Astronomy & Astrophysics

Mapping the stellar age of the Milky Way bulge with the VVV[*]

I. The method

F. Surot[1,4], E. Valenti[1], S. L. Hidalgo[2,3], M. Zoccali[4,5], E. Sökmen[2,3], M. Rejkuba[1,6], D. Minniti[5,7,8],
O. A. Gonzalez[9], S. Cassisi[10,11], A. Renzini[12], and A. Weiss[13]

[1] European Southern Observatory, Karl Schwarzschild-Straße 2, 85748 Garching bei München, Germany
  e-mail: frsurot@uc.cl
[2] Instituto de Astrofísica de Canarias, Vía Láctea S/N, 8200 La Laguna, Tenerife, Spain
[3] Department of Astrophysics, University of La Laguna, 38200 La Laguna, Tenerife, Spain
[4] Instituto de Astrofísica, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna, 4860 Santiago, Chile
[5] Millennium Institute of Astrophysics, Av. Vicuña Mackenna 4860, 782-0436 Macul, Santiago, Chile
[6] Excellence Cluster Universe, Boltzmann-Straße 2, 85748 Garching bei München, Germany
[7] Departamento de Ciencias Físicas, Universidad Andrés Bello, República 220, Santiago, Chile
[8] Vatican Observatory, 00120 Vatican City, Italy
[9] UK Astronomy Technology Centre, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
[10] INAF – Astronomical Observatory of Abruzzo, Via M. Maggini sn, 64100 Teramo, Italy
[11] INFN, Sezione di Pisa, Largo Pontecorvo 3, 56127 Pisa, Italy
[12] INAF – Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, 35122 Padova, Italy
[13] Max-Planck-Institut für Astrophysik, Karl Schwarzschild-Straße 1, 85748 Garching bei München, Germany

ABSTRACT

Context. Recent observational programs are providing a global view of the Milky Way bulge that serves as a template for detailed comparison with models and extragalactic bulges. A number of surveys (VVV, GIBS, GES, ARGOS, BRAVA, APOGEE) are producing comprehensive and detailed extinction, metallicity, kinematics, and stellar density maps of the Galactic bulge with unprecedented accuracy. However, the still missing key ingredient is the distribution of stellar ages across the bulge.
Aims. To overcome this limitation, we aim to age-date the stellar population in several bulge fields with the ultimate goal of deriving an age map of the bulge. This paper presents the methodology and the first results obtained for a field along the bulge minor axis, at $b = -6°$.
Methods. We use a new PSF-fitting photometry of the VISTA Variables in the Vía Láctea (VVV) survey data to construct deep color–magnitude diagrams of the bulge stellar population down to ~2 mag below the main sequence turnoff. To address the contamination by foreground disk stars we adopt a statistical approach by using control-disk fields located at different latitudes (spanning approximately the bulge's range) and longitudes $-30°$ and $+20°$. We generate synthetic photometric catalogs of complex stellar populations with different age and metallicity distributions, including the observational errors and completeness. The comparison between the color–magnitude diagrams of synthetic and observed disk-decontaminated bulge populations provides constraints on the stellar ages in the observed field.
Results. We find the bulk of the bulge stellar population in the observed field along the minor axis to be older than ~7.5 Gyr. In particular, when the metallicity distribution function spectroscopically derived by GIBS is used, the best fit to the data is obtained with a combination of synthetic populations with ages in between ~7.5 Gyr and 11 Gyr. However, the fraction of stars younger than ~10 Gyr strongly depends upon the number of blue straggler stars present in the bulge. Simulations show that the observed color–magnitude diagram of the bulge in the field along the minor axis is incompatible with the presence of a conspicuous population of intermediate-age and young ($\lesssim$5 Gyr) stars.

Key words. Galaxy: structure – Galaxy: bulge

1. Introduction

The recent photometric surveys (e.g., VVV, Minniti et al. 2010; OGLE IV, Udalski et al. 2015) and spectroscopic surveys (e.g., BRAVA, Rich et al. 2007; Howard et al. 2008; ARGOS, Freeman et al. 2013; GIBS, Zoccali et al. 2014; GES, Gilmore et al. 2012; Randich et al. 2013; APOGEE, Majewski 2012; Majewski et al. 2017) have drastically changed our

understanding of the Milky Way bulge. The emerging picture is much more complicated than we believed until about one decade ago based on the stellar properties observed in a few small regions, mostly located along the bulge minor axis.

Owing to the systematic and detailed study of the red clump (RC) stellar population across the whole bulge we have finally reached a comprehensive view of the 3D structure of the bulge (see Zoccali & Valenti 2016, for a recent review). It has been established that our bulge hosts a bar with an orientation with respect to the Sun-Galactic center line of sight of ~27° (Wegg & Gerhard 2013), and whose near-side points toward the

first Galactic quadrant. The bar has a boxy/peanut/X-shape structure in its outer regions (McWilliam & Zoccali 2010; Nataf et al. 2010; Saito et al. 2011; Wegg & Gerhard 2013; Ness & Lang 2016), a typical morphology of bulges formed from the natural evolution of disk galaxies, and as result of disk dynamical instabilities and vertical buckling of the bar (see, e.g., Debattista et al. 2006, and references therein). The RC distribution also hints at the presence of an axisymmetric structure in the innermost ~250 pc (Gonzalez et al. 2011; Gerhard & Martinez-Valpuesta 2012; Valenti et al. 2016), and at a long bar with semi-major axis of ~4.6 kpc, which appears to be the natural thin extension of the main bar at larger radii (Wegg et al. 2015).

The spectroscopic metallicity distribution function (MDF) observed in different fields across the bulge is quite broad and complex ($-1.2 \lesssim$ [Fe/H] $\lesssim +0.7$ dex; Zoccali et al. 2017, and references therein). It is best represented by two components, metal-rich (MR) and metal-poor (MP), whose relative fraction changes along different lines of sight. In particular, MP stars dominate the outer regions, and are characterized by a *spherical* spatial distribution. On the other hand, the MR population becomes progressively more abundant toward the plane, showing instead a remarkably boxy distribution.

Although spectroscopic surveys of M-giants (e.g., BRAVA) and RC stars (e.g., ARGOS, GIBS, GES) have shown that while the bulge rotates cylindrically like a bar (Rich et al. 2007; Kunder et al. 2012; Ness et al. 2013b, 2016; Zoccali et al. 2014), there is compelling observational evidence that demonstrates that MR and MP stars have different kinematics (Babusiaux et al. 2010; Kunder et al. 2016; Zasowski et al. 2016; Zoccali et al. 2017; Rojas-Arriagada et al. 2017; Clarkson et al. 2018). MR stars show a steep velocity dispersion gradient as a function of the latitude, from $\sigma \sim 50$ km s^{-1} at $b = -8°$ to $\sigma \sim 140$ km s^{-1} at $b = -1°$. Instead, the MP components have a dispersion that ranges from ~80 km s^{-1} in the outer region to ~120 km s^{-1} at $b = -1°$.

While there is a general agreement regarding the properties of the bulge morphology (i.e., 3D structure), metallicity, and kinematics, a unanimous consensus on the age is still missing. In recent years the age of the bulge stellar population has been the most controversial problem because of a number of contradicting results based on different approaches.

Dating bulge stars is a very complicated task, made more difficult by stellar crowding, the large and high differential extinction, the uncertainties on the distance modulus, the distance spread due to the spatial depth of the bulge/bar along the line of sight, the metallicity dispersion, and finally the contamination by foreground disk stars. The different contribution of all these factors prevents us from determining an accurate location in terms of magnitude and color of the main sequence turnoff (MS-TO) of the bulge population, which are among the most reliable age diagnostics (Renzini & Fusi Pecci 1988). Historically, the earliest age constraint by van den Bergh & Herbst (1974) in the Plaut field along the bulge minor axis at $b = -8°$ (~1 kpc) indicated a globular cluster-like age. Terndrup (1988) fitted the photometry of bulge fields at a range of latitudes with globular cluster isochrones of varying metallicity, but because there was no certain distance for the bulge, only a weak age constraint (11–14 Gyr) was derived. Ortolani et al. (1995) solved the problem of the distance uncertainties by comparing the bulge population with the two clusters NGC 6528 and NGC 6553. Specifically, by measuring the difference between the RC and the MS-TO magnitude in the bulge field and in the clusters, they showed, for the first time, that the relative ages of the bulge and metal-rich cluster population could not differ by more than 5%.

Feltzing & Gilmore (2000) used the counts of stars brighter and fainter than the MS-TO observed in their HST-based photometry of Baade's window and another low extinction field known as the Sgr-I (at $l = 1.25°$ and $b = 2.65°$) to argue in favor of an old age. The case for a *purely old* bulge has been further strengthened by later studies based on more accurate HST and ground-based photometry of different bulge fields located mostly along the minor axis (Kuijken & Rich 2002; Clarkson et al. 2008, 2011; Zoccali et al. 2003), but also at the edge of the bar (Valenti et al. 2013). Unlike previous works, the problem of contamination from foreground disk stars was tackled either kinematically by using proper motions (Kuijken & Rich 2002; Clarkson et al. 2008, 2011; Calamida et al. 2015a,b) or statistically by considering control disk fields (Zoccali et al. 2003; Valenti et al. 2013). From the analysis of optical and near-IR decontaminated color–magnitude diagrams (CMDs), these studies found the bulk of the bulge stellar population to be old (>10 Gyr), with no evidence of significant age differences between the field and old Milky Way (MW) cluster population. In particular, Clarkson et al. (2011) provided an upper limit of ~3.4% for a bulge component younger than 5 Gyr, although they argued that the majority of the stars brighter than the old MS-TO in their CMDs could be blue straggler stars (BSS).

