Draft version March 11, 2023
Typeset using LaTeX **twocolumn** style in AASTeX631

Gaia Parallax Distances

Michael Tauraso

ABSTRACT

With the unique and revolutionary measurements of the Gaia era affecting many areas of the study of the structure and history of the Milky Way, this paper looks at the parallax distances and explores the high level causes of spurious parallaxes in Gaia DR3, as well as exploring some approaches for using these parallaxes and the astrometric solution as a whole to measure distance.

1. INTRODUCTION

ESA's Gaia mission has been a boon to the study of the Milky Way. At time of writing Data Release 3 (DR3) has a publically searchable catalog of some 1.8 billion sources with 88% of those having a high quality astrometric solution, and some $Y\%$ containing a published astrometric parallax with. In comparison with Gaia's predecessor sattellite, Hipparcos, this is an increase of $10x$??? in the number of sources, and a $x\%$?? increase in precision.

The detailed study of the structure of the milky way benefits greatly from the enhanced precision of these parallax measures; however Gaia's astrometric parallaxes contain an estimated ($y\%$???) spurious astrometric solutions including ($z\%$???) negative valued parallaxes. The interpretation of these parallax values as distances is the subject of some study, and their presence is in some ways a commitment to a certain predictability in the structure of astrometric data processing, which has independent scientific value outside of the accuracy of the data themselves.

Section 2 will sketch the process by which Gaia parallaxes are determined, and in Section 3 I will discuss some of the ways this process can lead to a erronious (or even negative) parallax. Section 4 will characterize the parallax values published in DR3, and section 5 will sketch and compare efforts to derive distances from DR3 parallaxes. Section 6 will discuss improvements expected in Gaia's upcoming data release (DR4), as well as comparing DR3 with the prior major data release (DR2). Finally, Section 7 will conclude by summarizing an unorthodox use of the astrometric solution which are enabled by Gaia's data processing discipline.

2. HOW GAIA DOES PARALLAX

The standard introductory treatment of parallax as the apex angle of a giant triangle subtended by two ends of the earth's orbit and a distant star, while pedagogically interesting and trigonometrically correct, pales in comparison to the complexity measuring nearly 2 billion parallaxes on a modern space telescope. The Gaia spacecraft sits in a Lissajous orbit at the Earth-Sun L2 Lagrange point. Gaia's two telescopes, separated by a basic angle shown in figure 1, scan the sky as the satellite rotates. As the telescope rotates, sources of light across the universe shine on to CCD detectors for each telescope as shown by the yellow star in figure 1. The times of these transits and the response of the CCDs are recorded.

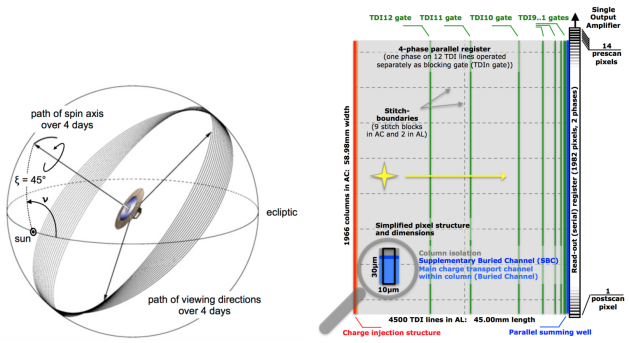

Figure 1. Gaia Scan illustration (left) and Gaia CCD schematic (right) (Collaboration 2016)

The first steps of this analysis are focused on constructing a frame of reference for the astronomic observations and accurately placing the spacecraft and its instruments within that reference frame. Well known astronomic sources with predictable behavior are identified algorithmically and used as points of reference to calibrate detailed mathematial models of the spacecraft's location over time. Once this model of the spacecraft is calibrated, it is used to convert timed transit events to sky locations and times where a source was observed. The time that a source passes by each telescope is critical, because the time that each source appears and leaves the field of view of the spacecraft is an additional piece of information that allows its sky location to be inferred with greater accuracy

