ITC 3/53
Information Technology and Control
Vol. 53 / No. 3/ 2024
pp. 846-864
DOI 10.5755/j01.itc.53.3.35062

A CEEMDAN-based Stacking Ensemble Learning Method for SO_2 Emission Forecast in a Wet FGD Process

Received 2023/09/09 | Accepted after revision 2023/11/19

HOW TO CITE: Lu, J. (2024). Li, X., Liu, Q., Wang, K. (2024). A CEEMDAN-based Stacking Ensemble Learning Method for SO_2 Emission Forecast in a Wet FGD Process. *Information Technology and Control*, 53(3), 846-864. https://doi.org/10.5755/j01.itc.53.3.35062

A CEEMDAN-based Stacking Ensemble Learning Method for SO_2 Emission Forecast in a Wet FGD Process

Xiaoli Li, Quanbo Liu, Kang Wang

Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
e-mails: lixiaolibjut@bjut.edu.cn; liuquanbobjut@163.com; wangkang@bjut.edu.cn

Corresponding author: liuquanbobjut@163.com

There has recently been increasing attention paid to sulphur dioxide (SO_2) pollution owing to its hazardous effect on both human health and atmospheric environment. To handle this problem, the wet flue gas desulphurization (FGD) system has found wide applications in SO_2 emitting industries. Accurate prediction of SO_2 emissions in treated flue gas serves the purpose of providing timely operating guidance for the FGD system. However, the wet FGD process is characterized by highly nonlinear dynamics and non-stationarity, which poses significant difficulties and limitations for traditional modeling methods. To address above issues, in this article, an integrated model is proposed to perform SO_2 emission forecasting for an FGD process. Our integrated model comprises a multiplicity of techniques, including complete ensemble empirical mode decomposition (EMD) with adaptive noise CEEMDAN stacking ensemble learning (SEL) and permutation-based entropy (PEN). The CEEMDAN serves as decomposing SO_2 emission signal, then the complexity of each decomposed sub-series is analyzed by PEN and ones with similar scores are combined, finally a stacking-based ensemble learning model which incorporates different types of member models are developed for modeling purposes. The proposed method was validated and evaluated by measurements of a real FGD system in a 600MW coal-fired unit, and experimental results illustrate the superiority of our method.

KEYWORDS: Wet flue gas desulfurization, Ensemble learning, Stacking, CEEMDAN-forecasting, Neural network.

1. Introduction

Due to massive coal combustion in electricity-generating plants, SO_2 contained in flue gas has become one of the primary atmospheric pollutants, which negatively affects both human health and environment. To abate SO_2 emissions arise from industrial processes, a series of strict SO_2 emission standards have issued by governments worldwide, and the limit on the maximum concentration of outlet SO_2 is determined to be 35mg·m^{-3} in China [27]. To meet such an ultralow SO_2 emission regulation, the post-combustion wet FGD technology is extensively employed in fossil-fueled power stations owing to its cost-effectiveness and high SO_2 removal efficiency. Accurate prediction of SO_2 concentration in treated stack gas serves two purposes: 1) remind the public to take precautionary measures in cases where the predicted SO_2 concentration exceeds the limit and 2) provide guidance for the operation of an FGD system. However, the wet FGD process is characterized by nonlinear dynamics of high complexity and non-stationarity, which pose great challenges to develop an accurate model used for FGD process description and further for SO_2 emissions prediction. The research for FGD process modeling is of fundamental importance to controller design and desulfurization system optimization. According to a wealth of studies carried out in this direction, the modeling approach is broadly categorized into two groups: first-principles modeling approaches and black-box modeling approaches.

The first-principles modeling approach is physical-chemical relationship based, which includes reaction kinetics, thermodynamic, as well as mass/energy conservation relationships. [9] developed a SO_2 removal model with the use of double-film theory, chemical improvement coefficients and a control mechanism for sulfate concentration. Then the established model was employed for predicting desulphurization efficiency, and experimental reselts show that predicted values closely approximate targets. A simplified model for wet FGD system with absorber was developed in [30], the model is built based upon movement equations of limestone slurry and material balance, and used to simulate the mass transfer of SO_2 in the absorption zone. Experimental results demonstrate the superiority of proposed modeling approach, which is also of great practical value for the FGD system design. [31] developed a Eulerian model on the basis of chemistry in the SO_2 absorption phase, and the double-film theory was also incorporated into the model derivation process. The effectiveness of proposed model is verified by measurements of a real FGD process. A rate-based FGD system model is suggested by [32], where dynamic characteristics of the studied FGD process is derived based on the knowledge of gas-liquid equilibrium, and experimental studies were conducted to verify the established model.

Due to high complexity of the FGD process, the practical implementation of mechanism-based modeling approaches is quite limited. On the other hand, advances in sensor technology enable acquisition of massive amounts of operation data. For above reasons, data-driven approaches are paid wide attention in the FGD modeling field. Based on historical running data from a real FGD process, a regression model was developed in [11] with the least square approach. Then the resulting model was used for multi-objective programming purpose, so as to ensure the FGD system operates in a safe and economical fashion. Artificial neural network, as a universal nonlinearity approximator, is typically employed to model the FGD process. A combined control strategy is proposed in [22] to perform pH control in FGD plants, the strategy incorporates a network-based controller and a feedback controller, and the network is introduced to approximate the nonlinear mapping from the manipulated variable space to the controlled variable space. In [21], to model the FGD process under study, a neural model was constructed by taking multiple FGD process variables (e.g. inlet SO_2 temperature) as inputs, and the established model was in turn used to investigate optimum operating ranges of FGD variables. [10] designed a combined FGD model that is composed of a mechanistic model and a network-based model to predict SO_2 emissions, experimental results suggest that satisfactory prediction performance can be achieved with the proposed method.

Apart from the CEEMDAN method, empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD) and variational mode decomposition (VMD) are representative of signal decomposition techniques. The major distinction between

EMD based techniques and VMD is that the former performs the signal decomposition adaptively and in a data-driven fashion, while VMD is a nonrecursive decomposition method and decomposed modes are tight their center frequency. Despite decomposition mechanisms for above-mentioned methods are slightly different, they all essentially serve to decompose the signal into intrinsic mode functions (IMFs) and a residue with various frequencies and scales. In recent years, signal decomposition techniques are increasingly incorporated into a prediction model to handle time series modelling problems in different contexts. Among these, to mention but a few, [1] adopted EMD technique to augment model training data, and a nonlinear auto-regressive network with exogenous inputs (NARX) is employed to validate this method. Experimental results show that predictive performance of NARX is improved with the EMD-based data augmentation method. In [6], a noise cancellation method is innovatively designed based upon the EMD method, and the proposed approach is proven to have better denoising performance than classical methods by jitter analysis. Authors in [8] presented an improved EMD method named BoostEMD, which has the ability to obtain IMFs with higher frequencies, and electromyography signal is employed to verify its enhancement in denoising performance. VMD method is employed to process multi-frequency signals in [15] such that the satisfactory imaging effect could be achieved, and the proposed method is successfully applied to electrical impedance tomography in practice. [20] integrated VMD, empirical wavelet transform (EWT) and long short-term memory (LSTM) to form a prediction model, which is used to predict cutterhead torque for shield tunneling machine, and high-precision prediction can be achieved with the hybrid model.

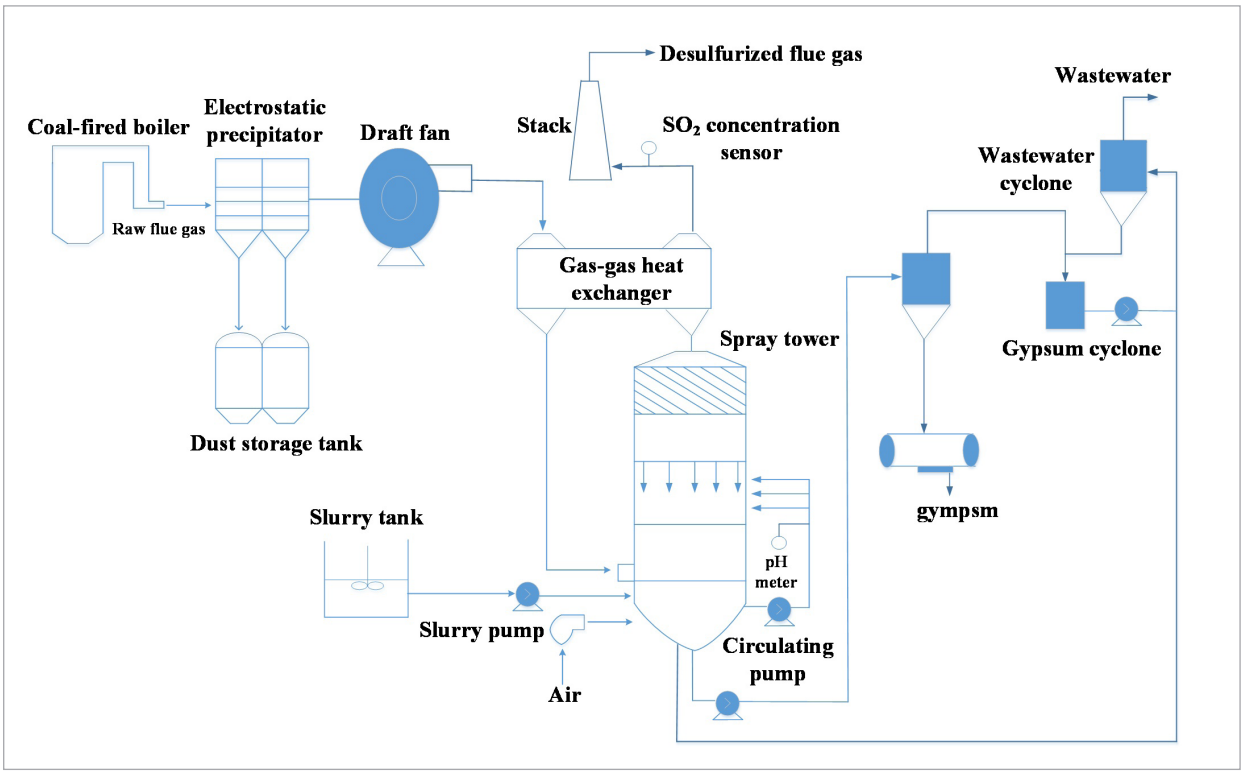
In this study, a dynamic model of the flue gas desulfurization process is developed, which innovatively hybridizes the signal decomposition technique CEEMDAN, artificial intelligence (AI)-based models, permutation entropy and stacking ensemble learning (SEL), is proposed. To our knowledge, this is the first work to combine time-frequency data analysis method with SEL strategy and used for modeling a real FGD process. The outstanding forecasting performance indicates that our approach can provide a promising solution for SO_2 emission forecasting, besides the significant flexibility structure of our model makes it applicable to various industrial process modeling scenarios.

