

RESEARCH

Open Access

# Tropospheric ozone column retrieval from OMI data by means of neural networks: a validation exercise with ozone soundings over Europe

Antonio Di Noia<sup>1,5\*</sup>, Pasquale Sellitto<sup>2</sup>, Fabio Del Frate<sup>1</sup>, Marco Cervino<sup>3</sup>,  
Marco Iarlori<sup>4</sup> and Vincenzo Rizi<sup>4</sup>

## Abstract

The retrieval of the tropospheric ozone column from satellite data is very important for the characterization of tropospheric chemical and physical properties. However, the task of retrieving tropospheric ozone from space has to face with one fundamental difficulty: the contribution of the tropospheric ozone to the measured radiances is overwhelmed by a much stronger stratospheric signal, which has to be reliably filtered. The Tor Vergata University Earth Observation Laboratory has recently addressed this issue by developing a neural network (NN) algorithm for tropospheric ozone retrieval from NASA-Aura Ozone Monitoring Instrument (OMI) data. The performances of this algorithm were proven comparable to those of more consolidated algorithms, such as Tropospheric Ozone Residual and Optimal Estimation. In this article, the results of a validation of this algorithm with measurements performed at six European ozonesonde sites are shown and critically discussed. The results indicate that systematic errors, related to the tropopause pressure, are present in the current version of the algorithm, and that including the tropopause pressure in the NN input vector can compensate for these errors, enhancing the retrieval accuracy significantly.

**Keywords:** Neural networks, Remote sensing, Satellite, Ozone

## 1 Introduction

Tropospheric ozone is a key player in a number of atmospheric processes that affect both climate and air quality. Its climatic impact is expressed by a radiative forcing of about  $0.35 \text{ W/m}^2$ , as estimated by the intergovernmental panel on climate change (IPCC) fourth assessment report [1]. Such radiative forcing makes tropospheric ozone the fourth atmospheric greenhouse gas by importance, following water vapor, carbon dioxide and methane [1]. As for the air quality, tropospheric ozone has both a positive and a negative role; its positive role lies in the fact that it acts as a precursor of the hydroxyl radical, which is able to remove several pollutants from the middle troposphere through oxidation reactions [2]; its negative role lies in its

toxicity for both humans and crops when it reaches high concentrations near the Earth's surface [3,4].

Monitoring the concentration of tropospheric ozone from a satellite platform offers the advantage of a temporally and spatially continuous observation, allowing the identification of long-range transport processes [5,6], and the generation of temporally extended records, which are useful for the investigation of long term trends [7-9].

In the last two decades, the advent of a new generation of satellite hyperspectral atmospheric sounders, which make simultaneous radiance measurements with high spectral resolution and sampling rate, covering the ultraviolet (UV), visible (VIS) and infrared (IR) spectral ranges, has greatly enhanced our capability to detect and quantify several tropospheric trace gases, including ozone [10].

Among the tropospheric gases that can be monitored from space, ozone is one of the most problematic ones. In fact, the contribution of tropospheric ozone to the measured radiance signal must be separated from the contribution of stratospheric ozone, which is much larger,

\*Correspondence: dinoia@disp.uniroma2.it

<sup>1</sup> Earth Observation Laboratory, Department of Civil and Computer Engineering, Tor Vergata University, Via del Politecnico 1, 00133 Rome, Italy

<sup>5</sup> Now at: SRON Netherlands Institute for Space Research, Sorbonnelaan 2, 3584 CA Utrecht, The Netherlands

Full list of author information is available at the end of the article

due to the fact that most of the atmospheric ozone is found in the stratosphere. In order to accomplish this, several techniques were developed during the last 20 years. The rationale behind the first tropospheric ozone retrieval algorithms was to isolate the stratospheric ozone column by means of limb measurements [11,12] or total ozone retrievals over high-altitude clouds [13,14], and then subtract it from a co-located or neighboring measurement of the total ozone column. In the case of limb measurements, the separation between stratosphere and the troposphere is achieved thanks to the limb viewing geometry, whose line of sight does not encounter the atmospheric layers located beneath the upper troposphere/lower stratosphere (UTLS). In the case of measurements over high clouds, it is assumed that such clouds shield the underlying troposphere, and that the stratospheric ozone field does not have a significant horizontal variability within a certain number of neighboring pixels. If these assumptions hold true, it is possible to say that total ozone column retrievals over high altitude clouds actually represent stratospheric columns, which can be subtracted from total ozone columns retrieved over neighboring clear-sky pixels to yield an approximated value for the tropospheric ozone column. This type of approach has been mainly used over the Tropics, where high convective clouds are more frequent.

During the last decade, the improved sensitivity to the lower tropospheric layers that was achieved with new satellite instruments—including the Global Ozone Monitoring Experiment (GOME), the SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY) and the Ozone Monitoring Instrument (OMI)—has enabled the development of algorithms that directly derive tropospheric ozone information from ozone profiles retrieved through an optimal estimation (OE) scheme [15-18].

OE retrieval schemes make use of forward radiative transfer models (RTMs), which are computationally intensive and require a consistent characterization of the whole atmospheric state, including the properties of clouds, aerosols and spectrally interfering trace gases (i.e., gases that have absorption features in the same spectral region as the trace gas of interest), which in most cases must be assumed *a priori*. This can cause the retrieval process to be slow and sensitive to wrong *a priori* assumptions, as well as to forward modeling errors [19].

An alternative approach to the direct determination of tropospheric ozone from satellite measurements is represented by neural network (NN) algorithms. Instead of explicitly using a forward model, NNs attempt to approximate the relationship between the measured radiance and the atmospheric parameter of interest directly by means of a nonlinear regression on a given training set [20]. In the case of atmospheric retrievals, the training set for a NN

algorithm will consist of simultaneous realizations of the radiometric measurements and the geophysical process of interest. In addition, other parameters that can be useful to better constrain the relationship between the radiance measurements and the parameter to be retrieved (e.g., information on the observation geometry, other atmospheric parameters) can be given as inputs to a NN. For the training of a NN to be successful, a large and comprehensive training set must be built, possibly covering all the atmospheric situations that can be encountered in reality (e.g., heavy pollution events, tropopause folds).

Although the training process can be slow, a trained NN is able to operate very quickly, which is an attractive feature for operational retrievals. Furthermore, NNs allow to handle heterogeneous data in an easy way. This is an important feature when a complex model relating a large number of different quantities (e.g., atmospheric optical thickness, tropopause height and tropospheric ozone column) cannot be explicitly formulated, although it is known that a physical correlation between these quantities exists. On the other hand, a disadvantage of NNs lies in the difficult interpretation of their results. Such difficulty arises from the fact that the physical relationships underlying the retrieval process are represented by a NN in a purely numerical form, without any reference to the causal relationships that link the observed data. Because of this, NN retrieval schemes do not provide diagnostics that measure the relative contribution of each atmospheric layer to the retrievals and the number of independent pieces of information provided by the algorithm—such as the averaging kernels and the degrees of freedom for signal (DFS) [19]—whose computation requires an RTM.

NNs have been successfully applied in several branches of atmospheric remote sensing [21], including retrievals of ozone profiles [22,23], total ozone [24] and tropospheric ozone column [25-27]. Recently, a new NN algorithm for tropospheric ozone retrieval over the northern mid-latitudes from OMI data—named OMI tropospheric ozone column neural network (OMI-TOC NN)—has been proposed [26]. In the present article, the results of a validation of this latter algorithm with ozone soundings performed at a number of European stations are presented.

The article is organized as follows. In Section 2, a brief overview of the NASA Aura-OMI mission is given. In Section 3, a description of the OMI-TOC NN algorithm is given. In Section 4, the ozonesonde sites used for this validation and the co-location criteria are described. In Section 5, the validation results are shown, the temporal trends in the retrieval errors are discussed, and the importance of a parameter which was not originally used in the NN input vector—namely, the tropopause pressure—is demonstrated. The conclusions are drawn in Section 7.

## 2 The NASA-Aura mission and the OMI instrument

The NASA EOS Aura mission [28], started in 2004 with the launch of the homonymous satellite, aims at the study of the atmospheric composition, chemistry and dynamics. The scientific instrumentation onboard the Aura satellite includes the OMI instrument, as well as the tropospheric emission spectrometer (TES), the microwave limb sounder (MLS) and the High resolution dynamics limb sounder (HIRDLS).

