Open Research Online

The Open University's repository of research publications
and other research outputs

The size distribution of Jupiter Family comet nuclei

Journal Item

For guidance on citations see FAQs.

oro.open.ac.uk

The size distribution of Jupiter Family comet nuclei

C. Snodgrass[1,2]⋆, A. Fitzsimmons[3], S. C. Lowry[4] and P. Weissman[5]

[1]*Max Planck Institute for Solar System Research, Max-Planck-Str. 2, 37191 Katlenburg-Lindau, Germany*
[2]*European Southern Observatory, Alonso de Cordova 3107, Vitacura, Santiago, Chile*
[3]*Astrophysics Research Centre, Queen's University, Belfast, UK*
[4]*Centre for Astrophysics and Planetary Science, The University of Kent, Canterbury, UK*
[5]*Jet Propulsion Laboratory, Pasadena, California, USA*

ABSTRACT

We present an updated cumulative size distribution (CSD) for Jupiter Family comet (JFC) nuclei, including a rigourous assessment of the uncertainty on the slope of the CSD. The CSD is expressed as a power law, $N(> r_{\rm N}) \propto r_{\rm N}^{-q}$, where $r_{\rm N}$ is the radius of the nuclei and q is the slope. We include a large number of optical observations published by ourselves and others since the comprehensive review in the *Comets II* book (Lamy et al. 2004), and make use of an improved fitting method. We assess the uncertainty on the CSD due to all of the unknowns and uncertainties involved (photometric uncertainty, assumed phase function, albedo and shape of the nucleus) by means of Monte Carlo simulations. In order to do this we also briefly review the current measurements of these parameters for JFCs. Our final CSD has a slope $q = 1.92 \pm 0.20$ for nuclei with radius $r_{\rm N} \geqslant 1.25$ km.

Key words: comets.

1 INTRODUCTION

The size distribution of any population of solar system small bodies is of critical importance in constraining their formation and subsequent collisional evolution. Dohnanyi (1969) showed that a collisionally relaxed population of self-similar bodies, with the same strength per unit mass, has a characteristic power law size distribution with a slope of 2.5. For a collisional population of gravity controlled (strengthless) bodies O'Brien & Greenberg (2003) demonstrated that the expected size distribution has a shallower slope, 2.04. Happily, the size distribution is also one of the more straightforward characteristics of the population to determine, as at least reasonable estimates of the size of bodies can be made with snap-shot observations. Time-series data allow better measurements of their sizes, as this removes uncertainties due to their rotational light curve. For Jupiter Family comets (JFCs) these observations are generally made when the comet is at a large distance from the Sun, and therefore more likely to be inactive, so a brightness measurement for the bare nucleus can be made. Converting the measured optical magnitude to the size of an object depends on its albedo and phase function. There are only 7 JFCs for which both of these are independently well measured (2P/Encke, 9P/Tempel 1, 10P/Tempel 2, 19P/Borrelly, 28P/Neujmin 1, 67P/Churyumov-Gerasimenko and 81P/Wild 2; three of

which have a size measurement from resolved imaging by spacecraft [see Tables 3 and 4 for references]). A few more comets have measurements of one or the other. Where these parameters are not known, we are forced to assume values for them; typically 4% for the albedo and a linear phase function of $\beta = 0.035$ mag deg^{-1}. Previous measurements of the size distribution of JFCs have followed similar assumptions.

The distribution is generally plotted (fig. 1) as a cumulative size distribution (CSD), expressed in terms

$$N(> r_{\rm N}) \propto r_{\rm N}^{-q} \qquad (1)$$

N is the number of nuclei with radius, $r_{\rm N}$, larger than $r_{\rm N}$, and q is the slope. A number of measurements of these distribution co-efficients have been made, generally based on snap-shot observations of a large number of nuclei. Lowry et al. (2003) estimated $q = 1.6 \pm 0.1$ based on a sample of 33 comets, and Weissman & Lowry (2003) updated this to $q = 1.59 \pm 0.03$, based on 41 JFCs with $r_{\rm N} \geqslant 1.4$ km. They chose to fit the size distribution to only those nuclei with $r_{\rm N} \geqslant 1.4$ km as the slope of the CSD is approximately constant above this radius, while below it there is a sharp cut-off. This break may imply a relative paucity of small nuclei compared with the expected number from a continuation of the power law from larger sizes; Meech et al. (2004) and Fitzsimmons et al. (2011) show that such a break is most likely real and not due to observational biases (as suggested by Lamy et al. (2004)) by modelling realistic observational surveys. A more recent size distribution from Weiss-

Table 1. Previous JFC size distribution estimates.

| Reference	$N_{\mathrm{fit}}/N_{\mathrm{tot}}$	r_{N} range	q
Fernández et al. (1999)	12/64	$2.1 - 3.3$	2.7 ± 0.3
Lowry et al. (2003)	16/19	$1.4 - 3.6$	1.6 ± 0.1
Weissman & Lowry (2003)	41/54	$1.4 - 15$	1.59 ± 0.03
Meech et al. (2004)	38/48	$1.0 - 10$	1.45 ± 0.05
Meech et al. (2004)	21/48	$2.0 - 5$	1.91 ± 0.06
Lamy et al. (2004)	29/65	$1.6 - 15$	1.9 ± 0.3
Weissman et al. (2009)	41/67	$1.4 - 6.0$	1.94 ± 0.07
Tancredi et al. (2006)	32/72	$1.7 - 4.5$	2.7 ± 0.3

Figure 1. Reference CSD with our data set and the usual assumptions. This shows the normal shape of the CSD, with a linear part and a cut-off. The error bars shown are calculated using the Kaplan-Meier statistic. Each fit within a MC run produces a CSD like this with a slightly different distribution of points.

man et al. (2009) has a value of $q = 1.94 \pm 0.07$ based on 41 JFCs with $r_{\mathrm{N}} > 1.4$ km. Another estimate comes from Fernández et al. (1999), who used selected data in quality classes 1-3 (uncertainties on $m_V(1,1,0)$ up to ± 1 mag.) from the catalogue presented by Tancredi et al. (2000), with cut-offs in both absolute magnitude and perihelion distance. The discrepancy between Fernández et al.'s estimate of $q = 2.65 \pm 0.25$ and those of Lowry et al. can be explained by these cut-offs (leaving only 12 comets) and the large uncertainties on magnitudes in Tancredi et al.'s catalogue. Tancredi et al. (2006) presented an updated catalogue and find $q = 2.7 \pm 0.3$ for $r_{\mathrm{N}} \geqslant 1.5$ km. Meech et al. (2004) estimate $q = 1.45 \pm 0.05$ over the range $1 \leqslant r_{\mathrm{N}} \leqslant 10$ km, and a steeper $q = 1.91 \pm 0.06$ in the range $2 \leqslant r_{\mathrm{N}} \leqslant 5$ km, showing the large dependence on the choice of size range. Hicks et al. (2007) estimate $q = 1.50 \pm 0.08$ from Near Earth Asteroid Tracking (NEAT) survey data, although this makes use of observations of active comets and a coma subtraction technique, so is not included with the other surveys of distant inactive comets in Table 1. Finally, Lamy et al. (2004) collated the data from most of these catalogues, together with their own unpublished results, those from Licandro et al. (2000) and also from other papers on individual comets, to calculate $q = 1.9 \pm 0.3$ for JFC nuclei with $r_{\mathrm{N}} \geqslant 1.6$ km.

These previous estimates are listed in Table 1, where we list the number of comets included in the fit (and the total number considered in the survey), the range in sizes over which the authors fit the linear part of the CSD, and the resulting slope q. The uncertainty on q is that quoted by each author, and is generally the formal uncertainty from a least squares (or similar) fit to the line, despite the fact that technically one cannot use such a fitting technique as the points in a cumulative distribution are not independent. These previous works also make no attempt to assess the uncertainty on q contributed by the assumptions on phase function and albedo or the uncertainty from the photometry. In this paper we present an updated size distribution based on new photometry of distant JFC nuclei, using a censored data analysis technique to produce the CSD, and make a rigourous assessment of the uncertainty on q due to all the unknown factors.

2 METHOD

2.1 Censored data analysis

The majority of previous measurements of the comet nucleus CSD follow a similar procedure. For each nucleus a
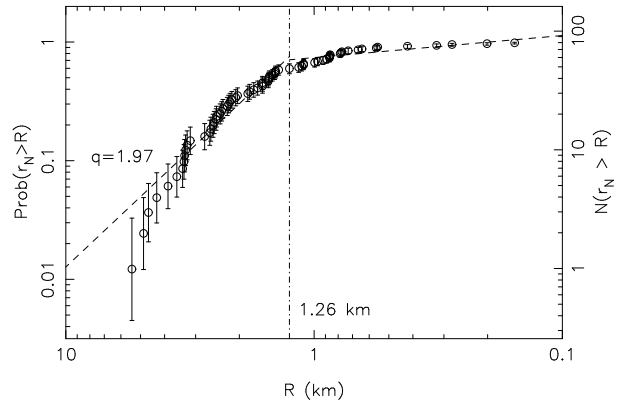
best estimate of the intrinsic radius is arrived at either from the investigators own data or though consideration of several reported detections. The resulting radii are combined to form a CSD and a straight line is fit through the CSD using linear regression. Although this is relatively straightforward, there are intrinsic problems to this method. Aside from the incorrect use of slope uncertainties from linear regression techniques mentioned above, a more important drawback is that it ignores non-detections or upper limits to nuclear radii. Not only is this is an important source of information, ignoring these non-detections can act to instill significant bias. For example, consider a single imaging survey of inactive nuclei with a well defined limiting magnitude for detection. Nearby comets might all be detected, but the fraction of comets observed would decrease as their geocentric and heliocentric distances increased. The resulting CSD of nuclei would contain contain a progressively smaller fraction of the true population at smaller sizes, giving an observed CSD with a shallower power-law distribution than in reality. In the past, breaks in observed luminosity distributions have often been interpreted as due to the onset of such incompleteness. To our knowledge, the only previous study that has accounted for upper limits was that of Meech et al. (2004).

