The EUMETSAT Network of Satellite Application Facilities

NWP SAF
Numerical Weather Prediction

Document: NWPSAF-EC-TR-044

Version: 1.0.3

Date: 30/05/2025

Diverse profile datasets from the ECMWF CAMS 137-level short range forecasts

Emma Turner

ECMWF

Met Office ECMWF DWD METEO FRANCE

Diverse profile datasets from the ECMWF CAMS 137-level short range forecasts

Emma Turner

ECMWF

This documentation was developed within the context of the EUMETSAT Satellite Application Facility on Numerical Weather Prediction (NWP SAF), under the Cooperation Agreement dated 7 September 2021, 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, DWD and Météo France.

Change record			
Version	Date	Author / changed by	Remarks
1.0.0	07/04/2025	ET	First draft
1.0.1	29/04/2025	ET	Updates after FC comments
1.0.2	30/04/2025	ET	Updates after TB comments
1.0.3	30/05/2025	ET	Updates after reprocessed Cy49r1 e-suite dataset

Diverse profile datasets from the ECMWF CAMS 137-level short range forecasts

Emma Turner

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

Overview

A new dataset of 100,000 diverse atmospheric profiles covering the period March 2024 – February 2025 has been produced from Copernicus Atmospheric Monitoring Service (CAMS) global short-range forecasts at The European Centre for Medium-range Weather Forecast (ECWMF). The data are forecast profiles from the current Cy49r1 operational cycle, which is calculated on 137 vertical levels extending from 1013.25 to 0.01 hPa and an N256 (~40 km) reduced Gaussian horizontal grid. Profiles of meteorological variables and atmospheric composition, the latter comprising greenhouse gases, aerosols and reactive gases have been simultaneously derived for thermodynamic consistency, apart from two greenhouse gases, CO_2 and CH_4, which are from the CAMS greenhouse gas model as this is a more accurate source for these species. A comparison of equivalent results from the operational Integrated Forecasting System (IFS) is presented for meteorological variables. Profiles are sampled in such a way to preserve the statistical properties of the total distribution, following the method refined in the previous datasets (Eresmaa and McNally, 2012, 2014 & 2016). The dataset consolidates and extends the previous three releases, comprising 20 subsets of 5000 profiles, each with dominant variability for one of the five original NWP (meteorological) variables (temperature, water vapour, ozone, cloud condensate and precipitation), seven reactive gases (CO_2, CH_4, SO_2, CO, NO_2, CH_2O with N_2O new), and eight aerosol species (sea salt, desert dust, black carbon, organic matter, sulphate with nitrate, ammonium and secondary organic aerosols new). Each subset also includes the corresponding profiles of all other species at the locations of the dominant species, as well as additional surface based variables. The profile files are available in netcdf4 format.

Table of Contents

Introduction and scope

Datasets comprising thousands of diverse atmospheric profiles of meteorological variables and gaseous species are required by many satellite related science applications and other disciplines. Designed to represent the full statistical variability of the underlying Numerical Weather Prediction (NWP) models in typically 5000 profiles per subset, these datasets can be used to train statistical modules, such as regression or machine learning, or can be input to radiative transfer models, such as the NWP SAF Radiance Simulator wrapper to RTTOV, to test its sensitivity to multiple variables (Hocking, 2023). NWP models have developed rapidly in recent decades and are regularly updated to new 'cycles', resulting in the previous profile datasets becoming out-of-date and necessitating a new product be created, which is the subject of this document.

This is the fifth diverse profile dataset produced by ECMWF for NWP and in this instance it also amalgamates Atmospheric Composition (AC) profiles into a single product consisting of 100,000 profiles (20 variables). The two previous NWP sets have focused on sampling a handful of variables that exert the greatest effect on NWP radiances: temperature, humidity, ozone, cloud and precipitation. The last release of NWP profiles was over 10 years ago (Eresmaa and McNally, 2014) and comprised forecast profiles produced by model version Cy40r1 of the high-resolution Integrated forecasting System (IFS) at T1279 horizontal resolution (~16 km grid spacing) on 137 vertical levels. This set is referred to as NWP v4 in the text. The last release of AC profiles (AC v3) was slightly more recent (Eresmaa and McNally 2016) and provided datasets of so-called 'reactive gases': carbon monoxide (CO), nitrogen dioxide (NO_2), sulphur dioxide (SO_2), and formaldehyde (CH_2O), using forecasts produced by version Cy41r1 of the Copernicus Atmospheric Monitoring Service (CAMS) at T511 horizontal resolution (~40 km grid spacing) on 60 vertical levels. The AC release before this (Eresmaa and McNally 2012) focused on greenhouse gases and aerosols: carbon dioxide (CO_2), methane (CH_4), sulphate, organic matter, black carbon, sea salt and desert dust, using forecasts produced by the EU FP7 project "Monitoring Atmospheric Composition and Climate (MACC). MACC was a predecessor to CAMS that was not produced in real time and had the same spatial resolution as the CAMS model described above.

The current dataset is produced by CAMS global atmospheric composition (CAMS AC) forecasts and incorporates all previous variables and some new ones, specifically nitrous oxide (N_2O) and three aerosols: ammonium, nitrate and secondary organic matter. The decision to use all CAMS data is based on a desire for thermodynamic consistency between variables. In other words all profiles provided are produced by the same model and are therefore physically consistent with each other, which would not be the case if we were to mix IFS NWP variables with CAMS AC variables. The only two species for which this condition is relaxed are CO_2 and CH_4, as these are from CAMS global greenhouse gas (GG) forecasts[1] because they are modelled more accurately. The horizontal resolution for CAMS GG forecasts is higher (O1280) so these fields must be interpolated to the same resolution as the CAMS AC forecasts. A separate comparison of CAMS NWP profiles with the equivalent diverse product derived from the operational IFS is presented for validation.

[1] DOI: 10.24381/93910310

The selection process for obtaining the optimum 5000 profiles for each variable is largely unchanged since the previous release of NWP v4, which places emphasis on creating a representative sample of the 'parent' model through prioritising random selection. Earlier releases, such as Eresmaa and McNally (2012), used a different process where 100% of the first step was derived from a selection algorithm that sought to identify dissimilar profiles, which placed a higher weight on the extremes of the parent model and hence the statistics diverged. Eresmaa and McNally (2014) show how even a relatively small reliance on a diverse selection algorithm, such as just 10% of the number of profiles, creates a relatively unrepresentative sample for some variables. But applying the random selection step to 90% of the data *twice* reduces the selected number sufficiently to retain the statistics of the wider envelope of all profiles, and hence this is how the current dataset derived.

A table summarising the key features of the current and previous datasets including the selection processes is provided in Appendix A.

This document describes the selection methodology and presents the CAMS-L137 v1 dataset for each variable with comparisons to the last previous dataset where available. There is brief discussion of the profiles differences, but explaining the mechanisms behind these changes, apart from the effects of the selection algorithm, is beyond the scope of this work and the interested reader is referred to the technical documents in the footnotes and references for more details of the various model upgrades.

1 Profile selection

1.1 Methodology

1.1.1 Selection algorithm

The selection algorithm is, in essence, nearly identical to the one described in the previous NWP v4 documentation, and is reproduced below. The sampling method is based on Chevallier et al. (2006). Pools of input and output profiles, S_I and S_O, are defined respectively. S_I is arranged in random order and the first profile is saved in S_O. The next candidate profile, s_i is compared with the profile in S_O (s_j), and a measure of inter-profile difference, D, is calculated as below.

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

Formally, the comparison is based on the sum of squared normalised departures of profile quantities over number of levels and variable bins, weighting each one equally. k and m are the indices of variable and level, respectively, and K and M are the corresponding total number of variables and levels to be considered. $\theta_{ik}(m)$ and $\theta_{jk}(m)$ are the values of

variable k on level m for the two profiles and $\sigma_k(m)$ is the associated standard deviation. A candidate profile is saved to S_O if the following inequality is true for all of the profiles s_j

$$D(s_i, s_j) > t \; \forall \; s_j \in S_O \qquad 2$$

The process repeats until there are no profiles left in S_I. t is a threshold that is empirically tuned such that the total number of profiles saved to S_O is as desired after *all* input profiles are considered. Note that it was necessary to apply a scaling factor of 0.01 to D for all variables to achieve a suitable number of output profiles using thresholds with magnitudes similar to those given in previous documents.

Before the selection algorithm is applied to each step, 90% of the profiles required are selected by random. To make the random selection non-reproducible a random seed generator was used ahead of a random number selector, which seeds the calculation with random data retrieved from the operating system and means a different set of numbers are obtained each time the code is run.

1.1.2 Implementation

The Copernicus Atmosphere Monitoring Service[2] is a component of the European Earth Observation programme, Copernicus (Peuch et al., 2022), which has been fully operational since 2015 and is based on the operational IFS. The initial conditions of the forecast are obtained by combining a previous forecast with satellite observations of aerosol optical depth, O_3, CO, NO_2, and SO_2 through the 4D-VAR data assimilation process. CAMS provides daily analyses and forecasts of reactive trace gases, greenhouse gases and aerosol concentrations. Forecasts up to five days are initialised twice per day at 0000 and 1200 UTC and variables are available every hour for surface fields and every three hours for model level fields.

The spatial resolution remains at T511 in the horizontal dimension but the number of levels were increased from 60 to 137 in 2016 (Cy46r1). There have been 13 cycle upgrades to the CAMS model since 2015 (starting at Cy40r1). The operational version is currently at Cy49r1 as of 12th November 2024 and prior to this was Cy48r1[3], which was implemented on 27th June 2023 (Eskes et al., 2024). The latest upgrade brought significant changes, which are mostly limited to aerosols and SO_2[4]. The sample year that the new profile dataset is drawn from traverses the recent change of cycle, so in order to retain the more recent developments profiles before 12th November 2024 are obtained from the early delivery suite (e-suite), which is identical to Cy49r1 but runs for several months prior to the official release. The experiment version id is 0079 for the e-suite and the id for the operational suite (o-suite) is 0001.

[2] http://atmosphere.copernicus.eu
[3] https://atmosphere.copernicus.eu/sites/default/files/publications/CAMS2_82_2022SC1_D82.3.2.1-2023Q2_Cy48R1_upgrade_evaluation.pdf
[4] https://atmosphere.copernicus.eu/sites/default/files/publications/CAMS2_82_2023SC2_D82.3.2.1-2024Q3_Cy49R1_upgrade_evaluation.pdf

The profile database is compiled from the short-range forecasts spanning the time period 1st March 2024 to 20th February 2025. The selection of dates and times from the archive follows the pattern set by the last two NWP datasets, and the subsequent thinning seeks to drastically reduce data volume, while preserving the annual and diurnal variability of the model fields. Only forecasts initialised at 0000 UTC on the 1st, 10th and 20th day of the month with forecast lead times of 36, 42, 48 and 54 hours are retained, resulting in 144 snapshots of the model. The horizontal spherical harmonic model field representation with a spectral truncation of T511 corresponds to a reduced Gaussian grid of N256, which means there are 256 equally spaced latitude lines between the poles and equator, and a decreasing number of equally spaced longitude points from 1024 at the equator to 18 at the poles, resulting in 348,528 grid points per level. The reduced Gaussian grid is retained for the selection algorithm rather than re-gridding to a regular latitude, longitude arrangement to keep the density of model grid points roughly equally in all areas of the globe. Results from this model are labelled as CAMS-N256 in the text. CO_2 and CH_4 fields from the CAMS GG model are retrieved at the same dates and times and interpolated to a N256 grid from their octahedral reduced Gaussian grid at O1280, which corresponds to 6,599,680 points.

The first step of the selection process seeks to retain approximately 1500 profiles from each forecast time, totalling around 216,000 per variable. Following the established methodology 1350 are selected at random and then approximately 150 are selected via the process described in section 1.1.1. The second step collates all of the profiles from the first step as the new input pool and retains exactly 5000 for each profile where 4500 are randomly selected and 500 are determined via the selection algorithm.

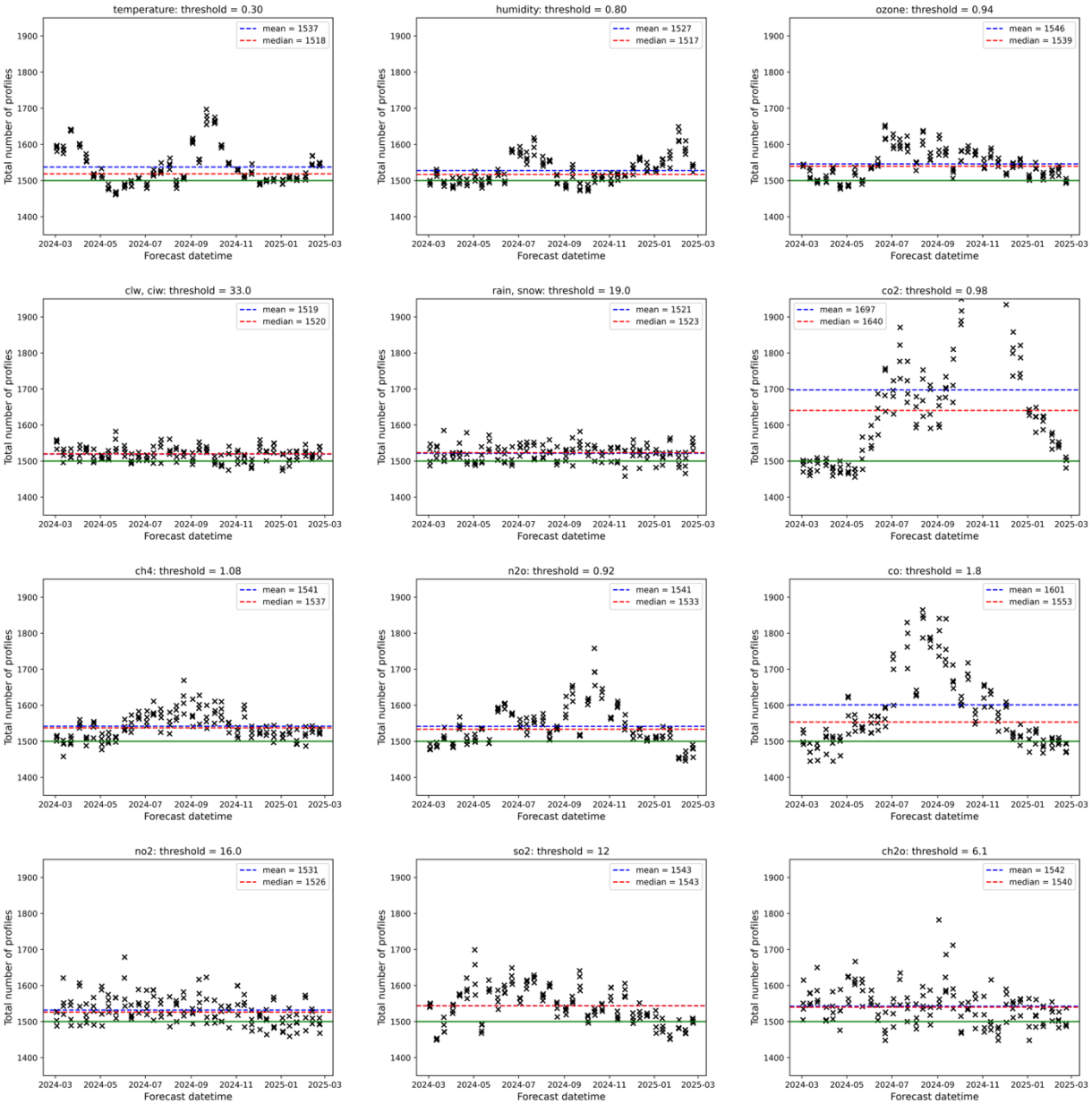
The final set of CAMS profiles are compared with the equivalent product from the operational IFS model for the NWP fields: temperature, humidity, ozone, cloud condensate and precipitation. The IFS currently has a vertical resolution of 137 levels and a T1279 spherical harmonics horizontal resolution (~9 km) with an octahedral reduced Gaussian grid of O1280 corresponding to 6,599,680 points. The IFS model was also upgraded to Cy49r1 on 12th November 2024 with a suite of changes[5] so data from the corresponding e-suite was also used prior to this. The IFS e-suite began on 13th March 2024 so the first eight forecasts are missing. IFS forecasts are initialised four times daily at 0000, 0600, 1200 and 1800 UTC, where the 0000 and 1200 UTC baseline forecasts are available up to day 15 (day 10 in Cy48r1) and the 0600 and 1800 UTC initialised forecasts are available up to day 6 (~day 4 in Cy48r1). The selection process for the IFS was reproduced in exactly the same way as CAMS to give a completely independent test of the resulting profiles, which is shown in section 4. Results from this model are labelled IFS-O1280 in this report.

1.1.3 First step

The selection algorithm presented in section 1.1.1 cannot be solved analytically, so the process was repeated multiple times with different thresholds to achieve the desired number of dissimilar profiles, which for the first step is about 150, after 1350 have been extracted randomly. The first round was labour intensive as each of the 144 files produces a different

[5] https://www.ecmwf.int/en/about/media-centre/news/2024/forecast-upgrade-improves-wind-and-temperature-predictions https://www.ecmwf.int/en/publications/ifs-documentation

number of selected files for the same threshold, due to the annual variability of the atmospheric profiles. This required all time steps to be processed for every candidate threshold to assess the full distribution as a whole, as no one snapshot would necessarily represent all of them. See Figure 1 for the distribution of the number of selected profiles produced for the final threshold for each variable. Note the y-axis shows the total number of profiles including the first 1350 which are randomly selected. Some variables like precipitation (rain and snow) are relativity stable in time but these are few and the annual variability is particularly high for temperature, carbon dioxide and carbon monoxide. Other species show some more abrupt changes that is more likely related to isolated events or possibly model changes, such as ammonium or nitrate aerosol.



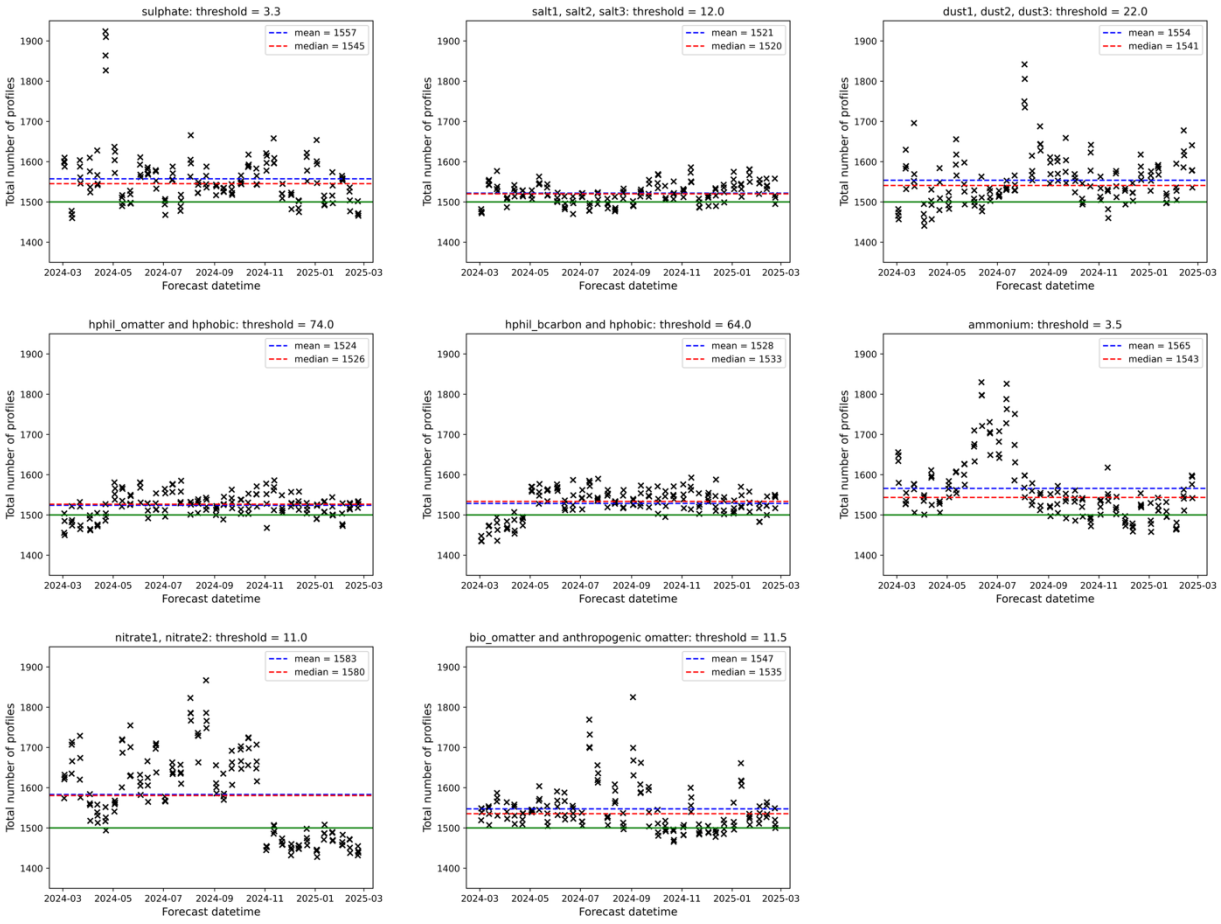


Figure 1. The total (random + selected) number of profiles retained from step one of the selection process, for each of the 144 samples (shown at their forecast date and time). Each panel displays a different variable with the name given in the title alongside the chosen threshold. The green line indicates 1500 profiles (the theoretical aim), and the red and blue dashed lines are the median and mean of the data, respectively.

