Sifting Through the Static: Exploring the Space Beyond Neptune with Digital Tracking

Hayden Smotherman

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:
Andrew Connolly, Chair
Mario Juric
Rory Barnes

Program Authorized to Offer Degree:
Astronomy

University of Washington

Abstract

Sifting Through the Static: Exploring the Space Beyond Neptune with Digital Tracking

Hayden Smotherman

Chair of the Supervisory Committee:
Dr. Andrew Connolly
Astronomy

Trans-Neptunian Objects (TNOs) provide a window into the history of the Solar System, but they can be challenging to observe due to their distance from the Sun and relatively low brightness. Digital tracking helps address these challenges by algorithmically searching many possible TNO trajectories in a stack of images, enabling detection of TNOs too faint to detect in single images. Here we report the detection and characterization of 86 classical TNOs, 5 detached TNOs, 6 resonant TNOs, and 2 scattering TNOs that we could not link to any other known objects. We report measurements of semi-major axis, eccentricity, inclination, longitude of ascending node, argument of pericenter, and time of pericenter passage for these 99 objects. We also report values for the absolute magnitude H in the VR band with a largest measured H value of $H = 9.63$. These objects are dynamically classified using 10 Myr `Rebound` orbital integrations. Additionally, we report the detection of 75 moving objects with short $\sim$ 4 day arcs that we could not link to any other known objects and place constraints on the barycentric distance, inclination, and longitude of ascending node of these objects. We describe extensions to the Kernel-Based Moving Object Detection (`KBMOD`) software that helped enable these detections, including an in-line graphics processing unit (GPU) filter, a convolutional neural network (CNN) stamp filter, and an astrometric and photometric post-processing tool. These tools enable `KBMOD` to take advantage of difference images and help ready `KBMOD` for deployment on future big data surveys such as LSST.

TABLE OF CONTENTS

LIST OF FIGURES

GLOSSARY

CONVOLUTIONAL NEURAL NETWORK (CNN): A neural network algorithm using convolutional layers that is capable of "learning" specific features in images. For example, CNNs can be trained to help reject false positive candidates in `KBMOD`.

DIGITAL TRACKING: A similar method to "shift-and-stack" or "track-before-detect" where images are not necessarily coadded, but rather trajectories may be evaluated based on an aperture that moves in image-space along a potential trajectory.

KERNEL-BASED MOVING OBJECT DETECTION (KBMOD): A "digital tracking" algorithm that uses GPU acceleration to quickly search a large grid of rates and angles.

KUIPER BELT: A belt of small bodies orbiting the Sun with a semi-major axis between about 35 au and 50 au, with low-to-moderate eccentricity and inclination. Sometimes called the Edgeworth-Kuiper Belt.

LAGRANGE POINT: Equilibrium points in the orbits of two massive bodies (e.g. Jupiter and the Sun) where the gravitational forces and the centrifugal force are balanced. The Lagrange 4 (L4) and Lagrange 5 (L5) points are stable equilibrium points 60° ahead (L4) and behind (L5) the orbit of the smaller mass body (e.g. Jupiter).

NEAR EARTH OBJECT (NEO): A small body orbiting the Sun that passes relatively close to the Earth.

POTENTIALLY HAZARDOUS ASTEROID: A small body orbiting the Sun that has the potential to impact the Earth.

SHIFT-AND-STACK: A method of image coaddition where images are shifted along a grid of on-sky speeds and angles then stacked. This allows faint moving objects with a trajectory corresponding to one of the given rates and angles to be recovered with greater signal to noise than in an individual image.

TRANS-NEPTUNIAN OBJECT (TNO): A small body orbiting the Sun with a semi-major axis greater than that of the planet Neptune ($a > 30.1$ au).

TROJAN: A small body typically (but not always) orbiting the Sun that resides in either the L4 or L5 point of a larger co-orbital body.

ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to his supervisory committee and the University of Washington. The author wishes to thank the graduate students, postdoctoral researchers, and professors who worked alongside him throughout the course of his graduate studies. The author wishes to thank his parents for supporting his love of learning. The author wishes to acknowledge all of the teachers who have taught, guided, and challenged him throughout the course of his education. The author wishes to thank his wife for her love and support, without which this work would not have been possible. The author acknowledges support from NASA award 80NSSC21K1528 and NSF award AST-2107800.

DEDICATION

To my dear wife, Stephanie.

Chapter 1

INTRODUCTION

"Things are only impossible until they are not." - Jean-Luc Picard [113]

A doctoral dissertation is an exercise in perseverance with the goal of pushing through many inevitable challenges in order to leave something new behind for those who will come after. Contained herein is my small contribution to the onward march of human knowledge.

This dissertation seeks to explain one method for detecting faint objects in the outer Solar System and why we should care to do so. This introduction will serve to give some needed background information that will help contextualize the rest of the work. This includes a little bit of orbital dynamics, a very brief overview of small bodies in the Solar System, a slightly less brief overview of the objects in the outer Solar System, some things we can learn from studying these populations, and a description of the detection method used in this work.

Much of the work discussed herein is presented in [119] and will be presented in [12], [91], and [120].

1.1 A brief introduction to magnitudes and orbital dynamics

Before discussing different populations of small bodies and their place in the Solar System, we must first discuss some general characteristics of objects moving in elliptical orbits around the Sun.

"How bright is that?" This is perhaps the first and most obvious question one can answer with regard to a light in the sky. This question can be refined by dividing it into two categories. "How bright does that look to me?" and "How bright is that in reality?" To answer these questions, astronomers have long used a logarithmic magnitude scale. This

tradition dates back to Hipparchus, who cataloged how bright the stars in the sky looked to him [24]. We now define the apparent magnitude m to be

$$m = -\frac{5}{2} \log_{10} \frac{F}{F_0}, \tag{1.1}$$

where F is the flux of the object and F_0 is a reference zero point [108].

Of course, we now know that each star is a specific and unique distance away from the Earth. Just like a candle up close may appear as bright as the lights of a distant city, so too will the brightness of a star depend on distance. To account for this, astronomers define the absolute magnitude M to be the apparent magnitude of the objects if it were 10 pc away [24].

Solar System objects, which are much closer to the Earth than even the closest star, have an analogous but different absolute magnitude system. The absolute magnitude H is defined to be the brightness of the object at 1 au. The visual band brightness of a Solar System object is primarily determined by the reflected light from the Sun. Just as the moon changes phase, so too do other Solar System objects change brightness based on the angle of observation. Additionally, when viewed at opposition[1], the brightness of Solar System objects can be different (typically brighter) compared to a smooth, evenly-scattering spherical surface [e.g. 46]. In 1985, to account for this "opposition effect", the IAU added a slope term G to the magnitude system for Solar System objects[2]. Astronomers therefore often use a value for H and G when discussing the absolute magnitude of Solar System objects, where H is given by

[1]Opposition is the point in an orbit when the Sun, Earth, and minor planet are in a line and are in that order. It is the orbital configuration that maximizes the amount of sunlight that is reflected off of the minor planet back to the Earth.

[2]The MPC has an archived version of the IAU general assembly notes discussing this change. See more here: `https://minorplanetcenter.net/iau/ECS/MPCArchive/1985/MPC_19851227.pdf`

the following equations:

$$H = m + 2.5\log_{10}(q(\alpha)) - 5\log_{10}(rd) \tag{1.2}$$

$$q(\alpha) = (1-G)\phi_1 + G\phi_2 \tag{1.3}$$

$$\phi_i = \exp\left[-A_i\left(\tan\frac{\alpha}{2}\right)^{B_i}\right], \tag{1.4}$$

where α is the phase angle, r is the geocentric distance in au, d is the barycentric distance in au, and $A_1 = 3.33$, $A_2 = 1.862$, $B_1 = 0.631$, $B_2 = 1.218$ are fixed quantities (see footnote for the MPC archived version containing these values).

To break down the phrase "moving in elliptical orbits around the Sun" is to cover hundreds of years of history. It has long been known that some of the objects in the sky are "moving". Most of the planets have been observed since ancient times. However, it was not until the 16th century that Nicolaus Copernicus theorized that objects moved "around the Sun" rather than around the Earth and that the Earth itself is a moving object. His theories were not published until just before his death in 1543 [29]. Later, in 1609, Johannes Kepler extended Copernicus' theories to describe the "elliptical" motion of Mars around the Sun [73][3].

Thanks to the works of Copernicus, Kepler, and many others, we can now describe the general motion of two objects bound together by gravitation as they move through space. Namely, they move in an elliptical orbit where the center of mass of the system is one foci of the ellipse. In the case of the Solar System, where the Sun is much, much more massive than other bodies, we can approximate the Sun as being at the foci of this elliptical orbit. In two dimensional space, this motion can be described with just three numbers: a, e, and M. The semi-major axis a describes the size of the ellipse along its longest axis. For objects in the Solar System, a is often described in astronomical units (au). The eccentricity e describes how elongated the ellipse is; e is between zero and one ($0 < e < 1$) for bound orbits. The amount of time it takes for the object to complete one full orbit is called the orbital period

[3]NASA has a brief summary of Kepler, including a short biography, available here: `https://www.nasa.gov/kepler/education/johannes`

P. P is intrinsically related to a by Kepler's third law:

$$P^2 = \frac{G(M+m)}{4\pi^2}a^3, \tag{1.5}$$

where G is Newton's gravitational constant, M is the mass of the larger body (in this case the Sun), and m is the mass of the smaller body (e.g. the Earth). The mean anomaly M describes the fraction of the orbital period that has elapsed since the object has made its closest approach to the center of mass. This point of closest approach is called the pericenter q, while the point of furthest separation from the center of mass is called the apocenter.

Three more numbers (i, Ω, and ω) can describe the orientation of this orbital plane in three dimensional space. The inclination i defines the tilt from some reference plane, which is typically the plane perpendicular to the vector of the total angular momentum of the Solar System. The longitude of ascending node Ω defines the rotation of the orbital plane relative to some reference direction. The argument of pericenter ω defines the position of pericenter along the orbit. These angles are shown in Figure 1.1[4].

These orbital elements do not precisely describe the real motion of objects in the Solar System. Objects do not move around the Sun, strictly speaking. They move around the center of mass of the Solar System, which is called the barycenter. Even the Sun moves around the Solar System barycenter, although since the barycenter is within about two solar radii from the center of the Sun, it admittedly does not move very much when compared to the motion of the planets. Even this is not exactly accurate because objects in the Solar System do not have a fixed set of Keplerian elements. They perturb each other, causing minute changes in their trajectories at any given time. To determine the actual positions and future motion of a given object in the Solar System requires complex numerical simulations where each timestep accounts for the gravitational force of each body on every other body. However, it is nevertheless useful to describe the orbit of objects using the Keplerian elements of what their orbit would be at an instant in time if they were simply orbiting the Sun (or

[4]Lasunncty at the English Wikipedia, CC BY-SA 3.0 `http://creativecommons.org/licenses/by-sa/3.0/`, via Wikimedia Commons

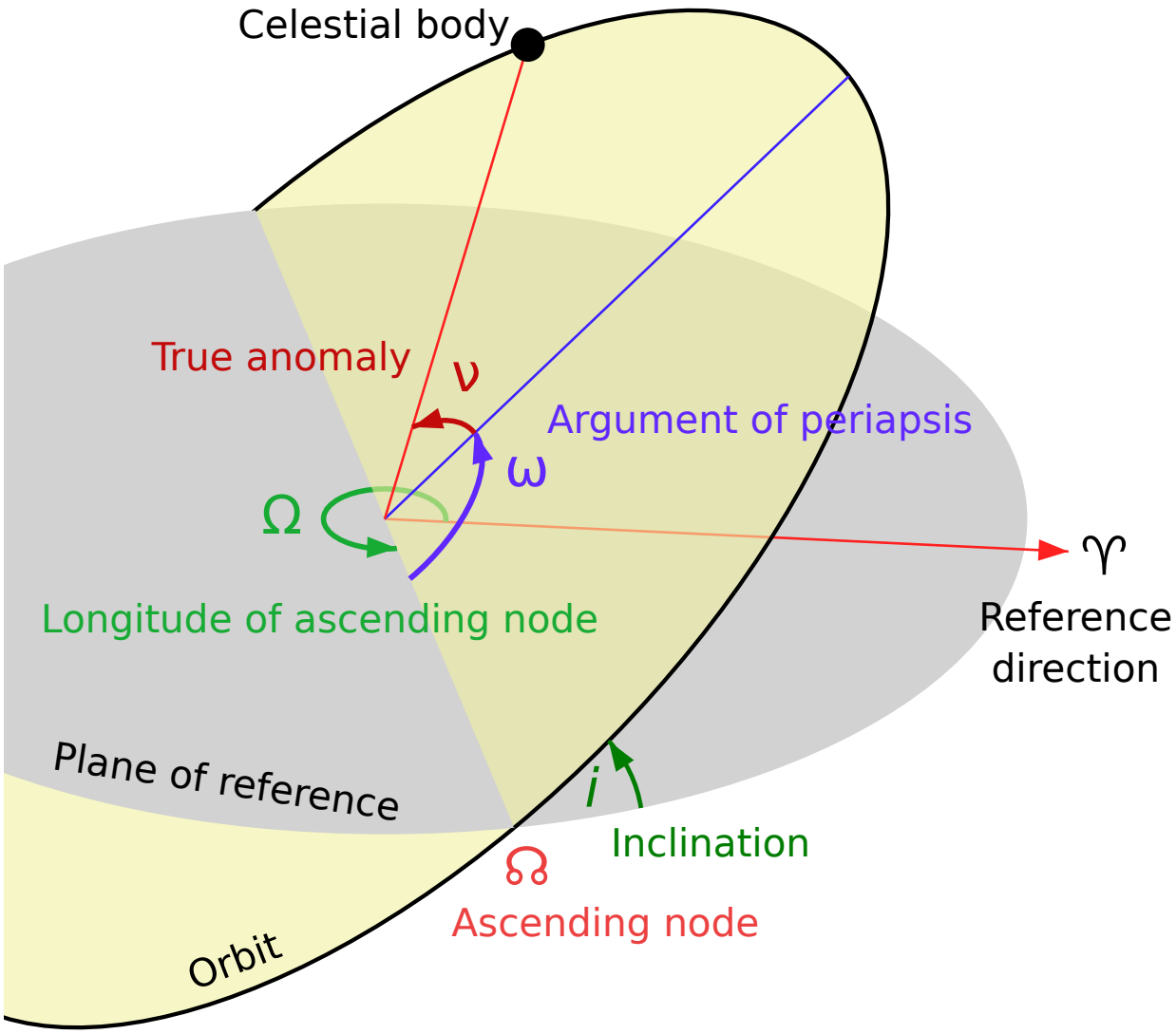


Figure 1.1: A visualization of the angles of the Keplerian elements in three dimensional space. Ω, ω, and i describe the orientation of the plane of the elliptical orbit in three dimensional space.[4]

the barycenter) with no other perturbing bodies.

These elements, sometimes called osculating elements, help to separate small bodies into distinct populations. For example, one can define a number of different types of small bodies based primarily on their orbital parameters. In the next section, we will briefly describe some of the different groups of small bodies, as well as provide a bit of historical context and interesting tidbits about each group.

1.2 A brief summary of small bodies in the Solar System

Small bodies are the final frontier in the study of flux-limited populations in the Solar System. While these objects are primarily very small and often very distant, it is nevertheless critical that we strive to understand them. Some can tell us about the formation of the Solar System and its early history. Others have the potential to violently impact the Earth, destroying cities, countries, or even entire species. In this section, we will briefly describe many of the main classes of small bodies in the Solar System, very loosely starting with objects closest to Earth and moving to more distant populations.

While it is not the purpose of this work, we will first present a very brief summary of the formation of small bodies in the Solar System and their role in the formation and evolution of the Earth [131]. Like many stars, the Sun likely formed with a protoplanetary disk largely made up of H_2 and He, with a smaller amount of dust. The Sun would have formed from nebular material with a gas-to-dust ratio of about 100, but disks generally have smaller gas-to-dust ratios and hence proportionately more dust [5]. Dust particles grow to cm sizes, at which point the streaming instability begins to dominate [138]. The streaming instability can be understood as follows. Gas rotates at sub-Keplerian speeds (i.e. slower than an orbiting body). Thus, the cm size particles experience a headwind and begin to migrate inwards towards the star. Where many of these particles accumulate, they speed up the local gas, lessening the headwind, which in turn makes the cm particles migrate more slowly. This allows for a greater pile up of particles, which can eventually become self-gravitating and form planetesimals of 100 km sizes [88]. The gaseous disk around

the early Sun likely dissipated after only a few Myr [133]. Within about 100 Myr after the dissipation of the gaseous disk, the terrestrial planets are thought to have formed from collisions between planetesimals and the forming planetary embryos [134]. The Earth is thought to have formed with minimal *in situ* water content[5], in part because the ratio of deuterium to hydrogen (called the D/H ratio) is much higher in the Earth's oceans than in protosolar matter [34]. After formation, water and other volatiles are thought to have been delivered via impacts primarily by C-type asteroids (carbonaceous chondrites), with a $< 10\%$ contribution (possibly $< 1\%$) from comets [131, 87, 83]. See [131] for a comprehensive review. Impacts therefore played an important role in the formation of the planets as we know them today. However, these impacts can still occur today, sometimes with devastating consequences.

Near Earth Objects (NEOs) are small bodies that come relatively close to Earth. More specifically, they are objects with a pericenter of less than 1.3 au[6]. As of February 2022, the JPL small body database lists 28317 NEOs[7]. These objects are of interest to many people, both astronomers and the general public, for many reasons, not least of which is the fact that some of these objects come close enough to the Earth that they could potentially impact our planet.

The subclass of NEOs that could potentially impact Earth are known as Potentially Hazardous Asteroids (PHAs), provided they are sufficiently bright. As defined by JPL, PHAs must have an Earth "Mininum Orbit Intersection Distance" (MOID) of less than 0.05 au. Additionally, they must have an H magnitude of $H \geq 22.0$, which corresponds to a size greater than about 140m[6]. As of February 2022, the JPL small body database lists 2227 PHAs. These objects have the potential to cause regional-to-global devastation if they

---

[5][131] considers this to be the scientific concensus, but they note that there is some discussion of *in situ* water accumulation and formation [e.g. 33].

[6]JPL has a summary page of NEOs here: `https://cneos.jpl.nasa.gov/about/neo_groups.html`

[7]The JPL Small Body Database provides a convenient query tool for small bodies, allowing a user to query the names, orbital elements, and physical properties of small bodies in the Solar System. It can be found here: `https://ssd.jpl.nasa.gov/sb/`

were to impact the Earth. The most well-known example of such a destructive impact is the meteor that is believed to have caused the extinction of the dinosaurs. This meteor, sometimes called the Chicxulub impactor, was quite large with a diameter of about 10 km [32].

However, even a much smaller asteroid could cause city-wide devastation upon impact. In fact, on June 30, 1908, an object that was likely smaller than 140 m exploded over the forests in the Tunguska region of Siberia, releasing about 20 Mt of TNT worth of energy and damaging about 2,000 square km of forest [135, 6]. Over one hundred years later, on February 15th, 2013 a meteoroid that was only about 15-20 m exploded in the atmosphere over the Russian city of Chelyabinsk releasing about 300 kt of TNT worth of energy and shattering windows over a 5,000 square kilometer area [6]. Not all small bodies in the inner Solar System come so close to Earth, however.

Main Belt Asteroids (MBAs) have a semi-major axis between Mars and Jupiter. As of February 2022, the JPL small body database lists 1045489 MBAs. Although there are a large number of MBAs, the total mass of the asteroid belt is not actually that large. Estimates place the total mass around $18 \times 10^{-10} M_{\odot}$, which is under 5% the mass of the Earth's moon. Furthermore, using the values for GM from the JPL small body database[8], the three largest asteroids (Ceres, Vesta, and Pallas) in the asteroid belt account for about 39% of the total mass.

Trojan asteroids are in stable equilibrium points 60° ahead and 60° behind an orbiting body. These points are called the L4 and L5 Lagrange points [77]. Lagrange points are orbital configurations in the three body problem where, assuming one of the bodies is of negligible mass, the gravitational force of the other two more massive bodies is balanced by the centrifugal force of the negligible-mass body. The L4 point is 60° ahead of the main body, while the L5 point is 60° behind the main body.

It was over a century after the theoretical description of Lagrange points before an object

[8]The JPL small body database directly reports GM, where G is Newton's gravitational constant and M is the mass of the asteroid.

was found in one of these points. An object was discovered in the jovian L4 point by German astronomer Max Wolf in 1906 in Heidlberg. It was soon named "Achilles" and was followed later that year by an object named "Patroclus", then the next year by "Hektor" [1]. Other small bodies detected in the jovian Lagrange points followed the same naming scheme, all with names from Homer's *Iliad*. Objects in the leading L4 point are named after Greeks, while those in the L5 point are named after Trojans, except for "Hektor" and "Patroclus", which are in the L4 and L5 points respectively[9].

Trojans have been detected for most planets. Jupiter has the greatest number of known trojans by far, with 11631 listed by the JPL small body database as of February 2022, but every planet besides Saturn and Mercury has at least one known trojan. Earth has two known trojans, both in the L4 point [28, 114]. Uranus trojans are thought to be largely unstable, and the two known Uranus trojans are only stable for 100 kyr timescales [98, 4, 31]. In addition to the planetary trojans, several smaller bodies in the Solar System have their own trojans, including two of the moons of Saturn [117]. Telesto and Calypso are two saturnian moons that are in the L4 and L5 points of the larger saturnian moon Tethys. Helene and Polydeuces are similarly in the L4 and L5 points of the saturnian moon Dione. The two most massive asteroids in the asteroid belt even have their own temporary trojans, although they are only stable for about 2 Myr [27].

Comets are perhaps some of the most spectacular small bodies. As of February 2022, the JPL small body database lists 3782 comets. Comets are icy objects that spend most of their time in the outer Solar System, but they have orbital trajectories that bring them into the inner Solar System. When this occurs, they heat up and sometimes obtain dramatic and bright tails. In fact, of all of the small bodies mentioned here, comets were the first to have been observed and recorded. Chinese oracle bones dating to between 1400 B.C. and 1200 B.C. provide the earliest written record of a comet observed by humans[10].

---

[9] A MPC list of Jupiter trojans can be found here: `https://minorplanetcenter.net//iau/lists/JupiterTrojans.html`

[10] An archived page discussing Chinese oracle bones and their relationship to comets can be found here: `https://web.archive.org/web/20131005100532/http://www.lib.cam.ac.uk/mulu/oracle.html`

Perhaps the most famous comet is Halley's Comet. This object bears the name of Edmund Halley, who discussed with Sir Isaac Newton the comet observations of 1531, 1607, and 1682, theorizing that they were all observations of a single object which reappeared about every 75 years. Although unknown to Edmund Halley, this comet had actually been observed by humans for hundreds of years, with possible (although uncertain) recorded observations as far back as 466 B.C. [56]. Halley's comet is still around today, most recently visible with the naked eye in 1986 and due for another appearance in 2061.

Centaurs are objects that, like comets, are interlopers into the orbital space between planets. Unlike comets, however, they are constrained strictly to the outer Solar System. The JPL small body database lists 703 objects with a semi-major axis of 5.5 au $< a <$ 30.1 au. [50] defines a centaur as having a $T_j > 3.05$, $q > 7.35$ au, and $a < 30.1$ au, where T_J is the Tisserand parameter with respect to Jupiter. T_J is given by Equation 1 in [50], reproduced here:

$$T_J \equiv \frac{a_J}{a} + 2\sqrt{\frac{a_J}{a}(1-e^2)\cos i}, \tag{1.6}$$

where a_J is the semi-major axis of Jupiter. The Tisserand parameter quantifies how strongly a small body interacts with a large body, such as Jupiter. Therefore, the centaur definition from [50] defines these objects to be in the outer Solar System, not strongly influenced by Jupiter, but not beyond the orbit of Neptune (and therefore not trans-Neptunian Objects).

Far in the outer reaches of the Solar System, often beyond the planets themselves, there is a wealth of small bodies. Trans-Neptunian Objects (TNOs) are, as the name implies, simply small bodies with orbits beyond that of the planet Neptune. As of February 2022, the JPL small body database lists 4159 TNOs. Despite the fact that we call them "small bodies", they are only small with respect to the planets. These icy objects can actually be quite large in size, with diameters of over a thousand km. Furthermore, they take hundreds of years to complete a single orbit around the Sun. The rest of this work will focus primarily on TNOs.

Many of the populations listed here can be further sub-divided on, for example, orbital parameters or composition. Because this work primarily concerns itself with the slow-moving

bodies in the outer Solar System, we will take some time to describe TNOs and other distant objects in more detail, separating them into orbital classes and describing some key physical characteristics of these bodies.

1.3 Populations of small bodies beyond Neptune

As discussed above, there are thousands of objects beyond the orbit of Neptune, many of which belong clearly to distinct subpopulations. These subpopulations, as well as the objects that are difficult to classify, can tell us much about the origin and history of the Solar System. This will be the focus of Section 1.4, but first we need to describe these populations of TNOs. They will provide the scientific background for the work discussed in Chapters 2, 3, and 4.

One of the main populations in the outer Solar System is called the Kuiper Belt. The Kuiper Belt is named after the astronomer Gerard Kuiper, who wrote about planetary formation hypotheses in the mid 20th century [75, 76]. Although he did not predict the Kuiper Belt as we know it today, he did discuss formation methods that would form small bodies out beyond the orbit of Neptune. From [76]:

> The outermost region of the solar nebula, from 38 to 50 astr. units (i.e., just outside proto-Neptune), must have had a surface density below the limit set by equation (7)... Condensation products (ices of H20, NH3, CH4, etc.) must have formed, and the flakes must have slowly collected and formed larger aggregates, estimated to range up to 1 km. or more in size. The total condensable mass is about 10^{29} g., but not all of this could be collected. These condensations appear to account for the comets, in size, number and composition.

The Kuiper Belt is sometimes called the Edgeworth-Kuiper Belt, after both Kenneth Edgeworth and Gerard Kuiper. The contribution of Edgeworth is noteworthy because he predicted the existence of belt of objects beyond Neptune in 1949, predating the publications of [75, 76]. From [36]:

> In the region outside the orbit of Neptune the material would also be highly attenuated, and here again condensations would be small and numerous, but the progress of evolution was slower, and the region is probably populated by a very large number of small clusters.

The very first Kuiper Belt Object (KBO) was discovered on February 18th, 1930 by Clyde Tombaugh at the Lowell Observatory in Flagstaff, Arizona. The object was of course Pluto, and it was not until much later that it became known as a KBO. Originally, Pluto was called the ninth planet in the Solar System. Pluto has a diameter of 2376 km and a mass of 1.3029×10^{22}, making it smaller than Earth's moon. Now called a dwarf planet, Pluto has 5 known natural satellites, the largest of which is call Charon. According to the JPL small body database, Charon has a mass that is about 12% that of Pluto. Pluto is in a 3:2 mean-motion resonance with Neptune.

The next KBO was not discovered until 1992 [66]. The object was named 1992 QB1, and later renamed Albion. This object has an inclination of about $i = 2.2°$ and an eccentricity of about $e = 0.11$. Because of this, other KBOs with similarly low inclination and eccentricity are sometimes called "cubewanos", after the initial name "QB1". Since the discovery of Albion, over 4000 unique TNOs have been discovered, leading to both a broader and deeper understanding of objects in the outer Solar System.

Objects in the Kuiper Belt may still be called "small bodies", but they are not actually all that small. TNOs in general are only "small" relative to the planets. The JPL small body database lists 2869 TNOs with $H \leq 8$, which corresponds to a size of approximately greater than 135 km, depending on the albedo. Known TNOs extend in size all the way up to Pluto, with a diameter of 2376 km, although there is some reason to suspect there may still be larger undiscovered TNOs (see Section 1.5). In orbital space, the classical Kuiper Belt extends from about 30 au to about 50 au, although the outer edge is loosely defined.

TNOs can be further subdivided into classical, resonant, detached, and scattered (or scattering) populations [50, 74]. The classical Kuiper Belt can be further separated into

dynamically "hot" and dynamically "cold" populations. Additionally, there are 25 TNOs listed by the JPL small body database with $a > 250$ au and $q > 30$ au, which are sometimes called "Extreme TNOs" [e.g. 11]. We will briefly describe each of these subpopulations, both in terms of their orbital dynamics as well as some of their other properties.

1.3.1 Classical KBOs

Using the classification scheme of [50], the classical Kuiper Belt consists of non-resonant TNOs with $e < 0.24$. [50] divides the classical Kiuper Belt on semi-major axis, separating the classical Kuiper Belt into "inner", "main", and "outer" sections. They place this division at $a > 39.4$ au, $39.4 \leq a \leq 48.4$ au, and $a > 48.4$ au respectively.

The classical Kuiper Belt is also sometimes separated into dynamically "hot" and dynamically "cold". The cold classical Kuiper Belt is often defined as objects where $i < 5°$, while the hot classical Kuiper Belt consists of objects in the classical Kuiper Belt with $i > 5°$ [21, 88].

Although this cut on inclination is not always used (e.g. [50, 74]), there are a number of other observables that suggest that there are two different subpopulations in the classical Kiuper Belt, each with unique origins. Specifically, color, size distribution, and binary fraction seem to indicate at least two physically distinct subpopulations.

Cold classicals tend to be redder than hot classicals. In 1998, [122] was the first paper to identify two subpopulations of KBO colors. Using only 11 KBOs, they did not identify any dependence on "inclination angle, eccentricity, perihelion distance, semi-major axis, or absolute magnitude". However, by 2002, [126] used color data from 60 KBOs to identify an inclination dependence in the Kuiper Belt; namely, lower inclination KBOs are redder than higher inclination KBOs. Further studies substantiated this relationship, and show that cold classical KBOs are nearly all relatively red, while more dynamically excited populations have a broader "grey" distribution of colors [106, 22]. It is worth mentioning that the inclination of $i < 5°$ used to divide the hot and cold populations dynamically does not agree with the inclination break of $i < 12°$ that [106] finds in the color distribution. Recent work ([41, 115])

suggests that three progenitor populations may reconcile KBO dynamical inclination classes with KBO color distributions.

Cold classicals are lower-mass and have a different size distribution than both hot classicals and Jupiter trojans. [15] identified that KBO magnitudes, both "cold" and "excited", follow a double power law. [42] extends this to the absolute magnitude H, and shows that the hot and cold populations do not come from the same size distribution. [42] also estimates the total mass of both the hot and cold classical populations at about 0.01 M_E and 0.0003 M_E respectively, where M_E is the mass of the Earth.

Cold classicals have a relatively high fraction of binary systems. In fact, [41] states quite clearly that all objects that formed in the region of the present Kuiper Belt would have formed as binaries. There is, however, a lack of widely-separated binaries in the dynamically excited TNO populations (including but not limited to hot classical KBOs) [102]. KBO binarity provides an important constraint on formation hypotheses. Although the hot populations can be strongly dynamically perturbed (interrupting any widely-separated binaries), any formation model must reproduce and preserve the large fraction of binaries in the cold classical population, as well as the distribution of prograde versus retrograde binaries. For example, [94] leverages this information to provide further support for the streaming instability planetesimal formation model, which they find reproduces the large fraction of cold classical binaries as well as the high (approximately 80%) fraction of prograde binaries.

Hot classicals share some similarities to Jupiter trojan asteroids. More specifically, [137] demonstrates that dynamically hot KBOs and the Jupiter trojans have a statistically indistinguishable color-magnitude distribution. Similarly, [42] is unable to rule out the possibility that hot KBOs share a size distribution with the Jovian trojans. This similarity suggests that both the hot KBOs and the Jovian trojans may have come from the same progenitor population. This observation has important impacts on theories concerning the origin of the current outerplanetary configuration, as we will discuss in Section 1.4.

1.3.2 Resonant TNOs

We can define a specific kind of orbital relationship between two bodies orbiting a third parent body. An object is in a mean-motion resonance (MMR) with another object if their orbital periods are some integer ratio of each other. In this paper, we well describe resonances as the number of orbits of the inner object to the number of orbits of the outer object. For example, if the inner object completes 3 orbits in the time it takes the outer object to complete 2 orbits, we call this a 3:2 resonance.

Some trans-Neptunian objects are in resonance with one of the giant planets, typically Neptune. Resonant TNOs are interesting for a few reasons. TNOs in resonance can have high eccentricity, yet remain stable for Gyr timescales [50]. The relative amount of resonant objects, as well as their physical properties, can tell us about the orbital evolution of Neptune in the early Solar System [88]. See Section 1.4.

The most commonly discussed TNO resonances are the 3:2 and 2:1 resonances. However, [50] lists 17 different mean-motion resonances with at least one TNO that occupies the resonance. An integer orbital period ratio is necessary but not sufficient for a TNO to be in resonance with Neptune. Rather, there must be a particular agreement of Ω, ω, and M [50]. In particular, using the example of [50], a TNO in the 9:4 resonance with Neptune must be such that the resonant argument oscillates around a particular value. The resonant argument Φ_{94} for 9:4 MMR is given by Equation 2 in [50], repeated here:

$$\Phi_{94} = 9\lambda_N - 4\lambda - 5\varpi, \tag{1.7}$$

where $\lambda = \Omega + \omega + M$ is the mean longitude of Neptune (λ_N) and the TNO (λ), and $\varpi = \Omega + \omega$ is the longitude of pericenter. Numerical simulations are usually required to show that a particular TNO is in a mean-motion resonance.

Pluto is the most famous example of a resonant TNO, falling in the 3:2 resonance, and is both the largest TNO and the first TNO to be discovered. For this reason, other TNOs in the 3:2 resonance are sometimes called “plutinos”.

1.3.3 Scattered disk objects, the detached population, extreme TNOs, and the Oort Cloud

In addition to the resonant TNOs and classical KBOs, some TNOs are highly eccentric, highly inclined, very distant, or all three. These objects can be categorized in a number of different ways, which we will summarize below.

The first category of objects is called "scattered disk objects" (SDO). These objects interact strongly with Neptune. The dynamics of these objects were first discussed in [35], who carried out 4 Gyr orbital integrations of 2200 massless test particles that interacted with Neptune. They found that about 1% of these objects remained after the 4 Gyr simulation time. [50] makes a point to call these objects "scattering" rather than "scattered" in order to specify that they are defining these objects to be currently interacting with and scattering off of Neptune, rather than that they were scattered sometime in the past.

In addition to the population currently scattering off of Neptune, a "detached" population of TNOs is also commonly defined [50]. [88] calls this population the "fossilized scattered disk", which makes a suggestion as to the origin of this population. [50] defines the detached population to be TNOs with eccentricity $e > 0.24$ that do not have sufficiently large semi-major axis to be influenced by extra-solar forces. [50] puts this limit at around $a > 2000$ au, corresponding to the approximate start of the inner Oort Cloud.

Eris is perhaps the most famous example of a fossilized scattered disk or detached object, although it is sometimes also referred to as simply a scattered disk object. It has a semi-major axis of $a = 68.1$ au, an inclination of $i = 43.8°$, and an eccentricity of $e = 0.43$, making it significantly more distant, more eccentric, and more highly inclined than classical KBOs[7]. It is actually the most massive TNO, despite being smaller than Pluto, and was initially announced to be the tenth planet in the Solar System[11]. However, both Eris and Pluto were subsequently classified as "dwarf planets"[12].

---

[11]The archived version of the JPL press release is available here: `https://web.archive.org/web/20070514171359/http://www.jpl.nasa.gov/news/news.cfm?release=2005-126`

[12]The archived IAU resolution defining dwarf planets and reclassifying Pluto is available here: `https://web.archive.org/web/20060928081931/http://www.iau.org/fileadmin/content/pdfs/Resolution_GA26-5-6.pdf`

[88] notes that, using a definition similar to [50], there appear to be two distinct sub-populations of detached TNOs. At lower perihelion distances of $q < 45$ au, the detached population appears to be a more distant, more excited extension of the hot population of the classical Kuiper Belt. However, in addition to this population, there are currently three known objects with $q > 60$ and $a > 250$ au. The first such object detected is the object Sedna, which is thought to be a member of a larger population of sednoids [23]. Sedna is an object with a perihelion of $q = 76 \pm 4$ au but a semi-major axis of $a = 480 \pm 40$ au. In fact, Sedna has the second largest perihelion (after 2012 VP113) of any detected Solar System object. Therefore, it spends well over 90% of the time on its orbit beyond the detection limit of the survey that discovered it. This suggests that there may be a numerous and massive population of objects at high q.

[11] defines a general category of "Extreme TNOs" (eTNOs) to be those objects with $a > 250$ au and $q > 30$ au. This population would include the three known sednoids and an additional 22 other distant and highly-eccentric. Further surveys of the outer Solar System, such as LSST and DEEP, will be needed in order to determine whether or not the detached population should be subdivided and if so, how. However, this subsection has defined several of the commonly-used subclassifications for the very distant, very eccentric, and/or very inclined populations of TNOs.

Well beyond the edge of the classical Kuiper Belt lies the Oort Cloud. [50] places the dividing line between the Kuiper Belt and the Oort Cloud at $a = 2000$ au. This distance is chosen such that extra-solar effects like passing stars and galactic tides can have an appreciable effect on the orbital evolution of these objects. There are currently no known objects with $a > 2000$ au and $q > 30$ au. This region is, therefore, a currently un-analyzed region of TNO parameter space. It is worth noting that there are 175 comets and 2 asteroids with $a > 2000$ au and no limit on q, but such low q objects are not the primary topic of this work.

1.4 Models of the early outer Solar System

TNOs contain many dynamically-unperturbed relics from shortly after the formation of the Solar System [81]. They provide a window into the early history of the Solar System and enable tests of planetary formation and migration hypotheses. In this section, we will discuss several models of planetary migration in roughly historical order, with a focus on the movement of the giant planets after the gaseous disk dissipated. One of the earlier and more famous models, the Nice Model [127], suggests that all the giant planets formed well-interior to 20 au and migrated outwards due to interactions with each other and with a large disk of planetesimals. Later models built on this concept, refining the details to meet observational constraints. In this section, we will start with a description of the original Nice Model then move on to more modern planetary migration models.

The Nice Model was the first to predict a giant planet instability in the early Solar System [127]. In this initial model, [127] started a numerical orbital integration with the giant planets all inside of 17 au. They also include a massive belt of planetesimals from the edge of the outer planetary configuration (about 17 au) to $30 - 35$ au. This belt had a mass between 30 and 50 M_E, where M_E is the mass of the Earth, and was made up of 1,000-5,000 equal mass bodies. (As we will discuss later, subsequent research based on the capture probability and current population of Jupiter trojans places this disk mass closer to $15 - 20$ M_E [99]).

Using these simulations, the Nice Model predicts that Jupiter and Saturn enter a 2:1 MMR, which causes Neptune and Uranus to dramatically "swap" positions [127]. This happens because the giant planets exchange angular momentum with the massive disk of planetesimals, during which time Jupiter drifts inwards while Saturn, Uranus, and Neptune drift outwards. When this drift occurs sufficiently close to a MMR between Jupiter and Saturn, the eccentricities of these planets are driven up, which in turn drives up the inclinations and eccentricities of Uranus and Neptune, sending them out into the massive belt of planetesimals. At this point, dynamical friction [e.g. 25] causes the circularization of the orbits of

Uranus and Neptune, leaving them in the outer Solar System. In cases where Saturn interacts with one of the ice giants, [127] found agreement between their models and the current outer planetary configuration to within two standard deviations of their model estimates for a, e, and i. This explains the relatively high inclination and eccentricity of the outer planets relative to what might be expected if they had formed in situ in a protoplanetary disk.

The initial jump in the eccentricity of Jupiter and Saturn is expected due to the divergent nature of the resonance crossing. As mentioned in Section 1.3.2, to be in a stable resonance, the resonance argument Φ must oscillate or "librate" about a particular point. For example, [82] describes the mechanisms for the resonant capture of Pluto due to Neptune's outward migration. However, there are regions in $e\cos\Phi$ and $e\sin\Phi$ space where capture into this stable libration does not occur. In short, as the two planetary semi-major axes approach exact MMR resonance, the bodies move from a region of low-eccentricity "internal circulation" to a high-eccentricity "external circulation". See the Supplementary Material from [127] for a description of this as applied to the Nice Model. See [60] for a more in-depth description of the mechanics of this divergent resonance crossing as well as an application to the formation of the Kirkwood Gaps in the asteroid belt.

The Nice Model was initially invoked to explain the Late Heavy Bombardment (LHB), which is thought to have occurred about 700 Myr after the formation of the giant planets [52]. [52] found that they can delay the crossing of the Jupiter-Saturn 2:1 MMR by up to 1.1 Gyr depending on the location and mass of the outer disk of planetesimals, as well as the semi-major axes of Saturn, Uranus, and Neptune. Specifically, if the inner edge of the disk of planetesimals is close to the initial orbit of Uranus, the instability will happen quickly, while if it is far away, the instability will happen more slowly.

Current studies suggest that the timing delay described in [52] likely could not have occurred in the Solar System. Observations of the Jupiter trojan binary Patroclus–Menoetius and subsequent dynamical simulations show that this binary pair would likely have been disrupted if the predicted giant planet instability occurred more than about 100 Myr after the dispersal of the gas disk [96]. This timeline is determined by the rate at which binaries in a

massive planetesimal disk from $20-30$ au would have been disrupted by collisional grinding. After a period of giant planet instability, the Jupiter trojans and the hot population are thought to be derived from this planetesimal disk. Therefore, since the survival probability of a binary similar to Patroclus-Menoetius drops below 10% after 100 Myr (and below 1% after 200 Myr), there could not have been a large (approximately 700 Myr) delay between the dispersal of the protoplanetary gaseous disk and the predicted giant planet instability. [96] therefore concludes that the giant planet instability could not have caused the LHB.

Newer studies of the orbital dynamics of early Solar System formation models differ quantitatively from the original Nice Model presented in [127], but they generally retain the theme of a sudden giant planet instability causing outward radial migration. In particular, [89] discusses the orbital evolution of the giant planets in the presence of a gaseous protoplanetary disk, where the giant planets could have been captured into a resonant chain, then undergone an instability qualitatively similar to [127] once the gas dissipated. This approach has the advantage of using more physically-motivated starting conditions. It is worth noting that [89] still links the dynamical instability to the LHB, a concept which is not supported by the more modern work of [96].

Other studies investigate the possibility of more than 4 giant planets in the early Solar System. [10] and [95] both invoke the existence of a fifth giant planet in an initial resonant chain with Jupiter, Saturn, Uranus, and Neptune. This fifth planet is often ejected during this instability, as required by the current state of the Solar System. [10] finds that this 5 planet configuration is allowed, but not necessarily preferred over 4 giant planets, while [95] finds the 5 planet configurations preferable to the 4 planet configuration.

Furthermore, the existence of a fifth giant planet could help meet constraints set by the inner Solar System. If an ice giant scattered off of Jupiter shortly after the gaseous disk dissipated, it could cause Jupiter's semi-major axis to jump by a significant fraction of an au. This jump ensures that the inner Solar System is not overly disrupted during the early period of giant planetary evolution [99].

The exact details of an early Solar System giant planet instability have changed signifi-

cantly from the original Nice Model and are a matter of ongoing research. Regardless, many of the overall themes remain the same. [99] offers a comprehensive review of this subject. In short, the current theory is that shortly (< 100 Myr) after the gaseous protoplanetary disk dissipated, the giant planets interacted which each other and with a massive ($15 - 20\ M_E$) and numerous planetesimal disk from about $20 - 30$ au. This disk was strongly disrupted, with a small percentage ($\ll 10\%$) of members ending up on orbits in the outer Solar System. During this time, Neptune slowly migrated from about $20 - 25$ au to its present orbit at 30.1 au. A fifth ice giant may have been ejected from the Solar System during this time.

1.5 Modern TNO research

In this section, we will build on the discussion from Section 1.4. We will provide a summary of important observable features of TNO subpopulations, how they inform models of the orbital history of the outer Solar System, and what still remains to be studied. We will place a particular emphasis on areas that contain open questions which could be addressed by detecting more TNOs.

1.5.1 Hot classicals and the hot populations

Hot classicals are thought to be remnants of a massive protoplanetary disk that were embedded into the modern-day Kuiper Belt during the period of giant planet instability. Because of this, parameters such as the inclination, size, and color distribution of the hot population, especially compared to the Jupiter trojan population, serve as important constraints for any dynamical model of the early solar system.

The observational data that suggest the hot classicals and the Jupiter trojans are similar in size and color therefore create a strong argument in support of theories of early giant planet instability. (See Section 1.3.1 for more detail). As discussed in Section 1.4, the idea is that a massive belt of planetesimals interacted with the giant planets, helping to drive the planets towards their present orbits. This belt would have been dynamically excited by this interaction, driving up inclination and eccentricity. During the time of instability, the

giant planets would have lost any primordial trojan populations they may have had, and then would have gotten a new population of trojans from the massive disk of planetesimals [97, 55]. This analysis allows studies of the Jupiter trojans (a population at about 5.2 au) to help inform our understanding of the hot population of TNOs (a population beyond 30 au).

Direct observation of hot classicals also serve as critical model constraints. For example, [92] use the inclination distribution of the hot classical Kuiper Belt to suggest that the migration of Neptune must have been relatively slow, with characteristic timescales of > 10 Myr. Shorter migration timescales do not sufficiently raise the inclination and eccentricties of hot TNOs. Their model estimates that the initial mass of the disk of planetesimals would have been between 15 M_E and 20 M_E, which is consistent with the mass predicted from studies of the capture rate and current population of Jupiter trojans [99].

It is sometimes useful to discuss all of the hot populations of TNOs together, as opposed to specifically hot classicals. In this context, the hot populations of TNOs include hot classicals, detached objects, eTNOs, and sometimes scattering objects. Although the classical Kuiper Belt is separated into hot and cold populations, and we therefore often discuss the hot classicals as a separate category from the detached population, the two populations are closely related. In fact, it is not always clear that there is a distanced-based separation between the two populations. [50] separates these two subpopulations at $e = 0.24$ au. [51] places an edge to the hot classical Kuiper Belt around 50 au, but notes that low e objects beyond this distance can be considered to be an extension of the hot classical Kuiper Belt. Dynamically, hot classicals and detached objects should have been generated by similar mechanisms, although more distant detached objects may be subjected to unique effects (see Section 1.5.3).

The hot populations are interesting in part because there may still be large undiscovered objects with diameters greater than that of Pluto. Because the sizes of the hot populations are characteristically larger than the cold population, and because the hot population can be said to extend out to the inner Oort Cloud, there may be planet-sized objects in the hot population that remain to be detected simply because they are very faint. Specifically, [51]

postulates that there may still be Mars-sized objects hiding in the outer Solar System. They perform a quick estimate to motivate this prediction. Formation models predict $\mathcal{O}(1000)$ Pluto-sized objects were present in the early massive disk of planetesimals [88]. The efficiency with which members of the planetesimal disk are implanted into the hot populations (orbits beyond 30 au) is between 0.1% and 1% [99]. [51] therefore suppose that it is likely that there exist hot TNOs larger than Pluto but smaller than Earth, hiding somewhere in the outer Solar System, possibly in the inner Oort Cloud, scattering, detached, or eTNO populations.

1.5.2 Cold classicals

The cold classical Kuiper Belt is different from both the hot population and the Jupiter trojans. The color and size distribution of the cold classical Kuiper Belt [106, 42, 41, 115], the low inclination and eccentricity, the lower total population mass [42], and the high fraction of widely separated binaries [41, 102] all suggest that this population formed separately from the hot population (likely in situ) and was not strongly disrupted by any giant planet instability.

With regards to formation models, the un-excited, un-disrupted nature of the cold classical Kuiper Belt serves as an important constraint. Any model seeking to explain the history of the Solar System must not significantly raise the eccentricity or inclination of this population, nor disrupt the large fraction of widely-separated binary pairs. For example, in their simulations investigating the inclination distribution of the hot classical Kuiper Belt, [92] uses the method of [53, 54] to ensure that Neptune's eccentricity is never particularly large ($e < 0.1$), as required to avoid disrupting the cold classical Kuiper Belt.

Another constraining feature of cold classicals is the Kuiper Belt "kernel". The kernel is an apparent overdensity of cold classical KBOs at about 44 au [107]. [93] finds that this kernel could be explained if Neptune's semi-major axis suddenly jumped, rather than migrated smoothly outwards. This occurred in a simulation from [95] where Neptune has a close encounter with the simulated fifth ice giant, which was later ejected from the Solar System. [51] notes that a jumping Neptune would predict a population of cold classicals

with 37 au $< q <$ 39 au and that TNOs in the 2:1 resonance would outnumber those in the 3:2 resonance. Neither of these predictions are consistent with observations, although TNOs in the 2:1 resonance would be more difficult to detect than those in the 3:2 resonance due to their increased distance [51]. This introduces the question of observational bias, and is itself a reason for further investigation into TNOs. Further characterization of the Kuiper Belt kernel and the resonant populations, including ruling out observational bias, will help to test this hypothesis. At the furthest extreme, observational bias can be ruled out entirely once the inventory of TNOs down to an H magnitude of about 8 to 9 out to a distance of about 50 au is complete or nearly-complete. Once we have detected and characterized every (or nearly every) large TNO in the region of interest (in this case about 30 au to 50 au), the question of observational bias will be completely resolved. This would require a multi-year survey with high detection efficiency of the entire ecliptic down to an approximate r band magnitude of about 26. This is a large and challenging task; however, each additional smaller (and numerically incomplete) survey of the Kuiper Belt (e.g. [107, 8, 14]) helps to rule out the likelihood of observational bias accounting for the Kuiper Belt kernel, particularly if the surveys generate debiased TNO population models and have good detection efficiency out to about 50 au.

Finally, it is an open question why and where the cold classical Kuiper Belt ends. [51] discusses this question in detail. There appears to be a drop in the number density of cold classicals beyond the kernel, and either another drop in number density or else a hard edge at the 2:1 resonance at 48 au. Since the cold classicals formed in situ, there is no reason for there to be a hard edge corresponding to the 2:1 resonance. It is possible that the kernel represents the true edge of the cold classicals, and objects between 44 au and 48 au were dragged outwards via resonance drop-off [49]. [51] provide further hypotheses for truncating the edge of the cold classicals, either at 44 au or 48 au, as well as a discussion concerning the possible existence of cold classicals around 90 au.

In short, the exact nature of the distribution of cold classicals remains to be studied. Suggested over-densities, edges, and even very distant re-appearances of the cold classicals

can all be informed by improved detection of TNOs in these parameter spaces.

1.5.3 SDOs, eTNOs, and detached TNOs

Due to their extreme distance, the population of scattered (scattering) disk objects, detached TNOs, and extreme TNOs (sometimes used interchangably) are some of the most poorly characterized regions of trans-Neptunian space. These objects can provide constraints on a number of scientific questions.

If the Sun formed in a stellar cluster, the forces exerted by passing stars would have been more pronounced due to the greater stellar density in the natal cluster. In this case, these forces could have raised the perihelion distance of objects with a semi-major axis of only a few hundred au, such as Sedna [19, 20, 88]. For comparison, extra-solar forces today begin to act appreciably on TNOs around 2000 au, for which reason [50] adopts 2000 au as the beginning of the inner Oort Cloud and the end of the detached population. A more dense natal cluster (and thus a more compact proto-Oort Cloud) could therefore explain a unique population of eTNOs and could link them mechanistically to the current Oort cloud. For this reason, Sedna is sometimes used as an observational constraint on formation models of the inner Oort Cloud [e.g. 19, 20]. If we could increase the number of known sednoids and other eTNOs, they would provide observational constraints on the formation environment of the Sun [19] and the Sun's dynamic history in the Milky Way after leaving its formation environment [70].

[9] used observed clustering of the pericenter position of eTNOs to argue for the existence of a current $\mathcal{O}(10)$ M_E planet in the outer Solar System. This was based on the parameters of the 6 then-known TNOs with $q > 30$ which did not strongly interact with Neptune (and therefore were not part of the scattering population). However, [14] used TNO detections from the Dark Energy Survey to revisit this problem. They detected 16 TNOs with $q > 30$ au and $a > 150$ au, but found minimal evidence for any clustering in Ω, ω, or ϖ. It is therefore unclear whether or not any angular alignment of eTNOs exists. Further detection and characterization of eTNOs will help to answer this question.

In short, better knowledge of the dynamical and physical populations of TNOs would enable tests of additional and alternative hypotheses regarding the dynamical history of the Solar System. These hypotheses include smooth migration of Neptune [57, 93, 88], a stellar flyby [72], and rogue planetary embryos [48]. Improving our understanding of TNO size and orbital distributions, especially in the low mass and extremely distant regime where they are more poorly constrained, will be critical for our understanding of these and other hypotheses.

