Review of “Cloud detection from multi-angular polarimetric satellite measurements using a neural-network ensemble approach” by Z. Yuan et al.

This manuscript describes a novel neural-network (NN) approach for cloud detection from on multi-angular polarimetric (MAP) measurements, which is here applied to the POLDER-3/ PARASOL instrument. The main motivations put forward by the authors are mainly to use this product as a cloud mask for aerosol retrievals in order to avoid biases in case of multi-layer or coincident conditions within the same pixel (6°x6° for PARASOL). The NN approach is first presented and sensitivity analyses of some of its thresholds are discussed. Retrievals corresponding to one year of POLDER observations are shown and compared to MODIS. Comparisons to the more standard cloud masking approach “Goodness-of-fit” are also shown for two days. Finally, the impact of the choice of the mask (NN, MODIS or Goodness-of-fit) when doing aerosol retrievals is assessed by comparison to AERONET.

Advanced NN (or machine learning) approaches to detect and categorise clouds are increasingly used and excellent alternatives to replace or complement more traditional decision trees. Such an application for MAP measurement is in my knowledge new and is worth exploring. The method is timely considering the launch of PACE and 3MI in the very near-future. In this sense, this work is of scientific importance and fall within the scope of AMT. The overall NN approach seems sound and well constructed, although I must admit not being an expert.

The writing of the manuscript is fluent but some explanations (in particular of the method) lack clarity. Some figures are difficult to read, and overall lack a clear labelling of the panels; which makes it difficult sometimes to relate them to the text. The description of the figures is sometimes insufficient and overall there is a clear lack of quantitative analyses in the text. Some strong conclusions are taken without clear justification in my opinion. All these aspect unfortunately makes it difficult for the reader to properly understand the results of the new method and their added value to the other approaches illustrated here. Considering all these points, I advise for major revisions following the comments and suggestions below before considering this manuscript for publication.

**General comments to the authors**

1. A few comments regarding the method (described in section 3.1): Why are uniform distributions used to represent liquid and ice cloud effective radius, global dataset would clearly show a normal or log-normal distribution of these properties. What is the consequence of not considering the presence of both liquid and ice together on your method (and what is your reason for not considering them at all?) - these situations occur very often!
2. The conclusions of the comparisons to MODIS are too optimistic in my opinion. For instance l. 244-245, it is clear from the maps that the zonal fraction distribution does not agree very well. NN misses high cloud fractions in the tropics (along the ITCZ and in the warm pool region), see Fig 2. This seems to be related to convective clouds not being well identified by the NN approach. That can again be verified by looking at convective clouds over land during hemispheric summers not being identified by NN (e.g. above South America in the top row of Fig 3). Could you provide an explanation for this issue? And in any case it might be worth better discussing these limitations (that one and possible others) in your conclusion. You partly attribute the differences to the incorrect detection of aerosol as clouds by MODIS, which is true, but can’t be the only reason. Also, the argument of cloud moving during the acquisition seems strange, how long does the acquisition take?
3. Have you considered providing the ice and liquid cloud fraction separately? This would make your method even more attractive and help to understand the performance of your method for instance by comparison to MODIS. One strong benefit of MAP measurements is the phase detection. It feels to me that this would be a small effort to add to your NN method considering the way it is constructed for a very high gain. In any case it is important to you characterise better what cloud type contribute to your uncertainties in CF estimates.
4. The main conclusion should be more clear. It seems to me that the NN approach leads to similar result to the goodness-of-fit method but that the latter is very computationally expensive. In that sense, a complete transition to the NN approach could be encouraged for future missions, is that correct?

**Specific comments**

1. There is no section describing the satellite data and its availability, as required in the AMT author guidelines. Additionally it would be useful to indicate in section 2.3 what version of the MODIS cloud mask is used.
2. Overall, there are large multi-panel figures that would require a labelling of the panels to be better understood, please add these and adjust the text discussions accordingly. Note that this is a requirement in AMT guidelines to authors.
3. l. 15: “Climate change, which refers to long-term changes in temperature and weather patterns” - this statement is too limiting, please revise it.
4. l. 23: Ice crystals are not directly formed from condensation nuclei, this is a bit misleading.
5. l. 57: Surely there are other reasons for retrievals not to fit (other non-retrieved parameters from the forward models). Could you be a bit more specific on how this method attributes the misfit to the presence of a cloud?
6. l. 78: “PARASOL contain unique sensitivity to clouds” - please be more specific.
7. Section 2.3: What version of the MODIS Cloud Mask have you used? In general, the paper lacks a clear description of the dataset that are used (including full name, DOI and access)
8. l. 217: “The higher the effectiveness, the fewer aerosol retrievals are attempted on cloudy pixels”: please be more specific on what is meant here, is there more than one aerosol retrieval attempted on a cloudy pixel?
9. l. 255: What have you done to mitigate this effect? Be more specific.
10. Fig. 6: Very little can be seen on this figure in my opinion, I’d suggest removing it. The “NN seems more blue” (l. 284) is not quantitative enough for the standards here.
11. Fig. 8: This figure contains a lot of information and it is not clear what the point beyond the fact that there are little differences (at least notable in this figure) between POLDER using different masks and AERONET. Fig. 9 should be sufficient to make that point.
12. l. 320: “For both AOD and AE, applying a cloud mask (either MODIS or PARASOL-NN) can improve the percentage of retrievals within the requirements”: it does not seem so clear from the figure, please be more quantitative in your analyses to justify this conclusion. There are some differences, but are they significant?
13. l. 351: Could you add some information on what you expect improvements would be for the method as applied to 3MI and PACE? Increased precision?