CARBON MONOXIDE RETRIEVAL WITHIN THE OPERATIONAL IASI LEVEL 2 PROCESSOR, TRAINING AND VALIDATION RESULTS

Thomas August, Peter Schlüssel, Rosemary Munro, Tim Hultberg, Olusoji Oduleye, Xavier Calbet, Arlindo Arriaga

EUMETSAT – Eumetsat-Allee, 1 - 64295 Darmstadt - Germany

Abstract

The carbon monoxide total column retrieval operated in the Infrared Atmospheric Sounding Interferometer (IASI) L2 processor was revised and the upgraded algorithm is presented here. The modifications aimed at improving the retrieval accuracy, at correcting an angular dependency and enabling inversions over specific elevated or desert areas. We describe the algorithmic changes in this paper and present their theoretical performances together with initial validation results performed with external CO products from the Measurements Of Pollution In The Troposphere (MOPITT) mission. They confirm the overall improvements with typical differences of approximately 0.25×10^{18} molecules/cm^2.

INTRODUCTION

IASI is an advanced hyper-spectral Infra-Red (IR) sounder, launched onboard Metop in October 2006. It covers the [3.62 µm – 15.5 µm] IR range with a spectral sampling of 0.25cm^{-1} allowing among other atmospheric products the quantification of chemical components such as ozone, carbon dioxide, carbon monoxide, methane and nitrous oxide. The CO total column is retrieved in near-real time by the IASI L2 Product Processing Facility (PPF) operated at EUMETSAT and distributed via GTS in a pre-operational mode within the TRG BUFR product since April 2008.

The retrieval is based on an artificial neural network (ANN). A preliminary validation cycle of the initial CO retrieval algorithm, including comparisons with model outputs and external instrument products, concluded on a significant positive bias as well as a scan angle dependency and some artefacts over particular areas such as elevated terrains, deserts and ice coverage. The root causes were analysed such that a first upgrade mainly addressing the abundance overestimation and the surface elevation was applied to the IASI L2 PPF and became pre-operational in January 2009 with the revision 4.3.2.

We present here further improvements made to the CO retrieval scheme, consisting in additional predictors and a modified and extended use of the spectral information, the training approach, new theoretical performances, and an intercomparison with CO columns as derived from MOPITT. The learning error amounts to about 0.24×10^{18} molecules/cm^2, which typically translates into 7% to 11% in relative terms.

This network, which will become operational in the first quarter 2010 with the revision 5.0 of the IASI L2 PPF, was applied to real measurements covering the last week of August 2008. The retrievals were analysed and inter-compared with MOPITT products of the same period. We present here some initial validation results which show a flat response to the scanning angle as well as spatial coherence and departures from MOPITT total columns consistent with the theoretical figures.

THE INITIAL RETRIEVAL SCHEME AND SKILLS

Such ANN retrievals are basically statistical non-linear regressions between a targeted output, here the CO total column, and a collection of inputs meant to capture all the required physics it relates to. The structure considered here is a multi-layer perceptron (MLP) with two hidden layers. An overview of the algorithm is recalled in Figure 1 and formalised in the equation (1).

$$CO = D \cdot g\left(\sum_{k=1}^{S_2} w_{pk}^{3} . f\left(\sum_{j=1}^{S_1} w_{kj}^{2} . f\left(\sum_{i=1}^{N_E} w_{ji}^{1} . E_i + b_j^{1} \right) + b_k^{2} \right) + b^{3} \right) \quad (1)$$

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 CO column while E_i is the i^{th} normalised input computed as follows:

$$E_i = 0.9 \times \frac{\text{Input(i)} - C_1}{C_2} \qquad (2)$$

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

The regression is performed with a set of coefficients triggering the connections between the inputs and outputs, namely the weights and biases. The input and output usually are normalised parameters in the scope of the neural network itself. The weights and biases governing the retrievals are adjusted in a so-called training phase, where teaching pairs of input and output are iteratively presented to the network.

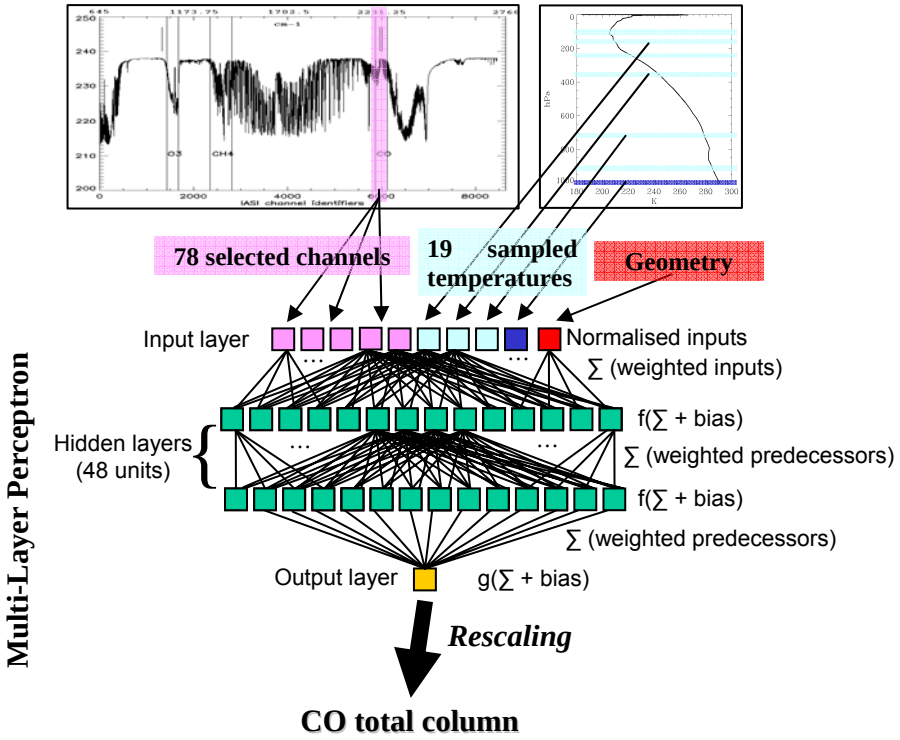


Figure 1 : Overview of the ANN CO retrieval scheme implemented in the IASI L2 PPFv5

