

# Observational bias and young massive cluster characterization – II. Can *Gaia* accurately observe young clusters and associations?

Anne S. M. Buckner<sup>1</sup>,<sup>1</sup>★ Tim Naylor,<sup>1</sup> Clare L. Dobbs,<sup>1</sup> Steven Rieder<sup>2</sup> and Thomas J. R. Bending<sup>1</sup>

<sup>1</sup>*School of Physics and Astronomy, University of Exeter, Stocker Road, Exeter EX4 4QL, UK*

<sup>2</sup>*Geneva Observatory, University of Geneva, Chemin Pegasi 51, CH-1290 Sauverny, Switzerland*

Accepted 2023 October 26. Received 2023 October 26; in original form 2023 September 11

## ABSTRACT

Observations of clusters suffer from issues such as completeness, projection effects, resolving individual stars, and extinction. As such, how accurate measurements and conclusions are likely to be? Here, we take cluster simulations (Westerlund2- and Orion-type), synthetically observe them to obtain luminosities, accounting for extinction, and the inherent limits of *Gaia*, then place them within the real *Gaia* DR3 catalogue. We then attempt to rediscover the clusters at distances of between 500 and 4300 pc. We show the spatial and kinematic criteria that are best able to pick out the simulated clusters, maximizing completeness, and minimizing contamination. We then compare the properties of the ‘observed’ clusters with the original simulations. We looked at the degree of clustering, the identification of clusters and subclusters within the data sets, and whether the clusters are expanding or contracting. Even with a high level of incompleteness (e.g. < 2 per cent stellar members identified), similar qualitative conclusions tend to be reached compared to the original data set, but most quantitative conclusions are likely to be inaccurate. Accurate determination of the number, stellar membership, and kinematic properties of subclusters are the most problematic to correctly determine, particularly at larger distances due to the disappearance of cluster substructure as the data become more incomplete, but also at smaller distances where the misidentification of asterisms as true structure can be problematic. Unsurprisingly, we tend to obtain better quantitative agreement of properties for our more massive Westerlund2-type cluster. We also make optical style images of the clusters over our range of distances.

**Key words:** methods: data analysis – methods: numerical – methods: observational – methods: statistical – stars: statistics – open clusters and associations: general.

## 1 INTRODUCTION

Star clusters and associations are the principal sites of star formation in galaxies, so studying them is important to understand the conditions stars are formed in, and the surrounding conditions of stars and planets as they evolve. Young massive clusters (YMCs) are of particular interest as they host massive stars, which ionize and shape the surrounding gas (Dale & Bonnell 2011; Motte, Bontemps & Louvet 2018; Panwar et al. 2020).

Observational studies of clusters are subject to non-trivial biases and uncertainties, including stellar membership identification, data incompleteness, and 2D projection effects due to the clusters’ 3D orientation in the line of sight (LoS; Ascenso, Alves & Lago 2009; Pfalzner, Kaczmarek & Olczak 2012; Buckner et al. 2022a,b). Detailed studies of high-mass star-forming regions within the Milky Way are therefore typically limited to within a few kiloparsecs to reduce the severity of these issues, which means we tend to focus on a few well-studied objects (such as the Orion, Carina, and Rosette nebulae for example). Nevertheless, it remains difficult to estimate how much impact these biases have on the properties we derive for them.

The aim of this paper series is to assess the potential accuracy of the properties and physics of YMCs derived from observational-based studies, given the above limitations. In Paper I (Buckner et al. 2022a), we considered a cluster formed from two colliding clouds, to investigate issues arising from 2D projection effects. We found that generally correct qualitative conclusions are obtained when viewing a cluster from various orientations, though the accuracy of specific values are unreliable. The properties most likely to be interpreted incorrectly were whether a cluster is expanding or contracting, and the stellar membership of substructure. Other factors affecting observed measurements involve considering gas and dust along the LoS, and the limitations of the telescope. In this paper, we produce synthetic observations of clusters to examine these remaining factors.

Arguably, the current optimal instrument to study clusters is *Gaia* (Gaia Collaboration 2016, 2018, 2021a, 2022), due to the all-sky coverage and high precision measurements. The availability of *Gaia*’s all-sky position, parallax, and proper motions has allowed many photometrically identified clusters to be definitively confirmed as real or asterisms, their membership lists culled and/or expanded, and the refinement of region properties (Cantat-Gaudin et al. 2018; Monteiro & Dias 2019; Dias et al. 2021, 2022; Mahmudunnobe et al. 2021; Rain, Ahumada & Carraro 2021; Cantat-Gaudin 2022; Göppel & Preibisch 2022). Large-scale searches making use of *Gaia*’s

\* E-mail: [a.buckner@exeter.ac.uk](mailto:a.buckner@exeter.ac.uk)

parallax and kinematic data are also proving fruitful, particularly for identifying clusters that have been previously missed due to a lower density contrast with the field (Cantat-Gaudin et al. 2019, 2022; Castro-Ginard et al. 2019, 2020; Sim et al. 2019; Ferreira et al. 2020; Kounkel; Covey & Stassun 2020; Ferreira et al. 2021; Quintana & Wright 2021; Hao et al. 2022; He et al. 2022; Li et al. 2022). However, infrared photometric searches remain useful as *Gaia* is an optical survey so it is vulnerable to extinction (due to both local nebulosity and the foreground interstellar medium). For YMCs whose members are still embedded deep within the natal cloud, this limitation is a particular problem as highly incomplete *Gaia* membership lists are to be expected, with further complexity introduced to the data set by fainter stars that suffer from large parameter uncertainties (Gaia Collaboration 2022). Therefore, despite *Gaia*'s unprecedented accuracy, with incomplete membership lists potentially containing significant uncertainties and interloping field stars, how reliable are the properties we derive for clusters using it?

Here, we attempt to answer this question by placing a simulated cluster within the real *Gaia* EDR3 catalogue at various distances. We attempt to (i) distinguish the cluster's stellar members and non-member stars and (ii) recover the cluster's spatial distribution, structure, and kinematic properties. We also make synthetic images of our clusters as if they were observed, again including all the field stars within our field of view. Kounkel et al. (2018) also use synthetic clusters to test how well they can characterize structure in their observations of the Orion star-forming complex. In contrast to our approach of using simulated clusters, they begin with assumptions about the properties of the populations of field and cluster stars, and draw samples from them.

This paper is structured as follows. Section 2 details our cluster simulations, and Section 3 explains our methodology to convert them into realistic *Gaia* observations. In Section 4, we show optical images of the cluster observations and we describe how stellar members are identified in Section 5. Section 6 details our analysis methods. Our results are presented in Section 7 and discussed in Section 8.

## 2 CLUSTER DATA FROM SIMULATIONS

We take data from two simulations from Dobbs et al. (2022), labelled M1R1FB and M5R1FB in that paper. The simulations follow the formation of clusters and associations in a region of spiral arm. They both simulate the same region, but the initial mass of gas is  $10^5 M_{\odot}$  in M1R1FB, and  $5 \times 10^5 M_{\odot}$  in M5R1FB. Both include feedback in the form of ionization and supernovae, but for the time frames used in this paper, few if any supernovae have occurred.

The evolution of the two regions and the types of clusters formed are quite different in the two simulations. In the lower gas mass simulation, M1R1FB, a loose association consisting of multiple groups of stars forms. This association is not dissimilar in terms of mass, physical size, and substructure to the Orion star-forming region. By contrast, M1R5FB produces stellar clusters that are gravitationally bound and more similar to YMCs.

For the analysis presented here, we take the times from M1R1FB and M5R1FB when the older stars have ages around 3 Myr, and a further time frame from M1R1FB when the older stars have ages around 5.5 Myr. In reality, star formation will be ongoing but we use a single age parameter to simplify our comparisons. The M5R1FB simulation was not run for as long, so we cannot compare both at the later time.

The simulations contain both sink as gas particles. The sinks typically represent groups of stars rather than individual stars and

**Table 1.** List of simulations with the age of the stars, total stellar mass, and number of stars.

Name	Age (Myr)	Mass ( $10^3 M_{\odot}$ )	No. of stars
Orion-type-3	3	0.89	1635
Orion-type-5.5	5.5	2.30	5105
Wd2-type	3	30.14	50 696

we use the grouped star formation tool GRETA<sup>1</sup> (GRoupEd sTAR formation; Liow et al. 2022) to convert them to stars, sampling from a Kroupa initial mass function (Kroupa 2001) with a mass range of 0.1–100  $M_{\odot}$ , then placed randomly within a 10 pc radius of each sink particle, resulting in an approximately uniform distribution (Rieder et al. 2022). A value of 10 pc was chosen as it is representative of the distance between sink particles in the simulation. The gas particles are used to calculate the line-of-sight extinction due to the natal nebulosity for each star (Section 3.1).

For a clearer comparison with observations, from now on we refer to the M1R1FB simulation as ‘Orion-type’ and the M5R1FB simulation as ‘Wd2-type’. M5R1FB is comparable to the more massive young clusters in the Milky Way, e.g. Westerlund 1, Westerlund 2, NGC 3603. In terms of morphology, M5R1FB contains two clear clusters and likewise Westerlund 2 has two groups of stars. Hence, we relate the region in M5R1FB to Westerlund 2 and name this Wd2-type. Thus, we can determine how reliably regions, or clusters, like these are observed, and how well we could observe them at larger distances from the Sun.

Table 1 summarizes the stellar populations for the Orion-type simulation at the two different times (where the appendices ‘-3’ and ‘-5.5’ in the names represent the ages), and for the Wd2-type simulation.

The stars in each simulation have full 6D phase space (3D positions and velocities) data for each star, and membership is absolute – that is all stars are genuine members of the original data sets and there is no non-member contamination.

