Supplement of

Alicenet - An Italian network of Automated Lidar-Ceilometers for 4D aerosol monitoring: infrastructure, data processing, and applications

S1 Cloud screening

In order to avoid cloud contamination within the aerosol retrievals, the ALC signals are filtered from 500 m below the Cloud Base Height (CBH, identified by the ALC) up. This threshold was setted to avoid both the presence of cloud droplets and virga frequently observed below CBH. Furthermore, ALC signals are filtered out 15 min before and after the cloud detection. As an example of the effectiveness of the cloud screening procedure, Fig. S1 shows both the 1-hour averaged cloud screened and cloud unscreened total attenuated backscatter profiles derived from the Alicenet CHM15k in Aosta in three different cloudy conditions. These were characterised by the presence of low (Fig. S1a), medium (Fig. S1b), and high (Fig. S1c) clouds.

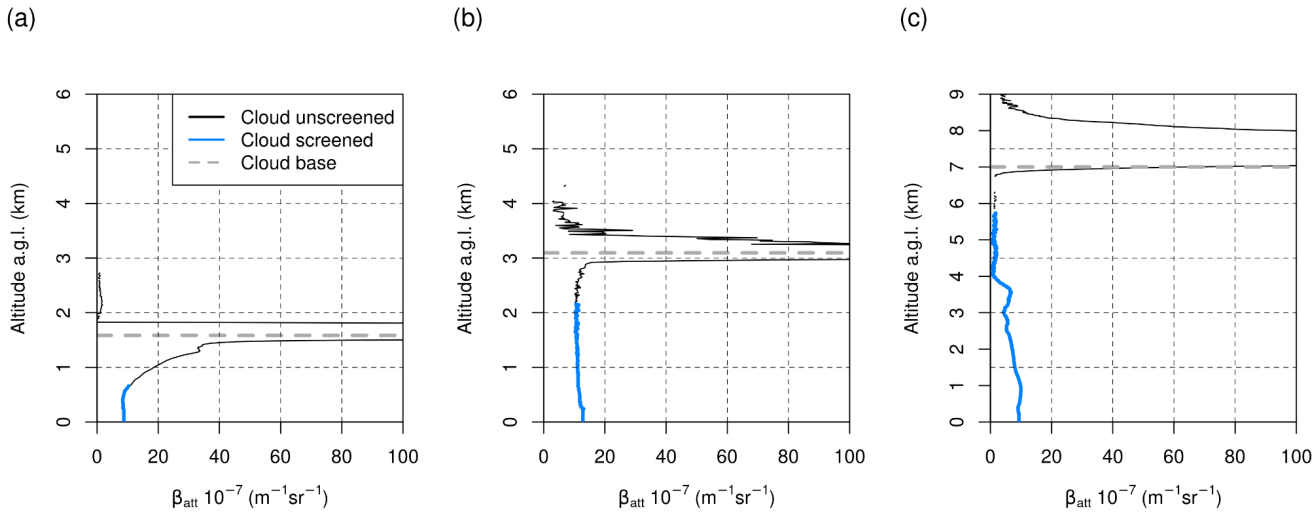

Figure S1: Cloud (un)screened total attenuated backscatter profiles derived from the CHM15K signals in Aosta: (a) 19/06/2022 4-5 UTC, (b) 27/06/2022 11-12 UTC, (c) 18/06/2022 20-21 UTC. Values cat at 100 m-1 sr-1 to highlight aerosol profiles.

S2 Overlap correction

Within Alicenet, the overlap correction applied to CHM15k systems is based on the procedure of Hervo et al. (2016). This procedure first selects time windows reasonably associated with the presence of an homogeneous aerosol layer near the surface by considering the ALC signal in the lowermost levels with complete overlap (typically > 800-1200 m for CHM15k systems). For each window an overlap correction factor ($f_c(r)$) for the overlap function provided by the manufacturer ($Ovl_{man}(r)$) is derived by imposing the ALC signal to be constant down to the ground. The overlap function resulting from the application of $f_c(r)$ can thus be expressed as:

$$Ovl(r) = \frac{Ovl_{man}(r)}{f\,c(r)} \quad (1)$$

The relative difference between the Ovl-corrected and the Ovl_{man}-corrected ALC signals is thus $RD(r) = f_c(r) - 1$.

