

## Article

# Coal Thickness Prediction Method Based on VMD and LSTM

Yaping Huang<sup>1,\*</sup> , Lei Yan<sup>1</sup>, Yan Cheng<sup>2</sup>, Xuemei Qi<sup>1,\*</sup> and Zhixiong Li<sup>3</sup>

<sup>1</sup> School of Resources and Geosciences, China University of Mining and Technology, Xuzhou 221116, China; ts20010069a31@cumt.edu.cn

<sup>2</sup> China National Administration of Coal Geology, Beijing 100038, China; cumt\_cheng@sohu.com

<sup>3</sup> Yonsei Frontier Lab, Yonsei University, Seoul 03722, Korea; zhixiong.li@yonsei.ac.kr

\* Correspondence: yphuang@cumt.edu.cn (Y.H.); qixuemei@cumt.edu.cn (X.Q.)

**Abstract:** The change in coal seam thickness has an important influence on coal mine safety and efficient mining. It is very important to predict coal thickness accurately. However, the accuracy of borehole interpolation and BP neural network is not high. Variational mode decomposition (VMD) has strong denoising ability, and the long short-term memory neural network (LSTM) is especially suitable for the prediction of complex sequences. This paper presents a method of coal thickness prediction using VMD and LSTM. Firstly, empirical mode decomposition (EMD) and VMD methods are used to denoise simple signals, and the denoising effect of the VMD method is verified. Then, the wedge-shaped coal thickness model is constructed, and the seismic forward modeling and analysis are carried out. The results show that the coal thickness prediction based on seismic attributes is feasible. On the basis of VMD denoising of the original 3D seismic data, VMD-LSTM is used to predict coal thickness and compared with the prediction results of the traditional BP neural network. The coal thickness prediction method proposed in this paper has high accuracy and basically coincides with the coal seam information exposed by existing boreholes. The minimum absolute error of the predicted coal thickness is only 0.08 m, and the maximum absolute error is 0.48 m. This indicates that VMD-LSTM has high accuracy in predicting coal thickness. The proposed coal thickness prediction method can be used as a new method for coal thickness prediction.

**Keywords:** VMD; LSTM; coal thickness; seismic attribute



**Citation:** Huang, Y.; Yan, L.; Cheng, Y.; Qi, X.; Li, Z. Coal Thickness

Prediction Method Based on VMD and LSTM. *Electronics* **2022**, *11*, 232.

<https://doi.org/10.3390/electronics11020232>

electronics11020232

Academic Editors:

Darius Andriukaitis, Yongjun Pan and Peter Brida

Received: 8 December 2021

Accepted: 10 January 2022

Published: 12 January 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/>).

## 1. Introduction

In the construction and production process of large-scale mines, all aspects of coal mine safety production need to accurately calculate the change in coal thickness [1]. When the coal thickness of the working face changes dramatically, the reserves per unit length of the working face will change greatly. If the same mining speed is adopted, the monthly coal output will be unbalanced, and the mine production will be greatly affected. According to statistics, when the actual coal thickness is 10–20% thinner than the designed coal thickness, the coal output will decrease by 35–40% [2]. In general, the thicker the coal is, the more gas is formed in the coalification process, and the greater the probability of gas leakage in the mining process is. The impact risk degree is closely related to the coal thickness and its changes. A large number of field observations and in situ stress measurements have found that in the coal thickness change area, the abnormal phenomenon of in situ stress field occurs, and the stress concentration degree is high, which is easy to induce strong mine earthquakes and rock bursts [3]. Therefore, coal thickness is an essential type of data in the process of coal mine design and mining [4]. Accurately predicting coal thickness can not only improve the economic benefit of coal mines, but also provide a strong geological guarantee for the safe mining of coal mines [5].

With the development of coal mining, in the seismic exploration of coal fields, it is necessary not only to find out the structure in the mining area, but also to provide the change of the coal seam thickness [6]. However, because most coal seams are typically

thin layers, their vertical resolution is very low, which is difficult to distinguish in seismic records and cannot meet the requirements of solving coal thickness, so the quantitative prediction of coal thickness has been one of the recognized problems in the geophysical field. The traditional method of coal thickness is to compare and interpolate the borehole data, and the accuracy of the coal thickness data provided is limited. In particular, when the coal seam is missed, stripped, and bifurcated, the prediction error of coal thickness is large [7].

In recent years, scholars have conducted a lot of research on coal thickness prediction and have achieved more significant results and progress. Some scholars use seismic attributes to predict coal thickness and have achieved good results. Zou et al. [8] used the staggered grid finite difference method to establish a wedge model to study coal thickness. The results show that when the coal seam is thin, the amplitude attribute is correlated with the coal thickness, and the law can be used to predict the coal thickness qualitatively. Guo et al. [9] analyzed the correlation between seismic attributes at the borehole and coal thickness. They obtained the regression equation between each attribute and coal seam thickness through optimal seismic attributes and achieved good application effects in coal thickness prediction. Lin et al. [10] used logging data combined with 3D seismic data and their interpretation results to conduct wave impedance inversion under multi-well constraints. The corresponding wave impedance layer of the coal seam was determined through horizon calibration, and then the thickness of the wave impedance layer was extracted and matched with the thickness of the coal seam exposed by the borehole. Finally, the distribution law of coal thickness in the 3D seismic survey area was obtained. Some scholars have introduced geostatistics into coal thickness prediction. Du et al. [11] used the Cokriging method to predict coal thickness by taking the thickness of drilled coal seam and seismic amplitude as regional variables, which reduced the error of coal thickness prediction and significantly improved the prediction accuracy. Cheng et al. [12] combined the variogram and the Kriging method to fully consider well data and seismic attribute data. He assigned certain weight coefficients to the two data types and performed a weighted average to calculate the thickness of the coal. The in-seam wave is a kind of guided wave propagating in the coal seam. It has the advantages of long propagation distance, carrying much information about the coal seam, roof, and floor; strong energy; and obvious dispersion characteristics. Based on these advantages, the in-seam wave can be used to study coal thickness. Liu et al. [13] found that the periodicity of the refraction P-wave in the in-seam wave is proportional to the thickness of coal, and the thickness of coal can be predicted by using the period of the refraction P-wave. Li et al. [14] chose to collect the appropriate frequency at the mining face in the transmission method of love waves with wave CT tomography, according to the wave velocity and thickness of the negative correlation and dispersion curve. Combined with the tunnel exposing coal thickness, the wave velocity contour map can be converted to a coal thickness contour map. Implementing the coal thickness quantitative prediction method proved high accuracy. With the development of artificial intelligence, more and more artificial intelligence technology is gradually applied to coal thickness prediction. The mature development of deep neural network technology urges mining workers to study the geological prediction method based on data and forms geological prediction technology driven by data [15,16]. Sun et al. [17] used non-linear BP neural network technology to predict coal thickness by extracting seismic attributes in the wavelet domain of different frequency bands according to the changing characteristics of coal thickness. Wu et al. [18] combined the least square support vector machine (LSSVM) with the Kriging method to predict coal thickness. They used the strong non-linear fitting and generalization capabilities of LSSVM to adaptively fit the experimental variogram, which overcame the disadvantages of the traditional variation function, such as difficulty in solving and strong subjectivity, and greatly improved the prediction accuracy. Guo et al. [19] proposed the coal thickness prediction method by combining genetic algorithm (GA) and simultaneous iterative reconstruction technology (SIRT), which can improve the accuracy of coal thickness inversion. However, the above methods are basically coal

thickness prediction methods directly based on seismic data, and discussion on the actual situation that seismic data usually contains noise is limited. Seismic data contains a lot of noise, which is a practical problem that needs to be faced when predicting coal seam thickness.

