MONTHLY NOTICES of the Royal Astronomical Society

On the completeness-corrected mass function of star clusters

Journal:	*Monthly Notices of the Royal Astronomical Society*
Manuscript ID	Draft
Manuscript type:	Letter
Date Submitted by the Author:	n/a
Complete List of Authors:	Khorrami, Zeinab; Cardiff University School of Physics and Astronomy, School of Physics and Astronomy Buckner, Anne; University of Leeds, School of Physics and Astronomy Gonzalez, Marta; IPAG Clark, Paul; Cardiff University, School of Physics and Astronomy Lumsden, Stuart; University of Leeds, School of Physics and Astronomy Moraux, Estelle Joncour, Isabelle Oudmaijer, Rene de la Calle, Ignacio Hacar, Alvaro; Leiden University Herrera, Jose Blanco , Jose Valero, Luis
Keywords:	stars: luminosity function, mass function < Stars, (Galaxy:) open clusters and associations: general < The Galaxy, methods: numerical < Astronomical instrumentation, methods, and techniques, techniques: photometric < Astronomical instrumentation, methods, and techniques, methods: statistical < Astronomical instrumentation, methods, and techniques

MNRAS 000, 1–5 (2019) Preprint 20 September 2019 Compiled using MNRAS LaTeX style file v3.0

On the completeness-corrected mass function of star clusters

Zeinab Khorrami,[1★] Anne S. M. Buckner,[2] Marta Gonzalez,[3] Paul C. Clark,[1] Stuart Lumsden,[2] Estelle Moraux,[3] Isabelle Joncour,[3,4] Rene D. Oudmaijer,[2] Ignacio de la Calle,[5] Alvaro Hacar,[6] Jose M. Herrera-Fernandez,[5] Jose Manuel Blanco,[5] and Luis Valero-Martin[5]

[1]School of Physics and Astronomy, Cardiff University, The Parade, Cardiff CF24 3AA, UK
[2]School of Physics and Astronomy, University of Leeds, Leeds LS2 9JT, UK
[3]Universite Grenoble Alpes, CNRS, IPAG, F-38000 Grenoble, France
[4]Department of Astronomy, University of Maryland, College Park, MD 20742, USA
[5]Quasar Science Resources, S.L., Edificio Ceudas, Ctra. de La Coruna, Km 22.300, E-28232 Las Rozas de Madrid, Madrid, Spain
[6]Leiden Observatory, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Observations of Young Massive Clusters (YMCs) play a central role in star formation research, as they represent the most extreme end of the star formation process. The initial mass function (IMF) in these clusters is of particular importance, as any variations from the standard field IMF can hint at changes in the dominant physical processes that regulate the mass of stars during their formation. However, the IMF is very difficult to observe in YMCs, and to correct for the effects of crowding, one needs to work out the completeness limits as a function of mass. This is typically done by inserting artificial stars randomly into the observed image, and seeing how often they can be recovered by the source extraction process. However, we show here that the standard technique for introducing artificial stars in the images is woefully inadequate, underestimating the completeness towards the peak of the IMF by as much as 50%. We introduce a new Monte-Carlo-Markov-Chain technique that does a much better job of estimating completeness, and thus produces more reliable IMF determinations in YMCs. Our study makes use our new MYOSOTIS tool, which is designed to make synthetic observations of star cluster simulations.

Key words: stars: luminosity function, mass function – open clusters and associations: general – methods: numerical – techniques: photometric – methods: statistical

1 INTRODUCTION

The Initial Mass Function (IMF) in Young Massive Clusters (YMCs) is of particular importance as any variation from the standard field IMF can hint at changes in the dominant physical processes that regulate the mass of stars during their formation (Clark, Klessen & Bonnell 2007; Bonnell, Larson & Zinnecker 2007; Krumholz 2014). However estimating the stellar masses is extremely uncertain, not only because of the limitations in our current atmosphere and evolutionary models (e.g. Stancliffe et al. 2016) but also because of the observational difficulties. The latter is the combination of observational instrumental limitations – lack of angular resolution, pixel sampling, sensitivity and contrast of the instrument – as well as challenges posed by the target cluster itself; newly formed star clusters are still immersed in their natal clouds and have varied extinction across the Field Of View (FOV), and the contrast is an issue when the presence of the bright massive stars start to mask the fainter stars.

