Mass segregation and elongation of the starburst cluster Westerlund 1*

M. Gennaro,1†‡ W. Brandner,1 A. Stolte2 and Th. Henning1

1Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
2Argelander Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany

Accepted 2010 November 23. Received 2010 November 22; in original form 2010 October 14

ABSTRACT
Massive stellar clusters are the best available laboratories to study the mass function of stars. Based on NTT/SoI near-infrared photometry, we have investigated the properties of the massive young cluster Westerlund 1. From comparison with stellar models, we derived an extinction $A_{K_s} = 0.91 \pm 0.05$ mag, an age $\tau = 4 \pm 0.5$ Myr and a distance $\delta = 4.0 \pm 0.2$ kpc for Westerlund 1, as well as a total mass of $M_{\text{Wd1}} = 4.91^{+1.79}_{-0.49} \times 10^4 \, M_\odot$. Using spatially-dependent completeness corrections, we performed a two-dimensional study of the cluster's initial mass function (IMF) and, in addition, of the stellar density profiles of the cluster as a function of mass. From both IMF slope variations and stellar density, we find strong evidence of mass segregation. For a cluster with some $10^3$ stars, this is not expected at such a young age as the result of two-body relaxation alone. We also confirm previous findings on the elongation of Westerlund 1; assuming an elliptical density profile, we found an axial ratio of $a:b = 3:2$. Rapid mass segregation and elongation could be well explained as the results of subclusters merging during the formation of Westerlund 1.

Key words: stars: evolution – stars: formation – Hertzsprung–Russell and colour–magnitude diagrams – stars: luminosity function, mass function – stars: pre-main-sequence – open clusters and associations: individual: Westerlund 1.

1 INTRODUCTION
Westerlund 1 (Wd 1) is among the most-massive young clusters in the Local Group. Recent studies have revived interest in this cluster discovered almost 50 yr ago (Westerlund 1961). Several of these studies focus on the rich population of massive stars that are spectroscopically identified as Wd 1 members (see e.g. Clark et al. 2005; Negueruela & Clark 2005; Crowther et al. 2006; Negueruela, Clark & Ritchie 2010). Among this population, it has been possible to find Wolf–Rayet stars, evolved OB stars and short-lived transitional objects, like luminous blue variables and yellow hypergiants (YHGs). Wd 1 is the only case in which such a rich population of these very rare objects is observable. This makes Wd 1 one of the most-important templates for understanding the evolution of very massive stars after they leave the main sequence (MS). One of the great advantages is that the progenitor’s mass of the evolved massive stars can be inferred from the observable MS turn-off. While the massive stars are bright enough to be observed at optical wavelengths, observation of the intermediate- and low-mass stellar populations is best performed in the near-infrared (near-IR), given the high extinction towards Wd 1 of $A_V \sim 10–12$ mag (Piatti, Bica & Claria 1998). A study of the Wd 1 population below $\sim 30 M_\odot$ has been recently carried out by Brandner et al. (2008) (hereinafter Paper I). In this paper, we present comprehensive analysis of the data described in Paper I.

With the present estimates of its mass, $5 \times 10^4$–$1.5 \times 10^5 \, M_\odot$, and age, 3–6 Myr (see Clark et al. 2005; Crowther et al. 2006; Paper I; Mengel & Tacconi-Garman 2009; Negueruela et al. 2010), Wd 1 represents probably the best template in the Milky Way to understand the cluster mode of star formation that can be observed in other galaxies, like the Antennae Galaxies, where super star clusters with masses larger $10^5 \, M_\odot$ have been detected (see e.g. Whitmore et al. 2010).

In addition to the study of the intriguing formation scenario of such massive extragalactic clusters, Wd 1 may also serve as a template to understand the interplay between evolution of massive stars and dynamical processes that may lead to the formation of stable, bound and relaxed globular clusters. Given its mass, Wd 1 may indeed be able to retain a substantial fraction of its initial stellar population, even though, according to Munu et al. (2006), it has probably undergone $\sim 65$ supernova events. These, in addition to stellar winds and ionizing radiation from the most massive stars, have dispersed the residual gas reservoir of the cluster, decreasing the gravitational binding energy of the system. If massive enough to resist disruption, Wd 1 will eventually turn into a closed, virialized system. A study of the dynamical status of Wd 1 has been made by Mengel & Tacconi-Garman (2009). The authors use the

*Based on observations collected at the European Southern Observatory, La Silla, Chile, and retrieved from the ESO archive (Program ID 67.C-0514).
†E-mail: gennaro@mpia.de
‡Member of the International Max Planck Research School for Astronomy and Cosmic Physics at the University of Heidelberg, IMPRS-HD, Germany.
measured radial velocity dispersion for a group of 10 massive stars to infer a dynamical mass of $1.5 \times 10^5 M_{\odot}$, on the upper end of the Wd 1 mass estimates available in the literature. To derive this number, the authors assume virial equilibrium and isotropy of the stellar motions; hence, their estimate is an upper limit. The analysis of star clusters’ dynamical and structural parameters often assumes spherical symmetry. Hence, the cluster properties, like the initial mass function (IMF) slope, the stellar density profiles and the stars’ velocity distributions, are described as one-dimensional functions depending on the distance from the centre of the cluster. However, the spherical symmetry assumption may not be valid and this is the case of Wd 1. Several studies have already shown that Wd 1 has indeed an elongated shape, based on X-ray diffuse emission (Muno et al. 2006) and stellar counts (Paper I). Therefore, an unbiased study, which does not assume a priori any symmetry for the geometry of Wd 1, is needed to properly investigate the spatial properties of the cluster.

We focus our attention on the study of mass segregation, global and spatially varying IMF, and the overall cluster shape as can be derived by the study of density profiles. These macroscopic properties are, in turn, related to the formation history of the cluster, its internal dynamical evolution and its global interactions with the rest of the Galaxy. We developed new analysis techniques to take into account the observational biases related to the presence of many very bright objects that can hamper quantitative determination of both the IMF slope and the stellar density profiles. The most important improvement compared to Paper I is that we drop any spherical symmetry assumption regarding the cluster structure. Hence, the completeness maps, the photometric errors and the density profiles are all obtained in a two-dimensional approach. In addition, new stellar evolutionary models are used for comparison with observations. A probabilistic approach is developed to determine cluster memberships, using a nearby off-cluster image as a control frame for the field population. Stellar masses are derived using a maximum-likelihood technique, taking into account realistic photometric errors and their correlations. IMF slopes are inferred using an approach which does not require any binning, but makes use of all the information contained in each star’s mass-probability distribution. We use two-dimensional elliptical generalization of the radial density profiles by Elson, Fall & Freeman (1987) (hereinafter EFF87) to obtain shape properties of Wd 1 (e.g. its semimajor axis, elongation and orientation).

The layout of this paper is as follows. In Section 2, we describe the data set used. The technique to build completeness maps is introduced in Section 3. In Section 4, we use simulated stars to obtain photometric errors and their correlation. A statistical field subtraction method is introduced in Section 5. After the description of the adopted stellar models (Section 6), we use them and the clean colour–magnitude diagram (CMD) of Wd 1 to infer its properties, like extinction, age and distance (Section 7). An approach to obtain, for each star, its mass-probability distribution (given the adopted models) is given in Section 8, where we also derive the global IMF slope and the variation in the IMF slope across the cluster. In Section 9, we build cluster density profiles and analyse them using elliptical models. We also quantify the extent of mass segregation. The last section deals with our conclusions.

2 THE DATA

The data set used, the reduction process, and the photometric analysis and calibration have been extensively described in Paper I; hence, we will only provide a short summary here. NTT/SoFI J-

and $K_s$-broad-band observations of Wd 1 [centred on RA (J2000) = $16^h47^m03^s$, Dec. (J2000) = $-45^\circ50'37''$] and of a nearby comparison field (offset by $\approx 7$ arcmin to the east and $\approx 13$ arcmin to the south of Wd 1), each covering an area of $4.5 \times 4.5$ arcmin$^2$, were retrieved from the ESO archive (PI: J. Alves).

Data reduction was performed using the eclipse jitter routines (Devillard 2001). Point spread function (PSF) fitting photometry was derived using the IRAF implementation of DAOPHOT (Stetson 1987). The number of objects detected in both the J and the $K_s$ bands is $\approx 7000$ for the Wd 1 field and $\approx 5300$ for the comparison field. Photometric zero-points and colour terms were computed by comparison of instrumental magnitudes of relatively isolated, bright sources with counterparts in the 2MASS Point Source Catalogue (Skrutskie et al. 2006).

3 TWO-DIMENSIONAL COMPLETENESS MAPS

To obtain a correct cluster IMF and for the analysis of Wd 1 density profile, it is necessary to derive appropriate incompleteness corrections. In Paper I, the authors considered completeness correction as a function of magnitude and distance from the centre of Wd 1. In this work, we drop the assumption of radial symmetry and build incompleteness correction maps as a function of the position on the chip and of the magnitude.

The main source of incompleteness in our case is crowding, which severely affects seeing-limited observations (see e.g. Eisenhauer et al. 1998). The effects of crowding on the detection of point sources change according to two quantities: the average stellar density and the magnitude contrast between the given point source and its neighbours. Both these quantities may not follow a radially symmetric or regular distribution. Very bright objects are normally scattered over the field in a non-uniform way. Even when they have a regular distribution, they still can cause sudden and very well localized drops in the completeness. In addition, each of them has its own brightness and causes lack of detections in areas of different angular widths over the chip. Stellar density itself does not as a priori have to follow a symmetric distribution; indeed, the actual number of stars for a given position is determined by an interplay of several factors, for example, the intrinsic spatial distribution of stars within the cluster, varying extinction pattern (in the foreground, but also within the cluster) or changes in the foreground and background population characteristics, for example, within the spiral arms. For these reasons, we think that an approach that does not assume any spatial distribution in the completeness characteristics of an observed field is preferable, in contrast to integrated or averaged cluster characteristics, and is definitely recommendable when spatial properties have to be investigated. For each photometric band, we built a function with three variables:

$$C_f \equiv C(M_f | x, y, \mu),$$

where $\mu$ is the actual value of the magnitude (in the $M_f$ band) and $(x, y)$ is the position at which completeness is evaluated. It is then possible to associate an incompleteness correction to each star for each photometric band. The total incompleteness correction for a star detected in both the J and the $K_s$ bands is the product of the single corrections in each band. The reason is that each of these corrections represents the probability of detecting that given star in that specific band and detections in each band are independent of each other. The $C_f$ completeness maps have been obtained in several steps, which are detailed in Appendix A.
Mass segregation and elongation of Wd 1

Figure 1. SofI $K_S$-band image of Wd 1. Superimposed are $K_S$ 50 per cent completeness contours. The labels correspond to the $K_S$ magnitudes for which completeness is 50 per cent along the contour.

Figure 2. Comparison of the completeness values for the Wd 1 stars (black dots) and the stars in the off-cluster frame (red dots) for the $J$ and $K_S$ bands. The green dots represent stars in the Wd 1 frame, with angular distance from the cluster’s centre larger than 2 arcmin.

A visualization of the completeness pattern for Wd 1 is shown in Fig. 1. We display the $K_S$-band image of the cluster with superimposed 50 per cent completeness magnitude loci. The contours are labelled with the corresponding values of $K_S$ magnitudes for which completeness drops to 50 per cent. Such contours follow the general distribution of stars, but also show peaks around the brightest stars, as expected; from Fig. 1, it is clear that radial symmetry is not a perfect assumption for the completeness distribution of Wd 1.

A comparison of the completeness values between the Wd 1 frame and the off-cluster frame, for both photometric bands, is shown in Fig. 2. Given the spatial dependence of the completeness for the Wd 1 frame’s stars, for them, there is not a unique value of the completeness at a given magnitude; for what concerns the off-cluster frame, we assumed spatial uniformity for the completeness, so the off-cluster frame stars (red dots) have unique values of the completeness as a function of magnitude (see Appendix A). Fig. 2 shows that the completeness for the off-frame stars is always higher, at a given magnitude, than the average completeness for the Wd 1 frame stars. Similarly, 50 per cent incompleteness is reached for the control field at ≈1 mag fainter than the average 50 per cent incompleteness for the Wd 1 field. The cause of this difference may be found in the different degree of crowding of the two fields. The green dots in the figure represent stars in the Wd 1 frame located at more than 2 arcmin from the centre of the cluster, corresponding to ~2.3 pc at the cluster’s distance of 4 kpc (see Section 7.2). Even though these latter stars show – as expected – the highest completeness values for the Wd 1 frame, they still have slightly lower completeness than the off-cluster frame stars. This is a reason to believe that crowding in the ‘peripheral’ regions of the cluster frame is still higher than in the off-cluster frame, a hint to the presence of a low-mass cluster stellar population extending quite far away from the cluster centre. In Section 9, we will show evidence that the low-mass stars of Wd 1 may indeed occupy a region with a radius of the order or even larger than 3 pc.

