ANN OZONE RETRIEVAL WITHIN THE OPERATIONAL IASI LEVEL 2 PROCESSOR

Thomas August, Peter Schlüssel, Rosemary Munro, Xavier Calbet, Olusoji Oduleye, Ariindo Arriaga, Tim Hultberg, Juliette Hadji-Lazaro, Solène Turquety, Cathy Clerbaux

1EUMETSAT, Am Kavalleriesand 31, 64295 Darmstadt, Germany
2Service d’Aéronomie, Université Paris 6, 75252 Paris Cédex 05, France

Abstract

Four ozone columnar amounts are currently retrieved by the EUMETSAT Polar System’s Level 2 (L2) Product Processing Facility (PPF), operated at EUMETSAT, from the Infrared Atmospheric Sounding Interferometer (IASI) measurements. Still under calibration and validation, they are distributed in a pre-operational mode since Spring 2008, in near real time (NRT) by EUMETSAT via EUMETCast on a pixel basis and sub-sampled via GTS within the OZO BUFR packets.

Artificial Neural Networks have been chosen to fill the L2 products since the 12 August 2008 and the installation of IASI L2 PPF v4.3. We present here a training strategy relying on real data as an alternative to synthetic learning datasets based on numerical radiative transfer models. The learning root mean square was as low as 11 DU for ozone total column error at non-polar latitudes, translating into 3.8 % relative error. Additionally, this good performance remains stable with scan angle and land/sea surface type. Limited to three distinct days, the representativeness of this learning set has been tested with a collection of data covering a longer period of 9 months and involving ECMWF analyses as well as GOME-2 assimilated products. Despite a certain bias, the standard deviations remain of the order of the training errors for the higher columns: <16 km and total amount. The accuracy of the tropospheric columns, <12 km and particularly <6 km is however not as stable yet, especially over land.

INTRODUCTION, ANN O3 RETRIEVAL WITHIN THE IASI L2 PPF

The Infrared Atmospheric Sounding Interferometer (IASI) is an advanced hyperspectral Infra-Red (IR) sounder, launched onboard Metop in October 2006. It covers the [15.5μm – 3.62μm] IR range with a spectral sampling of 0.25cm-1 allowing the quantification of atmospheric chemical components – such as ozone, carbon dioxide, methane, etc. – which play a basic role in tropospheric phenomena and the climate change.

Two statistical methods can be run within the PPF to retrieve the ozone components, respectively involving Empirical Orthogonal Functions (EOF) and Artificial Neural Networks (ANN). The results are available in L2 products for clear cases and for areas of moderate elevation variability only. The ability of ANNs to retrieve trace gas columnar amounts from hyperspectral sounder data has been demonstrated and the operational processor implements the results of studies performed for EUMETSAT by the "Service d’Aéronomie, Université Paris VI" until 2002. This technique applies to other trace gases amounts, namely the CO₂, N₂O, CO, CH₄ total columns, also forming part of the IASI L2 products. The trace gases amounts are entered as input state vector parameters to a one dimension variational retrieval performing the temperature and humidity profile retrievals within the PPF. Ultimately, ozone retrieval could be activated as well within this iterative method.

Following a series of acceptance tests on the IASI L1c spectra, a selection of vertical temperatures and of radiances, subtracted from the baseline, are entered as inputs to a multi-layer perceptron. The net, of the feed-forward type, is composed of two hidden layers between the input and output ones. The number of channels and neurons are configurable. The actual network inputs and outputs are normalised such that the global transfer function for the ozone column \( p \) reads:

\[
O_p = D_p \cdot g \left( \sum_{k=1}^{S_2} w_{pk} \cdot f \left( \sum_{j=1}^{S_1} w_{kj} \cdot f \left( \sum_{i=1}^{N_f} w_{ij} \cdot E_i + b_j \right) + b_k \right) + b_p \right) \tag{1}
\]
where \( b \) and \( w \) are respectively the biases and the weights of the neurons and their connections. \( f(x) = \tanh(x) \) and \( g(x) = x \) are the transfer functions of the hidden and output layers, respectively. \( D \) is the output rescaling coefficient to the final ozone column while \( E \) is the normalised input \( i \) computed as follows:

\[
E_i = 0.9 \times \frac{\text{Input}(i) - C_1}{C_2}
\]  

(2)

where \( C_1 \) and \( C_2 \) are normalisation coefficients defined globally for all input radiances and temperature respectively.

