



# Exploring the Solar System with the NOIRLab Source Catalog I: Detecting Objects with CANFind

Katie M. Fasbender<sup>1</sup>  and David L. Nidever<sup>1,2</sup> 

<sup>1</sup> Department of Physics, Montana State University, P.O. Box 173840, Bozeman, MT 59717-3840 USA; [katiefasbender@montana.edu](mailto:katiefasbender@montana.edu)

<sup>2</sup> NSF's National Optical-Infrared Astronomy Research Laboratory, 950 North Cherry Avenue, Tucson, AZ 85719 USA

Received 2020 December 17; revised 2021 August 25; accepted 2021 August 26; published 2021 November 16

## Abstract

Despite extensive searches and the relative proximity of solar system objects (SSOs) to Earth, many remain undiscovered and there is still much to learn about their properties and interactions. This work is the first in a series dedicated to detecting and analyzing SSOs in the all-sky NOIRLab Source Catalog (NSC). We search the first data release of the NSC with CANFind, a Computationally Automated NSC tracklet Finder. NSC DR1 contains 34 billion measurements of 2.9 billion unique objects, which CANFind categorizes as belonging to “stationary” (distant stars, galaxies) or moving (SSOs) objects via an iterative clustering method. Detections of stationary bodies for proper-motion  $\mu \leq 2.5 \text{ hr}^{-1}$  ( $0.017 \text{ day}^{-1}$ ) are identified and analyzed separately. Remaining detections belonging to high- $\mu$  objects are clustered together over single nights to form “tracklets.” Each tracklet contains detections of an individual moving object, and is validated based on spatial linearity and motion through time. Proper motions are then calculated and used to connect tracklets and other unassociated measurements over multiple nights by predicting their locations at common times, forming “tracks.” This method extracted 527,055 tracklets from NSC DR1 in an area covering 29,971 square degrees of the sky. The data show distinct groups of objects with similar observed  $\mu$  in ecliptic coordinates, namely Main Belt Asteroids, Jupiter Trojans, and Kuiper Belt Objects. Apparent magnitudes range from 10 to 25 mag in the  $ugrizY$  and VR bands. Color–color diagrams show a bimodality of tracklets between primarily carbonaceous and siliceous groups, supporting prior studies.

*Unified Astronomy Thesaurus concepts:* [Astronomy data analysis \(1858\)](#); [Solar system \(1528\)](#)

## 1. Introduction

Although the sky has been extensively searched for solar system objects (SSOs) over the last few hundred years, new bodies are still being found today. Long before the first discovery of an asteroid (Ceres) at the Palermo Observatory in 1801 (Cunningham et al. 2009) until as recently as 1991, the process of identifying SSOs depended entirely on the human eye. Through the origin of astronomical surveys specifically intended for SSO detection, and later the addition of charge-coupled devices (CCDs) for astronomy purposes by the Spacewatch Telescope at the Steward Observatory (Gehrels 1991), SSOs were spotted either by direct observation or by “blinking” between physical exposures of the sky in a lengthy process. This changed when Rabinowitz (1991) created the first automated method used for detecting Earth-approaching asteroids in near real time using the Spacewatch Telescope. With sequential exposures taken by the telescope, the detection software identified trails that moving objects (MOs) left in the first two exposures, and displacements between them and a third exposure. This new method of MO detection marked the beginning of a striking growth in the rate of asteroid discovery, and of our knowledge of the solar system.

SSOs fall into categories based on their orbital parameters and composition, called “groups” and “families.” While groups have similar orbits, families are more closely related as they often consist of fragments from past collisions. Distinct to us by their high proper-motion and perihelia  $\leq 1.3 \text{ au}$ , Near Earth Objects (NEOs) are potential impactors due to their proximity. It was not until 2008 that the first detection of an asteroid before its impact was made by the Catalina Sky Survey (CSS, Boattini et al. 2009). Various organizations and warning systems like the Asteroid Terrestrial-impact Last Alert System

(ATLAS, Tonry et al. 2018) are dedicated to the NEO search, as even a relatively small asteroid has the potential for mass destruction (Dearborn et al. 2020) and lasting effects on the Earth’s environment (Marchi et al. 2019; Koeberl & Ivanov 2019). To address the problem, in 1991 the Spaceguard Survey Network was established by NASA in response to direction from the United States Congress to increase the detection rate of asteroids that cross Earth’s orbit (Morrison 1992). Spaceguard has led to a significant increase in surveys dedicated to discovering NEOs, and in 2011 the NEO subset of the Wide-Field Infrared Survey Explorer (NEOWISE, Mainzer et al. 2011) reported the completion of the initiative to identify 90% of NEOs greater than 1 km in diameter. The search expanded in 2005 when NASA was instructed to find 90% of NEOs with diameters greater than 140 m within 15 yr, a time window that is closing. Despite CSS, NEOWISE, ATLAS, the Pan-STARRS Survey (Denneau et al. 2013), and further programs and surveys dedicated to reaching this goal, surprise impacts continue; the 2013 event in Chelyabinsk was caused by an undetected NEO with an estimated diameter of 20 m (Popova et al. 2013). This goes to show that blind spots still exist in our tracking methods and additional SSO exploration is, at the very least, beneficial.

Observational limitations also make it particularly difficult to detect many of the smaller subsets within the NEO population, including Atens (Earth-crossing orbit), Atiras (orbits within Earth’s), Vatiras (orbits within Venus’s), and Earth Trojans (co-orbiting with Earth). Due to their close proximity to the Sun, members of these groups elude observation and it is estimated that only 5%–7% of Atiras with absolute magnitudes less than 20 ( $d \approx 300 \text{ m}$ ) have been found (Ye et al. 2020). The least studied groups are the Vatiras and the Earth Trojans, as only one of each has been discovered as of 2020 (Connors et al. 2011;

Greenstreet 2020). Studies such as that by Markwardt et al. (2020),<sup>3</sup> who searched for Earth Trojans using exposures from DECam, sometimes require confining the area included in their search in order to optimize the chances of detection. This and other surveys that constrain their fields to small sections of the sky to improve detection prospects, like those performed by Terai et al. (2013) and Ivezić et al. (2001), often do not allow for orbit determination as the observation arcs observed were too short to extrapolate.

Further out beyond the orbit of Mars, the Main Belt is filled with dozens of dynamic asteroid groups and families that can be constrained by searches through large catalogs (Lemaître 2005). Main Belt Asteroids (MBAs) observed by the Sloan Digital Sky Survey are shown to have a bimodality of carbonaceous and rocky siliceous objects (Ivezić et al. 2001), with size distributions estimated assuming consistent albedos throughout each population. MBA size and color can be relatively well studied because of their abundance and our closeness, but contrasting results raise questions. The steepest estimated size distribution is a simple power law  $\sim 0.9$  from the 2015 High cadence Transient Survey (HiTS) campaign (Peña et al. 2020), similar only to the results of Parker et al. (2008), who also use Sloan Digital Sky Survey (SDSS) data. However, the 2014 HiTS campaign reported a slope of  $\sim 0.76$  in the  $g$  band for brighter objects and  $\sim 0.28$  for faint objects (Peña et al. 2020), which is closer to more common results like those found by Gladman et al. (2009). According to Peña et al. (2020), this is likely due to an interesting result described by Bhattacharya et al. (2010), who reported that objects at ecliptic latitudes  $b \geq 5^\circ$  observed by the Spitzer Space Telescope were 30% fainter than SSOs in the ecliptic plane (the 2015 HiTS campaign observed fields at a higher ecliptic latitude than the 2014 campaign). Additional SSOs must be observed to model a more accurate size distribution, and to provide more insight on its relation to ecliptic latitude.

At the far reaches of our solar system, Trans-Neptunian Objects (TNOs) are distant enough (semimajor axis  $a \gtrsim 30$  au) that they do not feel the effects of the main planets as much as objects in the inner solar system. This means that they are particularly susceptible to external forces like galactic tides, passing stars, or perhaps even a distant unknown planet. These interactions with extrasolar forces provide insight into the dynamics, formation, and history of the solar system. TNOs, which include Kuiper Belt Objects (KBOs) and Oort Cloud Objects (OCOs), are essentially remnants of the primordial solar system, and studying their properties will provide stronger constraints on planet formation theories.

The potential presence of an unseen planet or mass has been explored by several authors after the discovery of Neptune and its perturbed orbit. Pluto's discovery in 1930 confirmed the existence of TNOs and alluded to the possibility of another larger object, dubbed "Planet X" by astronomer Percival Lowell. The most recent and promising hypothesis was established when Brown et al. (2004) discovered 90377Sedna (Sedna), the most distant SSO known at the time. Sedna has an unusually eccentric and inclined orbit, and one of the proposed causes of this is the gravitational scattering by an unseen planet. This hypothesis was further developed when Trujillo & Sheppard (2014) discovered the second Sedna-like body, 2012 VP<sub>133</sub>, orbiting at the very edge of the Kuiper Belt using the

Dark Energy Camera (DECam, Flaugher et al. 2015). They found that both Sedna and 2012 VP<sub>133</sub> showed clustering in the argument of perihelion ( $\omega$ ), the angle between an inclined SSO's perihelion and the point where its orbital plane intersects with the ecliptic plane. The mechanism by which this clustering is thought to be achieved is referred to as "gravitational shepherding," which describes the effect of a large body on the orbits of nearby smaller bodies. As a small body passes by a larger one, the latter's nonzero net gravitational force on the former alters its trajectory and shifts its orbit. To this day, the gravitational shepherding due to a distant massive planet, often referred to as "Planet 9," remains by far the best explanation for the clustering seen in extra-TNOs.

Batygin & Brown (2016) added to this hypothesis by showing a clustering in not only  $\omega$  of Kuiper Belt orbits, but a physical clustering of their perihelia as well. They also report that the orbits of known KBOs with semimajor axis  $a \geq 227$  au all approach perihelion within  $94^\circ$  of longitude of each other, and all lie in an orbital plane  $\sim 22^\circ$  from the ecliptic (Brown & Batygin 2016). These numbers are supported by the recent discovery of 541132 Leleākūhonua (2015 TG387) by Sheppard et al. (2019). Two distinct groups of TNOs with similar  $\omega$  have allowed astronomers to derive theoretical orbits of Planet 9.

Alternative explanations have been proposed and rejected; Khain et al. (2018) showed that the clustering may be due in part to the primordial state of the solar system, but would not have produced an effect as substantial as the one observed. Brown (2017) effectively ruled out observational bias as the cause, estimating that the probability that the 10 KBOs known at the time with  $a \geq 230$  au were part of a larger population with uniform  $\omega$  was 1.2%. A later work (Brown & Batygin 2019) reported a 0.02% probability that distant KBOs would be clustered as strongly as observed in both  $\omega$  and in orbital pole position if the sole reason for this clustering was a detection bias.

Constraints on a mass and orbit for Planet 9 have changed over the years, from Hogg et al. (1991) placing a detectable limit on its mass of about  $70 M_\oplus$ , to Batygin et al. (2019) more recently constricting it to  $\sim 5\text{--}10 M_\oplus$  thanks to the greater number of known TNOs. Through a numerical simulation, Brown & Batygin (2016) suggested that the perihelion of Planet 9 would be between 150 and 350 au, with an inclination within  $30^\circ$  of the ecliptic.

