Draft version March 1, 2023
Typeset using LaTeX twocolumn style in AASTeX631

An improved classification scheme for distinguishing NEOs from MBAs
ASTR 597A Final Project

Tom Wagg[1]

[1]Department of Astronomy, University of Washington, Seattle, WA, 98195

(Received March 1, 2023)

1. INTRODUCTION

Near-Earth Objects (NEOs) are asteroids and comets that have a perihelion distance less than 1.3 au. It is estimated that approximately one fifth of this population passes close enough to Earth that small perturbations in their orbit may lead to intersections with the Earth's orbit and potential collisions (e.g. Jones et al. 2018). A subset of these objects are known as Potentially Hazardous Asteroids (PHAs), these objects are defined as being at least 140m in diameter that pass within 0.05au of the Earth[1]. PHAs are large enough to make it through the Earth's atmosphere and still cause continent scale damage through impact. Given the threat posed by these objects, a world-wide effort[2] has been ongoing to catalogue and determine the orbits and sizes of NEOs including identifying any posing a hazard to the Earth.

The Minor Planet Center maintains a catalogue of known NEOs and their orbits[3], as well as the NEO confirmation page (NEOCP[4]). The NEOCP is a continously updated web page listing newly discovered NEO candidates that should be prioritised for additional observations by the NEO follow-up community. These follow-up observations contribute additional astrometric observations necessary to more accurately determine the orbit of the candidate, as well as photometry to constrain its size. An object is only listed on the NEOCP when it has a high probability of being an NEO. This probability is quantified using the `digest2` code (Keys et al. 2019). `digest2` assigns a score between 0 and 100 based on potential orbits that fit the observations and only objects with a score of 65 or more are listed on the page. Currently, on average around two dozen objects are added to the NEOCP on each night.

The Rubin Observatory Legacy Survey of Space and Time (LSST, Ivezić et al. 2019) will rapidly increase the rate at which NEO candidates are identified and reported to the NEOCP. Jones et al. (2018) showed that at the end of the 10-year LSST baseline survey the completeness of NEOs with an absolute magnitude of $H \leq 22$ would be 73%. Most of these objects will be discovered using "tracklet linking": a computational technique where at least three pairs of observations ("tracklets") observed over a 15-night period are identified as belonging to the same object (Jurić et al. 2017; Heinze et al., in prep). The orbits of objects discovered with this technique will typically be reasonably well known, and in need of no immediate follow-up. However, this tracklet linking comes at a cost: the object is not identified as interesting until the third tracklet is imaged – at best, two nights after the first observation or, at worst, nearly two weeks later. This means that potentially interesting (or hazardous) objects may be missed until it's too late to observe (or react to) them.

The number of objects submitted to the NEOCP will increase by several orders of magnitude, and a large fraction of these submissions will be main belt asteroids (MBAs) (Wagg et al. in prep). This would present difficulties for community follow-up, with too many available candidates (and of low purity) passing the current submission criteria.

Due to the vast MBA background, `digest2` performs poorly in identifying NEO candidates, as is demonstrated in Figure 1. We show due to the extreme number of MBA submissions, only 1.5% of submissions meeting the current criteria (a score of least 65) are NEOs, and only 10% of objects with a perfect score of

Corresponding author: Tom Wagg
tomwagg@uw.edu

[1]https://cneos.jpl.nasa.gov/about/neo_groups.html
[2]E.g. https://www.unoosa.org/oosa/en/ourwork/topics/neos/index.html
[3]https://www.minorplanetcenter.net/iau/MPCORB/NEA.txt
[4]https://www.minorplanetcenter.net/iau/NEO/toconfirm_tabular.html