The OMI instrument [29] is a nadir UV/VIS imaging spectrometer, that measures direct and backscattered solar radiation in three channels; namely, the UV1 channel (270–310 nm), the UV2 channel (310–365 nm) and the VIS channel (365–500 nm). The UV1 and UV2 channels are the most important ones for ozone monitoring, because they cover the Hartley and Huggins absorption bands of the ozone molecule. The VIS channel is used for observations of clouds, aerosols and other atmospheric trace gases (e.g., nitrogen dioxide, formaldehyde). However, it does not cover the region of the ozone Chappuis absorption bands where the ozone absorption cross section is largest (i.e., about 530–610 nm), and thus it cannot be directly exploited in ozone retrievals.

OMI can observe the Earth's atmosphere in three observation modes. In the main mode—called the Global measurement mode—OMI has a swath width of 2600 km, a nadir pixel size of  $13 \times 48 \text{ km}^2$  (along-  $\times$  across-track) for the UV1 channel and  $13 \times 24 \text{ km}^2$  for the UV2 and VIS channels. The pixel size increases in the swath direction for increasing distances from the satellite ground track. The OMI average spectral resolution is of about 0.4 nm in the UV1 and UV2 channels and about 0.6 nm in the VIS channel. The OMI Global measurement mode provides almost global coverage in one day. In principle, a complete daily global coverage is possible at midlatitudes. However, a complex instrumental effect, called *row anomaly*—which started to appear in the Level 1B data on June 25th of 2007—creates some gaps in the instrumental coverage. More informations on this effect are available from the Royal Dutch Meteorological Institute (Koninklijk Nederlands Meteorologisch Instituut (KNMI)) website [30].

In addition to the Global mode, two so-called “zoom-in” observation modes are available. In both modes the nadir pixel size is reduced to  $13 \times 12 \text{ km}^2$ . In the Spatial zoom-in mode the pixel size is reduced at the expense of the swath width, which decreases to 725 km; in the Spectral zoom-in mode the reduction comes at the expense of the wavelength range, which is limited to 306–432 nm [29]. Zoom-in observations are only performed during selected orbits.

## 3 The OMI-TOC NN algorithm

Recently, a NN algorithm for tropospheric ozone column retrieval from OMI reflectance measurements has

been proposed [26]. From now on, this algorithm will be referred to as OMI-TOC NN. The design and optimization stages of the algorithm are thoroughly discussed in [26]. The OMI-TOC NN was trained and tested with an extended set of ozonesonde measurements taken at the northern midlatitudes between 2004 and 2008. The ozonesonde stations whose data were used in the training set are listed in Table 1. The OMI-TOC NN performances were found to be comparable, and in some cases slightly better, with respect to those of the trajectory enhanced tropospheric ozone residual (TTOR) [12] and OE [17] algorithms over a set of co-located ozonesonde measurements [26]. These results suggest that the OMI-TOC NN is a valuable alternative method for tropospheric ozone retrievals from OMI data.

The input vector for the OMI-TOC NN consists of OMI spectral reflectances at 19 selected wavelengths, extracted from OMI Level 1b data; the solar zenith angle (SZA) and the total ozone column taken from the operational OMI Level 2 product. Only Global measurement mode data were used, because only this observation mode provides daily global coverage. The 19 wavelengths were selected according to an extended pruning (EP) technique [31]. This technique aims at reducing the dimensionality of an input vector for a NN by retaining only the most informative inputs, i.e., those who have the strongest influence on the NN output. Six of the selected wavelengths belong to the 305–307 nm range (covered by the OMI UV1 channel), while the remaining 13 wavelengths lie in the 322–325 nm range (covered by the OMI UV2 channel). The spectroscopic relevance of these two spectral ranges in the context of ozone retrievals is discussed in [27].

The dimensionality reduction of the reflectance spectra is useful for a number of reasons. First, using full spectra would lead to a very big input vector, which would in turn cause a need for a larger training dataset and longer training times. Second, there would be the risk of including irrelevant information in the input vector, which may compromise the learning capabilities of the NN (e.g., by causing overfitting).

In order to homogenize the spatial resolution of the input spectra, the UV2 reflectances were degraded to the spatial resolution of the OMI UV1 channel (see Section 2). The resolution degradation was performed through simple arithmetical averages between pairs of adjacent spatial pixels in the across-track direction.

The output quantity for the NN, i.e., the retrieved parameter, is the integrated ozone column between the surface and the 200 hPa pressure level. From now on, the name tropospheric ozone column (TOC) will be used when referring to this quantity. However, it must be pointed out that the choice of a static upper integration limit in the definition of the TOC—regardless of the actual tropopause height—might be rather inaccurate.

**Table 1 Summary of the stations used in the training set of the OMI-TOC NN. After [26]**

Station	Country	Latitude	Longitude	N. co-locations
Barajas (Madrid)	Spain	40.46°N	3.65°W	107
Boulder	United States	40.09°N	105.25°W	95
Bratt's Lake	Canada	50.21°N	104.71°W	108
Churchill	Canada	58.75°N	94.07°W	91
Egbert	Canada	44.23°N	79.78°W	103
Goose Bay	Canada	53.30°N	60.36°W	126
Hohenpeissenberg	Germany	47.80°N	11.02°E	360
Huntsville	United States	34.72°N	86.64°W	82
Isfahan	Iran	32.48°N	51.43°E	36
Kelowna	Canada	49.93°N	119.40°W	116
Legionowo	Poland	52.40°N	20.97°E	179
Lindenberg	Germany	52.21°N	14.12°E	161
Payerne	Switzerland	46.69°N	6.57°E	463
Sapporo	Japan	43.06°N	141.33°E	117
Tateno	Japan	36.06°N	140.10°E	120
Trinidad Head	United States	40.80°N	124.16°W	57
Uccle	Belgium	50.80°N	4.35°E	390
Wallops Island	United States	37.89°N	75.48°W	57

The problems that can arise as a consequence of this choice are shown and critically discussed in Section 5.

#### 4 Validation set and intercomparison methodology

Six European ozonesonde stations were used in the validation of the OMI-TOC NN: Ankara (Turkey), Izaña (Canary Islands, Spain), Lerwick (Shetland Islands, United Kingdom), Valentia Island (Republic of Ireland), L'Aquila and San Pietro Capofiume (Italy). No data from such stations were used during the training of the OMI-TOC NN. Data acquired between October 2004 and December 2008 were considered in this validation exercise. This is the same period that is covered by the training dataset of the OMI-TOC NN. This choice was made in order to ensure that eventual problems in the algorithm are not caused by instrumental changes that may have occurred after the period covered by the training set.

The data for Ankara, Lerwick and Valentia Observatory were taken from the World Ozone and Ultraviolet Data Center (WOUDC) archive. The data for Izaña were taken from the public archive of the network for the detection of atmospheric composition change (NDACC).

In addition to the data available from WOUDC and NDACC, data from two the two Italian ozonesonde stations of L'Aquila and San Pietro Capofiume were used.

The L'Aquila ozone soundings were performed by the University of L'Aquila and the Centre of Excellence for the integration of remote sensing techniques and modeling for the forecast of severe weather (Centro di Eccellenza di Telerilevamento e Modellistica numerica per la Previsione di eventi Severi (CETEMPS)). The ozonesonde station is located at the CETEMPS atmospheric observatory, Casale Calore di San Vittorino (42.3°N, 13.31°E, 683 m a.s.l.), near the town of L'Aquila. The ozonesondes are SPC-6A type electrochemical concentration cell

**Table 2 Summary of the stations used in the validation presented in this article**

Station	Country	Latitude	Longitude	N. co-locations
Ankara	Turkey	39.95°N	32.85°E	112
Izaña	Spain	28.29°N	16.49°W	227
L'Aquila	Italy	42.38°N	13.31°E	76
Lerwick	United Kingdom	60.13°N	1.18°W	240
San Pietro Capofiume	Italy	44.65°N	11.62°E	12
Valentia Observatory	Republic of Ireland	51.93°N	10.25°W	141