In the next section, we will discuss a class of detection methods called "digital tracking" algorithms, which are used to detect faint moving objects. Such methods are generally required to detect ultra-faint TNOs with apparent magnitudes $m > 26$ and will be crucial to the characterization of faint TNO populations over the next decade.

1.6 Digital tracking algorithms

"Digital Tracking" is a broad term to describe methods of detecting moving objects that are too faint to detect in a single image by testing many candidate trajectories. It is not known beforehand if an object actually exists along the candidate trajectory. Rather, a large grid of potential trajectories is searched and those trajectories that lead to a high signal-to-noise ratio are selected for further analysis. The grid is selected such that the searched trajectories correspond to the likely range of motion of the class of object of interest. "Shift-and-stack" is a specific type of digital tracking that directly coadds images along various grids. "Track-before-detect" and "synthetic tracking" are both other names for "digital tracking".

Broadly, digital tracking can be compared to static image coaddition techniques performed to detect faint stars and galaxies. However, coadding moving objects poses unique challenges compared to coadding stars. Because stars move very slowly compared to most survey cadences, coaddition of a stack of aligned images usually increases the limiting magnitude for stars compared to single images. Solar System objects, however, generally move at on-sky velocities of > 1 arcsec hr^{-1}, due to both the proper motion of the objects and the reflex motion caused by the Earth's orbit. This means that traditional image coaddition typically does not increase the number of detectable moving objects. Known moving objects

may be tracked and aligned along their orbits to improve the quality of the detection, but to use image coaddition to detect new objects with unknown orbits, digital tracking is required.

The first successful attempt to use digital tracking on KBOs dates back to 2004. [15] used a technique known as "shift-and-stack". This work used the Hubble Space Telescope (HST) to search a 0.02 square degree region of the sky for TNOs. They detected three new objects with F606W magnitudes between 26.95 and 28.38. Each of these objects was consistent with the cold classical Kuiper Belt. The relatively small number of objects detected (three new objects and one previously-known object) was in stark contrast to the expected 85 detection based on an extrapolation of the bright-end power law and suggested a break in the TNO size distribution, which was characterized in detail 10 years later in [42].

New and upcoming approaches to survey astronomy provide exciting opportunities for the study of the populations of the outer Solar System. For example, the upcoming Legacy Survey of Space and Time [LSST; lsst.org; 65] expects to survey over 18,000 square degrees of the sky 825 times over a period of 10 years, generating about 20 TB of data every 24 hours.

LSST plans to detect Solar System objects from individual images, with a single-visit limiting magnitude in the r band of 24.7, and link these detections to measure orbits. Current projections ([80]) show that LSST is expected to detect about 40,000 TNOs, which is by itself a large increase over the currently-known 4159 TNOs. However, if we could coadd the images to increase the signal-to-noise ratio (SNR) of the Solar System detections, then we could recover significantly more objects. Following the formula $\Delta m = (5/2) \log \sqrt{N}$, coadding just three months of LSST data would increase the limiting magnitude in the r band from 24.7 to 26.1. This increase in depth means LSST would detect approximately 8 times more TNOs compared to a single image, assuming the single power-law r band KBO distribution of [43]. Instead of 40,000 TNOs, we could detect up to 320,000 TNOs. If we could coadd a year of LSST data, this increases to over 520,000 new TNOs detected (given our simplified assumptions). None of this requires any more data than LSST will already acquire.

However, new approaches are required in order to use digital tracking algorithms with LSST data, given the wide survey footprint of LSST. Unlike the small field of fiew of [15], image coaddition on moving objects must be able to span thousands of square degrees to take full advantage of the data provided by LSST. Furthermore, new algorithms must be as robust to outliers as possible, minimizing the number of false positives that humans have to vet. One such algorithm is called `KBMOD`.

The Kernel-Based Moving Object Detection [`KBMOD`; 136] algorithm takes a time series of images of the same RA and Dec, uses a digital tracking approach to account for the potential motion of objects on an image, and then sums the coadded likelihood of a candidate trajectory, thereby increasing the SNR of objects with that candidate trajectory. To sample all possible orbital parameters requires searching billions of candidate trajectories even within the footprint of a single charge-coupled device (CCD). Consequently, current implementations of digital tracking have generally been restricted to narrow-field surveys [15]. `KBMOD` addresses this by using GPU-accelerated computing to search over a wide range of trajectories for a stack of CCDs in $\mathcal{O}(10)$ minutes, increasing the possible field of view of surveys to which digital tracking algorithms can be applied.

`KBMOD` generates images of likelihood (Ψ_i) and variance (Φ_i) from a series of CCD images as described in [136]. These images are described by Equations 19 and 20 of [136], which state

$$\Psi(y) = \sum_j \frac{1}{\sigma_j^2} n(x_j) T(x_j - y) \tag{1.8}$$

$$\Phi(y) = \sum_j \frac{1}{\sigma_j^2} T(x_j - y)^2, \tag{1.9}$$

where σ_j is the variance at pixel j, $n(x_j)$ is the number of counts in pixel j at position x_j, and $T(x_j - y)$ is the PSF centered at the true position y of a given point source. Assuming a Gaussian likelihood function, a stack of Ψ_i and Φ_i images can then be shifted along a potential asteroid trajectory and summed in order to get the coadded likelihood of a detection. Thus,

we get

$$\Psi_{\text{coadd}} = \sum_i \Psi_i \tag{1.10}$$

$$\Phi_{\text{coadd}} = \sum_i \Phi_i. \tag{1.11}$$

See [103] for the optimal approach for source detection with Poisson noise. We define a SNR ν for a detection such that $\nu_{\text{coadd}} = \Psi_{\text{coadd}}/\sqrt{\Phi_{\text{coadd}}}$. In this ν image, generated for each given angle and velocity vector, any points above some threshold m can be considered to be m-sigma detections of a moving source. For a single trajectory, we can define the summed likelihood that there is a source along a given trajectory as

$$\sum LH = \nu_{\text{coadd}}^{\text{trajectory}}. \tag{1.12}$$

The interested reader is directed to [136] for more detail.

The large number ($\gg 10^9$) of potential TNO trajectories means that these Ψ_i and Φ_i images must be searched many times over. For this reason, `KBMOD` uses massively-parallel GPU computing for the core computations. The current software allows a user to search over 10^{10} potential moving object trajectories in a stack of 10-15 4K x 4K images in under a minute using a consumer-grade GPU (e.g., Nvidia 1080 Ti) [136]. Our pointer-arithmetic approach means that we never actually shift and stack images. Rather, we merely sum the previously-calculated likelihoods, utilizing thousands of concurrent GPU threads to keep the computation feasible on consumer-grade hardware.

1.7 Linking and orbit fitting

The generalized problem of taking a set of single observations on the sky and converting them to a set of unique objects in space with characterized orbital elements falls into the category of linking and orbit fitting. Linking and orbit fitting are two separate processes that each pose their own unique set of challenges and constraints. They are critical to any Solar System survey because without these steps it is impossible to map detections to physical locations in space or to say anything with precision about their future motion.

1.7.1 Linking

Linking is the process of connecting candidate tracklets into a set of observations that will be used as inputs for orbit fitting. A tracklet is simply a set of astrometric observations, often pairs or triplets, that likely belong to a single moving object. These observations must be time-separated enough that they are clearly not static sources, but also not too time-separated such that it becomes unclear if they belong to a single object [62].

One method of linking involves moving to heliocentric space [62]. This method takes tracklets, moves into a heliocentric reference frame, propagates to a common time, and identifies clusters corresponding to linked objects. To apply this method, one must select a distance and a radial velocity (or inverse distance γ and radial velocity scaled by inverse distance $\dot{\gamma}$ in the framework of [16] and [62]) likely to correspond to an orbital class of interest. One advantage of this linking algorithm is the tractable $\mathcal{O}(N \log N)$ scaling.

As has been mentioned, the observations used to generate tracklets imposes a strong cadence requirement on surveys. The observations must be taken far enough apart that objects of interest have moved but not moved too much. This time separation varies somewhat depending on the class of moving object and becomes an optimization problem for large survey operations.

One method of linking, called the "Tracklet-less Heliocentric Orbit Recovery" (`THOR`) algorithm, helps to alleviate this problem [86]. As the name implies, `THOR` is a cadence-independent linking algorithm. Unlike [62], which requires tracklets that are transformed into heliocentric space, `THOR` can operate solely on individual detections. This is a large advantage of `THOR` over most other linking algorithms. To do this, `THOR` moves into the corotating frame of reference of a test orbit and identifies clusters in this space. This requires searching over many possible test orbits, the exact number of which depends on the orbital class of interest. Generally, orbital classes close to the Sun, like MBAs and NEOs, require a finer resolution of test orbits compared to more distant classes like TNOs.

The method of linking presented in [13] is specifically designed for TNOs. Like [62], it

uses the orbit-fitting framework of [16], which is discussed in more detail in Section 1.7.2. This method moves single detections into a heliocentric space at a common reference time by binning in inverse distance (γ). [13] then identifies possible pairs, grows the pairs to triplets, grows triplets into n-lets, and finally fits full orbits following the method of [16]. This linking method is applied to the candidates detected in Chapter 4.

1.7.2 Orbit fitting

Orbit fitting is the process of taking a set of (usually) linked detections and parameterizing both the best-fit orbital elements as well as the goodness of that fit. This can be a challenging problem for astronomers because the observed quantities are typically only right ascension and declination, which leaves distance and radial velocity without measured values. However, due to orbital geometry, the state vector and thus the orbit can still be determined.

One of the oldest orbit fitting methods is called Gauss' Method. This method was developed by Carl Friedrich Gauss in 1801 in response to the intial discovery of a then-unidentified moving object by Giuseppe Piazzi [39]. This object is now known as the dwarf planet Ceres. A notable feature of Gauss' Method is that it requires only three astrometric observations of an object and requires only that the orbit be a conic section. Using this method, Gauss's observed orbital parameters for Ceres were remarkably accurate and precise. For example, Gauss found Ceres had an eccentricity of $e = 0.083$ and an inclination of $i = 10.62°$, while the JPL Small Body Database reports present values of $e = 0.079$ and $i = 10.59°$. See e.g. [30] for a presentation of Gauss' Method as applied to geocentric orbits.

A number of factors contribute to inaccuracy in Gauss' Method. For example, the method makes numerical approximations that can be eliminated by numerically solving Kepler's Equation. Even by improving the numerical precision of the problem, Gauss' Method inherently relies on a two-body orbital functional form. This means that many perturbing effects are ignored, including Solar radiation pressure, perturbations from bodies other than the central body of the orbit, and even non-uniformity in the central body and atmospheric drag (in the case of geocentric orbits) [30]. More exact methods of orbit fitting require using

numerical integration techniques such as N-body simulations to account for these perturbations.

One method of orbit fitting used specifically for TNOs is that of [16]. Because TNOs are quite far from the Sun and move much more slowly than small bodies in the inner Solar System, over observational time baselines of months or even a couple of years, their motion can be considered largely inertial and gravitational effects can be treated as perturbations to their motion. In this formalism, shown in Equation 5 of [16], they define an α, β, and γ parameter space where $\alpha \equiv x_0/z_0$, $\beta \equiv y_0/z_0$, and $\gamma \equiv 1/z_0$. These values have associated time derivatives of $\dot{\alpha} \equiv \dot{x_0}/z_0$, $\dot{\beta} \equiv \dot{y_0}/z_0$, and $\dot{\gamma} \equiv \dot{z_0}/z_0$. Here, x_0, y_0, and z_0 are cartesian coordinates from the orbit, which is parameterized by Equation 2 of [16] in the form $\boldsymbol{x}(t) = \boldsymbol{x_0} + \dot{\boldsymbol{x_0}}t + \boldsymbol{g}(t)$. $\boldsymbol{g}(t)$ is a gravitational perturbation term given by Equations 3 and 4 of [16]. In the method of [16], $\boldsymbol{g}(t)$ includes gravitational perturbations from the Sun and the giant planets. Note that this parameterization is similar to scaling the state vector by distance, with the main difference being that γ is directly inverse distance. This parameterization makes conceptual sense for TNOs because the reflex motion is the dominant driver of on-sky motion. In turn, this means that distance is the primary quantity affecting the apparent motion in right ascension and declination.

1.8 Chapter summaries

Chapter 2 discusses various numerical methods and filtering improvements that have been created to use with `KBMOD`. This includes changes and additional options for stamp generation, a statistically-robust σ_G-based clipped median lightcurve filter, in-line GPU filtering, and a convolutional neural network stamp filter.

Chapter 3 discusses the first application of `KBMOD` to difference images. The data used for this analysis was the DECam NEO Survey data, P.I. Lori Allen. In this analysis, we detect 24 known KBOs as well as 75 objects we could not link to any known objects. Due to the short time baselines, we are only able to place constraints on the barycentric distance, i, and Ω. We compare these objects to published KBO inclination and magnitude distributions

and find that our detections are consistent.

Chapter 4 discusses the application of KBMOD to the DEEP survey. This survey was specifically designed for digital tracking methods. The repeat visits to a single field over multiple years allow us to fit detailed orbits to detected objects. The injection of a large and dynamically varied population of synthetic objects (or fakes) into this data set allows for detailed analysis of the performance of KBMOD. In this analysis, we detect and characterize 99 TNOs that we could not link to any known TNOs. We measure and report H magnitudes and osculating elements for these objects and perform dynamical classification.

In Chapter 5 we discuss the implications of the work presented herein as well as future work that remains to be done with KBMOD.

Chapter 2

NUMERICAL METHODS, ALGORITHMS, AND FILTERING: IMPROVEMENTS TO KBMOD

In this chapter, we present a number of new software tools and enhancements to existing tools that have been added to KBMOD. These improvements enable KBMOD to run more effectively at scale, more robustly filter out false positives, and apply more generally to differing data sets.

Several of the tools discussed herein were developed partially in response to the processing of the DECam NEO data set discussed in Chapter 3. These data presented unique filtering challenges compared to [136] due to the increased number of potentially-valid trajectories, the short intra-night cadence, and image differencing artifacts. In [136], detected sources appearing in the same position in 2 or more images, pixels with counts above 120 counts, and other mask flags set by the DECam community pipeline or the LSST software stack were all masked. In the current data set, we use image differencing to subtract static sources. This enables us to decrease the masked area of the image, only masking sources flagged as detected if they appear in 10 or more images, as opposed to the more restrictive masking from [136]. However, despite reducing the number of detected individual sources on the image by a factor of about 18 (see Figure 3.3), leaving most of the image unmasked (coupled with difference imaging artifacts) increases the number of trajectories with $\sum LH > 10$ by a factor of 10. This problem is worsened by the intranight cadence. The average time between images within a single night is about 5 minutes. This means that for a characteristic trajectory with a velocity of 100 px/day, objects will move by less than 1 pixel between images. Conversely, this also means that if a static source appears along the potential trajectory, flux from this object will most likely be present in at least 5 trajectory data points,

introducing repeated outliers into the trajectory. There is a trade-off between decreased image masking (increasing area) and increased false positives. False positives, however, can often be filtered out algorithmically. We therefore focused heavily on developing improved filtering methods to apply to these data.

Although some of these tools were developed in response to the processing challenges of the DECam NEO data, they are generally applicable to other data sets. For example, the quantile limits of σ_G filtering discussed in Section 2.1 are variable and user-programmable. Similarly, although we used median coadds for the central moment filter, as discussed in Section 2.3, this option is also user-programmable, and can be reverted to a simple sum as in [136]. Tooling is also in-place to develop new options for coadded stamp generation.

Other tools, such as the joint-fit discussed in Section 2.6, were developed to serve as characterization tools to complement `KBMOD` (which is itself primarly a detection tool). The joint-fit was applied to the data discussed in Chapter 4, and significantly improved the astrometric and photometric accuracy and precision for candidate detections. Still other algorithms, such as the barycentric reprojection discussed in Section 2.5, are primarily proof-of-concept analysis waiting to be refined and applied to future data sets like LSST.

2.1 σ_G filtering

In order to deal with the increased number of high likelihood trajectories in the DECam NEO Survey data (i.e. 10^7 with $\sum LH > 10$) and to decrease the number of false positives that require human review, we developed faster, more-effective filtering. First, we altered how the GPU and C++ code handed off data to the Python-based filtering, leading to a speed increase of up to 300%. This was a relatively small code rewrite which led to significant optimization. Second, we replaced the Kalman filter used in [136] with a more statistically-robust quantile-based filtering method. We describe this new filtering method below.

With a traditional quantile-based filter, the filter rejects data points that are greater than $n\sigma$ from the central value of the distribution, where σ is a measure of the spread of the distribution. In the case of a Gaussian distribution, σ might be estimated by computing

the standard deviation of the data and the central value estimated by computing the mean of the data. If we take $n = 1$, then this simple filter would reject any data points that are greater than 1σ from the mean.

In the presence of significant outliers, the mean and standard deviation become biased estimators for the central value and the spread of the underlying Gaussian distribution. Following the approach of [64], we adopt a robust estimator for the central value and the true standard deviation of a Gaussian distribution with outliers. Consider the cumulative distribution function (CDF) of a Gaussian distribution

$$f(x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x-\mu}{\sigma_G\sqrt{2}}\right)\right], \tag{2.1}$$

where μ is the mean, σ_G is the standard deviation of the Gaussian, and erf is the error function. The inverse, then, is given by

$$x = \mu + \sigma_G\sqrt{2}\ \text{erf}^{-1}\left[2f(x) - 1\right]. \tag{2.2}$$

By sampling the Gaussian distribution at two quantiles $f(x_i)$ and $f(x_j)$, we can estimate σ_G. To do this, we take the difference of the inverted CDF

$$x_j - x_i = \sigma_G\sqrt{2}\left(\text{erf}^{-1}\left[2f(x_j) - 1\right] - \text{erf}^{-1}\left[2f(x_i) - 1\right]\right) \tag{2.3}$$

$$\implies \sigma_G = \frac{1}{\text{erf}^{-1}\left[2f(x_j) - 1\right] - \text{erf}^{-1}\left[2f(x_i) - 1\right]}\left(x_j - x_i\right) \tag{2.4}$$

$$\implies \sigma_G = C\left[x_j - x_i\right]. \tag{2.5}$$

Here, C is a coefficient dependent only on the choice of quantiles. x_j and x_i are estimated by selecting values from the lightcurve closest to the desired quantile. The choice of upper and lower quantiles is user-determinable. Here, we estimate σ_G using data from the 25th to 75th percentiles, for a coefficient of $C_{25,75} \approx 0.7413$. Then, we can estimate x_{25} and x_{75} from the data by selecting the 25th and 75th percentile values respectively from the data. We can then estimate the standard deviation of the underlying Gaussian distribution with

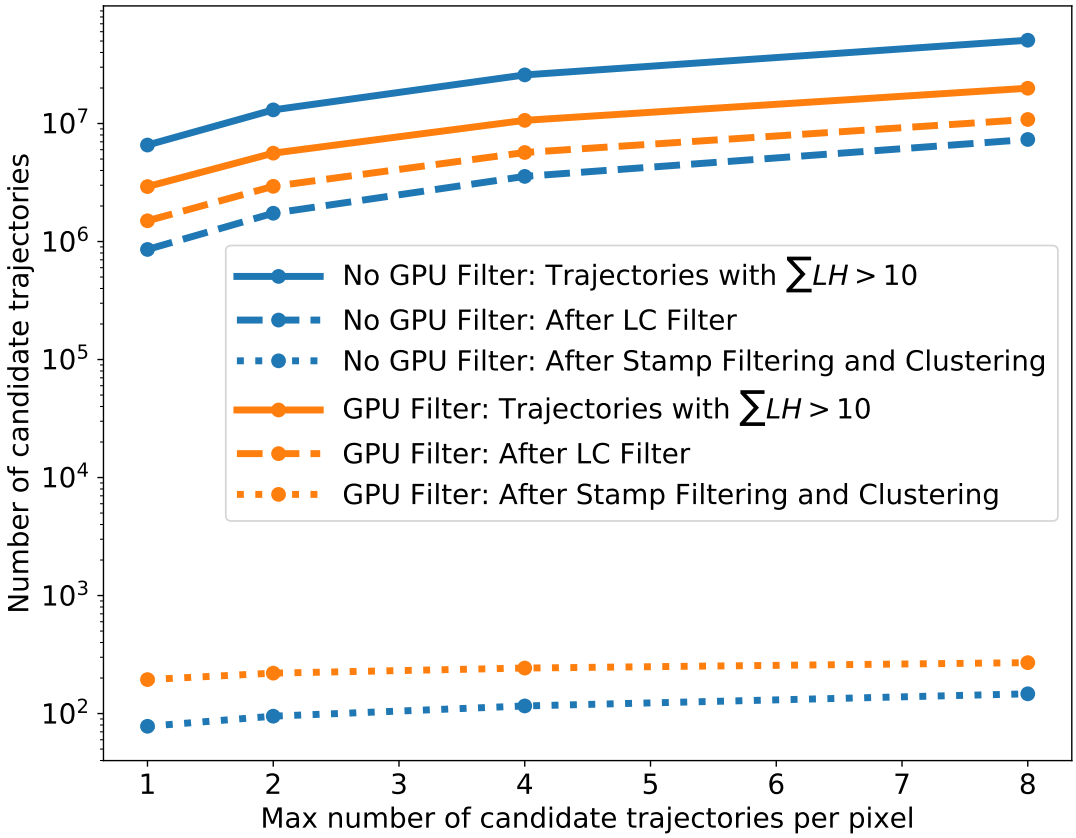

Number of candidate trajectories
10^7
10^6
10^5
10^4
10^3
10^2
No GPU Filter: Trajectories with $\sum LH > 10$
No GPU Filter: After LC Filter
No GPU Filter: After Stamp Filtering and Clustering
GPU Filter: Trajectories with $\sum LH > 10$
GPU Filter: After LC Filter
GPU Filter: After Stamp Filtering and Clustering
1
2
3
4
5
6
7
8
Max number of candidate trajectories per pixel

Figure 2.1: Number of candidate trajectories at various stages of processing for variable numbers of results per pixel. The solid line shows the number of candidate trajectories with $\sum LH > 10$ returned from the GPU for subsequent filtering. The dashed and dotted lines show the number of candidate trajectories passing CPU σ_G filtering (dashed) and central moment stamp filtering and clustering (dotted). GPU filtering decreases the total number of candidate trajectories with $\sum LH > 10$, but increases the number of candidate trajectories that pass subsequent lightcurve filtering and stamp filtering and clustering. Because GPU memory constraints limit the number of candidate trajectories per starting pixel that can be saved for subsequent analysis, using a GPU filter means that the results that are passed out of the GPU are more likely to be potentially-valid. These results are then processed with the CNN filter and subject to human review. These data come from repeated reprocessings of pointing group 023, CCD 35, which is part of the analysis of the DECam NEO data set presented in Chapter 3. This pointing group is not part of the search sample. Here, it is used primarily to compare the relative number of candidate trajectories at various stages of processing.

$\sigma \approx 0.7413\,(x_{75} - x_{25})$. Given a robust estimator of the spread of the distribution (i.e. σ_G), we apply a filter that rejects any points that are not within $\pm n\sigma_G$ (e.g. $2\sigma_G$) of the median of the data.

We can apply this method to the likelihood and/or flux values of each trajectory. In the data presented in Chapter 3, we apply this filter to both the likelihood and the flux. We then recompute $\sum LH$ for the trajectory values that pass the filter and reject the trajectory if the recomputed likelihood ($\sum LH'$) is less than 10. In practice, this filtering method successfully rejects of order 10^6 erroneous candidate trajectories in approximately 60s using 30 central processing unit (CPU) cores.

2.2 In-line GPU filtering

As commercially-available GPUs become more advanced, it is increasingly possible to get a single GPU with more onboard RAM. For example, the data processed in Chapter 3 was processed in part with an Nvidia 2080 Ti, with 11GB of onboard RAM. In comparison, by the time we finished processing the data described in Chapter 4, we were able to use Nvidia A40 GPUs, each with 48 GB of RAM. This increase is likely to continue as technology progresses, however it is likely that GPU RAM will remain a constant constraint for `KBMOD`. As mentioned, `KBMOD` searches over a stack of aligned images, which are 2K x 4K in the data from Chapter 3 and Chapter 4, or two 2K x 4K images stitched together (for a total image of 4K x 4K) in [136]. If there were no constraint at all on GPU RAM, in principle we could have run a single search on the entire 60 2K x 4K CCD focal plane of DECam, instead of searching each CCD individually. The advantage to doing so would be to eliminate the constraint that a moving object must largely remain within the footprint of a single CCD in order to be detected. This effect becomes increasingly important for sparse lightcurves and long baselines, as will be the case when running `KBMOD` on LSST data. Therefore, the best-case scenario will likely always be to fit as much imaging data as possible onto the GPU running the processing.

This problem extends not just to image size, but also to the number of candidate results

passed from the GPU to subsequent CPU filtering. In [136], KBMOD passed the 4 trajectories per pixel with the highest $\sum LH$ from the GPU to the CPU. Other trajectories with the same starting pixel were discarded. Because KBMOD searches of order 10^{12} trajectories for a 2K x 4K image, it is computationally infeasible to keep the results of all evaluated trajectories in GPU RAM. The disadvantage of this approach is that lower-likelihood trajectories may get removed from the search even if they are valid trajectories of true objects. With reduced masking, there are many erroneous candidate trajectories with high likelihood. This means that removing the masks may have increased the probability of discarding valid trajectories.

Applying a variant of the σ_G filter in the GPU while the search is running, instead of in post-processing, increases the number of potentially-valid trajectories returned to the CPU by KBMOD. In-line GPU filtering helps mitigate the problem of limited results per pixel by applying the filtering method to compute $\sum LH'$ before the trajectory is passed back to the CPU. This in-line GPU filter means that if a trajectory has a high $\sum LH$ only due to an outlier in the data, that trajectory is unlikely to supplant another valid trajectory when GPU results are passed back to the CPU. As shown in Figure 2.1, after CPU σ_G filtering, stamp filtering, and clustering, processing with the GPU filter returns about 2 times more results per search than with no GPU filter. We also increased the number of returned results per pixel from 4 to 8. This means that we were able to process about 4 times as many results per pixel compared to [136]. In principle, we would have liked to increase this number as high as possible, or even to have removed this constraint entirely, but hardware limitations at the time of processing required a limit. Qualitatively, the slope of the orange curves (using GPU filtering) in Figure 2.1 is lower between 4 results and 8 results than it is between one result and two results or two results and 4 results. As one might expect, this suggests that there may be diminishing effects to further increasing this number. The in-line GPU filtering uses a single GPU and is about 10% faster than comparable CPU filtering using 30 CPU cores, although in practice both GPU and CPU filtering is run. Figure 2.1 demonstrates how the in-line GPU filter returns more potentially-valid trajectories for a given number of trajectories per pixel.

2.3 Median stamp coadd generation

As shown in Figure 3.2, saturated cores and small image misalignments leave a number of artifacts in the difference image that also have to be accounted for in the filtering process. As in [136], we computed the central moments of postage stamps for candidate trajectories. The central moments of an image are a computationally inexpensive way to characterize some of the properties of the data in that image. For example, the x moment is a measure of how centered the data in the image is with respect to the x axis. Stamps were rejected if they did not have central moments that were consistent with a Gaussian. In this data, we required that the x, y, xy, xx, and yy moments be strictly less than 0.5, 0.5, 1.5, 36.5, and 36.5 respectively. These values were chosen empirically based on the central moments of known KBOs. We generated coadded stamps by computing the median pixel value for each pixel along the trajectory. This mitigates the effect of image differencing artifacts, improving the performance of the central moment filter.

2.4 CNN filtering

The basic idea of an artificial neural network (ANN) is to assign a probability that a given input belongs to a given class. This is done by inputting a set of labelled training data. The network, a connection of different nodes each with an associated weight value, then takes the input value, breaks it down into some numerical components, scales them by the weight at each node, and assigns a probability that the input belongs to a class. The network then uses the label and back-propagates the associated error to each node, adjusting each associated weight. This allows the ANN to "learn" the weights from the input data. This approach allows the ANN to efficiently perform convex optimization over a very high-dimensional parameter space (such as the parameter space of a hand-written word, which varies from person to person).

The specifics of neural networks vary widely from network to network. There are infinitely many different possible configurations of nodes, forming different networks which perform

differently on different data. Additionally, there are various ways to penalize error and a multitude of hyperparameters that affect the training of a specific network. A full treatment of neural networks is beyond the scope of this section, but we will give some historical background and describe the particular neural network we train for the classification of `KBMOD` candidates.

The origins of ANNs date back as far as the 1940s. [84] (originally printed in 1943) and [59] (published in 1949) studied organic neural networks, paving the way for their work to be applied to the new and exciting device that would come to be called the "computer". By the 1980s, computers had become fast enough that they began to have practical scientific applications (e.g. [110] applied neural networks to protein modelling). By the early 2000s, GPU applications of ANNs began to be developed, leading to further speedups (e.g. [26]). Today, neural networks can be used in a wide range of applications, from text-to-speech, to protein struture modelling [68, 128], to stock market predictions, to the game of Go [118].

One specific kind of ANN is called a Convolutional Neural Network (CNN) [44, 45, 78]. CNNs take tensor inputs, such as RGB images where each color (red, green, and blue) forms a layer of image data. As the name implies, CNNs have one or more convolutional layers, where the value of the specific convolutional kernel is itself a learned parameter. This is in contrast to, for example, astronomical imaging which might convolve an image with a PSF of a fixed functional form. CNNs are able to "learn" complicated features in images, identifying animals, stop signs, and, as we will show, TNOs.

Despite the layers of filtering that have been described so far, `KBMOD` can still return tens or even thousands of candidate trajectories depending on the $\sum LH$ level. In the DECam NEO data set described in Chapter 3, after all filtering described so far there are still characteristically hundreds to about a thousand candidates per CCD with $\sum LH > 10$ and tens to hundreds of candidates per CCD with $\sum LH > 15$. Each CCD corresponds to a `KBMOD` search. This is shown in more detail in Figure 2.2. Furthermore, in the DEEP data discussed in Chapter 4, which uses a lower threshold of $\sum LH > 7$ and detects fainter objects, there are tens of thousands of total candidates that would require human vetting

(see Figure 2.3). Another filtering method is required if a single person is to be able to vet the candidates results from `KBMOD`.

To further reduce the number of false positives, we filter using a CNN. We built a Residual Network with 50 layers (ResNet50[1]) [58]. Residual networks are a type of CNN that add "shortcut connections" into the network architecture, which help to train deeper networks. Training a CNN requires a large amount of representative data. In this case, we needed a large ($> 10^4$) labelled set of 21 x 21 stamps containing approximately equal numbers of false positives and true positives. To generate false positives, we ran an untargeted search (with similar grid spacing as described in 3.3.1) with a coadded likelihood limit of $\sum LH > 10$ along trajectories unlikely to correspond to real objects (approximately 90° from the direction of the ecliptic)[2]. We ran a total of 53 searches with data from 34 unique pointing groups. These pointing groups were not constrained to the search sample. These searches yielded 113,549 21 x 21 false positive postage stamps. Because KBOs are relatively rare, we could not use real recovered objects to generate the thousands of true positives needed to train the CNN. To circumvent this limitation, we generated 44,950 simulated true positives. To make these stamps, we retrieved 25 21 x 21 postage stamps from a CCD along a semi-random trajectory. Next, we drew a random brightness from an exponential distribution (with dimmer objects being the most likely). Using this brightness, we added a Gaussian to each background stamp with a random standard deviation (1 to 2.1 pixels), a random central offset (< 2 pixels), and a random linear offset (< 2 pixels over the image time baseline). To train the CNN, we cut the false positive stamps and simulated true stamps down to 40,000 randomly-selected coadded stamps each. We used 70% of the data for training, 20% for validation, and the

[1]See `https://github.com/priya-dwivedi/Deep-Learning/blob/master/resnet_keras/Residual_Networks_yourself.ipynb`

[2]Another option would be to run a time-shuffled search, where image times are randomly switched among the search images. There is a small possibility that this randomization will still result in some signal from real objects when searching prograde trajectories. This is more likely in cases such as the data set discussed in Chapter 4, where images do not move very far between images and thus at any given image time there is more likely to be some overlap between the PSF and the source. However, in general, this time shuffling is likely a good approach and should be considered for future work, especially in cases where reflex motion is not the dominant source of on-sky motion (e.g. when searching for MBAs).

remaining 10% for testing. After 20 epochs, the training set accuracy was about 99%, while the validation set accuracy was about 96%. After training, the test set accuracy was also about 96%.

This CNN returns a predicted probability that a coadded postage stamp contains a simulated object. Because the stamps of simulated objects differ from the stamps of real objects, this probability is not a perfect representation of the likelihood that a coadded stamp contains a real object. However, it creates a user-programmable threshold that can be used to reduce false positives enough that the remaining candidate trajectories can be analyzed by-eye. In the work presented in Chapter 3, we reject any stamps with a CNN probability of true less than 75%. In the work presented in Chapter 4, we lower the threshold to 50%. We lower the threshold in part because the network was trained on image backgrounds and using the cadence from the data in Chapter 3. As shown in Figure 2.2, when reviewing only objects with a $\sum LH > 15$ and using this CNN filter, there are generally fewer than 10 candidate trajectories per CCD that require human by-eye confirmation or rejection. Because a human can review about 1 image per second, $\mathcal{O}(10^1)$ is an acceptable number of candidates per CCD[3].

We further analyze this CNN using the DEEP data set described in Chapter 4. We inject a large and dynamically-varied synthetic (or fake) TNO population into the DEEP data set. We call these synthetic implanted objects "fakes"[4]. We made no attempt to replicate the actual physical population of TNOs when selecting our fakes. Rather, we chose these fakes to cover a wide brightness range and dynamical range such that both known and potentially unknown classes of TNOs will be included in this population. (See Section 4.3.3 for more detail). This allows us to perform a more robust characterization of the CNN efficiency both

[3]For clarity, we emphasize that this target of $\mathcal{O}(10^1)$ is completely separate from the maximum number of results per pixel discussed in Section 2.2. Here, the target of $\mathcal{O}(10^1)$ final candidate trajectories per CCD exists to try to minimize the time it takes for a human to vet all candidate trajectories.

[4]We stress that "fakes" are different from "false positives". A false positive is a spurious detection that should be filtered out and/or rejected by human vetting. A fake should be detected and accepted by human vetting, as it corresponds to a simulated object implanted in the data set. A fake does not, however, correspond to a real object in space.

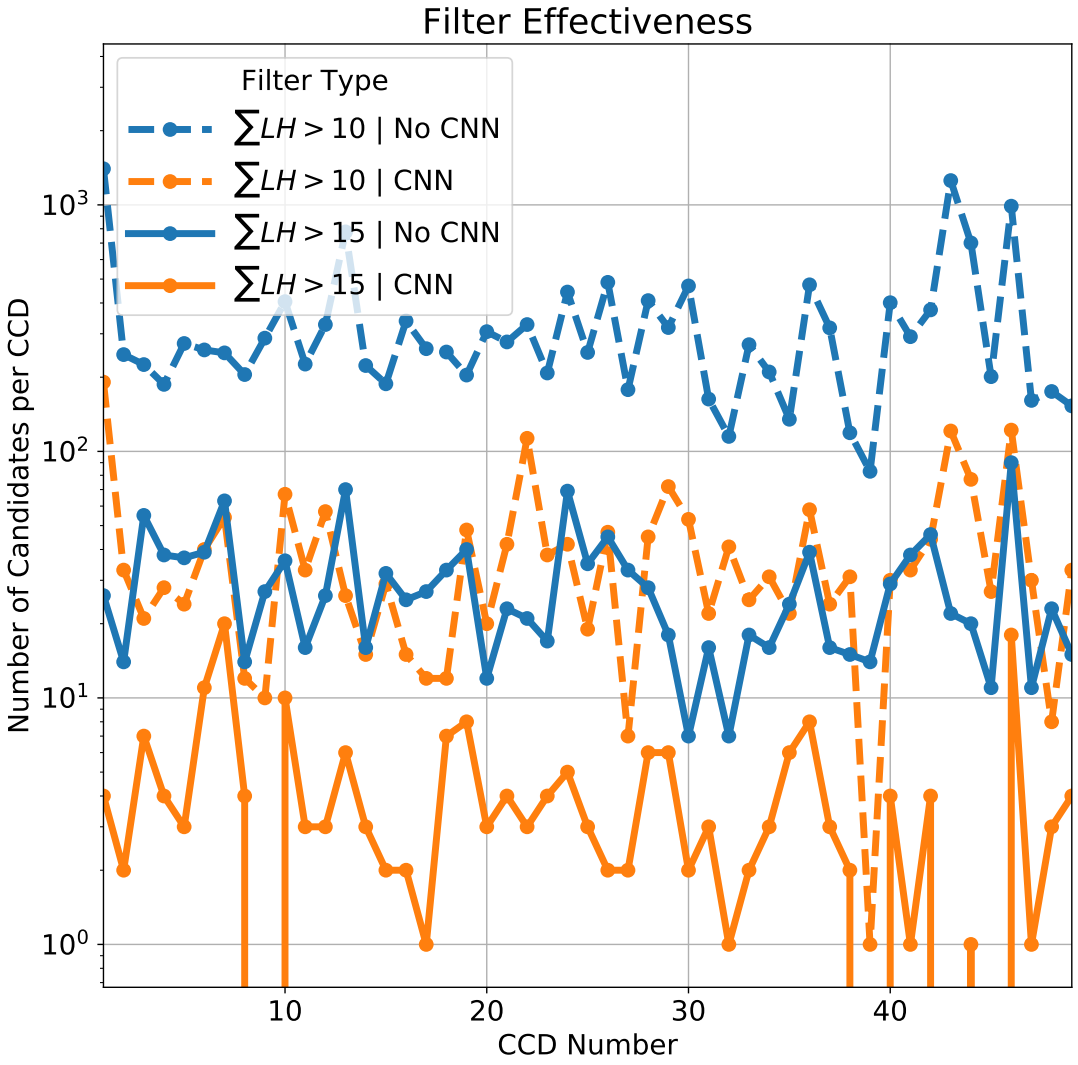

Filter Effectiveness
Filter Type
$\sum LH > 10$ | No CNN
$\sum LH > 10$ | CNN
$\sum LH > 15$ | No CNN
$\sum LH > 15$ | CNN
Number of Candidates per CCD
10^3
10^2
10^1
10^0
10
20
30
40
CCD Number

Figure 2.2: Number of results per CCD image stack from pointing group 190 requiring by-eye confirmation or rejection (hereafter "candidate trajectories") for likelihood limits of 10 (dashed line) and 15 (solid line), with (orange) and without (blue) the RESNET 50 CNN filter described in Section 2.4. These results are from pointing group 190, one of the search sample pointing groups. Here, the CNN was set to filter out any candidate trajectories with a probability of true that was less than 75%. When using a LH limit of 15 and the CNN, the number of candidate trajectories per CCD was reduced to 11 or less. 11 candidates per CCD can be reviewed by a human in about as many seconds and is therefore an acceptable number of trajectories for a human to review.

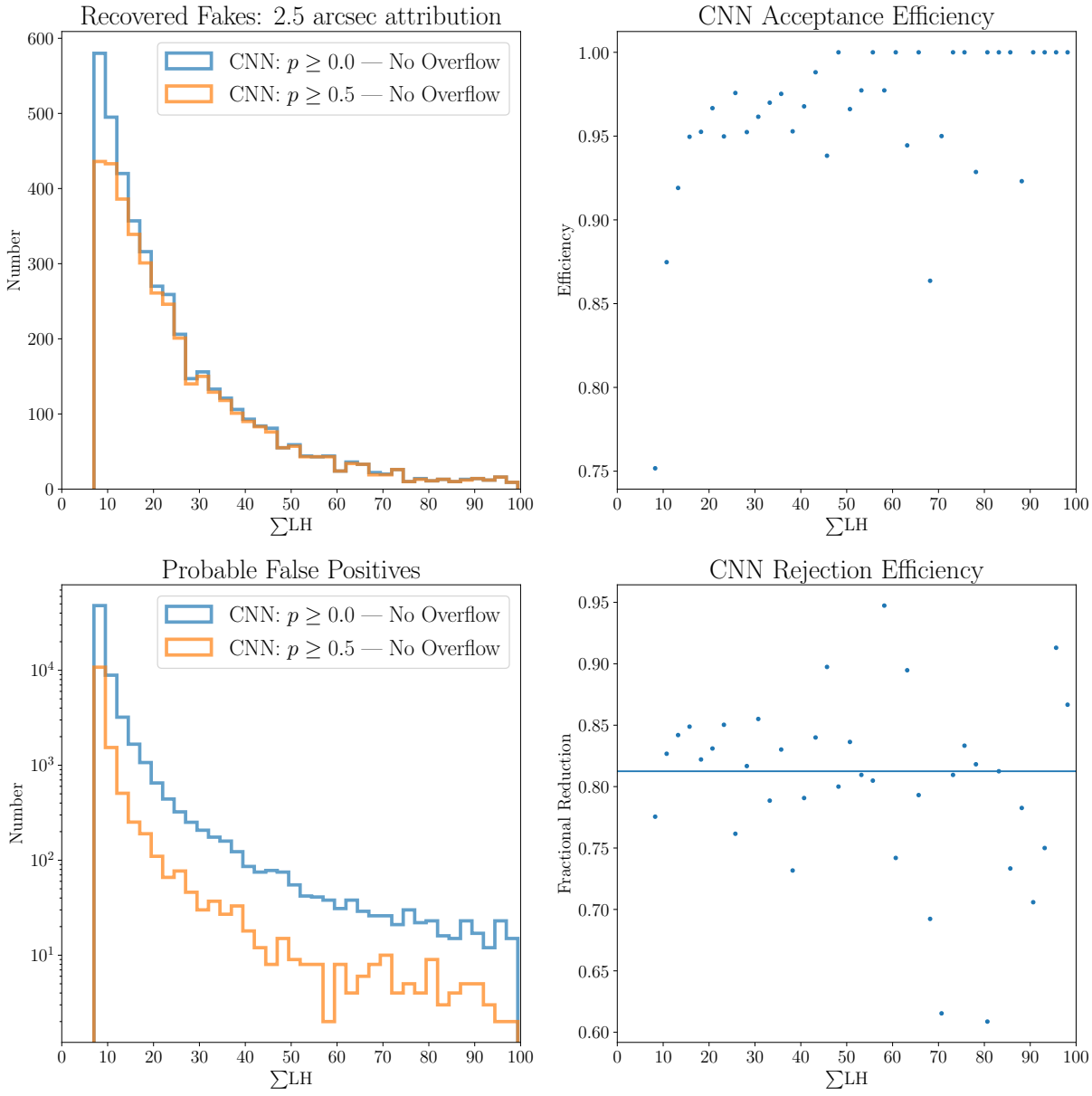

Recovered Fakes: 2.5 arcsec attribution
CNN: $p \geq 0.0$ — No Overflow
CNN: $p \geq 0.5$ — No Overflow
Number
∑LH
CNN Acceptance Efficiency
Efficiency
∑LH
Probable False Positives
CNN: $p \geq 0.0$ — No Overflow
CNN: $p \geq 0.5$ — No Overflow
Number
∑LH
CNN Rejection Efficiency
Fractional Reduction
∑LH

Figure 2.3: Efficiency of the CNN as applied to DEEP data. For this figure, we removed the "overflow" criteria, where the minimum likelihood threshold was raised from $\sum LH > 7$ to $\sum LH > 10$ if a CCD search returned more than 50 results with $\sum LH > 7$. See Section 4.3 for more information. For visibility, data were cut to include only objects with $\sum LH < 100$. Top Left: The number of fakes (implanted synthetic objects) found with "overflow" disabled with a CNN threshold p of $p \geq 0.0$ (blue) and $p \geq 0.5$ (orange). The CNN assigns a probability p that a stamp contains a TNO. This value p can then be used as a user-programmable threshold to accept or reject a canddiate trajectory. "overflow" is a data post-processing step used in Chapter 4 and is disabled here to better characterize the effectiveness of the CNN. See Section 2.4 and Chapter 4 for more detail. Fakes are automatically attributed based on their RA and Dec in the first exposure in the `KBMOD` search with a minimum attribution distance d of $d < 2.5''$. Top Right: The efficiency with which the CNN accepts the fakes. These data points were generated based on the number of fakes in each bin in the orange distribution divided by the number of fakes in each bin in the blue distribution. Bottom Left: The number of probable false positives returned by the DEEP `KBMOD` forward searches with $p \geq 0.0$ (blue) and $p \geq 0.5$ (orange). For these distributions, we removed objects that were attributed to fakes, or which subsequently passed by-eye vetting by Smotherman in the DEEP data set. We therefore label these objects "Probable False Positives", because of two possibilities. First, there is the potential for mislabelling by-eye. See Chapter 4 for labelling false positive rates. Second, there is the possibility that some objects which are present in the blue distribution but not the orange distribution (i.e. objects rejected by the CNN) are in fact true positives that are erroneously rejected. The top right plot quantifies the rate at which this happens as a function of $\sum LH$. Bottom Right: The efficiency with which the CNN rejects probable false positives. The horizontal blue line at $y = 0.813$ represents the median of the data.

in accepting real TNOs and rejecting false positives. Furthermore, this demonstrates that this CNN filter can be effectively applied to new data sets apart from the one on which it was trained[5].

To analyze the performance of the CNN on DEEP data, we changed some software parameters relative to what is used to generate the results in Chapter 4. We disabled the "overflow" functionality in the DEEP data. This "overflow" functionality increased the minimum likelihood threshold for a candidate trajectory to be considered from $\sum LH > 7$ to $\sum LH > 10$ if there were more than 50 candidates per search with $\sum LH > 7$. Then, we decreased the automatic fake attribution distance d from $d < 5.0$ arcsec to $d < 2.5$ arcsec. d is the maximum in-image distance between a candidate object and a fake. d is measured based on the candidate position in the first exposure of a `KBMOD` search. This decrease was done to reduce spurious attributions between the `KBMOD` candidate trajectories and the fakes. In the DEEP data, we ran a reverse search centered on the direction of increasing ecliptic longitude (opposite the direction both the real and fake TNOs are expected to move). We call the candidates from these searches "reverse search candidates". (See Section 4.3.2 for further discussion on the reverse search candidates). Making the assumption that all reverse search candidates are false positives (or more specifically do not correspond to any of the fakes), we quantify the false association rate by taking the fraction of reverse search candidate trajectories that link to the fakes. Using $d < 2.5$ arcsec, in the case of CNN threshold p with $p \geq 0.0$, this false association rate was about 0.11%. For $p \geq 0.5$, the false association rate was about 0.07%.

Overall, this CNN accepts consistently over 90% of the fakes in the DEEP data, except at the lower likelihood end of the distribution where one would expect slightly worse perfor-

[5] "Transfer learning" ([18, 109]) may help increase efficacy of apply this CNN to future data. Transfer learning freezes the weights of the early layers in the neural network. The network then updates the weights of the later layers using a smaller amount of training data (relative to the amount needed to train the neural network initially). Future work should investigate the number of early layers to be frozen and the amount of training data needed to update the CNN.

mance, while it rejects an average of 81.3% of false positives[6]. In short, the CNN reduces the number of candidates a human has to vet by tens of thousands of candidates in the case of DEEP in exchange for an acceptable loss in efficiency at the very faint end. This is particularly noteworthy because the CNN was trained using image background and a survey cadence from the data in Chapter 3 but still performs as described here for the data in Chapter 4. See Figure 2.3 for a visualization of the CNN acceptance and rejection efficiency.

2.5 Barycentric space

In this section, we will describe a method for searching TNO trajectories with time baselines of a few months by transforming our trajectory into the frame of the Solar System barycenter. We will motivate why we would want to make a transformation into barycentric space and how we can do so, in both `KBMOD` and in the joint-fit (described in Section 2.6). We will then characterize the effectiveness of this barycentric transformation, discussing the range of time baselines this transformation will allow us to search as well as the required grid spacing in barycentric distance.

`KBMOD` currently searches a grid of linear trajectories in image space, that is to say topocentric space. The frame of reference of these trajectories is the frame of reference of the telescope that took the images (the Blanco telescope at CTIO in Chile, in the case of Chapters 3 and 4). The trajectories of moving objects in the Solar System quickly become non-linear in topocentric space. The speed at which they diverge from linearity depends in part on their distance. In the case of TNOs, an object may remain on a linear trajectory with less than 1 arcsec (approximately 1 PSF FWHM) deviation for about a week, depending on the precise orbit. However, within a few weeks, the maximum deviation from the best-fit linear trajectory is typically much more than an arcsecond, meaning that searching a linear trajectory in topocentric space is no longer possible over the entire time baseline.

However, it is possible to instead parameterize each candidate trajectory as a great circle

[6]These acceptance and rejection efficiency values will vary depending on the choice of user-programmable CNN threshold p. These values are for $p \geq 0.5$.

in barycentric space instead. This is similar to the approach of [62] who transform candidates to heliocentric space and propagate them to a common epoch in order to more effectively link moving objects that are bright enough to be detected in a single exposure. Here, of course, the goal is to detect moving objects that are generally too faint to detect in a single exposure, rather than link existing detections.

The motion of a TNO on the sky is driven primarily by the reflex motion of the Earth. Due to the changing angles between the Solar System barycenter (about which the TNO orbits), the TNO, and the Earth, the deviation of a TNO's trajectory from a topocentric linear trajectory is primarily caused by the Earth's motion. This problem is furthered because the desired search time baseline is often a significant fraction of the Earth's orbital period. For example, when searching over a time baseline of 4 hours, a topocentric linear trajectory works well because 4 hours is much, much less than 365.25 days. However, when searching over 1 month, 3 months, or 6 months, the Earth drives strong non-linearity in the topocentric position of TNOs.

This challenge can be greatly ameliorated with a conceptually straightforward transformation into barycentric space. Because we know the position of the observatory relative to the Solar System barycenter, it is possible to remove the Earth's motion when modelling a trajectory. To define the cartesian position of the TNO, we must assume a barycentric distance to the TNO[7]. Then, using the RA and Dec of the TNO, we can determine the TNO's cartesian barycentric position. Once we have the relevant positions, the transformation is simply

$$\boldsymbol{r}_{\mathrm{TNO}} = \boldsymbol{r}'_{\mathrm{TNO}} + \boldsymbol{r}_{\mathrm{Obs}}, \tag{2.6}$$

where $\boldsymbol{r}_{\mathrm{Obs}}$ is the barycentric position of the observatory, $\boldsymbol{r}'_{\mathrm{TNO}}$ is the topocentric position of the TNO, and $\boldsymbol{r}_{\mathrm{TNO}}$ is the barycentric position of the TNO. We can then model trajectories in barycentric RA and Dec.

[7]In order to apply this method of barycentric reprojection to a digital tracking algorithm, we will need to assume a range of distances with some particular grid spacing. The range of distances will depend on the population of interest (e.g. about 38 au to 47 au to search the Kuiper Belt). We will discuss estimates for the required grid spacing later in this section.

In principle, this method could be extended to more complicated topocentric trajectories by adding additional vector terms to Equation 2.6. For example, the Hubble Space Telescope has additional motion due to its orbit around the Earth. With accurate positions for the Hubble Space Telescope, it may be possible to repeat the results of [15] with `KBMOD`, possibly allowing for longer time baselines or increased search grid resolution.

At this point, we must introduce the concept of a great circle. A great circle is the intersection of the surface of a sphere and a planar cross-section through the center of the sphere. It is parameterized along the surface of the given sphere. For example, approximating the Earth as a sphere, the equator of the Earth is a great circle. In astronomical terms, lines of constant RA form great circles. We will denote RA as α and Dec as δ. Any two points $P_1 = (\alpha_1, \delta_1)$ and $P_2 = (\alpha_2, \delta_2)$ on the sky can be connected by a great circle. The angle between these two points (with the center of the sphere as the vertex) is called the central angle σ_{12}. We can then define an angular velocity along a great circle as $d\sigma/dt$.

In barycentric space, Keplerian orbits move along great circles, with the Solar System barycenter at the center. This is true for any Keplerian orbit, no matter the eccentricity, inclination, semi-major axis, or any other orbital parameter. However, eccentricity, for example, will affect $d\sigma/dt$. Circular Keplerian orbits will move with a constant angular velocity, while eccentric orbits will have variable angular velocity over the course of the orbit.

