Atmos. Meas. Tech. Discuss., 7, C4076–C4081, 2014
www.atmos-meas-tech-discuss.net/7/C4076/2014/

Atmospheric Measurement Techniques
Discussions

Interactive comment on "Use of neural networks in ground-based aerosol retrievals from multi-angle spectropolarimetric observations" *by* A. Di Noia et al.

A. Di Noia et al.

a.di.noia@sron.nl

Received and published: 4 December 2014

We thank Reviewer 1 for his/her comments and for his/her appreciation of the paper. Here is our point-by-point reply. The Reviewer's comments are highlighted in bold, our replies are in plain text.

Section 2.3, line 9 (grammar) Change to "The forward model provides a simulation of ...

We have changed the text as suggested by the Reviewer.

Section 4, page 9057, lines 6-8: Were the tested noisy data real observations or

simulated data with noise added?

The test data are simulations with noise added. To make this more clear, in the revised text we have added the adjective "simulated" after "noisy" at page 9057, L6.

Section 4, page 9057, line 20: Were separate y vectors used for reflectance and DLP? If so, why? Isn't the retrieval algorithm really using a concatenated version of the two? Does this make the determination of the number of PC for reflectance and DLP simpler in some way?

Actually, we could have even applied the PCA to the entire observation vector (reflectance + DLP). This topic is slightly controversial, and conflicting opinions can be encountered in literature on this topic. We decided to apply the PCA separately in order not to mix two physically "inhomogeneous" quantities, because we feel that this approach is more "traceable" from a physical standpoint, even though we recognize that this does not necessarily lead to better results. The rationale may be described as follows. The physical variables of the retrieval problem are the vector of reflectances and the vector of DLPs (together with the SZA and the surface pressure). Since the reflectance vector and the DLP vector are representative of inherently different physical quantities, we consider them as living in two separate vector spaces, and therefore we look for a minimal representation of each quantity in its own vector space, with the PCA providing the vector basis for such representation in both cases.

Section 5, page 9060, line 15: Change the third to last word from "worth" to "worthwhile".

Done.

Section 5.1, page 9061, line 27: The non retrieved aerosol parameters have been "randomly perturbed". In what way, and by how much? This is a little vague.

In the revised version of the manuscript we have added the details of the random perturbation. In particular, we have extended the last sentence at page 9061 by adding the following:

"... , by replacing the assumed values with values drawn from a uniform distribution (between 0.1 and 0.3 for the effective variance of both modes, between 0.03 and 0.06 for the surface albedo at 440 and 675 nm, between 0 and 1 for the fraction of spherical particles of the coarse mode)."

Section 5.1, page 9062, 9063, Tables 4 and 5. I'm a bit overwhelmed by the all the ways in which error, bias, correlation, etc. are represented. Which parameter is the best metric for success, and are presenting all these values a bit of confusing overkill? For example, I originally compared RMSE in table 5, and was confused because this does not support the conclusions in item 1 on page 9062. I then realized that these conclusions were made with the correlation coefficient. Which is a better metric, and why?

We are aware that Table 4 and 5 condense a lot of information, and therefore may be confusing. However, our opinion is that each of the four statistical parameters that are listed in those tables tells something different about the data, and for this reason it is not possible to designate a best metric for success. The RMSE is perhaps the most widely used indication of the overall agreement between two datasets, and can be seen as a sort of "L2 distance" between them. It encompasses the bias and the error standard deviation. In cases of small biases (as in our case), the RMSE is dominated by the error standard deviation, that gives the most immediate feeling of whether the retrieved data reproduce the reference data in a satisfactory way. The shortcoming of the RMSE is that it may be sensitive to outliers because of the presence of square powers, and for this reason we also give the mean absolute error (MAE), that conveys a similar type of information (it can be seen as a "L1 distance" between the datasets) but is much more robust to outliers. In principle we may drop the RMSE, but we would prefer to leave it, because it is still the most intuitive metric, and a reader may want to check it first and then use the MAE just as a control metric. The bias indicates if there is a systematic shift between the datasets, and is important on its own. In our case, it is perhaps the

least interesting parameter, because it seems to be small everywhere. On one hand, it could maybe make sense not to report it and just mention in the text the reason why it is not reported, but on the other hand including it in the tables gives a more convincing proof that this is the case. The correlation coefficient gives an immediate idea of how well the time variations of the reference data are followed by the retrieved data, and for this reason we think that it is also worthwhile to report it. Table 4 and especially Table 5 were structured in this way in order to give the reader a synoptic view of the performance statistics, in such a way that the reader does not have to wander from one page to another to get the global picture. While we recognize that this is not necessarily the best option for everyone, it is still the best we could conceive.

As for the conclusions in item 1 on page 9062, item 1 says that NN+PT outperforms LUT+PT in all the aerosol parameters, with differences that are particularly relevant in the fine mode effective radius and complex refractive index and in the coarse mode imaginary refractive index. It seems that this can be also said by looking at the decrease in the RMSE and in the MAE, right? The rise in the correlation coefficients may be more immediate, but it seems to us that also the other metrics are consistent. In the revised version of the paper, we have tried to emphasize this point more explicitly.

Figure 4: perhaps it would be useful to indicate on this figure the chi-squared error threshold for a successfully converged solution? Also, considering the range of the results, is this better represented on a log-log plot?

We have followed the Reviewer's suggestion of representing the values on a log-log plot. The resulting figure is indeed much more clear. We are not sure that indicating the χ^2 threshold would be pertinent in this figure, because the message we would like to convey is more general. The absolute value of χ^2 also depends on the assumed value for the measurement error, and changing the error shifts χ^2 up or down. But the message of this figure is that most of the retrievals performed using the LUT+PT scheme end up with a worse χ^2 than the corresponding retrievals performed using the NN+PT scheme (regardless of whether their final χ^2 is above or below the chosen threshold).

Section 5.2, general: The presumption in the comparisons to AERONET data is that AERONET represent the well calibrated best representation of truth. This is probably the case for total optical depth, but the retrievals actually have less information content than polarization sensitive ground-SPEX, right? This should probably be pointed out in some way – one explanation of differences between groundSPEX and AERONET could also be that groundSPEX is more accurately representing geophysical 'truth'. I would think this is most likely the case for the imaginary refractive index, which are shown (Dubovik et al JGR 2000, I think?) to be not particularly accurate.

We agree with the Referee. In the revised version of the paper we have included a statement that discusses this fact.

Figures 5-9: I think the time series plots are useful, but it might also be useful to have a scatterplot of matching AERONET/groundSPEX data points. That way you can represent the entire retrieval dataset rather than just four days, and look for systematic differences. Of course, this does challenge my previous comments about the use of AERONET as truth, so that would need to be discussed as well.

The four days of data shown in the time series plots encompass all the AERONET/groundSPEX matchings we currently have. Representing these data in scatter plots is a legitimate alternative to the time series plots, but we think this would not add very much to the plots that are already present. Therefore we would prefer not to add new plots to this manuscript. For the interested reader, some scatter plots of AERONET versus groundSPEX retrievals, limited to the case of the NN first guess, can be found in a paper by van Harten et al. (2014), published on AMTD and recently accepted for AMT.

Appendix A, page 9072, line 7 (grammar): change "worth" to "worthwhile"

Done.

REFERENCES

van Harten et al. (2014), "Atmospheric aerosol characterization with a ground-based SPEX spectropolarimetric instrument", Atmos. Meas. Tech. Discuss., 7, 5741-5768, doi: 10.5194/amtd-7-5741-2014

Interactive comment on Atmos. Meas. Tech. Discuss., 7, 9047, 2014.