Technical Note

Rayleigh Lidar Signal Denoising Method Combined with WT, EEMD and LOWESS to Improve Retrieval Accuracy

Yijian Zhang [1], **Tong Wu** [2], **Xianzhong Zhang** [1], **Yue Sun** [1], **Yu Wang** [1], **Shijie Li** [1], **Xinqi Li** [1], **Kai Zhong** [1,*], **Zhaoai Yan** [3], **Degang Xu** [1] and **Jianquan Yao** [1]

1   Key Laboratory of Optoelectronic Information Science and Technology (Ministry of Education),
    School of Precision Instruments and Opto-Electronic Engineering, Tianjin University, Tianjin 300072, China;
    zhangyijian1895_@tju.edu.cn (Y.Z.); zxzhitr@tju.edu.cn (X.Z.); sunyue102695@tju.edu.cn (Y.S.);
    2020202055@tju.edu.cn (Y.W.); li_shije@tju.edu.cn (S.L.); qixinli19990605@tju.edu.cn (X.L.);
    xudegang@tju.edu.cn (D.X.); jqyao@tju.edu.cn (J.Y.)
2   School of Marine Science and Technology, Tianjin University, Tianjin 300072, China; wutong_101@tju.edu.cn
3   Key Laboratory of Science and Technology on Environmental Space Situation Awareness,
    National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China; yanza@nssc.ac.cn
*   Correspondence: zhongkai@tju.edu.cn

Abstract: Lidar is an active remote sensing technology that has many advantages, but the echo lidar signal is extremely susceptible to noise and complex atmospheric environment, which affects the effective detection range and retrieval accuracy. In this paper, a wavelet transform (WT) and locally weighted scatterplot smoothing (LOWESS) based on ensemble empirical mode decomposition (EEMD) for Rayleigh lidar signal denoising was proposed. The WT method was used to remove the noise in the signal with a signal-to-noise ratio (SNR) higher than 16 dB. The EEMD method was applied to decompose the remaining signal into a series of intrinsic modal functions (IMFs), and then detrended fluctuation analysis (DFA) was conducted to determine the threshold for distinguishing whether noise or signal was the main component of the IMFs. Moreover, the LOWESS method was adopted to remove the noise in the IMFs component containing the signal, and thus, finely extract the signal. The simulation results showed that the denoising effect of the proposed WT-EEMD-LOWESS method was superior to EEMD-WT, EEMD-SVD and VMD-WOA. Finally, the use of WT-EEMD-LOWESS on the measured lidar signal led to significant improvement in retrieval accuracy. The maximum error of density and temperature retrievals was decreased from 1.36% and 125.79 K to 1.1% and 13.84 K, respectively.

Keywords: rayleigh lidar signal denoising; ensemble empirical mode decomposition; wavelet transform; locally weighted scatterplot smoothing; detrended fluctuation analysis; density retrieval; temperature retrieval

1. Introduction

As an active remote sensing technology, lidar is widely used in autonomous driving, target detection and atmospheric detection, which has the characteristics of high spatial resolution and detection sensitivity [1–4]. In recent years, the middle atmosphere has received increasing attention. Rayleigh lidar, employing Rayleigh backscattering from atmospheric molecules, has become an effective means for studying middle atmosphere [5–7]. However, both the exponential decrease in number density with altitude and the spherical scattered light result in a decrease in lidar echo signal with the altitude squared, severely limiting the effective detection range and retrieval accuracy [8]. Moreover, strong background sunlight noise, dark current noise, and thermal noise interfere with the lidar echo signal [9]. Consequently, the signal-to-noise ratio at high altitudes is too low to retrieve correctly, so the lidar backscattering signal must be denoised.

In the last few decades, many researchers have paid attention to the processing of nonlinear and non-stationary lidar echo signal. The moving averages technique, a low-pass

filtering method, is the simplest and most direct denoising method. However, on the one hand, the ideal filter will inevitably depend on the particular scene. On the other hand, this method will create extra errors due to the smearing of resolvable atmospheric structure [10]. Han et al. used the Kalman filtering method to derive water vapor profiles using surface in situ and Raman measurements, but this method may cause retrieval errors in cases of rapidly changing environments [11]. The WT method can decompose the signal into different frequency terms through multiscale analysis, but the main limitations are the selection of suitable wavelet basis functions and the decomposition level according to a particular scene [12,13].

Empirical mode decomposition (EMD) is widely used in the processing of non-stationary signals, due to the fact that its decomposition entirely depends on the signal itself and does not need any priori information [14,15]. Through EMD, a complex signal can be decomposed into a series of IMFs and a residual. The research on EMD mostly focuses on distinguishing whether noise or signal is the main component of the IMFs, or overcoming the shortcomings of EMD, such as modal aliasing and end effects. To solve these problems, Tian et al. suggested typical range (TR) and low-frequency fraction (LFF) as the reference principles to decide how many high-frequency IMFs should be removed as noise [16], and Wu et al. proposed EEMD to effectively suppress modal aliasing [17].

Variational mode decomposition (VMD) is a new method for decomposing signals adaptively [18]. Compared with EMD, VMD is a non-recursive decomposition algorithm, which can overcome the modal aliasing problem and embody Wiener filtering characteristics. In the VMD technique, the selection of the decomposition mode number and quadratic penalty is performed through an optimization algorithm [19]. In addition, the application of machine learning and neural networks for signal denoising has received extensive attention due to their adaptability to include prior information [20]. Sreekanth et al. adopted the dictionary learning technique in denoising lidar signal [21].

In the past, most studies focused on the lidar echo signal with a low signal-to-noise ratio, while ignoring the overall relevance. However, in engineering, it is always necessary to process the lidar echo signal within a certain height range, and therefore the signal-to-noise ratio has a large dynamic range. Since the single denoising method has its own advantages and shortcomings as described above, it is necessary to combine several denoising algorithms for improving the denoising performance. Furthermore, it is more convincing to use the simulated lidar echo signal to verify the effectiveness of the proposed algorithm than to use the synthetic signal.

In this paper, a WT and LOWESS based on the EEMD algorithm, called WT-EEMD-LOWESS, was proposed to extend the effective detection range and improve retrieval accuracy. The proposed method utilized the WT method to remove the noise in the signal with a signal-to-noise ratio higher than 16 dB. The EEMD method was applied to decompose the remaining signal into a series of IMFs, and the signal was reconstructed by using the DFA. Moreover, the LOWESS method was adopted to finely extract the signal in the reconstructed signal. The lidar simulation system with a typical integration time of 1200 s validated the feasibility of the WT-EEMD-LOWESS model. Then, the simulation experiment verified that our algorithm can be effectively applied to the lidar echo signal with different integration times. Finally, the denoising algorithm was investigated on the real Rayleigh lidar signal and experimental results showed that the WT-EEMD-LOWESS method can be directly adapted to engineering.