This study is structured as follows. The "FGD process description" section provides a comprehensive description of the wet FGD process under study. In "Methodology" section, we give a general description of related techniques utilized in our method. The design of proposed modeling framework is detailed in "Proposed modeling approach" section. "Experiments and analysis" tests effectiveness of our approach through experiments designed based upon measurements of a real FGD process. Conclusions are drawn in the "Conclusion" section.

2. FGD Process Description

A schematic of a typical wet FGD process is shown in Figure 1 below, where key components are indicated in bold. The countercurrent spray tower plays a central role in the flue gas cleaning system, according to chemical reactions taking place during the whole desulphurization process, the removal of SO_2 in flue gas can be decomposed into four phases: calcium carbonate ($CaCO_3$) dissolution, sulphur dioxide absorption, sulfite oxidation and calcium sulfate ($CaSO_4$) crystallisation. The process of FGD can be summarized as follows: The raw flue gas is first produced by a coal-fired boiler, then it passes through an electrostatic precipitator so as to remove contained fine particles. After that, the flue gas with no dust flows into the gas-gas heat exchanger, where heat transfer is carried out between the inlet and treated stack gas. Specifically, the inlet stack gas temperature would decrease from around 100°C to below 80°C, which can decrease PM2.5 produced in the FGD tower by a factor of about 10 to 100; and at the same time, temperature rise in the treated stack gas is beneficial to gypsum rain elimination. Finally, the processed flue gas is fed to the desulphurizing tower from the top. As for the flue gas desulfurizer limestone, it is crushed and mixed with water to produce limestone slurry. The resulting slurry is stored in a slurry tank for subsequent feeding to the FGD tower, and it is transferred through slurry pumps that are mounted at the lower part of the spray scrubber. As it is shown, limestone desulfurizer is sprayed down via upper nozzles, while the stack gas flows countercurrently through slurry

Figure 1
A schematic illustration of a wet FGD process



circulation pumps, then the chemical reaction between limestone and the SO_2 in flue gas would take place. According to reaction shown in Equation (1), the primary product between SO_2 and $CaCO_3$ is hydrated $CaSO_3$.

$$SO_{2(g)} + CaCO_{3(s)} + 0.5H_2O_{(aq)} \rightarrow CO_{2(g)} + CaSO_3 \cdot 0.5H_2O_{(aq)} \quad (1)$$

At the bottom part of the absorber, the reaction product can be handled in two different ways: natural oxidation and forced oxidation. The main difference between two oxidation modes is whether oxygen is sparged into the scrubber purposely. As for nature-based oxidization, its primary products are $CaSO_3 \cdot \frac{1}{2}H_2O$ and $CaSO_4 \cdot 2H_2O$ that are mixed in the sludge, rendering the dewatering task a truly challenge; by contrast, the end product of forced oxidation is 90% (see Equation (2)), which can minimize the scaling problem to the highest possible extent.

$$\frac{1}{2}O_{2(g)} + CaSO_3 \cdot \frac{1}{2}H_2O_{(aq)} + \frac{3}{2}H_2O_{(aq)} \rightarrow CaSO_4 \cdot 2H_2O_{(aq)} \quad (2)$$

The reaction product is FGD-gypsum which is a synthetic solid byproduct, and finds applications in both plaster and wallboard manufacture. Moreover, the whole FGD process consumes considerable amounts of water, which includes crystallization water for gypsum and wash water for removing chlorides. As a consequence, cyclones are installed to separate gypsum from the water that is recycled and circulated around the FGD loop.

3. Methodology

A timeseries is defined as observed values that are ordered based on chronological sequence of time. In this study, SO_2 emissions processed by an FGD process is a univariate time series that can be represented as

$Y=[y_0,y_1,\cdots,y_N]^T$, where y_t denotes the observation at instant t. Suppose we are given an input sequence $X=(x_0,x_1,\cdots,x_N)^T$, we wish to predict corresponding outputs $y_0,y_1,\cdots,y_N$ at each time. This is realized through some mathematical model, more formally, a sequence model is any mapping $f:X^{N+1}\to Y^{N+1}$ that produces the mapping $\hat{y}_0,y_1,\cdots,y_N=f(x_0,\cdots,x_N)$. In the generic case f is approximated by only a single model, which greatly increases the risk of model insufficiency and misspecification. This paper presents a general methodology to establish forecasting model for SO_2 emissions, the motivation behind it is twofold: 1) improve the forecasting ability by integrating AI-based models with the stacking ensemble learning, 2) decrease the forecasting difficulty with signal processing techniques and information theory. This section is devoted to providing brief introductions to core components in the proposed model.

3.1. CEEMDAN

CEEMDAN is a data-driven approach for non-linear and nonstationary signal analysis, it is developed based upon EMD and ensemble EMD. The EMD and its variants take advantage of sifting processes to enable decomposition of the timeseries into a multiplicity of modulated subseries termed as intrinsic mode functions (IMFs), then IMFs can be employed as bases to restore the original signal. From the perspective of time series modeling, IMF components are relatively stationary subseries and possess more regularity and stability than original series, which can substantially simplify the modeling problem and hence EMD-based modeling techniques find wide applications in various fields (e.g. wind power, crude oil and water quality). Despite EMD shows exceptional performance in signal decomposition, it always subject to "mode mixing" and "aliasing" problems. In order to handle above problems, ensemble EMD is proposed in [25], whose basic idea is to add varied realizations of white-noise signals to the signal to be decomposed and calculate mean IMFs. CEEMDAN is an improvement of ensemble EMD, which can not only ensure decomposition completeness, but also reduce the sifting iteration number and computational cost. Let $E_j(\cdot)$ be an operator that calculates the jth mode by EMD, $w_i(t)\sim N(0,1)$ denotes white noise with Gaussian distribution and $S_i(t)$ is the time series to be processed. The is performed in the following steps:

1 Given a time series $S_i(t)=S(t)+\varepsilon_0 w_i(t)\,(i=1,2,\cdots,I)$, use EMD to extract its first mode IMF defined as $\overline{IMF_1}=\sum_{i=1}^{I}IMF_{i1}/I$, and ε_0 is a noise coefficient. Then the first residue is calculated as $r_1(t)=S(t)-\overline{IMF_1}$.

2 Construct the noise term on the basis of $\overline{IMF_1}$, white noise $w_i(t)$ and noise coefficient ε_1. Then add the noise term to $r_1(t)$ and a new time series is built, the second mode $\overline{IMF_2}$ is extracted from the new time series as $\overline{IMF_2}=\sum_{i=1}^{I}E_1(r_1(t)+\varepsilon_1E_1(w_i(t)))/I$, the second residue $r_2(t)=r_1(t)-\overline{IMF_2}$.

3 As with step (2), with white noise $w_i(t)$ $(i=1,2,\cdots,I)$ and noise coefficient ε_k ($k=2,3,\cdots,K$, K is the total number of IMFs), develop new noise terms and thus new time series to be decomposed. Then the kth IMF is $\overline{IMF_k}=\sum_{i=1}^{I}E_1(r_{k-1}(t)+\varepsilon_{k-1}E_{k-1}(w_i(t)))/I$ and the corresponding residue $r_k(t)=r_{k-1}(t)-\overline{IMF_k}$.

4 When the residue cannot be decomposed by EMD, the decomposition process terminates. The final residue is $R(t)=S(t)-\sum_{k=1}^{K}\overline{IMF_k}$, which represents the trend of time series $S(t)$.

3.2. Permutation-based Entropy

As a statistical measure that is used for quantifying complexity of a signal, PEN [2] benefits from low computational cost, simplicity and robustness with respect to non–linear monotonous transformations, and thus has found wide applications in many research fields. Given a time series $\{x(i),i=1,\cdots,N\}$, according to the Takens–Maine theorem, its phase space is formulated as

$$X(i)=\{x(i),x(i+\tau),\cdots,x(i+(m-1)\tau)\}\quad i=1,2,\cdots,N-(m-1)\tau \tag{3}$$

in which m and τ are the embedded dimension and time delay, respectively. As for each $X(i)$, its m real valued elements are sorted in descending order as

$$\{x(i+(j_1-1)\tau)\le x(i+(j_2-1)\tau)\le\cdots\le x(i+(j_m-1)\tau)\} \tag{4}$$

In cases where elements in $X(i)$ are of the same value (e.g. $x(i+(j_1-1)\tau)=x(i+(j_2-1)\tau)$), we have $j_1\le j_2$, $x(i+(j_1-1)\tau)\le x(i+(j_2-1)\tau)$. As a result, a mapping

between $X(i)$ and $(j_1, j_2, \cdots j_m)$ can be established, where $(j_1, j_2, \cdots j_m)$ is one of symbol permutations. With $P_1, P_2, \cdots P_k$ ($k \le m!$) denoting the probability distribution for different symbols, then the PEN of order m for $\{x(i), i = 1, \cdots, N\}$ is defined as the Shannon entropy for K varied symbols [3] and computed as

$$H_p(m) = -\sum_{l}^{K} P_l \ln P_l \tag{5}$$

$H_p(m)$ reaches its maximum value $\ln(m!)$ when all symbols have the same distribution as $P_l = 1/m!$. For convenience, $H_p(m)$ is always normalized by $\ln(m!)$ and is written as $0 \le H_p = H_p / \ln(m!) \le 1$. H_p can be employed to measure the randomness degree of $\{x(i), i = 1, \cdots, N\}$, to be specific, H_p equals unity when the time series is far from randomness while zero in the case where $\{x(i), i = 1, \cdots, N\}$ belongs to the white noise.