Studies of the Kuiper Belt and searches for TNOs have yielded small bodies and dwarf planets, but no body large enough to cause the observed clustering of TNOs has been found. Although the existence of Planet 9 itself remains unconfirmed, the discovery of additional TNOs through the work described in this paper could provide further evidence. Estimations of its parameters require a working knowledge of the most distant SSOs, hence the need to continue the search for them. Similarities between new and known objects may support the evidence of Planet 9 and contribute to estimations of its properties, while differences may lead to new hypotheses for unidentified forces acting on TNOs.

Two common obstacles that surveys face include limitations on the brightness of what can be detected (typical depth  $\approx 19\text{--}24$  mag, Denneau et al. 2013) and the area of the sky covered by observations. SSO-based surveys are often constrained to the ecliptic where the bulk of the solar system's mass is found, meaning more surveys and catalogs with more sky coverage must be searched to detect objects with high orbital inclinations to explore relationships such as the latitude-magnitude dependence shown by Peña et al. (2020) and Bhattacharya et al. (2010). Several

<sup>3</sup> Data from Markwardt et al. (2020) are included in the NSC and were analyzed as a part of this study.

**Table 1**  
Instrument Specifications and Coverage in NSC DR1

Instrument	Observatory	Coverage(MMYYYY)	$N_{\text{exposures}}$	FOV (°)	CCDs	Pixel Spacing (")
DECam <sup>a</sup>	CTIO Blanco 4 m	092012–082017	195,489	3	62	0.26
Mosaic-3 <sup>b</sup>	KPNO Mayall 4 m	022016–082017	40,224	3	4	0.26
90Prime <sup>c</sup>	Steward Bok 2.3 m	022016–062017	19,741	1	4	0.45

**Notes.**

<sup>a</sup> Dark Energy Camera; Dey et al. (2019); <http://www.ctio.noao.edu/noao/node/1033>.  
<sup>b</sup> Mosaic-3 Wide Field Imager; Dey et al. (2016); <https://www.noao.edu/kpno/mosaic/>.  
<sup>c</sup> Prime Focus CCD Mosaic Camera; Zou et al. (2017).

**Table 2**  
General Information for the 11 Top Contributing Surveys in NSC DR1, Expanded from Nidever et al. (2018) Table 1

Survey Name	Instrument	PI	$N_{\text{exposures}}$	Bands	Depth (mag)	PSF FWHM (")
DES <sup>a</sup>	Blanco/DECam	Friedman	58,546	<i>grizY</i>	24.5, 24.3, 23.5, 22.9, 21.7	1.2, 1.1, 1.0, 0.9, 0.9
Legacy Surveys: <sup>b</sup>						
DECaLS <sup>c</sup>	Blanco/DECam	Schlegel & Dey	17,533	<i>grz</i>	24.0, 23.5, 22.5	1.3, 1.2, 1.1
BASS <sup>d</sup>	Bok/90Prime	Zou & Fan	19,741	<i>gr</i>	23.7, 23.1	1.6, 1.45
MzLS <sup>e</sup>	Mayall/Mosaic-3	Dey	40,224	<i>z</i>	22.6	1.01
NEO <sup>f</sup>	Blanco/DECam	Allen	11,800	<i>VR</i>	23.0	~1
DECaPS <sup>g</sup>	Blanco/DECam	Finkbeiner	7590	<i>grizY</i>	23.7, 22.8, 22.2, 21.8, 21.0	~1
Bulge Surveys:						
BDBS <sup>h</sup>	Blanco/DECam	Rich	3849	<i>ugrizY</i>	23.5, 23.8, 23.5, 23.1, 22.5, 21.8	~0.8–1.8
DSSGB <sup>i</sup>	...	Saha	3837	<i>ugriz</i>	...	...
Light Echoes <sup>j</sup>	Blanco/DECam	Rest	7622	<i>rVR</i>	...	...
SMASH <sup>k</sup>	Blanco/DECam	Nidever	6645	<i>ugriz</i>	23.9, 24.8, 24.5, 24.2, 23.5	1.2, 1.1, 1.0, 1.0, 0.9
BLISS <sup>l</sup>	Blanco/DECam	Soares-Santos	3049	<i>griz</i>	...	...
NSC DR1	...	Nidever	255,454	<i>ugrizYVR</i>	23.1, 23.3, 23.2, 22.9, 22.2, 21.0, 23.1	1.4, 1.3, 1.2, 1.0, 1.0, 1.0, 1.2

**Notes.**

<sup>a</sup> Dark Energy Survey (Dark Energy Survey Collaboration et al. 2016); <https://www.darkenergysurvey.org>.

<sup>b</sup> Dey et al. (2019); <http://legacysurvey.org>.

<sup>c</sup> DECam Legacy Survey; <http://legacysurvey.org/decamls/>.

<sup>d</sup> Beijing-Arizona Sky Survey (Zou et al. 2017).

<sup>e</sup> Mayall z-band Legacy Survey (Dey et al. 2016).

<sup>f</sup> DECam Near Earth Object survey (Allen et al. 2016).

<sup>g</sup> DECam Plane Survey (Schlafly et al. 2018); <http://decaps.skymaps.info>.

<sup>h</sup> Blanco DECam Galactic Bulge Survey (Rich 2015, Johnson et al. 2020).

<sup>i</sup> Deep Synoptic Study of the Galactic Bulge.

<sup>j</sup> Light Echoes of Galactic Explosions (Rest et al. 2015).

<sup>k</sup> Survey of the Magellanic Stellar History (Nidever et al. 2017); <http://datalab.noao.edu/smash/smash.php>.

<sup>l</sup> Blanco Imaging of the Southern Sky.

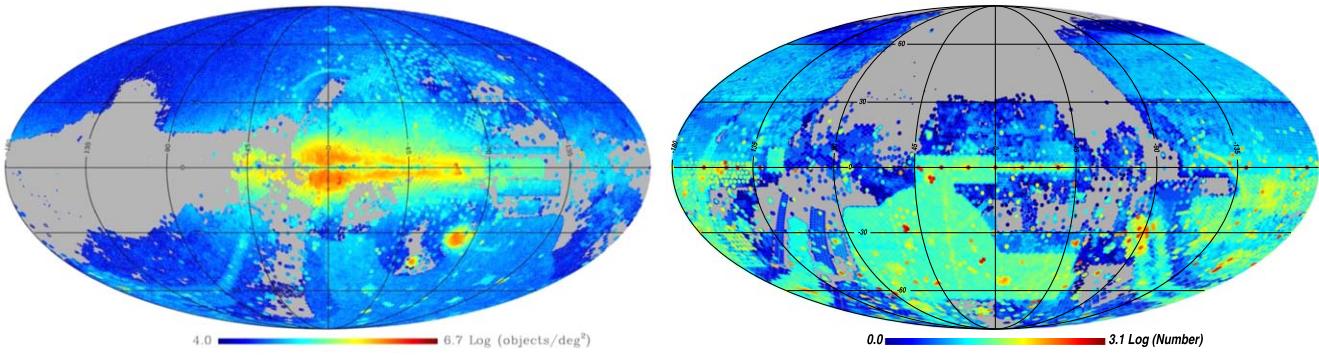
wide-field surveys have been used to achieve this, such as CSS, the Near-Earth Asteroid Tracking program (NEAT; Pravdo et al. 1999), CCS, and the Lincoln near-Earth asteroid program (LINEAR; Stokes et al. 2000), although both NEAT and LINEAR focus mainly on the northern celestial hemisphere due to the location of the telescopes used. The upcoming Legacy Survey of Space and Time (LSST; Ivezić et al. 2019) with the Rubin Observatory will extensively monitor the southern sky and is predicted to increase the known number of all small body populations by  $10\text{--}100 \times$  (Jones et al. 2016), and is expected to be a significant detector of NEOs (Veres & Chesley 2017). However, much of the northern sky will remain unmonitored at the faint magnitudes.

Independent endeavors to create detection algorithms that can be applied to any astronomical data, for example the NEARBY Platform (Stefanut et al. 2019), the Moving Object Processing System (MOPS) developed for Pan-STARRS, and other methods (e.g., Lieu et al. 2019; Perdelwitz et al. 2018,

and Holman et al. 2018) have varying degrees of success and can typically be tuned to find specific families of SSOs with similar orbits. A recent example of this is THOR (Tracklet-less Heliocentric Orbit Recovery), an algorithm developed with detections from the Zwicky Transient Facility to extract SSO detections by applying test orbits to the data (Moeyens & Juric 2019).

This work describes the background and implementation of the Computationally Automated NSC tracklet Finder (CANFind), an approach designed to identify SSOs in photometric catalogs with individual-epoch measurements like the NOIRLab Source Catalog (NSC)<sup>4</sup>. The NSC is a crowd-sourced catalog covering 3/4 of the sky, with 34 billion observations; most regions are covered by at least 3 exposures. We test CANFind on the first data release of the NSC and detect over 500,000 tracklets to magnitudes of 24 mag. Of these, 50,222 tracklets have  $g-r$ ,  $r-i$

<sup>4</sup> Formerly known as the NOAO Source Catalog.



**Figure 1.** (Left) Density plot (number of objects per square degree) of the objects in NSC DR1 (Nidever et al. 2018) in galactic coordinates. Three-quarters of the sky is covered, primarily the southern celestial hemisphere. (Right) Number of exposures overlapping any piece of sky in NSC DR1, on a logarithmic scale and in equatorial coordinates. Most areas have tens of exposures, while some have thousands.

color–color information showing a distinct bimodality of two general asteroid types (carbonaceous and silicic). In all, 85.9% of tracklets were found within  $\pm 30^\circ$  of the ecliptic plane, and 60% have more than three measurements.

The paper is organized as follows: Section 2 describes the first data release of the NOIRLab Source Catalog and its benefits for the use of SSO detection. CANFind is explained in Section 3, an overview of the detection efficiency simulation is in Section 4, the results are discussed in Section 5, and future work to be done with NSC tracklets is described in Section 6. This introduces a set of planned publications all involving SSO detection in the NSC and subsequent investigation. Upcoming projects will connect tracklets with more spatially disperse measurements, fit NSC tracklets to SSO orbits, and analyze recovered SSO populations, morphology, comet activity, and streaked detections.

## 2. NOIRLab Source Catalog

This project uses the first data release of the NOIRLab Source Catalog (NSC; Nidever et al. 2018) to detect SSOs. The NSC is a crowd-sourced catalog using public data from the NOIRLab Science Archive processed in a uniform manner. The NSC has 34 billion individual observations of 2.9 billion distinct objects from 255,000 images spanning five years of observation from 2012–2017. Data come from DECam on CTIO 4 m, KPNO 4 m+Mosaic-3, and Bok 2.3 m+90Prime from the DESI Legacy Survey (Dey et al. 2019). Table 1 lists several specifications from the instruments used by the surveys included in the catalog. The NSC contains a large amount of time-series information that is useful for searches of moving objects and photometrically variable objects. Table 2 lists information from the 11 top contributing surveys in DR1.

The NSC data are publicly available through the Astro Data Lab<sup>5</sup>, and can be accessed via SQL queries. The catalog will be updated periodically, with the second data release containing over 68 billion measurements of 3.9 billion objects (Nidever et al. 2020). Figure 1 shows the object density (left) and exposure coverage (right) of NSC DR1. An overview of NSC DR1 can be found on the Astro Data Lab site,<sup>6</sup> and a complete description is available in Nidever et al. (2018).

