The EUMETSAT Network of Satellite Application Facilities

NWP SAF
Numerical Weather Prediction

Document NWPSAF-EC-TR-015

Version 1.0

21 February 2012

Diverse profile database of aerosol and trace gas concentrations from the Monitoring Atmospheric Composition and Climate short-range forecasts

Reima Eresmaa, Angela Benedetti and Anthony P. McNally

European Centre for Medium-range Weather Forecasts
Shinfield Park, Reading, RG2 9AX, United Kingdom

Met Office

ECMWF

Royal Netherlands Meteorological Institute
Ministry of Transport, Public Works and Water Management

METEO FRANCE
Toujours un temps d'avance

Diverse profile database of aerosol and trace gas concentrations from MACC short-range forecasts

Doc ID : NWPSAF-EC-TR-015
Version : 1.0
Date : 21 February 2012

This documentation was developed within the context of the EUMETSAT Satellite Application Facility on Numerical Weather Prediction (NWP SAF), under the Cooperation Agreement dated 1 December, 2006, between EUMETSAT and the Met Office, UK, by one or more partners within the NWP SAF. The partners in the NWP SAF are the Met Office, ECMWF, KNMI and Météo France.

Change record			
Version	Date	Author / changed by	Remarks
1.0	21.2.12	Reima Eresmaa, Angela Benedetti and Anthony P. McNally	

Diverse profile database of aerosol and trace gas concentrations from the Monitoring Atmospheric Composition and Climate short-range forecasts

Reima Eresmaa, Angela Benedetti and Anthony P. McNally

European Centre for Medium-range Weather Forecasts
Shinfield Park, Reading, RG2 9AX, United Kingdom

Overview

A database consisting of vertical profiles of atmospheric parameters has been created at the European Centre for Medium-range Weather Forecasts (ECMWF) using short-range forecasts originally produced for the Monitoring Atmospheric Composition and Climate (MACC) project. The database consists of 40,000 vertical profiles, given on 60 model levels, and it is divided in ten subsets, each representing spatio-temporal variability of one sampling variable in the global modelling system. The ten sampled variables include temperature and specific humidity, mixing ratios of ozone, carbon dioxide and methane, as well as aerosol mixing ratios of sulphate, organic matter, black carbon, sea salt, and desert dust.

1 Introduction

Following development of their operational NWP system, the European Centre for Medium-range Weather Forecasts (ECMWF) has in the past published several databases consisting of vertical profiles of meteorological parameters in heterogeneous atmospheric conditions. Such databases aim at representing the full spectrum of meteorological situations that occur in the model atmosphere. The previous release of the diverse profile database (Chevallier et al., 2006; hereafter IFS-91 database) consists of 25,000 vertical profiles, given on the 91-level grid of the Integrated Forecast System. The IFS-91 database is constituted by five subsets sampling atmospheric fields of temperature, specific humidity, ozone mixing ratio, cloud condensate, and precipitation.

Active research towards improved understanding of atmospheric chemical processes is taking place at ECMWF. In particular, ECMWF coordinates the EU FP7 project “Monitoring Atmospheric Composition and Climate” (MACC). As part of the project activities, global analyses and forecasts are produced at ECMWF in near-real-time for trace gas abundances and aerosol concentrations. This report introduces a new diverse profile database, that has been compiled using short-range forecast data, that is originally produced and archived for the MACC project. This is the first time to include trace gas and aerosol mixing ratio data in the history of ECMWF diverse profile databases.

Properties of the new MACC-based profile database (hereafter MACC-60 database, as profile data in the new database is given on 60 model levels) are summarized in this report. In Section 2, the method used for collecting the profile data is briefly reviewed. Properties of the MACC-60 database are discussed in Section 3, including listing of atmospheric and surface parameters included in the database.

2 Sampling method

The sampling method described in Chevallier et al. (2006) is applied as such. In this method, difference between vertical profiles s_i and s_j is quantified as

$$D(s_i, s_j) \quad = \quad \sum_{k=1}^{K} \sum_{m=1}^{M} \left(\frac{\theta_{ik}(m) - \theta_{jk}(m)}{\sigma_k(m)} \right)^2, \tag{1}$$

where k and m, respectively, are indices of variable and model level, while K and M are numbers of variables and model levels, $\theta_{ik}(m)$ represents value of variable k on model level m in profile i, and $\sigma_k(m)$ represents standard deviation of variable k on model level m. The selection process is started by picking up a random profile from the input pool S_I, that consists of all available profiles, and saving this profile as s_1 in the output pool S_O. On subsequent steps, profiles available in S_I are processed one-by-one in a random order. A candidate profile s_c is saved in S_O if, and only if, the inequality

$$D(s_c, s_j) > t \ \ \forall \ \ s_j \in S_O \tag{2}$$

is true. Number of profiles in S_O at the end of the selection process can be regulated either by tuning the threshold value t or by specifying a maximum number for selected profiles (in which case some profiles in S_I are not necessarily processed at all).