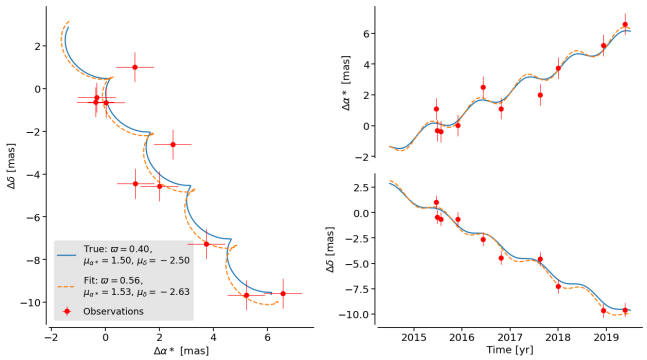

Figure 2. A simplified but representative astrometric curve fit. Simulated star location is in blue, simulated measurements are in red, and the fit curve is orange. The error bars are of representative size to Gaia's data. Actual and fit parameters for proper motion and Parallax are reported on the plot. Simulation was generated by author with code from Luri et al. (2018)

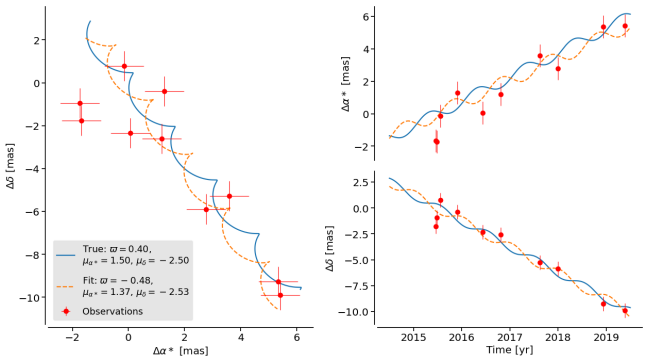

Figure 3. Same as figure 2, except the fit is to a negative parallax value. Note the orange curve (fit) differs from the blue (actual) by a phase of π. Figure generated from code in Luri et al. (2018)

These observations must then be grouped algorithmically, such that each group of observations is identified as being from the same source in the night sky. Parallaxes then flow from a least-squares fit of a linear model to each group of observations identified with a particular source. A simplified and simulated view of such a curve fit is shown in Figure **??**. Each gaia source includes around 200 observations, and the published astrometric parameters, errors and correlations flow directly from this curve fit. The "happy path" for this pipeline is the 5-parameter fit, so called because the 5 astrometric parameters Right Ascension (α), Declination (δ), Right Ascension Proper Motion (μ_α), Declination Proper Motion (μ_δ) and Parallax ($\bar{\omega}$) are the output of the least squares fit.(Lindegren et al. 2021)

There can be issues with the veracity of this fit though. Gaia's CCD instrument and telescope have non-linear response to different frequences of light, and it is well known that objects in the sky are not monochromatic. These two effects together affect the calibration of the instrument, because there are chromatic effects that affect where, when, and how much a given source is recorded on the CCD. Gaia corrects for these effects using a model that takes as input a single color, reported in the published data as `nu_eff_used_in_astrometry`. In the case where this color can be determined from spectral observations, and the resulting 5 parameter fit is of sufficiently high quality, the 5-parameter fit is reported.

In the case that the 5-parameter fit is not high enough quality to be reported, Gaia Astrometry falls back to either a 6 parameter fit or a 2 parameter fit. The 6 parameter fit treats `nu_eff_used_in_astrometry` as an additional unknown taking part of the least squares fit. The value found by the curve fit is reported in the database as `pseudocolor`. Typically very bright and very dim sources require this treatement for different reasons. Dim sources because there is not enough light to determine their true color. Bright sources end up having a 6-parameter fit because their 5-parameter fit is uncertain as their light is washing out much of the positional precision that Gaia would otherwise have. 2-parameter fits (only α and δ) are reported when neither 5-parameter nor 6-parameter fits reach the desired level of quality.