Variational mode decomposition (VMD) is a non-recursive modal decomposition method proposed by Dragomiretskiy in 2014 [20]. The VMD algorithm decomposes the signal according to the center frequency, divides the signal around a certain center frequency into one component, and repeats this process to achieve the decomposition of the original signal. When VMD is used to process non-stationary signals, it can effectively avoid the mode-aliasing effect and the endpoint effect caused by the empirical mode decomposition (EMD) algorithm, and the intrinsic mode function (IMF) is redefined as an amplitude modulation–frequency modulation signal. It has a solid mathematical basis. More importantly, the VMD method can effectively attenuate a large amount of random noise contained in seismic data. The long short-term memory neural network (LSTM) was proposed by Hochreiter et al. in 1997 and further improved in the following years [21]. LSTM is designed to learn long-term memory, and the model can overcome the inherent problems of the recurrent neural network (RNN), namely “vanishing gradients” and “exploding gradients”. LSTM is one of the most successful recurrent neural networks, and its prediction accuracy is high.

Based on the good denoising ability of VMD and good prediction ability for complex sequences of LSTM, this paper proposes VMD-LSTM method for coal thickness prediction. Firstly, the denoising effect of the VMD method is verified. Then, the coal thickness wedge model is used to illustrate that the seismic attribute can predict the coal thickness. The VMD-LSTM method is used for coal thickness prediction of actual seismic data. The prediction results have high accuracy and basically coincide with the coal seam information exposed by existing boreholes, which shows that the coal thickness prediction method proposed in this paper can be used as a new method for coal thickness prediction.

## 2. Basic Principle of VMD and LSTM for Predicting Coal Thickness

### 2.1. Basic Principles of VMD

During the late 1990s, Huang introduced the algorithm called EMD, which is widely used today to recursively decompose a signal into different modes of unknown but separate spectral bands [22]. However, due to the lack of mathematical theory and freedom, the robustness of the algorithm is very low, which leaves room for the development and improvement of EMD. VMD is an adaptive, non-recursive decomposition method that can decompose signals into the sum of finite component signals [23]. It is a new decomposition algorithm based on the Wiener filter, Hilbert transform, and heterodyne demodulation. It has a solid mathematical and theoretical foundation. VMD decomposes the signal according to the central frequency, divides the signal around a certain central frequency into a component, and repeats this process [24]. Finally, the original signal is decomposed. When VMD is used to process non-stationary signals, it can effectively avoid the mode-aliasing effect and the endpoint effect caused by the EMD algorithm, and IMF is redefined as an amplitude modulation–frequency modulation signal.

#### 2.1.1. Construction of Variational Problems

The goal of VMD is to decompose a real-valued input signal into a discrete number of IMF. The IMF expression is as follows:

$$u_k(t) = A_k(t) \cos(\varphi_k(t)) \quad (1)$$

In Equation (1),  $A_k(t)$  is the non-negative envelope, and  $\varphi_k(t)$  is the phase of the signal.

For each mode, compute the associated analytic signal by means of the Hilbert transform in order to obtain a unilateral frequency spectrum:

$$[\delta(t) + \frac{j}{\pi t}] * u_k(t) \tag{2}$$

In Equation (2),  $\delta(t)$  is a Dirac delta function,  $j$  is an imaginary unit, and  $*$  is the convolution symbol.

For each mode, shift the mode’s frequency spectrum to “baseband” by mixing with an exponential tuned to the respective estimated center frequency.

$$[(\delta(t) + \frac{j}{\pi t}) * u_k(t)]e^{-jw_k t} \tag{3}$$

In Equation (3),  $e^{-jw_k t}$  is the center frequency index.

According to the estimated bandwidth of IMF, the variational constraint equation is obtained:

$$\begin{aligned} \min_{\{u_k\}, \{w_k\}} \{ \sum_k \|\partial_t [(\delta(t) + \frac{j}{\pi t}) * u_k(t)]e^{-jw_k t}\|_2^2 \} \\ \text{s.t. } \sum_k u_k = f \end{aligned} \tag{4}$$

$\{u_k\} := \{u_1, u_2 \dots, u_K\}$  and  $\{w_k\} := \{w_1, w_2 \dots, w_K\}$  are shorthand notations for the set of all modes and their center frequencies, respectively.  $\sum_k := \sum_{k=1}^K$  is understood as the summation over all modes.

### 2.1.2. Solution of Variational Problem

In order to obtain the optimal solution, the augmented Lagrange expression is obtained by introducing the quadratic penalty factor  $\alpha$  and the Lagrange multiplication operator  $\gamma$ . The purpose of introducing  $\alpha$  is to ensure the signal reconstruction accuracy under the influence of noise, and the purpose of introducing  $\gamma$  is to maintain strict constraints in the solving process.

$$\begin{aligned} L(\{u_k\}, \{w_k\}, \gamma) = \alpha \left\{ \sum_{k=1}^K \|\partial_t [(\delta(t) + \frac{j}{\pi t}) * u_k(t)]e^{-jw_k t}\|_2^2 \right\} + \\ \sum (f(t) - \sum_{k=1}^K u_k(t))\|_2^2 - \left\langle \gamma(t), f(t) - \sum_{k=1}^K u_k(t) \right\rangle \end{aligned} \tag{5}$$

The alternating direction method of multipliers (ADMM) is used to solve the formula. In order to achieve the purpose of complete signal decomposition, the central frequency of each IMF is divided according to the frequency domain characteristics of the original signal, and then the central frequency of IMF is updated. The updated IMF expression is as follows:

$$u_k^{n+1} = \underset{u_k \in X}{\operatorname{argmin}} \left\{ \alpha \|\partial_t [(\delta(t) + \frac{j}{\pi t}) * u_k(t)]e^{-jw_k t}\|_2^2 + \|f(t) - \sum_i u_i(t) + \frac{\gamma(t)}{2}\|_2^2 \right\} \tag{6}$$

Fourier transform is used to transform the formula from the time domain to the frequency domain, and then the non-negative frequency domain interval is integrated to obtain the optimal solution. Among them, the frequency domain calculation criterion expression of each IMF is as follows:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\gamma}(\omega)}{2}}{1 + 2\alpha(\omega - w_k)^2} \tag{7}$$

The updated iteration formula of IMF center frequency is:

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|_2 d\omega} \quad (8)$$

In Equation (8),  $\hat{u}_k^{n+1}(\omega)$  is the Wiener filter of  $\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega)$ , and  $\omega_k^{n+1}$  is the frequency spectrum center of each IMF.

