# Answers to referee's report

We would like to thank the referee once again for his/her report. We have addressed all the corrections/comments/suggestions made by the referee.

The Major Point was divided into 6 sub points addressed separately for clarity.

## Major point 10

1.

> *You confirmed in the answer MI-43 that the membership probability is indeed used as weight in the fit, but for my surprise, it has a smaller effect in the fit, whereas a large number of faint stars with higher photometric uncertainties and lower membership probability have a larger effect in the fit.*

Although the membership probabilities (MPs) are taken into account in the fundamental parameters fitting process, they have a limited influence. This is because the likelihood function employed (*Dolphin 2002*[1], Eq 10) is a binned statistic, meaning bins (in our case 3-dimensional cells, since we use a 3D photometric space: G, BV, VI) are compared between the observed and synthetic clusters. The MPs are currently used to weight each bin by the mean of the MPs of all stars that fall within it. We added this weighting to the likelihood when developing the code. If you look at Eq 10 in Dolphin (2002) you'll notice that MPs (or weights of any kind) play no role there. Seeing that the weights results in a very small influence over the final estimation -and complicate the likelihood process-, in the upcoming version of ASteCA (v0.3.0) they will be removed.
In other words: MPs are mainly used to clean the sample, and although they are also used in the fitting process their impact here is minimal.

An approach that would give more importance to the individual MPs in the final parameter estimation is using the per-star likelihood function developed by Tolstoy & Saha (1996)[2] -which ASteCA also supports-, instead of the current Dolphin (2002) likelihood. The reason for not using this likelihood is because it has two drawbacks: 1. it is over 10x slower than a binned statistic (given that it compares each observed star within the cluster region with each star in a synthetic cluster), and 2. it is not able to fit the mass of the cluster. The performance could perhaps be improved by the use of a Python just-in-time compiler (like Numba) but this is not guaranteed.

---

[1] "Numerical methods of star formation history measurement and ...."
https://ui.adsabs.harvard.edu/abs/2002MNRAS.332...91D/abstract
[2] "The Interpretation of Color-Magnitude Diagrams through ...."
http://adsabs.harvard.edu/full/1996ApJ...462..672T

2.

*> As you said in MA-11, your concern of avoid biasing the results by hand-picking stars was present in the updated analysis, but still you have shown that if you cut off the fainter stars from NGC4349 and if you select only high-probable members of RUP87 the isochrone plot seems to represent better the CMDs, with a significant change in the fitted parameters.*

The case of RUP87 shows the issue we mentioned in the previous report (and also in the manuscript): one must be very careful with the selection of probable members within the cluster region, particularly when dealing with sparse clusters. Different choices for cleaning the photometric diagram of field stars (MP cut, magnitude cut, number of stars selected, etc..) could lead to wildly different (purported) “clean” sparse cluster sequences, and thus to the estimation of dissimilar final parameters. This is seldom (if never) taken into account in the usual analysis of stellar clusters. Some researchers resort to applying no cleaning method at all, but this is even a worse choice -again, particularly for sparse clusters- as one could end up fitting mostly a sequence associated with the surrounding field.

The current selection of stars in the manuscript for RUP87 (MPs>0.6 and G>18) was done at the request of the referee to see if a more “clean” sequence could be fitted. As shown, indeed it can; but this doesn’t mean that there is a cluster there.

The aforementioned issues are why we prefer to apply density-based methods of stars removal within the cluster region (included in ASteCA). Although these might result in somewhat “dirtier” sequences (particularly for sparse clusters) they are based on a physical motivation, rather than a by-eye search for a (purported) hidden cluster sequence. There is clearly still work to be done in this area of the analysis in order to make the results more robust.

3.