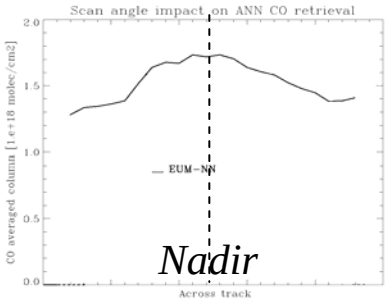


Figure 2 : Angular variation of the CO total column retrieved over the Pacific Ocean on the 25 August 2008 with IASI L2 PPFv4. Mean abundances have been computed for successive scan positions around the Nadir, placed in the centre

The abilities of such techniques at retrieving trace gas columns from infrared spectra in Nadir views had been demonstrated[6,12] and these results served as a basis for the earlier implementations in IASI L2 PPFv4[3,11,13]. The input vector was composed of a coarse temperature profile coming on a fixed grid and of the remainders of the radiances measured in selected CO lines subtracted from the upwelling background radiance. The latter was inferred from a couple of so-called baseline channels together with a fixed surface emissivity, set to 0.975 for this particular gas. These CO retrievals therefore were not suited for barren areas with specific surface emissivities like deserts and ice covers, neither could they apply to elevated terrains. No account is being made for the local sensing geometry, therefore the

retrieved CO columns exhibit a strong dependency to the scan angle[7], with an amplitude of about 25% between the Nadir and the swath edge, as illustrated in Figure 2.

MODIFICATIONS IN THE IASI L2 PPF AND THEORETICAL PERFORMANCES

The version 5 implements several algorithmic changes in the area of ANN CO retrievals, the first of which deals with a refined handling of the inputs and outputs. The artificial nets are indeed advantageously taught with patterns which have previously been centred and normalised for a faster and more accurate learning[8]. The first improvement, which had already been tested with benefits in the ozone retrieval[1], consists in an individualised scaling to account for the dynamic range of each input/output parameter. They initially were normalised with the same coefficient despite their much different ranges illustrated in the Figure 3. The coefficients used in (2) are now defined as:

$$C_1(i) = (\text{percentile(element(i),98.5\%)} + \text{percentile(element(i),1.5\%)}) / 2 \quad (3)$$

$$C_2(i) = (\text{percentile(element(i),98.5\%)} - \text{percentile(element(i),1.5\%)}) / 2 \quad (4)$$

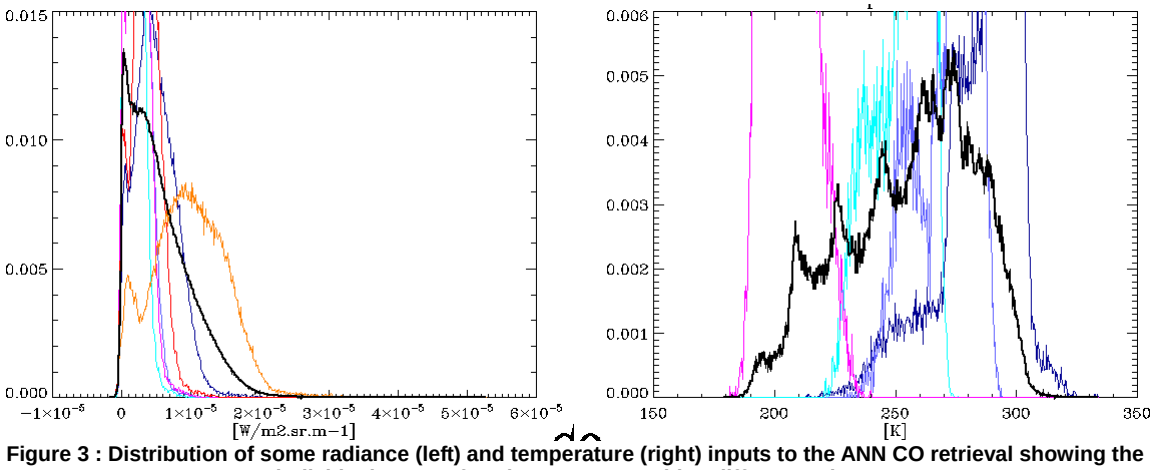


Figure 3 : Distribution of some radiance (left) and temperature (right) inputs to the ANN CO retrieval showing the individual range of each component with a different colour.

As for the scan angle dependency and the retrievals over elevated area, it is proposed in this revision to explicitly handle the local sensing geometry with two additional predictors, namely, the satellite zenith angle and the surface pressure. As the input temperature is expressed on a fixed grid, some of the levels actually fall below the surface with elevated surfaces. Such situations are handled with an isothermal extension of the temperature profile (at the brightness temperature of the surface) to the subsurface levels, which enables the training and subsequently the retrievals with all altitudes.