This variability means the threshold is therefore a somewhat arbitrary choice. It was decided to favour values producing overall higher numbers of profiles, with means and median above 1500, in order to have at least a reasonable number of disparate profiles at most timesteps.

1.1.4 Second step

It was a simpler task to determine the threshold for the second step as only one set of 5000 profiles is produced, of which 500 are selected, allowing the tuning to be more accurate. The random seed operator meant that the final number of profiles would vary slightly (within around 20 profiles) every time the process was initiated, even for the same threshold, because a different set had already been selected at random. Once a threshold was found to give a number of selected profiles between 500 and 520 the process was terminated, and all selected profiles were retained at the expense of a few random ones, to give 5000 profiles exactly.

Applied inter-profile departure threshold parameter values for both steps and counts of profiles remaining after the first step, are listed in Table 1.

Table 1. Applied threshold parameters for the first and second rounds of running the selection algorithm and total counts of selected profiles after step one, for each subset of the CAMS-L137 v1 database.

Long name	**Short name**	**Bins**	t **step 1**	$\sum S_O$ **step 1**	t **step 2**
Temperature	temp	1	0.30	221,395	0.57
Specific humidity	hum	1	0.80	219,975	1.24
Ozone	oz	1	0.94	222,644	1.50
Cloud condensate	clw, ciw	2	33.0	218,814	9.7
Precipitation	rain, snow	2	19.0	219,148	2.82
Carbon dioxide	co2	1	0.98	244,445	1.30
Methane	ch4	1	1.08	222,021	1.315
Nitrous oxide	n2o	1	0.92	221,987	1.10
Carbon monoxide	co	1	1.80	230,565	3.05
Nitrogen dioxide	no2	1	16.0	220,562	5.35
Sulphur dioxide	so2	1	12.0	222,289	9.0
Formaldehyde	ch2o	1	6.1	222,134	4.33
Sea salt aerosol	salt1, salt2, salt3	3	27.0	219,111	11.1
Desert dust aerosol	dust1, dust2, dust3	3	22.0	223,776	16.45
Sulphate aerosol	sulphate	1	3.3	224,293	3.86
Organic matter	hphil_om, hphob_om	2	74.0	219,464	30.3
Black carbon	hphil_bc, hphob_bc	2	64.0	220,106	24.7
Ammonium	ammonium	1	3.5	225,479	3.32
Nitrate aerosol	nitrate1, nitrate2	2	11.0	227,963	9.8
Secondary organic	bio_om, anthr_om	2	9.5	222,820	10.52

2 NWP profiles

In the following sections spatial distributions and statistical quantiles plus the minimum, maximum, median (and mean where appropriate) of the selected profiles for each variable, are displayed against the equivalent quantities from the last available dataset. Differences between them are highlighted in the text.

In order to compare statistical quantiles between datasets with a different number of vertical levels, model levels are converted to pressure levels using the appropriate a and b coefficients[6] and the surface pressure. A surface pressure of 1013.25 hPa is applied to all profiles to produce values on every level, which is required to calculate the quantiles, minimum, maximum and median, which would not be possible if the actual surface pressure of each profile were used. Therefore the statistical quantile plots should be used as a tool for comparison, rather than as an accurate depiction of individual profiles. See Appendix B for details on the model level to pressure conversion, using CO_2 as an example.

2.1 Meteorological variables

2.1.1 Temperature

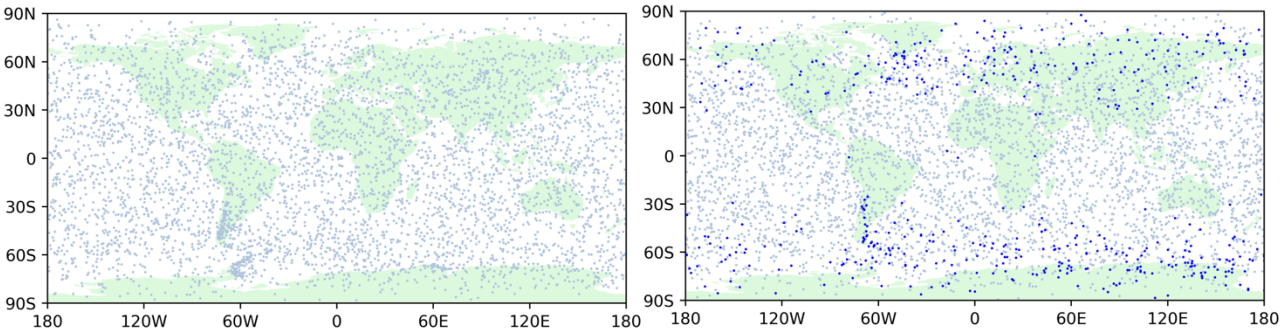


Figure 2. Locations of temperature profiles in the sampled subsets of the previous NWP v4 (left) database from the IFS in 2014 compared with the equivalent set produced from CAMS-L137 v1 (right). For CAMS the final 500 profiles derived from the selection algorithm are shown in blue pixels and the 4500 randomly selected profiles are shown in blue grey pixels.

On the whole, the spatial distribution of temperature profiles show similar patterns between NWP v4 and CAMS-L137 v1 (Figure 2), apart from a tighter cluster of pixels visible across the southern tip of America and the Antarctic Peninsula in the NWP v4 dataset. This feature is also in the latest operational version of the IFS (see Figure 44) suggesting the selection algorithm picks up more disparate profiles in the IFS forecasts due to the higher horizontal resolution. The selected profiles are not distinguished from the random profiles in the previous datasets so this cannot be proved definitively. There is evidence of colder profiles in the previous IFS at levels above the tropopause around 10 hPa in the log scale pressure panel of Figure 3, but the similar quantiles indicate this is not a systematic issue and therefore could just be a few isolated profiles. Overall the median profile from CAMS is up to 1 K warmer in the troposphere than the previous dataset, and one reason for this could the global temperature increase in the past 10 years. Conversely, the mesosphere has

[6] https://confluence.ecmwf.int/display/UDOC/L137+model+level+definitions

reduced in temperature up to 5 K in the interim time, which is more likely to do with model development.

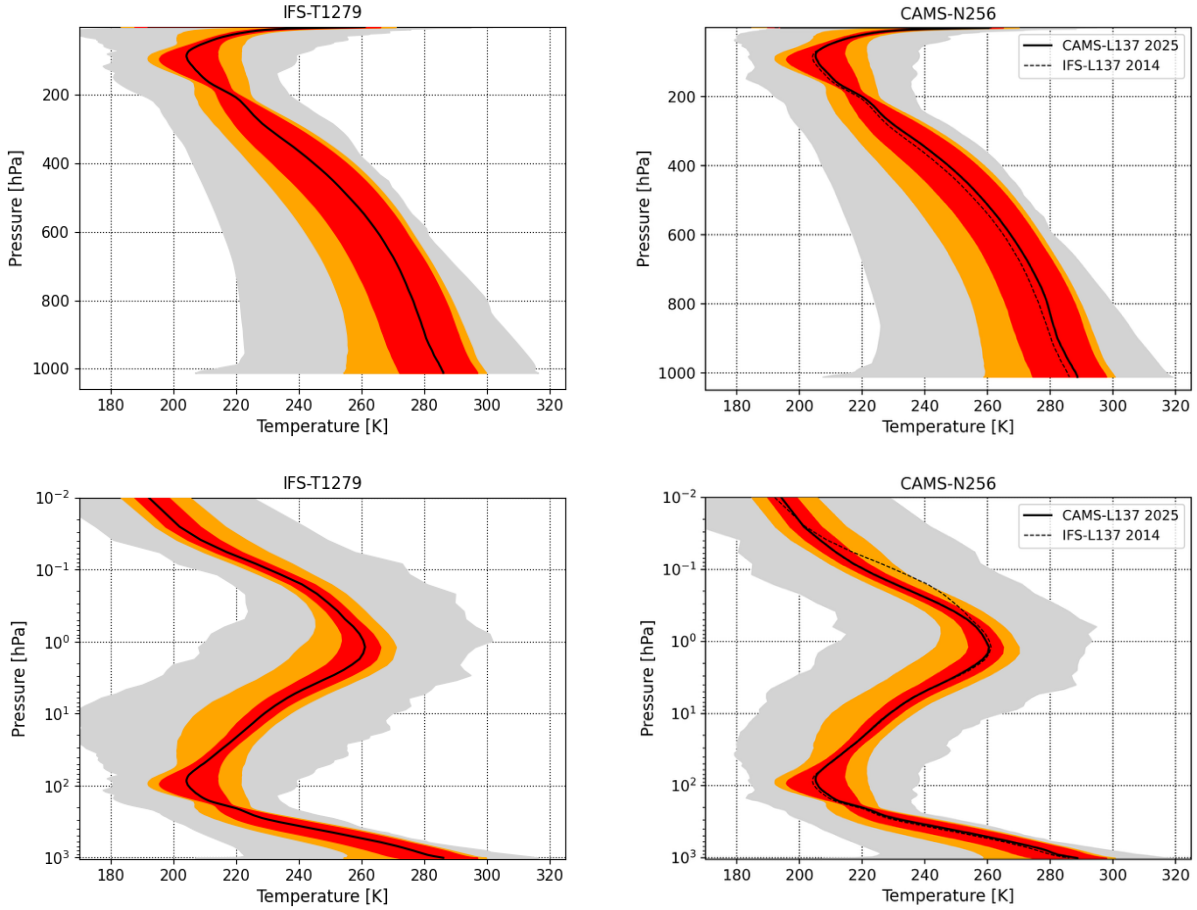


Figure 3. Distribution of temperature profiles in the respective subsets of the previous NWP v4 IFS (left) and the current CAMS (right) databases. Top panels have a linear pressure scale and bottom have a log pressure scale. Grey shading indicates the range constrained by the minimum and maximum values, orange shading is the range constrained by the 10th and 90th percentiles, and red is the lower and upper quartiles (25th and 75th). The black line is the median profile and the right plot displays medians from both datasets for comparison.

2.1.2 Humidity

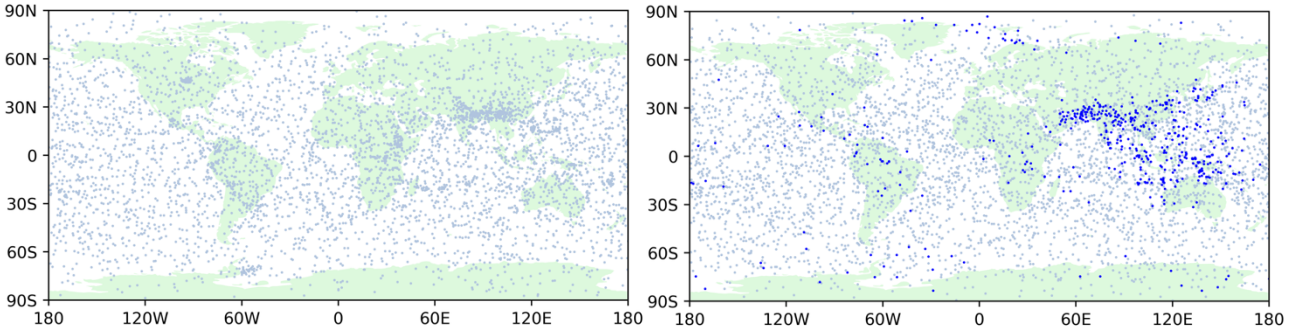


Figure 4. Same as Figure 2 but for specific humidity profiles. NWP v4 (left) and CAMS-L137 v1 (right).

There are clusters of selected profiles over Southern Asia and the Indo-Pacific ocean in the current CAMS dataset, which is likely related to dissimilar profiles produced by convective activity and monsoon events (Figure 4). Overall the profile statistics seems relatively unchanged since the NWP v4 dataset 10 years previous, apart from perhaps slightly lower humidity around 500 hPa in the CAMS-L137 v1 dataset (Figure 5).

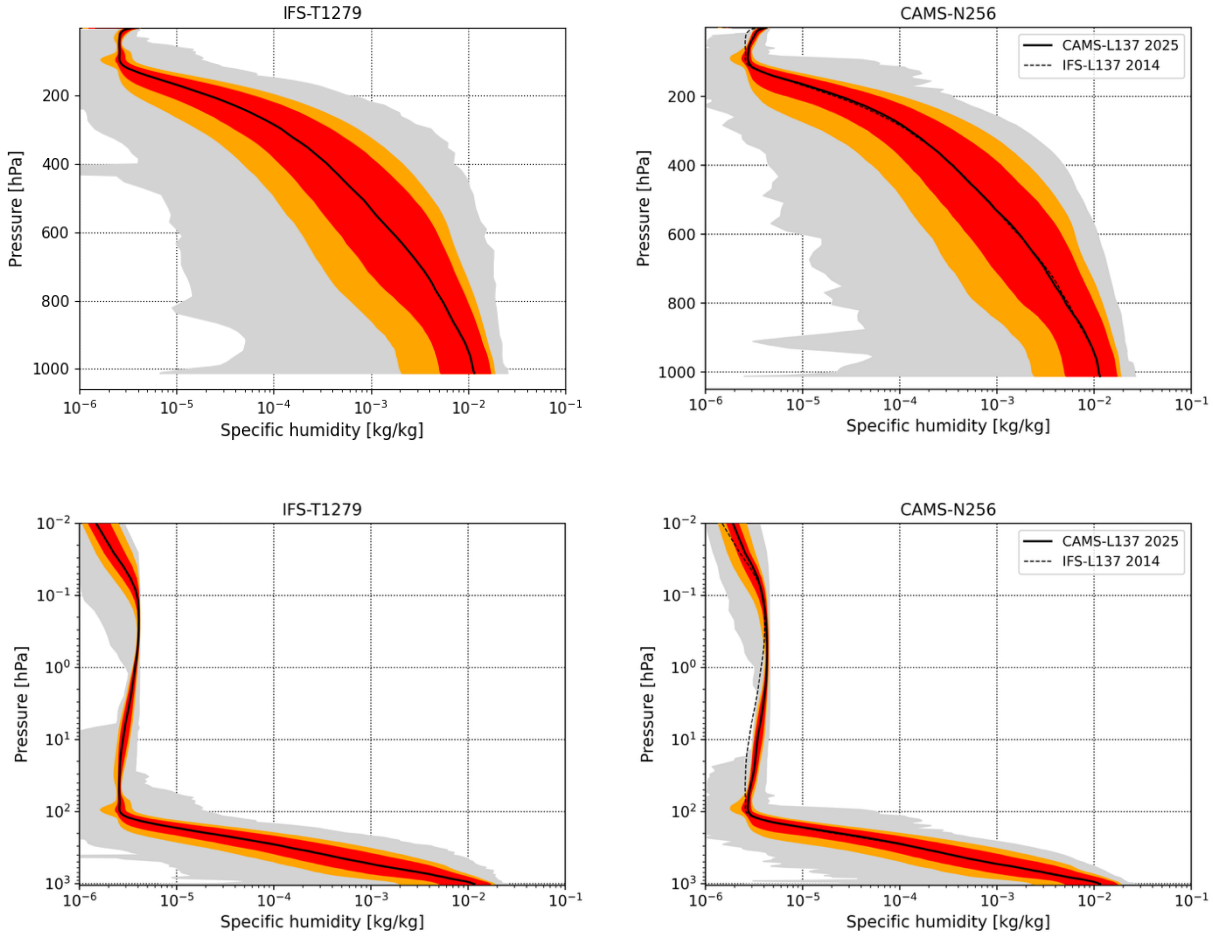


Figure 5. Same as Figure 3 but for specific humidity.

2.1.3 Ozone

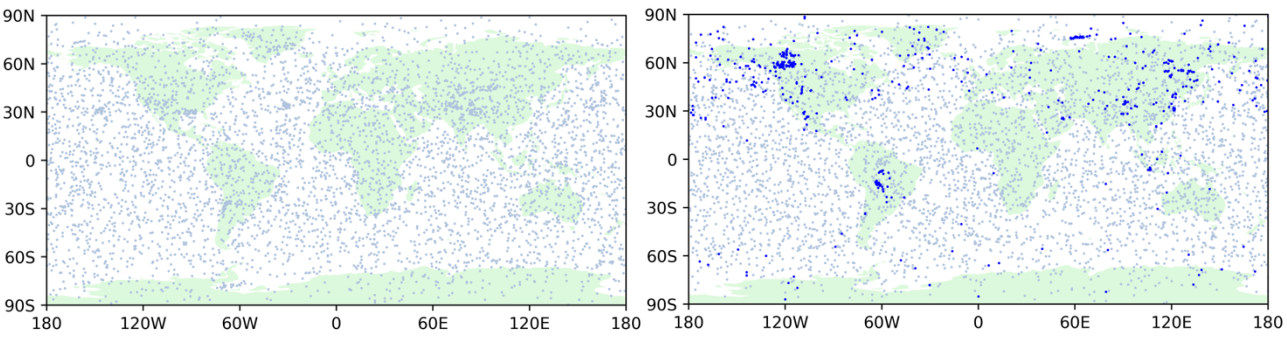


Figure 6. Same as Figure 2 but for ozone profiles. NWP v4 IFS (left) and CAMS (right).

Disparate ozone profiles seem to be preferentially selected over the northern hemisphere in the current CAMS subset, with clusters visible over Northern Canada, Russia and Asia (Figure 6). There are some noticeable differences in ozone profiles at levels above 10 hPa shown in Figure 7. There appears to be a wider range of concentrations in the CAMS-L137 v1 dataset and the second peak, which is visible in the profile envelope is higher - around 0.5 hPa relative to around 1 hPa in the 2014 equivalent.

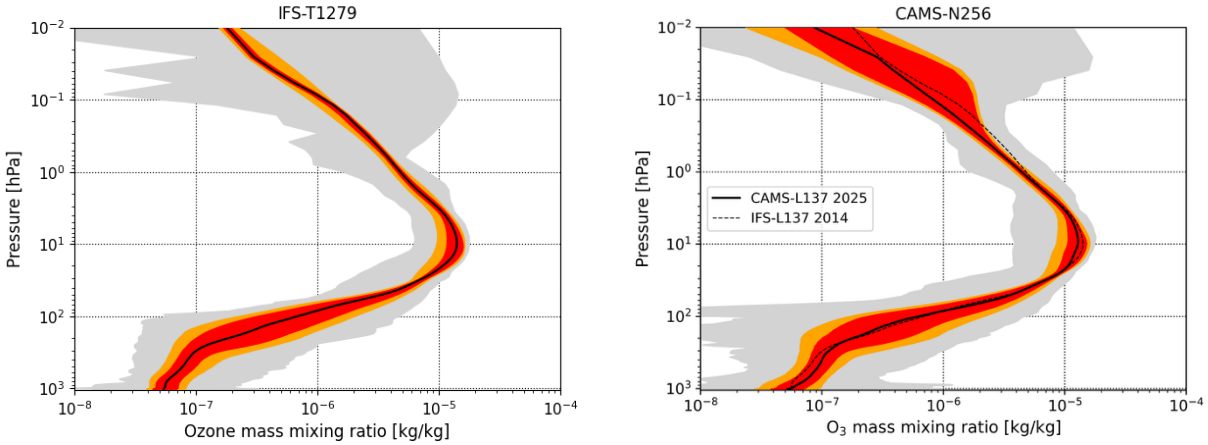


Figure 7. Same as Figure 3 (bottom row) but for ozone.

This second peak is known as the Tertiary Ozone Maximum (TOM), which is observed to occur sporadically in the polar winter and spring at an altitude of around 72 km due to a build-up of ozone in the absence of ultraviolet radiation at night-time. See Hartogh et al. (2004) for more detail of the generative mechanism. It is known as the third peak as there is a second stronger peak that occurs much higher at around 90 km (see Schranz et al. 2018). The locations of all profiles in the current CAMS ozone dataset with a TOM are shown in Figure 8. This confirms that all TOM profiles occur in the polar regions in the boreal and austral autumn and winter (but none in spring), and the majority are chosen by the selection algorithm. TOM profiles comprise just under 5% of the total number of profiles in the dataset.

