Icarus 412 (2024) 115949

Contents lists available at ScienceDirect

Icarus

journal homepage: www.elsevier.com/locate/icarus

Research Paper

Sparse multi-apparition linkages in large datasets

Óscar Rodríguez [a,b,*], Giovanni F. Gronchi [c], Giulio Baù [c], Robert Jedicke [d]

[a] Dept. Matemàtiques, Universitat Politècnica de Catalunya, Av Diagonal 647, 08028 Barcelona, Spain
[b] IMTech, Universitat Politècnica de Catalunya, Pau Gargallo 14, 08028 Barcelona, Spain
[c] Dipartimento di Matematica, Università di Pisa, Largo B. Pontecorvo 5, 56127 Pisa, Italy
[d] Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822-1839, USA

ARTICLE INFO

Keywords:
Orbit determination
Keplerian integrals methods
Linkage problem
Asteroid surveys

ABSTRACT

We present a new procedure to identify observations of known objects in large data sets of unlinked detections. It begins with a Keplerian integrals method that allows us to link two tracklets, computing preliminary orbits, even when the tracklets are separated in time by a few years. In the second step, we represent the results in a 'graph' where the tracklets are the nodes and the preliminary orbits are the edges. Then, acceptable '3-cycles' are identified and a least squares orbit is computed for each of them. Finally, we construct sequences of $n \geq 4$ tracklets by searching through the orbits of nearby 3-cycles and attempting to attribute the remaining tracklets. We calculate the technique's efficiency at identifying unknown objects using real detections that attempt to mimic key parameters of the Minor Planet Center's (MPC) Isolated Tracklet File (ITF) and then apply the procedure to the ITF. This procedure enables the recovery of several orbits, despite some having few tracklets per apparition. The MPC accepted > 95% of our linkages and most of the non-accepted linkages are 2-apparition linkages even when those linkages contained more than half a dozen tracklets.

1. Introduction

In recent years, there have been significant developments in the observational techniques employed for detecting asteroids which have resulted in a marked increase in the number of asteroid detections. This trend is anticipated to continue with the forthcoming surveys, such as the Rubin Observatory Legacy Survey of Space and Time (LSST) (Ivezić et al., 2019), which will survey the sky more comprehensively and deeply than previous endeavors. The LSST is expected to detect millions of asteroids, including many that are too small or too faint to be detected by current surveys. This will provide a wealth of data and will help us to better understand the population of asteroids in our solar system.

Detections of asteroids are usually grouped into tracklets of very short arcs, each referring to the same observed object. These tracklets are collected over a few days and are used to compute the orbit of an asteroid. If the tracklets are not successfully used to compute an orbit they are stored in the isolated tracklet file[1] (ITF), a database maintained by the Minor Planet Center. The data in the ITF are mainly provided by the Pan-STARRS1 (Denneau et al., 2013) and Catalina surveys (Christensen et al., 2016) which are both large programs that have been successful in discovering and tracking asteroids. These two observatories have provided more than 4 and 2 million observations, respectively. With the work done in recent years, (e.g., see Sansaturio and Arratia, 2012; Holman et al., 2018; Weryk et al., 2020), the size of the ITF has been considerably reduced. Sansaturio and Arratia (2012) used an identification technique called 'attribution type' to compute the orbits of asteroids, particularly near-Earth asteroids (NEAs), by taking into account their higher apparent rates of motion. HelioLinC (Holman et al., 2018) employed a tracklet clustering technique to define an algorithm with a complexity of $\mathcal{O}(N \log N)$, where N is the total number of tracklets. Weryk et al. (2020) developed techniques to optimize the multi-apparition linking of tracklets based on their apparent rates of motion despite being far from their predicted locations.

Most initial orbit determination methods (e.g., see Laplace, 1780; Lagrange, 1783; Gauss, 1809; Gronchi et al., 2021) are based on the two-body equations of motion and rely on Taylor's series expansions around a central time. If the detections are widely spaced in time the initial orbit may not be accurate or computable.

The Keplerian integrals methods (Gronchi et al., 2015, 2017) impose the conservation laws of Kepler's dynamics (angular momentum, Laplace-Lenz vector, and energy) to calculate a preliminary orbit from the information contained in two or three tracklets. The main advantage of these methods is that they do not impose constraints on the time separation between the tracklets. The idea of using conservation

* Corresponding author at: Dept. Matemàtiques, Universitat Politècnica de Catalunya, Av Diagonal 647, 08028 Barcelona, Spain.
E-mail address: oscar.rodriguez@upc.edu (Ó. Rodríguez).

[1] Minor Planet Center - Isolated Tracklet File: https://www.minorplanetcenter.net/iau/ITF/itf.txt.gz.

https://doi.org/10.1016/j.icarus.2024.115949
Received 11 October 2023; Received in revised form 5 January 2024; Accepted 5 January 2024
Available online 22 January 2024

laws was introduced by Taff and Hall, who imposed the conservation of angular momentum and energy to solve the problem of linking tracklets and computing preliminary orbits (Taff and Hall, 1977; Taff, 1984), but did not fully exploit the algebraic character of the resulting equations, even if it was observed that they could be expressed in polynomial form. In these references, the high sensitivity of the results to astrometric error was noted. Later, Gronchi et al. (2010, 2011) and Gronchi et al. (2015) derived polynomial equations of degree 48, 20, and 9, respectively, from the Keplerian conservation laws for the purpose of linking two very short arcs. Then, Gronchi et al. (2017) demonstrated that the polynomial of degree 9 introduced in Gronchi et al. (2015) is optimal in some sense and derived an equation of degree 8 for the linkage of three very short arcs. Rodríguez et al. (2022) examined the numerical behavior of two Keplerian integrals algorithms introduced in Gronchi et al. (2015) and Gronchi et al. (2017), referred to as `link2` and `link3`, respectively. Although these methods are sensitive to astrometric error, their analysis showed that solutions with moderate error are promising. In addition to their ability to link tracklets that are widely spaced in time, `link2` and `link3` have the advantage of being computationally efficient due to their polynomial formulation.

In this study, we propose a procedure for computing least squares (LS) orbits using ITF detections submitted by Pan-STARRS1 (hereafter denoted by its observatory code F51) which is known for its small astrometric errors (Carpino et al., 2003). The procedure first links pairs of tracklets using the `link2` algorithm, then constructs '3-cycles' composed of 3 tracklets that have been successfully linked in the previous step. For each 3-cycle, a 'norm' is calculated with all the orbits obtained by `link2` using the 3 pairs of tracklets within the 3-cycle, and only 3-cycles with a norm below a certain threshold are retained. For each accepted 3-cycle, we compute a least squares orbit along with its root mean square (rms) astrometric error. Finally, we construct 'n-ids', sequences of $n \geq 4$ tracklets that were successfully linked by `link2`, by identifying additional candidate tracklets to the 3-cycles and applying differential corrections.

This article is structured as follows. In Section 2 we present the proposed procedure for linking tracklets, including the relevant indicators to assess the quality of the results. In Section 3 the values of the indicators are tuned to optimize the performance of the algorithm by applying it to a test data set constructed from real observations of main belt asteroids and some NEAs. Finally, in Section 4, the procedure is applied to all the F51 tracklets contained in the ITF with at least 3 observations each.

2. The procedure

Let us consider a list T of N tracklets, each composed of at least 3 observations, with the goal of identifying all the tracklets that belong to the same objects and determining their orbits. Our procedure follows three major steps described in the next sub-sections.

2.1. First step: `link2` exploration

The first step is to attempt to link all possible pairs of tracklets in the list $T = \{t_1, \dots, t_N\}$. Since this step has a quadratic cost, i.e., $\mathcal{O}(N^2)$, a highly efficient method is necessary if the number of tracklets N is large, as in the case of the ITF. Additionally, the linking method must be able to join tracklets even if they are separated by several years.

The Keplerian integrals algorithm `link2`, introduced in Gronchi et al. (2015) and tested in Rodríguez et al. (2022), is well-suited for this purpose. It is based on solving a univariate polynomial equation of degree 9 in the radial distance of the observed object at the mean epoch of one of the two tracklets being linked. The use of a polynomial with a relatively low degree makes this method fast compared to others. The `link2` algorithm is not symmetric, i.e., it is sensitive to inverting the order of the two tracklets, so both options must be considered to ensure that all possible linkages are computed.

The method may result in more than one solution since it depends on the computation of the roots of a polynomial of degree 9, but multiple solutions also occur with classical orbit determination methods, e.g., Gauss' method can have up to 3 solutions (Gronchi, 2009). All the solutions should be considered in the absence of other information. While a detailed analysis of the occurrence of multiple solutions of the `link2` algorithm has not been performed, we have established that the likelihood of multiple solutions decreases with the number of solutions (see Figures 3 and 22 in Rodríguez et al., 2022).

In the following, we refer to a *linkage* between two tracklets as a successful join using the `link2` algorithm, without any quality control. However, when using a Keplerian integrals method in practice, it is possible to obtain a preliminary orbit even if the tracklets do not belong to the same object, and to obtain a poor quality preliminary orbit even if they do. It is desirable to maximize the number of *true* linkages (where the tracklets belong to the same object), while minimizing the number of *false* ones.

The χ_4 and *rms* metrics (see Appendix A) are useful to quantify the quality of the solutions and select the best ones for the next step (Rodríguez et al., 2022). Thresholds for χ_4 and *rms* were determined based on testing with real observations of known objects (Section 3). `link2` linkages that satisfy the threshold values of both metrics are referred to as *accepted*. We denote the kth accepted preliminary orbit obtained with tracklets t_i, $t_j \in T$ by $o_{ij}^{(k)}$ (see Section 2.2.2).

Even after filtering solutions with the χ_4 and *rms* metrics it is possible to get false accepted linkages, i.e., accepted linkages between tracklets that do not belong to the same object. Additionally, even in the case of true linkages, the computed orbits are often not sufficiently accurate. To address these issues, the next two steps of the procedure are applied to join together more than two tracklets.

2.2. Second step: Constructing LS orbits using 3 tracklets

The next step is to group sets of three tracklets $\{t_i, t_j, t_k\}$ using the information obtained from `link2` such that each pair within the set is an accepted linkage.