Further to this, we propose a modified spectral sampling, which is the key information driving the retrievals. The channels split into two categories: a selection within the various CO absorption lines and a smaller set of so-called baseline channels which describe the background radiance before it is attenuated by CO. The channel selection for both absorption and baseline channels have been extended (see Figure 4) and the baseline channels become explicit additional inputs to the net. The latter selection was extended to increase the signal-to-noise ratio but also to include pure absorption lines of two species interfering with CO in this spectral region: water vapour and N_2O. The surface temperature is kept as an input parameter such that the surface emissivity, which is no longer a static value, as well as potential interference with other species are assumed now to be implicitly covered by the combination of the inputs and expected to be statistically captured during the network learning phase. Some tests however showed that the surface emissivity further helped the CO retrieval if explicitly added to the input vector. This parameter, also derived from IASI spectra, could become a valuable additional descriptor of the physics fed into the neural net.

Config. / Gas	Initial settings	PPFv5 improvements + initial channel selection	PPFv5 improvements + modified channel selection
Training error 10^{18} molec/cm^2	0.370[a]	0.270[b]	0.235[b]

[a] *elevated ground (Ps < 980hPa) excluded* [b] *all elevations are included in the training sets*

Table 1 : Evolution of the ANN CO training error with the successive algorithmic and configuration improvements.

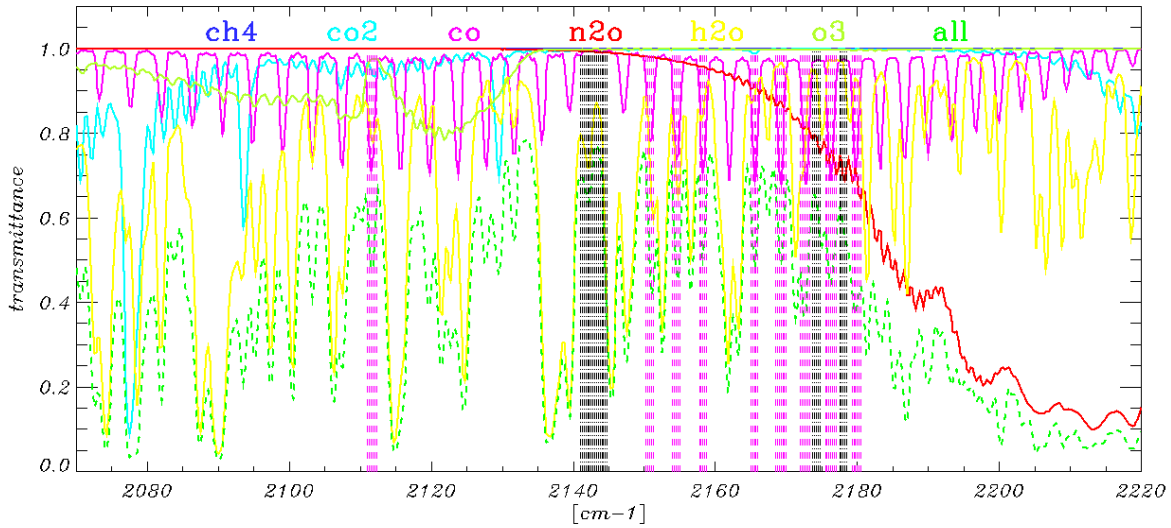


Figure 4 : Modified channel selection for PPFv5.0. Pink and black vertical dash-lines respectively show the absorption and baseline channels involved in the ANN CO retrieval. The various transmittances were computed for a tropical case.

The training database contains approximately 200 000 patterns made of atmospheric state vector and their associated synthetic IASI spectra computed with RTIASI-5.3 for all viewing geometries. The synthetic spectra were degraded with the instrument noise characteristics and so were the auxiliary input temperature to account for their associated errors in real operations. The atmospheric temperature, humidity, surface pressure and wind components are based on a climatology sub-sampled from the ECMWF 40 years re-analysis. Neither clouds nor aerosols were included such that the networks learnt pure clear cases only, which subsequently also is their only domain of validity. The trace gas profiles were forged to cover the whole range of expected situations with random variations around standard profiles. In the case of the CO, the vertical distributions are based on 43 original profiles sampled from the MOZART 3D chemical transport model calculations[2], obtained from D. Cunnold's runs in 2001. Over water, the surface emissivity was computed following the Masuda's model and its extension by Watts[3]. Ground emissivities were derived from the MODIS UCSB emissivity library (www.icess.ucsb.edu/modis/EMIS/html/em.html). Based on the spectra of pure surfaces, composite surfaces have been generated with random contributions from up to three different types, however, excluding some combinations like snow/ice at tropical temperatures.

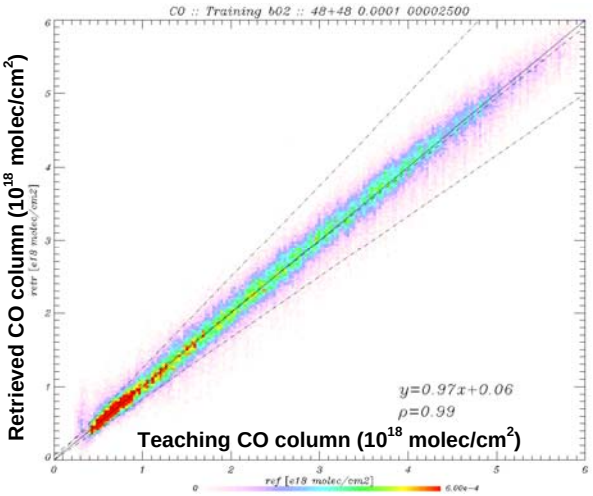


Figure 5 : Linear fit of the retrieved CO total columns to their associated target after training with noisy inputs. Dash-dot lines are the ± 20 % showing EURD thresholds.