Considering all these limitations and difficulties, the observed MF is unlikely to reflect the actual IMF in the cluster. In order to correct the MF for observational biases and crowding, one needs to work out the completeness limits as a function of magnitude/mass. This is typically done by inserting artificial stars randomly into the observed image to see how often they can be recovered by the source extraction process (e.g. Stolte et al. 2002; Brandner et al. 2008;

★ E-mail: KhorramiZ@cardiff.ac.uk

Hußmann et al. 2012; Khorrami, et al. 2016; Khorrami, et al. 2017). One issue of this technique is that it does not take into account the spatial distribution of the stars in the cluster: observed images are typically sensitive enough to show that there is often (depending on the FOV) a spatial variability in the density of stars. To overcome this problem, this method is improved by adding artificial stars in different radii –according to the observed density profile of the cluster– in the image (e.g. Tosi, et al. 2001; Perina, et al. 2009; Hosek, et al. 2019). However, we still do not know how good these performed since they were on observed data and the "true" completeness is unknown.

In this work we introduce a new technique to test for completeness that uses the observed image (and catalogue) to create a Probability Density Function (PDF) of stars, that can be used for choosing the location of the artificial test stars. We compare the uniformly random distribution of the artificial stars versus a technique that employs a Monte-Carlo-Markov-Chain (MCMC) to sample from a stellar density PDF, derived from the source extraction. Rather than performing this test on real cluster data, where the true completeness is unknown, we instead run our completeness tests (including source extraction) on synthetic data, obtained by post-processing N-body simulations with the new MYOSOTIS[1] tool (Khorrami et al. 2019). We can hence accurately compare the completeness from the random and MCMC test star recovery with the "true" stellar population in the N-body simulation.

Our MCMC PDF sampling technique is described in Section 2. The results from the synthetic observation analysis is presented in Section 3. Section 4 shows the example of completeness-corrected MFs and how the MF is sensitive to the completeness-correction method. The summary of the results is provided in Section 5.

2 COMPLETENESS TEST METHODS

The test for completeness is usually performed by seeing if the source extraction technique is able to recover artificial stars that are inserted into the image with a *uniformly-random distribution*. The artificial stars are inserted into the image one-by-one, and one records the fraction that are recovered by photometric analysis. By doing this for artificial stars of different magnitudes, the experiment can provide an estimate of the completeness as a function of observed magnitude.

However, the ability of the photometry to recover the artificial stars depends strongly on their chosen position within the cluster; stars placed in a crowded region will be more difficult to recover (due to blending and the high background flux), than stars placed in empty areas in the image. If the cluster properties are spatially invariant over the image FOV – for example the distribution of stars of all masses is equal and the density of stars is also equal, such that the mean flux in the image is roughly uniform everywhere – then this is not a problem for the standard technique. However any sub-clustering is going to introduce a problem: the sampling is being done assuming a uniform distribution of stars, which does not reflect the observed distribution in the cluster. As we shall demonstrate here, this can lead to a significant overestimate of the completeness.

The key parameter in the completeness analysis is the PDF which is used to choose the positions of the artificial star. In the standard technique, the PDF is simply a constant value everywhere (i.e. a uniform deviate). In this paper we suggest that a PDF derived from the image itself be used, using MCMC to draw from the PDF to provide sample locations. In our new technique [2], different PDFs are generated per magnitude, depending on the shape and brightness of the star cluster in that magnitude and below (i.e. brighter).

The spatially-varying 2D PDF that we use is the stellar density map of the extracted sources in a given magnitude range.[3] The (surface) density in the PDF is defined by,

$$\Sigma_{\Delta m} = \frac{N_{\Delta m}}{R_N^2} \quad (1)$$

where Δm denotes the magnitude range of the stars being used to create the PDF, and R_N is the distance to the nearest $N_{\Delta m}$ of the stars in equation, to that point in the PDF. We adopt an optimal value of $N_{\Delta m} = 5$ in this study, and select our magnitude ranges such that there is always at least 5 stars in the map.

