Reply to Anonymous Referee #2

We would like to thank the Referee #2 for his/her insightful comments. His/her suggestion of assessing the informative contribution of the OMI reflectances by training a second NN with ancillary data adds substantial value to the paper, and also other comments have been helpful in making the paper more clear. We agree less upon the fact that the results presented here are not new/extensive enough to justify a publication, as explained in more detail later.

MAJOR COMMENTS

Neural networks are used as a statistical retrieval tool taking as input a significant amount of ancillary data. The actual contribution of OMI reflectance to the obtained TCO compared to the ancillary data (note the strong correlations between actual TCO, temperature profile, tropopause pressure and TCO climatology) should be properly assessed. This can be done by training a second NN with only the ancillary data (without OMI data as input) and comparing the results obtained with OMITROPO3-NN.

We have followed the suggestion given by the Referee. By comparing the NN trained with only the ancillary data with OMITROPO3-NN, we observed that the second NN performs considerably worse with respect to estimation of the TCO anomaly. In the revised version of our manuscript be allowed, we have included this analysis in a subsection of Section 5.

The TCO from the NN algorithm are only validated with ozonesonde data (i.e. the same type of data used for the NN training) and it is limited to a few locations over land. The comparison with model is restricted to only two days. This limited validation is not sufficient to justify a publication of an improved NN algorithm that has been al-ready presented for OMI (Sellito et at., 2011) and SCIAMACHY (Sellito et at., 2012).
On top of that, the aim of the proposed algorithm is "Global tropospheric ozone column retrievals", therefore the validations should be done on a *global* scale using appropriated tropospheric ozone data from other instruments like for example MLS or other OMI retrieval algorithms like for example Liu et at., 2010.

This comment can be summarized by the following propositions, that will be addressed separately in our reply.

P1 – Ozonesonde data are of the same type of the data used to train the NN.
P2 – A limited validation effort has been carried out.
P3 – In this paper, not much else than the validation of an already published algorithm is being presented, together with some improvements. This makes the amount of work insufficient for a publication.
P4 – A validation against other satellite products is compulsory in order to justify the adoption of the adjective "global" in the title.

P1. P1 seems to imply that using data measured by the same type of instrument as those included in the training set takes something away from the significance of our validation. Our opinion is that what makes a validation of a retrieval algorithm scientifically significant is the fulfillment of the two following main requirements:

- R1) the reference data should not have been used to train the retrieval algorithm (this point is specifical for statistical techniques as NNs);
- R2) the reference data should be considered more reliable than the retrieval results, so that the discrepancies can be attributed to the retrieval method and not to the data used as reference.

Our validation dataset satisfies R1, because the data used for validation have not been included in the training set, and also a separation between ozonesonde stations has been performed (which is even more than necessary in order to fulfill R1). Furthermore, our validation data satisfy R2, because to our awareness ozonesonde data are the most reliable reference for tropospheric ozone.

P2. We think that the term “limited” associated with our validation effort is debatable. We have encountered other published papers where retrieved tropospheric ozone columns or ozone profiles are compared to ozonesonde data only (e.g. in Liu et al. (2005) the tropospheric ozone columns are compared to ozonesonde data, whereas limb satellite data are used to validate the ozone profiles in the stratosphere – which makes sense, because satellite retrieved stratospheric profiles are more reliable than satellite tropospheric ozone columns; in Müller et al. (2003) the retrieved ozone profiles are compared to ozonesonde data, whereas satellite data are used for the total columns, and again this is a meaningful choice, because satellite total columns can be considered as very good reference data). Therefore, we do not really think that our validation effort is below the standards for a publication. As for the comparison with other satellite tropospheric ozone estimates, we do see some merit in this task, and we would like to carry out an extensive study in the future (see also our reply to Referee 1), but we would not put such a task at the same priority level as a comparisons against ozonesondes. Other satellite TCO products usually have uncertainties that are of the same order as those observed in our analyses, therefore we doubt that we could regard other satellite observations as a real “truth” for a proper validation.
As for the comparisons with the CTM, they are just meant as single examples, and not as an extended validation. In the revised version of our manuscript, we have added that analyses carried out in some other dates during October 2006 (presented at the ESA Atmospheric Science Conference 2012 and at the 9th International Symposium on Tropospheric Profiling) have shown a similar behaviour in terms of RMSE and correlation coefficient. We have added references to the extended abstracts by Di Noia et al. (2012a,b). Unfortunately we do not have the processing capabilities needed in order to carry out more systematic analyses like global comparisons between monthly means: this would require too much time.

