CAMS Service Evolution

D1.2 Report on first assimilation of LIDAR/Ceilometer data into the IFS

Due date of deliverable	30. June 2025
Submission date	30 June 2025
File Name	CAMAERA-D1.2-V1.1
Work Package /Task	1/1.2
Organisation Responsible of Deliverable	ECMWF
Author name(s)	Michael Kahnert
Revision number	1.1
Status	
Dissemination Level	PUBLIC

The CAMAERA project (grant agreement No 101134927) is funded by the European Union.
Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Commission. Neither the European Union nor the granting authority can be held responsible for them.

1 Executive Summary

In WP 1 task 1.2 the aim is to develop the capacity to assimilate lidar and ceilometer observations into the global CAMS system, and to evaluate its performance when assimilating only lidar/ceilometer data. As a data source, the European E-profile network of ground-based lidar/ceilometer stations is used. One potential issue is to deal with clouds and other unwanted signals in the observations. Another problem is the non-sphericity of dust aerosols, which can make it challenging to simulate the attenuated backscatter signal in the observation operator. During the first 18 months of the project the assimilation system was successfully set up, and the observation operator, tangent linear, and adjoint codes were extensively tested. An option of replacing spherical with spheroidal model particles for the backscattering properties of mineral dust aerosols was implemented and tested. Adequate pre-processing routines of the lidar/ceilometer data were developed and evaluated, which also contained a method for masking out clouds, precipitation, fog, and noise. A first evaluation of the system, in which only NWP and E-profile data were assimilated, revealed that the vertical correlations assumed in the wavelet representation of the background error covariance matrix can significantly impact the system performance. Replacing the lidar ratio of dust aerosols simulated by spherical particles with that computed by spheroidal model particles can also have a significant effect on the analysis and forecast. All tests show that assimilation of lidar/ceilometer data improves the agreement between model-simulated lidar signals and independent observations. When using spherical dust optics and wavelets with zero vertical correlations, the forecast for AOD and PM2.5 is improved by assimilating lidar/ceilometer data, but the forecasts for PM10 degrades. When using spherical and, even more so, spheroidal dust optics and wavelets with non-zero vertical correlations, the forecasts for AOD and PM2.5 are degraded by the assimilation, while the forecasts for PM10 is improved.

Table of Contents

2 Introduction

2.1 Background

The European Union's flagship Space programme Copernicus provides a key service to the European society, turning investments in space-infrastructure into high-quality information products. The Copernicus Atmosphere Monitoring Service (CAMS, https://atmosphere.copernicus.eu) exploits the information content of Earth-Observation data to monitor the composition of the atmosphere. By combining satellite observations with numerical modelling by means of data assimilation and inversion techniques, CAMS provides in near-real time a wealth of information to answer questions related to air quality, climate change and air pollution and its mitigation, energy, agriculture, etc. CAMS provides both global atmospheric composition products, using the Integrated Forecasting System (IFS) of ECMWF - hereafter denoted the global production system -, and regional European products, provided by an ensemble of eleven regional models - the regional production system.

The CAMS AERosol Advancement (CAMAERA) project will provide strong improvements of the aerosol modelling capabilities of the regional and global systems, on the assimilation of new sources of data, and on a better representation of secondary aerosols and their precursor gases. In this way CAMAERA will enhance the quality of key products of the CAMS service and therefore help CAMS to better respond to user needs such as air pollutant monitoring, along with the fulfilment of sustainable development goals. To achieve this purpose CAMAERA will develop new prototype service elements of CAMS, beyond the current state-of-art. It will do so in very close collaboration with the CAMS service providers, as well as other tier-3 projects. In particular CAMAERA will complement research topics addressed in CAMEO, which focuses on the preparation for novel satellite data, improvements of the data assimilation and inversion capabilities of the CAMS production system, and the provision of uncertainty information of CAMS products.

2.2 Scope of this deliverable

2.2.1 Objectives of this deliverables

This deliverable report describes the ongoing work to set up and test an assimilation system in the IFS for lidar and ceilometer data from the E-profile network. The goal up to month 18 was to make use of and expand the existing IFS infrastructure for assimilating lidar observations, and to evaluate the system performance when assimilating only aerosol observations from lidar and ceilometer stations.

2.2.2 Work performed in this deliverable

The essential steps that led toward this goal were (i) acquisition, conversion, and pre-processing of the E-profile data for ingestion in the IFS; (ii) masking the data for clouds, precipitation, fog, and noise; (iii) parametrisation of molecular absorption at the ceilometer wavelength of 910 nm (at which a large number of E-profile instruments observe); (iv) implementation and testing of the observation operator, tangent linear, and adjoint models at this wavelength; and (v) assimilation and evaluation of lidar and ceilometer observations, excluding other CAMS data (especially AOD).

The work progressed according to the plan in the Description of Action (DoA, WP1 T2).

2.2.3 Deviations and counter measures

No deviations have been encountered.

2.2.4 CAMAERA Project Partners:

HYGEOS	HYGEOS SARL
ECMWF	EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS
Met Norway	METEOROLOGISK INSTITUTT
RC.io	RESEARCHCONCEPTS IO
BSC	BARCELONA SUPERCOMPUTING CENTER-CENTRO NACIONAL DE SUPERCOMPUTACION
KNMI	KONINKLIJK NEDERLANDS METEOROLOGISCH INSTITUUT-KNMi
SMHI	SVERIGES METEOROLOGISKA OCH HYDROLOGISKA INSTITUT
FMI	ILMATIETEEN LAITOS
MF	METEO-FRANCE
TNO	NEDERLANDSE ORGANISATIE VOOR TOEGEPAST NATUURWETENSCHAPPELIJK ONDERZOEK TNO
INERIS	INSTITUT NATIONAL DE L ENVIRONNEMENT INDUSTRIEL ET DES RISQUES - INERIS
IOS-PIB	INSTYTUT OCHRONY SRODOWISKA - PANSTWOWY INSTYTUT BADAWCZY
FZJ	FORSCHUNGSZENTRUM JULICH GMBH
AU	AARHUS UNIVERSITET
ENEA	AGENZIA NAZIONALE PER LE NUOVE TECNOLOGIE, L'ENERGIA E LO SVILUPPO ECONOMICO SOSTENIBILE

3 Pre-processing of E-profile data for use in the CAMS system

The E-profile network comprises several instrument types observing at different wavelengths. The most abundant ones are listed in Table 1.

Table 1: Lidar and ceilometer instruments.

Instrument	Type	manufacturer	Wavelength [nm]	Vertical range [km]	Vertical resolution [m]
CL31	Ceilometer, diode laser	Vaisala	910	≤ 7.6	5-10
CL51	Ceilometer, diode laser	Vaisala	910	≤ 15	10
CHM15k	Ceilometer, Nd:YAG solid state laser	Lufft	1064	7-15	5
Mini-MPL	Elastic lidar	Micro Pulse LiDAR	532	≤ 30	5-75

The observations are provided in netCDF format and can be downloaded from www.ceda.ac.uk. A script has been implemented to automatise navigation of the CEDA database and downloading of data for any prescribed period and selection of stations. Currently, data from a total of 373 stations in the E-profile network are being used for assimilation and evaluation.

Prior to assimilation of the ceilometer data, the following pre-processing steps are performed.

- Masking out unwanted data.
- Reduction of the temporal and vertical resolution.
- Conversion of units to those used by the IFS.
- Reordering of vertical profiles from bottom-up to top-down ordering.
- Pooling of the data and exporting into ODB format.

3.1 Data masking

The masking of unwanted data largely follows recommendations based on prior experiences by Meteo-France. More specifically, we apply a vertical cut-off, a cloud mask, rain mask, fog mask, and noise mask.

Altitude range. Typically, the ceilometers are capable of detecting clouds within a vertical range close to that claimed by the manufacturers (see above). However, the aerosol signal tends to become noisy and unreliable at altitudes significantly lower than that. For this reason, the following reduced altitude ranges are assimilated, depending on the instrument:

CL31: 0 – 3 km

CL51: 0 – 5.3 km

CHM15k: 0 – 7.5 km

miniMPL: 0.25 – 30 km

Cloud mask. The ability of lidar and ceilometer instruments to detect backscattered signals from multiple cloud and/or aerosol layers depends on the source power and the extinction along the optical path. An important pre-processing step involves distinguishing clouds from aerosol layers. Although clouds are typically optically thicker than aerosol plumes, their optical thickness ranges can significantly overlap. Consequently, the strength of the backscattered signal is generally insufficient for reliable discrimination. However, the standard deviation of the attenuated backscatter coefficient has proven to be an effective parameter for differentiating between clouds and aerosols. Here, a measurement β_{att} of the attenuated backscatter coefficient at altitude z_i and time step t_j is regarded as a cloud, if

$$\text{std}[\ \beta_{att}\ (t, z)] > \text{Threshold},$$

where std denotes the standard deviation of all data within a neighbourhood $t_{i-2} \le t \le t_{i+2}$ and $z_{j-3} \le z \le z_{j+3}$. The assumed threshold is 1e-6 m^{-1} sr^{-1} for CHM15k instruments, and 1.5e-6 m^{-1} sr^{-1} for all other instruments.

Precipitation mask. First, the number of cloud layers (up to three) is determined at each time step t_i. If the difference between two consecutive levels with clouds is ≥ 35 m, then the upper level is counted as a new cloud layer. For each cloud with a base altitude $z_c \ge 1$ km, one considers the quantity

$$\overline{\beta}_{att}\ (t_i, z_{min}, z_{max}) = \text{mean}[\beta_{att}\ (t_i, z)]$$

where the mean is taken over the vertical range $z_{min} \le z \le z_{max}$.

It is tested if

$$\overline{\beta}_{att}\ (t_i, z_{min}, z_{max}) > 2.5\text{e-}6\ m^{-1}\ sr^{-1}$$

where the mean is computed over all altitudes below the cloud in the range $z_{min} = z_c - 1$ km and $z_{max} = z_c - 0.5$ km. If the condition is fulfilled, then all data below the cloud base at time step t_i are masked as precipitation.

Fog mask. Observations below 250 m altitude are masked as fog if any of the following two conditions is satisfied:

$$\overline{\beta}_{att}\ (t_i, 0\ m, 250m) > 2.5\text{e-}6\ m^{-1}\ sr^{-1}$$

$$\overline{\beta}_{att}\ (t_i, 0\ m, 250m) > 2\ \overline{\beta}_{att}\ (t_i, 250\ m, 500m).$$

Noise mask. An observation at altitude z_i and time step t_j is masked out as noise, if

$$\frac{\text{mean}[\beta att\ (t,z)]}{\text{std}[\ \beta att\ (t,z)]} < 3,$$

where the mean and standard deviation are computed over a neighbourhood $t_{i-2} \le t \le t_{i+2}$ and $z_{j-2} \le z \le z_{j+2}$. Also, if this neighbourhood contains less than 13 unmasked observations, then the observation at z_i and t_j is removed.

Figure 1 shows an illustration of these masks for the Mini-MPL station near Toulouse on 17 September 2023. The top panel shows the cloud-base heights contained in the E-profile data files, in which up to a maximum of three cloud layers are identified. The cloud mask in the second panel from top shows that there is generally a good correspondence between the cloud mask used here and the cloud identification in the E-profile data. The third panel illustrates several precipitation events identified by the rain mask on that day, whereas no fog

could be detected (fourth panel). Most of the data above the clouds are masked out as too noisy (fifth panel).

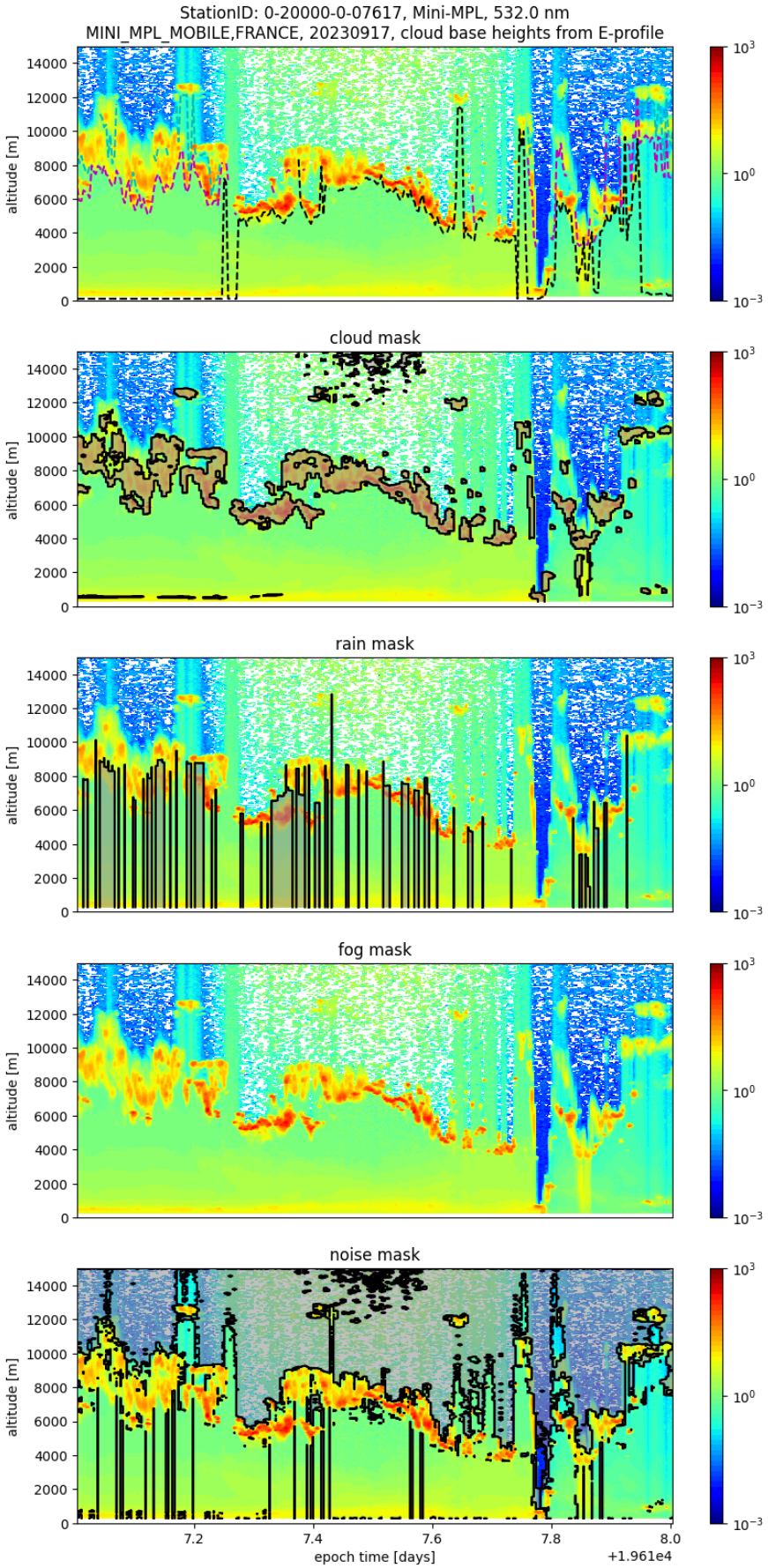

Figure 1. From top to bottom: Cloud base heights from E-profile data, cloud, rain, fog, noise masks.

3.2 Reduction of temporal and vertical resolution

To reduce the data volume prior to assimilation, both averaging and data thinning by selection have been considered. As expected, taking arithmetic mean values over appropriate intervals of time and altitude can reduce noise in the data, while data thinning by selection is somewhat faster. Since the difference in computation time between these two approaches was not found to be substantial, it was decided to reduce the data volume by taking arithmetic mean values over 30-minute time intervals and over appropriate altitude intervals. For the latter the geometric altitudes have been used of the 137 IFS model levels for a US Standard Atmosphere (National Oceanic and Atmospheric Administration, 1976). Averages where only computed if at least 10 % of the data in the averaging interval where not flagged as missing values. Otherwise, the average value was also flagged as missing value. The effect of this reduction approach is illustrated in Figure 2.