The theoretical performance of such artificial neural networks can be characterised with the synthetic dataset that served in the training process by comparing the retrieved columns to the targets. Table 1 summarises the learning skills improvement with the successive introduction of the algorithmic changes and the modified channel selection. Results for the proposed network are detailed in Figure 5 to Figure 7. The correlation between retrieved and target columns is 0.99 and the linear relation is very

close to identity. Due to the non-linear nature of the MLP, the errors are not statistically Gaussian-like, as can be seen in Figure 7. The rms of the absolute errors typically ranges from 0.2 to 0.24 x 10^{18} molec/cm^2 and appear to be quite independent of the column density itself. As a result, the relative errors are amplified for the thinnest columns. However, generally they remain well below the thresholds specified in the EPS End User Requirements Document (EURD), set to 20 %. Excluding the 7 % faintest columns (< 0.7x10^{18} molec/cm^2) from the error characterisation, which mostly correspond to the extremely elevated surfaces like that of Antarctica or Himalaya, does not change the absolute errors and gives an estimation of the relative error between 7 and 11 %.

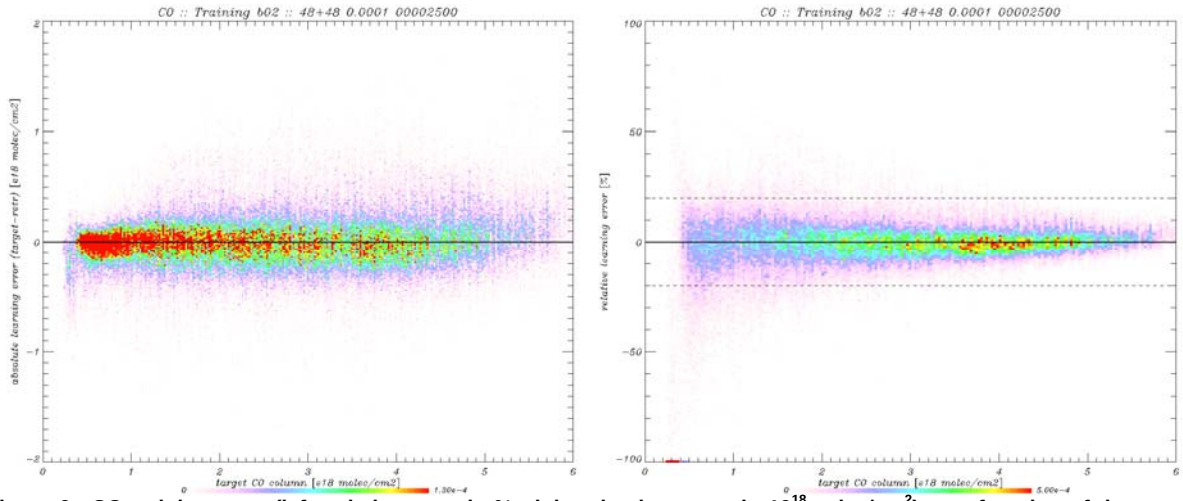


Figure 6 : CO training error (left: relative error in %, right: absolute error in 10^{18}molec/cm^2) as a function of the target CO column. Noisy inputs were used for training. Dash lines at ± 20 % show the thresholds after EURD.

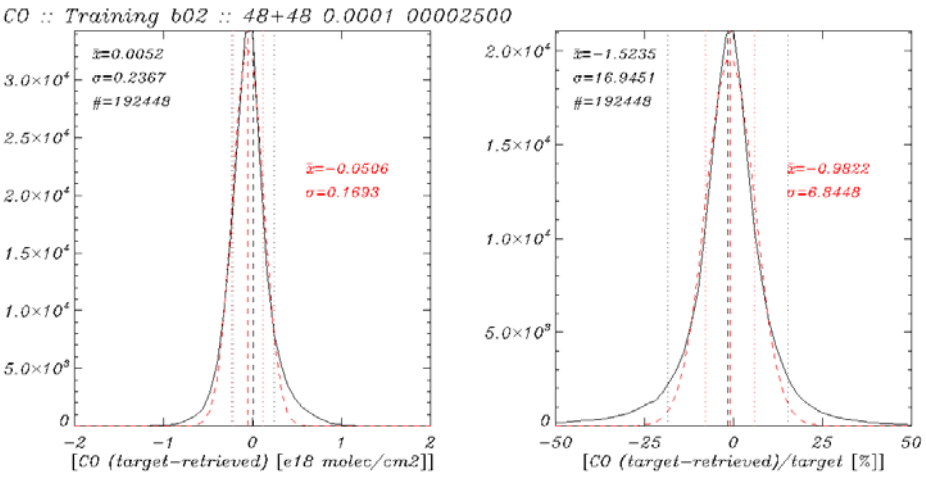
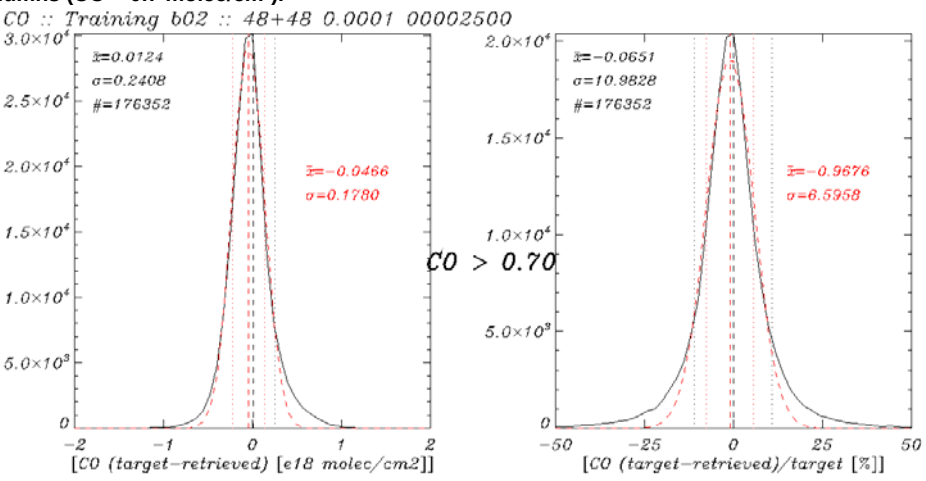


Figure 7 : CO absolute (left) and relative (right) training errors. The overall statistics are displayed in black while the most fitting Gaussian distributions and associated numbers are shown in red. Below, the same statistics without the 7% faintest columns (CO < 0.7 molec/cm^2).



