Icarus **151**, 150–159 (2001)
doi:10.1006/icar.2001.6594, available online at http://www.idealibrary.com on IDEAL®

The Asteroid Identification Problem IV: Attributions

Andrea Milani

Dipartimento di Matematica, Università di Pisa, Via Buonarroti 2, 56127 Pisa, Italy
E-mail: milani@dm.unipi.it

Maria Eugenia Sansaturio

E.T.S. de Ingenieros Industriales, University of Valladolid, Paseo del Cauce s/n 47011 Valladolid, Spain

and

Steven R. Chesley

Navigation & Mission Design Section, Jet Propulsion Laboratory, Pasadena, California 91109

Received June 20, 2000; Revised January 8, 2001; Posted online April 23, 2001

Existing archives of asteroid observations contain many objects with very short observed arcs. In this paper we present a method that we have used with considerable success to attribute these short arc "discoveries" to other objects with better defined orbits. The method consists of a three-stage filtering process whereby several billion possible attribution/orbit pairs are systematically analyzed with more and more exact algorithms, at each stage rejecting improbable cases. The first stage compares an attributable, by definition a synthetic observation representative of all the observations over a short arc, with the predicted observation for each available orbit. The second stage compares the proposed attributable observations with predicted positions from the known orbit using conventional linear covariance techniques, considering both the position and motion on the celestial sphere. In the final filter we attempt to compute a best-fitting orbit by differential corrections using the combined dataset. With this algorithm we have found 1675 attributions in approximately one year of operations, in addition to 902 identifications found with another algorithm. We discuss the lessons learned from this one-year experiment and the possibilities of further improvement and automation of the procedure.

1. INTRODUCTION

The general problem of asteroid identification aims at detecting cases in which the same physical object has been observed in two arcs, separated in time, without an a priori knowledge that they were the same. The algorithms being used, by our group as well as by others, to achieve identifications are many and very different. The main distinction is whether the possible identifications are selected by proximity in the observation space or by proximity in the phase space of orbital elements. Since the algorithms in these two cases are very different, we prefer to use distinct words, following the definitions given in the first paper of this series (Milani 1999) and corresponding to the distinction between different identification procedures in Marsden (1985). An *orbit identification* is a case in which the observational data of both arcs are sufficient to separately solve for two orbits, one for each arc; then the input data are two sets of orbital elements, and the identifications are proposed on the basis of some measure of similarity of the orbits (which is not just the size of the difference in the orbital elements; Milani *et al.* 2000a). In contrast, a *linkage* joins together two observed arcs, each one over a too short time span and/or with too few observations to allow for a meaningful orbit determination, in such a way that the two together are enough to compute an orbit. The third case is mixed: an *attribution* is obtained by comparing an amount of observational data, insufficient to compute a reliable orbit for one arc, to the orbit already computed for the other arc. As a matter of principle, this goal is obtained by predicting observations, based upon the orbit which is known for one arc, and then comparing with the observed data for a second arc.

It is important to realize that attributions sometimes are very simple and sometimes are very complicated. When the time span of the better observed arc is very long, covering several apparitions, the predictions of the observations can be very accurate and the conjectured attribution is obtained by a simple comparison: the data either are or are not consistent with the prediction, within the expected observational errors. We find this kind of situation when, for example, observations are attributed to numbered asteroid orbits. In contrast, the shorter the time span of the arc used to compute an orbit, the larger the uncertainty of the predictions. For short arc orbits the comparison between predicted and observed values cannot be done meaningfully unless the prediction uncertainty is accounted for in the comparison algorithm.

0019-1035/01 $35.00

If all the asteroids detected (or at least the vast majority of them) were followed up for a time span long enough to compute a reasonable orbit, then orbit identification would be a more effective algorithm because it indeed exploits more information; in intuitive terms, the comparison taking place in a six-dimensional space can provide tighter constraints than a comparison in a two-dimensional space (the celestial sphere). Unfortunately the majority of asteroid discoveries are not followed up after either one or just a few nights. This fact results from priority being given only to discovery observations, as opposed to observations allowing recovery and good orbit determination, in giving credit to the observers. It is also a consequence of the pressure to preferentially perform follow-up observations of unusual objects.

Let us take as an example the dataset of asteroid astrometric observations made public by the Minor Planet Center (MPC) on April 18, 2000. Out of 135,435 provisional designations (corresponding to independent discoveries of asteroids still unnumbered) there were 14,416 with observed arc shorter than 12 hours (pre-1992 "one night stands"), 19,062 longer than 12 but shorter than 36 hours (observations in two consecutive nights), and 18,435 longer than 36 hours and shorter than 60. The reasons for a larger number of two-nighters has to do with the criterion used by the MPC to assign a new designation, which was modified in 1992 to the effect that a one night stand is no longer provided with a provisional designation. Essentially none of the one night stands since 1992 are currently made available by the MPC for outside analysis.

Although we have successfully applied an orbit identification algorithm even to orbits based upon observed arcs of 2–4 days, for roughly three out of four of the designations with arcs between 2 and 4 days we are not currently computing any orbit at all.[1] Thus, limiting the identification algorithms to the orbit identification case would represent a significant decrease of the opportunities of identification.

As long as the residuals (observations, conjectured to belong to the same object, minus prediction) are small enough, the linear formalism of normal and covariance matrices provides a rigorous and comparatively efficient algorithm to assess the likelihood of an attribution. Nonlinear methods are available to deal with cases in which the residuals are well outside the linearity region, e.g., the semilinear method of Milani (1999). Indeed, these methods have been successfully used when the problem is to investigate the suitability of a specific attribution, as is the case when the observations belong to a targeted recovery of the same object. Much more challenging is to *propose attributions*, that is, to start from a large catalog of orbits and a large catalog of short arc observations and to select a reasonable number of couplings between them, to be confirmed by a more detailed analysis of each case. Using again the data from the April 2000 monthly update, we have computed 67,324 unnumbered asteroid orbits and have generated 74,656 *attributables* (this term is better explained in Section 2) that might be linked to the unnumbered orbits. This means we should, in principle, perform about 5×10^9 comparisons.

