Citation: Xiang, C.; Zheng, Y.; Liang, A.; Li, R. Echo-Signal De-Noising of CO\textsubscript{2}-DIAL Based on the Ensemble Empirical Mode Decomposition. Atmosphere 2022, 13, 1361. https://doi.org/10.3390/atmos13091361

Received: 8 August 2022
Accepted: 23 August 2022
Published: 25 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Check for updates

Abstract: The carbon dioxide (CO\textsubscript{2}) differential absorption lidar echo signal is susceptible to noise and must satisfy the high demand for signal-retrieval precision. Thus, a proper de-noising method should be selected to improve the inversion result. In this paper, we simultaneously decompose three signal pairs into different intrinsic mode functions (IMFs) using the method of ensemble empirical mode decomposition (EEMD). Further, the correlation coefficients of the IMFs with the same temporal scale are regarded as the criterion to determine the components that need removal. This method not only retains the useful information effectively but also removes the noise component. A significant improvement in the R\textsuperscript{2} of the differential absorption optical depth (DAOD) of the de-noised signals is obtained. The results of the simulated and observed analysis signal demonstrated improvement both in the SNR and in the retrieval precision.

Keywords: CO\textsubscript{2}-DIAL; echo-signal de-noising; ensemble empirical mode decomposition

1. Introduction

Since the Industrial Revolution, the concentration of atmospheric carbon dioxide (CO\textsubscript{2}) has continued to rise due to the massive combustion of fossil fuels [1,2]. Research has shown that part of the CO\textsubscript{2} generated by human activities has inexplicably disappeared, which indicates that some unknown carbon sinks may exist [3]. To determine the distribution of carbon sources and sinks, many CO\textsubscript{2} observation methods have been presented [4–9]. Differential absorption lidar (DIAL) is considered the most promising method for observing CO\textsubscript{2} concentration [10]. In contrast to passive remote sensing, DIAL can provide better precision to retrieve atmospheric CO\textsubscript{2} concentration [11]. Moreover, a ground-based DIAL can provide vertical profiles of CO\textsubscript{2}, which can be of great significance to the research of the carbon cycle [12–14].

The echo signal of a ground-based CO\textsubscript{2}-DIAL mainly relies on the scattering of aerosol and atmospheric molecules [15]. Therefore, the signal intensity is relatively weak. Because a lidar signal is susceptible to noise such as background light, dark current, and thermal noise, any small perturbation will result in poor signal inversion [16]. Moreover, because the signal-to-noise ratio (SNR) rapidly decreases with the increase in range, useful information will be overwhelmed by the noise over long distances [17]. Signal de-noising of the CO\textsubscript{2}-DIAL will inevitably improve the effective detection range and will play an important role in the signal inversion precision. Therefore, a proper de-noising method must be selected for an important performance.

Until now, several methods have been developed, and some are in the development process for application in lidar signal de-noising [18–21]. The multi-pulse average method has been widely used in signal de-noising by calculating the mean of the multiple sets of profiles. However, multiple averages smoothen the fast-changing components of aerosols. The wavelet transform (WT) has been proven quite useful in extracting signals from data generated in noisy, nonlinear, and nonstationary processes. Zhou et al. applied it to reduce...
lidar echo signal noise and improve the system SNR [22]. WTs are founded on basis functions, which inevitably have poor adaptability. Mao et al. put forward an alternative approach to obtain a de-noised signal as a by-product of combining the ensemble Kalman filter and the Fernald method [23].

Wu et al. and Tian et al. applied the empirical mode decomposition (EMD) to lidar signal de-noising [24,25]. The EMD developed by Huang et al. decomposed the signal on the basis of the local characteristic time scale of the signal itself and dispensed with any basis functions [26]. As discussed by Huang et al., a major drawback in the original EMD is the frequent appearance of mode mixing, which makes the physical meaning of the individual intrinsic mode functions (IMFs) unclear. To overcome this problem, a new noise-assisted data analysis (NADA) method is proposed, namely the ensemble EMD (EEMD), which adds a white noise with finite amplitude into the signal and performs EMD [27,28].

