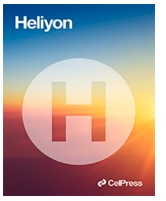
Heliyon 10 (2024) e28112

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Gas concentration prediction by LSTM network combined with wavelet thresholding denoising and phase space reconstruction

Kun Gao [a], ZuoJin Zhou [b,*], YaHui Qin [c]

[a] College of Mining Engineering, Liaoning Technical University, Fuxin, Liaoning, 123000, China
[b] College of Safety Science & Engineering, Liaoning Technical University, Fuxin, Liaoning, 123000, China
[c] Coal Science and Technology Research Institute, China Coal Science and Industry Group, Beijing, 100000, China

ARTICLE INFO

Keywords:
Time series
Wavelet threshold denoising
Phase space reconstruction
LSTM neural network
Gas concentration prediction

ABSTRACT

The Long Short-Term Memory neural network is a specialized architecture designed for handling time series data, extensively applied in the field of predicting gas concentrations. In the harsh conditions prevalent in coal mines, the time series data of gas concentrations collected by sensors are susceptible to noise interference. Directly inputting such noisy data into a neural network for training would significantly reduce predictive accuracy and lead to deviations from the actual values. The Empirical Mode Decomposition method, commonly employed in gas concentration prediction, faces challenges in practical engineering applications due to the substantial influence of newly acquired data on the initial decomposition subsequence values. Consequently, it is difficult to use this method as intended. Conversely, the Wavelet Threshold Denoising method does not encounter this issue. Furthermore, gas concentration sequences exhibit chaotic characteristics. Performing phase space reconstruction allows for the extraction of additional valuable hidden information. In light of these factors, a prediction model is proposed, integrating WTD, Phase Space Reconstruction, and LSTM neural networks. Initially, the gas concentration sequence itself is subjected to wavelet threshold denoising. Subsequently, phase space reconstruction is performed, and the resulting reconstructed phase space matrix serves as the input for the LSTM neural network. The outcomes from the final LSTM neural network reveal that the PS method indeed extracts more valuable information. The Mean Absolute Error and Root Mean Square Error are reduced by 35.1% and 25%, respectively. Additionally, when compared to the PS-LSTM model without utilizing the WTD method, the WTD-PS-LSTM predictive model showcases reductions of 77.1% and 80% in MAE and RMSE, respectively. Compared with the LSTM model, the MAE and RMSE of the WTD-PS-LSTM prediction model were reduced by 81.4% and 82.6%, respectively. This greatly improves the credibility of whether or not a response related to coal mine safety management is implemented.

1. Introduction

China not only leads the world in coal consumption but also relies on coal for over half of its primary energy consumption. The rapid development of China's economy is underpinned by a predominantly coal-based energy structure. While the trend is moving towards green energy, coal remains the primary source of energy consumption in China [1,2]. Simultaneously, China also stands out as

* Corresponding author.
E-mail address: 2261251242@qq.com (Z. Zhou).

https://doi.org/10.1016/j.heliyon.2024.e28112
Received 7 September 2023; Received in revised form 26 February 2024; Accepted 12 March 2024
Available online 16 March 2024

one of the countries with the most severe coal mine accidents globally, with gas incidents being one of the two major accident types [3]. The hazards of gas incidents manifest in various ways, including asphyxiation due to excessive gas concentration, gas explosions, gas combustion, and coal and gas outbursts. All these dangers are directly or indirectly linked to gas concentration. Accurately predicting gas concentration data for a future period can serve as a crucial tool for coal mine safety warning systems. It enables timely alerts and provides safety control personnel with the lead time needed to make informed decisions, thereby reducing the financial losses resulting from unexpected production stoppages.

Currently, there are three well-established and relatively novel approaches for predicting gas concentration:1. The first method involves treating gas concentration time series as chaotic sequences. The gas sequence is reconstructed in phase space, and predictive models are derived from fitting functions to various phase points.2. The second method utilizes grey system control theory to establish relevant models. This approach dynamically analyzes the changing patterns of coal mine gas concentration, predicting the evolving states of gas concentration.3. The third method involves summarizing the various factors influencing gas concentration variations. It selects key feature vectors as inputs, constructs neural network models, trains the network's parameter structure, and obtains predictive models for gas concentration. This method enables the prediction of the evolving trends in gas concentration. These approaches represent mature and innovative ways to forecast gas concentrations.

In the realm of gas concentration prediction, research combining phase space reconstruction and neural networks remains relatively scarce. Phase Space Reconstruction (PS) was employed in conjunction with particle swarm neural networks for precise gas concentration prediction [4]. Phase space reconstruction was used in combination with Long Short-Term Memory (LSTM) neural networks to achieve favorable predictive outcomes [5]. However, such integration is less explored in this field compared to other domains. For instance, in other domains, Wang et al. [6,7] established flight risk and photovoltaic power prediction models by reconstructing time series into phase space and then feeding them into Radial Basis Function (RBF) neural networks. Wang et al. [8] applied phase space reconstruction to predicted data and relevant influencing factors, followed by inputting them into an LSTM neural network to create a pollutant concentration prediction model. Nevertheless, the mentioned methods did not account for the interference of noise with neural network input information. Although He et al. [9,10] did address noise-related issues, their denoising methods were based on Empirical Mode Decomposition (EMD). Predictive models employing the EMD method typically involve strategies where the decomposed EMD sub-sequences serve as multivariate inputs for a predictive model [11], or each decomposed sub-sequence is used to establish separate prediction models based on data characteristics, which are then combined for an overall prediction [12].

