**Atmospheric
Measurement
Techniques** Open Access
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.

A. Di Noia (Referee)

a.di.noia@sron.nl

Received and published: 13 June 2014

Dear Authors,

I would like to thank you for the detailed answers to my previous comments, and for taking the effort of producing additional statistics and graphs to support your statements. Overall, I think this study is interesting and your proposed modifications have the potential to improve the manuscript, but based on your reply I think that there are still some points that would benefit from some further clarification. As I will further detail in

this report, I think that the paper presents two fundamental critical areas:

1. Since all the data used to quantify the performances of the algorithm were – in a way or in another, as we will see later – involved in the specification of the NN algorithm, we are left without a convincing demonstration that the algorithm really works when applied to a "random" scenario over the Saharan region. My opinion is that such a demonstration should be included in this paper, because this is part of the work that has to be done when a new methodology to perform a given task is proposed.

2. Regarding on the discussion of the performances of the NN with respect to each aerosol parameter, I see the risk that some of the statistics shown may be overinterpreted.

Furthermore, I thing that the presentation of the results can be improved in at least two respects:

1. In the introduction it should be made clear why MODIS has been chosen for the development of this method, despite the fact that other satellite instruments exist that are known to be better suited than MODIS for the retrieval of the aerosol microphysical properties.

2. Some of the statistics shown in the original version of the paper might perhaps be omitted, because they do not add much to the presentation of the method.

Here is my complete report.

Best regards,

Antonio Di Noia

MAJOR ISSUES

1. Model selection. In your reply to my previous comment you mention that you chose

train+validation MSE as the criterion to optimize the NN architecture because it led to a better performance on the Dakar test site with respect to using the validation MSE. If I have correctly understood your reply it seems to me that you performed the model selection exercise twice. Initially, you used the common validation MSE as a metric, which led to a certain solution. Then, you arbitrarily performed the exercise again by using training + validation MSE, and this led you to another solution. Then, you decided that this latter solution was the better one, because it led to a lower MSE on Dakar. But isn't this equivalent to saying that you chose the best NN architecture by minimizing the MSE on Dakar?

If this is the case, I would say that this approach, even with its limitations (Dakar is only a single site, and obtaining the best performance on this site does not necessarily mean that the same would be the case anywhere), is more fair than including the training MSE in the metric for the model selection. However, this also creates another problem. If also the Dakar data have been involved in the NN selection, as appears from your reply, then I do not see any validation of the algorithm on data that have been neither used to train the net nor to optimize its architecture. Therefore, I would suggest you to perform at least one of the following additional tests, that I would not postpone to a future paper:

(a) Validate your NN on another Saharan site that has not been used for training or model selection, if you still have some data available

(b) Use a CTM to select a number of different situations (at least 2) that have occurred over the Saharan region in the past, and invert the corresponding MODIS imagery using your algorithm, comparing (even only qualitatively) the spatial distributions of some of the retrieved aerosol parameters to those predicted by the CTM. Are the resulting spatial patterns reasonable?

In my opinion, only this type of test can lead us to substantial conclusions about the general feasibility of the proposed approach.

<div align="center">C4879</div>

2. As I suggested in my previous comment, you tested the mean absolute values of the NN-AERONET differences against the same statistic for the difference AERONET-training set mean. The table shows that the mean absolute (NN-AERONET) difference is lower than the mean absloute (AERONET-mean) difference for all the aerosol parameters. Ok, but how significant are these differences? It must be kept in mind that the AERONET estimates of the aerosol microphysical parameters are not error-free. For instance, in the case of desert dust aerosols, the standard error on the real part of the complex refractive index is quantified by Dubovik et al. (2000) in 0.04 for AOD(440 nm) larger than 0.5 (and presumably larger for smaller AODs), whereas for the imaginary part a relative error of 50% is reported. If this is the case, then I would doubt that, for example, the numbers you report for CRI-R(440) – 0.041283 vs. 0.040864 – are indicative of a real difference between estimating CRI-R(440) using the NN and estimating it using its average value on the training set. A similar line of reasoning applies to many of the retrieved aerosol parameters, the only noticeable exception being the coarse mode volume, for which it is clear that the NN estimate gives us more knowledge on this parameter than its simple average on the training set does. In other words, it seems to me that many of the aerosol parameters are not really retrieved by the NN, and this fact also relates to what I wrote in my previous comment about which aerosol parameters should be included in the NN output vector: my opinion is that, if you include all these parameters in the output vector and deliver them as a product without any sort of flagging, a data user might erroneously think that their values reflect the evidence of a real situation (e.g. spatial distribution of refractive index, or SSA, etc.) going on over the Saharan region, while perhaps they are not much different from random variation around a "climatological" mean. Maybe in this paper you should leave the output vector as it is, but I would at least suggest you to present this work more clearly as an exploratory investigation aimed at verifying which aerosol parameters can be inferred from MODIS AODs over the Saharan region and which aerosol parameters cannot.