There is, however, a clear discrepancy between the ages inferred from the determination of the MS-TO in the observed CMDs and those derived by the microlensing project of Bensby and collaborators (Bensby et al. 2013, 2017), which derives individual stellar ages from the effective temperature and gravity (i.e., from isochrones in the [T_{eff}, log g] plane) as obtained from high resolution spectra. Based on a sample of 90 F and G dwarfs, turnoff, and subgiant stars in the bulge ($|l| \lesssim 6°$ and $-6° < b < 1°$) observed during microlensing, Bensby et al. (2017, hereafter B17) found that about 35% of the MR stars ([Fe/H] > 0) span ages between 8 Gyr and 2 Gyr, whereas the vast majority of MP ([Fe/H] $\lesssim -0.5$) are 10 Gyr or older. In addition, from the derived age-metallicity and age-α element distribution, the authors concluded that the bulge must have experienced several significant star formation episodes about 3, 6, 8, and 12 Gyr ago. Comparable results have been found by Schultheis et al. (2017), who presented the age distribution of 74 giants in Baade's Window as a function of stellar metallicity. Specifically, the relation of Martig et al. (2016) calibrated on asteroseismic data has been used to link the [C/N] abundances measured from APOGEE spectra to the stellar age. While the age distribution of the MP ([Fe/H] < −0.1) giants peaks at 10 Gyr with a decreasing tail toward younger ages (as young as 2 Gyr), MR ([Fe/H] > −0.1) stars can be either young or old; their age distribution appears bimodal, with two peaks at 4 and 11 Gyr.

Different concepts have been explored and proposed to partially reconcile the spectroscopic and photometric ages. In this respect the first attempt was presented by Nataf & Gould (2012) and Nataf (2016) who proposed a higher helium enrichment factor than currently adopted for the MR isochrones. The use of standard isochrones on He-enhanced stellar populations would lead to photometric and spectroscopic ages that are over- and under-estimated, respectively. Therefore, the discrepancy could be interpreted under the assumption that the chemical evolution of the bulge is He-enhanced. On the other hand, Haywood et al. (2016) suggested that the discrepancy is caused by the effect of the age-metallicity degeneracy that makes it hard to distinguish in the CMD a young MR star from an old MP one. They compared the MS-TO color spread observed in the CMD of Clarkson et al. (2011) with that of synthetic CMDs obtained by using two different age-metallicity relations, one

presented by Bensby et al. (2013), based on a total sample of 59 microlensed dwarfs, and one that extends from [Fe/H] = −1.35 dex at 13.5 Gyr to [Fe/H] = +0.5 dex at 10 Gyr. When taking into account distance, reddening, and metallicity effects, Haywood et al. (2016) showed that the MS-TO color spread of a *purely* old stellar population would be wider than what is observed, and thus advocating for the presence in the bulge of a conspicuous population of young and intermediate-age stars. Very similar results are presented by Bernard et al. (2018) who calculated the star formation history of four bulge fields, including that of Clarkson et al. (2011). Their findings suggest that over 80% of the stars are older than 8 Gyr, but also suggest the presence of star formation as recent as ∼1 Gyr.

Clearly, currently the age distribution of the bulge is still not universally understood, and in particular its spatial variation across the large area of the bulge has not been fully explored. In this framework, the use of near-IR deep photometry provided by the VVV survey represents a unique opportunity. With the ultimate ambitious goal of deriving the age map of the stellar population across the bulge, in this paper we present the methodology used to derive stellar ages from the CMDs and the first results obtained on a field located along the bulge minor axis, at $b = -6°$.

2. Observations and data reduction

This analysis makes use of a combination of J and K_s single-epoch observations of bulge and disk fields collected with the wide field near-infrared imager VIRCAM mounted at the VISTA telescope at the ESO Paranal Observatory. The bulge data for a field at $(0°, -6°)$ have been taken from the VVV survey. In addition, eight disk fields with latitudes in the range $-8° < b < +4°$ and longitude $-30°$ and $+20°$ (see Table 1) were observed in Service Mode as part of the program 095.B-0368(A) (PI: Valenti) using exactly the same VVV observing strategy.

The VISTA/VIRCAM focal plane is a mosaic of 16 detectors with gaps on the order of the size of the detectors between them. A single image is called pawprint, which comprises a single exposure of all 16 detectors, producing an image with gaps. Additionally, a jitter of ∼60 pixels is applied to subtract sky background. In order to fill the gaps in the individual exposures, a series of 6 pointings are combined in a 2×3 dither *pawprint pattern*, which ensures more or less homogeneous coverage in a $\sim 1.5 \times 1.2$ deg^2 field of view. This strategy also helps to diminish the effects of defects in particular detectors by having each pixel exposed at least twice.

However, the overlap areas between individual pawprints and edges of the tiles have 2–6 times higher exposures, meaning that the noise distribution within a tile varies strongly with position in the sky. This drove the decision to run the photometry on stacked pawprint images, rather than the composite tile, as described in Table 1 below. All raw data has been processed by the VISTA data flow system pipeline (Lewis et al. 2010) at CASU, which among a number of final products[1] provides the tiled and stacked pawprint images.

We carried out photometry including PSF modeling on each single detector from the pawprint image by using a customized pipeline based on DAOPHOT (Stetson 1987) and ALLFRAME (Stetson 1994). The single-detector catalogs derived in this way are combined together using error weighted means in case of overlap to produce a single-band catalog of a tile. Then, for each observed bulge and disk fields, we derive a photometric catalog

---

[1] http://casu.ast.cam.ac.uk/vistasp/

Table 1. Observed bulge and disk fields.

Name	(l, b)	$\langle E(J - K_s) \rangle^a$	Detected stars
b249	$(-0.45°, -6.39°)$	0.150 ± 0.047	2 728 265
c001	$(-30.00°, -2.97°)$	−	2 069 471
c002	$(-30.00°, -5.97°)$	−	1 153 139
c003	$(-30.00°, -7.98°)$	−	723 786
c004	$(-30.00°, +3.98°)$	−	1 537 790
c005	$(+19.99°, -2.97°)$	−	2 555 073
c006	$(+20.00°, -5.98°)$	−	1 524 750
c007	$(+19.99°, -7.98°)$	−	1 044 536
c008	$(+19.99°, +3.97°)$	−	1 711 455

Notes. [a] Color excess from Gonzalez et al. (2012), average and standard deviation are taken over the whole tile.

listing the instrumental J and K_s magnitudes through cross-correlation of the single-band catalogs. Finally, the CASU catalogs have been used to perform the absolute calibration (by means a of simple magnitude shift using several thousands stars in common) and astrometrization onto the VIRCAM photometric and astrometric system. More technical details specific to catalogs cross-match and internal detector-by-detector calibration will be provided in a companion paper (Surot Madrid et al., in prep.) where the release of the PSF-fitting photometry of the entire VVV bulge area ($-10° < l < 10°$ and $-10° < b < +5°$) is presented.

We estimated the photometric and systematic errors affecting the derived magnitudes by means of artificial star experiments. Specifically, ∼200 000 artificial stars per detector were added to the stacked pawprint images with random magnitudes J and K_s compatible with those observed in the instrumental CMDs. In other words, the artificial stars were added along the observed CMD sequences, including their spread. In each independent experiment, the artificial stars were simulated by using the same PSF model obtained during the photometric reduction process, and spatially distributed around a grid properly customized to avoid artificially increasing the crowding (see Zoccali et al. 2000, for further details).

As shown in Fig. 1, the combined effect of systematics and photometric uncertainties produces a spread in the recovered versus injected magnitudes (red solid line) that is considerably larger than would be expected from the photometric error alone (black dots). It also evidences the known issue related to the saturation of bright stars in the VVV K_s-band images ($K_s \sim 12$). There is also saturation in the J-band, but the uncertainties up to $J \sim 10$ are well in line with the rest of the profile, from both photometric and simulated sources.

For each targeted field, the simulations yield the fraction $p = p(J, K_s)$ defined as the number of injected stars recovered with $|m^{in} - m^{rec}| < 0.75$ mag, to the total number of injected stars per color–magnitude bin. Therefore, $p = p(J, K_s)$ is the probability of observing a given star with J and K_s magnitudes (i.e., magnitude and color) within a specified bin, hence providing the completeness of the derived catalogs. As expected, p decreases for fainter magnitudes and redder colors; however, we also found that it changes on a detector-by-detector basis, with absolute differences on the order of $\Delta p = 0.1$ around the faint magnitude end of the CMD. In the case of the less crowded fields ($b \lesssim -6°$) the simulations show that the photometric catalogs of the observed bulge and disk fields are more than 50% complete above $J \sim 18.5$ and $K_s \sim 18$ (see middle panel of
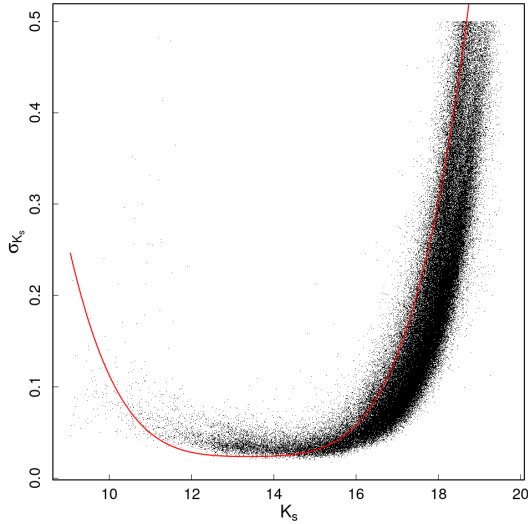
Fig. 1. Distribution of photometric errors as a function of magnitude (black dots) in the K_s band of detector #12 for the bulge field b249. With the exception of the magnitude range $15 \gtrsim K_s \gtrsim 11$, the photometric errors are well below the dispersion calculated from simulations (red line). Also worth noticing is the effect from saturation for stars with $K_s \lesssim 12$.