In the present study, we assume that the signals in a short time have the same useful information and that their differences are only caused by noise in stable atmospheric conditions. We apply the EEMD method to decompose the signals into different IMFs with frequencies ranging from high to low. Generally, we consider the noise existing in the high-frequency parts and implement signal de-noising by removing them. The correlation coefficients of the IMFs with the same temporal scale are regarded as the criterion to determine the components that need removal. Finally, we regard the linear fitting parameters of the differential absorption optical depth (DAOD) as the criterion for evaluating the de-noising effect.

In the next section, the system and theoretical foundation of the CO$_2$-DIAL and the EEMD method are presented in detail. Further, we study the feasibility of the method by processing simulated signals. Two parts were selected for the analysis, and an obvious improvement in the result was obtained. Moreover, we apply the above-mentioned method to the observed signal, and the r-square increased from 0.444 to 0.755, which is doubtless beneficial to the inversion precision. In conclusion, the method turns out to be applicable to simulated and observed signals.

2. Data and Methods

2.1. System and Theoretical Foundation of CO$_2$-DIAL

Our group is devoted to the development of a ground-based CO$_2$-DIAL system. The system located in Wuhan University involves four parts, namely, laser-emitting, frequency-stabilization, optical-receiving, and signal-acquisition systems. The system was designed to work at ~1.57 μm with a pulse repetition frequency of 20 Hz. Two close wavelengths were selected, and they were at the peak and wing of the absorption line, namely, on-wavelength (λ$_{on}$) and off-wavelength (λ$_{off}$), respectively. The sampling frequency is 20 MHz with a vertical range resolution of 7.5 m. The configuration of the lidar system is shown in the Figure 1.

The foundation of the CO$_2$-DIAL is the backscatter Lidar equation, as expressed in the following:

$$P_{on,off}(R) = \frac{P_0 A \beta(R) \eta c \tau}{2R^2} \cdot \exp\left(-2 \int_0^R (\alpha(R) + N\sigma_{on,off}(R)) dr\right)$$  \hspace{1cm} (1)$$

where $R$ is the detection range, $P$ is the received power of range $R$, $P_0$ is the transmitted power, $A$ is the effective receiver area, $c$ is the speed of light, $\tau$ is the laser pulse width, $\eta$ is the receiver efficiency, $\beta(R)$ is the atmospheric backscatter coefficient, $\alpha(R)$ is the atmospheric extinction coefficient, $N$ is the number density of CO$_2$, and $\sigma$ is the absorption cross-section of CO$_2$. Because the on-line and off-line wavelengths are very close, $\beta(R)$ and $\alpha(R)$ can be considered equivalent at two similar wavelengths. From Han et al., we can ignore the effect of water vapor through strict wavelength selection [15]. Given that the inversion of the CO$_2$ concentration is mainly based on the absorption difference of the
on-line and off-line wavelength, a slight noise may cause a poor result. The signal intensity exponentially attenuates with range. Therefore, the amplitude of the signal affected by noise considerably varies at different ranges.

\[ y = \sum_{i=1}^{n} c_i + r \]  

(2)

where \( y \) is the signal, \( c_i \) represents the IMFs that include different frequency bands ranging from high to low, and \( r \) is the final residue that indicates the trend of the signal. Compared with previous methods, EMD requires no preliminary knowledge and possesses better adaptation. However, the method still suffers from the drawback of mode mixing, which makes the physical meaning of the IMFs unclear. To solve this problem, Wu and Huang [28] proposed a new NADA method named EEMD, which adds a white noise series to the targeted data and then decomposes the data into IMFs. By repeating the above-mentioned process with different white noise series every time, we can obtain multiple IMFs with the same temporal scale. Ultimately, we treat the mean of the corresponding IMFs as the final result.

The EEMD algorithm can be directly defined by the following steps:

1. Add random white noise \( n_i(t) \) with magnitude \( \varepsilon \) to the original signal \( s(t) \) to generate a new signal \( y_i(t) \) as

\[ y_i(t) = s(t) + \varepsilon n_i(t) \]  

(3)

2. Use the EMD method to decompose the new generated signal \( y_i(t) \) as

\[ y_i(t) = \sum_{j=1}^{m} c_{ij} + r_{i,m} \]  

(4)

3. Calculate the mean of the corresponding IMFs and the residues as the final result as

\[ c_j = \frac{1}{n} \sum_{i=1}^{n} c_{ij} \]  

(5)

Figure 1. Configurations of ground-based CO₂-DIAL system.

