The Solar System Notification Alert Processing System (SNAPS): Design, Architecture, and First Data Release (SNAPShot1)

David E. Trilling[1,2] , Michael Gowanlock[2,1] , Daniel Kramer[2,1] , Andrew McNeill[1,3] , Brian Donnelly[2], Nat Butler[4] , and
John Kececioglu[5]

[1] Department of Astronomy and Planetary Science, Northern Arizona University, P.O. Box 6010 Flagstaff, AZ, 86011, USA
[2] School of Informatics, Computing, and Cyber Systems, Northern Arizona University, P.O. Box 5693, Flagstaff, AZ, 86011, USA
[3] Department of Physics, Lehigh University, 16 Memorial Drive East, Bethlehem, PA, 18015, USA
[4] School of Earth and Space Exploration, Arizona State University, Tempe, AZ, 85287, USA
[5] Department of Computer Science, University of Arizona, Tucson, AZ, 85721, USA

Abstract

We present here the design, architecture, and first data release for the Solar System Notification Alert Processing System (SNAPS). SNAPS is a solar system broker that ingests alert data from all-sky surveys. At present, we ingest data from the Zwicky Transient Facility (ZTF) public survey, and we will ingest data from the forthcoming Legacy Survey of Space and Time (LSST) when it comes online. SNAPS is an official LSST downstream broker. In this paper we present the SNAPS design goals and requirements. We describe the details of our automatic pipeline processing in which the physical properties of asteroids are derived. We present SNAPShot1, our first data release, which contains 5,458,459 observations of 31,693 asteroids observed by ZTF from 2018 July to 2020 May. By comparing a number of derived properties for this ensemble to previously published results for overlapping objects we show that our automatic processing is highly reliable. We present a short list of science results, among many that will be enabled by our SNAPS catalog: (1) we demonstrate that there are no known asteroids with very short periods and high amplitudes, which clearly indicates that in general asteroids in the size range 0.3–20 km are strengthless; (2) we find no difference in the period distributions of Jupiter Trojan asteroids, implying that the L4 and L5 clouds have different shape distributions; and (3) we highlight several individual asteroids of interest. Finally, we describe future work for SNAPS and our ability to operate at LSST scale.

Unified Astronomy Thesaurus concepts: Asteroids (72); Small Solar System bodies (1469); Sky surveys (1464); Catalogs (205)

1. Introduction

1.1. Asteroid Science from Sky Surveys

Asteroids are the most numerous objects in the solar system and as such act as tracers of the dynamical and physical evolution of our planetary system. At present, there are more than one million known asteroids. A comprehensive understanding of the evolution of our solar system can therefore use the individual properties of asteroids as fine-grained measurements from which broader conclusions can be drawn. Asteroids can be used to understand the dynamical and collisional history of the solar system, the intrinsic material properties of primitive bodies, and help unravel the origin of life on Earth through understanding the degree to which asteroids brought water and organics to the early Earth. Individual asteroids are akin to pixels in a large image, and by building up our knowledge, one pixel at a time, the entire picture is revealed.

To gain this global understanding of the solar system through asteroids, large data sets are required. The Vera C. Rubin Observatory will carry out the Legacy Survey of Space and Time (LSST) starting in 2024 and will revolutionize many fields of astronomy. LSST will observe more than five million main belt asteroids (LSST Science Collaboration et al. 2009). It is not only the number of asteroids to be observed that is

significant, but the fact that most asteroids will be observed many hundreds of times, under relatively uniform conditions. This data set will be far richer—and far more scientifically valuable—than existing catalogs of asteroid measurements. However, the LSST data by itself (time, brightness, position) will not provide new insight into the processes that have driven the evolution of our solar system. To reap the scientific reward from this data set, the scientific context and physical properties of the entire asteroid population must be understood.

The current LSST schedule of first light in 2023 and science operations in 2024 means that the next steps in understanding the history of the solar system are not far away. This has motivated us to develop the necessary infrastructure now, to be prepared for the arrival of LSST data. Fortunately, the ongoing Zwicky Transient Facility (ZTF; Bellm et al. 2019) survey is a present-day analog of and precursor to LSST. We are using data from ZTF to develop, hone, and prepare for the coming LSST era.

All-sky surveys enable two important research goals: (1) understanding global trends and (2) looking for rare events, as follows. In category 1, by deriving the physical properties of millions of asteroids we can understand the global properties of the asteroid belt that reveal information about the formation and evolution of our solar system. These are signatures that are only revealed through very large sample sizes. Examples here could include measuring asteroid lightcurve periods and amplitudes as a function of color (and implied composition), where the required densities and strengths may differ between rocky S-type asteroids and more primitive C-type asteroids.

Large data sets can enable detecting subtle signatures, or slicing data sets (example: as a function of size) to reveal trends.

Objects detected in category 2—detecting and characterizing rare events—also can place constraints on solar system formation through probing subtle signatures that can indicate important processes. We refer to objects that exhibit rare behaviors as outliers, and there are two general kinds. Real-time outliers (category 2a) are objects whose properties change on short timescales. An example of this case is an asteroid that is found to be active. Population outliers (category 2b) are objects whose fundamental properties are unusual compared to the entire population. Examples here could include asteroids with very large lightcurve amplitudes, very short or long rotation periods, or asteroids with very unusual colors.

Studying outlier asteroids allows us to understand these rare events. As an example in category 2a, it is generally thought that most activity among asteroids is driven by the sublimation of volatiles, but the occurrence rate of active asteroids is not well known (e.g., Jewitt et al. 2015). Measuring this rate from a uniform survey would improve our understanding of the volatile content of the asteroid belt and provide important new evidence for understanding the formation of the solar system and the origin of life on Earth. (The LOOK project (Lister et al. 2022) has some goals and approaches that overlap with SNAPS, though LOOK has an overall greater emphasis on cometary science.) Thus, real-time outlier detections provide a direct impact on these important outstanding science questions. Population outliers (category 2b) may also provide insight into the formation and history of the solar system. For example, a small number of asteroids have very high-amplitude light-curves or very long rotation periods, both of which may indicate the need for significant internal strength (e.g., McNeill et al. 2018a and references therein). This too has implications for how asteroids formed, and how the solar system formed.

1.2. ZTF and LSST

ZTF does, and LSST will, broadcast alerts that report transient objects on the sky. These variable sources include supernovae, variable stars, and moving objects (and many other kinds of astrophysically variable objects). Moving objects are considered variable here because the brightness at a given sky location changes as a function of time, as an asteroid passes to and from that location. An alert includes measured data, some properties of the observed source, a postage stamp, and metadata, and alerts are distributed through a Kafka stream. LSST will broadcast alerts for around 10,000 variable sources detected in each 30 s visit.[6] Moving objects may be 10%–50% of all variable sources in a given LSST visit, depending primarily on the ecliptic latitude. The ZTF data flow is about one tenth of this rate, with some 1000 alerts, and 100 asteroids, per visit, which corresponds to around 10,000 asteroids each night. The alerts (will) include measured properties of moving objects, including date/time, photometric properties (magnitude, uncertainty), a real/bogus score (likelihood that the source is real and is a point source), and properties of the point-spread function (PSF; e.g., elongation or extendedness). In the work described here, we focus only on known objects; in this case, the identity of the object is also transmitted with the measurement data. The orbital elements of each reported object are therefore known; in the work presented here, the orbital

elements are not transmitted with the measurements and we obtain them through cross-referencing the Minor Planet Center's catalogs.

At full scale, LSST will report measurements of 10^5–10^6 asteroids every night, for ten years. The total observational record, over ten years, will be a few billion unique measurements of more than five million unique asteroids. LSST is a carefully defined project, and key definitions for the LSST project—what LSST will and will not do—are clearly defined (see Ivezić et al. 2019 and many documents at the LSST Documentation Hub[7]). This data stream will enable science, but will not produce science. In order to carry out a wide range of deep scientific investigations, the detailed physical properties of these millions of asteroids must be derived outside of the LSST project.

This flood of data can only be handled with sophisticated automated software. The general approach to handling solar system objects, or indeed any transient or time-varying science, is through brokers: software packages that automatically ingest the LSST (or ZTF) alert stream. Brokers have some intelligence and may have the ability to derive target properties, recognize outlier behavior, attach external measurements to create value-added data records, and issue further alerts. There are a number of brokers that are under development and that are oriented toward sidereal or astrophysical sources, or are general and will include some basic processing of solar system observations.[8]

Individual asteroid measurements such as magnitude or position by themselves are generally not scientifically interesting (except for extendedness, as described below). However, for asteroids with enough measurements, additional properties can be derived, including solar system absolute magnitude (H), photometric slope parameter (G), lightcurve period and amplitude, color, taxonomic classification, as estimated from color, and even albedo and diameter, as estimated from color (Ivezić & Ivezić 2021).

1.3. This Paper: SNAPS

In this paper we present SNAPS—the Solar System Notification Alert Processing System. SNAPS was designed from the outset to be a moving object broker and hence has a number of capabilities that specifically enable solar system science. SNAPS has been designated an official downstream broker. Here we present, in Section 2, the SNAPS design and architecture and our technical work completed to date. In Section 3 we present SNAPShot1, our first data release. In Section 4 we demonstrate the accuracy of our automatically derived results. Section 5 presents several science results, among the many that will be enabled by our SNAPS catalog. In Section 6 we discuss the scalability of our approach to LSST operations, and Section 7 describes the next steps for SNAPS.

2. SNAPS Design

The SNAPS broker has been designed to ingest the ZTF alert stream and enable science, as well as act as a testbed for broker needs for the LSST era. Over the past four years, since the beginning of the ZTF real-time alert broadcast in 2018 July, we have ingested nearly 20 million observations of nearly
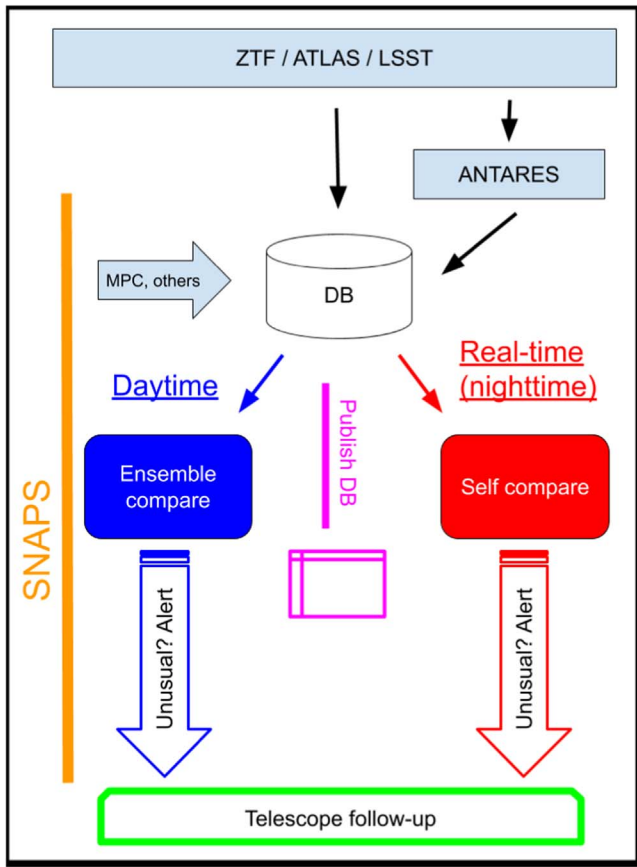

Figure 1. Flowchart for SNAPS. The scope of the overall SNAPS project is indicated by the orange bar on the side. The main SNAPS facets are (a) ingesting data, from a survey or from an upstream broker; (b) detecting real-time outliers [red path here]; (c) detecting population outliers [blue path here]; and (d) disseminating results. Telescope follow-up may be appropriate but is not formally a part of SNAPS.

