.

Statistical validation of Gaia Archive

prepared by: E. Antiche, R. Gutiérrez
approved by: X. Luri
reference: GAIA-C9-TN-UB-ELA-017
issue: 1
revision: 0
date: 2015-02-18
status: Issued

Abstract

We describe the statistics and validations performed to the Gaia Archive ingestion system. A GOG catalogue in gbin format was ingested into the Archive. The Gaia Analysis Tool (GAT) and Gaia Archive tools and have been used to validate the data before and after the ingestion, in order to detect differences.

Document History

Issue	Revision	Date	Author	Comment
I	0	2015-02-18	ELA	Ready to be uploaded at Livelink. Issued status
D	8	2015-02-16	ELA	Comments from G. Gracia. Added 'differences' column to the validation tables
D	7	2015-02-13	ELA	Sent to CU9 CCB
D	6	2015-02-12	RGS	Gaia Archive alphaError and deltaError histograms replaced
D	5	2015-02-11	ELA	Errors fixed. GOG alphaError and deltaError histograms replaced
D	4	2015-02-06	RGS	Section revision, some typos
D	3	2015-01-21	ELA	Added GAT statistics and major changes
D	2	2015-01-14	ELA	First draft in DPAC format
D	1	2015-01-14	RGS	Document creation in Word with Gaia Archive statistics

Contents

1 Introduction

This document describes the methodology used and the results obtained in the validation of Gaia Archive Core Systems ingestion, from now on, the Gaia Archive. The integrated results of the DPAC processing of the Gaia data are stored in the Main Data Base. To publish the data and make it available to the scientific comunity they are ingested in the Gaia Archive Core System (GACS) DB. The Gaia Archive are a set of subsystems developed for the Gaia project that allows to search, download and explore the Gaia data JSB-001.

For its first version, a GOG catalogue has been selected as test case. This catalogue was transferred to the ESAC systems in gbin format, using the MDB version 15.2.4. Later on, the catalogue was ingested into the Gaia Archive database. The GOG catalogue contains true data and combined end-of-mission data, in the NSL mode. Please refer to the GOG manual for more information regarding GOG and the error models YI-003.

The purpose of this statistical validation is to detect differences in the data due to the ingestion of the GOG catalogue in the Gaia Archive. We have built several sets of validation and statistics in order to compare the data before and after ingestion. The GOG catalogue in gbin format will be considered as the data before the ingestion and the catalogue located in the Gaia Archive as the data after ingestion.

References

[YI-003], Antiche, E., Julbe, F., Borrachero, R., Luri, X., 2013, *GOG User Guide*, GAIA-C2-UG-UB-YI-003, URL http://www.rssd.esa.int/cs/livelink/open/2775659

[GAIA-CU9-UG-UB-ELA-007-1], Antiche, E.B., R; Martinez, O.I., Y; Luri, X., 2014, *Description and usage of the GATCore library*, GAIA-CU9-UG-UB-ELA-007-1, URL http://www.rssd.esa.int/cs/livelink/open/3259832

Gutierrez, R., 2014, Statistical plots in the Gaia Archive, http://gaia.esac.esa.int/dpacsvn/DPAC/meetings/CU9/Plenary2014/GACS_SatisticalPlots_RGutierrez.pdf, [Online; accessed 13-February-2015]

[JSB-001], Salgado, J., 2013, *GACS Internal Releases until First Public Version*, GAIA-C9-PL-ESAC-JSB-001, URL http://www.rssd.esa.int/cs/livelink/open/3212802

Definitions, acronyms, and abbreviations

Acronym	Description
CU	Coordination Unit (in DPAC)
DPAC	Data Processing and Analysis Consortium
DPC	Data Processing Center
DU	Development Unit (in DPAC)
GACS	Gaia Archive Core Systems
GOG	Gaia Object Generator
GaiaSimu	Gaia Simulator (Universe model)
GAT	Gaia Analysis tool
MDB	Main DataBase
NSL	Nominal Scanning Law
SVN	SubVersioN

2 Input Data Description

The GOG catalogue has been simulated using the MDB/CU1/Integrated/CompleteSource format to then be converted into DPC/CU9/ArchiveArchitecture/Core/CatalogueSource format, as this table has the catalogue release format. The current analysis is focused in the following fields belonging to the CatalogueSource table:

Field	Units	Description	Data type
alpha	deg	Right ascension of object in ICRS at reference epoch	double
alphaError	mas	Standard error of right ascension of object in ICRS at reference epoch	double
delta	deg	Declination of object in ICRS at reference epoch	double
deltaError	mas	Declination of object in ICRS at reference epoch	double
muAlphaStar	mas/year	Proper motion component in alpha direction	double
mualphaStarError	mas/year	Standard error of muAlpha in alpha direction	double

muDelta	mas/year	Proper motion component in delta direction	double
parallax	mas	Parallax	double
parallaxError	mas	Standard error of the parallax	double
magGMean	mag	Mean G magnitude	double
radialVelocity	km/s	Radial Velocity	double
radialVelocityError	km/s	Standard error of the Radial Velocity	double

Table 2: Fields validated

Please refer to section 5, to for futher details about the issues found in the GOG catalogue.

3 Methodology

The analysis performed does not intend to be an exhaustive test, but a way of having a general picture of the data ingested in the archive, that can be compared with the same picture of the original data. In this document we compare the statistics generated before and after the ingestion in GACS. Statistical analysis of the data is performed in several steps. Statistics generated before the ingestion are generated with the Gaia Analysis Tool (GAT) reading the gbins from the MDB while other set of figures is generated with the statistics tools provided in GACS. Please refer to GAIA-CU9-UG-UB-ELA-007-1 for more information about GAT and to (Gutierrez, 2014) for more information about the methodology used in the Gaia Archive for the statistical plots generation.

While the whole set of data is used to compute general figures (mean, standard deviation, etc), limits are applied to generate more useful histograms.

The GOG catalogue contains exactly 1881159998 sources. The simulation to obtain this catalogue was performed up to magnitude 20, including single stars and system sources.

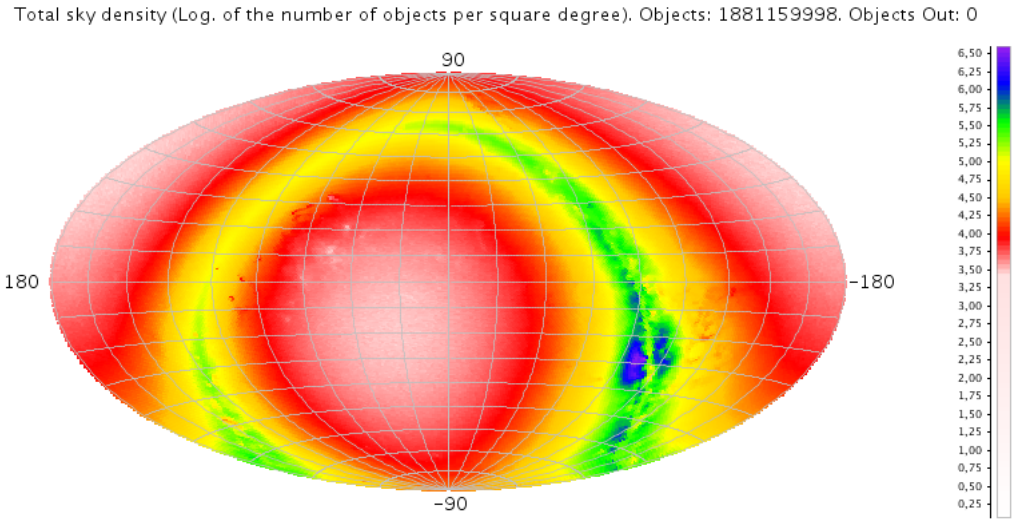

FIGURE 1: Total sky density (Log. of the number of objects per square degree)

4 Validation Results

In this section the results obtained for each of the analysed columns are presented.

4.1 Validation of alpha

	GOG	Archive	Difference
Counts	1881159998	1881159998	0
Mean	227.2592319386482	227.259231938259	3.8920689E-10
Std. dev.	82.4606795297589	82.4606795530568	-2.32979005E-8
Min. value detected	1.7054035960928776E-07	1.70540359609288E-07	-2.3822802E-22
Max. value detected	359.99999883648815	359.999998836488	1.7053026E-13
Monitored minimum	0.0 degrees	0.0 degrees	
Monitored maximum	360.0 degrees	360.0 degrees	

Table 3: Range monitored and detection results (using 1881159998 objects) of alpha