2. Rayleigh Lidar Simulation System

To select different yet available denoising methods for echo signal with different signal-to-noise ratios, the noisy lidar echo signal was simulated for comparison experiments. In the case of the Rayleigh lidar system, the return signal can be described as follows:

$$N_R(\lambda, R) = \left(\frac{E_0}{hc/\lambda}\right)[\sigma_{180}(\lambda, R)n(R)\Delta R]\frac{A}{R^2}\left[\eta(\lambda)T_tT_rT^2(\lambda, R)G(R)\right] + N_B \qquad (1)$$

where $N_R(\lambda, R)$ is the number of received photons, when the wavelength of emitted laser is λ and the detection distance is R; E_0 is pulse energy of the emitted laser; h is Plank's constant; $\sigma_{180}(\lambda, R)$ is the backscatter cross-section, $\sigma_{180}(\lambda, R) = \frac{1}{4\pi} \times \frac{3}{2} \times 5.16 \times 10^{-31}$ m^{-2}; $n(R)$ is the number density of atmospheric molecules; ΔR is range resolution; A is receiving area of telescope; $\eta(\lambda)$ is detector quantum efficiency; T_t and T_r represent transmitting and receiving optical efficiencies, respectively; $T^2(\lambda, R)$ is two-way atmospheric transmittance; $G(R)$ is geometric overlap factor; N_B is the noise. The first term of the expression represents the number of emitted photons, the second term indicates the probability of backscattering of atmospheric molecules, the third term represents the ability of the detector to accept backscattered light and the last term stands for the system efficiency and transmittance. The schematic of the Rayleigh lidar system is illustrated in Figure 1.
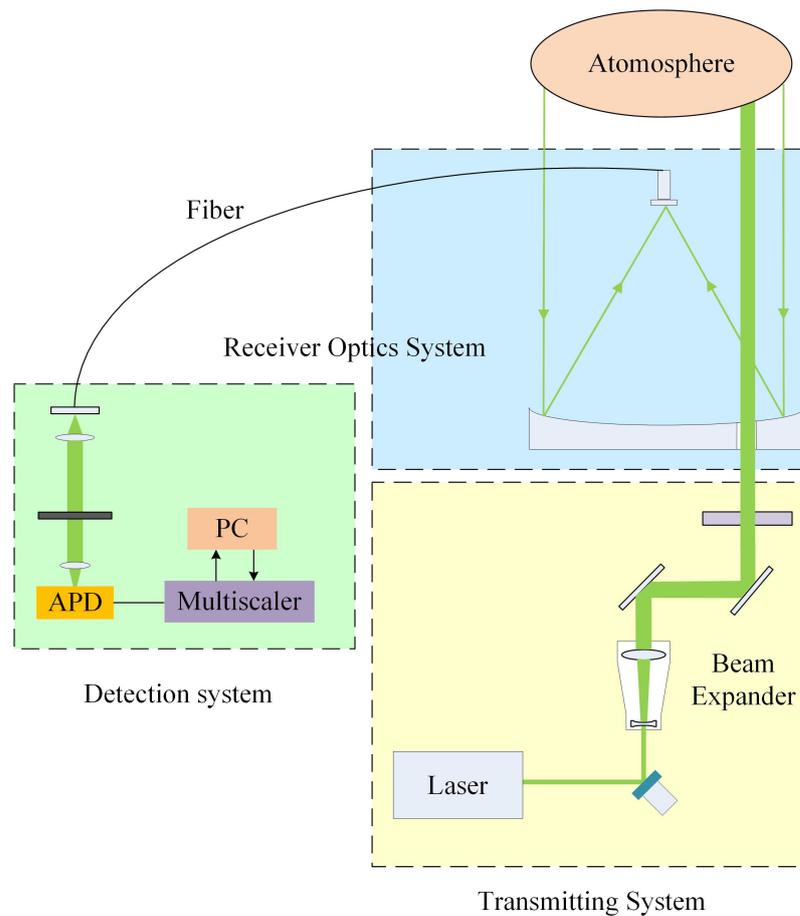

Figure 1. Schematic diagram of the Rayleigh lidar system.

The lidar signal is inevitably affected by noise during the detection process. The main sources of noise are background radiation noise, detector thermal noise and shot noise. The calculation formulas of background radiation noise N_b and detector thermal noise N_d are as follows:

$$N_{\mathrm{b}} = \frac{\eta\lambda}{hc} P_b \left(\frac{\theta_R}{2}\right)^2 \Delta\lambda A \Delta t T_{\mathrm{r}} \tag{2}$$

$$N_{\mathrm{d}} = CPS \cdot \Delta t \tag{3}$$

where P_b is the radiances of the sky, which were calculated using the MODTRAN software [22]; θ_R is the field of view of the receiving telescope; $\Delta\lambda$ is the bandwidth of the filter; CPS is the dark count of the detector; Δt is the acquisition time of the system. Table 1 lists the parameters of our simulation system.

Table 1. Simulation parameters of lidar system.

System Parameter	Value
Laser pulse energy/mJ	40
Telescope diameter/mm	350
Field of view/μrad	165
Detection quantum efficiency	0.5
Total optical transmittance	0.4
Dark counts/counts per second	50
Range resolution/m	100
Filter bandwidth/nm	0.25
Pulse repetition frequency/Hz	50
Platform height/km	20

Shot noise follows a Poisson distribution. Therefore, when using the lidar equation to simulate the echo signal, Poisson noise needs to be added. Since the Rayleigh signal per laser pulse is too weak, averaging the pulses or integrating them is often used to increase the signal-to-noise ratio. Figure 2 shows the ideal echo signal obtained from the simulation and the noisy signal after subtracting the background radiation noise under the condition that the platform height is 20 km, the measurement altitude range is 30–70 km, and the integration time is 1200 s.
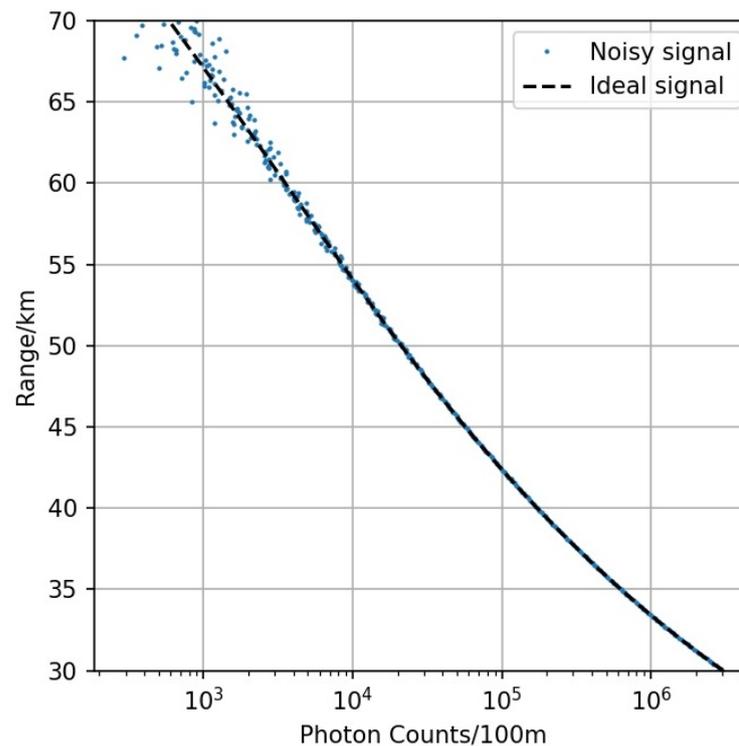
Figure 2. The ideal echo signals and the noisy signals in the case of integration time of 1200 s.

