Referee's comments and suggestions

GI Perren, AE Piatti, RA Vázquez

October 21, 2016

Abstract

We thank the referee ~~very much~~ for their detailed revision of our article. Below are the responses to the referee's comments and suggestions, ~~keyed to each one as presented in the referee's report.~~

General concerns

A

As the editor points out, this was done by ~~their~~ office to shrink the size of the PDF. All the images used in the manuscript are of high quality.

For the referee and the editor's convenience, we make all images available in the following link at their full size: `https://www.dropbox.com/sh/jlrwttw2220adfp/AABwqBNNt7GzcID0CKi6s[illegible]a?dl=0`

This url will be accessible until the ~~referring~~ process finishes.

B

Taking into account the additions done to address the referee's concerns, the new version of the manuscript was shortened as shown in the following table:

—— **FINISH** ——

	Original	New version
Words	14759	XXX
Lines	3085	XXX

Comments and questions

1

The sentence was changed to improve its readability.

2

The referee is ~~absolutely~~ correct, there are *many* articles where the analysis of star clusters is not done by-eye. In Sect 2.9 of Paper I, Perren et al. 2015, we mention approximately thirty of these articles. It was certainly not our intention to convey ~~that~~ *all* studies but this one are done by-eye~~!~~

A few sentences were added to make this more clear.

3

Added a sentence to explain why these clusters were used.

4

Corrected in all instances.

5

Corrected.

6

The maximum mass was selected after a first rough estimate of the masses of the clusters in our dataset (a first pass with ASteCA), using 1×10^4 $M_\odot$ as the limit. In all but a handful of cases, the masses where estimated below 5×10^3 $M_\odot$; making the 1×10^4 $M_\odot$ Mo max limit a reasonable value. For the 15 clusters which showed total masses close to 1×10^4 $M_\odot$, we extended this limit to 3×10^4 $M_\odot$.

These details were added to the second to last paragraph of Sect 3.1.

6.1 Add

1. The 3×10^4 $M_\odot$ maximum value proved to be a very reasonable limit for the most massive clusters, as shown later in Sect 5.2.1. There, we demonstrate that even if no field star cleaning is performed (i.e., we use *all* stars in the cluster region when searching for the best synthetic cluster match), and using a max mass limit of 5×10^5 $M_\odot$, the largest cluster (NGC419) is assigned a total mass of 5.5×10^4 $M_\odot$. We know that this is an overestimated value, since we fitted not only the cluster stars but *all contaminating field stars too*, which shows that 3×10^4 $M_\odot$ is a reasonable limit.

2. The step defines the grid's density much like the width of a histogram's bin, and it is fixed throughout the best match process. The value of the step is defined by the user along with the min and max values of the parameter's range (in the `params_input.dat` file). This is true not only for the mass but for all the fitted parameters (age, metallicity, etc.) Picking a larger step would only decrease the accuracy of the parameter estimate, and it would have very little (if not no) impact on the final selected mass value (or any other parameter).

7

First: added a portion of the observed cluster region above the image, as requested. The caption was modified to reflect this change, and to make the description of the overall plot clearer.

Second: that must be related to the compression performed by the editorial office. Please see the corresponding image shared in the url presented in Sect. A.

8

The limiting magnitude is taken individually for each observed/processed cluster. It is the limiting magnitude of the observed cluster itself, ~~meaning it will vary~~ with each observed cluster. This is explained in Paper I, Sect 2.9.1. At the beginning of Sect 3.4 (footnote 17) we sate that the details of the code's built-in functions are explained in Paper I. We did not repeat them here to avoid redundancy.

We added a sentence in the caption of Fig. 2 explaining this.

9

As explained in Paper I Sect 2.9.2, we employ a genetic algorithm (GA) to find the best synthetic cluster match for each observed cluster. The GA explores the entire paramete~~rs~~ space, which is composed – in this case – of approximately 2.3×10^7 possible models/solutions (i.e: synthetic CMDs), as shown in Table 3. In this work, each observed cluster was compared by the GA to roughly 1.5×10^6 models (including those explored during the bootstrap runs). For each of these models, a *unique* random IMF sampling is ~~performed t~~o populate its isochrone.

We added a sentence in this paragraph, and at the end of the section explaining this.

10

Changed "usually" to "often" as suggested.

11

Changed "versus" to "of" as suggested.

12

The fonts of labels and values In fig. 3 have been increased, as requested.

13

We use z instead of [Fe/H] for two reasons. The ~~primary~~ reason is that this is the default form in which the evolutionary tracks are ~~generated~~ by Girardi's CMD

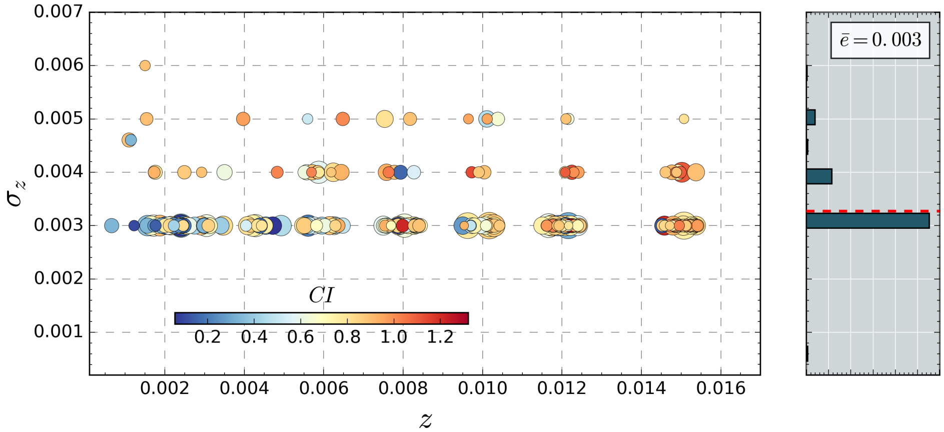

Figure 1: Error in z.