2.2.1. Constructing 3-cycles

The results of the first step can be represented as a graph $G = G(V, E)$, where the set of vertices $V = \{1, 2, \dots, N\}$ corresponds to the set of tracklets $T = \{t_1, \dots, t_N\}$, and the set of edges E corresponds to the accepted linkages. Specifically, $e_{ij} \in E$, with $i, j \in V$, $i > j$, if and only if a linkage between tracklets t_i and t_j is found with at least one ordering of the tracklets and with acceptable values of χ_4 and *rms*. Therefore, each edge in E represents an accepted linkage. Even if there is more than one solution for a pair of tracklets, t_i and t_j, we consider only one edge for the linkage, i.e., we consider a *simple* graph. Moreover, we treat the edges as having a direction, from i to j, so it is possible to interpret the graph as a directed graph. The ordering is introduced only to simplify the computations.

As previously mentioned, our goal is to search for sets of three tracklets $\{t_i, t_j, t_k\}$ such that each pair of tracklets is an accepted linkage. This is equivalent to searching for sets of vertices $i, j, k \in V$ such that $e_{ij}, e_{ik}, e_{jk} \in E$ or, in other words, searching for 3-cycles in the graph G.

To find all the 3-cycles, for each vertex we select all the adjacent vertices in descending order $i = N, N-1, \dots, 2$, i.e., the set

$$N_G(i) = \{l \in V \mid e_{il} \in E,\ l < i\}.$$

In addition, the elements of the set $N_G(i)$ are considered in descending order, that is $N_G(i) = \{l_1^{(i)}, l_2^{(i)}, \dots, l_{p_i}^{(i)}\}$, with $l_1^{(i)} > l_2^{(i)} > \cdots > l_{p_i}^{(i)}$ (see Fig. 1). Using an adjacency list can save a significant amount of space compared to other graph representations, such as an adjacency matrix. The reduction in the space required to store the graph is especially important when dealing with sparse graphs, as is our case. It is also easy to insert or delete elements in the linked list.

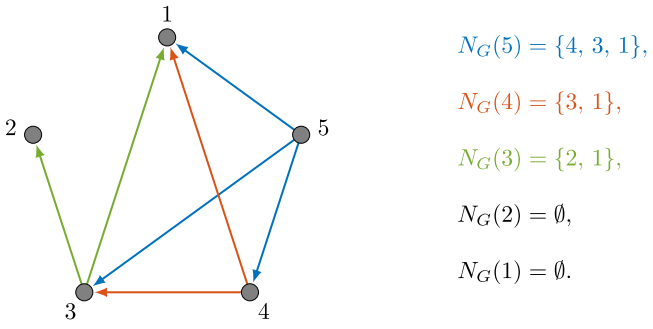

Fig. 1. Example of graph and sets $N_G(i)$.

Finally, this representation is useful to find all the 3-cycles by the following procedure: for each $i = N, N-1, \dots, 3$, we find the set of neighbors $N_G(i)$ and for each $j \in N_G(i) \setminus \{l_{p_i}^{(i)}\} = \{l_1^{(i)}, \dots, l_{p_i-1}^{(i)}\}$, we consider the set of neighbors $N_G(j)$. Then, we search for indices $k \neq i, j$ such that $k \in N_G(i) \cap N_G(j)$. For each tracklet k that satisfies this condition we obtain the 3-cycle $\{i, j, k\}$ (see Fig. 2 where we display the 3-cycles of Fig. 1). This classical procedure for searching 3-cycles is detailed in Algorithm 1.

Algorithm 1 Finding 3-cycles

```
V = {1, ..., N}
for i = N, N − 1, ..., 3 do
    for j ∈ N_G(i) \ {l_{p_i}^{(i)}} do
        Save the sets {i, j, k} with k ∈ N_G(i) ∩ N_G(j).
    end for
end for
```

We make the following remarks.

Remark 1. In order to find the 3-cycles in an efficient way it is important to sort the set of vertices. This is useful to avoid searching for equivalent 3-cycles, e.g., $\{i, j, k\}$ and $\{j, i, k\}$.

Remark 2. Selecting the tracklets in descending order allows us to avoid the exploration of the entire graph if we add new tracklets to the data set. In particular, let $t_{N+1}, \dots, t_{N+M}$ be the M new tracklets added to T. The addition of these new tracklets corresponds to the inclusion in G of the vertices $N+1, \dots, N+M$ and their corresponding edges e_{ij}, with $i \in \{N+1, \dots, N+M\}$ and $j \in \{2, \dots, N+M-1\}$. To find the new 3-cycles it is only necessary to perform the first loop of Algorithm 1 for $i = N+M, \dots, N+1$.

Remark 3. Algorithm 1 is easily parallelizable by distributing the values of i among the different nodes.

2.2.2. Angular momentum norm and LS orbits

For correct linkages Rodríguez et al. (2022) showed that the angular momentum of orbits, c, computed using `link2` is accurate. As a result, we employ the angular momentum as a measure of the quality of the 3-cycles. We recall that for each pair of tracklets `link2` is applied with both orderings resulting in possible multiple solutions. These solutions possess distinct values of the χ_4 and *rms* metrics that are taken into account when evaluating the quality of the 3-cycles. To quantify the quality of the 3-cycles we define the *angular momentum norm* as follows:

$$\|\{t_i, t_j, t_k\}\|_M = \min_{h,\ell,p} \left\{ m(o_{ij}^{(h)}, o_{ik}^{(\ell)}) + m(o_{ij}^{(h)}, o_{jk}^{(p)}) + m(o_{ik}^{(\ell)}, o_{jk}^{(p)}) \right\}, \quad (1)$$

with

$$m(o_A, o_B) = \frac{|c_A - c_B|}{\sqrt{|c_A||c_B|}} \left(\chi_{4,A} + \chi_{4,B}\right) \left(rms_A + rms_B\right),$$

where the subscripts A, B refer to the orbits o_A, o_B. This norm simply measures the difference between the angular momenta of the preliminary solutions using the indicators as weights.

After computing norm (1) for all the 3-cycles, we sort them by its values in ascending order, and accept only the 3-cycles with the norm below a threshold that will be determined later.

2.2.3. LS orbits

Finally, for each 3-cycle we construct an orbit by means of the least squares (LS) method starting with the preliminary orbits from `link2` and `link3`, because sometimes `link3` provides a better initial orbit for the differential corrections than `link2`.

Not all the 3-cycles will yield a LS orbit because the differential corrections algorithm may not converge. Furthermore, a solution will not be retained if the rms of the residuals of the resulting LS orbit is not sufficiently small.

We denote by C the set of triplets of tracklets $\{t_{i_1}, t_{i_2}, t_{i_3}\}$ with an acceptable LS orbit o_i ordered by the angular momentum norm. It is important to note that the same tracklet can be present in multiple 3-cycles.

Algorithm 2 Joining 4 or more tracklets

```
T = {t_1, ..., t_N}
for i = 1, ..., m do
    S = {t_{i_1}, t_{i_2}, t_{i_3}}
    if S ⊆ T then
        for j = i + 1, ..., m do
            if ({t_{j_1}, t_{j_2}, t_{j_3}} ⊆ T ) and ({t_{j_1}, t_{j_2}, t_{j_3}} ⊄ S) and (o_i, o_j close enough) then
                S* = S ∪ {t_{j_1}, t_{j_2}, t_{j_3}}
                o ← difCor(o_i, S*)
                if successful difCor then
                    o_i = o
                    S = S*
                end if
            end if
        end for
        for t_j ∈ T \ S do
            if t_j is close enough to o_i then
                o ← difCor(o_i, S ∪ {t_j})
                if successful difCor then
                    o_i = o
                    S = S ∪ {t_j}
                end if
            end if
        end for
        if size(S) ≥ 4 then
            Save S and o_i.
            T = T \ {S}
        end if
    end if
end for
```

2.3. Third step: Joining 4 or more tracklets

The last step is to attribute at least one additional tracklet to the LS orbit. The general idea is that for each triplet in C we attempt to identify other triplets of C that have orbits close to that of the considered triplet, and then try to attribute the new tracklets to the original triplet.

After applying the second step above we obtained m triplets of tracklets with a LS orbit, i.e., we have $C = \{C_i = (\{t_{i_1}, t_{i_2}, t_{i_3}\}, o_i)$ with $i = 1, \dots, m\}$. T will denote here the set of tracklets that have not been assigned to a LS orbit with 4 or more tracklets. Before applying the third step, T coincides with the set of the N available tracklets, and

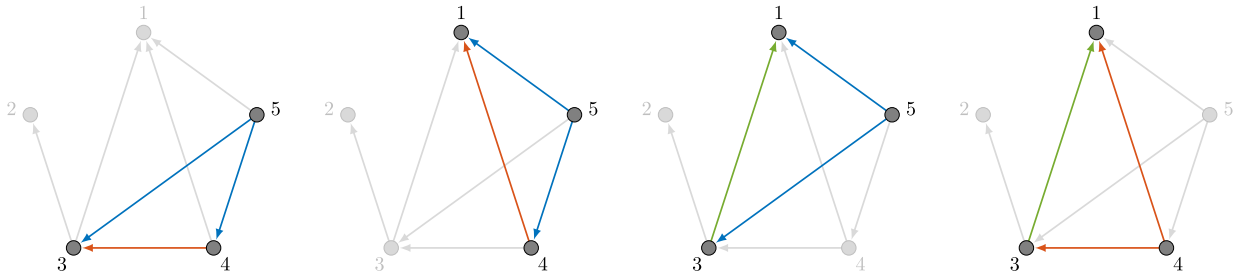

Fig. 2. The 3-cycles of the graph of Fig. 1.

as we obtain LS orbits with 4 or more tracklets, these will be removed from T.

We select the elements C_i of C following the order in which they appear in C and, if the three tracklets are in T, we consider the set of tracklets $S = \{t_{i_1}, t_{i_2}, t_{i_3}\}$ and the orbit o_i. Then, for each element $C_j \in C$ with $j = i+1, \ldots, m$ we check if all tracklets $t_{j_1}, t_{j_2}, t_{j_3} \in T$ and $\{t_{j_1}, t_{j_2}, t_{j_3}\} \nsubseteq S$. If both these conditions are satisfied, we check whether the orbit o_j is *close enough* to the orbit o_i. We say that the orbits o_A and o_B are close enough if their orbital elements satisfy

$$\begin{aligned} &\left|\frac{a_A - a_B}{a_A}\right| < \varepsilon_a, \quad |e_A - e_B| < \varepsilon_e, \quad |i_A - i_B| < \varepsilon_i, \\ &|\Omega_A - \Omega_B| < \varepsilon_\Omega, \quad |\omega_A - \omega_B| < \varepsilon_\omega, \end{aligned} \tag{2}$$

for some sufficiently small values of the ε thresholds, and if the tracklets belonging to the orbit o_B and not to o_A are close to the ones simulated from the orbit o_A.[2] If the two orbits are close enough we try to calculate a LS orbit for the detections contained in the tracklets in S and the new detections using the orbit o_i as the initial guess. If the differential corrections converge, the orbit o_i is updated with the LS orbit and the new tracklets are added to S. The differential corrections are successful if a solution is obtained using all the observations and the rms of the residuals of the LS orbit is below a certain threshold. This procedure is applied for all the remaining C_j taking into consideration that the orbit o_i and the set of tracklets S may have been updated if differential corrections were successful (see Algorithm 2 for more details).