The faintest magnitude in the magnitude range is that of the artificial stars that we want to test for. The brightest magnitude is that of the brightest star recovered by the original photometry. This means that when testing for the very bright stars, we use a stellar density map derived from only the sources of a similar magnitude. In contrast, when testing for the completeness of the faint stars, the stellar density PDF is created from essentially all the sources recovered in the original source extraction. For each magnitude we draw in excess of 10^6 positions using MCMC from the PDF (i.e. using the stellar density map as the target distribution). We use a 2D gaussian as the proposal distribution, with $\sigma = 20$ pixels (angular resolution is about 2.2 pixels). We then randomly sample from these 10^6 positions for our random draws of position. This ensures that the positions are not spatially correlated.

When testing for the completion of a given magnitude, we place 500 artificial stars of this magnitude on the original image using both the *Uniform Distribution Model (UDM)* and the *Spatially Varying Model (SVM)*. The photometric analysis is done using the same tool (STARFINDER, Diolaiti et al. 2000) and criteria as we used to analyse the original image. The input point spread function (PSF) for photometry is the one we used for creating and detecting the artificial stars. The minimum correlation between the input PSF and the detected source is 0.9 and the photometric threshold is one σ of the sky. The completeness value (C_f) is defined by the ratio between the number of recovered stars and the number of input stars (500 in this work) for each magnitude.

[1] https://github.com/zkhorrami/MYOSOTIS

[2] https://github.com/zkhorrami/SpatiallyVaryingModel.git

[3] Note that the stellar density map is not a strict PDF since it is not normalised (i.e. the values do not represent probability). However, MCMC works on the ratios of the map at two given locations, and so the issue of normalisation is irrelevant. We will refer to the stellar density map as the PDF here for simplicity.

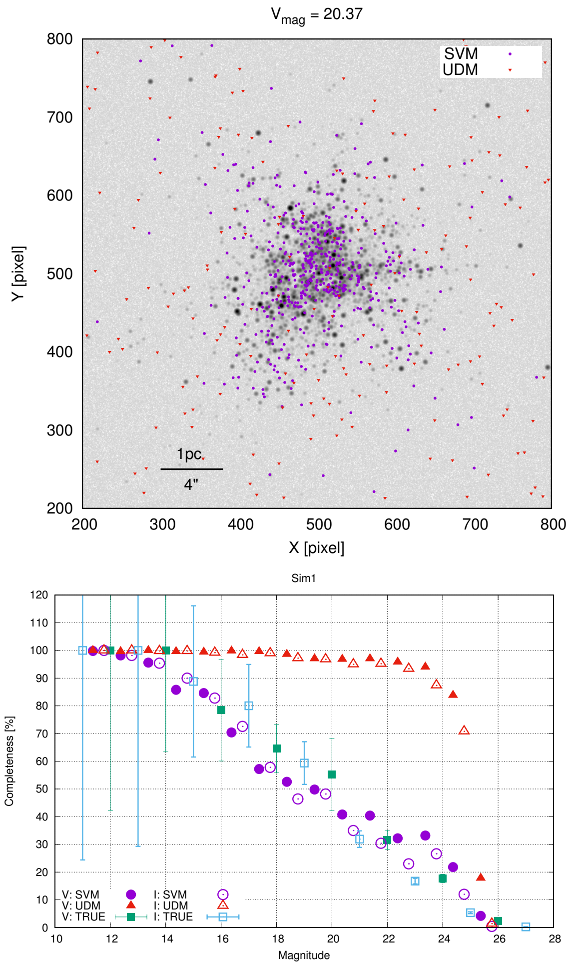

Figure 1. Top: synthetic image of Sim1 (see Table 1). Violet dots shows the position of the artificial stars with Vmag=20.37 randomly selected from the MCMC probability distribution and Red triangles shows the position of artificial stars randomly distributed in the FOV. Bottom: completeness value in different magnitudes using artificial star experiment distributed via SVM (violet) or UDM (red) methods, compared to the real completeness (green and blue). Completeness are shown for synthetic images in V-band (filled) and I-band (hollow).