4.1.1 Archive histogram

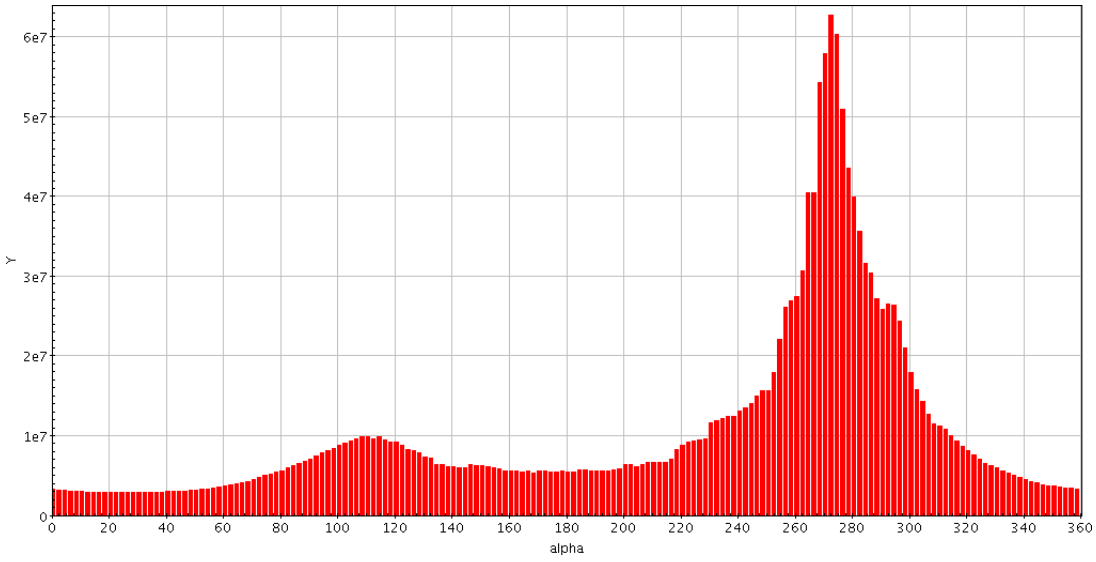

FIGURE 2: Gaia archive alpha histogram

4.1.2 GOG histogram

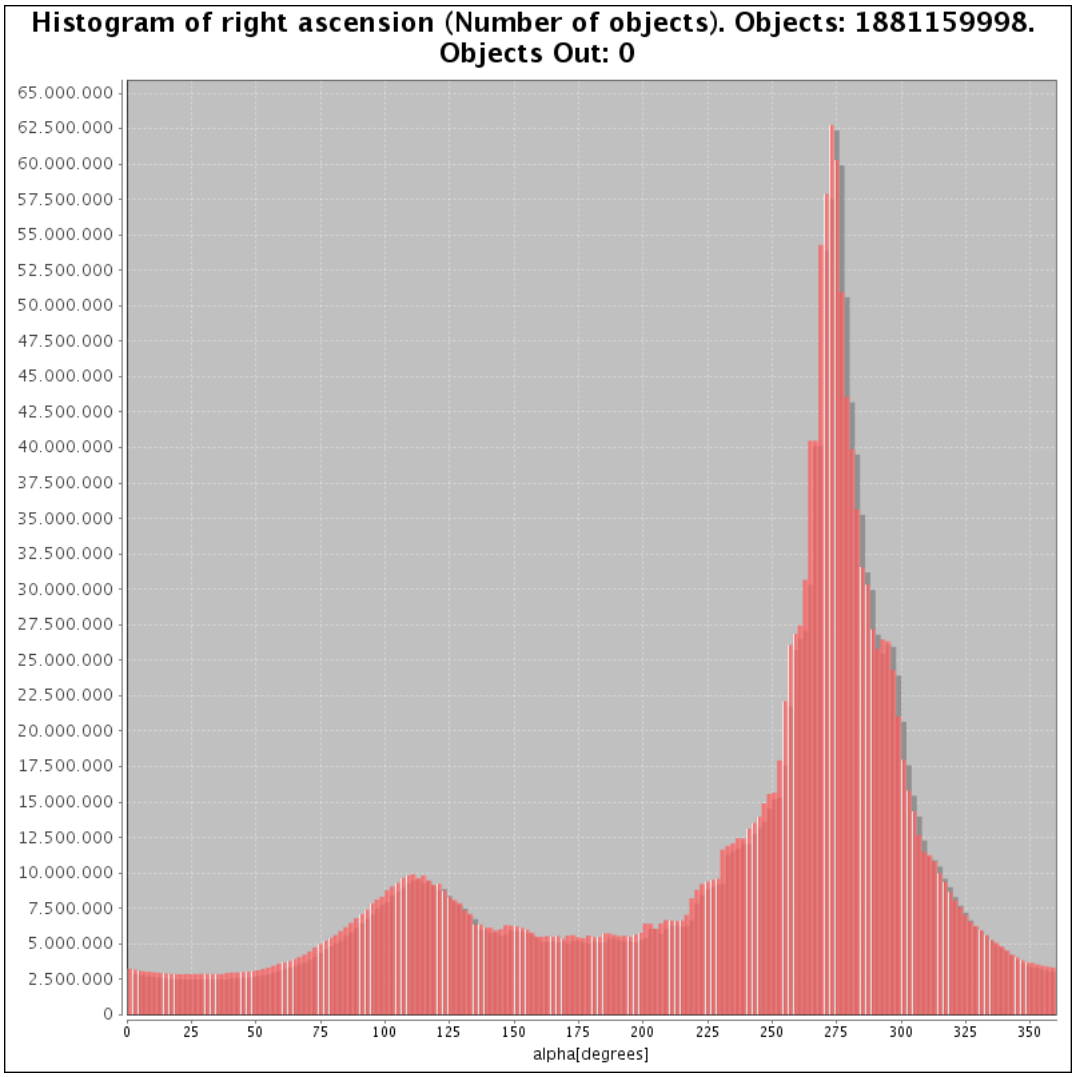

FIGURE 3: GOG alpha histogram

4.2 Validation of alphaError

	GOG	Archive	Difference
Counts	1881159998	1881159998	0
Mean	0.1138968139915306	0.113896813991839	-3.0840608E-13
Std. dev.	0.07209604473189782	0.0720960447505569	-1.8659088E-11
Min. value detected	0.00229548619377242	0.00229548619377242	0
Max. value detected	3.3151596022591927	3.3151596022592	-7.1054273E-15
Monitored minimum	0.0 mas	0.0 mas	
Monitored maximum	3.0 mas	3.0 mas	

Table 4: Range monitored and detection results (using 1881159998 objects) of alphaError

4.2.1 Gaia Archive histogram

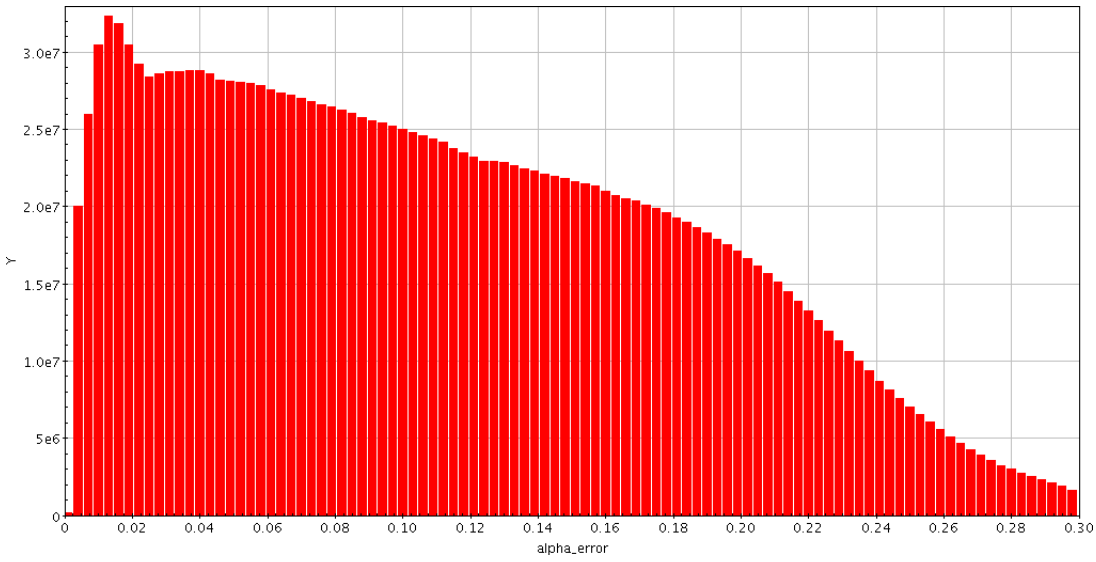

FIGURE 4: Gaia Archive alphaError histogram

4.2.2 GOG histogram

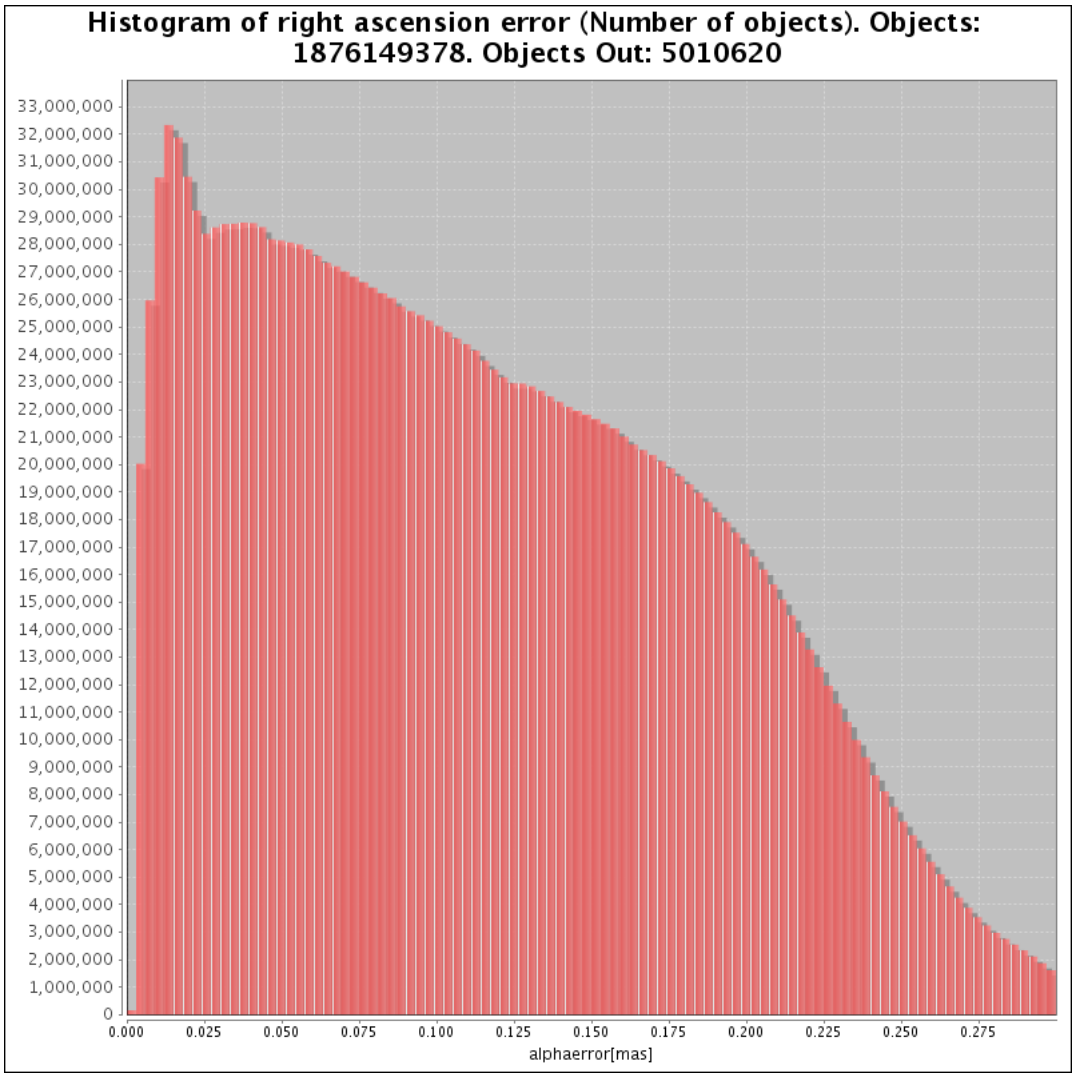

FIGURE 5: GOG alphaError histogram

4.3 Validation of delta

	GOG	Archive	Difference
Counts	1881159998	1881159998	0
Mean	-13.458781554583716	-13.4587815544862	-9.7516661E-11
Std. dev.	37.533988409470055	37.533988419469	-9.99894212E-9
Min. value detected	-89.99857786232435	-89.9985778623244	4.2632564E-14
Max. value detected	89.99037595594407	89.9903759559441	-2.8421709E-14
Monitored minimum	-90.0 degrees	-90.0 degrees	
Monitored maximum	90.0 degrees	90.0 degrees	

Table 5: Range monitored and detection results (using 1881159998 objects) of delta

