








The spatial evolution of star clusters in NGC 628 with *JWST*

Anne S. M. Buckner¹ , ¹★ Ana Duarte-Cabral¹ , Angela Adamo,² Sean T. Linden,³ Michele Cignoni^{4,5,6} , Varun Bajaj,⁷ Arjan Bik,² Giacomo Bortolini² , Daniela Calzetti,⁸ Matteo Correnti,^{5,9} Bruce G. Elmegreen,¹⁰ Debra M. Elmegreen,¹¹ Helena Faustino Vieira,² John S. Gallagher, III,¹² Kathryn Grasha^{13,14} , Benjamin Gregg,¹⁵ Rob A. Gutermuth,⁸ Kelsey Johnson,¹⁶ Mark Krumholz¹³ , Drew Lapeer,⁸ Matteo Messa¹⁵ , Göran Östlin,² Alex Pedrini,² Jenna E. Ryon,⁷ Linda J. Smith¹⁷ and Monica Tosi⁵

¹Cardiff Hub for Astrophysics Research and Technology (CHART), School of Physics & Astronomy, Cardiff University, The Parade, CF24 3AA Cardiff, UK

²Department of Astronomy, Oskar Klein Centre, Stockholm University, AlbaNova University Center, SE-106 91 Stockholm, Sweden

³Steward Observatory, University of Arizona, 933 N Cherry Avenue, Tucson, AZ 85721, USA

⁴Dipartimento di Fisica, Università di Pisa, Largo Bruno Pontecorvo 3, 56127 Pisa, Italy

⁵INAF–OAS, Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Gobetti 93/3, I-40129 Bologna, Italy

⁶INFN, Largo B. Pontecorvo 3, 56127 Pisa, Italy

⁷Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

⁸Department of Astronomy, University of Massachusetts, 710 North Pleasant Street, Amherst, MA 01003, USA

⁹ASI-Space Science Data Center, Via del Politecnico, I-00133, Rome, Italy

¹⁰Katonah, NY 10536, USA

¹¹Vassar College, Department of Physics & Astronomy, Poughkeepsie, NY 12604, USA

¹²Department of Astronomy, University of Wisconsin, 475 North Charter St., Madison, WI 53706, USA

¹³Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia

¹⁴ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia

¹⁵Department of Astronomy, University of Massachusetts Amherst, Amherst, MA 01003, USA

¹⁶Department of Astronomy, University of Virginia, Charlottesville, VA, USA

¹⁷Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

Accepted 2025 November 7. Received 2025 November 4; in original form 2025 July 23

ABSTRACT

We examine the spatial distribution of star clusters in NGC628 using the statistical tool INDICATE to quantify clustering tendencies. Our sample, based on *Hubble Space Telescope* and *James Webb Space Telescope* observations, is the most complete to date, spanning ages from 1 to >100 Myr. We find cluster spatial behaviour varies with galactic position, age, and mass. Most emerging young clusters are tightly spatially associated with each other, while fully emerged clusters are in ~ 1.5 times looser spatial associations, irrespective of age. Young Massive Clusters (YMCs $\geq 10^4 M_\odot$) tend to associate with lower-mass clusters but not strongly with other YMCs, implying that intense star formation regions produce a few YMCs alongside many lower-mass clusters rather than multiple YMCs together. Young concentrated clusters show a wide radial distribution in the galactic disc, which narrows with age; with concentrated clusters > 100 Myr mostly residing between 2 and 6 kpc. This pattern may reflect either faster dispersal of isolated tight cluster spatial ‘structure’ in a lower gas density outer disc or gradual inside-out growth, with the formation of this structure shifting outwards over time. We also detect distinct spatial behaviours for clusters within 2 kpc, linked to the inner Lindblad resonance (≤ 1 kpc), nuclear ring (~ 0.5 –1 kpc), and the start of spiral arms (~ 1.25 –2 kpc), suggesting these regions exhibit strong radial motions that could hinder clusters from forming and remaining in tight concentrations. Our results highlight how spatially resolved studies of clusters can reveal the influence of galactic dynamics on star formation and cluster evolution.

Key words: galaxies: individual: NGC 628 – galaxies: spiral – galaxies: star clusters: general – galaxies: statistics – galaxies: stellar content – galaxies: structure.

1 INTRODUCTION

Star clusters are the fundamental building blocks of spiral galaxies as most stars form in clustered environments. Unfortunately, the

mechanisms of cluster formation, effect of local and galactic environmental conditions have on this, as well as their longer-term evolution, remain unclear despite considerable efforts by both the theoretical and observational communities. To better constrain the role of galactic environment requires a multipronged analysis of cluster populations, comparing variation in their properties (age,

* E-mail: bucknera@cardiff.ac.uk

mass, stellar composition) as a function of galactic environment and conditions. Young Massive Clusters (YMCs) – characterized as having ages < 100 Myr and masses $\geq 10^4 M_\odot$ – are of particular interest as they provide important clues about the physics underpinning massive cluster and star formation. Thus, when examined as part of larger cluster population studies, the influence of the galactic environment on both the early and longer-term evolution of massive clusters can be explored. The advent of high-resolution infrared data from the *James Webb Space Telescope* (*JWST*; J. P. Gardner et al. 2006) provides the opportunity to greatly advance our understanding of these processes, as many of the youngest clusters – still associated with their natal clouds – are observable for the first time.

Here, we characterize the spatial properties of the star cluster population of spiral galaxy NGC 628. Also known as Messier 74, it is a multiple arm galaxy of type SA(s)c with two main spiral arms and is the principal galaxy of the M74 galaxy group. NGC 628 has an estimated total stellar mass of $2.2 \times 10^{10} M_\odot$ (A. K. Leroy et al. 2021) is nearby (9.84 Mpc; R. B. Tully et al. 2009), bright ($V = 9.46$ mag; D. O. Cook et al. 2014), face-on ($i = 8.9^\circ \pm 12.2^\circ$; A. K. Leroy et al. 2021), and is actively forming stars with an estimated star formation rate (SFR) of $1.74 M_\odot \text{yr}^{-1}$ (A. K. Leroy et al. 2021). NGC 628 has several multiwavelength data sets available, including recent *JWST* (J. P. Gardner et al. 2023) observations, making it, unsurprisingly, one of the best studied nearby spiral galaxies. It has had studies devoted to its star formation history and efficiency as a function of environment (K. Kreckel et al. 2016; K. Kreckel et al. 2018; M. Lomaeva et al. 2022; J. Kim et al. 2023; F. Demachi et al. 2024); the distribution and properties of giant molecular clouds (C. N. Herrera et al. 2020; F. Demachi et al. 2023; H. Faustino Vieira et al. 2025); regularly-spaced $8\mu\text{m}$ cores along the spiral arms, H II regions, Polycyclic Aromatic Hydrocarbons and dust (L. Rousseau-Nepton et al. 2018; B. G. Elmegreen & D. M. Elmegreen 2019; D. A. Dale et al. 2023; O. V. Egorov et al. 2023; D. A. Thilker et al. 2023; D. Calzetti et al. 2024; B. Gregg et al. 2024; K. Matsumoto et al. 2024; A. Pedrini et al. 2024); feedback-driven bubbles (A. T. Barnes et al. 2023; E. J. Watkins et al. 2023); magnetic structure (D. D. Mulcahy, R. Beck & G. H. Heald 2017; M. Weżgowiec et al. 2022; M. Zhao et al. 2024); and the broader structure/evolution of the galaxy (S. Aniyani et al. 2018; A. A. Marchuk 2018; S. Inoue et al. 2021; K. Ujjwal et al. 2022).

Numerous studies have focused on the demographics of NGC 628’s star cluster population, extracting key characteristics using a wide-variety of statistically-based approaches. For example, both box-counting and surface brightness threshold methodologies have shown that the cumulative size distribution function of young stellar regions is a power law with a slope of approximately -1.5 (B. G. Elmegreen et al. 2006; A. S. Gusev 2014), while Bayesian modelling has revealed the absence of mass-dependent disruption in the galaxy (J. Tang, K. Grasha & M. R. Krumholz 2024). The use of cumulative distribution functions found that there is no age gradient across the spiral arms (F. Shabani et al. 2018) but clusters do display age-dependent spatial behaviour, such that the youngest strongly clustered together but the older population is non-clustered, as identified using the two-point correlation function (K. Grasha et al. 2015; K. Grasha et al. 2017; F. Shabani et al. 2018). Although these statistics are distinct, they are also similar in that they look at the ‘big picture’ trends across the population i.e. a single parameter is derived for a group of clusters, rather than individual clusters. Advantageously this allows typical galaxy-scale behavioural trends to be described (as a function of average age, mass, etc.). However, by design they do not describe the behaviours of individual clusters within each group, so cannot identify local-scale variances in behaviour, nor

quantify how various physical factors have impacted these. In this paper we will use the novel statistical clustering tool ‘INDICATE’ (INdex to Define Inherent Clustering And TEndencies; A. S. M. Buckner et al. 2019) to describe the spatial behaviour of each cluster in NGC 628’s population individually. We will quantify, for the first time, how small differences in cluster age, mass, and galactic position influence spatial behaviour of stellar clusters in NGC 628. Our study includes a significant number of recently discovered clusters still embedded within their natal clouds, which were found with high-resolution Near InfraRed (NIR) *JWST* images of NGC 628 obtained by the Feedback in Emerging extrAGalactic Star ClusTers (FEAST, GO-1783, PI A. Adamo) collaboration.

This paper is structured as follows. In Section 2, we describe our cluster catalogue, and in Section 3, we detail the methodology used to study their spatial properties. The results of our analysis for clusters’ spatial behaviour, including trends with galactic position, age, and mass, are presented in Section 4, and our conclusions given in Section 5.

2 DATA

In this work, we analyse the clusters of NGC 628, using a combined sample consisting of candidates identified using both *Hubble Space Telescope* (*HST*) and *JWST* observations. Below we describe how our sample was constructed and detail our selection criteria.

2.1 Initial Sample Selection

To ensure a complete (as possible) stellar cluster population for our analysis, we created our sample by combining the star cluster catalogues produced within the FEAST collaboration. The first catalogue consists of young clusters in the emergence phase (emerging young stellar clusters, eYSCs) which have been extracted in *JWST*/NIRCam imaging. The second catalogue contains stellar clusters extracted in optical *HST* imaging using a well-established approach that has led to the production of many star clusters catalogues in local Universe galaxies over the years (see e.g. M. R. Krumholz, C. F. McKee & J. Bland-Hawthorn 2019a for a review). Hereafter, we will refer to the sample of clusters extracted in this way as optical YSCs for simplicity, although we note that these optical clusters are not necessarily young. The two samples combined cover the entire cluster population from the earliest phases to clusters > 100 Myr old.

2.1.1 eYSC catalogue

The eYSC catalogue presented in Adamo et al. (in preparation) consists of 1576 emerging cluster candidates which are still associated with their H II regions and photodissociation regions. They were extracted from *JWST* Cycle 1 observations under the FEAST program (GO-1783), with additional archival data from program GO-2107 (PI J. Lee) included. We opted to utilize this catalogue for our analysis as it represents the most complete dedicated emergence stage cluster catalogue for NGC 628 to date. Full descriptions of observations, cluster identification, classification, and fitting procedures will be made available in Adamo et al. (in preparation), Linden et al. (in preparation), and A. Pedrini et al. (2025). In brief, the GO-1783 observations were carried out in the F115W, F150W, F187N, F200W, F277W, F335M, F405N, F444W, and GO-2107 observations in F300M and F360M filters. Data was processed using the calibration files 1.11.4, and the final mosaics have been aligned to *Gaia* and have a final pixel scale of $0.04 \text{ arcsec pixel}^{-1}$. eYSCs were identified as

candidates which have compact morphologies in Pa α , and if present, in Br α . Further differentiation was made between eYSCs of two evolutionary classes, Type 1 (eYSCI) and Type 2 (eYSCII), based on whether they have compact morphologies in 3.3 μ m Polycyclic Aromatic Hydrocarbons (PAH) or not, respectively (A. Pedrini et al. 2024; B. Gregg et al. 2024). This classification resulted in 835 and 741 entries respectively. Physical properties were obtained for the eYSCs using SED fitting with CIGALE (M. Boquien et al. 2019). CIGALE operates by generating a grid of models based on the user's input parameters, comparing a cluster's photometry with each model of the grid, and calculating the χ^2 value to determine the goodness-of-fit. The χ^2 value is then converted into a likelihood via $\exp(-\chi^2/2)$, which is then used as a Bayesian prior estimate for the physical properties of each source, where 1σ uncertainties are then determined by the likelihood-weighted standard deviation of all the models. Once CIGALE fits a cluster's spectral energy distribution (SED) based on the age and reddening, the mass is appropriately scaled based on the cluster's luminosity assuming a fully-sampled IMF. A Milky Way-like extinction curve from J. A. Cardelli, G. C. Clayton & J. S. Mathis (1989) was assumed, with an extinction factor of $R_V = 3.1$, solar metallicity, internal reddening in the range from $E(B-V) = 0$ to 5 mag in 0.01 mag step, ages from 1 to 10 Myr with $\Delta T = 1$ Myr, and total covering fractions which range from 0.01–1. The adopted upper age limit of 10 Myr follows from the FEAST selection, which identifies embedded candidates via strong hydrogen recombination emission (Pa α /Br α); the equivalent widths of the near-IR hydrogen lines (e.g. Pa β , Br γ) decline to negligible levels by ~ 10 Myr in Starburst99 models (e.g. C. Leitherer et al. 1999, 2014), so older clusters would not be selected by this tracer.