## 2.2. Basic Principles of LSTM

Hochreiter et al. proposed LSTM on the basis of RNN. RNN takes the sequence data as the input of the model, calculates the influence on the final output by learning the changes of a series of data, and modifies the weight matrix to optimize the network by calculating the gradient through positive feedback and negative feedback so as to realize the learning of time series. However, RNN systems are problematic in practice, despite their stability. During training, they suffer from well-documented issues, known as “vanishing gradients” and “exploding gradients” [25]. These difficulties become pronounced when the dependencies in the target subsequence span a large number of samples, requiring the window of the unrolled RNN model to be commensurately wide in order to capture these long-range dependencies [26]. LSTM is a good solution to the problem of “vanishing gradients” and “exploding gradients” and can be efficiently learned through memory cells and “gates” and for long-term memory of useful information. The structure of LSTM is similar to a traditional neural network, which has one input layer, one or more hidden layers, and one output layer. The hidden layer of the LSTM neural network contains many neurons called memory cells, and each of these memory cells has three “gates”, whose function is to maintain and adjust the state of the memory cells. These three “gates” are called the forget gate, input gate, and output gate, respectively. The essence of the control gate of LSTM is a fully connected layer. The input is a vector  $x$ , which is calculated with weight vector  $W$  and bias item  $b$  and filtered by the activation function. The output will obtain a real vector of 0–1, whose calculation formula is as follows:

$$g(x) = \sigma(wx + b) \quad (9)$$

The activation function of the gates in the hidden layer of LSTM is the sigmoid function with a range of (0, 1). The output of the gates is multiplied by the input vector to determine how much information is passed. When the output of the gate is 0, the product of the output vector and the input vector results in 0, and the information cannot pass. When the output is 1, the input vector is unaffected, and the gate is open. The gate is always ajar because of the sigmoid function range properties.

LSTM controls the unit state  $c$  with the forget gate and the input gate. The forget gate determines how much of the unit state  $c_{t-1}$  at the previous moment is retained to the current moment  $c_t$ , and the input gate determines how much of the network input  $x_t$  at the current moment is saved to the unit state  $c_t$ . LSTM uses the output gate to control how much output the unit state  $c_t$  has to the current output of LSTM  $h_t$ .

Updating a cell can be broken down into the following steps [24].

- (1) Calculate the value of the forget gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (10)$$

In Equation (10),  $W_f$  is the weight matrix of the forget gate,  $[h_{t-1}, x_t]$  is the joining of two vectors into a longer vector,  $b_f$  is the offset of the forget gate, and  $\sigma$  is the sigmoid function.

- (2) Calculate the value of the input gate:

In Equation (11),  $W_i$  is the weight matrix of the forget gate, and  $b_i$  is the offset of the forget gate.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (11)$$

- (3) Calculate the value of  $\tilde{c}_t$ . It is used to describe the unit state of the current input, calculated from the previous output and the current input:

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (12)$$

- (4) Calculates the cell state at the current time  $c_t$ :

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (13)$$

In Equation (13),  $\odot$  is the symbol for multiplying matrices by bits, and  $c_t$  is the result of combining current and long-term memories.

- (5) Calculate the value of the output gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (14)$$

In Equation (14),  $W_o$  is the weight matrix of the output gate, and  $b_o$  is the offset of the forget gate.

- (6) Calculate the current cell output of LSTM:

$$h_t = o_t \odot \tanh(c_t) \quad (15)$$

### 2.3. Coal Thickness Prediction Process of VMD-LSTM

Firstly, the wedge model of coal thickness is constructed and simulated by seismic forward modeling, and the sensitive seismic attributes of coal thickness such as instantaneous amplitude, instantaneous frequency, and relative wave impedance are extracted, then the feasibility of coal thickness prediction using seismic attributes is analyzed. Secondly, the seismic data were decomposed into IMFs with different characteristics and frequencies by using the VMD method, the correlation coefficients between seismic attributes of seismic data IMF1 and coal thickness were calculated, and the appropriate IMFs were selected in order to effectively remove random noise and highlight effective signals. Finally, the LSTM network is used to predict coal thickness. The specific flow chart is shown in Figure 1.

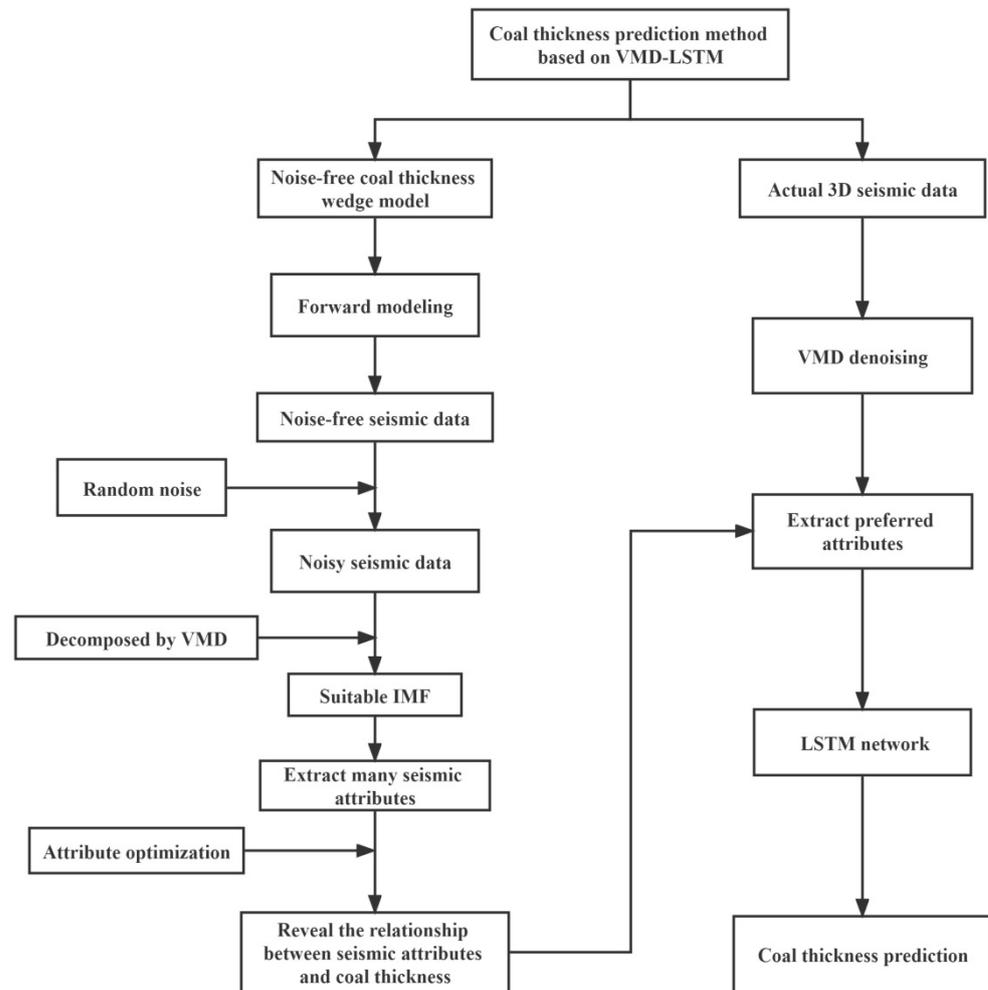


Figure 1. Flow chart of coal thickness prediction by VMD and LSTM.