From the NSC DR1 exposures table, the average exposure time is 96.28 s, with a minimum of 0.01 s and a maximum of 1300 s. Figure 2 shows the number density for the full range of exposure times, broken down by instrument. As described in Section 3.2, the

NSC is only searched for SSOs in areas covered by at least three exposures; the difference between the total exposure list and the searched exposure list is also shown in Figure 2. All exposures are run through the NOIRLab Community Pipeline. Flaugher et al. (2015) report a shutter accuracy of  $\leq 1$  ms for DECam, which provides both the largest sample of exposures and the widest range of exposure times. Images with long exposure times may contain detections of faint TNOs detected as point sources as well as streaking from SSOs exhibiting higher proper motion; for an exposure time of 1300 s, an object moving  $\mu = 20''$  relative to the background would leave a  $\sim 7.2''$  trail, within the  $\sim 1080''$  range of a typical single DECam CCD (2048  $\times$  4096 pixels with pixel scale 0.26).

The NSC measurements table (nsc\_dr1.meas) provides the main source of data for this work. Each source-extracted measurement is represented as a row of data in the table, with a total of 34,658,213,888 rows. Thirty-one columns of detection parameters include position and positional uncertainty (RA, Dec), exposure filter, apparent magnitude (mag\_auto), photometry information, and the observation Modified Julian Date (MJD). An overview of all columns of the measurements table table is also available at Astro Data Lab.<sup>7</sup>

Use of the NSC may provide solutions to some of the limitations common to SSO searches described in the previous section. One issue is the potential for observational bias in the regions covered. The focus of surveys is often away from the galactic plane, due to the higher density of objects in that area polluting the recovered SSO population. NSC DR1 covers 3/4 of the sky and approximately half of the Galactic plane is included in the observed region, meaning this project searches portions of the sky that are under-examined for SSOs.

A drawback of the NSC is that exposures of the most dense areas are often oversaturated, resulting in a lack of distinction between real sources and image artifacts, and therefore a higher rate of false positives in SSO searches near the Galactic plane ( $\delta \leq 15^\circ$ ) and the Magellanic Clouds ( $5^\circ$  around LMC and  $3^\circ$  around SMC). The NOIRLab Community Pipeline used in the NSC calibration process masks saturated stars and their bleed trails, however some effects from bright saturated star still get through, including scattered light in the periphery. These generally do not line up in any preferred direction, and their effect must be accounted for in the validation of any tracklets found in the NSC.

The limiting magnitude of a survey is also very important: distant SSOs are quite faint and will be undetectable to most telescopes during the bulk of their orbits due to low albedo or large

<sup>5</sup> <https://datalab.noirlab.edu>

<sup>6</sup> <https://datalab.noirlab.edu/nscdr1/index.php>

<sup>7</sup> [https://datalab.noirlab.edu/query.php?name=nsc\\_dr1.meas](https://datalab.noirlab.edu/query.php?name=nsc_dr1.meas)

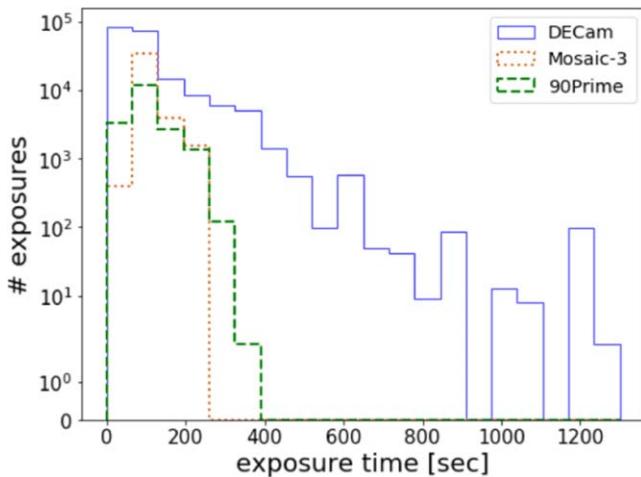


Figure 2. Histogram of exposure times for all exposures in the NSC.

distance. The depth of most NSC exposures is approximately 23 mag, an adequate value for detecting TNOs whose magnitudes are typically fainter than 20 mag. The NSC also contains multiband information that provide colors of objects and will shed further light on their properties. In a study of the colors of almost 100 TNOs, Schwamb et al. (2019) found two taxonomic groups, and the NSC should contain measurements of enough TNOs for a similar investigation (see Section 5).

Any newly discovered objects may also clarify the size distribution of solar system bodies. For instance, Sheppard & Trujillo (2010) showed a lack of discovered intermediate and small Neptune Trojans. Missing objects may be uncovered in the NSC after orbit determination, or may otherwise support these missing Trojans. By investigating the range of asteroid sizes, new parameters can be established regarding the accretion and formation history of the solar system.

Finally, the NSC covers much of the southern celestial hemisphere, which has been largely under-searched in SSO surveys. Five years of observations indicates the likelihood of orbiting bodies appearing in the NSC multiple times. An issue is the intermittent observational pattern; certain areas of the NSC have insufficient exposure coverage for SSO detection, as the contributing surveys were designed for various purposes. At least three exposures of a field are required to detect tracklets as defined in this work. Section 3.2 describes the selection of areas with adequate NSC coverage.

Because of these reasons, the deep, multiband, time-series NSC information will allow this work to markedly improve the census of solar system bodies and help our understanding of planet formation. Detections of Planet 9 may be hiding in the NSC, and, at the very least, any KBOs found in the catalog will provide further investigation into its proposed existence.

### 3. Moving Object Detection Method

Here, we describe the CANFind method used to detect SSO moving objects in the NSC catalog. The source code is written in Python and uses the `scikit-learn`<sup>8</sup> (Pedregosa et al. 2011) machine-learning package. CANFind relies on two `scikit-learn` algorithms: the Density-Based Spatial Clustering of Applications with Noise (DBSCAN; Ester et al. 1996) and RANdom SAmple Consensus (RANSAC; Fischler & Bolles 1981), a linear regressor that is robust to

outliers. Beginning with the NSC `meas` table of individual exposure measurements, the main steps are:

1. Initial spatial clustering with DBSCAN ( $\text{eps} = 0.^{\circ}5$ ) to detect and remove stationary objects and initial detection of moving objects with moderate angular velocities.
2. Second spatial clustering with DBSCAN ( $\text{eps} = 11.^{\circ}$ ) to identify detections of SSOs that form tracklets.
3. Validation of tracklets obtained in steps 1 and 2 using robust linear regression (RANSAC) and the Pearson Correlation Coefficient (PCC). Proper motions of valid tracklets are measured.
4. Track formation and validation by linking tracklets and unclustered measurements by using proper motion to predict object position at common epochs.

A flowchart outlining the process described in this section can be seen in Figure 3.

#### 3.1. Spatial Clustering with DBSCAN on Hyalite

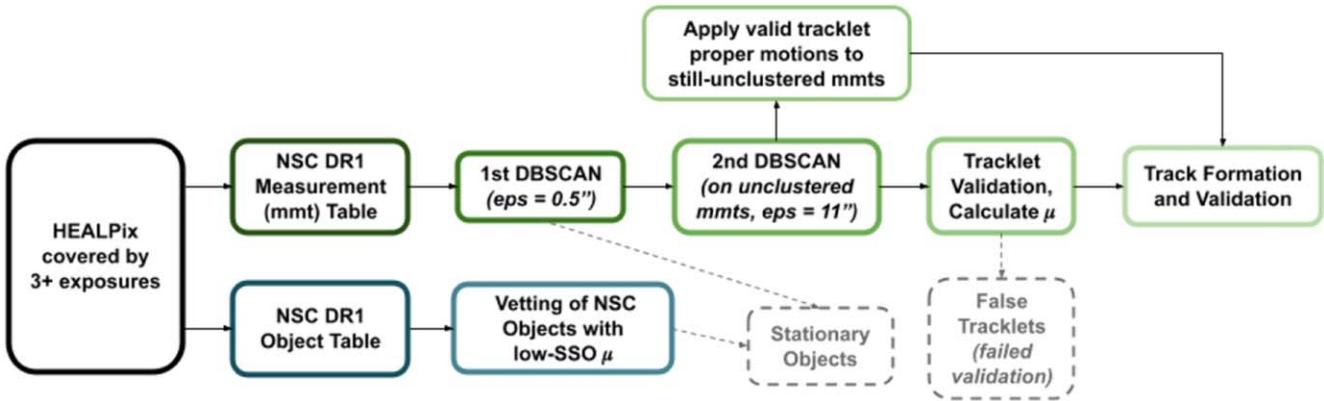
DBSCAN is used in an iterative process to assign measurements to unique objects, both stationary and moving. DBSCAN clusters data in areas of high density, and was developed in order to categorize the objects of a database into relevant groups.

This particular algorithm was selected for a number of reasons, the most important being that minimal knowledge of the data set is required to establish the two DBSCAN input parameters set for this project; the  $\text{eps}$ -neighborhood ( $\text{eps}$ ) of a point, which is the maximum distance between two points for one to be considered as in the neighborhood of the other, and  $n_{\text{points}}$ , the minimum number of points required to define a cluster. In the case of this work, DBSCAN is applied to measurements using only the information of spatial location, and  $\text{eps}$  and  $n_{\text{points}}$  are determined based on the type of object being clustered. DBSCAN also does not require a set number of priors, which is advantageous as there is no way of knowing how many objects, stationary or otherwise, will be in any area of the sky. One caveat is that DBSCAN does not take into account the positional uncertainties of measurements; they are, however, included in the calculation of tracklet proper motion.

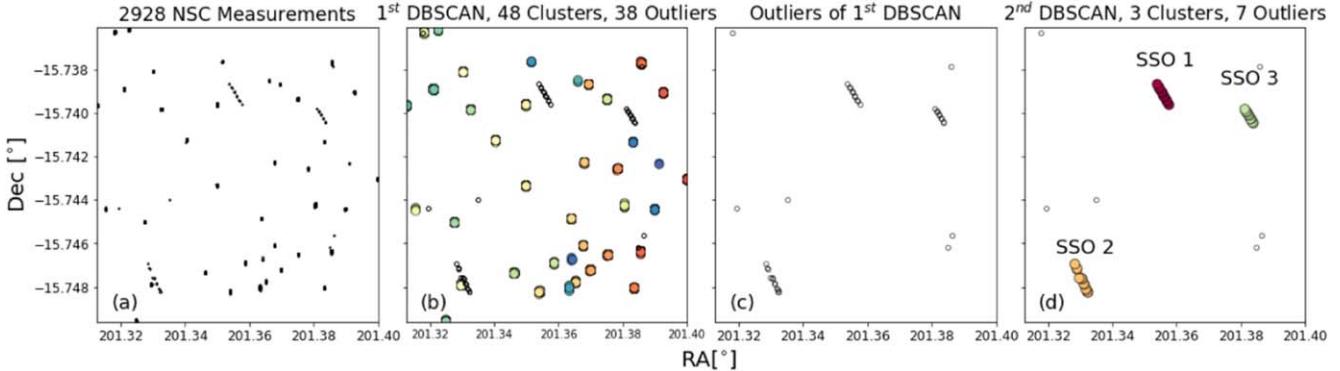
The second reason for the selection of DBSCAN is its ability to identify clusters with arbitrary shape. This allows one clustering method to be applied to all types of objects in the NSC identified in this work. It can cluster the measurements of both stationary objects (SOs), which lie in close proximity to each other in a roughly spherical pattern, as well as the more linearly distributed measurements of moving objects. As DBSCAN is only used to connect intra-night detections, more complex SSO motion can be ignored for the extent of its use. Furthermore, the capability of DBSCAN to identify outlier points is essential to the method of application, as described in the following sections. Some of `scikit-learn`'s other clustering algorithms offer this feature; for example the Variational Bayesian Gaussian Mixture Model (VBGMM) was also considered as it can identify clusters with both strong and weak linear correlations. Ultimately, the number of necessary parameters and the slow computation speeds (this algorithm is not good for very large data sets) eliminated VBGMM in favor of DBSCAN.