service (`http://stev.oapd.inaf.it/cgi-bin/cmd`). ~~These *Padova* isochrones are the ones `SteCA` uses to produce its synthetic CMDs.~~ ~~In a later stage of the code, I'd~~ like to add other sets of tracks (Darmouth, Geneva, BaSTI, MIST-MESA, etc.), ~~but for now this set is the only one supported.~~ The ~~secondary~~ reason is that it is easier to develop the code when all fundamental parameters are strictly positive.

The errors are displayed using [Fe/H] to make the reader aware of the issue that [Fe/H] errors will increase for low metallicities, for a mathematical reason. If we had plotted the errors using z instead, we'd get a diagram that is not much informative, as the errors in z are (as stated in the article) very clustered around the $z = 0.003$ value. This can be seen in Fig. 1.

A footnote was added in paragraph 3 of Section 4, to address this concern.

14

We meant "low" compared to solar; perhaps this wasn't the best choice of words. The "confirmation bias" effect explains that researchers will tend to select values that have already been used, which explains the clustering of [Fe/H] around the -0.4 dex and -0.7 dex values in the literature.

We've re-written that paragraph adding a more detailed explanation of the effect.

15

The referee is correct, we did not mean to imply that there was not a good reason for doing this.

We've added this information to the paragraph.

16

We are not saying that the distance moduli is fixed *in all studies*. We are referring to the "literature" articles (the 19 shown in Table 1), were the distance moduli were indeed always assumed to be fixed ~~constants.~~ That's what we meant by saying "in the latter", referring to the previous sentence which mentioned these literature articles.

We have changed the wording to explicitly mention the "literature articles".

17

Corrected here, and also in the second to last paragraph of Sect 5.2 where the same name for the plot was ~~also~~ used.

18

On the determination of cluster masses

The mass estimation done by `ASteCA` via the GA relies on the same exact method used to estimate the other fundamental parameters (metallicity, age, etc.) This is explained in Paper I, Sect. 2.9.1 and 2.9.2. The only thing that changes from Paper I to this one, is the likelihood function. The Dolphin (2002) likelihood is used here, because it allows the estimation of mass through the comparison of the number of stars in cells (2D bins) of the CMD's histogram. Basically, this amounts to a comparison between the "Hess diagrams" of an observed cluster versus that of the synthetic clusters ~~generated.~~

When a synthetic CMD of given metallicity and age is created via a theoretical isochrone (in our case, we used the Bressan et al. (2012) set), the *entire* theoretical isochrone is taken into account, down to the faintest magnitude. Given a total mass value, this isochrone is populated through sampling of the selected IMF (in the case of this work, we used the Chabrier (2001) log-normal IMF, as mentioned in Sect. 3.4) until that fixed total mass value is achieved. This is: the synthetic CMD is generated accounting for *all* stars in the theoretical isochrone~~, even those low mass stars beyond the observed cluster's magnitude limit.~~

After this, the synthetic CMD is perturbed by the same processes that affect the observed cluster, i.e., limiting observed magnitude, incompleteness of observed stars ~~at the fainter end~~, and photometrical errors. This is done to make the synthetic CMD comparable to the observed one. The two first processes will effectively remove stars from the synthetic CMD, particularly the limiting magnitude effect. These stars are removed because *there is nothing to compare them to*, since the observed cluster does not contain stars beyond its magnitude limit. Even though these stars are removed prior to the fitting of a synthetic CMD (with the observed CMD), the mass associated with this cluster *remains the total mass used to create it*.

For example: a synthetic CMD of 10000 $M_{\odot}$ is generated, and after the removal of faint stars beyond the magnitude limit its mass is "reduced" to

2000 $M_\odot$. I.e., the mass of the stars that are visible in the synthetic CMD, **after** the magnitude limit removal, is 2000 $M_\odot$. If this synthetic CMD results the one with the best match to the observed cluster (i.e., its likelihood value is the maximum found for all the synthetic CMDs processed), the estimated mass assigned to the observed cluster will be 10000 $M_\odot$, **not** 2000 $M_\odot$, because that is the actual mass that was used to generate the synthetic CMD.

To be clear, at no point during our analysis do we make use of either integrated magnitudes or M/L relations to estimate masses. The mass estimation depends exclusively on the maximization of ~~the~~ the PLR. There is no extrapolation performed since none is needed: *all* stars are taken into account at the moment of the synthetic CMD's generation. Fig. 8 in Paper I shows the full process of generating a synthetic CMD.

On the mass fraction outside the radius

We are not entirely sure what the referee is ~~discussing~~ when they mention the "fraction of the cluster's mass that is located outside the cluster radius". Are they talking about the dynamical effect of mass loss due to evaporation or two-body relaxation? If this is the case, given that mass loss tends to affect low-mass stars preferentially (Vesperini 2010), this issue will have a minor impact on our mass determination. This is because we asses the total cluster mass by directly comparing the visible portion of the cluster's CMD with a large number of synthetic CMDs (after affecting these synthetic CMDs with similar processes to that which affect the observed cluster, like the limiting magnitude for example). Since the portion of lowest mass stars is not observed (due to the limiting magnitude), it will not affect the matching process.