600,000 unique asteroids. In this section we present various elements of the overall design of SNAPS.

2.1. Overview

The high-level description of SNAPS is that it must ingest data from an alert stream, update the database with this recent data, search for both real-time and population outliers, and disseminate results through both an alert stream and a publicly accessible web portal. The architecture for SNAPS is shown in Figure 1. Our broker has four primary facets: (a) we listen to an alert stream, either directly from the source (ZTF/LSST), or through an intermediate broker (in this case, ANTARES; Matheson et al. 2021). We also incorporate adjacent databases (e.g., MPC, SDSS, NEOWISE, 2MASS, etc.); (b) we analyze in real time new observations, and issue alerts if unusual properties are detected (red path in Figure 1). Typically this done by comparing the object's observed behavior with its expected behavior; (c) during the daytime each object is compared to the ensemble, and alerts are triggered about objects with unusual properties; and (d) a value-added database with derived properties for all asteroids is published and world-accessible. This database enables a wide range of science investigations.

Element (a) is fully mature and is presented here. Element (b) is under development, and will be reported in a future paper. Element (c) is presently carried out, though in a manual, rather

than automated or algorithmic, way. Element (d) is partially complete: the value-added database exists, and the public database portal and alert stream are under development.

2.2. Description of SNAPS Elements

2.2.1. Element (a): Ingesting Data

ZTF broadcasts an alert stream[9] with measurements and metadata for all transient sources detected in an image. Around 10% of all transient sources are solar system objects. We could listen directly to this alert stream, and discard the 90% of alerts that are irrelevant for us, but we choose a different approach instead. The ANTARES broker (Matheson et al. 2021) is broadly interested in all astrophysical transients, though not solar system objects. ANTARES listens to the full ZTF alert stream. As part of their processing, they rebroadcast solar system alerts on a dedicated Kafka stream. It is this stream that we ingest.

One of the first steps that ANTARES carries out, before rebroadcasting solar system objects, is removing only observations with real/bogus scores <0.55. ZTF defines its real/bogus scores in the range [0, 1] where 0 is not a point source and 1 is a point source (Duev et al. 2019). Since the ZTF real/bogus classifier was trained on point sources, real extended objects can receive low real/bogus scores and would not be distributed by ANTARES. To ingest these "lost" observations, we download and ingest the data from ZTF's alert archive.[10] For the data presented in this paper, however, we are not searching for activity, and the ANTARES rebroadcast is sufficient.

Each individual measurement has a large number of metadata properties associated with it; the relevant ones that we store in our database are shown in Table 1.

2.2.2. Element (b): Real-time Analysis

After ingesting the data, we retrieve object ephemerides from JPL Horizons[11] by using the jplhorizons module in astroquery (Ginsburg et al. 2019). Using geocentric and heliocentric distance and phase angle we calculate the object's absolute magnitude in the Bowell HG system (Bowell et al. 1989) using either a previously derived G value if one exists, or else by assuming G = 0.15. Light time corrections are also retrieved, to be used in our downstream analysis. These values are stored in our database (Table 1).

One goal of our real-time analysis is to compare a new observation with the prediction for that observation and to identify significant differences. The measured properties to be considered here include magnitude and PSF shape. This element of SNAPS is not fully mature at this time.

To calculate the predicted magnitude, we will use our derived solar system absolute magnitude (H) and slope parameter (G) and the observational geometry to define the expected magnitude, which is then modulated by the object's lightcurve (described in the next section). Clearly, our predictions are more accurate for objects with a large number of observations, where all of H, G, and lightcurve are well determined.

The vast majority of objects have PSF shapes that are round; this corresponds to ZTF elong values of 1.0. Therefore, an

---

[9] https://www.ztf.caltech.edu/ztf-alert-stream.html
[10] https://ztf.uw.edu/alerts/public/
[11] https://ssd.jpl.nasa.gov/horizons/

Table 1
Properties of Individual Observations That are Stored in the SNAPS Database

Parameter	Definition
Image information	
jd	Julian date
fid	Filter ID (a numberical value showing g,r,i)
ra	Right ascension
dec	Decl.
night	Date (YYYMMDD) of observation
Source information	
magpsf	Magnitude from PSF fitting
sigmagpsf	1σ uncertainty in magpsf
chipsf	Reduced χ^2 for PSF fit
magap	Magnitude in 8 pixel aperture
sigmagap	1σ uncertainty in magap
magapbig	Magnitude in 18 pixel aperture
sigmagapbig	1σ uncertainty in magapbig
elong	Ratio of semimajor to semiminor axis of the source
rb	Real/bogus score [0 bad, 1 good]—likelihood of true point source
ssnamenr	Name of nearest known solar system object
id	ZTF identifier for this observation
Values from Horizons	
obsdist	Distance from source to observer
heliodist	Distance from source to Sun
phaseangle	Observer-source-Sun angle
ltc	Light time correction—calculated

Note. Image and source information are extracted from the ZTF (or LSST) alert; the descriptions for the ZTF columns are available at https://zwickytransientfacility.github.io/ztf-avro-alert/schema.html. Using ssnamenr we retrieve relevant geometry information from JPL's Horizons service.

observation with an elong value that differs significantly from unity may be of interest. It is also possible that an object that is known to be active has significant changes in its elong values, indicating either a significant increase or decrease in activity. In principle, these changes to the PSF shape can be determined with a very small number of observations in our database, and the algorithmic implementation is relatively straightforward and is under way.

As described above, the ZTF real/bogus algorithm was trained on point sources. Therefore, asteroids that show signs of activity—in other words, objects that are expected to be point sources but are not—may receive low rb scores, and these anomalous observations will also be investigated.

2.2.3. Element (c): Daytime Analysis

The daytime analysis is designed to enable science results that are not highly time critical. We (will) produce results both for individual objects and for the entire population.

Individual object processing. Objects with more than 50 observations are processed with two separate pipelines to calculate lightcurve properties and phase curves. (We do not attempt to derive properties for objects with fewer than 50 observations.) The lightcurve pipeline is used to derive an object's color and lightcurve properties (period and amplitude). The phase curve calculation pipeline is used to derive the absolute magnitude and phase parameter.

Deriving lightcurves. Lightcurves are calculated in the following way.

1. Observations are split into groups by filter.
2. Each filter group is processed through a Lomb–Scargle periodogram (LSP) approach (Lomb 1976; Scargle 1982) using a uniform period grid of 10^6 periods, ranging from 1 to 2500 hr. These values are lightcurve periods; since an asteroid typically executes two lightcurve periods in a single rotation period, the rotation period search range is 2–5000 hr. (It would be very interesting to identify asteroids with rotation periods less than 2 hr—so-called super fast rotators (SFRs), which are unusual and generally require nonzero strength to avoid catastrophic disruption. However, the sparse ZTF cadence is not well suited to finding SFRs, and our synthetic population studies, described below, show many false positives rotation periods less than 2 hr. We therefore use a lower limit of 2 hr in this paper, and defer the more complex topic of searching for SFRs in ZTF/SNAPS to a subsequent paper.)
3. The resulting periods derived separately for each filter are compared. If they are the same, that period is then used as the object's period. If they are not the same, then the filter with the most observations (for ZTF, this is usually r) is used as the object's period.
4. A Fourier model is then fit to the g and r data separately using the derived period. Using the median of each Fourier model, a g − r color can be derived.
5. Using the derived color, all data are combined (that is, the g data are shifted using the derived g − r color to overlay it on the unshifted r band data). These combined data are used to derive a refined period solution.
6. Another Fourier model (LSP) is fit to the combined multiband data and the amplitude of this solution is treated as the amplitude of the object's lightcurve.
7. The derived period, amplitude, Fourier parameters, and g − r color are then stored in our database (Table 2).

All ground-based surveys are subject to aliasing in period searches due to the unavoidable 24 hr day/night cycling of potential observing times. Nearly 20% of our lightcurve period solutions have values in the ranges 11–13 hr, 15–17 hr, 23–25 hr, or 47–49 hr, with the largest contribution from periods near 16 hr. That the dominant alias is at 16 hr, and not 12 or 24 hr, is likely due to details of the ZTF cadence. Our analysis of these alias features, and our overall approach to de-aliasing, will be presented in a forthcoming paper (D. Kramer et al. 2023, in preparation); for this data release, no attempt at de-aliasing is made. Users are cautioned that a substantial fraction of solutions near these common periods (12, 16, 24, 48 hr) are likely to be aliases and not the true rotation periods, although certainly there are some asteroids whose true periods are near these values. At present, it is impossible to distinguish (false) alias solutions from (true) rotation periods near these common periods (12, 16, 24, 48 hr). Nevertheless, for all of these objects the g − r color and amplitude are most likely correct.

Phase curve calculations. We carry out the following steps to derive phase curve solutions for each object. Steps 1 and 2 are carried out separately for the data in each filter.

1. Our initial solution for each asteroid is H = 10 and G = 0.15. From these values we iteratively adjust and

Table 2
Derived Properties of Individual Objects and Values from External Catalogs
That are Stored in the SNAPS Database

Parameter	Definition
Derived properties	
H	Solar system absolute magnitude in g and r, with uncertainties
G	Solar system phase parameter in g and r, with uncertainties
g-r color	Mean color and uncertainty
lcper	Derived lightcurve period
lcamp	Derived lightcurve amplitude
peakpower	Peak power in the periodogram
residPeriod	The derived LS period for the residuals of the generated lightcurve model
mag18/mag8	The max, mean, standard deviation, and most recent mag18/mag8 value for the object
Information from external catalogs	
ugriz photometry	SDSS
JHK photometry	2MASS
diameter and albedo	WISE/NEOWISE
orbital elements; H, G	Minor Planet Center
Rotation period and quality code	Lightcurve database

derive best-fit H and G values using SciPy's minimize function, $1 - r_P^2$, where r_P is the Pearson correlation coefficient (Freedman et al. 2007). The search bounds for G are (−0.302, 0.907), which are 10% (less than, greater than) the (minimum, maximum) values from Vereš et al. (2015). (The minimization process also requires bounds on H; we use [4.7, 29.3], well beyond the expected values for any SNAPS targets.) The minimize function returns, among other values, G, H, and the success of the minimization.

2. Step 1 is repeated 30 times. For each of these trials, the magnitudes for each observation are randomized with a Gaussian distribution within the reported observational errors. The standard deviations of the results from these 30 trials are used as σ_H and σ_G.

3. If the minimizations in both filters succeed, then the two derived G values are combined, using an average weighted by the number of observations in each filter. This step does not add much in the case of ZTF with only two filters but could be useful for other data sets (i.e., LSST) where solutions exist in three or more filters.