For our sample, we removed sources detected in less than 4 bands, and/or poor SED fit ($\chi^2 \geq 20$, $\leq 200 M_\odot$). This selection leaves 1216 clusters remaining, of which 653/563 are eYSCI/II, respectively.

2.1.2 Optical YSC catalogue

The FEAST optical YSC catalogue (Adamo et al. in preparation) consists of 16 645 cluster candidates which were extracted from archival *HST* data including the Legacy ExtraGalactic UV Survey (LEGUS; D. Calzetti et al. 2015) Hubble treasury program, GO-9796, and GO-13773. We used this catalogue for our analysis, as it is the largest extraction of the HST F555W FoV clusters to date, surpassing the previously published reference catalogue of LEGUS (A. Adamo et al. 2017). A comparison of this and the LEGUS catalogue is extensively discussed in Adamo et al. (in preparation). The FEAST catalogue shows clear improvements over the LEGUS catalogue due to a different adopted aperture correction method and the inclusion of narrow-band filters, capturing H α and Pa α , and NIR broadband capturing the stellar continuum.

We refer the reader to the original catalogue paper for full details of the extraction, classification, and fitting procedures. In brief, after the initial extraction was performed, final mosaics in WFC3 F275W, F336W, F657N, and ACS/WFC F435W, F555W, F606W, F658N, and F814W were drizzled to a common pixel scale of 0.04 arcsec pixel $^{-1}$ and registered to the Gaia reference frame. Properties for the candidates were derived by fitting their SED from 0.3 to 2 μ m (*HST* F275W, F336W, F657N, F435W, F555W, F606W, F658N, and F814W and NIRCам F115W, F187N, F200W) with Yggdrasil spectral evolutionary models (E. Zackrisson et al. 2011), with a metallicity of $Z = 0.02$ and a P. Kroupa (2001) initial mass function.

Nebular emission was included assuming a covering factor of 50 per cent and $n_e = 100 \text{ cm}^{-3}$.

For our analysis, we removed sources detected in less than 4 bands, flagged as 'star or spurious object', and/or poor SED fit ($\chi^2 \geq 65$, $\leq 200 M_\odot$), resulting in a total of 5806 optical YSCs remaining.

2.2 Final sample

We combine the selections of optical YSCs and eYSCs to create our final sample, made necessary as (i) high extinctions associated with eYSCs mean a significant portion are expected to be below the detection threshold of *HST* and thus absent from the optical YSC catalogue; (ii) despite there being significant overlap between the IR and optical coverage, due to the *JWST* mosaic orientation eYSCs are detected in some regions where there is no *HST* coverage (Fig. 1). To remove duplicate entries from the two catalogues, a cross-match is performed with the criteria that any eYSCs within a 4 pixel tolerance (0.16 arcsec) of an optical YSC (i.e. the aperture radius used to extract photometry) is considered a match. In the cases where more than 1 match is found for a single source, only the closest pairing is considered. A total of 136 matches were found, with four instances of an eYSC being paired with two YSCs, leaving the final combined sample of 6890 clusters. For the cross-matched YSCs we use the properties (age, mass) from the eYSC catalogue. The age and mass ranges for the sample are shown in Fig. 2 and the top panels of Fig. A2. Finally, each cluster is assigned (i) an environmental flag denoting their spatial position in NGC 628's centre, spiral arms or interarm regions, using the environmental mask of NGC 628 from M. Querejeta et al. (2021), as shown in Fig. 3 and (ii) an origin flag denoting which catalogue(s) the cluster originates from, namely:

- (i) YSC – optical YSC catalogue
- (ii) eYSC [NIR] – eYSC catalogue
- (iii) eYSC [optical] – optical YSC and eYSC catalogue.

A comparison between the physical properties obtained for clusters using CIGALE and Yggdrasil SED fitting has been undertaken by Linden et al. (in preparation), who show that they are in very good agreement within the error bars. Therefore, any differences we find in the spatial properties of eYSCs and optical YSCs during our analysis are not attributable to how the physical properties were derived.

3 METHOD

3.1 INDICATE

To describe the spatial distribution of the stellar cluster population we use the 2 + D INDEX to Define Inherent Clustering And Tendencies (INDICATE) tool introduced by A. S. M. Buckner et al. (2019). INDICATE belongs to a class of statistical descriptors known as Local Indicators of Spatial Association (LISA; L. Anselin 1995), which are defined as characterizing discrete distributions by calculating a parameter value for every unique point, rather than a single parameter for the collective whole. Hereafter, we use 'association' strictly in the statistical sense of *spatial association*, and not to denote OB-type stellar groupings such as the Scorpius–Centaurus association.

INDICATE quantifies the degree of association of each and every cluster in our catalogue. No a priori knowledge is required, nor assumptions made about, any sub-structure present in the data set. Extensive statistical testing by A. S. M. Buckner et al. (2019) showed it to be robust against outliers and edge effects, as well as independent of a data set's size, shape and density. A. S. M. Buckner et al. (2022b) demonstrated INDICATE is robust against

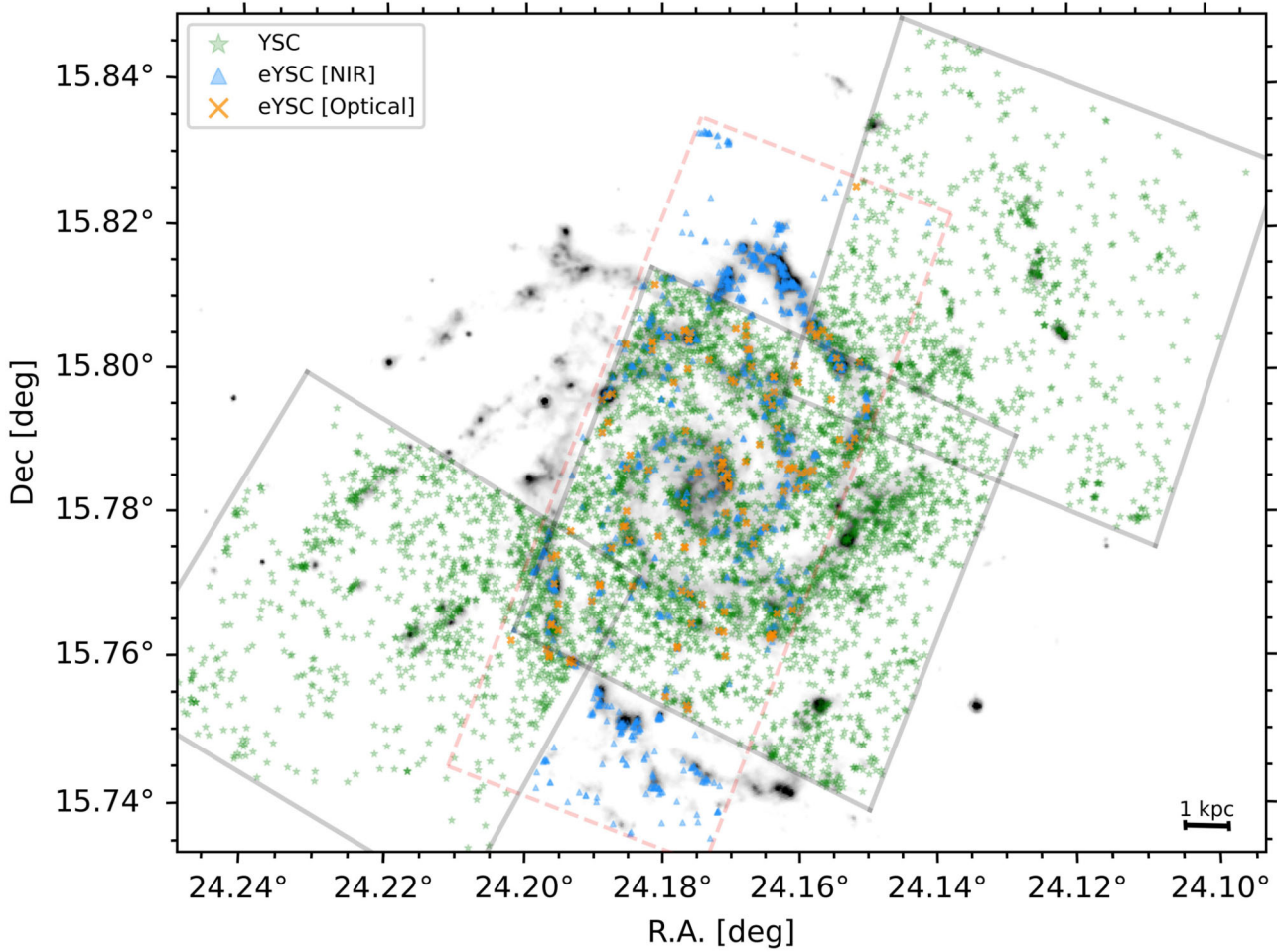


Figure 1. Final cluster sample as described in Section 2.2, overlaid on the *Spitzer* Infrared Nearby Galaxies Survey $8\ \mu\text{m}$ image of NGC 628. Coloured markers denote which catalogue the cluster originates from, as specified by the legend. The grey and red-dashed panels denote the position of the *HST* and *JWST* images, respectively.

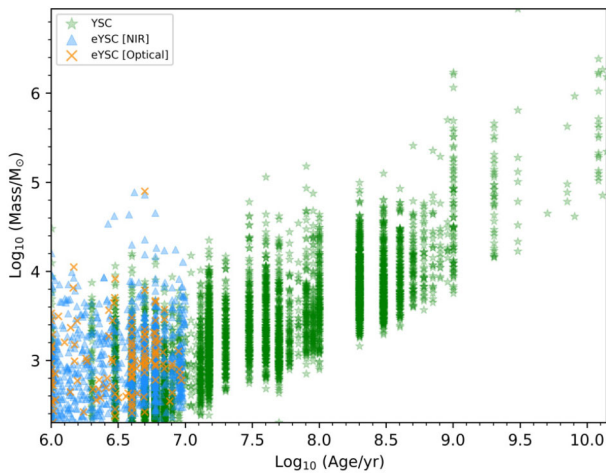


Figure 2. Plot of age against mass for our cluster catalogue. Green stars represent optical YSCs, orange crosses eYSCs that are optically visible, and blue triangles eYSCs that are not optically visible. See Section 2.1 for details.

incompleteness up to 83.3 per cent, whereas A. S. M. Buckner et al. (2024) showed that conclusions about the spatial properties of a data set are unchanged even when there are few true members (< 2 per cent) or significant (> 20 per cent) contamination by non-members that have been mistaken as true members of the data set. INDICATE can be applied to any discrete astronomical data set, observed or simulated, on any scale, and the results obtained for data sets of different densities, size, shape, and parameter space can be directly compared. Previous applications include describing spatial behaviours of: observed Young Stellar Objects as a function of evolutionary stage in NGC 2264 (A. S. M. Buckner et al. 2020); stars formed from a cloud–cloud collision simulation over a period of 1 Myr (A. S. M. Buckner et al. 2022a); the observed stellar population of Orion and their potential evolution over the next 1 Myr (simulated; A. Sills et al. 2023); mass segregation across various individual star clusters and forming complexes (A. S. M. Buckner et al. 2019; G. A. Blaylock-Squibbs et al. 2022); differences in the degree of spatial clustering of the stellar component of the bar, inner spiral arm, outer arm, and inter-arm region of a simulated Milky-Way type galaxy (A. A. Ali et al. 2023).