## 3 PLACING THE SIMULATED DATA SETS IN THE GAIA CATALOGUE

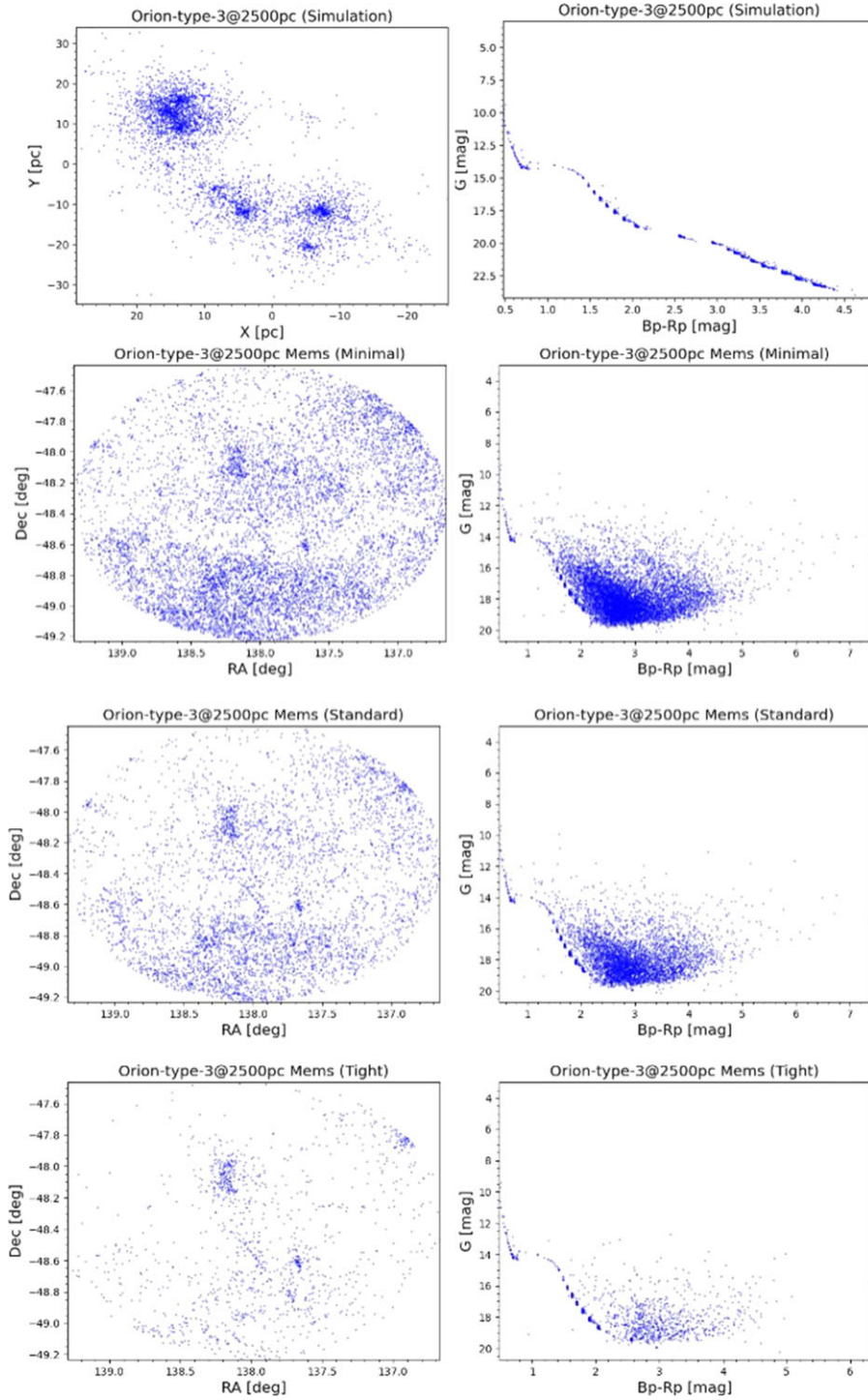
The steps we use to place our simulated clusters in the real *Gaia* DR3 survey<sup>2</sup> (Gaia Collaboration 2022) are as follows:

- (i) We generate observational properties of the stars from the simulated data using a *Gaia* simulator.
- (ii) Assign uncertainties to the simulated stars through random sampling of stars with comparable magnitude in the real *Gaia* catalogue.
- (iii) Use Gaussian sampling to perturb the data values, about these uncertainties.
- (iv) The simulated stars are then combined with the *Gaia* catalogue.
- (v) Stars that would not be detected or resolved by *Gaia* are removed.

These steps are described fully in the following sections. We repeat these procedures assuming the simulated stars lie at distances between 0.5 and 4.0 kpc in 0.5 kpc increments and at 4.3 kpc, the original distance the stars would be in the galaxy simulation (i.e. 9 total observations per data set), centred at coordinates ( $l = 270^{\circ}$ ,

<sup>1</sup><https://github.com/kyliow/greta>

<sup>2</sup><https://gea.esac.esa.int/archive/>



**Figure 1.** These panels show the various stages in creating a synthetic observation and selecting stellar members, as demonstrated for Orion-type-3 placed at a distance of 2500 pc. Position and colour–magnitude diagrams are plotted for (top row:) the SPH simulation, (second row:) members selected using the minimal criteria, (third row:) members-only selected using the standard criteria, (bottom row:) members-only selected using the tight criteria. The clustering of members in the colour–magnitude diagrams is because their properties are taken from the model stars with the closest masses and ages to those given by the simulation (for details see Khorrami et al. 2019). (See Section 5 for selection details).

$b = 0^\circ$ ). For similar real objects, we note that the distance to Orion is  $\sim 400$  pc (Jeffries 2007; Menten et al. 2007), while the distance to Westerlund 2 is uncertain but estimated to be around 3–4 kpc (Carraro et al. 2013; Hur et al. 2015).

We show the positions and colour–magnitude diagrams for the simulated data set from Orion-type-3 before and after these steps in the top two panels of Fig. 1, where the simulated data set is assumed to lie 2.5 kpc away.

### 3.1 Gaia simulator

To generate *Gaia* observations of the simulated stars, we use the *Gaia* simulator<sup>3</sup> tool of synthetic observation generator MYOSOTIS<sup>4</sup> (Make Your Own Synthetic ObservaTIonS; Khorrami et al. 2019). The simulator works by taking the relative positions, ages, and masses of the stars and, using stellar evolutionary and atmosphere models, derives their magnitudes ( $G$ ,  $G_{BP}$ ,  $G_{RP}$ ), parallaxes and proper motions. Distance, line-of-sight (LoS), and variable extinction – due to the relative depth of stars within the simulated gas cloud along the LoS – are taken into consideration when performing the calculations.

We calculate the extinction of the stars by considering (i) the column densities of the simulations’ natal gas cloud and (ii) the presence of dust and gas in the Milky Way’s disc, along the LoS to our cloud. For the latter, we use the canonical value of  $A_V = 0.7 \text{ mag kpc}^{-1}$  (Froeblich et al. 2010). To derive stellar fluxes, we used the Dmodel extinction model with  $R_V = 3.1$ , a solar metallicity ( $Z = 0.015$ ), and turned on the ‘OBtreatment’ option for high-mass stars ( $T_{\text{eff}} > 15 \text{ kK}$ ) to ensure proper spectral energy distributions (SEDs) were selected for these stars.

For each synthetic observation, we removed members whose apparent magnitude was calculated to be outside the sensitivity limit of *Gaia* ( $3 \text{ mag} \leq G \leq 20.7 \text{ mag}$ ).

### 3.2 Introducing uncertainties

Values of proper motions, parallaxes, and radial velocities assigned to stars by the *Gaia* simulator are exact, being scaled from simulation parameters. In reality, however, the values for stars in the *Gaia* catalogues represent the most probable value and include an uncertainty. The values for our clusters are therefore too precise and require perturbing to simulate the effects of observational errors

We began by creating a sample that consisted of sources drawn from the DR3 survey in the LoS of the clusters ( $l = 270^\circ$ ,  $b = 0^\circ$ ) and radius that encompassed all the synthetic observations ( $8.9^\circ$ ). No quality constraints were imposed at this stage as all sources detected by *Gaia* were required, including those with parameters whose solution was a poor fit to the astrometric model, not available, and/or with large uncertainties. The stars’ synthetic observation values from the *Gaia* Simulator were then adjusted as follows:

- (i) The DR3 sample is binned by  $G$ -band magnitude, in increments of 0.5 mag.
- (ii) Each simulated star  $i$  is randomly assigned a renormalized unit weight error (RUWE), and the uncertainties in parallax and proper motion of a DR3 star in the same magnitude bin.
- (iii) A value is randomly drawn from a Gaussian distribution whose mean is equal to the simulated parallax value of star  $i$  and whose standard deviation of the associated uncertainty assigned in step (ii). This is the adjusted parallax value for star  $i$ .
- (iv) Steps (iii) and (iv) are repeated to obtain star  $i$ ’s adjusted proper motion values.

Fig. 2 shows an example of the uncertainties obtained as a function of magnitude for one of our data sets, and also the offset spread for stellar parallax and proper motion. Due to the limited availability of radial velocity measurements in *Gaia* to draw from, we do not derive uncertainties for these values.

### 3.3 Adding field stars

Many stars along the LoS will be field stars rather than members of a given cluster, so to see how our regions would appear as part of an observed data set, we added field stars. This was achieved by merging the synthetic data set obtained at the end of (Section 3.2) with the DR3 data. No quality constraints were applied as all field stars detected by *Gaia* should be included at this stage.

Parallax measurements were used to identify field stars which were foreground, background, and within the dimensions of the simulated cluster’s natal cloud at each proposed distance. Stars within the cloud boundary were kept to represent an expected older star population, which is not included in our simulation data. The LoS  $V$ -band extinction due to the presence of the cloud was calculated for those identified as being within or background to the cloud, from column densities using a modified routine of the *Gaia* Simulator. These were then converted to a  $G$ -band extinction using the extinction law  $A_G/A_V = 0.789$  (Wang & Chen 2019) and the stars’ catalogued  $G$ -band magnitudes were adjusted accordingly. Those now fainter than  $G = 20.7 \text{ mag}$  were removed from the sample. No modification to the catalogues parameters of stars identified as foreground, or which had no or negative DR3 parallax measurements, were made. Fig. 3 shows an example of the changes in  $G$ -band magnitudes obtained for the DR3 survey samples.

We note that although *Gaia* provides the highest precision stellar parallax measurements to date, there are still non-negligible uncertainties on stars’ true  $z$ -axis (LoS distance) positions. This is because the parallax and associated errors of the *Gaia* catalogues represent the mean and standard deviation in a Gaussian distribution of potential true parallaxes, meaning there is a small probability that a star has a true parallax significantly larger or smaller than the given value. Thus, it is expected that some stars we have identified as foreground, within or behind the cloud have been inadvertently mislabelled.

We justify taking the parallax measures at face value rather than use a posterior probability distribution (like, e.g. Bailer-Jones et al. 2021) as the number of stars mislabelled is expected to be small, and statistically insignificant w.r.t. to populating our parameter space with a plausible field star distribution. Furthermore, we chose not to assign distances to stars that had no or negative parallaxes as parallax, rather than distance, will be used to identify cluster members (Section 5), i.e. these stars would be excluded during that stage regardless.

Fig. 1 shows the positions and colour–magnitude diagram for one of our data sets, Orion-type-3, before (top) and after (second panel) the field stars are added, assuming a distance of 2.5 kpc. As would be expected, the field stars engulf the simulated region; however, it is still substantial enough to be visible both in position space and as a sequence in the colour–magnitude diagram.

### 3.4 Resolution and quality limitations

Having merged our simulated data sets with real stars from *Gaia*, we now remove stars that would not be observed due to resolution and/or quality limitations.