2.1 Implementation

Vertical profile data is retrieved from MACC forecasts run in delayed mode and archived at ECMWF. In January 2012, the archived MACC data cover analysis and forecast data for time period of 1 May 2010 – 31 March 2011. The data is archived in resolution T511, corresponding roughly to 40 km grid spacing. 96-hour forecasts are initialized each day at 00 UTC. However, some of the required parameters are only saved up to 12 hours at forecasts initialized each day at 06 and 18 UTC. Therefore, only four daily forecasts are used in this work. These include 6- and 12-hour forecasts initialized each day at 06 and 18 UTC. By sampling 5th, 15th, and 25th day of each month (to ensure good representation of seasonal cycle, but limiting the data volume to a manageable level) and the four daily forecasts (to ensure representation of diurnal cycle), the total number of forecasts considered in this work is 132. The T511 resolution results in 348,528 individual profiles in each forecast field, making the total number of 46,005,696 individual profiles considered in the sampling.

Sampling of profile data is repeated ten times, each time producing a subset that represents variability of one particular variable only. The ten sampled variables are temperature, specific humidity, mixing ratios of ozone, carbon dioxide, and methane, and aerosol mixing ratios of sulphate, organic matter, black carbon, sea salt, and desert dust. In case of organic matter and black carbon, mixing ratios of hydrophobic and hydrophilic components are sampled together (i.e., K=2 in Eq. (1)), whereas for both sea salt and desert dust, mixing ratios in three size bins are sampled together (i.e., K=3). For all other variables, K=1. Vertical variability is explicitly accounted for in the sampling (i.e., M=60). For the aerosol subsets, however, mixing ratios on model levels higher than 10 hPa are not taken into account, because upper-stratospheric aerosol concentrations are considered unlikely to be realistic (see the disclaimer in Section 3.3). Consequently, M=46 for these subsets.

Each subset of the MACC-60 database consists of 4,000 profiles. The number of profiles in the final database is completely subjective and it is here chosen such that total size of the output database (i.e., including the ten subsets of equal sizes and accounting for the different vertical grids) is on the same order of magnitude as that of the IFS-91 database. For each subset, selection of profiles is carried out in three steps. In the first step, the threshold distance t is specified such that approximately 1000 profiles (out of the 348,528 available ones) from each forecast are selected to the output pool. The applied threshold distances and numbers of profiles selected in the first selection step are given in Table 1. In the second selection step, the output pool of the first step becomes the input pool, and exactly 4,000 profiles are drawn from this pool by random selection. In the third step, 40 extreme profiles (i.e., profiles

Subset	Symbol	Threshold	Profiles out
temperature	T	0.07	158,650
specific humidity	q	0.28	142,208
ozone	O_3	0.17	103,496
carbon dioxide	CO_2	0.24	255,368
methane	CH_4	0.25	153,189
sulphate aerosol	SUA	0.9	123,595
organic matter aerosol	OMA	1.7	148,482
black carbon aerosol	BCA	1.6	137,097
sea salt aerosol	SSA	1.1	158,346
desert dust aerosol	DDA	1.1	149,250

Table 1: Threshold distances used and numbers of profiles selected in the first selection step for different subsets of the MACC-60 database.

deviating as much as possible from the pool of selected profiles) are searched from those profiles that were discarded in the second step. The extreme profiles are chosen one-by-one to account for the changes in the pool of selected profiles during the process. The purpose of the third selection step is to increase variability in the output subsets, as the second selection step is based on random selection rather than a preferential selection algorithm.

3 Validation

3.1 Statistical properties of sampled variables

Figures 1–10 show vertical profiles of mean and standard deviation, as well as curves of minimum and maximum values, of each sampled variable in the corresponding subset of the MACC-60 database. Additionally, profiles of mean and standard deviation are shown for a reference sample that consists of 4,000 randomly selected profiles.

