https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

Assimilation of ground based lidar and ceilometer observations of aerosols from the European E-Profile network into ECMWF's Integrated Forecasting System (IFS-COMPO, CY49R1)

Michael Kahnert[1], Melanie Ades[2], Mickaël Bacles[3], Johannes Flemming[1], Vincent Guidard[3], Alexander Haefele[4], Robin J. Hogan[2], Samuel Rémy[5], and Eric Sauvageat[4]

[1]European Centre for Medium-Range Weather Forecasts, Bonn, Germany
[2]European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom
[3]Météo-France, CNRS, Univ. Toulouse, CNRM, Toulouse, France
[4]Federal Office of Meteorology and Climatology MeteoSwiss, Payerne, Switzerland
[5]HYGEOS, Lille, France

Correspondence: Michael Kahnert (michael.kahnert@ecmwf.int)

Abstract. The Integrated Forecasting System with its extension for atmospheric composition (IFS-COMPO) provides global forecasts of atmospheric trace gases and aerosols for the Copernicus Atmosphere Monitoring Service (CAMS). The present system constrains aerosol concentrations by assimilating aerosol optical depth (AOD) from different satellites. Here, we explore the possibility of assimilating, in addition, ground-based lidar and ceilometer observations from the European E-Profile network. The system performance is evaluated by comparison to non-assimilated E-Profile stations, AOD observations from Aeronet, and aerosol surface concentrations from AirBase. Assimilation of E-Profile data significantly reduces biases and root mean square errors (RMSE) of model-equivalent vertical profiles of the attenuated backscatter coefficient. Without assimilation of E-Profile, surface concentrations of particles smaller than 2.5 μm (PM2.5) are frequently overestimated during summer, while corresponding concentrations of particles smaller than 10 μm (PM10) tend to be underestimated. Assimilation of E-Profile can reduce the RMSE of PM2.5 by up to 50% and of PM10 by up to 10 %. Since the present analysis system uses the total aerosol mass mixing ratio as control variable, it cannot simultaneously reduce the positive PM2.5 bias and the negative PM10 bias. It typically reduces the PM2.5 bias at the expense of PM10, since fine particles make the dominant contribution to the optical cross sections per mass. Tests of different assimilation-system configurations reveal that the best overall performance is obtained by treating optical properties of dust with a spheroid model, suppressing vertical correlations in the background error covariances, and applying a relatively aggressive cloud mask.

1 Introduction

The Copernicus Atmospheric Monitoring Services (CAMS, https://atmosphere.copernicus.eu), established by the European Commission in 2014, provides long-term reanalysis (Inness et al., 2019) as well as near-real time forecasts and analyses of atmospheric composition. The modelling system and methodology in CAMS evolved over a series of collaborative programmes starting with GEMS (Hollingsworth, 2008); it contains a global (Peuch et al., 2022) and a regional component (Colette et al.,

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

2025). The global production system relies on the Integrated Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecast (ECMWF) with its extension for greenhouse gases (GHG) and atmospheric composition (COMPO). The IFS-COMPO provides global forecasts and analyses to users, as well as boundary conditions for the CAMS regional ensemble production system. An important component of the global system is the capacity for data assimilation of reactive gases (Inness et al., 2013) and aerosols (Benedetti et al., 2009). At the moment, aerosols are constrained by assimilating aerosol optical depth (AOD) from different satellite sensors, which only provides a proxy for the aerosol load in the entire atmospheric column.

Lidar and ceilometer observations provide vertical profiles of the laser power backscattered by aerosols and gases. These observations are potentially useful for constraining the vertical distribution of aerosols in forecasting models. Several studies have been performed in which observations from ground-based and satellite-borne lidar instruments have been assimilated (e.g. Yang et al. (2023) and references therein). In this article we present a pilot study on assimilating profiles of the attenuated backscatter coefficient observed by lidar and ceilometer instruments in the European E-Profile network (EUMETNET, 2025) into the IFS-COMPO.

2 Model and observations

2.1 CAMS model and data assimilation system

The IFS-COMPO is an online model, i.e., meteorological and compositional forecasts and analyses are provided alongside each other, allowing for an integrated approach by accounting for the radiative impact of dynamic aerosol fields on meteorological processes. The IFS-COMPO contains different options for simulating reactive gases and aerosols, and it has been continuously updated. In this article, we use the operational configuration of IFS Cycle 49R1 (hereafter CY49R1), (ECMWF, 2024)) for CAMS forecasts of reactive gases and aerosols (Flemming, 2015; Huijnen et al., 2016). The atmosphere is discretised by 137 vertical levels with a horizontal resolution of $0.35^\circ \times 0.35^\circ$, spectral truncation T511. The tropospheric and stratospheric chemistry schemes are based on Williams et al. (2022) and Chabrillat et al. (2025), respectively, while the aerosol scheme is described and evaluated in Rémy et al. (2022, 2024).

There are in total 16 prognostic aerosol variables: three size bins of sea salt, three size bins of dust, hydrophobic and hydrophilic black carbon, hydrophobic and hydrophilic organic carbon, sulfate, fine and coarse nitrate, ammonium, biogenic secondary organic aerosols (SOA), and anthropogenic SOA. A detailed description of the chemistry and aerosol schemes, emission inventories, online emissions for dust particles, marine aerosols, and biogenic fluxes, as well as transport and removal processes can be found in the IFS documentation (ECMWF, 2024).

The IFS employs an incremental formulation of the four-dimensional variational (4DVAR) approach (Courtier et al., 1994) with 12-hour assimilation windows, which in CAMS range from 03:00 to 15:00 and from 15:00 to 03:00 UTC. The initial control vector is optimised by minimising a cost-function

$$J(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^T \cdot \mathbf{B}^{-1} \cdot \delta\mathbf{x} + \frac{1}{2}\sum_{i=0}^{n}(\mathbf{H}_i \cdot \delta\mathbf{x}(t_i) - \mathbf{d}_i)^T \cdot \mathbf{R}_i^{-1} \cdot (\mathbf{H}_i \cdot \delta\mathbf{x}(t_i) - \mathbf{d}_i), \quad (1)$$

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

where $\delta\mathbf{x}$ denotes the control vector at initial time, and $\mathbf{B}$ represents the background error covariance matrix. The sum extends over discrete time steps covering the assimilation window, and $\delta\mathbf{x}(t_i)$, $\mathbf{H}_i$, and $\mathbf{d}_i$ denote, respectively, the control vector, linearised observation operator, and innovations at time step i. The latter are given by

$$\mathbf{d}_i = \mathbf{y}_i - H_i\,\mathbf{x}_b(t_i), \tag{2}$$

where $\mathbf{x}_b(t_i)$ is the background model state propagated to time t_i, H_i is the observation operator at that time, and $\mathbf{y}_i$ is the vector of observations at time index i. The cost function measures the difference between the model field and the observations during the assimilation window. The model trajectory $\mathbf{x}_b(t_i)$ in Eq. (2) is computed with the full nonlinear forecast model, and it is mapped into observation space with a generally nonlinear observation operator H_i. For these calculations the full model resolution is used. By contrast, the time-consuming iterative minimisation procedure of Eq. (1) is performed at a lower resolution, where $\delta\mathbf{x}(t_i)$ is computed from the initial state by the tangent linear model, and it is mapped into observation space by a tangent linear approximation $\mathbf{H}_i$ of the observation operator.

The minimum of the cost function is the analysis increment $\delta\mathbf{x}_a$ at initial time, which is added to the background field to obtain an analysis for the initial state,

$$\mathbf{x}_a = \mathbf{x}_b + \delta\mathbf{x}_a. \tag{3}$$

The control vector $\delta\mathbf{x}$ contains both meteorological and compositional variables. For aerosols, only the total mass mixing ratio (MMR) is included in the control vector. The analysis increment for the aerosol MMR is partitioned among the 16 aerosol species according to their proportions in the background aerosol field (Benedetti et al., 2009).

There is an extensive number of satellite products of reactive gases and aerosol optical depth (AOD) that are actively assimilated or passively monitored in the operational suite (o-suite) of IFS-COMPO, as listed in ECMWF (2024). Specifically for aerosols, AOD is assimilated from MODIS on both Aqua and Terra, VIIRS on both SNPP and NOAA-20, and from the PMAP product based on METOP-B and METOP-C. Being an online model, compositional satellite products are assimilated alongside observations for the numerical weather prediction (NWP) part of the IFS. This setup is used in the global component of the CAMS forecasting system. In the experiments reported in this study, this system will be employed for the control experiments, but using slightly different background errors than in the o-suite (as explained later). The potential usefulness of ground-based ceilometer and lidar data will be tested by assimilating attenuated backscatter coefficients from the E-Profile network in addition to all other observations in the o-suite of the IFS-COMPO.

A description of the observation operator for the attenuated backscatter coefficient is given in App. A. In CY49R1 the lidar wavelengths 1064, 532, and 355 nm are available. For the assimilation of E-Profile data it is necessary to extend the observation operator to include a wavelength of 910 nm. This extension, and specifically the parameterisation of molecular absorption at that wavelength, is described in App. A1. A new optics model for the lidar ratio, in which dust aerosols are described by spheroidal particles, is presented in App. A2. A comparison of the system performance with dust-optics models based on spheres and spheroids is given in App. D1, and the overall improvements obtained with the new model are discussed in Apps. D1 and D5.

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

instrument	type	manufacturer	λ [nm]	vertical range [km]	vertical resolution [m]	No.
CL31	ceilometer, diode laser	Vaisala	910	≤ 7.6	5–10	153
CL51	ceilometer, diode laser	Vaisala	910	≤ 15	10	34
CHM15k	ceilometer, Nd:YAG solid-state laser	Lufft	1064	7–15	5	128
Mini-MPL	lidar, Nd:YAG solid-state laser	Micro Pulse LiDAR	532	≤ 30	5–75	6

Table 1. Instruments from which data have been assimilated. The maximum vertical ranges and resolutions are based on the manufacturers' specifications. In the actual analysis experiments the vertical ranges have been capped and the vertical resolution was reduced as explained in Sect. 2.2.3. The last column shows the total number of stations that were used in this study for assimilation and validation.

The IFS uses a wavelet representation of background error covariances. Here we are using artificial wavelets in which vertical correlations have been suppressed. Further details are provided in App. B. A comparison of the system performance with and without vertical correlations can be found in Apps. D3 and D4.

2.2 E-Profile

E-Profile data are operationally available in near-real time (https://e-profile.eu), typically with a latency of three hours between observation and delivery to ECMWF. This short latency is a key motivation for investigating the potential benefits of assimilating these data into the global CAMS system, as near-real time availability is a fundamental requirement for inclusion in the operational IFS runs.

2.2.1 Station network and instruments

The E-Profile network comprises a growing number of stations using different lidar and ceilometer instruments (Haefele et al., 2022). For the analysis experiments, two periods have been selected, SON 2023 and JJA 2024. Data from a total of 321 stations are used for assimilation and evaluation. The types of instruments and number of stations from which data were available during these periods are listed in Table 1. As illustrated in Fig. 1, the stations were partitioned into two subsets, a larger one for assimilation and a smaller one for evaluation. The latter comprises 30 stations with CHM15k instruments and 44 stations with either CL31 or CL51 instruments.

2.2.2 Processing of E-Profile data

E-Profile lidar and ceilometer measurements are processed centrally and provided in a common file format for the different instrument types. The vertical profiles of attenuated backscatter are averaged on 5 minutes, whereas the altitude range and vertical resolution depend on the instrument type. The data are provided with a vertical resolution of either 20 (CL31 and CL51) or 30 meters (CHM15k and Mini-MPL). Overlap correction is performed by the manufacturers and varies depending on the instrument type.

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

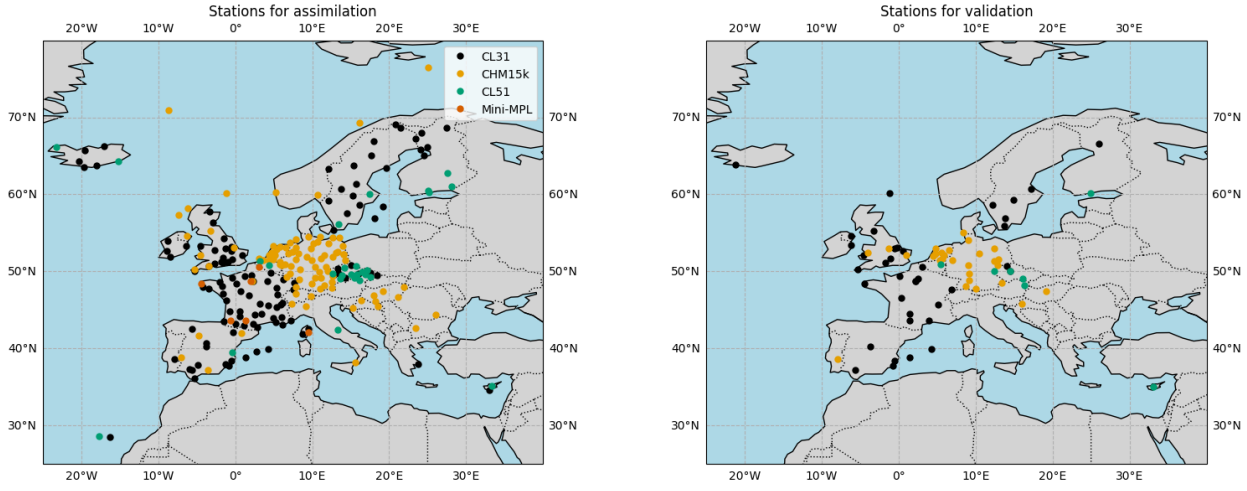

Figure 1. E-Profile stations used in this study for assimilation (left) and validation (right). The colours represent different types of instruments as explained in the legend and detailed in Tab. 1. Two CHM15k stations are outside the plotted area, one in Southern Ontario, the other in the Dutch Antilles.

For the Vaisala instruments (CL31 and CL51), the calibration of the attenuated backscatter profile is provided directly by the manufacturer and is constant in time. However, for the CHM15k and the Mini-MPL, E-Profile applies its own calibration procedure, based on the Rayleigh calibration technique (e.g. Wiegner and Geiß (2012)) and usually updated on a monthly basis. Following recent improvements to the calibration procedure, we reprocessed the CHM15k time series for the two assimilation experiment periods, using a daily calibration value. It should reflect the current quality of the data distributed in near real-time.

2.2.3 Preprocessing of E-Profile data for assimilation into the CAMS system

The E-Profile data are available in netCDF format and can be downloaded from CEDA Archive (2018). A script has been created for automatic navigation and downloading of data for any desired period and list of stations, which has been used to download the CL31 and CL51 stations. The reprocessed CHM15k and Mini-MPL data have been provided by MeteoSwiss.

To prepare the data for assimilation the following preprocessing steps are performed.

- Masking of unwanted data.
- Reduction of temporal and vertical resolution.
- Unit conversions.
- Reordering of bottom-up to top-down vertical profiles.
- Pooling of the data and export into ODB format.

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

A critical step is the masking of clouds, precipitation, fog, and noise. In the body of the paper, we show results for a masking approach that aggressively removes clouds, while being relatively conservative with noise masking. This approach proved to be most reliable as compared to other masking strategies. More details on masking of unwanted data and other preprocessing operations are provided in App. C, and a comparison of two different masking approaches is provided in Apps. D2 and D4.

3 Analysis experiments

The starting point for the analysis experiments reported here is the o-suite of the IFS-COMPO with assimilation of all NWP observations and satellite products for reactive gases and AOD used in the global CAMS production system. This setup is used for the control experiments. The only difference in our control experiments to the setup in the o-suite is that we use wavelets with vertical correlations being disabled. To evaluate the potential benefits of lidar and ceilometer data, attenuated backscatter coefficients from the E-Profile network are additionally assimilated, and the outcome is compared to the control runs. Only the stations indicated in the left panel of Fig. 1 are assimilated.

The experiments were run for two 3-month periods, SON 2023 (2023-09-01 – 2023-11-30) and JJA 2024 (2024-06-01 – 2024-08-31). For each period a 1-month spin-up was used. We use a configuration of the E-Profile assimilation that proved to give optimum results. In this approach the optical properties of dust were derived from spheroids, vertical correlations in the background error covariances were suppressed, and an aggessive cloud-masking approach was applied to the input data. A more comprehensive evaluation of the assimilation system with different configurations is provided in App. D.