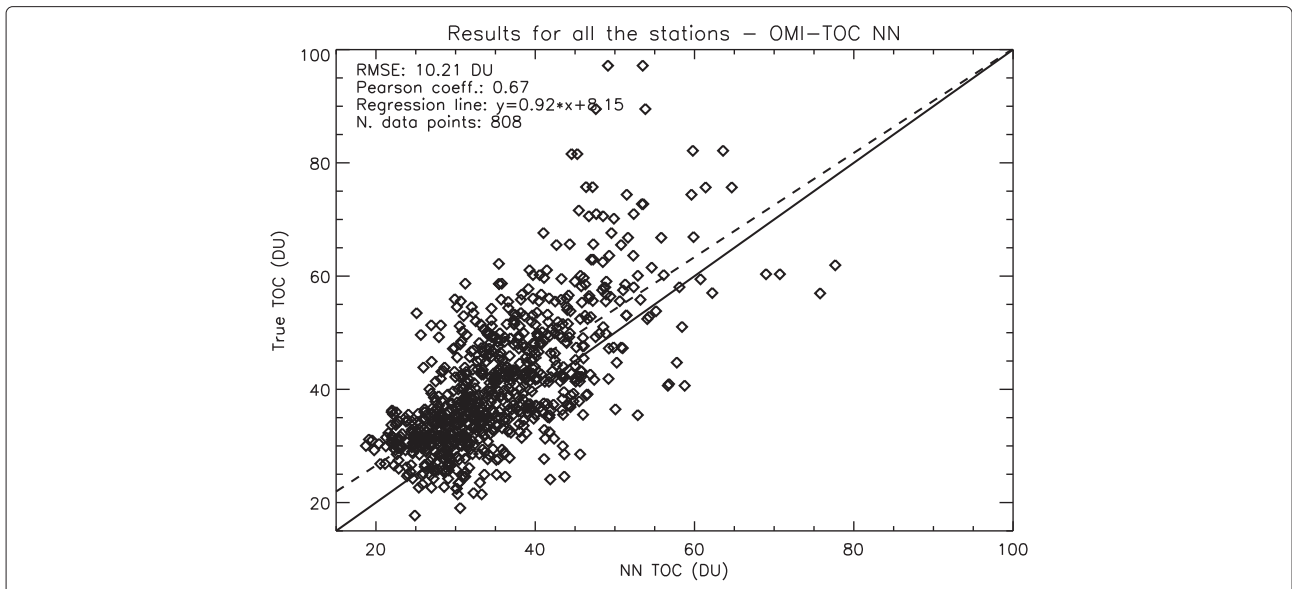


Figure 1 Scatter plot of true versus retrieved TOCs for all the stations considered in this study.

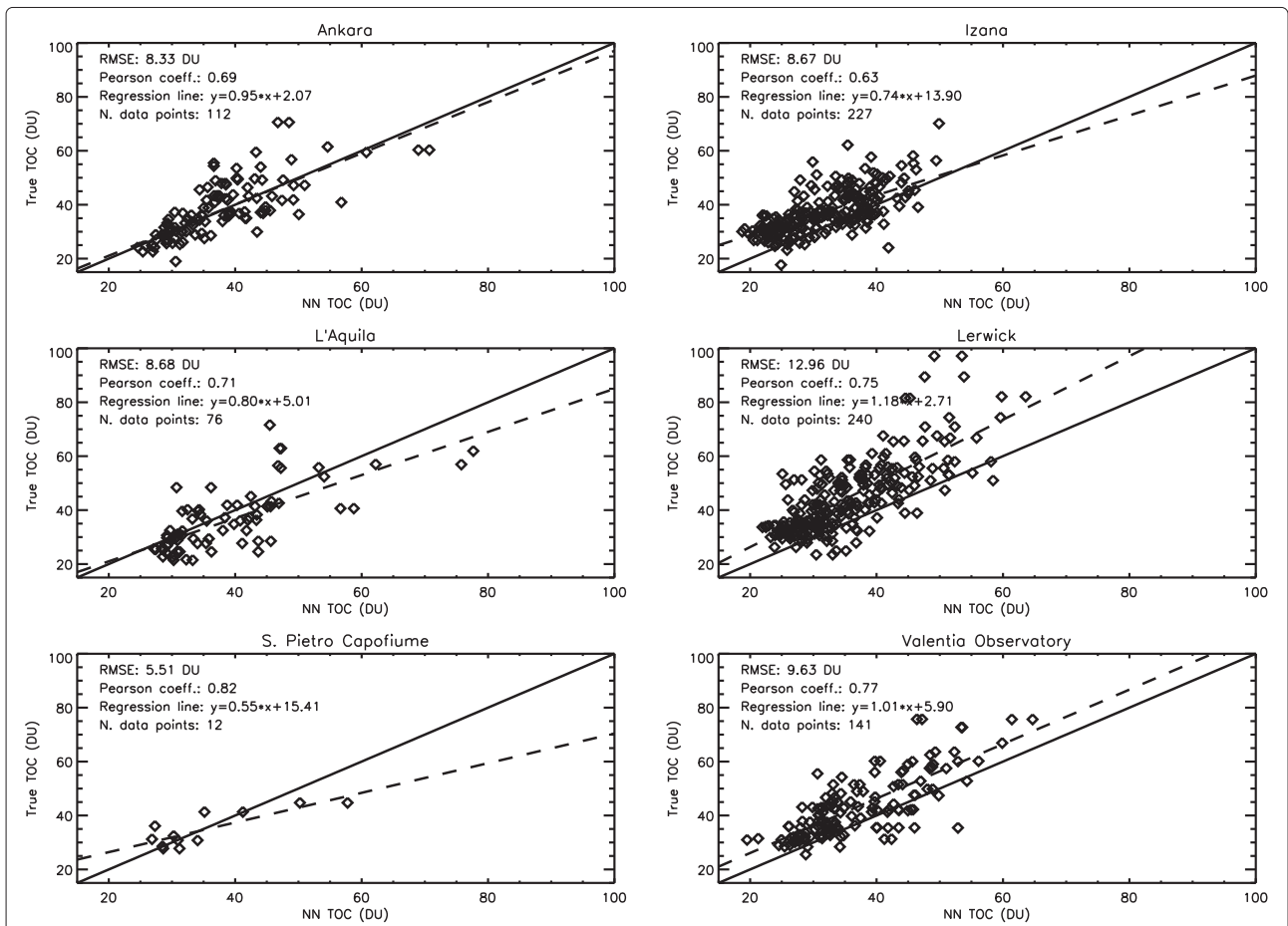
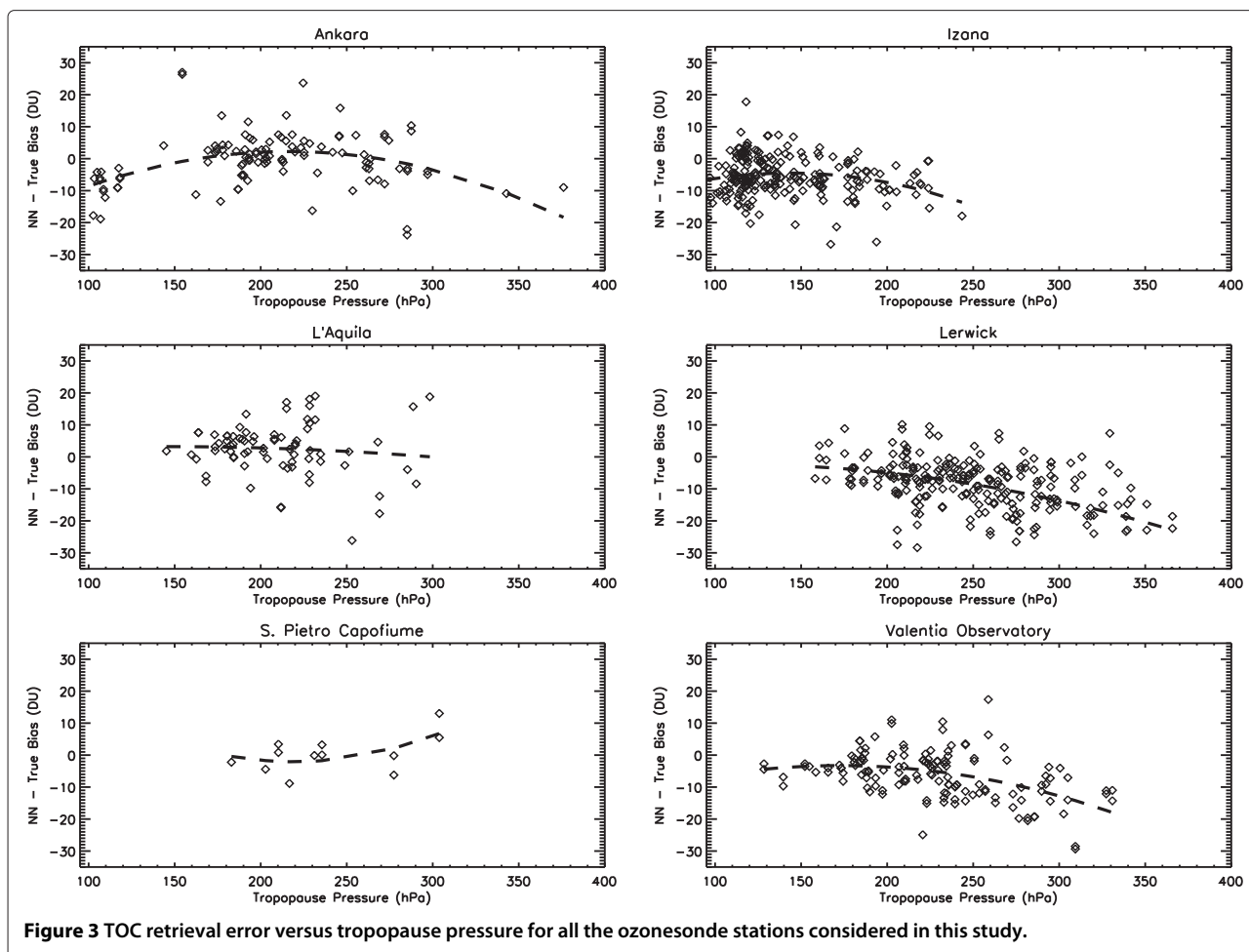


Figure 2 Scatter plots of true versus retrieved TOCs divided by ozonesonde station.



