

# The Alma catalogue of OB stars – II. A cross-match with *Gaia* DR2 and an updated map of the solar neighbourhood

M. Pantaleoni González <sup>1,2</sup>★ J. Maíz Apellániz <sup>1</sup> R. H. Barbá <sup>3</sup> and B. Cameron Reed<sup>4</sup>

<sup>1</sup>Centro de Astrobiología, CSIC-INTA, Campus ESAC, Camino bajo del castillo s/n, E-28692 Villanueva de la Cañada, Madrid, Spain

<sup>2</sup>Departamento de Astrofísica y Física de la Atmósfera, Universidad Complutense de Madrid, E-28040 Madrid, Spain

<sup>3</sup>Departamento de Astronomía, Universidad de La Serena, Av. Cisternas 1200 Norte, La Serena, Chile

<sup>4</sup>Department of Physics (Emeritus), Alma College, Alma, MI 48801, USA

Accepted 2021 March 3. Received 2021 February 18; in original form 2020 November 30

## ABSTRACT

We cross-match the Alma catalogue of OB stars with *Gaia* DR2 astrometry and photometry as a first step towards producing a clean sample of massive stars in the solar neighbourhood with a high degree of completeness. We analyse the resulting colour–absolute magnitude diagram to divide our sample into categories and compare extinction estimates from two sources, finding problems with both of them. The distances obtained with three different priors are found to have few differences among them, indicating that *Gaia* DR2 distances are robust. An analysis of the 3D distribution of massive stars in the solar neighbourhood is presented. We show that a kinematically distinct structure we dub the Cepheus spur extends from the Orion–Cygnus spiral arm towards the Perseus arm and is located above the Galactic mid-plane, likely being related to the recently discovered Radcliffe wave. We propose that this corrugation pattern in the Galactic disc may be responsible for the recent enhanced star formation at its crests and troughs. We also discuss our plans to extend this work in the immediate future.

**Key words:** stars: massive – catalogues – astrometry – (Galaxy:) solar neighbourhood – Galaxy: structure.

## 1 INTRODUCTION

The short lifespans of massive OB stars imply that they can be used as tracers of the large stellar formation regions and spiral arms of the Galaxy (Morgan, Sharpless & Osterbrock 1952; Reed 1993b; Bouy & Alves 2015; Ward, Kruijssen & Rix 2020), with the exception of a small fraction of massive runaway stars (Blaauw 1993; Maíz Apellániz et al. 2018). Going back to the studies of the 1950s, OB stars have been preferentially identified from spectroscopy. Photometric identifications can be problematic, especially for O stars, unless one has good-quality photometry to the left of the Balmer jump (Maíz Apellániz & Sota 2008; Maíz Apellániz et al. 2014), where the atmosphere complicates calibration for ground-based observations (Maíz Apellániz 2006). One solution to this problem is to leave the atmosphere behind and that is one of the reasons for the exquisite calibration of *Gaia* photometry (Jordi et al. 2010). However, the DR2 (2018) and EDR3 (2020) *Gaia* versions do not provide the full spectrophotometry that will become available in DR3 (expected for 2022) and therefore there is little information about the spectral energy distribution (SED) to the left of the Balmer jump in the current data (Maíz Apellániz & Weiler 2018 and Appendix B here).

The ‘Alma Luminous Star’ (ALS) catalogue (Reed 2003, hereafter Paper I) is currently the largest published compilation of known Galactic luminous stars with available  $UBV\beta$  photometric data. The ALS catalogue was primarily built on the ‘Case-Hamburg’ (C-S) survey for Galactic luminous stars, by joining data from Stephenson & Sanduleak (1971) in the south, with the northern

component of the survey, published in six separate volumes (Hardorp et al. 1959; Stock, Nassau & Stephenson 1960; Nassau & Stephenson 1963; Hardorp, Theile & Voigt 1964, 1965; Nassau, Stephenson & McConnell 1965). Since its first appearance (Reed 1993a), the ALS catalogue has grown to include more than 680 references. Since the publication of Paper I, over a decade and a half ago, a tremendous amount of information about OB stars has become available so this is an opportune moment to update the ALS catalogue. In this second paper of the series we cross-match the sample from Paper I with the astrometric and photometric information from *Gaia* DR2. In future papers of the series we will update the *Gaia* information with that of subsequent data releases and include data from other photometric and spectroscopic surveys such as spectra and uniform spectral classifications.

In the next section, we describe our data and methods: how we have built the new version of the catalogue and what its contents are. We then present our results about the colour–colour and colour–absolute magnitude diagrams of OB stars, a comparison between different extinction and distance estimates, a study of the distribution of OB stars in the solar neighbourhood, and an analysis of a new structure we dub the Cepheus spur. We end the paper with a description of our plans for future papers of the series.

## 2 DATA AND METHODS

### 2.1 Building the new version of the catalogue

Cross-matching a compilation of old sources with a modern and uniform data base such as *Gaia* DR2 is not straightforward, as there

\* E-mail: mpantaleoni@cab.inta-csic.es

**Table 1.** Number of objects by category in this paper. The first block details the cleaning process until arriving to the final sample of 15 662 stars and the second block the breakdown of the final sample by categories.

Cat.	Description	#
	Objects in the 2005 version of the ALS catalogue	18 693
	After eliminating 393 duplicates (D)	18 300
	After eliminating 211 unmatched objects (U)	18 089
	After eliminating 2336 stars with bad astrometry (A)	15 753
	After eliminating 91 objects with bad colours (C)	15 662
M	Likely massive stars	13 762
I	High/intermediate-mass stars	1506
L	Intermediate/low-mass stars	260
H	High-gravity stars	127
E	Extragalactic stars	7

are different ways in which we can have false positives and false negatives: low-quality old coordinates, objects with large proper motions, duplicates, and erroneous identifications are some of them. To those, one has to add the cases where the star should not have been in the original catalogues in the first place because it did not really belong to its purported class of objects (which was unknown at the time). Therefore, there is a significant amount of detective work involved in the process if one desires a clean new version. A description of what we have done is provided in this subsection and details about the different types of problems are provided in Appendix A.

We use here the traditional definition of an OB star: a massive star ( $\gtrsim 8 M_{\odot}$ ) with an O or B spectral type. This corresponds to spectral types up to B2 for dwarfs, B5 for giants, and B9 for supergiants. In principle, we exclude from the definition low-luminosity hot stars such as white dwarfs and subdwarfs, intermediate-type (AF) and late-type (GKM) supergiants, Wolf–Rayet stars, and others. Also, as we intend the catalogue to be a Galactic one we plan to separate extragalactic objects from the main sample. Of course, excluding from the definition is not the same as excluding from the sample, the latter being more difficult due to the diverse quality of the data. For that reason, we will proceed with caution on this series as to how we determine what each object is, placing them in this first paper in temporary categories based on *Gaia* DR2 photometric and astrometric information alone and leaving for subsequent papers a permanent assignment based on further spectroscopic and photometric data. Also, objects determined to be e.g. A supergiants will not be completely excluded from the catalogue in future papers but instead placed on appropriate supplements. Our goal in this first paper then is to produce a clean version of the ALS catalogue with good-quality photometric and astrometric information and where we can use it to produce a preliminary classification of the objects into likely massive stars and other types.

We start the cleaning process with the current version of the ALS catalogue (updated in 2005), which contains 18 693 entries (Table 1). Most of the objects correspond to OB stars but  $\sim 10$  per cent of the entries correspond to other types of objects, which Paper I indicates they are mostly A–G supergiants, white dwarfs (WDs), planetary nebula nuclei (PNNi), and Wolf–Rayet stars (WRs). 393 of these are duplicates: 298 flagged as so in the ALS catalogue itself, 72 matched by the CDS on Simbad’s set of identifiers, and 23 additional duplicates found by us in this work. Not all duplicates are pairs of identifiers, the term ‘duplicate’ here includes six triplets in the original catalogue. We remove the duplicates from the new version of the catalogue and flag them with a D, leaving the number of unique objects at 18 300. These 393 duplicates do not always reflect identical

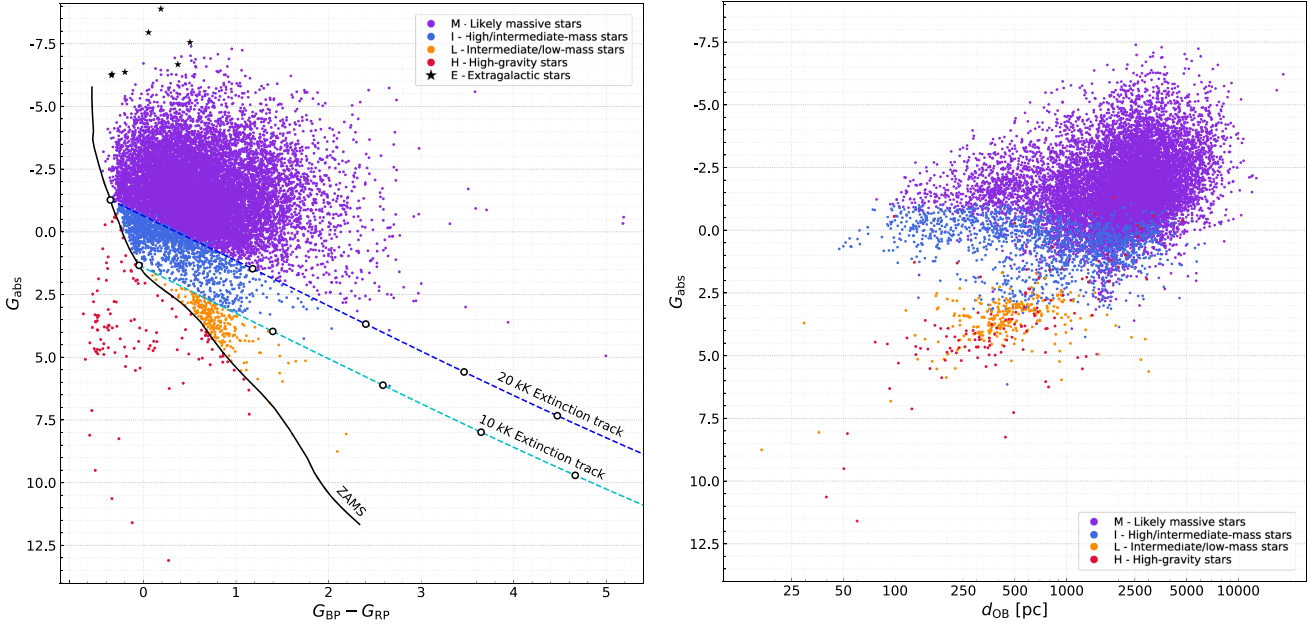
instances of the same object, as they are drawn from different parts of the literature and can also appear in situations where modern references account for multiple objects (or a particular component of a multiple system) while older references could not resolve the parts of the system, in which case we consider the identifier linked to the old reference as a duplicate of the brightest component in the set of resolved sources. Multiplicity, spectroscopic or visual, is a ubiquitous characteristic of massive stars (Mason et al. 1998; Sota et al. 2014; Maíz Apellániz et al. 2019b) and this leads to the question of how to name multiple components in the catalogue (with independent entries or with A,B,C...extensions). We will not address it in this paper but will do so in future articles in this series.

After eliminating duplicates, we proceed to cross-match the ALS catalogue with *Gaia* DR2 sources. Currently, all ALS identifiers can be queried in Simbad with the exception of ALS 1823, ALS 12 636, ALS 15 196, ALS 15 858, and ALS 19 764, which are recognized in Simbad as CD –58 3529, BD +61 2352, Trumpler 14 9, CPD –58 2655, and LS 4723, respectively. Including those five objects by hand, 16 912 of the 18 300 unique ALS entries are currently cross-matched in *Gaia* DR2 by Simbad. For the remaining 1388 entries we search for *Gaia* DR1, *Hipparcos*, *Tycho*, *2MASS*, and *WISE* identifiers in Simbad and then we use the cross-match products of the *Gaia* Archive to retrieve a *Gaia* DR2 source identifier. This can be successfully done for an additional 684 ALS identifiers. Finally, we address the remaining 704 unmatched cases manually by examining the available references attached to the ALS and by searching for astrophotometrically similar sources within a few arcseconds and small magnitude differences (the margins of these criteria are a function of crowding). To accomplish those matches we used Aladin and for high proper motion objects the back-propagation of the trajectories was executed before the separation threshold was applied. In this way, 493 additional ALS sources were successfully matched with *Gaia* DR2. This leaves us with only 211 unique unmatched sources and a total of 18 089 matched ones (98.8 per cent of the unique sources).