4.3.1 Gaia Archive histogram

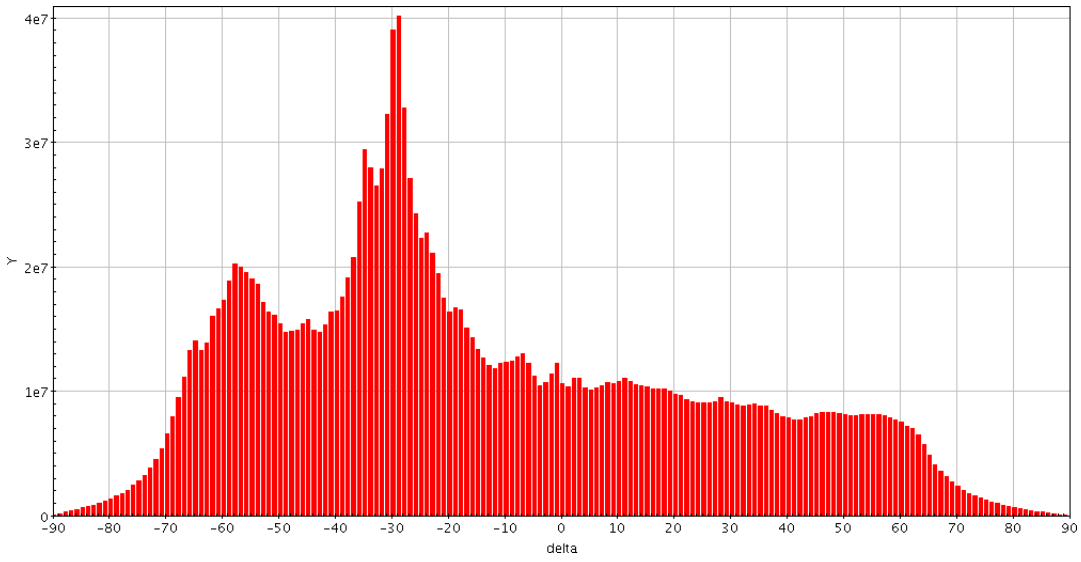

FIGURE 6: Gaia Archive delta histogram

4.3.2 GOG histogram

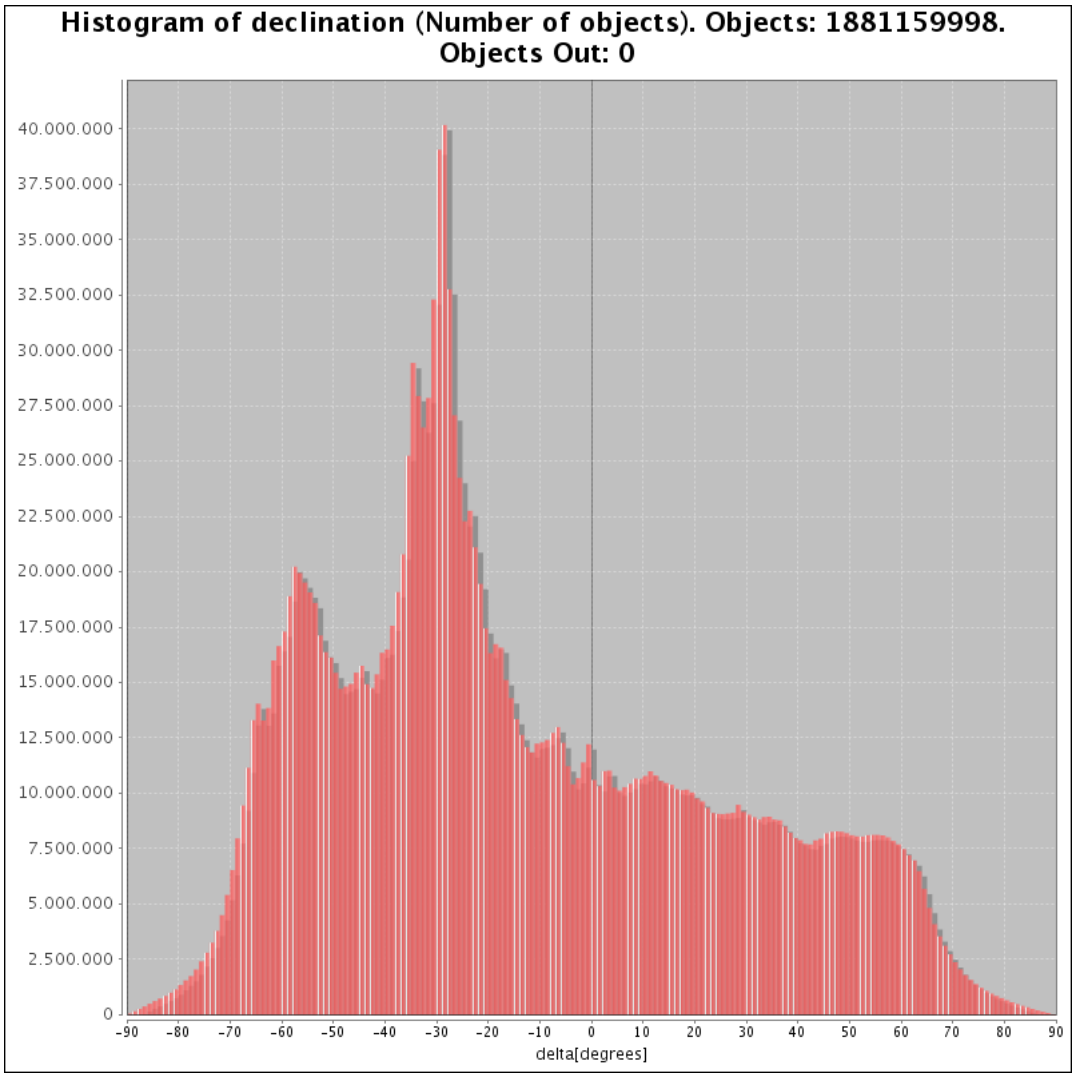

FIGURE 7: GOG delta histogram

4.4 Validation of deltaError

	GOG	Archive	Difference
Counts	1881159998	1881159998	0
Mean	0.10116121090228783	0.101161210902447	-1.5916435E-13
Std. dev.	0.06403447937416139	0.0640344793909611	-1.6799714E-11
Min. value detected	0.002038811752791514	0.00203881175279152	-5.6378513E-18
Max. value detected	2.9444683125529547	2.94446831255296	-5.3290705E-15
Monitored minimum	0.0 mas	0.0 mas	
Monitored maximum	3.0 mas	3.0 mas	

Table 6: Range monitored and detection results (using 1881159998 objects) of deltaError

4.4.1 Gaia Archive histogram

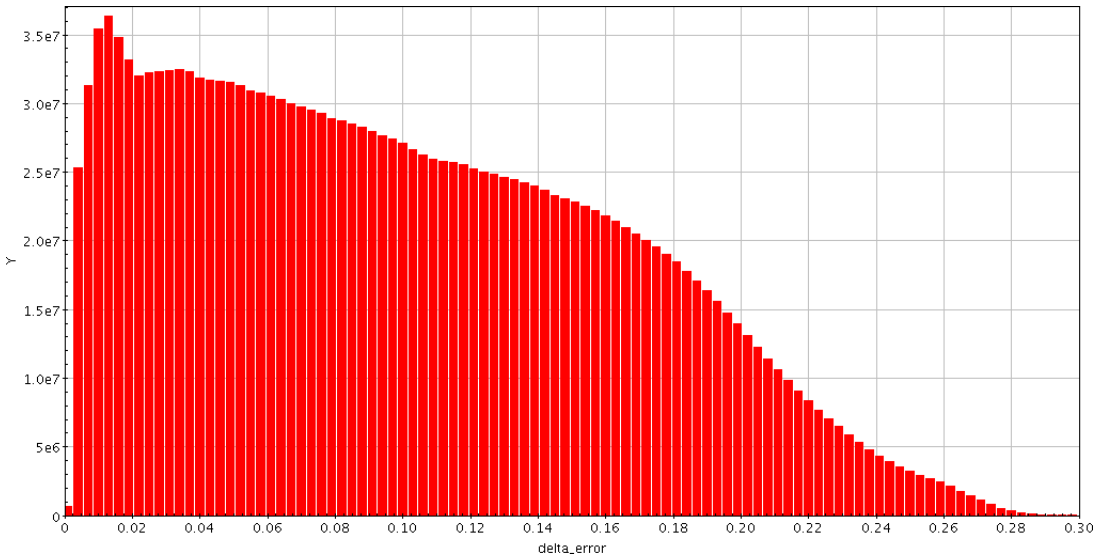

FIGURE 8: Gaia Archive deltaError histogram

4.4.2 GOG histogram

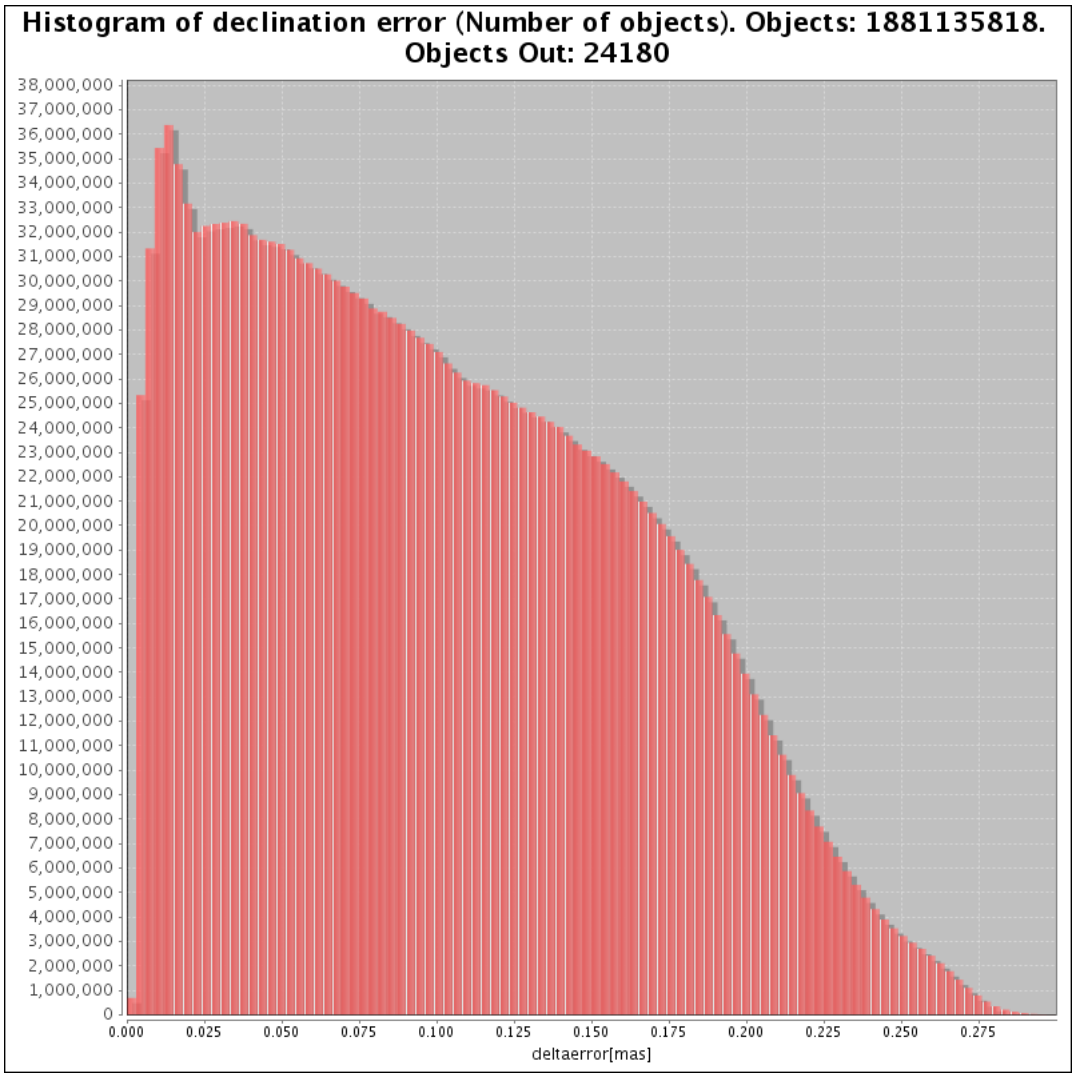

FIGURE 9: GOG deltaError histogram

4.5 Validation of muAlphaStar

	GOG	Archive	Difference
Counts	1881159998	1881159998	0
Mean	-1.9251091100486102	-1.92510911003903	-9.5801145E-12
Std. dev.	5.491292540663813	5.49129254211467	-1.45085721E-9
Min. value detected	-9039.741508901167	-9039.74150890117	3.6379788E-12
Max. value detected	8110.362963212998	8110.362963213	-1.8189894E-12
Monitored minimum	-90.0 mas/yr	-90.0 mas/yr	
Monitored maximum	90.0 mas/yr	90.0 mas/yr	