3.3. Long Short-term Memory Network

The flue gas desulfurization process has characteristics of strongly nonlinear dynamics and large time delay, making it difficult to accurately model the FGD process. Long short-term memory (LSTM) network, which was first proposed in [11] and benefited from three control gates (i.e. input gate, forget gate and output gate) and a memory cell, can effectively prevent the vanishing/ exploding gradient problem during the training process. Due to this fact, LSTM has powerful ability not only for learning nonlinear process, but also for capturing potential long-range dependencies, and consequently a desirable candidate for modeling the flue gas desulphurization process. Provided x_t represents the input vector at the moment t, c_t denotes a cell state, h_t and h_{t-1} are, respectively, the output of the LSTM unit at time t and t-1. The schematic is shown in Figure 2, and detailed calculation formulae are shown in Equations (6)-(11) progressively.

1 Calculate candidate cell state $\tilde{c}_t$ using both weight W_c and bias b_c.

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{6}$$

2 The input gate i_t serves to control the cell state update by taking the unit output at time t-1 and input at time t to a sigmoidal function σ, with weighting matrix W_i and bias b_i introduced, the above process is formulated as

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{7}$$

3 The forget gate f_t is employed to quantify the content kept in the previous cell state, the corresponding weighting matrix and bias are respectively respresented as W_f and b_f.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{8}$$

4 The cell state at the current time c_t is updated using $\tilde{c}_t$ and cell state at the previous moment c_{t-1}, which is controlled by both input and forget gate.

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \tag{9}$$

5 The output gate o_t decides what information can be used as the output based on the cell state, W_o is the connection weighting matrix and denotes the bias.

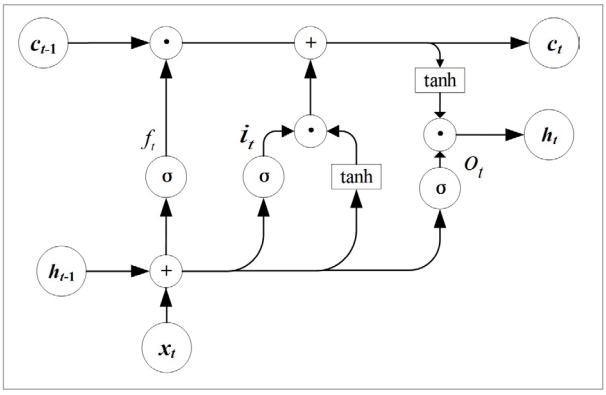
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{10}$$

6 Finally, the output of LSTM unit h_t is calculated as

$$h_t = o_t * \tanh(c_t) \tag{11}$$

where $\tanh(\cdot)$ is the hyperbolic tangent function. In this research study, LSTM serves as one of base learners in the SEL-based framework that is discussed at length in the following section.

Figure 2
Schematic of an LSTM unit



3.4. Elman Neural Network

Unlike the traditional feedforward network, feedback configuration is introduced in a RNN which makes it particularly suited for the sequence learning task. In [18], it was proved that the RNN has the capability to

make an approximation to any dynamical system with an arbitrary accuracy, and this was confirmed through many practical problems. The Elman neural network (ENN) is a local RNN which comprises the following units: a context unit, an input unit, a hidden unit and an output unit, where last three types of layers are exactly the same as those in a feedforward network; while the context layer receives outputs of the hidden layer in previous time steps, which can be treated as a one-step delay operator and thus ENN is endowed with the dynamic modeling capability. The output of the ith hidden unit $H_i(t)$ is formulated in Equation (12) as:

$$H_i(t) = g\left(\sum_{k=1}^{K} W_{ik} C_k(t-1) + \sum_{j=1}^{J} V_{ij} I_j(t-1)\right) \tag{12}$$

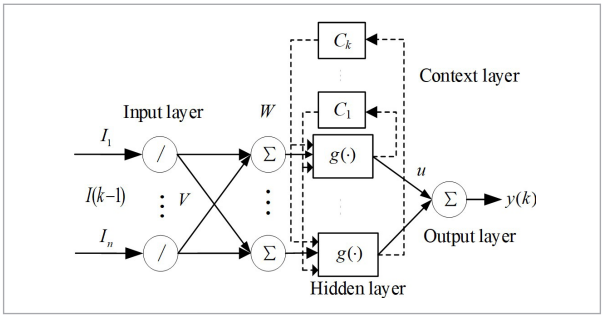
with $I_j(t)$ and $C_k(t)$ being the output of jth input unit and kth context unit, respectively. V_{ik} and W_{ik} represent corresponding weights with respect to the ith context unit, and $g(\cdot)$ is a sigmoidal activation function. For simplicity we shall consider only one output, then we have

$$y(t) = f\left(\sum_{i=1}^{K} H_i(t) z_i\right), \tag{13}$$

where z_i is the ith output weight, $f(\cdot)$ represents the output transfer function, which is typically chosen as a linear function. Figure 3 depicts the schematic architecture of an ENN.

The ARIMA model, proposed in [16], underlies nearly all statistical time series forecasting methods, and it shows superior short-term forecasting capability as compared with complex structural models. As for an ARIMA model, the output at any future time is realized by combining its past values and white noise terms in a linear fashion, besides differencing of a certain order is utilized such that the stationarity of a time series is achieved. For notational convenience, the structure of an ARIMA model is represented as ARIMA (p,d,q) through the rest of this research, where p represents auto-regressive order, d stands for differencing order and q for moving-average order.

Figure 3
A typical ENN architecture



3.5. Component Model Comparison

In our approach, the forecasting of SO_2 emission is performed jointly by an LSTM, an ENN and an ARIMA model in a stacking ensemble learning framework. As model diversity is a crucial factor to ensure the forecasting performance of an ensemble learning model, here characteristics of each component model are compared and discussed, which is beneficial to better understand strengths of the proposed modeling approach that will be elaborate upon later.

- As a linear dynamic model, ARIMA model is relatively easy to interpret and explain, compared to non-linear approximators. On the other hand, its simple structure lends itself to parameter tuning. However, owing to its linear structure, it may perform poorly when modeling non-linear processes.
- Because of the existence of context layer, structurally speaking, an ENN can be regarded both as a feedforward network and as a recurrent network. In this sense, ENN has the capability for memorizing both static and dynamic information of a process. However, as mentioned earlier, the recurrent network is subject to vanishing (or exploding) gradient issue which limits its ability to remember information that appears long time period earlier.
- LSTM, as a variant of RNN with gating mechanism and memory cells, is particularly suited to handle forecasting problems that requires the use of long range contextual information. However, LSTMs are computationally expensive and memory requirements for training are high.

The above discussion indicates that chosen component models are structurally different which can ensure the diversity of ensemble model when combining them together. As each of component model has its unique merits and weaknesses, it is hoped that the proposed ensemble model could not only benefit from their advantages, but also alleviate their drawbacks, and our end goal is to establish a well-performing forecasting model for SO_2 emissions in an FGD process.

