frontiers | Frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 25 April 2022
doi: 10.3389/fevo.2022.855606

Prediction of Air Pollution Interval Based on Data Preprocessing and Multi-Objective Dragonfly Optimization Algorithm

*Jiyang Wang, Jingrui Li and Zhiwu Li**

Institute of Systems Engineering, Macao University of Science and Technology, Macao, China

With the rapid development of global industrialization and urbanization, as well as the continuous expansion of the population, large amounts of industrial exhaust gases and automobile exhaust are released. To better sound an early warning of air pollution, researchers have proposed many pollution prediction methods. However, the traditional point prediction methods cannot effectively analyze the volatility and uncertainty of pollution. To fill this gap, we propose a combined prediction system based on fuzzy granulation, multi-objective dragonfly optimization algorithm and probability interval, which can effectively analyze the volatility and uncertainty of pollution. Experimental results show that the combined prediction system can not only effectively predict the changing trend of pollution data and analyze local characteristics but also provide strong technical support for the early warning of air pollution.

Keywords: atmospheric contamination prediction, temporal convolution network, fuzzy information granulation, multi-objective dragonfly optimization algorithm, interval prediction

OPEN ACCESS

Edited by:
Wendong Yang,
Shandong University of Finance and Economics, China

Reviewed by:
He Jiang,
Jiangxi University of Finance and Economics, China
Feng Liu,
University of Technology Sydney, Australia

***Correspondence:**
Zhiwu Li
zwli@must.edu.mo

Specialty section:
This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Ecology and Evolution

Received: *15 January 2022*
Accepted: *07 March 2022*
Published: *25 April 2022*

Citation:
Wang J, Li J and Li Z (2022) Prediction of Air Pollution Interval Based on Data Preprocessing and Multi-Objective Dragonfly Optimization Algorithm. Front. Ecol. Evol. 10:855606. doi: 10.3389/fevo.2022.855606

INTRODUCTION

With the continuous development of the economy and the rising living standards, the deterioration of the environment, land desertification, greenhouse effect, and other problems have begun to plague us. In addition, the United States and other developed countries have classified indoor air pollution into the five environmental factors that endanger human health. The Health Effects Institute from the US released "State of Global Air (2020)," which indicates that, at least 6.7 million people worldwide, will die from chronic exposure to air pollution in 2019 (State of Global Air, 2020).

Up to now, many studies have been conducted to study the problem of air pollution. Recently, to accurately measure the quality of air, particulate matter (PM) has become a significant and common index to be monitored (Beaulant et al., 2008). $PM_{2.5}$ is one type of PM, which means that the particulate matter in the ambient air has an aerodynamic equivalent diameter less than or equal to 2.5 μm (van Donkelaar et al., 2006). It can be suspended in the air for a long period, and the higher the concentration of its content in the air, the more serious the air contamination. Compared with coarser atmospheric particles, $PM_{2.5}$ has the following features: small particle size, large area, strong activity, easy adhesion, and long residence time in the atmosphere; thus, it has a greater impact on human health and the quality of the atmospheric environment (Sun and Li, 2020).

As a result, $PM_{2.5}$ has become a worldwide problem to be solved, and many institutions have established various methods to accurately monitor $PM_{2.5}$ concentrations (McKeen et al., 2007; Borrego et al., 2011; Air Quality Expert Group, 2012; Bergen et al., 2013; Wakamatsu et al., 2013).

Bai et al. (2019) proposed the DL-SSAE method as an autoencoder model to consider the advantages of seasonal analysis and deep feature learning to predict the hourly $PM_{2.5}$ concentrations. Irina et al. developed the Community Multiscale Air Quality method with five different data preprocessing strategies to analyze the concentrations of $PM_{2.5}$, and they found that the Kalman filter correction could compute the most precise results (Djalalova et al., 2015). In addition, Samia et al. (2012) combined autoregressive integrated moving average (ARIMA) and Ann to enhance predicting performance, and the results show that the proposed hybrid system could be used to efficiently forecast and provide useful air quality information. In addition, multiple linear regression have been utilized to forecast $PM_{2.5}$ or PM_{10} concentrations in the air to make decisions related to traffic restrictions in the future or support the control of air quality (Akyüz and Çabuk, 2009; Genc et al., 2010). Moreover, Banik et al. (2020) employed long short-term memory (LSTM) to analyze wind speed in various seasons, and they concluded that LSTM performs better than Elman and non-linear auto-regressive models. Osowski and Garanty (2007) used support vector machine (SVM) to decompose the original data and to predict the air quality of Poland based on wavelet representation. Another common method is the gray model (GM), which was employed by Pai et al. After comparing with other models for predicting the performance of $PM_{2.5}$ and PM_{10} concentrations of Taipei, they demonstrated that GM (1, 1) could be a useful early warning system for nearby citizens (Pai et al., 2013).

Additionally, the temporal convolutional network (TCN) is widely used to achieve more accurate performance. For instance, Zhu et al. (2020) solved the problem of long-term dependencies and performance degradation of a deep convolutional model by TCN, which shows that the power system with TCN performs better and more stably compared with others. Li et al. (2018) predicted oil consumption with various parameters according to TCN and found that the proposed model could obtain more satisfying results and help make decisions for the energy market. Wei designs a convolutional spiking neural network to deal with temporal datasets, which corrects and optimizes the historical performance, and more accurately forecasts wind speed. Also, this method could quantify the differences in predicting the performances that resulted from uncertainties (Wei et al., 2021). Chen et al. (2020) established a structure with the convolutional neural network (CNN) to forecast associated sequences and to handle more complex seasonal problems, which helps make useful decisions to assess power generation by providing more evidence. Yang W. et al. (2020) combined empirical mode decomposition (EMD) and **TCN** to forecast the remaining useful life and reduce the cost during the operation. Tian and Wang (2021) applied the temporal convolution networks with the quantile regression (TCNQR) method to judge the period of health and operation. In this study, we used TCN as one of the forecasting tools to obtain the results of air quality.

The recently developed approaches mainly belong to point forecasting, which includes some disadvantages and limitations (Wang J. et al., 2021). For example, Wang et al. (2022b) have pointed out that the point predicting approaches produce an unavoidable error during the operation, which might result in immense risks for an electric power system since it only depends on the accurate results. In addition, a considerable amount of time and high cost will be wasted if precise information cannot be provided, which is also a loss to the entire power system.

Unlike point prediction, which gives a "specific numerical prediction," the interval forecast aims for a future period and gives an interval in which the predicted values are likely to occur, with a prediction interval corresponding to the expected probability. The interval forecast gives more prediction information than the point forecast, which means that we can get the value of the point forecast within a certain interval based on a certain probability, thus more scientifically characterizing the uncertainty of the model forecast.

As for data preprocessing, information granulation (IG) is a technique for studying the formation and representation of information grains and for information pre-processing. Fuzzy information granulation (FIG) is one type of IG first proposed by Zadeh (1997) to discuss how to deal with fuzzy datasets. FIG has been employed to acquire original data of fluctuating traffic and construct a traffic flow, predicting the approach with interval forecasting (Guo et al., 2018). Zhang and Na (2018) applied FIG to transform the historical agricultural price into FIG particles, and the forecasting results show that the proposed price predicting system model performs more efficiently with better accuracy. FIG could also be used in the power system. For example, the authors utilize FIG to remove the variability of the historical series of wind and solar energy, and the experimental results demonstrate that the developed approach performs efficiently and could help decision-makers stabilize the energy system (He et al., 2019b). Additionally, to forecast the actual streamflow data, FIG is combined with support vector regression (SVR) to provide more precise computation and eliminate the fluctuation of the streamflow, which means that the proposed model has a more accurate prediction interval of the hydrologic system (He et al., 2019a).

According to the existing research about $PM_{2.5}$ concentrations and forecasting, we found that the majority of the models are combined models. Compared with the traditional single model, the combined models avoid the error of individual approaches and yield more accurate results. Therefore, more researchers have adopted combined models for prediction. For example, Wang S. et al. (2021) applied a novel wind power combined predicting system to obtain more precise performance, which supports further research in wind generation. Wang et al. (2022a) used four foundation models and optimized the weight coefficient using a multi-objective water cycle algorithm (MOWCA) to predict hourly $PM_{2.5}$ concentrations. Details of a single model and combined models are summarized in **Table 1**. In **Table 2** for detailed nomenclature in the article.

Based on the analyses above, this study employs a novel combined predicting system to monitor $PM_{2.5}$ and PM_{10}. It integrates FIG, TCN, ARIMA, and LSTM to forecast $PM_{2.5}$ and PM_{10} concentrations, then uses a weight generation structure to compute each coefficient, and finally combines the single approaches to achieve a better result of the experiments.

The primary contributions and innovations of this study are shown as follows.

TABLE 1 | Summary of predicting approach types.

Category	References	Advantages	Disadvantages
Physical model	Xiao et al., 2019	Physical model concentrates on long-term predicting data series where it could perform better	Physical models have difficulties in predicting data series in short term, and they need to collect sufficient data as an initial dataset
Statistical approach			
Auto-regressive moving average (ARma)	Wang et al., 2012; Wang and Hu, 2015	Statistical approaches achieve satisfying performance in dealing with linear data	The assumption of statistical approach is tough to realize, and they cannot operate well in non-linear patterns
Auto-regressive integrated moving average (ARIMA)			
Grey method (GM)	Ding, 2019		
Artificial intelligence			
Artificial neural network (Ann)	Wang et al., 2016		
Support vector machine (SVM)	Zhang et al., 2019		
Fuzzy logic (FL)	Sfetsos, 2000		
Back propagation neural network (BPnn)	Bin et al., 2014	They are suitable to handle non-linear data series and could obtain better results and stability.	Artificial intelligence approaches cannot perform well due to the over-fitting problem, and single approaches have some limitations.
General regression neural network (GRnn)	Majumder and Maity, 2018		
Long short-term memory (LSTM)	Banik et al., 2020		
Particle swarm optimization (PSO)	Liu W. et al., 2019		
Combined model			
WPD-PSO-BPnn	Liu H. et al., 2019		
FT-CS	Yang H. et al., 2020	Combined models consider the strengths and weaknesses of single approaches to achieve more precise results	The running period of the combined models is longer than other single models, and the weight coefficient needs to be considered
VMD-BEGA-LSTM	Mencar and Fanelli, 2008		
FIG-SVR	He et al., 2019a		

(1) *To avoid the limitations of point predicting methods, this study proposes a useful interval predicting approach.* This technique could deal with the fluctuation associated with the $PM_{2.5}$ and PM_{10} concentrations by quantifying the information of the original dataset. The performance of interval predicting is shown to be more effective than that of other point forecasting approaches.
(2) *According to the decomposition and reconstruction techniques, this study applies the data pretreatment method to eliminate the negative influence of the initial data series.* As a superior data preprocessing strategy, FIG is used to decrease high-frequency noise and to reconstruct the novel data sequences to acquire the significant elements of the historical data and facilitate the smooth implementation of the next phase.
(3) *A combined model is developed in the predicting section to obtain the results of $PM_{2.5}$ concentrations.* It obtains more accurate prediction results when compared with the traditional $PM_{2.5}$ and PM_{10} concentrations of prediction approaches.
(4) *The developed model could be employed in air quality monitoring.* The proposed system and the predicting results are clearly improved by providing more useful information on air quality to people and analyzing and predicting $PM_{2.5}$ and PM_{10} concentrations even in more complicated conditions.

The rest of this article is organized as follows. Section "Forecasting System Development" touches upon the design of the forecasting system, including data fuzzy information granulation and the proposed combined forecasting system. Section "Framework of the Proposed Forecasting System and Parameter" describes the framework and parameters of the proposed prediction system. To further verify the accuracy and effectiveness of the proposed combined model from various aspects, detailed experimental results and analysis are presented in section "Experimental Results and Discussion". Finally, section "Conclusion" concludes this research.

FORECASTING SYSTEM DEVELOPMENT

This section develops an innovative combined predicting system to predict the $PM_{2.5}$ and PM_{10} concentrations in the air, which enhances the performance of the results by a data denoising strategy and a predicting approach.

TABLE 2 | Nomenclature.