*> ...it still calls the attention that depending on whether you make a selection of stars before the fit or not, the parameters change by more than their error bars. And as far as I understand, the strength of ASTECA is exactly the attempt to be as blind and objective as possible during the fitting process, which is notable, and, in my opinion, should be kept to strengthen the results of the paper.*

The referee is correct in noticing that ASteCA found two different sets of fundamental parameter values when these two clusters were re-analyzed. It is worth noting nonetheless that both sets overlap in their 1-sigma region (~16th-84th percentiles) for most of the parameters, and in their 2-sigma region (~2th-98th percentiles) in the most extreme cases. This can be seen in **Fig 1** for NGC4349, where the parameter distributions for both runs are shown (first run on the left, second run on the right). The case of RUP87 is similar and thus not shown.

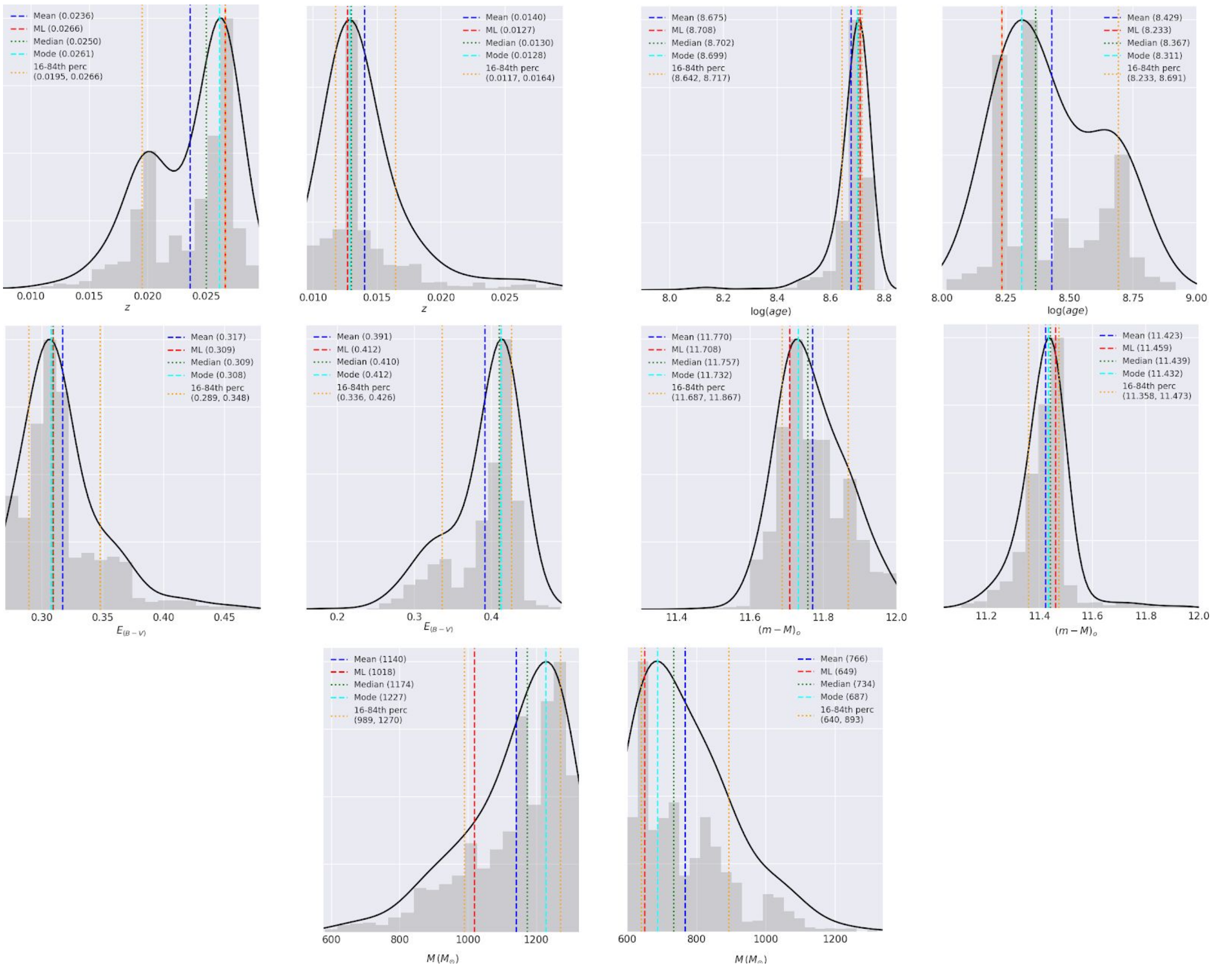


**Fig 1: parameters distribution for the first run (left) and second run (right) of the cluster NGC4349**

The answer to why ASteCA returns different parameter values for these two runs (even though, as stated above, they are still within their 2-sigma regions at most) is simple: different results were obtained because different input samples were analyzed. If we were to re-process the *exact same* sample, ASteCA would give pretty much the exact same final estimates (assuming of course proper convergence of the method). In this case though, we did not re-process the *exact same* samples: we modified them by applying magnitude and/or MPs cuts, prior to the fundamental parameters analysis. This means that a different input was fed to the code in each run.

Although the parameters of the first and second runs overlap in their 1-2 sigma regions in the case of NGC4349 (which is a desirable feature), we have come to the conclusion the second run is a sub-standard result (a local minima). More on this below.

4.

> *Could you please explain why ASTECA did not recover parameters (represented by an isochrone) that reasonably match the high-probable members stars in the first version of the paper, in particular for NGC4349 and RUP87?*

The case of RUP87 is simple: photometric diagrams for this cluster show a very dispersed group of stars with a smaller subset of stars with larger MPs that *might* resemble a poorly populated cluster sequence. This dispersion clearly affects the synthetic cluster match process. It was only after these diagrams were extensively cleaned that the method was able to fit a “cluster sequence” made up of large MPs stars. Notice that even if the fit might “look” better, the final conclusion on this stellar region is not altered (i.e., it is not a real cluster).