[1] Added in proof: We have improved our preliminary orbit algorithms, and we now compute a least squares orbit for more than half of the two-nighters.

This large number of pairs to be tested implies that there are two problems to be considered distinctly. On one hand we would like to define rigorous algorithms to filter out, among the huge number of orbit-attributable pairs, the ones that could correspond to the same object. On the other hand, the computational efficiency of the filtering procedure is critical to apply the algorithms with reasonable computing resources.

This paper contains the definition of a procedure for selecting orbit-attributable pairs and reports on the practical tests performed in about one year of operation. The main feature of this procedure is that it consists of a three-stage filtering, with the pairs passing the first filter further analyzed with the second filter and finally the ones passed by the second filter confirmed by accurate least-squares fit. Thus, unlike the procedure for orbit identifications of Milani *et al.* (2000a), the three-stages are qualitatively different, and they require different strategies for optimization.

Orbit identification algorithms are discussed in Milani *et al.* (2000a), but there is one interaction with attributions that needs to be discussed here: after an identification has been found, with any one of the two algorithms, this new orbit can be used to search for attributions to it. Since in most cases the orbit resulting from an identification is multi-opposition, the uncertainty will be much smaller and additional attributions, if they indeed exist in the observations archives, are comparatively easy to add. However, this scheme requires the implementation of a recursive procedure with considerable efficiency problems. We discuss here the extent to which the entire sequence of procedures for orbit identification, attribution, and attribution to identification can be automated.

This paper is organized as follows: Section 2 describes the three-stage filtering procedure. Section 3 discusses the results of the application of this procedure for about one year and analyzes the properties of the identifications found in this way. Section 4 contains an overall appraisal of the results, a comparison with the results obtained by other groups, and some indications on the open problems to be solved to further increase the efficiency and the productivity of searches for identifications.

2. THE FILTERING PROCEDURE

To give a global view of the way our three-stage procedure operates, Table I shows an example (referring to the April 2000 update) of the number of pairs passing through the stages of our filtering procedure. For the four classes of orbits, corresponding to arcs longer than 180 days, between 20 and 180 days, between 4 and 20 days, and between 2 and 4 days, respectively, we give the number of orbits available, which is multiplied by the number of attributables to give the number of pairs, the number of pairs passed by each filtering stage, the number of submissions, and the number of attributions published.

TABLE I
The Attribution Filtering Efficiency

Arc length (days)	>180	20–180	4–20	2–4
Number of orbits	20872	18179	24188	4085
Number of pairs	1.56×10^9	1.36×10^9	1.81×10^9	0.30×10^9
Number passing 1st filter	978,903	691,076	1,420,594	203,150
Number passing 2nd filter	374	7026	396,795	97,744
Number passing 3rd filter	183	1208	7010	1701
Number submitted	68	40	101	23
Number published	67	40	85	22

Note. Data are from the April 18, 2000 MPC observational dataset update. The number of attributables was 74,656.

Given the large number of orbit-attributable pairs, the first stage of the filtering procedure needs to be as computationally simple as possible. It does not matter if the first filtering stage selects a large number of "false positives," provided that this number is much smaller than the total number of pairs being examined. From Table I we can see that the number of pairs passed by the first filter (described in Section 2.1) is less by more than a factor 1,000 with respect to the total number of pairs entering the filter.

The second stage of filtering, described in Section 2.2, is also effective in reducing the number of pairs to be considered, but in a very uneven way. The predictive value of the quantity σ we compute as a measure of similarity in the observations space is very good for long arc orbits (180 days and more, which in most cases means multi-opposition) and is satisfactory for medium arc orbits (between 20 and 180 days, which in most cases means extensive observation during one opposition). For short arc orbits the second filter is not very effective and, as discussed in Section 4, this is likely to be a point on which further improvements are possible.

The results of the third stage, described in Section 2.3, are more difficult to assess in a quantitative way because, out of the successful differential corrections, only a fraction are submitted as identifications to the MPC. The attributions passed by the three selection stages still need to be assessed, by means of a formally defined protocol, discussed in Section 3.1; in many cases a visual inspection of the residuals is also necessary to evaluate the reliability of the proposed identification. The submissions also depend upon the outcome of the competition with other groups of identifiers; that is, an identification we have found is not submitted if it has been already found by someone else. However, the speed of the computation is an achievement in itself: if others are faster in submitting identifications, they deserve the credit.

2.1. First Selection

The first step of the filtering procedure consists of making a rough comparison of the available short arc observations with the predictions resulting from the available nominal orbit. This means we need two catalogs to start with. One catalog contains the nominal orbits, obtained by a least-squares fit to the observations of arc 1. The second catalog contains the *attributables*, which are single observations computed in such a way that they can represent *all* of the observations of arc 2. Before giving a more precise description of the content of an attributable, we need to understand how it is going to be used, both in the first and in the second filtering stages.

In the first step, the nominal orbits have to be propagated to the epoch of the attributables. The problem is that the number of attributables is the number of short arc discoveries. We use all the arcs containing at least two observations[2] not degraded in accuracy and with a time span not exceeding 8–10 days. Thus, the total number of attributables is usually more than 70,000 (e.g., in the April 2000 update, it was 74,656); moreover, they are unevenly distributed in time, being much denser in recent times. Propagating each orbit to each one of these times would be very inefficient. Typically we perform about 5,000 integration steps for each orbit, and we would be forced either to perform integration steps shorter than the ones optimal for computational efficiency or to use interpolation procedures which are also computationally expensive. Therefore in the attributable record we include an extrapolated observation for the nearest time which is an integer multiple of a fixed Δt (we currently use 10 days).