Nomenclature			
Abbreviate			
Anns	Artificial neural networks	**WPD**	Wavelet packet decomposition
BPnn	Back propagation neural Network	**FT**	Fuzzy theory
ARIMA	Auto-regressive integrated Moving Average	**CS**	Cuckoo search
ARma	Auto-regressive moving Average	**BEGA**	Binary encoding genetic optimization Algorithm
GM	Gray method	**VMD**	Variational mode decomposition
GRnn	General regression neural Network	**MSE**	Mean squared error
LSTM	Long short-term memory Network	**MAPE**	Average absolute percent error
PSO	Particle swarm optimization	**MAE**	Mean absolute error of *N* predicting results
SVM	Support vector machine	**RMSE**	Root of mean error squares
TCN	Temporal convolution network	**SSE**	Sum of squared error
PICP	Prediction interval coverage probability	**PBW**	Prediction band width
FL	Fuzzy logic	**MOWCA**	Multi-objective Water Cycle Algorithm
SVR	Support vector regression	**FIG**	Fuzzy information granulation
IG	Information granulation	**TCNQR**	Temporal convolution network with the quantile regression
EMD	Empirical mode decomposition	**MLR**	Multiple linear regression
PM	Particulate matter	**MODOA**	Multi-objective dragonfly optimization algorithm
CNN	Convolutional neural network		

Decomposition and Denoising Strategies

Fuzzy information granules (FIG) construct information granules by building fuzzy sets for each subsequence formed by discretizing the time series (Mencar and Fanelli, 2008). The core of fuzzy information granulation is to complete the fuzzification process after the window is created, which mainly includes window division and information fuzzification.

The window division is to convert the time series $\overline{R} = \left\{\overline{R}_1, \overline{R}_2, \ldots, \overline{R}_\gamma\right\}$ into the granular time series $\bar{\bar{\Theta}} = \left\{\bar{\bar{\Theta}}_1, \bar{\bar{\Theta}}_2, \ldots, \bar{\bar{\Theta}}_\varsigma\right\}$ after information granulation by setting the time granularity $\widehat{\mathbf{E}}$ to divide $\overline{R} = \left\{\overline{R}_1, \overline{R}_2, \ldots, \overline{R}_\gamma\right\}$ into $\underline{\mathbf{H}}$ subseries $\bar{\bar{\Theta}} = \left\{\bar{\bar{\Theta}}_1, \bar{\bar{\Theta}}_2, \ldots, \bar{\bar{\Theta}}_\varsigma\right\}$, where $\underline{\mathbf{H}} = \frac{\gamma}{\widehat{\mathbf{E}}}$ and the ηth subseries is $\bar{\bar{\Theta}}_\eta = \left[\overline{\overline{R}}_1^{(\eta)}, \overline{\overline{R}}_2^{(\eta)}, \ldots, \overline{\overline{R}}_{\widehat{\mathbf{E}}}^{(\eta)}\right]$.

$$\left\{\overline{R}_1, \overline{R}_2, \ldots, \overline{R}_\gamma\right\} \Rightarrow \left\{\left[\overline{\overline{R}}_1^{(1)}, \overline{\overline{R}}_2^{(1)}, \ldots, \overline{\overline{R}}_{\widehat{\mathbf{E}}}^{(1)}\right], \ldots, \left[\overline{\overline{R}}_1^{(\underline{\mathbf{H}})}, \overline{\overline{R}}_2^{(\underline{\mathbf{H}})}, \ldots, \overline{\overline{R}}_{\widehat{\mathbf{E}}}^{(\underline{\mathbf{H}})}\right]\right\} \quad (2-1-1)$$

The information granulation of the time series $\overline{R} = \left\{\overline{R}_1, \overline{R}_2, \ldots, \overline{R}_\gamma\right\}$ is to construct the information particles $\tilde{\Gamma} = \left\{\tilde{\Gamma}_1', \tilde{\Gamma}_2', \ldots, \tilde{\Gamma}_\varsigma'\right\}$ using the fuzzy method for each of the $\underline{\mathbf{H}}$ subsequences $\bar{\bar{\Theta}} = \left\{\bar{\bar{\Theta}}_1, \bar{\bar{\Theta}}_2, \ldots, \bar{\bar{\Theta}}_\varsigma\right\}$ formed by the discretization operation.

Suppose that $\mathbf{Z}$ is a given theoretical domain, then a fuzzy subset $\Lambda = \{\chi, \Omega(\chi) \,|\, \chi \in \mathbf{Z}\}$ on $\mathbf{Z}$, $\Omega(\chi) : \chi \rightarrow [0, 1]$ represents the affiliation function of Λ. Two fuzzy subsets, Φ and Ξ, are equal, denoted by $\Phi = \Xi$, if they have the same affiliation function, i.e., $\widehat{\Omega}'_{\Phi}(\chi) = \widehat{\Omega}''_{\Xi}(\chi)$.

In this study, the triangular fuzzy particles are chosen to construct the information grain and its affiliation function is as follows:

$$\mathbf{A}_{Tf}(\mathbf{x}) = \begin{cases} \frac{\mathbf{x}-\mathbf{I}_{Tf}}{\mathbf{K}_{Tf}-\mathbf{I}_{Tf}}, \mathbf{I}_{Tf} \leq \mathbf{x} \leq \mathbf{K}_{Tf} \\ 0, \mathbf{x} < \mathbf{I}_{Tf} \cup \mathbf{x} > \mathbf{N}_{Tf} \\ \frac{\mathbf{N}_{Tf}-\mathbf{x}}{\mathbf{N}_{Tf}-\mathbf{K}_{Tf}}, \mathbf{K}_{Tf} < \mathbf{x} \leq \mathbf{N}_{Tf} \end{cases} \quad (2-1-2),$$

where x is the variable in the theoretical domain and $\mathbf{I}_{Tf}$, K_{Tf}, and $\mathbf{N}_{Tf}$ are the three parameters of the triangular type fuzzy example affiliation function, which correspond to the lower boundary, average level, and upper boundary of the window after fuzzy particleization, respectively.

Fuzzy sets get rid of the either-or duality in classical set theory and extend the value domain of the affiliation function from the binary {0, 1} to the multi-valued interval [0, 1], which is a kind of extension of the set theory. Information fuzzification is the fuzzification of each information grain, and the fuzzification of a single sub-window, $\bar{\bar{\Theta}}_\mu$, generates multiple fuzzy sets $\tilde{\Gamma}_\mu' = \left[\tilde{\Gamma}_{\mu;1}'', \tilde{\Gamma}_{\mu;2}'', \tilde{\Gamma}_{\mu;3}''\right]$.

Considering the single-window problem, $\bar{\bar{\Theta}}_\mu = \left[\bar{\bar{\mathbf{T}}}_1^{(\mu)}, \bar{\bar{\mathbf{T}}}_2^{(\mu)}, \ldots, \bar{\bar{\mathbf{T}}}_{\widehat{\mathbf{E}}}^{(\mu)}\right]$ should first be viewed as a window for fuzzification. The task of fuzzification is to build a triangular fuzzy particle TFP on $\bar{\bar{\Theta}}_\mu = \left[\bar{\bar{\mathbf{T}}}_1^{(\mu)}, \bar{\bar{\mathbf{T}}}_2^{(\mu)}, \ldots, \bar{\bar{\mathbf{T}}}_{\widehat{\mathbf{E}}}^{(\mu)}\right]$, which can reasonably explain the fuzzy concept M of $\bar{\bar{\Theta}}_\mu$. The fuzzy particle $\tilde{\Gamma}_\mu' = \left[\tilde{\Gamma}_{\mu;1}'' = \widehat{\mathbf{I}}_{Tf}^{\mu}, \tilde{\Gamma}_{\mu;2}'' = \widehat{\mathrm{K}}_{Tf}^{\mu}, \tilde{\Gamma}_{\mu;3}'' = \widehat{\mathbf{N}}_{Tf}^{\mu}\right]$ can be constructed by the relevant parameters in the determined affiliation function (2-1-2) of the triangular fuzzy particle.

Predicting Algorithm

In this section, the basic theory and equations of some forecasting approaches are described.

Auto-Regressive Integrated Moving Average Model

The AR($\boldsymbol{p}$) model means the auto-regressive approach of the $\boldsymbol{p}^{th}$ order, expressed as (Hamilton James, 2015):

$$y_t = \boldsymbol{c} + \boldsymbol{a}_1 y_{t-1} + \cdots + \boldsymbol{a_p} y_{t-p} + u_t \quad (2-2-1),$$

where $\boldsymbol{a_1}, \cdots, \boldsymbol{a_p}$ are indicators; $\boldsymbol{c}$ is a constant; and u_t is referred to as the random variable.

Besides, the MA($\boldsymbol{q}$) model represents the moving average model of the $\boldsymbol{p}^{th}$ order, which is defined as:

$$y_t = \boldsymbol{a} + u_t + \boldsymbol{m}_1 u_{t-1} + \cdots + \boldsymbol{m_q} u_{t-q} \quad (2-2-2),$$

where $\boldsymbol{m}_1, \cdots, \boldsymbol{m}_q$are the factors of the approach; y_t is always set as 0, and the expectation of y_t can be written as $\boldsymbol{a}$. Also, u_t, u_{t-1} and u_{t-q} could describe the white noise error terms of the initial series.

Then, the **ARMA**($\boldsymbol{p}$, $\boldsymbol{q}$) combines the two approaches listed above, which is shown in the following formula:

$$y_t = c + \boldsymbol{a_1} y_{t-1} + \cdots + \boldsymbol{a_p} y_{t-p} + u_t + \boldsymbol{m_1} u_{t-1} + \cdots + \boldsymbol{m_q} u_{t-q} \quad (2-2-3)$$

If these three approaches are employed in dealing with samples with non-stationarity evidence, we could consider taking various steps to decrease this limitation, which is regarded as **ARIMA** ($\boldsymbol{p,d,q}$), where d is the degree of differencing.

Deep Learning Using Long Short-Term Memory Recurrent Neural Networks

Recurrent neural network (RNN) is one type of ANN, and the combinations of various samples become a directed cycle. LSTM is proposed to handle long series. Both LSTM and RNNs could employ some gates to fix the gradient problem. Some scientists have proved that RNNs are included in the hidden layer, which is one of the features of LSTM (Gers et al., 2000). The three layers of RNNs with LSTM demonstrated in **Figure 1** present the memory cell functions.

$$f_t = \boldsymbol{\sigma}(W_f \bullet [C_{t-1}, h_{t-1}, x_t] + b_f) \quad (2-2-4)$$

$$i_t = \boldsymbol{\sigma}(W_i \bullet [C_{t-1}, h_{t-1}, x_t] + b_i) \quad (2-2-5)$$

$$o_t = \boldsymbol{\sigma}(W_o \bullet [C_t, h_{t-1}, x_t] + b_o) \quad (2-2-6)$$

$$C_t = f_t \times C_{t-1} + i_t \times \widetilde{C}_t \quad (2-2-7)$$

$$\widetilde{C}_t = \boldsymbol{tanh}(W_c \bullet [h_{t-1}, x_t] + b_c) \quad (2-2-8)$$

$$h_t = o_t \times \boldsymbol{tanh}(C_t) \quad (2-2-9),$$

where x_t is the input value; h_t is the output vector; C_t represents the cell state variable; $\boldsymbol{W}$ and $\boldsymbol{b}$ are indicator matrices and indicator; f_t, i_t, and o_t are forget, input, and output gate variables, respectively. In the equation, σ means the sigmoid formula and ***tanh*** refers to a rescale logistic sigmoid function belonging to (−1, 1).

