

The generation of RTTOV  
regression coefficients for IASI and  
AIRS using a new profile training  
set and a new line-by-line  
database

Marco Matricardi

Research Department

May 2008

Final Report  
EUMETSAT Contract No. EUM/CO/06/1504/PS

This paper has not been published and should be regarded as an Internal Report from ECMWF.  
Permission to quote from it should be obtained from the ECMWF.



**Series: ECMWF Technical Memoranda**

A full list of ECMWF Publications can be found on our web site under:

<http://www.ecmwf.int/publications/>

Contact: [library@ecmwf.int](mailto:library@ecmwf.int)

**© Copyright 2008**

European Centre for Medium Range Weather Forecasts  
Shinfield Park, Reading, Berkshire RG2 9AX, England

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. This publication is not to be reprinted or translated in whole or in part without the written permission of the Director. Appropriate non-commercial use will normally be granted under the condition that reference is made to ECMWF.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error, omission and for loss or damage arising from its use.

## 1. Introduction

Several versions of the RTIASI (Matricardi and Saunders, 1999) fast radiative transfer model have developed over the years for the exploitation of the IASI datasets. The most recent version of the code, RTIASI-5 (Matricardi, 2005) includes, amongst other features, variable trace gases, solar radiation in the short wave region of the spectrum and a parameterization of multiple scattering for different types of clouds and aerosols. More recently, all the features of RTIASI have been implemented into the RTTOV fast radiative transfer model and all the development work shifted towards the latter model. The core of RTTOV is a fast model of the transmittances of the atmospheric gases that is generated from a database of accurate line-by-line (LBL) transmittances computed for a set of diverse atmospheric profiles. All the LBL databases generated at ECMWF are based on the profile set described in Matricardi and Saunders (1999) using the GENLN2 (Edwards, 1994) LBL model although regression coefficients for a number of sensors are available for RTTOV that are based on the 52 profile set described in Chevallier (2003) using the kCARTA (Strow et al. 1998) forward model. The GENLN2 model was adopted for use at ECMWF more than ten years ago and since then no major version of the code has been released. Since the long term maintenance of GENLN2 has become an issue (there is no apparent commitment to further develop the code) we have considered the possibility of a different choice of line-by-line model for the training of RTTOV. To this end we have recently carried out a study (Matricardi, 2007) where the accuracy and computational efficiency of a number of line-by-line models has been assessed. As a result of the study we have adopted LBLRTM (Clough et al. 1992) as the LBL model to be used for the training of RTTOV. Among the main advantages of LBLRTM there are its computational efficiency and the fact that the model is updated on a timely basis (for instance, the latest version of LBLRTM includes CO<sub>2</sub> v2 and v3 band line mixing). The study also suggested that an optimal database of molecular parameters could be envisaged by blending line data from the HITRAN (Rothman et al. 2005) and GEISA (Husson et al. 2005) databases. The quality of the RTTOV forward radiances is of paramount importance in the assimilation of satellite data into a NWP system. Since these radiances depend crucially on the quality of the spectroscopy used in the LBL computations, we have generated a new database of LBL transmittances to be used for the generation of RTTOV's IASI and AIRS regression coefficients using the latest release of the LBLRTM model. For the generation of the database we have selected a new training set of 83 diverse atmospheric profiles and, as mentioned earlier, a compilation of molecular parameters based on the most recent line data available in the HITRAN and GEISA compilations. In section 2 of the paper we give a description of the LBLRTM model. In section 3 we describe the selection of a new training set of atmospheric profiles of temperature, water vapor, ozone, carbon dioxide, methane, carbon monoxide and nitrogen dioxide. In section 4 we describe the generation of the transmittance database whereas in section 5 we give a description of the optimal selection we made of the predictors used in the fast transmittance model to avoid unphysical features in the trace gas Jacobians and, finally, in section 5 we assess the accuracy of the fast transmittance model by comparing RTTOV radiances to LBL equivalents.

## 2. The line-by-line model

The database described in this paper was calculated using version 11.1 of the LBLRTM (LBLRTM\_v\_11.1) line-by-line atmospheric transmittance and radiance model. The LBLRTM\_v\_11.1 line-by-line model has been developed at AER and is derived from the Fast Atmospheric Signature Code (FASCODE) (Clough et al. 1981). To describe the effects of pressure and Doppler line broadening the Voigt line shape is used at all atmospheric levels with an algorithm based on a linear combination of approximating functions. LBLRTM\_v\_11.1

incorporates the self- and foreign-broadened water vapor continuum model MT\_CKD\_2.1 as well as continua for carbon dioxide and for the collision induced bands of oxygen at 1600 cm<sup>-1</sup> (Thibault et. al 1996) and nitrogen at 2350 cm<sup>-1</sup> (Lafferty et al. 1996). Since RTTOV uses a dedicated model for the far wing water vapor continuum, we have not included the water continuum contribution in the LBLRTM\_v\_11.1 computations. In line with the methodology described in Matricardi (2003) we have instead created a separate database of monochromatic water continuum transmittances and then generated regression coefficients by linear regression of the model transmittances versus the predictor values calculated from the profile variables. The regression coefficients are then used by RTTOV to predict the water continuum transmittances. Since we think it is a useful feature to have different continuum models available for RTTOV, in addition to the regression coefficients based on the MT\_CKD\_2.1 model we have also generated coefficients based on an earlier version, CKD\_2.4, of the continuum model. Although the recommendation for the RTTOV user is to use the latest available version of the continuum (MT\_CKD\_2.1), the availability of an alternative version of the continuum would allow, for instance, carrying out a validation study where the comparison of synthetic and measured satellite radiances could reveal deficiencies in either the continuum models.

For the water vapour continuum model adopted in LBLRTM\_v\_11.1 (Clough et al. 1989) the total continuum contribution to the absorption coefficient,  $k_c(\nu)$ , at wavenumber  $\nu$  can be written in the form

$$k_c(\nu) = \nu \tanh\left(\frac{h\nu}{2kT}\right) \left(\frac{T_o}{T}\right) \left(\frac{p_s}{p_o} C_s^o(\nu, T) + \frac{(p - p_s)}{p_o} C_f^o(\nu, T)\right) \quad (1)$$

where  $T$  is temperature,  $p$  is pressure,  $p_s$  is water vapour partial pressure,  $T_o$  and  $p_o$  are reference temperature (296K) and pressure (1013 hPa) values and  $h$ ,  $c$  and  $k$  are the Planck constant, speed of light and Boltzmann constant respectively.  $C_s^o$  and  $C_f^o$  are the self-broadening coefficients and the foreign-broadening coefficients defined at the reference temperature and pressure. To illustrate the differences between the CKD\_2.4 and MTK\_CD\_2.1 models we have plotted  $C_s^o$  in Figure 1 and  $C_f^o$  in Figure 2.

Water vapour continuum plays an essential role in the absorption process. In the 10 μm window region between the water vapor 6.3 μm band and the water vapor rotation band the continuum absorption is stronger than the line absorption. The dominant source of continuum in this region is the self broadening (H<sub>2</sub>O-H<sub>2</sub>O collisions) continuum whereas the dominant source of continuum in the 6.3 μm water vapor band is the foreign broadening (H<sub>2</sub>O-N<sub>2</sub> collisions) continuum. The foreign broadening continuum is particularly important between 1350 and 2100 cm<sup>-1</sup>. From Figure 1 it can be seen that in the window region, the self broadening coefficients for CKD\_2.4 are larger than the MT\_CKD\_2.1 coefficients for wave numbers less than 950 cm<sup>-1</sup> and smaller for wave number greater than 950 cm<sup>-1</sup>. Figure 2 shows that foreign broadening coefficients for CKD-2.4 are smaller than MTK\_CKD.2.1 coefficients between 1400 and 1900 cm<sup>-1</sup> and between 2100 and 2200 cm<sup>-1</sup> and larger between 1900 and 2100 cm<sup>-1</sup>.

For the LBLRTM\_v\_11.1 calculations line coupling parameters for the CO<sub>2</sub> P/Q/R branches are provided for the first and second isotopes (Niro et al., 2005). These line coupling/mixing coefficients were created by Niro et al. using lines from HITRAN\_2000. It should be noted that since the CO<sub>2</sub> continuum and the *CHI* factor depend on the line coupling, to obtain accurate results the Niro et al. line coupling must be used in LBLRTM\_v\_11.1 using

the line parameters on which the line mixing coefficients are based. It is then mandatory that the CO<sub>2</sub> line data used in the LBLRTM\_11.1 calculations be those contained in the aer\_v\_2.0/2.1 line parameter database.

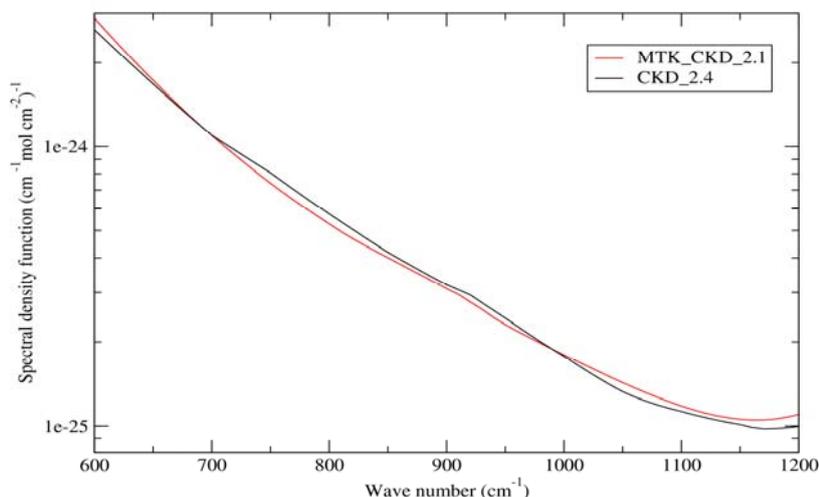


Figure 1: Spectral density function for the water vapor self broadening coefficients at 296K and 1013hPa.

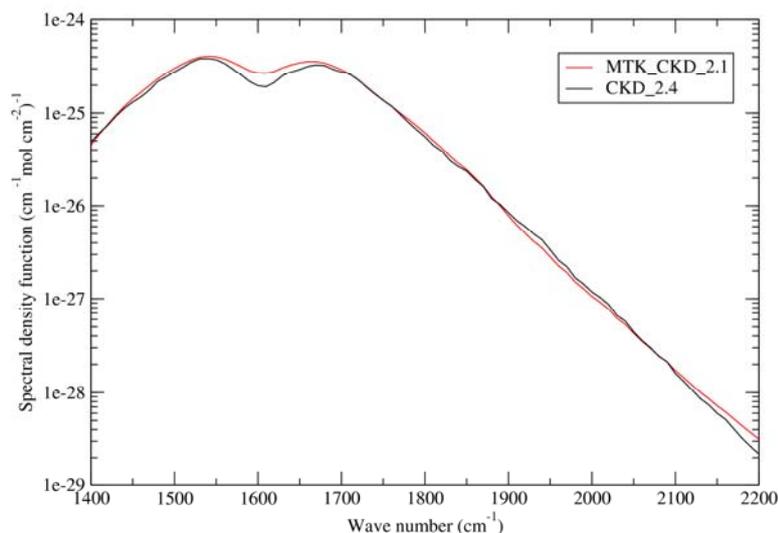


Figure 2: Spectral density function for the water vapor foreign broadening coefficients at 296K and 1013hPa.

LBLRTM\_v\_11.1 can use temperature dependent cross section data to model the absorption due to heavy molecules. The pressure dependence of the cross sections is treated by performing a convolution of the cross section spectrum with an appropriate Lorentz function. LBLRTM\_v\_11.1 has the capability to perform radiative transfer calculations for down looking scenes with a Lambertian surface by obtaining the down welling flux from a radiance calculation at the diffusivity angle of 53.21 degrees. An algorithm (linear in tau) is used for the treatment of the variations of the Planck function within a vertically inhomogeneous atmosphere. Finally, computations can be performed for atmospheric layers which are not in local thermodynamical equilibrium, a solar term can be included and a recently added feature is the capability of performing the computation of analytic derivatives/jacobians.

As discussed previously, the Voigt line shape incorporates the effects of both Doppler (velocity) and collisional line broadening; collisional broadening dominates from the surface to a certain altitude and velocity broadening dominates above that altitude with an intervening transition region. As a consequence, for a given temperature profile, one could devise an optimal sampling of the spectrum that is dependent on wave number and layer pressure, i.e. a sampling interval larger at the surface and smaller at the top of the atmosphere. In LBLRTM\_v\_11.1 a spectral sampling is used that is optimal for the atmospheric layer and the spectral regime under consideration. To achieve a monochromatic accuracy of better than 0.5% this sampling interval has been chosen to be 1/4 of the line halfwidth based on an analysis of the errors in the reconstruction of the Lorentz line as a function of the sampling interval using the four point interpolation scheme utilized in LBLRTM\_v\_11.1. The sampling interval  $DV$  is defined as:

$$DV = \bar{\alpha}_v / SAMPLE \quad (2)$$

where  $\bar{\alpha}_v$  is the average value of the Voigt halfwidth for the layer and  $SAMPLE$  is set by default to 4 as stated previously. The average value of the Voigt halfwidth  $\bar{\alpha}_v$  can be written as:

$$\bar{\alpha}_v = \frac{\bar{\alpha}_L}{2} + \left[ \left( \frac{\bar{\alpha}_L}{2} \right)^2 + \bar{\alpha}_D^2 \right]^{1/2} \quad (3)$$

where  $\bar{\alpha}_L$  and  $\bar{\alpha}_D$  are the average layer values of the Lorentz and Doppler halfwidths respectively. The mean Doppler halfwidth can be written in the form:

$$\bar{\alpha}_D = \frac{\bar{\nu}}{c} + \left[ 2 \ln 2 \frac{kT}{\bar{M} / N_o} \right]^{1/2} \quad (4)$$

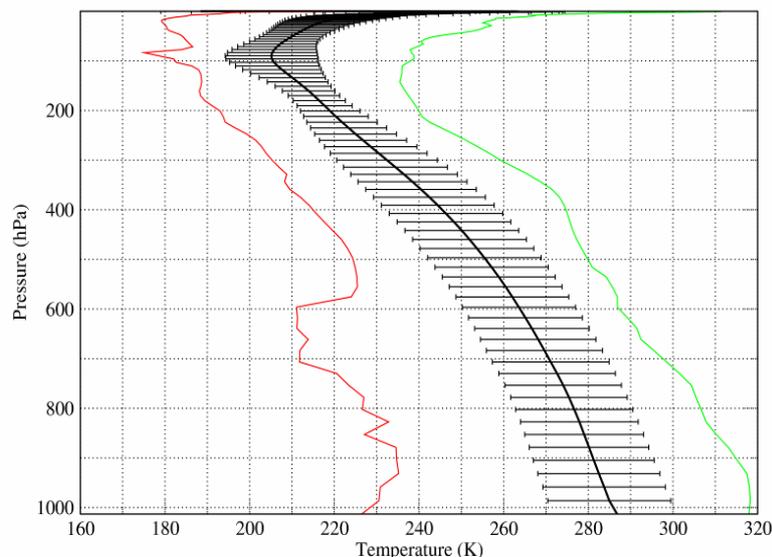
where  $\bar{\nu}$  is the average of the beginning and ending wave number values for the problem,  $N_o$  is the Avogadro's number and  $\bar{M}$  is a representative gram molecular weight for the constituents in the path. Since the sampling interval is based on the average of the beginning and ending frequencies, the accuracy of the code depends on the spectral domain used in the calculations. For this reason LBLRTM\_v\_11.1 has been structured in such a way that the maximum value of  $\bar{\nu}$  is set so 2000  $\text{cm}^{-1}$ , a trade off value with respect to core size, disk space and computational time. The implication of this choice is that it is not possible to perform a single run computation that covers the whole near infrared spectrum (e.g. 600 to 3000  $\text{cm}^{-1}$ ). Our approach is to perform multiple runs using a shell script with appropriate naming conventions. It should be stressed that the optimal sampling used in LBLRTM\_v\_11.1 makes the code significantly more computationally efficient than GENLN2 where the spectral sampling is defined by a fixed interval.

### 3. Diverse Profile Dataset

For each gas allowed to vary, the profiles used to compute the database of line-by-line transmittances are chosen to represent the range of variations in temperature and absorber amount found in the real atmosphere. The transmittances computed for the diverse profiles become data points in the regression. To derive the water vapor, ozone and fixed gases fast transmittance coefficients we have developed a new training set of 83 profiles selected from the database described in Chevallier et al. (2006). This database was sampled from a large profile dataset

containing 121,462,560 profiles generated using the experimental suite (cycle 30R2) of the ECMWF forecasting system. Due to the sampling strategy adopted by Chevallier et al. (i.e. the sampling of the different variables is performed separately), three different databases are available for temperature, water vapor and ozone, each containing 5000 profiles. A feature of all the cycles from cycle 30R2 onward is a vertical discretization of the atmosphere into a grid of 91 pressure levels. The spacing of the grid follows the orography of the terrain and while the top level is fixed at 0.01 hPa, the other levels are assigned a pressure value that depends on the value of the surface pressure.

During the course of the study we examined the profiles in the database and found a number of anomalies in the form of large unphysical oscillations of the ozone values in the troposphere and anomalously large ozone values in the stratosphere. Anomalous ozone profiles were present in all three databases. A detailed investigation subsequently revealed that the large oscillations in the troposphere were the result of the correlation between vorticity and ozone introduced by mistake (and lately corrected) in the experimental suite whereas the anomalies in the stratosphere could be considered an intrinsic (albeit undesirable) feature of the ozone assimilation system. To correct the anomalies in the tropospheric ozone, a new database was generated (Di Michele, ECMWF, personal communication) using the operational suite of the ECMWF forecasting system (cycle 32R2) sampling the profiles during the period July 2006-June 2007. This effort resulted in a new database made of 14,586 profiles. Since the anomalous large values in the stratosphere were still a feature of the ozone profiles we further thinned the database by excluding all the ozone profiles that in the stratosphere exceeded a threshold we set based on climatology. This brought down the total number of profiles in the database to 12,564 (4,680 profiles in the temperature dataset, 5,171 profiles in the humidity dataset and 2,713 profiles in the ozone dataset). The statistical characteristics of the profiles in the database are shown in figures 3, 4, and 5. Data are shown for the total dataset of 12,564 profiles since the selection of the training set, as described later, is based on the global ensemble of profiles. Note that the maximum value of the surface pressure shown in figures 3, 4, and 5 has been restricted to 1,020 hPa. In fact, although profiles with a lower surface pressure are present in the database, the number of profiles is too small to compute any meaningful statistics.



*Figure 3: Statistics of the temperature profiles in the 91 level sampled database. The minimum and maximum values are shown in red and green respectively. The thick black curve is the mean value and the error bars have a width of twice the standard deviation.*

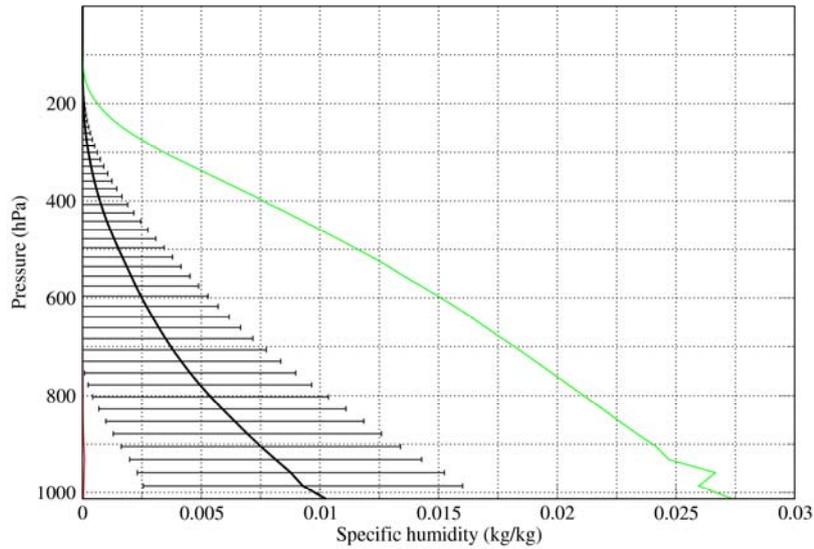


Figure 4: Statistics of the water vapor profiles in the 91 level sampled database. The minimum and maximum values are shown in red and green respectively. The thick black curve is the mean value and the error bars have a width of twice the standard deviation.