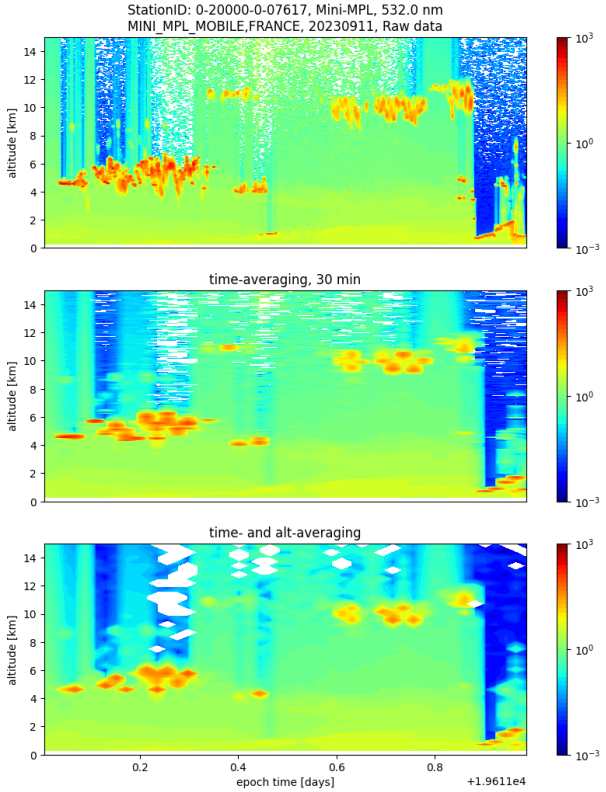

Figure 2. E-profile observations at the original resolution (top), after time averaging (centre), and after time and altitude averaging (bottom).

3.3 Combining the masks and data-reduction operations

Masking and resolution-reduction by averaging are non-commuting operations. Thus, an important question is whether the masking should be done on the high- or the low-resolution data. Figure 3 demonstrates how the chosen order of operations affects the results.

The left row shows that when one first computes the masks and then performs the averaging (2nd row), then one obtains a rather small number of masked data points. If the order is reversed (3rd row) one obtains a significantly more aggressive mask. If the main priority is to exclude all points suspected of cloud contamination, one would favour the latter approach.

However, there can be significant drawbacks with this approach. The example in the right column shows a dust plume originating from a Sahara dust storm. Around mid-day, at higher altitudes, it becomes evident that performing the mask first (centre) excludes a significant number of noisy pixels. By contrast, when averaging is performed first one obtains a more

defensive masking of data regions with noisy data. However, the most important observation in this example is that in the former approach (masking first) the dust plume is not masked out, while in the latter approach (averaging first) the cloud mask becomes so aggressive that it misclassifies the dust plume as a cloud.

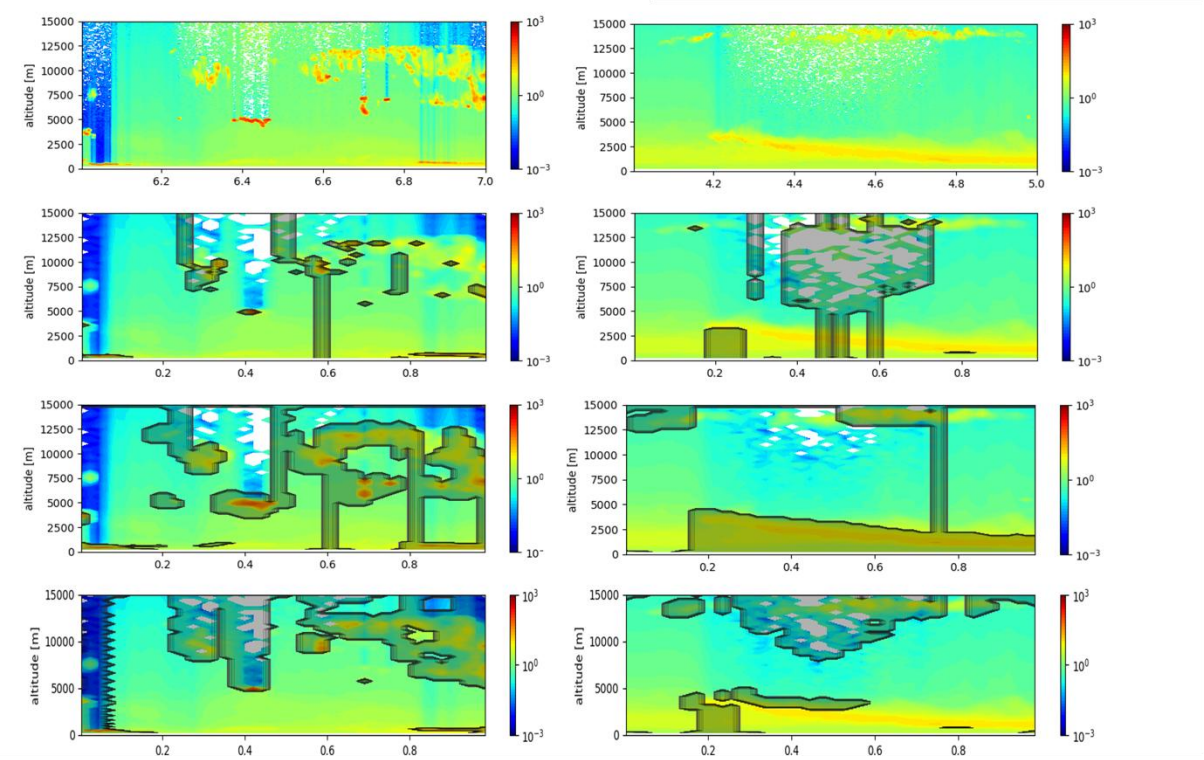

Figure 3. Observations from the MiniMPL lidar near Toulouse on a partially cloudy day (16 September 2023, left) and during a Sahara dust-storm even (4 September 2023, right). The cloud/rain/fog/noise mask is shown for the cases that masking is done prior to averaging (2nd row), after averaging (3rd row), and for the case that cloud/rain/fog masking is done prior, while noise masking is done after averaging (4th row).

A mixed approach is illustrated in the fourth row, which results in a reasonably aggressive cloud mask without the drawback of misclassifying dust plumes. Here, one first computes the cloud, rain, and fog mask. Subsequently, one performs time and altitude averaging. Finally, one computes the noise mask on the low-resolution data. This is the approach that has been adopted for the pre-processing workflow.

3.4 Miscellaneous pre-processing operations

Unit conversions. The E-profile data files provide attenuated backscatter coefficients in units of Mm^{-1} sr^{-1}, and time is given as epoch date since 1 January 1970. Prior to ingestion into the IFS, these need to be converted into m^{-1} sr^{-1} and datetime (YYYYMMDD:HHMMSS), respectively.

Ordering of Profiles. An important pre-processing step is to convert the bottom-up ordering of altitudes native to E-profile into top-down ordering required by the IFS.

Pooling of data. The ingestion of experimental observations in ODB-like ascii format in the IFS requires that large data files are split into pools. Here, the maximum number of lines per

pool was limited to 12 000. The pre-processed data belonging to a given day and analysis time-step where split into pools accordingly and saved as ODB-like ascii files.

4 Preparation of the IFS for assimilation of E-profile observations

Lidar observation operator for $\lambda = 910$ nm. For lidar observations, the IFS (CY49R1) contains observation operators and diagnostic routines for computing the attenuated backscatter coefficient observed from the ground or from space, the unattenuated backscatter coefficient, and the extinction coefficient. These quantities are computed at the wavelengths 1064 nm (typical for a Nd:YAG solid state laser) and at the frequency-doubled and -tripled harmonics 532 nm and 355 nm, respectively.

In the E-profile network instruments measuring at 1064 nm (CHM15k) and 532 nm (miniMPL) are being used (see Table 1). However, a significant number of stations employ CL31 and CL51 instruments, which operate at a wavelength of 910 nm. Thus, a main requirement for assimilating E-profile data was to implement an extra wavelength of 910 nm into the IFS. To this end, both the optical properties of aerosols and the molecular absorption need to be simulated.

For aerosols, one needs to specify the mass absorption cross section, the total mass scattering cross section, and the lidar ratio for each aerosol species (which, for hydrophilic species depend on relative humidity). In the present cycle 49R1, a look-up table of aerosol optical properties at 331 wavelengths in the range 200 nm – 2 mm is used. The aerosol optical properties at 910 nm are computed in the IFS by linear interpolation between the nearest wavelengths in the look-up table.

Molecular extinction is due to absorption and Rayleigh scattering. The latter is parametrised in the IFS for arbitrary wavelengths. (It is the same for all gas species.) At 355 nm, molecular absorption is negligible. At 532 nm the dominant absorber is ozone, at 1064 nm it is molecular oxygen, and at 910 nm it is water vapour. Absorption by other gases can generally be neglected. While parametrisations for absorption by O_3 at 532 nm and O_2 at 1064 nm are implemented in the IFS, a corresponding parametrisation for absorption by H_2O vapour at 910 nm first had to be developed.

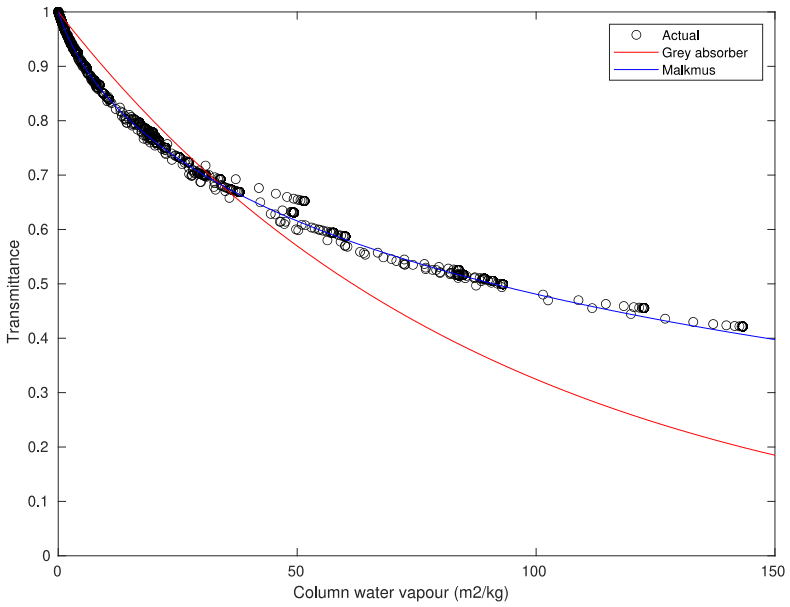

Figure 4. Two-way transmittance versus column water vapour, based on detailed line-by-line computations (symbols), a grey absorber (red), and a fitted Malkmus model (blue line).

To this end, detailed line-by-line radiative transfer computations for several different scenarios of vertical humidity profiles were analysed and taken as a reference. The computational results were taken from the Correlated K-Distribution Model Intercomparison Project (Hogan and Matricardi, 2020) and averaged over a spectral range of ±10 nm, typical for a ceilometer detector. Figure 4 shows the two-way transmittance

$$T(z) = \exp[-2\tau(z)]$$

(where τ denotes the optical depth) as a function of the column water vapour *X(z)*, each taken between the ground and altitude *z*. For a grey absorber (one for which the optical depth has no spectral dependence), one would have

$$\tau(z) = \int_0^z q(z')\rho_{air}(z')k[T(z'), p(z'), q(z')]dz',$$

Where *q* is the specific humidity, ρ_{air} is the air density, and *k* represents the mass absorption cross section of water vapour (in m^2/kg). This model, shown in red, does not fit the line-by-line computations very well. Instead, a Malkmus band model was used (e.g. Geleyn et al., 2005) by parametrising the effective optical depth according to

$$\tau_{eff} = 2\,C\,[\sqrt{1 + K \cdot X/C} - 1].$$

The best fit of the two-way transmittance (blue solid line in Figure 4) was achieved by *C* = 0.034, *K* = 0.013509 m^2/kg.

The new parametrisation for water-vapour absorption at 910 nm has been added to the diagnostic routine and to the lidar observation operator. A corresponding tangent linear (TL) model and adjoint operator (AD) have been implemented, and the consistency of the TL and AD versions of the observation operator has been ensured by performing standardised adjoint tests of the IFS.

For the time being, experimental grib codes have been implemented for saving the attenuated backscatter coefficient and other lidar quantities at 910 nm.

These additions related to the ceilometer wavelength of 910 nm have recently been included in the IFS, CY49R2 and CY50R1.

Aerosol optics model for nonspherical dust particles. In the IFS, aerosol optical properties are generally modelled under the assumption that all aerosol components are externally mixed and can be represented by homogeneous spheres. However, in Cycle 49R1 size- and wavelength-dependent scaling factors have been introduced that account for the asphericity of desert dust aerosols. The scaling factors are computed by use of the online tool MOP-SWAP (Gasteiger and Wiegner, 2018); they are applied to the extinction coefficient, single-scattering albedo, and asymmetry parameter. The computation of the scaling factors is based on an optics model that employs homogeneous prolate spheroids with an aspect ratio of 1.6, which is a median value derived from laboratory studies (Kandler et al., 2007; 2009). More details on the scaling factors and its variation with wavelength and dust-size bin can be found in the documentation of CY49R1 (ECMWF, 2024a).

As of Cycle 49R1, the lidar ratio of dust aerosols was still modelled by use of the homogeneous-sphere model. Here, it was investigated if replacing the lidar ratios based on spheres by lidar ratios based on spheroids had a noticeable effect on the assimilation results. To this end, appropriate size- and wavelength-dependent scaling factors have been applied to the lidar ratios.

Wavelet representation of the background error covariance matrix. In the IFS the B matrix is represented by wavelets. The wavelet files are produced from state files that are either obtained from forecast experiments (known as the NMC method) or from an ensemble data

assimilation (EDA) run. The resulting wavelets express both horizontal and vertical background correlations and variances. When assimilating vertical profiles, the effect of vertical correlations is of particular interest. To evaluate this, two sets of wavelet files were used: first, the standard wavelet files derived using the NMC method, which incorporate vertical correlations; and second, artificial wavelet files in which the vertical correlations were set to zero, thereby suppressing vertical information spreading in the analysis. For brevity, these wavelets will from now on be referred to as *standard* and *artificial* wavelets, respectively.

5 Analysis Experiments

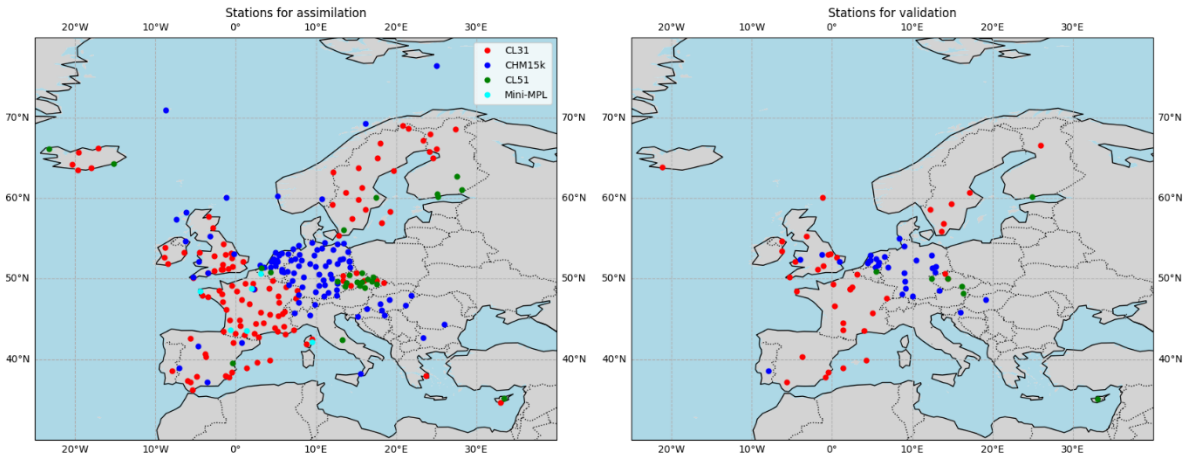

Figure 5. E-profile stations selected for assimilation (left) and for validation (right), showing CL31 (red), CHM15k (blue), CL51 (green), and Mini-MPL instruments (cyan).