<div align="center">C4880</div>

PRESENTATION ISSUES

1. I think that in general it is fair to show how using additional information in the NN input vector improves the retrievals. However, I would suggest to restructure the paper so as to emphasize in advance that this "empirical sensitivity analysis" is one of actual goals of this work. But on top of that, is the full analysis of the CASE 1 NN (with all the statistics shown in Tables 2 and 3) really informative for the reader? Is it not trivial that trying to estimate 7 uncorrelated quantities from only one input variable cannot lead to useful results? My personal opinion is that discussing the results of the CASE 1 NN in detail distracts the attention of the reader, because poor performances are, in a sense, to be expected in advance for such a NN. Would it not be enough to say that a PCA applied to the AERONET AODs only gave rise to a significant principal component, and that this is clearly inadequate to perform the task you are aiming at? Why go ahead, train a dedicated NN for such a case and discuss its results? I must say that also the best NN, that should be the CASE 4, suffers from this problem, but at least in that case you have 3 uncorrelated components in the input, so in that case the discussion of what type of information about the aerosol parameters you can retrieve becomes more interesting. I think that also for that case, the scope of your experiment should made clearer in the paper, by explaining that there is no hope to retrieve all the 7 principal components of the aerosol parameters from a neural model that is driven by only 3 uncorrelated (that does not necessarily mean independent) inputs. With respect to that, it might also be interesting to see what would happen if you apply a PCA to the set of the aerosol parameters retrieved by the NN, and compare the number of significant principal components (perhaps using the same criterion of 98% variance) to that of the AERONET values (7). I have never tried this myself, but I suspect that this number should be less than or equal to 3, and should give you a rough estimate of the actual "degrees of freedom" in your retrievals (i.e. how many parameters, or combinations of parameters, you really retrieve).

2. I would strongly recommend to remove Section 4.4 about the statistics of the com-

pliance between your retrievals and the accuracy requirements by Mishchenko et al. (2007). As I said in my previous comment, I think that the test set you are using is too case-specific to claim that, and that "small difference from AERONET" does not mean "certainty", especially if these differences are mostly evaluated in "average" situations.

3. I would suggest to include in the revised manuscript the analysis of the mean absolute error you presented in your reply to my previous comment. I think that this is an important part of the verification process. However, as I said above, in interpreting the differences between (NN-AERONET) and (AERONET-mean) I would recommend you to acknlowledge that also AERONET aerosol parameters often have non-negligible uncertainties (based on Dubovik et al., 2000), so that the significance in those differences should not be overinterpreted. For the same reason, I would suggest you to avoid to use many decimal digits to represent the mean absolute errors, because I am quite sure that the least significant digits do not have a real meaning.

_____

MINOR COMMENTS

-Abstract, P10597, L1. I do not think it is correct to say that the aerosol parameters were "previously inaccessible" from space, because methods for the retrieval of the aerosol refractive index and the aerosol size distribution have been previously reported for POLDER.

-Introduction. P10597, L10. Perhaps you might replace the reference to IPCC (2007) with the more recent one of the 5AR (IPCC, 2013).

-P 10597, L13. I would move the reference to Dubovik and King (2000) from after "robust inversion algorithms" to after "retrieval of aerosol parameters".

-In the introduction it should be mentioned that algorithms for the retrieval of the aerosol microphysical properties from space have been developed for the POLDER instrument on the platforms ADEOS-1 (Deuzé et al., 2000, 2001) and PARASOL (Dubovik et al.,

2011; Hasekamp et al., 2011; Waquet et al., 2014).

-P10959, L24-26. I do not think that AERONET aerosol microphysical properties retrievals are based on multivariate regression. As far as I know, such retrievals are performed by iterative fitting of a radiative transfer model using Tikhonov regularization, as described by Dubovik and King (2000).