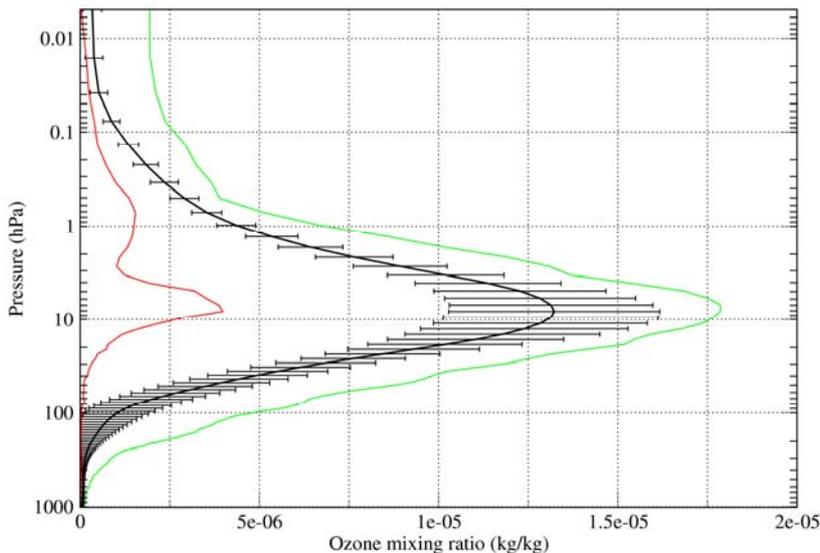


Figure 5: Statistics of the ozone profiles in the 91 level sampled database. The minimum and maximum values are shown in red and green respectively. The thick black curve is the mean value and the error bars have a width of twice the standard deviation.

As suggested in Chevallier (2002) the reduction of the number of profiles to a size that is manageable for line-by-line computations could be achieved by randomly sampling the profiles since this would not change the distribution of the profiles contained in each of the original datasets. An obvious feature of random sampling is that extreme values are very difficult to be selected. Although this can be corrected by manual intervention, we think that additional considerations should guide the selection of the training profiles. In fact, it is important that the profiles do represent the whole range of variation of the atmospheric temperature and of the atmospheric

constituents but it is also important that the profiles are distributed in a way that the regression at each pressure level be well constrained. In particular, to optimize the fit of the line-by-line optical depths to the curve defined by the predictors used in the regression, it would be desirable to have the training profiles at each pressure level distributed as uniformly as possible across the range covered by the profiles at that level. This is a difficult task to achieve if a strong thermodynamical coupling exists between any of the variables, as for instance in the case of humidity and temperature or ozone and temperature. To illustrate this point we have plotted in Figure 6 the histogram of frequency by bin number and pressure level of the temperature profiles associated to the water vapor profiles in the water vapor dataset. The histogram is obtained by dividing the temperature range into a number of bins and computing the number of profiles in each bin. By following the same procedure, in Figure 7 we plot the histogram of frequency by bin number of the water vapor profiles associated to the temperature profiles in the temperature dataset. It can be seen that when water vapor is the sampled variable a disproportionate number of temperature profiles accumulates in a ridge-like region oscillating between the extremes of the range whereas when temperature is the sampled variable, in the important tropospheric region a very large number of water vapor profiles accumulates in the dry end of the range. This is in net contrast with the situation shown in figures 8 and 9 where we have plotted the histogram of the temperature and water vapor profiles taken from the datasets where temperature and humidity are the respective sampled variables. Incidentally, features similar to those observed in figures 6 and 7 can also be seen in the distribution of temperature and water vapor profiles taken from the ozone dataset. We expect that all these features will be in the end reproduced in any pure random selection of the training profiles. In particular, the disproportionate number of dry water vapor training profiles is something that we want, as far as possible, to avoid.

Although we think it is unavoidable that any strategy adopted for the sub-sampling of the datasets will retain, to some extent, these features, nevertheless, under these restrictions, we think it is still possible to make a selection of the training profiles that would not result in too many dry water vapor profiles and too few profiles covering the moister regions of the range. To this end we have devised a strategy that involves a constrained random selection of the profiles. Prior to the selection we have interpolated all the profiles to the grid of 101 vertical pressure levels tabulated in Table 1. During the interpolation process, if the surface pressure was found to be less than 1100 hPa the temperature profile was extrapolated to the surface assuming a constant value whereas the water vapor profile was extrapolated assuming adiabatic expansion. All the profiles were then checked for supersaturation and corrected if necessary.

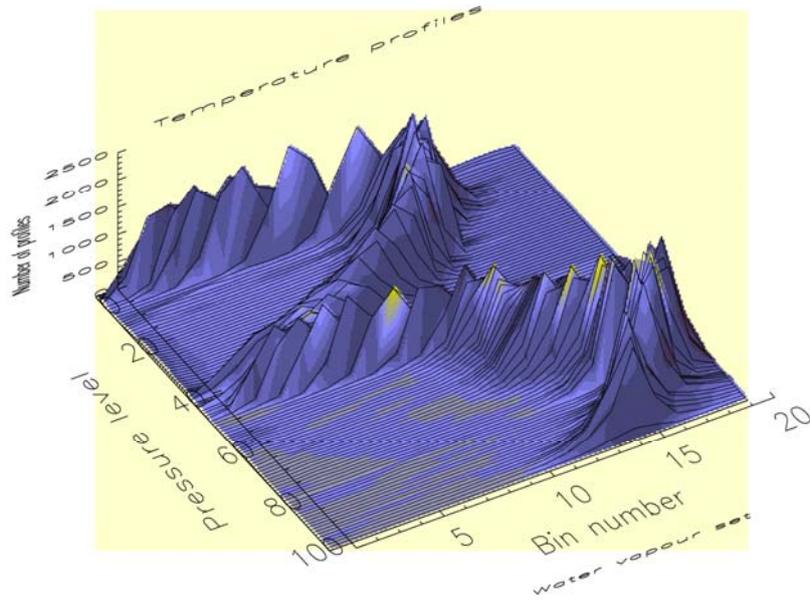


Figure 6: Histogram of frequency by bin number and pressure level for the temperature profiles in the water vapor dataset.

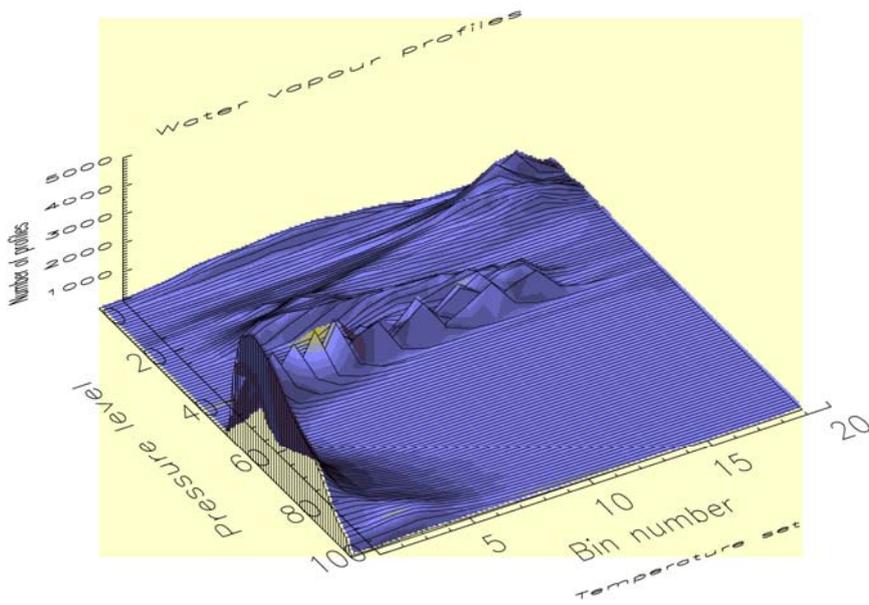


Figure 7: Histogram of frequency by bin number and pressure level for the water vapor profiles in the temperature dataset.

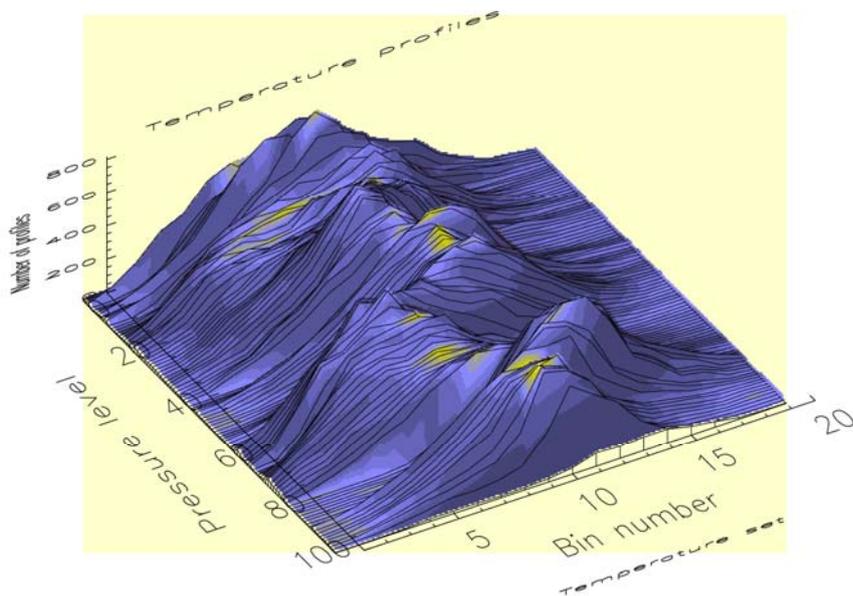


Figure 8: Histogram of frequency by bin number and pressure level for the temperature profiles in the temperature dataset.

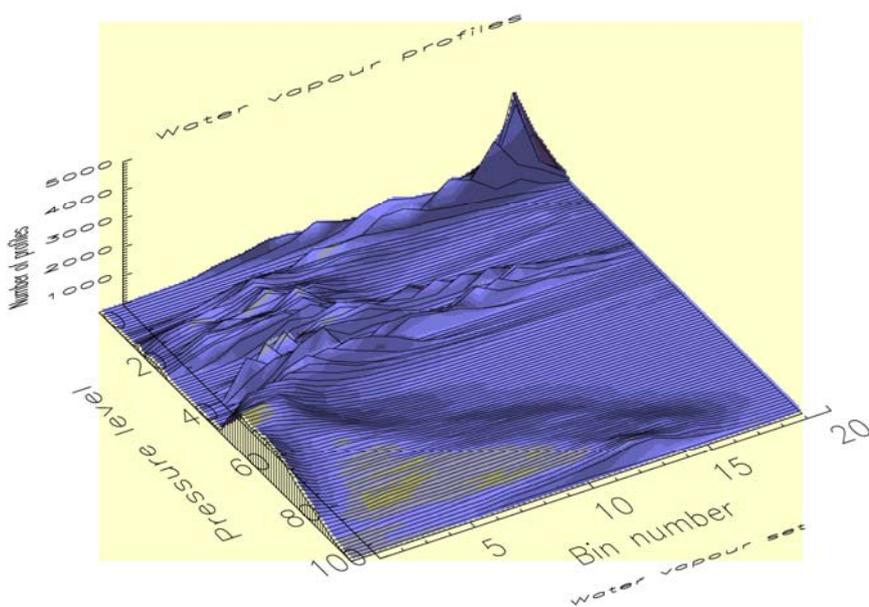


Figure 9: Histogram of frequency by bin number and pressure level for the water vapor profiles in the water vapor dataset.

Level	Pressure	Level	Pressure	Level	Pressure	Level	Pressure
1	0.005	30	29.121	59	235.234	88	753.628
2	0.016	31	32.274	60	247.408	89	777.790
3	0.038	32	35.651	61	259.969	90	802.371
4	0.077	33	39.257	62	272.919	91	827.371
5	0.137	34	43.100	63	286.262	92	852.788
6	0.224	35	47.188	64	300.000	93	878.620
7	0.345	36	51.528	65	314.140	94	904.866
8	0.506	37	56.126	66	328.675	95	931.524
9	0.714	38	60.990	67	343.618	96	958.591
10	0.975	39	66.125	68	358.966	97	986.067
11	1.297	40	71.540	69	374.725	98	1013.948
12	1.687	41	77.240	70	390.893	99	1042.232
13	2.153	42	83.231	71	407.474	100	1070.917
14	2.700	43	89.520	72	424.470	101	1100
15	3.340	44	96.114	73	441.882		
16	4.077	45	103.017	74	459.712		
17	4.920	46	110.237	75	477.961		
18	5.878	47	117.778	76	496.630		
19	6.957	48	125.646	77	515.720		
20	8.166	49	133.846	78	535.232		
21	9.512	50	142.385	79	555.167		
22	11.004	51	151.266	80	575.525		
23	12.649	52	160.496	81	596.306		
24	14.456	53	170.078	82	617.511		
25	16.432	54	180.018	83	639.140		
26	18.585	55	190.320	84	661.192		
27	20.922	56	200.989	85	683.667		
28	23.453	57	212.028	86	706.565		
29	26.183	58	223.442	87	729.886		

Table 1: Fixed pressure levels (in hPa)

To start the selection process, we fixed to 70 the number of profiles to be randomly selected and considered all the possible random permutations of 12,564 profiles taken 70 at a time (the number of 70 profiles is based on a recommendation made by the ITSC working group on radiative transfer. Since the number of permutations ( $\sim 10^{186}$ ) is too large to be manageable we restricted the number of random permutations to  $10^{10}$ . We then divided the temperature, water vapor and ozone interval at each pressure level into 70 bins and for each pressure level  $i$  we computed the number of profiles  $N_{ij}^{Water\ vapor}$ ,  $N_{ij}^{Ozone}$ , and,  $N_{ij}^{Temperature}$  in each bin  $j$ . A perfectly balanced distribution (i.e. each bin contains a single different profile) would result in a zero value of the parameter  $D$  defined as:

$$D = \sum_{i=1}^{101} \sum_{j=1}^{70} |(N_{ij}^{Water\ vapor} - 1)| + \sum_{i=1}^{101} \sum_{j=1}^{70} |(N_{ij}^{Ozone} - 1)| + \sum_{i=1}^{101} \sum_{j=1}^{70} |(N_{ij}^{Temperature} - 1)| \quad (5)$$

To select what we consider the optimal set we looked for the combination of profiles that resulted in the minimum value of  $D$ . During this stage no effort was made to exclude profiles that occurred over high elevations since we deem important that these profiles are included in the training set since they usually represent extreme situations. Following the automatic selection of the 70 profiles we added two profiles (minimum and maximum profile) representing the envelope of the 12,564 profile set. A further 10 profiles were added manually to improve the coverage of the regions in the proximity of the minimum and maximum values (mainly for temperature) and finally we added one more profile computed as the average of the 82 profiles to serve as a reference profile in the regression. The histogram of frequency by bin number of the 83 profiles in the training set is shown in figures 10, 11, and 12 for temperature, water vapor and ozone respectively. From these figures it can be seen that the distribution of the profiles in the training has many features in common with the distribution of the profiles in the global set. However, the main reason why we performed a constrained random selection of the profiles is that above all we wanted the distribution of the water vapor profiles in the training set to look like the distribution in figure 7 rather than that in figure 9. Figure 11 shows that on average this goal has been achieved. By comparison, in figures 13, 14, and 15 we show the equivalents of figures 10, 11, and, 12 for the training set of 52 profiles described in Chevallier (2002). What can be observed from these figures is that the water vapor profiles in the 83 profile set are more uniformly distributed than the profiles in the 52 profiles set. This is all the more evident if the profiles are plotted on a single point basis as a function of pressure. This is shown in figures 16 and 17 for the 52 and 83 profile set respectively. It is evident that the coverage of the water vapor profiles in the 52 profile set is punctuated by large gaps whereas the profiles in the 83 profile set fills more regularly the range of variability of humidity. It could be argued that this is the mere results of the 82 profile set containing a larger number of profiles. However, it should be noted that the percentage of moist profiles in the 52 profile set is significantly lower than the percentage in the 82 profile set, a feature that we think can be largely ascribed to the fact that during the selection process we have forced the water vapor profiles to better cover the moister regions. For completeness, the point distribution is also shown for temperature and ozone in figures 18, 19, 20 and 21 for both training sets. These figures confirm what has been already noticed for the water vapor profiles, i.e. the profile coverage in the 82 profile set is more uniform than the coverage in the 52 profile set.

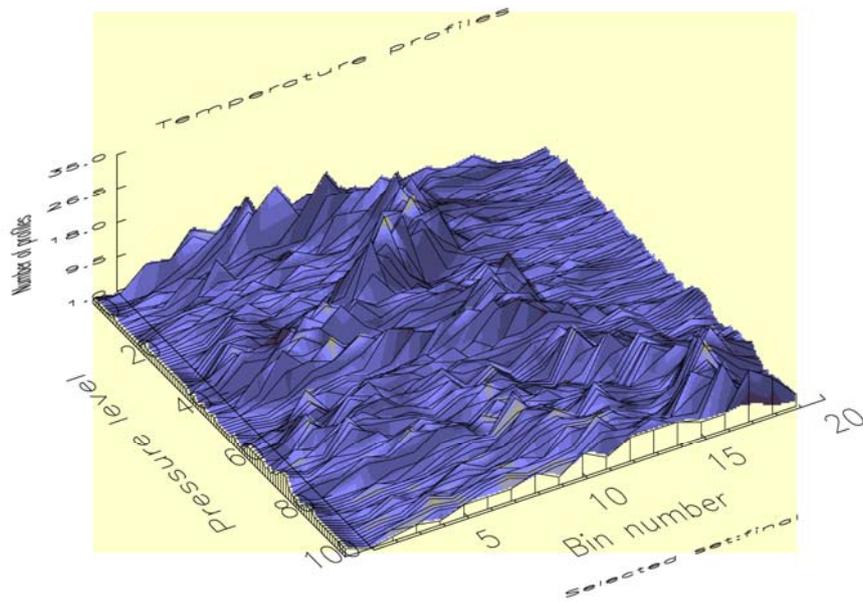


Figure 10: Histogram of frequency by bin number and pressure level for the temperature profiles in the 83 profile training set.

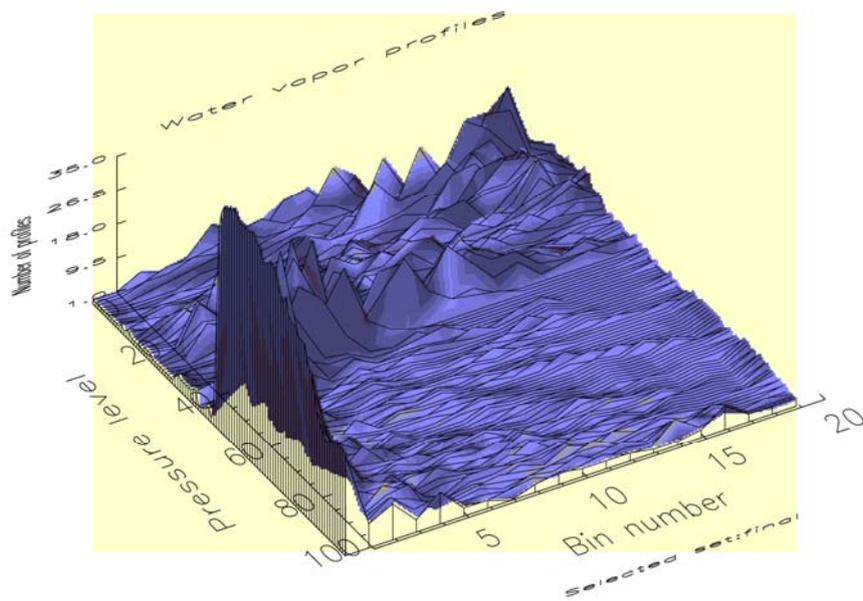


Figure 11: Histogram of frequency by bin number and pressure level for the water vapor profiles in the 83 profile training set.