A total of 243 stations were assimilated (Figure 5, left), while 74 stations were excluded and retained for validation (Figure 5, right). The colour coding refers to the four different types of instruments as listed in Table 1.[1]

All analysis experiments presented here are performed by use of the IFS-COMPO CY49R1, in which the atmospheric composition model is modelled as part of the Integrated Forecasting System (IFS), which is a global numerical weather prediction (NWP) model. In all experiments, the NWP part of the model is run by assimilating the full set of observations (from ground stations, airplane, ship, satellite, buoys, sondes, etc) with which the operational suite of the IFS is run. However, composition-related observations, in particular satellite observations of aerosol optical depth (AOD), have been excluded from the assimilation. For constraining atmospheric composition only E-profile lidar/ceilometer data have been assimilated. (Assimilation of E-profile data in conjunction with the full CAMS setup, including AOD, is planned to be performed during months 19 - 36 of CAMAERA). More details on the NWP data that are being assimilated can be found in the IFS documentation (ECMWF, 2024b,2024c).

Two periods were considered, 2023/09/01 to 2023/11/30 (for brevity, referred to as SON 2023), and 2024/06/01 to 2024/8/31 (JJA 2024). During the first period, there was a Sahara dust-storm event that reached Northern Spain and South-Western France on September 4 – 5, and the British Isles on September 6. During the second period, smoke from Canadian wildfires was transported across the Atlantic and reached Europe between 17 – 21 August.

[1] Note that the left map in Figure 5 shows three additional Swiss stations, for which data were not yet available at the time of the experiments. These extra stations will be part of the next round of analysis experiments with the full CAMS setup.

Table 2: Analysis experiments for NWP and E-profile assimilation

Experiment ID	Period	Wavelets	Description
ioer	SON 2023	standard	Control run
ioeq	SON 2023	standard	Analysis, spherical dust
ioep	SON 2023	standard	Analysis, spheroidal dust
io8f	JJA 2024	standard	Control run
io98	JJA 2024	standard	Analysis, spherical dust
io9f	JJA 2024	standard	Analysis, spheroidal dust
inx6	SON 2023	artificial	Control run
incg	SON 2023	artificial	Analysis, spherical dust
in85	SON 2023	artificial	Analysis, spheroidal dust

The experiments were run on a linear reduced Gaussian grid with a spectral truncation T511 and with 137 vertical model levels. In most experiments the CY49R1 standard wavelet files for the spectral background error correlations were used. For the autumn period, SON 2023, an extra set of experiments was performed, replacing the standard wavelet files by artificial wavelet files in which all vertical correlations were set to zero. The aim of these experiments was to test to what extent the vertical correlations determine how strongly the ceilometer observations constrain the vertical distribution of the aerosols. Table 2 summarises the two periods, the three types of model-runs, and the wavelets employed in each experiment.

Results in observation space. As a first illustration, Figure 6 shows a Sahara dust-storm episode in early September 2023, visible in the CAMS ensemble forecast (left). The storm peaked around 4 – 5 September and reached the British Isles on 6 September. Vertical profiles of the attenuated backscatter coefficient are shown for 6 September at two non-assimilated stations, namely, Palaiseau near Paris (top) and Cranwell (UK) (bottom). The lines show the observations (black) and control experiment (green), which are the same in the central and rightmost panels. The analysis experiments are shown in blue (spherical dust model) and red (spheroidal dust model), where the results in the central panels are based on using regular wavelet files, while those in the rightmost panels are based on using artificial wavelet files with zero vertical correlations. The experiments in the top panel show a 3-hour forecast after the analysis, the bottom panels present a 15-hour forecast. At these two stations the control experiment (green) overestimates the observations (black). The analysis experiments, especially the one based on using spheroidal dust particles for the optics model (red), significantly improves the control-run results. Comparison of the central and right columns suggests that the use of the artificial wavelet files allows the analysis results to follow the observed vertical profiles somewhat more faithfully, although the effect is, arguably, not very strong.

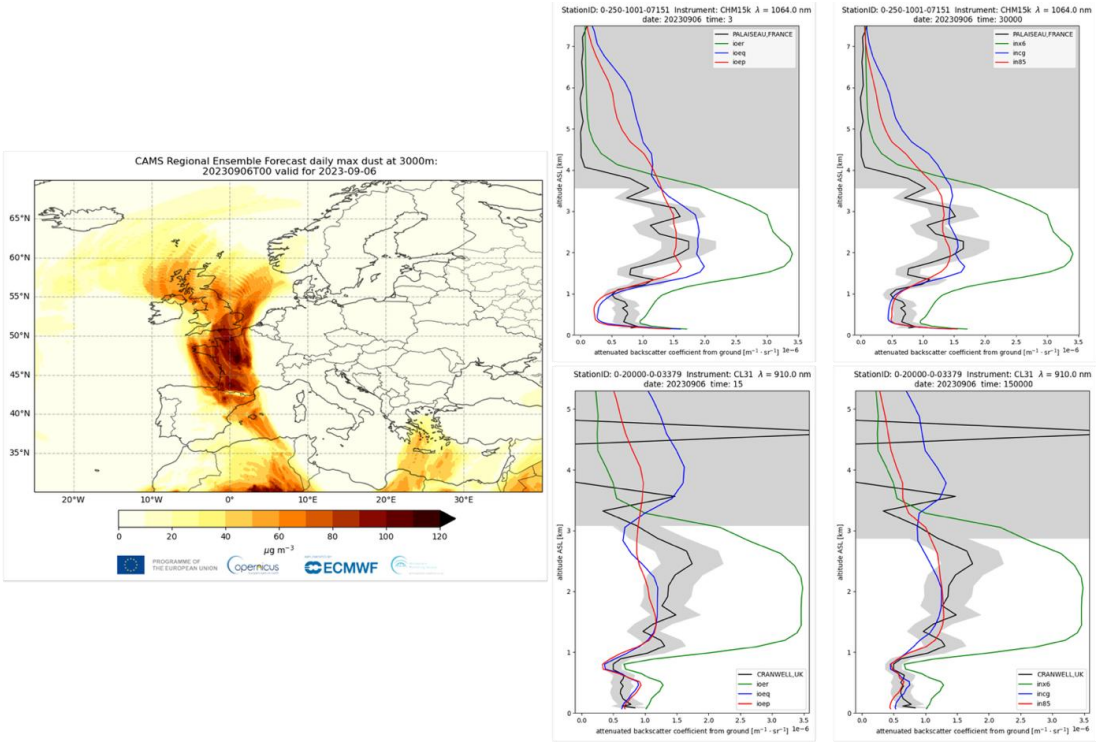

Figure 6. Left: CAMS regional ensemble forecast of daily dust maximum concentration at 3000 m, showing a Sahara dust plume on 6 September 2023 0:00 UTC. Centre and Right: Vertical profiles of the attenuated backscatter coefficient in Palaiseau (top) and Cranwell (bottom), where analysis experiments were carried out with standard (centre) and artificial wavelets (right). The lines show the observations (black), control run (green), and analysis experiments with the spherical (blue) and spheroidal dust model (red), respectively. The grey shading indicates the observation error. Shadings that extend over the entire range of the x-axis indicate observations that have been masked out. Station, instrument, wavelength, and observation time are indicated in the headings.

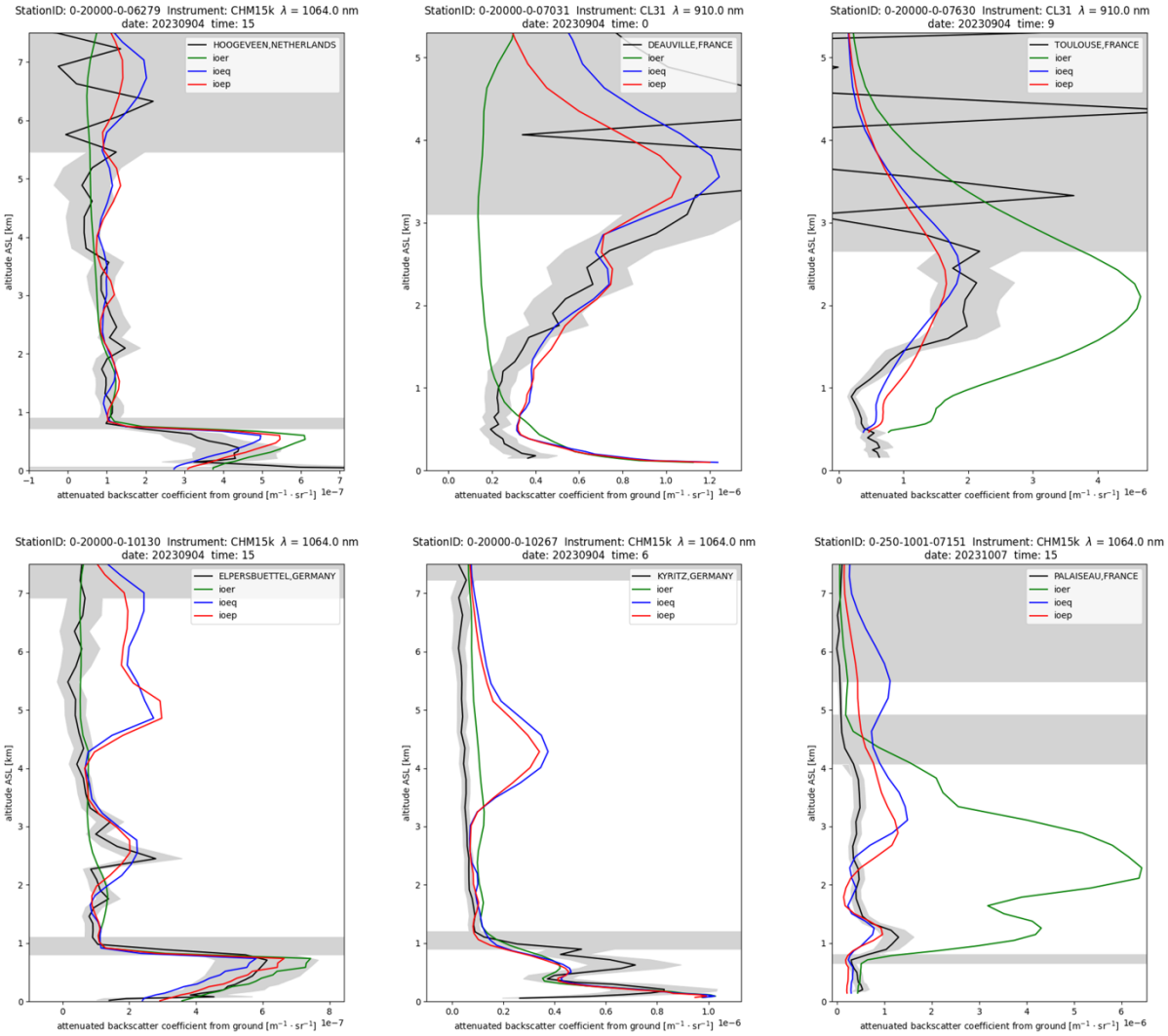

Figure 7. Non-assimilated profiles of the attenuated backscatter coefficient. Station, instrument, wavelength and observation time are indicated in the headings. The line colours are as Figure 6. All results are based on using regular wavelet files.

Figure 7 presents more examples analogous to those in Figure 6 with various forecast lengths. All but the lower right profile show results for 4 September 2023, on which day the Sahara dust storm apparent in Figure 6 (left) peaked. In some cases, the assimilation of ceilometer observations results in moderate (top left) or substantial improvements (top, centre and right, bottom right). The top centre plot shows a case in which the control experiment significantly underestimates the observations. The right plots (top and bottom) show the opposite scenario, in which the control run overestimates the observed attenuated backscatter coefficient. In either case assimilation of ceilometer data brings the forecasts much closer to the observations.

During the dust storm the control experiment underestimates the observations in Deauville, Normandy (top centre), and overestimates the observations in Toulouse (top right). This may indicate that the control experiment does not capture the path and/or timing of the plume sufficiently well. The other locations (top left, bottom left, bottom centre) are located outside the path of the plume.

The bottom left plot shows one example with rather mixed results. While in the lowermost 2 km the analysis experiments (red and blue) only give minor improvements over the control run (green), the aerosol layer around 2.5 km is missed by the control run but shows up in the analysis experiments. However, above 4 km one observes a degradation in the analysis

experiments. Comparison with the corresponding results based on using artificial wavelet files (not shown) show that suppression of vertical correlations slightly, but not entirely, alleviates this problem.

The bottom centre plot shows an example in which the analysis experiments do not differ appreciably from the control run in the lowermost 3 km, but the results between 3 – 6 km are degraded. At about 4.3 km the attenuated backscatter coefficient reaches a peak value of nearly 4e-7 m^{-1} sr^{-1}. No such peak is seen in the observations nor control run. The use of artificial wavelets reduces this peak to about 2.3e-7 m^{-1} sr^{-1} but it does not eliminate the problem.

Comparison of the blue and red lines reveal differences in results obtained by use of dust lidar ratios based on spheroidal (red) and spherical model particles (blue). But these are usually small compared to the differences between each of the analysis experiments and the control run.

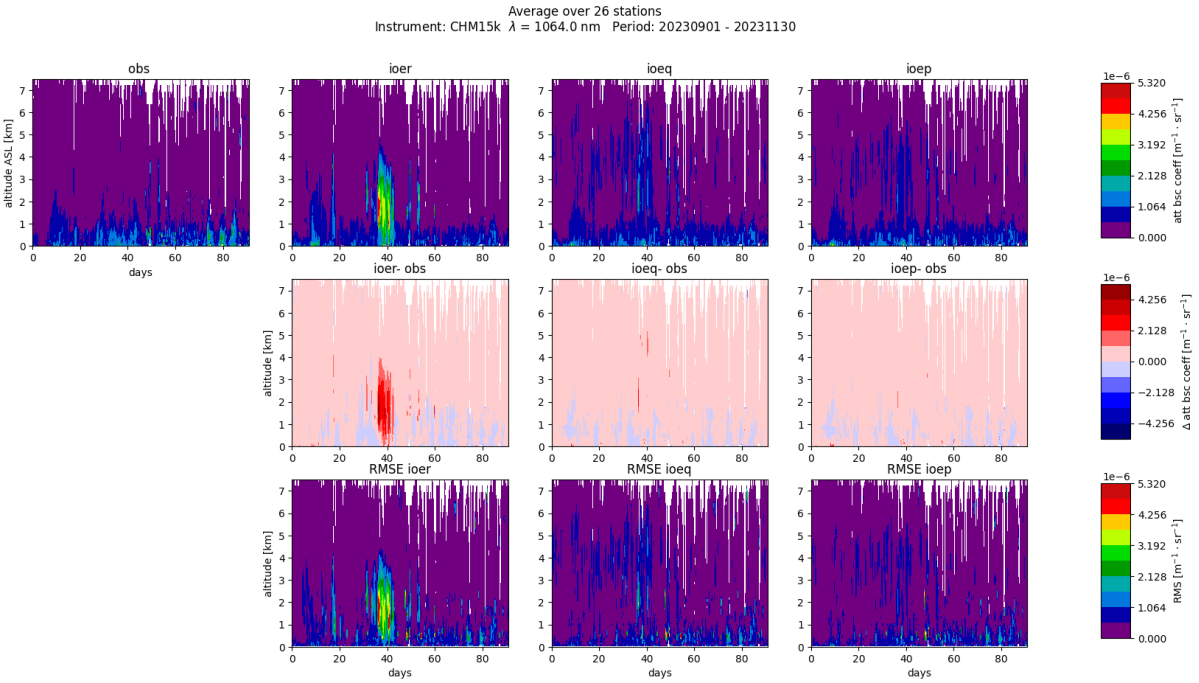

Figure 8. Time-series of vertical profiles of the attenuated backscatter coefficient (top row), difference between forecast and observations (centre row), and root mean square error (bottom row). Results are shown for the observations (1st column), control run (2nd column), and analysis experiments with a spherical (3rd column) and spheroidal dust model (4th column). The figure shows the mean value over 28 CHM15k stations (observing at 1064 nm) that were excluded from the assimilation. The runs were performed with standard wavelet files for the period SON 2023.

Figure 8 shows the mean over 26 CHM15k-stations that were withheld from the assimilation for the whole period September to November 2023 (see the blue dots in Figure 5, right). The top left panel shows the time-altitude dependency of the mean observed attenuated backscatter coefficient at 1064 nm wavelength. The other top-row panels show, from left to right, corresponding results from the control experiment and the assimilation experiments with a spherical and spheroidal dust model, respectively. The second row shows the corresponding biases (forecast minus observation, FC – OBS), and the bottom row displays the root mean square error (RMSE). The analysis experiments employed standard wavelet files. A most prominent feature is a significant overestimation by the control run around 7 – 10 October (days 37 – 40) up to an altitude of 4 km (compare with the bottom right panel in Figure 7). This problem is significantly alleviated in the analysis runs, especially in the one based on using spheroidal dust particles (rightmost column). A less dramatic events is the elevated

backscatter signal observed in the lowermost 1 km around 10 – 19 November (days 71 – 80), which is underestimated by the control run. In this case, neither analysis experiment achieves a significant improvement.

