Supplementary materials

Integration and Comparative Analysis of Remote Sensing and In situ Observations of Aerosol Optical Characteristics Beneath Clouds

Jing Chen[a], Jing Duan[b,*], Ling Yang[a], Yong Chen[c], Lijun Guo[b] and Juan Cai[d]

[a] College of Electronic Engineering, Chengdu University of Information Technology, Chengdu 610225, China; chen-jing7292022@163.com (J.C.); cimyang@cuit.edu.cn (L.Y.)
[b] China Meteorological Administration Cloud-Precipitation Physics and Weather Modification Key Laboratory, BeiJing 100081, China; ljguo@cma.gov.cn (L.G.)
[c] State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences, BeiJing 100029, China; chenyong@mail.iap.ac.cn
[d] Weather Modification Center of Jiangxi Province, Nanchang 330096, China; caijuan0103@163.com
[*] Correspondence: duanjing@cma.gov.cn

Figure S1. Schematic of the Lidar cube and visibility meter (in red box).

Method S1：Adaptive wavelet threshold denoising

To maximise the suppression of the interference of the noise signal to the useful signal, ensuring the accuracy and reliability of the subsequent signal processing. In this research, the wavelet threshold denoising algorithm with the number of decomposition layers determined adaptively is used. The expression of the cross-correlation coefficient as follows:

$$\rho_m = \frac{Cov(cA_m, cA_{m-1})}{\sqrt{D(cA_m)}\sqrt{D(cA_{m-1})}} = \frac{\sum_{i-1}^{n}(cA_m - \overline{cA_m})\,(cA_{m-1} - \overline{cA_{m-1}})}{\sqrt{\sum_{i=1}^{n}\left((cA_m - \overline{cA_m})\right)^2}\sqrt{\sum_{i=1}^{n}\left((cA_{m-1} - \overline{cA_{m-1}})\right)^2}} \quad (S1)$$

In Eq. (S1), cA_m and cA_{m-1} represent the approximation components after wavelet decomposition at the levels m and m-1, respectively. If $\rho_{m+1} - \rho_m > \rho_m - \rho_{m-1}$, indicates a higher cross-correlation coefficient between the approximation components of neighbouring decomposition levels, implying that noise signals in the approximate components is dominant and need further decomposition at the next level. Conversely, if $\rho_{m+1} - \rho_m < \rho_m - \rho_{m-1}$, the decrease in the cross-correlation coefficient slope indicates that useful signals is dominant, and further decomposition may lead to signal distortion. Then, the decomposition level m is considered the optimal decomposition level.

Select the Daubechies (db5) wavelet functional as wavelet basis, which has orthogonality, tight support, and similar to the Lidar signal waveform. Signal reconstruction with the better adapted soft threshold function and VisuShrink threshold criteria. To verify the denoising efficiency of the method, a simulation experiment with N=1~4 levels of wavelet decomposition was performed on a simulated signal with random noise. Fig. S2 shows the approximation and detail components of each level after wavelet decomposition. As the decomposition levels increase, the resolution of the signal increase in both frequency and time, which helps to enhance the detail information of the signal. However, an excessive amount of decomposition layers may result in more noise being introduced into the detail components as well. Calculation the cross-correlation coefficients between adjacent decomposition level approximation components with Eq. (S1) is: $\rho_1 = 0.95$, $\rho_2 = 0.97$, $\rho_3 = 0.98$, $\rho_3 - \rho_2 < \rho_2 - \rho_1$. This indicates that after the third level decomposition, the content of useful information in the approximation component shows a decreasing trend compared with the second level decomposition. Thus, N=3 is the optimal decomposition level. Table S1 lists the signal-to-noise ratio (SNR) and root mean square error (RMSE) for levels of N=1~4. When N=3, the SNR reaches a relative maximum of 14.77dB and the RMSE is relatively minimum at 0.07. This shows the effectiveness of the proposed algorithm in determining the optimal decomposition level and improving the signal quality.