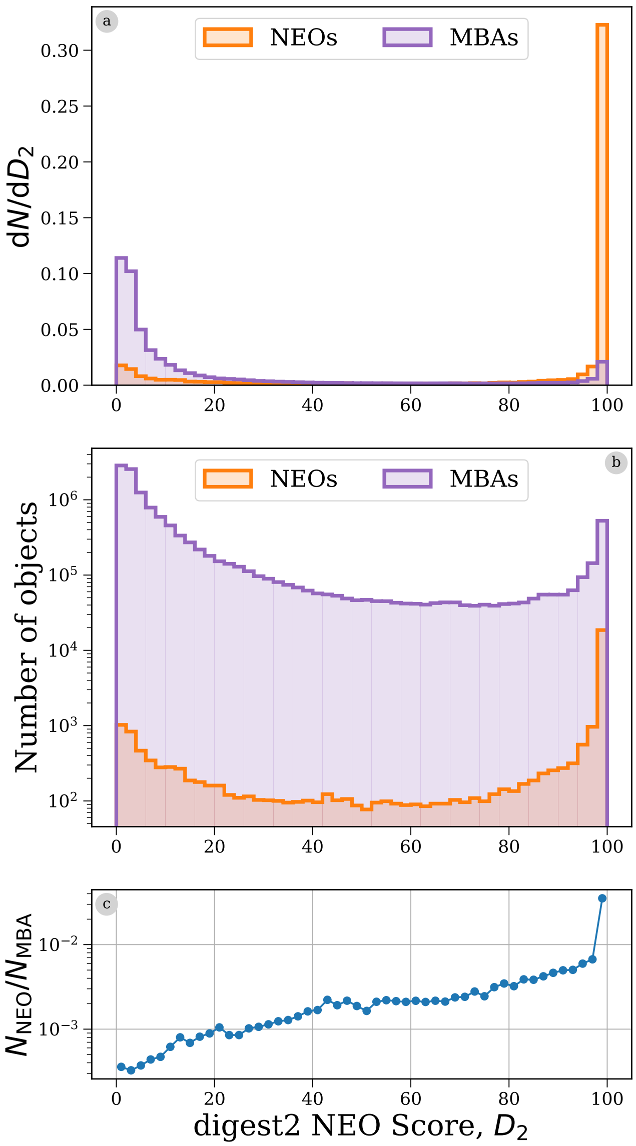

Figure 1. `digest2` scores for all NEOs and MBAs observed in the first year of our simulated LSST observations. **(a)** normalised histograms of `digest2` scores, **(b)** the same histograms un-normalised **(c)** ratio of the histograms in (b). Note that the latter two panels are on a logarithmic scale.

100 are actually NEOs. Additionally, a large fraction of objects obtain a score of exactly 100 and it can be difficult to discriminate the best candidates among this group. It is clear that this classification scheme requires improvement.

The aim of this paper is to augment `digest2` in classifying NEOs in the presence of a significant MBA background. We simulate one year of LSST observations on a synthetic solar system catalogue and calculate `digest2` scores for each tracklet. We investigate three additional parameters - ecliptic latitude, direction of motion relative to the ecliptic and apparent magnitude - as potential further discriminators between NEOs and MBAs in addition to the `digest2` score. We consider various combinations of these parameters to further refine submissions to the NEOCP and highlight how sorting submissions by these parameters can better identify high probability NEO candidates.

2. METHODS

2.1. *Simulated tracklet population*

We create a population of simulated tracklets by performing mock LSST observations on the Synthetic Solar System Model (Grav et al. 2011). We use a modified "Baseline v2.1"[5] 10 year scheduler simulation strategy (Naghib et al. 2019; Cornwall et al. 2020). These observations account for both scheduled and unscheduled downtime and simulate the current baseline observing strategy that will be followed by LSST. The resulting simulations span 3653 days, and have over 50 million observations. The modification consists of shifting the start of LSST to March 2022, around the epoch of the MPCORB catalog used to generate the input hybrid catalogue dataset.

We select observations from the simulated LSST dataset that correspond to tracklets. For a tracklet to be built, we require it to satisfy the following criteria:

1. **Number of observations:** We consider only objects which have at least 3 observations on a given night.

2. **Maximum time separation:** We set the maximum time between observations to 90 minutes.

3. **Minimum arc length:** We ensure that each tracklet is at least 1 arcsecond in length.

The motivation behind these cuts is to ensure that the tracklet constrains the on-sky motion of the object sufficiently well, so that its position at a later time can be easily extrapolated. With fewer observations or shorter tracklets, many different orbits could reproduce the same motion on the sky. Moreover, observations that are separated too significantly in time may be spurious linkages, where observations of multiple objects are incorrectly assumed to be of the same source.