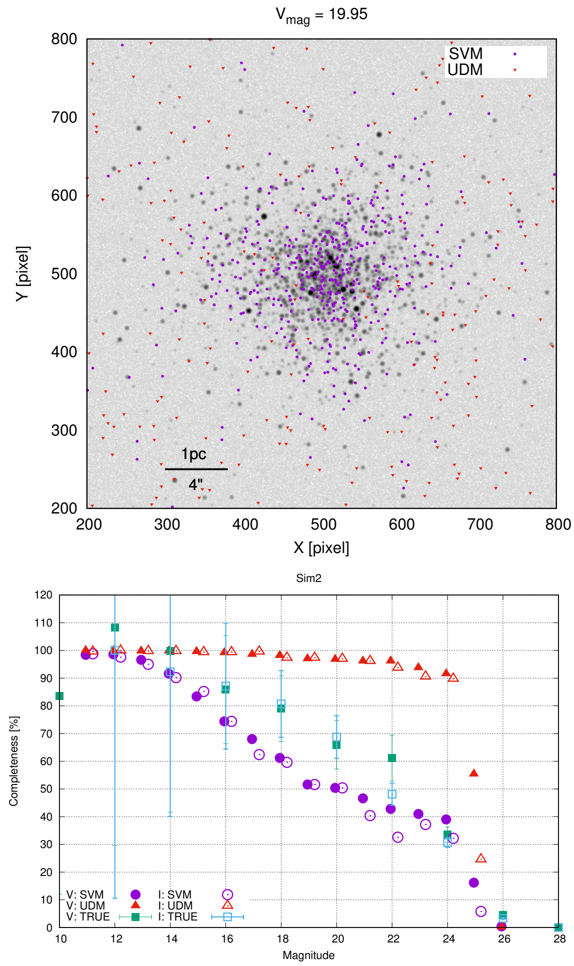

Figure 2. Same as Fig. 1, but for Sim2 and Vmag=19.95

3 SYNTHETIC OBSERVATIONS

To assess the performance of our new completeness test, we make use of synthetic observations of model clusters. Since the issues of completeness manifest more readily when both the contrast and the background in the image are high, we chose young massive clusters for our models. The physical properties of the clusters used in our study are given in Table 1. These clusters set up using MCLUSTER tool (Kupper et al. 2011). Sim1 is a star cluster with the total mass of 10^4 $M_\odot$ (14562 stars), half-mass radius (R_h) of 0.5 pc and fractal dimension (2.0. Cluster Sim2, with the total mass of 8.5×10^3 $M_\odot$ (13941 stars), R_h of 1.1 pc, and FD of 3.0, is evolved dynamically for 5 Myr using NBODY6 (Aarseth 1999), to demonstrate that the results we present here are not due to a quirk in the cluster setup. Sim3 is the combination of two fractalised clusters with R_h of 1.0 pc, separated about 2.8 pc from each other in 3 dimensions. Note that the values presented in Table 1 represent the properties at the time at which the synthetic observations are made.

The synthetic observational imaging data are simulated using MYOSOTIS tool. The clusters are given an "observed" distance similar to the Large Magellanic Cloud (LMC, 50 kpc). Simulating clusters at LMC distance is computationally cheaper than Galactic ones and we can cover larger FOV. The FOV is set to 50" × 50" ($12.5 \times 12.5\,pc^2$) for Sim1 and Sim2, and 35" × 35" ($8.75 \times 8.75\,pc^2$) for Sim3. All simulations have the pixel sampling of 50 mas/pixel and the mean full width at half maximum (FWHM) of the PSF of 0.11 arcsec, mimicking Hubble Space Telescope (HST) data by WFPC2-PC camera in the visible (V, F555W) and near-infrared (I, F814W). As we need to re-estimate observed stellar sources in these images (see next section), creating synthetic images in two filters is necessary.

