Moving Object Discovery Engine for LSST Data:
Report from the IPAC Participation in the JPL Study of Asteroid Detection with LSST

George Helou, Frank Masci, Cate Liu, John Dailey, Carrie Nugent
(Caltech/IPAC)
June 2017

Executive Summary

We have adapted the new software package "Moving Object Discovery Engine" (MODE) to run on simulated Large Synoptic Survey Telescope (LSST) asteroid detections. The LSST simulated data are based on a typical LSST cadence, and include photometric and astrometric noise, spurious detections, as well as weather effects. The adaptation of MODE included algorithmic speed-ups, and a processing architecture that thins the input data iteratively to improve completeness. We find that in general MODE is able to recover asteroids, both Main Belt and Near-Earth Objects, from the LSST data with good completeness and high reliability (>95%). An important caveat is that because the realized LSST cadence and coverage are a strong function of position on the sky, MODE performance varies when used with fixed parameters. When its parameters are tuned to the local coverage, MODE achieves over 90% completeness (efficiency) and over 95% reliability (purity). We identify future MODE enhancements that will improve these results.

We did not find clear evidence that the baseline LSST cadence is the limiting factor globally in finding near-Earth objects. However, the realized cadence can make the recovery significantly more challenging in some patches of the sky compared to others, depending on the amount of clumping in time of the visits. A systematic exploration is needed of this and other LSST survey simulations in order to assess whether the basic LSST cadence may lead to patches where finding Near-Earth Objects is inefficient, and how to mitigate that possible deficiency. Such an exploration will help evolve a MODE processing architecture with the flexibility to deal with the significant variations in realized local coverage.

1. Introduction

The overview document by Chesley describes a study to provide an independent estimate of the performance of the LSST for discovery of Near-Earth Objects (NEO). The IPAC group participated in the study, but with limited interaction with the JPL group and the LSST group at University of Washington (UW) in view of the novel algorithmic approach used by IPAC. The IPAC activity aims at (1) the basic question of the suitability of the LSST survey scheme for finding NEOs within a time frame of days to weeks, and (2) the question of suitability of a new software system called "Moving Object Discovery Engine" (MODE) for processing LSST data for NEOs.

As described by Chesley, linking the stream of potential moving object detections generated by LSST is one of the more stressing elements computationally for finding asteroids, in view of the density of LSST moving object detections on the sky. The JPL and UW groups used different incarnations of the pre-existing Moving Object Processing System (MOPS) to link detections extracted from the same simulated LSST data. The IPAC group used the MODE system described below; the rationale for this choice is given at the end of Section 1.1. It is worth noting that the IPAC effort did not address the question of LSST efficiency for characterizing the overall NEO population over the lifetime of the survey; this is addressed in the Chesley & Veres report (JPL Publication 16-11).

This section (1) provides technical details on MODE and on its adaptation to the LSST simulation, and highlights some of the key differences between MODE and MOPS. Section 2 describes the simulated LSST data sets used in this study, and the subsets used for rapid exploration. Section 3 describes the performance of MODE under various conditions. Section 4 discusses the LSST cadence and its suitability for NEO finding, and section 5 offers summary and conclusions.

1.1. Moving Object Discovery Engine (MODE)

MODE was developed at IPAC for data from the Palomar Transient Factory (PTF, Law et al 2009, PASP, 121, 1395) and is planned for use on its successor, the Zwicky Transient Facility (ZTF, Bellm 2016; see Sec 6). It was developed for use on extractions from difference images, with provisions for scaling up to meet LSST's computational demands by optimally leveraging readily available hardware architectures. It uses a new algorithm, related to ideas presented in Waszczak et al (2013, MNRAS, 433, 3115), and conceived after a detailed study of the classic MOPS package.

MODE has been running in the nightly production pipeline on PTF data, with on-going refinements guided by its on-sky performance from PTF, and a goal to support ZTF, which enters nominal survey operations in January 2018. From studies of the recovery fraction of known objects (with well known orbits), MODE can detect moving objects at a completeness (efficiency) of > 90%, and a reliability (accuracy) of > 98% to moderately faint flux levels (R ~ 20 for PTF).

The most significant difference between MODE and MOPS is the first step in linking candidate moving object detections into building blocks of candidate tracks. MODE requires a minimum of three detections to form a moving-object stringlet, whereas MOPS starts out by forming two-detection "tuples". MODE forms stringlets by matching the relative velocities of two adjacent pairs of detections that share a common middle detection. Their velocities are matched within some tolerance that depends on the time separation of the detections. After all possible stringlets in the detection stream have been identified, they are linked using velocity matching on a

coarser grid to create moving-object candidate tracks. Two types of velocities are used in the final stringlet-merging step: (i) the mean velocity of the (intra-stringlet) detection-pairs; and (ii) the relative (inter-stringlet) velocities between the average position of detections in each of the stringlets. A schematic is shown in Figure 1. The candidate tracks from MODE are then vetted using a number of quality metrics, including an orbit fit to test if the candidate track is dynamically plausible.

By building three-detection stringlets, MODE moves part of the combinatorial challenge earlier in the processing. It helps eliminate spurious two-detection "tuples" that would otherwise add to the load at later stages of candidate track construction. Moreover, it significantly reduces the number of possible combinations that need to be processed and merged downstream, as opposed to MOPS which carries along all possible two-detection "tuples". The MODE design is therefore more efficient in regions of high source density for a given input timespan.

1.2 Processing Approach

In what follows, we will refer to the version of MODE specialized for the LSST case as LMODE. As with MOPS, a major challenge to LMODE is the spatial density of detections considered to be candidate moving object detections. In early test runs, we found the completeness to drop off from near 90% to around 30% as the density of detections went up from the PTF-like level to the LSST simulated levels.

To reduce the incidence of false linkages due to high source confusion when generating candidate tracks, we iteratively thin-out the detection candidate pool by removing detections associated with reliable candidate tracks identified at each iteration. Each iteration progressively allows for less stringent matching tolerances on velocity to capture cases with less linear motion. The matching tolerances are progressively increased, starting with a search for the most-linear and slowest-moving objects (i.e., within the tightest tolerances) because these are less subject to contamination from confusion during the linking process and can be reliably identified. Removing their associated detections from the input pool reduces the number of potential contaminants for the next (less stringent) iteration. The thinning-out is designed to first remove the more dominant Main-Belt Asteroids (MBAs) from the pool, in order to reveal more of the NEO population at later iterations. A reliable candidate track consists of five or more linked detections fitting accurately a dynamically valid orbit. The identification of reliable candidate tracks is based on the Find_Orb orbit-fitting software, which has been adapted and tuned to process candidate tracks from LMODE. We found that requiring five or more linked detections per candidate track (as opposed to six) consistently yielded a higher fraction of reliable candidate tracks and recovered more of the input simulated objects (i.e., a higher completeness).

