# *Gaia* and massive stars in 2025

## J. Maíz Apellániz

Centro de Astrobiología, Campus ESAC, CSIC-INTA. 28 692 Villanueva de la Cañada, Madrid, Spain

**Abstract.** I present some of the highlights of the *Gaia* mission on massive stars and discuss what the fourth data release (DR4) will bring in late 2026. In the first part of the contribution I describe the different types of *Gaia* products available now and for DR4 and their caveats. In the second part I present the most significant results on massive stars regarding parameter determination, binaries, photometric variability, stellar groups, and the sample and spatial distribution.

**Keywords.** Astrometry, Gaia, Galactic structure, Multiple stars, Photometry, Spectroscopy, Stellar clusters and OB associations, Stellar variability

---

## 1. The *Gaia* mission: products and caveats

The *Gaia* mission (Prusti et al. 2016) took data for ten and a half years from 24 July 2014 to 15 January 2025, yielding the most complete ever high-dynamic-range, astrometric+photometric+spectroscopic survey of the whole sky. The third data release (DR3) took place in two phases in 2020 and 2022 (Brown et al. 2021; Vallenari et al. 2023) and the next and fourth one (DR4) is expected to take place in December 2026†. Table 1 summarizes the different data products available in DR3 and DR4. In this section we describe the characteristics and caveats of the different types.

### 1.1. *Astrometry*

*Gaia* DR3 astrometry is described in Lindegren et al. (2021b,a) as part of the "official" early-DR3 (EDR3) papers from DPAC (*Gaia* Data Processing and Analysis Consortium‡). Further contributions from Cantat-Gaudin & Brandt (2021); Maíz Apellániz et al. (2021); Maíz Apellániz (2022) have expanded and refined that analysis. The most important points that need to be considered for DR3 astrometry are:

• Distances depend on the chosen prior (Maíz Apellániz 2005; Luri et al. 2018). The commonly used Bailer-Jones et al. (2021) distances assume an underlying spatial distribution for low- and intermediate-mass stars. Pantaleoni González et al. (2025) instead assumes a spatial distribution for Galactic OB stars (Maíz Apellániz 2001; Maíz Apellániz et al. 2008).

• Catalog parallaxes require a zero point correction (Lindegren et al. 2021a; Maíz Apellániz 2022). Cantat-Gaudin & Brandt (2021) derive a similar correction for proper motions.

• Catalog (or internal) parallax uncertainties are underestimated by up to a factor of three (Fabricius et al. 2021; Maíz Apellániz 2022, see also https://www.cosmos.esa.int/documents/29201/1770596/Lindegren_GaiaDR2_Astrometry_extended.pdf).

• There is a residual systematic uncertainty ($\sigma_{\rm s} = 10.3\,\mu$as) that yields a minimum uncertainty for distances of $\approx d\%$, where $d$ is the distance in kpc (Maíz Apellániz et al. 2021).

• The residual systematic uncertainty has an angular correlation (Lindegren et al. 2018; Maíz Apellániz et al. 2021).

• Poor-quality data (RUWE > 1.4 or six-parameter solutions) can still be used with a larger uncertainty (Maíz Apellániz 2022).

† https://www.cosmos.esa.int/web/gaia/release
‡ https://www.cosmos.esa.int/web/gaia/dpac/consortium

| Type | Details | DR3 | DR4 |
|---|---|---|---|
| All | Release dates+data span | 2020/2022, 34 months | Dec 2026, 66 months |
| Astrometry | Coordinates + parallaxes + proper motions | Mean + uncertainties | Mean + uncertainties + epoch data |
| Photometry | $G$+ $G_{BP}$+ $G_{RP}$+ $G_{RVS}$ | Mean + uncertainties + epoch data ($\sim 1\%$) | Mean + uncertainties + epoch data |
| Spectrophotometry | 3300-10 200 Å low + variable resolution | Mean basis coefficients for $\sim 10\%$ | Mean + epoch basis coefficients for X% |
| Spectroscopy | 8450-8740 Å $R \sim 11\,500$ | Mean spectra for $\sim 0.1\%$ + Mean RV for $\sim 10\%$ | Mean + epoch spectra + RV for X% |