Here we use INDICATE to examine how the spatial behaviour of NGC 628’s clusters varies with age, mass and galactic position. The tool works by examining the local neighbourhood of a stellar cluster,

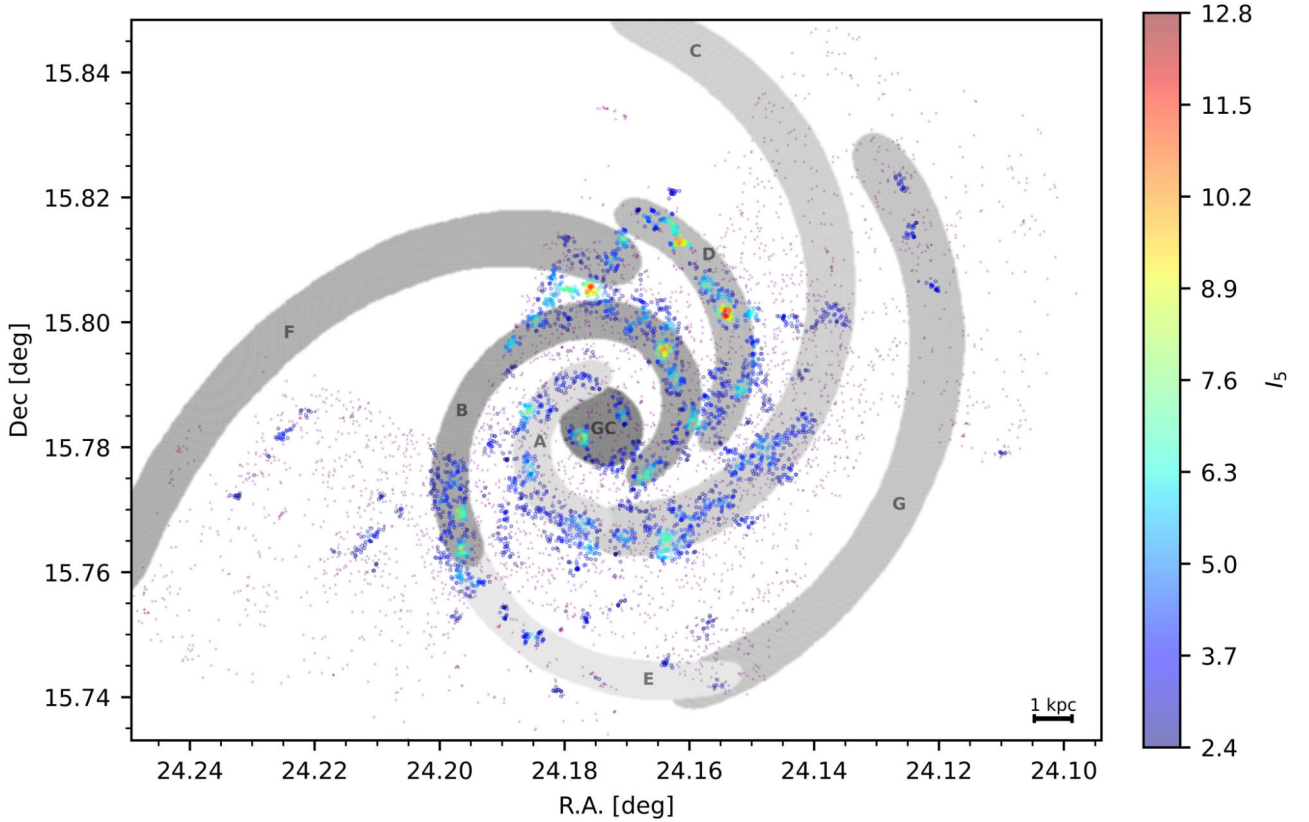


Figure 3. Plots of index values, I_5 , calculated by INDICATE for NGC 628, with the environmental mask by M. Querejeta et al. (2021) overlaid showing the positions of Spiral Arms A-G (marked by various shades of grey), and Galactic Centre (dark grey circle, ‘GC’). Clusters with an index value above the significance threshold ($I_5 > 2.4$) are represented by circle markers, colour-coded by their I_5 index, where higher values denote greater degrees of spatial concentration. Purple triangles are clusters with an index value below the significance threshold.

comparing the number of clusters found there to that expected if it was in a random distribution, and assigning an index to denote whether the cluster is spatially concentrated – and if so, quantifies the degree of concentration.

The full details on INDICATE can be found in A. S. M. Buckner et al. (2019), but we provide here a summary of how the technique works in practice. The first step is to convert the catalogue’s J2000 sky coordinates to a local X–Y coordinate frame to ensure all dimensions have the same scaling. The number density of the catalogue is then calculated by dividing the total number of stellar clusters, N_{tot} , by the total 2D surface area of the parameter space. Next, an evenly spaced uniform (i.e. definitively non-clustered) control distribution with the same number density across the parameter space is generated. For each stellar cluster j the Euclidean distance, r_j , to its 5th nearest neighbour in the control distribution is measured:

$$r_j = \sqrt{(x_j - x_5^{\text{con}})^2 + (y_j - y_5^{\text{con}})^2}, \quad (1)$$

where (x_j, y_j) and $(x_5^{\text{con}}, y_5^{\text{con}})$ are the respective coordinates of the stellar cluster and 5th nearest neighbour in the control distribution. The average for the catalogue, \bar{r} , is calculated as

$$\bar{r} = \frac{\sum_{j=1}^{N_{\text{tot}}} r_j}{N_{\text{tot}}}. \quad (2)$$

Finally, the actual number of stellar clusters within a radius \bar{r} to stellar cluster j , $N_{\bar{r}}$, is then counted and the index, $I_{j,5}$, calculated

$$I_{j,5} = \frac{N_{\bar{r}}}{5}, \quad (3)$$

where $I_{j,5}$ is unit-less and has a value in the range of $0 \leq I_{j,5} \leq \frac{N_{\text{tot}}-1}{5}$. Significant values are identified by calibrating the index with 100 random distributions of size N_{tot} in the parameter space of the catalogue to determine the ‘significance threshold’, I_{sig} , defined as

$$I_{\text{sig}} = \bar{I}_5^{\text{random}} + 3\sigma, \quad (4)$$

where $\bar{I}_5^{\text{random}}$ is the mean index derived for the points of the random distribution and σ the standard deviation. Using this definition, 99.7 percent of the index values derived for randomly distributed points will have $I_5^{\text{random}} < I_{\text{sig}}$. Values greater than the significance threshold denote spatial clustering, and for these, higher values denote greater degrees of spatial association. It is important to note that the index is not a descriptor of density, such that values derived for data sets of disparate densities can be directly compared, and will be unchanged if the spatial distributions are identical (see section 2.1.4. of A. S. M. Buckner et al. 2019 for a discussion). The minimum sample size for INDICATE is 50 data points (A. S. M. Buckner et al. 2019). The use of a uniform control field (rather than one where we follow an underlying non-uniform distribution) is adequate for our purpose, as we are exploring how the spatial association of the stellar component vary across the galactic disc. In other words, we aim to assess clusters’ non-random spatial relationships, then directly compare, quantify, and identify small-

scale local variations in these. By changing the control field to a non-random distribution (for example, one that follows the underlying stellar distribution of NGC 628) removes our ability to identify random spatial relationships and thus quantify and compare clusters' spatial associations across the galaxy.

3.2 Accounting for the observational age–mass bias

As shown in Fig. 2, there is an age–mass bias in our catalogue which is well known in the literature and is due to the limited depth of the data combined with the fading of star cluster light due to stellar evolution (e.g. A. Adamo et al. 2017, among many others). This means that for clusters older than 10 Myr, the minimum mass that we are able to detect increases steadily with age, producing a lower limit with a rising slope as a function of age. We note that no artificial spreading was applied to the figure at <10 Myr: the vertical banding for the optical YSCs arises directly from the Yggdrasil spectral evolutionary models, whereas the eYSC ages were derived using CIGALE (see Sections 2.1.1 and 2.1.2 for details), and the logarithmic age axis can additionally enhance the visual prominence of such discrete loci.

At all ages, we also see that the number of massive clusters is steadily decreasing due to the shape of the cluster mass function: a power law with slope -2 , implies that it is 10 times more likely to sample a cluster of $10^3 M_\odot$ than one at $10^4 M_\odot$ and so on.

This, combined with the age binning of the sample, produces a visual increasing upper mass limit with age: since the x -axis on Fig. 2 is in logarithmic scale, the linear-age interval increases to the right in the diagram. Therefore even if NGC 628's star/cluster formation rate remains constant with time, the total number of clusters that form in each equal interval of $\log(\text{Age})$ increases. Thus, the likelihood of a very high mass cluster being within a given $\log(\text{Age})$ -bin increases (see e.g. D. A. Hunter et al. 2003).

We note that eYSCs densely populate the age bins <10 Myr. This population of recently formed (and still partially embedded) star clusters has largely been missed by HST surveys due to the lack of IR coverage, which is less affected by dust attenuation. Because of the clustering nature of star formation, eYSCs naturally include associations that are not gravitationally bound and dissolve on time-scales of 10 Myr. For example, we know from observational studies of our own Galaxy between 93 – 96 per cent of eYSCs will not survive the emergence phase to become bound star clusters (C. J. Lada & E. A. Lada 2003; Q. E. Goddard, N. Bastian & R. C. Kennicutt 2010), and with a value of ~ 85 –90 per cent found for the nearby spiral galaxy NGC 6946 (Q. E. Goddard et al. 2010), it is not unreasonable to anticipate a value similar to these for NGC 628. By comparing the number of eYSCs and those that are classified as likely bound in the optical YSCs catalogues (e.g. class 1 and 2), Adamo et al. (in preparation) find that only ~ 20 per cent of eYSCs are expected to survive within 10 Myr in NGC 628. That is, it is expected that the eYSC population would be more densely populated than the recently emerged YSC population. Moreover, only higher mass YSCs are expected to survive to ages of 100 Myr + , due to mass-loss from internal and external disruption events, so we would not expect to see a large low-mass population in older age bins (S. G. Boutloukos & H. J. G. L. M. Lamers 2003; M. Gieles 2010; M. R. Krumholz et al. 2019a).

Despite these issues, INDICATE is, by design, robust against both uniform and patchy incompleteness in data sets. Extensive testing has shown that when it is applied to the full and incomplete versions of a data set, only small variations between the index values occurs (for an incompleteness of up to 83.3 per cent) and there is no change in the conclusions regarding the spatial behaviours of a data set (A.

Table 1. Table detailing for each region: total number of clusters present, the percentage of clusters found to be spatially concentrated (i.e. $I_5 > I_{\text{sig}}$) and their median index value (\tilde{I}_5^C).

Region	Total clusters	% Concentrated	\tilde{I}_5^C
All	6890	51.8	3.8
GC	206	74.3	3.8
Inter Arms	3224	30.6	3.2
Spiral Arms	3460	70.3	4.0
SA-A	409	74.8	3.8
SA-B	932	81.9	4.4
SA-C	868	63.8	3.8
SA-D	599	83.3	4.8
SA-E	223	60.5	3.6

S. M. Buckner et al. 2022b, 2024). As a check, we compared the index derived for the full sample and with various mass and age cuts, which are presented in Appendix A1 concluding that results do not change significantly when cuts are applied. Therefore, for the purpose of this study, we do not need to apply age or mass cuts to the cluster sample to compensate for the age-mass completeness bias. One exception is that, for the study of the spatial behaviours for different mass clusters (Section 4.3), we apply a mass-age cut in order to minimize the potential propagation of an evolutionary trend on to the mass trend.

4 SPATIAL BEHAVIOUR OF CLUSTERS IN NGC 628

4.1 Galactic regions

INDICATE was applied to the combined sample of stellar clusters, and index values were derived for each of the 6890 clusters, as shown in Fig. 3. Table 1 gives statistics on the index values derived for each region of NGC 628. The significance threshold was determined using equation (4) to be $I_{\text{sig}} = 2.4$.

Only 51.8 per cent of the clusters in the galaxy are spatially concentrated, with a median value of $\tilde{I}_5^C = 3.8$. The proportion of the population in each region which is spatially dispersed varies, with the lowest proportion in the spiral arms (29.7 per cent) and galactic centre (25.7 per cent), and the largest proportion in the interarm regions (69.4 per cent). Interestingly, the degree of association of the concentrated population also varies by region, with the spiral arms showing the greatest (median index $\tilde{I}_5^C = 4.0$), followed by the galactic centre ($\tilde{I}_5^C = 3.8$), and the interarm regions typically demonstrating the lowest ($\tilde{I}_5^C = 3.2$).

Fig. 3 shows that all spiral arms host strongly spatially concentrated cluster populations, with typical INDICATE index values above the galaxy-wide average. However, the clusters' spatial behaviour varies both between the spiral arms and along the length of each arm. Arm D hosts both the greatest proportion of spatially concentrated clusters (83.3 per cent) and typical degree of association ($\tilde{I}_5^C = 4.8$). Two areas were found to have particularly high degrees of association, with index values here being the greatest found in the galaxy with a maximum value of $I_5 = 12.8$ (i.e. in the local neighbourhood of the clusters there are 5.3 times the number of clusters above random expectation, and 3.4 times more than typical for the non-dispersed cluster population in NGC 628). Arm B hosts the second greatest proportion of spatially concentrated clusters (81.9 per cent) and typical degree of association ($\tilde{I}_5^C = 4.4$). Like in D, there are a few areas of high spatial concentration (albeit slightly lower intensity, max. $I_5 = 10.4$), with a high-intensity area

found in between the designated bounds of B and F. The remaining spiral arm segments (for which we have coverage) contain a few areas of higher degrees of association of similar intensity, but the typical association here is comparable to that found for the general concentrated population in NGC 628 and the proportion of spatially concentrated clusters is significantly lower than that found in D and B (60.5–74.8 per cent).