3. Methodology Descriptions

3.1. Brief Description of the WT Algorithm

Wavelet transform is developed on the basis of short-time Fourier transform and has good time-frequency localization characteristics. Compared with Fourier transform, which only has sine or cosine function as a basis, the types of wavelet basis are diverse. Previous experiments have proved that "*dB*4" and "*Sym*4" have achieved good results in the field of lidar signal processing.

In the wavelet threshold denoising, the raw signal is divided into approximation and detail coefficients. The orthogonal wavelet basis can mainly concentrate useful signal

energy in some large wavelet coefficients. Therefore, the employment of threshold can retain the large useful signal coefficients and zero or shrink the the small noise coefficients, so as to denoise. In practice, the selection of the wavelet basis, the decomposition level and threshold value is a key point in algorithm design. The specific steps of wavelet threshold denoising are as follows:

Step 1: According to the characteristics of Rayleigh scattering echo signal, the proper wavelet basis and decomposition layer number N are selected, and the original return lidar data is decomposed with N layers based on wavelet transformation.

Step 2: Selecting the soft threshold method to carry on the threshold quantization processing for the high frequency coefficients from layer 1 to layer N by different threshold value of each decomposition layer.

Step 3: The low-frequency coefficient of the Nth layer and all high-frequency coefficients of each layer after wavelet threshold processing are reconstructed by inverse wavelet transformation, and the required signal is obtained.

3.2. Basic Theory of EEMD

EEMD is an improved algorithm of EMD, so the EMD algorithm is introduced first. The EMD method can adaptively decompose signals into a series of IMFs and a residual component based on the local characteristic time scale of the signal itself, which is given by

$$x(t) = \sum_{i=1}^{n} imf_i + r_n \tag{4}$$

Each IMF stands for the information on different scales of the original signal and must satisfy two conditions: (1) the number of extrema and zero crossings must either be equal or differ by one at most; (2) at any point, the mean value of the envelope defined by the local maxima and minima is zero. The first condition is a common signal analysis requirement for narrowband signals and the second condition ensures the local symmetry of the IMFs.

For any original signals $x(t)$, the procedure of EMD can be described as follows:

1. Identifying the local extrema of $x(t)$.
2. Using the cubic spline interpolation function to form the upper envelope $U(t)$ and the lower envelope $L(t)$.
3. Calculating the mean value $m(t)$ of the upper and lower envelopes.

$$m(t) = [U(t) + L(t)]/2 \tag{5}$$

4. Denoting the difference in value between $x(t)$ and $m(t)$ as $h(t)$. If $h(t)$ satisfies conditions for IMFs, then $h(t)$ is defined as the first IMF. If not, $x(t)$ is represented as $h(t)$. Repeating the steps 1 to 3 *K* times until $h_K(t) = h_{K-1}(t) - m_{K-1}(t)$ meets IMF conditions and $h_K(t)$ will be the first IMF of the signal.

$$h(t) = x(t) - m(t) \tag{6}$$

5. Separating the first IMF from $x(t)$. The residue item $r(t)$ can be computed with the following equation:

$$r(t) = x(t) - \text{imf}_1(t) \tag{7}$$

If $r(t)$ does not become a monotone function or the number of extrema is less than one or equal to one, $x(t)$ is replaced by $r(t)$. Repeating steps 1 to 4 *n* times, the *n* IMFs of $x(t)$ will be computed.

In order to solve modal aliasing problems of EMD, the EEMD algorithm, an effective noise-aided data analysis (NADA) method, adds white noise to the original signal by using the zero-mean characteristic of white noise. Because the white noise spectrum is evenly distributed, the white-noise signals will be automatically distributed to the appropriate reference scales. The specific procedure of the EEMD method is expressed as follows:

1. Adding white noise, which follows a normal distribution with mean 0 and standard deviation σ, N times to the original signal $x(t)$. Then, the new data series can be computed as follows:

$$x_i(t) = x(t) + \sum_{j=1}^{N} \sigma n_j(t), n_j(t) \sim N(0,1) \tag{8}$$

2. Decomposing the signal added with Gaussian white noise by EMD to obtain the *i*th set of decomposition results, which includes IMF components $c_{ij}(t)$ (j = 1, 2, ..., m) and a residual $r_i(t)$.
3. Repeating the above steps n times, the overall average values $c_j(t)$ and $r(t)$ are used as the results of the *j*th IMF component and the residual decomposed by EEMD, respectively.

$$c_j(t) = \frac{1}{n} \sum_{i=1}^{n} c_{ij}(t) \tag{9}$$

$$r(t) = \frac{1}{n} \sum_{i=1}^{n} r_i(t) \tag{10}$$

3.3. Principle of DFA

Selecting specific IMF components for superposition can form high-pass, low-pass and band-pass filters. Because the signal is concentrated on the low-frequency IMFs while noise is devoted to the high-frequency IMFs, denoising can be achieved by discarding some high-frequency components. DFA is a scaling analysis method to study long-range dependency for the non-stationary signal. DFA is utilized as a reliable metric to determine noisy IMFs resulting from the EMD [23]. Therefore, in this paper, DFA is used to distinguish whether noise or signal is the main component of the IMFs.

For a given time series $\{v(i), i = 1, 2, \ldots, N\}$ with the length N, the main procedure of the DFA method is expressed as follows :

1. Integrating the time series of the mean detached signal and the cumulative series $y(n)$ can be computed as follows:

$$y(k) = \sum_{i=1}^{N} [v(i) - \bar{v}], k = 1, 2, \ldots, N \tag{11}$$

where $\bar{v}$ is the average of $v(i)$, $\bar{v} = \frac{1}{N} \sum_{i=1}^{N} v(i)$.

2. The integrated series $y(n)$ was then divided into equal and non-overlapping windows of length n. For each window, the local linear trend $y_n(k)$ is calculated by a polynomial fitting of order l. The root mean square fluctuation $F(n)$ is given by the equation

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} [y(k) - y_n(k)]^2} \tag{12}$$

3. The scaling exponent α is calculated as a slope of the curve $log(F(n))/log(n)$ as follows:

$$F(n) \propto n^{\alpha} \tag{13}$$

The scaling exponent α is related to the Hurst exponent H, which reflects the correlation of time series. The value $\alpha = 0.5$ indicates that the time series is uncorrelated (white noise). $0 < \alpha < 0.5$ indicates that the time series is noise, which has short-range correlation and poor self-similarity between local and global data sequences. $\alpha > 0.5$ represents the time series signal, which has long-range correlation.