It is generally expected that standard deviation of sampled variables is larger when computed among the corresponding subset of the database than when computed from the random sample. According to Figs. 1–10, this is often the case, but in most cases some exceptions to this behaviour are found in some parts of the vertical profiles. In case of the T-sampled subset, there is a region near tropical tropopause, where the random selection shows more variability than the T-sampled subset of the database. This is interpreted as a reflection from uneven geographical distribution of data, as relatively few tropical profiles are included in the subset. In support to this hypothesis, mean temperature curve of the random selection shows more shape of a tropical profile than does the mean curve of the T-sampled subset. A different picture is found in the q-sampled subset. Also this feature can be attributed to geographical characteristics of the subset, as there are relatively many tropical profiles in the q-sampled subset. In case of the O_3-sampled subset, maximum values in low latitudes are not very well represented in the subset, and this results in standard deviations of the subset and the random selection being similar between 100–10 hPa, while in the upper stratosphere and in troposphere the subsets shows more variability.

In case of aerosol-sampled subsets, as well as for CO_2- and CH_4-sampled subsets, standard deviation in each subset exceeds that in the random sample systematically below 100 hPa. This gives us some confidence that implementation of the sampling is fairly successful in finding a diverse set of profile data for these variables.

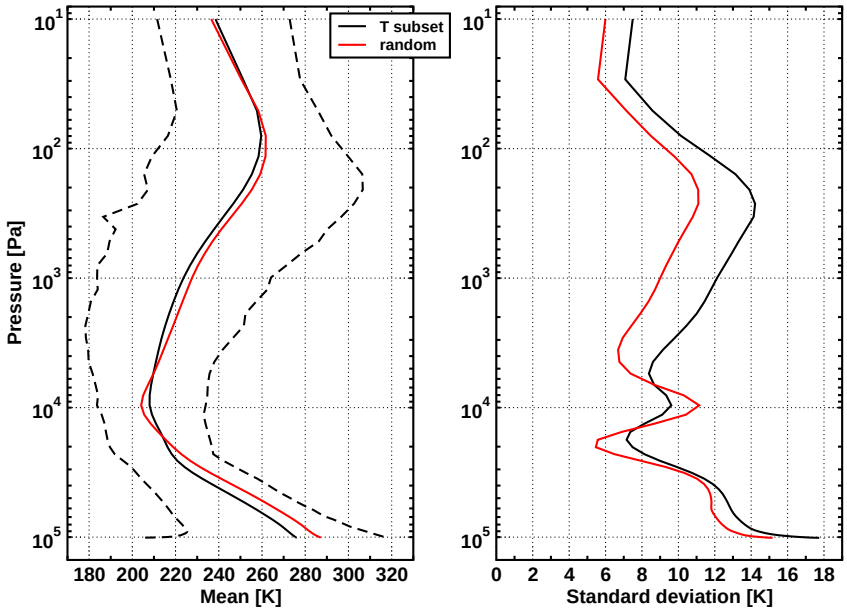

Figure 1: Vertical profiles of mean (left panel) and standard deviation (right panel) of temperature in the T-sampled subset of the database (black lines) and in a random sample of 4,000 profiles (red lines). Dashed lines in the left panel indicate minimum and maximum values in the subset of the database.

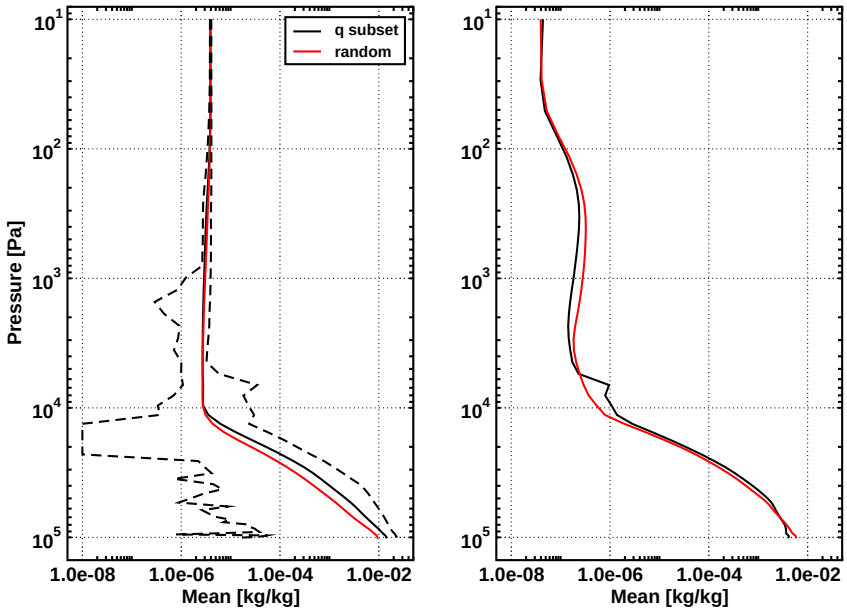

Figure 2: As Fig. 1, but for specific humidity in the q-sampled subset of the database (black lines) and in the random sample (red lines).