4 PHOTOMETRIC ERRORS

As shown in Paper I, the DAOPHOT photometric errors are usually an underestimate of the true errors. DAOPHOT errors are connected to the residuals in the PSF fitting of the stellar counts. This error estimate is absolutely correct for isolated stars only – so that the photons are coming from the source of interest alone – and only if the analytical PSF model chosen for PSF fitting is the correct representation of the true PSF. In this ideal case, the errors would come only from the Poisson noise in stellar counts. In crowded fields, however, there are additional sources of uncertainty. The primary one is the presence of bright objects. Even though the light from these sources is iteratively subtracted from the frame by the PSF fitting algorithm, the unsubtracted noise in the wings of these objects can still affect the magnitude estimate of nearby faint stars. Stellar crowding itself can cause problems when the algorithm has to disentangle very close sources even when they have similar magnitudes. We use simulated stars to estimate realistic errors as a function of magnitude and position of the stars. The new estimates of the photometric errors are derived from the difference between the inserted and recovered magnitude of the simulated stars. In addition, we examine the correlation between the estimated magnitude errors in the $J$ and $K_S$ bands. The details of error evaluation are given in Appendix B. Our error estimates are shown in Fig. 3 as a function of magnitude.

5 SUBTRACTION OF THE FIELD STARS

We developed a technique for field subtraction based on a probabilistic approach. The technique takes into account the photometric
errors, their correlation and the information about completeness. The natural space for our approach is an N-dimensional magnitude space. The technique is quite general and, as long as photometric errors in different bands and their correlations are evaluated, does not have to be limited to two bands. In the case of Wd 1, we used only the J and Ks bands; hence, we will explicitly refer to them.

In the ideal case, a cluster magnitude–magnitude diagram (MMD) would look exactly the same as in the off-cluster field, plus additional stars belonging to the cluster, possibly following a separate sequence in the diagram, along an isochrone. It should be possible to compute the stellar densities at each MMD position for both the on-cluster and the off-cluster frames and compare them. Regions with an overdensity of stars would correspond to regions occupied by cluster members. The difficult part in the on-field versus off-field comparison is to compute a proper density. Usually this is accomplished by gridding the CMD and by computing a density at each grid cell. Then, according to the numbers in the cluster cells and in the off-field cells, some stars are subtracted, usually by making use of Monte Carlo techniques. This approach has been very successful in many applications, also in Paper I. Anyway, any gridding or binning procedure always implies a loss of information. Gridding is usually performed using equal cells and this does not take into account, for example, the fact that photometric errors increase with magnitude, making it less obvious to which cell a faint star should belong. On the bright parts of the CMD, the grid size may instead be very large compared to the photometric errors. In this case, the gridding would result in combining stars that, if errors would be reliable, are very distant from each other – in units of their $\sigma_{\text{phot}}$ – and then should not be considered ‘similar’ and assigned to the same cell. We decided to change this approach and calculate the density of stars locally, at each position in the cluster’s MMD where a star is located. We then calculated the density at the same point of the MMD, but for the off-frame population. The ratio of the two densities is a measure of the membership probability of the star that is in that position in the cluster’s MMD.

According to its photometric errors, each star is not a single point in the MMD, but a multidimensional Gaussian cloud of probability, representing the chance of observing that object in that position. In our two-dimensional case, these Gaussians have an elliptical symmetry with the semiaxis represented by $\sigma_j$ and $\sigma_{Ks}$, and a tilt in the MMD related to the correlation between the two magnitude errors. Since Gaussian probability is greater than 0 everywhere in the MMD, each star contributes a bit to the total density at each MMD position, the closest stars to that position having higher weight. Given a star with magnitudes ($J_i$, $K_{s,i}$), we define the density at its position in the MMD in the following way:

$$
\rho(J_i, K_{s,i}) = \sum_j \frac{1}{C_{J,j} C_{Ks,j}} \times \frac{1}{2\pi |\Sigma|^{1/2}} \times \frac{1}{2\pi |\Sigma|^{1/2}} \times \exp \left[ -\frac{(M - \mu_j)^2}{2\Sigma_{\sigma_j}^{-1}(M - \mu_j)} \right] \exp \left[ -\frac{(M - \mu_j)^2}{2\Sigma_{\sigma_j}^{-1}(M - \mu_j)} \right] \times \int dM;
$$

where the asterisk refers to the star at whose position the density is evaluated, such that $C_{J,j}$ and $C_{Ks,j}$ are the completeness fractions for that object, while $C_{J,j}$ and $C_{Ks,j}$ are the completeness fractions for the other stars. The density is calculated in the both on-field and off-field MMDs; hence, the index $i$ may run, respectively, on the stars in one or the other field. The $\mu$ vectors and the $\Sigma$ matrix are, respectively, the measured magnitudes and the covariance matrix associated to them:

$$
\mu_{j/i} = \begin{pmatrix} J_{s/i} \\ K_{s+1} \end{pmatrix}; \quad \Sigma = \begin{pmatrix} \sigma_j^2 & r \sigma_j \sigma_{Ks} \\ r \sigma_j \sigma_{Ks} & \sigma_{Ks}^2 \end{pmatrix}.
$$

$|\Sigma|$ is the determinant of the correlation matrix and $r$ is Pearson’s correlation coefficient (equation B1). The $M$ vector is the vector of coordinates ($J$, $Ks$) over which the integration is actually performed. The integration is ideally performed in the whole (infinite) magnitude space. For obvious reasons, we limit the numerical integration around each star to a region within $\pm 5\sigma_j$ for each coordinate.

Equation (1) deserves several comments. Its meaning is as follows: the contribution of the $i$th star to the density at the ($J_i$, $K_{s,i}$) position is the integral of the product of that star’s probability distribution, convolved with the probability distribution of the $j$th star. Then the total density in that position is the sum over all the $i$ stars either on-field or off-field. The probability of each single star is normalized to 1, as it has to be, but it is important to consider the completeness factors $\frac{1}{C_{J,j}}$ and $\frac{1}{C_{Ks,j}}$ for $M = J$, $Ks$ that appear in equation (1). These factors account for the missing detections in both the science and the control field. It is easy to understand why such correction is necessary. Imagine a star in the cluster field, with completeness factor 0.25; it means that if we detect that object, then (in statistical sense) there are three other similar objects that we are unable to detect. Now imagine that at the same position in the off-cluster MMD we would detect two objects both with completeness factors equal to 1. Neglect for a moment the real ‘cloud’ shape of the stars’ density-probability distributions and consider them, ideally, as points in the MMD. By computing densities without the completeness corrections, we would obtain $\rho^{\text{off}} = 1$ and $\rho^{\text{off}} = 2$. Hence, we would oversubtract that star from the cluster’s MMD. Conversely, the completeness factor tells us that the actual value of $\rho^{\text{off}}$ is not 1 but 4 and then we would subtract that object only in 2/4 = 0.5 cases or, better said, we would assign to that star a 50 per cent membership probability (see also below).

Once we have both the on-field and off-field densities at a given star’s location in the MMD, we can compare them. The ratio $\mathcal{R}_{ij} = \rho^{\text{off}} / \rho^{\text{on}}$ defines a rejection probability; the higher the contrast in the two densities – the lower $\mathcal{R}_{ij}$ – the more likely the object being a member. On the opposite side, when we are in a region of the MMD where no cluster members are present, this number approaches 1. Hence, each detected object has its associated membership probability. To decide whether or not to keep it in the catalogue of member stars, we extracted uniform random numbers $\xi \in [0, 1]$. Then, if $\xi < \mathcal{R}_{ij}$, then we discard the object; otherwise, we keep it. This also means that in the following analysis, the actual catalogues that we used may differ from one another, because some stars may be sometimes excluded or included according to this random selection. The uncertainties related to this selection directly propagate into, for example, the IMF slope evaluation. To account for this, we used multiple catalogue realizations and evaluated the uncertainties in the results as the scatter in the results (e.g. the IMF slope, see Section 8.2 for more details).

In Fig. 4, we show the CMDs\(^1\) of the Wd 1 frame and of the control frame, used as a reference for the field population, together with the results of the subtraction process. The colour-coding in

---

\(^1\) Even though the complete procedure is performed in the MMD, for the sake of clarity, we show the most commonly used CMDs, where the usual characteristics of a cluster population are better visible.
the lower panels indicates the rejection probability, $R_{\text{rej}}$. As already mentioned in Paper I, and as is clearly visible in the upper panels of the figure, the foreground and, especially, the background populations in the two frames do not look really similar. A possible cause for the differences might be different amount of extinction along the different lines of sight on-field and off-field. This population difference causes an undersubtraction of stars in certain regions of the CMD. However, it is clear that the most likely members in the lower left-hand panel (the red points) follow a well-defined cluster sequence; nevertheless, some isolated foreground and background stars in the cluster’s frame also show an artificially high membership probability. The reason is that there are no objects in the off-field MMD at the same position.

To avoid such artificial contamination, in addition to the subtraction process, we used $\sigma$-clipping of our CMD (see Appendix C). After finding the best-fitting isochrone (see Sections 6 and 7), we decided to keep only stars that lie within $3\sigma$ from it, that is, those stars that satisfy the criteria $|J_* - J_{\text{isoc}}| < 3\sigma(J_*)$ and $|K_{S,*} - K_{S,\text{isoc}}| < 3\sigma(K_{S,*})$ for at least one point $(J_{\text{isoc}}, K_{S,\text{isoc}})$ on the isochrone. An example of CMD after $\sigma$-clipping is shown in Fig. 5, together with the best-fitting isochrone.

6 THE STELLAR MODELS

In the following analysis, we use a combination of Padova MS isochrones (Marigo et al. 2008) and Pisa pre-MS (PMS) models (Degl’Innocenti et al. 2008). Padova models are accessible on the Internet2 and are already provided in the 2MASS photometric system. For Pisa isochrones, we performed the conversion from the theoretical Hertzsprung–Russell (HR) diagram to the observational 2MASS CMD ourselves. We used Brott & Hauschildt (2005) spectra, calculated with the PHOENIX model atmosphere code for the lowest-temperature regions and Castelli & Kurucz (2003) spectra, based on ATLAS9 model atmospheres for the highest temperature in the PMS isochrones (see Table 1). As in Paper I, we assumed a solar chemical composition for Wd 1; hence both the MS and the PMS models used here have this composition. Nevertheless, given the intrinsic differences in the evolutionary codes (opacity tables, equation of state, heavy elements mixture) and also given the fact that the ‘solar’ composition is not exactly the same in the two sets of

---

2 http://stev.oapd.inaf.it/cgi-bin/cmd
The knowledge of the interstellar extinction law provides the missing $A_j/A_{K_S}$ ratio. While in Paper I we adopted the widely used Rieke & Lebofsky (1985) (hereinafter RL85) extinction law, in this work, we use the much more recent Nishiyama et al. (2006) one (hereinafter N06). The authors make use of a large number of red-clump stars located in the Galactic plane. These stars have intrinsically similar colours; hence, the observed differences in colour are related to different amount of interstellar absorption. Red-clump stars describe a straight line in the $(H - K_S, J - H)$ diagram parallel to the reddening vector. Hence, the slope of this line can be used to determine the $A_j/A_H : A_j/A_{K_S}$ selective absorption ratios. In addition to the largely improved statistics, as compared to the few sources available in RL85, the N06 selective absorption has the advantage of having been measured using a $K_S$ filter, while the RL85 selective absorption used $K$. Hence, the former provides a result that is in the same photometric system as our data. The $J$-to-$K_S$ selective absorption ratio in the N06 case is given by $A_j/A_{K_S} = 3.021$, slightly higher than $A_j/A_{K_S} = 2.518$ from RL85. We checked that the obtained $A_{K_S}$ value actually does not depend on the age of the adopted isochrone. Our best-fitting isochrone of 4 Myr (see also Section 7.2) provides a value of $A_{K_S} = 0.907$ mag; if isochrones in the range 3–8 Myr are used, the scatter in the inferred $A_{K_S}$ is less than 0.01 mag. To estimate the error on the extinction value, we followed this reasoning. The absolute scatter in $J - K_S$ colour of the UMS stars used for the reddening fitting described above is about 0.2 mag. This means that a reasonable estimate for the reddening fitting error is 0.1 mag. From this, and using the N06 reddening law coefficients, it follows that the error on the inferred total extinction can be estimated as $\Delta A_{K_S} = 0.05$ mag.