3.4. Brief Description of the LOWESS Method

When processing lidar echo signals, it is expected to be less sensitive to photon fluctuations mainly caused by shot noise without distorting the signal. The LOWESS algorithm, first proposed by Cleveland in 1979 [24], effectively solves this problem. With the LOWESS method, the value of each point is an iterative weighted least square regression through a nearby data point within a specified span. The steps of the LOWESS method are as follows:

1. Estimating the span width, which usually contains more than 13 data points and the initial weight of points (x_i, y_i) within the span width is defined as a cubic weighting function as follows:

$$w_i = \left\{ \begin{array}{c} \left(1 - \left|\frac{x-x_i}{d(x)}\right|^3, \left|\frac{x-x_i}{d(x)}\right| \leq 1\right) \\ 0, \left|\frac{x-x_i}{d(x)}\right| > 1 \end{array} \right\} \tag{14}$$

where $d(x)$ is the furthest distance between x and x_i in the specified span width.

2. The first-order polynomial fit with the formula $y = ax + b$ is used to estimate the value of $\hat{y}$ at x. The slope a and constant b are defined as:

$$a = \frac{\sum w_i (x - \overline{x}) \cdot (y - \overline{y})}{\sum w_i^2 (x - \overline{x})^2} \tag{15}$$

$$b = \overline{y} - a\overline{x} \tag{16}$$

where $\overline{x}$ and $\overline{y}$ are the weighted averages of x and y, respectively.

3. The robust weight function is defined to calculate the new weight using the residual of the estimated formula. The robust weighting is:

$$\delta_i = \left\{ \begin{array}{c} \left(1 - |e_i|^2\right)^2, e_i \leq 1 \\ 0, otherwise \end{array} \right\} \tag{17}$$

where $e_i = \frac{r_i}{6Median(|r_1|,|r_2|,\cdots,|r_n|)}$ and $r_i = y_i - \hat{y}_i$ is the error of each y value.

4. Repeating steps 2–3 with the new weighting so that the smoothed value of any point (x, y) could be obtained after several cycles. The theoretical cycles were repeated less than four times.

3.5. Proposed WT-EEMD-LOWESS Algorithm

A new denoising method, called the WT-EEMD-LOWESS method, for Rayleigh lidar echo signals with a large SNR range is proposed. Figure 3 shows the flowchart of the WT-EEMD-LOWESS algorithm and the specific steps is expressed as follows:

1. The echo signal is divided into two segments according to Equation (20), and the high SNR signals are imported into the WT method for denoising.
2. The low SNR signals are decomposed by EEMD and the scaling exponent α is calculated through the DFA algorithm to select IMFs for the signal reconstruction. Then a LOWESS algorithm is performed on the reconstructed signal for further denoising.
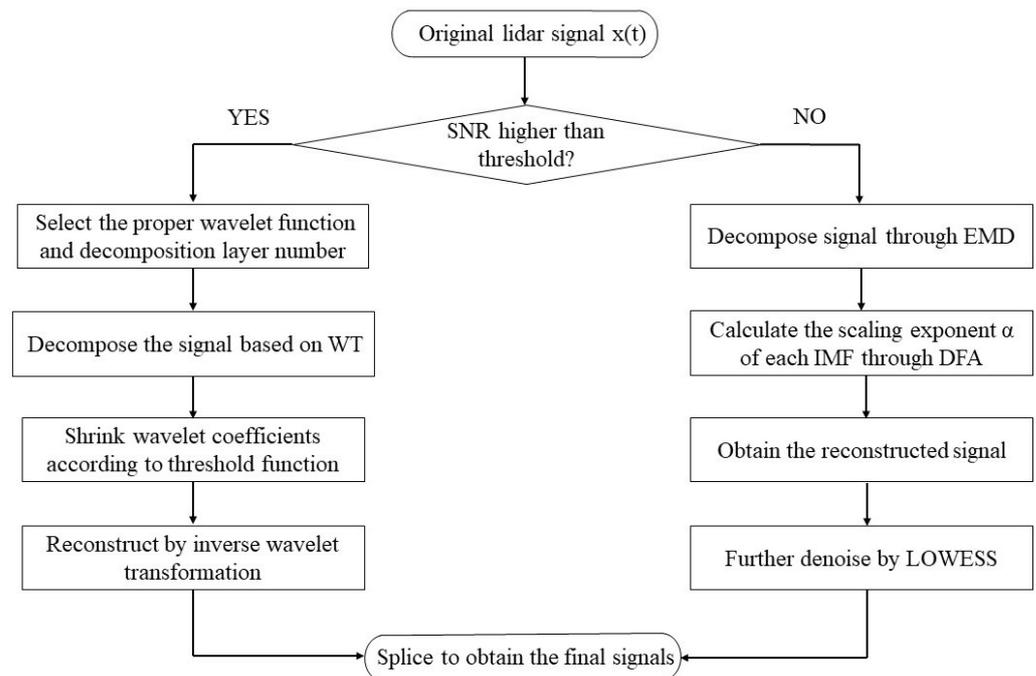3. The two denoised signals are spliced to obtain the final signals.

Figure 3. Flowchart of the proposed WT-EEMD-LOWESS algorithm.

4. Results and Discussion

4.1. Experiments with Simulated Signals

In order to evaluate and compare the performance of the proposed method and the other signal processed methods, the evaluation indexes were set as signal-to-noise ratio (SNR) and the root mean square error (RMSE).

$$SNR = 10\lg\left(\frac{\sum\limits_{n=1}^{N} f^2(n)}{\sum\limits_{n=1}^{N} [f'(n) - f(n)]^2}\right) \qquad (18)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{n=1}^{N} [f'(n) - f(n)]^2} \qquad (19)$$

where $f(n)$ is the original signal, $f'(n)$ is the denoised signal and N is the length of the signal.

In engineering, the SNR of the lidar signal that needs to be processed often has a large dynamic range, and single denoising methods do not have good performance for the whole echo profile. So we plan to divide the signal to be processed into two parts according to the SNR and explore which method is more suitable in the case of integration time of 1200 s. The denoising performance of different algorithms, namely, moving averages (MA), wavelet hard thresholding (WTH), wavelet soft thresholding (WTS), EMD combined with DFA (EMD), EEMD combined with DFA (EEMD) and VMD combined with whale optimization algorithm (VMD-WOA) were compared for different evaluation indexes. As shown as Table 2, the WT method could achieve better results when dealing with high SNR signal, while the decomposition algorithm had good performance when processing low SNR signal.

