Anonymous Referee #1

This paper describes the result of applying a Neural Network approach to UV-VIS spectral measurements from the OMI satellite to determine tropospheric ozone. Retrieving tropospheric ozone from UV-VIS measurements is at the edge of what is possible both fundamentally but also instrument-wise. Hence, any information on what can be achieved is highly valuable and appreciated. The NN approach is a fundamentally different approach than more standard methodologies that have been developed over the last decade or so (residual methods, optimal estimation). It is therefore interesting and relevant to evaluate the results from the NN approach with those from other methods, which is presented in this paper. The paper is well written, concise and well organized, and in general I have no major remarks or criticism with what is presented.

Thank you for the kind words.

However, I have three more general comments which with probably just a little bit of effort could help improve the paper. Furthermore, there is a small list of minor issues and/or suggestions.

GENERAL COMMENTS

A) Distribution of ozone sonde stations:
The ozone sonde locations used in this study are all from Northern Hemisphere midlatitudes. The question arises whether or not campaign data from the Northern Hemisphere can be really considered a true “out of sample” test. After all, conditions in the Northern Hemisphere to some extent are similar: often a thick stratospheric ozone layer and lots of ozone variability – both chemically and dynamically – in the troposphere and in the stratosphere. Measurements from the tropics and the Southern Hemisphere provide quite different conditions. In the tropics stratospheric ozone is rather constant, and variability in tropospheric ozone occurs on longer timescales than outside of the tropics. Furthermore, the tropical tropopause is much higher and there are places where little to no tropospheric ozone is present (tropical Pacific). In the Southern Hemisphere one has to deal with ozone depleted stratospheric layers in the vicinity of Antarctica and in general a thinner stratosphere. I do not think it is necessary to perform analysis for those regions in this study – the contents of this paper are sufficient to present a case for the NN as a proper functioning algorithm, but at some point comparison with data from the tropics and the Southern Hemisphere provides a highly interesting and probably more independent “out of sample” test. Hence, I would strongly recommend some discussion of such work in the “conclusions”, for example as “future work”. Could be done in a couple of sentences.

A) Also with reference to (Anonymous Referee #2, Major Issue A), we decided to underline that our algorithm has only been trained and tested over northern mid-latitudes by changing the title to “Tropospheric ozone column retrieval at northern mid-latitudes from the Ozone Monitoring Instrument by means of a neural network algorithm”. In general, it is not simple to ensure a sufficient generalization to use our NNs in areas with very different climatologies. To take this aspect into account, we also included a comment at the end of the conclusions: “Studied are planned to extend the training set with tropical and Southern Hemisphere samples and/or to develop dedicated algorithms which can be effectively used at different regions, i.e. with different climatologies.”

B) Validation statistics and collocation criteria.
Selection criteria for what is considered a collocation of ozone sondes are missing, for example on page 4/5, paragraph 2.2/2.3. Please provide how a collocation is selected, i.e. within what spatial and temporal interval is a sonde measurement considered to be collocated with an OMI measurement. This is also important information that is lacking for the comparison with sonde measurements later in section 3. Furthermore, it is unclear to what extent a change of the criteria improves the validation as presented in section three.
Some information on this would be highly appreciated. For example, it is stated in section 4.1 that the results appear to be insensitive to ozone enhancements in the upper troposphere. Although it is well known that the sensitivity of UV-VIS ozone profile retrievals for the upper troposphere is limited, such a difference could – in theory – also be the result of collocation issues, i.e. such enhancements may have a certain horizontal size but due to the collocation criteria the enhancement may be missed by OMI as it is not measuring at the geographical location where the enhancement occurs (think: a stratospheri-troposphere exchange event). If the results are robust (or not) with regard to the collocation criteria then that is an indication that this is not the case (or could be the case). Either way, it is valuable information.

B) Please note that matching criteria are briefly discussed at page 2498, lines 9-12: “A matching procedure is implemented to create the input-output pairs for the OMI-TOC NN. OMI pixels are considered to match with ozone sonde launches if data are collected within a maximum time interval of 12 hours with respect to the OMI overpass, and if the station is contained within the OMI pixel”. The input-output pairs are obtained in the same manner for both the training and the test dataset, so we think it's not worth re-discussing these criteria in Section 3. Regarding these criteria, they’ve been chosen to maximize the performances of the NN. In fact, a fair compromise has to be found between the number of the samples and their quality; of course, while in general the quality is expected to get better with stricter matching rules, the number of samples decreases. We didn't make a systematic study of the dependency of the validation results on the collocation criteria – it is likely to be a very long work and we preferred to use good sense, instead - but we expect to have a good compromise with the adopted criteria.

