Applied Soft Computing Journal 102 (2021) 106957

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

Review article

Intelligent modeling strategies for forecasting air quality time series: A review

Hui Liu *, Guangxi Yan, Zhu Duan, Chao Chen

Institute of Artificial Intelligence and Robotics (IAIR), Key Laboratory of Traffic Safety on Track of Ministry of Education, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China

ARTICLE INFO

Article history:
Received 27 April 2020
Received in revised form 25 September 2020
Accepted 27 November 2020
Available online 20 January 2021

Keywords:
Air quality forecasting
Intelligent predictors
Hybrid modeling strategies

ABSTRACT

In recent years, the deterioration of air quality, the frequent events of the air contaminants, and the health impacts from that have caused continuous attention by the government and the public. Based on that, suitable and effective forecasting tools are urgently needed in scientific research. In this study, the basic forecasting algorithms are introduced as the simple forecasting models with their background, applications, advantages, and limitations, which include shallow predictors and deep learning predictors. Then, to enhance the forecasting ability, the data processing methods and two commonly used auxiliary methods (the ensemble learning and the metaheuristic optimization) in the hybrid models have been reviewed. The recent articles of the spatiotemporal aspects have also brought changes in both the analysis and the modeling methods. Furthermore, the representative models are summarized to present the structures of efficient predictive models. Some possible research directions of the air pollution forecasting are given at the end. This review aims to provide a comprehensive literature summary of the intelligent modeling strategies in the air quality forecasting, which may be helpful for subsequent study.

Contents

* Correspondence to: Institute of Artificial Intelligence and Robotics (IAIR), Key Laboratory of Traffic Safety on Track of Ministry of Education, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, Hunan, China.
E-mail address: csuliuhui@csu.edu.cn (H. Liu).

https://doi.org/10.1016/j.asoc.2020.106957
1568-4946/

1. Introduction

With the rapid progress of modern industry and traffic, the explosive inflation of the population, and the process of urbanization, air pollution has become a global problem [1,2]. Various air pollutants will exacerbate the environment and cause serious environmental disasters like the greenhouse effect, the ozone hole, and photo-chemical smog [3], leading to increased worldwide health risks and diseases burdens, such as the human respiratory diseases, chronic diseases, cardiovascular diseases, or even cancer [4,5]. What is more, the haze air pollution has a strong impact on the economic development [6] worldwide and policies, such as GDP decrease in developing countries [7], energy consumption and energy structure [8], and the social welfare burden in developed countries [9]. Air pollution is a gradual process and it may lead to catastrophic consequences without effective control. So reducing air pollution can be an effective measure to control the above diseases and benefit society [10].

1.1. The main pollutants

The major pollutants include SO_2, TSP (dust, PM_{10}, $PM_{2.5}$), NOx, CO, O_3, which are a mixture of particles and gases and turn to be a growing problem [11,12]. The particulate matters (PM) are not a single type of particles but a mixture of small particles in the atmosphere, which leads to deterioration of air quality and the frequent occurrence of haze weather. Researches proved that the 10μm diameter particles can reach the upper respiratory tract and the particles under 5μm diameter can enter into the deep part of the bronchioles [13]. What is more, the particles under 1μm can even penetrate the alveoli [13]. The sulfur dioxide (SO_2) is concentrated by volcanic eruptions and industrial development. Fossil fuels, such as coal and petroleum, contain sulfur. The sulfur dioxide can be formed during the consumption of fossil energy. The SO_2 in the atmosphere may also cause the acid rains to damage the environment. And SO_2 also has great effects on increasing the rate of respiratory diseases [14].

Most of the emissions of NO_2 are from the combustion of fossil fuels in industry and vehicles [3]. NO_2 is also regarded as one of the major causes of the acid rains and it also plays an important role in the form of ozone [15]. Moreover, NO_2 has negative effects on human health and ecosystems, leading to a high incidence of lung diseases. Due to the strong oxidizing properties, the ozone is a kind of harmful gas and can even damage human health at ground level. By high concentration, it is easy to cause lesions in the upper respiratory tract and irritation to the skin, eyes and nose [16]. A high concentration of O_3 has a negative influence on food crop yields and the crop loss may be higher in the future [17].

1.2. The current research of air quality forecasting models

The air quality forecasting can provide effective data of the environmental quality to the society and the government and reflect the trends of environmental pollution in advance. There are some fundamental indexes for the approximate classification of air quality forecasting models: the time-scale classification of forecasting models, forecasting approach, and the type of inputs. By the data resolution of time, air quality prediction includes very short-term, short-term, medium-term, and long-term categories, as shown in Fig. 1. The forecasting data in the reviewed articles are mainly in hourly and daily resolution. The daily values are usually from the data for years, which represent long-term trends of the contamination development [18].

In general, a shorter forecasting temporal extent can achieve more detailed and accurate results and longer forecasting temporal extent provides long-term information for research. This kind of application of the prediction research helps to realize long-term pollution control processes, while hourly prediction could contribute to short-term air quality monitoring and management of higher accuracy. The establishment of the air quality modeling is a complex process of system engineering and a difficult point in environmental science research, which helps to study the relevance of the causes and results of contaminants and contribute to the mitigation solutions in the future with effective analysis.

In recent years, air quality forecasting models have been continuously improved and expanded. So that various methods and approaches for time series data can be taken to facilitate targeted control of the atmospheric pollution and to prevent serious pollution incidents.

Nomenclature

Abbreviations

AdaBoost	Adaptive Boosting
AMI	Adaptive multiple input
ANFIS	Adaptive neural network fuzzy inference system
ANN	Artificial neural network
AQI	Air quality index
ARIMA	Autoregressive integrated moving average model
BA	Bat algorithm
BPNN	Backpropagation neural network
CAMX	Comprehensive air quality model with extensions
CEEMD	Complementary empirical mode decomposition
CEEMDAN	Complete ensemble empirical mode decomposition with adaptive noise
CMAQ	Community multiscale air quality
CNN	Convolutional neural network
CPSOGSA	Chaotic particle swarm optimization method and gravitation search algorithm
CS	Cuckoo search
DBN	Deep belief network
DE	Differential evolution
DES	Double exponential smoothing
DESVM	Differential evolution optimized support vector machine
DWT	Discrete wavelet transform
EEMD	Ensemble empirical mode decomposition
ELM	Extreme learning machine
EMD	Empirical mode decomposition
ENN	Elman neural network
EWT	Empirical wavelet transform
FCM	Fuzzy C-Means
FFNN	Feedforward neural network
FKM	Fuzzy K-Medoid
FLM	Fuzzy logic model
FTS	Fuzzy time series
GA	Genetic algorithm
GABP	Genetic algorithm optimized backpropagation
GCN	Graph convolutional networks
GK	Gustafson-Kessel
GM	Grey model
GRNN	Generalized regression neural network
GRU	Gated recurrent unit
GWO	Grey wolf optimizer
ICEEMDAN	Improved complete ensemble empirical mode decomposition with adaptive noise
IA	Indices of agreement
IMF	Intrinsic mode function
KNN	K-nearest neighbor
LASSO	Least absolute shrinkage and selection operator
LLE	Locally linear embedding
LSSVM	Least squares support vector machine
LSTM	Long short-term memory
MADRID	Model of aerosol dynamics, reaction, ionization, and dissolution
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MCSDE	Modified cuckoo search and differential evolution algorithm
MGWO	modified grey wolf optimization
MI	Mutual information
ML	Manifold learning
MLP	Multilayer perceptron
MLR	Multiple linear regression
MM	Mesoscale model
MOALO	Multi-objective ant lion optimizer
MODWT	Maximum overlap discrete wavelet transform
MOHHO	Multi-objective salp swarm algorithm
MOSSA	Modified cuckoo search and differential evolution algorithm
MSE	Mean square error
MTL	Multitask learning
OLDBN	Online deep belief network
PCA	Principal component analysis
PCR	Principal component regression
PSO	Particle swarm optimization
PSOGSA	Particle swarm optimization and gravitational search algorithm
RBF	Radial basis function
RBFNN	Radial basis function neural network
RBM	recurrent neural network
RELU	Rectified linear unit
RF	Random forest
RFELM	Random fourier extreme learning machine
RSP	Respirable Suspended Particle
RMSE	Root mean square error
RNN	recurrent neural network
SAE	Stacked auto-encoder
SD	Secondary-decomposition
SE	Sample entropy
SPM	Suspended particulate matter
SSA	Singular spectrum analysis
STELSTM	Spatiotemporal ensemble strategy
STFV	Spatiotemporal feature vector
STSVR	Space–time support vector regression
SVM	Support vector machine
SVR	Support vector regression
TSP	Total suspended particles
VM	Variational mode
VMD	Variational mode decomposition
WELM	Weighted extreme learning machine
WPD	Wavelet packet decomposition
WRF	Weather research and forecasting

WRF-Chem	Weather research and forecasting/ chemistry
WM	Winning-Model
WNN	Wavelet Neural Networks
WT	Wavelet transform
XGBOOST	Extreme gradient Boosting

The chemical transport models (CTMs) aim at describing chemical and meteorological processes in the atmosphere, focusing on the emission, transport, and mixing of air pollutant concentrations to establish a corresponding mathematical model [19,20]. The weather research and forecasting (WRF) based models are utilized for atmospheric research and applications in prediction, such as WRF-Chem [21] and WRF/Chem-MADRID [22]. Other deterministic methods like CAMx [23], CMAQ model [24], and LOTOS-EUROS [25] are also used in air pollutants forecasting by scholars. A lot of chemical dynamic conditions, reaction index, and chemical products should be taken into consideration. The accuracy of these deterministic methods relies on the plenty of information and data of pollutant sources and explicit description of chemical reactions and physical processes [26]. The key points for further improvement may be the acquisition of the nonlinearity between the concentration of the pollutants and the sources of their transport and diffusion [27,28], especially by application in area with complex terrain [29].

The statistical methods mainly reflect the statistical connections between various impacts and air pollutants in time series, applying the historical data to predict air quality instead of physical, chemical, and biological processes. They are data-based models with the theory of mathematical statistics, probability, and stochastic processes. The traditional statistical models used in air pollution forecasting are autoregressive integrated moving average (ARIMA) [30], grey model (GM) [31], and other regression models. The prediction accuracy of the grey model highly relies on the data characteristics and the grey parameters. The regression models used for the prediction of pollutant concentrations mainly include the stepwise regression [32], principal component regression (PCR) [33], and multiple linear regression (MLR) [34]. The statistical models work on a principle of describing the relationship between variables based on possibility and statistical average and they may generally achieve satisfactory accuracy of the concentration level for future predictions [35]. However, the statistical models still have space for improvement on the accuracy while the behavior of air contaminations and other regional features could be complex, disordered, and highly nonlinear [36]. Therefore, effective approaches are still needed to account for handling the modeling of air quality.

With the rapid progress of artificial intelligence and machine learning in recent years, artificial intelligence-based forecasting models are getting popular and attract more attention. The intelligent predictors can achieve better performance of accuracy in approaches when dealing with nonlinearities and interactive relationships in air pollution modeling [37]. The main advantages of the models are they can handle the non-linear elements, conduct the operation of large-scale data volumes to solve problems. The application does not require an in-depth understanding of the dynamic and chemical processes between air contamination levels and other relative variables in the atmosphere [38]. A commonly-used predictor is an artificial neural network (ANN), which simulates the system of the human brain and nervous to model nonlinear series. The neural networks are also conducted improvement in years of application, resulting in more evolutionary versions for the air pollution prediction, such as the backpropagation neural network (BPNN) [39], the radial basis function neural network (RBFNN) [40], the generalized regression neural network (GRNN) [41], the wavelet neural network (WNN) [42,43]. Other models, which are also popular in air pollutant prediction, are the support vector machine (SVM) and fuzzy logic. Similar to ANN, both SVM and the fuzzy logic model (FLM) have improved versions of better ability to deal with complex variables and increase accuracies like the least squares support vector machine (LSSVM) [44] and the adaptive neuro-fuzzy inference system (ANFIS) [45,46]. Besides, the deep learning models based on forecasting algorithms can realize the functions with multiple layers. And the effectiveness of these models has been presented by comparison in experiments of the deep belief network (DBN) [47], the convolutional neural network (CNN) [48], and the long short-term memory (LSTM) [49].

Recently, much effort has been made to review and research the characteristics of different types of intelligent predictive models in the air quality field [50–52]. It was generally concluded that no single predictors can be appropriate for in all aspects of modeling because of their limitations and there was no individual intelligent approach appropriate for all specific problems. The hybrid models generally mean not only the combination of different algorithms or methods but also the combination of advantages of each component, which leads to better performances. The hybrid models mainly include several modeling strategies [53]. The simple hybrid model, Mishra, and Goyal built a novel model by PCA and ANN for NO_2 concentrations prediction [54], which combines the feature extraction method with predictors. Bai et al. used wavelet transform (WT) to preprocess the relevant data into a series of sublayers as data decomposition and then used the BPNN to perform the forecasting and the forecasting accuracy is higher than that of the simple BPNN model [55]. The state-of-the-art intelligent hybrid models may combine more data processing or optimization algorithms. Liu et al. proposed a hybrid forecasting model [53]. The wavelet packet decomposition (WPD) is applied to process the input data and the BPNN is used to conduct the three-step prediction, which will be optimized by the particle swarm optimization (PSO) and further strengthened by the adaptive boosting (Adaboost) algorithm. From the above examples, it can be seen that besides the predictors and the data processing as the components of hybrid models [56–58], the application of metaheuristic optimization [59–62] and ensemble learning [63–65] are also popular in application. The metaheuristic optimizations search for the optimal parameters at an acceptable range and ensemble learning integrates multiple models to achieve a model with better performance.

Moreover, as continuous, and advanced development of air pollution research, the research areas and targets should not be limited to a small scope of locations or certain urban areas. A broader range of overall air pollution trends, which contain a huge amount of input datasets in space, should be taken into consideration to conduct precise and comprehensive analysis for air quality of many places in a big region or multiple periods. Preprocessing of multiple data [66–68], improving individual models for spatiotemporal interaction [69,70] and a combination of spatial and temporal predictors in hybrid models [71,72] are three main aspects in this section.