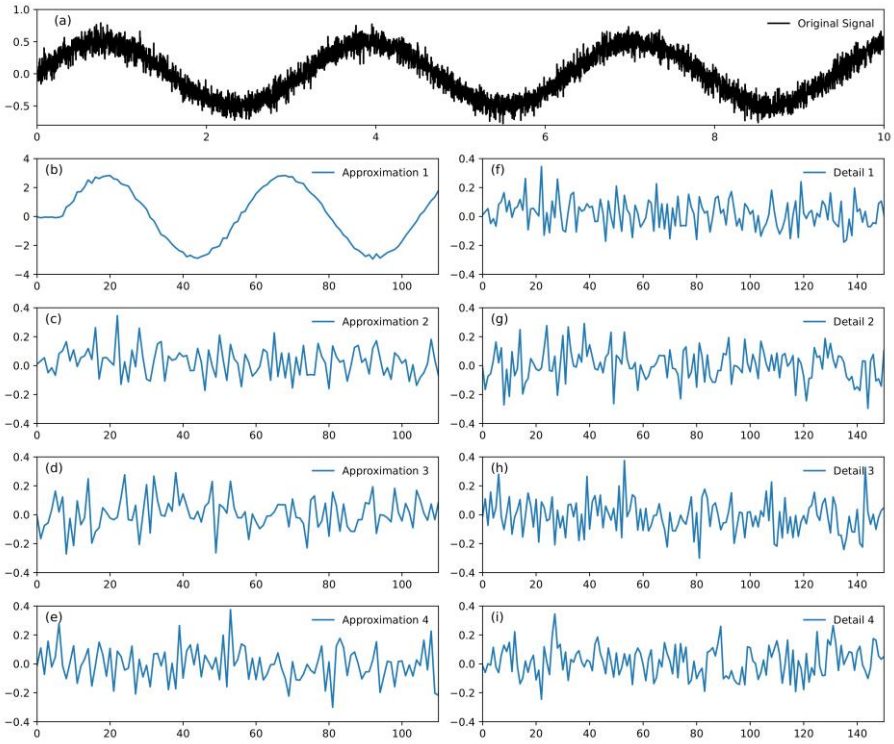

Figure S2. Four-level wavelet decomposition of noise signal: (a) original signal; (b)~(e) approximate parts a1~e4; (f)~(i) detail parts f1~i4.

Table S1. Reconstructed SNR and RMSE at different decomposition levels.

Parameters	First Level	Second Level	Third Level	Fourth Level
SNR(dB)	12.47	13.15	14.77	13.88
RMSE	0.09	0.08	0.07	0.07

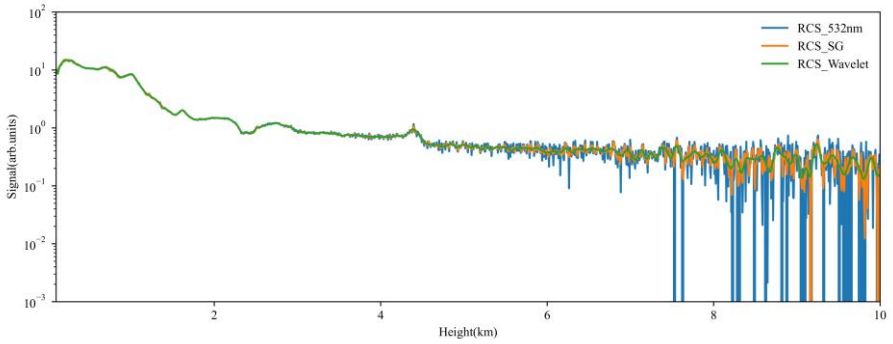

Figure S3. Lidar signal before and after denoising by SG and adaptive wavelet threshold.

The Savitzky-Golay filter and the adaptive wavelet threshold method are used for denoising on the distance-corrected signals at a certain moment in time, respectively, and the comparison is shown in Fig. S3. Before denoising, the signal has a strong fluctuation in intensity above 6km, and most of signals are drowned in the noise. After SG filtering, the effective detection distance increased to approximately 7 km. Meanwhile, when the adaptive wavelet threshold was applied, the high-frequency noise components were greatly reduced, increasing the effective detection range to over 10 km, and the denoised signal was smoother. Within the effective detection range, the SNR corresponding to the SG filter and the adaptive wavelet threshold were 16.19 dB and 32.47 dB, respectively. The SNR of the 532 nm signal improved by a factor of 2 after denoising with the adaptive wavelet threshold method. This indicates the wavelet adaptive threshold method can effectively filter out noise while preserving the features of useful signals, improving data quality and detection capabilities.