4. The derived G, H, and σ values and the success of each minimization are stored in the database.

Population studies. There are very few examples of outlier asteroids in the literature. Consequently, we use an unsupervised approach to detect asteroids that are outliers relative to the population of objects in our ZTF database. We create a feature vector for each asteroid in our database using 15 properties selected from Tables 1 and 2, where example feature vectors are given in Table 3. We use these features as input to several unsupervised outlier detection algorithms. At present, each object is assigned an outlier ranking, $r_i ä [1, N_{objects}]$ (here, $N_{objects} = 31,693$). Object i with $r_i = 1$ refers to

the object with the greatest outlier score, whereas the object with $r_i = 31,693$ is denoted as the most typical object in the database.

SNAPS currently ranks each object using two outlier detection methods, although we will expand on the number of outlier detection methods in the future. Our aim is to use an ensemble of methods for outlier detection, as each method will derive disparate sets of outliers, as they are sensitive to different aspects of the data. We combine the output of these methods to derive an average outlier score for each object.

We use k-nearest neighbors (kNN) as an outlier detection method (Hautamaki et al. 2004). Intuitively, an object in the feature space will be located in a sparsely populated region if it is an outlier, and it will be located in a dense region if it is an inlier. We use kNN to detect outliers using two different methods:

1. Mean Distance to Neighbors: we search the database and derive the kNN for each object. For each of the k neighbors found, we compute each object's mean distance to its neighbors. We assign each object an outlier ranking (described above) where the object with the greatest distance to its k neighbors is assigned $r_i = 1$.

2. Reverse kNN: consider that objects that are frequently in the set of kNN for many other objects are likely to be inliers, whereas those objects that are rarely found in the set of kNN for other objects are outliers. This is computed using the reverse kNN (RkNN), where a query feature vector finds all of the instances where it is found in the kNN set of all of the other feature vectors in the database (Tao et al. 2007). This is also known as the in-degree of a kNN graph. Each object is assigned a ranking described above, where $r_i = 1$ denotes the object with the smallest in-degree in the data set.

In addition to using kNN, we also use distance similarity searches to derive outlier scores. This algorithm simply searches a fixed radius, ò, and returns those neighbors that are found within the search radius (Gowanlock & Karsin 2019). The object with the fewest number of neighbors within ò has the greatest outlier score, $r_i = 1$, and the object with the greatest number of neighbors within ò is the most typical object in the database.

While the kNN and distance similarity search methods output an outlier ranking, we also plan to use other methods to detect outliers, including those that return a binary classification (inlier or outlier). An example is density-based clustering (Gowanlock 2019), where inliers are assigned to clusters, and outliers are assigned to a set of noise points.

Performance is critical for outlier detection tasks, as all-pairs searches have a worst-case quadratic time complexity. We are leveraging our prior work in this area to perform fast kNN searches, distance similarity searches, and clustering, which have been parallelized using GPUs, and has been published in the computer science literature (Gallet & Gowanlock 2019; Gowanlock 2019; Gowanlock & Karsin 2019; Gowanlock et al. 2021).

2.2.4. Element (d): Disseminate Results

We attach data from external databases to the ZTF derived data catalog to create a value-added product. At present we include data from the following static databases: 2MASS (JHK photometry of around 20,000 asteroids; Sykes et al. 2020)

Table 3
Four Example Feature Vectors in SNAPShot1, Where a Subset of 15 Features from Tables 1 and 2 are Shown Here

Object	lcamp (mag)	lcper (hr)	g-r color (mag)	peakpower ()	diameter (km)	K	albedo
152	0.229	6.245	0.679	0.695	57.8		0.247
168	0.161	46.973	0.472	0.595	144.1		0.045
198	0.151	8.532	0.641	0.428	50.9		0.293
227	0.165	15.969	0.473	0.343	109.3		0.046

and SDSS (ugriz photometry of around 40,000 asteroids; Ivezic et al. 2020). We also include data from two dynamic (that is, continually or occasionally updated) databases: WISE/NEO-WISE (two or four band thermal infrared photometry and derived diameter and albedo for around 100,000 asteroids; Mainzer et al. 2019); and the Minor Planet Center (orbital elements and absolute magnitudes for all known solar system objects[12]). We have also ingested solutions[13] from the (asteroid) Lightcurve Database (LCDB; Warner et al. 2009), though this compendium is not maintained and updated as uniformly as the other catalogs listed above.

To disseminate our results we will deploy a web interface as well as an API; these elements are not fully mature as of this writing. The final development step, which will begin later this year, will be to broadcast alerts for objects that are identified as outliers. These alerts can be used by us and others to carry out telescopic confirmation and follow-up of these objects of interest.

2.3. Requirements and Current Performance

Two timing requirements are imposed by the cadence of the sky surveys that are providing data to us: (1) our data ingest and real-time processing cannot fall behind, which implies processing on average around 100 (ZTF) and 1000 (LSST) asteroids in the 30 s exposure times used by both surveys; and (2) our population outlier detection scheme cannot fall behind, which means that all computations that are carried out during the following day must be completed between sunrise and sunset at the observatory.

For the first constraint, there are two key steps that must be carried out within a 30 s window: ingesting the data and identifying real-time outliers. Ingesting the data is fast (<1 s), and our present real-time outlier detection implementations—where we simply query observed properties such as elongation of the measured source—are also fast (<1 s). As we develop more sophisticated multidimensional outlier detection schemes we will ensure that the computational performance meets the requirements.

For the second constraint, we must compare the processing time for the entire database—or at least the subset that has changed since the previous day—to the available time, which could be as little as 10 hr during winter observing. For 10,000 objects, calculating all lightcurve-related properties (color, period, amplitude) takes about 83 CPU-hours on a single AMD EPYC 7542 CPU core. The average number of asteroids observed in a full ZTF night may be about 10 times this amount, and LSST another factor of 10 greater. This computational need would take more than one day if just a single CPU were available. Thus, our SNAPS pipelines run on

NAU's High Performance Computer (Monsoon) with multiple CPU/GPU nodes in order to complete the processing within a day. Almost all the 83 CPU hour effort is dedicated to LSP processing. To address this computational load, we have created a GPU LS implementation (Gowanlock et al. 2021) that is described in Section 6.2.1. With this GPU approach, we can readily complete the daytime processing within one day, using our project-specific node on Monsoon that has 2 AMD EPYC 7542 CPUs (32 cores each) and 4 Nvidia A100 GPUs.

3. SNAPShot1: Our First Data Release

Since late 2018 we have been ingesting ZTF moving object alerts in real time, every night. We ingest data from a Kafka data stream provided by ANTARES; this is a substream rebroadcast of solar system objects that appear in the main ZTF stream. (As described above, we have also retroactively ingested alerts for objects with low real/bogus scores.) As of this writing, our database contains some 17 million observations of some 469,000 unique asteroids. This Oracle database of measurements is 3.9 GiB as of this writing.[14]

For the purposes of this paper we define SNAPShot1, which is our first data release. SNAPShot1 includes data for numbered, known solar system objects reported in ZTF alerts during 2018 July 19 to 2020 May 19. Future SNAPShots will include unnumbered objects and a broader date range. SNAPShot1 includes 5,458,459 observations of 31,693 asteroids. Figure 2 shows the histogram of the number of observations for targets in SNAPShot1.

For each ZTF asteroid with more than 50 observations we derive its color, lightcurve period, and lightcurve amplitude, as described above, as well as its absolute magnitude and phase coefficient. Figures 3 through 6 show histograms of the derived properties for the objects in SNAPShot1. Figures 7 through 10 show various combinations of the SNAPS properties. Finally, Figures 11 through 14 show the SNAPS properties in relation to properties from external catalogs. Several of these figures are discussed in the following sections.

Not shown here are several parameter comparisons where nothing new is revealed. Grossly, we find that the lightcurve period and lightcurve amplitude do not significantly depend on g − r color (i.e., asteroid taxonomy). There is no apparent dependence of the lightcure period or amplitude on semimajor axis. And, as expected, we find that the inner belt has more objects that have redder g − r colors (S-types), and the outer belt is dominated by objects that have less red g − r colors (C-types).

---

[12] https://www.minorplanetcenter.net/data

[13] https://www.minorplanet.info/php/lcdb.php

[14] Note that 1 GiB is 2^{30} bytes, whereas 1 GB is 10^9 bytes. Throughout the paper, when reporting data sizes, we use GiB, which uses base 2 and is how computers measure storage (and which is also the standard unit in the computer science literature). Since 1 GB is equal to ≃0.93 GB, they are nearly equivalent, and a reader can acceptably interpret GiB as GB without misunderstanding our results.
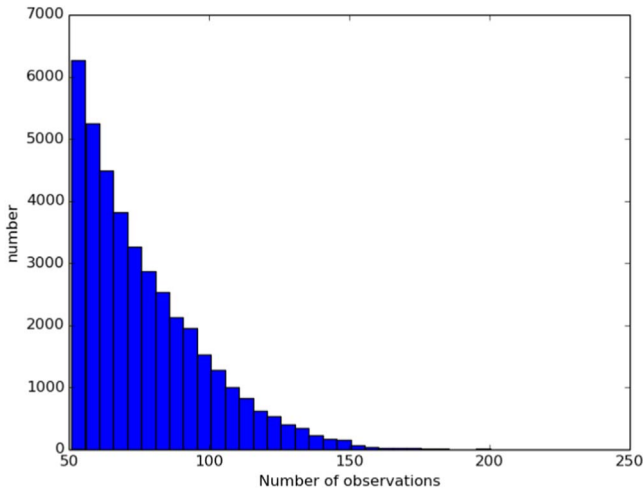
Figure 2. Histogram of number of observations of asteroids in SNAPShot1. All 40,000 asteroids presented in SNAPShot1 have more than 50 observations; nearly 6000 asteroids have more than 100 observations.
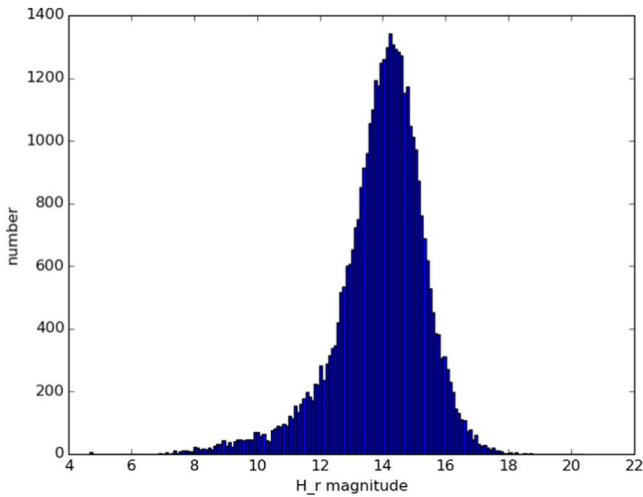

Figure 3. Histogram of the derived H_r magnitudes (the absolute magnitudes of the objects, in r band). The histogram of H_g looks very similar, with a small shift due to the g − r color of each object. H_r = 15 corresponds to a diameter of around 3 km (for an albedo of 0.25).

