DEVELOPING A NEURAL NETWORK ALGORITHM TO CALCULATE TPW AND LAYER PRECIPITABLE WATER FROM GPM MICROWAVE IMAGERY

Miguel A. Martinez[1], Francisco J. Tapiador[2], Cecilia Marcos[1], Antonio Rodríguez[1]

(1) Agencia Estatal de Meteorología (AEMET), Spain
(2) Universidad de Castilla-La Mancha, SPAIN

Abstract

The Global Precipitation Measurement (GPM) core satellite will be launched on 2013. The GPM Microwave Imagery (GMI) instrument is a multi-channel, conical-scanning, microwave radiometer. It will provide information on fourteen channels: ten channels between 10.65 to 89 Ghz with H and V polarization, similar to TMI of TRMM, and four channels more with frequency 165.5 GHz (2 channels: V, H polarization) and 183.31 GHz (2 channels: ± 3 GHz, ± 9 GHz, V polarization). As a preparatory task to a comprehensive modelling, the calculation of simulated radiances from several profile datasets has been initiated with the use of RTTOV and the RTTOV coefficients from equivalent channels on other instruments as TMI on board TRMM and MHS on board METOP. As an example of the use of these simulated radiances datasets, here we explore the training of neural networks to obtain the Total Precipitable Water (TPW) and Layer Precipitable Water (LPW) following the methodology of PGE6 to 8 of NWC SAF for SEVIRI. This will allow us to check the possibility of having a very fast algorithm to get TPW, LPW and CLWP from GMI data. A second objective is to show the performance of the neural networks for TPW and LPW from GMI synthetic data over synthetic TMI and pseudo-GMI brightness temperatures calculated from ECMWF analysis.
Neural networks to calculate TPW and LPW at 5 different layers have been trained. Clear air synthetic brightness temperatures have been calculated using the profiles of 60L_SD dataset as input to the radiative transfer model RTTOV-9.3; 60L_SD and RTTOV9.3 have been provided by the NWP SAF. TMI and MHS coefficients have been used in order to get a proxy of GMI channels. The 5 layers of LPW are the same as in NWCSAF/MSG PGE06 and PGE07 products: BL is defined as the precipitable water at a layer between surface and 840 hPa, ML is defined as the precipitable water at a layer between 840 hPa and 437 hPa, HL as the precipitable water at a layer between 437 hPa and the top of the atmosphere. The new layers: M1 (840 hPa < p < 703 hPa) and M2 (703 hPa < p < 437 hPa), that are under developing, have also been trained and some examples are displayed here.

INTRODUCTION

There are several studies (e.g. Gairaola (2006); Mallet (2002) and Singh (2005)) that have made use of neural networks to calculate TPW from MW; in this paper the possibility of also obtaining the vertical distribution of precipitable water from MW data is explored. The objective of this paper is to explore the possibility of generalizing for MW the algorithm used for TPW and LPW in the NWCSAF/MSG software package based on neural networks. A process similar to that undertaken to train the neural network for the METEOSAT Second Generation IR radiances has been followed. In this paper, the early stage in the development of NWCSAF/MSG algorithm has been repeated for the GMI instrument. The process is based on generating synthetic BT for several MW instruments using the NWP SAF 60L_SD profile dataset and the NWP SAF radiative transfer model RTTOV-9.3.

The 60L-SD is a sampled database of 60-level atmospheric profiles from ECMWF analysis. It summarises the characteristics of a diverse profile data set from the ECMWF 40-year re-analysis (ERA-40). This kind of databases is an interesting alternative to the use of radiosonde databanks like the TIGR (Thermodynamic Initial Guess Retrieval) database. The reason is that "the analyses of the NWP centres are homogeneous and cover all latitudes, longitudes and days of the year. Moreover they provide a whole set of variables consistent with each other for each profile, like layer temperature, water vapour, cloud cover and surface characteristics" contributed by Chevallier (2001). The sampling method used by Chevallier extracts temperature, specific humidity and ozone and the entire sampled database includes 13495 profiles. The 60L-SD database contains the same number of samples from seven subgroups differing by the total precipitable water vapour content of the profiles, as it can be observed in figure 1. This is the main reason why it is a suitable database to implement TPW and LPW algorithms. The file with the database profiles "60L_SD.dat" is freely available on the website of the NWP SAF http://www.nwpsaf.org. A direct link to the 60L_SD files is not provided because new users must register before accessing NWP SAF non-public web pages. Once a user has logged on to NWP SAF internal web page, the 60L_SD is available on the "RTTOV & profiles" deliverables link.

