiple families, as highlighted by the part of Vesta assigned captured in Baptistina in Fig. 5. Despite this, we were interested in providing a membership list anyway. To do so, we created a matrix where each row was an asteroid, each column was a family, and each element was the v_c at which the body is assigned to the corresponding family. Then,

we subtracted each column by the v_c at which the family began clustering the second body. After this, we assigned each body to the family based on the minimum value. For instance, if body X is assigned to Flora at 50 m s^{-1} and Vesta at 40 m s^{-1}, the body in the end will be assigned to Vesta. These results are displayed in Fig. A.2.

Originally, we did not normalize to the velocity at which the second body is assigned to the cluster. However, membership was not well assigned, especially in the case of Nysa and Massalia, which became confused with one another. Shifting by the second body gathers more of the essence of the cluster's center rather than the central body that takes the families name.

One suboptimal result is that this method is not able to parse Baptistina from Flora. Only four bodies are assigned to Baptistina. Thus, we believe separating these families beyond the cores is more involved and future work is needed to tackle this problem.

Appendix B: Membership of planetesimals and ancient family members

The lists of the ancient family, planetesimals, and the bodies that could belong to either are presented in Tables B.1, 1, and B.3, respectively. Additionally, these objects are labeled with red, magenta, and yellow markers in the left panel of Fig. 7. Lastly, these are the bodies that were assigned to each group at a 1σ significance level, as described in section 3.2. The distribution of the ancient family members is shown in Fig. B.1.

Table B.1. List of objects within the lobes of the V-shape used to identify the ancient family.

(30) Urania	(470) Kilia	(1365) Henyey
(60) Echo	(584) Semiramis	(1375) Alfreda
(115) Thyra	(753) Tiflis	(1419) Danzig
(161) Athor	(822) Lalage	(1594) Danjon
(169) Zelia	(889) Erynia	(1629) Pecker
(172) Baucis	(896) Sphinx	(1643) Brown
(186) Celuta	(939) Isberga	(1689) Floris-Jan
(219) Thusnelda	(1063) Aquilegia	(1797) Schaumasse
(234) Barbara	(1078) Mentha	(1988) Delores
(287) Nephthys	(1117) Reginita	(2013) Tucapel
(299) Thora	(1137) Raissa	(2159) Kukkamaki
(306) Unitas	(1147) Stavropolis	(2286) Fesenkov
(337) Devosa	(1155) Aenna	(3385) Bronnina
(432) Pythia	(1224) Fantasia	(5676) Voltaire

Notes. These bodies are those of Fig. 7. Mass and diameter estimates are available for: 30, 60, 115, 172, 287, 337, 432, 584, and 939 from Vernazza et al. (2021), which were used to calculate a density and in turn obtain one of the age estimations.

**Table B.2.** List of nominal planetesimals.

(6) Hebe	(79) Eurynome	(192) Nausikaa
(7) Iris	(80) Sappho	(198) Ampella
(9) Metis	(118) Peitho	(317) Roxane
(11) Parthenope	(126) Velleda	(364) Isara
(12) Victoria	(131) Vala	(376) Geometria
(17) Thetis	(135) Hertha	(556) Phyllis
(18) Melpomene	(136) Austria	(722) Frieda
(21) Lutetia	(138) Tolosa	(813) Baumeia
(40) Harmonia	(149) Medusa	(857) Glasenappia
(42) Isis	(178) Belisana	(1182) Ilona
(63) Ausonia	(189) Phthia	(2616) Lesya
		(2675) Tolkien

Notes. These asteroids lie below the V-shape in Fig. 7.