2.2. EEMD

The EMD proposed by Huang et al. is a non-linear, non-stationary, and self-adaptive decomposition technology that, unlike the traditional methods (Fourier transform, WT, etc.), dispenses with any basis functions and can decompose the signal into a series of IMFs.
and

\[ r_m = \frac{1}{n} \sum_{i=1}^{n} r_{i,m} \]  \hspace{1cm} (6)

where \( y'_t(t) \) is the new generated signal, \( s(t) \) is the original signal, \( n_i(t) \) stands for different white noise series, \( c_{ij} \) represents the IMFs, \( c_i \) is the mean of the corresponding IMFs of the decompositions, \( n \) is the ensemble number, \( r_{i,m} \) is the residue after \( n \)-times of decomposition, and \( r_n \) is the mean of the residues. By calculating the means of the IMFs and the residues, the added white noise series will finally cancel each other, and the problem of mode mixing can very well be solved.

As explained above, a series of IMFs with frequency ranging from high to low can be obtained. We generally consider the noise to exist in the high-frequency component. By removing the first few IMFs, we can achieve signal de-noising. Then, the key problem is how many IMFs should be removed.

In this work, we consider that the signals within a short time are similar to one another. In other words, the useful information in these signals is the same. Therefore, we simultaneously select several echo signals to decompose by EEMD and calculate the correlation coefficient of the IMFs with the same temporal scale. In the absence of the noise effect, the IMFs with the same temporal scale will have a strong correlation; otherwise, they have lesser relativity. According to the correlation coefficient, we can determine the number of removed IMFs.

The methodology shown in Figure 2 consists of two parts: signal de-noising and performance evaluation. The input matrix \( Y \) represents the original data, which includes three signal pairs. The first part starts by decomposing each segment into different time scales, \( c_{ij}^s \) (\( s = 1, 2, \ldots, S \)), through EEMD and then calculating the correlation coefficient \( r_{s,j}^R \). Finally, the signal is reconstructed on the basis of the threshold. The second part mainly evaluates the de-noising performance by calculating the DAOD.

![Figure 2](image)

**Figure 2.** Methodology of the CO\(_2\)-DIAL signal de-noising: (a) shows the signal de-noising by EEMD, and (b) shows the performance evaluation.

3. Simulated Signal Analysis

To verify the feasibility of the method, we analyzed the simulated CO\(_2\)-DIAL signal. Figure 3a shows the pure signal of the on-line and off-line wavelengths. The initial CO\(_2\) concentration was assumed to be 400 parts per million, and the temperature and pressure profiles were obtained by interpolating the sounding data of the temperature and pressure at Wuhan in March 2015. Compared with the off-line wavelength, the on-line wavelength fell even more rapidly due to CO\(_2\) absorption. We introduced a white noise into the simulated pure signal and considered their combination as the observed signal [Figure 3b]. Because of the influence of noise, the intensity of on-line wavelength echo signal may be stronger than off-line wavelength and can impair the absorption phenomenon. Thus, this problem can influence CO\(_2\) retrieval precision. Hence, we introduced the above-mentioned method to process the echo signal.
Figure 3. Simulated CO$_2$–DIAL signal. (a) Pure signal of the on-line and off-line wavelengths. (b) Signal with white noise.

We proposed false SNR (FSNR) to approximately estimate the real SNR (RSNR) of the signal.

$$FSNR = 10 \times \log \frac{\sum_{i=1}^{n} r^2}{\sum_{i=1}^{N} (C_1)^2} \quad (7)$$

and

$$RSNR = 10 \times \log \frac{\sum_{i=1}^{n} S_p^2}{\sum_{i=1}^{N} (S_n - S_p)^2} \quad (8)$$

where $r$ is final residue and can be viewed as useful information of the signal, $C_1$ is the first IMF and can approximately represent the noise of the signal, $S_p$ and $S_n$ represent the pure signal and noise of a signal, respectively. As Figure 4 shows, we have calculated the FSNRs and RSNRs of ten simulated signals. The difference between FSNR and RSNR is smaller, so we can substitute FSNR for RSNR.

Figure 4. FSNR and RSNR of simulated signals.