Table 7: Range monitored and detection results (using 1881159998 objects) of muAlphaStar

4.5.1 Gaia Archive histogram

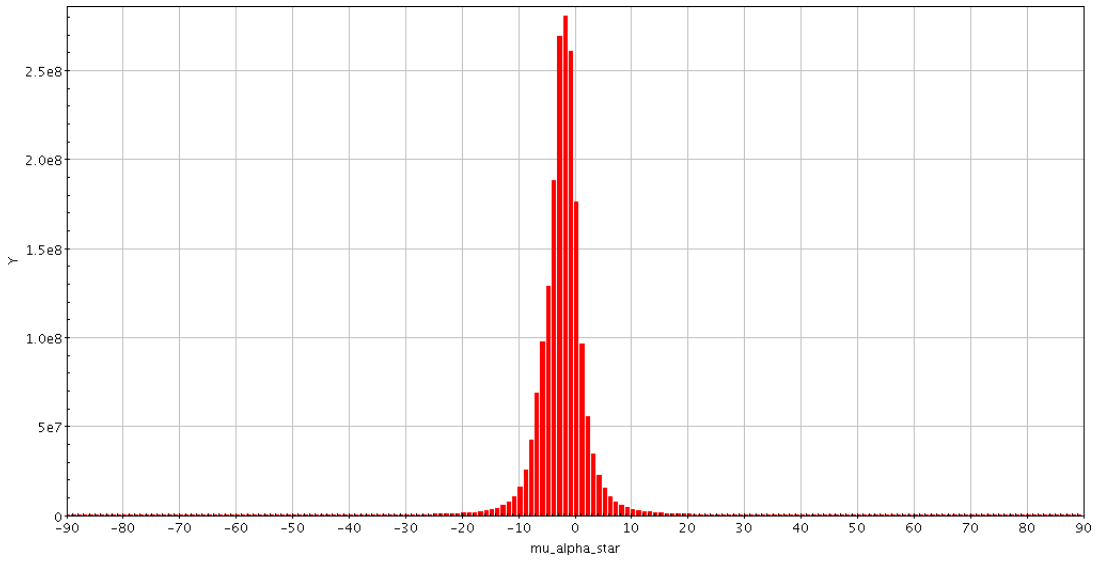

FIGURE 10: Gaia Archive muAlphaStar histogram

4.5.2 GOG histogram

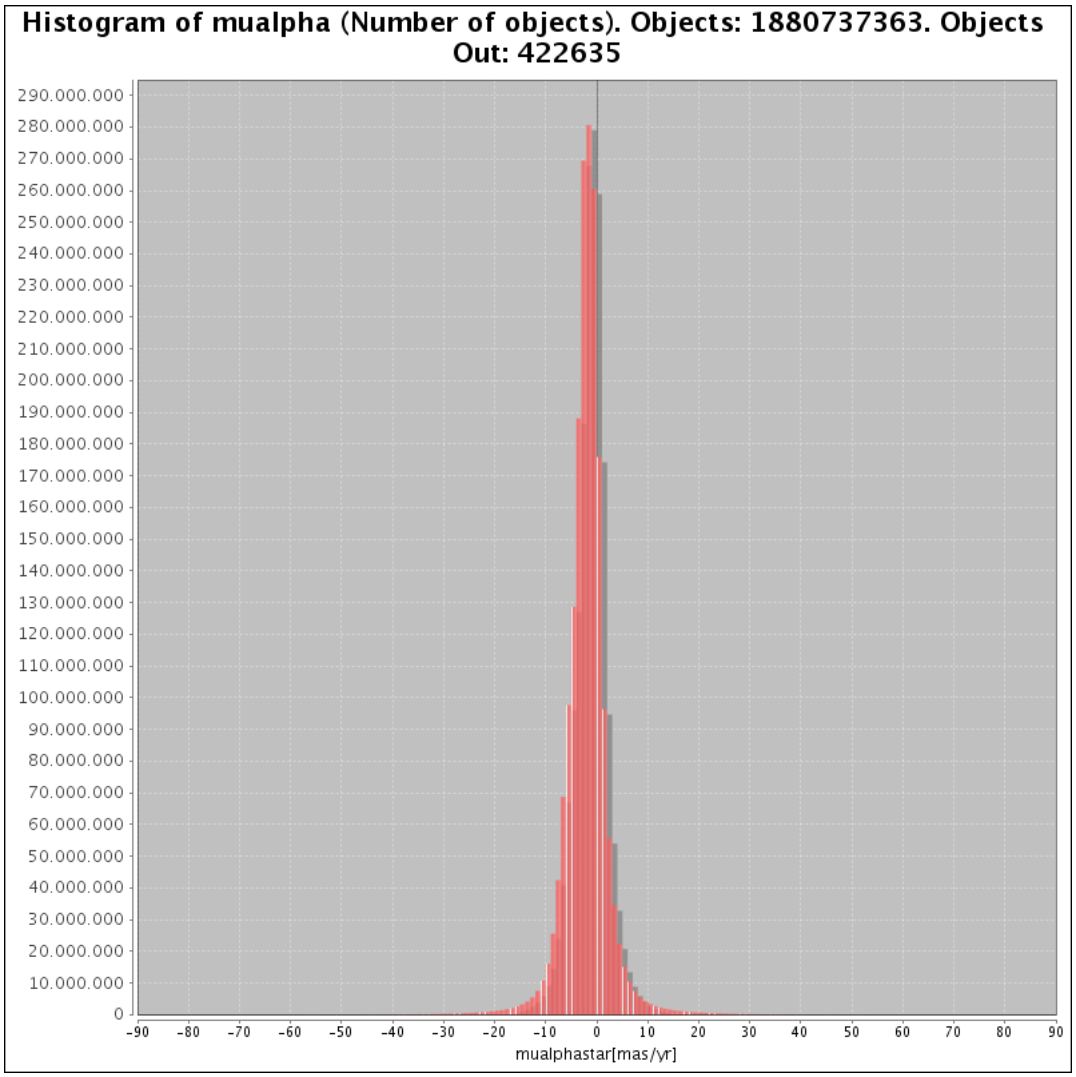

FIGURE 11: GOG muAlphaStar histogram

4.6 Validation of muAlphaStarError

	GOG	Archive	Difference
Counts	1881159998	1881159998	0
Mean	0.07178249013959273	0.0717824901396993	-1.0656753E-13
Std. dev.	0.0454375080239808	0.0454379138452838	-4.05821303E-7
Min. value detected	0.001446710485528745	0.00144671048552874	4.98733E-18
Max. value detected	2.089350905616979	2.08935090561698	-1.3322676E-15
Monitored minimum	0.0 mas/yr	0.0 mas/yr	
Monitored maximum	0.3 mas/yr	0.3 mas/yr	

Table 8: Range monitored and detection results (using 1881159998 objects) of muAlphaStarError