To assess whether these differences are statistically significant, we applied two-sample Kolmogorov–Smirnov (K-S) tests, which evaluate whether two samples are drawn from the same parent distribution and therefore test for differences in their overall distributions (significance threshold $p < 0.05$). We found that clusters in the spiral arms and galactic centre have a similar underlying spatial distribution ($p > 0.05$), but clusters in the interarm region have significantly different spatial behaviour ($p < 0.05$). Clusters in spiral arms B and D have distinct clustering tendencies ($p < 0.05$), but A, C and E have similar clustering tendencies ($p > 0.05$). This result confirms that the spatial behaviour of clusters varies with galactic position. We explore the possibility of variations with evolutionary stage and mass in the next sections.

4.2 Evolutionary Stage

4.2.1 Emerging YSCs versus optical YSCs (≤ 10 Myr)

Here we explore the differences in the spatial distributions of the emerging young clusters still (partially) embedded in their natal clouds (eYSCs) and the non-embedded optical YSCs (≤ 10 Myr) which we will refer to in this section as recently emerged clusters (RECs). We define RECs as clusters that are ≤ 10 Myr and not in the eYSC catalogue. For consistency in cross-sample comparisons, we cap the RECs at a maximum age of 10 Myr when comparing to eYSCs, matching the eYSC catalogue’s upper age limit. Given that both subsamples span the same ages, differences on their relative spatial distribution could be indicative of the spatial evolution of the clusters as they emerge from their natal clouds. Unfortunately, as the eYSCs were identified from the *JWST* image, and the RECs from the *HST* images, there is only a relatively small overlap area containing both samples as shown on Fig. 4. Therefore, we restrict our comparison to clusters inside this area (as, for example, it is unknown whether RECs have similarly high degrees of association as found for the eYSCs towards the northern end of Arm D). This overlap region contains 750 eYSCs and 785 optical RECs. Their respective age and mass distributions are shown in Fig. A3.

We find that the proportion of eYSCs that are concentrated (42.9 per cent) is marginally higher than that of the REC population (38.6 per cent), and that eYSCs also are typically a factor of ~ 1.3 times more tightly associated with each other than their emerged counterparts. From the left panels of Fig. 4 the galactic distribution of the two concentrated populations share some commonality, as well as some differences in the specific areas of high and low degrees of association which vary across the disc. Plotting a histogram of the indices (right panels, Fig. 4), both distributions have a similar range on index values, especially within the overlap region, but the shape of the distributions is different (with one smoothly decreasing, while the other presents strong troughs and peaks). To determine the significance of these disparities, we compare the cumulative distribution functions of the index values of (i) all and (ii) only above the significance threshold, for all clusters in each group inside the overlap area using a 2 sample K-S test, using a significance boundary of $p < 0.05$. Finding $p < 0.05$ in both tests, we conclude eYSCs and RECs (≤ 10 Myr) do not have the same underlying spatial

distribution. This suggests that there is a distinct change in the spatial behaviour of clusters as they emerge from the natal clouds, with them typically becoming significantly less tightly associated than when they are still (partially) embedded. This could potentially be due to clusters moving away from their birth place as they clear their natal cloud (and therefore moving away from other nearby young clusters), or even due to the fact that some clusters may not survive once the natal cloud has been cleared (as they lose the potential well from the gas), thus dispersing their stellar population and ‘disappearing’ from the population of recently emerged clusters. Such rapid early dissolution (*‘infant mortality’*) and its environmental dependence have been both theoretically predicted and observationally observed in other system (e.g. B. G. Elmegreen & D. A. Hunter 2010; N. Bastian et al. 2012; M. R. Krumholz et al. 2019b; S. T. Linden et al. 2022).

4.2.2 Age

In order to investigate the spatial behaviour of stellar clusters as a function of age (and including before and after the emerging phase), we consider our eYSC sample as per Section 4.2.1, and we divide the optical (emerged) clusters of our sample into three groups: Young (< 21 Myr), Intermediate (21–100 Myr), and Old (> 100 Myr). The upper age limit of the Young sample was chosen to include all potentially unbound clusters with pre-main sequence stars that have not yet dissolved (T. Santos-Silva & J. Gregorio-Hetem 2012). INDICATE was run on each of these subsamples to ascertain the tendencies of clusters at a similar evolutionary stage to spatially associate with each other, and the strength of these associations.

Fig. 5 shows the distribution of the index values derived for each sample across the galactic disc, and Table 2 contains a summary of the statistics. At all evolutionary stages the concentrated population is primarily found across the spiral arms, but there are marked differences in their locations. As expected from the analysis of the distribution of RECs in Section 4.2.1, the Young optical clusters (which here include up to 20 Myr-old clusters) display areas of high degrees of association in Arms D, B, and B/F as indicated in the top panel of Fig. 5, which are somewhat similar to the distribution of eYSCs (albeit with different levels of concentration). However, this Young population also has higher degrees of association in A and C that were not seen in the eYSC population. Intermediate clusters have areas of higher degrees of association in Arm E, and D (but more southerly than the Young/eYSC populations). Old clusters have just one area of intense association in Arm C which was not observed in either the young or intermediate populations.

Regarding the proportion of spatially concentrated clusters within each age group, eYSCs have the greatest at 55.1 per cent across the entire sample. A steady decrease with increasing cluster age is then observed, culminating in only 35.6 per cent of Old clusters across the entire disc found to be spatially concentrated. The intensity of these concentrations is also greatest for the eYSCs at $\bar{I}_5^C = 4.8$. Young, Intermediate, and Old clusters are found to have similar intensities, at a factor of ~ 1.5 times less than that of the eYSCs. We check the significance of these findings by running 2 sample K-S tests on the index values for all clusters in each age group against those of the other age groups, with a significance boundary of $p < 0.05$. Subsequently, we reject the null hypotheses and verify that the exhibited spatial association behaviour of clusters at each evolutionary stage is distinct.

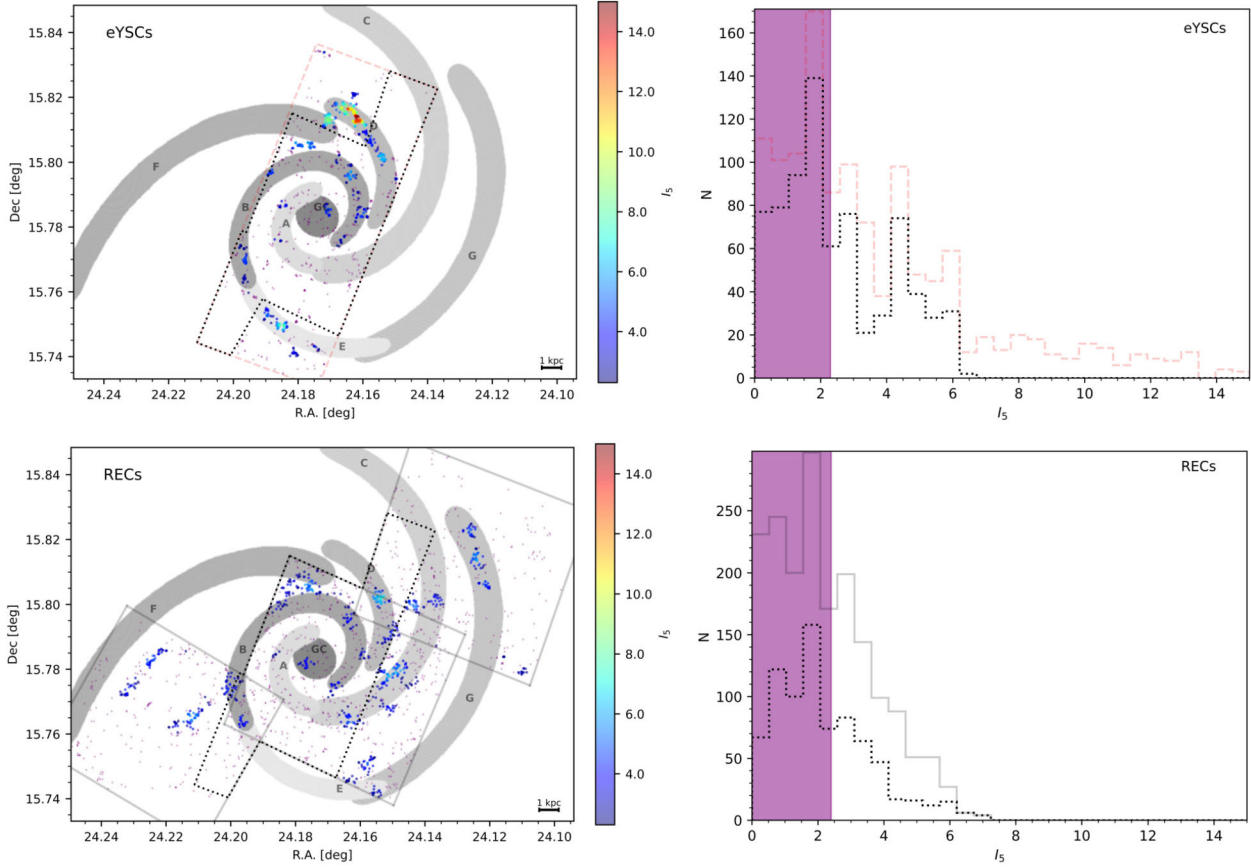


Figure 4. Index values, I_5 , plotted as a function of galactic position (left) and as a histogram (right), for the (top) eYSCs and (bottom) recently emerged clusters (RECs), defined as the optical YSCs < 10 Myr, discussed in Section 4.2.1. The index values represent the degree of association clusters have with their counterparts of similar age. On the left panels, clusters with an index value above their sample’s significance threshold are represented by circle markers, colour-coded by their I_5 value, as per the respective colour bars. Clusters with index values below the sample’s significance threshold are marked by purple triangles on the left panels, and shaded areas on the histograms. On the left panels, the environmental mask by M. Querejeta et al. (2021) is overlaid to show the location of Spiral Arms A–G (marked by various shades of grey), and Galactic Centre (dark grey circle, labelled ‘GC’). The grey, red-dashed, and black-dotted boxes denote the position of the *HST* mosaics, the *JWST* coverage, and their overlap region, respectively. Line colour and style in the histograms on the right correspond to that of clusters within the respective coverage and overlap areas, as denoted in the position plots.

4.2.3 Galactocentric distribution

Due to the aforementioned variations across the disc, we examined how the spatial behaviour of the emerged clusters varies with galactocentric distance, R_{GC} . We omit eYSCs from this part of the analysis as the difference in the shapes of the *JWST/HST* coverage area means we cannot be certain that any differences identified are attributable to real physical effects (i.e. evolutionary) or variations in the properties of regions that are not in both coverage areas (i.e. R_{GC} does not necessarily correspond to the same regions).

Fig. 6 shows the frequency distribution of clusters in each group as a function of galactocentric distance for (i) all and (ii) concentrated clusters. For the former, there is no discernible difference in the radial distribution of clusters across the age groups. For the concentrated clusters, the median radial distance remains constant but the spread decreases with increasing cluster age narrowing to predominantly the $R_{GC} \sim 2 - 6$ kpc range. Fig. 7 shows the variation in the proportion of clusters spatially concentrated and their index values, for each age group as a function of galactocentric distance. Between $R_{GC} \sim 2 - 6$ kpc a similar proportion and concentration strength is observed irrespective of cluster age. However, outside this narrow

range, variation is found with Young clusters exhibiting the greatest concentration behaviours.