Lastly, DBSCAN is very efficient for large databases with significantly more than several thousand objects (Ester et al. 1996). As we analyze billions of measurements, it is essential that CANFind is as efficient as possible to maintain reasonable computation times. This is aided by Montana State University's

<sup>8</sup> <https://scikit-learn.org/>



**Figure 3.** CANFind flowchart, beginning with HEALPix covered by 3+ exposures (left). The top row follows the main CANFind process from the NSC Measurement Table to Track Formation and Validation. The bottom row indicates the NSC objects that were analyzed as potential SSOs, referred to as “Vetting NSC Objects.”



**Figure 4.** An example spatial map of NSC measurements showing clear detections of three SSOs. The SSOs move rapidly ( $\sim 40'' \text{ hr}^{-1}$  or  $0.27 \text{ day}^{-1}$ ) and show a “tracklet,” while background objects keep a fixed position (panel a). Measurements belonging to fixed objects are clustered (panel b) and removed, leaving behind outlier measurements (panel c) to be assigned to moving objects. A second iteration of clustering identifies the SSO tracklets (panel d).

Hyalite computer cluster, enabling the application of these methods on the vast extent of the NSC through a parallelized framework. Hyalite has 1300 cores and 77 computing nodes with either 64 or 256 GB of RAM. By running CANFind on the Hyalite cluster, many areas of the sky can be searched simultaneously and the computation time is dramatically reduced.

We take advantage of the NSC coming pre-divided into HEALPix<sup>9</sup> (sections of equal area, Górski et al. 2005) with NSIDE = 265 indexed in the `ring265` column of the object table, and query 4 neighboring HEALPix at a time to combine them into an area that takes CANFind an average of 5 s to analyze, depending on the number of sources and measurements present in the field. The resulting combined HEALPix, with NSIDE=128 for a resolution of approximately  $0.5'$ , is run through the CANFind process with several of its neighboring HEALPix as a job on Hyalite’s `slurm` queue. Although Hyalite allows several HEALPix to be processed simultaneously, computation speed is hampered slightly by PostgreSQL query limitations. Fewer HEALPix can be analyzed per job in fields with high object density, as this increases the query processing time and slows down the job’s progression. Additionally, jobs had to be submitted in incremental batches as the database can only handle a limited amount of queries at a time. Depending on the field density, 1–10 jobs were submitted every 10–60 minutes for 1–5 HEALPix per job. Using roughly

10 nodes at a time with 32 or 40 CPU per node, the NSC took approximately one week to analyze on Hyalite.

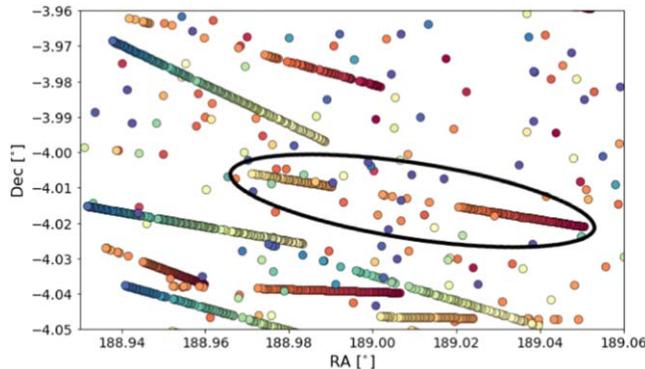
The rest of this section describes how DBSCAN is used to assign NSC measurements to unique objects, visualized Figure 4. The first iteration of DBSCAN identifies and removes stationary objects moving less than  $2.5'' \text{ hr}^{-1}$  ( $0.017 \text{ day}^{-1}$ ) for a separate vetting process, leaving remaining measurements to be clustered into distinct moving objects during the second iteration. These data will be fit to solar system orbits to create object candidates, and the properties of these candidates will be derived and analyzed.

### 3.2. Selecting Areas of the Sky for Initial Tracklet Search

The NSC has a very high object density within the Galactic plane of the Milky Way. Exposures from the Galactic plane also contain a high number of very bright and saturated stars that can cause various artifacts in the source catalog. Because these artifacts heavily impair the performance of the DBSCAN detection method, fewer objects are expected to be extracted within  $\pm 15^\circ$  of the Galactic plane. This is exacerbated by the high density of background objects in both the Galactic plane and the Magellanic Clouds, as crowding hampers the ability of DBSCAN to differentiate unique sources. A sample of tracklets extracted from these areas by CANFind is examined in finer detail in Section 5.

We used the HEALPix scheme with NSIDE = 128 to break up the sky into smaller chunks to be analyzed separately. For the initial search, only areas of sky with adequate cadence to capture

<sup>9</sup> <https://HEALPix.jpl.nasa.gov/>



**Figure 5.** An example of track formation, with color as time of observation (over two nights, MJD = 56746–56747). The two circled tracklets, along with various unclustered measurements between them, are linked together to form a track using the position-prediction method described in Section 3.5.

sequential displacements of the faster-moving SSOs were used. The selection criterion was that the HEALPix must have coverage from at least three or more exposures taken within roughly 20 minutes of each other. This reduced the 153,472 HEALPix covered by the NSC to 132,884 to be studied with CANFind.

### 3.3. Vetting NSC Objects

Because some SSOs have proper motions low enough to be represented in the NSC object table, we run an additional test on all NSC objects with at least three detections and a recorded NSC proper motion above  $0.^{\circ}0041 \text{ hr}^{-1}$  ( $2.7 \text{e-}5 \text{ day}^{-1}$ ). This value corresponds to the average proper motion of 2015 TG387, a TNO with the second-largest semimajor axis known to date (1085 au) and an orbit stable over the history of the solar system (Sheppard et al. 2019).

Every object above the minimum proper motion is analyzed to estimate the likelihood that it is an SSO. With a set residual threshold of  $0.^{\circ}001$ , RANSAC is applied to the spatial and temporal object data to remove galaxies with several NSC detections over multiple years. The `curve_fit` function from `scipy.optimize` is applied to the measurements of each object to calculate proper motion in R.A. and decl. [ $.^{\circ} \text{ hr}^{-1}$ ].

A signal-to-noise ratio cut of  $S/N \geq 4$  was made on net proper motion in order to reduce the contamination of stars and galaxies among true detections of SSOs. To further reduce this contamination, a cut on the reduced chi squared value was made, keeping all objects with  $\chi^2_{\nu} \leq 1.5$ . Despite quality cuts, it is expected that a greater number of false tracklets are present in the 2095 objects that passed this validation step. Nonetheless, they are considered to be tracks or tracklets, and therefore will be tested as SSO candidates with the rest of the tracklets detected by CANFind.

### 3.4. Stationary Object Identification

DBSCAN is first used to identify measurements belonging to slow-moving or stationary objects using a relatively small  $\text{eps} = 0.^{\circ}5$  and  $n_{\text{points}} = 2$ . The  $\text{eps}$  value was selected to reflect the original object identification of the NSC. The NSC object table was created by associating points within  $0.^{\circ}5$  of each other, but this did not account for the faster proper-motion ( $\mu$ ) of most SSOs. For this reason, only the NSC measurement table was used for this step. Panels a and b of Figure 4 demonstrate the application of DBSCAN on 2928 measurements from the NSC in HEALPix pixel 124,958 positioned a few

degrees from the ecliptic plane, forming 48 clusters (SOs) and leaving 38 outliers. The proper motions of many of these objects are already given by the NSC catalog and were referenced in the analysis for objects exhibiting solar system–like motions. The DBSCAN outliers are retained to be assigned to faster-moving objects in the next step. Because the source-extracted measurements of resolved galaxies may be more than  $0.^{\circ}5$  apart, nearby galaxies are often not identified in this first step, and must be removed later during the validation process.

### 3.5. Identifying Tracklets and Tracks

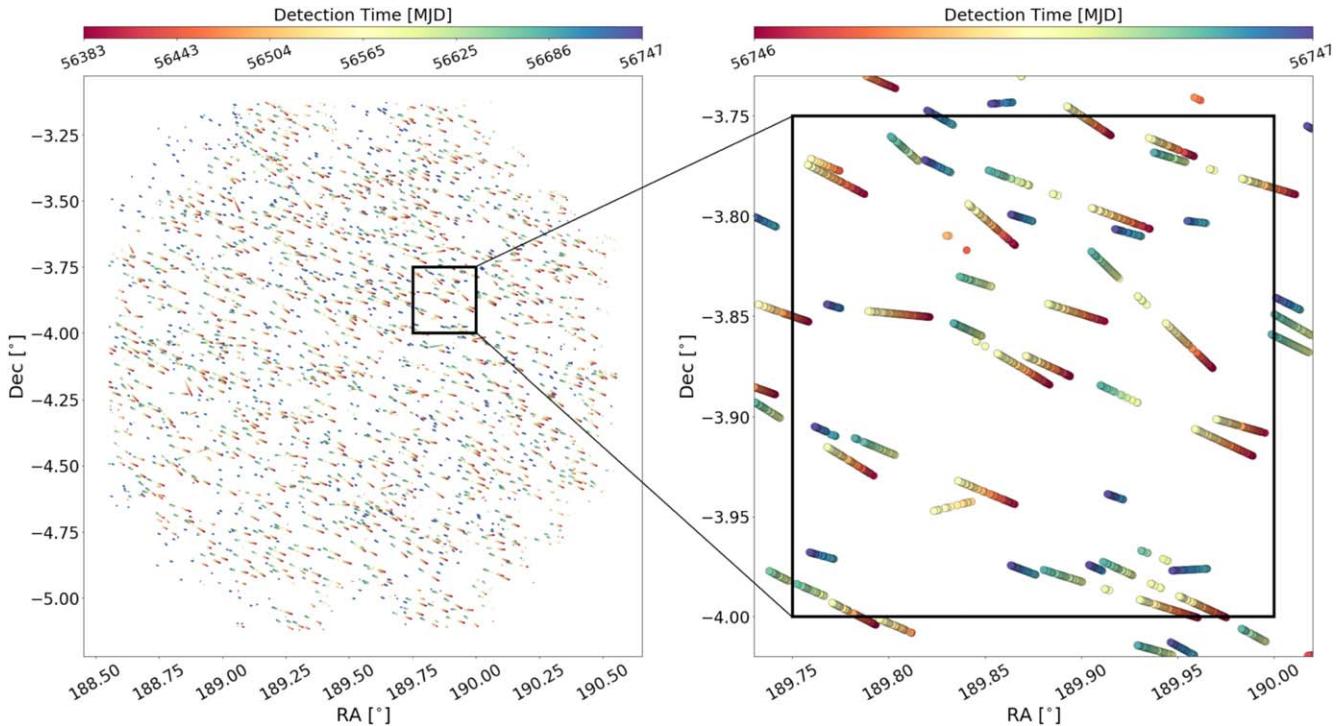
The identification of tracklets and tracks is performed in three steps:

*Linking Measurements into Tracklets:* Once measurements of SOs are identified and removed, the remaining measurements are assigned to unique moving objects by the second iteration of DBSCAN. When closely spaced exposures ( $\sim 20$  minutes apart in this example) are analyzed, measurements of fast moving objects form spatially linear structures called “tracklets.” Because these measurements are farther apart from each other than those belonging to SOs, a new  $\text{eps}$  of  $11^{\circ}$  is used, as seen in panel d of Figure 4. Each DBSCAN cluster must now contain  $n_{\text{points}} \geq 3$  detections, according to our definition of a tracklet as a collection of at least three displaced detections of a unique object. Therefore, each cluster is a tracklet that may represent the detections of an SSO, and is referred to as a “CANFind tracklet.”