However, the predictive models involving the EMD method mentioned above did not fully address noise. In response, leveraging the characteristic that EMD decomposed sub-sequences have decreasing frequencies, Guo et al. [13] simply excluded the first sub-sequence from training. Although this method might not completely eliminate noise, it achieved satisfactory results due to the powerful fitting capabilities of neural networks. Jia et al. [14] introduced an approach on top of the EMD method, which used the correlation coefficient between the original sequence and the decomposed sub-sequences to determine the boundary between noise and signal. The signal above this boundary was treated as noise and removed. However, Wang et al. [15] demonstrated that the EMD method has limitations in practical prediction. During training, to avoid propagating prediction errors, actual values obtained from new measurements are used to replace predicted values. Adding newly acquired data points to the sequence significantly impacts the magnitudes of the atoms in the EMD decomposition columns, potentially rendering existing prediction models for the original sub-sequences unsuitable for the new decomposed sub-sequences. To tackle this issue, Wavelet Threshold Denoising (WTD) [16] presents a viable solution. The role of WTD is similar to the noise reduction concept based on the EMD method – eliminating high-frequency noise from the signal, thereby enhancing the predictability of the time series [17,18]. However, unlike the EMD method, WTD employs fixed basis functions for decomposition and reconstruction, meaning it is not influenced by the introduction of new data while processing and does not affect the decomposition of the original data.

Building upon the aforementioned discussions, this study employs Wavelet Threshold Denoising (WTD) to overcome the limitations of EMD reconstruction denoising, which hinder its normal usage in gas concentration prediction. The use of phase space reconstruction techniques helps uncover more valuable information from gas concentration time series, thereby enhancing predictive accuracy. The proposed approach introduces a WTD-PS-LSTM gas concentration prediction model, which combines Wavelet Threshold Denoising, Phase Space Reconstruction, and LSTM neural networks.

Two major innovations are presented in this paper: (1) Utilizing the WTD method to reduce the noise of the gas time data, which solves the problem that the commonly used EMD method does not work properly in the task of autoregressive prediction of gas concentration. (2) Combining the phase space reconstruction technique to process the data further improves the prediction accuracy of the model. In summary, the establishment of the WTD-PS-LSTM prediction model combines the advantages of these two methods, and is able to provide accurate gas concentration prediction, which provides an accurate basis for coal mine safety warning.

This paper is divided into six sections. Section 1 describes the significance of gas concentration prediction with the background of current research on gas concentration and the existing problems. Section 2 introduces the theory of the WTD method and the main points of the modeling implementation. Section 3 presents the theoretical points of Phase space reconstruction (PS). Section 4 first introduces the LSTM network, followed by coupling the WTD and PS methods and proposes a simulation implementation of the WTD-PS-LSTM model. Section 5 presents the data situation, the setting of all relevant parameters in the model, and the model simulation results and analysis. In section 6, the analysis of the simulation results is summarized and the limitations of the current approach of this paper and future work are discussed.

2. Wavelet threshold denoising

In the mining environment, various complex noise factors such as temperature, humidity, and pressure can lead to a certain degree of distortion in the data collected by sensors. Neural networks demand high-quality data for effective training. Despite their exceptional fitting capabilities, neural networks can easily learn irrelevant information due to the presence of noise. Moreover, the introduction of noise can decrease the predictability of time series data. Therefore, it is essential to apply denoising techniques to the data.

The fundamental idea behind Wavelet Threshold Denoising is to perform a wavelet decomposition of the data first and then set a threshold for the wavelet coefficients. Coefficients below this threshold are discarded, and the remaining coefficients are subjected to a wavelet inverse transform to obtain a denoised signal. Thus, the process can be divided into four sequential steps (The focus of parameter selection is the first three steps).

1. Selecting a Suitable Wavelet Basis. it is essential to choose an appropriate wavelet basis for decomposing the original data.
2. Choosing a Threshold. an appropriate threshold value needs to be selected as the threshold limit for noise in the wavelet domain.
3. Applying the Thresholding Function. The thresholding function is then applied to the wavelet coefficients containing noise to filter out the noise coefficients.
4. Wavelet Reconstruction. the denoised wavelet coefficients are subjected to a wavelet reconstruction to obtain the denoised signal.

This process effectively separates noise from the signal by operating in the wavelet domain, allowing for improved data quality and enhanced predictability, particularly when used in conjunction with neural networks for tasks such as gas concentration prediction. The flowchart of the process is shown in Fig. 1.

