Atmos. Meas. Tech. Discuss., 6, C4787–C4794, 2014
www.atmos-meas-tech-discuss.net/6/C4787/2014/
© Author(s) 2014. This work is distributed under
the Creative Commons Attribute 3.0 License.

Atmospheric Measurement Techniques Discussions

# *Interactive comment on* "Satellite retrieval of aerosol microphysical and optical parameters using neural networks: a new methodology applied to the Sahara desert dust peak" *by* M. Taylor et al.

**M. Taylor et al.**

patternizer@gmail.com

Received and published: 25 April 2014

We would like to thank the reviewer for their detailed review of the manuscript and for supporting the publication of the work:

"General comments:

This paper discusses the new methodology to retrieve aerosol microphysical and optical parameters from satellite remote sensing based on the artificial neural networks.



Global retrieval of aerosol microphysical properties is well-recognized and challenging task. This work suggests an alternative to well-established methods of aerosol properties retrieval by utilizing the methodology of neural networks. NNs are widely used in many other fields but still remain exotic in aerosol remote sensing. The paper is well written, has a good structure and clear. It has good and detailed description of the methodology, data selection and the discussion of results. Additional attention was paid to such NN-specific topics like training dataset preparation, selection of NN architecture and training rules. At the same time, I spotted few inaccurate and unclear statements in the paper that should be corrected before the publication.

Therefore, in my opinion, the paper can be accepted for publication in Atmospheric measurements Techniques "after minor revision". I have outlined below few comments for the authors consideration".

In answer to the reviewer's specific and technical comments:

"Specific comments:

Page 10959, line 24. "sensors by performing multivariate regression" that's not exactly correct as AERONET retrieval performs a numerical inversion of the observations."

We thank the reviewer for this clarification and we agree. In the revised manuscript we have changed the sentence on Page 10959, lines 24-27 to:

"AERONET's latest Level 2.0 Version 2 inversion algorithm retrieves all of the aforementioned aerosol microphysical and optical parameters from ground-based sensors by performing a numerical inversion of the observations – which must be performed for each case."

"Page 10966, line 25. "normalize all input and output variables". I think authors should provide a more detail here. First of all NN is a rather new technique in the atmospheric remote sensing and not everyone is familiar with the fact that the values of input and output data should be in the certain range, if I'm not mistaken from 0 to 1. Secondly it is

not clear how the normalization is performed, I assume it is applied for each value of the vector. I think this should be outlined because input dataset contains size distribution and it could be mistaken if normalization should be applied to each value or to the whole distribution. Maybe term "scaling" would be more appropriate for this operation. Thirdly, as I understand, to perform such normalization we need to know the range of possible values of each variable (min and max) since they all have different scales. It is not clear which ranges were used, if they were defined for aerosol in general, or for dust aerosol type. You can add a reference here."

We thank the reviewer for pointing out the importance of outlining in detail the normalization process used. On Page 10967, line 4 we have added the following paragraph:

"Normalization of the input and output variables was achieved as follows. For each input and output variable data vector X, we calculated the mean $\mu$X and standard deviation $\sigma$X. The vector means and standard deviations were then used to map (or shift and scale) the input and output data vectors onto their z-score values: zX = (X - $\mu$X)/ $\sigma$X (i.e. standard normal values having a mean = 0 and a standard deviation = 1)."

In relation to your third comment, we agree that more clarification is needed. In the revised manuscript on Page 10967, line and following on from the above paragraph, we have added the sentence:

"In this study, we consider the min-max values to be those available in our training dataset, and that they arecharacteristic of dust in the Northern Africa region."

"Page 10966, line 26. "apply principal components analysis", I assume this is a well known and widely used operation, however I think authors could provide a bit more information on this operation, at least give a reference to some papers describing PCA itself."

Again, we are grateful to the reviewer for pointing out the importance of outlining in more detail the PCA process and the rationale behind applying it to the data used in



this work. On Page 10967, line 4 (following directly on from the inserted paragraph above), we have added the following paragraph:

"The application of PCA is done in our study to reduce the redundancy in the input and output variables. PCA is an effective procedure for removing this redundancy: it orthogonalizes the components of the data vectors (so that they are uncorrelated with each other), and it orders the resulting orthogonal components (principal components or "PCs") so that those with the largest variation come first - allowing us to eliminate the components that contribute the least to the variation in the data set. The application of PCA requires normalization of the variables prior to application of the method due to the fact that different variables have very different value ranges and would bias the measurement of the variance (Abdi and Williams, 2010). PCA was applied separately to the input and output variables and the extracted PCs were ordered. Best results were obtained by retaining the top ranked PCs that accounted for 98% of the total variation in the input and output data."

âĂć Abdi, H, & Williams, L.J.: 2010: Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2: 433–459. doi:10.1002/wics.101.

The relevant reference has also been added in the revised manuscript.

"Page 10969, line 8. "the outputs produced 7 PCs (see Sect 2.2 for details)" this comment directly connected with the previous one, it is not very clear according to the section 2.2 how 116 values were reduced to 7 and vice versa."

We thank the reviewer for their comment. We would like to point out that the PCA actually reduces the number of output variables from 38 to 7 without changing the number of data records (116 in the case of Dakar). The result of PCA with regard to this dramatic reduction in dimensionality (38 $\rightarrow$ 7) is due to the fact that the 29 "missing" components contribute a total of only 2% to the variation of the 38 outputs. We hope that the extra explanatory paragraph is answer to their previous comment will help to clarify this. With the reviewers' permission, we believe that it is not necessary to make

a change in the text here.