In the case of NGC4349, the first run of ASteCA *did* in fact recover a reasonable synthetic cluster match with an associated isochrone (whose trace should **not** be used to test the goodness of the fit, as stated previously) that follows closely almost all of the stars with large MPs. It is only for a handful (around 5) of the brightest stars that the isochrone can be said to be located at some distance. The algorithm searches for a solution that produces the best possible fit taking the *entire sequence* into account. One single star (even if it is bright and has a large MP) can not determine the fit of the entire sequence. That is what a researcher would normally do in a by-eye fit, which is not a valid statistical criteria no matter how “good” the isochrone might look.

With the removal of stars with G>18 mag in the second run the code appeared to have found a solution where the isochrone traced perfectly the bright star with large MP, but we’ve now realized that this is just a local minima solution[3]. If we let the code run for more steps, it moves out of that local minima to a solution much like the one found in the first run[4].

The root of the difficulty in fitting NGC4349 is the known fact that young clusters generally lack RC and RGB regions. This makes it very hard to estimate the position of the post-main sequence section of the isochrone, mostly determined by age and distance. It is important to note that this is *not* a shortcoming of the code, but rather an inevitable consequence of the degeneration of the solution space which is inherent to the physics of the problem. When a researcher makes a by-eye isochrone fit they are entirely disregarding all these issues.

---

[3] In our haste to respond as soon as possible to the previous reviewer’s report, we concluded the second run not noticing this. This was an oversight on our part and we apologize for the inconvenience.

[4] When employing any numerical optimizer (Genetic Algorithm) settling on a local minima is a risk that cannot be avoided, as there is no mathematical way of knowing that the global minima is ever found. The only remedy against this is to run the code for more steps and starting from different positions in the solution space. If it consistently converges to the same value, you can be pretty confident that you’ve found the global minima or at least something very close to it. The second run solution is within the 1 sigma region of the first run solution as we noted earlier, which means it is not completely “wrong” just not entirely optimal.