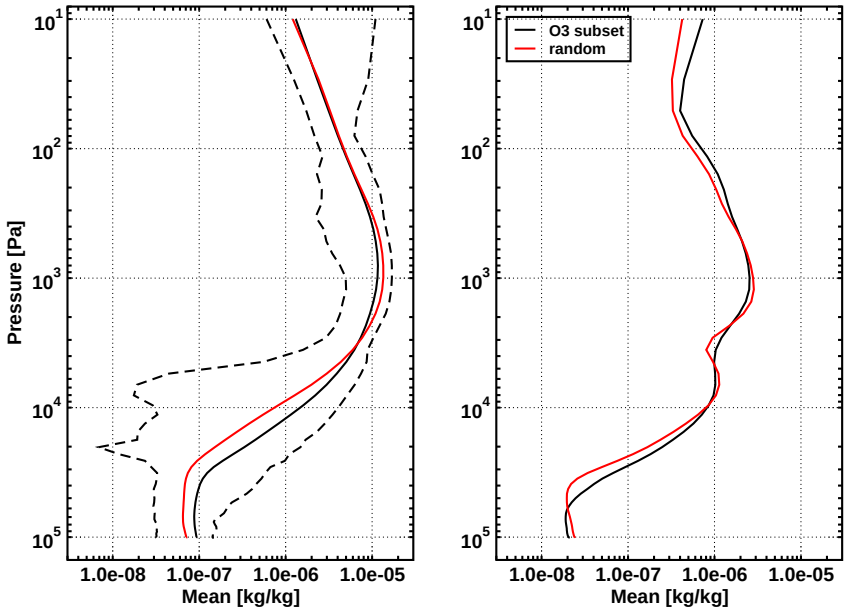

Figure 3: As Fig. 1, but for ozone mixing ratio in the O_3-sampled subset of the database (black lines) and in the random sample (red lines).

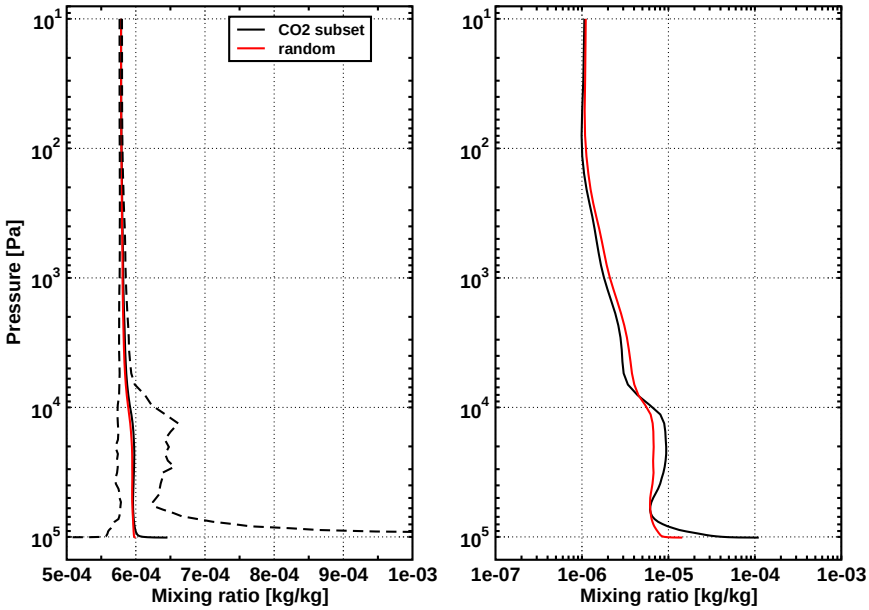

Figure 4: Vertical profiles of mean (left panel) and standard deviation (right panel) of carbon dioxide mixing ratio in the CO_2-sampled subset of the database (black lines) and in the random sample of 4,000 profiles (red lines). Dashed lines in the left panel indicate minimum and maximum values in the subset of the database.