3. WHAT CAN GO WRONG?

The process of source identification is possibly the most error-prone step in the entire astrometric pipeline, involving both components on-spacecraft that initially identify sources and a large amount of earthbound data analysis.

In addition to source identification issues, there are astrophysical phenomena that can cause a correctly identified source to not fit the linear 5 parameter model. Some of these have a somewhat mean-reverting property to linear motion such as binary systems, gravitationally lensed sources, and sources with dark companions. Others, such as stellar close encounters, and exceptionally fast sources have motion that simply diverges from the underlying model. These types of systems can cause a low quality or even a spurious curve fit depending on the magnitude of the effect and the timing of observations.

Finally, even if source identification works and the star itself does move linearly over the observation timescale, you can still have a curve fit that is simply wrong. Many negative parallaxes fall into this category. Figure 3 shows a bad fit for a simulated star, yielding a negative parallax.

4. WHAT DATA IS IN DR3?

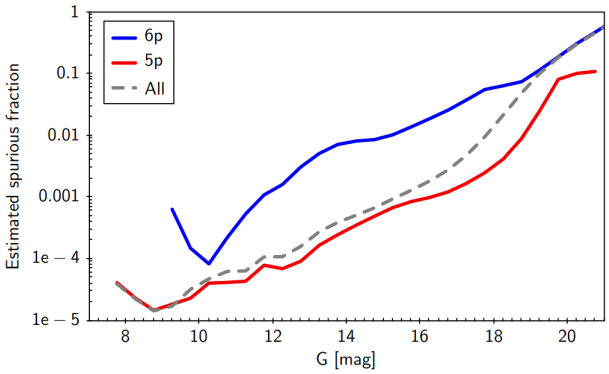

Figure 4. Fraction of Spurious Astrometric solutions for 5 and 6-parameter solutions in EDR3(Fabricius et al. 2021)

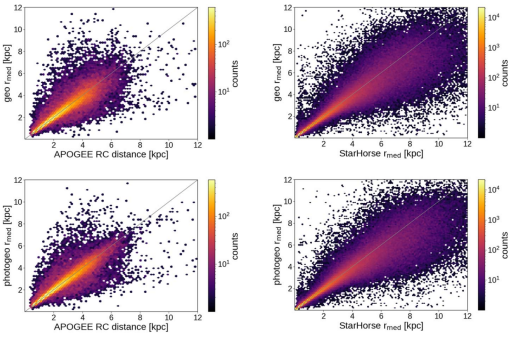

Figure 6. Comparison of Bailer Jones distances to other methods (Bailer-Jones et al. 2021)

5. REASONING DISTANCES FROM PARALLAX

6. IMPROVEMENT EFFORTS

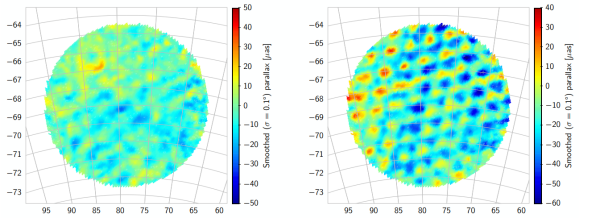

Fig. 14. Smoothed maps of parallaxes in the LMC area, visualising small-scale systematics (the 'checkered pattern') in *Gaia* EDR3 and DR2. *Left:* Smoothed parallaxes in EDR3 for sources in the magnitude range G = 16–18 (median G = 17.4), kinematically selected as probable members of the system (see Appendix B in Lindegren et al. 2020 for details). *Right:* Smoothed parallaxes in DR2 for the same sample of sources. Both maps were smoothed using a Gaussian kernel with standard deviation 0.1°. While the sample includes about 730 000 sources within 5° radius of the adopted centre, only smoothed points within a radius of 4.5° are shown to avoid unwanted edge effects. Comparison between the two diagrams is facilitated by the use of the same colour scale, only shifted by 10 μas to compensate for the mean difference in parallax between DR2 and EDR3.