**Figure 3** TOC retrieval error versus tropopause pressure for all the ozonesonde stations considered in this study.

(ECC) sondes [32,33], interfaced with Vaisala RS-92 PTH (Pressure, Temperature, Humidity) radiosondes.

The ozone sounding activity at L'Aquila is performed within the framework of a collaboration between CETEMPS, L'Aquila University and the Italian Ministry for the Environment and Territory. The first soundings were performed in 1994. Since 2004, about two soundings per month have been regularly carried out on average. In the past, L'Aquila ozonesonde data were used in the validation of ozone profiles retrieved by the Michelson interferometer for passive atmospheric sounding (MIPAS), onboard Envisat [34,35].

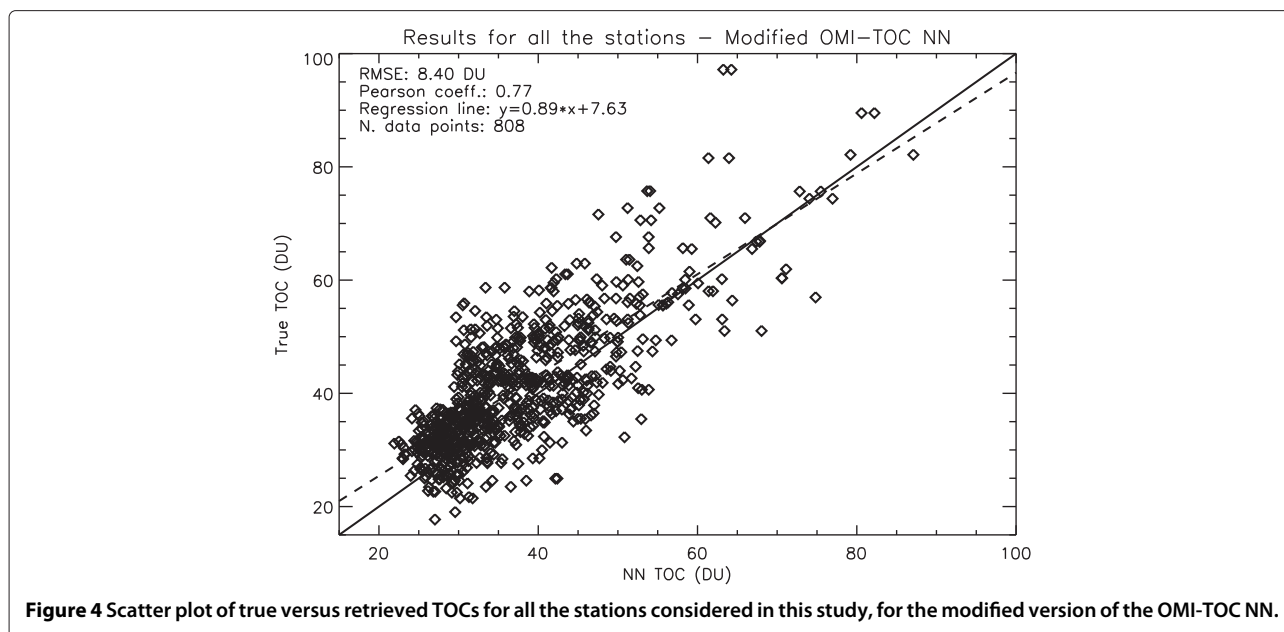
**Table 3** RMSE on training, test and validation sets for the OMI-TOC NN and its modified version

Dataset	TOC Std. [DU]	RMSE [DU]	
		OMI-TOC NN	Mod. OMI-TOC NN
Training	11.49	7.43	6.22
Test	11.23	8.03	7.20
Validation	11.83	10.21	8.40

The San Pietro Capofiume ozone soundings were performed under the responsibility of the Italian National Research Council (Consiglio Nazionale delle Ricerche (CNR)) Institute for Atmospheric and Climate Sciences (Istituto di Scienze dell'Atmosfera e del Clima (ISAC)). The San Pietro Capofiume ozonesondes are ENSCI-Z type ECC sondes, interfaced with Vaisala RS-80 PTH radiosondes.

In the past, ozone soundings were regularly performed at San Pietro Capofiume from 1991 to 1995 [36,37], and a specific campaign was organized in 1997 [37]. In 2004 and 2005, a sporadic sounding activity was carried out. However, it was subsequently interrupted due to scarcity of research funds. The data acquired during 2004 and 2005 were used in this study.

Within the above mentioned set of locations, different climatological characteristics are represented. This allows the geographical generalization capabilities of the OMI-TOC NN algorithm to be assessed, even at the upper and lower boundaries of the latitudinal range covered by the training set. Izaña is close to the African continent and not far from the Tropic of Cancer, and thus can



be regarded as an hybrid midlatitude/subtropical station, being influenced by air masses coming from both the mid-latitudes and the subtropics [38]. Lerwick and Valentia are characterized by an oceanic climate, and are subjected to advections of both midlatitude and polar air masses [39]. Hence, these stations can either behave as polar or midlatitude stations depending on the location of the polar front. Ankara, L'Aquila and San Pietro Capofiume can be regarded as typical midlatitude stations. Furthermore, all the stations are located in geographical areas that are not covered by the training set of the OMI-TOC NN algorithm. For this reason, validating the algorithm over this set of locations can give a reliable insight on its geographical generalization capabilities, as well as on its limitations.

In order to generate the validation set, the same co-location criteria as those used in the development of the OMI-TOC NN algorithm [26] were followed. Specifically, an ozone sounding and an OMI pixel were considered as co-located if two criteria were met: (i) the nominal coordinates of the ozonesonde station and those of the pixel center were no more than  $\pm 1^\circ$  apart; and (ii) no more than 12 hours had elapsed between the ozone sounding and the Aura overpass on the ozonesonde station.

By using these criteria, a total of 808 input-output pairs for validation were created. The number of co-locations obtained for each station is given in Table 2. An exiguous number of co-locations was obtained for San Pietro Capofiume. However, such data have been included in the present study for sake of completeness.

**Table 4** Error statistics for the OMI-TOC NN and its modified version for all the ozonesonde stations considered in this article

Station	N. data	Station statistics		OMI-TOC NN		Modified OMI-TOC NN	
		Mean TOC [DU]	Std. dev. [DU]	RMSE [DU]	Mean bias [DU]	RMSE [DU]	Mean bias [DU]
Ankara	112	38.04	11.54	8.33	-0.21	7.68	0.40
Izaña	227	37.38	8.31	8.67	-5.46	8.71	-5.97
L'Aquila	76	36.00	11.59	8.68	2.50	7.35	1.42
Lerwick	240	44.80	13.41	12.96	-9.28	8.60	-3.46
San Pietro Capofiume	12	34.82	6.55	5.51	0.35	3.91	1.17
Valentia	141	43.31	11.54	9.63	-6.18	8.88	-5.54
Observatory							
Overall	808	40.54	11.83	10.21	-5.17	8.40	-3.47

## 5 Validation results

The validation results from October 2004 to December 2008 are shown in the scatter plot in Figure 1. The retrieved TOCs are given as the abscissa, the true TOCs are given as the ordinate. A root mean square error (RMSE) of 10.21 DU was found. This value is definitely larger than that found in the validation results shown in [26], over a different set of ozonesonde stations. Furthermore, from a visual inspection of the scatter plot, it is evident that the algorithm has a systematic tendency to underestimate tropospheric ozone values larger than about 60 DU and overestimate values smaller than about 25 DU. Some quantitative statistics confirm this impression: 29 out of 33 TOCs larger than 25 DU are overestimated, and 42 out of 48 TOCs larger than 60 DU are underestimated. In order to assess whether this behavior displays a geographical dependence, the validation results were separately analyzed for each station.