Fig. 2, and right panels of Fig. 3). The 50% completeness line moves to brighter magnitudes the closer the field is to the galactic plane.

3. Observed bulge color–magnitude diagram

The observed $(J-K_s, K_s)$ CMD of the selected VVV field (b249) is shown in the left panel of Fig. 2. When using standard point-scatter plots, the large number of detected stars (see Table 1) saturates the plots for $K_s > 15$ preventing us from distinguishing different evolutionary sequences, such as the subgiant branch (SGB) and the MS-TO. Therefore, we show the CMDs in Fig. 2 as Hess density diagrams.

The observed CMD shows a well-populated RC at $K_s \sim 13$ and $(J - K_s) \sim 0.8$, a prominent red giant branch (RGB) easily traceable down to $K_s \sim 16$, and the MS-TO at $K_s \sim 17$ and $(J - K_s) \sim 0.5$. On the other hand, the SGB is barely visible because it is heavily contaminated by the foreground disk population. The two vertical blue sequences departing from the bulge MS-TO and RC upward correspond, respectively, to the foreground disk MS and its RC descendants.

In the magnitude range $10 \leq K_s \leq 17.5$, the observed spread in all the sampled evolutionary sequences, is mostly due to the combination of metallicity and depth effect, whereas in the fainter range the contribution of the photometric errors becomes predominant. In b249 the differential reddening is not dramatically large ($\Delta E(J - K_s) \sim 0.047$ mag, see Table 1); when we correct for the extinction by using the VVV-based extinction map from Gonzalez et al. (2012) the dereddened CMD shows a very similar magnitude spread (e.g., compare left- and rightmost panels in Fig. 2).

Finally, as expected, the characteristic double RC signature of the well-known X-shape bulge structure is detected at $K_{s0} = 12.75$ and 13.27 (see Fig. 4 for a zoom-in). As demonstrated by several authors (see Sect. 1) the X-shape structure is detected only in the outer bulge, at latitudes $|b| > 5.5°$ and longitudes $|l| \leq 4.5°$.

4. Disk decontamination procedure

As shown in Fig. 2, the MS of the disk hits the bulge CMD exactly on top of its MS-TO, hence preventing any reliable age determination. Therefore, prior to any attempt to age-dating the bulge population from its observed CMD, special care must be taken to remove the contribution of the intervening disk population along the line of sight. This can be done either by using proper motions (see Clarkson et al. 2008, 2011; Kuijken & Rich 2002; Bernard et al. 2018) or statistically with a disk control-field (e.g., Zoccali et al. 2003; Valenti et al. 2013). Both methods require some assumptions, and therefore have their own pros and cons.

In principle, the proper motions selection yields more accurate decontamination providing that bulge and disk populations have clear distinct kinematics along the line of sight of interest, which is not always the case. As is evident in Fig. 4 in Bernard et al. (2018), along different lines of sight the proper motions of disk and bulge largely overlap, which means that there is no simple clear cut in μ_l and/or μ_b that would differentiate disk from bulge; although a filter in proper motions can still be applied, it would inevitably end with an equally problematic cross-contamination between the populations. In addition, from the observations point of view the determination of proper motions is very time consuming. The need for precise astrometric measurements and long time-baselines for error reduction effectively limits this type of observations to expensive and accurate observatories such as HST, which has a small field of view ($\lesssim 3 \times 3$ sq. arcmin). Finally, it is worth mentioning that the proper motions provided by *Gaia* (DR2) do not allow a proper *cleaning* of the CMDs. Being severely limited by both crowding and reddening already at $b = -6°$, the *Gaia* photometry is too shallow and does not fully sample the old MS-TO.

On the other hand, the statistical approach best suits the case of large surveyed areas such as that presented here, but it relies on the assumption that the adopted control-field is representative of the disk population along the bulge line of sight. In this respect, the selection of the disk control-field is crucial and must take into account that the contamination of the bulge CMDs from foreground disk stars strongly depends on latitude.

4.1. Comparable populations

Figure 3 shows the derived Hess density diagrams in the $(K_s, J - K_s)$ plane of the eight disk control-fields studied here, together with the corresponding completeness map obtained from the artificial stars experiment (see Sect. 2).

On the blue side of the diagram for $(J - K_s) \lesssim 0.5$, the very well populated MS can be easily identified, while the redder vertical sequence corresponds to the evolved RC population. At fixed latitude, the sampled disk population in the fields at longitude $+20°$ and $-30°$ has a fairly similar CMD, with the major difference being the overall reddening, which appears to be less at $-30°$. As a consequence, the color of the MS and RC stars is generally bluer than is observed at $+20°$.

As expected from an exponential disk density profile at a fixed longitude, the number of detected stars increases in fields closer to the Galactic plane, whereas at a given latitude the fields at $+20°$ are systematically more populous than their counterpart $-30°$ (see Table 1).

The first step in the decontamination process is the selection of the field that best represents the disk population observed along the bulge line of sight. This is done by comparing the bright portion of the CMDs ($K_s \lesssim 15$) in the bulge and disk
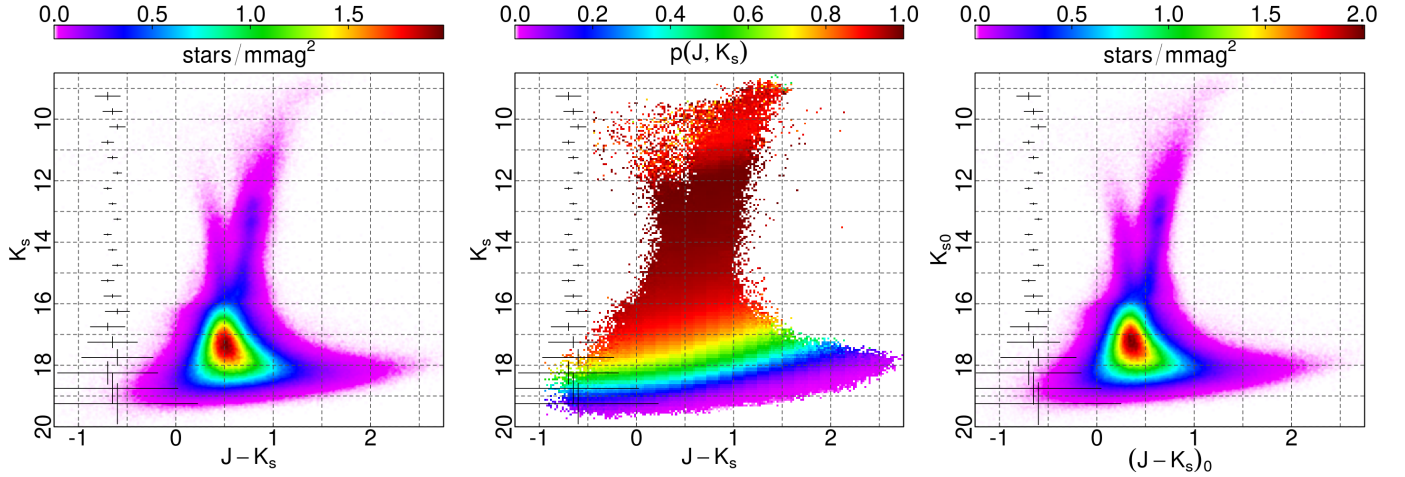
Fig. 2. *Left panel*: observed CMDs of the VVV field (b249) shown as a Hess density diagram, with the corresponding photometric completeness map (*middle panel*) as detailed in Sect. 2. *Right panel*: reddening-corrected CMD of the same field. Typical errors on the color and magnitude are shown as crosses for all cases.

fields. Specifically, we trace the profile of the young disk MS and the RGB by means of a series of Gaussian fits to their $(J - K_s)$ color distribution per K_s magnitude bin (see Fig. 4). It is worth mentioning that these sequences lie on the bright and most complete part of the CMD, and as such have the smallest errors in general.

Shifts in color and magnitude along the reddening vector (see Nishiyama et al. 2009; Gonzalez et al. 2012) are applied to all disk control-fields to match the profile of the young blue MS of each control-field with that observed in the b249 field. The shifts applied here are carried over for all subsequent analyses. In doing this, we are ignoring distance distribution differences between the bulge and control fields. However, because of the lack of secure and recognizable standard distance features in the disk sequences, there is no way to properly account for them. We compare the relative differences between the profiled young MS in b249 and in all disk fields, and select the one with the lowest dispersion of the residuals, which are in turn defined by the separation of the profiled young MS of the control field, interpolated onto the observed bulge counterpart. Following this procedure, the disk field c002 located at approximately the same latitude as the target bulge region turns out to be the most appropriate for decontamination purposes.

After choosing the best candidate control field, the next important step is to make sure that its dispersion in color and magnitude due to the combination of systematics and photometric uncertainties (see Sect. 2) is comparable to that observed in the bulge field. We have in principle two kinds of uncertainties. The first is related solely to the PSF fitting procedure, and to the counts of the individual star profiles in an image taken with the photometric filter M: σ_M ($M = J, K_s$). This is our photometric error, which is tied to the shape and brightness of the individual stars. The second comes from the measured dispersion in the completeness experiments, how artificial stars with similar injected magnitude m^{in} disperse into randomly different recovered magnitudes m^{rec}. We call this $\Sigma_M = \Sigma_M^{field}(M)$, which is mostly a function of J, K_s, unique to each *field* (detector and image), and likely comprising the systematics in our data. To estimate this source of uncertainty, for each detector we use a fourth-degree polynomial to fit the K_s-binned m_{in} versus the MAD($m^{in} - m^{rec}$) profile, where MAD is the adjusted median absolute deviation (1-MAD is equivalent to $1 - \sigma$ from a normal distribution).