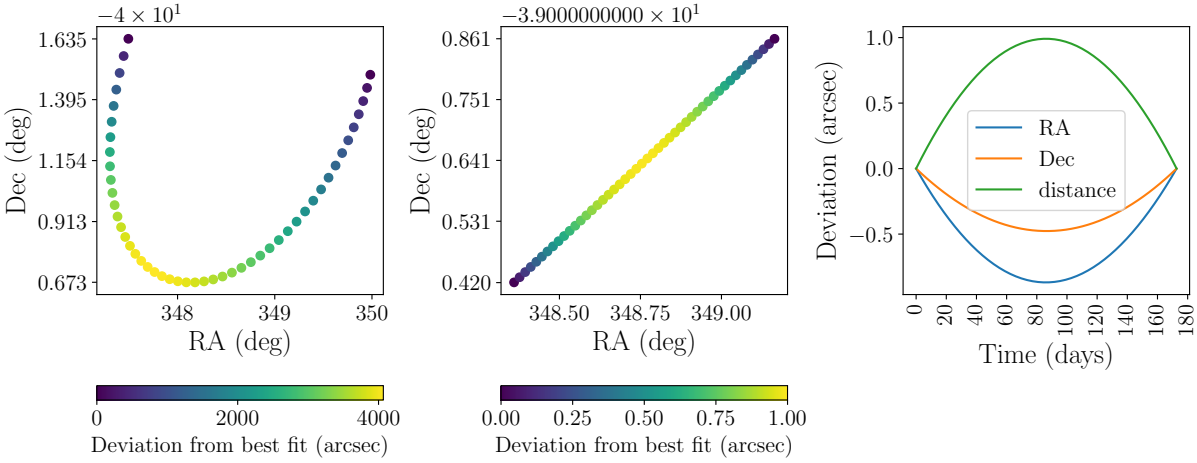
In practice, we must model TNO trajectories as barycentric great circles with constant $d\sigma/dt$. The goal of this reprojection is to allow `KBMOD` to search trajectories over several months, while adding only one more parameter to the `KBMOD` search grid. Adding more than one additional parameter to the search grid is not computationally feasible. We must already assume a barycentric distance for any given object of interest, so we cannot also assume an eccentricity without a prohibitive increase in search dimensionality and computational cost.

Using JPL Horizons[8], we can show that this reprojection may allow us to model TNO trajectories for up to 6 months with less than 1 arcsec deviation from the actual trajectory of

[8]JPL Horizons, along with documentation, can be found here: `https://ssd.jpl.nasa.gov/horizons/`

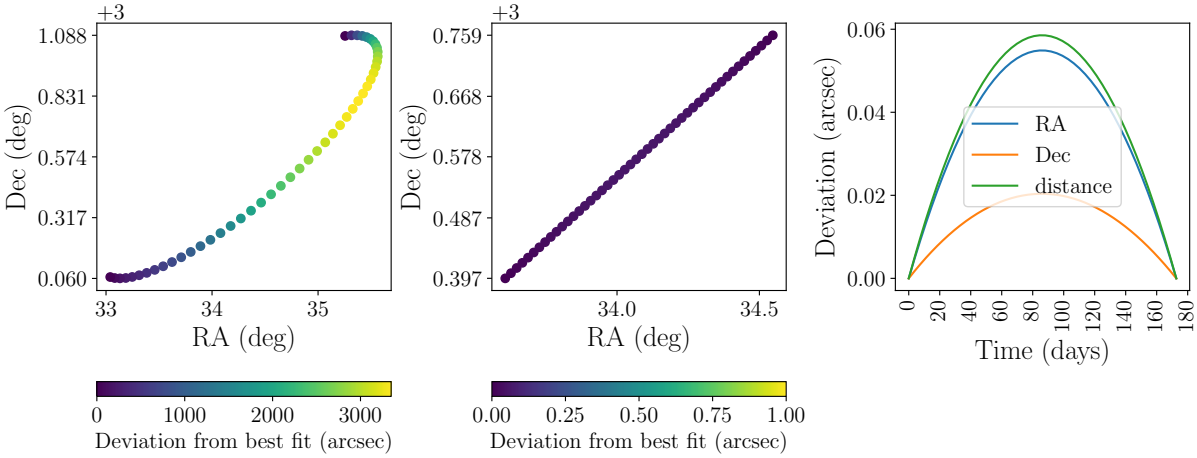
2014 QV510 : $a = 43.27$ au | $e = 0.152$ | $i = 31.50°$ | $\Omega = 47.83°$ | $\omega = 324.43°$ | $M = 329.04°$ | $d = 37.97$ au

Max Deviation: 4068.30 (arcsec) Max Deviation: 0.99 (arcsec) Deviation from best fit



2011 SO277 : $a = 30.16$ au | $e = 0.012$ | $i = 9.64°$ | $\Omega = 113.52°$ | $\omega = 117.38°$ | $M = 161.08°$ | $d = 30.50$ au

Max Deviation: 3348.87 (arcsec) Max Deviation: 0.06 (arcsec) Deviation from best fit



2014 US277 : $a = 99.46$ au | $e = 0.453$ | $i = 36.37°$ | $\Omega = 143.33°$ | $\omega = 351.27°$ | $M = 320.56°$ | $d = 78.63$ au

Max Deviation: 1943.30 (arcsec) Max Deviation: 0.45 (arcsec) Deviation from best fit

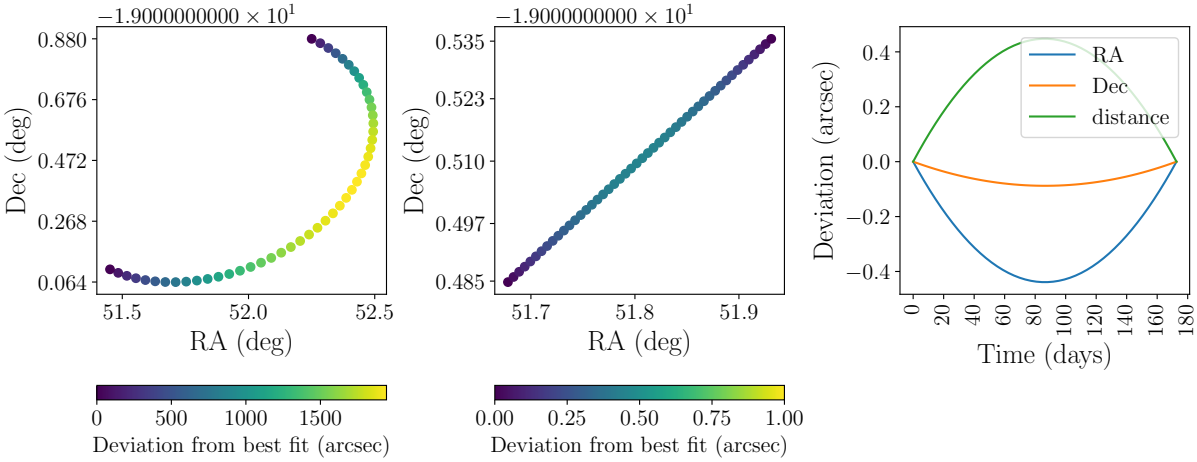


Figure 2.4: The effectiveness of reprojecting to barycentric space for three objects over a 180 day time baseline. The top, middle, and bottom row correspond to objects 2014 QV510, 2011 SO277, and 2014 US277 respectively. We queried JPL Horizons directly to obtain barycentric ephemerides for these objects. We model the trajectory as a great circle parameterized by the starting and ending RA and Dec with a constant velocity in the central angle. We do this in both topocentric and barycentric space and visualize the deviation between the great circle trajectory and the JPL ephemeris. *Left column:* The ephemeris over 180 days in topocentric space. The color axis shows the deviation from the modelled trajectory. *Middle column:* The ephemeris over 180 days in barycentric space. The color axis shows the deviation from the modelled trajectory. *Right column:* The deviation of the modelled trajectory from the barycentric ephemeris. "distance" is total distance in arcsec between the predicted position from the modelled trajectory and the barycentric ephemeris.

a TNO. We do this by directly querying the barycentric ephemerides over 180 days for three different objects: 2014 QV510, 2011 SO277, and 2014 US277. We then use the starting and ending points (α_1, δ_1) and (α_2, δ_2) of the ephemerides to parameterize great circle trajectories for each object. Finally, we model the entire object trajectory with $d\sigma/dt = \sigma_{12}/\Delta t$, where $\Delta t = 180$ days. This allows us to compute arbitrary points $P(t) = (\alpha(t), \delta(t))$ along this great circle.[9]

The objects 2014 QV510, 2011 SO277, and 2014 US277 were chosen to be approximately representative of three main classes of interest. 2014 QV510 has $a = 43.27$ au, $e = 0.152$, and $i = 31.50°$. It is therefore fairly dynamically hot and represents a somewhat more challenging TNO to model. 2011 SO277 is a Neptune trojan with $a = 30.16$ au, $e = 0.012$, and $i = 9.64°$. It is dynamically colder, with low inclination and eccentricity. It is therefore a good candidate for modelling its trajectory as a barycentric great circle of constant angular velocity. Finally, 2014 US277 is an eTNO with $a = 99.46$ au, $e = 0.453$, and $i = 36.37°$. This is by far the most eccentric of the three objects, as well as the most distant. It represents a class of TNOs that is generally difficult to detect and is of great scientific interest (see Section 1.5).

The effectiveness of modelling the trajectory of these objects as barycentric great circles is shown in Figure 2.4. We model these trajectories for 180 days and show the residuals between the modelled trajectory and the JPL reported barycentric ephemerides. In order to quantify the effectiveness of the reprojection, we quantify the maximum deviation of the modelled trajectory compared to the JPL ephemeris over the 180 day time baseline. We consider a 1 arcsec deviation to be a very approximate metric for whether or not the reprojection error would be unacceptably large. For a single observation, once the overlap becomes small between the true topocentric position of an object and a PSF placed at the predicted position (for example, at distances greater than 1 PSF FWHM), that observation contributes little signal to the coadded likelihood of a hypothetical `KBMOD` search. In practice,

[9] A complete write-up of the trigonometry involved in computing σ and $(\alpha(t), \delta(t))$ is available here: `https://en.wikipedia.org/w/index.php?title=Great-circle_navigation&oldid=1084967593`

the exact allowable reprojection error will depend on image quality, pixel scale, the fraction of the trajectory with high residuals between the predicted and true location, and likely more parameters besides. For our purposes, we want a constant metric to use among the various reprojected objects considered here, and we want it to be of the same order of magnitude as the expected seeing. To this end, 1 arcsec is a reasonable choice to use as a threshold between "acceptable" and "unacceptable" reprojection error.

All three objects considered in Figure 2.4 have a maximum deviation of below 1 arcsec. 2014 QV510 has the largest deviation between the modelled trajectory and the JPL ephemeris, with a maximum deviation of 0.99 arcsec. 2011 SO277, the dynamically cold object, has a maximum deviation of just 0.06 arcsec. Finally, despite being the most eccentric of the three TNOs, 2014 US277 is not the worst-modelled of the three. It has a maximum deviation of 0.45 arcsec. This is because the increased barycentric distance (relative to the other two objects) means that its angular velocity is slower than the other two objects. In turn, the separation between P_1 and P_2 is relatively small. In other words, 2014 US277 moves a smaller fraction of its orbit over 180 days than the other two TNOs. This helps to offset the trajectory modelling error that would be expected for more eccentric objects.

This exercise shows that it could be possible to precisely follow the trajectory of a TNO for up to 6 months by modelling its trajectory in barycentric space. However, there are several limitations to this methodology. Some of these limitations are accounted for in the results of Figure 2.4, but others are not. We will discuss some of these limitations in the following paragraphs.

First and most fundamentally, TNOs do not move along Keplerian orbits. Instead, we report the osculating elements that correspond to what the Keplerian orbital elements would be in a two-body problem. In principle, when considering only the two-body problem of a TNO orbiting the Sun, a TNO should always move along great circle trajectories in barycentric RA and Dec. In practice, small perturbations from the giant planets and other Solar System bodies will cause some deviation of the TNO orbit from the expected great circle. The exact magnitude and timescale of this variation will depend heavily on the specific orbit

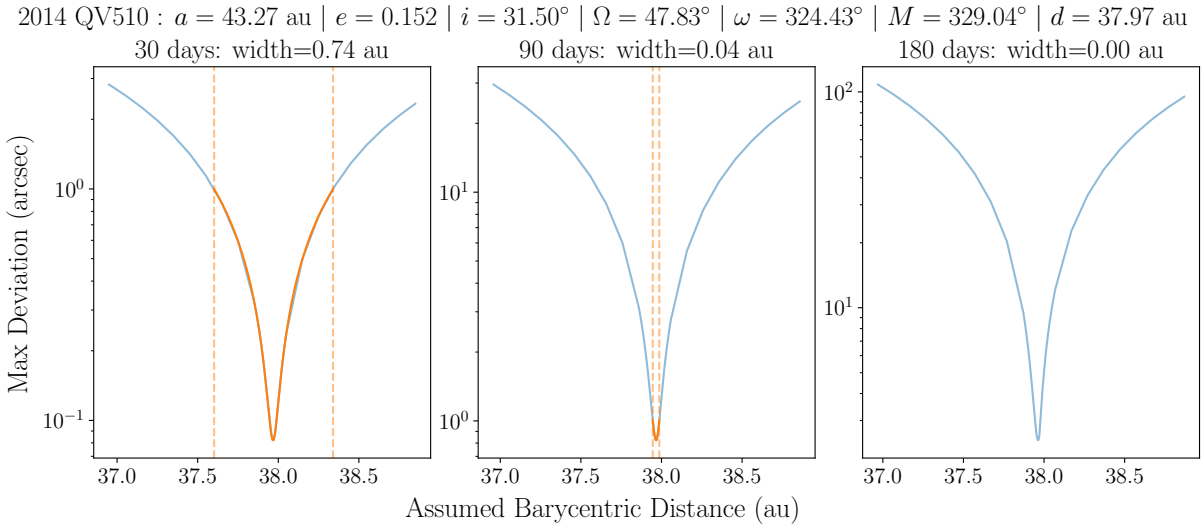
of a given TNO. For example, resonant TNOs will vary in semi-major axis differently from scattering TNOs which will vary differently from cold classical TNOs (See e.g. Figure 3 of [51]). These variations in semi-major axis often occur over thousands of years. This effect is included in the results of Figure 2.4.

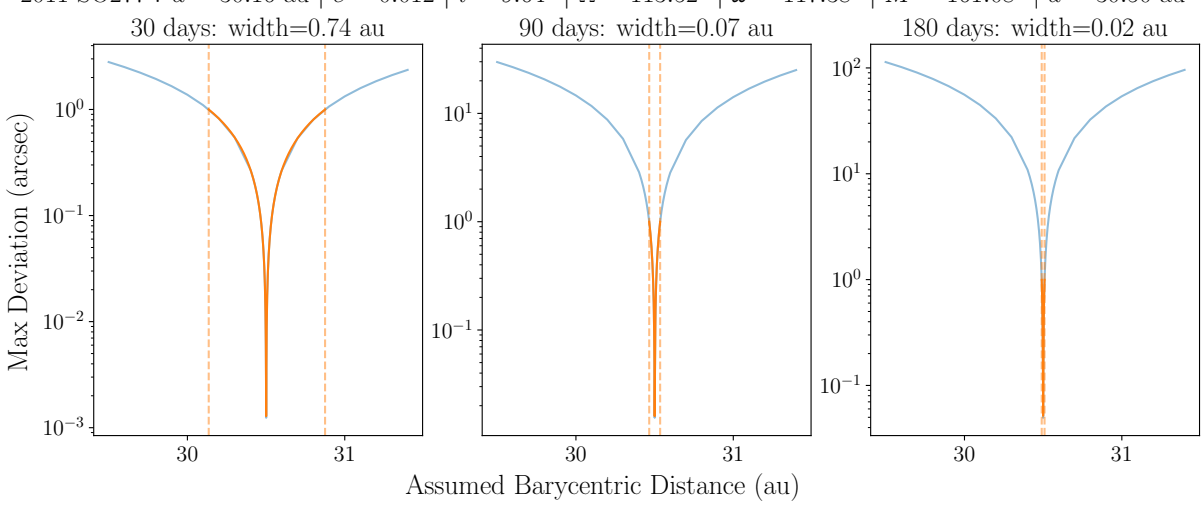
Second, as we discussed above, eccentric objects break the assumption of constant angular velocity. The degree to which this assumption is violated depends on the fraction of the TNO's orbital period that is being considered, as well as the proximity to pericenter. This can help to explain why 2014 QV510 has a greater maximum deviation than 2014 US277, despite being less eccentric. This effect is included in the results of Figure 2.4.

Third, in addition to the position of the Earth, observatory, and Solar System barycenter, we must also be able to model the position of a TNO by assuming a single barycentric distance. We can model the topocentric position (RA and Dec) with a grid of velocities and angles corresponding to the search space of interest. However, modelling the position of a TNO in three dimensions requires assuming a distance as well. An incorrect distance estimate will lead to an incorrect transformation into barycentric space. Therefore, the precision of the grid spacing for barycentric distance will be important. This effect is not included in the results of Figure 2.4.

Finally, even without orbital evolution, it is inaccurate to assume a constant barycentric distance over a series of observations of a potential TNO. Any TNO with a nonzero eccentricity will have a barycentric distance that varies over its orbit. This variation means that there is no single barycentric distance that will perfectly fit the trajectory. This effect is not included in the results of Figure 2.4.

We can investigate the effect of these last two limitations by implementing our own barycentric reprojection, rather than using the "best case" reprojection from JPL. We do this by assuming a single value for barycentric distance. We sample over a range of barycentric distances. We can then plot the maximum deviation of the great circle model to the computed barycentric RA and Dec for a grid of barycentric distances. This simultaneously tests the effect of assuming a single value for barycentric distance and quantifies the barycentric


2014 QV510 : $a = 43.27$ au | $e = 0.152$ | $i = 31.50°$ | $\Omega = 47.83°$ | $\omega = 324.43°$ | $M = 329.04°$ | $d = 37.97$ au
30 days: width=0.74 au
90 days: width=0.04 au
180 days: width=0.00 au
Max Deviation (arcsec)
Assumed Barycentric Distance (au)


2011 SO277 : $a = 30.16$ au | $e = 0.012$ | $i = 9.64°$ | $\Omega = 113.52°$ | $\omega = 117.38°$ | $M = 161.08°$ | $d = 30.50$ au
30 days: width=0.74 au
90 days: width=0.07 au
180 days: width=0.02 au
Max Deviation (arcsec)
Assumed Barycentric Distance (au)

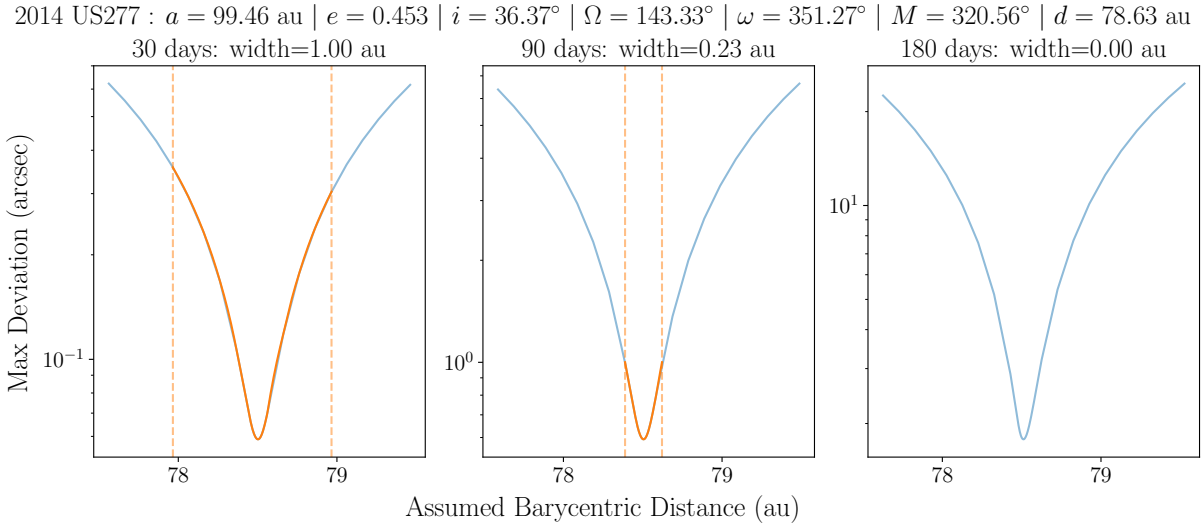

2014 US277 : $a = 99.46$ au | $e = 0.453$ | $i = 36.37°$ | $\Omega = 143.33°$ | $\omega = 351.27°$ | $M = 320.56°$ | $d = 78.63$ au
30 days: width=1.00 au
90 days: width=0.23 au
180 days: width=0.00 au
Max Deviation (arcsec)
Assumed Barycentric Distance (au)

Figure 2.5: The maximum deviation between a great circle model and the barycentric RA and Dec for three objects as a function of the assumed barycentric distance. A single and time-invariant barycentric distance is used to perform the reprojection into barycentric RA and Dec. We perform this reprojection over 30 days, 90 days, and 180 days. For each plot, blue lines correspond to barycentric distances where the maximum deviation is greater than 1 arcsec. Orange lines correspond to barycentric distances where the maximum deviation is greater than 1 arcsec. The orange dashed lines define the boundary in barycentric distance between the blue and orange regions. "Width" describes the distance region bounded by the orange dashed lines and corresponds to the approximate barycentric grid spacing that would be required to reproject these objects with a maximum deviation of < 1 arcsec.

distance grid precision necessary to maintain a maximum deviation of less than one arcsec.

We perform this investigation using 2014 QV510, 2011 SO277, and 2014 US277. We sample a grid of barycentric distances d centered on the true barycentric distance d_{true} (as reported by JPL Horizons) at time t_0, corresponding to the start of the ephemeris. We set the initial grid spacing to be 0.1 au and sample from $d_{\text{true}} - 1$ au $< d < d_{\text{true}} + 1$ au. We use these barycentric distances to reproject the ephemerides into barycentric RA and Dec, model the trajectory as a great circle, and compute the maximum deviation between the great circle model and the barycentric RA and Dec. We define the barycentric distance with the smallest maximum deviation to be d', and resample the parameter space around d' with a grid spacing of 0.001 au and $d' - 0.1$ au $< d < d' + 0.1$ au. We then update d'. We now have a range of barycentric distances and maximum deviations, with fine sampling near the smallest maximum deviation.

We then find the range of barycentric distances that yielded maximum deviations of less than 1 arcsec. To do this, we use a first order spline interpolation to find the barycentric distance above and below d' where the maximum deviation reaches 1 arcsec. We define the difference between this upper and lower barycentric distance to be the grid width. This grid width approximately corresponds to the grid spacing in barycentric distance that `KBMOD` would need in order to recover these objects.

We perform this investigation for 30 day, 90 day, and 180 day time baselines and show the results in Figure 2.5. Over 30 days, the barycentric reprojection achieves sub-arcsec precision over a wide range of barycentric distances. The grid widths over 30 days were between 0.74 au and 1.00 au. Over 90 days, grid widths decrease significantly to between 0.04 au and 0.23 au. Over 180 days, the two more eccentric objects no longer reproject with sub-arcsec maximum deviation. 2011 SO277, which is a low eccentricity object, has a grid width of just 0.02 au.

We therefore conclude that using this technique of barycentric reprojection will likely work well in practice over time baselines of less than 90 days, but it may be challenging to use this technique for time baselines greater than 90 days. Despite the limitations discussed above,

transforming the search coordinates from topocentric space to barycentric space significantly extends the time baseline over which KBMOD can search.

A version of this barycentric transformation has already been implemented into KBMOD by Stephen Portillo. The current version searches linear trajectories in barycentric RA and Dec, approximating the observatory positions as the Earth's geocenter. It should be possible to implement a complete great circle model and to include observatory positions. A more straightforward addition would be to search linear trajectories in barycentric $\alpha \cos \delta$ and δ, although a full great circle model will be more accurate and precise. Although these additions are not necessary for the work presented here, they will be necessary to apply KBMOD to LSST. We will discuss this in more detail in Chapter 5.

2.6 The joint-fit

One shortcoming of the grid-based approach of KBMOD is the lack of fine astrometric precision. Because the grid spacing of KBMOD is such that each pixel is the starting point of a trajectory, there is a fundamental limit to the astrometric precision achievable by KBMOD governed by the size of each pixel. Then, there is a further limitation due to the grid spacing chosen for the search[10]. As shown in Figure 2.6 and Figure 3.7, the fine grid spacing of KBMOD as applied to DECam data with a pixel scale of 0.263 arcsec per pixel does allow for sub-arcsecond precision, with a median offset of typically below two pixels (0.526 arcsec). However, improving the astrometric positions of candidates detected with KBMOD will improve the linking, characterization, orbit fitting, precovery, and magnitude measurements of the candidates. In particular, for the short time baselines typical of KBMOD searches, which have so far typically ranged from hours to days, any improvement we can make to the astrometry becomes important for followup and characterization.

To this end, we developed an extension to KBMOD meant to improve the measured as-

[10]Grid spacing is user-programmable. It is therefore not a fundamental limitation, as a user can improve the search grid spacing at the cost of increased computational cost. However, given the limitation on starting position, where each trajectory search starts from a single pixel, improving grid spacing has diminishing returns on improving the recovered starting pixel location of an object.

trometric positions of candidate objects. We call this extension the "joint-fit". The goal here is to take the individual stamps of each candidate that passes all prior levels of filtering and human vetting, then properly minimize the negative log likelihood of a source along a trajectory through each candidate's stamps rather than rely on sampling from a grid of trajectories. Full processing of all 23,336 candidates in DEEP which pass the CNN filtering (but have not yet been vetted by humans) takes about 2 days using 3 28 core nodes on a high performance computing (HPC) cluster, from loading the data to getting final magnitude estimates. In comparison to `KBMOD`, which can search, characterize, and filter 256 x 512 x 2048 x 4096 candidate trajectories in a few minutes on a GPU, it becomes apparent why we use the joint-fit only in the final step of candidate characterization[11].

The joint-fit requires a number of additional inputs which are not used directly by `KBMOD`. The joint-fit uses a unique PSF for each image derived by the LSST Science Pipelines, as opposed to a constant-width Gaussian PSF as used by `KBMOD`. As inputs to this fit, we load 31 x 31 pixel warped difference imaging and calexp cutout stamps, variance planes, PSFs, magnitude zero points, stamp centers in RA/Dec and pixel space, and associated `astropy` WCSs. The stamp size is user-selectable, and was intentionally chosen to be larger than the 21 x 21 pixel stamps used by default in `KBMOD`. As we will discuss later, we can place bounds on the trajectories searched within these stamps in later stages of the joint-fit. Loading these data from the LSST Science Pipelines butler and from the images saved on disk was about one third of the overall joint-fit processing time for the DEEP data set, even after optimization efforts.

---

[11]The joint-fit would be challenging to fully implement inside a `KBMOD` search, although some of the lessons learned in developing the joint-fit may be applicable. For example, the improved PSF model could be directly implemented into `KBMOD`, and future work should investigate this improvement. Not all parts of the joint-fit can be applied to the core `KBMOD` search. Running a minimization technique inside of the core `KBMOD` grid search would likely be prohibitively computationally expensive. However, there would likely be performance improvements to tying the joint-fit more closely into `KBMOD`. One could imagine a case where the joint-fit is still run at the end of a `KBMOD` search (after all filtering), but takes advantage of the fact that all images are already loaded into CPU RAM (for the `KBMOD` search). As another example, perhaps the joint-fit could be optimized to take advantage of GPU acceleration, since a GPU is already a hardware requirement to run both `KBMOD` and the CNN filter.

Next, we rescale the stamps to a common magnitude zero point (e.g. $m_{VR,0} = 31$ in the case of the DEEP data in Chapter 4). We apply this rescaling by defining a scaling constant k such that

$$k = 10^{-(m_0 - m_0')/2.5}, \tag{2.7}$$

where m_0 is the magnitude zero point of the image and m_0' is the new magnitude zero point to which the image is being rescaled. In the case of the DEEP data, $m_0' = m_{VR,0} = 31$. Then, the counts in each candidate postage stamp are multiplied by k and the corresponding variance in each pixel are multiplied by k^2.

After loading the data, we run a high-precision fit to minimize the negative log likelihood that there is a source along a given trajectory. This fit follows the same formalism as `KBMOD`. Namely, we still follow the coaddition approach described in Equations 1.10, 1.11, and 1.12. However, instead of generating each Ψ and Φ image by convolving the image with a fixed-width Gaussian, we instead convolve each image separately with the best-fit PSF returned by the LSST Science Pipelines.

Directly loading the WCS, stamp centers, and LSST PSFs allows for flexibility in the nature of the trajectory that is being fit. As mentioned, this eliminates the limitation imposed by starting candidate trajectories on a per-pixel basis, as well as searching a fixed grid of angles and velocities. In addition, it also allows users to define and search more complicated trajectories as needed for a given data set. In the DEEP data, we applied a topocentric linear fit characterized by the starting and ending pixel position. This fit was appropriate given the very short time baseline of about 4 hours for each search. However, for future applications of `KBMOD` to LSST and other data sets, where the search time baseline may be months, this tool allows a user to jointly fit other parameters alongside starting and ending positions. For example, we have included an option that fits the best linear trajectory in barycentric space, including a fit for barycentric distance.

The minimization methodology for finding the minimum negative log likelihood uses `scipy.optimize.minimize`, which includes a number of minimization methods. We find the 'L-BFGS-B' method ([142]) to be particularly well-suited to our use case. This method

allows for bounds to be placed on the input fit parameters, allowing a user to focus the joint-fit closer to the center of the stamps, for example. It also returns the inverse hessian matrix for the fit. From this, a user can compute the standard deviation of the fit for each parameter by taking the square root of the diagonals of the inverse hessian matrix. In essence, this means that this fit method returns uncertainties for each parameter.

We apply this joint-fit method to the DEEP data discussed in Chapter 4. The large population of injected fakes in the DEEP data allows us to characterize the performance of the joint-fit method. Figure 2.6 shows the astrometric residuals of these fakes before and after applying the joint-fit tool using a topocentric linear trajectory. The joint-fit tool reduces the mean residual in RA from $\mu = 0.2633$ arcsec to $\mu = -0.0572$ arcsec and the mean residual in Dec from $\mu = 0.2840$ arcsec to $\mu = -0.0178$ arcsec. This is a reduction of a factor of about 4 in RA and and 16 in Dec. The standard deviation of the residuals in RA and Dec reduces from $\sigma = 0.3389$ arcsec to $\sigma = 0.1873$ arcsec in RA and from $\sigma = 0.5427$ arcsec to $\sigma = 0.2047$ arcsec in Dec. More notably, the value for σ_G for the distribution of residuals in RA and Dec reduces from $\sigma_G = 0.2520$ arcsec to $\sigma_G = 0.0703$ arcsec in RA and from $\sigma_G = 0.4361$ arcsec to $\sigma_G = 0.0542$ arcsec in Dec. Because σ_G is more robust to outliers, this reduction in σ_G may give a better idea of the "typical" decrease in scatter one might expect from the joint-fit. Finally, there is a visible population of candidates with a systematic offset in RA and Dec before the joint-fit, which can be seen in Figure 2.6 as a cloud of candidates shifted up and right (positive RA and Dec residuals). After the joint-fit, these systematic outliers are re-centered.

However it is parameterized, it is clear that the joint-fit tool improves the precision and accuracy of the astrometric positions of `KBMOD` candidates. The main drawback to the joint-fit, of course, is that it requires candidates to already be detected by `KBMOD`. This is to say, it is strictly a characterization technique rather than a detection technique. This is by design, as `KBMOD` is primarily a detection technique.

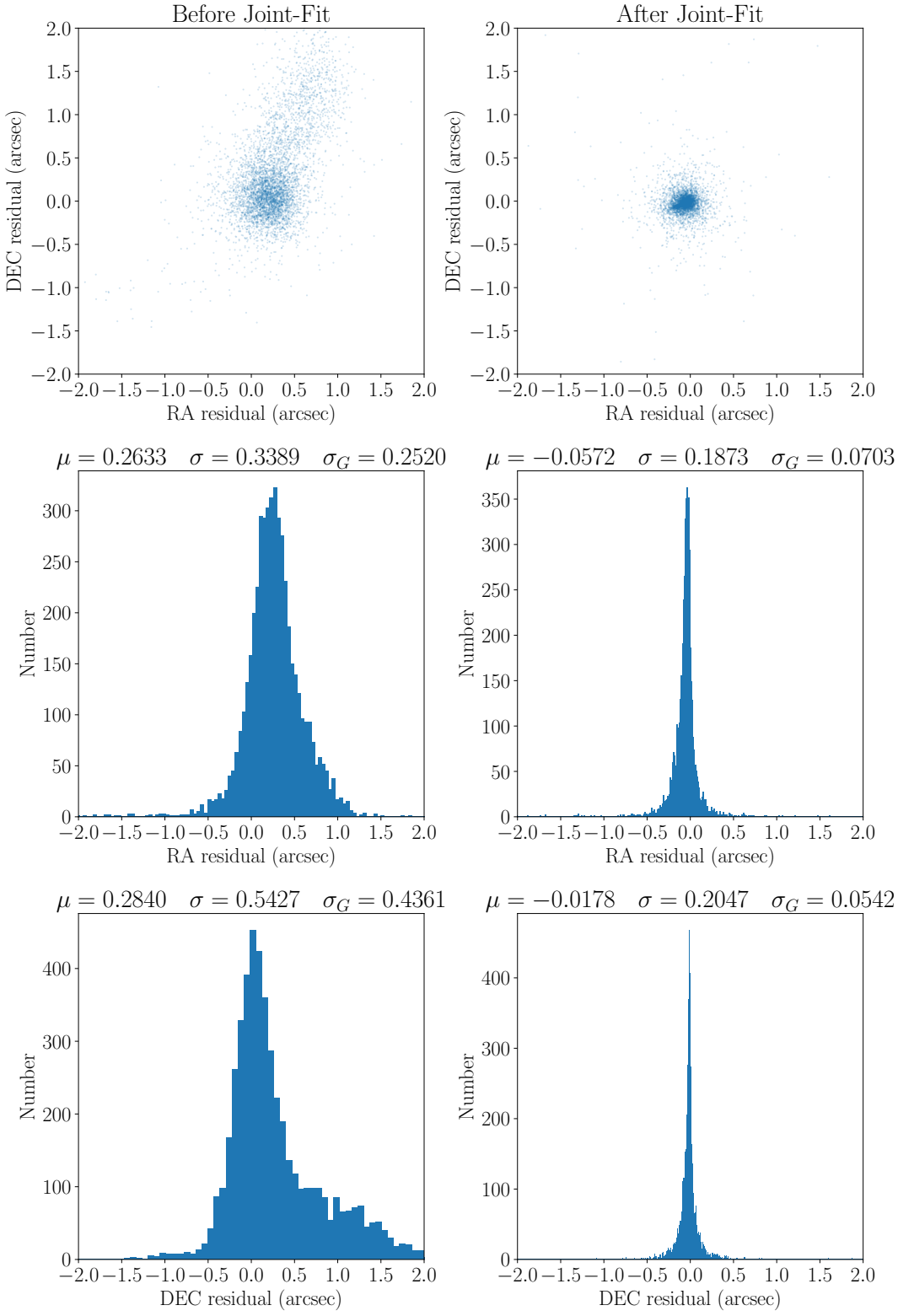

Before Joint-Fit
After Joint-Fit
DEC residual (arcsec)
RA residual (arcsec)
$\mu = 0.2633$ $\sigma = 0.3389$ $\sigma_G = 0.2520$
$\mu = -0.0572$ $\sigma = 0.1873$ $\sigma_G = 0.0703$
$\mu = 0.2840$ $\sigma = 0.5427$ $\sigma_G = 0.4361$
$\mu = -0.0178$ $\sigma = 0.2047$ $\sigma_G = 0.0542$
Number
DEC residual (arcsec)

Figure 2.6: Astrometric residuals of the fakes (implanted synthetic objects) in the DEEP B1 fields before (left column) and after (right column) a topocentric joint-fit. Residuals are measured as the difference between the known injected RA and Dec and the measured RA and Dec of the first exposure of each DEEP B1 field and night. See Table 4.1. The first row shows the Dec residuals vs the RA residuals. The second row shows a histogram of the RA residuals. The bottom row shows a histogram of the Dec residuals. The title of each histogram shows the mean μ, standard deviation σ, and σ_G value of the given distribution. The joint-fit improves the astrometric positions by reducing the systematic offset (shown by μ) as well as the spread of the distribution (shown by σ and σ_G).

Chapter 3

KBMOD APPLIED TO THE DECAM NEO DATA SURVEY

3.1 Introduction

In Chapter 2, we presented a number of algorithmic improvements to KBMOD that allow us to search for moving objects in difference images. Here, we use the Dark Energy Camera (DECam) NEO Data Survey to validate our improvements. This is a larger survey with a longer and more irregular cadence than KBMOD has been applied to in the past [e.g. 136]. Successfully running on difference images and a more complicated survey shows that KBMOD is beginning to be applicable at the scale needed for upcoming big data surveys like LSST. In Section 3.2, we discuss the DECam NEO Data Survey and the processing we applied to it using the LSST Software Stack. In Section 3.3, we discuss the results from our analysis, including the detection of 75 unidentified outer Solar System objects.

3.2 Data

3.2.1 The DECam NEO Survey data

The DECam NEO Data Survey covered an area on the sky of greater than 2000 square degrees. The ∼6.7 TB data set from the DECam NEO Data Survey (PI Lori Allen) uses the Dark Energy Camera on the 4m Blanco telescope at the Cerro Tololo Inter-American Observatory (CTIO) [37]. The DECam NEO Data Survey consists of 32 nights of data. In the first 10-night observing run in 2014, [123] found 235 unique NEOs.

Each individual image taken by DECam is a composite of 62 2K x 4K science CCDs, with a fill factor of 0.8 [61]. Each CCD image covers an area of ∼0.04 square degrees with a pixel scale of 0.27 arcsec. This results in a total field of view for DECam of about 3 square degrees. The CCDs are 250 μm thick fully depleted devices, with a peak quantum efficiency

above 85% at $\sim$ 650 nm [37]. Gaps between CCDs are between 153 pixels (columns) and 201 pixels (rows). Observations for this data set were taken in the *VR* filter, a broad optical filter extending from 500 to 760 nm.

We separate this data set into 782 pointing groups based on RA and Dec. CCD 01 and 61 had no data in our images, leading to a set of 60 CCDs per pointing group. We define a pointing group as a set of DECam exposures within 25 arcsec of a common RA and Dec and define a pointing as an individual DECam exposure (i.e. a set of 60 CCDs) in a pointing group. Most pointing groups contain between 5 and 25 pointings. Pointing groups characteristically have 5 pointings per night, with all data taken over nearly-consecutive nights. The intra-night pointings are taken about 5 minutes apart for a total intra-night timespan of approximately 25 minutes.

24 pointing groups had a high stellar number density, with more than 10,000 sources detected in a CCD. When astrometrically calibrating these images (see Section 3.2.2), these pointing groups exceeded the memory limits of the available computational resources and were therefore excluded. The current limitations regarding the processing of dense fields with LSST Science Pipelines are described in [121]. Detectability of moving objects with `KBMOD`, however, is driven strongly by the quality of the difference images.

372 pointing groups contained data from at least 4 unique survey nights. Because of the short intra-night image cadence, which can cause slow-moving objects to exhibit minimal motion within a night, we only search over pointing groups with at least 4 unique survey nights. This ensures that any given `KBMOD` trajectory will search a sufficiently large number of unique on-sky positions, thereby reducing the probability of linking of static objects.

In order to efficiently focus computational resources and human effort on higher-quality data, we selected 43 pointing groups from the set of 372 pointing groups. These 43 pointing groups have a total effective search area of approximately 132 square degrees. We refer to these 43 pointing groups as the "search sample". These pointing groups are shown in Figure 3.1. This down select from 372 pointing groups was as follows. 40 pointing groups existed where all pointings in the pointing group had a maximum seeing full width at half maximum

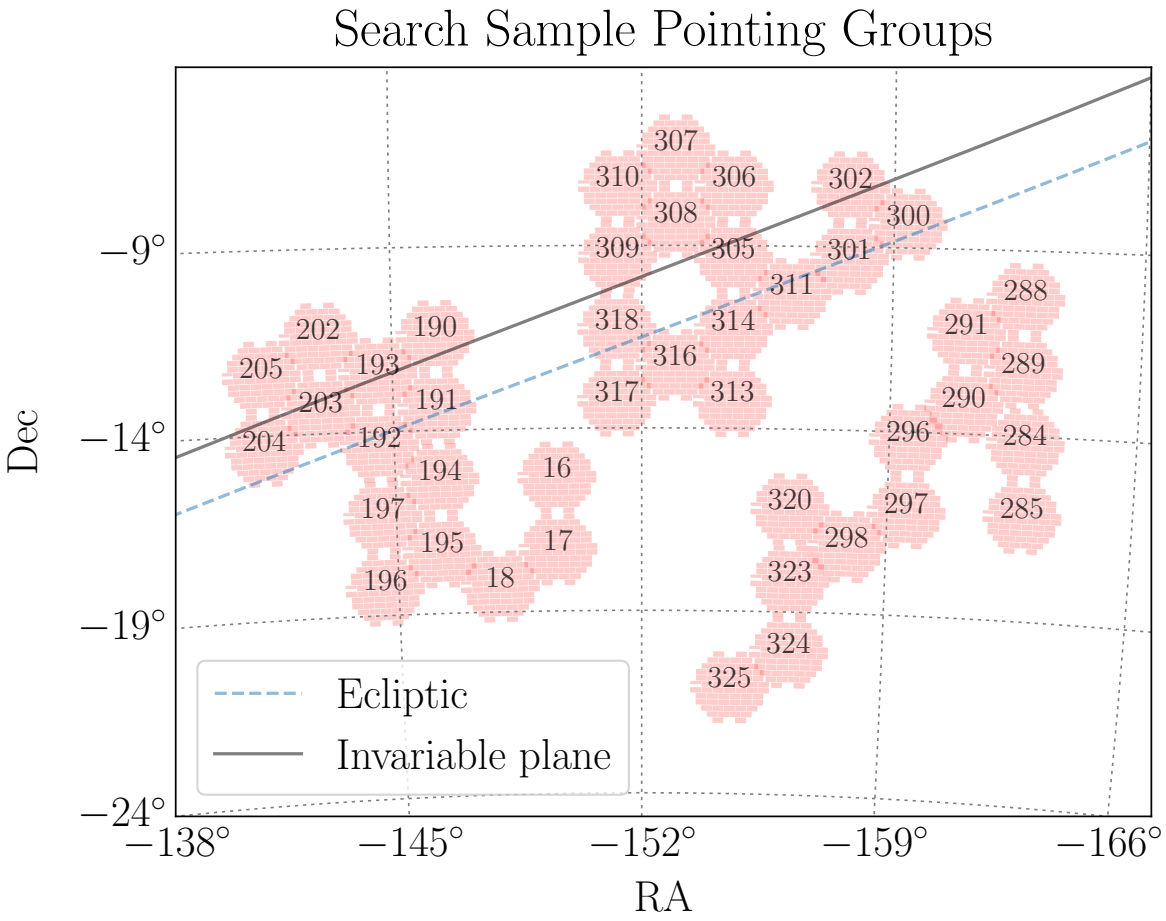


Figure 3.1: A visualization of the RA and Dec of the 43 pointing groups that make up the search sample. The red shapes show the DECam camera footprint plotted at the location of the first image in each pointing group in the search sample. The pointing group number is annotated over each corresponding camera footprint. The ecliptic is shown as a blue dashed line. The invariable plane is shown as a black solid line.

(FWHM) of 1.25 arcsec. These 40 pointing groups make up the bulk of the search sample. There were an additional 12 pointing groups that had over 20 total pointings, but with only 20 pointings with seeing < 1.25 arcsec. These pointing groups returned a greater number of erroneous candidate trajectories that required by-eye rejection. This is possibly due to the inclusion of poor-seeing images in the image differencing template (see Section 3.2.2). Due to computational limitations, we elected to run `KBMOD` on only 3 of these pointing groups, focusing our GPU resources on the 40 pointing groups where all 20 pointings had the required seeing limits. These 3 pointing groups make up the remainder of the search sample.

3.2.2 Processing the DECam data

The raw DECam images were processed by the DECam Community Pipeline [129] resulting in a set of InstCal PROCTYPE images, as defined in the NOAO Data Handbook [101]. These images are bias and linearity corrected, flat-fielded, and sky-subtracted by the community pipeline. Data quality masks and inverse variance arrays were provided. We downloaded the compressed InstCal data from the NOAO Data Archive between July and November of 2017.

Prior to running the `KBMOD` pipeline, we first astrometrically calibrate the images in all 782 pointing groups. This was undertaken using the LSST Science Pipelines Software [69]. Sources were detected in the individual images. Sources with a SNR ≥ 40 were matched to the data from the GAIA Data Release 1 (DR1). The median astrometric scatter for the sources used to fit the CCD world coordinate systems (WCS) was 25 mas; 373 CCDs had an astrometric scatter worse than 100 mas. The median number of sources detected per CCD was 3575.

As a followup to [136], we use image differencing to remove non-variable and non-moving sources within an image (as opposed to just masking the sources). We used the LSST Stack to difference the images in the pointing groups using a method based on [3]. For the DECam NEO Survey data, we difference each pointing against a coadded template. Given the short intra-night time separation between images of a given pointing group, objects moving slower

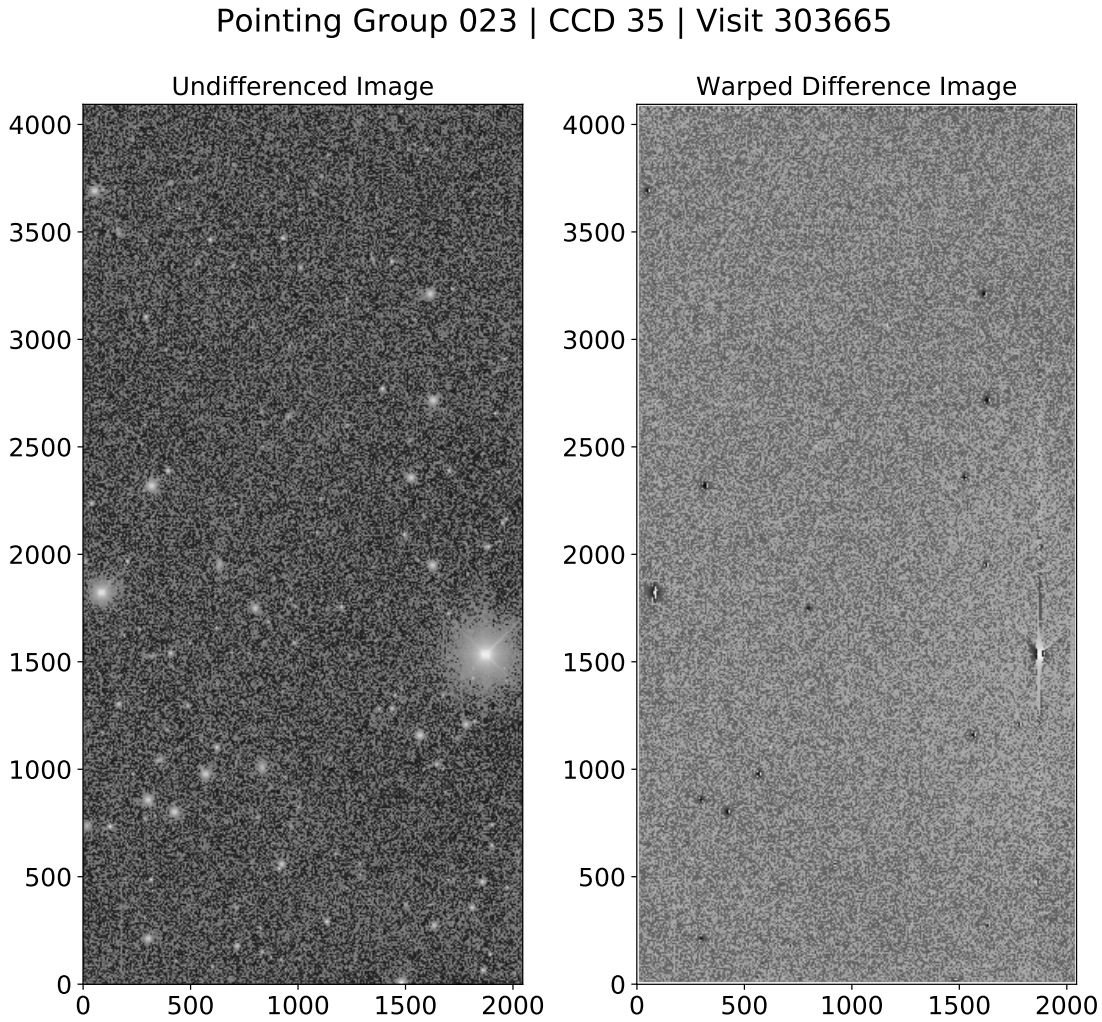


Figure 3.2: Single pointing (pointing group 011, CCD 29, visit 303605) before (left) and after (right) image differencing. Similar to DS9, we applied an arcsinh filter to the pixels in this example in order to better show objects in each image.

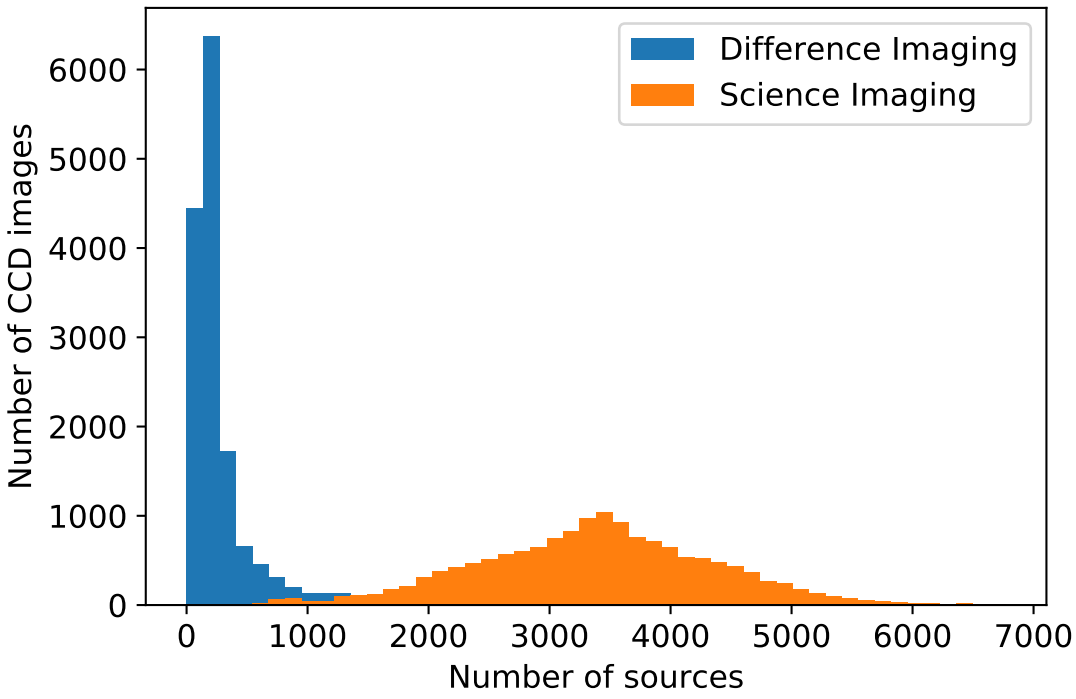


Figure 3.3: Number of sources per CCD image for each visit in 10 pointing groups (pointing group 091 to 100). The median number of sources per science image (orange) is 3396 per CCD image. The median number of sources per difference image (blue) is 180 per CCD image. Differencing the science images therefore reduces the number of static sources in the image by a factor of about 18.

than of order 1 arcsec hr^{-1} will not move a full psf width over a single night. We therefore separate a pointing group into two approximately equal groups such that each image in the first group will be separated in time from each image in the second group by at least twelve hours. A coadded template was independently generated from each group and used to difference the opposite group. Because our minimum search velocity is ≥ 92 px/day (≥ 24 arcsec per day), this guarantees that objects of interest will be much greater than one PSF away from where they were in the coadded template. This means that pointings in the middle of a pointing group—with respect to time—will have the shortest image differencing baseline, and will therefore set a theoretical limit on the slowest-moving objects we can detect.