Temporal Convolutional Network

A TCN, an interval predicting method, is a special kind of CNN (Shelhamer et al., 2017). It includes three sections: causal convolution, dilated convolution, and residual network. The first section makes sure of the result at time β, and we assume the input value $\overleftrightarrow{\boldsymbol{\mu}} \in \underrightarrow{\widetilde{\boldsymbol{\Theta}}}^{\mathrm{B}}$and a filter $\pi : \{0, \ldots, k-1\} \to \underrightarrow{\widetilde{\boldsymbol{\Theta}}}$. The historical convolutional layer is stated by two equations: $\underleftrightarrow{\widehat{\boldsymbol{\Lambda}}}(\overleftrightarrow{\boldsymbol{\mu}}_{\beta}) = (\overleftrightarrow{\boldsymbol{\mu}} \odot \pi)(\beta) = \sum_{j=0}^{k-1} \pi_j \overleftrightarrow{\boldsymbol{\mu}}_{\beta-j}$and $\widetilde{ouse} = (\underleftrightarrow{\widehat{\boldsymbol{\Lambda}}}(\overleftrightarrow{\boldsymbol{\mu}}_1), \underleftrightarrow{\widehat{\boldsymbol{\Lambda}}}(\overleftrightarrow{\boldsymbol{\mu}}_2), \cdots, \underleftrightarrow{\widehat{\boldsymbol{\Lambda}}}(\overleftrightarrow{\boldsymbol{\mu}}_{\mathrm{B}}))$, where k is the size of convolutional kernel, $\widetilde{ouse}$ is the output series, and $\underleftrightarrow{\widehat{\boldsymbol{\Lambda}}}(\bullet)$ is the process of convolution.

The second part uses a hyperparameter to jump some input values; thus, a range longer than it used to be could be accepted by the filter. In detail, if the causal convolution mixes, the m^{th} layer dilated convolution can be described by: $\underleftrightarrow{\widehat{\boldsymbol{\Lambda}}}(\overleftrightarrow{\boldsymbol{\mu}}_{\beta}) = (\overleftrightarrow{\boldsymbol{\mu}} \odot d_m \pi)(\beta) = \sum_{j=0}^{k-1} \pi_j \overleftrightarrow{\boldsymbol{\mu}}_{\beta - d_m j}$ and $\widetilde{ouse} = (\underleftrightarrow{\widehat{\boldsymbol{\Lambda}}}(\overleftrightarrow{\boldsymbol{\mu}}_1), \underleftrightarrow{\widehat{\boldsymbol{\Lambda}}}(\overleftrightarrow{\boldsymbol{\mu}}_2), \ldots, \underleftrightarrow{\widehat{\boldsymbol{\Lambda}}}(\overleftrightarrow{\boldsymbol{\mu}}_{\mathrm{B}}))$, where d_m means the dilation indicator of the m^{th} layer and the range could be set to $2^{m-1}2^{m-1}$. Here, $\beta - d_m j$ represents the historical direction. The second formula is a temporal convolutional layer, which constructs TCN in many layers.

If the layers are deep, to deal with the issue of decreased efficiency of the CNN results, a residual block is utilized. During the training procedure, we added a residual connection into the block to ensure normal operation in the deep layers. Moreover, TCN prevents the over-fitting problem by introducing the dropout layer after each dilated convolution (Srivastava et al., 2014).

Combined Model

Combining forecasts has long been recognized as an effective and a simple way to improve forecast stability, an improvement over a single model. This study proposes a new combined forecasting model that fuses ARIMA, neural networks, and the non-positive constraint theory.

The traditional forecasting combination method attempts to find the best weight of the combined models based on minimizing SSE:

$$\min \mathbf{F}=\mathbf{D}^T\mathbf{E}\mathbf{D}=\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} d_i d_j e_{it} e_{jt} \begin{cases} \boldsymbol{R^T D} = 1 \\ \boldsymbol{D} \geq 0 \end{cases} (2-3-1),$$

where $\mathbf{D} = \left(d_1, d_2, \ldots, d_m\right)^T$ is the weight vector; $\mathbf{R} = (1, 1, \ldots, 1)^T$ is a column vector where all elements are 1; and $\mathbf{E} = \left(\mathbf{E} =_{ij}\right)_{m \times m}$is called the error information matrix ($\mathbf{E}_{ij} = e_i^T e_j$, $e_i = (e_{i1}, e_{i2}, \ldots, e_{iN})$).

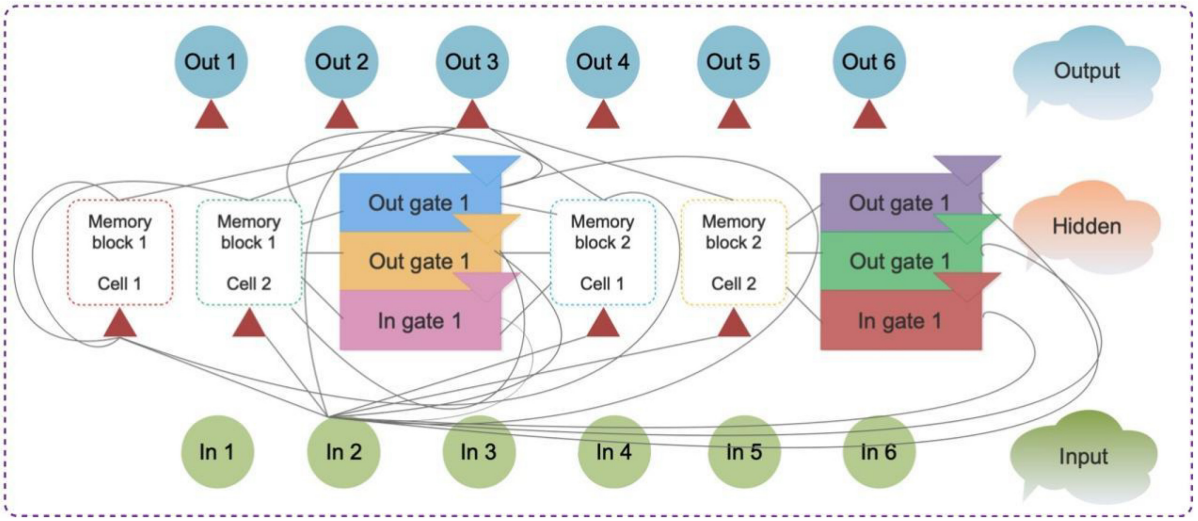

FIGURE 1 | Three-layer long short-term memory (LSTM) of two LSTM memory blocks (Gers et al., 2000).

An improvement of the traditional combination method based on the non-negative constraint theory (TCM-NNCT) and non-positive constraint theory is given as follows:

$$\min \mathbf{J}=\mathbf{D}^T\mathbf{E}\mathbf{D}=\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} d_i d_j e_{it} e_{jt} \qquad (2-3-2)$$

$$\text{st } \mathbf{R}^T\mathbf{D}=1 \qquad (2-3-3)$$

In Eq. (2-3-1), the weight vector has no limitation in the range [0,1]. The experiment results show that the combination model can obtain desirable results if the weight vector has a value in the range of [−2,2]. This section provides a weight-determined method that will be assessed by experimental simulation rather than a theoretical proof.

Interval Prediction Based on Temporal Convolutional Network

In this section, we constructed a novel multidimensional time series CNN prediction model for air contamination forecasting and uncertainty analysis.

Interval Prediction Module

There is no need to presuppose an error distribution in an interval forecasting model that, based on a linear model, is expressed as:

$$\begin{aligned}\mathbf{Y} &= \tilde{\beta}_0(\theta)+\tilde{\beta}_1(\theta)\mathbf{X_1}+\cdots+\tilde{\beta}_n(\theta)\mathbf{X_n}+u\\ &= \tilde{Q}(\theta;\mathbf{X})+\mathbf{u} \qquad (2-4-1),\end{aligned}$$

where $\boldsymbol{X}$ is n explanatory variables, $\tilde{\beta}_i$ is a vector that can be determined based on $\boldsymbol{X}$ and θ, and $\tilde{\beta}_i$ is a vector that can be identified based on $\boldsymbol{X}$ and θ. Given K samples, the vector is statistically estimated: $\tilde{\beta}_i(\theta)=\arg\min_{\beta}\sum_{i=1}^{K} f(\theta)\left[\mathbf{Y}_i-\tilde{Q}(\theta;\mathbf{X})\right](i=1,2,\ldots,K)$, where $f(\theta)$ is a piecewise linear loss function that can be defined as:

$$f(\theta)=\begin{cases}\theta u, & u\geq 0\\ (\theta-1)u, & u<0\end{cases} \qquad (2-4-2)$$

Evaluation Index of Prediction Model and Interval Forecasting

Due to the advanced non-linear characteristics and uncertainty of air contamination data, the prediction error of a single model is usually possible. For this case, the calculation of the prediction interval, i.e., the higher and lower bounds for predicting future values, is appropriate for the prediction of air contamination data. Below the given prediction interval, the predictions are often created higher and therefore the stability of prediction can be additionally improved. Different indicators can judge the prediction results and quality of the interval prediction model, such as the following common evaluation indicators.

Prediction interval coverage probability (PICP) is the most important index to measure the quality of the prediction interval, which reflects the probability result of the observed value falling into the prediction interval, namely reliability. In other words, the greater the probability value, the more the observations covered by the prediction interval and vice versa. In general, within the established prediction interval, the calculated probability p (PICP) should be higher than the rated confidence level, namely: $p=P\left(f_i\in\left[L\left(\overline{\overline{\mathbf{X}_i}}\right),U\left(\overline{\overline{\mathbf{X}_i}}\right)\right]\right)\geq\boldsymbol{\mu}$, $i=1,2,\ldots,K$, where $P(\bullet)$ is the expressed probability; $L\left(\overline{\overline{\mathbf{X}_i}}\right)$, $U\left(\overline{\overline{\mathbf{X}_i}}\right)$ represent the lower and upper bounds of the prediction interval predicted by $\overline{\overline{\mathbf{X}_i}}$, respectively; f_i is the predicted value; and $\boldsymbol{\mu}$ is a given confidence level. According to Bernoulli' law of huge numbers, $\tilde{\chi}_{cp}$ will be expressed by the frequency

that the prediction interval covers the determined value, and its likelihood converges to P, namely:

$$\tilde{\chi}_{cp} = \frac{1}{D}\sum_{i=1}^{D} c_i \times 100\% \xrightarrow{P} P\left(f_i \in \left[L\left(\overline{\overline{X_i}}\right), U\left(\overline{\overline{X_i}}\right)\right]\right) \quad (2-4-3)$$

where D is the predicted sample size and C_i is a Boolean quantity.

$$C_i = \begin{cases} 1, & \textbf{if } f_i \in \left[L\left(\overline{\overline{X_i}}\right), U\left(\overline{\overline{X_i}}\right)\right] \\ 0, & \textbf{if } f_i \notin \left[L\left(\overline{\overline{X_i}}\right), U\left(\overline{\overline{X_i}}\right)\right] \end{cases} \quad (2-4-4)$$

If $\tilde{\chi}_{cp} \geq \mu$, it indicates that the established prediction interval is valid; otherwise, it indicates that the established prediction interval is invalid, and it should be reestablished.

To decide the prediction interval more reasonably, it is necessary to depend on the prediction interval mean width percentage (**PIMWP**), which is the parameter basis for evaluating the prediction interval. If the forecast interval is wide enough, the coverage of the forecast interval will be on the brink of 100%. However, such a good interval cannot effectively provide the uncertainty data of the predicted value, rendering the results of the forecast interval meaningless.

If $\tilde{\chi}_{cp}$ is larger and $\bar{\tilde{\chi}}_{MWP}$ is smaller, the prediction interval of the model is more accurate and the performance is better.

$$\bar{\tilde{\chi}}_{MWP} = \frac{1}{D}\sum_{i=1}^{D} \frac{U\left(\overline{\overline{X_i}}\right) - L\left(\overline{\overline{X_i}}\right)}{f_i} \times 100\% \quad (2-4-5)$$

In addition, single high reliability and high clarity cannot reflect the performance of the interval prediction model, which is one of the biggest differences with the evaluation index of the deterministic prediction model. $\tilde{\chi}_{cp}$The performance evaluation indexes, $\tilde{\chi}_{cp}$ and $\bar{\tilde{\chi}}_{MWP}$, are often used to predict interval models. However, if some special situations occur, these two indicators cannot achieve a reasonable and scientific performance evaluation of the interval prediction model. For example, if the observed value is not within the prediction interval and if there is a small difference between $\tilde{\chi}_{cp}$ and $\bar{\tilde{\chi}}_{MWP}$ at the same time, it is impossible to measure the degree of deviation of the observed value from the prediction interval. The extent to which observations deviate from the predicted interval is immeasurable. To compensate for the shortcomings of $\tilde{\chi}_{cp}$ and $\bar{\tilde{\chi}}_{MWP}$, this study introduces another evaluation index of the prediction interval model, namely accumulated width deviation (AWD), which can clearly measure the deviation degree of observed values outside the prediction interval. Here, $\overline{\overline{\chi}}_{AWD} = \sum_{i=1}^{D} \zeta_i$, where ζ_i represents the degree to which the observed value deviates from the upper and lower bounds of the predicted interval.

$$\zeta_i = \begin{cases} \frac{L\left(\overline{\overline{X_i}}\right) - f_i}{f_i}, & \textbf{if } f_i < L\left(\overline{\overline{X_i}}\right) \\ 0, & \textbf{if } f_i \in \left[L\left(\overline{\overline{X_i}}\right), U\left(\overline{\overline{X_i}}\right)\right] \\ \frac{f_i - U\left(\overline{\overline{X_i}}\right)}{f_i}, & \textbf{if } f_i > U\left(\overline{\overline{X_i}}\right) \end{cases} \quad (2-4-6)$$

Under the condition of the same $\tilde{\chi}_{cp}$ and $\bar{\tilde{\chi}}_{MWP}$, the smaller the value of $\overline{\overline{\chi}}_{AWD}$, the higher the quality of the prediction interval.