P3. As for the novelty of the algorithm, we recall that we have introduced the following improvements: (i) Extension of the algorithm to all the latitude bands; (ii) removal of the static upper limit of 200 hPa in the definition of tropospheric ozone (which makes a major difference in the definition of the output product at midlatitudes); (iii) use of ancillary data from external sources in order to yield more accurate TCO estimates by performing a fusion between these data and OMI reflectances; (iv) a different preprocessing of the OMI reflectances. We would argue that the only thing this algorithm has in common with that shown in Sellitto et al. (2011) is that it is based on neural networks. For the rest, it has been redesigned from scratch.

P4. The adjective “global” in the title is used to emphasize that the algorithm can be applied to all latitudes. In our opinion, this feature should be pointed out. If the title of the paper by Sellitto et al. (2011) mentions “tropospheric ozone column retrieval at northern midlatitudes”, then we think that it is legitimate to mention “global tropospheric ozone column retrievals” here.

MINOR COMMENTS

P 7676, L2
"a new NN algorithm ..." this is not really a new algorithm, but an improvement of an al-gorithm already presented by Sellito et at. (2011 for OMI) and (2012 for SCIAMACHY).

We still think that the number and the significance of the improvements that have been introduced make this a "new NN algorithm", and not just "an improvement of an algorithm already presented". Therefore, we would prefer to leave this statement as it is.

P 7677, L 26
The work of Valks et al. is related to GOME/ERS-2 and not TOMS.

We are aware of this. The sentence was originally meant to cite a list of papers describing cloud slicing methods, without dividing them by instrument. However, we have to admit that the sentence is a bit misleading, because Valks et al. (2003) is the only non-TOMS paper included in the list.Therefore we have expanded our brief review on other tropospheric ozone retrieval methods separating the cited papers based on the instrument they refer to, also in view of the citations suggested by the Referee #1.

P 7678, L 18-19
Add a reference to a validation paper showing that the RMS of OMI-TOC NN is 8 DU.

We have added references to Sellitto et al. (2011) and a paper by Di Noia et al. (2013), recently accepted for publication on EURASIP Journal on Advances in Signal Processing.

P 7678, L 11-12
Explain in detail the "several refinements" used for quality control and filtering.

The most important refinements refer to: 1) Full implementation of the KNMI technical recommendations on the interpretation of OMI quality flags (in OMI-TOC NN all the spectral pixels having nonzero QF were discarded); 2) Solution of some issues related to the spectral interpolation of OMI reflectances, which could have caused some spectral misalignment in the processing chain of the OMI-TOC NN; 3) Introduction of a cloud filtering threshold. Since these are quite technical refinements (especially 1 and 2), we are not sure that it is really worth to list them in the text. As for the cloud threshold, this is already mentioned later.

P 7680, Eq. (1)
The symbol Phi is not explained.

Phi_W is the symbol of a function parameterized by W. To make this more clear, we have modified L9 by writing "...through a nonlinear model Phi_W, such that..."

P 7681, L 17-19
What is the link between the text after ii) and the MLP used in this paper?

Since in that section we describe the general approximation properties of MLPs, we just wanted to point out that MLP is not the only type of NN that has these properties. However, we have removed this sentence, because it is not really fundamental in our framework.

P 7682, L 18
If a component of x has no effect on y then this component shouldn't be used as part of the state vector.

We have removed this sentence.

P 7684, Eq (7)
Define the x^ on the left side of the equations.

We have changed L6 by writing "...a non parametric estimator for x, here denoted with x^..."

P 7684, L 26
There are other shortcomings from NN statistical retrievals that are not mentioned like (1) the solution for a particular observation may not be optimal, the cost function for a single measurement can be better minimized with classical retrieval techniques and (2) a proper error propagation can not be done.

We agree on point (1) and we have included this in the revised version of our manuscript. We would be less categorical on point (2), as significant research has been carried out on error propagation in NNs. We have mentioned that, to our awareness, most of the error propagation methods for NNs rely on the assumption that the suboptimality of the training process and of the NN model has a negligible impact on the estimated error budget.

P 7686, L 16
Add a reference to the "early stopping cross-validation"

We have added a reference to the book by Haykin (1999).

P 7686, L 26
Justify why the surface albedo is not used as input

In the revised version of the manuscript, we have specified that we have extended the wavelength range of the algorithm beyond the ozone Huggins absorption bands, so as to include wavelengths containing information on aerosols (around 342 nm) and surface albedo (around 345 nm).

P 7687, L 8
What is the source of the cloud fraction? Why is the cloud-top height not used in the retrieval?

The cloud fraction comes from the OMI Rotational raman (OMCLDRR) product.
We have tried to include the cloud-top height in the input vector, but this did not improve our retrievals. We have inserted this information in the revised text.

P 7689, L 15-21
Is the measurement time not used as co-location criteria?