To obtain this, we first compute the straight lines in the (t, α) and in the (t, δ) planes best fitting to the data of arc 2. Then we define a *central time* t_m as the arithmetic mean of the observation times, neglecting the observations which are degraded in accuracy (with weights corresponding to RMS errors larger than 4 arcsec.). We then identify the time t_{step}, which is a multiple of Δt and closest to t_m, and compute the value on the best-fitting lines of $\alpha(t_{step})$, $\delta(t_{step})$.

The use of a linear fit over a time span of up to $\Delta t/2 = 5$ days is inaccurate, but the first filtering stage uses a control $d < R$ on the distance d between prediction and observation which is large, e.g., $R = 1.5$ degrees (d is computed by the usual metric on the celestial sphere: $d^2 = \Delta\delta^2 + \cos^2\delta\,\Delta\alpha^2$).

Of course, the use of such a loose control results in many false positives, but the area of the region within 1.5 degrees from a given position on the celestial sphere is only $A_1 \simeq 7$ square degrees. Taking into account that most asteroid detections take place in a band around the ecliptic which has an area of the order of 10,000 square degrees, this explains the ratio between pairs examined and those passing the first filter. Thus the value of 1.5 degrees is appropriate to make the second stage of our procedure efficient.

We have expended a significant effort to make the orbit computation efficient by optimizing the orbit propagation routines of the `OrbFit` software. Nevertheless, the first filter is a CPU-intensive step of the attributions procedure since the orbit computation must be accurate and the attributables are spread over

[2] There were (in the April 2000 update) 7,464 designations corresponding to only one observation, and we cannot use them, unless they can be attributed to an orbit which is already very good, such as a numbered one.

a total time span of about 100 years. As an example, during the May 2000 update we have used for this stage about 50 CPU hours (spread over three different computers). However, it is not necessary to search for attributions to all the orbits for all the attributables every month, as we did in April and May 2000; only the orbits changed since last month need to be tested on all attributables, and the new attributables need to be tested for all the orbits.

2.2. The Checking Algorithm

In the second filtering stage we propagate the orbit to the exact time of the attributable, rather than the rounded time used in the first stage. As the time of the attributable we use the central time t_m defined above. The attributable contains the values of $\alpha(t_m)$ and $\delta(t_m)$ as derived from the linear fit, but also the slopes of the same fit, that is, the proper motions $\dot\alpha(t_m)$ and $\dot\delta(t_m)$, which provide the rate of motion on the plane tangent to the sky. This information allows us to perform the comparison with the predictions in a four-dimensional space.

The orbit is propagated with the full variational equations, in such a way that the covariance of the predictions can be computed (see Milani 1999, Section 3.1). Let $\tilde\Gamma_{obs}$ be the 2×2 covariance matrix of the predicted angles (α, δ) and $\tilde C_{obs} = \tilde\Gamma_{obs}^{-1}$ the corresponding normal matrix. Since it is better at this stage to use a prudent estimate, we assume that the observation errors could be comparatively large, e.g., with an RMS $\sigma_{obs} = 2$ arcsec. Then, the covariance matrix expressing this assumed uncertainty in the (α, δ) plane is

$$\Gamma_{obs} = \begin{bmatrix} \sigma_{obs}^2/\cos^2\delta & 0 \\ 0 & \sigma_{obs}^2 \end{bmatrix}.$$

Let $C_{obs} = \Gamma_{obs}^{-1}$ be the corresponding normal matrix. Then, the likelihood that the predicted and the attributable observations are indeed the same can be expressed by the two-dimensional *penalty* (Milani *et al.* 2000a)

$$\Delta Q_{obs} = \begin{bmatrix} \alpha(t_m) - \alpha \\ \delta(t_m) - \delta \end{bmatrix} \cdot [C_{obs} - C_{obs}(C_{obs} + \tilde C_{obs})^{-1} C_{obs}] \times \begin{bmatrix} \alpha(t_m) - \alpha \\ \delta(t_m) - \delta \end{bmatrix}.$$

Since the proper motion data are also available, let $\tilde\Gamma_{vel}$ be the covariance matrix of the predicted proper motions $(\dot\alpha, \dot\delta)$, and let $\tilde C_{vel} = \tilde\Gamma_{vel}^{-1}$ be the normal matrix. The uncertainty of the proper motions as estimated from observations depends upon the length Δt of the observed arc: $\sigma_{vel} = \sqrt{2}\sigma_{obs}/\Delta t$. Then the covariance matrix expressing this assumed uncertainty in the $(\dot\alpha, \dot\delta)$ plane is

$$\Gamma_{vel} = \begin{bmatrix} \sigma_{vel}^2/\cos^2\delta & 0 \\ 0 & \sigma_{vel}^2 \end{bmatrix}$$

and, with $C_{vel} = \Gamma_{vel}^{-1}$, we can use the same two-dimensional penalty:

$$\Delta Q_{vel} = \begin{bmatrix} \dot\alpha(t_m) - \dot\alpha \\ \dot\delta(t_m) - \dot\delta \end{bmatrix} \cdot [C_{vel} - C_{vel}(C_{vel} + \tilde C_{vel})^{-1} C_{vel}] \times \begin{bmatrix} \dot\alpha(t_m) - \dot\alpha \\ \dot\delta(t_m) - \dot\delta \end{bmatrix}.$$

The two penalties have to be combined to assess the likelihood of the attribution: we use

$$\sigma = \sqrt{\Delta Q_{obs} + \Delta Q_{vel}}$$

because each penalty is an increase in the value of the target function, which is related to the square of the residuals (in this case the residuals are the difference between the prediction for the given orbit and the observation of the given attributable). It would be possible to use a full four-dimensional penalty, taking also into account the correlation between predicted angles and proper motions, but this does not appear to be necessary.

The computational cost of the second stage is much less than that of the first one, essentially because the number of pairs orbit-attributable to be tested has been decreased by three orders of magnitude by the first filter. As an example, during the May 2000 update we have used about 7 CPU hours.