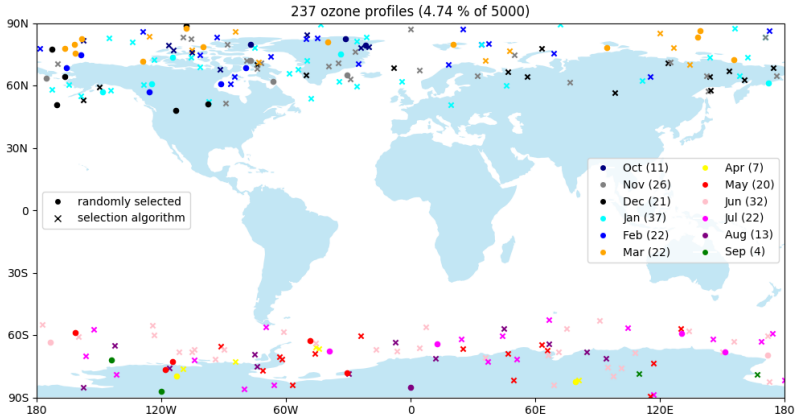


Figure 8. Locations of all ozone profiles in the current CAMS ozone dataset with a Tertiary Ozone Maximum (identified as having ozone concentrations over 2 ppmv in any of the first five levels). Month of occurrence is colour coded, and profiles that are randomly selected are show in circles, whereas those picked via the selection algorithm are shown in crosses.

2.1.4 Cloud condensate

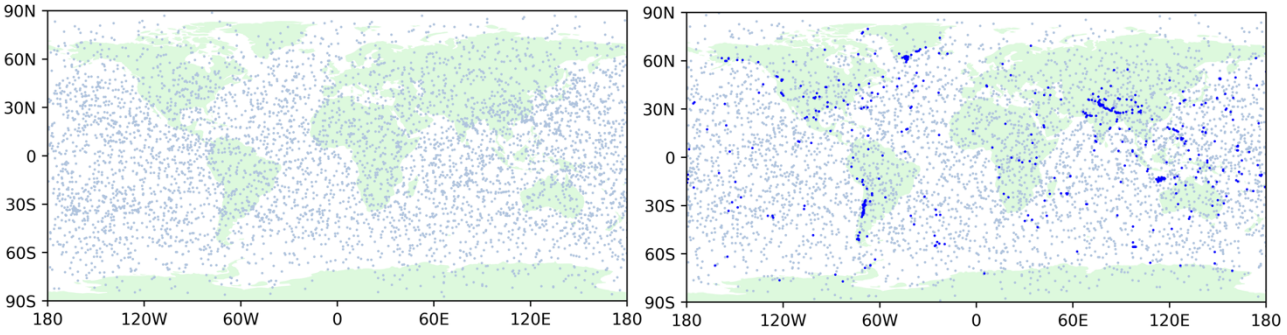


Figure 9. Same as Figure 2 but for cloud liquid water and cloud ice water profiles (both have the same locations as they are binned together). NWP v4 IFS (left) and CAMS-L137 v1 (right).

The selection algorithm picks up a band of profiles above the Himalayan mountains around the Tibetan Plateau in CAMS indicating large variability in the cloud profiles above this area that is not evident in the NWP v4 dataset (see Figure 9). There are also several clusters selected around the Indo-Pacific ocean, which is unsurprising given the prevalence of deep convective clouds in this region. There are also more cloud profiles being selected along the coast of America and the eastern coast of Greenland compared to the previous dataset.

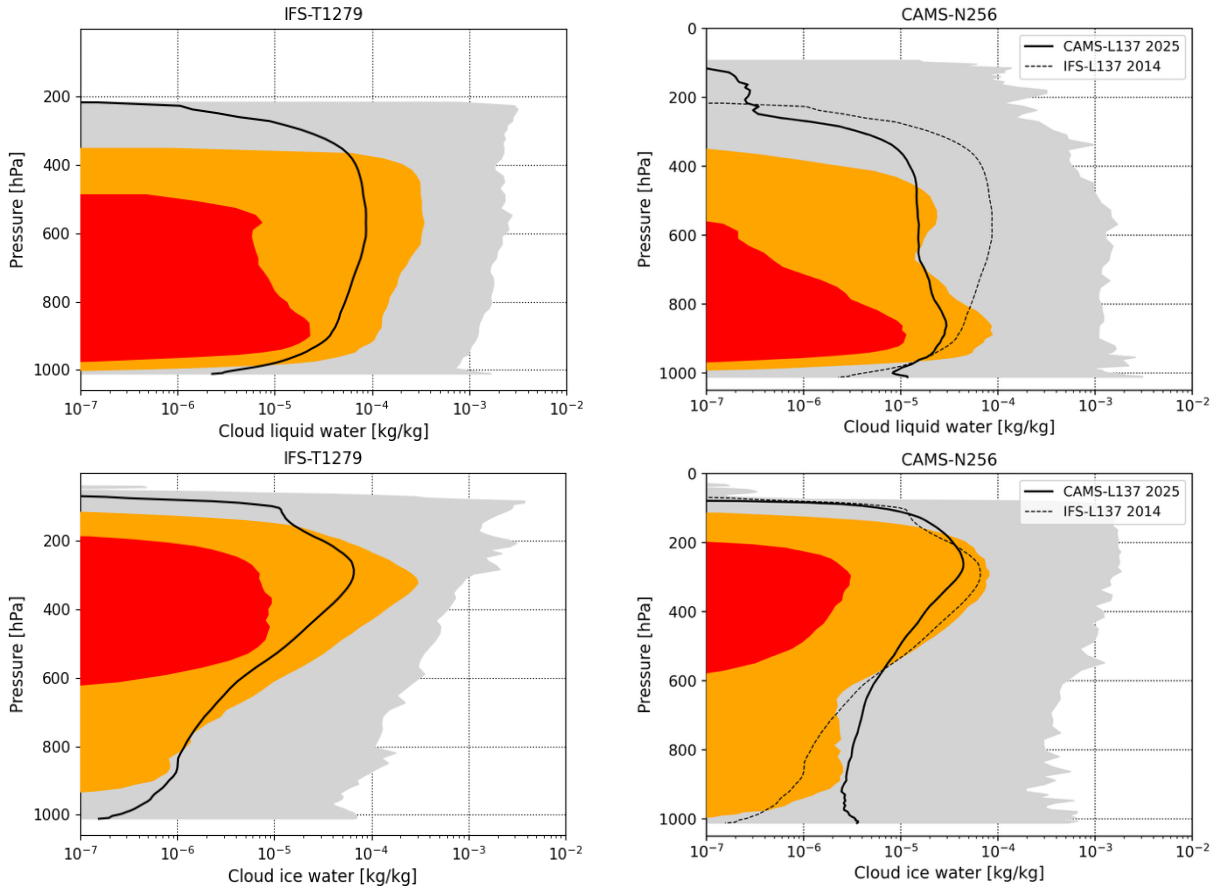


Figure 10. Same as Figure 3 but for cloud liquid water (top) and cloud ice water (bottom). Note that black lines represent the mean rather than the median in these panels.

Note that for clouds and precipitation the black line in the quantile figures represents the mean profile rather than the median, as the high prevalence of clear sky profiles with no hydrometeors renders the median near zero and is therefore not a useful tool for comparison. There is a layer of liquid clouds in CAMS at levels above 200 hPa, which is absent in the previous IFS (Figure 10) but visible in the latest operational IFS (see Figure 48) indicating a model development. In general there are lower concentrations of liquid clouds in CAMS in the mid to upper troposphere relative to the IFS suggesting this is a feature of horizontal resolution. There are also less ice clouds in CAMS at upper levels but seemingly more below around 600 hPa.

2.1.5 Precipitation

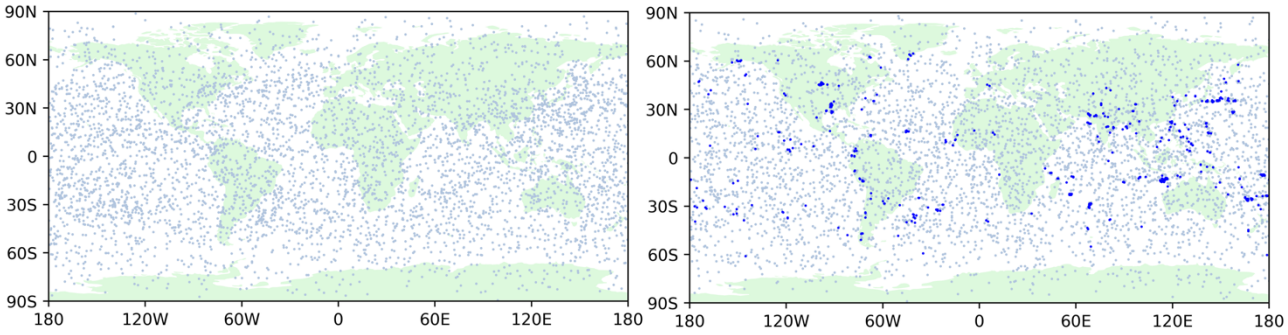
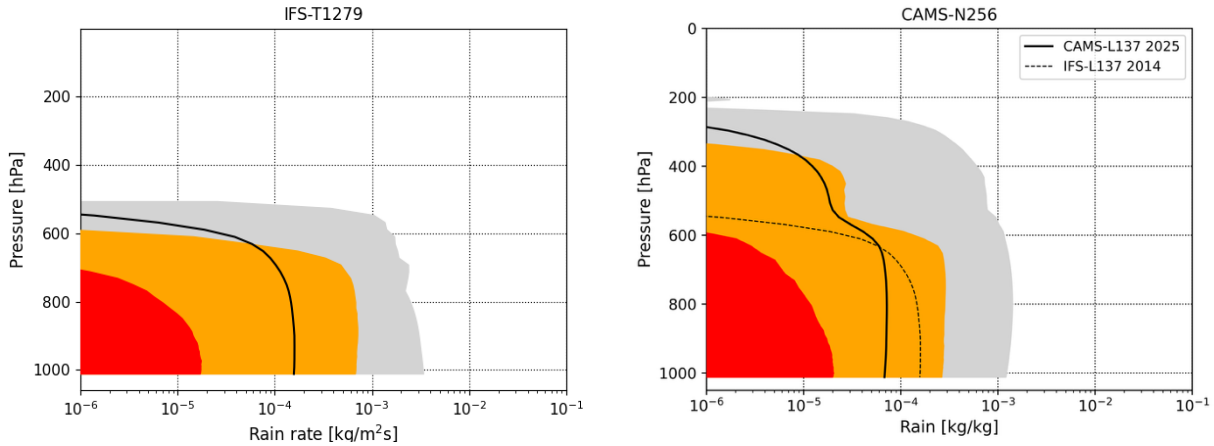


Figure 11. Same as Figure 2 but for rain and snow (both have the same locations as they are binned together). NWP v4 IFS (left) and CAMS-L137 v1 (right).

It is not obvious that there is any noticeable change in spatial precipitation patterns between precipitation datasets from the global distribution apart from perhaps a horizontal band of disparate precipitation profiles east of Japan in the CAMS dataset, which is not evident in the IFS dataset (Figure 11), so this could be from an isolated event in the latter time period. Figure 12 shows the presence of rain in CAMS between 500 and nearly 200 hPa, which is completely absent in the NWP v4 dataset. This is compensated somewhat by a lower mass mixing ratio at lower levels. There is more snow overall in CAMS shown in all statistics.



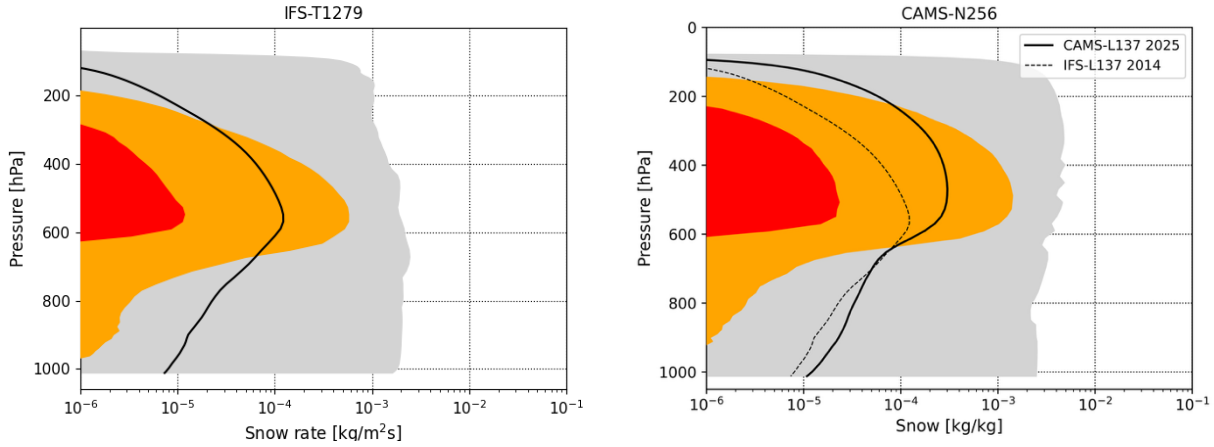


Figure 12. Same as Figure 10 but for rain (top) and snow (bottom). Note that precipitation used to be provided in units of rate [kg/m^{2}s] but is now in mass mixing ratio [kg/kg], consistent with all other gaseous species.

3 AC profiles

3.1 Greenhouse gases

The first two species shown in this section, carbon dioxide and methane, are obtained from the CAMS global greenhouse gas (CAMS GG) model based on the advice from the CAMS team at ECWMF. This model is designed to be as accurate as possible for these quantities. Nitrous oxide is not present in CAMS GG so is obtained from the CAMS AC model.

3.1.1 Carbon dioxide

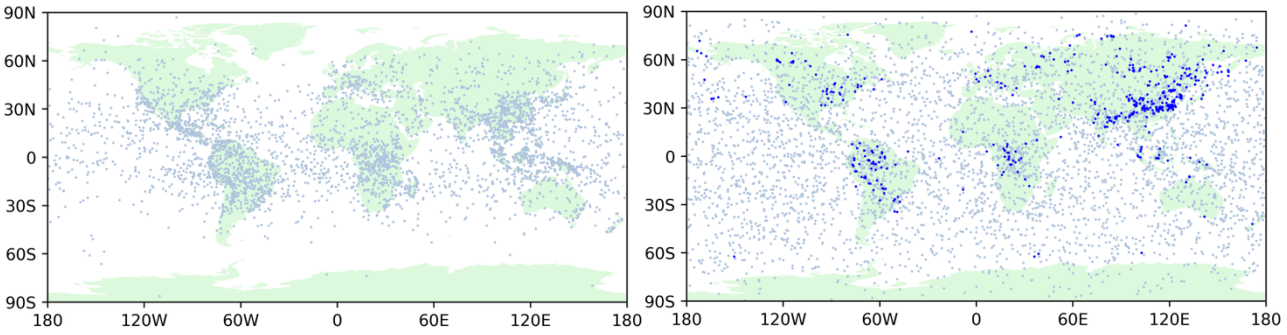


Figure 13. Locations of carbon dioxide profiles in the sampled subsets of the previous MACC (left) databases compared with the equivalent set produced from the CAMS GG. For CAMS the final 500 profiles derived from the selection algorithm are shown in blue pixels, whereas the 4500 randomly selected profiles are shown in blue grey pixels.

The selection algorithm used to create the MACC dataset placed a strong emphasis on selecting dissimilar profiles (see Appendix A) hence the spatial differences between the maps in Figure 13 reflect the different algorithms, with a less even distribution in the MACC dataset. Even so, in the CAMS dataset it is clear that there are several clusters of highly variable profiles around China picked up by the selection algorithm, which could be a result

of high urban emissions, and also around South America and Africa, which could be due to biomass burning.

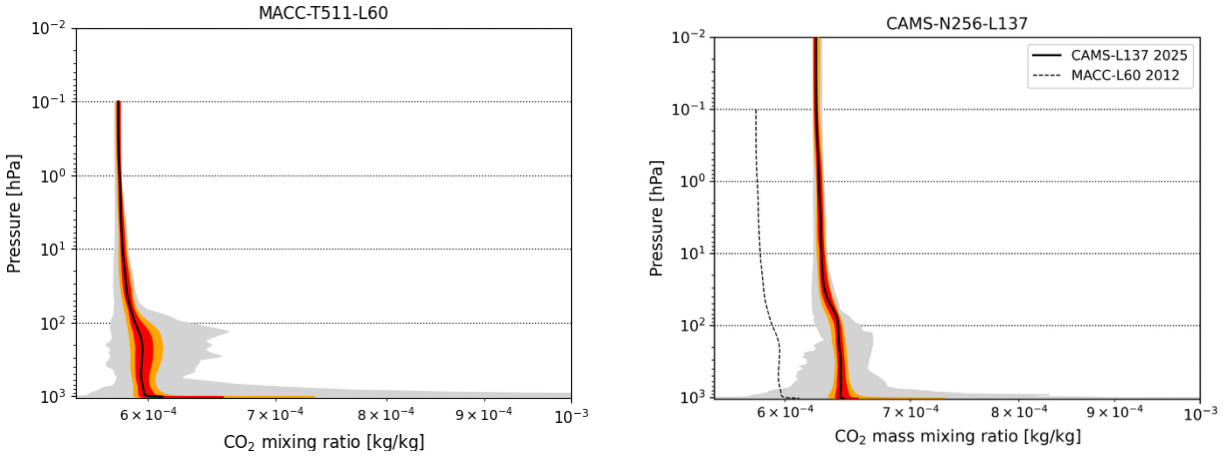


Figure 14. Distribution of carbon dioxide profiles in the respective subsets of the previous MACC (left) and CAMS (right) databases. Grey shading indicates the range constrained by the minimum and maximum values, orange shading is the range constrained by the 10th and 90th percentiles, and red is the lower and upper quartiles (25th and 75th). The black lines are the median profiles.

In terms of vertical profile distribution the concentration of carbon dioxide has obviously increased in in the intervening 13 year time period since MACC and the quantile distribution appears tighter in the more recent subset (Figure 14). As mass mixing ratio is less a common place unit for presenting carbon dioxide profiles see Appendix B: Model level to pressure conversion where the data is also presented in ppmv.

3.1.2 Methane

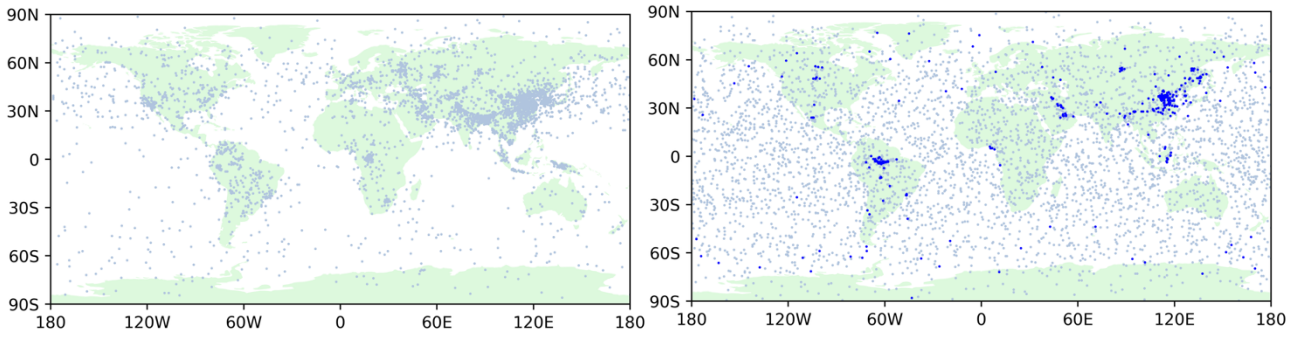


Figure 15. Same as Figure 13 but for methane. MACC (left) and CAMS GG (right).

The spatial distribution of profiles for methane is somewhat different to carbon dioxide, showing a visible cluster of selected profiles above the Middle East, for example (Figure 15). The statistics of the profiles are reasonably similar between MACC and CAMS (Figure 16), apart from the obvious extension to higher levels and an increase in tropospheric concentrations in the latter.

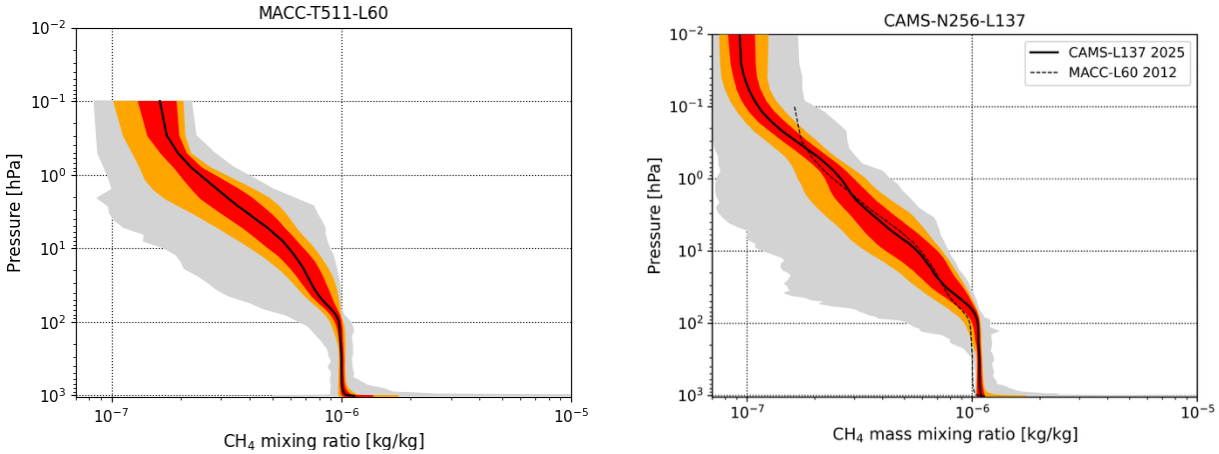


Figure 16. Same as Figure 14 but for methane.