This iterative thinning of the detection pool is another significant difference with the default usage of MOPS. A schematic of the process is shown in Figure 2. Due to the

limited computational resources available to this study, it was necessary to subset the initial simulated master detection list (that covers $\sim 3\pi$ steradians and one lunation or ~ 28 days) both spatially and temporally. This partitioning of the detection list also allows for a parallel processing architecture. For this study, we were limited to four computing nodes only. To enable repeated testing and tuning of the full parameter space with a reasonable turnaround time for the runs, we focused on three spatial regions with a limited time span as detailed in section 2 below.

As shown in Figure 2, the goal is to first eliminate from the detection pool as many as possible of the most abundant "contaminating" moving objects, the MBAs. The velocity-matching tolerances are gradually increased for the first N_mba (~ 30) iterations, where we target the velocity range characteristic of MBAs ($0.05 <\sim v <\sim 0.4$ deg/day). The Find_Orb thresholds that are used to select good quality orbits (and hence reliable candidate tracks) are held fixed during this MBA-thinning phase. When the number of reliable MBA-like candidate tracks drops significantly to several or a few (typically after ~ 30 iterations), the iterations are rerun with relaxed search tolerances for LMODE and Find_Orb to attempt to recover NEOs (against a thinned-out MBA sky). The velocity range searched for now is up to ~ 2 deg/day. Additional MBAs with non-linear and/or faster motions are also recovered during this phase. The penalty of running with relaxed tolerances (particularly for Find_Orb) is a slightly higher incidence of unreliable candidate tracks, but this is traded off against potentially more NEOs -- the objects sought in this study. Not shown in Figure 2 is special processing that occurs for the final two (NEO-based) iterations. Here, the Find_Orb thresholds to find good quality orbits are relaxed even further in an attempt to catch those NEOs with the most non-linear trajectories (again, at the expense of admitting more unreliable candidate tracks).

1.3 Updates and Tuning of LMODE

Considerable updates were made to the LMODE track-finding software. These updates improved runtime and memory use, but also allowed for a more generic search of candidate stringlets (MODE's building blocks) and their linkages to construct candidate tracks (see Section 1.1). More debug diagnostics and graphical outputs were added to facilitate the visual tracking of individual objects through the linking process. This information was crucial in tuning the iterative process described in Section 1.2, particularly LMODE's interplay with the Find_Orb orbit-fitting software. The author of Find_Orb (Bill Gray) also provided valuable advice and streamlined the software for this study.

More importantly, the LMODE input parameter set was tuned to match the LSST cadence used to simulate the input master detection list. The primary parameters here were constraints on the time-separation of the detections (observation epochs) used to construct the stringlets (Section 1.1). Two constraints were imposed when finding the candidate moving-object stringlets: (i) one of the detection-pairs in the

stringlet must span ≤10 hours (i.e., the same night); (ii) the other pair must span ≤ 3.5 days (however, see sec 3.2 below).

2. Data Sets and Assumptions

Input data were provided by the LSST team in file sample_noDD_noSingletons1.csv. This file was derived from the LSST survey simulation named Enigma1189, and contains only detections of asteroids. The details of the simulation are discussed in Chesley & Veres (JPL Publication 16-11).

Each detection had an associated detection identifier ("det_id"), modified Julian Date ("epoch_mjd"), position in right ascension and declination in degrees ("ra_deg" and "dec_deg" respectively), magnitude ("mag"), simulated observation filter ("filter_id"), field identification ("field_id"), the index of nights ("nn"), object provisional name ("object_name"), which began with "S1" for MBAs and "S0" for NEOs, the length of the semi-major axis of the detected source profile ("length_deg"), and the detected source profile's position angle East from North in degrees ("orient_deg"). The file contained 13,123,961 detections.

2.1 Data Sampling and Test Regions

The data set was sampled for initial exploration and verification of LMODE capabilities, forming subsets of detections to use as input. Sampling was carefully designed to avoid introduction of bias into results. Two types of sampling were explored.

In early exploration, the whole input file was "thinned out" to achieve reductions in detection counts by randomly selecting objects to keep. The goal was to run LMODE on files of manageable size given the computing hardware available. This led to samples with full sky coverage preserving the simulated properties but with a lower density of detections on the sky. Several such thinned all-sky samples were produced, with total detection counts ranging from about 50,000 to 500,000. We do not show results here from these early experiments, because they were primarily useful for improving LMODE.

The later runs relied on spatial sampling, which keeps all detections that form chains of 6 or more within circular sky regions spanning a predetermined range of roughly two weeks of data. Within those two weeks, data were available from twelve or fewer nights, since the simulation includes weather effects. This subsetting accommodates the simulated LSST cadence of three pairs of same-night visits occurring over a period as long as two weeks. Three different centers were selected, and each center was combined with three different radii for the selection, resulting in samples of 25 thousand, 50 thousand and 100 thousand detections respectively associated with each center. We will use the lables 25K, 50K, 100K for these samples. This approach preserves the density of detections to a large extent,

and tests LMODE against varying input file sizes. Region 1 is centered at ct1 with RA=325deg, Dec=-32deg [elon, elat = 317, -17deg], Region 2 is centered at ct2 with RA=210deg, Dec= -5deg [elon, elat = 210, +7deg], and Region 3 is centered at ct3 with RA=200deg, Dec=-30deg [elon, elat = 211, -20deg]. These regions were identified as representative of various object densities and asteroid sky motions. Region 1 was selected for its very high density, roughly ten times higher than the density in Region 3. The density of detections reflects the density of observable asteroids, but is affected by cadence and survey scheme, sensitivity of observations and other factors entering the simulation. The solar elongation of Regions 1, 2, and 3 is 110, 140, and 137 degrees, respectively, with Region 1 in the same hemisphere as Earth motion vector and Regions 2 and 3 in the hemisphere opposite to Earth motion vector.

Table 1 summarizes the samples, their centers, the number of detections and objects in each, and the radius of each region/sample combinations. Plots illustrating one of the sampled fields are shown in Figures 3 and 4.