4.6.1 Gaia Archive histogram

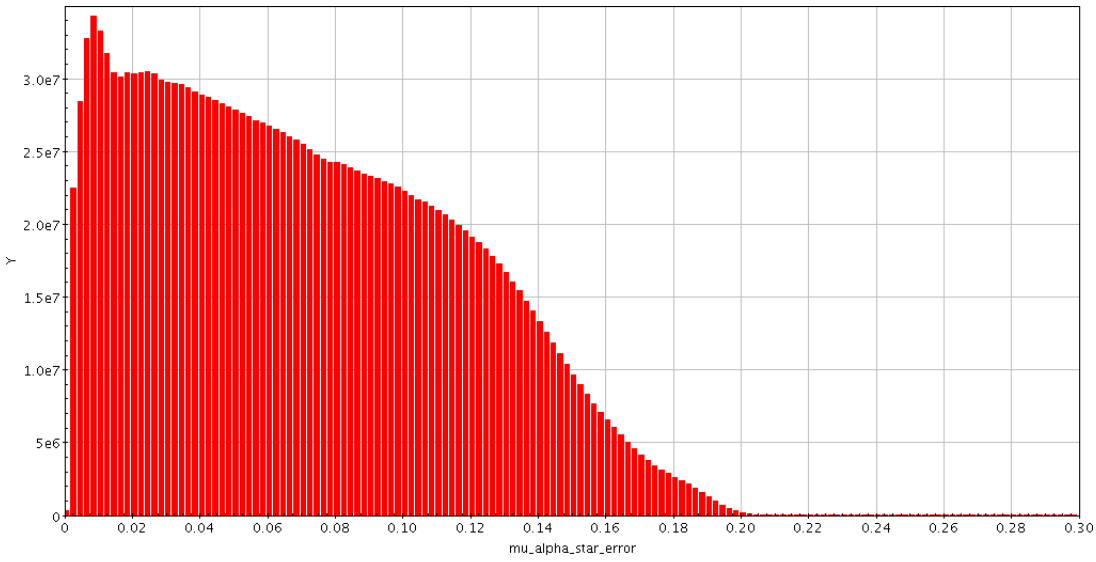

FIGURE 12: Gaia Ardhive muAlphaStarError histogram

4.6.2 GOG histogram

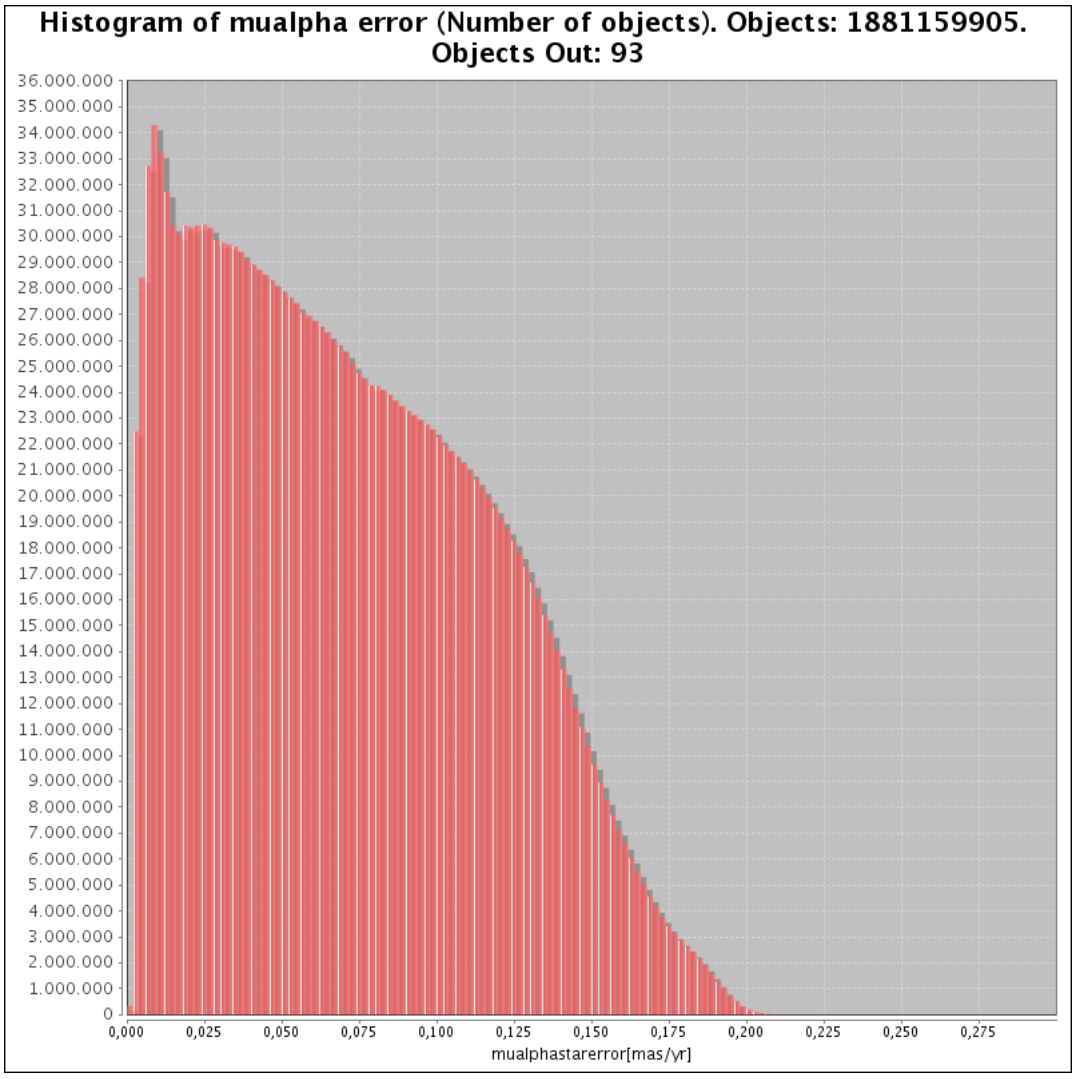

FIGURE 13: GOG muAlphaStarError histogram

4.7 Validation of muDelta

	GOG	Archive	Difference
Counts	1881159998	1881159998	
Mean	-2.1926858570348866	-2.19268585702899	-5.8966165E-12
Std. dev.	5.673317476904544	5.67331747840695	-1.5024062E-9
Min. value detected	-5572.44643654689	-5572.44643654689	0
Max. value detected	10996.92966134515	10996.9296613451	4.9112714E-11
Monitored minimum	-150.0 mas/yr	-150.0 mas/yr	
Monitored maximum	150.0 mas/yr	150.0 mas/yr	

Table 9: Range monitored and detection results (using 1881159998 objects) of muDelta

4.7.1 Gaia Archive histogram

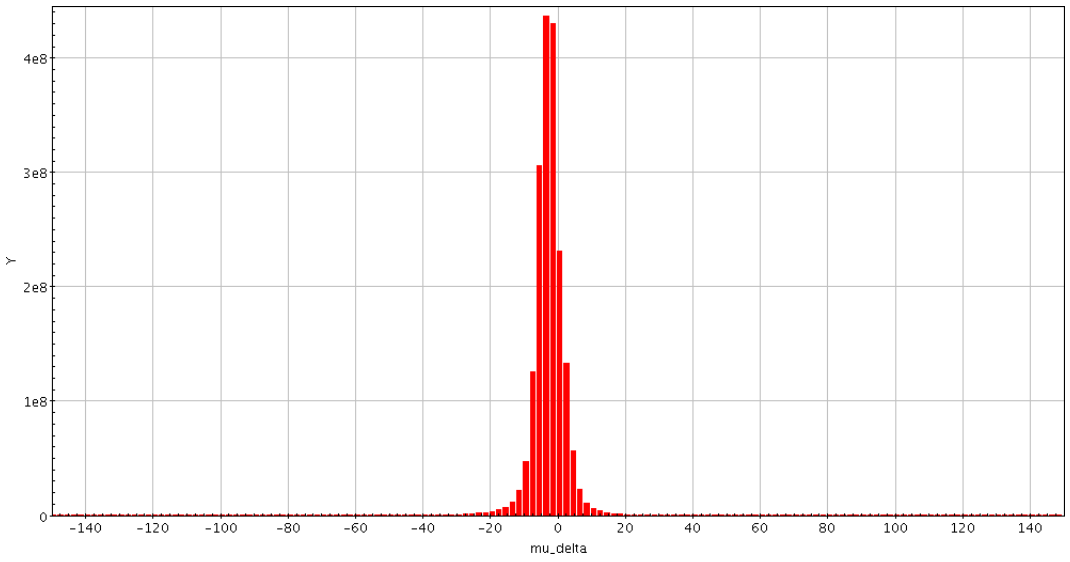

FIGURE 14: Gaia Archive muDelta histogram

4.7.2 GOG histogram

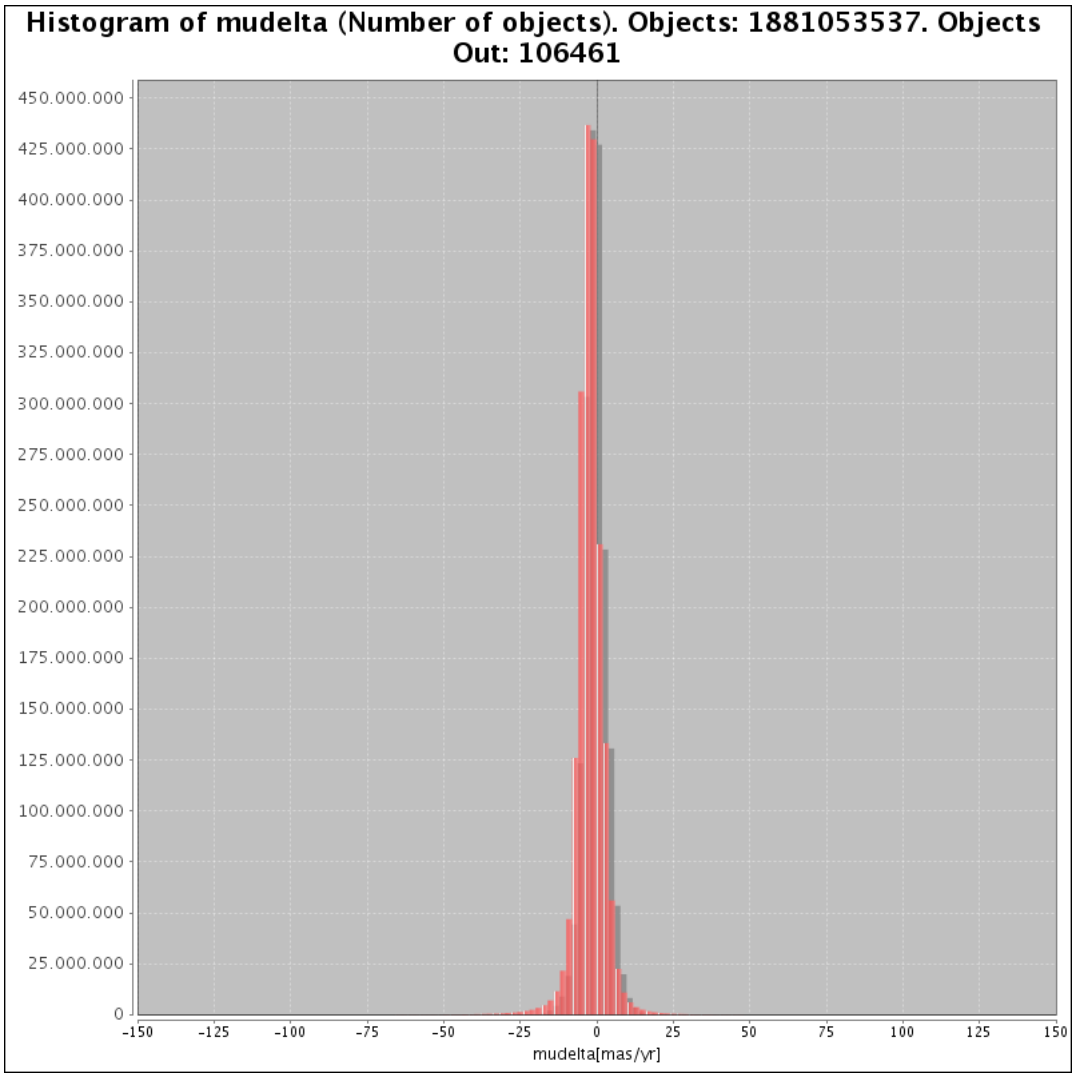

FIGURE 15: GOG muDelta histogram

4.8 Validation of parallax

	GOG	Archive	Difference
Counts	1881159998	1881159998	0
Mean	0.4124361448962173	0.412436144895801	4.1633363E-13
Std. dev.	0.6179052243070141	0.617905224473051	-1.66036851E-10
Min. value detected	-5.418328663617543	-5.41832866361754	-2.6645352E-15
Max. value detected	415.5901607591285	415.590160759128	4.54747351E-13
Monitored minimum	-2.0 mas	-2.0 mas	
Monitored maximum	10.0 mas	10.0 mas	

Table 10: Range monitored and detection results (using 1881159998 objects) of parallax

