

A million asteroid observations in the Sloan Digital Sky Survey[★]

Alexey V. Sergeev^{id} and Benoit Carry^{id}

Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, France
e-mail: alexey.sergeev@oca.eu, alexey.v.sergeev@gmail.com

Received 27 January 2021 / Accepted 22 April 2021

ABSTRACT

Context. The populations of small bodies of the Solar System (asteroids, comets, Kuiper-Belt objects) are used to constrain the origin and evolution of the Solar System. Both their orbital distribution and composition distribution are required to track the dynamical pathway from their regions of formation to their current locations.

Aims. We aim at increasing the sample of Solar System objects that have multi-filter photometry and compositional taxonomy.

Methods. We search for moving objects in the archive of the Sloan Digital Sky Survey. We attempt at maximizing the number of detections by using loose constraints on the extraction. We then apply a suite of filters to remove false-positive detections (stars or galaxies) and mark out spurious photometry and astrometry.

Results. We release a catalog of 1 542 522 entries, consisting of 1 036 322 observations of 379 714 known and unique SSOs together with 506 200 observations of moving sources not linked with any known SSOs. The catalog completeness is estimated to be about 95% and the purity to be above 95% for known SSOs.

Key words. methods: data analysis – minor planets, asteroids: general – catalogs – surveys

1. Introduction

Considered as nuisances for the trails they imprint on photographic plates, asteroids (i.e., the “vermin of the sky”) are prime trackers of the dynamical events that shaped our Solar System, in particular migrations of the giant planets (Morbidelli et al. 2015; Clement et al. 2020). While their orbital distribution has been widely used, the distribution of their compositions provides further constraints on the models that describe the evolution of our Solar System (Levison et al. 2009; Vokrouhlický et al. 2016; Raymond & Izidoro 2017).

In this context, it is critical to understand their composition in detail and their location as well as the timing of formation. This is achieved through the laboratory study of meteorites (Bottke et al. 2006; Vernazza et al. 2014; Scott et al. 2018), of which the near-Earth asteroids (NEAs) are the progenitors (Granvik & Brown 2018; Granvik et al. 2018), and the compositional mapping of asteroids in the main belt (MB) between Mars and Jupiter (DeMeo & Carry 2014).

The composition, and more broadly the classification, of asteroids has been based for decades on multi-filter photometry in the visible (e.g., Zellner et al. 1985). With the advent of charge coupled devices (CCDs) in the 1990s, multi-filter photometry moved toward low-resolution spectroscopy in the visible (Xu et al. 1995; Bus & Binzel 2002a) and was later extended to the near-infrared (DeMeo et al. 2009).

After three decades of targeted observations, spectra (visible, near-infrared, or covering both wave ranges) have been collected for about 9000 asteroids (albeit disseminated over a myriad of articles, e.g., Bus & Binzel 2002b; Lazzaro et al. 2004; Fornasier et al. 2014), including around one thousand NEAs

(mainly from the NEOSHIIELD2, MANOS, and MITHNEOS surveys; Perna et al. 2018; Devogèle et al. 2019; Binzel et al. 2019). However, compared to the current census of over 20 000 NEAs and more than 860 000 MB asteroids, these 9000 are only the tip of the iceberg.

In the current era, these “vermin” have become more appreciated, being serendipitously imaged by large sky surveys from the ultraviolet to far-infrared ranges. The most prominent examples of such un-targeted observations are the scientific exploitations of the Sloan Digital Sky Survey (SDSS) – which provides multi-filter photometry in the visible, constraining the composition (Ivezić et al. 2001) – and the Wide-field Infrared Survey Explorer (WISE) – which provides mid-infrared photometry and hence diameter and albedo determinations (Masiero et al. 2011).

However, most imaging surveys do not report moving objects as direct products. They must be specifically searched for on released data, generally by external teams. The recent works of Popescu et al. (2016) on the European Southern Observatory (ESO) Visible and Infrared Survey Telescope for Astronomy (VISTA) Hemisphere Survey and Mahlke et al. (2018) on the ESO VST (Very Large Telescope (VLT) Survey Telescope) Kilo-Degree Survey are typical examples. This situation is, fortunately, evolving. The current European Space Agency (ESA) *Gaia* mission includes a specific processing dedicated to Solar System objects (SSOs) for which a data release already occurred (Spoto et al. 2018), and the upcoming ESA *Euclid* mission and the Legacy Survey of Space and Time (LSST) of the Vera C. Rubin observatory include SSOs releases in their respective baselines (see Carry 2018; Jones et al. 2009).

We focus here on broadband photometry, in the visible, as it provides strong constraints on asteroid composition and classification. With over 100 000 reported individual asteroids with five filters in the visible (u , g , r , i , and z ; Gunn et al. 1998), the SDSS has been the main source of compositional information for almost two decades. Its multi-filter photometry

[★] The catalog is available at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via <http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/652/A59>

has been analyzed both in its own color space (Szabó et al. 2004; Nesvorný et al. 2005) and mapped into a classification compliant with visible and near-infrared spectroscopy (Carvano et al. 2010; DeMeo & Carry 2013). SDSS photometry has been the source of a variety of studies, including regarding the selection of candidates for dedicated spectroscopic surveys (Moskovitz et al. 2008; Oszkiewicz et al. 2014; DeMeo et al. 2014, 2019), dynamical families and surface aging through space weathering (Parker et al. 2008; Thomas et al. 2012; Graves et al. 2018), the orbital distribution of hydrated C-types (Rivkin 2012), and the general structure of the asteroid belt (DeMeo & Carry 2014).

Even though the SDSS Moving Object Catalog (MOC) has been widely used, its current release, which is called the ADR4 Ivezic et al. (2010) and was published in 2008, only contains a fraction of the entire data set. First, only 211 138 moving sources were identified (i.e., linked with a known SSO) out of the 471 569 detected moving sources. As the number of known SSOs has increased from about 550 000 to 960 000 since 2008, many of the identifications missing from the ADR4 may be recovered (such as was done by Carry et al. 2016 who identified 57 646 observations of 36 730 asteroids). Second, the data selection at the origin of the SDSS MOC imposed strict lower and upper limits on the apparent velocity of sources between frames. This led to the rejection of any object moving faster than 0.5 deg day^{-1} , that is, effectively filtering out most NEAs present in the images (Solano et al. 2014). Finally, the last observing run included in the ADR4 release dated from March 2007, while the SDSS imaging survey lasted until November 2008.

The present article aims to increase the number of asteroids with SDSS astrometry and multi-filter photometry by identifying previously unidentified moving sources and searching for moving sources in frames not included in the latest SDSS MOC release. The article is organized as follows. In Sect. 2 we describe the approach we take to identify potential moving objects from the SDSS archive, aiming at maximizing completeness. We detail the filters applied to the sample to increase its purity (i.e., to minimize the number of false-positive sources) in Sect. 3. The identification of known sources among these candidates is described in Sect. 4. The completeness and purity of the catalog is estimated in Sect. 5. We then address suspicious individual measurements (astrometry or photometry) in Sect. 6. We classify the objects consistently with the DeMeo et al. (2009) taxonomic classification in Sect. 7. Finally, we summarize the released asteroid sample in Sect. 8.

2. Extraction of Solar System object candidates

As noted earlier, the latest release (ADR4) of the SDSS MOC took place before the end of operations of the SDSS imaging survey. We thus compared the list of 519 runs (number, which identifies the specific scan)¹ included in the ADR4 with the 832 runs of SDSS Data Release 16 (DR16; Ahumada et al. 2020) and the 303 runs of Stripe 82 (Adelman-McCarthy et al. 2008), together known as the SDSS Legacy Survey.

The main SDSS imaging survey covered 7500 deg^2 in the northern Galactic cap and 750 deg^2 in the southern Galactic cap, where each patch of the sky was imaged in five filters in a single visit. The 270 deg^2 of Stripe 82 (a thin long area within $21 \text{ h} < \text{RA} < 04 \text{ h}$ and $-1.25^\circ < \text{Dec} < 1.25^\circ$; see Fig. 1) were repeatedly imaged (about 80 times) as a time-domain sur-

¹ An SDSS run encompasses all the contiguous observations along a great circle.