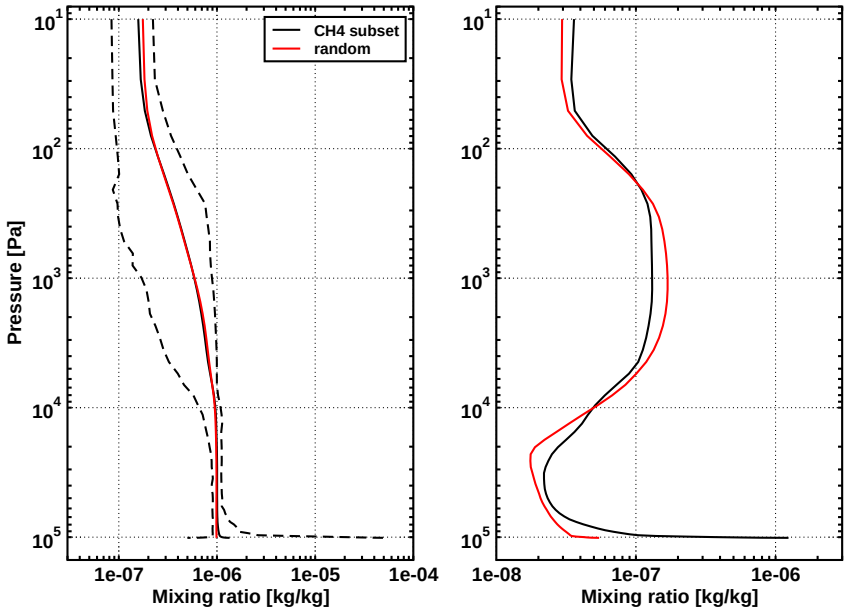

Figure 5: As Fig. 4, but for methane mixing ratio in the CH_4-sampled subset of the database (black lines) and in the random sample (red lines).

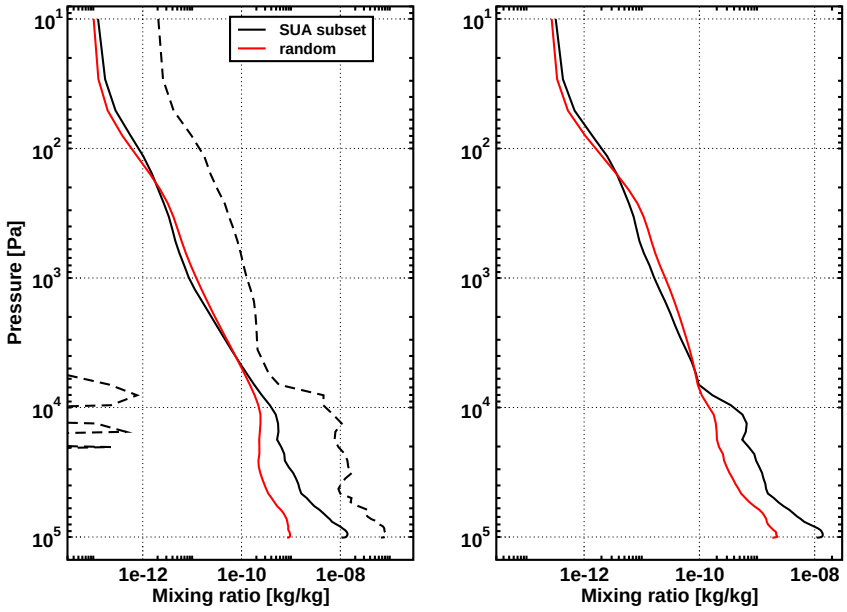

Figure 6: As Fig. 4, but for sulphate aerosol mixing ratio in the SUA-sampled subset of the database (black lines) and in the random sample (red lines).