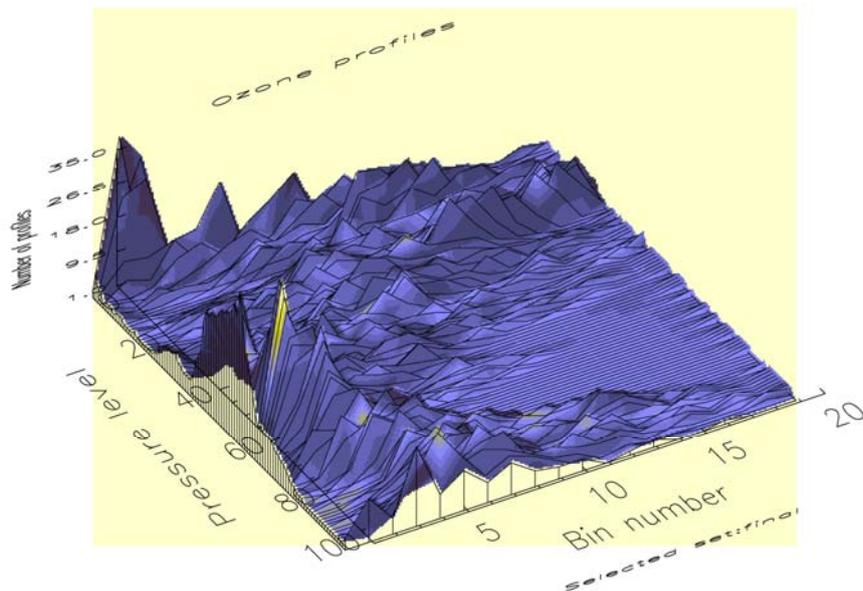


Figure 12: Histogram of frequency by bin number and pressure level for the ozone profiles in the 83 profile training set.

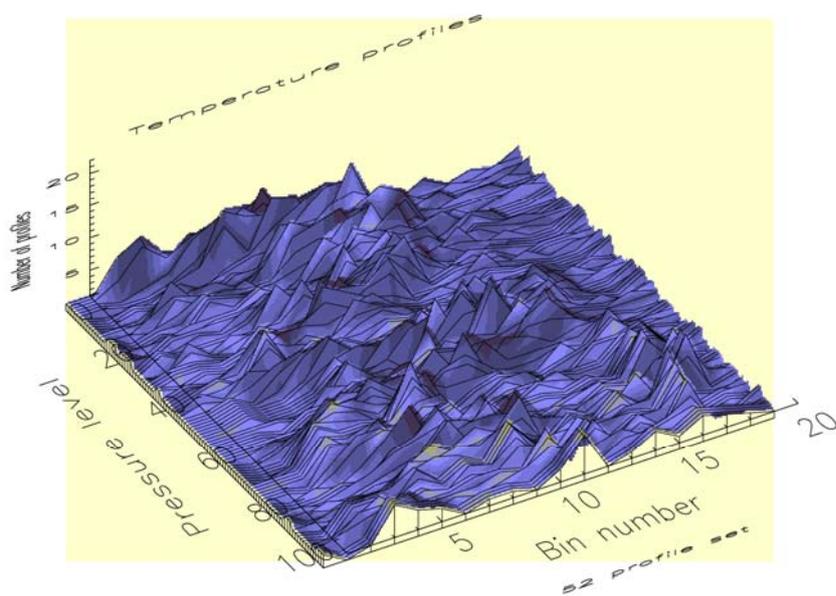


Figure 13: Histogram of frequency by bin number and pressure level for the temperature profiles in the 52 profile training set.

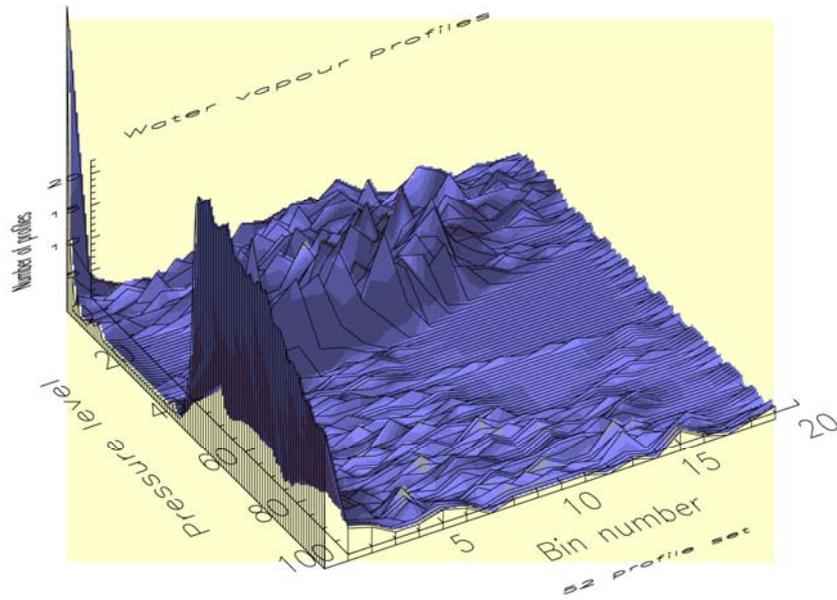


Figure 14: Histogram of frequency by bin number and pressure level for the water vapor profiles in the 52 profile training set.

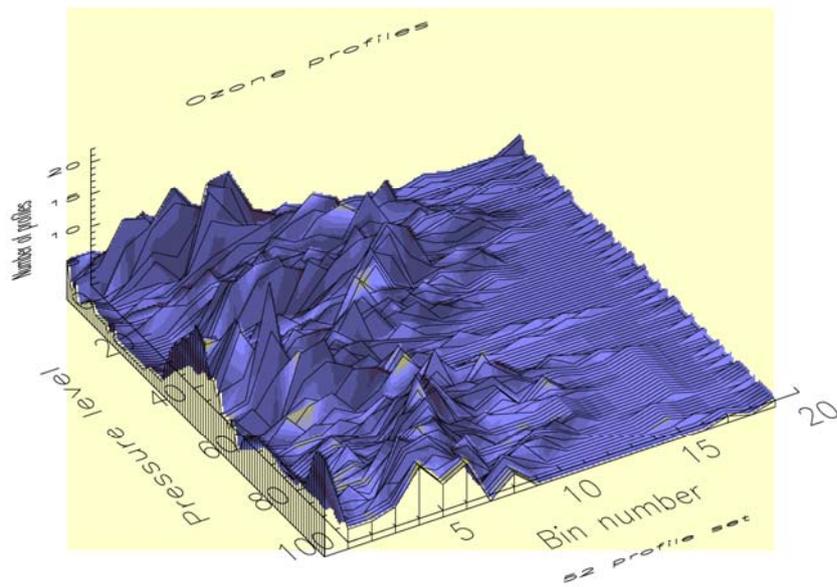


Figure 15: Histogram of frequency by bin number and pressure level for the ozone profiles in the 52 profile training set.

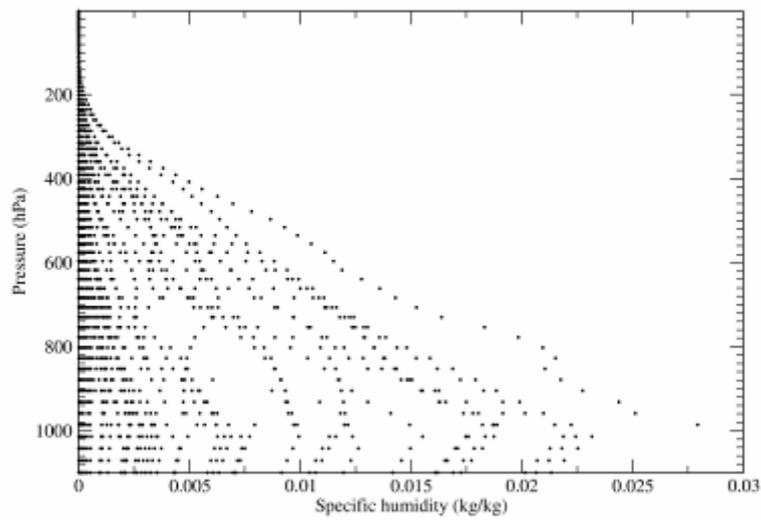


Figure 16: The point distribution of the water vapor profiles in the 52 profile set.

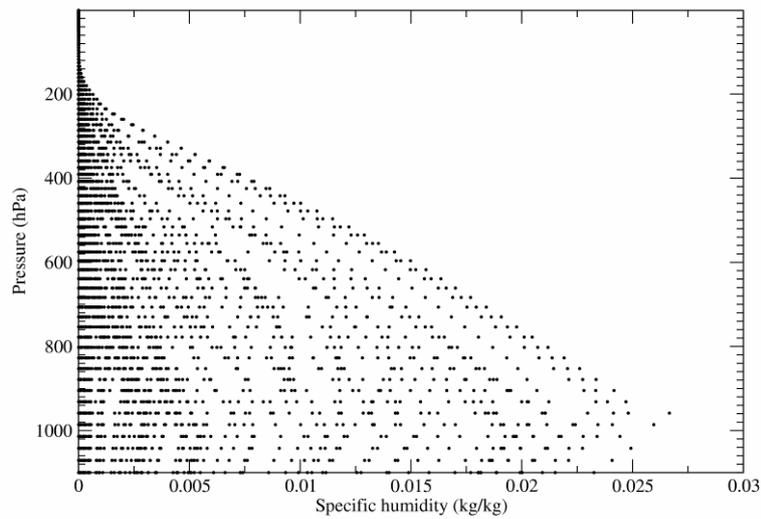


Figure 17: The point distribution of the water vapor profiles in the 83 profile set.

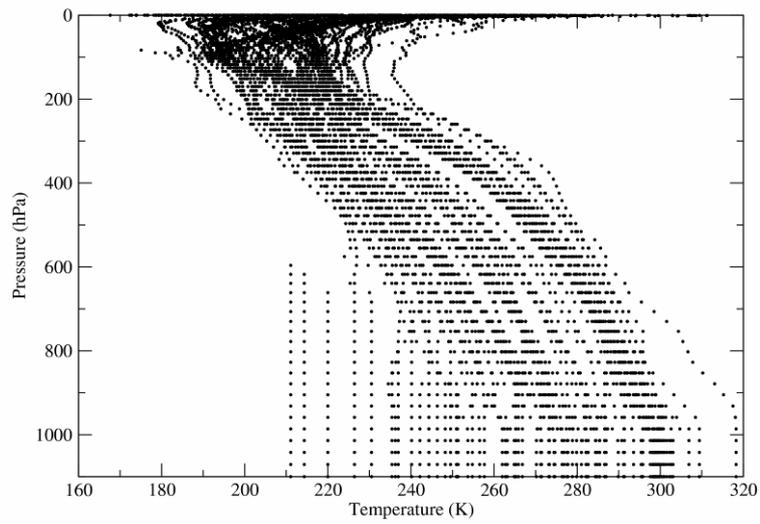


Figure 18: The point distribution of the temperature profiles in the 83 profile set.

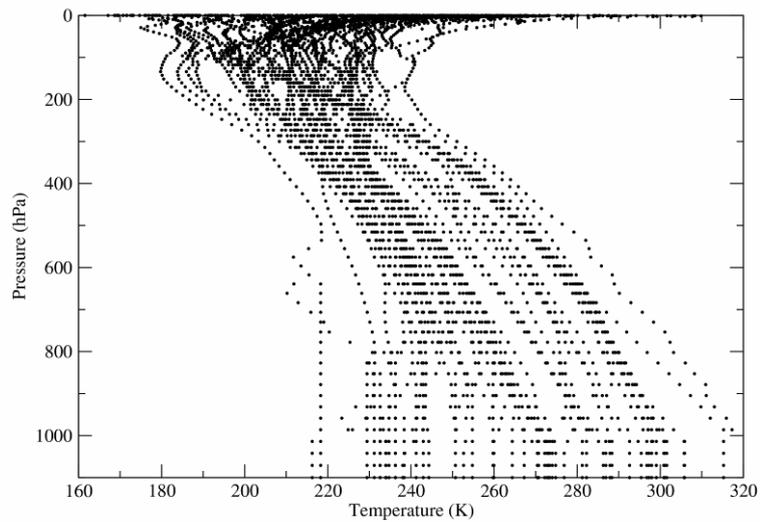


Figure 19: The point distribution of the temperature profiles in the 52 profile set.

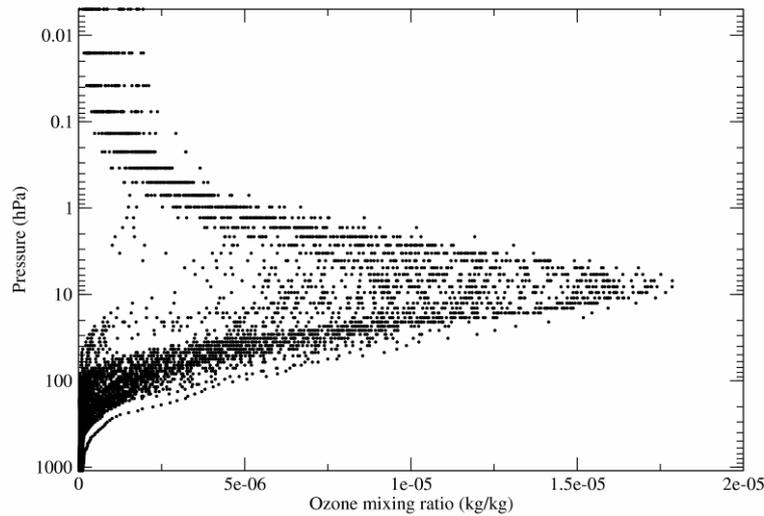


Figure 20: The point distribution of the ozone profiles in the 83 profile set.

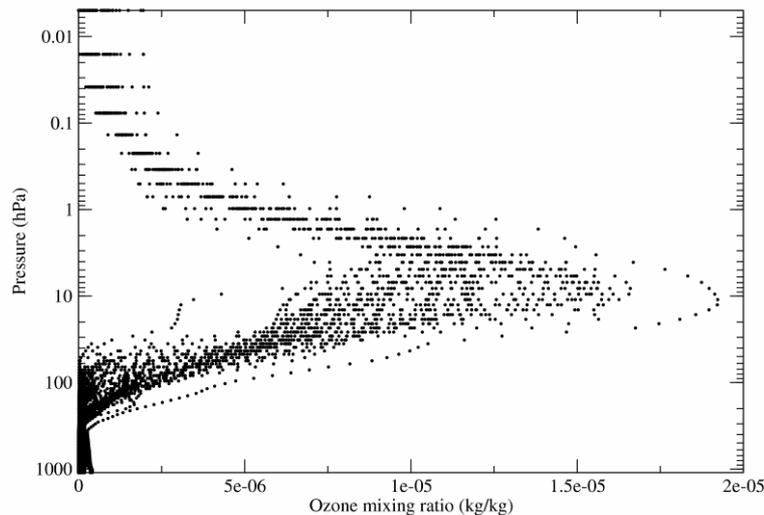


Figure 21: The point distribution of the ozone profiles in the 52 profile set.

The statistical characteristics of the profiles in the 83 profile training set are shown in figures 22, 23, and 24. As discussed previously the maximum and minimum profiles in the training set consists of the envelope formed by all the profiles contained in the global set on which the training set is based. Differences that exist between the maxima and minima in figures 22, 23, 24 and figures 3, 4, 5 are due either to extrapolations or adjustments made to correct for supersaturation. For instance the minimum temperature profile in figure 22 differs markedly from the minimum temperature profile in figure 3 because the minimum value around 600 hPa was found for a profile over a very high elevation and the temperature profile had to be extrapolated to the surface assuming the constant

value at 600 hPa. Although the minimum profile features artificially low values in this region, this is not detrimental for the regression since the whole range of values at each level is well covered by the profiles in the training set as shown in figure 18. Apart from the obvious feature of the lowest minima in the troposphere, a comparison of the statistics of the temperature profiles in the global and training set shows that the mean value of the temperature in the training set is shifted towards smaller values (above all in the troposphere) whereas larger values of the standard deviation are observed. A larger standard deviation and a shift towards smaller average values is also a feature of the water vapor and ozone profiles in the training set. This pattern (i.e. smaller mean value, larger standard deviation) is indicative of the fact that the selection process of the training profiles acted towards a thinning of the profile population around the mean value.

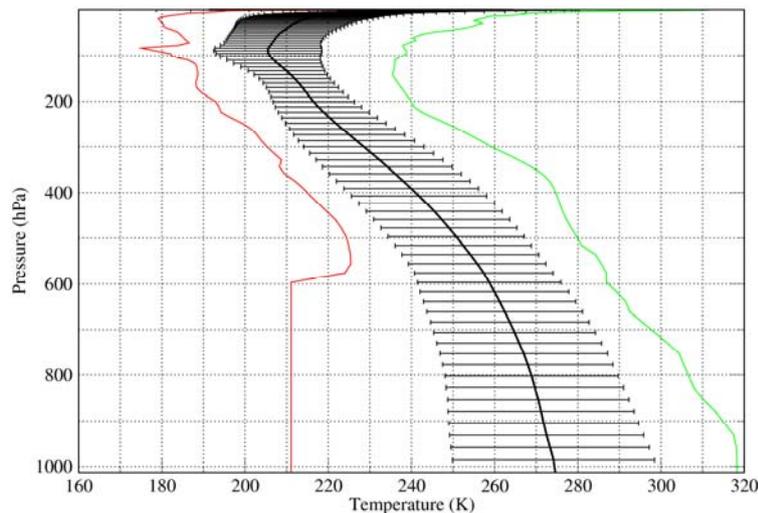


Figure 22: Statistics of the temperature profiles in the 83 profile training set. The minimum and maximum values are shown in red and green respectively. The thick black curve is the mean value and the error bars have a width of twice the standard deviation.

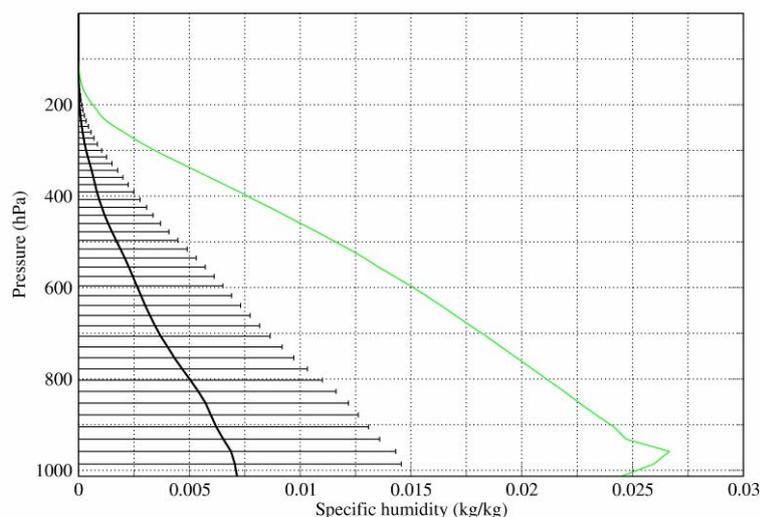


Figure 23: Statistics of the water vapor profiles in the 83 profile training set. The minimum and maximum values are shown in red and green respectively. The thick black curve is the mean value and the error bars have a width of twice the standard deviation.

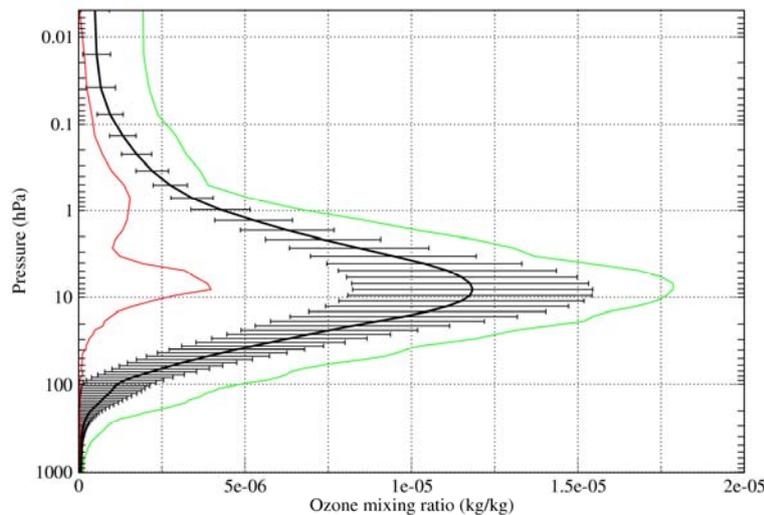


Figure 24: Statistics of the ozone profiles in the 83 profile training set. The minimum and maximum values are shown in red and green respectively. The thick black curve is the mean value and the error bars have a width of twice the standard deviation.