1.3. The motivation of the review

Recently, some review papers have focused on the predictive methodologies of air quality. The main content of reviews from 2015 to 2020 is briefly summarized as follows. Bai et al. had reviewed statistical forecasting methods, artificial intelligence methods, and numerical forecasting methods for air quality prediction, but did not discuss the intelligent models and the

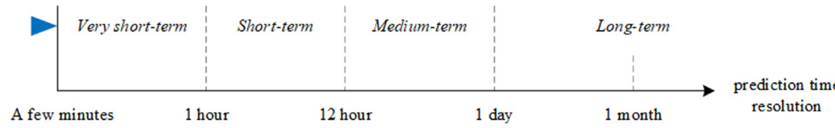

Fig. 1. The time-scale classification of air quality forecasting.

combination of them in depth [52]. Cabaneros et al. surveyed mainly the application of ANN with some hybrid strategies for air pollution prediction [38]. Taheri Shahraiyni and Sodoudi listed the statistical modelings of PM_{10} in urban areas [73]. Rybarczyk and Zalakeviciute summarized the machine learning models to predict air pollution and compared these methods to traditional approaches [51]. Masih reviewed the main techniques of machine learning applied for pollutant concentration estimation and forecasting [50]. Wang et al. reviewed some forecasting models of air quality, but the main purpose was on the interaction of air pollution and terrain nexus with the impacts of energy generation and consumption [74]. Gulia et al. summarized the air quality management plan (UAQMP) strategies in urban air pollution management and monitoring air quality in different countries [75]. Casazza et al. concentrated on spatial monitoring and modeling and listed assessment and forecasting models for urban port planning [76].

Although the abovementioned reviews excellently summarized and analyzed the research status and prospect of air pollutants predictive study from different aspects, the in-depth reviews of the intelligent modeling strategies are rare and the extensive reviews for the advanced models are still needed. The data-driven strategies have developed rapidly in the last decades. They have been the hotspot of current prediction research and various improved intelligent methods can be applied to pollution forecasting [77–91]. Therefore, new theories should be concluded and clarified. This paper focuses on these research gaps and presents an integrated review of the intelligent modeling methods for air quality forecasting. The commonly used data processing methods and the auxiliary methods are applied in hybrid intelligent models to increase the predictive ability of intelligent predictors. With the trends of increasing input data and a broader range of air quality research, we also summarized the adopted spatiotemporal methods and search for the way they raised the predictive ability of hybrid models. The main contribution of the review is to comprehensively present and classify the intelligent modeling methods for air quality forecasting, as well as the theory and statistical results. We hope to complement the existing technologies of the scientific research aspect.

1.4. Organization and the evaluation indexes

This study focuses on the above deficiencies and offers a detailed review of the applications of intelligent modeling strategies in air quality forecasting. More specifically, we summarize the progressive improvement in air quality forecasting. In this paper, different models will be classified and compared. The remainder of the review is presented as follows: Section 2 gives an in-depth introduction and classification of intelligent predictors for predictive application and summarizes their advantages and limitations. Section 3 gives a review of the methods in data processing, which includes decomposition and feature extraction. Section 4 and Section 5 present auxiliary methods of intelligent hybrid models (i.e., metaheuristic optimization and ensemble learning). Section 6 reviews the recent research papers in the spatial and temporal aspects. Section 7 presents an extensive discussion of these methods and the trends and challenges of air quality research in the future. Section 8 concludes to end this paper.

Many evaluation indexes have been used in the field of error assessment. The definitions and formulas of the indexes involved in this review are shown in Table 1, where n is the total number of instances, P_i represents the predicted value, and A_i represents the actual value, $\bar{A}$ and $\bar{P}$ stand for the mean actual and mean predicted value.

1.5. Literature search and select strategy

The literature search was conducted in systematical research highly indexed database Google Scholar and Web of Science journal publications for relevant literature published mainly in the past decade. They are among the few databases which compile the most significant science databases such as IEEE Xplore, Science Direct, and Springer, where the major articles can be found in the air quality research.

The key search terms included “Air quality”, “prediction”, “forecasting”, “simulation”, “urban area”, “modeling”, “hybrid model” and “machine learning” with different combinations. The main purpose is on recent studies of the most credible, authoritative, and reliable science research work covering worldwide scope. Some conference papers were also selected. The search process was repeated until the relevant citation stopped. The list of references for the chosen articles was also analyzed to identify further references. Only the literature published in the English language were finally selected. The articles were then categorized according to the expression of the necessary information (e.g., scientific background, input variables, modeling process, and accuracy of the prediction). Besides the citations of original methodology papers, the main content of application papers is from the year 2000 to 2020 and about 90% of them are from 2010 to 2020. The whole searching process is shown in Fig. 2.

As a result of the literature search, 209 articles were chosen. The second step was to read the title and abstract of all articles and checking the quartiles of the journals. Based on this filtration 20 papers were reduced from the selection list because; (1) they did not address the topic; (2) the works were not based on computational models; (3) the similar works by the same authors. Another limitation of 12 papers was to focus on the recent quartiles of the journals according to the Journal Citation Reports (JCR), which conclude the Q1–Q4 categories. The result of the second step provides us with 177 documents in the review content. In the last step, 66 papers of the original methodology and the principle introduction instead of application papers were excluded. Finally, a total number of 111 manuscripts were included for the qualitative and quantitative analysis of the review after careful reading. Fig. 3 represents the flow diagram of the article selection for the systematic review, which is based on the PRISMA guideline.

2. Effective intelligent predictors

After a detailed and comprehensive literature survey, the commonly-used intelligent predictors for air quality prediction are shallow predictors and deep learning predictors. Each predictor has been optimized in continuous development. Due to the similarities in structure and algorithm, the updated simple intelligent models will be classified and discussed in this section.

Table 1
The definitions and formulas of indexes applied in this review.

Index	Definition	Equation
MAE	Mean absolute error	$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} \lvert P_i - A_i \rvert$
MSE	Mean squared error	$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} (P_i - A_i)^2$
RMSE	Root mean square error	$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (P_i - A_i)^2}$
MAPE	Mean absolute percentage error	$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left\lvert \frac{A_i - P_i}{A_i} \right\rvert \times 100\%$
IA	Index of agreement	$\mathrm{IA} = 1 - \frac{\sum_{i=1}^{n}(P_i - A_i)^2}{\sum_{i=1}^{n}(\lvert P_i - \bar{A}\rvert + \lvert A_i - \bar{A}\rvert)^2}$
CC	Correlation coefficient	$\mathrm{CC} = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(A_i - \bar{A})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(A_i - \bar{A})^2}}$

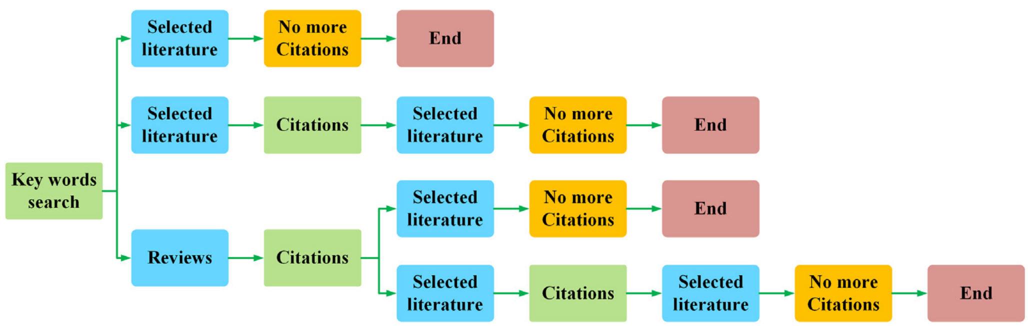

Fig. 2. The process of the methodology used for searching the literature.

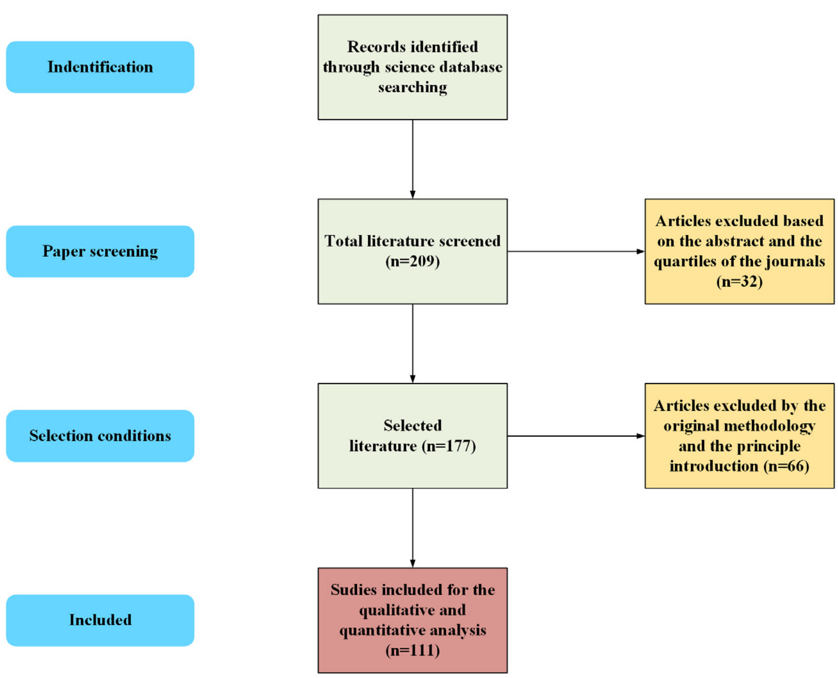

Fig. 3. PRISMA flowchart for the systematic selection of the relevant articles.

2.1. *Shallow intelligent predictors*

Currently, the commonly utilized shallow predictors for air quality prediction could be further separated into four subcategories: the artificial neural network (ANN), extreme learning machine (ELM), support vector machine (SVM), and fuzzy logic model (FLM).

2.1.1. *Artificial neural network*

The ANN simulates the neuron framework of the human brain to build a simplified model with different networks according to the application requirements. The basic ANN is a predictive model of massively connected nodes, in which each node represents an activation function. The ANN is the most predominant algorithm used in various engineering fields and can use a huge amount

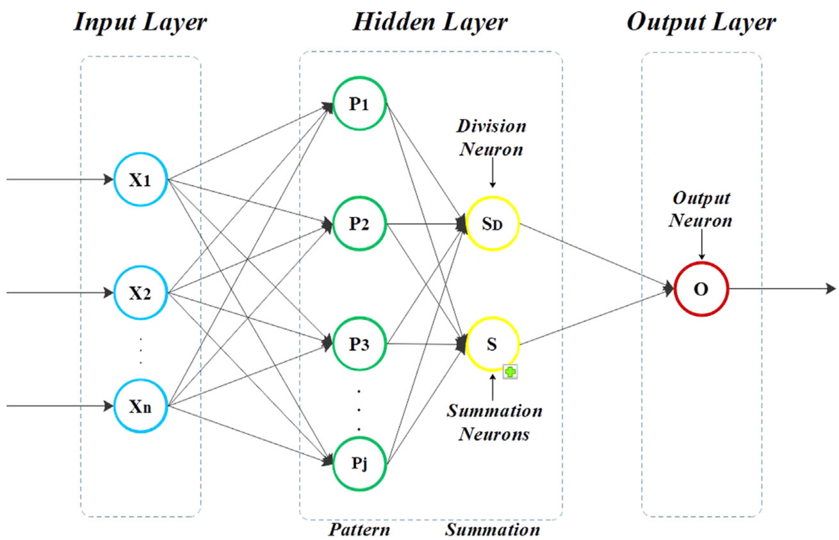

Fig. 4. A basic structure of GRNN.

of neurons to learn the nonlinear information from the input data and infer the complex relationship between the unknown data to build models so that the models can generalize and predict unknown data of air pollutants [92,93]. With continuous application and updating, many improved models of traditional ANN have been proposed and used for air quality prediction.

The backpropagation neural network (BPNN) model is a widely used machine learning method with feed-forward multilayer networks. The BP algorithm has been used to improve ANN. In the BPNN structure, it is a 3-layer predictor, in which the input data are delivered by artificial neurons from the input layer to the hidden layer so that the information flows forward and the results will be transmitted to the output layer. The errors of the network, however, will be propagated backward as the feedback [39]. BPNN model also has been applied to solve the problems of air quality prediction [94,95].

The radial basis function neural network (RBFNN) uses the radial basis function as the activation function. The main difference of RBFNN with other ANNs is that the number of hidden neurons in the hidden layer could be adjusted appropriately to meet the convergence to the planned goal. Zheng and Shang compared two types of neural network BPNN and RBFNN for the prediction of $PM_{2.5}$ concentration, and the results proved that RBFNN has higher accuracy (MAE = 0.0040 and CC = 0.9851) than BPNN in the prediction [96]. The generalized regression neural network (GRNN) is an improved algorithm based on RBFNN and has good nonlinear approximation performance and fast convergence. GRNN has four layers, in which the hidden layer has been divided into the pattern layer and the summation layer [97]. The summation layer has two kinds of neurons, the summation neurons, and a division neuron. The number of summation neurons always keeps the same as the number of GRNN output neurons, as shown in Fig. 3. Zheng et al. compared GRNN and traditional BPNN by the predicted NOx emissions and the actual values for better convergence rate, predictive accuracy, and less computing time about 1/6 of BPNN [98]. Antanasijević et al. used GRNN to forecast PM_{10} concentration and compared with the conventional principal component regression (PCR) model, the results of which have shown that the GRNN has better predictive capabilities than the PCR model with the same datasets and input data [99]. A similar comparison of BPNN and GRNN is conducted by Sun et al. for forecasting the NH_3, H_2S, CO_2, and PM_{10} levels and demonstrates the advantages of GRNN over BPNN, including, faster running time, better approximation, and extra stability [100].

Combined with the ANN and wavelet analysis, the wavelet neural networks (WNN) utilizes a wavelet basis function as the activation function. So that it can take advantage of both networks to achieve better learning ability and higher accuracy for air pollution prediction [43,101]. Elman proposed the Elman neural network (ENN) in 1990 [102]. By adding a context layer to the hidden layer as a one-step delay operator, ENN achieves the goal of better memory and the ability to adjust to time-varying identity and improve the stability of the network. Many researchers used ENN for the concentrations of pollutants forecasting and optimized the ENN model for better predictive performance [103–105].

2.1.2. Support vector machine

The support vector machine (SVM) is a machine learning method based on the theory of structural risk minimization (SRM) to minimize the generalization error from the training error. However, the mathematical structures and processing methods of SVM are quite different from ANNs. SVM constructs hyperplanes for separating different classes. For a continuous output variable, the regression analysis can be used to replace classification. Nieto et al. used support vector regression (SVR), which is the type of optimized modeling to find an approximate solution of highly nonlinear problems and to research the air quality in the Oviedo urban area of Northern Spain with success [106]. A basic function structure of SVR is shown in Fig. 5, where ξ_i and ξ_i^* are lack variables. Liu et al. have taken input data to predict Beijing's air quality index (AQI) using the SVR algorithm, aiming at increasing the AQI forecast accuracy [107]. The work of Sánchez et al. also showed that the SVR model had high accuracy in prediction of the dependence between the main pollutants such as CO, NO_2, SO_2 in the Avilés urban area [108].