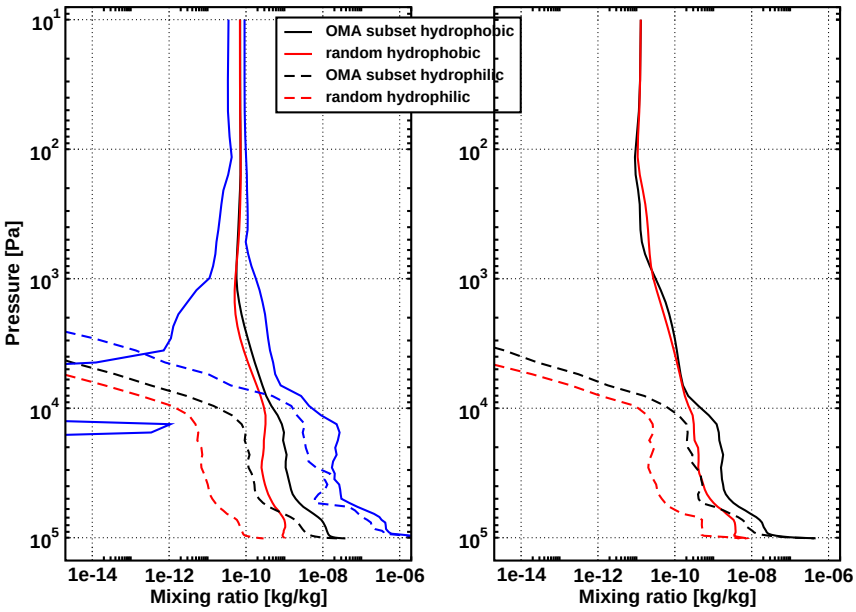

Figure 7: Vertical profiles of mean (left panel) and standard deviation (right panel) of hydrophobic (solid black lines) and hydrophilic (dashed black lines) organic matter aerosol mixing ratios in the OMA-sampled subset of the database. Red lines indicate the corresponding statistics in the random sample of 4,000 profiles. Blue lines indicate minimum and maximum values in the subset of the database.

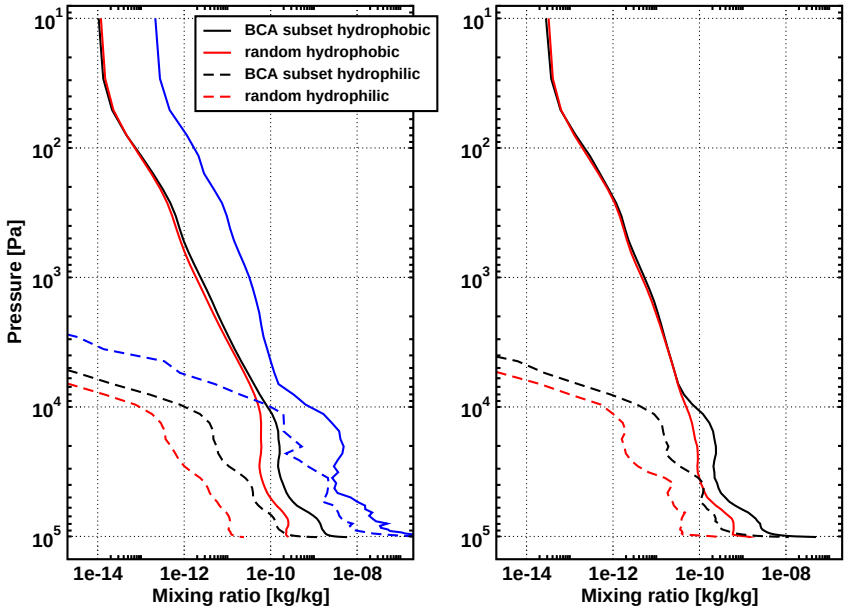

Figure 8: As Fig. 7, but for black carbon aerosol mixing ratios in the BCA-sampled subset of the database (black lines) and in the random sample (red lines).

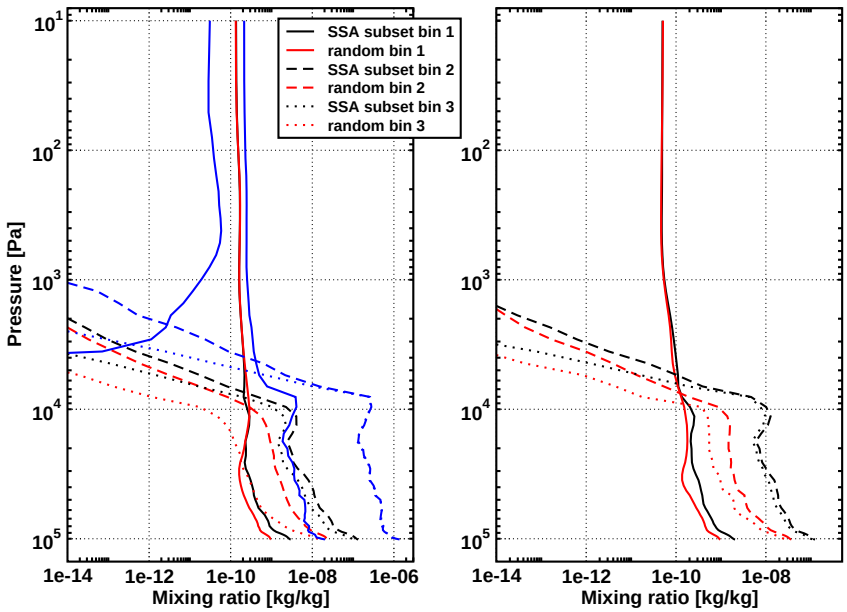

Figure 9: Vertical profiles of mean (left panel) and standard deviation (right panel) of sea salt aerosol mixing ratios in the smallest (solid black lines), mid-size (dashed black lines), and largest (dotted black lines) size bins in the SSA-sampled subset of the database. Red lines indicate the corresponding statistics in the random sample of 4,000 profiles. Blue lines indicate minimum and maximum values in the subset of the database.