3.1 Impact of the analysis on vertical profiles of the attenuated backscatter coefficient

It is instructive to first look at two specific examples, shown in Figs. 2 and 3. Figure 2 (top left) shows the concentration of coarse dust aerosols at 800 hPa on September 6, 2023. The image represents the 00+03 hour forecast of the control experiment. There is an aerosol plume originating from a Sahara dust storm that passes over the Iberian peninsula, western parts of France, and reaches the UK. The analysis increments are shown for the control run (assimilating only AOD, centre left), and for assimilating both AOD and E-Profile observations (bottom left). The E-Profile assimilation results in a region of negative analysis increments ranging from the south of France to the south of the UK (bottom left), which is almost absent in the control experiment (centre left). For illustrative purposes, four profiles of the attenuated backscatter coefficient observed at non-assimilated E-Profile stations are selected that lie within the path of the dust plume: Gourdon, FR, (September 4 15:00 UTC, top centre), Palaiseau, FR (September 6 03:00 UTC, top right), Nottingham, UK (September 6 12:00 UTC, bottom centre), and Cranwell, UK (September 6 12:00 UTC, bottom right). The observations are shown in black, the control run in orange (dashed line), and the analysis experiment with E-Profile assimilation in light blue (dash-dotted line). Near the surface all experiments agree reasonably well with the observations. But between 1–3 km altitude, where the dust plume manifests itself in the backscattered signal, the control run significantly overestimates the observations. By contrast, the analysis experiment brings the model-equivalent backscattering analysis into much closer agreement with the observations.

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

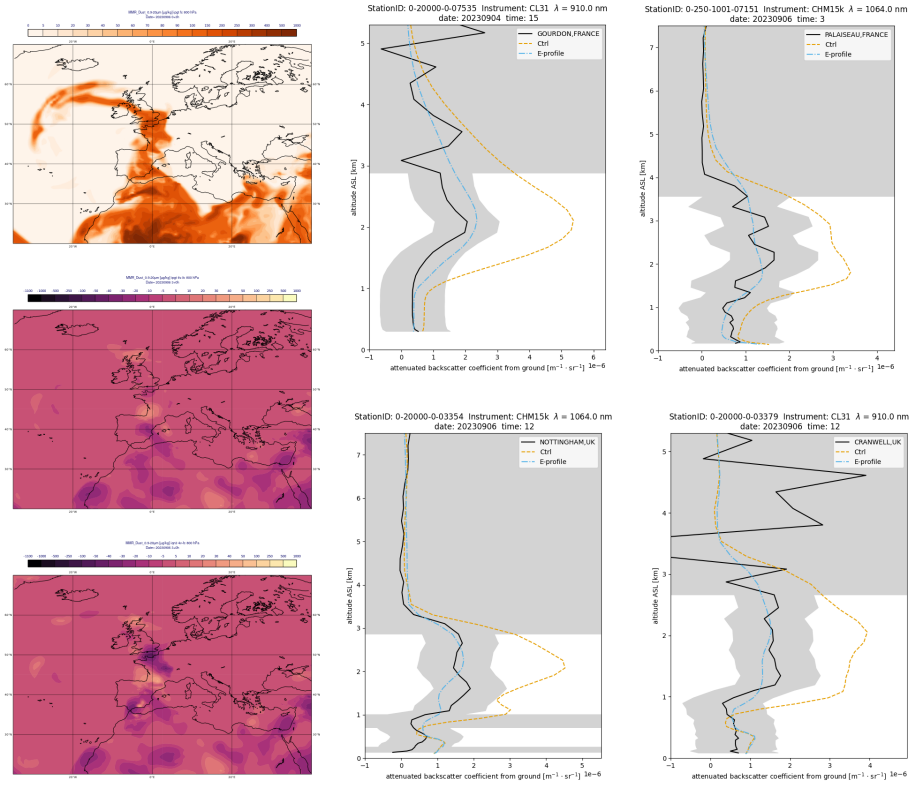

Figure 2. Control-run forecast of the coarse-dust concentration at 800 hPa during a Sahara dust storm on 6 September 2023 at forecast time 00+03h UTC (top left). Analysis increment at 03:00h UTC for the control run (AOD only assimilation, centre left), and for AOD + E-Profile assimilation (bottom left). Attenuated backscatter coefficients at selected non-assimilated stations at different times during the event: Gourdon, FR (top centre), Palaiseau, FR (top right), Nottingham, UK (bottom centre), and Cranwell, UK (bottom right). The lines show observations (black), the control experiment (orange dashed line), and analysis experiment with E-Profile assimilation (light blue dash-dotted line).

Figure 3 (top left) shows the concentration of hydrophilic organic matter at 800 hPa on 19 August 2024, taken from the 12+03h forecast of the control experiment. One can see a smoke plume originating from Canadian wildfires and reaching parts of Western Europe. The analysis increment in the control run (centre left) reveals both positive and negative increments over North America and the Atlantic, but nearly zero increments over Europe. By contrast, assimilation of AOD and E-Profile observations (bottom left) produces both positive and negative analysis increments over Europe. Four selected, non-assimilated profiles are shown that lie in the path of the plume (all on August 19): Nottingham, UK (00:00 UTC, top centre), Deauville, FR (12:00 UTC, top right), Lelystad, NL (15:00 UTC, bottom centre), and List, DE (21:00 UTC, bottom right). The observations (black) are mostly peaking around 2-3 km, which is well captured by the control experiment (orange dashed line). However,

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

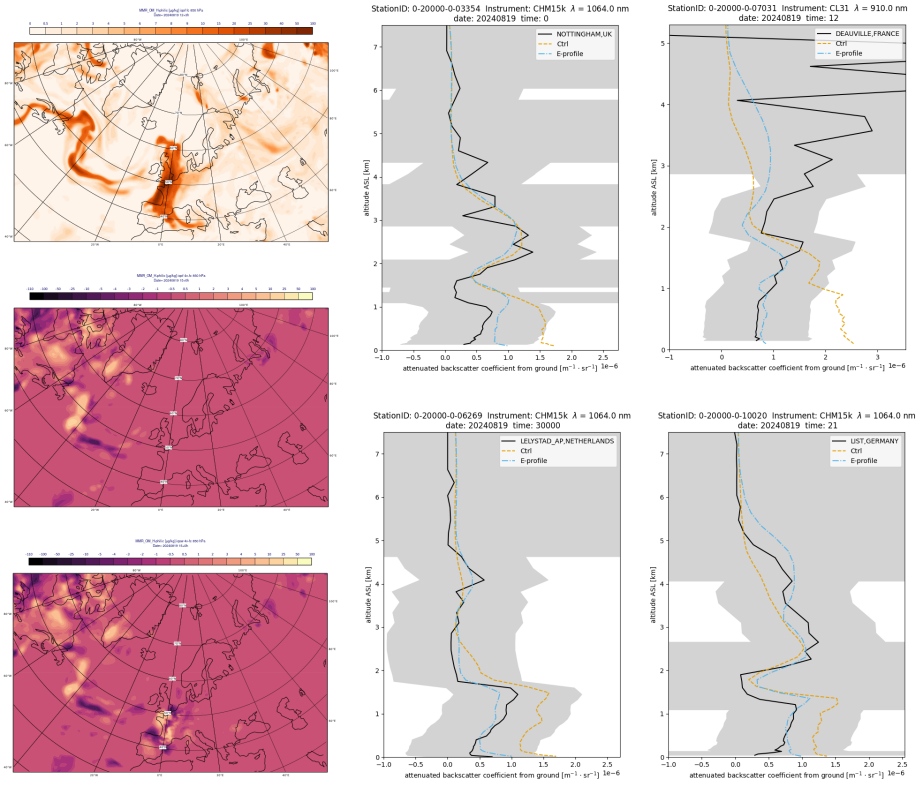

Figure 3. Control-run forecast of the hydrophilic organic matter concentration at 800 hPa during a Canadian wildfire on 19 August 2024 at forecast time 12+03h (top left). Analysis increment at 15:00h UTC for the control run (AOD only assimilation, centre left), and for AOD + E-Profile assimilation with spheroidal dust optics (bottom left). Attenuated backscatter coefficients at selected non-assimilated stations at different times during the event: Nottingham, UK (top centre), Deauville, FR (top right), Lelystad, NL (bottom centre), and List, DE (bottom right). The line colours and analysis experiments are as in Fig. 2.

the control run overestimates, in these examples, backscattering in the lowermost 1.5 km. The E-Profile analysis experiment (light blue dash-dotted line) fits the observations more faithfully.

Can these findings be generalised? Figure 4 shows the attenuated backscatter coefficient at 1064 nm wavelength for the entire autumn period (SON 2023) averaged over all 30 non-assimilated CHM15k stations. The top row shows, from left to right, the observations, the control run, and the E-Profile analysis experiment. The corresponding biases and root-mean square errors (RMSE) are shown in the second and third row, respectively. There are episodes of large discrepancies between the control run and the observations during which the control experiment mostly overestimates the backscattering signal, notably around day 40 (10 October), but also at day 18, 48, 52, and others. Near the surface, the control experiment often underestimates the observations. These discrepancies also manifest themselves in the bias and RMSE. Assimilation of E-Profile data significantly reduce both the bias and the RMSE (rightmost column).

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

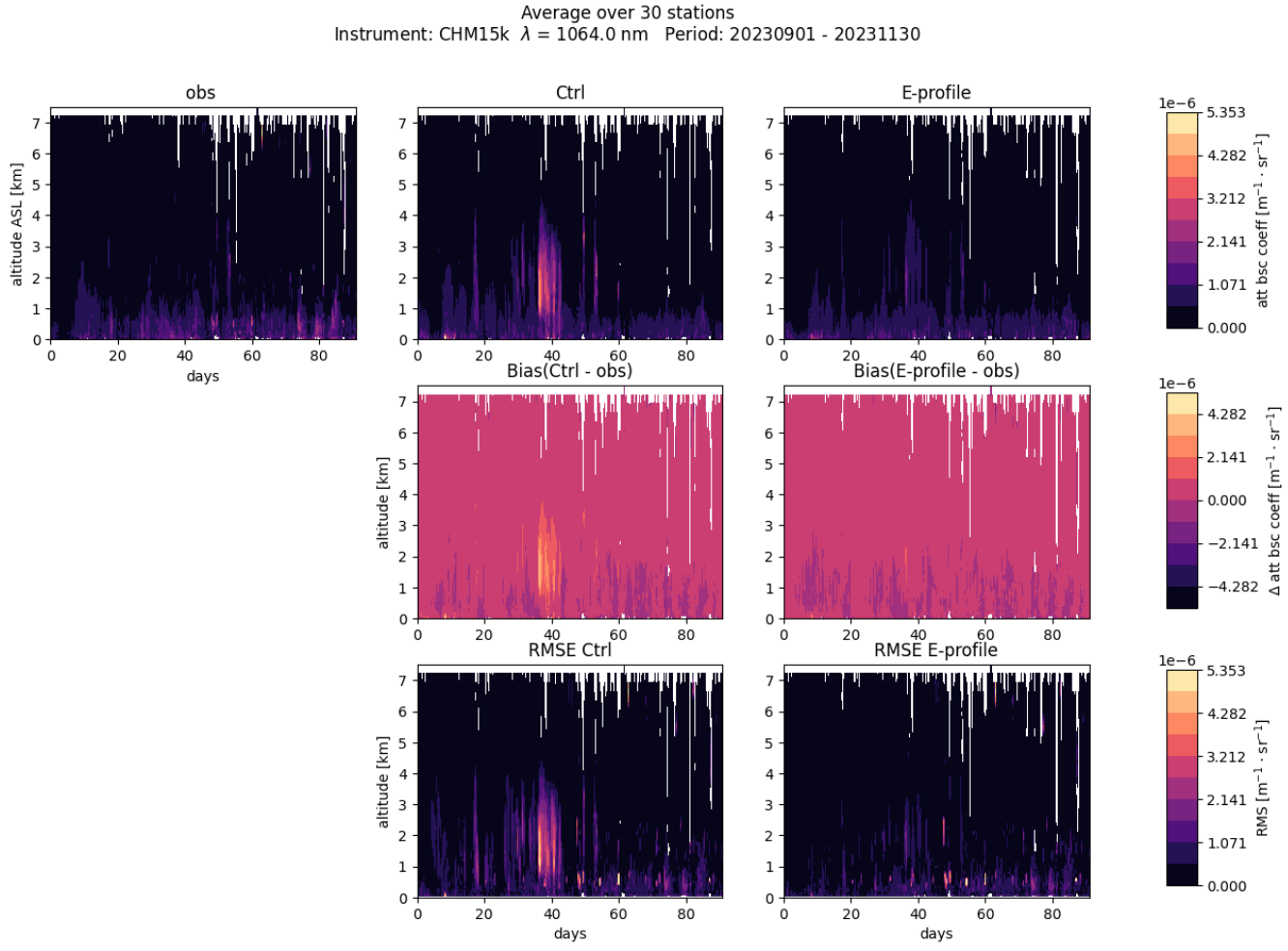

Figure 4. Attenuated backscatter coefficient at 1064 nm wavelength during SON 2023, averaged over 30 non-assimilated CHM15k stations. Top row (from left to right): observations, control experiment, and E-Profile analysis experiment. Second row: bias (model - observations). Third row: root mean square error.

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

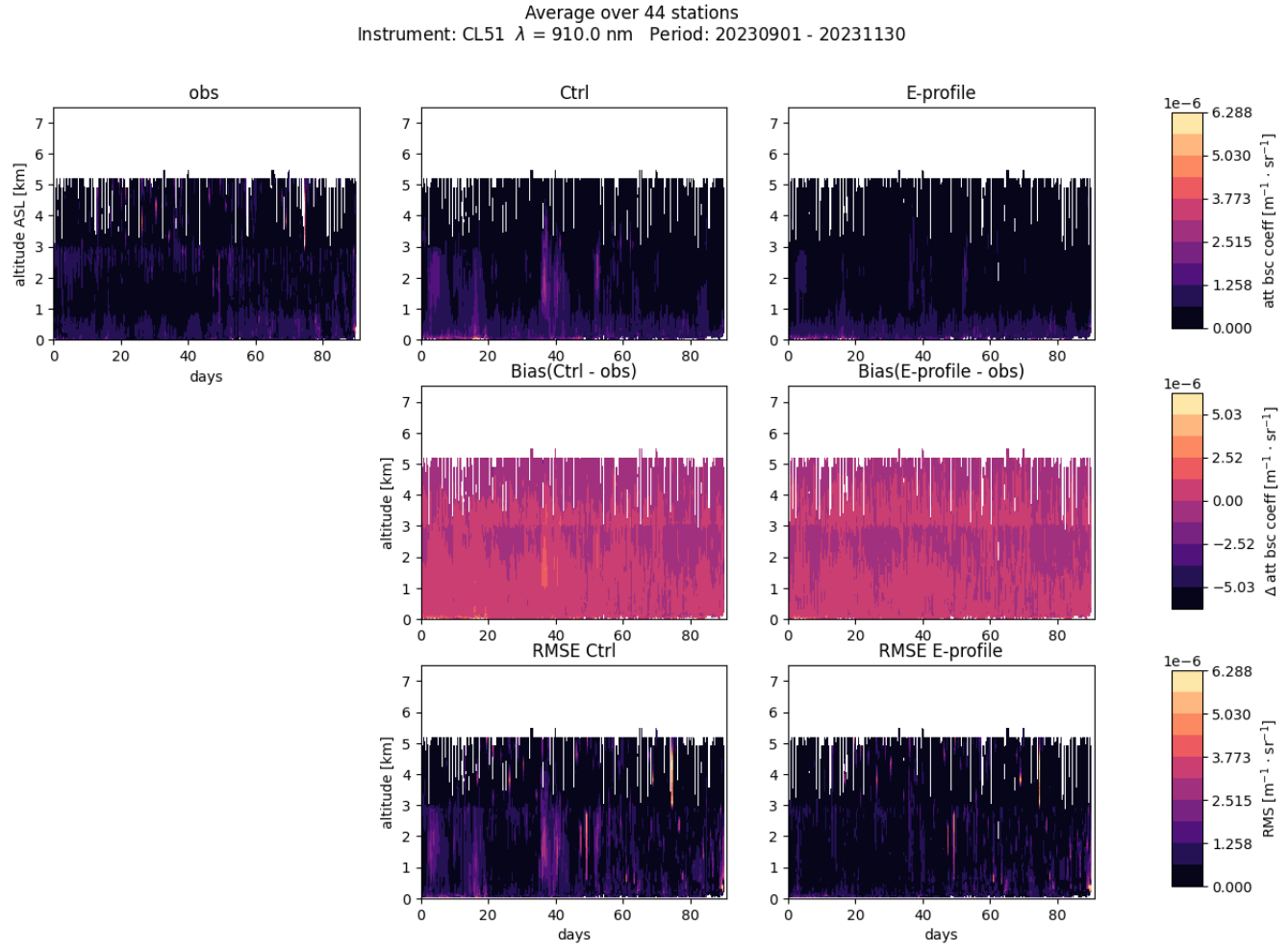

Figure 5. As Fig. 4, but averaged over 44 non-assimilated CL31 and CL51 stations.

The results in Fig. 4 were based on averaging over non-assimilated CHM15k stations, which observe at 1064 nm wavelength. As seen in Fig. 1, CHM15k tend to be clustered in Germany, Switzerland and the Netherlands, while CL31 and CL51 stations are more common elsewhere. Figure 5 shows results analogous to those in Fig. 4, but averaged over 44 non-assimilated CL31 and CL51 stations. The general trends are closely analogous to those observed for the CHM15k stations. The control run periodically overestimates the observed attenuated backscattering coefficient, which is accompanied by high RMSE. Assimilation of E-Profile observations reduces the bias and substantially lowers the RMSE. However, the observations seem to have an artifact around 3 km altitude. This is because the Figure shows an average over both CL31 and CL51 stations. The former were capped at 3 km, the latter at 5.3 km — see Sect. C. The aerosol backscatter signal of the CL31 may be too unreliable for altitudes around 3 km and should, possibly, be capped at 2–2.5 km.