SVM represents to be efficient with good performance, and much research on air quality prediction is based on the SVM model [109] and analysis. Lu and Wang made comparisons of SVM and radial basis function (RBF) in predicting pollutant levels of CO, SO_2, NOx, O_3, and respirable suspended particles (RSP), while SVM has shown better generalization performance and smaller error. SVM has been also used with different kernel models to forecast pollutants in regression and time-series [110]. Saxena and Shekhawat have also developed an SVM model, in which the kernel parameter and bias are handled by grey wolf optimizer (GWO) [111]. Least Squares Support Vector Machine

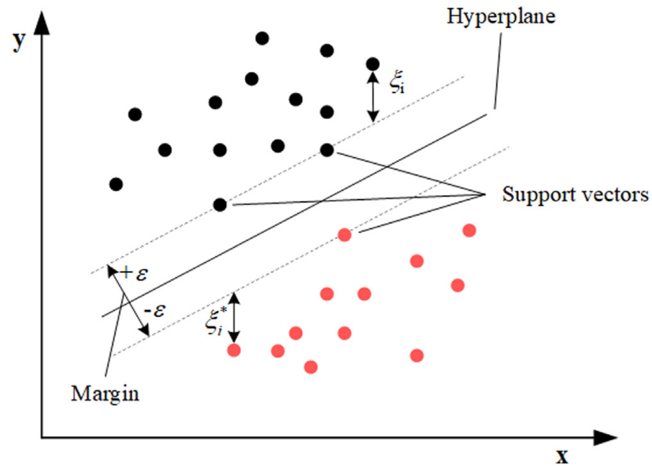

Fig. 5. A basic structure of the division of a dataset by SVR.

(LSSVM) was proposed by Suykens and Vandewalle to solve problems by pattern classification and function estimation [112]. The traditional SVM uses the quadratic programming method, but the LSSVM applies the least squares linear system as the loss function instead. Some researchers utilized LSSVM to predict air pollutant levels [44,113].

2.1.3. Fuzzy logic model

The fuzzy logic model (FLM) is an approach of multiple-valued logic to imitate the uncertainty of the human brain and the form for unknown or undetermined models. It is based on the concept of a fuzzy set and it utilizes the concept of membership function to handle fuzzy relations and solve various types of uncertain problems. Song and Chissom proposed a prediction model for fuzzy time series and started the research of fuzzy time series theory and application. In general, time series prediction is one of the main goals of time series analysis [114]. Based on fuzzy logic, the fuzzy time series (FTS) algorithm is composed of the corresponding fuzzy sets and suitable for the uncertainty in the data, which reduces the influence of inaccurate data on the accuracy of the model. Dincer and Akkuş used an improved FTS model with fuzzy k-medoids (FKM) algorithm in SO_2 concentrations forecasting and compared the results with improved FTS models by the Gustafson-Kessel (GK) and fuzzy c-means (FCM) algorithms [115]. Domańska and Wojtylak established an air pollution forecasting model of FTS to obtain a set of pollutant data by fuzzy grouping [116].

The adaptive neuro-fuzzy inference system (ANFIS), developed by Jang, is a new system composed of a fuzzy inference system and ANN [117]. Fig. 4 shows the general architecture of ANFIS is composed of six layers. ANFIS can be utilized to adjust the premise parameters and conclusion parameters, and can automatically obtain the If-Then rules. Moreover, the neuro-fuzzy technique can handle and analyze various types of information and achieve self-learning properties, which will further improve forecasting quality [118]. The neuro-fuzzy models have been widely used in forecasting of air pollutants level by scholars [46,119,120].

2.1.4. Extreme learning machine

Extreme learning machine (ELM) is a kind of machine learning system on feedforward neuron network (FFNN), which is suitable for supervised learning and unsupervised learning problems. Its prominent identity is that the weights of hidden layer nodes are random or artificially given and do not require updating after index setting [121]. The traditional ELM has a single hidden layer, whose activation function can be a linear function or a sigmoid function. Fig. 7 shows the basic frame of the ELM with n input neurons, j hidden neurons, m output w is the weight matrix between the input layer and the hidden layer and β is the output matrix between the hidden layer and the output layer. When compared with other shallow learning systems, such as single-layer perceptron and SVM, it is considered to have a better learning rate and usage in the application of earth and environmental sciences [121]. Zhao et al. proposed an ELM-based to predict the air quality with hourly records around Helsinki [122]. Zhang and Ding studied the concentration of air pollutants in Hongkong and evaluated the pollutants from meteorological and time parameters [123]. After the comparison of the forecasting performance of MLR, FFANN-BP, and ELM, the results showed that ELM performs well in aspects of precision, robustness, and generalization.

The online sequential extreme learning machine (OSELM) was proposed to obtain the data chunk-by-chunk [124]. Bueno et al. conducted the air quality prediction in São Paulo, Brazil, and compare OS-ELMs with ELMs, showing higher accuracy of OSELM under both individual and ensemble processes [125]. The similar research across Canada on the hourly concentration forecasts of the ozone, $PM_{2.5}$, and NO_2 was proposed by Peng et al. in which the OSELM generally obtained better predictive performance to the MLPNN, ELM and updated-MLR, and updated-ELM) in the MAE and correlation scores [126]. To enhance the insufficient robust ability of ELM against outliers, Deng et al. proposed an improved ELM, called the weighted regularized extreme learning machine (WRELM) [127]. Because of the excellent performance in generalization and robustness, the WRELM is also selected in the field of the air pollutant prediction [128].

2.2. Deep learning predictors

Deep learning is one subfield of machine learning and a class of algorithms, which use artificial neural networks as a basic structure to characterize and learn data. Compared to traditional methods, which increase many neurons to solve problems, the deep learning models utilize multiple layers to achieve the functions gradually. The purpose of deep learning is to extract more features by modeling with more hidden layers and a large sum of training data, so to enhance the predictive accuracy ultimately. The commonly-used deep learning models in air quality prediction are restricted Boltzmann machine, convolutional neural network, and recurrent neural network.

2.2.1. Deep belief network

The restricted Boltzmann machine (RBM) was invented by Smolensky, which can acquire the probability distribution through the input data [132]. The RBM unit is composed of a visual layer and a hidden layer and it is also a stack-component of deep belief network (DBN) [133]. The structure is described in Fig. 5, in which only the top layer is undirected. The scholars often combined algorithms to form the individual models for further improving the data generating, the prediction accuracy, and efficiency [47,131].

Xie developed a DBN-based model combined with manifold learning (ML) or locally linear embedding (LLE) algorithm, which processes nonlinear data and enables dimension reduction, to overcome the shortage of traditional neural networks by the shallow structure and obtain low-dimensional input for the deep neural network [47]. Li et al. also constructed a DBN model with the multitask learning (MTL) algorithm [131]. The advantage of MTL is that it can explore the relationship between subtasks, so the correlations between air pollutants ($PM_{2.5}$, SO_2, and NO_2) have been analyzed in the process to prove the system suitability of the prediction model.

Table 2
Part of deep learning models in air quality forecasting.

Deep learning models	Pollutants	Cities	Data resolution	Contrastive models	Degree of improvement[a]
CNN [48]	O_3	Seoul	hourly	ANN, SAE, LSTM	11.19% (MAE at peaks), 10.43% (MAE at daytime), 17.58% (MAE at nighttime), 6.76% (CC),
LSTM [129]	O_3, $PM_{2.5}$, NOx, CO	NCT-Delhi	hourly	SVM, ANN, M5P, and REPTree	O_3: 22.4% (CC), $PM_{2.5}$:8.99% (CC), NOx: 17.7% (CC), CO: 9.5% (CC)
LSTM [130]	PM_{10}	Seoul	daily	LR, RNN	4.85% (MSE), 5.1% (RMSE)
LSTM [49]	O_3	Kuwait	hourly	FFNN, ARIMA	98.46% (MAE)
DBN (LLE) [47]	$PM_{2.5}$	Chongqing	daily	DBN, BPNN	8.34% (MAPE), 3.59% (RMSE)
DBN (MTL) [131]	$PM_{2.5}$, SO_2, NO_2	Beijing	hourly	OLDBN, DBN, WM	$PM_{2.5}$: 21.6% (MAE), 17.76% (MAPE) SO_2: 3.63% (MAE), 34.66% (MAPE) NO_2: 2.22% (MAE), 25.54% (MAPE)

[a]The improvement degree is from the error indexes of the proposed model and the best contrastive model.

2.2.2. Convolutional neural network

Unlike other intelligent prediction models, the convolutional neural network (CNN) is one kind of FFNN formed by deep structures and calculation of convolution to analyze the input images and simulate the visual perception of biological organisms.

Zhang et al. proposed a CNN-based model including nine convolution layers, two pooling layers, and two dropout layers with the rectified linear unit (ReLU) as activation. The dropout layers can be utilized to prevent overfitting problems [134]. Rijal et al. developed three CNN models with different structures as the base learners, combined with the feedforward network, to evaluate $PM_{2.5}$ levels from images [135]. Eslami et al. provided a CNN model with five convolutional layers, a fully connected layer and output layer in the real-time prediction of O_3 concentrations across Seoul in South Korea, and compared with a long short-term memory (LSTM) model, ANN, and a stacked auto-encoder (SAE) model [48]. The experimental results showed the CNN model had the best indices of agreement (0.87) and lowest MAE (10.3).

2.2.3. Recurrent neural network

Different from the designed construction for image data in CNN, the recurrent neural networks (RNN) use a kind of time series data in loop structures, which transfers the information of data in circular ways. Nevertheless, the single traditional RNNs are difficult to capture long-term correlation of the air pollutant information by gradient explosion or gradient disappearance. Hence, the following two improved RNN models are more practical in the application of air quality forecasting: long short-term memory (LSTM) [49,129,130,136–138], gated recurrent unit (GRU) [139,140].

The key of LSTM is the state of cells, which is like a conveyor belt and runs over the whole process with only a few interactions but also unable to control the memorized information. LSTM controls the state of the cells by the gates (input, forget and output gate) to transmit and transform the status, which passes through different gates to be removed or added to the cell state. The basic structure of LSTM is presented in Fig. 9. GRU is an improved RNN that is like LSTM and obtain a simpler form. It has only a rest gate and an update gate and cannot control or preserve internal memory.

2.3. Epilog

The purpose of this section is a review of the basic intelligent predictors in air quality forecasting. These predictors have been summarized from research papers and organized in this section: ANN, SVM, FLM, ELM, DBN, CNN, and RNN. Although they are the fundamental of intelligent forecasting, there are still differences in forecasting capabilities. Table 2 describes part of the reviewed results of the deep learning methods. The deep learning has developed rapidly in recent years, but from the related papers on air quality prediction, its application in this field is not extensive. After the review, the conclusion can be drawn that deep learning methods can reach higher accuracy faster speed, and better efficiency in prediction. For example, by Freeman et al. the mean absolute error (MAE) of the LSTM model decreased by almost 98.46% in comparison with that of simple FFNN [49].

Table 3
The percent of the application frequency of the shallow predictors.

Model type	Frequency	Percent
ANN	12	40%
SVM	7	23.33%
FLM	6	20%
ELM	5	16.67%
Total articles	30	100%

Table 4
The percent of the application frequency of the deep learning predictors.

Model type	Frequency	Percent
DBN	2	15.38%
CNN	3	23.08%
RNN	8	61.54%
Total articles	13	100%

Tables 3 and 4 show the application frequency of the single forecasting models. According to the tables, the current trends could be found that the ANN as shallow predictors and the RNN as deep learning predictors are used more frequently than others. That may be taken into the consideration for future studies among modeling strategies.

Table 5 demonstrates detailed survey results of predictive models for air quality, including merits and demerits, types of air pollutants, and corresponding algorithms.

ANN is a simplified model according to the structure of the human brain neuron network for information processing. There are also many improvements over the years. ANNs have high accuracy in classification and is easy to learn and construct models for nonlinear connections, leading to good accuracy in air quality forecasting. However, ANNs also have the limitations that neural networks require many hyperparameters to be adjusted that could affect the forecasting results and the learning process cannot be observed. A huge amount of neurons in models may cause extension of learning time.

In the structure of SVM, kernel functions are used to map to high dimensional spaces and solve the nonlinear classification,

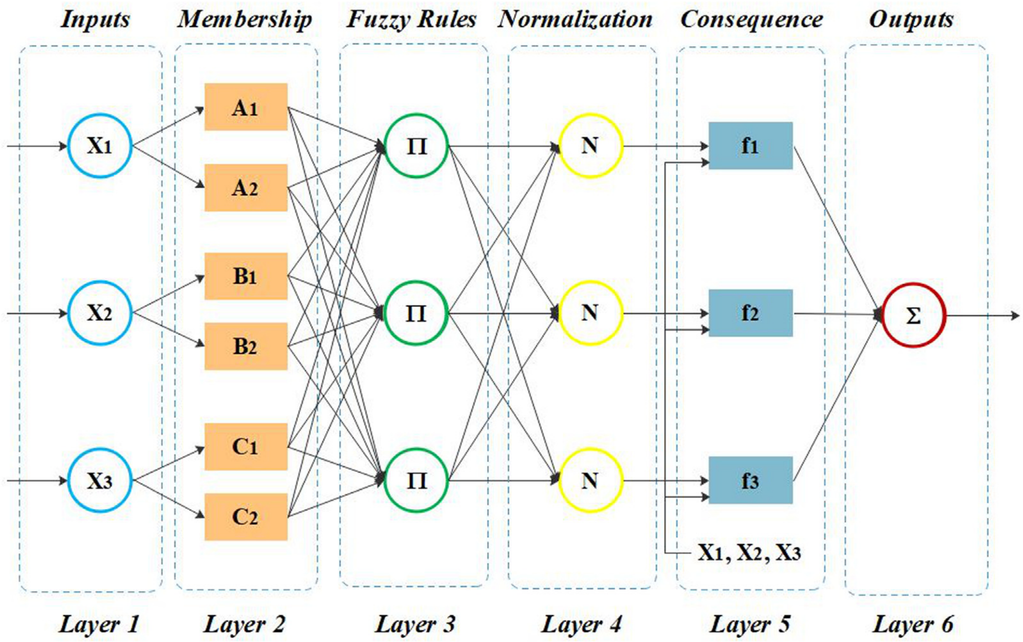

Fig. 6. A general structure of ANFIS.