The above three evaluation indicators, $\tilde{\chi}_{cp}$, $\bar{\tilde{\chi}}_{MWP}$, and $\overline{\overline{\chi}}_{AWD}$, are independent of each other, and only a certain feature of the prediction interval is considered. However, if only one evaluation index is selected, it is not enough to explain the quality and performance of the prediction interval. A high-quality prediction interval should conform to the confidence level requirements, i.e., $\tilde{\chi}_{cp}$ should be as high as possible while $\bar{\tilde{\chi}}_{MWP}$ and $\overline{\overline{\chi}}_{AWD}$ should be as low as possible. However, the definitions of $\tilde{\chi}_{cp}$, $\bar{\tilde{\chi}}_{MWP}$, and χ_{AWD} show that these three metrics are conflicting with each other: the higher the $\tilde{\chi}_{cp}$, the higher the $\bar{\tilde{\chi}}_{MWP}$; the lower $\bar{\tilde{\chi}}_{MWP}$, the lower $\tilde{\chi}_{cp}$ and the higher $\overline{\overline{\chi}}_{AWD}$; the lower the $\overline{\overline{\chi}}_{AWD}$ is, the higher the $\bar{\tilde{\chi}}_{MWP}$ is. Therefore, taking these three indicators into consideration, this study proposes a comprehensive index that can quantitatively evaluate the prediction interval, namely, prediction interval satisfaction index (PISI). It can be calculated by:

$$\chi_{PISI} = \left[1 - \left(1 + \lambda \times \overline{\overline{\chi}}_{AWD}\right) \bar{\tilde{\chi}}_{MWP} \times \left(1 + e^{-\eta\left(\chi_{cp} - \mu\right)}\right)\right] \times 100\% \quad (2-4-7),$$

where λ is the penalty factor of $\overline{\overline{\chi}}_{AWD}$, η is the penalty factor of $\tilde{\chi}_{cp}$, and μ (95%) is the given confidence level. In this study, we choose $\lambda = 0.5$ and $\eta = 50$.

If $\tilde{\chi}_{cp}$ is greater than the given confidence level μ, the curve of χ_{PISI} is flat and the value of χ_{PISI} tends to 1. At this point, χ_{PISI} is mainly determined by $\bar{\tilde{\chi}}_{MWP}$ and $\overline{\overline{\chi}}_{AWD}$. If $\tilde{\chi}_{cp}$ is less than the given confidence level μ, the value of χ_{PISI} changes according to the difference between $\tilde{\chi}_{cp}$ and μ, and χ_{PISI} is mainly determined by $\tilde{\chi}_{cp}$ at this time. Therefore, χ_{PISI} can further reflect the quality of the prediction interval by combining $\tilde{\chi}_{cp}$, $\bar{\tilde{\chi}}_{MWP}$, and $\overline{\overline{\chi}}_{AWD}$, making the evaluation of the prediction interval more effective and accurate.

FRAMEWORK OF THE PROPOSED FORECASTING SYSTEM AND PARAMETER

This section presents the description of the material analyzed (section "Dataset Description") and the entire probabilistic forecasting system applied in this study (section "Flow of the Proposed Ensemble Probabilistic Forecasting System").

Dataset Description

This study took the $PM_{2.5}$ and PM_{10} pollution data of Beijing, Shanghai, and Shenzhen as the experimental data set, which are daily data from January 2020 to December 2021. From each dataset, we extracted 4,386 point values as experimental sequences and selected 80% of the total length as training sets. The remaining 20% points were divided into test sets as shown in **Figure 2**.

FIGURE 2 | Information of the research areas.

Flow of the Proposed Ensemble Probabilistic Forecasting System

In accordance with the aforementioned data processing approaches and forecasting models, the proposed forecasting system includes Fuzzy information granulation, ARIMA, LSTM, TCN, multi-objective optimization, and interval prediction.

Step 1: The original three data sets were divided into a training set and a test set. A total of 4,386 pieces of data were collected. There were 3,500 pieces of data in training sets and 877 pieces of data in test sets.

Step 2: The pollution values of $PM_{2.5}$ and PM_{10} are reconstructed by graining Fuzzy information granulation and the data after noise reduction has been obtained.

Step 3: ARIMA, **LSTM,** and **TCN** were used for forecasting, and they were used as the comparative models of the multi-target dragonfly combination prediction results in the fourth step.

Step 4: The prediction results of **ARIMA**, **LSTM,** and **TCN** were combined with a multi-objective Dragonfly algorithm for optimization.

Step 5: Probabilistic forecasting module: The upper and lower bounds and the prediction interval were obtained by using interval prediction to forecast the progress of $PM_{2.5}$ and PM_{10} data.

By constructing the prediction interval, the probability prediction of air pollution is carried out. To determine the distribution of forecast errors resulting from point forecasts, three metrics were used: the **PICP**, the **BW**, and the **PINAW**. Furthermore, interval forecasts were created by combining upper and lower bounds with an optimal distribution with a design confidence level of 95%.

Model Selection and Parameter Setting

In general, a hybrid forecasting system adopts a decomposition strategy using a shallow neural network; all of the ARIMA and LSTM have satisfactory performance in solving regression problems. In DL, the TCN based on multidimensional time series is sensitive to the prediction of statistical data. Therefore, we selected the multi-objective dragonfly optimization algorithm based on the multidimensional time series for interval prediction. The model naming and argument details of the other models are presented in **Table 3**.

Evaluation Index

In this study, five evaluation indexes [such as the mean absolute percentage error (MAPE) and root mean square error

TABLE 3 | Compare the parameter settings of each model.

Model	Symbol	Meaning	Value	Reason
LSTM	n_i	Number of input layer nodes	5	Number of feature inputs
	n_h	Number of hidden layer nodes	[100, 100]	Trial-and-error manner
	n_0	Number of output layer nodes	1	Number of feature inputs
	e_t	Epochs of training	5000	Trial-and-error manner
ARIMA	p	Auto-Regressive term	[p-site1, p-site2, p-site3] = [5,5,3]	**AIC** and **BIC**
	d	Integrated term	1	**ADF** test
	q	Moving Average term	[q-site1, q-site2, q-site3] = [3,6,3]	**AIC** and **BIC**
TNC	$\overleftrightarrow{\mu}$	input value	DataSet	Preset
	$\widetilde{\widetilde{ouse}}$	output series	\	Preset
	$\overleftrightarrow{\hat{\Lambda}}(\bullet)$	The process of convolution	\	Preset
	d_m	dilation indicator	\	Preset

(RMSE)] were used to assess the prediction system stability and accuracy, and other indicators were used to evaluate the interval prediction capability. **Table 4** presents the specific equations and definitions.

EXPERIMENTAL RESULTS AND DISCUSSION

This section discusses in detail the fuzzy granulation strategy based on multi-dimensional time series, the multi-objective dragonfly optimization algorithm, and the simulation results of interval prediction. To further improve the prediction results, the prediction efficiency (FE) and improvement rate (IR) of the proposed combined prediction model and interval prediction, as well as sensitivity, are analyzed in the study.

Data Pre-processing: Fuzzy Information Granulation

Through fuzzy information granulation, the pollution data of $PM_{2}.5$ and PM_{10} are processed.

Specific steps are as follows:

(1). To confirm sample extraction and fuzzification processing, sample information needs to be extracted to a certain extent. Then, the specific size of the window can be understood through the extracted data. Later, fuzzy information granulation processing is carried out according to the formula (2-1-1).

(2). The minimum, average, and maximum values are normalized after granulation treatment. The processing formula is:

$$\boldsymbol{p}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}},$$

where $\boldsymbol{p}_i$ is a variable data in the sample data; x_i is the normalized data coefficient; $x_{\min}$ is the minimum value of the extracted data; and $x_{\max}$is the maximum value in the sample.

In the subsequent combined prediction model, we use the granulated average $\boldsymbol{R}$ as the input for training and testing. The comparison result of fuzzy granulation with the original data is shown in **Figure 3**.

TABLE 4 | Evaluation metrics applied in this study.

Metric	Equation
RMSE	$\boldsymbol{RMSE} = \sqrt{\left[\sum_{i=1}^{N} \boldsymbol{PP_i} - \boldsymbol{AP_i}\right] / \boldsymbol{N}}$
MAPE	$\boldsymbol{MAPE} = \frac{1}{N}\sum_{i=1}^{N} \left\|(\boldsymbol{AP_i} - \boldsymbol{PP_i}) / \boldsymbol{AP}\right\| \times 100\%$
MSE	$\boldsymbol{MSE} = \frac{1}{n}\sum_{i=1}^{n} (\boldsymbol{y_i} - \widehat{\boldsymbol{y}}_{\boldsymbol{i}})^2$
Adjusted *R* square	$\boldsymbol{R} = 1 - \frac{\sum_i (\widehat{y}_i - y_i)^2}{\sum_i (\overline{y}_i - y_i)^2}$
SSE	$\boldsymbol{SSE} = \sum_{i=1}^{n} (\boldsymbol{y_i} - \widehat{\boldsymbol{y}}_{\boldsymbol{i}})^2$
PICP	$\tilde{\chi}_{cp} = \frac{1}{D}\sum_{i=1}^{D} \boldsymbol{c_i} \times 100\%$
BW	$\overline{\overline{X}}_{BW} = \sum_{i=1}^{D} \xi_i$
PINAW	$\overline{\overline{X}}_{PINAW} = \left[1 - (1 + \lambda \times \bar{\bar{X}}_{AWD}) X_{MWP} \times (1 + e^{-\eta(X_{cp} - \mu)})\right] \times 100\%$

Multi-Objective Optimization Combination Forecasting and Comparison Model

This section compares the proposed combined forecasting model with the commonly used single-point forecasting models. The single models of point prediction include ARIMA, LSTM, and T-convolutional neural network.

Experiment I: $PM_{2}.5$ **and** PM_{10} **forecasting.**

In this experiment, three traditional single models, **ARIMA**, **LSTM,** and **TCN**, are utilized to compare with the proposed system. The prediction results are shown in **Tables 5**, **6**.

(a) From **Tables 5**, **6**, we can see that the proposed model has achieved significant improvements compared with the three single models. In the forecast of $PM_{2}.5$ daily concentration in the three cities, the MAPE ($\times$100%) of the model in this study are **17.53130124**, **11.52643852,** and **6.00510985,** respectively, and the MAPE of PM_{10} are **20.10103656**, **19.61939713,** and **9.348984687,** respectively. In addition, there are substantial improvements in other data comparisons, which demonstrate the superior predictive power of the proposed model in simulating air contamination series.