Given the errors and the results of Paper I, with $A_{K_S} = 1.13 \pm 0.03$ mag, it may seem that our new findings are inconsistent with the previous ones. Nevertheless, one always has to keep in mind two crucial sources of systematic uncertainty in the method used and that are not included in the error estimates above. One is of course the choice of the stellar models, which may differ from one another both in the theoretical HR diagram and in the transformations used to convert temperatures and luminosities into colours and magnitudes. A difference of 0.05 mag in the intrinsic near-IR colours of UMS stars is anything but unexpected. We compared the Padova models used in this work with the Geneva models used in Paper I (Lejeune & Schaerer 2001), using in both cases solar metallicity and ages of 4 and 3.9 Myr, respectively. We observed differences in $J - K_S$ intrinsic colour ranging from 0.03 to 0.1 mag, at a given magnitude, in the mass interval 5–30 $M_\odot$, used for the reddening estimate. The other source of systematic uncertainty is the aforementioned choice in such a CMD simply looks like a vertical line. Therefore, it is possible to estimate the reddening towards Wd 1 by fitting the $J - K_S$ colour of the UMS. To perform the fit, we used the stars for which $K_S < 13.5$ mag and $1.2 < J - K_S < 2.0$ mag and minimized the quantity:

$$\sum_j [(J - K_S)_j - (J - K_S)_{\text{loc}}].$$

where $j$ runs over the selected stars and the isochrone colour is taken at the same $K_S$ of the $j$th star. The $J - K_S$ colour selection reduces the contamination by stars clearly belonging to the foreground or background population. Once the $J - K_S$ reddening has been estimated, extinction $A_{K_S}$ is computed using an extinction law. Since, by definition, $E_{J/K} = A_j - A_{K_S}$, we have

$$A_{K_S} = \frac{E_{J/K}}{(A_J/A_{K_S}) - 1}.$$

models, they show some differences in the region of overlap. Small differences are also present between the sets of PMS isochrones transformed with PHOENIX and ATLAS9 model atmospheres. We have carefully chosen the masses for the transition from one set of models to the other, in order to minimize the differences in colour between them. The colour differences are shown in Table 1, together with the mass and temperature ranges in which we adopt each model. The PS–ATLAS9 isochrones have been shifted in order to match the Padova isochrones at 4 $M_\odot$ and the PS–PHOENIX isochrones have been shifted to match the PS–ATLAS9 ones at 2 $M_\odot$. Table 1 shows that the offsets are quite small, especially when compared to the expected absolute precision in our photometry, which, taking into account the zero-point errors, is of the order of 0.05–0.1 mag.

### 7 FUNDAMENTAL PARAMETERS OF WD 1

Before proceeding with the spatially-dependent analysis, we derived the global, average properties of Wd 1 using the combined isochrones described above.

#### 7.1 reddening and extinction

For high-mass stars on the MS, the near-IR part of the spectrum is very well approximated by the Rayleigh–Jeans tail of a blackbody with temperature $T_{\text{eff}}$. Then, for masses above $\sim 5 M_\odot$, given that the SED shape is almost unchanged, the near-IR, $J - K_S$, colours stay constant (at around zero magnitude). The upper MS (UMS)
of the reddening law. In Paper I, we used the RL85 law and, given that the selective absorption ratios are quite different between RL85 and N06, this explains the difference in our previous and new results for the total extinction.

To compare our findings with those by other authors, our best $A_{K_S}$ value cannot be directly converted into an $A_V$ value using only the N06 law. This law has indeed been obtained only from the $J$-band redwards (see also Nishiyama et al. 2009, for the extension of the N06 reddening law towards photometric bands redder than $K_S$). Hence, we use a combination of the $A_J/A_{K_S} = 3.021$ ratio from N06 and the $A_V/A_J = 3.546$ ratio from RL85 to obtain $A_V = 9.7$ mag. As already noted in Paper I, different authors report values of $A_V$ that vary in the range from 9.4 to ~12.0 mag, so our final value is included well within this range.

Recently, Negueruela et al. (2010) have observed the presence of differential reddening across Wd 1. They report a range of $\Delta E_{K_S} \approx 1.4$ mag. This range can be converted into a range of $\Delta E_{J-K_S} = 0.51$ mag, again using a combination of RL85 and N06 laws, matched at the $J$-band. The observed colour range for the UMS members in our data set is somewhat smaller than this and part of this spread is probably also due to photometric errors and undetected binarity. Hence, differential reddening across the cluster cannot be excluded, but Negueruela et al. (2010) extinction spread has to be regarded as an upper limit.

7.2 Distance and age

As illustrated in Paper I, the morphology of the PMS–MS transition region and of the whole PMS can be combined as a good age indicator for young clusters. Since extinction is determined independently (see Section 7.1), the distance modulus, DM, and the age, $\tau$, can be determined without having extinction as a free parameter.

Good age constraint is provided by those stars that have just entered the MS. These stars are located at the base of the vertical MS and have $14.9 \lesssim K_S \lesssim 15.1$ mag and $1.6 \lesssim J - K_S \lesssim 1.8$ mag. No cluster members are present at magnitudes immediately fainter than that (see the lower left-hand panel in Fig. 4). This zero-age main-sequence (ZAMS) region is very well identifiable in the cluster’s CMD and can be used to anchor the isochrones position. It is worth mentioning that the determination of the age and DM in this paper is not done by a real fitting procedure, but through the conventional superposition of different isochrones for several values of the pair (DM, $\tau$). The DM and age values would be degenerate if only the ZAMS position would have been used for their determination. A slightly older isochrone would have an intrinsically fainter ZAMS point and this could be compensated by a reduction in the DM. Isochrones of different ages, however, also show different colours for the PMS branch, the younger, the redder. Hence, in our comparison, after trying to reproduce the ZAMS point, we also take into account the shape of the PMS-to-MS transition region and the PMS colour. The uncertainty in the DM determination can be reasonably quantified as $\DeltaDM = 0.1$ mag from the magnitude extension of the ZAMS region. The minimum age uncertainty that we can quote is instead half of the spacing between the different isochrones in our grid, that is, 0.5 Myr. By isochrone superposition, we obtain our fiducial values of DM = 13.0 ± 0.1 mag (corresponding to a distance $d = 4.0 \pm 0.2$ kpc) and $\tau = 4 \pm 0.5$ Myr.

In Paper I, we found values of DM = 12.75 ± 0.10 mag ($d = 3.55 \pm 0.17$ kpc) and $\tau_{\text{PMS}} = 3.2$ Myr for the PMS population, while the MS stars provided weaker constraints on the age with $\tau_{\text{MS}}$ between 3 and 5 Myr. The use of more recent PMS models partially reconciles our findings with those of other authors. For example, Crowther et al. (2006), by comparing the number of Wolf–Rayet stars and of cool hypergiants, find DM = 13.4 ($d = 4.8$ kpc) and $\tau = 4.5$ or 5 Myr. From observations of H I, Kothes & Dougherty (2007) find a distance $d = 3.9 \pm 0.7$ kpc. Negueruela et al. (2010), from a comparison of their spectroscopically classified objects with models by Meynet & Maeder (2000), favour values of $d \gtrsim 5$ kpc and $\tau \gtrsim 5$ Myr. The authors point out the difficulties in spectral classification for several objects, the approximate character of the $T_{\text{eff}}$ scale, the uncertainties in $M_V$ values and, finally, the uncertainty in stellar evolutionary models for massive stars. The values of $d \sim 5$ kpc and $\tau \sim 5$ Myr are also supported by Ritchie et al. (2010), where the authors derive constraints on these quantities from the study of a massive, interacting, eclipsing binary. Clearly, there are still difficulties in the determination of the distance and age for Wd 1 with different methods providing slightly different values. Nevertheless, with this paper, the differences between the values inferred using the intermediate- and low-mass ends of the stellar population, on one hand, and those inferred using the high-mass end, on the other hand, are somehow reduced.

8 THE IMF OF WD 1

The comparison of observed magnitudes with isochrones allows the determination of stellar masses. The mass-probability distribution for each star was determined by taking into account the magnitude errors and their correlation. The distributions for the single stars are then combined to build the IMF of Wd 1. The detailed information on the completeness pattern across the field allows us to explore the variations in the IMF slope within Wd 1. In the following, we consider all our objects as single stars; nevertheless, we are aware of the possible biases introduced by neglecting the presence of binaries (see Maiz Apellániz 2009). This will be accounted for in an upcoming paper.

8.1 The mass of the single stars

Given our best-fitting isochrone (see Section 7), we used a maximum-likelihood approach to determine the mass of the member stars. Again we work in the magnitude–magnitude space. There a star is characterized by its average magnitudes, by its photometric errors and by the correlation between them. Isochrones in the MMD are curves parameterized by the mass value of the star, $m$. Hence, the probability of a star with mass $m$ and magnitudes $M(m) = (J(m), K_S(m))$ to be observed at the $\mu_s = (J_s, K_{S_s})$ location in the MMD is

$$p(m) = \frac{1}{2\pi|\Sigma|^{|1/2}} \times \exp \left\{ -\frac{1}{2} [M(m) - \mu_s]^T \Sigma_s^{-1} [M(m) - \mu_s] \right\}$$

(3)

(see equations 1 and 2 for the definition of the symbols).

Note that $\int_0^\infty p(m) dm = 1$; hence, $p(m)$ represents a probability distribution. With this approach, we can determine not only the most likely mass for each star, by maximizing $p(m)$, but also the reliability of the mass value obtained. If a star is indeed located very far from the best-fitting isochrone (in units of its photometric $\sigma$), then its $p(m)$ will be a very broad function, with a poorly determined peak. On the contrary, if the star lies exactly on the isochrone, then, ideally, $p(m)$ will be a Dirac $\delta$ function.
10 M is the projected half-mass radius, M is the mass function for Wd 1; the dashed box indicates the region that is used for the fit of the IMF slope. Right-hand panel: zoomed version for the dashed box region; γ = 2.44 and A = 12200 are our best estimates of the power-law coefficient and IMF normalization constant, respectively. Red lines correspond to the completeness-corrected function, the uncorrected function is shown in black for comparison. The blue line in the right-hand panel is the best-fitting power law.

8.2 IMF slope and total mass determination

A standard approach to evaluate the IMF slope of a cluster is to build a histogram of the stellar masses and then fit a power law (or a lognormal distribution) to the histogram. It is known, however, that the value of the slope is quite sensitive to the way the binning is performed and even to the space in which the fitting is done, that is, a linear or logarithmic space for the mass coordinate (see e.g. Maiz Apellániz 2009, for an exhaustive description of the subject). These problems were also discussed in Paper I where we showed that the cumulative mass distribution, not requiring any binning, can be used to give stronger constraints on the IMF slope. Here, we introduce an alternative method that does not require any binning and makes use of the fundamental information on the mass-probability distribution, p(m), which is always ignored when only the best-mass values are used, even without binning. Given p(m) for each star in Wd 1, we define the observed mass function:

\[ \frac{dN(m)}{dm} = \sum_i \frac{1}{C_{K,i}} \times \frac{1}{C_{K,1}} \times p_i(m). \] (4)

The \( \frac{dN(m)}{dm} \) function for the whole Wd 1 population is shown in Fig. 6. We used a restricted range of masses to determine the global slope of the IMF, \( \gamma \), where \( \frac{dN(m)}{dm} = A \times m^{-\gamma} \) with normalization constant A and \( \gamma = 2.3 \) for a typical Salpeter or Kroupa IMF, in the mass regime above 0.5 M⊙ (Salpeter 1955; Kroupa 2001). The lower mass limit for the slope fit is chosen to be \( m_{\text{min}} = 3.5 \) M⊙. At this mass, we have 50 per cent global completeness on the whole frame. Locally, this value could be different. For example, in the very centre of the cluster, high incompleteness is reached at high values of the stellar mass (see the lower panels of Fig. 8, shown later). This may cause some additional uncertainty on the derived IMF slope. The effects of spatially varying incompleteness are investigated in detail in Section 8.3, where the potential bias in the cluster centre is also analysed. The upper mass limit for the slope fit is chosen to be \( m_{\text{max}} = 27 \) M⊙. The reasons for this limit are: (i) the magnitude limit of our data set. Stars more massive than this are above the linearity regime of the NTT/SofI observations that we have used; (ii) stars above this mass are close to the turn-off region, according to Padova isochrones. Hence, the determination of their initial masses starts to be age-dependent and the complex post-MS evolution of such massive stars is quite uncertain from the theoretical point of view; and (iii) the fitting procedure: above this mass value, the numbers become so small that statistical fluctuation is not negligible and could lead to a bad fit.