“Page 10975, line 8. “There appears to be a play-off . . . ” In terms of NN learning it seems that the learning dataset wasn’t big enough for NN to learn all the possible combinations of the output.”

We agree with the reviewer and with Dr Tsekeri’s related comment above. In a future publication we will present the results of applying this method on the global scale via the use of cluster analysis to extract large datasets of homogeneous AERONET inversion data containing tens of thousands of data points. On Page 10975, line 10 in the revised manuscript we have added the phrase:

“A possible explanation for this is that the dataset was not big enough for the NN to learn all possible output variable combinations. While this is beyond the scope of the present paper whose aim is to present the methodology and validate it on a small but well-defined data sample whose aerosol properties are well understood (i.e. desert dust in the Sahara), in a future study we will apply the methodology developed here on a larger spatial scale and to other aerosol type regions in order to investigate this effect more fully.”

“Page 10975, line 12. “unsupervised (testing) mode” I think the term “unsupervised” is not very suitable, usually in NN terminology it is used for one of the paradigms of NN learning, alternate to “supervised learning” used in this paper. This could lead to misunderstanding of the NN test, as during this process no weights are adapted.”

We thank the reviewer for this observation and we agree. We have removed the terminology “supervised” and “unsupervised” from the manuscript, retaining instead the terminology “training” and “testing” throughout so as to simplify and clarify the presentation. Relevant changes have been made on Page 10970, lines 10-14 and on Page 10975 line 12.

“Page 10976–10977, lines 23–5. The differences between MODIS OMI and AERONET



are discussed. Conclusion made that differences are explained by the different angular variations of aerosol in ground-based and space measurements. In my opinion, starting from the phrase “ . . . whereas MODIS’s spectro-radiometers measure the intensity of solar radiation reflected vertically by the earth system”, it is more logical to assume that it is much harder to differentiate the contributions of the surface reflection and aerosol extinction to the total measured radiance in the case of measurements from orbit, which is also worsen by the fact that deserts where most of the dust measurements are performed could have comparatively bright surfaces. So the mentioned systematical error could be explained by the influence of the surface reflectance.”

We are grateful to the reviewer for kindly sharing their expert knowledge. On Page 10977, line 3 we have inserted the sentence:

“Furthermore, in the case of measurements from orbit, the separation of the effect of surface reflectance and the effect of aerosol extinction on the total measured radiance is a much more difficult task, especially over deserts which can have bright surface pixels.”

“Technical comments:

“Page 10957, line 10. Could also add reference to IPCC 2013.”

We thank the reviewer and have also added the 2013 IPCC AR5 to the list of references and a citation in the text on Page 10957, line 10.

“Page 10975, line 14. “The testing performance of the trained NN was tested”, I assume word “testing” is a duplicate and could be deleted.”

Thank you. We have removed the word “testing” before performance so that the start of this sentence now reads:

“The performance of the trained NNs was tested by . . .”

“Page 10977, line 13. “and also the location of the coarse mode peak (radial bin 15)”

I think the term “location” is more suitable for the radius, which is defined, maybe it is better to use term “value”, “height” or “magnitude”.”

Thank you for this comment. We used the “location” to try to emphasize that the peak occurs at radial bin 15 – i.e. it is positioned in this particle size interval. We are worried that magnitude-related words might confuse the reader here. With the reviewer’s permission we propose leaving this unchanged.

“Page 10983, line 13. “aerosol microphysical and optics parameters”, replace “optics” with “optical”.”

We thank the reviewer for spotting this and apologize that we did not pick this up at submission stage. This typo has been corrected in the revised manuscript.

“Page 10984, lines 9–13. “In assessing the performance of the NN model”. The whole sentence is too big and hard to understand, try to split it in several simpler sentences.”

Thank you for this observation. We agree that it is not clear enough. We have changed the sentence to:

“In assessing the performance of the NN model, the uncertainty accuracy requirements provided by Mishchenko et al. (2007) were shown to be very helpful. In Table 5, it is seen that they provide limits of validity for the comparison of our test results against AERONET data.”

“Page 10992. Table 2. The line delimiting “inputs” and “outputs” as well as “optics” and “microphysics” in the first and second columns should be one row upper.”

We are very grateful to the reviewer for spotting this formatting error which was not picked up by us during the typesetting stage. This has been corrected in the revised manuscript.

“Page 10996. Table 5. I don’t think “Mishchenko parameter” is the best headline for this column, it gives the impression that these parameters were invented by Mishchenko or



were named to his honor, maybe “parameter” or “parameter according to Mishchenko et al.” would be better.”

We agree. In the revised manuscript we have opted for the simpler “Parameter” as the heading of column 1 in Table 5.

“Page 11004. Figure 4d. I think the x axis label should be “epoch” not “15 epochs” ”

Thank you, you are correct. The x-axis should read “epoch”. This figure has been reproduced with this edit in the revised manuscript.

“Pages 11004–11010. Figures 4–10. General remark. I don’t know what will be the sizes of these figures in the published article, but consider making axis labels and values as well as legends bigger for better readability.”

We thank the reviewer also for this important point pertaining to the clarity of the figures. In this AMTD version, the figure quality appears a little worse than the 300 dpi figures we uploaded. We are happy to increase the font size of the axis labels and values and will coordinate with production department to try to ensure the best way of achieving this. The reviewer’s comments have substantially helped improve the flow, clarity and presentation of the manuscript and we are grateful to them for their attention to detail.

---

Interactive comment on Atmos. Meas. Tech. Discuss., 6, 10955, 2013.