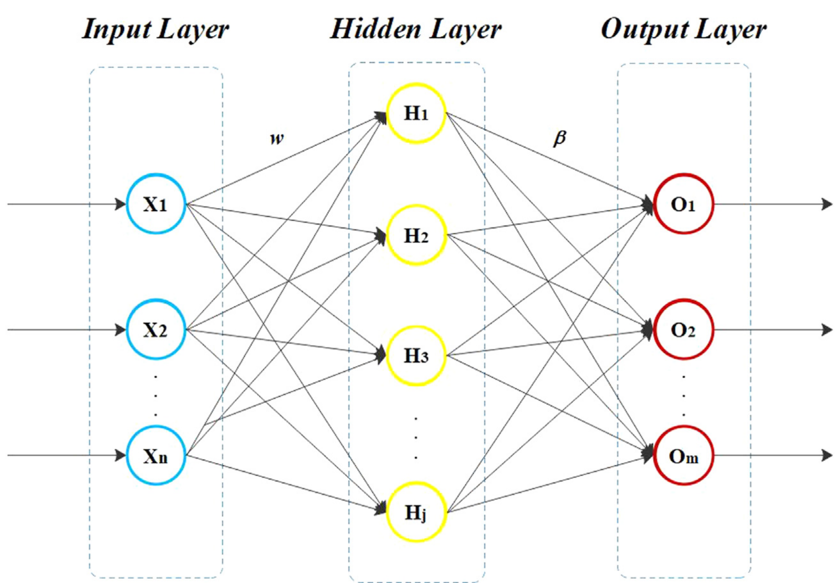

Fig. 7. A basic structure of ELM.

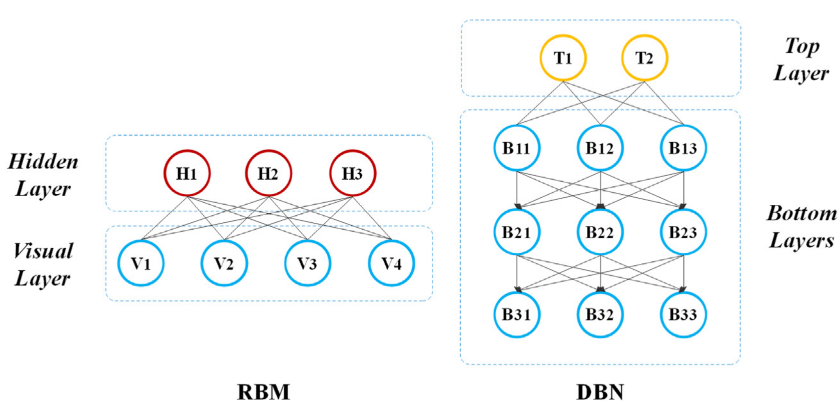

Fig. 8. The graphical depictions of RBM and DBN.

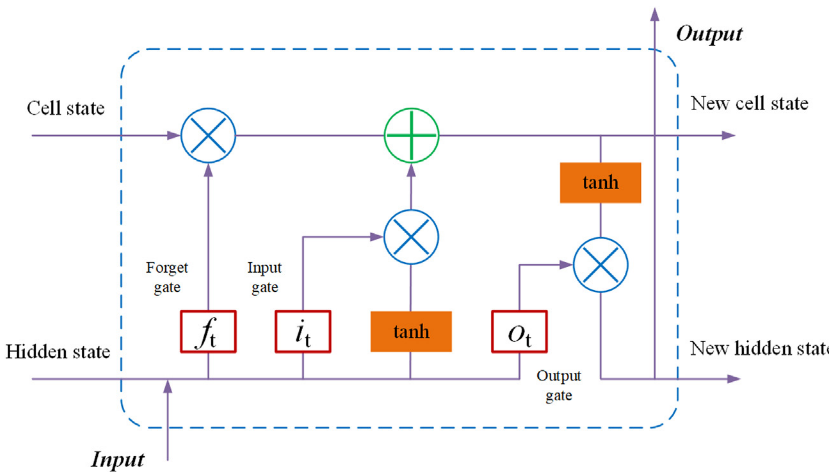

Fig. 9. The basic structure of the LSTM network.

so it is quite important and sensitive for SVM to select the parameters and kernel functions. The structure selection of neural networks and local minimum point problems can be avoided. However, SVM is difficult to train large-scale data and cannot efficiently support multiple classifications.

With the application of fuzzy rules, the FLM has merits in adaptability and interaction. However, the fuzzy process of data may compromise the control accuracy and dynamic quality of the model. Because of less systematization, it also makes the fuzzy control of complex systems difficult to work.

The ELMs have the advantages of relatively simple structure, less parameter adjustment, and lower computational complexity. The characters of ELM are appropriate for short-term air pollution prediction. However, the random initialization of input weight and offset of hidden elements in ELM is still controversial. For a higher level in various applications of ELM, the reasonable structural design and research on the generation type of hidden nodes are essential.

Deep learning forecasting models have developed rapidly that improvements over traditional visual and speech recognition have been achieved by continuous updating. Although deep learning forecasting models have obvious merits like great learning ability and adaptability, the problems of complex verification and training speed cannot be ignored.

3. Data processing for modeling strategies

From the analysis of intelligent models, it could conclude that each model has its advantages and disadvantages. Due to the complexity of atmospheric pollutant levels, the influence of multiple factors, and the complex changing trends, it is very difficult to utilize a simple predictor for effective forecasting. Taking advantage of various intelligent predictors, the predictive information of different models could be combined to form a hybrid model to effectively improve the fitting ability and increase the accuracy of prediction. In this section, the classification and introduction of components in air quality forecasting models from the published literature will be mainly described and discussed. including the commonly-used data processing methods: data decomposition and feature extraction.

3.1. Feature extraction

In air pollution forecasting, the principal components analysis (PCA) is a very common linear feature extraction method of dimension reduction. Finding new orthogonal mappings is to achieve the goal that the variance of the mapping points of the data in the direction is as large as possible after transformation, thus containing a larger amount of information. That can also be regarded as base transformation, which eliminates the new base direction with less information and implements dimensionality reduction.

The application of PCA in air quality forecasting is usually in hybrid models. Mishra and Goyal combined PCA with ANN in the forecasting of time series data fo hourly NO_2 concentrations, in which PCA was used to select the important input variables. Then the variables would be taken as input to conduct prediction by ANN [54]. Sun et al. proposed the PCA to extract important information and to reduce the dimension of original variables as the first step in the hybrid process to predict $PM_{2.5}$ pollution [77]. Lu et al. proposed as the PCA-RBF network to predict the RSP, NOx, and NO_2 concentrations of hourly time series, in which the results of the comparison between models demonstrate that PCA-RBF can achieve higher accuracy and faster training speed than the single models [78]. Part of the hybrid models is listed in Table 6. It can be concluded that the prediction of hybrid approaches outperforms their componential methods.

3.2. Data decomposition

Nevertheless, the precision of intelligent models can be further improved by using data pre-processing methods. In the reviewed literature, the current trend is that researchers are taking advantage of various hybrid models, which mostly start with data decomposition methods that simplify the sequence of complex components. The framework, as shown in Fig. 6, is normally the decomposition and integration process. It uses decomposition methods to separate the original input data of time series into several subseries and then generates relatively independent forecasting results by establishing individual forecasting models, which greatly improved the accuracy of air quality prediction. After combining all the forecasting results, the outcomes can be obtained. With the literature survey, the main decomposition strategies applied in air quality prediction are displayed as follows (see Figs. 10 and 11).

3.2.1. Wavelet decomposition

The wavelet transform (WT) is an excellent transform analysis method of information science and technology. Based on Fourier transform, it is appropriate for analysis and process of signal time–frequency, in which the signal is separated into different frequency bands and extract information from the signal to adapt to the requirements of the analysis.

Table 5
Partial summary of intelligent predictive models for air quality.

Categories	Advantages	Disadvantages	Data resolution	Main air pollutants	Partial used algorithms and references
Artificial neural network	• Good learning and building abilities to solve nonlinear complex problems • high accuracy of classification	• Many hyperparameters to be adjusted that could affect the forecasting results • The learning process cannot be observed • Possibility of extending the learning time	Hourly, daily	$PM_{2.5}$, PM_{10}, O_3, SO_2, CO_2, NO_2 and CO	BPNN [39,94,95], RBFNN [40,96], GRNN [41,99], WNN [42,43], ENN [103–105]
Support vector machine	• Using kernel functions to solve the dimension problem • Avoiding structure selection of NN and local minimum point problem	• Difficult for training large-scale data • Sensitive to the selection of parameters and kernel functions	Hourly, daily	$PM_{2.5}$, PM_{10}, O_3, SO_2, NO, NO_2 and CO	SVM [109–111], SVR [106,108], LSSVM [44]
Fuzzy logic model	• Less dependence on accurate mathematical models • Good fault tolerance and robustness • Easy to connect with human–machine interface	• The fuzzy process may reduce control accuracy and dynamic quality	Hourly, daily	$PM_{2.5}$, O_3, SO_2, CO, and NO_2	FTS [115,116], ANFIS [45,46,118,119]
Extreme learning machine	• Suitable for nonlinear activation functions • Rapid convergence and faster learning speed	• The random initialization is controversial	Hourly, daily	AQI, NO, NO_2, O_3, PM_{10} and $PM_{2.5}$	ELM [122]; OSELM [125,126]
Deep learning	• Good transfer learning property and adaptability • Great learning ability	• Complex verification of model correctness • Insufficient interpretability of the middle layers • Slow training speed	Hourly, daily	$PM_{2.5}$, PM_{10}, CO, NO_2, NO, O_3, SO_2 and NH_3	CNN [48,134,135], DBN [47,131,141], LSTM [49,129,130,138], GRU [139,140]

Table 6
Part of simple hybrid models in air quality forecasting.

Hybrid models	Pollutants	Cities	Data resolution	Contrastive models	Degree of improvement[a]
PCA-ANN [54]	NO_2	Taj Mahal, Agra	Hourly	MLP	31.88% (CC)
PCA-RBF [78]	RSP, NOx, NO_2	Hong Kong	Hourly	RBF	RSP: 31.49% (MAE), 31.49% (RMSE) NOx: 34.24% (MAE), 34.23% (RMSE) NO_2: 13.18% (MAE), 17.46% (RMSE)
PCA-CS-LSSVM[77]	$PM_{2.5}$	Baoding	Hourly	GRNN, LSSVM	Fuyong: 5.23% (MAE), 3.43% (RMSE) Longhua: 2.14% (MAE), 2.15% (RMSE)

[a]The improvement degree is from the error indexes of the proposed model and the best contrastive model.

Cheng et al. proposed hybrid models for forecasting $PM_{2.5}$. Based on WT the models decompose the original data, then combine with simple predictors like ANN and SVM to forecast data for the series [79]. Finally, the wavelet will be used to reconstruct the forecasting results of series to obtain the outcomes. The improvements of MAE, RMSE, and R^2 in hybrid forecasting models are visible in the prediction of different cities when compared with traditional ANN and SVM. In practical time-series prediction, the discrete wavelet transform (DWT) is processed in computational operation instead of continuous WT and used by many scholars [18,56,142]. A decomposition of input data can be divided into parts of different frequencies. The parts of low-frequency A_1, A_2, $\ldots A_n$ are approximations of input data, and the parts of high-frequency D_1, D_2, $\ldots D_n$ is the specifics of input data. However, the DWT has a problem with the input signal that the stability of the coefficients in the transform may be affected by the analysis area in original data. To overcome this problem, an improved version of DWT, the maximum overlap discrete wavelet transform (MODWT), also has been applied in air quality prediction [80].

In the WT process, only the low-frequency components of each layer will be decomposed. So in many cases, the useful information of time series in the high-frequency parts may be ignored. Compared with the WT method, the wavelet packet decomposition (WPD) can be used to decompose the appropriate coefficients and the detailed coefficients for increasing accuracy [81]. The comparison between WT and WPD in the process of the decomposing structure is shown in Fig. 7 as the simplified binary tree with three layers. Liu et al. applied a WPD-GBRT-LPBoost-MLP-DPMM hybrid model in $PM_{2.5}$ prediction [82]. In the combination process, the gradient boost regression tree (GBRT) is used to select the most important part of the original feature sets. The linear programming boosting (LPBoost) algorithm is an optimization algorithm for the weight of the predictor and the MLP could conduct $PM_{2.5}$ deterministic prediction while the Dirichlet process mixture model (DPMM) could obtain $PM_{2.5}$ probabilistic prediction. This model was tested with real-time pollutant

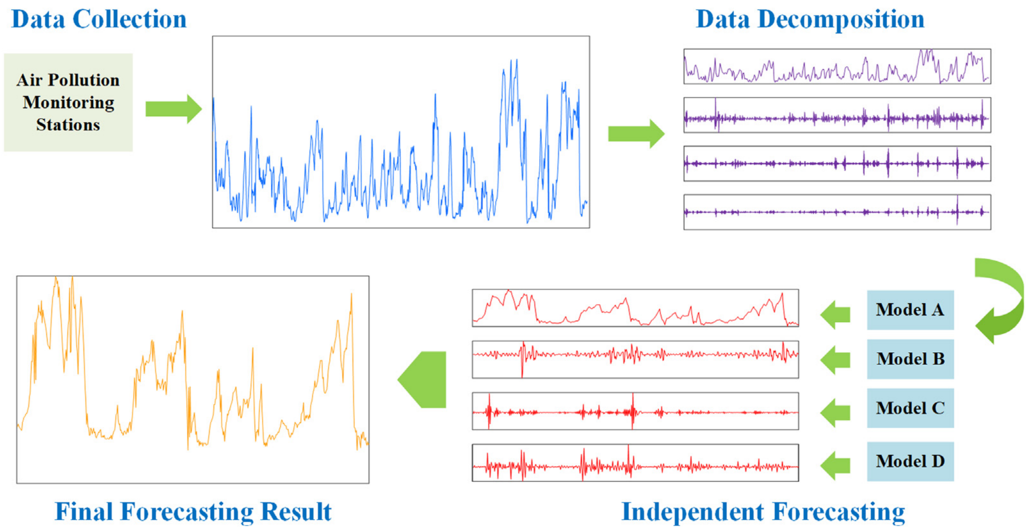

Fig. 10. The decomposition and integration process in air quality forecasting.

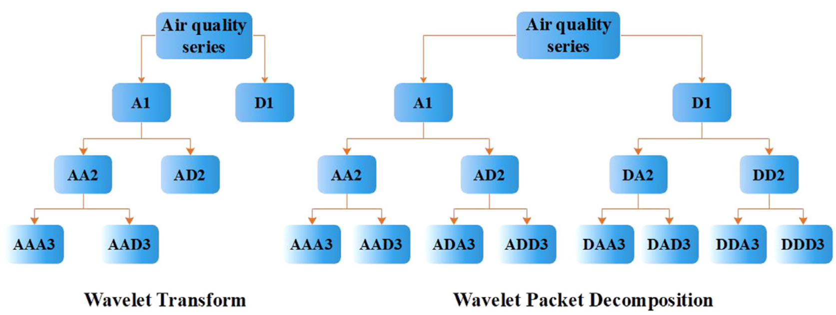

Fig. 11. The comparison between three-layer WT and WPD of decomposing structure.

data and compared with other algorithms and hybrid models, such as MLP, ARMA, WPD-MLP, WPD-LPBoost-MLP, WPD-GBRT-MLP, WPD-GBRT-LPBoost-MLP for the deterministic prediction and ARMA, ARMA-GARCH, WPD-GBRT-LPBoost-MLP-Gaussia for the probabilistic prediction. The results in Table 7 show that the proposed model has the best forecasting accuracy.