Why are there 211 unmatched ALS sources in *Gaia* DR2? There are different reasons. Some are too bright in *G* to be included in *Gaia* DR2 (e.g. ALS 14 793 =  $\zeta$  Ori A) and some are too faint (e.g. ALS 19 589). Others are in crowded regions like NGC 3603, where in some cases it is unclear which source the original paper refers to. We list those objects in a supplemental table but leave them out of our final sample for this paper. It is our plan to incorporate at least some of them into the catalogue in future papers.

The next step is the astrometric quality control. There are 2343 remaining objects with either (a) no *Gaia* DR2 parallaxes ( $\varpi$ ), (b) large uncertainties in the parallax ( $\varpi/\sigma_{\varpi} < 3$ ), or (c) bad astrometric solutions (RUWE  $> 1.4$ ). We eliminate all of those from the sample except for seven which are located in the Magellanic Clouds (see below) to arrive at a number of 15 753 objects.

The last step to arrive at the final sample is the photometric quality control, that is, the elimination of targets with incomplete (one or more missing) or bad-quality *Gaia* DR2  $G_{BP} + G + G_{RP}$  magnitudes. With respect to what we consider as bad-quality photometry,  $G_{BP}$  and  $G_{RP}$  magnitudes are expected to be contaminated in crowded fields and bright sources may be saturated in one or more of the three bands. Those two are the reasons we have to apply this step, for which we selected the criterion  $|d_{CC}| > 0.15$  and eliminated a total of 91 objects with bad colours.  $d_{CC}$  is the distance from the stellar locus in the  $G_{BP}-G' + G' - G_{RP}$  plane and  $G'$  is the *Gaia* DR2 corrected *G* magnitude (Maíz Apellániz & Weiler 2018; Maíz Apellániz 2019). In future papers, with information from other sources and new *Gaia*



**Figure 1.** (left) *Gaia* DR2  $G_{BP}-G' + G_{abs}$  colour–absolute magnitude diagram for the final sample in this paper. The sample is colour-coded by category. The three boundaries used to divide the sample in the four Galactic categories are shown. (right) Distance–absolute magnitude diagram for the same sample without the seven extragalactic stars and ALS 2481 (= Gliese 440, a white dwarf at a distance of 4.6 pc). Markings on the extinction curves correspond to  $E(4405 - 5495) = 0, 1, 2, 3,$  and 4 mag.

data releases, most of these objects should be reintegrated into the main part of the catalogue.

Having arrived at our final sample, we calculate  $G_{abs}$  using the distance  $d_{OB}$  described below in subsection 3.3.  $G_{abs}$  is computed from  $G'$  but is not corrected from extinction. Exceptions for the distance are made for seven extragalactic objects in the Magellanic Clouds that were inadvertently included in the original catalogue (ALS 15 895, ALS 15 896, ALS 18 185, ALS 18 840, ALS 18 845, ALS 19 597, and ALS 19 598), for which we use distances of 62 kpc (SMC, ALS 19 597) and 50 kpc (LMC, rest of the sample). We use the  $G_{BP} - G_{RP}$  versus  $G_{abs}$  colour–absolute magnitude diagram in Fig. 1 to classify the final sample into five categories: likely massive stars (M), objects in the high/intermediate mass regime (I), objects that are likely to be intermediate- or low-mass stars (L), high-gravity stars (H), and extragalactic objects (E). The four Galactic categories are defined using as boundaries the zero-age main sequence (ZAMS) from the solar metallicity grid of Maíz Apellániz (2013b) and the extinction tracks for ZAMS stars with 10 and 20 kK with the  $R_{5495} = 3.0$  extinction law of Maíz Apellániz et al. (2014). The  $R_{5495} = 3.0$  value is chosen from Maíz Apellániz & Barbá (2018) as representative of the typical intermediate-to-high extinction for OB stars (note that in the low-extinction case higher values are more common). The ZAMS separates high-gravity stars (white dwarfs and subdwarfs) from the rest of the sample. The values for the  $T_{eff}$  of the extinction tracks are selected to represent the approximate value for the A0 V and B2.5 V spectral subtypes in the MS, respectively. Note that evolved OB stars such as intermediate-B giants and later type supergiants are also located above the 20 kK track independent of their extinction due to their intrinsic  $G_{abs}$  values. We repeat here that these five are temporary categories and that the membership to each one of them will be reanalysed in subsequent papers. For example, some type L stars may be extinguished H-type objects. For the purposes of this paper we will use the type M stars as our

‘OB massive star sample’ but we remind the reader that it has a few contaminants in the form of later type supergiants.

## 2.2 Catalogue description

The result of the cleaning and classification process are presented in the form of a main catalogue and three supplements. The main catalogue lists our final sample of 15 662 stars and the three supplements lists the 2427 stars excluded due to bad astrometry or colours, the 211 unmatched objects, and the 393 duplicates. The information in the main catalogue is described in Table 2. The supplements give the same information as the catalogue whenever applicable. The catalogue and supplements are available in electronic format only due to the large number of columns.

The codes used for the category in Table 2 are the same as those in Table 1. The codes for the Sflag indicate if Simbad successfully matches the ALS identifier with the correct *Gaia* DR2 identifier (S), if the match is different from ours (W), or if it did not perform a match at all (N). The codes for the Pflag indicate the photometric type in the ALS: V when the magnitude was retrieved from photoelectric measurements, P when it was obtained from photographic plates, and N when the original ALS lacked photometry.

## 3 RESULTS

### 3.1 The colour–colour and colour–absolute magnitude diagrams

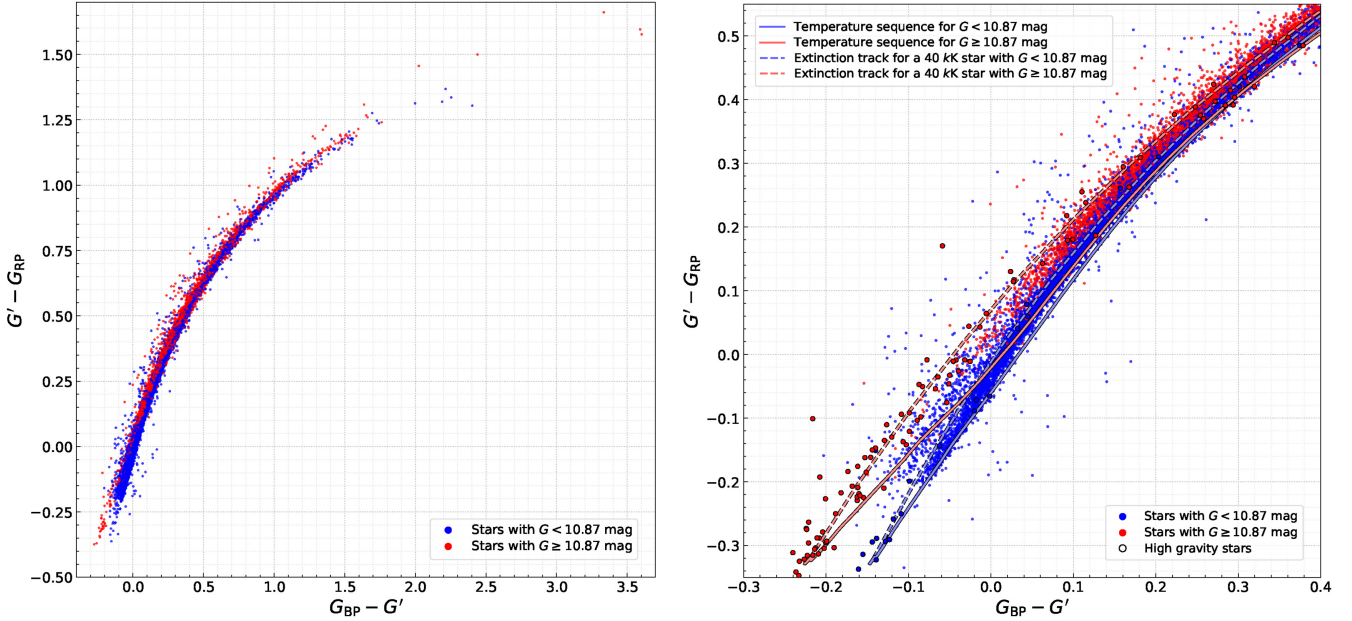
We plot in the left-hand panel of Fig. 2 the  $G_{BP}-G' + G'-G_{RP}$  colour–colour diagram for the full clean sample of 15 662 stars in the paper, colour-coded by their membership to the bright or faint magnitude ranges. The intrinsic colours of the vast majority of the sample are negative (lower left part of the plot) and the curved

**Table 2.** Description of the contents in the table sent to the CDS with the final sample.

Column name	Units	Description
		Category and identifiers
Cat	–	Category (M/IL/HE)
ID_ALS	–	Alma Luminous Star number
ID_DR2	–	<i>Gaia</i> DR2 source identifier
ID_GOSC	–	Galactic O-Star Catalogue identifier
		Coordinates
RA_ALS	hh:mm:ss	Right ascension (J2000.0) from original ALS catalogue
DEC_ALS	dd:mm:ss	Declination (J2000.0) from original ALS catalogue
RA_DR2	hh:mm:ss	Right ascension (Epoch 2015.5) from <i>Gaia</i> DR2
DEC_DR2	dd:mm:ss	Declination (Epoch 2015.5) from <i>Gaia</i> DR2
GLON	deg	Galactic longitude from <i>Gaia</i> DR2
GLAT	deg	Galactic latitude from <i>Gaia</i> DR2
		Photometry
Vmag	mag	<i>V</i> magnitude from original ALS catalogue
Gmag	mag	<i>G</i> magnitude in <i>Gaia</i> DR2
Gmag_cor	mag	$G'$ , corrected <i>G</i> magnitude
BPmag	mag	$G_{BP}$ magnitude in <i>Gaia</i> DR2
RPmag	mag	$G_{RP}$ magnitude in <i>Gaia</i> DR2
BPmag-Gmag_cor	mag	$G_{BP}-G'$ colour
Gmag_cor-RPmag	mag	$G'-G_{RP}$ colour
		Quality indicators
CCDIST	mag	$d_{CC}$ , distance to the <i>Gaia</i> DR2 colour–colour main locus
Sep_ast	arcsec	Separation between ALS and <i>Gaia</i> DR2 positions
RUWE	–	Renormalized Unit Weight Error in <i>Gaia</i> DR2
		Proper motions
PM_RA	mas/a	Proper motion in right ascension in <i>Gaia</i> DR2
PM_RA_err	mas/a	Proper motion uncertainty in right ascension in <i>Gaia</i> DR2
PM_DEC	mas/a	Proper motion in declination in <i>Gaia</i> DR2
PM_DEC_err	mas/a	Proper motion uncertainty in declination in <i>Gaia</i> DR2
		Parallax and distance
Plx	mas	$\varpi$ , parallax in <i>Gaia</i> DR2
Plx_err	mas	$\sigma_{\varpi}$ , parallax uncertainty in <i>Gaia</i> DR2
ALS2_dist_mode	pc	Mode of the posterior distribution of distances
ALS2_dist_P50	pc	$d_{OB}$ , median of the posterior distribution of distances
ALS2_dist_mean	pc	Mean of the posterior distribution of distances
ALS2_dist_HDl	pc	Lower limit of the highest density interval for $p = 68$ per cent
ALS2_dist_HDh	pc	Upper limit of the highest density interval for $p = 68$ per cent
ALS2_dist_P16	pc	16th percentile of the posterior distribution of distances
ALS2_dist_P84	pc	84th percentile of the posterior distribution of distances
BJ_dist	pc	$d_{BJ}$ , Bailer-Jones distance estimate
BJ_dist_l	pc	Bailer-Jones lower distance estimate
BJ_dist_h	pc	Bailer-Jones upper distance estimate
SH_dist	pc	$d_{SH}$ , StarHorse distance estimate
SH_dist_l	pc	StarHorse lower distance estimate
SH_dist_h	pc	StarHorse upper distance estimate
		Absolute magnitude
Gabs_mode	mag	$G_{abs}$ from the mode of the posterior as a distance estimate
Gabs_mode_err	mag	$G_{abs}$ uncertainty from the highest density interval for $p = 68$ per cent
Gabs_P50	mag	$G_{abs}$ from the median of the posterior as a distance estimate
Gabs_P1684	mag	$G_{abs}$ uncertainty from the 16th and 84th percentiles
		Other information
SpT_ALS	–	Spectral classifications from original ALS catalogue
Sflag	–	Simbad cross-match flag (N/S/W)
Pflag	–	<i>V</i> magnitude type from original ALS catalogue (N/P/V)
Other_crossmatch	–	Other <i>Gaia</i> DR2 cross-match candidates
Comments	–	Notes

sequence we see in the plot is for the most part an extinction sequence. The curvature is caused by the need to integrate over very broad-band filters to accurately calculate extinction. If non-linear extinction effects are ignored (e.g. by assuming that  $A_G$  is

proportional to  $E(G_{BP} - G_{RP})$ ), significant biases can be introduced in the result (Maíz Apellániz 2013a; Maíz Apellániz & Barbá 2018; Maíz Apellániz et al. 2020a). Furthermore, extinction not only reddens colours but also makes stars fainter overall (again, non-



**Figure 2.** *Gaia* DR2  $G_{BP}-G' + G'-G_{RP}$  colour-colour diagram for the final sample in this paper. The points are colour-coded depending on their value of  $G$  (bright or faint stars). The left-hand panel shows the full range spanned by the sample, while the right-hand panel is a zoom into the low-extinction region. In the right-hand panel the high-gravity subsample is shown with additional black circumferences and the lines show the (a)  $T_{\text{eff}}$  MS zero-extinction sequences and (b) the  $R_{5495} = 3.0$  extinction tracks for an MS star with  $T_{\text{eff}} = 40$  kK for both the bright and faint magnitude ranges using the calibration of Maíz Apellániz & Weiler (2018).

linearly, the extinction tracks in Fig. 1 are slightly curved). For that reason, most stars in the lower left part of the plot are bright while faint stars concentrate in the central and upper right regions (of course, there are also selection effects involved). The right-hand panel of Fig. 2 is a zoom into the bottom left region of the left-hand panel, where all OB stars should be if it were not for extinction, and the differences seen there as a function of magnitude range are a *Gaia* DR2 calibration issue, as described in Appendix B.