**Table 1.** Summary of *Gaia* data products for DR3 and DR4

### 1.2. *Photometry*

*Gaia* DR3 photometry is described in Riello et al. (2021) as part of the DPAC papers. In Maíz Apellániz & Weiler (2025) we reanalyse it and provide a model to use it when comparing to synthetic photometry:

• *Gaia* optical photometry is unparalleled in coverage, dynamic range and uniformity but the passbands have evolved with each data release (from DR1 to DR3 and likely DR4, see Maíz Apellániz 2017; Maíz Apellániz & Weiler 2018) likely as a result of water deposition in optical surfaces at the beginning of the mission (Prusti et al. 2016) and differences in processing between data releases.

• *G* catalog magnitudes require a correction.

• *G* magnitudes in DR2 and DR3 require two different passbands with a cutoff at $G = 13.0$ and $G_{BP}$ and $G_{RP}$ magnitudes require two different passbands in DR2 with a cutoff at $G = 10.87$. This is likely a result of the electronic gate-based data processing (Prusti et al. 2016).

• If the Maíz Apellániz & Weiler (2025) procedure is followed it is possible to attain an absolute calibration between 3.4 and 5.7 mmag for *Gaia* DR3 photometry.
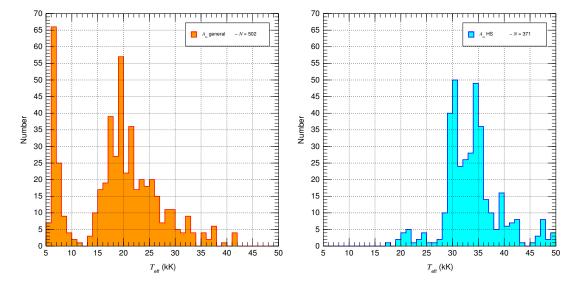
### 1.3. *Spectrophotometry*

*Gaia* spectrophotometry results from combining BP and RP prism data to cover the 3300-10 200 Å range. The combination of low and variable resolution complicates the interpretation of the results. The output is given as a series of coefficients of an expansion in basis functions, but the choice of basis functions partially determines the derived $f_\lambda(\lambda)$ and possibly introduces artifacts if the real $f_\lambda(\lambda)$ cannot be realistically expressed as a combination of the basis functions (Weiler et al. 2020). This is especially true for extinguished OB stars. Under those conditions, synthetic photometry and line measurements are not straightforward (Weiler et al. 2023). The reader is encouraged to read those papers before using the data.

### 1.4. *Spectroscopy*

Data from the *Gaia* Radial Velocity Spectrometer (RVS) has a resolution of 11 500 and covers the Calcium triplet window, where few lines exist for OB stars other than the broad Paschen lines. Furthermore, very few spectra of massive stars were published in DR3 and the papers on "hot stars" in DR3 refers to mostly intermediate- and low-mass stars (Blomme et al. 2023; Frémat et al. 2023). Hopefully in DR4 there will more RVS data for massive stars.

### 1.5. *Multi-origin processed data products*

*Gaia* also provides a number of processed products obtained by combining different data types. An example is the *Gaia* astrophysical parameters inference system (Apsis, Bailer-Jones et al. 2013), for which the results for OB stars in DR3 are described below.

**Figure 1.** Values for $T_{eff}$ determined by Apsis for 947 O stars with accurate GOSSS spectral classifications (Maíz Apellániz et al. 2011) using GSP-Phot (left) and ESP-HS (right).
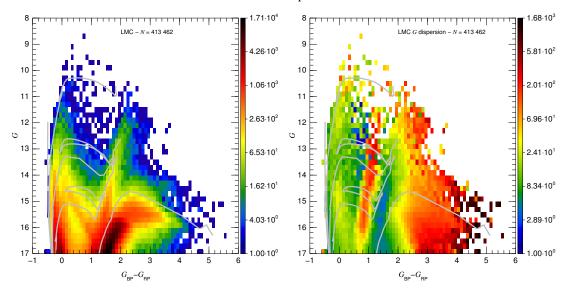
## 2. Parameter determination