The procedure is generally applied over a > 1-year ALC dataset. From an ensemble of overlap correction factors and associated median instrument internal temperatures (T_m), a temperature-dependent overlap model Ovl(r,T) is derived assuming a linear dependence of RD(r) on T_m.

Within Alicenet, prior to the derivation of the overlap model, the ensemble of overlap correction factors is quality checked to filter those reasonably derived in inhomogeneous conditions. To this purpose, for each $f_c(r)$ a metric (M_{ov}) function of RD(r) and T_m is calculated:

$$M_{ov} = \sum \left(\frac{\left| \overline{RD(r) - RD_{near}(r)} \right|}{\left| T_m - T_{near} \right|} \right) / N \qquad (2)$$

where the sum is done over the N nearest (in temperature, i.e., $\Delta T_m < 5$ K) relative differences (RD_{near}) and associated instrument internal temperatures (T_{near}), and the overline denotes the median over the range 225-1200 m. The overlap correction factors associated with $M_{ov} < 0.05$ are filtered out.

Finally, if the number of non-rejected overlap correction factors is > 20, the overlap model is derived with a robust linear fit (rlm R package). As an example, the overlap models derived for the Alicenet CHM15k in Rome and Aosta exploiting the ALC 2019-2021 datasets are shown in Fig. S2.

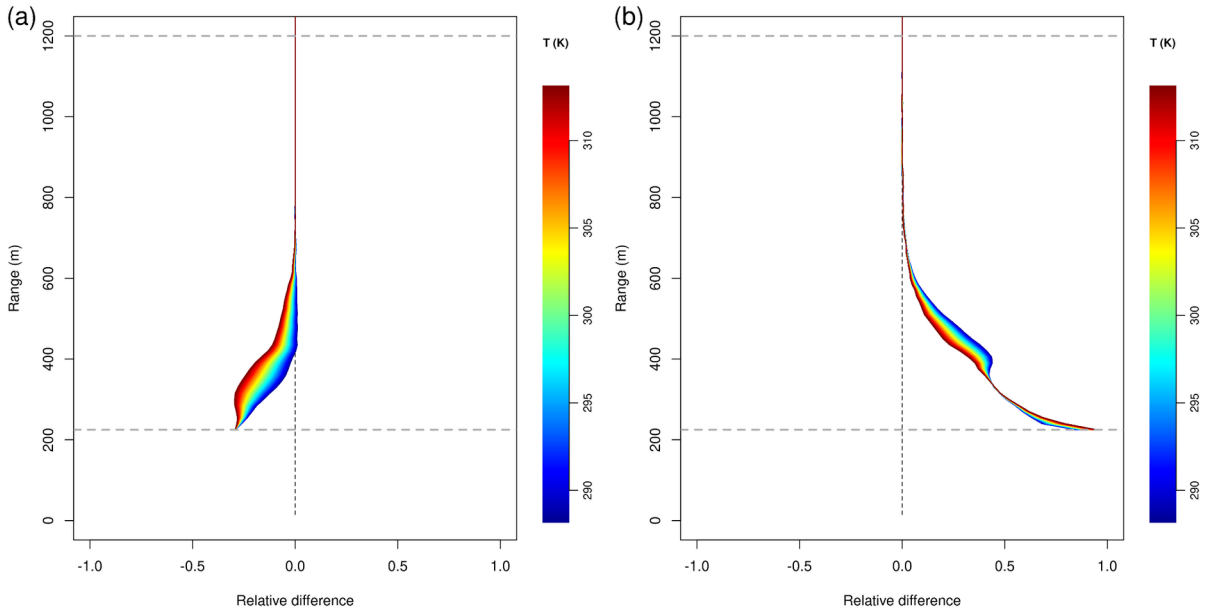

Figure S2: Overlap models derived for the Alicenet CHM15k in (a) Rome, and (b) Aosta.