*Gaia*’s ability to resolve two visually close stars as separate sources varies as a function of angular separation,  $s$ , and  $G$ -band magnitude contrast. Using close pairs from the full EDR3 catalogue, *Gaia* Collaboration (2021b) empirically defined this function as

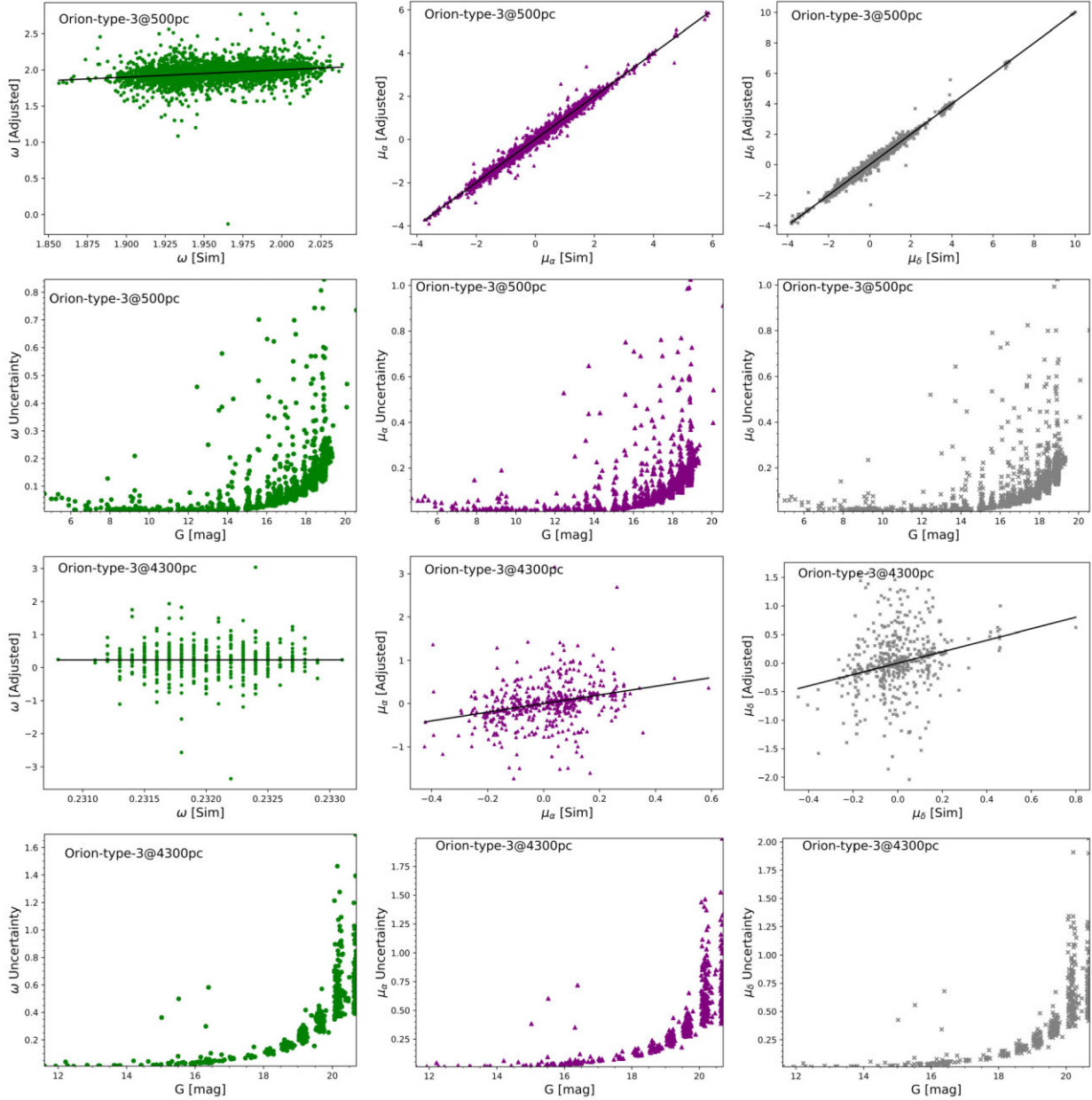
$$s_{\text{min}} = 0.532728 + 0.075526\Delta G + 0.014981(\Delta G)^2 \quad (1)$$

where  $s_{\text{min}}$  is the minimum angular separation in arcseconds between two stars needed by *Gaia* to be resolved as separate sources, and  $\Delta G$  is the difference in their  $G$ -band magnitudes. Cases where  $s < s_{\text{min}}$

<sup>3</sup><https://github.com/zkhorrami/gaiaSimulations>

<sup>4</sup><https://github.com/zkhorrami/MYOSOTIS>





**Figure 2.** The parallax and proper motion uncertainties, and adjusted values, shown as functions of  $G$ -magnitude and original value from the simulations for stars in Orion-type-3 within the magnitude limit of *Gaia* at (rows 1 and 2) 500 pc and (rows 3 and 4) 4300 pc. The solid black line shows a 1:1 relationship.

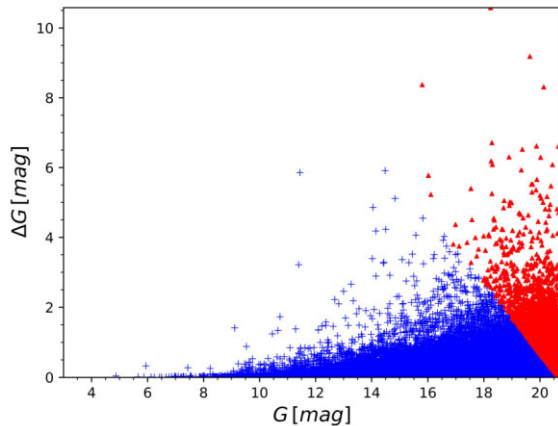
will include astrometric binaries that are detected as a single source, which will not fit a single-star astrometric solution resulting in a poor fit of the astrometric model and hence would be given a high RUWE by the *Gaia* reduction pipeline (Wang & Chen 2019).

We combined our synthetic observations (Section 3.2) with the DR3 field star selections (Section 3.3) to create a synthetic plus field catalogue. To identify unresolved cluster stars, the angular separation between all stars in the sample was calculated. When  $s < s_{\min}$  was found, both stars were assigned a  $\text{RUWE} > 1.4$ . We did not combine the fluxes of unresolved pairs as the high RUWE values resulted in their removal from the sample in the next quality-based selection cut. Of course, not all such stars would be detected as astrometric binaries, and although removing them all does not affect our results presented in Section 7, it does mean our CMDs in Fig. 1 lack the scatter often seen in *Gaia* data at faint magnitudes caused by unresolved physical

and line-of-sight binaries. For all remaining synthetic observation stars we assume the RUWE assigned in Section 3.2 ( $\approx 1.0$  for most stars), and for the remaining DR3 stars an unchanged RUWE from the listed value. Finally, a selection cut of  $\text{RUWE} < 1.4$  was made for all stars (synthetic and DR3 sources).

### 3.5 Kinematic data

Corrections to the proper motions of cluster stars in real observed catalogues are necessary to compensate for (i) perspective contraction, caused by members radial motions and (ii) perspective point convergence, caused by members sharing a common motion (see e.g. Buckner et al. 2020). As the *Gaia* simulator derives proper motion values by simply scaling the simulated  $X$  and  $Y$  velocities in the local standard of rest, these perspective effects are not present in



**Figure 3.** This figure shows the  $G$ -band magnitudes of stars in our DR3 sample identified as within or background to the SPH cloud of Orion-type-3, when the cluster is placed at a distance of 500 pc. On the  $x$ -axis are stars' real  $G$ -band magnitudes (as listed in the DR3 catalogue), and on the  $y$ -axis is the difference between the real and artificially increased  $G$ -band magnitudes (due to the presence of the cloud in their LoS). Crosses represent stars that are still seen by *Gaia* after the increase, and triangles are stars which are now fall below *Gaia*'s magnitude limit.

the synthetic observations and no correction is needed for it. We add the additional velocity to the stars owing to the galactic rotation curve at each observed distance, assuming a disc rotation of  $220 \text{ km s}^{-1}$ .

The proper motions of DR3 field stars comprise of two components, which are their true perpendicular motion to the observer's LoS (which includes the galactic rotation curve) and a component of velocity owing to solar rotation, i.e.

$$\mu = \frac{U_p + V_0 \cos l}{r}, \quad (2)$$

where  $\mu$  is the proper motion,  $U_p$  perpendicular velocity,  $r$  distance of a star,  $l$  galactic longitude, and  $V_0$  is the solar rotation velocity. As our LoS is along  $l = 270^\circ$  (Section 3.1), the term  $V_0 \cos l$  is zero, so no correction to the proper motions of DR3 stars are needed.

We consider *Gaia* spacecraft's scanning law, which causes small systematic errors in the proper motion measurements that increase with decreasing angular scale. Following the approach of Buckner et al. (2020), we use the values of these errors given in table 7 of Lindegren et al. 2021 to exclude all stars from our samples (simulated + real) with a proper motion within uncertainties smaller than the systematic error associated with the angular dimension of each synthetic observation.

To convert between  $\text{mas yr}^{-1}$  to  $\text{km s}^{-1}$  at each cluster distance, we use transformation factor  $\kappa$ , which has a value of 4.7405 at 1000 pc.

#### 4 OPTICAL IMAGES OF THE CLUSTER CATALOGUE

Although *Gaia* is not suited to producing images of star clusters, we wished to see how our data sets would appear visually, and (for example) whether the simulated clusters are readily apparent. We use the AMUSE-FRESCO software (Portegies Zwart & McMillan 2018; Rieder & Pelupessy 2023) to make optical images. We first calculate the flux of the stars in the Johnson–Cousins  $V$ ,  $R$ , and  $I$  bands from the *Gaia*  $G$ ,  $G_{BP}$ , and  $G_{RP}$  magnitudes. These magnitudes have already been adjusted for extinction from the gas. FRESCO then creates an image for each of these bands, using the positions of the stars (RA, Dec. mapped to  $x$ ,  $y$ ). This image is then convolved with

a point-spread function with a core, halo and diffraction spikes to simulate an image as it would be observed from a ground-based telescope. We create the point-spread function using a Moffat (1969) function with a full width at half-maximum of an arcsecond, and then add diffraction spikes as narrow two-dimensional Gaussians aligned along the cardinal directions. The resulting  $V$ ,  $R$ , and  $I$  images are used as blue, green, and red channels in the final image. Finally, we add a background image showing the column density of the gas to the red channel.

We show the optical images of the clusters in Fig. 4. The Orion-type cluster is visible as an overdensity of stars, but is difficult to see at the larger distances of 2500 and 4300 pc, and for the later time. The Wd2-type cluster is easily visible at all distances, which is not surprising as this clusters is so massive, and is also associated with substantial gas.