The DR3 DPAC determination of stellar parameters using Apsis is described in Creevey et al. (2023); Fouesneau et al. (2023) and the specific case of "hot stars" (defined as those with $T_{eff} > 7.5$ kK) is discussed in Frémat (2024). There are different pipelines involved in Apsis, including a general one called GSP-Phot and another one for hot stars called ESP-HS. Frémat (2024) shows poor results in the confusion matrix that compares the spectral types predicted by ESP-HS and the ones from Simbad (which have their own problems, see Maíz Apellániz et al. 2016), with a large fraction of misclassified O and B stars, including a majority of O stars. I did a similar study (compatible with the Frémat 2024 results), partially presented as Maíz Apellániz (2024) comparing the DR3 Apsis $T_{eff}$ values with O stars with accurate GOSSS spectral types (Maíz Apellániz et al. 2011). Figure 1 shows how GSP-Phot only correctly identifies a few O stars as such, with a significant fraction being assigned $T_{eff} < 10$ kK. ESP-HS does a much better job but still underestimates $T_{eff}$ for most O stars.

Therefore, at least for DR3, Apsis does a poor job at determining the properties for OB stars but why? There are at least two reasons:

[1] Apsis (and *Gaia* in general) was designed mostly with cool stars in mind. To determine the properties of hot stars (a) the blue-violet range is much better that the Calcium triplet window of RVS and (b) the Balmer jump is difficult to measure with BP spectrophotometry because it is located close to its wavelength edge and also likely because the predetermined basis functions do not work well with extinguished OB stars.

[2] Apsis uses the Fitzpatrick (1999) family of extinction laws instead of other existing better choices and, more importantly, fixes the value of $R$ to 3.1, which only works properly for a minority of OB stars (Maíz Apellániz et al. 2014; Maíz Apellániz & Barbá 2018). As a consequence, the determined extinguished SED is inconsistent with the same *Gaia* spectrophotometry used to establish it (see Fig. 9 in Maíz Apellániz 2024). The incorrectly determined extinction quite likely interferes with an accurate $T_{eff}$ determination.

## 3. Binaries

The largest DR3-based study of binaries is Mowlavi et al. (2023), who analysed over 2 million eclipsing binaries from the Rimoldini et al. (2023) sample, see below. *Gaia* DR3 typically

**Figure 2.** (left) CMD and (right) mean *G*-band photometric dispersion in mmag for the LMC *Gaia* DR3 sample of Maíz Apellániz et al. (2023). The most prominent variability regions correspond (in increasing $G_{BP} - G_{RP}$) to Be stars, cepheids, and long-period variables, respectively.
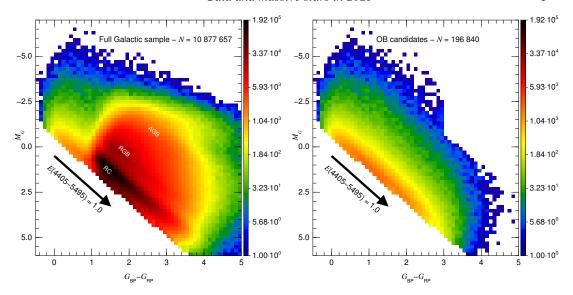
included 50 photometric epochs (the exact number depends primarily on sky position) and the high photometric accuracy, spatial resolution, and stability of *Gaia* photometry is a good complement to other surveys such as TESS, which has a much better sampling but poorer spatial resolution and long-term photometric accuracy. See Holgado et al. (2025) for an example of how to combine the best of both worlds to study massive eclipsing binaries. DR4 will provide $\sim 100$ photometric epochs for a much larger sample and should allow the discovery and study of even more eclipsing binaries (massive and non-massive).

Of course, the most interesting massive binaries are those that are expected to be discovered through the combination of epoch astrometry and radial velocities, especially those with a BH or NS component where the motion of the photocentre is most prominent (Andrews et al. 2019; Janssens et al. 2022). However, as *Gaia* DR3 published very limited epoch data, only two systems with BHs were discovered there (El-Badry et al. 2023a,b). A third one was discovered using pre-release DR4 astrometry (Panuzzo et al. 2024). DR4 is expected to yield a large number of systems with dark companions and a review similar to this one in the next beach meeting will likely be very different.