### 3. Numerical Calculation

#### 3.1. Simple Signal Test

Formula (16) is adopted to synthesize the signal, and Gaussian white noise with a signal-to-noise ratio (SNR) of 30 dB is added to the signal to form a signal containing noise (Figure 2). The EMD and VMD methods are used to denoise simple signals to test their denoising ability. The denoising results are shown in Figures 3 and 4. It can be seen from Figure 3 that there is an obvious mode-aliasing problem in EMD decomposition, which cannot effectively decompose the random noise information in simple signals. Figure 4 shows that VMD has a better decomposition effect on random noise, IMF1 corresponds to the main components of the signal, and the correlation coefficient with the noise-free signal reaches 0.9998. The result shows that VMD can be used in signal denoising, and the denoising effect is good.

$$\begin{cases} y_1 = \cos(30\pi * t) \\ y_2 = 0.5 * \sin(30\pi * t) \\ y = y_1 + y_2 \end{cases} \quad (16)$$

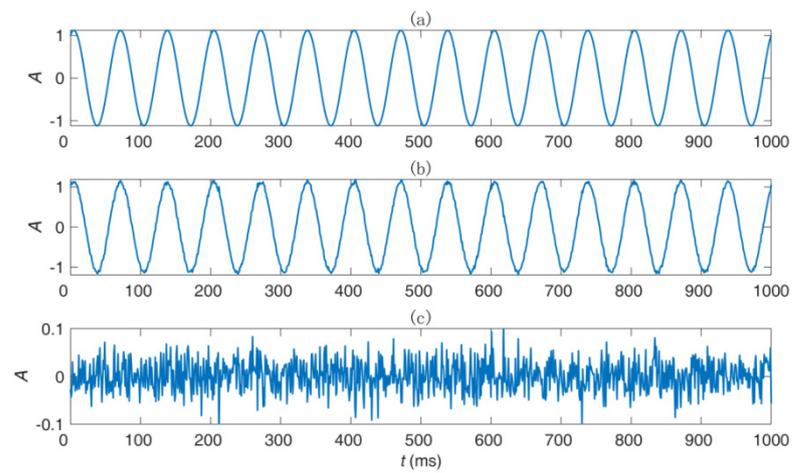


Figure 2. Simple signal and noisy signal: (a) simple signal; (b) noisy signal; (c) noise.

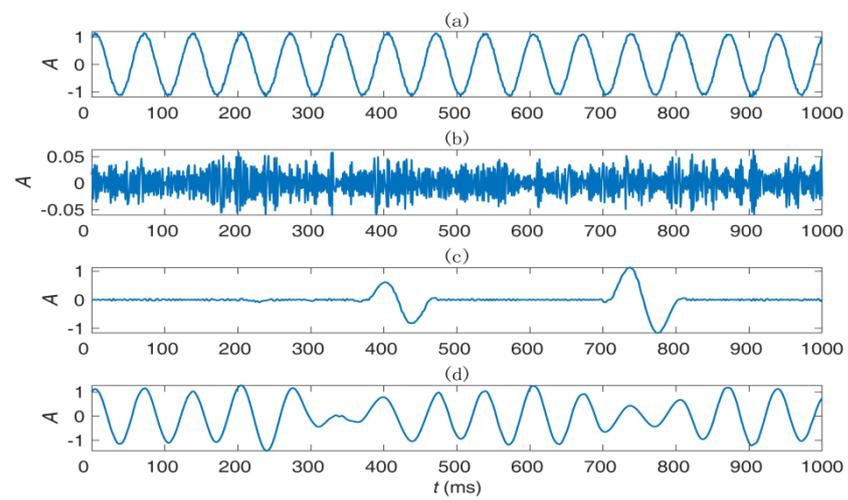


Figure 3. EMD decomposition results of noisy signal: (a) noisy signal; (b) IMF 1; (c) IMF 2; (d) IMF 3.

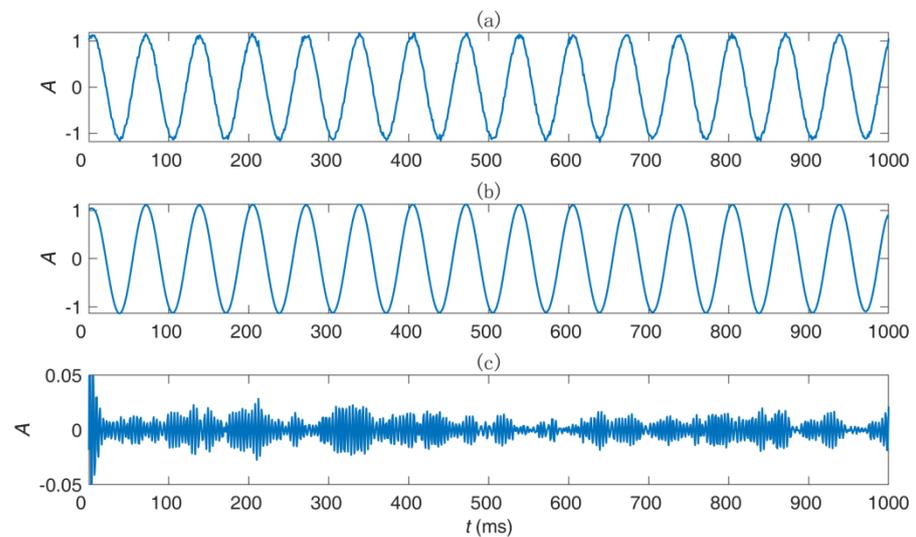
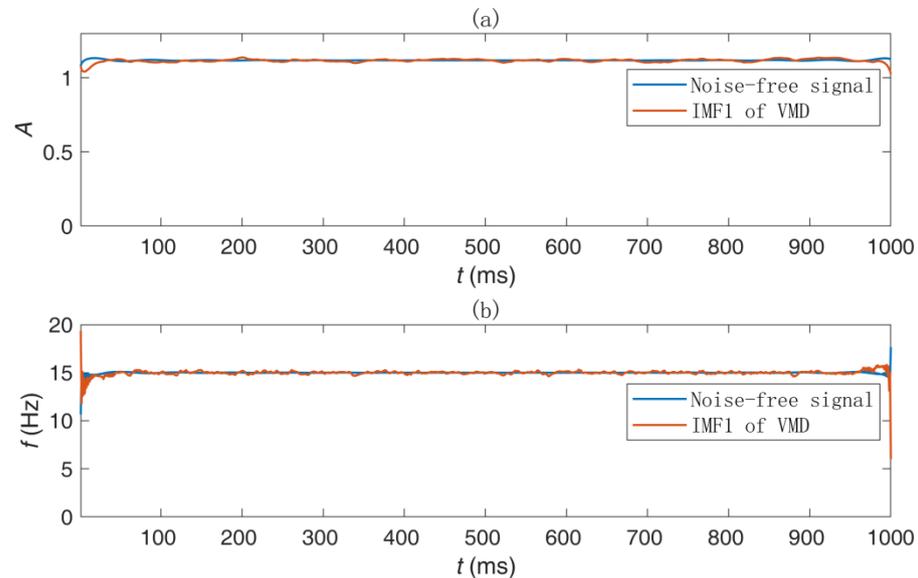


Figure 4. VMD decomposition results of noisy signal: (a) noisy signal; (b) IMF 1; (c) IMF 2.