The scatter plots of true versus retrieved TOCs for each ozonesonde station are shown in Figure 2. It can be noticed that, whilst the Ankara and L'Aquila scatterplots have a fairly symmetrical shape, the scatter plots for Izaña,

Lerwick and Valentia Observatory exhibit a quite pronounced underestimation tendency throughout the whole dynamical range of the TOC values.

One possible reason for the systematic underestimation of TOCs higher than 60 DU lies in the choice of 200 hPa as a static upper integration limit for the retrieved ozone column. In fact, if this TOC definition is used, extreme TOC values can be expected when the actual tropopause pressure exceeds 200 hPa (i.e., when the actual tropopause height is lower than the upper integration limit used in the OMI-TOC NN), because a large portion of stratospheric air—which is very rich in ozone—is included in the column over which the ozone profile is integrated in order to derive TOC. As a result, including the tropopause pressure in the input vector can help the NN discriminate such cases of enhanced TOC, and hence improve the overall retrieval accuracy.

In order to check the correctness of this hypothesis, an analysis of the retrieval error versus the actual tropopause pressure was carried out for each station. The tropopause pressure data were taken from the NCEP/NCAR Reanalysis 1 [40]. Plots representing the

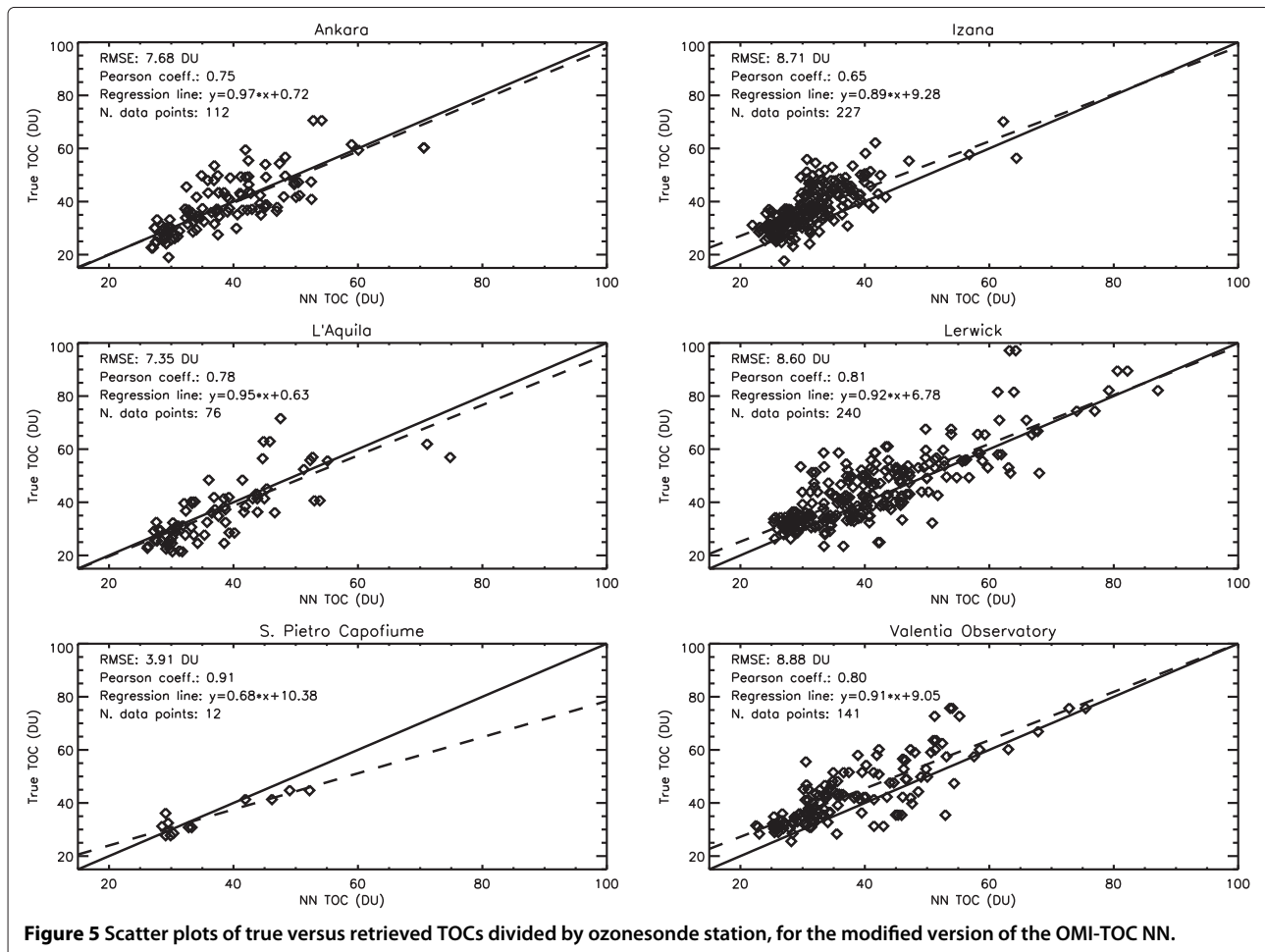


Figure 5 Scatter plots of true versus retrieved TOCs divided by ozonesonde station, for the modified version of the OMI-TOC NN.

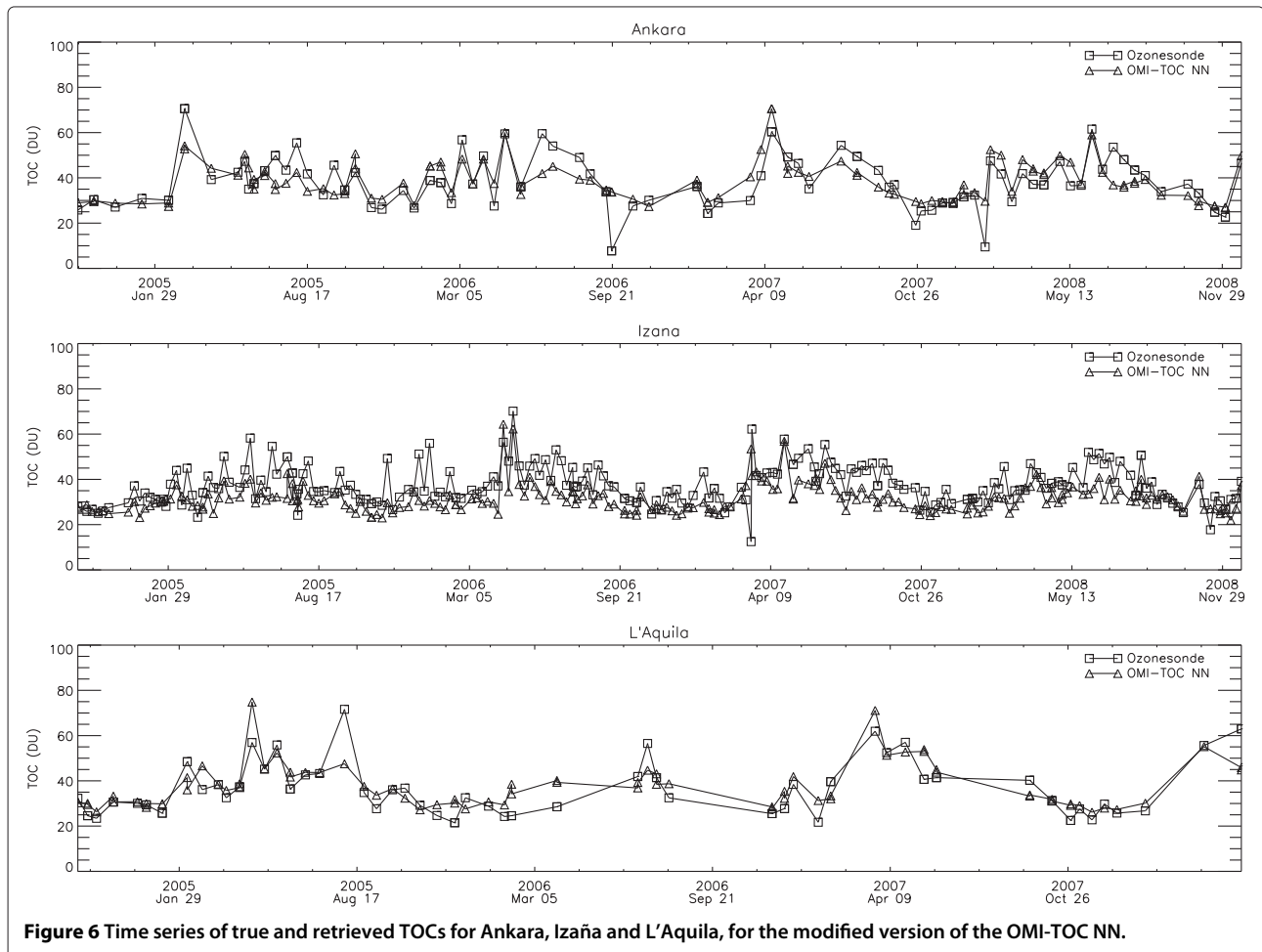
retrieval error against the tropopause pressure for each station are shown in Figure 3. A trend line, resulting from a quadratic fit of the retrieval error versus the tropopause pressure, is superimposed on each plot. It can be seen that the error trend is particularly clear on Lerwick and Valentia Observatory, where cases of tropopause pressures considerably greater than 200 hPa are most frequent.