It is interesting to look at the differences between the statistics of the profiles in the 82 and 52 profile set. It should be observed that the global set on which the 52 profile set is based was generated by the ERA-40 assimilation system using a forecast model with a horizontal resolution of about 125 km and a vertical pressure grid of 60 levels with the top level placed at 0.1 hPa. In contrast, the horizontal resolution of the forecast model used for the 82 profile set is about 25 km and the vertical levels are 91 with the top level at 0.01 hPa. It should also be stressed that since the release of the 52 profiles set the forecast model has been improved and the assimilation system has benefited from a new wealth of radiance data. Both training sets use the envelope of the global databases as extreme profiles and consequently any difference in the range should be ascribed to differences in the original sampled databases. The statistical characteristics of the profiles in the 52 profile training set are shown in figures 25, 26, and 27. Looking at the temperature profiles some features can be noticed. The lowest minima in the lower troposphere observed in the 82 profile training set are a result of including more profiles over higher elevations and the consequent extrapolation of these profiles to the surface. If we exclude this region, the minimum values in the 82 profiles training set are generally larger than the minimum values in the 52 profile training set. Although in some limited regions this result is influenced by the fact that a number of ozone profiles (and associated temperature and water vapor profiles) have been excluded from the global set because of the anomalies in the upper stratosphere, larger minimum values are in general a genuine feature of the 82 profile training set. A result that is not influenced by the exclusion of the ozone profiles is that the 82 profiles set features larger maximum values in the troposphere (pressures lower than 200 hPa) and smaller maximum values in the stratosphere. Since these features reflect almost exactly those of the global databases on which the different training sets are based, the conclusion is that in most of the regions the envelope of the global set used to select the 83 profiles is shifted towards larger values. This is also true for the water vapor profiles: maximum values in the 83 profile set are significantly larger than the maximum values in the 52 profile set at any pressure. For ozone we observe the opposite pattern, i.e. maximum and minimum values are shifted towards smaller values. The interpretation of these results is difficult because the global dataset on which the two training

sets are based have been obtained using different sampling strategies and different forecast/assimilation systems and any in-depth investigation would be outside the scope of this paper.

In terms of mean value and standard deviation, the 83 profile temperature set exhibits larger mean values in the troposphere and smaller values above 200 hPa. Values of the standard deviation are comparable to those observed in the 52 profiles set up to 300 hPa and tend to be smaller above this level. For water vapor, mean values in the 82 profiles are larger at any pressure level whereas standard deviations are of comparable or larger magnitude. Finally, for ozone, mean values are of comparable magnitude up to 10 hPa. Above this pressure values in the 83 profiles set tend to be smaller than values in the 52 profiles set. Standard deviations for the profiles in the former set are larger than the values in the latter set at any pressure.

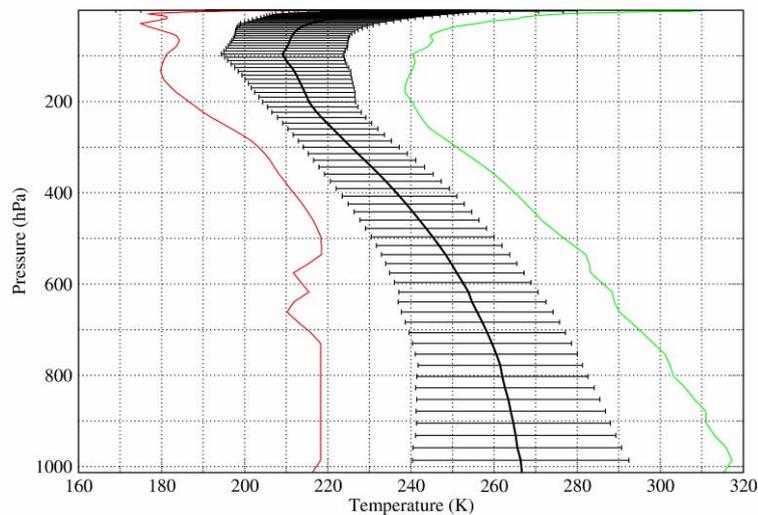


Figure 25: Statistics of the temperature profiles in the 52 profile training set. The minimum and maximum values are shown in red and green respectively. The thick black curve is the mean value and the error bars have a width of twice the standard deviation.

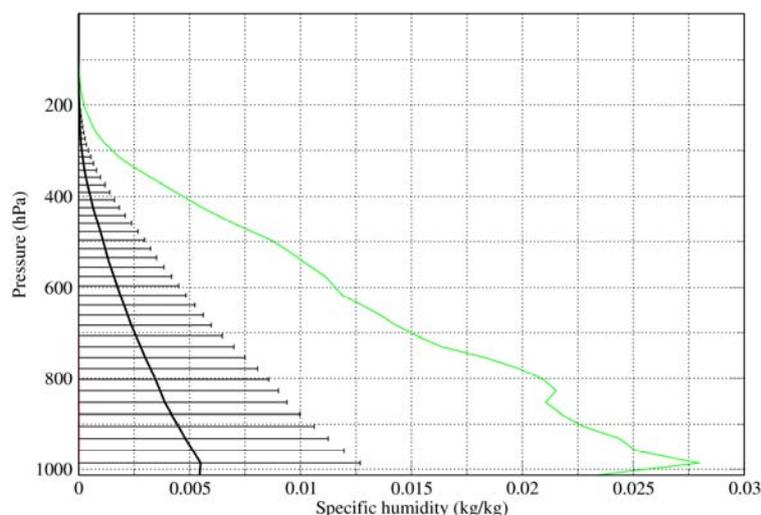


Figure 26: Statistics of the water vapor profiles in the 52 profile training set. The minimum and maximum values are shown in red and green respectively. The thick black curve is the mean value and the error bars have a width of twice the standard deviation.

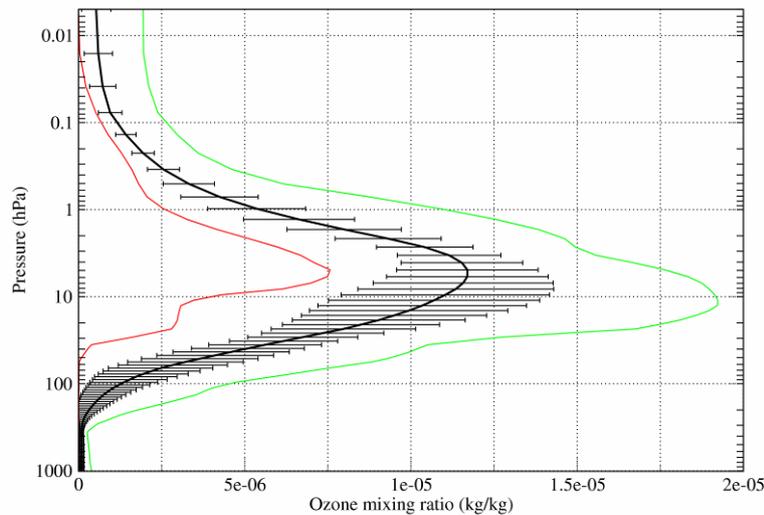


Figure 27: Statistics of the ozone profiles in the 52 profile training set. The minimum and maximum values are shown in red and green respectively. The thick black curve is the mean value and the error bars have a width of twice the standard deviation.

In addition to temperature, water vapor and ozone profiles, the training of RTTOV requires a set of diverse profiles for each of the trace gases that are allowed to vary, namely  $\text{CO}_2$ ,  $\text{N}_2\text{O}$ ,  $\text{CO}$  and  $\text{CH}_4$ .

For the generation of the  $\text{N}_2\text{O}$  profiles we used the methods described in Matricardi (2003). The  $\text{N}_2\text{O}$  level in the lower troposphere is monitored by the Climate Monitoring and Diagnostic Laboratory (CMDL) of the National Oceanic and Atmospheric Administration (NOAA) and by the Advanced Global Atmospheric Gas Experiment (AGAGE) program through measurements made at a world-wide distributed network of stations (for further information see <http://www.cmdl.noaa.gov> and <http://www.cdiac.ornl.gov/ndps/alegag.html>). The  $\text{N}_2\text{O}$  profile set was generated using stratospheric profiles retrieved from spectra measured by the Cryogenic Limb Array Etalon Spectrometer (CLAES) and from surface measurements made at a number of stations of the CMDL and AGAGE network. The CLAES instrument has flown on board the Upper Atmosphere Research Satellite (UARS) (Reber et al. 1993) to provide stratospheric mixing ratios for 23 different molecular species. Profiles for  $\text{N}_2\text{O}$  are available from 19-1-1992 to 13-3-1993 and are given at 4 degree latitude intervals on the UARS pressure grid. To generate the  $\text{N}_2\text{O}$  profile set, stratospheric profiles from CLAES were selected for the location nearest to each of the 83 temperature/water vapor/ozone profiles described previously. Depending on season and latitude, these stratospheric profiles were then joined by parabolae to a constant tropospheric mixing ratio from the tropopause to the surface based on the latest measurements made by the CMDL and AGAGE network. A preliminary profile set was then generated using the stratospheric monthly mean profile at each location. The final profile set was obtained by replacing in turn each of the monthly mean profiles with either the minimum or maximum profile if the latter was found to extend the range of variability of the set based on the monthly mean profiles. Finally, all the profiles were adjusted to reflect the mixing ratio forecasted for the year 2008. This was done assuming an increase of 0.72 ppb/year.

For the generation of the  $\text{CO}_2$ ,  $\text{CO}$  and  $\text{CH}_4$  datasets we have adopted a strategy different from that used in Matricardi (2003). The profiles in this study have in fact been selected using data from a number of forecast experiments performed at ECMWF within the context of the GEMS project. It should be stressed that a database

of sampled profiles like that available for temperature, water vapor and ozone is not yet available for trace gas/reactive gases and consequently the selection of the trace gas profiles has been performed by extracting archived data corresponding to the time and location of each of the 83 training profiles based on the assumption that the geographical spread of the training profiles would result in a distribution that covers well the range of variability of the trace gas species.

Forecasted profiles of CO<sub>2</sub> (Engelen, ECMWF, personal communication), CO (Flemming, ECMWF, personal communication) and CH<sub>4</sub> (Serrar, ECMWF, personal communication) were available for the entire year 2003. Archived profiles were extracted for the day/month and location of each training profile. For this purpose we used short range forecast (3 hours) data averaging the profiles generated at 00Z and 12Z hours. All the profiles were then interpolated to the standard 101 levels from the GEMS 60 level vertical pressure grid.

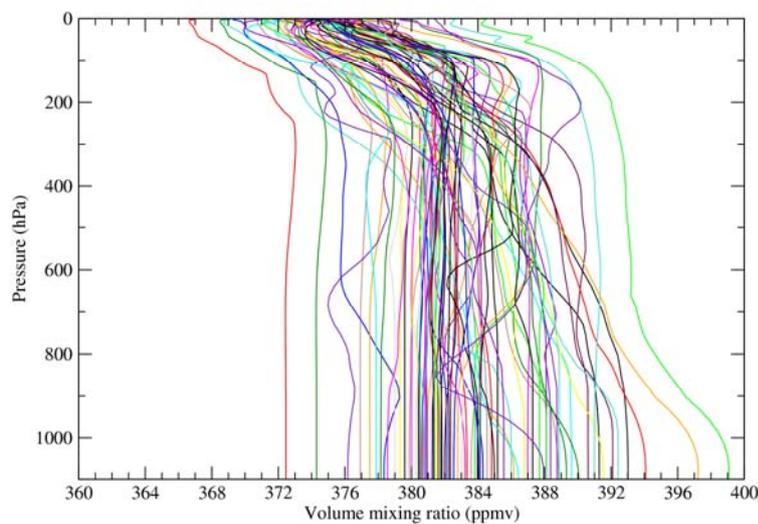


Figure 28: The training set of CO<sub>2</sub> profiles.

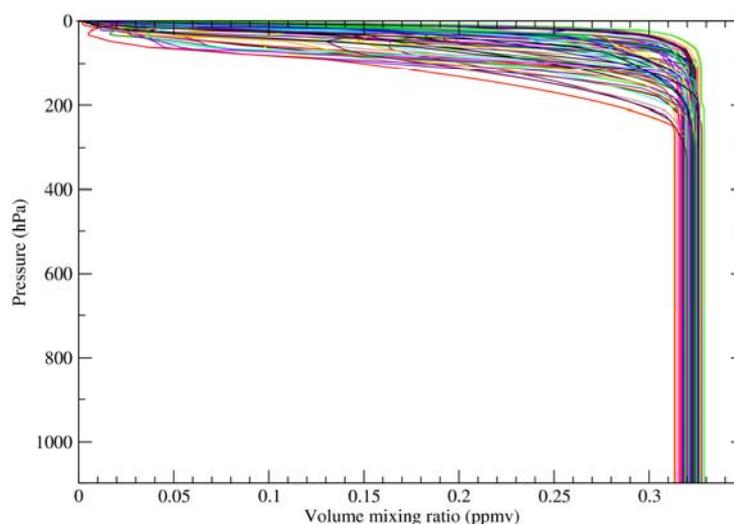


Figure 29: The training set of N<sub>2</sub>O profiles.

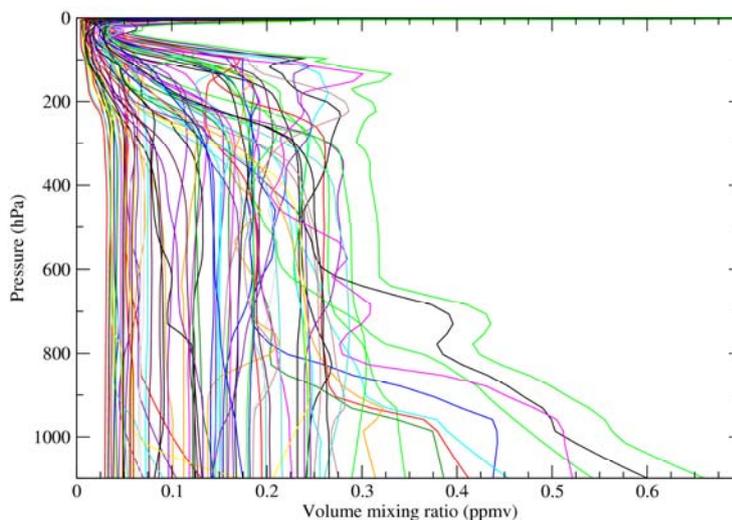


Figure 30: The training set of CO profiles.

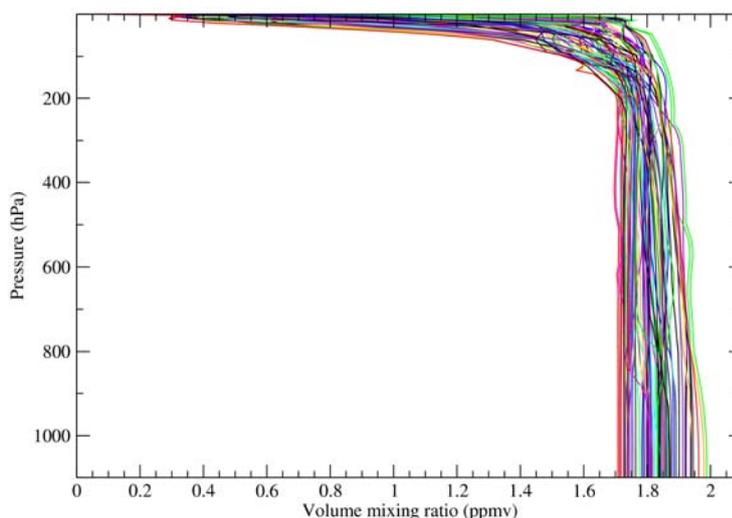


Figure 31: The training set of CH<sub>4</sub> profiles.

Since this sampling does not guarantee that profiles representative of extreme values are included in the database, we have scaled each profile to a value based on measurements made by the closest station of the CMDL and AGAGE network. To this end we have collected CMDL and AGAGE surface data and computed monthly mean, minimum and maximum values for each station. We have then scaled each profile to a value that was either the monthly mean or the minimum or maximum value in such a way that the resulting profile distribution was as uniform as possible and the whole range of variability seen in the CMDL and AGAGE data is represented in the training set.

Finally, the concentrations for CH<sub>4</sub>, N<sub>2</sub>O, and CO<sub>2</sub> profiles were scaled to year 2009 assuming a rate of increase of 1.5ppbv/year, 0.72ppbv/year and 1.85ppmv/year respectively.

#### 4. The line-by-line database

As discussed earlier, the LBL database in this paper was generated using LBLRTM. In our LBL calculations we included 17 atmospheric constituents. While the concentration of H<sub>2</sub>O, O<sub>3</sub>, CO<sub>2</sub>, N<sub>2</sub>O, CO and CH<sub>4</sub> is allowed to vary, the other gases included in the LBL computations are held constant. They are not allowed to vary in the fast RT model because it is assumed that their spatial and temporal concentration variations do not contribute significantly to the observed radiances and will be referred to as fixed gases. Fixed gases included in the LBL computations are NO<sub>2</sub>, SO<sub>2</sub>, NO, N<sub>2</sub>, O<sub>2</sub>, HNO<sub>3</sub>, OCS, CCl<sub>4</sub>, CF<sub>4</sub>, CCl<sub>3</sub>F (CFC-11) and CCl<sub>2</sub>F<sub>2</sub> (CFC-12). Among these species, NO<sub>2</sub>, SO<sub>2</sub>, and NO were not featured in any of the LBL database previously released. We used AFGL atmospheric constituent profiles for all gases with the exception of HNO<sub>3</sub> and NO for which we used profiles generated using the MOZART chemical transport model (Clerbaux, Service d'Aéronomie, personal communication). The profiles of CCl<sub>4</sub>, CF<sub>4</sub>, CCl<sub>3</sub>F (CFC-11) and CCl<sub>2</sub>F<sub>2</sub> (CFC-12) were scaled to reflect present-day concentrations using the value tabulated in Table 2.

Species	Concentrations in ppb
CFC-11	246
CFC-12	540
CCl <sub>4</sub>	74
CF <sub>4</sub>	90

Table 2 Present-day concentrations for the CFC and Halons included in the LBLRTM computations.

Since RTTOV radiances are computed using the polychromatic approximation (i.e. it is assumed that the convolution of the monochromatic radiances can be approximated by the radiance computed using the convolved transmittances) to reduce errors introduced by the separation of the gas transmittance after the convolution (the convolution of the total transmittance differs from the product of the transmittance of the single gases convolved individually) we have defined 16 frequency intervals and computed LBL transmittances,  $T$ , for all the different combinations of gases shown in Table 3. Note how in each interval all terms but one cancels out. The term that is left is the correct value for the total convolved transmittance. It should be stressed that in each interval we have only included the species that are radiatively active in that interval. Because of the complex nature of its absorption line spectrum, H<sub>2</sub>O is the only specie that is featured in every interval. In Table 3, the circumflex over the symbols denote convolution with the appropriate instrumental spectral response function,  $i$  is the channel number and  $j$  is the pressure level.