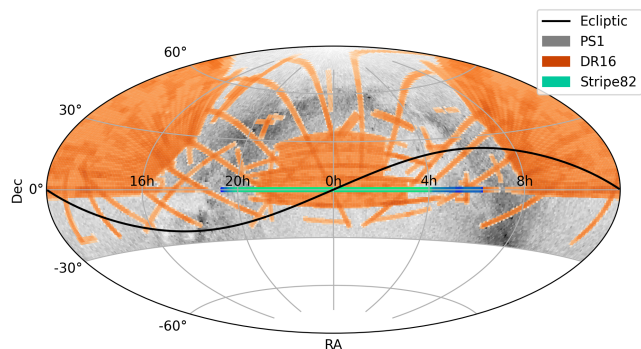


Fig. 1. Sky coverage of the main SDSS survey (orange) and Stripe 82 (light blue) in equatorial coordinates (Aitoff projection). The background image is generated from the Pan-STARRS stellar catalog (Chambers et al. 2016). The black line represents the ecliptic plane.

vey (Sako et al. 2008; Frieman et al. 2008). Both surveys include 80 common runs. We only kept 756 runs from DR16, rejecting the runs that were affected by bad weather (e.g., clouds, bad seeing) and thus had a quality score below 0.1. In total, we analyzed 938 unique runs, about twice as much as those listed in the ADR4.

The median 5σ depth of SDSS photometric observations, based on the formal uncertainties from point spread function (PSF) photometry on point sources, are, in AB magnitudes: $u = 22.15$, $g = 23.13$, $r = 22.20$, $i = 22.20$, and $z = 20.71$ (Ahumada et al. 2020). We thus set a brightness limit of 22.2 in r (0.7 magnitude fainter than the selection criterion of Ivezić et al. 2001).

Contrary to the ADR4 extraction we did not set an upper limit on the apparent velocity, to avoid the rejection of NEAs. Furthermore, as sources moving faster than $0.5^\circ \text{ day}^{-1}$ may appear trailed in the images, we did not restrict the object type to stars only, but allowed both stars (type=6) and galaxies (type=3). The larger number of runs, combined with our more relaxed selection criteria to maximize the completeness (the SQL (Structured Query Language) code is provided in Appendix B for reproducibility), leads to a dramatic increase in the number of SSO candidates compared to the ADR4: 4 804 003 in DR16 and 4 071 153 in Stripe 82.

3. Rejection of false-positive candidates

The above extraction returned almost 9 million SSO candidates. As the selection criteria aimed at completeness, many if not most candidates could be stars or galaxies (i.e., false-positive identification rather than genuine SSOs). We describe below the successive steps we performed to improve the purity of the sample by removing these false-positive detections.

3.1. Identification of background sources with Pan-STARRS

The Pan-STARRS catalog provides a homogeneous, multiband survey that covers the sky north of declination -30° (Fig. 1). It probes the sky with similar filters, depth, and seeing quality as the SDSS (Flewelling et al. 2020). The mean 5σ point source limiting sensitivities in the Pan-STARRS stacked images reach up to a 23.3 mag limit. The epochs of observation separated by years between the SDSS and Pan-STARRS provide a practical means to identify fixed-coordinate objects (stars and galaxies) among the SSO candidates.

We used the Pan-STARRS Python search API (MAST²) to compare our catalog of candidates with the Pan-STARRS Data Release 2 (DR2) catalog. We excluded candidates matched within 2'' by Pan-STARRS DR2 objects if the latter were detected at least twice by Pan-STARRS. After this filtering, the number of SSO candidates dropped to 3 238 202.

3.2. Exclusion of duplicates

The SDSS PhotoObjAll catalog we queried (Appendix B) contains duplicated measurements in some cases (i.e., the same observation could appear a few times in the catalog). This is a consequence of several factors.

The first is the methodology of SDSS photometry. Identical objects could be measured in different fields because part of the frames has cross-covering. The second is the fact that while the Stripe 82 and DR16 catalogs have the same runs, they were processed by different versions of the SDSS photometry pipeline and therefore could have slightly different photometry. The third factor is due to the moving nature of observed SSOs. For a high apparent rate, the same object can be identified as two or more stationary sources (see Sect. 6.3 for a description of our extraction of photometry for fast-moving asteroids such as NEAs). For each duplicated entry, we grouped objects by run number within a 5'' radius and selected the most accurate photometry for each band.

3.3. Identification of high-proper-motion stars with Gaia

While the cross-match with Pan-STARRS allows the identification and rejection of most stationary sources, high-proper-motion stars may have moved farther than our cross-match radius of 2'' in 20 years. We selected all stars listed in *Gaia* DR2 (Brown et al. 2018) with proper motion above 0.05 arcsec year⁻¹, fainter than $G = 14.0$, and with declination above -30° . For each of the 2 278 339 stars we retrieved, we computed its position at the SDSS mid-observing epoch, January 1, 2004 (see Appendix B). We then cross-matched this catalog of high-proper-motion stars with the catalog of SSO candidates. We identified 67 additional stars missed by the Pan-STARRS comparison above and rejected them from the sample.

3.4. Rejection of dubious frames

Among all SDSS fields, we found frames with a suspiciously high number of SSOs. A visual inspection of these fields revealed issues with the astrometry across the different filters. As a result, sources have different positions in each filter, leading to erroneous classification as moving sources by the SDSS software. We removed 312 such fields from DR16 and 414 from Stripe 82. We also removed all 300 frames of Run 2505, which were totally crowded by Milky Way stars. In this step, we rejected 6701 candidates from our catalog.

4. Identification of known Solar System objects

The extraction and selection presented in the previous section resulted in a catalog of 1 257 408 observations. To identify the detected objects, we performed 938 046 (DR16) and 586 983 (Stripe 82) cone-search queries with SkyBoT (the Sky Body Tracker; Berthier et al. 2006), one for each of the frames

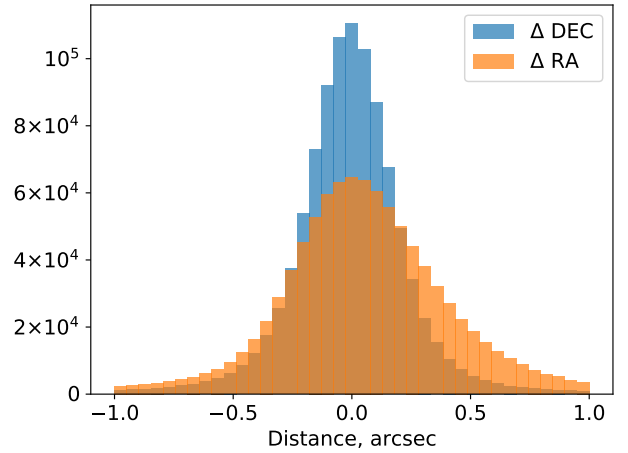


Fig. 2. Distribution of distances between SkyBoT predictions and SDSS measurements.

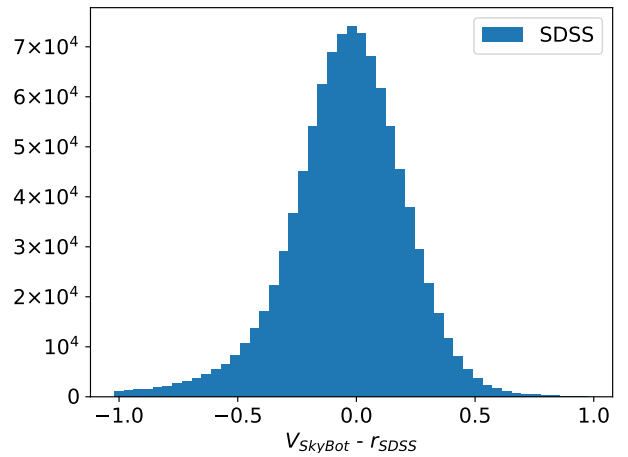


Fig. 3. Distributions of SkyBoT V magnitudes and transformed SDSS SSOs.

(13.5' \times 9.0') included in our catalog. SkyBoT is a Virtual Observatory web service that provides cone-search utilities for SSOs. It is thus a useful tool for seeking and identifying moving objects in astronomical images.

In 812 573 of our cases, a known SSO was predicted by SkyBoT to be located near the source we extracted from the SDSS archive (Sect. 2). The mean distance between the prediction and the detection is $-0.05 \pm 0.45''$ in right ascension and $-0.01 \pm 0.22''$ in declination (Fig. 2). The $V-r$ color, presented in Fig. 3, is 0.01 ± 0.24 (i.e., slightly redder than Solar color; Willmer 2018). We associated the SkyBoT prediction with the SDSS measurement whenever the angular distance was smaller than 30'', the $|V-r|$ magnitude difference was smaller than 5, and the difference in apparent motion was within 5'' h⁻¹.