As already mentioned, the colour-absolute magnitude diagram in Fig. 1 is used to classify the Galactic part of the sample into four categories, with seven additional objects being located in the Magellanic Clouds. The numbers in each category are given in Table 1. As expected, the vast majority (13 762/15 662 or 87.9 per cent) are classified as type M and most of the rest (9.6 per cent) are of type I, that is, objects near the real boundary between the two categories. Only 2.5 per cent of the sample is clearly excluded, indicating that the original ALS catalogue was a relatively clean sample but not a perfect one, especially considering that our procedure does not allow us to discriminate between OB stars and later type supergiants, which are known to be contaminants.

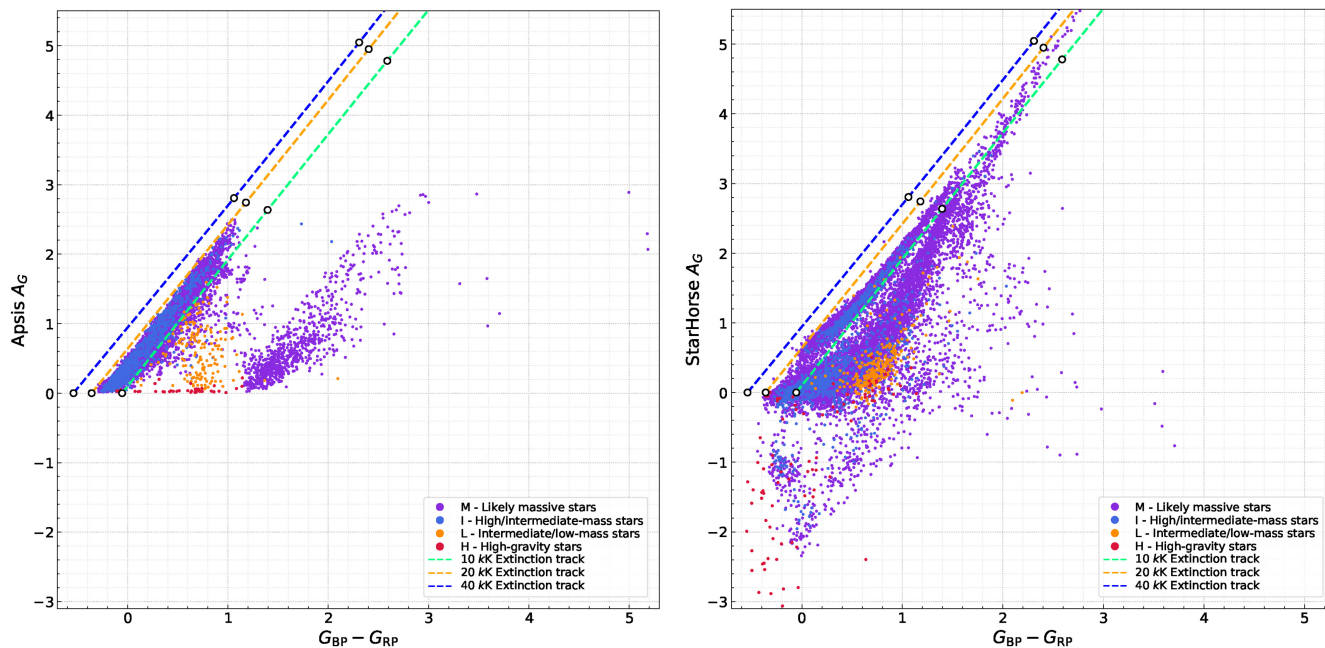
The spread on the M- (and I-) type objects in the horizontal direction in Fig. 1 is due to a combination of two effects. The first one is the spread in intrinsic colours caused by the different values of  $T_{\text{eff}}$  for OB stars (plus the later type supergiant contaminants). However, A0 stars have  $G_{BP} - G_{RP} \sim 0.0$  so this effect is a minor one. The second effect is the largest one and is extinction, which is analysed in more detail in the next subsection.

Another feature seen in the colour-absolute magnitude diagram is the different behaviour of the I- and M-type objects near the main sequence. Intermediate-mass stars are seen relatively close to the ZAMS with a small gap caused by the stars experiencing small amounts of extinction or evolution from their zero age. As we move to M-type objects the gap widens significantly as a combination of four

effects: (a) As seen in the right-hand panel of Fig. 1, close to the Sun I-type objects are more abundant than M-type ones, thus increasing the chances of finding more low-extinction I-types than M-types. (b) Stars near the top of the diagram are more likely to be B supergiants than extinguished O stars because the bolometric correction makes them intrinsically brighter in  $G$  at constant luminosity. (c) The IMF and the MS lifetimes decrease as we move upwards in the diagram. (d) Finally, O-type stars near the ZAMS are hard to find anywhere in the Galaxy, especially the earlier subtypes (Holgado et al. 2020). These four effects will be discussed in future instalments of this series.

The position of the seven MC objects in the colour-absolute magnitude diagram also deserves comment. Four of them are among the brightest objects in  $G_{\text{abs}}$ , which we remind the reader is not extinction-corrected, and the other three are in the gap close to the ZAMS described in the previous paragraph. In both cases the low extinction of most stars in the Magellanic Clouds compared to the OB Galactic sample plays a role in placing those stars towards the upper left corner of the diagram. In the case of the first four objects a second explanation is that they are of spectral type B1a, i.e. very luminous supergiants with small bolometric corrections in the  $G$  band. In the case of the last three objects, they are two early O-stars and an O + O binary, placing them close to the extinction-free leftmost possible location in the colour-absolute magnitude diagram.

Finally, it is instructive to compare our colour-absolute magnitude diagram with fig. 1 of Babusiaux et al. (2018), which is the equivalent diagram using the 65 921 112 stars with good-quality *Gaia* DR2 data independently of their spectral types. As a first-order approximation, both diagrams are complementary: the region with a high density of objects in Fig. 1 here appears as a low-density region in the diagram of the general *Gaia* DR2 population, with most stars there located towards the right and bottom. This is a manifestation of what dominates the *Gaia* DR2 population: low-mass stars near the



**Figure 3.** Comparison between the values for the extinction  $A_G$  as a function of  $G_{BP}-G_{RP}$  obtained using Apsis (left) and StarHorse (right) for the final sample. In both cases we also plot the expected relationship between the two quantities for MS stars of 10, 20, and 40 kK and a extinction law from Maíz Apellániz et al. (2014) with  $R_{5495} = 3.0$ . Markings on the extinction curves correspond to  $E(4405 - 5495) = 0, 1, \text{ and } 2$  mag.

main sequence and red giants (most noticeably in the diagram, red-clump stars, clearly seen as an extinction sequence). This in turn is a consequence of the very large numbers of the first type and of the right combination of high (but not as large as the first) numbers and high luminosities for the second type. However, a second, more subtle effect is present in the comparison between the two diagrams. For luminous stars the transition from low density to high density is rather abrupt in the Babusiaux et al. (2018) around  $G_{BP} - G_{RP} = 1.0-1.2$  due to the appearance of the first low-extinction red-clump stars but why should that be accompanied by a complementary transition among the M + I type stars in our diagram? The reason is that OB stars are relatively easy to identify by their colours as long as they are not too extinguished. When they are, they of course become fainter but being intrinsically luminous objects they should still be detected. The real problem is that around  $G_{BP} - G_{RP} = 1.0-1.2$  (or their equivalent in other photometric systems) the much more numerous population of Galactic red giants makes their appearance and severely hampers the identification of OB stars. That is why compilations of relatively old catalogues like the previous version of the ALS have only a few high-extinction objects. Detecting OB stars in the Galactic plane is like finding a needle in a haystack.

### 3.2 Comparing extinction estimates

As discussed in the previous subsection, most of the spread in the horizontal direction in Fig. 1 is caused by extinction. We have previously measured the extinction properties of part of the ALS sample (Maíz Apellániz & Barbá 2018; Maíz Apellániz et al. 2021) but only for a small fraction of it, so we will not consider those results here and we will leave our extinction analysis for a future instalment of this series. Instead, here we analyse the extinction estimates in the  $G$  band ( $A_G$ ) by Andrae et al. (2018) using Apsis (Bailer-Jones et al. 2013) from *Gaia* DR2 alone and by Anders et al. (2019) using StarHorse (Queiroz et al. 2018) combining *Gaia* DR2 information

with Pan-STARRS1, 2MASS, and WISE. We have cross-matched our sample with both of those results and found numbers of 8701 for Apsis and 13 844 for StarHorse, with 7799 objects in common between the three papers. We plot in Fig. 3  $A_G$  as a function of  $G_{BP}-G_{RP}$  for both techniques.

We first analyse the Apsis results in Fig. 3. On the positive side, there is a linear trend between  $G_{BP}-G_{RP}$  and  $A_G$  for the bulk of the stars, indicating that the non-linearity effect of extinction is correctly accounted for. Also, high-gravity stars have low extinctions, stars of type L are in the region of the plot where they are supposed to be, and there are no objects with negative extinctions (a condition expressly imposed by Andrae et al. 2018). On the negative side, Andrae et al. (2018) trained their algorithm using models up to 20 kK for calculating extinction (and up to 10 kK for their calculation of  $T_{\text{eff}}$ ) and, as a consequence, the bulk of their stars is contained within the relationships for 10 and 20 kK, while most of the stars there actually have values of  $T_{\text{eff}}$  between 20 and 40 kK. Therefore, the Andrae et al. (2018)  $A_G$  values are consistently underestimated for most of the M-type stars (and also possibly for I-type stars) with  $G_{BP}-G_{RP} \lesssim 1.1$  in our sample by several tenths of a magnitude. A corollary of this is that the bluest objects in the sample have no Apsis extinctions, as they would require negative values of  $A_G$  that the algorithm does not allow. Another problematic aspect is that the main trend (located in the region expected for stars with  $T_{\text{eff}}$  in the range 10–20 kK) stops around  $G_{BP}-G_{RP} \sim 1.1$  and continues in a parallel track that goes back to zero extinction starting at that same colour and containing about 800 stars. We have verified that a few of those 800 stars are some of the later type supergiants still present among the M-type stars but the majority of them are bona fide OB stars, including a significant number of O stars with accurate GOSSS (Maíz Apellániz et al. 2011) spectral types. The likely explanation for this effect is that the Apsis algorithm considers that the vast majority of objects with  $G_{BP}-G_{RP} \gtrsim 1.1$  cannot be OB stars and are instead assigned a low value of  $T_{\text{eff}}$  and, hence, a lower value of  $A_G$ .

The structures seen in the StarHorse plot of Fig. 3 are very different from those seen in the Apsis plot. In the first place, there are 1502 stars (10.8 per cent of the total) with negative extinction values (non-negativity was not imposed as a condition) extending to values of  $A_G \sim -3$  mag, indicating that at least some OB stars were erroneously identified by the algorithm as being of later spectral type. On the other extreme, we find  $A_G$  values of more than  $\sim 5$  mag for stars with estimated  $T_{\text{eff}}$  above 10 kK. This is as expected but we note such extinction values are missing on the Apsis results (see above). In between the two extinction extremes, two different trends are seen in the  $A_G = 0-3$  mag, one above the other and both with a curvature. The origin of the curvature is likely to be an inaccurate treatment of non-linear extinction effects. As it happened with the Apsis results, there are no stars in the region of the diagram expected for 20–40 kK stars, leading us to suspect that  $A_G$  is underestimated for most of the stars.