-P10959, L25 till P10960, L1. What does it mean that NN retrievals can be performed "without having to recalculate each day"?

-P10960, L3. I would replace "inverse function" with something like "a nonlinear regression function yielding an estimate for the atmospheric state given the measurement vector". In fact, the forward function is usually not invertible, therefore the concept of "inverse function" does not really apply here.

-P10960, L6-7. The reason why the calculation of the inverse function (or of the regression function) takes time is not the need for running a grid of NNs, because this step is not strictly necessary (you decided to do it, but not everyone does), but simply because the training times can be very long (also depending on how many training data are used).

-P10973 and so on. While I find it legitimate to say that a correlation coefficient higher than 0.95 indicates a "very high" precision, I find it a little bit arbitrary to say that correlation coefficients smaller than 0.5 indicate a "moderate" agreement. Why is this "moderate" and not "poor"?

-P10983, L20-22. I would remove the adjective "new". Determining the NN architecture via exhaustive search over the space of NN architectures is a standard technique, and there is a large number of works where this approach is discussed or at least mentioned. Here are just a few references:

Curry, B., and P. H. Morgan (2006), "Model selection in Neural Networks: Some difficulties", Eur. J. Oper. Res., 170, 567-577, doi: 10.1016/j.ejor.2004.05.026

Gorr, W. L., D.Nagin, D., J. Szczypula (1994), "Comparative study of artificial neural network and statistical models for predicting student grade point averages", Int. J. Forecasting, 10, 17-34, doi: 10.1016/0169-2070(94)90046-9

Lawrence, S., C. L. Giles, and A. C. Tsoi (1996), "What size neural network gives optimal generalization? Convergence properties of backpropagation, Technical Report UMIACS- TR-96-22 and CS-TR-3617, Institute for Advanced Computer Studies, University of Maryland, College Park, MD, United States

Stathakis, D. (2009), "How many hidden layers and nodes?", Int. J. Remote Sens., 30(8), 2133-2147, doi: 10.1080/01431160802549278

-P10999, Table 7. As I said previously, I find the "mapping" from a value of the correlation coefficient to an adjective (good, moderate, poor, etc.) a bit arbitrary (e.g., the statement that anything above 0.5 is "very good" might be questionable). Wouldn't it be better to just fill the table with the correlation coefficients? Keeping the color scale might also be a good idea to make the results more immediate, but I would use a different scaling (perhaps a scaling that is linear with respect to R^2 might make more sense, given the statistical meaning of R^2, e.g. $R^2 > 0.8$= dark green, 0.6-0.8=light green, 0.4-0.6=yellow, 0.2-0.4=orange, <0.2=red).

———————————————————————————————

REFERENCES

Deuzé, J. L., et al. (2000), "Estimate of the aerosol properties over the ocean with POLDER", J. Geophys. Res., 105(D12), 15329–15346, doi: 10.1029/2000JD900148

Deuzé, J. L., et al. (2001), "Remote sensing of aerosols over land surfaces from POLDER-ADEOS-1 polarized measurements", J. Geophys. Res., 106(D5), 4913–4926, doi: 10.1029/2000JD900364

Dubovik, O., and King, M. D. (2000), "A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements", J. Geophys,

Res., 105(D16), 20673-20696, doi: 10.1029/2000JD900282

Dubovik, O., et al. (2000), "Accuracy assessments of aerosol optical properties retrieved from Aerosol Robotic Network (AERONET) sun and sky radiance measurements", J. Geophys. Res., 105, 9791-9806, doi: 10.1029/2000JD900040

Dubovik, O., et al. (2011), "Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations", Atmos. Meas. Tech. 4, 975-1018, doi: 10.5194/amt-4-975-2011

Hasekamp, O. P., et al. (2011), "Aerosol properties over the ocean from PARASOL multiangle photopolarimetric measurements", J. Geophys. Res., 116, D14204, doi: 10.1029/2010JD015469

IPCC (2013), "Climate change 2013, the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change", Cambridge University Press, New York, NY, United States, 2013

Waquet, F., et al. (2014), "Retrieval of the Eyjafjallajökull volcanic aerosol optical and microphysical properties from POLDER/PARASOL measurements", Atmos. Chem. Phys., 14, 1755-1768, doi: 10.5194/acp-14-1755-2014

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

C4885