Double-blind tests program for astrometric planet detection with GAIA. Progress report I: results for tests T0a and T0b.

S. Casertano
Space Telescope Science Institute
M. G. Lattanzi, R. Morbidelli, R. Pannunzio, & A. Spagna
Osservatorio Astronomico di Torino
S. Jancart & D. Pourbaix
Université Libre de Bruxelles
A. Sozzetti
University of Pittsburgh and Harvard-Smithsonian Center for Astrophysics

September 14, 2003

Abstract

We present a preliminary report on the results of the Phase 0 of the double-blind tests aimed at quantifying GAIA's ability to find and measure planetary companions of nearby stars via an extensive simulation and analysis program. The Phase 0 tests have the primary aim of establishing the integrity of the simulation, solution and analysis pipeline, including the transfer of information, the compatibility of the software, and the mutual understanding of definitions and conventions. Phase 0 has now been brought to a successful conclusion.

1 Introduction

In Document PSWG-OAT-001, the GAIA Planets Working Group established a tentative protocol for a comprehensive test of GAIA's ability to detect and measure exoplanets orbiting nearby stars from GAIA astrometric observations. The test aims at verifying 1) the expected sensitivity of GAIA's observations to exoplanets, as a function of mission parameters where applicable; 2) the compatibility and readiness level of the simulation and solution software that different groups have been developing in recent years; and 3) the ability of these groups to communicate information effectively and to understand the possible difficulties in the development of a shared analysis network. The tests will be carried out in double-blind format to ensure the reliability of the results.

Phase 0 of the test is designed primarily to establish a basic level of readiness in the simulation, analysis, and information exchange. Because of its exploratory level, the double-blind protocol is enforced less restrictively, and Solvers are given more information about the simulations than in later stages. Phase 0 is composed of two steps. Step T0a consists of the analysis of time series simulating GAIA observations of 1000 stars, approximately half of which have a Jupiter mass planet with different signal levels. Solvers are asked to identify which stars have a planet on the basis of criteria of their own choice, with the understanding that a reasonable level of false positives and false negatives may be encountered. Step T0b consists of 1000 time series, each representing the GAIA observations of a star+planet combination with strong nominal signal-to-noise ratio; Solvers are asked to estimate the orbital parameters of the planet. Because of the exploratory nature of this test, Solvers could choose to ask for reasonable starting values for the planet's parameters. Of the three Solvers that provided a solution for T0b, only one asked for such information (see below).

This report presents a basic analysis of the results provided by the Solvers. The short summary is that the Solvers achieved high quality results, and Phase 0 of the double-blind tests can be considered successfully concluded.

2 The simulations

Simulations were provided by the group at the Osservatorio Astronomico di Torino. The simulations were made available via Web as plain text files.

Each simulation consists of a time series of observations of a set of stars with given (catalog) low-accuracy astrometric parameters; each observation consists of a one-dimensional coordinate on the instantaneous great circle followed by GAIA at that instant. Such data indicate already a substantial level of processing: both the instantaneous great circle solution and the abscissa solution have been carried out, with nominal zero error (see PSWG-OAT-001).
The scanning law for GAIA has been updated to the most recent expectations, which result in fewer observations and possibly less ability to disentangle near-degenerate solutions than with the original scanning law. Typically, each star is observed between 25 and 40 times; note that the simplest star+planet solution has 12 parameters, and therefore the redundancy of the information may not be as good as desirable.

3 The solutions

Completely independent solutions were produced by Alessandro Sozzetti, Dimitri Pourbaix and Sylvie Jancart. The fact that all three Solvers achieve essentially similar results using potentially different methods and independent codes is especially significant.

3.1 Step T0a

The simulations produced for Step T0a included 480 stars without planet and 520 stars with a Jupiter-mass planet, with distance ranging between 8 and 71 pc and semi-major axis ranging between 0.6 and 2.9 AU. The nominal signal $\alpha = M_p a/(M_* d)$ ranges from 40 to 80 μas, for a single-observation S/N between 5 and 10 (the single-observation noise was set at 8 μas). The number of observations ranges from 22 to 91.

