

OCAAT: automated analysis of star cluster colour-magnitude diagrams for gauging the local distance scale

Gabriel I. Perren¹, Ruben A. Vázquez¹, Andrés E. Piatti²
and André Moitinho³

¹Facultad de Ciencias Astronómicas y Geofísicas (UNLP)
IALP-CONICET, La Plata, Argentina
email: gperren@fcaglp.unlp.edu.ar

²Observatorio Astronómico, Universidad Nacional de Córdoba
Córdoba, Argentina

³SIM, Faculdade de Ciências da Universidade de Lisboa
Ed. C8, Campo Grande, 1749-016 Lisboa, Portugal

Abstract. Star clusters are among the fundamental astrophysical objects used in setting the local distance scale. Despite its crucial importance, the accurate determination of the distances to the Magellanic Clouds (SMC/LMC) remains a fuzzy step in the cosmological distance ladder. The exquisite astrometry of the recently launched ESA Gaia mission is expected to deliver extremely accurate statistical parallaxes, and thus distances, to the SMC/LMC. However, an independent SMC/LMC distance determination via main sequence fitting of star clusters provides an important validation check point for the Gaia distances. This has been a valuable lesson learnt from the famous Hipparcos Pleiades distance discrepancy problem. Current observations will allow hundreds of LMC/SMC clusters to be analyzed in this light.

Today, the most common approach for star cluster main sequence fitting is still by eye. The process is intrinsically subjective and affected by large uncertainties, especially when applied to poorly populated clusters. It is also, clearly, not an efficient route for addressing the analysis of hundreds, or thousands, of star clusters. These concerns, together with a new attitude towards advanced statistical techniques in astronomy and the availability of powerful computers, have led to the emergence of software packages designed for analyzing star cluster photometry. With a few rare exceptions, those packages are not publicly available.

Here we present OCAAT (Open Cluster Automated Analysis Tool), a suite of publicly available open source tools that fully automatise cluster isochrone fitting. The code will be applied to a large set of hundreds of open clusters observed in the Washington system, located in the Milky Way and the Magellanic Clouds. This will allow us to generate an objective and homogeneous catalog of distances up to ~ 60 kpc along with its associated reddening, ages and metallicities and uncertainty estimates.

Keywords. galaxies: star clusters, methods: data analysis, methods: statistical, cosmology: distance scale

1. Introduction

The aim of OCAAT is to bring together many of the algorithms which are usually applied on open clusters (OCs) throughout the literature, in a single easy to use package. OCAAT is able to assign a precise center, calculate a limiting radius, reject stars based on their errors, obtain the cluster's luminosity function and its integrated color, assign a probability of the cluster being a real system and not a random overlapping of field stars, assign membership probabilities for each star within the cluster region and

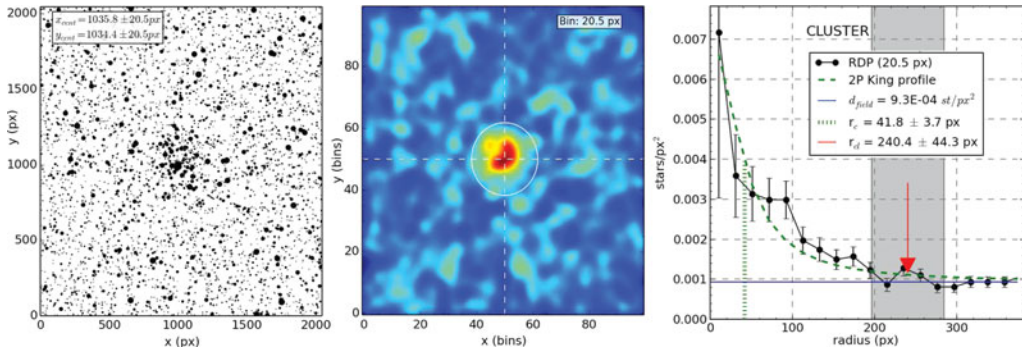


Figure 1. *Left:* Finding chart of synthetic cluster. *Center:* Center determination via a KDE spatial density function. *Right:* radius assignment, obtained as the value where the RDP reaches the field density level.

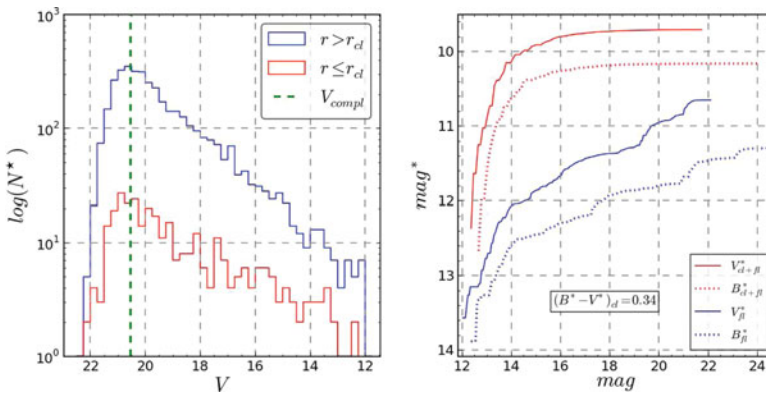


Figure 2. *Left:* Cluster and average field luminosity functions curves in red and blue respectively. *Right:* Integrated B and V magnitudes for the cluster region and averaged field regions.

estimate the cluster’s parameters: metallicity, age, distance and extinction along with their uncertainties.

