# ON THE ROLE OF VIS RADIATION FOR THE OZONE INFORMATION RETRIEVAL FROM SCIAMACHY DATA BY MEANS OF NEURAL NETWORK ALGORITHMS

**Pasquale Sellitto, Antonio Di Noia, Fabio Del Frate, and Domenico Solimini**

*Tor Vergata University, Via del Politecnico 1, 00133 Rome, Italy*

## ABSTRACT

In this paper we discuss the design, implementation and performance of two neural network algorithms, one for ozone concentration profile, and the second for tropospheric ozone column retrieval from ESA-Envisat SCIAMACHY Level 1b data. The performances of the two algorithms were checked through tests against ozone sondes measurements. Both algorithms make use of visible radiation whose role has been investigated, by alternatively considering the overall UV/VIS information or only the UV band. In both cases, the algorithms operating with UV/VIS radiance spectra exhibited better performances than those using only UV spectra, especially in troposphere. These results suggest the possibility of improving the accuracy of ozone profile and tropospheric ozone column retrievals by including VIS radiances in the input vector of suitably designed neural network algorithms.

Key words: Ozone, SCIAMACHY, Neural Networks, VIS radiation.

## 1. INTRODUCTION

The systematic and continuous observation of ozone concentration levels from satellite observations is crucial to understand and monitor processes that can play a fundamental role for the life on Earth. Nadir UV/VIS spectra from dedicated sensors, as, e.g., SCIAMACHY, OMI or GOME-2, can provide useful height-resolved information about ozone. The most interesting spectral intervals for ozone remote sensing are the ultraviolet Hartley and Huggins bands and the visible Chappuis band. Inversion schemes based on Optimal Estimation (OE) usually do not consider the latter spectral interval, given the enhanced interaction of VIS radiation with aerosols, clouds and surface. Indeed, the uncertainties in the knowledge of the parameters characterizing such interfering effects can yield large errors in radiative transfer modeling.

Neural Network (NN) algorithms represent a useful alternative approach to OE for inversion problems in remote sensing. The effectiveness of NNs in retrieving ozone concentrations from satellite data has been shown in recent studies [1, 2, 3]. Moreover, NNs determine the input-output relationship directly from the data, hence an explicit modeling of all radiative interactions can be avoided with this approach.

The aim of this paper is to demonstrate that the use of VIS wavelengths in a NN algorithm can significantly improve the accuracy of ozone retrievals. To achieve this goal two different algorithms were designed, the first trained to retrieve ozone concentration profiles (see Sect. 3), and the second to directly retrieve Tropospheric Ozone Columns (TOCs) (see Sect. 4), both from nadir SCIAMACHY Level 1b data. Satellite data were matched with ozone sondes measurements to obtain input-output pairs to train and test the nets. Comparisons between two neural architectures, one using only UV wavelengths and one using both UV and VIS wavelengths, were made for the two algorithms. The design stage of the NNs, including the preparation of training/test datasets and input selection by means of an Extended Pruning (EP) technique, is also discussed.

## 2. NEURAL NETWORK ALGORITHMS FOR OZONE RETRIEVAL FROM SCIAMACHY DATA: GENERAL DESCRIPTION OF THE METHODOLOGY

The NN algorithms used by our research group for the retrieval of ozone profiles and tropospheric ozone column share a common design framework. In the present section the characteristics of such general methodology will be summarized.

### 2.1. Training dataset

The input-output pairs forming the training datasets for the NNs consist of normalized nadir SCIAMACHY radiance spectra, acquired in UV for one network and in UV/VIS for the other network, and co-located ozone profiles or TOCs measured by ozone sondes. Ozone sonde data were obtained from WOUDC [4] and SHADOZ datasets [5]. A map displaying the geographical distribution of the ozone sonde stations we used is reported in

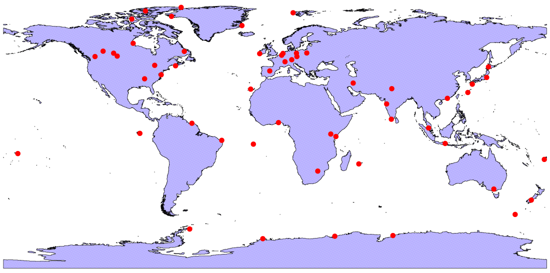


*Figure 1. Geographical distribution of the ozone sonde stations used to generate our dataset.*