All Solvers define the planet detection process in terms of the quality of the single-star solution. If the single-star solution is statistically acceptable, as measured by some statistical test, then the planet is considered not detected; if the single-star solution can be rejected, then the planet is detected. Of course the detection probability depends on the statistical test and the threshold used; the probability of false detections varies roughly in the opposite direction.

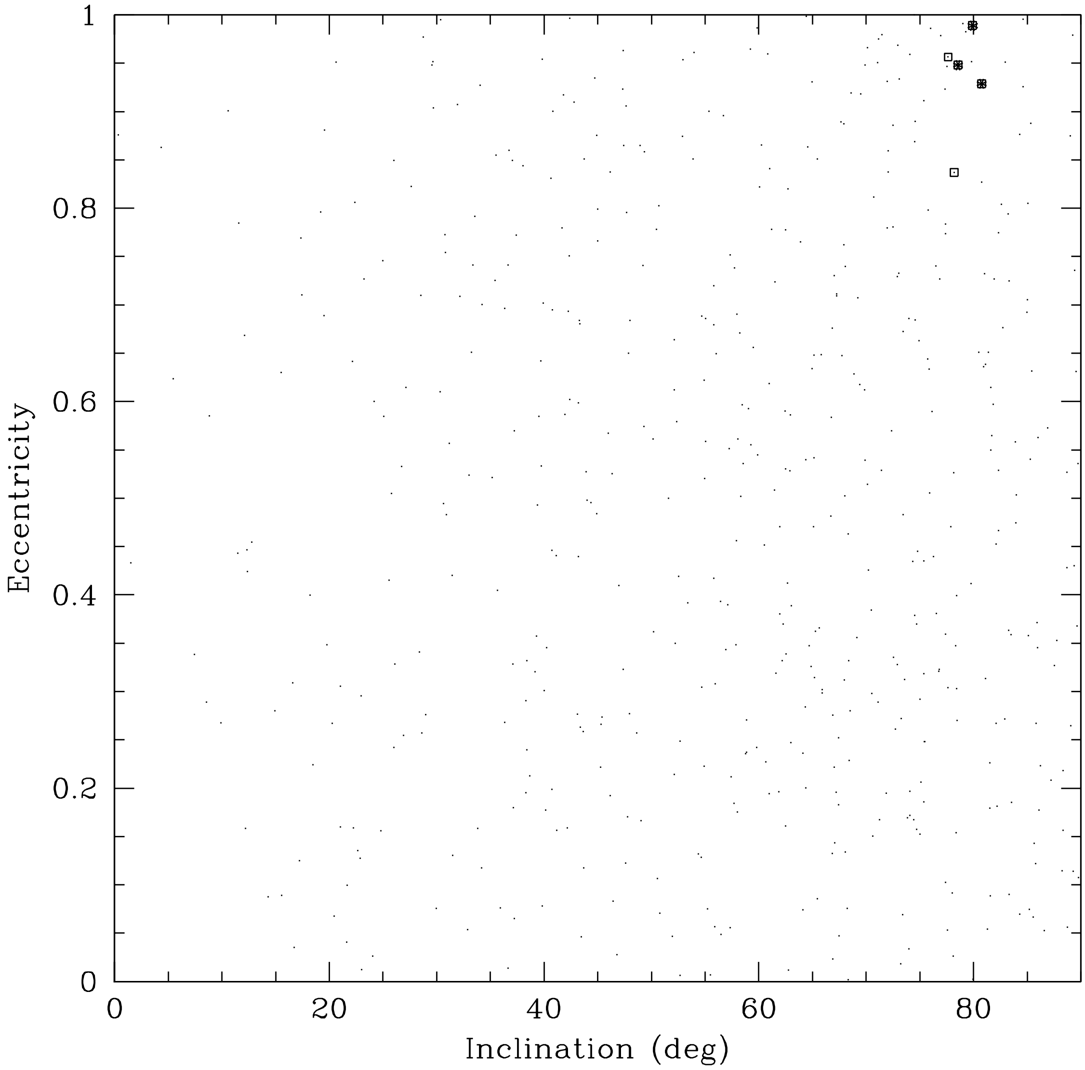

Figure 1: Inclination and eccentricity of T0a planets. Asterisks mark the 3 planets not detected by Sozzetti; open boxes the 5 planets not detected by Pourbaix and Jancart, which include those not detected by Sozzetti).