Table 1: Details of thinning based on region

Sample Name: Center-Count	Region center	Number of detections	Number of objects	Number of NEOs	Number of MBAs	Radius (degree)
C1-25K	ct1	24,713	2433	7	2426	3.11
C2-25K	ct2	24,943	3277	4	3273	5.89
C3-25K	ct3	25,182	3546	12	3534	10.44
C1-50K	ct1	48,905	4879	23	4856	4.33
C2-50K	ct2	49,447	6474	8	6466	8.77
C3-50K	ct3	50,040	7180	32	7148	14.39
C1-100K	ct1	98,910	9485	39	9446	6.17
C2-100K	ct2	99,579	12801	18	12783	10.67
C3-100K	ct3	100,099	14503	56	14447	17.67

Figure 5 shows histograms of the numbers of detections occurring in each night for each of the first three samples and the last three samples in Table 1. Note the very different distributions of the detections over nights among the three regions, and note that these differences persist even as the solid angle of the region increases by a factor of 3 to 4.

2.2 Assumptions

A significant simplifying assumption for this study was that false positive control would be achieved at a high level of confidence by filtering input detections before this step of moving object finding. Experience with PTF data is strongly encouraging, where machine-learned filters are able to suppress false positives significantly (Masci et al 2017, PASP 129, 014002). For the test runs, a minimum of 10% false-positive detections were added to simulate unfiltered noise detections.

Runs with additional false positives were also completed to test for robustness in case the 10% assumption proves too optimistic. The 25k cases in all three regions were repeated with 50% and 100% false positives injected, and the results are reported below.

2.3 Definition of Reliability and Completeness

Reliability is defined as the fraction of candidate tracks produced by the software (LMODE) that successfully recover a moving object in the simulation. A track is considered successful in recovering an object if it contains 5 or more detections of that object. For reliability, uniqueness of an object in the output set is ignored. Reliability is a term describing the success or failure of each track, and applies to the output set of analysis without consideration of type of object, i.e. MBA or NEO.

Completeness is defined as the fraction successfully recovered by the software (LMODE) of objects meeting the minimum requirements for detection in our input set. The minimum requirements for detection are five or more detections of the object present in the input. Success in recovering an object has the same criteria for calculating completeness as it does for reliability. Completeness measures the performance of object detection as it relates to the known input set. As such, it can be used to measure performance on any sub-population of that input set, and can be reported separately for NEOs and MBAs, or other subsets of the population.

3. MODE Performance

3.1 Completeness and Reliability for Nominal Runs

The results for the nominal runs are presented in the three panels of Figure 6. Each panel refers to a given sample size (25K, 50K or 100K) and shows the resuts for each of the three regions described in Table 1. The cumulative completeness for all objects brighter than a given magnitude is plotted as a function of magnitude. The reliability is given as a single number for each Region and sample size in the legend. It is immediately evident that reliability runs at or above 95% in most cases. While Region 1 shows the highest reliability, there are no strong trends as a function of Region or sample size.

Completeness shows more variability, with strong dependence on magnitude, and also on Region. For Regions 1 and 3, completeness runs mostly above 90%, dipping below that line only at magnitudes fainter than 23 for the 25K and 50K runs, and at magnitudes fainter than ~21 for the 100K runs.

Region 2 reports less complete results than Regions 1 and 3, running between 60 and 80%. After analyzing various possibilities, we ascribe this behavior to the significantly different distribution over time of the detections for Region 2. Figure 5

shows that Region 2 is unique in having a stretch of 6 nights during which no detections are reported. Region 1 has a 4-night detection gap, and Region 3 a 2-night gap. Interestingly, even as the sample size is enlarged from 25K to 100K, the gaps do not shrink, and Region 2 remains more incomplete than 1 or 3.

3.2 The Impact of Cadence

The distinct behavior of Region 2 suggested pursuing the question of the relative importance of cadence vs. location on the sky. We created three new samples of detections by varying the time sampling while keeping the location fixed. The detections were gathered from Region 2 in Table 1, within the radius of 5.89 deg originally chosen to yield 25K detections. Three time windows were chosen, each encompassing 12 nights, and again only objects detected 6 times or more within that time window were retained. The three time windows were picked to span the lunation as simulated, with a small amount of time overlap among them: w1=[21, 32], w2=[29, 40], and w3=[37, 48], where the numbers in square brackets refer to start day and end day for each time window. The resulting samples contained approximately 42K, 4K and 12K detections each.

We ran LMODE on each of these new samples. The results in completeness and reliability are shown in Figure 7. Overall completeness for these time-window samples shows significant improvement, as does reliability. Moreover, the sample size varies significantly both up and down from the original 25K, and that variation does not seem to affect the performance of LMODE in recovering asteroids. The tentative conclusion is that cadence is more important than sky location, with more investigation needed to quantify cadence metrics that would correlate well with LMODE performance in asteroid and NEO recovery.

We also explored varying some parameters and exercising other features in LMODE to improve its performance on the data at ct2. One parameter relevant to coverage gaps in time is the search window for constructing 3-detection stringlets (Fig 1). Another such parameter is the search window for connecting stringlets (Fig 1). We increased both of these parameters and re-ran the processing of Region 2. There was very limited improvement in the completeness.

LMODE has an option (not used in the main study here) to keep more than one stringlet with the same middle detection and slightly degraded linearity, as opposed to retaining the "best" (most linear) stringlet. Another option (again not used in the main study) is to allow the same stringlet to pair wih multiple other stringlets rather than to the "best match". Both of these options were exercised in dedicated test runs on Region 2. The results were promising but not conclusive.

3.3 The Impact of False Positives

In most of the analysis we injected into the simulation a false-positive population of detections that amounts to 10% of the real detections (sec 2.2 above). However, we did experiment with increasing that rate to test the performance of LMODE against such an increase. We ran for each of the 25K cases in each Region 2 additional cases, one with 50% false positives and one with 100% false positives. The intent is to represent all possible sources of false positives, from artifacts in image differencing to real transients and source variability. Since the distribution over magnitude of these various populations of false postivies will be complicated to model, we adopted a simple distribution which is flat over magnitude.

The results are shown in Figure 8. The effects of an increased rate of false positives is very modest. Reliability remains the same, and completeness is degraded by barely a few percent at the faint end, meaning for the whole sample, since we plot cumulative completeness. The degradation in completeness is slightly greater at the bright end because the distribution is uniform in magnitude, and therefore the effective false-positive rate is much higher at the bright end.