Figure 9 shows analogous results for an average over 40 CL31 and CL51 stations withheld from the assimilation (see the red and green dots in Figure 5, right). At this wavelength (910 nm), the overestimation by the control run of the backscattered signal around 7 – 10 October (days 37 – 40) is still visible. In addition, somewhat weaker overestimation events around days 4, 15, and 51 become apparent. The analysis experiments, especially the one based on the use of the spheroidal dust model, substantially reduce the bias, and they reduce the RMSE.

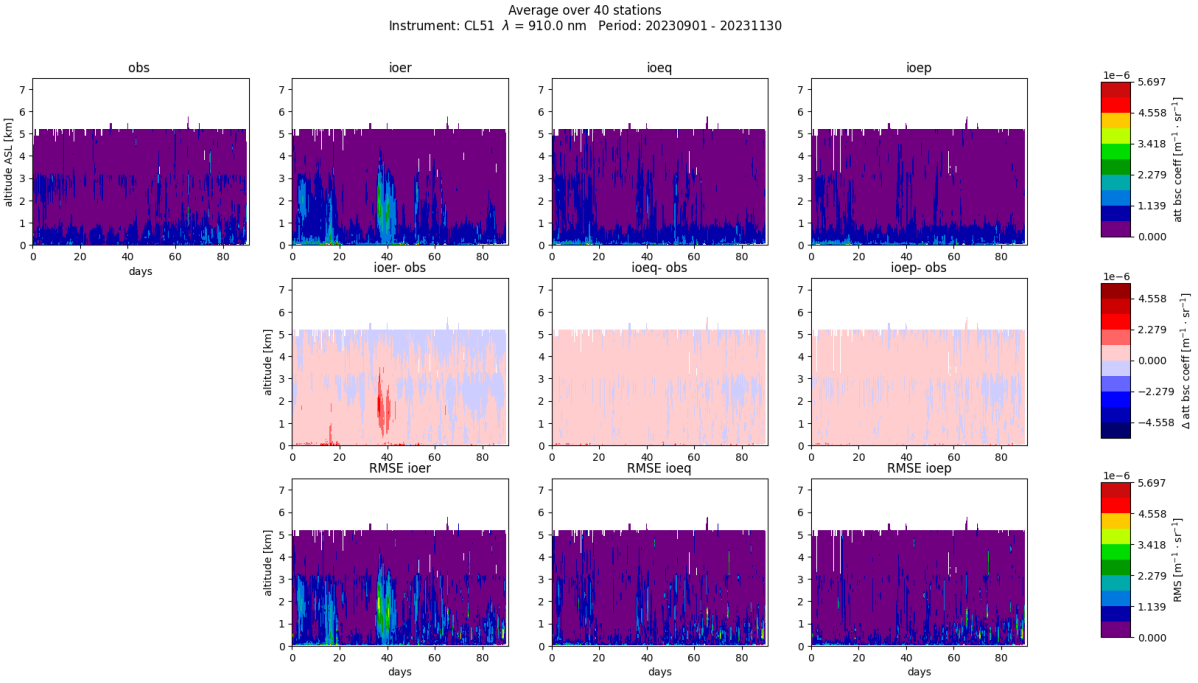

Figure 9. As Figure 8, but averaged over 40 CL31 and CL51 stations, observing at 910 nm wavelength.

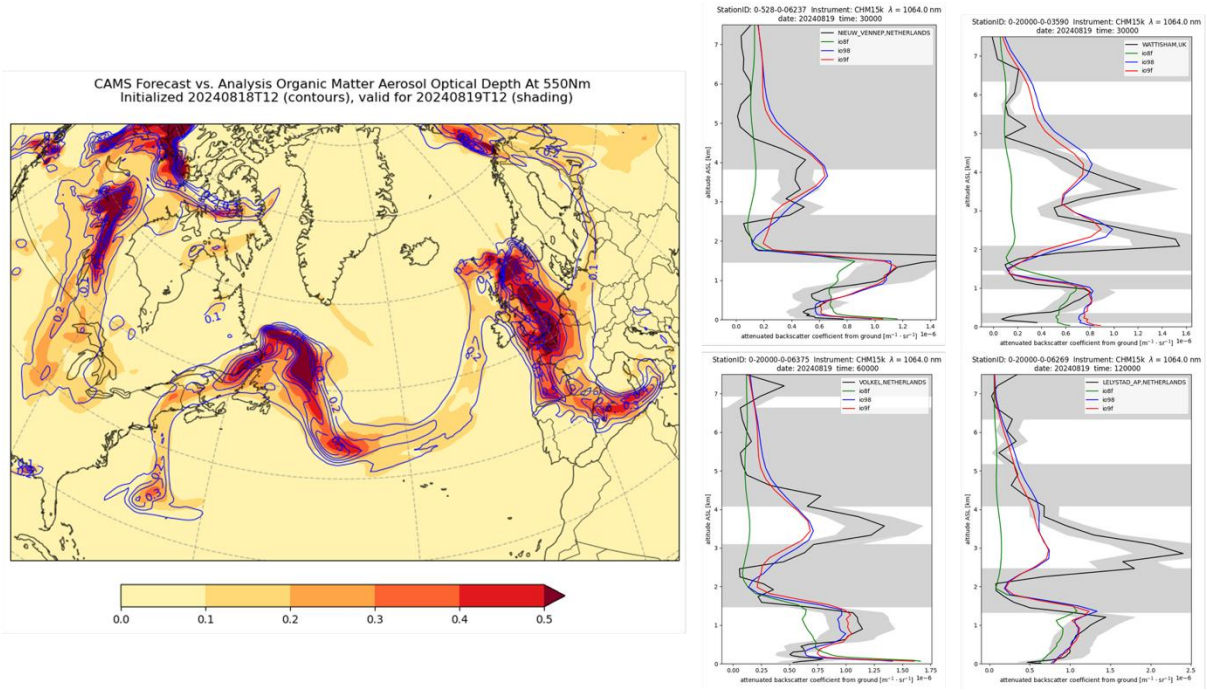

Figure 10. Left: CAMS forecast and analysis of smoke plume, originating from Canadian forest fires, long-range transported to Europe on 19 August 2024 12:00 UTC. Right: Profiles of observed and modelled attenuated backscatter coefficient from the same day (colour coding as in Figure 7).

The second study period is JJA 2024. A notable event during this period was the smoke plume transported to Europe from Canada in the second half of August 2024. The plume is clearly visible in the CAMS forecast of the organic matter aerosol optical depth (Figure 10, left). The control run for this scenario significantly underestimates the observed attenuated backscatter coefficient at affected observation sites at altitudes around 2 – 5 km (green lines in Figure 10, centre and right panels). (Recall that the control run here was performed without AOD assimilation.) By contrast, the analysis experiments (blue and red) substantially improve the simulated backscattered signal aloft.

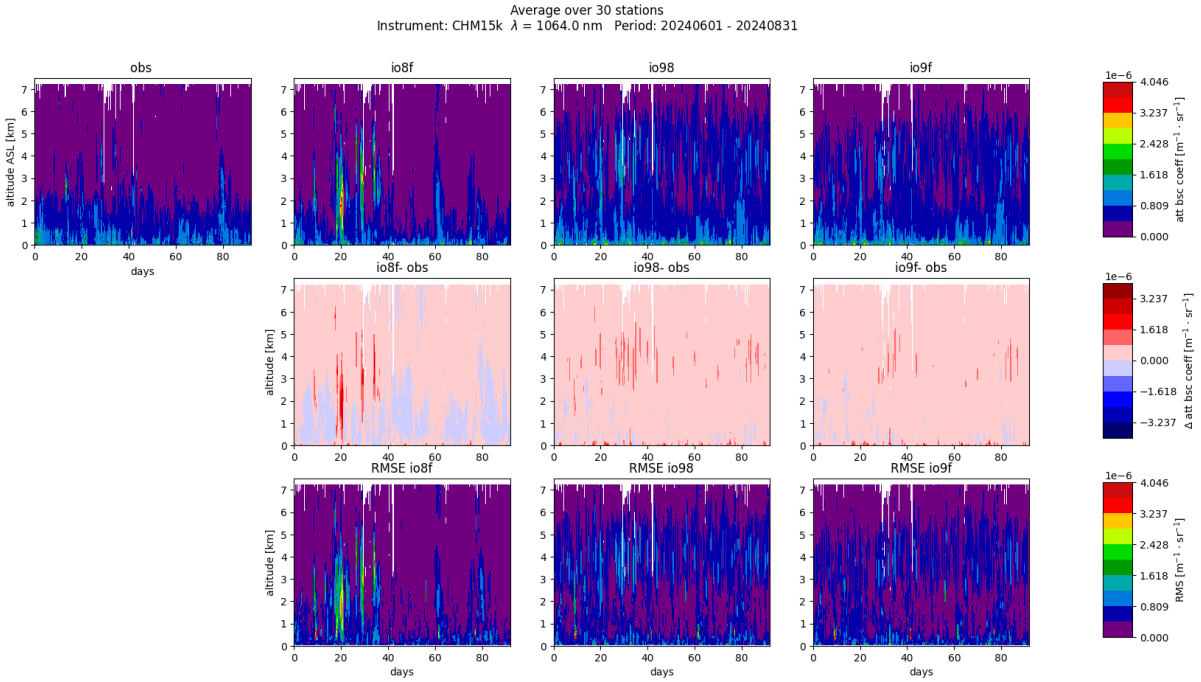

Figure 11. As Figure 8, but showing an average over 30 CHM15k stations for the period JJA 2024.

Figure 11 shows the corresponding time series for this period, where the observations and observation-equivalent model results have been averaged over 30 non-assimilated CHM15k stations. The long-range transported smoke plume around 19 August (day 80) is visible in the observations (top left), but hardly visible in the control run (2nd top). The two analysis experiments result in a better representation of this event (3rd and 4th top). On other occasions, the control experiment significantly overestimates the observations and/or displays elevated RMSE, such as around days 9, 20, 29, 33, 62, and 75. This often occurs at altitudes between 1 – 5 km. Background values in the range 0 – 2 km are often underestimated by the control experiment. These problems are alleviated in the analysis experiment using spherical dust particles (3rd column), and even more so in the one using spheroidal dust particles in the observation operator (4th column). However, the analysis experiments also tend to produce backscattering at altitudes 3 – 6 km that is slightly higher that that observed by the instruments.

Figure 12 shows analogous results averaged over 44 CL31 and CL51 stations, observing at 910 nm wavelength. Several episodes can be identified during which the control experiment overestimates the observations, e.g. around days 6, 20, 25, 29 35, 40, 50, 60, 84, and 90. The background values in the lowermost 2 km are generally well reproduced at 910 nm wavelength. The two analysis experiments, especially the one based on spheroidal dust particles, reduced the bias and RMSE.

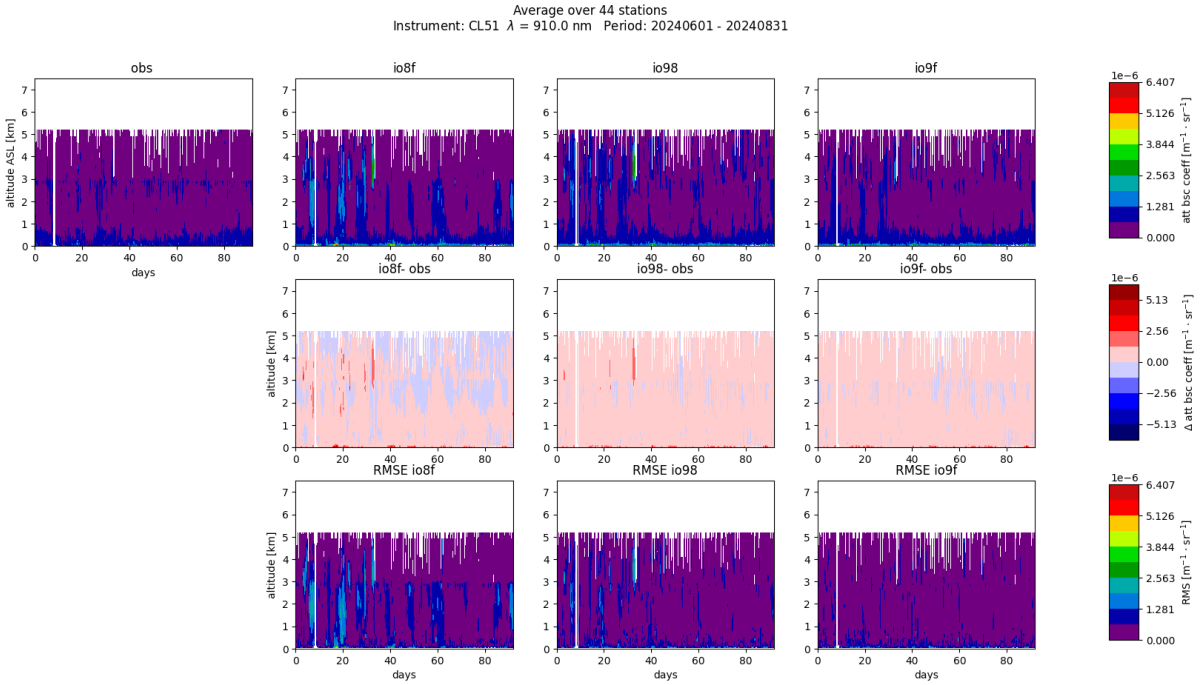

Figure 12. As Figure 11, but averaged over 44 CL31 and CL51 stations, observing at 910 nm wavelength.

Comparison with AERONET level 2 Aerosol Optical Depth (AOD). In this comparison, the focus is on AOD observations at 870 nm wavelength, which is the AERONET channel closest to the assimilated ceilometer measurements at 910 and 1064 nm. Figure 13 presents a comparison of level 2 AERONET observations with the three experiments run for SON 2023 (see Table 2). The time series of 24-hour mean values averaged over 102 European AERONET sites is shown (top left), as well as the bias, FC-OBS (centre left) and the RMSE (bottom left). The scatter plots to the right show all 24-hour mean observations compared to the corresponding model results, the control run on top (ioer), the one based on the spherical dust optics model in the centre (ioeq), and the one based on the spheroidal dust optics model on the bottom (ioep).

Mostly, the control experiment (green lines) slightly underestimates the observations, which is also apparent by comparing the mean values in the top right scatter plot. The analysis experiments overestimate the observations, more so with the spheroidal dust model (grey) than with the spherical one (red). A notable exception is when the Sahara dust storm reaches parts of South-Western Europe on 4 – 5 September 2023. The control run does not capture the elevated AOD observations, while the analysis experiments, especially the one based on spheroidal dust (grey), somewhat reduce the negative bias. The RMSE is consistently lower for the control run than for the analysis experiments. The high positive bias and RMSE observed for the analysis experiments seems to be particularly pronounced during the first month. It is possible that the spin-up time applied to these experiments, which was 10 days, was too short.

For this reason, the second period, JJA 2024, was run with a spin-up time of a full month. The results are presented in Figure 14. Although there is no apparent spin-up effect in this case, there is an even more pronounced degradation of the AOD results in the analysis experiments compared to the control run, which manifests itself by a relatively high bias and RMSE. The analysis based on spheroids yields a slightly higher correlation, but also a higher bias than that based on spheres (compare legends in the central and bottom scatter plots). This can be seen for both SON 2023 and JJA 2024. The differences in the mean values between model

results and observations, as well as the corresponding correlations, are summarised for all experiments in Table 3.