We show the results in bilogarithmic plots, but the actual fit has been performed in a linear space. The global IMF slope we obtain is \( \gamma = 2.44^{+0.20}_{-0.08} \), slightly steeper than an ordinary Salpeter/Kroupa IMF. We will explore in Section 8.3 local departures from this behaviour. For the normalization constant, we found \( A = 1.23^{+0.56}_{-0.14} \times 10^4 \). The best values and uncertainty of \( \gamma \) and A are evaluated by using a bootstrap technique, as detailed in Appendix D. Given the couple of values (\( \gamma, A \)) obtained from a single bootstrap sample, it is possible to associate with them a value of the total mass and total number of stars for Wd 1. We extrapolate the power law with index \( \gamma \), in the range \( m(M_\odot) \in [0.5, 120] \). The upper mass limit is a reasonable estimate of the highest stellar mass that is expected to form in a massive cluster as Wd 1. From Padova isochrones, we have that stars with initial masses larger than \( \sim 65 \) M⊙ are supposed to have already undergone supernova explosions at the estimate cluster age of 4 Myr. Hence, our results are estimates of the total initial mass and total initial number of stars for the cluster, under the assumption that the present-day mass function is a representative of the IMF value. For masses below 0.5 M⊙ and down to the hydrogen burning limit, that is, 0.08 M⊙, we used the Kroupa IMF slope for this stellar regime, with \( \gamma = 1.3 \). In Appendix D, we also show how we derived the best estimates for the total number of stars and the total mass of the cluster, given the set of \( N_{\text{ind}}, i \) and \( M_{\text{tot}}, i \) from the different bootstrap samples. The total number of stars is \( N_{\text{ind}} = 1.04^{+0.60}_{-0.33} \times 10^5 \), while the total mass of the cluster is estimated to be \( M_{\text{tot}} = 4.91^{+1.79}_{-0.49} \times 10^4 \) M⊙.

Our present findings, based on a more complete and thorough approach, confirm the findings of Paper I and are on the lower end of the recent literature estimates for the mass of Wd 1. Using the MS turn-off mass and the identified post-MS member, by extrapolation of a Kroupa IMF down to lower masses, Clark et al. (2005) found a somewhat higher value for the total mass of \( \sim 10^5 \) M⊙. Part of this discrepancy could be ascribed to the model-dependent uncertainties in the determination of the progenitor mass for the post-MS-identified members. Additionally, one has to be cautious when counting only the very massive stars to normalize the Kroupa IMF and then extrapolate it all the way down to low-mass stars. Only few young clusters in the Milky Way are known for which the IMF can be actually measured up to this masses; hence, the nature of the IMF and its exact form is not known with great certainty in this regime. Moreover, also in the case that a standard IMF is valid for the very massive stars, high stochastic (Poissonian) fluctuations are expected when the numbers become small as towards the very high mass end of the Wd 1 population. A completely different approach was used by Mengel & Tacconi-Garman (2009) to determine a gravitating mass of Wd 1, \( M_{\text{dyn}} = 1.5^{+0.7}_{-0.5} \times 10^4 \) M⊙. The authors measured the radial velocity of \( \sim 10 \) stars from their spectra. From the dispersion of these velocity measurements, the total mass of the system is derived, under the hypothesis of virial equilibrium, using the following equation:

\[ M_{\text{dyn}} = \frac{\eta \sigma^2 r_{\text{hp}}}{G}, \]

where \( r_{\text{hp}} \) is the projected half-mass radius, \( \sigma \) the velocity dispersion and \( \eta \) is a factor that the authors use under the additional assumption of isotropy. Possible pulsations in the five YHGs of the sample, which would cause a wrong estimate of their radial velocities, may cause an overestimate of the true \( \sigma \). Ritchie et al. (2009), indeed, demonstrate that one of the YHG observed in Wd 1, W243, shows...
a very complex, time-varying spectrum with signs of pulsation and mass-loss that may hamper a precise determination of the radial velocity. This star is not in the Mengel & Tacconi-Garman (2009) sample, but it exemplifies that velocity dispersions derived from radial velocity measurements of evolved stars can lead to an overestimate of the true dispersion. In addition to this, we think that part of the discrepancy in the inferred dynamical mass could derive from the fact that Wd 1 is actually non-spherical (see Section 9) and this anisotropy might also be reflected in the stellar motions. Therefore, the η factor used by the authors should be slightly modified, possibly giving better agreement with other findings. Indeed, the velocity distribution seems to be non-isotropic from our preliminary analysis of stellar proper motion using multi-epoch near-IR Adaptive Optics data (Kudryavtseva et al., in preparation). Conversely, Fleck et al. (2006) showed that the η parameter is a time-dependent quantity, which changes rapidly, especially in very rich clusters, due to the effects of mass segregation. The authors found that the use of η ≈ 10, like in Mengel & Tacconi-Garman (2009), may lead to underestimates of clusters masses. We will show in the following that Wd 1 is mass segregated. In such a case, an increase in η is needed to correctly estimate its dynamical mass. This would lead to an even stronger discrepancy with our photometric mass estimate. A possible interpretation of this difference could be that Wd 1 is indeed out of virial equilibrium, with stellar motions still not relaxed after the gas-expulsion phase that followed the first supernova explosions. An effect that could balance the effects of mass segregation on the η value is the inclusion of binaries in the estimates of this parameter. Binary orbital motions increase the measured value of the velocity dispersion; consequently, the true mass of a cluster is overestimated if the binary contribution is not properly taken into account. Kouwenhoven & de Grijs (2008) showed the dependency of the η value on binary properties and cluster density. For the densest clusters (N ≥ 10^2 stars pc^{-1}), the bulk motions dominate the total value of σ^2, while for the sparsest clusters, with N ~ 0.1 stars pc^{-1}, the velocity dispersion is fully dominated by orbital motions. Wd 1 density is in between these two extreme values. In this case, the dynamical mass can be overestimated by 10–100 per cent, depending on the properties of the binary population. Gieles, Sana & Portegies Zwart (2010a), including mass-dependent mass-to-light ratio of stars and the intrinsically different binary properties of massive stars, found that the contribution to σ^2 from binary orbital motions is already very important for young (~10 Myr), moderately massive (M ~ 10^4 M☉) and compact (r_{ap} ~ 1 pc) star clusters, comparable to Wd 1.

8.3 Spatial variability of the IMF

In Paper I, we have shown that, considering concentric annuli centred on the Wd 1 centre and computing the slope of the IMF for the stars in the annuli, there is a tendency for a flattening of the IMF when going closer to the centre. The IMF slope was computed using stars more massive than 3.4 M☉ only to avoid any bias due to the lower degree of completeness in the crowded centre of the cluster. Still, close to the brightest stars, 50 per cent incompleteness is reached already at higher masses of up to 6 M☉. This may still cause an artificially flatter IMF in the central parts of Wd 1, because, even though the IMF slope is obtained using the incompleteness corrected number of detected stars, the correction itself becomes quite uncertain when one uses it for much lower levels than 50 per cent completeness. We have also shown that the completeness pattern in Wd 1 is not really radially symmetric and we will show in the following Section 9 that the shape of the cluster itself is elongated; hence, using concentric annuli can smooth out some of the intrinsic spatial variations of the IMF. With our new approach, we determine the IMF slope locally, in order to follow its real pattern within Wd 1.

To calculate the IMF slope at each position, we used a moving box, 200 pixels in size. The slope was obtained by the same technique as described in Section 8.2, applied only to the stars in the box. At each position, we additionally selected stars such that the total completeness factor C_J × C_KS is always higher than some fixed threshold values. In this way, we can compare results for varying completeness thresholds. Hence, at each position, the minimum mass considered can be different. The upper mass limit is determined by the stochastic distribution of non-saturated high-mass stars within the moving box. The fit is performed only when the number of stars inside the box is larger than 30. The calculation is repeated at each pixel. Anyway we are forced to use a moving box that is much bigger than the sampling scale, because we need enough stars to perform the IMF slope fit. Hence, the adjacent-pixel slope values are not independent of each other. The final maps are obtained by convolving the adjacent-pixel slope values with a Gaussian kernel of full width at half-maximum (FWHM) = 100 pixels (half-box size). At the distance of Wd 1 (~4 kpc) and with a plate scale of 0.29 arcsec pixel^{-1}, 100 pixels corresponds to ~0.5 pc in linear scale.

Results for three values of C_J × C_KS, 0.125, 0.25 and 0.375, are shown in Fig. 7. With our definitions, the conventional 50 per cent threshold in one band is replaced by a 0.5 × 0.5 =~ 0.25 combined threshold. The green areas correspond to an ordinary Kroupa-like slope (γ ≈ 2.3). The yellow-red areas in the centre indicate regions with a flatter IMF, that is, with more high-mass stars than what is predicted by a Kroupa IMF. Blue-purple areas are areas with a paucity of massive stars. Hence, when fitting a power law we obtain a very steep function due to the overabundance of low-mass stars. The contours in the three panels are the contours for the C_Kcorr × C_KS function, that is, the product of the J and Ks 50 per cent completeness magnitudes. These contours trace the shape of the total completeness correction factor.

The overall pattern of the IMF slope remains unchanged among the three different maps of Fig. 7. Nevertheless, some differences can be noted. Going to lower and lower completeness thresholds (i.e. from the right-hand to left-hand side in the figure), the yellow region in the centre of the image ‘shrinks’, leaving space for regions of slightly steeper IMF around it. Hence, when completeness corrections are properly taken into account, there are strong hints that low-mass stars are overabundant outside the very centre of Wd 1. Conversely, the yellow–red regions still visible in the centre of the maps, even for the lowest completeness threshold, indicate an overabundance of massive stars that is very likely to be intrinsic and not just a result of missing detections in the low-mass end. A similar ‘shrinking’ behaviour is observed for the ‘purple’ outer regions that are very pronounced in the rightmost panels of Fig. 7 and less in the leftmost panels. In this case, the effect is due to the difference in the mass intervals used for the fit of the IMF slope. At higher completeness thresholds, only few mass points are available and the differences in number counts between the low-mass and high mass limits within the fitting interval are very high when few high-mass stars are present. When lowering the completeness threshold, star counts are added at lower masses; hence, the observed mass function becomes more ‘regular’ and the results of the fit of the slope are less extreme.

Given the low number of stars towards the high-mass end, statistical fluctuations in this regime may increase the uncertainty of the
Figure 7. Two-dimensional maps of the IMF slope of Wd 1. Within our definition, the colour-coding corresponds to values of $-\gamma$. The three maps are built using only stars with a completeness factor, $C_J \times C_{K_S}$, down to 0.125, 0.25 and 0.375 (from the left-hand to right-hand side); overplotted are the contours of $J_{\text{half}} \times K_{\text{half}}$. The x- and y-axes correspond to RA and Dec. offsets, in arcmin, relative to the centre of the reduced image, (RA, Dec.) = ($16^h 47^m 06^s$, $-45^\circ 50' 33''$).

Figure 8. Upper panels: differences in the $\gamma$ values using stars between the completeness-threshold mass and $7 M_\odot$ and using all stars above the threshold mass. Lower panels: values of the completeness-threshold mass. Columns from the left-hand to right-hand side correspond to $C_J \times C_{K_S}$ = 0.125, 0.25 and 0.5, respectively. Grey areas are areas with not enough stars to perform a reliable fit of the IMF slope. The meaning of the x- and y-axes is the same as in Fig. 7.
IMF slope. To see whether this effect is important, we compared the sample IMF slopes for all stars with $m > m_{\text{lim}}$ and the sample IMF slope for all stars with $m_{\text{lim}} < m < 7 M_\odot$ only. The results are shown in the upper panels of Fig. 8. The grey regions in the centre are regions where the number of stars in the fitting interval was too low to perform a good fit. Excluding these regions, it is clear that the difference between the two slopes is almost everywhere zero. This tells us that the fit is dominated by the low-mass regime of the fitting interval, where the stars are more numerous and where the overall shape of the IMF is very well determined, since statistical fluctuations are less pronounced. The only differences between the two slopes are observed in the very centre, where $m_{\text{lim}}$ becomes very close to the upper mass limit of 7 $M_\odot$. In these regions, indicated by a cyan colour, the inferred IMF is flatter when the high-mass end is neglected. Anyway, the incompleteness level in the very centre is high; hence, these small differences ($\Delta \gamma \lesssim 0.3$) cannot be considered significant.