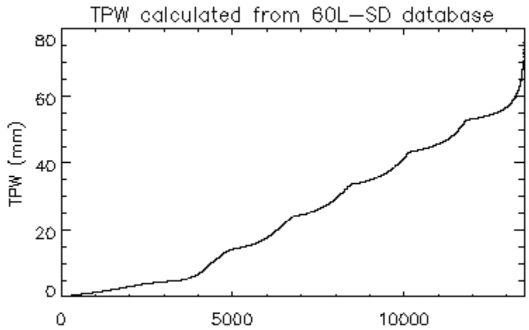

Figure 1: Distribution of the total precipitable water in the 60L-SD database.

The radiative transfer model RTTOV is also available on the above same user registered NWP SAF web page. The last RTTOV version, RTTOV-9.3 has been used in this paper to calculate the synthetic brightness temperatures. RTTOV-9.3 has coefficient files to calculate synthetic radiances for most of the IR and MW instruments on board operational meteorological satellites. In the case of the GMI instrument, the RTTOV coefficients file for the GMI is not yet available; but it is foreseen that it will be available in the near future, before the GPM launch. Therefore, this study is a preliminary and theoretical early work. To build a proxy of the GMI channels, similar channels present in other MW instruments as TMI on board TRMM and MHS on board METOP have been used.

Neural networks for TPW and LPW parameters with the PSEUDO-TMI and PSEUDO-GMI proxies have been trained and validated. The process can be divided into: training datasets generation, training of the neural networks and validation of the neural networks. Finally the application of the trained neural networks to a case study is made in the second half of this paper.

TRAINING OF TPW AND LPW NEURAL NETWORKS FOR PSEUDO-TMI AND PSEUDO-GMI PROXYS

TRAINING DATASETS GENERATION

The first task is to create training datasets. This task can be divided into two parts: a) to obtain the synthetic brightness temperature in clear air from the profiles of 60L_SD dataset b) to calculate the parameters of TPW and layer precipitable water directly from temperature and moisture profiles of 60L_SD dataset.

The task to obtain the synthetic brightness temperature has been divided into the following steps:
Step a.1) To read every profile (60 hybrid pressure levels) present on file "60L_SD.dat" and to interpolate it to the 43 RTTOV pressure levels. RTTOV-9.3 can operate with profiles in the vertical levels of the user, but due to the fact that RTTOV works internally by interpolation/extrapolation to 43 RTTOV pressure levels; it has been considered better to work directly at the 43 RTTOV pressure levels and to perform the interpolation of the 60 hybrids levels to the 43 RTTOV pressure levels.

Step a.2) To select sea profiles with latitude lower than 70º.

Step a.3) To obtain synthetic brightness temperature for TMI and MHS instruments using all sea profiles with latitude lower than 70º. RTTOV-9.3 can be considered as a library of radiative transfer model routines; routines to calculate radiance/brightness temperature, Jacobian, etc are available. An interface program between our data and the RTTOV library must be built; in the RTTOV-9.3 sources there are some sample programs to be used as a basis. The sea profiles are used as input to the FORTRAN program built to call RTTOV-9.3 forward routine (rttov_direct) from the 60L_SD. Here, the developed FORTRAN-90 program: reads every profile, builds a Fortran structure with temperature and specific humidity profiles, skin temperature, u and v 10 meter winds components, zenith angle, cloud fraction, cloud pressure level, etc and makes a call to the rttov_direct (forward model) routine. As zenith angle, the fixed value of 53º has been used. Since the aim of this paper is to investigate the training of the neural network in clear air pixel the cloudy parameters have been fixed to 0.0. Since emissivity values are fixed to 0 in the call to RTTOV and option *calcemis* is also fixed equal true (see RTTOV-9 user guide for details); then, RTTOV-9.3 calculates emissivity values using the FASTEM-3 model. As FASTEM-3 requires the surface wind-speed to be provided in the state vector, 10 meters wind is used as input to RTTOV-9.3.

In order to allow for a synthetic brightness temperature calculation for different satellites, RTTOV-9.3 has coefficients files for almost every instrument on IR and MW spectra domain. Due to the fact that the GMI RTTOV coefficient file is not yet available, the TMI (TRMM satellite) RTTOV coefficient and MHS (METOP satellite) RTTOV coefficients have been used here.