## 4. Photometric variability

*Gaia* DR3 contained epoch photometry for only $\sim 1\%$ of the sample, with the targets selected on the basis of their (suspected) variability. That epoch photometry is the basis for the classification of 12.4 million photometric variables by Rimoldini et al. (2023), which included classes (mostly) specific to massive stars such as $\alpha$ Cyg, $\beta$ Cep, and Be (+ other emission line stars) and other more generic ones such as eclipsing binaries (see above) and long-period variables that included some massive stars. A comparison with TESS data reveals that *Gaia* photometry is highly useful for the classification of variable stars (Hey & Aerts 2024).

One problem with using the Rimoldini et al. (2023) results is that their sample is heavily biased by design towards highly variable stars, leaving no variability data for most stars and without the possibility of doing global mean-variability studies for most regions of the CMD, at least until DR4 becomes available. To address those issues, Maíz Apellániz et al. (2023) devised a method to obtain the photometric dispersion (or variabilty amplitude) in the three

**Figure 3.** (left) CAMD for Galactic stars with good-quality *Gaia* DR3 + 2MASS data located above the 20kK ZAMS extinction track with $R_{5495} = 3.1$. The arrow marks the extinction direction for $E(4405 - 5495) = 1.0$ mag and the labels indicate the location of the RC, RGB, and AGB stars, as smeared by extinction, (right) Same CAMD after applying *Gaia* + 2MASS filters to eliminate non-OB stars.

$G+G_{BP}+G_{RP}$ bands for the sample of *Gaia* stars with $G > 17$ mag and five-parameter astrometric solutions (the majority of them). An example of their results is shown in Fig. 2, where the mean $G$ photometric dispersion for 413 462 bright LMC stars is shown, allowing for the identification of variable and non-variable stars alike. As a comparison, less than 13% of those stars are in Rimoldini et al. (2023).
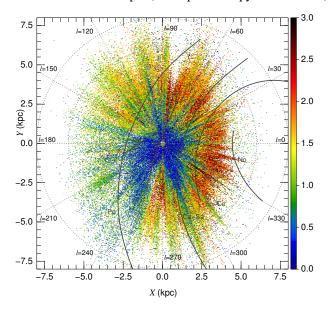
## 5. Stellar groups

*Gaia* has had a tremendous impact on the study of stellar groups (clusters and associations). One important area has been the census of open clusters and the determination of their characteristics (Cantat-Gaudin et al. 2018; Castro-Ginard et al. 2020, 2022; Hunt & Reffert 2023, to name some of the relevant studies). Another one has been the determination of the internal kinematics of stellar groups, with the evidence piling up against a monolithic model of star formation of clusters followed by a posterior dispersion and in favour of a hierarchical model with stars being born both in clusters of different sizes and in looser associations (Carraro 2020; Wright 2020; Ward et al. 2020; Quintana & Wright 2022; Vaher et al. 2023).

Specifically regarding stellar groups with massive stars, the Villafranca project (Maíz Apellániz et al. 2020, 2022b, 2025) is obtaining the membership of the most relevant targets with OB stars and properly deriving their distances using the calibration described in subsection 1.1. One of its most significant results is the discovery that the Bermuda cluster, the origin of the ionising stars of the North America and Pelican nebulae, is dissolving not because of the loss of its gas but because of three different dynamical events that ejected its most massive stars, leaving behind a expanding orphan cluster (Maíz Apellániz et al. 2022a).

## 6. Sample and spatial distribution

Up to DR3, the most important contribution of *Gaia* to massive-star research has been the calculation of accurate and precise distances to stars in the solar neighborhood. This in turn has led to studies of their spatial distribution and of their sample, as it is now possible to build relatively accurate color-absolute magnitude diagrams (CAMDs) and from there to identify

new massive stars to some degree. At the same time, an increased confusion in terminology has appeared. Traditionally, most† massive stars were identified as OB stars, meaning O-B2 dwarfs, O-B5 giants, and O-B8 supergiants, leaving the much more numerous mid- and late-B dwarfs out, as they are for the most part of intermediate mass (Maíz Apellániz et al. 2026). However, *Gaia* has prompted a number of searches for stars in the upper left part of the CAMD (Chen et al. 2019; Liu et al. 2019, 2024; Poggio et al. 2021; Xu et al. 2021; Zari et al. 2021; Li & Binney 2022; Xiang et al. 2022; Khalatyan et al. 2024; Quintana et al. 2025) under names such as "Upper Main Sequence stars", "OBA stars", "B-type candidates", and even as "OB stars" when most of their sample is actually made out of late B dwarfs. The reader is referred to ALS III (Pantaleoni González et al. 2025), where we have compared those studies, indicating which ones are aware of that distinction and/or try to eliminate non-massive stars from the sample (with spectroscopy or otherwise) and which ones are/do not.



