

# Towards Efficient and Precise Queries Over Ten Million Asteroid Trajectory Models

Yusra AlSayyad, K. Simon Krughoff, Bill Howe, Andrew J. Connolly,  
Magdalena Balazinska, Lynne Jones

University of Washington, Seattle, WA  
{yusra@astro, krughoff@astro, billhowe@cs, ajc@astro, magda@cs, ljones@  
astro}.washington.edu

## 1 Introduction

The new generation of telescopes under construction return to the same area of the sky with sufficient frequency to enable tracking of moving objects such as asteroids, near-earth objects, and comets [4, 5]. To detect these moving objects, one image may be subtracted from another (separated by several days or weeks) to differentiate variable and moving sources from the dense background of stars and galaxies. Moving sources may then be identified by querying against a database of expected positions of known asteroids. At a high-level, this task maps onto executing the query: “Return all known asteroids that are expected to be located within a given region at a given time.” We consider the problem of querying for asteroids in a specified interval in space and time, specifically as applied to populating the simulations of the data flow from the Large Synoptic Survey Telescope (LSST).

Spatio-temporal databases have been well studied[3], however this problem introduces new challenges. (1) *Number of Objects*: A characteristic Solar System model<sup>1</sup> contains over 11 million asteroids[2]. (2) *Complex Trajectory Models*: An asteroid’s trajectory can be calculated precisely<sup>2</sup> at any given point in time, but evaluation of the model describing its position is prohibitively expensive at query time. (3) *Accuracy Constraints*: The LSST simulated images require accuracy within 5 milliarcseconds<sup>3</sup> (mas), precluding the use of coarse statistical approximation methods. Additional requirements include 30-second query response time for a 9.6 sq. degree circular search area and a 10TB storage limit.

Within this context, we develop and evaluate two approaches. (1) Model each trajectory with a *bounding envelope*, use a spatial index to reduce the search space, and evaluate the exact positions of the asteroids at the given epoch using the ephemeris calculation code. (2) Model each trajectory by a set of positions sampled frequently enough to *interpolate* their positions within a given accuracy threshold at runtime. We implement all solutions in a relational database and evaluate them on a dataset of asteroid trajectories. Please refer to our technical report for details[1].

<sup>1</sup> representing the asteroids potentially observable by LSST and Pan-STARRS.

<sup>2</sup> from its orbital elements - six variables that completely describe an orbit

<sup>3</sup> For comparison, 1 milliarcsecond (mas) corresponds to the apparent diameter of a dime about 3700 kilometers away.

## 2 Evaluation

We find that the bounding envelope method provides exact accuracy and tunable storage requirements, but at the cost of lengthy query times. Total response time is extended by the necessary calculation of exact positions within the returned envelopes. The interpolation method can drastically reduce query time, but at the expense of either accuracy or storage space. With third degree polynomial interpolation, a sampling rate meeting the accuracy requirements exceeds our available storage. Table 1 summarizes a subset of the storage, accuracy and query time tradeoffs for “worst-case” searches in the densest region of the sky (returns 20K objects per 9.6 sq. deg. search area). This dense region forms a narrow strip across the sky. Typical searches in less dense regions (at least  $20^\circ$  from the ecliptic) return  $<1000$  objects in under 2s for both methods.

**Table 1.** Preliminary Evaluation Summary

		Storage <sup>a</sup>	Accuracy	Response Time	
		(billion rows)	Error (mas)	Query <sup>b</sup> (s)	Evaluate positions <sup>c</sup> (s)
<b>Bounding Envelope</b>	Biweekly <sup>d</sup>	45	exact	20	+15
	Daily <sup>e</sup>	80	exact	20	+4
<b>Interpolation</b>	Inaccurate	200	$<1000$	3	-
	Accurate	900	$<5$	3	-

<sup>a</sup> Storage required to support queries over a 10yr range

<sup>b</sup> Bounding env. uses MS Spatial Index; interpolation a hierarchical triangular mesh.

<sup>c</sup> Time to evaluate 20K exact positions with ephemeris calculation code

<sup>d</sup> Store daily bounding envelopes and biweekly orbital elements

<sup>e</sup> Store daily bounding envelopes and daily orbital elements

The challenge of predicting the positions of sources extends beyond the question of tracking asteroids. In general, our techniques are applicable to a class of spatio-temporal trajectory search problems, where the true positions of the objects can be predicted by the evaluations of complex, often non-linear models that are extremely accurate but computationally expensive.

**Acknowledgments** This work is funded by the NSF Cluster Exploratory (CluE) grant (IIS-0844580) and NSF CAREER award (IIS-0845397).

## References

1. Y. AlSayyad, K. S. Krughoff, B. Howe, A. J. Connolly, M. Balazinska, and L. Jones. Technical Report UW-CSE-11-04-03, University of Washington, April 2011.
2. T. Grav, R. Jedicke, L. Denneau, M. J. Holman, T. Spahr, and Pan-STARRS Moving Object Processing System Team. The Pan-STARRS Synthetic Solar System Model and its Applications. volume 38 of *Bulletin of the American Astronomical Society*, pages 807–+, Dec. 2007.
3. R. Güting and M. Schneider. *Moving Objects Databases*. Morgan Kaufmann, 2005.
4. N. Kaiser and et al. Pan-STARRS: A Large Synoptic Survey Telescope Array. volume 4836 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 154–164, Dec. 2002.
5. LSST Science Collaborations. LSST Science Book, V2.0. *ArXiv*, Dec. 2009.