2.1. Discrete wavelet decomposition

The first step of the Wavelet Threshold Denoising process involves the mathematical procedure of decomposing the original data $x(t)$ through wavelet decomposition as follows [19]:

$$x(t) = \sum_n c_{m,n} 2^{-\frac{m}{2}} \psi(2^{-m}t - n) + \sum_n d_{m,n} 2^{-\frac{m}{2}} \varphi(2^{-m}t - n) \tag{1}$$

$$c_{m,n} = \langle x(t), \psi_{m,n}(t) \rangle = 2^{-\frac{m}{2}} \int_{-\infty}^{+\infty} x(t) \overline{\psi(2^{-m}t - n)} dt \tag{2}$$

$$d_{m,n} = \langle x(t), \varphi_{m,n}(t) \rangle = 2^{-\frac{m}{2}} \int_{-\infty}^{+\infty} x(t) \overline{\varphi(2^{-m}t - n)} dt \tag{3}$$

In the equation, $\psi(\cdot)$ represents the mother wavelet function of $x(t)$ in the V_m space, and $\varphi(\cdot)$ represents the mother wavelet function of $x(t)$ in the W_m space. W_m is the complementary space of V_{m-1}. The term $c_{m,n}$ corresponds to the low-frequency coefficients (approximation coefficients), and $d_{m,n}$ corresponds to the high-frequency coefficients (detail coefficients). m is the hierarchy of low and high frequency coefficients; n is the number of coefficients in each hierarchy.

It's worth noting that the choice of different wavelet functions can affect the effectiveness of signal decomposition and reconstruction. For signals that are not sensitive to phase information, selecting a wavelet from the db family that produces stable and smooth transformed signals is generally suitable. However, since gas concentration sequences are time series data, using wavelets with linear phase characteristics is important to reduce nonlinear phase-induced time-domain aliasing distortions during reconstruction. Therefore, an improved version of the db family, specifically the sym4 wavelet from the *sym* family, is chosen due to its approximately linear phase characteristics.

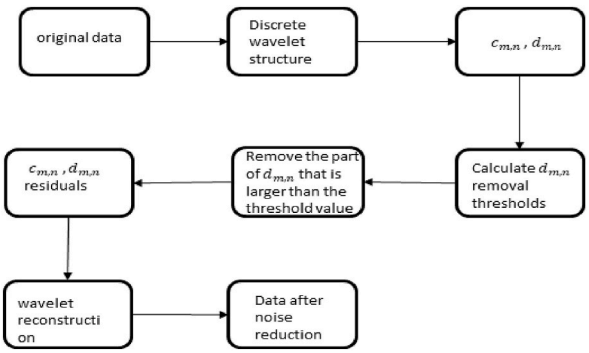


Fig. 1. Wavelet thresholding method denoising flowchart.

2.2. Selection of wavelet denoising threshold

The unbiased risk estimation method is selected to determine the threshold. Unbiased risk estimation method (rigrsure): A threshold selection method based on the principle of unbiased likelihood estimation that accurately estimates the risk of a parameter or model. It calculates the risk value for each threshold and selects the one with the minimum risk. This approach ensures the preservation of useful signal components in a cautious manner to prevent data distortion. The specific algorithm for the "rigrsure" method is as follows.

Step 1: Take the absolute value of each element in the signal $x(t)$ of length L, sort them in ascending order, and then calculate the average for each element:

$$f(I) = (sort(|x(t)|))^2, (I = 0, 1, \ldots, L-1) \tag{4}$$

Step 2: Set the threshold as the square root of the k-th element of $f(I)$:

$$\lambda_I = \sqrt{f(I)}, (I = 0, 1, \ldots, L-1) \tag{5}$$

The resulting risk is calculated as follows:

$$Rish(I) = \left[L - 2I + \sum_{i=1}^{I} f(i) + (L-I)f(L-I) \right] \Big/ L \tag{6}$$

Step 3: Identify the value I_{min} corresponding to the minimum point on the risk curve $Rish(I)$ and obtain the threshold:

$$\lambda_I = \sqrt{f(I_{min})} \tag{7}$$

2.3. Selection of threshold function

Once the threshold value is determined, the next step involves choosing a function to filter the wavelet coefficients. Two commonly used threshold functions are the hard threshold and the soft threshold. Their formulas are as follows:

Hard threshold function:

$$w_{thr} = \begin{cases} w & |w| \geq thr \\ 0 & |w| < thr \end{cases} \tag{8}$$

Soft Threshold Functions:

$$w_{thr} = \begin{cases} [sgn(w)](|w| - thr) & |w| \geq thr \\ 0 & |w| < thr \end{cases} \tag{9}$$

Signals denoised using the soft threshold function tend to appear smoother in the time domain compared to signals processed with the hard threshold function, resulting in better filtering effects. However, in terms of mean square error, the hard threshold function is superior to the soft threshold function. Soft threshold denoising can compress signals and introduce certain biases, directly affecting how closely the reconstructed signal approximates the true signal. Therefore, it's important to strike a balance. Excessively pursuing filtering effectiveness might inadvertently remove useful information and lead to distortion. Taking into account the robust fitting capability of neural networks to noise, adopting the hard threshold filtering method is a prudent choice.

3. Phase space reconstruction of gas concentration sequences

Phase space reconstruction can unearth valuable information from chaotic time series, revealing the trajectories of chaotic motion and enhancing the predictability of the data. For the phase space reconstruction of chaotic time series, in 1980, Packard, Farmer, and others proposed the theory of phase space reconstruction through the methods of derivative reconstruction and time delay coordinate reconstruction. The latter involves determining the time delay and embedding dimension to decompose the original one-dimensional time series into a phase space matrix. Determining whether a time series is chaotic and predicting its behavior often occurs in this reconstructed phase space [20]. The phase space reconstruction matrix is as follows:

$$\begin{aligned} V &= [V_1, V_2, \cdots V_{N-1}, V_N] \\ V_1 &: x(t_1), x(t_1+\tau), \cdots, x(t_1+(r-1)\tau) \\ V_2 &: x(t_2), x(t_2+\tau), \cdots, x(t_2+(r-1)\tau) \\ V_3 &: x(t_3), x(t_3+\tau), \cdots, x(t_3+(r-1)\tau) \\ &\vdots \\ V_N &: x(t_N), x(t_N+\tau), \cdots, x(t_N+(r-1)\tau) \end{aligned} \tag{10}$$