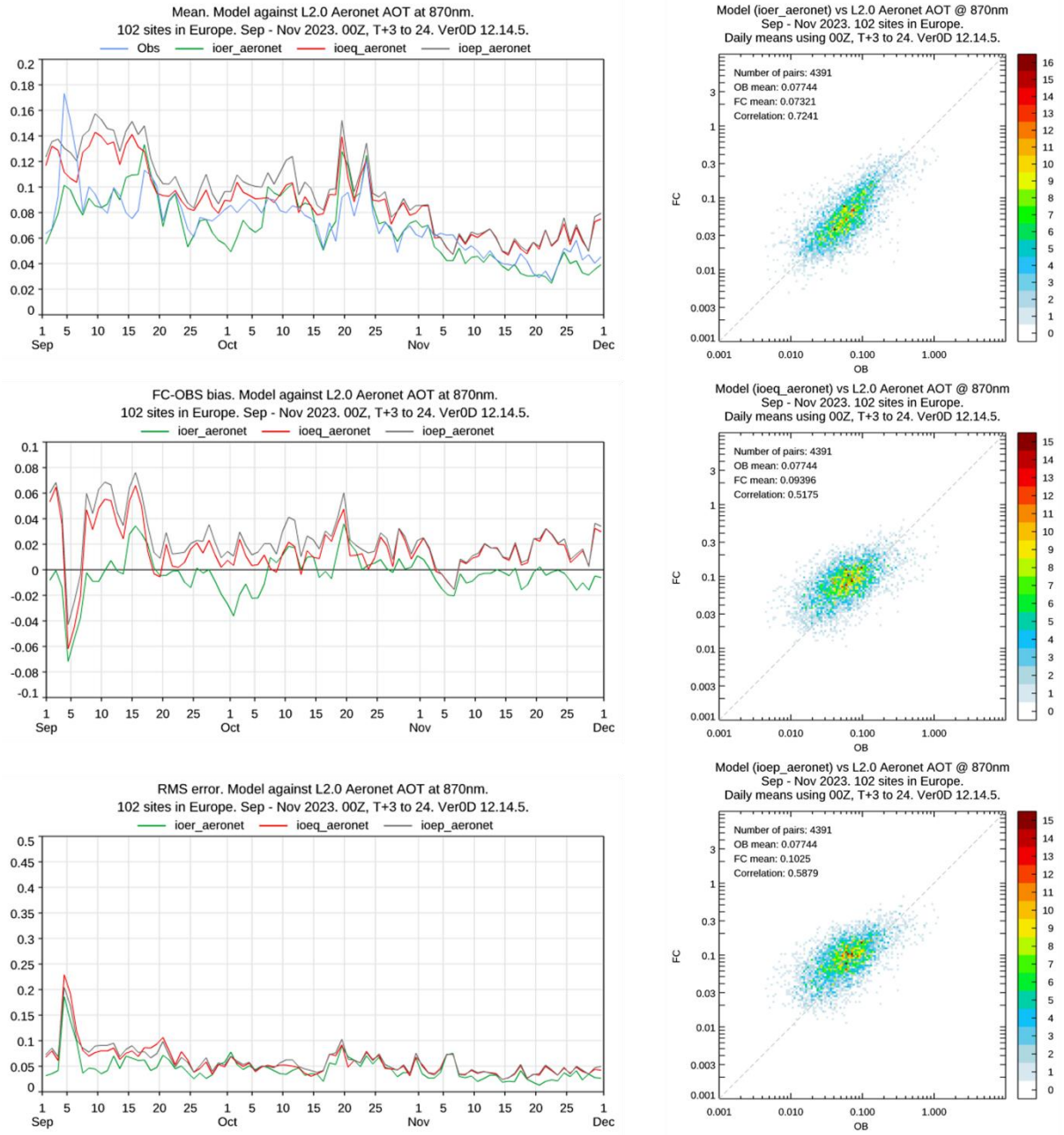

Figure 13. Time series (SON 2023) of AOD at 870 nm (top left), FC-OBS (centre left), RMSE (bottom left), where the lines show a mean value over 102 Aeronet sites in Europe (blue), the corresponding results from the control experiment (green, ioer), and analysis experiments with dust optics based on spheres (red, ioeq) and spheroids (grey, ioep). The right column shows the corresponding scatter plots for these three experiments with the mean values and correlations indicated in the legends. The analysis was performed by using standard wavelet files.

The period SON 2023 was rerun with artificial wavelet files; the results, shown in Figure 15, should be compared to those in Figure 13. The control run for this set of experiments (inx6) yields identical results to that of the first set (ioer), as it should. (Since no CAMS observations were assimilated in the control experiments, they are unaffected by the change of wavelets.) In both analysis experiments (incg and in85) the biases and RMSE are significantly reduced compared to the runs with standard wavelets (ioeq and ioep), and the correlations are

increased (see also Table 3). In the analysis experiment based on the spherical dust optics model (incg), the AOD bias is only 0.001, which is lower by magnitude than that of the control run, -0.004. The RMSE is insignificantly higher in the analysis experiments (0.017 and 0.018, respectively) than in the control experiment (0.016). The correlation is 0.64 and 0.69 in the respective analysis runs, while that of the control run is 0.72. Overall, the use of artificial wavelets substantially improves the analyses (incg and in85) compared to the corresponding analyses with standard wavelets (ioeq and ioep).

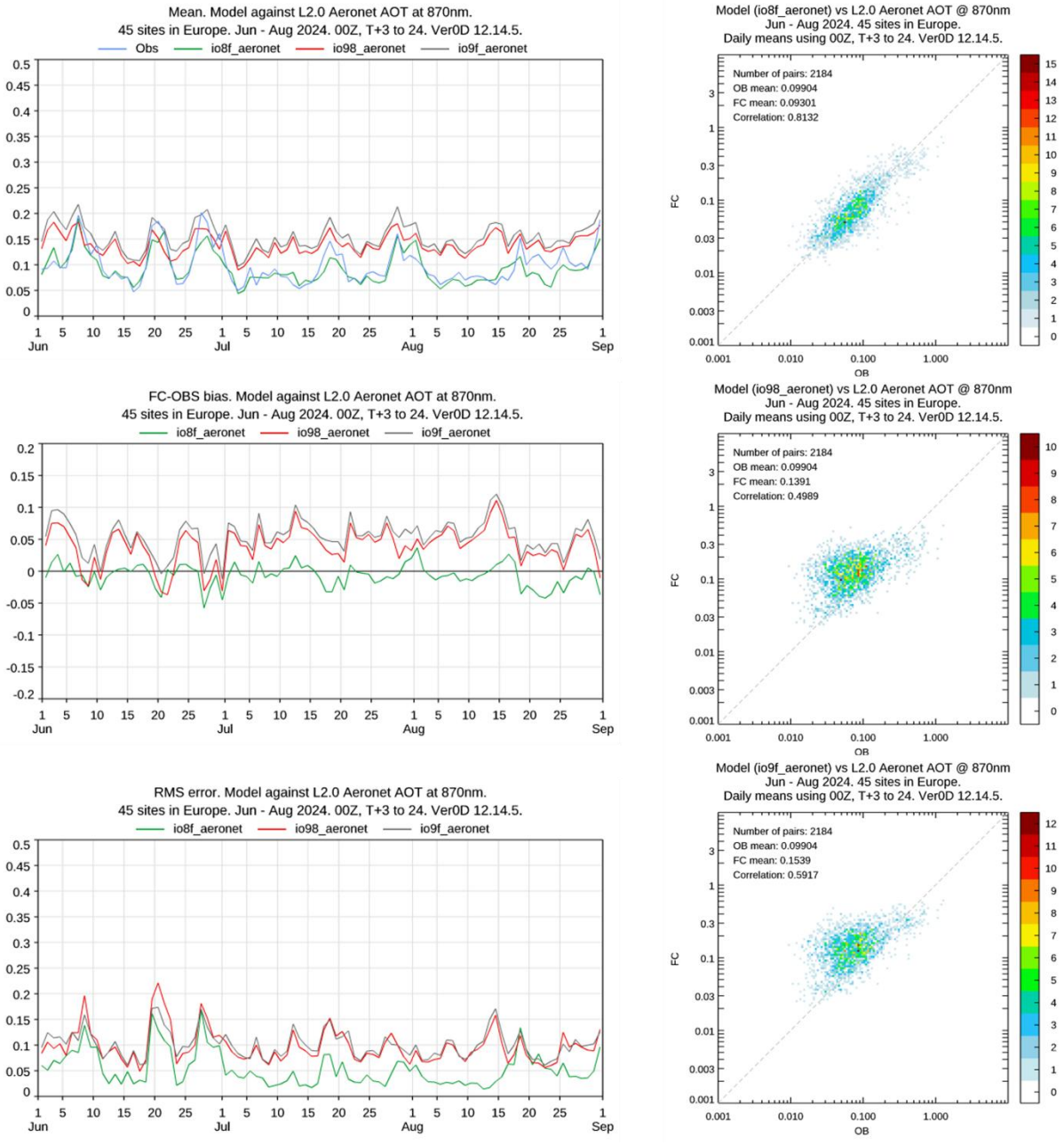

Figure 14. As Figure 13, but for the period JJA 2024, where the mean over 45 stations is shown.

Generally, all analysis experiments result in a positive AOD bias in comparison to the observations, while the control experiment yields, on average, a negative AOD bias. The lowest AOD bias by magnitude is achieved in the analysis experiment in which artificial

wavelet files and a spherical dust optics model has been used. The correlations are, generally, degraded by the analysis; however, the degradation can be reduced significantly using artificial wavelets. The differences between spheres and spheroids are relatively small. The spheres seem to yield slightly lower biases and RMSE than the spheroids, while the spheroids seem to give slightly higher correlations than the spheres.

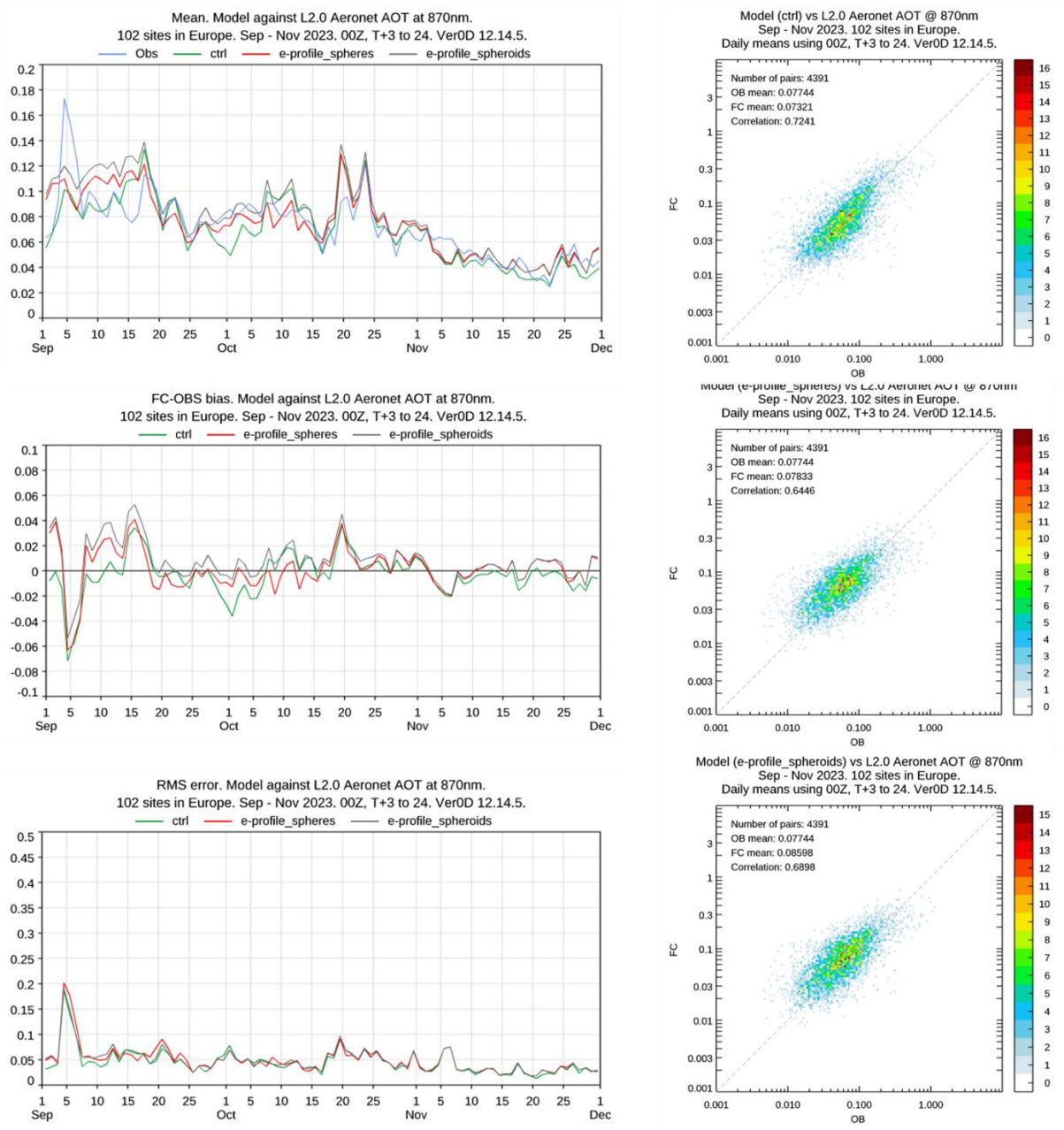

Figure 15. As Figure 13, but showing the analysis results based on using artificial (zero vertical correlation) wavelet files.

Comparison with AERONET level 2 Ångström exponent. The spectral dependence of AOD can often be described by the empirical relation

$$\tau(\lambda) = \beta \cdot \lambda^{-\alpha},$$

where β is a proportionality constant, and α denotes the extinction Ångström exponent. Given observations and model results at two wavelengths λ_1 and λ_2, the Ångström exponent can be obtained by

$$\alpha = -\frac{\ln\left(\frac{\tau(\lambda_1)}{\tau(\lambda_2)}\right)}{\ln\left(\frac{\lambda_1}{\lambda_2}\right)}$$

AERONET observations are available at λ_1=440 nm and λ_2=870 nm, while model results have been archived at λ_1=440 nm and λ_2=865 nm, which are close enough for comparison.

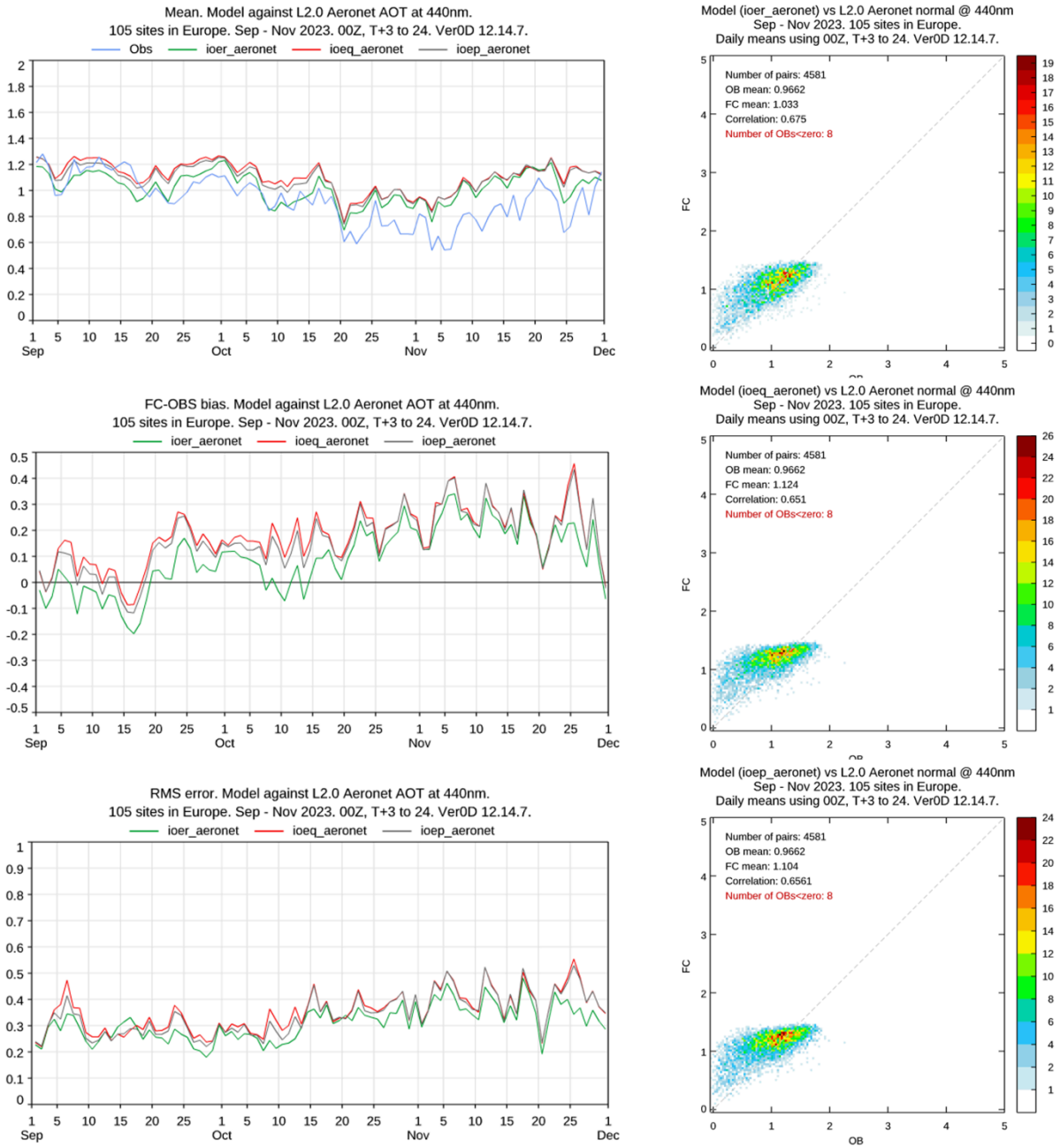

Figure 16. As Figure 13, but for the Ångström exponent averaged over 105 stations.