4.8.1 Gaia Archive histogram

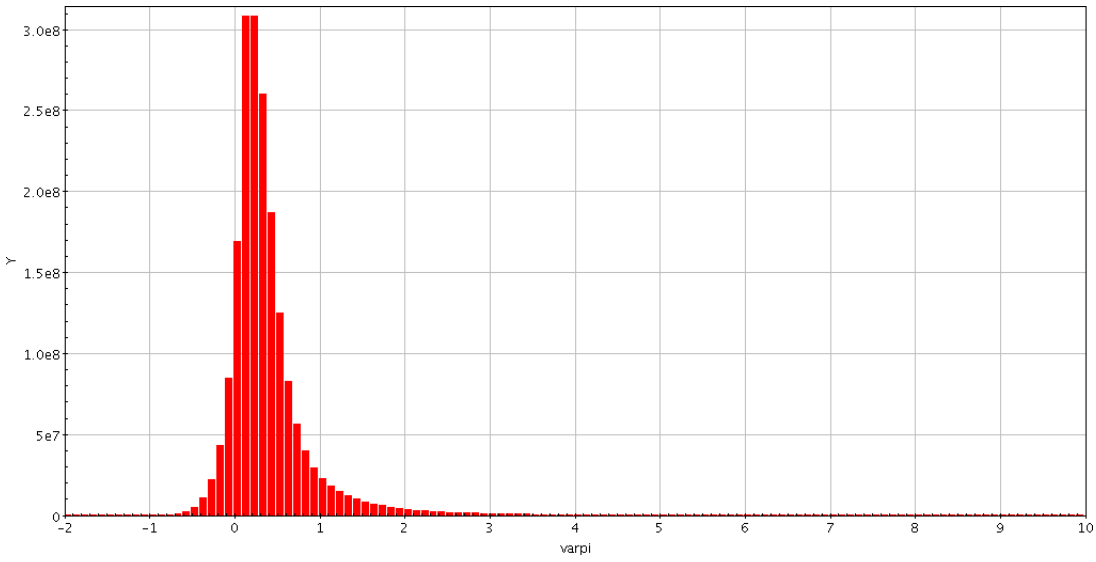

FIGURE 16: Gaia Archive parallax histogram

4.8.2 GOG histogram

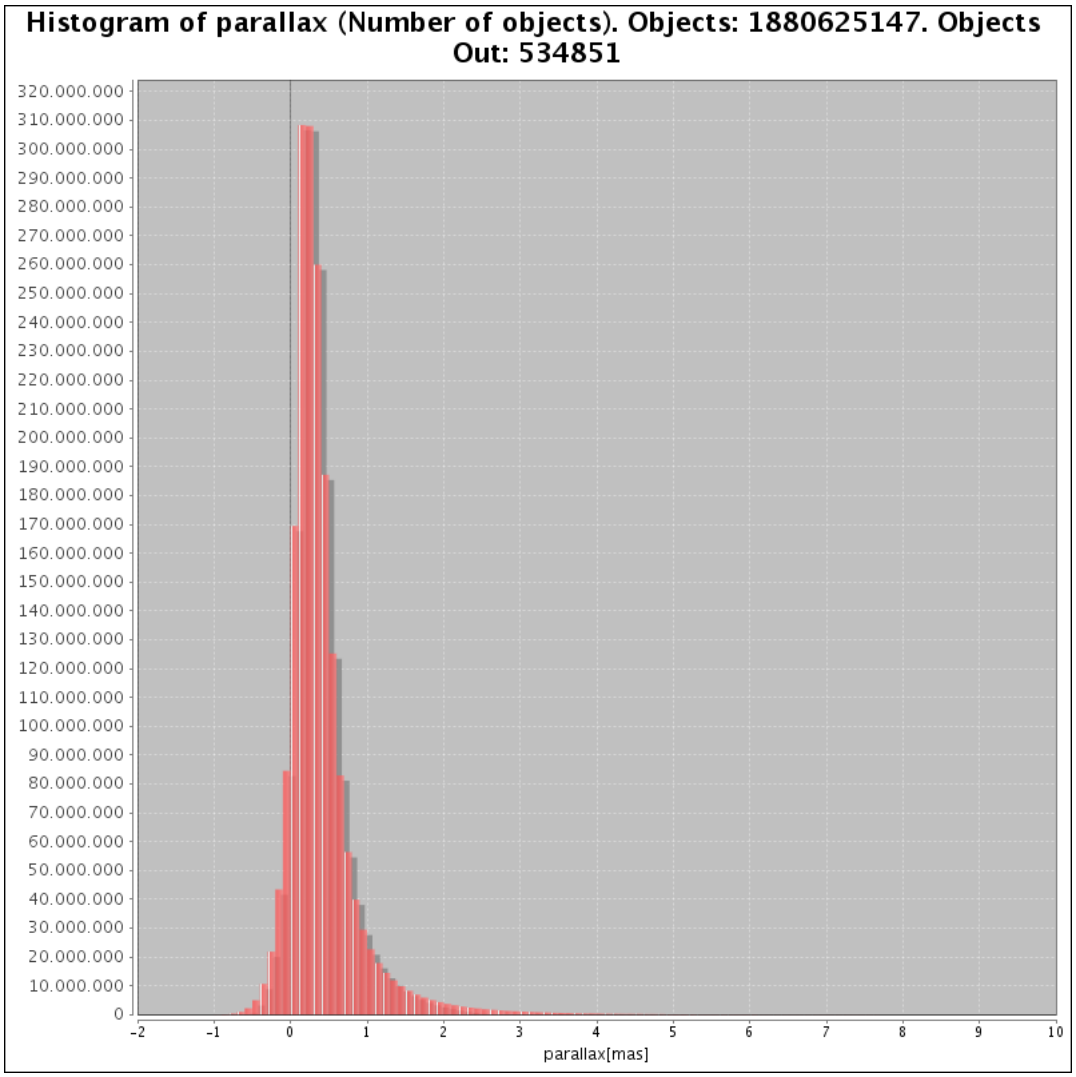

FIGURE 17: GOG parallax histogram

4.9 Validation of parallaxError

	GOG	Archive	Difference
Counts	1881159998	1881159998	0
Mean	0.14472276237817108	0.144722762378244	-7.2913897E-14
Std. dev.	0.0916086972444877	0.091608697268865	-2.4377306E-11
Min. value detected	0.002916755011146663	0.00291675501114666	3.0357661-18
Max. value detected	4.21240101938907	4.21240101938907	0
Monitored minimum	0.0 mas	0.0 mas	
Monitored maximum	0.5 mas	0.5 mas	

Table 11: Range monitored and detection results (using 1881159998 objects) of parallax

4.9.1 Gaia Archive histogram

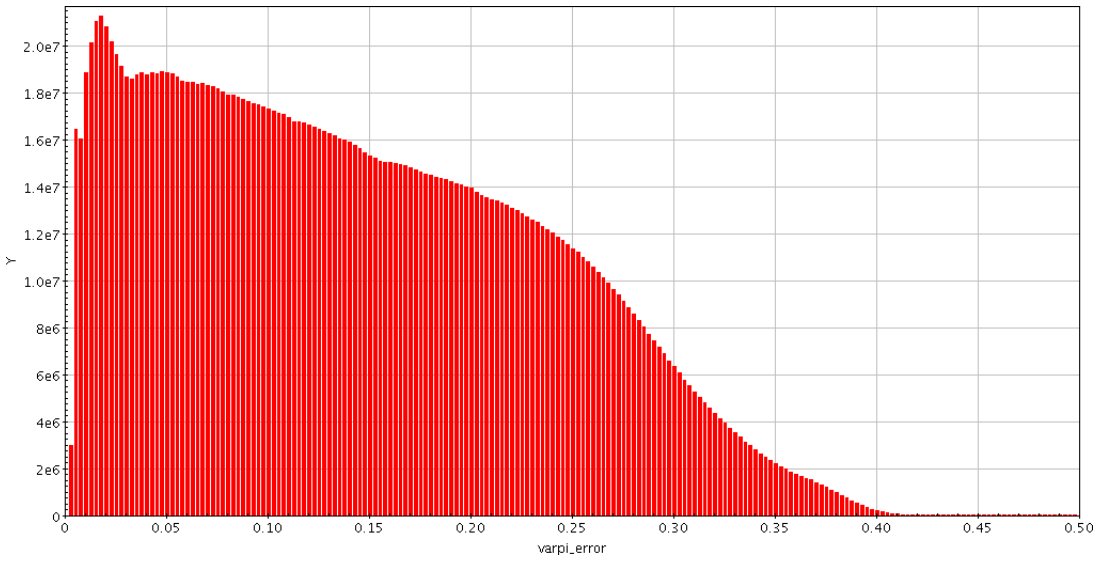

FIGURE 18: Gaia Archive parallaxError histogram

4.9.2 GOG histogram

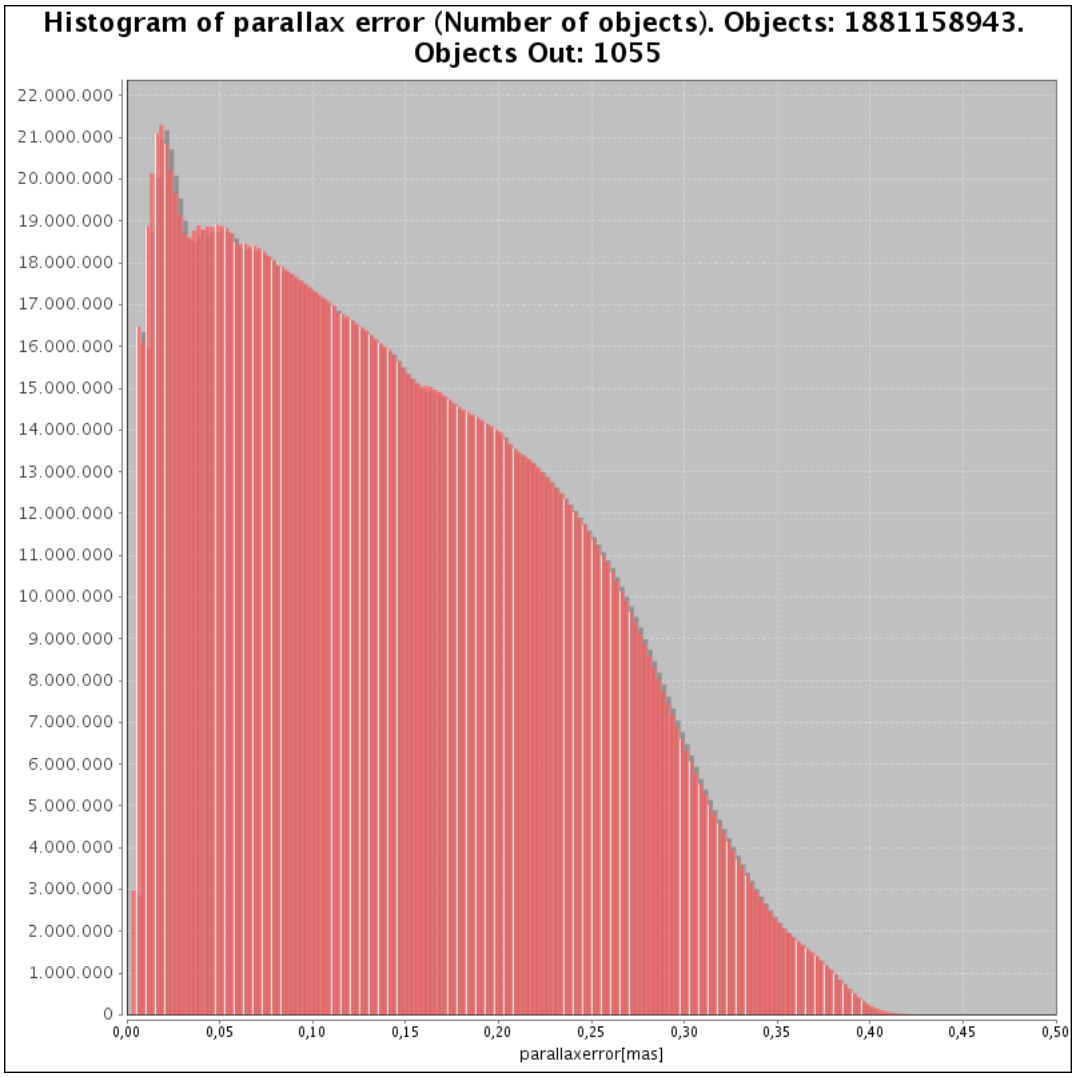

FIGURE 19: GOG parallaxError histogram

4.10 Validation of magGMean

	GOG	Archive	Difference
Counts	1881159998	1881159998	0
Mean	18.19317233669397	18.1931723366898	4.1708859E-12
Std. dev.	1.672232823322565	1.67223282370672	-3.8415515E-10
Min. value detected	5.492620229587953	5.49262022958795	2.6645352E-15
Max. value detected	23.417210359279277	23.4172103592793	-2.1316282E-14
Monitored minimum	5.0 mag	5.0 mag	
Monitored maximum	22.0 mag	22.0 mag	