The geometrical meaning of this method is the following. The observation as given in the attributable has an unknown error; the real one could be anywhere within a distance of the order of σ_{obs}, along great circles on the sphere (the factor $\cos^2\delta$ correctly accounts for the metric on the sphere); this is analytically described by the normal matrix C_{obs}. The prediction is in turn uncertain, its confidence region being the ellipse defined by the normal matrix $\tilde C_{obs}$. We are looking for intersections of the two confidence regions. This situation is mathematically the same as the identification problem in the elements space, only in two dimensions. The same argument applies to proper motions.

It is not easy to decide a priori what control value σ needs to be used for confirming the proposed attribution and passing it to the final differential correction procedure. Figure 1 shows the values of σ for the attributions which have been accepted by the MPC. Note that during the April update we have ourselves selected for differential corrections all the cases with $\sigma < 20$, but only $\simeq$25% of these have $\sigma \geq 8$. Among the pairs later passing the third filter, only 2% had $\sigma \geq 8$. This means that the control value can be kept very low, and we actually plan to use a lower value in the future. The number of good cases that could be missed by decreasing the control value would be small: as shown in Fig. 1, the number of published attributions with $\sigma > 15$ is just 5 (including one case off of scale in the plot, with $\sigma \simeq 25$).

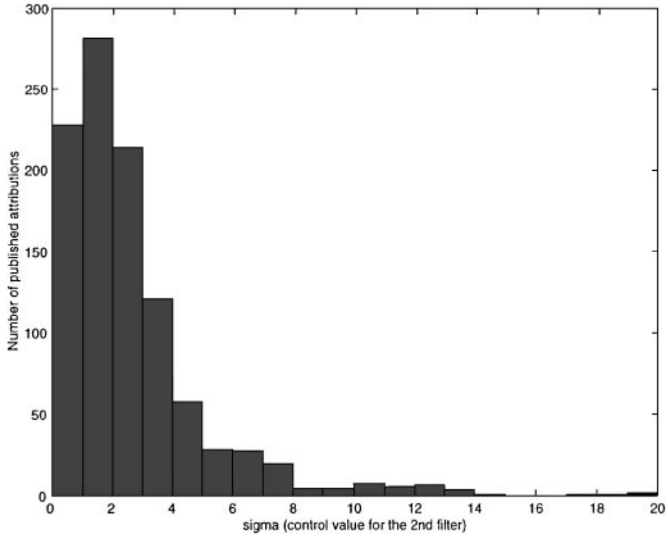

FIG. 1. Histogram of the number of attributions, submitted by us and published by the MPC, as a function of the control parameter σ of the checking algorithm.

From Table I we can infer that passing the second filtering stage with a low value of σ has a good predictive value for attributions to multi-opposition and medium arc (>20 days) orbits; that is, a significant fraction of the pairs passing the second filter are also passing the third filter. However, for shorter arcs the second filter is not so effective and a very significant computational effort has to be used in the third stage; this can be understood as follows.

For short arc orbits, the confidence boundary of an observation prediction, many years after the asteroid has been lost, is typically several degrees long. Then the ratio between the area in which the attributables are passed by the first filter ($A_1 \simeq \pi R^2 \simeq 7$ square degrees) and the area A_2 of the confidence region acceptable for the second filter is roughly

$$\frac{A_1}{A_2} \simeq \frac{\pi\, R^2}{2\, R\, \sigma_w\, \sigma},$$

where σ_w is the width of the confidence region, which can be computed as the square root of the lower eigenvalue of Γ_{obs}. For $R = 1.5$ degrees and a control value of the second filter $\sigma = 20$

$$\frac{A_1}{A_2} \simeq \frac{7'}{\sigma_w}.$$

This simple order of magnitude computation shows that for a width of the confidence region of the order of several arcmin the second filter becomes ineffective. If the available computing resources are not enough, a decrease in the control value of σ is an acceptable compromise, in particular for short arc orbits: it would result in a significant decrease of the computational load of the third stage, with the possible loss of a small fraction of the real attributions detected.

2.3. The Confirmation Algorithm

The third step is the application of the classical differential correction procedure to all the orbit-attributable pairs selected by the second stage. We try to fit (in the least-squares sense) an orbit to the observations of both arcs, by using as first guess the orbit of the first arc. For this we use the `OrbFit` software, which has been thoroughly optimized, but still this is the most CPU time consuming part of the attributions work; for the May 2000 update we have used more than 100 hours of CPU. As is clear from Table I, this is mostly due to the short arc orbits for which the second filter is not selective enough.

The differential correction procedure we use has three features which are relevant for this paper. First, the differential correction is iterated until convergence is achieved; this feature is especially important because fitting an attribution to a single opposition orbit may need a comparatively large correction, well outside of the linearity region. For controlling the convergence of the iterations, we use as termination condition

$$\|\Delta X\| = \sqrt{\Delta X \cdot C \Delta X / 6} < 10^{-3},$$

where the norm $\|\Delta X\|$ of the last correction to the orbital elements is defined by means of C, the normal matrix in the elements space, in such a way that a correction has a small norm if it is much less than the corresponding uncertainty.

Second, the algorithm normally used to fit asteroid orbits to the observations, e.g., to prepare the orbit catalog we use as starting point, includes an automated outlier rejection procedure (Carpino *et al.* 2001). However, we disable the outlier rejection for the attribution confirmation procedure. The outliers which have already been rejected during the fit of the first arc remain rejected, but no others are added, in particular no rejections are performed among the observations of the second arc. This way of acting is in general appropriate, although once an attribution is accepted it is sometimes possible to further improve the fit by outlier rejection. On the other hand, if we were to perform outlier rejection for attributions which are still dubious, we would often end up rejecting proportionally too many of the few observations of the second arc, even rejecting the second arc entirely and going back to the first arc orbit.

Third, the differential correction algorithm we use minimizes a *weighted RMS* of the residuals,

$$\|\xi\| = \sqrt{\frac{1}{m} \sum_i \xi_i^2 w_i},$$

where $m/2$ is the number of observations, ξ_i are the residuals, and the weight w_i is computed for each observation as a function of the performance of the same observatory over a time span long enough to allow for statistical analysis (Carpino *et al.* 2001).