4. Accuracy of our Derived Properties

4.1. Accuracy of the Derived Properties—Synthetic Asteroids

To determine the accuracy of our derived physical properties we have created a population of synthetic asteroids; a detailed description of this synthetic asteroid population is given in the Appendix. Briefly, these asteroids are randomly assigned colors, lightcurve periods, lightcurve amplitudes, and several other properties from distributions that are similar to distributions in the literature (generally, from the LCDB; Warner et al. 2009). These synthetic asteroids are "observed"—that is, a synthetic observational record is produced—with ZTF-like cadences and ZTF-like photometric errors. These data sets are then passed through our processing pipeline, which derives the various properties. Finally, we compare the derived properties to the input properties to assess the accuracy of our approaches.

Some results from analyzing this synthetic population are shown in Figures A3–A5. For our derived colors, around 60% of the solutions are within 10% of the assigned colors, and 90% of the solutions are within around 25% of the assigned colors.

Figure 4. Histogram of the derived g − r colors for the objects in SNAPShot1.

For our derived lightcurve amplitudes, around 60% of the solutions are within 10% of the assigned lightcurve amplitudes, and 90% of the solutions are within around 20% of the assigned amplitude. Finally, for our derived lightcurve periods, around 70% of the solutions are within 10% of the assigned periods (here we do not include aliases as "matching"), and 90% of the derived periods are within around a factor of two. Based on this analysis we conclude that in general our pipeline is producing reliable results for our derived properties. However, as with any tool that carries out bulk processing of large data sets, any individual asteroid result may have a derived property that is not correct. The SNAPS database can readily be used for bulk analysis (i.e., distributions of periods or amplitudes), and solutions for most individual asteroids are correct. However, individual objects that are found to be unusual in our processing (example: unusually high amplitudes, or extreme colors, or long periods) should be examined carefully.

4.2. Accuracy of the Derived Properties—Comparison to Other Published Results

We compare our SNAPS results to other published results. Figure 15 shows our derived SNAPS H magnitudes (in g and r bands) compared to the H values (taken to be V band) in the current version of MPCORB.DAT from the Minor Planet Center. The correspondence is excellent, with $\langle H_g/H_V \rangle$ = 1.03 and $\langle H_r/H_V \rangle$ = 0.99 where $\langle \rangle$ indicates the mean, for all objects, of the ratio of the absolute magnitudes. The standard deviation for both ratios is around 0.21 mag, presumably with components of around 0.1–0.15 mag, more or less, from each of the MPC and SNAPS catalogs—typical uncertainties for survey measurements. The offset between the two comparisons (H_g and H_r compared to H_V) simply reflects the mean color of all asteroids.

We compare our derived g − r colors to those from SDSS MOC4 for objects that appear in both catalogs. We find $\langle (g - r)_{SNAPS}/(g - r)_{SDSS} \rangle$ to be 0.98, with a standard deviation of 0.12 mag, again showing excellent agreement between our results and previously published results.

We compare our derived periods to values in the current version[15] of the LCDB (Warner et al. 2009). (Amplitudes

---

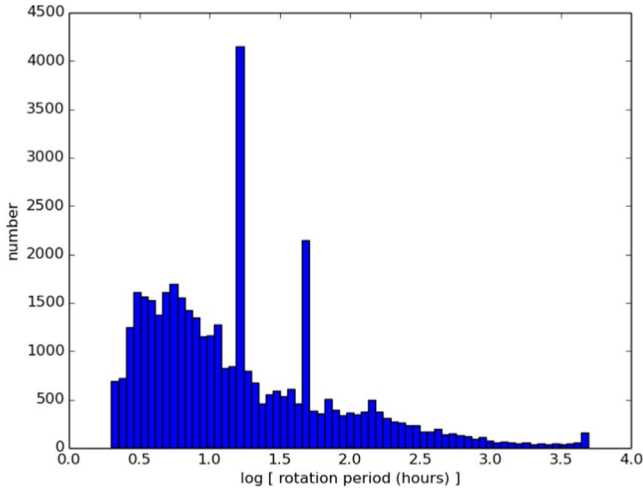[15] https://www.minorplanet.info/php/lcdb.php

Figure 5. Histogram of the derived rotation periods. Our search range is 2–5000 hr. This is a relatively smooth distribution except for the peaks at 16 and 48 hr, which are aliases that arise from the natural diurnal observational cadence of ZTF. However, it is clear from interpolating across these alias-affected bins that some 10% of those solutions are likely correct, but it is impossible to identify which objects these are. The small peak in the very largest bin may represent objects either that have periods longer than 5000 hr or objects with very low-amplitude lightcurves, where the "best" period simply reflects the search window.

Figure 6. Histogram of the derived lightcurve amplitudes. This is a very smooth distribution, with the largest values around 1.4 mag. The peak of this histogram is around 0.15 mag, with relatively few objects having derived amplitudes smaller than this. This lack of low amplitudes is almost certainly an observational bias, as typical ZTF photometric uncertainties are in the range 1%–10% such that detecting amplitudes smaller than 0.15 mag would be difficult.

cannot readily be compared since they may vary substantially from epoch to epoch due to changes in the pole orientation with respect to the observer.) There are 8748 objects that appear in both SNAPShot1 and the LCDB. Our period comparisons are shown in Figure 16. There are several clear signatures in this plot: (i) the 1:1 line where the two solutions agree is readily apparent; (ii) there are many SNAPS solutions at periods of 16 and 48 hr (see Figure 5). These are alias solutions that are products of the ZTF observing cadence and, in general, not the actual rotation periods for these asteroids, although some of these solutions are presumably correct (near where these alias distributions cross the 1:1 line); (iii) there is a line parallel to

Figure 7. SNAPS g − r color as a function of SNAPS H_r. Redder objects are detected to slightly smaller sizes (larger H values) than less red objects. This is presumably due to an observational bias, probably because red objects have higher albedos (see Figure 14) where a given limiting optical band magnitude corresponds to larger H (smaller asteroid).

Figure 8. SNAPS g − r color as a function of the SNAPS slope parameter G in r band. Two clear groups are shown, as expected: C-type asteroids have bluer colors and smaller slope parameters, and S-type asteroids have redder colors and larger slope parameters. The solutions at G = 0.9 and G = − 0.3 indicate a small number of poor solutions rather than significant populations of asteroids with extreme slope parameter values. Note that the commonly used default G value of 0.15 does not really correspond to an "average" asteroid but rather a C-type asteroid.

the 1:1 line, but shifted upward by a factor of two (in period space)—this is a presumably another alias where SNAPS has derived a period that is twice that of the LCDB period; and (iv) several curved lines can be seen that together form a quasi-triangular shape with one vertex in the lower left. These are "pseudo-aliases" (VanderPlas 2018) that result from interactions between the real period of an object and the strongest alias. These curves are defined by $1/P_D = ((1/P_R) \pm (1/24 \text{ hr}))^{-1}$, where P_R is the real period for the object and P_D is the derived period (Ďurech et al. 2022).

Solutions in the LCDB are assigned quality codes, with "2" suggesting solutions that may be uncertain at the 30% level and "3-" or better indicating solutions that generally are reliable. Almost 97% of the targets that appear in both SNAPShot1 and
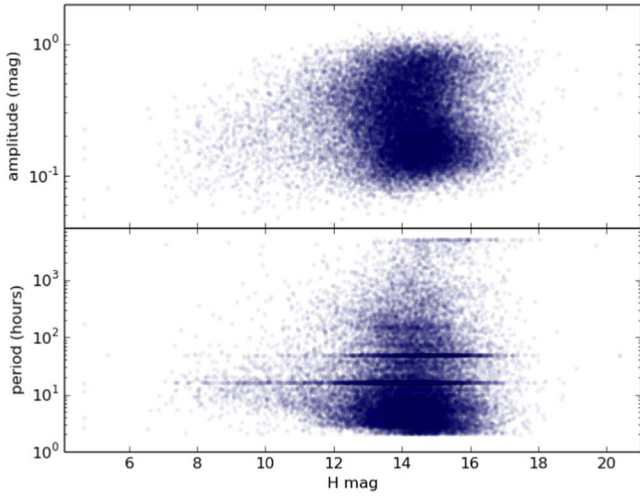
Figure 9. SNAPS lightcurve amplitude (top) and period (bottom) as a function of SNAPS H magnitude. There are few large (small H) asteroids with amplitudes around 1 mag, showing that large asteroids are generally more spherical than small asteroids. Alias period solutions are evident in the lower panel. There are few long periods among the largest objects, but the total number of large (H < 10) objects is small and this finding may not be significant.
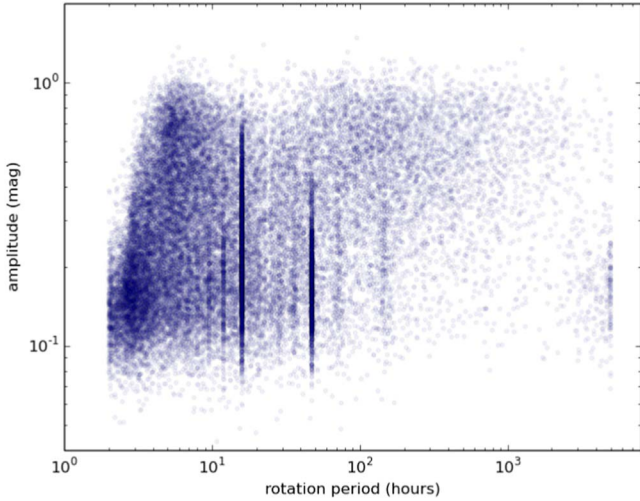

Figure 11. SNAPS H magnitude as a function of semimajor axis (from the MPC). The dynamical structure of the asteroid belt is clearly evident (Kirkwood gaps, Hildas and Trojans, etc.). As expected, ZTF is sensitive to smaller objects in the inner belt than in the outer belt and beyond: our smallest main bet asteroids have H ≈ 17.5 (1 km) whereas the smallest Trojans have H ≈ 12.5 (10 km). There are a small number of outer solar system objects in SNAPShot1.

Figure 10. SNAPS amplitude as a function of SNAPS period. Given this kind of sparse data set, very long periods with low amplitudes would be difficult to detect, and there is a relative lack of objects in this part of the figure. The lack of objects with short periods and high amplitudes is real and discussed further in Section 5.1.

Figure 12. 2MASS J − H color as a function of SNAPS g − r color. Despite the large scatter in the 2MASS colors, two clumps are evident, as expected: C-type asteroids are bluer in both colors and S-type asteroids are redder in both colors. Similar signatures exist for 2MASS J − K and H − K as a function of SNAPS g − r (plots not shown here).

the LCDB have LCDB quality codes of 3- or better. We find that around 60% of the SNAPS periods are within 10% of the LCDB period given a quality code of 2 or better; for LCDB code of 3- or better, the 10% agreement rises to 67%. Of the remaining solutions, almost 20% are in the 16 and 48 hr aliases, and most of the rest are in the pseudo-aliases.

5. Science Results to Date

The primary goal of the SNAPS project is to enable a wide range of science cases. We present here several science results that represent only the beginning of what can be accomplished with the large, uniform catalog that we are producing.