Pourbaix and Jancart define the detection criterion in terms of the quality of the single-star fit as defined by the $F2$ indicator (see the Hipparcos catalogue, vol. 1, p. 112). The $F2$ parameter is expected to follow a normal distribution with mean 0 and dispersion 1; both Solvers define as non-acceptable values of $F2$ that exceed 3 in absolute value, i.e., a 3-σ criterion (Jancart classifies as single a star with $F2 = -3.003$). On this basis, one can expect about 0.3% of false positives, while the detection probability depends of course on the strength of the signal. Both Pourbaix and Jancart find 485 single stars, of which 480 are in fact single; conversely, 515 stars are classified non-single, all of which do in fact have a planet. The sets of stars identified as single is exactly the same for both, and their $F2$ values are in very close agreement for the single stars, with a typical difference of 0.1 (Jancart does not report the χ^2 or $F2$ value for stars identified as non-single). Therefore Pourbaix and Jancart find *no* false positives and fail to identify only 5 (false negatives) of the 520 planets, all of which have very high inclination and eccentricity. The non-detected planets do not seem to have an especially large period or small number of observations. The finding of no false positives (potentially one in Jancart's case) is in line with the expected number for a sample of this size. There is also no doubt that the detection performance is very good. Larger samples will be needed for a meaningful verification of the expected performance in terms of both false positives and failed detections.

Sozzetti proposes two different criteria for planet detection, one based on the χ^2 test of the single-star solution and the other on the F test (a third criterion, based on the Kolmogorov-Smirnov test on the deviations between best fit single-star solution and observations, is discarded by Sozzetti as less well-behaved). For the χ^2 test, Sozzetti suggests a 99% confidence criterion, i.e., a star is deemed to have a planet if the χ^2 exceeds the value it would assume 99% of the time. On that basis, Sozzetti identifies 480 stars as single, of which 477 are in fact single, and assigns a planet to 520 stars, of which 3 are single. (The F-test with a threshold of 0.1% identifies exactly the same stars.) Therefore Sozzetti's test finds 3 false positives, and fails to detect 3 planets. The number of false positives is in line with the ~ 5 expected with a 99% confidence criterion (which implies a 1% incidence of false positives). As with the other Solvers, the planets that were not detected have high eccentricity and high inclination, and in fact all three of Sozzetti's failed detections were also missed by the other Solvers. The higher detection rate and larger false positive rate in Sozzetti's detections are consistent with the lower detection threshold he adopted. In any event, the differences between the solutions are very minor, and all three solutions were extremely successful in the detection test.

3.2 T0b

The T0b test includes 1000 star+planet system with periods ranging from 0.5 to 5 years and distances from 0.7 to 8 pc. Since the stars are ~ 10 times closer than the T0a test, the typical signal is 10 times larger, with α ranging from 400 to 800 μas(S/N = 50 to 100). Detection should therefore not be an issue. However, some of the longer orbits may not be well sampled during the mission, especially if the eccentricity and inclinations are high.

Both Pourbaix and Sozzetti found the best-fitting orbital solution without any help with the starting values[1], while the solution obtained by Jancart was based on initial values provided by the Simulators.

In the vast majority of the cases, the best-fitting solution reported by all three Solvers matches very well with the truth. We use the period P as an indicator of the quality of the solution, and define $\Delta P = P_{\rm fitted} - P_{\rm true}$ as the difference between the period $P_{\rm fitted}$ determined by each of the three Solvers and the input value $P_{\rm true}$ used by the Simulators. With the exception of a few outliers, the dispersion of ΔP is 0.0002 years for periods 1 year and shorter, and increases gradually to 0.05 years for periods between 4 and 5 years. In fact, the dispersion in ΔP scales roughly with $10^{0.5P}$, as can be seen in Figure 2. The histogram of ΔP is shown in Figure 3 for the Pourbaix solution; the distribution for the other Solvers is essentially the same.