S3 Absolute calibration

The Rayleigh calibration implemented within Alicenet is based on the comparison between the ALC signal, temporally averaged over a nighttime window of 3-6 hours (depending on cloudiness), and the theoretical molecular attenuated (by molecules) backscatter profile at altitude ranges of 3-7 km altitude. Along this vertical range, a fine grid of 'potential' molecular windows centred at different altitudes and with variable amplitudes (from 600 to 3000 m, at steps of 30 m) is constructed, and for each potential range-window (i.e. combination of central altitude and amplitude) the linear fit between the ALC signal and the theoretical molecular attenuated (by molecules) backscatter profile, and the Breusch-Godfrey test (BG test; bgtest R package) for the calculation of autocorrelation in residuals are performed. The molecular window selected for the calibration is the one maximising a metric (M_{ray}) function of the adjusted R^2 (adjR2) and the intercept (b) of the linear fit, defined as follows:

$$M_{ray} = \frac{adjR\,2 + (1 - |b|)}{std(b)} \quad (3)$$

where std(b) denotes the standard deviation of the intercept of the linear fit over the grid of potential molecular windows.
The quality control QC2 is conceived to filter an erroneously selected window associated with the presence of an homogeneous aerosol layer within its borders. To this purpose, the autocorrelation (BG test) and the cumulative sign in residuals at the window borders are calculated (± 200 m from each border). If the p-value of the BG test is > 0.05 and the cumulative sign in residuals is < 0, the window is rejected.
The uncertainty of the instrument calibration coefficient C_L is defined as follows:

$$E_{CL} = \left(\frac{ster(C_L\,slope)}{C_L\,slope}\right) + \left(\frac{std(C_L)}{C_L}\right) \quad (4)$$

In the first term, C_L^{slope} represents the slope of the fit between the ALC signal and the theoretical molecular attenuated backscatter profile within the selected molecular window, and ster(C_L^{slope}) the standard error of the slope. In the second term, C_L represents the calibration coefficient (more specifically, the median value of C_L along the molecular window) derived following Wiegner and Geiβ (2014), and std(C_L) the standard deviation of the calibration coefficient over the molecular window. The calibration coefficients associated with relative uncertainty ((E_{CL} / C_L) 100) > 40% are rejected.

S4 Retrieval of aerosol properties

Within Alicenet, the aerosol extinction coefficient $\alpha_p(r)$ is retrieved through a specific functional relationship linking this quantity to the aerosol backscatter coefficient $\beta_p(r)$ as described by Dionisi et al. (2018). The Aerosol Optical Depth is then derived by integrating $\alpha_p(r)$ over the ALC-probed column. Figure S3 shows the long-term (2016-2022), multi-site (Aosta, Roma, Messina) comparison between the AOD derived from Alicenet ALCs (at 1064 nm) and AERONET/SKYNET sun photometers (at 1020 nm). Data pairs associated with a photometer-derived Angstrom Exponent (AE) < 0.5, i.e., with coarse-mode dominated aerosol types such as desert dust or sea-salt particles, are highlighted in red.