**Figure 4.** Spatial distribution of the sample in the right panel of Fig. 3 colour-coded by $G_{BP} - G_{RP}$ as a proxy for extinction.

Most of the studies mentioned above easily detect the Carina-Sagittarius, Local, and Perseus spiral arms (the latter, only clearly seen for $l < 140°$, Negueruela & Marco 2008). The Poggio et al. (2021) and Zari et al. (2021) samples, though dominated by intermediate-mass stars, also detect the Scutum-Centaurus spiral arm. Pantaleoni González et al. (2021) discovered the Cepheus arm as an interarm structure elevated over the Galactic plane. Kuhn et al. (2021) detected a similar but shorter high-pitch angle structure in the Carina-Sagittarius arm.

ALS III lists 15 648 high-probability Galactic massive stars. Other studies that rely as ours on filtering with spectroscopy or other methods (Chen et al. 2019; Xu et al. 2021) reach similar or lower numbers. Others that are more lax on their selection criteria can reach values more than an order of magnitude higher† (see Table I in ALS III). The question is that, given that the Galactic massive-star population is expected to be about two orders of magnitude higher than that, how can we use *Gaia* to significantly increase the known sample? One answer (Maíz Apellániz in prep.) is to combine *Gaia* photometry and astrometry with 2MASS photometry to filter out the low+intermediate mass evolved population that dominates the upper right part of the Galactic CAMD in the form of red clump (RC), red giant branch (RGB), and asymptotic giant branch (AGB) stars. The effect of such a filter is seen by comparing the two panels in Fig. 3: after applying the filters we are left with the expected distribution of a population of OB stars first extinguished and then limited in magnitude as we move towards redder colours. Even if the contamination from intermediate-mass stars is of the order of 50%, the filtering process yields $\sim 10^5$ stars. Figure 4 plots the filtered population projected onto the Galactic Plane, colour-coded by $G_{BP} - G_{RP}$, where the increase in extinction at the location of the Scutum-Centurus arm is easily visible. With the better quality of DR4 data, the number of stars filtered by this method is expected to be several times higher, thus increasing the number of known Galactic massive stars by an order of magnitude from the current number.

---

† To these one has to add (most) A-M supergiants, (most) WR stars, and a zoo of other minor classes.
    † Some are explicit about their sample being dominated by non-massive stars, others are not.

## 7. Summary

I have reviewed the previous and future contributions of *Gaia* to the study of massive stars: parameter determination, binaries, photometric variability, stellar groups, and sample and spatial distribution. Of the previous ones, likely the most important one has been the ability to derive for the first time precise and accurate distances to massive stars within several kpc of the Sun, thus eliminating many uncertainties in their properties. Of the ones expected for DR4, two stand out: the likely discovery of a significant number of NS and BH companions and the increase of the sample of known Galactic massive stars by an order of magnitude.