*Tracklet Validation:* Each clustered tracklet undergoes a validation process to ensure that its measurements are consistently linear in space and time, and to eliminate any falsely clustered data. Some nearby, spatially resolved galaxies with a roughly linear distribution of measurements are mistakenly clustered as tracklets in this step, as DBSCAN often fails to identify them as stationary objects if that distribution is too large. However, they are easily identified through examination of their temporal information. In the creation of the NSC, multiple source detections from a single galaxy were sometimes extracted from one exposure, so resulting DBSCAN clusters do not produce a realistic trajectory for an SSO.

To test each tracklet, its measurements are analyzed using RANSAC on both the spatial coordinates and the tracklet’s normalized (*position, time*) information. Any outlier found by RANSAC is likely an incorrectly clustered measurement, and is eliminated from the tracklet. As an additional step, the Pearson Correlation Coefficient (PCC) is calculated for each tracklet. The assumption that a true tracklet has measurements that fall into a straight line motivates the PCC cutoff value of  $\geq |0.9|$  used by CANFind. If the spread of tracklet measurements is too large, then the cluster likely only contains measurements of a galaxy, not an SSO. This validation process disposes of many erroneous tracklets, making the next step of linking tracklets together a less computationally expensive process.

*Linking Measurements and Tracklets into Tracks:* Track formation in CANFind is meant to associate tracklets with both other tracklets and unclustered measurements left from the previous clustering steps. After a tracklet is verified, the slope values determined by RANSAC are saved as the proper motion in R.A. and decl. [ $^{\circ} \text{ day}^{-1}$ ]. Using these values, a position-prediction method is applied to the tracklets and remaining unclustered measurements within the HEALPix, allowing them to be connected into longer “tracks,” which consist of at least two tracklets and typically contain detections from multiple nights. As



**Figure 6.** A sample of 2,941 tracklets, color coded by measurement detection time, in one DECam field in the ecliptic plane. Field depth goes to 25.7 mag.

the assumption of linear motion starts to fail for trajectories spanning so much time, the track validation step is less restrictive than tracklet validation; that is to say, the positional uncertainty allowed for predicted positions of tracklet measurements is large ( $0^{\circ}0.005$ ) compared to measurement uncertainty to account for the deviation from linearity that SSOs display over multiple nights.

The proper motion of a tracklet is used to project its position to the observation time of other tracklets and unclustered measurements. This is done twice. First, the calculated proper motion of a single tracklet is applied to all other tracklets and unclustered measurements in the healpix, and all points are moved to the location they would hypothetically be at a common time. If the positions coincide within a few times their measurement uncertainties, then the tracklets and unassigned measurements are linked together into tracks. This is repeated for every tracklet in the healpix. A second “projection” process is applied for comparison, this time with each tracklet moving at its own calculated velocity to a common epoch. An example of track formation is shown in Figure 5.

To demonstrate the typical distribution of tracklets within the ecliptic plane, Figure 6 roughly covers the area of one DECam field, with color as detection time.

#### 4. Efficiency Simulations

To test the efficiency and completeness of CANFind and the NSC Object Vetting process to detect moving objects, a simulation was created that generated synthetic tracklets with linear trajectories and a range of proper motions that are characteristic of SSOs. These tracklets had between 3 and 10 measurements, and the average tracklet cadence was between 20 minutes (CANFind) up to several hours (NSC Object Vetting). Measurements were given equatorial coordinates, an observation date, and a magnitude between 14 and 25 with an uncertainty (`magerr_auto`) that was then used to calculate the positional error in (R.A., decl.). These generated tracklets

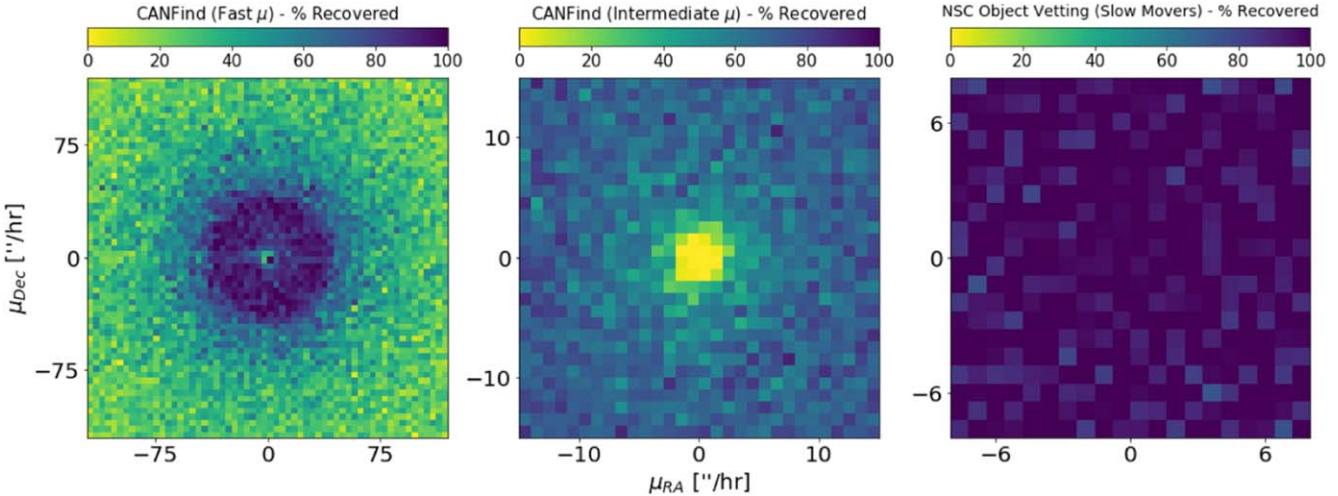
were separately added to unanalyzed NSC measurements from a HEALPix (NSIDE=128) in a field of average object density.

The first set of tests varied the range of input tracklet proper motion with the goal of determining biases set by CANFind’s use of DBSCAN. Simulations of 100,000 tracklets each were run for different ranges of  $\mu$  to explore the upper and lower limits of CANFind’s range. Results for  $0 \leq \mu \leq 100'' \text{ hr}^{-1}$  ( $0.67 \text{ day}^{-1}$ ),  $0 \leq \mu \leq 15'' \text{ hr}^{-1}$  ( $0.1 \text{ day}^{-1}$ ), and  $0 \leq \mu \leq 8'' \text{ hr}^{-1}$  ( $0.053 \text{ day}^{-1}$ ) are shown in Figure 7.

The performance of CANFind for a range of magnitude uncertainties was explored in a second set of tests using an average FWHM value = 1.18 from the NSC exposure table. The range of `magerr_auto` was sampled from several areas including the Galactic plane, ecliptic plane, and in a field away from both planes. Between simulations of 100,000 tracklets each for three ranges of magnitude uncertainties (0.0002–0.002 mag, 0.002–0.02 mag, and 0.02–0.2 mag) CANFind’s detection efficiency only changed slightly (97.6%, 97.2%, and 95.5% for the three ranges, respectively).

In addition to the 600,000 tracklets tested over varying  $\mu$  and `magerr_auto`, a cumulative 200,000 tracklet test over full proper motion and uncertainty ranges found that CANFind recovered 96.5% of synthetic tracklets with average measurement spacing ( $\Delta s$ ) between  $0''5$  and  $10''9$  for proper motions less than  $1.4'' \text{ day}^{-1}$  ( $209'' \text{ hr}^{-1}$ ) to within  $1'' \text{ hr}^{-1}$  of their input proper motion. The limits of CANFind are a result of the parameters used in both instances of DBSCAN. The first iteration ( $\text{eps} = 0''5$ ) to dispose of stationary objects sets the lower limit, (center panel of Figure 7), and the upper limit (left panel of Figure 7) is set by the second iteration ( $\text{eps} = 11''$ ). As a result, this search optimizes CANFind to detect tracklets with cadences between  $0''5$  and  $11''$  ( $0.00012$ – $0.003$ ).

CANFind’s false-positive rate was also tested in fields of varying object density. Fake, linear sky-plane tracklets were created with measurements spanning one night or less and



**Figure 7.** Density plots from the simulation showing three general ranges of tracklet proper-motion  $\mu$  [ $'' \text{ hr}^{-1}$ ]. CANFind results show the algorithm’s upper (left) and lower (center) limits set by the parameters used for DBSCAN. CANFind leaves a gap around  $\mu = 0$  of about  $2.5'' \text{ hr}^{-1}$  ( $0.017 \text{ day}^{-1}$ ) in radius, which is filled in by the NSC object vetting process (right) described in Section 3.3.

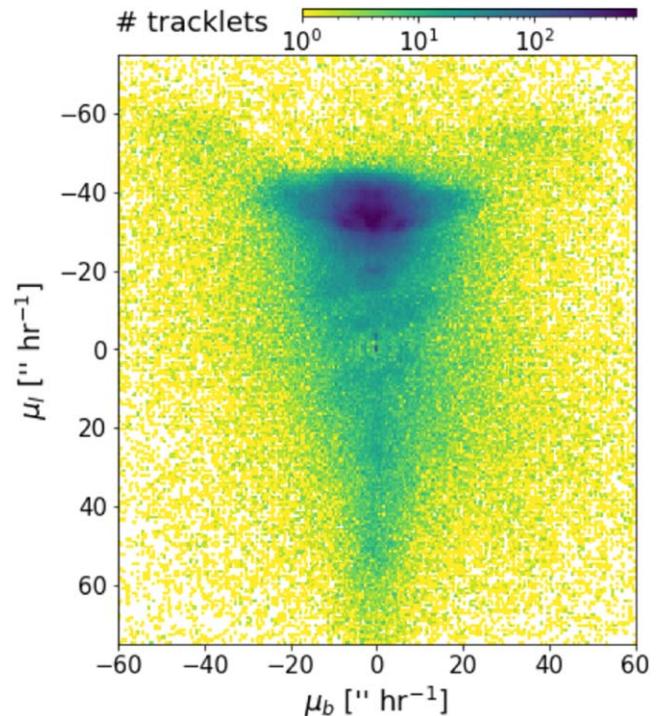
given random, realistic positional uncertainties and observation times. One set of 1,000,000 synthetic “fake” tracklets was analyzed in a HEALPix  $5^\circ$  above the Galactic plane (35,993 objects, 159,191 exposures) and in a HEALPix  $6^\circ$  above the ecliptic (7990 objects, 248,004 measurements). The fake tracklets, like the “true” tracklets created for previous simulations, consisted of 3 to 10 measurements spaced less than  $0.003^\circ$  ( $10''$ ) from each other. In both fields, 3% of fake tracklets were falsely recovered by CANFind.