(b) MAPE and RMSE are mainly used to measure the prediction error of each model. The smaller the value, the better the model prediction performance. In addition, the $\boldsymbol{R}$ index mainly evaluates the fit consistency between the

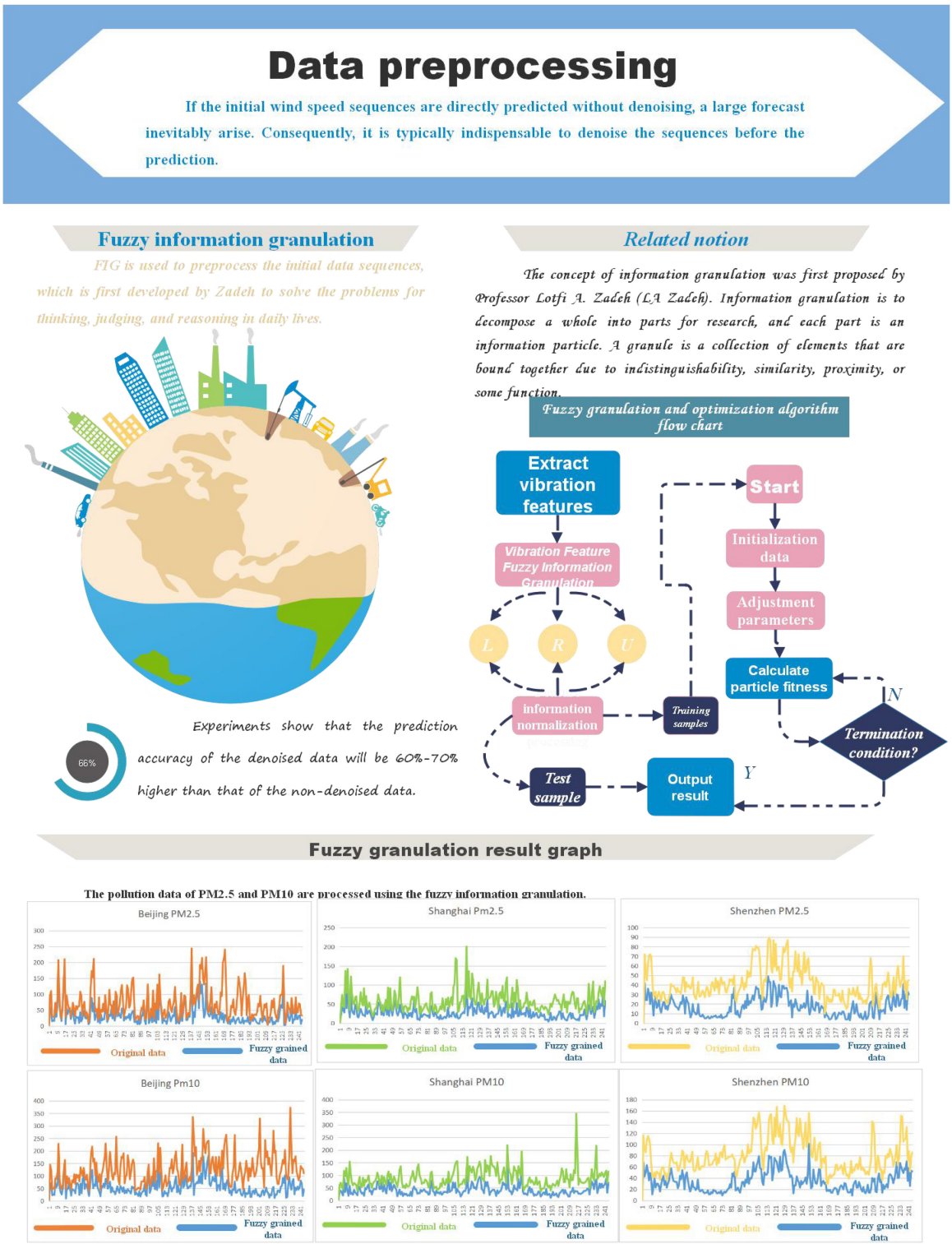

FIGURE 3 | Data preprocessing flowchart.

original value and the predicted value. The index values of the system are all larger than the reference model, indicating that the system has a better simulation effect on the air pollution sequence. The ***R*** value is negative, indicating that the model is not suitable for simulating the air pollution series.

TABLE 5 | Statistical errors of the proposed system and three traditional models for daily $PM_{2.5}$ concentrations of Beijing, Shanghai, and Shenzhen.

Dataset	Model	MAPE (×100%)	MSE	Adjusted *R* square	RMSE	SSE
Beijing	ARIMA	86.97079038	290.8284471	0.447545833148643	17.053693062077780	5.816568941111197e + 04
	LSTM	136.01785	413.20679	0.215077461604307	20.327488086028843	8.264135437752891e + 04
	TCN	142.5970955592992	483.03770834970896	0.1626560139686989	21.9781	66176.16604391013
	Proposed model	**17.53130124**	**148.870600556075**	**0.658175732872686**	**25.2317663**	**29774.1201112151**
Shanghai	ARIMA	43.06089753	139.425587	0.448273075306167	11.807861237106376	2.788511739895187e + 04
	LSTM	44.672726	173.59648	0.313053969451486	13.175601796860805	3.471929654188834e + 04
	TCN	60.826491189904864	265.8895658979947	−0.8598887306068743	16.3061	36426.870528025276
	Proposed model	**11.52643852**	**67.0458435333302**	**0.667495893927825**	**15.29028516**	**13409.1687066660**
Shenzhen	**ARIMA**	40.30779213	37.85861391	0.624570379270827	6.152935389900167	7.571722782457183e + 03
	LSTM	66.276283	89.870964	0.108783389745464	9.480030256857855	1.797419473418808e + 04
	TCN	60.826401222797664	84.42822174344232	0.0699984144048289	9.1885	11566.666378851598
	Proposed model	**6.00510985**	**17.5222223250933**	**0.985588304**	**7.806397025**	**3504.44446501865**

The indicators can be defined as $\boldsymbol{SSE} = \sum_{i=1}^{N} (\boldsymbol{PP}_i - \boldsymbol{AP}_i)^2$, $\boldsymbol{RMSE} = \sqrt{\left[\sum_{i=1}^{N} \boldsymbol{PP}_i - \boldsymbol{AP}_i\right] / N}$, *and* $\boldsymbol{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left|(\boldsymbol{AP}_i - \boldsymbol{PP}_i) / \boldsymbol{AP}\right| \times 100\%$. *The most satisfactory results are shown in bold.*

TABLE 6 | Statistical errors of the proposed system and three traditional models for daily PM_{10} concentrations of Beijing, Shanghai, and Shenzhen.

Dataset	Model	MAPE (×100%)	MSE	Adjusted *R* square	RMSE	SSE
Beijing	ARIMA	63.52736357	487.0786852	0.174520697390467	22.069859201835650	9.741573703776996e + 04
	LSTM	196.7689707455670	2.432077115750104e + 03	−3.121776177956965	49.316093881714760	4.864154231500208e + 05
	TCN	87.22617049732669	706.49757233221	−0.30037075679791525	26.5800 RMSE	96790.16740951277
	Proposed model	**20.10103656**	**172.374212971936**	**0.403054937**	**25.04239173**	**34474.8425943872**
Shanghai	ARIMA	29.09318151	176.4140437	0.528676587826571	13.282094853290085	3.528280873835899e + 04
	LSTM	35.3052930993369	2.069021408276871e + 02	0.447221882339130	14.384093326577350	4.138042816553741e + 04
	TCN	40.82371479881204	552.8902300983436	−0.09493369083894865	23.5136	75745.96152347307
	Proposed model	**19.61939713**	**98.0218221960804**	**0.342456465189497**	**19.8011941252118**	**19604.3644392161**
Shenzhen	ARIMA	25.09504498	88.93366541	0.659983745777606	9.430464750332934	1.778673308145440e + 04
	LSTM	28.8580663433921	1.252390110215527e + 02	0.521179080890740	11.191023680680540	2.504780220431053e + 04
	TCN	33.353551605811546	158.35503196357078	0.29102131681963195	12.5839	21694.639379009197
	Proposed model	**9.348984687**	**35.9146836500957**	**0.71286274**	**11.84927859**	**7182.93673001914**

The indicators can be defined as $\boldsymbol{SSE} = \sum_{i=1}^{N} (\boldsymbol{PP}_i - \boldsymbol{AP}_i)^2$, $\boldsymbol{RMSE} = \sqrt{\left[\sum_{i=1}^{N} \boldsymbol{PP}_i - \boldsymbol{AP}_i\right] / N}$, *and* $\boldsymbol{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left|(\boldsymbol{AP}_i - \boldsymbol{PP}_i) / \boldsymbol{AP}\right| \times 100\%$. *The most satisfactory results are shown in bold.*

Experiment II: SO2 and CO forecasting.

In Experiment **I**, we performed prediction experiments using PM2.5 and PM10 data, which achieved good results. To further verify the effectiveness of the prediction system, in Experiment 2, the SO_2 and CO data of three cities were used to conduct the experiment again.

Therefore, in this part, we selected the CO value and sulfur dioxide data of three cities for comparative experiments. The detailed results are shown in **Table 7**.

To further explore the application of the point prediction system, this experiment used SO_2 and CO daily datasets in Beijing, Shanghai, and Shenzhen to examine the superiority and applicability of the developed system. The results showed that the model proposed in this study not only exhibits the best prediction performance, indicating that the prediction system in this study is not only suitable for the prediction of $PM_{2.5}$ and PM_{10}, but also for the prediction of other air pollutants. Although the randomness and complexity of different datasets are different, the results show that the proposed model has strong applicability and effectiveness for the prediction of various air pollution and has potential application prospects in air pollution monitoring.

Indexes of Prediction Model and Interval Forecasting

Point forecasting only provides each forecast point for the target and does not show the probability of correct forecasting. However, in several problems, it is necessary to quantify the accuracy of estimates using countermeasures. Once the extent of uncertainty increases, the dependability of the point prediction decreases significantly. In contrast to point forecasting, prediction intervals not only provide the location in which observations are presumably made but also conjointly provide an indicator of capability known as the confidence level. Since interval forecasting is more reliable and informative than the settled point forecast, it is helpful to investigate and evaluate the data.

Experiment III: Interval forecasting and evaluation index.

To comprehensively evaluate the forecast results of the prediction model, four analysis indexes are adopted in the study, including the prediction interval coverage probability (**PICP**), coverage width criterion (**CWC**), prediction band width (**PBW**), and PI normalized averaged width (**PINAW**).

Prediction interval coverage probability is the basic evaluation index to assess the overall probability of the actual value falling into the **PBW**, and it is expressed as follows:

$$\mathbf{PICP} = \left(\frac{1}{n}\sum_{i=1}^{n}\overline{\overline{C}}_i\right) \times 100\% \qquad (4-3-1)$$

$$\overline{\overline{C}}_i = \begin{cases} 0\ \mathbf{y}_i \notin (\boldsymbol{L}_i, \boldsymbol{U}_\mathbf{I}) \\ 1\ \mathbf{y}_i \in (\boldsymbol{L}_\mathbf{I}, \boldsymbol{U}_\mathbf{I}) \end{cases} \qquad (4-3-2),$$

Where the variable y_i is the actual air contamination value. U_i and L_i represent the upper and lower bounds, respectively, and n is the number of samples.

If the PI width is sufficiently large, the PICP can easily reach 100%. Considering that the PICP meets the prediction interval nominal confidence (**PINC**) of the required prediction interval, the PBW should be as small as possible to guarantee the prediction effect.

Due to the contradiction between **PICP** and **PINRW**, **CWC** is used as a comprehensive evaluation index. In addition, because **PICP** is a basic evaluation indicator compared with **PINAW** (or **PINRW**) and is expected to achieve the desired nominal confidence level $\boldsymbol{\mu}$, an improved **CWC** design can assess the prediction effect better.

The modified CWC used in the experiment is defined as follows:

$$\mathrm{CWC}_{proposed} = \begin{cases} \boldsymbol{\eta}_1 \cdot \mathbf{PINAW}(\mathrm{PICP} \geq \boldsymbol{\mu}) \\ (0.1 + \boldsymbol{\mu}_1 \cdot \mathbf{PINAW})[1 + \exp(\boldsymbol{\eta}_2(\boldsymbol{\mu} - \mathbf{PICP}))] \\ (\mathrm{PICP} < \boldsymbol{\mu}) \qquad (4-3-3) \end{cases}$$

PBW, **PICP**, **PINAW**, and **CWC** are used to assess the **IP** performance. The detailed forecasting results of the proposed hybrid forecasting system and the comparative models are presented in **Figure 4** and **Table 8**.