3.4 The Performance on NEOs

The numbers of NEOs from the LSST simulation are relatively small in the region samples (Table 1), so the results are subject to small number statistics. The completeness estimates are shown in Table 2, and they range from 53% to 100%, with seven out of nine cases reporting 2/3 or better completeness. These data are also plotted in Figure 9. The last column in Table 2 shows the completeness range formally associated with the recovery and input numbers. The Poisson distribution interval corresponding to 16% to 84% confidence, i.e. equivalent to $\pm 1\sigma$ for a normal distribution, was evaluated for the recovered number of NEOs and shown as a fraction of the input number.

Interestingly, Region 2 does not lag behind Regions 1 and 3 in completeness, but is rather intermediate, with Region 3 the clear outlier in this respect. Several factors could play into this reversal, but small number statistics are quite unlikely to provide the explanation, since completeness does not depend on sample size. The strongest dependence is between completeness and Region, pointing again to the importance of cadence in determining the recovery rate. See Section 4.1 for a qualitative discussion of this dependence.

Table 2: Completeness of NEO recovery per test region

Sample Name: Center-Count	Number of NEOs (input)	Nr of NEOs Recovered	Completeness	$\pm 1\sigma$ equivalent interval
C1-25K	7	7	1.00	>0.57
C2-25K	4	3	0.75	>0.33

C3-25K	12	8	0.67	>0.38
C1-50K	23	20	0.87	>0.65
C2-50K	8	6	0.75	>0.38
C3-50K	32	17	0.53	0.39—0.66
C1-100K	39	31	0.79	0.64—0.92
C2-100K	18	13	0.72	0.50—0.89
C3-100K	56	32	0.57	0.46—0.66

4. Discussion

4.1 Cadence

The input simulated master detection list used here was constructed from an early simulation run that followed the originally proposed LSST cadence: two intra-night visits occurring three times in a two-week window. For comparison, the PTF survey used two intra-night visits on consecutive nights (depending on the observing program, weather, etc.). Despite PTF reaching depths much shallower than those envisaged for LSST (factors of ~100 less), MODE was still able to recover an appreciable fraction of moving objects in high source-density regions (e.g., the galactic plane), with tracks consisting of typically 4 or 5 linked detections. These regions are of comparable density to LSST's high galactic-latitude sky, though the PSF in LSST is expected to run 2 to 3 times smaller than in ZTF. The reason for a relatively higher recovery rate using the (shorter) PTF-based cadence is that the search-space was smaller, since the apparent distance travelled by objects on the sky over 2-3 days is shorter. This shorter span limits the search area and hence number of potential detections to test when constructing candidate stringlets and tracks through velocity matching. The number of input detections scales with area. A larger search area leads to a higher rate of false linkages and strains compute resources when forming and testing all possible linkages from the detection stream. Our experiments show that LSST's longer observing cadence and its depth compound the underlying combinatorial problem when generating reliable tracks. The originally proposed LSST cadence can indeed recover an appreciable fraction of moving objects using a delicately tuned discovery system, as we show, but at the expense of computational resources.

The post-facto LSST cadence is clearly not uniform on the sky, at least in this simulation. LSST is likely to evolve their survey cadence guidelines, resulting in further non-uniformities in post-facto cadence, both in space and in time. As the gaps between visits grows larger, recovery becomes more challenging.

4.2 Towards A Realistic Implementation

In designing a realistic system for mining the LSST data stream for asteroids and NEOs, it will be necessary to improve on this prototype work in a number of ways:

(1) Better orbit fitting, both faster algorithms and higher accuracy.
(2) Weeding out known asteroids from the input stream, with benefits increasing as the survey progresses and more faint asteroids are found.
(3) Optimized partitioning of the sky into search zones, to make the search more tractable computationally, especially that each zone may require slightly different tuning of the processing parameters.
(4) In parallel with the partitioning of the sky, the time windows for processing will also need to be optimized, with rolling windows adapted to position on the sky and master LSST cadence.
(5) Further tuning of the algorithms with real sky data, and the exploration of options such as those described in sec 3.2, with potential to improve performance markedly in areas with challenging cadences.
(6) General speed-up of algorithms and streamlining of computing architecture to support production configurations rather than the exploration work described here.
(7) Methods to detect and utilize trailed image detections (already in place for ZTF).
(8) Additional vetting of candidate tracks before they are forwarded to MPC.

5. Conclusions

Are there fundamental barriers to MODE as the search engine to asteroids/NEOs in the LSST data stream generated by the nominal survey? The answer is no, based on our ability to achieve high reliability and completeness for subsets of the simulated data, albeit with some variations. Bearing in mind that that MODE is a significantly less mature product than the current benchmark (MOPS), this simulation exercise represents great progress towards a full implementation.

Is the default cadence as implemented in the LSST simulation data set capable of supporting an asteroid/NEO search? The answer is overall yes, assuming the specific simulation used here is representative. While certain time and sky location combinations proved challenging, it was possible to make inroads and improve LMODE performance. These challenging spots were a minor component of the data set, and could be mitigated by modifying the time window for subsetting the data.

Tuning to actual data sets may become the pacing item, especially if cadence varies significantly during the survey. How hard is that? We found that running with one set of parameters and options generated quite good results overall. If the sliding time window can address the challenging spots as found here, it may well be that

little tuning is needed on an on-going basis. On the other hand, the modest amount of parameter and option exploration in this exercise suggests that MODE has wide latitude for adjusting to variable cadence and difficult spots.

Is LMODE capable of scaling up to meet a full implementation challenge? Based on the current codebase running on limited hardware, LMODE can meet the full load of the LSST survey data stream. Moreover, ZTF will start its science survey in January 2018, providing a real-world benchmark to enable a more quantitative evaluation, with projected data rate and alert rate about 10% of the LSST rates.