In 285 114 other cases, a known SSO was predicted by SkyBoT but did not correspond to any of the 1 257 408 sources we extracted (Sect. 2). For each of these predictions, we extracted the closest source in the SDSS archive. The majority of these predictions have small apparent rates (hence were not marked as moving objects by the SDSS pipeline). We then applied the suite of filters described above (Sect. 3) and added all these recovered sources to the catalog.

In total, we identified 1 036 322 observations of 379 714 known SSOs. The distribution of the number of asteroid observations is presented in Fig. 4. Nearly half of the asteroids were

² http://ps1images.stsci.edu/ps1_dr2_api.html

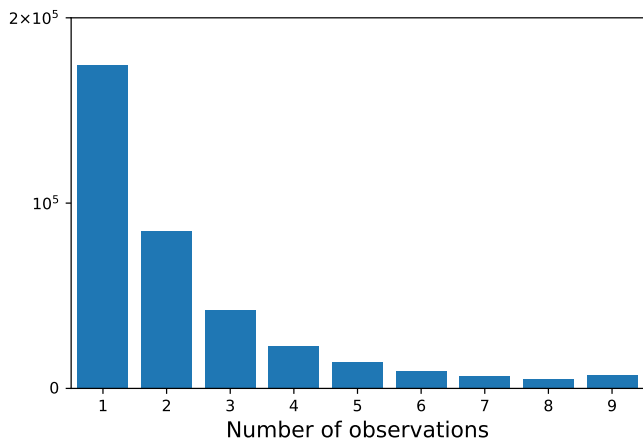


Fig. 4. Distribution of the number of observations per SSO.

Table 1. Number of observations ($\#_{\text{obs}}$) and unique SSOs ($\#_{\text{SSOs}}$) for each dynamical class.

Dynamical class	$\#_{\text{SSOs}}$	$\#_{\text{obs}}$
NEA Aten	52	68
NEA Apollo	836	1451
NEA Amor	764	1355
Mars-Crosser	4242	9024
Hungaria	6362	12 841
MB Inner	105 273	276 899
MB Middle	134 878	370 154
MB Outer	119 392	337 668
MB Cybele	1722	5070
MB Hilda	1923	5021
Trojan	3929	8721
Centaur	123	522
KBO	1024	3143
Comet	233	420
Total	379 714	1 036 322

observed only once; however, the mean value of the observations of individual asteroids is 2.7. Moreover, 13 882 asteroids have more than ten observations.

The catalog contains 1 542 522 observations, including 506 200 observations of moving objects not linked with any known SSO. Most of the identified SSOs are located in the main asteroid belt. All dynamical classes are, however, present in the catalog, including comets (Table 1).

5. Purity and completeness

We estimated the completeness of our catalog using the association with known SSOs presented above (Sect. 4). The SSOs predicted by SkyBoT but not present in our catalog provide an estimate of its completeness, albeit not an unbiased one. The current census of SSOs may itself be biased (Mahlke et al. 2018), and a 100% completeness with known SSOs may not necessarily imply a full completeness of sources on sky. As an illustration, about a third of all SSOs reported in the present catalog are not identified. Hence, this comparison provides an estimate on the fraction of objects removed by the different filters we applied in Sect. 3.

We treated the cases of MB asteroids, NEAs, and Kuiper belt objects (KBOs) separately owing to their vastly different appar-

ent motion. In all cases, we studied the completeness as a function of the apparent magnitude (as reported in V by SkyBoT) and motion (in $'' \text{h}^{-1}$). We present the estimated completeness in Fig. 5. The completeness is typically around 95% until $V \approx 22$, where it drops to about 70% at $V \approx 23$, and then to 30% at $V \approx 24$. Such a completeness is similar to that of the ADR4 (Ivezić et al. 2001).

We estimated the purity of the catalog by visually inspecting RGB (red, green, blue) images made by combining the g , r , and i frames (similar to the insets of Fig. 6). We inspected all KBOs, 2000 randomly selected MB asteroids, and 2915 NEAs. The purity of the catalog is typically above 97% down to magnitude $V \approx 23$, where it suddenly drops (Fig. 5). This represents a clear improvement with respect to the ADR4, for which a 94% purity was quoted for MB asteroids, dropping to 50% for both fast and slow motion SSOs (Ivezić et al. 2001).

6. Identification of spurious measurements

The filters described in Sect. 3 limit the number of false positives among the 1 542 522 observations we report (Sect. 5). We describe below further tests for identifying potential issues with the astrometry or photometry of measurements in individual bands.

6.1. Spurious astrometry from linearity of motion

As the SDSS imaging camera scanned the sky, the same region of the sky was sequentially imaged in r , u , i , z , and g , with a time interval of 17.7 s between each 54 s exposure frame. From the reported position offsets in each band with respect to the position in r , we built the footsteps of each SSO to check the linearity of its motion (Mahlke et al. 2019). We used the Siegel estimator, a linear regression that is robust to outliers (Siegel 1982). We find that 92% of the SSOs have a coefficient of determination, R^2 , larger than 0.9. Considering that an offset position in a single band significantly lowers R^2 , such high values provide a strong indication of linear motion, and hence a confirmation of the genuineness of these SSOs. We used this criterion to identify spurious astrometry in the catalog (see Appendix A). Using g and r as reference (the two deepest filters; Sect. 2), we predicted the expected position in z , u , and i . We marked as dubious each position in these bands that is located farther than one pixel ($0.4''$) from the prediction (Fig. 6).

6.2. Identification of slow-moving SSOs

The extraction of SSO candidates (Sect. 2) imposes a minimum apparent motion of $0.05^\circ \text{day}^{-1}$ to reject all stationary sources, such as stars and galaxies (similar to the ADR4 extraction; Ivezić et al. 2001). This threshold converts to a minimum change in position of $0.6''$ between the r and g frames, which provide the largest temporal baseline, 286.8 s. This effectively precludes the identification of most moving objects from the outer Solar System (only 41 SSOs with a semimajor axis above 5.5 au were reported by Ivezić et al. 2001).

As is visible in Fig. 7, however, 216 348 objects do not display apparent motion. These can be slow-moving objects, such as KBOs or asteroids at their turnaround point between direct and retrograde motion. However, the false-positive rate may be higher among these objects because the identification of these sources as moving objects is more prone to uncertainty (the linearity of motion cannot be checked, for instance).