3.1.3 Nitrous Oxide

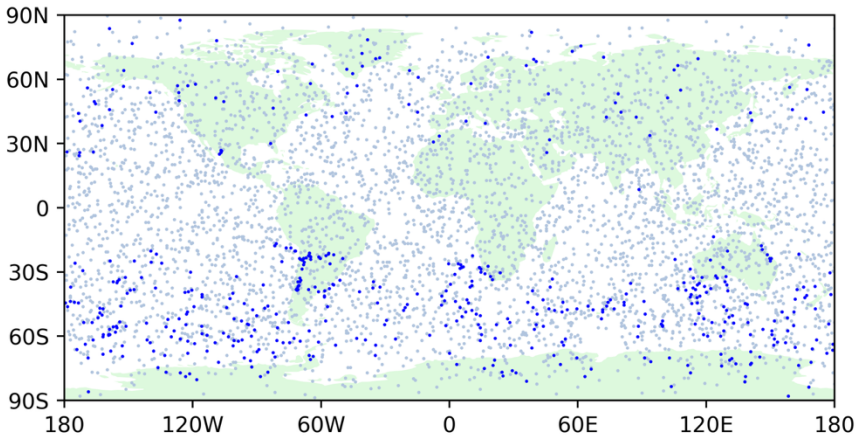


Figure 17. Locations of nitrous oxide profiles from CAMS AC. The final 500 profiles derived from the selection algorithm are shown in blue pixels and the 4500 randomly selected profiles are shown in blue grey pixels.

As N_2O has not featured in a previous profile dataset there is no comparison possible here. The spatial distribution shows that disparate profiles (blue pixels) appear to be selected from mainly over the Southern Ocean, which is interesting given the major source of N_2O is land (Figure 17). The maximum concentration is close to a fixed constant in the tropospheric plume, which spreads out and reduces steadily at upper levels (Figure 18). Note that unlike the other two greenhouse gases CO_2 and CH_4, N_2O was only available from the CAMS AC model.

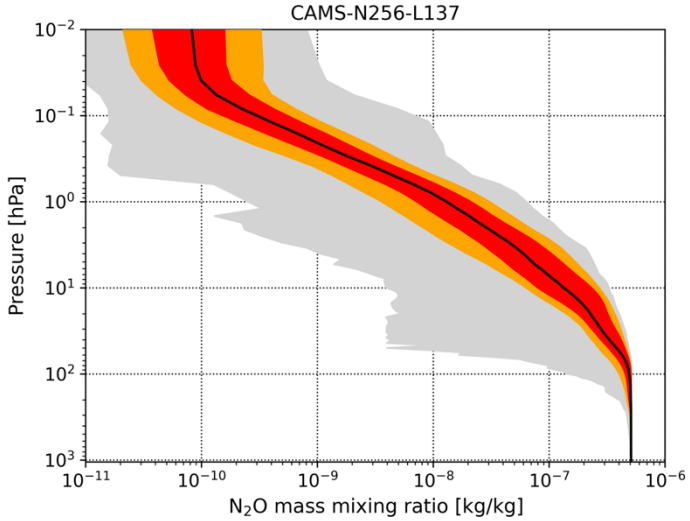


Figure 18. Distribution of nitrous oxide profiles in the CAMS database.

3.2 Reactive gases

3.2.1 Carbon monoxide

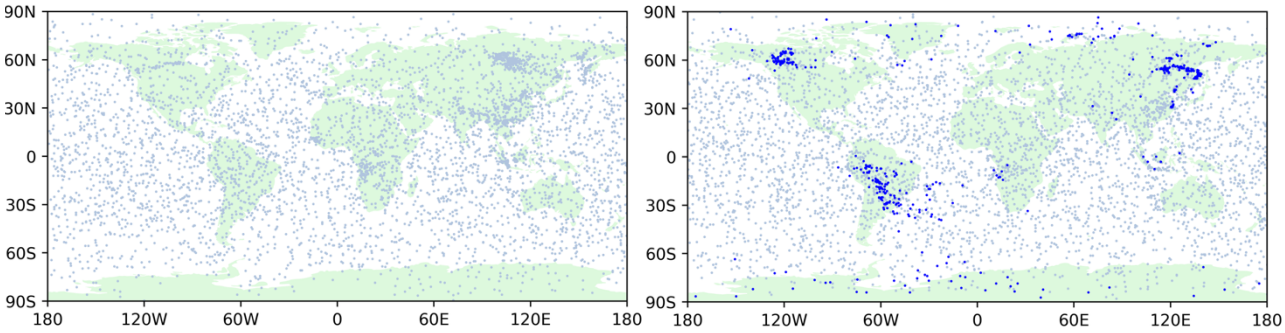


Figure 19. Locations of carbon monoxide profiles in the sampled subsets of the previous CAMS 60 level (left) database from 2016 compared with the equivalent set produced from current CAMS-L137 v1 (right). The final 500 profiles derived from the selection algorithm are shown in blue pixels and the 4500 randomly selected profiles are shown in blue grey pixels.

Figure 19 shows there is a strong cluster of carbon monoxide profiles that have been selected for their differences in North-West Canada, Southern America and parts of Russia, which could be due to wildfires and are therefore highly variable in time, but this may not be the only source.

The vertical profiles distributions of CO look quite different in the two versions of CAMS, mainly due to the increase and extension at levels above around 100 hPa in the latest dataset (Figure 20). The increased concentrations between 1 and 0.1 hPa to beyond surface values is particularly striking.

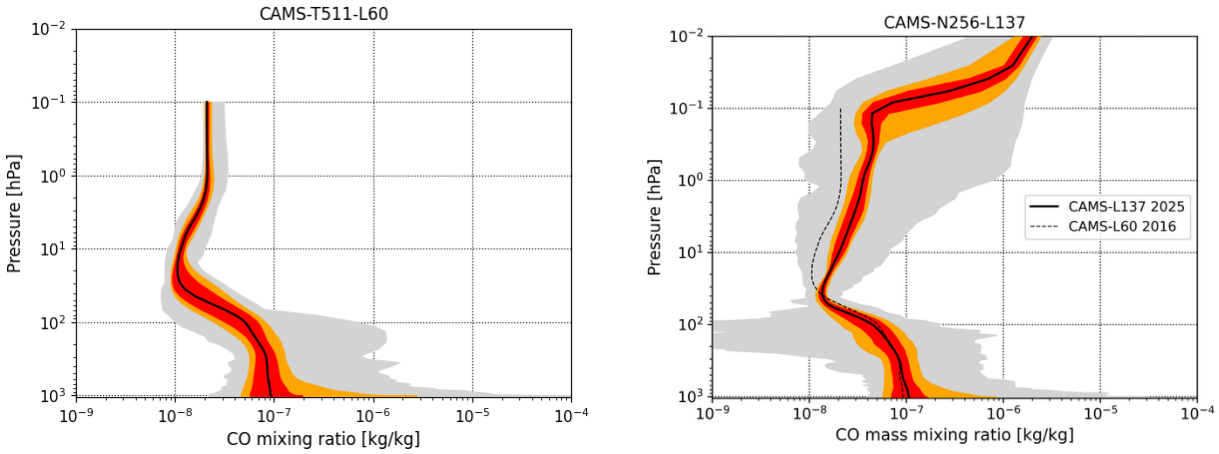


Figure 20. Distribution of carbon monoxide profiles in the respective subsets of the previous CAMS-L60 (left) and the present CAMS-L137 v1 (right) databases. Grey shading indicates the range constrained by the minimum and maximum values, orange shading is the range constrained by the 10th and 90th percentiles, and red is the lower and upper quartiles (25th and 75th). The black lines are the median profiles.

3.2.2 Nitrogen Dioxide

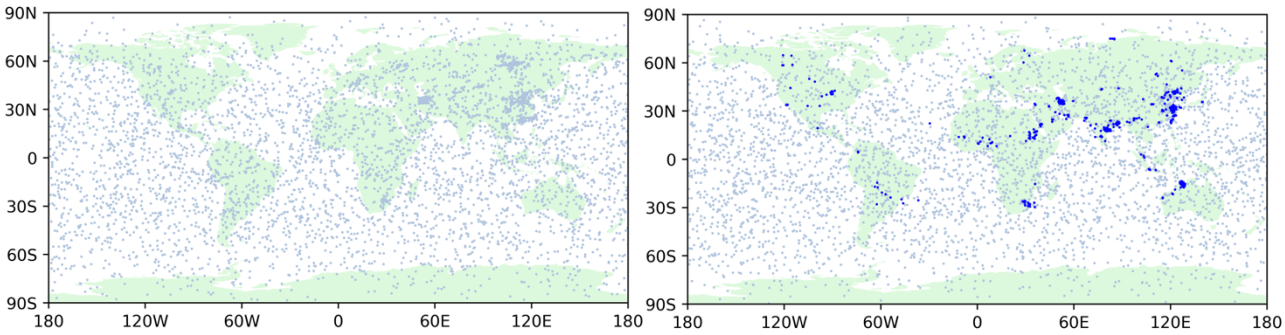


Figure 21. Same as Figure 19 but for nitrogen dioxide. CAMS-60L from 2016 (left) and CAMS-137L v1 (right).

Patterns of NO_2 profiles look relatively similar between previous and present CAMS shown in Figure 21, apart from perhaps the strong cluster in Northern Australia in the current dataset, which could be due to recent industrial emissions. The shape of the profiles has changed significantly between past and present CAMS (Figure 22). The previous inversion between 300 and 50 hPa is now absent and instead the concentration is shown to increase up to around 10 hPa before reducing with altitude but still exceeding values from the former dataset. In a sense it looks like it has been stretched vertically.

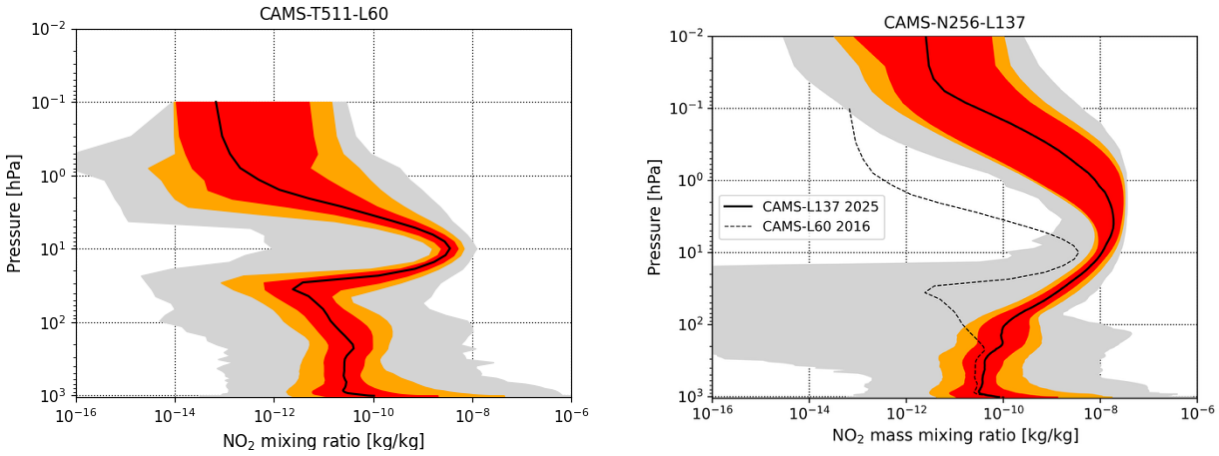


Figure 22. Same as Figure 20 but for nitrogen dioxide.

3.2.3 Sulphur dioxide

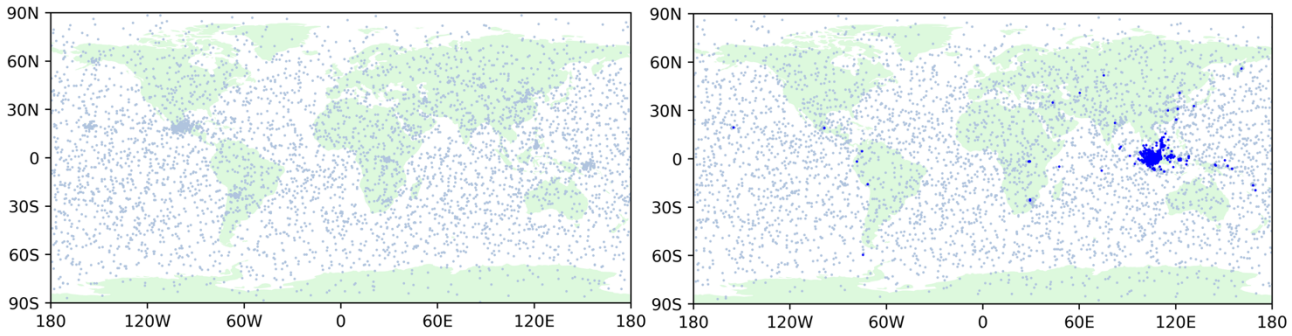


Figure 23. Same as Figure 19 but for sulphur dioxide. CAMS-60L (left) and CAMS-137L v1 (right).

There one strong cluster of sulphur dioxide profiles in each dataset that does not appear in the other. The latest version shows a strong concentration of disparate profiles over Indonesia, whereas the previous CAMS has a cluster over Mexico instead (Figure 23). Again, the shape of the profile statistics is very different in the latest CAMS (Figure 24), which has higher concentrations and stronger variability at all levels relative to CAMS-L60 database. The former profiles all reduce to a constant near zero value above the troposphere whereas the current profiles increase to near surface concentrations by 0.1 hPa.

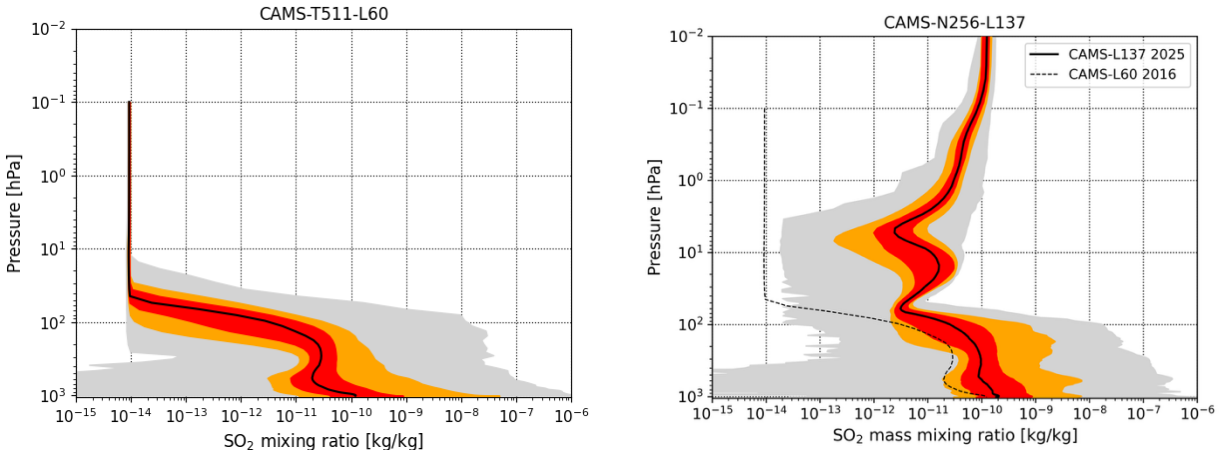


Figure 24. Same as Figure 20 but for sulphur dioxide.

3.2.4 Formaldehyde

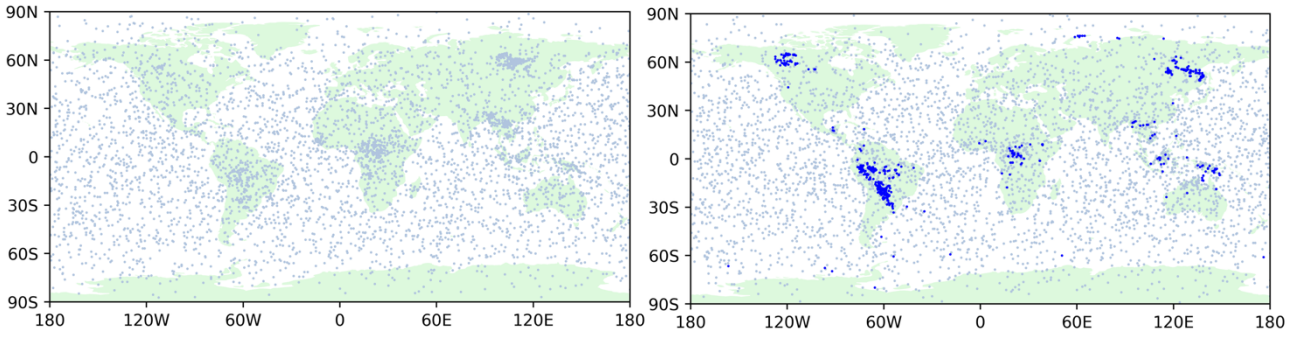


Figure 25. Same as Figure 19 but for formaldehyde.

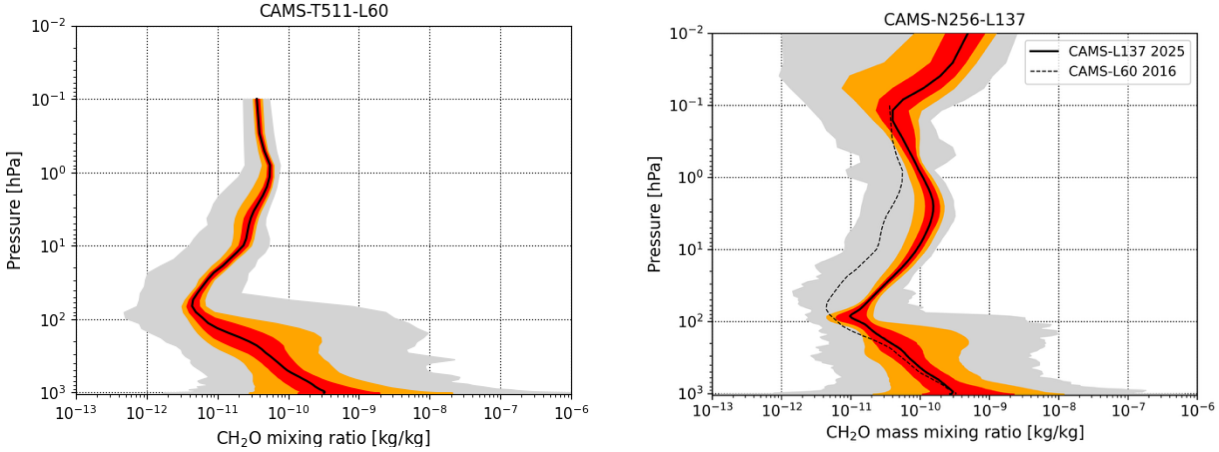


Figure 26. Same as Figure 20 but for formaldehyde.

There are similar spatial (and quantile) patterns in the CAMS-L137 v1 formaldehyde maps (Figure 25) to those of carbon monoxide. Similar to the other reactive gases, concentrations at upper levels in the latest CH_2O dataset show high variability and increasing values above 100 hPa relative to the previous CAMS subset (Figure 26).

3.3 Aerosols

Over the past few years aerosol profiles have benefited from improved representation so unlike the previous MACC release no upper levels are cautioned against in the current CAMS-L137 v1 dataset. The mass mixing ratio range has also been extended towards lower values from 10^{-20} to 10^{-25} kg/kg for all species.