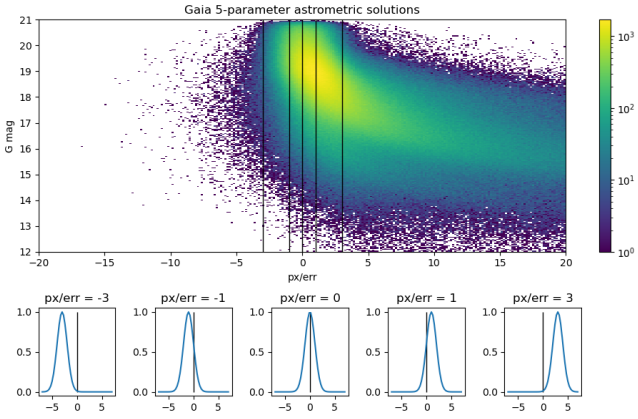

Figure 5. Heatmap of Gaia DR3 5 parameter solutions by `parallax_over_error` measure and G magnitude. Lower plots show example normal distributions for parallax measurements implied by selected `parallax_over_error` values. These same selected values are marked on the main plot by the black vertical lines. Data from Collaboration et al. (2022)

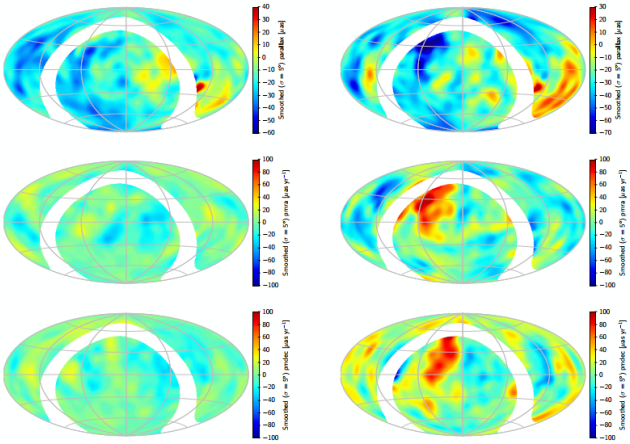

Figure 7. Comparison of DR2 and EDR3 from Lindegren et al. (2021)

7. UNORTHODOX USE OF THE ASTROMETRIC SOLUTION

REFERENCES

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, AJ, 161, 147, doi: 10.3847/1538-3881/abd806

Collaboration, G. 2016, doi: 10.1051/0004-6361/201629272

Collaboration, G., Vallenari, A., Brown, A. G. A., et al. 2022, doi: 10.48550/arXiv.2208.00211

Fabricius, C., Luri, X., Arenou, F., et al. 2021, A&A, 649, A5, doi: 10.1051/0004-6361/202039834