VALIDATION WITH SATELLITE DATA

A week worth of IASI measurements covering the last week of August 2008 was reprocessed with this ANN inversion scheme over clear scenes as identified by cloud detection tests in the IASI L2 PPF. The known CO sources[5] are qualitatively well retrieved and can be seen in Figure 8. They essentially split into two main classes: the biomass burning below the Equator in Africa and South America on the one hand and the agricultural fires and industrial emission over Northern India and China on the other hand. The latter plumes are transported out over the Pacific Ocean beyond Japan[10]. IASI ANN retrievals were compared over that period on an IFOV basis to MOPITT L3 Daily gridded (1°x1°) CO products (v3) downloaded from the NOAA WIST server (http://wist.echo.nasa.gov).

Mean IASI CO columns were computed for each viewing position (EFOVs) during that period and for day and night time respectively. Averaged over the about 100 orbits, they are expected to be approximately constant with scan angle. Furthermore, densities retrieved at daytime are also expected to match those retrieved during night. The short period addressed here does not offer daylight sensing over Antarctica and conversely, no night counterparts are available for North Pole measurements. Therefore the polar caps were excluded from this particular angular analysis. As can be seen in Figure 9, the implementation in PPFv4.3 shows a significant angular dependence, with increasing columns towards the Nadir. The amplitude of the variations is as high as 0.3×10^{18} molec/cm^2 at daytime and even 0.4×10^{18} molec/cm^2 for night cases. The proposed new algorithm is showing a much flatter profile in general, with a small decrease inferior to 0.1×10^{18} molec/cm^2 at the swath edges for both day and night retrievals.

With an atmospheric lifetime of about two months, day and night time CO retrievals are expected to generally match with each other. With the exception of some desert regions in Australia, Mongolia or Sahara, the day/night contrast in IASI products is globally relatively small in average and fluctuates around 0.05×10^{18} molec/cm^2. This number is essentially indicative because clear land cases are less represented at night and because the exact same Earth points were not necessarily associated with retrievals from the day and night overpasses. Especially the two main sources in South America and Central Africa are mostly retrieved at day time and were more discarded at night time by the cloud detection in both the IASI and MOPITT products. In the same way, low land and sea retrievals should present smooth transitions in coastal areas. This is in general achieved with only a few exceptions, e.g. off the coasts of Morocco and Mauritania. In that particular case however, MOPITT reported the area as cloudy and consequently did not provide any CO measurements.

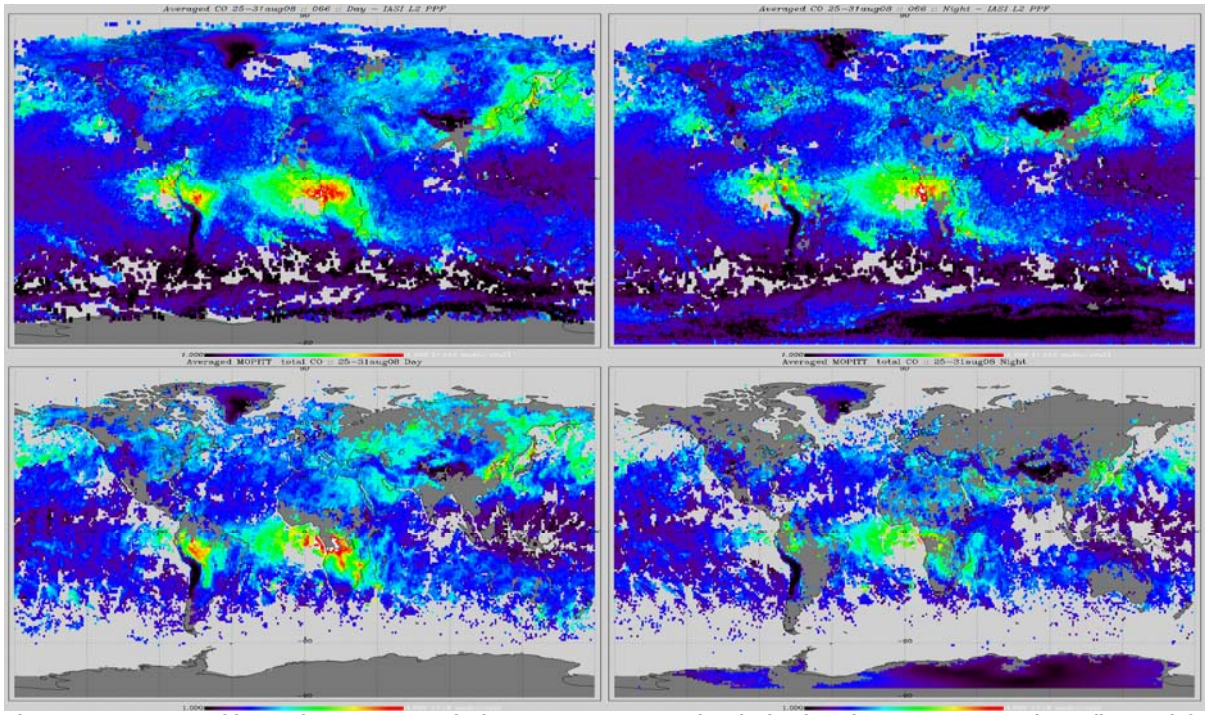


Figure 8 : CO L3 weekly products computed after IASI ANN CO retrievals (top) and MOPITT CO products (bottom) for the period 25-31 August 2008. Day and night maps are shown on the left and right hand-side, respectively.