RTTOV forward routine provides as output one FORTRAN-90 structure with several fields; the field *bt_clear* has been used here to create the training dataset. *Bt_clear* field represents the synthetic brightness temperature obtained through integration of the radiative transfer equation from the pressure at surface to the top of atmosphere, assuming the pixel is free of clouds. In the case of TMI RTTOV-9.3 execution, *bt_clear* array provides the synthetic clear air BT for channels 10.65 H, 10.65V, 18.70H, 18.70V, 23.80, 36.5H, 36.5V, 89.0H and 89.0V GHz. In the case of MHS RTTOV-9.3 execution, *bt_clear* array provides the synthetic clear air BT for channels 89.05, 150.09, 183.4H, 183.4V and 190.4 GHz. In order to build a proxy for the GMI instrument, two options have been tested: ***pseudo-TMI proxy*** that uses only TMI bt_clear synthetic BTs for all TMI channles and ***pseudo-GMI proxy*** that uses TMI *bt_clear* channels plus MHS *bt_clear* at 150.096, 183.4H and 183.4V channels.

On the other hand, it is necessary to calculate the parameters of TPW and layer precipitable water directly from the temperature and moisture profiles of the 60L_SD dataset. This task has been made by directly calculating the TPW and Layer Precipitable Water parameters from the specific humidity profile interpolated to the 43 RTTOV pressure levels. To fix the limits of the layers, the same RTTOV pressure levels used in the NWCSAF/MSG at PGE07 Layer Precipitable Water product have been chosen. Then, in this first version of the GMI algorithm, the following layers have been considered:

1. TPW: Total precipitable water from the humidity retrieved profiles.
2. Precipitable water in three layers LPW from the humidity retrieved profiles:
    a. BL: Surface - 840 hPa,
    b. ML: 840 – 437 hPa,
    c. HL: 437 – TOP

As the limits of the layer are RTTOV pressure levels, it is not needed to make interpolation. Then, to calculate Layer Precipitable Water parameters is equivalent to “integrate” the precipitable water between the pressure levels limits. The sum of the Layer Precipitable Water parameters should be equal to the TPW.

Together with previous Layer Precipitable Water, two new layers have been tested. In previous convection studies, it has been identified that the ML layer could be too wide; new Layers called M1 (840 – 703 hPa) and M2 (703 – 437 hPa) that are the result of splitting the ML Layer at 703 hPa are tested in the NWCSAF/MSG project before being included in any future NWCSAF/MSG software package. Due to this fact, neural networks for M1 and M2 have also been trained and tested.

Once the TMI and MHS synthetic BT and TPW and LPW values are available, the training dataset can be generated. The format of SNNS version 4.2 has been selected to write the training dataset files. SNNS data files are ASCII files and they have a header component and a data component. The header defines the number of patterns that the file contains as well as the dimensionality of the input and target vectors. The data component of the pattern file is simply a listing of numbers (one line for every pattern) that represent the activations of the input and output units.

In neural network training, normalization of input and output is important. To write the normalized inputs for every channel in pseudo-TMI or pseudo-GMI model the following process is undertaken: i) the mean of synthetic BTs for 60L_SD is calculated, ii) to this a random noise of ±0.5 K is added, iii) the formula (bt-mean(bt)+31)/62 has been applied.

To write the normalized output, the TPW or LPW values are divided by a predetermined value. These predetermined values are: 70 for TPW, 45 for ML, 35 for BL, 20 for M1 and M2 and 4.5 for HL.

In neural network training it is usual to split the available data between the validation and the training dataset. In our case, 9 out of every 10 patterns have been randomly extracted for the training dataset and 1 out of every 10 patterns has been randomly extracted for the validation dataset.

TPW and LPW NEURAL NETWORK TRAINING

Once the training datasets have been generated, the Stuttgart Neural Network Simulator (SNNS) version 4.2 is used to train the neural Networks. SNNS program is freely available in http://www.ra.cs.uni-tuebingen.de/SNNS/.

For every parameter (TPW or one of the 5 LPW) of pseudo-TMI or pseudo-GMI proxies, the training has been made using the SNNS version 4.2. Between all neural network types the Multilayer Perceptron has been selected as being the most adequate. The topology selected has one hidden layer with 12 neurons. Then, the topology of every neural network has 9 (in the pseudo-TMI) or 12 (in the pseudo-GMI) neurons in the input layer, 12 neurons in the hidden layer and 1 neuron (TPW or one of the 5 LPW) in the output layer.

To speed up the training process, instead of using the interactive SNNS GUI, several scripts have been built in order to use the “batchman” tool that allows the training in batch.

VALIDATION of the NEURAL NETWORK WITH 60L_SD DATASET

In order to assess the performance of the neural networks the statistical parameters of the neural networks have been calculated over the validation and training datasets. In figures 2 and 3, the scatter plots of the TPW and LPW for pseudo-TMI and pseudo-GMI proxies are shown. The performance of the neural networks is excellent and the low spread suggests that this kind of fast and NWP independent algorithm could be developed for the GMI instrument. Comparison of pseudo-GMI performance with pseudo-TMI shows the added value of the extra channels and that in every case the extra channels improve the statistical coefficients and reduces the spread. The greatest improvement and reduction of the spread can be seen for the high layer (HL).