Summarizing, we can say that the overall IMF slope is consistent with a Salpeter or Kroupa Galactic IMF in the range of masses between 3.5 and 27 $M_\odot$. This slope is the spatial average of a slope that varies across Wd 1. A trend in the local IMF slope values can be observed in Fig. 7, with central regions having flatter IMF compared to the outer regions of the cluster. This is a robust indication that Wd 1 is mass segregated. We will show additional evidence of this mass segregation in Section 9, where we will also discuss its possible origins.

9 MORPHOLOGY OF Wd 1

Several recent studies indicate that Wd 1 is elongated (see e.g. Munó et al. 2006). In Paper I, assuming an elliptical shape with $a$ and $b$ as the semimajor and semiminor axes, respectively, we found an ellipticity of the cluster, $\eta = 1 - \frac{b}{a} = 0.19$, when stars with masses in the range 10–32 $M_\odot$ were considered. The value slightly decreased, to $\eta = 0.15$, using masses between 3.5 and 10 $M_\odot$. Elongation was computed by calculating the half-mass radius as a function of the position angle (PA), considering for each PA only stars within $\pm 45^\circ$ around the PA and, correspondingly, around PA $+ 180^\circ$. We also showed that the overall surface mass density profile of the cluster follows a $\Sigma(r) \propto [1 + (r/a)^2]^{-\beta}$ radial law (see EFF87), with the core radius related to the $\alpha$ parameter by EFF87 equation (22), that is, $r_0 \approx \alpha (2/\beta - 1)\beta/2$ and $\beta = 2$ for Wd 1. At large distances from the centre, the three-dimensional density profile goes like $\rho(r) \propto r^{-2\beta-1}$—see EFF87 equations (13a) and (13b). Hence, an index $\beta = 2$ for the two-dimensional density profile implies a three-dimensional density that goes like $r^{-2}$, which corresponds to a Plummer (1911) model. An index $\beta = 0.5$, instead, corresponds to an isothermal sphere with three-dimensional density going like $r^{-2}$.

The density profile of Wd 1 falls more rapidly compared to the case of the R136 cluster in the Large Magellanic Cloud. This cluster has a mass comparable to that of Wd 1 and slightly younger age of $\sim 3$ Myr, but shows a profile that is closer to isothermal, with $\beta \approx 0.8$ (Andersen et al. 2009; Campbell et al. 2010).

Our two-dimensional incompleteness mapping enables a study of the cluster’s two-dimensional stellar density distribution. We calculated the surface number density for several mass ranges and used four values for the completeness threshold. Given a lower mass threshold, $m_{\text{lim}}$, and a completeness threshold, $C_w$, we considered all the stars above these thresholds for calculating the stellar surface number density. The number density was computed using a moving box 100 pixels in size, which was moved pixel by pixel. After counting the stars at each position, we convolved the counts with a Gaussian kernel of FWHM = 50 pixels, that is, half-box-size, to account for the fact that the density value computed at each pixel position is not independent of the values computed at nearby pixels.

In this way, we have been able to build smooth number density maps for Wd 1. These density profiles are always elongated; hence, we decided to perform a fit by using an elliptical generalization of the EFF87 profile, a natural extension of the work done in Paper I. We will refer to this profile as the GEFF profile.

The GEFF profile can be described in the following way:

$$\Sigma_{\text{GEFF}} = \Sigma_{\text{BG}} + \Sigma_{1}(1 + L^2)^{-\Gamma},$$

where $\Sigma_{\text{BG}}$ is a stellar background density, $\Sigma_1$ is the density in the centre and $\Gamma$ represents the density decay for large distances from the centre ($L \gg 1$).

The quantity $L^2$ is given by

$$L^2 = \left( \frac{x'}{a} \right)^2 + \left( \frac{y'}{b} \right)^2.$$

In analogy with the EFF87 $\alpha$ parameter, which is related to the core radius, $a$ and $b$ are related to the core semimajor and semiminor axes of the elliptically symmetric GEFF profile. The quantities $x'$ and $y'$ are given by

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x - x_c \\ y - y_c \end{pmatrix},$$

where $(x_c, y_c)$ are the pixel coordinates of the centre of the ellipse and $\theta$ is the tilt angle between the semimajor axis and the $x$-axis, measured counterclockwise.

Summarizing, a GEFF profile has eight different parameters: $P = (\Sigma_{\text{BG}}, \Sigma_1, \Gamma, a, b, x_c, y_c, \theta)$. In the fit, we left all of them free, apart from the exponent $\Gamma$. Since the equivalent exponent for an EFF87 profile was found to be $\beta = 2$ in Paper I, we constrained $\Gamma$ to stay between 1 and 3.

Moreover, given that the completeness correction in the very centre of Wd 1 may be uncertain, we performed the fit by neglecting the region in which the stars have, on average, a completeness factor smaller than 0.25. An example of stellar density contours and the relative GEFF fit is given in Fig. 9.

9.1 Results of the GEFF fit

We calculated the stellar density for several combinations of the mass and completeness thresholds; the values used are $m [M_\odot] = [2.5, 3.5, 4.5, 6.0, 7.5, 10.0, 12.5, 15.0]$ and $C = [0.125, 0.250, 0.375, 0.500]$. A summary of the outcome of the GEFF fit for all these combinations can be found in Fig. 10. For clarity, we emphasize that, given the value of the mass threshold, we consider all the stars with $m > m_{\text{low}}$. In Fig. 10, going from right to left along the mass axis in each plot, it is possible to see the cumulative effect of including stars with lower masses. The different symbols represent different completeness thresholds, as indicated.

In the left-hand panel of the figure, we show the eccentricity $e = \sqrt{1 - (\frac{b}{a})^2}$. With this definition, the eccentricity values, $\eta = 0.15$ and 0.19, of Paper I become $e = 0.53$ and 0.59, respectively.
Figure 9. Example of surface density contours and a fit using the GEFF profile, in pixel coordinates on the frame. Left-hand panel: stars with \( m > 7.5 \, M_\odot \) and \( C > 0.375 \); completeness factors for the single stars are colour-coded. Right-hand panel: the corresponding density contours (in colours). The density is in arbitrary units and the colour-coding goes from the minimum to the maximum density. The grey area corresponds to the area where the average completeness for the whole cluster’s population is below 0.25; this area is masked out when performing the GEFF profile fit. The results of the fit are displayed as black elliptical contours.

Figure 10. Left-hand panel: eccentricity values as a function of the minimum mass; middle panel: semimajor axis values as a function of the minimum mass; the quoted \( k \) and \( q \) values are obtained by fitting functions as described in equation (8); and right-hand panel: tilt angle between the semimajor axis and \( x \)-axis, measured counterclockwise; each of the concentric rings helps to distinguish the different values as a function of the minimum mass used. Different symbols correspond to different completeness thresholds.

The eccentricity values are almost constant with mass. Their average values are somewhat higher than what found in Paper I, with \( \epsilon \approx 0.75 \) indicating an axial ratio \( a:b = 3:2 \). The fact that the numbers are slightly different compared to Paper I is not surprising. The adoption of a radially symmetric completeness correction in Paper I has partially smoothed out some of the asymmetry and intrinsic elongation of the cluster. Our new results clearly reveal the elongated two-dimensional density distribution of MS stars with masses between \( \approx 3 \) and \( \approx 30 \, M_\odot \), with higher elongation observed for lower mass stars. This might be related to the fact that more massive stars are also more centrally concentrated (see below); hence, their average collision time is shorter than that of the less-massive stars. Consequently, massive stars undergo more dynamical interactions and their momenta become more isotropic.

For completeness values between 0.125 and 0.5, we also performed a least-squares fit of the semimajor axis values, using the functional relation

\[
a = k \times [\log(m)]^q .
\]

The results of the fits are shown in the middle panel of Fig. 10. In general, the cumulative semimajor axis decreases almost like \( 1/\log(m) \). We already found an indication of mass segregation by investigating the IMF spatial variations (see Section 8). The finding here confirms that massive stars are more centrally concentrated. The figure also shows that by adopting a lower completeness threshold, the actual size of the semimajor axis decreases, at fixed mass. The reason is the inclusion of more and more stars in the centre of the cluster, where, of course, the total completeness is lower. Consequently, going to lower completeness thresholds, the stellar density has a more pronounced peak in the centre, while the density in the outer regions of the cluster does not change as much. Since \( a \) is a measure of the length-scale of the density decay with distance from
the centre, we obtain lower values of $a$ when the density contrast between the centre and the outskirts is more pronounced.

From Fig. 10 (right-hand panel), it is also very interesting to note that Wd 1’s direction of elongation lies very close to the galactic plane.

### 9.2 Possible sources of elongation

In the following, we carry out a qualitative discussion of possible sources of elongation. The typical orbital period for a star at 1 pc distance from a central point source with mass $M = 10^6 M_\odot$ is about $t_o \approx 3 \times 10^5$ Myr. This time-scale is much shorter than the typical half-mass relaxation time of about $t_h \approx 10^5 - 10^6$ yr, as determined for a typical globular cluster, comparable in stellar mass to Wd 1, and defined as the time required for the central half of the cluster mass to reach equilibrium (see section 1.1 in Spitzer 1987). Given the difference in the two time-scales, it is clear that the observed deviation from spherical symmetry cannot be ascribed to the global evolution of the cluster as an isolated system; after few orbits and encounters, the phase-space distribution of stars for an isolated system is expected to be isotropic (in $r$) and spherically symmetric (in $r$). The deviation from the spherical cluster shape must be explained either by unusual initial conditions still reflected in the present cluster appearance or by some interaction with the rest of the Galaxy.

A net angular momentum of the giant molecular cloud forming Wd 1 or a formation of Wd 1 out of two or more subclusters either might be responsible for its elongated shape. Differential Galactic rotation exerts a shear on molecular clouds which might lead to a net angular momentum. According to a recent study by Ballesteros-Paredes et al. (2009), galactic shear and tides have rather strong effects on initially elongated clouds, eventually quenching star formation and disrupting the clouds. Hence, differential rotation is an unlikely source for the elongation of Wd 1.

Another intriguing possibility would be a ‘hierarchical’ formation scenario, with merging of two or more smaller subclusters. The existence of a non-negligible fraction of possible binary clusters is supported by observations (see e.g. de la Fuente Marcos & de la Fuente Marcos 2008). A hierarchical organization of the interstellar medium and of young stellar groups and clusters is indeed observed on many scales (Elmegreen 2009). Negueruela et al. (2010) report the presence of a subclump of massive stars in the south-eastern region of Wd 1, even though they also warn that fluctuations in the star counts could be responsible for this observed clump. Clark et al. (2005) suggest that an age spread within the Wd 1 massive star population is really unlikely; hence, any possible merging or capture event must have happened in the very beginning of the cluster’s history. Otherwise, this episode could have also happened more recently, but under the condition that the subclusters are coeval, that is, star formation has occurred at almost the same time in different regions of the giant molecular cloud. If this scenario would be true, then the modest amount of dynamical crossing times occurred from Wd 1 formation (age $\approx 10^9$ yr) may be the reason why the stellar motion have not yet reached an isotropic distribution. No dynamical simulations that include such a large number of particles as Wd 1 members have been performed so far. Nevertheless, recent studies suggest that merging is indeed possible over a wide range of initial conditions (see Portegies Zwart & Rasio 2007; de la Fuente Marcos & de la Fuente Marcos 2010, and references therein).