Many fields of astronomical research face a similar problem of censored data, where a sample contains both directly measured values and upper limits. The combination of observed and censored data was first tackled by Avni et al. (1980) via a maximum likelihood method. Established statistical methods as applied to censored astronomical datasets were subsequently described by Schmitt (1985) and Feigelson & Nelson (1985). These authors showed how the cumulative distribution function of a censored dataset can be constrained via the Kaplan-Meier estimator (Kaplan & Meier 1958). This statistic normally estimates the probability of measuring a value for an observable less than or equal to a given value at the non-censored values of the dataset, or in our case for radii of R km, $P(r_N \leqslant R)$. As long as the observed sample is randomly picked from the population, then this will be equivalent to estimating the relative number of comets with radii $\leqslant R$. Inverting this function then gives the normal form of the CSD.

For cometary nuclei at large heliocentric distances the probability of non-detection will be larger as nucleus size decreases, hence at first sight the sample will be biased towards large nuclei. However, other effects act to randomise the censoring. First, comets are observed at various positions in their orbits, which when combined with different sized telescopes, imaging cameras and exposure times leads to a wide variation in effective detection limits. As an example, a 600-second exposure at R-band with the 4.2m William Herschel Telescope can detect a point source at $m_R = 24.5$ at a signal-to-noise of ~ 6. Assuming an optical albedo of 0.04, this allows direct detection of a $r_N = 0.8$km nucleus at opposition at $R_h = 5.2$ AU, or the same nucleus at opposition at $\Delta = R_h = 4$AU but at a phase angle of $\simeq 20°$. Second, when comets become active the shielding effect of their dust comae can result in censoring of even large cometary nuclei. As the strength and onset of activity appears to strongly vary from comet to comet, this further randomises whether an inert nucleus will be directly imaged or not. In this study we therefore assume that the observed cometary nuclei approximates a randomised sample of the JFC population.

The use of the Kaplan-Meier statistic is accurate as long as the quantity being measured and the censoring variable (here the nuclear radius and the limiting magnitude) are independent. The limiting magnitude is set by the instrumental setup, exposure times, and activity level of the comet, while the individual comets surveyed will normally depend on their orbital visibility at a particular date. Therefore this requirement should be true of most cometary observations where non-detection is due to large distance leading to a faint apparent magnitude. For nucleus measurements censored by activity, it will also hold. Although weak correlations of active area with orbital parameters have previously been suspected (A'Hearn et al. 1995), there appears to be no link between nuclear radius and orbit (Lamy et al. 2004), and the orbital positions of comets are randomised by the dates of observation. Finally, an important aspect of the Kaplan-Meier statistic is that in principle it allows an estimate of the uncertainty of the luminosity function at each value where an uncensored measurement is made. One method of calculating this is given in equations (11) and (12) of Feigelson & Nelson (1985), which we adopt in this work. We therefore plot the CSD in terms of $P(r_N > R)$ with uncertainties at each point from this statistic, which is equivalent to $N(r_N > R)$ and gives the slope q we are interested in. Fitzsimmons et al. (2011) give more details on this method and its application to comet size distributions.

2.2 Fitting the CSD

We assume that the CSD is well described by two power laws (i.e. straight lines on a log-log plot), as we expect a break in the power law at small sizes. The power law describing the CSD at larger sizes is the one of primary interest, and is the one we refer to when discussing the slope q elsewhere in this paper, as the data at smaller sizes suffers more from observational incompleteness. The choice of radius to define this break is important, as some of the variation of slopes found by the various papers listed in Table 1 can be explained by the different radius ranges over which the authors chose to fit the CSD. We attempt to independently choose the correct break radius by testing all possible radii in the range
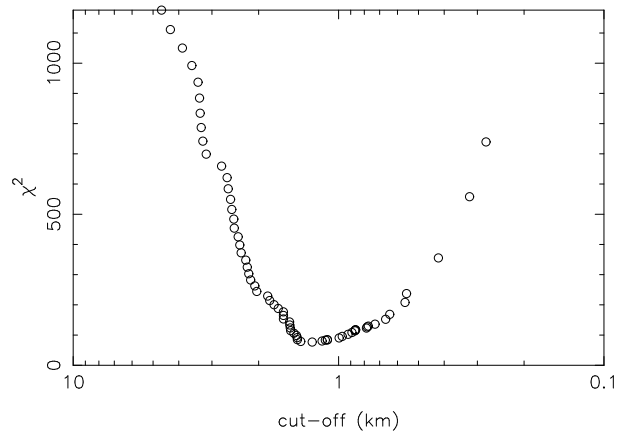

Figure 2. Combined χ^2 for two power law fits either side of a cut-off, as a function of the selected cut-off radius, for the reference CSD shown in fig. 1. We select the cut-off radius with the minimum χ^2 as the best description of the overall shape of the CSD, and therefore the corresponding q value from the fit to the CSD at larger radii as the best slope for that particular CSD. In this case, the minimum is at a radius of 1.26 km.

of the comet sizes. We fit two straight lines on either side of the break, and compute the χ^2 statistic for the fit of this model to the CSD, for each test radius (fig. 2). The minimum $\chi^2(r)$ then gives us the optimum break radius, which tends to fall around 1.2 km.

While strictly speaking a χ^2 fit is not appropriate in this case, as a cumulative distribution does not have independent points, our goal is to find a simple description of the fit that gives a reasonable measure in an individual case. The formal uncertainty on each fit is not used, since we find the overall uncertainty from the standard deviation around the average of all fits. We investigated the results from a number of fitting techniques, to be sure that the result from the χ^2 technique is representative. We performed these tests on our reference size distribution (fig. 1; see section 3). The χ^2 minimisation technique gives a value of $q = 1.97$ for this data set.

One common approach to fitting a line to a cumulative distribution is to use a maximum-likelihood fit. Using the formulae given by Badenes et al. (2010) for the maximum likelihood power law index, we find $q = 1.95 \pm 0.29$, a nearly identical result to the χ^2 method.

An alternative, as used by Lamy et al. (2004), is to test the observed distribution against various model distributions using a Kolmogorov-Smirnov test. In this case the best fit is the one for which the probability that the samples are drawn from the same distribution is the highest. Applying this approach to our test size distribution returns a best fit value of $q = 2.03$ with a probability P_{KS} of 0.99, for the same cut off radius. This method has a significant draw back though, as the Kolmogorov-Smirnov test is not valid when the parameters of the model (in this case the cut-off radius and normalisation of the power law) are estimated from the data, which is necessary in this work.

A further disadvantage with both the maximum-likelihood and K-S techniques is that they fail to take advantage of the fact that the K-M statistics return uncertainties on each radius, which include information from upper limits and should be taken into account in an ideal fit. On the other

hand, while the χ^2 fit is not appropriate due to the fact the points are not independent, the use of K-M statistics mean that the error bars give proper weighting to each point. In conclusion, while none of these statistical methods are fully appropriate to these data, the fact that they all result in a similar value for q gives us confidence that the result from χ^2 is a reasonable estimate of the power-law exponent for the size distribution over this size range.

2.3 Monte Carlo techniques

The effective radius of the nucleus r_N is given by

$$A_R r_N^2 = 2.238 \times 10^{22} R_h^2 \Delta^2 10^{0.4(m_\odot - m_R + \beta\alpha)} \qquad (2)$$

where A_R is the geometric albedo, R_h, Δ and α are the heliocentric and geocentric distances (AU) and phase angle (degrees) at the time of observation, β is the linear phase coefficient (mag deg^{-1}) and m_R and $m_\odot = -27.09$ are the apparent magnitudes of the comet and the Sun, both in the R-band (Russell 1916). The orbital position parameters (R_h, Δ, α) for a given observation and the magnitude of the Sun are known to a high level of precision, but the remaining parameters all contribute to the uncertainty on the reported radius. In addition, this equation converts a cross-sectional area (found as the area is directly proportional to the reflected flux) to an effective radius assuming a spherical nucleus, while the shape is also generally unknown (see section 4.4). We examine the effect of each of these sources of uncertainty by applying a Monte Carlo (MC) technique. For each comet observation in our data set (Table 2) we select a random value for each of the unknown parameters (β, A_R) from a distribution of possible values and a magnitude within the range $m_R \pm \sigma_R$, and use this to generate a radius. We correct this radius for the effects of the non-spherical shape of the nucleus by selecting an ellipsoidal shape from a suitable distribution of axial ratios. We then generate a CSD by fitting the resulting radii as described in the previous subsections. We repeat this procedure many times (generally, $N = 1000$), selecting a different random value for the parameters for each comet each time, which allows us to measure the effect of varying the parameters within our chosen distributions. We then measure the average value of $< q >$ found from the N different CSDs; the standard deviation on this $\sigma(q)$ then incorporates the uncertainty on the varied parameters. In section 4 we vary each parameter in turn (while keeping the other unknown values fixed at default values) to assess the relative contribution to the total uncertainty of each parameter. Finally we allow all of the parameters to vary simultaneously to measure the overall uncertainty on q.

3 DATA SET

The data used for this investigation are selected from all JFC nucleus photometry found in the literature (only properly calibrated professional data sets), including our own previously published work. The majority come from the surveys listed in the introduction that made their own estimates of the CSD, which were all collated in the review by Lamy et al. (2004). We also include the many observations published in the six years since the publication of that work,

which have added new comets to the database and improved the accuracy of the radius measurements of others (in many cases measuring light curves where previously only snap-shot measurements had been taken). We include the surveys of Lowry & Fitzsimmons (2005), Snodgrass et al. (2005, 2006, 2008), Mazzotta Epifani et al. (2007, 2008) and Weissman et al. (2008), and also papers on individual comets from Fernández et al. (2005) [2P/Encke], Jewitt (2006) [D/1819 W1 (Blanpain)], Snodgrass et al. (2010) [103P/Hartley 2], Tubiana et al. (2008) [67P/Churyumov-Gerasimenko] and Pittichová et al. (2008) [21P/Giacobini-Zinner]. The references for each comet are given in Table 2.