Fig. 1. The following co-location criteria were applied: SCIAMACHY spectra and WOUDC/SHADOZ profiles were considered as co-located if acquired in locations separated by no more than 2.5° in latitude and 5° in longitude (referring to center pixel coordinates for the SCIAMACHY measurements), and with a temporal delay not higher than 12 h. The data were acquired during 2003 and 2004. The input vectors for the networks consist of SCIAMACHY normalized radiance spectra, Solar Zenith Angles (SZAs) and ozone slant columns, this latter obtained applying the TIDAS algorithm [6] to the input radiance spectra. The NNs were trained with uncalibrated data, in order to reduce the computational burden of the calibration procedure. We point out that use of uncalibrated data is made possible by the intrinsic robustness of NN algorithms to calibration errors, provided these are accounted for in the training. The output vectors were obtained from ozone sondes concentration profiles, by a) re-sampling the ozone concentrations at specific altitudes, for the profile NN (see section 3), and b) integrating the ozone concentrations from surface to P=200.0 hPa, for the TOC NN (see section 4); for more details please refer to next sections.

The resulting training dataset consists of 10050 examples. The same examples were used to train both the NN algorithms. The following partition of the training dataset was adopted: 5000 pairs were used for automatic wavelength selection, 3500 were used for the training phase and 1550 were used for the test phase.

### 2.2. Automatic wavelength selection

Since SCIAMACHY measures radiance at a large number of wavelengths, a dimensionality reduction technique is needed to select the most informative wavelengths to use in the retrieval, minimizing the ill-conditioning of the inversion problem. This end was achieved by applying an EP procedure [1]. EP can be summarized as follows: 1) a net with a maximum input dimensionality is trained; 2) the synapse with the lowest magnitude is pruned; 3) the net is retrained without re-initialization and with a lower number of training cycles; 4) the training error $E$ is evaluated: if $E$ is larger than a threshold value $E_{\mathrm{thr}}$ then the pruning in stopped, otherwise the algorithm restarts from point 2).

In designing the NN algorithms, the following parameters for the EP procedure were selected: first training with 1000 cycles, subsequent trainings with 100 cycles, error threshold $E_{\mathrm{thr}}$ equal to the 10% of the error obtained at the end of the previous re-training phase.

### 2.3. Topology selection

Since no deterministic criteria are available to select the number of hidden layers and the number of neurons per hidden layer, empirical criteria were adopted to select the topology of the nets. Specifically, only networks with a single hidden layer were considered, and the number of hidden neurons was selected repeatedly training the neural networks with different cardinalities of the hidden layer, and selecting the cardinality which yielded the minimum test error.

## 3. USE OF VIS WAVELENGTHS IN NEURAL NETWORKS FOR OZONE PROFILE RETRIEVAL

To assess the potential of VIS wavelengths in increasing the accuracy of ozone profile retrievals by NN algorithms, two NNs were trained: one having as inputs nadir SCIAMACHY radiance spectra measured only in the UV intervals and one having as inputs UV/VIS radiance spectra. A comparison between the performances of the two algorithms was carried out by comparing the retrieval results against the corresponding ozone sondes measurements. In this section the design stage of the two algorithms will be presented, and the experimental results will be discussed.

### 3.1. Design of the neural networks

The first problem addressed during the design of NNs for ozone profile retrievals was the choice of the number of output neurons, *i.e.* the number of heights at which ozone retrieval is carried out. This selection was based on the experience gained with analyses of GOME data made at the Rutherford Appleton Laboratory [7]. Such analyses showed that the maximum vertical resolution of GOME ozone profile retrievals approximately amounts to 4 km in stratosphere and 7 km in troposphere. According to these results, 9 heights have been selected for ozone concentration retrieval: 2 tropospheric and 7 stratospheric levels. Specifically, the pressure levels selected are the following (in hPa): 1000.00, 421.70, 177.83, 100.00, 56.23, 31.62, 17.78, 10.00 and 5.62. Ozone sondes profiles were re-sampled accordingly.