The mechanism causing the lack of concentrated Intermediate/Old clusters at $R_{GC} > 6$ kpc is unclear. One possibility is that conditions in the outer disc favour the dissolution or decorrelation of spatial ‘structure’, perhaps in a similar process that leads to the disintegration of strong two-arm spiral structure often seen outside co-rotation (B. G. Elmegreen, D. M. Elmegreen & L. Montenegro 1992; D. M. Elmegreen & B. G. Elmegreen 1995). However, this would contrast with J. Tang et al. (2024), who found clusters inside the co-rotation radius ($R_{GC} = 6.3$ kpc; F. Sakhibov et al. 2021) to be more easily disrupted. It may instead be that young clusters in the outer galaxy arise from isolated, localized star-formation events, such that their initial groupings disperse rapidly as they move away from their natal regions. Alternatively, this pattern could reflect gradual inside-out growth: over the past $\sim 10 - 100$ Myr the outer disc may have only recently achieved the gas surface densities necessary to form the observed structure. In this scenario, older tight concentrations were simply never formed here, whereas the inner disc (< 6.3 kpc) has sustained star formation over many tens to hundreds of Myr, producing the Intermediate and Old cluster concentrations now truncated near the co-rotation radius. A similar inside-out trend

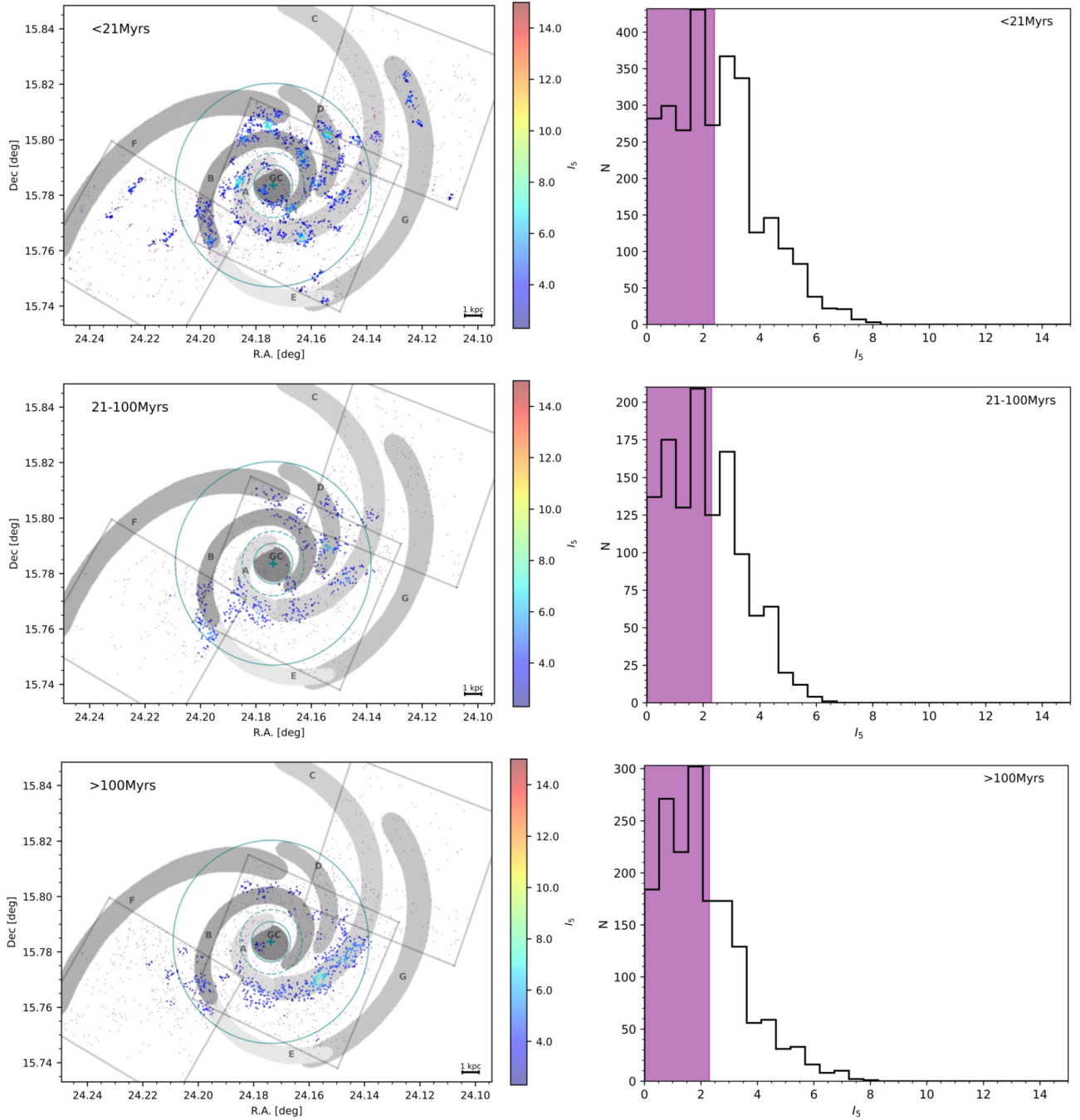


Figure 5. Index values, I_5 , for the Young (top), Intermediate (middle), and Old clusters (bottom), as discussed in Section 4.2.2. The index values represent the degree of association clusters have with their counterparts of similar age. Left panels show I_5 as a function of galactic position, and right panels as a histogram. The grey boxes on the left panel denote the position of the *HST* coverage, with the environmental mask by M. Querejeta et al. (2021) overlaid. The teal cross and ellipses mark the galactic centre, a radius of 1.25 kpc, 2 kpc and the galactic co-rotation radius (6.3 kpc; F. Sakhibov, A. S. Gusev & C. Hemmerich 2021), respectively. Clusters with an index value above their sample’s significance threshold are represented by circle markers, colour-coded by their index value. Clusters with index values below the sample’s significance threshold are marked by purple triangles and shaded on the position and histogram plots.

was reported by S. M. Gogarten et al. (2010) for NGC 300: using resolved-star formation histories from colour–magnitude diagrams and N -body simulations, they found an increasing fraction of young stars with radius and negligible impact from radial migration, thereby favouring a growth-driven rather than scattering-driven scenario. The similarity in both the proportion of clusters that are concentrated and the strength of those concentrations across the age groups between

2 and 6 kpc provides further evidence supporting a growth-driven origin for NGC 628.

For $R_{GC} < 2$ kpc, we find a more complex set of behaviours. For instance, from Fig. 7, we can see that while the fraction of concentrated Young clusters remains relatively high (at ~ 50 per cent), there is a steady decrease of the fraction of concentrated Intermediate clusters towards smaller R_{GC} . The Old clusters show yet a different

Table 2. Table detailing the spatial behaviour of clusters by evolutionary stage. Shows the total number of clusters, percentage found to be spatially concentrated with those of a similar age, and their respective median index values. Stages are as defined in Section 4.2.2. Bracketed values represent those within the *HST/JWST* overlap area shown in Fig. 4.

Stage	Age	Total clusters	% Concentrated	\bar{I}_5^C
eYSCs	≤ 10 Myr	1216 (750)	55.1 (42.9)	4.8 (4.4)
RECs	≤ 10 Myr	1813 (785)	41.5 (38.6)	3.4 (3.4)
Young	< 21 Myr	2805	49.4	3.4
Intermediate	21–100 Myr	1201	41.2	3.2
Old	> 100 Myr	1668	35.6	3.4

behaviour, with a reasonably high fraction of concentrated clusters for $R_{GC} < 1$ kpc, but a clear dip between $\sim 1 - 2$ kpc. We verified that the double-peaked radial profile of the old cluster population persists when restricting the sample to clusters aged between 100 Myr and 1 Gyr. This check was performed to test whether the inner peak could be driven by contamination from globular clusters (ages $\gtrsim 1$ Gyr). Although removing these objects markedly lowers the amplitude of the inner peak, the bimodal structure remains, confirming that the trend is intrinsic to the population of clusters younger than 1 Gyr.

The intensity of the spatial concentrations also decreases substantially for the older age groups, contrary to that found generally across the disc (Section 4.2.2). More specifically, the Old and Intermediate population, when concentrated, are only very mildly so, with I_5 indices very close to the threshold, thus reducing the significance of the large fraction of Old concentrated clusters at $R_{GC} \sim 1$ kpc. For the Young clusters, however, there are significant peaks in the concentration indices for $R_{GC} \sim 0.5 - 1$ kpc and $R_{GC} \sim 2$ kpc.

Our interpretation for these observed trends, lies in a comparison to some of the known dynamical features of this inner region. In particular, as shown in Fig. 5, a radius of ~ 1.25 kpc corresponds to the transition between the galactic centre and the beginning of the spiral arms. Just below this radii there is an inner Lindblad resonance, thought to be linked to the origin of the nuclear ring at

$\sim 0.5 - 1$ kpc. The location of the first peak of concentrated Young clusters at $\sim 0.5 - 1$ kpc, would thus suggest this population of Young concentrated clusters being formed in the nuclear ring, but that the highly dynamical galactic centre quickly disperses these clusters as they evolve. Just outside of the inner Lindblad resonance ($R_{GC} \simeq 1.25$ kpc) is an area which has been previously observed to have no H II regions, with a velocity drop around the galaxy’s nuclear ring and inflowing (colder) gas (K. Fathi et al. 2007). Thus, it is reasonable to expect very little cluster structure to be formed here, and this is consistent with the dip in the concentration indices of the Young population at these radii. Moving outwards, spiral arms (which are regions of reduced shear) start to provide again the conditions to form tightly concentrated clusters, and we see this effect with a higher concentration index for Young clusters formed in the tips of the spiral arms. However, perhaps due to the harsh environment outside the spiral arm tips in this region (e.g. from strong shear and strong inward motions from the Lindblad resonance), any cluster structure would be prone to a very rapid dissolution as it emerges from the arms, so the overall proportion (and concentration index) of the Intermediate and Old cluster population at < 2 kpc radius remains lower than across the galactic disc.

4.3 Mass

To quantify how the clusters’ spatial behaviour might change as a function of their mass, we compare the distribution of clusters with high ($\geq 10^4 M_\odot$) and lower ($< 10^4 M_\odot$). Two quite different realizations of mass segregation are explored here. The classic ‘Type 1’ definition examines if massive clusters are spatially concentrated with other massive clusters, significantly more so than seen for their lower mass counterparts. Another definition, ‘Type 2’, looks at whether massive clusters are in concentrated regions, but are not necessarily concentrated with other massive clusters. Due to the age-mass bias (Section 3.2) it is incorrect to simply divide the catalogue into ‘high’ and ‘lower’ mass, as lower mass clusters will inherently represent a younger population and high mass clusters an older population. Instead, we apply an age-mass cut to the catalogue to remove this bias. The new sample consists of clusters with an age of ≤ 100 Myr and a mass larger than the minimum mass of clusters

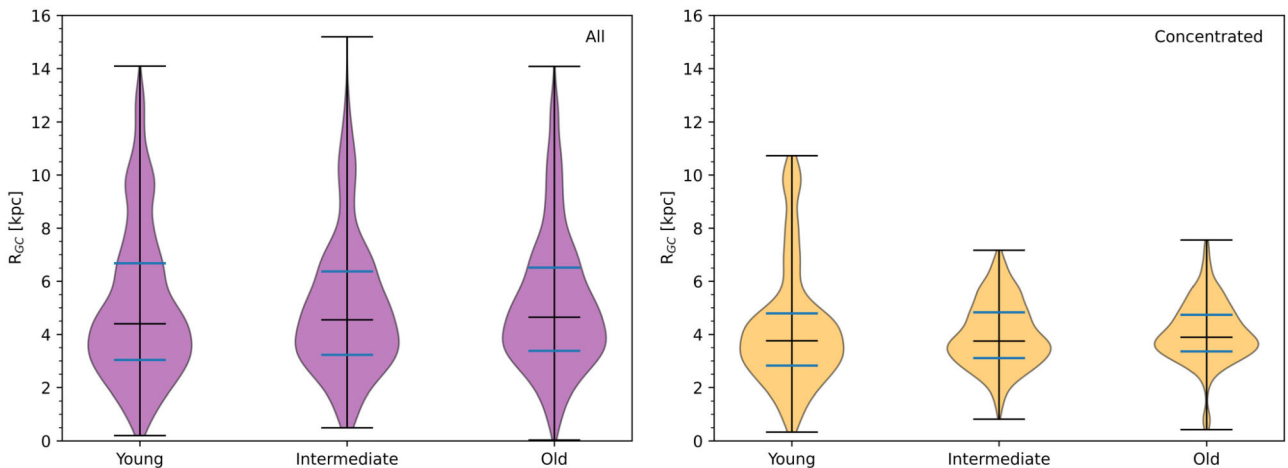


Figure 6. Violin plots depicting the one-dimensional radial distributions of optical clusters for (left) all and (right) spatially concentrated of the Young, Intermediate and Old groups, as discussed in Section 4.2.2. The vertical extent of each violin corresponds to galactocentric radius, R_{GC} , while the horizontal width represents the relative number density (probability density) of clusters at a given radius. Spatially concentrated clusters are defined as having an index value above their sample’s significance threshold i.e. a degree of association above random expectation with their counterparts of a similar age. Black lines represent the minimum, median, and maximum of the distribution. The blue and black horizontal lines represent the 25th, 50th, and 75th percentiles.