The instantaneous amplitude and frequency of IMF1 denoised by VMD are calculated (Figure 5). As shown in Figure 5, the correlation coefficients of the instantaneous amplitude and instantaneous frequency between the noise-free signal and IMF1 are high, reaching 0.9861 and 0.9709.

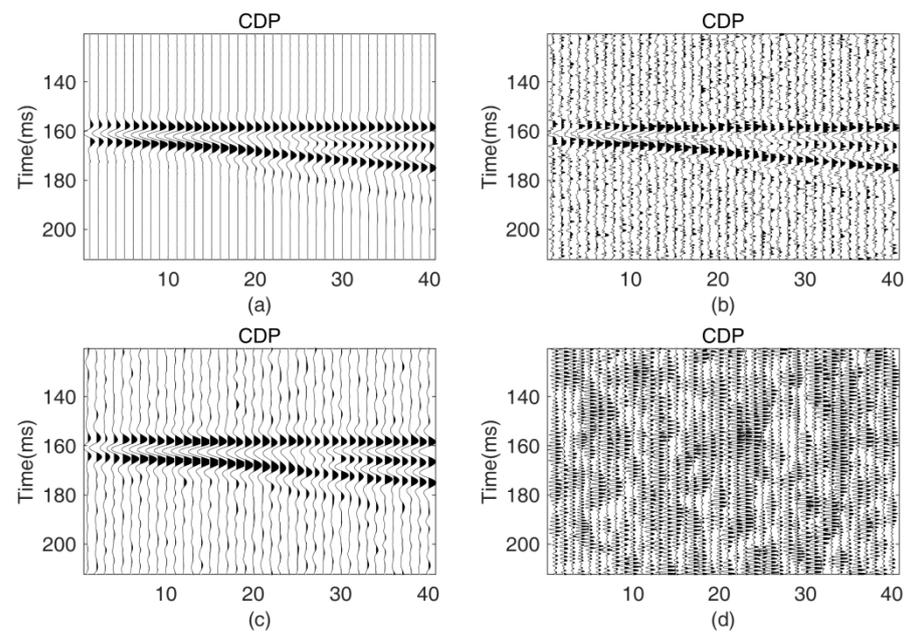


**Figure 5.** VMD decomposition results of noisy signal: (a) instantaneous amplitude of noise-free signal and IMF1 of VMD; (b) instantaneous frequency of noise-free signal and IMF1 of VMD.

The application results of VMD in the simple signal show that VMD has good denoising ability, the noisy signal processed by VMD can accurately contain the effective components of the noise-free signal, and the instantaneous amplitude and instantaneous frequency extracted from IMF1 have a high correlation with the instantaneous frequency and instantaneous amplitude of the noise-free signal.

### 3.2. VMD Decomposition of Coal Thickness Wedge Model

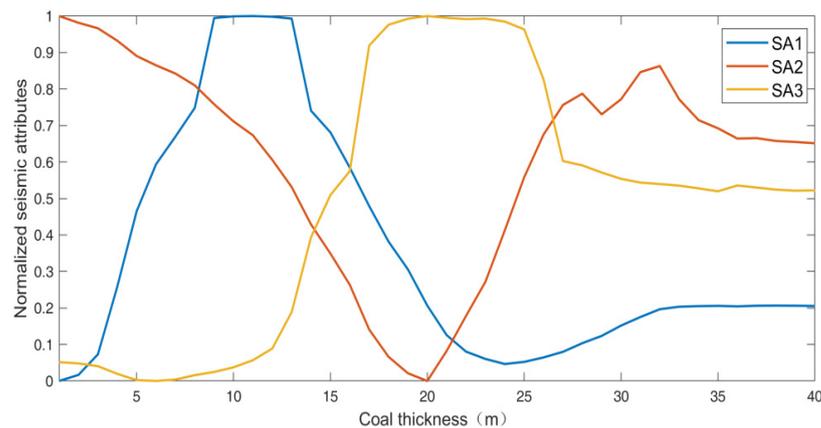
Firstly, a wedge model, which is commonly used in the study of coal thickness, is constructed. The model mainly consists of three layers: sandstone, mudstone, and coal seam. The coal thickness varies from 0 m to 40 m, and the velocity, density, and thickness of each layer are obtained from [5]. A Ricker wavelet with a main frequency of 50 Hz was used for forward modeling, and the distance between the receiving channels was 10 m, with a total of 40 receiving channels. The forward modeling section is shown in Figure 6a. Random noise of 40 dB was added to the forward record of the wedge model, as shown in Figure 6b. As can be seen from Figure 6, seismic records are significantly affected by noise, and the continuity of the in-phase axis becomes poor, which is not conducive to the subsequent prediction of coal thickness. VMD was carried out for noisy signals. The experiment found that when the decomposed IMF number  $K$  was greater than 2, the center frequency was very close, so  $K$  was set to 2. The results of the seismic records of the wedge model of the coal seam decomposed by VMD are shown in Figure 6c,d. The correlation coefficients between the noise-free signal and IMF1 seismic tracks are calculated, with the maximum value reaching 0.9659 and the average value 0.9258. It shows that VMD has strong denoising ability, and IMF1 decomposed by VMD is basically consistent with the noise-free synthetic seismic record.



**Figure 6.** Wedge model seismic record and VMD decomposition result: (a) forward seismic record of the wedge-shaped model; (b) seismic record of the model after adding random noise; (c) IMF1 after VMD denoising of noisy signals; (d) IMF2 after VMD denoising of noisy signal.

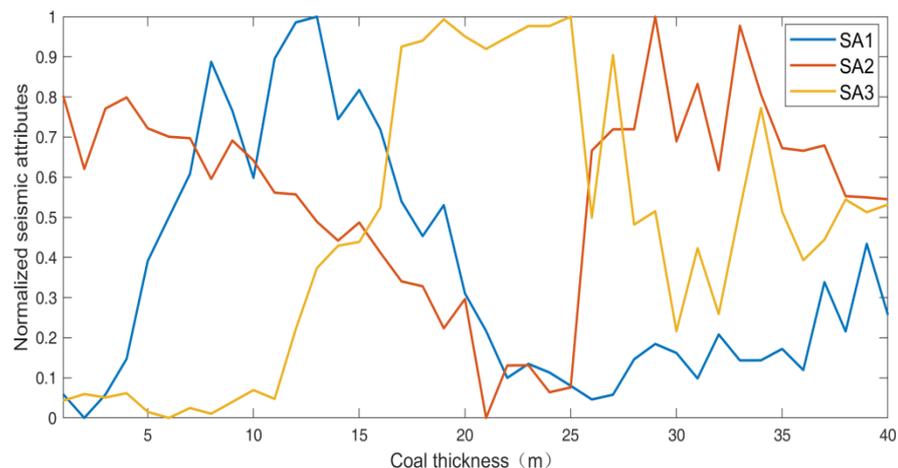
### 3.3. Seismic Attributes Analysis of Wedge Model Seismic Records