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

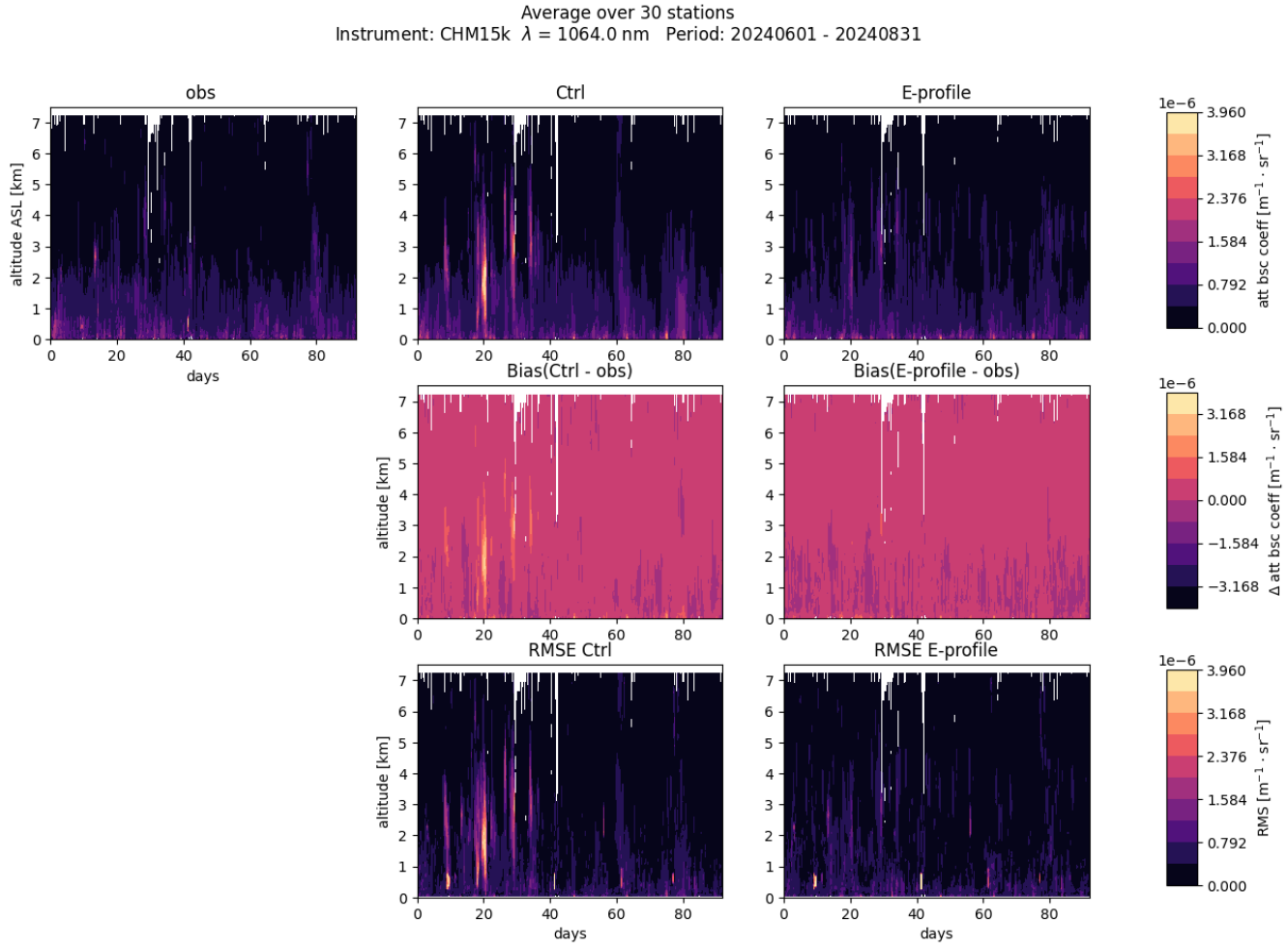

Figure 6. As Fig. 4, but for JJA 2024.

Figures 6 and 7 show results analogous to those in Figs. 4 and 5, but for the summer period (JJA 2024). Overall, the observed attenuated backscatter coefficient is higher and vertically more extensive than in the autumn period. As before, there are several episodes during which the control run overestimates the observations aloft, e.g. around days 20, 27, 33, 62, and 79. The E-Profile analysis experiment consistently reduces the bias and, even more so, the RMSE.

3.2 Impact of the analysis on aerosol mass concentrations at the surface and on AOD

Figure 8 illustrates the effect of assimilating E-Profile observations on aerosol mass concentrations at the surface and on AOD. The examples are based on SON2023. PM2.5 (top) and PM10 results (2nd row) are compared to 447 (PM2.5) and 901 (PM10) AirBase sites in Europe, while AOD at 870 nm wavelength (3rd row) is compared to level-2 data from 115 Aeronet stations in

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

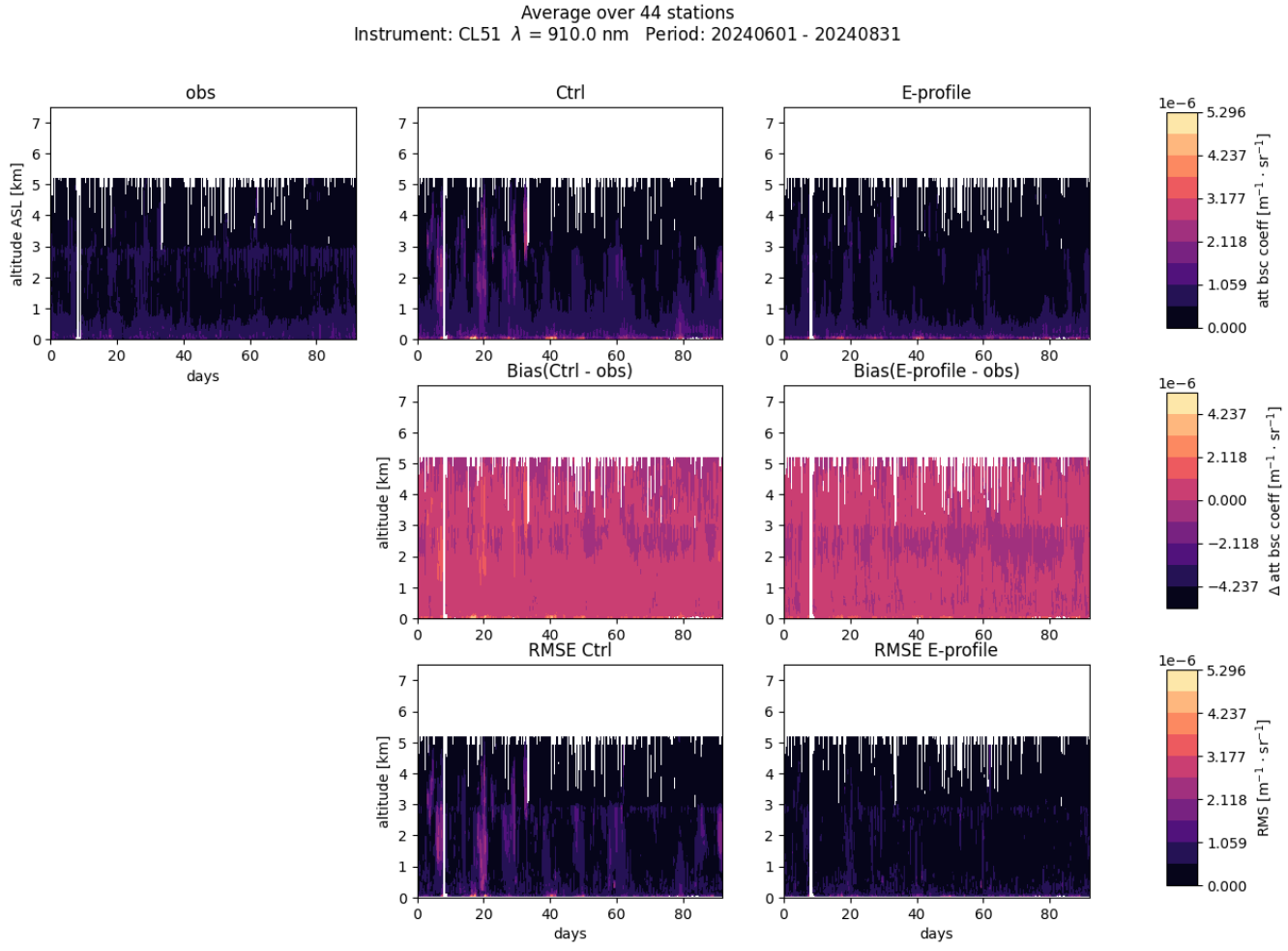

Figure 7. As Fig. 5, but for JJA 2024.

Europe. This Aeronet wavelength lies closest to 910 nm and 1064 nm, at which most of the E-Profile measurements are taken. The results are averaged over all stations, and the time series of the bias (left) and RMSE (right) are shown.

The control experiment (green) mostly overestimates PM2.5 (top left) during the first two months, and underestimates it during the last month. PM10 (2nd row left) is mostly underestimated by the control experiment. The E-Profile assimilation experiment periodically decreases the PM2.5 and PM10 biases, while at other times there is a degradation. By contrast, the RMSE of PM2.5 (top right) and PM10 (2nd row right) is mostly reduced by the assimilation of E-Profile observations.

The magnitude of the bias in AOD (3rd row left) is mostly reduced by the assimilation of E-Profile data. This is rather unexpected. AOD is the only observation constraining the aerosols in the control analysis experiment. By contrast, when the analysis has to minimise the cost function by accommodating both AOD *and* ceilometer observations, then one would expect that the overall match with independent AOD observations may be degraded compared to the AOD-only control run.

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

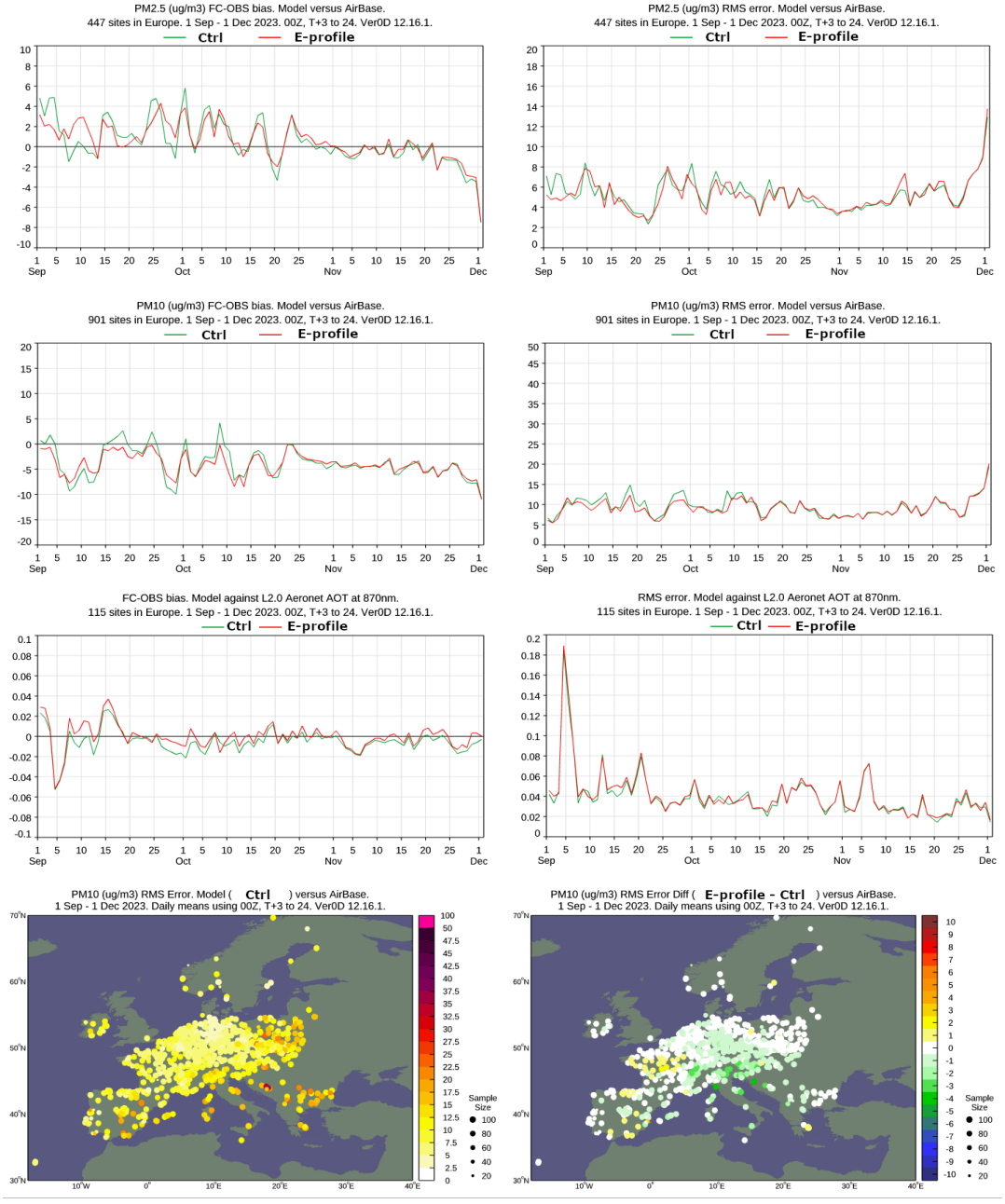

Figure 8. Top row: bias (left) and RMSE (right) of PM2.5 surface concentrations during SON2023, averaged over 447 European AirBase stations. The control experiment is shown in green, the E-Profile analysis experiment in red. Second row: analogous results for PM10, averaged over 901 European AirBase stations. Third row: analogous results for AOD, averaged over 115 European Aeronet stations. Bottom row: PM10 RMSE of the control experiment (left) and RMSE difference (right) between the control and the E-Profile assimilation experiment.

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

The bottom row shows the seasonally averaged RMSE in PM10 at 901 AirBase sites for the control experiment (left), and the RMSE differences between the control and the E-Profile assimilation run (right), where green shades indicate that the E-Profile assimilation reduces the RMSE. Comparison with Fig. 1 suggests that regions in which the RMSE is most strongly reduced correlate with those in which the observational network is dominated by instruments of higher quality (CHM15k, CL51, and miniMPL), while regions with a higher density of CL31 instruments or no instruments show less positive impact, no impact, or even a mild degradation.

Figure 9 Shows similar results to those in Fig. 8, but for JJA2024. PM2.5 (top) and PM10 results (2nd row) are compared to 441 (PM2.5) and 981 (PM10) AirBase sites in Europe, while AOD at 870 nm wavelength (3rd row) is compared to level-2 data from 83 Aeronet stations in Europe. The control experiment (green) generally overestimates PM2.5 (top left), and it underestimates PM10 (2nd row left). Assimilation of E-Profile data results in a significant reduction of the PM2.5 bias and RMSE. For PM10 (2nd row) the bias (left) is degraded, while the RMSE (right) is reduced by the assimilation of E-Profile data.

This opposite effect on PM2.5 and PM10 biases can be understood as follows. It was apparent in Fig. 6 that the control run mostly overestimates the observed backscattering signal. Accumulation-mode aerosols tend to make the dominant contribution to the optical cross sections per mass. Therefore, the overestimation of the backscattering signal is likely to be caused by an overestimation of fine particles, which is consistent with the overestimation of PM2.5 at the surface that is evident in Fig. 9. The model-equivalent attenuated backscatter coefficient can be brought into closer agreement with the observations by reducing the mass concentration of fine particles. However, the control variable is the *total* aerosol mass concentration. Therefore, a reduction in fine aerosols also entails a reduction in coarse aerosols. At the surface, the latter are underestimated, so that a reduction in total aerosol mass by the analysis results in an improvement for PM2.5 and, inevitably, a degradation for PM10.

AOD (third row) shows a degradation. This is different from what was observed for SON2023, but, as discussed there, more in line with what one would expect for an observing system in which the analysis has to simultaneously fit AOD and backscattering observations.

The fourth row shows the seasonally averaged PM2.5 bias in the control experiment at the 441 AirBase sites (left), and the difference in biases between the control experiment and E-Profile assimilation experiment (right). Positive biases (left) are found in Northern Germany, the Benelux countries, Northeastern France, Northern Italy, Corsica, Sardinia, along the northern and western coast of the Iberian Peninsula, and at some Eastern European stations. These are largely the regions in which analysis of E-Profile data leads to a pronounced reduction in the bias (right).