3.2.2. EMD decomposition

Compared to the choosing of wavelet functions and the setting of decomposition levels by wavelet decomposition in prediction, the empirical mode decomposition (EMD) separates the signals by the characteristics of the data without pre-setting of basic functions [143]. Therefore, the complex signals will be decomposed adaptively into a finite number of the intrinsic mode functions (IMFs), containing different local time-scales characters of original signals.

Zhu et al. used hybrid models to predict the pollution levels in Xingtai. Based on EMD, IMFs with high frequency will be removed and the remain IMFs will be used as input in LS-SVR and seasonal ARIMA model for prediction (EMD-SVR-hybrid) [83]. Another model is EMD-IMFs-hybrid, in which the IMFs will be separately modeled and predicted.

Ensemble empirical mode decomposition (EEMD) is an improved EMD that appends white noise into the signal to be decomposed and takes advantage of the uniform distribution in the white noise spectrum so that the white noise background can distribute in the whole time–frequency space by the application. The noises will counteract each other after multiple averaging and the results of the calculation could be regarded as the outcome. In air quality prediction, EEMD could help to determine different distinct information scales in the original data and avoid mode mixing effectively [84,85]. Bai et al. proposed an EEMD-LSTM model for $PM_{2.5}$ concentration forecasting and Niu et al. proposed EEMD-PSR-LSSVM for day-ahead $PM_{2.5}$ concentration forecasting, in which the results were satisfied compared to the separate single models.

By adding noise for analysis in EEMD, it may also bring a certain degree of damage to the original signal and the residual noise cannot be completely neutralized, which raises the computational complexities and weakens the efficiency. To solve that problem, the complementary ensemble empirical mode decomposition (CEEMD) is utilized to improve the preprocessing. Both the positive and negative white noise with the same amplitude have been appended in pairs so that the redundant noise is largely eliminated when the signal is reconstructed. CEEMD has been used in papers to decompose air pollutants data in hybrid models and reach better prediction accuracy than EMD and EEMD [86,87], as shown in the following Table 8, Niu et al. proposed the model for short-term PM2.5 concentration forecasting, in which the SVR optimized by grey wolf optimization (GWO) algorithm is the predictor [144].

The complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) is the upgraded version of CEEMD. The main character of CEEMDAN is the addition of adaptive noise comparing with EEMD, but it still has problems that the modes possess some residual noise and several spurious modes stand in the early decomposing phase by inadequate parameters. So the improved CEEMDAN (ICEEMDAN) has been introduced and applied to decompose input data of air pollutant levels [145,146]. Sharma et al. applied the hybrid models for hourly prediction in

Table 7
The hourly deterministic forecasting results of the models in the 1-step [82].

Models	MAE	MAPE	RMSE	IA
MLP	15.4043	24.0740	20.9419	0.8784
WPD-MLP	16.1768	26.5835	21.6145	0.8704
WPD-LPBoost-MLP	12.1657	18.6309	16.2762	0.9265
WPD-GBRT-MLP	13.6102	21.4844	18.3437	0.9067
WPD-GBRT-LPBoost-MLP	10.4672	16.7833	13.5670	0.9490

Table 8
Comparison of the forecasting results using different models [144].

Models	City 1			City 2		
	MSE	MAPE	MAE	MSE	MAPE	MAE
SVR-GWO	0.2196	23.1163	5.5300	0.1303	27.1406	9.0079
EMD-SVR-GWO	0.1233	19.7902	4.2689	0.1444	27.4051	8.7655
EEMD-SVR-GWO	0.0648	14.7810	3.1177	0.0473	16.3081	5.3451
CEEMD-SVR-GWO	0.0412	10.6033	2.4103	0.0186	10.4454	3.1741

Table 9
Improving percentages of comparison groups in the 1-step prediction [88].

The proposed model vs.	PMSE(%)	PMAPE (%)	PRMSE(%)
AMI-H/LSAE-DESN	80.047	78.654	81.040
VMD-H/LSAE-DESN	82.343	80.975	82.622
VMD-AMI-DESN	61.004	60.578	63.517

Table 10
The percent of the application frequency of the decomposition methods.

Decomposition type	Frequency	Percent
WT	8	30.77%
EMD	9	34.62%
VMD	2	7.69%
EWT	3	11.54%
Hybrid	4	15.38%
Total articles	26	100%

Australia combining ICEEMDAN with different predictors [147]. In hybrid predictive models, the scholars compared and chose with EMD, EEMD, CEEMDAN, and ICEEMDAN according to their practical needs and prediction effects. The improving process of the EMD and its variations are presented in Fig. 8.

3.2.3. Other single decomposition

A prevalent and effective decomposition method is the variational mode decomposition (VMD) [148]. Compared to EMD decomposition, VMD has excellent noise resistance, better decomposing performance, and stability and can be also utilized for feature extraction and fault diagnosis. In the practical air quality forecasting process, VMD can perform with other optimization methods to achieve precise separation results [58,149]. Xu et al. constructed the VMD-AMI-H/LSAE-DESN model and conducted a comparative analysis, and the results in Table 9 showed each component also contributed to the performance of the model [88].

The empirical wavelet transform (EWT) is a self-adaptive decomposition approach coupling the advantages of EMD and wavelet transform [150], which can autonomously divide the Fourier spectrum and separate the different modes and construct adaptive band-pass filters to decompose the data and extract different useful components. Due to its good adaptiveness and improvement for the learning ability, some references trend to take EWT as an important part of complex models to decompose the air pollutants series [151,152]. Liu and Chen constructed the HI-EWT-NNA-WRELM-IEWT hybrid model for the forecasting of $PM_{2.5}$ [128]. The processing steps of this hybrid model are as follows: the first step used the Hampel identifier (HI) to correct the outlier correction of the original data, then EWT decomposes the data into several subseries. WRELM conducted multi-step forecasting for each subsequence as the third step with and optimization of the neural network algorithm (NNA). At last, the IEWT reconstructed the subseries to generate the final prediction results. Moreover, the experimental results show that the forecasting accuracy of this model is much higher than other benchmark models, For example, the improvements of the #1 series are 66.70% (MAPE), 66.70% (MAE), 68.00% (RMSE).

3.2.4. Hybrid decomposition

To achieve better performance of the decomposition in the prediction process, many researchers developed hybrid decomposition methods, as secondary-decomposition (SD). Normally, through two-layer decomposition, the hybrid decomposition algorithms can extract more information from the original data series than simple decomposition so that the total prediction model can reach higher accuracy and achieve better performance. To choose the appropriate decomposed components is an essential impact in hybrid decomposition. Gan et al. proposed a hybrid model combining WPD and CEEMD as decomposition [153]. The WPD decomposes the original data series into components of both low frequency and high frequency. The CEEMD is used as the second-layer decomposition technique to continually decompose the high-frequency components. In this way, the hybrid models of a two-layer structure can be effective solving the decomposition problem.

Other scholars also proposed many hybrid decomposition methods. By Wang et al. the CEEMD is used as the first layer to decompose the air quality index series into IMFs of different frequencies and the VMD will follow to decompose the high-frequency IMFs into variational modes (VMs) to eliminate unsatisfied forecasting results [154]. Wang et al. also employed VMD as the second-layer decomposition, but WT as the first layer is set to disassemble the original $PM_{2.5}$ concentration series [155]. The singular spectrum analysis (SSA) was applied to optimize the components and to remove redundant noise in the original signal [89]. In the reference, simple EEMD decomposition and none decomposition in benchmark models were used for comparison experiments. It is known from the experimental results indicated that the hybrid decomposition model has better forecasting accuracy than others do (see Fig. 12).

Table 10 present the application frequency of the decomposition methods. According to the tables, the current trends could be found that WT and EMD are used more frequently than others. Moreover, the hybrid strategies have been taken into the consideration for studies in recent years. The above hybrid decomposition strategies have been proven effective in the aspect of accuracy and stability. However, with the increase of subseries and layers of model structure, the computation burden

Table 11
A summary of decomposition approaches in hybrid models of air quality forecasting.

Category	Subcategories	Specialties	Applied approaches and references
Decomposition	Wavelet decomposition	• Good localization covering the time domain and frequency domain for complete descriptions • Multiresolution features are suitable for time series analysis • Rely on the decomposition structure formed by wavelet function and decomposition levels	WT [79], DWT [18,56,142,156], MODWT [80], WPD [53,82]
	EMD decomposition	• Suitable for analyzing nonlinear and non-stationary signal sequences • Drawbacks of endpoint and mode mixing in EMD • Improved versions have been developed to enhance the prediction ability	EMD [83], EEMD [84], CEEMD [57,86,144], ICEEMDAN [145–147]
	Other simple decomposition	• Excellent noise resistance, decomposing performance, and stability • EWT combines the advantages of EMD and WT	VMD [58,149,154,155], EWT [128,151,152]
	Hybrid decomposition	• Combine with different methods as secondary-decomposition • Achieve higher accuracy and better forecasting performance • May lead to more computation time	WPD + CEEMD [153], CEEMD + VMD [154], WT + VMD [155],VMD + CEEMD [149], EEMD + SSA [89]

may also remarkably increase. So, it is necessary to compromise between calculation accuracy and running time in the application of secondary decomposition approaches. Table 11 concludes the above methods with the subcategories, specialties, and applied approaches in references.

4. Auxiliary method I: Metaheuristic optimization

Although the simple predictive models and data processing methods can greatly enhance the abilities of intelligent models, the forecasting performance can be further improved in the structure and hyperparameters. The metaheuristic optimization algorithms aim at further optimization of original data. The whole process of input data is a typical feature selection, in which metaheuristic algorithms are used to generate input data and the data will be used as training samples for forecasting results.

4.1. Heuristic and metaheuristic algorithm

The heuristic algorithms are a kind of intuitively or empirically constructed algorithms that give a feasible solution for specific problems to be solved. Metaheuristic algorithms are the improvement of heuristic algorithms, which is the combination of random algorithms and local search algorithms as general heuristic strategies. Many metaheuristic algorithms simulate the biological or physical phenomena of nature into mathematical structures to solve problems [157]. Some metaheuristic algorithms used in air quality forecasting models are demonstrated in Fig. 13.

Heuristic algorithms are the methods that depend on the problem. Therefore, they usually adapt to the current problem and try to take full advantage of the particularity of this problem. However, they usually fall into a local optimal state, and therefore usually cannot obtain a globally optimal solution. Although the metaheuristic algorithms are different in the mechanism, they are problem-independent methods. They are not so greedy that enable them to explore the solution space more thoroughly and repeat until the convergence criterion is good enough to obtain the optimal solutions [158]. It is still necessary to make some adjustments to its internal parameters especially the hyperparameters, such as number of iterations, number of hidden layers, number of neurons in each layer, learning rate, and so on [86].

The genetic algorithm (GA) is a method to search optimal solutions by simulating natural evolutionary selection, aiming at all individuals of the population. The particle swarm optimization (PSO) is also a population-based optimization method developed by simulating the collective behavior of birds to obtain the optimal solution [59,159]. In Ref. [160], similar to GA, differential evolution algorithm (DE) is used for the global search of optimization in hybrid models. Other metaheuristic algorithms, such as cuckoo search (CS) [62,161], bat algorithm (BA) [162], and grey wolf optimizer (GWO) [90], also have been commonly utilized to enhance the performance of prediction. The combination of metaheuristic algorithms can further improve the parameters based on the complexity of hybrid models, such as Particle swarm optimization and gravitational search algorithm (PSOGSA) and Modified cuckoo search and differential evolution algorithm (MCSDE).

4.2. Classification by optimized objects

According to the optimized objects of the application in air quality forecasting models, the metaheuristic optimization algorithms can be further classified into two kinds: (1) optimization methods of combination weights and (2) optimization methods of parameters in predictors.

4.2.1. Optimization methods of combination weights

The application of hybrid models is to take advantage of every single forecasting model while avoiding their weaknesses. One method is to obtain the combination of the weight coefficient of every single model. The metaheuristic optimization algorithm is applied to search for optimal combination weights of the individual models, so the optimum forecasted results of air quality forecasting can be achieved. The flowchart of the process is shown in Fig. 14.

Yang et al. proposed a combination forecasting system of data decomposition, metaheuristic optimization, and individual forecasting models [161]. The base predictors of the proposed model include three predictors: BPNN, ELM, and Double exponential smoothing (DES). The cuckoo search (CS) algorithm used in the paper can determine the combination weights for model aggregation and the results show that the CS optimization can further increase the forecasting accuracy and stability of the model. The analysis of sensitivity has also been conducted to analyze the impact of hyperparameters on the results and their optimal configuration, including the population size, maximum number of iterations, and discovery rate.

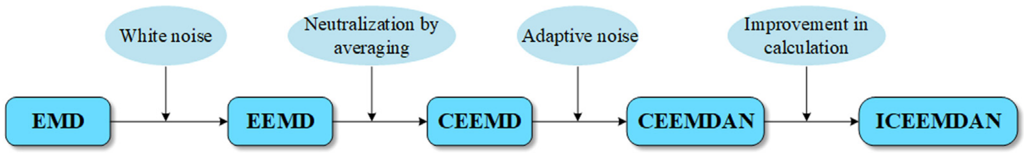

Fig. 12. The development process of the EMD and improved versions.

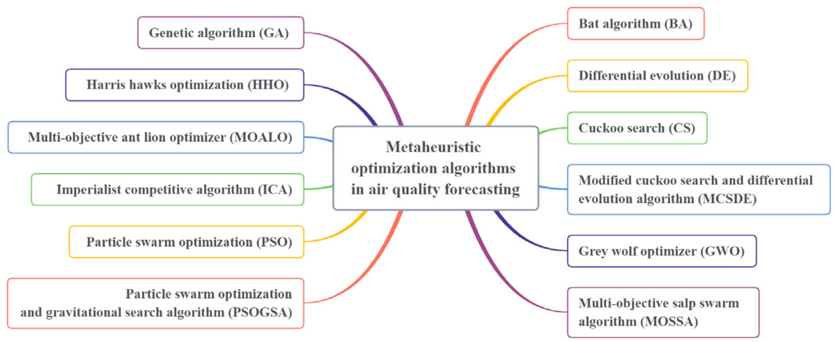

Fig. 13. The mechanisms of utilized metaheuristic algorithms in air quality forecasting models.