3.3.1 Sea salt

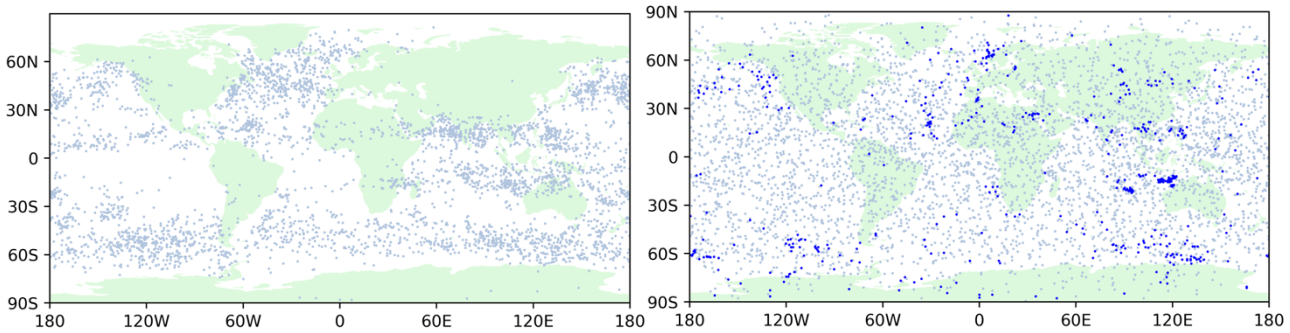


Figure 27. Locations of sea salt aerosol profiles in the sampled subsets of the previous MACC (left) databases compared with the equivalent set produced from CAMS-L137 v1. For CAMS the final 500 profiles derived from the selection algorithm are shown in blue pixels and the 4500 randomly selected profiles are shown in blue grey pixels.

Sea salt profiles are split into three bins based on particle size. Bin one contains all particles between 0.03 - 0.5 μm, bin two is 0.5 - 5 μm and bin three is 5 - 20 μm. Profiles from all bins are considered together in the selection algorithm so the global distribution for all three are the same. The selection algorithm used to create the MACC dataset placed a stronger emphasis on selecting disparate profiles (see Appendix A), so the spatial differences between the maps in Figure 27 reflect the different algorithms. The MACC profiles are almost all over ocean as the only source of sea salt hence where the wider range of profiles are likely to be, whereas the CAMS profiles are spread more evenly over land and sea reflecting the high proportion of random profiles selected.

The vertical distribution of sea salt has significantly changed for all size bins since the MACC database was released (Figure 28). Overall concentrations have decreased in the troposphere. However, at higher levels the two bins with larger particles now display a narrow plume with a constant value that tapers off at around 1 hPa and 10 hPa, for bin 2 and 3, respectively, relative to nearly zero concentrations in the MACC datasets. Bin 1 is the only one whose profiles extend to 0.01 hPa, with a slightly reducing concentration, whereas the previous subset comprised more vertical profiles with a near constant and higher value.

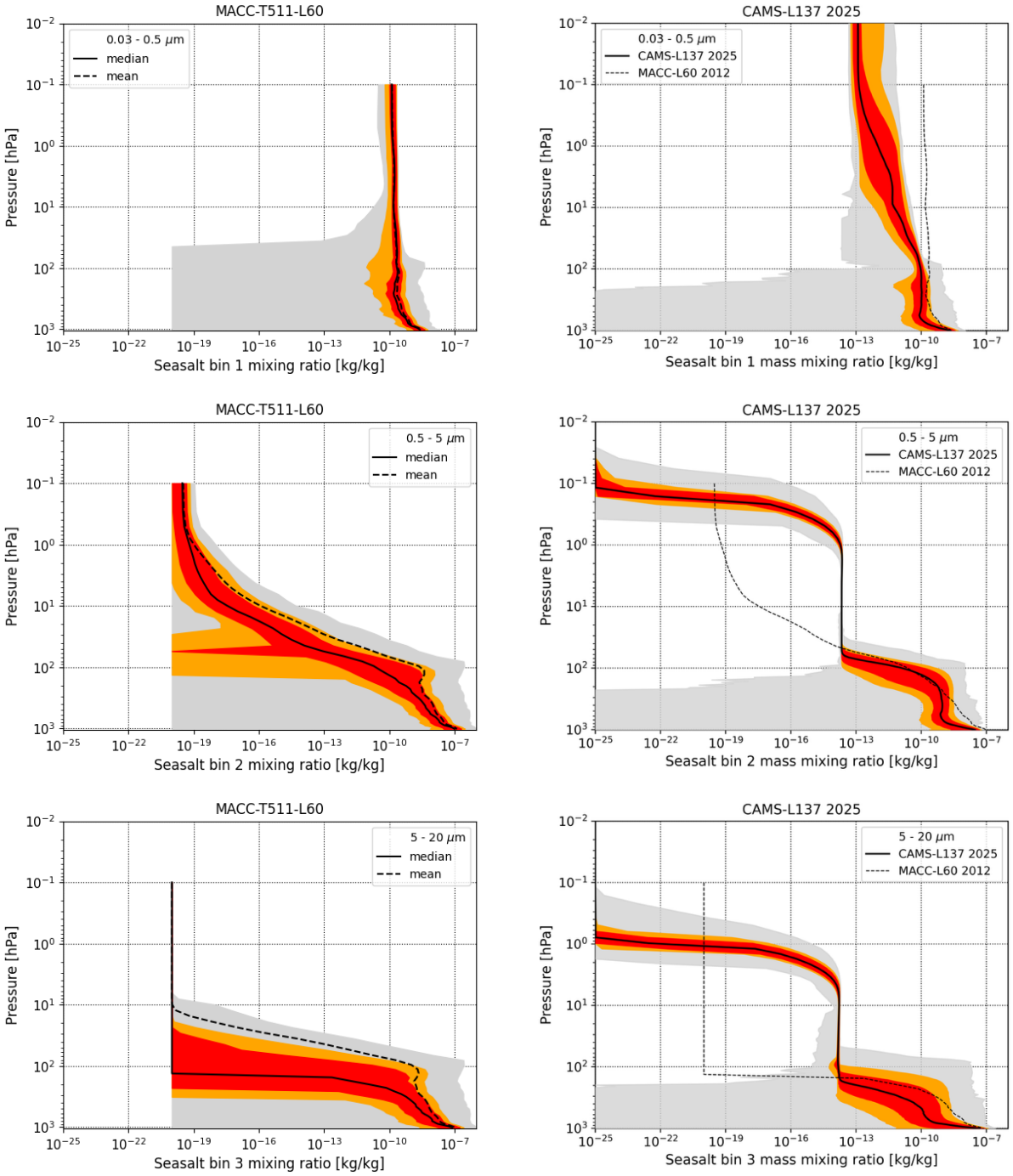


Figure 28. Distribution of sea salt profiles in the respective three particle size bins (top to bottom) of the previous MACC (left) and the CAMS (right) databases. Grey shading indicates the range constrained by the minimum and maximum values, orange shading is the range constrained by the 10th and 90th percentiles, and red is the lower and upper quartiles (25th and 75th). For MACC the black solid line is the median profile and the black dashed line is the mean. For CAMS both solid and dashed lines are medians.

3.3.2 Desert Dust

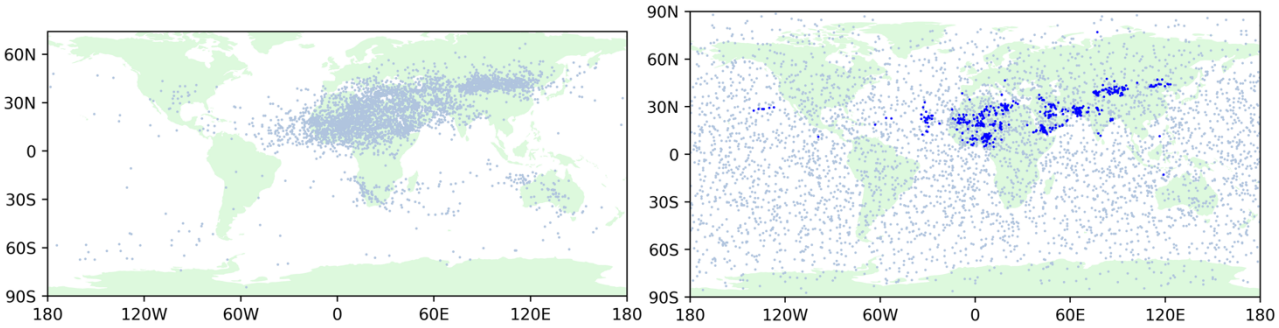
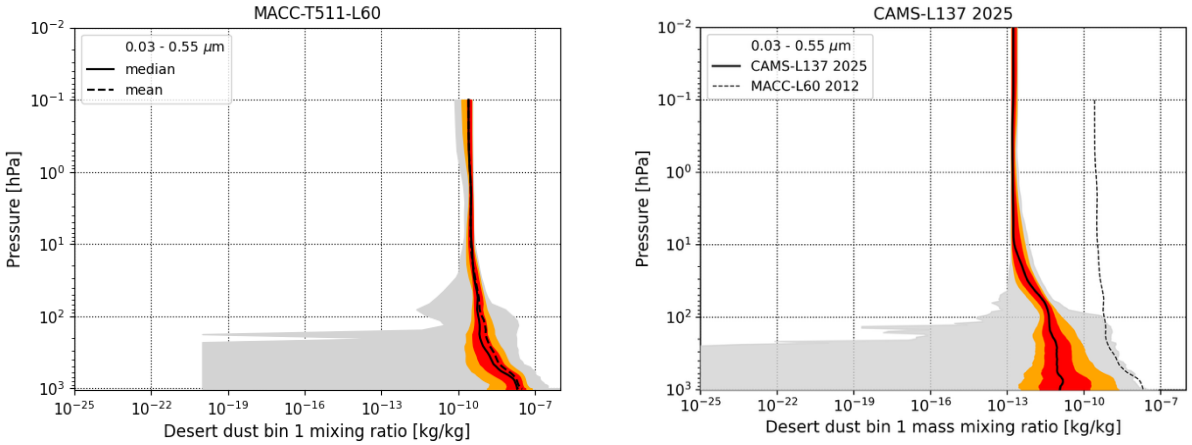


Figure 29. Same as Figure 27 but for desert dust. MACC-60L from 2012 (left) and CAMS-137L v1 (right).

Like sea salt, desert dust is also split into three bins based on particle size. Bin one contains all particles between 0.03 - 0.55 μm, bin two is 0.55 – 0.9 μm and bin three is 0.9 - 20 μm. The primary sources of desert dust are Northern Africa, the Middle East and parts of Asia, which is evident in both the MACC and CAMS dataset, the latter from selection algorithm clusters indicating more disparate profiles over these regions.

In bins 1 and 2 desert dust concentrations have decreased since the last MACC dataset, particularly in the former. However, bins 2 and 3 have become more tightly concentrated at higher levels and now have a similar structure to bin 1. Concentrations above 10 hPa in bin 3 were practically zero in the past and now there is a plume with a constant higher value from 100 to 0.1 hPa.



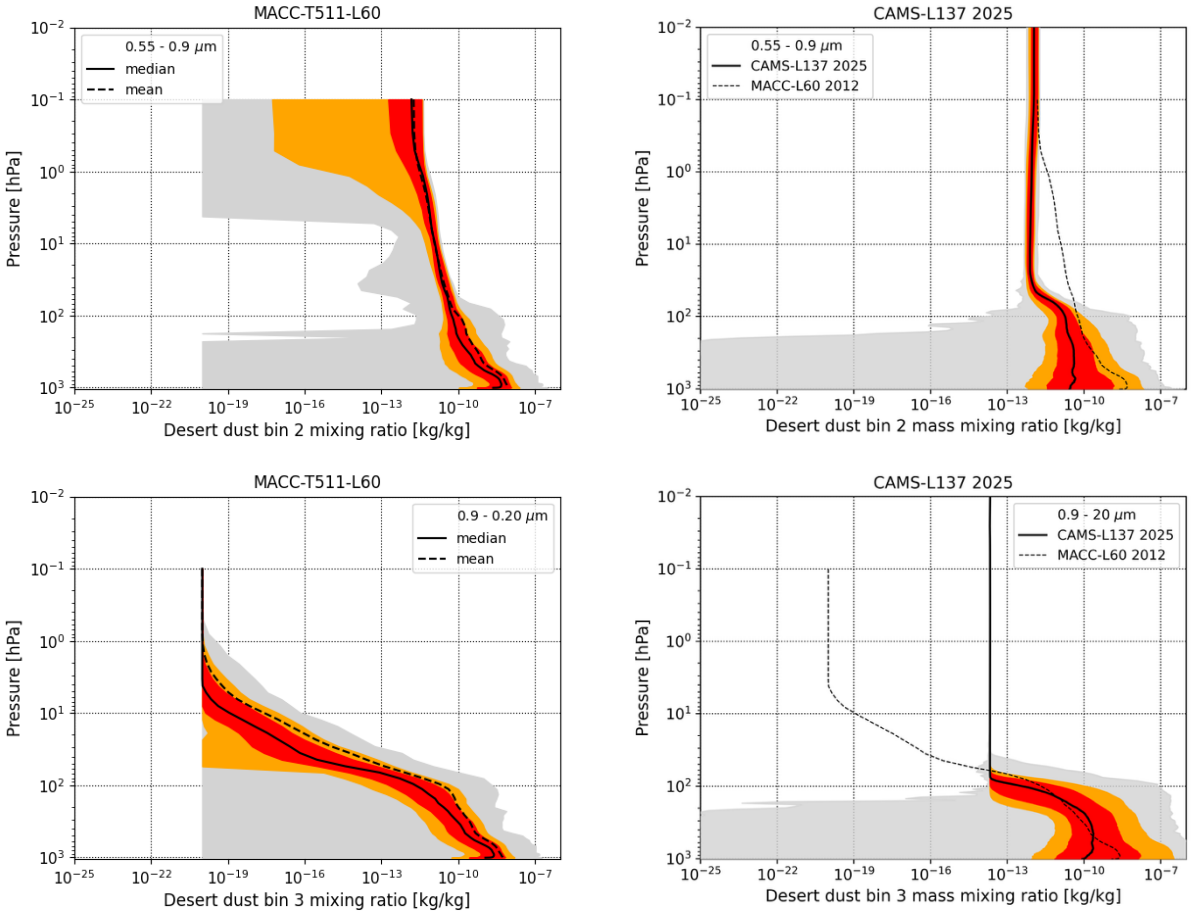


Figure 30. Same as Figure 28 but for desert dust.

3.3.3 Sulphate

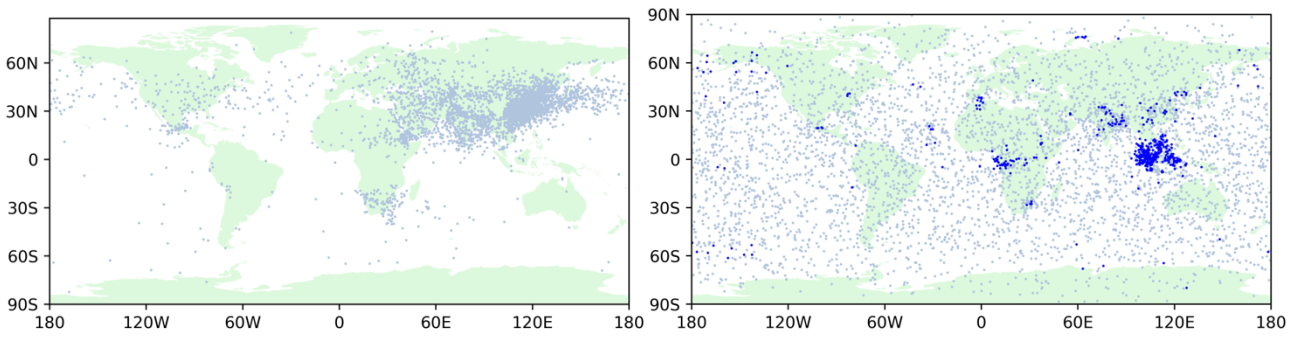


Figure 31. Same as Figure 27 but for sulphate aerosol. MACC-60L from 2012 (left) and CAMS-137L v1 (right).

Sulphate profiles identified by the selection algorithm appear to have changed in at least one spatial area between MACC and CAMS (Figure 31), where previously most were over China these have reduced significantly and a cluster has appeared over Indonesia instead.

The location of the recent cluster is consistent with that seen in SO_2 and could possibly be attributed to the volcanic eruption of Mount Ruang in North Sulawesi on 16th April 2024.

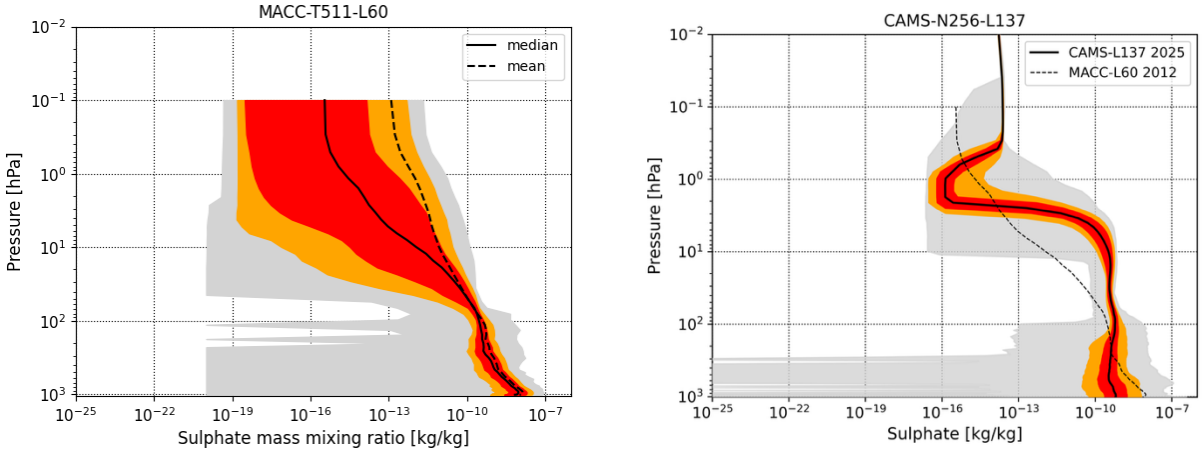


Figure 32. Distribution of sulphate profiles in the previous MACC-60L (left) and the CAMS-L137 v1 (right) databases. Grey shading indicates the range constrained by the minimum and maximum values, orange shading is the range constrained by the 10th and 90th percentiles, and red is the lower and upper quartiles (25th and 75th). For MACC the black solid line is the median profile and the black dashed line is the mean. For CAMS both solid and dashed lines are medians.

The statistics of the vertical distribution of sulphate mass mixing ratio profiles are significantly different between datasets (Figure 32). Above the troposphere the profiles show a tighter range of concentrations with a pronounced inversion around 10 hPa in the latest dataset. This feature is not present in the previous MACC profiles.

3.3.4 Organic matter

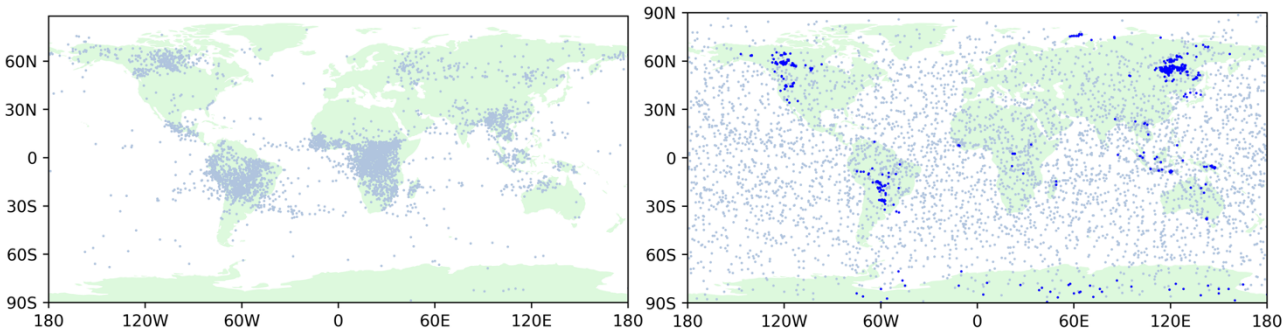


Figure 33. Same as Figure 27 but for organic matter. MACC-60L from 2012 (left) and CAMS-137L v1 (right).

Organic matter is split into two bins, a hydrophilic (particle size varies with humidity) and a hydrophobic (no variation with humidity) component. The spatial distribution in the new CAMS dataset is more evenly spread out than MACC apart from a new cluster of profiles selected for their differences in Russia above Japan that isn't present in MACC (Figure 33).