#### 5 STELLAR MEMBERSHIP

So far, our catalogue includes all observable stars (real *Gaia* and simulated) in the field of view of the simulated data sets. We now identify members of our clusters, or associations, from the catalogue as an observer might do assuming that, as our regions are quite substantial, they would already be known and, for example, listed in an earlier cluster catalogue such as DAML02<sup>5</sup>. Therefore, rather than search the samples for potential clusters, we attempt to identify members of clusters around their 'known' 5D ( $l$ ,  $b$ ,  $\omega$ ,  $\mu_{\alpha*}$ ,  $\mu_\delta$ ) position.

Three different membership criteria were explored.

- (i) *Tight* – have proper motions within  $0.5 \text{ mas yr}^{-1}$ , and parallax within  $0.3 \text{ mas}$ , of catalogued values.
- (ii) *Standard* – have proper motions within  $2 \text{ mas yr}^{-1}$ , and parallax within  $0.3 \text{ mas}$ , of catalogued values.
- (iii) *Minimal* – have a parallax within  $0.3 \text{ mas}$  of catalogued values (no proper motion constraints).

For all three sets of criteria, stars had to be within the catalogued angular radius, about the central longitude and latitude coordinates. The radii used were the maximum angular distance of a simulation star from the cluster centre, plus 10 per cent.

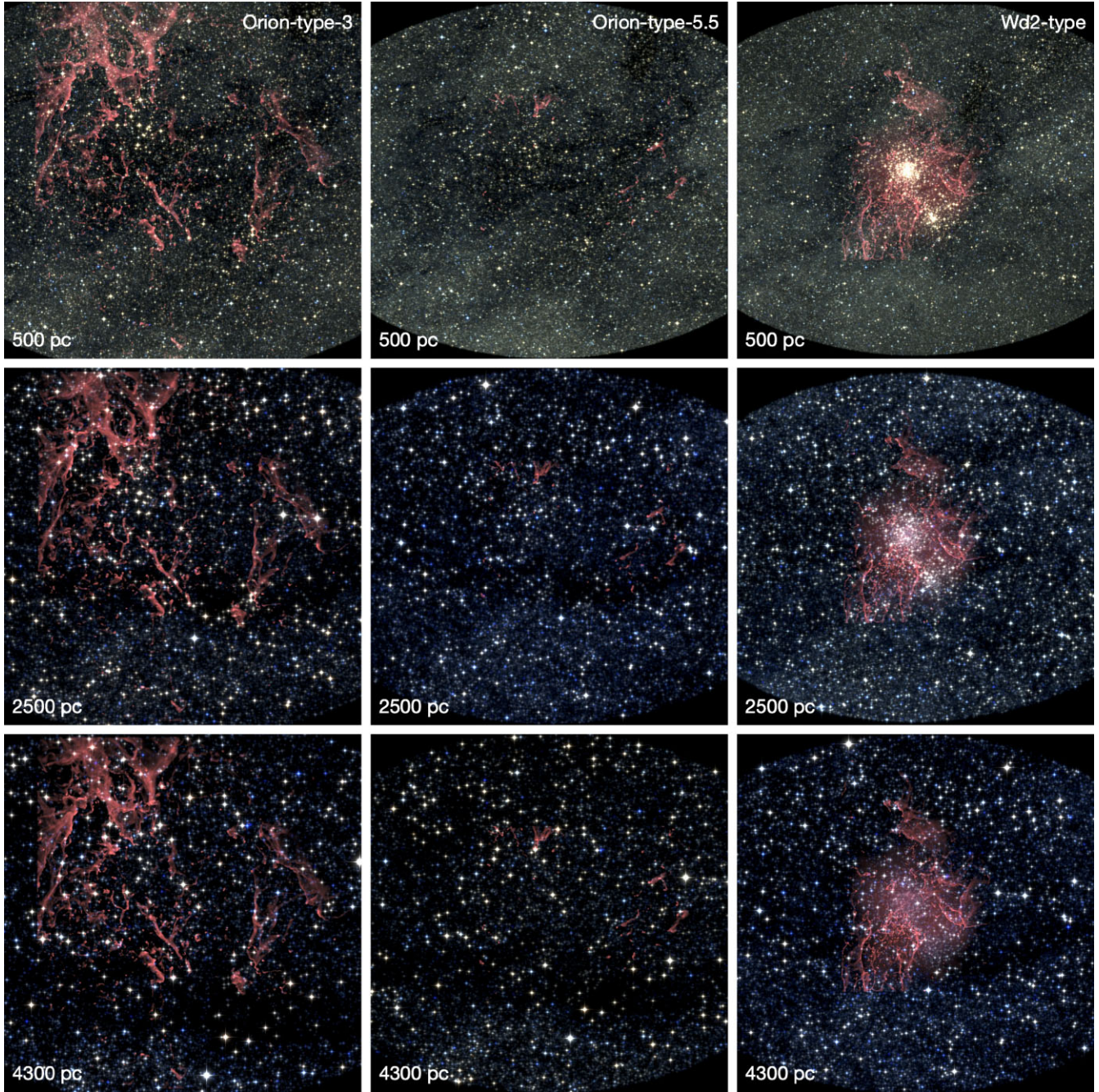
The criteria loosely follow the prescription of Cantat-Gaudin et al. (2018) who adopted these in conjunction with their unsupervised membership assignment code, UPMASK, to identify cluster members in the *Gaia* DR2 catalogue. The authors applied the Standard criteria to most clusters, with tight and minimal criteria used for compact and dispersed clusters in proper motion space clusters, respectively. They concluded applying the criteria in this way struck a good balance between keeping field contamination to a minimum but the criteria remained greater than the apparent dispersion of members for those clusters.

While we do not use UPMASK to identify cluster members, these criteria are a reasonable starting place to compile a membership list. Advantageously, in this work we know *a priori* which stars in our membership lists are true members and which are field contamination, so can quantify the effectiveness of each set of criteria for use in future observational studies. To this end, we applied all three sets of criteria to the clusters in our catalogue, regardless of their distribution in the proper motion space.

Cantat-Gaudin et al. (2018) also excluded stars fainter than  $G = 18$ . We experimented with this cut, but found that for our clusters this

<sup>5</sup><https://wilton.unifei.edu.br/ocdb/>





**Figure 4.** The clusters are shown as would be observed using an optical telescope, Orion-type-3 (left), Orion-type-5.5 (centre), and Wd2-type (right). The Orion-type cluster is visible at 500 pc, but difficult to identify by eye at 4300 pc, and for the later time. The Wd2-type cluster is readily observable at all distances. Gas is shown in red.

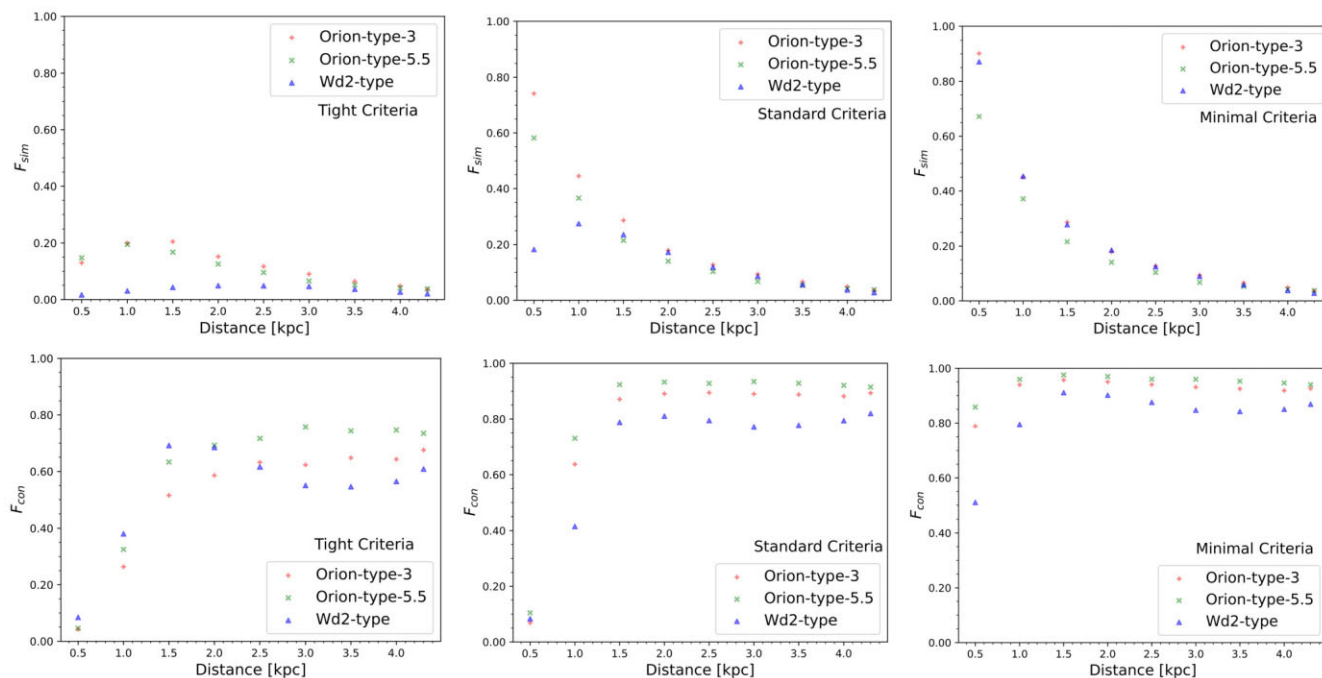
significantly reduced the number of members correctly identified for only marginal decreases in the contamination. Hence, we imposed no cut in magnitude.

An example of stars selected as members for the Orion-type-3 model using the minimal, standard and tight criteria is shown in Fig. 1 (lower panels). As shown, the minimal criteria do a poor job of selecting stellar members, including many stars which are clearly part of the field population. The standard and tight criteria recover stellar members with more success, both in position space and in the colour magnitude diagram, but overexclude stars, with the tight criterion making the greatest cuts.

In Fig. 5, we explore this in a more quantitative way using two metrics. The first is the number of actual members selected as members expressed as a fraction of the number of observable by *Gaia* ( $F_{\text{sim}}$ , upper panel). Of course, the members selected could include many stars that are not actual members, so our second metric is the fraction of the membership list that is actually non-members ( $F_{\text{con}}$ , lower panel). Our ideal criteria would capture a very large number of members (*i.e.* a large  $F_{\text{sim}}$ ) with a very low contamination ( $F_{\text{con}}$ ).

We can see that for the Orion-type models, the tight criteria achieves the best balance, with  $F_{\text{sim}}$  typically between 5 and 20





**Figure 5.** The panels show (top row:) fraction,  $F_{sim}$ , of observable simulation stars correctly identified as members using the different membership criteria and (bottom row:) contamination of membership lists by non-member field stars as given by the number of incorrectly assigned members as a fraction of the total assigned members for clusters. For the minimal criteria up to around 80 per cent of stars in the cluster are wrongly attributed, but the tight criteria is too restrictive particularly at small distances.

per cent and  $F_{con}$  below 75 per cent. Removing the proper motion cuts entirely for the Orion-type models (right-hand panels) results in unacceptably high contamination (around 95 per cent), while loosening them (middle panels) increases the contamination from around 70 to 95 per cent, with no notable gain in true members at distances greater than 1.5 kpc.