We need to take into account differences in observing conditions and crowding for the observed disk and bulge fields, which result in differences in PSF fitting and the associated error profiles. Given these differences between images, and when considering individual detectors, we must be sure that the ith star in the control field catalog, with magnitude M^i, has an error similar to that of an equivalent star in the bulge catalog. To do this, we define a value ς_M^i for the ith star in the control field, as the maximum between its photometric error, completeness dispersion $\Sigma_M^{c002}(M^i)$ and the corresponding $\Sigma_M^{b249}(M^i)$ from the observed bulge field:

$$\varsigma_M^i = \max\left(\sigma_M^i, \Sigma_M^{c002}(M^i), \Sigma_M^{b249}(M^i)\right). \tag{1}$$

Similarly, as mentioned in Sect. 2, each field and detector has its own completeness $p^{field}(J, K_s)$; therefore, a further step to guarantee a proper statistical subtraction is to correct the control field by its completeness and then apply the completeness of the observed bulge field. In practice, this can be achieved by assigning a weight ω^i to the ith star in the control catalog, defined as

$$\omega^i = p^{b249}(J^i, K_s^i)/p^{c002}(J^i, K_s^i). \tag{2}$$

Finally, we must calculate the bulge-to-disk normalization factor, which gives us the number of stars to be removed from b249 for each given disk star observed in c002. To do so, we select a region in the b249 CMD where only the disk population ($(J - K_s) \lesssim 0.5$ is likely to be found, and $K_s \lesssim 16$), as was done by Zoccali et al. (2003) and Valenti et al. (2013). This factor is a single scalar applied uniformly throughout the control CMD.

4.2. Kernel approximation and subtraction

To take into account the effects of the error bars and systematics in the bulge and disk CMDs, we adopted a bivariate Gaussian kernel smoothing map. This is similar to modeling any given star in the $[K_s, J - K_s]$ plane as a bivariate normal distribution, whose centroid is just the color–magnitude position of the star, and its covariance matrix constructed from the errors of J and K_s. We then stack/add all the Gaussians, and evaluate the result on a finer grid so that now the integral of this yields the expected number of stars in any given region of the CMD. As σ_{K_s} and σ_J, we take the corresponding ς_M from (1), and approaching the problem as if all stars had errors defined by this quantity.
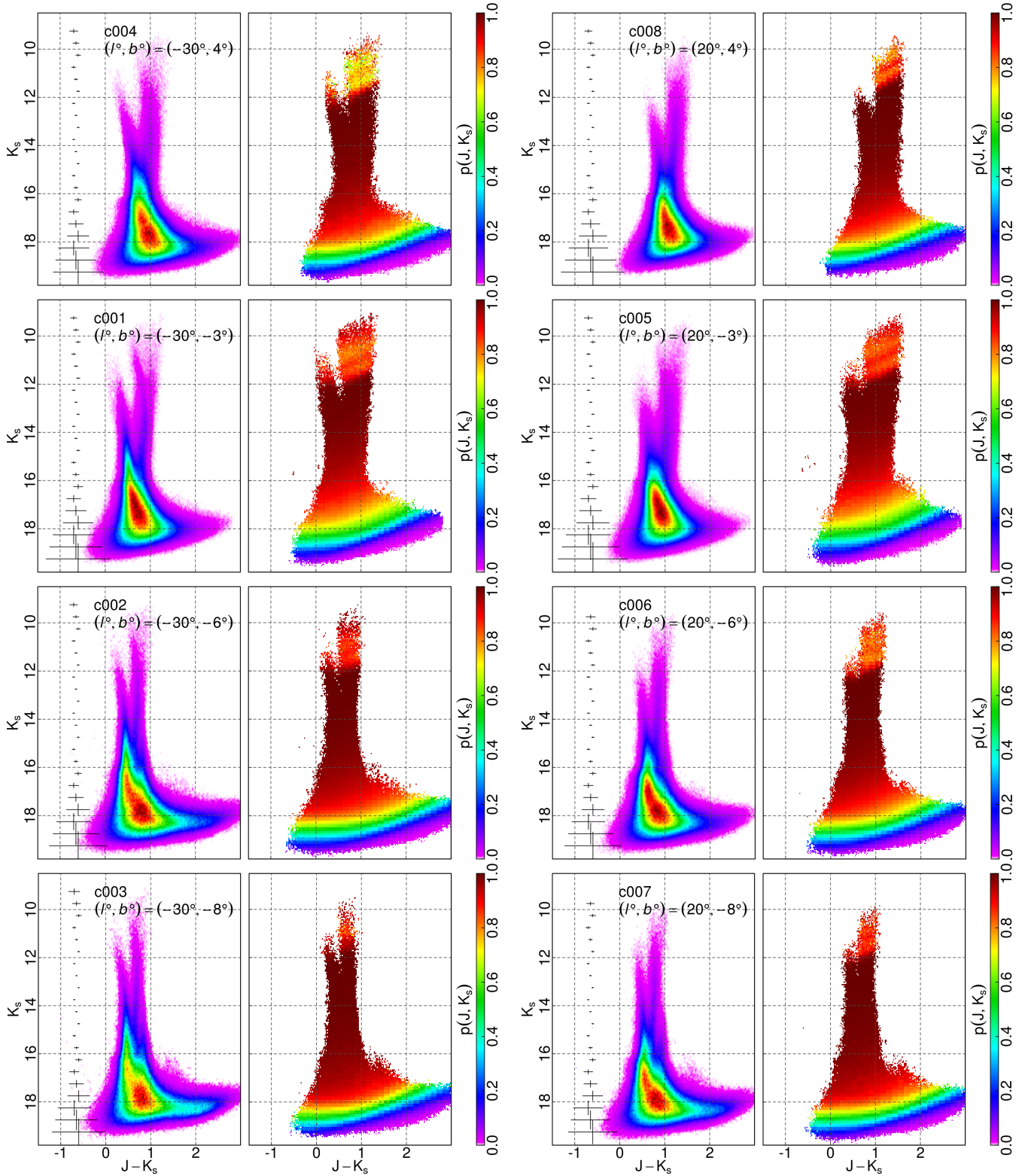
Fig. 3. Observed CMDs of the eight disk-control fields (*left panels*) and the corresponding photometric completeness map (*right panels*).

Because the errors and dispersion are a function of the *J* and K_s magnitudes for any given star, there is no unique kernel valid for any given catalog. Therefore, we divide the dataset in $\sigma_J - \sigma_{K_s}$ bins, calculate the kernel map for each, and add them together. Once the kernel map is constructed, we scale it by using the bulge-to-disk normalization factor (see Sect. 4.1) in order to ensure that the expected number of stars in the kernel map and in the observed bulge CMD is the same in the defined disk-only window. After estimating the expected number of stars from the kernel map in small regular bins across the color–magnitude space, we extract the corresponding number of stars by randomly picking the catalog entries from each bin. We then count the number of stars within the defined window for the remaining stars and compare it with the original observed
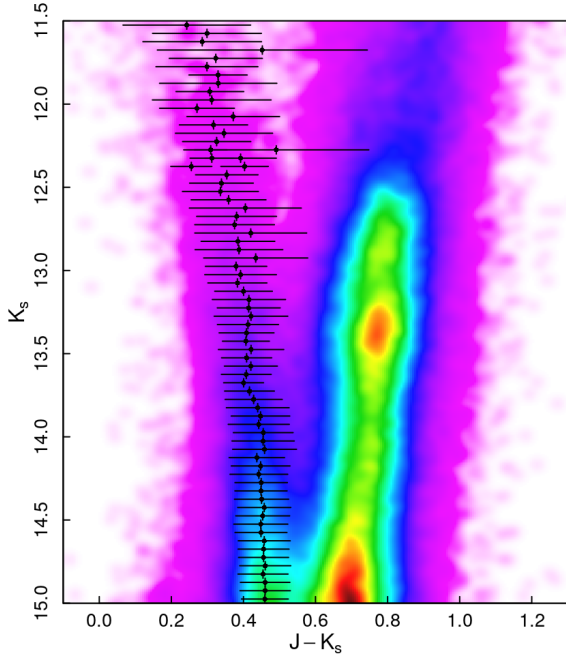
Fig. 4. Color-coded Hess density diagram of b249 compared to the Gaussian young MS profile (black dots). Horizontal and vertical crosses refer respectively to the estimated sequence width and bin size.
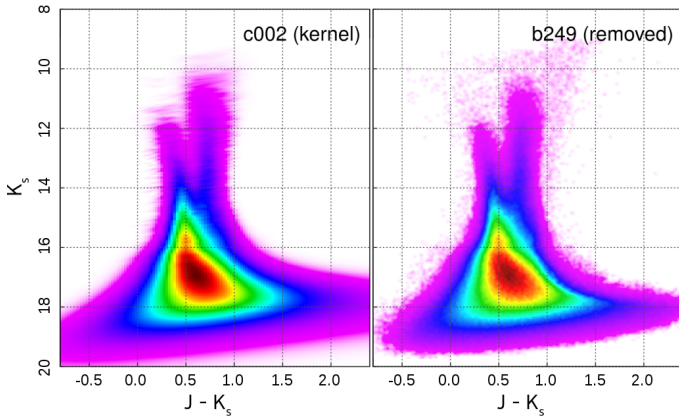

Fig. 5. *Left panel*: intensity kernel map for c002, as constructed from b249 dispersion. *Right panel*: color–magnitude diagram of the stars removed from b249 by using the kernel map shown in the *left panel*.