To apply our MC method, it is necessary that we choose one observation at a given R_h, Δ and α for each comet, and use equation 2 to generate the corresponding radius for a chosen set of parameters. For some comets there have been multiple observations at different epochs, and we therefore require suitable criteria to select the 'best' observation. We must do this rather than attempting to take an average value of the radius for the comet from separate observations (as was done by Lamy et al. (2004), for example) as such a combination would require an assumed phase function to correct for the different orbital positions, and the effect of the unknown phase function is one of the parameters we investigate.

The criteria we use are:

(i) Is there a detection of the inactive nucleus? If not, we use the strongest upper limit, and flag the value as a limit (except in the cases where space based observations allowed an accurate coma subtraction).

(ii) Is there a light curve? Where there is a direct nucleus detection we prefer light curves over snap-shots, as this removes some of the uncertainty due to the unknown shape.

(iii) Finally, where we are left with a choice between snapshot observations, we use the result with the smallest photometric uncertainty and smallest phase angle to minimise uncertainty.

The final list of selected observations is given in Table 2. The table also includes the radii calculated under the standard assumptions ($\beta = 0.035$ mag deg^{-1} and $A_R = 0.04$ for all comets; photometric values are precise and correspond directly to the size of a spherical nucleus). This is for reference only as the radius found for each comet varies in each MC run, over a range set by the variable parameters, throughout the rest of the paper. The data set contains 86 comets. 71 are nucleus detections, 15 are upper limits from either active or non-detected comets. 21 have light curves, and another 2 have a constraint on a magnitude range but not a full light curve. 12 have known β (although these are often very rough estimates), 11 have known albedo and 7 have both of these parameters measured.

For reference, fitting a CSD to this data set without varying any parameters (i.e. following the procedure applied to previous works; taking the reference radii based on the standard assumptions listed in Table 2) gives $q = 1.97$, with a cut off at 1.26 km (fig. 1). For comparison, using the measured phase functions and albedos for the comets where these values are known (and using the assumed values given above for other comets) gives $q = 2.00$, and the same cut off. There are relatively few comets with known β and/or A_R (see Tables 3 and 4), and these tend to be the larger comets

Table 2. Photometry database used for input

Comet	$R_{\rm h}$	Δ	α	m_R	σ_R	$\Delta m_R{}^a$	$r_{\rm N}{}^b$	Reference
2P/Encke[c]	3.97	2.97	1.20	19.34	0.00	0.35	4.65	Fernández et al. (2005)
4P/Faye	2.96	2.00	5.60	20.12	0.03	–	1.75	Lamy et al. (2009)
6P/dArrest	2.83	2.19	18.19	20.74	0.07	0.18	1.69	Lowry & Weissman (2003)
7P/Pons-Winnecke	4.69	4.31	11.60	22.46	0.02	0.30	2.24	Snodgrass et al. (2005)
9P/Tempel 1	3.52	4.03	13.30	21.30[d]	0.04	0.50	2.72	Lamy et al. (2007)
10P/Tempel 2	3.99	3.16	9.00	19.66	0.05	0.35	4.86	Jewitt & Luu (1989)
14P/Wolf	5.51	4.96	8.90	22.28	0.01	0.55	3.15	Snodgrass et al. (2005)
17P/Holmes	4.66	3.92	9.00	22.86	0.02	0.30	1.61	Snodgrass et al. (2006)
19P/Borrelly	1.40	0.62	38.00	16.77	0.05	1.00	2.03	Lamy et al. (1998)
21P/Giacobini-Zinner	3.80	3.21	13.60	21.89	0.04	–	1.82	Pittichová et al. (2008)
22P/Kopff	4.49	3.80	9.93	21.23	0.10	0.58	3.25	Lowry & Weissman (2003)
26P/Grigg-Skjellerup	3.82	2.85	5.00	21.60	0.07	–	1.61	Boehnhardt et al. (1999)
28P/Neujmin 1	7.66	6.67	1.53	20.74	0.02	0.45	10.65	Delahodde et al. (2001)
29P/Schwassmann-Wachmann 1	5.89	4.97	4.62	$\geqslant$18.00	0.01	–	$\leqslant$ 22.63	Meech et al. (1993)
31P/Schwassmann-Wachmann 2	4.58	3.69	6.35	21.05	0.06	0.50	3.29	Luu & Jewitt (1992)
36P/Whipple	4.78	3.79	1.00	21.57	0.01	0.70	2.55	Snodgrass et al. (2008)
37P/Forbes	2.27	1.39	14.80	20.83	0.04	–	0.78	Lamy et al. (2009)
40P/Vaisala 1	4.58	3.68	6.40	$\geqslant$22.08	0.02	–	$\leqslant$ 2.05	Snodgrass et al. (2008)
43P/Wolf-Harrington	4.43	3.44	3.80	$\geqslant$21.45	0.06	–	$\leqslant$ 2.38	Lowry & Fitzsimmons (2005)
44P/Reinmuth 2	4.51	3.84	10.10	22.49	0.06	–	1.85	Snodgrass et al. (2008)
45P/Honda-Mrkos-Pajdusakova	5.14	4.17	3.69	23.27	0.86	–	1.44	Lowry et al. (2003)
46P/Wirtanen	4.98	4.00	3.50	25.15	0.15	0.38	0.56	Boehnhardt et al. (2002)
47P/Ashbrook-Jackson	5.42	4.48	3.50	21.68	0.01	0.45	3.39	Snodgrass et al. (2006)
48P/Johnson	3.87	2.94	6.10	21.11	0.02	0.32	2.14	Jewitt & Sheppard (2004)
49P/Arend-Rigaux	3.34	2.78	16.01	19.51	0.05	–	4.30	Lowry et al. (2003)
50P/Arend	2.37	1.47	11.80	20.58	0.04	–	0.92	Lamy et al. (2009)
51P/Harrington	5.30	5.73	9.22	$\geqslant$23.50	0.00	–	$\leqslant$ 2.01	Lowry & Fitzsimmons (2001)
52P/Harrington-Abell	2.83	1.86	5.50	20.30	0.20	–	1.43	Licandro et al. (2000)
53P/Van Biesbroeck	8.31	7.34	1.38	23.65	0.06	–	3.32	Meech et al. (2004)
54P/de Vico-Swift	5.39	5.70	9.68	$\geqslant$23.20	0.00	–	$\leqslant$ 2.35	Lowry & Fitzsimmons (2001)
56P/Slaughter-Burnham	7.42	7.58	7.41	25.37	0.12	–	1.53	Meech et al. (2004)
57P/duToit-Neujmin-Delporte	5.10	4.35	7.82	$\geqslant$24.00	0.00	–	$\leqslant$ 1.14	Lowry & Fitzsimmons (2001)
59P/Kearns-Kwee	2.52	1.54	3.30	20.86	0.05	–	0.79	Lamy et al. (2009)
61P/Shajn-Schaldach	4.39	3.40	3.14	23.27	0.86	–	1.00	Lowry et al. (2003)
63P/Wild 1	2.27	1.30	9.20	19.17	0.02	–	1.43	Lamy et al. (2009)
64P/Swift-Gehrels	3.63	2.79	9.50	21.60[e]	0.10	–	1.61	Licandro et al. (2000)
65P/Gunn	4.43	3.55	6.36	$\geqslant$17.74	0.04	–	$\leqslant$ 14.09	Lowry & Fitzsimmons (2001)
67P/Churyumov-Gerasimenko	5.60	4.60	0.50	22.46	0.01	0.40	2.39	Tubiana et al. (2008)
69P/Taylor	4.89	3.99	6.10	21.00	0.40	–	3.88	Lowry et al. (1999)
70P/Kojima	4.84	4.26	10.60	22.53	0.15	–	2.18	Snodgrass et al. (2008)
71P/Clark	2.71	1.76	7.50	21.83	0.07	–	0.66	Lamy et al. (2009)
73P/Schwassmann-Wachmann 3	3.03	2.35	15.00	21.57	0.05	–	1.26	Boehnhardt et al. (1999)
74P/Smirnova-Chernykh	3.55	2.58	4.20	20.52	0.10	0.14	2.21	Lamy et al. (2004)
75P/Kohoutek	3.48	2.73	12.36	21.76	0.05	–	1.47	Weissman et al. (2008)
76P/West-Kohoutek-Ikemura	3.09	2.26	12.10	24.39	0.10	0.42	0.32	Lamy et al. (2004)
78P/Gehrels 2	5.46	4.49	3.54	23.44	0.19	–	1.52	Lowry & Weissman (2003)
79P/duToit-Hartley	4.74	4.29	11.50	23.30	0.40	–	1.53	Lowry et al. (1999)
81P/Wild 2	5.29	4.31	2.06	22.13	0.21	–	2.52	Weissman et al. (2008)
82P/Gehrels 3	3.73	2.75	0.60	23.04	0.10	0.51	0.73	Lamy et al. (2004)
84P/Giclas	2.21	1.37	16.90	20.59	0.04	–	0.86	Lamy et al. (2009)
86P/Wild 3	4.95	4.10	6.23	25.00	0.10	–	0.64	Meech et al. (2004)
87P/Bus	2.45	1.43	2.60	22.87	0.10	0.86	0.28	Lamy et al. (2004)
92P/Sanguin	4.46	3.58	6.30	21.94	0.01	0.60	2.07	Snodgrass et al. (2005)
94P/Russell 4	4.14	3.18	5.30	$\geqslant$20.97	0.01	–	$\leqslant$ 2.62	Snodgrass et al. (2008)
97P/Metcalf-Brewington	4.76	4.11	10.11	22.23	0.40	–	2.36	Lowry et al. (2003)
98P/Takamizawa	3.28	2.53	12.70	24.20	0.30	–	0.42	Mazzotta Epifani et al. (2008)

[a] The observed variation in magnitude, corresponding to a minimum axial ratio.
[b] The radius (km) calculated for a *fixed* $\beta = 0.035$ mag deg^{-1} and albedo of 4%, for reference.
[c] We include 2P/Encke, which is an ecliptic comet although not technically a JFC by the Tisserand parameter definition.
[d] From reported V-magnitude and $(V - R) = 0.50 \pm 0.01$ from Li et al. (2007a)
[e] From reported V-magnitude and assumed $(V - R) = 0.50$.