The Ångström exponent gives some indication on the effective diameter of the aerosols. In the optical part of the spectrum, $\alpha \gtrsim 1$ typically indicates the dominance of fine particles, while $\alpha \lesssim 1$ is usually an indication for the presence of coarse particles. Thus, comparing observed and modelled Ångström exponents can help diagnosing problems in the model related to an incorrect partitioning between fine and coarse particles.

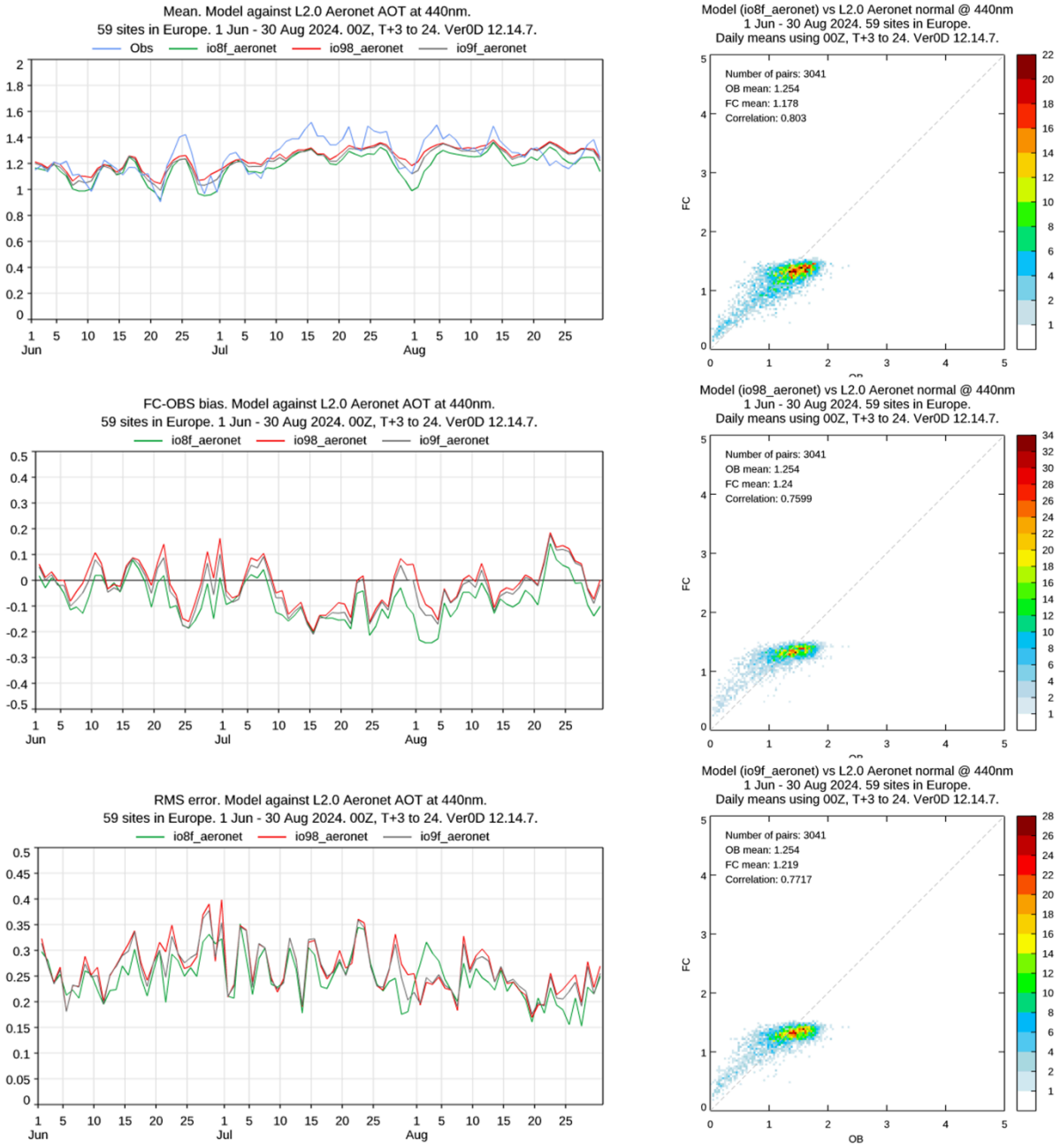

Figure 17. As Figure 16, but for the period JJA 2024, where the mean over 59 stations is shown.

During SON 2023 (Figure 16) the observed Ångström exponent is largely overestimated, more so by the analysis experiments than by the control run. Also, the RMSE is generally lowest for the control experiment. During JJA 2024 (Figure 17) the situation is reversed. The observed Ångström exponent is often underestimated by the experiments, and the control run has the highest negative bias. Thus, the analysis experiments result in a bias reduction during the summer period, and in a degradation during autumn. The positive bias during autumn may

indicate that the model underestimates the coarse-to-fine ratio in the atmospheric column, while the negative summer bias indicates an overestimation of the coarse-to-fine ratio.

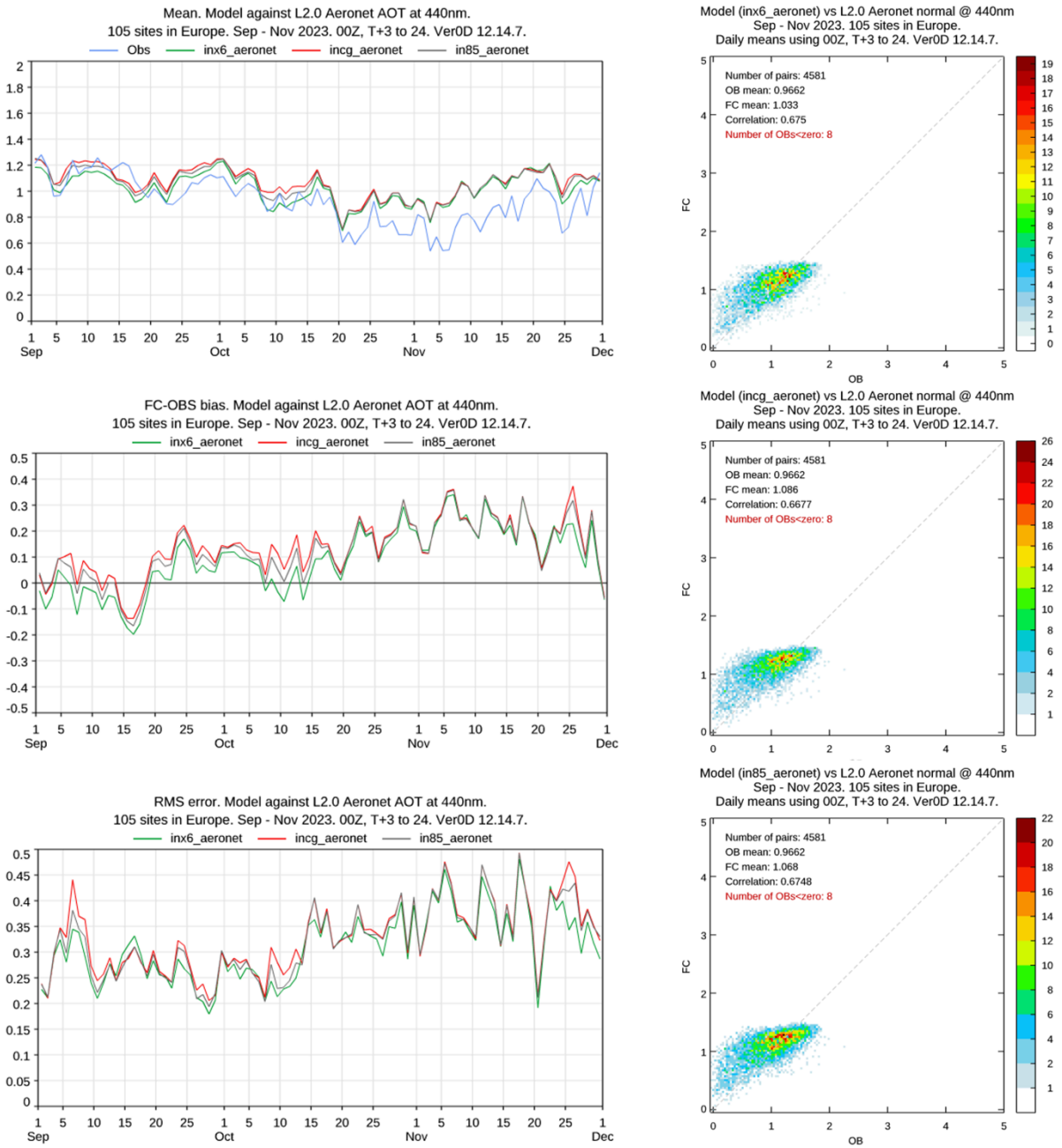

Figure 18. As Figure 16, but showing the analysis results based on using artificial (zero vertical correlation) wavelet files.

The autumn results for the Ångström exponent obtained with artificial wavelets are presented in Figure 18; they should be compared to the corresponding results in Figure 16 obtained by using standard wavelets. The artificial wavelets result in a significant reduction in the bias and RMSE in the analysis experiments (red and grey lines), but the control run still outperforms both analysis experiments. The time-averaged bias, RMSE, and correlation are summarised in Table 3.

Comparison with AirBase PM2.5 and PM10 ground observations. The effect on PM ground concentrations of assimilating ceilometer data is rather diverse. Results for the autumn

period in 2023 are shown in Figure 19 and Figure 20 for PM2.5 and PM10, respectively. In Figure 19 one can identify several episodes of high PM2.5 ground concentrations, e.g., around 10 and 26 September and 11 October. The experiments typically overestimate the intensity of these events. On 10 September, the control run lies closer to the observations, while on 25 September and 11 October the analysis experiments give a better agreement with the measurements. Comparing the histograms reveals that both the control and the analysis experiments yield mean values that are higher than that of the observations, where the mean of the spheroid-based analysis experiment (ioep) is highest, followed by the sphere-based one (ioeq) and the control experiments (ioer). Also, the analysis experiments display a somewhat lower correlation than the control run. However, the degradation is less pronounced than that observed for AOD. The averaged RMSE, given in Table 3, obtained with the sphere-based dust-optics model (ioeq) is lower than that in the control experiment (ioer), while that obtained with the spheroid-based dust-optics model (ioep) is higher.

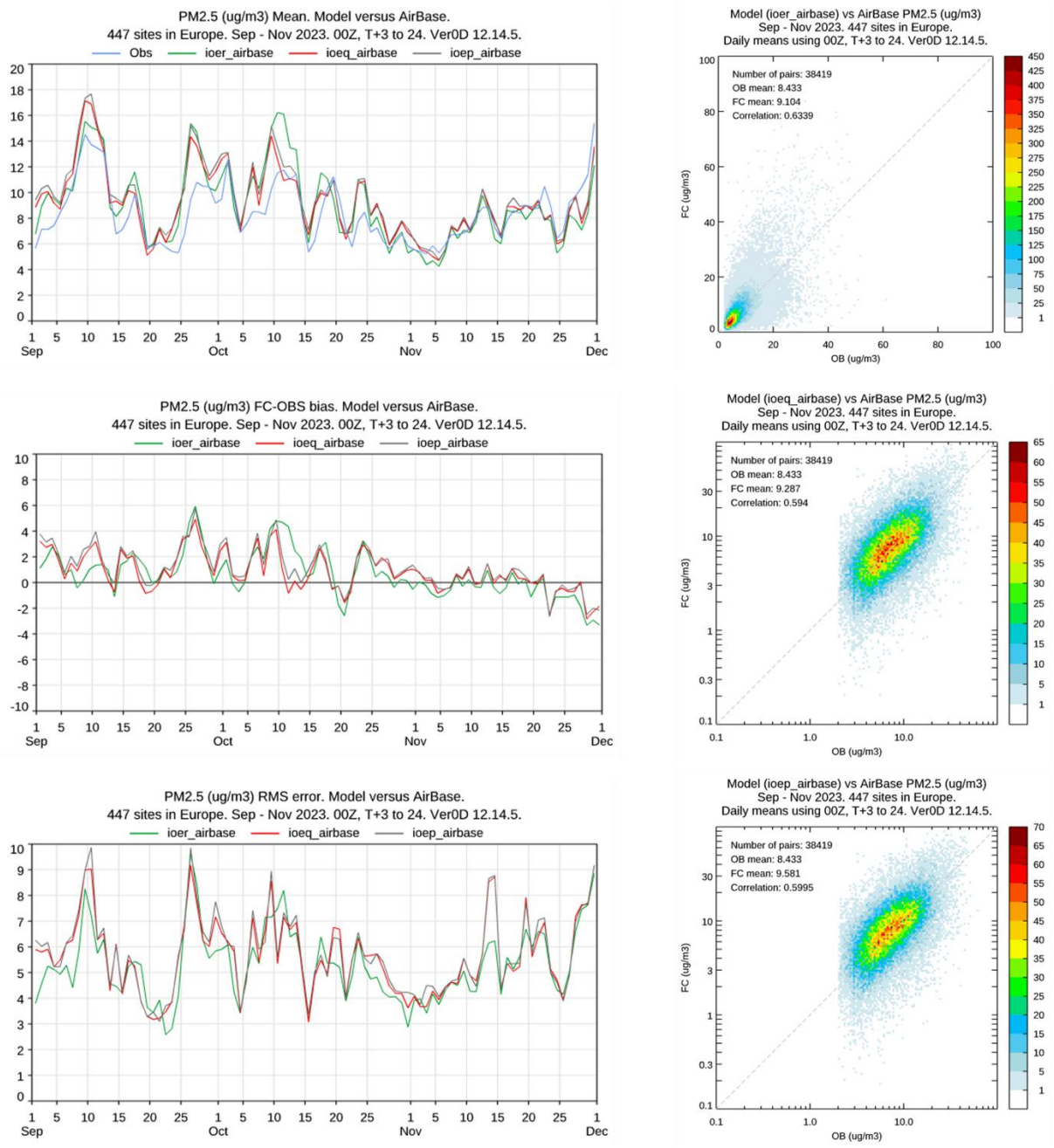

Figure 19. As Figure 13, but for PM2.5 ground concentrations, averaged over 447 AirBase sites in Europe.

Note that the dust-storm event on 4 September, which is clearly visible in the AOD observations (cf. Figure 13), barely manifests itself in the PM2.5 ground observations. This is because the long-range transported dust is mainly found at higher altitudes (depending on location between 1.5 – 5 km, see Figure 6, and Figure 7 top centre and top right).

Figure 20 shows that the model runs mostly underestimate the observed PM10 concentrations. In many instances, the analysis experiments do not deviate significantly from the control run, especially in the latter half of the season. Notable exceptions are the elevated PM10 concentrations observed around 17 September and October 11, during which the control run overestimates, while the analysis experiments tend to underestimate the observations, ioep (spheroids) less so than ioeq (spheres). Comparing the histograms and the summary in Table 3, one finds that the mean bias of the analysis run, ioep (spheroids) is slightly degraded compared the control run, ioer, while the RMSE is improved by the analysis. For PM10, the spheroid model outperforms the corresponding analysis with the sphere model (ioeq).