Table B.3. Asteroids that are either planetesimals or ancient family members

(230) Athamantis	(525) Adelaide	(548) Kressida
(749) Malzovia	(1396) Outeniqua	

Notes. These objects are displayed as yellow squares with black edges in Fig. 7.
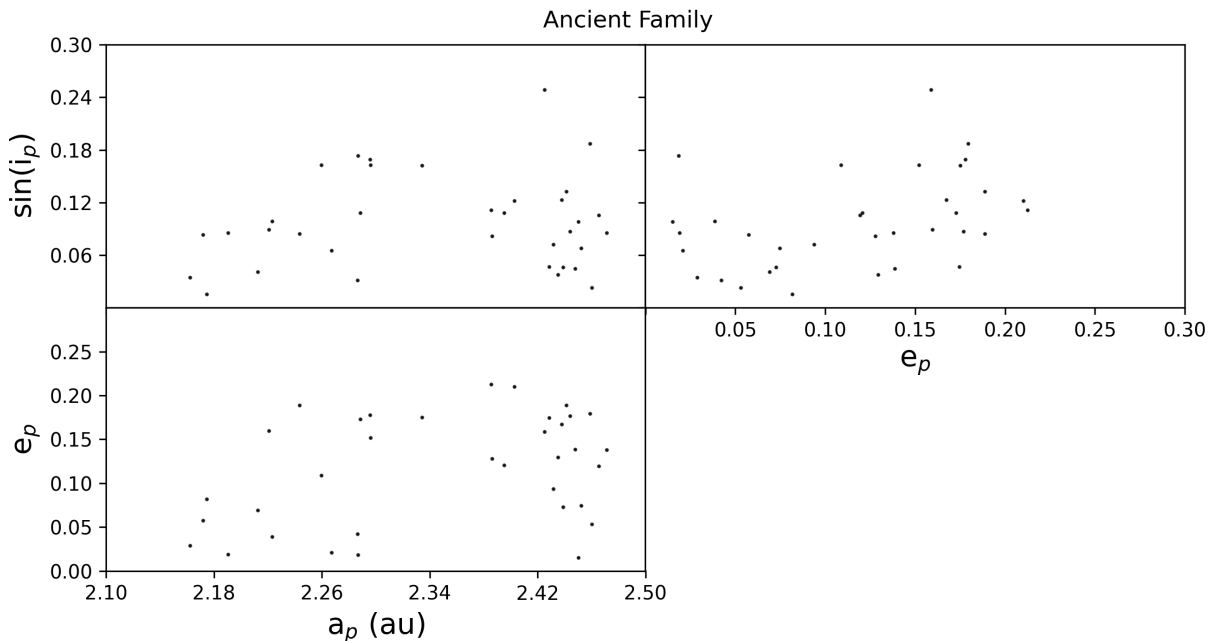
Fig. B.1. Proper elements of the ancient family members reported in Table B.1.

Table B.4. Complete list of planetesimals used to create the SFD in the bottom panel of Fig. 10. These objects come from Table B.2 as well as the planetesimals from Delbo et al. (2017) and Delbo et al. (2019).

|					
(4) Vesta	(20) Massalia	(80) Sappho	(178) Belisana	(336) Lacadiera	(689) Zita
(6) Hebe	(21) Lutetia	(118) Peitho	(189) Phthia	(337) Devosa	(722) Frieda
(7) Iris	(27) Euterpe	(126) Velleda	(192) Nausikaa	(345) Tercidina	(813) Baumeia
(8) Flora	(40) Harmonia	(131) Vala	(198) Ampella	(364) Isara	(857) Glasenappia
(9) Metis	(42) Isis	(135) Hertha	(207) Hedda	(376) Geometria	(1182) Ilona
(11) Parthenope	(51) Nemausa	(136) Austria	(261) Prymno	(435) Ella	(1892) Lucienne
(12) Victoria	(63) Ausonia	(138) Tolosa	(298) Baptistina	(556) Phyllis	(2616) Lesya
(17) Thetis	(72) Feronia	(149) Medusa	(317) Roxane	(572) Rebekka	(2675) Tolkien
(18) Melpomene	(79) Eurynome	(161) Athor	(326) Tamara	(654) Zelinda	(13977) Frisch

Appendix C: Planetesimal counts compensation

As shown in Fig. C.1, the identification method for finding the planetesimals has a bias. Namely, no bodies can be found within V-shapes, yet *a priori* there is no reason that planetesimals cannot exist within these areas. Though we cannot identify which of the asteroids within the V are likely planetesimals with the dynamical and size information alone, we can estimate how many we could expect to exist within the V-shapes. In turn, we can increase the number of counts on a per asteroid basis by taking the ratio of the total linear length, in semi-major axis, divided by the available search length, at a given asteroid diameter. The available search length exists outside the V. The total search length is always equal to 0.4 au, which is the limit of the inner main belt between 2.1 and 2.5 au. This, of course, is also equal to the sum of lengths interior and exterior to the V. For example, for the asteroid in Fig. C.1, this equates to 3.6.

