Report on the article: “pyUPMASK: an improved unsupervised clustering algorithm”, reference (AA/2020/40252), second version.

General comments:

The paper’s presentation has been notably improved and I thank the authors for the work they have done. Nevertheless, I believe that a few points still need to be clarified before giving it the green light. I also discuss a few new issues that have arisen from the authors’ answers.

1.- Authors’ answer: *Both the articles for UPMASK (Krone-Martins & Moitinho 2014; Sect. 1. ‘Introduction’) and ASteCA (Perren et al. 2015; Sect. 2.8 ‘ Field star decontamination’) contain large sections commenting on the history of membership estimation in star clusters with lots of references. We have added to the Introduction a few of the articles mentioned by the referee, and directed the reader to those two references for a more detailed recount.*

R: I think that the historical development of the algorithms for separating between field and cluster stars is now better defined and its inclusion gives the necessary context to the potential reader who is not familiar with this methodology. However, I still believe there is a need for a more detailed discussion of the methods, developed up until now, for carrying out this classification including photometric data. Why am I being so insistent about this? Because of the special nature of this type of data (their pdfs make it difficult to describe them using a simple mathematical formulation) and due to the scarcity of algorithms that have used it to separate between cluster and field stars. UPMASK was designed to work on this space of variables (along with the positions – I will return to this in point 2); ASteCA was also designed for this purpose and for these variables; and now pyUPMASK (an extension of UPMASK) expands the space of variables but also includes the subspace of positions+photometry. It is crucial, that a reader who is searching for a tool to separate cluster and field stars be able to understand the differences between these methods and be able to choose the one that suits best depending on the data available and their scientific objectives, particularly if the data available are only positions and photometry, something that surveys like PanSTARRS are providing more and more frequently.

2.- Authors’ answer:

- *“ Since UPMASK only uses positions and photometry” this is not correct. Both methods were tested using the same sets of synthetic data, i.e. PHOT and PM. UPMASK is not limited to photometric data, even though that was the only data employed in the Krone-Martins & Moitinho (2014) article. All the performance analysis values shown (including of course Fig 6) are comparable between both methods, because both methods analyzed the exact same set of synthetic clusters (PHOT + PM). We have made this point more clear in Sect 3.1*

This fact notwithstanding we have added an appendix showing Fig 6 but segregated between results for the PHOT and PM datasets, for clarity. This is mentioned in Sect. 4

R: The first sentence of *item* 2 (*"Since UPMASK only uses positions and photometry" this is not correct)* is a very emphatic answer that made me doubt whether I had correctly understood the two articles (pyUPMASK and UPMASK). I reread the original UPMASK paper, published in A&A 561, 57 (2014), by Krone-Martin & Moitinho, entitled "UPMASK: unsupervised **photometric** membership assignment in stellar clusters", and I found that the first sentence of the abstract states: "*Aims: We **develop** a method for membership assignment in stellar clusters **using only photometry and positions**."*

It does not say that the method has been applied to a case study "***using only photometry and positions***". Rather, it says that the method was **developed** "*for membership assignment in stellar clusters **using only photometry and positions***". Although in scientific papers sometimes, sentences are included that can lead to misunderstanding and require very close and even repeated reading, this particular sentence seemed to me to be so descriptive (clear and concise) of the variables which the method was developed for, and not open to misinterpretation, that I accepted what it said: UPMASK is designed to work only with positions and photometry. It may be that what the authors had in mind was for UPMASK to be extendable to other types of data, but in the original paper abstract they did not state this, and the case study that they considered was not based on variables other than positions+photometry either. In order to prevent other potential readers from making such mistakes, the following points should be clarified in this article:

a) Indicate the published studies that have made use of UPMASK with other additional (or substitute) variables to position and photometry. These papers should be cited here as proof of UPMASK's applicability to other variables, since this example of use was not tested in the original 2014 article.
b) If such an implementation has not existed until this pyUPMASK work, this point needs to be clearly made. In other words, you should include a sentence that says something like: "In this study we apply UPMASK for the first time to data types other than position and photometry, replacing photometry with proper motions, which demonstrates the versatility of our code…", or something similar. I believe that, aside from making the potential of UPMASK clearer, this would increase the value of both studies.
c) It would also be a good idea to send a "corrigendum" to A&A, correcting and expanding the 2014 article abstract, including the important information that the method is also applicable to other types of variables that meet certain conditions, even though only the case of positions+photometry is presented there. This last issue is only a suggestion.

The separation of the comparison of results between the different sets of variables used (PHOT and PM) is very helpful for understanding the capabilities and potential of the two methods. I think the new tables (which had been included in an appendix) should take the place of Figure 6, which should be removed since it is

now redundant (it is enough to add up the corresponding cells of the new tables to obtain the above).