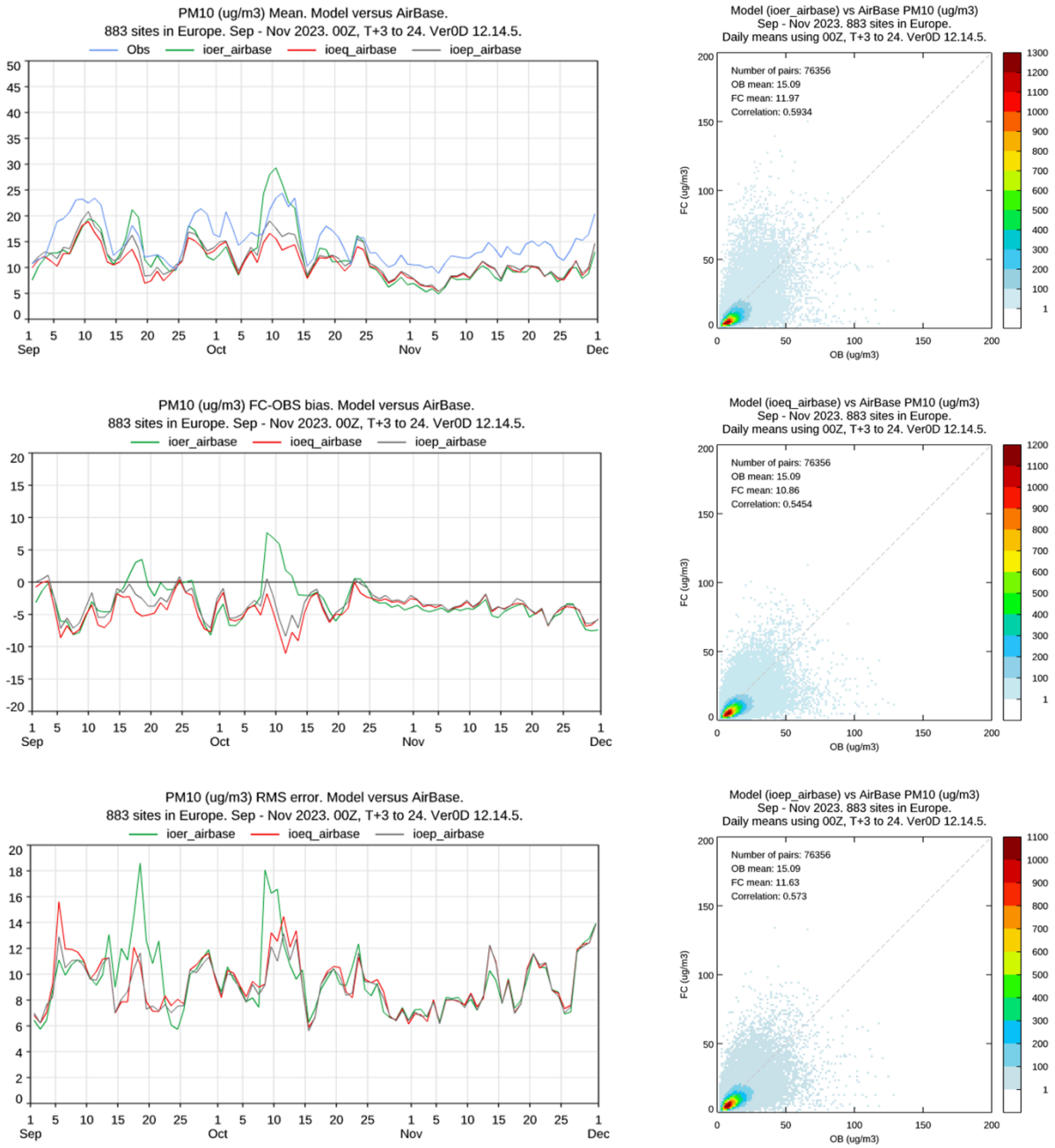

Figure 20. As Figure 19, but for PM10, averaged over 883 AirBase sites.

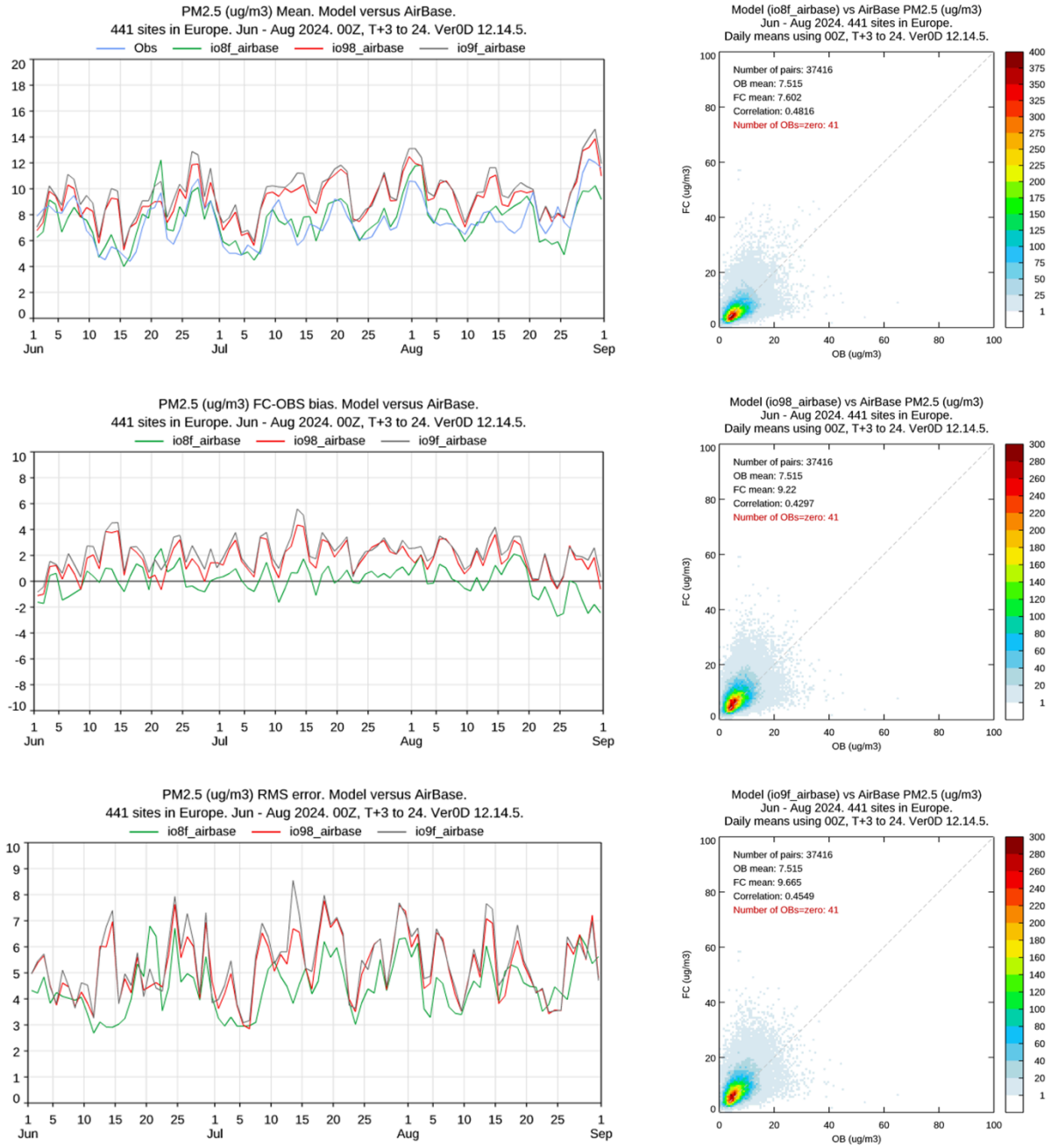

Figure 21. As Figure 19, but for JJA 2024.

The PM2.5 results for JJA 2024 are shown in and Figure 21. The analysis experiments yield a significant degradation in terms of biases, RMSE, and correlations – see also Table 3. By contrast, the negative PM10 biases and the RMSE (Figure 22Figure 21) are significantly reduced in the analyses compared to the control run. The spheroid model outperforms the sphere model in terms of bias, RMSE, and correlation. Overall, the experiments for JJA 2024 show similar trends as those for SON 2023, i.e., the analyses yield higher AOD biases and

RMSE than the control runs, higher PM2.5 biases and mostly higher RMSE, and mostly lower PM10 RMSE. For AOD and PM2.5 the spheroids score worse than the spheres, but for PM10 the spheroids score better than the spheres.

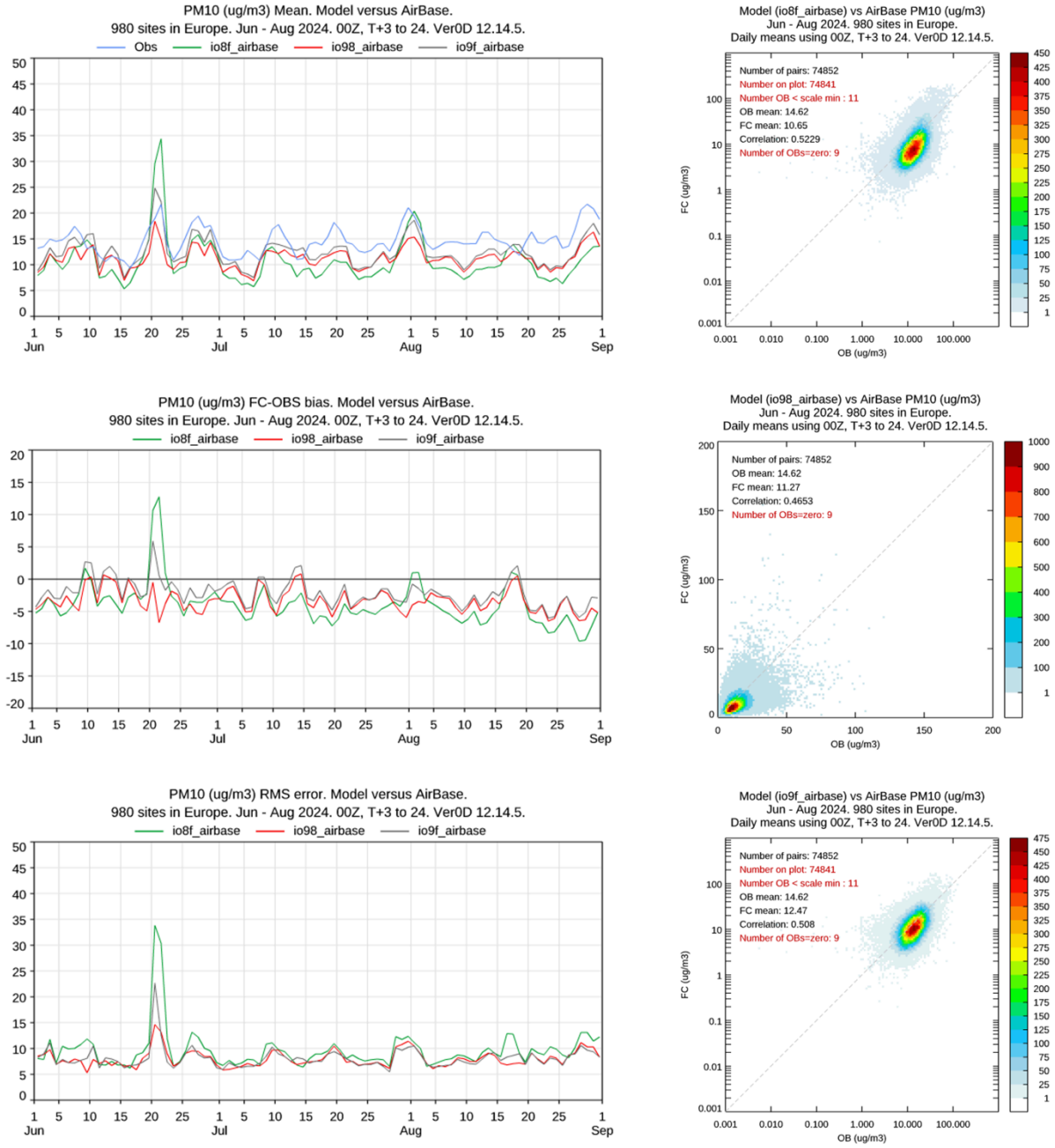

Figure 22. As Figure 21, but for PM10, averaged over 980 AirBase sites.

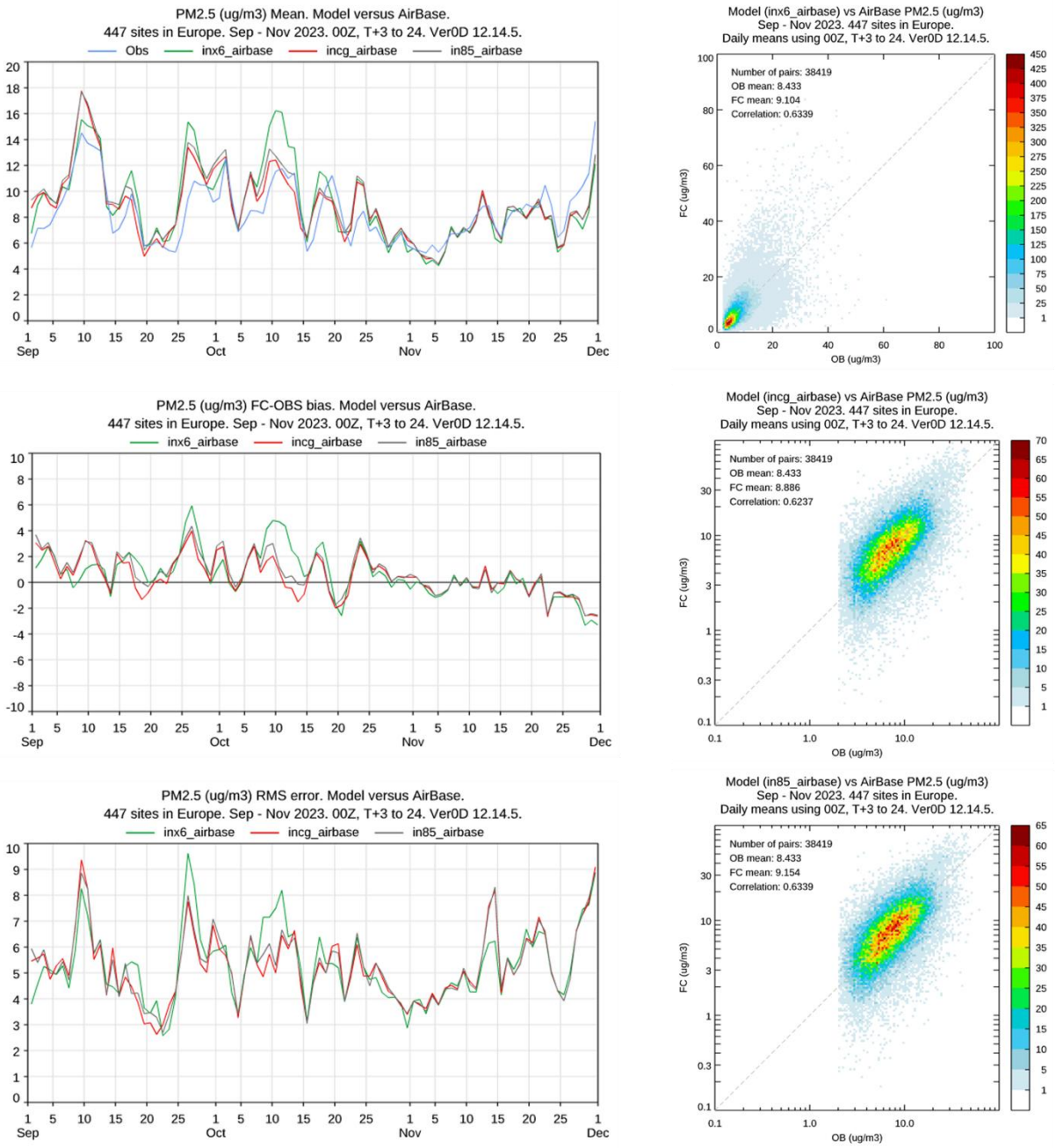

Figure 23. As Figure 19, but showing the analysis results based on using artificial (zero vertical correlation) wavelet files.