We evaluate whether tidal effects in the Galactic central field could be responsible for the shape of Wd 1. Under the simplifying assumption of a circular restricted three-body problem, with primary mass at the Galactic Centre position and secondary at the cluster centre, we find that $r_l$, that is, the distance of the inner Lagrangian point from the Wd 1 centre is

$$r_l = \frac{1}{3} \frac{M_{\text{Wd1}}}{M_{\text{MW}}} R_0^3,$$

where $M_{\text{Wd1}} = 5 \times 10^4 M_\odot$ is our Wd 1 mass estimate, $M_{\text{MW}} = 6 \times 10^{11} M_\odot$ is the mass of the Milky Way and $R_0 = 4$ kpc is the Wd 1 galactocentric distance. With these numbers, we obtain $r_l = 12$ pc; given that $M_{\text{Wd1}} \ll M_{\text{MW}}$, this is also the distance for the external Lagrangian point from the Wd 1 centre. This estimate for the tidal radius is a lower limit, since it was assumed that all the mass of the Galaxy is confined within the orbit of Wd 1. A more correct estimate, taking into account only the enclosed Galactic mass at radius $R_0$, would lead to an even larger value of $r_l$. Consequently, it is clear that Wd 1 is far from filling its Roche lobe, while we measure elongation already on a scale of $\approx 1$ pc from the cluster core. Hence, tidal distortion from the whole Galaxy is unlikely the reason for the elongation of Wd 1.

Another important tidal effect could be caused by the Galactic disc gravitational field. The disc potential is constant far away from the Galactic mid-plane, where the matter density distribution of the disc drops to zero. On the contrary, within the disc, the potential has a non-zero divergence in the direction perpendicular to the plane ($z$-axis). This divergence causes a net acceleration of the stellar motions in the $z$-direction and, as a result, an initially spherical and isotropic cluster moving across the disc mid-plane is compressed along the $z$-direction. This phenomenon is known as ‘compressive gravitational shock’ and an analytical solution is presented in chapter 5 of Spitzer (1987). Unfortunately, the conditions that are required to apply this analytical treatment do not hold for Wd 1. In Spitzer (1987), the Galactic plane tidal field is treated as a fast time-dependent perturbation to the motion of stars within globular clusters, which cross the mid-plane at a speed of hundreds of km s$^{-1}$. Hence, the duration of the perturbation is short compared to the stellar orbital period around the cluster centre. However, Wd 1 is moving much slower in the $z$-direction. From a preliminary analysis of our Adaptive Optics multi-epoch observation, we can set a limit on the net bulk motion of $\lesssim 10$ km s$^{-1}$ along the Galactic longitude coordinate $b$ (Kudryavtseva et al., in preparation). For this reason, the Spitzer (1987) analytical solution cannot be used here, but it could be worth to investigate the effects of the disc tidal field in detail with dynamical simulations.

### 9.3 The effective cumulative radius

As an alternative to the study of mass segregation, in addition to the estimate of the GEFF best-fitting semimajor axis, $a$, we used an independent quantity that we call the effective cumulative radius:

$$r_{\text{eff}}(m) = \frac{\sum w_i > m \left( \frac{r_i}{C_i \times C_{\text{eff}}} \right)^2}{\sum w_i > m \left( \frac{1}{C_i \times C_{\text{eff}}} \right)^2}. \quad (9)$$

This quantity is obtained by taking all the stars with mass $m_i > m$ and computing their geometric-averaged distance from the centre of the cluster. The distance from the centre for the single stars is $r_i$; the completeness factors, $C_i$, are needed to take into account, in a statistical sense, the undetected sources.

Fig. 11 shows the quantity, $r_{\text{eff}}$, as a function of mass. Looking from the right-hand to left-hand side, it is evident that the inclusion of less-massive stars in the computation of $r_{\text{eff}}$ leads to an increase
in the average distance from the centre of the cluster, meaning that more-massive stars are located on average at smaller effective radii compared to less-massive stars. The first few points on the right-hand side of the panel do not follow this relation. This is due to the fact that the definition of an average distance for the most-massive stars is problematic, given the low numbers considered. Indeed, the most-massive star in our data set is a bit off-centre; hence, the effective radius for this star is quite large; this star also affects the effective radius of the first ~10 points on the right-hand side of the diagram, because its distance from the centre enters the computation of the average distance for the other stars. As long as the number of stars included in the computation of \(r_{\text{eff}}\) increases, the results converge towards a more stable averaged distance.

The smooth increase in \(r_{\text{eff}}\) with decreasing mass confirms the findings for the semimajor axis length of Section 9.1. Hence, we can state that Wd 1 is clearly mass segregated.

9.4 The origin of mass segregation for Wd 1

Mass segregation is a phenomenon observed in many young clusters; some examples are the Orion Nebula Cluster (Hillenbrand & Hartmann 1998), the NGC 3603 Young Cluster (Stolte et al. 2006; Harayama, Eisenhauer & Martins 2008) and the Arches Cluster (Stolte et al. 2005; Kim et al. 2006; Espinoza, Selman & Melnick 2009), and debate is still open whether this phenomenon is either primordial or due to dynamical evolution. If only two-body encounters are considered, the half-mass relaxation time for a cluster is given by (Binney & Tremaine 1987):

\[
t_{\text{rh}} = \frac{6.5 \times 10^8 \text{ yr}}{\ln(0.4N)} \left(\frac{M}{10^3 \text{ M}_\odot}\right)^{1/2} \left(\frac{1 \text{ M}_\odot}{(m)}\right)^{1/2} \left(\frac{r_\odot}{1 \text{ pc}}\right)^{3/2},
\]

where \(N\) is the total number of stars, \(m\) is the mean stellar mass and \(r_\odot\) is the deprojected half-mass radius equal to 4/3 of the projected half-mass radius. Considering \(N = 10^5\), \(M = 5 \times 10^4 \text{ M}_\odot\), \(r_\odot = 4/3 \times 1.1 \text{ pc}\) (see Paper I) and \(m = 0.6 \text{ M}_\odot\), we get an estimate of \(t_{\text{rh}} \approx 130\) Myr, much larger than the age of the cluster. This discrepancy between relaxation time and age is common to many clusters and has been used as an argument in favour of the primordial segregation scenario (Bonnell & Davies 1998). Nevertheless, one has to consider that the time for a star with mass \(m_*\) to drift towards the cluster centre due to dynamical friction is

\[
t_{\text{df}} = \frac{(m_*)}{m_\odot} \times t_{\text{rh}}.
\]

In the case of Wd 1, this segregation time can be as short as 2 per cent of \(t_{\text{rh}}\), that is, 2.6 Myr for a star of \(\sim 30\) \text{ M}_\odot, that is, the most-massive stars in our sample. Hence, primordial segregation would not be necessary to explain the observed mass segregation. Furthermore, the evolution of Wd 1 has probably been more complex and the value of \(t_{\text{rh}}\) might have changed in time. Mass-loss from stellar winds, supernova explosions and gas removal might have caused a global expansion of Wd 1, meaning that \(t_{\text{rh}}\) was shorter in the past (Gieles et al. 2010b). Gürkan, Freitag & Rasio (2004) and Portegies Zwart et al. (2004) showed that the core-collapse time for massive clusters is about 0.1–0.2\(t_{\text{rh}}\). It is therefore not unlikely that the core of Wd 1 has undergone a dynamical collapse, which is then also followed by expansion that could increase the relaxation time. This would push down the mass limit which we expect to be affected by mass segregation. McMillan, Vesperini & Portegies Zwart (2007) have proposed an alternative scenario to the primordial-segregation one, in order to explain mass segregation observed in young clusters. This scenario predicts that mass-segregated young, massive clusters could be the product of merging of several smaller subclusters. Substructure in molecular clouds is observed in both density and kinematics (Williams 1999; Williams, Blitz & McKee 2000), and this substructure is reflected as well in young clusters (Larson 1995; Testi et al. 2000; Gutermuth et al. 2005; Allen et al. 2007). In a hierarchical formation scenario, a massive cluster could be formed by the merging of several subclusters and still might show mass segregation. Given their smaller \(N\), the subclusters can rapidly reach a mass-segregated status before they merge and regardless of the initial spatial distribution of stars. The mass segregation in these smaller clusters is favoured by shorter \(t_{\text{rh}}\). In addition, Allison et al. (2009, 2010) show that clusters may undergo an initial collapse phase which can significantly accelerate mass segregation. In this phase, the cluster forms a very dense and clumpy core, where the massive stars can rapidly segregate, given the short crossing time and large number of encounters. The time-scale for such a process in a cluster with \(N \sim 10^5\) is \(\lesssim\) 1 Myr. The simulations by McMillan et al. (2007) additionally show that mass segregation developed by single subclusters is preserved during merging. Consequently, the final massive cluster exhibits mass segregation at an age much smaller than its global relaxation time. While merging might not be required to explain mass segregation, it could also explain the elongated shape of Wd 1. Hence, we think that this is a very interesting scenario for Wd 1 formation. We point out that McMillan et al. (2007) have carried out pure \(N\)-body simulations. Recently, Bate (2009) has performed hydrodynamic simulation of star-forming regions that include gas drag and gas accretion on to stars, in addition to the mutual gravity between them. The final cluster that is formed is the result of merging of five subclusters. The author finds no evidence for mass segregation. However, the number of stars formed in his simulations is of the order of \(10^4\) with the most-massive star of only \(\approx 5\) \text{ M}_\odot. Hence, the simulated cluster cannot be directly compared to Wd 1. More recently, using the final stellar positions of Bate (2009), and applying their own segregation detection method, Moeckel & Bonnell (2009) found evidence of segregation at least for the most-massive stars. The last scenario, in which only a few most-massive members are found in the cluster’s core, is more similar to what is observed for the Trapezium stars in the Orion Nebula Cluster than to what we observe in Wd 1, that is, an evidence of continuous mass segregation across the whole stellar mass spectrum.
10 CONCLUSIONS

We have presented a new, thorough analysis of near-IR data for the intermediate- and low-mass stellar populations of the massive young cluster Wd 1. Using artificial stars, we have been able to sample spatial variations of photometric completeness on a scale of few stellar FWHM. The same artificial stars have been used to infer realistic photometric error estimates, as well as the correlation between errors in different bands. Incompleteness corrections and errors were used to apply a novel statistical field subtraction technique to the data. Using a nearby control field, we obtained a clean sample of cluster members. The clean catalogue of stars, together with state-of-the-art stellar models, has been used to determine the cluster’s properties. We derived an extinction $A_{Ks} = 0.91 \pm 0.05$ mag, an age $t = 4 \pm 0.5$ Myr and a distance $d = 4.0 \pm 0.2$ kpc.

We investigated the cluster’s IMF slope using a new approach to stellar mass determination. The information on magnitude errors and their correlation has been used to derive the mass–probability distribution for each star, given the best-fitting isochrone. The completeness-corrected IMF has a slope of $\gamma = 2.44^{+0.20}_{-0.08}$ slightly steeper than the Salpeter or Kroupa IMF; this slight discrepancy could be partially reconciled if we consider that, for the sake of simplicity, we are neglecting the influence of (unknown) undetected binaries; hence, our quoted error is probably an underestimate of the total, statistical plus systematic, error (Maíz Apellániz 2009). From the IMF slope and its normalization constant, we found a total mass for the cluster of $M_{Wd1} = 4.91^{+0.70}_{-0.49} \times 10^4 M_\odot$.

The spatially varying completeness, combined with the probabilistic mass determination, enabled us to investigate the spatial variations of the IMF. The Wd 1 starburst cluster is mass segregated, with massive stars more centrally concentrated. Other indications of mass segregation come from the analysis of the stellar density distribution. In order to study the two-dimensional density distribution as a function of stellar mass, we fitted two-dimensional elliptical distributions. In order to study the two-dimensional density distribution, we fitted two-dimensional elliptical profiles. This analysis revealed a tight dependency of the ellipse’s semimajor-axis length on mass: $a(m) \propto 1/\log(m)$. Given the young age of Wd 1, its global mass segregation cannot be explained by simple two-body relaxation. Interestingly, from the density distribution analysis, we found that Wd 1 is elongated along the Galactic plane with an axial ratio $a/b = 3.2$. The mass segregation and the elongation together hint at a formation scenario involving the merging of multiple subclusters formed almost coevely in the parental giant molecular cloud.

ACKNOWLEDGMENTS

The authors would like to thank E. Tognelli (Università’ di Pisa, Pisa, Italy) for his support in the calculation of the Pisa stellar models. We made large use of the IDL, programming language, adopting The IDL Astronomy User’s Library, the Coyote Library and the mpfit package. MG would like to thank the developers and those who maintain these very useful tools for making them publicly available. MG would also like to thank M. Fang, B. Vaidya and R. Andrae (all at Max-Planck-Institut für Astronomie, Heidelberg, Germany) for useful discussions.