2.2. *Additional NEO scoring parameters*

[5] https://community.lsst.org/t/survey-simulations-v2-1-april-2022/6538

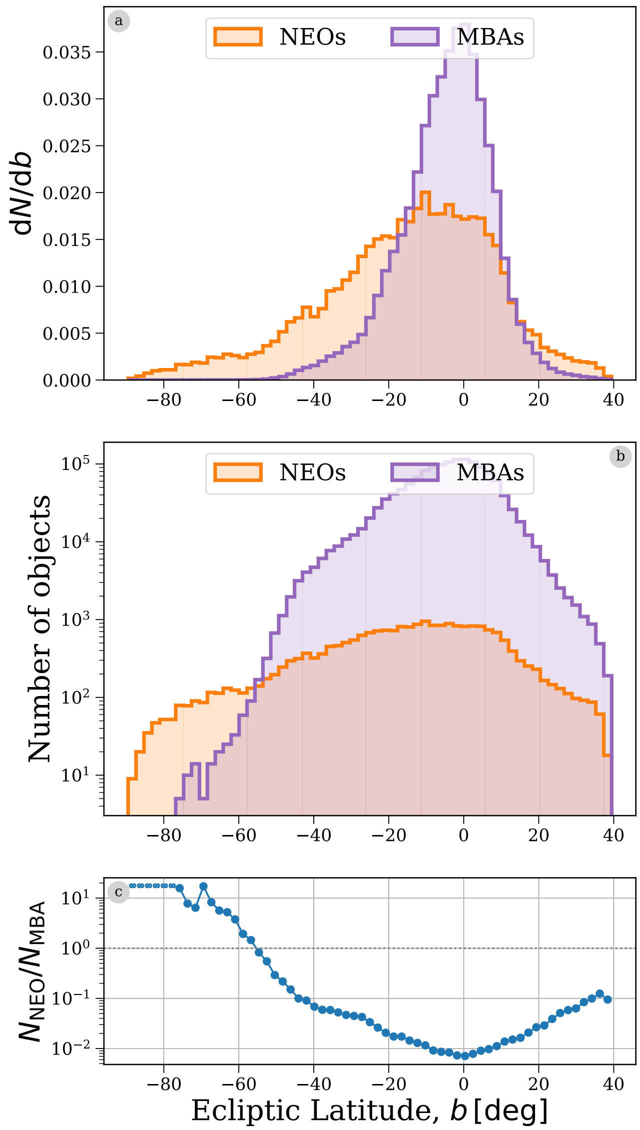

Figure 2. As Figure 1, but for ecliptic latitude and only for observations with a `digest2` score of at least 65.

In addition to the orbital parameters used in computing the `digest2` score, we propose 3 additional parameters to consider in scoring NEOs: ecliptic latitude, direction of motion relative to the ecliptic plane and apparent magnitude. These parameters are motivated by the fact that MBAs are constrained to lie within the ecliptic plane around 3au from the Sun, whereas NEOs are often much closer and are not constrained to any location in the ecliptic.

For these reasons, one would predict that an object with a large ecliptic latitude is less likely to be a MBA. For each tracklet, we convert the mean sky position to an ecliptic latitude, b (and ecliptic longitude, l) using `Astropy` (Astropy Collaboration et al. 2013, 2018, 2022).

In Figure 2, we show the distribution of ecliptic latitudes for objects with a `digest2` score of at least 65. As expected, we find that MBAs are concentrated around the ecliptic plane, whereas NEOs extend to the survey limits in either direction. In particular, we note that even without normalising the histogram, NEOs dominate over MBAs below an ecliptic latitude of $-55°$.