On the x axis the pattern (the TPW or LPW value directly calculated from the 43 RTTOV level 60L-SD specific humidity profile) is always represented versus the denormalized value of the neural network output on the y axis. To denormalize the output of the neural network (in the range [0,1]), the outputs are multiplied by the TPW or LPW normalization value. The outputs from the validation dataset, which have not been used in the training of the neural network, are represented by red dots.

CASE STUDY: 17th September 2006 at 12 UTC

In order to test the spatial performance of the trained neural networks for TPW and LPW, they were applied to synthetic brightness temperatures from one ECMWF GRIB analysis. On the selected date, 17th September 2006 at 12 UTC, the circular pattern of the hurricanes Helene and Gordon can be shown in the middle of the Atlantic Ocean.

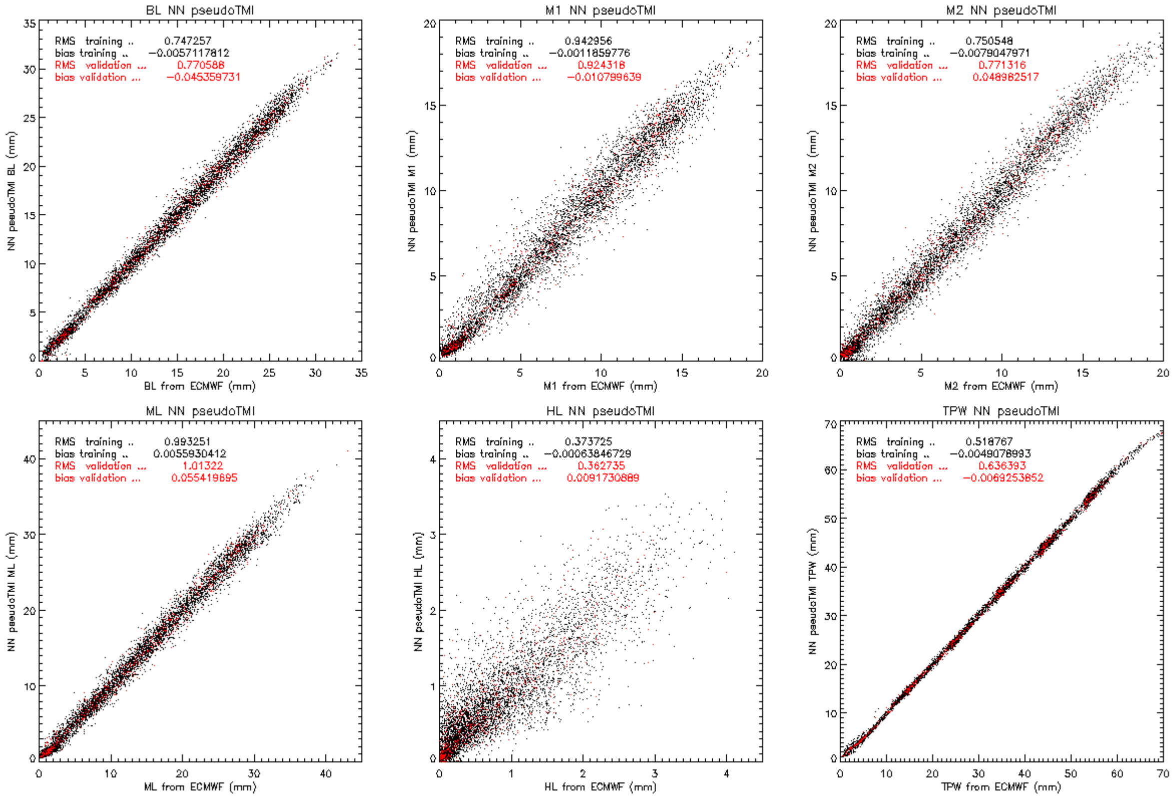

Figure 2. TPW and LPW scatter plots for the pseudo-TMI model (use only of TMI channels as Proxy for GMI). From left to right and top to bottom BL, M1, M2, ML, HL and TPW scatter plots. Red dots are from validation datasets, black dots are from training dataset.

