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.)

Two years of daily Gaia data processing

J. Portell, C. Fabricius, N. Garralda, J. J. González-Vidal, J. Castañeda, M. Clotet, and J. Torra

Dept. Física Quàntica i Astrofísica, Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, E08028 Barcelona, Spain

Abstract

Gaia has been in the nominal operations phase since July 2014. It has observed millions of astronomical sources every day, which have been processed by the Initial Data Treatment system (IDT), coordinated by the IEEC Gaia team at the University of Barcelona. This system feeds the whole Data Processing and Analysis Consortium of Gaia (DPAC), ultimately leading to the Gaia data releases. In this paper we briefly describe the main IDT algorithms and monitoring tools, and present some results obtained during the first two years of nominal operations, covering both technical and scientific aspects.

1 Introduction

After half a year of commissioning, the Gaia mission [5] of the European Space Agency entered nominal operations on July 2014. During two years it has been observing, on average, more than 50 million transits of astronomical sources every day, meaning more than 60 GB of raw data that must be timely processed. This is a requirement both on the technical side, to avoid accumulating all mission data for later offline processing, and on the scientific and mission health side, to continuously diagnose the spacecraft while obtaining preliminary science outputs. The Initial Data Treatment (IDT), coordinated by the IEEC-UB Gaia team in Barcelona, is the software system carrying out these near real-time duties. It must reconstruct self-contained scientific measurements from a variety of instrumental outputs, combining sample data with time and position information and measurement configurations. These raw measurements are then processed to generate intermediate-level data products, including image centroids, fluxes in several bands, sky positions, preliminary proper motions and quality indicators. The spacecraft attitude must also be reconstructed with enough accuracy to allow a reliable cross-matching of observations against a reference star catalogue. All these outputs have been essential for the execution of the main data reduction systems that have led to the first Gaia data release [4].

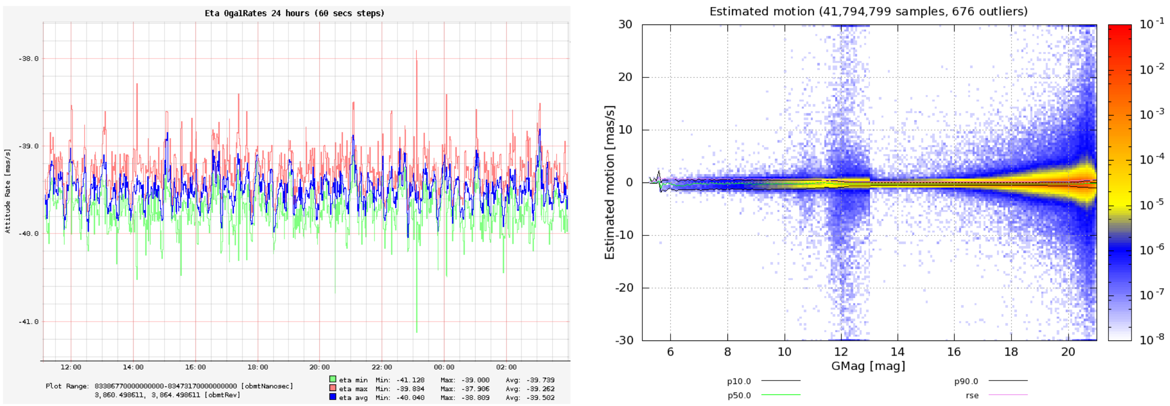

Figure 1: Left panel shows the Gaia spin rate difference vs. the nominal spin rate as determined by IDT from its first attitude refinement algorithm (OGA1), which is accurate to 1 mas/s thus making possible good estimations of object motions (right panel).

2 Overview of the IDT system

IDT [3] is one of the first data processing elements of the Gaia ground segment, most of which are implemented and handled by the Data Processing and Analysis Consortium (DPAC [4]). IDT is actually the first large data processing system generating scientific outputs, furthermore with a reasonably good quality — specially considering its prompt execution. Typically, the delay between an on-board measurement and its corresponding IDT output ranges from about two hours up to one day, reaching almost two weeks in the worst of the cases. IDT must detect all incoming inputs and trigger the execution of the adequate algorithms depending on the data type. It is thus a *data-driven* system, rather than a schedule-driven system based on some given schedule of the satellite downlink.