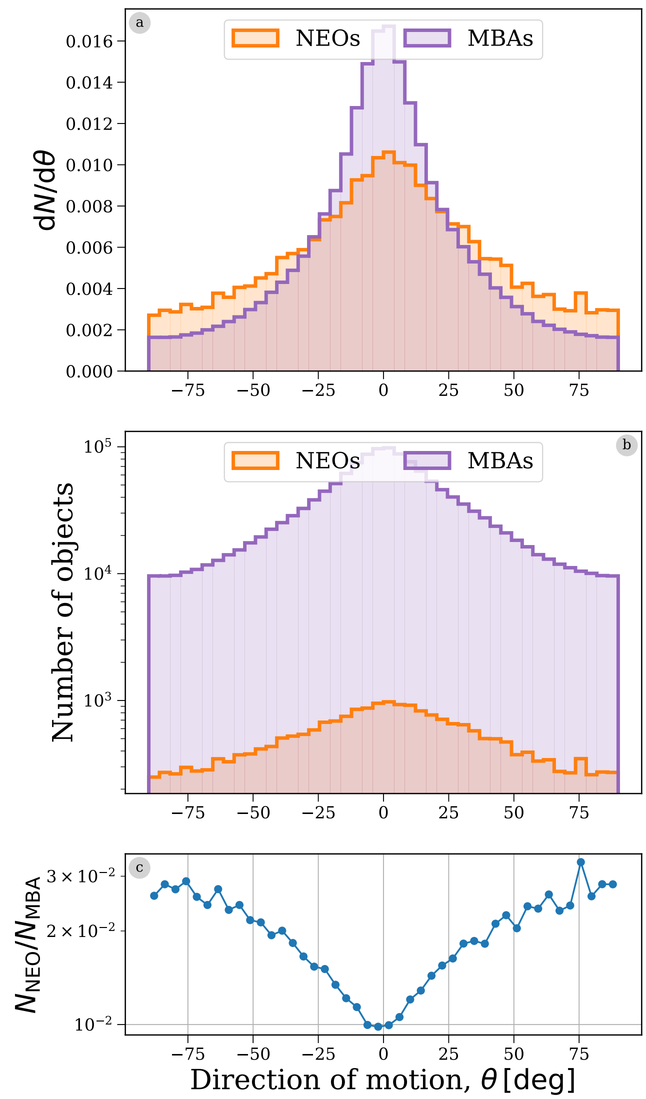

Figure 3. As Figure 1, but for direction of motion relative to the ecliptic plane and only for observations with a `digest2` score of at least 65.

Given that MBAs are constrained to the ecliptic plane and are further away than NEOs, one would predict that their motion is generally along the ecliptic plane, whilst NEOs could move in any direction. To quantify this, we consider

$$\theta = \arctan\left(\frac{l_f - l_i}{b_f - b_i}\right), \tag{1}$$

which gives the angle of the motion relative to the ecliptic plane. We plot the distributions of θ in Figure 3. The separation between NEOs and MBAs in distributions is less clear cut than with the ecliptic latitude but we still find that objects moving away from the ecliptic plane have a larger fraction of NEOs than those moving along the plane.

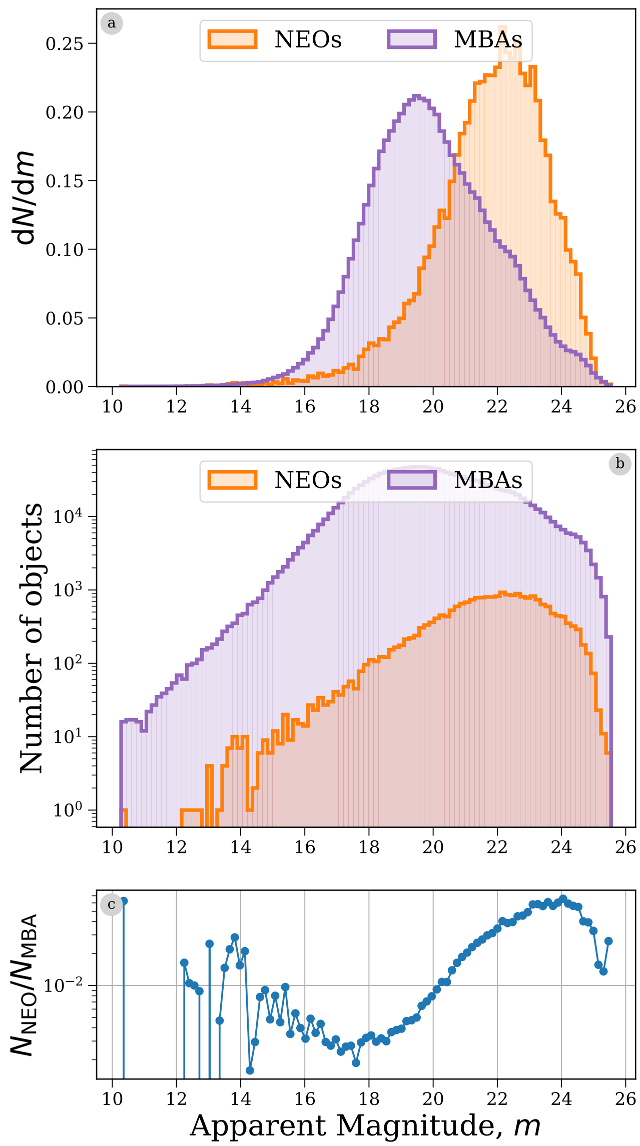

Figure 4. As Figure 1, but for apparent V-band magnitude and only for observations with a `digest2` score of at least 65.