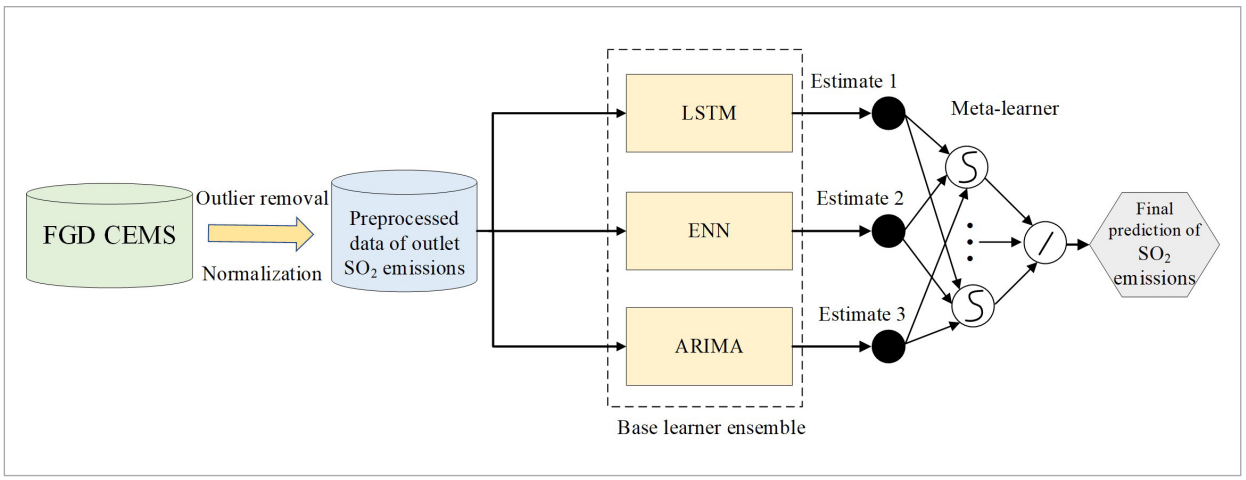
3.6. Proposed Modeling Approach

As indicated above, modeling SO_2 emission in an FGD process is by no means a easy task. This section presents details of the CEEMDAN-PEN-SEL modeling approach proposed in this study. Prior to introducing the proposed approach, the stacking ensemble strategy is first described. Over the years, ensemble learning approaches have raised great concern and found wide applications [27]. Ensemble learning approaches attempt to address a particular task with the use of multiple base models, to be specific, base learners are established separately and their results are combined in a certain manner to generate the final result. It is shown in [12] that the use of ensemble learning can obtain improved generalization performance as compared to using an individual base learner alone. From the point of view of base learners, ensemble learning is classified into two broad categories, homogeneous ensemble or heterogeneous ensemble. For the homogeneous ensemble learning, models are developed with the use of identical induction algorithm. Otherwise, the ensemble learning is said to be heterogeneous. The SO_2 emission for the desulfurized flue gas heavily depends on dynamic characteristics of the FGD process, in this sense, the forecasting model should fully catch real dynamics of the process. It is known that past observations are primary source to reflect dynamics of a process, this requires massive amounts of data which is not feasible in practice. In most cases, only a small amount of observations are available and the selected forecasting model may not be appropriate. On the other hand, using a single model is typically inclined to incorrect assumptions, and biases in parameter estimation often arises, which would negatively affect the satisfactory forecasting results. The stacking ensemble learning method is chosen as the combining method in our study, stacking makes use of a two-layer learning architecture, the first layer consists of heterogenous base learners and predictions are made by each of them; while the meta-learner in the second layer takes all predictions as inputs and learn how to combine them together optimally. Figure 4 depicts the block diagram of SEL framework employed in our study, where the continuous emission monitoring system (CEMS) is mounted around the exit of an FGD absorber so as to acquire SO_2 emissions at different instants; LSTM, ENN and ARIMA are constituent base learners; A feedforward single-layer network serves as the meta-learner that takes outputs of all base learners as its inputs. Put formally, for the time series $Y=[y_1,y_2,\cdots,y_N]^T$, corresponding estimates for the ith base learner are denoted as $\hat{Y}^{(i)}=[\hat{y}_1^{(i)},y_2^{(i)},\cdots,y_N^{(i)}]^T\ (i=1,2,\cdots,n)$. Let f be a function that serves to combine n predictions, it can be any operator such as the average or a model driven by a learning algorithm, then the output of stacking ensemble model is expressed as

$$\hat{y}_k=f(\hat{y}_k^{(1)},\cdots,\hat{y}_k^{(n)}) \qquad \forall k=1,2,\cdots,N \tag{14}$$

Figure 4

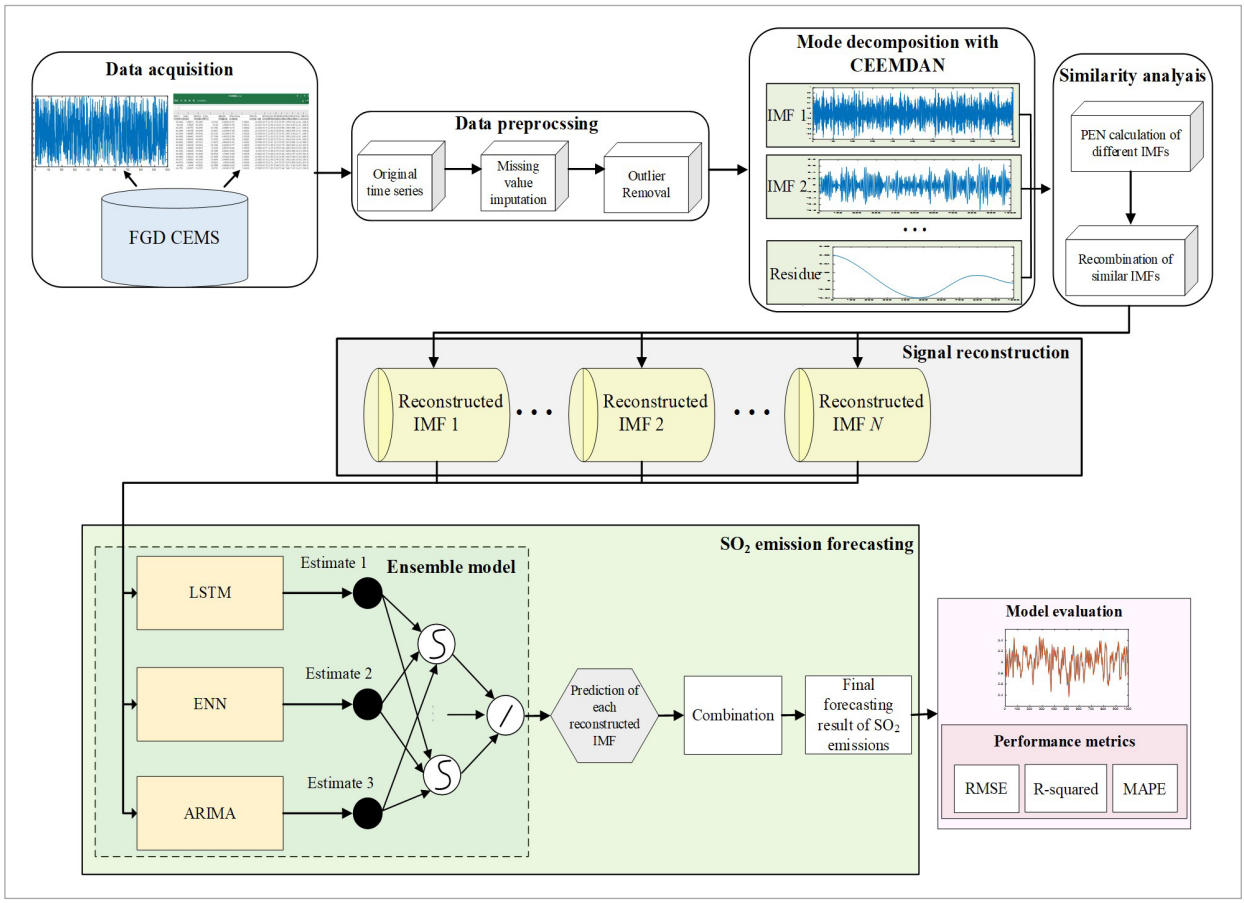
Design of the stacking ensemble learning framework for forecasting SO_2 emissions



On the basis of concepts described above, the hybrid model proposed in this study can be developed. A pictorial description of our method is presented in Figure 5.

As the industrial FGD is a process with highly nonlinear dynamics and complexity, we attack the SO_2 emission forecasting problem from both the signal and model perspective. To be specific, the technique is first employed to decrease the complexity of SO_2 emission time series. which aims at decomposing original time series into some subseries with greater simplicity and stationarity. However, a forecasting model should be established for each IMF, which is computationally expensive. In our method, PEN is employed as a tool to evaluate the complexity of each decomposed IMFs, and it is shown in [22] that integrate IMFs with similar complexity can effectively decrease the computational effort, speed up modeling and avoid overfitting problems. Quantitatively, the computational complexity for our method is determined by three part: 1) signal decomposition with CEEMDAN, 2) calculation of signal complexity by PEN and 3) training of learners in the ensemble. As for CEEMDAN, it is shown in [23] that the computational complexity of empirical mode decomposition is $O(n\log n)$, where n denotes the total number of sample instants in the signal to be decomposed. CEEMDAN is the ensemble version of empirical mode decomposition (EMD), which can effectively overcome the problem of mode mixing problem that frequently appears in EMD, and its computational complexity can be directly derived based on that of EMD by multiplying NE, where NE is the specified ensemble number. The computational complexity of PEN is expressed as $O(N-(m-1)\tau)+O(N*m\log m)+O(nk)$ [7], where design parameters m and represent respec-

Figure 5
A flowchart of the proposed CEEMDAN-PEN-SEL approach



tively the embedding dimension and time delay, n is the dimension number and is taken as unity in our study, k corresponds to the number of possible permutations in a delay vector and is equal to , N denotes the series length. There are a total of three types of learners in the ensemble, they are autoregressive integrated moving average (ARIMA) model, recurrent neural network and Long short-term memory (LSTM) network. According to [24], the computational complexity of ARIMA modeling is $O(n^2T)$, where n is the number of parameters (i.e. the summation of auto-regressive order, differencing order and moving-average order), and T is the length of the time series of interest. In [17], it is shown that the computational complexity of LSTM and recurrent neural network is identical and is expressed as $O(Td_h^2 + Td_h d_i)$, where T is the sequence length, d_h and d_i are dimensions of the hidden state and the input, respectively. The principle for the reduction of computational cost with our method is that direct use of ensemble learning for each decomposed component is quite time-consuming and tedious, where three learners should be trained separately and recombined together. However, when PEN is introduced to optimize the decomposition component, the number of components to be modelled is greatly reduced, along with the fact that the computational complexity of PEN is considerably less than that of training three types of learners for each IMF, and hence the required computational effort is significantly decreased with the proposed method. From the point of modeling, the SEL model (see Figure 4) is anticipated to achieve superior forecasting performance over the single model. With each reconstructed IMF and residue modeled by an SEL model, the predicted SO_2 emissions is calculated as

$$\hat{y}_k = \widehat{\mathrm{IMF}}_1 + \widehat{\mathrm{IMF}}_2 + \cdots + \widehat{\mathrm{IMF}}_m + \hat{r}\,, \tag{15}$$

where at time instant k, $\widehat{\mathrm{IMF}}_i$ represents the ith predicted IMF, and there are a total of m reconstructed IMFs. $\hat{r}$ is the predicted residue. In our method, the signal decomposition technique serves to decrease the complexity of SO_2 emission series, AI-based models are employed to learn and approximate FGD dynamics in a data-driven fashion, and stacking ensemble learning plays the role of enhancing model's generalization performance and adaptivity in multiple process operating conditions. As for applicability of our approach, it can be readily extended to other industrial processes, which can be explained from two aspects: first of all, the proposed modeling approach is purely data-driven, easy-to-implement and easy-to-understand, hence the corresponding model can be established straightforwardly by following the procedure introduced in our research with available measurements of the studied process Second, our method focuses on the complex and representative flue gas desulfurization process, whose characteristics (e.g. highly nonlinear dynamic, multiple working conditions, etc.) are shared or partly shared by most industrial processes, and hence same modeling difficulties would be encountered in those scenarios and can be handled effectively with our method. In this sense, the proposed method is sufficiently general and effective, making it applicable to different industrial processes, and desirable forecasting performance is expected to achieve just with minor modifications to associated hyper-parameters in the model. The pseudocode of the proposed CEEMDAN-PEN-SEL method is presented below.