FSNRs of ten measured signals have been calculated and they have minor difference between any two adjacent signals (Table 1). Therefore, three similar simulated signal groups with different SNRs, (99,100,101), were selected as our research objects (Figure 5). They can be considered as observed signals collected within a short time in stable atmospheric conditions. We segmented the signals into two parts to verify the performance of the method under different signal intensities. One part was approximately from 300 to 1500 m, and the other was from 1500 to 3000 m.
Table 1. The differences of FSNRs of measured signals.

| Number | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       |
|--------|---------|---------|---------|---------|---------|---------|---------|---------|
| FSNR   | 97.198  | 96.637  | 98.711  | 96.151  | 97.165  | 98.817  | 99.130  | 99.419  |
| Differences of FSNRs | 0.561   | 2.074   | 2.560   | 1.014   | 1.653   | 0.313   | 0.313   | 0.289   |
| Mean of the differences | 1.209   |         |         |         |         |         |         |         |

Figure 5. Similar simulated signals, which can be considered as measured signal collected within a short time: (a) previous time of the signal that needs de–noising; (b) signal that needs de–noising, and (c) following time of signal that needs de–noising.

Figure 6 shows the IMFs of the selected simulated signals of the on-line wavelength from 300 to 1500 m decomposed by EEMD. The frequency of the IMF systematically decreased. We generally consider that IMFs with high frequency contain little useful information, and most of them are occupied by noise. Therefore, we assumed the IMFs with high frequency to be completely submerged in noise. We adopted the method of removing high-frequency IMFs to realize CO₂-DIAL signal de-noising. Therefore, the key to the problem was how to determine the IMFs that need to be removed. To solve this problem, the correlation coefficients of the IMFs with the same temporal scale were calculated.

Figure 6. IMFs of the three selected simulated signals of the on-line wavelength. (a) IMFs of the signal at previous time; (b) IMFs of the signal at a certain time, and (c) IMFs of the signal at following time.
We determined the IMFs that need to be removed and reconstructed the signal according to the following equations:

\[ y_j = \sum_{s=1}^{S} K_s(t) C_{js}(t) \]  

(9)

and

\[ K_s(t) = \begin{cases} 
1 & \text{CR} \geq 0.5 \\
0 & \text{CR} < 0.5 
\end{cases} \]  

(10)

Signal \( y_j \) is reconstructed by several IMFs on individual scale \( C_{js}(t) \) (\( s = 1, 2, \ldots, S \)) with a relevance factor \( K_s(t) \), where CR represents the correlation coefficients of the IMFs with the same temporal scale. We considered that the IMF contains more noise than useful information on the condition that half of all correlation coefficients are less than 0.5. The 0.5 correlation threshold is determined by a statistical experiment. The main idea of the experiment is to determine the threshold when the signal gets the maximum SNR and conduct a count in various thresholds. In order to verify the selection of threshold in various signal quality, we have simulated 50 signals with various SNR (51, 52, \ldots, 100). We use the method proposed in this paper to analyze the signals in the sets of three and it totals 48 groups. The result showed that the signals obtained the maximum SNR when 0.3, 0.4, 0.6, 0.7 were selected as the criterion in some cases (Figure 7). Although sometimes several thresholds can also get the highest SNR, only the 0.5 correlation threshold can apply to almost all of the cases (45 groups in whole 48 groups). Therefore, a 0.5 correlation threshold is optimal.

![Figure 7. Number of groups getting the highest SNR at different thresholds.](image-url)

The following table (Table 2) lists the correlation coefficients of the IMFs with the same temporal scale of the on-line and off-line wavelengths. The correlation coefficients of the first IMF were comparatively low (i.e., below 0.5), whereas those of IMFs 2–7 reached more than 0.5. Therefore, we removed the first IMF and reconstructed the remaining components.

The ultimate purpose of de-noising is to improve the inversion accuracy; thus, we simultaneously analyzed the off-line wavelength.