In order to difference the science images against the coadded template [3, 140, 139, 141], we need to find a convolution kernel K such that for a science image $I(x)$ and a coadded template $\Phi(x)$, $I(x) = K \otimes \Phi(x)$. Following the approach of [3], we separated the template into local spatial cells of 128 x 128 pixels. We detected sources in both images, and grouped them into the spatial cells. Stamps of these sources were created with sizes between 21 x 21 pixels and 35 x 35 pixels, depending on the FWHM of the source. Stamps in each cell were used to find the local spatially-invariant convolution kernel solutions of each stamp. The local convolutional kernel was modelled as a set of Gaussian functions multiplied with a polynomial. The coefficients of the kernel were then found by solving a least-squares problem. One source (and thus one stamp) was selected for each grid cell based on the clipped mean of all the kernel solutions in the cell. This gave the local convolution kernel for that cell. In order to account for spatial variation over the gootprint of an entire CCD, a spatially-variant global convolutional kernel is required. Chebyshev polynomials of the first kind were fit to the local kernel coefficients in order to determine a model for spatially-variant global convolution kernel coefficients. This global kernel was then used to match the PSF of the coadded template to that of the science image. We matched the template to the science image, rather than the other way around, because the template has less noise and the convolution correlates noise. The two images were then subtracted. Finally, a decorrelation algorithm was run to remove the correlation in the noise of the difference

image. Decorrelation is particularly important when coadded templates are relatively low SNR, as is the case with these data. After differencing the image, we warp all images in a pointing group to the sky plane of the first pointing in the pointing group. This ensures that a pixel in one pointing will correspond to the same RA and Dec as that of the same pixel in another pointing.

As an example, Figure 3.2 shows pointing group 023, CCD 35, visit 303665 before and after image differencing and warping. The final image size for the KBMOD image is set by the intersection of the image and the template that is subtracted. Slight misalignments of the pointings in a pointing group may reduce the final image sizes. All pointing groups were, however, aligned to within 50 arcsec in RA and Dec, with all but 28 pointing groups aligned to better than 25 arcsec in both RA and Dec. Therefore the reduction in image area was minimal.

The KBMOD TNO search was run for each aligned stack of DECam CCDs independently; we did not search trajectories across CCD boundaries. In principle, we might like to search all plausible TNO trajectories across the largest footprint allowed by the data. Algorithmic and hardware limitations preclude this possibility, as KBMOD currently requires that all images and initial search results must fit inside GPU memory. We will discuss this in more detail in Section 5.4.1. For the data described in this Chapter, the effective area on which we are able to search for moving objects is, therefore, about 0.04 square degrees. In other words, a necessary requirement for the detection of a moving object with the KBMOD algorithm is that the object stays within the field of view of an individual CCD for at least two pointings. In practice, we require that an object stay in the field for at least 3 nights (typically 15 pointings). This means that an object must move slower than about 15 arcsec hr^{-1} to be detected by KBMOD. As we will discuss in the next section, our KBMOD search already sets an upper bound at 6.19 arcsec hr^{-1} (consistent with typical KBO topocentric speeds). Therefore, this limitation of the data (which sets an approximate upper bound at 15 arcsec hr^{-1}) imposes no additional restrictions on detection ability.

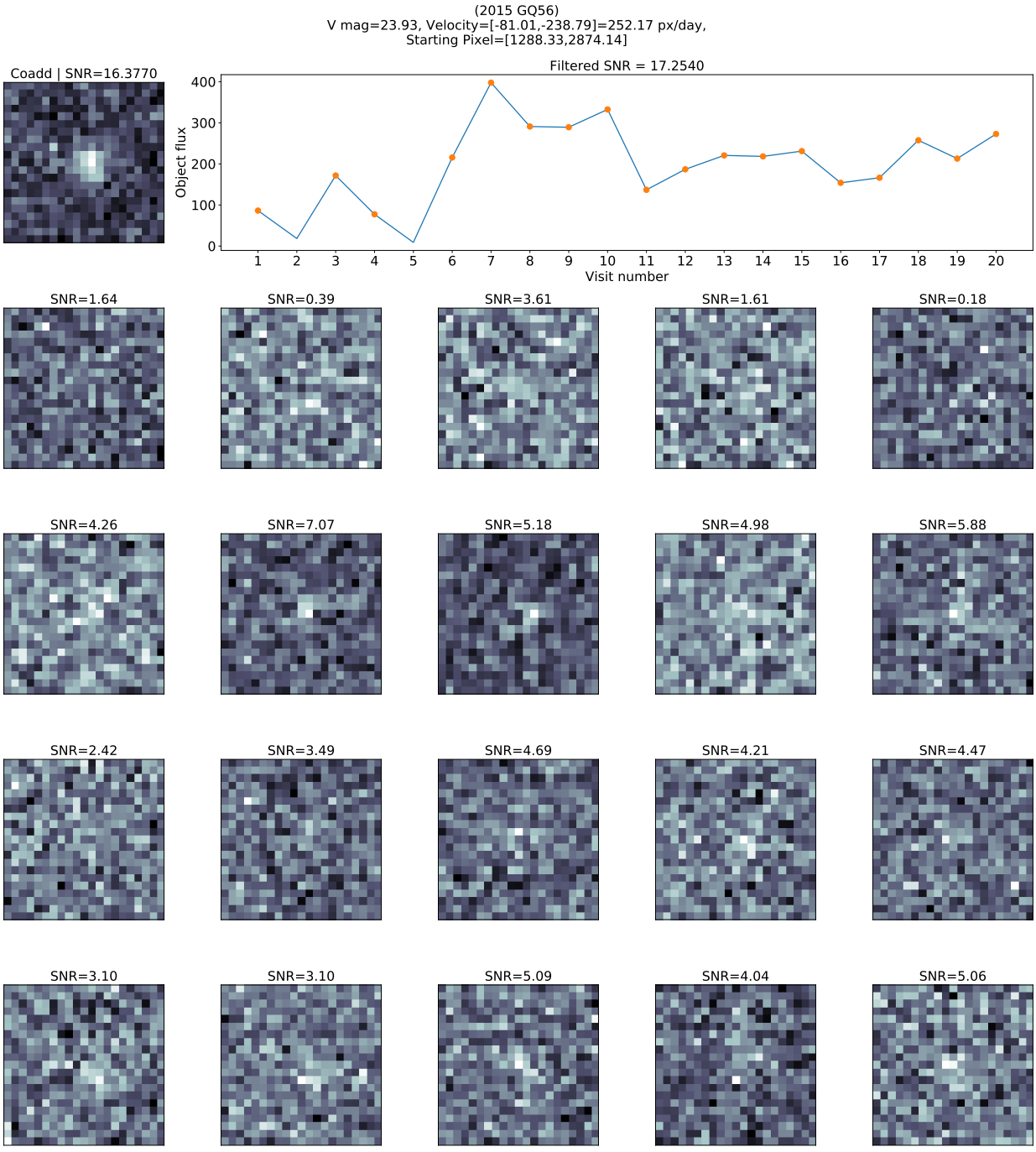

(2015 GQ56)
V mag=23.93, Velocity=[-81.01,-238.79]=252.17 px/day,
Starting Pixel=[1288.33,2874.14]
Coadd | SNR=16.3770
Filtered SNR = 17.2540
Object flux
Visit number
SNR=1.64
SNR=0.39
SNR=3.61
SNR=1.61
SNR=0.18
SNR=4.26
SNR=7.07
SNR=5.18
SNR=4.98
SNR=5.88
SNR=2.42
SNR=3.49
SNR=4.69
SNR=4.21
SNR=4.47
SNR=3.10
SNR=3.10
SNR=5.09
SNR=4.04
SNR=5.06

Figure 3.4: An example output for object 2015 GQ56 (pointing group 300, CCD 30) when using trajectory estimates from the JPL Horizons service. The first row shows the coadded stamp (left) and the flux lightcurve (right). Orange points in the flux lightcurve are points that pass σ_G lightcurve filtering. The remaining rows show the postage stamps for 2015 GQ56 in each individual image. The coadded stamp was generated by taking the median value at each pixel; this effectively removes image differencing artifacts. This trajectory was generated using orbital values from JPL Horizons. These figures were generated for all known KBOs in search sample in order to determine the unfiltered $\sum LH$, as well as for debugging purposes. Each stamp shows the estimated SNR ν of that stamp.

3.3 Results

3.3.1 Search, detection, and recovery

We ran an untargeted KBMOD search on each stack of CCDs in the search sample for a total of 2580 searches. Similar to [136], an untargeted search looks for linear trajectories with velocities between 92 and 550 px/day (1.04 arcsec hr^{-1} to 6.19 arcsec hr^{-1}) with angles of $\pm\pi/10$ from the ecliptic angle. Compared to [136], we doubled the resolution of the grid spacing from 256 velocity steps and 128 angle steps to 512 velocity steps and 256 angle steps. This ensured that trajectories would end up separated by no more than about two PSF FWHM from neighboring trajectories.

In order to test the efficiency of these new filtering methods, we generated a list of known objects in the search sample. We used Skybot [17] and JPL Horizons [47] to find all KBOs that were present in the search sample, with the additional requirement that they be present in the first image of the pointing group. We impose this requirement because a KBMOD search requires that an object be present in the first image[1]. We generated 21 x 21 pixel postage stamps of the object in each image in which it is present. We developed a variant of KBMOD that computes the likelihoods along a single trajectory then runs the aforementioned quantile-based filtering, and computed the central moments of the postage stamps. Figure 3.4 shows these results for pointing group 300, CCD 30, object 2015 GQ56. We removed

[1]This is a current algorithmic limitation due to the way that KBMOD handles the results of each candidate trajectory. Results are generated and stored based on the starting pixel location. Trajectories which begin off of the CCD are therefore not considered. Future work should investigate ways to mitigate this limitation.

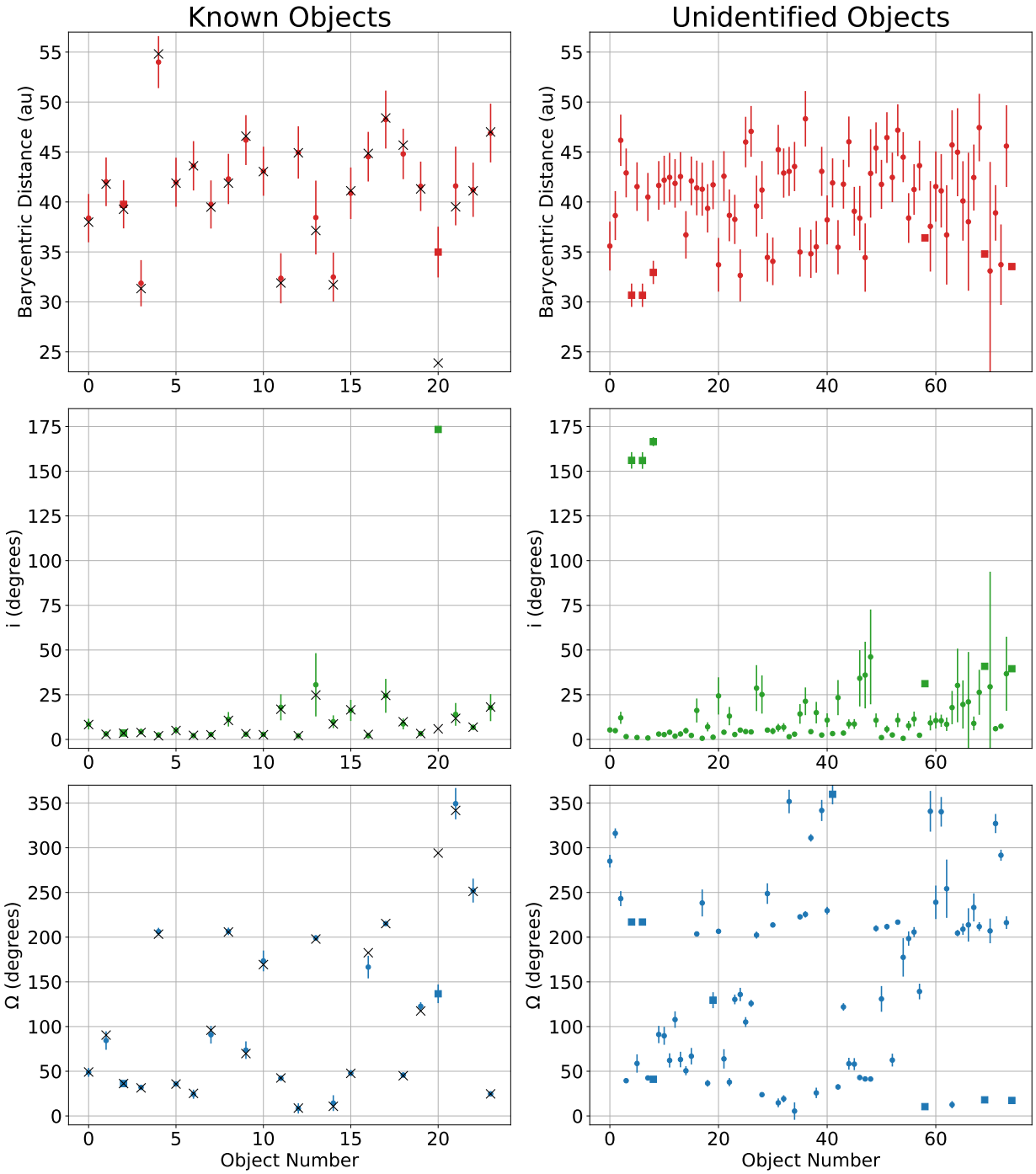

Known Objects
Unidentified Objects
Barycentric Distance (au)
i (degrees)
Ω (degrees)
Object Number

Figure 3.5: Best-fit barycentric distance r_0, inclination i, and longitude of ascending node Ω (dots) with respective standard deviations (lines) of the detected known objects (left) and unidentified objects (right) using the method of [16]. r_0, i, and Ω were also fit with Find_Orb. When the value from Find_Orb is inconsistent with [16] within 1σ, we show the best-fit value from [16] with a square instead of a dot. For the known objects, the JPL Horizons value of the corresponding parameter is overplotted with an x marker. The short time baseline of the observations allows us only to constrain initial barycentric distance, inclination, and longitude of ascending node. The medians of the absolute value of the residuals between the best-fit values and the JPL Horizons values are 0.36 au, 0.32 degrees, and 0.92 degrees for r_0, i, and Ω respectively. As reported by JPL Horizons, the median values of the known objects for r_0 and i are $\widetilde{r_0} = 41.55$ au and $\widetilde{i} = 5.46^\circ$ respectively. The median values of the unidentified objects for the best-fit r_0 and i are $\widetilde{r_0} = 41.28$ au and $\widetilde{i} = 7.67^\circ$.

KBOs with an unfiltered $\sum LH < 15$[2]. This left us with a "recovery sample" of 26 KBOs.

In the untargeted search of the search sample, we recovered 22 out of 26 (or 84.6%) of the known objects in the recovery sample after all filtering was applied (see Figure 3.6). The CNN probability threshold was kept at 75%. Recovery statistics for these objects are shown in Figure 3.7. For object recovery, we discarded any trajectories that had a starting position more than 5 pixels (approximately 1.35 arcsec, or one PSF FWHM) from the predicted location or had a velocity difference from the known velocity of more than 5 px/day (approximately 0.056 arcsec hr^{-1}). The median position and speed residuals were 0.427 arcsec and 0.0036 arcsec hr^{-1} respectively, significantly below the chosen cutoff values. This velocity error corresponds to approximately a 1.27 pixel position error over 4 days. Using the NOAO DECam Exposure Time Calculator (ETC), we estimate the single-image 10σ depth to be at most 22.75V. Because the pointing groups contain data from different nights, we computed this limit assuming a new Moon. It is therefore an upper limit. 18 of the recovered objects were fainter than the upper-limit single-image 10σ depth. Repeating this exercise at the 5σ level estimates the single-image 5σ depth to be around 23.53V. There were 10 recovered objects with known V magnitudes greater than this approximate depth of 23.53V. This confirms that `KBMOD` is able to use difference images to find moving KBOs that are too dim to detect in a single image at the 10σ level, extending the result of [136] to difference images.

We investigated each of the missed known objects individually. 2013 GY136 (pointing group 204, CCD 57) failed to process due to a CCD that failed image differencing. This

[2]As of July 2022, about a year after this initial analysis, there are about 310 KBOs within 90 arcmin of each pointing group center according to Skybot. There are about 68 KBOs present on the CCD in the first visit with $\sum LH > 0$, and another 7 with $\sum LH < 0$. These numbers serve to give an estimate of the scale of the down-sampling from nearby KBOs, to KBOs that should be present on the image, and finally to KBOs with recoverable $\sum LH$. We specify the time because it is possible that updates to skybot may have altered the returned list of KBOs since these data were originally processed (e.g. in the case of new discoveries or updated orbits). However, the maximum discovery year for all KBOs within 90 arcmin of the search sample pointing group centers (as determined from the object name) is 2015. This may help indicate that there are no newly-discovered KBOs, since the time of the initial data processing. Furthermore, the number of known objects meeting the criteria for the recovery sample remains at 26, as of July 2022.

reduced the total number of images in CCD 57 to fewer than 20, and CCD 57 was therefore not reprocessed. 2013 GZ137 (pointing group 202, CCD 52) failed CNN filtering with a threshold of 75%, but passes with a threshold of 50%. 2015 GY55 (pointing group 306, CCD 26) starts within 4 pixels of the chip edge, causing this trajectory not to be searched by `KBMOD`. 2013 GH137 (pointing group 192, CCD 41) has two fully-masked stamps, and two more with partial masking, which may have caused it to be filtered out by stamp filtering regardless of its $\sum LH$ value.

In addition to the detected 22 known objects in the recovery sample, we detected 2 additional known KBOs. These KBOs had an unfiltered $\sum LH < 15$ along the JPL Horizons trajectories, and were therefore not included in the recovery sample. The best `KBMOD` trajectories for these objects had a filtered $\sum LH' > 15$. We then linked these objects back with known KBOs.

3.3.2 Orbit fitting and analysis

We detected 75 moving objects that we were unable to link to existing objects. We refer to these objects as "unidentified" rather than "previously unkown" because it is likely that some of these 75 objects are previously known objects that Skybot was unable to propagate to the epoch of these data with sufficient precision for them to be included in the recovery sample of known objects. This could happen if, for example, a TNO was initially discovered with a short arc and has not yet been followed up. Trajectories with $\sum LH > 15$ that passed all filtering were accepted or rejected with a by-eye examination of the individual stamps, the coadded stamp, and the flux lightcurve.

As shown in Figure 3.3.1, we used the method described in [16] to fit barycentric distance r_0, inclination i, and longitude of ascending node Ω of both the recovered known objects and the unidentified objects. For the known objects, we compared the orbital parameters fit to the `KBMOD` trajectory with their respective parameters as reported by JPL Horizons. The medians of the absolute value of the residuals between the best-fit values and the JPL Horizons values are 0.36 au, 0.32 degrees, and 0.92 degrees for r_0, i, and Ω respectively. The

median values for r_0 and i of the known objects reported by JPL Horizons are $\widetilde{r_0} = 41.55$ au and $\widetilde{i} = 5.46°$ respectively. The median values of the unidentified objects for the best-fit r_0 and i are $\widetilde{r_0} = 41.28$ au and $\widetilde{i} = 7.67°$. The three parameters (shown in Figure 3.3.1) that are well-fit with our data constrain the plane of the orbit and the initial distance of the object from the Solar System barycenter. Individual values are shown in Table A.1.

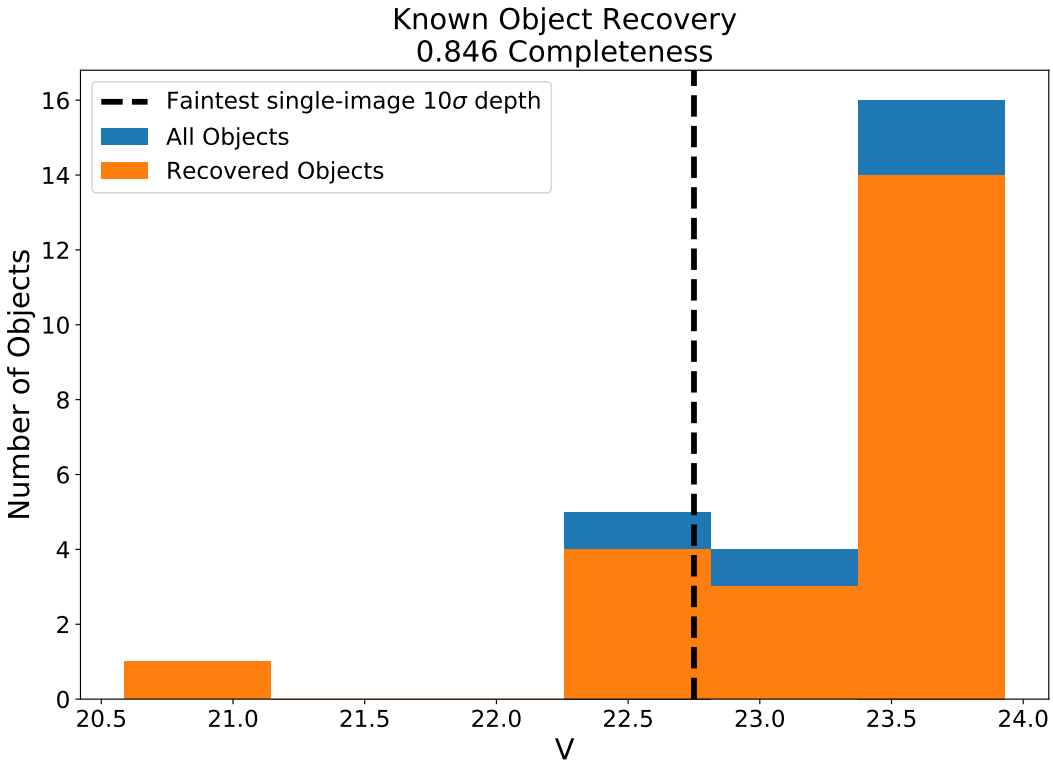


Figure 3.6: Recovered known objects as a function of reported magnitude. We ran an untargeted `KBMOD` search on all CCDs in the search sample that had known KBOs on them. Figure 3.7 shows the recovery statistics for the recovered objects. 18 of the recovered objects were below the approximate upper-limit single-image 10σ limiting magnitude.

In addition to the method of [16], we used Find_Orb[3] to fit r_0, i, and Ω. This allowed us to compare the best-fit values between the two orbit fitting codes. When best-fit values from Find_Orb were not within 1σ of the best-fit value from the [16] code, we show the value as a square in Figure 3.3.1. We discarded values with inconsistent inclinations from the remainder of the orbit analysis.

For the known objects with consistent barycentric distance, we find that our analysis with the Bernstein TNO code overestimates our barycentric distance uncertainties. The median of the absolute value of the barycentric distance residual is 0.29 au, while the median of the uncertainty in barycentric distance is 2.50 au. Our estimates of barycentric distance, therefore, are generally better than the error bars in Figure 3.3.1 would imply.

There were a few noteworthy limitations to our data set and apparent outliers in our best-fit values. Because of the relatively short time baseline of about 4 days, we were unable to place any meaningful constraints on the other Keplerian elements individually. For three unidentified objects (unidentified object numbers 58, 69, and 74), the orbit fitting code did not return uncertainties. We therefore consider them inconsistent between [16] and Find_Orb. Unidentified object numbers 4, 6, and 8 have a best-fit inclination of $i_{\rm fit} > 90°$. Similarly, known object number 20 (2000 EE173) has a best-fit inclination of $i_{\rm fit} = 173.36° \pm 0.54°$ but a JPL Horizons inclination of $i_{\rm Horizons} = 5.95°$[4]. However, these 4 objects are all marked as inconsistent between Find_Orb and [16]. As such, their best-fit values are removed from further orbital analysis.

To evaluate the consistency of the properties of our detected asteroids with published distributions, we apply the analysis of [136] to the detected objects with consistent inclinations reported in this paper. We compared our observed inclination distribution with that

[3] `https://github.com/Bill-Gray/find_orb`

[4] This object in particular can be better understood by looking at both the barycentric distance and the inclination together. For a retrograde object, the barycentric object motion adds constructively with the reflex motion. For a given topocentric speed, it is possible for there to be two consistent solutions: a retrograde orbit at greater barycentric distance and a prograde orbit at smaller barycentric distance. In this case, the orbit fitting code of [16] selected the orbital solution with a distance consistent with TNOs, rather than the correct prograde solution with a barycentric distance less than 25 au.

of [21] by using a one-sided Kuiper variant of the Kolmogorov-Smirnov (K-S) test. We use a test statistic of $D\sqrt{N}$ where N is the number of objects, and D is given by Equation 30 in [136]:

$$D = \max\left(P_j - j/N\right). \tag{3.1}$$

P_j is the probability for a given inclination distribution that an object j has an inclination equal to or below the actual inclination i_j. P_j is also a function of the inclination distribution (in this case the double Gaussian of [21]) and the discovery latitude. As discussed in [136], P_j varies uniformly from 0 to 1. We therefore call the distribution of P_j the "expected uniform distribution", as shown in Figure 3.8. (See Section 4.2.1 of [136] and Section 3.3 of [21] for more detail). Some TNO sub-populations have non-uniform inclination distributions around the ecliptic. This is an unmodelled systematic in our test statistic. We compute P_j using Monte Carlo methods. We take 10^5 inclinations from the [21] distribution, place them randomly along circular orbits and take all objects within $\pm 0.5^\circ$ of the ecliptic latitude β_j of discovery. These values allow us to find P_j by calculating the probability that an object with a given β_j has an inclination at or below i_j. We run 1000 Monte Carlo simulations, using the mean $D\sqrt{N}$ as our test statistic. See Section 4.2.1 of [136] and Section 3 of [21] for more detail.

Our mean value for $D\sqrt{N}$ was 1.40. As shown in Figure 3.8, we reject the null hypothesis that our observed inclinations come from the distribution of [21] with only 76.6% confidence, which is less than the 1σ confidence level of 84.1% ($D\sqrt{N} = 1.47$). This is to say that we cannot confidently reject the null hypothesis. We can therefore say that our observed inclinations are consistent with [21].

We repeated the further comparison of [136], using an approximate survey simulation to identify the distribution of objects with a given inclination that we would expect to find given the central RA and Dec of our search sample. We modelled the DECam field of view as a circle with a diameter of 2.2°. We selected the inclinations and orbits from the Monte Carlo simulations used to generate Figure 3.8 and recorded the objects visible within the simulated

camera footprint. We then normalized this simulated object distribution to the number of detected objects in the search sample. Figure 3.9 shows the simulated distribution (blue) and the observed distribution (orange). The χ^2 value between the simulated and expected distributions was 8.58, corresponding to a p-value of 0.48. We therefore again say that our observed inclinations are consistent with [21].

3.3.3 Magnitude estimation and analysis

Figure 3.10 shows our estimates of the VR magnitude of the known and unidentified objects detected with `KBMOD`. To fit the VR magnitudes, we generated 25 x 25 pixel postage stamps in the undifferenced science images following the `KBMOD` linear trajectory. In each stamp, we fit for the location of the object by maximizing the value of the flux minus the stamp background. The flux was calculated by summing the counts within a circular top-hat psf with a radius of twice the FWHM of the stamp. The local stamp background was estimated from the region outside of this psf. The magnitude zero point was obtained from the InstCal images. We then took the median magnitude value from each set of 15 to 20 magnitude estimates.

As we did in [136], we compared our joint magnitude distribution with the apparent magnitude luminosity function presented in [43], adjusting for ecliptic latitude by using the inclination distribution of [21]. We use the mean $VR - R$ KBO color reported by [43]. We note, however, that the DECam VR filter of our observations differs somewhat from the Mosaic2 VR filter used in [43]. They have similar central wavelengths but different filter response curves. Individual magnitudes are shown in Table A.1. The magnitude uncertainties listed in Table A.1 are reported as σ_G uncertainties estimated from each set of magnitude estimates.

We approximate the camera footprint as a 3 square degree circle. In practice, our trajectories do not cover the entire camera footprint. Each individual `KBMOD` search only uses data from a single CCD, requiring 60 individual searches to cover a full camera footprint. We further require that each candidate trajectory has at least 15 observations, corresponding to

a time baseline of about 3 days. Depending on the search velocity and angle, this means that any objects that start near a CCD edge will not be searched, as the trajectory will go off the CCD edge before the trajectory has the requisite 15 observations. We define an effective search fill factor as the fraction of the CCD that is actually searched with `KBMOD`. For our search parameters, the search fill factor varies from about 0.5 to about 0.9. Assuming a typical KBO speed and angle of 275 px/day with an in-image angle of 4.4 radians gives a typical search fill factor of around 0.7. Multiplying this by the camera active-pixel fill factor (e.g. [61]) of about 0.8 gives a typical net fill factor of approximately 0.55. The net fill factor therefore refers to the total fraction of a DECam focal plane that is searchable with `KBMOD` relative to a circle with an area of 3 square degrees.

Figure 3.11 shows a histogram of our observed VR magnitudes along with the number expected from [43] assuming a fill factor of 1.0 and 0.55. Our joint magnitude distribution is largely inconsistent with [43] assuming a fill factor of 1.0, but is consistent to within uncertainties up to about $VR = 23.25$ assuming a fill factor of 0.55.

In Chapter 5, we discuss some of the conclusions we can make from this analysis. We discuss some of the limitations of the work presented in this chapter, the key parameters we are able to characterize, and how this work fits into the broader effort of validating `KBMOD`, applying it to current data, and preparing it for future surveys.

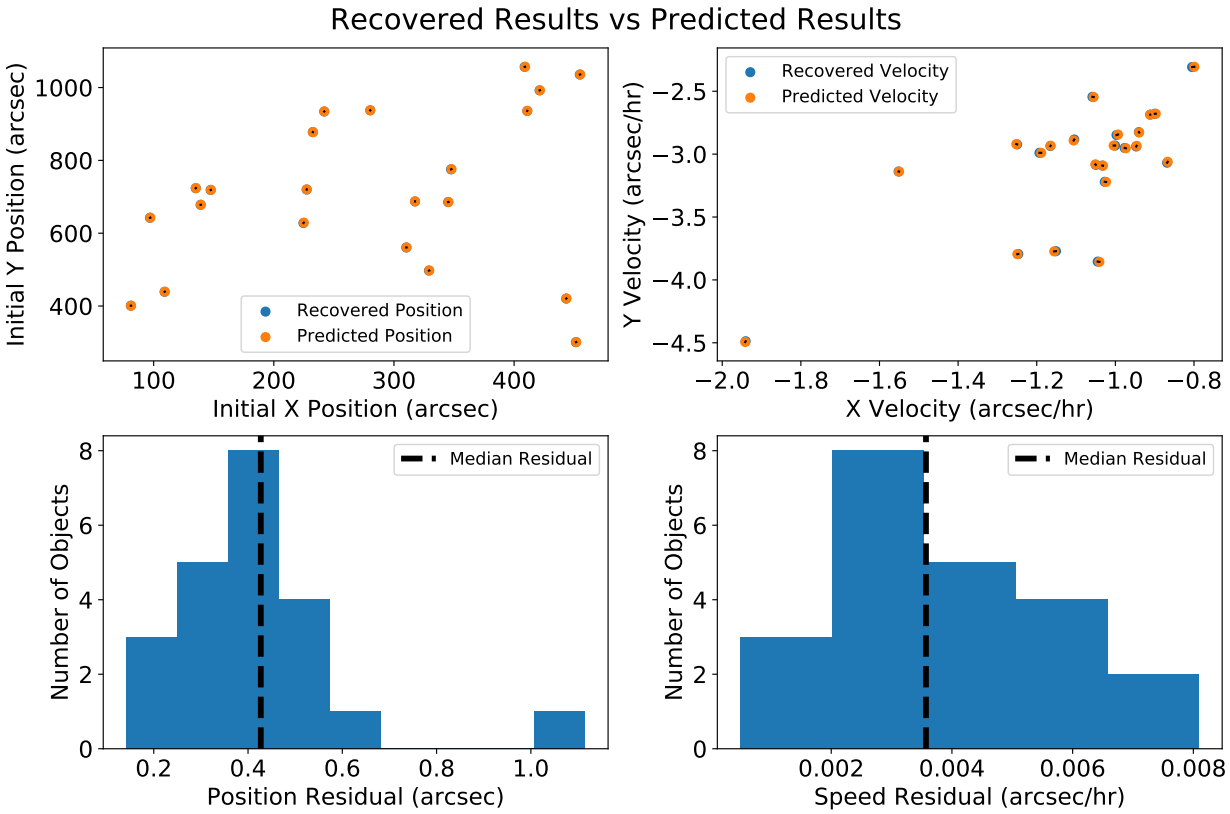


Figure 3.7: Statistics for the known objects that were recovered with an untargeted KBMOD search on CCDs with known objects in the search sample. For object recovery, we discarded any results that had a starting position more than 5 pixels (approximately 1.35", or one PSF FWHM) from the predicted location or had a velocity difference of more than 5 px/day (approximately 0.056" hr^{-1}). The velocity cutoff was chosen based on the recovery distribution. As shown in bottom left and bottom right, respectively, the median difference between predicted and recovered position and speed was significantly lower than these cutoff values. The upper left plot shows each trajectory's initial predicted and recovered position on the CCD image for each object. The upper right plot shows each trajectory's predicted and recovered x and y velocity on the CCD image for each object.

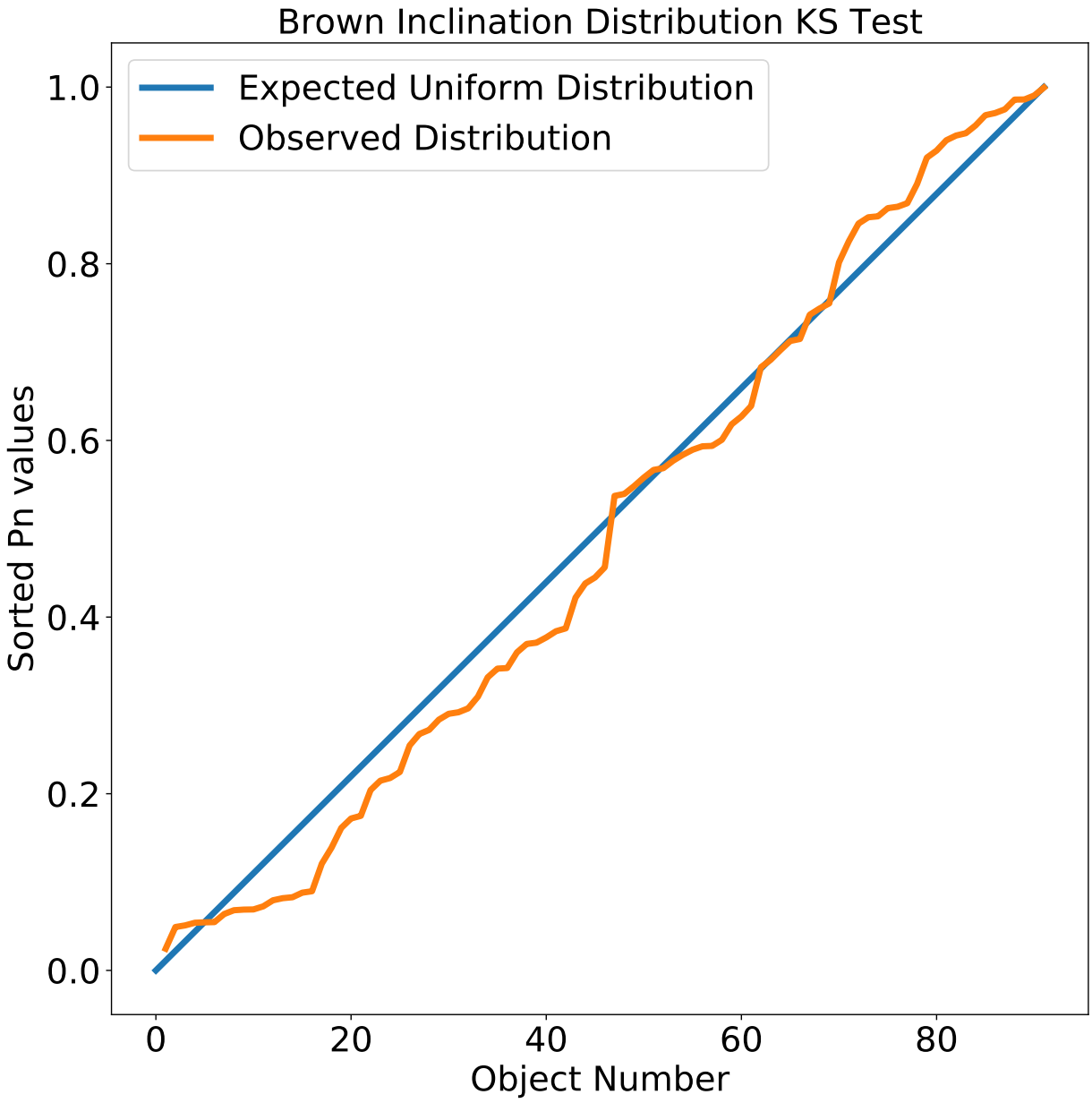


Figure 3.8: One-sided Kuiper variant of the Kolmogorov-Smirnov (K-S) test comparing our recovered inclinations with the inclination distribution predicted by [21]. We reject the null hypothesis that our inclinations came from the distribution of [21] with only 76.6% confidence (less than 1σ). We therefore consider our observed inclinations to be consistent with the distribution predicted by [21].

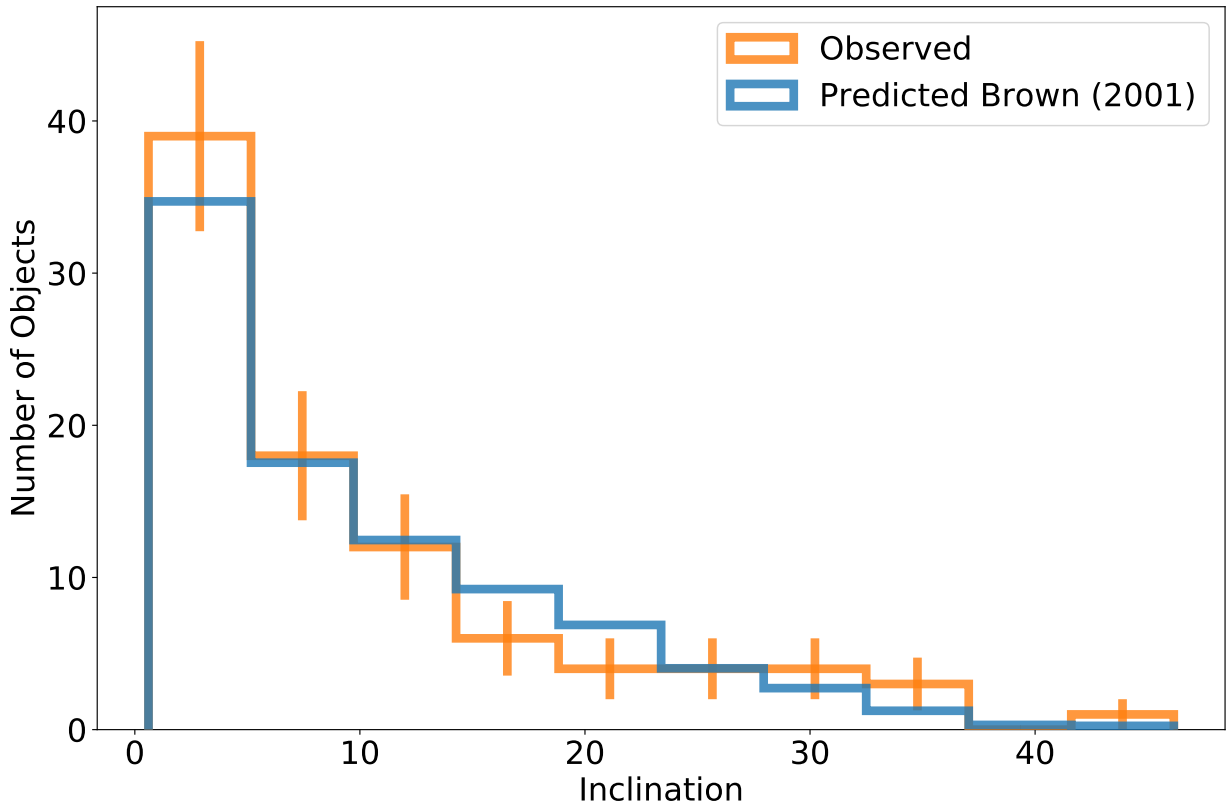


Figure 3.9: Inclination distributions of our detected objects (orange) and the distribution predicted based on [21] (blue), after accounting for the search sample ecliptic latitudes. The [21] is weighted to the number of objects recovered by `KBMOD`. Uncertainties in the orange histogram are calculated as 1σ Poisson intervals of $\sqrt{N}$.

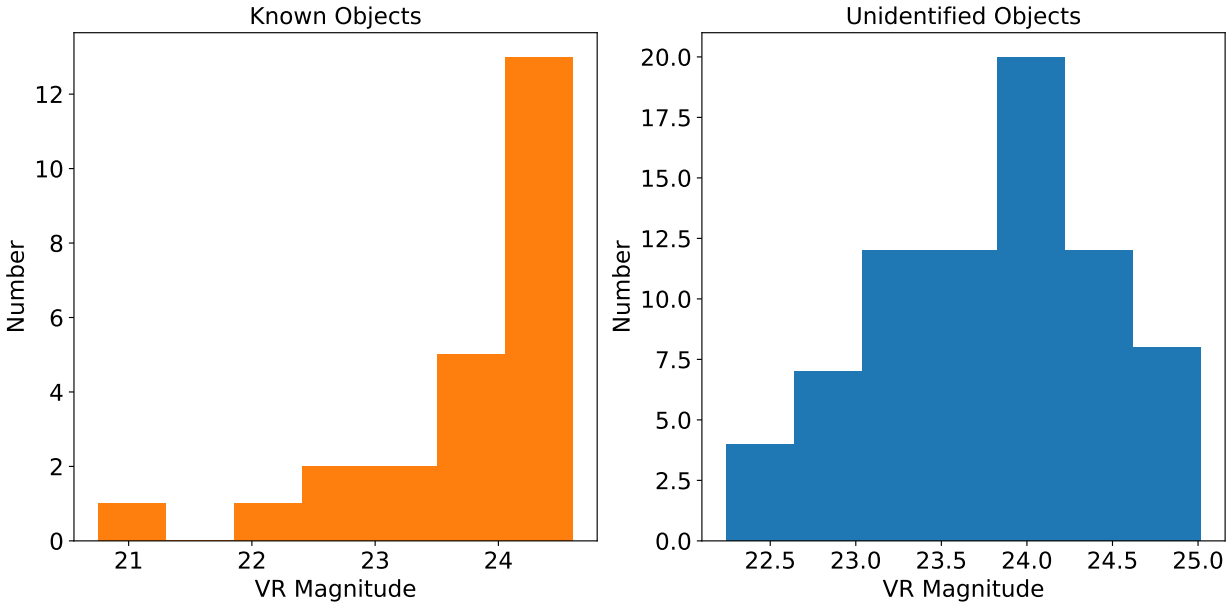


Figure 3.10: Best-fit *VR* magnitudes for the previously-known objects (left) and unidentified objects (right).

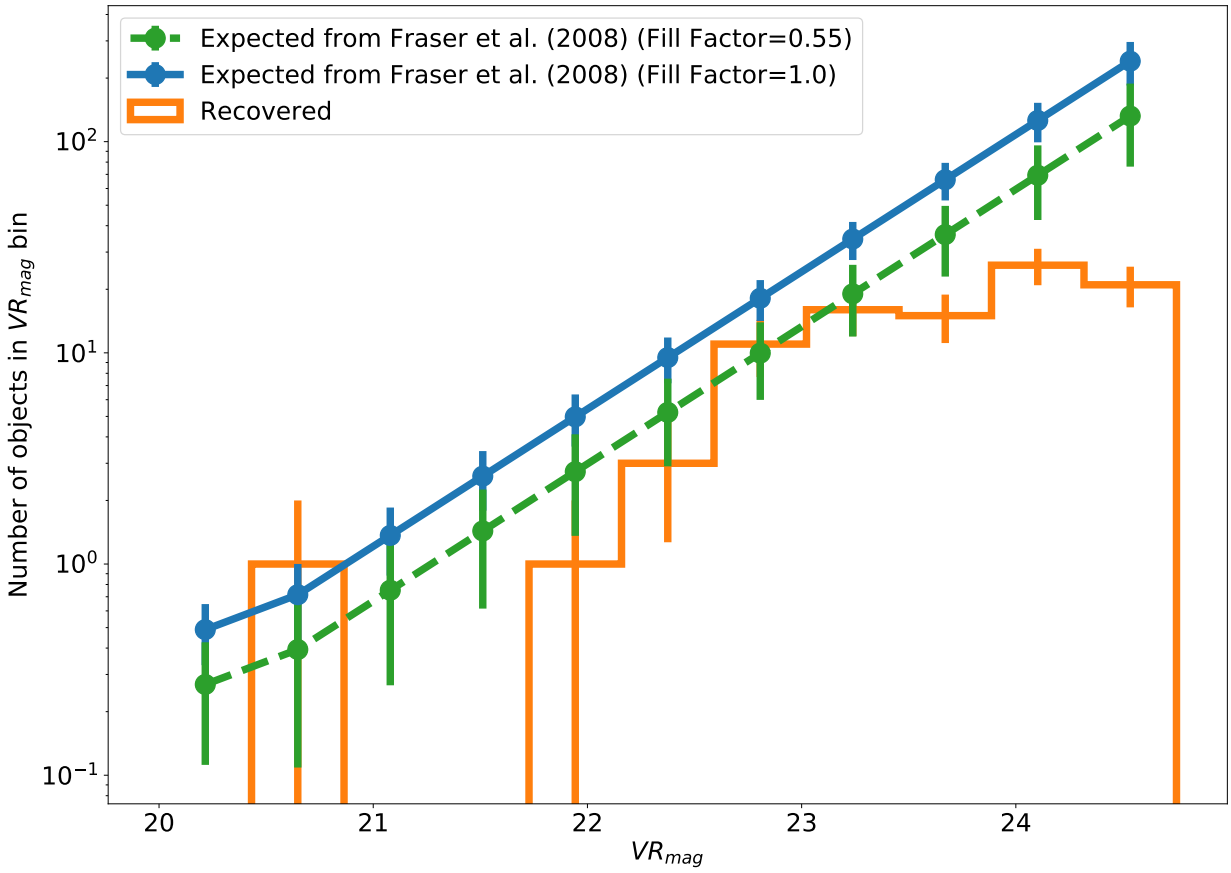


Figure 3.11: *VR* magnitude distribution (orange) of recovered objects (known and unidentified), along with the number of objects predicted by [43] assuming a circular camera footprint of 3 square degrees with no fill factor (blue) and a fill factor of 0.55 (green). Uncertainties in the orange histogram are calculated as 1σ Poisson intervals of $\sqrt{N}$.

Chapter 4

KBMOD APPLIED TO THE DEEP SURVEY

4.1 Introduction

In this chapter, we discuss the application of KBMOD to another DECam data set called the DECam Ecliptic Exploration Project (DEEP). The DEEP survey was designed specifically with digital tracking in mind. As we will discuss in Section 4.4, this enables detections that are over a magnitude fainter than any detected in Chapter 3. Furthermore, the cadence of the data allows for objects to be linked across multiple years, leading to accurate and precise measurements of the osculating orbital elements for many objects. In particular, the total baseline is sufficient to break the eccentricity-semi-major axis degeneracy which plagues short time baseline observations.

The work presented here will also be presented in a series of papers by the DEEP Collaboration [124, 125, 12, 91, 120].

4.2 Data

DEEP is a survey using the Dark Energy Camera on the 4m Blanco telescope at the Cerro Tololo Inter-American Observatory (CTIO) [37]. As with the data set in Chapter 3, the data are taken in the DECam *VR* band, which has a similar band center to (for example) Pan-STARRS r_{p1} but a wider overall bandpass. This makes it a good detection band, although it is less preferred for precision photometry.

DEEP is setup with a series of "fields" and "nights". For this survey, we define a "field" to be a particular RA and Dec observed with DECam. Thus, a field is made up of 60 CCDs and has a field of view of about 3 square degrees. Many fields are revisited multiple times over a period of about two years and as of the time of writing, more data are still being taken.

Therefore, each field also has a "night" associated with it, based on the `NITE` keyword in the FITS header. The night is simply a string date identifier based on the night of observation. Each night is associated with a "long stare", or a series of between 52 and 130 consecutive or nearly-consecutive exposures. We run `KBMOD` on each CCD in a given long stare. We will therefore describe a set of long stare observations using a combination of field and night. For example, Field B1b Night 2020-10-15 is a set of 100 exposures taken one after another on the night of October 15, 2020. Each exposure in a long stare has a 120 second exposure time.

Furthermore, DEEP is separated into 4 quadrants, indicated by the first two letters of the field name. These quadrants are "A0", "A1", "B0", and "B1". Each quadrant is composed of a number of fields that spread out in a triangular pattern to capture the dispersion of TNOs over time. This will be described in more detail in [124, 125]. This work focuses on the 2019, 2020, and 2021 data of the B1 quadrant, which has 10 unique fields and 29 total nights of data. The total footprint of the DEEP B1 data covers about 30 square degrees and is shown in Figure 4.1. The fields, nights, and number of exposures of all the data in the B1 quandrant are shown in the following table[1]. For example, for Field B1a Night 2019-08-28, there are 102 total exposures in the long stare. The total summed exposure time is therefore 12240 seconds, or 3 hours 24 minutes, long. This long effective exposure time, combined with digital tracking, is what allows DEEP to detect such faint TNOs from the ground.

[1]The DECam NEO data set discussed in Chapter 3 has a total of 43 pointing groups. For comparison, the DEEP data has 29 unique long stares, spread out over 10 unique RA and Dec positions. Therefore, the DEEP data set covers less total area on the sky. However, the DEEP data set is well-suited to the purposes of digital tracking detection of TNOs due to the greater number of images, greater exposure time per image, and repeat visits over two years.

Table 4.1: Fields, Nights, RA, Dec, and Number of Exposures for all of the nights in the DEEP B1 quadrant that is searched with `KBMOD`. RA and Dec correspond to the central RA and Dec of DECam for the first exposure in the long stare. Each exposure in each long stare is 120 seconds long.

Field	Night	RA	Dec	Number of Observations
B1a	2019-08-28	351.877200	-5.035722	102
B1a	2019-09-26	351.380700	-5.240305	95
B1a	2020-10-18	351.121033	-5.346222	91
B1a	2021-09-09	351.713325	-5.103250	65
B1b	2019-08-29	353.618654	-5.297667	101
B1b	2019-09-27	353.120700	-5.502861	97
B1b	2020-10-15	352.863196	-5.609639	100
B1b	2021-09-12	353.455079	-5.365278	70
B1b	2021-10-01	353.058450	-5.529917	81
B1c	2019-08-27	352.919908	-3.623028	103
B1c	2019-09-28	352.424117	-3.827972	96
B1c	2020-10-16	352.165621	-3.934944	96
B1c	2021-09-06	352.756529	-3.690555	53
B1c	2021-10-04	352.360275	-3.854361	101
B1d	2020-10-19	354.606779	-5.868861	99
B1d	2020-10-21	354.606525	-5.869916	99
B1d	2021-09-27	354.800983	-5.788139	89
B1e	2020-10-17	353.904871	-4.195444	96
B1e	2021-09-04	354.495283	-3.950028	59

Continued on next page

Field	Night	RA	Dec	Number of Observations
B1e	2021-10-06	354.098908	-4.115194	97
B1f	2020-10-20	353.205779	-2.521444	98
B1f	2021-09-03	353.795867	-2.276583	61
B1f	2021-10-02	353.400358	-2.441222	89
B1g	2021-09-28	356.545621	-6.043333	92
B1h	2021-09-08	356.235704	-4.206055	70
B1h	2021-10-05	355.839704	-4.371417	97
B1i	2021-09-05	355.531650	-2.533694	52
B1i	2021-10-03	355.136983	-2.698944	100
B1j	2021-09-30	354.436454	-1.026361	85

The DEEP survey strategy is designed to follow candidate TNOs over two or more years. As shown in Table 4.1, the B1 data discussed in this chapter extend from 2019 to 2021. In 2019, the DEEP survey took data on Fields B1a to B1c. In 2020, the DEEP survey took data on Fields B1a to B1f. In 2021, the DEEP survey took data on the entire footprint of the DEEP B1 quadrant, namely Fields B1a to B1j. This dispersal pattern was chosen so that a candidate TNO observed in (for example) Field B1a Night 2019-08-28 would be re-observed in the follow-up observations despite potentially leaving the B1a field. The survey strategy will be described in more detail in [124, 125]. The overall survey time baselines means that candidates that can be linked over the entire survey time baseline will have observational arcs extending for over two years (three oppositions). As of the time of writing, data are also scheduled to be taken in 2022, which may further extend the total time baseline of the fields.