This shouldn't be confused with the assumption that low-mass stars are not considered at all in the mass estimation process. They are indeed taken into account at the moment of the synthetic CMD's generation, when the fixed total mass of the synthetic cluster is distributed along its theoretical isochrone via IMF sampling.

If the referee is concerned about luminosity being lost outside a certain measurement aperture, and thus affecting the mass estimation, this is not an issue for `ASteCA`. The observed cluster's mass is obtained (as stated in the subsection above) without making use of integrated magnitudes or M/L relationships. The code relies entirely on the comparison of the distribution and number of stars in the observed CMD, with that of many synthetic CMDs.

If the question was aimed ~~towards~~ the more simple issue of some stars being left out of the selected cluster region, we do not believe this ~~to be an effect of much importance.~~ Because we employ the radius of the cluster as the limit where the star density falls to that of the field, we do not expect to be missing any significant portion of the observed cluster.

We added a couple of sentences to the first, second, and third paragraphs of Sect. 3.4 in the new version of the article, which should address the doubts regarding total mass estimation.

19

On the M/L relation of the H03 and P12 articles

If we ~~understand~~ correctly, the referee is asking us to calculate a normalized mass via a relation of the form

$$M_{P12,norm} = M_{P12}\frac{(M/L)_{H03}}{(M/L)_{P12}} = \frac{M_{H03}L_{P12}}{L_{H03}} \quad (1)$$

where M_{H03}, $(M/L)_{H03}$, M_{P12}, $(M/L)_{P12}$ are the masses and mass-luminosity relations of H03 and P12 respectively, and $M_{P12,norm}$ is the mass of P12 normalized to the $(M/L)_{H03}$ relation.

If this is the request, we must point out that P12 used the same database of integrated magnitudes as H03, which means that the luminosities used for each cluster are equal ($L_{P12} = L_{H03}$). This would reduce Eq. 1 to

$$M_{P12,norm} = M_{H03} \quad (2)$$

meaning the mass differences would be zero for all clusters. If this is *not* what the referee ~~had in mind~~, we ask ~~them~~ to please expand a bit on ~~their~~ request.

It is also important to point out that P12 ~~does~~ not employ a standard mass-luminosity relationship to estimate masses. Rather, the authors use their MASSCLEAN*age* package (which in turn depends on their MASSCLEAN*colors* package, Popescu & Hanson 2010a,b). This tool allows the simultaneous determination of the most probable age-mass values, from the observed integrated magnitudes matched in a hyper-space of synthetic integrated magnitudes. These articles by Popescu et al. also highlight the enormous degenerations in the solutions that arise, when simple integrated magnitude methods are used for mass and age estimations.

On the mass discrepancy of NGC 419, 1917, 1751

We thank the referee ~~very much~~ for taking the time to track down and calculate the mass values presented in this section, for these three clusters.

Obtaining mass and age estimates from integrated magnitudes is known to be a process affected by large stochasticity (see for example the Popescu & Hanson articles). Below we show diagrams for each of these three clusters, where we explain in detail how different mass values affect the generated synthetic CMD.. This should help understand why we believe our mass estimates are ~~correct~~, and the large mass values given in the integrated magnitude articles are overestimated.

* ***NGC 1917***, Fig. 2

- Panel **(a)**: CMD of the observed cluster region. In our original analysis we used a radius of 27.4 arcsec which resulted in the 4000 $M_{\odot}$ total mass estimation. Here, we use instead a larger radius of 62 arcsec, equal to the aperture size used by van den Bergh (1981). N_{accpt} is the number of stars that were not rejected due to large photometric errors, and n_{memb} is the number of approximated cluster members. This last value is obtained by averaging the number of stars in the surrounding field (next panel) and subtracting that value from the total number of stars within the cluster region. The final result is ~169 expected cluster members.
- Panel **(b)**: CMD of ten combined surrounding field regions, each with an area equal to that of the cluster.
- Panel **(c)**: CMD of the cluster region after the decontamination algorithm (DA) was applied. Stars are colored according to the membership probabilities (MPS) assigned, semi-transparent stars are those removed by the cell-by-cell density based decontamination process. N_{fit} is the number of stars that are left after this removal. Ideally, n_{memb} and N_{fit} should be similar, since they are both rough (independent) estimates of the (visible) number of cluster members.
- Panel **(d)**: best match synthetic cluster found using the larger radius of 62 arcsec. Its fundamental parameter values are fixed to z=0.008, log(age)=9.15, $(m-M)$=18.48, and $E_{(B-V)}$=0.08, as obtained originally by `ASteCA`. The mass is allowed to vary in a range between [500, 1×10^5] $M_{\odot}$. N_{synth} is the number of visible stars in this best match synthetic CMD (notice that this value is rather close to N_{fit}, the approximated number of cluster members), and M=$1.1\times10^4\,M_{\odot}$ is the *total mass used to generate this synthetic CMD* (**not** to be confused with the portion of mass visible in the CMD, which is much smaller). We see that this mass is larger than the 4000 $M_{\odot}$ originally estimated. This is because the used radius is now more than twice as large, and because for relatively low mass clusters the stochasticity in their mass assignment is non-negligible.
- Panel **(e)**: best match synthetic CMD using the same fundamental parameter values and mass range shown in **(d)**, where the *full cluster region* with no decontamination procedure applied was used. I.e., we fit here the ~ 1700 stars in the observed cluster region, which will obviously include a majority of field stars. The best match estimated mass increases to $4.9\times10^4\,M_{\odot}$. This shows that even if no field star cleaning is performed, the estimated mass would still not be close to the mass given in articles where integrated magnitudes were used.
- Panel **(f)**: synthetic CMD generated using the same fundamental parameters mentioned in **d**, and a fixed mass of $8\times10^4\,M_{\odot}$ (average mass value estimated by H03, P12, and the referee's own calculations). The number of stars in this CMD is more than two times larger than the stars present in the CMD of the observed cluster region, *even if all contaminating field stars are taken into account.* It is visible at plain sight, with no statistical estimator needed, that the large