Here, V_N stands for phase point, and the sequence corresponding to each phase point forms its coordinates. The dimension of the

coordinates for each phase point corresponds to the embedding dimension. r stands for embedding dimension. $x(t)$ stands for time series. Its length is denoted by L. t stands for time. N stands for the number of phase points, $N = L - (r-1)\tau$. In terms of coordinates, this can be visualized as an $N \times r$ matrix. In this study, the correlation coefficient method (cc method) [21] is used to compute the embedding dimension r and time delay τ.

4. Prediction method based on WTD-PS-LSTM

The gas concentration prediction method based on the combination of Wavelet Threshold Denoising (WTD), Phase Space Reconstruction (PS), and LSTM neural networks involves several steps. First, the data is denoised using the Wavelet Threshold Denoising method. Then, the denoised data undergoes phase space reconstruction. In this process, each row of the reconstructed phase space matrix serves as an input to the LSTM neural network. Essentially, each column of the reconstructed matrix corresponds to a new feature vector, with a total of r columns representing the embedding dimension. This approach effectively transforms the data into a format suitable for input into the LSTM network for prediction.

4.1. Long Short-Term Memory (LSTM) neural network

While traditional backpropagation (BP) neural networks possess excellent linear and nonlinear approximation capabilities, they face limitations when dealing with sequential data, as in the case of gas concentration time series. Traditional feedforward neural network models struggle to establish connections between previous time steps. In contrast, Long Short-Term Memory (LSTM) neural networks, unlike their traditional counterparts, can learn to capture hidden states within the entire preceding time series, thereby establishing connections with previous time steps [22]. The architecture of an LSTM neural network is depicted in Fig. 2.

The operational process of the LSTM network is as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{11}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{12}$$

$$C_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{13}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot C_t \tag{14}$$

$$o_t = \sigma(W_0 \cdot [h_{t-1}, x_t] + b_0) \tag{15}$$

$$h_t = o_t \cdot tanh(C_t) \tag{16}$$

Where Equation (11) represents the process of the forget gate, Equations (12) and (13) denote the input gate process, Equation (14) captures the update of the cell information state, and Equations (15) and (16) describe the output gate process. In the above equations, x_t is the input at moment t and h_{t-1} is the output of the neural unit at moment $t-1$; W_f, W_i, W_0, W_c, b_f, b_i, b_0, and b_c are weight matrices and biases updated for each of these processes; C_t is the new information at moment t and C_t is the updated cell state; σ is the "sigmod activation function" of the neural unit; *tanh* is the activation function.

4.2. Construction of the WTD-PS-LSTM model

The model's phase space reconstruction, wavelet threshold denoising, and parameter calculations were all implemented in MATLAB 2014b. Following the method outlined in Section 1 using the wavelet threshold denoising approach, a comparison of data

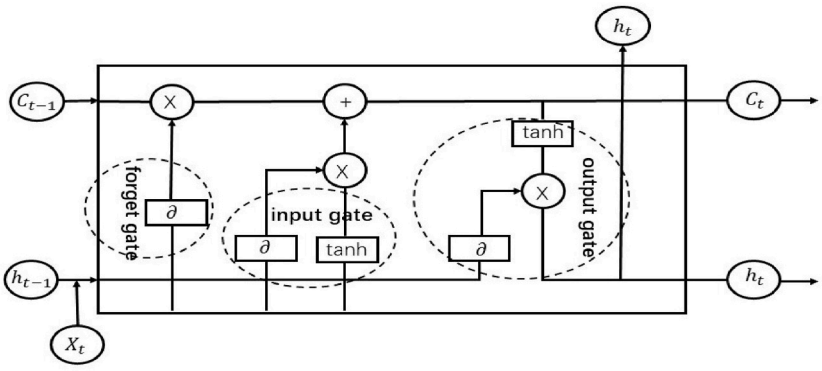


Fig. 2. LSTM structure diagram [23].

before and after filtering is shown in Fig. 3. Observing the comparison between the denoised and original data, it is evident that the denoised waveform still generally maintains the original trend of variation, with significant reduction in baseline spikes. In the figure, several points with high energy levels can be seen to have their abruptness and energy greatly diminished after filtering. From the perspective of signal processing science, these points are precisely where noise interference is severe. Consequently, the data's predictability has been significantly enhanced.

Using the correlation coefficient method (cc method) [21], the phase space reconstruction yielded a time delay of $\tau = 6$ and an embedding dimension of $r = 3$. Consequently, a reconstructed phase space matrix with $N = L - (r-1)\tau$ rows and r columns can be obtained.