At this point, the set S may have 3 or more tracklets and the next step consists in trying to attribute the tracklets remaining in T to the LS orbits that have already been computed. It is not feasible to apply the differential corrections algorithm for each possible attribution since it is computationally expensive because it involves an iterative process including n-body propagation (see Milani and Gronchi, 2010, Chap. 5). To minimize the number of candidates for a given LS orbit we employ a criterion based on the attributables. From the orbit we can compute a propagated attributable $(\alpha_p, \delta_p, \dot{\alpha}_p, \dot{\delta}_p)$ at the epoch of the attributable $(\alpha, \delta, \dot{\alpha}, \dot{\delta})$ associated with a tracklet. We check whether the difference between the two attributables is sufficiently small by requiring

$$\begin{aligned} &\left(\cos\alpha\cos\delta - \cos\alpha_p\cos\delta_p\right)^2 + \left(\sin\alpha\cos\delta - \sin\alpha_p\cos\delta_p\right)^2 \\ &\quad + \left(\sin\delta - \sin\delta_p\right)^2 < \varepsilon_1, \end{aligned} \tag{3a}$$

$$\frac{(\dot{\alpha} - \dot{\alpha}_p)^2 + (\dot{\delta} - \dot{\delta}_p)^2}{\sqrt{(\dot{\alpha}^2 + \dot{\delta}^2)(\dot{\alpha}_p^2 + \dot{\delta}_p^2)}} < \varepsilon_2, \tag{3b}$$

for some small values of $\varepsilon_1, \varepsilon_2$ and then proceed with the attribution. Eq. (3a) measures the Cartesian distance between the two points on the celestial sphere, while Eq. (3b) is related to the difference in their velocities.

[2] The latter condition is checked by comparing the values of the detections (α_i, δ_i) at epochs t_i of the tracklets related to o_B with the simulated detections obtained by propagation of the orbit o_A at the same epochs t_i.

If the differential corrections are successful we update o_i and add the new tracklet to S. The iteration continues with the tracklets that are yet to be explored but now accounting for the fact that S incorporates the new tracklet t (i.e., $S \leftarrow S \cup t$) and the propagated orbit is the one obtained with the new set S. It is worth noting that the tracklets in T can be ordered chronologically to reduce the computational cost of the propagation so that propagation times can be reduced by using the results of previous propagations. This procedure is consistent since each tracklet can be assigned to only one orbit.

Finally, if S contains 4 or more tracklets, we save the orbit o_i and the set S of tracklets that were used to compute it and remove the tracklets in S from T. The schematic idea of the procedure is described in Algorithm 2.

Algorithm 2 can be implemented as a sequence of two separate steps to facilitate its parallelization. In the first part, we search for the 3-cycles whose orbits are close enough and compute a least squares orbit. Moreover, we remove the tracklets from the leftover database T when they have been used to construct an orbit with 4 or more tracklets. In the second part, we try to attribute the tracklets in T to the orbits obtained in the previous step. It might happen that in the first part of the algorithm the same tracklet is employed in the computation of different orbits. In these cases we keep the orbit obtained from the 3-cycle with the smallest value of the M norm; the common tracklet is removed from the set of tracklets of the other orbit(s) and if at least 4 tracklets remain we try to compute a new least squares orbit, otherwise the tracklets are added to T. Note that at the end of the first part of the algorithm we may have orbits obtained from only 3 tracklets. These orbits will be discarded unless at least one additional tracklet is not attributed to them in the second part.

Finally, when parallelizing the second part of the procedure we may also obtain inconsistencies due to tracklets that have been attributed to more than one orbit. However, we can easily eliminate these inconsistencies a posteriori.

3. Testing the procedure

To define the values of the thresholds of the norms described in Section 2 we apply the procedure to a set of real F51 observations of known asteroids.

3.1. The test dataset

To test our linking algorithm and determine the thresholds for the ITF processing we extracted a realistic set of ITF-like tracklets from actual F51 observations.

The test data is composed of real F51 observations of 1021 asteroids with ≥ 6 tracklets each, where each tracklet contains ≥ 4 detections acquired between 2010 and 2022 inclusive. The minimum reported detection magnitude was $m_{min} = 21$ (we use 'm' to indicate a generic filter magnitude for PS1 which typically uses the r_{P1} or i_{P1} filters depending on the phase of the moon, Schlafly et al., 2012). We then randomly selected 6 tracklets from the set of tracklets for each object and randomly selected 4 detections from tracklets with > 4 detections.

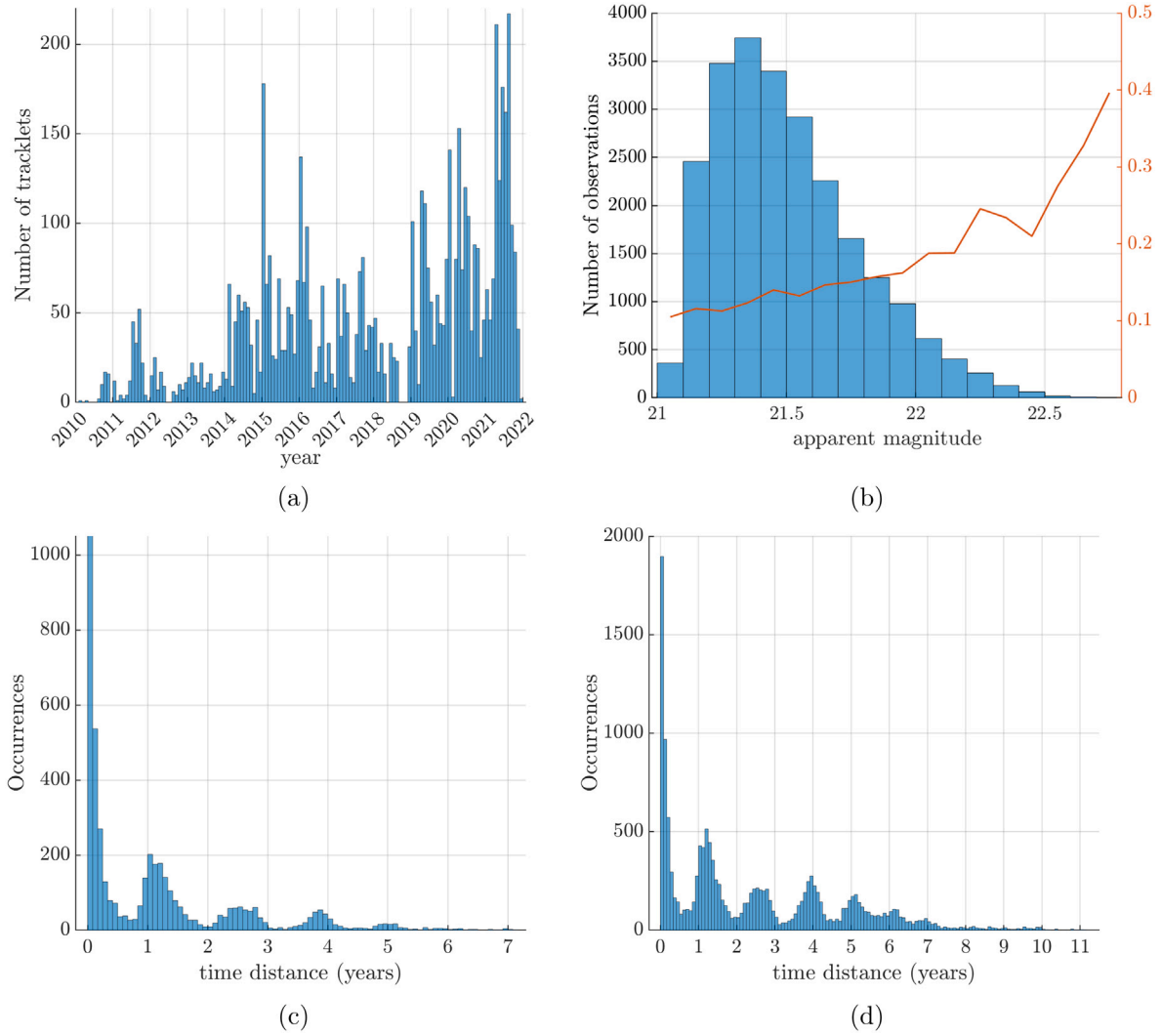

Fig. 3. (a) Time of observation of all 6 tracklets for the test data sample, (b) their reported apparent magnitude distribution and (in red) average astrometric error, (c) time between the closest two pairs of tracklets for each object, and (d) time between all pairs of tracklets for each object. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The $m \geq 21$ requirement was imposed on each detection in an attempt to match the apparent magnitude distribution of our test data to the apparent magnitude distribution of F51 observations in the ITF. This is important because the astrometric uncertainty depends on the apparent magnitude of the detections and has an impact on the linking and orbit determination efficiencies. The algorithm had no difficulty extracting the required tracklets for 1000 main belt objects but there were only 18 NEOs that met the requirements.

The time distribution of the randomly selected tracklets (Fig. 3(a)) shows that the number of tracklets increases with time. This is due to a major shift in the F51 system's survey strategy about five years after operations began and also an improvement in the system's capabilities with time. This distribution should mimic F51's contribution rate to the ITF under the assumption that the fraction of detected tracklets that are 'isolated' is relatively constant.

The brightest detections in our test data have $m \sim 21$ by design, while the faintest objects have reported apparent magnitudes $m > 22.5$ (Fig. 3(b)). The mode of $m = 21.3$ and median of $m = 21.4$ of the test data detections are about a half magnitude brighter than the mode and median of the real F51 ITF detections of $m = 21.9$ and 21.8, respectively. We were able to calculate the astrometric error for each detection (Fig. 3(b)) because these objects are main belt asteroids with precise and accurate orbital elements. The mean astrometric error is less than 0.17" for $m \lesssim 22$ and increases quickly to fainter magnitudes (Fig. 3(b)), consistent with Vereš et al. (2017).