4 Conclusions

The main conclusion is that assimilation of E-Profile attenuated backscatter coefficient, β_{att}, into the IFS-COMPO significantly reduces the bias and RMSE of the model-equivalent vertical profiles when compared to independent, non-assimilated E-Profile observations. The control experiments frequently overestimate β_{att} aloft, and they often misrepresent to overall shape of the vertical profiles, especially during episodic events. This can be strongly alleviated by assimilation of the lidar and ceilometer

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

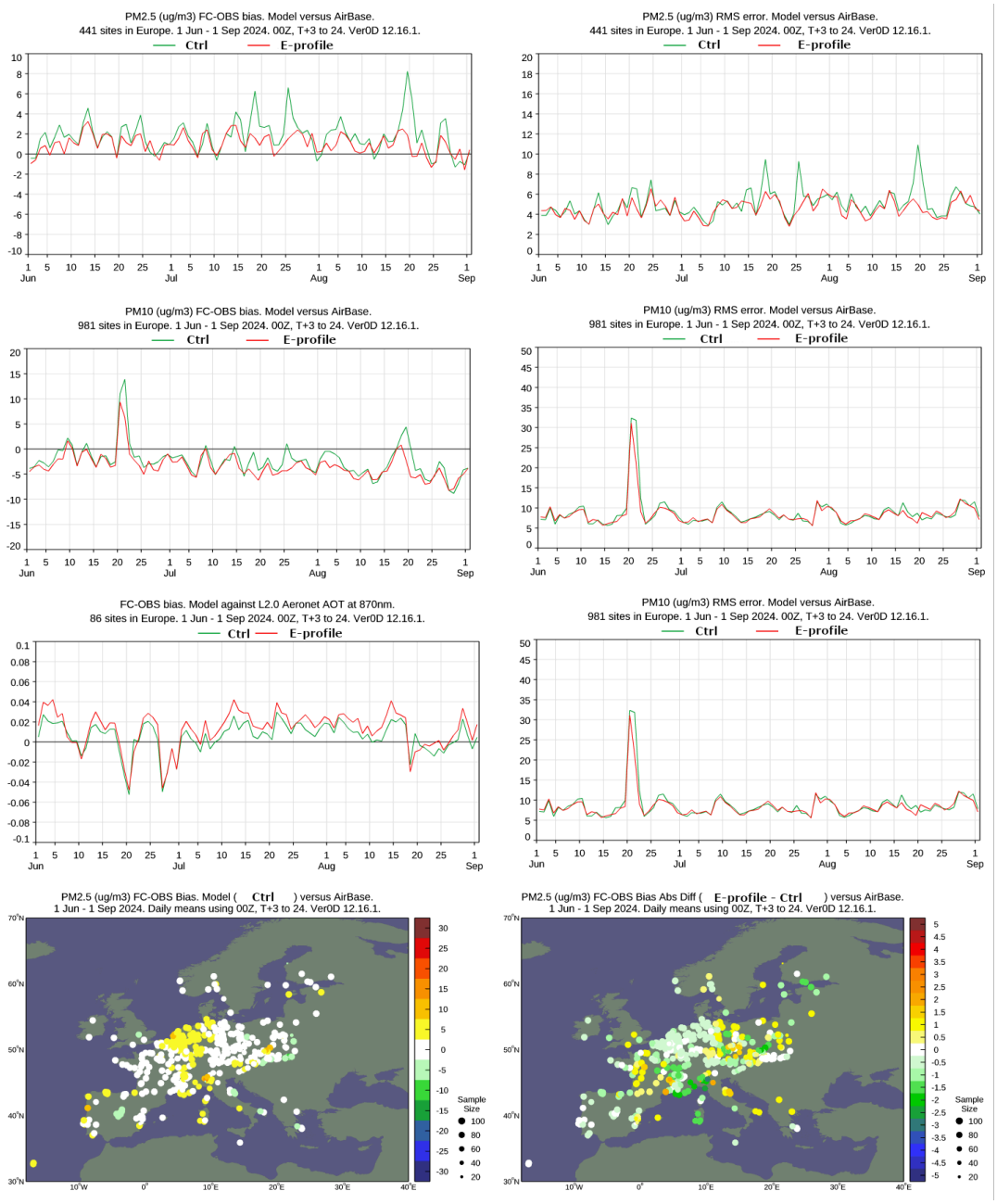

Figure 9. Similar to Fig. 8, but for JJA2024. Top row: bias (left) and RMSE (right) of PM2.5 surface concentrations, averaged over 441 European AirBase stations. Second row: analogous results for PM10, averaged over 981 European AirBase stations. Third row: analogous results for AOD, averaged over 83 European Aeronet stations. Bottom row: PM2.5 bias of the control experiment (left) and bias difference (right) between the control and the E-Profile assimilation experiment.

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

data. This was consistently found for both the summer and autumn seasons by comparison against both CHM15k instruments at 1064 nm wavelength and CL31 and CL51 instruments at 910 nm wavelength.

Different configurations of the analysis system have been tested — see App. D. The results suggest that assimilation of E-Profile observations gives the most favourable results when using an aggressive cloud-masking approach combined with a dust-optics model derived from spheroidal model particles and suppressing vertical correlations in the background error covariances. The effects of E-Profile assimilation become particularly apparent when prioritizing the improvement of vertical profiles and of PM2.5 ground concentrations.

At 3 km altitude CL31 instruments record slightly higher values of β_{att} than CL51 instruments. The CL51 use InGaAs diode lasers with an average pulse power of 27 W, while the CL31 instruments use the same type of laser source, but with an average pulse power of only 11 W. Thus the signal detected by the CL51 instruments is stronger and more reliable; the elevated signal seen by the CL31 instrument around 3 km is probably due to too much noise. In future assimilation experiments the CL31 observations should be capped at 2–2.5 km rather than 3 km, as we did here.

A main limitation of the present analysis system in the IFS-COMPO is the use of a single control variable for aerosols, namely, the total aerosol mass mixing ratio. This makes it impossible to simultaneously reduce the frequent overestimation of PM2.5 concentrations while increasing the often underestimated PM10 concentrations. In the current operational CAMS system, only AOD at a wavelength of 550 nm is assimilated. The limited information contained in these observations would hardly justify to constrain more than a single control variable. However, it is well understood that supplementing extinction with backscattering observations or having multi-channel observations can increase the information contents of observations, thus allowing to constrain more model variables (e.g. Burton et al. (2016); Kahnert and Andersson (2017)).

The need to extend the number of control variables may also arise when assimilating, in addition to AOD 550, lidar data from EarthCARE (Donovan et al., 2024) or multi-channel AOD observations from, e.g., the 3MI instrument (EUMETSAT, 2025). Different approaches to multiple control variables have been pursued (e.g. Kahnert (2009); Li et al. (2013); Liang et al. (2020); Sun et al. (2020)), mostly based on educated guesses on which aerosol control variables may be most suited for an observation system. A more systematic method for selecting optimal control variables based on quantifying the information content of the observations has been proposed (Kahnert, 2018). However, a pragmatic next step beyond the use of the total aerosol mass may simply be to use the fine-mode and coarse-mode mass mixing ratios as control variables.

A main concern for global modelling is the currently limited geographic coverage of the E-Profile network. Extending the E-Profile data with observations from additional regions, such as those provided by NASA's Micro-Pulse Lidar Network (MPLNET) (Welton et al., 2018), would significantly enhance coverage. A main requirement for assimilation in the IFS is the availability of data in near-real time.

. **Code and data availability.** Codes developed at ECMWF are the intellectual property of ECMWF and its member states and can, therefore, not be made publicly available. However, ECMWF member-state weather services and their approved partners will get access granted. Also, access to an open version of the IFS code may be obtained from ECMWF under an OpenIFS license (OpenIFS); this

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

includes cycle CY43R3 IFS(AER), see https://doi.org/10.5194/gmd-15-6221-2022 (Huijnen et al., 2022). More details can be found at https://confluence.ecmwf.int/display/OIFS/About+OpenIFS (ECMWF, 2025). An updated IFS Fortran routine for computing the attenuated backscatter signal at 910 nm can be found under https://doi.org/10.5281/zenodo.18803400 (Kahnert et al., 2026). In the same archive, we have published the python programs for pre-processing E-profile data prior to ingestion into the IFS, and those for post-processing model output for plotting and comparison with observations. E-Profile data can be downloaded from the CEDA archive, https://data.ceda.ac.uk/badc/eprofile/data/daily_files (CEDA Archive, 2018). The reprocessed data from CHM15k and Mini-MPL instruments is publicly available on Zenodo (E. Sauvageat, 2026).

Appendix A: Observation operator for the attenuated backscatter coefficient

The attenuated backscatter coefficient at altitude z is defined by

$$\beta_{\mathrm{att}}(z) = [\beta_{\mathrm{a}}(z) + \beta_{\mathrm{mol}}(z)]e^{-2\tau(z)}, \tag{A1}$$

where $\beta_{\mathrm{a}}(z)$ is the unattenuated backscatter coefficient [m^{-1} sr^{-1}] of aerosols at z. $\tau(z)$ is the optical depth along the optical path from the source to the point of backscattering, which is equal to that along the return path to the detector, thus the factor of 2. The unattenuated molecular backscatter coefficient, β_{mol}, is parametrised as a function of wavelength following Chiriaco et al. (2006). The 16 aerosol species are assumed to be externally mixed, in which case β_{a} can be computed by linear combination of the optical properties of each individual aerosol species appropriately weighted by their mass mixing ratios. More specifically, for a layer l between pressure levels $p(z_l)$ and $p(z_l) + \Delta p(z_l)$,

$$\beta_{\mathrm{a}}(z_l) = \sum_{i=1}^{16} X_i(z_l)\, k_i^{\mathrm{m}}\, \frac{\Delta p(z_l)}{g \Delta z_l} \,/\, S_i. \tag{A2}$$

Here, $X_i(z_l)$ are prognostic variables of the IFS-COMPO, namely the mass mixing ratios [kg (aerosol species i) / kg (air)]. k_i^{m} is the mass extinction cross section [m^2 / kg (aerosol species i)], and S_i is the lidar ratio [sr], which expresses the ratio of the aerosols' extinction cross section to their backscatter cross section. g is the Earth's acceleration, so the term $\Delta p(z_l)/(g\Delta z_l)$ yield the air density at that pressure level. Analogously, the aerosol optical depth τ_{a} of a population of externally mixed aerosols along the optical path from the ground to altitude z_L can be obtained by

$$\tau_{\mathrm{a}}(z_L) = \sum_{i=1}^{16} \sum_{l=137}^{L} X_i(z_l)\, k_i^{\mathrm{m}}\, \frac{\Delta p(z_l)}{g}, \tag{A3}$$

where in the IFS layer $l = 137$ is the one closest to the ground.

The total optical depth in Eq. (A1) consists of contributions from aerosol extinction, Rayleigh scattering, and molecular absorption,

$$\tau(z_L) = \tau_{\mathrm{a}}(z_L) + \tau_{\mathrm{Ray}}(z_L) + \tau_{\mathrm{mol}}(z_L). \tag{A4}$$

The Rayleigh-scattering contribution is parametrised as a function of wavelength according to Bodhaine et al. (1997), while the molecular absorption optical depth, τ_{mol} is obtained by summation over the contributions of each molecular species. Which molecules contribute to absorption depends on wavelength.

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

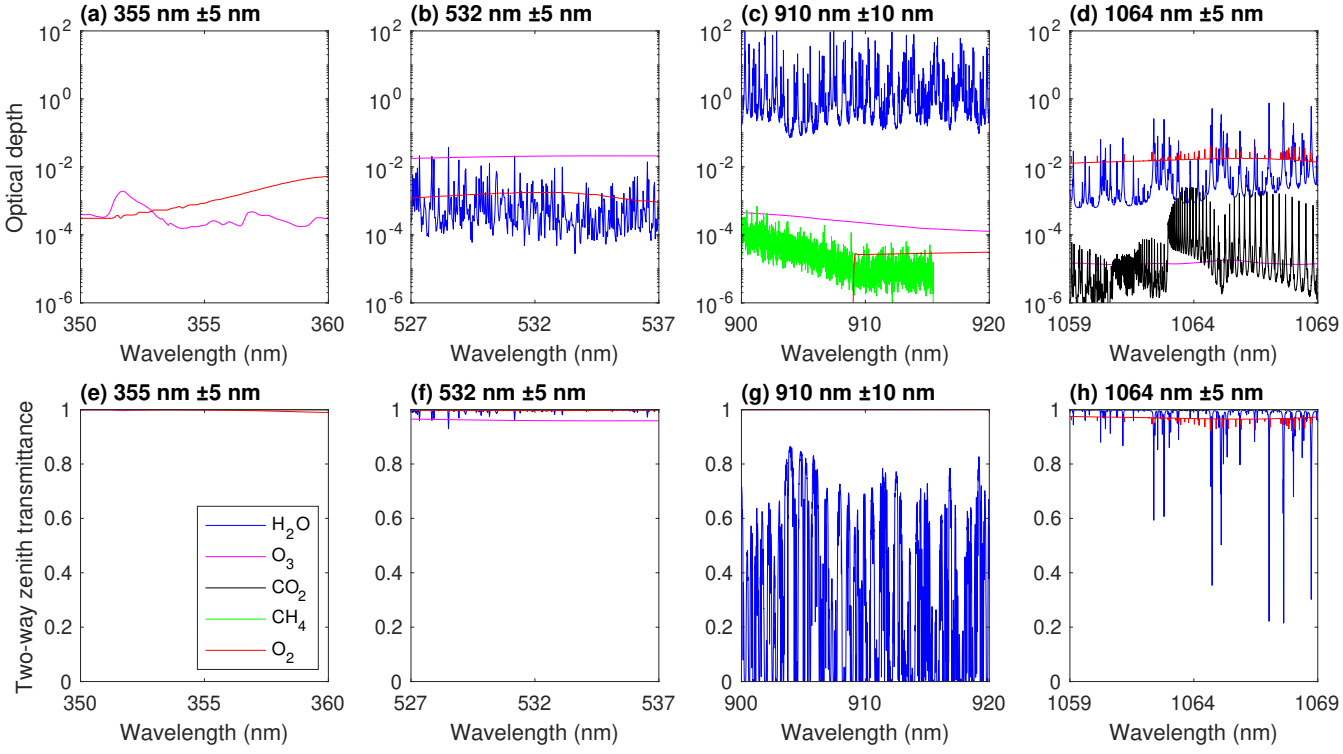

Figure A1. Total column optical depth (top) and two-way transmittance (bottom) of individual gases at 355, 532, 910, and 1064 nm, for the CKDMIP "median" atmosphere (Hogan and Matricardi, 2020). The gas components are water vapour (blue), ozone (magenta), carbon dioxide (black), methane (green), and molecular oxygen (red).

A1 Extension of the observation operator to 910 nm

In CY49R1 observation operators for the attenuated backscatter coefficient based on Eqs. (A1)–(A4) are implemented for three wavelengths, λ=355, 532, and 1064 nm. The code contains parametrisations for $\tau_{\mathrm{mol}}(\lambda)$ that account for O_3 and NO_2 at 355 and 532 nm, and for O_3, NO_2 CO_2, and O_2 at 1064 nm. However, many ceilometer instruments in the E-Profile network measure at a wavelength of 910 nm. Thus the observation operator in the IFS-COMPO needs to be extended to that wavelength. For aerosols, CY49R1 uses a look-up table of optical properties at 331 wavelengths in the range 200 nm – 2 mm. The mass extinction cross section and lidar ratio at 910 nm are determined by linear interpolation between the closest tabulated values (873 and 914 nm). The main challenge is to determine the molecular absorption at 910 nm. To this end, detailed line-by-line radiative transfer computations from the Correlated K-Distribution Model Intercomparison Project, CKDMIP (Hogan and Matricardi, 2020) are analysed. Figure A1 shows the total column optical depth (top) and corresponding two-way transmittance (bottom) for individual gases in the case of the CKDMIP "median" atmosphere. At 1064 nm (panels d and h), molecular absorption is dominated by molecular oxygen (red), while at 532 nm (panels b and f), the dominant absorber

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

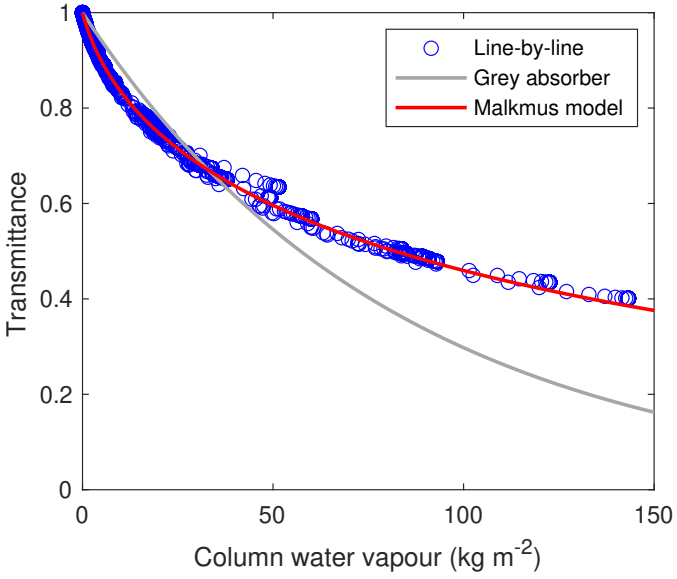

Figure A2. Two-way transmittance as a function of column water vapour obtained by spectrally averaging CKDMIP line-by-line results at 910 nm over a spectral interval ±10 nm (blue circles). Corresponding results are shown for a grey absorber (grey line) and a Malkmus model (red line).

is ozone (magenta). (Note the logarithmic optical-depth scale.) At 355 nm, molecular absorption is, in fact, dwarfed by the effect of Rayleigh scattering. Therefore, we simplified the observation operators by

- neglecting molecular absorption at 355 nm;
- only retaining the contribution of O_3 to the molecular absorption at 532 nm; and
- only retaining the contribution of O_2 to the molecular absorption at 1064 nm.