6. Ackowledgements

Figure 1. Schematic showing the construction of moving-object tracks in MODE (Section 1.1).
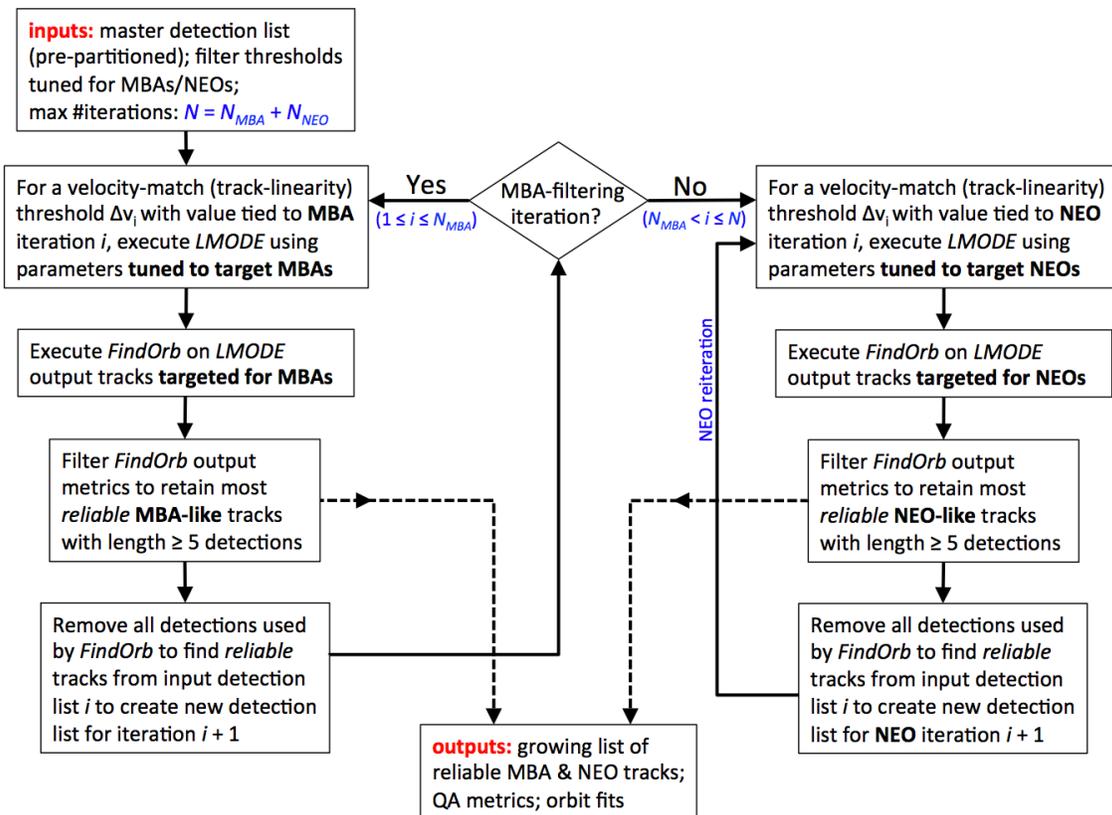
Figure 2. Overview of iterative-thinning approach using LMODE together with the orbit-fitting software, Find_Orb.
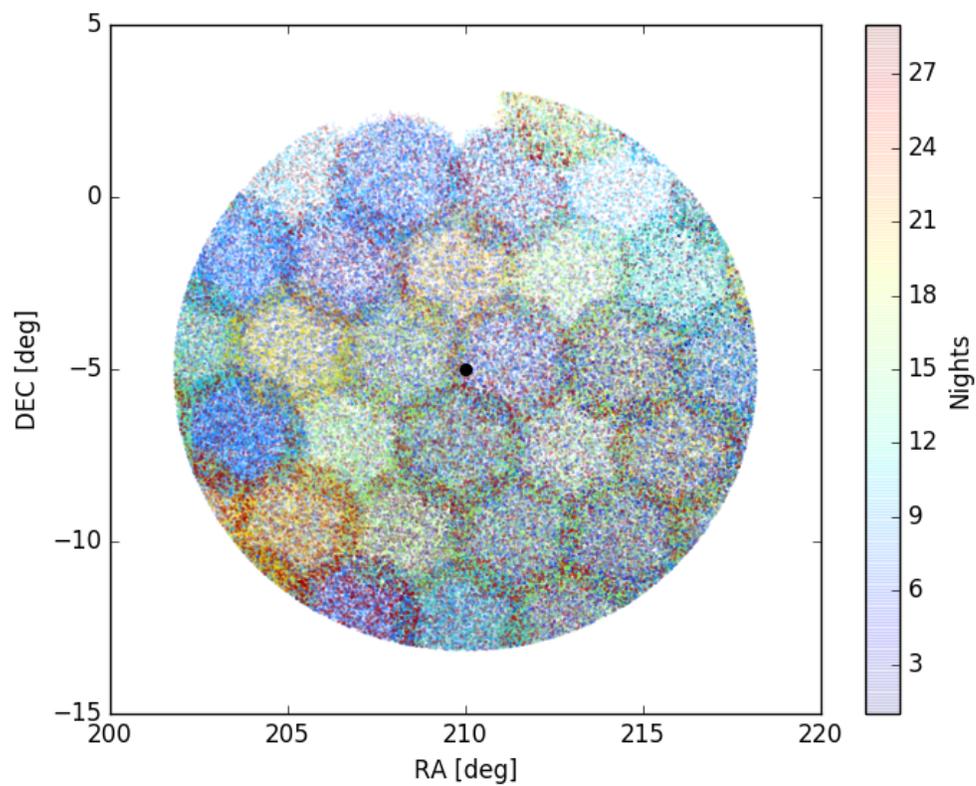
Figure 3. All detections contained in the original simulation data provided by LSST within a radius of 8.77 deg of the Region 2 center, shown as a black dot. Detections are color-coded by number of nights. The hexagonal patterning is presumably a reflection of observing pattern.
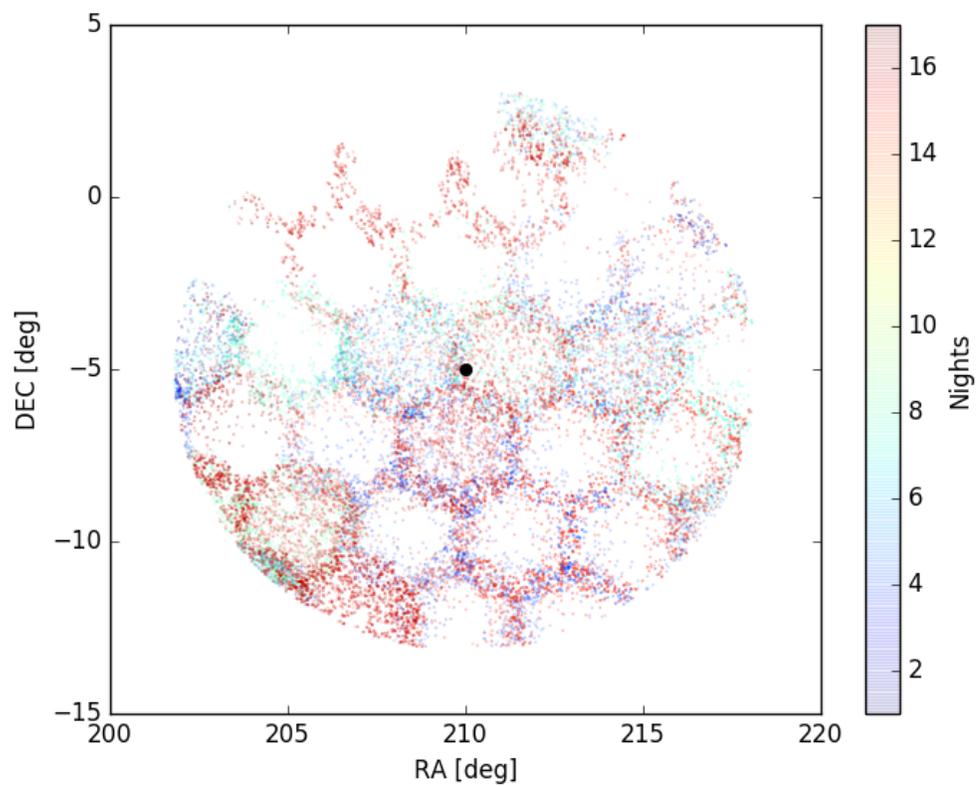
Figure 4. The detections retained in the 50k subset within a radius of 8.77 deg of the Region 2 center, shown as a black dot. Detections are color-coded by number of nights. The hexagonal patterning is a signature of the original data, and is presumably a reflection of observing pattern.