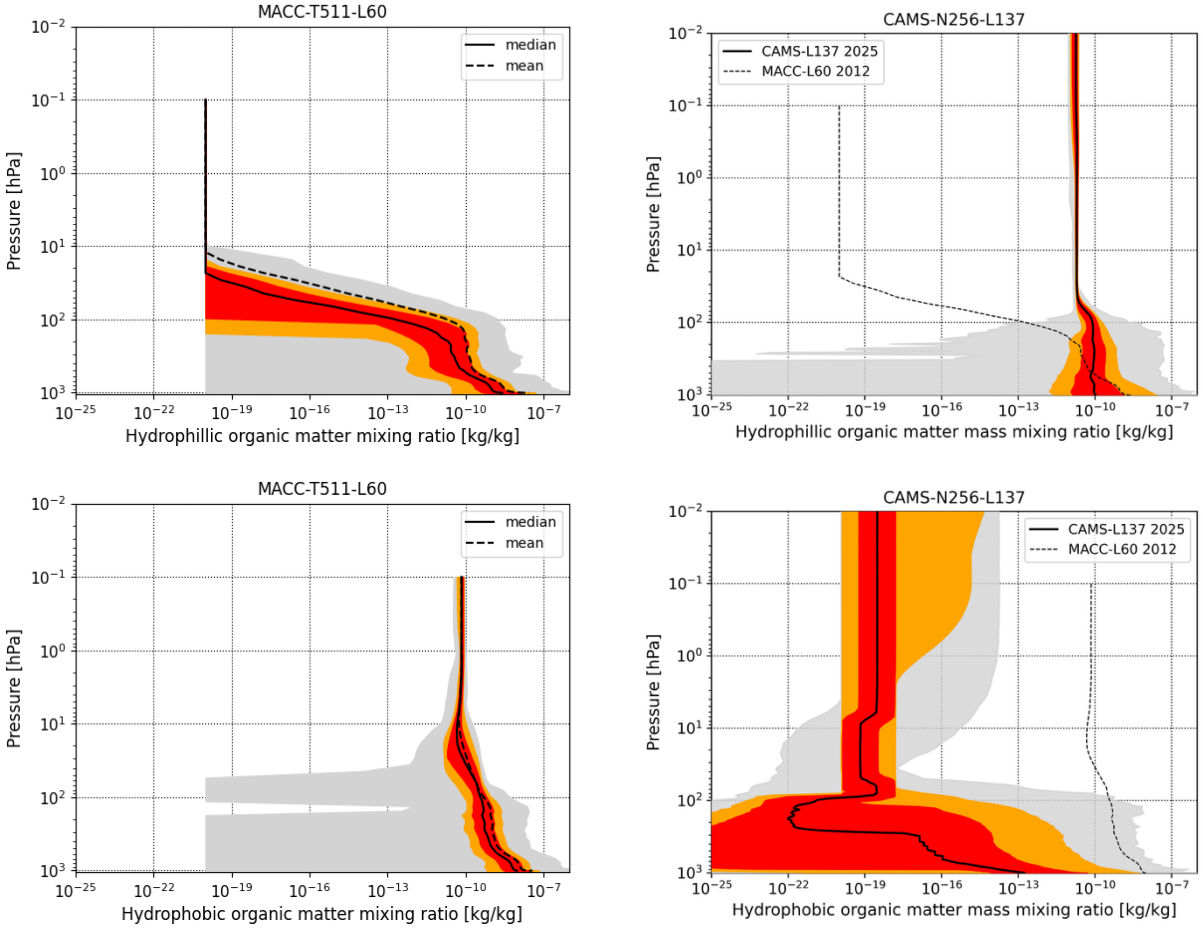


Figure 34. Distribution of organic matter profiles for the hydrophilic component (top) and the hydrophobic component (bottom) of the previous MACC (left) and the CAMS (right) databases. Grey shading indicates the range constrained by the minimum and maximum values, orange shading is the range constrained by the 10th and 90th percentiles, and red is the lower and upper quartiles (25th and 75th). For MACC the black solid line is the median profile and the black dashed line is the mean. For the CAMS panels both the solid and dashed lines are medians.

There has been a large shift in the behaviour of organic matter profiles between the two datasets. The hydrophilic element has significantly increased in value at upper levels above 100 hPa (Figure 34 – top row). For the hydrophobic element, by contrast, there is a strong reduction and an unusually blocky appearance in the new profiles. This suggests there may have been a change in the way the two elements are defined.

3.3.5 Black carbon

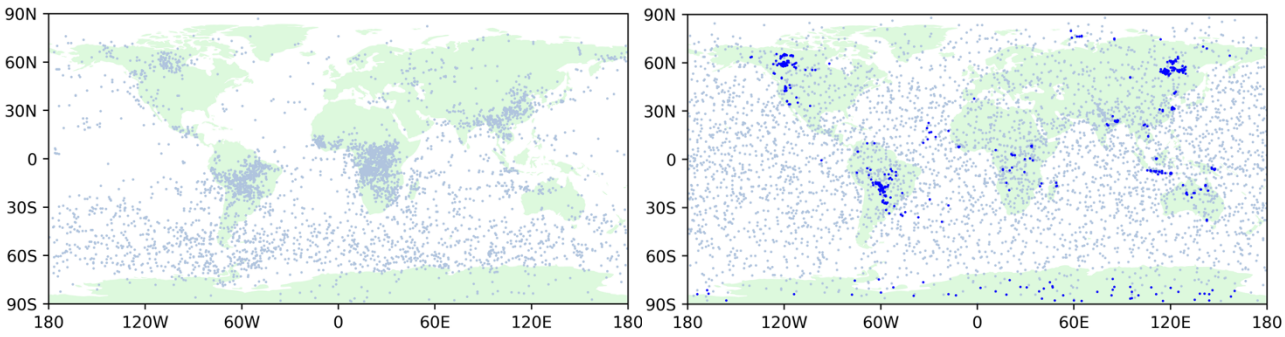


Figure 35. Same as Figure 27 but for black carbon. MACC-60L from 2012 (left) and CAMS-137L v1 (right).

Figure 35 shows that selected clusters of black carbon show a similar pattern to that of organic matter. The changes between profile statistics are also alike in structure (Figure 36) indicating a possible change in the way the hydrophilic and hydrophobic components are defined.

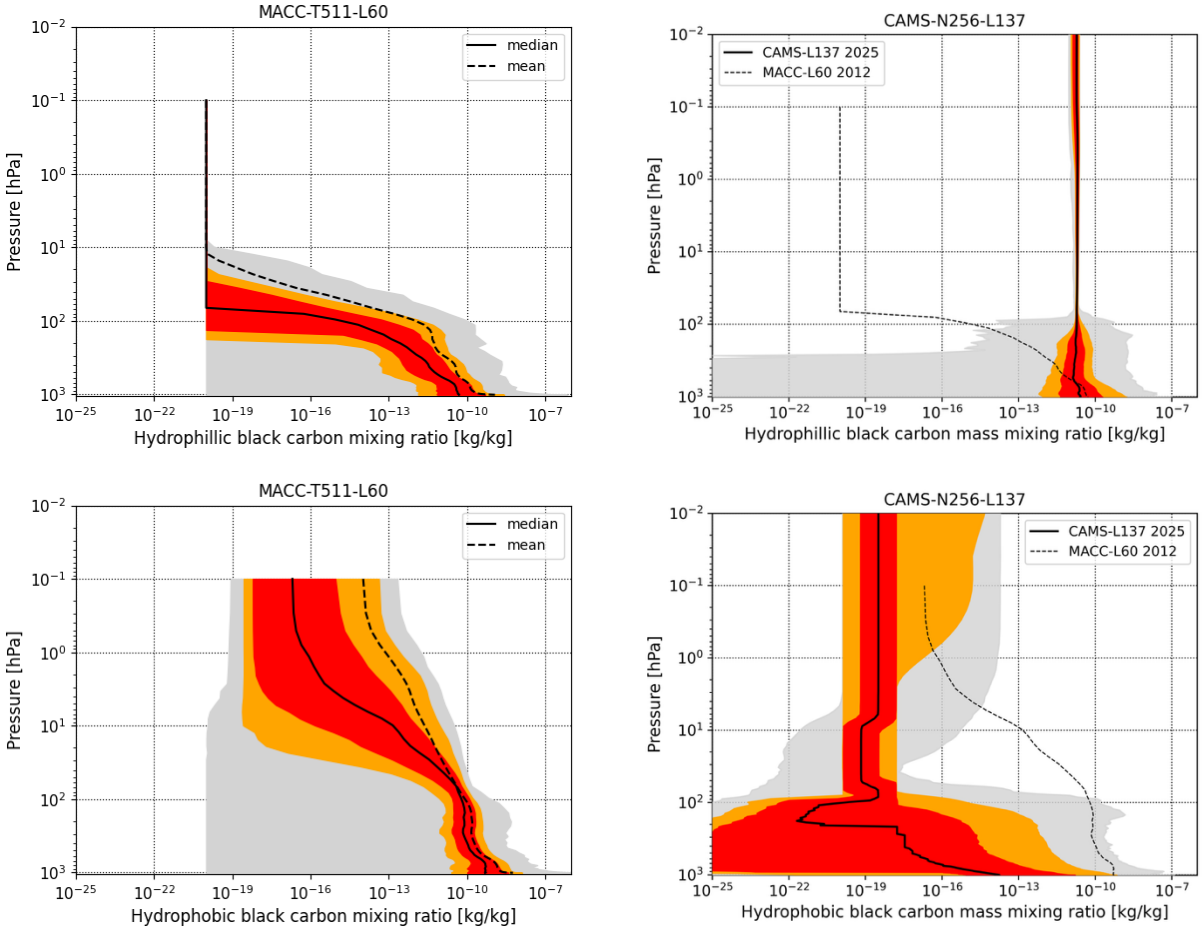


Figure 36. Same as Figure 34 but for hydrophilic (top) and hydrophobic (bottom) components of black carbon.

3.3.6 Ammonium

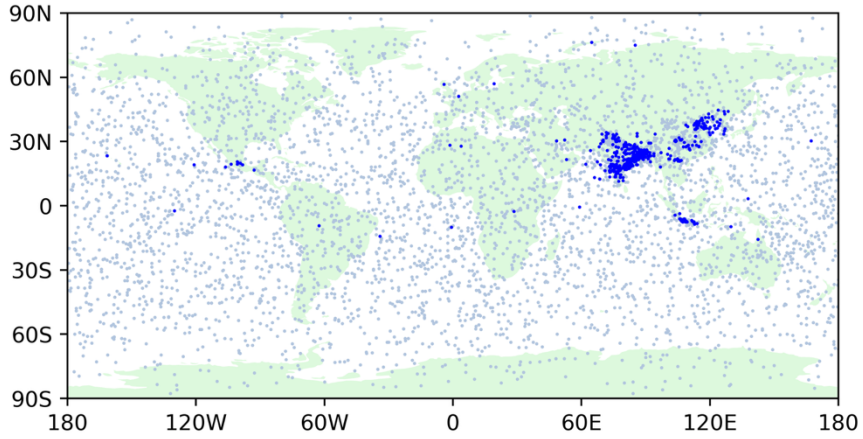


Figure 37. Locations of ammonium aerosol profiles in the CAMS-L137 v1 database. The final 500 profiles derived from the selection algorithm are shown in blue pixels and the 4500 randomly selected profiles are shown in blue grey pixels.

There is no predecessor ammonium database as it was added to CAMS in July 2019. The global distribution displayed in Figure 37 shows concentrated clusters of selected profiles over China and India, possibly being attributable to agricultural activities and the fertiliser industry.

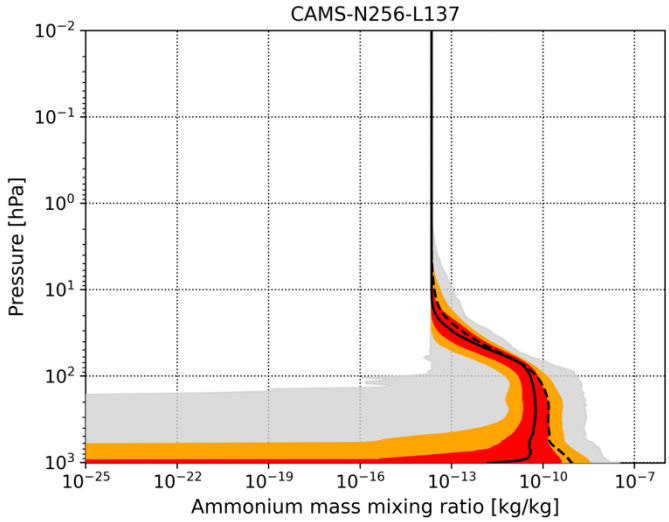


Figure 38. Distribution of ammonium profiles in the CAMS-L137 database. Grey shading indicates the range constrained by the minimum and maximum values, orange shading is the range constrained by the 10th and 90th percentiles, and red is the lower and upper quartiles (25th and 75th). The black solid line is the median profile and the black dashed line is the mean.

Ammonium tends to increase with height in the troposphere and then decreases to a plume with constant concentration at upper levels beyond 10 hPa (Figure 38).

3.3.7 Nitrate

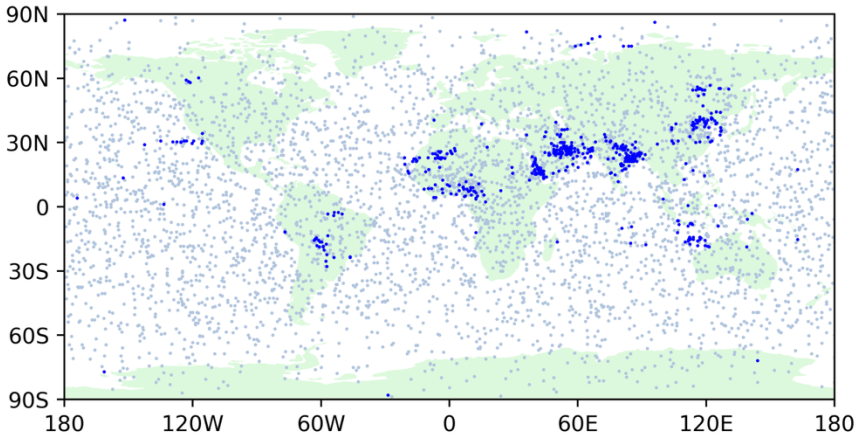


Figure 39. Same as Figure 37 but for nitrate aerosol.

There is no predecessor nitrate database as it was only added to CAMS in July 2019. Some clusters of selected profiles look similar to that of ammonium, which is not unlikely given the two are relates via ammonium nitrate (AN), a key component used in the mining and fertiliser industry. There are, however, clusters of variable nitrate profiles around the Middle East and Africa that are not apparent in the ammonium maps.

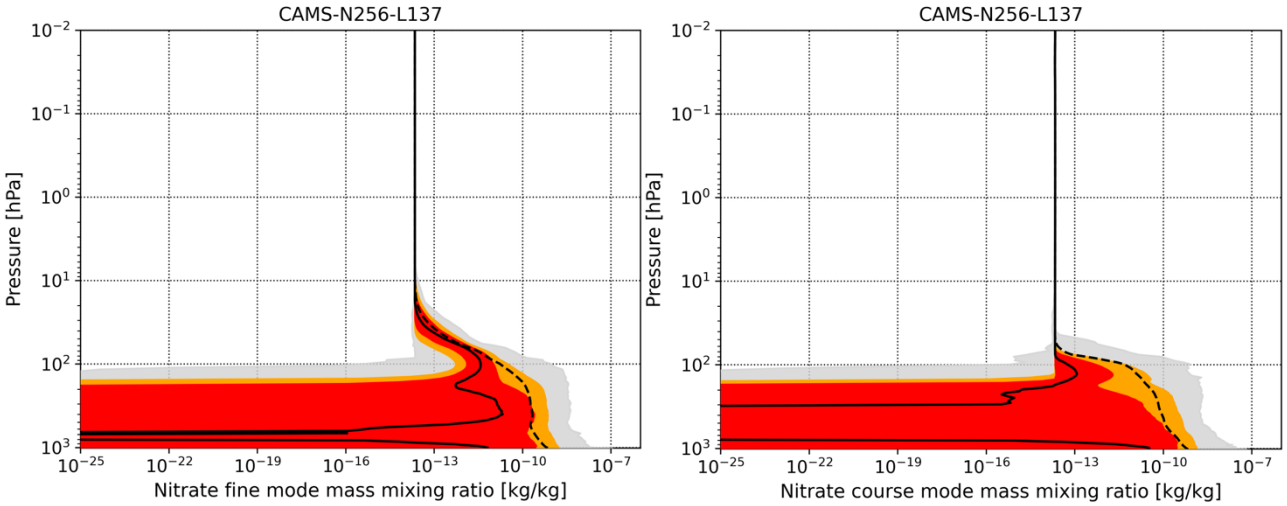


Figure 40. Same as Figure 38 but for nitrate fine mode (left) and nitrate course mode (right).

Both fine and course modes of nitrate have similar profile statistics to each other, with a lot of variability in the troposphere and a constant plume above 100 hPa for the course mode and above 10 hPa for the fine mode (Figure 40).

3.3.8 Secondary organic matter

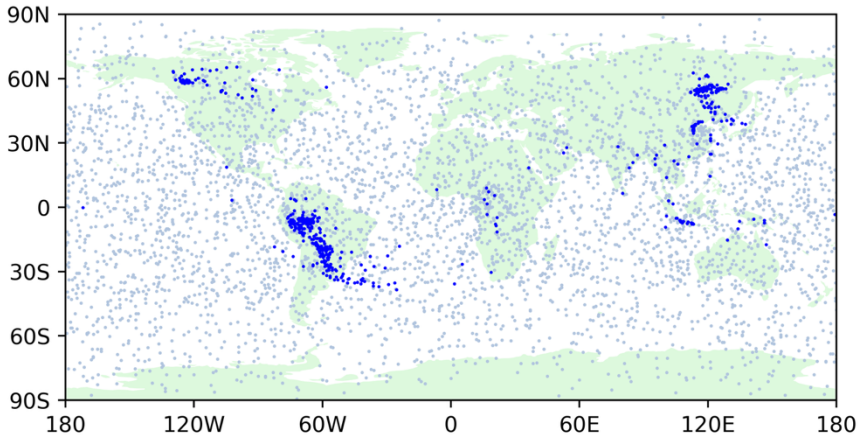


Figure 41. Same as Figure 37 but for secondary organic matter.

Secondary organic matter was implemented most recently in CAMS, in the Cy48r1 upgrade in June 2023. It is split into two bins from different sources, biogenic and anthropogenic. Note these were previously included in organic matter and that they are both hydrophilic. Secondary aerosols are formed in the atmosphere when precursor gases undergo chemical reactions, transforming into aerosol particles, unlike primary aerosols which are emitted directly from sources. Secondary organic aerosols are based on Volatile Organic Compound (VOC) emissions and whereas the biogenic component is naturally emitted by soil, plants and oceans the anthropogenic organic aerosols have man-made sources such as traffic exhaust. Figure 41 indicates there are concentrated clusters of variable profiles of secondary organic aerosol over parts of Southern America, North-West Canada, East Asia and Russia.

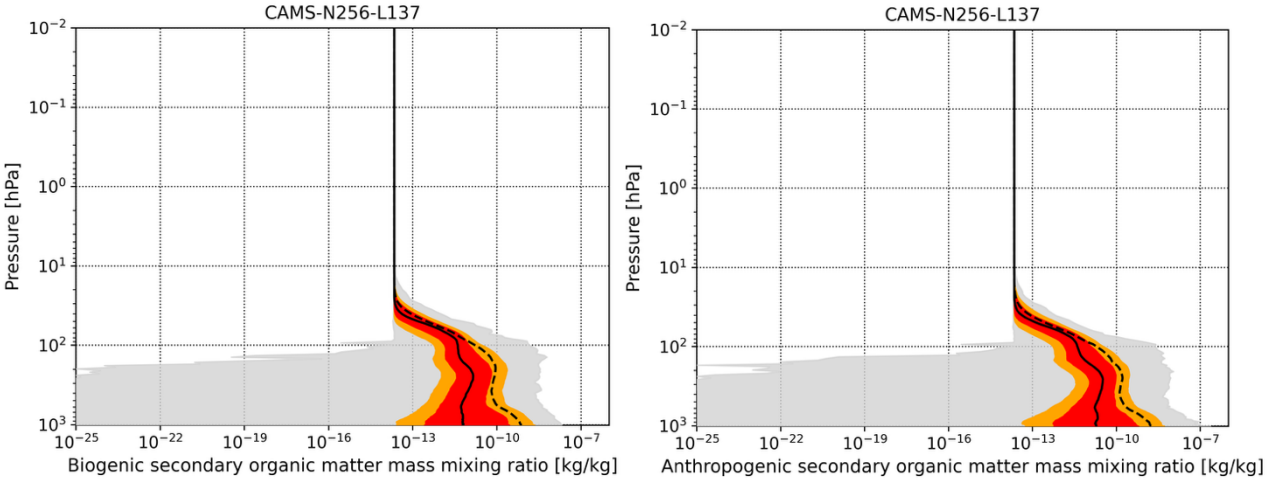


Figure 42. Same as Figure 38 but for biogenic secondary organic aerosol (left) and anthropogenic secondary organic aerosol (right).

Both components have almost identical profile statistics with a slight inversion in the troposphere and a narrow plume above 10 hPa, see Figure 42.

4 IFS O1280 profiles

4.1 Meteorological variables

What follows is a comparison of the CAMS-L137 v1 profiles with the equivalent set produced by the IFS in order to validate the meteorological quantities produced by CAMS, which could theoretically be seen as more accurate in the IFS due to the higher resolution. As is shown for ozone, however, the representation of the profile may not be as well represented in time as it comes from a climatology.