The final integration and implementation of the WTD-PS-LSTM gas concentration prediction model were carried out using the Python programming language and the Keras library. The output layer of the LSTM consists of 1 neuron. Based on the c-c method, the embedding dimension $r = 3$, resulting in an input dimension of 3 for the LSTM. The number of hidden layers, the number of neurons in each hidden layer, the dropout rate between recurrent layers, and the time step were determined through an iterative trial-and-error process. The optimal configuration was set to 2 hidden layers with 100 neurons each, a dropout rate of 0.1 between recurrent layers, and a time step of 10. The model was trained for 100 epochs with a batch size of 60, while other parameters were set to their default values. In the existing literature, optimization for models often involves a choice between the classic stochastic gradient descent (SGD) optimizer and the newer Adam optimizer. Due to factors related to data and network structure, SGD tends to yield better results for computer vision tasks. However, this study focuses on predicting time series using recurrent neural networks, which share similarities with NLP tasks based on the Transformer model in terms of data structure. Therefore, the use of the Adam optimizer is more effective. In the experiments, it was observed that Adam indeed performs significantly better, while SGD struggled to converge. Fig. 3 illustrates the loss curves for the training and testing sets of the WTD-PS-LSTM and WTD-LSTM models. It can be observed that both models' losses eventually stabilize close to zero. Furthermore, the close alignment of the testing and training loss curves after convergence indicates strong generalization capabilities of the models, suggesting that the chosen method and hyperparameters for the neural network are appropriate. Additionally, it's worth noting that Fig. 4 shows that the loss and convergence performance of the WTD-PS-LSTM model are excellent after just 100 epochs. Therefore, for the sake of overall computational efficiency, an epoch count of 100 was chosen.

5. Experiments and results

5.1. Case description

The data for this experiment is sourced from a mine in Shanxi Province, consisting of a total of 790 data points sampled at 15-min intervals. As a result, the reconstructed phase space matrix after phase space reconstruction is of size 778 rows by 3 columns. The last 300 rows of the reconstructed matrix are utilized as the input matrix for the test set. Given that the optimal time step is set at 10, meaning that the LSTM considers historical information from 10 phase points (10 rows of the reconstructed matrix), each step of prediction comprises 290 experimental points. During code implementation, the temporal ordering of data points within each row of the reconstructed phase space matrix is arranged in decreasing order. Consequently, when considering a single time step as input, the corresponding output label for each row is the data in the first column of the subsequent row. When considering p time steps, the output label corresponds to the data in the first column of the subsequent row at the p-th time step.

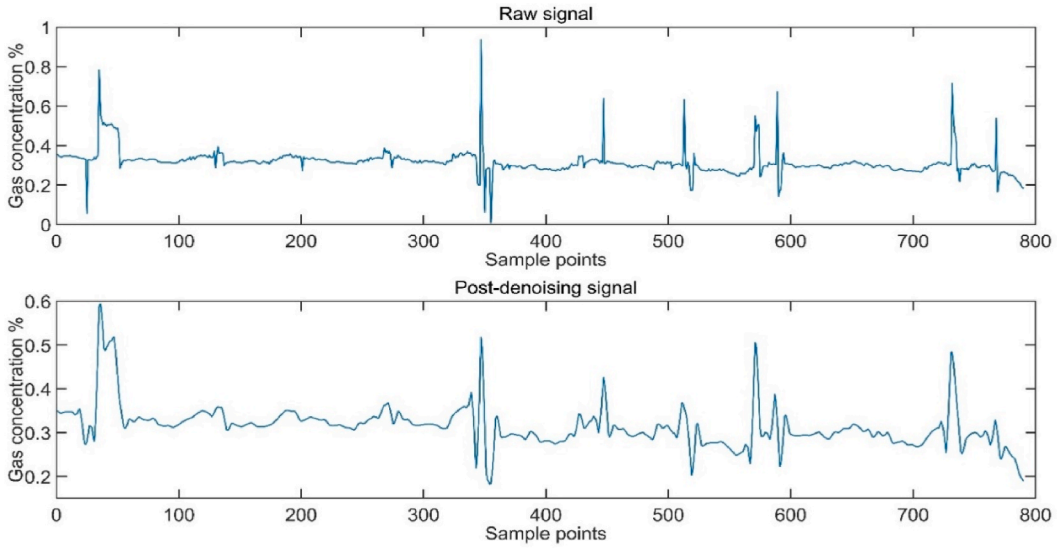
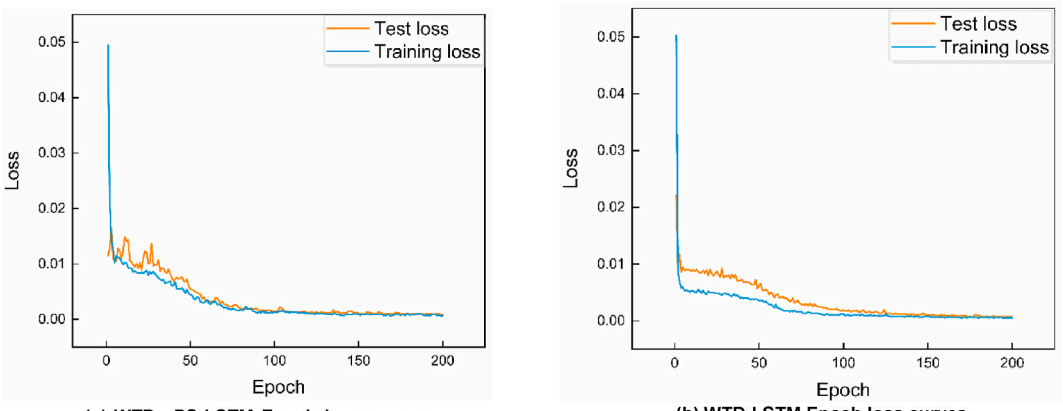


Fig. 3. Comparison before and after WTD filtering.