At 910 nm (panels c and g) the dominant absorber is water vapour (blue), which displays a rich spectral structure and absorbs much more strongly than any gas at the other wavelengths. We seek to parametrise the transmittance averaged across the spectral range corresponding to the band width of a typical 910-nm lidar, which is around ±10 nm for the CL31 and CL51 ceilometers deployed at several E-Profile stations. For each of the 50 diverse atmospheric profiles of the CKDMIP "Evaluation-1" dataset, we compute the spectrally averaged transmittance from the surface up to height z:

$$T_{\mathrm{w}}(z) = \frac{1}{\lambda_2 - \lambda_1} \int_{\lambda_1}^{\lambda_2} \exp\left[-2\tau_{\mathrm{w}}(\lambda, z)\right] d\lambda, \tag{A5}$$

where $\tau_{\mathrm{w}}(\lambda, z)$ is the water-vapour absorption optical depth at wavelength λ from the surface to height z. The blue circles in Fig. A2 show the transmittances computed for $\lambda_1 = 900$ nm and $\lambda_2 = 920$ nm as a function of the column water vapour content $x(z)$ up to the same height. With 50 profiles and each profile being described 55 points in height, we have 2750 points.

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

For a grey absorber, for which the optical depth has no spectral dependence, one would predict a column optical depth of

$$\tau_{\mathrm{w}}(z) = \int_0^z q(z')\rho_{\mathrm{air}}(z')\, k_{\mathrm{w}}^{\mathrm{m}}(z')\,\mathrm{d}z', \tag{A6}$$

where q is the specific humidity, ρ_{air} is the air density, and $k_{\mathrm{w}}^{\mathrm{m}}$ denotes the mass absorption cross section of water vapour (which in principle depends on temperature, pressure, and specific humidity). Applying $k_{\mathrm{w}}^{\mathrm{m}} = 0.015$ m^2 kg^{-1} leads to the grey line in Fig. A2; clearly neglecting the spectral variation of absorption leads to a very poor fit to the line-by-line computations.

Instead, a Malkmus model (e.g. Geleyn et al., 2005) was used: by assuming the spectrum to consist of randomly distributed overlapping lines, the Malkmus model predicts an effective absorption optical depth of water vapour of the form

$$\tau_{\mathrm{eff}}(z) = 2C\left[\sqrt{1 + Kx(z)/C} - 1\right]. \tag{A7}$$

The red line represents the transmittance calculated with C=0.034 and K=0.013509 m^2 kg^{-1}, and is an excellent fit to the line-by-line data.

The new parametrisation for water-vapour absorption as well as the simplifications at the other three lidar wavelengths have been added both to the diagnostic routines and to the lidar observation operator. A corresponding tangent linear (TL) model and adjoint operator (AD) have been implemented for 910 nm, and the consistency of the TL and AD codes have been ensured by use of standardised adjoint tests of the IFS. These changes will be included in CY50R1.

A2 Aerosol optics model for nonspherical dust aerosols

The observation operators of aerosol optical properties, such as AOD or attenuated backscatter coefficient, make use of pre-computed look-up tables of mass extinction cross section, single-scattering albedo, asymmetry parameter, and lidar ratio. In earlier IFS cycles these aerosol optical properties where based on models that assumed all aerosols to be homogeneous spheres. This assumption is known to be erroneous, in particular for desert dust which comes in a wide variety of shapes (e.g. Adebiyi et al., 2023) . As of CY49R1 a correction has been applied to the extinction to account for the nonsphericity of desert dust aerosols (Rémy et al., 2024), which led to an increase in UV and visible dust extinction by 5–20 % for fine and coarse size bins, and by 5–10 % for the super-coarse bin. The nonspherical dust extinction was modelled by assuming a prolate spheroidal shape with an aspect ratio of 1.6, derived from Kandler et al. (2009), which was found to be a shape representative of a majority of particles from the West Sahara. The online tool MOP-SMAP (Gasteiger and Wiegner (2018)) has been used to compute the non-spherical optical properties of desert dust. In CY49R1 IFS-COMPO, this approach has been applied to compute mass extinction, single scattering albedo and asymmetry parameter for the three desert dust bins, at the 20 wavelengths that are produced and archived. Here, this approach has been extended to also compute the lidar ratio S of dust aerosols, so as to take into account nonsphericity as well for this parameter.

Figure A3 shows the spectral dependence of the size-averaged lidar ratio of spheroids normalised by that of volume-equivalent homogeneous spheres. The size averages were taken over the three size bins of dust particles in CY49R1 (as explained in the caption). In the near infrared (NIR), where most E-Profile instruments operate, the lidar ratio of the spheroids

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

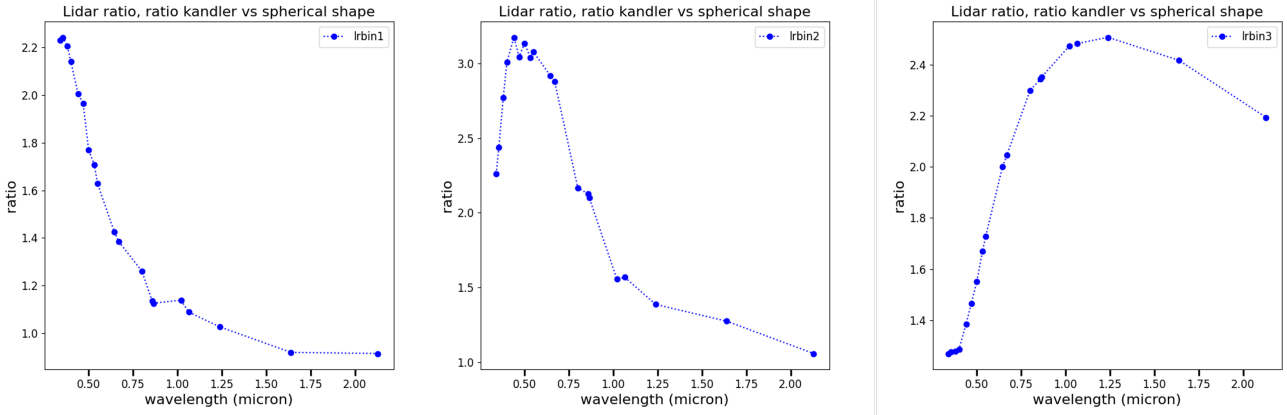

Figure A3. Lidar-ratio of prolate spheroids with an aspect ratio of 1.6 as a function of wavelength, normalised by the lidar ratio of volume-equivalent homogeneous spheres. The three panels show size-averaged results over each of the three size bins of dust particles used in CY49R1, namely, 0.03–0.55 μm radius (left), 0.55–0.9 μm (centre), and 0.9–20 μm (right).

is larger than that of spheres by about 10 % for the fine particles, 50 % for medium-size particles, and by a factor of 2.4 for coarse particles.

Appendix B: Wavelet representation of the background error covariance matrix

The background error covariance matrix ($\mathbf{B}$ in Eq. (1)) is given in wavelet formulation in the IFS (Fisher, 2004, 2006). The wavelet formulation allows for spatial variation of both the geographical and vertical correlations, whilst simultaneously allowing for spectral variations to be represented. In the standard CAMS system, the total aerosol mass mixing ratio background errors represented by the wavelet are calculated using the National Meteorological Center (NMC) method (Parrish and Derber, 1992). Figure B1 shows the resulting globally averaged standard deviation (left) and horizontal correlation (centre) at each of the model levels. The vertical correlations for a single point over Paris, averaged across wavenumbers (right), is shown to demonstrate how the wavelet formulation can spread information between model levels and specifically lead to increments, $\delta\mathbf{x}_a$ (Eq. (3)), being induced without or contrary to the observational information.

An alternative is to create a wavelet representation of the background errors with vertical correlations that are diagonal; i.e. there are no correlations between neighbouring levels. This can be done by setting the horizontal geographical correlations to be isotropic Gaussian functions with a user determined length scale. A vertical standard deviation can then be imposed on these identical horizontal correlations. For the 'artificial' background error wavelet representation used here, the standard deviation from the operational CAMS wavelet was imposed as shown in Fig B1 (left) and a lengthscale of 150km was used for the Gaussian lengthscale of the horizontal correlations (centre). 150km lengthscale was chosen since it reflects a balance between

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

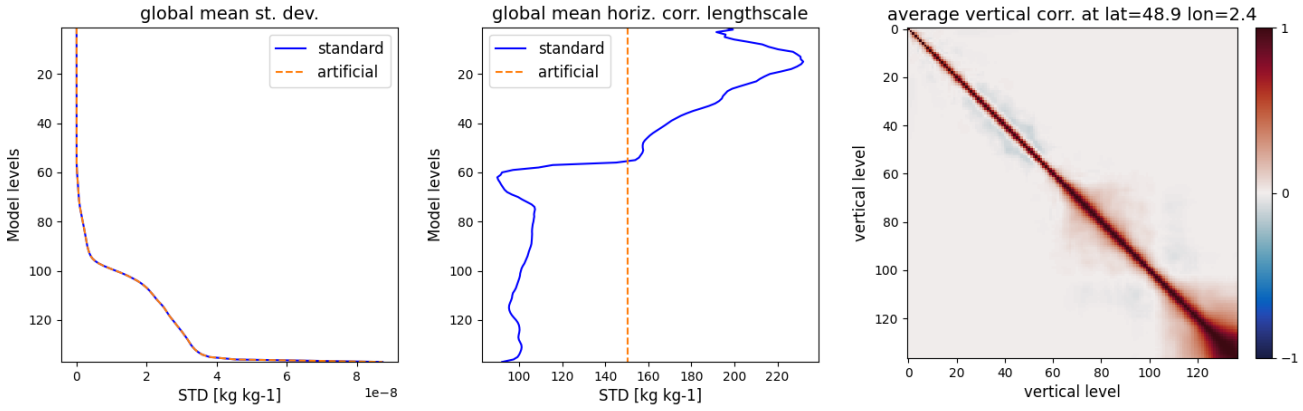

Figure B1. The globally averaged standard deviation at each model level (left), the globally averaged lengthscale of the horizontal correlation (center) and the vertical correlation averaged across wavenumbers for a point centered on Paris (right) from the CAMS standard wavelet representation of the background errors.

the mean lengthscale observed in the operational background errors at the different levels in the atmosphere and the ability to correctly capture the Gaussian correlation at the grid resolution of the minimisation (Inness et al., 2022).

Appendix C: Data preprocessing prior to assimilation

C1 Masking of unwanted data

The masking method has been adapted from Meteo-France. It consists of a truncation of the altitude range, a cloud mask, precipitation mask, fog mask, and noise mask.

Altitude range. The vertical ranges in Tab. 1 mainly reflect the instruments' limitations for cloud-base detection. For the much weaker and, potentially, noisier aerosol signal a more conservative vertical range is adopted, depending on the instrument:

CL31: 0–3 km

CL51: 0–5.3 km

CHM15k: 0–7.5 km

miniMPL: 0.25–15 km

All data outside these ranges are masked out.

Cloud mask. Even though backscattering from clouds is generally stronger than that of aerosol plumes, the signal strengths of clouds and aerosols can significantly overlap. Therefore, clouds and aerosols cannot be reliably discriminated based on the strength of the backscattered power. However, the standard deviation of the attenuated backscatter coefficient, std(β_{att}), has been found to be an effective discriminator. More specifically, an observation of β_{att} at altitude z_l and time step t_i is regarded

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

as a cloud, if

$$\mathrm{std}[\beta_{\mathrm{att}}(t_i, z_l)] > \mathrm{Threshold}, \tag{C1}$$

where the standard deviation is computed from all data in a neighbourhood $t_{i-2} \leq t_i \leq t_{i+2}$ and $z_{l-3} \leq z_l \leq z_{l+3}$. The threshold is taken to be 1e-6 $\mathrm{m}^{-1}\mathrm{sr}^{-1}$ for CHM15k instruments, and 1.5e-6 $\mathrm{m}^{-1}\mathrm{sr}^{-1}$ for all other instruments. Several thresholds have been tested for each instrument to determine the best value for detecting clouds without masking aerosols. To verify the validity of our results, we compared observations marked as clouds with satellite images. We also tested these thresholds in cases of desert dust episodes. Our results showed that using a threshold that is too low leads to an overestimation of clouds, which can mask aerosols. On the contrary, a threshold that is too high can result in the retention of a large number of observations that are in reality clouds.

Precipitation mask. At each time step t_i the number of cloud layers is determined up to a maximum of three layers. If there are cloud-free model layers between two cloudy layers, and if the altitude difference between the lower and the upper cloud layer is $\geq$ 35 m, then the upper layer is counted as a new cloud layer. For each cloud with base altitude $z_c \geq 1$ km, one considers the quantity

$$\overline{\beta}_{\mathrm{att}}(t_i, z_{\min}, z_{\max}) = \underset{z_{\min} \leq z \leq z_{\max}}{\mathrm{mean}} [\beta_{\mathrm{att}}(t_i, z)]. \tag{C2}$$

The precipitation mask is based on evaluating this quantity by taking the mean over all altitudes between $z_{\min} = z_c$-1 km and $z_{\max} = z_c$-0.5 km. If

$$\overline{\beta}_{\mathrm{att}}(t_i, z_c - 1\,\mathrm{km}, z_c - 0.5\,\mathrm{km}) > 2.5\mathrm{e}-6\,\mathrm{m}^{-1}\mathrm{sr}^{-1}, \tag{C3}$$

then all data at time step t_i below the cloud base z_c are masked out as precipitation.

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

$$\overline{\beta}_{\mathrm{att}}(t_i, 0, 250\,\mathrm{m}) > 2.5\mathrm{e}-6\,\mathrm{m}^{-1}\mathrm{sr}^{-1} \tag{C4}$$

$$\overline{\beta}_{\mathrm{att}}(t_i, 0, 250\,\mathrm{m}) > 2\overline{\beta}_{\mathrm{att}}(t_i, 250\,\mathrm{m}, 500\,\mathrm{m}). \tag{C5}$$

Noise mask. Data points are masked out as noise if the following criterion is satisfied:

$$\frac{\mathrm{mean}[\beta_{\mathrm{att}}[(t_i, z_l)]}{\mathrm{std}[\beta_{\mathrm{att}}[(t_i, z_l)]} < 3, \tag{C6}$$

where the mean and standard deviation are evaluated in a neighbourhood $t_{i-2} \leq t_i \leq t_{i+2}$, $z_{l-2} \leq z_l \leq z_{l+2}$. The data point is also removed if the number of unmasked data points in that neighbourhood is less than 13.

Figure C1 shows the attenuated backscatter coefficient observed at the lidar station in Toulouse on 17 September 2023. In the E-Profile data files up to three cloud layers are identified, which are represented in the top panel by dashed lines. The other panels illustrate the masks (grey shading) for clouds (second panel from top), precipitation (third panel), fog (fourth panel), and noise (bottom panel). The clouds identified by the mask closely agree with the cloud-base heights identified in the E-Profile data files. The precipitation mask identifies several rainy episodes during that day, while no fog has been detected. Most data points above the clouds are rejected by the noise mask.

