

# Response to Referee 2 on revised manuscript ‘Machine learning of cloud types in satellite observations and climate models’

Peter Kuma, Frida A.-M. Bender, Alex Schuddeboom, Adrian J. McDonald, and Øyvind Seland

December 5, 2022

Dear anonymous referee,

Thank you very much for the third round of comments. Please find our response below. In the following text, the original comments are in **bold**, followed by our response.

Kind regards,

Dr. Peter Kuma on behalf of the authors

**I appreciate that the authors have considered my previous comments and am satisfied with their revisions. I recommend acceptance after considering my remaining minor comments:**

- **I doubt that many readers will be familiar with receiver operating characteristic plots. What does “specificity” mean? Please provide more information to help readers interpret the plot.**

In the methods subsection on validation (Sect. 3.3), we added: ‘For a validation of the ANN, we calculate the receiver operating characteristic (ROC) diagram (Sect. 4.1). An explanation of the diagram is given for example by Wilks (2019) in Chapter 9.4.6. The diagram shows sensitivity (the true positive rate) and specificity (the true negative rate) of the prediction for a set of choices of thresholds for a positive prediction, represented on the diagram by points on a curve. The area under curve (AUC) is calculated by integrating the area under a curve, and can be interpreted as a goodness of the prediction.’

In the caption of Fig. 5, we added: ‘Shown is also the ROC of a random predictor. Sensitivity is the true positive rate (probability of a positive prediction if positive in reality), also called the hit rate. Specificity is the true negative rate (probability of a negative prediction if negative in reality). ‘1 - specificity’ is also called the false alarm rate. Area under curve (AUC) of the ROCs is in the label.’

- **L399: This could be a good place to remind the reader that the ANN is trained on ground-based cloud observations and so disagreement with satellite-based cloud data being larger for higher clouds might be expected.**

We added: ‘This may be due to the fact that our method is based on ground-based cloud observations, which are often not capable of identifying high and mid-level clouds. It can be expected that discrimination of high and mid-level clouds by the ANN is not as good as that of low clouds.’

- **L442: “close to CERES”**

We corrected this mistake.