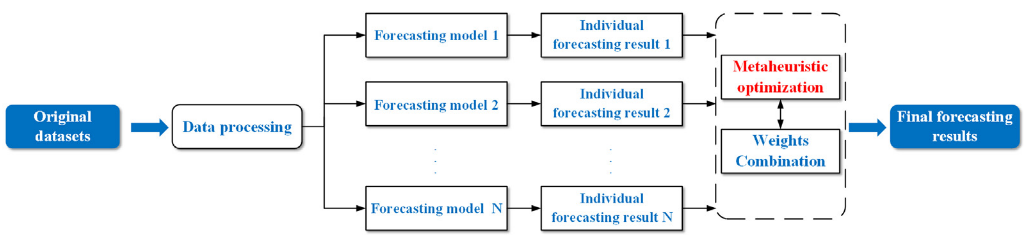

Fig. 14. The flowchart of the process for optimizing combination weights.

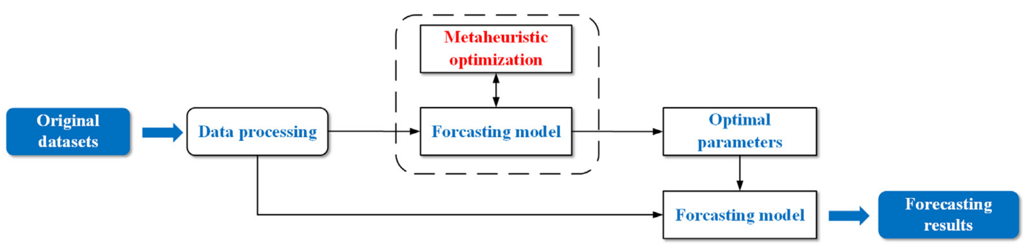

Fig. 15. The flowchart of the process for optimizing model parameters.

4.2.2. Optimization methods of parameters in predictors

The metaheuristic optimization can also be applied in the combination with other predictors, whose parameters are optimized with metaheuristic methods, to make the predictors appropriate for the input datasets. The input dataset is usually divided into three parts: the training set, the validation set, and the test set. The parameters with the lowest error by metaheuristic optimization on the validation set are chosen as optimal parameters. The flowchart of the process is shown in Fig. 15.

Recently in the studies of air quality forecasting, most of the optimal models have been proposed with the theory of parameter combination. The initial weight and threshold are usually used as parameters for optimization. Du et al. proposed the multi-objective Harris hawks optimization (MOHHO) algorithm to optimize the weight and threshold of ELM [163]. The multi-objective ant lion optimizer (MOALO) is an improved version of the ant lion optimizer (ALO) to obtain a more accurate global optimal solution through many iterative calculations. The MOALO was used to build the SSA-EEMD-MOALO-$L_{2,1}$RFELM forecasting model, of which the $L_{2,1}$RFELM is an improved version of optimized for $PM_{2.5}$ forecasting by MOALO [89]. It can be concluded from the results in Table 12 that MOALO contributed to the increase of the hybrid model.

To improve the application ability of predictors, metaheuristic optimization algorithms have been also utilized to optimize other important parameters in the hybrid models. Wu et al. used the Bat algorithm (BA) to optimize the two parameters, namely the penalty coefficient c and the kernel parameter σ^2 in the LSSVM model for better adaptability and accuracy [162]. By Xing et al. the modified grey wolf optimization (MGWO) algorithm is proposed, which has good nonlinear convergence and computational robustness, to improve the DBN structure parameters [141]. Compared to the other models like BPNN, SVM, and random forest, the DBN model reached better results with the lowest MAE (17.604 $\mu g/m^3$) and MSE (410.266 $\mu g^2/m^6$). Sun et al. proposed cuckoo search (CS) in PCA-CS-LSSVM to optimize the two parameters of LSSVM and predict $PM_{2.5}$ pollution, and PCA extracts important information and reduces the dimension of original variables as the first step in the hybrid process [77]. The LSSVM improved

Table 12
Comparison of the forecasting results of $PM_{2.5}$ of different models [89].

Models	City 1			City 2		
	MAPE	MAE	RMSE	MAPE	MAE	RMSE
$L_{2,1}$RFELM	9.8328	2.9938	4.4955	5.6780	1.1164	1.6233
MOALO-$L_{2,1}$RFELM	9.6955	2.9843	4.4070	5.5808	1.1004	1.6095
EEMD-MOALO-$L_{2,1}$RFELM	3.7076	1.0804	1.6058	2.7503	0.5191	0.6755
SSA-EEMD-MOALO- $L_{2,1}$RFELM	3.6465	1.0598	1.4796	2.6965	0.5028	0.6698

Table 13
The optimized parameters in predictors in air quality forecasting.

Optimized parameters	Predictors
Initial weight and threshold	BPNN [62,89,146,160,164], ELM [163,165], $L_{2.1}$RFELM [89], ENN cite90,149
Penalty coefficient and kernel parameter	SVM [59,159], SVR [57,86,144], LSSVM [77,162,166]
Others	Structure parameters of DBN [141]

by CS seems to be very attractive and shows a great degree of improvement in the results of 23.53% in MAE, 26.76% in MAPE, and 33.47% in RMSE. The optimized parameters in predictors are listed in Table 13.

5. Auxiliary method II: Ensemble learning

Ensemble learning is an important method in intelligent models for air quality forecasting, which builds multiple individual models by certain algorithms and then combines them with a certain strategy to obtain a powerful model to complete the task and improve forecast results. To achieve this goal, there are currently three popular methods applied in air quality prediction: Boosting, Bagging, and Stacking.

5.1. Bagging

The principle of Bagging is random sampling, which is to collect a fixed number of samples in our training set. These samples will be used in parallel training in weak models that will be integrated into a powerful model through specific strategies. Random forest (RF) is an evolutionary version of the Bagging algorithm, which can handle linear and nonlinear problems without extra concern of the independent or the dependent variables. Formed by a designed number of binary decision trees, RFs use sample subsets independently collected from single trees to aggregate and optimize results of final predictions for air pollution [167].

In the study to predict the air pollutant levels at the Region of Murcia (Spain), using the Bagging, Random Committee, Random Forest, a decision tree, and an instance-based technique were utilized to conduct the research with datasets of two years and compare the results through the indexes like MAE and RMSE [168]. Dotse et al. proposed a hybrid model GA-RF-BPNN to predict PM10 levels' exceedances [64]. The model combines GA and RF for preprocessing to select input variables before data training in BPNN and produces better results (Table 14) in developing a real-time forecasting system. Philibert et al. also adopted Random Forest to predict a greenhouse gas N_2O emission by using global meteorological and crop data [169]. For a precise validation, the results were compared to the regression model and the simple non-linear model. The best result in the RMSE of the improved model can reach 11.34% than the compared model.

Table 14
Comparison of the forecasting results using different models [64].

Models	Station 1			Station 2		
	CC	IA	RMSE	CC	IA	RMSE
BPNN	0.9223	0.9563	4.0057	0.9072	0.9468	5.5392
GA-BPNN	0.9266	0.9612	3.9080	0.9113	0.9490	5.4059
GA-RF-BPNN	0.9502	0.9727	3.2942	0.9397	0.9677	4.5346

5.2. Boosting

Boosting refers to converting multiple weak learners into a strong predictive model through a set of algorithms. Adaptive Boosting (AdaBoost) is a commonly used iterative algorithm of boosting. After each iteration, a new learner is generated and the samples will be predicted. Based on the training performance, the samples with forecasting error could be assigned a higher weight that the higher the weight is, the greater the proportion of the next iteration is. Liu et al. developed the WPD-PSO-BP-Adaboost model in $PM_{2.5}$ concentration prediction, in which Adaboost trains the weak learner made by the PSO-BP model [53]. Compared with the WPD-PSO-BP model, the MAPE was reduced by 18.79%, 20.89%, and 6.28% by 1-step, 2-step, and 3-step prediction and that means the proposed hybrid model has a significant increase in accuracy while ensuring stability by adding Adaboost. Liu and Chen proposed a novel hybrid model for prediction of air quality index [170], in which Adaboost.MRT, as an upgraded version of Adaboost, was used to improve the performance of base predictor ORELM in each subseries. Then the results from all subseries are combined to achieve the final forecasting result. The structure of the Boosting strategy is presented in Fig. 16.

5.3. Stacking

Stacking is an algorithm different from bagging and boosting. The main characteristic is to integrate the results of previous algorithms, that is, a framework based on the weak learners, which combines the previous weak learners with a machine learning algorithm.

Sun and Li proposed a Stacking-driven ensemble model to obtain accurate forecasting results of base predictors and the results are regarded as the input of meta-model SVR, to get the final prediction results [171]. Zhai and Chen also designed a stacking ensemble model to predict $PM_{2.5}$ in Beijing, China [65]. Individual models, including least absolute shrinkage and selection operator (LASSO), Adaboost, extreme gradient boosting (XGBoost), and GA-MLP, are to conduct data training by layers and the output will be integrated into SVR by the stacked process. The results show that the ensemble model has better performance, comparing to the best results of simple models in the test set including IA (1.04%), MAE (18.98%), RMSE (6.91%). The structure of the Stacking strategy is demonstrated in Fig. 17. Table 15 present the application frequency of the ensemble learning methods. According to the tables, the current trends could be found that the frequency of the Bagging application is equal to the sum of Boosting and Stacking. In recent articles, the scholars trend to choose the Boosting and Stacking for the hybrid strategies [53,65,170,171].

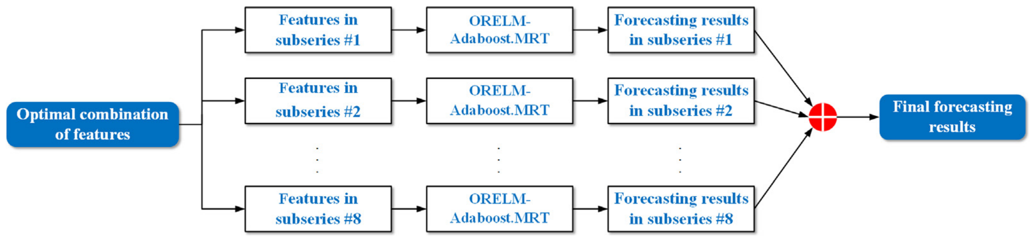

Fig. 16. The proposed structure of the Boosting strategy in Ref. [170].

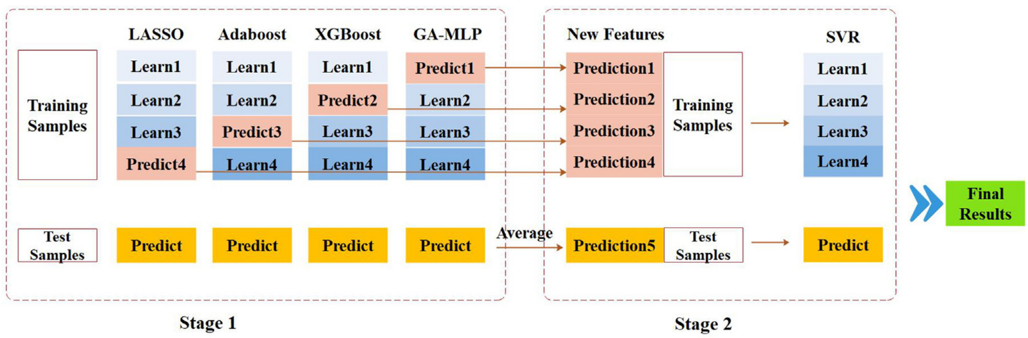

Fig. 17. The proposed structure of the Stacking strategy in Ref. [65].

Table 15
The percent of the best results of ensemble learning.

Method type	Frequency	Percent
Bagging	4	50%
Boosting	2	25%
Stacking	2	25%
Total articles	8	100%

6. Spatiotemporal analysis for modeling strategies

Due to the complexity of air pollution, the prediction of pollutant concentrations is easily affected by many factors at specific times and locations. Previous studies for air pollutant concentrations prediction rely on the time series data of air quality index (AQI) issued regularly by air quality monitoring stations from specific cities or regions. Because air pollutants have the property of regional diffusion, air quality research based on individual cities has great limitations, which could be affected by geography process, atmospheric situation, or other factors.

Therefore, it is necessary to analyze the internal correlation of air quality between the target cities and the surrounding environment from a regional or national perspective, namely both temporal and spatial characteristics of air quality-related meteorological information should be taken into consideration. The most important impacts in spatiotemporal models of published papers are multiple data analysis, improved individual models, and spatiotemporal hybrid models, and the partial summary of the quoted papers in this review is introduced in Fig. 18.

6.1. Multiple data analysis

Multiple data from many stations in the target area offer a variety of choices and comparative possibilities for forecasting. Awad et al. used the data from 368 monitors of 12 years in the New England States (USA) to analyze the spatiotemporal interaction of ambient Black Carbon (BC) concentration [66]. With the great amount of data from 368 monitoring stations in 12 years, the experiment and analysis can be conducted that more repeated measurements and monitoring lead to completer obtaining temporal characteristics. A land-use regression model is improved by the application of a nu-SVR and a generalized additive model is applied to refit residuals from nu-SVR. The spatial analysis of transportation, topographical, and neighborhood characteristics also contributed to the prediction together with temporal terms. The model is tested in cold and warm seasons by comparison to actual data.

As for the neighborhood or distance analysis, Kurt and Oktay described three geographic forecasting models using ANN by the pollutant data from 10 different monitoring stations in Istanbul, as shown in Fig. 19 [67]. The single-site neighborhood model used the air pollution data from one or more nearby districts as extra input for the target district. However, the two-site neighborhood model took two neighboring districts as a whole target object instead of one. Different from the above two, the distance-based model averaged the air pollutant data with the weight of distances in three neighboring districts to calculate the combined influence to the center space, which also indicated the best performance than other models during the comparative analysis.

In order to better divide the research area and apply air pollutants data, a $0.1^{\circ} \times 0.1^{\circ}$ grid (98341 cells in all) across China was applied for the daily prediction of O_3 in Ref. [68]. The random forest was proposed to predict the spatiotemporal distributions O_3 in China by weather information from 1608 monitoring sites over the country in 2015 and the results would be further summarized and weighted by seasons, regions, and districts of great population. The meteorological variables account for 65% of the predictive accuracy for the predictive features. In this paper, anthropogenic emissions (NH3, Organic Carbon, CO, and NOx) present comparably lower importance than meteorology, and lower accuracy is presented for the area with few monitoring stations. Hence, the precision depends on the complexity of the network.

