Highlights on Spanish Astrophysics IX, Proceedings of the XII Scientific Meeting of the Spanish Astronomical Society held on July 18 – 22, 2016, in Bilbao, Spain. S. Arribas, A. Alonso-Herrero, F. Figueras, C. Hernández-Monteagudo, A. Sánchez-Lavega, S. Pérez-Hoyos (eds.)

Validation of *Gaia* data

C. Fabricius[1], F. Arenou[2], C. Jordi[1], F. Figueras[1], S. Soria[1], M. Romero-Gómez[1], and J. M. Carrasco[1]

[1] Institut de Ciències del Cosmos, Universitat de Barcelona (IEEC-UB), C. Martí i Franquès 1, 08028 Barcelona
[2] GEPI, Observatoire de Paris, PSL Research University, CNRS, Univ. Paris Diderot, Sorbonne Paris Cité, 5 Place Jules Janssen, 92190 Meudon, France

Abstract

In preparation of the first *Gaia* data release the data has undergone a large number of tests, in order to check the validity of positions, parallaxes, proper motions, and magnitudes. Tests carried out by the *Gaia* consortium – with participation of the University of Barcelona – include independent error estimations from the negative parallax tail, comparisons with existing catalogues, internal consistency of open clusters, distributions of the various quantities, etc. This validation has led to final conclusions on which *Gaia* solutions to accept for publication.

1 Introduction

The first release of data from the *Gaia* mission, *Gaia* DR1, is due a few weeks after the XII Scientific SEA Meeting. It covers only 14 months of observations, and therefore only gives positions and G magnitudes for the majority of the 1142 million sources. However for a subset of two million sources, the *Tycho-Gaia* astrometric solution (TGAS), where *Tycho-2*[4] and Hipparcos[8] positions are used as prior information, it also gives proper motions and parallaxes [2]. Before the release, the data have undergone extensive validations, to check that the published uncertainties are realistic, to check any systematic errors, and to check the catalogue completeness. This validation process involved a large group of people and is described in detail in [1]. Here, we will only discuss some main points and some illustrative examples.

2 Astrometric data

The original TGAS proposal [6] promises parallaxes with 0.2–0.3 mas uncertainties for HIPPARCOS stars and about twice that for *Tycho-2* stars. The fulfilment of this promise was checked in two ways, partly from the negative parallax tail and partly using a set of QSOs.

The first observation is that the negative tail for TGAS parallaxes is much shorter than for HIPPARCOS, cf. [5]. This already is a good result. A deconvolution of the tail is presented in [1] for ranges of standard uncertainty. It shows a very good agreement between the formal and reconstructed uncertainty as long as the uncertainty is below 1 mas. The same analysis showed disagreements for larger error estimates, and these solutions were therefore excluded from *Gaia* DR1.

A special astrometric solution was made for QSOs [5], where the negligible proper motions for these objects were used as prior information. A test is shown in [1] where median QSO parallaxes are calculated in regions of radius two degrees across the sky. The conclusion is that systematic parallax errors are below 0.3 mas over 90% of the sky, and that there is a zero-point offset of -0.04 mas. Possible reasons for these errors are discussed in [5].

TGAS proper motions for the general *Tycho-2* stars have uncertainties similar in size to the ones given in *Tycho-2*, but are on a much more solid foundation. They are based on two space missions 24 years apart and avoid the complications of incorporating the Astrographic Catalogue as was done for *Tycho-2*. Comparisons between the two sets of proper motions are presented in both [1] and [5] and the differences show clearly a pattern coinciding with the zones of the Astrographic Catalogue, thus confirming the better quality of TGAS.

3 Photometric data

The photometric data in *Gaia* DR1 consist basically of fluxes and magnitudes in the broad G band, defined by the average response of *Gaia* detectors and optical elements. The validation is described in [1] and [3], and apart from internal consistency tests, consists to a large part in comparisons with different photometric catalogues. Such comparisons are tricky, partly because they involve transformations between photometric systems, and partly because each catalogue only covers part of the magnitude interval. The general conclusion is that no trend ascribable to *Gaia* is found.

Photometric consistency checks on a pre-release catalogue showed interesting features. In particular, sources with no more than 10 CCD transits, i.e. much below the average, showed a very peculiar colour distribution, not seen for more frequently observed sources, and therefore it was concluded that the very low number of transits must be due to a technical processing issue, and this set of sources were therefore not included in this release.

The exclusion of these sources also helped remove the vast majority of a puzzling group of sources, viz. very faint TGAS stars. By construction, TGAS stars are not expected to be much fainter than 13 mag, yet there were several thousands such sources. After the final filtering, we are still left with a few hundreds, which are mainly either misidentified or spurious *Tycho-2* stars [1].