**THE \( \text{O}_3 \) ANN TRAINING**

Until the IASI L2 PPF v4.2, the auxiliary coefficients driving the ANN and EOF resulted from a teaching performed with synthetic spectra. These patterns were computed with a sub-sample of ECMWF reanalysis atmospheric state vectors\(^6\) and the Radiative transfer model RTIASI-5. Although giving satisfying results with the synthetic learning base, the actual retrievals appeared biased when compared to ECMWF analyses and GOME-2 ozone total column, which on the other hand agree well with each other. Both statistical methods tended to overestimate the actual amounts within the IASI L2 PPF (see Figure 1, left). Moreover, a slight scan angular was reported as well as a degradation of the retrieval quality over land. As shown too on Figure 1, the synthetic training set was eventually not representative enough of the actual ozone vertical distribution variability.

---

**Figure 1:** Representativeness of the retrieved and synthetic ozone data sets. On the left, the histograms of ozone total column retrieved by the two IASI L2 PPF statistical methods, trained with synthetic data, plotted together with the synthetic training abundances and real statistics computed with 3 days of GOME-2 products and ECMWF analyses data. To the middle and the right, ozone profile covariance is represented for the synthetic training set and ECMWF data, respectively.

For these reasons, a training set based on real data was built to assess the ability of a single network to retrieve the ozone columns with the sought accuracy over land, sea and at all scanning angles. It is expected to capture better the actual land surface emissivity variability. The input vectors were built with IASI L1c measured radiances and with ECMWF analyses temperatures for the thermal component. The corresponding teaching output total columns were taken from assimilated GOME-2 products\(^4,9\), available on-line from the Tropospheric Emission Monitoring Internet Service (TEMIS), [http://www.temis.nl](http://www.temis.nl) while the partial columns were integrated from ECMWF analyses ozone profiles. We restricted in a first step this study to non-polar latitudes. The teaching method was the simple online standard backpropagation where weights are updated after each presentation of an input/output pair. A cycle (epoch) is completed when all teaching patterns have been presented to the net, which is done in a random manner. We also successively implemented two changes to speed-up and improve the learning. The first one gives each layer an individual learning rate\(^6\) while the second and the most important attributes to each input and output element a specific pair of normalisation coefficients. They are used with eq.2 and let the network benefit from a much higher dynamic with all input/output elements while learning. In that case, \( C_1 \) and \( C_2 \) were redefined as follows:

\[
C_1(i) = \frac{\text{percentile}(\text{element}(i),98.5\%) + \text{percentile}(\text{element}(i),1.5\%)}{2}
\]  

(3)

\[
C_2(i) = \frac{\text{percentile}(\text{element}(i),98.5\%) - \text{percentile}(\text{element}(i),1.5\%)}{2}
\]  

(4)
Figure 2: Example of ozone total column training and verification errors with various network configurations as a function of the learning cycles (epochs). (a) Net trained with synthetic data and a single pair of scaling coefficients: \(c_1 = \text{min}, \ c_2 = [\text{max}-\text{min}]\). (b) same as (a), with a different learning rate defined for each layer. (c) One pair of normalisation coefficient \((c_1, c_2)\) for each input and output. (d) same network as (b), trained with real data (21/06, 08/09 & 07/12/2007) and tested with real data from the 22/06/2007. (e) similar net settings as in (c), taught with real data (21/06, 08/09 & 07/12/2007).

Figure 3: Statistics and distributions of the ozone total column training error with real data. Top left and right histograms relate to the absolute and relative error, respectively. The teaching vs. learned correlation (of about 0.95) is plotted in the bottom left corner while the map to the right shows the geographical distribution of the learning error.