Table 2. Denoising performance with different heights for each algorithm.

Height Range/km	RAW	MA	WTH	WTS	EEMD	VMD-WOA
30–40	59.7109/1233.5	38.9091/1352.8	58.1694/1473.0	**60.7532/1094.0**	57.1453/1657.4	23.8776/7634.9
30–50	60.1029/837.7	34.2279/1647.4	60.9346/761.2	**62.7611/616.8**	56.6552/837.7	22.4360/6403.4
30–60	58.9623/780.6	32.1997/1700.4	59.1537/763.6	**60.3729/663.6**	55.7952/1124.0	22.4334/5234.4
30–70	59.2095/657.3	31.2823/1637.3	59.6589/624.2	**61.2158/521.7**	55.5927/996.8	22.4416/4530.8
40–70	42.6718/349.4	34.0989/937.5	44.6919/276.9	**47.8397/192.7**	47.2425/206.4	24.6075/2795.9
50–70	28.4629/297.1	35.5781/130.9	32.1896/193.4	**35.8858/132.5**	35.0250/139.6	26.3373/379.4
60–70	16.3705/282.9	26.1785/91.4	18.1742/229.8	20.8954/167.9	23.7415/121.1	**26.1806/91.4**

In order to make the algorithm as simple and efficient as possible on the basis of the above conclusions, we continued to explore the maximum height range that the WTS method can effectively denoise under the condition that the absolute error of density retrieval was less than 5% and the absolute error of temperature retrieval was less than ± 10 K by the CH method [25]. The following paragraphs were also calculated under this condition. Here, the WT method used the "$dB4$" wavelet basis and had three decomposition layers. Since the lidar echo signal obey Poisson distribution so that the signal are random, the mean value of the absolute errors of 100 profiles determined denoising performance of the following paragraphs. As shown in Figure 4, when the maximum detection range was 59 km and the SNR_m of the echo signal was generally above 16 dB, the WTS method could obtain valid data with the maximum absolute error of density retrieval of 1.02% and the maximum absolute error of temperature retrieval of 8.36 K.
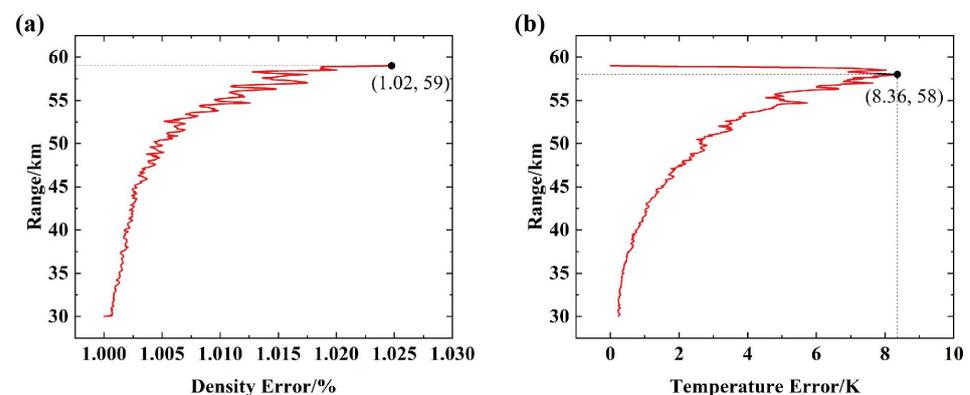
Figure 4. The absolute error of (**a**) density and (**b**) temperature retrieval with the WT method.

It should be noted that the calculation of the SNR here is different from the previous one, due to the ideal signal not being able to be acquired in actual measurement. Considering the statistical properties of the detector, the SNR under Poisson distribution is defined as the ratio of the target backscatter echo signal to the square root of the original echo signal, where the original echo signal includes the target backscatter echo signal and the background noise. Usually, after a certain distance, the background counts of the lidar do not change with height, then it is considered that the photon counts collected by lidar is only the background noise. Thus, the average value of the photon counts after the certain distance can be taken as the background noise. The expression of SNR_m is as follows:

$$SNR_m = \frac{P_{raw}(z) - BG}{\sqrt{P_{raw}(z)}} \qquad (20)$$

where $P_{raw}(z)$ is the original echo signal and BG is the background noise. Figure 5 shows the change SNR_m with height.
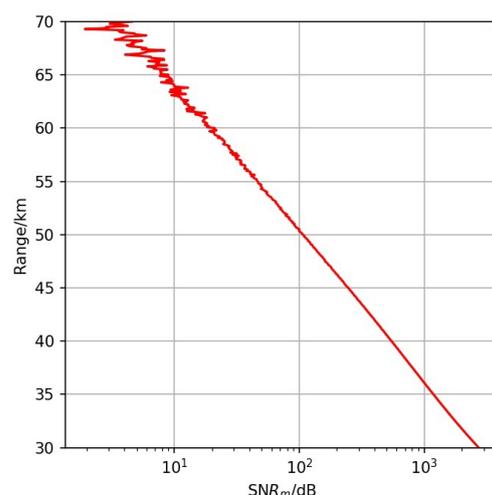
Figure 5. The SNR_m changes with height.

Taking the simulated signal with the SNR_m below 16 dB in Section 2 as an example, the DFA method was used to calculate the scaling exponent α_i of of each IMF for verification, and the results are given in Table 3. Here, the length of non-overlapping windows was an eighth of the length of the input signal. The scaling exponent of $IMF_1 - IMF_3$ were all less than 0.5, so they were considered as noise; The scaling exponent of $IMF_4 - IMF_6$ were all higher than 0.5, so they were considered as signal.

Table 3. The scaling exponent α of each IMF.

	IMF_1	IMF_2	IMF_3	IMF_4	IMF_5	IMF_6
α	0.19	0.12	0.42	**1.00**	**1.48**	**1.99**

For the purpose of validating the accuracy of the DFA method in distinguishing whether noise or signal was the main component of the IMFs, it was accumulated layer-by-layer from IMF_2 to IMF_6. Meanwhile, SNR and RMSE were calculated, respectively. Comparing the original signal with the accumulated signal shown in Table 4, it was found that the accumulated signal of 4–6 layer IMFs was the closest to the original signal with the largest SNR and the smallest RMSE. This result is consistent with the results in Table 3, proving that the DFA method could be used as a reliable algorithm for qualification between noise and signal.

Table 4. Denoising performance of layer-by-layer accumulating of IMFs.