An orbit-attribution pair is considered passed by the third-stage filter when the weighted RMS of the residuals $\|\xi\|$ is below some critical value M_ξ. For a successful orbit determination,

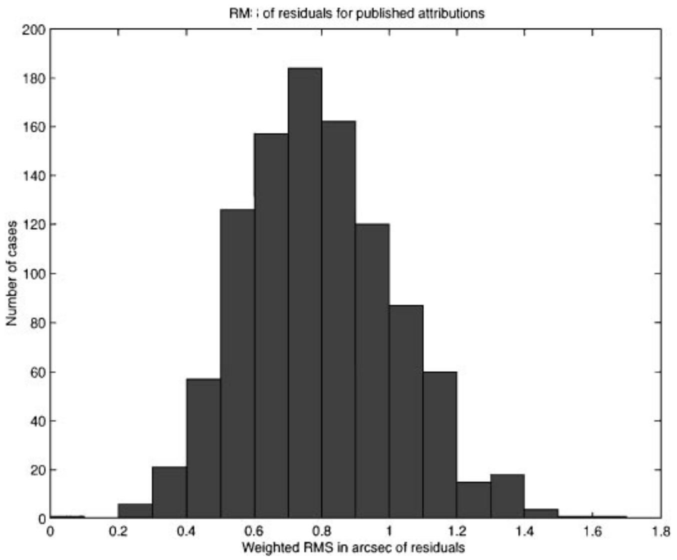

FIG. 2. Histogram of the number of attributions, submitted by us and published by the MPC, as a function of the weighted RMS of the residuals of the least squares fit for the observations of both arcs.

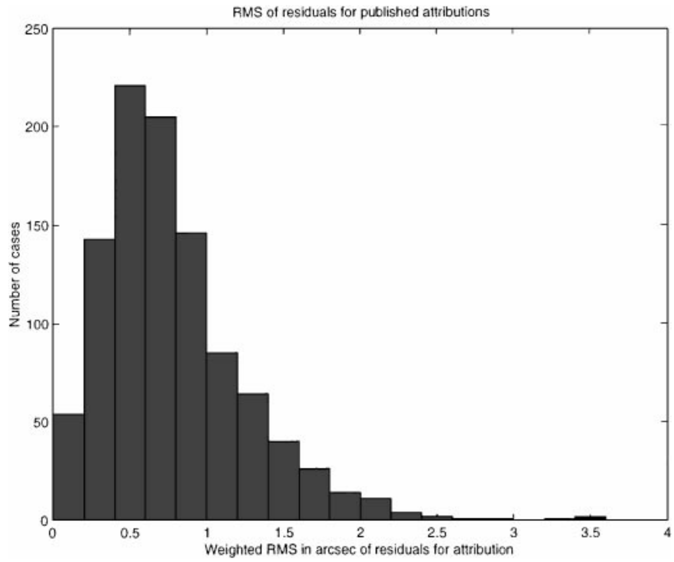

FIG. 3. Histogram of the number of attributions, submitted by us and published by the MPC, as a function of the weighted RMS of the residuals of the attributed observations only.

we expect that $\|\xi\| \lesssim 1$, but this is true only for good orbits and after outlier rejection. Thus, for the attribution confirmation procedure, a somewhat higher value is acceptable. In Fig. 2 we can see that most of the attributions already accepted by the MPC have $\|\xi\| < 1$, but a fraction have a somewhat higher value. For the first few months of operations, we were using $M_\xi = 1.8$. If we had used $M_\xi = 1.5$, we would have lost only 2 published attributions; thus we later changed the control value to $M_\xi = 1.5$.

The fraction of pairs passing the third checking stage, that is, with a meaningful least-squares fit, is small, especially for short arc orbits. Nevertheless, it is not the case that all of them correspond to real identifications, and other parameters have to be analyzed to assess the quality of the fit. Another interesting indicator is $\|\xi_{att}\|$, the weighted RMS of the residuals of the attributed observations only, since the RMS of all the residuals could be dominated by the residuals of the first arc, especially when it has a much larger number of observations. We have been using a control $\|\xi_{att}\| < 3.5$; as shown in Fig. 3, we could have used $\|\xi_{att}\| < 2.5$ without a significant loss of attributions. Indeed, for the April 2000 update, the number of pairs passing the third filtering stage, as shown in Table I, was obtained by using the controls $\|\xi\| < 1.5$ and $\|\xi_{att}\| < 2.5$.

Another useful criterion is the detection of signatures in the residuals for the attributed observations. Since they belong to a short arc, a simple linear fit to the residuals in both α and δ as functions of time gives enough information in most cases. Separately for right ascension and for declination we compute the *bias*, which is the value of the best-fitting line at the central time t_m of the attributed arc, and the *span*, which is the difference between the values of the best-fitting line at the beginning and at the end times of the same arc.

Figures 4 and 5 show the values of such quantities for the attributions submitted by us and published by the MPC. From these figures it is apparent that an additional control, based upon these values, could be used. Starting from May 2000, we began using such an additional control based upon bias and span to select the attributions to be submitted. Let the biases be B_α and B_δ, and let the spans be S_α and S_δ; we require

$$\sqrt{B_\delta^2 + \cos^2\delta\, B_\alpha^2} < 3 \text{ arcsec} \quad \sqrt{S_\delta^2 + \cos^2\delta\, S_\alpha^2} < 4 \text{ arcsec}$$

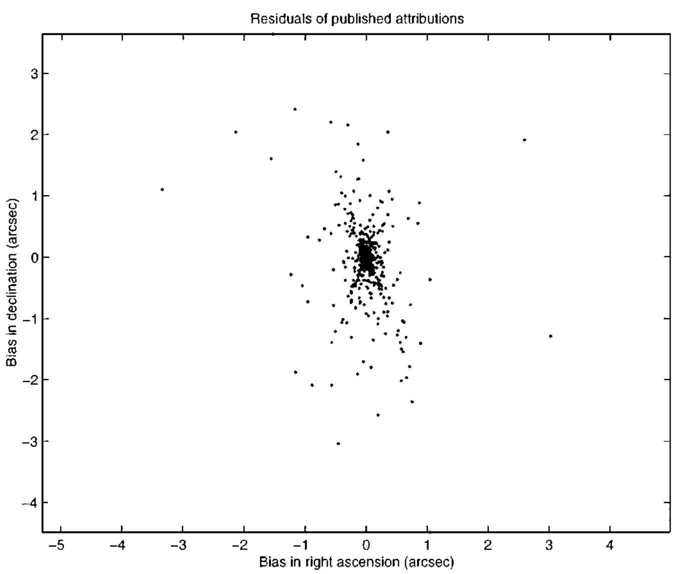

FIG. 4. Biases of the attributions, submitted by us and published by the MPC.