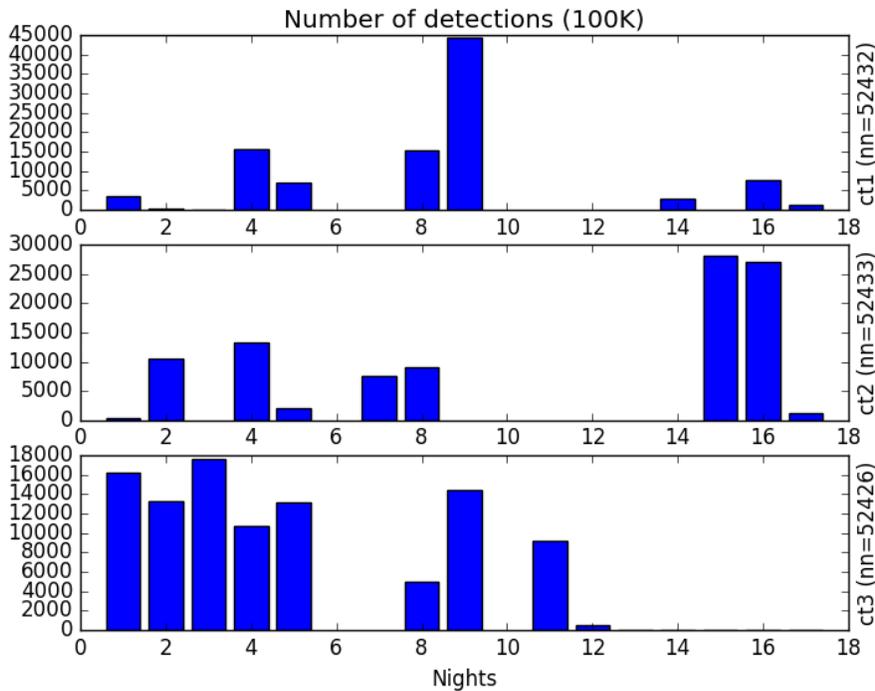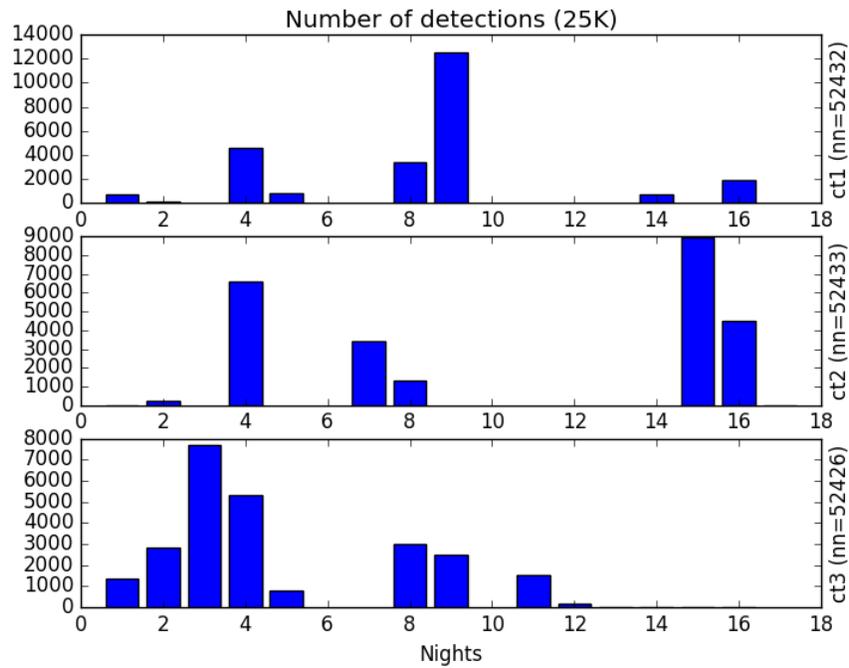
Figure 5. The histograms show the numbers of detections occurring in each night for each of three regions used in exploring the LSST simulation. The regions are identified by ct1, ct2 or ct3 (right-hand side vertical label). The upper group of three frames is for the small radius regions containing about 25,000 detections each, whereas the lower group is for the large radius regions containing about 100,000 detections. See Table 1 for more details. Note the very different distributions of the detections over nights among the three regions, and notice that they persist even as the radius grows to >10deg.
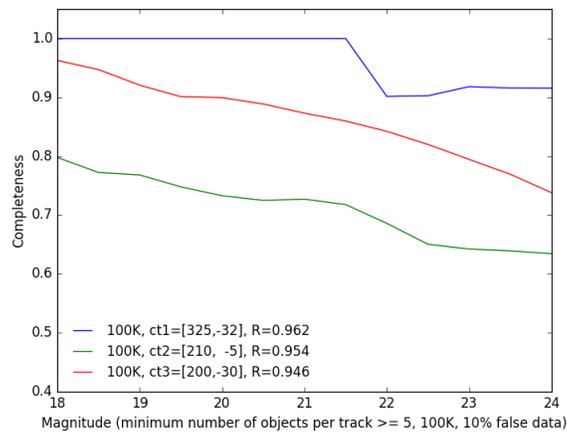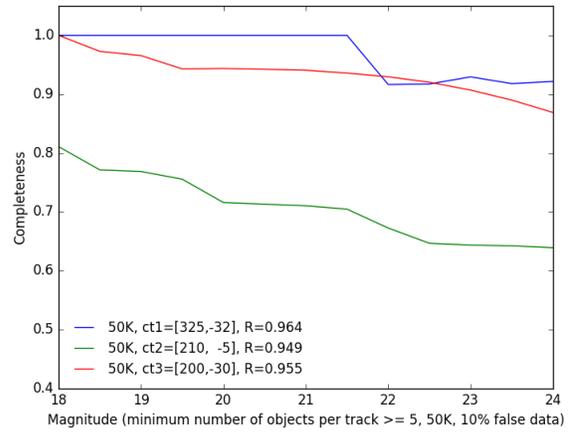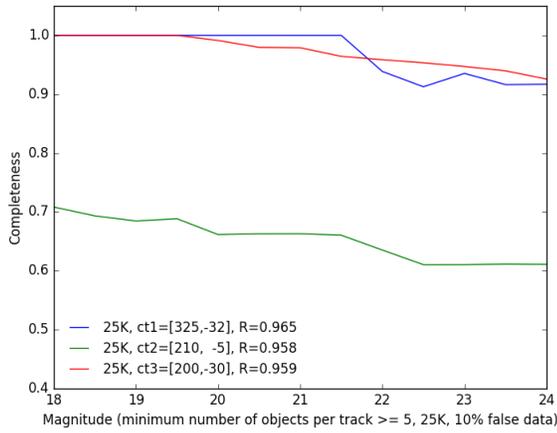