We forgot to mention in the text that we used 6h as temporal matching criterion. We have corrected for this in the revised manuscript.

P 7690, L 11
Discuss the errors induced by using just a linear interpolation

The reason why we have chosen a linear interpolation instead of more powerful techniques like cubic splines is the processing time. We have tested that using cubic splines would have multiplied the processing time by a factor 4 on our computers. Unfortunately we cannot provide a quantitative assessment of the errors introduced by the interpolation method.

P 7691, L 18
Add a quantitative description of the reconstruction errors due to using 20 PCs

We have mentioned that using 20 PC a RMS reconstruction error of less than 0.1% was found in the reflectances.

P 7693, L 7
Define the Pearson correlation coefficient or provide a reference

The Pearson correlation coefficient is the ordinary *r* correlation coefficient between two variables. We think that it is too standard as a statistic to justify an explicit definition or a reference (it would be like explicitly defining the mean or the standard deviation).

P 7694, L 15
Provide at least a preliminary explanation for the underperformance at the Arctic region

Our initial idea, already described at page 7694, L24-25, was that it could be related to cloudy pixels that have not been detected as such. However, the analysis proposed by Referee 2 revealed that also the NN trained with ancillary data underestimates over the Arctic, even more than the OMITROPO3-NN. Therefore, another possible cause might lie in possible problems in the tropopause definition, since in the Arctic it could be difficult to identify the tropopause precisely. We have inserted this consideration in the revised text.

P 7696, L 21
Indicate which ozone data is used as input for TM5

According to Williams et al. (2012), the constraint on stratospheric ozone comes from the Fortuin and Kelder (1998) climatology, rescaled to a total ozone reanalysis. We are not sure, however, that this is the information requested by the Referee. Since we could not add much more with respect to Williams et al. (2012) in this explanation, we would prefer to leave the reference to that paper as a pointer to the relevant information.

P 7709, Table 1
Indicate the wavelength coverage of UV1 and UV2

We have done that in the caption.

P 7717-7718
Figures 5 and 6 could be merged.

We would prefer to leave them as they are, because maybe it gives more flexibility in the typesetting of the manuscript.

P 7724-7725
What is the value added of Fig 13 with respect to Fig 12? Consider showing only one these figures.

We have removed Fig. 13.

P 7726, Fig. 14
Why are the regions without data much smaller in this figure than in Fig. 9 (large no retrieval areas) or Fig. 15?

Fig. 9 shows TCO maps at full OMI spatial resolution. Same holds for Fig. 15. In Fig. 14, in order to compute the differences, we had to remap the OMI TCO fields on the TM5 resolution, that is coarser. For each TM5 grid cell, the remapping method takes all the OMI pixels whose centers lie in the grid cell and computes the median between the non-missing TCOs. Since the TM5 pixels are much larger than the OMI pixels, it is more likely for a TM5 pixel to contain some non-missing data. This explains the smaller no retrieval areas. We have changed the description of the remapping method in the text, because it was erroneously reported that a nearest neighbour resampling had been used.

References

Di Noia, A. et al. (2012a), “Global tropospheric ozone retrievals from OMI data by means of neural networks”. Proceedings of ESA Atmospheric Science Conference 2012, Bruges, Belgium, 18-22 June 2012, in press

Di Noia, A. et al. (2012b), “A neural network algorithm to retrieve tropospheric ozone from the Ozone Monitoring Instrument: Design and validation”. Proceedings of the 9th International Symposium on Tropospheric Profiling, L'Aquila, Italy, 3-7 September 2012

Di Noia, A. et al. (2013), “Tropospheric ozone column retrieval from OMI data by means of neural networks: A validation exercise with ozone soundings over Europe”, EURASIP J. Adv. Sig. Pr., accepted for publication

Haykin, S. (1999), “Neural Networks: A Comprehensive Foundation”, Prentice Hall

Liu, X. et al. (2005), “Ozone profile and tropospheric ozone retrievals from the Global Ozone Monitoring Experiment: Algorithm description and validation”, J. Geophys. Res. 110, D20307, doi: 10.1029/2005JD006240

Müller, M. D. et al. (2003), “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, doi: 10.1029/2002JD002784

Sellitto, P. et al. (2011), “Tropospheric ozone column retrieval at northern mid-latitudes from the Ozone Monitoring Instrument by means of a neural network algorithm”, Atmos. Meas. Tech. 4, 2375-2388, doi: 10.5194/amt-4-2375-2011

Williams, J. E. et al. (2012), “The application of the Modified Band Approach for the calculation of on-line photodissociation rate constants in TM5: implications for oxidative capacity”, Geosci. Model Dev. 5, 15-35, doi: 10.5194/gmd-5-15-2012