5.1. Asteroid Strengths

Figure 10 shows the derived rotation periods and lightcurve amplitudes for the minor planets from SNAPS. There are two regions on this plot where the density of objects is interestingly low. The first is the relative lack of objects showing a long rotation period with a low amplitude. This is highly likely to be an observational bias as solving for these solutions with sparse data is difficult. The second, more interesting, effect cannot simply be explained by observational biases: we observe a dearth of objects at short periods (P < 3 hr) with amplitudes larger than around 0.25 mag. This appears as an "empty triangle" in the upper left of Figure 10.
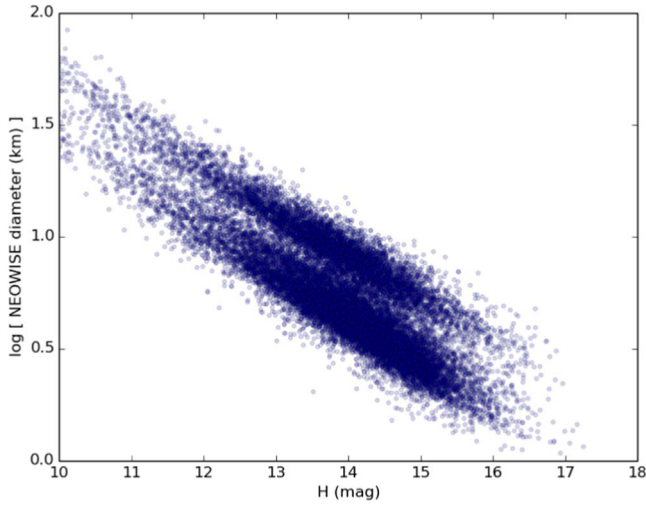
Figure 13. NEOWISE diameter as a function of the derived SNAPS H_r magnitude. The two parallel tracks are because S-types (lower track) and C-types (upper track) have different albedos (high and low, respectively).
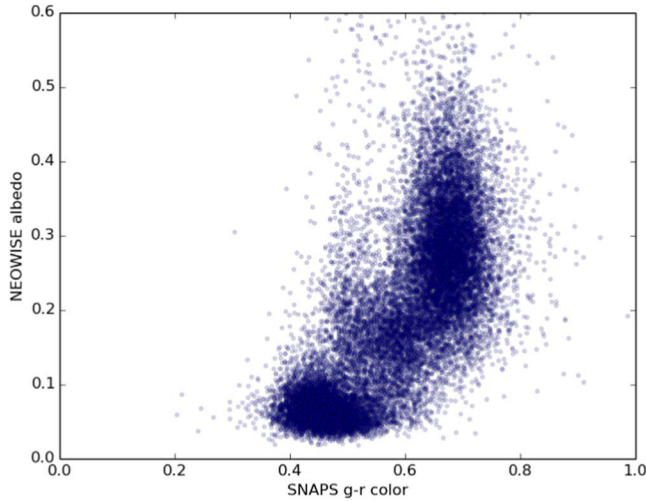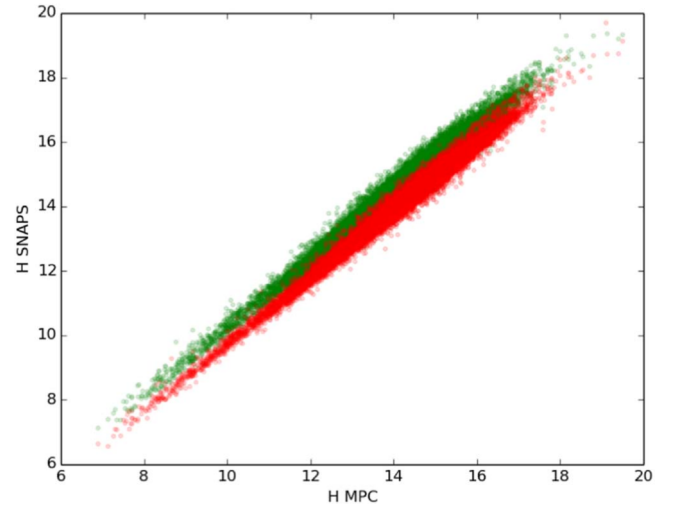

Figure 15. SNAPS H magnitudes in g and r (green and red symbols, respectively) as a function of H magnitude from the MPC, which is taken to be H_V. The offset between the green and red locus of points shows the mean g − r color of the asteroids. The overall agreement is excellent.

Figure 14. NEOWISE albedo as a function of SNAPS g − r color. As expected, there are two broad clumps of objects present: C-type asteroids generally have less red colors and low albedos, and S-type asteroids have redder colors and higher albedos.
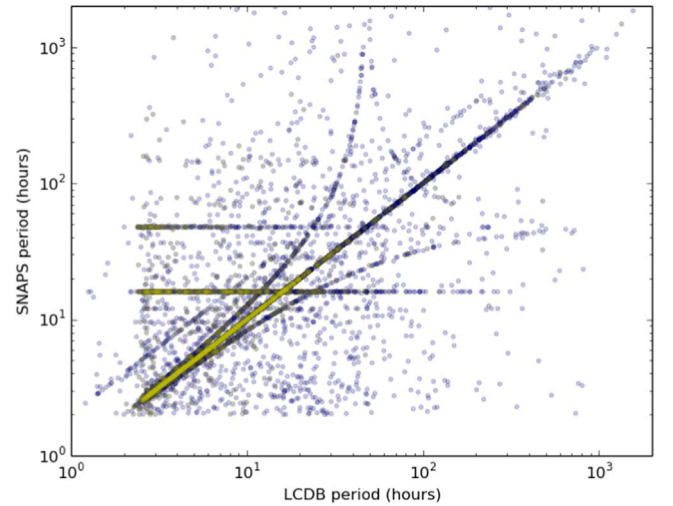

Figure 16. SNAPS period as a function of period in the LCDB (Warner et al. 2009). Blue indicates all LCDB solutions with a quality code of 2- or better, whereas cyan shows objects with LCDB quality codes of 3- or better. Around 60% of the SNAPS periods are within 10% of the LCDB periods; when restricting the comparison data set to LCDB objects with quality codes of 3- or better this match fraction increases to 67%. Most of the remaining "unmatched" SNAPS periods correspond to known aliases at 16 and 48 hr (see Figure 5); to objects where the mismatch is a factor of two (another common failure mode); and "pseudo-aliases," which are the curved shapes in this figure (VanderPlas 2018; Ďurech et al. 2022), as described in the text.

The "spin barrier" discussed in the context of asteroid rotation represents the critical spin rate at which a strengthless rubble pile of average density would undergo rotational fission (Pravec & Harris 2000). This effect will also be shape-dependent, as areas of the surface of an elongated body are much farther from the center of gravity of the object than would be the case for a more spherical object. It follows then that objects with shorter rotation periods would have a more stringent limit on elongation than an objects with longer periods.

To examine this in more detail we focus here on objects with estimated diameters of 0.3 < D < 20 km to ensure that the subpopulation is likely to consist of primarily rubble piles (both smaller and larger objects may be substantially monolithic, though for different reasons). In SNAPShot1 there are 14,784 objects with 19.5 > H > 10.5 (where we use these H magnitudes to correspond approximately to 0.3 and 20 km, respectively) and P < 10 hr. Using the methodology of

McNeill et al. (2018b), for each object we can derive a lower limit on the required strength needed to resist rotational fission.

Although we can carry out the strength modeling for every object in this subset, the derived cohesive strength is always a lower limit and for most objects the value yielded is 0 Pa. As a lower limit, 0 Pa does not provide any useful constraint, and we do not present these findings here. However, 1067 of the 14,784 objects in this subset require nonzero cohesive strength if they are indeed rubble piles held together only by self-gravity and the friction between constituent parts. This is a larger sample of strengths than has been previously determined. A
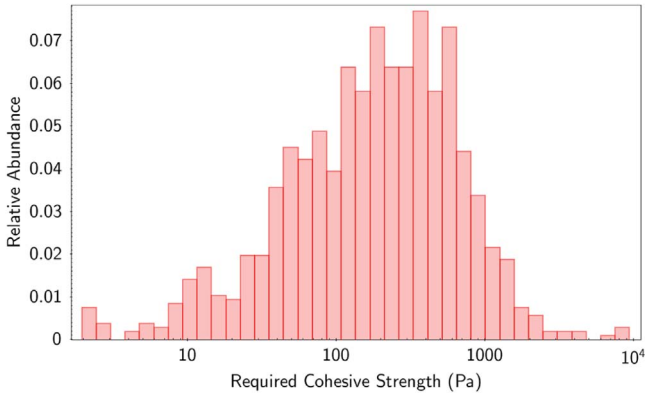
Figure 17. Histogram showing all of the nonzero strength limits determined from SNAPS objects with D < 20 km, as described in the text. The majority of these objects require only minimal cohesive strength to resist rotational fission at their current shape–spin configuration, but 6% require strengths greater than 1 kPa.
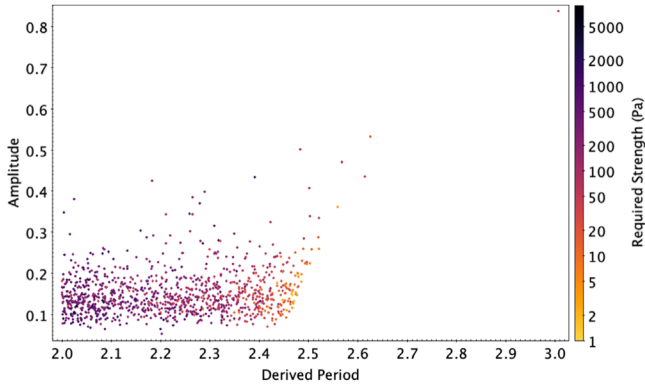

Figure 18. The derived rotation periods and amplitudes for objects in SNAPS with derived nonzero strength limits (as shown in Figure 17). The auxiliary scale shows the required internal strength (in Pascals) for the object to resist fission. The empty region in the lower right shows where objects with zero strength may exist.

Figure 19. The strength (yellow) and strengthless (maroon) regimes for a population of synthetic asteroids, each of which has a 3 hr rotation period and a density of 2500 kg m^{-3}.

histogram of these nonzero strengths is given in Figure 17. Note that for this calculation all objects are assumed to have a typical S-type bulk density of 2500 kg m^{-3}.