mass value given in other articles produces a number of stars in the best match synthetic CMD, that is **not compatible with the observed cluster**.

* ***NGC 419***, Fig. 3

- Panels **(a)**, **(b)**, and **(c)** are equivalent to those from Fig. 2. The radius used in Goudfrooij et al. (2014) to derive their mass estimate of $\sim 2.2\times10^5\,M_\odot$, is 50 arcsec. In our original analysis of this cluster we employed a larger radius of 85 arcsec, which is also used here.
- Panel **(d)**: original mass estimate, where we see that the number of stars in the synthetic CMD is rather close to the estimated number of cluster members.
- Panel **(e)**: best match estimated mass using the full cluster region, with no field star cleaning performed. The fundamental parameters were fixed to z=0.012, log(age)=8.95, $(m-M)$=18.92, and $E_{(B-V)}$=0.02, as obtained originally by `ASteCA`. The mass is allowed to vary in a range between [500, 1×10^5] $M_\odot$. The best match mass is twice as large as the mass value obtained using the decontaminated cluster region.
- Panel **(f)**: how the CMD *should look* if this cluster had a mass of $\sim 2.7\times10^5\,M_\odot$, as estimated on average by H03, and Goudfrooij et al. Again, we see that the number of stars in this large mass synthetic CMD, is **not compatible with the observed cluster**.

* ***NGC 1751***, Fig. 4

- Panels **(a)**, **(b)**, and **(c)** are equivalent to those from Fig. 2. The radius used in Goudfrooij et al. (2014) is 50 arcsec. We use here a radius of 60 arcsec, the same employed in our original analysis of this cluster.
- Panel **(d)**: original mass estimate.
- Panel **(e)**: best match estimated mass using the full cluster region, with no field star cleaning performed. The fundamental parameters were fixed to z=0.012, log(age)=9.1, $(m-M)$=18.6, and $E_{(B-V)}$=0.04, as obtained originally by `ASteCA`. The mass is allowed to vary in a range between [500, 1×10^5] $M_\odot$. The best match mass is four times the mass value obtained using the decontaminated cluster region.
- Panel **(f)**: how the CMD *should look* if this cluster had a mass of $\sim 7.2\times10^4\,M_\odot$, as estimated on average by P12, H03, and Goudfrooij et al. Once more, the number of stars in this large mass synthetic CMD, is **not compatible with the observed cluster**.