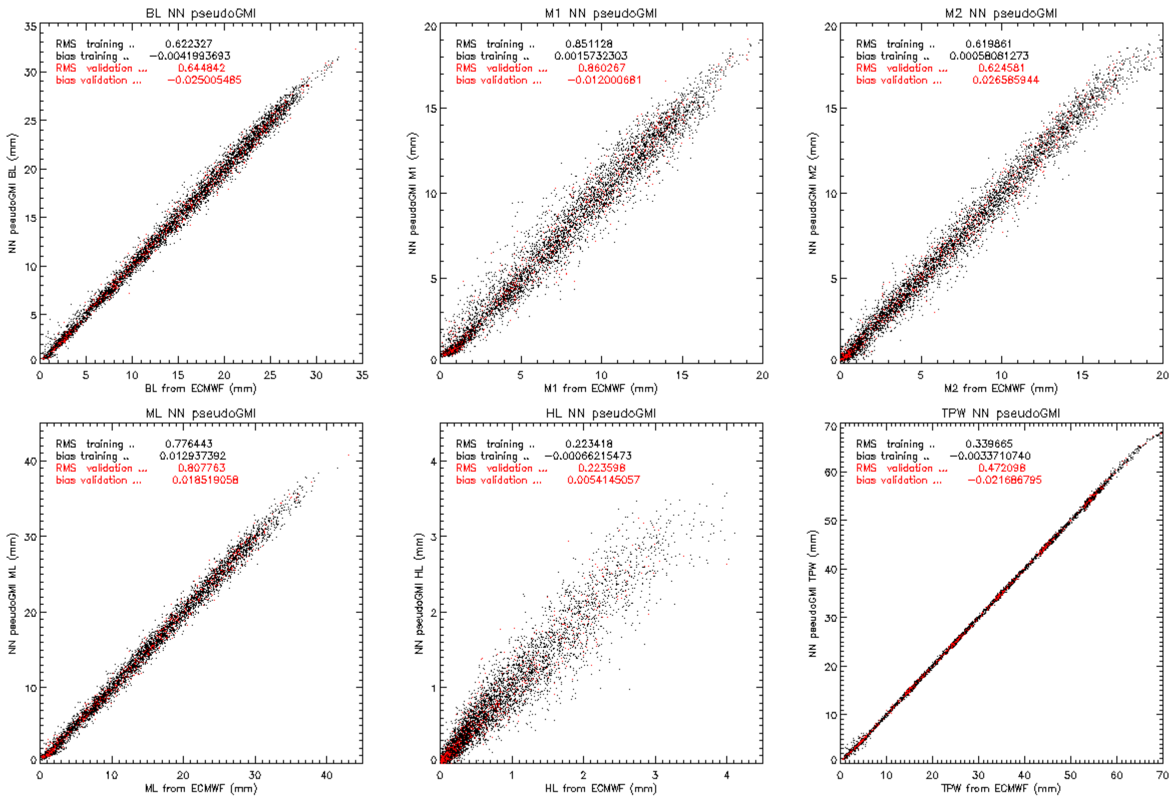

Figure 3. TPW and LPW scatter plots for the pseudo-GMI model (TMI channels and 3 channels from MHS as proxy for GMI). From left to right and top to bottom BL, M1, M2, ML, HL and TPW scatter plots. Red dots are from validation datasets, black dots are from the training dataset.

One FORTRAN interface program that uses as input ECMWF GRIB files and makes the call to RTTOV forward routine has been used. This program was developed for simulation and validation of METEOSAT Second Generation algorithms but it has been used here after one slight modification (it has been added as input the 10 meters wind). The configuration file has been also slightly modified: the zenith angle has been fixed to 53º and the name of the RTTOV coefficients are now TMI or METOP. The program allows to archive on binary file the temperature and specific humidity profiles interpolated to the 43 RTTOV pressure levels and the surface pressure. This binary file has been used to calculate TPW and LPW directly from the ECMWF analysis; see figure 5.

The 2D pseudo-TMI and pseudo-GMI proxies have been built from the 2D synthetic clear air brightness temperatures arrays after TMI RTTOV-9.3 execution and MHS RTTOV-9.3 execution. Pseudo-TMI proxy is built with the nine TMI channels (10.65 H, 10.65V, 18.70H, 18.70V, 23.80, 36.5H, 36.5V, 89.0H and 89.0V GHz); see figure 4. Pseudo-GMI proxy is built with the nine TMI channels and the 2nd, 3rd and 4th outputs from MHS. The two proxies are ideal clear air simulations.