Lastly, given that NEOs are detected closer to the Earth, one may predict that the faintest NEOs can be detected whilst MBAs of equally faint magnitudes would not be found. We therefore convert the magnitude of each observation from its filter to a V-band magnitude using the colours from Jones et al. (2018) and assign each tracklet its mean apparent magnitude. In Figure 4 we explore the distributions of these magnitudes. As expected, NEOs are peaked at fainter magnitudes than MBAs, but when considering the size of the MBA background (in the lower panels) this difference is not as significant.

2.3. *New Score Calculation*

We define a new score as

$$S = \frac{1}{3}\left[\frac{|b|}{0.9} + \frac{|\theta|}{0.9} + \frac{(m - m_{\rm min}) \cdot 100}{m_{\rm max} - m_{\rm min}}\right], \tag{2}$$

which results in a score that equally weights the three parameters and has a range of 0 to 100. In the following sections we explore the use of this score as a threshold for the NEOCP, or potentially more effectively as a way to sort the page.

3. RESULTS

3.1. *New score as a threshold*

In Figure 5, we demonstrate the effect that applying a threshold based on our new score would have on the number of NEOs and MBAs present on the NEOCP and therefore how the purity of the page would change. Applying even a small threshold to objects that pass the initial `digest2` cutoff can drastically decrease the MBA population with a relatively small effect on the NEOs. For example, a threshold of 30 would reduce the number of MBAs submitted by around 550,000 whilst only around 2000 NEOs would be removed.

However, given the relative sizes of the populations this still means that the average purity of the page increases more slowly. In order to increase the purity by a factor of 10 we would decrease the size of the NEO population by around a factor of 7.

For these reasons using this new score as a threshold may not be well-suited to improving NEO follow-up. Nevertheless, objects with the highest scores are preferentially NEOs and so this could be a good way to sort the page and discriminate between objects with the highest `digest2` scores.

3.2. *New score as a sorting metric*

Since the `digest2` score cannot exceed 100, a significant fraction of submitted objects will have a perfect score of 100. We find that on average ∼500 tracklets with at least 3 observations would be submitted to the page and be assigned a score of 100 per night in the first year of LSST. Currently the page is sorted by default by the time at which each object was last updated and so can be fairly random.

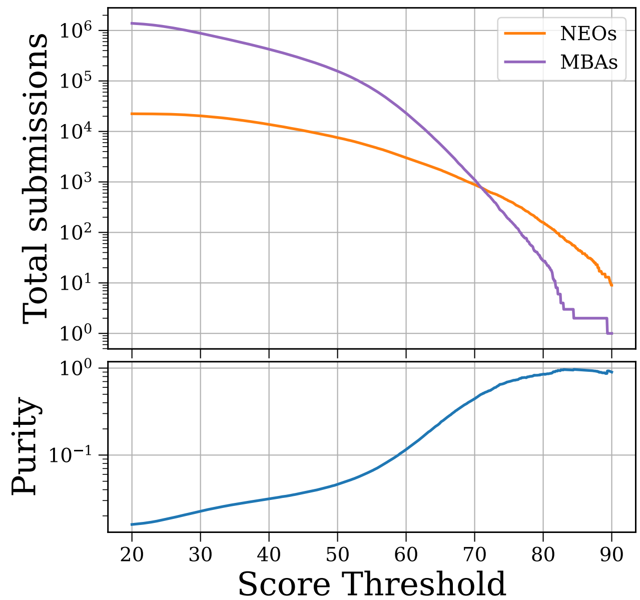

Figure 5. Performance of the new score for different thresholds. **Top:** The total number of submissions to the NEOCP in the first year of LSST that pass the given score threshold. **Bottom:** Purity of submissions that pass the threshold (fraction of objects above the threshold that are NEOs).

Instead of deciding randomly between these objects with perfect scores, it would be more effective to first sort them based on the new score we have calculated. To investigate this we perform bootstrap sampling on the population of NEOs and MBAs that have scores of 100. We draw 52 NEOs and 468 MBAs for each sample since these are the average number of each objects that would be submitted to the NEOCP nightly and attain a score of 100.

For each of these samples we explore the effect of having a randomly ordered list or one sorted by the new score. Given that the community will only have capacity to follow-up a certain number of objects we can assess how many NEOs would be found when taking only the first $N_{\rm max}$ objects on the list.