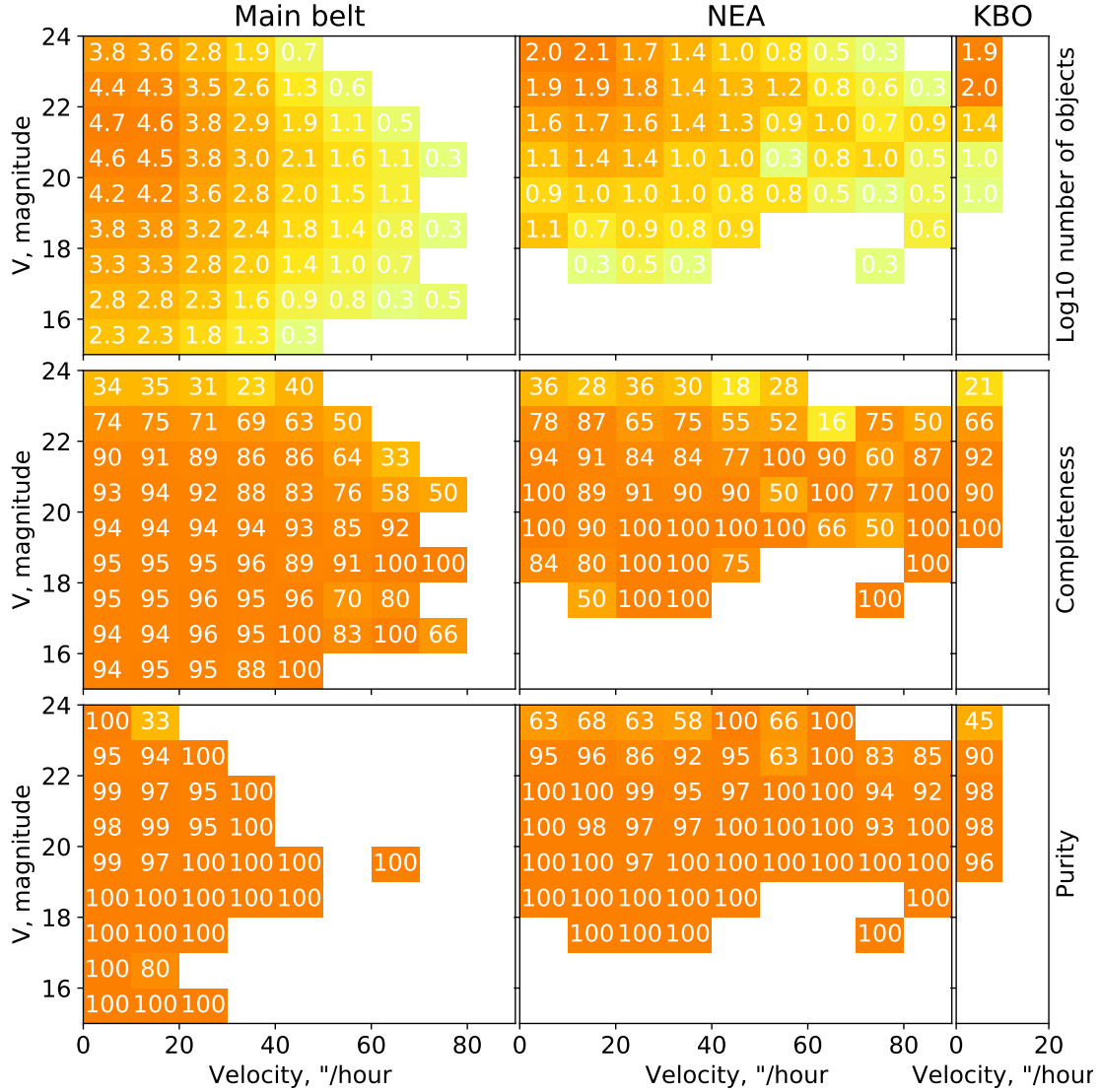


Fig. 5. Distribution of the population (*top*), completeness (*middle*), and purity (*bottom*) of asteroids (MB asteroids and NEAs) and KBOs.

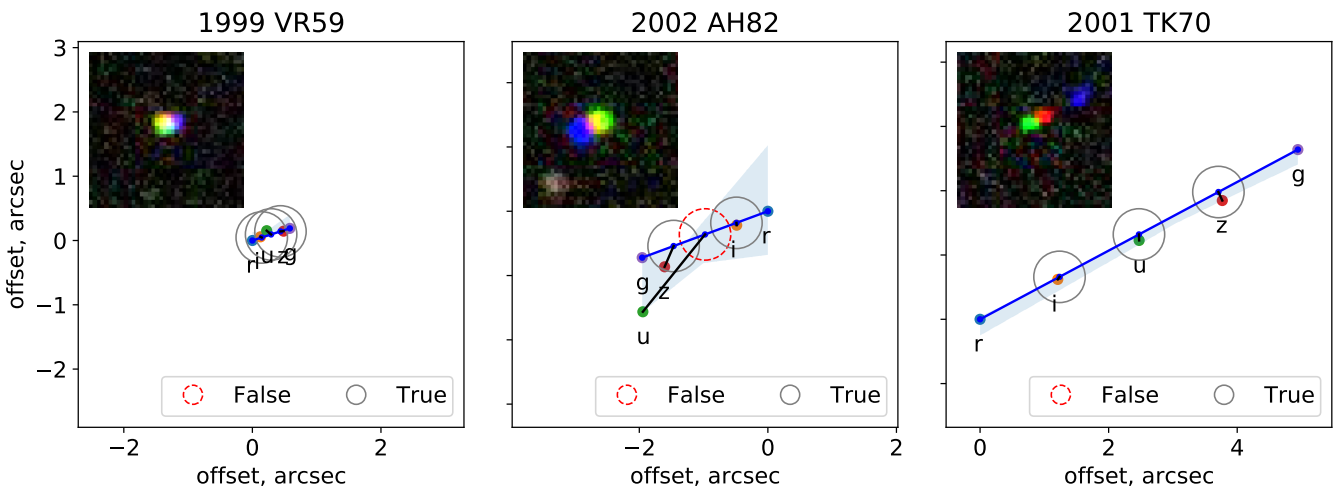


Fig. 6. Examples of asteroid signatures in SDSS frames: slow-moving ($0.05^\circ \text{ day}^{-1}$, *left*), typical motion ($\sim 0.25^\circ \text{ day}^{-1}$, *center*), and fast-moving ($0.5^\circ \text{ day}^{-1}$, *right*). For each, the observed position offsets (red points) and expected position offsets (blue points) with respect to *r* are shown. The solid gray and dashed red circles with 0.4 arcsec (1 pixel) radii indicate validated and spurious astrometry, respectively. For each object, we also present a *g r i* tri-color image at the top-left corner.

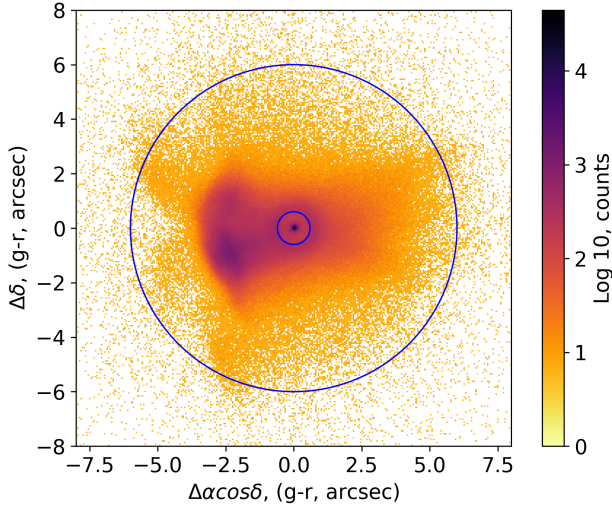


Fig. 7. Position offset between r and g . The inner black circle represents the lower limit of $0.05^\circ \text{ day}^{-1}$ that we use to extract SSO candidates. The outer blue circle shows the upper limit of $0.5^\circ \text{ day}^{-1}$ from the ADR4 that we do not use here.

We thus marked 53 980 objects in the catalog (Appendix A) as suspicious: Their change in position between g and r is less than $0.6''$ and there are no known SSOs predicted within $5''$. We, however, did not flag out the 162 368 slow-moving SSOs associated with a SkyBoT counterpart (Sect. 4).

6.3. Retrieving the photometry of fast-moving SSOs

As a feature of the SDSS catalog creation, moving objects with a large apparent motion may not be recognized as such (Ivezić et al. 2001). Their successive appearances in the different filters may be cataloged as different SDSS sources. We thus applied an improved source extraction for 9134 SSOs with a proper motion above $60'' \text{ h}^{-1}$. This apparent velocity corresponds to an apparent displacement of $1''$ (the mean seeing value) over the 54 s integration of each filter.

The extraction procedure includes extracting all sources in the predicted moving distance from the SDSS catalog for each band. We then checked sources along with predicted positions and combined them into one moving source.

For each filter, we replaced the astrometry and photometry from the previous extraction with those of the SDSS source located the closest to the SkyBoT prediction. We then reapplied all the filters described in Sect. 3.

6.4. Spurious photometric uncertainty

As is visible in Fig. 8, the distribution of photometric uncertainties in each filter is too complex to be due to photon noise only. There are in particular two spurious behaviors, which are present in all filters. We marked the photometry as suspicious in the catalog (Appendix A) using the following empirical threshold:

$$\text{psfMagErr}_{\text{max}} \stackrel{\text{min}}{=} \begin{cases} 1.05 - \text{Exp} \left[\frac{\text{psfMag} - \text{Mag}_{\text{lim}}}{0.6} \right] \\ \exp \left[\frac{\text{psfMag} - (\text{Mag}_{\text{lim}} - \sigma)}{\sigma} \right] + 0.05 \end{cases}, \quad (1)$$

where the limiting magnitude, Mag_{lim} , and the curvature, σ , are filter dependent and listed in Table 2.