4.1.1 Thresholds

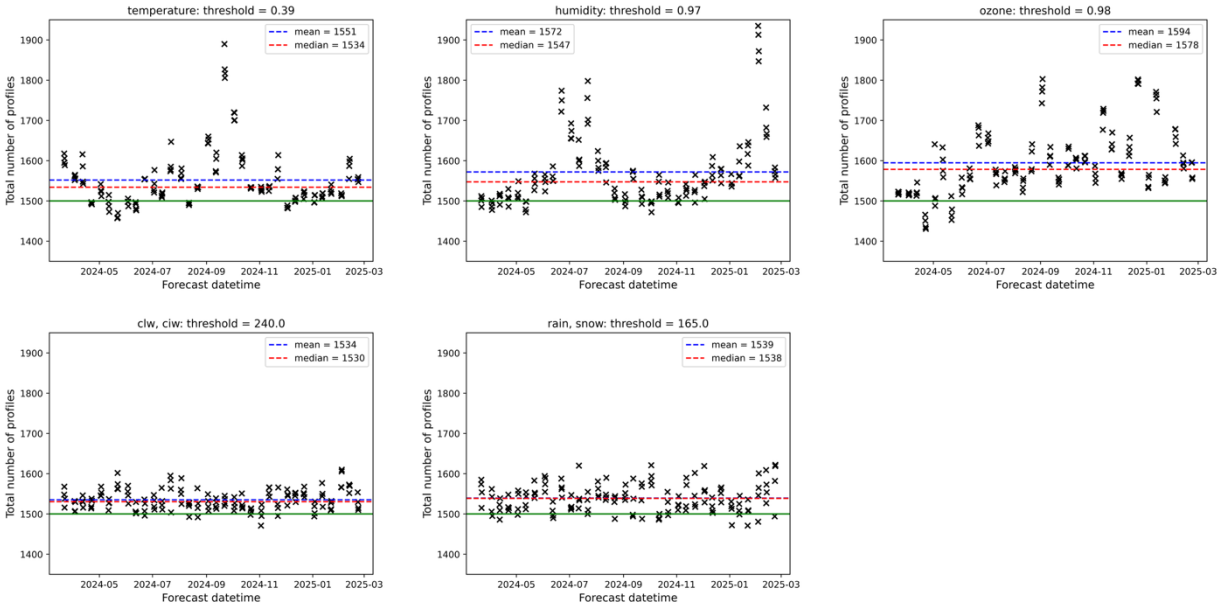


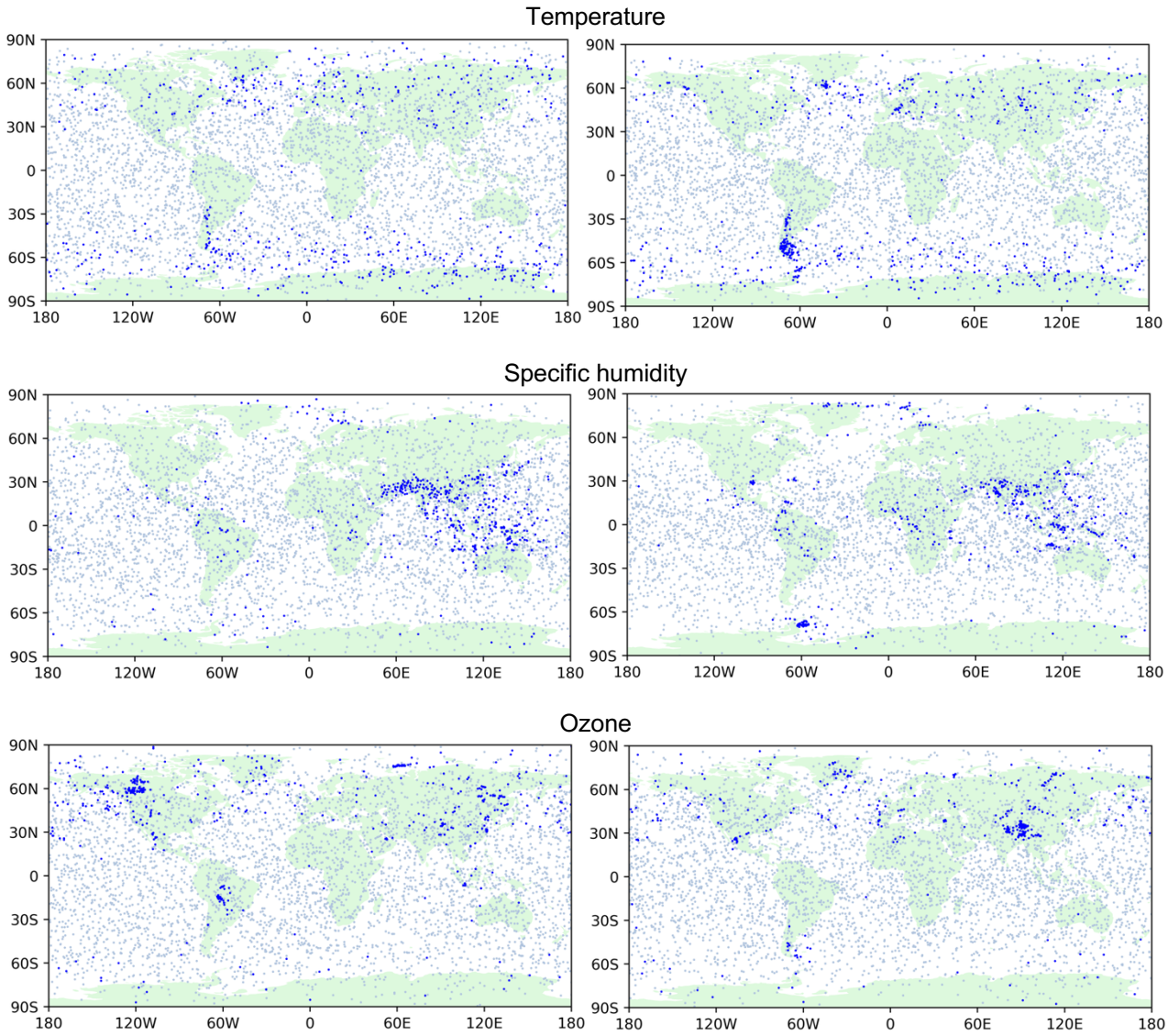
Figure 43. The total (random + selected) number of profiles retained from step one of the selection process, for each of the 136 samples (shown at their forecast date and time). Each panel displays a different variable with the name given in the title alongside the chosen threshold. The green line indicates 1500 profiles (the theoretical aim), and the red and blue dashed lines are the median and mean of the data, respectively.

The selection procedure for the IFS dataset followed the exact same procedure as the CAMS dataset to provide a completely independent test for the meteorological variables. Note that March 1st and 10th 2024 were not available for the Cy49r1 e-suite as it began on the 13th March 2024, so there are only 136 samples available for step 1. Table 2 lists the equivalent thresholds chosen for the IFS for both selection rounds. All are higher than the corresponding values for CAMS due to the higher horizontal resolution of the IFS. Figure 43 shows the distribution of the number of profiles selected at each forecast step for the chosen threshold in the first step. The variability of ozone total numbers is notably different to the CAMS equivalent.

Table 2. Applied threshold parameters for the first and second rounds of running the selection algorithm and total counts of selected profiles after step one, for each subset of the IFS version of the database.

Name	Short name	Bins	t step 1	$\sum S_O$ step 1	t step 2
Temperature	temp	1	0.39	211,059	0.72
Specific humidity	hum	1	0.97	213,811	1.62
Ozone	oz	1	0.98	216,884	1.45
Cloud condensate	clw, ciw	2	240.0	208,753	9.7
Precipitation	rain, snow	2	165.0	209,337	3.8

4.1.2 Spatial distribution of profiles



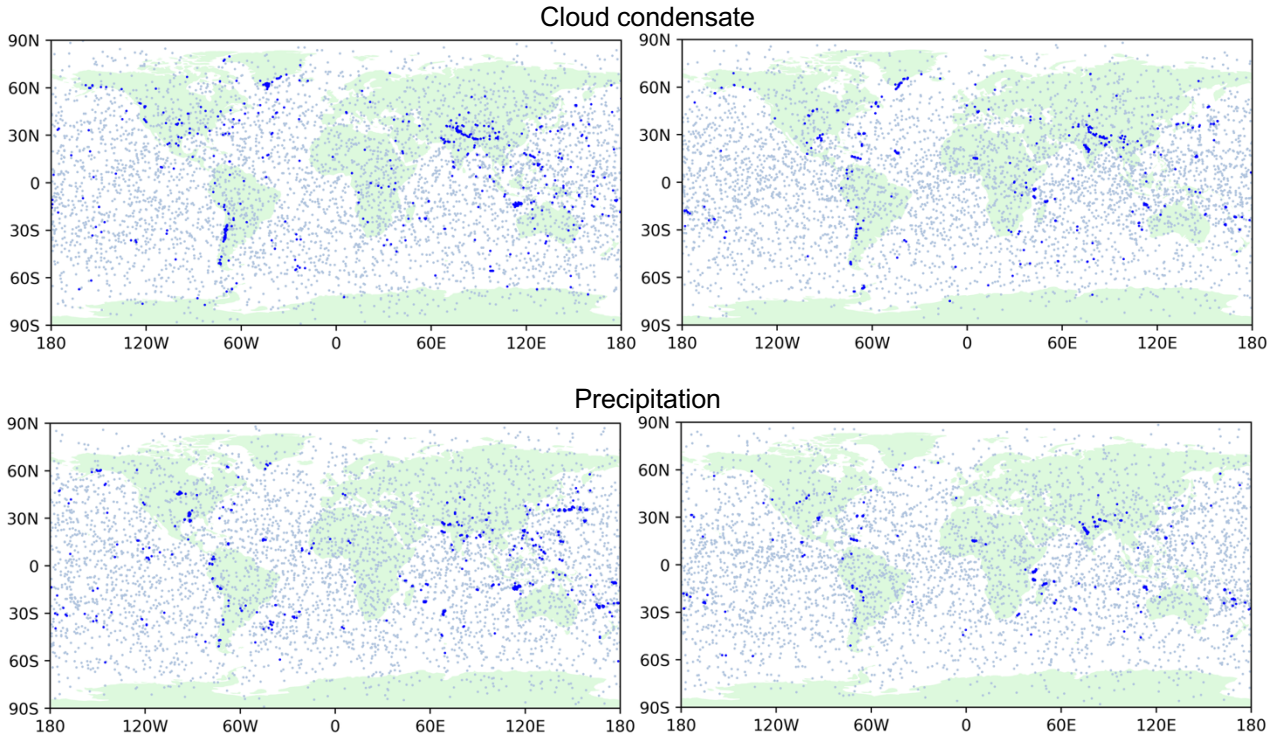


Figure 44. Locations of selected profiles in the sampled subsets of the CAMS-N256 (left) databases compared with the equivalent set produced from the IFS-O1280 (right). Variables are from top to bottom: temperature, specific humidity, ozone, cloud condensate (clw and ciw) and precipitation (rain and snow). The final 500 profiles derived from the selection algorithm are shown in blue pixels and the 4500 randomly selected profiles are shown in blue grey pixels.

On the whole the distribution of profiles show similar patterns for hydrometeor variables between CAMS and IFS in Figure 44, however there are some clusters of difference in temperature, humidity and ozone. The effects of resolution are seemingly evident for the temperature locations across the southern tip of America and the Antarctic Peninsula (the latter more so in the humidity maps), where the selection algorithm picks up more profiles in the IFS forecasts, indicating there are sufficiently different profiles produced when these areas are resolved at higher horizontal resolution. Ozone profile selection looks quite different between datasets in its spatial distribution. There is also a larger cluster of disparate ozone profiles selected above the Tibetan Plateau in the IFS that is not present in CAMS, which selects clusters above North-West Canada instead. Clouds and precipitation clusters are reasonably similar between datasets.

4.1.3 Profile statistics of sampled variables

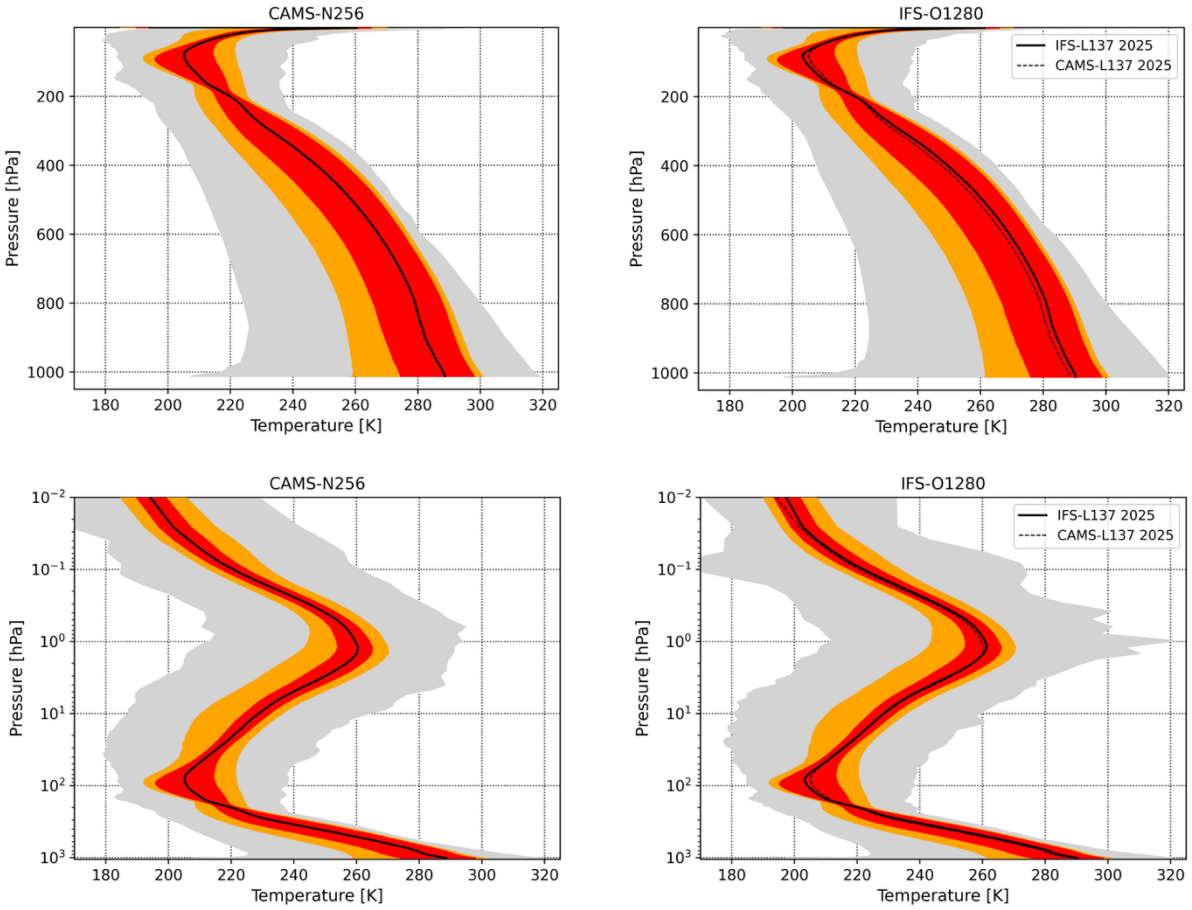


Figure 45. Distribution of temperature profiles in the respective subsets of the CAMS-N256-L137 (left) and the IFS-O1280-L137 (right) databases. Top panels have a linear pressure axis and bottom have a log pressure axis. Grey shading indicates the range constrained by the minimum and maximum values, orange shading is the range constrained by the 10th and 90th percentiles, and red is the lower and upper quartiles (25th and 75th). The black lines are the median profiles.

There are small differences between statistical profile quantiles of temperature and humidity. The median profile from the IFS is up to 1 K warmer than CAMS in the troposphere and shows a wider range of temperatures around the stratopause and mesosphere (Figure 45). For humidity there are also very slightly moister profiles in the troposphere (Figure 46).

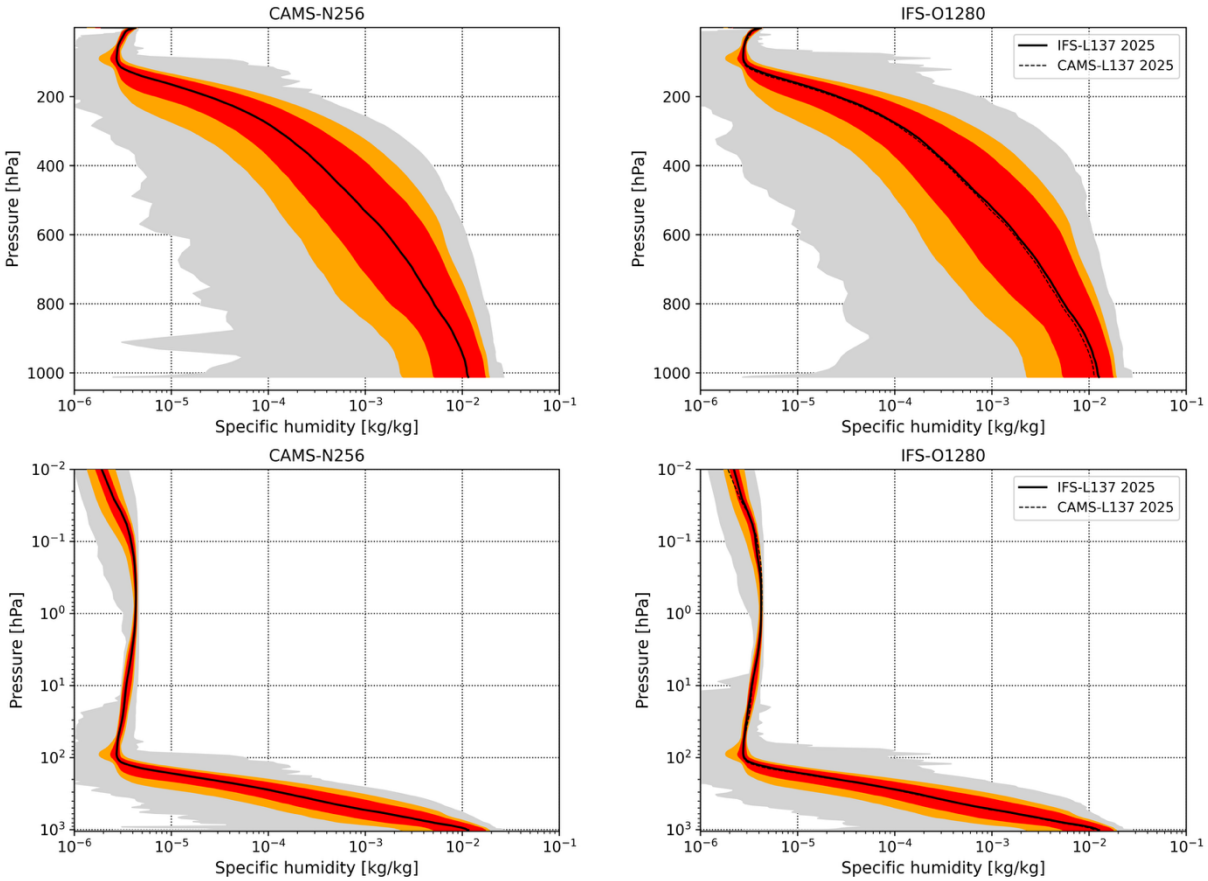


Figure 46. Same as Figure 45 but for humidity profiles.

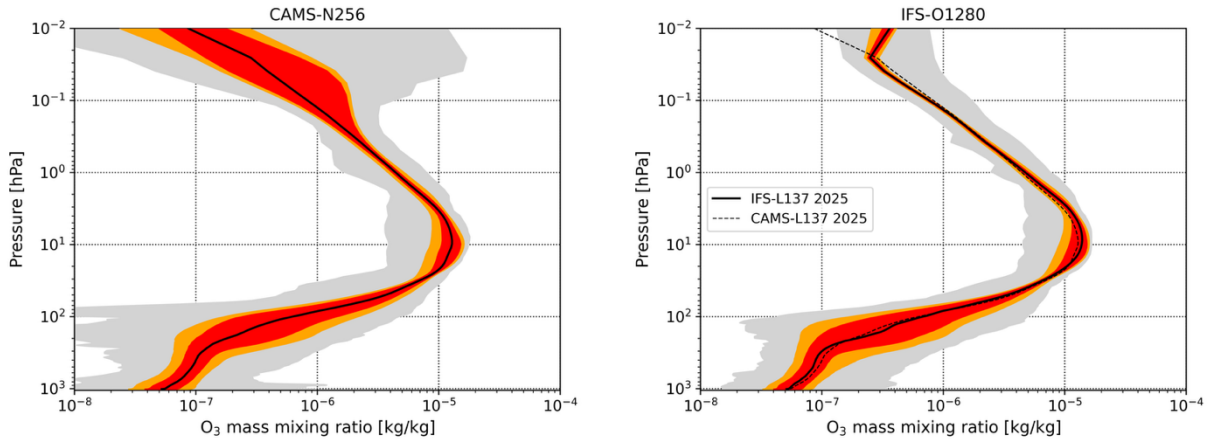


Figure 47. Same as Figure 45 but for ozone profiles with a log pressure scale.