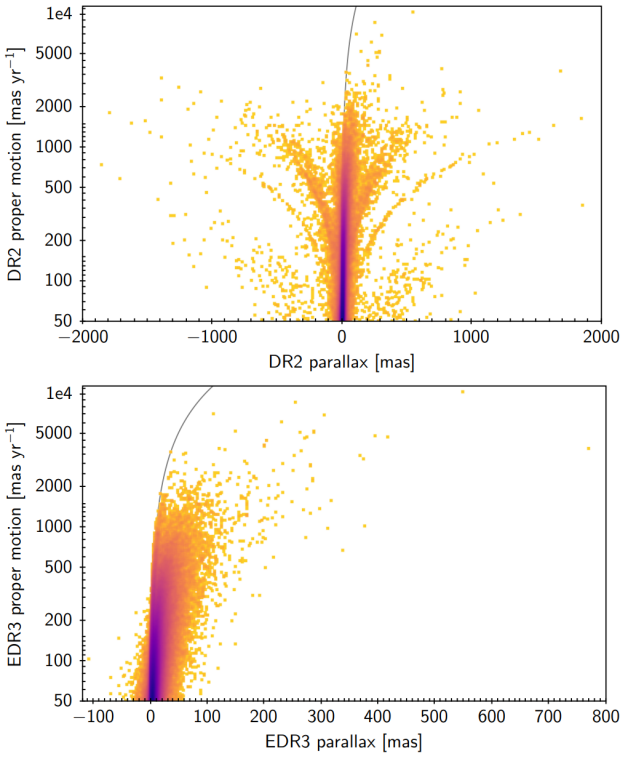

Fig. 10. Proper motion versus parallax for large proper motions. *Top*: in *Gaia* DR2. *Bottom*: in *Gaia* EDR3. The grey line shows the locus of tangential velocity 500 km s^{-1}.

Figure 8. Figure from Fabricius et al. (2021)

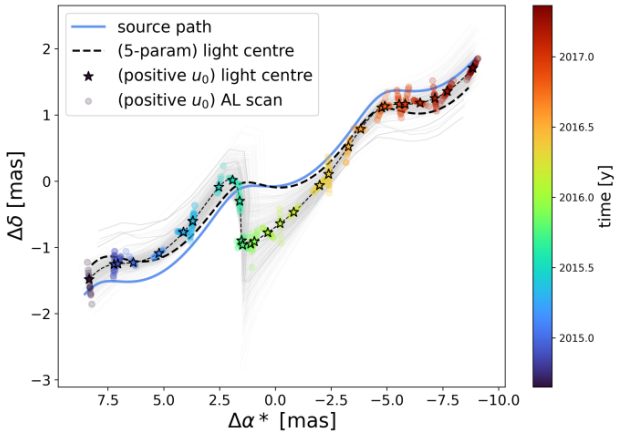

Figure 9. Figure from Jabłońska et al. (2022)

Jabłońska, M., Wyrzykowski, Ł., Rybicki, K. A., et al. 2022, A&A, 666, L16, doi: 10.1051/0004-6361/202244656

Lindegren, L., Klioner, S. A., Hernández, J., et al. 2021, A&A, 649, A2, doi: 10.1051/0004-6361/202039709

Luri, X., Brown, A. G. A., Sarro, L. M., et al. 2018, A&A, 616, A9, doi: 10.1051/0004-6361/201832964

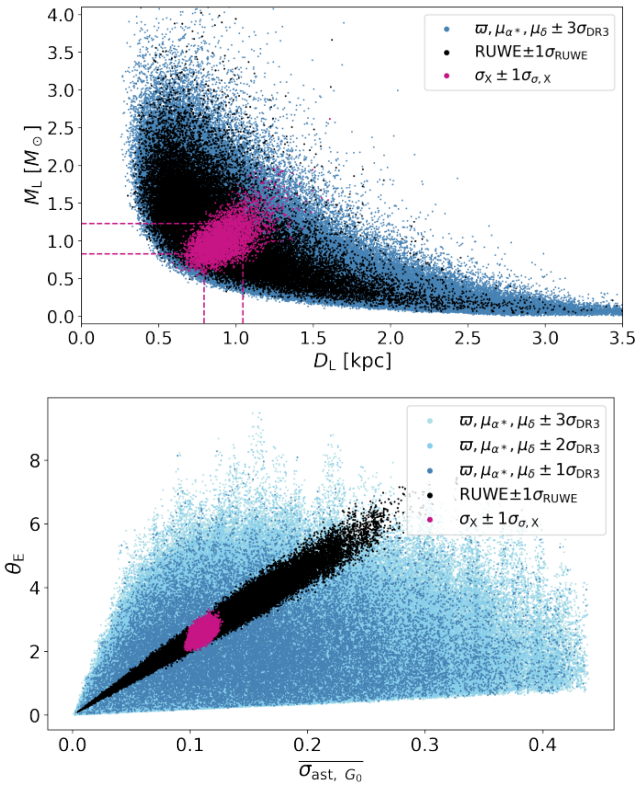

Figure 10. Figure from Jabłońska et al. (2022)