Wen et al. also utilized a method to efficiently extract spatiotemporal correlations and improve prediction ability [172]. The data were collected from 1233 air quality monitoring stations in two years period. Based on the k-nearest neighbor (KNN) algorithm, the adaptive k-nearest neighboring stations in single stations were selected to ensure the relatedness so that the CNN and

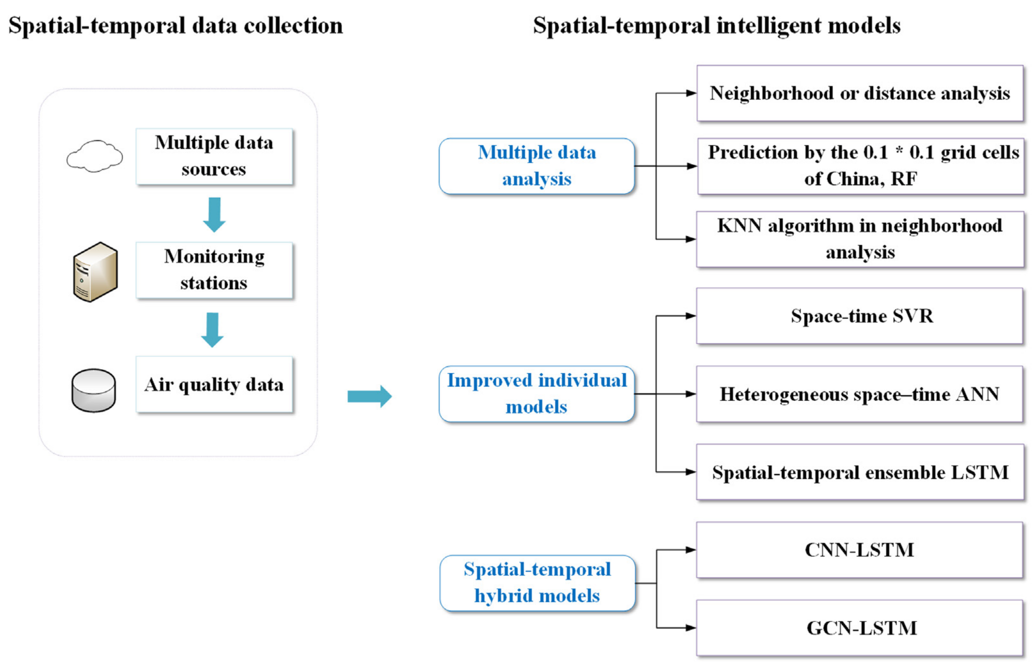

Fig. 18. A partial summary of spatiotemporal intelligent models in air quality forecasting.

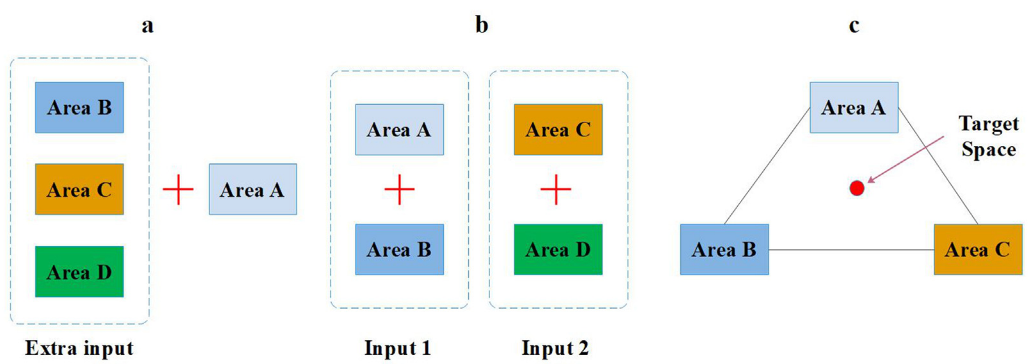

Fig. 19. A graphical summary of three geographic forecasting models. (a) Single-site neighborhood model (b) Two-site neighborhood model (c) Distance-based model [67].

LSTM models can be further utilized to extract spatiotemporal features.

The air pollution data can be collected by monitoring stations across the country, Fig. 20 shows the current locations of 1635 air quality monitoring stations in China, which are like the above studies and perform a better understanding of the geographic distribution and air quality. The sites can be randomly selected as the forecasting targets to test the effectiveness of the proposed model Ref. [68]. They are in Beijing, Shanghai, Chengdu, Guangzhou, Lasa, and Urumqi cities whose codes are 1004A, 1143A, 1352A, 1435A, 1460A, and 1493A.

6.2. Improved individual models

In recent research on air pollutants, the study of spatial correlation is a non-negligible direction of prediction. Starting with the methodology, many scholars offered novel frameworks of individual models based on spatial dependence. An extension from the model of non-spatial data to geospatial data was proposed in Ref. [69]. With the consideration of spatial factors, a space–time support vector regression (STSVR) model was developed that spatial clustering analysis and spatial autocorrelation variables were added into the structure, aiming at increasing spatial correlation and dependence. A heterogeneous space–time ANN was designed for the prediction of $PM_{2.5}$ that both space–time dependence and heterogeneity were added into the framework of feedback structure, resulting in better evaluation values than traditional ANN and hybrid model [91].

Wang and Song proposed a 3-step model with a deep spatiotemporal ensemble strategy (STELSTM) [173]. The first step is to establish an ensemble method based on weather patterns and partitioning strategy. Then the second step is exploring the spatial correlation of neighbor stations. At last, the LSTM combines the local and neighbor data to work as the temporal predictor. Ma et al. have set a Geo-layer into the LSTM neural network for comprehensive spatial analysis that the spatiotemporal correlation of relative monitoring stations could be integrated into the model [70]. After the selection of the Geo-layer, the information of 37 monitoring stations in the area, which have weaker correlations, will be removed. The Geo-LSTM model can interpolate the spatial distribution of air pollutants, which means both the spatial and temporal impacts of air pollutants will be inclusive at the same time. Furthermore, compared with other models, such as ANN, RNN, RF, SVR, ordinary LSTM, the proposed Geo-LSTM model has the lowest RMSE. The RMSE improvements to the ordinary LSTM reached over 42%.

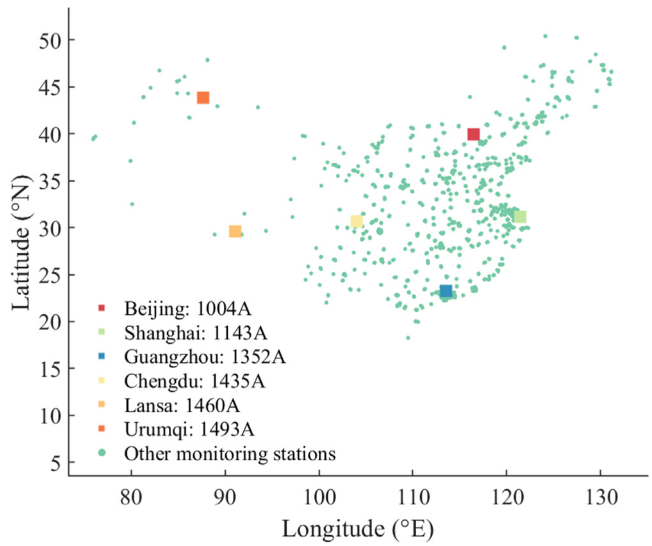

Fig. 20. Locations of air pollution monitoring stations.

Table 16
The percent of the best results of the ELM.

Model type	Frequency	Percent
Multiple data analysis	4	36.36%
Improved individual models	4	36.36%
Spatiotemporal hybrid models	3	27.28%
Total articles	11	100%

6.3. Spatiotemporal hybrid models

Besides data preprocessing and improved individual models, there are some other hybrid models in the application. Each part of the model is designed to conduct specific functions or aim at specific types of information. The general process is present in Fig. 15. Pak et al. used the CNN-LSTM hybrid model to forecast the daily $PM_{2.5}$ levels. After the analysis of spatiotemporal correlation by the mutual information (MI) estimator, the output spatiotemporal feature vector (STFV) was constructed [71]. Then the CNN started to extract the important inherent features from the input and the LSTM was used to handle the integration of feature time series passed CNN to represent the long-term historical process of input temporal data and form the final output of prediction.

Based on the ability of image recognition to spatial data processing and great performance in time-series data processing, some researchers directly combined CNN and LSTM as a hybrid model to conduct experiments of air quality forecasting. Qin et al. utilized CNN to extract spatial features and reduce redundancy from the dataset of Shanghai that contains data in three years as the first step, then LSTM can handle time-series information [72]. Compared with other models (BPNN, CNN, RNN, LSTM), results of experiments (epochs = 100) presented the proposed CNN-LSTM achieved the lowest RMSE (14.3) and highest correlation coefficient (0.97). There is also a similar model formed by LSTM and graph convolutional networks (GCN), which has better performance in the non-linear and spatial study than CNN [174]. GCN also extracted the spatial correlation of many stations and LSTM still worked on the temporal dependency for prediction. Table 16 present the application frequency of the spatiotemporal methods. According to the tables, all three methods currently have almost the same in the modeling application. In the future study, the choice of the intelligent spatiotemporal model may depend on the actual research requirements (see Fig. 21).

7. Discussion and recommendations for future work

A general analysis of data-driven methods in air quality forecasting has been demonstrated in this paper. The intelligent predictors, data processing methods, metaheuristic optimization, ensemble learning, and spatiotemporal analysis are important parts in forecasting strategies, while the mathematical structures of the models range from simple to complex structures. The shallow predictors and deep learning predictors have transformed from independent predictors to hybrid prediction combined with multiple popular techniques, aiming at better performance in forecasting. In this review, the current air pollution forecasting models are thoroughly summarized by the purpose, application, and characteristics of various methods. In this section, different modeling methods will be compared and evaluated, the predictive models in artificial intelligence and some possible future development trends are discussed and proposed.

7.1. The representative structure of predictive models and comparative analysis

To achieve high accuracy in the field of air quality forecasting, many predictive models are hybrid models based on intelligent algorithms, such as ANN, SVM, Fuzzy logic, ELM, and deep learning predictors. With the development of hybrid models, these predictors relate to more algorithms and methods, adapting to various forecasting requirements. In this paper, the data processing, metaheuristic optimization, ensemble learning and are integrated into the hybrid structures and Table 17 lists some representative predictive models.

The trend of the combination of different methods in air quality forecasting is obvious and the decomposition is the most used method among data processing methods because of the effective increase in the forecasting accuracy. Some models can reduce the forecasting error to 40%–60% [144,147] by different decomposition methods when compared to the models without decomposition algorithms. From the comparative analysis of the above sections, it indicated that the metaheuristic optimization, ensemble learning of auxiliary methods, and spatiotemporal models also contribute to the increase of the forecasting accuracy.

To further compare the different methods and select the efficient models from all methods to the above conclusions, by

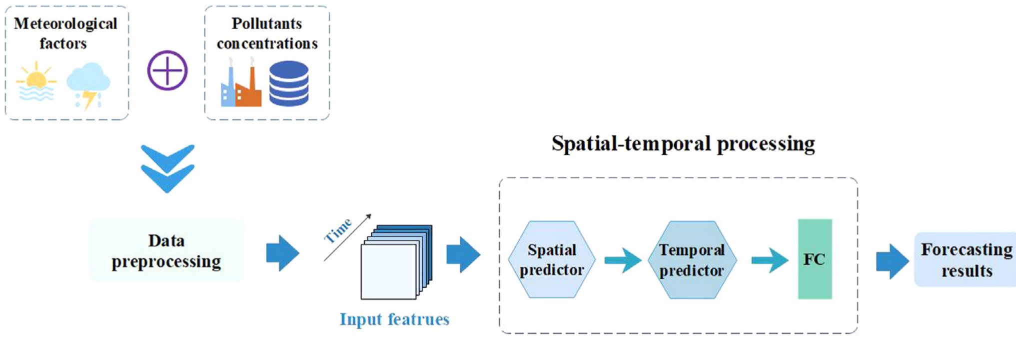

Fig. 21. The general framework of spatiotemporal hybrid modeling [72].

Table 17
Partial representative structures of hybrid models in air quality forecasting.

Model framework	Methods	Subjects	Error index	Reference
Decomposition + predictor	EEMD + LSTM	$PM_{2.5}$	19.604% (MAPE)	[85]
Hybrid decomposition + predictor	VMD + SE + LSTM	AQI	9.09% (MAPE)	[58]
Secondary decomposition + predictor + metaheuristic optimization	WPD + CEEMD + LSSVR + CPSOGSA	$PM_{2.5}$	Shenyang: 9.83% (MAPE) Chengdu: 8.41% (MAPE)	[153]
Decomposition + metaheuristic optimization + predictors	CEEMD + PSOGSA + SVR + GRNN	$PM_{2.5}$	Chongqing: 3.9374 (RMSE) Harbin: 4.0263 (RMSE) Jinan: 6.2995 (RMSE)	[86]
Metaheuristic optimization + ensemble learning + predictor	GA + RF + BPNN	PM_{10}	Brunei-Muara: 2.4032 (MAE) Temburong: 3.1072 (MAE) Belait: 7.5557 (MAE) Tutong: 8.2211 (MAE)	[64]
Metaheuristic optimization + predictor	PSO + ELM	CO_2	71.11% (RMSE), 73.73% (MAPE)	[165]
Decomposition + metaheuristic optimization + predictor + ensemble learning	WPD + PSO + BPNN + Adaboost	$PM_{2.5}$	Case 1(1-step): 9.04% (MAPE) Case 2(1-step): 6.69% (MAPE)	[53]
Feature extraction + metaheuristic optimization + predictor	PCA + CS + LSSVM	$PM_{2.5}$	12.56% (MAPE)	[175]
Spatial predictor + temporal predictor	CNN + LSTM	$PM_{2.5}$	14.3 (RMSE)	[72]

the experimental data in [153], the MAPEs of the three models WPD-LSSVR-CPSOGSA, CEEMD-LSSVR-CPSOGSA, and WPT-CEEMD-LSSVR-CPSOGSA were compared. It can be seen from Fig. 22 that the accuracy improvement capability of LSSVR-CPSOGSA is gradually improved when combined with the WPT, CEEMD, and hybrid decomposition.

According to the experimental data in [53], the hybrid strategies can improve the forecasting ability of the benchmark models. The comparison in BP, WPD-BP, WPD-PSO-BP, and WPD-PSO-BP-Adaboost was used as the benchmark models to achieve the MAPEs of 1-step prediction. It can be seen from Fig. 23 that the predictive models can be improved by a combination of the decomposition algorithms, the metaheuristic optimization, and the ensemble learning. The forecasting abilities of the benchmark models are also gradually increased. So, the conclusion can be drawn that the hybrid models with the abovementioned methods could reach the high level in air quality forecasting and the spatial analysis for the data and the predictors may be included in the future exploration.

The hybrid models can take advantage of their subordinate components to achieve the satisfied forecasting results with higher accuracy, better robustness, and fewer errors, which can be concluded by the analysis of comparative experiments in the review. However, the complex constructions of models also bring more costs in the computing process. There is no absolute perfect method, and the air quality cannot rely on a few strategies to handle a variety of prediction situations. The hybrid models of complex frameworks often need more running time because of the structure. The time efficiency of the model may decrease while the prediction accuracy has been improved so that they are not suitable for the high real-time forecasting demands. Therefore, the balance between modeling factors should be considered in further research.