In summary, both Andrae et al. (2018) and Anders et al. (2019)  $A_G$  estimates are not optimized for OB stars and for the ALS sample they are in general underestimates and in many cases simply wrong. Both papers do not consider the possibility of stars having  $T_{\text{eff}}$  above 20 kK (which is the case for most OB stars according to the traditional definition) with Andrae et al. (2018) erroneously assigning low values for  $A_G$  for OB stars with high extinction and Anders et al. (2019) yielding erroneous relationships between  $G_{\text{BP}}-G_{\text{RP}}$  and  $A_G$  for OB stars in general. Therefore, we recommend their  $A_G$  values are not used to correct the extinction of OB stars.

### 3.3 Comparing distance estimates

It has been known for a long time (Lutz & Kelker 1973) that using the inverse of the observed parallax to estimate distances leads to biased or even absurd results (as observed parallaxes can be negative). For the case of OB stars, one of us (JMA) developed a Bayesian formalism (Maíz Apellániz 2001, 2005) with a prior based on a self-consistent vertical distribution of the Galactic disc population measured from *Hipparcos* parallaxes. The prior consists of a self-gravitating isothermal thin disc and an extended halo population of runaway stars and depends on Galactic latitude but not on Galactic longitude. Here, we use it to calculate distances for the final ALS sample with the parameters from Maíz Apellániz, Alfaro & Sota (2008) and the following details:

(i) We apply a parallax zero-point of  $40 \mu\text{as}$  and we add  $10 \mu\text{as}$  in quadrature to the *Gaia* DR2 uncertainties (Maíz Apellániz et al. 2020b).

(ii) For objects farther away than  $1^\circ$  from the Galactic plane, the standard mixture of an isothermal thin disc and a halo (with a fraction of stars of 3.9 per cent) is used i.e. we consider both possibilities for any given star, that it is not a runaway or that it is.

(iii) For objects less than  $1^\circ$  from the Galactic plane, we use just an isothermal thin disc (i.e. we neglect the possibility of the star being a runaway) and we use the prior corresponding to  $b = 1^\circ$  for stars in the northern Galactic hemisphere and the prior corresponding to  $b = -1^\circ$  for stars in the southern Galactic hemisphere. This has to be done because the prior was developed using data only from the solar neighbourhood and does not consider the possibility that the Galactic disc has a finite extent.

Alternatively, Bailer-Jones et al. (2018) and Anders et al. (2019) use their own Bayesian formalisms to develop more complex priors that depend on both Galactic coordinates. However, those priors are primarily based on the spatial distribution of late-type stars and, in principle, should not be strictly applicable to OB stars, as their

distribution in the vertical Galactic direction is narrower than that of late-type stars. In summary, we have three different priors but each one of them has different limitations regarding their applicability to the sample in this paper and therefore should yield different distances. We now analyse how large those differences are.

We want to compare the posterior distributions using the three priors described above. We define  $d_{\text{OB}}$ ,  $d_{\text{BJ}}$ , and  $d_{\text{SH}}$  as the median of the posterior for each star using Maíz Apellániz et al. (2008), Bailer-Jones et al. (2018), and Anders et al. (2019), respectively.<sup>1</sup> Given how we have selected the final sample, we can calculate  $d_{\text{OB}}$  for all of the objects in our final sample except for the seven stars in the Magellanic Clouds, that is, for 15 656 stars. The same is true for  $d_{\text{BJ}}$ . However, for  $d_{\text{SH}}$  values can be calculated only for 13 845 stars. Fig. 4 shows the comparison between  $d_{\text{OB}}$  and  $d_{\text{BJ}}$  (left-hand panel) and between  $d_{\text{OB}}$  and  $d_{\text{SH}}$  (right-hand panel).

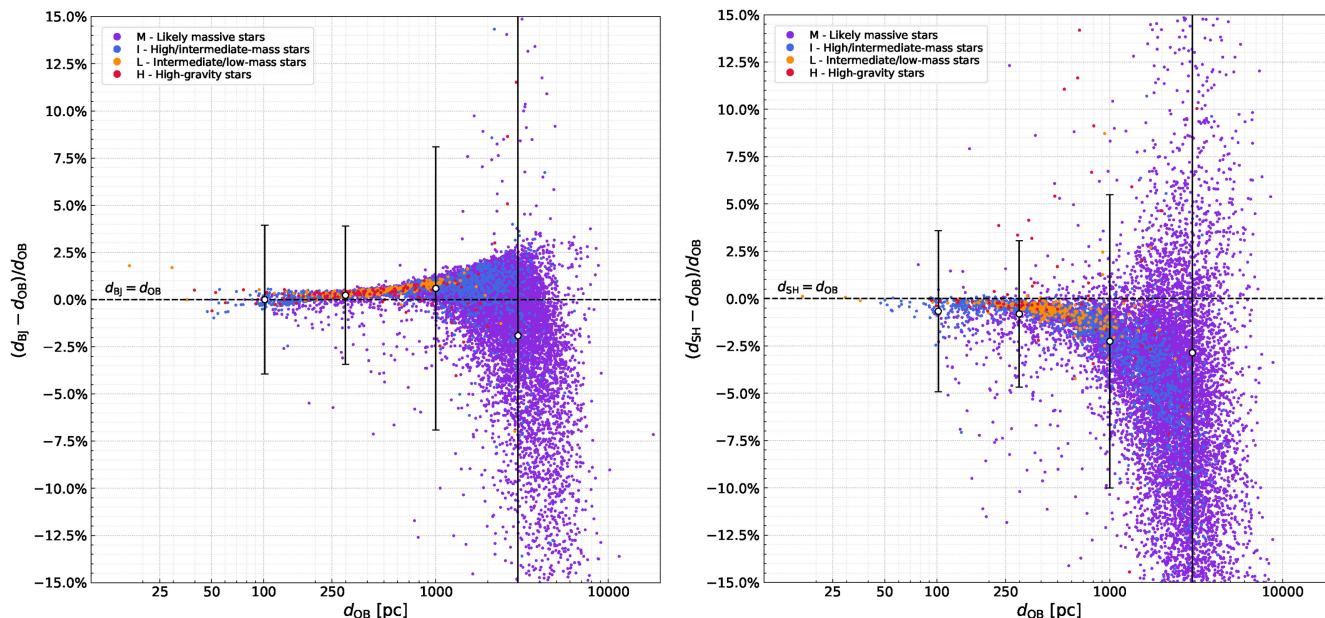
The most important result of the comparisons is that all three distances are very similar. In the range between 0 and 1 kpc the differences between  $d_{\text{OB}}$  and  $d_{\text{BJ}}$  grow from 0 per cent to  $\sim 1$  per cent and between  $d_{\text{OB}}$  and  $d_{\text{SH}}$  they do it from 0 per cent and 2 per cent. The effects are systematic and of opposite sign:  $d_{\text{BJ}}$  values are larger than  $d_{\text{OB}}$  ones and these are in turn larger than  $d_{\text{SH}}$  ones. However, the typical standard deviations of the posterior distributions (in relative terms) are  $\sim 2$  per cent around 100 pc and  $\sim 5$  per cent around 1000 pc (the plotted values in Fig. 4 are  $\sim \sqrt{2}$  larger to consider the contribution from both measurements) i.e. significantly larger. Between 1 and 3 kpc the dispersion of the difference between the compared values increases considerably but so do the standard deviations. In the comparison between  $d_{\text{OB}}$  and  $d_{\text{BJ}}$  there is no significant bias (indeed, the trend with distance seen between 0 and 1 kpc is reversed until the bias neutralizes around 3 kpc). In the comparison between  $d_{\text{OB}}$  and  $d_{\text{SH}}$  the values for the second become clearly smaller (but within the typical standard deviations of the posterior distribution). Beyond 3 kpc the posterior distributions become very broad but the comparison between  $d_{\text{OB}}$  and  $d_{\text{BJ}}$  values is reasonably good. Therefore, our conclusion is that the distances obtained from the OB prior and from the Bailer-Jones prior are very similar and that for the StarHorse distances the differences are somewhat larger but the comparison is still reasonably close. The lesson is that if your prior is a reasonable approximation to reality, your distances will be mostly independent of the details of the prior itself.

### 3.4 Mapping the solar neighbourhood

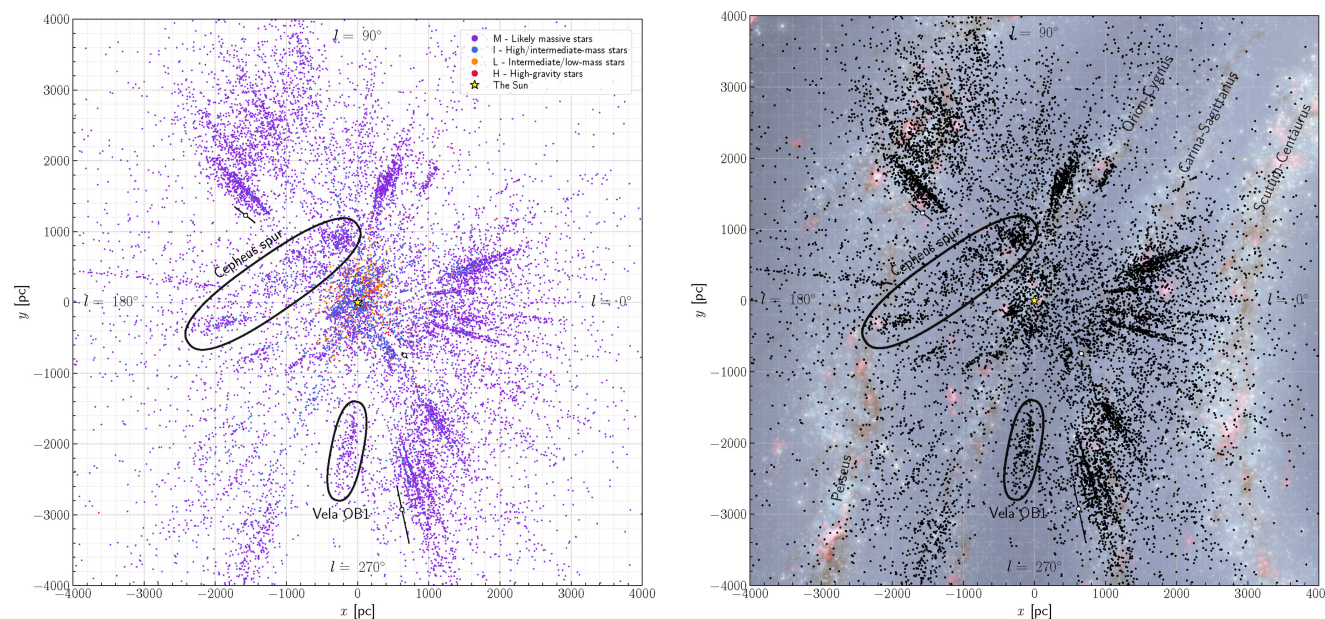
In this subsection, we use the  $d_{\text{OB}}$  distances from the previous one to map the location of the OB stars in the solar neighbourhood. Two previous papers have done a similar study cross-matching the previous version of the ALS catalogue with *Gaia* DR2. Xu et al. (2018) applied a straightforward cross-match with a 1 arcsec radius and obtained a sample of 5772 objects. Ward et al. (2020) used a larger search radius of 5 arcsec that was later cleaned using a  $V - G$  cut, leaving them with a sample of 11 844 stars. Those numbers should be compared with our significantly larger sample of 15 662 objects obtained using the procedure described above. A third paper, Zari et al. (2018), analyses the distribution of young populations in the solar neighbourhood but its scope is different from ours, as it deals mostly with intermediate- and low-mass stars within 500 pc.

We first describe the distribution of our sample in the Galactic plane, using as reference Fig. 5. In addition, we also provide an

<sup>1</sup>In Table 2, we give our distance estimates based on the mean, median, and mode, and  $G_{\text{abs}}$  from the median and the mode.



**Figure 4.** Comparison between distances obtained using the OB and the Bailer-Jones priors (left) and between the OB and the StarHorse priors (right) for the final ALS sample colour-coded as in Fig. 1. The comparison is made between the medians of the posterior distribution. The error bars, on the other hand, correspond to one standard deviation of the posterior distribution for typical stars at 100, 300, 1000, and 3000 pc, respectively. The error bars for 3000 kpc are larger than the plotted range and amount to 18 per cent (left-hand panel) and 21 per cent (right-hand panel).