Table 2. Photometry database used for input (continued)

Comet	$R_{\rm h}$	Δ	α	m_R	σ_R	Δm_R	$r_{\rm N}$	Reference
100P/Hartley 1	4.84	4.06	8.20	22.06	0.06	–	2.48	Weissman et al. (2008)
103P/Hartley 2	5.60	4.70	4.50	$\geqslant$24.50	0.04	–	$\leqslant$ 1.02	Snodgrass et al. (2010)
104P/Kowal 2	3.94	3.31	12.66	23.05	0.93	–	1.12	Lowry et al. (2003)
106P/Schuster	1.67	0.78	23.00	18.91	0.03	–	0.89	Lamy et al. (2009)
110P/Hartley 3	3.60	2.66	6.58	20.58	0.10	–	2.33	Weissman et al. (2008)
111P/Helin-Roman-Crockett	3.47	2.53	5.60	21.48	0.05	–	1.39	Mazzotta Epifani et al. (2007)
112P/Urata-Niijima	2.30	1.50	19.20	20.96	0.04	–	0.86	Lamy et al. (2009)
114P/Wiseman-Skiff	3.75	2.95	10.90	22.95	0.11	–	0.97	Snodgrass et al. (2008)
115P/Maury	6.38	5.43	2.56	24.86	0.07	–	1.10	Meech et al. (2004)
116P/Wild 4	4.72	3.82	6.00	$\geqslant$21.36	0.06	–	$\leqslant$ 3.03	Weissman et al. (2008)
117P/Helin-Roman-Alu 1	3.29	2.56	13.50	$\geqslant$16.31	0.03	–	$\leqslant$ 16.36	Mazzotta Epifani et al. (2008)
118P/Shoemaker-Levy 4	4.71	3.72	3.50	21.54	0.20	–	2.61	Lowry et al. (2003)
120P/Mueller 1	3.89	3.05	8.80	23.52	0.13	–	0.77	Snodgrass et al. (2008)
121P/Shoemaker-Holt 2	3.92	3.43	13.50	20.77	0.01	0.15	3.34	Snodgrass et al. (2008)
128P/Shoemaker-Holt 1	4.99	4.16	7.29	22.02	0.09	–	2.63	Lowry & Weissman (2003)
131P/Mueller 2	3.48	2.65	10.10	22.77	0.09	–	0.87	Snodgrass et al. (2008)
136P/Mueller 3	4.83	4.04	8.00	23.70	0.20	–	1.15	Mazzotta Epifani et al. (2008)
137P/Shoemaker-Levy 2	6.95	6.17	5.30	22.86	0.03	–	3.57	Snodgrass et al. (2006)
143P/Kowal-Mrkos	3.40	2.47	8.20	18.52	0.02	–	5.42	Jewitt (2002)
147P/Kushida-Muramatsu	2.83	2.30	18.70	25.48	0.10	0.46	0.20	Lamy et al. (2004)
160P/LINEAR	3.98	3.41	12.50	23.69	0.17	–	0.87	Snodgrass et al. (2008)
179P/Jedicke	5.52	4.71	6.90	22.63	0.07	–	2.47	Snodgrass et al. (2008)
243P/NEAT	3.97	3.68	14.30	22.70	0.30	–	1.52	Mazzotta Epifani et al. (2008)
D/1819 W1 (Blanpain)	1.64	0.72	20.70	22.40	0.10	–	0.16	Jewitt (2006)
P/2001 H5 (NEAT)	4.61	3.62	3.13	23.29	0.11	–	1.10	Lowry & Fitzsimmons (2005)
P/2002 T5 (LINEAR)	5.24	4.46	7.50	$\geqslant$19.10	0.30	–	$\leqslant$ 11.40	Mazzotta Epifani et al. (2008)
P/2003 S1 (NEAT)	3.54	3.05	15.20	$\geqslant$21.37	0.03	–	$\leqslant$ 2.10	Mazzotta Epifani et al. (2008)
P/2004 DO29 (Spacewatch-LINEAR)	4.22	3.33	7.00	$\geqslant$20.13	0.03	–	$\leqslant$ 4.23	Mazzotta Epifani et al. (2008)
P/2004 H2 (Larsen)	3.71	3.04	13.10	$\geqslant$21.59	0.01	–	$\leqslant$ 1.91	Snodgrass et al. (2008)
P/2004 H3 (Larsen)	3.71	2.97	12.00	24.19	0.21	–	0.55	Snodgrass et al. (2008)

which have a weak affect on the CSD, so the difference is not large.

4   VARIATION OF PARAMETERS

4.1   Photometric uncertainty

The first source of uncertainty in the radius, and the most straight forward to account for, is the uncertainty σ_R in the measured magnitude of the comet m_R. This includes the Poisson noise from the CCD photometry, uncertainty from the standard star calibration, etc. The true value can fall anywhere within the Gaussian distribution centred on the reported value, with the standard deviation set by the reported error bar. In most cases $\sigma_R \leqslant 0.1$ mag, and this is not the most significant uncertainty on $r_{\rm N}$. The result of a MC run allowing the comet magnitudes to vary within the Gaussian distributions defined by the reported m_R, σ_R for each, and holding all other parameters fixed at the default values given above, gives $q = 1.93 \pm 0.10$ and an average cut-off of 1.2 km.

4.2   Phase function

For all previous size distributions, a constant Solar phase function has been assumed for all comets. The work presented in this paper was motivated in part by a simple

Table 3. Phase function measurements for JFCs.

Comet	β	Reference.
2P	0.060 ± 0.005	Fernández et al. (2000)
"	0.049 ± 0.004	Boehnhardt et al. (2008)
"	*0.053 ± 0.003*	*Weighted mean*
9P	0.046 ± 0.007	Li et al. (2007a)
10P	0.037 ± 0.004	Sekanina (1991)
19P	0.043 ± 0.009	Li et al. (2007b)
28P	0.025 ± 0.006	Delahodde et al. (2001)
36P	0.060 ± 0.019	Snodgrass et al. (2008)
45P	~ 0.06	Lamy et al. (2004)
47P	0.083 ± 0.006	Snodgrass et al. (2008)
48P	0.059 ± 0.002	Jewitt & Sheppard (2004)
67P	0.076 ± 0.003	Tubiana et al. (2008)
81P	0.0513 ± 0.0002	Li et al. (2009)
143P	0.043 ± 0.001	Jewitt et al. (2003)

demonstration that changing the assumed value changed the resulting value of q: increasing the assumed β increases q (Snodgrass 2006). To investigate the more realistic situation of β varying between comets within some distribution we developed the MC technique presented here. We recreate the earlier result when choosing the simplest distribution; a constant value for all comets.

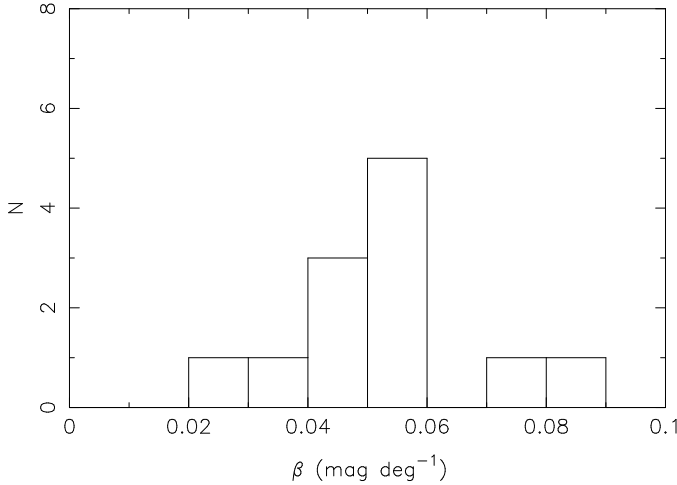To choose an appropriate distribution we briefly review the measured values. Phase functions for comets are diffi-

Figure 3. Distribution of measured phase functions for JFCs.

Table 4. Albedo measurements for JFCs.

Comet	A_R	Reference
2P	0.050 ± 0.030	Fernández et al. (2000)
9P	0.064 ± 0.013	Li et al. (2007a)
"	0.046 ± 0.015	Lisse et al. (2005)
"	0.072 ± 0.016	Fernández et al. (2003)
"	*0.061 ± 0.008*	*Weighted mean*
10P	0.030 ± 0.012	A'Hearn et al. (1989)
19P	0.029 ± 0.006	Buratti et al. (2004)
"	0.072 ± 0.020	Li et al. (2007b)
"	*0.033 ± 0.006*	*Weighted mean*
22P	0.048 ± 0.010	Lamy et al. (2002)
28P	0.03 ± 0.01	Jewitt & Meech (1988)
49P	0.045 ± 0.019	Campins et al. (1995)
67P	0.054 ± 0.006	Kelley et al. (2009)
81P	0.064 ± 0.010	Li et al. (2009)
103P	0.028 ± 0.009^a	Lisse et al. (2009)
162P	0.037 ± 0.014	Fernández et al. (2006)

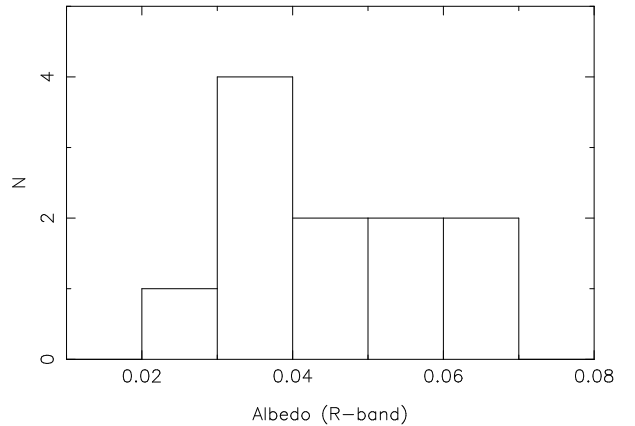[a] Uncertain due to difficulties with the optical photometry, see Snodgrass et al. (2010)

Figure 4. Distribution of observed R-band albedos.

cult to measure, since it is necessary to obtain light curves at multiple epochs to remove rotational variations from the photometry and leave the variation due to changing phase angle. Even relatively well measured phase curves, such as that for 28P/Neujmin 1 (Delahodde et al. 2001), are still quite uncertain, and rough estimates from only a few epochs are very approximate even if formally good fits to the data (e.g. Snodgrass et al. 2008). The best results come from spacecraft data, although we note that spacecraft may measure phase functions to high α, while ground based measurements generally only cover $\alpha \lesssim 15°$. We list all JFCs with estimates in Table 3, and plot the distribution of values in fig. 3. It is noticeable that the measured phase functions are generally steeper than the normally assumed values of $\beta = 0.035$ or 0.04 mag deg^{-1}; the mean value is 0.053 ± 0.016. This is due to a number of recent measurements of steep phase functions for JFCs, while the larger (brighter, easier to observe) nuclei observed previously all had shallower slopes. We tested for any correlation between size of nucleus and phase function, but none exists. The β values are approximately normally distributed around the mean, so we use a Gaussian distribution for input into the MC simulations.