Fig. 6. HESS diagram of the bulge b249 field statistically decontaminated from the foreground disk population. Solid black contours are isodensity curves spaced by 1% and 5%, and then from 1/12 to 11/12 of the maximum density in steps of 1/6. Also plotted are BaSTI isochrones for ages (bluest to reddest) 5, 7.5, 10, and 11 Gyr, following the metal-poor regime from GIBS for the oldest (gray lines), and metal-rich for all others (dotted line). The isochrones have been shifted so that their RC and RGB roughly match the observed ones.

catalog in order to get a new scale factor (usually a small correction on the order of 10%) and repeat once more the subtraction.

The remaining stars represent the effective bulge clean sample. It is important to note that this procedure yields bulge-like and disk-like catalogs according to their CMDs only. In other words, if we had additional information (e.g., kinematics) that would allow us to clearly distinguish disk and bulge stars individually, and used that on either clean or removed catalogs, neither would show a pure disk or bulge population, but the resulting CMDs would look as if they were, regardless of their true precedence.

5. Bulge clean sample

We performed the decontamination procedure described in Sect. 4 on the observed bulge field b249 by using c002 as disk control-field.
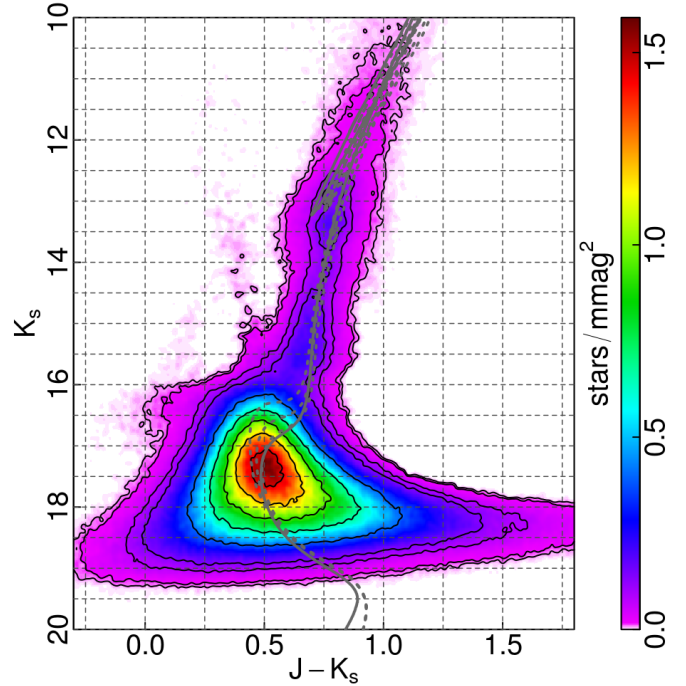
In Fig. 5 we show the kernel map of c002 (left panel) and the density map of the stars subtracted from b249 (right panel). The overall similarity of the two maps provides a first sanity check, demonstrating that the procedure worked as intended. Indeed, for $K_s \lesssim 19$ and $0 \lesssim J - K_s \lesssim 1.2$ the maps are virtually indistinguishable, except for the two faintest magnitude bins where the kernel shows a wing extending redward that is not reflected in the removed set. In this case, the wing extends toward rapidly declining completeness (see Fig. 2), which is not taken into account in the kernel map but only considered in the weight of the entries of the control catalog.

The final result of the decontamination procedure is presented in Fig. 6, where we show the CMD of the bulge b249 clean sample, consisting of 1 654 603 stars. By removing the foreground disk population, it is now possible to more easily recognize the bulge main evolutionary features: the double RC $K_s = 12.83, 13.35$ and $(J - K_s) = 0.79$); the RGB ($K_s \lesssim 15.5, (J - K_s) \gtrsim 0.5$); the SGB ($16.5 \gtrsim K_s \gtrsim 15.5$, and $0.4 \lesssim (J - K_s) \lesssim 0.7$); and the hot spot near $K_s \sim 17$ with a shape resembling a MS-TO. The halos silhouetting the subtracted young disk MS are outliers that could not be removed completely. However, their signal in the Hess density diagram is very weak (i.e., ~4000 stars, <0.2% of the maximum density value, and ~11% of the original observed in the same region); therefore, they can be safely neglected. Finally, the blue feature near $(J - K_s) \sim 0.1$ and $K_s \sim 16.5$ is likely an artifact due to stars mostly located at a corner of detectors 2, 8, and 16. The diagnostic values (shape and χ^2 from PSF-fitting) of these artifacts have different distributions to that of normal stars, but they still overlap, meaning that a usual cut (e.g., in χ^2) cannot clean

them effectively. This feature was also present in the observed CMD (Fig. 2), but since it is not real nor is it present in the control field, it became highlighted after the decontamination.

6. Stellar ages determination

To determine the age of the bulge stellar population in the b249 field a simple isochrone fitting of the clean sample does not yield meaningful results because of the spread in color and magnitude induced by the observational effects around the expected MS-TO (see Fig. 6). In addition, because we are studying a complex stellar population over a large surveyed area, the metallicity and large distance spread introduce further dispersion in color and magnitude that cannot be qualitatively taken into account by using simply isochrones. Instead, we resort to the comparison of the bulge clean sample with synthetic populations with known age and metallicity.

The synthetic populations are obtained with the IAC-star code (Aparicio & Gallart 2004), employing the BaSTI stellar evolution models (Pietrinferni et al. 2004, 2006, 2013, 2014), and by using the empirical IMF of Calamida et al. (2015b) with a 35% binary fraction.

We adopt the spectroscopic MDF provided by the GIBS survey (Zoccali et al. 2017), and based on 213 RC stars observed in the b249 field. The observed b249 MDF is best represented by two Gaussian components: the MR distribution peaks at $[Fe/H]_{MR} = +0.08 \pm 0.16$ dex, while the MP at $[Fe/H]_{MP} = -0.35 \pm 0.17$ dex. Moreover, to further constrain the synthetic populations, we adopt a 1:1.2 ratio of MP to MR, as suggested by Zoccali et al. (2017), and an α-element enhancement of +0.31 for the MP population (Ness et al. 2013a; Gonzalez et al. 2015; McWilliam 2016, and references therein). We note that this consideration toward differing $[\alpha/Fe]$ is merely an approximation. A more careful approach should account for $[Fe/H]–[\alpha/Fe]$ relations for both metallicity regimes.

We also add the reddening to the synthetic population from a random draw, emulating the color excess distribution of the stars in the b249 field according to the extinction map of Gonzalez et al. (2012). To account for the depth effect we use the split RC information of the observed field. We combine two identical synthetic populations with a gray magnitude shift equal to the distance between the split RC Gaussians, and add to each the corresponding gray Gaussian spread obtained from the quadratic difference between the Gaussians fits of the RC in observed and synthetic CMDs. The combination is constructed so that the star count ratios between the RC Gaussians in the synthetic CMD is equal to that in the observed CMD. Based on the RC distribution, which as we pointed out in Sect. 5 includes two overdensities, the distance probed in the observed bulge field is approximately between 6 and 11 kpc.

Theoretically, once the metallicity, α-enhancement, distance, and reddening of the synthetic population are constrained, and a given binary fraction, reddening law, and IMF are assumed, the age becomes the only free parameter. Hence, in principle we could let the age vary until we find the best match to the observed clean sample. In practice, however, the situation is further complicated by the dispersion due to the observational uncertainties (i.e., photometric errors and completeness). Any comparison between synthetic and observed CMDs aimed at providing reliable age estimates must take into account the observational effects. This is best done by dispersing the synthetic populations according to a new completeness map specifically tailored for this purpose. We recall here that the artificial stars experiment described in Sect. 2 has provided a completeness ($p = p(J, K_s)$)

for all the stars in the observed catalog alone. If we want the age of the synthetic population to vary freely we need to quantify the completeness and the distribution of the differences ($m^{in} - m^{rec}$), also for some regions that are not populated in the observed CMD. We achieve that by adding ~1 000 000 artificial stars per detector, distributed around the same spatial grid as in Sect. 2, but covering a specific region of the CMD ($20 \lesssim K_s \lesssim 9$, and $-0.2 \lesssim (J - K_s) \lesssim 1.4$) where we know model stars lie, and then constructing a new completeness map[2] following the same procedure outlined in Sect. 2.

According to the models, the observed RC shape suggests that the MP component must be quite old, between 11 and 12 Gyr; therefore, in our simulations we only allow the age of the MR component to vary. We note that this is also in line with all previous studies (see Sect. 1) that found the MP bulge stellar population to be consistently older than ~10 Gyr. If a significant fraction of a younger population is present then it is likely to be more metal rich than the older one. Although our simulations explored a fairly broad age range (Age (Gyr) $\lesssim 13.5$), in the present paper we present and discuss only a subsample that can be considered the most representative of the entire set of simulations. Specifically, for the age of the two components we considered six different scenarios:

- S1: MP = (11 ± 0.3) Gyr, MR = (10 ± 0.3) Gyr
- S2: MP = (11 ± 0.3) Gyr , MR = (7.5 ± 0.3) Gyr
- S3: MP = (11 ± 0.3) Gyr, MR = (5 ± 0.3) Gyr
- S4: age-metallicity distribution from B17 as obtained from their Fig. 14 (i.e., from individual stars).
- S5: MP = (11 ± 0.3) Gyr,
  MR = $40\%(10 \pm 0.3)$ Gyr + $60\%(7.5 \pm 0.3)$ Gyr
- S6: MP = (11 ± 0.3) Gyr,
  MR = $60\%(10 \pm 0.3)$ Gyr + $40\%(7.5 \pm 0.3)$ Gyr.

The ages are always flat distributions, and for all simulations but S4 we adopt the MDF spectroscopically derived by the GIBS survey that is best represented by two Gaussians for the MR and MP components. In addition, as mentioned before, we adopt the MR-to-MP ratio found in GIBS implying that the number of RC stars in the MR and MP domains has the ratio 1.2:1. The results of the simulations for the six scenarios are shown in Fig. A.1, where in the left panels we compare the Hess density diagrams of the synthetic population with the isodensity contours corresponding to the clean observed sample as derived in Fig. 6.