Figure 8 shows the linear regression of the de-noised signal as a function of the pure signal. The good linear relationship demonstrates that the signals have a high degree of similarity. Meanwhile, the SNR and mean square error (MSE) were calculated (Figure 9). A better result was obtained after removing the first IMF.
Table 2. Correlation coefficients of the IMFs with same temporal scale. (Left) On-line wavelength. (Right) Off-line wavelength.

|       | \(\lambda_{on}\) a | \(\lambda_{on}\) b | \(\lambda_{on}\) c | \(\lambda_{off}\) a | \(\lambda_{off}\) b | \(\lambda_{off}\) c |
|-------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| C1    | a 1.0000            | -0.0595             | -0.0254             | a 1.0000            | 0.0235              | 0.0029              |
|       | b -0.0595           | 1.0000              | -0.0443             | b 0.0235            | 1.0000              | -0.0579             |
|       | c -0.0254           | -0.0443             | 1.0000              | c 0.0029            | -0.0579             | 1.0000              |
| C2    | a 1.0000            | 0.5718              | 0.5765              | a 1.0000            | 0.6571              | 0.6984              |
|       | b 0.5718            | 1.0000              | 0.7691              | b 0.6571            | 1.0000              | 0.7428              |
|       | c 0.5765            | 0.7691              | 1.0000              | c 0.6984            | 0.7428              | 1.0000              |
| C3    | a 1.0000            | 0.9123              | 0.9709              | a 1.0000            | 0.9206              | 0.9902              |
|       | b 0.9123            | 1.0000              | 0.9710              | b 0.9206            | 1.0000              | 0.9067              |
|       | c 0.9709            | 0.9710              | 1.0000              | c 0.9902            | 0.9067              | 1.0000              |
| C4    | a 1.0000            | 1.0000              | 0.9998              | a 1.0000            | 0.9998              | 0.9998              |
|       | b 1.0000            | 1.0000              | 0.9997              | b 0.9998            | 1.0000              | 0.9992              |
|       | c 0.9998            | 0.9997              | 1.0000              | c 0.9998            | 0.9992              | 1.0000              |
| C5    | a 1.0000            | 0.9852              | 0.4979              | a 1.0000            | 0.9609              | 0.4460              |
|       | b 0.9852            | 1.0000              | 0.6376              | b 0.9609            | 1.0000              | 0.6764              |
|       | c 0.4979            | 0.6376              | 1.0000              | c 0.4460            | 0.6764              | 1.0000              |
| C6    | a 1.0000            | 1.0000              | 1.0000              | a 1.0000            | 1.0000              | 1.0000              |
|       | b 1.0000            | 1.0000              | 1.0000              | b 1.0000            | 1.0000              | 1.0000              |
|       | c 1.0000            | 1.0000              | 1.0000              | c 1.0000            | 1.0000              | 1.0000              |
| C7    | a 1.0000            | 0.9999              | 0.9997              | a 1.0000            | 0.9998              | 1.0000              |
|       | b 0.9999            | 1.0000              | 0.9993              | b 0.9998            | 1.0000              | 0.9999              |
|       | c 0.9997            | 0.9993              | 1.0000              | c 1.0000            | 0.9999              | 1.0000              |

From the above discussion, we can conclude that the method works well for on-line or off-line wavelength echo signals. However, evaluating the log ratio of two de-noised signals is more significant, which is crucial to the inversion precision. The following equation is the inversion formula for the CO\(_2\) concentration [29,30]:

\[
N_{\text{CO}_2} = \frac{1}{2\left(\sigma_{on} - \sigma_{off}\right)\Delta R} \ln \frac{P(\lambda_{on}, R_{bottom}) \cdot P(\lambda_{off}, R_{top})}{P(\lambda_{off}, R_{bottom}) \cdot P(\lambda_{on}, R_{top})}
\] (11)

where \(N_{\text{CO}_2}\) represents the concentration of CO\(_2\), \(\sigma\) is the absorption cross-section of CO\(_2\), and \(\Delta R\) is the range resolution.
Figure 8. Comparison between the de-noised and pure signals. (Left) On-line wavelength. (Right) Off-line wavelength (the de-noised signal is reconstructed by IMFs 2–7).

Figure 9. SNR and MSE of the de-noised signal. (Left) On-line wavelength. (Right) Off-line wavelength. The x-axis label indicates the sum of the IMFs and C represents IMF. For example, the first is the sum of IMFs 1–7, the second is the sum of IMFs 2–7, and the last is IMF 7.