### 6 Correction of tropopause related errors

The results shown in Section 5 confirm the hypothesis that a relationship between the retrieval accuracy of the OMI-TOC NN and the tropopause pressure exists. Furthermore, they suggest that the use of tropopause information as an input for the algorithm has the potential to enhance the retrieval accuracy. For this reason, a first attempt was made to design a new NN algorithm receiving such information as an input. The OMI Level 1B data were co-located with the NCEP/NCAR tropopause pressure fields in order to generate training, testing and validation sets for the new NN. The same stations used in

the OMI-TOC NN were used to train the new NN. A comparison between the two NNs in terms of training, test and validation RMSE is shown in Table 3. The standard deviations of the sonde TCOs in the three sets are also reported. It can be observed that the new NN has a lower RMSE with respect to the previous one on all the three sets.

The overall results for the set of ozonesonde stations considered in this article are shown in Figure 4. A significant reduction in the both RMSE and the bias is evident. Particularly significant is the reduction in the underestimation tendency for high values of TCO. Out of the 48 TCOs larger than 60 DU, 25 were found to be underestimated by the modified OMI-TOC NN, in contrast with the 42 underestimations found for the original NN (Section 5). In more formal terms, if  $TOC_{retr}$  is the retrieved TOC and  $TOC_{sonde}$  is the TOC measured by an ozonesonde, we can say that the conditional probability  $Prob(TOC_{retr} < 60 DU | TOC_{sonde} \geq 60 DU)$  on the validation dataset can be estimated in about 88% for the original OMI-TOC NN described in [26] and about 52% for the modified NN proposed in this article.



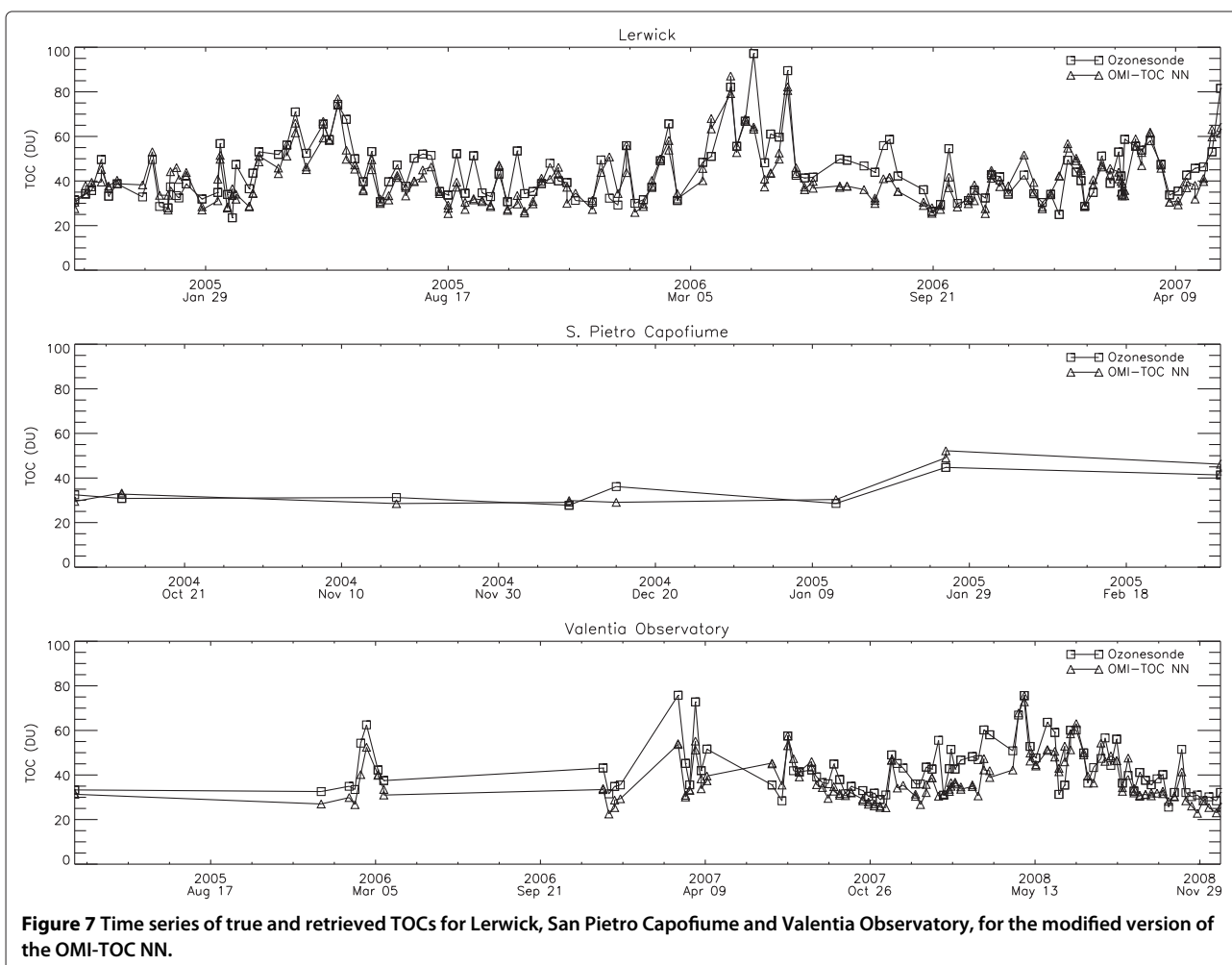
**Figure 6** Time series of true and retrieved TOCs for Ankara, Izaña and L'Aquila, for the modified version of the OMI-TOC NN.

Table 4 summarizes the performances of both the OMI-TOC NN and its modified version in terms of RMSE and mean bias. Mean and standard deviation of the TOCs measured by the ozonesondes are also reported, in order to facilitate the interpretation of the validation results. The results divided by station are also shown, in form of scatter plots, in Figure 5.

The improvements are evident on Ankara and L'Aquila, and dramatic on Lerwick and Valentia Observatory. Such improvements were not found on Izaña, which still appears to be the most problematic station amongst those shown in this article. From a visual inspection of the upper right panel of Figure 3, it is evident that the tropopause pressures over Izaña were most often far below 150 hPa (i.e., the tropopause was considerably higher than the corresponding altitude level) with regard to the ozone soundings used in this validation exercise. This suggests that Izaña mostly behaved as a tropical station, and thus portends poor performances of the OMI-TOC NN with air masses of tropical origin. This behavior appears

reasonable, because the OMI-TOC NN was trained using only midlatitude data. Anyway, further investigations are ongoing in order to interpret this result.

In Figures 6 and 7, time series of true and retrieved TOC over the six stations considered in this article are shown for the modified OMI-TOC NN. Apart from the above mentioned case of Izaña, where a strong negative bias of the NN versus the ozonesonde data exists, a slight underestimation tendency can be observed over all the six stations considered in this article. Such tendency appears to be strongest during the summer months, as evident from the results on Ankara (Figure 6, above panel) and L'Aquila (Figure 6, below panel). Specifically, it appears that the OMI-TOC NN is not able to reproduce situations of enhanced TOCs that occur during the summer. It is still not clear whether this fact is caused by a lack of sensitivity of the algorithm to the lowest atmospheric layers. Appropriate actions, aimed at reducing this effect, should be taken in the development of further versions of the OMI-TOC NN algorithm.