(a) WTD—PS-LSTM Epoch loss curves

(b) WTD-LSTM Epoch loss curves

Fig. 4. Loss curve.

5.2. Model performance evaluation metrics

In this study, the chosen performance evaluation metrics for prediction are the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The formulas for RMSE and MAE are provided in Equation (17):

$$\begin{aligned} Rmse &= \sqrt{\frac{1}{T}\sum_{i=1}^{T}|y_i - \hat{y}_i|^2} \\ Mae &= \frac{1}{T}\sum_{i=1}^{T}|y_i - \hat{y}_i| \end{aligned} \tag{17}$$

In the equation, y_i represents the i-th actual value; $\hat{y}_i$ represents the i-th corresponding predicted value; T is the total number of labels in the test set.

5.3. Predictive results analysis

In order to validate the effectiveness of the proposed WTD-PS-LSTM model and the necessity and correctness of the PS-LSTM algorithm module for phase space reconstruction and wavelet threshold denoising, a comparison is conducted with the WTD-LSTM model and the PS-LSTM control model.

Fig. 4 (b) presents the loss curve for the WTD-LSTM model. It can be qualitatively observed that the loss approaches the minimum and stabilizes around 150 epochs. Additionally, the final losses for the test and training sets are less close to 0 compared to the losses shown in Fig. 4 (a), and there exists a noticeable gap between the test and training loss curves. This indicates that the inclusion of the PS method enhances the visibility of data information in the model, allowing it to achieve better results with fewer epochs. Furthermore, without the PS method, the model's performance on the test set is inferior to that on the training set, suggesting relatively weaker generalization ability. For more quantitative experimental results, refer to Table 1. In Table 1, the performance of WTD-PS-LSTM models with 1, 10, 30, and 50 steps is compared with the WTD-LSTM model with 13 steps. The term "1-step" for WTD-PS-LSTM indicates that the model reviews the history of 1 phase point coordinate in the phase space. In contrast, the term "13-dimensional" refers to the WTD-LSTM model without phase space reconstruction, where a sliding window of length 13 is used as input directly without the segmentation process of time delay $\tau = 6$ and embedding dimension $r = 3$. Comparing these two scenarios, it is evident that the

Table 1
Phase space reconstruction analysis.

PS versus no PS experimental group	Mae	Rmse
WTD-PS-LSTM 1 step	0.038	0.065
WTD-LSTM 13 dimensions	0.053	0.075
WTD-PS-LSTM 10 steps	0.024	0.039
WTD-LSTM 22 dimensional	0.037	0.052
WTD-PS-LSTM 30 steps	0.031	0.047
WTD-LSTM 42 dimensions	0.043	0.056
WTD-PS-LSTM 50 steps	0.036	0.057
WTD-LSTM 62 dimensions	0.046	0.068

predictive accuracy of the WTD-LSTM model with no phase space reconstruction is lower. Similar comparisons are made for other corresponding time window dimensions, as shown in Table 1. Comprehensive analysis of Table 1 reveals that even though the non-reconstructed group's time windows may contain more or an equal number of data points compared to the reconstructed group's corresponding time steps, the reconstructed group achieves better performance. This implies that the time delay τ and embedding dimension r play a role in phase space reconstruction. The time delay τ and embedding dimension r contribute to the segmentation and extraction of directly useful information from the original data, thus enhancing data predictability.

Fig. 5 presents a comparison of multi-step prediction performance among all models. Due to the extreme sensitivity of chaotic systems to initial conditions, even a tiny difference in initial values can lead to significant divergence after a few iterations. This divergence becomes more pronounced in systems with stronger chaotic behavior. As a result, accurate long-term prediction is impossible for chaotic systems. With increasing prediction time, the accuracy of predictions becomes increasingly inaccurate, and only limited short-term predictions are feasible. Therefore, it is expected that the multi-step predictions of all models will become less accurate as the prediction steps increase. However, upon comparing the respective prediction steps of the models, it is observed that WTD-PS-LSTM slightly outperforms WTD-LSTM. This indicates that phase space reconstruction contributes to enhancing prediction accuracy by extracting information from chaotic systems. Furthermore, the prediction accuracy of both models is significantly better than that of the PS-LSTM model, indicating the beneficial impact of WTD denoising on improving prediction accuracy. Table 2 provides a quantitative performance evaluation of the one-step prediction model among the aforementioned models.

6. Conclusion

In this paper, a gas concentration prediction model based on WTD-PS-LSTM is proposed considering the use of phase space