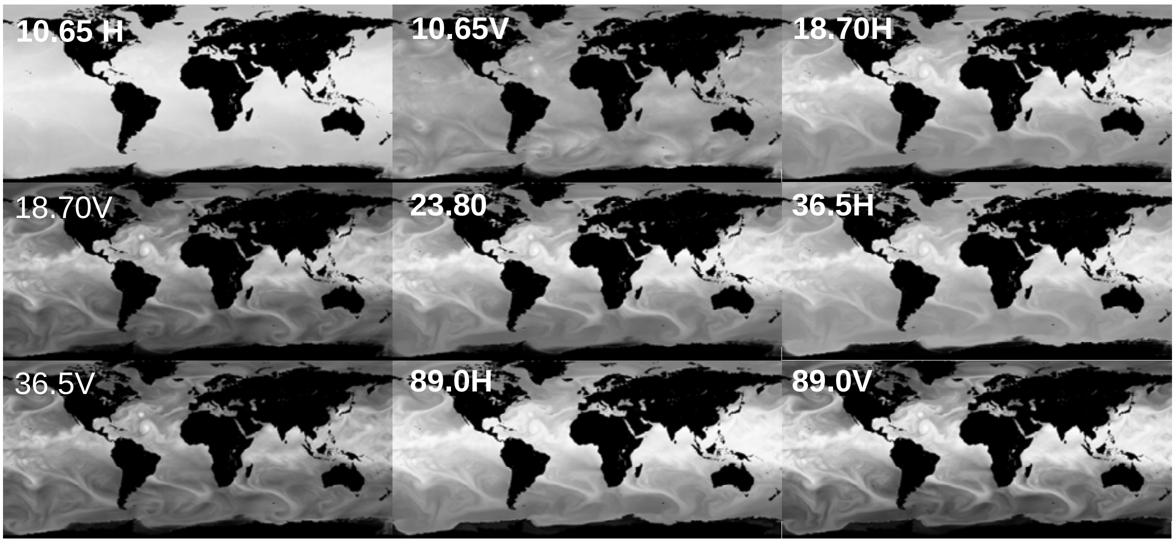

Figure 4. Pseudo-TMI proxy for 17th September 2006 at 12 UTC. It is built with the nine clear air BT from TMI channels.

In order to apply the neural networks, the pseudo-TMI and pseudo-GMI 2D brightness temperature are normalized with the same mean values used in the generation of the training datasets. After applying the neural networks the outputs are denormalized by multiplying the outputs by the TPW or LPW normalization value. TPW and LPW for pseudo-TMI neural networks are shown in figure 6. TPW and LPW for pseudo-GMI neural networks are shown in the figure 7. The ECMWF land-sea mask has been used to filter out the land pixels; to calculate the statistics values only sea points (160590 sea points) have been used. The statistical values for pseudo-TMI and pseudo-GMI for 17th September 2006 at 12 UTC have been summarised in Table 1 and Table 2. As in 60L-SD validation, the comparison of pseudo-GMI performance with pseudo-TMI one shows that pseudo-GMI is better, and the largest improvement is obtained in HL Layer. This fact can be also confirmed after HL images comparison.

Ps-TMI@Sea points	TPW	BL	ML	M1	M2	HL
RMSE (mm)	0.32	0.57	0.78	0.84	0.60	0.28
BIAS (mm)	0.02	-0.17	0.26	0.05	0.15	-0.04
Correlation	0.99	0.99	0.99	0.97	0.98	0.90

Table 1. Pseudo-TMI statistical values for 17th September 2006 at 12 UTC.

Ps-GMI@Sea points	TPW	BL	ML	M1	M2	HL
RMSE (mm)	0.21	0.53	0.56	0.76	0.50	0.13
BIAS (mm)	-0.02	-0.17	0.14	0.08	0.03	-0.02
Correlation	0.99	0.99	0.99	0.98	0.99	0.97

Table 2. Pseudo-GMI statistical values for 17th September 2006 at 12 UTC.