Figures 1 to 3 show the synthetic HST/V images and the completeness functions using both uniform random (red triangles) and the spatially varying (purple circles) techniques. As an example, the position of the artificial stars around V_{mag} = 20 is shown on top of the synthetic images for uniform random (red) and spatially variable (purple). We see that the completeness as a function of magnitude is very different using these two techniques. For all the simulations (in both V and I filters), the uniform random techniques

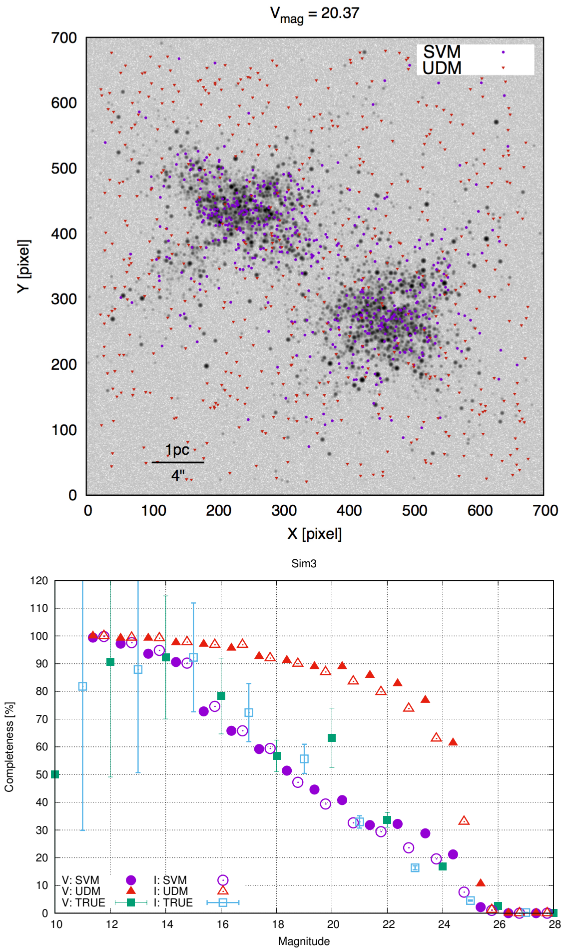

Figure 3. Same as Fig. 1, but for Sim3

gives the completeness above 90% for mag = 20. In contrast, the spatially varying technique recovers a completeness value of around 50% for this magnitude.

Since we are using simulations, we can assess which technique is closer to the "true" completeness of the source extraction. The true completeness values are estimated by comparing the synthetic catalogues with the total number of input stars in the synthetic images (green (V-band) and blue (I-band) squares in bottom plots). In the Figures 1, 2 and 3, we see that the completeness values estimated by the spatially varying method are much closer to the true value than those recovered using the uniform random technique. This is especially true in the case of Sim3 (Figure 3), where the star cluster has more structure. We stress that differences between the uniform random and spatially varying techniques for inserting the artificial stars depends on the variation in the stellar density within the image. We note that both methods are sensitive to the size of the FOV. To check if the FOV is large enough to cover the cluster's structure and its sub-clustering the INDICATE tool by Buckner, et al. 2019 can be utilised (Khorrami et al., in prep).

Table 1. Physical properties of the star clusters used for synthetic observation simulations. Total mass (M_{tot}) of the star clusters and their total stellar population (N_{tot}) are shown in the second and third columns. Fourth column is the FD of these clusters. Central density (ρ_{cen}) and the age of the clusters are given in the last two columns.

ID	M_{tot} [$M_\odot$]	N_{tot}	R_h [pc]	FD	ρ_{cen} [$M_\odot/pc^3$]	age [Myr]
Sim1	10^4	14562	0.5	2.0	1.6×10^3	2.0
Sim2	8.5×10^3	13941	1.1	3.0	1.4×10^3	5.0
Sim3	2.0×10^4	28113	-	-	-	2.0

