

# Preparation of the Gaia Data Processing: First Astrometric Results

X. Luri, W. O'Mullane, U. Lammers, L. Lindegren, and E. Masana

**Abstract** The Gaia mission [4] of the European Space Agency (ESA) will produce high-precision astrometry for  $10^9$  objects up to 20th magnitude. The volume of data generated (about 150TB of compressed raw data) and the complexity of the relationships between them make the scientific processing a challenging task. This paper presents the basic concepts of the core of the astrometric data reduction, the AGIS system, its present status and the first test results using simulated data.

## 1 Introduction

The Gaia mission [4] of the European Space Agency (ESA) will produce high-precision astrometry for  $10^9$  objects up to 20th magnitude. In addition to astrometry the mission will also provide spectrophotometry and, for a part of the objects, spectra. The volume of data generated (about 150TB of compressed raw data) and the complexity of the relationships between them make the scientific processing a challenging task. This paper presents the basic concepts of the core of the astrometric data reduction, the AGIS system, its present status and the first test results using simulated data.

---

X. Luri, E. Masana

Dept. d'Astronomia i Meteorologia and IEEC-UB-ICC, C/ Martí i Franquès 1, 08028 Barcelona  
e-mail: xluri@am.ub.es

W. OMullane, U. Lammers

ESA / European Space Astronomy Centre, Urbanización de Villafranca del Castillo Avda. de los Castillos s/n, E-28692 Villanueva de la Canada (Madrid)

L. Lindegren

Lund Observatory, Box 43, SE-221 00 Lund (Sweden)

## 2 AGIS

The Gaia astrometric problem can be formulated very simply: which set of astrometric parameters best predict the positions of the star images in the Gaia focal plane at the given observations times? Therefore the derived question: what is the mapping (model, transformation or relation) between astrometric parameters and the position in the focal plane at  $t_{obs}$ ? The main goal of AGIS, the Astrometric Global Iterative Solution, is the determination of the astrometry and the parameters of this mapping. AGIS achieves this through an iterative processing of the Gaia data.

AGIS is just one of many parts of the Gaia processing, a central or core part certainly, but still a part. In the overall design of the Gaia Processing System [2] the Main Database [1] is the central repository of all information.

AGIS has been implemented by the ESAC Gaia team. This implementation has not yet reached its final complexity, but is already running and producing its first results: AGIS has to solve, and provides, four different types of data:

- **Source data:** the astrometric parameters of the sources observed by Gaia, that is  $(\varpi, \alpha, \delta, \mu_\alpha, \mu_\delta, \nu_r)$ . This is the main science product of AGIS.
- **Attitude data:** the 3D orientation of the satellite with respect to the ICRS as a function of time
- **Calibration data:** the calibration of the Gaia instruments (geometry, photometric response, etc.) as a function of time.
- **Global data:** other data globally affecting the astrometry of the objects, and in particular the parameters of the PPN relativistic model used

## 3 The simulated data

AGIS, like the rest of the Gaia data processing software, is being tried and tested using simulated data produced by one of the units of the DPAC. A complex simulation software has been developed allowing the generation of, among other products, realistic simulated Gaia telemetry [3] & [5].

This (simulated) telemetry is first processed through another of the Gaia data processing systems (IDT, Initial Data Treatment) and ingested in the Main Database mentioned above, where it is ready for AGIS processing.

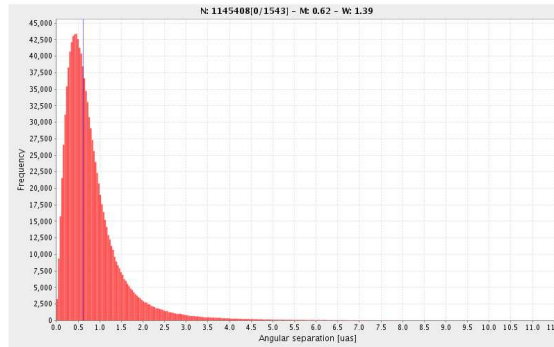
AGIS then takes the satellite data from this database and starts its iterative processing until convergence. Datasets of up to 15 million objects have been produced by the simulator up to now and processed by AGIS.