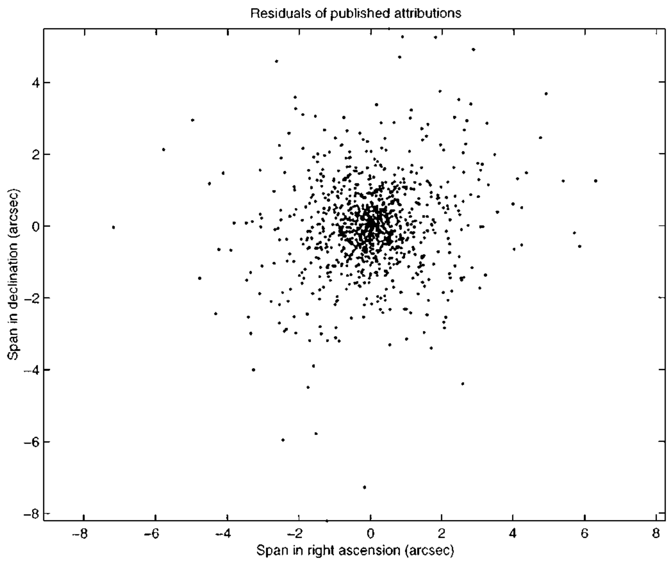

FIG. 5. Spans of the attributions, submitted by us and published by the MPC.

for multiple night attributions and

$$\sqrt{B_\delta^2 + \cos^2\delta\, B_\alpha^2} < 2\ \text{arcsec} \quad \sqrt{S_\delta^2 + \cos^2\delta\, S_\alpha^2} < 2\ \text{arcsec}$$

for one night stands with more than 2 observations; the span has to be less than 1.6 arcsec for one night stands with only 2 observations. All these limit values have been empirically determined from experience.

3. RESULTS

The three-stage filtering procedure to propose attributions has been regularly used by our group since March 1999 and has resulted so far in 1,675 attributions accepted and published by the MPC. In this section we discuss the procedure for selecting, among the attributions confirmed by least-squares fit, the ones to be submitted to the MPC, and we present an analysis of the submissions and publications until May 2000. Table II contains a summary of all of the object identifications submitted by our group since March 1999 and of those published by the MPC.

3.1. Handling of the Confirmed Attributions

The attributions, even if confirmed by a successful least-squares fit with low RMS of the residuals, cannot always be considered certain. To assess the reality of a proposed attribution we have to take into account, besides the RMS and the other fit parameters discussed in Section 2.3, the presence of systematic signatures and/or outliers in the residuals and the length of both observed arcs. Although our three-stage filtering procedure is sophisticated, it can only provide a necessary condition for identification. At present, the intervention of an experienced human is still essential, and in many cases the residuals have to be inspected visually to decide if there is some signature suggesting an erroneous identification.

The main criterion to assess the likelihood of an identification is the length of the two observed arcs (and the number of observations). As an example, for asteroids already observed at multiple apparitions, a good fit for an attributed arc is reliable even when this second arc is short, e.g., a one night stand. In contrast, when both arcs are short, an ostensibly good fit could be deceiving.

It is therefore a fact that some proposed attributions can be rated neither true nor false but have to remain as "possible," in some cases even "probable," but cannot be considered certain. The MPC is presently using a cautious policy, which appears sensible, trying to publish only identifications that appear almost certain, to minimize the number of "erroneous identifications." The others are left "pending," either because they have been submitted to the MPC and not published, or because they have been rated too uncertain and therefore not been submitted by us. This pending status could be better formalized.

There is one way to remove a proposed identification from this pending status and that is to find other observations that could belong to the same object. As a matter of principle one could propagate the best-fitting orbit resulting from the identification, predict observability conditions, and then use a

TABLE II
Identifications Results in One Year of Operation

Monthly update		Attributions	Attributions to attributions	Orbit IDs	Attributions to orbit IDs
Mar 2, 1999	Submitted	90	0	100	8
	Published	82	0	94	8
Apr 2, 1999	Submitted	115	0	27	7
	Published	86	0	21	6
May 4, 1999	Submitted	498	0	19	7
	Published	329	0	16	6
Aug 31, 1999	Submitted	91	3	13	2
	Published	60	2	9	1
Sep 28, 1999	Submitted	82	0	42	11
	Published	70	0	42	11
Oct 26, 1999	Submitted	42	3	6	0
	Published	15	2	6	0
Nov 23, 1999	Submitted	80	3	25	7
	Published	42	3	24	7
Dec 22, 1999	Submitted	58	6	11	7
	Published	55	6	10	7
Jan 24, 2000	Submitted	109	16	192	46
	Published	108	16	183	46
Feb 22, 2000	Submitted	125	27	90	18
	Published	114	24	85	18
Mar 20, 2000	Submitted	41	7	147	23
	Published	39	7	144	22
Apr 18, 2000	Submitted	232	27	135	25
	Published	214	23	131	24
May 23, 2000	Submitted	183	32	142	17
	Published	179	26	137	17
TOTAL	Submitted	1746	124	949	178
	Published	1393	109	902	173

telescope (and/or a plate archive search) to either confirm or contradict the reality of the identification (contradiction requires a procedure of *negative observation*, as discussed in Milani *et al.* (2000b)). In practice, nowadays this method can be applied only for especially important cases, such as Near Earth Asteroids; in the future, with better automation, it could be used more extensively.