As might be expected, some outliers appear especially at longer periods. These outliers are likely to be due to a combination of two effects: insufficient sampling of the orbit, which can lead to a degeneracy between solutions or to strong sensitivity to noise, and bad initial values. The latter possibility applies only to the Pourbaix and Sozzetti solutions, since Jancart was provided with very good initial guesses as part of the solution process.

Of Sozzetti's best-fit solutions, 23 correspond to periods that differ from the nominal value by more than 10%, well outside the normal distribution of deviations. On the other hand, only 4 of Pourbaix's solutions differ by that amount, possibly indicating that the initial guess used by Pourbaix may be better than Sozzetti's. Many, but not all, of Sozzetti's outliers can be

[1] In this respect, the software of these two Solvers is already compliant with the next phases of the double-blind test campaign.

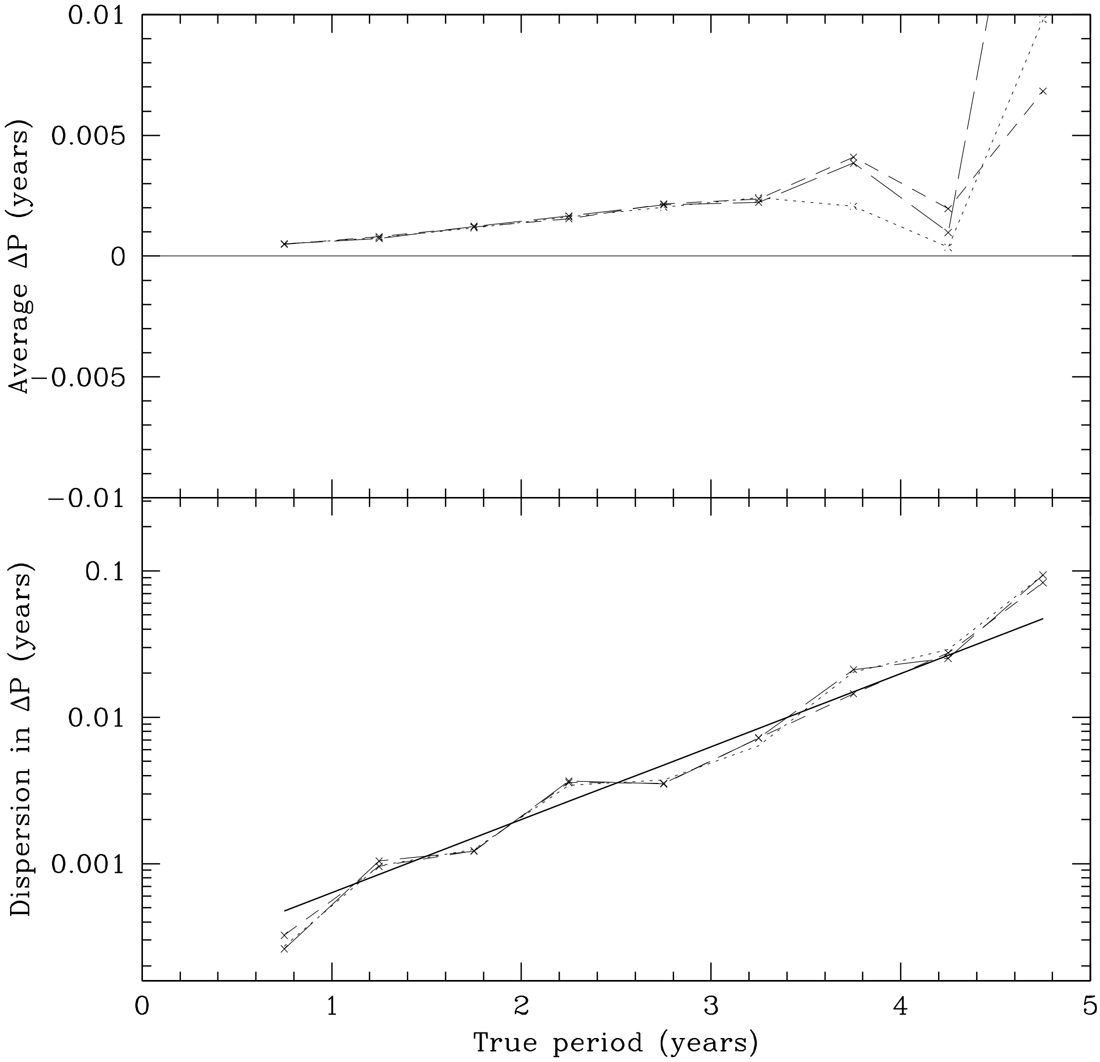

Figure 2: Mean and dispersion of $\Delta P = P_{\rm fitted} - P_{\rm true}$ in half-year bins of true period, computed after iterative 3-σ clipping. The results for the three Solvers (crosses and dashed lines) are very similar. Period residuals tend to have a slight positive bias. The precision of the period measurement, after subtracting this bias, scales with period approximately as $10^{0.5P}$ (solid line in the lower panel).