**Figure 5.** (left) Positions of our final sample projected on to the Galactic plane using  $d_{OB}$  as the distance and applying the same colour coding by category as in previous figures. Error bars are used to show typical uncertainties for stars located at 1, 2, and 3 kpc, respectively, and can be used to assess how much of the spread in the radial direction is caused by the uncertainties in  $d_{OB}$ . (right) Version of the left-hand panel with only objects of type M and using the artistic impression of the Milky Way by Robert Hurt as background.

animation as supplementary material to better visualize the 3D structures seen in the data (see Fig. 9 for a frame of the animation). We establish a coordinate system with  $x$  and  $y$  centred at the Sun’s position and where the Galactic Centre is at (8.178,0) kpc (Abuter et al. 2019) and where  $+y$  is the direction of Galactic rotation. The origin in  $z$  is fixed at a position of 20 pc below the Sun’s position (Maíz Apellániz et al. 2008) so that it corresponds to the Galactic

mid-plane. The left-hand panel shows the distribution of the stars in our final sample and the right-hand panel only that of M-type objects with the artistic impression of the Milky Way by Robert Hurt in the background (see <https://www.eso.org/public/images/eso1339g/>).

The first description of the Galactic spiral structure using OB stars was done by Morgan et al. (1952). Since then, tracers at different wavelengths have been used to study the configuration of



the spiral arms (Vallée 2017, 2020). In our type-M sample we can see delineated three spiral arms. Outside of the solar circle, the Perseus arm is well traced by OB stars in the  $l = 100^\circ$ – $140^\circ$  range but from that point there are few objects (Negueruela & Marco 2003). The Orion–Cygnus or local arm is seen in opposite directions of the sky, extending well into the third quadrant in the range  $l = 240^\circ$ – $250^\circ$ , in agreement with Vázquez et al. (2008), and not merging with Perseus arm as previously thought, in agreement with Xu et al. (2018). Inside the solar circle, the Carina–Sagittarius arm is the best traced of the three (once one accounts for the artificial spread in the radial direction caused by the uncertainties in  $d_{OB}$ ), as expected by the richness of the different star formation episodes present within it (e.g. Sota et al. 2014; Maíz Apellániz et al. 2020b). Beyond the Carina–Sagittarius arm, the Scutum–Centaurus arm is not seen in the distribution of OB stars in the ALS catalogue mostly due to the strong extinction present in the directions close to the Galactic Centre. These three spiral arms are the same ones that are seen in the spatial distribution of the nearby OB associations in fig. 19 of Wright (2020). Note, however, that Fig. 5 here reaches a distance twice that of the one shown in that paper.

There are at least two interarm structures seen in Fig. 5. One is seen at a distance of  $\sim 2$  kpc around  $l = 265^\circ$  and is produced by the Vela OB1 association. Its ease of identification is likely favoured by the small amount of dust present in this interarm sightline but note that, as with other structures at those distances from the Sun, the elongation in the radial direction is an artefact of the individual distance uncertainties for each star (see the plotted typical error bars). Note, however, that these structures in Vela have been confused in the past with the continuation of the Orion–Cygnus arm (Vázquez et al. 2008 and references therein). The second one is a spur that starts from the Cygnus arm at around  $l = 90^\circ$  and moves out towards the Perseus arm, meeting it around  $l = 190^\circ$ . This structure is delineated by six OB associations: Cep OB2+OB3 + OB4, Cam OB1, Aur OB1, and Gem OB1 and, to our knowledge, it has never been identified before. We name it the Cepheus spur and we discuss it further below. In this case the elongation in the radial direction is smaller because most of its stars are closer than 2 kpc.

We now describe the distribution of OB stars in the vertical direction, which one of us analysed previously using *Hipparcos* data (Maíz Apellániz 2001) but without paying attention to the possible changes as a function of position in the Galactic disc. In that respect, the most noticeable structures expected in the Galactic disc are (a) the warp previously detected in positions and proper motions (Reed 1996; Poggio et al. 2018) and (b) possible corrugation patterns, which should be more prominent in young populations than in old ones (Matthews & Uson 2008). The measured line of nodes of the warp (Poggio et al. 2020) is not distant from the anticentre position, with the second quadrant bent upwards and the third quadrant downwards. To check the possible presence of the Galactic warp in our data, we show the vertical distribution as a function of Galactocentric distance in Fig. 6. OB stars in the second quadrant indeed appear to be located preferentially above the plane at large Galactocentric distances and the disc itself appears to become broader as we move outwards in the Galaxy there, in accordance with the predictions of the Galactic warp model of Poggio et al. (2020). The situation in the third quadrant is less clear, as the vertical distribution seems to bifurcate into an upper and a lower components. It is possible that a small part of the bifurcation is caused by extinction but that is not likely to be all of the story, as dust is scarce in the outer disc.

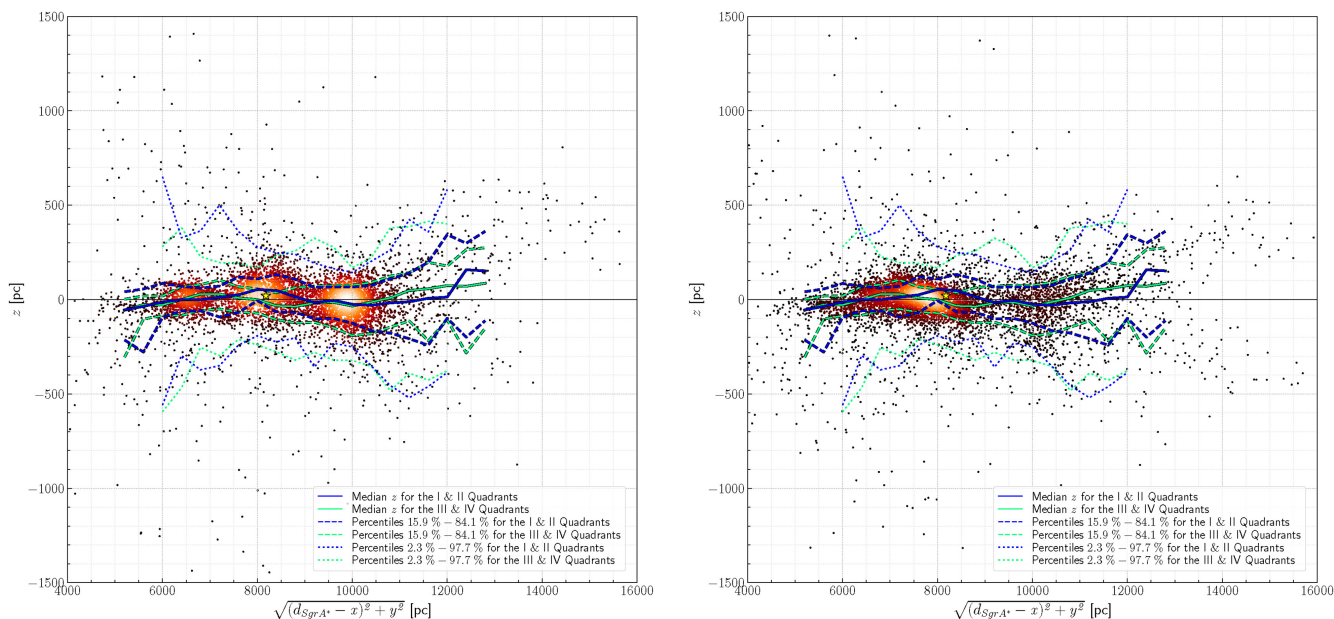
There are other effects seen in Fig. 6 that deserve attention. Inside the solar circle the distribution reaches shorter distances than in the opposite direction, a consequence of the rapid increase in extinction

towards the inner disc. Also, the distribution shows an increased number at higher latitudes in those directions, indicating there are more runaways in the first and fourth quadrants. This is a likely consequence of the overall higher number of OB stars within the solar circle: extinction may not let us see the OB stars close to the Galactic plane but can do little to avoid the detection of the ejected stars far from it. Another effect in Fig. 6 is the different radial distribution between the two panels. In the right one the population is dominated by the Carina–Sagittarius arm, while in the left-hand panel the three main arms discussed in this paper (Carina–Sagittarius, Orion–Cygnus, and Perseus) are seen as distinct concentrations at different average Galactocentric distances. A third effect is related to the recent discovery by Alves et al. (2020) of a structure they dubbed the Radcliffe wave. It is a damped sinusoidal 1D vertical oscillation that traces the Orion–Cygnus arm, with a region with positive  $z$  close to the Sun being above the mid-Galactic plane and a region with negative  $z$  in the opposite direction (that includes Ori OB1) below the mid-Galactic plane. Fleck (2020) has proposed that such a wave may arise as a result of a Kelvin–Helmholtz instability between the Galactic disc and the Galactic halo. The effect of the Radcliffe wave can be seen in the different behaviour of the median height close to the Sun in the two panels: above average in the left-hand panel and below average in the right-hand panel.<sup>2</sup>

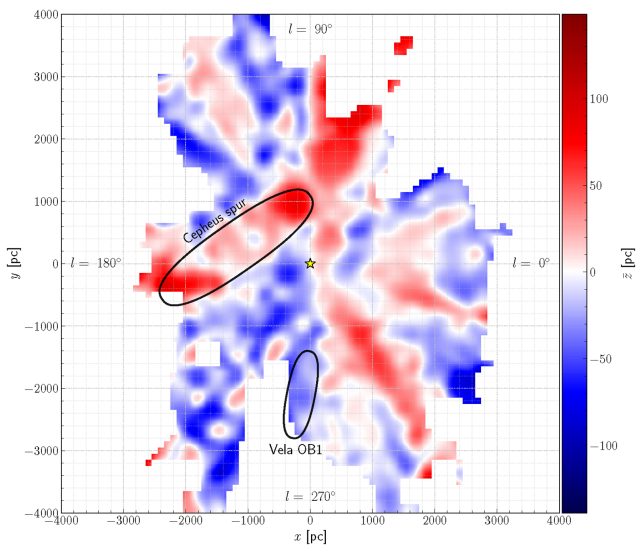
To study in more detail the variations of the distribution of OB stars in the vertical direction, we show the average value of the height with respect to the Galactic mid-plane in Fig. 7. Here, we see that the distribution in the vertical direction within a few kpc has a complex behaviour along the  $x$ – $y$  plane and does not appear to be dominated by the warp, which can be instead clearly seen for OB stars farther from the Sun (Fig. 6). Instead, the pattern we see is more consistent with being dominated by corrugation effects. A different behaviour in the vertical direction between samples of ‘OB’ and RGB stars is also observed by Romero-Gómez et al. (2019). The left-hand panel of their Fig. 5 is the equivalent to our Fig. 7 but note that their ‘OB’ sample is quite different from ours: much larger in size but including objects with ages up to 1 Ga, while ours contains only ages up to 30 Ma. In other words, most of their sample is made out of intermediate-mass late-B dwarfs, which are not included in the traditional definition of OB stars: they are what we classify as I-type objects.

There are other structures seen in Fig. 7 but their reality is unclear at this point without further data. For example, the Cygnus–X region of the Orion–Cygnus arm is also located above the mid-plane but the strong extinction present there questions whether that is just a partial obscuration effect. Two other elevations above the mid-plane in the fourth quadrant are suspiciously elongated in the radial direction (something that does not happen for the Cepheus spur), indicating a possible origin in obscuring clouds on the Carina–Sagittarius arm blocking stars at negative values of  $z$  behind them. On the other hand, the already commented Vela OB1 association is likely to be indeed located below the mid-plane, as in that direction there is little extinction. We note that the structures we see in our sample are quite similar to those seen in fig. 1 of Alfaro, Cabrera-Cano & Delgado (1991), who did an analysis similar to ours but using young open clusters. We plan to analyse these corrugation patterns in more detail with *Gaia* EDR3 data and future versions of the ALS catalogue with more OB stars.

<sup>2</sup>This is aided by the fact that the ALS sample is more complete within 1 kpc of the Sun than at longer distances.



**Figure 6.** Positions of the M-type sample as a function of distance from the Galactic Centre and height with respect to the mid-Galactic Plane (with the Sun located 20 pc above it). The left-hand panel shows the positions for objects in the first two Galactic quadrants and the right-hand panel the equivalent for objects in the last two Galactic quadrants. The running median plus  $1\sigma$  and  $2\sigma$  equivalent percentiles are shown in both plots (blue lines for the distribution in the first two quadrants, green lines for the distribution in the last two). The yellow star marks the Sun’s position.



**Figure 7.** Average height with respect to the Galactic mid-plane of the M-type sample as a function of position on the plane. The values have been calculated using a 2D Gaussian smoothing kernel with  $\sigma = 100$  pc. Pixels with a small number of stars have been whitened out to show only regions where the average is calculated using a significant number of stars.