Figure 4.1 shows specific nights of the DEEP B1 fields. The nights shown in Figure 4.1 are the first night observed in the 2021 observing run for each individual field. Over various nights, a given field can vary by up to 0.817 degrees from other observations of that field

with a median difference of 0.3947 degrees. These intra-field variations were designed in part to mitigate the effect of chip gaps and gaps between fields. The DEEP B1 fields are aligned with the invariable plane (the solid black line in Figure 4.1) of the Solar System. Within a particular long stare, however, the images are well-aligned. The median standard deviation in RA and Dec within a long stare is $\sigma_{\mathrm{RA}} = 0.629$ arcsec and $\sigma_{\mathrm{Dec}} = 0.070$ arcsec. As in Chapter 3, images are further aligned to a common WCS before running `KBMOD`.

4.3 Techniques

In this section, we describe how we apply the `KBMOD` pipeline to the B1 DEEP data [136, 119].

4.3.1 Image pre-processing

As in Chapter 3, we use the LSST Science Pipelines to process the DEEP B1 data [69]. For these data, we start with DECam raw image files, flat images, and bias images. All but two long stares have associated flats and biases from the same night. For the two exceptions, Field B1b Night 2020-10-15 and Field B1d Night 2020-10-19, we use flats and biases from time-adjacent Night 2020-10-16 and Night 2020-10-18 respectively. We then use the LSST Science Pipelines to perform instrument signal removal (ISR), photometric calibration, astrometric calibration, and image differencing. Photometric calibration is performed using Pan-STARRs data release 1 data [38] as a reference catalog, mapping the *VR* band to r_{p1}. Astrometric calibration is performed using GAIA dr2 data [79] as a reference catalog. The LSST Science Pipelines refer to these photometrically and astrometrically calibrated exposures as 'calexps'.

After calibration, we inject a population of synthetic TNOs into the calexps. We refer to these implanted synthetic objects as "fakes". The goal of this population is to test all possible bound TNO orbits, rather than to attempt to reproduce any real existing TNO population structure. To accomplish this, we use `DESTNOSIM`[2] to create a joint distribution from a KBO-like population, a moderately-excited population, and an isotropic distribution

[2]The software can be found here:`https://github.com/bernardinelli/DESTNOSIM`

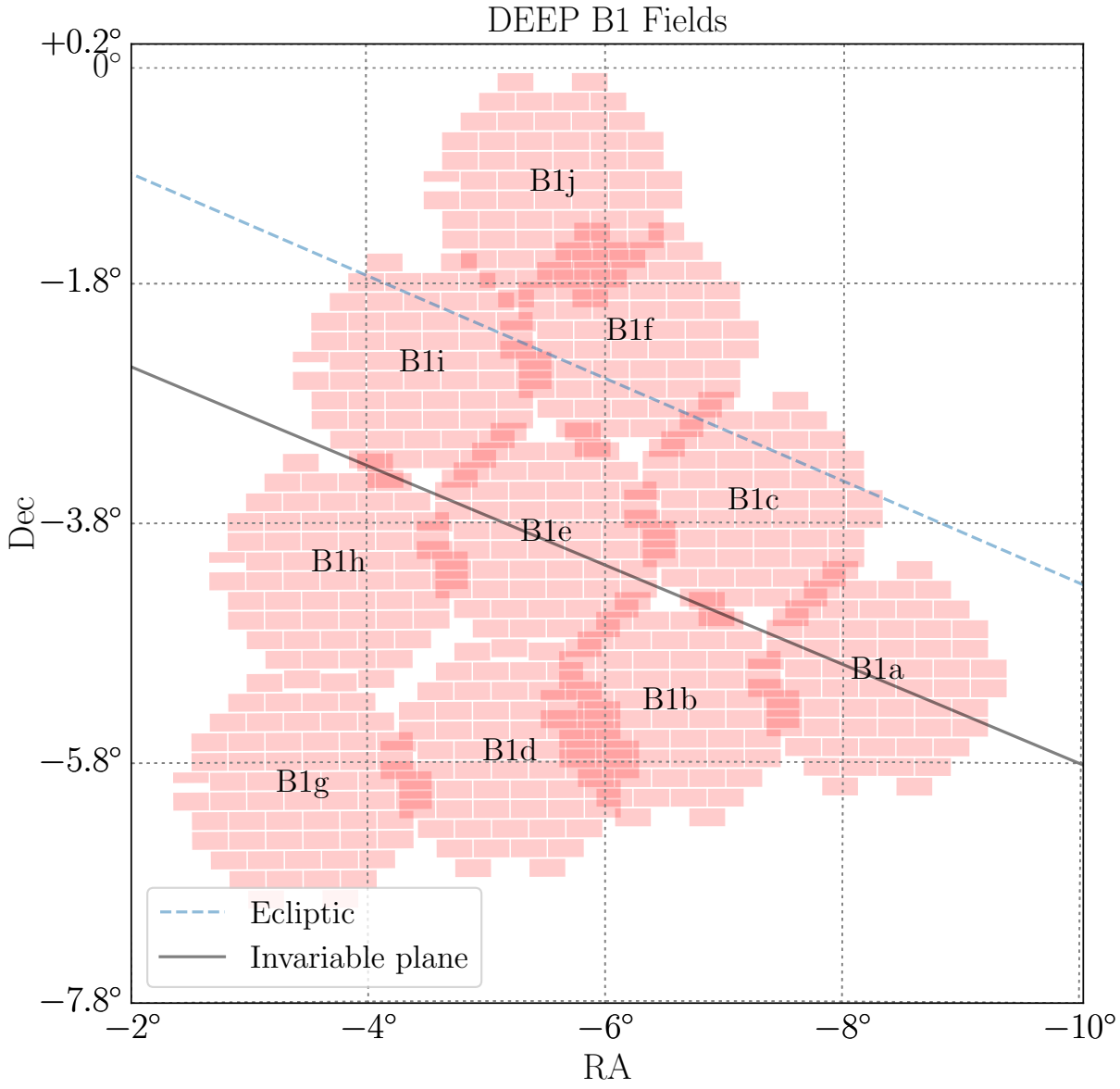

DEEP B1 Fields
+0.2°
0°
−1.8°
−3.8°
−5.8°
−7.8°
Dec
−2°
−4°
−6°
−8°
−10°
RA
B1j
B1f
B1i
B1c
B1e
B1h
B1a
B1b
B1d
B1g
Ecliptic
Invariable plane

Figure 4.1: An example of the DEEP B1 fields. The fields included here are the first observation of each field taken in 2021. For example, the B1b field shown here is Field B1b Night 2021-09-12. Each field is centered on the RA and Dec of the first exposure in the long stare. Different nights of the same field vary positionally as shown in Table 4.1. This is done in part to minimize the effect of chip gaps and gaps between fields. The 2019 observing run took data in Fields B1a to B1c. The 2020 observing run took data in Fields B1a to B1f. The 2021 observing run took data in all fields shown here (Fields B1a to B1j). This strategy helps ensure that TNOs detected in e.g. Field B1a Night 2019-08-28 will be re-observed over the next two years. In turn, this allows these objects to be linked and for accurate and precise orbits to be fit. The solid line shows the invariable plane while the dashed line shows the ecliptic.

with randomly-sampled x, y, and z velocities. The KBO-like population is constrained such that 30 au $< a <$ 80 au, $0.0 < e < 0.3$, and $0° < i < 60°$. The moderately-excited population is constrained such that 20 au $< a <$ 2000 au, $10 < q < 100$, and $0° < i < 90°$. For objects with large semi-major axis and pericenter distance ($a > 150$ au and $q > 30$ au), mean anomaly M is chosen such that $|M| < 30°$. This helps ensure that these objects are not placed at large heliocentric distances, where topocentric speeds become so low that detection is highly unlikely over the approximately 4 hour time baseline of DEEP. We place limits on the isotropic distribution such that the barycentric distance r_0 is distributed between 25 au and 1000 au. Half of the isotropic population is distributed logrithmically from 80 au $< r_0 <$ 1000 au and half is distributed uniformly from 25 au $< r_0 <$ 80 au. We propagate these objects to the epoch of each exposure, adding objects that intersected with the plane of a CCD to the synthetic catalog. This creates a catalog of 5737 unique fakes that each overlap with a CCD image between 1 and 9 times, inclusively. Figure 4.4 shows the distributions of osculating orbital elements for the fakes, taken from the first visit in each long stare, along with the associated distributions for the fakes recovered with `KBMOD`. Figure 4.5 shows the distribution of barycentric distances and topocentric speeds. See Section 4.3.3 for more detail.

We generate a wide range of possible apparent magnitudes (m_{VR}) for this distribution of fakes such that we cover the entire parameter space of interest for DEEP. We assign values for m_{VR} from two uniform distributions, $20 < m_{VR} < 24$ and $24 < m_{VR} < 28$, with 20% and 80% of the total objects in the brighter and fainter distributions respectively. The distribution of magnitudes, taken from the first visit in each long stare, is shown in Figure 4.5, along with the distribution of fakes recovered by `KBMOD`. See Section 4.3.3 for more detail. We propagate magnitudes with respect to changing geocentric distance. Half of the synthetic population included simulated lightcurves with periods between 2 and 100 hours and amplitudes between 0 and 0.5 mag. See [12] for more detail on the generation and distribution of this population, with respect both to osculating orbital elements and magnitudes.

After injecting synthetic TNOs, we generate coadded image differencing templates from each night of long stare data and subsequently subtract each calexp to create difference images. Chapter 3 describes in more detail the astrometric performance of the LSST Science Pipelines applied to DECam data as well as the image differencing technique. Notably for this data set, we use the default LSST Science Pipelines coaddition statistic of MEAN. Because we generate coadded templates from the images of each long stare, we are removing some flux from the TNOs, both real and synthetic, in the individual images. This removal will depend on the speed of the TNO and the time baseline of the specific long stare. This has the effect of reducing our limiting magnitude (reported here as m_{25}) and complicating the magnitude measurement for each candidate object. For detected candidate objects, we model and account for this flux loss, as described in Section 4.3.2. Future DEEP data releases will investigate the use of a non-default coaddition statistic such as MEDIAN or MEANCLIP.

Finally, we warp each CCD exposure to the WCS of the first CCD exposure in the long stare using the `lanczos4` method in the LSST Science Pipelines. This alignment ensures that, for example, pixel [234,1847] in CCD 10 corresponds to the same RA and Dec for every exposure in the long stare that will be searched with `KBMOD`. As discussed in Section 4.2, the long stares are already well-aligned, so this is a small (< 1 arcsec) adjustment in most cases.

4.3.2 Single-night detection and characterization

After generating calibrated, aligned difference images and injecting a large population of fakes, we run the digital tracking algorithm `KBMOD` on each stack of CCD exposures in each long stare. `KBMOD` is a GPU-accelerated pipeline that uses pre-convolved likelihood images (referred to as Ψ and Φ images) to estimate the likelihood that there is a source along a given trajectory. This likelihood is given by Equations 24, 25, and 26 of [136] and is a simple sum, division, and square-root to compute the SNR of an object along a trajectory.

This formalism allows the core GPU algorithm in `KBMOD` to evaluate more than 10^{10} candidate trajectories in a few minutes using consumer-grade GPUs [136]. For DEEP, a `KBMOD` search of a stack of $\mathcal{O}(100)$ images for a single CCD has a median runtime of 576s

using an NVIDIA A40 GPU. This includes loading the data from disk, running the core GPU algorithm, filtering false positives, and generating stamps. Our total runtime for the forward searches was about 282 hours. A "forward search" is a search in the direction of decreasing ecliptic longitude (i.e. the direction a TNO would be expected to move due to Earth's reflex motion). In this section, we define a "search" to mean a `KBMOD` processing run of a set of images from a single CCD (e.g. DECam CCD 20). Processing a given field and night, therefore, will require a 60 individual searches. With 60 searches for each of the 29 long stares considered here, we ran a total of 1740 searches with `KBMOD`.

Our grid choice for `KBMOD` as applied to DEEP is as follows. Velocity ranges from 90 px/day to 400 px/day (1.0125 arcsec/hr to 4.5 arcsec/hr), with 50 uniform steps. Angles are centered on the direction of decreasing ecliptic longitude, with offsets of $\pm 45°$, with 30 uniform steps. These values were chosen so that the maximum separation between neighboring trajectories would be less than about 2 PSF FWHM over a 4 hour time baseline[3]. The bounds in velocity and angle were chosen to adequately cover the Kuiper Belt. Of the 6445 single-night observations of implanted fakes where the barycentric distance is between 30 au and 60 au, only 9 have topocentric speeds outside of the 150 px/day and 400 px/day bounds. Recovery of objects given this grid selection is discussed in Section 4.3.3. We set the minimum sum likelihood threshold for a candidate trajectory to be considered to $\sum LH > 7$. Although we refer to this threshold as a "minimum sum likelihood", both in this work and in [119], it is analogous to an approximate SNR threshold.

Due to the number of false positives at very slow speeds[4], we discarded candidate trajectories with velocities between 90 px/day and 130 px/day prior to human vetting. For

[3]This means that a given TNO trajectory within the searched angle and velocity ranges should only be about 1 PSF FWHM away from the ending position of a search trajectory, even if it falls exactly between two neighboring grid spacing. Even in this worst-case scenario, the starting position is spaced every pixel (i.e. every 0.263", which is much closer together than a PSFW FWHM in these data). Furthermore, the joint-fit has no grid spacing dependency and does a good job of improving the astrometry for a given trajectory (see Figure 2.6). We therefore consider this grid spacing to be sufficient.

[4]The increased number of false positives is often due to the increasing presence of static image artifacts and static sources within the searched trajectory. The slower trajectories imply lesser total motion across an image. This means that static sources are present in an increasing fraction of a time-series of images.

candidate trajectories with velocities between 130 px/day and 150 px/day, we increased the likelihood threshold from $\sum LH > 7$ to $\sum LH > 10^5$. For searches which had more than 50 candidates per search with speed > 150 px/day and $\sum LH > 7$, we increase the likelihood threshold to $\sum LH > 10$. This "overflow" occurred in 51/1680 of the forward searches.

In order to characterize our false positive rate, we also run `KBMOD` searches in the direction of increasing ecliptic longitude, which is opposite from the direction of a TNO due to the reflect motion of the Earth. We refer to these searches as "reverse searches". These reverse searches are used to select the appropriate SNR cutoff limit for a candidate trajectory to be labelled "good". This is described in more detail in Section 4.4.

Finally, we apply the same CNN from [119], discarding candidates with a CNN-assigned probability of "good" of less than 0.5. The effectiveness of this CNN, both at accepting good candidates and rejecting false positives, is described in more detail in Section 2.4. We note that in practice the CNN threshold is merely a user-defined cutoff, rather than any kind of robust statistical likelihood. For the candidate trajectories that remain after the aforementioned cuts, we assign a randomized candidate Id and generate images for human vetting. This is done with the combined results from both the forward and the reverse searches. We show an example of the images used for human vetting in Figure 4.2. The rate at which we mislabel reverse search candidates as "good" is about 7.2% for candidates with $7 < \sum LH < 10$, 1.2% for candidates with $10 < \sum LH < 12.5$, and generally $< 1\%$ for candidates with $12.5 < \sum LH$.

After human vetting, we apply the joint-fit analysis tool described in Section 2.6 to improve the quality of the astrometric fit and to characterize our errors in starting and ending position, flux, and magnitude. We rescale each stamp for each candidate object to a common magnitude zero point of $m_{VR,0} = 31$. We use the same formalism as `KBMOD` for flux and likelihood estimation, along with the `scipy.optimize.minimize` routine, to find

[5]This special treatment between 130 px/day and 150 px/day was done in an attempt to detect objects at as slow of speeds as possible. However, as we will show in Figure 4.5, our recovery efficiencies drop quickly for speeds slower than 150 px/day.


LH: 24.69 | FLUX: 176.22 | X: 345.00 | Y: 1235.00 | V: 183.00
sum
32.20
median
26.55

Figure 4.2: This is an implanted fake recovered with KBMOD with an injected magnitude $m_{VR} = 24.9$ and a barycentric distance of 60 au. These images are used for human vetting. The title includes information from the KBMOD search, including $\sum LH$, flux, starting x and y pixel location, and speed. We include the speed, rather than the x and y velocities, because x and y velocities easily allow a human to differentiate between the forward searches and the reverse searches. The two postage stamps in the upper left are sum and median coadds, as well as an approximate SNR estimated from these postage stamps. The graph in the upper right shows an approximate flux lightcurve. The postage stamps included below are cutouts from individual images in the long stare.

the topocentric best-fit line that minimizes the negative log likelihood. We use the "L-BFGS-B" minimization method and a 3-point jacobian. The best-fit line is parameterized in starting and ending positions, which are each bounded to be no more than 10 pixels from the center. We derive astrometric uncertainties for the starting and ending locations from the square-root of the diagonals of the inverse hessian matrix returned by the minimization method.

Finally, we measure the magnitude of each candidate object. When fitting best-fit fluxes, flux error, and magnitudes, we adjust the PSF of each image to model the effect of the "negative well" around TNOs in the difference image. This negative well is caused by generating coaddition templates from the images used in the `KBMOD` search itself. In the coadded template, the moving TNOs generate a streak as the TNO moves over the time baseline of a long stare. When subtracting the coadded template from a calexp science image, this creates a negative "streak" that we are calling the negative well. We model this effect by generating an effective PSF p_{eff} which includes this negative well. This negative well is modelled for a given exposure i by subtracting the LSST PSF p of each long stare exposure j scaled by the number of images N in the long stare. There is an implicit assumption of constant brightness in this method. For exposure i, $p_{i,\text{eff}}$ is given by the following equation.

$$p_{i,\text{eff}}(x_i, y_i) = p_i(x_i, y_i) - \sum_{j=0}^{N} p_j(x_j, y_j)/N, \tag{4.1}$$

where $p_i(x_i, y_i)$ is PSF p_i of an exposure i centered at the joint-fit predicted object location (x_i, y_i). Similarly, $p_j(x_j, y_j)$ is PSF of image j centered at the joint-fit predicted object position (x_j, y_j). To clarify, (x_j, y_j) is the predicted pixel position of the object in exposure j. This propagation works in pixel space because images in the long stare are warped to a common WCS. This subtraction occurs along the trajectory the candidate object would take through the coadded template, thereby modelling the effect of the lost flux due to the moving object's presence in the coadded template. Each exposure has a different p_{eff} because of the dependence of the location of the object in each image j. In short, we are subtracting the "streak" that the object would make in the coadded template from the image PSF.

Once we have p_{eff} for each image, we then measure the flux in each individual cutout in a single candidate trajectory. With these individual flux measurements, we use the `astropy.stats.sigma_clip` routine with "stdfunc='mad_std'" to iteratively reject outliers more than 3σ from the median. 3σ is the default value used by `astropy.stats.sigma_clip`. As used here, the function iterates until no additional points are removed from the distribution, with a cutoff at 100 total iterations. For the cutouts that are not clipped, we then measure the variance-weighted flux and flux uncertainty as shown in the equations below.

$$f = \frac{\sum_i \Psi_i}{\sum_i \Phi_i} \tag{4.2}$$

$$\sigma_f = \frac{1}{\sum_i \Phi_i}, \tag{4.3}$$

where f is the best-fit flux, σ_f is the standard deviation in flux, and Ψ_i and Φ_i are defined as in [136]. However, the sum in Ψ_i and Φ_i only uses data from the stamps that are not rejected in sigma clipping. Finally, we add an additional flux error term in quadrature to the flux uncertainties such that the sigma-clipped standard deviation of the distribution of the residuals of the implanted and measured fluxes for the simulated objects is 1.00. This helps to ensure we do not underestimate our uncertainties in flux. Finally, we convert these fluxes into m_{VR} values using Equation 1.1.

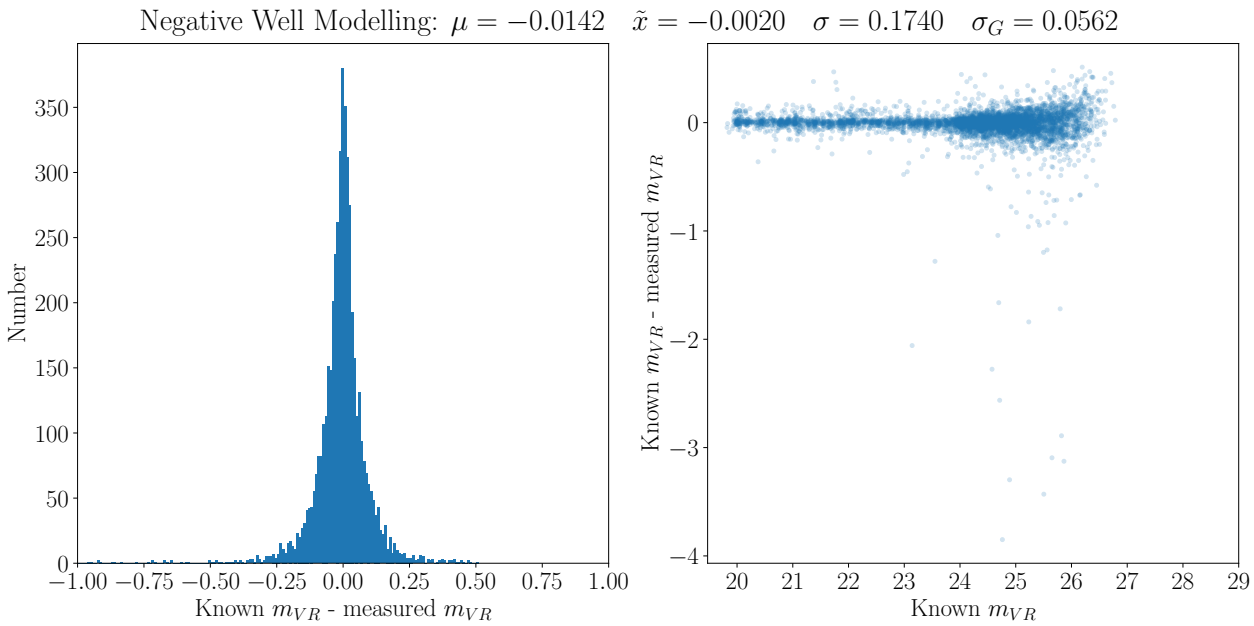
4.3.3 Population of fakes

With the population of fakes, we are able to robustly describe the performance of `KBMOD`, CNN filtering, and the joint-fit. CNN filtering and the astrometric results of the joint-fit are described in Sections 2.4 and 2.6 respectively. Because the method of modelling the negative well in the difference image in order to fit magnitudes (described in Section 4.3.2) is unique to this data set, we characterize the technique in this chapter instead of Chapter 2.

First, we must determine which detected candidates are fakes that were artificially added to the data. After the human vetting described above, we cross-match the starting position

of all candidates labelled "good" with the catalog of fakes using the measured joint-fit RA and Dec and the catalog RA and Dec. Here, we will refer to the catalog values as "known" values and the observed values from either `KBMOD` or the joint-fit as "measured" values. We attribute a candidate trajectory with a fake if the distance between the measured RA and Dec and the known RA and Dec is less than 5 arcsec. Out of the 350 reverse search candidate trajectories erroneously labelled "good", only 3 are are linked to fakes, implying a false association rate of about than 0.86%. In Section 2.4, which also removes the overflow criteria, we find the false association rate to be as low as 0.07% when using a cross-matching distance of 2.5 arcsec. However, we find this higher threshold of 5 arcsec to be appropriate in this analysis, because we would prefer to miss a small percent of real candidate objects (due to false association) than claim a detection of a real candidate object that is actually a fake. 5 arcsec (about 19 pixels) is nearly the size of the 21 x 21 postage stamp, meaning that the actual starting position of a fake would have to be nearly off of the postage stamp for it not to be associated with its corresponding catalog value. Furthermore, from Figure 2.6 which also uses a 5 arcsec association distance, we find that the standard deviation of the observed RA and Dec residuals are about $\sigma_{\mathrm{RA}} = 0.1873$ and $\sigma_{\mathrm{Dec}} = 0.2047$. In the case of minimizing the RA and Dec distance from the edge of a 5 arcsec circle, we still guarantee at least a $5 \times 1/\sqrt{2}$ arcsec association distance in either RA or Dec. Thus, for us to mis-classify a fake as a real candidate, it would have to be at least a 17σ outlier in RA or Dec residual. We acknowledge the potential for non-Gaussian outliers, for which σ_{RA} and σ_{Dec} are not an appropriate metric, but any further increase in the fake association distance of 5 arcsec would correspondingly increase the false association probability from the current 0.86%. We therefore consider this an acceptable fake association distance. To avoid any erroneous associations in our characterization of the recovered fakes, we only use cross-matched candidates from the forward searches in the following analysis.

Using the fakes, we are able to characterize the performance of our magnitude fitting technique both with and without the modelling of the negative well. We repeat the flux estimation method described above, including iterative sigma clipping, using the PSF pro-


Negative Well Modelling: $\mu = -0.0142$ $\tilde{x} = -0.0020$ $\sigma = 0.1740$ $\sigma_G = 0.0562$
Number
Known m_{VR} - measured m_{VR}
Known m_{VR} - measured m_{VR}
Known m_{VR}

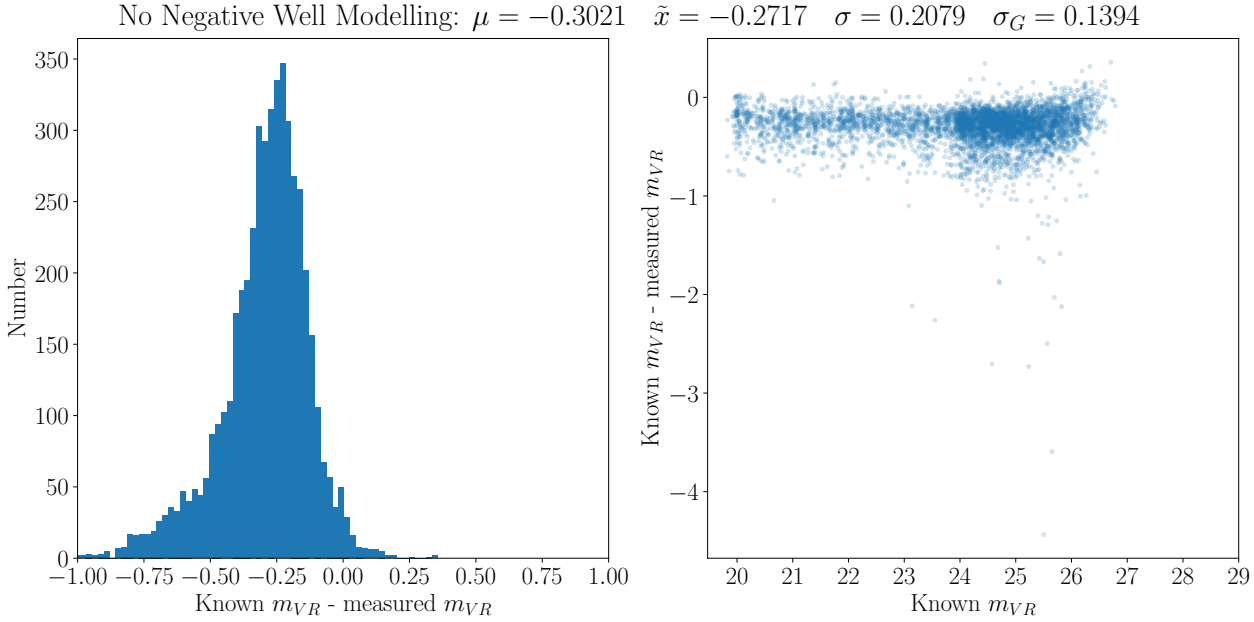

No Negative Well Modelling: $\mu = -0.3021$ $\tilde{x} = -0.2717$ $\sigma = 0.2079$ $\sigma_G = 0.1394$
Number
Known m_{VR} - measured m_{VR}
Known m_{VR} - measured m_{VR}
Known m_{VR}

Figure 4.3: The known and measured magnitudes of the fakes in the DEEP data. The top row shows the results of magnitude fitting including a model for the negative wells that occur in the images due to the coaddition method used in image differencing (see Section 4.3.2 for more detail). The bottom row shows the results of magnitude fitting without modelling the negative wells. In each row, the left plot shows a histogram of the residuals of the known m_{VR} of the fakes and the measured m_{VR} of the fakes. For visibility, the limits on the x axis were set to [-1,1].. Bins are set using the `numpy` "bins='auto'" method. In each row, the right plot shows a scatter plot of the magnitude residuals (known m_{VR} − measured m_{VR}) vs the known m_{VR}. The title of each row includes the mean μ, median $\tilde{x}$, standard deviation σ, and σ_G value for the histogram (left) in each row. As shown here, including a model for the negative well reduces the systematic offset of the residual (shown by μ and $\tilde{x}$) as well as decreases both the overall scatter (shown by σ) and the scatter of the core distribution (shown by σ_G).

vided by the LSST Science Pipelines without any adjustment to account for the negative well in each image. By analyzing the residual of the known m_{VR} and the measured m_{VR} both with and without a model for the negative well, we are able to characterize the improvement offered by modelling the negative well. As shown in Figure 4.3, modelling the negative well in the data shifts the mean of the residual from $\mu = -0.3021$ to $\mu = -0.0142$, effectively fixing a 0.3 magnitude systematic offset. Furthermore, both the standard deviation σ and σ_G (defined in Section 2.1) decrease significantly. Given the non-trivial number of outliers in each data set, σ_G is the better choice for estimating the characteristic scatter of the core distribution of the residuals. This makes it noteworthy that σ_G decreases from $\sigma_G = 0.1394$ to $\sigma_G = 0.0562$, more than a factor of two reduction. Finally, as a qualitative observation, the symmetry of the distribution is much improved by modelling the negative well.

The photometric scatter at the bright end does not quite perform as expected. For photometric scatter that is proportional to the inverse of SNR, one would expect the photometric scatter at (for example) $m_{VR} = 20$ to be significantly less than that at $m_{VR} = 23$, but this does not appear to be the case. This is likely related to the astrometric error that remains even after the joint-fit (see Figure 2.6). The astrometric scatter is roughly constant along the bright end. To characterize this, we can consider the distance residuals of the implanted fakes with SNR > 4.1. For $20 < m_{VR} < 21.5$, the standard deviation of the distance residuals is about 0.068 arcsec. For $21.5 < m_{VR} < 23$, the standard deviation of the distance residuals is a similar 0.87 arcsec. Then, for $23 < m_{VR} < 24.5$, the standard deviation of the distance residuals increases significantly to 0.152 arcsec. Thus, astrometric scatter is fairly consistent between $20 < m_{VR} < 23$, which likely contributes to characteristics of the photometric scatter in this regime (i.e. the bright end of our data).

We use the fakes to visualize our recovered distribution of fakes versus the total distribution of fakes implanted into the images in the DEEP survey as a function of osculating orbital elements, barycentric distance, limiting magnitude, and topocentric speed. Figure 4.4 shows the distributions of the recovered fakes and the entire population of fakes as a function of the osculating elements. This figure is particularly noteworthy because it demon-

strates the difficulty of mapping topocentric recovery efficiencies to osculating elements. For example, there appears to be an unexpected loss of efficiency in eccentricity in the range of $0.4 < e < 0.75$ and in inclination in the range of $140° < i < 175°$. Similarly, time of pericenter passage appears to have a strong and noticeable cutoff in recovery efficiency around $t_p < -350$ yr. However, as we will demonstrate, these effects are largely a function of the interactions between the osculating elements, rather than a dependence solely on eccentricity or solely on inclination. Regardless, we find this figure critical to include for a number of reasons. First, it demonstrates the distribution of all of the implanted fakes. Figure 4.4 shows that the population of fakes implanted into the images cover a wide range of parameter space which we can then narrow down to quantify specific features of our efficiency function. Second, osculating elements can be used to identify populations of small bodies in the Solar System (e.g. eTNOs or the Kuiper Belt kernel) and are commonly among the first parameters used to describe detected Solar System objects. Third, as mentioned, these plots serve as motivation for the upcoming cuts and deeper analyses of our survey efficiency.

Instead of parameterizing our distributions based on osculating orbital elements, we can plot both the distributions and the direct recovery efficiencies based on parameters that we expect to have an impact on detectablility, including barycentric distance, topocentric speed, and m_{VR}. As shown in the first row of Figure 4.5, barycentric distance has an obvious effect on detectability, with nearly zero fakes detected at a barycentric distance r_0 of $r_0 > 100$ au. Because TNO topocentric motion is driven primarily by the reflex motion of the Earth, there is a strong causal relationship between barycentric distance and topocentric speed for these data. We therefore also plot the distributions of topocentric speed. 175 of 14112 total observations of the fakes go off of the CCD footprint prior to the end of the long stare. These fakes are excluded from the distributions of topocentric speed discussed here. As shown in the second row of Figure 4.5, the recovery efficiency for the fakes is roughly constant in the primary region searched with `KBMOD` (150 px/day $< v_{\rm topo} <$ 400 px/day). There are fakes with $v_{\rm topo} > 400$ px/day, and recovery efficiency understandably drops above this limit (since we do not search speeds above this limit with `KBMOD`). In future DEEP analysis, we will

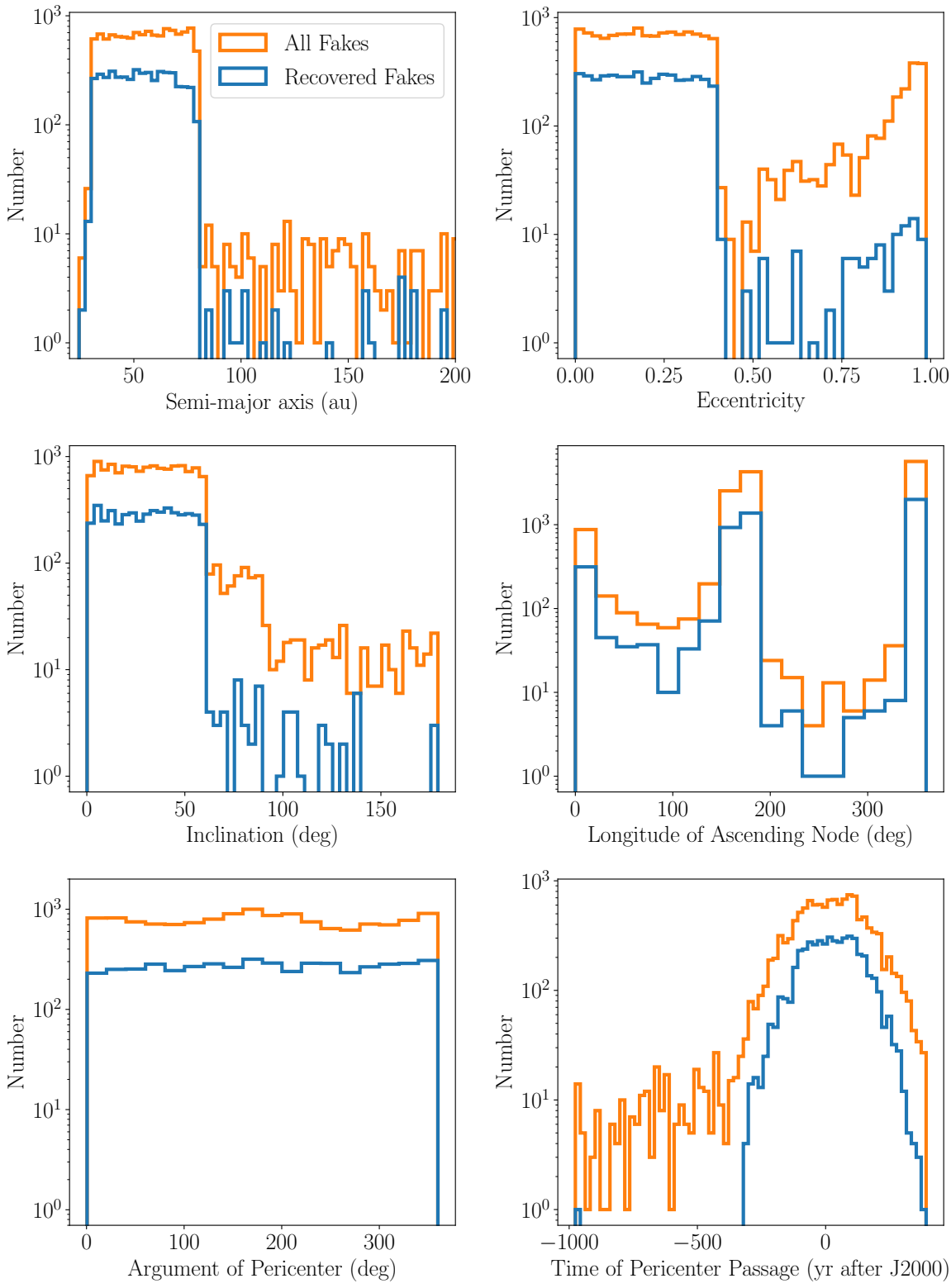

All Fakes
Recovered Fakes
Number
Semi-major axis (au)
Eccentricity
Inclination (deg)
Longitude of Ascending Node (deg)
Argument of Pericenter (deg)
Time of Pericenter Passage (yr after J2000)

Figure 4.4: Histograms of both the recovered fakes (blue) and all of the fakes (orange) implanted into the images in the DEEP survey. The histogram of semi-major axis (top left) is bounded between 20 au and 200 au for visibility. Bins are selected with the "bins='auto'" method in `numpy`.

increase the upper speed limit to $v_{\text{topo}} = 500$ px/day. However, this increase is not critical in order to cover the Kuiper Belt. Out of 6445 total observations of fakes with topocentric distance d such that 30 au $< d <$ 60 au, only 9 observations fall outside of the bounds of 150 px/day $< v_{\text{topo}} <$ 400 px/day, indicating that we efficiently cover the Kuiper Belt. For candidates with 130 px/day $< v_{\text{topo}} <$ 150 px/day, we made an attempt to recover these candidates by increasing the $\sum LH$ threshold from $\sum LH > 7$ to $\sum LH > 10$, but it is clear from Figure 4.5 that efficiency drops off quickly below $v_{\text{topo}} = 150$ px/day. Finally, as expected, efficiency as a function of m_{VR} is approximately constant until it quickly tapers off. With our understanding of the effects of topocentric speed and m_{VR} on discovery efficiency, we can now treat each parameter separately to get a more detailed picture of each.

At this stage of analysis, the efficiencies reported in Figure 4.5 are useful only for their characteristic shape, including the points at which efficiency drops and the speed with which it drops. Their amplitudes are not particularly informative because they depend entirely on the choice of the underlying distribution of fakes. Consider, for example, the case of the recovery efficiency in topocentric speed. One could potentially add 10,000 new fakes with a $m_{VR} > 30$, using the same osculating orbital element distributions as before, and none of them would be detected. However, the amplitude of the recovery efficiency in topocentric speed would be driven towards zero, even though the shape of the efficiency distribution would remain approximately the same. Similarly, in the case of magnitude, one could add 10,000 new fakes with $r_0 > 200$ au, and nearly zero would be detected, but the magnitude efficiency would be driven towards zero. Therefore, for the amplitude of a given efficiency distribution to be useful, we have to isolate the parameter in question. For example, see the argument of pericenter subplot in Figure 4.4 for a visualization of both the number of fakes implanted and the number of fakes detected.

To better characterize `KBMOD`'s efficiency as a function of magnitude, we cut on topocentric speed such that 150 px/day $< v_{\text{topo}} <$ 400 px/day. As mentioned, these speeds effectively cover the region of objects with 30 au $< d <$ 60 au, which includes the Kuiper Belt. These are the topocentric speed limits of the grid searched by `KBMOD` when using $\sum LH > 7$, as

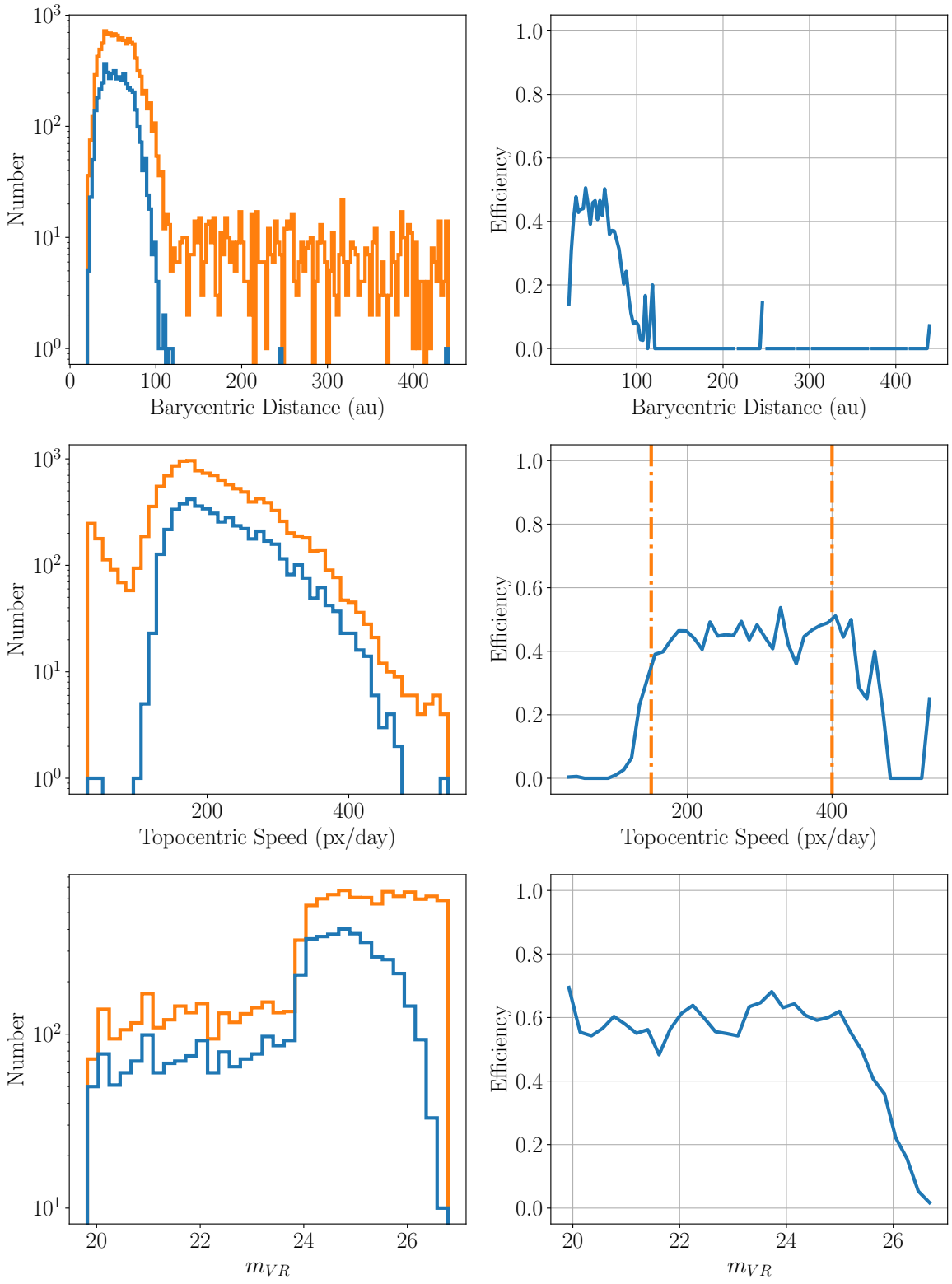

Number
Barycentric Distance (au)
Efficiency
Barycentric Distance (au)
Number
Topocentric Speed (px/day)
Efficiency
Topocentric Speed (px/day)
Number
m_{VR}
Efficiency
m_{VR}

Figure 4.5: Visualization of both the recovered fakes (blue) and all of the fakes (orange) implanted into the images in the DEEP survey. The left column shows a histogram of two distributions of fakes. Bins are selected with the "bins='auto'" method in `numpy`. The right column shows the recovery efficiency determined by the fraction of recovered fakes versus all implanted fakes in each histogram bin from the right column. In the "Topocentric Speed" efficiency plot, the two orange dash-dot vertical lines correspond to the limits of topocentric speed v_{topo} for the vetting candidates output by `KBMOD`. This limit of $150 < v_{\text{topo}} < 400$ px/day corresponds to the parameter space which used a cutoff of $\sum LH > 7$.

described in Section 4.3.2. As discussed above, this cut is justified in part by the stable recovery efficiency within these bounds. Following the approach of [67], we can then fit a probability $p(m|m_{25}, c, k_1, k_2)$ of observing an object with magnitude m using the function

$$p(m|m_{25}, c, k_1, k_2) = \frac{c}{[1 + \exp(k_1(m - m_{25}))][1 + \exp(k_2(m - m_{25}))]}, \tag{4.4}$$

where m_{25} is the magnitude at which the efficiency has dropped to 0.25 of the maximum efficiency c. k_1 and k_2 are sharpness parameters. Equation 4.4 is equivalent to Equation 1 of [67], which is written in terms of hyperbolic tangent functions instead of exponential functions.

We show the results of the best-fit values of Equation 4.4 in Figure 4.6. The primary blue line shows the best-fit function, which has values of $m_{25} = 26.22$, $c = 0.7784$, $k_1 = 2.30$, and $k_2 = 4.54$. The orange data points in Figure 4.6 show the approximate efficiency in each bin in m_{VR}, along with $1/\sqrt{N}$ Poisson errors, where N is the total number of fakes in each bin. These points are generally consistent with the best-fit function to within the Poisson error. For visualization purposes, we also include the approximate single exposure m_{25}, shown by the green dotted line in Figure 4.6. This line is generated using the best-fit $m_{25} = 26.22$ and assuming a $\sqrt{N}$ increase in flux for N images due to digital tracking. We set $N = 87.38$, which is the mean number of images per long stare. We then compute the position of the green line using the simplified assumption that $m_{25,\text{single}} = m_{25} - 5/2 \log_{10} \sqrt{N}$, where $m_{25,\text{single}}$ is the magnitude at which we draw the green line. This line approximates the increased depth yielded by digital tracking methods.

Figure 4.6 also shows a fit to data from single long stares, instead of the total analyzed DEEP B1 quadrant data. The best-fit functional forms for each long stare are shown as transparent blue lines in Figure 4.6. Although the net efficiency is the primary distribution of interest, the individual distributions give an idea of the varying quality of different fields and nights. Similarly to the main dotted green line, the transparent green lines correspond to the approximate $m_{25,\text{single}}$ considering only the data from an individual long stare.

We can visualize the effect of both the number of images and the image quality per

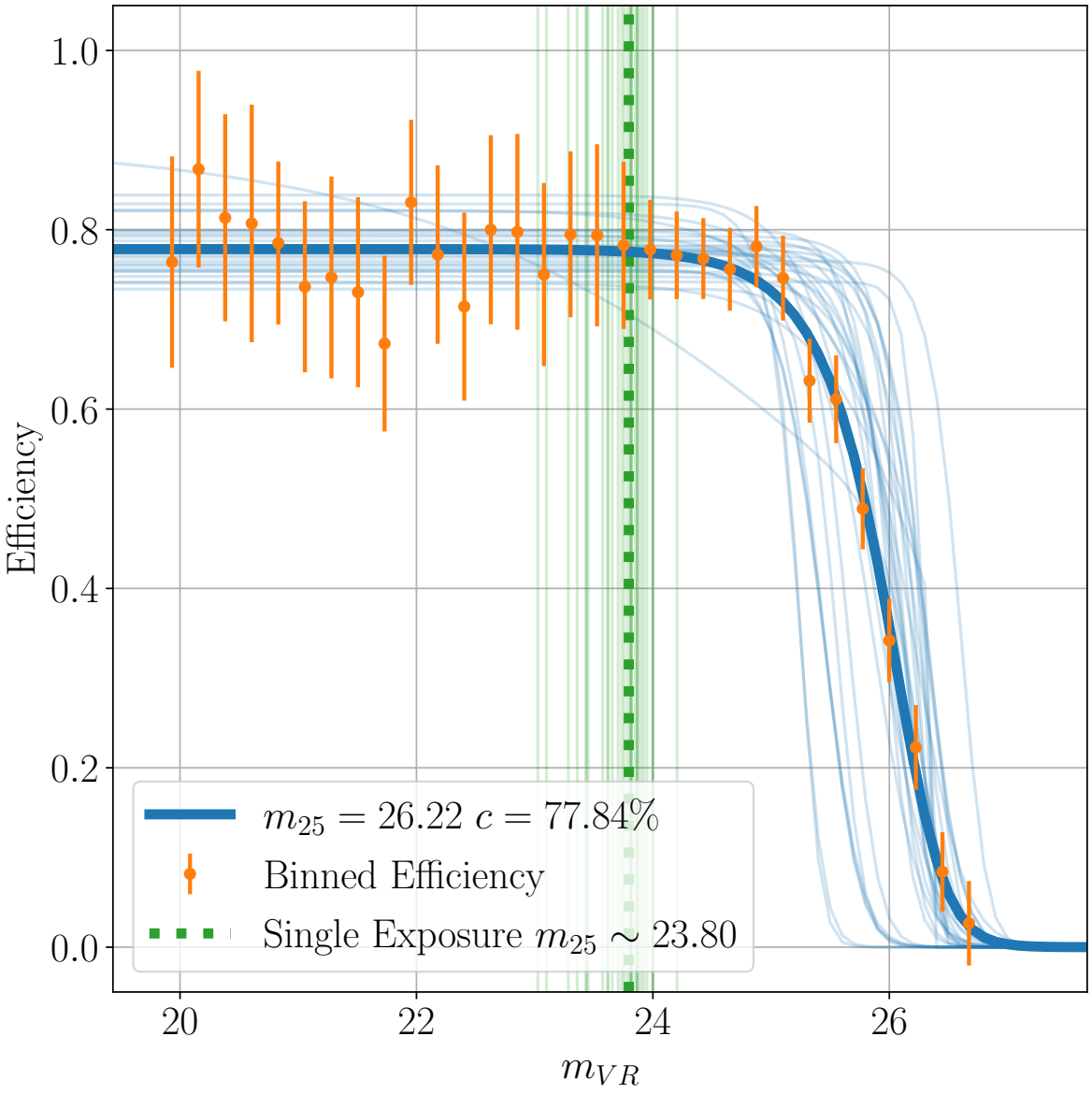

1.0
0.8
0.6
0.4
0.2
0.0
Efficiency
20
22
24
26
m_{VR}
$m_{25} = 26.22$ $c = 77.84\%$
Binned Efficiency
Single Exposure $m_{25} \sim 23.80$

Figure 4.6: Recovery efficiency of the fakes in DEEP as a function of m_{VR}. Only fakes with 150 px/day $< v_{\rm topo} <$ 400 px/day are included. The orange data points are the binned fraction of recovered fakes divided by all implanted fakes. Bins are set with the `numpy` option "bins='auto'". Error bars are $1/\sqrt{N}$ Poisson errors. The main blue line is the best-fit functional form to all data. m_{25} is the magnitude at which efficiency is equal to 0.25 of the maximum efficiency c. The primary dotted green line corresponds to the approximate m_{25} for single exposure data instead of digital tracking. The transparent thin blue lines correspond to the best-fit functional forms considering only individual long stare data, rather than the entire DEEP B1 quadrant data. Similarly, the transparent thin green lines correspond to the approximate m_{25} for single exposure data using only data from individual long stares. For these lines, N is set to the number of images in the corresponding long stare.