We show the improvements gained by sorting the list in Figure 6. From top to bottom we assume values of 40, 100 and 200 for $N_{\rm max}$ and for each distribution we fit for the mean and standard deviation (mostly using Gaussians but also a Poisson for the distributions centred closer to zero). In each case we find that sorting the list results in more NEOs being followed-up each night. In particular, assuming an $N_{\rm max}$ of 100 would on average result in 7 more NEOs being follow-up each night, which would mean $\sim$2500 additional NEOs would receive follow-up observations.

This improvement is particularly important to consider in cases where the NEO would no longer be visible

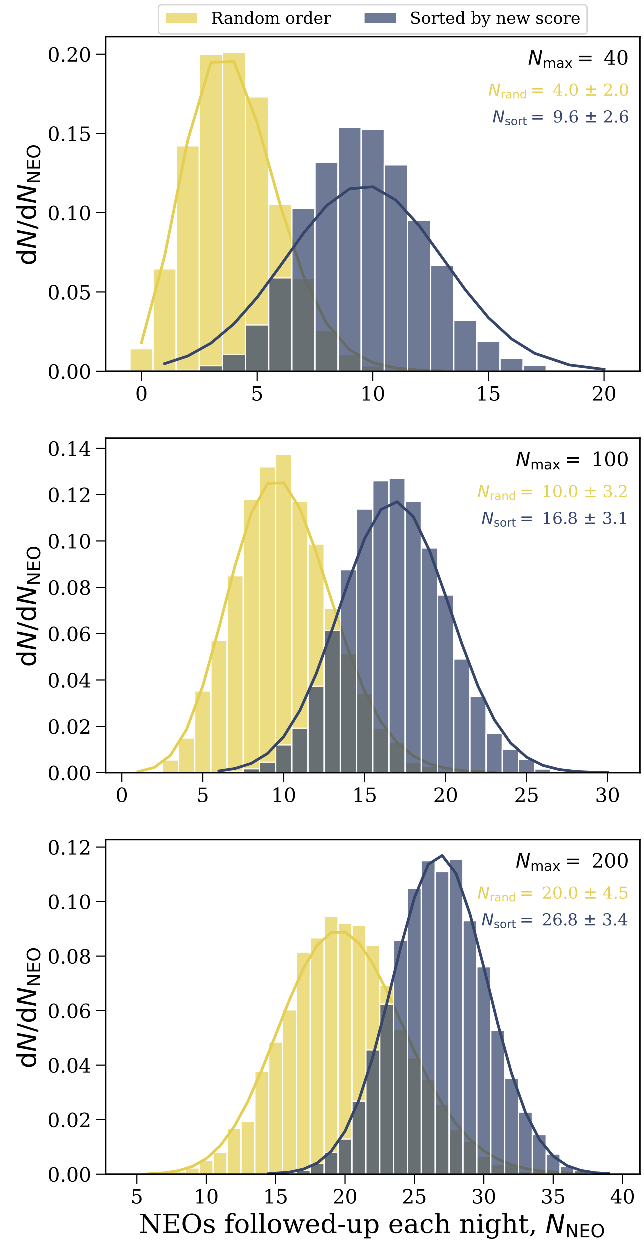

Figure 6. A demonstration of the benefit of sorting the NEOCP by our score. Histograms show the average number of NEOs that would be followed-up each night without sorting the NEOCP and using our score to sort respectively. Each panel corresponds to a different number of maximum follow-ups that could be attempted in a single night. Each panel is annotated with the mean number of follow-ups and the standard deviation.

on a following night - identifying it on that night using our sorting method could make the difference of whether the orbit of the object would be well constrained.

We recommend that the NEOCP continue to use `digest2` as a threshold for the page, but use this new score as a sorting metric for objects that receive the same `digest2` score. Moreover, this technique will work well in conjunction with the LSST self-follow-up probability algorithm presented in (Wagg et al. in prep) as these probabilities, which can be used as further thresholds beyond the `digest2` score, would also benefit from being sorted in this way.