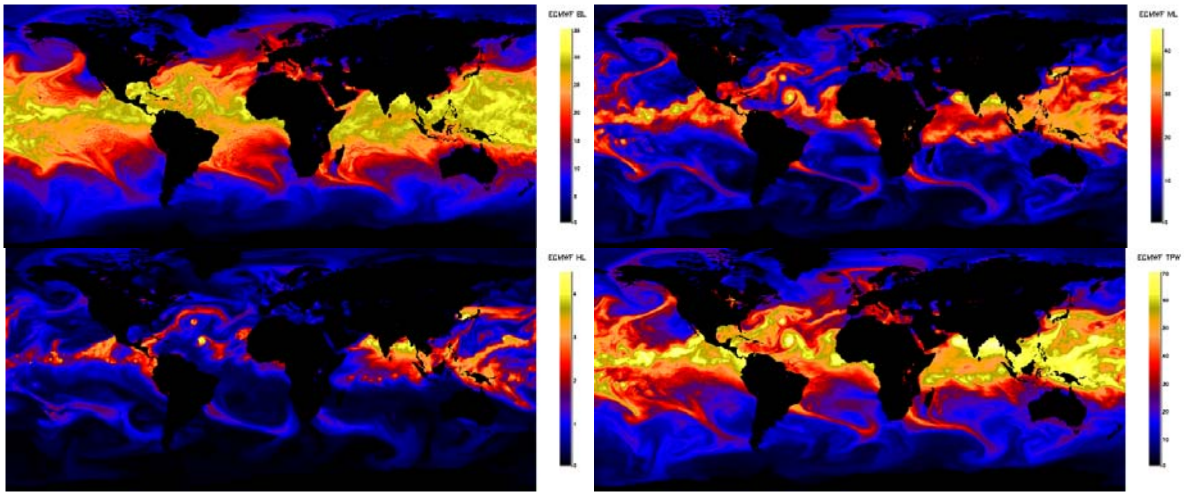

Figure 5. TPW and LPW directly calculated from the ECMWF GRIB analysis. 17th September 2006 at 12 UTC.

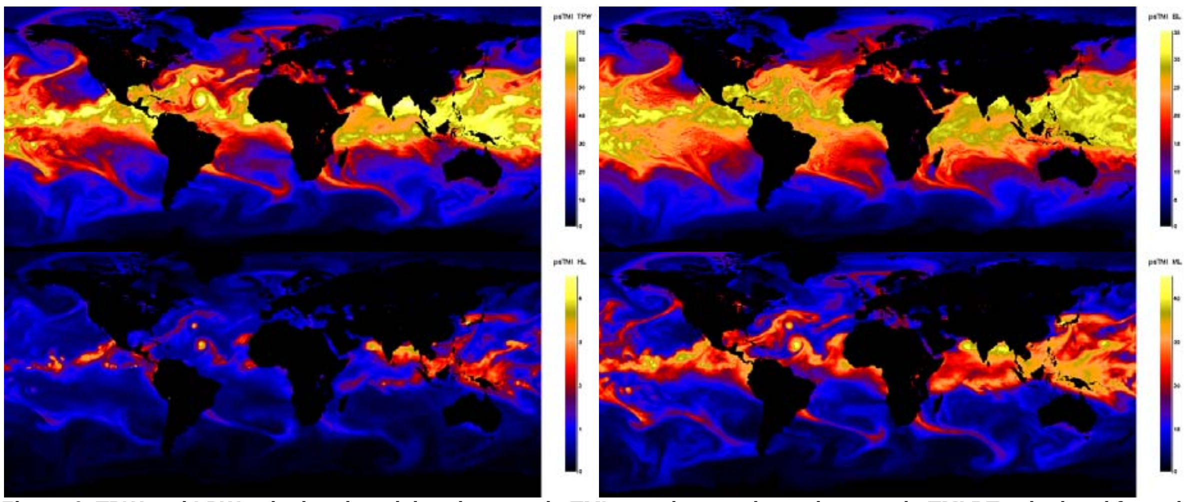

Figure 6. TPW and LPW calculated applying the pseudo TMI neural networks to the pseudo TMI BT calculated from the ECMWF grib analysis. 17th September 2006 at 12 UTC.