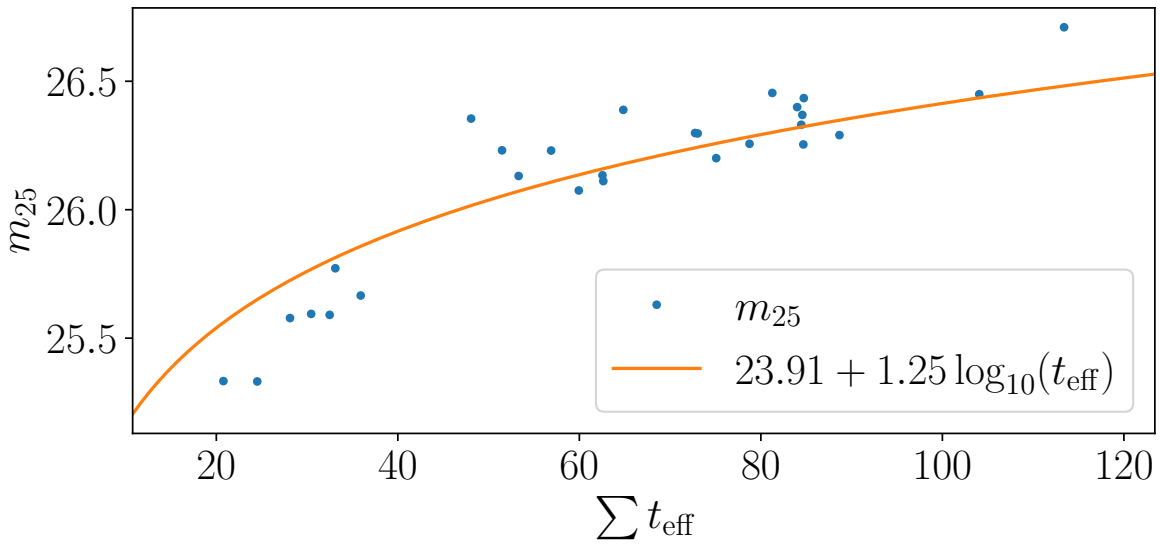


Figure 4.7: m_{25} versus $\sum t_{\rm eff}$ for each long stare. The orange line is the line following $m_0 + 1.25 \log \sum t_{\rm eff}$, with a best-fit value of $m_0 = 23.91$. $\sum t_{\rm eff}$ was not available for Field B1a Night 20210909, and is therefore not included in this plot.

individual long stare on each long stare m_{25} value. We do this using the parameter $t_{\rm eff}$ defined in [90]. In short, $t_{\rm eff}$ is a parameter that quantifies the quality of an image. $t_{\rm eff} = 1$ corresponds to an image of fiducial quality (defined in [90]). In this case, higher $t_{\rm eff}$ corresponds to a better image. We can therefore compare the $\sum t_{\rm eff}$ of each long stare to the best-fit m_{25} for that long stare. We show the result of this comparison in Figure 4.7. As expected, there is a clear correlation between m_{25} and $\sum t_{\rm eff}$, approximately following the expected relationship of $m_0 + 1.25 \log \sum t_{\rm eff}$, where the best-fit value of m_0 is 23.91.

To further support our use of Equation 4.4, we test two other functional forms for $p(m)$ and compare the Bayesian Information Criteria [116]. The Bayesian Information Criteria (BIC) is a metric used to determine which model out of a set of models should be preferred. Because adding more components to a model may yield a better fit simply because it has more components, the BIC adds a penalization term for the number of model parameters, as shown in the following equation

$$\mathrm{BIC} = k \ln N - 2 \ln L, \tag{4.5}$$

where k is the number of model parameters (4 in the case of Equation 4.4), N is the number of data points, and L is the maximum likelihood of the model given the data. In general, the model with the lowest BIC is preferred.

To test alternative models, we fit a version of Equation 4.4 with a single exponential (following the approach of [13]) as well a triple exponential, which is included primarily as a higher-dimensional alternative to Equation 4.4. These have the form

$$p(m|m_{50}, c, k) = \frac{c}{[1 + \exp(k_1(m - m_{50}))]} \tag{4.6}$$

$$p(m|m_{12.5}, c, k_1, k_2, k_3) = \frac{c}{\prod_{i=1}^{i=3}[1 + \exp(k_i(m - m_{12.5}))]}. \tag{4.7}$$

Of the three functional forms for $p(m)$ presented here, Equation 4.4 has the lowest BIC of BIC = 8055.14, while Equation 4.6 and Equation 4.7 have BIC = 8071.99 and BIC = 8061.12 respectively. The difference between two BIC values is an approximation of minus one half times the logarithm of the Bayes Factor (i.e. a constant multiple of the logarithm of the

Bayes Factor). (See [71] for a discussion). Therefore, among a set of models (model 1, model 2, and model 3), we can estimate the odds that a given model fits the data best. If we call model 1 the model with the lowest BIC, then the odds O that model i best fits the data is approximately

$$O = \exp\left[(\mathrm{BIC}_1 - \mathrm{BIC}_i)/2\right], \tag{4.8}$$

where BIC_1 is the BIC of model 1 and BIC_i is the BIC of model i. The odds against Equation 4.6 fitting the data better than Equation 4.4 are approximately 4570:1 and the odds against Equation 4.7 fitting the data better than Equation 4.4 are approximately 20:1. We have therefore confirmed that the model given by Equation 4.4 is likely the best model to report of the three considered here.

However, we stress that the BIC is an approximate metric and that the choice of model among these three models does not qualitatively alter our peak efficiency. Among Equations 4.6, 4.4, and 4.7, peak efficiency is 0.7631, 0.7784, and 0.7749 respectively. Similarly, Equation 4.6 and Equation 4.4 predict similar drop-off points. Equation 4.6 fits a value for m_{50} (the magnitude at which the efficiency has dropped to 0.50 of the maximum efficiency) of $m_{50} = 25.93$. Converting the best-fit m_{25} from Equation 4.4 to an m_{50} value, we get a converted m_{50} of 25.96. Our results are therefore only weakly sensitive to the choice between Equation 4.4 and Equation 4.6.

Having established the justification and reasoning for making cuts to the implanted distributions of fakes, we can revisit the question of there being any particular relationship between recovery efficiency and osculating elements. As mentioned before, the three elements in Figure 4.4 that appear to have some causal relationship with recovery efficiency are eccentricity, inclination, and time of pericenter passage. As above, we cut distribution of implanted and recovered fakes on topocentric speed such that 150 px/day $< v_{\mathrm{topo}} <$ 400 px/day. Using the magnitude efficiency function derived for Figure 4.6, we compute the magnitude at which our efficiency has dropped to 90% of the maximum recovery efficiency. We find this point to be $m_{90} = 25.22$. We then cut the distribution of implanted and recovered fakes on magnitude such that $m < m_{90}$.

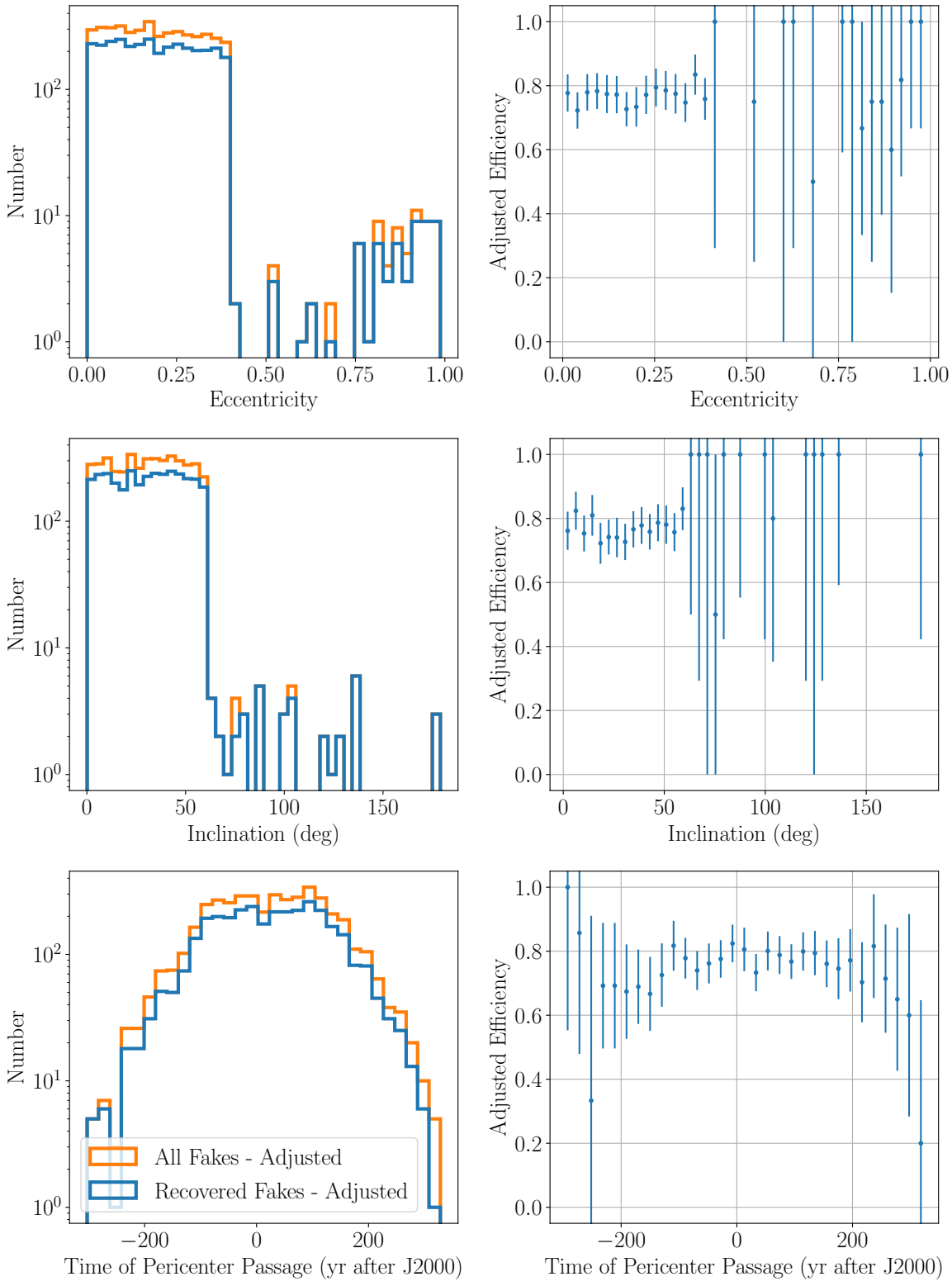

Number
Eccentricity
Adjusted Efficiency
Eccentricity
Number
Inclination (deg)
Adjusted Efficiency
Inclination (deg)
Number
All Fakes - Adjusted
Recovered Fakes - Adjusted
Time of Pericenter Passage (yr after J2000)
Adjusted Efficiency
Time of Pericenter Passage (yr after J2000)

Figure 4.8: Visualization of both the recovered fakes (blue) and all of the fakes (orange) implanted into the images in the DEEP survey after applying a cut on topocentric speed and magnitude. In this plot, we only consider fakes with 150 px/day $< v_{\text{topo}} <$ 400 px/day and $m_{VR} < 25.22$. The left column shows a histogram of two distributions of fakes. Bins are selected with the "bins='auto'" method in `numpy`. The right column shows the recovery efficiency determined by the fraction of recovered fakes versus all implanted fakes in each histogram bin from the right column. Error bars are $1/\sqrt{N}$ Poisson errors for each bin.

We visualize the distribution of implanted fakes and recovered fakes after the cuts have been applied. Specifically, we show these distributions with respect to eccentricity, inclination, and time of pericenter passage in Figure 4.8. From this figure, we see that any potential loss in efficiency from $0.4 < e < 0.75$ is merely due to the fact that there are few to no fakes in the implanted distribution with these eccentricities and the above limitations on $v_{\rm topo}$ and m_{VR}. Figure 4.8 shows a similar effect in the range of $140^\circ < i < 175^\circ$ and $t_p < -350$ yr. After the cuts are applied, the adjusted efficiency in eccentricity, inclination, and time of pericenter passage is uniformly consistent with the predicted recovery efficiency of 77.84% shown in Figure 4.6. In summary, the driving factor in efficiency is $v_{\rm topo}$ and m_{VR}, as is expected when using a digital tracking algorithm like `KBMOD`.

4.4 Results

In this section, we will discuss the set of candidate detections which we can quantify as unique physical objects in space. We will apply and discuss additional SNR clipping, linking, orbit fitting, and filtering of fit orbits. These steps allow us to create a list of objects with multi-year linkages and detailed orbits. Finally, we will dynamically classify these multi-year objects that we have detected and quantified.

4.4.1 SNR clipping of candidate single-night detections

Now that we have characterized our recovery efficiency and the performance of `KBMOD` and the joint-fit, we can analyze those detected candidates that are likely to correspond to real moving objects in space. We refer to candidates that are potential moving objects in space (as opposed to implanted fakes or false positives) as "real". First, to further ensure no fakes are in this sample, we run fake association on both starting and ending RA and Dec. From our list of candidate objects that pass human vetting, we remove any candidate object that is within 5 arcsec of the starting position or the ending position of a fake in the implanted fakes catalog. The intersection between using starting or ending position versus just starting position is negligible. Only 6 additional candidates are associated with fakes when using

either starting or ending position (rather than solely starting position). Regardless, we exclude these 6 additional candidates from the real distribution in the following analysis.

First, having run the joint-fit on all candidates that went into human vetting and having established that the reported flux values are accurate and precise (see Figure 4.3), we are able to assign a signal-to-noise ratio to each detection. Here, SNR is functionally identical to the $\sum LH$ value discussed so far, except it uses the Φ and Ψ values from the joint-fit and is therefore more closely aligned to the actual SNR of the source. Namely, SNR $= \sum_i \Psi_i / \sum_i \sqrt{\Phi_i}$, where Ψ_i and Φ_i are generated using the negative well modelling described in Section 4.3.

In order to find the ideal SNR cutoff, above which we can classify a non-fake forward search candidate as real, we use our recovered fake distribution and our reverse search distribution. To characterize our false positive rate, we ran reverse searches on all stacks of CCD exposures in the long stares. This resulted in a total of 6973 reverse search candidates included in human vetting. Of these, we erroneously label 350 as good. Similarly, to characterize our vetting efficiency, we use the distribution of fakes before and after human vetting. These distributions are shown in Figure 4.9 as a function of SNR.

Our goal is to find an SNR cutoff that keeps the greatest number of fakes labelled good while rejecting the greatest number of reverse search candidates labelled good[6]. We quantify this cutoff by iteratively increasing the SNR cutoff until the marginal decrease in the number of fakes kept is consistently greater than the marginal decrease in the number of reverse search candidates kept. Here, we define "consistently" to mean that the aforementioned comparison happens three times in a row. Above this cutoff, increasing the SNR discards more fakes labelled good than it discards reverse search candidates labelled good. For the purpose of finding the SNR cutoff, we consider the range of SNR from $2 < \text{SNR} < 10$, with an SNR

[6]As discussed, reverse search candidates are generated by searching the opposite of the direction TNOs are expected to move. Because TNO motion is heavily influenced by Earth's reflex motion, they are expected to move in the direction of decreasing ecliptic longitude. For the reverse searches, we search in the direction of increasing ecliptic longitude. The candidates detected in the reverse search are therefore dominated by false positives. As such, we seek to minimize the number of accepted reverse search candidates.

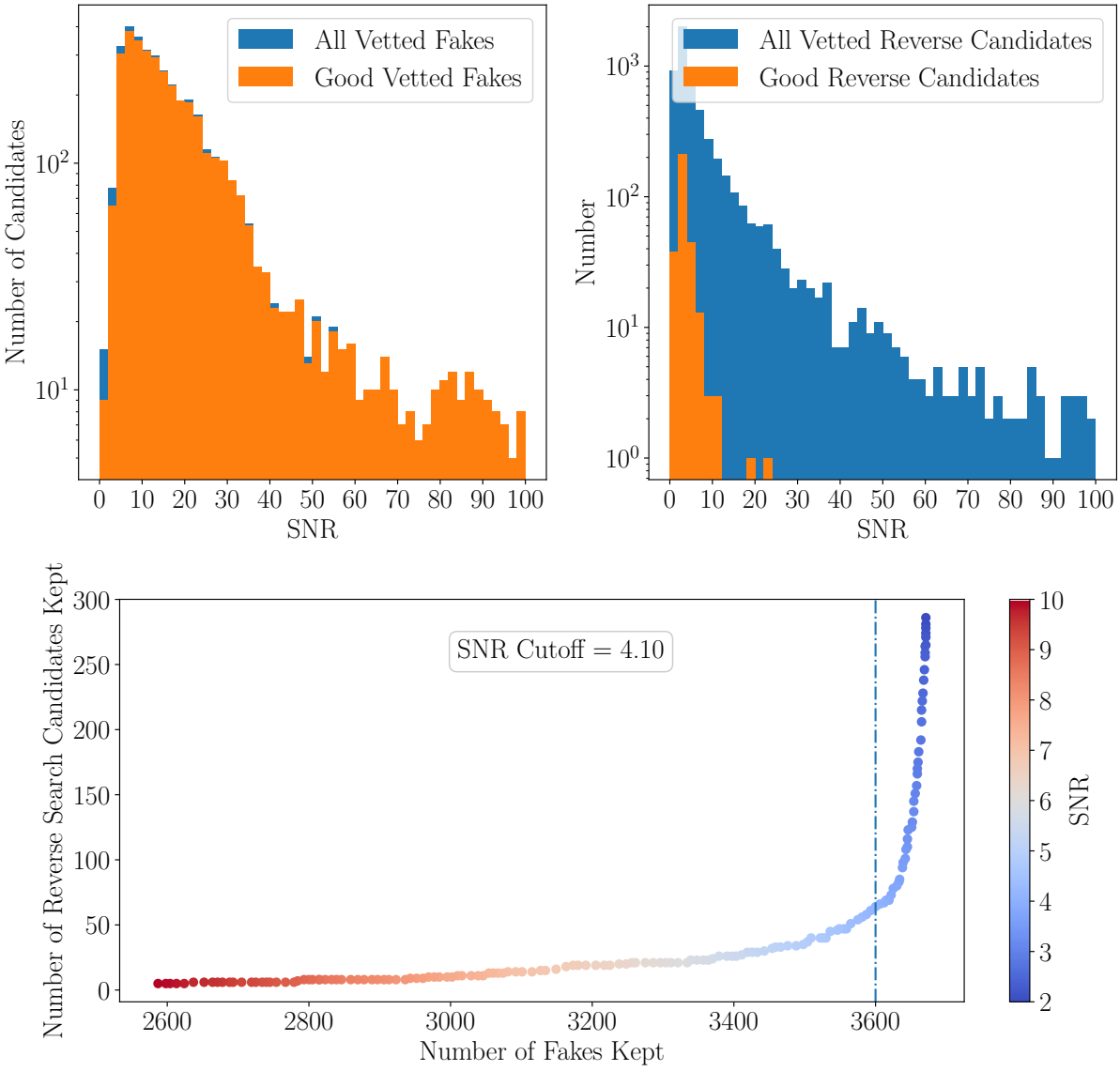

All Vetted Fakes
Good Vetted Fakes
Number of Candidates
SNR
All Vetted Reverse Candidates
Good Reverse Candidates
Number
SNR
SNR Cutoff = 4.10
Number of Reverse Search Candidates Kept
Number of Fakes Kept
SNR

Figure 4.9: A visualization of the signal-to-noise ratio cutoff used to determine if a candidate should be considered real. *Top left:* The number of all fakes that were included in vetting (blue) and the number of fakes that were labelled "good" in human vetting. The ideal case is that the orange histogram matches the blue histogram. SNR is bounded to $0 < \mathrm{SNR} < 100$ for visibility. Bins are selected using the `numpy` "bins='auto'" method based on the distribution of all implanted fakes included in vetting. *Top right:* The number of all reverse search candidates that were included in vetting (blue) and the number of reverse search candidates that were labelled "good" in human vetting. The ideal case is that the orange histogram goes to zero. SNR is bounded to $0 < \mathrm{SNR} < 100$ for visibility. *Bottom:* A parametric plot of the number of fakes kept after human vetting (higher is better) versus the number of reverse search candidates kept after human vetting (lower is better) as a function of SNR. The blue vertical line corresponds to the selected SNR cutoff value of 4.10.

step size of 0.05. As shown in Figure 4.9[7], we find this cutoff point to be SNR = 4.10. We will use this cutoff as one of several cuts in linking and orbit fitting.

4.4.2 Linking, Orbit Fitting and Dynamical Classification

Now that we have validated single-night recoveries by comparing to our population of fakes detected in a single-night long stare, we can link our single-night observations to fit detailed orbits. Furthermore, we can analyze the population of linked fakes to validate the real objects and real linked orbits that we detect. In this section, we will describe this process and the results it begets.

We parameterize each single-night detection as a tracklet of two observations. The two observations are set as the starting and ending positions using the joint-fit astrometric RA and Dec. Errors are derived from the covariance matrix from the joint-fit and are set using the square root of the diagonal terms in the covariance matrix. The greater of the RA error and the Dec error is used for linking. Tracklets for which the joint-fit failed to return a covariance matrix are given a default uncertainty in starting and ending RA and Dec of 270 mas (approximately 1 pixel), based on the median `KBMOD` residuals from Figure 2.6. This occurred in about 9.58% of cases. We then input these astrometric positions into linking and orbit fitting.

We use the method of [14] to link all candidate single-night detections (real and fake). This method first finds pairs of candidate detections that have approximately linear motion in a barycentric frame. It then grows pairs into triplets. At this point, an orbit is fit to the "n-let" using the method of [16]. Points within 4σ of the orbital error ellipse are added to the "n-let", creating an "(n+1)-let". The orbit is then refit and the process iterates until no remaining single-night candidates are consistent with the "n-let". See Section 2.8 of [14] for more detail on this method of linking and orbit fitting.

Once we have fit candidate orbits for our single-night detections, we must apply cuts to

[7]The number of accepted reverse search candidates versus the number of recovered fakes is functionally similar to a receiver operating characteristic (ROC) curve.

the large table of all possible orbits in order to create a list of trustworthy multi-night linked orbits. We require ARC > 0.8 yr and ARCCUT > 0.5 yr. ARC is the time between the first and last detection in an orbit, and ARCCUT is the ARC after removing any single night of detections from the orbit. Using ARCCUT as a cut helps ensure that the orbit does not rely solely on the potentially spurious linkage of single detection months or years later to extend an otherwise very short ARC. We also require NUNIQUE ≥ 4 and NDETECT ≥ 8, where NUNIQUE is the number of unique nights in a candidate orbit and NDETECT is the number of individual observations in a candidate orbit. We then further require that all individual observations link to their tracklet pair. This has the effect of requiring NDETECT $= 2 \times$ NUNIQUE and causing NUNIQUE to be equal to the number of tracklets in an orbit.

Next, we apply filtering based on the properties of the component single-night detections. We check if a candidate orbit has at least 4 single-night component tracklets with SNR $>$ 4.1 and remove the candidate orbit entirely if it does not[8]. We separate real and fake candidate orbits. We require each real candidate orbit to be composed solely of real single-night detections and each fake candidate orbit to be composed solely of fake (implanted synthetic) single-night detections. We label candidate orbits with a mix of real and fake component tracklets as "bad", because no correct linkage can be composed of both real and fake component tracklets.

Next, we iteratively remove possible duplicate orbits. We iterate through the list of all real orbits, fake orbits, and bad orbits separately. For each orbit in the list, we find any other orbits that share at least one tracklet with the orbit. From this group, we save the orbit with the lowest χ^2/ν and discard any other orbits that share tracklets with this orbit. If there are any remaining orbits, we again check if there are any shared tracklets and repeat the process if so. Then, we remove any candidate orbits in the bad population if that candidate orbit shares any single-night component tracklets with an orbit in the de-duplicated real or fake orbit populations. This de-duplication requires that each single-night long stare tracklet

[8]After all cuts, none of the 103 reported real objects with linked orbits contain any tracklets with SNR ≤ 4.1

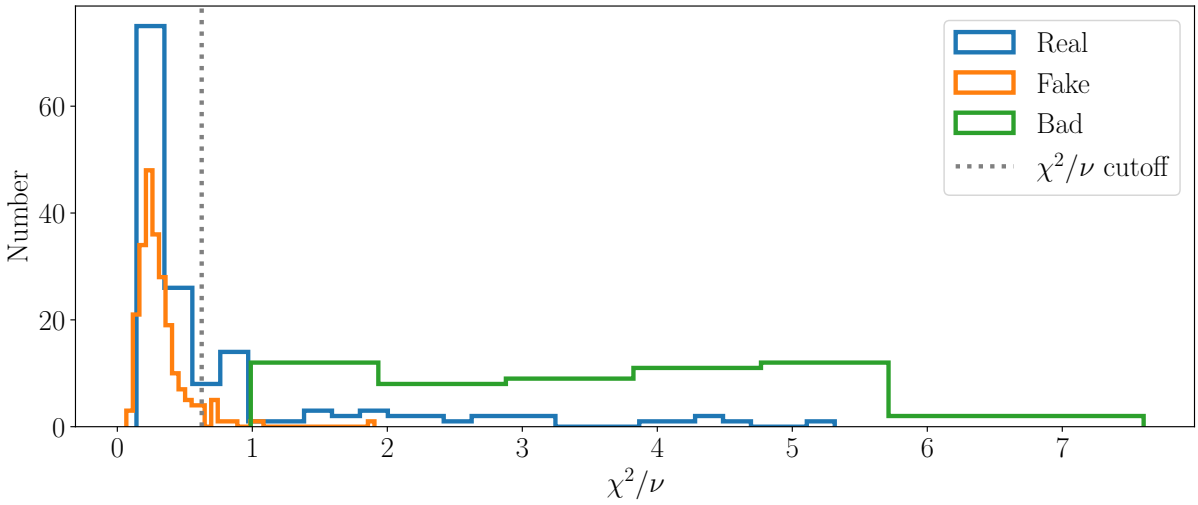


Figure 4.10: The reduced chi squared distribution (χ^2/ν) for the real (blue), fake (orange), and bad (green) linked orbits that have passed all described cuts. The grey dotted line corresponds to our chosen χ^2/ν cutoff of 0.625. Bins are set with the "bins='auto'" method.

corresponds to only one orbit in any of the three (real, fake, or bad) populations.

At this point, we investigate which of our fakes are properly linked and which are mislinked. A mislinked fake orbit occurs when single-night tracklets from two different implanted fakes link to a single orbit. Out of 250 candidate orbits, 20 are mislinked fakes and 230 are properly-linked fakes. We add these 20 mislinked fakes to the "bad" population. With these populations in mind, we can identify one final cut before investigating the real candidate orbits.

Following [14], we cut on χ^2/ν, where $\nu \equiv (2 \times \texttt{NDETECT} - 6)$ is the number of degrees of freedom of the orbit and `NDETECT` is the number of detections in an orbit. We reject candidate orbits with $\chi^2/\nu \geq 0.625$. This threshold was chosen such that it keeps 95% of the properly-linked fakes. This is a lower threshold than that used in [14] (who removed candidate orbits with $\chi^2/\nu \geq 4$). Lowering this threshold from 4 to 0.625 removes 12 additional correct fake orbits but also removes 11 erroneous linkages of fake objects. We consider this a worthwhile

trade to reduce the false positive linking rate. This threshold of 0.625 removes all orbits in the bad population (removing all the mislinked fakes and all the linkages with a mix of real and fake component tracklets) and keeps 218 correctly linked fakes. Finally, this threshold is in a small local minima in the distribution of real candidate orbits, which increase slightly where $0.65 < \chi^2/\nu < 1$.

The choice of cut depends in part on the intended scientific goals. For example, one could slightly increase this cut in order to accept more real orbits at the expense of potentially accepting false positives within the real orbital distribution. The minimum χ^2/ν from the bad population is 0.989. Therefore, we could use a cut of $\chi^2/\nu < 0.989$, which would keep the number of accepted bad orbits at 0 while increasing the number of properly-linked fakes to 227. However, within about $0.65 < \chi^2/\nu < 1$ there is a small increase in the number of real candidate orbits, diverging slightly from the characteristics of the properly-linked fake orbits. We therefore prefer a slightly more cautious limit of $\chi^2/\nu < 0.625$ derived from the 95th percentile of the χ^2/ν values of properly-linked fake orbits. Future followup (for example with the scheduled 2022 DEEP B1 data) may be able to confirm many of these real candidate orbits with $0.65 < \chi^2/\nu < 1$. The χ^2/ν distributions for the real, fake, and bad orbit populations can be seen in Figure 4.10, along with our chosen cutoff at $\chi^2/\nu = 0.625$.

After these cuts, we have a list of 218 correctly-linked fake objects, with 0 incorrectly linked fake objects. There are a total of 242 fakes in the data input into linking with at least 4 detected tracklets, `ARC` > 0.8 yr and `ARCCUT` > 0.5 yr. We therefore recover 218/242 fake orbits, with 0 mislinked fake orbits, placing our effective linking efficiency at about 90.1%. While our false positive linkage rate is numerically consistent with 0%, we emphasize that this does not inherently mean that there are identically 0 false positives in the real objects. Unanticipated differences between the real and fake populations of objects could cause them to have slightly different false positive linkage rates. Therefore, some of the real orbits could potentially be bad linkages. However, 32/38 of the orbits in the bad population are unbound ($e > 1$). As we will show, none of the real orbits have $e > 1$, increasing our confidence in our real orbits. Future follow up may help to further constrain the false positive rate of the real

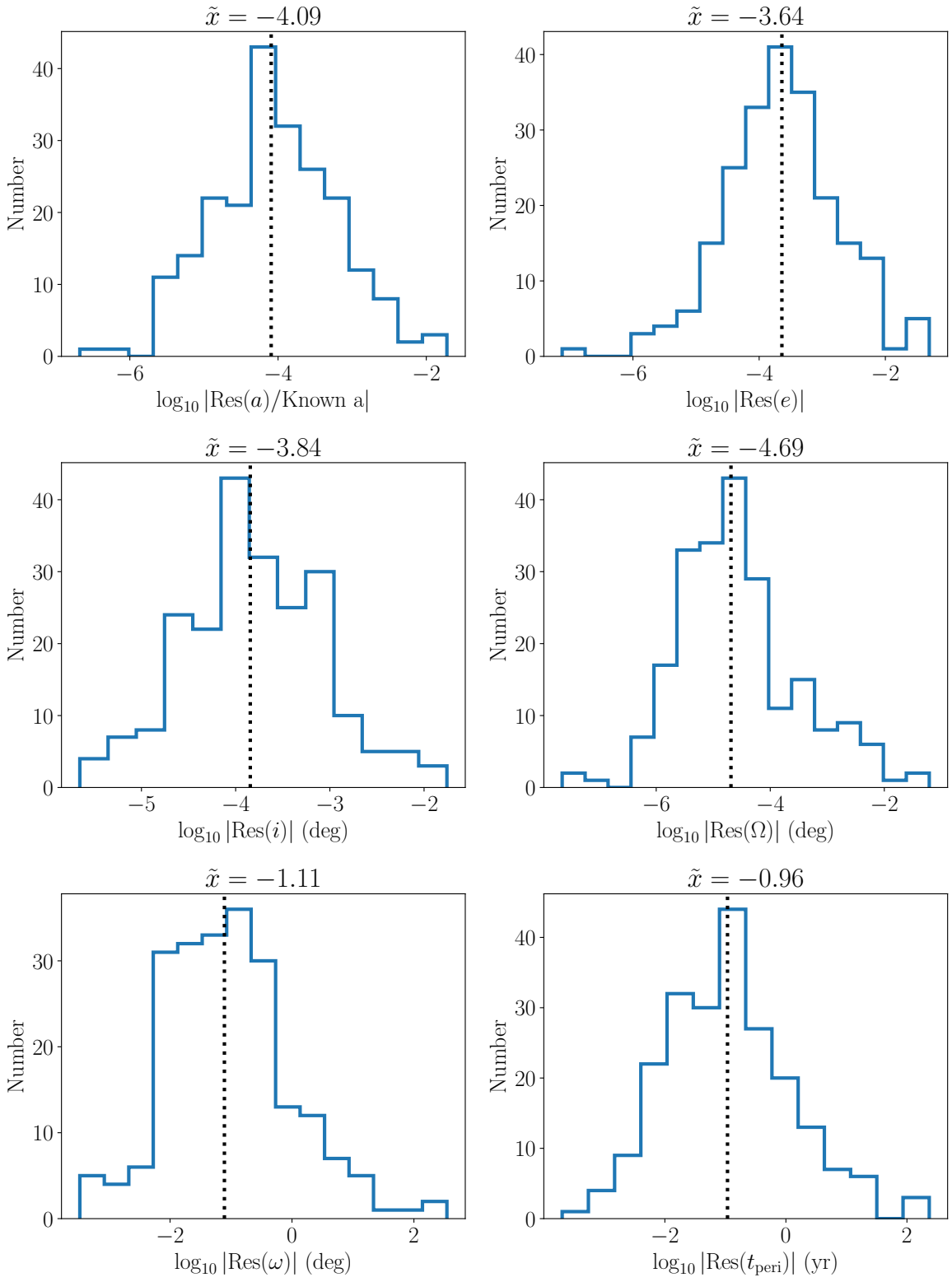

$\tilde{x} = -4.09$
Number
$\log_{10}|\mathrm{Res}(a)/\mathrm{Known\ a}|$
$\tilde{x} = -3.64$
Number
$\log_{10}|\mathrm{Res}(e)|$
$\tilde{x} = -3.84$
Number
$\log_{10}|\mathrm{Res}(i)|$ (deg)
$\tilde{x} = -4.69$
Number
$\log_{10}|\mathrm{Res}(\Omega)|$ (deg)
$\tilde{x} = -1.11$
Number
$\log_{10}|\mathrm{Res}(\omega)|$ (deg)
$\tilde{x} = -0.96$
Number
$\log_{10}|\mathrm{Res}(t_{\mathrm{peri}})|$ (yr)

Figure 4.11: Logarithmic residual plots for the 218 recovered fake orbits for semi-major axis a, eccentricity e, inclination i, longitude of ascending node Ω, argument of pericenter ω, and time of pericenter passage t_{peri}. Res() corresponds to the residual, calculated as the known value minus the measured value. The dotted black line corresponds to $\tilde{x}$, where $\tilde{x}$ is the median of the logarithmic distribution. a is scaled by the known semi-major axis to give a better understanding of orbit quality. By example, this is because a hypothetical fake object at 40 au with a 1 au residual is more poorly constrained than an object at 90 au with a 1 au residual. Bins are set with the "bins='auto'" method.

orbits.

Unlike [14], we elect not to cut orbits with FPR > 0.02 because there does not seem to be a strong benefit to using this cut with our data, and we lose about 7% of the fake linkages when applying this cut. FPR is the false positive rate of the individual linkage (as opposed to the measured false positive rate of all linking). FPR is given by

$$\mathrm{FPR} = \sum_j \mathrm{A}_{j,\mathrm{search}} n_j. \tag{4.9}$$

$\mathrm{A}_{j,\mathrm{search}}$ is the area in exposure j consistent within 4σ of the orbit, and n_j is the transient density. Using a cut on FPR decreases the number of recovered fake orbits with $\chi^2/\nu < 0.625$ from 218/242 to 202/242 (90.1% to 83.5% linking efficiency). In both cases, the number of mislinked fake orbits with $\chi^2/\nu < 0.625$ is 0. We therefore elect not to use this cut for this analysis.

We visualize the quality of the correctly-linked fake orbits and show the results in Figure 4.11. In this figure, we show the residuals of the Keplerian elements, computed as the known value minus the measured value. We take the logarithm of the absolute value of these residuals and plot the distributions. These distributions indicate that we can accurately and precisely measure all Keplerian elements. Semi-major axis, eccentricity, inclination, and longitude of ascending node are particularly well-constrained. Even for argument of pericenter and time of pericenter passage, which are slightly more challenging to constrain, the medians of the residual distributions correspond to residuals of about 0.08 degrees and 0.11 years respectively. For reference, TNO orbital periods are longer than 164 years (the orbital period of an object at 30 au).

For semi-major axis a, we scale the residuals by the known value of semi-major axis. We find that the median residual in a is about $10^{-4.09}$ of the known value for a, or about 0.01% of a. Figure 16 of [13] shows a similar (but not identical) quality parameter for OSSOS ([8]) and DES, two large and comprehensive TNO surveys. In this figure, they use data from real objects and show estimated uncertainty in a divided by a. They call this parameter σ_a/a. The median of the distribution of σ_a/a is $\mathcal{O}(10^{-4})$ for OSSOS and DES. Repeating this exercise

for our correctly-linked fake orbits, we find a median value of $\sigma_a/a = 10^{-2.91}$, which is larger than that of OSSOS and DES. However, the disagreement between $\log_{10}|\text{Res}(a)|/\text{Known a}$ and σ_a/a suggests that we are overestimating our errors. This is corroborated by our χ^2/ν distributions for our real and fake orbits. Among those orbits with $\chi^2/\nu < 0.625$, the median χ^2/ν is 0.268 and 0.266 for the real and fake orbits respectively. These distributions should peak near 1. We therefore appear to overestimate our positional errors by about a factor of 2. This may be due to correlation in the tracklet errors, which we treat here as uncorrelated. See Section 5.3 for further discussion and planned future improvements.

After all cuts, we have a list of 103 detected real objects with linkages. These 103 real objects have a median $\sigma_a/a = 10^{-3.30}$, closer to the results of OSSOS and DES, but still somewhat worse. We use `skybot` to identify known objects in the DEEP B1 data. We associate all single-night candidates input into linking with KBOs reported with `skybot`. If both the starting RA and the starting Dec of the detection are within 20 arcsec of the RA and Dec predicted by `skybot`, then we consider a candidate associated with a known KBO. We then apply the cuts on `ARC`, `ARCCUT`, and `NUNIQUE`. This leaves us with 4 known KBOs (2003 QL91, 2002 PV170, 2002 PX170, and 1999 RX215) all of which are members of the 103 detected real objects with linkages. In the table in Appendix A.2, they have Id 8, 11, 12, 39. These known KBOs are all classical KBOs.

In Figure 4.12, we show the median topocentric discovery distance for the 218 correctly-linked fake orbits and the 103 real orbits which pass all orbital cuts. This distance is defined as the median of the topocentric distances reported by the orbit fit at each long stare for which the object was observed. As expected, these results are consistent with the distances for the fake objects detected in single long stares, shown in Figure 4.5. The distribution of discovery distances for the fake orbits encompasses the distribution of the discovery distances of the 103 real orbits, as one would expect for correctly-linked real orbits.

We follow the classification scheme of [14], which is similar to [50] and [74]. We create 20 clones of each of the 103 real objects. We then use `Rebound` ([112]) with the WHFast symplectic integrator ([111]) to integrate the clones for 10 Myr from the starting time t_0.

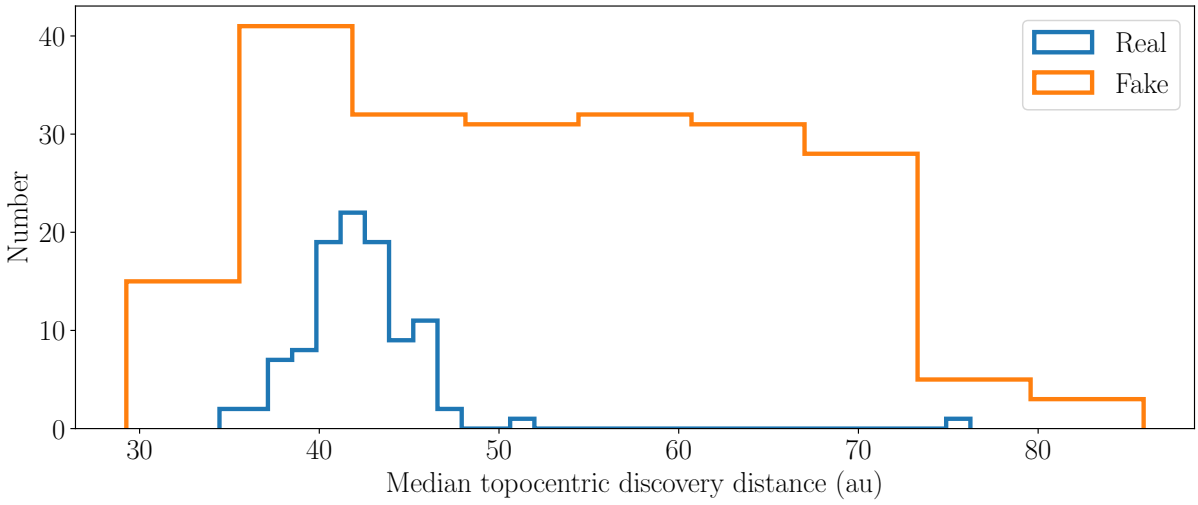


Figure 4.12: The median topocentric discovery distance for all correctly-linked fake orbits and all 103 real orbits that pass all orbital cuts. The median is taken from the distribution of topocentric distances for each long stare (and thus each tracklet) in the linked orbit. Bins are selected using the `numpy` "bins='auto'" method.

Clones that librate around the resonance argument for more than 90% of the simulation time are considered resonant. Clones where the semi-major axis varies by more than 3.75% from $a(t_0)$ are considered scattering. Clones where $a(t_0) < a_N$ are considered inner centaurs, where a_N is the semi-major axis of Neptune. Clones where $q(t_0) < a_N$ but $a(t_0) \geq a_N$ are considered outer centaurs. Clones with $e(t_0) > 0.24$ are considered detached. Other clones are considered classicals.

Objects are classified based on the classification of the majority of their clones ($> 50\%$). Objects where more than 80% of the clones are resonant in the same resonant number are classified as securely resonant. Objects where more than 50% of the clones are resonant in the same resonant number are classified as insecurely resonant. Objects where no single classification reaches a majority are considered insecurely classified. All 103 detected and linked real orbits are shown in Figure 4.13.

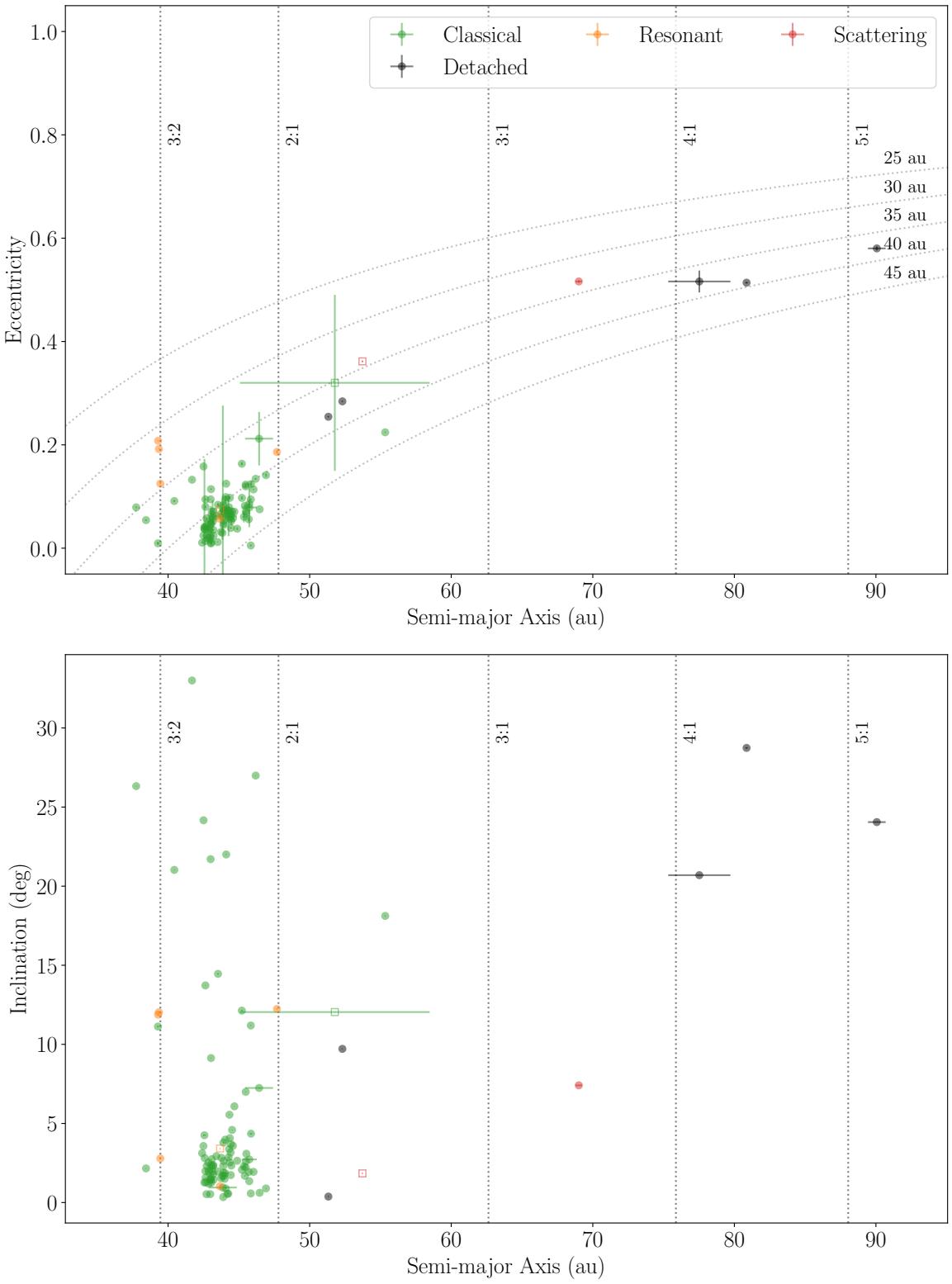

Classical
Resonant
Scattering
Detached
3:2
2:1
3:1
4:1
5:1
25 au
30 au
35 au
40 au
45 au
Eccentricity
Semi-major Axis (au)
Inclination (deg)
Semi-major Axis (au)

Figure 4.13: Eccentricity versus semi-major axis (top) and inclination versus semi-major axis (bottom) for the 103 detected objects with good orbits from the DEEP B1 fields. Color denotes dynamical classifications, which were determined following the approach [50] and [14]. Shaded circles denote secure dynamical classifications. Open squares denote insecure dynamical classifications. Vertical dotted lines correspond to the approximate location of select Neptunian p:q MMR. Curved dotted lines (top plot only) correspond to lines of constant pericenter.

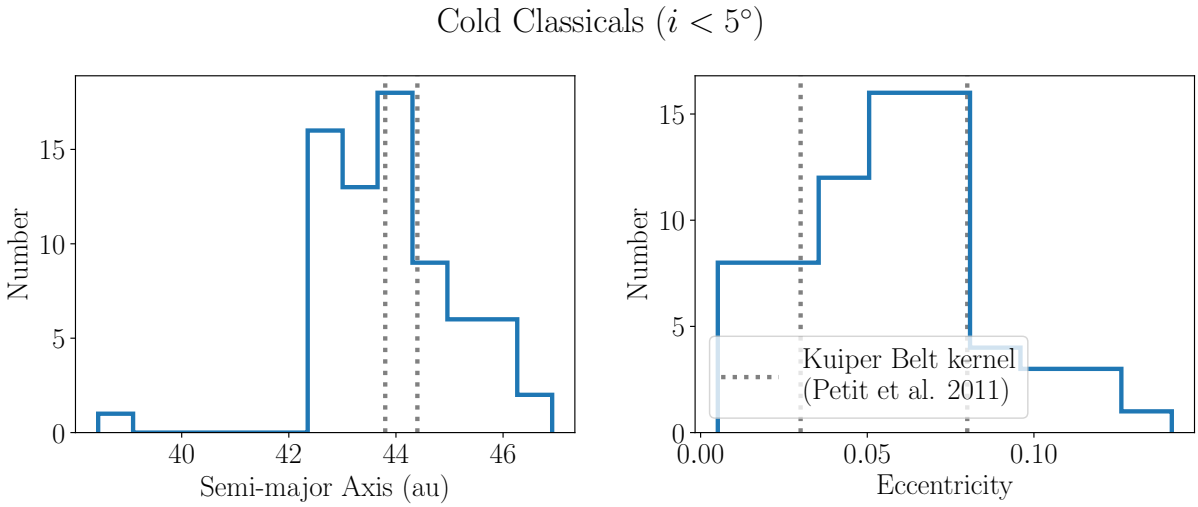


Figure 4.14: The distribution of semi-major axis and eccentricity for the detected and classified cold classical TNOs ($i < 5°$). Grey dotted lines show the bounds of the Kuiper Belt kernel population modelled by [107]. Bins are selected using the `numpy` "bins='auto'" method.

We identify 6 resonant TNOs, 1 of which is insecurely classified. The securely-resonant objects are along the 3:2, 7:4, 2:1 resonances, with more objects in the 3:2. This is in line with previous surveys (e.g. [14]). The survey geometry (often close to Neptune) is such that the DEEP B1 fields are generally far from the 3:2, 7:4, and 2:1 resonant populations, which explains the proportionately-smaller number of resonant objects compared to (for example) [14].

We identify 90 classical TNOs, 1 of which is insecurely classified. The 4 recovered known objects are in this population. [107] found a component of cold classical TNOs with $i < 5°$, 43.8 au $< a <$ 44.4 au, and $0.03 < e < 0.08$. Figure 4.14 shows our population of cold classical TNOs ($i < 5°$) along with the bounds of the kernel population from [107]. We appear to match well in eccentricity, and we seem to reproduce the sharp drop-off around $a = 44.4$ au predicted by the kernel population. The kernel population is often discussed and informs theoretical models of the Solar System. For example, it could be explained by a

sudden jump in Neptune's semi-major axis, possibly due to an encounter with a now-ejected fifth giant planet [93, 95, 99, 51].

We identify 5 detached objects, 0 of which are insecurely classified. One of these detached objects ($a = 77.53$ au, $e = 0.52$, $i = 20.69°$) was detected with a median topocentric discovery distance of 76.23 au. In our data, it is the object with the most distant discovery distance.

We identify 2 scattering objects, 1 of which is insecurely classified. With the detection of these 2 scattering objects, we demonstrate that we are able to detect and characterize the 4 main populations of TNOs (classical, scattering, detached, and resonant).

Finally, we investigate our observed H magnitude distributions. First, we visualize the correctly-linked fakes, comparing the residuals of the simulated values for median H magnitude and the median of the observed values for H magnitude. Because we did not incorporate phase information when computing the implanted value of H for the fakes, we likewise do not incorporate phase information in the measured value of H for the fakes. We show the results in the top plot of 4.15. The median of the residuals is $\tilde{x} = -0.001$, indicating that there is no observable systematic offset in the residuals. The standard deviation is $\sigma = 0.086$ and the value of σ_G is even smaller, with $\sigma_G = 0.031$.

For the 103 real orbits, we use an $H - G$ model to calculate H as described in Equation 1.2. This gives us a number of measurements for H equal to the number of tracklets in the orbit. We then compute the median value for H. We show the results in the bottom plot of 4.15. In our data, the object with the largest H value has a median H of $H_{VR} = 9.63$. Unless otherwise specified, we are reporting H_{VR}, as our data are taken in VR. This means that there is not a direct comparison to the JPL reported H values values, but we can make an approximate estimate by assuming a TNO color. Assuming an average TNO color of $V - R = 0.6$ ([15]) and $VR - R = 0.03$ ([43]), we find that $H_V = H_{VR} + 0.57$. For approximate comparison with JPL, our faintest object would have $H_V \sim 10.20$. This would make it the object with the largest H_V among those TNOs with $q > 30$ au and an observational arc greater than 1 year, according to the JPL small body database as of June 2022. Currently, 2015 VL166 and 2015 VM165, which fit the aforementioned constraints,

both report $H_V = 10.10$ on the JPL small body database.

We will discuss these results, as well as some of the limitations and next steps, in more detail in Chapter 5. A table of these detected objects is listed in Appendix A.2.