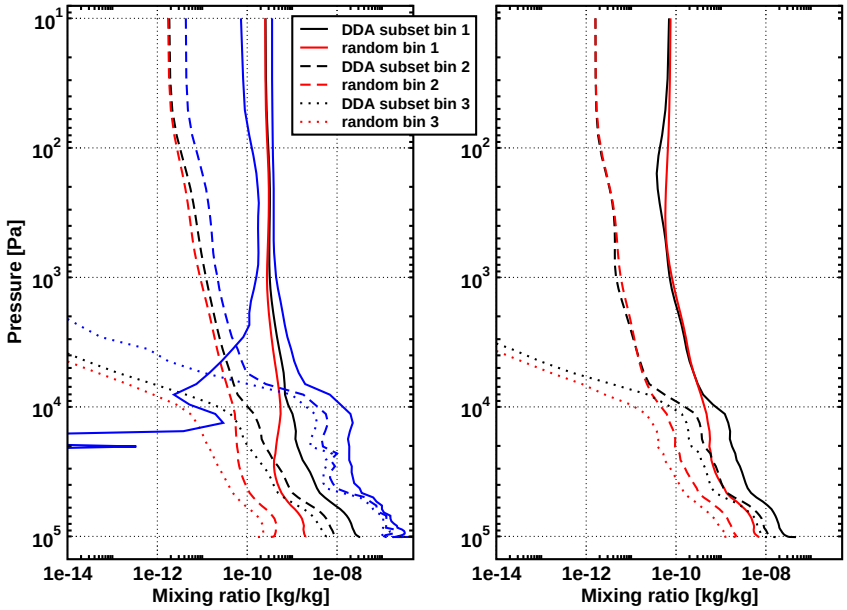

Figure 10: As Fig. 9, but for desert dust aerosol mixing ratios in the DDA-sampled subset of the database (black lines) and in the random sample (red lines).

3.2 Constituents of the database

Atmospheric and surface variables that are included in the MACC-60 database, along with their units, are listed in Table 2. The atmospheric variables are given on 60 model levels. Table 3 gives pressure on each model level for surface pressure 1013.25 hPa. It should be noted that pressure on each model level varies from one profile to another, depending on surface pressure. Additionally to the atmospheric and surface variables, identification of each profile is given in terms of geographical coordinates, forecast base time and forecast lead time. Instructions on how to gain access to the database are given in the Appendix.

3.3 Disclaimer

The current version of the MACC aerosol model does not include stratospheric aerosols. Furthermore, some unrealistic features like persistence of fine mode aerosol particles at high altitudes due to erroneous removal rates and large convective uplifting, have been observed in the aerosol profiles. Hence, use of aerosol profiles above 200 hPa is not recommended, even though profiles of aerosol mixing ratios extend up to 0.1 hPa in the MACC-60 database.

4 Conclusions

A new diverse profile database (the MACC-60 database) is compiled from the 11-month archive of global short-range forecasts originally run for the MACC project. The database consists of 40,000 individual vertical profiles and it is divided in ten subsets of equal sizes to ensure even representation of different sampling variables. In addition to temperature, specific humidity and ozone mixing ratio, that were used as sampling variables in the previous release of the ECMWF diverse profile database (IFS-91 database; Chevallier et al., 2006), samples of carbon dioxide and methane mixing ratios, as well as aerosol mixing ratios of sulphate, organic matter, black carbon, sea salt, and desert dust, are included in the new database. As compared with a random sample of 4,000 profiles, statistical properties of each sampled variable in the corresponding subset suggest that the sampling method has been successfully implemented, and diverse sets of profile data have therefore been found.

Acknowledgements

The author has received funding for this work through the EUMETSAT NWP SAF Programme. Sabatino Di Michele, Rossana Dragani, Richard Engelen, Marco Matricardi, and Jean-Jacques Morcrette are thanked for providing useful comments and general assistance throughout the work.