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

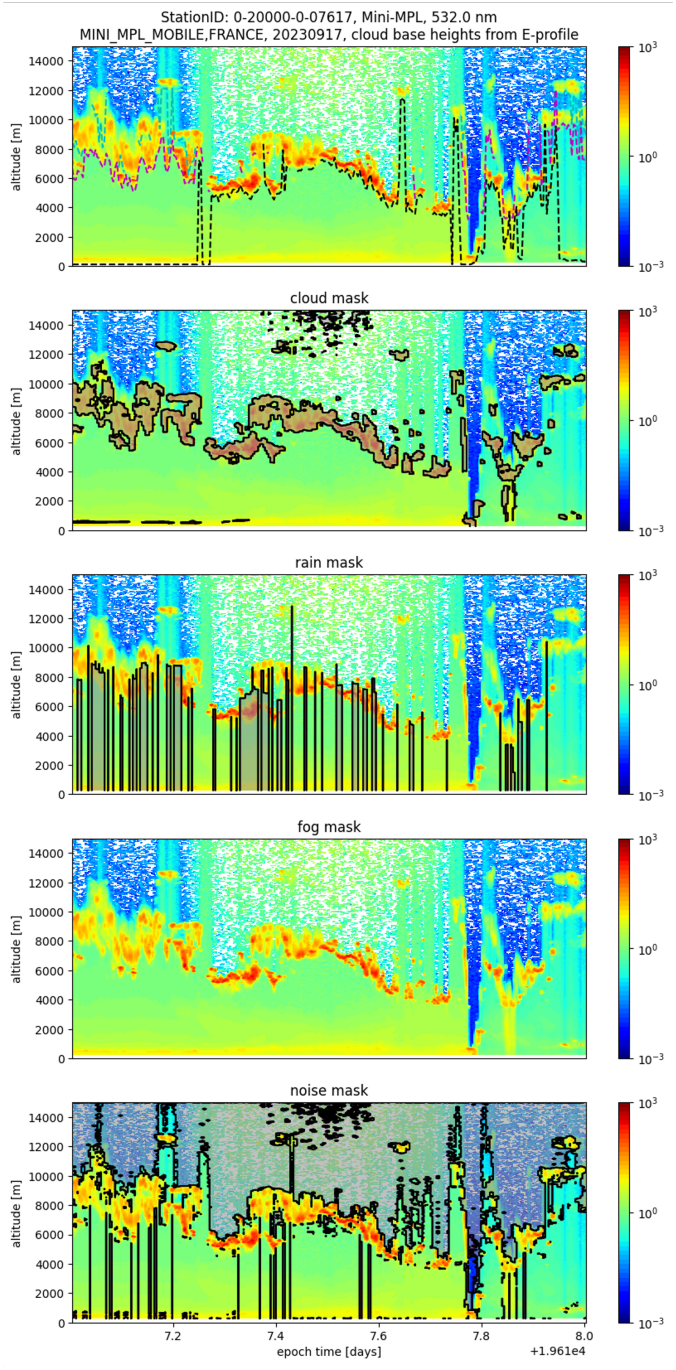

Figure C1. Attenuated backscatter coefficient at 532 nm wavelength, observed on 17 September 2023 by the Mini-MPL lidar instrument in Toulouse, France. The top panel shows the cloud-base heights (dashed lines) provided in the E-Profile data files. Second panel from top: cloud mask; third panel: precipitation mask; fourth panel: fog mask; bottom panel: noise mask.

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

C2 Combining masking and data-reduction operations

Pre-averaging over time and altitude. To reduce the data volume prior to assimilation, both averaging data and thinning data by selection have been considered. Averaging can reduce noise in the data, while thinning by selection is somewhat faster. Since the extra time required for averaging was not found to be substantial, it was decided to reduce the amount of data by averaging over 30-minute time intervals and over appropriate altitude intervals. For the latter the geometric altitudes of the 137 IFS model levels of a US standard atmosphere have been employed. If more than 90 % of the data points in an averaging interval where flagged as missing values, then the average was not computed; instead, the low-resolution data point was also flagged as a missing value.

Ordering of masking and resolution-reduction. Data masking and resolution reduction are non-commuting operations. Therefore, it is important to determine in which order these operation have to be applied. This is illustrated in Fig. C2. The top panels show observations by the Mini-MPL instrument in Toulouse on a partially cloudy day (16 September 2023, left), and during a Sahara dust-storm event (4 September 2023, right). The second row shows the effect of applying the cloud/rain/fog/noise mask first and time/altitude averaging second. Only few data points are identified as clouds, noise, or precipitation (left), while most of the dust plume at 1-2 km altitude is identified as valid data (right). The third panel shows the effect of applying time/altitude averaging first and the cloud/rain/fog/noise mask second. This results in a much more aggressive cloud mask (left), which may be desirable if one wants to be sure not to assimilate any cloud-contaminated data. However, this approach is so aggressive that even the dust plume is misclassified as a cloud (right). To obtain a reasonably aggressive cloud mask that does not misidentify optically thick aerosol plumes as clouds, the following approach has been tested. The masking operations are split up; the cloud/rain/fog mask is applied first, time/altitude averaging is applied second, and the noise mask is applied last. The effect of this approach is seen in the bottom row. Clouds and noise are securely masked out (left), while the dust plume passes the mask unharmed (right). For brevity, we refer to the combined mask/averaging operation illustrated in the second row of Fig. C2 as the *conservative mask*, and to that shown in the fourth row as the *aggressive mask*. Symbolically,

$$(\text{conservative mask}) \;=\; (\text{averaging}) \circ (\text{noise mask}) \circ (\text{fog mask}) \circ (\text{rain mask}) \circ (\text{cloud mask}) \tag{C7}$$

$$(\text{aggressive mask}) \;=\; (\text{noise mask}) \circ (\text{averaging}) \circ (\text{fog mask}) \circ (\text{rain mask}) \circ (\text{cloud mask}), \tag{C8}$$

where the operators are applied in the order from right to left. In the analysis experiments performed in this study, we tested both approaches.

C3 Miscellaneous preprocessing steps

Unit conversion. The E-Profile data files provide the attenuated backscatter coefficient in units of $\mathrm{Mm^{-1}sr^{-1}}$ and time as epoch date (i.e., number of days) since 1 January 1970. To conform to IFS standards these have to be converted into $\mathrm{m^{-1}sr^{-1}}$ and date-time (YYYYMMDD:HHMMSS), respectively.

Reordering of bottom-up to top-down vertical profiles. The E-Profile data are ordered from bottom to top in altitude. However, the IFS requires profiles in top-down format. It is essential to reorder the data accordingly prior to assimilation.

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

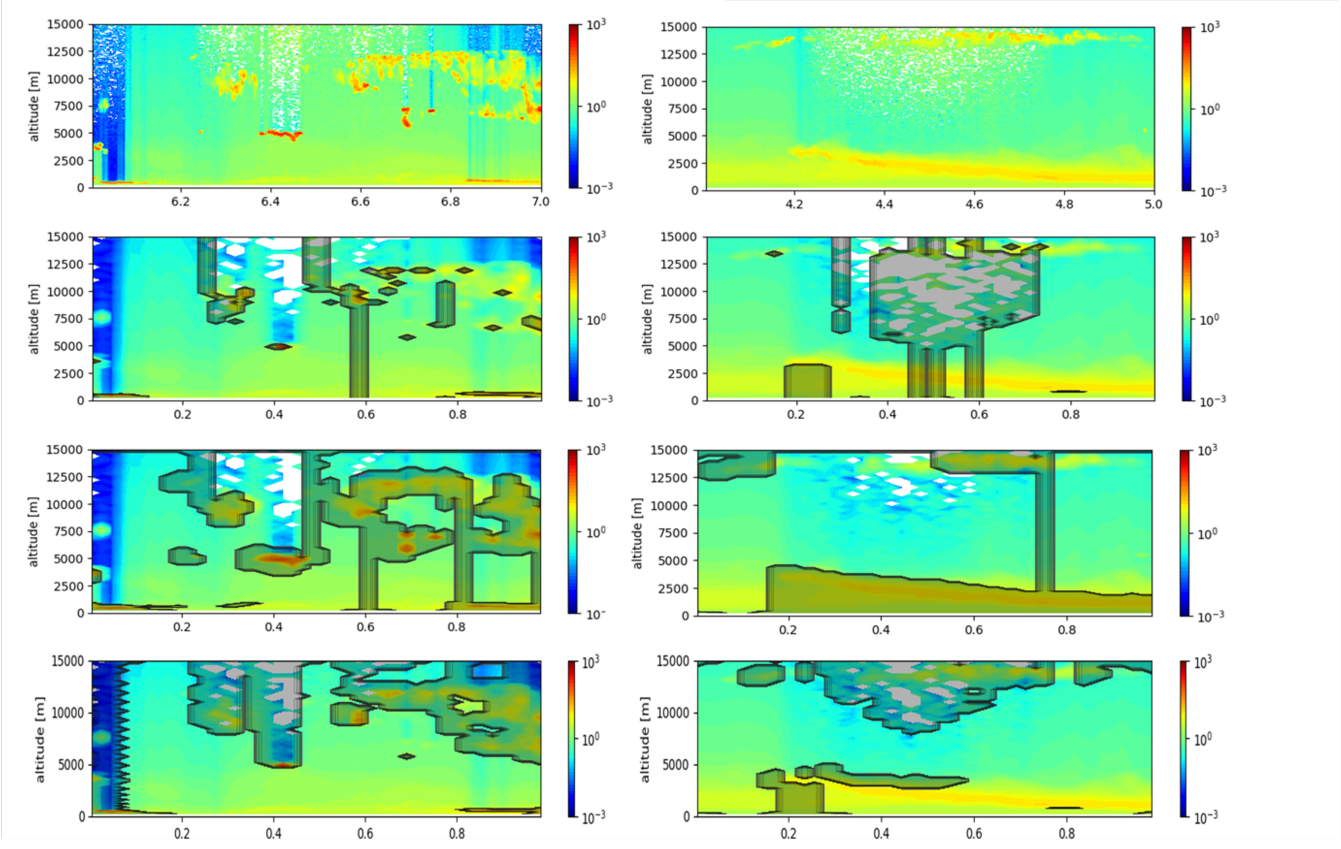

Figure C2. Observation from the same station as in Fig. C1, but on 16 September 2023 (top left) and 4 September 2023 (top right). The other rows illustrate the effect of applying masking operations and time/altitude averaging in different orders; second row: (averaging) ∘ (noise mask) ∘ (fog mask) ∘ (rain mask) ∘ (cloud mask) (where the rightmost operation is applied first); third row: (noise mask) ∘ (fog mask)∘(rain mask)∘(cloud mask)∘(averaging); fourth row: (noise mask)∘(averaging)∘(fog mask)∘(rain mask)∘(cloud mask).

Otherwise, the IFS will misinterpret the vertical structure, placing boundary-layer aerosols on top of the atmosphere, resulting in significant analysis errors. To prevent such errors, a validation step has been implemented in the lidar observation operator to ensure that all input data comply with the required top-down format. This extra safety measure will be included in CY50R1.

Pooling of the data and export in ODB format. New experimental observations can be ingested by the IFS by use of the *simulobs2odb* routine. This requires that large data files are split into pools and converted into ascii files following the Observational Data Base (ODB) format. Here, the maximum number of lines per pool was set to 12,000, and the preprocessed data for each day and analysis time-step where split into pools and saved as ODB-like ascii files.

Observation errors. The E-Profile data files come with a rough estimate for the observation error standard deviation $\sigma_o(t_i, z_l)$ for each data entry $\beta_{\mathrm{att}}(t_i, z_l)$. In the analysis experiments reported here these values are basically adopted, but we assume that the standard deviation is never smaller than 1e-6 $\mathrm{m}^{-1}\mathrm{sr}^{-1}$. Thus, a standard deviation is assumed according

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

to $\sigma'_o(t_i, z_l) = \max[\sigma_o(t_i, z_l), 10^{-6}\mathrm{m}^{-1}\mathrm{sr}^{-1}]$. This is based on earlier tests of assimilating Earlinet data into the IFS. If σ_o becomes much smaller than that threshold, then the algorithm for minimising the cost function in Eq. (1) can sometimes fail to converge. It is further assumed that the observation error covariance matrix in Eq. (1) is diagonal, i.e. $\mathbf{R} = \mathrm{diag}\{\sigma'^2_o(t_i, z_l)\}$.

Appendix D: Comparing different configurations of the assimilation system

For the assimilation of E-Profile data, different assumptions in the assimilation system were tested:

- Two different dust-optics models were employed (Sec. A2), one in which the lidar ratio was simulated by use of homogeneous spheres, the other in which it was modelled by use of homogeneous prolate spheroids with an aspect ratio of 1.6.
- Two sets of background error covariances were used (Sec. B), the standard wavelets from CY49R1, and artificial wavelets in which vertical correlations were suppressed.
- The data were preprocessed with two types of masks, the conservative mask, Eq. (C7), and the aggressive mask, Eq. (C8) (where *conservative* and *aggressive* refers to the masking effect on clouds).
- The experiments were run for two 3-month periods, SON 2023 (2023-09-01 – 2023-11-30) and JJA 2024 (2024-06-01 – 2024-08-31). For each period a 1-month spinup was used.

Combining these four, we have $2^4 = 16$ different experiments, and four control experiment without E-Profile assimilation were performed (two seasons and two sets of wavelets), resulting in a total of 20 experiments.

D1 Impact of dust-optics model on assimilation results

Figure D1 shows the same observations (1st column) and control experiment (2nd column) as in Fig. 4; however, the 3rd and 4th columns present a comparison of the E-Profile assimilation experiment using dust optical properties derived from spheres and spheroids, respectively. In both configurations the high positive biases observed in the control run as well as the RMSE are reduced by the assimilation of E-Profile data. However, the spheroid-based approach outperforms the sphere-based method. This effect is more pronounced for the RMSE than for the bias. This finding held for both seasons under study and was evident even in the CL31 and CL51 measurements (not shown).

D2 Impact of masking approach

The results in Fig. D1 were based on preprocessing the observations with the aggressive cloud-masking approach according to Eq. (C8). Figure D2 presents analogous results obtained by preprocessing the observations with the conservative cloud masking method in Eq. (C7). As in the previous examples, the analyses experiments, especially the one with the spheroidal dust optics, significantly reduce both the bias and the RMSE, resulting in a much improved vertical distribution of the model-equivalent backscatter signal. However, the observations (top left) reveal a few episodes of high backscatter signals at altitudes of 6–7

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

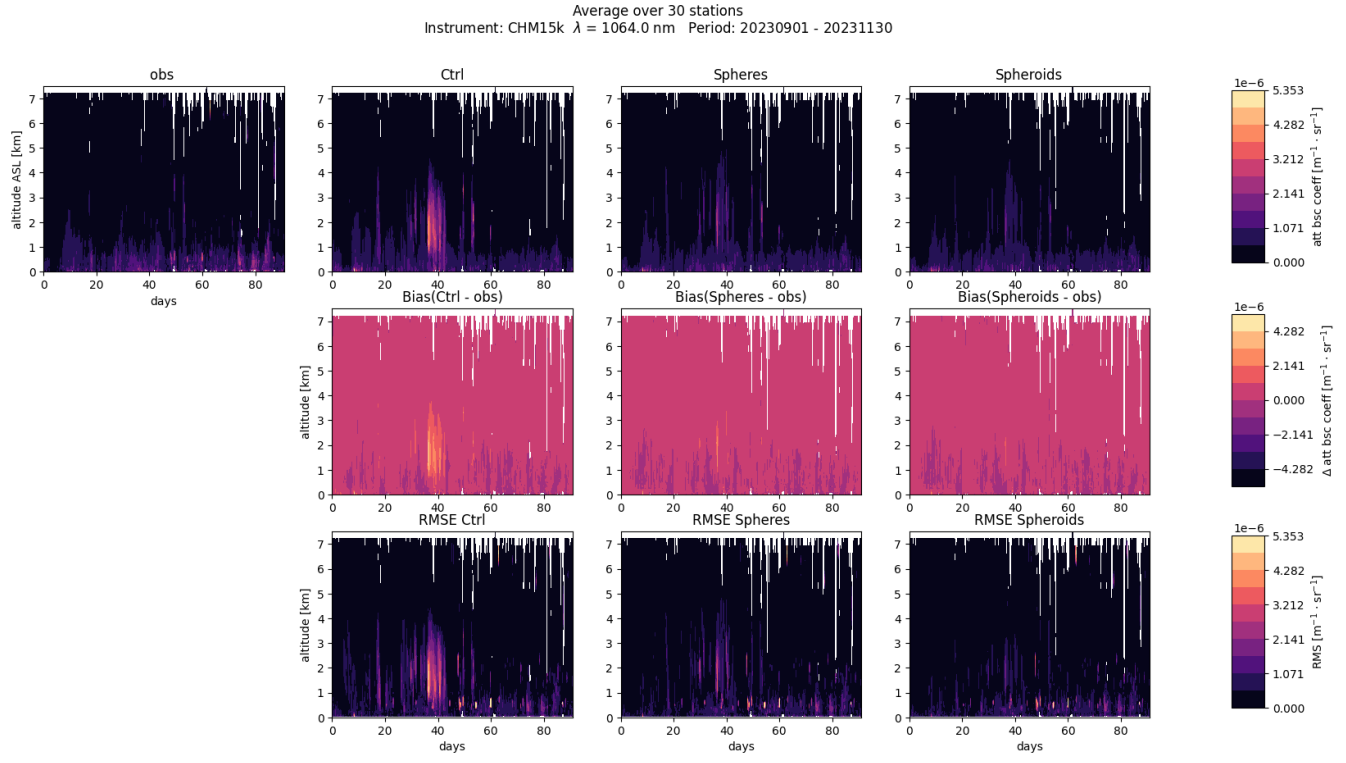

Figure D1. Attenuated backscatter coefficient at 1064 nm wavelength during SON 2023, averaged over 30 non-assimilated CHM15k stations. Top row (from left to right): observations, control experiment, and E-Profile analysis experiments with spherical and spheroidal model particles for dust optical properties. Second row: bias (model - observations). Third row: root mean square error.

km, for example around days 45 and 76. These signals are likely to originate from clouds that the conservative mask has not captured. Another striking difference to Fig. D1 is that data above 5.5 km altitude are essentially missing. The reason is that in this approach noise masking is applied prior altitude/time averaging — see Eq. (C7). Thus, much of the noisier data at higher altitudes are rejected by the mask. In the aggressive cloud mask averaging is applied prior to noise masking — see Eq. (C8). But averaging reduces noise, which lets more data points at higher altitudes pass the noise mask. Therefore, the aggressive cloud mask acts, in fact, more conservatively on noise. Aggressive cloud masking and conservative noise masking are both desirable properties.