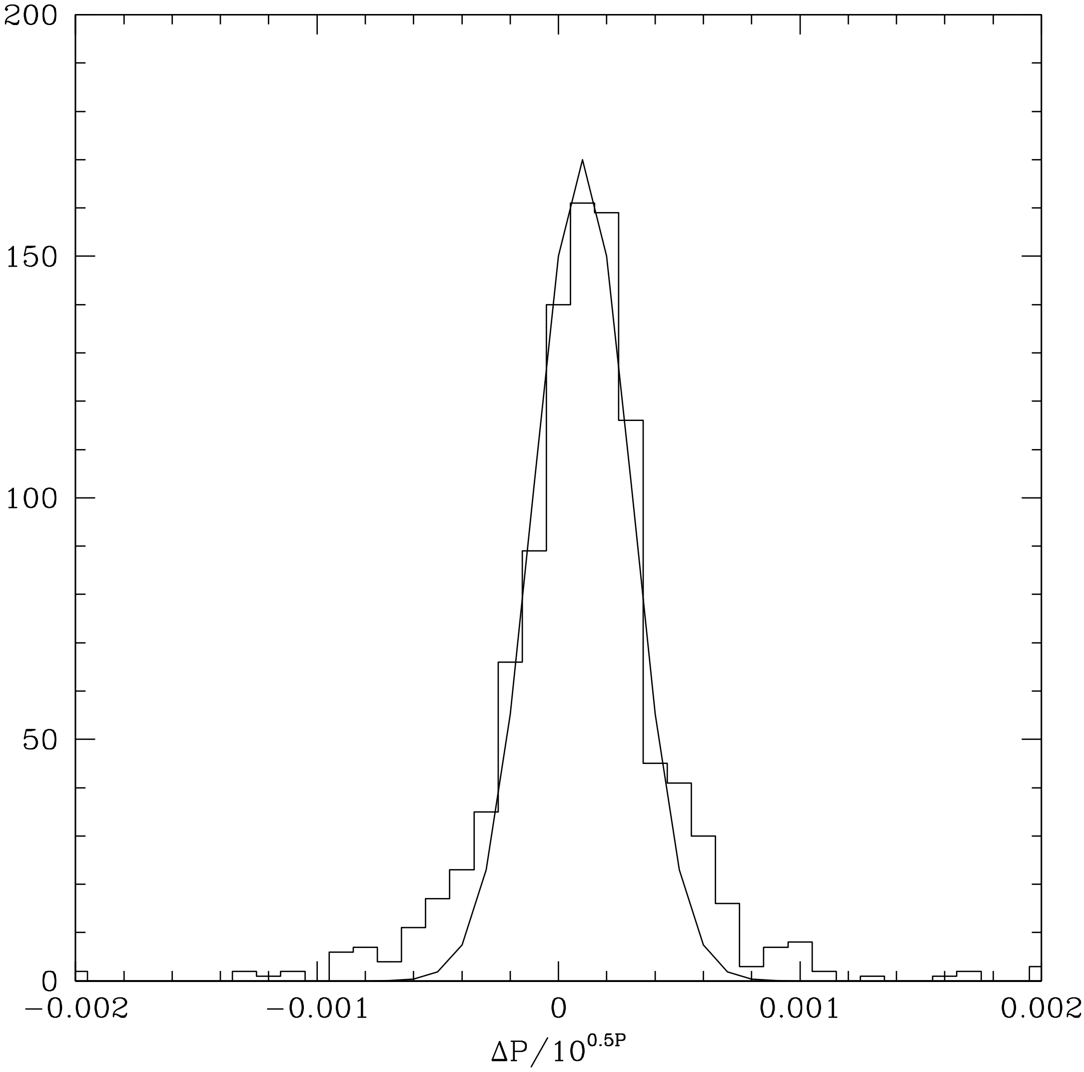

Figure 3: Histogram of the period residuals from the Pourbaix solution scaled by $10^{0.5P}$. A Gaussian with mean 0.0001 and dispersion 0.0002 is shown for comparison. The distribution of period residuals has a non-Gaussian tail, presumably resulting in part from the mixture of simulated planets with a variety of eccentricities and inclinations.

identified from a poor value of the χ^2 of the solution. Note than only 2 outliers are in common between the two sets, indicating some dependence of the solution performance on the initial guess.

Despite the very good initial parameters provided to Jancart, her solution also has three planets with period more than 10% off the correct value, all of which are among the four outliers in Pourbaix's solution. (For the fourth Pourbaix outlier, Jancart obtains a correct solution.) All three planets are in highly eccentric orbits, $e > 0.9$, with $P > 4$ years; two are in orbits with inclination $i > 89$ deg. Note that the fourth outlier in Pourbaix's solution has $P_{\rm true} = 1.2$yr and is not otherwise anomalous, while $P_{\rm fitted} = 0.089$yr.

These results clearly indicate that all three Solvers can reproduce the orbital parameters of the vast majority of the planets simulated with high accuracy. Therefore the T0b test can also be considered very successful. A more detailed analysis of the derived orbital elements and of the few outliers in the solution is outside the scope of the T0 test.

4 Conclusions

After an initial period of alignment of the procedures, formats and definitions, Phase 0 of the GAIA Planet Detection Double Blind Tests has concluded successfully. Three Solvers have independently and efficiently carried out both detection and orbit-fitting solutions on simulated data, and the quality of the results is well in line with what could be expected. Some questions of detail still remain, such as whether a better definition of the starting values for the Sozzetti solution might improve the quality of the orbital fits, or what may be the optimal balance between false positives and failed detections, but these small refinements in no way limit the very good quality of the results. We nonetheless encourage the Solvers to analyze the individual cases where the solution may not have performed quite as well as expected, and understand whether they are in fact in line with statistical expectations or there are areas where the solution process can be profitably enhanced.