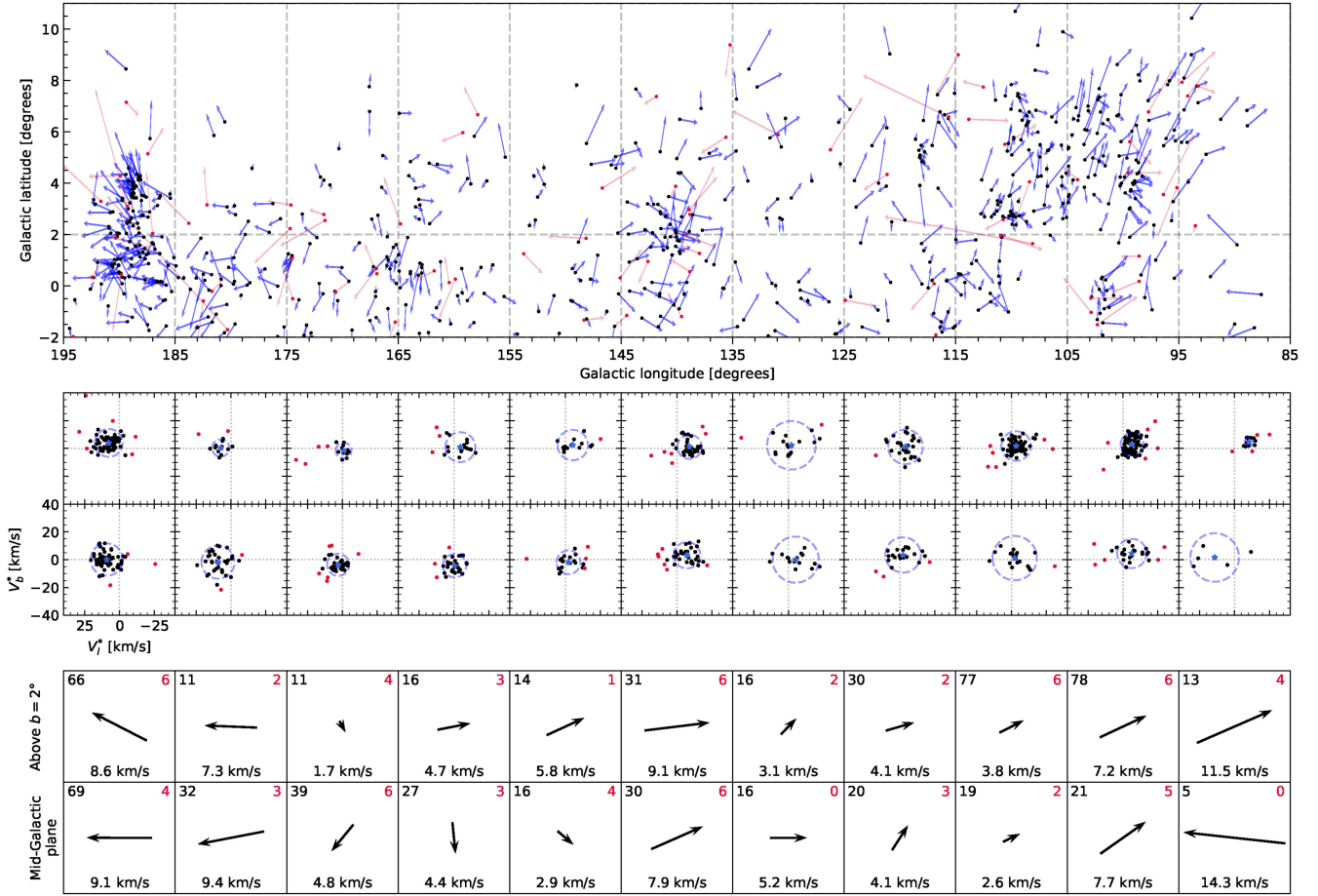
### 3.5 The Cepheus spur

The Cepheus spur is a conspicuous structure in Fig. 7 (see also Fig. 9). It starts around (0,1 kpc),  $\sim 500$  pc beyond where the oscillation of the Radcliffe wave begins, and maintains an elevated position 50–100 pc above the plane as it moves diagonally towards the lower left direction (the real height of the structure is likely to be closer to 100 pc than to 50 pc, as we have not excluded objects that lie below it). Therefore, it looks like the Radcliffe wave is not a 1D

oscillation that creates a 2D structure in a vertical plane but instead a full 3D structure extended towards larger Galactocentric radii, with the Cepheus spur being a region of recent enhanced star formation at the crest of the wave. Considering that Ori OB1 is located at the opposite side of the wave with respect to the Galactic mid-plane (it is the blue region just to the lower left of the Sun in Fig. 7), we tentatively propose that this type of oscillation in the Galactic disc may be responsible for the enhanced star formation at the crests and troughs of the wave. Note that the initial rise of the Cepheus spur between Cygnus and Cepheus was noticed as early as by Hubble (1934), who called attention to the existence of molecular clouds some distance away from the plane and dubbed that part of the structure the Cepheus flare. The associated molecular gas extends well above the heights above the plane where OB stars are found (Kun, Kiss & Balog 2008).

To verify the identity of the Cepheus spur we have done an analysis of the peculiar velocities in the plane of the sky for the OB stars in that region of the Milky Way derived from the *Gaia* DR2 proper motions, which is shown in Fig. 8. We selected the sample from the oval region in Fig. 5 and divided it into an  $11 \times 2$  grid in Galactic longitude and latitude. The top row corresponds approximately to the Cepheus spur itself, while the bottom row is the normal mid-Galactic plane population (we do not show the  $b = -2^\circ$  to  $-11^\circ$  range because there are few stars in there). As a reference, the dividing  $b = 2^\circ$  line corresponds to 35 pc at a distance of 1 kpc (right part of the plot) and to 70 pc at a distance of 2 kpc (left part of the plot).

We start the analysis of the Cepheus spur with the peculiar latitudinal motion which, for practical purposes, is very similar to the peculiar motion in the vertical Galactic direction given the small angular distances from the Galactic plane involved. The average value for the stars above  $b = 2^\circ$  is  $2.3 \text{ km s}^{-1}$ , while that for the stars below that value is  $-0.2 \text{ km s}^{-1}$ . This indicates that the mid-Galactic plane population has the expected behaviour of a near-zero vertical



**Figure 8.** Kinematic analysis of the Cepheus spur. (top) A sky chart of the M and I catalogue stars inside the Cepheus spur oval selection of Fig. 5, with arrows showing the peculiar velocities in the plane of the sky. To calculate the velocities we first corrected for the peculiar velocity of the Sun with respect to the Local Standard of Rest and then for a flat Galactic rotation curve model. The red dots and arrows are considered kinematical outliers (possible runaways) and are excluded from the analysis. The grey dashed lines show the limits of the bins used to separate the mid-Galactic plane ( $|b| < 2^\circ$ ) and the Cepheus spur ( $b > 2^\circ$ ) above it in 11 longitudinal tiles. (middle) A representation of the longitudinal and latitudinal components of the peculiar velocity in the 22 bin selections defined in the top chart. The red dots are again the cases excluded during the robust mean calculation, while the black dots are used to give an average motion for each bin, which is represented with a blue star. The blue dashed circles correspond to 3 standard deviations from the selected sample of each bin. (bottom) A representation of the robust average corrected transverse velocities of each bin. For each cell the number of stars used for the final average is shown in black, and the number of outliers excluded from the analysis in red. In the lower part of each cell the absolute value of the robust average corrected transverse velocity is given.

motion, while the above-the-plane population is differentiated by having an average upwards motion.

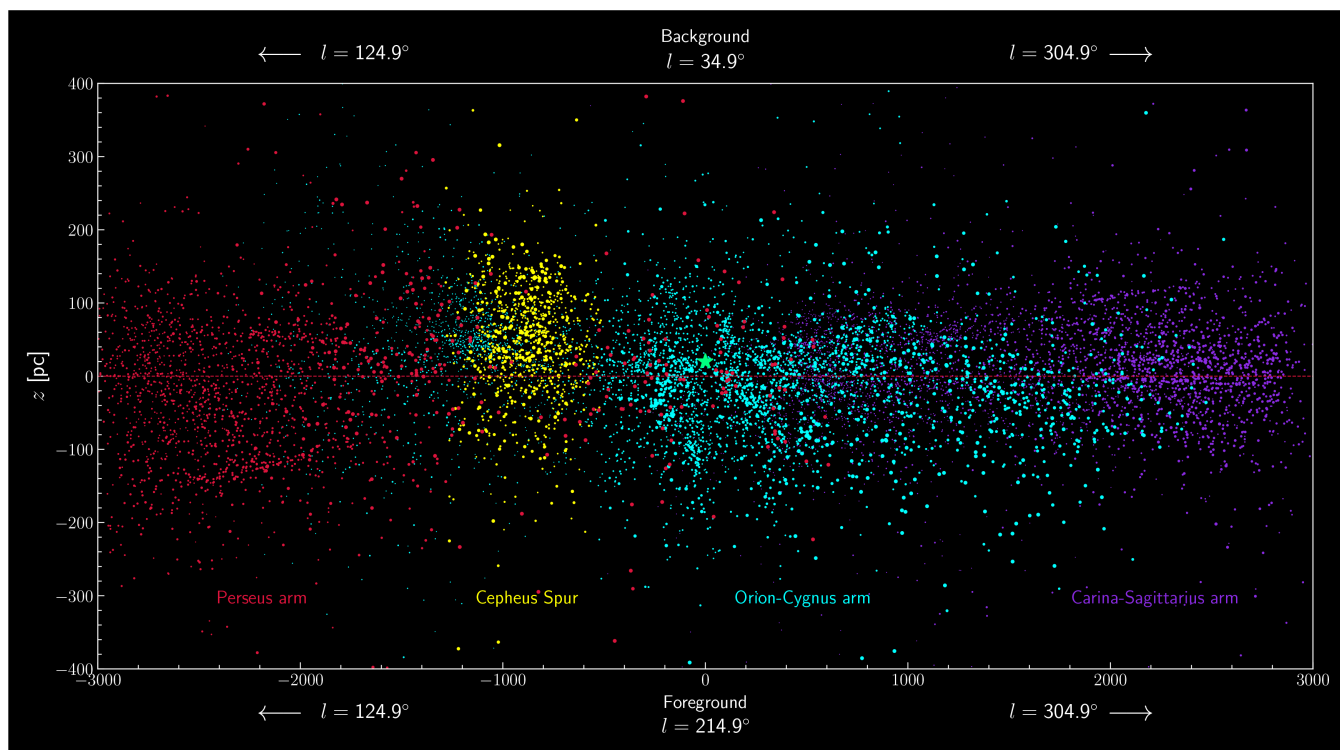
We now turn to the motion in the longitudinal direction. What we see there for  $b > 2^\circ$  is the signature of a radiant point similar to that of a meteor shower i.e. the projection of a 3D velocity vector into a spherical coordinate system. Values are small in the central region of the upper panels of Fig. 8 (with the exception of the  $l = 135^\circ$ – $145^\circ$  bin) and increase in opposite directions as we move towards lower and higher Galactic longitudes. If we assume a radiant point at  $l = 150^\circ$  and a peculiar velocity in the Galactic plane of  $12 \text{ km s}^{-1}$  toward us, we would see a peculiar longitudinal velocity of  $+7.7 \text{ km s}^{-1}$  at  $l = 190^\circ$  and of  $-10.4 \text{ km s}^{-1}$  at  $l = 90^\circ$ , which are nearly identical to the measured values in the two extreme bins of  $+7.7$  and  $-10.5 \text{ km s}^{-1}$ , respectively. For the stars with  $b < 2^\circ$  a similar pattern in longitude is seen, so it is possible that the Cepheus spur is a coherent structure that is on average above the plane but extends to a larger range of Galactic latitudes and is characterized by approaching us with a peculiar velocity close to  $12 \text{ km s}^{-1}$  and from a Galactic longitude of  $\sim 150^\circ$ . A corollary of this result is that peculiar velocities in the

radial direction in the Cepheus spur are expected to be negative. We plan to test that in the future and we also plan to analyse the motion derived from the stellar clusters in the region.

Finally, we analyse the spatial distribution of stars above and below the  $b = 2^\circ$  dividing line. For  $l < 145^\circ$  there are 245 final stars (after excluding possible runaways) above the line and 111 below, a difference of more than a factor of two that is the main reason for the prominence of the Cepheus spur in Figs 7 and 9. In the  $l = 145^\circ$ – $185^\circ$  range there are relatively few stars but the number rises again in the last bin ( $l = 185^\circ$ – $195^\circ$ ) due to the presence of Gem OB1. Therefore, we conclude that there are three lines of evidence that support the existence of the Cepheus spur: an overdensity in the  $xy$  map, an anomalous average height over the Galactic plane, and kinematics consistent with an overall common peculiar motion.