In this paper, DAOD refers to the difference between on-line and off-line wavelength in the CO2-DIAL system, which is used to represent the difference between the two laser echo signals caused by CO2 absorption. Then, Equation (11) can also be expressed as the following equation by DAOD:

$$ N_{CO2} = \frac{DAOD}{2\left(\sigma_{on} - \sigma_{off}\right)\Delta R} $$

(12)

Figure 10 shows the DAOD of the pure, original, and de-noised signal and directly shows that the data quality significantly improves and has a better fitting effect. The r-square listed in Table 3 is consistent with this. According to Equation (12), the slope of the DAOD is the ratio of the DAOD and $\Delta R$, which is important for the inversion of the CO2 concentration. By comparing the slopes of the three data types, the slope of the de-noised signal is closer to the expected value, which demonstrates the validity of the de-noising method.

Table 3. Linear fitting parameters of the DAOD for simulated signal from 300 to 1500 m.

|                  | Pure     | Original | De-Noised |
|------------------|----------|----------|-----------|
| $R^2$            | 0.998    | 0.751    | 0.841     |
| Slope            | $1.21 \times 10^{-4}$ | $1.16 \times 10^{-4}$ | $1.23 \times 10^{-4}$ |
wavelength signals was manifested; thus, the large changes in the log ratio of the two works well for single signals. 

Table 4. Linear fitting parameters of the DAOD for simulated signal from 1500 to 3000 m.

|          | Pure         | Original     | De-Noised   |
|----------|--------------|--------------|-------------|
| $R^2$    | 0.997        | 0.027        | 0.835       |
| Slope    | $8.88 \times 10^{-5}$ | $1.11 \times 10^{-4}$ | $9.25 \times 10^{-5}$ |

Figure 10. Comparison of the DAOD from 300 to 1500 m. (Left) Pure and original signal. (Right) Pure and de-noised signal.

The above simulation results indicate that the method performed well in the range from 300 to 1500 m, which can be considered to have little noise influence. Further analysis should be implemented to assess the capability of the de-noising method. We performed the same steps specified above to analyze the range from 1500 to 3000 m, which is more readily affected by noise. Seven IMFs were obtained after EEMD, and we carried out signal reconstruction by removing the first four IMFs that did not meet the criteria (Table 4). Compared with the pure signal, the de-noised signal had good consistency (Figure 11). Therefore, we can deduce that the de-noising process destroys little useful information from the original signal. The highest SNR and the lowest MSE were obtained when the first four IMFs were removed (Figure 12). These results suggest that the de-noising method works well for single signals.

Table 4. Linear fitting parameters of the DAOD for simulated signal from 1500 to 3000 m.

|          | Pure         | Original     | De-Noised   |
|----------|--------------|--------------|-------------|
| $R^2$    | 0.997        | 0.027        | 0.835       |
| Slope    | $8.88 \times 10^{-5}$ | $1.11 \times 10^{-4}$ | $9.25 \times 10^{-5}$ |

Figure 11. Comparison between the de-noised and pure signals. (Left) On-line wavelength. (Right) Off-line wavelength. The de-noised signal is reconstructed by IMFs 5–7.

Because of the influence of noise, the signal intensity of the on-line and off-line wavelength signals was manifested; thus, the large changes in the log ratio of the two signals produced many negative values (Figure 13, left). These values impaired the validity and credibility of the linear fitting and led to a poor result. Nevertheless, a good result was obtained after implementing the de-noising method (Figure 13, right). The figure shows that no negative value was obtained, which fully illustrates that the errors of the two echo signals were greatly corrected. Because the DAOD represents the log ratio of two signals significantly smoothed after de-noising, some smooth fluctuations in the DAOD are acceptable. Moreover, the slope of the de-noised data was closer to the slope of the pure signal than that of the original signal (Table 4), which improved the inversion precision.
Figure 12. (Right axis) SNR and (left axis) MSE of the de-noised signals. (Left) On-line wavelength. (Right) Off-line wavelength.

Figure 13. Comparison of the DAOD from 1500 to 3000 m. (Left) Pure and original signal. (Right) Pure and de-noised signal.

In order to further prove the superiority of the method in this paper, we compared the proposed method with several de-noising methods (Figure 14). The multi-pulse average method contributed little to the result. Compared with the EEMD method, the EMD and WT methods did not sufficiently improve the performance. From the data listed in Table 5, the EEMD method is more representative of the other methods.