	1–6 IMFs	2–6 IMFs	3–6 IMFs	4–6 IMFs	5–6 IMFs	6th IMF
SNR/dB	18.22	21.43	25.11	**28.88**	25.00	24.73
RMSE	286.97	198.32	129.87	**84.11**	131.53	135.67

To explore which decomposition algorithm was more suitable for low SNR_m signal, the EMD, EEMD, CEEMDAN, and VMD-WOA algorithms were selected for comparison. In the experiments with simulated signal, it was found that since the echo signals obey Poisson distribution and have a certain randomness, there was no method that always had the best denoising performance, which was also related to the selected SNR threshold. So as to solve this problem and find a generally suitable denoising method, the average SNR (SNR_{ave}) was used as the evaluation index to determine the denoising performance on the basis of multiple profiles. When selecting the signal with SNR_m below 16 dB, after 100 statistics, the EEMD algorithm showed the best performance in 55 of them, and ranked first with 28.56 dB as shown in Table 5. For EEMD, the standard deviation of Gaussian white noise was 0.1, and Gaussian white noise was added 50 times.

**Table 5.** Denoising performance of different algorithms for low SNR_m signals.

	RAW	EMD	EEMD	CEEMDAN	VMD-WOA
SNR_{ave}/dB	18.56	27.34	**28.56**	27.71	26.47
Best times		12	**55**	18	15

Based on the above findings, we proposed to combine EEMD and LOWESS to achieve better denoising performance. Table 6 shows that the proposed EEMD-LOWESS outperformed all the other proposed methods [26] which could retain more effective signal. In addition, it was found that it was more effective at removing noise in low-frequency IMFs than extracting signal in high-frequency IMFs.

Table 6. Denoising performance of proposed algorithms compared with other methods.

	RAW	EEMD	EEMD-NLM	EEMD-WTS	EEMD-SVD	EEMD-LOWESS
SNR/dB	18.56	28.56	20.92	28.97	26.46	**30.54**

The denoising performance of the WT-EEMD-LOWESS algorithm is shown in Figure 6. The echo signals after denoising were smoother, more stable and closer to the ideal signal. The denoised data obtained through the above process were used as the input data of the CH method, and the density and temperature retrieval of the simulated signal could be gained, as presented in Figure 7a,b and then in contrast with the model temperature, the reliability of the proposed algorithm could be evaluated. It was not practical to calculate the absolute value of density due to several parameters fluctuating with measurement, so the relative density was calculated instead. As shown in Figure 8, under the condition that the error of density and temperature retrieval does not exceed 5% and 10 K, the maximum detection range was 67.5 km and 58.4 km after denoising, which were 12.8 km and 11.1 km higher than those without denoising, respectively.
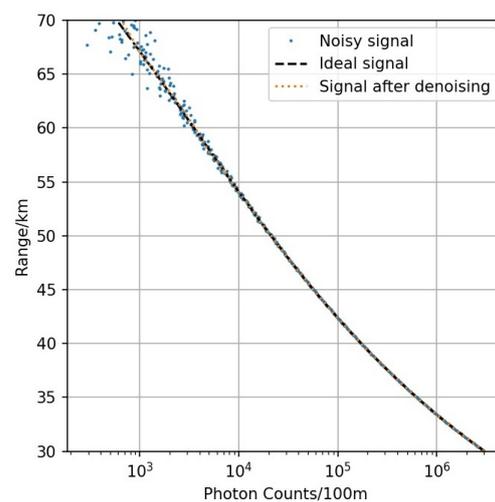
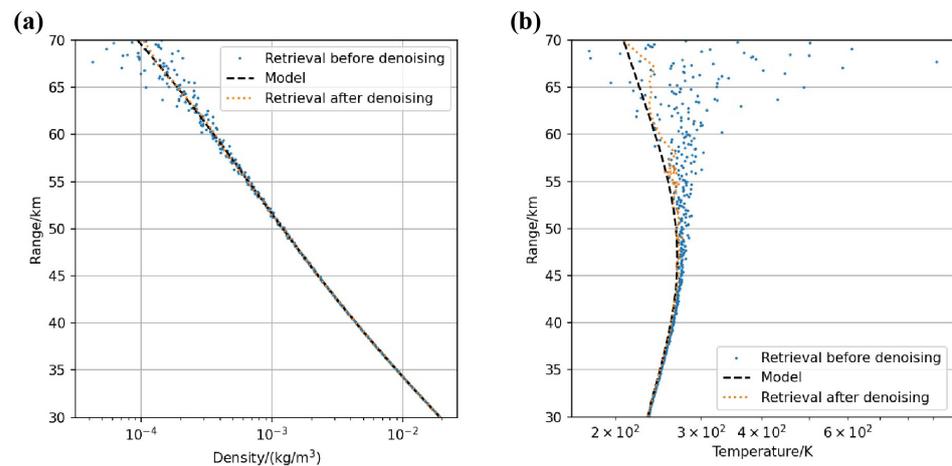**Figure 6.** Denoising performance of WT-EEMD-LOWESS algorithm.

Figure 7. The variation of (**a**) density and (**b**) temperature retrievals with WT-EEMD-LOWESS algorithm.
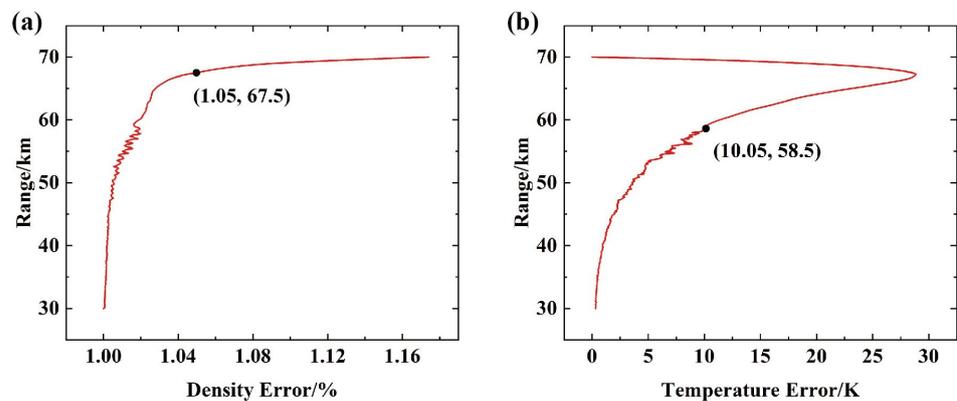

Figure 8. The absolute error of (**a**) density and (**b**) temperature retrievals with the WT-EEMD-LOWESS method.