#### 4 FUTURE WORK

In the immediate future we plan to incorporate the information from future *Gaia* releases. EDR3 will improve the quality of the



**Figure 9.** One of the frames of the supplementary animation associated with this article, where a full rotation around the solar mid-Galactic plane projection point is completed with the Galactic disc shown edge-on. The camera position is at infinity (located in the  $l = 214.9^\circ$  direction and pointing towards  $l = 34.9^\circ$ ) so that the y axis represents the height above the mid-Galactic plane without field-of-view projection distortions. The stars from the M and I catalogues closer than 3 kpc are depicted for the Perseus arm (crimson), the Orion–Cygnus arm (cyan), the Carina–Sagittarius arm (violet), and the Cepheus spur region (yellow, note that we do not distinguish between objects above and below  $b = 2^\circ$  as we do in Fig. 8). The position of the Sun is represented by a green star (20 pc above the mid-Galactic plane) and the position of the Galactic Centre is marked in the animation by a red star. The y axis has been exaggerated by a factor of 3.6 to better illustrate the distributions of heights in the Galactic disc, with the Cepheus Spur oval selection showing the anomalous average height of this structure.

parallaxes and of the  $G_{BP} + G + G_{RP}$  photometry, allowing us to add existing ALS stars to the clean sample in this paper. In DR3 spectrophotometry will become available and with it a better discrimination between source types and a measurement of extinction properties. Also, for some sources (primarily the cooler spectral subtypes) radial velocities will be provided.

We also plan to add accurate spectral classifications from spectroscopic surveys. For bright stars we have already obtained  $R \sim 2500$  blue-violet spectroscopy with GOSSS (Maíz Apellániz et al. 2011) for several thousands of stars. Some of them have already been published (most of them in Sota et al. 2011, 2014; Maíz Apellániz et al. 2016) and can be retrieved from the Galactic O Star Catalogue (Maíz Apellániz et al. 2004) web site: <https://gosc.cab.inta-csic.es>. Most of the rest corresponds to B stars and their spectral classifications will be added to future versions of the ALS catalogue. We will also incorporate high-resolution spectroscopic results from LiLiMaRlin (Maíz Apellániz et al. 2019a), including radial velocities derived from multi-epoch data. For faint stars spectral types for many OB stars will become available with WEAVE (Dalton 2016) in the Northern hemisphere and with 4MOST (de Jong et al. 2019) in the Southern hemisphere.

A third type of addition to the ALS catalogue will come from ground-based photometric surveys, which can contribute with the  $u$ -band photometry necessary to identify and characterize OB stars. For this purpose we will use GALANTE (Lorenzo-Gutiérrez et al. 2019, 2020), IGAPS (Monguió et al. 2020), and VPHAS + (Drew et al. 2014). We may also add *Gaia*-identified members from the

Villafranca catalogue of Galactic OB groups (Maíz Apellániz et al. 2020b) that are bright enough to be OB stars.

The most direct effect of those contributions to the ALS catalogue will be a larger sample with many new objects and a more precise characterization of their properties. That in turn will produce an improved knowledge of the spatial distribution of the OB stars in the solar neighbourhood, allowing us to detect finer structures and new runaway stars and to extend the range at which we can discover and study new stellar clusters and associations with OB stars.

## ACKNOWLEDGEMENTS

We thank the referee for a constructive report that helped us to improve the paper. MPG and JMA acknowledge support from the Ministerio de Ciencia, Innovación y Universidades through grant PGC2018-095049-B-C22. RHB acknowledges support from DIDULS Project 18 143 and the ESAC Faculty Visitor Program. This work has made use of data from the European Space Agency (ESA) *Gaia* mission, processed by the *Gaia* Data Processing and Analysis Consortium (DPAC). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. This research has made use of the SIMBAD data base, Aladin Sky Atlas, and the VizieR catalogue access tool, operated and developed by the Centre de Données astronomiques (CDS), Strasbourg Observatory, France, as well as TOPCAT, an interactive graphical viewer and editor for tabular data,

and ASTROPY, a community-developed core Python package for Astronomy.

## DATA AVAILABILITY

The data from this paper are available from the CDS.

## REFERENCES

- Abuter R. et al., 2019, *A&A*, 625, L10
- Alfaro E. J., Cabrera-Cano J., Delgado A. J., 1991, *ApJ*, 378, 106
- Alves J. et al., 2020, *Nature*, 578, 237
- Anders F. et al., 2019, *A&A*, 628, A94
- Andrae R. et al., 2018, *A&A*, 616, A8
- Babusiaux C. et al., 2018, *A&A*, 616, A10
- Bailer-Jones C. A. L. et al., 2013, *A&A*, 559, A74
- Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Mantelet G., Andrae R., 2018, *AJ*, 156, 58
- Barbier M., Bernard A., Bigay J. H., Garnier R., 1973, *A&A*, 27, 421
- Blaauw A., 1993, in Cassinelli J. P., Churchwell E. B., eds, *ASP Conf. Ser.*, Vol. 35, *Massive Stars: Their Lives in the Interstellar Medium*. Astron. Soc. Pac., San Francisco. p. 207
- Bouy H., Alves J., 2015, *A&A*, 584, A26
- Chini R., Elsaesser H., Neckel T., 1980, *A&A*, 91, 186
- Clark J. S., Neuguera I., Crowther P. A., Goodwin S. P., 2005, *A&A*, 434, 949
- Dalton G., 2016, in Skillen I., Balcells M., Trager S., eds, *ASP Conf. Ser.*, Vol. 507, *Multi-Object Spectroscopy in the Next Decade: Big Questions, Large Surveys, and Wide Fields*. Astron. Soc. Pac., San Francisco. p. 97
- de Jong R. S. et al., 2019, *The Messenger*, 175, 3
- Drew J. E. et al., 2014, *MNRAS*, 440, 2036
- Evans C. J. et al., 2005, *A&A*, 437, 467
- Fleck R., 2020, *Nature*, 583, E24
- Hardorp J., Rohlf K., Slettebak A., Stock J., 1959, *Hamburger Sternw. Warner & Swasey Obs.*, C01, 0
- Hardorp J., Theile I., Voigt H. H., 1964, *Hamburger Sternw. Warner & Swasey Obs.*, C03, 0
- Hardorp J., Theile I., Voigt H. H., 1965, *Hamburger Sternw. Warner & Swasey Obs.*, C05, 0
- Holgado G. et al., 2020, *A&A*, 638, A157
- Hubble E., 1934, *Contributions from the Mount Wilson Observatory / Carnegie Institution of Washington*, 485, 1
- Johnson H. L., Morgan W. W., 1953, *ApJ*, 117, 313
- Jordi C. et al., 2010, *A&A*, 523, A48 +
- Kun M., Kiss Z. T., Balog Z., 2008, preprint ([arXiv:0809.4761](https://arxiv.org/abs/0809.4761))
- Lorenzo-Gutiérrez A. et al., 2019, *MNRAS*, 486, 966
- Lorenzo-Gutiérrez A. et al., 2020, *MNRAS*, 494, 3342
- Lutz T. E., Kelker D. H., 1973, *PASP*, 85, 573
- Maíz Apellániz J., 2001, *AJ*, 121, 2737
- Maíz Apellániz J., 2005, in Turon C., O’Flaherty K. S., Perryman M. A. C., eds, *ESA SP-576: The Three-Dimensional Universe with Gaia*. ESA, Noordwijk, p. 179
- Maíz Apellániz J., 2006, *AJ*, 131, 1184
- Maíz Apellániz J., 2013a, in *HSA 7*. p. 583
- Maíz Apellániz J., 2013b, in *HSA 7*. p. 657
- Maíz Apellániz J., 2019, *A&A*, 630, A119
- Maíz Apellániz J., Barbá R. H., 2018, *A&A*, 613, A9
- Maíz Apellániz J., Sota A., 2008, in Benaglia P., Bosch G. L., Cappa C. E., eds, *RMxAC*, Vol. 33, p. 44
- Maíz Apellániz J., Weiler M., 2018, *A&A*, 619, A180
- Maíz Apellániz J., Walborn N. R., Galué H. Á., Wei L. H., 2004, *ApJS*, 151, 103
- Maíz Apellániz J., Alfaro E. J., Sota A., 2008, preprint ([arXiv:0804.2553](https://arxiv.org/abs/0804.2553))
- Maíz Apellániz J., Sota A., Walborn N. R., Alfaro E. J., Barbá R. H., Morrell N. I., Gamen R. C., Arias J. I., 2011, in *HSA 6*. p. 467
- Maíz Apellániz J. et al., 2014, *A&A*, 564, A63
- Maíz Apellániz J. et al., 2016, *ApJS*, 224, 4 (GOSSS III)
- Maíz Apellániz J., Pantaleoni González M., Barbá R. H., Simón-Díaz S., Neuguera I., Lennon D. J., Sota A., Trigueros Páez E., 2018, *A&A*, 616, A149
- Maíz Apellániz J., Trigueros Páez E., Jiménez Martínez I., Barbá R. H., Simón-Díaz S., Pellerin A., Neuguera I., Souza Leão J. R., 2019a, in *HSA 10*. p. 420 (LiLiMaRlin)
- Maíz Apellániz J. et al., 2019b, *A&A*, 626, A20 (MONOS I)
- Maíz Apellániz J., Pantaleoni González M., Barbá R. H., García-Lario P., Nogueras-Lara F., 2020a, *MNRAS*, 496, 4951
- Maíz Apellániz J., Crespo Bellido P., Barbá R. H., Fernández Aranda R., Sota A., 2020b, *A&A*, 643, A138
- Maíz Apellániz J., Barbá R. H., Caballero J. A., Bohlin R. C., Fariña C., 2021, *MNRAS*, 501, 2487
- Martin N., 1972, *A&A*, 17, 253
- Mason B. D., Gies D. R., Hartkopf W. I., Bagnuolo W. G., Brummelaar T. T., McAlister H. A., 1998, *AJ*, 115, 821
- Matthews L. D., Uson J. M., 2008, *ApJ*, 688, 237
- Moffat A. F. J., Jackson P. D., Fitzgerald M. P., 1979, *A&AS*, 38, 197
- Moffat A. F. J., Drissen L., Shara M. M., 1994, *ApJ*, 436, 183
- Monguió M. et al., 2020, *A&A*, 638, A18
- Morgan W. W., Sharpless S., Osterbrock D., 1952, *AJ*, 57, 3
- Nassau J. J., Stephenson C. B., 1963, *Hamburger Sternw. Warner & Swasey Obs.*, C04, 0
- Nassau J. J., Stephenson C. B., McConnell D. J., 1965, *Hamburger Sternw. Warner & Swasey Obs.*, C06, 0
- Neuguera I., Marco A., 2003, *A&A*, 406, 119
- Orsatti A. M., Muzzio J. C., 1980, *AJ*, 85, 265
- Poggio E. et al., 2018, *MNRAS*, 481, L21
- Poggio E., Drimmel R., Andrae R., Bailer-Jones C. A. L., Fouesneau M., Lattanzi M. G., Smart R. L., Spagna A., 2020, *Nat. Astron.*, 4, 590
- Queiroz A. B. A. et al., 2018, *MNRAS*, 476, 2556
- Reed B. C., 1993a, *ApJS*, 87, 367
- Reed B. C., 1993b, *PASP*, 105, 1465
- Reed B. C., 1996, *AJ*, 111, 804
- Reed B. C., 2003, *AJ*, 125, 2531 (Paper I)
- Romero-Gómez M., Mateu C., Aguilar L., Figueras F., Castro-Ginard A., 2019, *A&A*, 627, A150
- Sota A., Maíz Apellániz J., Walborn N. R., Alfaro E. J., Barbá R. H., Morrell N. I., Gamen R. C., Arias J. I., 2011, *ApJS*, 193, 24 (GOSSS I)
- Sota A., Maíz Apellániz J., Morrell N. I., Barbá R. H., Walborn N. R., Gamen R. C., Arias J. I., Alfaro E. J., 2014, *ApJS*, 211, 10 (GOSSS II)
- Stephenson C. B., Sanduleak N., 1971, *Publications of the Warner & Swasey Observatory*, 1, 1
- Stock J., Nassau J. J., Stephenson C. B., 1960, *Hamburger Sternw. Warner & Swasey Obs.*, C02, 0
- Vallée J. P., 2017, *Astron. Rev.*, 13, 113
- Vallée J. P., 2020, *ApJ*, 896, 19
- Vázquez R. A., May J., Carraro G., Bronfman L., Moitinho A., Baume G., 2008, *ApJ*, 672, 930
- Wade G. A. et al., 2015, *MNRAS*, 447, 2551
- Walborn N. R., Sota A., Maíz Apellániz J., Alfaro E. J., Morrell N. I., Barbá R. H., Arias J. I., Gamen R. C., 2010, *ApJ*, 711, L143
- Ward J. L., Kruijssen J. M. D., Rix H.-W., 2020, *MNRAS*, 495, 663
- Wramdemark S., 1976, *A&AS*, 26, 31
- Wramdemark S., 1980, *A&AS*, 41, 33
- Wright N. J., 2020, *New Astron. Rev.*, 90, 101549
- Xu Y. et al., 2018, *A&A*, 616, L15
- Zari E., Hashemi H., Brown A. G. A., Jardine K., de Zeeuw P. T., 2018, *A&A*, 620, A172

## SUPPORTING INFORMATION

Supplementary data are available at [MNRAS](https://www.mnras.org/) online.