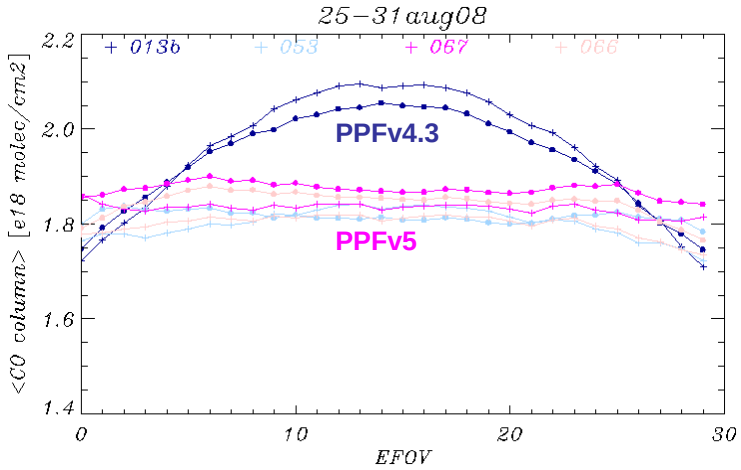


Figure 9 : Angular variation of the mean IASI CO total column retrieved with PPFv4.3 (blue) and v5 (pink) between the 25 and the 31 August 2008.

The inter-comparison with MOPITT products summarised hereafter is more exhaustively displayed in Figure 10 and Figure 11. The latter retrievals are based on an optimal estimation method and were handled here without their averaging kernels. Such products contain both the vertical profile and its integration to the total column. They also detail the altitude dependence of the final retrieval (CO profile) on the a priori information. An additional filter was applied to discard all retrievals constrained more than 50% by the background used in the MOPITT product generation. The study was broken down into day/night and geographical classes covering the Poles (|lat.| > 60°), mid-latitudes (30° < |lat.| < 60°) and a wide tropical-equatorial band from 30°S to 30°N.

The departures of the proposed implementation from MOPITT products are similar to the theoretical errors with standard deviations varying between 0.2 and 0.3 x 10^{18} molec/cm^2 in general, up to 0.39 in the Northern Hemisphere at night. This translates into approximately 11 to 18 % in relative terms. Their distributions are usually not symmetrical and tails where IASI ANN CO is in excess are visible, of which root cause is still under investigation. Although the standard deviation of the inter-comparison only varies by a few percent in the various geographical and illumination classes, the bias can be very different from one class to another. It reaches 25% in Antarctica while amounting to 13% in the Southern Hemisphere, 7% in the North Pole and almost vanishes between 30°S and 60°N. In terms of correlations, the best results are obtained at daytime.

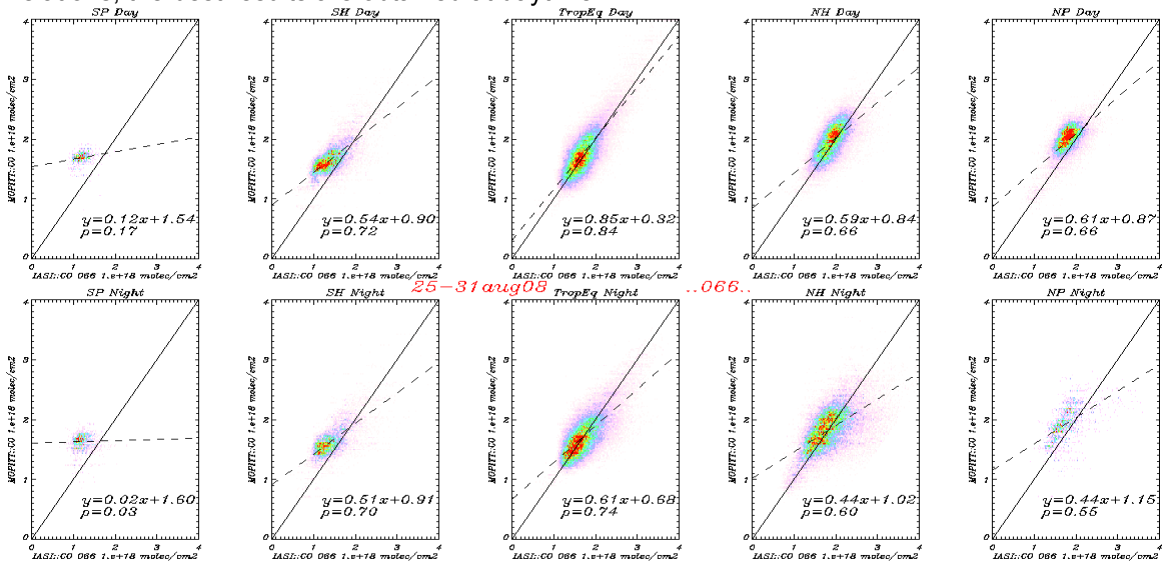


Figure 10 : Day/night and geographical breakdown of the correlation between IASI and MOPITT CO averages retrieved for the last week of August 2008. From left to right the latitudes are grouped in the following bands: SP [-90° ;-60°], SH [-60° ;-30°], TropEq [-30° ;30°], NH [30° ; 60°] and NP [60° ; 90°].