3.- Authors' Answer: *ASteCA was purposely left out of the article because the method employed by this code to assess membership probabilities is not directly comparable with either UPMASK or pyUPMASK. Unlike these two methods which are unsupervised , ASteCA requires that the cluster region and field region are a priori defined. This is done by estimating the center and radius of the cluster (and a surrounding field region) before the membership algorithm can be applied. This means that ASteCA uses a supervised method of membership estimation, as one of the classes (field stars) must be clearly identified and the other (cluster stars) approximately identified (as it is contaminated by field stars). Being non-comparable methods, we do not believe that adding ASteCA to the analysis is reasonable. We have explained this point clearly in Sect 2*

R: In the abstract of the article, "ASteCA: Automated Stellar Cluster Analysis", published in A&A 576, A6 (2015) by Perren, Vázquez, and Piatti, it says: "*The set of functions included in the code* ***make use of positional and photometric data*** *to* ***obtain precise and objective values for a given cluster's center coordinates, radius,*** *luminosity function and integrated color magnitude, as well as characterizing through a statistical estimator its probability of being a true physical cluster rather than a random overdensity of field stars.* ***ASteCA incorporates a Bayesian field star decontamination algorithm capable of assigning membership probabilities using photometric data alone.***"

The ASteCA abstract does not say that the centre and radius of the cluster should be known *a priori*. Rather, it states that by choosing an area of sky, the code is capable of *characterizing through a statistical estimator its probability of being a true physical cluster rather than a random overdensity of field stars* and ***obtain precise and objective values for a given cluster's center coordinates, radius,*** *luminosity function and integrated color magnitude,* all by using only the positions and photometry. That is to say, it obtains the cluster centre and radius from the position and photometry data; it does not need them beforehand to begin the analysis. ASteCA basically has the same scientific objectives as pyUPMASK, and the latter method can work with different datasets, including positions and photometry. My question is therefore: why can we not analyse the PHOT synthetic cluster data shown in Figures 3 and 4 of this article with ASteCA and compare the results with those obtained by pyUPMASK? Or even better, why not take the 320 synthetic clusters of the PHOT subsample and analyse it with ASteCA? Frankly, I believe that this would considerably help to understand what the array of methods we have available is and make it clear what pyUPMASK contributes compared to the two previous methods when we only wish to use (or only have access to) photometric and position data.

Specific issues

1.- I thank the authors for their answers to the majority of the specific questions raised. They have helped me to better understand the study. Above all, the issue raised in item (e), with the following response:

- "…one can infer that the second dataset (means for PM sample) contains PMs and photometry.", that is correct it does.
- "For the PM set we used only the proper motions, and no photometry" this means that although the PM set contains photometry in addition to proper motions, we did not use this photometric data in the analysis.

So, a (PM) dataset with positions, photometry and proper motions has been created to separate between cluster and field stars. I have reviewed Section 3.1 to see how it reads now, and I quote it here:
"We will refer to this sub-set as PHOT hereinafter. The second sub-set contains 280 clusters generated adding synthetic proper motions to all the stars in the frame; we will refer to this sub-set as PM hereinafter. The idea is then to see how both algorithms handle the case were only photometry is available (i.e., the PHOT dataset), and the increasingly common case (thanks to the Gaia mission) where proper motions with very reasonable quality are available (i.e., the PM dataset).

It seems to me that it would be a good idea to include the adverb **only** in the sentence, "*…, and the increasingly common case (thanks to the Gaia mission) where **only** proper motions with very reasonable quality are available (i.e., the PM dataset)"* in order to properly understand how the PM sample is built.

In addition, since you also have photometry for the PM sample, have you confirmed whether the colour-magnitude diagrams of the member stars, obtained with the proper motions, follow the loci expected for the cluster?

2.- Point f) of my first report was as follows:
"*f) Also on p.17, it says: "Proper motions are generally regarded as better cluster members discriminators than photometry. We were able to confirm this by checking that the results (with either UPMASK or pyUPMASK) degraded if photometry was added to the proper motions as input data for the PM set." This is a very interesting conclusion, which leads us to think that the comparative analysis shown in Fig. 6 does not provide the appropriate information, and that the results of the performances of the two methods, when UPMASK only works with PHOT and pyUPMASK with PHOT or PM, should not be mixed.*"

I have received an answer that I don't understand particularly well, which is as follows:

"● 'when UPMASK only works with PHOT' this is not correct, as mentioned in 'General revisions ' b) ."

Allow me to explain myself better. UPMASK was designed to work with positions and proper motions (see point 2 a,b in General Comments), but could be applied to other

datasets provided they met the condition that the cluster stars were more concentrated than the field stars in that space. This, in principle, occurs with any method based on the pdf empirical estimation, but it needs to be specified how to carry this out for each set of variables that have their own characteristics. In this study (unless the authors show me otherwise with the appropriate references, as I state in 2a), it is the first time that UPMASK is used with positions and proper motions. The results are therefore compared between pyUPMASK and UPMASK both for PHOT and for PM. The explanation, "*Proper motions are generally regarded as better cluster members discriminators than photometry. We were able to confirm this by checking that the results (with either UPMASK or pyUPMASK) degraded if photometry was added to the proper motions as input data for the PM set*", seems to indicate that other trials have been carried out with the variables positions+photometry+PMs, and that the results were worse than using only positions+PMs. I now understand better why the PM subsample also included photometry. I believe that this result is very interesting and should be given greater strength. For example, I wonder what the results obtained with the variables positions+photometry and positions+photometry+PMs are like. A question like this would not be difficult to answer, since it appears that UPMASK and pyUPMASK have been applied to three different sets of variables: PHOT, PM and PM+photometry. One sentence on the comparison between the results for PM and PM+photometry is included, but the conclusions of the comparison between the results for PHOT and PM+photometry should also be communicated.