D3 Impact of vertical correlation in the background error covariances

The effect of vertical correlations in the wavelets is best illustrated by inspecting individual profiles. Figure D3 shows two selected profiles from the dust-storm event in Fig. 2 (left) and the wild-fire episode in Fig. 3 (right). There are now two control experiments, one using standard CY49R1 wavelets (light blue, dash-dotted line), the other using artificial wavelets with zero vertical correlations (orange dashed). The corresponding E-Profile assimilation experiments with standard (yellow solid) and

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

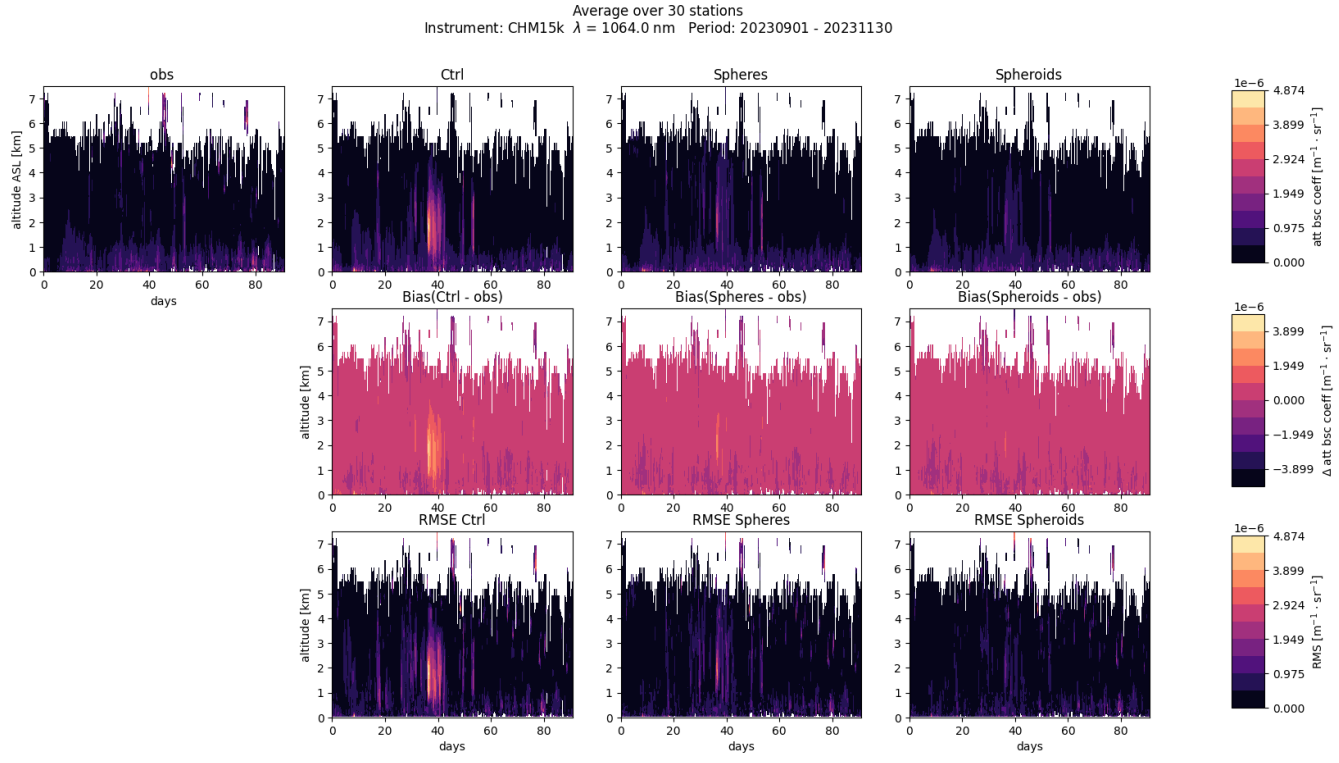

Figure D2. As Fig. D1, but preprocessing the observations with the conservative cloud mask in Eq. (C7).

artificial wavelets (dotted green) are also shown; both are based on using the aggressive cloud mask and spheroidal dust optics. In the dust-storm case (left) the control experiments give similar results. Both ceilometer assimilation runs substantially improve the vertical profiles in comparison to the observations (black). Above 3.6 km altitude the observations are flagged as too noisy. At those altitudes the assimilation experiment based on artificial wavelets (orange dashed) closely matches the control experiment. By contrast, the one based on standard wavelets (yellow solid) displays higher backscatter values in that altitude range, presumably due to stronger vertical spreading of information from lower altitudes. However, this effect is not very strong. In the wild-fire case (right) the two control experiments differ substantially. But the corresponding E-Profile assimilation experiments match the observations about equally well. This has been found to be rather typical. However, as we shall see in the next section, the two sets of wavelets can give very different results for surface concentrations of particulate matter.

D4 Comparison of all experiments

A concise summary of the performance of all 20 experiments is given in Figure D4, showing biases (left) and RMSE (right) for PM2.5 (top), PM10 (centre), and AOD (bottom). The biases and RMSEs are averaged over all validation stations and over the entire season. The columns in each matrix show results for (from left to right) 1. the control run, the aggressive mask with

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

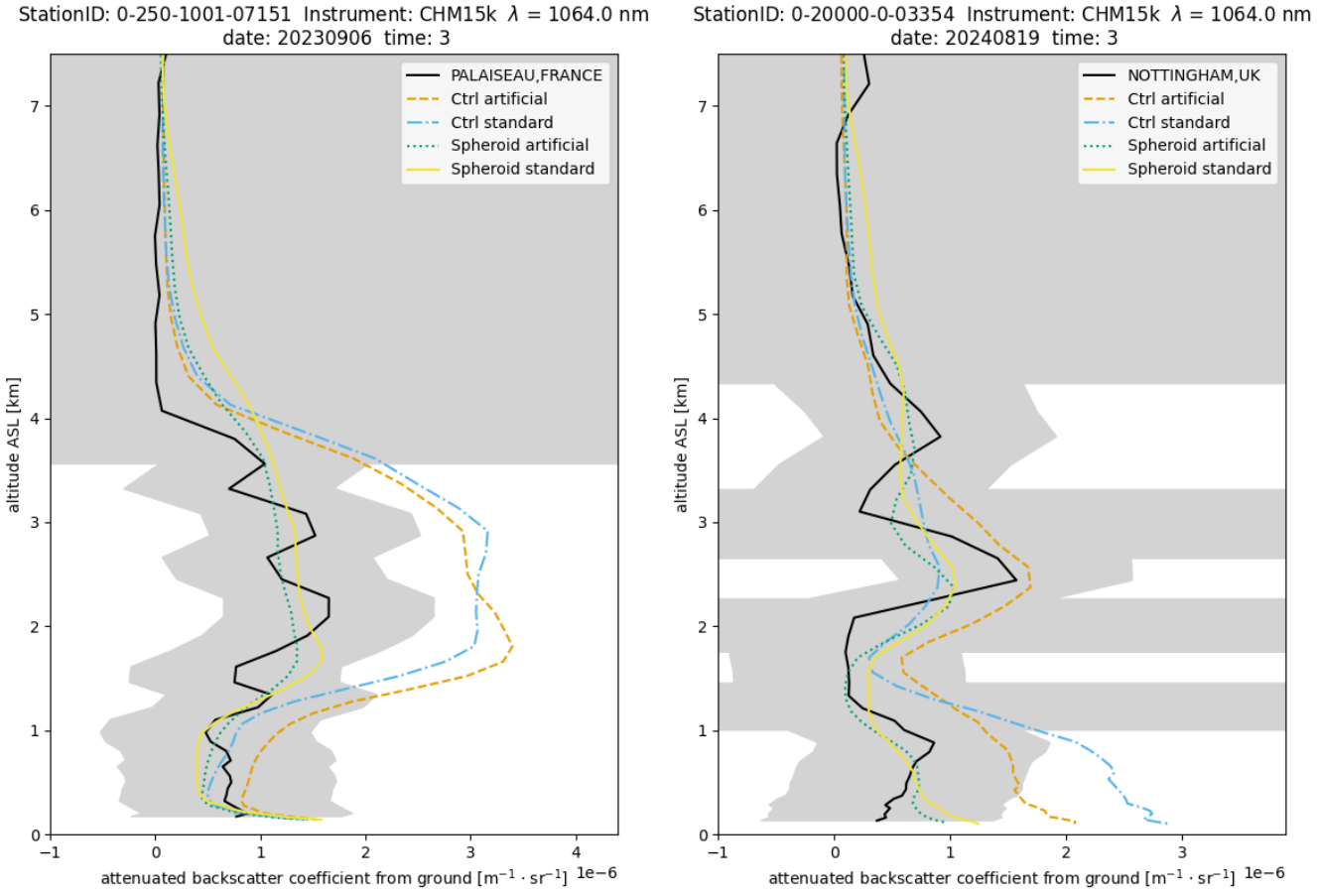

Figure D3. Vertical profiles of β_{att} in Palaiseau, FR, during the dust-storm in Fig. 2 (left), and in Nottingham, UK, during the wild-fire in Fig. 3. The lines show the control runs with standard (light blue, dash-dotted line) and artificial wavelets (orange dashed), and the corresponding ceilometer assimilation experiments for standard (yellow solid) and artificial wavelets (greed dotted), respectively.

2. spherical and 3. spheroidal dust optics, and the conservative mask with 4. spherical and 5. spheroidal dust optics. The rows present results for (from top to bottom) JJA 2024 with 1. artificial and 2. standard wavelets, and for SON 2023 with 3. artificial and 4. standard wavelets.

For PM2.5, biases are generally higher in the summer than in the autumn period. The artificial wavelets consistently yield lower biases than the standard wavelets, even for the control experiments. The aggressive and conservative masking approaches do not result in significantly different biases, while the experiments based on spherical dust optics yield lower biases than those based on spheroidal dust optics. The PM2.5 RMSE in the control runs is higher in the summer than in the autumn period, but the E-Profile analysis experiments reduce the RMSE very effectively in the summer and makes it lower than the corresponding RMSE values in the autumn period. The RMSE values of the aggressive masking methods are on the same order as those of the conservative masking method. Spherical dust optics and artificial wavelets yield the lowest RMSE values.

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

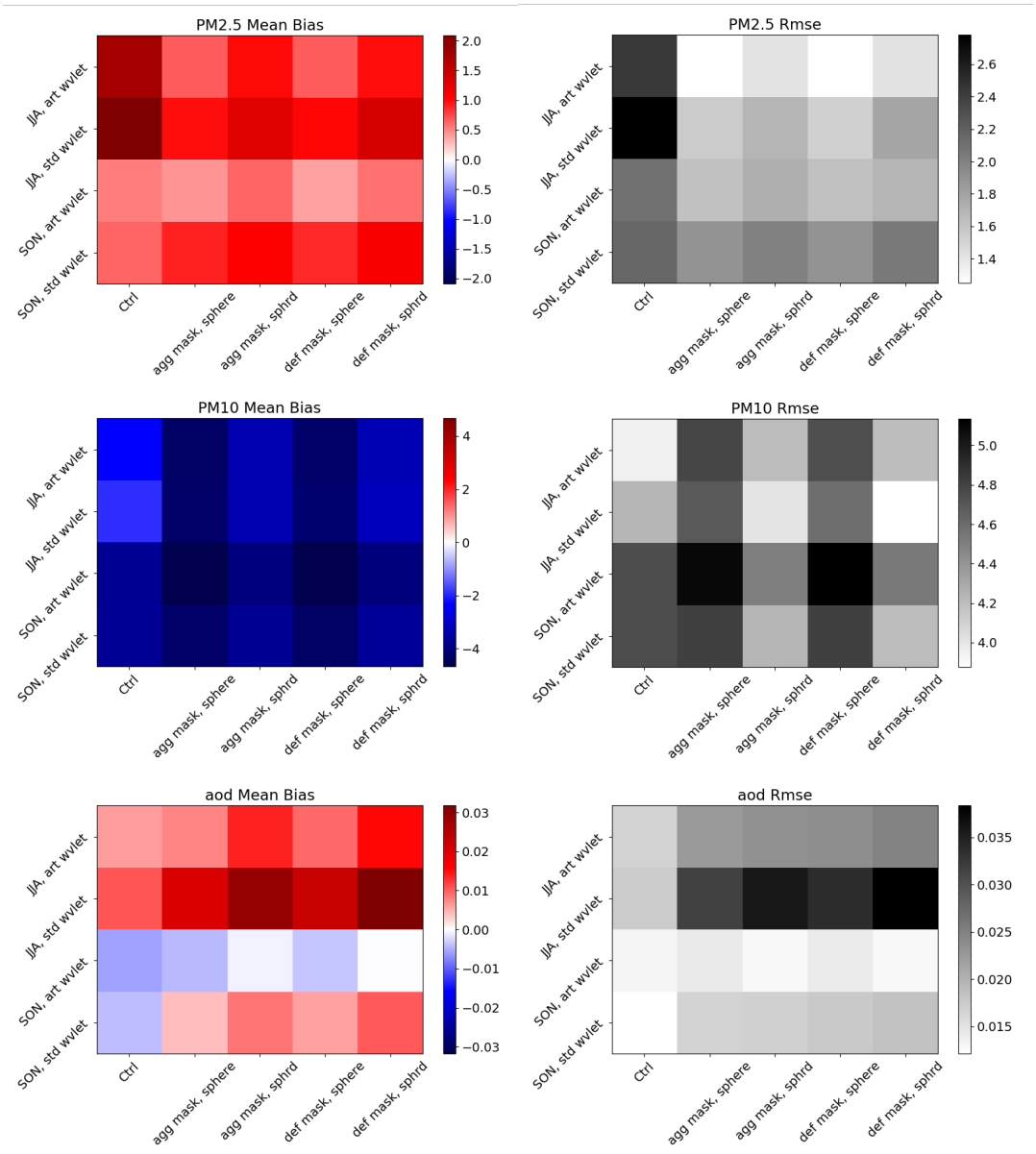

Figure D4. Bias (left) and RMSE (right) for PM2.5 (top), PM10 (centre), and AOD (bottom) averaged over all validation stations and over 3-month periods. In each matrix the control experiments are shown in the first column; the other columns show E-Profile analysis experiments using the aggressive cloud mask with spherical (2nd column) and spheroidal dust optics (3rd column), and the conservative cloud mask with spherical (4th column) and spheroidal dust optics (5th column). The rows show from top to bottom experiments for JJA 2024 using artificial (1st row) and standard wavelets (2nd row), and SON 2023 using artificial (3rd row) and standard wavelets (4th row).

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

PM10 (centre left) is underestimated by all experiments, but less so in the summer period. Spheroidal dust optics lead to slightly lower absolute biases, and in many cases to significant reductions of the RMSE, while spherical dust optics degrades the RMSE. A possible explanation is that the optical properties of coarse dust particles are particularly sensitive to shape. Therefore, while the spheroid model gives less improvements for fine particles compared to spheres, there is a much more pronounced effect on RMSE of PM10.

For AOD one would usually expect that the control experiments yield the lowest bias and RMSE for reasons already mentioned above. This is confirmed in most experiments, except for the autumn period run with artificial wavelets. In those experiments, assimilation of E-Profile data reduces the bias, especially in the experiments based on using spheroidal dust optics.

D5 Identifying an optimum configuration of the analysis system

Dust-optics model. Two different aerosol optics models were tested for representing the lidar ratio of dust particles, one based on spheres, the other based on spheroids. The reduction of biases and RMSE in the vertical profiles are more pronounced for the spheroid model than for the sphere model. For surface concentrations, spheroids reduce, in most cases, the RMSE for PM10, while spheres result in a degradation. For PM2.5 both spheres and spheroids mostly give a reduction of biases and RMSE, where spheres slightly outperform spheroids. All things considered, the spheroid model seems to be slightly superior to the sphere model, especially for improving vertical profiles.

Masking method. The masking method that has a more conservative effect on clouds appears to occasionally miss high-altitude clouds above 6 km. Further, for ceilometers observing at 1064 nm it rejects most data above 5.5 km as too noisy. Both properties are undesirable. The more aggressive cloud mask securely masks out the high-altitude clouds. Since in this approach time/altitude averaging is performed prior to noise masking, it retains a lot more data above 5.5 km. One can conclude that the aggressive cloud mask defined in Eq. (C8) proved to be superior to the conservative mask in Eq. (C7), at least for masking data in the upper troposphere. No significant differences were observed between these two masking approaches when evaluating their effect on surface concentrations in the E-Profile assimilation experiments.