There are some ancillary data types that indicate the on-board configuration used for the astrometric, photometric and spectroscopic measurements. Such ancillary data is combined with the main measurements (the pixels from the CCD focal plane) to generate the so-called *Observation* records, with one record per object transit through the focal plane. These records are the main input to the downstream DPAC systems, including the core astrometric solution (AGIS [7]) and the cyclic re-processing of raw astrometric data (IDU [1]). The satellite attitude is also determined by IDT with an accuracy of about 100 mas, to allow for a preliminary cross-matching of star transits with a reference catalogue (later refined in IDU [2]). Finally, raw observation records are processed using the latest instrumental calibrations available (some of which are also determined by IDU [1]), leading to image parameters such as centroids or fluxes in different broad bands. These are later refined by IDU as well [1]. It is also worth mentioning that IDT determines variations in the basic angle between the two telescopes with μas accuracy, which is required not only to monitor the spacecraft but also to provide better inputs to the astrometric data reduction systems.

The Gaia team at IEEC-UB has coordinated the implementation, test and operation of IDT since its conception in 2006. It has also designed and implemented its overall execution

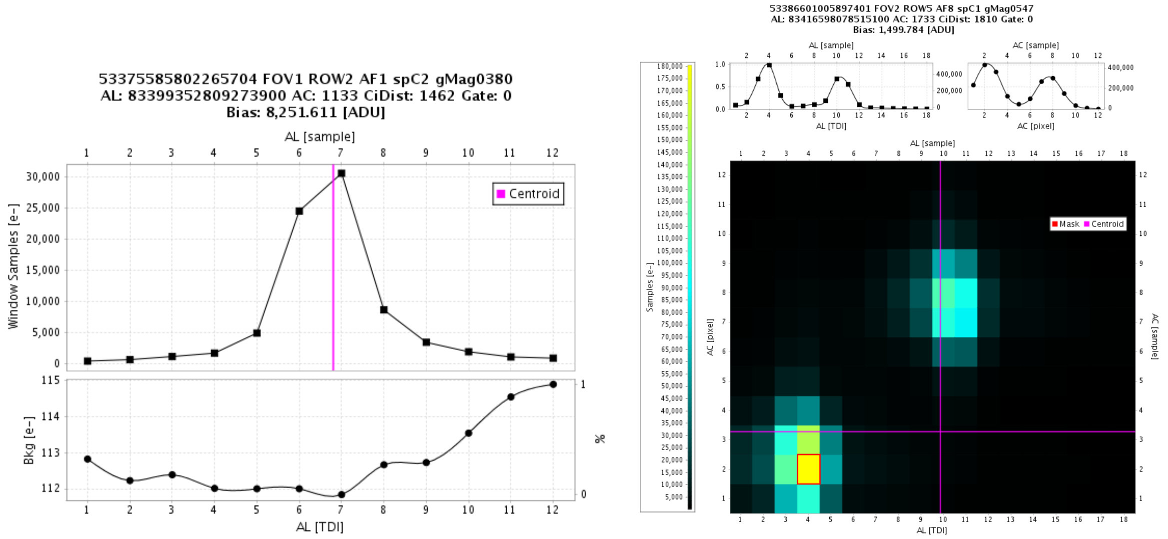

Figure 2: Two examples of raw measurements processed by IDT. Left: 15^{mag} star. Right: a double star with separation 1.1 arcsec with the fainter component being 12.5^{mag}.

framework, as well as the raw data reconstruction algorithms, the first attitude refinement algorithm, the preliminary cross-matching (including blacklisting of on-board spurious detections [6]) and the exhaustive near-realtime monitoring system.