(a) After optimization using the combined prediction algorithm, the interval prediction can estimate the upper and lower bounds of the probability prediction. Then, **PICP**, **PINAW**, and **CWC** indicators are selected to measure the performance of interval prediction of air pollution series. **PICP** mainly measures the probability of the original data entering the prediction interval, and **PINAW** is used to evaluate the normalized average width of the interval. This section adopts the interval prediction of $PM_{2.5}$ and PM_{10}, and the obtained results provide a practical application for analyzing the uncertainty of air pollution.

TABLE 7 | The forecasting performances of various models for SO_2 and CO in three cities.

Dataset (SO_2)	Model	MAPE (×100%)	MSE	Adjusted *R* square	RMSE	SSE
Beijing	ARIMA	16.9679992	0.594247371	0.214374179	0.77087442	1.19E + 02
	LSTM	62.5043641	3.575070593	−3.7264286	1.890785708	7.15E + 02
	TCN	21.19875252	0.358244808	−1.246512939	0.5985	48.7212939
	Proposed model	**8.801941086**	**0.017101509**	**−4.775512824**	**0.307953961**	**3.129576183**
Shanghai	ARIMA	16.804447	1.47013432	0.242891519	1.212490957	2.94E + 02
	LSTM	33.6266108	4.642820638	−1.391018856	2.154720548	9.29E + 02
	TCN	17.30624383	1.627779555	−0.25163582	1.2758	221.3780194
	Proposed model	**13.3948766**	**0.120567162**	**−3.960236439**	**0.817679472**	**22.06379071**
Shenzhen	ARIMA	13.93960103	0.183750043	0.68076782	0.428660755	36.75000861
	LSTM	7.4051157	0.003636823	0.659593382	0.060306074	0.727364522
	TCN	7.401322957	0.323880616	0.047690701	0.5691	44.0477638
	Proposed model	**4.5874185**	**0.143997793**	**−7.824128059**	**0.893606857**	**26.35159606**
Dataset (CO)						
Beijing	ARIMA	55.82851595	0.054564355	0.005877223	0.233590143	10.91287095
	LSTM	63.6736338	0.123604838	−1.251990073	0.351574798	24.72096768
	TCN	39.54360038	0.062681733	−0.71777545	0.2504	8.524715657
	Proposed model	**36.0314606**	**0.027527125**	**−3.06026958**	**0.390705028**	**5.037463826**
Shanghai	ARIMA	15.8161825	0.017936156	0.224697526	0.133925933	3.587231123
	LSTM	23.8757882	0.041810611	−0.807291964	0.204476431	8.362122138
	TCN	18.00573321	0.033547157	−0.834057427	0.1832	4.562413404
	Proposed model	**11.03447391**	**0.011493365**	**−5.983131272**	**0.07983478**	**0.210328539**
Shenzhen	ARIMA	22.46848191	0.379235401	0.341147671	0.615820916	75.84708014
	LSTM	9.8188329	0.005360446	0.498262188	0.07321507	1.072089292
	TCN	11.01559594	0.007725412	0.046786078	0.0879	1.050656042
	Proposed model	**6.6615594**	**0.003953843**	**−1.79785567**	**0.148073826**	**0.723553317**

The indicators can be defined as $\mathbf{SSE} = \sum_{i=1}^{N}(\mathbf{PP}_i - \mathbf{AP}_i)^2$, $\mathbf{RMSE} = \sqrt{\left[\sum_{i=1}^{N}\mathbf{PP}_i - \mathbf{AP}_i\right]/N}$, *and* $\mathbf{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|(\mathbf{AP}_i - \mathbf{PP}_i)/\mathbf{AP}\right| \times 100\%$. *The most satisfactory results are shown in bold.*

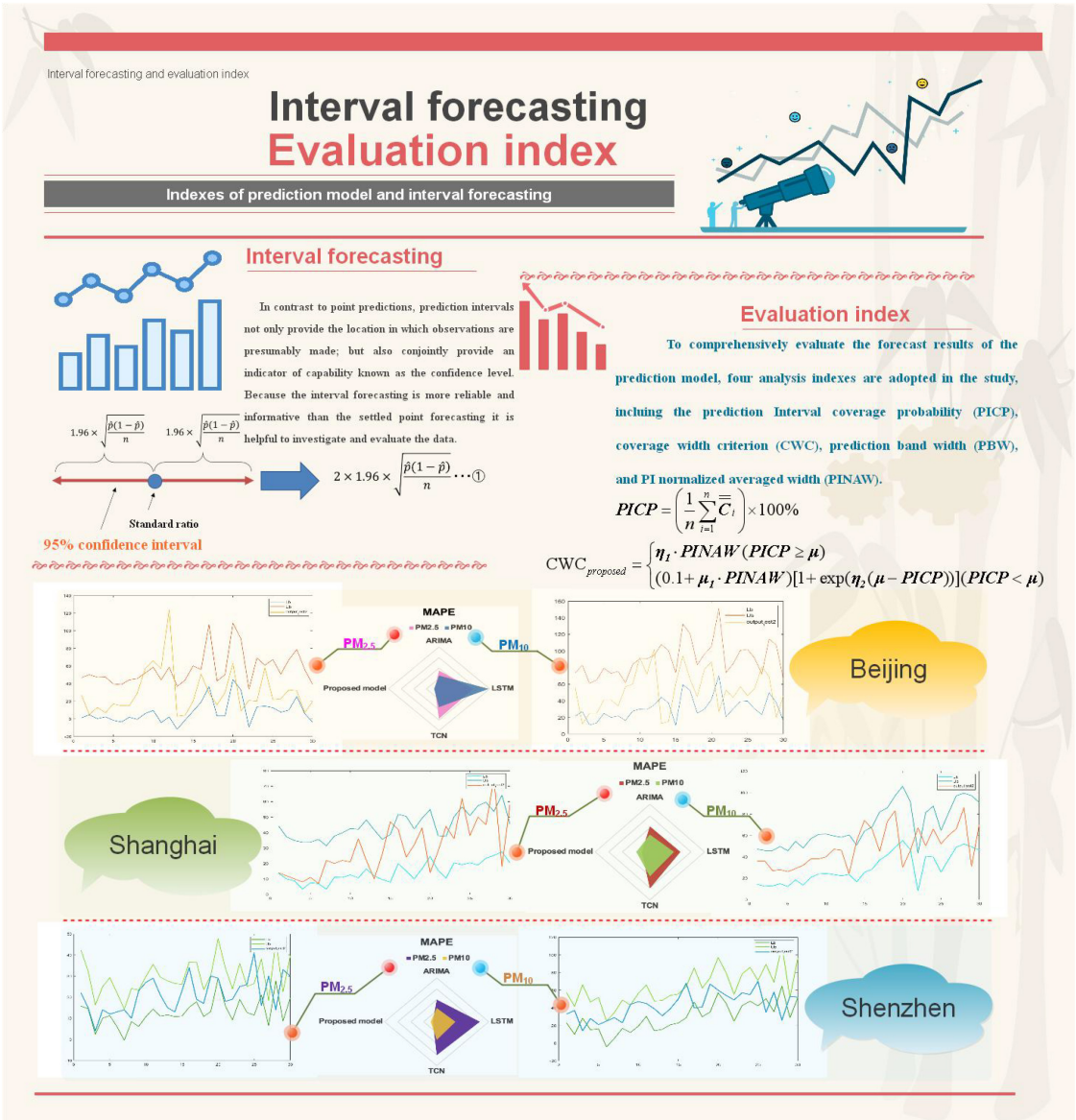

FIGURE 4 | A graph of interval prediction results.

(b) **Table 8** show the daily probability prediction evaluation results of $PM_{2}.5$ and PM_{10} in Beijing, Shanghai, and Shenzhen. It can be seen that the results of a single model interval are not good because the result of the width is too narrow and the coverage rate is low, while the interval coverage rate of the combined forecasting system proposed in this study is higher than that of a single model and the results are more accurate.

(c) It is difficult to satisfy all the optimal conditions due to the large number of indicators that measure the performance of interval prediction. However, the higher the confidence, the greater the coverage probability and the wider the interval. Therefore, the probability forecast has a certain prediction interval, which provides a reference for the actual application of air pollution monitoring.

Discussion

In this section, we used three methods to discuss the performance of the proposed combined forecasting system: forecasting effectiveness (FE), stability analysis (SA), and improvement ratio (IR).

Forecasting Effectiveness

To verify the availability of the relevant prediction system, the finite element method (Banik et al., 2020) is adopted in this study. This may be determined using the expected result of the prediction accuracy series, that is, the deviation between the expected value and normal deviation. The indicator is explained as follows.

Count the d-**th** order predicting availability element $m^d = \sum_{i=1}^{n} Q_i A_i^d$, where A_i is the prediction accuracy, Q_i is the discrete

TABLE 8 | Interval forecasting results of the proposed system for daily $PM_{2.5}$ and PM_{10}concentrations of Beijing, Shanghai, and Shenzhen.

$PM_{2.5}$	*Indicators*	Proposed model	ARIMA	LSTM	TCN
Site 1	**BW**	**27.27709623**	5.588932739	5.400260609	9.399792741
	CP	**0.8**	0.1	0.2	0.4
	CWC	**967.5573339**	4.62E + 17	3.30E + 15	4.05863E + 11
	PINAW	**0.534845024**	0.160988443	0.170554714	0.462680511
Site 2	**BW**	**25.83786828**	3.101276071	1.711007768	2.246018338
	CP	**0.7**	0.2	0.1	0.1
	CWC	**1216.366544**	1.23E + 15	9.11E + 16	3.79E + 17
	PINAW	**0.453295935**	0.063429076	0.031777656	0.132020441
Site 3	**BW**	**13.02474235**	1.822308173	2.649576346	2.393098477
	CP	**0.7**	0.1	0.1	0.2
	CWC	**1294.458165**	2.17E + 17	−2.06E + 17	1.50E + 15
	PINAW	**0.482397865**	0.075831053	−0.071902246	0.077871318
PM_{10}	***Indicators***	**Proposed model**	ARIMA	LSTM	TCN
Site 1	**BW**	**40.37286522**	2.864634888	2.867020268	7.299477064
	CP	**0.8**	0.1	0.3	0.2
	CWC	**1197.315173**	3.11E + 17	2.5209E + 13	5.96E + 15
	PINAW	**0.66185025**	0.108333022	0.193635382	0.30842717
Site 2	**BW**	**39.53693477**	3.204081672	18.80110089	4.136207351
	CP	**0.8**	0.3	0.4	0.3
	CWC	**1300.436217**	7.86722E + 12	3.06843E + 11	3.62306E + 13
	PINAW	**0.718853359**	0.060429758	0.349798846	0.278294498
Site 3	**BW**	**36.49380793**	5.868071455	4.509674695	2.006110094
	CP	**0.8**	0.3	0.1	0.1
	CWC	**1500.428327**	2.04898E + 13	1.78E + 17	3.38E + 17
	PINAW	**0.829404726**	0.157386217	0.062204966	0.11778856

The indicators can be defined as $\tilde{\lambda}_{cp} = \frac{1}{D}\sum_{i=1}^{D} c_i \times 100\%$,$X_{MWP} = \frac{1}{D}\sum_{i=1}^{D} \frac{U(X_i)-L(X_i)}{\hat{f}_i} \times 100\%$, $\overline{\overline{X}}_{BW} = \sum_{i=1}^{D} \xi_i$, *and* $\overline{\overline{X}}_{PINAW} = \left[1 - (1 + \lambda \times \bar{\bar{X}}_{AWD}) X_{MWP} \times (1 + e^{-\eta(X_{cp}-\mu)})\right] \times 100\%$. *The most satisfactory results are shown in bold.*

TABLE 9 | Forecasting effectiveness (FE) of different models ($PM_{2.5}$).