The tight criteria also works best for the Wd2-type models, except when the cluster is closer than 1.5 kpc. This is because the cluster has a relatively large velocity dispersion, because of its large mass, and hence many members are lost in the proper motion cuts. This is demonstrated in the upper right panel of Fig. 5, where removing the proper motion cuts allows us to recover most members. However, this gain is bought at the cost of an increased contamination rate (bottom right panel), showing that there is no good solution for a nearby Wd2-type cluster; its high-velocity dispersion makes proper motion selection problematical. This mirrors the conclusion Cantat-Gaudin et al. (2018) reached from real data that even for modest clusters at distances closer than the ones considered here, it is advantageous to drop proper motion cuts.

Although the tight criteria, on balance, provided the more favourable outcome for our clusters the contamination rate was surprisingly high. Plotting the positions of ‘members’ found using the tight criteria showed in nearly all cases clear structure at the centre (the real cluster) surrounded by an extended dispersed population, demonstrating clear evidence of contamination in the sample. Mindful that observers do not have *a priori* knowledge of the true appearance of a cluster, we attempted to reduce the contamination by making cuts in the RA/Dec. space to select regions with apparent structure present (Figs 6 and 7). This approach worked quite well, reducing the contamination between 42.4 and 91.9 per cent at a cost of typically less than a 5 per cent loss in true members.

Cantat-Gaudin et al. (2018) designed what we have called the standard criteria to be sufficiently wide to include most members, while eliminating a large number of non-members. With the advantage of a ‘ground truth’ we can see that the tight criteria with additional cuts in the position space after member selection produces relatively little contamination, and using this set of criteria the CMD and position plots seem reasonable, at least for the clusters we have simulated, and so we will use them for the remainder of the paper. We note that when placing our clusters outside the mid-plane (where stellar density is significantly lower) the standard criteria produced the best balance of members/contamination and convincing plots. Therefore we recommend careful consideration is given by observers to the galactic location and velocity dispersion of their cluster, as well as inspections of the CMD/position plots trialling both the tight and standard criteria, when considering which constraints works best, and if further constraints are needed, in their specific case.

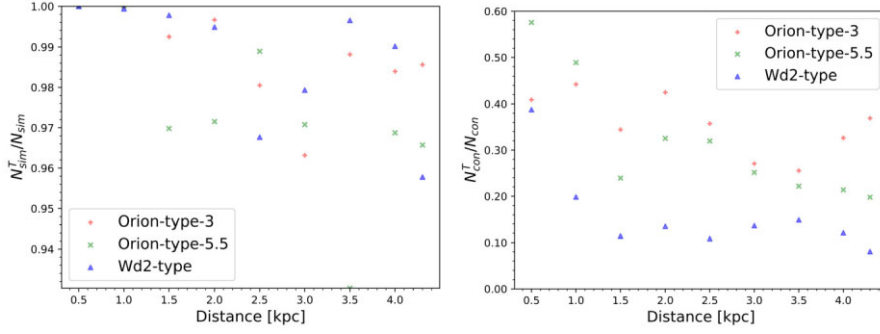
## 6 ANALYSIS METHODS

We analyse the spatial and dynamic properties of the clusters using the same techniques as in Paper I (Buckner et al. 2022a). We outline the techniques below, but for a full description see Paper I.

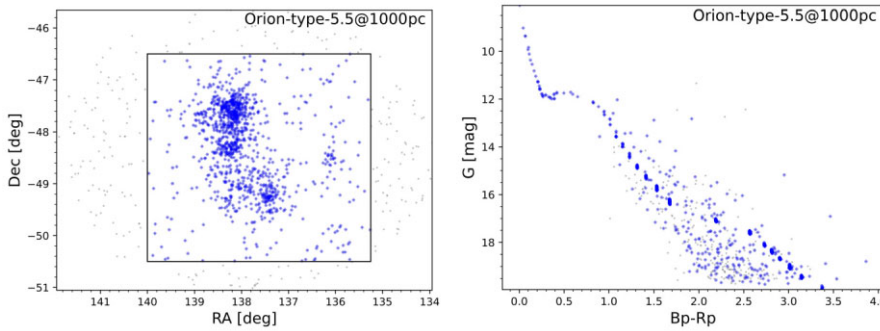
We assess the spatial association of cluster members using the 2 + D statistical clustering tool INDICATE<sup>6</sup> (INDEX to Define Inherent Clustering And Tendencies; Buckner et al. 2019). INDICATE is a local statistic which quantifies each star’s degree of association with other stars, given by an index  $I$ . This index is a ratio of the expected number of neighbours,  $N$ , for a star  $j$  if it was not spatially clustered and its

<sup>6</sup><https://github.com/abuckner89/INDICATE>





**Figure 6.** Plots show (left:) fraction of observable simulation stars correctly identified using the tight criteria,  $N_{\text{sim}}^T$ , still included as members after the position cut  $N_{\text{sim}}^T$ , (right:) fraction of contamination of membership lists by non-member field using the tight criteria,  $N_{\text{con}}^T$ , still included as members after the position cut assigned members for clusters  $N_{\text{con}}^T$ . By applying the position cuts, based on visible structure in the RA/Dec. plots, we have been able to reduce the contamination by up to 91.2 per cent with minimal impact on the identification of true members.



**Figure 7.** Example of the (left:) position cuts made to members identified using the tight criteria and (right:) the resulting CMD. Blue crosses and grey dots represent members included and excluded by the trim, respectively.

actual number of members,  $N_r$ . Following Buckner et al. (2020), we set  $N = 5$ , such that

$$I_{5,j} = \frac{N_r}{5} \quad (3)$$

The index is unit-less with a maximum value of  $\frac{S-1}{5}$ , where  $S$  is the total number of stars in the sample. It is calibrated against random distributions to identify significant values, with values greater than  $3\sigma$  above the mean value for random distributions representative of spatial clustering and higher values denoting progressively greater degrees of spatial association.

We also use the  $Q$  parameter (Cartwright & Whitworth 2004) to quantify spatial structure. Unlike INDICATE, this parameter provides a global, rather than local, assessment of the stellar spatial distribution based on the 2D positions of members. Values of  $Q < 0.8$ ,  $Q \approx 0.8$ , and  $Q > 0.8$  denote the cluster has a fractal, random or radial density gradient configuration.

We identify discrete clusters within the data using the Hierarchical Density-Based Spatial Clustering of ApplicationNs (HDBSCAN\*; Campello, Moulavi & Sander 2013) algorithm. HDBSCAN\* is similar to DBSCAN, except that for DBSCAN there is a constant search radius  $\epsilon$  which is used to determine whether stars within this distance are part of the same cluster. With HDBSCAN\*,  $\epsilon$  is variable enabling clusters of differing densities to be found. The only input parameter required was ‘*min samples*’. This provides a measure of how conservative the clustering is, and we determined the appropriate value on a cluster-by-cluster basis by checking that the clusters agreed with (i) those identified of visual inspection and (ii) the spatial distribution described by INDICATE.

In terms of the dynamics, we determine the median 2D velocity w.r.t. the system centre,  $v_{\text{out}}^*$ , of each cluster identified with HDBSCAN\*, where

$$v_{\text{out}}^* = \vec{v}_* \cdot \hat{r} \quad (4)$$

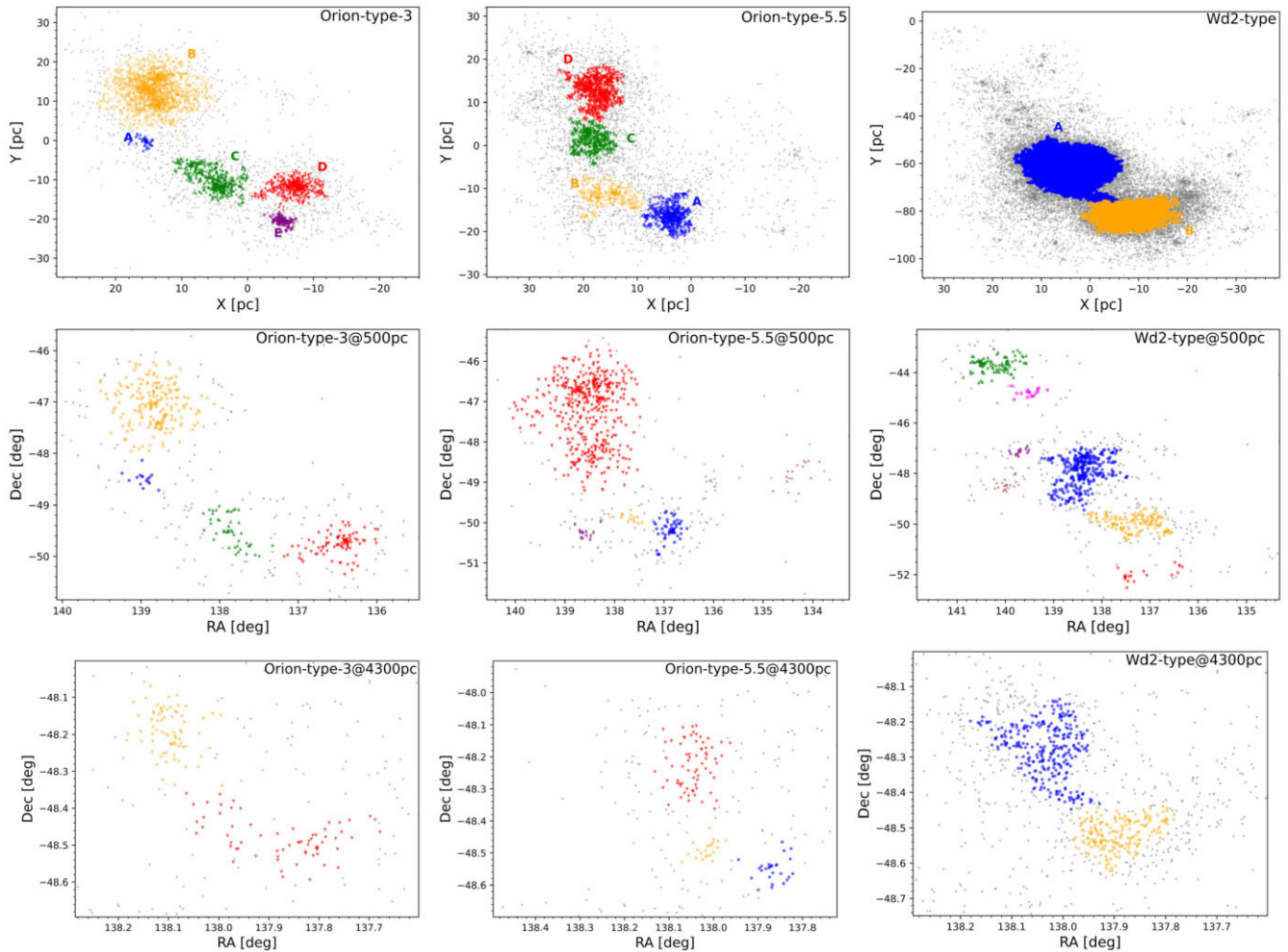
and  $\hat{r}$  is the 2D outward component of velocity.  $v_{\text{out}}^*$  will be positive if the cluster is expanding, negative if the cluster is contracting, and zero if static.