3 Scientific and technical results

One of the most important elements of IDT is the determination of the satellite attitude. It must be accurate enough for the preliminary cross-matching, but also stable enough to allow for a preliminary estimation of the proper motions of the objects being observed. The latter is specially interesting for one of the downstream systems, the short-term Solar System Objects (SSO) processing of Coordination Unit 4 (CU4 [5]). Specifically, IDT provides them information to determine if a given transit is an SSO candidate. Figure 1 illustrates the along-scan attitude rates determined by IDT (left panel), whereas right panel shows the distribution of estimated motions. The latter have been externally validated by the SSO team revealing an excellent motion estimation by IDT in the vast majority of the cases. It is worth mentioning that CU4 has also validated the sky coordinates determined by IDT, comparing them against ICRF sources and assessing that the accuracy achieved is typically better than 200 mas. This is improved by a factor 1000 in later processing stages of DPAC.

The heaviest computing load of IDT corresponds to the raw data reconstruction and, specially, the image parameters determination. Figure 2 illustrates some of the raw data handled by IDT. In the left panel we can see an otherwise typical *window* or measurement, similar to the ~500 million measurements processed by IDT daily. Its upper part shows the values of the raw samples, each corresponding to the sum of 12 pixels in the across-scan direction. It is thus the *Line Spread Function* (LSF) corresponding to that specific

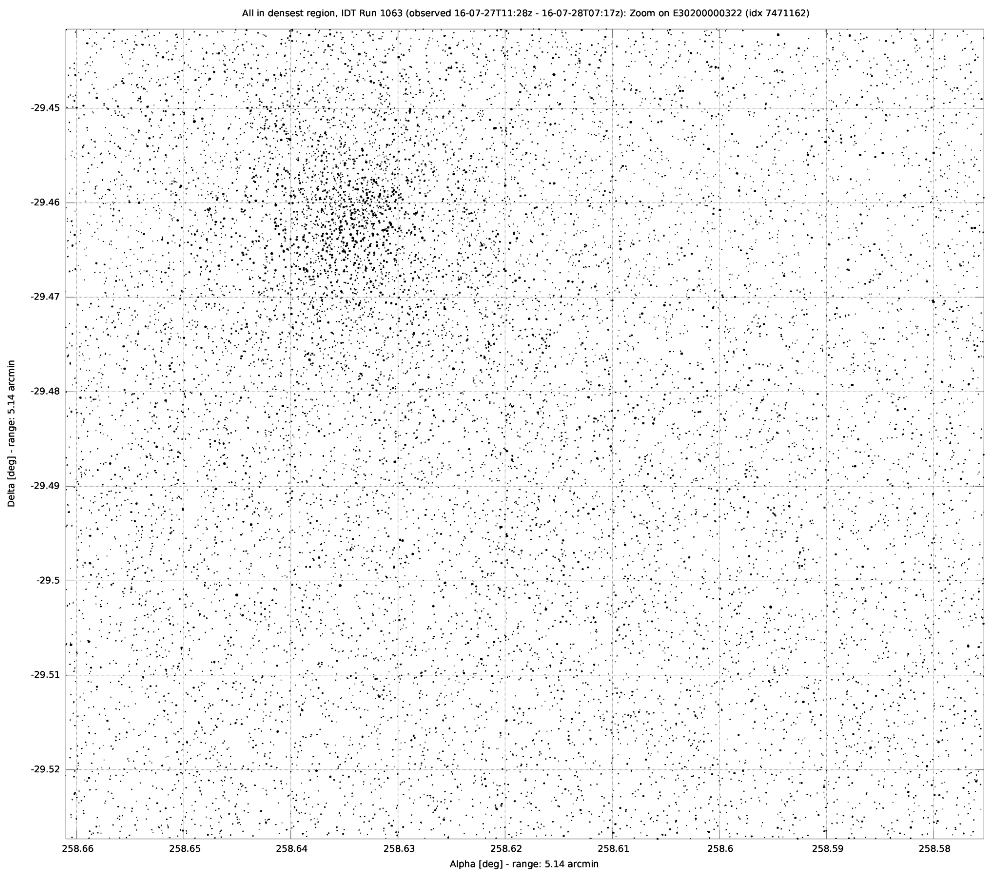

Figure 3: NGC 6304 seen by Gaia and processed by IDT.

measurement. The bottom part shows the astrophysical and instrumental background level for this measurement, which presents small variations in this specific case. The right panel of Fig. 2 shows a more difficult case, corresponding to a binary object (either "true" or visual), bright enough to trigger the two-dimensional sampling on-board. In this case, since IDT only focuses on single objects, the centroiding could not be correctly determined. Image parameters are obtained by fitting the adequate LSF or PSF model to the samples [1]. It leads to a centroiding precision from $1/10^{th}$ to $1/1000^{th}$ of a pixel, depending on the brightness of the star. That is, even with this preliminary processing we can reach a precision of up to 0.1 mas for individual measurements.