## 4 Results

The effort of running AGIS tests should not be underestimated. These are not the sort of tests which can be run on a desktop machine in an evening. Rather it requires quite more hardware, time and planning. The machine used for the Gaia tests so far has been purchased in installments by ESA. The initial machine for the first tests in 2005 consisted of twelve nodes each with 6 GB of RAM and two processors (3.6GHz Xeon EM64T). An EMC storage area network (SAN) with 5 TB of disk is attached to the nodes using fiber optic cable and the nodes are also connected via gigabit Ethernet cards in a local area network. AGIS has also been tested, and shown to run as fast, using a cost-effective Rack Server Network Attached Storage device of 6 TB. This cluster was upgraded to 18 nodes in 2006 and a further 4 nodes but with quad core processors were added in 2007.

The very first results from AGIS were published within DPAC early in 2006 Lammers et al. (UL-015) This was with one million sources simulated over just eighteen months of the mission. This behaved very well making a major adjustment to the initial noisy source positions already in the first iteration. Figure 1 shows the source position updates for the fourth iteration of this first published AGIS test, showing that these updates (corrections) to the source positions are already of the order of a few  $\mu\text{as}$  at this early iteration, indicating a good convergence of the solution. This was actually the 373rd run of the AGIS system since work started at the end of 2005. All updates in the cycle were in line with expectations from the simulation and this was a very satisfactory initial result for AGIS.

**Fig. 1** Source position updates for the fourth iteration of the first AGIS test. The fact that these updates are already down to a few  $\mu\text{as}$  indicates a good convergence even at this early stage of the iterative process.



From testing the initial implementation in 2006 a basic confidence in the stability and suitability of the system was gained. Next some more involved test campaigns were designed to identify and fix other potential problems in the system. Each time a test campaign was undertaken an initial test with noise-free inputs was performed to ensure no coding error had introduced some strange behaviour in the system.

The ESAC team has now been busy testing AGIS with various simulation data sets for over three years. Each set has improved AGIS and demonstrated that it can

retrieve or smooth out the sort of effects expected in the Gaia astrometric instrument. Perhaps most importantly, the tests have proven a Global Iterative solution is both viable and useful for the Gaia Astrometry. With the AGIS system as of 2008 running an iteration over  $10^6$  sources in one hour on a modest multiprocessor system it would be possible to process  $10^8$  sources in about three months. We may safely deduce this as the time to completion for AGIS is linear with the number of sources included. Undoubtedly the algorithms will become more complex but the computers will become faster. It is in any case the intention to have a far more powerful machine for the final system. No other approach to the Gaia astrometry has proved viable to the level demonstrated by AGIS and briefly summarised here. The testing is by no means finished and will continue to ensure the best astrometry may be extracted from the Gaia data.

Looking at the future, the actual AGIS usage on mission telemetry will take one hundred times more data and have five times more processing to do, that is, a five hundred scaling factor. On the current 140 GFLOPS machine this scaled up AGIS would take about 1000 days. The 10 TFLOPS machine expected to be available during the mission, however, will be 70 times more powerful than the current system so it could do the scaled up AGIS in 15 days, assuming I/O is dealt with as efficiently as in the current system. Even allowing another factor two for I/O slowdown and problems this would lead to around a 30 day run time for the final AGIS. This is a perfectly acceptable run time. Furthermore, it has been demonstrated that AGIS is capable of reducing Gaia like observations to provide the reference frame for the Gaia data. A novel approach has been employed in the solution of a unique problem. It has been equally difficult and rewarding, and the Gaia flight data is eagerly awaited!

**Acknowledgements** The work presented in this paper is the basis for the ongoing PhD thesis "Implementing the Gaia Astrometric Solution" by W. O'Mullane, under the direction of L. Lindegren and X. Luri, that is near completion and will be presented in early 2009. The simulations used in this work have been done in the Mare Nostrum supercomputer at the Barcelona Supercomputing Center - Centro Nacional de Supercomputaci3n. We wish to thank the DPAC Coordination Unit 2 for providing them. This research has been supported by MCyT under contracts ESP2006-26356-E, ESP2006-13855-C02-01 and AYA2006-15623-C02-02.

## References

1. Hernandez, J., ter Linden, M., GAIA-C1-SP-ESAC-JH-004, "Main Database Software Requirements Specification"
2. Hoar, J., GAIA-C1-TN-ESAC-JSH-010, "Gaia Science Operations Scenario"
3. Masana, E. et al., poster, Reunin cientfica SEA 2008 (this book), "El simulador de la misi3n Gaia: diseo y resultados"
4. Perryman, M. et al., A&A 369, 339P (2001), "GAIA: Composition, formation and evolution of the Galaxy"
5. Portell, J., charla, Reunin cientfica SEA 2008 (this book), "Massive parallelization of a space mission simulator and data processing system"