Finally, the efficiency of the track linking process was tested at a range of velocities between  $0 \leq \mu \leq 60'' \text{ hr}^{-1}$  ( $0.4 \text{ day}^{-1}$ ). A total of 100,000 tracks were created for this test by extending a synthetic tracklet beyond the time frame of one night, either creating a new tracklet or adding unlinked “detections” of the object “observed” at a later time. Track measurements were given the same information given in tracklet simulations, and all complete tracklets created to form tracks were correctly recovered by CANFind. In all, 80% of the synthetic tracks were completely identified by CANFind.

## 5. Results

The CANFind search for SSOs yielded 527,055 tracklets in 35,710 HEALPix of NSIDE = 128 covering 29,971 square degrees of the sky. The density of tracklets in ecliptic proper motion can be seen in Figure 8. The 2094 tracklets detected by the NSC object vetting process fill the range between  $0 \leq \mu \leq 2.5'' \text{ hr}^{-1}$  ( $0.017 \text{ day}^{-1}$ ) left by CANFind (Figure 9). The distribution of tracklets in ecliptic coordinates is shown in Figure 10. A total of 177,949 tracklets (40%) contain only three detections; 170,584 tracklets were detected within 5% of the ecliptic, and 380,189 have negative ecliptic longitudinal proper motions. Tracklets detected at high ecliptic latitudes tend to have high proper-motion values ( $\sim 100'' \text{ hr}^{-1}$ , or  $0.67 \text{ day}^{-1}$ , and above). Figure 11 shows these objects following a color profile similar to stars and galaxies, suggesting a higher rate of false tracklets in regions further from the ecliptic.

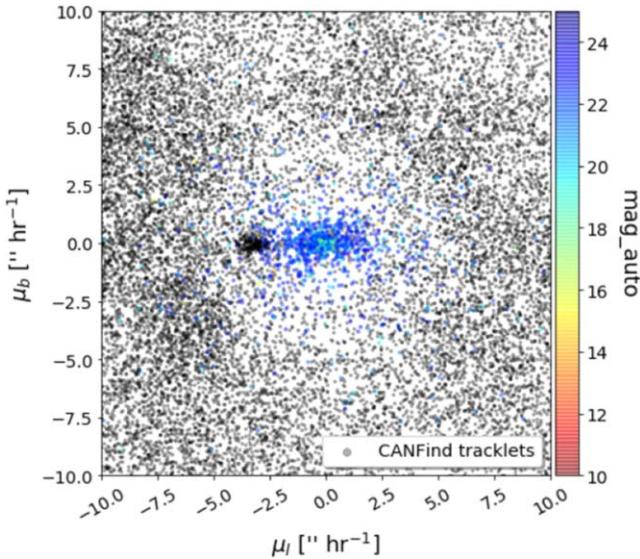
Samples of tracklets from high-density areas (Galactic latitude  $\leq 15^\circ$ ,  $5^\circ$  around LMC,  $3^\circ$  around SMC) were examined in more detail. Despite greater NSC exposure coverage in the Galactic plane, CANFind detected more tracklets near the Magellanic Clouds. A visual inspection of tracklets pulled from



**Figure 8.** Density plot of the proper motions [ $'' \text{ hr}^{-1}$ ] in ecliptic coordinates of tracklets found in the NSC using CANFind. Distinct groups of families are present, most distinctly the main-belt group ( $\mu_l \sim -40'' \text{ hr}^{-1}$  or  $-0.27 \text{ day}^{-1}$ ), Jupiter Trojans ( $\sim -20'' \text{ hr}^{-1}$  or  $-0.13 \text{ day}^{-1}$ ), and some Kuiper Belt Objects ( $\sim -3'' \text{ hr}^{-1}$  or  $0.02 \text{ day}^{-1}$ ).

this area revealed a higher number of false positives detected near the MCs. It is likely that fewer tracklets were found in the Galactic plane due to the crowding of sources and the optimal spacing between exposures in the MCs.

Several clear structures in proper-motion space can be seen in Figure 8, reflecting different groups of SSOs. Most tracklets have proper motions that indicate main-belt orbits with several other general asteroid groups represented as well. Most prominent, centered at roughly  $\mu_l = -40'' \text{ hr}^{-1}$  ( $-0.27 \text{ day}^{-1}$ ),  $\mu_b = 0'' \text{ hr}^{-1}$  in  $\mu$  space is a group of tracklets consistent with MBA proper motion. The presence of Jupiter Trojans and Trans-Neptunian



**Figure 9.** Proper motions of tracklets detected with the main CANFind algorithm (gray), and objects identified by the vetting process (Section 3.4) with color as apparent magnitude. The gap around  $\mu = 0$  left by CANFind (see center panel of Figure 7) is filled in by this extra step.

objects is evident at  $\mu_l = -20'' \text{ hr}^{-1}$  ( $-0.13 \text{ day}^{-1}$ ) and  $\mu_b = 3'' \text{ hr}^{-1}$  ( $0.02 \text{ day}^{-1}$ ). A noticeable feature of the proper-motion distribution are the “wings,” which extend from the MBA group. Terai et al. (2013) describe these structures as a result of the variety of inclination and semimajor axes values of MBAs. A more detailed analysis of each group will require orbital fitting (see Section 6).

Tracklets were detected in all photometric bands included in the NSC ( $ugrizY$  and VR). If a single tracklet contained multiple detections in the same band, their mean value was recorded as the tracklet’s magnitude in that band. Figure 12 shows the normalized distribution of magnitudes for the  $ugr$  bands extending to  $\sim 24$  mag. Table 3 gives the number of tracklets with data in each band. This overall distribution of tracklet detections in the various bands largely reflects the underlying distribution of exposures in the NSC, but may also be affected by the asteroid compositions.

Table 4 lists the six most populous colors represented in the tracklet information. Interestingly, NSC tracklets show a bimodality in color identified in Ivezić et al. (2001), most notably in  $g-r$  versus  $r-i$  (see Figure 13). Objects that fall in the redder group are mainly *S*-type rocky (siliceous) asteroids, while those in the bluer group are *C*-type (carbonaceous) asteroids. A total of 45.5% of tracklets fall roughly into the *C*-type group, with the remaining fraction closer to *S*-type.

A fraction of the tracklet measurements were found in images with exposure times long enough to produce a trailing effect, as no morphology cuts are made on tracklet measurements to remove elongated detections. NSC measurements are found with Source Extractor (Bertin & Arnouts 1996), which can identify objects of various shapes including stars, elongated galaxies, and streaked SSOs. As a result, some trailed measurements of SSOs are identified and present in the NSC measurements table. Figure 14 gives an example of trails left by a tracklet detected by CANFind. For reference, 119 of the 255,454 NSC DR1 images have exposure times  $\geq 1000$  seconds, and 881 have exposure times  $\geq 600$  s. We find an approximate upper limit on the proper motion of

CANFind tracklets containing measurements from these images, proportional to exposure time. This corresponds to the limit on trailed detections represented as single source detections in the NSC. Trails longer than  $\sim 13'$  are not identified by CANFind, as the source extraction process will incorrectly identify multiple sources from a single streak, all with the same observation date. Future analysis will properly identify longer trails.

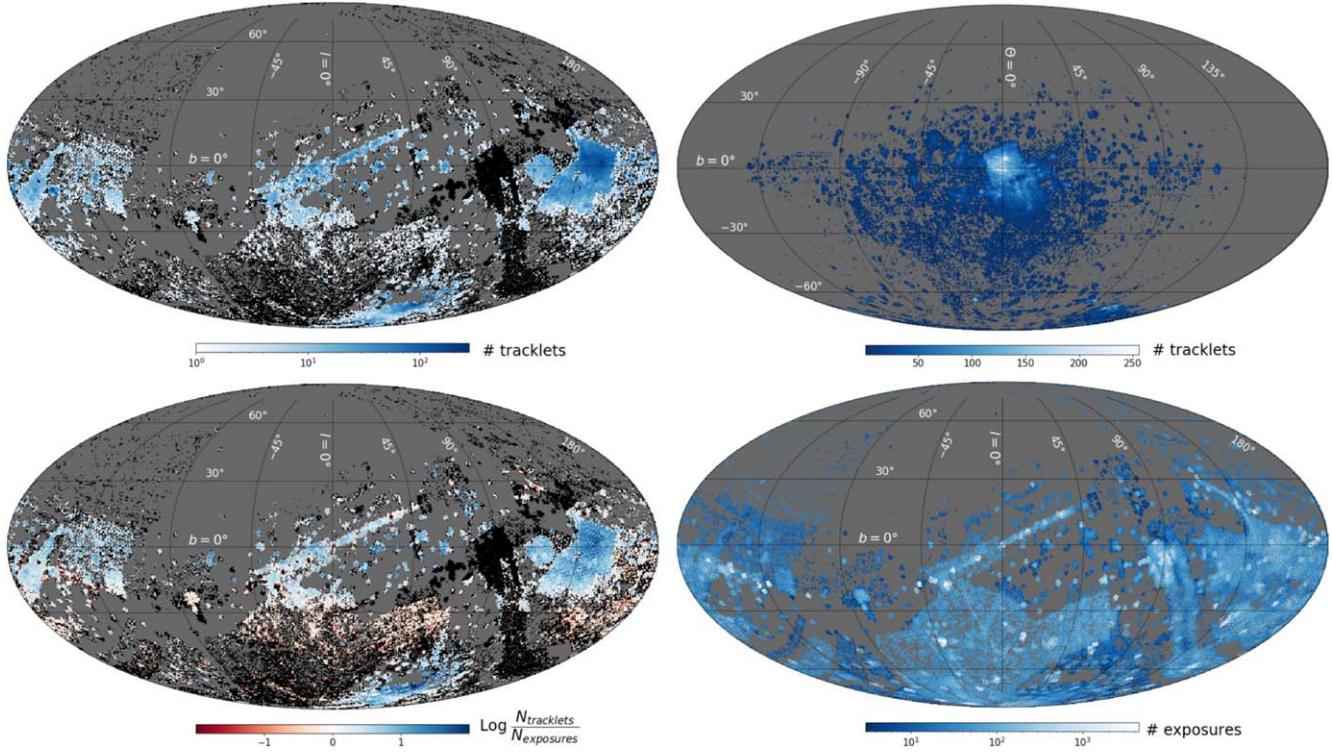
## 6. Future Work

Recovered tracklet measurements, along with measurements not associated with stationary objects, are stored in a PostgreSQL database at MSU for ease of access, similar to the one used to house the NSC in the Astro Data Lab. In the future, we plan to run CANFind on the second data release of the NSC and the same techniques described in this paper will be applied to the new data.