## 7 RESULTS

In this section, we discuss the 2D spatial and kinematic properties of the three cluster simulations, then compare these to those derived for our cluster catalogues to assess if these appear qualitatively similar, or whether the biases of *Gaia* derived membership lists are significant. We show two different measures of the spatial (Figs 8 and 9), and the kinematic (Fig. 10), characteristics of the stars in the original data set, and placed in the *Gaia* catalogue at distances of 500 and 4300 pc, the most extreme distances we used. Properties for the full catalogue are available in Tables A1–A4, and the original simulation in Table 2.

### 7.1 Spatial properties of the stars

HDBSCAN\* does a reasonable job identifying the rough substructure in our clusters, even at 4300 pc, despite relatively few stars being observable (Fig. 8). For the more massive Wd2-type region, field overdensities are erroneously detected as additional smaller clusters, only at distances less than 1.5 kpc. For the Orion-type-5.5 region,



**Figure 8.** Stellar subclusters found for the simulations and observations (smallest and greatest distance) with HDBSCAN\*. Members are denoted by same colours and symbols. Grey crosses represent stars not found to be in a subcluster.

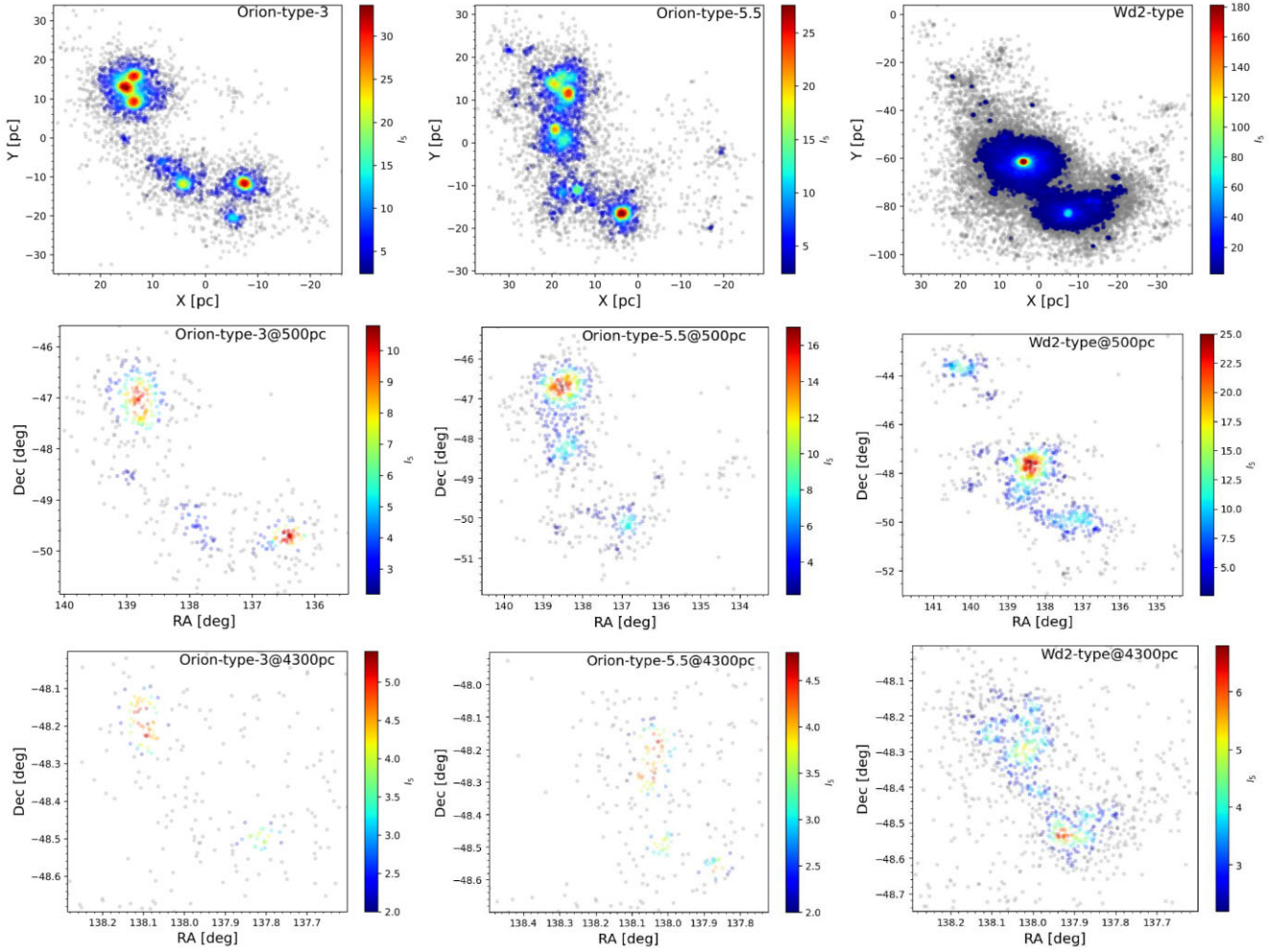
we see that two additional subclusters are picked out at 500 pc, and one additional cluster is found at 3000 pc, compared to the original data set. At the remaining distances, typically only three clusters are found, and clusters C and D are merged in all observations. In the Orion-type-3 cluster, larger subclusters B, C, and D are most commonly identified, with the smaller A and E clusters less commonly found and additional subclusters (asterisks) are rare. By 4300 pc, we see that only the largest clusters B and D are identifiable.

Fig. 9 shows the INDICATE  $I_5$  parameter that indicates the degree of spatial association for each star with its neighbours. All panels show the presence of spatially clustered stellar population in our clusters. The values of  $I_5$  are highest for the simulations of the Wd2-type cluster and for the Orion-type region at the earlier time compared to the later time. Similarly, the proportion of member stars that are found to be clustered (rather than dispersed) is highest for the Wd2-type and Orion-type-3 simulations. For the Orion-type cluster, this suggests that initially the majority of stars in these simulations are tightly spatially clustered, then some dispersal of stars occurs between the two time frames, with those still clustered in the later snapshot now more loosely so.

These trends for each cluster are mostly preserved when seen with *Gaia* at the two distances, though the relative intensities

between subclusters are not necessarily preserved due to confusion selecting their true members by HDBSCAN\*. The main differences with increasing distance in the Orion-type clusters are (i) an overall decrease in values of  $I_5$  (ii) a loss of resolution of fine structures, due to missing data, and (iii) the dispersal of stars between the two time frames is not observed at all distances. An example of (ii) is shown in Fig. 9, in the region of the Orion-type-3 simulation at  $(x, y) \approx (15, 15)$ , which has the greatest degree of association of the cluster and a complex substructure consisting of three smaller radial concentrations. When observed at both 500 and 4300 pc, this region is still perceived to have the greatest degree of association, but the complex substructure is lost. The Westerlund2-type cluster does not contain fine structures so we are unable to confirm the universality of this result. A similar overall decrease in values of  $I_5$  is observed in this cluster as with the Orion-types. Its two subclusters A and B are both consistently correctly identified as spatially clustered and their relative degrees of association ascertained (with the exception of 4300 pc where they appear to be the same; Table A4).

Regarding the global structure, the highest  $Q$  parameter is found for the Wd2-type simulation ( $Q = 0.9$ ). In the Orion-type cluster, there is an increase with time from 3 Myr ( $Q = 0.6$ ) to 5.5 Myr ( $Q = 0.7$ ). Similar behaviour is seen both in our observed catalogue,



**Figure 9.** Index values assigned to cluster members by INDICATE for the (top row:) simulations and observed by *Gaia* at (middle row:) 500 pc, and (bottom row:) 4300 pc (see text for details). Grey points indicate stars that are not clustered. The effects of crowding is apparent in bottom panels where there are missing stars ‘holes’ in the regions with the highest index values.

but typically the Orion-type cluster appears to be less substructured and differences between the two times are sometimes absent (Table A1). The largest change is observed for the Wd2-type cluster at 500 pc, which coupled with the HDBSCAN\* analysis may result in the cluster perceived to be more heavily substructured than reality.

The most obvious difference apparent in Figs 8 and 9 when observed is simply that there are far fewer stars. Fig. 5 indicates that typically fewer than 20 per cent of the stars are recovered. Our results indicate that despite this measures of the spatial structure of the stars are fairly robust.

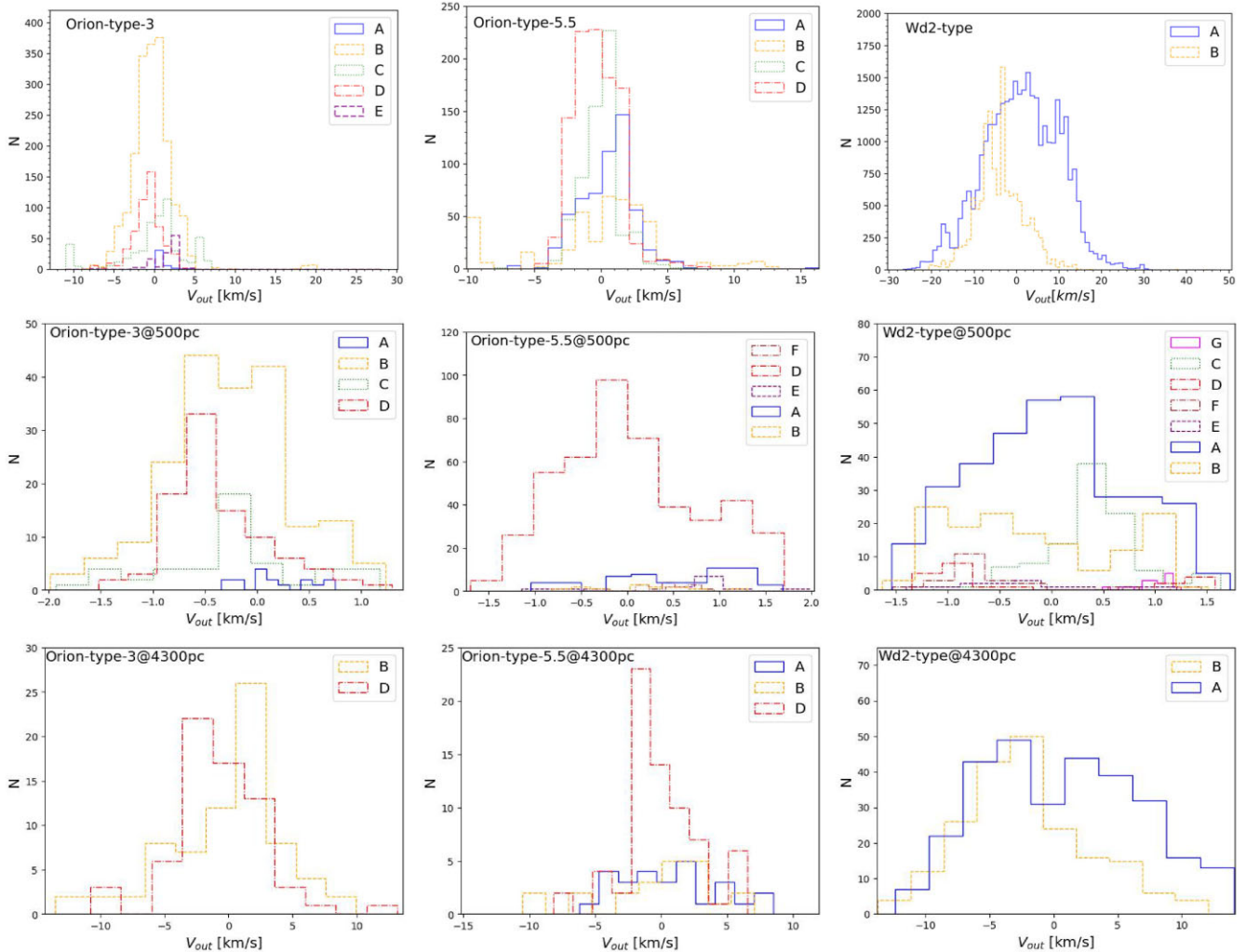
## 7.2 Kinematic properties of the stars