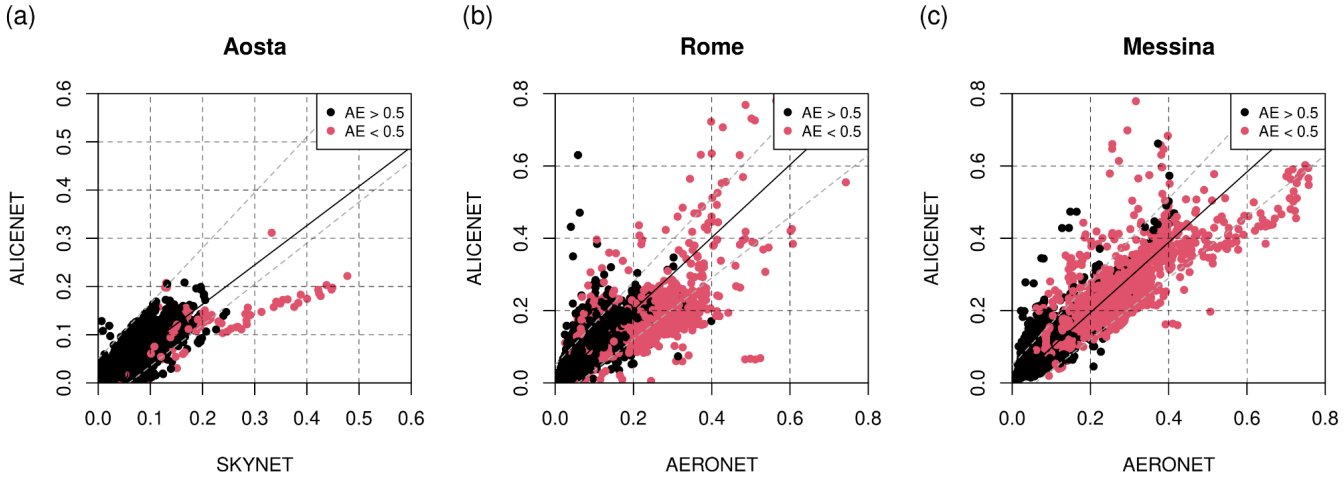

Figure S3: Comparison between the AOD derived by Alicenet (at 1064 nm) and AERONET/SKYNET sun photometers (at 1020 nm) in (a) Aosta, (b) Rome Tor Vergata, and (c) Messina over the 2016-2022 dataset. Black line is the linear fit. Gray dashed lines indicate a AOD-Alicenet deviation to the 1:1 line of delta= ± 0.01 ± 0.15*AODsunphotometer. Data pairs associated with a photometer-derived Angstrom Exponent (AE) < 0.5 are highlighted in red.

The aerosol mass concentration (M_p) is estimated from the Alicenet-derived aerosol volume (V_p) as $M_p = \rho_p V_p$, where ρ_p is an a-priori aerosol density. For comparison with in-situ instrumentation (as those operating in AQMN in compliance to the EU AQ Directive) generally drying the aerosol samples, M_p can be corrected to 'dry' aerosol mass (M_p^{dry}) by taking aerosol hygroscopicity into account. The comparison between the Alicenet-derived M_p^{dry} and in-situ PM_{10} measurements was conducted exploiting the Alicenet CHM15k in Aosta and an OPC operating at the high altitude (3500 m a.s.l.) station Testa Grigia - Plateau Rosa (data courtesy of Stefania Gilaroni CNR-ISP). The map with the relative location of the Aosta and Testa Grigia - Plateau Rosa stations is shown in Fig. S4.

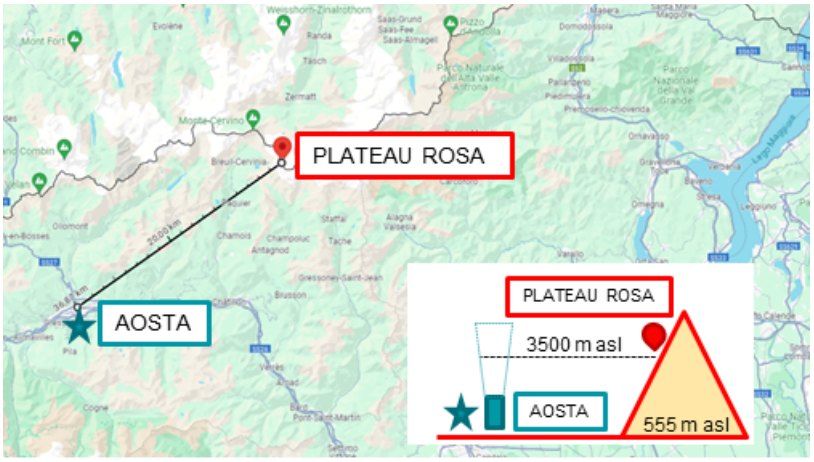

Figure S4: Map with the locations of the Aosta and Testa Grigia - Plateau Rosa stations. In the bottom-right corner, a scheme with the station altitudes is reported (map credits: © Google Maps).

S5 Identification of aerosol layers

Within Alicenet, the identification of the main aerosol layer types (i.e., continuous, mixed, and elevated aerosol layers, CAL, MAL, and EAL, respectively) is solely based on ALC data and takes advantage of specific procedures. In the following, details are provided concerning the identification of:

a) the Mixed Aerosol Layer (MAL).

The search of the MAL height (MALH) starts 1 hour before sunrise and stops 1 hour after sunset (the site- and season-dependent sunrise/sunset hours are calculated using the Michalsky algorithm; Michalsky, 1988), and is performed within maximum/minimum heights defined by the user (generally, a maximum height of 3.5 km is setted during spring and summer, and a minimum height of 250 m is setted all over the year).