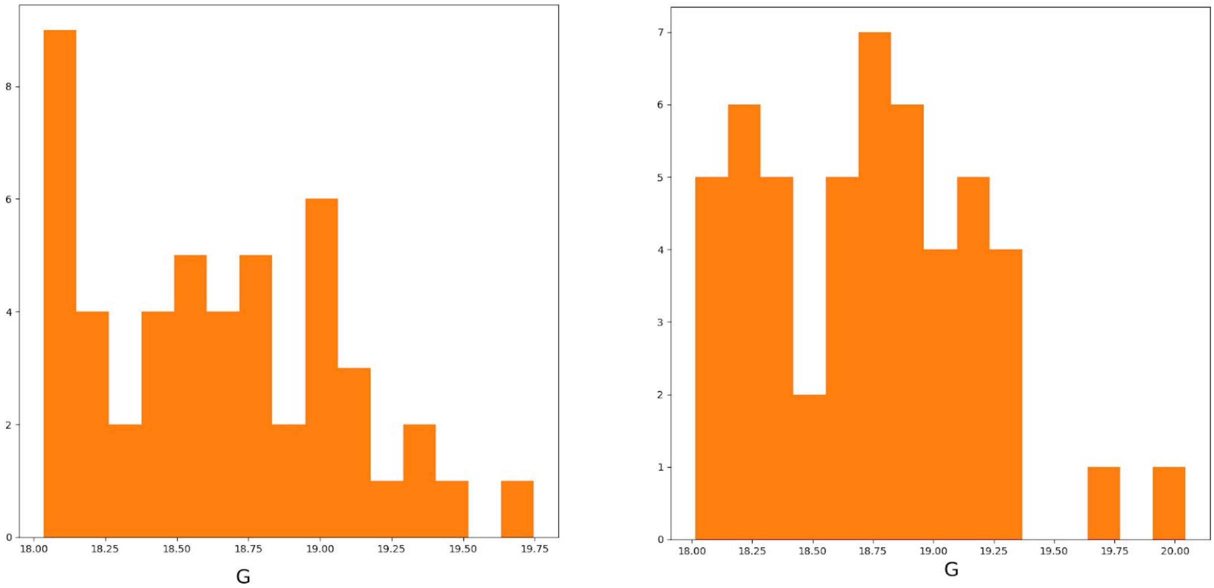


**Fig 2: G magnitude histogram for NGC4349 synthetic clusters of equal masses and remaining parameters as in the first run (left) and second run (right)**

With all the above in mind, we performed a third run for NGC4349 (included in the new manuscript) this time increasing the radius by 50%. We discovered that this allows for a small RC to show in the CMDs, which makes the best synthetic match algorithm better constrain the position of the post-main sequence section. The changes in the final parameters are small, within the 1-sigma regions of both the 1st and 2nd runs, but we feel this result is more robust. We apologize for making a non-requested change to the manuscript at this point in the review process, but we felt it was necessary.

5.

> *... the differences between the isochrones in this magnitude limit is negligible, which leads to the conclusion that the fainter magnitudes and colours did not affect the fitting considerably.*

The referee is correct that the *isochrones* are very similar in this magnitude range, but the code *does not* fit isochrones; it fits synthetic clusters populated sampling a selected IMF. We generated synthetic clusters with the parameters recovered in the first and second runs of NGC4349 using equivalent masses of 1000 $M_\odot$, to demonstrate how they can differ in this range. In **Fig 2** we show the G magnitude histograms for both synthetic clusters.

As can be seen, even though the isochrones might look very similar in this range their *physical stellar distribution* is not. These differences, albeit small, have an impact on the search for an optimal solution even if both isochrones look *geometrically* the same. The referee is nonetheless correct that in this particular case -as shown above- the stars with G>18 mag were not causing an incorrect fit (as the second run solution is in fact a local minima).