We calculate the overall  $v_{\text{out}}$  for each simulated region, and also for each subcluster within the region, with their properties listed in Table 2. Overall, the Orion-type region is undergoing expansion at both ages (3 and 5.5 Myr), with the overall expansion increasing from  $0.01 \text{ km s}^{-1}$  at the earlier time to  $0.19 \text{ km s}^{-1}$  later. The synthetic observations (Table A1) of both Orion-type clusters give mixed results, with some appearing to contract and others expand. At some distances, the expansion increase is observed, and at others the

expansion appears to be slowing or an increased rate of contraction is occurring. On the other hand, the Westerlund2-type simulation is undergoing contraction at  $-0.65 \text{ km s}^{-1}$ , which is correctly identified within an order of magnitude in the observations  $> 1 \text{ kpc}$  and two orders of magnitude at 500 pc.

The net expansion or contraction of the subclusters identified in Section 7.1 can be misidentified. The distributions of the velocities, simulated and observed, are shown in Fig. 10 and properties of the latter are detailed in Tables A2–A4. Generally, the simulated velocities of the clusters have normal distributions suggesting they are true kinematic structures rather than asterisms. The lower panels of Fig. 10 show the 500 and 4300 pc observations. For the Orion-type-3 region, the subclusters still tend to have normal distributions, but these become flatter with distance and some (such as A and E) may be mistaken as not real kinematic structures. The presence of outliers is clear at 4300 pc, and is an issue in clusters at distances  $> 2500 \text{ pc}$ . Their large values suggest they represent stars that have been mistakenly identified as members (Fig. 10), which is true in some cases, but in others are real members with either inaccurate velocity measurements, appear more prominent due to missing data in the distribution, and/or assigned to the incorrect subcluster. Similar results are found for the Orion-type-5.5 region, and asterisms E and





**Figure 10.** Directional 2D velocity histograms (w.r.t. the system centre) for the stellar subclusters found in the simulations and observations (smallest and greatest distance) using HDBSCAN\* defined in Fig. 8.

**Table 2.** Properties of subclusters found in the simulations.

Cluster	Name	Total members	$\bar{V}_{\text{out}}$ (km s $^{-1}$ )	$\bar{I}_5$
Orion-type-3	A	41	0.86	3.8
	B	1918	-0.21	11.0
	C	541	0.71	6.8
	D	524	-0.72	8.6
	E	113	2.07	9.2
Orion-type-5.5	A	565	0.60	8.2
	B	452	0.71	4.8
	C	604	0.09	8.4
	D	1038	-0.48	10.7
Wd2-type	A	33 001	1.28	19.4
	B	12 379	-4.04	9.6

F are clearly identifiable from their velocity distributions, which are flat.

The Wd2-type region consists of two subclusters, one of which is expanding and the other is contracting. Again, the velocities follow roughly normal distributions. The mean kinematic behaviour of A and B is correctly identified at distances  $>1500$  pc, most likely as

they contain many stellar members and exhibit overall broad velocity distributions that are merely narrowed by the selection criteria. Most of the asterisms found at 500 pc are clearly identifiable as such from their velocity distributions, but the larger subcluster C found at both 500 and 1000 pc may be mistaken as a true structure.

## 8 DISCUSSION AND CONCLUSIONS

Despite more stars having full 3D spatial and kinematic information than ever before, these remain a minority with most stars lacking good radial positions within the cluster and radial velocities. *Gaia* provides radial velocities only for brighter stars and even the new generation of wide-field high-multiplex fibre-fed spectrographs can only obtain  $\sim 10^3$  radial velocities in an hour. Parallaxes are more readily available, but even *Gaia* parallaxes still cannot pinpoint the position of a star within a star-forming region at 500 pc accurately enough to yield a true 3D view. Hence, in Paper I, we explored the impact of perspective effects on the accurate derivation of cluster properties, finding that while it is possible to obtain the correct qualitative conclusions, quantitative conclusions are likely to be inaccurate. However, observations of real clusters also suffer from observational biases such as *Gaia*'s ability to detect and

resolve stars, the effects of extinction and the ability of observers to differentiate between true members and foreground or background stars. Therefore, in this paper, we have addressed the effects of these problems on the results of 2D spatial-kinematic star cluster analyses.

We have taken the results of two simulations that produce two quite different types of object, and placed them, as close as possible to do so, in the real sky, using the *Gaia* catalogue of Milky Way stars. One simulation produces a stellar association similar to the Orion OB1 association, which we study at two different time frames, or ages. The other simulation produces two dense clusters, not dissimilar to Westerlund 2. Thus, we can make reasonable predictions about our ability to observe Orion-type associations, and YMCs like Westerlund 1 and 2, at different distances from the Sun.

Unsurprisingly, given its mass and density, the Westerlund 2 type cluster is easy to observe at all the distances we tested, up to 4.3 kpc. This is true even though this region is relatively young and still contains gas. We were able to identify that the cluster consists of two components, or subclusters, and the correct expansion/contraction of each at large distances. In fact, the Wd2-type cluster is harder to correctly observe at smaller distances from the Sun, due to the large spread of velocities, and the application of proper motion cuts in order to select members of the cluster.

The Orion-type association is more typical of regions closer to the Sun, and the types of objects that will be harder to study at larger distances. However, even though the number of stars detected at large distances in some cases comprises less than 2 per cent of the original member list, we still recover the basic properties of the region, including the identification of three or four clusters. As the distance to the association increases however, the ability to identify substructures decreases, and we see examples of clusters that are missed, or merged together into a single feature (compared to the original simulation data), but ultimately the cluster still appears to have some substructure. Likewise, the degree of clustering, as indicated by the INDICATE analysis decreases, but the majority of trends are preserved even when the effects of crowding cause ‘holes’ to appear in the densest regions (in agreement with Buckner et al. 2022b).

Generally, the spatial characteristics of the clusters are significantly more robust than the kinematics. With a relatively high contamination to member ratio and less than 20 per cent of real members identified as such we, perhaps unsurprisingly, find it can be difficult to correctly determine whether subclusters are expanding or contracting, compared to the original simulation data, and in the case of the Orion-type cluster the region as a whole as well. We also see that the velocity distributions become flatter with distance, which could lead to clusters mistakenly being identified as asterisms. Reassuringly, we find most asterisms misidentified as subclusters in the spatial analysis are typically easily identifiable as such when their velocity distributions are examined.

All the results above assume that the ‘tight criteria’, based on the criteria suggested by Cantat-Gaudin & Anders (2020), with additional position-based trimming of the membership list, are used to identify the larger scale cluster or association. Our analysis strongly supports using this criteria, rather than the minimal or standard spatial and kinematic criteria for young clusters in the galactic mid-plane and/or regions with a high stellar density. With the standard criteria, the contamination of the membership lists was significant, particularly for our smaller simulations at larger distances, with between  $\sim 7$  and 93 per cent of ‘members’ actually being foreground or background to the cluster. For the minimal criteria, the contamination of the membership lists was between  $\sim 51$  and 98 per cent, high-

lighting the importance of proper motion cuts during membership selection.

Overall, we find that observational biases do play a role in the accuracy of perceived 2D properties of clusters. Qualitative conclusions can reasonably be obtained for the spatial properties of YMCs and large Orion-scale associations up to distances of 4.3 kpc using *Gaia*, but observers should take care when interpreting the significance of any given result. While it is possible to obtain reasonable kinematic properties, these can be unreliable and lead to the wrong conclusions of contraction/expansion. Our findings here compound the results from Paper I meaning that, although found properties should be mostly indicative of true behaviour, the majority are unlikely to be quantitatively reliable.

In this paper, we have considered two regions that would be relatively easy to observe in our Galaxy, due to their size and age. In further work, we will consider clusters or associations that are more challenging to observe, and see whether such objects could be detected with *Gaia*, or indeed other instruments.

## ACKNOWLEDGEMENTS

AB, CLD, and TN were funded by the European Research Council H2020-EU.1.1 ICYBOB project (grant no. 818940). SR acknowledges funding from the European Research Council Horizon 2020 research and innovation programme (grant no. 833925, project STAREX). This work has made use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

## DATA AVAILABILITY

The final stage field-of-view files, from which stellar members are selected, at 500, 2500, and 4300 pc for all three clusters are available to download from the Zenodo repository at <https://zenodo.org/records/10053996>. Additional distances and analysis data underlying this article will be shared on reasonable request to the corresponding author.

The simulation data underlying this article were provided by CLD by permission, which will be shared on request to the corresponding author with permission of CLD.

