Review of "Evaluation of VIIRS Neural Network Cloud Detection against Current Operational Cloud Masks"

This Manuscript describes a new methodology to estimate cloud mask using the observations of VIIRS instruments. This method use artificial intelligence tools such neural networks to perform the distinction between pixels contaminated with clouds and the pixels without clouds. Three years of CALIPSO/CALIOP data are used to train, validate and test the neural network. Then, the authors compare the performances of their methods with two other one. Hence, this study shows that neural network tools are adapted to retrieve cloud mask in passive remote sensing with good consistency with active remote sensing. The work is well presented and worthy of publication in Atmospheric Measurement Techniques after answering one major comment and some minor modifications.

Major Comment:

What is exactly, is given in the input of the main neural network? Is it radiances, or reflectances and brightness temperature? For each radiances, is it the differences with climatology (or simulations) and the observation, or raw radiances?

My concern is that the neural network description lack about the physic that is behind such the nature of the input. Also that important information are dispatch in all the study to explain fairly some results, but they are still necessary to be mentioned in the neural network description.

This is more a thought for the conclusion: How does neural network methods will react in the context of global warming and the fast modification of some surfaces? Does it mean that despite the benefits of the accuracy that provide neural network, they are countered by the fact they will need regular updates?

Minor modifications

It would help to provide a table of the VIIRS band.

Page 4, section 2.2: You mentioned in the discussion section (page 16, line 505) that you use ancillary data. But it is poorly described in this section 2.2 (linked to my major comment).

Page 6, lines 171-181: The second part of the section "3.1 Pseudo-Labelling Procedure" is hard to understand at some points. In this section, it is about the neural network that help to account for sun glint. What information is provided

by this neural network to detect sun glint? Is this information provided to the main neural network to not perform a cloud mask, or does it simulate input that are supposed to appear in sun glint condition for the main neural network? Where comes from the information of true sun glint conditions, to be reproduced? Why the 15th day of every month in 2018?

Page 7, line 202: what is the meaning of "binary cross-entropy"?

Page 10, lines 279- 284: Seeing the Figure 4, the difference between TPR of MVCM and the one for neural network is really small. It is most likely that their performances for low broken clouds are similar.

Page 10, line 314-322: In Figure 6, the cloud mask with neural network is less sensitive to variation of latitudes.

Page 11, line 323: "All the of the previous" A word is missing!

Page 12, lines352-354: "This is surprising … a land or water surface." This is really important information it should be mentioned in the description of the neural network input section. (major comment)

Page 14, line 428: "is subject **to** a large"

Page 14, line 447-page 15, line 459: I suggest you put this section and Figure 12, with the section "3.1Pseudo-Labelling Procedure", as it makes the understanding of the pseudo-labelling more clear. Also, because this section is quiet independent of all the analysis of the neural network performances.

Page 23, Figure 1: This paper would benefits of a better scheme that describe the neural network. Better description of the input vector with geo-localisation information. In relation with section 3.2 page 6 and 7, can you say more about the meaning of dropout X% between each layer?

Page 34, Figure 12: There are obvious difference in the behave of the cloud mask from neural network without pseudo-label and the one with pseudo-label. The second cloud mask is more "binary" (i.e. values equal to 0 or 1) than the first one. Can you comment this result? Which neural network of figure 12, have you compared during your paper?