	Beijing		Shanghai		Shenzhen	
	1-Order	**2-Order**	**1-Order**	**2-Order**	**1-Order**	**2-Order**
Proposed model	**0.713962612**	**0.499107698**	**0.833677634**	**0.674130951**	**0.827274422**	**0.715910344**
Model ($PM_{2.5}$)						
ARIMA	0.537758525	0.365927564	0.626486448	0.434454329	0.687655842	0.501227795
LSTM	0.419353392	0.262811053	0.620248042	0.442276111	0.552618069	0.371521413
TCN	0.380004231	0.25407457	0.509236577	0.344293544	0.58990613	0.408951676
Model (PM_{10})						
Proposed model	**0.820324675**	**0.612407931**	**0.892528922**	**0.780386906**	**0.882115673**	**0.798110303**
ARIMA	0.584812738	0.406227792	0.71977323	0.539768394	0.776748883	0.625024691
LSTM	0.21938475	0.147504326	0.676791129	0.488168419	0.740723134	0.585249795
TCN	0.488147467	0.319256144	0.56964758	0.390169063	0.633711865	0.453582124

probability distribution, and $\sum_{i=1}^{n} \boldsymbol{Q_i} = \mathbf{1}$, $\boldsymbol{Q_i} > 0$. Since we could not obtain any prior information on $\boldsymbol{Q_i}$, it is determined as $\boldsymbol{Q_i} = 1/\boldsymbol{n}$, $\boldsymbol{i} = 1, 2, \ldots, \boldsymbol{n}$. The other $\boldsymbol{A_i}$ is calculated using $\boldsymbol{A_i} = 1 - \left|\overline{\overline{\xi}}_{\boldsymbol{i}}\right|$, in which $\overline{\overline{\xi}}_i$ is expressed as:

$$\overline{\overline{\xi}}_i = \begin{cases} -1, & (AP_i - PP_i)/AP_i < -1 \\ (AP_i - PP_i)/AP_i, & -1 \leq (AP_i - PP_i)/AP_i < 1 \\ 1, & (AP_i - PP_i)/AP_i > 1 \end{cases}, (4-4-1)$$

where $\boldsymbol{PP_i}$ and $\boldsymbol{AP_i}$ indicate the $\boldsymbol{i}$**-th** point forecast value and observation quantitative value, respectively.

Thereafter, the continuous function $\overline{\overline{\boldsymbol{L}}}\left(\boldsymbol{g}^1, \boldsymbol{g}^2, \ldots, \boldsymbol{g}^d\right)$ of a ***d-order*** unit is introduced to assess the $\boldsymbol{d}$**-th** order predicting availability. While there is only one variable in the equation $\overline{\overline{\boldsymbol{L}}}(\boldsymbol{z}) = \boldsymbol{z}$, the *first*-order FE can be expressed as $\overline{\overline{\boldsymbol{L}}}\left(g^1\right) = g^1$.

If there are two variables in this equation, for instance, $\overline{\overline{L}}(i,j) = i\left(1-\sqrt{j-i^2}\right)$, the second-order FE can be expressed as follows:

$$\overline{\overline{L}}\left(g^1, g^2\right) = g^1\left(1-\sqrt{g^2-\left(g^1\right)^2}\right) \quad (4-4-2)$$

According to the FE definition, the higher the value of $\overline{\overline{L}}$, the better the prediction performance of the models.

Therefore, the ***d-th*** order FE is expressed as $\boldsymbol{H}\left(\boldsymbol{m}^1, \boldsymbol{m}^2, \cdots, \boldsymbol{m}^k\right)$. Thus, the *first*-order prediction effectiveness is defined as $\boldsymbol{H}\left(\boldsymbol{m}^1\right) = \boldsymbol{m}^1$. If there are two variables in the equation, the second-order FE is given by $\boldsymbol{H}\left(\boldsymbol{m}^1, \boldsymbol{m}^2\right) = \boldsymbol{m}^1\left(1-\sqrt{\boldsymbol{m}^2-\left(\boldsymbol{m}^1\right)^2}\right)$.

By comparing the FE values with those of other related models, it can be easily concluded that the proposed system obtains the highest index value in both the first-order and second-order calculations, which shows that its performance in air pollution prediction exceeds that of other models. Specifically, we took Beijing PM_{10} data as an example in one-step, $\boldsymbol{H}_1^{\boldsymbol{Beijing}} = \mathbf{0.820324675}$ and $\boldsymbol{H}_2^{\boldsymbol{Beijing}} = \mathbf{0.612407931}$ in two orders, and the FE is much larger than that of other models. In other predictions, our proposed system exhibits the best forecast performance compared with the other models. The specific experimental results of the other models are listed in **Table 9**.

Sensitivity Analysis

In this section, the sensitivity of the proposed prediction system is analyzed experimentally. Since the weight determination method plays an important role in the final prediction, this study discusses the prediction sensitivity of the combined prediction model by adjusting the optimization parameters. In the parameter setting stage, the important parameter of the population size has a great influence on the optimization performance. Therefore, the experiment adopts the method of changing one parameter to examine its influence on the prediction result. Here, the size of the population is set to 40, 60, 80, and 100 in turn. The specific experimental results are shown in **Table 10**. The relevant conclusions are summarized as follows:

From **Table 10**, it can be seen that the performance of the proposed mode is different under various parameter settings. For example, in the Shenzhen PM2.5 forecast, MAPE values range from 5.9821 to 6.4834%.

Consequently, the fluctuation range of the forecast values in the three regions is small, indicating that the forecast system is less sensitive to the two modes and has a good stability in practical applications.

Improvement Ratio

In this section, the effectiveness of the combined forecasting model system is analyzed by the percentage improvement of MAPE and MSE. We proposed an index $\boldsymbol{IR_{MAPE}}$ to measure the improvement in the **PCFM** prediction accuracy. $\boldsymbol{IR_{MAPE}}$ can be

TABLE 10 | Experimental results of forecasting results under different population sizes.

Dataset	PM2.5				PM10			
	40	60	80	100	40	60	80	100
Beijing								
MAPE (×100%)	16.7949636457552	17.7356139438864	16.5110612195599	14.5260609900006	19.9055947149062	18.9704302780156	20.0524797526017	21.1438211511366
RMSE	24.6970552160974	25.3686235730786	24.4857948286102	23.3129364646248	24.9616400286964	24.2687122670146	25.1103732743030	26.0590974636506
SSE	30497.2268173483	32178.3530996280	29977.7074194398	27174.6503301816	31154.1736461110	29448.5197549573	31526.5422987416	33953.8280310020
Shanghai								
MAPE (×100%)	11.7258101883320	11.5430390511560	11.6253007536679	11.5970879035290	25.7764408550232	22.7232812706143	14.6954959388701	13.8943981605036
RMSE	15.6570375896180	15.3272904374202	15.5085690707090	15.4803567790693	29.5040374762424	26.4801510616650	18.4758225872704	17.5545394655169
SSE	12257.1413041355	11746.2916076516	12025.7857310475	11982.0723003638	43524.4113699757	35059.9200124298	17067.8010138146	15408.0927923195
Shenzhen								
MAPE (×100%)	5.9821028059 6730	6.1100433517 6550	6.4834299242 8906	6.2499767640 4358	9.3955940572 2180	9.6249946829 7147	10.1251000479952	10.0495620103120
RMSE	7.7788584591 5623	8.0177187303 4480	8.5561268049 5358	8.2522204788 6010	11.8654999185498	12.0187644430554	12.4795681613774	12.4454433453428
SSE	3025.5319463 7932	3214.1906819 4609	3660.3652951 2225	3404.9571415 8590	7039.5044158 5532	7222.5349368 8267	7786.9810747 2327	7744.4530031 0681

The indicators can be defined as $\boldsymbol{SSE} = \sum_{i=1}^{N}(\boldsymbol{PP}_i - \boldsymbol{AP}_i)^2$, $\boldsymbol{RMSE} = \sqrt{\left[\sum_{i=1}^{N}\boldsymbol{PP}_i - \boldsymbol{AP}_i\right]/\boldsymbol{N}}$, *and* $\boldsymbol{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\left(\boldsymbol{AP}_i - \boldsymbol{PP}_i\right)/\boldsymbol{AP}\right| \times 100\%$. *The most satisfactory results are shown in bold.*

TABLE 11 | Improvement ratio (IR) for CO of different models (×100%).

	Model	Beijing	Shanghai	Shenzhen
$\bar{\bar{P}}_{MAPE}$	**ARIMA**	35.460472%	30.233%	70.3515%
	LSTM	43.412275%	53.7838%	32.155%
	TCN	8.8817%	38.71689%	39.526%
$\bar{\bar{P}}_{MSE}$	**ARIMA**	49.551%	35.92069%	98.957%
	LSTM	77.729%	72.51089%	26.2404%
	TCN	56.084%	65.7396%	48.8203%

The ***MAPE*** *×100% of the* ***proposed prediction system is*** $MAPE^{Beijing} = 36.0314606$, $MAPE^{Shanghai} = 11.03447391$ *and* $MAPE^{Shenzhen} = 6.6615594$.

expressed as:

$$\bar{\bar{P}}_{MAPE} = \left[\left(MAPE^{com} - MAPE^{pro}\right) / MAPE^{com}\right] \times 100\%$$

$$\bar{\bar{P}}_{MSE} = \left[\left(MSE^{com} - MSE^{pro}\right) / MSE^{com}\right] \times 100\%,$$

where $MAPE^{com}$ is the compared model **MAPE** values and $MAPE^{pro}$ indicates **MAPE** values of the prediction system. Moreover, the three models are compared with different indicators, which shows the superiority of the combined forecasting model system. The detailed calculation results are shown in **Table 11**.

(a) The model is improved by **MAPE**, which verifies the superiority of the proposed prediction system. Compared with the **ARIMA** model, this model improves by 35.460472%. For the **LSTM** model, the combined prediction system achieves a 43.412275% improvement in **MAPE**. The results show that the system has a good prediction effect on $PM_{2.5}$ and PM_{10}.

(b) For the three urban datasets considering mean square error (**MSE**) and **MAPE**, the proposed prediction system still achieves a significant improvement in prediction accuracy. Experiments show that, compared with the **TCN** model (taking Beijing as an example), the **MAPE** of the combined model is improved by 56.084%.

CONCLUSION

Predicting air quality plays a vital role in the environment and economy of energy development, which is widely discussed worldwide. In recent years, more researchers have focused on the methods to forecast $PM_{2.5}$ and PM_{10} concentrations and provide useful information for the citizens in their daily lives. However, to overcome the limitations and negative effects of an individual approach, this study develops a novel combined forecasting system that takes advantage of data preprocessing, single models, and the interval predicting approach.

The developed system includes an advanced data denoising technique, three single forecasting algorithms, and an optimization approach to predict the $PM_{2.5}$ and PM_{10} concentrations. Based on the experiments, we concluded that the combined model has the following advantages: (1) as for data denoising strategy, the combined system computes the data series without fluctuation and uncertainty by **FIG**, which yields better performance compared with single models by decomposing and reconstructing the initial data. (2) In the comparative experiments, to predict the $PM_{2.5}$ and PM_{10} concentrations of three cities, we found that the $PM_{2.5}$ **MAPE** (×100%) values of the proposed system are **17.53130124**, **11.52643852**, and **6.00510985**, which provide more satisfying results than the **ARIMA** models (**86.97079038**, **43.06089753**, and **40.30779213**). (3) Consequently, **MODOA** is utilized as an advanced optimization algorithm to determine the weight of every single model and to obtain the forecasting values of $PM_{2.5}$ and PM_{10} concentrations.

The proposed early warning system has many practical applications, such as warning and guiding the public before the occurrence of harmful air pollutants and mining the characteristics of air pollutants.

(1) The fuzzy preference rough set was applied to the early warning system to determine the main pollutants suitable for different cities. Attribute selection simplifies the process of early warning systems and makes the prediction of pollutants more effective. In addition, these results can help decision-makers in relevant sectors to monitor and analyze certain polluting pollutants, which play a crucial role in formulating effective strategies for each city.

(2) In the developed early warning system, the interval forecast based on deterministic forecast provides the forecast range and the confidence level, which can be used to analyze and monitor the uncertainty information of the future value of pollutants. Air quality warning systems trigger alerts when air pollution exceeds an upper limit. According to the forecast range, different early warning levels can also be divided as a guide for daily life.