REFERENCES

Allen L. et al., 2007, in Reipurth B., Jewitt D., Keil K., eds, Protostars and Planets V. University of Arizona Press, Tucson, p. 361
Allison R. J., Goodwin S. P., Parker R. J., de Grijs R., Portegies Zwart S. F., Kouwenhoven M. B. N., 2009, ApJ, 700, L99
Allison R. J., Goodwin S. P., Parker R. J., de Grijs R., Portegies Zwart S. F., de Grijs R., 2010, MNRAS, 407, 1098
Andersen M., Zinnecker H., Moneti A., McCaughey M. J., Brandl B., Brandner W., Meylan G., Hunter D., 2009, ApJ, 707, 1347
Andrae R., 2010, preprint (arXiv:1009.2755)
Ballesteros-Paredes J., Gómez G. C., Richiardo B., Vázquez-Semadeni E., 2009, MNRAS, 393, 1563
Bate M. R., 2009, MNRAS, 392, 590
Binney J., Tremaine S., 1987, Galactic Dynamics. Princeton Univ. Press, Princeton, NJ
Bonnett I. A. D., Davies M. B., 1998, MNRAS, 295, 691
Brandner W., Clark J. S., Stolte A., Waters R., Negueruela I., Goodwin S. P., 2008, A&A, 478, 137 (Paper I)
Brott I., Hauschildt P. H., 2005, in Turon C., O’Flaherty K. S., Perryman M. A. C., eds, The Three-Dimensional Universe with Gaia. ESA Publications Division, Noordwijk, p. 565
Campbell M. A., Evans C. J., Mackey A. D., Gieles M., Alves J., Ascenso J., Bastian N., Longmore A. J., 2010, MNRAS, 405, 421
Castelli F., Kurucz R. L., 2003, in Piskunov N., Weiss W. W., Gray D. F., eds, Proc. IAU Symp. 210, Modelling of Stellar Atmospheres. Astron. Soc. Pac., San Francisco, p. A20
Clark J. S., Negueruela I., Crowther P. A., Goodwin S. P., 2005, A&A, 434, 949
Crowther P. A., Hadfield L. J., Clark J. S., Negueruela I., Vacca W. D., 2006, MNRAS, 372, 1407
de la Fuente Marcos R., de la Fuente Marcos C., 2008, ApJ, 672, 342
de la Fuente Marcos R., de la Fuente Marcos C., 2010, ApJ, 719, 104
Degl’Innocenti S., Prada Moroni P. G., Marconi M., Ruoppo A., 2008, Ap&SS, 316, 25
Devillard N., 2001, in Harned F. R., Jr, Primini F. A., Payne H. E., eds, ASP Conf. Ser. Vol. 238, Astronomical Data Analysis Software and Systems X. Astron. Soc. Pac., San Francisco, p. 525
Efron B., 1979, Ann. Stat., 7, 1
Eisenhauer F., Quirrenbach A., Zinnecker H., Genzel R., 1998, ApJ, 498, 278
Elmegreen B. G., 2009, in Richtler T., Larsen S., eds, Globular Clusters – Guides to Galaxies, Hierarchical Formation of Galactic Clusters. Springer Verlag, Berlin, p. 87
Elson R. A. W., Fall S. M., Freeman K. C., 1987, ApJ, 323, 54 (EFS87)
Esposino P., Selman F. J., Melnick J., 2009, A&A, 501, 563
Fleck J., Boily C. M., Lancón A., Deiters S., 2006, MNRAS, 369, 1392
Gieles M., Sana H., Portegies Zwart S. F., 2010a, MNRAS, 402, 1750
Gieles M., Baumgardt H., Heggie D. C., Lamers H. J. G. L. M., 2010b, MNRAS, 408, L16
Gürkan M. A., Freitag M., Rasio F. A., 2004, ApJ, 604, 632
Gutermuth R. A., Megeath S. T., Pipher J. L., Williams J. P., Allen L. E., Myers P. C., Rains S. N., 2005, ApJ, 632, 397
Harayama Y., Eisenhauer F., Martins F., 2008, ApJ, 675, 1319
Hartig T., Tibshirani R., Friedman J., 2009, The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science+Business Media, New York
Hillenbrand L. A., Hartmann L. W., 1998, ApJ, 492, 540
Kim S. S., Gill F. D., Kudritzki R. P., Najjaro F., 2006, ApJ, 653, L113
Kothes R., Dougherty S. M., 2007, A&A, 468, 993
Kouwenhoven M. B. N., de Grijs R., 2008, A&A, 480, 103
Kroupa P., 2001, MNRAS, 322, 231
Larson R. B., 1995, MNRAS, 272, 213
Lejeune T., Schaerer D., 2001, A&A, 366, 538
Mamajek E. E., 2008, A&A, 494, 1049
McMillan S. L. W., Vesperini E., Portegies Zwart S. F., 2007, ApJ, 655, L45
Maíz Apellániz J., 2009, Ap&SS, 324, 95
Marigo P., Girardi L., Bressan A., Groenewegen M. A. T., Silva L., Granato G. L., 2008, A&A, 482, 883
Mengel S., Tacconi-Garman L. E., 2009, Ap&SS, 324, 321
Meynet G., Maeder A., 2000, A&A, 361, 101
Moeckel N., Bonnell I. A., 2009, MNRAS, 400, 657
Muno M. P., Law C., Clark J. S., Dougherty S. M., de Grijs R., Portegies Zwart S., Yusuf-Zadeh F., 2006, ApJ, 650, 203
Negueruela I., Clark J. S., 2005, A&A, 436, 541
Mass segregation and elongation of Wd 1 2483
APPENDIX A: COMPLETENESS MAPS

In this appendix, we will illustrate, step by step, how the two-dimensional completeness maps for Wd I were obtained.

A1 Adding and detecting artificial stars

The basic idea is to use the same PSF as was obtained by PSF fitting with DAOPHOT to add stars (using the DAOPHOT addstar task) in the reduced images and then run the same PSF-fitting photometry scheme to see whether artificial stars can be recovered or not. 50 stars per run were added, in order not to change the crowding characteristic of the frame. Stars are positioned randomly on the frame and have a uniform distribution in magnitude. To achieve sufficient spatial resolution, we iterate the procedure until we have added 4500 stars per filter.

With an effective detector area of eff x eff = (Nobs/2225) ≈ 8 pixel^2, it means that we sample the whole frame on a scale which is about 3.5 times the PSF FWHM. The effective sampling scale is a bit larger, due to the use of a certain number of nearest neighbours to calculate the local value of completeness at the position of each simulated star. The natural limit, that is, the minimum length-scale at which completeness can be sampled by our method, is the FWHM of the PSF itself, which characterizes the ability to distinguish two different point sources. The resolution of the incompleteness map could not be improved further, even if the number of simulated stars would be increased in order to obtain a spatial sampling smaller than the PSF FWHM. Our choice of the total number of stars is a compromise between a short sampling scale and a reasonable number of simulations.

A2 Building the two-dimensional maps

Each simulated star can be either recovered or not by DAOPHOT PSF fitting, meaning that for that specific star, completeness is either 0 or 1. Starting from this series of sparsely sampled 0s and 1s, several steps are necessary to obtain a smooth function, which is determined at each point on the frame. In the following, we will indicate the position of simulated stars with a hat symbol, (x, y), while the coordinate grid on which we actually calculated the function will be simply (x, y), which corresponds to the pixels grid of the chip. We will refer in this section only to a single magnitude bin and to a single photometric band; interpolation in the magnitude dimension will be treated in Section A3.

The first step is to create average completeness values at each (x_i, y_i) for i = 1 and N_{sim}. This is accomplished by considering a certain number, v, of nearest neighbours to the i-th simulated star and defining the completeness fraction as

\[ C_v(x_i, y_i) = \frac{\text{recovered stars}}{\text{v} + 1} \]  

where the ‘recovered stars’ are counted among the v neighbours of the i-th simulated star, which is also included, hence +1 in the denominator. The actual value of v is somewhat arbitrary and has to satisfy two opposite requirements: the higher it is, the less the completeness values will be affected by statistical noise. On the other hand, a too large value would imply a loss in spatial resolution for our completeness maps. As mentioned in Section A1, the effective sampling scale is not the (d) of equation (A1) but more precisely

\[ d_{eff} = (d) \times \sqrt{v} \]  

After several experiments, we decided to use v = 16, which degrades our completeness sampling scale by a factor of 4, giving \[ d_{eff} \approx 43.2 \text{ pixels} \], corresponding to about 14 times the FWHM of the image PSF.

At this stage, the \[ C_v \] is known only point-wise in the set of (x, y) positions occupied by the simulated stars. The next step is to interpolate this function into a regular grid of points. This is accomplished via the IDL procedure GRIDDATA, using the Kriging method of interpolation with an exponentially decreasing model for the variogram. Kriging allows to interpolate a random field known in a set of positions into another set, under some assumptions about its...
covariance. In our case, the random field is the completeness itself, with its Poissonian error due to the finite number of simulated stars considered in equation (A2). An exponential model for the covariance is appropriate here, because the estimated values of $C_0$ at some location $(\tilde{x}, \tilde{y})$ are correlated with those for other stars and the correlation is stronger for closer simulated stars than for those separated by a large distance. We have chosen an e-folding scale equal to $d_{\text{eff}}$.

After the interpolation, we performed a smoothing of the completeness. The grid used for the interpolation is indeed finer than $d_{\text{eff}}$, meaning that the interpolated function may show artificial variations on a scale smaller than our minimum size, which would be unrealistic. That is why we additionally smoothed the maps with a boxcar kernel with a size of $d_{\text{eff}}$. The boxcar model is appropriate, given the uniform spatial distribution of simulated stars.

**A3 Interpolation in magnitude**

Once the maps are available in magnitude layers, we enforced that at each location, $(x, y)$, completeness is a decreasing function of magnitude. We fitted pixel by pixel a monotonically decreasing function of Fermi-like type:

$$C_i(x, y) = \frac{\alpha(x, y)}{e^{\frac{\alpha(x, y)}{\mu(x, y)}} + 1}.$$  \hspace{1cm} (A3)

The meaning of the three coefficients is as follows:

(i) $\alpha$ is the normalization and is always $\leq 1$.

(ii) $\beta$ is the magnitude for which completeness is $\alpha/2$.

(iii) $\gamma$ represents the rapidity with which $C_i$ drops down.

Once the ($\alpha, \beta, \gamma$) coefficient triplets are calculated, it is straightforward to assign to each real star its completeness value using the coefficients evaluated at the star’s position.

**A4 Completeness for the control field**

The offset field that we used for field decontamination of the CMD is also affected by incompleteness. In this case, however, it is not necessary to investigate the two-dimensional structure of the completeness pattern; under the assumption that the spatial distribution of the stars in the control field is uniform, we only consider spatially uniform incompleteness correction.

When using a control field for decontamination of a star cluster’s CMD, one implicitly assumes that the stars are, on average, representatives of the foreground/background population in the cluster’s field. This assumption has a series of shortcomings. For example, the copious cluster population itself may partially ‘shield’ background stars. In addition to that, in the Galactic plane, variable extinction may cause differences in the observed population of stars. Furthermore, the population along different lines of sight could be intrinsically different, due to the different Galactocentric distances sampled at the same heliocentric distance or changes within the spiral arms. The two latter problems are reduced by choosing nearby fields, so that the foreground/background populations show similar distributions in age and extinction — hence in magnitude and colour — along the cluster’s line of sight. Hence, the choice of the control field is done in order to have a population that on average looks like the contaminating population in the cluster frame.

For these reasons, it is not necessary to try the same two-dimensional approach to assign completeness values to the offset cluster frame stars. We only populated the whole frame in a uniform way with 250 stars per 0.5-mag-wide bin; only 50 stars were added in each run, not to alter the crowding characteristics of the field. Then, we computed the fraction of recovered stars over the total number of simulated ones and fitted a function like that of equation (A3), this time without any spatial dependence. The last step was to assign the single stars in the control field their corresponding completeness values in each photometric band.

**APPENDIX B: EVALUATION OF THE PHOTOMETRIC ERRORS AND THEIR CORRELATIONS**

In Paper I, we used simulated stars to estimate the photometric errors. We showed that for stars with known input magnitudes, the output magnitudes were often in disagreement at a level of more than 1σ, where the the `daophot` fitting errors were taken as $\sigma$ values. Hence, the difference between input and output magnitudes seemed to be a more conservative and robust estimate of the real photometric error. Simulated stars are also used here to estimate the correlation between magnitude errors. The photon counts associated to an isolated star in two different bands are uncorrelated. In reality, even though the photon counts are independent, the inferred magnitude values may not be. The reason why $J$ and $K_s$ magnitude errors are correlated is the presence of bright stars or, more generally, crowding. When a faint star is located close to a bright star, the residuals of the PSF-fitting procedure of the bright star (which is usually bright in both bands) may lead to magnitude errors in both bands. If the bright star’s wings are not properly subtracted, then there will be an excess in the flux that is assigned to the nearby faint star. The bright star’s wings may also be oversubtracted (e.g. because the core is not well fitted), leading to too small flux estimates. This can lead to a correlation of the photometric errors. Crowding from stars of comparable magnitudes will lead to a similar behaviour.