This kind of plots are automatically generated by the IDT monitoring system on a tiny set of "interesting" measurements every day. Figure 3 shows another kind of automatic monitoring plots. In this case, IDT determines the densest sky region processed each day, as well as the region leading to most "unmatched" transits (that is, measurements for which a catalogue counterpart could not be found). In this example we can actually see the NGC 6304 cluster. The plot, with larger dots for brighter sources, also gives an idea of the angular resolution of Gaia. The separation between two detections can be as small as ~500 mas.

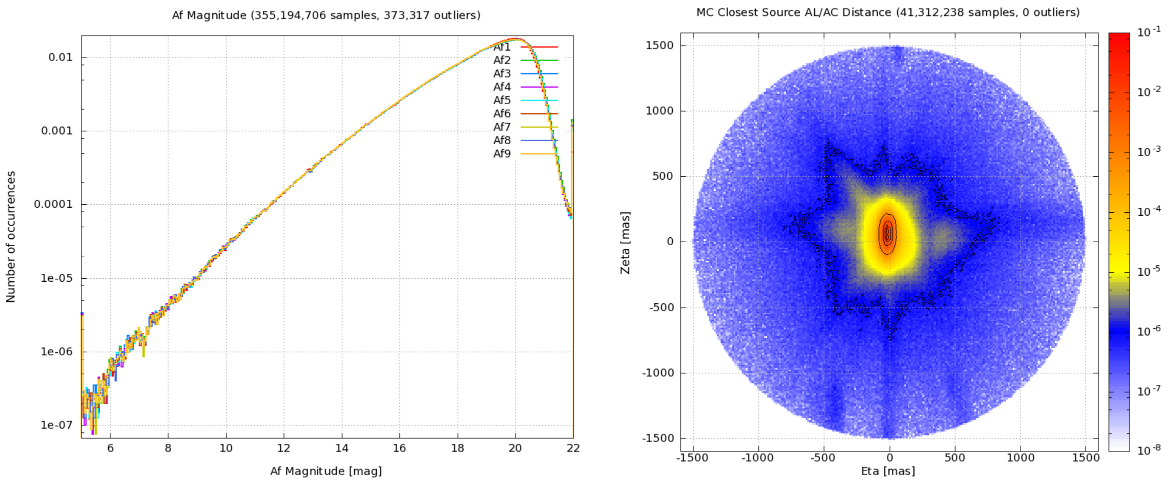

Figure 4: Left: magnitude distribution in the *Gaia* band for one given day and the 9 strips of the on-board astrometric field CCDs (AF). Right: distribution of cross-match distances between detections and catalogue sources.

Figure 4 illustrates some further diagnostics and quality of IDT outputs. The left panel shows that the magnitude distribution is as expected according to the Galaxy model and the Gaia detection limit, whereas the right panel shows that the match distance is better than 200 mas in almost all cases. Note that the search radius in IDT is 1.5 arcsec. Finally, Fig. 5 shows the density of star transits observed by Gaia and processed by IDT so far, covering about 27 months of nominal operations. The total number of transits processed exceeds 57 billion, each of which includes 10 astrometric measurements and two spectro-photometric measurements. About 10% of these also include three spectroscopic measurements.

IDT runs at the European Space Astronomy Centre (ESAC) near Madrid. About 20 computing nodes are used by the IDT software, as well as a large database node with 1 TB memory. These computing resources are more than enough for IDT, even during heavy days such as when Gaia scans the Galactic Plane tangentially or when recovering from a downtime for ground systems maintenance. The peak performance of IDT exceeds 200 million transits per day, which means that the system is typically idle several hours every day.