To populate the training set, clear IFOVs, as identified by IASI in the L2 PPF, were selected from all orbits of those three days: 21 June, 8 September and 7 December 2007. A further restriction towards the latitudes (<60°) and the surface pressure (>980hPa) was applied such that the learning was based on about 300,000 cases eventually. Similarly, a control set was built with all orbits of the 22 June 2007, which contained about clear 100,000 clear cases. The training errors variation along the learning process is plotted in Figure 2 for the various configurations detailed in the previous paragraph. The verification error for ozone total column – (d) in dark blue –, includes both land and sea cases of the control set. It is 1 to 2 DU smaller than the error obtained over oceans with a net trained on synthetic data and 6 DU below the same error over lands. A further gain of 2 to 3 DU, observed in the training error with the individual scaling scheme exposed in eq. 3 and 4, could be expected in actual retrievals but still needs to be confirmed with other verification dates. An example of a training correlation and error is given for ozone total column in Figure 3. In particular, we can verify the continuity of the good retrievals over land and sea as well as the absence of any scan angle effect. However, the polar ozone amounts are not well inferred from the training set, which excluded polar cases. Table 1 summarizes the statistics for the training and verification errors for all four ozone columns.
VALIDITY OVER A LONGER PERIOD

Given the limited number of dates involved in the learning phase, an extended verification exercise was performed, which covered the period from June 2007 to February 2008. This comprises 36 orbits of IASI measurements, centred on ECMWF regular analysis times with the corresponding ozone data derived from the assimilated GOME-2 products as well as the integrated ECMWF analyses profiles. The match-up time between computed IASI L2 ozone amounts and ECMWF analyses does not exceed 1 h. The IASI L1c spectra were reprocessed with the version v4.3 of the L2 PPF. Only clear cases as identified by IASI within the L2 processor were kept for this comparison (~300 000 cases) and the nearest neighbours in the GOME and ECMWF horizontal grids were linearly interpolated to the IFOV centres. The training days were excluded from this exercise which aims to assess the validity of the retrieval over a longer period of time and a wider range of conditions.

As shown on Figure 4 for sea cases, the correlation between the retrieved ozone total column and the reference based on GOME-2 products remains good, of the order of the correlation obtained during training. Similarly the standard deviation of 4 % is comparable to the 3.7 % obtained with the learning set. A minor population of outliers was ignored in these statistics which are still under investigations. They particularly affected the partial columns and are most likely due to cloud contaminations. The bias however is no longer negligible and amounts to about 2 %. This behaviour applies to the two biggest columns, both with ECMWF and TEMIS products. Statistics have been computed separately for land and sea scenes which show a better match over oceans than over continents. The difference is smaller than 1 % when GOME-2 is taken as a reference (standard deviation of 4.0 % and 4.9 % for sea and land, respectively) but increases to several percents when ECMWF analyses are used as the reference. Especially the partial columns over land show major deviations as presented in Table 2 where the statistics for all land/sea, column and reference configurations are detailed.

Eventually, the ozone column errors were plotted as a function of time (see Figure 5). No seasonal trend appears clearly but we can see that the retrieval accuracy is affected in the same way for all columns. This suggests that the training sets were not representative enough for all real configurations. A too big time interval between the measured radiances and the teaching amounts could also account for it since ECMWF integrated ozone columns were up to 3 h apart and since the TEMIS products are generated for 12.00 UTC. Also, the nets were trained with low scenes only and has by essence a limited validity over high elevations, which particularly impacts the below 6 and 12 km. Additionally, the retrieval of the low tropospheric ozone requires good thermal contrast conditions, which are not systematically matched and which affect the actual sensitivity in the lower troposphere. Eventually, the scan angle does not seem to impact the accuracy of the retrieval with this verification dataset, which would confirm the results obtained during the training phase.