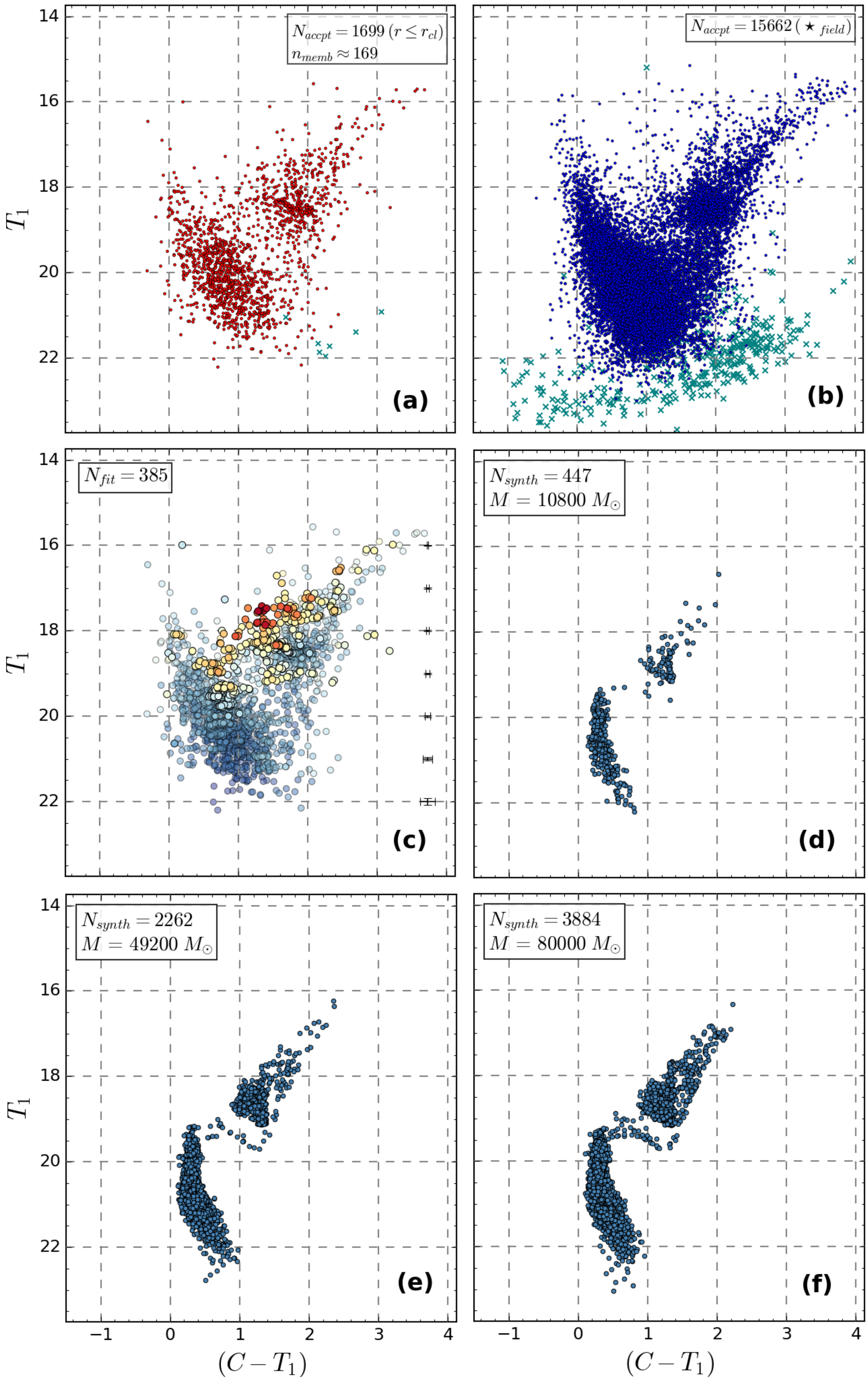

Figure 2: Analysis of NGC1917.

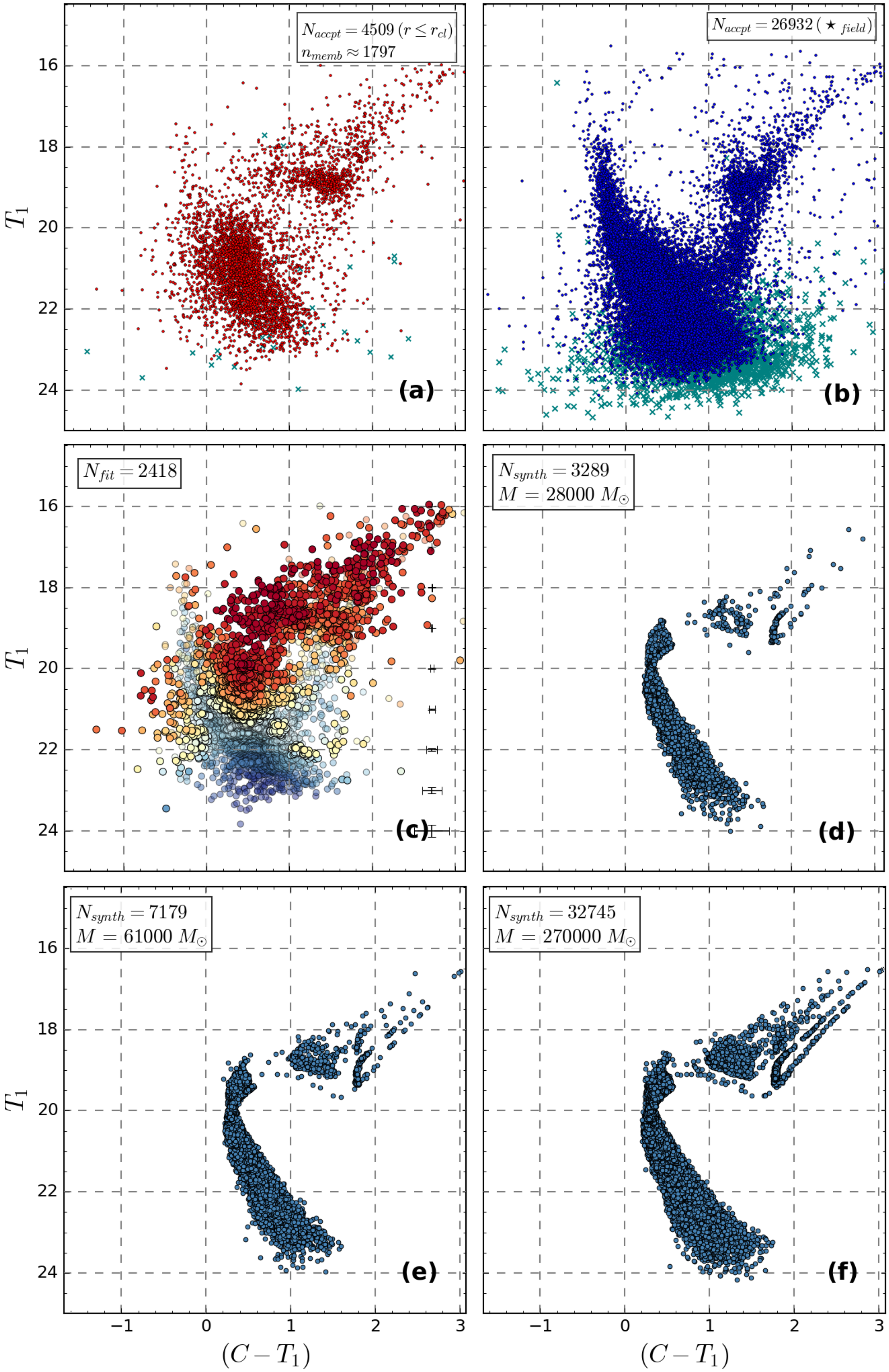

Figure 3: Analysis of NGC419.