A MC run with a distribution matching this observed one gives $q = 1.95 \pm 0.07$. Increasing the mean β increases q, and decreasing it decreases q, as expected. Changing the width of the distribution also affects the CSD slope, a narrower distribution ($\beta = 0.053 \pm 0.005$) gives a steeper slope and a smaller uncertainty, $q = 1.99 \pm 0.02$, while a wider distribution ($\sigma = 0.025$) gives a shallower slope and larger uncertainty ($q = 1.91 \pm 0.11$).

4.3 Albedo

The conversion from a reflected flux to a radius requires knowledge of the albedo, which generally must be assumed for comets. Under the normal assumption that all comets have the same albedo the CSD slope is directly proportional to the slope of the 'luminosity' (reflected flux) distribution ($q_S = 5q_L$, Irwin et al. (1995)), and so a change of assumed albedo will not affect q. However, this does not hold for a distribution of albedo values that vary between comets, so

we use our MC technique to assess the affect on the CSD of a variable albedo.

The distribution of albedos for JFC nuclei is poorly constrained. Only when the radius can be determined independently of the flux can the albedo be found, and this has been done for very few comets (generally by simultaneous observation in the optical and thermal infrared, where a thermal model of the nucleus returns the size for comparison with the reflected optical flux, or from spacecraft observations). We list all published JFC albedos in Table 4, converting albedos to R-band where necessary, and plot the distribution of R-band values in fig. 4. We note that the values obtained by Li et al. (2007a,b, 2009), from modelling of disk resolved photometry from spacecraft data, are higher than all but one of the other measurements. This suggests some underlying problem in comparing results from this technique with radiometric methods, but we take them at face value for the purposes of this study. The mean value is 0.044 ± 0.013, close to the usually assumed 4%. The small number of known values means that the shape of the distribution if not well defined; we choose to use a Gaussian distribution as this seems reasonable.

For a Gaussian with the observed mean and standard deviation, the CSD has $q = 1.85 \pm 0.15$. For a lower average albedo (2%) we find a shallow slope, $q = 1.64 \pm 0.21$, with a larger uncertainty as some objects in such a distribution will occasionally have $A_R \approx 0$, giving very large sizes and therefore a large spread in slopes. A higher average albedo (6%) gives $q = 1.89 \pm 0.12$, a steeper and less uncertain slope. A narrower distribution ($A_R = 0.04 \pm 0.005$) gives $q = 1.93 \pm 0.09$, while a broad distribution ($A_R = 0.04 \pm 0.02$) gives $q = 1.72 \pm 0.19$.

4.4 Correction for shape

4.4.1 Relating true effective radii and snap-shot radii

Comets are generally found to be non-spherical, and the varying cross-sectional area of a rotating elongated body reflects a changing amount of light, which would imply a different effective radius if a 'snap-shot' measurement was made at different rotational phase. Nuclei are often described as tri-axial ellipsoids, with semi-major axes $a \geqslant b = c$, as this is the simplest non-spherical body. Except for comets that have been visited by spacecraft, there generally is not sufficient data to make a more complex model of the shape. When producing a size distribution of these bodies, we are interested in the effective radius of a sphere with the same volume, given by $r_N = \sqrt[3]{ab^2}$, but what we measure using equation 2 is the radius corresponding to the projected cross sectional area at the time of observation. The projected area of a tri-axial ellipsoid, at rotational phase ϕ and with the pole orientated at ϵ to the line of sight, is given by (Lamy et al. 2004):

$$S(\phi, \epsilon) = \pi ab^2 \sqrt{\left(\frac{\sin^2 \phi}{a^2} + \frac{\cos^2 \phi}{b^2} \right) \sin^2 \epsilon + \frac{\cos^2 \epsilon}{b^2}}. \qquad (3)$$

The radius found from a snap-shot observation r_{ss} is the one corresponding to this area. Substituting $S(\phi, \epsilon) = \pi r_{ss}^2$ and $ab^2 = r_N^3$ into equation 3 we find that the true effective radius is related to the snap-shot radius by:

$$r_N = r_{ss} \left(\frac{a}{b} \right)^{-\frac{1}{6}} \left[\left(\frac{\sin^2 \phi}{\frac{a}{b}^2} + \cos^2 \phi \right) \sin^2 \epsilon + \cos^2 \epsilon \right]^{-\frac{1}{4}} \qquad (4)$$

This equation can be used to find the *average* relation between r_N and r_{ss} by integrating over $0 \leqslant \phi \leqslant 180°$ and $0 \leqslant \epsilon \leqslant 90°$, which gives a correspondence that is a weak function of a/b:

$$r_N = r_{ss} \left(\frac{a}{b} \right)^{-\frac{1}{6}} \sqrt{2} \left[3 + \left(\frac{a}{b} \right)^{-2} \right]^{-\frac{1}{4}} \qquad (5)$$

For an $a/b = 1.5$ this gives $r_N = 0.97 r_{ss}$, and for a more extreme $a/b = 3$ we find $r_N = 0.89 r_{ss}$. This implies that snapshots will, *on average*, give a very reasonable measurement of the effective radius, with the true radius being between 90-100% of the reported radius from a single snap-shot.

For individual measurements this does not hold; sometimes an observation will happen to fall at light curve minimum, and sometimes it will fall at a maximum. Instead of applying a statistical correction to the whole database of snap-shots under the assumption that they will average out, we use MC simulations to generate the 'true' radius of the
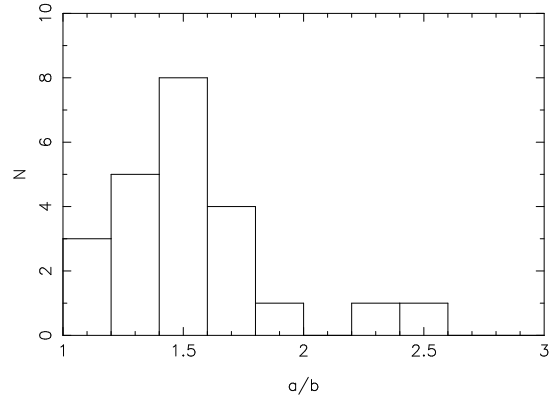

Figure 5. Distribution of observed *minimum* axial ratios for cometary nuclei, from the magnitude ranges listed in table 2.

nucleus from a snap-shot observation using equation 4. We do this by selecting a random shape (a/b) and orientation (ϕ, ϵ) for the nucleus, together with the radius r_{ss} from the snap-shot magnitude (and all other assumed parameters) given by equation 2. Of course the 'true' radius generated this way may not match reality for any individual case, but over the N MC runs this will generate a distribution of the possible effective radii that match the snap-shot photometry. This method has the advantage that it doesn't simply apply a correction for snap-shots that shift the radii up or down uniformly, but produces a family of CSDs describing how the overall distribution is affected by the possible variation in individual comet radii due to shape, allowing us to measure the uncertainty introduced.

The MC method requires that we choose appropriate distributions for the random parameters. For ϕ and ϵ this is trivial; any value in the ranges $0 \leqslant \phi \leqslant 180°$ and $0 \leqslant \epsilon \leqslant 90°$ respectively is equally likely as the rotational phase and pole orientation is entirely random for comets, as torques from jets can change the spin state in any direction on a short timescale. When a light curve is measured, one obtains the magnitude averaged over rotational phase, but ϵ is generally unknown. In this case we follow the same technique, but set $\phi = 0$ and only randomise ϵ. The choice of axial ratio is more difficult, as we need to understand the distribution of nucleus shapes. In the next subsection we discuss the current knowledge of this distribution.

4.4.2 Nucleus shapes

Some constraint can be placed on the shape of an individual nucleus from its light curve, as a *minimum* value of a/b can be found from the amplitude of the light curve Δm:

$$\left(\frac{a}{b} \right)_{min} = \frac{\pi ab}{\pi b^2} = \frac{Area_{max}}{Area_{min}} = \frac{Flux_{max}}{Flux_{min}} = 10^{0.4 \Delta m} \qquad (6)$$

The distribution of minimum axial ratios from the light curves available in our catalogue is shown in figure 5; it is peaked at $a/b \sim 1.5$ and can be reasonably well described as having a decrease in N as a/b increases, with a lack of bodies with $a/b \approx 1$. However, this describes only the observed distribution of minimum axial ratios, not the distribution of actual shapes.
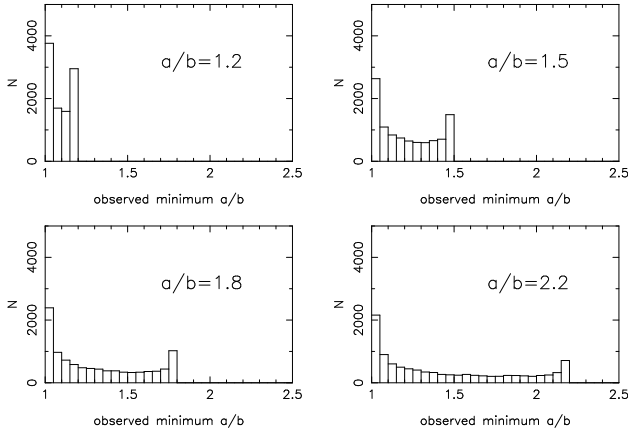
Figure 6. Distributions of apparent minimum a/b for 10,000 'observations' of nuclei with fixed axial-ratio at random ϵ. Input values of true $a/b = 1.2, 1.5, 1.8, 2.2$.