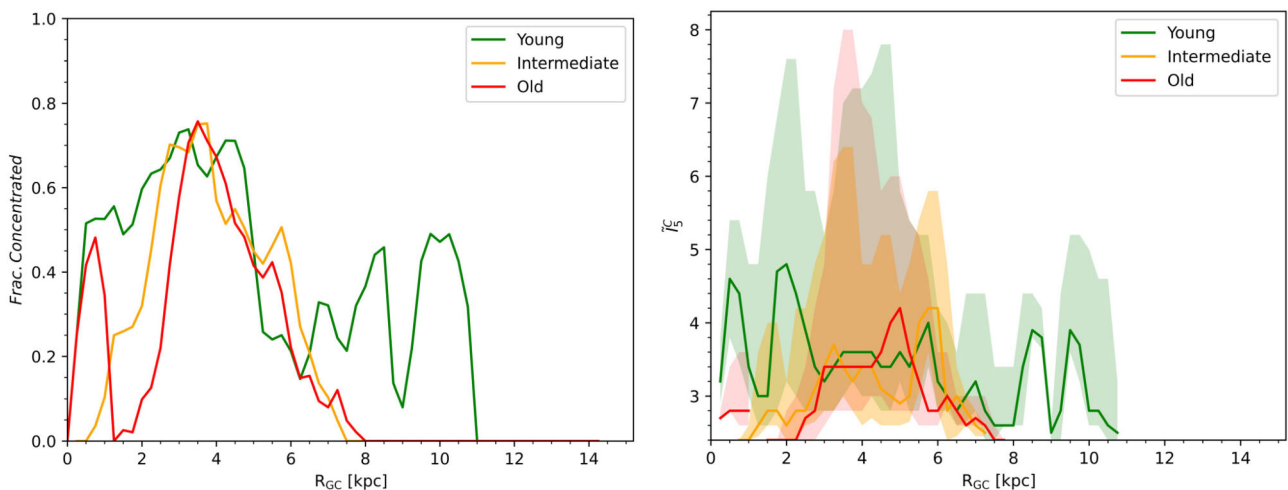


Figure 7. Spatial clustering behaviour of Young, Intermediate, and Old clusters as a function of galactocentric distance. The left panel shows the fraction of clusters found to be spatially concentrated with their counterparts of a similar age as discussed in Section 4.2.2; and the right panel shows the running 25th, 50th, and 100th percentile of the I_5 index values of the spatially concentrated clusters.

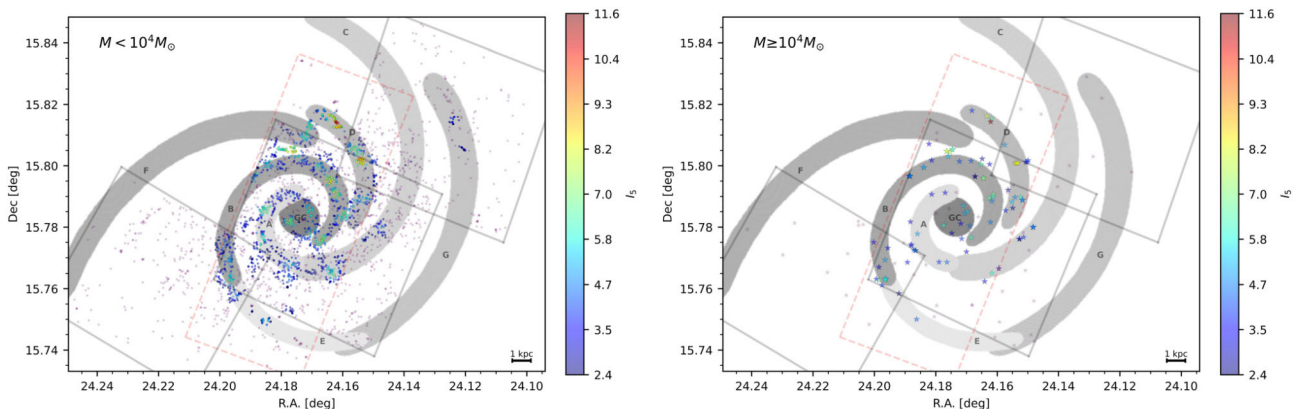


Figure 8. Plots of index values, I_5 , calculated by INDICATE for (left) lower mass and (right): YMCs, to investigate if Type 2 mass segregation is present as described in Section 4.3. As in previous figures, the grey and red-dashed panels denote the position of the *HST* and *JWST* images, respectively. Clusters with an index value above the significance threshold ($I_5 > 2.4$) are colour-coded by their I_5 index, where higher values denote greater degrees of spatial concentration. Purple markers are clusters with an index value below the significance threshold.

> 100 Myr ($\sim 790 M_\odot$), visualized in the right panel of Fig. A1. A total of 3087 clusters meet this criteria, of which 137 are high mass (thus corresponding to YMCs).

To determine if Type 1 mass segregation is present, we divide the sample into two subsamples (‘high’ and ‘lower’ mass), and run INDICATE on both separately. If mass segregation is present, we would expect the high-mass sample (YMCs) to exhibit significantly greater tendencies to spatially associate than the lower mass sample. No signatures were found: YMCs do not have a tendency to concentrate together in NGC 628. In fact, the I_5 values suggest the opposite – massive clusters tend to be more spatially dispersed w.r.t. each other than lower mass clusters. A 2 sample K-S test confirms that while Type 1 mass segregation is not present, YMCs and lower mass clusters do have different spatial behaviour ($p < 0.05$). Moreover, we then examined the behaviour of the Old high mass clusters (> 100 Myr) finding that while there is no significant change in the tightness of the concentrations, the proportion of the population that are concentrated decreases by 17.5 per cent.

To determine if Type 2 mass segregation is present, INDICATE is run on the full sample (within our age-mass cut). Again, if

present, YMCs should exhibit significantly greater tendencies to spatially associate than the lower mass clusters. We find that a greater proportion of massive clusters (67.2 per cent) are found in concentrated regions compared to lower mass clusters (57.7 per cent). As shown in Fig. 8, the degree of association is typically higher for YMCs compared to their lower mass counterparts (2-3 clusters more in their local neighbourhood) because they tend to be located in regions where the lower mass population exhibits greatest spatial association. Moreover, massive clusters that have signatures of Type 2 mass segregation are almost exclusively located within the spiral arms and galactic centre – whereas those in the interarm regions typically do not. This result suggests most of the massive clusters are forming in the high-density, shock-compressed spiral arms, in regions of intense star formation activity. Occasionally, an exceptionally massive GMC in the inter-arm region will collapse and form a massive cluster but, as overall star formation intensity is low there, the cluster has very few neighbours nearby and relatively isolated (no Type 2 signal). A 2 sample K-S test comparing the index values of the two populations confirms the found signatures of Type 2 mass segregation are real ($p < 0.05$).

5 CONCLUSIONS

We have characterized the spatial distributions of star clusters in the nearby face-on spiral galaxy NGC 628. The aim of our study was to quantify the impact galactic environment has on the formation, early and longer-term evolution of stellar clusters, and YMCs in particular. This was achieved through the analysis of a complete sample, from FEAST collaboration, of the cluster population consisting of 6890 candidates from the (partially) embedded phase to > 100 Myr.

The powerful novel statistical clustering tool INDICATE by A. S. M. Buckner et al. (2019) was used to characterize the spatial behaviour of the clusters. INDICATE is a local indicator of spatial association, meaning that it quantifies the clustering tendencies of each cluster individually. As such, it is possible to not only study general variations with cluster properties (age, mass) in the galaxy, but also to compare the impact disc position has between clusters with similar ages and masses.

We found that the spatial behaviour of clusters in NGC 628 varies with galactic position, evolutionary stage and mass. The majority of young clusters ≤ 10 Myr, which are still embedded within their natal clouds, are spatially concentrated. Typically, they tend to be in the tightest spatial associations, located in multiple ‘hot spots’ across the spiral arms, with the most intense concentrations occurring in arm D. These locations correlate well with regions that have the highest molecular star formation efficiency¹ as found by M. Lomaeva, A. Saintonge & I. De Looze (2024), suggesting multiple clusters are forming in these tight ‘structures’ and that the spatial behaviour we identify is a star formation signature rather than indicative of early dynamical evolution. Once emerged, clusters of a similar age do still tend to concentrate together within the spiral arms, but in looser spatial associations than observed for the embedded population. We speculate that this could be due to clusters moving away from their birth place as their natal cloud is cleared, and/or due to the fact that some clusters may become dispersed once they lose the gravitational potential from the gas of their natal cloud. While the median degree of association remains approximately constant with increasing cluster age after the clusters are fully emerged, the proportion of clusters which are spatially concentrated decreases, reaching a minimum at ages > 100 Myr. This result is in good agreement with K. Grasha et al. (2015) who, by applying the 2-point correlation function to an earlier (and less complete) version of our cluster sample, found that the strength of the clustering decreases as age increases and that clusters > 40 Myr tend to be randomly distributed.

The radial distribution of all emerged clusters across the disc is similar irrespective of cluster age, but that of the concentrated population progressively narrows to within a radius of $2 - 6$ kpc of the galactic centre with increasing age. Within this radius a similar proportion of clusters that are concentrated, and the strength of this concentration, is found across all age groups. This suggests either tight cluster spatial ‘structure’ in the outer disc (1) is more easily disrupted than the inner disc (not supported by previous studies); (2) is formed in isolated, localized star-formation events, such that their initial groupings disperse rapidly; and (3) is a result of a gradual inside-out growth scenario for NGC 628, such that the disc outside the galactic co-rotation radius (> 6.3 kpc) has built up over the past $\sim 10 - 100$ Myr to gas surface densities capable of forming these structures, leading to the observed truncation beyond the co-rotation radius. We also found noticeable changes in the spatial behaviour of

clusters of different ages in the galactic centre, for radii < 2 kpc, which we interpret as being related to the location of an inner Lindblad resonance (at $\lesssim 1$ kpc), the nuclear ring (at $\sim 0.5 - 1$ kpc), and the start of the spiral arms (at $\sim 1.25 - 2$ kpc), and suggest that this area’s shear and particularly strong radial motions, make it hard not only for clusters to form in tight concentrations but even more so in remaining tightly concentrated as they evolve.

Finally, we looked for signatures of mass segregation in NGC 628. Massive clusters ($> 10^4 M_{\odot}$) were found to typically be spatially dispersed w.r.t. each other, in agreement with K. Grasha et al. (2015), and this was most pronounced at ages > 100 Myr. By using INDICATE, we were able to extend this analysis to examine the spatial relationship between massive cluster and their lower mass counterparts. We found that YMCs in the spiral arm and galactic centre regions tended to be associated with the most intense spatial concentrations of lower mass cluster population, whereas those in the interarm regions tended to be relatively isolated.

Our results suggest a ‘big picture’ scenario in NGC 628 where the process of hierarchical fragmentation and collapse of GMCs tend to produce a higher mass cluster and multiple lower mass clusters in relatively close proximity in regions that have a high star formation activity, especially in the inner galaxy (< 6.3 kpc). In regions of lower star formation activity, massive clusters will form in relative isolation, potentially due to the collapse of a somewhat rare high-density cloud. Once clusters have fully emerged, this structure is mostly retained but diluted, as clusters are dynamical dispersed from their birth place and/or disrupted due to insufficient stellar mass or stellar depletion. This may reflect either faster dispersal of isolated tight cluster spatial ‘structure’ in a lower gas density outer disc or gradual inside-out growth, with the formation of this structure shifting outward over time. A comprehensive comparison between the spatial distribution of young embedded clusters and that of the molecular gas is required to better understand these observations, and will be presented in a forthcoming companion paper.

ACKNOWLEDGEMENTS

We thank the anonymous referee for their insightful comments and helpful suggestions, which greatly improved this manuscript. ASMB and ADC acknowledge the support from the Royal Society University Research Fellowship URF/R1/191609 (PI: ADC). The calculations used in this work were performed using the supercomputing facilities at Cardiff University operated by Advanced Research Computing at Cardiff (ARCCA) on behalf of the Cardiff Supercomputing Facility and the HPC Wales and Supercomputing Wales (SCW) projects. This work is based in part on observations made with the NASA/ESA/CSA *JWST*, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5–03127. The data were obtained from the Mikulski Archive for Space Telescopes (MAST) at the Space Telescope Science Institute. These observations are associated with program #1783 (GO-01783). Support for program #1783 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5–03127. The specific observations analysed can be accessed via [doi:10.17909/f4vm-c771](https://doi.org/10.17909/f4vm-c771). Additional archival data was used from *JWST* program GO-2107 (PI J. Lee); and HST programs GO-9796 and GO-13773. AA and AP acknowledge support from the Swedish National Space Agency (SNSA) through the grant 2021–00108. AA and HFV acknowledge support from SNSA 2023–00260. KG is supported by the Australian Research Council through the Discovery Early Career Researcher

¹ defined as the ratio of the SFR from $H\alpha$ emission to molecular gas mass surface density, i.e. $\sum \text{SFR}(H\alpha) / \sum m_{\text{mol}}$

Award (DECRA) Fellowship (project number DE220100766) funded by the Australian Government. MM acknowledge financial support through INAF GO Grant 2024 “Mapping Star Cluster Feedback in a Galaxy 450 Myr after the Big Bang” and by the European Union – NextGenerationEU within PRIN 2022 project n.20229YBSAN. MK acknowledges support from the Australian Research Council through Laureate Fellowship FL220100020.