Therefore, we concluded that the proposed combined predicting system enhances the forecasting capacity and accuracy of $PM_{2.5}$ and PM_{10} concentrations by conducting and analyzing the experiments. Accurate forecasts not only reduce the cost and risk of dealing with air pollution systems but also help policymakers come up with effective strategies.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://www.tianqihoubao.com.

AUTHOR CONTRIBUTIONS

ZL contributed to the conception and design of the study. JW organized the model development and the experimental analysis and wrote the first draft of the manuscript. JL wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This study was supported by the Science and Technology Development Fund, FDCT, Macau SAR, under Grant 0064/2021/A2.

REFERENCES

Air Quality Expert Group (2012). *Fine particulate matter (PM2.5) in the UK. Technical report*. Available online at: http://uk-air.defra.gov.uk [accessed on March 19, 2022]

Akyüz, M., and Çabuk, H. (2009). Meteorological variations of PM2.5/PM10 concentrations and particle-associated polycyclic aromatic hydrocarbons in the atmospheric environment of Zonguldak, Turkey. *J. Hazard. Mater.* 170, 13–21. doi: 10.1016/j.jhazmat.2009.05.029

Bai, Y., Li, Y., Zeng, B., Li, C., and Zhang, J. (2019). Hourly PM2. 5 concentration forecast using stacked autoencoder model with emphasis on seasonality. *J. Clean. Prod.* 224, 739–750. doi: 10.1016/j.jclepro.2019.03.253

Banik, A., Behera, C., Sarathkumar, T. V., and Goswami, A. K. (2020). Uncertain wind power forecasting using LSTM-based prediction interval. *IET Renew. Power Gener.* 14, 2657–2667. doi: 10.1049/iet-rpg.2019.1238

Beaulant, A. L., Perron, G., Kleinpeter, J., Weber, C., Ranchin, T., and Wald, L. (2008). Adding virtual measuring stations to a network for urban air pollution mapping. *Environ. Int.* 34, 599–605. doi: 10.1016/j.envint.2007.12.004

Bergen, S., Sheppard, L., Sampson, P. D., Kim, S. Y., Richards, M., Vedal, S., et al. (2013). A national prediction model for PM2.5 component exposures and measurement error-corrected health effect inference. *Environ. Health Perspect.* 121, 1017–1025. doi: 10.1289/ehp.1206010

Bin, H., Zu, Y. X., and Zhang, C. (2014). A forecasting method of short-term electric power load based on BP neural network. *Appl. Mech. Mater.* 538, 247–250. doi: 10.4028/www.scientific.net/AMM.538.247

Borrego, C., Cascão, P., Lopes, M., Amorim, J. H., Tavares, R., Rodrigues, V., et al. (2011). Impact of urban planning alternatives on air quality: URBAIR model application. *WIT Trans. Ecol. Environ.* 147, 13–24. doi: 10.2495/AIR110021

Chen, Y., Kang, Y., Chen, Y., and Wang, Z. (2020). Probabilistic forecasting with temporal convolutional neural network. *Neurocomputing* 399, 491–501. doi: 10.1016/j.neucom.2020.03.011

Ding, S. (2019). A novel discrete grey multivariable model and its application in forecasting the output value of China's high-tech industries. *Comput. Ind. Eng.* 127, 749–760. doi: 10.1016/j.cie.2018.11.016

Djalalova, I., Delle Monache, L., and Wilczak, J. (2015). PM2. 5 analog forecast and Kalman filter post-processing for the Community Multiscale Air Quality (CMAQ) model. *Atmos. Environ.* 108, 76–87. doi: 10.1016/j.atmosenv.2015.02.021

Genc, D. D., Yesilyurt, C., and Tuncel, G. (2010). Air pollution forecasting in Ankara, Turkey using air pollution index and its relation to assimilative capacity of the atmosphere. *Environ. Monit. Assess.* 166, 11–27. doi: 10.1007/s10661-009-0981-y

Gers, F. A., Schmidhuber, J., and Cummins, F. (2000). Learning to forget: continual prediction with LSTM. *Neural Comput.* 12, 2451–2471. doi: 10.1162/089976600300015015

Guo, J., Liu, Z., Huang, W., Wei, Y., and Cao, J. (2018). Short-term traffic flow prediction using fuzzy information granulation approach under different time intervals. *IET Intell. Transp. Syst.* 12, 143–150. doi: 10.1049/iet-its.2017.0144

Hamilton James, D. (2015). *Time Series Analysis*. Beijing: China Renmin University Press.

He, Y., Yan, Y., Wang, X., and Wang, C. (2019a). Uncertainty forecasting for streamflow based on support vector regression method with fuzzy information granulation. *Energy Procedia* 158, 6189–6194. doi: 10.1016/j.egypro.2019.01.489

He, Y., Yan, Y., and Xu, Q. (2019b). Wind and solar power probability density prediction via fuzzy information granulation and support vector quantile regression. *Int. J. Electr. Power Energy Syst.* 113, 515–527. doi: 10.1016/j.ijepes.2019.05.075

Li, J., Wang, R., Wang, J., and Li, Y. (2018). Analysis and forecasting of the oil consumption in China based on combination models optimized by artificial intelligence algorithms. *Energy* 144, 243–264. doi: 10.1016/j.energy.2017.12.042

Liu, H., Jin, K., and Duan, Z. (2019). Air PM2.5 concentration multi-step forecasting using a new hybrid modeling method: comparing cases for four cities in China. *Atmos. Pollut. Res.* 10, 1588–1600. doi: 10.1016/j.apr.2019.05.007

Liu, W., Guo, G., Chen, F., and Chen, Y. (2019). Meteorological pattern analysis assisted daily PM2.5 grades prediction using SVM optimized by PSO algorithm. *Atmos. Pollut. Res.* 10, 1482–1491. doi: 10.1016/j.apr.2019.04.005

Majumder, H., and Maity, K. (2018). Application of GRNN and multivariate hybrid approach to predict and optimize WEDM responses for Ni-Ti shape memory alloy. *Appl. Soft Comput.* 70, 665–679. doi: 10.1016/j.asoc.2018.06.026

McKeen, S. A., Chung, S. H., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., et al. (2007). Evaluation of several PM2.5 forecast models using data collected during the ICARTT/NEAQS 2004 field study. *J. Geophys. Res. Atmos* 112, 1–20. doi: 10.1029/2006JD007608

Mencar, C., and Fanelli, A. M. (2008). Interpretability constraints for fuzzy information granulation. *Inf. Sci.* 178, 4585–4618. doi: 10.1016/j.ins.2008.08.015

Osowski, S., and Garanty, K. (2007). Forecasting of the daily meteorological pollution using wavelets and support vector machine. *Eng. Appl. Artif. Intell.* 20, 745–755. doi: 10.1016/j.engappai.2006.10.008

Pai, T. Y., Hanaki, K., and Chiou, R. J. (2013). Forecasting hourly roadside particulate matter in taipei county of taiwan based on first-order and one-variable grey model. *Clean Soil Air Water.* 41, 737–742. doi: 10.1002/clen.201000402

Samia, A., Kaouther, N., and Abdelwahed, T. (2012). A hybrid ARIMA and artificial neural networks model to forecast air quality in urban areas: case of Tunisia. *Adv. Mater. Res.* 518, 2969–2979. doi: 10.4028/www.scientific.net/AMR.518-523.2969

Sfetsos, A. (2000). A comparison of various forecasting techniques applied to mean hourly wind speed time series. *Renew. Energy* 21, 23–35. doi: 10.1016/S0960-1481(99)00125-1

Shelhamer, E., Long, J., and Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 3431–3440. doi: 10.1109/TPAMI.2016.2572683

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1929–1958.

State of Global Air (2020). *Health Effects Institute*. Available online at: https://www.stateofglobalair.org [accessed on October 22, 2020]

Sun, W., and Li, Z. (2020). Hourly PM2.5 concentration forecasting based on feature extraction and stacking-driven ensemble model for the winter of the Beijing-Tianjin-Hebei area. *Atmos. Pollut. Res.* 11, 110–121. doi: 10.1016/j.apr.2020.02.022

Tian, Q., and Wang, H. (2021). Predicting remaining useful life of rolling bearings based on reliable degradation indicator and temporal convolution network with the quantile regression. *Appl. Sci.* 11:4773. doi: 10.3390/app11114773

van Donkelaar, A., Martin, R. V., and Park, R. J. (2006). Estimating ground-level PM2.5 using aerosol optical depth determined from satellite remote sensing. *J. Geophys. Res. Atmos* 111:D21201. doi: 10.1029/2005JD006996

Wakamatsu, S., Morikawa, T., and Ito, A. (2013). Air pollution trends in Japan between 1970 and 2012 and impact of urban air pollution countermeasures. *Asian J. Atmos. Environ.* 7, 177–190. doi: 10.5572/ajae.2013.7.4.177

Wang, J., and Hu, J. (2015). A robust combination approach for short-term wind speed forecasting and analysis - Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. *Energy* 93, 41–56. doi: 10.1016/j.energy.2015.08.045

Wang, J., Zhang, L., and Li, Z. (2022b). Interval forecasting system for electricity load based on data pre-processing strategy and multi-objective optimization algorithm. *Appl. Energy* 305:117911. doi: 10.1016/j.apenergy.2021.117911

Wang, J., Wang, R., and Li, Z. (2022a). A combined forecasting system based on multi-objective optimization and feature extraction strategy for hourly PM2.5 concentration. *Appl. Soft Comput.* 114:108034. doi: 10.1016/j.asoc.2021.108034

Wang, J., Zhang, L., Wang, C., and Liu, Z. (2021). A regional pretraining-classification-selection forecasting system for wind power point forecasting and interval forecasting. *Appl. Soft Comput.* 113:107941. doi: 10.1016/j.asoc.2021.107941

Wang, S., Wang, J., Lu, H., and Zhao, W. (2021). A novel combined model for wind speed prediction – Combination of linear model, shallow neural networks, and deep learning approaches. *Energy* 234:121275. doi: 10.1016/j.energy.2021.121275

Wang, Y., Wang, J., Zhao, G., and Dong, Y. (2012). Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: a case study of China. *Energy Policy* 48, 284–294. doi: 10.1016/j.enpol.2012.05.026

Wang, Z., Wang, C., and Wu, J. (2016). Wind energy potential assessment and forecasting research based on the data pre-processing technique and swarm intelligent optimization algorithms. *Sustainability* 8:1191. doi: 10.3390/su8111191

Wei, D., Wang, J., Niu, X., and Li, Z. (2021). Wind speed forecasting system based on gated recurrent units and convolutional spiking neural networks. *Appl. Energy* 292:116842. doi: 10.1016/j.apenergy.2021.116842

Xiao, X., Xie, W., Zhou, Y., Zhao, W., Liu, X., and Zhang, C. (2019). Prediction and analysis of energy demand of high energy density AC/DC park based on spatial static load forecasting method. *J. Eng.* 2019, 3388–3391. doi: 10.1049/joe.2018.8389

Yang, H., Zhu, Z., Li, C., and Li, R. (2020). A novel combined forecasting system for air pollutants concentration based on fuzzy theory and optimization of aggregation weight. *Appl. Soft Comput. J.* 87:105972. doi: 10.1016/j.asoc.2019.105972

Yang, W., Yao, Q., Ye, K., and Xu, C. Z. (2020). Empirical mode decomposition and temporal convolutional networks for remaining useful life estimation. *Int. J. Parallel Program.* 48, 61–79. doi: 10.1007/s10766-019-00650-1

Zadeh, L. A. (1997). Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst.* 90, 111–127. doi: 10.1016/S0165-0114(97)00077-8

Zhang, X., Wang, J., and Gao, Y. (2019). A hybrid short-term electricity price forecasting framework: cuckoo search-based feature selection with singular spectrum analysis and SVM. *Energy Econ.* 81, 899–913. doi: 10.1016/j.eneco.2019.05.026

Zhang, Y., and Na, S. (2018). A novel agricultural commodity price forecasting model based on fuzzy information granulation and MEA-SVM model. *Math. Probl. Eng.* 2018, 1–10. doi: 10.1155/2018/2540681

Zhu, R., Liao, W., and Wang, Y. (2020). Short-term prediction for wind power based on temporal convolutional network. *Energy Rep.* 6, 424–429. doi: 10.1016/j.egyr.2020.11.219

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.