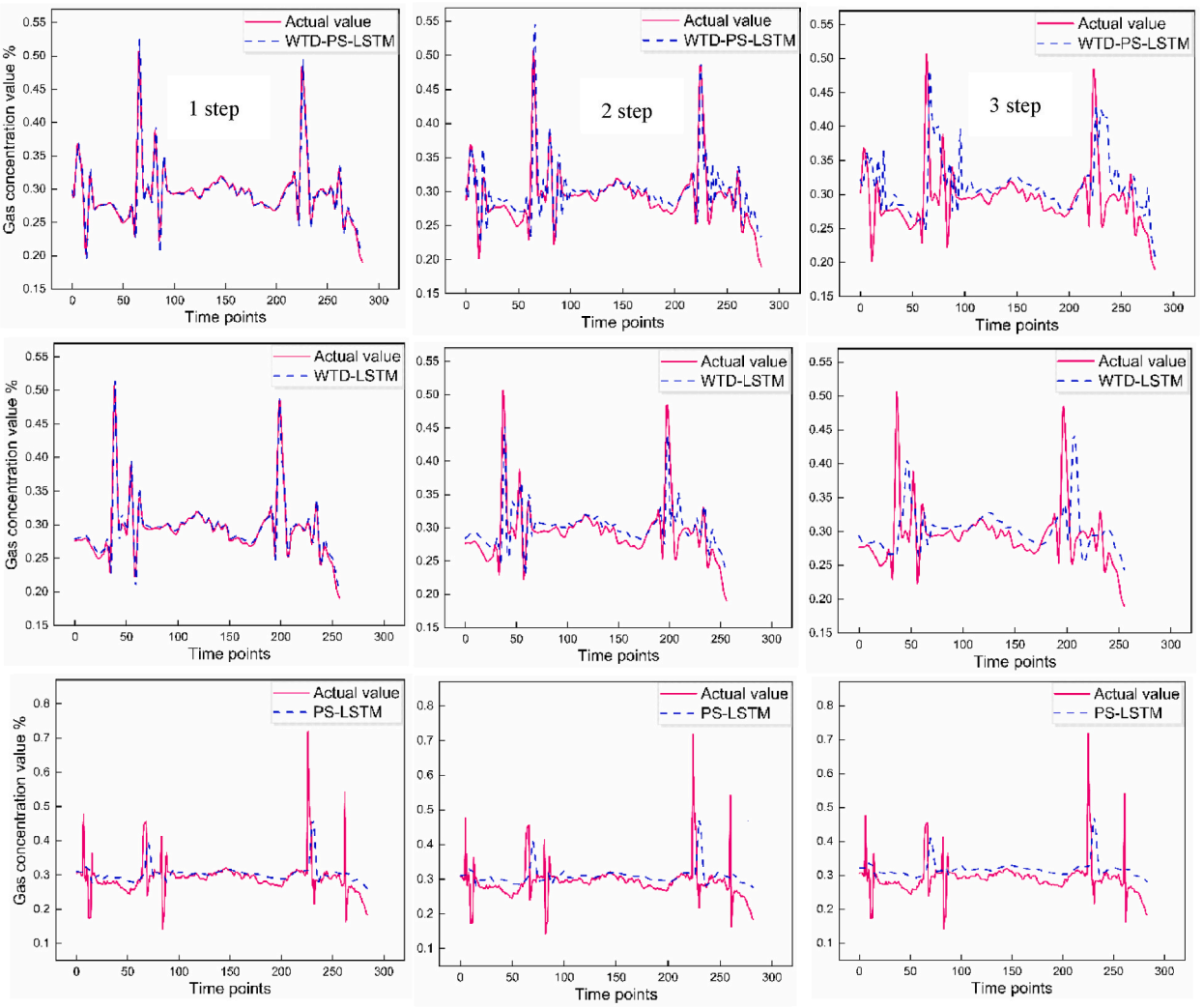


Fig. 5. Multistep prediction.

Table 2
One-step predictive model performance evaluation.

Predictive Models	WTD-PS-LSTM	WTD-LSTM	PS-LSTM	LSTM
Mae	0.024	0.037	0.105	0.129
Rmse	0.039	0.052	0.195	0.225

reconstruction (PS) and the problem that the traditional EMD denoising method cannot be used properly in reality. The model utilizes the WTD method to solve the problem that EMD method denoising cannot be used normally in reality, while the simulation experiment results are tabulated: under one-step prediction experiments, the PS-LSTM model, the WTD -LSTM model, the LSTM model, and the WTD-PS-LSTM prediction model Mae (Mean Absolute Error) and Rmse (Root Mean Squared Error) were reduced by 77.1%, 77.4%, 35.1%, 81.4%, and 81.4%, respectively, 35.1%, and 81.4%, with 80%, 25%, and 82.6%, respectively. Therefore, it can predict the trend of gas concentration more accurately.

The purpose of the WTD method used in this paper is to solve the problem that the EMD method noise reduction does not work properly in real time series prediction tasks as envisioned in the simulation, so the optimization algorithm of the WTD method has not been investigated in depth. The same is true for the PS method used in this paper; currently there are several algorithms with different perspectives for the PS method, and only one of the algorithms from one of these perspectives is utilized in this paper. On the basis of the model in this paper, in order to obtain better filtering effect as well as prediction accuracy, it can be improved from the optimization point of view of WTD and PS methods.

Ethics declarations

All participants/patients (or their proxies/legal guardians) provided informed consent to participate in the study.

Additional information

No additional information is available for this paper.