## 7 Conclusions

In this article, the results of a validation of a NN algorithm for tropospheric ozone column retrieval from OMI data—named the OMI-TOC NN—are shown. The validation was performed over six ozonesonde stations distributed across the European continent. This validation set is considered as a benchmark for the retrieval performances of the algorithm, as it represents a number of climatological situations that can be encountered over Europe.

A good agreement over Ankara, L'Aquila and San Pietro Capofiume—the most central stations in terms of latitude—was found. However, strong negative biases are present over Lerwick, Valentia Observatory and Izaña, especially in conditions of high TOC values. In order to investigate the reasons for this problem, the retrieval bias of the OMI-TOC NN algorithm was analyzed as a function of the tropopause pressure values taken from the NCEP/NCAR Reanalysis 1. A significant correlation between tropopause pressure and retrieval error was found. As a consequence, a new version of the OMI-TOC NN, having the NCEP/NCAR tropopause pressures in its input vector, was designed, and its results were evaluated over the same validation set.

The modified OMI-TOC NN algorithm exhibited a considerably improved retrieval accuracy, in terms of RMSE, over the whole validation set. The improvements were found to be most significant on the northernmost stations of Lerwick and Valentia Observatory, where cases of low tropopauses (i.e., high tropopause pressures) are most frequent. However, no improvements were observed on Izaña, where tropopause pressures larger than 200 hPa are quite unlikely. The results of the modified OMI-TOC NN on Izaña also suggest that using the tropopause pressure as an input for the algorithm is still not sufficient to improve the retrieval accuracy in cases of high tropopauses. In the future, this issue will be addressed by including tropical ozonesonde stations in the training set.

A major point that might be raised on the basis of these results is that using 200 hPa as upper integration limit in the TOC definition is not a sensible choice in order to characterize the tropospheric ozone column. Further versions of the OMI-TOC NN algorithm should provide estimates of the ozone column up to the actual tropopause, whether it be defined based on the NCEP/NCAR Reanalysis or by other means (e.g., dynamical tropopause).

### Competing interests

The authors declare that they have no competing interests.

### Acknowledgements

This study had been carried out in the framework of the ESA Project C1P.2930. All the Pls of the ozonesonde stations whose data are used in this study are gratefully acknowledged. The ozone monitoring activities of CETEMPS are partly funded by the Italian Ministero dell'Ambiente e della Tutela del Territorio e del Mare. The data for Ankara, Lerwick and Valentia Observatory ozonesonde stations were retrieved from the World Ultraviolet Data Center (WOUDC) archive (<http://www.woudc.org>). The data for the Izaña ozonesonde

station were obtained as part of the Network for the Detection of Atmospheric Composition Change (NDACC) and are publicly available (see <http://www.ndacc.org>). The NCEP Reanalysis data are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at <http://www.esrl.noaa.gov/psd/>.

### Author details

<sup>1</sup>Earth Observation Laboratory, Department of Civil and Computer Engineering, Tor Vergata University, Via del Politecnico 1, 00133 Rome, Italy. <sup>2</sup>Laboratoire Inter-universitaire des Systèmes Atmosphériques, UMR7583, CNRS—Universités Paris-Est et Paris Diderot, 61 Avenue du Général de Gaulle, 94010 Créteil, France. <sup>3</sup>Istituto di Scienze dell'Atmosfera e del Clima, Consiglio Nazionale delle Ricerche, via Gobetti 101, 40129, Bologna, Italy. <sup>4</sup>CETEMPS, Department of Physics, University of L'Aquila, Via Vetoio 1, 67100, Coppito-L'Aquila, Italy. <sup>5</sup>Now at: SRON Netherlands Institute for Space Research, Sorbonnelaan 2, 3584 CA Utrecht, The Netherlands.

Received: 15 July 2011 Accepted: 18 January 2013

Published: 19 February 2013

### References

1. IPCC: Summary for policymakers, in *Climate Change 2007: the Physical Science Basis*, vol. 2, ed. by S Solomon, D Qin, M Manning, Z Chen, M Marquis, KB Averyt, M Tignor, and HL Miller (Cambridge University Press, Cambridge, 2007), pp. 53–57
2. DJ Jacob, *Introduction to Atmospheric Chemistry*. (Princeton University Press, New Jersey, 1999)
3. SC Anenberg, JJ West, AM Fiore, DA Jaffe, MJ Prather, D Bergmann, K Cuvelier, FJ Dentener, BN Duncan, M Gauss, P Hess, JE Jonson, A Lupu, IA Mackenzie, E Marmer, RJ Park, MG Sanderson, M Schultz, DT Shindell, S Szopa, M García Vivanco, O Wild, G Zeng, Intercontinental impacts of ozone pollution on human mortality. *Environ. Sci. Technol.* **43**, 6482–6487 (2009). doi:10.1021/es900518z
4. F Booker, R Munifering, M McGrath, K Burkey, D Decoteau, E Fiscus, W Manning, S Krupa, A Chappelka, D Grantz, The ozone component of global change: Potential effects on agricultural and horticultural plant yield, product quality and interactions with invasive species. *J. Int. Plant Biol.* **51**(4), 337–351 (2009). doi:10.1111/j.1744-7909.2008.00805.x
5. JK Creilson, J Fishman, AE Wozniak, Intercontinental transport of tropospheric ozone: a study of its seasonal variability across the North Atlantic utilizing tropospheric ozone residuals and its relationship to the North Atlantic Oscillation. *Atmos. Chem. Phys.* **3**, 2053–2066. doi:10.5194/acp-3-2053-2003
6. JK Creilson, J Fishman, AE Wozniak, Arctic Oscillation-induced variability in satellite-derived tropospheric ozone. *Geophys. Res. Lett.* **32**, L14822 (2005). doi:10.1029/2005GL023016
7. JH Kim, MJ Newchurch, Climatology and trends of tropospheric ozone over the eastern Pacific Ocean: The influences of biomass burning and tropospheric dynamics. *Geophys. Res. Lett.* **23**(25), 3723–3726 (1996). doi:10.1029/96GL03615
8. P Valks, RBA Koelemeijer, M van Weele, P van Velthoven, JPF Fortuin, H Kelder, Variability in tropical tropospheric ozone: Analysis with Global Ozone Monitoring Experiment observations and a global model. *J. Geophys. Res.* **108**(D11), 4328 (2003). doi:10.1029/2002JD002894
9. J Fishman, JK Creilson, AE Wozniak, PJ Crutzen, Interannual variability of stratospheric and tropospheric ozone determined from satellite measurements. *J. Geophys. Res.* **110**, D20306 (2005). doi:10.1029/2005JD005868
10. JP Burrows, U Platt, P Borrell, in *The Remote Sensing of Tropospheric Composition from Space*, ed. by JP Burrows, U Platt, and P Borrell. Tropospheric Remote Sensing from Space (Springer-Verlag, Berlin, Germany, 2011), pp. 1–65
11. J Fishman, CE Watson, JC Larsen, JA Logan, Distribution of tropospheric ozone determined from satellite data. *J. Geophys. Res.* **95**, 3599–3617 (1990). doi:10.1029/JD095iD04p03599
12. MR Schoeberl, JR Ziemke, B Bojkov, N Livesey, B Duncan, S Strahan, L Froidevaux, S Kulawik, PK Bhartia, S Chandra, PF Levelt, JC Witte, AM Thompson, E Cuevas, A Redondas, DW Tarasick, J Davies, G Bodeker, G Hansen, BJ Johnson, SJ Oltmans, H Vömel, M Allaart, H Kelder, M Newchurch, S Godin-Beekmann, G Ancellet, H Claude, SB Andersen, E Kyrö, M Parrondos, M Yela, G Zablocki, D Moore, H Dier, P von der Gathen,