According to the existing research, seismic attributes such as instantaneous amplitude, instantaneous frequency, and relative wave impedance have a certain correlation with the thickness of coal. Therefore, the three seismic attributes mentioned above are extracted, and their relationship with the thickness of coal is analyzed. Figure 7 shows the relationship between the instantaneous amplitude, instantaneous frequency, and relative wave impedance of seismic records in the noise-free wedge model and coal thickness. When the coal seam thickness is less than 10 m, the instantaneous amplitude attribute shows an obvious increasing trend with the increase in the coal seam thickness; in other words, the instantaneous amplitude attribute has a good positive correlation with the coal thickness. When the coal thickness is greater than 10 m, the instantaneous amplitude attribute gradually decreases and tends to change gently. When the thickness of coal is less than 20 m, the instantaneous frequency attribute gradually decreases and tends to the minimum value with the increase of coal thickness. The instantaneous frequency attribute has a good negative correlation with the coal thickness. With the continuous increase in coal thickness, the instantaneous frequency attribute gradually increases and tends to be stable. When the thickness of coal is less than 20 m, the relative wave impedance attribute increases gradually with the increase in the coal thickness and has a good positive correlation with the coal thickness. When the thickness of coal is greater than 20 m, the relative wave impedance attribute decreases and tends to be stable with the change in coal thickness. In conclusion, when the thickness of coal is small, the relationship between the attributes of the instantaneous amplitude, instantaneous frequency and relative wave impedance, and coal thickness is simple, and the coal thickness can be predicted by using seismic attributes.



**Figure 7.** Relationship between seismic attributes and coal thickness of noise-free wedge model.

VMD was used to decompose the seismic record of the wedge model with noise, extract instantaneous amplitude, instantaneous frequency, and relative wave impedance of IMF and draw the relationship between the three attributes and coal seam thickness (Figure 8). It can be seen from Figure 8 that the variation trend of seismic attributes extracted after denoising from the wedge model with noise is basically consistent with that extracted from the model without noise.



**Figure 8.** Relationship between the instantaneous amplitude, instantaneous frequency, and relative wave impedance of IMF and coal thickness.

#### 4. Application of VMD-LSTM Method in Coal Thickness Prediction

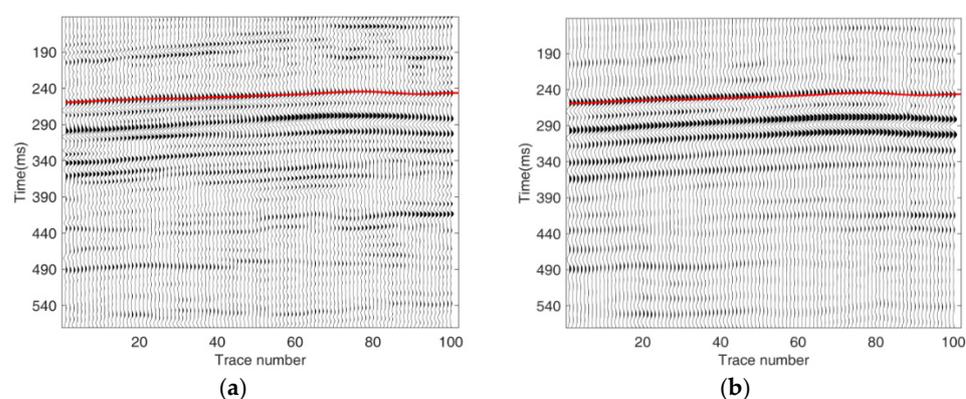
##### 4.1. Geological Survey of the Working Area

The working area is located in Ordos Basin. The coal-bearing strata in the mine field are the Carboniferous Taiyuan Formation and the Lower Permian Shanxi Formation. The total thickness of the coal-bearing strata revealed by drilling is 118.26–192.88 m, with an average of 142.74 m. The #4 coal seam is located in the upper part of the second rock member of the Lower Permian Shanxi Formation. The natural thickness of the coal seam is 0–5.65 m, with an average thickness of 3.82 m. The thickness of the working area is 0.86–3.79 m, 2.44 m on average. The structure of the coal seam is complex. There are 0–9 layers of partings, generally three layers. Most partings are located in the middle of the coal seam. The coefficient of variation of coal thickness is 32%. The coal seam gradually thickens from west to east. The #4 coal seam is a relatively stable coal seam with reliable comparison and most of the mining area. The thickness of the #4 coal seam in the working area varies from 2.30 to 4.70 m.

#### 4.2. Coal Thickness Prediction and Result Analysis

Seismic data processing mainly adopts conventional processes such as preprocessing, filtering, deconvolution, velocity analysis, normal moveout correction and static correction, stack, and migration. The focus of this paper is to use the VMD method to attenuate the noise of post-stack migration 3D seismic data at any time so as to provide seismic data with a high SNR for coal thickness prediction. The VMD-LSTM method was used to predict the thickness of the #4 coal seam in this area. The parameters of LSMT are as follows: the solver is 'adam', the gradient threshold is 1, the maximum number of rounds is 500, and the execution environment is specified as 'cpu'.

Figure 9 shows the original seismic record and the seismic record after VMD denoising of the #4 coal seam. It can be seen that the #4 coal seam is between 240 ms and 290 ms of the seismic record. It can be seen from the comparison between Figure 9a,b that the original seismic record is greatly affected by noise, and the seismic event of the #4 coal seam is not clear. After VMD denoising, the seismic event of the #4 coal seam is clearer, and the energy is significantly enhanced, which is conducive to the subsequent coal thickness prediction.



**Figure 9.** Original seismic record and seismic record after VMD denoising of #4 coal seam: (a) original seismic record of #4 coal seam; (b) seismic record after VMD denoising of #4 coal seam.

Instantaneous amplitude, instantaneous frequency, and relative wave impedance attributes of the seismic record after VMD denoising were extracted (Figures 10–12). The relationship between the coal thickness and instantaneous amplitude attribute derived from the wedge model shows that the coal seam in the western part of the area may be thinner, while the coal seam in the eastern part may be thicker. The instantaneous frequency and relative wave impedance attributes also have similar characteristics, but the response characteristics of each seismic attribute are quite different. The results of a single seismic attribute reflect the possible distribution characteristics of the coal seam, but it is impossible to quantitatively predict the coal thickness.