4. Experiments and Analysis

In this part, the prediction performance of proposed modeling approach is evaluated, and comparative studies are conducted to demonstrate its superiority over models in existing works.

4.1. Data Description

To verify the model performance for SO_2 emission prediction, a case study is performed on the basis of historical measurements from a real FGD system in a 600MW coal-fired power plant in Hebei Province, China. The photograph of absorber under study is shown in Figure 6 below. A total of 12962 SO_2 emission samples from 05/09/2022-05/18/2022 with a sampling period of one minute are extracted from the continuous emission monitoring system, Table 1 presents the statistical analysis result of the dataset. For notational brevity, in Table 1, Num denotes the total number of samples in each type of dataset; Min and Max respectively represent the minimum and maximum in the set; Ave and SD indicate the average and standard deviation of the data, respectively. Out of all experimental data, 70 percent were used for training, 15% for validation and 15% for test. Among

Figure 6
An FGD tower in the FGD process under study

Table 1
Statistical information of experimental data

Dataset	Num	Min	Max	Ave	SD
Training set	9073	2.357	9.961	6.714	6.83
Validation set	1944	2.795	8.494	6.148	6.05
Testing set	1945	2.415	9.225	6.347	6.37

three types of datasets, the training dataset is used for parameter estimation for a given model structure; the validation set is used to provide the criterion for early stopping during the model training process, meanwhile its performance can provide guidance for selecting optimal hyper-parameter combinations; The test set serves to evaluate the final model prediction performance.

4.2. Experimental Settings

This part gives a detailed description of our experimental settings, which includes hardware/software environments, performance metrics, model training and structure determination. All algorithms are programmed by Matlab 2023a and experiments are performed on a PC server with an Intel® Core™ i7-10750 at 2.60 GHz and 32GB RAM. To evaluate the prediction performance of the presented model, mean squared error (RMSE), R squared (R^2) and mean absolute percent error (MAPE) are applied. They are written as

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e(i)^2} \tag{16}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{e(i)}{y(i)}\right| \tag{17}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(y(i)-\hat{y}(i))^2}{\sum_{i=1}^{n}(y(i)-\overline{y}(i))} \quad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y(i), \tag{18}$$

where $e(i) = y(i) - \hat{y}(i)$, $y(i)$ and $\hat{y}(i)$ are respectively actual and forecast values at instant i; n is the sample size. Among three metrics, R^2 is a measure to indicate the goodness-of-fit of a prediction model, it takes a value between 0 and 1, and a greater R^2 (closer to 1) suggests a good fit to the data. RMSE, as a basic forecasting performance indicator, can reflect the degree of deviation between the predicted value of the model and the real value, and a lower RMSE score implies a higher data approximation ability of the model. MAPE is the average of absolute percentage error, it is one of the most extensively used accuracy measure of forecast accuracy which has benefits of scale-independency and interpretability. As with RMSE, the forecasting performance of a model increases with the decreasing MAPE value. In our approach, LSTM, ENN and ARIMA are constituent models, and maximum likelihood estimation method is employed for parameter estimation of an ARIMA model. As for neural models, the Adam optimizer, which adapts the learning gain for each parameter on the basis of gradients is employed; besides the backpropagation through time approach should be employed when training an ENN. Mini-batch training, with a batch size of 64, was used, and mean squared error is chosen as the loss function to be optimized. The learning gain is set equal to 0.002. One of problems that typically arise during neural network training is called overfitting, which implies the network offers outstanding performance on the training set but performs poorly on the test set. Early stopping is one of the commonly used regularization

techniques to address the over-fitting problem and enhance model's generalization performance, and it has been successfully employed in many applications, such as natural language processing, emotion recognition, and so on. With the early-stopping technique, a portion of experimental data is assigned as the validation set, and the validation error is monitored during the model training. During the initial stage of training, the validation and training error is dropped simultaneously; but when overfitting begins, the training error is constantly decrescent while the validation error start to rise in subsequent epochs. Once continuous increase in validation error is monitored, the training terminates and model is determined as the one with the smallest validation error. In our experimental study, 15% of experimental data is chosen to constitute the validation set, and the training process of each learner is terminated provided the error on the validation set does not drop for six consecutive epochs. In addition, initial weights are randomized in the range [-0.1, 0.1] and the training data were normalized between -1 and 1 so as to bring faster convergence.

The hyperparameters play a crucial role in the model performance, to obtain the optimal architecture of a model, the grid search technique is adopted here. The grid search enables an exhaustive evaluation of all possible hyperparameter combinations within a prescribed hyper-space. With preliminary experiments, rough boundaries for different hyperparameters can be obtained. As for LSTM network, hyper-parameters and the corresponding range are listed as follows: the size of sliding window $s_w = \{1,2,3,5\}$, LSTM layer units $L_u = \{8,16,32,64\}$ and stacked layer number $l_n = \{1,2,3,4\}$. For ENN, number of hidden layers $n_d = \{1,2,3\}$, hidden units number $h_u = \{4,8,16,32\}$. In the ARIMA case, autoregressive polynomial degree $p = \{2,3,4,5\}$, differencing order $d = \{0,1,2\}$, and moving average polynomial degree $q = \{2,3,4\}$. For each parameter combination, we run 10 trials and in each trial different weights are assigned, then calculate the MAPE score on the validation set for each run. Then scores are averaged to give the final evaluation outcome, and the model structure with the best result is chosen. By this approach, the chosen hyper-parameter combinations of three base learners are shown in Table 2 below.

Table 2
Optimal structure for each base learner

Model type	Hyper-parameters
ARIMA	p=4, d=1, q=3
LSTM	s_w=3, $L_u = \{8,32,64\}$, $l_n = 3$
ENN	$n_d = 3$, $h_u = \{16,32,8\}$

Note: The notation $\{n_1,n_2,...,n_m\}$ denotes the network has hidden layers containing n_1, n_2, ..., n_m units, respectively.

4.3. SO_2 Emission Series Decomposition and Optimization

Due to properties of non-stationarity and complexity, the original SO_2 emission series cannot be satisfactorily modeled by a single predicting model. Consequently, CEEMDAN method is applied to make the decomposition of the SO_2 emission series such that multiple and a residual part are generated. The standard deviation of Gaussian white noise, ensemble size and maximum sifting iteration number are set equal to 0.2, 100 and infinity. Figure 7 presents the SO_2 emission series and its decomposed parts, 8 IMFs and a residue from top to bottom, and it is easily seen that the fluctuation degree of IMFs gradually decreases. As mentioned before, the redundancy problem may arise if decomposed modes are directly used for model development. Consequently, our study makes use of the PEN index to measure each IMF's complexity, then decrease the redundancy of decomposed modes by aggregating modes with approximate complexity. With embedded dimension m and time delay τ respectively setting equal to 3 and 2, the PEN score for each mode is calculated and summarized in Table 3 below. Based upon PEN values of each mode, it is observed that scores of the first three modes far exceed others', and thus possess relatively high complexity. By comparison, scores of last three modes are all below 0.40, suggesting trend information of the original SO_2 emission series is reflected by them. Scores of remaining three modes IMF4-IMF6 are close to each other, and hence can be integrated into one mode. At this point, three optimized portions (modes) are established: high-frequency portion OIMF1, low-frequency portion OIMF2 and trend portion OIMF3. Optimization results are presented in both Table 3 and Figure 8.