Futhermore, the algorithm was applied to simulated the lidar echo signal with integration times of 300–3000 s. The variation of the maximum detection range with the integration time is shown in Figure 9. It could be found that the maximum detection range increased with the accumulation of the integration time, which was conformed to the law. When the error of density retrieval was less than 5%, the maximum detection range increased first fast and then slowly with the integration time. The reason is that when the integration time was too long, the detection range was close to the limit of 70 km. The maximum detection range changed slowly with the integration time at 60 km when the error of temperature retrievals was less than 10 K. The reason is that the measurement was inaccurate within 10 km below the seed temperature, which is an inherent shortcoming of the CH method.
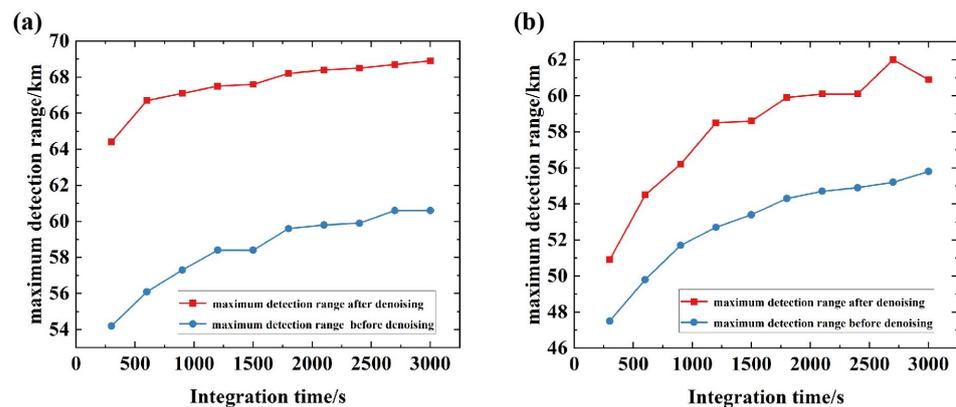
(a)

(b)

Figure 9. The variation of the maximum detection range with the integration time under the condition that (**a**) the error of density retrieval was less than 5% and (**b**) the error of temperature retrieval was less than 10 K.

4.2. Real Experiment on a Lidar Echo Signal

In order to further investigate the denoising performance of the algorithm proposed in this paper, we used the measurement data provided by the National Space Science Center, CAS. The signal was obtained by continuously observing the atmosphere at 40°18′N latitude and 116°40′E longitude by ground-based lidar and the parameters of lidar system are shown in Table 7 below. The observation time was 3 September 2018, 17:30. The measurement results and denoising performance are shown in Figure 10.

Table 7. Ground-based lidar system.

System Parameter	Value
Laser pulse energy/mJ	500
Telescope diameter/m	1
Average power/W	>10
Wavelength/nm	532
Pulse width/ns	8
Quantum efficiency	0.4
Filter transmittance	0.8
System optical efficiency	0.02
Pulse repetition frequency/Hz	30
Degree from zenith/°	0

Based on the denoised echo signal, the density and temperature retrievals of the measured data can be obtained by the CH method, as shown in Figure 11. The maximum error between density retrieval and model was reduced from 1.36% to 1.1%, and the maximum error of temperature retrieval and model was reduced from 125.79 K to 13.84 K. The improvement of the maximum detection distance was not obvious with error of density and temperature retrievals less than 5% and 10 K separately, due to the fact that the measured original signal is different from the model when the SNR is high. It can also be seen that when the SNR is low, the performance improvement of the proposed algorithm is more obvious, and when the SNR is high, the discrepancy between the retrievals and model depends more on the original data itself.
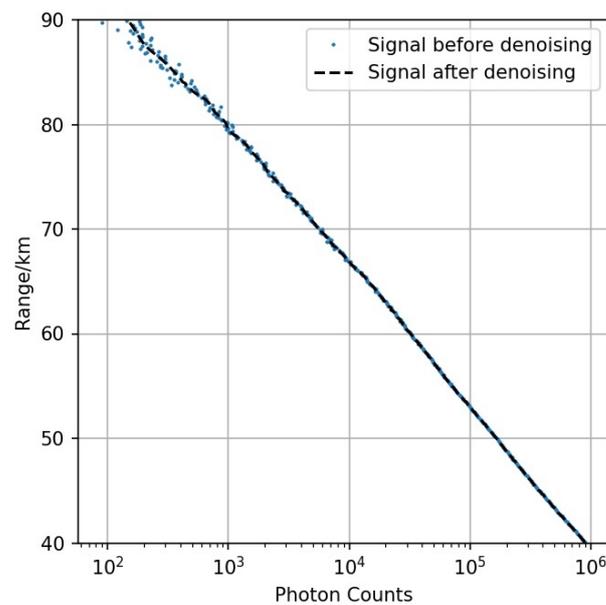
Figure 10. The noisy echo signal and the signal after denoising of ground-based lidar.
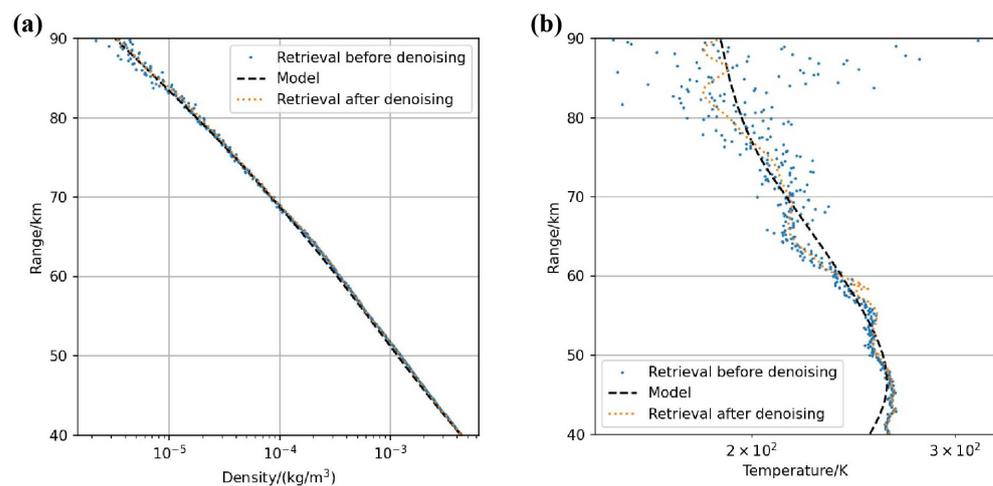

Figure 11. The variation of (**a**) density and (**b**) temperature retrievals with WT-EEMD-LOWESS algorithm of ground-based lidar.

5. Conclusions

Based on WT and EEMD, a new WT-EEMD-LOWESS denoising method for signal with large SNR dynamic range was proposed by using the simulated lidar signal. The echo signal was divided into two parts with SNR of 16 dB as the boundary. The high SNR part was denoised by the WT method; the low SNR part was decomposed by the EEMD algorithm first, and the scaling exponent α of each IMF was calculated by DFA, then the IMF with $\alpha > 0.5$ was used to obtain the reconstructed signal. The reconstructed signal still contained noise, so the LOWESS method was performed on the reconstructed signal to futher extract the signal. Moreover, in order to verify the effectiveness of the algorithm, the simulated denoised signal with a typical integration time of 1200 s was used as the input data of the CH method for density and temperature retrievals. In the case that the absolute error of density retrieval was less than 5% and the absolute error of temperature retrieval was less than 5 K, the maximum detection range improved 12.8 km and 11.1 km, respectively. This algorithm was also verified to be effective with an integration time of 300–3000 s. In addition, the WT-EEMD-LOWESS method was applied to the measured lidar echo signal. The experimental results showed that the algorithm achieved a good denoising

performance in the low SNR signal part on the basis of not distorting the signal. The maximum density error was reduced from 1.36% to 1.1% and the maximum temperature error was reduced from 125.79 K to 13.84 K which indicated that the proposed algorithm has good application value in Rayleigh scattering lidar signal processing.