In order to study the variation of coal thickness in this area more accurately, based on the coal thickness information and seismic attributes of 10 known boreholes in this area, the instantaneous amplitude, instantaneous frequency, and relative wave impedance attributes of the borehole side channel were used as the input of the training samples, and the known coal thickness was used as the output of the training samples. The coal thickness predicted by the traditional BP neural network method is shown in Figure 13. The LSTM network was also used for training, which has six layers: an input layer, an output layer, and four hidden layers. In order to prevent over-fitting, the dropout layer was introduced into the model, and the dropout value was set to 0.1. The root mean square error is used for the loss function, and the optimizer is Adam. After obtaining satisfactory training results, the coal thickness of the whole area was predicted, as shown in Figure 14. Comparing Figures 13 and 14, it can be seen that the LSTM prediction results have better regularity. Figure 14 shows that the area of the southeast coal seam is thick, between 3 and 5 m, and the southwest coal seam is thin, about 3 m. The predicted coal thickness information is basically

consistent with the trend that the coal seam in this area gradually thickens from west to east and from south to north. This regularity is not obvious in Figure 13. The A10 borehole reveals a coal thickness of 4.7 m. Figure 14 shows that the coal seam in the area where A10 is located is thicker and the surrounding area is stable, but there is a certain deviation in the prediction results in Figure 13. The A9 borehole reveals a coal thickness of 2.3 m. Figure 14 also shows that the coal seam in the area where the borehole is located is thinner, and the coal seam is thinner in a certain range around. The analysis of the prediction results of coal thickness at 10 boreholes shows that the minimum absolute error of the BP neural network coal thickness prediction is 0.35 m, and the maximum absolute error is 1.27 m. However, the minimum absolute error of the LSMT coal thickness prediction is only 0.08 m, and the maximum absolute error is 0.48 m. The application results of actual data show that the coal thickness prediction method proposed in this paper has high accuracy, and the predicted coal thickness results have important guiding significance and reference value for later coal mining.

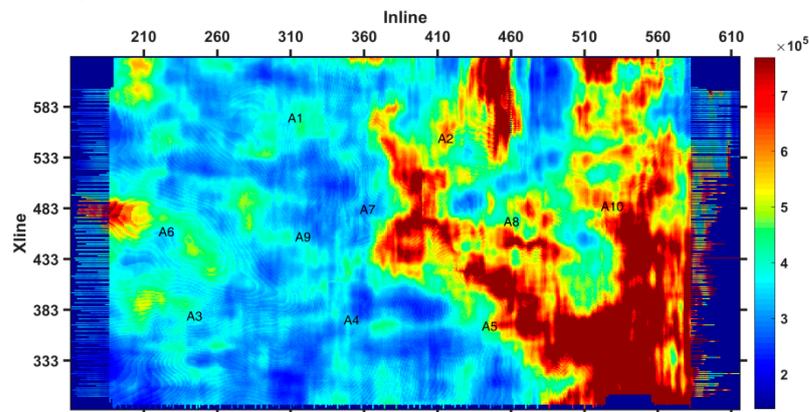


Figure 10. Instantaneous amplitude attribute of #4 coal seam.

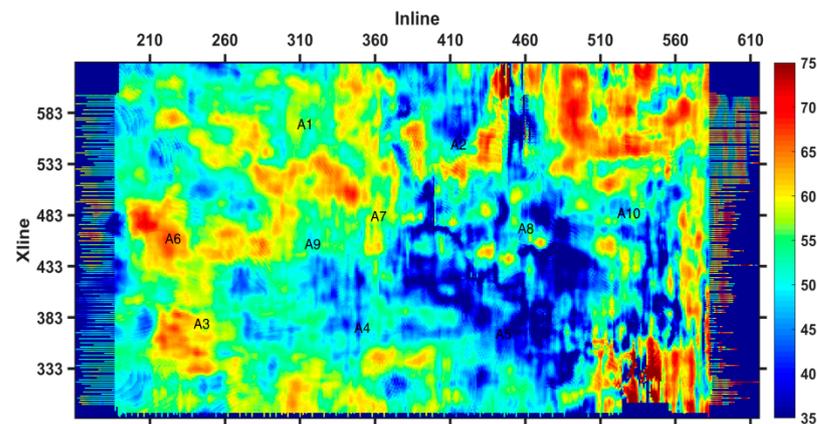


Figure 11. Instantaneous frequency attribute of #4 coal seam.

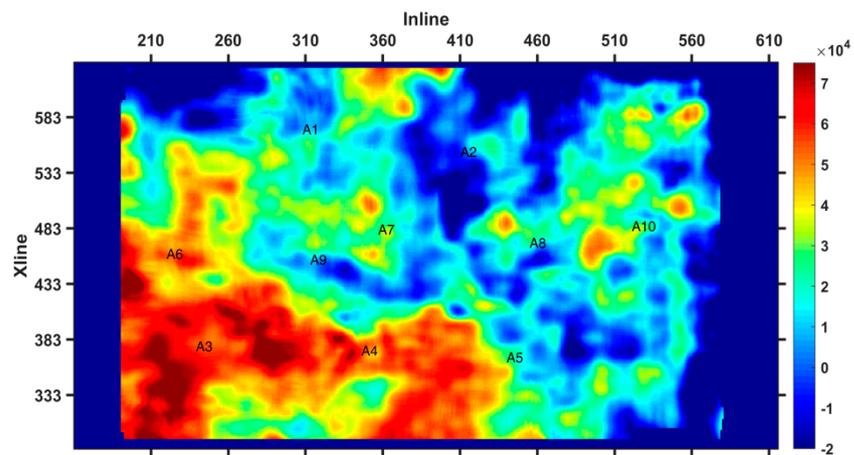


Figure 12. Relative wave impedance amplitude attribute of #4 coal seam.

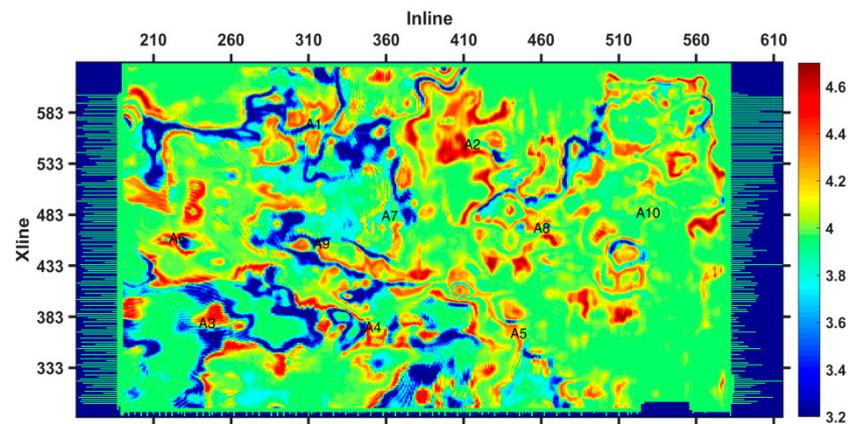


Figure 13. Coal thickness prediction results based on BP neural network.

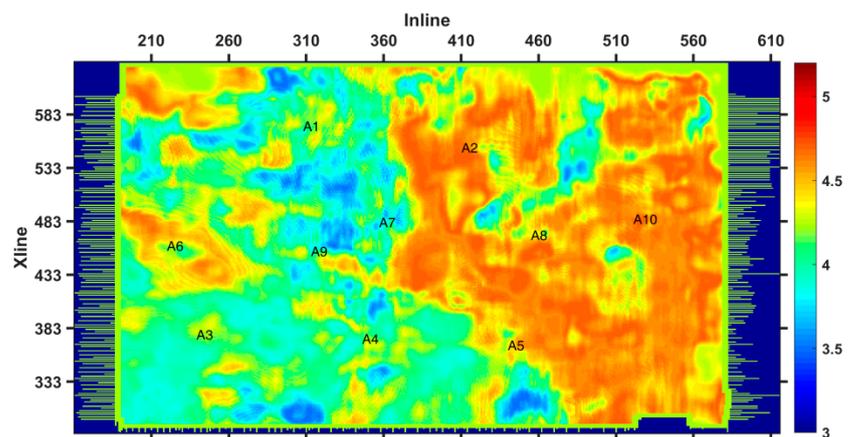


Figure 14. Coal thickness prediction results based on VMD-LSMT method.