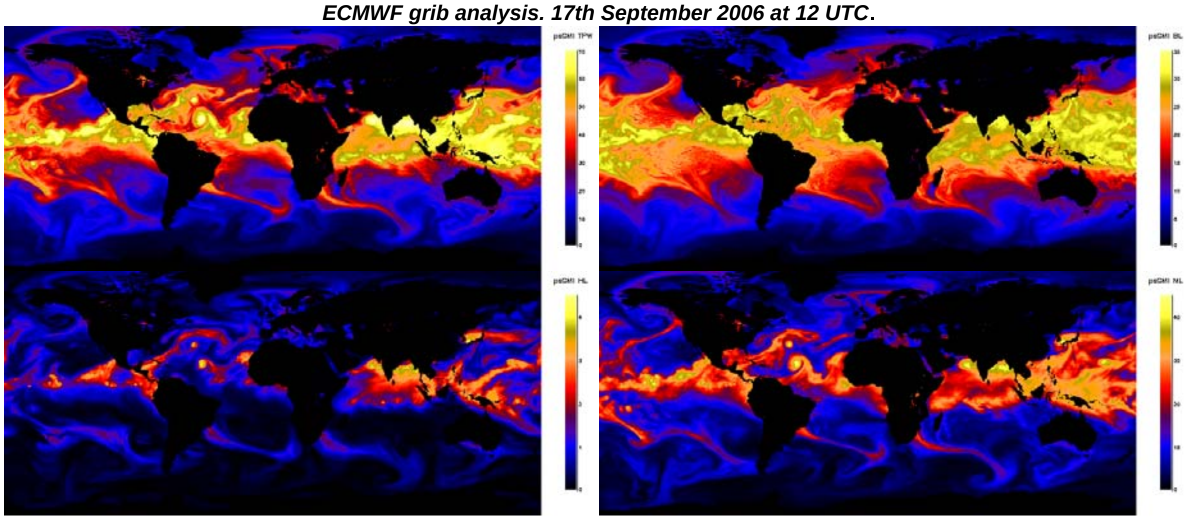

Figure 7. LPWs and TPW calculated applying the pseudo GMI neural network to the pseudo GMI BT calculated from the ECMWF grib analysis. 17th September 2006 at 12 UTC.

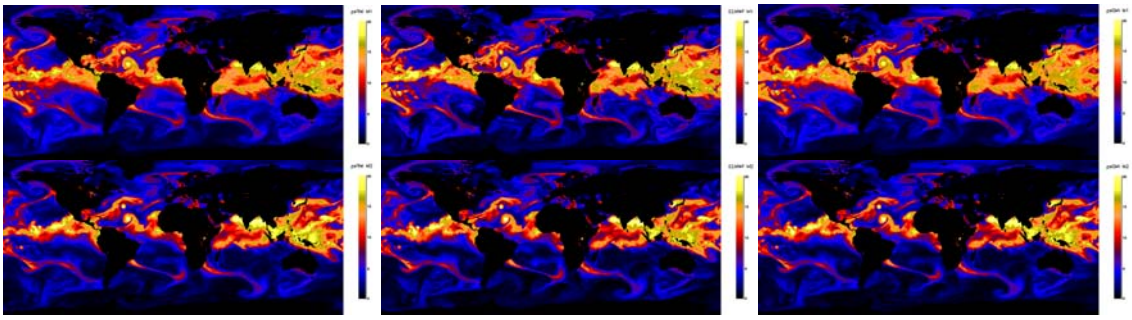

***Figure 8. New M1 Layer (top) and M2 Layer (bottom) comparison. Calculated from the ECMWF profiles (center); using pseudo-TMI (left) and from pseudo-GMI neural network (right). 17th September 2006 at 12 UTC*.**

CONCLUSIONS

Neural networks for TPW and LPW have been trained from the GMI instrument proxies. The performance of the neural networks is excellent and the possibility to obtain, at least on clear air pixels, the vertical distribution of the precipitable water from MW data has been demonstrated. The advantages of this algorithm would be to obtain a fast algorithm independent of NWP.

The process followed in this paper can be repeated with slight modifications to train neural networks for other parameters, such as LWP, IWP, etc. The work described in this paper is only at an early stage. In order to get one operational algorithm additional issues (such as bias between measured and synthetic brightness temperatures removal, generalization of the algorithm to cloudy and rainy brightness temperatures issue, etc.) should be solved.

ACKNOWLEDGEMENTS

This study was carried out within the Spanish CICYT activities framework (Project CGL2006-03611).

REFERENCES

Eyre, J.R., (1991). *A fast radiative transfer model for satellite sounding systems*. ECMWF Tech. Memo. 176 (28 pp.).

Chevallier, F., 2001. *Sampled databases of 60-level atmospheric profiles from the ECMWF analysis.* Research Report No. 4, EUMETSAT/ECMWF SAF Program.

Gairola, R. M. et al, 2006. *Multiparameter microwave retrieval algorithms: performance of neural networks*. CURRENT SCIENCE, VOL. 91, NO. 10, 25

C. Mallet et al., 2002. *Determination of integrated cloud liquid water path and total precipitable water from SSM/I data using a neural network algorithm.* International Journal Remote Sensing, 2002, vol. 23, no. 4, 661–674

Martinez, M.A. et al., 2007. *Application of LPW and SAI SAFNWC/MSG satellite products in pre-convective environments.* Atmospheric Research 83 (2007) 366–379

Saunders, R., Matricardi, M., Brunel, P., 1999. *An improved fast radiative transfer model for assimilation of satellite radiance observation.* Quart. J. Roy. Meteor. Soc., 125, 1407–1425.

D. Singh et al., 2005. *A neural network based algorithm for the retrieval of TPW from AMSU measurements.* http://cimss.ssec.wisc.edu/itwg/itsc/itsc14/presentations/session2/2_1_singh.pdf