5 Suggestions for future tests

As a result of the first-pass comparison of the simulations and solutions obtained by the Solvers, we can suggest a few areas where small improvements could result in more efficient and complete analysis of the results.

> Solvers are encouraged to report physically meaningful results from their solutions, e.g., semi-major axis etc., instead of—or together with—the parameters that result directly from their solution, e.g., Thiele-Innes representations or similar. Although the covariance for the latter may be smaller, the quantities that can be compared meaningfully with the simulations are the former.
>
> Solvers are encouraged to report expected errors for their fit parameters, including—if possible—the full covariance matrix. Although some solution methods may not produce this information as a matter of course, it will be of course critical in the evaluation of solutions based on real data, and therefore it is very desirable that the process of understanding parameter errors be fully tested in double-blind fashion. This process will become even more important in later stages, when the data may include more realistic error (and correlation) estimates based on full sphere solutions.
>
> Solvers are encouraged to either make the full fitted observations available, possibly through the same web interface used by the simulators, or in any event to retain this information for use in the post-mortem. The evaluators can and should identify which characteristics of the observations may lead to the rare imperfect result, such as an incorrect fitted period; afterwards, it is up to the Solvers to carry out an analytical review of their solution to understand where the errors may have arisen, and whether they are avoidable with enhanced methods. For this purpose, it is best to have the full set of fitted observations available as a result of the original fitting process, rather than having to reconstruct it *a posteriori*.