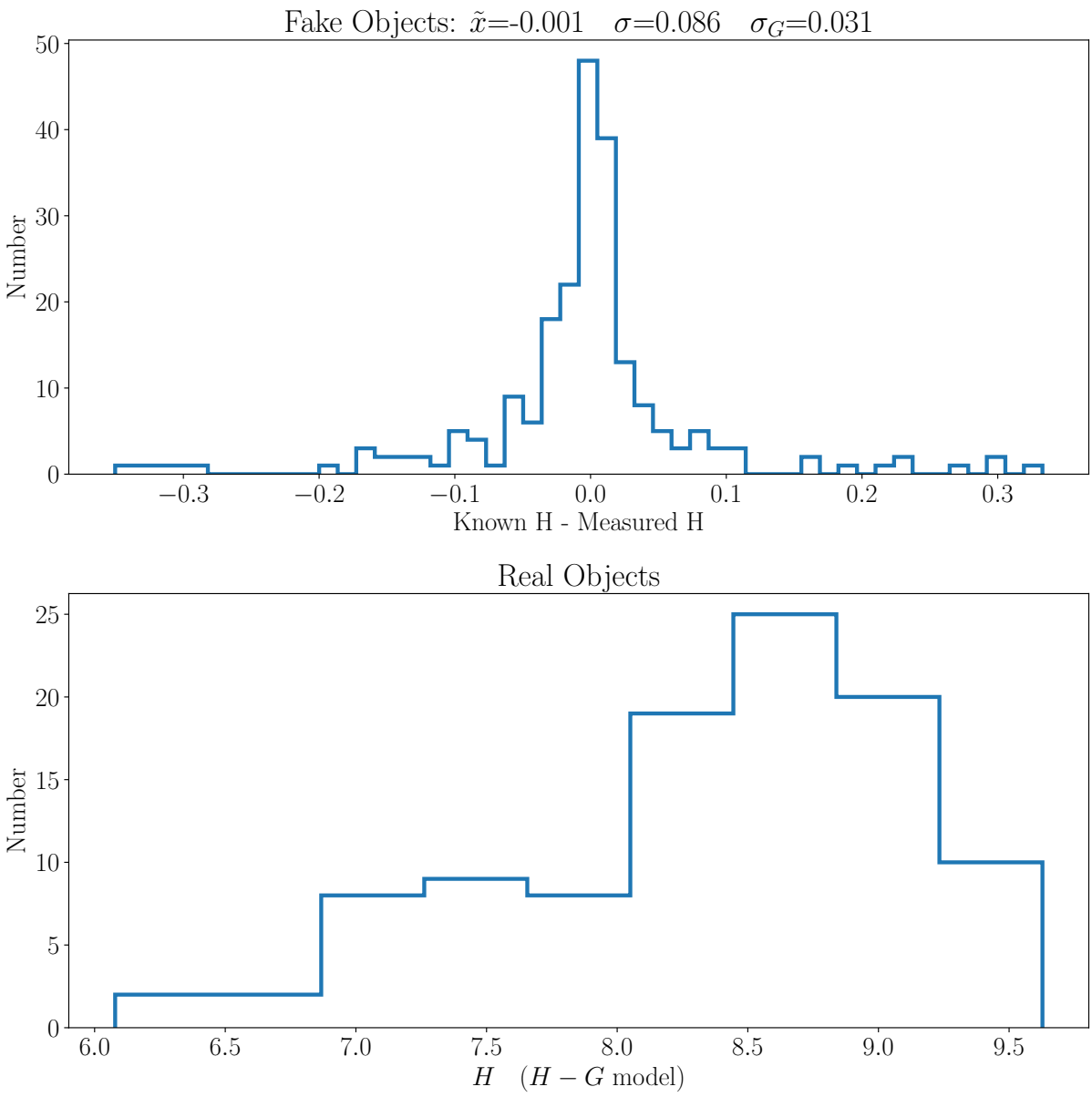


Figure 4.15: *Top:* The residuals in median H magnitude for the correctly-linked fake orbits. The title includes the median $\tilde{x}$, standard deviation σ, and σ_G value for the distribution. H is computed without an H-G model in order to conform to the known value for H from `DESTNOSIM`. *Bottom:* The median H magnitudes of the 103 detected real objects. This is computed using an H-G model (Equation 1.2) with an assumed value of G=0.15.

Chapter 5
DISCUSSION

In this chapter, we will summarize and discuss the impacts from the work in Chapter 2, Chapter 3, and Chapter 4. We will also discuss the next steps for KBMOD, including the possibility of applying KBMOD to upcoming surveys such as LSST.

5.1 Discussion of Chapter 2

In Chapter 2, we present a number of improvements and extensions to KBMOD. σ_G-based lightcurve filters robustly remove outliers, while providing user-programmable percentile thresholds that allow the filter to be tailored to specific data sets. Running the σ_G filter in the core GPU algorithm helps to ensure that the most promising trajectories are passed to the CPU for further analysis, in cases where the number of results per pixel are maxed out. For the analysis of the DECam NEO data set, Figure 2.1 shows an example of the number of candidate trajectories at three stages of processing: after they are output from the KBMOD GPU subroutine, after σ_G lightcurve filtering, and after stamp filtering and clustering. It shows this both with and without GPU filtering and as a function of the maximum number of results per pixel. Characteristically, these filters reduce the number of candidate trajectories output by KBMOD from $\mathcal{O}(10^7)$ candidate results per CCD search to $\mathcal{O}(10^2)$ candidate results per CCD search. CNNs allow for novel filtering of false positives. As shown in Figure 2.2, the CNN filter we developed is able to lower the total number of candidate results per CCD by about 1 order of magnitude, yielding $\mathcal{O}(10^1)$ candidate results per CCD search for subsequent human review. Reprojecting the search space into a barycentric frame may help extend KBMOD's searchable time baselines by an order of magnitude. Finally, the joint-fit allows precise characterization of KBMOD results, which was critical to the linking and orbit

fitting process in Chapter 4.

These improvements enabled KBMOD to process and characterize the data sets discussed in Chapter 3 and Chapter 4. In Section 5.4, we will discuss planned extensions to KBMOD which will build on this work and help to enable KBMOD to process LSST data.

5.2 Discussion of Chapter 3

The improvements presented in Chapter 2 and Chapter 3 helped enable KBMOD to detect 22 out of 26 known objects in the recovery sample. The trajectories of these known objects were recovered with a median error in starting position of less than two pixels. Furthermore, KBMOD was able to detect 75 objects that we were unable to link with any previously-known objects. Although the time baseline of the data was short, we were able to fit the barycentric distance, inclination, and longitude of ascending node of both the known objects and the unidentified objects. The inclination distribution of the recovered objects is consistent with the distribution from [21]. The number of objects detected as a function of magnitude is consistent with the distribution from [43] assuming a net fill factor of 0.55.

[136] validated KBMOD on the High Cadence Transient Survey (HiTS) [40]. Chapter 3 validates algorithmic improvements to KBMOD filtering with a survey that has a time baseline of up to 4 nights, compared to the three nights used in [136]. Furthermore, this work validates KBMOD as applied to images that have been differenced with a coadded template.

In so doing, we demonstrated that KBMOD can recover KBOs in difference images from a survey with a longer time baseline and an irregular cadence. However, this required more robust filtering methods. By adding GPU filtering, we have increased the effective number of potentially-valid candidate trajectories that can be passed out of the GPU for further filtering and analysis. After σ_G lightcurve filtering, stamp filtering, and clustering, a search with the GPU filtering enabled returns about 2 times more candidate results than a search without the GPU filtering enabled. With the σ_G-based filtering, we have also implemented more robust lightcurve filtering that improves filtering with an irregular image cadence. The CNN ResNet50 stamp filter shows great promise for future stamp filtering methods.

Given the relatively low inclination (median value of $\widetilde{i} = 7.67°$) and barycentric distances between 30 au and 50 au (median value of $\widetilde{r_0} = 41.28$ au), we find it likely that the majority of the unidentified objects presented in Figure 3.3.1 are Kuiper Belt objects. However, because the short arcs prevent us from placing accurate constraints on semi-major axis and eccentricity, we are unable to confirm this prediction with the current data. Future follow-up or precovery attempts[1] for these objects may be able to extend the observational arcs enough to accurately constrain them to the Kuiper Belt, and perhaps place them within a Kuiper Belt subpopulation (e.g. the cold classical Kuiper Belt).

We will discuss some current limitations of `KBMOD` in Section 5.4.

5.3 Discussion of Chapter 4

In Chapter 4, we present the first time that results from `KBMOD` have been used to detect TNOs with sufficient depth and orbital arcs to fit detailed orbits and perform dynamical classification. This step indicates that `KBMOD` (and digital tracking in general) is ready to serve as a detection tool for state-of-the-art TNO surveys, and indeed has done so already by being applied to DEEP.

σ_a/a, a metric of orbital quality, is characteristically worse in the DEEP B1 data than in OSSOS and DES. However, σ_a/a is also greater than Res(a)/Known a, indicating that this may be partially because we are overestimating our errors. The distribution of χ^2/ν, which peaks below 1, also suggests we are overestimating our errors. To address this limitation, we plan to implement a new fitting method in the orbit fitting code of [16] to allow it to propagate the tracklet covariance matrices into the fit, rather than relying on uncorrelated errors between observations.

Using the residuals (which can only be calculated for the fakes), the median value for the

[1]Follow-up or precovery would likely be challenging for these objects. Precovery would require an existing data set to overlap with the predicted positions of these objects. Because these objects have short observational arcs, targeted follow-up may no longer be feasible. Follow-up would therefore likely need to be similar to precovery efforts, requiring an existing data set that already overlaps with the predicted positions of these objects. The closer the precovery or follow-up data are to the search sample data (which were taken in 2014 and 2015), the more feasible these attempts become.

distribution of $\log_{10}|\text{Res}(a)/\text{Known } a|$ is -4.09. This is consistent with the $\mathcal{O}(10^{-4})$ values for σ_a/a reported in [13]. This suggests that if we can more properly propagate our covariant errors through the orbit fitting algorithm, our values of σ_a/a (which should then more closely agree with the residuals) may reach similar values as those in OSSOS and DES.

We report the detection of 99 real objects with orbits that we were unable to link with previously detected objects, as well as the recovery of 4 known TNOs. These numbers are expected to grow as more data from B1 is taken. We have 1935 real single-night detections with SNR > 4.1 which were not linked to any of the 103 real objects with characterized orbits. We may be able to link some of these single-night detections once the scheduled 2022 DEEP B1 data are taken and analyzed. There are also 3 more quadrants of data in the DEEP survey (the A0, A1, and B0) fields, although both the number and quality of images in B1 suggest that it may remain the most productive quadrant for TNO discovery in the DEEP survey.

We are able to dynamically classify these 103 detected objects with characterized orbits, recovering classical TNOs, resonant TNOs, scattering TNOs, and detached TNOs. Because we can accurately and precisely characterize the orbits, we are also able to compute H magnitudes for these objects.

Future work on DEEP will use the `DESTNOSIM` survey simulator to create a debiased TNO population model. This will allow DEEP to model parameters such as the absolute size distribution of the Kuiper Belt.

5.4 Next steps and future surveys

Next-generation astronomy surveys will soon be current-generation. This imminent wealth of data will require new computational tools in order to access its full potential. `KBMOD` has the potential to increase the number of TNOs detected with LSST from $\sim 40,000$ to $\sim 320,000$ as well as investigate the faint and mysterious class of objects at the very edge of our Solar System. (See Section 1.6 for the assumptions used to estimate these numbers). In terms of probing the sednoids, with three months of coadded data we could detect a Sedna-like

object at opposition at over 290 au, as opposed to ~ 210 au for a single image. With a year of coadded data, 290 au increases to 310 au. If we could coadd the entire LSST survey, 310 au increases to over 415 au. Note that objects on elliptical orbits spend much more of their time further from the Sun. If Sedna, which was detected near its perihelion around 90 au [23], is representative of a larger population of sedoids, then most of these objects should be closer to apocenter than pericenter. Therefore, a linear increase in detection distance should yield a super-linear increase in the number of detected objects on a similar orbit. With this coaddition approach, it might even be possible to detect inner Oort Cloud objects with perihelion near 400 au, and aphelion well beyond.

Further work is needed before `KBMOD` will be able to run on LSST. This work has presented many improvements to `KBMOD`, but there are more improvements still to come. In this section, we will discuss some planned improvements for `KBMOD`, motivate their potential impact, and provide guidelines for how they might be implemented.

5.4.1 Cross-CCD searches and GPU memory limitations

Enabling `KBMOD` to search across multiple CCDs will increase the effective fill factor, enabling greater completeness and longer time baselines. CCD chip gaps and camera edges will always keep the fill factor below 1.0 (relative to a circular footprint). However, with a CCD chip gap between 153 (columns) and 201 (rows) pixels, a KBO would move past the chip gap and onto the next CCD in about one night, assuming a typical KBO velocity of 275 px/day.

More importantly, cross-CCD searches will be required in order to search for TNOs over 90 day time baselines. To achieve the estimated increase in TNOs detected by LSST (from about 40,000 to 320,000 TNOs), we will need to be able to effectively run digital tracking on about 3 months of LSST data. The barycentric reprojection discussed in Section 2.5 demonstrates that it is possible to efficiently model a TNO trajectory over this time baseline, if we can search trajectories that span many CCDs. Finite GPU memory causes this to be a tricky book keeping challenge of loading the relevant images into the GPU at the right stage of computation in a way that minimizes the effect of memory bandwidth bottlenecks.

Image compression may help alleviate the problem of GPU memory and GPU memory bandwidth, but it is unlikely to solve it entirely. To demonstrate this, we can first place a very approximate estimate on the data volumes needed to run a KBMOD search using LSST data. The LSST camera will have 189 4K x 4K CCDs. An uncompressed 4096 x 4096 image of 32 bit floating point numbers (e.g. the image plane) is about 67.1 MB. An LSST calexp FITS file has an image plane, a mask plane (which is not 32 bit) and a variance plane, plus associated metadata. KBMOD takes this information and turns it into 2 32 bit Φ and Ψ images, which are sent to the GPU to run the core KBMOD algorithm. It does this for N images in a search. Thus, a search of M CCDs and N epochs will have a total data volume of at least $M \times N \times 2 \times 67.1$ MB. LSST will take about 12 images of a single field in r band over 90 days. The total data size of these images, as sent to the GPU for KBMOD, would be about 304.4 GB. This data volume does not include result data, which can be as large as $67.1 \text{ MB} \times 8 \times M \times 6$, when saving 8 results per pixel with each result containing 6 32 bit floating point numbers. The 6 numbers are the x and y starting position, flux, likelihood, velocity, and angle. x and y starting pixel values are integer values, but we treat them as floating point numbers for the purpose of this rough estimate. Given these simplified assumptions, trajectory results could add up to 608.8 GB, for a total of about 913.2 GB. This is well above the available memory of current GPUs like the Nvidia A100, which has up to 80 GB of GPU RAM. However, it should be possible to significantly reduce the GPU memory footprint of KBMOD.

One improvement to KBMOD would be to stream results out of GPU RAM into CPU RAM (and possibly to disk, if necessary). Although KBMOD results can be up to 8 results per pixel, in practice KBMOD seldom maxes out this limit for every possible pixel. The exact number of results returned by KBMOD depends on factors such as image differencing quality, field density, and KBMOD grid selection. The above 608.8 GB is therefore an upper limit; however, even 608.8 GB could be fit into CPU RAM, which can currently be several TB. Streaming KBMOD results out of the GPU and into CPU RAM could therefore greatly help with the problem of limited GPU RAM.

Another improvement to KBMOD would be to allow variable floating point precision for

images and results and/or various forms of image compression. Switching to 16 bit floating point numbers reduces the memory footprint by a factor of 2. Switching to 8 bit integer values reduces the memory footprint by a factor of 4. Using 8 bit integer values would require applying some scaling to capture the numerical resolution required for each different variable. Combined with streaming results out of the GPU, this could reduce the GPU RAM required to 152.2 GB (in the case of 16 bit floats) or 76.1 (in the case of 8 bit int). Subdividing the LSST focal plane into halves or quarters then further reduces the `KBMOD` GPU memory footprint to tractable levels of about 40 GB to 80 GB.

It is also worth mentioning that GPU capabilities are rapidly improving. 10 years ago, 2 GB GPUs were the norm, with some 6 GB workstation GPUs like the Quadro 6000 available at the high-end [2]. Nowadays, the NVIDIA A100 with 80 GB of GPU RAM represents one of the best available GPUs. In another few years, there may be workstation and server GPUs with well over 100 GB of GPU RAM.

Despite these proposed memory improvements, many TNO trajectories will still span two or more LSST focal planes over 90 or 180 days. As shown in Figure 2.4, a TNO may span several square degrees over 180 days. Even with the 9.6 $\deg^2$ field of view of the LSST camera, any TNOs that are not well-centered in the LSST focal plane will be likely to drift into adjacent exposures over 90 to 180 days. Given the size estimates of processing even a single LSST focal plane, it is unlikely that it would be possible to fit several LSST focal planes into the memory of a single GPU. Therefore, the book keeping problem of loading relevant images into and out of GPU memory during the `KBMOD` search is not going away, but being able to load a wider field of Φ and Ψ images at once should make it easier.

5.4.2 The computational cost of barycentric searches for LSST

Reprojecting into barycentric space significantly changes the computational complexity of `KBMOD`. We strongly emphasize that a full characterization of the computational complexity

[2] See `https://en.wikipedia.org/w/index.php?title=List_of_Nvidia_graphics_processing_units&oldid=1092199809` for a partial list of Nvidia GPUs.

of a barycentric search will require real scaling tests with precursor data sets. However, the following quick back-of-the-envelope estimates may shed some light on what those scaling tests might reveal. Figure 2.5 suggests that barycentric reprojection may require grid spacing as fine as 0.04 au to cover the Kuiper Belt. Searching from 30 au to 50 au would therefore require about 500 grid points assuming uniform grid spacing. For circular orbits, we could assume an exact orbital velocity, completely negating the need to sample from a grid of speeds. For example, using Kepler's Third Law we find that a circular orbit at 40 au has a barycentric angular speed of about 14.03 arcsec day^{-1}. This 40 au TNO would therefore move an angular distance D about 1262.3 arcsec over 90 days. Using the law of cosines, we can estimate the angular grid resolution θ required in order to ensure that each trajectory is less than some value ϕ from neighboring trajectories at the end of a time baseline T. This is given by the following equation:

$$\theta = \arccos \frac{2D^2 - \phi^2}{2D^2}. \tag{5.1}$$

With a maximum separation $\phi = 2$ arcsec[3] and $D = 1262.3$ arcsec, we find $\theta = 0.128°$. To cover the full range of allowed objects, angles would need to be sampled from 0° to 360°. This means that we would need 2804 angular grid points, for a total grid of $500 \times 1 \times 2804 = 1402077$. For comparison, the search done in Chapter 3 had $256 \times 512 = 131072$ grid points. Therefore, just to cover the circular and nearly-circular orbits of the Kuiper Belt would require searching about 10.7 times as many grid points as was done in Chapter 3.

We can use these numbers to estimate that it may be computationally feasible to process LSST data with `KBMOD`. Each stack of 20 2K x 4K CCD searches took about 1300 seconds to run the core GPU `KBMOD` algorithm for the data in Chapter 3. Then, to process a full LSST focal plane of 189 4K x 4K CCDs, assuming 20 images in the stack, we might expect a GPU processing time of about $1300 \text{ s} \times 189 \times 2 \times 10.7 = 1460$ hr. To process 90 days of the entire $18,000$ deg^2 of LSST would therefore take about $2628253 = \mathcal{O}(10^6)$ GPU hours, just for the

[3]A trajectory separation of 2 arcsec means that a given moving object would be no more than about 1 arcsec (or approximately 1 PSF FWHM) from a searched trajectory.

core GPU algorithm. It would require a similar order of magnitude of CPU hours for post processing and filtering.

However, these numbers depend heavily on the choice of parameter space. We can repeat this exercise to instead cover circular orbits from 80 au to 100 au. The barycentric grid spacing is then approximately 0.23 au, $D = 446.3$ arcsec at 80 au, so the grid spacing becomes $87 \times 1 \times 991 = 495709$, which is only 3.78 times the number of grid points of the search from Chapter 3. The total process time decreases by a factor of 2 to about 929226 GPU hours to process 90 days of LSST data with the core GPU algorithm of `KBMOD` (again, with similar processing time for CPU post processing and filtering).

Depending on the simplified assumptions we use, processing time can scale dramatically. First, we assume processing scales linearly with the number of grid points, which ignores scaling in memory bandwidth and other factors. Next, we assume circularity, which collapses the velocity to a single dimension. To search (for example) 4 grid points in velocity would increase processing time by a factor of 4. However, we also assume that we would search angles between 0° and 360°. We could instead focus on objects with inclination $0° < i < 30°$, for example, which would reduce the computational cost by a factor of 6. Finally, as we directly demonstrated, the distance range of interest can vary computational cost by an order of magnitude.

These estimates are not a substitute for real scaling tests. One of the primary next steps for `KBMOD` before it is ready for processingg with LSST will be to use precursor data sets and/or early LSST data to directly measure the scaling of `KBMOD` with barycentric reprojection. As discussed, searching different regions of the outer Solar System changes the computational cost dramatically. One initial approach might involve getting $\mathcal{O}(10^5)$ GPU hours on a cluster such as XSEDE (a sizable but realistic allocation) and processing 10% to 25% of LSST data over 90 days, focusing on the class of low-eccentricity orbits. For distance, one could select an interesting feature with narrow distance ranges, such as the Kuiper Belt kernel from 43 au to 45 au. Alternatively, one could focus first on more distant objects, such as those between 80 and 100 au, then move to progressively closer distance regimes. After

demonstrating that KBMOD can process LSST data, one could then begin to request funding for the very large allocation of GPU resources required to process all LSST data for the entire outer Solar System.

5.4.3 Other steps

CPU RAM usage can become large during the filtering stages of KBMOD. Although CPU RAM can be several TB, and is therefore not as big of a concern as GPU RAM limitations, optimizing the RAM usage for KBMOD could allow it to run more easily on non-server workstations. This would improve the usability of the code for users who want to run on desktop PCs. Furthermore, server GPU resources are sometimes partitioned with relatively small amounts of CPU RAM (e.g. 128 GB). Future work should investigate the feasibility of creating a "low RAM" mode for KBMOD, where more processing can be efficiently cached to disk. One area of focus might be during KBMOD's stamp filtering, when millions of 21 x 21 stamps can be generated and stored in system RAM until final clustering is complete.

Improving the PSF model from a Gaussian to a full PSF model will improve the KBMOD $\sum LH$ estimate. Currently, implementations of KBMOD have used a Gaussian with constant standard deviation as the PSF model for creating the Φ and Ψ images. Future work will investigate the benefits of using a full PSF model from the LSST Science Pipelines. In particular, using a separate PSF for each image, rather than a single PSF for the entire search stack, should lead to more accurate and precise estimates for $\sum LH$.

Early termination of search trajectories should allow for significant time savings in some cases. In a Bayesian framework, a given search should be stopped when the data for a given trajectory demonstrate that the absence of a source is much more likely than the presence of a source. We can quantify this termination criteria by using a parameter Y. Then, the search should be stopped when $L\left(\hat{f}|0\right) > Y \int L\left(\hat{f}|f\right)\pi\left(f\right)df$, where $L\left(\hat{f}|f\right)$ is the likelihood of observing $\hat{f}$ given that a source of flux f exists. For the simple prior $\pi\left(f\right) = \delta\left(f - f_0\right)$, where f_0 is the source flux, this inequality simplifies to $\Psi f_0 > 0.5 f_0^2 \Phi - \log Y$. In practice, we can use $\log Y$ instead of Y as the stopping parameter. This is more computationally efficient

and should decrease any overhead added to the algorithm by computing this inequality. For 30 images, with $f_0 = 10\sigma$ and $\log Y = 3$, early estimates indicate time savings up to a factor of twenty, for individual trajectories that do not include a source. Future work will implement early termination in `KBMOD` and characterize the time savings potential with real world scaling tests.

Applying transfer learning to the CNN filter should allow it to continue to be applicable to new surveys. As demonstrated by Figure 2.2, we were able to apply the CNN trained for the data in Chapter 3 to the data in Chapter 4. This is partially due to creating generalized training data and selecting a deep CNN architecture that is able to learn generalized features. This is particularly noteworthy because the data in Chapter 4 has negative wells that were not present in Chapter 3, but the CNN filter still performs well. However, it is very likely that retraining the CNN for each new data set would improve the results of the CNN. This is especially true if the training data consists of injected fakes that cover the speed, brightness, and trajectories of the objects of interest. Large populations of fakes can be computationally expensive to inject, search, and recover. One way to mitigate the number of fakes required to train the CNN would be to apply transfer learning, where the pre-trained weights of an old CNN are used as the initial state of training. This helps reduce the amount of training data required. Future work will investigate using transfer learning to quickly and efficiently update old CNNs to new data sets.

Applying the CNN filter before clustering (replacing the current stamp filter) may improve `KBMOD`. If a bright false positive makes it through the stamp filtering stage of processing to the clustering stage, it may cause nearby faint objects to be removed before human vetting. One way to mitigate this effect is to improve the performance of the stamp filter. We have demonstrated that CNN filtering can be highly effective when applied after the clustering stage. Future work will investigate the speed and effectiveness of replacing existing central moment stamp filters with CNN stamp filters.

We do not currently address the "look-elsewhere" effect in our search algorithm [e.g. 132]. The "look-elsewhere" effect describes the possibility of randomly obtaining a high

likelihood false positive because one investigated a very large search volume. However, our false positives are already dominated by image artifacts and real sources. Even after filtering, trajectories require human by-eye confirmation or rejection. Because of this requirement of human review, we consider this an acceptable limitation. Future work will further investigate necessary algorithmic improvements to enable machine-only object confirmation, including addressing the "look-elsewhere" effect.

Improving image astrometry and image differencing is likely to reduce the number of image differencing artifacts, thereby reducing the number of candidate trajectories requiring by-eye detection. In Chapter 3, the non-uniformity of the image time baseline in this survey means that artifacts appeared in approximately the same location in up to 5 images. This posed a unique challenge to filtering out artifacts from candidate trajectories. In Chapter 4, the time baseline and the use of a "mean" coaddition statistic created deep negative wells that required careful modelling to account for. Even after this modelling, these negative wells likely reduced our m_{25} by about 0.25 to 0.5 mags (depending on the velocity of the class of objects in question). Future applications of `KBMOD` should apply the lessons learned in this work, including trying and validating alternative coaddition statistics such as "median" or clipped mean.

Although more work remains (as is ever the case with scientific research), this work has demonstrated numerous novel improvements to `KBMOD`. It has validated `KBMOD` in difference imaging. It has characterized the efficiency of `KBMOD` on real data with a detailed population of implanted fakes. Finally, it has proven that `KBMOD` can detect real objects in a large multi-year survey with sufficient precision to fit detailed orbits and dynamically classify the detections. This work therefore demonstrates that `KBMOD` is a powerful detection tool capable of producing state-of-the-art science.

Software: `KBMOD` [136], LSST Science Pipelines [69], astropy [7], scikit-image [130], numpy [104], CUDA [100], scikit-learn [105], pandas [85], matplotlib [63], tensorflow [2]

BIBLIOGRAPHY

[1] *(588) Achilles*, pages 61–61. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[2] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensorflow: A system for large-scale machine learning. In *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, OSDI'16, page 265–283, USA, 2016. USENIX Association.

[3] C. Alard and Robert H. Lupton. A method for optimal image subtraction. *The Astrophysical Journal*, 503(1):325, 1998.

[4] Mike Alexandersen, Brett Gladman, Sarah Greenstreet, J. J. Kavelaars, Jean-Marc Petit, and Stephen Gwyn. A uranian trojan and the frequency of temporary giant-planet co-orbitals. *Science*, 341(6149):994–997, 2013.

[5] M. Ansdell, J. P. Williams, N. van der Marel, J. M. Carpenter, G. Guidi, M. Hogerheijde, G. S. Mathews, C. F. Manara, A. Miotello, A. Natta, I. Oliveira, M. Tazzari, L. Testi, E. F. van Dishoeck, and S. E. van Terwisga. ALMA SURVEY OF LUPUS PROTOPLANETARY DISKS. i. DUST AND GAS MASSES. *The Astrophysical Journal*, 828(1):46, aug 2016.

[6] Natalia A. Artemieva and Valery V. Shuvalov. From tunguska to chelyabinsk via jupiter. *Annual Review of Earth and Planetary Sciences*, 44(1):37–56, 2016.

[7] Astropy Collaboration, T. P. Robitaille, E. J. Tollerud, P. Greenfield, M. Droettboom,

E. Bray, T. Aldcroft, M. Davis, A. Ginsburg, A. M. Price-Whelan, W. E. Kerzendorf, A. Conley, N. Crighton, K. Barbary, D. Muna, H. Ferguson, F. Grollier, M. M. Parikh, P. H. Nair, H. M. Unther, C. Deil, J. Woillez, S. Conseil, R. Kramer, J. E. H. Turner, L. Singer, R. Fox, B. A. Weaver, V. Zabalza, Z. I. Edwards, K. Azalee Bostroem, D. J. Burke, A. R. Casey, S. M. Crawford, N. Dencheva, J. Ely, T. Jenness, K. Labrie, P. L. Lim, F. Pierfederici, A. Pontzen, A. Ptak, B. Refsdal, M. Servillat, and O. Streicher. Astropy: A community Python package for astronomy. *Astronomy and Astrophysics*, 558:A33, October 2013.

[8] Michele T. Bannister, Brett J. Gladman, J. J. Kavelaars, Jean-Marc Petit, Kathryn Volk, Ying-Tung Chen, Mike Alexandersen, Stephen D. J. Gwyn, Megan E. Schwamb, Edward Ashton, Susan D. Benecchi, Nahuel Cabral, Rebekah I. Dawson, Audrey Delsanti, Wesley C. Fraser, Mikael Granvik, Sarah Greenstreet, Aurélie Guilbert-Lepoutre, Wing-Huen Ip, Marian Jakubik, R. Lynne Jones, Nathan A. Kaib, Pedro Lacerda, Christa Van Laerhoven, Samantha Lawler, Matthew J. Lehner, Hsing Wen Lin, Patryk Sofia Lykawka, Michaël Marsset, Ruth Murray-Clay, Rosemary E. Pike, Philippe Rousselot, Cory Shankman, Audrey Thirouin, Pierre Vernazza, and Shiang-Yu Wang. OSSOS. VII. 800+ trans-neptunian objects—the complete data release. *The Astrophysical Journal Supplement Series*, 236(1):18, may 2018.

[9] Konstantin Batygin and Michael E. Brown. EVIDENCE FOR a DISTANT GIANT PLANET IN THE SOLAR SYSTEM. *The Astronomical Journal*, 151(2):22, jan 2016.

[10] Konstantin Batygin, Michael E. Brown, and Hayden Betts. INSTABILITY-DRIVEN DYNAMICAL EVOLUTION MODEL OF a PRIMORDIALLY FIVE-PLANET OUTER SOLAR SYSTEM. *The Astrophysical Journal*, 744(1):L3, dec 2011.

[11] J. C. Becker, T. Khain, S. J. Hamilton, F. C. Adams, D. W. Gerdes, L. Zullo, K. Franson, S. Millholland, G. M. Bernstein, M. Sako, P. Bernardinelli, K. Napier, L. Markwardt, Hsing Wen Lin, W. Wester, F. B. Abdalla, S. Allam, J. Annis, S. Avila,

E. Bertin, D. Brooks, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, C. Davis, J. De Vicente, H. T. Diehl, P. Doel, T. F. Eifler, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, E. Gaztanaga, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, W. G. Hartley, D. L. Hollowood, K. Honscheid, D. J. James, K. Kuehn, N. Kuropatkin, M. A. G. Maia, M. March, J. L. Marshall, F. Menanteau, R. Miquel, R. L. C. Ogando, A. A. Plazas, E. Sanchez, V. Scarpine, R. Schindler, I. Sevilla-Noarbe, M. Smith, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, and A. R. Walker and. Discovery and dynamical analysis of an extreme trans-neptunian object with a high orbital inclination. *The Astronomical Journal*, 156(2):81, aug 2018.

[12] Pedro Bernardinelli and The DEEP Collaboration. The DECam Ecliptic Exploration Project (DEEP) III: Survey simulation methods, 2022. Manuscript in Preparation.

[13] Pedro H. Bernardinelli, Gary M. Bernstein, Masao Sako, Tongtian Liu, William R. Saunders, Tali Khain, Hsing Wen Lin, David W. Gerdes, Dillon Brout, Fred C. Adams, Matthew Belyakov, Aditya Inada Somasundaram, Lakshay Sharma, Jennifer Locke, Kyle Franson, Juliette C. Becker, Kevin Napier, Larissa Markwardt, James Annis, T. M. C. Abbott, S. Avila, D. Brooks, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, F. J. Castander, L. N. da Costa, J. De Vicente, S. Desai, H. T. Diehl, P. Doel, S. Everett, B. Flaugher, J. García-Bellido, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, D. L. Hollowood, D. J. James, M. W. G. Johnson, M. D. Johnson, E. Krause, N. Kuropatkin, M. A. G. Maia, M. March, R. Miquel, F. Paz-Chinchón, A. A. Plazas, A. K. Romer, E. S. Rykoff, C. Sánchez, E. Sanchez, V. Scarpine, S. Serrano, I. Sevilla-Noarbe, M. Smith, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, A. R. Walker, W. Wester, and Y. Zhang and. Trans-neptunian objects found in the first four years of the dark energy survey. *The Astrophysical Journal Supplement Series*, 247(1):32, mar 2020.

[14] Pedro H. Bernardinelli, Gary M. Bernstein, Masao Sako, Brian Yanny, M. Aguena,

S. Allam, F. Andrade-Oliveira, E. Bertin, D. Brooks, E. Buckley-Geer, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, C. Conselice, M. Costanzi, L. N. da Costa, J. De Vicente, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, K. Eckert, S. Everett, I. Ferrero, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, D. W. Gerdes, D. Gruen, R. A. Gruendl, J. Gschwend, S. R. Hinton, D. L. Hollowood, K. Honscheid, D. J. James, S. Kent, K. Kuehn, N. Kuropatkin, O. Lahav, M. A. G. Maia, M. March, F. Menanteau, R. Miquel, R. Morgan, J. Myles, R. L. C. Ogando, A. Palmese, F. Paz-Chinchón, A. Pieres, A. A. Plazas Malagón, A. K. Romer, A. Roodman, E. Sanchez, V. Scarpine, M. Schubnell, S. Serrano, I. Sevilla-Noarbe, M. Smith, M. Soares-Santos, E. Suchyta, M. E. C. Swanson, G. Tarle, C. To, T. N. Varga, and A. R. Walker. A search of the full six years of the dark energy survey for outer solar system objects. *The Astrophysical Journal Supplement Series*, 258(2):41, feb 2022.

[15] G. M. Bernstein, D. E. Trilling, R. L. Allen, M. E. Brown, M. Holman, and R. Malhotra. The size distribution of trans-neptunian bodies. *The Astronomical Journal*, 128(3):1364–1390, sep 2004.

[16] Gary Bernstein and Bharat Khushalani. Orbit fitting and uncertainties for kuiper belt objects. *The Astronomical journal*, 120(6):3323–3332, 2000.

[17] J. Berthier, F. Vachier, W. Thuillot, P. Fernique, F. Ochsenbein, F. Genova, V. Lainey, and J.-E. Arlot. SkyBoT, a new VO service to identify Solar System objects. In C. Gabriel, C. Arviset, D. Ponz, and S. Enrique, editors, *Astronomical Data Analysis Software and Systems XV*, volume 351 of *Astronomical Society of the Pacific Conference Series*, page 367, July 2006.

[18] Stevo Bozinovski and Ante Fulgosi. The influence of pattern similarity and transfer learning upon training of a base perceptron b2. In *Proceedings of Symposium Informatica*, volume 3, pages 121–126, 1976. Original in Croatian.

[19] R. Brasser, M.J. Duncan, and H.F. Levison. Embedded star clusters and the formation of the oort cloud. *Icarus*, 184(1):59–82, 2006.

[20] R. Brasser, M.J. Duncan, H.F. Levison, M.E. Schwamb, and M.E. Brown. Reassessing the formation of the inner oort cloud in an embedded star cluster. *Icarus*, 217(1):1–19, 2012.

[21] M. E. Brown. The Inclination Distribution of the Kuiper Belt. *Astronomical Journal*, 121:2804–2814, May 2001.

[22] Michael E. Brown. The compositions of kuiper belt objects. *Annual Review of Earth and Planetary Sciences*, 40(1):467–494, 2012.

[23] Michael E. Brown, Chadwick Trujillo, and David Rabinowitz. Discovery of a candidate inner oort cloud planetoid. *The Astrophysical Journal*, 617(1):645–649, dec 2004.

[24] Bradley W. Carroll and Dale A. Ostile. An introduction to modern astrophysics, 2007.

[25] S. Chandrasekhar. Dynamical Friction. I. General Considerations: the Coefficient of Dynamical Friction. *Astrophysical Journal*, 97:255, March 1943.

[26] Kumar Chellapilla, Sidd Puri, and Patrice Simard. High Performance Convolutional Neural Networks for Document Processing. In Guy Lorette, editor, *Tenth International Workshop on Frontiers in Handwriting Recognition*, La Baule (France), October 2006. Université de Rennes 1, Suvisoft. http://www.suvisoft.com.

[27] Apostolos A. Christou and Paul Wiegert. A population of main belt asteroids co-orbiting with ceres and vesta. *Icarus*, 217(1):27–42, 2012.

[28] Martin Connors, Paul Wiegert, and Christian Veillet. Earth's trojan asteroid. *Nature*, 475(7357):481–483, Jul 2011.

[29] Nicolaus Copernicus. De revolutionibus orbium coelestium, 1543.

[30] Howard D. Curtis. Chapter 5 - preliminary orbit determination. In Howard D. Curtis, editor, *Orbital Mechanics for Engineering Students (Third Edition)*, pages 239–298. Butterworth-Heinemann, Boston, third edition edition, 2014.

[31] C. de la Fuente Marcos and R. de la Fuente Marcos. Asteroid 2014 YX49: a large transient Trojan of Uranus. *Monthly Notices of the Royal Astronomical Society*, 467(2):1561–1568, 01 2017.

[32] Steve Desch, Alan Jackson, Jessica Noviello, and Ariel Anbar. The Chicxulub impactor: comet or asteroid? *Astronomy & Geophysics*, 62(3):3.34–3.37, 06 2021.

[33] Michael J. DRAKE. Origin of water in the terrestrial planets. *Meteoritics & Planetary Science*, 40(4):519–527, 2005.

[34] Michael J. Drake and Kevin Righter. Determining the composition of the earth. *Nature*, 416(6876):39–44, Mar 2002.

[35] Martin J. Duncan and Harold F. Levison. A disk of scattered icy objects and the origin of jupiter-family comets. *Science*, 276:1670+, Jun 1997. 5319.

[36] K. E. Edgeworth. The Origin and Evolution of the Solar System. *Monthly Notices of the Royal Astronomical Society*, 109(5):600–609, 10 1949.

[37] B. Flaugher, H. T. Diehl, K. Honscheid, T. M. C. Abbott, O. Alvarez, R. Angstadt, J. T. Annis, M. Antonik, O. Ballester, L. Beaufore, G. M. Bernstein, R. A. Bernstein, B. Bigelow, M. Bonati, D. Boprie, D. Brooks, E. J. Buckley-Geer, J. Campa, L. Cardiel-Sas, F. J. Castander, J. Castilla, H. Cease, J. M. Cela-Ruiz, S. Chappa, E. Chi, C. Cooper, L. N. da Costa, E. Dede, G. Derylo, D. L. DePoy, J. de Vicente, P. Doel, A. Drlica-Wagner, J. Eiting, A. E. Elliott, J. Emes, J. Estrada, A. Fausti Neto, D. A. Finley, R. Flores, J. Frieman, D. Gerdes, M. D. Gladders, B. Gregory, G. R. Gutierrez, J. Hao, S. E. Holland, S. Holm, D. Huffman, C. Jackson, D. J. James, M. Jonas, A. Karcher, I. Karliner, S. Kent, R. Kessler, M. Kozlovsky, R. G. Kron,

D. Kubik, K. Kuehn, S. Kuhlmann, K. Kuk, O. Lahav, A. Lathrop, J. Lee, M. E. Levi, P. Lewis, T. S. Li, I. Mandrichenko, J. L. Marshall, G. Martinez, K. W. Merritt, R. Miquel, F. Muñoz, E. H. Neilsen, R. C. Nichol, B. Nord, R. Ogando, J. Olsen, N. Palaio, K. Patton, J. Peoples, A. A. Plazas, J. Rauch, K. Reil, J.-P. Rheault, N. A. Roe, H. Rogers, A. Roodman, E. Sanchez, V. Scarpine, R. H. Schindler, R. Schmidt, R. Schmitt, M. Schubnell, K. Schultz, P. Schurter, L. Scott, S. Serrano, T. M. Shaw, R. C. Smith, M. Soares-Santos, A. Stefanik, W. Stuermer, E. Suchyta, A. Sypniewski, G. Tarle, J. Thaler, R. Tighe, C. Tran, D. Tucker, A. R. Walker, G. Wang, M. Watson, C. Weaverdyck, W. Wester, R. Woods, B. Yanny, and The DES Collaboration. The dark energy camera. *The Astronomical Journal*, 150(5):150, 2015.

[38] H. A. Flewelling, E. A. Magnier, K. C. Chambers, J. N. Heasley, C. Holmberg, M. E. Huber, W. Sweeney, C. Z. Waters, A. Calamida, S. Casertano, X. Chen, D. Farrow, G. Hasinger, R. Henderson, K. S. Long, N. Metcalfe, G. Narayan, M. A. Nieto-Santisteban, P. Norberg, A. Rest, R. P. Saglia, A. Szalay, A. R. Thakar, J. L. Tonry, J. Valenti, S. Werner, R. White, L. Denneau, P. W. Draper, K. W. Hodapp, R. Jedicke, N. Kaiser, R. P. Kudritzki, P. A. Price, R. J. Wainscoat, S. Chastel, B. McLean, M. Postman, and B. Shiao. The pan-STARRS1 database and data products. *The Astrophysical Journal Supplement Series*, 251(1):7, oct 2020.

[39] Eric G. Forbes. Gauss and the Discovery of Ceres. *Journal for the History of Astronomy*, 2:195, January 1971.

[40] F. Förster, J. C. Maureira, J. San Martín, M. Hamuy, J. Martínez, P. Huijse, G. Cabrera, L. Galbany, T. de Jaeger, S. González-Gaitán, J. P. Anderson, H. Kunkarayakti, G. Pignata, F. Bufano, J. Littín, F. Olivares, G. Medina, R. C. Smith, A. K. Vivas, P. A. Estévez, R. Muñoz, and E. Vera. The High Cadence Transient Survey (HITS). I. Survey Design and Supernova Shock Breakout Constraints. *Astrophysical Journal*, 832:155, December 2016.

[41] Wesley C. Fraser, Michele T. Bannister, Rosemary E. Pike, Michael Marsset, Megan E. Schwamb, J. J. Kavelaars, Pedro Lacerda, David Nesvorný, Kathryn Volk, Audrey Delsanti, Susan Benecchi, Matthew J. Lehner, Keith Noll, Brett Gladman, Jean-Marc Petit, Stephen Gwyn, Ying-Tung Chen, Shiang-Yu Wang, Mike Alexandersen, Todd Burdullis, Scott Sheppard, and Chad Trujillo. All planetesimals born near the kuiper belt formed as binaries. *Nature Astronomy*, 1(4):0088, Apr 2017.

[42] Wesley C. Fraser, Michael E. Brown, Alessandro Morbidelli, Alex Parker, and Konstantin Batygin. THE ABSOLUTE MAGNITUDE DISTRIBUTION OF KUIPER BELT OBJECTS. *The Astrophysical Journal*, 782(2):100, feb 2014.

[43] Wesley C. Fraser, J.J. Kavelaars, M.J. Holman, C.J. Pritchet, B.J. Gladman, T. Grav, R.L. Jones, J. MacWilliams, and J.-M. Petit. The kuiper belt luminosity function from mr=21 to 26. *Icarus*, 195(2):827 – 843, 2008.

[44] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202, Apr 1980.

[45] Kunihiko Fukushima, Sei Miyake, and Takayuki Ito. Neocognitron: A neural network model for a mechanism of visual pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(5):826–834, 1983.

[46] T. Gehrels, T. Coffeen, and D. Owings. Wavelength dependance of polarization. III. The lunar surface. *Astronomical Journal*, 69:826, December 1964.

[47] J. D. Giorgini. Status of the JPL Horizons Ephemeris System. *IAU General Assembly*, 22:2256293, August 2015.

[48] B. Gladman and C. Chan. Production of the Extended Scattered Disk by Rogue Planets. *Astrophysical Journal, Letters*, 643:L135–L138, June 2006.

[49] B. Gladman, S. M. Lawler, J.-M. Petit, J. Kavelaars, R. L. Jones, J. Wm. Parker, C. Van Laerhoven, P. Nicholson, P. Rousselot, A. Bieryla, and M. L. N. Ashby. THE RESONANT TRANS-NEPTUNIAN POPULATIONS. *The Astronomical Journal*, 144(1):23, jun 2012.

[50] B. Gladman, B. G. Marsden, and C. Vanlaerhoven. Nomenclature in the Outer Solar System. In M. A. Barucci, H. Boehnhardt, D. P. Cruikshank, A. Morbidelli, and Renee Dotson, editors, *The Solar System Beyond Neptune*, page 43. 2008.

[51] Brett Gladman and Kathryn Volk. Transneptunian space. *Annual Review of Astronomy and Astrophysics*, 59(1):203–246, 2021.

[52] R. Gomes, H. F. Levison, K. Tsiganis, and A. Morbidelli. Origin of the cataclysmic late heavy bombardment period of the terrestrial planets. *Nature*, 435(7041):466–469, May 2005.

[53] Rodney S Gomes. The origin of the kuiper belt high–inclination population. *Icarus*, 161(2):404–418, 2003.

[54] Rodney S. Gomes, Tabaré Gallardo, Julio A. Fernández, and Adrián Brunini. On the origin of the high-perihelion scattered disk: The role of the kozai mechanism and mean motion resonances. *Celestial Mechanics and Dynamical Astronomy*, 91(1):109–129, Jan 2005.

[55] Gomes, R. and Nesvorný, D. Neptune trojan formation during planetary instability and migration. *A&A*, 592:A146, 2016.

[56] Daniel W Graham. Anaxagoras and the comet. *Ancient philosophy (Pittsburgh, Pa.)*, 33(1):1–18, 2013.

[57] J. M. Hahn and R. Malhotra. Neptune's Migration into a Stirred-Up Kuiper Belt: A Detailed Comparison of Simulations to Observations. *Astronomical Journal*, 130:2392–2414, November 2005.

[58] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[59] D. O. (Donald Olding) Hebb. The organization of behavior : a neuropsychological theory, 1949.

[60] Jacques Henrard and Anne Lemaître. A mechanism of formation for the kirkwood gaps. *Icarus*, 55(3):482–494, 1983.

[61] K. Herner, J. Annis, D. Brout, M. Soares-Santos, R. Kessler, M. Sako, R. Butler, Z. Doctor, A. Palmese, S. Allam, D.L. Tucker, F. Sobreira, B. Yanny, H.T. Diehl, J. Frieman, N. Glaeser, A. Garcia, N.F. Sherman, K. Bechtol, E. Berger, H.Y. Chen, C.J. Conselice, E. Cook, P.S. Cowperthwaite, T.M. Davis, A. Drlica-Wagner, B. Farr, D. Finley, R.J. Foley, J. Garcia-Bellido, M.S.S. Gill, R.A. Gruendl, D.E. Holz, N. Kuropatkin, H. Lin, J. Marriner, J.L. Marshall, T. Matheson, E. Neilsen, F. Paz-Chinchón, M. Sauseda, D. Scolnic, P.K.G. Williams, S. Avila, E. Bertin, E. Buckley-Geer, D.L. Burke, A. Carnero Rosell, M. Carrasco-Kind, J. Carretero, L.N. da Costa, J. De Vicente, S. Desai, P. Doel, T.F. Eifler, S. Everett, P. Fosalba, E. Gaztanaga, D.W. Gerdes, J. Gschwend, G. Gutierrez, W.G. Hartley, D.L. Hollowood, K. Honscheid, D.J. James, E. Krause, K. Kuehn, O. Lahav, T.S. Li, M. Lima, M.A.G. Maia, M. March, F. Menanteau, R. Miquel, A.A. Plazas, E. Sanchez, V. Scarpine, M. Schubnell, S. Serrano, I. Sevilla-Noarbe, M. Smith, E. Suchyta, G. Tarle, W. Wester, and Y. Zhang. Optical follow-up of gravitational wave triggers with decam during the first two ligo/virgo observing runs. *Astronomy and Computing*, 33:100425, 2020.

[62] Matthew J. Holman, Matthew J. Payne, Paul Blankley, Ryan Janssen, and Scott Kuindersma. HelioLinC: A novel approach to the minor planet linking problem. *The Astronomical Journal*, 156(3):135, aug 2018.

[63] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007.

[64] Ž. Ivezić, A.J. Connolly, J.T. Vanderplas, and A. Gray. *Statistics, Data Mining and Machine Learning in Astronomy*. Princeton University Press, 2014.

[65] Željko Ivezić, Steven M. Kahn, J. Anthony Tyson, Bob Abel, Emily Acosta, Robyn Allsman, David Alonso, Yusra AlSayyad, Scott F. Anderson, John Andrew, James Roger P. Angel, George Z. Angeli, Reza Ansari, Pierre Antilogus, Constanza Araujo, Robert Armstrong, Kirk T. Arndt, Pierre Astier, Éric Aubourg, Nicole Auza, Tim S. Axelrod, Deborah J. Bard, Jeff D. Barr, Aurelian Barrau, James G. Bartlett, Amanda E. Bauer, Brian J. Bauman, Sylvain Baumont, Ellen Bechtol, Keith Bechtol, Andrew C. Becker, Jacek Becla, Cristina Beldica, Steve Bellavia, Federica B. Bianco, Rahul Biswas, Guillaume Blanc, Jonathan Blazek, Roger D. Blandford, Josh S. Bloom, Joanne Bogart, Tim W. Bond, Michael T. Booth, Anders W. Borgland, Kirk Borne, James F. Bosch, Dominique Boutigny, Craig A. Brackett, Andrew Bradshaw, William Nielsen Brandt, Michael E. Brown, James S. Bullock, Patricia Burchat, David L. Burke, Gianpietro Cagnoli, Daniel Calabrese, Shawn Callahan, Alice L. Callen, Jeffrey L. Carlin, Erin L. Carlson, Srinivasan Chandrasekharan, Glenaver Charles-Emerson, Steve Chesley, Elliott C. Cheu, Hsin-Fang Chiang, James Chiang, Carol Chirino, Derek Chow, David R. Ciardi, Charles F. Claver, Johann Cohen-Tanugi, Joseph J. Cockrum, Rebecca Coles, Andrew J. Connolly, Kem H. Cook, Asantha Cooray, Kevin R. Covey, Chris Cribbs, Wei Cui, Roc Cutri, Philip N. Daly, Scott F. Daniel, Felipe Daruich, Guillaume Daubard, Greg Daues, William Dawson, Francisco Delgado, Alfred Dellapenna, Robert de Peyster, Miguel de Val-Borro, Seth W. Digel, Peter Doherty, Richard Dubois, Gregory P. Dubois-Felsmann, Josef Durech, Frossie Economou, Tim Eifler, Michael Eracleous, Benjamin L. Emmons, Angelo Fausti Neto, Henry Ferguson, Enrique Figueroa, Merlin Fisher-Levine, Warren Focke, Michael D. Foss, James Frank, Michael D. Freemon, Emmanuel Gangler, Eric Gawiser, John C.