To assess the quality of the match between simulations and observations and to derive the best-fit model, we create for each simulation the corresponding residuals map (see right-hand panels in Fig. A.1). Specifically, the noise of the observed CMD is first modeled with a bootstrap approach. We repeatedly (~10 000 realizations) take a random resample of the complete observed catalogs, and save the residual between the subtraction of the original and resampled CMD density for each iteration. This effectively produces a residual dispersion per color–magnitude bin of the observed CMD, which is used to measure any significant deviation when comparing the synthetic populations (i.e., subtraction between the synthetic and observed CMD). Following this approach, the perfect match between simulation and observation should have a residuals map that is characterized only by noise, hence by the lack of any structures. We decided to favor a map comparison approach over a single-digit quantification of a goodness of fit (e.g., χ^2) because we expect the quantification of the discrepancies between the observed and

---

modeled datasets to be relevant, and we also wanted to know where in the CMD these discrepancies arise. For instance, a concentrated high-σ difference at $K_s < 16$ can be as important as a wider spread lower-σ difference at the $K_s > 17$ level.

The residual maps shown in Fig. A.1 have been scaled to the same color–magnitude window and identical sigma ranges, hence they can be directly compared against each other. Because the structures built upon the noise appear only above $|\sigma| \approx 6$, to be conservative the signal of a given feature in the map is considered significant if $|\sigma| \gtrsim 7$ (i.e., above green and below cyan according to the adopted color-coding in Fig. A.1). In addition, we adopted the simple convention where a positive $\sigma_{\Delta N}$ value implies excess observed stars with respect to the simulation, while negative values imply the opposite.

From Fig. A.1 we can rule out the possible presence in the bulge of a significant fraction of intermediate-age and young populations ($\lesssim 5$ Gyr). The residuals map of S3 and S4 shows a significant ($\sigma_{\Delta N} \gtrsim 8$) mismatch around the expected MS-TO (the roundish structure that coincides with the highest isodensity contours). In addition, the S4 scenario predicts the presence of a large number of stars brighter than $K_s \sim 16$ and $0.25 \lesssim J - K_s \lesssim 0.5$. that are not observed (shown as small black dots in the residuals map of S4 in Fig. A.1). Additionally, S4 has a slightly better fit for both RC and blue RGB, although this likely comes from the wider MDF used for this simulation.

On the other hand, the case of a purely very old bulge, represented by simulation S1, provides a good match of the region around the MS-TO, but significantly ($\sigma_{\Delta N} \gtrsim 12$) underestimates the number of stars observed at $16 \lesssim K_s \lesssim 16.5$ and $0.25 \lesssim J - K_s \lesssim 0.5$ (red area in the S1 residuals map in Fig. A.1). The region at the base of the RGB, $(J - K_s) \sim 0.75$ and $K_s \sim 16$, is also not well reproduced, but the presence of this mismatch in all simulations suggests that it is an artifact caused by the decontamination procedure. In fact, it corresponds to the bright end of the local M dwarf distribution clearly observed in the disk fields as the reddest vertical sequence around $(J - K_s) \sim 1$. As such, even if its intensity varies across the simulations, we refrain from taking it into account for the selection of the best-fit model.

When considering a slightly younger age (~ 7.5 Gyr) for the MR component, as in S2, S5, or S6, the overall match between the synthetic and observed population improves further, while the region around the MS-TO is still well matched. The deficit of synthetic stars present in S1 is considerably reduced, up to a factor of ~ 2 in the S2 case. This would make S2 the best-fitting scenario; however, it should be noted that in all simulations presented here we ignore the presence of BSS even though they have been observed in bulge fields (Clarkson et al. 2011). To qualitatively constrain their position in the $[K_s, J - K_s]$ plane, we have used the near-IR CMD of the bulge cluster NGC 6624 from Saracino et al. (2016), which includes BSS identified from UV and optical (Ferraro, priv. comm.). When accounting for the difference in distance and reddening between the cluster and the b249 field, we found that the BSS are approximately located in the region where S1, S2, S5, and S6 show a deficit of simulated stars with respect to the observed ones. Therefore, which simulation among scenarios S1, S2, S5, and S6 best fits the observed CMD depends upon the number of BSS present in the field.

According to Clarkson et al. (2011), the bulge field BSS frequency (S_{BSS}), defined as the number of BSS scaled to the number of RC stars, is between 0.3 and 1.23, while for the clusters Ferraro et al. (2003) found $0.1 \lesssim S_{BSS} \lesssim 1$. However, based on a photometric study of a sample of globular clusters and dwarf spheroidals, Santana et al. (2013, 2016) found that the number of BSS grows almost linearly with the total stellar mass of the

system; therefore, the BSS frequency in dwarf spheroidals is much higher than in clusters. In addition, when considering that the dynamical state of the bulge is expected to be more similar to that of dwarf spheroidals than of clusters, it is plausible that the bulge BSS frequency provided by Clarkson et al. (2011) could be an underestimation.

For the case of the b249 field, $S_{BSS} = 1.23$ would imply the presence of $\sim 32\,000$ BSS. To properly take into account the BSS in the simulation we would need to have robust information on their density distribution per color–magnitude bin, which at the moment is still lacking. In principle, this can be obtained by using a large and statistically robust sample of observed BSS (Ferraro et al., in prep.). Here we note that taking into account a population of $\sim 32\,000$ BSS uniformly distributed within the region defined by the cluster NGC 6624 would have the effect of completely removing the mismatch between the observed and simulated CMD for the S2. On the other hand, higher BSS frequency values $S_{BSS} = 1.66, 2.82, 3.03$ need to be assumed to remove the star deficit highlighted in the residuals map of scenarios S5, S6, and S1, respectively. Finally, the S3 and S4 scenarios are incompatible with any value of BSS frequency.

7. Discussion and conclusions

We used VVV images of the field b249 to perform new PSF-fitting photometry, allowing us to construct an accurate and deep CMD that samples most of the evolutionary sequences from the RGB down to ~ 2 mag below the old MS-TO. By using a disk control field, a statistical decontamination procedure that accounts for relative completeness, extinction, photometric, and systematic errors, as well as relative population fractions, was applied to extract a clean bulge sample. The corresponding CMD has been used to assess the stellar ages. Due to the complexity of the system, instead of simple isochrone fitting, we decided to rely on comprehensive simulated observations of composite populations exploring different scenarios in terms of stellar metallicity and ages.

When the MDF spectroscopically derived by GIBS is used, the best fit to the data is obtained for a composite population with ages in between 7.5 Gyr and 11 Gyr (scenarios S2, S5, and S6); therefore, arguing for a formation scenario on a relatively short timescale ($\lesssim 5$ Gyr) in which the bulk of the bar/bulge stellar population was already completely formed at least ~ 7.5 Gyr ago. However, although only qualitatively, we have also shown that when taking into account the presence of BSS in the bulge, the region of the CMD where one would expect to see intermediate-age or young stars (≤ 5 Gyr) gets populated.

In particular, we noted that the best fit to the data can also be obtained with a purely very old complex population (10 Gyr and 11 Gyr, as in S1) when assuming a BSS frequency of ~ 3, thus suggesting an even earlier formation for the whole bulge stellar population. This result would be in excellent agreement with the very recent study by Renzini et al. (2018) based on very deep HST CMDs of four fields located along the bulge minor axis and $-4 \lesssim b^\circ \lesssim -2$. By using a combination of UV, optical, and near-IR filters they have photometrically tagged all bulge field stars, and compared the luminosity function of the most MR and MP with simulated old and intermediate-age populations. The authors found that MR and MP populations appear to be essentially coeval, and consistent with a ~ 10 Gyr old population.

Because BSS mimic a rejuvenated population due to the mass transfer, ignoring their presence could partially be one of the reasons that led previous studies to advocate for the evidence of very extended star formation in the bulge (e.g.,

Bernard et al. 2018). Therefore, including BSS in the simulation of synthetic CMDs could potentially reduce, or even remove, the tension between different studies based on the photometric approach.

On the other hand, the present study still does not reconcile the discrepancy between the photometric and spectroscopic age measurements. When we allow for the presence of a significant fraction (>20%) of intermediate-age and young stellar populations (scenarios S3 and S4) the synthetic CMD does not provide a reasonably good fit of the observations. It should be noted here that the comparison between simulations and observations has been performed on the properties of the entire CMD, not only in terms of color spread of the MS-TO, as was done in some previous studies (e.g., Haywood et al. 2016). Stars as young as 1 Gyr up to ~3 Gyr occupy a region in the CMD that is not populated by the observations. Because our observed bulge sample is statistically very robust (>1.6×10^6 stars) due to the large surveyed area (~$1.8 \deg^2$), even a small fraction (≳10%) of such a young component, if present, would have been detected in the observed CMD. We note that the field b249 has only a partial overlap with the surveyed area by Bensby et al. (2017), thus we cannot directly contradict the results derived by them. However, for both results to be in agreement, a steep age gradient favoring younger stars toward the Galactic center would be necessary (i.e., effective at about a degree closer to the center from b249). Even though N-body simulations may predict such gradient (e.g., Ness et al. 2014; Debattista et al. 2017), the number of stars in Bensby et al. (2017) is insufficient to categorically establish the presence of a vertical gradient, especially at the outskirts of their surveyed area.

Finally, the method presented here will be applied to other VVV tiles in order to explore possible age distribution variations across the bulge. At the same time, we will compare the results with more sophisticated simulations using genetic algorithms that will ultimately lead to the reconstruction of the full star formation history.