In order to describe how the observed minimum a/b is related to the true axial ratio of the nucleus, we substitute equation 3 into the ratio of areas given in equation 6:

$$\left(\frac{a}{b}\right)_{\min} = \frac{S(\phi=0°)}{S(\phi=90°)} = \left[\left(\frac{a}{b}\right)_{\text{true}}^{-2} \sin^2 \epsilon + \cos^2 \epsilon\right]^{-\frac{1}{2}} \quad (7)$$

This returns the true a/b when the nucleus is viewed from the equator ($\epsilon = 90°$) and no light curve at all (observed $a/b = 1$) from the pole ($\epsilon = 0°$). The form of this equation means that the observed minimum a/b increases slowly from 1 as ϵ increases from 0, rising steeply towards the true value at $\epsilon \gtrsim 60°$ for highly elongated nuclei (see fig. 1 of Lamy et al. (2004)). A histogram of the observed minimum a/b for a sequence of light curves of an object, taken at random values of ϵ, is bimodal. It has a larger peak at $a/b \approx 1$ and a smaller one at the true a/b (fig. 6). We show histograms for four different nuclei with axial ratios of 1.2, 1.5, 1.8 and 2.2. As the true a/b increases, the difference in the relative size of the peaks in the histogram also increases; for nuclei with elongations $\gtrsim 1.5$ a constant $\sim 20\%$ of the 'observed' light curves give a minimum $a/b \approx 1$. The number of observations that return the true axial ratio decreases rapidly as the nucleus becomes more elongated, as the probability of seeing a more elongated body near enough to end on to get an accurate measurement of the true minimum area decreases.

For a collection of nuclei with varying shapes, the distribution of observed minimum a/b from light curves at random ϵ is a combination of these histograms. We modelled this by generating an input distribution of true shapes and finding the resulting minimum a/b distribution 'observed' in $N = 10,000$ light curves. We find that for *any* input distribution of true axial-ratios, the randomisation of pole orientation means that light curves return a distribution of observed minimum a/b that is strongly peaked at ~ 1 and falls off exponentially. Testing a number of input distributions (flat, Gaussian, exponentially decreasing and power laws with decreasing and even increasing slope) all produced qualitatively similar output distributions, with only slight differences in the tail of the distributions visible in fig. 7. The best fit power laws $(dN/d(a/b) \propto (a/b)^{-x})$ to the output distributions have $x \approx 7, 7, 11, 15,$ and 5 respec-
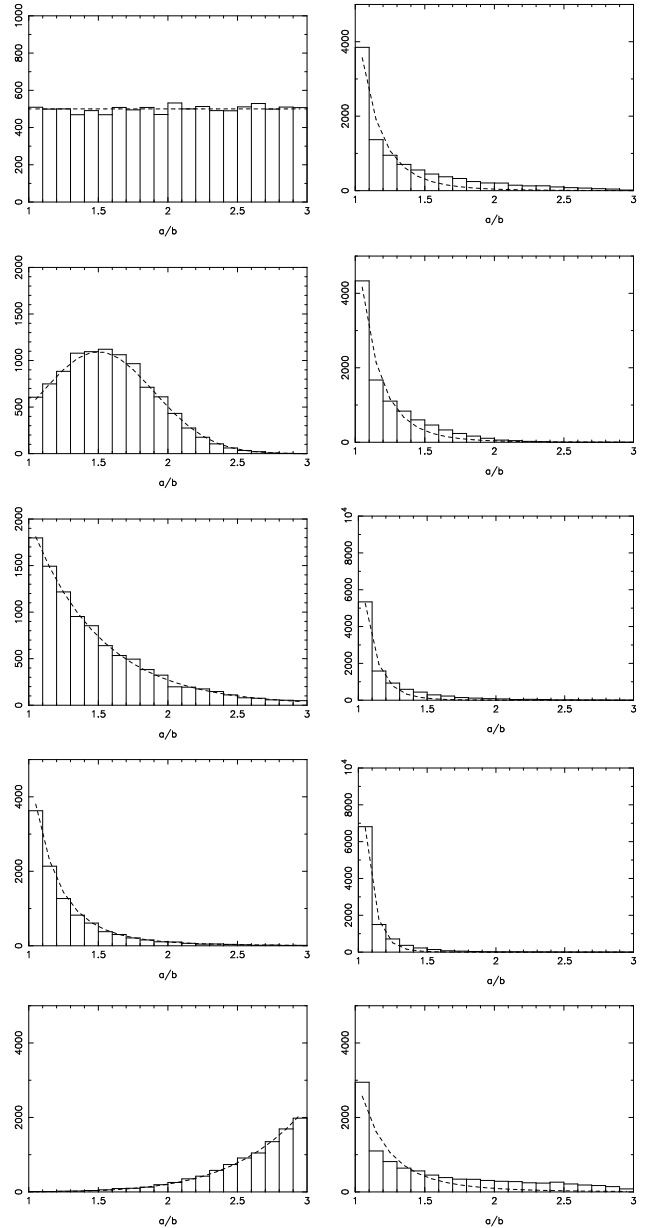

Figure 7. Left: Input 'true' a/b distributions. Right: Output 'observed' distributions based on random ϵ. From the top: Flat distribution between $1 \leqslant a/b \leqslant 3$; truncated Gaussian distribution ($< a/b > = 1.5$; $\sigma = 0.6$); exponential distribution ($\tau = 0.5$); power law ($x = 5.6$); power law ($x = -5.6$).

tively. Therefore it is impossible to constrain the underlying shape distribution from the distribution of the minimum axial-ratios observed in light curves.

In reality, there is an observational bias as nuclei with $a/b_{\min} \approx 1$ produce very small amplitude brightness variations, making the light curve un-measureable; it is impossible to know that the full range of the light curve has been observed if peaks and troughs are not recognisable, unless the rotation period of the nucleus is independently known. The difference between the appearance of the theoretical 'observed' distributions and that of fig. 5 is explained by this lack of reported light curves with $\Delta m \approx 0$.
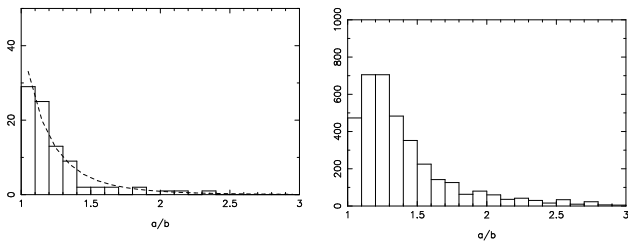
Figure 8. Shape distribution for asteroids. Left: True a/b distribution from shape models (Torppa et al. 2008). Right: Observed a/b distribution from LCDB.

4.4.3   Assumed distribution and MC results

As we cannot constrain the true distribution of axial ratios from the observed distribution, we are forced to assume a distribution for our M-C simulations: We choose to take asteroid shapes as a reasonable basis. The observed distribution of minimum axial ratios from 3666 light curves from the Asteroid Light Curve Data Base (LCDB: Warner et al. 2009) shows the same expected shape: A decrease in N as a/b increases but with a lack of objects at $a/b \approx 1$ where the light curve cannot be measured (fig. 8). However, for asteroids we can compare this with the distribution of 'true' shapes as a significant number have detailed shape models. Torppa et al. (2008) published a list of simplified descriptions of 87 models; the distribution of axial ratios from this work is best described by a power law with $x = 5.6$ (fig. 8), although we note that this sample is dominated by large asteroids.

Using this power law ($x = 5.6$, normalised over the range $1 \leqslant a/b \leqslant 3$) gives $q = 2.01 \pm 0.07$. Changing the normalisation to $1 \leqslant a/b \leqslant 5$ makes no difference. There is also only a very small difference ($\Delta q < 0.05$) for a range of power law distributions of the axial ratio from $x = 2.6$ to 8.6, although the uncertainty on the fit increases with decreasing x. This is most likely due to the smaller variation in shapes (and hence smaller variation in corrected size) at larger x, where a larger proportion of bodies are approximately spherical.

We also tested other shape distributions. Using a Gaussian distribution of a/b (average 1.5, $\sigma = 0.6$; a reasonable approximation to the observed distribution of *minimum a/b* values) instead gives $q = 1.95 \pm 0.10$. Changing the width of the distribution has no appreciable effect on q, while increasing the mean to $a/b = 2.0$ slightly decreases the slope to $q = 1.91 \pm 0.11$. A flat distribution of a/b values between 1 and 3 gives $q = 1.92 \pm 0.12$, extending it to $a/b = 5$ gives $q = 1.87 \pm 0.15$. In general, the slope of the CSD decreases as the distribution includes more elongated bodies, but due to the geometrical effects described above, there is little difference for any reasonable shape distribution. While we have shown that it is impossible to constrain the underlying shape distribution from light curve data alone due to random pole orientations, the same effect means that snap-shots tend to return approximately the correct effective radius without any knowledge of the true shape. Therefore the contribution to the uncertainty on the CSD due to the unknown shape of nuclei is not very large.
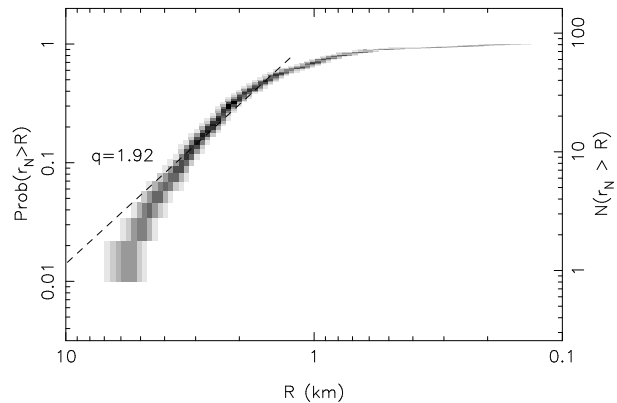

Figure 9. Probability map for MC run 39, which gives our final result by allowing all parameters to vary. The shading shows the average shape of the 10,000 CSDs: darker areas are the bins in R–$P(r_N > R)$ space where the majority of CSDs passed through, lighter areas show the outlying areas explored at the ends of the distributions. A dashed line showing the average slope from this run is over-plotted.