4 COMPLETENESS-CORRECTED MF

In this section, we show what affect the differences in the completeness can have on the derived mass function for the clusters. For simulations Sim1 and Sim2, we found the common sources between V- and I-band images. For these sources we estimated the stellar masses using isochrone fitting on the V-I Color Magnitude Diagram (CMD). We used PARSEC [4] evolutionary models (Bressan et al. 2012; Chen et al. 2014, 2015; Tang et al. 2014) at the age of the simulated star clusters (2 and 5 Myr). Table 2 shows the number of detected sources in V and I from photometric analysis and the common sources between these catalogues. We considered a photometric error of $\sigma_{\rm mag}$ = 0.2mag on the apparent magnitude of the detected common sources. This corresponds to the average flux error of the extracted sources and provides us with an error on the stellar masses (σ_m). The uncertainty in the mass of each star was accounted for when constructing the MF. We estimated the slope of the MF (Γ) defined by Eq.2

$$\log_{10}(N) = \Gamma\ \log_{10}(\frac{m}{M_\odot}) + \text{constant}. \qquad (2)$$

where m is the stellar mass and N is the number of stars.

To correct the MF for completeness we divide number of stars (N) by the completeness value (C_f) of the corresponding mass. Figure 4 shows the completeness-corrected MF for Sim1 (top) and Sim2 (bottom) star clusters, using uniformly random (red triangles) and spatially varying (purple circles) techniques for estimating completeness. The green stars in these plots are the true MF in the simulations and the blue squares are the 'observed' values, obtained via the source extraction and source classification described above. Green filled area shows the mass range we used to fit the MF and estimate Γ. The lower mass limit is chosen where the completeness is above 80% using the random technique (UDM) in all filters. For the sake of comparison we fix this lower-mass limit for the MF fit in other cases (SVM and real MF). The upper sub-plots show the Γ values derived from three points fit for a given mass. We used an implementation of the nonlinear least-squares Marquardt-Levenberg algorithm to calculate the value of Γ in all cases. The expected errors due to Poisson noise (shown in Figure 4) are considered in these fittings. As the uniform random method predicts the completeness to be above 90% for medium- and

[4] Version 1.2S and CMD 3.0, available at http://stev.oapd.inaf.it/cgi-bin/cmd

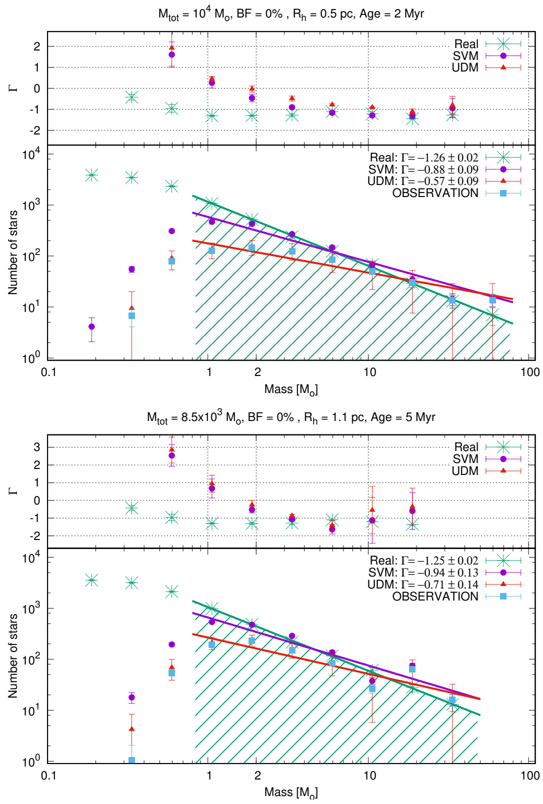

Figure 4. Completeness-corrected MF for Sim1 (top) and Sim2 (bottom), using *SVM* (purple) and *UDM* (red). True MF is shown in green stars, and the blue squares are the observed ones. The upper plots show the $\Gamma(dN/dm)$ values derived from three points fit for a given mass.

high-mass stars, the completeness-corrected MF is similar to its observed values for these mass ranges. However, for the spatially varying technique, the MF is corrected (shifted) significantly, even for masses above 1-2 $M_\odot$. This leads to estimates of the MF slope that are closer to the true values. These plots show that the small changes in the completeness function, can significantly change the MF even in the higher mass regime.