2. Functions description

Structural parameters. Traditionally, centers are inferred by eye after the identification of a visible spatial overdensity, seldom refined applying a simple histogram-based search for the maximum concentration in each axis separately. In OCAAT we employ a 2-dimensional Gaussian kernel density estimate (KDE) function applied to the entire region and the location of its maximum value. Fig. 1 shows to the left a field where a synthetic young cluster is located and to the center its center assignment. The radius is assigned as the value where the radial density profile (RDP) reaches the level of the field density contribution d_{field} , as shown to the left. The error is obtained as the standard deviation of the radius values estimated using different conditions to establish when the RDP has stabilized around d_{field} (grey area in the plot). A 3-parameter King profile is fitted to obtain the core and tidal radius, if the process does not converge then a 2-parameter fit is attempted to derive the core radius only.

Errors, LF and integrated color. A curve composed of a polynomial plus an exponential fit is used as an upper envelope for the error distribution in magnitude. Stars that are

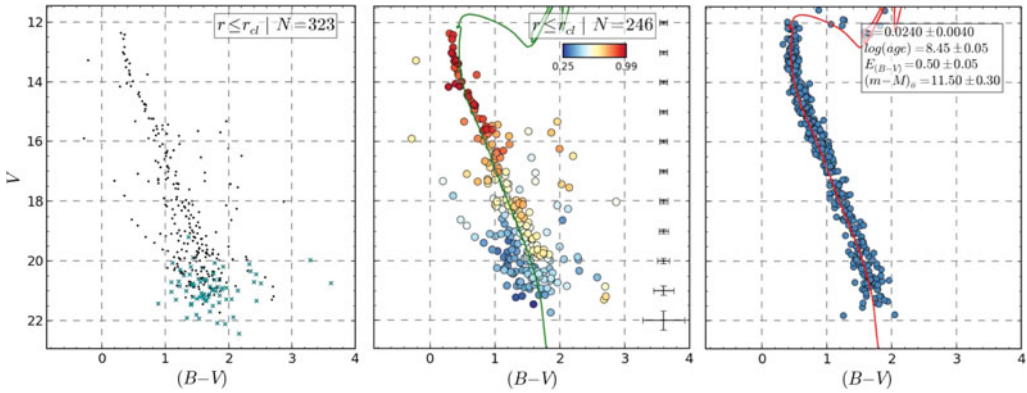


Figure 3. *Left:* CMD of synthetic cluster showing rejected stars (ie: stars above the maximum error limit) as green crosses. *Center:* CMD of the cluster region colored according to the membership probabilities. *Right:* Best fit synthetic cluster created from the red isochrone shown.

located above this curve or above a maximum error threshold in magnitude or color are rejected. The luminosity function curve is obtained for both the cluster region and the field region (scaled to the cluster's area) and the same is done to calculate the integrated color of the stars within the cluster region. Fig. 2 shows these curves for the cluster and field regions.

Cluster KDE probability. A KDE-based algorithm is used to assess the similarity between the arrangement of stars in the cluster region CMD versus CMDs of the field regions, quantified by p-values. The curves formed by these cluster vs field and field vs field region p-values overlap an amount proportional to the probability that the cluster and average field region were drawn from the same distribution. An overlap value close to 1 points to a high probability of the system being a true cluster.

Decontamination algorithm. A Bayesian algorithm inspired by the one presented in Cabrera-Cano & Alfaro (1990) is implemented to assign membership probabilities to stars within the cluster region, based on photometry data alone. Fig. 3 shows to the left all stars inside the defined cluster radius and at the center only those stars below the error cut-off, colored by the membership probabilities assigned by the code. Higher values are usually associated with brighter stars due to the usual field star contamination in the faint region of the CMD.

Synthetic cluster fit. Here is where we obtain the cluster distances. The membership probabilities calculated by the decontamination algorithm are used as weights in a likelihood equation that identifies the best observed vs synthetic cluster fit, based on the one presented in Hernandez & Valls-Gabaud (2008). Each synthetic cluster is generated from a theoretical isochrone accounting for errors, binarity and completeness effects. The search for the synthetic cluster which minimizes the likelihood can be done either via a brute force algorithm or a genetic algorithm (GA). In Fig. 3 (right) the best synthetic cluster found by the GA can be seen, with the underlying theoretical isochrone shown in red (same isochrone shown in green in the center plot) The uncertainties are estimated applying a bootstrap process.

References

- Cabrera-Cano, J. & Alfaro, E. J. 1990, *A&A*, 235, 94
 Hernandez, X. & Valls-Gabaud, D. 2008, *MNRAS*, 383, 1603