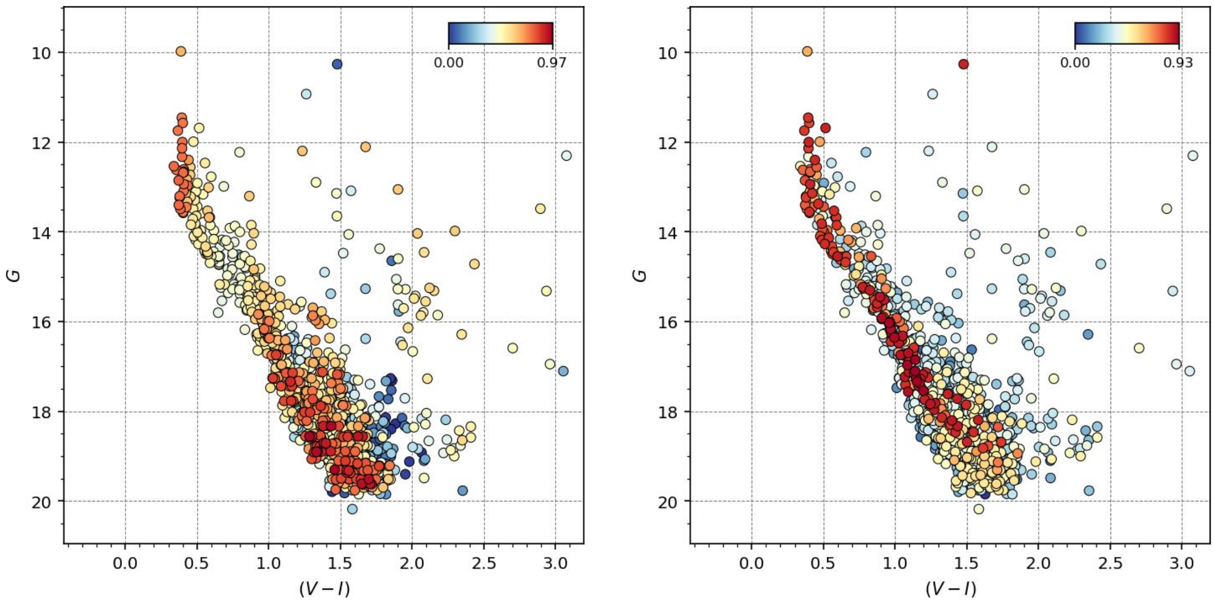


**Fig 3: membership probabilities obtained for NGC4349 using photometry (left) and parallax plus proper motions (right)**

6.

*> But briefly reading Perren et al. (2015,2017), it seems that the parameters estimation is done by fitting synthetic CMDs to the data, and the current improvement of including parallax and proper motion is done only for the step of assigning the membership probability. Moreover, on Sect. 4.2 you mention that the magnitudes are not used to estimate the memb. prob., only parallax and proper motion. Therefore I am tending to think that the memb. prob. do not necessarily generate a CMD with a sharp sequence of high-probable member stars (as a memb.prob. estimation based only on colour and magnitude would naturally do), and this may be the root of the answer to my question above. In other words, the apparent mismatch of the isochrones with the high-probable member stars would not be an issue with the isochrones (aka parameters), but with the membership probability.*

Photometry is widely considered an inferior feature compared to parallax and/or proper motions (the latter in particular) when it comes to disentangling the true members of a cluster from contaminating field stars. To demonstrate this we show in **Fig 3** (left) the result of using only photometry (magnitude and colors) in the membership probabilities estimation process. As can be seen, the cluster sequence is rather blurred with a much larger dispersion of the probability values when compared to the results obtained using parallax and proper motions (right).

# Minor points

## 1. MA-03

*> the coefficients of equation 2 have three to five significant figures, which seems too much. I wonder if you found that precision in the fit. Maybe adding the uncertainties and some quality factor of the fit would help clarifying.*

The standard deviations for each fitted coefficient the overall RMS of the fit have been added.

*> Please reduce the significant figures of the uncertainties to two figures and the parameter value accordingly also in Fig. 2.*

The significant figures have been reduced in Fig 2, Table 3, and the body of the article.