DATA AVAILABILITY

The INDICATE tool is available to download from GitHub at <https://github.com/abuckner89/INDICATE>. The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

- Adamo A. et al., 2017, *ApJ*, 841, 131
- Ali A. A., Dobbs C. L., Bending T. J. R., Buckner A. S. M., Pettitt A. R., 2023, *MNRAS*, 524, 555
- Aniyan S. et al., 2018, *MNRAS*, 476, 1909
- Anselin L., 1995, *Geogr. Anal.*, 27, 93
- Barnes A. T. et al., 2023, *ApJ*, 944, L22
- Bastian N. et al., 2012, *MNRAS*, 419, 2606
- Blaylock-Squibbs G. A., Parker R. J., Buckner A. S. M., Güdel M., 2022, *MNRAS*, 510, 2864
- Boquien M., Burgarella D., Roehlly Y., Buat V., Ciesla L., Corre D., Inoue A. K., Salas H., 2019, *A&A*, 622, A103
- Boutloukos S. G., Lamers H. J. G. L. M., 2003, *MNRAS*, 338, 717
- 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
- Buckner A. S. M., Naylor T., Dobbs C. L., Rieder S., Bending T. J. R., 2024, *MNRAS*, 527, 5448
- Calzetti D. et al., 2015, *AJ*, 149, 51
- Calzetti D. et al., 2024, *ApJ*, 971, 118
- Cardelli J. A., Clayton G. C., Mathis J. S., 1989, *ApJ*, 345, 245
- Cook D. O. et al., 2014, *MNRAS*, 445, 881
- Dale D. A. et al., 2023, *ApJ*, 944, L23
- F. Demachi, et al., Fukui Y., Yamada R. I., Tachihara K., Hayakawa T., Tokuda K., Fujita S., Kobayashi M. I. N. et al., 2023, in *ALMA at 10 years: Past, Present, and Future*, Puerto Varas, Chile
- Demachi F., et al., Fukui Y., Yamada R. I., Tachihara K., Hayakawa T., Tokuda K., Fujita S., Kobayashi M. I. N., et al., 2024, *PASJ*, 76, 1059
- Egorov O. V. et al., 2023, *ApJ*, 944, L16
- Elmegreen B. G., Elmegreen D. M., 2019, *ApJS*, 245, 14
- Elmegreen B. G., Hunter D. A., 2010, *ApJ*, 712, 604
- Elmegreen B. G., Elmegreen D. M., Montenegro L., 1992, *ApJS*, 79, 37
- Elmegreen B. G., Elmegreen D. M., Chandar R., Whitmore B., Regan M., 2006, *ApJ*, 644, 879
- Elmegreen D. M., Elmegreen B. G., 1995, *ApJ*, 445, 591
- Fathi K., Beckman J. E., Zurita A., Relaño M., Knapen J. H., Daigle O., Hernandez O., Carignan C., 2007, *A&A*, 466, 905
- Faustino Vieira H., Duarte-Cabral A., Smith M. W. L., Colombo D., Davis T. A., Bazzi Z., 2025, *MNRAS* 538 2744
- Gardner J. P. et al., 2006, *Space Sci. Rev.*, 123, 485
- Gardner J. P. et al., 2023, *PASP*, 135, 068001
- M. Gieles, , 2010, IAU Symp., 266 69,
- Goddard Q. E., Bastian N., Kennicutt R. C., 2010, *MNRAS*, 405, 857
- Gogarten S. M. et al., 2010, *ApJ*, 712, 858
- Grasha K. et al., 2015, *ApJ*, 815, 93
- Grasha K. et al., 2017, *ApJ*, 840, 113
- Gregg B. et al., 2024, *ApJ*, 971, 115
- Gusev A. S., 2014, *MNRAS*, 442, 3711
- Herrera C. N. et al., 2020, *A&A*, 634, A121
- Hunter D. A., Elmegreen B. G., Dupuy T. J., Mortonson M., 2003, *AJ*, 126, 1836
- Inoue S., Takagi T., Miyazaki A., Cooper E. M., Egusa F., Yajima H., 2021, *MNRAS*, 506, 84
- Kim J. et al., 2023, *ApJ*, 944, L20
- Kreckel K., Blanc G. A., Schinnerer E., Groves B., Adamo A., Hughes A., Meidt S., 2016, *ApJ*, 827, 103
- Kreckel K. et al., 2018, *ApJ*, 863, L21
- Kroupa P., 2001, *MNRAS*, 322, 231
- Krumholz M. R., McKee C. F., Bland-Hawthorn J., 2019a, *ARA&A*, 57, 227
- Krumholz M. R., McKee C. F., Bland-Hawthorn J., 2019b, *ARA&A*, 57, 227
- Lada C. J., Lada E. A., 2003, *ARA&A*, 41, 57
- Leitherer C. et al., 1999, *ApJS*, 123, 3
- Leitherer C., Ekström S., Meynet G., Schaerer D., Agienko K. B., Levesque E. M., 2014, *ApJS*, 212, 14
- Leroy A. K. et al., 2021, *ApJS*, 257, 43
- Linden S. T. et al., 2022, *ApJ*, 935, 166
- Lomaeva M., De Looze I., Saintonge A., Decleir M., 2022, *MNRAS*, 517, 3763
- Lomaeva M., Saintonge A., De Looze I., 2024, *MNRAS*, 531, 815
- Marchuk A. A., 2018, *MNRAS*, 476, 3591
- Matsumoto K., Hirashita H., Nagamine K., van der Giessen S., Romano L. E. C., Relaño M., De Looze I., Baes M., *A&A* 689, A79 et al., 2024,
- Mulcahy D. D., Beck R., Heald G. H., 2017, *A&A*, 600, A6
- Pedrini A. et al., 2024, *ApJ*, 971, 32
- Pedrini A. et al., 2025, *ApJ*, 992, 96
- Querejeta M. et al., 2021, *A&A*, 656, A133
- Rousseau-Nepton L., Robert C., Martin R. P., Drissen L., Martin T., 2018, *MNRAS*, 477, 4152
- Sakhibov F., Gusev A. S., Hemmerich C., 2021, *MNRAS*, 508, 912
- Santos-Silva T., Gregorio-Hetem J., 2012, *A&A*, 547, A107
- Shabani F. et al., 2018, *MNRAS*, 478, 3590
- Sills A., Rieder S., Buckner A. S. M., Hacar A., Portegies Zwart S., Teixeira P. S., 2023, *MNRAS*, 519, 4142
- Tang J., Grasha K., Krumholz M. R., 2024, *MNRAS*, 532, 4583
- Thieler D. A. et al., 2023, *ApJ*, 944, L13
- Tully R. B., Rizzi L., Shaya E. J., Courtois H. M., Makarov D. I., Jacobs B. A., 2009, *AJ*, 138, 323
- Ujjwal K., Kartha S. S., Subramanian S., George K., Thomas R., Mathew B., 2022, *MNRAS*, 516, 2171
- Watkins E. J. et al., 2023, *ApJ*, 944, L24
- Weżgowiec M., Beck R., Hanasz M., Soida M., Ehle M., Dettmar R. J., Urbanik M., 2022, *A&A*, 665, A64
- Zackrisson E., Rydberg C.-E., Schaerer D., Östlin G., Tuli M., 2011, *ApJ*, 740, 13
- Zhao M. et al., 2024, *ApJ*, 967, 18

APPENDIX A: AGE-MASS DISTRIBUTION

A1 Impact of age-mass bias on derived statistics

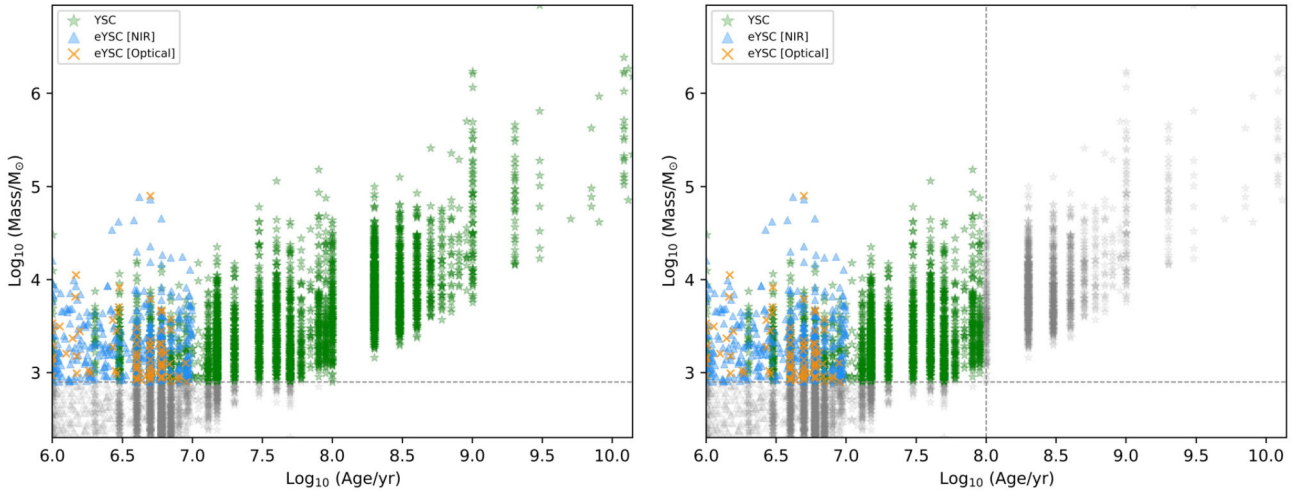
We tested the need to apply cuts to the cluster sample, owing to the age-mass bias present. INDICATE was run on three versions of the cluster sample:

- Full: No cuts were made.
- Mass restricted: Clusters with a mass less than the minimum mass of clusters that are > 100 Myr ($< 790.1 M_{\odot}$) were removed.
- Age restricted: An age limit of 100 Myr was applied to the mass restricted sample.

Figs A1 and A2 visualize the full and two restricted sample versions of the catalogue. Additionally, we created mass-restricted versions of our Young and Intermediate age samples (described in Section 4.2.2). Table A1 gives statistics on the index values derived for each sample.

Table A1. Table detailing the differences in the percentage of clusters found to be spatially concentrated and their median index value (\tilde{I}_5^C) for the full, mass-restricted and age-restricted samples.

Sample	Restriction	No. clusters	% Concentrated	\tilde{I}_5^C
Full	None	6890	51.8	3.8
	Mass-restricted	4755	51.8	3.8
	Age-restricted	3087	58.1	4.2
Young	None	2805	49.4	3.4
	Mass-restricted	1255	55.7	3.8
Intermediate	None	1201	41.2	3.2
	Mass-restricted	1143	44.3	3.2

**Figure A1.** Plot of age and mass distributions for our (left) mass-restricted and (right) age-restricted versions of our cluster catalogue as shown in Figs 2 and A2. Grey markers in the top panels represent clusters which were excluded under the restriction criteria.

In all samples, either no or small variations occurred in the median index values and/or % of spatially concentrated clusters. The conclusions drawn about the general spatial behaviour of the clusters are unchanged between the full and restricted versions of the sample.

In cases where a variation did occur, the strength of the spatial behaviour signatures increased rather than decreased. The greatest variations occurred in the age-restricted and Young samples. This was expected, as the most tightly spatially concentrated clusters tend to be young and thus the restrictions inherently impact the youngest clusters the most (increasing larger proportions of population were removed with decreasing cluster age). For the age-restricted sample (versus the respective full sample) this meant that the dispersed older population was removed, so stats represent the spatial behaviour of a younger population which tend to have greater spatial associations than their older counterparts (Section 4.2.2). For the Young samples, index values from the mass-restricted samples represent the spatial behaviour of a higher-mass population, which tend to

have greater spatial associations than their lower-mass counterparts (Section 4.3).

A2 Subsample statistics

We present diagnostic histograms of the age and mass distributions for the subsamples analysed in the main text. For each case, we show overlaid, probability–density histograms (unit area) of $\log_{10}(\text{age/yr})$ and $\log_{10}(M/M_{\odot})$ using identical binning to enable direct visual comparison. Specifically: (i) the eYSCs versus RECs comparison discussed in Section 4.2.1 (Fig. A3); (ii) the evolutionary subsamples (Young, Intermediate, Old) discussed in Section 4.2.2 (Fig. A4); and (iii) the mass-selected subsamples (‘high’ and ‘lower’ mass) discussed in Section 4.3 (Fig. A5). Legends identify the subsamples, and the y-axis shows probability density in all panels. For completeness, Table A2 summarizes the median $\log_{10}(\text{age/yr})$ and $\log_{10}(M/M_{\odot})$ (with 16th–84th percentile ranges) and sample sizes for all subsamples used in this paper.