- P Viatte, R Stübi, B Calpini, P Skrivanekova, V Dorokhov, H de Backer, FJ Schmidlin, G Coetzee, M Fujiwara, V Thouret, F Posny, G Morris, J Merrill, CP Leong, G Koenig-Langlo, E Joseph, *A trajectory-based estimate of the tropospheric ozone column using the residual method*, vol. 112, (2007), p. D24549
13. JR Ziemke, S Chandra, PK Bhartia, Two new methods for deriving tropospheric column ozone from TOMS measurements: The assimilated UARS MLS/HALOE and convective-cloud differential techniques. *J. Geophys. Res.* **103**, 22115–22127 (1998). doi:10.1029/98JD01567
  14. JR Ziemke, S Chandra, PK Bhartia, "Cloud slicing": a new technique to derive upper tropospheric ozone from satellite measurements. *J. Geophys. Res.* **106**, 9853–9867 (2001). doi:10.1029/2000JD900768
  15. R Munro, R Siddans, WJ Reburn, BJ Kerridge, Direct measurement of tropospheric ozone distributions from space. *Nature.* **392**, 168–171 (1998). doi:10.1038/32392
  16. X Liu, K Chance, RJD Spurr, TP Kurosu, RV Martin, MJ Newchurch, Ozone profile and tropospheric ozone retrievals from the Global Ozone Monitoring Experiment: Algorithm description and validation. *J. Geophys. Res.* **110**, D20307 (2005). doi:10.1029/2005JD006240
  17. X Liu, K Chance, CE Sioris, TP Kurosu, RJD Spurr, RV Martin, T Fu, JA Logan, DJ Jacob, PI Palmer, MJ Newchurch, IA Megretskaja, RB Chatfield, First directly retrieved global distribution of tropospheric column ozone from GOME: Comparison with the GEOS-CHEM model. *J. Geophys. Res.* **111**, D20308 (2006). doi:10.1029/2006JD007374
  18. X Liu, PK Bhartia, K Chance, RJD Spurr, TP Kurosu, Ozone profile retrievals from the Ozone Monitoring Instrument. *Atmos. Chem. Phys.* **10**, 2521–2537 (2010). doi:10.5194/acp-10-2521-2010
  19. CD Rodgers, *Inverse Methods for Atmospheric Sounding: Theory and Practice*. (World Scientific Publishing Company, London, UK, 2000)
  20. CM Bishop, *Neural Networks for Pattern Recognition*. (Oxford University Press, New York, USA, 1995)
  21. W Blackwell, FW Chen, *Neural Networks in Atmospheric Remote Sensing*. (Artech House, Norwood, MA, USA, 2009)
  22. F Del Frate, A Ortenzi, S Casadio, C Zehner, Application of neural algorithms for a real-time estimation of ozone profiles from GOME measurements. *IEEE Trans. Geosci. Rem. Sens.* **40**, 2263–2270 (2002). doi:10.1109/TGRS.2002.803622
  23. MD Müller, AK Kaifel, M Weber, S Tellmann, JP Burrows, D Loyola, 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 (2003). doi:10.1029/2002JD002784
  24. MD Müller, A Kaifel, M Weber, JP Burrows, Neural network scheme for the retrieval of total ozone from Global Ozone Monitoring Experiment data. *Appl. Opt.* **41**(24), 5051–5058 (2002). doi:10.1364/AO.41.005051
  25. P Sellitto, A Burini, F Del Frate, S Casadio, in *Proceedings of International Geosciences and Remote Sensing Symposium*. Dedicated neural network algorithms for direct estimation of tropospheric ozone from satellite measurements, (Barcelona, 2007), pp. 1685–1688
  26. P Sellitto, B Bojkov, X Liu, K Chance, F Del Frate, 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 (2011). doi:10.5194/amt-4-2375-2011
  27. P Sellitto, D Solimini, F Del Frate, S Casadio, Tropospheric ozone column retrieval from ESA-Envisat SCIAMACHY nadir UV/VIS measurements by means of a neural network algorithm. *IEEE Trans. Geosci. Rem. Sens.* **50**(3), 998–1011 (2012). doi:10.1109/TGRS.2011.2163198
  28. MR Schoeberl, AR Douglass, E Hilsenrath, PK Bhartia, R Beer, JW Waters, MR Gunson, L Froidevaux, JC Gille, JJ Barnett, PF Levelt, P DeCola, Overview of the EOS Aura mission. *IEEE Trans. Geosci. Rem. Sens.* **44**, 1066–1074 (2006). doi:10.1109/TGRS.2005.861950
  29. PF Levelt, GHJ van den Oord, MR Dobber, A Mälkki, H Visser, J de Vries, P Stammes, JOV Lundell, H Saari, The Ozone Monitoring Instrument. *IEEE Trans. Geosci. Rem. Sens.* **44**, 1093–1101 (2006). doi:10.1109/TGRS.2006.872333
  30. Background information about the Row Anomaly in OMI. [http://www.knmi.nl/omi/research/product/rowanomaly-background.php]
  31. F Del Frate, M Iapaolo, S Casadio, Neural Networks for the dimensionality reduction of GOME measurement vector in the estimation of ozone profiles. *J. Quant. Spectrosc. Radiat. Transfer.* **92**, 275–291 (2005). doi:10.1016/j.jqsrt.2004.07.028
  32. WD Komhyr, Electrochemical concentration cells for gas analysis. *Ann. Geophys.* **25**, 167–183 (1969)
  33. BJ Johnson, JH Vömel, SJ Oltmans, HGJ Smit, T Deshler, C Kroger, Electrochemical concentration cell (ECC) ozonesonde pump efficiency measurements and tests on the sensitivity to ozone of buffered and unbuffered ECC sensor cathode solutions. *J. Geophys. Res.* **107**(D19), 4393 (2002). doi:10.1029/2001JD000557
  34. V Rizi, V Iarlori, G Rocci, G Redaelli, G Visconti, in *Proceedings of Envisat Validation Workshop*, vol. 531. The setup and the performances of L'Aquila Raman lidar and standard PTU and PTO<sub>3</sub> balloon soundings for Envisat validation, (Frascati, Italy, 9–13 December 2002 ESA SP-531, 2003), pp. 45–1
  35. U Cortesi, JC Lambert, C De Clercq, G Bianchini, T Blumenstock, A Bracher, E Castelli, V Catoire, KV Chance, M De Mazière, P Demoulin, S Godin-Beekman, N Jones, N Jucks, C Keim, T Kerzenmacher, H Kuellmann, J Kuttippurath, M Iarlori, GY Liu, Y Liu, IS McDermid, YS Meijer, F Ravegnani, S Mikuteit, H Oelhaf, C Piccolo, M Pirre, P Raspolini, F Ravegnani, WJ Reburn, G Redaelli, JJ Remedios, H Sembhi, D Smale, T Steck, A Taddei, C Vartzos, C Vigouroux, A Waterfall, G Wetzel, S Wood, Geophysical validation of MIPAS-ENVISAT operational ozone data. *Atmos. Chem. Phys.* **7**, 4807–4867 (2007). doi:10.5194/acp-7-4807-2007
  36. M Banzi, C Carbonara, M Cervino, Ozone Observations at San Pietro Capofiume, Italy: preliminary results. *Geophys. Res. Lett.* **21**, 2231–2234 (1994). doi:10.1029/94GL01460
  37. G Giovanelli, G Lenzi, A Petritoli, F Ravegnani, M Cervino, D Bortoli, I Kostadinov, G Trivellone, in *Proceedings of XX Quadrennial Ozone Symposium*, vol. II, ed. by CS Zerefos. Study of vertical and temporal variability of ozone over Italy, by means of ozone-sounding activity at the S. Pietro Capofiume station (University of Athens, Kos, Greece, 2004), pp. 339–340
  38. SJ Oltmans, AS Lefohn, JM Harris, I Galbally, HE Scheel, G Bodeker, E Brunke, H Claude, D Tarasick, BJ Johnson, P Simmonds, D Shadwick, K Anlauf, K Hayden, F Schmidlin, T Fujimoto, K Akagi, C Meyer, S Nichol, J Davies, A Redondas, E Cuevas, Long-term changes in tropospheric ozone. *Atmos. Environ.* **40**, 3156–3173 (2006). doi:10.1016/j.atmosenv.2006.01.029
  39. A Mariotti, B Moustaooui, B Legras, H Teitelbaum, Comparison between vertical ozone soundings and reconstructed potential vorticity maps by contour advection with surgery. *J. Geophys. Res.* **102**(D5), 6131–6142 (1997). doi:10.1029/96JD03509
  40. E Kalnay, M Kanamitsu, R Kistler, W Collins, D Deaven, L Gandin, M Iredell, S Saha, G White, J Woollen, Y Zhu, M Chelliah, M Ebisuzaki, W Higgins, J Janowiak, KC Mo, C Ropelewski, J Wang, A Leetmaa, R Reynolds, R Jenne, D Joseph, The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteor. Soc.* **77**, 437–470 (1996). doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

doi:10.1186/1687-6180-2013-21

**Cite this article as:** Di Noia et al.: Tropospheric ozone column retrieval from OMI data by means of neural networks: a validation exercise with ozone soundings over Europe. *EURASIP Journal on Advances in Signal Processing* 2013 **2013**:21.

**Submit your manuscript to a SpringerOpen® journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)