References

Chevallier, F., S. Di Michele, and A. McNally, 2006: Diverse profile datasets from the ECMWF 91-level short-range forecasts. *NWP SAF Report No. NWPSAF-EC-TR-010, 14 p.*

Atmospheric variables (given on model levels)	
Variable name	Unit
Sea salt aerosol mixing ratio in size bin 0.03–0.55 um	kg/kg
Sea salt aerosol mixing ratio in size bin 0.55–0.9 um	kg/kg
Sea salt aerosol mixing ratio in size bin 0.9–20 um	kg/kg
Desert dust aerosol mixing ratio in size bin 0.03–0.55 um	kg/kg
Desert dust aerosol mixing ratio in size bin 0.55–0.9 um	kg/kg
Desert dust aerosol mixing ratio in size bin 0.9–20 um	kg/kg
Hydrophobic organic matter aerosol mixing ratio	kg/kg
Hydrophilic organic matter aerosol mixing ratio	kg/kg
Hydrophobic black carbon aerosol mixing ratio	kg/kg
Hydrophilic black carbon aerosol mixing ratio	kg/kg
Sulphate aerosol mixing ratio	kg/kg
Carbon dioxide mixing ratio	kg/kg
Methane mixing ratio	kg/kg
Temperature	K
Specific humidity	kg/kg
Ozone mixing ratio	kg/kg
Cloud cover	
Cloud liquid water content	kg/kg
Cloud ice water content	kg/kg
Specific rain water content	kg/kg
Specific snow water content	kg/kg
Vertical velocity	Pa/s

Surface variables	
Variable name	Unit
Logarithm of surface pressure in	Pa
Surface geopotential	m^2/s^2
Surface skin temperature	K
2-meter temperature	K
2-meter dew point temperature	K
10-meter wind speed U component	m/s
10-meter wind speed V component	m/s
Surface albedo	
Roughness length	m
Snow temperature	K
Snow depth	m
Land/sea mask	

Table 2: Atmospheric and surface parameters and their units provided for each vertical profile in the MACC-60 database.

Level	Pressure	Level	Pressure	Level	Pressure
1	0.100	21	44.335	41	577.375
2	0.292	22	54.624	42	616.042
3	0.510	23	66.623	43	654.273
4	0.796	24	80.397	44	691.752
5	1.151	25	95.978	45	728.163
6	1.575	26	113.421	46	763.205
7	2.077	27	132.758	47	796.588
8	2.666	28	153.995	48	828.047
9	3.362	29	177.118	49	857.342
10	4.193	30	202.086	50	884.266
11	5.201	31	228.839	51	908.651
12	6.444	32	257.356	52	930.370
13	7.984	33	287.638	53	949.349
14	9.892	34	319.631	54	965.567
15	12.257	35	353.226	55	979.063
16	15.186	36	388.270	56	989.944
17	18.815	37	424.571	57	998.385
18	23.311	38	461.900	58	1004.644
19	28.882	39	500.000	59	1009.056
20	35.784	40	538.591	60	1012.049

Table 3: Pressure (in hPa) on model levels of the MACC-60 database for surface pressure 1013.25 hPa.

A Instructions for using the MACC-60 database

The MACC-60 profile database consists of ten data files (subsets), that are written in ascii format and provided to users in a single compressed tar file. Assuming that the user has successfully copied the file `profiles_macc_60l.tar.gz` (from the web site of the NWP SAF) in a working directory, the data files can be extracted by entering commands

```
--> gunzip profiles_macc_60l.tar.gz
--> tar -xvf profiles_macc_60l.tar
```

in a linux shell.

In addition to the data files (nwpsaf_macc_sampled.dat), an example fortran program for reading the data files is supplied (readsaf_macc.f90) in the tar file. The example program needs to be compiled using a Fortran 90 compiler (pgf90 in this example). The resulting executable can then be run from the command line

```
--> pgf90 readsaf_macc.f90
--> ./a.out
```

When the program is run, the user is asked to enter identification of the subset he/she wants to read:

```
Enter the identification of the sampled variable:
- t (for temperature)
- q (for humidity)
- oz (for ozone)
- co2 (for carbon dioxide)
- ch4 (for methane)
- sulphate (for sulphate aerosol)
- organic_matter (for organic matter aerosol)
```

```
- black_carbon (for black carbon aerosol)
- seasalt (for sea salt aerosol)
- desertdust (for desert dust aerosol)
```

A confirmation of a successful run will be printed on the screen as the reading of data files is finished:

```
 Number of profiles found in the files:  4000
```

As the example program does not make any output files, users are encouraged to make their own modifications to example program to respond their needs.