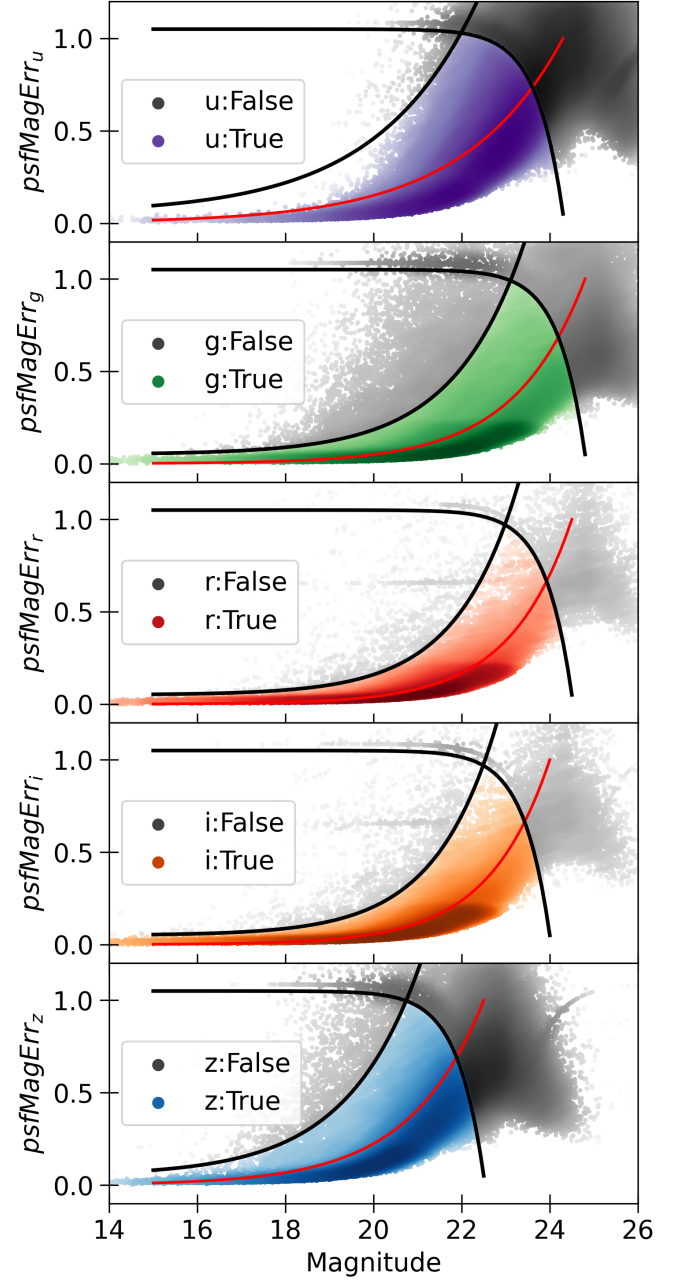


Fig. 8. Photometric uncertainty of SSOs in all bands. The photometry of the gray points is marked as suspicious. The black curves represent the acceptance limits and the red curve the global trend (see text).

The first criterion in Eq. (1) identifies the overemphasized uncertainty value of ~ 1.1 for bright sources, which exponentially decreases to 0 for faint sources. The second marks the photometry with uncertainties too large to be meaningful, including the overrepresented uncertainty value of 0.6 mag. We set the σ curvature by fitting the following equation to each band (red curves in Fig. 8):

$$\text{psfMagErr} = \exp \frac{\text{psfMag} - \text{Mag}_{\text{lim}}}{\sigma}. \quad (2)$$

As expected, the largest number of observations with suspicious photometry is presented in the u and z bands, while almost all observations in the r and i bands have accurate photometry (Table 2).

Table 2. Parameters (mag_{lim} , σ) for the selection of photometric uncertainties.

Parameter	u	g	r	i	z
mag_{lim}	24.3	24.8	24.5	24.0	22.5
σ	2.27	1.59	1.40	1.40	1.67
Suspicious	46%	14%	2%	2%	31%

Notes. We also report the fraction of measurements marked as suspicious.

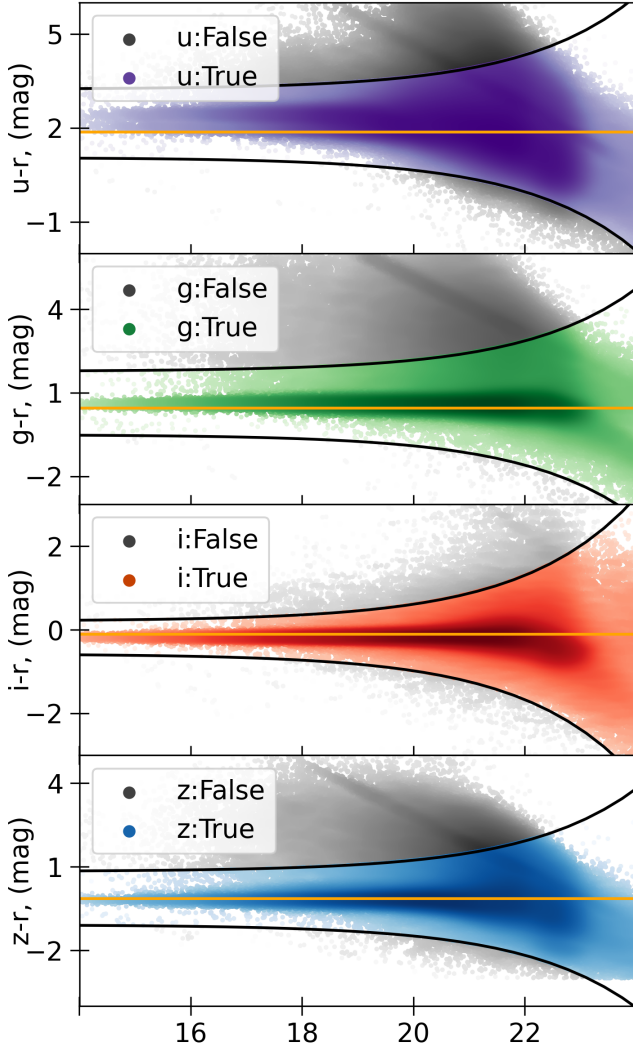


Fig. 9. Distribution of $u-r$, $g-r$, $i-r$, and $z-r$ colors as a function of magnitude. The rejected measurements are displayed in gray. An orange horizontal line demonstrates the Sun color.

6.5. Identification of spurious colors

We performed a final test on the reported photometry. In a first approximation, SSOs display Solar colors, and a significant departure from these colors can be used to identify potential issues with the photometry not identified in the previous steps. Using r as a reference, we analyzed how colors vary with magnitude (Fig. 9). The spread of colors increases with magnitude due to increasing photometric uncertainty. We flagged photometry resulting in a color farther than 3σ from the median color (Appendix A). The fraction of measurements marked as such amounts to 7%, 11%, 1%, and 8% for u , g , i , and z , respectively.

Table 3. Complex color boundaries.

Complex	gr_{min}	gr_{max}	iz_{min}	iz_{max}	gi_{min}	gi_{max}
A	0.70	0.90	-0.20	0.03	1.00	1.10
B	0.30	0.55	-0.10	0.00	0.20	0.60
C	0.35	0.60	0.00	0.15	0.50	0.65
D	0.53	0.85	0.11	0.34	0.70	1.00
K	0.55	0.63	-0.00	0.06	0.70	0.95
L	0.60	0.85	0.06	0.13	0.80	1.00
Q	0.60	0.65	-0.20	0.08	0.60	0.85
S	0.55	0.85	-0.20	0.06	0.65	1.00
V	0.55	0.90	-0.70	0.20	0.65	1.00
X	0.48	0.60	0.04	0.11	0.55	0.85

7. Taxonomic classification

We used the photometry to classify the asteroids within a scheme consistent with the widely used Bus-DeMeo taxonomy (DeMeo et al. 2009). We followed the approach in DeMeo & Carry (2013), according to which strict boundaries in the overall spectral slope and depth of the one-micron band are used to classify the SSOs into ten broad complexes: A, B, C, D, K, L, Q, S, V, and X. We, however, introduced several improvements.

Magnitude versus reflectance. We used the SDSS magnitudes, similarly to what was done in Carvano et al. (2010), rather than converting them into spectral reflectance, as was done by DeMeo & Carry (2013). Our aim was to remain as close as possible to the original data. Moreover, this conversion requires the colors of the Sun in the SDSS bands, the estimates of which differ between authors (see, e.g., the SDSS estimates³, Holmberg et al. 2006, or Ramírez et al. 2012). We thus converted the class boundaries from DeMeo & Carry (2013, Table 3) into limits in the 3D $g-r$, $r-i$, and $i-z$ color space (Table 3, Fig. 10). As in DeMeo & Carry (2013), some classes overlap in the 3D color space. We, however, deal differently with these overlaps, as described below.

Accounting for uncertainties. For each observation, we computed the volume it occupies in the 3D color space ($g-r$, $g-i$, $i-z$) based on a 3D Gaussian distribution, whose σ are set to color uncertainties.