Table 1: Training and control errors obtained after network teaching with real data.

| Columns | p | biasDU | σDU | <<% | σ% | p | biasDU | σDU | <<% | σ% |
|---------|---|--------|-----|-----|----|---|--------|-----|-----|----|----|
| <6 km   | 0.89 | -0.04 | 1.87 | -0.96 | 10.93 | 0.862 | -0.01 | 1.95 | -0.8 | 9.7 |
| <12 km  | 0.90 | -0.27 | 2.58 | -1.13 | 6.81 | 0.91 | -0.23 | 2.77 | -1.1 | 6.72 |
| <16 km  | 0.96 | 0.04 | 3.87 | -0.48 | 7.06 | 0.96 | 0.05 | 4.27 | -0.52 | 7.27 |
| TotalGOME | 0.95 | 1.91 | 10.74 | 0.55 | 3.77 | 0.94 | 2.88 | 11.63 | 0.81 | 3.98 |
| TotalECMWF | - | - | - | - | - | 0.94 | 3.83 | 11.26 | 1.22 | 3.96 |

Table 2: Validation errors obtained with artificial neural network trained with real data.

| Columns | p | biasDU | σDU | <<% | σ% | p | biasDU | σDU | <<% | σ% |
|---------|---|--------|-----|-----|----|---|--------|-----|-----|----|----|
| <6 km   | 0.15 | -1.89 | 15.93 | -11.2 | 101 | 0.67 | 0.59 | 2.48 | 1.27 | 10.34 |
| <12 km  | 0.65 | -1.96 | 5.24 | -6.41 | 13.35 | 0.73 | 0.86 | 3.19 | 1.53 | 8.13 |
| <16 km  | 0.85 | 0.56 | 10.32 | -1.87 | 13.48 | 0.90 | 1.42 | 4.80 | 2.0 | 8.42 |
| TotalGOME | 0.89 | 4.48 | 17.27 | 2.00 | 4.91 | 0.89 | 3.70 | 13.31 | 2.0 | 4.10 |
| TotalECMWF | 0.84 | 4.38 | 20.02 | 1.23 | 6.54 | 0.85 | 6.23 | 14.58 | 2.18 | 4.74 |
CONCLUSIONS AND FUTURE WORK

We trained single networks to retrieve ozone columns at all scan angles and under all surface conditions with real data limited to non-polar regions. The ability of such nets to perform this task was confirmed and training errors at IFOVs scale (12 km at nadir to 40 km at swath edge) are in line with the user requirements of 5% and 15% with a 250 km horizontal resolution for total and partial columns, respectively. This approach is considered as a good alternative to training with synthetic data where a precise and exhaustive surface modelisation is essential. Noticeably, such trained nets apply with the same accuracy over more difficult areas such as deserts.

The validity of the trained networks over a longer period and wider range of atmospheric conditions was assessed with 36 orbits covering June 2007 to February 2008. The actual vertical resolution of IASI, GOME-2 and the model profile was not taken into account in the comparisons. The quality of the ozone columns remains however stable with scan angle and retrieval errors over ocean are of the order of errors obtained at training. A bias of about 2% has however been introduced in the total
column and the quality of the land retrieval degrades for partial columns: particularly the lower tropospheric ozone computed with the ANN differ from the ECMWF reference. This points to some training set limitations in terms of seasonal and atmospheric configurations coverage.

Future work will include polar region in the training base, possibly with a dedicated network, and a collection of IASI L1c spectra and ozone amounts pairs covering a broader period of time with closer time coincidence. Further validations with ground-based measurements are currently being carried out and will be extended to products from the World Ozone and Ultraviolet radiation Data Centre (WOUDC). Eventually, an additional improvement of the order of 2 DU in the retrieval accuracy is expected from a normalisation scheme more specific to each input and output elements.

ACKNOWLEDGEMENTS

The authors acknowledge KNMI/TEMIS and all people operating the internet services and data assimilation. The authors thank Maya George for supporting the verification of the PPF implementation of the trace gases ANN retrieval as designed by the "Service d'Aéronomie, Uni. Paris 6".

REFERENCES