4.5   Including all parameters

A Monte Carlo run allowing all parameters to vary within our preferred distributions for each (Gaussians with the observed mean and σ values for A_R and β, and an $x = 5.6$ power law a/b distribution) represents our best description of what we expect the true size distribution to be, and gives an uncertainty including all parameters. The fit to this has $q = 1.92 \pm 0.20$, with an average cut off in radius at 1.25 km, and a best fit shallow slope of 0.18 ± 0.03 at sizes below this. The same run but fixing β and A_R at the 'known' values for the subset of comets with measurements gives $q = 2.01 \pm 0.20$. Note the relatively large uncertainty when including all parameters at once; we take this as a fair assessment of the true uncertainty on the CSD slope including all unknown factors.

5   DISCUSSION

Table 5 shows input parameters and results for all the MC runs, including all of the ones varying only a single parameter at a time described in the previous section and also runs allowing all parameters to vary. This shows that, relative to our reference CSD with $q = 1.97$, the largest change in slope is found when allowing the albedo to vary. This is mostly due to the very shallow slope that results from any albedo distribution that includes exceptionally dark ($A_R \approx 0$) nuclei, which are unlikely to be realistic. In any case, this demonstrates the importance of better constraining the albedo distribution of comets, an important result that the SEP-PCoN survey will provide (Fernández et al. 2008; Lowry et al. 2010). The only other MC runs with a large difference from the reference ($\Delta q \geqslant 0.1$) are the extreme shape distributions, which are also unlikely to represent reality. We can conclude from this that the uncertainty on the input parameter distributions actually has only a small effect, as the variation in slopes always falls within the typical uncertainty found when varying a single parameter at a time.

Allowing all parameters to vary within the best-fit distributions gives a final value of $q = 1.92 \pm 0.20$. This value

Table 5. Inputs and results from Monte Carlo runs.

#	Mag[a]	Phase function	Albedo	Shape[b]	Fix[c]	N	q_1[d]	q_2[e]	Cut-off
Fixed assumed values:									
1	N	0.035	0.04	Spherical	N	1	0.19	1.97	1.26
2	N	0.035	0.04	Spherical	Y	1	0.19	2.00	1.26
Adjust photometry:									
3	Y	0.035	0.04	Spherical	N	1000	0.18 ±0.01	1.93 ±0.10	1.24 ±0.07
Adjust phase function:									
4	N	0.025	0.04	Spherical	N	1	0.18	1.94	1.17
5	N	0.045	0.04	Spherical	N	1	0.19	1.98	1.35
6	N	0.065	0.04	Spherical	N	1	0.20	1.99	1.50
7	N	0.053 ± 0.016	0.04	Spherical	N	1000	0.19 ±0.01	1.95 ±0.07	1.38 ±0.07
8	N	0.053 ± 0.005	0.04	Spherical	N	1000	0.20 ±0.00	1.99 ±0.02	1.42 ±0.03
9	N	0.053 ± 0.025	0.04	Spherical	N	1000	0.19 ±0.02	1.91 ±0.11	1.35 ±0.09
10	N	0.053 ± 0.016	0.04	Spherical	Y	1000	0.19 ±0.01	2.05 ±0.06	1.39 ±0.06
11	N	0.010 ± 0.010	0.04	Spherical	N	1000	0.15 ±0.01	1.72 ±0.06	1.02 ±0.06
12	N	0.020 ± 0.010	0.04	Spherical	N	1000	0.16 ±0.02	1.81 ±0.11	1.13 ±0.11
13	N	0.030 ± 0.010	0.04	Spherical	N	1000	0.18 ±0.01	1.94 ±0.09	1.26 ±0.09
14	N	0.040 ± 0.010	0.04	Spherical	N	1000	0.19 ±0.01	1.97 ±0.05	1.32 ±0.06
15	N	0.050 ± 0.010	0.04	Spherical	N	1000	0.19 ±0.01	1.97 ±0.04	1.38 ±0.06
16	N	0.060 ± 0.010	0.04	Spherical	N	1000	0.20 ±0.01	1.97 ±0.05	1.43 ±0.06
17	N	0.070 ± 0.010	0.04	Spherical	N	1000	0.20 ±0.01	1.95 ±0.05	1.47 ±0.06
18	N	0.080 ± 0.010	0.04	Spherical	N	1000	0.20 ±0.01	1.92 ±0.05	1.49 ±0.05
19	N	0.090 ± 0.010	0.04	Spherical	N	1000	0.20 ±0.01	1.90 ±0.04	1.54 ±0.04
Adjust albedo:									
20	N	0.035	0.044 ± 0.013	Spherical	N	1000	0.17 ±0.02	1.85 ±0.15	1.16 ±0.10
21	N	0.035	0.020 ± 0.013	Spherical	N	1000	0.16 ±0.03	1.64 ±0.21	1.64 ±0.21
22	N	0.035	0.060 ± 0.013	Spherical	N	1000	0.18 ±0.02	1.89 ±0.12	1.01 ±0.08
23	N	0.035	0.040 ± 0.005	Spherical	N	1000	0.18 ±0.01	1.93 ±0.09	1.26 ±0.08
24	N	0.035	0.040 ± 0.020	Spherical	N	1000	0.16 ±0.03	1.72 ±0.19	1.18 ±0.13
25	N	0.035	0.044 ± 0.013	Spherical	Y	1000	0.17 ±0.02	1.85 ±0.14	1.16 ±0.10
Adjust shape:									
26	N	0.035	0.04	Flat, $1 \leqslant a/b \leqslant 3$	N	1000	0.18 ±0.02	1.92 ±0.12	1.17 ±0.09
27	N	0.035	0.04	Flat, $1 \leqslant a/b \leqslant 5$	N	1000	0.17 ±0.02	1.87 ±0.15	1.09 ±0.09
28	N	0.035	0.04	Gaussian, 1.5 ± 0.6	N	1000	0.18 ±0.01	1.95 ±0.10	1.22 ±0.08
29	N	0.035	0.04	Gaussian, 1.5 ± 0.3	N	1000	0.18 ±0.01	1.96 ±0.08	1.23 ±0.07
30	N	0.035	0.04	Gaussian, 1.5 ± 0.9	N	1000	0.18 ±0.01	1.94 ±0.11	1.20 ±0.08
31	N	0.035	0.04	Gaussian, 2 ± 0.6	N	1000	0.18 ±0.01	1.91 ±0.11	1.18 ±0.08
32	N	0.035	0.04	Power law, $x = -5.6$, $1 \leqslant a/b \leqslant 3$	N	1000	0.18 ±0.01	2.01 ±0.07	1.26 ±0.06
33	N	0.035	0.04	Power law, $x = -5.6$, $1 \leqslant a/b \leqslant 5$	N	1000	0.18 ±0.01	2.02 ±0.07	1.26 ±0.07
34	N	0.035	0.04	Power law, $x = -4.6$, $1 \leqslant a/b \leqslant 5$	N	1000	0.18 ±0.01	2.00 ±0.08	1.25 ±0.07
35	N	0.035	0.04	Power law, $x = -6.6$, $1 \leqslant a/b \leqslant 5$	N	1000	0.18 ±0.01	2.02 ±0.06	1.27 ±0.06
36	N	0.035	0.04	Power law, $x = -2.6$, $1 \leqslant a/b \leqslant 5$	N	1000	0.18 ±0.01	1.97 ±0.11	1.22 ±0.08
37	N	0.035	0.04	Power law, $x = -8.6$, $1 \leqslant a/b \leqslant 5$	N	1000	0.18 ±0.00	2.02 ±0.04	1.27 ±0.05
Adjust all parameters:									
38	Y	0.035 ± 0.020	0.040 ± 0.040	Gaussian, 1.5 ± 0.6	N	10000	0.16 ±0.04	1.51 ±0.22	1.17 ±0.19
39	Y	0.053 ± 0.016	0.044 ± 0.013	Power law, $x = -5.6$, $1 \leqslant a/b \leqslant 3$	N	10000	0.18 ±0.03	1.92 ±0.20	1.25 ±0.13
40	Y	0.053 ± 0.016	0.044 ± 0.013	Power law, $x = -5.6$, $1 \leqslant a/b \leqslant 3$	Y	10000	0.19 ±0.02	2.01 ±0.20	1.28 ±0.12

[a] Vary input magnitude within photometric error bars?

[b] Assumed shape distribution for nuclei.

[c] Use fixed values for β and A_R when known for a given comet? If no, all comets have values selected from the specified distribution, even those where these values are independently constrained.

[d] q_1 is the best fit at $r_N \leqslant$ cut-off.

[e] q_2 is the best fit at $r_N >$ cut-off, i.e. the quoted q.

is remarkably close to our reference value, again implying that the combined effect of the various uncertain parameters is actually relatively small. This helps to validate the assumptions made in previous CSD estimates, but the real significance of this work is that the uncertainty on this value has been rigorously determined by including all individual uncertainties on each radius. Figure 9 shows how the CSDs varied in this MC simulation.