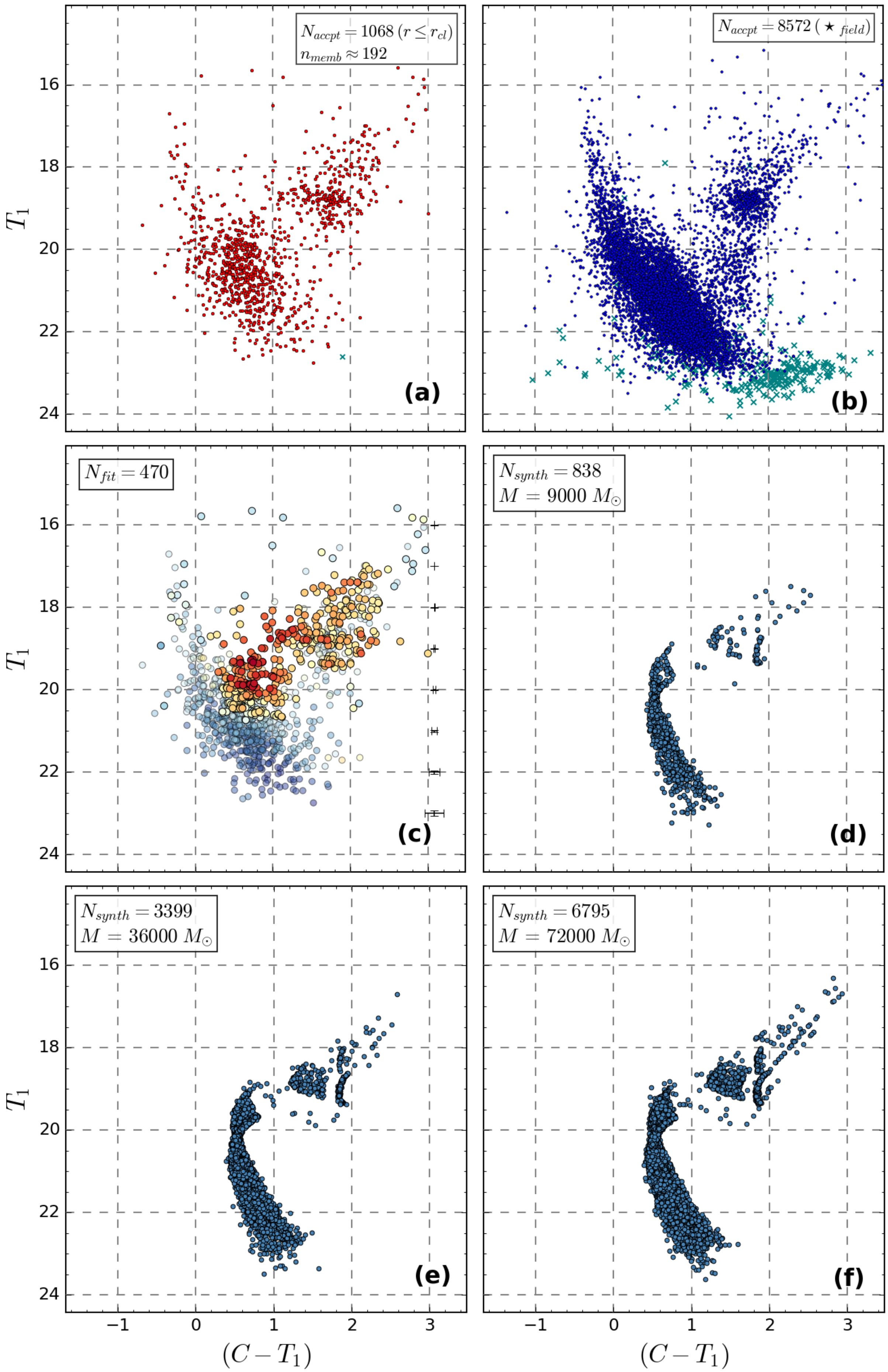

Figure 4: Analysis of NGC1751.

* *MASSCLEAN synthetic CMDs*

The referee could of course question the ability of `ASteCA` in producing these synthetic CMDs, of given metallicities, ages, and masses. To further strengthen our analysis, we present in Fig. 5 three synthetic CMDs generated with the MASSCLEAN tool (Popescu & Hanson 2009).

These CMDs were created using the same fundamental parameters assigned to NGC 1917, 419, and 1751 by `ASteCA`, but using an averaged total mass obtained from the values given to each of them by H03, P12, and the articles mentioned by the referee. This means that these CMDs are the equivalent of those in panels **(f)** of Figs. 2, 3, and 4, but produced with a completely independent tool.

MASSCLEAN does not support the Washington photometric system, hence the V versus $(B-V)$ diagrams. These synthetic CMDs are affected by a comparable limiting magnitude (basically $V_{lim} = T1_{lim} + 0.5$ mag) and incompleteness, as those affecting the observed CMDS for each of the three clusters. The differences with the synthetic CMDs produced by `ASteCA` are due to the stochastic nature of IMF sampling (MASSCLEAN uses a Kroupa-Salpeter IMF, we used the Chabrier 2001 log-normal IMF), the theoretical isochrones set (MASSCLEAN uses the old Marigo et al. 2008 set, while our code employs the newer Bressan et al. 2012 set), and binarity (MASSCLEAN does not support the addition of binaries).