Table 12: Range monitored and detection results (using 1881159998 objects) of magGMean

4.10.1 Gaia Archive histogram

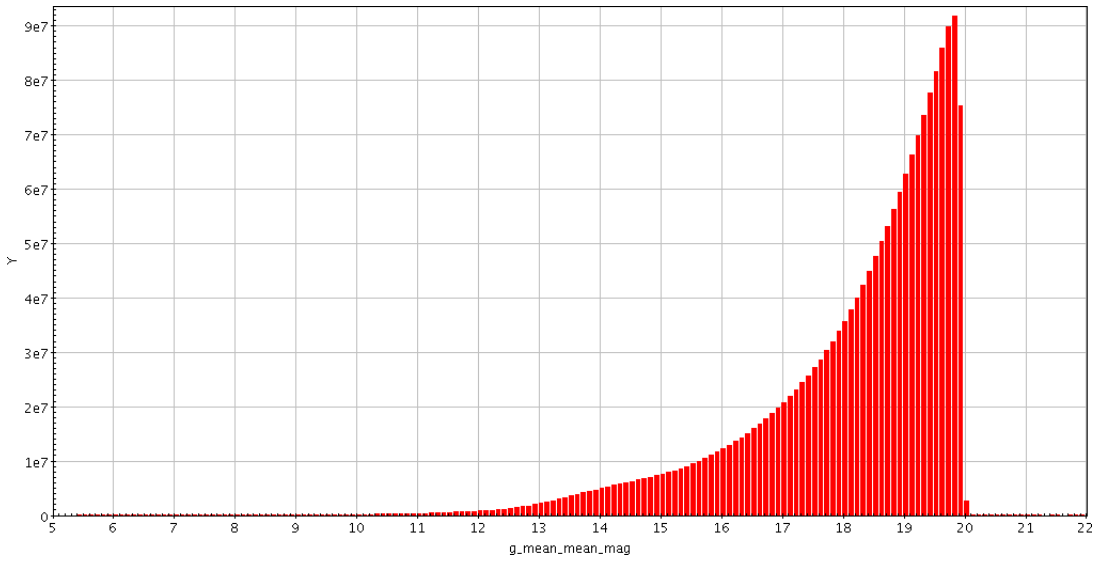

FIGURE 20: Gaia Archive magGMean histogram

4.10.2 GOG histogram

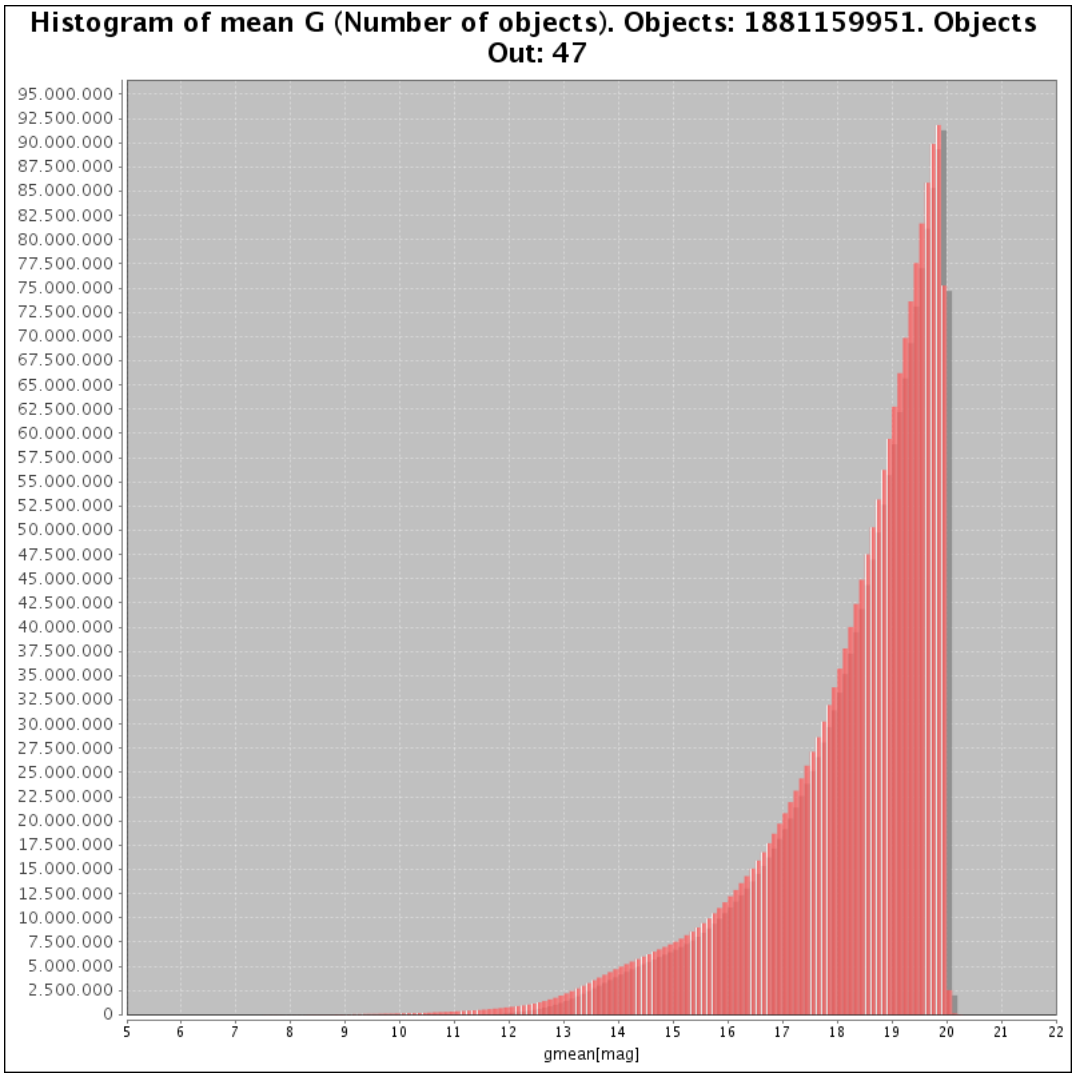

FIGURE 21: GOG magGMean histogram

4.11 Validation of radialVelocity

	GOG	Archive	Difference
Counts	502522117	502522117	0
Mean	-3.043814637358419	-3.04381463736421	5.7909233E-12
Std. dev.	77.99554078982418	77.9955408674368	-7.76126257E-8
Min. value detected	-2599.9870831649337	-2599.98708316493	-3.6379788E-12
Max. value detected	2884.329065354434	2884.32906535443	4.0927261E-12
Monitored minimum	-800.0 km/s	-800.0 km/s	
Monitored maximum	800.0 km/s	800.0 km/s	

Table 13: Range monitored and detection results (using 502522117 objects) of radialVelocity