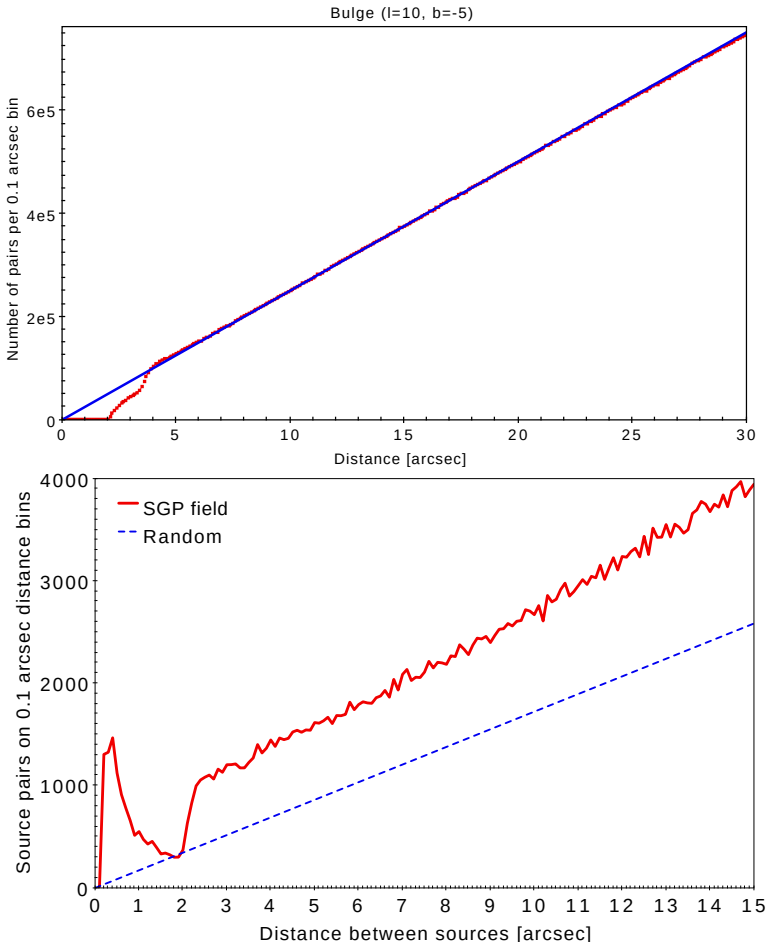

Figure 1: Distribution of distances between pairs of stars. *Top:* in a dense area of the bulge; *bottom:* in a sparse area near the south Galactic pole. The red points show the actual distribution in *Gaia* DR1, while the blue lines correspond to a random sky distribution.

4 Catalogue completeness

The catalogue completeness was evaluated in several ways, e.g. comparing with HST images or studying the magnitude distribution of globular clusters. We will here just mention a few internal tests. The first was checking for duplicated sources, where we would have two catalogue entries based on disjoint sets of observations, but still yielding nearly identical positions. In these cases, one solution was excluded from the release.

Another simple completeness test is to look at the distribution of distances between pairs of sources. If sources are randomly distributed over some area, the number of pairs in a given, small distance range is proportional to the distance with a known factor. If the distribution has several components, each with different density, the factor will typically be higher. On top we must add a contribution from binaries, which will show up at modest

distances. Fig. 1 shows examples from a very dense and a very sparse sky region. The dense field (top panel) is small and fairly uniform, but we clearly miss many pairs below 4 arcsec. This is a result of the photometric processing where colour could not be derived in crowded areas in this early stage of the mission. As a consequence, about a third of the sources – predominantly the fainter – in such fields did not reach *Gaia* DR1, but must wait till later releases. The sparse field (bottom panel) is less homogeneous, and due to the lack of crowding it maintains a good completeness down to about 2 arcsec separation, and even displays a small population of binaries.

5 Conclusions

Gaia DR1 is a first, preliminary release and still incomplete both in terms of source contents and in terms of parameters available for each source. Still, it provides an important set of proper motions and parallaxes much superior to existing catalogues. Parallaxes have, however, local systematic errors of order 0.3 mas, so correlations must be taken into account when forming mean parallaxes. Proper motions are of very high precision for the HIPPARCOS subset, while for other *Tycho-2* stars it is similar to *Tycho-2*, but with a much better accuracy.

Acknowledgments

This work has made use of results from the European Space Agency (ESA) space mission Gaia, the data from which were processed by the Gaia Data Processing and Analysis Consortium (DPAC), of which the authors are members. This work has financially been supported by: the Centre National dEtudes Spatiales (CNES), the MINECO (Spanish Ministry of Economy) - FEDER through grant ESP2014-55996-C2-1-R, MDM-2014-0369 of ICCUB (Unidad de Excelencia María de Maeztu) and the European Communitys Seventh Framework Programme (FP7/2007-2013) under grant agreement GENIUS FP7 - 606740. We have in this work made use of the software TOPCAT [7].

References

[1] Arenou, F., Luri, X., Babusiaux, C., et al. 2016, A&A, special *Gaia* volume

[2] Brown, A. G. A., Vallenari, A., et al. (*Gaia* Collaboration) 2016, A&A, special *Gaia* volume

[3] Evans, D. W., Riello, M., De Angeli, F., et al. 2016, A&A, special *Gaia* volume

[4] Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, A&A, 355, L27

[5] Lindegren, L., Lammers, U., Bastian, U., et al. 2016, A&A, special *Gaia* volume

[6] Michalik, D., Lindegren, L., & Hobbs, D. 2015, A&A, 574, A115

[7] Taylor, M. B. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 29

[8] van Leeuwen, F. 2007, Astrophysics and Space Science Library. Vol. 350 (Springer)