*> paragraph 6 of sect.3.1 discuss the differences between magnitudes without mentioning the r.m.s. of each mean. In all comparisons the offsets are compatible with zero within less than 1- sigma (incl. the figures sent only to the referee). Therefore it is pointless to discuss any offset. Maybe if only the brighter stars are considered on the mean, the r.m.s. would be smaller and the difference would or would not be consistent with zero.*

We’ve added the standard deviations to all the magnitude deltas.
While it is true that the offsets are small, our analysis in Sect 6 shows that the BV~0.015 mag offset found has a non-negligible effect on the final distance difference estimated comparing photometric versus Gaia parallaxes values. As it is a small added analysis that can easily be discarded if desired, we believe there’s no harm in keeping it in the article.

## 2. MA-07

*> Please mention the mathematical artifact that you explained also in the paper for clarity.*

Added to the caption of Fig 5.

*> Sect.4.2: it is ok to fix manually the radius in this case, if the King profile (or other) does not converge. But please explain in the paper how you reach the precise values of 2.23, 2.89, 1.49 arcmin etc... manually.*

The radii are estimated using the frames in pixels coordinates and *then* converted to arcmins. This explains the decimals used in these values. A footnote has been added in Sect 4.2.

*> I did not realise until now that you combined data from different instruments with different pixel scales and FOV for the clusters NGC4349 and Lynga15. Combining flux of images with different pixel scales is not a trivial task, if you did it please explain it briefly in Sect.3. Also, if the images with different FOV sizes were combined to generate Figs. 9 and K.2, please add this information in the figures at least, as this could explain the steep decrease in the field density for these two cases that stand out from the other clusters.*

We did not combine fluxes. The process employed to combine frames with different scales is much simpler: process both sets separately, identify a group of stars in common, derive a transformation equation (pixel to pixel) using these stars, and finally transform one frame to the coordinates system of the other. We believed this was a simple and direct process and thus not really worth mentioning.

The information about the combination of the frames was added in the sections of both clusters after the previous report:

- NGC4349: "*The observed frame's density map shows two regions with very distinct mean stellar densities of background. This is just an artifact generated by combining observations made with two different telescopes, as detailed in Sect. 3, and is the reason why the RDP shows such a strange shape, as seen in Fig. 9.*"
- Lynga15: "*In turn, ASteCA analysis of the spatial structure found an extended and irregular stellar density with no indication of a clear overdensity. The observed frame's density map shows two very distinct stellar densities, explained by the combination of observations made by two different telescopes detailed in Sect. 3 (same as NGC 4349).*"

## 3. MA-08

*> please add back more than 2 numbers in the colour bar. I know the scale is linear (and not logarithmic) because of the previous version of the paper, but the reader has no way to know that with only two numbers.*

We have added an explicit mention that the colorbar is drawn using a linear scale in the caption of Fig 5.

## 4. MI-01

*> still for aims (2): the 'results' say you found a variable level of disagreement (in other words, no systematic difference), and 'conclusions' say your results suggest that a systematic shift of 0.024 mas should be added to Gaia parallax.*

The “variable level of disagreement” statement means that not all the analyzed clusters show the *exact* same offset. We have added “average” to the conclusions to explicitly indicate that this is the result of averaging the (variable) offsets found.

## 5. MI-12

> *OK, but please add the unit information, i.e. "(degrees)" to both axis labels, or at least in the legend.*

Added to both plots.

## 6. MI-15

> *OK, but the reference should be to Fig.7 not A.4 (now that you chose a different cluster as first example).*

Fixed.

## 7. MI-17

> *OK, but please add reference to "King profile", is it King et al. 1962? 1966?.*

Added the required reference.

## 8. MI-20

> *OK, but now that the sentence is clarified, please make a final update for the sake of clarity:*
> *After -> During*
> *is finally compared -> is compared*

Fixed.

## 9. MI-29

*> OK, but please match the number of significant figures between the parameter and uncertainty. For example: 13.5±0.26 mag should be 13.50±0.26 mag, 5±0.6 kpc should be 5.0±0.6 kpc etc. Check the paper for that. The Gaia distance is correctly written 5.48±0.44 kpc.*

Fixed six instances of this issue.