For most proposed attributions to an orbit computed with the data of a single apparition and with the second arc of short duration, e.g., a one night stand, a certain confirmation can come only with a serendipitous rediscovery. Thus it is essential, whenever a possible attribution is found, to restart the same procedure of Section 2, looking for additional attributions. Since a single apparition orbit can be substantially changed by the addition of an attribution to the fit, the new orbit might pick up new attributions. This argument applies to all new orbits computed as a result of a proposed identification, both with the attribution algorithm and with the orbit identification one. In a few cases this procedure, applied recursively, gives spectacular results, with many arcs identified at once. Our record is

$$\begin{aligned}1998\ \mathrm{SP}_{61} &= 2000\ \mathrm{CK}_{108} = 1972\ \mathrm{LH} = 1984\ \mathrm{HF}_2\\ &= 1986\ \mathrm{VQ}_6 = 1988\ \mathrm{GL}_1 = 1990\ \mathrm{UQ}_1,\end{aligned}$$

in which the first identification was found with the orbit identification algorithm, with the other five attributions to the new multi-opposition orbit.

For this reason, Table II has four columns: the attributions to orbits computed on the basis of the already known identifications, the attributions to the orbits recomputed after finding a new attribution, the orbit identifications, and the attributions to the orbits recomputed after finding a new orbit identification.

Thus the procedure to process the orbit-attributable pairs selected by the differential correction stage needs to include the search for these additional attributions, and those which are found are relevant to the credibility of the proposed attribution. After some experimenting, we have decided as a standard procedure to submit to the MPC:

- all the orbit identifications with a good fit;
- all the attributions to multi-apparition orbits;
- all the attributions of arcs spanning more than one night;
- the attributions of one night stands to single apparition orbits only if we can find an additional attribution.

All the other attributions proposed by our algorithm remain in the pending status, because they are one night attributions to single apparition orbits for which no additional attribution has been found so far. After the May 2000 update, we have 2, 219 possible attributions pending (and not submitted to the MPC), all fulfilling the criteria discussed in Section 2.3. This pending status can change only in two ways: either the identification is contradicted by one of the asteroids in the pair being identified with a different one or it is confirmed by finding another attribution in a later monthly update.

Table II lists for each monthly update the identifications submitted by us to the MPC and those published by the MPC. Note that we are not including in the table the identifications we have sometimes proposed and that had already been published, even only the previous day. It is understandable that the different groups of identification hunters, even by using different methods, can end up proposing lists of identifications partially overlapping, and hence the credit goes to the fastest. We are now, before submitting, removing all the identifications published up to the same day.

It is clear from such a table that only a small fraction of the identifications we propose are not accepted, especially since December 1999 when we adopted the more restrictive protocol for submission described above.

3.2. Statistics of One Year of Operation

The total number of object identifications credited to our group in about one year of operations is 2,577. Out of these, 902 were orbit identifications; moreover, 173 were attributions to the newly identified orbits. Still the majority of the identifications (1,393 + 109) have been obtained by the use of only the attribution method. The changes from month to month reflect experiments we have been doing with new procedures, different order of operations, and even the use of different computer hardware. They also reflect the changes in the operations of the other groups hunting for identifications since, as mentioned in the previous section, we are not counting the cases in which we had found an identification and then discovered it had been proposed by others and published, often only a few days before.

Table II does not include a special case of attributions, namely the attributions to numbered asteroids. We have submitted 15 of these and all of them have been published.

A more detailed analysis is possible for the identifications we submitted in the past few months, since the protocol for submissions has been followed in a precise way. The results for the 654 attributions submitted by us and published by the MPC between January and May 2000 (excluding the attributions to other identifications; that is, taking only the first column of Table II) are summarized in Fig. 6 as a function of the length of both observed arcs. The cutoff at 10 days for the length of the second arc is the result of our choice of computing attributables only for arcs shorter than 10 days; for longer arcs, the orbit identification method is generally better. The comparatively small number of one night stand attributions published so far depends upon two factors: first, these single night attributions are rated uncertain and most of them are left pending; second, after 1992 the one night stand data have not been published by the MPC and are therefore not available for our search of identifications.

Out of these 654 attributions only 123 involved two arcs both observed before 1990; 302 involved two arcs both observed after 1997. These figures suggest that the attributions are mostly found among the new observations made available every month, which are identified with other recent observations.

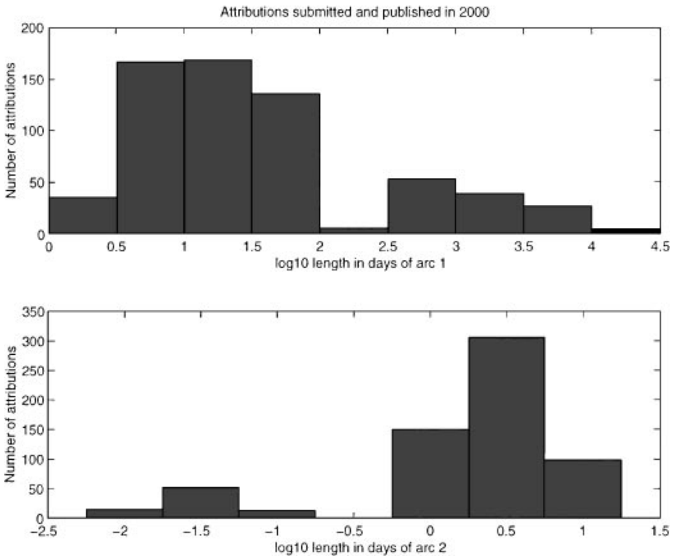

FIG. 6. Length of observed arcs for all the attributions proposed by our group and published by the MPC between January and May 2000. (Top) Length of the first arc (the one for which an orbit was computed and propagated). (Bottom) Length of the second arc, for which an attributable was prepared.