Figure 18 shows the periods and amplitudes for the objects shown in Figure 10 that require nonzero strengths to resist rotational fission. The auxiliary axis shows the calculated
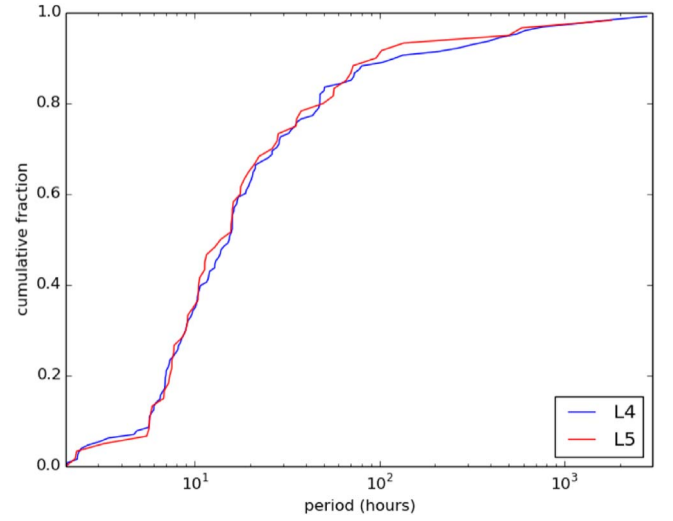

Figure 20. Cumulative fractions for the period distributions of the L4 (blue) and L5 (red) Jupiter Trojans. The period distribution of L5 Trojans is slightly and insignificantly shifted to shorter periods than for the L4 Trojans. This result rules out the possibility that L5 Trojans have significantly slower rotation rates than L4 Trojans, which increases the likelihood that the collisional argument of McNeill et al. (2021) is correct.
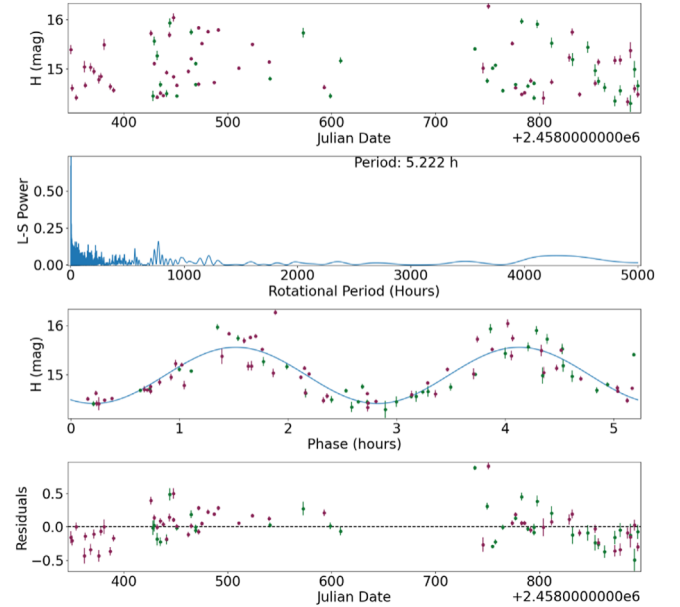

Figure 21. SNAPS results for (1620) Geographos. We produce analogous four-panel figures for each of our 31,693 targets using SNAPShot1 data. For all panels, the green points are observations in the g filter and red points are in the r filter; the g-band data are shifted by the derived g − r color. The top panel shows H magnitude for each individual observation. The second panel shows the LSP computed over a period grid from 2 to 5000 hr; the vertical axis is the normalized LS power. The best period (highest peak in the periodogram) is labeled. The third panel shows the H magnitudes shown in the top panel, folded to the best derived period (second panel), with a fitted sine curve. The phase (horizontal axis) is shown in hours, from zero to one full rotation. The bottom panel shows the residuals: folded data minus the fitted sine curve.

(minimum) strength value. This result implies a boundary dependent on both period and amplitude beyond which internal cohesive strength is required. Since this boundary depends on size and object density it will not appear as a clear cut-off in a
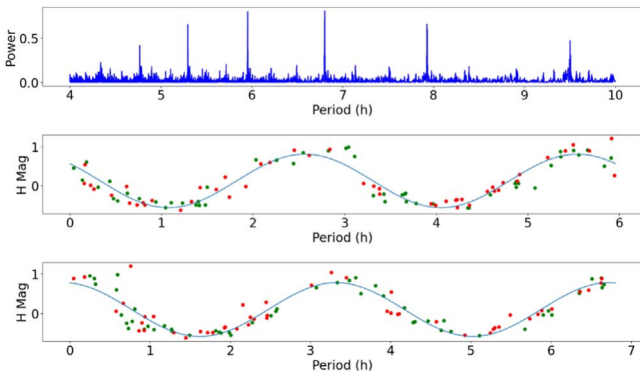
Figure 22. SNAPS results for (1865) Cerberus. The top panel shows our periodogram for this object in the range 4–10 hr. The peaks have a "picket fence" shape where the solutions in frequency space differ by a constant factor of (1/48 hr). This aliasing gives rise to the curved "pseudo-aliases" shown in Figure 16. The middle panel shows the data folded to our best solution of 5.95 hr, and the bottom panel shows the data folded to the literature value of 6.8 hr. Both folded solutions look acceptable. We present this object to demonstrate that our processing may find acceptable and good solutions that nevertheless contradict published values, or other data. Some caution is therefore urged in using results for any specific object from SNAPShot1.
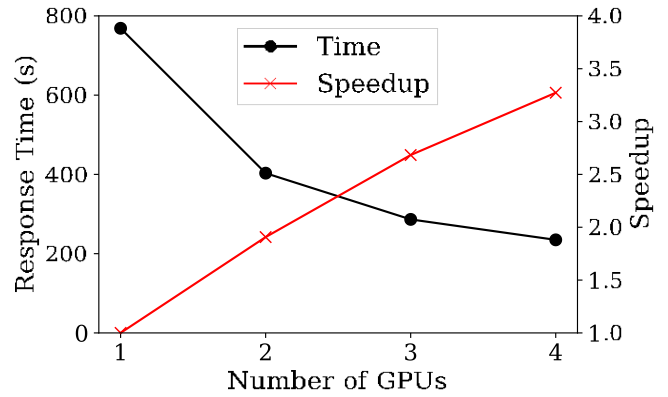

Figure 23. The response time (s) (left axis) and speedup (right axis) as a function of the number of GPUs for performing a period search using the LSPs. We compute the periods for 63,300 asteroids, using a uniform frequency grid having 10^6 frequencies. Our platform consists of two 2.6 GHz Intel Xeon Platinum 8358 processors with 64 total physical cores, 512 GiB of main memory, and four A100 GPUs, each with 40 GiB of on-card global memory.

population of mixed size and taxonomy, as is seen in the SNAPS results in Figure 10 and in the derived strengths of objects in Figure 18.

We further explore this result with a synthetic population of objects having a range of periods and amplitudes, but a fixed diameter and bulk density of 3 km and 2500 kg m^{-3}, respectively. Figure 19 shows the boundary between the strength and strengthless regimes for objects of this size and density.

The conclusion of this analysis is as follows. Some rapidly rotating, low-amplitude SNAPS asteroids require strengths as large as 5000 Pa, but most minimum strengths are hundreds of Pascals (Figure 17). These values establish a lower limit on the strength within these bodies. The strengths required are sufficient that we must consider that there are mechanisms beyond simple friction between constituent pieces of the rubble pile preventing rotational fission. The values obtained are significantly lower than the internal strength of solid rock so we do not necessarily believe that this implies that all of these objects are monolithic in nature, as opposed to being rubble piles. Instead the strengths are similar to that of lunar regolith (Mitchell et al. 1972) so we consider the possibility that these objects may have some lunar-like cohesion caused by regolith on their surfaces. The lack of SNAPS objects with large lightcurve amplitudes and short periods is a real effect. Objects with these parameters—that is, in the "empty triangle" of Figure 10—would require significant strength to maintain those elongations and rotation periods. The lack of objects in this area of parameter space indicates that indeed most asteroids have little or no strength.

Objects may be placed into this "empty triangle" corner of parameter space through (for example) collisions or the Yarkovsky–O'Keefe–Radzievskii–Paddack (YORP) effect, but unless they have sufficient internal strength they would soon shed material from their elongated ends, therefore increasing their rotation periods (through angular momentum loss) and decreasing their elongations. This would have the effect of moving those objects downward and to the right in Figure 10—out of the "empty triangle"—until they cross the stability boundary implied in that figure and in Figure 19. Objects that appear to require significantly large strength—in

other words, are found in or near the "empty triangle"—may be good targets to search for activity that is derived from mechanical mass loss.

5.2. Jupiter Trojan Asteroids

McNeill et al. (2021) used sky-survey data from ATLAS to derive the mean shape distributions of the L4 and L5 Jupiter Trojan asteroids. They conclude that L4 asteroids are, on average, more elongated than L5 asteroids, but only if their period distributions are similar; the observations could be explained without any shape difference if the L5 asteroids rotate significantly more slowly than the L4 asteroids.

We can compare period distributions using our SNAPS catalog. There are 188 Jupiter Trojan asteroids that appear in SNAPShot1 (128 L4, 60 L5). Figure 20 shows the cumulative fractions for the period distributions for the L4 and L5 Trojans. We find that the L4 Trojans are very slightly slower than the L5 Trojans (blue line slightly to the right of the red line). It is clear that the L5 Trojans are not significantly slower than the L4 Trojans (which would appear as a red line significantly to the right of the blue line), and in fact a Kolmogorov–Smirnov (K-S) test between these two data sets indicates that the period distributions are not significantly different. Therefore, we rule out one of the potential alternative explanations given by McNeill et al. (2021), increasing the likelihood that the two clouds do indeed have different shape distributions and that the collisional interpretation in that paper is correct.

5.3. Individual Objects of Interest

SNAPShot1, with more than 30,000 asteroids, contains a large number of asteroids that are not individually remarkable, but also a small number of objects that may be interesting. An exhaustive search for individual objects of interest is too cumbersome for any paper, or even any investigator; the data are available to enable a large number of people to carry out a wide range of science experiments. Here we present a few individual objects of note, to demonstrate several aspects of working with our database of derived properties.

5.3.1. (468861) 2013 LU28

(468861) 2013 LU28 is the object in SNAPShot1 with the largest semimajor axis: around 180 au (Figure 11). This object has an orbit that is Centaur- or comet-like, with an eccentricity of 0.95 (perihelion is 8.7 au) and an inclination of 125 degrees (in other words, a retrograde orbit). Its absolute magnitude is 8.0, which would make it one of the largest comets known, perhaps slightly or somewhat larger than the remarkable and recently discovered comet C/2014 UN271 (Bernardinelli–Bernstein). Our contribution to the understanding of this object is our derived lightcurve properties. We derive a rotation period of 19.5 hr and a lightcurve amplitude of 0.19 mag, suggesting a body that may be elongated or aspherical at the 10%–20% level, which is not remarkable. We cannot constrain the rotation pole. This object is presently around 10 au from the Sun, a distance at which many Centaurs and comets show activity (see, for example, Table 1 of Chandler et al. 2020). Therefore, further monitoring of this object—by classical observers, and through our SNAPS active object detection—may soon show interesting evidence of activity, potentially shedding light on the origin of this unusual body.

5.3.2. (1620) Geographos

(1620) Geographos is a near Earth object (NEO), one of only 95 in SNAPShot1. (NEOs generally have less regular observing seasons, so fewer of them will meet our criteria of 50 or more observations. Additionally, a relatively small percentage of NEOs are numbered, and SNAPShot1 only includes numbered objects. Future SNAPShots will have larger NEO samples.) This well-studied NEO has a known period of 5.22 hr and a reported amplitude in the range 1–2 mag in the LCDB.[16] (Amplitudes may vary from season to season depending on the relative orientation of the observer and the rotation pole.) We find a period of 5.22 hr and an amplitude of 1.14 mag (Figure 21), in excellent agreement with previously published values. This example is included here to show that our standard processing is acceptable even for high-amplitude NEO lightcurves, which generally present more complications than unremarkable main belt asteroids because of irregular observing seasons and wider phase angle ranges.

5.3.3. (1865) Cerberus

Asteroid (1865) Cerberus is also a well-studied NEO. For this object we derive a period of 5.96 hr and an amplitude of 1.4 mag, which is well within the range of amplitudes reported in the LCDB (0.56–2.3 mag). However, this NEO has a published period of 6.8 hr, as indicated by many entries in the LCDB, with the earliest reported measurement being from Harris & Young (1989). The SNAPShot1 data, folded at both periods, are shown in Figure 22. Both folded lightcurves appear acceptable. We present this object to demonstrate that our processing may find acceptable and good solutions that nevertheless contradict published values, or other data, and the SNAPS result may or may not be correct. Some caution is therefore urged in using results for any specific object from SNAPShot1. As demonstrated in Section 4, most of our solutions are correct, but even among solutions that are credible some may still be incorrect.