The time between the nearest pairs of tracklets for the same object has a strong peak at $\ll 1$ year because objects are most likely to be re-detected in the same lunation when they are bright or in a successive lunation (Fig. 3(c)). Surveys typically re-image the same area of sky even within a lunation and their field-of-regard is now so large that the same objects can appear in the data from lunation to lunation. Furthermore, most objects are brightest and most detectable at perihelion and less likely to be detected at their next few apparitions. The successive peaks at multiples of about 1.3 years are simply because the synodic period of main belt objects is ~ 1.3 years (for an object with semi-major axis of 2.5 au). The time difference between all pairs of tracklets for the same object also exhibits the 1.3 year synodic periodicity but the peak at $\ll 1$ year is reduced because the panel no longer selects the minimum time between pairs of tracklets (Fig. 3(d)).

3.2. *link2 exploration*

Recalling that a true linkage includes two tracklets belonging to the same object, now we define an *accurate* linkage as a true linkage yielding an orbit close to the correct/known one.

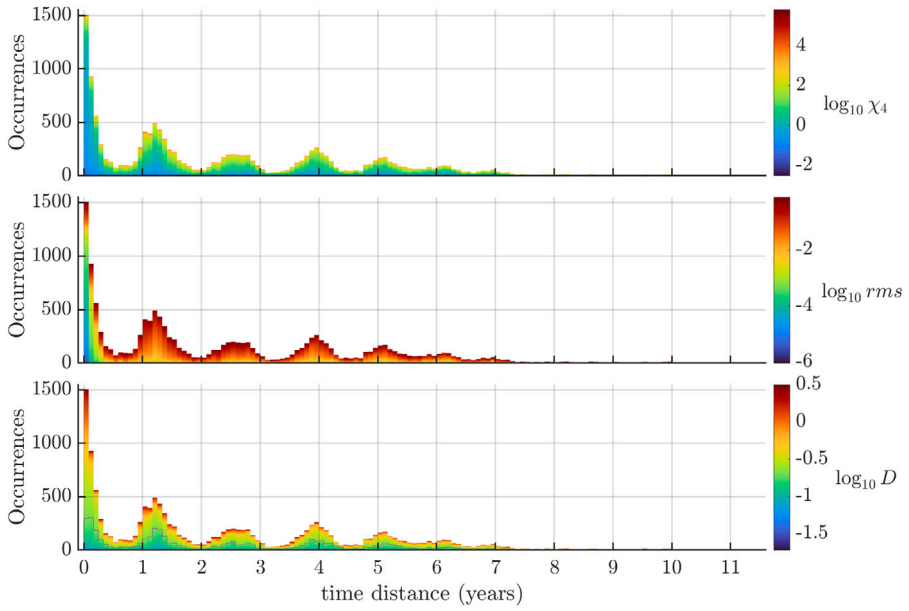

Fig. 4. Difference in time between any pair of tracklets belonging to the same object. The colors represent the values of $\log_{10}\chi_4$ (top), $\log_{10} rms$ (middle), and $\log_{10} D$ (bottom) of the `link2` solution with the best value of χ_4.

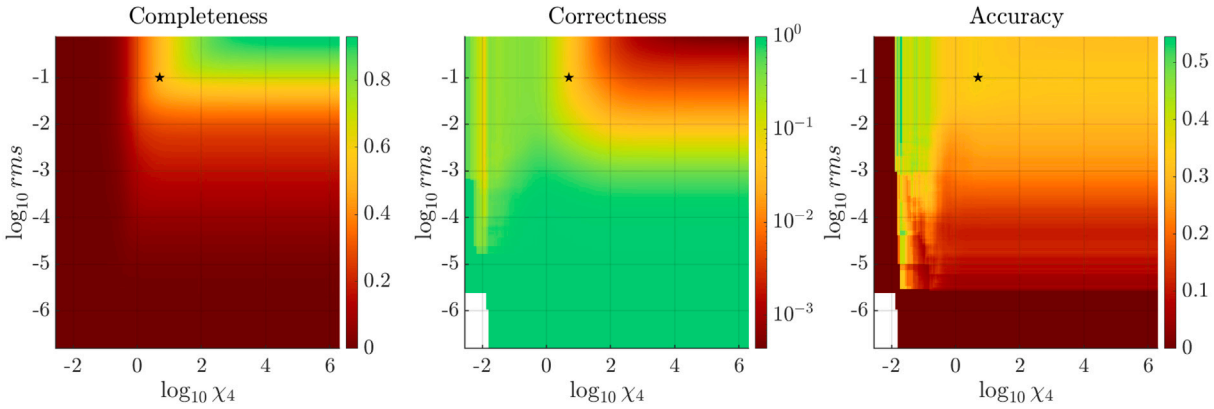

Fig. 5. Our algorithm's completeness (left), correctness (middle) and accuracy (right) as functions of the $\log_{10}\chi_4$, $\log_{10} rms$ thresholds for $D = 0.2$. The black star denotes the final implemented values. The white region in the middle and right panels corresponds to the cases where no linkages were found.

To quantify the proximity of two orbits we apply the D-criterion (Southworth and Hawkins, 1963; Drummond, 2000) which measures their distance, D, in the space of the orbital elements $(a, e, i, \Omega, \omega)$ (see Appendix A.2). We assume that the two sets of orbital elements are close enough to consider the linkage and orbit accurate if $D < 0.2$, a commonly used but somewhat arbitrary value in the literature. In a case where there are multiple solutions we use the preliminary solution with the smallest value of the χ_4 norm since, as discussed in Rodríguez et al. (2022), the values of D and χ_4 are correlated.

Our results (Fig. 4) obtained by applying `link2` to the dataset described in the previous section are worse than those reported in Rodríguez et al. (2022). Comparing Figs. 3(d) and 4 it is clear that we miss a considerable number of linkages. Moreover, the values of the χ_4 and rms indicators are higher than those in Rodríguez et al. (2022) due to the fact that here we only consider observations with an apparent magnitude ≥ 21 which have larger astrometric errors than brighter detections. In addition, a large number of solutions are lost when the time span between the mean epochs of the tracklets is short (see the leftmost bins in Figs. 3(d) and 4, with widths of 30 days). Nevertheless, the quality of the preliminary solutions obtained with `link2` remains good (Fig. 4).

We quantify the `link2` method's performance for observations of known objects with respect to the threshold values of χ_4 and rms with the following metrics:

$$\text{completeness} = \frac{\#\{\text{true linkages found}\}}{\#\{\text{total true linkages}\}},$$

$$\text{correctness} = \frac{\#\{\text{true linkages found}\}}{\#\{\text{all linkages found}\}},$$

$$\text{accuracy} = \frac{\#\{\text{accurate linkages found}\}}{\#\{\text{true linkages found}\}}.$$

The completeness, correctness and accuracy (Fig. 5) are consistent with expectations and with those presented in Rodríguez et al. (2022), where the numerical behavior of the `link2` and `link3` methods was analyzed with respect to the astrometric error of the observations, after accounting for the increased astrometric error in our current set

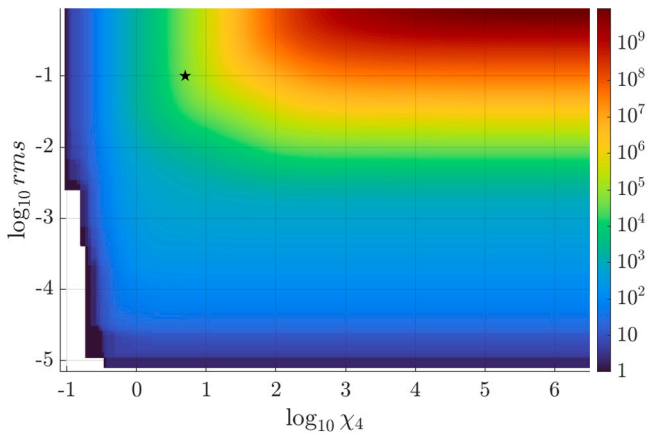

Fig. 6. Number of 3-cycles as a function of the maximum allowed $\log_{10} \chi_4$ and $\log_{10} rms$. The black star denotes the final implemented values.

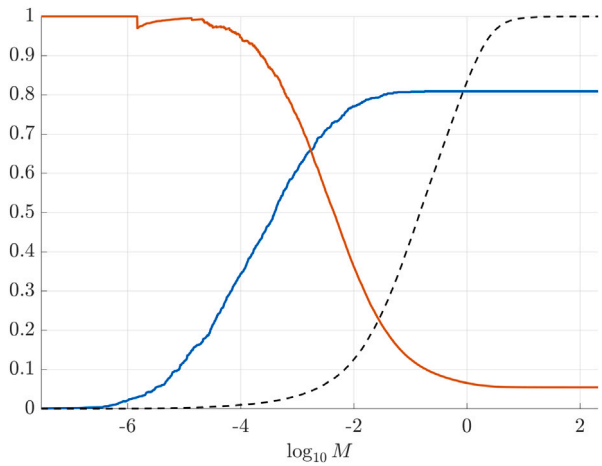

Fig. 7. Fraction of objects with at least one true 3-cycle (blue), fraction of true 3-cycles (red), and fraction of accepted 3-cycles (black/dashed) for $\chi_4 = 5$ and $rms = 0.1$ as a function of the angular momentum norm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of observations. A higher threshold for χ_4 and rms recovers a larger fraction of possible true linkages at the expense of increasing the number of false solutions.

3.3. Computing LS orbits from 3 tracklets

The next step of the procedure consists in constructing sets of 3 tracklets and computing an LS orbit for each set.

3.3.1. Constructing 3-cycles

The identification of 3-cycles was performed using Algorithm 1. Even working with a dataset containing only 6108 tracklets the total number of 3-cycles would be almost 10 billions if solutions were not discarded by means of the thresholds on χ_4 and rms (Fig. 6).

Since our ultimate goal was to apply this procedure to the ITF, we adopted tight thresholds for χ_4 and rms (see the next section) such that the total number of solutions was manageable.

3.3.2. Angular momentum norm

The angular momentum norm (1), hereafter denoted by M, was used to discard *false* 3-cycles without losing too many *true* 3-cycles, and the choice of the threshold value for M depends on the thresholds for χ_4 and rms (Fig. 7).

We set the χ_4 and rms thresholds to 5 and 0.1, respectively (Fig. 7), and the threshold value $\log_{10} M = -1.5$. The values were chosen empirically to produce manageable results with good efficiency and were operationally imposed at the beginning of the `link2` exploration (Section 3.2). With these values the procedure identifies at least one true 3-cycle for more than 80% of the asteroids and produces less than 20,000 3-cycles. Finally, the 3-cycles are ordered according to the value of M.