We then computed a score for each class, k , based on the volume of the intersection between each observation 3D Gaussian with the space occupied by each taxonomic complex (Fig. 10), normalized by the Gaussian volume:

$$V_{\sigma} = \prod_{j=1}^3 \left(\text{erf} \left[\frac{b_j - \mu_j}{\sqrt{2}\sigma_j} \right] - \text{erf} \left[\frac{a_j - \mu_j}{\sqrt{2}\sigma_j} \right] \right), \quad (3)$$

where $\text{erf}(z)$ is the error function, $\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$, the index j indicates the $g-r$, $i-z$, and $g-i$ colors, a_j and b_j are the color boundaries of the complexes, and μ_j and σ_j are the color and uncertainty of the SSO. Hence, for a given observation, the volumes of all intersections sum to one.

A probabilistic approach. These normalized volumes correspond to the probabilities, \mathcal{P}_k , of pertaining to each taxonomic class. An observation fully fitting inside a given class cuboid will have a probability of 1 of being of that class. Conversely, an

³ <https://www.sdss.org/dr16/algorithms/ugrizvegasun/>

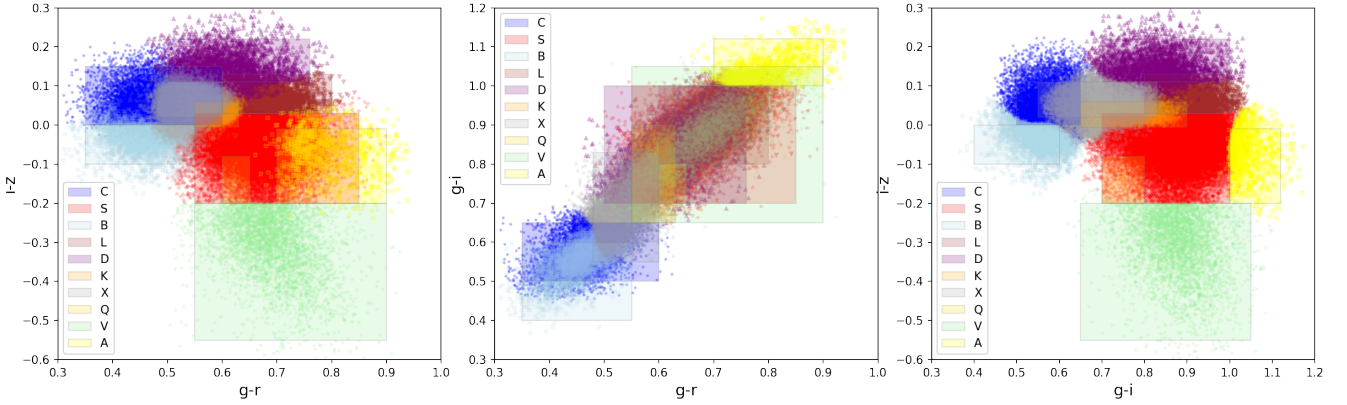


Fig. 10. Taxonomy of SSOs. Boxes are boundaries of the taxonomic classes. Color points mark the classification of the 72 043 asteroid measurements that have color uncertainties smaller than 0.05 mag.

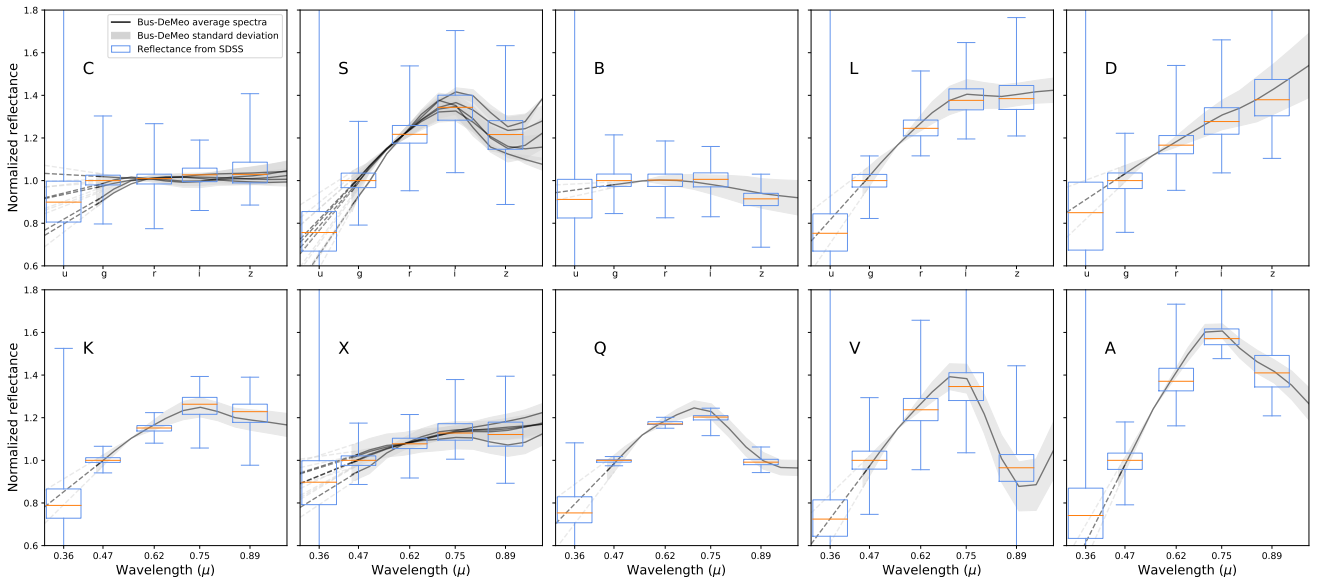


Fig. 11. Pseudo reflectance spectra of asteroids observed by the SDSS, grouped by taxonomic class. The distribution of values for each band is represented by whiskers (95% extrema, and the 25, 50, and 75% quartiles). For each, we also represent the associated template spectra of the Bus-DeMeo taxonomy (DeMeo et al. 2009). The Bus-DeMeo spectra do not cover the wavelength range covered by the u filter, so we present a simple linear extrapolation.

observation whose 3D Gaussian overlaps with two classes in a 1/3–2/3 proportion will have a 0.33 probability of being of the first class and a 0.66 probability of being of the second.

Handling of multiple observations. The probabilistic approach offers a straightforward solution to multiple observations of the same object. The total probability for each class (\mathcal{P}_k) is the sum of the probabilities for that class over all observations (j), weighted by the uncertainty of each observation ($V_{\sigma,j}$), where:

$$w_j = \frac{\sum V_{\sigma,j}^{-2}}{V_{\sigma,j}^{-2}}. \quad (4)$$

Assignment of a class to each asteroid. We assigned to each object its most probable class as derived from all its observations. The sole exception is the unknown class (U), which is only assigned if its probability is strictly equal to 1. Otherwise, whenever U is the most probable (due to poor photometric accuracy for instance), we assigned the second most probable class.

We report in the catalog the probabilities for each class for each observation, together with the most probable class for that

observation. We also list the uncertainty, V_{σ} , and the most probable class after the combination of all the observations of a given SSO.

We present in Fig. 11 the pseudo reflectance spectra of each class, computed using the Solar colors of Holmberg et al. (2006), compared with the template spectra of the associated Bus-DeMeo classes. The correspondence of the median SDSS spectra with the template spectra validates the approach described above. The spread of values is larger for the SDSS, but the sample size is four orders of magnitude larger than the 371 spectra that define the Bus-DeMeo taxonomy.

We also illustrate the catalog by presenting in Fig. 12 the orbital distribution of taxonomic classes. The general trend of S-types dominating the inner belt and C-types the outer belt is clearly visible. The dynamical families are also easy to identify.