References

Aparicio, A., & Gallart, C. 2004, AJ, 128, 1465
Babusiaux, C., Gómez, A., Hill, V., et al. 2010, A&A, 519, A77
Bensby, T., Yee, J. C., Feltzing, S., et al. 2013, A&A, 549, A147
Bensby, T., Feltzing, S., Gould, A., et al. 2017, A&A, 605, A89
Bernard, E. J., Schultheis, M., Di Matteo, P., et al. 2018, MNRAS, 477, 3507
Calamida, A., Sahu, K., Anderson, J., et al. 2015a, in Fifty Years of Wide Field Studies in the Southern Hemisphere: Resolved Stellar Populations of the Galactic Bulge and Magellanic Clouds, eds. S. Points, & A. Kunder, ASP Conf. Ser., 491, 160

Calamida, A., Sahu, K. C., Casertano, S., et al. 2015b, ApJ, 810, 8
Clarkson, W., Sahu, K., Anderson, J., et al. 2008, ApJ, 684, 1110
Clarkson, W. I., Sahu, K. C., Anderson, J., et al. 2011, ApJ, 735, 37
Clarkson, W. I., Calamida, A., Sahu, K. C., et al. 2018, ApJ, 858, 46
Debattista, V. P., Mayer, L., Carollo, C. M., et al. 2006, ApJ, 645, 209
Debattista, V. P., Ness, M., Gonzalez, O. A., et al. 2017, MNRAS, 469, 1587
Feltzing, S., & Gilmore, G. 2000, A&A, 355, 949
Ferraro, F. R., Sills, A., Rood, R. T., Paltrinieri, B., & Buonanno, R. 2003, ApJ, 588, 464
Freeman, K., Ness, M., Wylie-de-Boer, E., et al. 2013, MNRAS, 428, 3660
Gerhard, O., & Martinez-Valpuesta, I. 2012, ApJ, 744, L8
Gilmore, G., Randich, S., Asplund, M., et al. 2012, The Messenger, 147, 25
Gonzalez, O. A., Rejkuba, M., Minniti, D., et al. 2011, A&A, 534, L14
Gonzalez, O. A., Rejkuba, M., Zoccali, M., et al. 2012, A&A, 543, A13
Gonzalez, O. A., Zoccali, M., Vasquez, S., et al. 2015, A&A, 584, A46
Haywood, M., Di Matteo, P., Snaith, O., & Calamida, A. 2016, A&A, 593, A82
Howard, C. D., Rich, R. M., Reitzel, D. B., et al. 2008, ApJ, 688, 1060
Kuijken, K., & Rich, R. M. 2002, AJ, 124, 2054
Kunder, A., Koch, A., Rich, R. M., et al. 2012, AJ, 143, 57
Kunder, A., Rich, R. M., Koch, A., et al. 2016, ApJ, 821, L25
Lewis, J. R., Irwin, M., & Bunclark, P. 2010, in Astronomical Data Analysis Software and Systems XIX, eds. Y. Mizumoto, K. I. Morita, & M. Ohishi, ASP Conf. Ser., 434, 91
Majewski, S. R. 2012, Am. Astron. Soc. Meet. Abstr., 219, 205.06
Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, AJ, 154, 94
Martig, M., Fouesneau, M., Rix, H.-W., et al. 2016, MNRAS, 456, 3655
McWilliam, A. 2016, PASA, 33, e040
McWilliam, A., & Zoccali, M. 2010, ApJ, 724, 1491
Minniti, D., Lucas, P. W., Emerson, J. P., et al. 2010, Nat. Astron., 15, 433
Nataf, D. M. 2016, PASA, 33, e023
Nataf, D. M., & Gould, A. P. 2012, ApJ, 751, L39
Nataf, D. M., Udalski, A., Gould, A., Fouqué, P., & Stanek, K. Z. 2010, ApJ, 721, L28
Ness, M., & Lang, D. 2016, AJ, 152, 14
Ness, M., Freeman, K., Athanassoula, E., et al. 2013a, MNRAS, 430, 836
Ness, M., Freeman, K., Athanassoula, E., et al. 2013b, MNRAS, 432, 2092
Ness, M., Debattista, V. P., Bensby, T., et al. 2014, ApJ, 787, L19
Ness, M., Zasowski, G., Johnson, J. A., et al. 2016, ApJ, 819, 2
Nishiyama, S., Tamura, M., Hatano, H., et al. 2009, ApJ, 696, 1407
Ortolani, S., Renzini, A., Gilmozzi, R., et al. 1995, Nature, 377, 701
Pietrinferni, A., Cassisi, S., Salaris, M., & Castelli, F. 2004, ApJ, 612, 168
Pietrinferni, A., Cassisi, S., Salaris, M., & Castelli, F. 2006, ApJ, 642, 797
Pietrinferni, A., Cassisi, S., Salaris, M., & Hidalgo, S. 2013, A&A, 558, A46
Pietrinferni, A., Molinaro, M., Cassisi, S., et al. 2014, Astron. Comput., 7, 95
Randich, S., Gilmore, G., & Gaia-ESO Consortium 2013, The Messenger, 154, 47
Renzini, A., & Fusi Pecci, F. 1988, ARA&A, 26, 199
Renzini, A., Gennaro, M., Zoccali, M., et al. 2018, ApJ, submitted
Rich, R. M., Reitzel, D. B., Howard, C. D., & Zhao, H. 2007, ApJ, 658, L29
Rojas-Arriagada, A., Recio-Blanco, A., de Laverny, P., et al. 2017, A&A, 601, A140
Saito, R. K., Zoccali, M., McWilliam, A., et al. 2011, AJ, 142, 76
Santana, F. A., Muñoz, R. R., Geha, M., et al. 2013, ApJ, 774, 106
Santana, F. A., Muñoz, R. R., de Boer, T. J. L., et al. 2016, ApJ, 829, 86
Saracino, S., Dalessandro, E., Ferraro, F. R., et al. 2016, ApJ, 832, 48
Schultheis, M., Rojas-Arriagada, A., García Pérez, A. E., et al. 2017, A&A, 600, A14
Stetson, P. B. 1987, PASP, 99, 191
Stetson, P. B. 1994, PASP, 106, 250
Terndrup, D. M. 1988, AJ, 96, 884
Udalski, A., Szymański, M. K., & Szymański, G. 2015, Acta Astron., 65, 1
Valenti, E., Zoccali, M., Renzini, A., et al. 2013, A&A, 559, A98
Valenti, E., Zoccali, M., Gonzalez, O. A., et al. 2016, A&A, 587, L6
van den Bergh, S., & Herbst, E. 1974, AJ, 79, 603
Wegg, C., & Gerhard, O. 2013, MNRAS, 435, 1874
Wegg, C., Gerhard, O., & Portail, M. 2015, MNRAS, 450, 4050
Zasowski, G., Ness, M. K., García Pérez, A. E., et al. 2016, ApJ, 832, 132
Zoccali, M., & Valenti, E. 2016, PASA, 33, e025
Zoccali, M., Cassisi, S., Frogel, J. A., et al. 2000, ApJ, 530, 418
Zoccali, M., Renzini, A., Ortolani, S., et al. 2003, A&A, 399, 931
Zoccali, M., Gonzalez, O. A., Vasquez, S., et al. 2014, A&A, 562, A66
Zoccali, M., Vasquez, S., Gonzalez, O. A., et al. 2017, A&A, 599, A12