Although giving good indications, no definitive conclusions can be drown from this exercise in terms of absolute calibration. Recent studies indeed showed that the MOPITT CO is overestimated[2], which would imply that the proposed implementation could statistically also be positively biased. However,

assuming a mean error of about 10 % in MOPITT products, these results already suggest that such artificial neural network retrievals mostly match the accuracy defined in the EPS mission requirements.

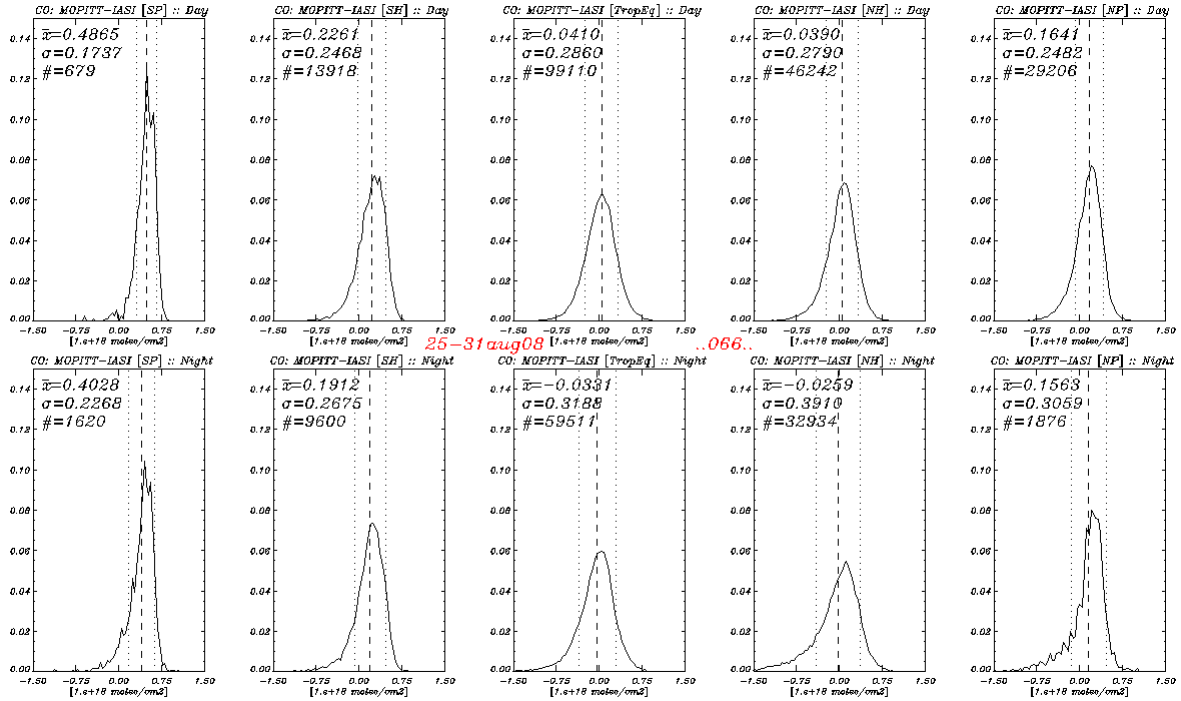
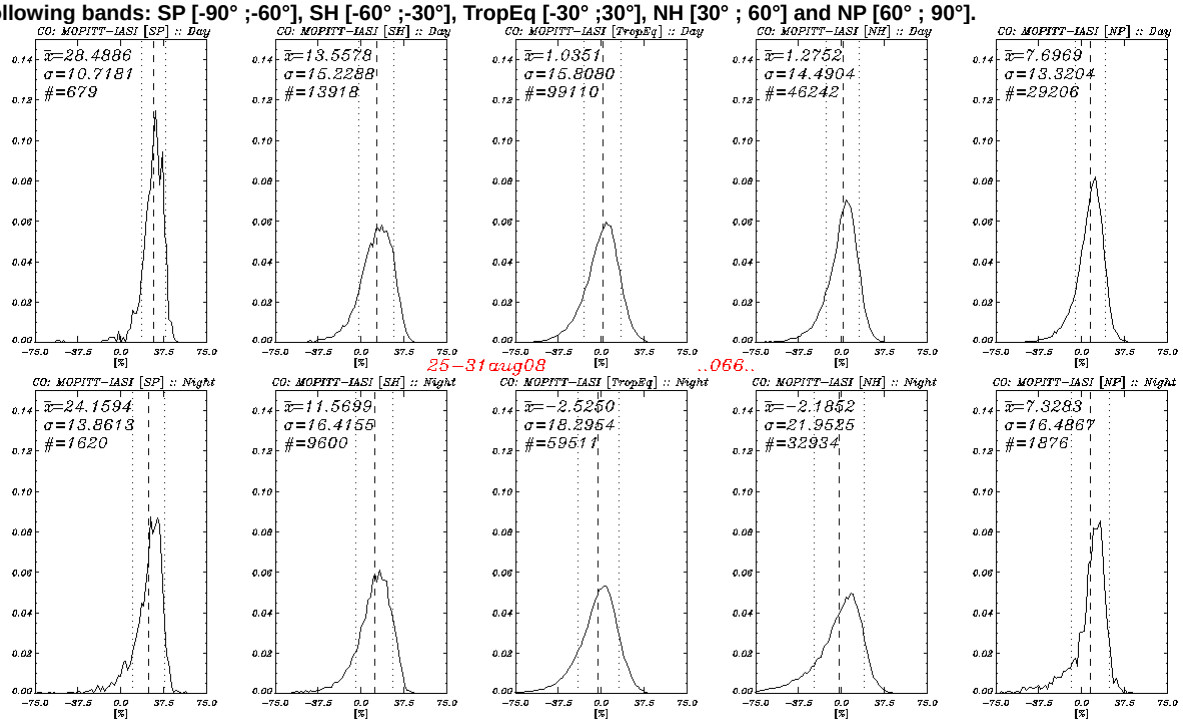