3.3.3. LS orbits

For each accepted 3-cycle in the previous step we try to construct an LS orbit from a preliminary orbit computed by either `link2` or `link3` (see Section 2.2.3) using all the observations within the 3-cycle's. Most (> 92%) of the false 3-cycles do not converge to an LS orbit but the majority (~ 92%) of the true 3-cycles do converge.

The quality of the LS orbits is assessed by the rms of the residuals, R_{LS} (see Appendix A.4), which is used to discard most of the false 3-cycles (Fig. 8). The maximums of the PDFs for the true and false LS orbits are well-separated but the tail of the false LS distribution overlaps almost completely with the R_{LS} for the true LS orbits. We selected a threshold value of $R_{LS} \leq 0.5$ to accept $\approx$ 99% of true LS orbits at the cost of also accepting $\approx$ 19% of false LS orbits. These false orbits

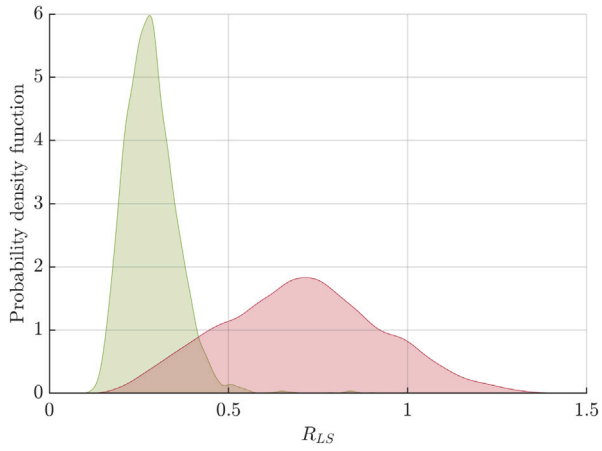

Fig. 8. PDF of the astrometric residuals, R_{LS}, for the false (red) and true (green) LS orbits computed from the 3-cycles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

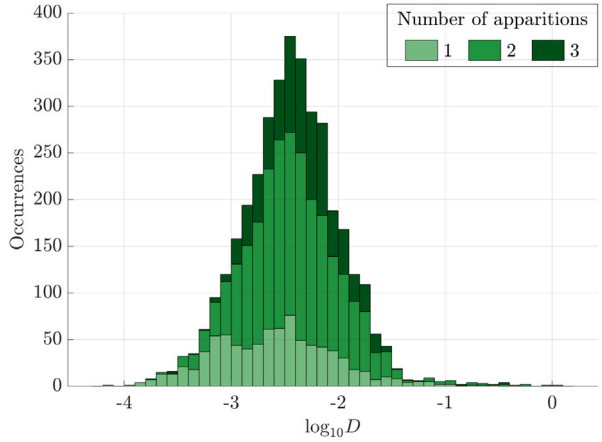

Fig. 9. Distribution of the values of D for the accepted LS orbits computed from the 3-cycles. The green color gradient indicates the number of apparitions included in the LS orbit. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

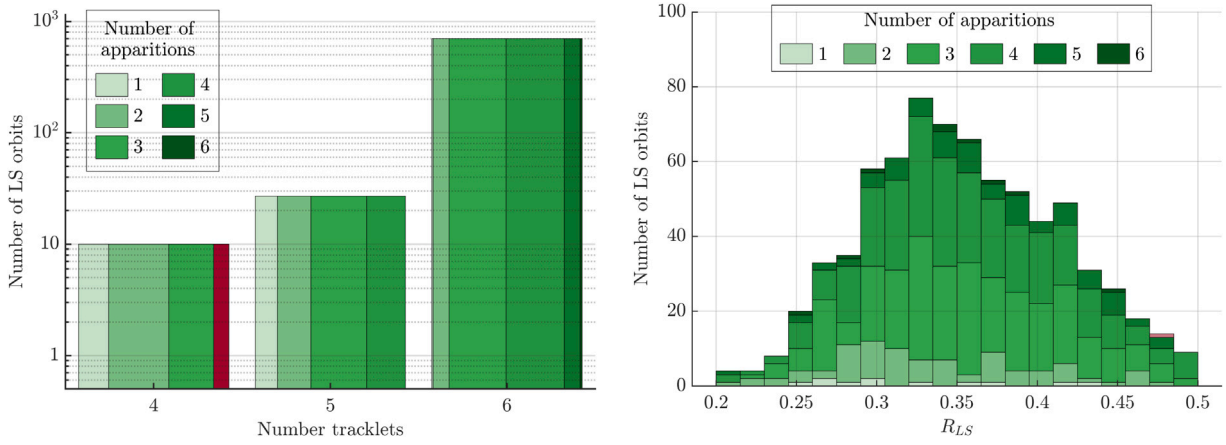

Fig. 10. (left) Number of accepted LS orbits computed with 4, 5, and 6 tracklets. The width of each sub-column indicates the fraction of orbits that are correct (green) or incorrect (red) solutions, while the green shading provides the number of apparitions spanned by the set of tracklets in each LS orbit. (right) Values of R_{LS} for the accepted orbits where green entries represent correct LS orbits and the single red entry indicates the single false orbit. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

are mostly eliminated in the next step (Section 3.4) by searching for additional isolated tracklets that are consistent with each orbit.

After selecting LS orbits according to their R_{LS} value, we find at least one LS orbit for ∼ 78% of the asteroids and the quality of these orbits is good as demonstrated with the D-criterion of our LS orbit compared to the known, high-accuracy orbit (Fig. 9). We find that ≈ 99.6% of the orbits have $D < 0.2$, the value we used above to determine if two sets of orbital elements were similar, and there is no obvious correlation between the quality of the orbit and the number of apparitions included in the set of observations.

3.4. Joining 4 or more tracklets

The final step is to apply Algorithm 2 to identify more tracklets and use all the detections contained in the tracklets to calculate an LS orbit. We set $\varepsilon_a = \varepsilon_e = \varepsilon_i = \varepsilon_\Omega = 0.02$, $\varepsilon_\omega = 0.05$, $\varepsilon_1 = 0.005$, $\varepsilon_2 = 0.05$ in Eq. (2) and an LS orbit is accepted if $R_{LS} < 0.5$.

The algorithm yielded 735 accepted LS orbits with 4 or more tracklets of the test sample (Fig. 10, left) and only one of them was false. The single incorrect orbit includes 3 tracklets corresponding to one asteroid in three consecutive nights in Aug 2021 and 1 tracklet of another object in Jan 2019. A total of 698 true orbits included all 6 possible tracklets in the test sample for each asteroid.

The average astrometric rms of the accepted orbits is ≈ 0.35" (Fig. 10, right), not much higher than the mean error of the detections in our test sample (Fig. 3b). The average D-criterion for the accepted orbits compared to the actual orbits is ∼ 0.0056, significantly better than the maximum value of 0.2 used as a threshold when setting our metric thresholds (Fig. 11).

The efficiency for recovering NEOs is 72^{+9}_{-11}% consistent with the 72.1 ± 0.1% MBA detection efficiency. The average D-criterion for the NEOs is 0.0026.

4. Application to the ITF

We applied the procedure outlined in the preceding sections to the Pan-STARRS observations in the ITF as of 2022 July 30, after the work of Holman et al. (2018) and Weryk et al. (2020), a dataset containing 3,760,777 F51 observations. Pan-STARRS data were used for this study because, (1) it provides the largest single-survey subset of tracklets in the ITF and a consistent operational mode, (2) we viewed this as a first foray into application of the procedure to the ITF, and (3) one of us was a Pan-STARRS team member and is intimately familiar with Pan-STARRS's capabilities and performance metrics.

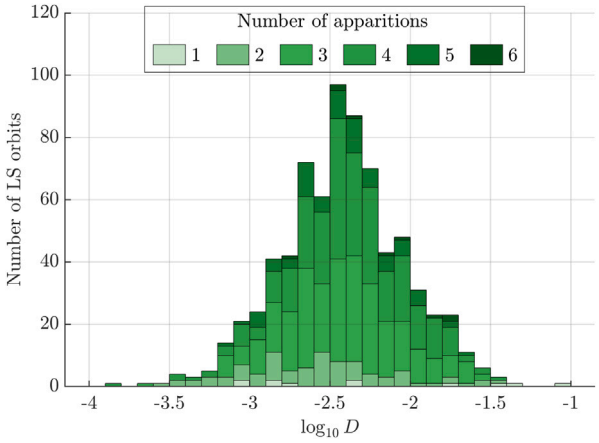

Fig. 11. Values of the D-criterion for the final accepted LS orbits. The green color gradient indicates the number of apparitions included in the orbits. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We first applied corrections for two types of inconsistencies in the data: (1) duplicate observations with the same RA and declination but at slightly different epochs and (2) tracklets spanning too long a time range.

There were only about half a dozen duplicate observations that had identical values of RA and declination at two or more times that differed by only a few seconds. Duplicate observations were combined into a single detection with the same RA and declination at the average time of observation.

Tracklets are generally a set of observations acquired over a short period of time within a single night so we split a tracklet into sub-tracklets if the time separation between two consecutive observations was > 0.5 days. In one extreme case, a single ITF tracklet contained observations spanning from 2014 June 21 to 2014 September 12. We restricted our tracklets to those containing ≥ 3 detections because the false tracklet rate is high for tracklets containing only 2 detections (i.e., a third confirmatory detection dramatically reduces the false tracklet rate) and most surveys to date only report tracklets containing ≥ 3 detections for unknown objects. After applying all our cuts on the dataset we were left with 1,072,171 tracklets containing 3,693,929 detections.