Comparing this with the previous CSD estimates listed in Table 1 shows that the value we find for q is in a similar range to recent results, if slightly steeper than the consensus value, and when we consider the true uncertainty we find the majority of results are encompassed within 2σ. While previous works assumed constant values for A_R and β and did not consider the uncertainties on these, we show that the CSD slope including all sources of uncertainty agrees with these earlier results. It is worth noting that our testing of various distributions for the assumed parameters never produced a slope as steep as the considerably larger q found by Fernández et al. (1999) and Tancredi et al. (2006), even for extreme distributions we regard as unlikely to be realistic. We suspect that the difference is due to the higher accuracy of the photometric measurements used in our catalogue, all of which come from large aperture telescopes. It is possible that a similar MC analysis of the uncertainty on q due to the photometric uncertainties in the Tancredi et al. (2006) catalogue would give a large error bar, making it consistent with the other results, however this catalogue does not include all the necessary information (in its currently published form) to apply our technique.

The CSD slope we find is very close to the theoretical value of $q = 2.04$ found by O'Brien & Greenberg (2003) for a collisionally relaxed population of strengthless bodies. However, the large uncertainty we find means that we cannot claim that this is evidence that JFC nuclei came from collisional disruption of larger bodies; we can simply state that the CSD slope is consistent with that interpretation. It is important to remember that the activity of comets alters their sizes as they lose significant amounts of material with each perihelion passage, and therefore we would not necessarily expect to see a CSD that matched a collisional one, even if the comets are collisional fragments. Both 'normal' cometary activity and major mass loss events (nucleus splitting or outbursts) are highly variable between comets and from one orbit to another, so the change in size due to mass loss from a nucleus is a highly complicated process. More detailed theoretical models describing the expected CSD of nuclei for various formation (and evolution) scenarios are required for comparison with observations, however the large uncertainty on q found by this paper shows that we must first provide better observational constraints on all physical properties of nuclei (albedos, phase functions and shape models) to produce observed CSDs with sufficient accuracy to compare with these models.

ACKNOWLEDGMENTS

REFERENCES

A'Hearn M. F., Campins H., Schleicher D. G., Millis R. L., 1989, ApJ, 347, 1155

A'Hearn M. F., Millis R. L., Schleicher D. G., Osip D. J., Birch P. V., 1995, Icarus, 118, 223

Avni Y., Soltan A., Tananbaum H., Zamorani G., 1980, ApJ, 238, 800

Badenes C., Maoz D., Draine B. T., 2010, MNRAS, 407, 1301

Boehnhardt H., Delahodde C., Sekiguchi T., Tozzi G. P., Amestica R., Hainaut O., Spyromilio J., Tarenghi M., West R. M., Schulz R., Schwehm G., 2002, A&A, 387, 1107

Boehnhardt H., Rainer N., Birkle K., Schwehm G., 1999, A&A, 341, 912

Boehnhardt H., Tozzi G. P., Bagnulo S., Muinonen K., Nathues A., Kolokolova L., 2008, A&A, 489, 1337

Buratti B. J., Hicks M. D., Soderblom L. A., Britt D., Oberst J., Hillier J. K., 2004, Icarus, 167, 16

Campins H., Osip D. J., Rieke G. H., Rieke M. J., 1995, Planet. Space Sci., 43, 733

Delahodde C. E., Meech K. J., Hainaut O. R., Dotto E., 2001, A&A, 376, 672

Dohnanyi J. W., 1969, J. Geophys. Res., 74, 2531

Feigelson E. D., Nelson P. I., 1985, ApJ, 293, 192

Fernández J. A., Tancredi G., Rickman H., Licandro J., 1999, A&A, 352, 327

Fernández Y. R., Campins H., Kassis M., Hergenrother C. W., Binzel R. P., Licandro J., Hora J. L., Adams J. D., 2006, AJ, 132, 1354

Fernández Y. R., Kelley M. S., Lamy P. L., Toth I., Groussin O., A'Hearn M. F., Bauer J. M., Campins H., Fitzsimmons A., Licandro J., Lisse C. M., Lowry S. C., Meech K. J., Pittichová J., Reach W. T., Snodgrass C., Weaver H. A., 2008, LPI Contributions, 1405, 8307

Fernández Y. R., Lisse C. M., Ulrich Käufl H., Peschke S. B., Weaver H. A., A'Hearn M. F., Lamy P. P., Livengood T. A., Kostiuk T., 2000, Icarus, 147, 145

Fernández Y. R., Lowry S. C., Weissman P. R., Mueller B. E. A., Samarasinha N. H., Belton M. J. S., Meech K. J., 2005, Icarus, 175, 194

Fernández Y. R., Meech K. J., Lisse C. M., A'Hearn M. F., Pittichová J., Belton M. J. S., 2003, Icarus, 164, 481

Fitzsimmons A., Snodgrass C., Lowry S., 2011, MNRAS, in prep.

Hicks M. D., Bambery R. J., Lawrence K. J., Kollipara P., 2007, Icarus, 188, 457

Irwin M., Tremaine S., Zytkow A. N., 1995, AJ, 110, 3082

Jewitt D. C., 2002, AJ, 123, 1039

—, 2006, AJ, 131, 2327

Jewitt D. C., Luu J. X., 1989, AJ, 97, 1766

Jewitt D. C., Meech K. J., 1988, ApJ, 328, 974

Jewitt D. C., Sheppard S., 2004, AJ, 127, 1784

Jewitt D. C., Sheppard S., Fernández Y., 2003, AJ, 125, 3366

Kaplan E. L., Meier P., 1958, J. Am. Statistical. Assoc., 53, 457

Kelley M. S., Wooden D. H., Tubiana C., Boehnhardt H., Woodward C. E., Harker D. E., 2009, AJ, 137, 4633

Lamy P. L., Toth I., A'Hearn M. F., Weaver H. A., Jorda L., 2007, Icarus, 187, 132

Lamy P. L., Toth I., Fernández Y. R., Weaver H. A., 2004, in Comets II, Festou M. C., Keller H. U., Weaver H. A., eds., University of Arizona Press, pp. 223–264

Lamy P. L., Toth I., Jorda L., Groussin O., A'Hearn M. F., Weaver H. A., 2002, Icarus, 156, 442

Lamy P. L., Toth I., Jorda L., Weaver H. A., A'Hearn M., 1998, A&A, 335, L25

Lamy P. L., Toth I., Weaver H. A., A'Hearn M. F., Jorda L., 2009, A&A, 508, 1045

Li J., A'Hearn M. F., Belton M. J. S., Crockett C. J., Farnham T. L., Lisse C. M., McFadden L. A., Meech K. J., Sunshine J. M., Thomas P. C., Veverka J., 2007a, Icarus, 187, 41

Li J., A'Hearn M. F., Farnham T. L., McFadden L. A., 2009, Icarus, 204, 209

Li J., A'Hearn M. F., McFadden L. A., Belton M. J. S., 2007b, Icarus, 188, 195

Licandro J., Tancredi G., Lindgren M., Rickman H., Hutton R. G., 2000, Icarus, 147, 161

Lisse C. M., A'Hearn M. F., Groussin O., Fernández Y. R., Belton M. J. S., van Cleve J. E., Charmandaris V., Meech K. J., McGleam C., 2005, ApJ, 625, L139

Lisse C. M., Fernández Y. R., Reach W. T., Bauer J. M., A'Hearn M. F., Farnham T. L., Groussin O., Belton M. J., Meech K. J., Snodgrass C. D., 2009, PASP, 121, 968

Lowry S. C., Fernández Y., Laird R., AHearn M. F., Bauer J. M., Campins H., Fitzsimmons A., Groussin O., Hsieh H., Kelley M., Lamy P., Licandro J., Lisse C. M., Meech K. J., Pittichová J., Reach W. T., Snodgrass C., Toth I., Weaver H. A., Weissman P., 2010, in European Planetary Science Congress

Lowry S. C., Fitzsimmons A., 2001, A&A, 365, 204

—, 2005, MNRAS, 358, 641

Lowry S. C., Fitzsimmons A., Cartwright I. M., Williams I. P., 1999, A&A, 349, 649

Lowry S. C., Fitzsimmons A., Collander-Brown S., 2003, A&A, 397, 329

Lowry S. C., Weissman P. R., 2003, Icarus, 164, 492

Luu J. X., Jewitt D. C., 1992, AJ, 104, 2243

Mazzotta Epifani E., Palumbo P., Capria M. T., Cremonese G., Fulle M., Colangeli L., 2007, MNRAS, 381, 713

—, 2008, MNRAS, 390, 265

Meech K. J., Belton M. J. S., Mueller B. E. A., Dicksion M. W., Li H. R., 1993, AJ, 106, 1222

Meech K. J., Hainaut O. R., Marsden B. G., 2004, Icarus, 170, 463

O'Brien D. P., Greenberg R., 2003, Icarus, 164, 334

Pittichová J., Woodward C. E., Kelley M. S., Reach W. T., 2008, AJ, 136, 1127

Russell H. N., 1916, ApJ, 43, 173

Schmitt J. H. M. M., 1985, ApJ, 293, 178

Sekanina Z., 1991, AJ, 102, 350

Snodgrass C., Fitzsimmons A., Lowry S. C., 2005, A&A, 444, 287

Snodgrass C., Lowry S. C., Fitzsimmons A., 2006, MNRAS, 373, 1590

—, 2008, MNRAS, 385, 737

Snodgrass C., Meech K., Hainaut O., 2010, A&A, 516, L9

Snodgrass C. D. B., 2006, PhD thesis, Queen's University Belfast

Tancredi G., Fernández J. A., Rickman H., Licandro J., 2000, A&AS, 146, 73

—, 2006, Icarus, 182, 527

Torppa J., Hentunen V., Pääkkönen P., Kehusmaa P., Muinonen K., 2008, Icarus, 198, 91

Tubiana C., Barrera L., Drahus M., Boehnhardt H., 2008, A&A, 490, 377

Warner B. D., Harris A. W., Pravec P., 2009, Icarus, 202, 134

Weissman P. R., Choi Y. J., Lowry S. C., 2008, in Bulletin of the American Astronomical Society, Vol. 40, p. 387

—, 2009, in IAU Symposium, Vol. 263, p. o36

Weissman P. R., Lowry S. C., 2003, in Lunar and Planetary Institute Conference Abstracts, p. 2003