If computing the SFD directly, we could simply increase the weight of each asteroid in the cumulative sum from 1 to the compensation value. However, the dynamical and collisional loss models requires that we work with equally weighted asteroids. To circumvent this problem, we generate a number of asteroids corresponding to the suggested compensation on a per-asteroid basis in a Monte Carlo fashion and append these generated asteroids to the total number of asteroids in Table B.4. To specify, and in the case of the asteroid in Fig. C.1, we append the list by two more asteroids, such that our count for this diameter reaches three. Next, we generate a random number from a uniform probability density distribution between 0 and 1. If this number is less than 0.6, then we add a third asteroid to the list. For each of these new bodies, in order to avoid duplicates, we give them new diameters by Gaussian sampling the diameter of the corresponding planetesimal given its uncertainty. This process is repeated for all of the asteroids. We also perform this random sampling 50 times. In each iteration, we typically add between 12 and 25 bodies. In general, the large objects have compensation values on the order of 1.03, which leads to small probabilities of adding large bodies; most of the new bodies have small diameters.

On a separate note, the V-compensation was applied according to the V-shape of the ancient family that was used to isolate the planetesimal. For instance, all of the planetesimals from Table B.2 use the V shape of the family identified in this work, whereas the planetesimals identified in Delbo et al. (2017) and Delbo et al. (2019) used the V-shape of the family used in those respective studies. For the parent bodies of asteroid families that were added, such as Flora, Vesta, Euterpe, and so on, no V-compensation was performed because they were not identified with this method.

Lastly, we plot all 50 iterations of the resulting SFD in the bottom panel of Fig. C.1. Interestingly, the variance of possible
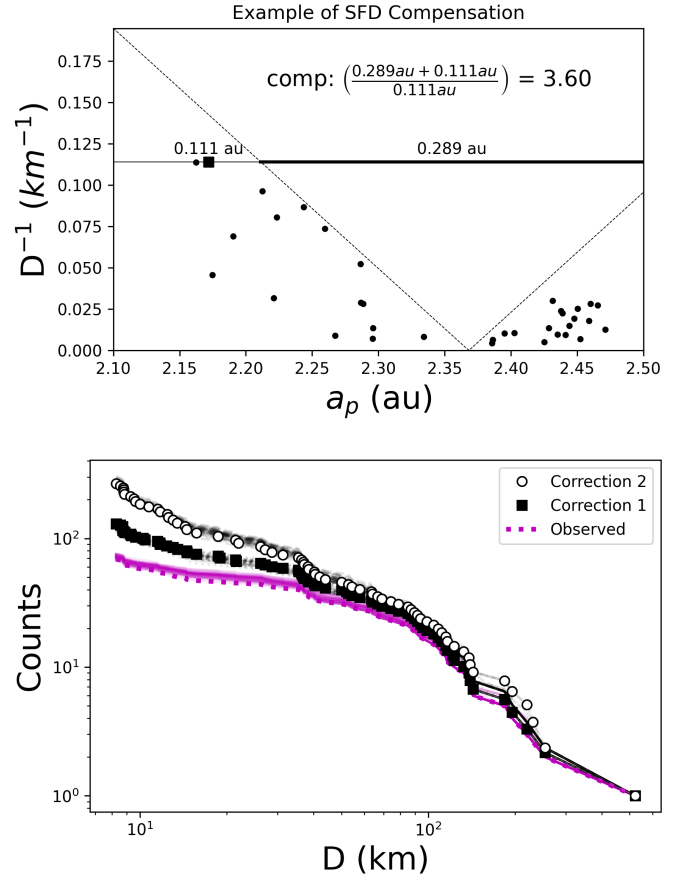
Fig. C.1. *Top:* Demonstration of compensating the number of planetesimal counts missed due to the ancient family covering a large portion of the search area. Here, we show the V-shape corresponding to the newly detected family in this work. The black points are the identified planetesimals from Table B.2. The black square is an example case, whose count is increased in proportion to the amount of search area that was obscured by the family's V-shape. Here, the counts will increase from one to three, with a 60% chance of being increased to four. See the text for more details. *Bottom:* Compensation was applied 50 times with Monte Carlo-style sampling and all iterations are over-plotted with highly transparent lines.

SFDs resulting from considering the Monte Carlo-style generation of new asteroids while considering the uncertainties of the diameters is much less than the variance from modeling the uncertainty used to correct the dynamical loss. For this reason, in the main plot of Fig. 10, where we fit the power-law slopes, we only present one iteration for clarity.