## REFERENCES

- Ascenso J., Alves J., Lago M. T. V. T., 2009, *Ap&SS*, 324, 113  
 Bailer-Jones C. A. L., Rybizki J., Foesneau M., Demleitner M., Andrae R., 2021, *AJ*, 161, 147  
 Buckner A. S. M. et al., 2019, *A&A*, 622, A184  
 Buckner A. S. M. et al., 2020, *A&A*, 636, A80  
 Buckner A. S. M., Liow K. Y., Dobbs C. L., Naylor T., Rieder S., 2022a, *MNRAS*, 514, 4087  
 Buckner A. S. M., Khorrami Z., González M., Lumsden S. L., Clark P., Moraux E., 2022b, *A&A*, 659, A72  
 Campello R. J. G. B., Moulavi D., Sander J., 2013, in Pei J., Tseng V. S., Cao L., Motoda H., Xu G. eds, *Advances in Knowledge Discovery and Data Mining*. Springer, Berlin/Heidelberg, p. 160  
 Cantat-Gaudin T., 2022, *Universe*, 8, 111  
 Cantat-Gaudin T., Anders F., 2020, *A&A*, 633, A99  
 Cantat-Gaudin T. et al., 2018, *A&A*, 618, A93

- Cantat-Gaudin T. et al., 2019, *A&A*, 624, A126
- Carraro G., Turner D., Majaess D., Baume G., 2013, *A&A*, 555, A50
- Cartwright A., Whitworth A. P., 2004, *MNRAS*, 348, 589
- Castro-Ginard A., Jordi C., Luri X., Cantat-Gaudin T., Balaguer-Núñez L., 2019, *A&A*, 627, A35
- Castro-Ginard A. et al., 2020, *A&A*, 635, A45
- Castro-Ginard A. et al., 2022, *A&A*, 661, A118
- Dale J. E., Bonnell I., 2011, *MNRAS*, 414, 321
- Dias W. S., Monteiro H., Moitinho A., Lépine J. R. D., Carraro G., Paunzen E., Alessi B., Vilella L., 2021, *MNRAS*, 504, 356
- Dias W. S., Monteiro H., Caetano T., Hickel G., Alves G., Sacchi M., Lépine J. R. D., Amarinho N., 2022, *MNRAS*, 512, 4464
- Dobbs C. L., Bending T. J. R., Pettitt A. R., Buckner A.S.M Bate M. R., 2022, *MNRAS*, 517, 675
- Ferreira F. A., Corradi W. J. B., Maia F. F. S., Angelo M. S., Santos J. F. C. J., 2020, *MNRAS*, 496, 2021
- Ferreira F. A., Corradi W. J. B., Maia F. F. S., Angelo M. S., Santos J. F. C. J., 2021, *MNRAS*, 502, L90
- Froebrich D., Schmeja S., Samuel D., Lucas P. W., 2010, *MNRAS*, 409, 1281
- Gaia Collaboration, 2016, *A&A*, 595, A2
- Gaia Collaboration, 2018, *A&A*, 616, A1
- Gaia Collaboration, 2021a, *A&A*, 649, A1
- Gaia Collaboration, 2021b, *A&A*, 649, A6
- Gaia Collaboration, 2023, *A&A*, 674, A1
- Göppel C., Preibisch T., 2022, *A&A*, 660, A11
- Hao C. J., Xu Y., Wu Z. Y., Lin Z. H., Liu D. J., Li Y. J., 2022, *A&A*, 660, A4
- He Z., Wang K., Luo Y., Li J., Liu X., Jiang Q., 2022, *ApJS*, 262, 7
- Hur H., Park B.-G., Sung H., Bessell M. S., Lim B., Chun M.-Y., Sohn S. T., 2015, *MNRAS*, 446, 3797
- Jeffries R. D., 2007, *MNRAS*, 376, 1109
- Khorrami Z. et al., 2019, *MNRAS*, 485, 3124
- Kounkel M. et al., 2018, *AJ*, 156, 84
- Kounkel M., Covey K., Stassun K. G., 2020, *AJ*, 160, 279
- Kroupa P., 2001, *MNRAS*, 322, 231
- Li Z. et al., 2022, *ApJS*, 259, 19
- Lindgren L. et al., 2021, *A&A*, 649, A2
- Liow K. Y., Rieder S., Dobbs C. L., Jaffa S. E., 2022, *MNRAS*, 510, 2657
- Mahmudunobe M., Hasan P., Raja M., Hasan S. N., 2021, *Eur. Phys. J. Spec. Top.*, 230, 2177
- Menten K. M., Reid M. J., Forbrich J., Brunthaler A., 2007, *A&A*, 474, 515
- Moffat A. F. J., 1969, *A&A*, 3, 455
- Monteiro H., Dias W. S., 2019, *MNRAS*, 487, 2385
- Motte F., Bontemps S., Louvet F., 2018, *ARA&A*, 56, 41
- Panwar N., Sharma S., Ojha D. K., Baug T., Dewangan L. K., Bhatt B. C., Pandey R., 2020, *ApJ*, 905, 61
- Pfalzner S., Kaczmarek T., Olczak C., 2012, *A&A*, 545, A122
- Portegies Zwart S., McMillan S., 2018, *Astrophysical Recipes: The art of AMUSE*. IOP Publishing, Bristol, UK
- Quintana A. L., Wright N. J., 2021, *MNRAS*, 508, 2370
- Rain M. J., Ahumada J. A., Carraro G., 2021, *A&A*, 650, A67
- Rieder S., Pelupessy I., 2023, rieder/Fresco: v2023.3.0
- Rieder S., Dobbs C., Bending T., Liow K. Y., Wurster J., 2022, *MNRAS*, 509, 6155
- Sim G., Lee S. H., Ann H. B., Kim S., 2019, *J. Korean Astron. Soc.*, 52, 145
- Wang S., Chen X., 2019, *ApJ*, 877, 116

## APPENDIX A:

**Table A1.** Table of observed  $\mathcal{Q}$  and  $V_{\text{out}}$  values for our cluster catalogue.

Distance (pc)	Cluster	$\mathcal{Q}$	$V_{\text{out}}$ (km s $^{-1}$ )
500	Orion-type-3	0.6	−0.21
	Orion-type-5.5	0.7	0.13
	Wd2-type	0.6	−0.01
1000	Orion-type-3	0.7	−0.11
	Orion-type-5.5	0.9	0.06
	Wd2-type	0.8	−0.09
1500	Orion-type-3	0.7	−0.09
	Orion-type-5.5	0.8	−0.05
	Wd2-type	0.8	−0.24
2000	Orion-type-3	0.7	0.03
	Orion-type-5.5	0.8	0.01
	Wd2-type	0.8	−0.31
2500	Orion-type-3	0.7	0.18
	Orion-type-5.5	0.7	0.31
	Wd2-type	0.8	−0.51
3000	Orion-type-3	0.7	0.11
	Orion-type-5.5	0.7	0.12
	Wd2-type	0.8	−0.41
3500	Orion-type-3	0.7	−0.14
	Orion-type-5.5	0.7	0.36
	Wd2-type	0.8	−0.56
4000	Orion-type-3	0.7	−0.09
	Orion-type-5.5	0.8	−0.18
	Wd2-type	0.8	−0.31
4300	Orion-type-3	0.7	−0.46
	Orion-type-5.5	0.8	−0.82
	Wd2-type	0.8	−0.77



**Table A2.** Table of subcluster properties found in the observations of Orion-type-3.

Distance	Cluster	Total members	$\tilde{V}_{\text{out}}$ (km s $^{-1}$ )	$\tilde{I}_5$
500	A	16	0.07	2.4
	B	196	-0.28	5.9
	C	47	-0.16	2.8
	D	94	-0.50	7.2
	E	-	-	-
1000	A	-	-	-
	B	407	-0.22	10.4
	C	118	0.40	4.4
	D	121	-0.60	7.0
	E	-	-	-
1500	A	-	-	-
	B	418	-0.25	5.8
	C	81	1.02	3.0
	D	105	-0.76	7.6
	E	17	1.76	2.8
2000	A	-	-	-
	B	330	-0.22	4.6
	C	185	0.76	1.2
	D	64	-1.15	7.5
	E	30	1.22	3.1
2500	A	-	-	-
	B	238	0.22	4.5
	C	95	0.61	1.6
	D	58	-1.69	6.3
	E	30	1.34	2.9
	F	31	-0.04	1.8
3000	A	-	-	-
	B	164	-0.12	4.6
	C	48	1.08	1.9
	D	69	-1.24	2.6
	E	-	-	-
3500	A	-	-	-
	B	116	0.04	4.2
	C	32	-0.13	1.5
	D	51	-1.51	4.8
	E	-	-	-
4000	A	-	-	-
	B	111	-0.04	5.2
	C	19	1.37	1.8
	D	37	-1.40	4.4
	E	-	-	-
4300	A	-	-	-
	B	73	0.75	4.0
	C	-	-	-
	D	66	-1.00	1.5
	E	-	-	-

**Table A3.** Table of subcluster properties found in the observations of Orion-type-5.5.

Distance	Cluster	Total members	$\tilde{V}_{\text{out}}$ (km s $^{-1}$ )	$\tilde{I}_5$
500	A	56	0.66	6.7
	B	13	0.15	2.6
	C	-	-	-
	D	458	-0.07	5.8
	E	13	0.79	2.4
	F	16	0.47	1.0
1000	A	160	0.77	3.6
	B	97	-0.10	8.4
	C	-	-	-
	D	364	-0.39	13.6
	A	383	0.37	1.8
1500	B	-	-	-
	C	-	-	-
	D	500	-0.26	2.8
	A	107	0.03	4.6
	B	103	0.60	3.0
2000	C	-	-	-
	D	360	-0.32	3.4
	A	88	0.15	3.4
	B	91	0.82	2.0
	C	-	-	-
2500	D	251	0.18	3.0
	A	64	0.33	2.2
	B	69	0.69	1.6
	C	-	-	-
	D	152	-0.11	3.2
3000	E	23	-1.42	1.4
	A	49	0.93	2.4
	B	59	1.17	2.0
	C	-	-	-
	D	120	0.08	3.5
3500	A	35	-0.23	3.0
	B	31	1.56	2.2
	C	-	-	-
	D	89	-0.45	3.4
	E	27	0.09	2.8
4000	A	27	0.09	2.8
	B	22	1.10	3.2
	C	-	-	-
	D	69	-0.50	3.6

**Table A4.** Table of subcluster properties found in the observations of Westerlund2-type.

Distance	Cluster	Total members	$\bar{V}_{\text{out}}$ (km s <sup>-1</sup> )	$\bar{I}_5$
500	A	332	-0.01	14.3
	B	143	-0.36	7.8
	C	105	0.42	7.6
	D	21	-0.95	2.0
	E	13	-0.27	3.2
	F	22	-0.73	3.4
	G	17	1.01	2.6
1000	A	721	0.08	11.0
	B	302	-0.85	7.0
	C	99	0.92	3.2
1500	A	1083	-0.02	9.2
	B	665	-1.38	6.8
2000	A	1139	0.04	9.4
	B	686	-1.82	8.2
2500	A	981	0.19	5.2
	B	608	-2.20	5.0
3000	A	987	0.31	5.2
	B	570	-2.45	4.2
3500	A	656	0.50	8.6
	B	512	-2.94	6.6
4000	A	465	0.89	6.8
	B	302	-2.76	6.4
4300	A	296	0.76	3.2
	B	200	-2.59	3.2

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