## References

Andrews, J. J. et al. 2019, ApJ, 886, 68
Bailer-Jones, C. A. L. et al. 2013, A&A, 559, A74
Bailer-Jones, C. A. L. et al. 2021, AJ, 161, 147
Blomme, R. et al. 2023, A&A, 674, A7
Brown, A. G. A. et al. 2021, A&A, 649, A1
Cantat-Gaudin, T. & Brandt, T. 2021, A&A, 649, A124
Cantat-Gaudin, T. et al. 2018, A&A, 618, A93
Carraro, G. 2020, BAAA, 61C, 11
Castro-Ginard, A. et al. 2020, A&A, 635, A45
Castro-Ginard, A. et al. 2022, A&A, 661, A118
Chen, B. Q. et al. 2019, MNRAS, 487, 1400
Creevey, O. L. et al. 2023, A&A, 674, A26
El-Badry, K. et al. 2023a, MNRAS, 521, 4323
El-Badry, K. et al. 2023b, MNRAS, 518, 1057
Fabricius, C. et al. 2021, A&A, 649, A5
Fitzpatrick, E. L. 1999, PASP, 111, 63
Fouesneau, M. et al. 2023, A&A, 674, A28
Frémat, Y. 2024, in EES2023, 59
Frémat, Y. et al. 2023, A&A, 674, A8
Hey, D. & Aerts, C. 2024, A&A, 688, A93
Holgado, G. et al. 2025, A&A, 701, A246
Hunt, E. L. & Reffert, S. 2023, A&A, 673, A114
Janssens, S. et al. 2022, A&A, 658, A129
Khalatyan, A. et al. 2024, A&A, 691, A98
Kuhn, M. A. et al. 2021, A&A, 651, L10
Li, C. & Binney, J. 2022, MNRAS, 516, 3454
Lindegren, L. et al. 2018, A&A, 616, A2
Lindegren, L. et al. 2021a, A&A, 649, A4
Lindegren, L. et al. 2021b, A&A, 649, A2
Liu, Z. et al. 2019, ApJS, 241, 32
Liu, Z. et al. 2024, ApJS, 275, 24
Luri, X. et al. 2018, A&A, 616, A9
Maíz Apellániz, J. 2001, AJ, 121, 2737
Maíz Apellániz, J. 2005, in ESASP, Vol. 576, 179
Maíz Apellániz, J. 2017, A&A, 608, L8

Maíz Apellániz, J. 2022, A&A, 657, A130
Maíz Apellániz, J. 2024, arXiv:2401.01116
Maíz Apellániz, J. & Barbá, R. H. 2018, A&A, 613, A9
Maíz Apellániz, J. & Weiler, M. 2018, A&A, 619, A180
Maíz Apellániz, J. & Weiler, M. 2025, in HSA XII, 223
Maíz Apellániz, J. et al. 2008, arXiv:0804.2553
Maíz Apellániz, J. et al. 2011, in HSA VI, 467–472
Maíz Apellániz, J. et al. 2014, A&A, 564, A63
Maíz Apellániz, J. et al. 2016, ApJS, 224, 4
Maíz Apellániz, J. et al. 2020, A&A, 643, A138
Maíz Apellániz, J. et al. 2021, A&A, 649, A13
Maíz Apellániz, J. et al. 2022a, A&A, 657, A72
Maíz Apellániz, J. et al. 2022b, A&A, 657, A131
Maíz Apellániz, J. et al. 2023, A&A, 677, A137
Maíz Apellániz, J. et al. 2025, in HSA XII, 225
Maíz Apellániz, J. et al. 2026, in Encyclopedia of Astrophysics, Vol. 2, 43–84
Mowlavi, N. et al. 2023, A&A, 674, A16
Negueruela, I. & Marco, A. 2008, A&A, 492, 441
Pantaleoni González, M. et al. 2021, MNRAS, 504, 2968
Pantaleoni González, M. et al. 2025, MNRAS, 543, 63
Panuzzo, P. et al. 2024, A&A, 686, L2
Poggio, E. et al. 2021, A&A, 651, A104
Prusti, T. et al. 2016, A&A, 595, A1
Quintana, A. L. & Wright, N. J. 2022, MNRAS, 515, 687
Quintana, A. L. et al. 2025, MNRAS, 538, 1367
Riello, M. et al. 2021, A&A, 649, A3
Rimoldini, L. et al. 2023, A&A, 674, A14
Vaher, E. et al. 2023, A&A, 679, A105
Vallenari, A. et al. 2023, A&A, 674, A1
Ward, J. L. et al. 2020, MNRAS, 495, 663
Weiler, M. et al. 2020, A&A, 637, A85
Weiler, M. et al. 2023, A&A, 671, A52
Wright, N. J. 2020, NAR, 90, 101549
Xiang, M. et al. 2022, A&A, 662, A66
Xu, Y. et al. 2021, A&A, 645, L8
Zari, E. et al. 2021, A&A, 650, A112