4.11.1 Gaia Archive histogram

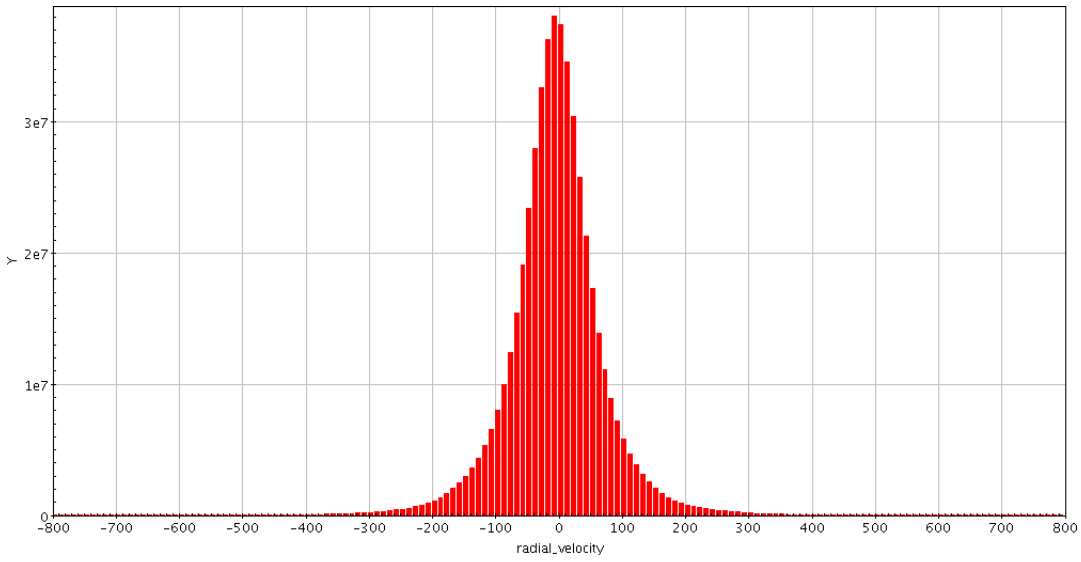

FIGURE 22: Gaia Archive radialVelocity histogram

4.11.2 GOG histogram

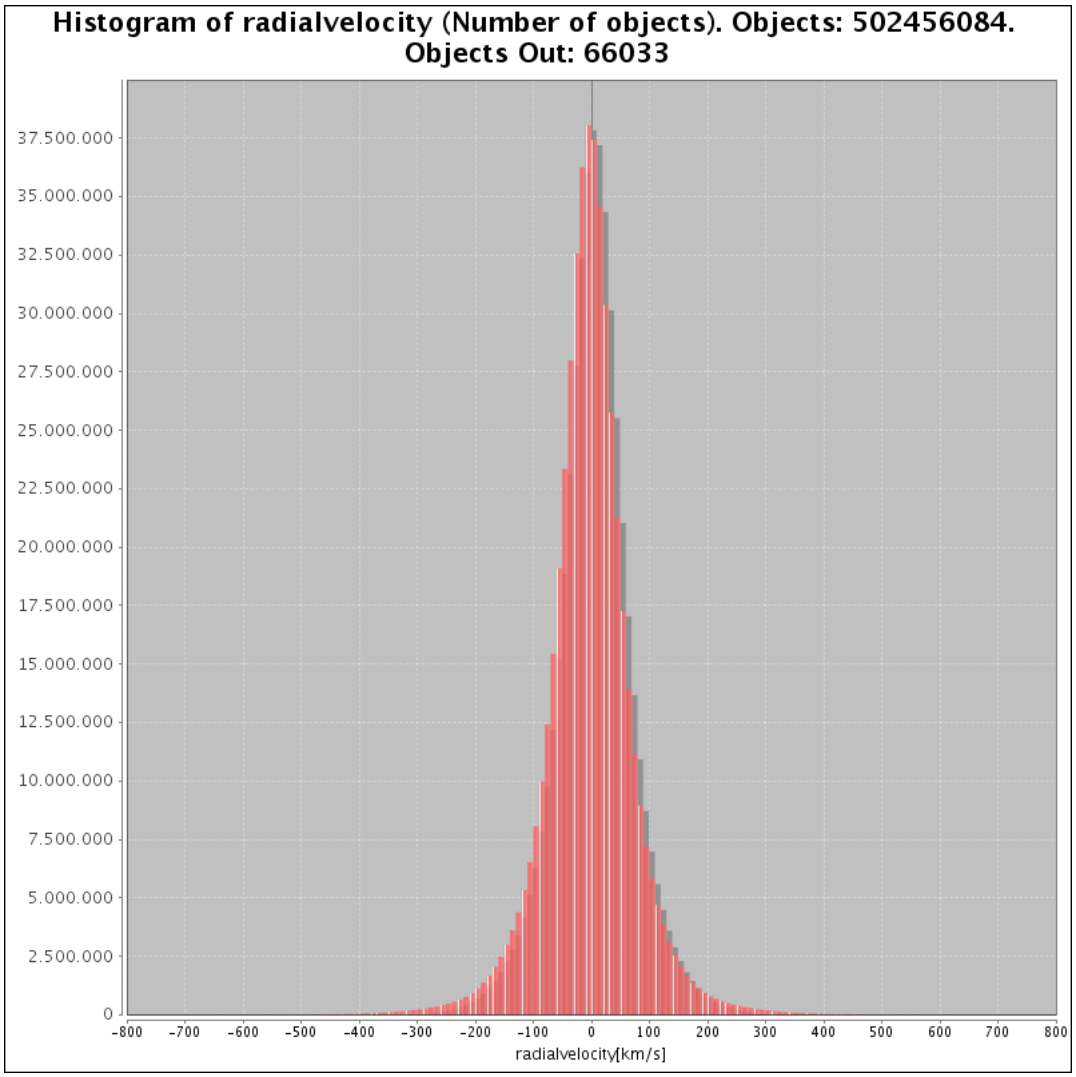

FIGURE 23: GOG radialVelocity histogram

4.12 Validation of radialVelocityError

	GOG	Archive	Difference
Counts	502522117	502522117	0
Mean	11.812608540416855	11.8126085403883	2.8554936E-11
Std. dev.	9.643497130762368	9.64349714039446	-9.63209246E-9
Min. value detected	1	1	0
Max. value detected	34.99999984179021	34.9999998417902	1.4210855E-14
Monitored minimum	0.0 km/s	0.0 km/s	
Monitored maximum	40.0 km/s	40.0 km/s	

Table 14: Range monitored and detection results (using 502522117 objects) of radialVelocityError

4.12.1 Gaia Archive histogram

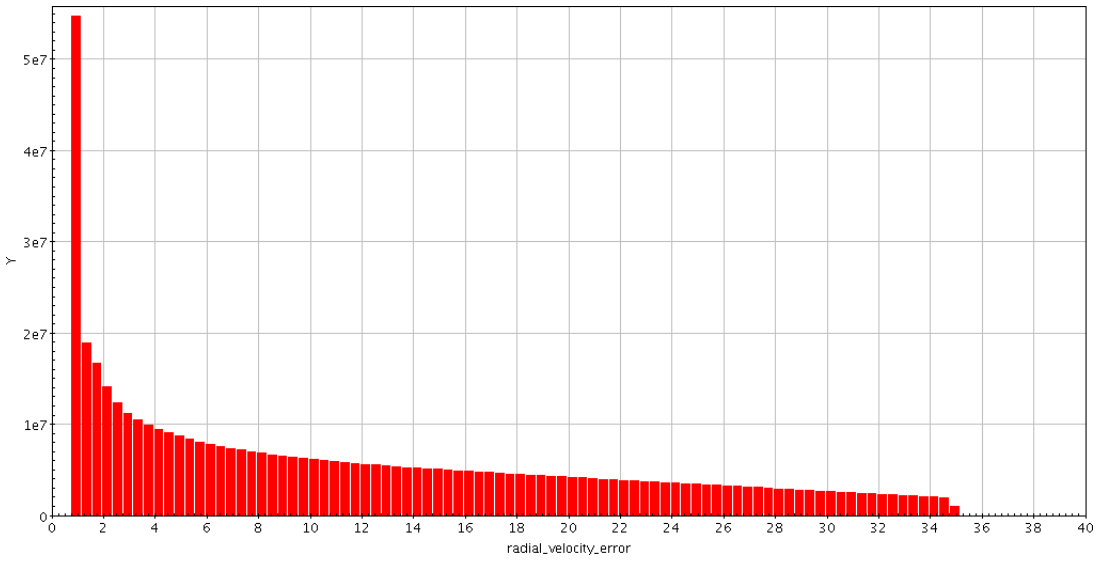

FIGURE 24: Gaia Archive radialVelocityError histogram

4.12.2 GOG histogram

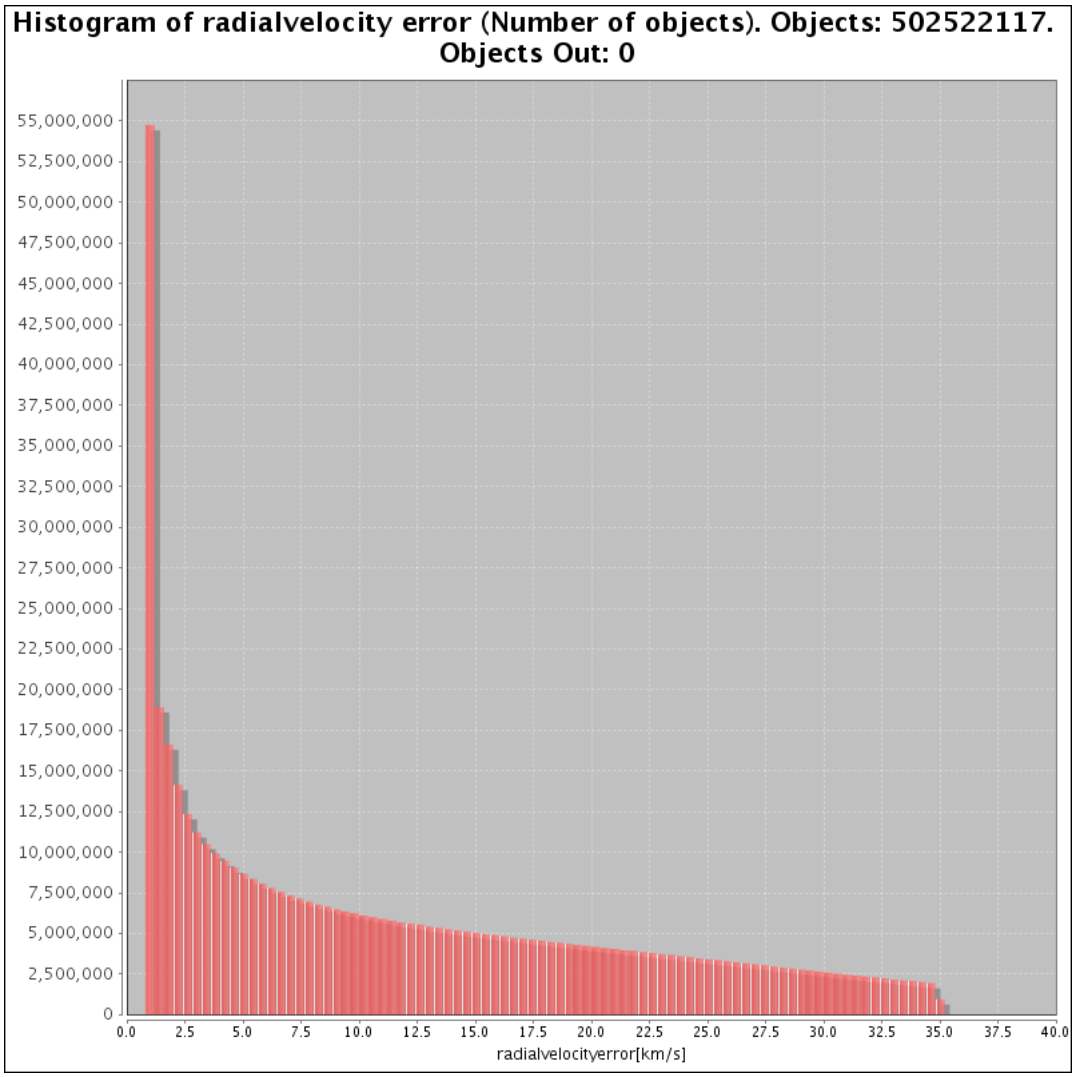

FIGURE 25: GOG radialVelocityError histogram

5 Known issues

The following issues have been detected inside the GOG catalogue:

- muDeltaError column is not filled, all values are zero.
- GOG performs a flux to magnitude conversion to fill the G, GBP and GRP magnitudes. However, the formula to convert the standar error from fluxes to magnitudes is wrong. The following fields are affected:
  - gMean (standard error)
  - bpMean (standard error)
  - rpMean (standard error)

As a result of this, values in magGMeanError and muDeltaError are erroneous, so the analysis has been omitted for these columns. All these issues have been reported in the following Mantis: `http://www.rssd.esa.int/mantisSI/view.php?id=30265`

6 Conclusions

When looking at the differences between GOG and the Gaia Archive values, we found that differences in means and standard devations are in general after the 8th digit behind the decimal point. For the maximum and minimum detected values, this is found to happen around the 12th digit after the decimal point. It is important to mention that, all the histograms in both frameworks were executed using the same bin width value, then we think that the origin of these differences are probably related with natural computational processes, e.g. the double precision inherent to the programming language or to the OS, rounding, etc. As the minimum and maximum values are not completely equals, it is expected to find differences between the means and standard deviation values. We consider that these differences are negligibles, and it is tremendously difficult (sometimes even impossible) to avoid them completely.

A Appendix

The GOG catalogue can be found in the Gaia Archive as public.gog_cataloguesource in `https://gaia.esac.esa.int/archive/`. The fields from the MDB CatalogueSource are equivalent to the following fields inside the Archive as:

CatalogueSource	**Gaia Archive gog_cataloguesource**
alpha	alpha
alphaError	alpha_error
delta	delta
deltaError	delta_error
muAlphaStar	mu_alpha_star
muAlphaStarError	mu_alpha_star_error
muDelta	mu_delta
parallax	varpi
parallaxError	varpi_error
magGMean	g_mean_mean_mag
radialVelocity	radial_velocity
radialVelocityError	radial_velocity_error

Table 15: Some of the CatalogueSource fields and their equivalents inside the Gaia Archive (public.gog_cataloguesource)