<b>Interval 1</b>	<b>[645 to 850 cm<sup>-1</sup>]</b>
$\hat{\Gamma}_{i,j}^{tot} = \hat{\Gamma}_{i,j}^{fix} \left( \frac{\hat{\Gamma}_{i,j}^{fix+CO_2}}{\hat{\Gamma}_{i,j}^{fix}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O+CO_2}}{\hat{\Gamma}_{i,j}^{fix+CO_2}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O+CO_2+O_3}}{\hat{\Gamma}_{i,j}^{fix+H_2O+CO_2}} \right)$	
<b>Interval 2</b>	<b>[850.25 to 949.75 cm<sup>-1</sup>]</b>
$\hat{\Gamma}_{i,j}^{tot} = \hat{\Gamma}_{i,j}^{fix} \left( \frac{\hat{\Gamma}_{i,j}^{fix+CO_2}}{\hat{\Gamma}_{i,j}^{fix}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O+CO_2}}{\hat{\Gamma}_{i,j}^{fix+CO_2}} \right)$	
<b>Interval 3</b>	<b>[950 to 1100 cm<sup>-1</sup>]</b>
$\hat{\Gamma}_{i,j}^{tot} = \hat{\Gamma}_{i,j}^{fix} \left( \frac{\hat{\Gamma}_{i,j}^{fix+CO_2}}{\hat{\Gamma}_{i,j}^{fix}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O+CO_2}}{\hat{\Gamma}_{i,j}^{fix+CO_2}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O+CO_2+O_3}}{\hat{\Gamma}_{i,j}^{fix+H_2O+CO_2}} \right)$	
<b>Interval 4</b>	<b>[1100.25 to 1199.75 cm<sup>-1</sup>]</b>
$\hat{\Gamma}_{i,j}^{tot} = \hat{\Gamma}_{i,j}^{fix} \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4}}{\hat{\Gamma}_{i,j}^{fix}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4+N_2O}}{\hat{\Gamma}_{i,j}^{fix+CH_4}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O+CH_4+N_2O}}{\hat{\Gamma}_{i,j}^{fix+CH_4+N_2O}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O+CH_4+N_2O+O_3}}{\hat{\Gamma}_{i,j}^{fix+H_2O+CH_4+N_2O}} \right)$	
<b>Interval 5</b>	<b>[1200 to 1250 cm<sup>-1</sup>]</b>
$\hat{\Gamma}_{i,j}^{tot} = \hat{\Gamma}_{i,j}^{fix} \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4}}{\hat{\Gamma}_{i,j}^{fix}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4+N_2O}}{\hat{\Gamma}_{i,j}^{fix+CH_4}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4+N_2O+CO_2}}{\hat{\Gamma}_{i,j}^{fix+CH_4+N_2O}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4+N_2O+CO_2+vw}}{\hat{\Gamma}_{i,j}^{fix+CH_4+N_2O+CO_2}} \right) \left( \frac{\hat{\tau}_{i,j}^{fix+CH_4+N_2O+CO_2+H_2O+O_3}}{\hat{\tau}_{i,j}^{fix+CH_4+N_2O+CO_2+H_2O}} \right)$	
<b>Interval 6</b>	<b>[1250.25 to 1350 cm<sup>-1</sup>]</b>
$\hat{\Gamma}_{i,j}^{tot} = \hat{\Gamma}_{i,j}^{fix} \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4}}{\hat{\Gamma}_{i,j}^{fix}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4+N_2O}}{\hat{\Gamma}_{i,j}^{fix+CH_4}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4+N_2O+CO_2}}{\hat{\Gamma}_{i,j}^{fix+CH_4+N_2O}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4+N_2O+CO_2+H_2O}}{\hat{\Gamma}_{i,j}^{fix+CH_4+N_2O+CO_2}} \right)$	
<b>Interval 7</b>	<b>[1350.25 to 1400 cm<sup>-1</sup>]</b>
$\hat{\Gamma}_{i,j}^{tot} = \hat{\Gamma}_{i,j}^{fix} \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4}}{\hat{\Gamma}_{i,j}^{fix}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4+CO_2}}{\hat{\Gamma}_{i,j}^{fix+CH_4}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O+CH_4+CO_2}}{\hat{\Gamma}_{i,j}^{fix+CH_4+CO_2}} \right)$	
<b>Interval 8</b>	<b>[1400.25 to 1649.75 cm<sup>-1</sup>]</b>
$\hat{\Gamma}_{i,j}^{tot} = \hat{\Gamma}_{i,j}^{fix} \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4}}{\hat{\Gamma}_{i,j}^{fix}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O+CH_4}}{\hat{\Gamma}_{i,j}^{fix+CH_4}} \right)$	
<b>Interval 9</b>	<b>[1650 to 1750 cm<sup>-1</sup>]</b>
$\hat{\Gamma}_{i,j}^{tot} = \hat{\Gamma}_{i,j}^{fix} \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4}}{\hat{\Gamma}_{i,j}^{fix}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O+CH_4}}{\hat{\Gamma}_{i,j}^{fix+CH_4}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O+CH_4+O_3}}{\hat{\Gamma}_{i,j}^{fix+H_2O+CH_4}} \right)$	
<b>Interval 10</b>	<b>[1750.25 to 1850 cm<sup>-1</sup>]</b>
$\hat{\Gamma}_{i,j}^{tot} = \hat{\Gamma}_{i,j}^{fix} \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O}}{\hat{\Gamma}_{i,j}^{fix}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O+O_3}}{\hat{\Gamma}_{i,j}^{fix+H_2O}} \right)$	

Table 3: The combination of gases used to compute the LBL transmittances

<b>Interval 11</b>	<b>[1850.25 to 1900 cm<sup>-1</sup>]</b>
$\hat{\Gamma}_{i,j}^{tot} = \hat{\Gamma}_{i,j}^{fix} \left( \frac{\hat{\Gamma}_{i,j}^{fix+N_2O}}{\hat{\Gamma}_{i,j}^{fix}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+CO_2+N_2O}}{\hat{\Gamma}_{i,j}^{fix+N_2O}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O+CO_2+N_2O}}{\hat{\Gamma}_{i,j}^{fix+CO_2+N_2O}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O+CO_2+N_2O+O_3}}{\hat{\Gamma}_{i,j}^{fix+H_2O+CO_2+N_2O}} \right)$	
<b>Interval 12</b>	<b>[1900.25 to 1994.75 cm<sup>-1</sup>]</b>
$\hat{\Gamma}_{i,j}^{tot} = \hat{\Gamma}_{i,j}^{fix} \left( \frac{\hat{\Gamma}_{i,j}^{fix+CO_2}}{\hat{\Gamma}_{i,j}^{fix}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O+CO_2}}{\hat{\Gamma}_{i,j}^{fix+CO_2}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O+CO_2+O_3}}{\hat{\Gamma}_{i,j}^{fix+H_2O+CO_2}} \right)$	
<b>Interval 13</b>	<b>[1995 to 2295 cm<sup>-1</sup>]</b>
$\hat{\Gamma}_{i,j}^{tot} = \hat{\Gamma}_{i,j}^{fix} \left( \frac{\hat{\Gamma}_{i,j}^{fix+CO}}{\hat{\Gamma}_{i,j}^{fix}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+CO+N_2O}}{\hat{\Gamma}_{i,j}^{fix+CO}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+CO+N_2O+CO_2}}{\hat{\Gamma}_{i,j}^{fix+CO+N_2O}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+CO+N_2O+CO_2+H_2O}}{\hat{\Gamma}_{i,j}^{fix+CO+N_2O+CO_2}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+CO+N_2O+CO_2+H_2O+O_3}}{\hat{\Gamma}_{i,j}^{fix+CO+N_2O+CO_2+H_2O}} \right)$	
<b>Interval 14</b>	<b>[2295.25 to 2359.75 cm<sup>-1</sup>]</b>
$\hat{\Gamma}_{i,j}^{tot} = \hat{\Gamma}_{i,j}^{fix} \left( \frac{\hat{\Gamma}_{i,j}^{fix+CO_2}}{\hat{\Gamma}_{i,j}^{fix}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O+CO_2}}{\hat{\Gamma}_{i,j}^{fix+CO_2}} \right)$	
<b>Interval 15</b>	<b>[2360 to 2660 cm<sup>-1</sup>]</b>
$\hat{\Gamma}_{i,j}^{tot} = \hat{\Gamma}_{i,j}^{fix} \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4}}{\hat{\Gamma}_{i,j}^{fix}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4+N_2O}}{\hat{\Gamma}_{i,j}^{fix+CH_4}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4+N_2O+CO_2}}{\hat{\Gamma}_{i,j}^{fix+CH_4+N_2O}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4+N_2O+CO_2+H_2O}}{\hat{\Gamma}_{i,j}^{fix+CH_4+N_2O+CO_2}} \right)$	
<b>Interval 16</b>	<b>[2660.25 to 2760 cm<sup>-1</sup>]</b>
$\hat{\Gamma}_{i,j}^{tot} = \hat{\Gamma}_{i,j}^{fix} \left( \frac{\hat{\Gamma}_{i,j}^{fix+CH_4}}{\hat{\Gamma}_{i,j}^{fix}} \right) \left( \frac{\hat{\Gamma}_{i,j}^{fix+H_2O+CH_4}}{\hat{\Gamma}_{i,j}^{fix+CH_4}} \right)$	

Table 3: The combination of gases used to compute the LBL transmittances (contd)

For each interval and combination of gases in Table 3 we computed LBLRTM monochromatic transmittances from 0.005 hPa to each of the standard pressure levels in Table 1. Spectra were computed for each atmospheric profile and six different path angles, namely the angles for which the secant has equally spaced values from 1 to 2.25. The spectra were then interpolated to a fixed wave number grid spaced 0.001 cm<sup>-1</sup>. Since the solar term in RTTOV requires transmittances for a wider range of angles, we have extended the database by computing transmittances for an additional number of path angles that include the values of 2.58, 3.94, 3.72, 4.83, 6.1, 7.2 and 9. This means that the solar term can be evaluated for solar zenith angles as large as ≈84°. Note that the additional database has been generated only for the spectral regions where we expect the solar term to be important (i.e. wave numbers greater than 2000 cm<sup>-1</sup>).

The spectral response function,  $f$ , of the IASI instrument (ISRF) for the level 1C radiances used operationally at NWP centres is the inverse Fourier transform of a truncated Gaussian function. This truncated Gaussian function

equals the direct Fourier transform of a  $0.5 \text{ cm}^{-1}$  full width at half height (FWHH) Gaussian in the interval  $[-OPD, +OPD]$  and zero elsewhere. The generation of the IASI level 1C transmittances,  $\hat{\Gamma}_{i,j}$ , can be very time consuming if it is performed directly in the wave number,  $\tilde{\nu}$ , space:

$$\hat{\Gamma}_{i,j} = \int_{-\infty}^{+\infty} \Gamma_j(\tilde{\nu}) f(\tilde{\nu}_i - \tilde{\nu}) d\tilde{\nu} \quad (6)$$

where  $\Gamma_j$  is the monochromatic transmittance,  $\tilde{\nu}_i$  is the central wave number of the IASI channel and  $f$  is the ISRF normalized to 1. For this reason, in the past we have convolved the spectra by applying the convolution theorem to Eq. (6): the fast Fourier transform (FFT) of the transmittance spectra is multiplied by the FFT of the  $0.5 \text{ cm}^{-1}$  FWHH Gaussian and the interferogram truncated at  $\pm OPD$ -cm; the inverse FFT of the truncated interferogram yields the convolved spectra.

Due to the nature of the IASI spectral response function the convolution of the monochromatic transmittances can result in unphysical negative values in spectral regions of strong absorption. The way these negative transmittances is dealt with in the regression is to heavily downweight the data correspondent to the negative transmittances because the radiance coming from the layers where negative transmittances occur has a very small contribution to the total radiance. Since the occurrence of negative transmittances does not allow Lambert's law [ $\Gamma = \exp(-\tau)$ ,  $\tau$  is the optical depth], to be satisfied, to reconcile LBL and fast model radiances, fast model transmittances are set to a constant small negative value if the transmittance of the layer above is less than a prescribed threshold.

During the course of this study we have looked further into the issue of negative transmittances and found that the convolution in the wave number space produces negative transmittances of much smaller magnitude than the negative transmittances generated by the FFT technique, a feature, we think, has to be ascribed to the numerical properties of the FFT algorithm. The implication of this finding is that the use of a threshold to switch to negative transmittances is no longer needed in the fast model because if the convolution is performed in the wave number space the resulting very small negative transmittances can now be replaced by the very small positive transmittances generated by the regression without loss of accuracy. In fact, in all the relevant spectral regions the fit of the fast model radiances to the LBL radiances is improved. Consequently we no longer use the Fourier technique to convolve the IASI spectra but perform the convolution directly in the wave number space. As mentioned earlier the convolution in the wave number space can be very time consuming if a large number of channels is involved. However, the use of LBLRTM and the latest computer facilities available at ECMWF have made this task manageable. This is in contrast to what happened in the past when the use of GENLN2 and the computer resources available at ECMWF would have made this task prohibitive in the framework of a medium term work-package. Finally, in line with the latest specifications of the IASI instrument the value of  $OPD$  used in this paper is  $1.9679466 \text{ cm}$ .

For practical reasons the integration domain in Eq.6 must be restricted to a finite interval and consequently the IASI ISRF has to be truncated. In this paper the ISRF has been calculated over an interval of  $\pm 32 \text{ cm}^{-1}$ . Although this is the interval originally prescribed by the CNES instrument team we have tested the impact that a larger integration domain would have on the accuracy of the IASI radiances to check whether or not any change should be made. To this end we convolved LBL spectra computed for the six US AFGL atmospheres using an interval

of  $\pm 32 \text{ cm}^{-1}$  and  $\pm 64 \text{ cm}^{-1}$ , spectra  $S1$  and  $S2$  respectively. The results are shown in Figure 32 where the difference between  $S1$  and  $S2$  is plotted for the six cases considered in this study. It can be seen that differences are well below  $0.02\text{K}$  for the vast majority of the channels and well below the IASI instrument noise (not shown in the figure) in all the frequency domain. Consequently there is no need to extend the integration domain to an interval larger than that originally envisaged. For the spectral response function of AIRS, we have used the latest data available from the AIRS science team (Strow, UMBC, personal communication).

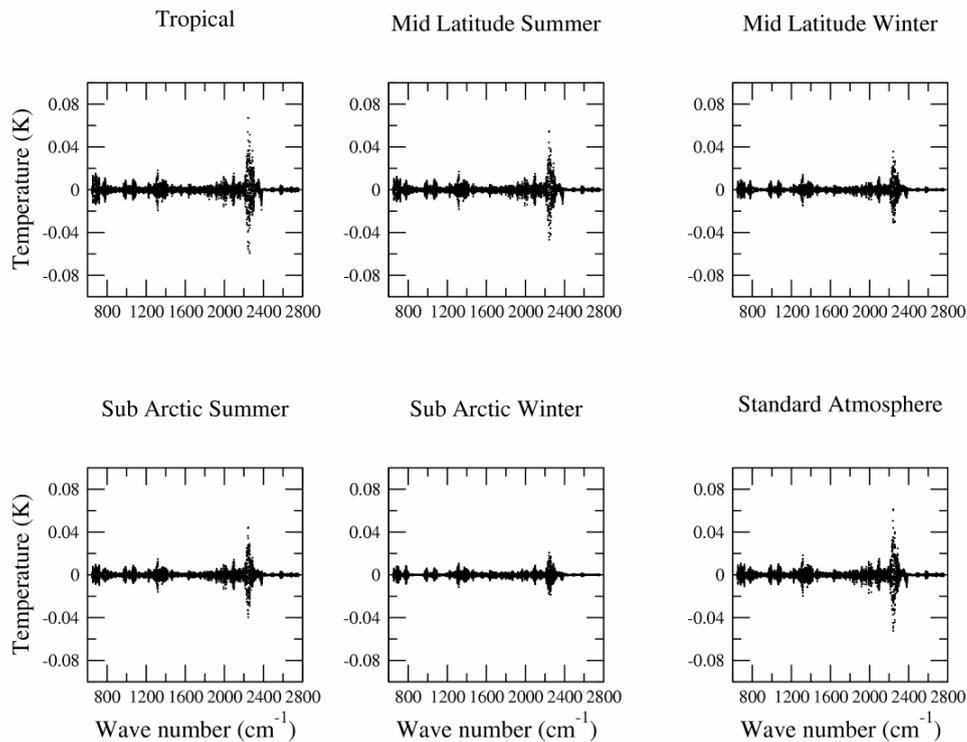


Figure 32: The difference between simulated IASI spectra using different definitions of the integration domain of the spectral response function.

Regarding the choice of molecular parameters to be used in the LBL computations we have followed the recommendation made in the study by Matricardi (2006). In this study it was suggested that the use of a database obtained from merging line data from the HITRAN and GEISA compilations could be envisaged. This suggestion was made based on the evidence that in a number of spectral regions the use of different line parameters can result in a better agreement of simulations with observations. For the database used in this study we have utilized as a baseline the line data contained in the database distributed with the LBLRTM model. This database is largely drawn from HITRAN2004 with updates from 2006 and as pointed out in a previous section it contains line parameters from HITRAN2000 that are consistent with the line mixing coefficients used in LBLRTM for the  $\text{CO}_2$  P/Q/R branches. The use of these line parameters is mandatory and consequently  $\text{CO}_2$  line data have not been merged. Details of the line data contained in the merged database are given in Table 4.

400-800 $\text{cm}^{-1}$	800-1000 $\text{cm}^{-1}$	1000-1100 $\text{cm}^{-1}$	1100-1700 $\text{cm}^{-1}$	1700-2400 $\text{cm}^{-1}$	2400-3000 $\text{cm}^{-1}$
GEISA2003	HITRAN2004/06	HITRAN2000	HITRAN2004/06	GEISA2003	HITRAN2004/06
400-3000 $\text{cm}^{-1}$					
Line data for $\text{CO}_2$ are from HITRAN2000					

Table 4: The line parameters used in the merged molecular database utilized in this study.

## 5. Optimal selection of the predictors

The functional dependence of the predictors used in RTTOV to parameterize the optical depths depends on factors such as the absorbing gas, spectral response function and spectral region although the order in which the gases are separated out and the layer thickness can also be important. An issue that we had to address when the compilation of the list of predictors was made (Matricardi 2003) is that the absorption by gas specie cannot be considered to be completely uncorrelated with absorption by other gases. In fact the quantity used in the regression is the “effective” transmittance defined in table 3 (i.e. the ratio of two transmittances) and this can differ significantly from the convolved transmittance computed for the single gas specie alone. Consequently in spectral regions where absorption lines for different molecules overlap, the introduction of predictors accounting for the variable concentration of a concomitant absorbing molecule has to be envisaged. In tables 5, 6 and 7 we give the list of predictors used in the fast transmittance model for channels not affected by solar radiation. The definition of the profile variables is given in table 8 and the channel number refers to IASI channels. From table 5 one can see for instance that the water vapor predictors can include terms that depend on the concentration of  $\text{CH}_4$  and  $\text{CO}_2$ . Predictors in RTTOV are not used on a single channel basis (i.e. each channel has its dedicated set of predictors) and consequently if predictors accounting for the variable concentration of a concomitant absorbing molecule have to be introduced, they are introduced for a whole set of channels. For many of the channels in the set, predictors based on the concentration of a concomitant absorbing molecule explain very little of the variance of the data and could, in principle, be excluded with virtually no impact on the accuracy of the transmittance model. Conversely, their use can result in the occurrence of unphysical oscillations in the Jacobians. Because of the way the effective transmittances are formed (see table 3) the concomitant absorbing molecule can only be a trace gas specie and, consequently, oscillations are only seen in trace gas Jacobians.

To address this problem we have performed a forward selection of the predictors adding one predictor at a time. In the first step we choose the single variable that is the best predictor (i.e. the predictor which gives the smallest residual sum of squares). We then add the predictor that give the best fit in conjunction with the first predictor (i.e. the predictor that results in the greatest reduction of the sum of squares when added to the current predictor). Further variables are then added in this recursive fashion, adding at each step the optimum variable, given the other variables already in the equation. The forward selection is based on an  $F$ -test and a critical value  $F_c$  is used to include or not include a predictor in the model. Consequently the model can have a number of predictors that is smaller than the number envisaged in RTTOV. Note that the selection was performed for each of the 100 layers used in the regression. A score was then computed for each predictor based on the occurrence of the predictor and the relative importance in each layer. We then looked at the score of each of the predictors based on the concentration of a concomitant absorbing molecule and excluded it from the regression if the score was found to be less than an empirically determined threshold. As a result the vast majority of the unphysical oscillations noticed in the  $\text{CH}_4$  and  $\text{CO}$  Jacobians were removed. This approach was effective in removing oscillations also for the  $\text{CO}_2$  Jacobians although for a number of moderate/weak absorption channels oscillations were still noticeable. For these channels oscillations could be removed by down-weighting the predictors  $X_{j,1}$  and  $X_{j,8}$  (see table 5 for details) in the regions not sensitive to a change of  $\text{CO}_2$  concentration. The forward selection analysis also allowed us to exclude the  $\text{N}_2\text{O}$   $N2O_{tw}(j)$  and  $\text{CH}_4$  predictors  $CH4_{tw}(j)$  from the regression since they have little or no impact on the accuracy of the transmittance model. Finally note that the forward selection was also performed for the predictors used in the short-wave for the computation of the solar term. Similar conclusions could be drawn and all the relevant predictors excluded accordingly.