6. Discussion

6.1. Other Recent SNAPS-related Results

Erasmus et al. (2021) present the discovery of very slowly rotating asteroids, with 39 objects each having periods greater than 1000 hr. This analysis was enabled by our SNAPS processing—in this case, our derived lightcurve periods. That paper presents the existence of such very slowly rotating objects, and estimated that very slow rotators must be at least 0.4% of all asteroids. In the current database we find that 633 asteroids (around 2%) have rotation periods greater than 1000 hr. We find 83 objects that have period solutions of exactly 5000 hr, our maximum allowed period, which could either indicate a true period that is longer than 5000 hr, or instead a low-amplitude lightcurve; excluding these objects yields 550 very slowly rotation asteroids, or around 1.7% of the total sample. These very slow rotators would be extremely difficult to detect in classical lightcurve programs (one person, one telescope, one target at a time).

As an outgrowth of our SNAPS work, Navarro-Meza et al. (2021) showed that the derived size and shape distributions for asteroids near the detection limit of any large-scale sky survey may be biased due to nonzero lightcurve amplitudes. The catalogs produced from ZTF and LSST will need to be debiased in order to derive accurate distributions of these properties.

6.2. Operating at LSST Scale

LSST will produce roughly an order of magnitude more data than ZTF. The two major computational tasks that need to be carried out are deriving the properties of asteroids, and outlier detection to find outlying asteroids. As has been described in other ZTF processing pipelines for the detection of transient events, deriving periods is often the most computationally expensive task (Coughlin et al. 2021), and we address our solutions to this problem in Section 6.2.1. In Section 6.2.2 we briefly discuss what SNAPS will store in its database, and the computational resources available to SNAPS that will be used during the first few years of the survey.

6.2.1. Fast Period Searches using LSPs

SNAPS will derive the lightcurves of small bodies in the solar system; however, period searches are a computationally expensive operation at LSST scale. We examine the expected performance of SNAPS for deriving asteroid rotation periods using the LSST synthetic moving object database.[17] The database contains ≈43 million objects. We derived periods for a subset of objects by selecting those that have at least 50 observations in two filters (i.e., those objects with sufficient observations to produce believable lightcurves). This yielded a total of 63,300 objects, with an average of 517 observations per object. As expected, the maximum observing window of an object is 10 yr (the length of the survey).

We examine period search performance at LSST scale using the GPU-accelerated LS algorithm developed by our team (Gowanlock et al. 2021). We derive the periods for the ≈63,000 asteroids described above, and search frequencies in the range 0.0150–150.796 day^{-1}, which corresponds to asteroid rotation periods of 2–20,000 hr, as LSST will enable

---

the search for very slowly rotating asteroids. We use the method described by Richards et al. (2011) for selecting the frequency grid spacing, Δf, which yields roughly one million frequencies over the abovementioned frequency range. Thus, we carry out a search using $N_f = 10^6$ searched frequencies for each object on a uniformly spaced frequency grid.

To assess the performance of period searches using the LSPs, we include the time to transfer all data to the GPU, and send the periodograms back to the host for each object. The total amount of memory required to store the periodograms in 64-bit double precision format is 471.6 GiB. To ensure that the on-card global memory is not exceeded on the GPU(s), the LS algorithm automatically batches the computation (see our prior work on our GPU-accelerated Super Smoother algorithm for additional detail on multi-GPU batching; Gowanlock et al. 2022). Figure 23 plots the response time as a function of the number of Nvidia A100 GPUs, where the speedup is shown on the right vertical axis. We find that we can derive the periods of all 63,300 objects in 768.6 s on a single GPU, and only 234.9 s on four GPUs. The speedup when using four GPUs is 3.27×, where we find that we do not obtain a perfect 4× speedup due to transferring periodograms back to the host. Further performance gains can be achieved if the entire periodogram does not need to be returned to the host, such as only transferring back those periods that have a significant peak in the periodogram. Consequently, this experiment shows the worst-case scenario for our algorithm.

Overall, we find that our system can derive periods very quickly, and we will be able to process a night's worth of data on the order of minutes. If we were to derive periods for all of the estimated 5.5 million asteroids that LSST will detect in 10 yr (without the observation and filter thresholds described above), this result indicates that this can be computed in less than one day on a single A100 GPU.

6.2.2. Database and Computational Resources

Our database of LSST objects will be limited to moving objects, which is overall a small (≲10%) fraction of all LSST targets. Furthermore, since ANTARES[18] and MARS[19] will store the postage stamps for the objects, our website and API will point to those resources, such that we do not need to store and host those images. Consequently, based on the size of our ZTF database, we estimate that our database of solar system objects in the LSST catalog will be O(1000) GiB over 10 yr. Our database is hosted on an enterprise database server at NAU.

Regarding computational resources, we have dedicated access for nighttime processing to the four A100 GPUs that were detailed in Section 6.2.1 and that are installed in a node in NAU's compute cluster, Monsoon. Furthermore, since the node has 512 GiB of main memory, we will be able to store the entire subset of the database needed for a given night in the main memory; therefore, we do not anticipate any performance-related issues regarding data movement from the database to the compute node.

As a contingency plan, if the memory resources are exhausted and performing all computation in-memory is no longer possible, we have an alternative plan to reduce the potential overhead of accessing the database. Monsoon uses a fast parallel filesystem for scratch storage. Before processing

the data each night, we will stage the data from the main SNAPS database on the scratch filesystem, and then directly access that data during the night of observing. This strategy will eliminate any slow on-demand database accesses.

7. Future Work

Our future SNAPS-related work over the next 1−2 yr will focus on the following topics:

1. We will continue to ingest ZTF alerts for known moving objects. (Indeed, this will continue for the duration of that survey.) This process runs with little maintenance or supervision.
2. We will continue to ingest relevant adjacent databases (e.g., the recent solar system observations catalog that is part of Gaia DR3 (Tanga et al. 2022)), both static and dynamic.
3. We will continue to develop our real-time analysis tools to detect (among other things) activity, eventually using a multidimensional machine-learning approach where `elong` will be one, but not the only, important parameter.
4. We will continue to develop our daytime unsupervised machine-learning analysis tools to detect population outliers within our sample. Part of this development will be to ensure that we can operate automatically at a speed appropriate for the expected LSST data volume.
5. We will continue to develop our community access tools. At present there is an alpha version (internal testing only) of a web portal that will move soon to a beta release (testing by a small number of outside users). Furthermore, we will develop and release alpha and beta versions of an API to interact directly with our database. When mature, both the web and API interfaces will be released to the public.
6. There is now a LSST Solar System Products Database (SSPDB[20]), which is a full-scale simulation of LSST moving object observations. At present, the SSPDB does include asteroid colors but not asteroid lightcurves, and we are working with the SSPDB authors to implement realistic asteroid lightcurves. When the next version of the SSPDB, with lightcurves, is released, we will carry out the synthetic population fidelity testing described in Section 4.1 and in the Appendix at LSST scale.

The goal of these future activities is to demonstrate our readiness to operate our high-fidelity pipeline at LSST scale in advance of the first LSST alerts in 2024.

8. Summary

In this paper we present SNAPS, a broker that ingests data from all-sky surveys and automatically derives properties for a large number of asteroids. We describe the architecture and demonstrate the fidelity of our derived properties. We present the current snapshot of our database—SNAPShot1—and the ensemble properties of 31,693 asteroids. In the near future we will continue to ingest ZTF alerts and develop additional community access tools. We will be ready to ingest LSST alerts and work at LSST scale when that survey begins to observe and report moving objects.

Appendix
Synthetic Data Processing

Here we present the discrete steps used to create the synthetic asteroid population that is used to test the fidelity of our SNAPS processing.

A.1. Assigned Asteroid Properties

The first step is to create the synthetic population of asteroids, each of which is assigned its fundamental properties, as follows:
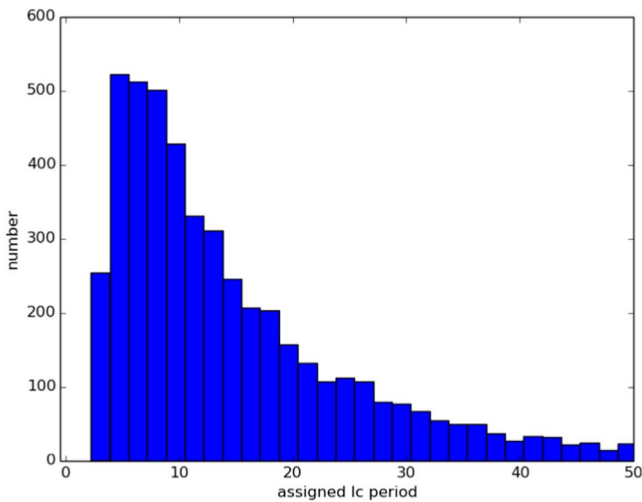

Figure A1. Histogram of assigned lightcurve periods (in hours) for our synthetic asteroids.
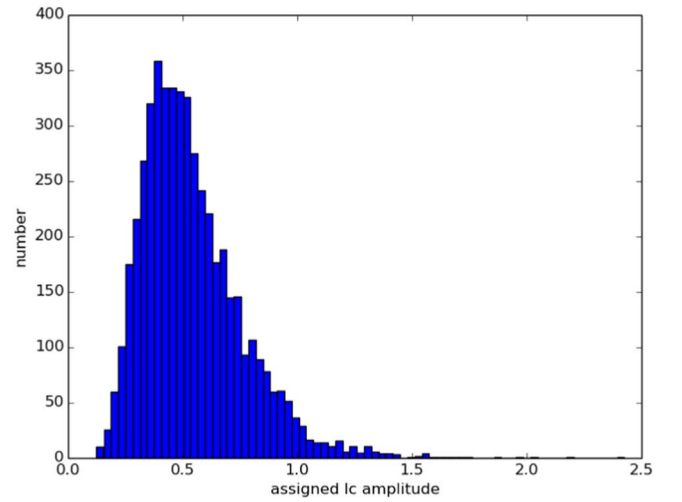

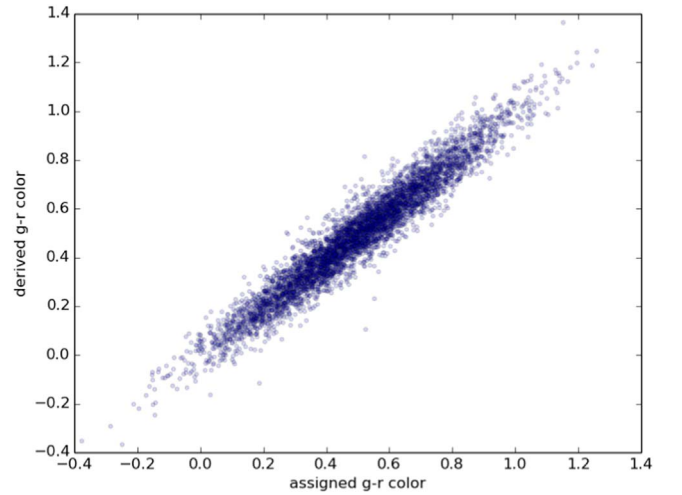Figure A2. Histogram of assigned lightcurve amplitudes (in magnitudes) for our synthetic asteroids.