Method S2. Coupling the visibility meter to determine the calibrated height of the near end.

Fig. S4 shows the temporal and spatial distribution of Lidar distance-corrected signals after data quality control, and the dots are the bottom cloud heights calculated by the differential zero-crossing method. It can be seen that before 10:00, there were clouds in the range of 1.8~4km, while it was clear-sky after 10:00. It is noteworthy that in the case of signals after overall cloud coverage, the laser energy is impaired or even unable to penetrate the thicker clouds due to the scattering and absorption of the laser beam by water droplets or ice crystals in the clouds. Thus, traditional calibration methods may no longer be applicable in the presence of clouds.

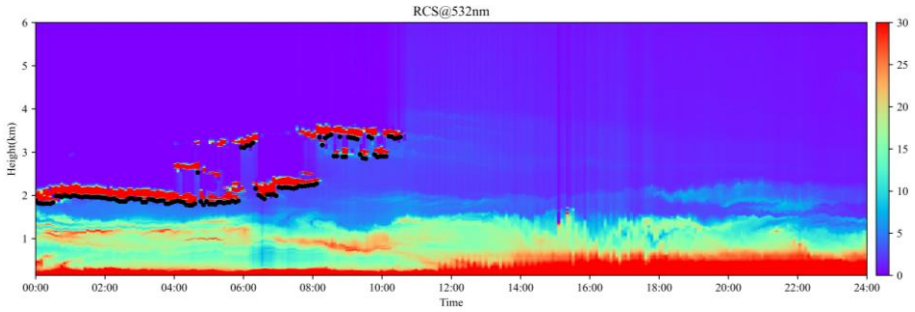

Figure S4. Distance-corrected signals and cloud base from the 532 nm of the Lidar (March 30, 2023).

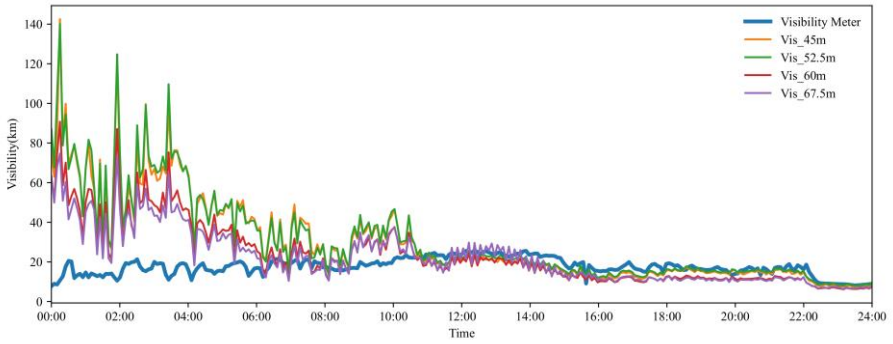

Figure S5. Compare inverted visibility from the Lidar data with visibility meter data.

To solve the above problems, this research chooses the near-end calibration method to determine the sum of the extinction contributions of aerosols and atmospheric molecules at calibrated heights based on the data from the visibility meter deployed on the Lidar. The remote calibration method is used to assist to determine the suitable calibrated height near the ground in clear-sky. The aerosol extinction coefficients were retrieved by remote calibrated method and the visibility at different heights near the ground was calculated, while the results of the calculations were compared and analysed with the data from the ground-based visibility meter. From Fig. S4 and Fig. S5, the time from midnight to 10:40 a.m. the clouds existed at a height of 2~4km, and the extinction coefficient beneath the clouds calculated with the remote calibrated method was lower than in clear-sky weather, and the visibility calculated by using this data also had a large error with the visibility meter. After 10:40 a.m., it turned to clear-sky, at which time the visibility calculated by Lidar at four different heights near the ground (45m, 52.5m, 60m, and 67.5m) was in good agreement with the visibility meter data.

Selected data during cloudless periods, a correlation analysis was performed between the visibility obtained at four different heights by the Lidar and the visibility meter. The results are shown in Fig. S6. The visibility calculated by the Lidar at 45m has the strongest correlation with the visibility meter(Pearson=0.94, R^2=0.89), which indicates a strong linear relationship between them, and the visibility data measured by Lidar at 45m can effectively approximate the visibility meter.