Author Contributions: Conceptualization, Y.Z. and D.X.; methodology, Y.Z. and K.Z.; software, Y.Z. and X.Z.; validation, Y.Z. and T.W.; investigation, Y.Z.; data curation, Z.Y.; writing—original draft preparation, Y.Z.; writing—review and editing, K.Z. and T.W.; visualization, Y.S., Y.W., S.L. and X.L.; project administration, K.Z. and J.Y.; funding acquisition, K.Z. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, Z.P.; Huang, X.; Cao, Y.; Wang, B.; Li, Y.H.; Jin, W.; Yu, C.; Zhang, J.; Zhang, Q.; Peng, C.Z.A. Single-photon computational 3D imaging at 45 km. *Photon. Res.* **2020**, *8*, 1532–1540. [CrossRef]
2. Bu, L.; Wang, Q.; Xu, J.; Zhu, S.; Tian, L. Validation of an Airborne High Spectral Resolution Lidar and Its Measurement for Aerosol Optical Properties over Qinhuangdao, China. *Opt. Express* **2020**, *28*, 24471–24488.
3. Xue, W.; Liu, L.; Dai, X.; Luo, Y. Moving target ranging method for a photon-counting system. *Opt. Express* **2018**, *26*, 34161–34178. [CrossRef]
4. Kaifler, B.; Rempel, D.; Roi, P.; Büdenbender, C.; Baturkin, V. A technical description of the Balloon Lidar Experiment BOLIDE. *Atmos. Meas. Tech.* **2020**, *13*, 5681–5695. [CrossRef]
5. Pan, D.; Zhang, T.; Chen, W.; Liu, J. Measurements of density, pressure and temperature in the middle atmosphere with Rayleigh lidar. In Proceedings of the International Symposium on Optoelectronic Technology and Application, Beijing, China, 9–11 May 2016.
6. Jalali, A.; Sica, R.J.; Haefele, A. Improvements to a long-term Rayleigh-scatter lidar temperature climatology by using an optimal estimation method. *Atmos. Meas. Tech.* **2018**, *11*, 6043–6058. [CrossRef]
7. Khanna, J.; Bandoro, J.; Sica, R.J.; Mcelroy, C.T. New technique for retrieval of atmospheric temperature profiles from Rayleigh-scatter lidar measurements using nonlinear inversion. *Appl. Opt.* **2012**, *51*, 7945–7952. [CrossRef]
8. Chang, J.; Zhu, L.; Li, H.; Fan, X.; Yang, Z. Noise reduction in Lidar signal using correlation-based EMD combined with soft thresholding and roughness penalty. *Opt. Commun.* **2018**, *407*, 290–295. [CrossRef]
9. Fang, H.T.; Huang, D.S. Noise reduction in lidar signal based on discrete wavelet transform. *Opt. Commun.* **2004**, *233*, 67–76. [CrossRef]
10. Sarvani, M.; Raghunath, K.; Rao, S. Lidar signal denoising methods- application to NARL Rayleigh lidar. *J. Opt.* **2015**, *44*, 164–171. [CrossRef]
11. Han, Y.; Westwater, E.R.; Ferrare, R.A. Applications of Kalman Filtering to Derive Water Vapor Profiles from Raman Lidar and Microwave Radiometers. In Proceedings of the International Geoscience & Remote Sensing Symposium, Honolulu, HI, USA, 25–30 July 2010.
12. Mao, J. Noise reduction for lidar returns using local threshold wavelet analysis. *Opt. Quantum Electron.* **2012**, *43*, 59–68. [CrossRef]
13. Zhou, Z.; Hua, D.; Wang, Y.; Yan, Q.; Li, S.; Yan, L.; Wang, H. Improvement of the signal to noise ratio of Lidar echo signal based on wavelet de-noising technique. *Opt. Lasers Eng.* **2013**, *51*, 961–966. [CrossRef]
14. Boudraa, A.O.; Cexus, J.C. EMD-based Signal Filtering. *IEEE Trans. Instrum. Meas.* **2007**, *56*, 2196–2202. [CrossRef]
15. Gong, W.; Li, J.; Mao, F.; Zhang, J. Comparison of simultaneous signals obtained from a dual-field-of-view lidar and its application to noise reduction based on empirical mode decomposition. *Chin. Opt. Lett.* **2011**, *9*, 4.
16. Tian, P.; Cao, X.; Liang, J.; Zhang, L.; Yi, N.; Wang, L.; Cheng, X. Improved empirical mode decomposition based denoising method for lidar signals. *Opt. Commun.* **2014**, *325*, 54–59. [CrossRef]
17. Zhaohua, W.U.; Huang, N.E. Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41.
18. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]
19. Li, H.; Chang, J.; Fan, X.; Liu, Z.; Liu, B. Efficient Lidar Signal Denoising Algorithm Using Variational Mode Decomposition Combined with a Whale Optimization Algorithm. *Remote Sens.* **2019**, *11*, 126. [CrossRef]
20. Sun, B.Y.; Huang, D.S.; Fang, H.T. Lidar signal denoising using least-squares support vector machine. *IEEE Signal Process. Lett.* **2005**, *12*, 101–104.
21. Sreekanth, V.S.; Raghunath, K.; Mishra, D. Dictionary learning technique and penalized maximum likelihood for extending measurement range of a Rayleigh lidar. *J. Appl. Remote Sens.* **2021**, *14*, 034529. [CrossRef]

22. Berk, A.; Conforti, P.; Kennett, R.; Perkins, T.; Hawes, F.; van den Bosch, J. MODTRAN® 6: A major upgrade of the MODTRAN® radiative transfer code. In Proceedings of the 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lausanne, Switzerland, 24–27 June 2014; pp. 1–4. [CrossRef]

23. Mert, A.; Akan, A. Detrended fluctuation analysis for empirical mode decomposition based denoising. In Proceedings of the 2014 22nd European Signal Processing Conference (EUSIPCO), Lisbon, Portugal, 1–5 September 2014.

24. Cleveland, W.S. Robust Locally Weighted Regression and Smoothing Scatterplots. *J. Am. Stat. Assoc.* **1979**, *74*, 829–836. [CrossRef]

25. Alain, H.; Marie-Lise, C. Density and temperature profiles obtained by lidar between 35 and 70 km. *Geophys. Res. Lett.* **1980**, *7*, 565–568.

26. Cheng, X.; Mao, J.; Li, J.; Zhao, H.; Rao, Z. An EEMD-SVD-LWT algorithm for denoising a lidar signal. *Measurement* **2021**, *168*, 108405. [CrossRef]