Predictor	H <sub>2</sub> O	CO <sub>2</sub>	O <sub>3</sub>
$X_{j,1}$	$(\sec(\theta)W_r(j))^2$	$\sec(\theta)CO_{2,r}(j)$	$\sec(\theta)O_r(j)$
$X_{j,2}$	$\sec(\theta)W_w(j)$	$T_r^2(j)$	$\sqrt{\sec(\theta)O_r(j)}$
$X_{j,3}$	$(\sec(\theta)W_w(j))^2$	$\sec(\theta)T_r(j)$	$\sec(\theta)O_r(j)\delta T(j)$
$X_{j,4}$	$\sec(\theta)W_r(j)\delta T(j)$	$\sec(\theta)T_r^2(j)$	$(\sec(\theta)O_r(j))^2$
$X_{j,5}$	$\sqrt{\sec(\theta)W_r(j)}$	$T_r(j)$	$\sqrt{\sec(\theta)O_r(j)}\delta T(j)$
$X_{j,6}$	$\sqrt[4]{\sec(\theta)W_r(j)}$	$\sec(\theta)$	$\sec(\theta)O_r^2(j)O_w(j)$
$X_{j,7}$	$\sec(\theta)W_r(j)$	$\sec(\theta)T_w(j)$	$\frac{O_r(j)}{O_w(j)}\sqrt{\sec(\theta)O_r(j)}$
$X_{j,8}$	$(\sec(\theta)W_r(j))^3$	$(\sec(\theta)CO_{2,w}(j))^2$	$\sec(\theta)O_r(j)O_w(j)$
$X_{j,9}$	$(\sec(\theta)W_r(j))^4$	$T_w^3$	$O_r(j)\sec(\theta)\sqrt{O_w(j)\sec(\theta)}$
$X_{j,10}$	$\sec(\theta)W_r(j)\delta T(j) \delta T(j) $	0	$\sec(\theta)O_w(j)$
$X_{j,11}$	$(\sqrt{\sec(\theta)W_r(j)})\delta T(j)$	0	$(\sec(\theta)O_w(j))^2$
$X_{j,12}$	$\frac{\sec(\theta)W_r^2(j)}{W_{tw}(j)}$	0	0
$X_{j,13}$	$\frac{\sqrt{\sec(\theta)W_r(j)W_r(j)}}{W_{tw}(j)}$	0	0
<b>For channels <math>\in [1802,4421]</math></b>		<b>For channels <math>\in [5401,6601]</math></b>	
$X_{j,14}$	$\sec(\theta)CH_{4,r}(j)$	$X_{j,10}\sec(\theta)CO_r(j)$	
$X_{j,15}$	$(\sec(\theta)CH_{4,r}(j))^2\delta T$	$X_{j,11}\sec(\theta)T_w(j)\sqrt{T_r(j)}$	
<b>For channels <math>\in [5022,5400]</math></b>		<b>For channels <math>\in [6602,8061]</math></b>	
$X_{j,14}$	$\sec(\theta)CO_{2,r}(j)$	$X_{j,10}\sec(\theta)T_w(j)\sqrt{T_r(j)}$	
<b>For channels <math>\in [5401,6601]</math></b>			
$X_{j,14}$	$\sec(\theta)CO_{2,r}(j)$		
$X_{j,15}$	$\sec(\theta)CO_r(j)$		

 Table 5: Predictors used for water vapour, CO<sub>2</sub> and ozone

Predictor	CO	N <sub>2</sub> O	CH <sub>4</sub>
$X_{j,1}$	$\sec(\theta) CO_r(j)$	$\sec(\theta)   N2O_r(j)$	$\sec(\theta)   CH4_r(j)$
$X_{j,2}$	$\sqrt{\sec(\theta) CO_r(j)}$	$\sqrt{\sec(\theta) N2O_r(j)}$	$\sqrt{\sec(\theta) CH4_r(j)}$
$X_{j,3}$	$\sec(\theta) CO_r(j) \delta T(j)$	$\sec(\theta) N2O_r(j) \delta T(j)$	$\sec(\theta) CH4_r(j) \delta T(j)$
$X_{j,4}$	$(\sec(\theta) CO_r(j))^2$	$(\sec(\theta) N2O_r(j))^2$	$(\sec(\theta) CH4_r(j))^2$
$X_{j,5}$	$\sqrt{\sec(\theta) CO_r(j)} \delta T(j)$	$N2O_r(j) \delta T(j)$	$CH4_r(j) \delta T(j)$
$X_{j,6}$	$\sqrt[4]{\sec(\theta) CO_r(j)}$	$\sqrt[4]{\sec(\theta) N2O_r(j)}$	$\sqrt[4]{\sec(\theta) CH4_r(j)}$
$X_{j,7}$	$\sec(\theta) CO_r(j) \delta T(j)  \delta T(j) $	$\sec(\theta) N2O_w(j)$	$\sec(\theta) CH4_w(j)$
$X_{j,8}$	$\frac{\sec(\theta) CO_r^2(j)}{CO_w(j)}$	$\sec(\theta) N2O_w(j)$	$CH4_w(j)$
$X_{j,9}$	$\frac{\sqrt{\sec(\theta) CO_r(j) CO_r(j)}}{CO_w(j)}$	$N2O_w(j)$	$(\sec(\theta) CH4_w(j))^2$
$X_{j,10}$	$\frac{\sec(\theta) CO_r^2(j)}{\sqrt{CO_w(j)}}$	$\frac{\sqrt{\sec(\theta) N2O_r(j) N2O_r(j)}}{N2O_w(j)}$	$\sec(\theta) CH4_w(j)$
$X_{j,11}$	$\frac{\sec(\theta) CO_r^2(j)}{\sqrt[4]{CO_w(j)}}$	0	$\frac{\sqrt{\sec(\theta) CH4_r(j) CH4_r(j)}}{CH4_w(j)}$

For channels  $\in$  [1621,2821]

		$X_{j,11} \sec(\theta) CH4_r(j)$	
		$X_{j,12} \sec(\theta) CH4_w(j)$	

For channels  $\in$  [5401,6601]

		$X_{j,11} \sec(\theta) CO_r(j)$	
		$X_{j,12} \sec(\theta)^2 CO_r(j) CO_w(j)$	

Table 6: Predictors used for for CO, N2O and CH4.

Predictor	Fixed Gases
$X_{j,1}$	$\sec(\theta)$
$X_{j,2}$	$\sec^2(\theta)$
$X_{j,3}$	$\sec(\theta) T_r(j)$
$X_{j,4}$	$\sec(\theta) T_r^2(j)$
$X_{j,5}$	$T_r(j)$
$X_{j,6}$	$T_r^2(j)$
$X_{j,7}$	$\sec(\theta) T_{fw}(j)$
$X_{j,8}$	$\sec(\theta) T_{fu}(j)$

Table 7: Predictors used for fixed gases.

$T(l) = [T^{profile}(l+1) + T^{profile}(l)] / 2$		$T^*(l) = [T^{reference}(l+1) + T^{reference}(l)] / 2$	
$W(l) = [W^{profile}(l+1) + W^{profile}(l)] / 2$		$W^*(l) = [W^{reference}(l+1) + W^{reference}(l)] / 2$	
$O(l) = [O^{profile}(l+1) + O^{profile}(l)] / 2$		$O^*(l) = [O^{reference}(l+1) + O^{reference}(l)] / 2$	
$CO(l) = [CO^{profile}(l+1) + CO^{profile}(l)] / 2$		$CO^*(l) = [CO^{reference}(l+1) + CO^{reference}(l)] / 2$	
$CH_4(l) = [CH_4^{profile}(l+1) + CH_4^{profile}(l)] / 2$		$CH_4^*(l) = [CH_4^{reference}(l+1) + CH_4^{reference}(l)] / 2$	
$N_2O(l) = [N_2O^{profile}(l+1) + N_2O^{profile}(l)] / 2$		$N_2O^*(l) = [N_2O^{reference}(l+1) + N_2O^{reference}(l)] / 2$	
$CO_2(l) = [CO_2^{profile}(l+1) + CO_2^{profile}(l)] / 2$		$CO_2^*(l) = [CO_2^{reference}(l+1) + CO_2^{reference}(l)] / 2$	
$T_r(l) = \frac{T(l)}{T^*(l)}$	$\delta T(l) = T(l) - T^*(l)$	$W_r(l) = \frac{W(l)}{W^*(l)}$	$O_r(l) = \frac{O(l)}{O^*(l)}$
$CO_r(l) = \frac{CO(l)}{CO^*(l)}$	$CH_{4r}(l) = \frac{CH_4(l)}{CH_4^*(l)}$	$N_2O_r(l) = \frac{N_2O(l)}{N_2O^*(l)}$	$CO_{2r}(l) = \frac{CO_2(l)}{CO_2^*(l)}$
$T_w(l) = \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] T(i) \} / \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] T^*(i) \}$			
$T_{fu}(l) = \{ \sum_{i=1}^l T(i) \} / \{ \sum_{i=1}^l T^*(i) \}$			
$T_{fw}(l) = \{ \sum_{i=2}^l T(i) \} / \{ \sum_{i=2}^l T^*(i) \}$			
$W_w(l) = \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] W(i) \} / \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] W^*(i) \}$			
$W_{rw}(l) = \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] T(i) W(i) \} / \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] T^*(i) W^*(i) \}$			
$O_w(l) = \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] O(i) \} / \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] O^*(i) \}$			
$CO_w(l) = \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] CO(i) \} / \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] CO^*(i) \}$			
$CO_{rw}(l) = \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] T(j) CO(i) \} / \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] T^*(j) CO^*(j) \}$			
$CH_{4w}(l) = \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] CH_4(i) \} / \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] CH_4^*(i) \}$			
$CH_{4rw}(l) = \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] T(j) CH_4(i) \} / \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] T^*(j) CH_4^*(j) \}$			
$N_2O_w(l) = \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] N_2O(i) \} / \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] N_2O^*(i) \}$			
$N_2O_{rw}(l) = \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] T(j) N_2O(i) \} / \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] T^*(j) N_2O^*(j) \}$			
$CO_{2w}(l) = \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] CO_2(i) \} / \{ \sum_{i=1}^l P(i) [P(i) - P(i-1)] CO_2^*(i) \}$			

Table 8: Definition of profile variables used in predictors defined in tables 5, 6 and 7.

The  $P(i)$ 's are the values of the pressure at each level.  $T^{profile}(l)$ ,  $W^{profile}(l)$ ,  $O^{profile}(l)$ ,  $CO^{profile}(l)$ ,  $CH_4^{profile}(l)$ ,  $N_2O^{profile}(l)$  and  $CO_2^{profile}(l)$  are the temperature and variable gases mixing ratio profiles.  $T^{reference}(l)$ ,  $W^{reference}(l)$ ,  $O^{reference}(l)$ ,  $CO^{reference}(l)$ ,  $CH_4^{reference}(l)$ ,  $N_2O^{reference}(l)$  and  $CO_2^{reference}(l)$  are corresponding reference profiles. For these variables  $l$  refers to the  $l$ th level; otherwise  $l$  is the  $l$ th layer, i.e. the layer below the  $l$ th level (layers are numbered from 1 to 100). Note that we take  $P(0) = 2P(1) - P(2)$  and  $T_{r1} = (P(2) - P(1))P(1)T(1) / T^*(1)$ . Here  $\theta$  is the zenith angle.

## 6. Performance of the fast model for simulation of IASI and AIRS radiances: results for the set of training profiles

The accuracy of RTTOV simulations can be assessed by a comparison of the transmittances and radiances computed by the fast model with the corresponding values from LBL models in different ways. Firstly the fast model transmittance profiles and top of the atmosphere radiances computed for the dependent set of profiles used to train the fast model can be compared with the LBL model equivalents to determine the accuracy of the fast model itself. Secondly a set of profiles independent of the regression coefficients can be used to allow uncertainties from different type of profiles to be included. The comparison of transmittances is more useful to understand how the model performs and to see where it needs to be improved, but the comparison of radiances is the most important as the radiances are what will be used. The analysis of the results discussed below concentrates on the error of RTTOV in terms of the bias, standard deviation and rms of the radiance and transmittances differences between the fast and LBL radiative transfer models.

The simulation of the layer optical depth is the essence of a regression based fast radiative transfer model. For the 83 profile dependent set the fast model transmittances for the IASI channels were compared to LBL equivalents computed using LBLRTM for the six scan angles used in the regression ( $0^\circ$  to  $64^\circ$ ). Results are shown in Fig. 33 where the maximum value of the rms of the difference between fast model and LBL layer-to-space transmittances is shown (the water vapour continuum contribution is not included).

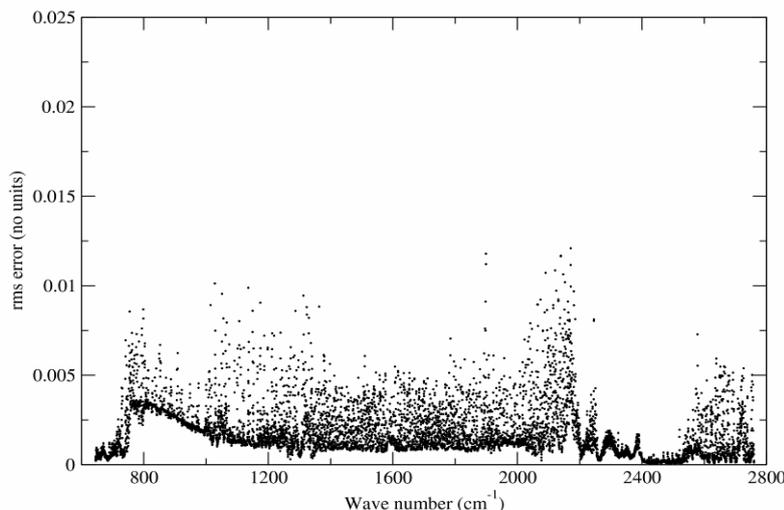


Figure 33: Maximum value of the rms of the difference between fast model and LBLRTM layer-to-space transmittances for 83 diverse profiles and 6 viewing angles. Results are shown for the IASI channels.

Figure 33 shows that the largest errors are found in the regions around  $800\text{ cm}^{-1}$ ,  $1200\text{ cm}^{-1}$  and  $2100\text{ cm}^{-1}$ . Errors in the  $800\text{ cm}^{-1}$  region are associated with interfering  $\text{CO}_2$  and  $\text{H}_2\text{O}$  lines whereas errors around the  $1200$  and  $2100\text{ cm}^{-1}$  region are in general associated with tropospheric water vapour channels characterized by a moderate absorption. It should also be noted that larger errors around  $1000\text{ cm}^{-1}$  are associated with ozone lines and that for a number of channels around  $1350\text{ cm}^{-1}$  some of the largest errors are associated with interfering  $\text{H}_2\text{O}$  and  $\text{CH}_4$  lines whereas for a number of channels around  $2100\text{ cm}^{-1}$  some of the largest errors are associated with interfering  $\text{H}_2\text{O}$ ,  $\text{N}_2\text{O}$  and  $\text{CO}$  lines.

The accuracy of the water continuum fast model was also tested by comparing the fast model transmittances with MTK\_CKD\_2.1 equivalents for the 83 training profiles and the six viewing angles used in the regression. Results are shown in Fig. 34 where the maximum value of the rms of the difference between fast model and MTK\_CKD\_2.1 layer-to-space water vapour continuum transmittances is plotted. Errors are in general less than 0.005 and are consistently larger in the spectral regions where the contribution of the foreign-continuum is predominant.

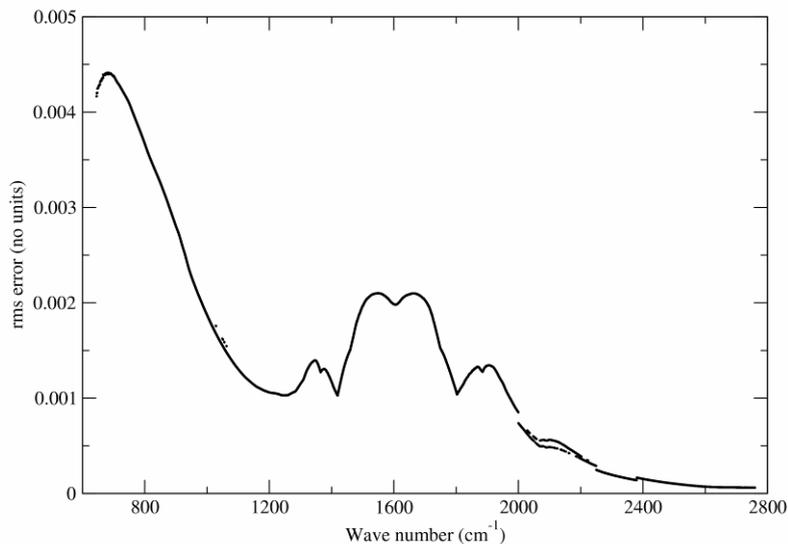


Figure 34: Maximum value of the rms of the difference between fast model and MTK\_CKD\_2.1 layer-to-space water vapour continuum transmittances for 83 diverse profiles and 6 viewing angles. Results are shown for the IASI channels.

The ability of the fast model to reproduce the LBL radiances has been assessed by comparing radiances computed using fast model transmittances and LBL equivalents. Results are shown in figure 35 where the bias, standard deviation and rms of the difference between fast model and LBL radiances is plotted for the IASI channels in units of equivalent black body temperature. From figure 35 it can be seen that biases are typically less than 0.05 K (absolute value) and contribute only fractionally to the rms error that, for the vast majority of the channels, is less than 0.1 K. Larger rms errors are observed for a small fraction of channels in the spectral regions where, as discussed previously, the simulation of the LBL transmittances is more problematic.

It should be stressed that in the 2000 to 2250  $\text{cm}^{-1}$  region, results shown in figure 35 have been obtained by using a revised version of the fast transmittance model. For wave numbers greater than 2000  $\text{cm}^{-1}$ , radiances are affected by solar radiation and consequently in this region we use a dedicated fast transmittance model that allows us to reproduce LBL transmittances for the wide range of angles required for the computation of the solar term (Matricardi 2003). However, between 2000 and 2380  $\text{cm}^{-1}$  the contamination of solar radiation can be important (of the order of a few Kelvin) only for a limited number of channels. In fact the region between 2250 and 2380  $\text{cm}^{-1}$  it totally opaque to solar radiation and no computation of the solar term is performed in RTTOV whereas between 2000 and 2250  $\text{cm}^{-1}$  the short-wave fast transmittance model is used for all the channels although, as mentioned before, only a fraction are affected by solar radiation. The predictors used in the short-wave water vapour fast transmittance model can reproduce very accurately the LBL transmittances for the channels contaminated by solar radiation but have a somewhat lesser skill for the more opaque channels where solar radiation is not important. In the 2000 to 2250  $\text{cm}^{-1}$  region this can result in larger errors as shown in figure

36 where the fit of fast model to LBL radiances is shown using the standard RTTOV fast transmittance model (i.e. short-wave predictors are used for all the channels in the 2000 to 2250  $\text{cm}^{-1}$  region). To improve the performance of RTTOV in this region we have identified all the channels that are not significantly affected by solar radiation (i.e. the signal is less than 0.1 K) and replaced the short-wave water vapour predictors with the standard predictors since for these channels there is no need to perform the computation of the solar term. As shown in figure 35, this result in a better fit to the LBL radiances. Finally, in figure 37 we show the bias, standard deviation and rms of the difference between fast model and LBL radiances for the AIRS channels in units of equivalent black body temperature.