Figure A3. Derived g − r color (magnitudes) vs. assigned g − r color (magnitudes) for our synthetic asteroids, showing that our solutions in general are good.
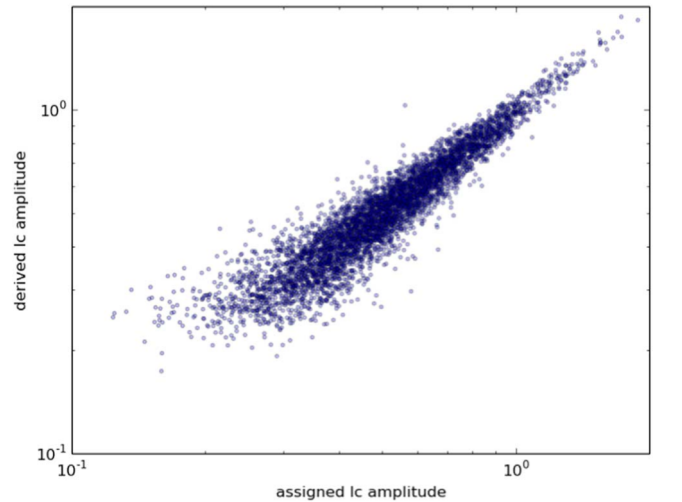

Figure A4. Derived lightcurve amplitude (magnitudes) vs. assigned lightcurve amplitude (magnitudes) for our synthetic asteroids, showing that our solutions in general are good.
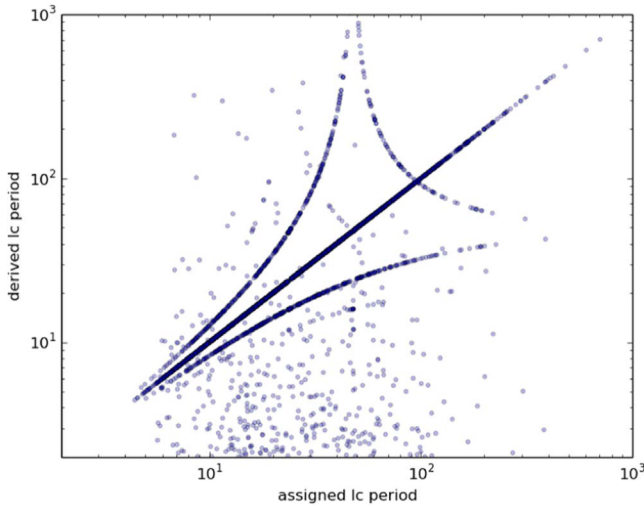
Figure A5. Derived lightcurve period (hours) vs. assigned lightcurve period (hours) for our synthetic asteroids. Most (around 70%) points are on the 1:1 line. The curved lines are "pseudo-aliases" identified by VanderPlas (2018), where the ratios between the assigned and derived periods obey certain relationships as identified by Ďurech et al. (2022).

(1) We create a distribution of g − r colors for asteroids by summing two Gaussians, one centered on g − r = 0.4 and the other on g − r = 0.6, to represent the C- and S-types, respectively. For simplicity, we assume equal number of C- and S-type asteroids.

(2) We assign a lightcurve period from a log-normal distribution with {mean, median, mode} of {19, 12, 6} hr. This distribution is shown in Figure A1 and is similar to the properties of the asteroid lightcurves in the LCDB.

(3) We assign a lightcurve amplitude from a log-normal distribution with {mean, median, mode} of {0.5, 0.5, 0.4} magnitudes. This distribution is shown in Figure A2 and is similar to the properties of the asteroid lightcurves in the LCDB.

(4) Finally, we assign a solar system absolute magnitude (H) randomly from a power-law distribution. However, for these synthetic ZTF observations, the absolute magnitude is arbitrary, as we do not determine the detectability of our synthetic asteroids. These H magnitudes may be useful in further fidelity testing for SNAPS or other projects.

A.2. Create an Observational Record

We now "observe" each synthetic asteroid with a relevant observing cadence.

(5) For each object we randomly choose a starting lightcurve phase that will be associated with the first observational timepoint.

(6) We ingest observational cadences (timestamps) from actual ZTF-observed asteroids. This preserves the number of observations and aliasing generated through the actual ZTF observations.

(7) We "observe" each asteroid (that is, calculate the asteroid's magnitude) according to the appropriate cadence, from the previous step. For each observation we determine the magnitude as a simple sine curve, using the assigned lightcurve period and amplitude, and the initial phase assigned in step 5.

(8) We assign two kinds of errors for each measurement; both errors are randomly drawn from an error distribution that is a log-normal distribution with a mode around 0.1 mag. This first error is applied as a random offset from the perfect sine curve. The second error we assign is a reported photometric error (uncertainty or error bar) for each data point, which is of similar magnitude to the offset but is not identical in value. Assigning two similar but unique errors thus allows us to approximately capture the real ZTF uncertainty distribution.

A.3. Ingest Synthetic Asteroids

Each object is treated identically to real asteroids that are reported by ZTF: we derive the period, amplitude, g − r color, etc.

A.4. Comparing the Derived Properties to Assigned Properties

For each synthetic object, we compare our derived properties to the assigned properties. The results are shown in Figures A3−A5. For our derived colors, around 60% of the solutions are within 10% of the assigned colors, and 90% of the solutions are within around 25% of the assigned colors. For our derived lightcurve amplitudes, around 60% of the solutions are within 10% of the assigned lightcurve amplitudes, and 90% of the solutions are within around 20% of the assigned amplitude. Finally, for our derived lightcurve periods, around 70% of the solutions are within 10% of the assigned periods (here we do not include aliases such as a 2:1 ratio as "matching"), and 90% of the derived periods are within around a factor of two.

A.5. Conclusion

Based on this analysis we conclude that our pipeline is producing reliable results for our derived properties. However, as with any tool that carries out bulk processing of large data sets, any individual asteroid may have a derived property that is not correct. The SNAPS database can readily be used for bulk analysis (i.e., distributions of periods or amplitudes), and solutions for most individual asteroids are correct. However, individual objects that are found to be unusual in our processing (example: unusually high amplitudes, extreme colors, or long periods) should be examined carefully.

ORCID iDs

David E. Trilling ⓘ https://orcid.org/0000-0003-4580-3790
Michael Gowanlock ⓘ https://orcid.org/0000-0002-0826-6204
Daniel Kramer ⓘ https://orcid.org/0000-0002-6676-1713
Nat Butler ⓘ https://orcid.org/0000-0002-9110-6673
John Kececioglu ⓘ https://orcid.org/0000-0003-1204-6535

References

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, PASP, 131, 018002
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bowell, E., Hapke, B., Domingue, D., et al. 1989, in Asteroids II, ed. R. P. Binzel, T. Gehrels, & M. S. Matthews (Tucson, AZ: Univ. Arizona Press), 524

Chandler, C. O., Kueny, J. K., Trujillo, C. A., Trilling, D. E., & Oldroyd, W. J. 2020, ApJL, 892, L38

Coughlin, M. W., Burdge, K., Duev, D. A., et al. 2021, MNRAS, 505, 2954

Duev, D. A., Mahabal, A., Masci, F. J., et al. 2019, MNRAS, 489, 3582

Ďurech, J., Vávra, M., Vančo, R., & Erasmus, N. 2022, FrASS, 9, 809771

Erasmus, N., Kramer, D., McNeill, A., et al. 2021, MNRAS, 506, 3872

Ferland, G. J., Porter, R. L., van Hoof, P. A. M., et al. 2013, RMxAA, 49, 137

Freedman, D., Pisani, R., & Purves, R. 2007, Statistics (4th edn.; New York: WW Norton & Company), https://wwnorton.com/books/Statistics

Gallet, B., & Gowanlock, M. 2019, in 2019 IEEE Int. Parallel and Distributed Processing Symp. Workshops (IPDPSW) (New York: IEEE), 396

Ginsburg, A., Sipőcz, B. M., Brasseur, C. E., et al. 2019, AJ, 157, 98

Gowanlock, M. 2019, in Proc. ACM Int. Conf. on Supercomputing, ICS '19 (New York: ACM), 35

Gowanlock, M., Butler, N. R., Trilling, D. E., & McNeill, A. 2022, A&C, 38, 100511

Gowanlock, M., & Karsin, B. 2019, JPDC, 133, 107

Gowanlock, M., Kramer, D., Trilling, D. E., Butler, N. R., & Donnelly, B. 2021, A&C, 36, 100472

Harris, A. W., & Young, J. W. 1989, Icar, 81, 314

Hautamaki, V., Karkkainen, I., & Franti, P. 2004, in Proc. 17th Int. Conf. on Pattern Recognition, Vol. 3 (New York: IEEE), 430

Ivezić, V., & Ivezić, Z. 2021, Icar, 357, 114262

Ivezic, Z., Juric, M., Lupton, R., et al. 2020, SDSS Moving Object Catalog V1.0, Technical Report, doi:10.26033/bv8r-xe89

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111

Jewitt, D., Hsieh, H., & Agarwal, J. 2015, Asteroids IV (Tucson, AZ: Univ. Arizona Press), 221

Lister, T., Kelley, M. S. P., Holt, C. E., et al. 2022, PSJ, 3, 173

Lomb, N. R. 1976, Ap&SS, 39, 447

LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv:0912.0201

Mainzer, A., Bauer, J., Cutri, R., et al. 2019, NEOWISE Diameters and Albedos V2.0, Technical Report, doi:10.26033/18S3-2Z54

Matheson, T., Stubens, C., Wolf, N., et al. 2021, AJ, 161, 107

McNeill, A., Erasmus, N., Trilling, D. E., et al. 2021, PSJ, 2, 6

McNeill, A., Fitzsimmons, A., Jedicke, R., et al. 2018a, AJ, 156, 282

McNeill, A., Trilling, D. E., & Mommert, M. 2018b, ApJL, 857, L1

Mitchell, J. K., Bromwell, L. G., Carrier, W. D., Costes, N. C., & Scott, R. F. 1972, JGR, 77, 5641

Navarro-Meza, S., Aadland, E., & Trilling, D. 2021, RNAAS, 5, 111

Pravec, P., & Harris, A. W. 2000, Icar, 148, 12

Richards, J. W., Starr, D. L., Butler, N. R., et al. 2011, ApJ, 733, 10

Scargle, J. D. 1982, ApJ, 263, 835

Sykes, M., Cutri, R., Skrutskie, M., et al. 2020, 2MASS Asteroid and Comet Survey V1.0, Technical Report, doi:10.26033/yj36-sn76

Tanga, P., Pauwels, T., Mignard, F., et al. 2022, arXiv:2206.05561

Tao, Y., Papadias, D., Lian, X., & Xiao, X. 2007, The VLDB Journal, 16, 293

VanderPlas, J. T. 2018, ApJS, 236, 16

Vereš, P., Jedicke, R., Fitzsimmons, A., et al. 2015, Icar, 261, 34

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261

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