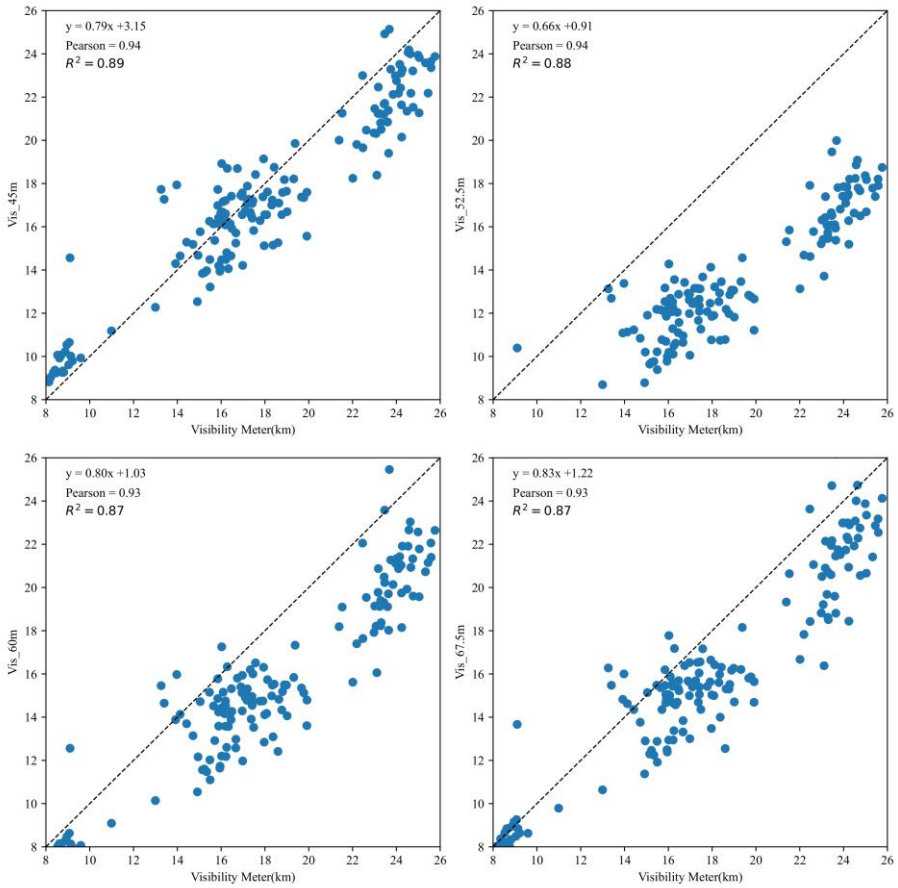

Figure S6. Correlation analysis of retrieved visibility of Lidar at different heights (45m, 52.5m, 60m, 67.5m) with the visibility meter.

In addition, in order to ensure the reliability of Lidar observation data, SNR of the original echo signal is calculated to determine the effective detection height in this research. After experimental trials, set 3 dB as the SNR threshold, and considered the signals able to provide more reliable observations if SNR is above the threshold. Consequently, the highest height with an SNR above this threshold is selected as the maximum effective detection height, and data above this height are considered invalid and excluded from subsequent inversion.

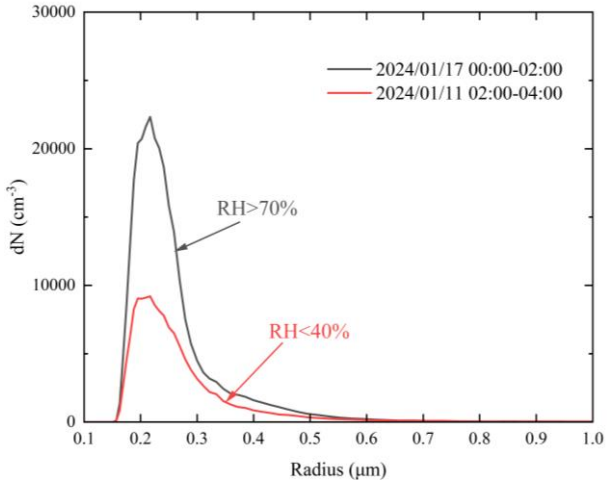

Figure S7. Variation in radius-number concentration of aerosol particles in clear-sky and cloudy-sky (Data from Welas).