For the search of the MALH, the Dynamic Time Warping algorithm (DTW, Giorgino et al., 2009) is applied to a sequence (> 30 min) of denoised β_{att} profiles at 1 min resolution. This algorithm computes the local stretch or compression to apply to each profile in order to optimally map the preceding into the following. Its output is a time- and range-dependent field (w_{DTW}) which can be interpreted as the local vertical displacement of the aerosol-loaded air parcels, and can be used to identify regions where aerosols are mixed by vertical turbulent fluxes. The standard deviation of w_{DTW} (w_{SD}) is thus calculated over 30 min intervals, and at each interval the MALH is identified as the height of the w_{SD} local minima (Z_{min}) which:

- lie in the vertical range (MALH(t-Δt)-300m, MALH(t-Δt)+1200m) during the morning (with Δt = 30 min), and (MALH(t-Δt)+300 m, MALH(t-Δt)-1200 m) during the afternoon;
- minimise a metric function of wSD and the vertical increment with respect to MALH(t-Δt). This metric is defined as follows:

$$M_{mal} = \left| Z_{min} - MALH\left(t - \Delta t\right) \right| + \frac{1}{w_{SD}} \quad (5)$$

where the overline represents the median along the vertical range (250 m, Z_{min}).

Although the DTW technique already links each time step to the previous one, the inclusion of additional continuity criteria within the procedure is foreseen in the next future.

If a 3D ultrasonic anemometer is co-located with the ALC, the MALH can be derived using a synergic approach. The vertical increment between MALH(t) and MALH(t-Δt) is derived as the average of the increment computed with the ALC-based procedure described above and the increment computed with ultrasonic anemometer data (Eddy-Covariance analysis). The increment computed with ultrasonic anemometer data is defined as (Stull, 1988):

$$dr = \frac{2\,HFX\,\Delta t}{\Delta T} \quad (6)$$

where HFX is the surface sensible heat flux and ΔT is the increment of the surface air temperature within Δt (Δt = 30 min).

b) the Elevated Aerosol Layers (EALs).

The EAL 'centres' are detected through a Continuous Wavelet Transform (CWT) technique. The MassSpecWavelet algorithm developed by Du et al. (2006) is used for this purpose. This algorithm takes as input cloud-screened and denoised β_{att} profiles at 30 min resolution, and exploits both CWT and β_{att} coefficients to identify peaks (here referred to as the EAL 'centres') attributable to aerosol layers and discriminate them from noise.

For each detected EAL 'centre', a grid of potential bottom and top semi-amplitudes (from 150 m to 1500 m, at steps of 60 m) is constructed (the procedure allows the selection of different semi-amplitudes for the layer bottom and top), and for each potential bottom/top combination, a metric function of β_{att} and CWT coefficients is calculated. This metric is defined as follows:

$$M_{eal} = \frac{I\left(\beta_{att}\right)}{I\left(\beta_{r}\right)} + Perc + G_{border} + CWT_{border} \quad (7)$$

In the first term (right-hand side), I represents the integral along the potential EAL top-bottom range, and β_r a reference total attenuated backscatter profile. In the second term, Perc represents the percentage of points with $\beta_{att} > \beta_{mol}$ within the top-bottom range. In the third term, G_{border} represents the mean gradient of β_{att} around the potential layer boundaries (± 90 m from

each border) normalised by the maximum value of the β_{att} gradient along the top-bottom range. In the fourth term, CWT_{border} represents the mean CWT coefficient around the potential layer boundaries (± 90 m from each border) normalised by the maximum value of the CWT coefficients along the top-bottom range.

The bottom/top combination maximising M_{eal} is selected and classified as the EAL vertical range. The selected EAL must verify specific criterias to be finally accepted:

- a percentage of points exceeding $\beta_r > 90\%$ along the top-bottom range;
- an integrated (long the top-bottom range) $\beta_{att} > 3$ times the integrated β_r;
- a negative (positive) β_{att} gradient at the layer top (bottom), and negative CWT coefficients at both borders (± 90 m from the bottom/top).

The threshold values 90% and 3 are user-defined.