Figure 14. Comparison of the four de-noising methods. (Upper left) Multi-pulse average method. (Top right) Wavelet de-noising method. (Left bottom) EMD. (Lower right) Proposed method (EEMD) in this paper.
4. Observed Signal Analysis

The CO$_2$-DIAL echo signal for analysis was collected in relatively stable atmospheric conditions and without acute changes on 29 December 2015. Thus, we assumed that the useful information of the signals is the same as one another; any difference could be due to the noise with a 30 s duration. Three signal pairs sequentially collected within 30 s were selected as our research objects; each pair includes on-line and off-line wavelength photonic signals of three signal pairs ranging from 1100 to 3000 m for presentation and various temporal scales.

In our CO$_2$-DIAL system, two acquisition modes were used, namely, photon-counting and analog modes. The former has a better SNR, and the latter can compensate for the limitation in collecting the near-range echo signal of the former. Then, we chose the photonic signals of three signal pairs ranging from 1100 to 3000 m for presentation and named them as T1, T2, and T3. In particular, we only implemented signal de-noising on T2, and T1 and T3 served as reference signals. Figure 15 shows the IMFs of the off-wavelength echo signals of T1, T2, and T3. We calculated the correlation coefficients of the IMFs with the same temporal scale. From Table 6, we can find that the correlation coefficients of the first two IMFs are lower than 0.5. Those of the remaining IMFs that are not listed are higher than 0.5. Thus, we deleted the two IMFs and reconstructed the residue components.

![Figure 15. IMFs of the off-wavelength echo signals of T1, T2, and T3 (C1, C2, …, C7 represent various temporal scales).](image)

A good de-noising result is shown in the following (Figure 16). We see that the de-noising signal became smoother as well as retained the characteristics of the original signal. To validate the effect of the de-noising method for the investigated signal, we calculated the DAOD of the original and the de-noised signals. A significant improvement was obtained. The DAOD of the original signal showed negative values and had large discreteness (Figure 17). In contrast to the original signal, the de-noised signal had good linearity, and the r-square after linear fitting was 0.31 higher than that of the original signal (Table 7). Moreover, the interference of the negative values was eliminated, which undoubtedly enhances the data reliability. Meanwhile, the slope of the DAOD improved to some extent.

|       | Average | Wavelet | EMD   | EEMD   |
|-------|---------|---------|-------|--------|
| $R^2$ | 0.028   | 0.433   | 0.311 | 0.835  |
| Slope | $6.01 \times 10^{-5}$ | $1.04 \times 10^{-4}$ | $1.02 \times 10^{-4}$ | $9.25 \times 10^{-5}$ |
Table 6. Correlation coefficients of the IMFs with the same temporal scale. (Left) On-line wavelength. (Right). Off-line wavelength.

|     | C1 | C1 |
|-----|----|----|
| $\lambda_{on}$ | T1 | T2 | T3 | $\lambda_{off}$ | T1 | T2 | T3 |
| T1  | 1.0000 | 0.0634 | 0.1116 | T1  | 1.0000 | -0.0057 | 0.0464 |
| T2  | 0.2929 | 1.0000 | 0.4128 | T2  | 0.2398 | 1.0000 | 0.0155 |
| T3  | 0.5217 | 0.4128 | 1.0000 | T3  | 0.3347 | 0.0155 | 1.0000 |

|     | C2 | C2 |
|-----|----|----|
| $\lambda_{on}$ | T1 | T2 | T3 | $\lambda_{off}$ | T1 | T2 | T3 |
| T1  | 1.0000 | 0.7089 | 0.7245 | T1  | 1.0000 | 0.7922 | 0.7482 |
| T2  | 0.7089 | 1.0000 | 0.7821 | T2  | 0.7922 | 1.0000 | 0.8087 |
| T3  | 0.7245 | 0.7821 | 1.0000 | T3  | 0.7482 | 0.8087 | 1.0000 |

Figure 16. Comparison of the original and de-noised signals. (Left) On-line wavelength. (Right) Off-line wavelength.

Figure 17. Comparison of the DAOD of the original and de-noised signals from 1100 to 3000 m. (Left) Original signal. (Right) De-noised signal.