The Bus-DeMeo classification scheme we applied is adapted for asteroids but not for comets nor KBOs, which generally display featureless spectra in the visible, with spectral slopes ranging from X-type to D-type and even redder (Snodgrass et al. 2006; Fulchignoni et al. 2008). We did not apply a different

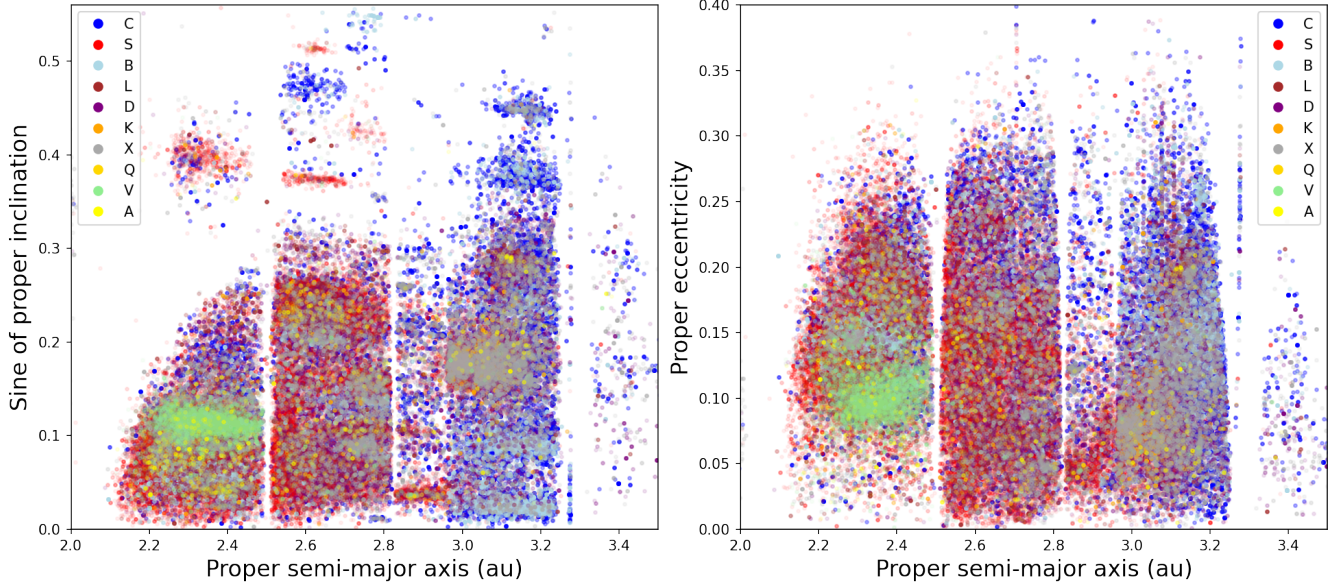


Fig. 12. Orbital distribution of the SSOs reported here, color-coded by taxonomic class.

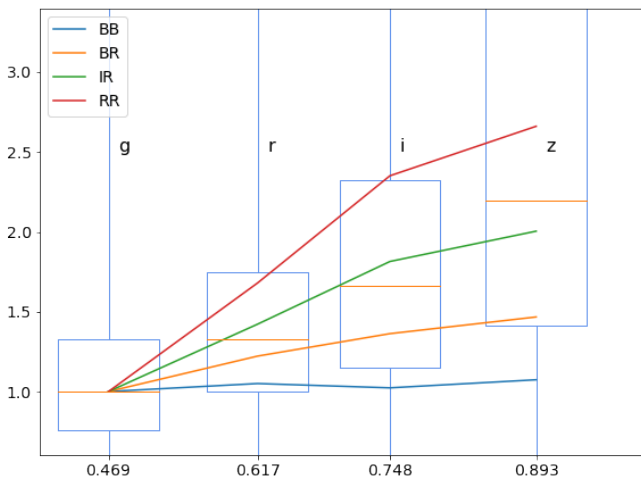


Fig. 13. Pseudo reflectance spectra of KBOs from the SDSS. The BB (blue), BR (intermediate blue-red), IR (moderately red), and RR (red) references are the median $B-V$, $V-R$, and $R-I$ colors from Fulchignoni et al. (2008), converted into SDSS colors.

scheme to comets and KBOs, but we present their pseudo reflectance spectra in Fig. 13 compared with template spectral classes.

8. Conclusions

Based on a new extraction of moving objects in the SDSS, and following a suite of filters to minimize contamination, we release a catalog of 1 542 522 entries, consisting of 1 036 322 observations of 379 714 known and unique SSOs together with 506 200 observations of moving sources not linked with any known SSO. The catalog contains the SDSS identification, astrometry, photometry, SSO identification, geometry of observation, taxonomy, and quality flag for each observation. Its content is fully described in Appendix A.

The catalog completeness is estimated to be about 95% and the purity to be above 95% for known SSOs (see Sect. 5 for details). The present catalog contains 399 502 (85%) of the

471 569 sources released in the ADR4 and 93% of the NEAs and Mars-crossers from Carry et al. (2016). The missing sources are found to be fixed-coordinate sources (identified by comparing their coordinates with the Pan-STARRS catalog using the procedure described in Sect. 3.1) in 58 381 cases. The remaining 13 686 sources were located in fields marked as affected by poor weather conditions, which we excluded from the present analysis (Sect. 2).

Acknowledgements. This research has been conducted within the NEOROCKS project, which has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 870403. Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org>. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington. The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen’s University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under Grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation Grant No. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation. This research made use of the cross-match service provided by CDS, Strasbourg (Pineau et al. 2017), the IMCCE’s SkyBoT and Skybot3D VO tools (Berthier et al. 2006, 2016), and TOPCAT/STILTS (Taylor 2005). Thanks to the developers.

References

- Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2008, *ApJS*, **175**, 297
- Ahumada, R., Allende Prieto, C., Almeida, A., et al. 2020, *ApJS*, **249**, 3
- Berthier, J., Vachier, F., Thuillot, W., et al. 2006, *ASP Conf. Ser.*, **351**, 367
- Berthier, J., Carry, B., Vachier, F., Eggl, S., & Santerne, A. 2016, *MNRAS*, **458**, 3394
- Binzel, R. P., DeMeo, F. E., Turtelboom, E. V., et al. 2019, *Icarus*, **324**, 41
- Bottke, W. F., Jr, Nesvorný, D., Grimm, R. E., Morbidelli, A., & O'Brien, D. P. 2006, *Nature*, **439**, 821
- Brown, A. G. A., Vallenari, A., Prusti, T., et al. 2018, *A&A*, **616**, A1
- Bus, S. J., & Binzel, R. P. 2002a, *Icarus*, **158**, 146
- Bus, S. J., & Binzel, R. P. 2002b, *Icarus*, **158**, 106
- Carry, B. 2018, *A&A*, **609**, A113
- Carry, B., Solano, E., Eggl, S., & DeMeo, F. 2016, *Icarus*, **268**, 340
- Carvano, J. M., Hasselmann, P. H., Lazzaro, D., & Mothé-Diniz, T. 2010, *A&A*, **510**, A43
- Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, ArXiv e-prints [arXiv:1612.05560]
- Clement, M. S., Morbidelli, A., Raymond, S. N., & Kaib, N. A. 2020, *MNRAS*, **492**, L56
- DeMeo, F. E., & Carry, B. 2013, *Icarus*, **226**, 723
- DeMeo, F., & Carry, B. 2014, *Nature*, **505**, 629
- DeMeo, F., Binzel, R. P., Slivan, S. M., & Bus, S. J. 2009, *Icarus*, **202**, 160
- DeMeo, F., Binzel, R. P., Carry, B., Polishook, D., & Moskovitz, N. A. 2014, *Icarus*, **229**, 392
- DeMeo, F., Polishook, D., Carry, B., et al. 2019, *Icarus*, **322**, 13
- Devogèle, M., Moskovitz, N., Thirouin, A., et al. 2019, *AJ*, **158**, 196
- Flewelling, H. A., Magnier, E. A., Chambers, K. C., et al. 2020, *ApJS*, **251**, 7
- Fornasier, S., Lantz, C., Barucci, M. A., & Lazzarin, M. 2014, *Icarus*, **233**, 163
- Frieman, J. A., Bassett, B., Becker, A., et al. 2008, *AJ*, **135**, 338
- Fulchignoni, M., Belskaya, I., Barucci, M. A., De Sanctis, M. C., & Doressoundiram, A. 2008, *Solar Syst. Beyond Neptune*, 181
- Granvik, M., & Brown, P. 2018, *Icarus*, **311**, 271
- Granvik, M., Morbidelli, A., Jedicke, R., et al. 2018, *Icarus*, **312**, 181
- Graves, K., Minton, D., Hirabayashi, M., DeMeo, F., & Carry, B. 2018, *Icarus*, **304**, 162
- Gunn, J. E., Carr, M., Rockosi, C., et al. 1998, *AJ*, **116**, 3040
- Holmberg, J., Flynn, C., & Portinari, L. 2006, *MNRAS*, **367**, 449
- Ivezić, Ž., Tabachnik, S., Rafikov, R., et al. 2001, *AJ*, **122**, 2749
- Ivezić, Z., Juric, M., Lupton, R. H., et al. 2010, *NASA Planetary Data System, EAR*
- Jones, R. L., Chesley, S. R., Connolly, A. J., et al. 2009, *Earth Moon Planets*, **105**, 101
- Lazzaro, D., Angeli, C. A., Carvano, J. M., et al. 2004, *Icarus*, **172**, 179
- Levison, H. F., Bottke, W. F., Gounelle, M., et al. 2009, *Nature*, **460**, 364
- Mahlke, M., Bouy, H., Altieri, B., et al. 2018, *A&A*, **610**, A21
- Mahlke, M., Solano, E., Bouy, H., et al. 2019, *Astron. Comput.*, **28**, e100289
- Masiero, J. R., Mainzer, A. K., Grav, T., et al. 2011, *ApJ*, **741**, 68
- Morbidelli, A., Walsh, K. J., O'Brien, D. P., Minton, D. A., & Bottke, W. F. 2015, *The Dynamical Evolution of the Asteroid Belt* (Tucson: University of Arizona Press), 493
- Moskovitz, N. A., Jedicke, R., Gaidos, E., et al. 2008, *Icarus*, **198**, 77
- Nesvorný, D., Jedicke, R., Whiteley, R. J., & Ivezić, Ž. 2005, *Icarus*, **173**, 132
- Oszkiewicz, D. A., Kwiatkowski, T., Tomov, T., et al. 2014, *A&A*, **572**, A29
- Parker, A., Ivezić, Ž., Jurić, M., et al. 2008, *Icarus*, **198**, 138
- Perna, D., Barucci, M. A., Fulchignoni, M., et al. 2018, *Planet. Space Sci.*, **157**, 82
- Pineau, F. X., Derriere, S., Motch, C., et al. 2017, *A&A*, **597**, A89
- Popescu, M., Licandro, J., Morate, D., et al. 2016, *A&A*, **591**, A115
- Ramírez, I., Michel, R., Sefako, R., et al. 2012, *ApJ*, **752**, 5
- Raymond, S. N., & Izidoro, A. 2017, *Sci. Adv.*, **3**
- Rivkin, A. S. 2012, *Icarus*, **221**, 744
- Sako, M., Bassett, B., Becker, A., et al. 2008, *AJ*, **135**, 348
- Scott, E. R. D., Krot, A. N., & Sanders, I. S. 2018, *ApJ*, **854**, 164
- Siegel, A. 1982, *Biometrika*, **69**, 242
- Snodgrass, C., Lowry, S. C., & Fitzsimmons, A. 2006, *MNRAS*, **373**, 1590
- Solano, E., Rodrigo, C., Pulido, R., & Carry, B. 2014, *Astron. Nachr.*, **335**, 142
- Spoto, F., Tanga, P., Mignard, F., et al. 2018, *A&A*, **616**, A13
- Szabó, G. M., Ivezić, Ž., Jurić, M., Lupton, R., & Kiss, L. L. 2004, *MNRAS*, **348**, 987
- Taylor, M. B. 2005, *ASP Conf. Ser.*, **347**, 29
- Thomas, C. A., Trilling, D. E., & Rivkin, A. S. 2012, *Icarus*, **219**, 505
- Vernazza, P., Zanda, B., Binzel, R. P., et al. 2014, *ApJ*, **791**, 120
- Vokrouhlický, D., Bottke, W. F., & Nesvorný, D. 2016, *AJ*, **152**, 39
- Willmer, C. N. A. 2018, *ApJS*, **236**, 47
- Xu, S., Binzel, R. P., Burbine, T. H., & Bus, S. J. 1995, *Icarus*, **115**, 1
- Zellner, B., Tholen, D. J., & Tedesco, E. F. 1985, *Icarus*, **61**, 355