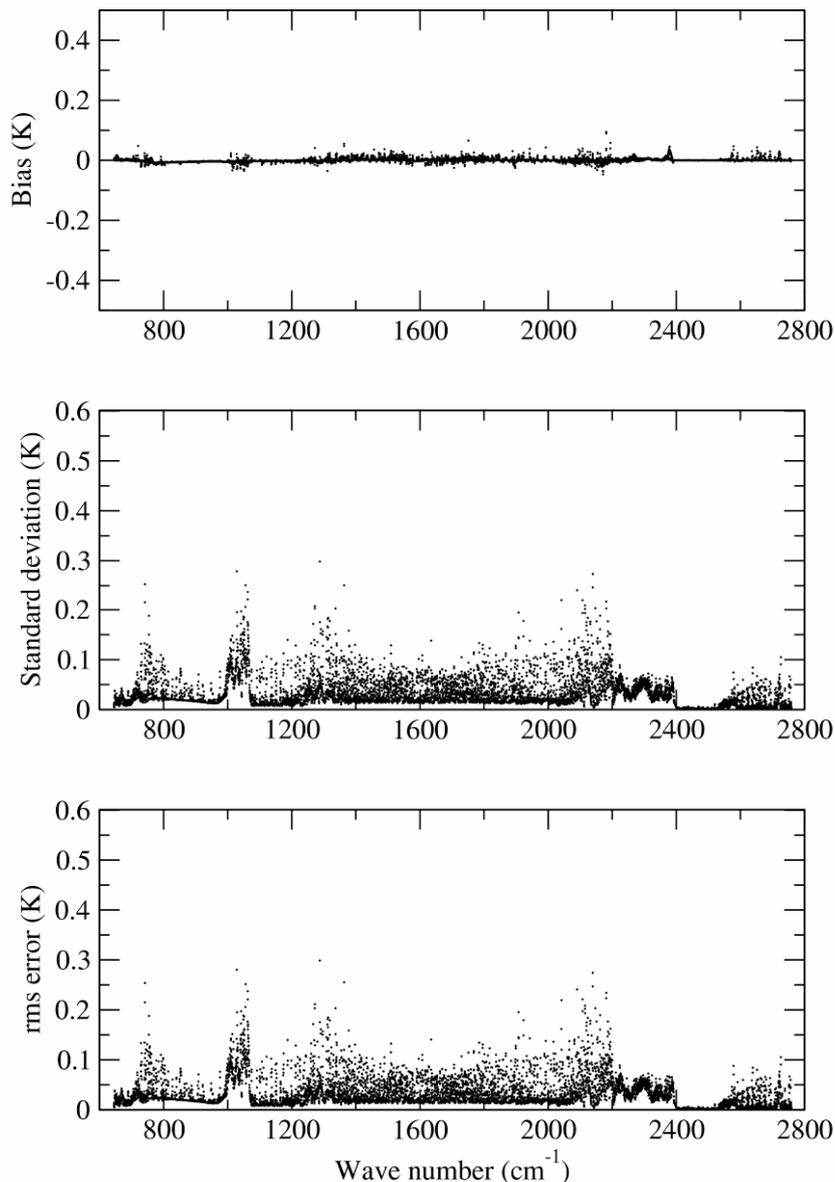


Figure 35: Mean value, standard deviation and rms of the difference between fast model and LBLRTM computed brightness temperatures for 83 diverse profiles and 6 viewing angles. Results are shown for the IASI channels.

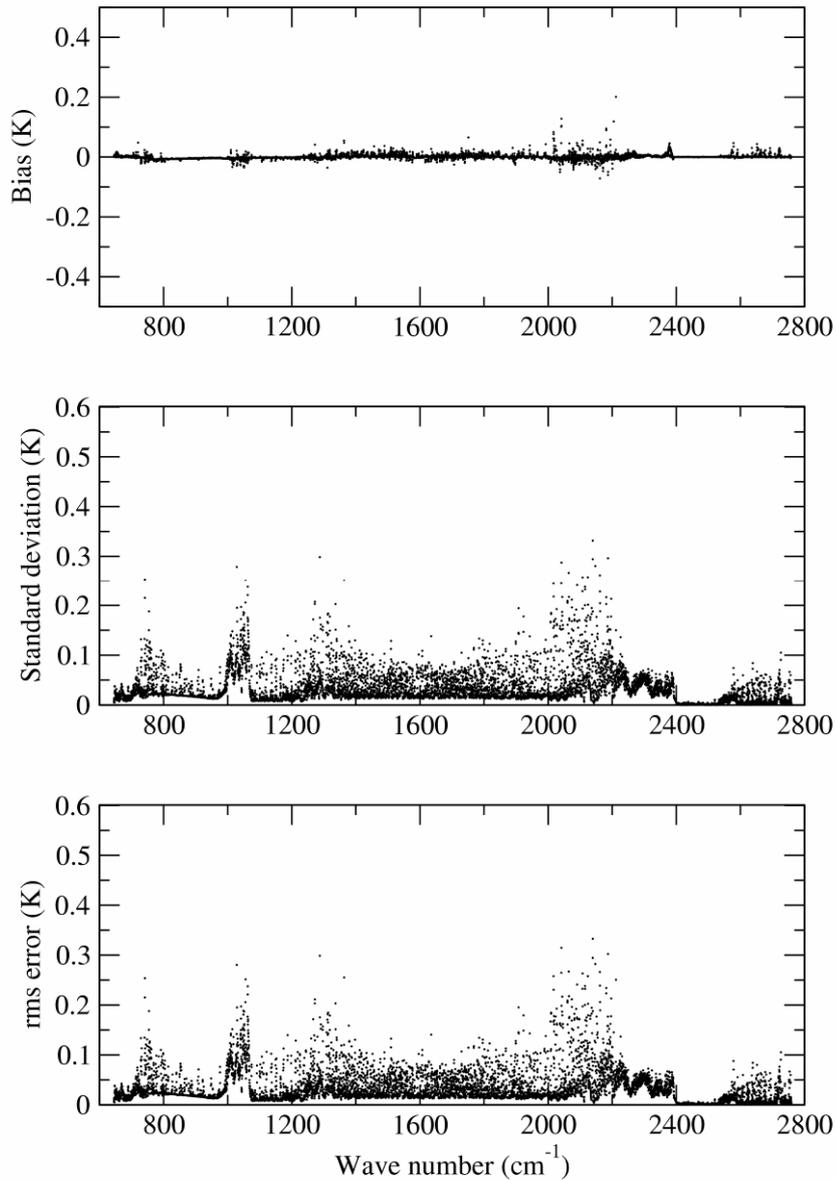


Figure 36: Mean value, standard deviation and rms of the difference between fast model and LBLRTM computed brightness temperatures for 83 diverse profiles and 6 viewing angles. Results are shown for the IASI channels using the standard fast transmittance model in the 2000 to 2250 cm<sup>-1</sup> region.

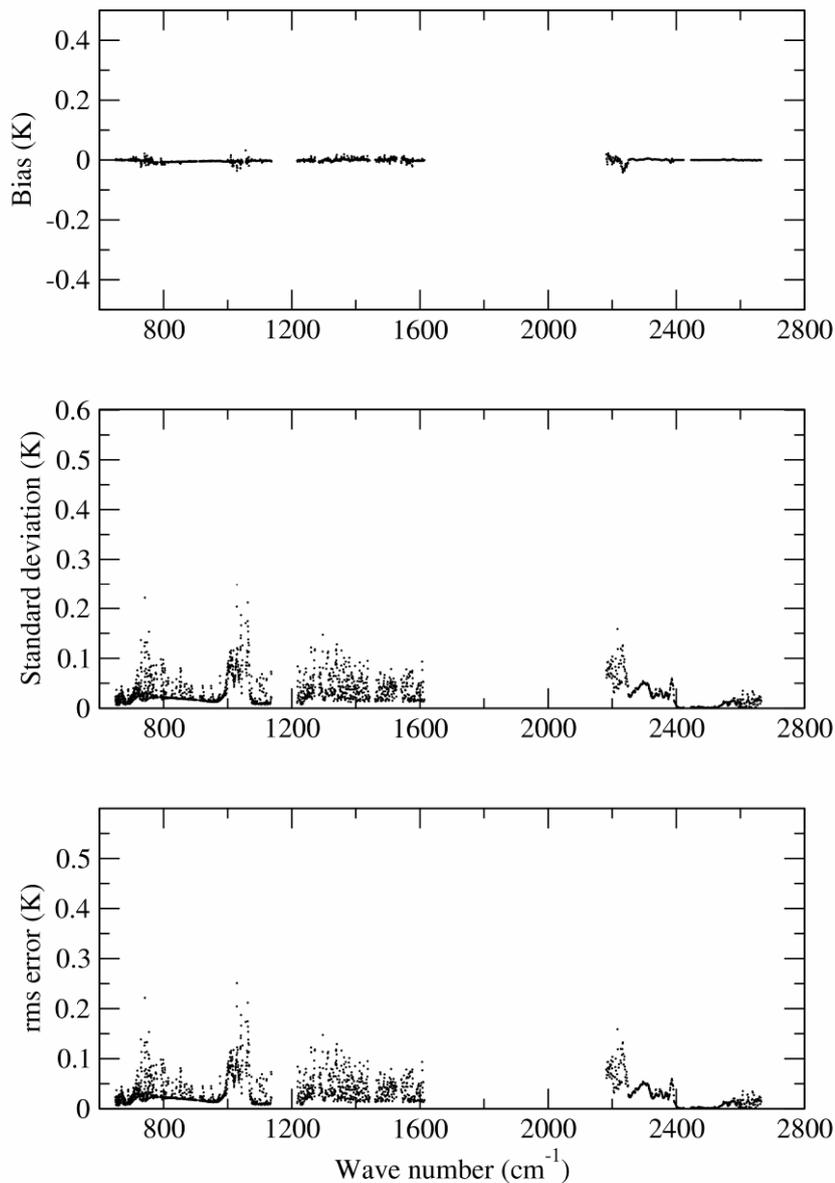


Figure 37: Mean value, standard deviation and rms of the difference between fast model and LBLRTM computed brightness temperatures for 83 diverse profiles and 6 viewing angles. Results are shown for the AIRS channels.

Results shown in figures 35 and 36 have been rearranged in figures 38 and 39 where the binned distribution of the rms error is plotted for IASI and AIRS channels respectively. Results for the two instruments are very similar and it can be seen that the channel distribution displays a very weak dependence on the viewing angle geometry. In fact, irrespective of the viewing geometry, almost 98% of the channels have rms errors less than 0.1 K.

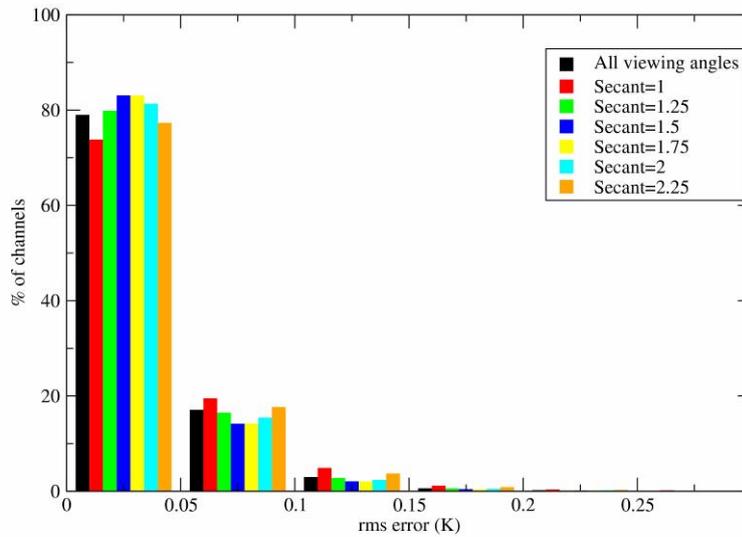


Figure 38: Histogram of the distribution of channels with rms error for six viewing angles. Results are shown for the IASI channels.

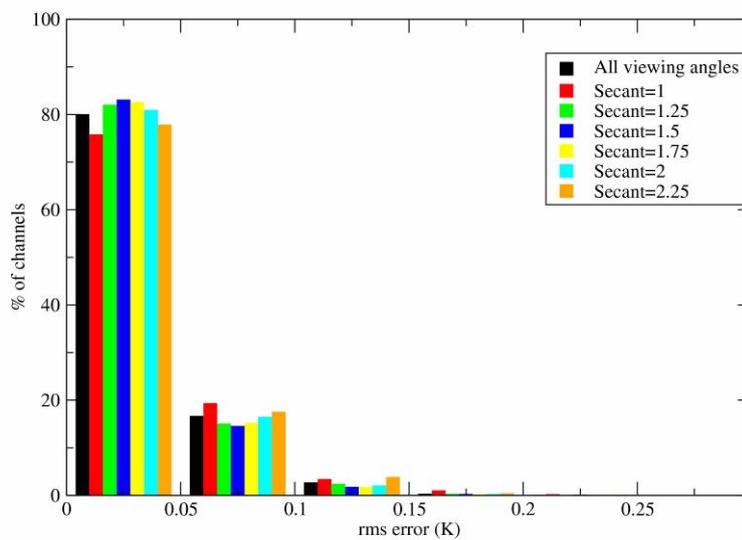


Figure 39: Histogram of the distribution of channels with rms error for six viewing angles. Results are shown for the AIRS channels.

## 7. Results for the independent set of profiles

A complete validation of the fast model requires the use of set of profiles that is independent to the regression coefficients. To this end we have utilized the 43 profile set used to generate previous LBL databases at ECMWF (Matricardi 2003). The main reason we have made this choice is that in this dataset the concentration of CO<sub>2</sub>, N<sub>2</sub>O, CO and CH<sub>4</sub> is variable. However, the dataset does not feature ozone that was added separately from the Fortuin and Langematz climatology (1994) depending on season and latitude. New LBL transmittances were

generated for this profile set using LBLRTM and the results used to compute the statistics of the error for IASI and AIRS.

Results for the fast transmittance model are shown in figures 40 and 41 for line absorption and continuum absorption respectively. The use of the independent profiles results in larger errors for the line absorption model above all for the channels that feature water vapour absorption lines whereas errors for the continuum model have the same magnitude or smaller magnitude as in the region around the centre of the water vapour band at  $1594\text{ cm}^{-1}$ . This latter result maybe indicative of the fact the highly accurate continuum model responds differently to the presence of a relative smaller number of moist profiles in the independent set.

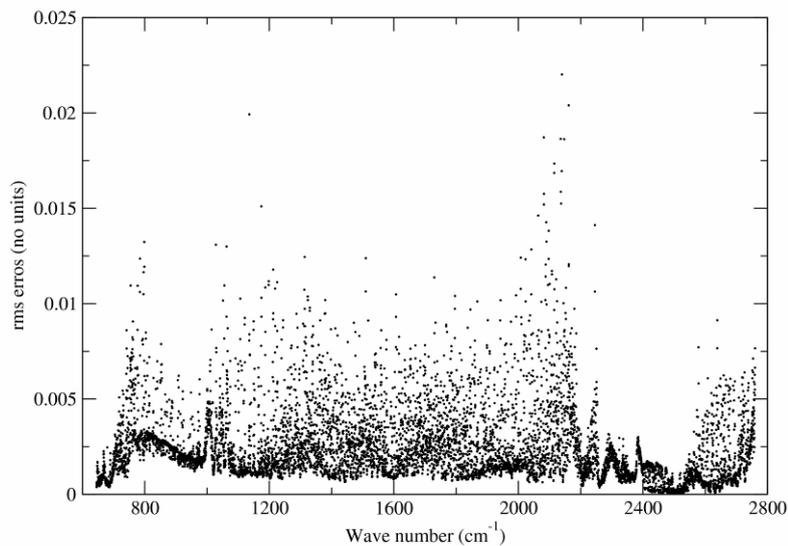


Figure 40: Maximum value of rms of the difference between fast model and LBLRTM layer-to-space transmittances for 43 independent profiles and 6 viewing angles. Results are shown for the IASI channels.

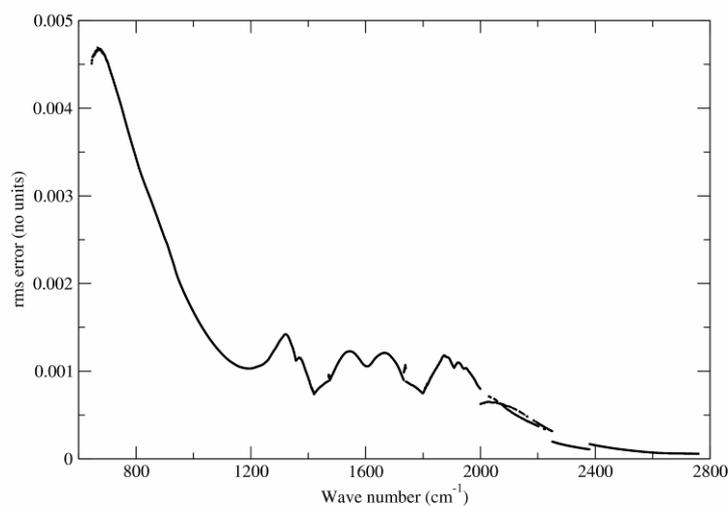


Figure 41: Maximum value of the rms of the difference between fast model and MTK\_CKD\_2.1 layer-to-space water vapour continuum transmittances for 43 diverse profiles and 6 viewing angles. Results are shown for the IASI channels

As done previously, transmittance errors can be translated into brightness temperature errors. Results for IASI are shown in figure 42 where the mean value (bias), standard deviation and rms of the difference between fast model and LBL radiances is plotted in units of equivalent black body brightness temperature. Figure 42 shows that errors are larger for the independent set. This is true above all for channels in spectral regions dominated by water vapour although a few outliers can be observed in the ozone band around  $1020\text{ cm}^{-1}$  and in the region around  $2100\text{ cm}^{-1}$  where water vapour lines interfere with  $\text{O}_3$ ,  $\text{N}_2\text{O}$  and  $\text{CO}$  lines. However it should be pointed out that rms errors are still below  $0.15\text{K}$  for the vast majority of the channels. For AIRS, results are shown in figure 43. Results in figure 42 for the  $2000$  to  $2250\text{ cm}^{-1}$  region were obtained using the revised transmittance model mentioned earlier. For completeness, results obtained using the standard transmittance model are shown in figure 44 where the inflation of the error resulting from the use of this model can be clearly seen. Finally, the binned distribution of the fitting errors is shown in figures 45 and 46 for IASI and AIRS respectively. As mentioned earlier, these figures clearly demonstrate that the rms error is below  $0.15\text{ K}$  for 95% of the channels for both instruments. Also noticeable is the fact that the dependence on the viewing angle geometry is very similar to that observed for the training set.

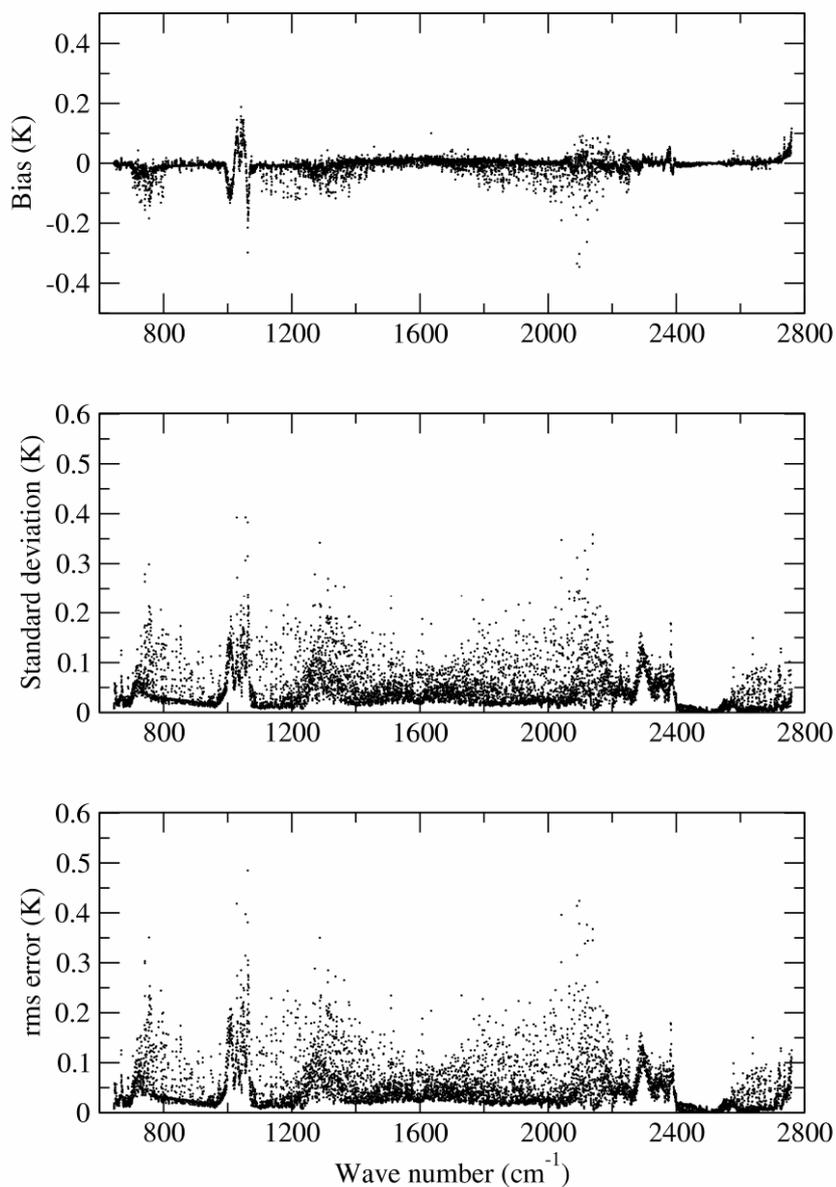


Figure 42: Mean value, standard deviation and rms of the difference between fast model and LBLRTM computed brightness temperatures for 43 diverse independent profiles and 6 viewing angles. Results are shown for the IASI channels.