Table 7. Linear fitting parameters of the DAOD for observed signal.

|       | Original | De-Noised |
|-------|----------|-----------|
| $R^2$ | 0.444    | 0.755     |
| Slope | $1.83 \times 10^{-4}$ | $1.77 \times 10^{-4}$ |
5. Conclusions

For the CO$_2$-DIAL, CO$_2$ inversion is an important process. However, the existence of noise severely hinders the improvement of inversion precision. In this paper, we used the EEMD algorithm to de-noise the echo signals of the ground-based CO$_2$-DIAL system and set the threshold by calculating the correlation coefficients of the same scale components of multiple groups of adjacent signals decomposed by EEMD so as to determine the number of basic mode components to be removed. This method not only retains the useful information effectively but also removes the noise component. We decomposed three adjacent signal pairs into different IMFs using EEMD and removed the IMFs according to their correlation coefficients with the same temporal scale. Then, we reconstructed the residual components to complete the signal de-noising.

To verify the feasibility of the method, we analyzed the simulated CO$_2$-DIAL signal. Three similar simulated signal groups with different SNRs were selected as our research objects. They can be considered as observed signals collected within a short time in stable atmospheric conditions. We segmented the signals into two parts to verify the performance of the method under different signal intensities. First, the de-noised signal and the pure signal were compared in this paper. The good linear relationship between the de-noised signal and the pure signal demonstrates that the signals have a high degree of similarity. Meanwhile, the SNR and MSE were calculated to prove the effectiveness of the method proposed in this paper. Then, the DAOD of the de-noised signal is calculated. The R$^2$ of the DAOD reaches 0.841 and 0.835 for the detected range of 300–1500 m and 1500–3000 m, indicating that the quality is high, which is conducive to the high-precision inversion of CO$_2$. In order to further prove the superiority of the method in this paper, we also compared the proposed method with several de-noising methods. The results show that our method is more representative of the other methods in CO$_2$-DIAL de-noising. Finally, the method was used to observe CO$_2$-DIAL signal de-noising. The de-noised signal became smoother as well as retained the characteristics of the original signal. In contrast to the original signal, the de-noised signal had good linearity, and the R$^2$ of the DAOD was improved from 0.444 of the original signal to 0.735 of the de-noised signal, which is beneficial to the enhancement of the inversion precision. Certainly, this work may have some shortcomings. For example, the threshold for selecting the correlation coefficients was determined by many experiments but lacked theoretical foundations. We simply discussed the effect of the method under stable atmospheric conditions, and other cases were not considered. The above-mentioned deficiencies still require further research, and the proposed method requires continuous improvement in the future.

**Author Contributions:** The study was completed with cooperation between all authors. C.X. and Y.Z. designed the research topic; C.X., Y.Z. and R.L. conducted the experiment; A.L. checked and analyzed the experimental results; C.X. wrote the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded in part by the Natural Science Foundation of Jiangsu Province, China under Grants BK20180809 and BK20190779; in part by the National Natural Science Foundation of China under Grants 41901274 and 42001273; and in part by the Talent Launch Fund of Nanjing University of Information Science and Technology under Grant 2017r066.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data set available on request to corresponding authors.

**Conflicts of Interest:** The authors declare no conflict of interest.
27. Wu, Z.; Huang, N.E. A study of the characteristics of white noise using the empirical mode decomposition method. Proc. R. Soc. Lond. A Math. Phys. Eng. Sci. 2004, 460, 1597–1611. [CrossRef]

28. Wu, Z.; Huang, N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. Adv. Adapt. Data Anal. 2009, 1, 1–41. [CrossRef]

29. Shi, T.; Han, G.; Ma, X.; Gong, W.; Chen, W.; Liu, J.; Zhang, X.; Pei, Z.; Gou, H.; Bu, L. Quantifying CO₂ uptakes over oceans using LIDAR: A tentative experiment in Bohai bay. Geophys. Res. Lett. 2021, 48, e2020GL091160. [CrossRef]

30. Shi, T.; Han, G.; Ma, X.; Zhang, M.; Pei, Z.; Xu, H.; Qiu, R.; Zhang, H.; Gong, W. An inversion method for estimating strong point carbon dioxide emissions using a differential absorption Lidar. J. Clean. Prod. 2020, 271, 122434. [CrossRef]