Figure 23 shows results for PM2.5 in SON 2023. In Figure 19 it was evident that the control run yields, on average, the lowest bias and highest correlation, while the use of the spheroid model for dust optics significantly degrades the bias. By contrast, Figure 23 demonstrates that the same analysis performed with zero vertical correlations in the wavelets can significantly reduce the bias and RMSE and increase the correlation of the analysis experiments. Specifically, the model based on spherical dust optics clearly outperforms the control experiment in terms of bias and RMSE, while the differences in correlations are insignificantly smaller – see also Table 3.

Figure 24 shows corresponding results for PM10 concentrations. Compared to Figure 20 the use of artificial wavelets with zero vertical correlation slightly degrades the bias and RMSE in the analyses, while the correlation is slightly increased (see also Table 3). Both for the standard and the artificial wavelets, the spheroid analysis yields smaller negative biases than

the spheres, but the negative biases of the spheroid analysis are higher than those of the control run. However, the spheroid analysis still outperforms the control run in terms of RMSE, even though the use of artificial wavelets results in some degradation of the RMSE compared to the experiments with the standard wavelets. In summary, the use of artificial wavelets in conjunction with the spherical dust-optics model produces an analysis that generally improves the control-run results for AOD and PM2.5, and it degrades the control-run results for PM10. The use of artificial wavelets in conjunction with the spheroid dust-optics model improves the control-run RMSE of PM2.5 and PM10, but it degrades the bias of PM2.5 and PM10, and it degrades both bias and RMSE of AOD.

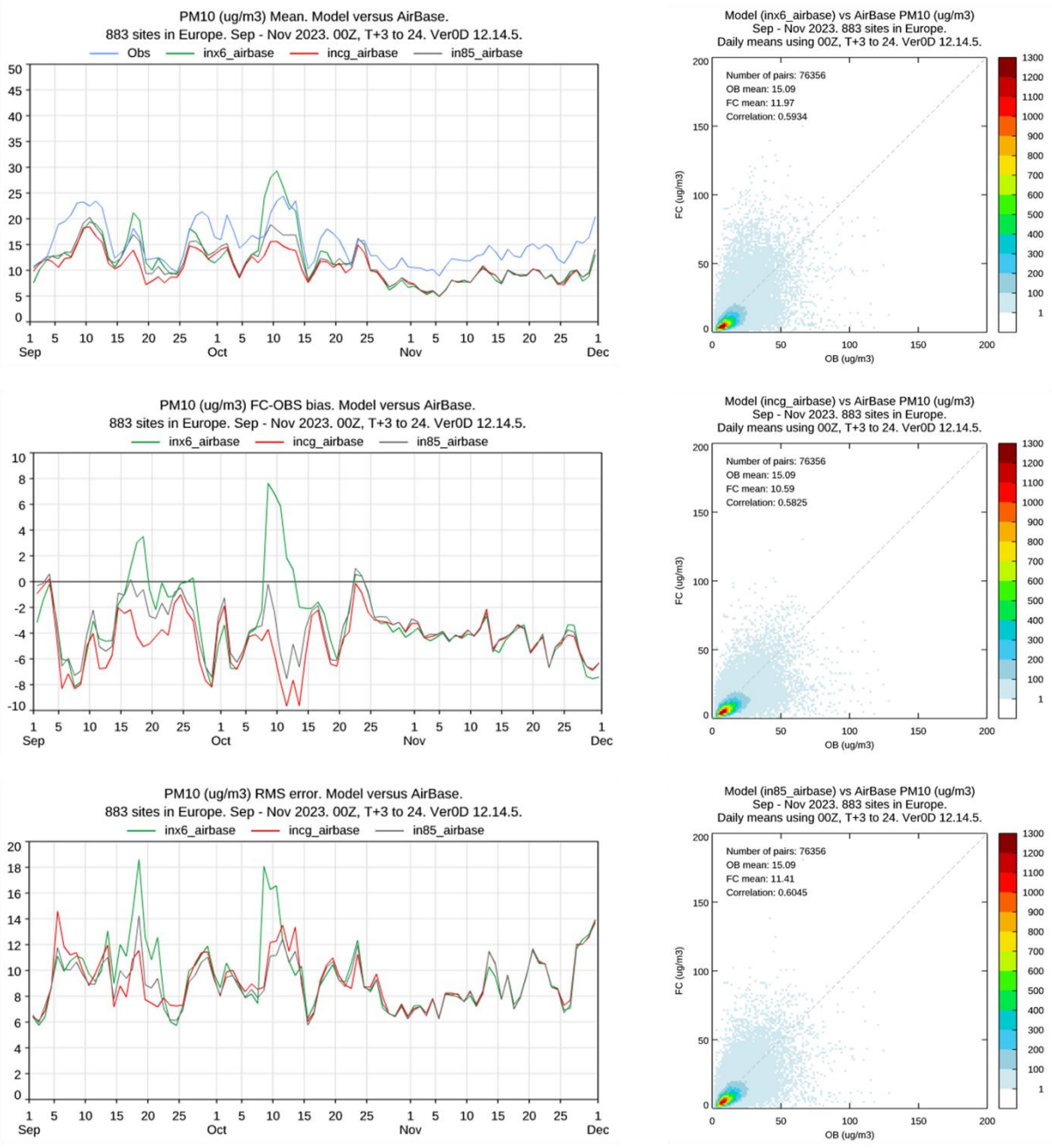

Figure 24. As Figure 20, but showing the analysis results based on using artificial (zero vertical correlation) wavelet files.

Table 3: Differences between mean values of model results and observations, and corresponding RMSE and correlations. The control experiments are marked by grey shading.

Period & wavelets		AOD 870			ÅE 440 - 870			PM2.5			PM10		
	exp ID	bias	rmse	corr	bias	rmse	corr	bias	rmse	corr	bias	rmse	corr
SON 2023 standard	ioer (ctrl)	-0.004	0.016	0.72	0.067	0.16	0.68	0.67	1.88	0.63	-3.1	4.31	0.59
	ioeq (sphr)	0.017	0.025	0.51	0.16	0.21	0.65	0.86	1.73	0.59	-4.2	4.71	0.55
	ioep (sphrd)	0.026	0.031	0.59	0.14	0.20	0.66	1.15	1.95	0.60	-3.5	3.97	0.57
JJA 2024 standard	io8f (ctrl)	-0.006	0.019	0.81	-0.076	0.11	0.80	0.08	1.09	0.48	-3.9	5.12	0.52
	io98 (sphr)	0.040	0.050	0.50	-0.014	0.085	0.76	1.70	2.11	0.43	-3.3	3.80	0.47
	io9f (sphrd)	0.051	0.061	0.59	-0.035	0.090	0.77	2.15	2.49	0.45	-2.2	3.01	0.51
SON 2023 artificial	inx6 (ctrl)	-0.004	0.016	0.72	0.067	0.16	0.68	0.67	1.88	0.63	-3.1	4.31	0.59
	incg (sphr)	0.001	0.017	0.64	0.12	0.18	0.67	0.46	1.51	0.62	-4.5	4.92	0.58
	in85 (sphrd)	0.009	0.018	0.69	0.10	0.17	0.68	0.72	1.67	0.63	-3.7	4.17	0.60

Note that in all experiments and seasons PM10 concentrations are consistently underestimated, while PM2.5 is consistently overestimated. It was also noted that the positive autumn bias in the Ångström exponent indicates that the ratio of coarse particles to fine particles is underestimated. However, during summer the opposite was the case; the negative bias in the Ångström exponent indicates that there is too much coarse mass in relation to fine mass, in contrast to the PM10 and PM2.5 surface biases. A possible explanation is that during summer the IFS produces too much PM2.5 and too little PM10 at the surface, while at higher altitudes their biases are reversed to such an extent that they dominate the bias in the Ångström exponent (which is determined by the particles in the entire atmospheric column).

6 Discussion

The experiments conducted during the first half of the project primarily served as a proof of concept to validate the technical developments, which constituted much of the work. However, the findings offer useful insights into factors that influence how the assimilation of ground-based ceilometer data can improve or degrade CAMS aerosol forecasts. These insights can be exploited for planning the analysis experiments for the second half of the project, in which ceilometer data will be assimilated as part of the full CAMS setup, including AOD assimilation. While in observation space there was an unambiguous improvement by the analysis over the control runs, the results for AOD and PM concentrations turned out to be quite sensitive to the wavelets and the optics model.

Vertical error correlations. The choice of wavelet files has a profound effect on the analysis results. Although the PM10 ground concentrations are somewhat worse than those obtained with the standard wavelets, the corresponding PM2.5 and AOD results are substantially improved. Overall, the use of the artificial wavelets seems to have a positive effect on the analysis, indicating that the standard wavelets may somewhat overestimate vertical correlations.

Optics model for dust aerosols. There are significant differences between analysis experiments depending on which optics model is assumed. This is rather remarkable, considering that only dust aerosols are affected by this assumption. While the homogeneous-sphere model performed better than the homogeneous-spheroid model for AOD and PM2.5, the opposite was the case for PM10. It remains to be seen how the two models perform in conjunction with AOD assimilation.

It should be noted that there are other types of non-spherical aerosols, such as sea salt at low relative humidity and black carbon, which are still modelled by homogeneous spheres. (Volcanic ash, which is also non-spherical, is presently not accounted for in the model.) Further, it is unclear if the spheroid model in the IFS-COMPO is based on the most appropriate choice of aspect ratio AR. The present model uses AR=1.6, which is a median value based on electron microscope studies of mineral dust samples (e.g. Kandler et al., 2007; 2009). Thus, it is meant to closely model the physical overall shape of typical dust aerosols. However, there is accumulating evidence that such a choice of spheroid model often does not well reproduce the optical properties of irregularly shaped mineral dust aerosols (e.g. Kahnert et al., 2014, and references therein). Rather, spheroids should be viewed as a flexible *ad hoc* optics model that can and should be tuned for each optical parameter and wavelength. The model tested here demonstrates the significant importance of taking particle shape into account, but it also leaves much room for further refinements.

There are other, potentially important factors that have not yet been tested in the experiments performed so far.

Masking of unwanted data points. As explained in Section 3.3 the ordering of masking and data-reduction operations can significantly impact the mask. Here a choice was made to use a rather aggressive cloud mask. But there is no evidence whether a more defensive mask, as that in the 2nd row of Figure 3, would improve or degrade the analysis. Such a mask would increase the number of data points used in the assimilation, but it may also increase the risk of assimilating cloud-contaminated data.

Observation errors. The observation error standard deviations were based on those provided in the E-profile data files. These are estimates of the measurement error statistics, but they do not account for the representativeness error, nor for errors introduced by the aerosol-optics forward model. Thus, the experiments may have used poor estimates of the total observation error variances. An alternative approach would be to refine the observation error variances by use of a diagnosis of observation error statistics (Desroziers et al. 2005).

PM2.5 / PM10 partitioning. The 4DVAR system in the IFS compo produces an analysis increment for the total aerosol mass concentrations, which is distributed to all aerosol components according to their *a priori* fractions (Benedetti et al., 2009). Thus, the assimilation changes neither the chemical composition nor the partitioning between PM2.5 and PM10, only the total mass concentration. This approach may be too simplistic for assimilation of ceilometer data for three reasons.

1. All experiments resulted in positive biases in PM2.5 and negative biases in PM10. This may indicate that the PM2.5 / PM10 partitioning is incorrect and would need to be corrected.
2. An incorrect distribution of mass between fine and coarse particles can also partially be responsible for a mismatch between modelled and observed AOD and backscattering. Thus, when only the total mass concentration is being incremented, it may not always be possible to find a mass increment that, mapped into observation space, is consistent with both AOD and ceilometer observations.
3. This point is further corroborated by an analysis of the information contents of aerosol-optics observables (Kahnert and Andersson, 2017), which indicates that by adding backscattering to extinction measurements the number of degrees of freedom for signal is approximately doubled. Based on this finding, one may hypothesise that

simultaneous assimilation of AOD and ceilometer data may contain sufficient information to use two independent control variables, such as PM2.5 and the coarse-mode mass concentration (PM10 – PM2.5).

Aerosol mixing state. The aerosol-optics model in the IFS-COMPO assumes that all aerosol components are externally mixed (Benedetti et al., 2009). Accounting for internal mixture and for the inhomogeneous morphology of such mixtures can significantly change the optical properties to an extent that may rival other sources of error in atmospheric composition models (Andersson and Kahnert, 2016). The external-mixture approximation may introduce systematic errors, and it can result in inconsistencies when simultaneously modelling AOD and backscattering.

7 Conclusion

The overarching goal of WP 1.2 of CAMAERA is to develop a data assimilation methodology in CAMS for E-profile lidar and ceilometer data, and to test the system performance in the IFS-COMPO 4DVar system. During the first 18 months of the project the method-development work was completed according to plan. Specifically, the E-profile data were converted into pre-ODB format, a mask for clouds, rain, fog, and noise was applied, the observation operator, tangent linear model, and adjoint model for molecular absorption at a ceilometer wavelength of 910 nm was parameterised, implemented, and tested in the IFS, and the 4DVar assimilation system was evaluated by performing a number of analysis experiments.

The experiments covered two 3-month periods. Two optics models for mineral dust particles were tested, and two sets of wavelet files were evaluated. Both mineral-dust optics and vertical correlations in the wavelets have a significant impact on the assimilation, especially on AOD and PM concentrations, while the effect in observation space is less pronounced. In the second half of the project, the impact of mineral-dust optics and vertical correlation on assimilation of ceilometer data will be tested within the full CAMS data assimilation setup. In addition, it will be evaluated if the use of a more defensive cloud-masking method may have a significant impact on the analysis experiments. One important scientific question is if these tests will provide more indications for a need to partition the PM control variable into a coarse and fine mode when simultaneously assimilating AOD and attenuated backscatter coefficient.

8 References

Andersson, E. and Kahnert, M., 2016: Coupling aerosol optics to MATCH (v.5.5.0) chemical transport model and SALSA (v1) aerosol microphysics module. Geosci. Model Dev. 9, 1803-1826.

Benedetti, A., Morcrette, J.-J., Boucher, O. et al., 2009 : Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: 2. Data assimilation. J. Geophys. Res. 114, D13205.

Desroziers, G., Berre, L., Chapnik, L. et al., 2005: Diagnosis of observation, background and analysis-error statistics in observation space. Q. J. R. Meteorol. Soc. 131, 3385-3396.

ECMWF, 2024a: Documentation CY49R1 – Part VIII : Atmospheric Composition. doi : 10.21957d13af18259.

ECMWF, 2024b: Documentation CY49R1 – Part I : Observations. doi : 10.21957/fd16c61484.

ECMWF, 2024c: Documentation CY49R1 – Part II : Data Assimilation. doi : 10.21957/105cb1333c.

Gasteiger, J. and Wiegner, M. 2028 : Mopsmap v1.0 : a versatile tool for the modelling of aerosol optical properties. Geosci. Model Dev. 11, 2739-2762.

Geleyn, J.-F., Benard, P. and Fournier, R., 2005: A general-purpose extension of the Malkmus band-model average equivalent width to the case of the Voigt line profile. Q. J. R. Meteorol. Soc. 131, 2757-2768.

Hogan, R. J. and Matricardi, M., 2020: Evaluating and improving the treatment of gases in radiation schemes: the Correlated K-Distribution Model Intercomparison Project (CKDMIP). Geosci. Model Dev. 13, 6501-6521.

Kahnert, M., Nousiainen, T., and Lindqvist, H., 2014: Model particles in atmospheric optics. J. Quant. Spectrosc. Radiat. Transfer 146, 41-58.

Kahnert, M. and Andersson, E., 2017: How much information do extinction and backscattering measurements contain about the chemical composition of atmospheric aerosol? Atmos. Chem. Phys. 17, 3423-2017.

Kandler, K., Benker, N., Bundke, U. et al., 2007: Chemical composition and complex refractive index of Saharan Minearl Dust at Izaña, Tenerife (Spain) derived by electron microscopy. Atmos. Env. 41, 8058-8074.

Kandler, K., Schütz, L., Deutscher, C. et al., 2009: Size distribution, mass concentration, chemical and mineralogical composition and derived optical parameters of the boundary layer aerosol at Tinfou, Marocco, during SAMUM 2006. Tellus 61B, 32-50.

Document History

Version	Author(s)	Date	Changes
1.0	Michael Kahnert	12/06/2025	Initial version
1.1	Michael Kahnert	26/06/2025	New version after internal reviews

Internal Review History

Internal Reviewers	Date	Comments
Thierry Elias	18/06/2025	Clarifications of the developments carried out
Samuel Remy	18/06/2025	Ask for Ångström exponent plots