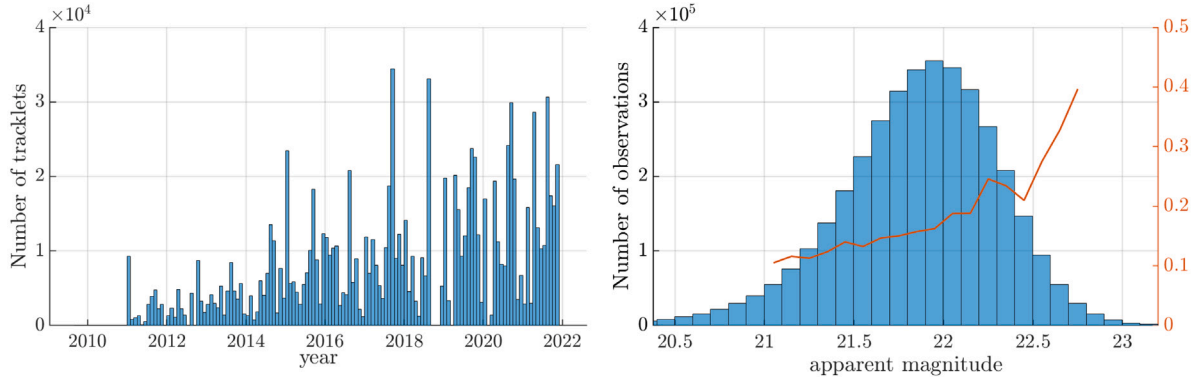

Fig. 12. (left) Times of observation of F51 tracklets in the ITF; (right) their reported apparent magnitude distribution and (in red) their average astrometric uncertainty as a function of the reported apparent magnitude as calculated from the 1000 object test sample. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

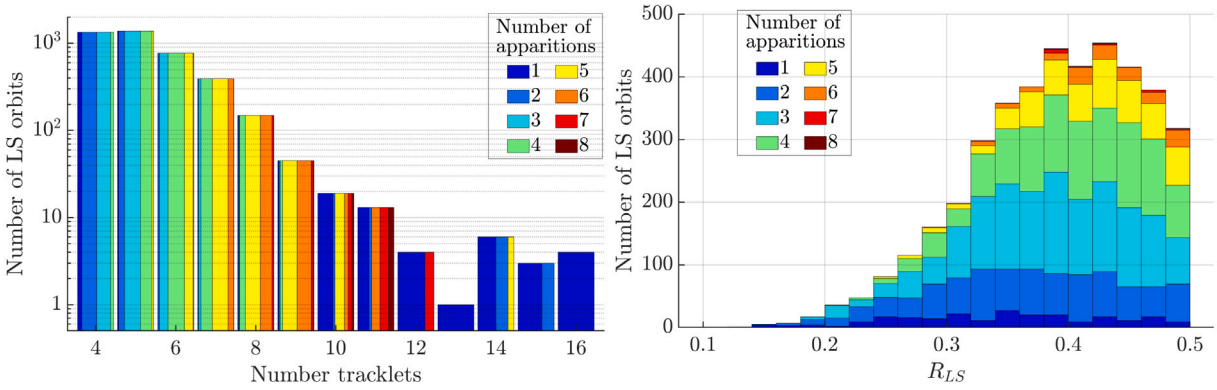

Fig. 13. (left) Number of tracklets per object in the final set of LS orbits. (right) Values of the residuals of the final set of LS orbits. The color indicates the number of apparitions included in each solution.

The distribution of the times of observations (Fig. 12, left) reflects the operations of the Pan-STARRS survey which began science operations in 2010 (Chambers et al., 2019) with an increasing fraction of time devoted to asteroid surveying as the years passed and gradual improvement in the system's asteroid detection efficiency.

The apparent magnitudes of the F51 ITF observations (Fig. 12) are typically greater than 21.7, the system's limiting magnitude in their most sensitive wide-band filter, w_{P1}, that was used for most asteroid surveying (Denneau et al., 2013). With $\gtrsim$ 59% of the observations greater than the system's limiting magnitude the astrometric uncertainty on these observations is much worse than the system's average rms uncertainty of $\sim$ 0.13" on observations of brighter targets (Milani et al., 2012).

The procedure described in this paper was applied to the cleaned F51 ITF observations and identified 4133 LS orbits that included 4 or more tracklets in 4 different nights. The majority of the orbits contain six or fewer tracklets but four of the orbits contain 16 tracklets (Fig. 13). Furthermore, most of the solutions span more than one apparition even though they are composed of a small number of tracklets.

The maximum time separation between two sequential tracklets in a single orbital solution spans a wide range (Fig. 14, left) from less than one year to almost 10 years, with peaks corresponding to the synodic periods of main belt objects. The total time span of the observations linked to a single object exhibits similar peaks with a maximum greater than 10 years and a mean greater than 5 years (Fig. 14, right).

The LS orbits have much higher astrometric residuals than typical of F51 because they only include detections at much fainter magnitudes (Fig. 15), with the peak of the distribution at $\sim$ 0.45", almost 50% higher than the test dataset's residuals that were specifically designed to match the apparent magnitudes of F51's ITF detections (Fig. 10). We think that the high residuals are not due to the presence of false solutions because (1) we only found one false solution out of 728 in our test dataset and (2) the orbital distribution of the final set of orbits is a good match to the orbital distribution of known objects as we will show below (Fig. 17).

There is evidence to support the conclusion that the 4133 LS orbits correspond to real objects because their orbit distribution reproduces the distribution of objects in the main belt including Kirkwood Gaps, Jupiter Trojans, and both collisional and dynamical asteroid families within the main belt (Fig. 17). Almost all the orbits correspond to MBAs but 2 represent already known NEOs (384P/Kowalski and 2019 KW3) and the most distant object with a semi-major axis of $\sim$ 7.7 au corresponds to a Centaur.

The absolute magnitudes of our linkages also provide evidence that they are legitimate (Fig. 15). They are strongly skewed to the faint end of the main belt values because bigger, brighter objects with smaller H are more likely to have been detected often in tracklets that have already been associated with known objects, while smaller, fainter objects with larger H are too faint to be detected regularly (i.e., the smallest objects might be detected in a serendipitous apparition when they are at perihelion near opposition but are unlikely to be detected in subsequent apparitions and therefore less likely to be identified with our algorithm).

More than 99% of the asteroids we identified in the inner belt ($a < 2.5$ au) are sub-km diameter asteroids with $H > 17.6$ assuming an S-class albedo of 0.17 typical of objects in the main belt (DeMeo et al., 2015; Wright et al., 2016). In 2009 it was suggested (Gladman et al.,

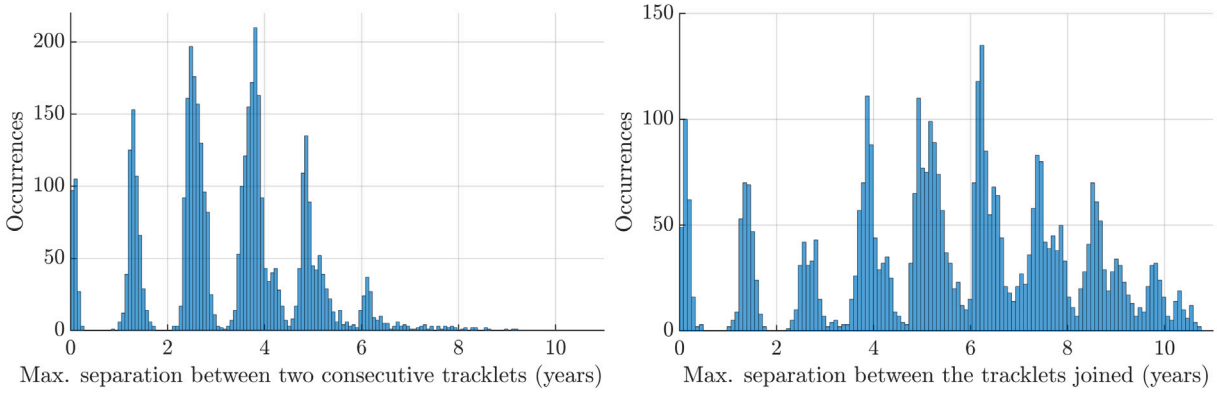

Fig. 14. (left) Maximum time between any two sequential tracklets within the set of tracklets belonging to a single orbit. (right) Total time span of the tracklets in the final set of orbits.

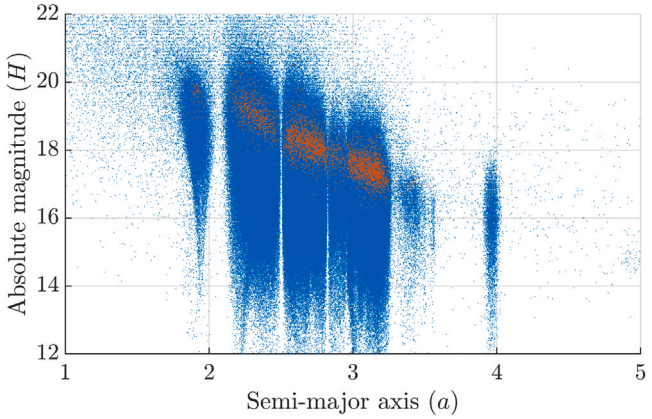

Fig. 15. Absolute magnitude versus semi-major axis of (blue) known objects from JPL Horizons and (orange) this work. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

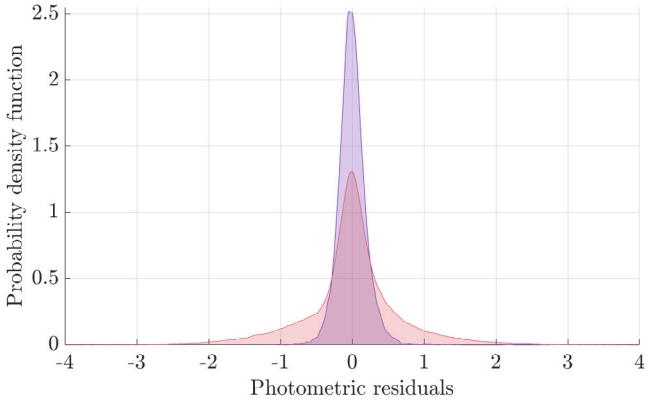

Fig. 16. Probability density function of the photometric residuals for (blue) the LS orbits obtained with the test data and (red) the LS orbits identified in the ITF data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2009) that the main belt population ($2.0\,\text{au} < a < 3.5\,\text{au}$) is completely known for $H < 15$ and only $\sim 0.02\%$ of the objects we identified fall into that absolute magnitude range, i.e., 2 objects, both in the outer region of the belt. About 25% of our main belt objects have $H < 17.5$, the completeness limit proposed in 2015 (Denneau et al., 2015), inhabiting the outer regions of the belt. Given that the outer belt is dominated by low albedo (typically ~ 0.03, see DeMeo et al., 2015; Wright et al., 2016) C-class asteroids, a 1 km diameter asteroid in the outer belt would have $H \sim 19.4$, suggesting that it will take some time till the main belt is effectively complete for km-scale asteroids.