CRediT authorship contribution statement

Kun Gao: Project administration, Conceptualization. **ZuoJin Zhou:** Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] F.B. Zhou, T.Q. Xia, X.X. Wang, et al., Recent developments in coal mine methane extraction and utilization in China: a review, J. Nat. Gas Sci. Eng. 31 (2016) 437–458, https://doi.org/10.1016/j.jngse.2016.03.027.
[2] M.Q. Li, H.L. Z, T.L. Zhen, et al., Characteristics and precursor information of electromagnetic signals of mining-induced coal and gas outburst, J. Loss Prev. Process. Ind. 54 (2018) 206–215, https://doi.org/10.1016/j.jlp.2018.04.004.
[3] Y.M. Zhou, On the disaster factors of coal mine safety accidents in China, Inner Mongolia Coal Economy (9) (2016) 82–83, https://doi.org/10.13487/j.cnki.imce.008064.
[4] Y. Geng, Gas concentration prediction based on chaotic particle swarm neural network, China Coal 43 (2017) 124–129, https://doi.org/10.19880/j.cnki.ccm.2017.03.027.
[5] H. Fu, Y.Z. Liu, N. Xu, et al., Research on gas concentration prediction based on multi-sensor-deep long short-term memory network fusion, Chin. J. Sensors Actuators 34 (2021) 784–790, https://doi.org/10.3969/j.issn.1004-1699.2021.06.011.
[6] Y.T. Wang, J.L. Li, R.P. Gu, Flight operation risk prediction model based on multivariate chaotic time series, Chinese Journal of Engineering 42 (2020) 1664–1673, https://doi.org/10.13374/j.issn2095-9389.2019.12.09.002.
[7] M. Ding, H.B. Yu, L. Liu, et al., Photovoltaic power prediction method based on multivariate phase spatial reconstruction and RBF neural network, J. Electron. Meas. Instrum. 34 (2020) 1–7, https://doi.org/10.13382/j.jemi.B1902772.
[8] S.G. Wang, P. Li, T.L. Lou, et al., PS-LSTM pollutant prediction model based on multivariate chaotic time series, Transducer and Microsystems 41 (2022) 117–120, https://doi.org/10.13873/J.1000-9787(2022)04-0117-04.
[9] K. He, X.Y. Li, Wind speed time series prediction based on neural network model improved by empirical mode decomposition and genetic algorithm, Electronic Measurement Technology 41 (8) (2018) 146–149, https://doi.org/10.19651/j.cnki.emt.1701311.
[10] J.N. Bai, Research on the Prediction Model of Gas Outburst Based on Improved GRU [D], Liaoning University of Engineering and Technology, 2022, pp. 37–55, https://doi.org/10.27210/d.cnki.glnju.2022.000282.
[11] C. Zhang, H. Wei, J. Zhao, T. Liu, T. Zhu, Short-term wind speed forecasting using empirical mode decomposition and feature selection, Renew. Energy 96 (2016) 727–737, https://doi.org/10.1016/j.renene.2016.05.023.
[12] H. Liu, C. Chen, H.Q. Tian, Y.F. Li, A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks, Renew. Energy 48 (2012) 545–556, https://doi.org/10.1016/j.renene.2012.06.012.
[13] Z. Guo, W. Zhao, H. Lu, J. Wang, Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model, Renew. Energy 37 (2011) 241–249, https://doi.org/10.1016/j.renene.2011.06.023.

[14] R.S. Jia, T.B. Zhao, H.M. Sun, et al., Noise reduction method for microseismic signals based on empirical modal decomposition and independent component analysis, J. Geophys. 58 (2015) 1013–1023, https://doi.org/10.6038/cjg20150326.
[15] Y. Wang, L. W, On practical challenges of decomposition-based hybrid forecasting algorithms for wind speed and solar irradiation, Energy 112 (2016) 208–220, https://doi.org/10.1016/j.energy.2016.06.075.
[16] Z. Peng, S. P, L. Fu, B. Lu, J. Tang, et al., A novel deep learning ensemble model with data denoising for short-term wind speed forecasting, Energy Convers. Manag. 207 (2020), https://doi.org/10.1016/j.enconman.2020.112524, 112524-112524.
[17] Z.F. Lei, W.B. Su, Q. Hu, Multimode decomposition and wavelet threshold denoising of mold level based on mutual information entropy, Entropy 21 (2019), https://doi.org/10.3390/e21020202, 202-202.
[18] X.W. Mi, H. Liu, Y.F. Li, Wind speed forecasting method using wavelet, extreme learning machine and outlier correction algorithm, Energy Convers. Manag. 151 (2017) 709–722, https://doi.org/10.1016/j.enconman.2017.09.034.
[19] V. Sharma, D. Yang, et al., Short term solar irradiance forecasting using a mixed wavelet neural network, Renew. Energy 90 (2016) 481–492, https://doi.org/10.1016/j.renene.2016.01.020.
[20] T.B. Sun, Y.H. Liu, Chaotic time series forecasting based on fuzzy information granulation and hybrid neural network, Inf. Control 51 (2022) 671–679, https://doi.org/10.13976/j.cnki.xk.2022.2018.
[21] Z.B. Lu, Z.M. Cai, K.Y. Jiang, Determination of embedding parameters for phase space reconstruction based on improved C-C method, J. Syst. Simul. 11 (2007) 2527–2529, https://doi.org/10.16182/j.cnki.joss.2007.11.036.
[22] Y.H. Mao, C.C. Sun, L.Y. Xu, et al., Review of time series prediction methods based on deep learning [J/OL], Microelectron. Comput. 4 (2023) 8–17, https://doi.org/10.19304/J.ISSN1000-7180.2022.0725.
[23] Y. Liu, C.Y. Yang, Multifactor-based LSTM gas concentration prediction model, China Safety Production Science and Technology 18 (1) (2022) 108–113.