The attribution algorithm appears to be less effective than the orbit identification one for the purpose of finding identifications among the old data, which have been in the archives for years or even decades.

An interesting special case is that of identifications involving Near Earth Asteroids (NEA). Only 4 such cases were found by us and published. Out of these, two were attributions to known NEA, 2000 CN_{101} and 1998 MR_{24}; another was an orbit identification between two known NEA, 1998 $EP_8 =$ 1999 RN_{45} (now, (16912) Rhiannon). The fourth one was an orbit identification of the asteroid 2000 GR_{127}, believed to be a NEA (that is, with nominal perihelion distance $q < 1.3$ AU) with a non-NEA asteroid 1997 GG_3, but the orbit resulting from the identification was not that of a NEA.

4. CONCLUSIONS AND FUTURE WORK

After more than one year of operations with our algorithms, including the one for attributions presented in this paper and the one for orbit identifications presented in Milani *et al.* (2000a), we think the results can be rated satisfactory. Recovering more than 2,500 asteroids using telescopes would have consumed significant resources and so obtaining the same result by pure computation frees the telescope and observer resources for the work which can only be done in that way.

The comparison of our results with those of other groups is not straightforward. The largest number of identifications is found by the MPC staff, as is natural since they have access to the new observations before anyone else. They disseminate the data with an average delay of 2–3 weeks and they efficiently use this time to scan for identifications. The very fact that we find some identifications implies that our algorithms are capable of finding nontrivial identifications that could escape the MPC scrutiny.

Another large fraction of the published identifications are credited to the so-called DANEOPS consortium (A. Doppler, A. Gnädig, G. Hahn, and others). We understand that their methods belong to the attribution class, but it is not easy to compare the efficiency of the two algorithms; probably their good results also depend upon a very efficient data flow and a rather advanced automation. Their purpose is to obtain the largest number of identifications in the shortest possible time after the data are available, and they are indeed very successful in this. Although we also try to be as efficient as possible and to conclude our computations in just a few days after the monthly update, speed is not our main concern. Our main goal is to define new algorithms, which are published in the scientific literature and therefore can be used by others.

Some improvements, which could result in a faster selection of the identifications to be proposed, are indeed suggested by the analysis of our results contained in this paper. As an example, the most computationally intensive steps of the identification search procedure should be performed in an incremental way, rather than by testing all the possible pairs each month. This way of acting could apply to the first filter, as pointed out in Section 2.1, and also to the third filter, which is even more time consuming, especially for short arc orbits. We are indeed experimenting with “cleanup routines” that remove from the list of pairs passed by the second filter the ones having already been submitted to the differential correction test in the previous months. In general, the entire procedure needs to move toward a “steady regime” in which only the orbits and attributables that are either new or have been changed are processed. However, such a regime cannot be adopted until the algorithm control values and parameters, as well as the algorithms themselves, are finalized and frozen.

The level of automation of the entire procedure we are using to propose identification is good but could be improved. The problem is that a fully automated procedure removes the possibility of applying human judgment and we are reluctant to give up the possibility of human intervention, at least in the final stage where the selection of the pairs to be submitted takes place. However, the procedure in the previous steps, including the three filtering stages and even the recursive procedure to search for additional attributions to newly identified orbits, can be automated. In the final selection of the pairs to be submitted, the number of cases to be examined is comparatively small and some human intervention is affordable.

The year of operations described in this paper has been the first and, of course, we have had to learn from experience. Nevertheless, in this time span our group has been the third largest contributor for identifications, supplying a significant fraction of the total number. This fact implies that our work is not only of theoretical interest but is a practical contribution to the problem of finding identifications, which can in turn play an important role in other issues, such as the prediction of close approaches (Milani and Valsecchi 1999) and the recovery of lost asteroids.

On the other hand, the number of identifications that is currently found by all groups is by far insufficient. Every month the number of new designations greatly exceeds the number of new identifications. The conclusion is not that so many new asteroids are actually discovered but only that the data archives are more and more full of short observational arcs, with corresponding very poor orbits (if any), and to which one should add the existing one night stands that are not published. Most of these short arcs should be identified, but they remain unidentified. We believe that in the long run this problem will be the main one: how to clean up the data archives from all these essentially unused data. It is very difficult to estimate the number of identifications remaining to be found, but there must be many tens of thousands and hence our contribution of more than 2,500 appears inadequate. It follows that more advanced algorithms need to be found, and this will be an important goal of our continuing research in this field.

ACKNOWLEDGMENTS

The `OrbFit` free software is maintained by a consortium led by A. Milani, M. Carpino, Z. Knežević, and G. B. Valsecchi; it is available at `http://newton.dm.unipi.it/asteroid/orbfit/`. This research has been supported by the Italian Space Agency under Grants ASI-ARS-98-240 and ASI-ARS-99-81 and by the North Atlantic Treaty Organization under a grant awarded in 1998.

REFERENCES

Carpino, M., A. Milani, and S. R. Chesley 2001. Error statistics of asteroid astrometric observations, in preparation.

Marsden, B. G. 1985. Identifications of minor planets. In *Asteroids Comets Meteors II* (C.-I. Lagerkvist *et al.* Eds.), pp. 3–12. Uppsala University, Uppsala.

Milani, A. 1999. The asteroid identification problem I. recovery of lost asteroids, *Icarus* **137**, 269–292.

Milani, A., and G. B. Valsecchi 1999. The asteroid identification problem II: Target plane confidence boundaries. *Icarus* **140**, 408–423.

Milani, A., A. La Spina, M. E. Sansaturio, and S. R. Chesley 2000. The asteroid identification problem III. Proposing identifications. *Icarus* **144**, 39–53.

Milani, A., S. R. Chesley, G. B. Valsecchi, and A. Boattini 2000. Virtual impactors: Search and destroy. *Icarus* **145**, 12–24.