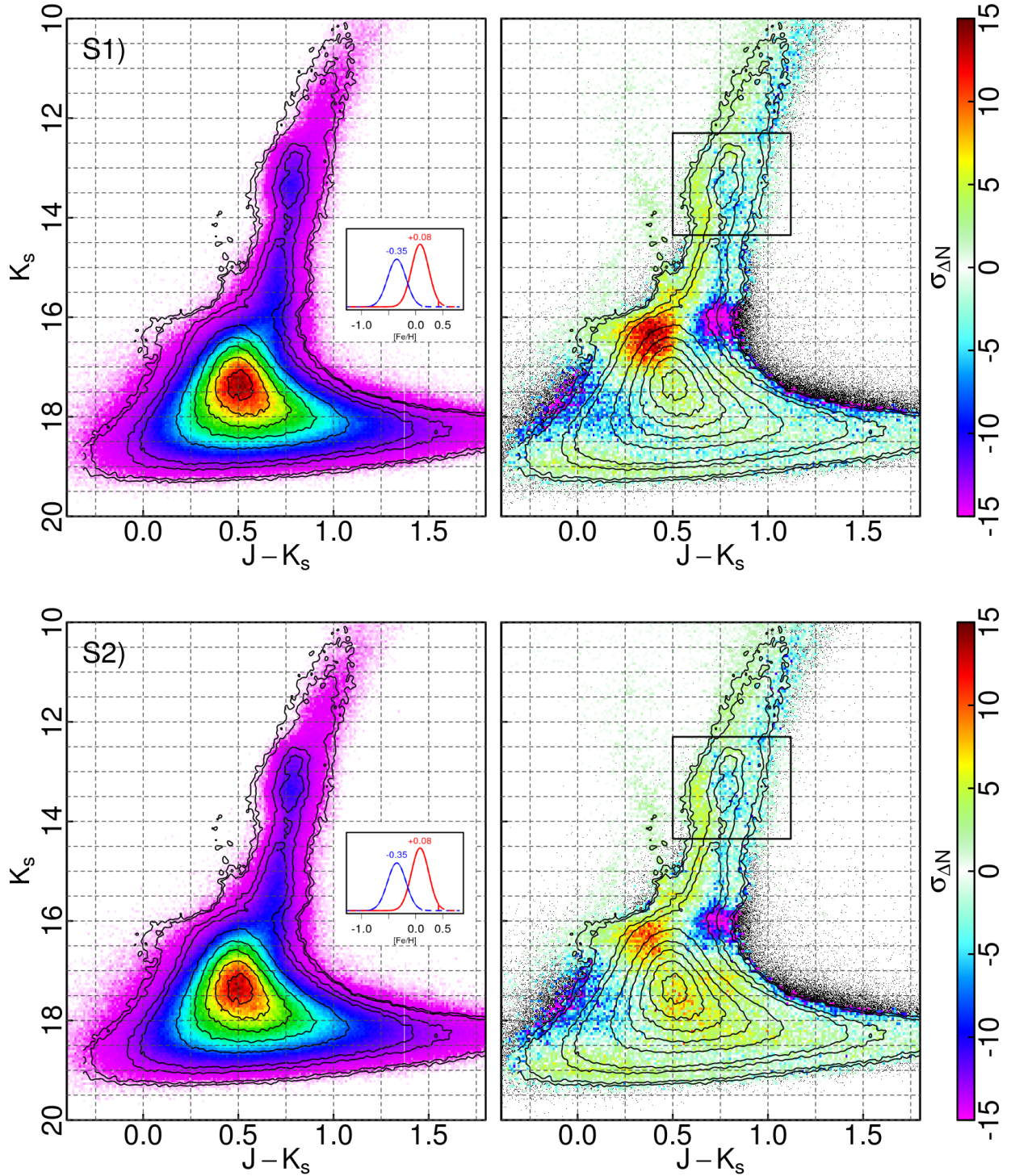## Appendix A: Additional figure

Fig. A.1. *Left panels*: synthetic CMDs corresponding to the six scenarios described in Sect. 6 compared to the isodensity curves of the observed clean sample as derived in Fig. 6 (solid black contours). The adopted MDF for all scenarios (except S4) is shown in the insets (see text). *Right panels*: residuals map providing the dispersion (e.g., the quality of the fit) between the synthetic and observed CMDs. The significant mismatch between the observations and simulations corresponds to $|\sigma_{\Delta N}| \gtrsim 7$. Regions of the CMD populated exclusively by simulated stars are indicated by black dots (see text). The boxes on the rightmost panels indicate the region of the CMD used to normalize the number of simulated stars to the observations.
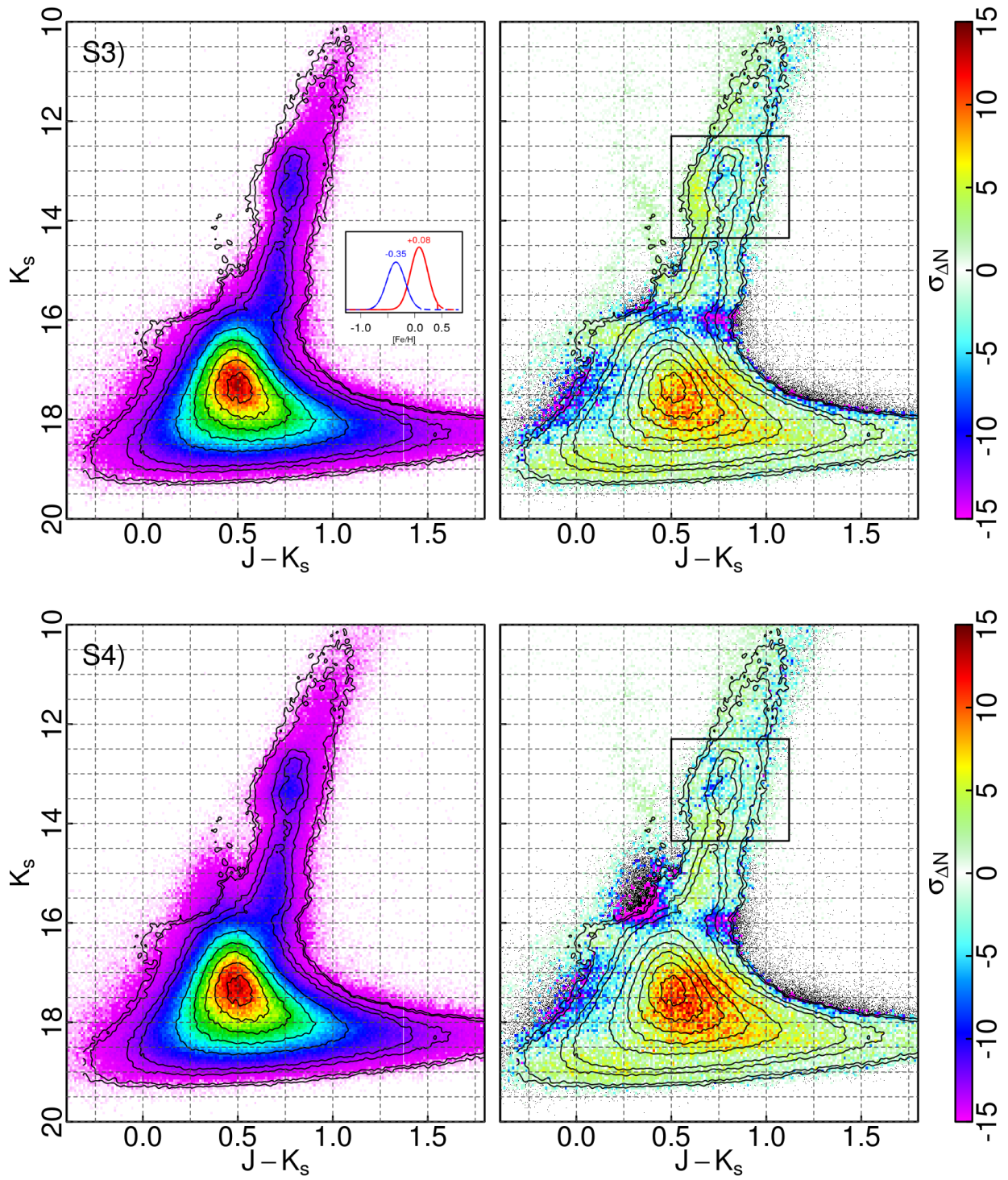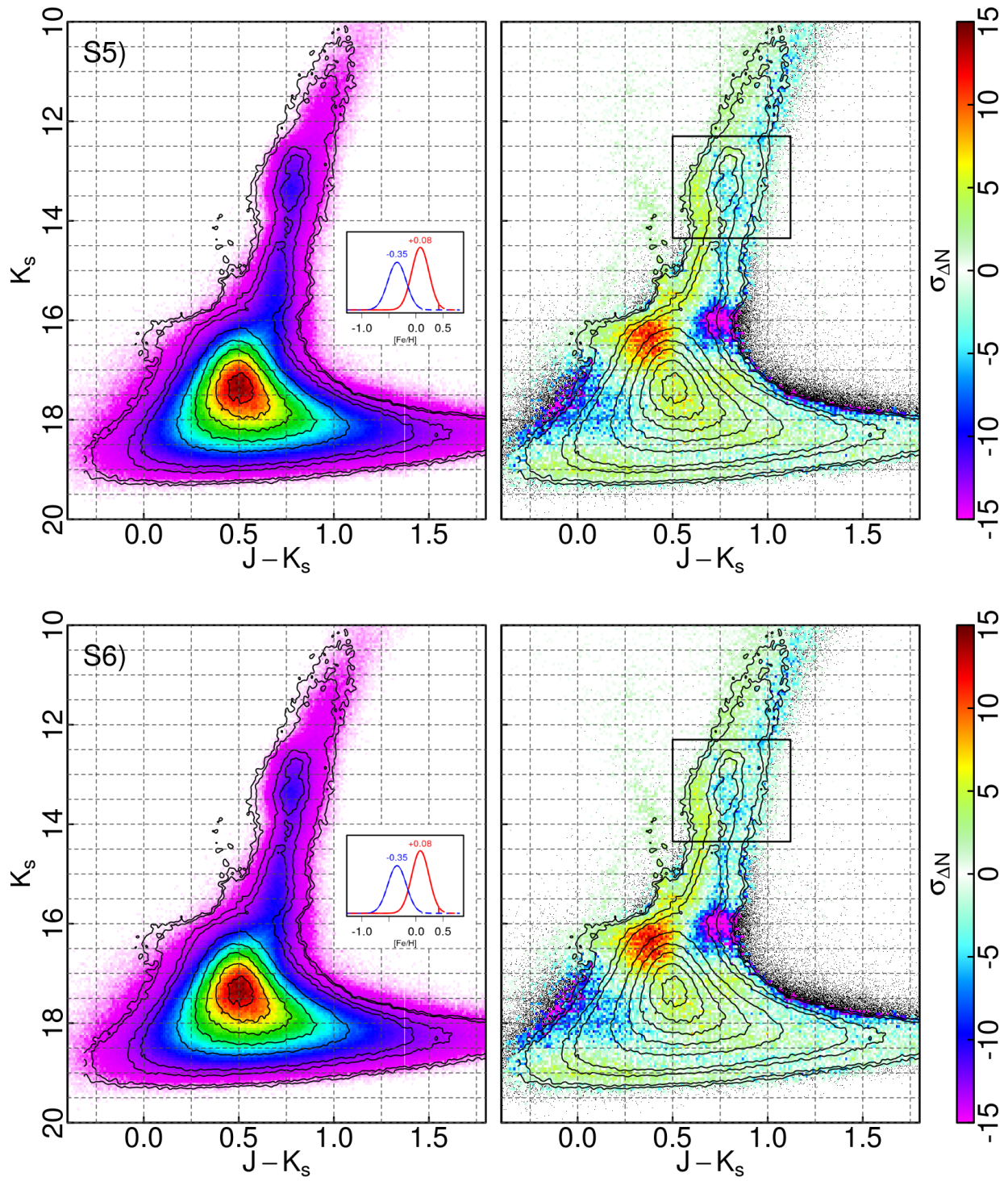
Fig. A.1. continued.

Fig. A.1. continued.