Appendix A: Description of the SSO catalog**Table A.1.** Description of the SDSS SSO catalog columns.

ID	Name	Unit	Description
1	objID		Unique SDSS identifier
2	run		SDSS Run number
3	camcol		SDSS Camcol number
4	field		SDSS Field number
5	ra	deg	J2000 Right Ascension (<i>r</i> -band)
6	dec	deg	J2000 Declination (<i>r</i> -band)
7	tai_r	day	MJD time of observation (TAI) in <i>r</i> filter
8	jd	day	Julian time (UTC)
9	rowv	arcsec/h	Row-component of object's velocity
10	rowvErr	arcsec/h	Row-component of object's velocity error
11	colv	arcsec/h	Column-component of object's velocity
12	colvErr	arcsec/h	Column-component of object's velocity error
13	vel	arcsec/h	Object's velocity
14	velErr	arcsec/h	Object's velocity error
15	type_r		SDSS Type classification of the object (3 – galaxy, 6 – star)
16	psfMag_u	mag	PSF magnitude in each filter
17	psfMag_g	mag	
18	psfMag_r	mag	
19	psfMag_i	mag	
20	psfMag_z	mag	
21	psfMagErr_u	mag	PSF magnitude error in each filter
22	psfMagErr_g	mag	
23	psfMagErr_r	mag	
24	psfMagErr_i	mag	
25	psfMagErr_z	mag	
26	offsetRa_u	arcsec	Filter positions RA and Dec minus final coordinates
27	offsetDec_u	arcsec	
28	offsetRa_g	arcsec	
29	offsetDec_g	arcsec	
30	offsetRa_r	arcsec	
31	offsetDec_r	arcsec	
32	offsetRa_i	arcsec	
33	offsetDec_i	arcsec	
34	offsetRa_z	arcsec	
35	offsetDec_z	arcsec	
36	Number		Solar System object number
37	Name		Solar System object name
38	ra_sb	deg	SkyBot predicted RA coordinate
39	dec_sb	deg	SkyBot predicted Dec coordinate
40	dynclass		SkyBot object dynamic class
41	V	mag	SkyBot predicted visual magnitude
42	posunc	arcsec	Positional uncertainty
43	centerdist	arcsec	Angular distance of target from cone center
44	ra_sb_rate	arcsec/h	RA rate of motion
45	dec_sb_rate	arcsec/h	Declination rate of motion
46	geodist	au	Geocentric distance of target
47	heliodist	au	Heliocentric distance of target
48	alpha	deg	Solar phase angle
49	elong	deg	Solar elongation angle
50	boffset		Offset flag

Notes. The catalog is available at the CDS.

Table A.1. continued.

ID	Name	Unit	Description
51	bphot_u		
52	bphot_g		
53	bphot_r		Photometry flags
54	bphot_i		
55	bphot_z		
56	bcolor_ru		
57	bcolor_rg		Color flags
58	bcolor_ri		
59	bcolor_rz		
60	bastrom_i		
61	bastrom_u		Astrometry flags
62	bastrom_z		
63	bknown		Known SSO
64	R2		Coefficient of determination
65	Volume		Color uncertainty
66	p[C, S, B, L, D, K, X, V, A, U]		Probability of complex values
67	complex		Most probably complex of individual observation
68	pcomplex		Complex probably value of individual observation
69	gr_complex		Most probably complex of SSO
70	gr_pcomplex		Complex probably value of SSO
71	proper_H	mag	Absolute magnitude (from astdys)
72	proper_a	au	Proper semimajor axis (from astdys)
73	proper_e		Proper eccentricity (from astdys)
74	proper_sinI		Sine of proper inclination (from astdys)
75	osc_H	mag	Absolute magnitude (from astorb)
76	osc_a	au	Semimajor axis (from astorb)
77	osc_inc		Eccentricity (from astorb)
78	osc_e	deg	Inclination (from astorb)

Appendix B: SQL code for moving object extraction

```

FROM dbo.PhotoObjAll as p
WHERE
  p.flags & dbo.fPhotoFlags('
  DEBLENDED_AS_MOVING') > 0
  AND p.flags & dbo.fPhotoFlags('SATURATED') =
  0
  AND p.flags & dbo.fPhotoFlags('BRIGHT') = 0
  AND p.flags & dbo.fPhotoFlags('EDGE') = 0
  AND (p.type_r = 6 OR p.type_r = 3)
  AND power(p.rowv,2) + power(p.colv, 2) >
  0.0025
  AND p.petroMag_r BETWEEN 14.5 AND 22.2

```

Listing 1. SDSS request.

```

SELECT source_id, ra, dec, coord1(prop) AS
  ra_2004, coord2(prop) AS dec_2004, pmra,
  pmdec, phot_g_mean_mag
FROM (SELECT gaia.source_id, ra, dec, pmra,
  pmdec, phot_g_mean_mag,
  EPOCH_PROP_POS(ra, dec, parallax,
  pmra, pmdec, 0, ref_epoch, 2004) AS prop
  FROM gaiadr2.gaia_source AS gaia
  phot_g_mean_mag>=14 AND
  sqrt(power(gaia.pmra, 2) + power(
  gaia.pmdec, 2)) > 10
  ) AS subquery

```

Listing 2. GAIA request.