Again, we can clearly see that these synthetic CMDs **do not** represent the CMDs of the observed clusters. If these observed clusters had the large masses estimated by other works, the number of stars that *should* be present in their CMDs would need to be *much larger* than the number we actually observe. On average, we should observe between 120%-570% more stars if we compared with the entire cluster region (no decontamination process applied), and between 1200%-1900% if we compare with the cluster region after removing the contaminating field stars.

* *Concluding remarks*

The conclusion that can be drawn from this analysis is thus equivalent to that presented in the article: these Magellanic clusters have had their masses overestimated by a substantial amount in the H03 and P12 works (and others too, i.e. the Goudfrooij et al. 2014, article mentioned by the referee). The link between these studies is the use of integrated magnitudes to derive masses.

Finally, to reinforce this notion we refer the referee to Appendix A of the original article. The validation there performed shows that `ASteCA` recovers masses with very reasonable accuracy. Furthermore, the masses are recovered with *increased accuracy*, as the clusters grow larger. Having used an external package (MASSCLEAN) to generate the almost 800 synthetic clusters used in this validation, there is no possibility for internal biases in the mass determination. Again, this leads to the same conclusion: `AsteCA` is performing a proper estimation of masses within its uncertainties, and the cause of the discrepancy between the masses must be located in those articles where the classical method of mass estimation via integrated magnitudes is employed.

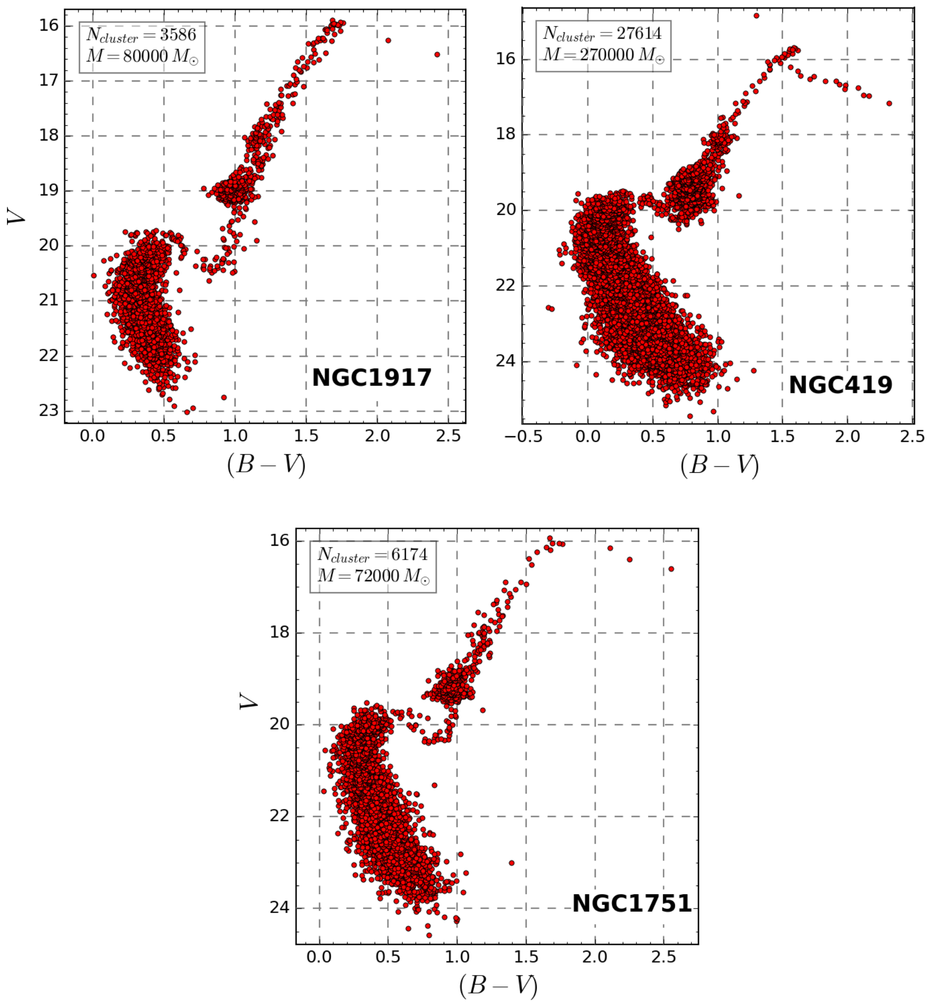

Figure 5: Synthetic CMDs for NGC 1917, 419, and 1751 produced by the MASSCLEAN package, using their large mass estimates.