Figure 7
CEEMDAN result for SO2 emission series

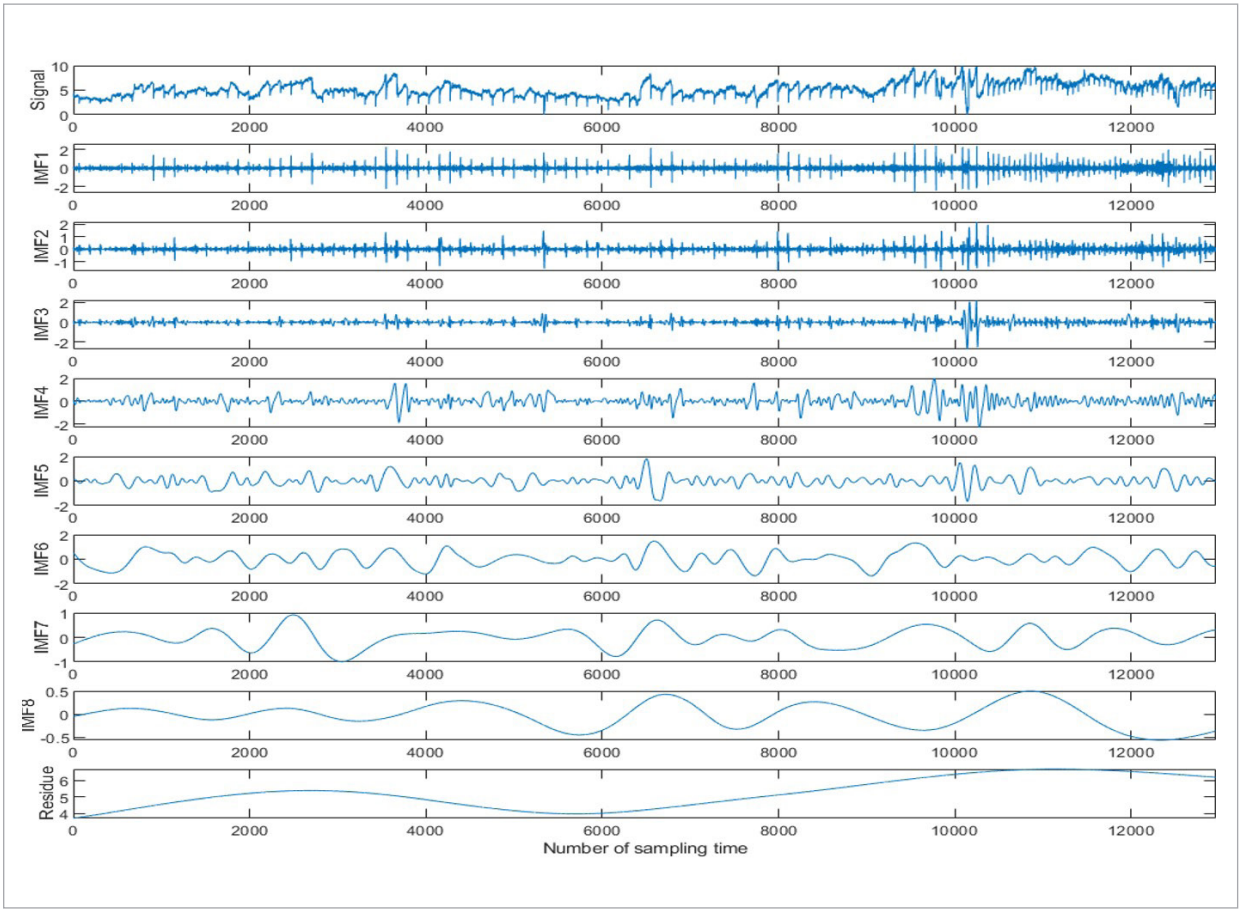
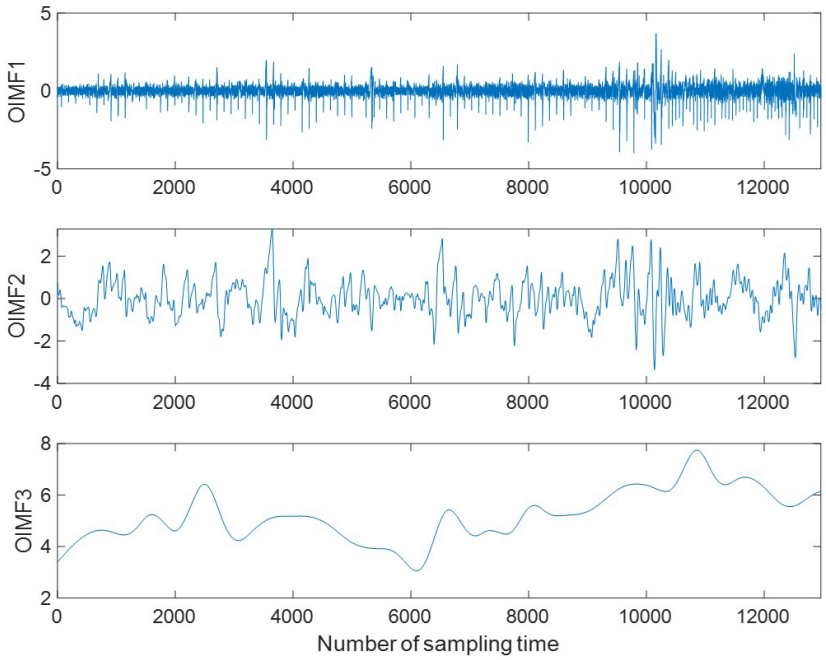


Table 3
PEN scores and optimization results of CEEMDAN modes

Result	Decomposed Mode								
	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	Residue
PEN score	0.998	0.957	0.782	0.534	0.471	0.423	0.369	0.335	0.317
Optimization	IMF1+IMF2+IMF3			IMF4+IMF5+IMF6			IMF7+IMF8+ Residue		
New mode	OIMF1			OIMF2			OIMF3		

Figure 8
Optimized modes by PEN



4.4. Forecasting Results of CEEMDAN-PEN-SEL Model

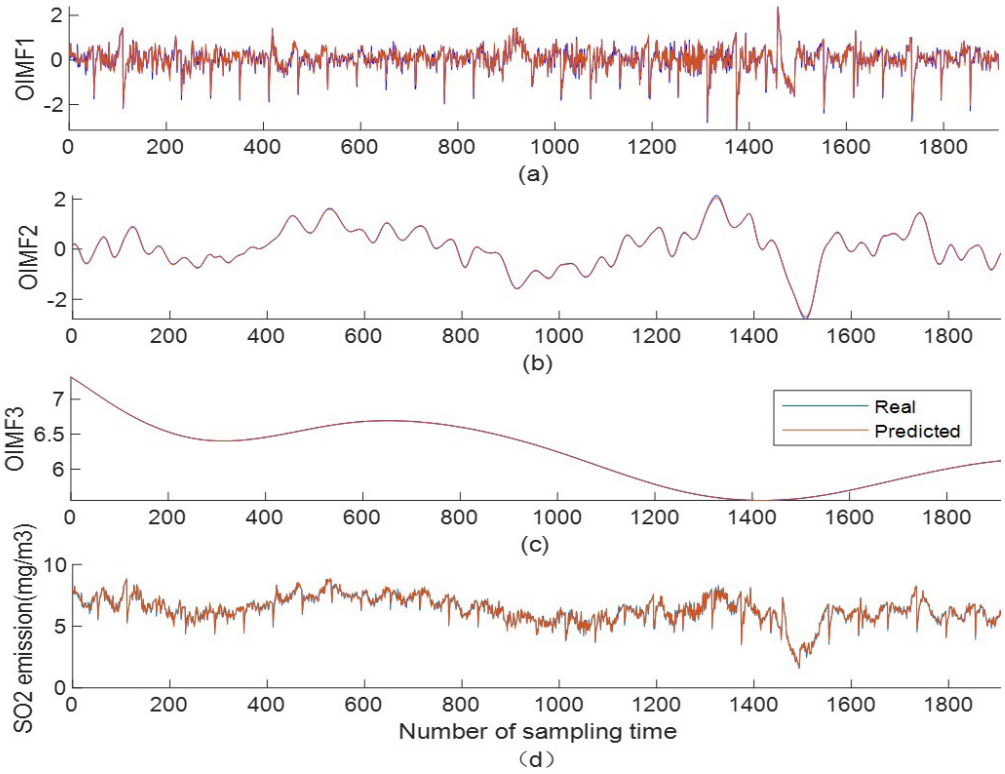
In this part, the performance of proposed CEEMDAN-PEN-SEL model is evaluated and analyzed. According to the procedure of establishing CEEMDAN-PEN-SEL model described earlier, SEL model should be developed for each optimized mode OMF1, OMF2 and OMF3. The determination of component model structure in each SEL model follows the same procedure as discussed in the Experimental settings part, where the grid search strategy is adopted on the basis model performance on the validation set, and the final model structure for each optimized mode are presented in Table 4.

Table 4
Optimal SEL model structure for each mode

Mode	ARIMA	LSTM	ENN	Meta-learner
OMF1	$p=3$, $d=0$, $q=2$	s_w=3, $L_u=\{64,32\}$, $l_n=2$	$n_d=2$, $h_u=\{16,32\}$	$h_d=6$
OMF2	$p=2$, $d=0$, $q=2$	s_w=2, $L_u=\{8,16\}$, $l_n=2$	$n_d=1$, $h_u=\{16\}$	$h_d=4$
OMF3	$p=1$, $d=1$, $q=1$	s_w=2, $L_u=\{16\}$, $l_n=1$	$n_d=1$, $h_u=\{8\}$	$h_d=3$

Figure 9(a)-(c) present prediction results of optimized modes on the test set. it can be clearly observed that our established SEL model can effectively capture the temporal dependencies of each mode. Specifically, despite high-frequency mode OIMF1 has higher complexity than the other modes, the prediction result is still in good agreement with the actual series. As for OIMF2 and OIMF3, the SEL model can provide outstanding prediction performance, and perfect fit is nearly achieved. By integrating all optimized modes together, the final forecasting result of SO_2 emissions can thus be obtained (see Figure 9 (d)). It can be observed that the forecast made by CEEMDAN-PEN-SEL model closely approximates

Figure 9
Prediction results of each optimized mode and SO_2 emissions



real values, which shows its effectiveness with respect to SO_2 emissions forecasting.