4. CONCLUSION

- Point out whether we did better :shrug:

REFERENCES

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, doi: 10.1051/0004-6361/201322068

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabc4f

Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, apj, 935, 167, doi: 10.3847/1538-4357/ac7c74

Cornwall, S., Eggl, S., Juric, M., & Jones, L. 2020, in AAS/Division for Planetary Sciences Meeting Abstracts, Vol. 52, AAS/Division for Planetary Sciences Meeting Abstracts, 110.04

Grav, T., Jedicke, R., Denneau, L., et al. 2011, PASP, 123, 423, doi: 10.1086/659833

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111, doi: 10.3847/1538-4357/ab042c

Jones, R. L., Slater, C. T., Moeyens, J., et al. 2018, Icarus, 303, 181, doi: 10.1016/j.icarus.2017.11.033

Jurić, M., Kantor, J., Lim, K. T., et al. 2017, in Astronomical Society of the Pacific Conference Series, Vol. 512, Astronomical Data Analysis Software and Systems XXV, ed. N. P. F. Lorente, K. Shortridge, & R. Wayth, 279, doi: 10.48550/arXiv.1512.07914

Keys, S., Vereš, P., Payne, M. J., et al. 2019, PASP, 131, 064501, doi: 10.1088/1538-3873/ab1157

Naghib, E., Yoachim, P., Vanderbei, R. J., Connolly, A. J., & Jones, R. L. 2019, AJ, 157, 151, doi: 10.3847/1538-3881/aafece

Wagg, T., Juric, M., Yoachim, P., et al. in prep. https://www.tomwagg.com/html/paper_posts/lsst_neocp.html

APPENDIX

A. DISTRIBUTIONS OF PARAMETERS FOR ALL OBSERVATIONS

The following plots show the distribution for the three parameters that we consider for the entire population of observations - in contrast to the earlier plots that only consider observations with a `digest2` score of at least 65. In general one can see that the distributions are less clear cut, in particular for the apparent magnitude distribution.

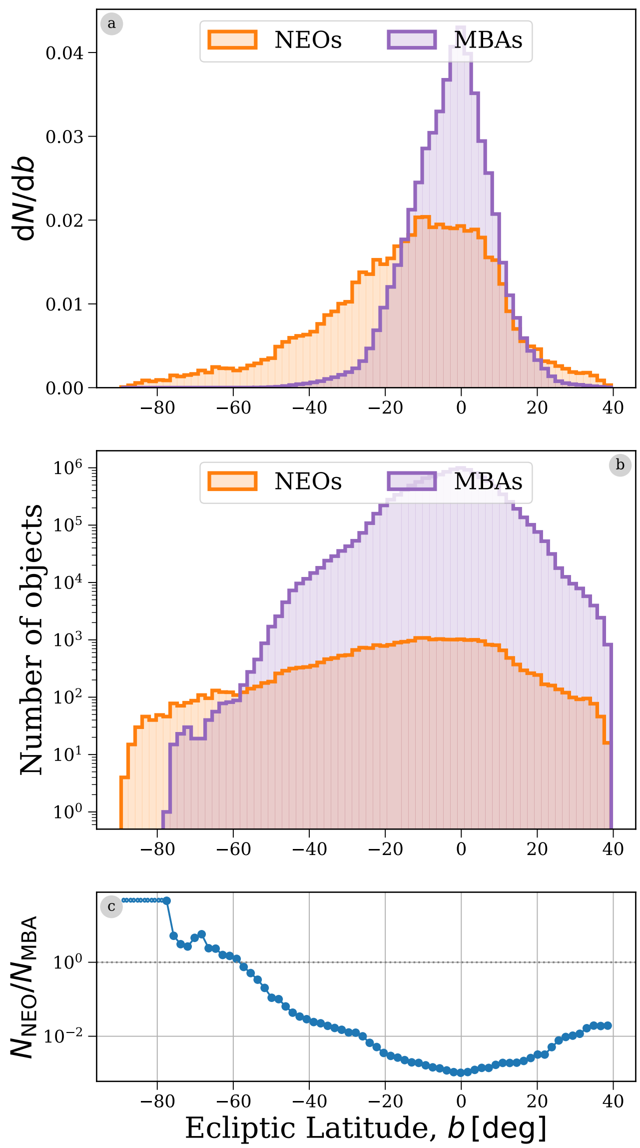

Figure 7. As Figure 1, but for ecliptic latitude.