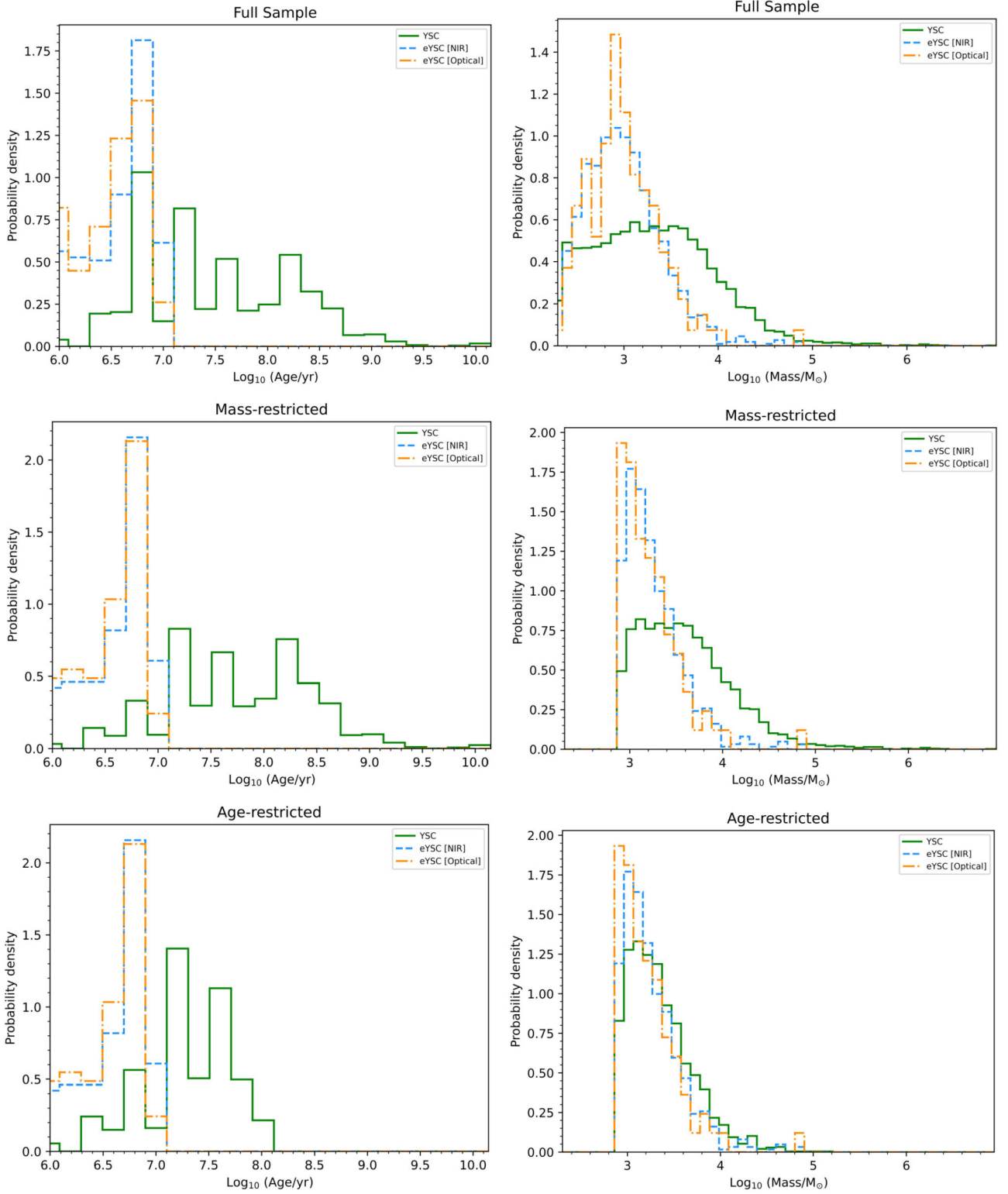


Figure A2. Histograms for the full, mass-restricted and age-restricted cluster samples showing (left:) age and (right) mass distributions. The histograms have been normalized as probability densities (such that the total area under the bars of each histogram equals one) to enable direct comparison of the age and mass distributions of the eYSC-NIR, eYSC-Optical, and YSC samples.

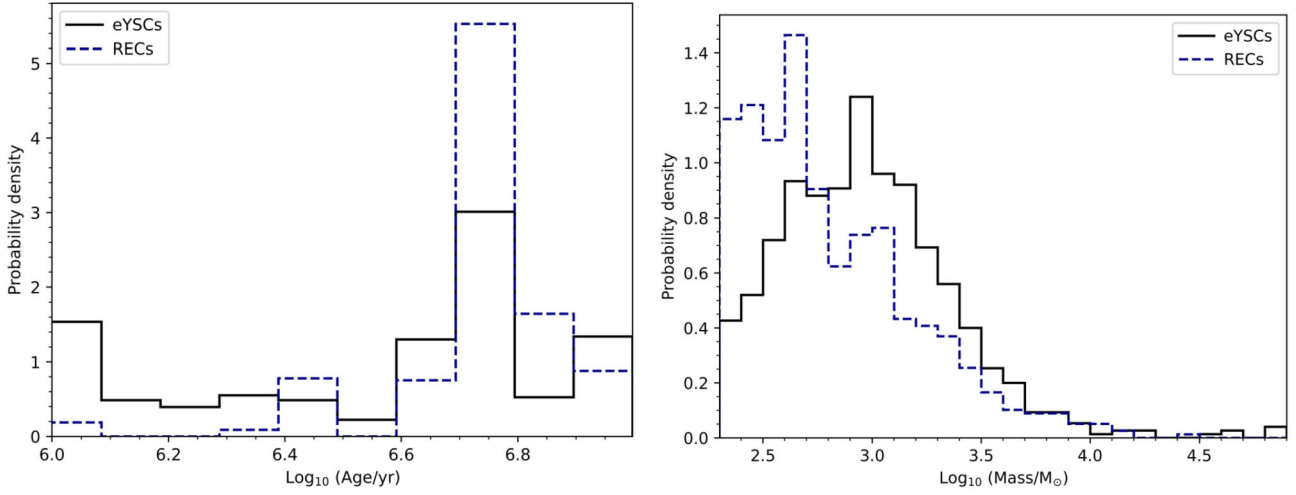


Figure A3. Probability–density histograms comparing the eYSCs and RECs: (left) $\log_{10}(\text{age yr}^{-1})$ and (right) $\log_{10}(M/M_{\odot})$. Both panels are normalized to unit area and use identical binning to enable direct comparison; the y-axis shows probability density. See Section 4.2.1.

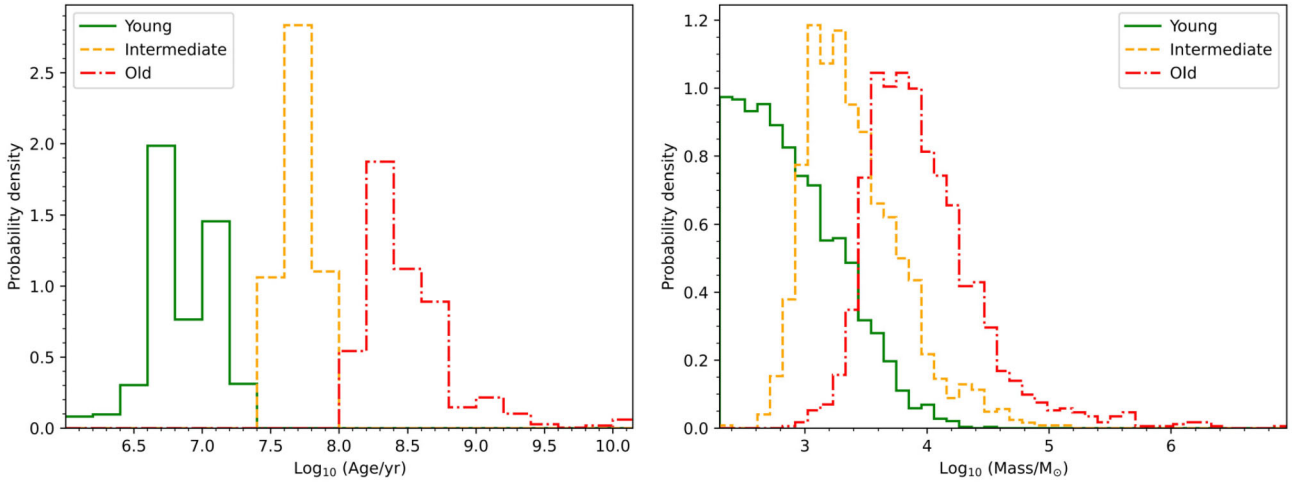


Figure A4. Probability–density histograms for the evolutionary subsamples (Young, Intermediate, Old): (left) $\log_{10}(\text{age/yr})$ and (right) $\log_{10}(M/M_{\odot})$. Both panels are normalized to unit area with identical binning; the y-axis shows probability density. See Section 4.2.2.

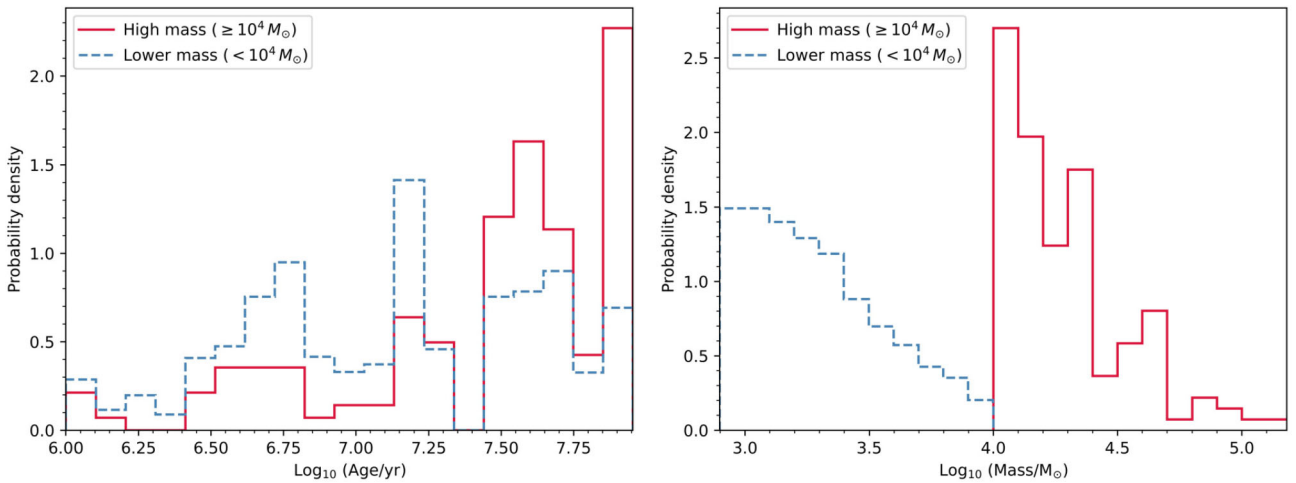


Figure A5. Probability–density histograms for the mass–selected subsamples (‘high’ and ‘lower’ mass): (left) $\log_{10}(\text{age/yr})$ and (right) $\log_{10}(M/M_{\odot})$. Both panels are normalized to unit area and use identical binning; the y-axis shows probability density. See Section 4.3.

Table A2. Median $\log_{10}(\text{age/yr})$ and $\log_{10}(M/M_{\odot})$ for each subsample, with 16th–84th percentile ranges. The final column lists the number of clusters used per subsample (identical for the age and mass estimates).

Sample	Age $\log_{10}(\text{yr})$	Mass $\log_{10}(M/M_{\odot})$	No. clusters
eYSC-NIR	$6.69^{+0.11}_{-0.52}$	$2.96^{+0.42}_{-0.35}$	1084
eYSC-Optical	$6.60^{+0.18}_{-0.55}$	$2.96^{+0.38}_{-0.36}$	132
YSC	$7.48^{+0.82}_{-0.70}$	$3.29^{+0.65}_{-0.65}$	5674
eYSCs	$6.69^{+0.15}_{-0.60}$	$2.95^{+0.38}_{-0.36}$	750
RECs	$6.78^{+0.07}_{-0.18}$	$2.71^{+0.49}_{-0.28}$	785
Young	$6.85^{+0.33}_{-0.24}$	$2.83^{+0.51}_{-0.36}$	2805
Intermediate	$7.70^{+0.20}_{-0.22}$	$3.34^{+0.45}_{-0.30}$	1201
Old	$8.48^{+0.22}_{-0.18}$	$3.90^{+0.47}_{-0.34}$	1668
High mass ($\geq 10^4 M_{\odot}$)	$7.60^{+0.30}_{-0.77}$	$4.24^{+0.31}_{-0.19}$	137
Lower mass ($< 10^4 M_{\odot}$)	$7.18^{+0.52}_{-0.57}$	$3.25^{+0.34}_{-0.24}$	2950

This paper has been typeset from a \LaTeX file prepared by the author.