

Reply to anonymous Referee 2 comments to *Neural network cloud top pressure and height for MODIS*

Nina Håkansson et al.

1 Overall quality of the discussion paper ("general comments"):

1.1 Referee comment:

In the paper a novel retrieval of cloud top pressure and height using neural networks is presented. The presented retrieval technique is state of the art and an accurate technique. To account for different availabilities of channels on different satellites, a few modification of the neural network are investigated revealing the information content of the different channels. The new algorithms are compared to two reference algorithms, the CTTH algorithm of the NWCSAF PPS-v2014 and the MODIS collection 6 L2 height product. Additionally the algorithms is compared to CALIOP and CPR measurements. The quality of the algorithm is evaluated in terms of the mean absolute error (MAE). The improved quality of the results is impressive. From my point of view, I would request for at least another quality measure like standard deviation (similar to Tables 6 and 7). In overall, good work!

Reply:

We thank Referee 2 for this positive comment, and for the other valuable comments that will help us improve the paper further.

Regarding adding standard deviation, which was also requested by Referee 1, we chose the MAE as evaluation metric, over bias and standard deviation of differences (SDD), for many good reasons. However, some of the good reasons became clear to us first when we where faced with the request to include them in the article. Most important is that including bias and SDD of the error distribution intuitively gives the reader the mental picture of a Gaussian error distribution, centred at the bias. However we are dealing with skewed and even bimodal distributions as shown in Figure 2 and the mean is not at the centre of the distribution; i.e. the bias is not located at the peak of the error distribution.

The overall standard deviation is much affected by the largest errors. Some large errors are expected due to the differences between the passive and active sensors and the different FOV (field of view). Therefore we argue that the MAE is a better measure of variation of the error compared to SDD. The largest errors are of course also interesting, but when investigating these some care should be made to separate true errors from expected differences due to for example cloud edges in the FOV. We will consider including the Median absolute deviation (MAD) or the interquartile range (IQR) as these are more robust measures of variability less sensitive to outliers compared to SDD.

See also the more detailed reply to comment 2.6 in the reply to Referee 1 regarding inclusion of bias and SDD. As the bias and SDD are traditionally used when evaluating cloud top height retrieval algorithms we will motivated in the article why they are not included.

2 Individual scientific questions/issues ("specific comments")

2.1 Referee comment:

p1 line 23: CTH might also be used in data assimilation of atmospheric motion vectors.

Reply:

We will consider adding this to the text.

2.2 Referee comment:

Introduction: A short description of the traditional technics to retrieve cloud top pressure and height could be added to the introduction. (or cite an overview paper like Hamann et al. "Remote sensing of cloud top pressure/height from SEVIRI: analysis of ten current retrieval algorithms." Atmospheric Measurement Techniques 7.9 (2014): 2839-2867.)

Reply:

We will add the suggested reference.

2.3 Referee comment:

The introduction should motivate, why it is expected that using machine learning, in particular neural networks, could improve the expected results.

Reply:

Many CTH retrieval algorithms including MODIS-C6 and PPS-v2014 include some fitting of temperatures to NWP temperature profiles. This is most difficult in the case of inversions both as one temperature occurs at several pressure heights in the profile but also as the inversions are often not captured accurately enough in the NWP temperature profile. Many different techniques are used to deal with this, for example PPS-v2014 will place the cloud at the inversion height if the temperature is not more than 0.5 to 2K lower than the temperature at the inversion. MODIS-C6 has another approach using climatological lapse rates over sea for clouds likely to be low. These kinds of fitting techniques a statistical machine learning technique could probably do better.

2.4 Referee comment:

Merge chapter 2.1.1 into chapter 3.2. (and skip the sentence (p3 line 10) “The MODIS Collection 6 cloud product were used as an independent. . .”, you said that before).

Reply:

We will remove section 2.1.1 and move the information from it to section 2.1 and 3.2.

2.5 Referee comment:

Chapter 2.2 Add a short sentence, why you chose the CALIOP 1km product and not also 5km or 10km product which are more sensitive to optically thin clouds.

Reply:

The 1km CALIOP product was selected because it has the resolution closest to the MODIS resolution. It is expected that the thinnest cloud seen by CALIOP lidar is invisible to the passive imagers, so it should not be a problem that the thinnest clouds are missing in the 1km data. However we have also done some tests using AVHRR-GAC data and CALIOP 5km (Version 4) resolution for training (this is outside the scope of this article). The first tests show that results improve if the thinnest (0.05 or 0.1 in optical depth) clouds are excluded from the training. If these networks (trained on AVHRR-GAC) are applied on the validation data (MODIS) of this article the MAEs of the retrievals are between 76hPa to 79hPa. We will add a sentence about why 1km data was chosen.