4 Conclusions

The Initial Data Treatment (IDT) is a near-realtime high-performance data processing system running on the raw data received daily from the Gaia spacecraft. After slightly more than two years of nominal operations, IDT has processed over 57 billion star transits without major problems. Scientific results confirm that Gaia is able to provide an excellent scientific return even with such a daily and preliminary data reduction. The satellite attitude and the sky coordinates of astronomical objects are determined by IDT with an accuracy of about 100 mas, and even proper motions are estimated with accuracies typically better

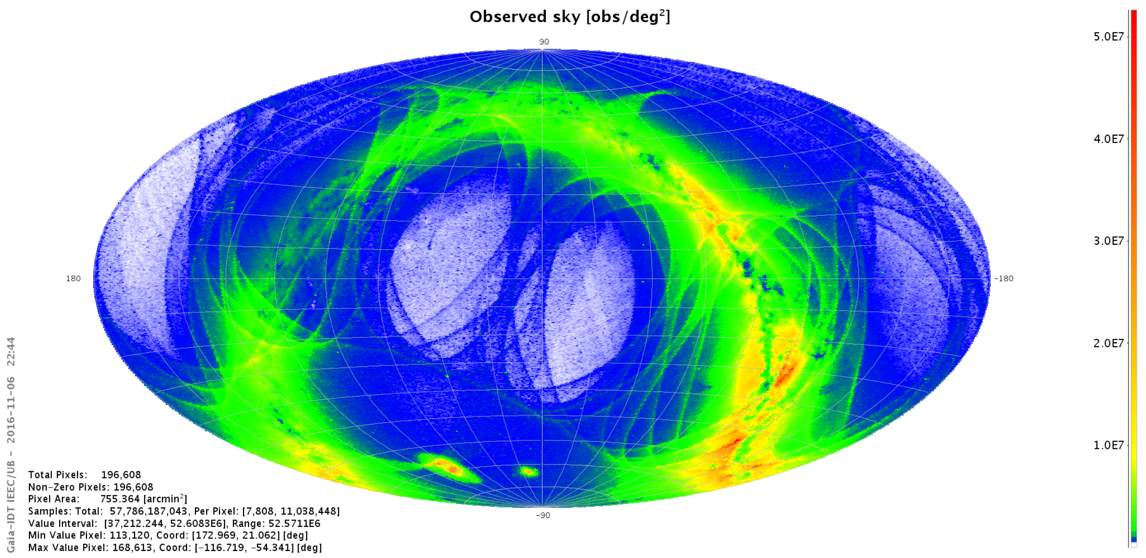

Figure 5: Mollweide projection (equatorial coordinates) of processed transits density, showing the combination of the actual star density with the satellite scanning law.

than 1 mas/s. The adequate execution and outputs of IDT are monitored daily through its automated monitoring system. The system keeps running stably and without problems, feeding downstream systems to make possible the future Gaia data releases.

Acknowledgments

This work was supported by the MINECO (Spanish Ministry of Economy) - FEDER through grant ESP2014-55996-C2-1-R and MDM-2014-0369 of ICCUB (Unidad de Excelencia 'María de Maeztu').

References

[1] Castañeda, J., Fabricius, C., Portell, J., Garralda, N., González-Vidal, J.J., Clotet, M. & Torra, J. 2016, in Highlights of Spanish Astrophysics IX (this volume).

[2] Clotet, M., González-Vidal, J.J., Castañeda, J., Garralda, N., Portell, J., Fabricius, C. & Torra, J. 2016, in Highlights of Spanish Astrophysics IX (this volume).

[3] Fabricius, C., Bastian, U., Portell, J., et al. 2016, A&A, in press.

[4] Gaia Collaboration (Brown et al.) 2016, A&A, in press.

[5] Gaia Collaboration (Prusti et al.) 2016, A&A, in press.

[6] Garralda, N., Fabricius, C., Castañeda, J., Portell, J., Clotet, M., González-Vidal, J.J. & Torra, J. 2016, in Highlights of Spanish Astrophysics IX (this volume)

[7] Lindegren, L., Lammers, U., Bastian, U., et al. 2016, A&A, in press.