Figure 47 shows that the quantile structure of ozone profiles in the IFS is very different to CAMS above 10 hPa. The IFS does not contain any profiles with ‘double peaks’, or more accurately Tertiary Ozone Maxima (see section 2.1.3), in the mesosphere. This is because

a different ozone field[7] was used for the most recent ozone zonal climatology in the IFS that does not include the TOM. The next update of the climatology should correct this.

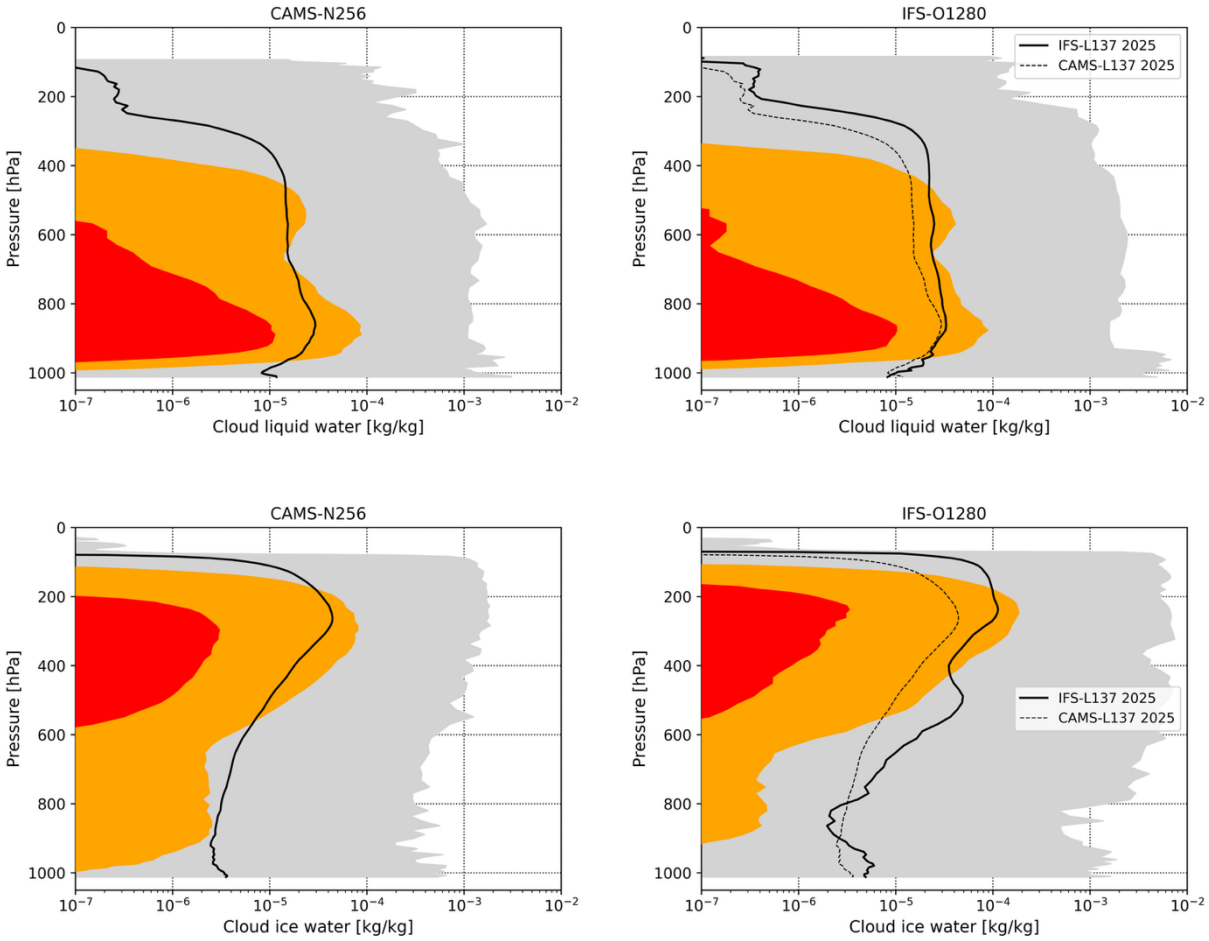


Figure 48. Same as Figure 45 but for cloud liquid water (top) and cloud ice water (bottom). Note that the black line now represents the mean rather than the median.

The IFS produces a higher concentration of liquid and ice clouds at upper levels in the IFS compared to CAMS (Figure 48), which is likely because clouds are strongly influenced by resolution. There are even more enhanced amounts of precipitation (Figure 49). This increase appears to be a trend as the last IFS dataset NWP v4 showed less precipitation than the current CAMS dataset in Figure 12, apart from lower tropospheric rain.

[7] The IFS currently includes gases and aerosols via the HLO scheme, which is based on the CAMS EAC4 reanalysis. The ozone parameter has ID *203* in the MARS archive. For CAMS parameter ID *210203* is advised which includes TOM.

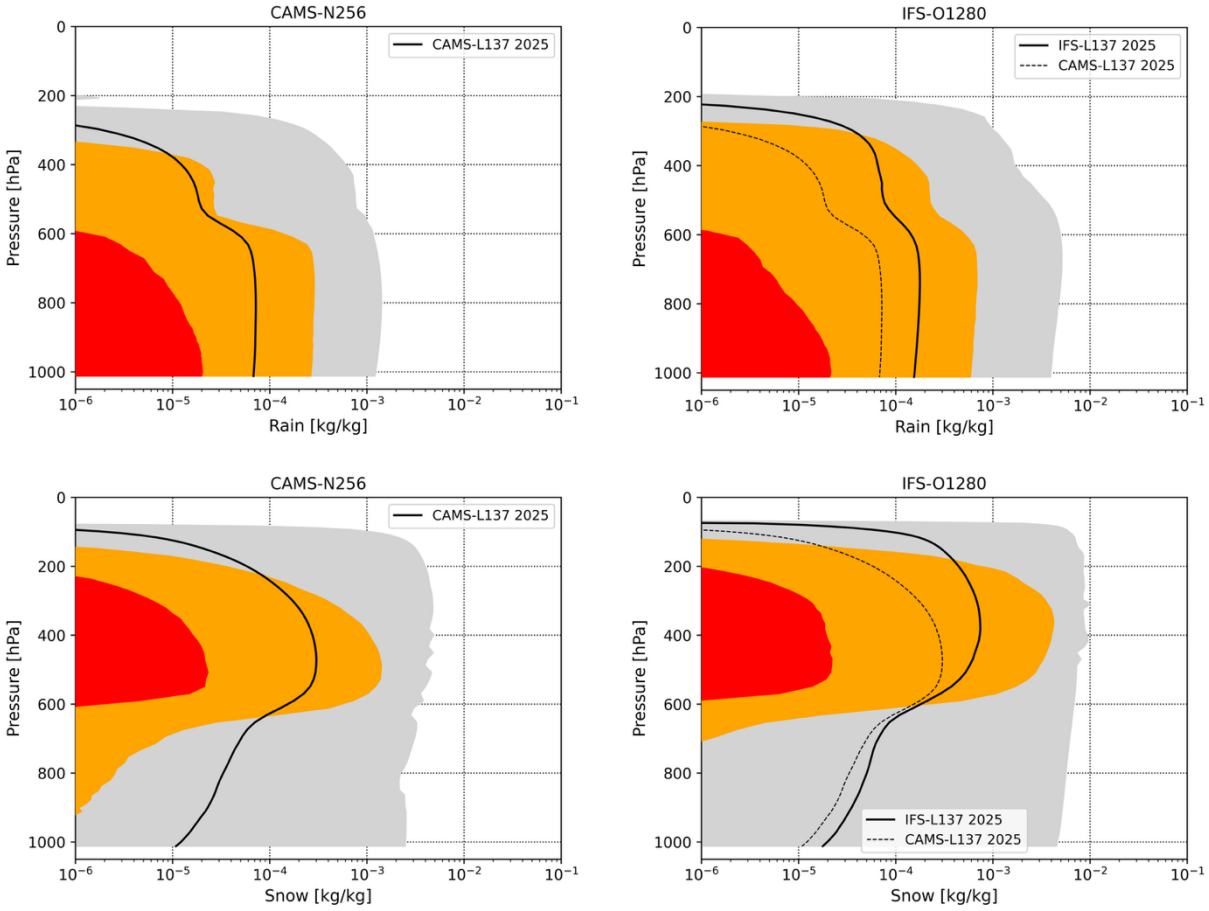


Figure 49. Same as Figure 45 but for rain (top) and snow (bottom). Note that the black line represents the mean.

5 Conclusions

A new diverse profile database (CAMS-137L v1) is compiled from 12 months of global short range forecasts between 1st March 2024 and 20th February 2025 from the ECMWF CAMS model at operational cycle Cy49r1. The database consists of 100,000 profiles divided into 20 subsets of equal size. It combines all previous variables present in previous datasets including meteorological, greenhouse gas, reactive gas and aerosols and brings the available selection up to date. There have been some major changes to the profile structure of some variables since the last datasets were released, particularly in the case of reactive gases and aerosols. Many of the reactive gases have increased in concentration, predominantly in the stratosphere and mesosphere. The vertical structure of aerosol profiles are very different in appearance to the previous MACC profiles released in 2012, and three new species: ammonium, nitrate and secondary organic aerosol have been included. Comparison of NWP profiles from CAMS with the operational IFS reveal minor differences in temperature and humidity and more clouds and precipitation in the latter due to the different horizontal resolution. There is also an absence of ozone profiles with Tertiary Ozone Maxima in the current IFS profiles that comprise nearly 5% of profiles in the CAMS equivalent.

6 Constituents of the database

Table 3. Vertical parameters provided for each profile in the CAMS-L137 v1 database

Atmospheric variables (model levels)	
Variable	**Unit**
Temperature	K
Specific humidity	kg/kg
Ozone mixing ratio	kg/kg
Fractional cloud clover	0-1
Cloud liquid water content	kg/kg
Cloud ice water content	kg/kg
Rain mixing ratio	kg/kg
Snow mixing ratio	kg/kg
Vertical velocity	Pa/s
Carbon dioxide mixing ratio	kg/kg
Methane mixing ratio	kg/kg
Nitrous oxide mixing ratio	kg/kg
Carbon monoxide mixing ratio	kg/kg
Nitrogen dioxide mixing ratio	kg/kg
Sulphur dioxide mixing ratio	kg/kg
Formaldehyde mixing ratio	kg/kg
Sea salt aerosol mixing ratio - (0.03 – 0.5 μm)	kg/kg
Sea salt aerosol mixing ratio - (0.5 – 5 μm)	kg/kg
Sea salt aerosol mixing ratio - (5 – 20 μm)	kg/kg
Desert dust aerosol mixing ratio - (0.03 – 0.55 μm)	kg/kg
Desert dust aerosol mixing ratio - (0.55 – 0.9 μm)	kg/kg
Desert dust aerosol mixing ratio - (0.9 – 20 μm)	kg/kg
Sulphate aerosol mixing ratio	kg/kg
Organic matter mixing ratio – hydrophilic	kg/kg
Organic matter mixing ratio – hydrophobic	kg/kg
Black carbon mixing ratio – hydrophilic	kg/kg
Black carbon mixing ratio – hydrophobic	kg/kg
Ammonium mixing ratio	kg/kg
Nitrate fine mode aerosol mixing ratio	kg/kg
Nitrate course mode aerosol mixing ratio	kg/kg
Biogenic secondary organic mixing ratio	kg/kg
Anthropogenic secondary organic mixing ratio	kg/kg

Each of the 20 subsets of variables is provided in a netcdf4 file where the title species is the ‘dominant’ variable whose variability is explicitly accounted for. All other variables included in the file can be thought of as ‘passive’ as they all correspond to the same grid points and times selected by the dominant variable. The final 500 dominant profiles are from the selection algorithm (second step) and the first 4500 are randomly chosen. Atmospheric and surface parameters included in the database are listed in Table 3 and Table 4, respectively.

Additionally latitude, longitude, year, month, day and forecast lead time are provided. The constituents of the database are made as far as possible similar to the previous NWP v4 database, however, there are some differences, specifically that vegetation and surface precipitation parameters are not available from the CAMS archived variables. It was also decided to restrict subsurface variables to the top layer as this is the most relevant parameter for the atmosphere.

Table 4. Surface parameters provided for each profile in the CAMS-L137 v1 database

Surface variables	
Variable	**Unit**
Surface pressure	Pa
Geopotential	m^2/s^2
Skin temperature	K
2-metre temperature	K
2-metre dew point temperature	K
2-metre specific humidity	kg/kg
10-metre wind speed U component	m/s
10-metre wind speed V component	m/s
Land sea mask	0-1
Sea ice area fraction	0-1
Surface albedo	0-1
Roughness length	m
Snow albedo	0-1
Snow density	kg/m^3
Snow depth	m
Soil temperature in the top layer (0-7 cm)	K
Soil water in the top layer (0-7 cm)	m^3/m^3
Ice temperature in the top layer (0-7 cm)	K

Acknowledgements

This work package was funded through the EUMETSAT NWP SAF programme. The accompanying dataset was generated using Copernicus Atmosphere Monitoring Service Information [2025]. Thank you for advice from Angela Benedetti, Reima Eresmaa, Tim Stockdale and to members of the CAMS team: Ana Agusti-Panereda and Johannes Flemming.

Appendix A: All diverse profile datasets

Table 5. Properties of current (2025) and previous profile datasets released by ECMWF.

Date of release	**2006**	**2012**	**2014**	**2016**	**2025**
Model source	IFS - Cy30r2	MACC	IFS - Cy40r1	CAMS - Cy41r1	CAMS - Cy49r1
Horizontal resolution	T799 (~25 km)	T511 (~40 km)	T1279 (~16 km)	T511 (~40 km)	T511 (~40km)
Number of levels	91	60	137	60	137
Time period	Jul05-Jun06	May10-Mar11	Sep13 - Aug14	Nov15 - Nov16	Mar24 – Feb25
Days selected	1st, 10th, 20th	5th, 15th, 25th	1st, 10th, 20th	9th, 19th, 29th	1st, 10th, 20th
Forecast initial	0000 UTC	0600, 1200 UTC	0000 UTC	0000 UTC	0000 UTC
Forecast steps	36, 42, 48, 54	6, 12	36, 42, 48, 54	12, 18, 24, 30	36, 42, 48, 54
Number of profiles	5000 per subset	4000 per subset	5000 per subset	5000 per subset	5000 per subset
Random selection	-	100% (2nd step)	90% (1st and 2nd steps)	87.5% (1st step), 90% (2nd step)	90% (1st and 2nd steps)
Quality control	-	-	> 25σ rejected	-	-
Selection algorithm	100% (1st and 2nd steps)	100% (1st step)	10% (1st and 2nd steps)	12.5% (1st step) 10 % (2nd step)	10% (1st and 2nd steps)
Extreme profiles	-	40 added at the end	-	-	-
Number of subsets	5	10	5	8	20
Subset variables	Temperature, humidity, O_3, cloud condensate, precipitation	Temperature, humidity, O_3, CO_2, CH_4, sulphate, sea salt, desert dust, organic matter, black carbon	Temperature, humidity, O_3, cloud condensate, precipitation	Temperature, humidity, O_3, CO, NO_2, SO_2, CH_2O, random	Temperature, humidity, O_3, cloud condensate, precipitation, CO_2, CH_4, N_2O, CO, NO_2, SO_2, CH_2O, sea salt, desert dust, sulphate, organic matter, black carbon, ammonium, nitrate, secondary organic matter

Appendix B: Model level to pressure conversion

Model variables are specified on model layers, which are defined by the pressures at the intervals between them known as 'half-levels'. These pressures $p_{l+1/2}$ are given by:

$$p_{l+1/2} = A_{l+1/2} + B_{l+1/2} p_s \quad (3)$$

where p_s is the surface pressure and A and B are coefficients[8] defining the model levels. The pressures that correspond to the values of the model variables are 'full-level' pressures p_l given by:

$$p_l = \frac{1}{2}\left(p_{l-1/2} + p_{l+1/2}\right) \quad (4)$$

Pressure levels for all profiles on the left panel of Figure 50 have been calculated assuming the lowest surface pressure is 1013.25 hPa, so every profiles will have the same vertical coverage, whereas the profiles on the right have been calculated with the real surface pressure associated with each so the vertical extent varies. The right panel has also been converted from mass mixing ratio (mmr) in units of kg/kg to volume mixing ratio (vmr) in units of ppmv, which is calculated by:

$$vmr_{CO_2} = mmr_{CO_2} * \frac{M_{air}}{M_{CO_2}} * 1.0e6 \quad (5)$$

where M_{air} is the molar mass of dry air = 29.964 g/mol and M_{CO_2} is the molar mass of carbon dioxide = 44.009 g/mol.

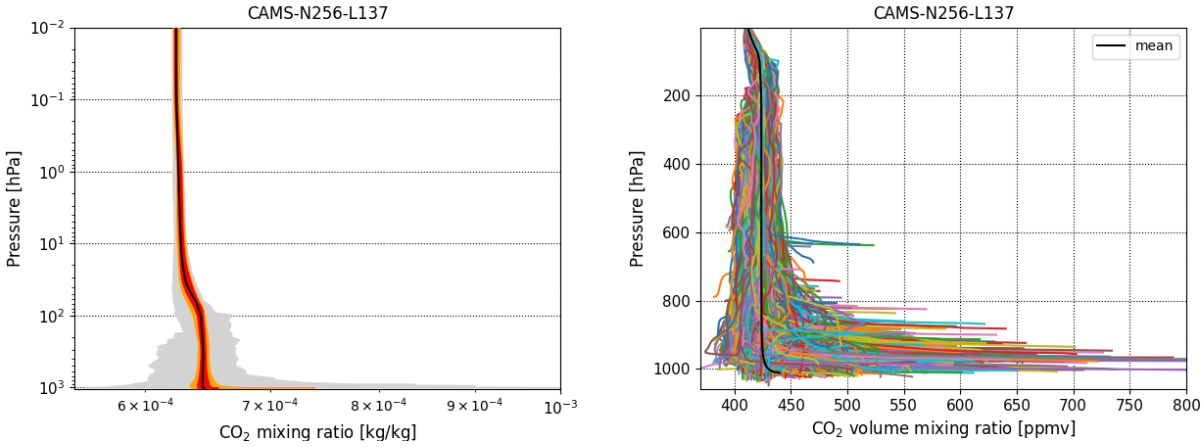


Figure 50. Left panel is the same as the right panel of Figure 14. Right panel shows the same profiles converted to volume mixing ratio and using the actual surface pressures for p_s in the model to pressure conversion in Equation 3.

[8] https://confluence.ecmwf.int/display/UDOC/L137+model+level+definitions

References

Chevalllier, 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, 14p.

Eresmaa, R., and A. McNally, (2016): Diverse profile datasets based on the CAMS atmospheric composition forecasting system. NWP SAF Report, 11 p.

Eresmaa, R., and A. McNally, (2014): Diverse profile datasets from the ECMWF 137-level short-range forecasts. NWP SAF Report No. NWPSAF-EC-TR-017, 12 p.

Eresmaa, R., A. Benedetti, and A. McNally, (2012): Diverse profile database of aerosol and trace gas concentrations from the Monitoring Atmospheric Composition and Climate short-range forecasts. NWP SAF Report No. NWPSAF-EC-TR-015, 12 p.

Eskes, H., Tsikerdekis, A., Ades, M., Alexe, M., Benedictow, A. C., Bennouna, Y., ... & Peuch, V. H. (2024). Evaluation of the Copernicus Atmosphere Monitoring Service Cy48R1 upgrade of June 2023. *Atmospheric Chemistry and Physics*, *24*(16), 9475-9514.

Hartogh, P., Jarchow, C., Sonneman, GR., Grygalashvyly, M. (2004). On the spatiotemporal behaviour of ozone within the upper mesosphere/mesopause region under nearly polar night conditions. *JGR Atmospheres*, 109(18) https://doi.org/10.1029/2004JD004576

Hocking, J. (2023). Radiance Simulator v3.2 User Guide. NWPSAF Report No. NWPSAF-MO-UD-051, 73 p.

Peuch, V. H., Engelen, R., Rixen, M., Dee, D., Flemming, J., Suttie, M., ... & Thépaut, J. N. (2022). The Copernicus atmosphere monitoring service: From research to operations. *Bulletin of the American Meteorological Society*, *103*(12), E2650-E2668.

Schranz, F., Fernandez, S., Kämpfer, N., and Palm, M. (2018). Diurnal variation in middle-atmospheric ozone observed by ground-based microwave radiometry at Ny-Ålesund over 1 year, *Atmos. Chem. Phys.*, 18, 4113–4130, https://doi.org/10.5194/acp-18-4113-2018