2.6 Referee comment:

Chapter 2.4 add the version number of the ECMWF model and add product name and version number of the OSISAF data used in this study.

Reply:

We will do that.

2.7 Referee comment:

You might consider to add the PPS-v2014 and MODIS C6 algorithm to table 3 and 4.

Reply:

We will add them to Table 4, this will give the clear view of what channels are used for which method also for them. We suggest that they are not included in Table 3 as it describes *Network specific variables*. There is also an error in table 4, the NN-OPAQUE uses channel 12 μ m as described in Table 3 not channel 11 μ m. We will correct this as well.

2.8 Referee comment:

Please make the order of algorithms in table 3, 4 and 5 consistent.

Reply:

We did this in the first revision, and can not find any remaining inconsistencies.

2.9 Referee comment:

p5 line 5: how often is a pressure lower than 70 hPa retrieved?

Reply:

It varies with each network from 0 up to 0.05%. We will report this in the text.

2.10 Referee comment:

p5 line 10: Why did you choose this number of levels? Is it sufficient to use 6 levels to represent the boundary layer inversions or other small scale features?

Reply:

Five of the levels (surface, 950, 850, 700, 500) were already used in the PPS-v2014, so this was our starting point. We tested to use the troposphere pressure, but then the networks became very sensitive to the type of NWP-data used. Instead we added the 250hPa level to have one more high level. We did test increasing the number of levels near the ground by adding levels at 800, 900 and 1000hPa, but the improvement was not large enough to motivate the extra computational time. One common problem for cloud height retrieval algorithms is that inversions are not represented accurately enough in the NWP data. As mentioned previously, MODIS-C6 instead uses climatological lapse-rates over sea to avoid this problem, other algorithms use sharpening techniques at the inversion. So it is not clear that more levels which would better represent the inversions in the NWP data would improve the neural network results, but this could be further investigated.

2.11 Referee comment:

p5 line 21: do you skip non cloudy pixels in the 5x5 pixel standard deviations?

Reply:

No, all pixels are included.

2.12 Referee comment:

p5 line 20: the $B_{3.7}$ has a solar component. Did you correct for this during day/night?

Reply:

No correction was made, that channel is used just like the others.

2.13 Referee comment:

p6 chapter 3.3.2: You chose to use specific days for training and others for validation. Given that you only use a limited number of days, wouldn't it be more to randomly select independent pixels from all available dates for training, validation, and testing to represent a larger variety of weather situations?

Reply:

Unfortunately all pixels in the dataset we have are not independent. A typical cloud is much larger than one pixel; there could be hundreds of pixels with almost identical data in the datasets. If we would randomly select independent pixels from all days for each dataset we would in practice use the same data all the time. This would cause the neural network to overtrain as the *during training validation data* would be the same as the *training data*. And in the last validation step results would be overly positive as also the *validation data* would be in practice the same as the *training data*.

Random sampling from all data has been suggested to us previously. We will therefore add the discussion to the article why it is not possible.

2.14 Referee comment:

p 6 line 19: Did you test other configurations that 30/15 neurons in the first/second layer? If yes, how was the performance?

Reply:

We did one test with 20/15 this network (NN-AVHRR) was 1 hPa worse and one with 30/45/45 (NN-AVHRR) this was 2.5hPa better but also took 5 times as long time to retrieve pressure.

2.15 Referee comment:

Chapter 3.3.3: May batch size and momentum be changed during the training process?

Reply:

No they can't.

2.16 Referee comment:

p 7 line 27: consider to discuss the solar component of the 3.7 μm channel. To my opinion this NN could perform better when corrected for that (e.g. adding the solar time as input variable).

Reply:

We considered adding the sun zenith angle as variable; however we can not decide how the neural network would use it. In the data we do not have all sun zenith angles present globally. It could be that the neural network would use the sun zenith angle to decide that during this time of day clouds of a particular height are most common. It does not have to be bad though and can be tested in future studies. The performance of NN-AVHRR1 could probably be improved if the solar component of 3.7 is treated explicitly. We will consider to add the discussion.