We note here that although these synthetic images are created from a very simple cluster model, containing no primordial binary systems, no extinction, no age-discrepancies, and no multiple-populations, we still find that the observed (completeness-corrected) MF cannot fully represent the true MF in the simulation. In real observational data, which is subject to all these extra features, we conclude that one must be extremely careful when using the MF to derive the physical properties of star clusters, such as the cluster's total mass or whether or not mass-segregation exists.

5 SUMMARY AND CONCLUSIONS

In this work we compared two methods for correcting stellar mass functions for completeness, using the "artificial star" technique. The artificial stars can be distributed in the original image uniformly random (uniform distribution model, UDM) or according to the shape of the star cluster (spatially varying model, SVM). In the latter method, we use

Table 2. The observed number of stars (N_{obs}) in V and I filters in the photometric analysis. N_{com} is the common sources between two catalogues which is used for stellar mass estimation. The lowest mass estimated using common sources is m_{low}. $\frac{N_{com}}{N_{(m>m_{low})}}$ shows the fraction of observed common sources versus the real number of stars more massive than m_{low}.

ID	$N_{obs}(V, I)$	N_{com}	m_{low}	$\frac{N_{com}}{N_{(m>m_{low})}}$
Sim1	(867,1566)	799	$0.52^{+0.02}_{-0.02}$	0.31
Sim2	(1053,1869)	1002	$0.68^{+0.02}_{-0.04}$	0.43

the stellar density map of the cluster as a PDF, and sample locations from it using a Monte-Carlo Markov-Chain (i.e. using the stellar density map as the target distribution).

To check which method is more accurate we applied these techniques on the synthetic observations of simulated clusters. We used our new tool (MYOSOTIS) to generate these synthetic observations for (HST/WFPC2) in the visible (F555W) and in the near-infrared (F814W). Analysis of these synthetic observations shows that stellar density map technique does a better job than the uniform random method at estimating the completeness. We also demonstrate that this technique yields completeness-corrected MFs that are closer to the true MF in the cluster than the standard uniform distribution of artificial stars. We therefore suggest that stellar density map be used to position artificial stars in completeness testing, especially when the observations reveal large variations in the stellar density.

Acknowledgments *The StarFormMapper project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 687528. PCC acknowledges support from the Science and Technology Facilities Council (under grant ST/N00706/1).*

REFERENCES

Aarseth S. J., 1999, PASP, 111, 1333
Bonnell I. A., Larson R. B., Zinnecker H., 2007, prpl.conf, 149
Bressan, A., Marigo, P., Girardi, L. et al. 2012, MNRAS, 427, 127
Brandner W., et al., 2008, A&A, 478, 137
Buckner A. S. M., et al., 2019, A&A, 622, A184
Chen, Y., Girardi, L., Bressan, A., et al. 2014, MNRAS, 444, 2525
Chen, Y., Bressan, A., Girardi, L., et al. 2015, MNRAS, 452, 1068
Clark P. C., Klessen R. S., Bonnell I. A., 2007, MNRAS, 379, 57
Diolaiti, E., et al. 2000, A&AS, 147, 335
Hosek M. W., et al., 2019, ApJ, 870, 44
Hußmann B., et al., 2012, A&A, 540, A57
Khorrami Z., Lanz T., Vakili F.,et al., 2016, A&A, 588, L7
Khorrami Z., Vakili F., Lanz T., et al., 2017, A&A, 602, A56
Khorrami, Z., et al. 2019, MNRAS, 485, 3124
Krumholz M. R., 2014, PhR, 539, 49
Kupper, A. H. W., et al., 2011, MNRAS, 417, 2300
Perina S., et al., 2009, A&A, 494, 933
Stancliffe R. J., et al., 2016, A&A, 586, A119
Stolte A., et al., 2002, A&A, 394, 459
Tang, J., et al. 2014, MNRAS, 445, 4287
Tosi M., Sabbi E., Bellazzini M., et al., 2001, AJ, 122, 1271

This paper has been typeset from a TEX/LATEX file prepared by the author.