C) Vertical Sensitivity:
The results clearly indicate that there is limited sensitivity to the lower troposphere or – my interpretation - to smaller changes in ozone throughout the troposphere. This is no surprise, as correctly noted by the authors, since others using different methods have come the same conclusions. However, this raises the particular question on how to actually use tropospheric ozone columns from the NN method. For other methods - residual methods or OE methods - it is possible to define a vertical sensitivity, either for the total ozone column used in residual methods or for the ozone profile, the so-called averaging kernel. When OMI tropospheric ozone measurements then are compared to other measurements or model data that vertical sensitivity should be taken into account – even though this is frequently and conveniently forgotten, in particular for the residual methods. Nevertheless, vertical sensitivity information is available. The question thus becomes how to address this with NN data. I could not find it here, but some discussion on how to do that would be useful for potential users. I'm not even sure if it is possible within the NN framework, but that conclusion then is also relevant for potential users. Please provide a short discussion of this issue.

C) As said by the Referee #1, an estimation of the vertical sensitivity in OE methods is generally given by means of the averaging kernels. NNs, being based on soft computing paradigms, simply doesn't have this information. We accounted for this aspect by adding the following sentence at line 2 of page 2500: “Please note that, differently from physically-based methods, NNs don't allow the calculation of averaging kernels and so a formal definition of the vertical sensitivity of the output cannot be given here.”

MINOR COMMENTS

All minor comments have been taken into account and changes into the text have been made accordingly. One point that needs attention:

1) Page 4/5, paragraph 2.2/2.3. I am missing something on the OMI row anomaly and if for the profile retrievals affected pixels are excluded. The row anomaly started halfway 2007, hence some part of the period considered here is affected. Providing a short explanation would do.
1) We agree and we added the following sentence at the end of section 2.2: “Pixels affected by row anomalies (http://www.knmi.nl/omi/research/product/rowanomaly-background.php), after May 2007, have been implicitly filtered out and then not used within our work, based on a threshold on fitting residuals for the correlative OE algorithm (see section 3).”

Anonymous Referee #2

This paper discusses the retrieval of tropospheric ozone from OMI using a statistically neural network algorithm. I regard this contribution as potentially important and may be published once the issues raised below are adequately solved.

MAJOR ISSUES

A) The statistical NN retrieval method presented in the paper is applicable only to the 30N-60N region. This limitation should be properly reflected in the title, abstract and conclusions.

A) Also with reference to (Anonymous Referee #1, General Comment A), we decided to change the title to “Tropospheric ozone column retrieval at northern mid-latitudes from the Ozone Monitoring Instrument by means of a neural network algorithm”. We also added two lines at the end of the conclusions to stress that our algorithm is trained, and then should be used, only at northern mid-latitudes, and that our future work will regard the development of dedicated algorithms for other climatologies and/or the extension of the training dataset of the present NN. We also think that the abstract contains already a reference to the area where our algorithm has been trained, and then we don’t think that further specifying is here necessary.

B) The analysis of the NN results is performed only in the same geographical region of the OS stations used for the NN training. The performance of the OMI-TOC NN over the complete 30N-60N region should be assessed by adding comparison of the NN to the OE and TOR results over the ocean regions (Pacific and Atlantic) as well as the Eurasia region where no OS stations are available.

B) This paper focuses on the development of the algorithm itself and a comparison with more established algorithms is only given to evaluate its performance where high quality correlative measurements, i.e. OS data, are present. Then we disagree with the Referee #2 on the need of extending the comparison and to do this also where OS measurements are not available. Please note that in the paper it is clearly stated that the comparison of our NN with OE and TOR is limited to some OS stations, and then our algorithm’s performances are only evaluated there. However, we agree that a more extended evaluation of our NN is desirable, and future work on this issue is planned.

C) In the same way, the validity of the NN results outside the time covered by the training period should be assessed, i.e. comparisons with measurements in 2009 and 2010 should be included.

C) Rather than using our NN outside the time covered by the training period, we base our algorithm on a regular refresh of the network coefficients by re-training the net with updated data sets. We’ve used this approach also for our NN for SCIAMACHY data [Sellitto et al.: Tropospheric ozone column retrieval from ESA-Envisat SCIAMACHY nadir UV/VIS radiance measurements by means of a neural network algorithm, Accepted for publication on IEEE Transaction on Geoscience and Remote Sensing, 2011]. For details, please refer to (Anonymous Referee #2, Detailed Comment 5).

DETAILED COMMENTS
Almost all minor revisions have been accepted and changes into the text and figures have been made accordingly. Some points that need attention:

1) Page 2493, lines 5-12. The meaning of the fist sentence is not clear. Limb and occultation sensors for obtaining “height-resolved atmospheric ozone” (e.g. MIPAS, GOMOS, etc.) are not mentioned at all.