The half-width at half-maximum (HWHM) of the distribution of photometric residuals for the main belt test data (Fig. 16) is about 0.2 mags implying an SNR ~ 5 for the detections in the test data. This is about what is expected for objects with the magnitude distribution having a mode of $V \sim 21.3$ (Fig. 3), almost half a magnitude brighter than the system limiting magnitude where each detection typically has SNR ~ 3. The HWHM of the distribution of the ITF objects with LS orbits is only 25% larger at ~ 0.25 mags but this comparison does not capture the different shapes of the two distributions. The photometric residuals for the ITF orbits have a much wider range of values extending out to ~ 2 mags due to the detections being typically fainter than the system limiting magnitude.

4.1. MPC processing

At the time of our first submission of identifications to the MPC they required that an ITF identification contained at least 1 tracklet observed on a minimum of 4 separate nights with an observational arc spanning at least 10 days if all the tracklets are in a single apparition. For identifications over multiple apparitions the MPC required that at least one of the apparitions contained at least 3 tracklets obtained on a minimum of 3 separate nights and the other apparitions had to contain at least 2 tracklets acquired on at least 2 nights. Only 112 out of our 4133 orbits satisfied this condition.

Later on, the MPC submission requirements for ITF identifications changed,[3] however some of our identifications did not fulfill the new

Table 1
Results of the MPC's processing of our linkages as a function of the number of apparitions included in the linkage: 'no solution' is the number of linkages for which the MPC was not able to calculate an orbit; 'mixed' is the number of linkages that include at least one incorrectly linked tracklet; 'accepted' is the number of accepted linkages in which all the tracklets are associated with the same object; '% accepted' is the percentage of accepted submissions including tracklets identified in the specified number of apparitions.

	Apparitions								
Category	1	2	3	4	5	6	7	8	Total
no solution	0	20	24	4	5	0	0	0	53
mixed	0	147	4	1	0	0	0	0	152
accepted	231	653	1314	1059	498	151	18	4	3928
% accepted	100.0	79.6	97.9	99.5	99.0	100.0	100.0	100.0	95.0

[3] see https://minorplanetcenter.net/mpcops/documentation/identifications/additional/.

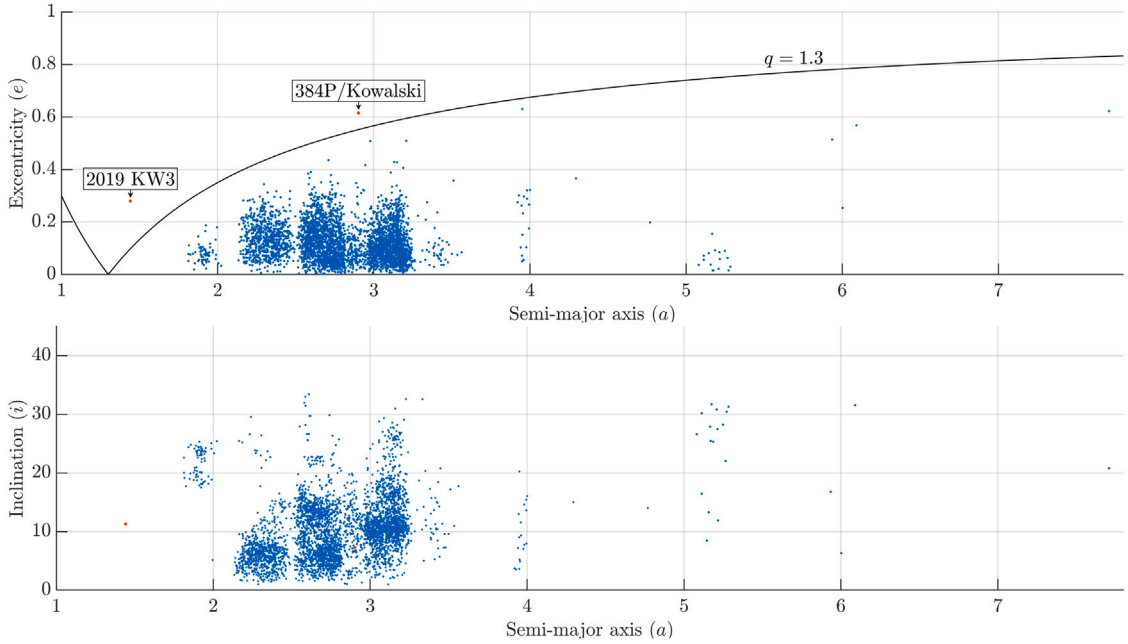

Fig. 17. Eccentricities (top) and inclinations (bottom, in degrees) versus semi-major axis of the accepted LS orbits. The red dots above the curve on the top panel correspond to NEOs.

Table 2
Number of our linkages for which the MPC was not able to fit an orbit as a function of the number of apparitions and number of tracklets included in the linkage. A blank cell means that the value is zero. We required that at least 4 tracklets are included in every linkage.

		# of tracklets in linkages with no MPC orbit solution											
		4	5	6	7	8	9	10	11	12	13	14	15
# of apparitions	1												
	2	15	3	1			1						
	3	22	1		1								
	4	2	2										
	5		3	2									

Table 3
Number of our linkages for which the MPC identified a mixture of tracklets corresponding to different objects as a function of the number of apparitions and number of tracklets included in the linkage. A blank cell means that the value is zero. We required that at least 4 tracklets are included in every linkage. A subsequent analysis of these results revealed that the three identifications denoted with an asterisk are correct. These linkages correspond to tracklets containing observations of 2018 NB63, 2017 UH213, and 2016 UQ171.

		# of tracklets with mixed linkages											
		4	5	6	7	8	9	10	11	12	13	14	15
# of apparitions	1												
	2	78	57	8	1	1		1					1
	3	2	1*		1*								
	4			1*									
	5												

acceptance criteria yet. The MPC agreed to process our entire set of identifications and accepted the majority of our linkages ($\gtrsim$ 95%) (Table 1). Nearly half of the orbits correspond to new identifications despite using observations spanning two or more years in the ITF (Fig. 14 left).

The MPC accepted all our identifications linking $\geq$ 4 tracklets in a single apparition, and $\gtrsim$ 98% of those linking tracklets in $\geq$ 3 apparitions (Table 1). On the other hand, the MPC accepted about 80% of the 2 apparition linkages, which is not surprising because most of these cases include only 4 or 5 tracklets.

There were cases for which the MPC could not fit an orbit even though we provided our LS orbit in the submission (Table 2) and cases where the MPC found a mixture of tracklets belonging to different objects (Table 3). The cases for which the MPC was not able to obtain a LS orbit (Table 2) include identifications with 4 and 5 apparitions and our post-submission analysis found that those cases have either 1 tracklet per apparition or 2 tracklets in 1 apparition and only 1 tracklet in the others. We also note that some values of the photometric residuals within these 53 linkages are high.

The MPC found that 152 of our identifications contained tracklets from different previously known objects (Table 3). In these cases, 147 of them link tracklets in only 2 apparitions, and in 142 cases all the tracklets except one belong to a single object in 1 apparition while the other belongs to a different object in a separate apparition. This includes the case of an orbit with 15 tracklets, 14 of which belong to an apparition of an object from 2019 while the incorrectly attributed tracklet is from 2012.

The majority of the non-accepted linkages (Tables 2 and 3) include identifications with only 4 or 5 tracklets. A subsequent analysis of these results revealed that three of those identifications (denoted in

Table 3 with an asterisk) are correct. Our linkages correspond to tracklets containing observations of 2018 NB63, 2017 UH213, and 2016 UQ171, and their identifications were published in the MPS 2072159, 2056275-2056276, and 2056085, respectively.

In general, we find a strong correlation between the acceptance of a linkage and both the number of apparitions and the number of tracklets included. We think that the fraction of non-accepted linkages could be sensibly reduced by implementing slightly different thresholds within our linking process and, perhaps, requiring a minimum number of tracklets in each apparition.

5. Conclusions

We presented a procedure to join tracklets in large datasets based on the Keplerian integrals method `link2` which allowed us to link tracklets that may be separated by years-long gaps in time. The quality of the accepted solutions is assessed by different norms to ensure that the final results are reliable. The procedure is fast enough that a complete exploration of a large dataset is computationally feasible and it was applied to F51 observations in the ITF yielding more than 4000 orbits, mostly MBAs, but also 2 NEOs.

The MPC's processing of our identifications found 2024 new orbits out of the 3928 accepted submissions (Table 1), the other orbits having been identified by other groups in the time after we downloaded the ITF for this study.

Future implementations of our algorithm should consider tuning the thresholds to account for the astrometric uncertainties of different surveys. We think that this procedure will be helpful with processing data from new surveys such as the LSST.

CRediT authorship contribution statement

Óscar Rodríguez: Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing, Resources. **Giovanni F. Gronchi:** Investigation, Supervision, Writing – original draft, Writing – review & editing, Software, Resources. **Giulio Baù:** Writing – review & editing, Writing – original draft. **Robert Jedicke:** Formal analysis, Investigation, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We wish to thank Peter Veres for processing our linkages at the Minor Planet Center and assessing their correctness. We also thank Steve Chesley and another anonymous reviewer for their useful comments that have improved this paper. This work was supported by the Spanish State Research Agency through the Severo Ochoa and María de Maeztu Program for Centers and Units of Excellence in R&D (CEX2020-001084-M) and through the H2020 MSCA ETN Stardust-Reloaded, Grant Agreement Number 813644. Ó. Rodríguez was supported by the Spanish grant PID2021-123968NB-I00 (AEI/FEDER/UE). G. F. Gronchi and G. Baù acknowledge the Italian project MIUR-PRIN 20178CJA2B "New frontiers of Celestial Mechanics: theory and applications" and the GNFM-INdAM, Italy.

Appendix A

From a very short arc of $m \geq 2$ optical observations of a celestial body $\{(\alpha_i, \delta_i) \mid i = 1, \dots, m\}$, where (α_i, δ_i) denote the right ascension and declination at epochs t_i, $i = 1, \dots, m$, respectively, we can calculate the *attributable* vector

$$\boldsymbol{\mathcal{A}} = (\alpha, \delta, \dot{\alpha}, \dot{\delta}),$$

which represents the angular position and angular rate of the body at the mean time $\bar{t} = \frac{1}{m}\sum_{i=1}^{m} t_i$. The `link2` method is capable of calculating a preliminary orbit using two attributables $\boldsymbol{\mathcal{A}}_1$, $\boldsymbol{\mathcal{A}}_2$ of the same object at different epochs.