4.5. Performance Comparison of the Models

In this part, the proposed model is contrasted with component models which are also mainstream models used for SO_2 emission forecast, so as to verify its superiority. All comparative models have appeared in past research works concerning SO_2 emission prediction, e.g. ARIMA, ENN and BP network. In addition, the SEL model proposed in our study is also involved. The structure for three component models are presented in Table 2, and the structure of BP network is determined as 16-64-32 (i.e. a network contains three hidden layer with 16, 64 and 32 units, respectively) using the same hyper-parameter optimization method introduced as before. Moreover,Ew we evaluate the performance of the proposed method versus current state-of-the-art SO_2 emission prediction models, they are LSTM network with attention mechanism (AM) in [15], CNN-LSTM-AM model in [28] and CNN-BiLSTM model in [13]. The hyperparameters of comparison models are set as recommended in original studies and some of key parameters are presented below. AM-LSTM: Adam optimizer is adopted, four layers in LSTM are respectively used with 256, 256, 128, 64 units, the dropout rate and learning rate is equal to 0.001 and 0.4, respectively; CNN-LSTM-AM: two layers in LSTM are respectively employed with 128 and 64 neurons, the number of neurons in the CNN layers is set as 80 and Adam optimizer is used; CNN-BiLSTM: two layers in CNN are respectively used with 48 and 32 filters, and corresponding filter sizes are [4,4] and [2,2], the number of units in three LSTM layers is 200, 100 and 100, respectively. The learning rate is set equal to 0.1 and Adam optimizer is applied during the model training process. The SO_2 emission prediction results of all experimental models on the test set are summarized in Table 5.

As seen from Table 5, both ARIMA model and BP network performs poorly on the SO_2 emission forecasting problem, whose MAPE and RMSE scores are higher than the other models'. This is mainly due to the fact that FGD is a process with highly nonlinear dynamics, and the ARIMA model has the ability to characterize linear dynamics but cannot handle the nonlinearity efffectively; BP network, as a static model, does a good job of approximating nonlineatity while works inefficiently for capturing dynamical relationship in a series. As for ENN model, in comparison with ARIMA and BP network, the MAPE score is respectively decreased by 26.76% and 24.09%, while the R^2 is respectively increased by 31.11% and 21.78%. A serious drawback of ENN is that its gradient either vanishes nor explodes when memorizing long-term context in a time series, prohibiting it from achieving outstanding prediction performance for SO_2 emissions. Common for all three state-of-the-art SO_2 emission prediction methods is that deep learning models and techniques are incorporated, which can effectively capture long-span temporal dependencies and complex dynamics in an FGD process, and thus is seen to performs considerably better than RNN which is the best-performing traditional SO_2 emission prediction model. Specifically, CNN-LSTM-AM achieves the best prediction performance, followed by CNN-BiLSTM and LSTM-AM, whose MAPE and RMSE score respectively achieve 0.0196 and 0.1535 while R^2 reachs a high value of 0.9683. As compared to CNN-LSTM-AM, our method attains a better prediction performance and experimental results show that MAPE and RMSE are respectively decreased by 35.21% and 36.22%, while R^2 is increased to 97.12%. Above results indicate that the proposed model has the best scores with respect to all three metrics among experimental models, suggesting it is the model that suits best for SO_2 emissions forecasting. Figure 10 summarizes the error histogram of the proposed method for the test set. It is observed from Figure 10 that the total error range produced by the proposed method has been partitioned into 20 bins, and the error varies between -0.3927 (leftmost bin) and 0.4135 (rightmost bin). To be specific, a total of 341 instances are within the range (-0.0320, 0.0104), followed by 326 and 311 instances are respectively in the range of (0.0104, 0.0528) and (-0.0320, -0.0744). All values in above three intervals closely approximates zero and

Algorithm. CEEMDAN-PEN-SEL forecasting method

Input: temporal sequence $\{x(t)\}$, hyper-parameters for three types of base learners: ARIMA: autoregressive polynomial degree p, differencing order d, moving average polynomial degree q; LSTM: stacked layer number l_n, units number L_u, sliding window size s_w; ENN: hidden layer number n_d, hidden units number h_u

Output: forecasting sequence $\{\hat{x}(t)\}$

1. Preprocess the series $\{x(t)\}$ with Pauta criterion to detect and eliminate outliers, Lagrange interpolation for missing data imputation and min-max normalization for mapping values within the range of [0,1].
2. Decompose the preprocessed time series into subseries (IMF_1, IMF_2, ..., IMF_n), whose frequencies are sorted in descending order, and a residue with the CEEMADAN technique.
3. Evaluate the complexity of each decomposed mode by calculating the PEN score, and modes with similar PEN scores are aggregated to form optimized modes ($OIMF_1$, $OIMF_2$, ..., $OIMF_m$)($m<n$).
4. For each optimized mode $OIMF_i$ (i=1,2...,m), train the base learner LSTM, ARIMA and ENN separately with specified hyper-parameters. With developed base learners, complete the training of meta-learner using series $OIMF_i$. Finally, the stacking ensemble model for $OIMF_i$ is established using well-trained base learners and meta-learner.
5. Compute the forecasting sequence $\{\hat{x}(t)\}$ using the following equation

$$\{\hat{x}(t)\} = \widehat{\mathrm{OIMF}_1} + \widehat{\mathrm{OIMF}_2} + \cdots + \widehat{\mathrm{OIMF}_m} \quad (16)$$

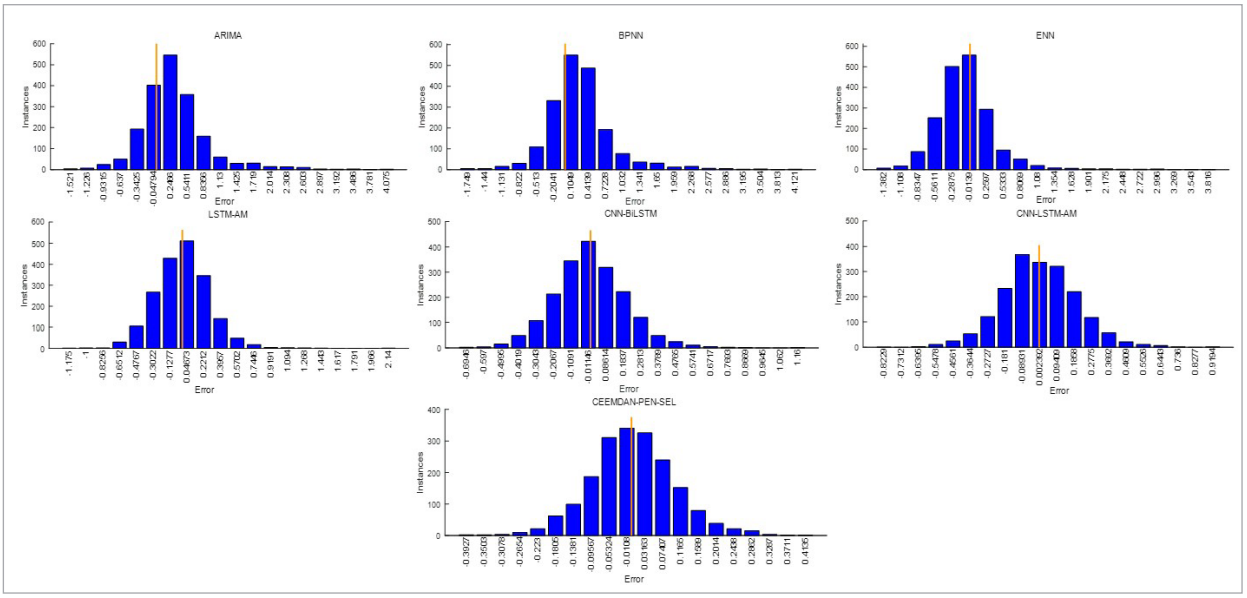
where $\widehat{\mathrm{OIMF}_i}$ (i=1,2...,m) denotes the output of i-th stacking ensemble model.

6. Evaluate the established stacking ensemble model on the test set using metrics like R squared (R^2), mean absolute percent error (MAPE) and root mean squared error (RMSE).

Table 5
Comparison of performance indicators of different forecasting models

Metrics	ARIMA	BPNN	ENN	CNN-BiLSTM	LSTM-AM	CNN-LSTM-AM	CEEMDAN-PEN-SEL
MAPE (%)	0.0908	0.0876	0.0665	0.0221	0.0358	0.0196	0.0127
RMSE	0.6938	0.6545	0.5158	0.1736	0.2711	0.1535	0.0979
$^{R}2$	0.5899	0.6351	0.7734	0.9613	0.9374	0.9683	0.9712

Figure 10
Histogram of the error with all experimental models



the total instance number therein accounts for more than a half of the total number of instances. On the whole, a significant portion of forecasting errors are distributed within a satisfactory range around zero, indicating the superiority of our approach.

5. Conclusions

SO_2 emission forecasting is of significant importance in an FGD process, which serves to offer early warning for the public as well as provide timely operating guidance for the desulfurizing system. However, the FGD is a process with highly nonlinear dynamics and high time delay, which makes the problem of SO_2 emission forecasting a formidable challenge. In this research, CEEMDAN technique is first adopted to extract modes at different scales, which are easier to be modeled. Then PEN is employed to quantify the complexity of each mode, and modes with approximate scores are integrated together to reduce the computational cost. Finally, SEL consisting of ARIMA, LSTM and ENN is developed and used to model each mode. The experimental results suggest that the proposed CEEMDAN-PEN-SEL has exceptional forecasting performance for SO_2 emissions. Using metrics like MAPE, RMSE and R squared, the performance of commonly used models and proposed model is evaluated quantitatively. According to experimental results, it is found that MAPE, RMSE and R^2 scores of our model can respectively achieve 0.0127, 0.0979 and 0.9712, all of which are best among all models. On this base, the effectiveness of the proposed hybrid model is thoroughly verified and it has a great utility for SO_2 emission forecasting in an FGD process. Although the

proposed approach achieved excellent performance in forecasting SO_2 emissions for a flue gas desulphurization process, the methods also present some potential limitations, such as difficulties in determining appropriate values of associated hyper-parameters in a rapid but effective method, incapability to handle multivariate sequences, and so on.