1) We limited our discussion to nadir measurements because we focus on the troposphere. Limb measurements, e.g., are not suited to make measurements of tropospheric ozone. To clarify this aspect we changed the sentence at line 5, page 2493: “Observing height-resolved atmospheric ozone, with a particular emphasis to the troposphere, from satellite is thus... "

2) Page 2493, lines 18-23. Only the advantages of the NN methods are listed. The disadvantages of statistically NN methods should be also discussed (e.g. need of external measurements for training, dependency of the results on the quality of the external measurements, limitation to geographical regions where training data is available, etc.).

2) We added the following sentence at page 2494, line 22: “...to operate in real time. However, their performances depend on the quality of the training set and a statistically complete set of external measurements is not always available”.

3) Page 2496, line 17. What is exactly mean with “to look between clouds”?

3) We changed the sentence at line 17, page 2496: “The small pixel size in principle enables OMI to have more cloud free pixels.”

4) Page 2496, lines 22-23. Properly justify why the total ozone is used as input to the NN. What is the difference on NN TOCs obtained with and without total ozone?

4) We added the following discussion at line 27, page 2499: “The value of the total column ozone is expected to act as a regularization parameter when used into the input vector of our net. We compared the results of our algorithm with those from a NN with the same input vector and topology definitions, and trained and tested over the same datasets, but without the total ozone in input. We found that the performances of this latter algorithm are more than 20% worse in terms of both the correlation coefficient and the RMS deviations of retrieved versus reference OS TOCs over the test dataset.”

5) Page 2499, lines 22-23. The viewing geometry is only partially described with the SZA. Properly justify why the viewing zenith angle and relative azimuth angles are not used. Page 2499. OMI as all UV instruments has a relative strong degradation. Using ratios of radiance to irradiances compensates only partially the degradation effects. Therefore the NN input vector should include time information to allow the NN to also compensate for degradation.

5) The main author discussed extensively why not using VZA, RAA and time into the input vector of a similar NN algorithm for SCIAMACHY data [Sellitto et al.: Tropospheric ozone column retrieval from ESA-Envisat SCIAMACHY nadir UV/VIS radiance measurements by means of a neural network algorithm, Accepted for publication on IEEE Transaction on Geoscience and Remote Sensing, 2011]. As for VZA and RAA, while in OE algorithms it is necessary to know exactly the geometry of observation to properly run the radiative transfer model with the aim of producing the synthetic spectra to be compared with the measured spectra to fit the ozone concentrations accordingly, in our method the rationale is that the information is entirely contained in the input-output patterns and that the NN learns the retrieval rule from the training process. Each measurement in the training set is taken with specific VZA and RAA, and it is not necessary to account for them explicitly into the input vector. Please note that we tried to train and test NNs also with these parameters into the input vector and we obtained no statistically significant differences in the results. SZA, on the contrary, must be included because it gives, along with Earth's radiance and solar irradianc, the complete reflectance information, which is necessary to retrieve trace
gases concentrations from satellite data [Lichtenberg et al.,: SCIAMACHY Level1 data: Calibration concept and in-flight calibration, Atmos. Chem. Phys., vol. 5, pp. 8925–8977, 2005]. We conclude that SZA does add information, whereas VZA and RAA are redundant with respect to the other inputs and for the considered neural algorithm. Based on this considerations, we do not include VZA and RAA into the the input vector. As for the time information, we think that adding a time channel to the input layer is not expected to be beneficial, since, apart from deterministic effects, the numerical values of the temporal variable do not carry specific information on the kind and amount of variations intervened in the sensor. Actually, a different approach has been reported in [Muller et al.,: Ozone profile retrieval from Global Ozone Monitoring Experiment (GOME) data using a neural network approach (Neural Network Ozone Retrieval SYstem (NNORSY)), J. Geophys. Res., 108(D16), 4497–4515, 2003], where, however, the authors don't clarify if and how time is helping in compensating radiometric calibration and degradation. We have doubts about the way the time numerical value injected into the nonlinear computational chain could succeed in counteracting unpredictable variations of the input-output relations caused by sensor changes. In addition, a validation of this approach, e.g., training a NN with or without time information over a period of time and then testing over different periods of time, e.g. years later, is not simple to address. Rather, although a detailed description of this approach is not a major objective of the present paper, we based our method on a regular re-training of the NN with updated data sets over the considered period of time. These considerations have an impact also for (Anonymous Referee #2, Major Issue C).

6) Pages 2513-2524. Add figures or tables showing the comparisons between the NN and TOR.

6) As the comparison is based on different data sets, we prefer to avoid the emphasis of a table and to keep the comments in a text form.

7) Pages 2523-2524. Put together Fig. 11 and Fig 12 as they show results for similar test cases.

7) In the revised manuscript we will put together figures 11 and 12.