Once the number of output neurons was selected, an automatic wavelength selection through EP, as indicated in Section 2, was made. The EP was applied to an initial network having an input layer with 224 units, an hidden layer with 80 neurons and an output layer with 9 neurons. The input vectors consisted of SCIAMACHY normalized radiance spectra measured at 222 wavelengths, SZAs and ozone slant columns. The EP led to the selection of 26 wavelengths. Specifically, 20 UV wavelengths and 6 VIS wavelengths were selected. 19 of the 20 selected UV wavelengths lie in Huggins band (310-350 nm), and only one lies in Hartley band (280-310 nm). A quite innovative result yielded by EP was the selection of 6 wavelengths in the visible range. This outcome suggests that the VIS measurements can carry information instrumental to the retrieval of ozone profiles.

Using the empirical procedure described in Section 2, an optimal value of 27 hidden neurons was determined.

### 3.2. Experimental results

To assess the retrieval enhancement contributed by VIS wavelength, a comparison was carried out between the accuracy yielded by the network described above and the one of an optimized network using only UV wavelengths. The optimal UV network has the same topology as the UV/VIS one, and used 26 wavelengths selected by the same EP procedure.

The ozone profiles retrieved by the two NNs were checked against an independent set of ozone profiles measured by ozone sondes. The Pearson correlation coefficient between retrieved and true (i.e., measured by ozone sondes) ozone concentration at the 9 retrieval heights was computed. The profile of the Pearson coefficient reported in Fig. 2 shows that the performances of the algorithm using UV/VIS wavelengths are significantly better than those of the UV only algorithm in troposphere, and slightly better in stratosphere. The results are highlighted in Figs. 3 and 4, which show examples of the scatter plots of retrieved vs. true ozone concentration at a tropospheric and stratospheric level, respectively. We add that the reasons of the sudden lowering of the Pearson coefficient observed at 31.62 and 5.62 hPa for both algorithms are unknown.

The significant improvement of retrieval accuracy obtained at the tropospheric levels by using VIS wavelengths can be partially explained considering that VIS radiation is more penetrating in atmosphere than the UV radiation, and thus it may carry more information on the ozone concentration at the lowest levels.

## 4. USE OF VIS WAVELENGTHS IN NEURAL NETWORKS FOR TROPOSPHERIC OZONE COLUMN RETRIEVAL

Following the same methodology described in Sect. 3, the influence of VIS wavelengths on the accuracy of TOC retrieval was investigated. Specifically, two NNs were

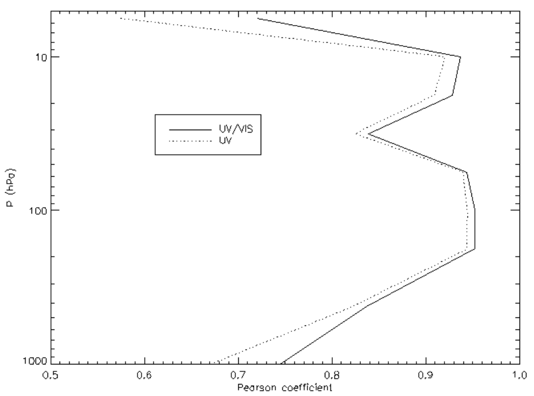



*Figure 2. Height dependence of the Pearson correlation coefficient between retrieved and true ozone concentrations, with and without the VIS.*

trained: one having as inputs nadir SCIAMACHY radiance spectra measured only in the UV intervals and one having as inputs UV/VIS radiance spectra. A comparison between the performances of the two algorithms was carried out by comparing the retrieval results against the corresponding ozone sondes measurements. In this section the design of the two neural networks and the results of TOC retrievals will be presented.