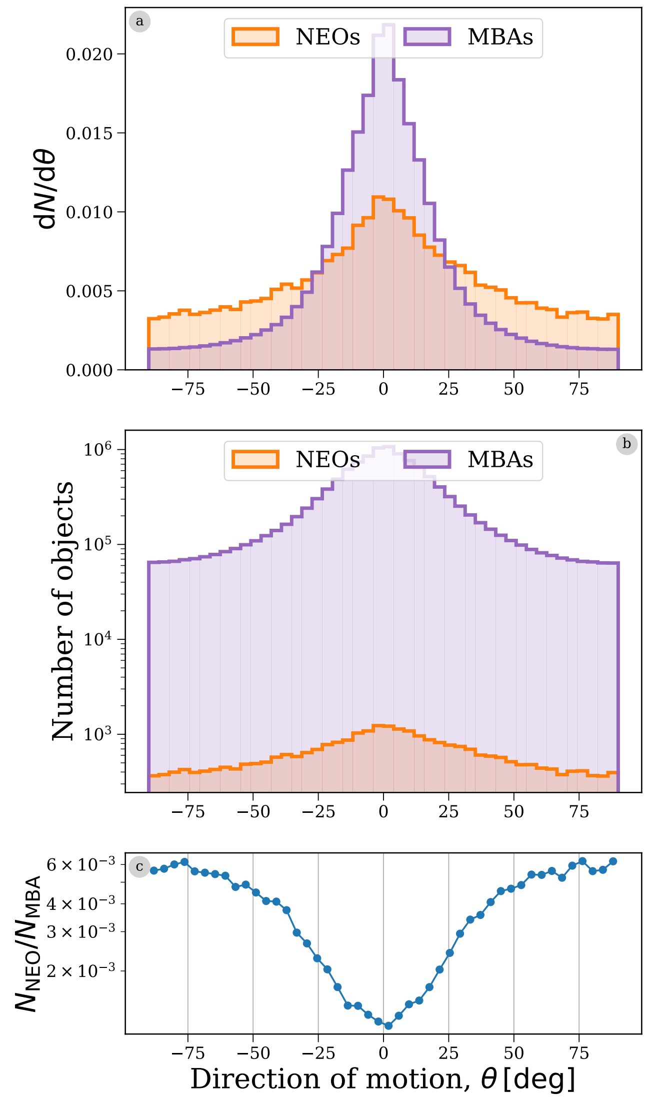

Figure 8. As Figure 1, but for direction of motion relative to the ecliptic plane.

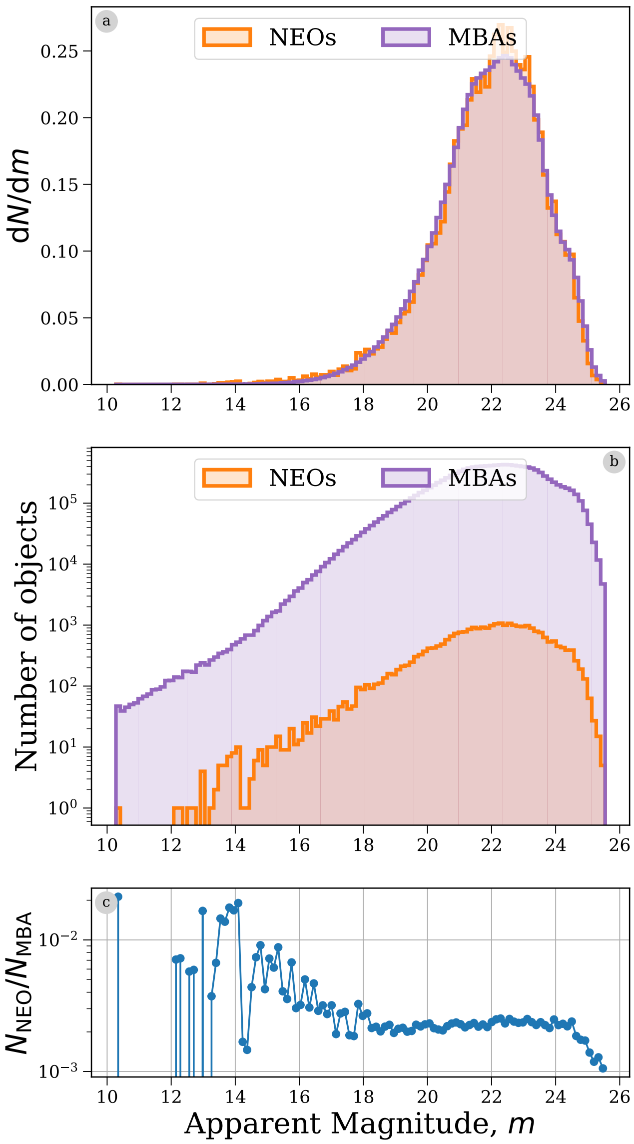

Figure 9. As Figure 1, but for apparent V-band magnitude.