Figure 6. The cumulative completeness and reliability for all asteroids the three test regions described in Table 1. The completeness is plotted as a function of magnitude. The reliability is given as a single number for each Region in the legend. Each plot refers to one sample size (25K, 50K or 100K) and presents results for each of the three regions.
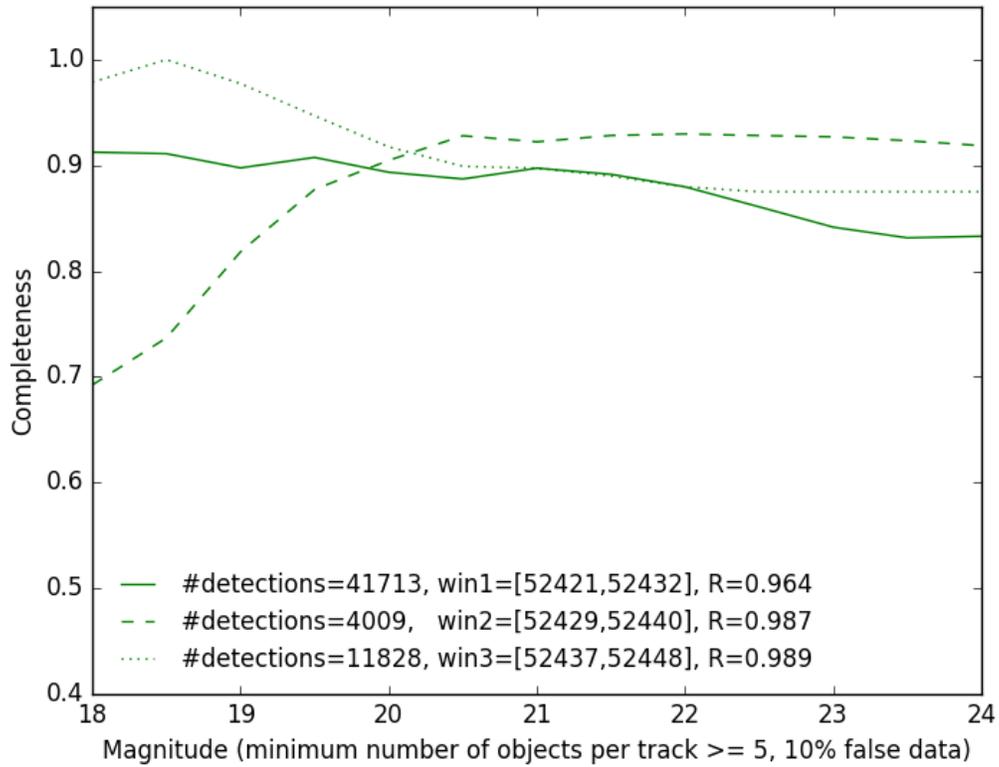
Figure 7. The cumulative completeness and reliability for the three time-window samples all centered on Region 2 with radius 5.89deg as described in Table 1. The legend gives the number of detections in each sample, the time window used and the reliability. These samples achieve marked improvement in completeness compared to the original sample, and a significant improvement in reliability.
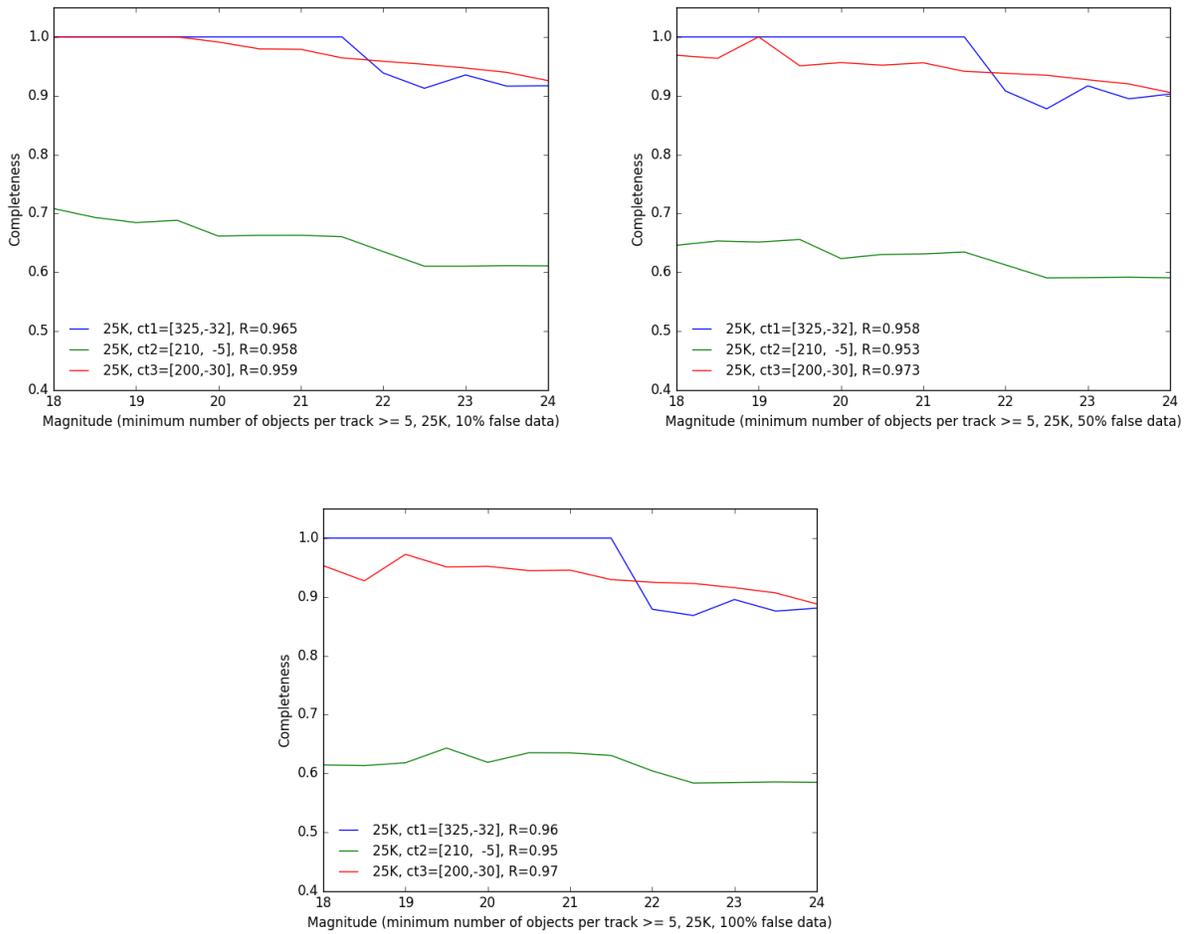