Vertical correlations. Two sets of wavelets have been tested, the standard wavelets in CY49R1, which contain vertical correlations, and an artificial set of wavelets in which vertical correlations had been suppressed. The latter gives the analysis more leeway to adjust the vertical profiles. This has remarkably little impact on model-equivalent backscatter signals aloft; but it can have a significant effect on PM2.5 surface concentrations. During both seasons, the assimilation experiments with artificial wavelets resulted in lower biases and lower RMSE compared to the standard wavelets. A similar improvement can be observed for AOD.

Based on these findings, the main body of the paper presented E-Profile assimilation experiments with suppressed vertical correlations in the wavelets, dust optical properties derived from spheroids, and with observations preprocessed by the aggressive cloud-masking approach,

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

. **Author Contribution:** MK implemented the 910 nm observation operator, developed and applied tools for E-Profile data preprocessing, designed the experiments, and drafted the paper with significant contributions from all authors. MB developed an algorithm for masking E-Profile data, which was taken up and improved by MK. VG contributed with conceptualisation and support in the development of the masking method for E-profile data. ES manages the E-Profile network and provided the reprocessed CHM15k data. MA conceptualised the assimilation study and supplied the artificial wavelet files and helped with setting up the experiments. SR provided the asphericity information for the lidar ratio in the optics file used in IFS-COMPO. RH developed the parametrisation for the 910 nm extinction by water vapour. AH manages the E-profile network and contributed to the provision of the data. JF contributed to the conceptualisation of the study and design of experiments.

. **Competing Interests:** At least one of the (co-)authors is a member of the editorial board of Geoscientific Model Development. The peer-review process is guided by an independent editor, and the authors also have no other competing interests to declare.

. **Acknowledgements.** MK acknowledges funding by the European Horizon project CAMAERA. A considerable number of datasets have been used in this assimilation study. The authors are grateful to all involved in creating and publicly supplying the following data sets: EUMETNET (E-Profile), EUMETSAT (PMAp), the European Environment Agency (AirBase), and NASA (MODIS, VIIRS, AERONET).

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

References

Adebiyi, A., Kok, J. F., Murray, B. J., Ryder, C. L., Stuut, J.-B. W., Kahn, R. A., Knippertz, P., Formenti, P., Mahowald, N. M., Pérez García-Pando, C., Klose, M., Ansmann, A., Samset, B. H., Ito, A., Balkanski, Y., Di Biagio, C., Romanias, M. N., Huang, Y., and Meng, J.: A review of coarse mineral dust in the Earth system, Aeolian Research, 60, 100 849, https://doi.org/https://doi.org/10.1016/j.aeolia.2022.100849, 2023.

Benedetti, A., Morcrette, M. J.-J., Boucher, O., Dethof, A., Engelen, R. J., Huneeus, M. F. H. F. N., Jones, L., Kaiser, J. W., Kinne, S., Mangold, A., Razinger, M., Simmons, A. J., and Suttie, M.: Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: 2. Data assimilation, J. Geophys. Res., 114, D13 205, 2009.

Bodhaine, B. A., Wood, N. B., Dutton, E. G., and Slusser, J. R.: On Rayleigh Optical Depth Calculations, J. Atmos. Ocean. Technol., 16, 1854–1861, 1997.

Burton, S. P., Chemyakin, E., Liu, X., Knobelspiesse, K., Stamnes, S., Sawamura, P., Moore, R. H., Hostetler, C. A., and Ferrare, R. A.: Information content and sensitivity of the $3\beta+2\alpha$ lidar measurement system for aerosol microphysical retrievals, Atmos. Meas. Techniques, 9, 5555–5574, 2016.

CEDA Archive: EUMETNET E-PROFILE: ceilometer cloud base height and aerosol profile data from a network covering most of Europe with additional sites worldwide, https://data.ceda.ac.uk/badc/eprofile/data/daily_files, 2018.

Chabrillat, S., Rémy, S., Huijnen, V., Bingen, C., Debosscher, J., Errera, Q., Metzger, S., Minganti, D., Opdebeek, M., Williams, J., et al.: Modelling stratospheric composition for the Copernicus Atmosphere Monitoring Service Cy49R1: polar ozone depletion and sulfate aerosols, Tech. rep., Copernicus Meetings, 2025.

Chiriaco, M., Vautard, R., Chepfer, H., Haeffelin, M., Dudhia, J., Wanherdrick, Y., Morille, Y., and Protat, A.: The Ability of MM5 to Simulate Ice Clouds: Systematic Comparison between Simulated and Measured Fluxes and Lidar/Radar Profiles at the SIRTA Atmospheric Observatory, Mon. Weather Rev., 134, 897–918, 2006.

Colette, A., Collin, G., Besson, F., Blot, E., Guidard, V., Meleux, F., Royer, A., Petiot, V., Miller, C., Fermond, O., Jeant, A., Adani, M., Arteta, J., Benedictow, A., Bergström, R., Bowdalo, D., Brandt, J., Briganti, G., Carvalho, A. C., Christensen, J. H., Couvidat, F., D'Elia, I., D'Isidoro, M., Denier van der Gon, H., Descombes, G., Di Tomaso, E., Douros, J., Escribano, J., Eskes, H., Fagerli, H., Fatahi, Y., Flemming, J., Friese, E., Frohn, L., Gauss, M., Geels, C., Guarnieri, G., Guevara, M., Guion, A., Guth, J., Hänninen, R., Hansen, K., Im, U., Janssen, R., Jeoffrion, M., Joly, M., Jones, L., Jorba, O., Kadantsev, E., Kahnert, M., Kaminski, J. W., Kouznetsov, R., Kranenburg, R., Kuenen, J., Lange, A. C., Langner, J., Lannuque, V., Macchia, F., Manders, A., Mircea, M., Nyiri, A., Olid, M., Pérez García-Pando, C., Palamarchuk, Y., Piersanti, A., Raux, B., Razinger, M., Robertson, L., Segers, A., Schaap, M., Siljamo, P., Simpson, D., Sofiev, M., Stangel, A., Struzewska, J., Tena, C., Timmermans, R., Tsikerdekis, T., Tsyro, S., Tyuryakov, S., Ung, A., Uppstu, A., Valdebenito, A., van Velthoven, P., Vitali, L., Ye, Z., Peuch, V.-H., and Rouïl, L.: Copernicus Atmosphere Monitoring Service – Regional Air Quality Production System v1.0, Geoscientific Model Development, 18, 6835–6883, https://doi.org/10.5194/gmd-18-6835-2025, 2025.

Courtier, P., Thépaut, J.-N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var using an incremental approach, Q. J. Roy. Meteorol. Soc., 120, 1367–1388, 1994.

Donovan, D. P., van Zadelhoff, G.-J., and Wang, P.: The EarthCARE lidar cloud and aerosol profile processor (A-PRO): the A-AER, A-EBD, A-TC, and A-ICE products, Atmos. Meas. Tech., 17, 5301–5340, https://doi.org/10.5194/amt-17-5301-2024, 2024.

E. Sauvageat, A. Haefele, E.: E-Profile reprocessing of ALC calibration constants for CHM15k and Mini-MPL instruments, https://doi.org/10.5281/zenodo.18739645, 2026.

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

ECMWF: IFS Documentation CY49R1 - Part VIII: Atmospheric Composition, chap. 8, ECMWF, https://doi.org/10.21957/d13af18259, 2024.

ECMWF: About OpenIFS, https://confluence.ecmwf.int/display/OIFS/About+OpenIFS, 2025.

EUMETNET: Vertical Profile Observations, https://eumetnet.eu/observations/vertical-profile-observations, 2025.

EUMETSAT: Metop-SG 3MI L1b and L1c data guide, https://user.eumetsat.int/resources/user-guides/metop-sg-3-mi-l1b-and-l1c-data-guide, 2025.

Fisher, M.: Generalized frames on the sphere with application to background error covariance modelling, Seminar on recent developments in numerical methods for atmospheric and ocean modelling, in: 6–10 September 2004, Proceedings, ECMWF, pp. 87–101, ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, UK, 2004.

Fisher, M.: Wavelet Jb – A new way to model the statistics of background errors, in: ECMWF Newsletter 106, pp. 23–28, ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, UK, 2006.

Flemming, J.: Tropospheric chemistry in the Integrated Forecasting System of ECMWF, Geosci. Model Dev., 8, 975–1003, 2015.

Gasteiger, J. and Wiegner, M.: MOPSMAP v1.0: a versatile tool for the modeling of aerosol optical properties, Geoscientific Model Development, 11, 2739–2762, https://doi.org/10.5194/gmd-11-2739-2018, 2018.

Geleyn, J.-F., Bénard, P., and Fournier, R.: 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, https://doi.org/https://doi.org/10.1256/qj.04.107, 2005.

Haefele, A., Bircher-Adrot, S., Rüfenacht, R., Lehmann, V., Mattis, I., Mortier, A., Cimini, D., and Turp, M.: EUMETNET's E-PROFILE network for thermodynamic profiling and the detection of airborne hazards, Tech. rep., Copernicus Meetings, 2022.

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

Hollingsworth, A.: Toward a monitoring and forecasting system for atmospheric composition: The GEMS Project, Bull. Amer. Meteor. Soc., 89, 1147–1164, 2008.

Huijnen, V., Flemming, J., Chabrillat, S., Errera, Q., Christophe, Y., Blechschmidt, A.-M., Richter, A., and Eskes, H.: C-IFS-CB05-BASCOE: stratospheric chemistry in the Integrated Forecasting System of ECMWF, Geoscientific Model Development, 9, 3071–3091, https://doi.org/10.5194/gmd-9-3071-2016, 2016.

Huijnen, V., Le Sager, P., Köhler, M. O., Carver, G., Rémy, S., Flemming, J., Chabrillat, S., Errera, Q., and van Noije, T.: OpenIFS/AC: atmospheric chemistry and aerosol in OpenIFS 43r3, Geoscientific Model Development, 15, 6221–6241, https://doi.org/10.5194/gmd-15-6221-2022, 2022.

Inness, A., Baier, F., Benedetti, A., Bouarar, I., Chabrillat, S., Clark, H., Clerbaux, C., Coheur, P., Engelen, R. J., Errera, Q., Flemming, J., George, M., Granier, C., Hadji-Lazaro, J., Huijnen, V., Hurtmans, D., Jones, L., Kaiser, J. W., Kapsomenakis, J., Lefever, K., Leitão, J., Razinger, M., Richter, A., Schultz, M. G., Simmons, A. J., Suttie, M., Stein, O., Thépaut, J.-N., Thouret, V., Vrekoussis, M., Zerefos, C., and the MACC team: The MACC reanalysis: an 8 yr data set of atmospheric composition, Atmospheric Chemistry and Physics, 13, 4073–4109, https://doi.org/10.5194/acp-13-4073-2013, 2013.

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., and Suttie, M.: The CAMS reanalysis of atmospheric composition, Atmospheric Chemistry and Physics, 19, 3515–3556, https://doi.org/10.5194/acp-19-3515-2019, 2019.

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

Inness, A., Ades, M., Balis, D., Efremenko, D., Flemming, J., Hedelt, P., Koukouli, M.-E., Loyola, D., and Ribas, R.: Evaluating the assimilation of S5P/TROPOMI near real-time SO2 columns and layer height data into the CAMS integrated forecasting system (CY47R1), based on a case study of the 2019 Raikoke eruption, Geoscientific Model Development, 15, 971–994, https://doi.org/10.5194/gmd-15-971-2022, 2022.

Kahnert, M.: On the observability of chemical and physical aerosol properties by optical observations: Inverse modelling with variational data assimilation, Tellus, 61B, 747–755, 2009.

Kahnert, M.: Information constraints in variational data assimilation, Q. J. R. Meteorol. Soc., 144, 2230–2244, 2018.

Kahnert, M. and Andersson, E.: How much information do extinction and backscattering measurements contain about the chemical composition of atmospheric aerosol, Atmos. Chem. Phys., 17, 3423–3444, 2017.

Kahnert, M., Ades, M., Bacles, M., Flemming, J., Guidard, V., Haefele, A., Hogan, R. J., Remy, S., and Sauvageat, E.: E-profile processing and observation modelling for the IFS-COMPO, CY49R1 (1.0.0), https://doi.org/10.5281/zenodo.18803400, 2026.

Kandler, K., Schütz, L., Deutscher, C., Hofmann, H., Jäckel, S., Knippertz, P., Lieke, K., Massling, A., Schladitz, A., Weinzierl, B., Zorn, S., Ebert, M., Jaenicke, R., Petzold, A., and Weinbruch, S.: Size distribution, mass concentration, chemical and mineralogical composition, and derived optical parameters of the boundary layer aerosol at Tinfou, Morocco, during SAMUM 2006, Tellus, 61B, 32–50, https://doi.org/10.1111/j.1600-0889.2008.00385.x, 2009.

Li, Z., Zang, Z., Li, Q. B., Chao, Y., Chen, D., Ye, Z., Liu, Y., and Liou, K. N.: A three-dimensional variational data assimilation system for multiple aerosol species with WRF/Chem and an application to $PM_{2.5}$ prediction, Atmospheric Chemistry and Physics, 13, 4265–4278, https://doi.org/10.5194/acp-13-4265-2013, 2013.

Liang, Y., Zang, Z., Liu, D., Yan, P., Hu, Y., Zhou, Y., and You, W.: Development of a three-dimensional variational assimilation system for lidar profile data based on a size-resolved aerosol model in WRF–Chem model v3.9.1 and its application in $PM_{2.5}$ forecasts across China, Geoscientific Model Development, 13, 6285–6301, https://doi.org/10.5194/gmd-13-6285-2020, 2020.

Parrish, D. F. and Derber, J. C.: The National Meteorological Centre's spectral statistical interpolation analysis system, Mon. Wea. Rev., 120, 1747–1763, 1992.

Peuch, V.-H., Engelen, R., Rixen, M., Dee, D., Flemming, J., Suttie, M., Ades, M., Agustí-Panareda, A., Ananasso, C., Andersson, E., Armstrong, D., Barré, J., Bousserez, N., Dominguez, J. J., Garrigues, S., Inness, A., Jones, L., Kipling, Z., Letertre-Danczak, J., Parrington, M., Razinger, M., Ribas, R., Vermoote, S., Yang, X., Simmons, A., de Marcilla, J. G., and Thépaut, J.-N.: The Copernicus Atmosphere Monitoring Service: From Research to Operations, Bulletin of the American Meteorological Society, 103, E2650 – E2668, https://doi.org/10.1175/BAMS-D-21-0314.1, 2022.

Rémy, S., Kipling, Z., Huijnen, V., Flemming, J., Nabat, P., Michou, M., Ades, M., Engelen, R., and Peuch, V.-H.: Description and evaluation of the tropospheric aerosol scheme in the Integrated Forecasting System (IFS-AER, cycle 47R1) of ECMWF, Geoscientific Model Development, 15, 4881–4912, https://doi.org/10.5194/gmd-15-4881-2022, 2022.

Rémy, S., Metzger, S., Huijnen, V., Williams, J. E., and Flemming, J.: An improved representation of aerosol in the ECMWF IFS-COMPO 49R1 through the integration of EQSAM4Climv12 – a first attempt at simulating aerosol acidity, Geoscientific Model Development, 17, 7539–7567, https://doi.org/10.5194/gmd-17-7539-2024, 2024.

Sun, W., Liu, Z., Chen, D., Zhao, P., and Chen, M.: Development and application of the WRFDA-Chem three-dimensional variational (3DVAR) system: aiming to improve air quality forecasting and diagnose model deficiencies, Atmospheric Chemistry and Physics, 20, 9311–9329, https://doi.org/10.5194/acp-20-9311-2020, 2020.

https://doi.org/10.5194/egusphere-2025-6077
Preprint. Discussion started: 11 March 2026

Welton, E. J., Stewart, S. A., Lewis, J. R., Belcher, L. R., Campbell, J. R., and Lolli, S.: Status of the NASA Micro Pulse Lidar Network (MPLNET): Overview of the network and future plans, new version 3 data products, and the polarized MPL, in: EPJ Web of Conferences, vol. 176, p. 09003, EDP Sciences, 2018.

Wiegner, M. and Geiß, A.: Aerosol profiling with the Jenoptik ceilometer CHM15kx, Atmospheric Measurement Techniques, 5, 1953–1964, https://doi.org/10.5194/amt-5-1953-2012, 2012.

Williams, J. E., Huijnen, V., Bouarar, I., Meziane, M., Schreurs, T., Pelletier, S., Marécal, V., Josse, B., and Flemming, J.: Regional evaluation of the performance of the global CAMS chemical modeling system over the United States (IFS cycle 47r1), Geoscientific Model Development, 15, 4657–4687, https://doi.org/10.5194/gmd-15-4657-2022, 2022.

Yang, T., Li, H., Wang, H., Sun, Y., Chen, X., Wang, F., Xu, L., and Wang, Z.: Vertical aerosol data assimilation technology and application based on satellite and ground lidar: A review and outlook, Journal of Environmental Sciences, 123, 292–305, https://doi.org/https://doi.org/10.1016/j.jes.2022.04.012, 2023.