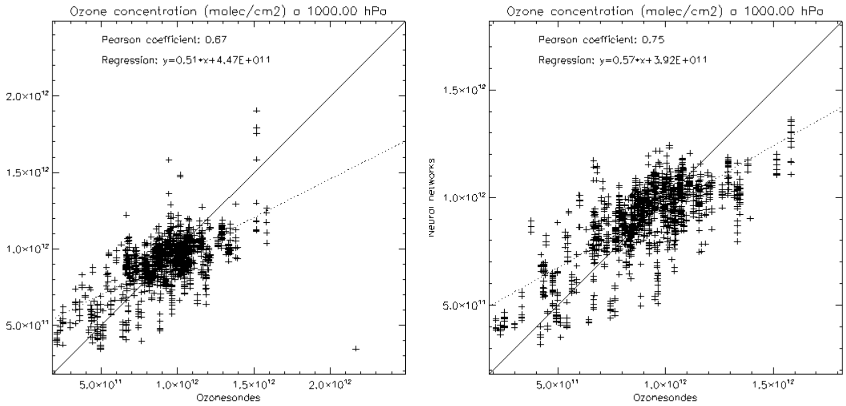
### 4.1. Design of the neural networks

Since in this case a single parameter is to be retrieved, the number of output neurons is equal to 1. Thus, the first design stage for the TOC retrieval algorithms was the automatic wavelength selection through EP. The EP led to the selection of 28 wavelengths. For the UV/VIS NN, 17 UV and 11 VIS wavelengths were selected. The resulting networks had 30 input units, due to the inclusion of SZA and ozone slant column in the input vector.

The optimization of the hidden layer cardinality was made according to the empirical procedure described in Section 2. Optimal values of 27 and 28 hidden units were found for the UV and the UV/VIS networks, respectively.

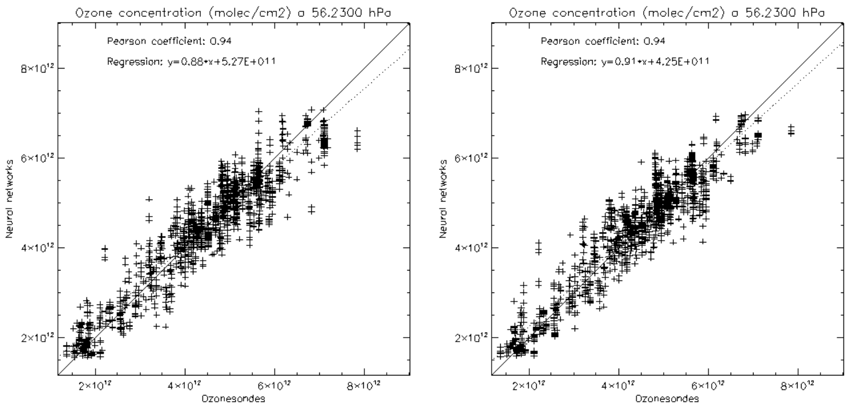
### 4.2. Experimental results

The TOC retrievals performed by the two NNs were checked against an independent set of TOCs derived by ozone sonde. The Pearson correlation coefficient between retrieved and true (i.e. obtained from ozone sonde profiles) TOCs was computed. In Fig. 5 two scatter plots of retrieved versus true TOC, one for each net, are shown. A significant increase (over the 10%) of the Pearson correlation coefficient is apparent in the network including VIS radiances in its input vector (Fig. 5(b)). This result confirms those found for ozone profiles, discussed in Sect. 3. Once again physically reasonable results have



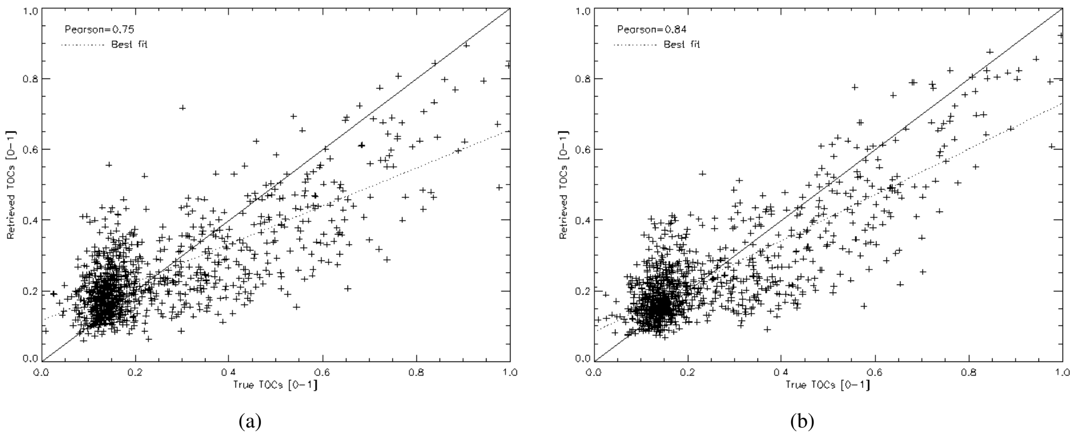

(a) (b)