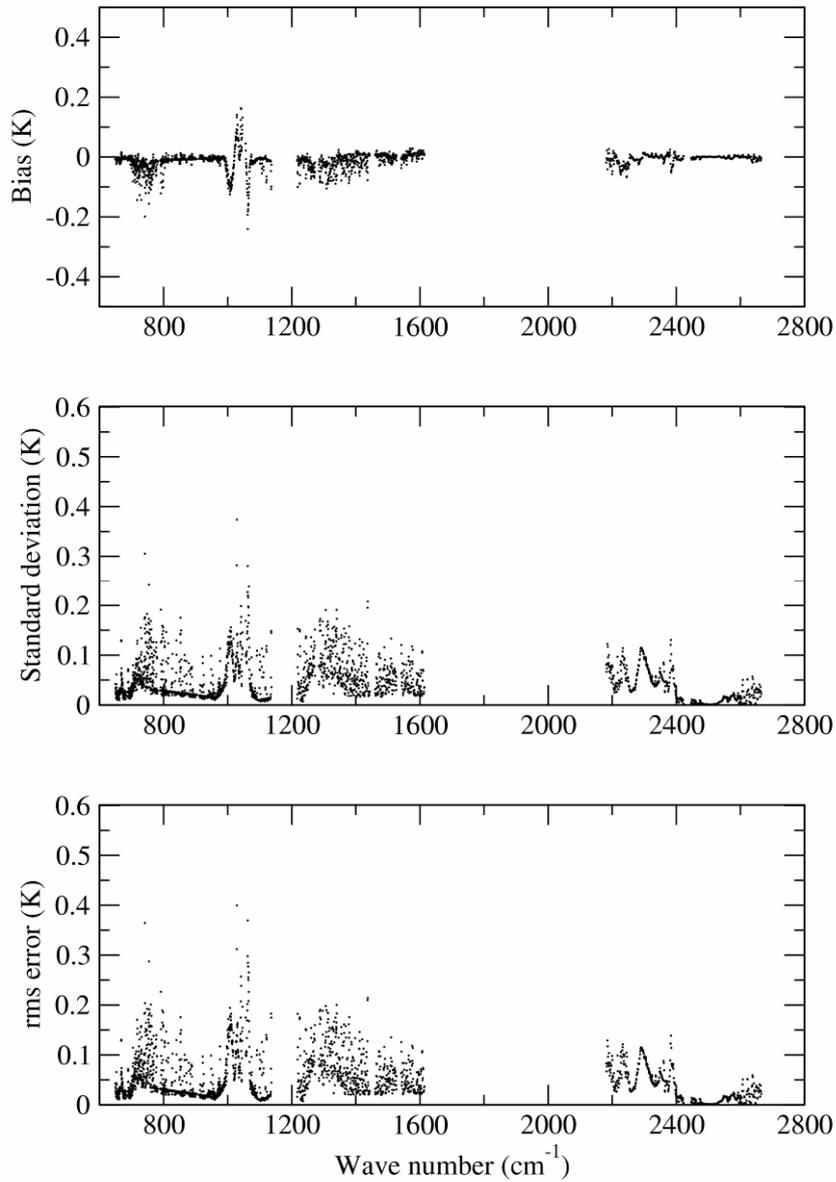


Figure 43: Mean value, standard deviation and rms of the difference between fast model and LBLRTM computed brightness temperatures for 43 diverse independent profiles and 6 viewing angles. Results are shown for the AIRS channels.

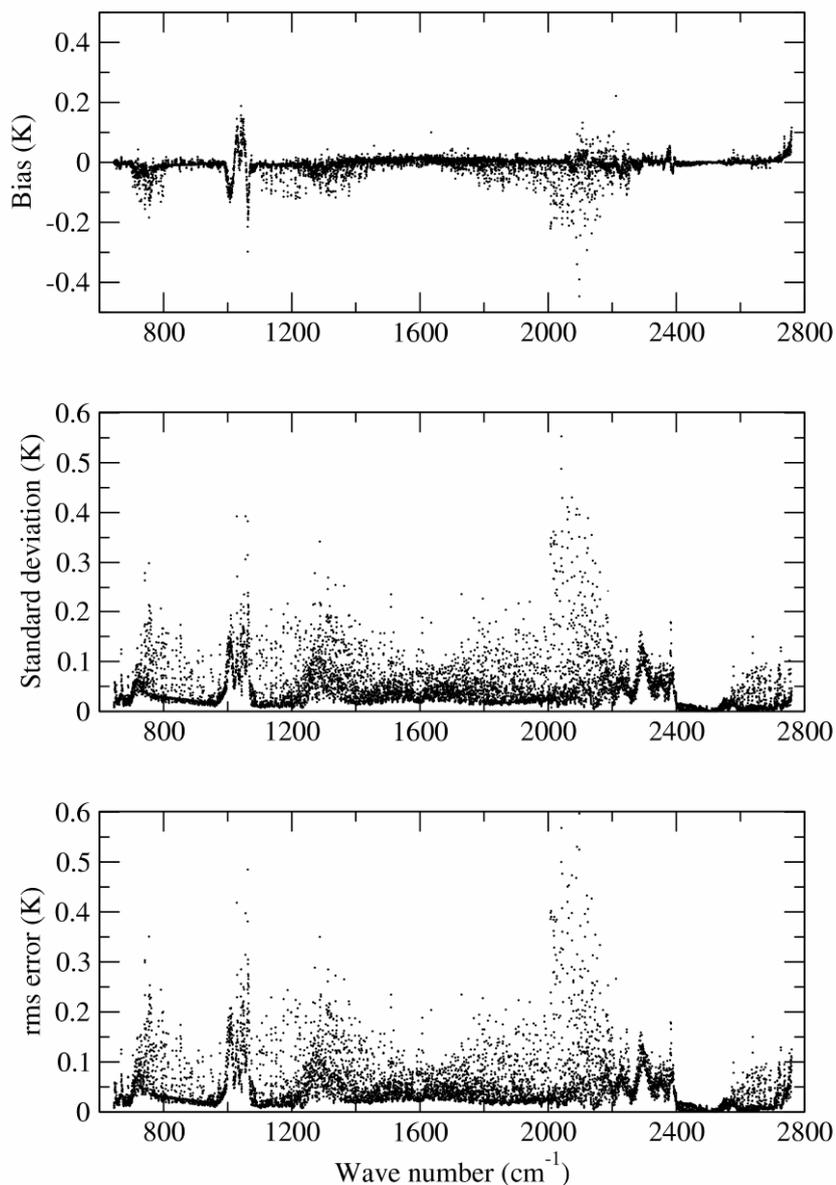


Figure 44: Mean value, standard deviation and rms of the difference between fast model and LBLRTM computed brightness temperatures for 43 diverse independent profiles and 6 viewing angles. Results are shown for the IASI channels using the standard fast transmittance model in the 2000 to 2250 cm<sup>-1</sup> region.

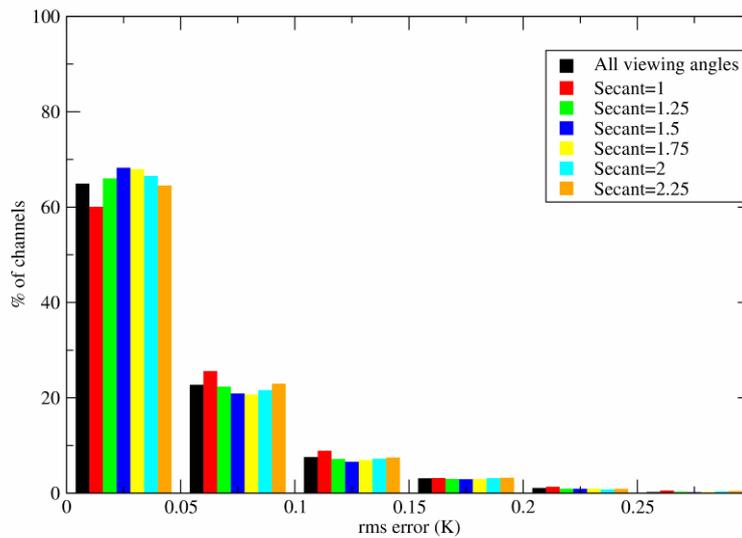


Figure 45: Histogram of the distribution of channels with rms error for six viewing angles. Results are shown for the IASI channels and the independent set.

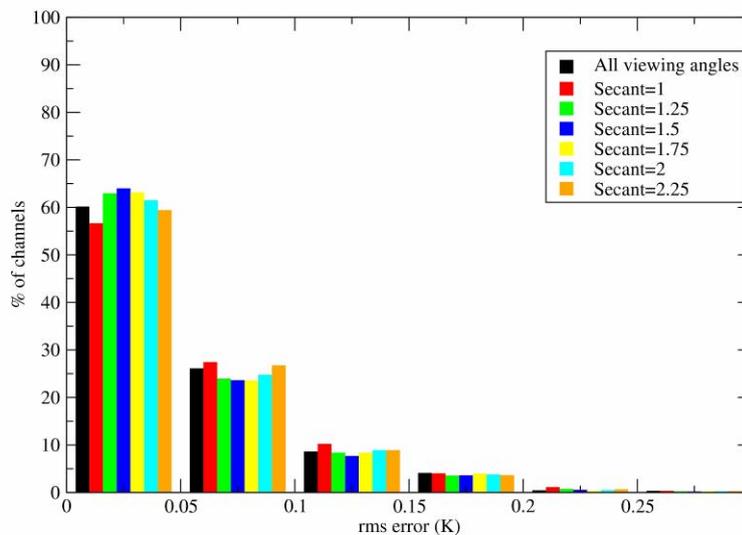


Figure 46: Histogram of the distribution of channels with rms error for six viewing angles. Results are shown for the AIRS channels and the independent set.

## 8. Conclusions

New RTTOV regression coefficients have been generated for IASI and AIRS from a database of line-by-line transmittances computed using the LBLRTM line-by-line model and a new set of 83 training profiles.

The LBL database was computed using version 11.1 of LBLRTM utilizing molecular parameters obtained by merging in different regions of the infrared spectrum data available from the HITRAN2000, HITRAN2004 and GEISA2003 databases. Line data for CO<sub>2</sub> have not been merged and are taken from HITRAN2000 since they must be consistent with the line mixing coefficients used in LBLRTM. Three new species (NO<sub>2</sub>, SO<sub>2</sub>, and NO) have been included in the computations that were not featured in any of the LBL databases previously released at

ECMWF. Regression coefficients for the water continuum have been computed for two different versions of the water continuum model, MTK\_CKD\_2.1 and CKD\_2.4.

To compute the database of line-by-line transmittances we have generated a new set of training profiles selected from a dataset of atmospheric profiles of temperature, water vapor and ozone produced using the operational suite of the ECMWF forecasting system during the period July 2006-June 2007. To optimize the fit of the line-by-line optical depths to the curve defined by the predictors used in the regression, the training profiles have been selected to cover as uniformly as possible the range of observed values at each pressure level. To achieve this we have performed a constrained random selection of the profiles from the ECMWF analyses. The training set is made of 83 profiles of which two profiles are the envelope of maximum and minimum values of the original set of ECMWF profiles and the 83<sup>th</sup> profiles is the average of the 82 profiles to serve as a reference in the regression. The profiles of temperature, water vapor and ozone have been supplemented by profiles of CO<sub>2</sub>, CO and CH<sub>4</sub> generated by a number of forecast experiments performed at ECMWF within the context of the GEMS project and by profiles of N<sub>2</sub>O generated by extrapolating to the surface profiles retrieved from radiances measured by the CLAES instrument. Consequently, the concentration of these atmospheric constituents is allowed to vary in the fast model.

Results for the dependent set of profiles used to train the fast model show that RTTOV can reproduce LBL radiances to a degree of accuracy that is below 0.1 K rms for 98 % of the channels of IASI and AIRS. Errors larger than 0.1K are observed for a small fraction of ozone channels, for tropospheric water vapor channels with moderate absorption and for channels where absorption lines due to CO<sub>2</sub>, CH<sub>4</sub>, N<sub>2</sub>O and CO interfere with absorption lines due to H<sub>2</sub>O. Although the statistics of the error for a set of profiles independent to the regression coefficients shows larger values, for the vast majority of the channels errors are still below 0.15 K rms. A number of significant outliers can be observed in the ozone band and in the regions where water vapor lines interfere with CH<sub>4</sub>, CO and N<sub>2</sub>O. For either the dependent or the independent set the statistics of the error does not show any marked dependency on the viewing geometry

Finally, to improve the simulation of line-by-line transmittances in the region between 2000 and 2250 cm<sup>-1</sup> we have replaced the standard RTTOV transmittance model with a new model that uses the predictors dedicated to the evaluation of the transmittances for the solar term only for those channels that are significantly contaminated by solar radiation.

## Acknowledgements

We want to thank S. Di Michele, ECMWF, for providing the revised database of atmospheric profiles sampled from the ECMWF analyses, R. Engelen, S. Serrar and J. Flemming, ECMWF, for making available forecast data from the GEMS project experiments and C. Clerbaux, Service d'Aéronomie, for providing trace gas profiles for NO, HNO<sub>3</sub> and SO<sub>2</sub>.

## References

- Chevallier, F, 2003, "Sampled database of 60-level atmospheric profiles from the ECMWF analyses", *NWP SAF Report No. NWPSAF-EC-TR-001*.
- Chevallier, F., Di Michele, S. and McNally, A.P., 2006, "Diverse profile datasets from the ECMWF 91-level short-range forecasts", *NWP SAF Report No. NWPSAF-EC-TR-010*.
- S.A. Clough, F.X. Kneizys, L.S., Rothman and W.O. Gallery, 1981: Atmospheric transmittance and radiance: FASCOD1B. *Proc. Soc. Photo Opt. Instrum. Eng*; 277, pp. 152-166.
- S.A. Clough, F.X. Kneizys and R.W. Davis, 1989: Line shape and the water vapour continuum. *Atmospheric Research*, **23**, 229-241.
- Clough, S.A., M.J. Iacono and J.-L. Moncet, 1992: Line by line calculation of atmospheric fluxes and cooling rates: application to water vapor. *J. Geophys. Res.*, **98**, 15761-15785.
- De Souza-Machado, S., Strow, L.L, Motteler, H.E., and Hannon, S.E., 2002, "kCARTA: An atmospheric radiative transfer algorithm using compressed lookup tables", *Tech. Rep. University of Maryland Baltimore County, Department of Physics*, <http://asl.umbc.edu/rta/lbl.html>.
- Edwards, D.P., 1992, "GENLN2. A general line-by-line atmospheric transmittance and radiance model", NCAR Technical note NCAR/TN-367+STR.
- Globalview-CO<sub>2</sub>: Cooperative atmospheric data integration project-carbon dioxide. CD-ROM, NOAA/CMDL, Boulder, Colorado. Also available on Internet via anonymous FTP to <ftp.cmdl.noaa.gov>, path:ccg/CO<sub>2</sub>/GLOBALVIEW, 1997.
- N. Jacquinet-Husson, N.A. Scott, A. Chedin, K. Garceran, R. Armante, A.A. Chursin, A. Barbe, M. Birk, L.R. Brown, C. Camy-Peyret, C. Claveau, C. Clerbaux, P.F. Coheur, V. Dana, L. Daumont, M.R. Debacker-Barilly, J.M. Flaud, A. Goldman, A. Hamdouni, M. Hess, D. Jacquemart, P. Kopke, J.Y. Mandin, S. Massie, S. Mikhailenko, V. Nemtchinov, A. Nikitin, D. Newnham, A. Perrin, V.I. Perevalov, L. Regalia-Jarlot, A. Rublev, F. Schreier, I. Schult, K.M. Smith, S.A. Tashkun, J.L. Teffo, R.A. Toth, V.I. Tyuterev, J. Vander Auwera, P. Varanasi, G. Wagner, 2005: The 2003 edition of the GEISA/IASI spectroscopic database. *J. Quant. Spectrosc. Radiat. Transfer*, **95**, 429-467.
- W.J. Lafferty, A.M. Solodov, A. Weber, W.B. Olson and J.-M. Hartmann, 1996: Infrared collision-induced absorption by N<sub>2</sub> near 4.3 μm for atmospheric applications: measurements and empirical modelling. *Applied Optics*, **35**, 5911-5917.
- Matricardi, M. and Saunders, R., 1999, "Fast radiative transfer model for simulation of infrared atmospheric sounding interferometer radiances", *Applied Optics*, 38, pp. 5679-5691.
- Matricardi, M, 2003, "RTIASI-4, a new version of the ECMWF fast radiative transfer model for the infrared atmospheric sounding interferometer", *ECMWF Technical Memorandum 425* (available from: <http://www.ecmwf.int/publications/>)

Matricardi, M, 2005, "The inclusion of aerosols and clouds in RTIASI, the ECMWF fast radiative transfer model for the infrared atmospheric sounding interferometer", *ECMWF Technical Memorandum 474* (available from: <http://www.ecmwf.int/publications/>)

Matricardi, M, 2007, "An inter-comparison of line-by-line radiative transfer models", *ECMWF Technical Memorandum 525* (available from: <http://www.ecmwf.int/publications/>)

F. Niro, K. Jucks and J.-M. Hartmann, 2005: Spectra calculations in central and wing regions of CO<sub>2</sub> IR bands. IV: software and database for the computation of atmospheric spectra. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **95**, 469-481.

Reber, C.A., Trevathan, C.E., McNeal, R.J., and Luther, M.R., 1993, "The Upper Atmosphere Research Satellite (UARS) Mission", *J. Geophys. Res.*, **98**, D6, 10643-10647.

L.S. Rothman, A. Barbe, D. Chris Benner, L.R. Brown, C. Camy-Peyret, M.R. Carleer, K. Chance, C. Clerbaux, V. Dana, V.M. Devi, A. Fayt, J.-M. Flaud, R.R. Gamache, A. Goldman, D. Jacquemart, K.W. Jucks, W.J. Lafferty, J.-Y. Mandin, S.T. Massie, V. Nemtchinov, D.A. Newnham, A. Perrin, C.P. Rinsland, J. Schroeder, K.M. Smith, M.A.H. Smith, K. Tang, R.A. Toth, J. Vander Auwera, P. Varanasi, K. Yoshino, 2003: The HITRAN molecular spectroscopic database: edition of 2000 including updates through 2001. *J. Quant. Spectrosc. Radiat. Transfer*, **82**, 5-44.

L.S. Rothman, D. Jacquemart, A. Barbe, D. Chris Benner, M. Birk, L.R. Brown, M.R. Carleer, C. Chackerian Jr., K. Chance, L.H. Coudert, V. Dana, V.M. Devi, J.-M. Flaud, R.R. Gamache, A. Goldman, J.-M. Hartmann, K.W. Jucks, A.G. Maki, J.-Y. Mandin, S.T. Massie, J. Orphal, A. Perrin, C.P. Rinsland, M.A.H. Smith, J. Tennyson, R.N. Tolchenov, R.A. Toth, J. Vander Auwera, P. Varanasi, G. Wagner, 2005: The HITRAN 2004 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transfer*, **96**, 139-204.

F. Thibault, V. Menoux, R. Le Doucen, L. Rosenman, J.-M. Hartmann, and Ch. Boulet, 1996: Infrared collision-induced absorption by O<sub>2</sub> near 6.4 microns for atmospheric applications: measurements and empirical modeling, *Appl. Optics*, **35**, 5911-5917.