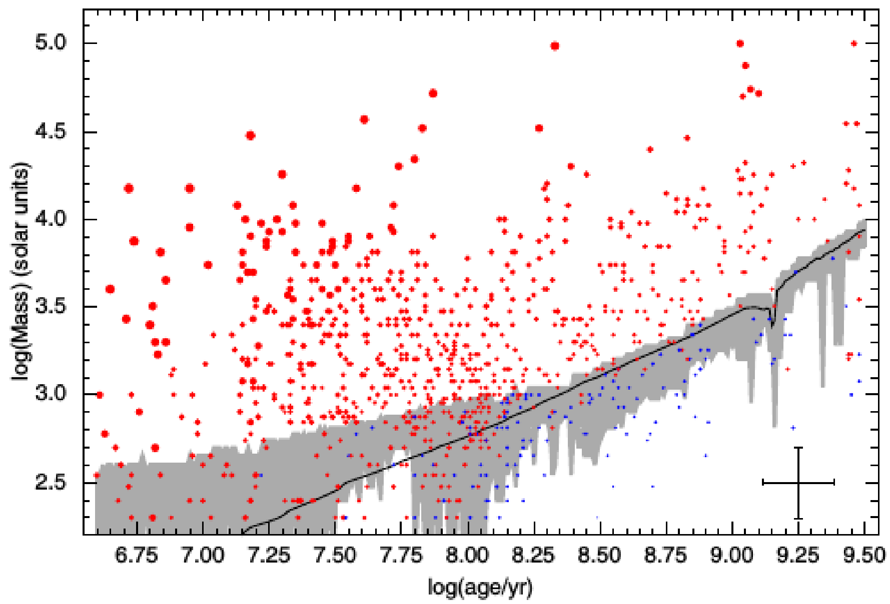

Figure 16. LMC clusters brighter than the $M_V = -4$ mag observed magnitude limit are represented as red dots, while the clusters fainter than this limit are shown as blue dots. The size of the dots is scaled with the M_V magnitude. The black line represents a traditional fading line. The gray zone represents the range over which clusters of that mass and age may or may not be seen to exceed the $M_V = -4$ mag observed limiting magnitude.

Figure 6: Figure taken from Popescu et al. (2012, Sect. 4)

20

Corrected, as suggested by the referee.

21

Corrected, as suggested by the referee.

22

As the luminosity of a star cluster decreases with age, the oldest clusters that can be detected ~~will be~~ those with the largest masses (and hence larger radii) This is discussed for example in Popescu et al. (2012, Sect. 4). Fig. 6 shows Fig. 16 from that article, where a fading limit is shown (black line) and the LMC clusters can be seen to increase in mass as their ages increase. We've added a reference to this effect and the above mentioned article.

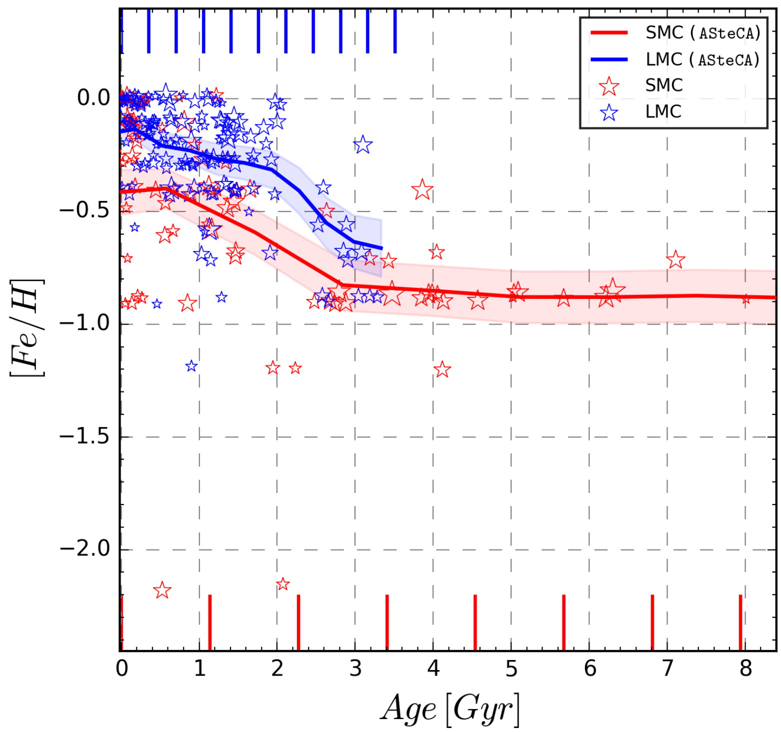

Figure 7: `ASteCA` AMR without the three old LMC clusters with high metallicities values.

23

The referee is right to point this out (this discussion was actually in the first draft of our article, but we cut it to save space). If this group of three "high" metallicity and "old" age ($\log(age) \geq 9.5$) LMC clusters is removed, the peak disappears and hence, so does the drop. In Fig. 7 we show how the LMC's AMR would look without these three clusters present. We've added a couple of sentences in the corresponding paragraph of the article, addressing this effect.

24

We believe we have addressed the issues brought up by the referee in this cover letter, particularly in Sects. 18 and 19, regarding the mass assignment performed by the code. As such, seeing no reason to alter them, we stand by our original conclusions.

References

Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127

Chabrier, G. 2001, ApJ, 554, 1274

Dolphin, A. E. 2002, Monthly Notices of the Royal Astronomical Society, 332, 91

Goudfrooij, P., Girardi, L., Kozhurina-Platais, V., et al. 2014, The Astrophysical Journal, 797, 35

Marigo, P., Girardi, L., Bressan, A., et al. 2008, A&A, 482, 883

Popescu, B. & Hanson, M. M. 2009, The Astronomical Journal, 138, 1724

Popescu, B. & Hanson, M. M. 2010a, ApJ, 724, 296

Popescu, B. & Hanson, M. M. 2010b, ApJ, 713, L21

Popescu, B., Hanson, M. M., & Elmegreen, B. G. 2012, ApJ, 751, 122

van den Bergh, S. 1981, A&AS, 46, 79

Vesperini, E. 2010, Philosophical Transactions of the Royal Society of London Series A, 368, 829