Geary, Perry Gee, Marla Geha, Charles J. B. Gessner, Robert R. Gibson, D. Kirk Gilmore, Thomas Glanzman, William Glick, Tatiana Goldina, Daniel A. Goldstein, Iain Goodenow, Melissa L. Graham, William J. Gressler, Philippe Gris, Leanne P. Guy, Augustin Guyonnet, Gunther Haller, Ron Harris, Patrick A. Hascall, Justine Haupt, Fabio Hernandez, Sven Herrmann, Edward Hileman, Joshua Hoblitt, John A. Hodgson, Craig Hogan, James D. Howard, Dajun Huang, Michael E. Huffer, Patrick Ingraham, Walter R. Innes, Suzanne H. Jacoby, Bhuvnesh Jain, Fabrice Jammes, James Jee, Tim Jenness, Garrett Jernigan, Darko Jevremović, Kenneth Johns, Anthony S. Johnson, Margaret W. G. Johnson, R. Lynne Jones, Claire Juramy-Gilles, Mario Jurić, Jason S. Kalirai, Nitya J. Kallivayalil, Bryce Kalmbach, Jeffrey P. Kantor, Pierre Karst, Mansi M. Kasliwal, Heather Kelly, Richard Kessler, Veronica Kinnison, David Kirkby, Lloyd Knox, Ivan V. Kotov, Victor L. Krabbendam, K. Simon Krughoff, Petr Kubánek, John Kuczewski, Shri Kulkarni, John Ku, Nadine R. Kurita, Craig S. Lage, Ron Lambert, Travis Lange, J. Brian Langton, Laurent Le Guillou, Deborah Levine, Ming Liang, Kian-Tat Lim, Chris J. Lintott, Kevin E. Long, Margaux Lopez, Paul J. Lotz, Robert H. Lupton, Nate B. Lust, Lauren A. MacArthur, Ashish Mahabal, Rachel Mandelbaum, Thomas W. Markiewicz, Darren S. Marsh, Philip J. Marshall, Stuart Marshall, Morgan May, Robert McKercher, Michelle McQueen, Joshua Meyers, Myriam Migliore, Michelle Miller, David J. Mills, Connor Miraval, Joachim Moeyens, Fred E. Moolekamp, David G. Monet, Marc Moniez, Serge Monkewitz, Christopher Montgomery, Christopher B. Morrison, Fritz Mueller, Gary P. Muller, Freddy Muñoz Arancibia, Douglas R. Neill, Scott P. Newbry, Jean-Yves Nief, Andrei Nomerotski, Martin Nordby, Paul O'Connor, John Oliver, Scot S. Olivier, Knut Olsen, William O'Mullane, Sandra Ortiz, Shawn Osier, Russell E. Owen, Reynald Pain, Paul E. Palecek, John K. Parejko, James B. Parsons, Nathan M. Pease, J. Matt Peterson, John R. Peterson, Donald L. Petravick, M. E. Libby Petrick, Cathy E. Petry, Francesco Pierfederici, Stephen Pietrowicz, Rob Pike, Philip A. Pinto, Raymond Plante, Stephen Plate, Joel P. Plutchak, Paul A. Price, Michael Prouza, Veljko Radeka, Jayadev Rajagopal,

Andrew P. Rasmussen, Nicolas Regnault, Kevin A. Reil, David J. Reiss, Michael A. Reuter, Stephen T. Ridgway, Vincent J. Riot, Steve Ritz, Sean Robinson, William Roby, Aaron Roodman, Wayne Rosing, Cecille Roucelle, Matthew R. Rumore, Stefano Russo, Abhijit Saha, Benoit Sassolas, Terry L. Schalk, Pim Schellart, Rafe H. Schindler, Samuel Schmidt, Donald P. Schneider, Michael D. Schneider, William Schoening, German Schumacher, Megan E. Schwamb, Jacques Sebag, Brian Selvy, Glenn H. Sembroski, Lynn G. Seppala, Andrew Serio, Eduardo Serrano, Richard A. Shaw, Ian Shipsey, Jonathan Sick, Nicole Silvestri, Colin T. Slater, J. Allyn Smith, R. Chris Smith, Shahram Sobhani, Christine Soldahl, Lisa Storrie-Lombardi, Edward Stover, Michael A. Strauss, Rachel A. Street, Christopher W. Stubbs, Ian S. Sullivan, Donald Sweeney, John D. Swinbank, Alexander Szalay, Peter Takacs, Stephen A. Tether, Jon J. Thaler, John Gregg Thayer, Sandrine Thomas, Adam J. Thornton, Vaikunth Thukral, Jeffrey Tice, David E. Trilling, Max Turri, Richard Van Berg, Daniel Vanden Berk, Kurt Vetter, Francoise Virieux, Tomislav Vucina, William Wahl, Lucianne Walkowicz, Brian Walsh, Christopher W. Walter, Daniel L. Wang, Shin-Yawn Wang, Michael Warner, Oliver Wiecha, Beth Willman, Scott E. Winters, David Wittman, Sidney C. Wolff, W. Michael Wood-Vasey, Xiuqin Wu, Bo Xin, Peter Yoachim, and Hu Zhan. LSST: From science drivers to reference design and anticipated data products. *The Astrophysical Journal*, 873(2):111, mar 2019.

[66] David Jewitt and Jane Luu. Discovery of the candidate kuiper belt object 1992 qb1. *Nature*, 362(6422):730–732, Apr 1993.

[67] R.L. Jones, B. Gladman, J.-M. Petit, P. Rousselot, O. Mousis, J.J. Kavelaars, A. Campo Bagatin, G. Bernabeu, P. Benavidez, J.Wm. Parker, P. Nicholson, M. Holman, T. Grav, A. Doressoundiram, C. Veillet, H. Scholl, and G. Mars. The cfeps kuiper belt survey: Strategy and presurvey results. *Icarus*, 185(2):508–522, 2006.

[68] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna

Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, Aug 2021.

[69] M. Jurić, J. Kantor, K. T. Lim, R. H. Lupton, G. Dubois-Felsmann, T. Jenness, T. S. Axelrod, J. Aleksić, R. A. Allsman, Y. AlSayyad, J. Alt, R. Armstrong, J. Basney, A. C. Becker, J. Becla, R. Biswas, J. Bosch, D. Boutigny, M. C. Kind, D. R. Ciardi, A. J. Connolly, S. F. Daniel, G. E. Daues, F. Economou, H. F. Chiang, A. Fausti, M. Fisher-Levine, D. M. Freemon, P. Gris, F. Hernandez, J. Hoblitt, Z. Ivezić, F. Jammes, D. Jevremović, R. L. Jones, J. B. Kalmbach, V. P. Kasliwal, K. S. Krughoff, J. Lurie, N. B. Lust, L. A. MacArthur, P. Melchior, J. Moeyens, D. L. Nidever, R. Owen, J. K. Parejko, J. M. Peterson, D. Petravick, S. R. Pietrowicz, P. A. Price, D. J. Reiss, R. A. Shaw, J. Sick, C. T. Slater, M. A. Strauss, I. S. Sullivan, J. D. Swinbank, S. Van Dyk, V. Vujčić, A. Withers, and P. Yoachim. The LSST Data Management System. In N. P. F. Lorente, K. Shortridge, and R. Wayth, editors, *Astronomical Data Analysis Software and Systems XXV*, volume 512 of *Astronomical Society of the Pacific Conference Series*, page 279, Dec 2017.

[70] Nathan A. Kaib, Rok Roškar, and Thomas Quinn. Sedna and the oort cloud around a migrating sun. *Icarus*, 215(2):491 – 507, 2011.

[71] Robert E. Kass and Adrian E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 2022/08/01/ 1995. Full publication date: Jun., 1995.

[72] S. J. Kenyon and B. C. Bromley. Stellar encounters as the origin of distant Solar System objects in highly eccentric orbits. *Nature*, 432:598–602, December 2004.

[73] Johannes Kepler. Astronomia nova, 1609.

[74] T. Khain, J. C. Becker, Hsing Wen Lin, D. W. Gerdes, F. C. Adams, P. Bernardinelli, G. M. Bernstein, K. Franson, L. Markwardt, S. Hamilton, K. Napier, M. Sako, T. M. C. Abbott, S. Avila, E. Bertin, D. Brooks, E. Buckley-Geer, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, L. N. da Costa, J. De Vicente, S. Desai, H. T. Diehl, P. Doel, B. Flaugher, J. Frieman, J. García-Bellido, E. Gaztanaga, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, D. L. Hollowood, K. Honscheid, D. J. James, N. Kuropatkin, M. A. G. Maia, J. L. Marshall, F. Menanteau, C. J. Miller, R. Miquel, A. A. Plazas, E. Sanchez, V. Scarpine, M. Schubnell, I. Sevilla-Noarbe, M. Smith, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, A. R. Walker, and W. Wester and. Dynamical classification of trans-neptunian objects detected by the dark energy survey. *The Astronomical Journal*, 159(4):133, feb 2020.

[75] G. P. Kuiper. On the evolution of the protoplanets. *Proceedings of the National Academy of Sciences of the United States of America*, 37(7):383–393, Jul 1951. 16578372[pmid].

[76] G. P. Kuiper. On the origin of the solar system. *Proceedings of the National Academy of Sciences of the United States of America*, 37(1):1–14, Jan 1951. 16588984[pmid].

[77] Joseph Louis Lagrange. Essai sur le problème des trois corps. *Oeuvres de Lagrange*, Tome 6:229–332, 1772.

[78] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[79] Lindegren, L., Hernández, J., Bombrun, A., Klioner, S., Bastian, U., Ramos-Lerate, M., de Torres, A., Steidelmüller, H., Stephenson, C., Hobbs, D., Lammers, U., Bier-

mann, M., Geyer, R., Hilger, T., Michalik, D., Stampa, U., McMillan, P.J., Castañeda, J., Clotet, M., Comoretto, G., Davidson, M., Fabricius, C., Gracia, G., Hambly, N.C., Hutton, A., Mora, A., Portell, J., van Leeuwen, F., Abbas, U., Abreu, A., Altmann, M., Andrei, A., Anglada, E., Balaguer-Núñez, L., Barache, C., Becciani, U., Bertone, S., Bianchi, L., Bouquillon, S., Bourda, G., Brüsemeister, T., Bucciarelli, B., Busonero, D., Buzzi, R., Cancelliere, R., Carlucci, T., Charlot, P., Cheek, N., Crosta, M., Crowley, C., de Bruijne, J., de Felice, F., Drimmel, R., Esquej, P., Fienga, A., Fraile, E., Gai, M., Garralda, N., González-Vidal, J.J., Guerra, R., Hauser, M., Hofmann, W., Holl, B., Jordan, S., Lattanzi, M.G., Lenhardt, H., Liao, S., Licata, E., Lister, T., Löffler, W., Marchant, J., Martin-Fleitas, J.-M., Messineo, R., Mignard, F., Morbidelli, R., Poggio, E., Riva, A., Rowell, N., Salguero, E., Sarasso, M., Sciacca, E., Siddiqui, H., Smart, R.L., Spagna, A., Steele, I., Taris, F., Torra, J., van Elteren, A., van Reeven, W., and Vecchiato, A. Gaia data release 2 - the astrometric solution. *A&A*, 616:A2, 2018.

[80] LSST Science Collaboration, Paul A. Abell, Julius Allison, Scott F. Anderson, John R. Andrew, J. Roger P. Angel, Lee Armus, David Arnett, S. J. Asztalos, Tim S. Axelrod, Stephen Bailey, D. R. Ballantyne, Justin R. Bankert, Wayne A. Barkhouse, Jeffrey D. Barr, L. Felipe Barrientos, Aaron J. Barth, James G. Bartlett, Andrew C. Becker, Jacek Becla, Timothy C. Beers, Joseph P. Bernstein, Rahul Biswas, Michael R. Blanton, Joshua S. Bloom, John J. Bochanski, Pat Boeshaar, Kirk D. Borne, Marusa Bradac, W. N. Brandt, Carrie R. Bridge, Michael E. Brown, Robert J. Brunner, James S. Bullock, Adam J. Burgasser, James H. Burge, David L. Burke, Phillip A. Cargile, Srinivasan Chandrasekharan, George Chartas, Steven R. Chesley, You-Hua Chu, David Cinabro, Mark W. Claire, Charles F. Claver, Douglas Clowe, A. J. Connolly, Kem H. Cook, Jeff Cooke, Asantha Cooray, Kevin R. Covey, Christopher S. Culliton, Roelof de Jong, Willem H. de Vries, Victor P. Debattista, Francisco Delgado, Ian P. Dell'Antonio, Saurav Dhital, Rosanne Di Stefano, Mark Dickinson, Benjamin

Dilday, S. G. Djorgovski, Gregory Dobler, Ciro Donalek, Gregory Dubois-Felsmann, Josef Durech, Ardis Eliasdottir, Michael Eracleous, Laurent Eyer, Emilio E. Falco, Xiaohui Fan, Christopher D. Fassnacht, Harry C. Ferguson, Yanga R. Fernandez, Brian D. Fields, Douglas Finkbeiner, Eduardo E. Figueroa, Derek B. Fox, Harold Francke, James S. Frank, Josh Frieman, Sebastien Fromenteau, Muhammad Furqan, Gaspar Galaz, A. Gal-Yam, Peter Garnavich, Eric Gawiser, John Geary, Perry Gee, Robert R. Gibson, Kirk Gilmore, Emily A. Grace, Richard F. Green, William J. Gressler, Carl J. Grillmair, Salman Habib, J. S. Haggerty, Mario Hamuy, Alan W. Harris, Suzanne L. Hawley, Alan F. Heavens, Leslie Hebb, Todd J. Henry, Edward Hileman, Eric J. Hilton, Keri Hoadley, J. B. Holberg, Matt J. Holman, Steve B. Howell, Leopoldo Infante, Zeljko Ivezic, Suzanne H. Jacoby, Bhuvnesh Jain, R, Jedicke, M. James Jee, J. Garrett Jernigan, Saurabh W. Jha, Kathryn V. Johnston, R. Lynne Jones, Mario Juric, Mikko Kaasalainen, Styliani, Kafka, Steven M. Kahn, Nathan A. Kaib, Jason Kalirai, Jeff Kantor, Mansi M. Kasliwal, Charles R. Keeton, Richard Kessler, Zoran Knezevic, Adam Kowalski, Victor L. Krabbendam, K. Simon Krughoff, Shrinivas Kulkarni, Stephen Kuhlman, Mark Lacy, Sebastien Lepine, Ming Liang, Amy Lien, Paulina Lira, Knox S. Long, Suzanne Lorenz, Jennifer M. Lotz, R. H. Lupton, Julie Lutz, Lucas M. Macri, Ashish A. Mahabal, Rachel Mandelbaum, Phil Marshall, Morgan May, Peregrine M. McGehee, Brian T. Meadows, Alan Meert, Andrea Milani, Christopher J. Miller, Michelle Miller, David Mills, Dante Minniti, David Monet, Anjum S. Mukadam, Ehud Nakar, Douglas R. Neill, Jeffrey A. Newman, Sergei Nikolaev, Martin Nordby, Paul O'Connor, Masamune Oguri, John Oliver, Scot S. Olivier, Julia K. Olsen, Knut Olsen, Edward W. Olszewski, Hakeem Oluseyi, Nelson D. Padilla, Alex Parker, Joshua Pepper, John R. Peterson, Catherine Petry, Philip A. Pinto, James L. Pizagno, Bogdan Popescu, Andrej Prsa, Veljko Radcka, M. Jordan Raddick, Andrew Rasmussen, Arne Rau, Jeonghee Rho, James E. Rhoads, Gordon T. Richards, Stephen T. Ridgway, Brant E. Robertson, Rok Roskar, Abhijit Saha, Ata Sarajedini, Evan Scannapieco, Terry Schalk, Rafe Schindler, Samuel Schmidt, Sarah

Schmidt, Donald P. Schneider, German Schumacher, Ryan Scranton, Jacques Sebag, Lynn G. Seppala, Ohad Shemmer, Joshua D. Simon, M. Sivertz, Howard A. Smith, J. Allyn Smith, Nathan Smith, Anna H. Spitz, Adam Stanford, Keivan G. Stassun, Jay Strader, Michael A. Strauss, Christopher W. Stubbs, Donald W. Sweeney, Alex Szalay, Paula Szkody, Masahiro Takada, Paul Thorman, David E. Trilling, Virginia Trimble, Anthony Tyson, Richard Van Berg, Daniel Vanden Berk, Jake VanderPlas, Licia Verde, Bojan Vrsnak, Lucianne M. Walkowicz, Benjamin D. Wandelt, Sheng Wang, Yun Wang, Michael Warner, Risa H. Wechsler, Andrew A. West, Oliver Wiecha, Benjamin F. Williams, Beth Willman, David Wittman, Sidney C. Wolff, W. Michael Wood-Vasey, Przemek Wozniak, Patrick Young, Andrew Zentner, and Hu Zhan. Lsst science book, version 2.0, 2009.

[81] Jane X. Luu and David C. Jewitt. Kuiper belt objects: Relics from the accretion disk of the sun. *Annual Review of Astronomy and Astrophysics*, 40(1):63–101, 2002.

[82] Renu Malhotra. The origin of pluto's peculiar orbit. *Nature*, 365(6449):819–821, Oct 1993.

[83] Bernard Marty, Guillaume Avice, Yuji Sano, Kathrin Altwegg, Hans Balsiger, Myrtha Hässig, Alessandro Morbidelli, Olivier Mousis, and Martin Rubin. Origins of volatile elements (h, c, n, noble gases) on earth and mars in light of recent results from the rosetta cometary mission. *Earth and Planetary Science Letters*, 441:91–102, May 2016.

[84] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology*, 52(1):99–115, 1990.

[85] Wes McKinney. Data structures for statistical computing in python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 51 – 56, 2010.

[86] Joachim Moeyens, Mario Jurić, Jes Ford, Dino Bektešević, Andrew J. Connolly,

Siegfried Eggl, Željko Ivezić, R. Lynne Jones, J. Bryce Kalmbach, and Hayden Smotherman. THOR: An algorithm for cadence-independent asteroid discovery. *The Astronomical Journal*, 162(4):143, sep 2021.

[87] A. Morbidelli, J. Chambers, J. I. Lunine, J. M. Petit, F. Robert, G. B. Valsecchi, and K. E. Cyr. Source regions and timescales for the delivery of water to the earth. *Meteoritics & Planetary Science*, 35(6):1309–1320, 2000.

[88] Alessandro Morbidelli and David Nesvorný. Chapter 2 - kuiper belt: Formation and evolution. In Dina Prialnik, M. Antonietta Barucci, and Leslie A. Young, editors, *The Trans-Neptunian Solar System*, pages 25–59. Elsevier, 2020.

[89] Alessandro Morbidelli, Kleomenis Tsiganis, Aurélien Crida, Harold F. Levison, and Rodney Gomes. Dynamics of the giant planets of the solar system in the gaseous protoplanetary disk and their relationship to the current orbital architecture. *The Astronomical Journal*, 134(5):1790–1798, sep 2007.

[90] E. Morganson, R. A. Gruendl, F. Menanteau, M. Carrasco Kind, Y.-C. Chen, G. Daues, A. Drlica-Wagner, D. N. Friedel, M. Gower, M. W. G. Johnson, M. D. Johnson, R. Kessler, F. Paz-Chinchón, D. Petravick, C. Pond, B. Yanny, S. Allam, R. Armstrong, W. Barkhouse, K. Bechtol, A. Benoit-Lévy, G. M. Bernstein, E. Bertin, E. Buckley-Geer, R. Covarrubias, S. Desai, H. T. Diehl, D. A. Goldstein, D. Gruen, T. S. Li, H. Lin, J. Marriner, J. J. Mohr, E. Neilsen, C.-C. Ngeow, K. Paech, E. S. Rykoff, M. Sako, I. Sevilla-Noarbe, E. Sheldon, F. Sobreira, D. L. Tucker, and W. Wester and. The dark energy survey image processing pipeline. *Publications of the Astronomical Society of the Pacific*, 130(989):074501, may 2018.

[91] Kevin J. Napier and The DEEP Collaboration. The DECam Ecliptic Exploration Project (DEEP): V: The Size Distribution of TNOs from Single-night Detections, 2022. Manuscript in Preparation.

[92] David Nesvorný. EVIDENCE FOR SLOW MIGRATION OF NEPTUNE FROM THE INCLINATION DISTRIBUTION OF KUIPER BELT OBJECTS. *The Astronomical Journal*, 150(3):73, aug 2015.

[93] David Nesvorný. Jumping neptune can explain the kuiper belt kernel. *The Astronomical journal*, 150(3):68, 2015.

[94] David Nesvorný, Rixin Li, Andrew N. Youdin, Jacob B. Simon, and William M. Grundy. Trans-neptunian binaries as evidence for planetesimal formation by the streaming instability. *Nature Astronomy*, 3(9):808–812, Sep 2019.

[95] David Nesvorný and Alessandro Morbidelli. STATISTICAL STUDY OF THE EARLY SOLAR SYSTEM's INSTABILITY WITH FOUR, FIVE, AND SIX GIANT PLANETS. *The Astronomical Journal*, 144(4):117, sep 2012.

[96] David Nesvorný, David Vokrouhlický, William F. Bottke, and Harold F. Levison. Evidence for very early migration of the solar system planets from the patroclus–menoetius binary jupiter trojan. *Nature Astronomy*, 2(11):878–882, Nov 2018.

[97] David Nesvorný, David Vokrouhlický, and Alessandro Morbidelli. CAPTURE OF TROJANS BY JUMPING JUPITER. *The Astrophysical Journal*, 768(1):45, apr 2013.

[98] D. Nesvorný and L. Dones. How long-lived are the hypothetical trojan populations of saturn, uranus, and neptune? *Icarus*, 160(2):271–288, 2002.

[99] David Nesvorný. Dynamical evolution of the early solar system. *Annual Review of Astronomy and Astrophysics*, 56(1):137–174, 2018.

[100] John Nickolls, Ian Buck, Michael Garland, and Kevin Skadron. Scalable parallel programming with cuda. *Queue*, 6(2):40–53, March 2008.

[101] NOAO. *NOAO Data Handbook*. National Optical Astronomy Observatory, May 2015.

[102] Keith S. Noll, William M. Grundy, David Nesvorný, and Audrey Thirouin. Chapter 9 - trans-neptunian binaries (2018). In Dina Prialnik, M. Antonietta Barucci, and Leslie A. Young, editors, *The Trans-Neptunian Solar System*, pages 205–224. Elsevier, 2020.

[103] Eran O. Ofek and Barak Zackay. Optimal Matched Filter in the Low-number Count Poisson Noise Regime and Implications for X-Ray Source Detection. *Astronomical Journal*, 155(4):169, April 2018.

[104] Travis E. Oliphant. *A guide to NumPy*. Trelgol Publishing, USA, 2006.

[105] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Müller, J. Nothman, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay. Scikit-learn: Machine Learning in Python. *ArXiv e-prints*, January 2012.

[106] Nuno Peixinho, Pedro Lacerda, and David Jewitt. COLOR-INCLINATION RELATION OF THE CLASSICAL KUIPER BELT OBJECTS. *The Astronomical Journal*, 136(5):1837–1845, sep 2008.

[107] J.-M. Petit, J. J. Kavelaars, B. J. Gladman, R. L. Jones, J. Wm. Parker, C. Van Laerhoven, P. Nicholson, G. Mars, P. Rousselot, O. Mousis, B. Marsden, A. Bieryla, M. Taylor, M. L. N. Ashby, P. Benavidez, A. Campo Bagatin, and G. Bernabeu. THE CANADA-FRANCE ECLIPTIC PLANE SURVEY—FULL DATA RELEASE: THE ORBITAL STRUCTURE OF THE KUIPER BELT. *The Astronomical Journal*, 142(4):131, sep 2011.

[108] Norman Pogson. Magnitudes of Thirty-six of the Minor Planets for the First Day of each Month of the Year 1857. *Monthly Notices of the Royal Astronomical Society*, 17(1):12–15, 11 1856.

[109] L. Y. Pratt. Discriminability-based transfer between neural networks. In S. Hanson, J. Cowan, and C. Giles, editors, *Advances in Neural Information Processing Systems*, volume 5. Morgan-Kaufmann, 1992.

[110] Ning Qian and Terrence J. Sejnowski. Predicting the secondary structure of globular proteins using neural network models. *Journal of Molecular Biology*, 202(4):865–884, 1988.

[111] Hanno Rein and Daniel Tamayo. whfast: a fast and unbiased implementation of a symplectic Wisdom–Holman integrator for long-term gravitational simulations. *Monthly Notices of the Royal Astronomical Society*, 452(1):376–388, 07 2015.

[112] Rein, H. and Liu, S.-F. Rebound: an open-source multi-purpose n-body code for collisional dynamics. *A&A*, 537:A128, 2012.

[113] Gene Roddenberry, Sandy Fries, and Hans Beimler. Star Trek: The Next Generation: Coming of Age, 1998.

[114] T. Santana-Ros, M. Micheli, L. Faggioli, R. Cennamo, M. Devogèle, A. Alvarez-Candal, D. Oszkiewicz, O. Ramírez, P.-Y. Liu, P. G. Benavidez, A. Campo Bagatin, E. J. Christensen, R. J. Wainscoat, R. Weryk, L. Fraga, C. Briceño, and L. Conversi. Orbital stability analysis and photometric characterization of the second earth trojan asteroid 2020 xl5. *Nature Communications*, 13(1):447, Feb 2022.

[115] Megan E. Schwamb, Wesley C. Fraser, Michele T. Bannister, Michaël Marsset, Rosemary E. Pike, J. J. Kavelaars, Susan D. Benecchi, Matthew J. Lehner, Shiang-Yu Wang, Audrey Thirouin, Audrey Delsanti, Nuno Peixinho, Kathryn Volk, Mike Alexandersen, Ying-Tung Chen, Brett Gladman, Stephen D. J. Gwyn, and Jean-Marc Petit. Col-OSSOS: The colors of the outer solar system origins survey. *The Astrophysical Journal Supplement Series*, 243(1):12, jul 2019.

[116] Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[117] P.K. Seidelmann, R.S. Harrington, D. Pascu, W.A. Baum, D.G. Currie, J.A. Westphal, and G.E. Danielson. Saturn satellite observations and orbits from the 1980 ring plane crossing. *Icarus*, 47(2):282–287, 1981.

[118] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, Jan 2016.

[119] Hayden Smotherman, Andrew J. Connolly, J. Bryce Kalmbach, Stephen K. N. Portillo, Dino Bektesevic, Siegfried Eggl, Mario Juric, Joachim Moeyens, and Peter J. Whidden. Sifting through the static: Moving object detection in difference images. *The Astronomical Journal*, 162(6):245, nov 2021.

[120] Hayden Smotherman and The DEEP Collaboration. The DECam Ecliptic Exploration Project (DEEP): VI. Multi-night Detection and Characterization, 2022. Manuscript in Preparation.

[121] Ian Sullivan and Eric Bellm. DMTN-171: Fall 2020 status of crowded field processing with the LSST Alert Production Pipelines. LSST Data Management Technical Note DMTN-171, LSST Data Management, February 2021.

[122] S. C Tegler and W Romanishin. Two distinct populations of kuiper-belt objects. *Nature (London)*, 392(6671):49–51, 1998.

[123] D. E. Trilling, F. Valdes, L. Allen, D. James, C. Fuentes, D. Herrera, T. Axelrod,

and J. Rajagopal. The size distribution of near-earth objects larger than 10 m. *The Astronomical Journal*, 154(4):170, 2017.

[124] David E. Trilling and The DEEP Collaboration. The DECam Ecliptic Exploration Project (DEEP): I. Survey description, science questions, and technical demonstration, 2022. Manuscript in Preparation.

[125] Chad Trujillo and The DEEP Collaboration. The DECam Ecliptic Exploration Project (DEEP) II: Survey strategy, 2022. Manuscript in Preparation.

[126] Chadwick A Trujillo and Michael E Brown. A correlation between inclination and color in the classical kuiper belt. *The Astrophysical journal*, 566(2):L125–L128, 2002.

[127] K. Tsiganis, R. Gomes, A. Morbidelli, and H. F. Levison. Origin of the orbital architecture of the giant planets of the solar system. *Nature*, 435(7041):459–61, May 26 2005. Copyright - Copyright Macmillan Journals Ltd. May 26, 2005; Document feature - references; graphs; Last updated - 2017-10-31; CODEN - NATUAS.

[128] Kathryn Tunyasuvunakool, Jonas Adler, Zachary Wu, Tim Green, Michal Zielinski, Augustin Žídek, Alex Bridgland, Andrew Cowie, Clemens Meyer, Agata Laydon, Sameer Velankar, Gerard J. Kleywegt, Alex Bateman, Richard Evans, Alexander Pritzel, Michael Figurnov, Olaf Ronneberger, Russ Bates, Simon A. A. Kohl, Anna Potapenko, Andrew J. Ballard, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Ellen Clancy, David Reiman, Stig Petersen, Andrew W. Senior, Koray Kavukcuoglu, Ewan Birney, Pushmeet Kohli, John Jumper, and Demis Hassabis. Highly accurate protein structure prediction for the human proteome. *Nature*, 596(7873):590–596, Aug 2021.

[129] F. Valdes, R. Gruendl, and DES Project. The DECam Community Pipeline. In N. Manset and P. Forshay, editors, *Astronomical Data Analysis Software and Systems*

XXIII, volume 485 of *Astronomical Society of the Pacific Conference Series*, page 379, May 2014.

[130] Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu, and the scikit-image contributors. scikit-image: image processing in Python. *PeerJ*, 2:e453, 6 2014.

[131] Julia Venturini, Maria Paula Ronco, and Octavio Miguel Guilera. Setting the stage: Planet formation and volatile delivery. *Space Science Reviews*, 216(5):86, Jul 2020.

[132] Ofer Vitells and Eilam Gross. Estimating the significance of a signal in a multi-dimensional search. *Astroparticle Physics*, 35(5):230–234, December 2011.

[133] Huapei Wang, Benjamin P. Weiss, Xue-Ning Bai, Brynna G. Downey, Jun Wang, Jiajun Wang, Clément Suavet, Roger R. Fu, and Maria E. Zucolotto. Lifetime of the solar nebula constrained by meteorite paleomagnetism. *Science*, 355(6325):623–627, 2017.

[134] George W. Wetherill. Formation of the earth. *Annual Review of Earth and Planetary Sciences*, 18(1):205–256, 1990.

[135] Lorien F. Wheeler and Donovan L. Mathias. Probabilistic assessment of tunguska-scale asteroid impacts. *Icarus*, 327:83–96, 2019. Tunguska.

[136] Peter J. Whidden, J. Bryce Kalmbach, Andrew J. Connolly, R. Lynne Jones, Hayden Smotherman, Dino Bektesevic, Colin Slater, Andrew C. Becker, Željko Ivezić, Mario Jurić, Bryce Bolin, Joachim Moeyens, Francisco Förster, and V. Zach Golkhou. Fast algorithms for slow moving asteroids: Constraints on the distribution of kuiper belt objects. *The Astronomical Journal*, 157(3):119, feb 2019.

[137] Ian Wong and Michael E. Brown. The bimodal color distribution of small kuiper belt objects. *The Astronomical Journal*, 153(4):145, mar 2017.

[138] Andrew N. Youdin and Jeremy Goodman. Streaming instabilities in protoplanetary disks. *The Astrophysical Journal*, 620(1):459–469, feb 2005.

[139] Barak Zackay and Eran O. Ofek. How to COAAD Images. I. Optimal Source Detection and Photometry of Point Sources Using Ensembles of Images. *Astrophysical Journal*, 836(2):187, February 2017.

[140] Barak Zackay and Eran O. Ofek. How to COAAD Images. II. A Coaddition Image that is Optimal for Any Purpose in the Background-dominated Noise Limit. *Astrophysical Journal*, 836(2):188, February 2017.

[141] Barak Zackay, Eran O. Ofek, and Avishay Gal-Yam. Proper Image Subtraction—Optimal Transient Detection, Photometry, and Hypothesis Testing. *Astrophysical Journal*, 830(1):27, October 2016.

[142] Ciyou Zhu, Richard H. Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.*, 23(4):550–560, dec 1997.

Appendix A

APPENDIX

A.1 Table of Detected Object Parameters from Chapter 3

Identifier (pg, ccd)	*VR* Mag	Barycentric Distance (au)	i (degrees)	Ω (degrees)	Linked to Known Object
(190,20)	24.46±0.22	38.38±2.42	8.58±2.99	48.73±4.99	True
(190,23)	24.13±0.82	42.02±2.43	3.10±0.50	84.35±10.33	True
(191,27)	24.27±0.23	39.76±2.41	3.60±1.33	36.21±0.22	True
(191,47)	24.18±0.40	31.87±2.31	4.34±1.69	31.89±1.58	True
(192,06)	24.51±0.27	54.00±2.60	2.52±0.92	206.20±4.28	True
(192,36)	24.48±0.40	41.97±2.46	5.16±1.89	35.82±0.84	True
(192,42)	24.37±0.29	43.62±2.47	2.14±0.78	24.31±5.10	True
(193,05)	24.61±0.29	39.75±2.41	2.75±0.35	90.86±9.91	True
(193,18)	24.44±0.37	42.30±2.51	11.27±4.05	206.48±4.19	True
(193,21)	24.48±0.36	46.19±2.49	2.90±0.67	73.62±9.82	True
(193,23)	23.72±0.19	43.08±2.46	2.82±0.59	173.42±11.60	True
(193,50)	24.31±0.29	32.36±2.51	17.95±7.29	42.20±1.84	True
(301,46)	23.76±0.57	44.95±2.61	2.02±0.74	8.43±5.74	True
(302,06)	22.83±0.29	38.44±3.69	30.55±17.70	199.08±2.98	True
(195,47)	22.97±0.15	32.48±2.45	9.90±3.56	14.35±8.85	True
(202,48)	23.26±0.17	40.88±2.58	16.22±5.92	47.86±3.56	True
(203,09)	24.18±0.68	44.53±2.48	2.08±0.36	166.51±12.63	True

Continued on next page

Identifier (pg, ccd)	*VR* Mag	Barycentric Distance (au)	i (degrees)	Ω (degrees)	Linked to Known Object
(203,11)	23.92±0.20	48.25±2.90	24.39±9.49	215.16±1.73	True
(203,43)	24.52±0.47	44.81±2.53	8.68±3.05	45.78±2.30	True
(205,18)	24.04±0.59	41.57±2.47	3.22±0.04	122.65±4.57	True
(284,29)	22.06±0.46	34.99±2.54	173.36±0.54	136.65±10.42	True
(285,22)	20.75±0.07	41.60±3.94	14.00±6.38	349.32±17.52	True
(296,28)	22.69±0.16	41.22±2.71	6.75±1.43	252.08±13.39	True
(300,30)	23.70±0.98	46.90±2.94	17.79±7.56	24.73±0.84	True
(017,46)	23.99±0.33	35.59±2.45	5.31±0.24	285.01±7.11	False
(018,20)	23.93±1.26	38.64±2.45	4.92±0.10	316.13±5.57	False
(018,52)	24.78±0.80	46.17±2.57	12.08±3.39	243.09±8.42	False
(191,19)	24.99±0.58	42.91±2.44	1.58±0.62	39.44±1.65	False
(191,50)	24.14±0.36	30.67±1.17	156.06±4.63	216.93±0.01	False
(192,05)	24.41±0.24	41.54±2.43	1.07±0.49	58.69±10.24	False
(192,08)	24.20±0.23	30.66±1.17	155.99±4.65	216.93±0.01	False
(192,08)	24.74±0.24	40.49±2.42	0.80±0.37	42.48±2.53	False
(192,54)	22.24±0.04	32.95±1.16	166.54±2.55	41.03±0.58	False
(193,07)	24.42±0.33	41.65±2.44	3.00±0.38	91.13±9.69	False
(193,10)	24.65±0.52	42.19±2.44	2.70±0.37	89.64±10.07	False
(193,14)	23.96±0.41	42.47±2.45	4.01±1.17	62.14±7.95	False
(193,14)	25.02±0.93	41.86±2.43	1.83±0.10	107.81±9.23	False
(193,26)	23.97±0.14	42.55±2.44	3.06±0.93	63.12±8.68	False
(193,32)	24.48±0.64	36.70±2.37	4.86±1.69	50.56±5.10	False
(193,40)	24.62±0.31	42.11±2.44	2.22±0.62	66.85±9.23	False
(301,30)	22.65±1.01	41.40±2.73	16.18±6.70	203.61±0.50	False

Continued on next page

Identifier (pg, ccd)	*VR* Mag	Barycentric Distance (au)	i (degrees)	Ω (degrees)	Linked to Known Object
(301,40)	23.69±0.40	41.28±2.65	0.63±0.24	238.25±15.08	False
(305,14)	23.27±0.34	39.36±2.42	7.04±2.54	36.55±3.74	False
(305,28)	23.40±0.43	41.71±2.44	1.24±0.04	129.48±8.93	False
(305,60)	23.23±0.46	33.71±2.69	24.33±10.40	206.56±0.24	False
(306,47)	23.42±0.28	42.58±2.51	4.02±0.99	63.85±10.83	False
(306,48)	22.90±0.54	38.66±2.59	13.04±5.12	37.81±4.62	False
(306,49)	22.93±0.69	38.26±2.46	2.75±0.07	130.35±5.44	False
(307,05)	23.36±0.69	32.65±2.60	5.20±0.24	135.87±7.47	False
(310,29)	22.95±0.18	46.00±2.53	4.40±0.11	105.05±5.51	False
(310,29)	23.01±0.55	47.06±2.54	4.18±0.03	125.85±3.95	False
(310,36)	23.93±0.68	39.59±3.05	28.70±12.84	202.30±3.92	False
(311,46)	22.38±0.19	41.20±2.90	25.12±10.68	23.80±0.88	False
(313,60)	22.85±0.24	34.45±2.43	5.22±1.24	248.71±11.50	False
(314,13)	23.21±0.26	34.06±2.38	4.65±1.81	213.63±2.65	False
(316,45)	23.17±0.51	45.23±2.49	6.53±2.19	14.69±4.92	False
(316,55)	23.06±0.22	42.89±2.46	6.75±2.36	19.17±3.99	False
(194,18)	24.61±0.68	43.06±2.45	1.51±0.34	351.73±13.15	False
(194,21)	24.42±0.26	43.55±2.46	2.94±0.83	5.44±9.74	False
(194,27)	24.52±...	34.99±2.46	14.26±5.49	222.73±2.60	False
(195,20)	24.14±0.21	48.32±2.78	21.31±7.74	225.49±3.55	False
(195,60)	24.84±0.70	34.81±2.41	4.34±0.03	311.05±4.15	False
(196,30)	24.61±0.57	35.51±2.58	14.99±6.01	25.81±5.82	False
(197,19)	24.46±0.49	43.06±2.46	2.46±0.34	341.70±11.82	False
(197,34)	22.84±0.23	38.20±2.45	10.77±3.82	229.64±4.24	False

Continued on next page

Identifier (pg, ccd)	*VR* Mag	Barycentric Distance (au)	i (degrees)	Ω (degrees)	Linked to Known Object
(197,36)	24.75±0.73	41.91±2.45	3.26±0.80	359.82±11.28	False
(197,58)	24.13±0.60	35.47±2.71	23.38±9.82	32.45±2.94	False
(202,05)	23.88±0.13	41.77±2.45	3.54±0.03	121.93±4.22	False
(202,20)	23.60±0.40	46.02±2.53	8.61±2.62	58.42±6.64	False
(202,27)	23.44±0.28	39.07±2.44	8.62±2.75	57.97±6.73	False
(202,36)	23.71±0.29	38.38±3.22	34.18±15.74	42.95±2.59	False
(202,40)	23.43±0.24	34.43±3.42	35.99±18.64	41.40±2.35	False
(202,42)	24.27±0.60	42.86±4.42	46.16±26.51	41.31±2.46	False
(203,12)	24.18±0.56	45.41±2.56	10.71±3.84	209.74±3.52	False
(203,43)	24.57±1.29	41.76±2.45	1.04±0.01	130.90±14.38	False
(204,21)	23.95±1.04	46.44±2.54	5.71±2.04	211.76±3.18	False
(204,21)	24.04±0.45	42.46±2.48	2.51±0.78	62.47±7.19	False
(204,27)	24.00±0.42	47.18±2.59	10.79±3.91	216.79±1.34	False
(204,41)	24.07±0.48	44.49±2.50	0.61±0.24	177.43±21.56	False
(205,22)	24.37±0.45	38.39±2.49	7.67±2.68	198.45±7.99	False
(205,23)	24.14±0.33	41.25±2.53	11.44±4.13	205.70±5.58	False
(205,49)	24.26±0.40	43.63±2.49	2.33±0.06	139.21±8.76	False
(284,13)	23.45±0.18	36.40±...	31.17±...	10.41±...	False
(284,29)	23.59±0.68	37.56±4.53	9.23±4.19	340.77±22.78	False
(284,42)	22.60±0.18	41.54±3.52	10.56±4.43	238.99±18.72	False
(284,52)	23.68±0.27	41.11±3.67	10.45±3.41	340.23±16.58	False
(284,59)	23.25±0.62	36.70±4.97	8.47±3.74	254.19±32.60	False
(288,15)	23.89±0.48	45.70±3.47	17.81±9.31	12.57±4.05	False
(288,29)	23.41±0.28	44.98±4.41	30.20±20.59	204.55±3.45	False

Continued on next page

Identifier (pg, ccd)	*VR* Mag	Barycentric Distance (au)	i (degrees)	Ω (degrees)	Linked to Known Object
(288,48)	23.54±0.27	40.10±3.98	19.55±13.47	208.92±6.35	False
(289,48)	23.51±0.41	38.02±6.91	21.02±27.89	213.72±18.67	False
(289,48)	23.82±0.51	42.45±3.30	8.96±3.85	233.24±15.66	False
(290,23)	23.84±0.37	47.45±3.38	26.35±12.60	211.83±4.95	False
(291,08)	23.25±0.26	34.80±...	40.90±...	18.00±...	False
(291,27)	23.85±0.61	33.10±10.91	29.45±64.35	206.97±13.80	False
(296,43)	23.60±0.31	38.91±2.78	6.00±0.71	327.10±10.71	False
(297,46)	23.33±0.27	33.72±4.03	7.33±0.07	291.61±6.21	False
(298,22)	23.95±0.41	45.59±4.09	36.73±20.72	216.21±7.06	False
(298,26)	22.38±0.59	33.54±...	39.53±...	17.32±...	False

Table A.1: Best-fit parameters estimated from the DECam NEO Survey data for the objects detected with `KBMOD`. Objects are identified based on their detected pointing group (pg) and CCD. Orbital values and uncertainties are found using the method of [16]. *VR* magnitudes are found as described in 3.3.3. *VR* magnitude uncertainties are σ_G uncertainties, estimated based on the individual stamp *VR* magnitude estimates. Parameters for which no uncertainty was returned are indicated with an ellipsis in the uncertainty value.

A.2 Table of Detected Objects from Chapter 4

Id	Class	χ^2/ν	a (au)	e	i (deg)	Ω (deg)	ω (deg)	$t_{\rm peri}$ (yr J200)	H
0	Classical	0.25	43.00±0.01	0.09	21.71	353.16	359.06	19.62	7.95
1	Resonant	0.26	43.66±0.01	0.06	1.03	329.57	235.42	-94.66	8.74
2	Classical	0.30	44.25±0.07	0.06	0.56	159.63	9.49	163.66	7.82
3	Classical	0.22	46.03±0.02	0.11	1.94	353.90	17.82	33.11	8.84
4	Detached	0.40	51.32±0.03	0.25	0.38	309.65	20.59	7.14	9.09
5	Classical	0.46	43.89±0.02	0.08	0.34	82.81	52.67	130.70	8.33
6	Classical	0.27	45.86±0.10	0.12	4.35	172.75	286.29	100.55	7.29
7	Classical	0.32	44.18±0.04	0.07	0.55	9.34	253.45	-46.29	8.46
8	Classical	0.23	43.02±0.01	0.01	1.54	164.59	201.57	30.51	6.76
9	Classical	0.22	45.84±0.03	0.09	11.19	160.22	224.06	44.46	8.45
10	Scattering	0.24	69.01±0.23	0.52	7.41	7.69	50.57	54.68	7.93
11	Classical	0.17	42.66±0.01	0.01	1.27	88.46	279.53	33.68	6.08
12	Classical	0.22	42.65±0.02	0.04	1.57	119.10	112.16	-68.53	7.16
13	Classical	0.22	42.56±0.02	0.04	1.27	68.70	163.24	-68.24	7.61
14	Classical	0.32	44.33±0.02	0.07	5.55	2.63	248.80	-54.68	8.95
15	Classical	0.17	46.19±0.02	0.13	26.99	352.53	22.29	35.61	8.72
16	Classical	0.19	45.22±0.02	0.10	2.07	146.68	121.58	-40.92	7.85
17	Classical	0.18	42.49±0.01	0.02	3.57	3.55	334.50	9.93	8.53
18	Classical	0.23	45.84±0.02	0.01	0.57	89.95	279.82	35.02	7.74
19	Classical	0.29	43.50±0.02	0.01	1.96	3.11	349.81	20.96	8.06
20	Classical	0.18	37.74±0.02	0.08	26.32	170.52	223.18	43.36	9.46
21	Classical	0.23	42.98±0.04	0.02	1.48	147.01	236.34	44.15	8.60
22	Classical	0.20	46.47±0.01	0.08	0.62	123.95	225.53	17.84	8.11

Continued on next page

Id	Class	χ^2/ν	a (au)	e	i (deg)	Ω (deg)	ω (deg)	$t_{\rm peri}$ (yr J200)	H
23	Classical	0.19	44.07±0.01	0.10	0.89	56.57	239.15	-18.89	8.26
24	Classical	0.24	42.98±0.01	0.02	0.53	87.25	262.91	18.45	7.11
25	Classical	0.36	43.99±0.02	0.07	1.50	138.62	238.80	38.01	9.29
26	Classical	0.38	43.53±0.06	0.08	14.46	353.56	154.01	141.86	9.12
27	Classical	0.29	43.87±0.99	0.06	0.95	69.81	28.64	100.59	8.37
28	Classical	0.46	43.02±0.14	0.05	1.42	38.71	265.27	-14.25	9.60
29	Resonant	0.27	47.69±0.05	0.19	12.24	356.46	226.33	-81.04	7.46
30	Classical	0.26	44.36±0.03	0.10	4.07	156.16	97.36	-51.75	8.79
31	Classical	0.36	42.64±0.02	0.09	13.72	356.41	61.54	63.46	7.27
32	Classical	0.24	43.03±0.02	0.11	9.13	359.12	21.51	37.80	9.45
33	Classical	0.30	43.15±0.09	0.06	1.95	30.28	39.17	75.89	8.98
34	Classical	0.62	45.53±0.02	0.07	3.08	150.48	24.53	-130.98	9.20
35	Classical	0.52	45.73±0.53	0.08	2.72	145.55	254.81	55.44	9.25
36	Classical	0.18	43.04±0.01	0.05	2.39	25.06	125.89	142.69	7.38
37	Classical	0.29	40.44±0.02	0.09	21.03	167.76	262.89	69.37	9.43
38	(Scattering)	0.42	53.73±0.04	0.36	1.84	149.08	160.29	-0.47	9.17
39	Classical	0.25	46.92±0.03	0.14	0.89	99.29	186.15	-25.64	7.22
40	Classical	0.19	44.68±0.01	0.07	6.08	160.52	164.53	1.25	8.74
41	Classical	0.61	42.86±0.07	0.03	2.00	136.03	309.49	90.69	9.12
42	Classical	0.19	43.94±0.01	0.05	1.70	104.47	232.05	8.71	7.10
43	Classical	0.51	45.36±0.11	0.06	2.29	134.82	290.85	77.67	8.60
44	Classical	0.19	45.73±0.01	0.09	1.35	83.34	225.22	-11.07	8.12
45	Classical	0.37	43.21±0.01	0.02	1.45	81.85	133.12	-86.30	8.71
46	Classical	0.23	41.69±0.01	0.13	33.00	353.20	171.22	148.34	8.15
47	Classical	0.17	44.43±0.02	0.04	1.74	103.80	292.19	53.96	8.00

Continued on next page

Id	Class	χ^2/ν	a (au)	e	i (deg)	Ω (deg)	ω (deg)	$t_{\rm peri}$ (yr J200)	H
48	Classical	0.23	45.48±0.03	0.12	7.00	4.85	24.11	44.64	8.80
49	Classical	0.23	45.21±0.06	0.16	12.13	164.57	305.98	107.09	7.60
50	Classical	0.36	45.48±0.04	0.12	2.22	133.57	222.82	23.77	8.72
51	Classical	0.30	45.42±0.04	0.08	1.69	101.66	0.84	107.27	8.36
52	Classical	0.17	44.58±0.01	0.06	3.58	19.76	239.45	-51.47	6.77
53	Detached	0.22	52.31±0.02	0.28	9.71	1.32	320.85	3.03	7.19
54	Classical	0.42	44.52±0.02	0.07	4.59	153.02	211.01	29.75	9.11
55	Resonant	0.14	39.29±0.02	0.21	11.87	356.29	161.41	130.65	6.90
56	Classical	0.45	44.04±0.03	0.09	3.97	14.82	262.18	-31.81	9.09
57	Classical	0.31	38.44±0.02	0.05	2.16	119.97	281.34	50.04	9.63
58	Classical	0.24	43.08±0.01	0.07	2.26	118.48	233.25	20.27	8.82
59	Classical	0.38	42.73±0.02	0.02	2.31	119.54	19.65	133.38	8.74
60	Classical	0.52	44.46±0.01	0.05	2.34	120.68	98.89	-85.17	8.54
61	(Resonant)	0.18	43.67±0.03	0.07	3.42	23.94	86.66	108.79	7.48
62	Classical	0.26	42.64±0.02	0.08	1.99	111.18	152.44	-41.55	8.48
63	Classical	0.44	43.84±0.03	0.03	1.68	80.22	179.69	-51.93	9.10
64	Classical	0.42	45.70±0.03	0.06	1.93	91.10	19.12	117.12	8.53
65	Classical	0.33	43.93±0.02	0.05	2.62	43.06	322.31	29.76	9.08
66	Classical	0.37	43.90±0.03	0.07	3.78	141.44	308.40	91.86	9.44
67	Classical	0.23	44.44±0.01	0.07	3.69	17.30	273.67	-23.94	8.48
68	Classical	0.21	42.50±0.01	0.16	24.16	355.02	344.58	13.81	9.09
69	Classical	0.20	43.20±0.02	0.04	1.79	50.11	131.44	-112.03	7.27
70	Classical	0.32	43.09±0.02	0.06	2.74	43.92	105.26	140.88	8.94
71	Classical	0.19	43.91±0.01	0.04	1.89	78.76	85.19	159.72	8.41
72	Classical	0.22	44.31±0.02	0.08	3.38	33.84	185.97	-82.28	7.10

Continued on next page

Id	Class	χ^2/ν	a (au)	e	i (deg)	Ω (deg)	ω (deg)	$t_{\rm peri}$ (yr J200)	H
73	Classical	0.20	43.89±0.02	0.08	2.31	116.55	164.86	-28.37	8.11
74	Classical	0.50	43.41±0.01	0.06	2.93	41.34	205.47	-58.48	8.86
75	Classical	0.23	42.90±0.06	0.04	1.89	84.11	286.86	35.25	7.49
76	Classical	0.55	43.17±0.02	0.05	2.32	104.38	164.07	-41.84	9.56
77	Resonant	0.26	39.35±0.03	0.19	12.03	162.87	10.86	-102.40	8.13
78	Classical	0.24	42.84±0.01	0.02	2.54	60.60	142.42	-95.81	8.32
79	Classical	0.53	45.49±0.01	0.07	2.59	56.50	268.66	-0.24	9.01
80	Classical	0.33	43.20±0.01	0.07	2.36	73.84	166.35	-62.35	8.53
81	Classical	0.30	44.10±0.03	0.12	22.00	165.79	184.20	18.34	9.33
82	Classical	0.28	44.88±0.01	0.04	2.64	71.74	147.89	-87.91	8.57
83	Classical	0.36	44.42±0.01	0.06	3.16	51.53	157.94	-94.18	8.32
84	Classical	0.26	55.34±0.03	0.22	18.12	1.18	350.71	19.76	7.25
85	Classical	0.31	39.28±0.07	0.01	11.12	158.94	34.75	-87.81	8.90
86	Classical	0.42	43.76±0.01	0.07	2.83	67.42	205.75	-37.44	9.02
87	Classical	0.44	42.40±0.02	0.01	3.12	102.91	43.36	137.75	8.57
88	Classical	0.33	46.43±0.99	0.21	7.24	146.62	247.13	44.13	8.93
89	Classical	0.26	42.71±0.02	0.06	0.53	35.38	280.50	-5.73	8.32
90	Classical	0.19	43.05±0.05	0.01	1.25	167.63	204.36	35.53	7.89
91	Classical	0.17	43.74±0.01	0.03	1.54	156.22	354.58	145.94	8.11
92	Detached	0.41	90.06±0.62	0.58	24.05	171.36	235.39	51.04	7.71
93	Resonant	0.25	39.45±0.03	0.12	2.78	160.08	15.04	-102.01	8.12
94	Classical	0.32	44.35±0.02	0.07	2.54	6.97	6.15	34.55	8.45
95	Classical	0.26	44.03±0.21	0.07	1.86	354.69	353.21	17.40	9.20
96	Detached	0.38	80.86±0.09	0.51	28.74	170.28	174.01	15.87	8.80
97	Detached	0.34	77.53±2.20	0.52	20.69	169.67	61.39	-92.22	6.21

Continued on next page

Id	Class	χ^2/ν	a	e	i	Ω	ω	$t_{\rm peri}$	H
			(au)		(deg)	(deg)	(deg)	(yr J200)	
98	Classical	0.25	42.59±0.03	0.04	2.81	157.28	200.75	24.48	8.19
99	Classical	0.28	42.87±0.04	0.05	1.19	28.20	2.28	47.39	8.57
100	Classical	0.27	44.27±0.11	0.06	2.85	102.92	236.19	9.85	8.37
101	Classical	0.29	42.57±0.10	0.04	4.25	39.14	184.70	-76.48	8.06
102	(Classical)	0.33	51.78±6.70	0.32	12.04	10.19	53.18	60.89	8.78

Table A.2: A table of the 103 detected real objects from Chapter 4 which passed all cuts. () surrounding the classification name denotes an insecure classification. Uncertainty estimates do not fit in the space allowed for all parameters, and are therefore only included for semi-major axis. Class "Inner" corresponds to "Inner Centaur".