A.1. The χ_4 norm

The `link2` algorithm calculates preliminary orbits along with their corresponding covariance matrices. Let us consider a preliminary orbit at epoch $\bar{t}_1$ of the first attributable $\boldsymbol{\mathcal{A}}_1$. This orbit, along with its covariance, is propagated to the epoch $\bar{t}_2$ of the second attributable $\boldsymbol{\mathcal{A}}_2$ and the associated covariance matrix Γ_2. Subsequently, the propagated attributable $\boldsymbol{\mathcal{A}}_p$ and its marginal covariance matrix Γ_p can be computed (Milani et al., 2001). We quantify the difference between the two attributables $\boldsymbol{\mathcal{A}}_p$ and $\boldsymbol{\mathcal{A}}_2$ using the χ_4 norm,

$$\chi_4 = \sqrt{(\boldsymbol{\mathcal{A}}_p - \boldsymbol{\mathcal{A}}_2) \cdot \left(C_2 - C_2\,\Gamma_0\,C_2\right)(\boldsymbol{\mathcal{A}}_p - \boldsymbol{\mathcal{A}}_2)},$$

where

$$C_2 = \Gamma_2^{-1}, \qquad \Gamma_0 = C_0^{-1},$$

with

$$C_0 = C_2 + C_p, \qquad C_p = \Gamma_p^{-1}.$$

A.2. The D-criterion

The D-criterion was introduced by Southworth and Hawkins (1963) to quantify the similarity between the orbits of two objects. Let us denote by q_k the perihelion distance, e_k the eccentricity, i_k the inclination, Ω_k the longitude of the node, and ω_k the argument of perihelion for the two orbits ($k = 1, 2$). These orbits may differ in four aspects: size, shape, inclination of the orbital plane, and perihelion direction. The value of D takes into account these four aspects and is defined as

$$D = \sqrt{d_1^2 + d_2^2 + d_3^2 + d_4^2},$$

where

$$d_1^2 = (q_1 - q_2)^2, \qquad d_2^2 = (e_1 - e_2)^2, \qquad d_3^2 = 4\sin^2(I/2),$$
$$d_4^2 = (e_1 + e_2)^2 \sin^2(\Pi/2),$$

with

$$I = \arccos\left(\cos i_1 \cos i_2 + \sin i_1 \sin i_2 \cos(\Omega_1 - \Omega_2)\right),$$

and

$$\Pi = \omega_1 - \omega_2 + 2\varepsilon \arcsin\left(\cos\left(\frac{i_1 + i_2}{2}\right)\sin\left(\frac{\Omega_1 - \Omega_2}{2}\right)\sec\left(\frac{I}{2}\right)\right),$$

where $\varepsilon = 1$ if $|\Omega_1 - \Omega_2| \leq 180°$ and $\varepsilon = -1$ otherwise.

A.3. The rms norm

Another metric employed to evaluate the quality of the preliminary orbits obtained with `link2` is the root mean square value of the angular separation between the observations and the predicted ephemeris:

$$rms = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\Delta_{\alpha_i}^2 \cos^2\delta_i + \Delta_{\delta_i}^2\right)},$$

where $\Delta_{\alpha_i} = \alpha_i - \alpha(\bar{t}_i)$, $\Delta_{\delta_i} = \delta_i - \delta(\bar{t}_i)$ are the residuals of the right ascension and declination of the observations, with $(\alpha(\bar{t}_i), \delta(\bar{t}_i))$ calculated with a 2-body propagation of the preliminary orbit.

A.4. Least squares norm (R_{LS})

After computing a preliminary orbit, a differential correction scheme (Milani and Gronchi, 2010, Chap. 5) can be employed to determine a least squares (LS) orbit.[4] If this scheme is successful, the quality of the LS orbit is assessed by the following norm. The definition of the *rms* norm is extended by introducing weights w_i on each set of observations (Farnocchia et al., 2015):

$$R_{LS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[w_i\left(\Delta_{\alpha_i}^2\cos^2\delta_i + \Delta_{\delta_i}^2\right)\right]},$$

where $\alpha(\bar{t}_i)$ and $\delta(\bar{t}_i)$ are obtained from the full n-body propagation of the LS orbit.

References

Carpino, M., et al., 2003. Error statistics of asteroid optical astrometric observations. Icarus 166 (2), 248–270. http://dx.doi.org/10.1016/S0019-1035(03)00051-4.

Chambers, K.C., et al., 2019. The Pan-STARRS1 surveys. arXiv:1612.05560.

Christensen, E.J., et al., 2016. The catalina sky survey for near-Earth objects. In: AAS/Division for Planetary Sciences Meeting Abstracts #48. In: AAS/Division for Planetary Sciences Meeting Abstracts, vol. 48, p. 405.01.

DeMeo, F.E., et al., 2015. The compositional structure of the asteroid belt. In: Michel, P., DeMeo, F.E., Bottke, W.F. (Eds.), Asteroids IV. pp. 13–41. http://dx.doi.org/10.2458/azu_uapress_9780816532131-ch002.

Denneau, L., et al., 2013. The Pan-STARRS moving object processing system. Publ. Astron. Soc. Pac. 125, 357–395. http://dx.doi.org/10.1086/670337, arXiv:1302.7281.

Denneau, L., et al., 2015. Observational constraints on the catastrophic disruption rate of small main belt asteroids. Icarus 245, 1–15. http://dx.doi.org/10.1016/j.icarus.2014.08.044, arXiv:1408.6807.

Drummond, J.D., 2000. The d discriminant and near-Earth asteroid streams. Icarus 146, http://dx.doi.org/10.1006/icar.2000.6401.

Farnocchia, D., et al., 2015. Star catalog position and proper motion corrections in asteroid astrometry. Icarus (ISSN: 0019-1035) 245, 94–111. http://dx.doi.org/10.1016/j.icarus.2014.07.033.

Gauss, C.F., 1809. Theoria Motus Corporum in Sectionibus Conicis Solem Ambientium. Reprinted by Dover publications in 1963.

Gladman, B.J., et al., 2009. On the asteroid belt's orbital and size distribution. Icarus 202, 104–118. http://dx.doi.org/10.1016/j.icarus.2009.02.012.

Gronchi, G.F., 2009. Multiple solutions in preliminary orbit determination from three observations. Cel. Mech. Dyn. Ast. (ISSN: 0923-2958,1572-9478) 103 (4), 301–326. http://dx.doi.org/10.1007/s10569-009-9201-x.

Gronchi, G.F., et al., 2010. Orbit determination with the two-body integrals. Cel. Mech. Dyn. Ast. 107/3, 299–318.

Gronchi, G.F., et al., 2011. Orbit determination with the two-body integrals. II. Cel. Mech. Dyn. Ast. 110/3, 257–270.

Gronchi, G.F., et al., 2015. Orbit determination with the two-body integrals. III. Cel. Mech. Dyn. Ast. 123/2, 105–122.

Gronchi, G.F., et al., 2017. Keplerian integrals, elimination theory and identification of very short arcs in a large database of optical observations. Cel. Mech. Dyn. Ast. 127/2, 211–232.

Gronchi, G.F., et al., 2021. Generalization of a method by mossotti for initial orbit determination. Cel. Mech. Dyn. Ast. 133 (9), 41. http://dx.doi.org/10.1007/s10569-021-10038-4.

Holman, M.J., et al., 2018. HelioLinC: A novel approach to the minor planet linking problem. Astron. J. 156 (3), 135. http://dx.doi.org/10.3847/1538-3881/aad69a.

Ivezić, Ž., et al., 2019. LSST: From science drivers to reference design and anticipated data products. Astrophys. J. 873 (2), 111. http://dx.doi.org/10.3847/1538-4357/ab042c, arXiv:0805.2366.

Lagrange, J.L., 1783. Sur le problème de la détermination des orbites des comètes d'aprés trois observations. Troisième mémoire. Nouveaux Mem. Acad. R. Sci. B.-Lett. Berl. Reprinted in *Œuvres de Lagrange*, Gauthier-Villars et fils, Paris (1869), volume 4, pp. 496–532.

Laplace, P.S., 1780. Mémoire sur la détermination des orbites des comètes. Mémoires de l'Académie royale des sciences de Paris Reprinted in *Œuvres complètes de Laplace*, Gauthier-Villars et fils, Paris (1894), volume 10, pp. 93–146.

Milani, A., Gronchi, G.F., 2010. Theory of Orbit Determination. Cambridge Univ. Press.

Milani, A., et al., 2001. The asteroid identification problem IV: Attributions. Icarus 151 (2), 150–159. http://dx.doi.org/10.1006/icar.2001.6594.

Milani, A., et al., 2012. Identification of known objects in solar system surveys. Icarus 220 (1), 114–123. http://dx.doi.org/10.1016/j.icarus.2012.03.026, arXiv:1201.2587.

Rodríguez, Ó, et al., 2022. Numerical behaviour of the Keplerian integrals methods for initial orbit determination. Icarus (ISSN: 0019-1035) 384, 115080. http://dx.doi.org/10.1016/j.icarus.2022.115080.

Sansaturio, M.E., Arratia, O., 2012. Mining knowledge in one night stands data sets. Mon. Not. R. Astron. Soc. 419 (4), 3399–3405. http://dx.doi.org/10.1111/j.1365-2966.2011.19978.x.

Schlafly, E.F., et al., 2012. Photometric calibration of the first 1.5 years of the Pan-STARRS1 survey. Astrophys. J. 756, 158. http://dx.doi.org/10.1088/0004-637X/756/2/158, arXiv:1201.2208.

Southworth, R.B., Hawkins, G.S., 1963. Statistics of meteor streams. Smithson. Contrib. Astrophys. 7, 261–285.

Taff, L.G., 1984. On initial orbit determination. Astron. J. 89, 1426–1428. http://dx.doi.org/10.1086/113644.

Taff, L.G., Hall, D.L., 1977. The use of angles and angular rates I: Initial orbit determination. Celestial Mech. 16 (4), 481–488. http://dx.doi.org/10.1007/BF01229289.

Vereš, P., et al., 2017. Statistical analysis of astrometric errors for the most productive asteroid surveys. Icarus 296, 139–149. http://dx.doi.org/10.1016/j.icarus.2017.05.021, arXiv:1703.03479.

Weryk, R., et al., 2020. Linking isolated tracklets to improve asteroid discovery. In: American Astronomical Society Meeting Abstracts #235. In: American Astronomical Society Meeting Abstracts, vol. 235, p. 329.05.

Wright, E.L., et al., 2016. The albedo distribution of near Earth asteroids. Astron. J. 152, 79. http://dx.doi.org/10.3847/0004-6256/152/4/79, arXiv:1606.07421.

[4] We utilized the `fdiff_cor` routine from the OrbFit package for the differential corrections (http://adams.dm.unipi.it/orbfit/).