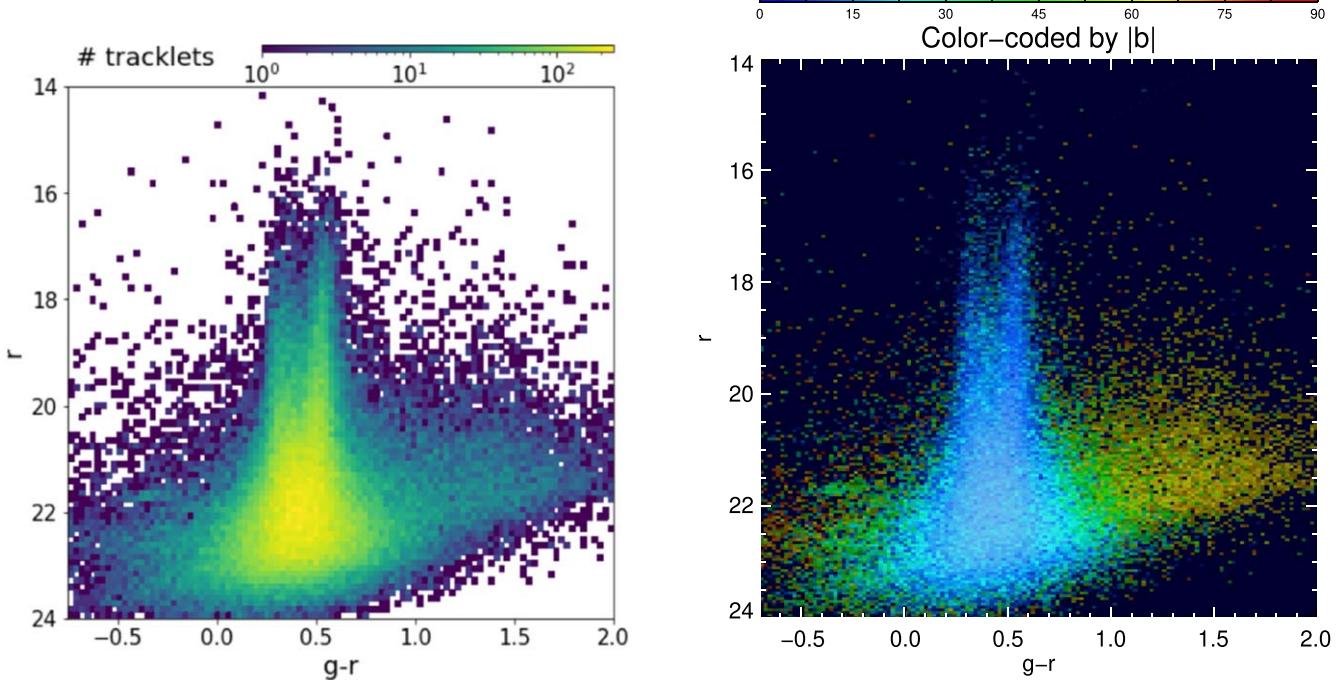
According to Weryk et al. (2020), the Isolated Tracklet File (ITF) of the Minor Planet Center, the organization responsible for the designation of all minor SSOs, holds over 16.6 billion unlinked measurements. Once further developed, CANFind may be applied to the ITF in hopes of reducing its volume.

The following steps will be taken next:

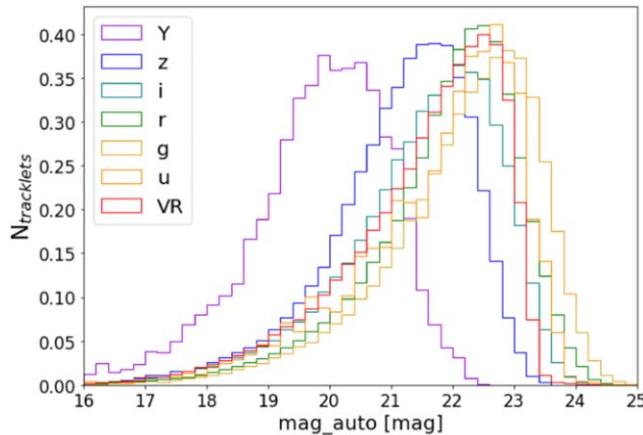
1. *Linking Dispersed Measurements of Moving Objects with the Hough Transform:* While DBSCAN effectively extracts tracklets with closely spaced measurements, it does not identify those with a more dispersed layout. These detections will be connected via the Hough Transform approach, a feature extraction technique traditionally used to isolate features in an image (Duda & Hart 1972), often straight lines or simple shapes. The process is related to the Radon Transform (Radon 1917) and works roughly as follows. Every NSC measurement will cast votes for all trajectory parameters ( $\rho, \theta$ ) consistent with that measurement and the votes from other measurements accumulated in a 2D parameter/accumulator array. Multiple measurements voting for the same line will suggest a candidate trajectory, identified by peaks in the accumulator array. The development of a modified version of the Hough Transform designed to work on data points rather than images is the subject of ongoing work.
2. *Analysis of Trailed Detections:* We plan to identify and analyze streaked detections in NSC exposures to determine further SSO candidates in NSC DR1. While several tracklets detected by CANFind contain trailed measurements, individual streaks remain unidentified. These will be identified using new or existing software on exposures and analyzed using point-spread function fitting.
3. *Orbit Determination:* Due to retrograde motion, linking tracks, tracklets, and unclustered measurements to form potential solar system orbits requires the use of an orbit fitting program designed to account for the effects of the Earth’s motion on the apparent trajectories of other moving objects. Groups of measurements connected by a possible solar system orbit will be designated as a solar system object candidate, and will be studied and reported to the Minor Planet Center. Properties of derived orbits will be analyzed against known objects. Many science questions can be addressed, especially considering the great quantity of observations available in the NSC. Once all tracks and tracklets in NSC DR1 have been identified using CANFind



**Figure 10.** Upper left: tracklet density (number of tracklets per HEALPix, NSIDE = 128) of NSC DR1 in ecliptic coordinates; gray areas had insufficient coverage to be searched, and black areas were searched but contained no tracklets. Upper right: tracklet density in opposition-centric ecliptic coordinates. In place of ecliptic longitude, the  $x$ -axis represents  $\Theta$  = angular separation from opposition, which is  $180^\circ$  from the apparent ecliptic longitude of the Sun at the time of tracklet observation. Tracklets found at high ecliptic latitudes (i.e., LMC, SMC) are more likely to be false positives (see Section 5). Lower right: tracklet exposure density in ecliptic coordinates. Lower left: density of the normalized tracklet/exposure ratio in ecliptic coordinates compares the relative amount of tracklet detection in an area to the amount of exposure coverage.



**Figure 11.** Color-magnitude diagrams of the 107,604 tracklets found by CANFind with  $g$ - and  $r$ -band photometric data. The left panel shows color as density, and the right panel shows saturation as density and color as angular distance from the ecliptic ( $|b|$ ,  $[\circ]$ ). Tracklets that are far from the ecliptic tend to be high in proper motion and do not follow the bimodal color profile characteristic of most SSOs.



**Figure 12.** Normalized histogram of the apparent magnitudes of NSC tracklets extracted by CANFind.

**Table 3**  
Number of NSC Tracklets With At Least One Detection in a Photometric Band

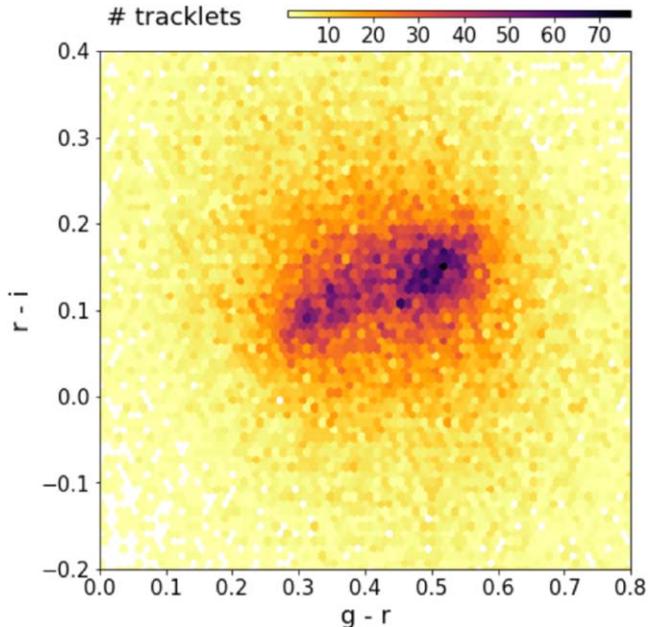
Band	# Tracklets
<i>u</i>	5400
<i>g</i>	138,903
<i>r</i>	161,229
<i>i</i>	129,610
<i>z</i>	47,948
<i>Y</i>	7570
VR	192,178

**Table 4**  
Six Most Populous Colors, “Multicolor” Groups

Color	# Tracklets	Multicolor	# Tracklets
<i>g-r</i>	107,604	<i>g+r+i</i>	50,222
<i>r-i</i>	70,557	<i>r+i+z</i>	16,323
<i>g-i</i>	70,161	<i>g+r+z</i>	9518
<i>i-z</i>	30,706	<i>g+i+z</i>	8505
<i>g-z</i>	13,335	<i>i+z+Y</i>	2806
<i>r-z</i>	22,563	<i>r+i+Y</i>	940

and the Hough Transform approach, they will be fit to possible SSO orbits. We hold off on this step until as many tracklets as possible can be identified in the NSC for the purpose of increasing the likelihood of correctly matching tracklets corresponding to unique objects.

4. *Comparison of Efficiency with THOR:* The effectiveness of these detection methods will be tested against the results of Tracklet-less Heliocentric Object Recovery (THOR; Moeyens & Juric 2019), which recovers known solar system objects by applying test orbits to object detections. THOR applies several orbits that are characteristic of distinct types of solar system objects to detections in a region of the sky, and objects placed in very similar orbits are evaluated and connected. When tested on data from the Zwicky Transient Facility, THOR recovered 97.4% of known asteroids (Moeyens & Juric 2019). However, THOR has not yet been tested or optimized for NEOs and cannot generate initial orbits. We will run THOR on the NSC and compare the



**Figure 13.** Color–color density plot of the 50,222 tracklets found by CANFind with *g*-, *r*-, and *i*-band measurements. In particular, two distinct groups (bluer carbonaceous, left and redder siliceous, right) are visible.

resulting set of solar system objects to those detected by the methods described above.

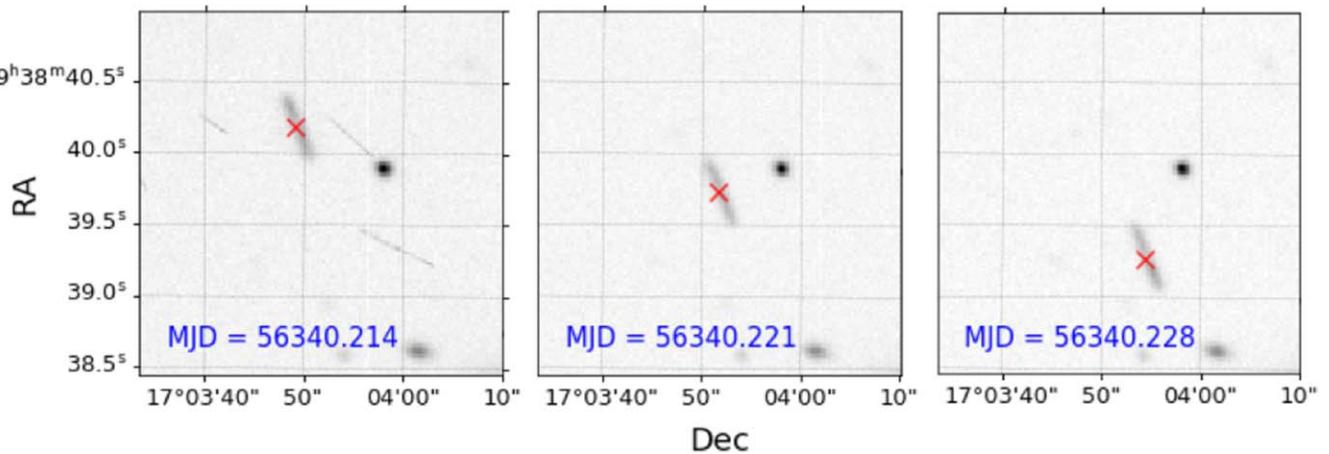
## 7. Summary

We describe CANFind, a new solar system object detection algorithm that we have applied on DR1 of the NOIRLab Source Catalog. We detected 524,501 tracklets in 35,644 HEALPix of NSIDE=128, covering 19% of the sky. The majority of tracklets have photometric data in the *g*, *r*, *VR*, and *i* bands, with fewer detections in *u*, *z*, and *y*. The 50,000 tracklets with measurements in the *g*, *r*, and *i* bands show a bimodality of siliceous and carbonaceous objects, supporting the results of prior studies. In proper-motion space, tracklets fall into groups with motion indicating the presence of Main Belt objects, Jupiter Trojans, Hilda asteroids, and Trans-Neptunian objects. Roughly one third of all tracklets were detected within 5° of the ecliptic.