2.17 Referee comment:

p8 line 5: Maybe express it positively: All NN can reproduce a clear bi-modal pdf very similar to CALIPSO, the pdf of PPS-v2014 deviates from this shape . . .

Reply:

We will change the formulation, thank you for the suggestion.

2.18 Referee comment:

p8 line 7: It is written “for the best performing network”. Did you train several networks for one channel configuration? If so, could you describe the number of trained networks in chapter 3.2.2, please?

Reply:

With the *the best performing networks* we meant that the NN-NWP, NN-OPAQUE, NN-BASIC and NN-BASIC-CIWV was excluded. We will clarify this in the text.

2.19 Referee comment:

p 8, line 11: according to my table 6, the NN-MetImage is better than the NN-MetImage- NoCO2.

Reply:

Yes it does! However the NN-MetImage does not perform well for higher satellite zenith angles. Only networks that perform well for all satellite zenith angles are discussed in this sentence. We will consider to reformulate to make it clearer.

2.20 Referee comment:

p 8, line 15: do you have an idea, why the MAE against CPR is larger than the MAE against CALIOP for NN-MetImage and NN-MetImage-NoCO₂?

Reply:

This we think it partly because the NN-MetImage and NN-MetImage-NoCO₂ have some skill in predicting very thin high clouds that are not detected by the CPR-Radar. We will add discussion about this.

2.21 Referee comment:

p 9, line 9: could you please describe a bit more in detail the differences seen in Figure 7?

Reply:

We will describe the differences in more detail. The blue squares for PPS-v2014 (c) are due to the temperature retrieval for 32x32 pixels in one go. We can see that a lot of high clouds are by NN-AVHRR placed higher (pixels that are blue in (c), are white in (a)). For NN-AVHRR in (a) we can see that the large area with low clouds in the lower left corner get a consistent cloud height (the same orange colour everywhere).

2.22 Referee comment:

Chapter 5 Discussion: Could you also comment on applying your NN technique on geostationary satellites? What would be the main differences/challenges?

Reply:

This technique should not be limited to polar orbiting satellites. As the instrument SEVIRI has the two most important channels at 11µm and 12µm it should be possible to apply the technique to SEVIRI data. More data (in terms of number of days) compared to MODIS may be needed to produce enough matches. As the SEVIRI resolution is coarser results might be degraded compared to MODIS. Matches of SEVIRI with CALIOP will occur at many different satellite zenith angles. This might make it possible to use the CO₂ channel on SEVIRI to improve results without losing performance skill at high satellite zenith angles. We will comment on using the NN-CTTH technique for geostationary satellites in the paper.

A compact listing of purely technical corrections ("technical corrections": typing errors, etc.)

Replies to a few of the technical corrections. For the ones not mentioned here we will follow the suggestions from Referee 2.

2.23 Referee comment:

- p5 line 29 and thereafter: don't write CO₂ with cursive letters

- consider to write NoCO2 (in MetImageNoCO2) not in cursive letters.

Reply:

We will keep the notation with subscript but without cursive letters.

2.24 Referee comment:

Figure 1 (p21): consider to have the figures in the same order as the algorithms are mentioned in table 3, 4, and 5.

Reply:

This is a reasonable request. However it is also nice to have the two AVHRR based algorithms next to each other so they can be compared. Also this order makes the two networks performing bad at high satellite zenith angles appear on the last row. And changing the order increases the risk to mix them up in later references. If someone refers to the bad satellite zenith angle behaviour in Figure 2 (h) in the discussion paper, and an interested reader by accident finds the final revised paper (assuming there will be one) and finds the result for NN-AVHRR1 in that sub figure that is not good. As there are also good reasons to keep the current order of sub figures we argue that their order should not be changed.

2.25 Referee comment:

- p 8, line 34 and following: avoid the abbreviation NN-CTTH, e.g. change: that the NN-CTTH all have -> that all NN retrievals have . . .
- avoid NN-CTTH abbreviation (which one do you mean? all NN retrievals or NN-MetImage or another one?)
- p 10, line 12: avoid NN-CTTH -> specify which retrieval you referring to

Reply:

We will choose to keep the NN-CTTH as the name for the neural network method in the paper. We will better present it as the name and avoid using it where it might be confusing.

2.26 Referee comment

p 10 line 16 “neither of which could be applied for the AVHRR1 instrument” can be skipped

Reply:

We will consider to reformulate. For the processing of longer climate data records it can be of interest to know if an algorithm can also be applied for AVHRR1.