Figure 11 : Day/night and geographical breakdown of the IASI absolute (above) and relative (below) departures from MOPITT CO averages retrieved for the last week of August 2008. From left to right the latitudes are grouped in the following bands: SP [-90° ;-60°], SH [-60° ;-30°], TropEq [-30° ;30°], NH [30° ; 60°] and NP [60° ; 90°].



CONCLUSION AND FUTURE WORK

The initial specifications for a neural network retrieval ran in a pre-operational mode for several months and showed some limitations, which were characterised. The main issues dealt with a strong scan angle dependency, a positive bias and failure over elevated areas as well as some specific surfaces (ice and deserts). A number of modifications were elaborated with a single architecture and presented here to enable global retrievals at all viewing angles. The improvements were confirmed in the course of verification with training data and a preliminary validation with real IASI measurements

and external satellite products (MOPITT). In terms of timeliness, this technique is compatible with operational processing in the EPS ground segment and these upgrades will form part of the IASI L2 PPF release v5.

It is acknowledged that the validation exercise with IASI and MOPITT data covered a limited period. It will be extended to other times of the year to confirm the seasonal stability of the retrievals and to ground-based measurements using Fourier transform infrared spectrometer, which are of a higher accuracy. The method has not been tested yet over extreme events, like wild fires, which could require a further diversification of the training base. With the exception of the Antarctica, where unrealistic CO variations are observed, first error characterisations were derived and show that CO is retrieved with a typical precision of 0.2 to 0.3 x 10^{18} molec/cm^2 in general (7 to 15% in relative terms), which meet the user requirements.

ACKNOWLEDGEMENTS

The authors acknowledge Atmospheric Science Data Center at NASA Langley Research Center for the provision of MOPITT data via the WIST interface https://wist.echo.nasa.gov/

REFERENCES

[1] August Th. et al. (2008) “ANN ozone retrieval within the operational IASI level 2 processor”, EUMETSAT User Conference 2008.
[2] Brasseur, G. P., D. A. Hauglustaine, S. Walters, P. J. Rasch, J.-F.Müller, C. Granier, and X. X. Tie: „MOZART: A global chemical transport model for ozone and related chemical tracers, Part 1. Model Description”, Journal of Geophysical Research, 103, 28,265-28,289, 1998.
[3] Clerbaux C., J.Hadji-Lazaro, S. Turquety, G. Mégie : Algorithme d’inversion gaz traces pour IASI, Technical Document, Oct. 2000
[4] Emmons L. K., D. P. Edwards, M. N. Deeter, J. C. Gille, T. Campos, P. Nédélec, P. Novelli, and G. Sachse, Measurements of Pollution In The Troposphere (MOPITT) validation through 2006, Atmos. Chem. Phys. Discuss., 8, 18091–18109, 2008
[5] de Laat, A. T. J., J. Lelieveld, G. J. Roelofs, R. R. Dickerson, and J. M. Lobert (2001), Source analysis of carbon monoxide pollution during INDOEX 1999, J. Geophys. Res., 106(D22), 28,481–495.
[6] Hadji-Lazaro J., Clerbaux C. and Thiria S. (1999) “An inversion algorithm using neural networks to retrieve atmospheric CO total columns from high-resolution nadir radiances”, JGR, **104**, D19, pp.23841-23854
[7] Hultberg T., T. August: CO and O3 total column retrieval comparison, EUM/OPS-EPS/TEN/08/0700
[8] LeCun Y., L.Bottou, G.B.Orr and K.-R. Müller, “Efficient backprop” in Neural Networks: tricks of the trade, Springer, 1998
[9] Masuda K., T. Takashima, Y. Takayama, 1988: Emissivity of pure water and sea waters for the sea surface in the infrared window regions, Remote Sensing of Environment, 24, 313-329
[10] Pochanart P., O. Wild and H. Akimoto, “Pollution Import to and Export from East Asia” in “Handbook of Environmental Chemistry”, Springer Berlin / Heidelberg, Volume 4G/2004, pp.99-130.
[11] Schlüssel, P., T.H. Hultberg, P.L. Phillips, T. August, X. Calbet, 2005: The operational IASI Level 2 Processor, Adv. in Space Res., 36, 982-988.
[12] Turquety S., Hadji-Lazaro J. and Clerbaux C., (2002) “First satellite ozone distributions retrieved from nadir high-resolution infrared spectra”, GRL, **vol.29**, n°24, 2198, pp.51.1–51.4
[13] Turquety, S., J. Hadji-Lazaro, C. Clerbaux, D. A. Hauglustaine, S.A. Clough, V. Cassé, P. Schlüssel, G. Mégie, 2004: Operational trace gas retrieval algorithm, for the Infrared Atmospheric Sounding Interferometer, *J. Geophys. Res*., 109, D21301, doi:10.1029/2004JD004821.
[14] Watts P.D., M.R. Allen, T.J. Nightingale, 1996: Wind speed effects on sea surface emission and reflection for the Along Track Scanning Radiometer, Journal of Atmospheric and Oceanic Technology, 13, 126-141.