Figure 8. The cumulative completeness and reliability for all asteroids in the three test regions described in Table 1, retaining only 25K detections, but increasing the number of false positive detections in the simulation. The rate of false positives goes from 10% in the left-hand upper panel to 50% in the right-hand upper panel, to 100% in the lower panel.
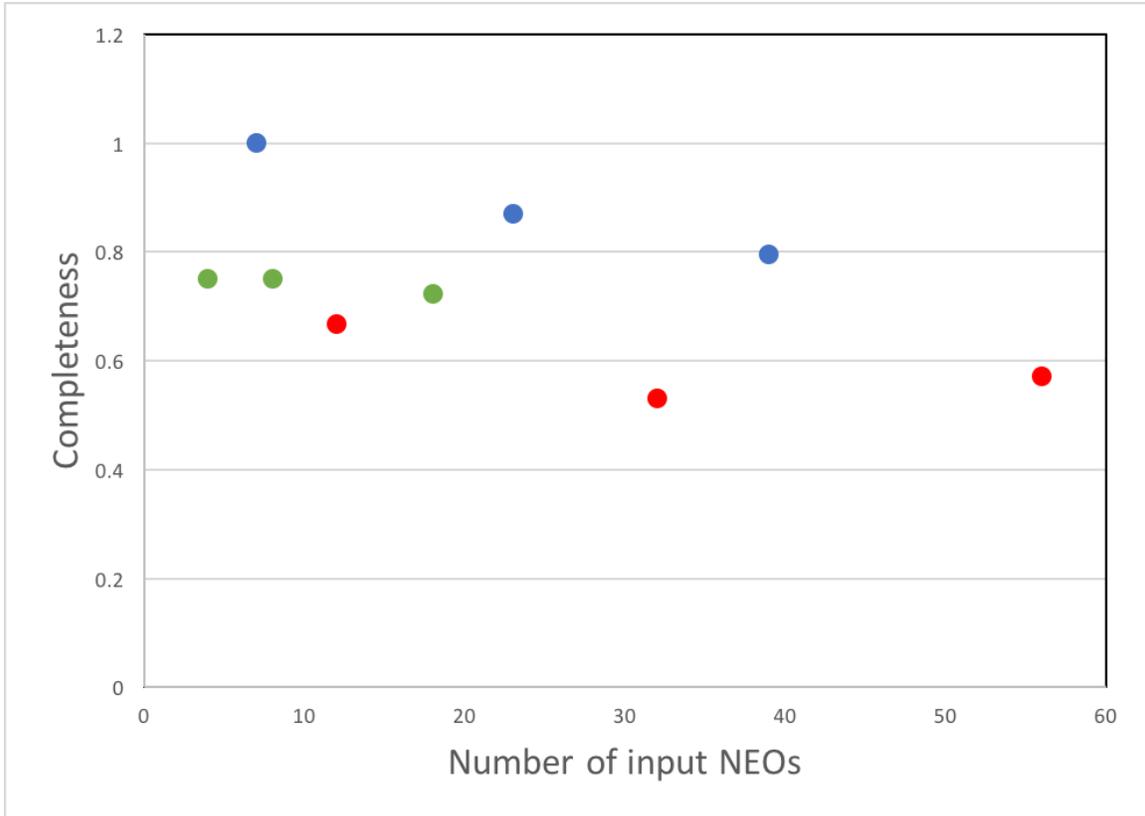
Figure 9. The completeness for NEOs, shown as one data point for each of the sample sizes in each of the three test regions as described in Table 1. Completeness is plotted against the number of NEOs in the input sample, to illustrate that small number statistics are an unlikely explanation for the observed behavior. The top three symbols (blue) refer to Region 1 samples, the three middle symbols (green) to Region 2, and the bottom 3 symbols (red) to Region 3. The rate of false positives is 10% all cases.