Acknowledgement

This study was supported by the 2022 Priority Research Project of Beijing Education Science 14th Five-Year Plan (Research on the Improvement of Graduate Education Quality in Capital Universities, project number: CDEA22009).

References

1. Abayomi-Alli, O. O., Sidekerskienė, T., Damaševičius, R., Siłka, J., Połap, D. Empirical Mode Decomposition Based Data Augmentation for Time Series Prediction Using NARX Network. In International Conference on Artificial Intelligence and Soft Computing, Springer, 2020, 702-711. https://doi.org/10.1007/978-3-030-61401-0_65
2. Bandt, C., Pompe, B. Permutation Entropy: A Natural Complexity Measure for Time Series. Physical Review Letters, 2002, 88(17), 174102-174107. https://doi.org/10.1103/PhysRevLett.88.174102
3. Cao, Y., Tung, W. W., Gao, J. B., Protopopescu, V. A., Hively, L. M. Detecting Dynamical Changes in Time Series Using the Permutation Entropy. Physical Review E, 2004, 70(4), 046217. https://doi.org/10.1103/PhysRevE.70.046217
4. Córdoba, P. Status of Flue Gas Desulphurisation (FGD) Systems from Coal-Fired Power Plants: Overview of the Physic-Chemical Control Processes of Wet Limestone FGDs. Fuel, 2015, 144, 274-286. https://doi.org/10.1016/j.fuel.2014.12.065
5. Damaševičius, R., Napoli, C., Sidekerskienė, T., Woźniak, M. IMF Mode Demixing in EMD for Jitter Analysis. Journal of Computational Science, 2017, 22, 240-52. https://doi.org/10.1016/j.jocs.2017.04.008
6. Damaševičius, R., Vasiljevas, M., Martisius, I., Jusas, V., Birvinskas, D., Woźniak, M. BoostEMD: An Extension of EMD Method and Its Application for Denoising of EMG Signals. Elektronika ir Elektrotechnika, 2015, 21(6), 57-61. https://doi.org/10.5755/j01.eee.21.6.13763
7. Deng, S., Huang, Z., Wang, X., Huang, G. Radio Frequency Fingerprint Extraction Based on Multidimension Permutation Entropy. International Journal of Antennas and Propagation, 2017, 2017, 1-6. https://doi.org/10.1155/2017/1538728
8. Dou, B., Pan, W., Jin, Q., Wang, W., Li, Y. Prediction of SO2 Removal Efficiency for Wet Flue Gas Desulfurization. Energy Conversion and Management, 2009, 50(10), 2547-2553. https://doi.org/10.1016/j.enconman.2009.06.012
9. Guo, Y., Xu, Z., Zheng, C., Shu, J., Dong, H., Zhang, Y. Modeling and Optimization of Wet Flue Gas Desulfurization System Based on a Hybrid Modeling Method. Journal of the Air & Waste Management Association, 2019, 69(5), 565-575. https://doi.org/10.1080/10962247.2018.1551252
10. Hou, P. F., Bai, J. Y., Yin, J. On-Line Monitoring and Optimization of Performance Indexes for Limestone Wet Desulfurization Technology. Applied Mechanics and Materials, 2013, 295, 1020-1028. https://doi.org/10.4028/www.scientific.net/AMM.295-298.1020
11. Hochreiter, S., Schmidhuber, J. Long Short-Term Memory. Neural Computation, 1997, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735
12. Li, QF., Song, ZM. High-Performance Concrete Strength Prediction Based on Ensemble Learning. Construction and Building Materials, 2022, 324, 126694. https://doi.org/10.1016/j.conbuildmat.2022.126694
13. Lu, Y., Li, K. Multistation Collaborative Prediction of Air Pollutants Based on the CNN-BiLSTM Model. Environmental Science and Pollution Research, 2023, 30, 92417-92435. https://doi.org/10.1007/s11356-023-28877-z
14. Pan, QX., Li, Y., Wang, N., Zhao, PF., Huang, L., Wang, ZY. Variational Mode Decomposition-Based Synchronous Multi-Frequency Electrical Impedance Tomography. Information Technology and Control, 2022, 51(3), 446-466. https://doi.org/10.5755/j01.itc.51.3.30014
15. Pang, C., Duan, D., Zhou, Z., Han, S., Yao, L., Zheng, C., Yang, J., Gao, X. An Integrated LSTM-AM and SPRT Method for Fault Early Detection of Forced-Oxidation System in Wet Flue Gas Desulfurization. Process Safety and Environmental Protection, 2022, 160, 242-254. https://doi.org/10.1016/j.psep.2022.01.062
16. Pack DJ. In Defense of ARIMA Modeling. International Journal of Forecasting, 1990, 6(2), 211-8. https://doi.org/10.1016/0169-2070(90)90006-W

17. Rotman, M., Wolf, L. Shuffling Recurrent Neural Networks. In Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 9428-9435. https://doi.org/10.1609/aaai.v35i11.17136

18. Schäfer, A. M., Zimmermann, H. G. Recurrent Neural Networks Are Universal Approximators. In Artificial Neural Networks-ICANN 2006: 16th International Conference, Athens, Greece, September 10-14, 2006, Proceedings, Part I 16, 632-640. https://doi.org/10.1007/11840817_66

19. Shi, G., Qin, C., Tao, J., Liu, C. A VMD-EWT-LSTM-Based Multi-Step Prediction Approach for Shield Tunneling Machine Cutterhead Torque. Knowledge-Based Systems, 2021, 228, 107213-107226. https://doi.org/10.1016/j.knosys.2021.107213

20. Uddin, G. M., Arafat, S. M., Ashraf, W. M., Asim, M., Bhutta, M. M. A., Jatoi, H. U. K., Niazi, S. G., Jamil, A., Farooq, M., Ghufran, M., Jawad, M., Hayat, N., Jie, W., Chaudry, I. A., Zeid, I. Artificial Intelligence-Based Emission Reduction Strategy for Limestone Forced Oxidation Flue Gas Desulfurization System. Journal of Energy Resources Technology, 2020, 142(9), 092103-092116. https://doi.org/10.1115/1.4046468

21. Villanueva Perales, A. L., Gutiérrez Ortiz, F. J., Vidal Barrero, F., Ollero, P. Using Neural Networks to Address Nonlinear pH Control in Wet Limestone Flue Gas Desulfurization Plants. Industrial & Engineering Chemistry Research, 2010, 49(5), 2263-2272. https://doi.org/10.1021/ie9007584

22. Wang, J., Sun, X., Cheng, Q., Cui, Q. An Innovative Random Forest-Based Nonlinear Ensemble Paradigm of Improved Feature Extraction and Deep Learning for Carbon Price Forecasting. Science of the Total Environment, 2021, 762, 143099-143112. https://doi.org/10.1016/j.scitotenv.2020.143099

23. Wang, Y. H., Yeh, C. H., Young, H. W., Hu, K., Lo, M. T. On the Computational Complexity of the Empirical Mode Decomposition Algorithm. Physica A: Statistical Mechanics and Its Applications, 2014, 400, 159-167. https://doi.org/10.1016/j.physa.2014.01.020

24. Wang, X., Kang, Y., Hyndman, RJ., Li, F. Distributed ARIMA Models for Ultra-Long Time Series. International Journal of Forecasting, 2023, 39, 1163-1184. https://doi.org/10.1016/j.ijforecast.2022.05.001

25. Wu, Z., Huang, N. E. Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method. Advances in Adaptive Data Analysis, 2009, 1(01), 1-41. https://doi.org/10.1142/S1793536909000047

26. Wan, Y., Hui, X., He, X., Xue, J., Feng, D., Chen, Z., Li, J., Liu, L., Xue, Q. Utilization of Flue Gas Desulfurization Gypsum to Produce Green Binder for Dredged Soil Solidification. Strength, Durability, and Planting Performance. Journal of Cleaner Production, 2022, 367, 133076-133088. https://doi.org/10.1016/j.jclepro.2022.133076

27. Yin, J., Li, N. Ensemble Learning Models with a Bayesian Optimization Algorithm for Mineral Prospectivity Mapping. Ore Geology Reviews, 2022, 145, 104916-104928. https://doi.org/10.1016/j.oregeorev.2022.104916

28. Yin, X., Sun, K., Li, S., Wang, X., Dong, Y., Cui, L. Enhancing Deep Learning for the Comprehensive Forecast Model in Flue Gas Desulfurization Systems. Control Engineering Practice, 2023, 138, 105587-105601. https://doi.org/10.1016/j.conengprac.2023.105587

29. Zhu, J., Ye, S. C., Bai, J., Wu, Z. Y., Liu, Z. H., Yang, Y. F. A Concise Algorithm for Calculating Absorption Height in Spray Tower for Wet Limestone-Gypsum Flue Gas Desulfurization. Fuel Processing Technology, 2015, 129, 15-23. https://doi.org/10.1016/j.fuproc.2014.07.002

30. Zhou, G., Zhong, W., Zhou, Y., Wang, J., Wang, T. 3D Simulation of Sintering Flue Gas Desulfurization and Denitration in a Bubbling Gas Absorbing Tower. Powder Technology, 2017, 314, 412-426. https://doi.org/10.1016/j.powtec.2016.09.051

31. Zhu, J., Zhao, P., Yang, S., Chen, L., Zhang, Q., Yan, Q. Continuous SO2 Absorption and Desorption in Regenerable Flue Gas Desulfurization with Ethylenediamine-Phosphoric Acid Solution: A Rate-Based Dynamic Modeling. Fuel, 2021, 292, 120263-120276. https://doi.org/10.1016/j.fuel.2021.120263