### 5. Conclusions

- (1) EMD and VMD were used to denoise simple signals. There is an obvious mode-aliasing problem in EMD decomposition, which cannot effectively decompose the random noise. VMD can be used in signal denoising, and the denoising effect is good.
- (2) It can be seen from the forward simulation of the coal thickness wedge model that there is a good positive correlation between the instantaneous amplitude attribute, the relative wave impedance attribute, and the coal thickness, while the instantaneous

- frequency attribute has a good negative correlation with the coal thickness, which indicates that the seismic attribute is feasible for coal thickness prediction.
- (3) It can be seen from the comparison with traditional BP neural network coal thickness prediction results that the VMD-LSTM method has higher prediction accuracy. The prediction results are in good agreement with the coal seam information exposed by the existing boreholes, which can be used as a new method for coal thickness prediction.
  - (4) The influence of different seismic attributes on coal thickness prediction needs to be further explored. In the process of using LSTM to predict, different weights can be assigned to each seismic attribute. This will help improve the accuracy of coal thickness prediction. Further in-depth research is needed in the future.

**Author Contributions:** Y.H., L.Y., and X.Q. conceived and designed the algorithms; Y.H., X.Q., Z.L., and L.Y. performed the algorithms; Y.H., X.Q., and Y.C. analyzed the data; Y.H. and L.Y. wrote the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Fundamental Research Funds for the Central Universities (Grant Number 2018QNA45) and a project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

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

## References

1. Sun, Z.W. Numerical simulation of geostress field distribution in local variation area of coal seam thickness. *Ground Press. Strata Control* **2003**, *5*, 95–97.
2. Zhong, Q.T. Research and Application of Coal Seam Thickness Inversion Method. Ph.D. Thesis, China University of Mining and Technology, Xuzhou, China, 2001.
3. Shang, X.Q.; Yang, H.Y.; Ai, G. Study on the evolution and impact of mining-induced stress in coal thickness variation area. *China Min. Mag.* **2020**, *29*, 148–151+157.
4. Cui, W.X.; Wang, B.L.; Wang, Y.H. High-precision inversion method of coal seam thickness based on transmission channel wave. *J. China Coal Soc.* **2020**, *45*, 2482–2490.
5. Zeng, A.P.; Zhang, J.W.; Ren, E.M.; Liu, T.; Jiang, F.; Liu, X.J.; Jiang, F. Research on coal thickness prediction method based on VMD and SVM. *Coal Geol. Explor.* **2021**, *49*, 243–250.
6. Li, Q.H.; Zhang, C.Z.; Li, K.X. Analysis of influencing factors of mining surface subsidence under huge thick unconsolidated strata. *Coal Sci. Technol.* **2020**, *49*, 191–199.
7. Dong, S.H.; Xu, Y.Z. Inversion of coal seam thickness from seismic data by spectral moment method. *J. Liaoning Tech. Univ.* **2005**, *24*, 38–40.
8. Zou, G.G.; Peng, S.P.; Hao, X.X.; Zhang, J.; Wang, L. The relationship between coal thickness and seismic amplitude is analyzed based on wedge model. *Coal Sci. Technol.* **2014**, *42*, 88–91+100.
9. Guo, Y.X.; Meng, Z.P.; Yang, R.Z.; Zhang, L.H.; Liu, Y.C.; Sun, X.Y.; Zhao, G.P. Seismic attribute and its application to coal seam thickness prediction. *J. China Univ. Min. Technol.* **2004**, *33*, 67–72.
10. Lin, J.D.; Huo, Q.M.; Wu, Y.F. Application of multi-well constrained 3D seismic inversion technique in coal thickness prediction. *Coal Geol. China* **2003**, *15*, 47–49.
11. Yu, S.J.; Wang, Z.S.; Liu, Y.X. Application of wavelet multiscale analysis in coal thickness detection. *Coal Geol. Explor.* **2005**, *33*, 73–75.
12. Du, W.F.; Peng, S.P. Use geostatistics to predict coal seam thickness. *Chin. J. Rock Mech. Eng.* **2010**, *29*, 2762–2767.
13. Cheng, Y. Geological statistics coal thickness prediction method and application analysis. *China Min. Mag.* **2019**, *28*, 245–249.
14. Liu, Z.L.; Wang, J. Periods of refracted P-waves in coal seams and their applications in coal thickness estimations. *Acta Geophys.* **2020**, *68*, 1753–1762. [[CrossRef](#)]
15. Brosig, A.; Knobloch, A.; Barth, A.; Legler, C.; Hielscher, P.; Noack, S.; Etzold, S.; Dickmayer, E.; Kaufmann, H.; Franke, D. Mineral predictive mapping in 2D, 2.5D and 3D using Artificial Neural Networks—Case study of Sn and W deposits in the Erzgebirge, Germany. *Miner. Prospect. Curr. Approaches Future Innov. Orléans* **2017**, *1*, 24–26.

16. Noack, S.; Barth, A.; Irkhin, A.; Bennewitz, E.; Schmidt, F. Spatial modeling of natural phenomena and events with Artificial Neural Networks and GIS. *Int. J. Appl. Geospat. Res.* **2012**, *3*, 1–20. [[CrossRef](#)]
17. Sun, Y.; Zhang, L.; Zhu, J.; Xue, T.; Ding, J. Application of seismic attribute parameters in coal seam thickness prediction. *Coal Geol. Explor.* **2008**, *36*, 58–60.
18. Wu, W.W.; Yang, Y.G.; Chen, Y.K. Prediction of coal thickness Change by Kriging Method based on LSSVM Optimization. *Coal Technol.* **2015**, *34*, 89–91.
19. Guo, C.F.; Zhen, Y.; Shuai, C.; Ren, T.; Yao, W.L. Precise Identification of Coal Thickness by Channel Wave Based on a Hybrid Algorithm. *Appl. Sci.* **2019**, *9*, 1493. [[CrossRef](#)]
20. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process* **2014**, *62*, 531–544. [[CrossRef](#)]
21. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
22. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc.* **1998**, *454*, 903–995. [[CrossRef](#)]
23. Lian, J.; Liu, Z.; Wang, H.; Dong, X. Adaptive variational mode decomposition method for signal processing based on mode characteristic. *Mech. Syst. Signal Process.* **2018**, *107*, 53–77. [[CrossRef](#)]
24. Hochreiter, S.; Yoshua, B.; Paolo, F.; Schmidhuber, J. Gradient flow in recurrent nets: The difficulty of learning long-term dependencies. In *A Field Guide To Dynamical Recurrent Neural Networks*; IEEE Press: Piscataway Township, NJ, USA, 2001.
25. Sherstinsky, A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *Phys. D Nonlinear Phenom.* **2020**, *404*, 1–45. [[CrossRef](#)]
26. Cao, C.F. Research and Application of Coal Mine Floor Water Inrush Warning Model Based on LSTM. Master's Thesis, Xi'an University of Architecture and Technology, Xi'an, China, 2017.