In the near future, CANFind will be run on NSC DR2, which contains 68 billion measurements of 3.9 billion unique objects (Nidever et al. 2020). The data will also be searched using a modified Hough Transform approach in order to link tracklets with detections that are farther spaced out. Tracklets found will be fit to potential solar system orbits and submitted to the Minor Planet Center.

The CANFind source code and a complete catalog of the tracklets detected in the NSC DR1 are publicly available at <https://github.com/katiefasbender/CANFind>. The tracklet catalog includes information about each tracklet’s mean (R.A., decl.) position, date of first observation, average magnitude in each of the NSC’s filters, and proper motion in both equatorial and ecliptic coordinates [° day<sup>-1</sup>] in FITS format.

This work is made possible by the National Science Foundation’s National Optical-Infrared Research Laboratory (<https://nationalastro.org>) and the National Optical Astronomy Observatory (<https://www.noao.edu>). Computational efforts



**Figure 14.** An example of a tracklet moving fast enough to leave streaked measurements in each 600 s exposure above. These CP-processed DECam exposures from NSC DR1 show 7" trails with their Source Extractor measurements overplotted in red crosses. The trail length for the exposure time matches the proper motion predicted by CANFind from the tracklet's trajectory,  $\mu = 42'' \text{ hr}^{-1}$ . Note: for those interested, the “stationary” object behind the tracklet is white dwarf SDSS J093839.94+170357.7.

were performed on the Hyalite High Performance Computing System, operated and supported by University Information Technology Research Cyber infrastructure at Montana State University.

We thank the anonymous reviewer for the detailed and thorough referee report; the overall quality and scientific merit of this manuscript is vastly improved by their recommendations.

*Facilities:* CTIO:Blanco (DECam), KPNO:Mayall (Mosaic-3), Steward:Bok (90Prime), Astro Data Lab.

*Software:* *Astropy* (The Astropy Collaboration et al. 2018), *IPython* (Pérez & Granger 2007), *matplotlib* (Hunter 2007), *numpy* (Walt et al. 2011), *scipy* (Virtanen et al. 2020), *healpy* (Zonca et al. 2019), *scikit-learn* (Pedregosa et al. 2011).

## ORCID iDs

Katie M. Fasbender  <https://orcid.org/0000-0003-1622-5901>  
David L. Nidever  <https://orcid.org/0000-0002-1793-3689>

## References

Allen, L., Valdes, F., Trilling, D., et al. 2016, AAS/DPS, **48**, 311.09  
 Batygin, K., Adams, F. C., Brown, M. E., & Becker, J. C. 2019, *PhR*, **805**, 1  
 Batygin, K., & Brown, M. E. 2016, *AJ*, **151**, 22  
 Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393  
 Bhattacharya, B., Noriega-Crespo, A., Penprase, B. E., et al. 2010, *ApJ*, **720**, 114  
 Boatmann, A., Beshore, E. C., Garradd, G. J., et al. 2009, AAS/DPS, **41**, 9.02  
 Brown, M. 2017, *AJ*, **154**, 65  
 Brown, M. E., & Batygin, K. 2016, *ApJL*, **824**, L23  
 Brown, M. E., & Batygin, K. 2019, *AJ*, **157**, 62  
 Brown, M. E., Trujillo, C., & Rabinowitz, D. 2004, *ApJ*, **617**, 645  
 Connors, M., Wiegert, P., & Veillet, C. 2011, *Natur*, **475**, 481  
 Cunningham, C. J., Marsden, B. G., & Orchiston, W. 2009, *JAHH*, **12**, 240  
 Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, *MNRAS*, **460**, 1270  
 Dearborn, D. S. P., Bruck Syal, M., Barbee, B. W., et al. 2020, *AcAau*, **166**, 290  
 Denneau, L., Jedicke, R., Grav, T., et al. 2013, *PASP*, **125**, 357  
 Dey, A., Rabinowitz, D., Karcher, A., et al. 2016, *Proc. SPIE*, **9908**, 99082C  
 Dey, A., Schlegel, D. J., Lang, D., et al. 2019, *AJ*, **157**, 168  
 Duda, R. O., & Hart, P. E. 1972, *Commun. ACM*, **15**, 11  
 Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, in Proc. Second Int. Conf. on Knowledge Discovery and Data Mining, KDD'96 (Portland, OR: AAAI Press), 226  
 Fischler, M. A., & Bolles, R. C. 1981, *Commun. ACM*, **24**, 381  
 Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, *AJ*, **150**, 150

Gehrels, T. 1991, *SSRv*, **58**, 347  
 Gladman, B., Davis, D., Neese, C., et al. 2009, *Icar*, **202**, 104  
 Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, *ApJ*, **622**, 759  
 Greenstreet, S. 2020, *MNRAS*, **493**, L129  
 Hogg, D. W., Quinlan, G. D., & Tremaine, S. 1991, *AJ*, **101**, 2274  
 Holman, M. J., Payne, M. J., Blankley, P., Janssen, R., & Kuindersma, S. 2018, *AJ*, **156**, 135  
 Hunter, J. D. 2007, *CSE*, **9**, 90  
 Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, **873**, 111  
 Ivezić, Ž., Tabachnik, S., Rafikov, R., et al. 2001, *AJ*, **122**, 2749  
 Johnson, C. I., Rich, R. M., Young, M. D., et al. 2020, *MNRAS*, **499**, 2357  
 Jones, R. L., Jurić, M., & Ivezić, Ž. 2016, in IAU Symp. 318, Asteroids: New Observations, New Models, ed. S. R. Chesley et al. (Cambridge: Cambridge Univ. Press), 282  
 Khain, T., Batygin, K., & Brown, M. E. 2018, *AJ*, **155**, 250  
 Koeberl, C., & Ivanov, B. A. 2019, *M&PS*, **54**, 2273  
 Lemaître, A. 2005, in IAU Colloq. 197, Dynamics of Populations of Planetary Systems, ed. Z. Knežević & A. Milani (Cambridge: Cambridge Univ. Press), 135  
 Lieu, M., Conversi, L., Altieri, B., & Carry, B. 2019, *MNRAS*, **485**, 5831  
 Mainzer, A., Grav, T., Bauer, J., et al. 2011, *ApJ*, **743**, 156  
 Marchi, S., Black, B., Drabon, N., et al. 2019, *LPICo*, **2134**, 1008  
 Markwardt, L., Gerdes, D. W., Malhotra, R., et al. 2020, *MNRAS*, **492**, 6105  
 Moeyens, J., & Juric, M. 2021, *AJ*, **162**, 143  
 Morrison, D. 1992, The Spaceguard Survey: Report of the NASA International Near-Earth-Object Detection Workshop, NASA STI/Recon Technical Report  
 Nidever, D. L., Olsen, K., Walker, A. R., et al. 2017, *AJ*, **154**, 199  
 Nidever, D. L., Dey, A., Fasbender, K., et al. 2021, *AJ*, **161**, 192  
 Nidever, D. L., Dey, A., Olsen, K., et al. 2018, *AJ*, **156**, 131  
 Parker, A., Ivezić, Ž., Jurić, M., et al. 2008, *Icar*, **198**, 138  
 Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, **12**, 2825  
 Peña, J., Fuentes, C., Förster, F., et al. 2020, *AJ*, **159**, 148  
 Perdelwitz, V., Völschow, M., & Müller, H. M. 2018, *A&A*, **615**, A159  
 Pérez, F., & Granger, B. E. 2007, *CSE*, **9**, 21  
 Popova, O. P., Jenniskens, P., Emel'yanenko, V., et al. 2013, *Sci*, **342**, 1069  
 Pravdo, S. H., Rabinowitz, D. L., Helin, E. F., et al. 1999, *AJ*, **117**, 1616  
 Rabinowitz, D. L. 1991, *AJ*, **101**, 1518  
 Radon, J. 1917, *IEEE Trans. Med. Imaging*, **5**, 170  
 Rest, A., Sinnott, B., Welch, D. L., et al. 2015, in ASP Conf. Ser. 491, Fifty Years of Wide Field Studies in the Southern Hemisphere: Resolved Stellar Populations of the Galactic Bulge and Magellanic Clouds, ed. S. Points & A. Kunder (San Francisco, CA: ASP), 247  
 Rich, R. M. 2015, in ASP Conf. Ser. 491, Fifty Years of Wide Field Studies in the Southern Hemisphere: Resolved Stellar Populations of the Galactic Bulge and Magellanic Clouds, ed. S. Points & A. Kunder (San Francisco, CA: ASP), 59  
 Schlafly, E. F., Green, G. M., Lang, D., et al. 2018, *ApJS*, **234**, 39  
 Schwamb, M. E., Fraser, W. C., Bannister, M. T., et al. 2019, *ApJS*, **243**, 12

Sheppard, S. S., & Trujillo, C. A. 2010, [ApJL](#), **723**, L233

Sheppard, S. S., Trujillo, C. A., Tholen, D. J., & Kaib, N. 2019, [AJ](#), **157**, 139

Stefanut, T., Bacu, V., Nandra, C., et al. 2019, arXiv:[1901.02545](#)

Stokes, G. H., Evans, J. B., Viggh, H. E. M., Shelly, F. C., & Pearce, E. C. 2000, [Icar](#), **148**, 21

Terai, T., Takahashi, J., & Itoh, Y. 2013, [AJ](#), **146**, 111

The Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, [AJ](#), **156**, 123

Tonry, J. L., Denneau, L., Heinze, A. N., et al. 2018, [PASP](#), **130**, 064505

Trujillo, C. A., & Sheppard, S. S. 2014, [Natur](#), **507**, 471

Veres, P., & Chesley, S. R. 2017, [AJ](#), **154**, 13

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, [NatMe](#), **17**, 261

Walt, S. V. D., Colbert, S. C., & Varoquaux, G. 2011, [CSE](#), **13**, 22

Weryk, R., Williams, G., & Wainscoat, R. 2020, AAS Meeting Abstracts, 235, 329.05

Ye, Q., Masci, F. J., Ip, W.-H., et al. 2020, [AJ](#), **159**, 70

Zonca, A., Singer, L., Lenz, D., et al. 2019, [JOSS](#), **4**, 1298

Zou, H., Zhou, X., Fan, X., et al. 2017, [PASP](#), **129**, 064101