That $J$ and $K_s$ magnitude errors are correlated is obvious from Fig. B1. In the left-hand panel, we show $J^{\text{out}} - J^{\text{in}}$ versus $K_s^{\text{out}} - K_s^{\text{in}}$ for the simulated stars. In the right-hand panel, we show $(J - K_s)^{\text{out}} - (J - K_s)^{\text{in}}$ versus $K_s^{\text{out}} - K_s^{\text{in}}$. Since the two magnitude estimates are correlated, the composed quantity $J - K_s$ is also correlated to the single-magnitude values. The coefficient $r$ in the figures is Pearson’s correlation coefficient for the whole sample of simulated stars, that is,

$$r_{X,Y} = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2} \sqrt{\sum_i (Y_i - \bar{Y})^2}}.$$  \hspace{1cm} (B1)

where $X$ and $Y$ are the respective abscissa and ordinate in the two panels. From its definition, it is clear that Pearson’s $r$ is equivalent to
the X and Y covariance divided by the product of X and Y standard deviations. A value of r very close to +1 (-1) indicates a very tight correlation (anticorrelation) between the two random variables, while two totally uncorrelated random variables would show a value of r = 0. The values of Pearson’s coefficients in Fig. B1 indicate a quite significant correlation of the magnitude errors as well as an even tighter anticorrelation between magnitude and colour errors. Given that r is not a robust, outlier-resistant quantity, the actual values were calculated by removing the outliers, that is, stars for which input and output magnitudes differ more than 1 mag in at least one band. The number of stars inside this limit is 97 per cent of the total number of simulated stars. Therefore, the exclusion of the outliers does not represent a shortcoming in the evaluation of a robust estimate for the overall correlation coefficient of the sample.

### B1 Assigning proper photometric errors and their correlation to each detected star

For each detected star, we selected at least seven simulated stars (using the same stars of Section A1) that were positioned in its neighbourhood. By neighbours we mean simulated stars whose distance from the position of the real star was not larger than 50 pixels and whose magnitude differs by no more than 1 mag – in each band – from that of the real star. The distance of 50 pixels represents the average radius of influence of the bright stars, that is, the typical extent of their haloes, as derived by the analysis of Wd 1 NTT/SoFi images. For each of the neighbours in the subsample, we calculated \( J^{\text{out}} - J^{\text{in}} \) and \( K^{\text{out}} - K^{\text{in}} \). The standard deviations of the two quantities, within the subsample, have been used as estimates for the photometric errors of the real stars. We also calculated Pearson’s coefficient between the two quantities in the neighbours’ subsample and assigned it to the detected star. The comparison between DAOPHOT errors and our newly estimated errors for the Wd 1 field are shown in Fig. B2. The new error estimates are, on average, larger than what predicted by DAOPHOT, especially in the \( K_s \) band.

Some of the real stars do not have a sufficient number of neighbours to perform this kind of estimate. This is true especially for faint stars, since the majority of the simulated stars at the faintest magnitudes cannot be recovered. Hence, these simulated stars cannot be used for the error estimate, because they do not have an \( M^{\text{out}} \) value. For stars without enough useful simulated neighbours, we used a different error estimate. We first divided the real stars for which the error determination worked fine in several magnitude bins. Then we calculated the mean error per magnitude bin and fitted an exponential relation to the mean error versus bin-magnitude points. This relation has been used to assign their errors to the stars that lack a sufficient number of neighbours. The new errors, as a function of the stars’ magnitudes, are shown in Fig. 3. The figure shows the exponential extrapolation used to determine the errors of the faintest stars. Together with the new errors for the cluster’s field stars, we show, in red, the new errors for the control field stars, whose derivation is illustrated in Section B2.

### B2 Photometric errors for the control field

A similar method was used to derive new photometric errors for the control field stars. Since we assume that these stars are uniformly distributed, there is no need to treat the spatial variation of the errors. Using the same simulated stars as in Section A4, we computed the \( \Delta M_j(i) = M_j^{\text{sim}}(i) - M_j^{\text{gal}}(i) \) for \( M_j = J, K_s \) and \( i \) running over the simulated stars; we then binned the stars in 0.5-mag-wide bins (in the input magnitudes) and for each bin we computed the standard deviation of \( \Delta M_j \) over the bin. The last step was to fit the \( \sigma(\Delta M_j, M_j[\text{bin}]) \) points with an exponential relation; here \( M_j[\text{bin}] \) is the central magnitude of the bin. This relation was used to assign an error to the real stars as a function of magnitude. The average value of the correlation between \( J \) and \( K_s \) was calculated for the whole sample and is \( r = 0.25 \). This value was assigned to each real star in the control field.

Fig. 3 shows that, on average, the photometric errors in the control field are smaller than those in the cluster’s field. This behaviour is expected and can be explained by the higher degree of crowding for Wd 1’s field. For the same reason, the detection limit for the control field is \( \sim 0.5 \) mag fainter than the Wd 1 field in both photometric bands.

### APPENDIX C: \( \sigma \)-CLIPPING

Because of some dissimilarities between the on-field and off-field foreground/background populations, the CMD of Wd 1, even after subtraction, does not look perfectly clean. For this reason, after having chosen the best-fitting isochrone, that is, the 4-Myr one, before any further analysis, we additionally subtracted those stars that lie more than 3\( \sigma \) away from the reference isochrone in the magnitude–magnitude space (see at the end of Section 5). After clipping, essentially all stars with colours and magnitudes consistent with the 4-Myr cluster population are included in the final source selection. Our clipping criteria may retain some arbitrariness; nevertheless, they do not affect our further analysis. The main reason is that the cut only affects the faint stars, with large photometric errors. Some of them could be excluded or included in the catalogues by slightly changing the \( \sigma \) threshold. Anyway, in the computation of the IMF (see Section 8) and of the stellar density (see Section 9), we only consider stars above a given completeness or mass threshold. Stars with uncertain membership are mostly excluded by these two additional cuts; hence, they do not affect the final results.

One realization of the clean cluster’s CMD is shown in Fig. 5, together with the best-fitting isochrone. The error bars shown in the diagram are the average \( J - K_s \) and \( K_s \) errors per magnitude bin. The colour errors for each star are calculated as

\[
\delta(J - K_s) = \sqrt{\sigma^2(J) + \sigma^2(K_s) + 2r_{JK} \sigma(J) \sigma(K_s)}.
\]  

(C1)

Pearson’s \( r \) is equal to the covariance divided by the product of the two standard deviations (see equation B1); hence, the third addendum on the right-hand side of equation (C1) is equal to twice the covariance of \( J \) and \( K_s \).

© 2011 The Authors, MNRAS 412, 2469–2488
Monthly Notices of the Royal Astronomical Society © 2011 RAS
Table C1. Detections in the on-field and off-field frames.  

| Field                           | Number of stars |
|--------------------------------|-----------------|
| On                             | 7036            |
| Off                            | 5381            |
| On (after subtraction)         | 5810 ± 25       |
| On (after subtraction and σ-clipping) | 4300 ± 23     |

In Table C1, we summarize the number of stars left after field subtraction and additional clipping. The mean values and their uncertainties are derived by iterating the probabilistic subtraction technique. We repeated the extraction of ζ for each star to generate 100 different catalogues (see Section 5). We then calculated mean and standard deviations of the number of members over the 100 samples.

APPENDIX D: BOOTSTRAP ESTIMATE OF THE IMF PARAMETERS AND THEIR ERRORS

Bootstrapping is a resampling technique for error estimation (see e.g. Efron 1979; Hastie, Tibshirani & Friedman 2009; Andrae 2010). Given a data set from which some parameters are estimated, bootstrapping consists in resampling the data to create alternative data sets. From these, it is possible to repeatedly estimate the parameters of interest, monitoring their distribution. We generated $10^5$ bootstrap samples to probe the parameter space of $(\gamma, A)$, assuming for the IMF the functional form $\frac{dN}{dm} = A \times m^{-\gamma}$. From our data set, we created 100 different realizations of the member catalogues. Each catalogue has a slightly different number of members $N_{c,j}$ with $j = 1, 100$, after statistical field subtraction and σ-clipping (see Appendix C). From the members of each jth catalogue, 1000 bootstrap samples were created. The new samples consist of the same number of stars as in the member catalogue, $N_{c,j}$, but the draw is made with replacement, that is, the same star can occur multiple number of times in a bootstrap sample. This sample of stars is then used to build the IMF as in equation (4) where now $i$ runs on the stars of the specific bootstrap sample. At each iteration, a power-law fit is performed to obtain a couple $(\gamma_{j,k}, A_{j,k})$ with $j = 1, 100$ and $k = 1, 1000$. As already detailed in Section 8.2, the fitting interval is restricted to $m \in [3.5, 27] \, M_\odot$. Given the $(\gamma_{j,k}, A_{j,k})$ values, we obtained the corresponding total mass, $M_{j,k}$, and total number of stars, $N_{j,k}$, by integrating the power law in the interval $m \in [0.08, 120] \, M_\odot$.

A two-dimensional density plot of the output values $(\gamma, A)$ is shown in Fig. D1. It is clear that the $\gamma$ and $A$ parameters are tightly correlated. This is easy to understand. For each bootstrap sample, we have a number, $N_{\text{fit}}$, of stars that are actually inside the fitting interval. Given the different catalogue realizations, this number can be slightly different, but is mostly in the interval $[1250, 1500]$. The IMF fit has to satisfy the condition

$$N_{\text{fit}} = A \times \int_{3.5}^{27} m^{-\gamma} \, dm.$$
from which we get
\[ A = \frac{N_{\text{fit}}(1 - \gamma)}{27^{1-\gamma} - 3.5^{1-\gamma}}. \]

This relation between \( A \) and \( \gamma \) is overplotted in Fig. D1 for \( N_{\text{fit}} = 1000, 1250, 1500 \) and 1750 (dotted lines). Given that the two-dimensional distribution of \((\gamma, A)\) pairs is clearly non-Gaussian, the definition of the best values and the confidence intervals for the two parameters is not straightforward. The maximum of the two-dimensional distribution is located at \((\gamma_{\text{max2D}}, A_{\text{max2D}}) = (2.46, 1.31 \times 10^4)\). With this pair of values, we obtain a total mass, \( M_{\text{max2D}} = 5.13 \times 10^4 \, \text{M}_\odot \), and a total number, \( N_{\text{max2D}} = 1.10 \times 10^5 \) stars.

Conversely, using the two-dimensional joint distribution is not the most suitable choice for defining the best values and confidence interval for the parameters \((\gamma, A)\), and for \( M_{\text{tot}} \) and \( N_{\text{tot}} \). For this purpose, in the case of \( \gamma \) and \( A \), we used the marginal distributions. These are obtained by integration of the joint distribution with respect to the other variable. Similarly, for \( M_{\text{tot}} \) and \( N_{\text{tot}} \), we used the distributions of \( M_{j,k} \) and \( N_{j,k} \) obtained after each bootstrap iteration. The best values are obtained by maximizing the distributions. The confidence intervals are obtained by integrating the distributions from the left and from the right until 16 per cent of the total area under the distribution is reached on each side. This means that the limits of the asymmetric confidence interval comprise 68 per cent of the total area under the distribution function. The marginal distributions for \( \gamma \) and \( A \), as well as the distributions for \( M_{j,k} \) and \( N_{j,k} \), are shown in Fig. D2. The best values and the confidence intervals are given in Table D1.

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

### Table D1. Best values and their confidence intervals for the IMF parameters, the total mass and the total number of stars of Wd 1.

| Quantity       | Best value | Lower limit | Upper limit |
|----------------|------------|-------------|-------------|
| \( \gamma \)   | 2.44       | 2.36        | 2.64        |
| \( A/10^4 \)   | 1.22       | 1.08        | 1.78        |
| \( M_{\text{tot}} (10^4 \, \text{M}_\odot) \) | 4.91       | 4.42        | 6.70        |
| \( N_{\text{tot}}/10^4 \) | 10.4       | 8.6         | 16.4        |