**ALS\_II\_Catalog.zip**

## ALS II Animation.mp4

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

## APPENDIX A: DIGGING INTO THE PREVIOUS ALS CATALOGUE AND CROSS-MATCHING WITH GAIA DR2

The original ALS catalogue was the result of years of painstaking data gathering and it provided the largest carefully built catalogue of its kind. It is no surprise that there are errors and other issues in the one-by-one elaboration of its 18 693 entries. Here, we acknowledge the diversity of the problems encountered by addressing different types.

*Crowding:* NGC 3603 is the richest very young stellar cluster accessible in the optical (Maíz Apellániz et al. 2020b) and, as a result, a severe example of crowding. There are 13 ALS sources in NGC 3603 and, except for ALS 2275, all of those cases lack photometry in the ALS. Since the coordinates were unreliable to perform certain matches, the coordinates shown in the original ALS catalogue were disposed in a grid-like fashion around the centre of NGC 3603. The cluster also includes two pairs of stars (ALS 19311/ALS 19314 and ALS 19310/ALS 19312) that, in the original catalogue, share the exact same coordinates and are not duplicates of each other. With the photographic plates referenced in the ALS, we were able to match four of these 13 entries with *Gaia* DR2 (which could not detect a number of the sources recognizable in this field due to crowding, see Maíz Apellániz et al. 2020b). ALS 2275 (= HD 97 950) in the LS-South catalogue refers to the cluster core and its multiple components (Moffat, Drissen & Shara 1994). Some variation of these problems are common in other dense clusters with ALS sources.

*Duplicates:* We have discovered 95 instances of duplicates that were not recognized as so in the original ALS catalogue, which was expected due to the large overlapping of some of the original references it was based on. In addition there are pairs of stars which are said to be duplicates but are not, such as ALS 11 110 and ALS 11 108, which are in fact the two components of the binary system HDE 228 827.

*Potentially misleading data:* In the ALS, some values for the *V* magnitude were intentionally taken from other photometric bands, to better compare the data with the LS photographic plates. Some confusion may arise for ALS 19 610 since the *V*-band photometry was instead taken from the column corresponding to the *B* band in Chini, Elsaesser & Neckel (1980). The same happened between the *V* and *B* bands for 10 stars in Westerlund 1, taken from Clark et al. (2005).

*Transcription errors:* Other cases arise from badly transcribed data, either from the reference sources into the ALS or from previous sources into the C-S and LS catalogues. For example, the star ALS 16 894 was mistakenly added to the ALS as CD −28 2561 due to a missing letter P in the originally targeted star, CPD −28 2561, which was correctly included as ALS 870.<sup>3</sup> In ALS 19 483, ALS 16 986, ALS 16 991, and many others the *V* photometry presented in the references was simply ignored. In ALS 9 528 the

coordinates from the LS catalogue erroneously substitute the 19 h in RA with an 18. Similarly, for ALS 17 479 the hour in the right ascension was swapped from 17 to 12 and in ALS 19 668 the 38 arcmin in declination were transcribed as 28 by mistake.

*Bad-quality coordinates:* Many ALS sources have coordinate uncertainties of a few arcseconds but some can be significantly large. For example, for ALS 12 636 (one of the few cases without a Simbad entry) the coordinates are suspiciously rounded to the arcminute. As it turns out, this is possibly a duplicate of ALS 12 639 or at least the result of a chain of badly transcribed coordinates from the LS and BD catalogues, which also show large errors in their coordinates.

*Simbad:* In some cases Simbad was the wrongdoer and therefore some re-examination of its reliability for the *Gaia* DR2 cross-match was needed. For example, ALS 15 862 was wrongly matched with *Gaia* DR2 5 350 363 910 256 783 488, which in turn is a better astrophotometric match for another ALS source, ALS 1820, while the best match for ALS 15 862 is *Gaia* DR2 5 350 363 875 897 024 256. Simbad also matched ALS 19 613 with Tyc 6265-1255-1, which in turn is matched with *Gaia* DR2 4 097 815 382 164 899 840 by the external cross-matches in the *Gaia* archive, an identifier which, according to Simbad, is itself matched to ALS 19 618. Some of these chains of inconsistent cross-matches between ALS references, Simbad, and *Gaia* DR2 cross-matches with external catalogues are found across the ALS. We have solved these issues adapting to what seemed to be the most plausible scenario in a case-by-case procedure, but in general we have assumed that *Gaia* DR2 cross-matches have better quality and consistency than those performed by Simbad.

*Detective cases:* Finally, in some cases we had to spend a significant amount of time to decipher what was going on. They usually have a combination of issues. Here are some examples.

In just 20 arcsec around ALS 18 476 there are 30 back-propagated *Gaia* DR2 sources, many of which are also good photometric matches for ALS 18 476. The ALS refers here to some photographic plates published by Wramdemark (1980), where the true match can be visually determined to be *Gaia* DR2 5 356 258 185 934 696 576. However, that paper gives a set of coordinates that are inconsistent with what is shown in the photographic plates, with a separation of 3.7 arcmin between them. On top of that, the ALS coordinates differ by as much as 4 arcmin from both the true source and the coordinates of its reference at the same epoch.

ALS 19 457 was entered into the ALS catalogue as star 42 of Orsatti & Muzzio (1980). However, this star is 1.23° away from the coordinates shown in the photographic plate, from which we derived the true match with *Gaia* DR2 5 877 156 797 428 541 952. The mistake comes from the reference using the declination of the star 43 for the star 42 and the declination of the star 42 for the star 41, while maintaining the correct right ascensions, thus wrongly tabulating the positions of their own photographic plates by displacing this column one entry. The ALS catalogue inherited this problem but also ignored the values for the photometry that are correctly shown in the reference, thus making the cross-match analysis even more subtle.

ALS 20 122 is designated as NGC 4755 302 in the ALS catalogue but there is no 302 star in the reference provided (Evans et al. 2005). The true source is in fact NGC 4755 093 in that reference since the classification and photometry coincide with what was really shown in the ALS, but still the coordinates are 7.4 arcmin off track here. So, not only the alternate identifier was wrong but also the coordinates were not correctly transcribed from the reference. But, just by changing the minute in the right ascension from 53 to 54, both ALS 20 122 and NGC 4755 093 end up in the same place.

<sup>3</sup>As it turned out, this object is a peculiar star, one of the few O stars with a significant magnetic field as signalled by being of Of?p type (Walborn et al. 2010; Wade et al. 2015).

Several mistakes in Wramdemark (1976) propagated into the previous version of the ALS catalogue. For example, ALS 18 375, which is the star number 3 in the reference, can be recognized as *Gaia* DR2 2 006 098 517 247 905 152 in the photographic plates, while in their own table it is located 53 arcsec away from that position. This is due to a misalignment between the photographic plate and the superimposed grid of coordinates, which apparently was displaced almost 1 arcmin to the south-west direction. This is easily recognizable for ALS 18 375, but when we consider ALS 12 201 it results in a problematic situation. ALS 12 201 is star number 5 in the reference, but its position in the photographic plates is better matched with star number 3 because of the aforementioned misalignment (when in reality we are dealing with a different star). To make things even worse, star 5 and 3 are very similar in terms of brightness, which could lead to even more confusion about their identity. Because of these issues, another entry, ALS 18 832, which was linked to the star number 162 in Barbier et al. (1973) by the ALS, was not flagged as a duplicate, while in reality it can be seen in the photographic plates that it is the same star as number 3 of Wramdemark (1976).

Ambiguous markings in photographic plates can also lead to wrong identifications, like in the case of ALS 18 099, also known as star number 77 in Martin (1972). The ALS coordinates would make a good match for *Gaia* DR2 2 012 999 082 277 360 512, but the real position of star 77 in the photographic plates of the reference is confusing due to the printed arrow pointing to this and to another star in a very crowded image. We believe that the real match here is *Gaia* DR2 2 012 993 554 649 161 344 which is the same star as ALS 13 252, and thus ALS 18 099 can be now recognized as its duplicate. Another problematic star is ALS 18 665, also known as S-237 4 in Moffat, Jackson & Fitzgerald (1979), because the photographic plate associated with it displays a configuration of stars that we could not confidently recognize inside the bright nebula.

## APPENDIX B: TESTING THE TWO GAIA DR2 $G_{BP}$ PASSBANDS

One of the features of the *Gaia* DR2 calibration by Maíz Apellániz & Weiler (2018) is the use of two different passbands and two different zero-points for  $G_{BP}$  depending on the value of  $G$  (not on the value of  $G_{BP}$ ) for the target. The need for this is explained by fig. 4 in that paper, where it is shown that a jump in the sample standard deviations for the  $G_{BP}$  flux happens at  $G = 10.87$  mag and that the size of the jump depends on the colour of the source: it is large for blue objects and small for red ones. The clean ALS sample obtained in this paper allows us to test this effect by removing the (vast) majority of intrinsically red stars in the full *Gaia* DR2 sample and leaving only blue stars with different degrees of extinction. Such a selection produces an easy-to-understand sample of intrinsically blue SEDs that follows (mostly) an extinction sequence.

According to Maíz Apellániz & Weiler (2018), the differences between the photometry for bright and faint stars can be described by two effects. First, for stars with Vega-like SEDs (*Gaia* magnitudes are expressed using Vega as the reference SED) the difference in  $G_{BP}$  corresponds to 26 mmag, in the sense that stars fainter than  $G = 10.87$  mag have measured  $G_{BP}$  magnitudes that are brighter

(for their SEDs) when compared with brighter stars. Putting it in another way, a faint Vega-like star has a *Gaia* DR2  $G_{BP}-G'$  colour that is 26 mmag bluer than that of its bright equivalent. Second, for faint stars  $G_{BP}$  is significantly more sensitive to the left of the Balmer jump than for bright stars. The effect is apparent in fig. 11 of Maíz Apellániz & Weiler (2018) by comparing the left-hand and right-hand panel. In principle, this could be quantified as a correction using a Johnson  $U - B$ -like colour (which measures the strength of the Balmer jump) but such a colour is very different to the ones available from current *Gaia* photometry. Both  $G_{BP}-G'$  and  $G'-G_{RP}$  are monotonously decreasing functions of  $T_{\text{eff}}$  but  $U - B$  has a more complex behaviour, decreasing for low and high values of  $T_{\text{eff}}$  but with the opposite behaviour at intermediate values that, furthermore, is a function of gravity. That was the primary reason why Maíz Apellániz & Weiler (2018) had to define two different bands for  $G_{BP}$ .

In order to further test the differences between the two magnitude ranges, we plot in the right-hand panel of Fig. 2 a zoom into the lower right corner of the left-hand panel with the addition of extinction tracks for 40 kK MS stars and  $T_{\text{eff}}$  sequences for zero-extinction MS stars. The first noticeable effect is that the two magnitude ranges are separated in the colour-colour diagram, with the faint stars to the left, as expected. Also as expected, the difference is larger for stars with bluer colours, as there is where the contribution to  $G_{BP}$  from photons to the left of the Balmer jump constitute a larger share of the total. As the SEDs become redder, the two sets become closer together and the large separation in  $G_{BP}-G'$  between the zero extinction sequence for O stars becomes smaller and smaller.

Another interesting effect in the right-hand panel of Fig. 2 is the separation between the extinction and temperature sequences, which is significantly larger for faint stars than for bright stars. This is a consequence of the larger sensitivity of  $G_{BP}$  to the left of the Balmer jump for faint stars. In a sense, for faint stars the  $G_{BP}-G' + G'-G_{RP}$  DR2 diagram crudely resembles the classical Johnson  $U - B + B - V$  diagram (Johnson & Morgan 1953), with separated sequences for extinction and temperature, while for bright stars the similarity is harder to notice. We say ‘crudely’ because the separation between sequences is much smaller for the *Gaia* diagram than for the Johnson one, as for  $G_{BP}$  (the  $U$  equivalent) most of the flux originates to the right, not to the left, of the Balmer jump. If we are able to see these effects is due to the extraordinary photometric quality of the *Gaia* data.

A final aspect of the right-hand panel of Fig. 2 is the location of the high-gravity stars. They belong mostly to the faint sample and they are the heavily dominant contribution to the bluest stars there. This is a consequence of the differences in luminosity between massive OB stars and subdwarfs/white dwarfs. The former are luminous objects located mostly in the Galactic plane. Therefore, if they have blue colours they must be relatively nearby and hence bright. If they are faint, they must have considerable extinction. On the other hand, high-gravity stars have a more uniform distribution in Galactic latitude and can be at the same time relatively nearby and faint.

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.