*Figure 3. Scatter plot of retrieved versus true ozone concentration at* 1000.0 hPa *for the UV-based algorithm (a) and the UV/VIS-based algorithm (b).*




(a) (b)

*Figure 4. Scatter plot of retrieved versus true ozone concentration at* 56.23 hPa *for the UV-based algorithm (a) and the UV/VIS-based algorithm (b).*




(a) (b)

*Figure 5. Scatter plot of retrieved versus true tropospheric ozone column for the UV-based algorithm (a) and the UV/VIS-based algorithm (b).*

been reached. Indeed, VIS radiation does carry non-negligible information on tropospheric ozone, but suffers from interfering effects. Retrieval algorithms like NNs, able to reduce the disturbance, can effectively exploit the VIS information, thus resulting in higher accuracy.

## 5. CONCLUSIONS

We reported the results of experiments on retrieving ozone concentration profiles and TOCs from SCIAMACHY measurements. The retrieval has been carried out by suitably designed Neural Network algorithms. An interesting result is that including visible radiation in the measurement set increases the retrieval accuracy. The improvement is particularly appreciable for the troposphere, both for ozone profiles and column, but is also present at stratospheric heights. In particular, the study indicates that the accuracy of the retrievals can be increased up to 20% by adding measurements in the visible band.

The capability of successfully exploiting the information contained in VIS wavelengths is a distinctive feature of retrieval algorithms based on NNs, and makes them useful tools for monitoring the ozone concentration at low altitudes. In this regard, a major target to pursue is the systematic investigation of the sensitivity of satellite measurements at UV/VIS wavelengths to the ozone concentration near the surface. This task appears crucial to progress towards monitoring air quality from satellite platforms.

## REFERENCES


1. F. Del Frate, M. Iapaolo, S. Casadio, et al., Neural Networks for the dimensionality reduction of GOME measurement vector in the estimation of ozone profiles. *J. Quant. Spectrosc. Radiat. Trans.*, 92, 275-291, 2005.

2. P. Sellitto, A. Burini, F. Del Frate, D. Solimini, S. Casadio, Dedicated neural networks algorithms for direct estimation of tropospheric ozone from satellite measurements. *Proc. IGARSS 2007*, Barcelona (Spain).

3. P. Sellitto, F. Del Frate, D. Solimini, et. al., Neural network algorithms for ozone profile retrieval from ESA-ENVISAT SCIAMACHY and NASA AURA-OMI satellite data. *Proc. IGARSS 2008*, Boston (USA).

4. World Ozone Ultraviolet Data Center. http://www.woudc.org

5. A. M. Thompson, J. C. Witte, SHADOZ (Southern Hemisphere ADditional OZonesondes): An Ozonesonde Network and Resource for Remote Sensing Research and Education. *Proc. 28th International Symposium on Remote Sensing of Environment*, 27-31 March 2000, Cape Town (South Africa).

6. C. Zehner, S. Casadio, A. di Sarra, E. Putz, Temperature Independent Differential Absorption Spectroscopy (TIDAS) and Simplified atmospheric Air Mass Factor (SAMF) techniques for the measurement of total ozone using GOME data. *Proc. ESA ERS-ENVISAT Symposium*, 16-20 October 2000, Gothenburg (Sweden).

7. R. Siddans, W. J. Reburn, B. J. Kerridge, R. Munro, Height Resolved Ozone Information in the Troposphere and Lower Stratosphere from GOME. *Proc. 3rd ERS Symposium "Space at the Service of our Environment"*, 615-620, 17-21 March 1997, Florence (Italy).