7.2. Possible future development trends and challenges

Compared to the abovementioned models, other research aspects can also be future development trends of critical impacts in air pollution prediction, such as the data collection and data pre-processing. Through the survey of the references, there are still some challenges in the recent applications, the viewpoints are given as follows:

7.2.1. Choose suitable predictors for subseries forecasting

In addition to the traditional algorithms summarized in the review, the search and application for new forecasting algorithms should still be conducted. Besides the traditional ANN, SVM, FLM, ELM, etc., the improved version of them have been utilized in the hybrid models like OSELM [147] and WRELM [128], as well as the new predictor like DESN [88]. Moreover, how to choose a suitable predictor for various air pollutants forecasting processes may lead to a breakthrough in increasing prediction performance.

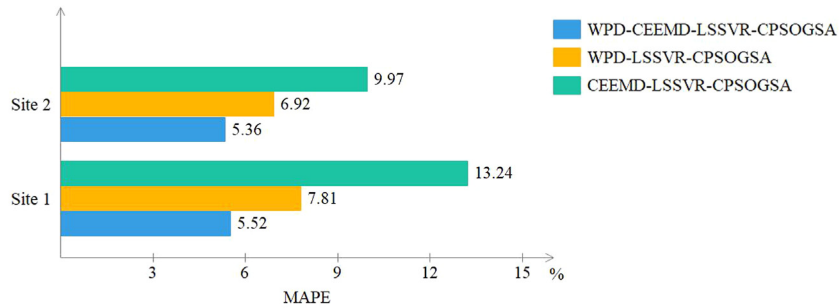

Fig. 22. The MAPEs of LSSVR-CPSOGSA by WPT, CEEMD, and hybrid decomposition [153].

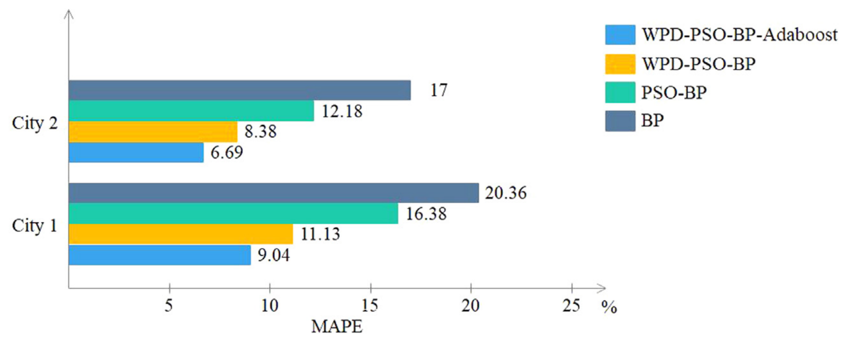

Fig. 23. The MAPEs comparative analysis in [53].

For example, the hybrid models with decomposition methods usually utilize the same predictor to forecast all subseries and the metaheuristic optimization methods aim to search for optimal combination weights of individual predictors in each subseries. But there is a lack of analysis for the diversity of subseries with different frequencies, which is still a problem worthy of research. With the consideration and choice of optimal predictors for subseries, the final forecasting results could be more precise.

7.2.2. The application of the IoT in the Smart City

Recently, the new concept of Smart City has brought possibilities to solve environmental and social problems and offer an improved model of life-based intelligent devices. The internet of things (IoT) devices can conduct smart communication to interact with each other and connect to the internet. With the strategies, all the individual units could be connected by intelligent systems to make the sustainable smart city [176]. And the Smart City system can handle a huge amount of real-time data from the city and give direction to the authorities to make better decisions [177]. The machine learning strategies and AI-based detection technique of smart air quality monitoring may lead to applications in the future studies of the IoT network [178].

7.2.3. The data collection challenges

Judging from the articles involved in the review, the various models in the review rely mostly on historical data from official publications, which are collected by national monitoring stations across the area. But it is still not enough in practical applications, the application of smart mobile pollution monitoring devices and low-cost air pollution sensors for real-time measurements have come up in recent study [179], including the PMS series sensors, MQ series sensors (especially MQ-135), Azimut, Smart Citizen Kit, NetAtmo and so on. The possible framework for the combination of the mobile data and station data could contribute to air quality prediction by the data resolution of time and the accurate spatial division and complete temporal analysis of the original data would further improve the scientific capacity and accuracy of the prediction, in which lots of advanced knowledge or problem-solved abilities are required. Besides, the massive knowledge of detection technique has great potential for further development.

7.2.4. Applications of distributed computing

With the increase of air quality data, the big data-driven modeling of a large sum of data has gradually gained more attention and been proved to be effective in improving the stability of hybrid model with sufficient information and features and increasing the training speed of the model by the distributed computing [88]. Currently, the Apache Spark framework was applied to predict big data series by some scholars [180], which is a distributed processing system commonly used in big data workloads with high-speed performance. By applying the cluster computing of the big data platform and storage system, such as the Hadoop Distributed File System (HDFS), the efficiency can be greatly increased. Combining the big-data and machine learning-based techniques for air quality forecasting may obtain greater achievement in the future development.

8. Conclusions

This paper comprehensively reviewed the recent modeling methodologies and algorithms in air pollution forecasting. At the beginning of this work, different types of simple intelligent predictors and their improved versions are classified. The models have their advantages and limitations and they are applied in the forecasting tasks corresponding to different variables.

As the research progresses and increasing of massive input data, simple models are becoming insufficient to support complex conditions. Besides the simple predictors, other components also have been combined into the models to form the hybrid predictive models. The main types of adopted components in the literature were summarized, namely the data processing methods and auxiliary methods: metaheuristic optimization and ensemble learning, whose characteristics are also summarized. With the concern for the complexity of air pollution and the overall consideration of the research space, the spatial analysis of the massive data has been added into the modeling structures with the improved predictors or the spatiotemporal hybrid models. At the end of this paper, some representative forecasting models are listed to describe the frameworks and characteristics of air quality forecasting. The comparative analysis for the selection of efficient models is also conducted. Moreover, the points of future development trends are given. The possible or potential research directions may lead to further improvement of the forecasting performance.

Declaration of competing interest

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

Acknowledgments

This study is fully supported by the National Natural Science Foundation of China (Grant No. 52072412), the Changsha Science & Technology Project, China (Grant No. KQ1707017), and the innovation driven project of the Central South University, China (Project No. 2019CX005).

References

[1] Y. Zhou, X. Zhao, K.-P. Lin, C.-H. Wang, L. Li, A Gaussian process mixture model-based hard-cut iterative learning algorithm for air quality prediction, Appl. Soft Comput. 85 (2019) 105789.
[2] M. Péres, G. Ruiz, S. Nesmachnow, A.C. Olivera, Multiobjective evolutionary optimization of traffic flow and pollution in Montevideo, Uruguay, Appl. Soft Comput. 70 (2018) 472–485.
[3] Y.S.H. Najjar, Gaseous pollutants formation and their harmful effects on health and environment, Innov. Energy Policies 1 (2011) 1–9.
[4] X. Han, Y. Liu, H. Gao, J. Ma, X. Mao, Y. Wang, X. Ma, Forecasting PM 2.5 induced male lung cancer morbidity in China using satellite retrieved PM 2.5 and spatial analysis, Sci. Total Environ. 607–608 (2017) 1009–1017.
[5] A.J. Cohen, M. Brauer, R. Burnett, H.R. Anderson, J. Frostad, K. Estep, K. Balakrishnan, B. Brunekreef, L. Dandona, R. Dandona, Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study, 2015, Lancet 389 (2017) 1907–1918.
[6] Y. Hao, H. Peng, T. Temulun, L. Liu, J. Mao, Z. Lu, H. Chen, How harmful is air pollution to economic development? New evidence from PM2.5 concentrations of Chinese cities, J. Cleaner Prod. 172 (2018) 743–757.
[7] Y. Xia, D. Guan, X. Jiang, L. Peng, H. Schroeder, Q. Zhang, Assessment of socioeconomic costs to China's air pollution, Atmos. Environ. 139 (2016) 147–156.
[8] W. Wei, L. Peng, H. Wang, M. Song, Quantifying the effects of air pollution control policies: A case of Shanxi province in China, Atmos. Pollut. Res. 9 (2017) 429–438.
[9] K.-M. Nam, N.E. Selin, J.M. Reilly, S. Paltsev, Measuring welfare loss caused by air pollution in Europe: A CGE analysis, Energy Policy 38 (2010) 5059–5071.
[10] W. Zhu, J. Wang, W. Zhang, D. Sun, Short-term effects of air pollution on lower respiratory diseases and forecasting by the group method of data handling, Atmos. Environ. 51 (2012) 29–38.
[11] F. Cereceda-Balic, M. Toledo, V. Vidal, F. Guerrero, L.A. Diaz-Robles, X. Petit-Breuilh, M. Lapuerta, Emission factors for PM2. 5, CO, CO2, NOx, SO2 and particle size distributions from the combustion of wood species using a new controlled combustion chamber 3CE, Sci. Total Environ. 584 (2017) 901–910.
[12] K. Xiao, Y. Wang, G. Wu, B. Fu, Y. Zhu, Spatiotemporal characteristics of air pollutants (PM10, PM2. 5, SO2, NO2, O3, and CO) in the inland basin city of Chengdu, southwest China, Atmosphere 9 (2018) 74.
[13] K.-H. Kim, E. Kabir, S. Kabir, A review on the human health impact of airborne particulate matter, Environ. Int. 74 (2015) 136–143.
[14] F. Zhu, R. Ding, R. Lei, H. Cheng, J. Liu, C. Shen, C. Zhang, Y. Xu, C. Xiao, X. Li, The short-term effects of air pollution on respiratory diseases and lung cancer mortality in hefei: A time-series analysis, Respir. Med. 146 (2019) 57–65.
[15] A. Sotomayor-Olmedo, M.A. Aceves-Fernández, E. Gorrostieta-Hurtado, C. Pedraza-Ortega, J.M. Ramos-Arreguín, J.E. Vargas-Soto, Forecast urban air pollution in Mexico City by using support vector machines: A kernel performance approach, Int. J. Intell. Sci. 3 (2013) 126.
[16] E. Velasco, A. Retama, Ozone's threat hits back Mexico City, Sustain. Cities Soc. 31 (2017) 260–263.
[17] S. Avnery, D.L. Mauzerall, J. Liu, L.W. Horowitz, Global crop yield reductions due to surface ozone exposure: 2. Year 2030 potential crop production losses and economic damage under two scenarios of O3 pollution, Atmos. Environ. 45 (2011) 2297–2309.
[18] S. Osowski, K. Garanty, Forecasting of the daily meteorological pollution using wavelets and support vector machine, Eng. Appl. Artif. Intell. 20 (2007) 745–755.
[19] I.B. Konovalov, M. Beekmann, F. Meleux, A. Dutot, G. Foret, Combining deterministic and statistical approaches for PM10 forecasting in Europe, Atmos. Environ. 43 (2009) 6425–6434.
[20] C. Seigneur, M. Moran, Chapter 8, Chemical-Transport Models, 2010.
[21] G. Zhou, J. Xu, Y. Xie, L. Chang, W. Gao, Y. Gu, J. Zhou, Numerical air quality forecasting over eastern China: An operational application of WRF-Chem, Atmos. Environ. 153 (2017) 94–108.
[22] M.-T. Chuang, Y. Zhang, D. Kang, Application of WRF/Chem-MADRID for real-time air quality forecasting over the Southeastern United States, Atmos. Environ. 45 (2011) 6241–6250.
[23] Y.-S. Koo, D.-R. Choi, H.-Y. Kwon, Y.-K. Jang, J.-S. Han, Improvement of PM10 prediction in East Asia using inverse modeling, Atmos. Environ. 106 (2015) 318–328.
[24] X. Qiao, Q. Ying, X. Li, H. Zhang, J. Hu, Y. Tang, X. Chen, Source apportionment of PM2.5 for 25 Chinese provincial capitals and municipalities using a source-oriented Community Multiscale Air Quality model, Sci. Total Environ. 612 (2018) 462–471.
[25] A.M.M. Manders, M. Schaap, R. Hoogerbrugge, Testing the capability of the chemistry transport model LOTOS-EUROS to forecast PM10 levels in the Netherlands, Atmos. Environ. 43 (2009) 4050–4059.
[26] X. Feng, Q. Li, Y. Zhu, J. Hou, L. Jin, J. Wang, Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation, Atmos. Environ. 107 (2015) 118–128.
[27] H. Shimadera, T. Kojima, A. Kondo, Evaluation of air quality model performance for simulating long-range transport and local pollution of PM2.5 in Japan, Adv. Meteorol. 2016 (2016) 13.
[28] J. Chen, H. Chen, Z. Wu, D. Hu, J.Z. Pan, Forecasting smog-related health hazard based on social media and physical sensor, Inf. Syst. 64 (2017) 281–291.
[29] M. Ritter, M.D. Müller, M.-Y. Tsai, E. Parlow, Air pollution modeling over very complex terrain: an evaluation of WRF-Chem over Switzerland for two 1-year periods, Atmos. Res. 132 (2013) 209–222.
[30] A. Kumar, P. Goyal, Forecasting of daily air quality index in Delhi, Sci. Total Environ. 409 (2011) 5517–5523.
[31] J. Zhou, J. Zhao, P. Li, Study on gray numerical model of air pollution in wuan city, in: 2010 International Conference on Challenges in Environmental Science and Computer Engineering, IEEE, 2010, pp. 321–323.
[32] E.L. Zvereva, M.V. Kozlov, Responses of terrestrial arthropods to air pollution: a meta-analysis, Environ. Sci. Pollut. Res. 17 (2010) 297–311.
[33] A. Kumar, P. Goyal, Forecasting of air quality in Delhi using principal component regression technique, Atmos. Pollut. Res. 2 (2011) 436–444.
[34] A. Donnelly, B. Misstear, B. Broderick, Real time air quality forecasting using integrated parametric and non-parametric regression techniques, Atmos. Environ. 103 (2015) 53–65.
[35] A. Russo, A.O. Soares, Hybrid model for urban air pollution forecasting: A stochastic spatiotemporal approach, Math. Geosci. 46 (2014) 75–93.
[36] A. Chaloulakou, M. Saisana, N. Spyrellis, Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens, Sci. Total Environ. 313 (2003) 1–13.
[37] Y. Zhang, M. Bocquet, V. Mallet, C. Seigneur, A. Baklanov, Real-time air quality forecasting, part I: History, techniques, and current status, Atmos. Environ. 60 (2012) 632–655.
[38] S.M.S. Cabaneros, J.K. Calautit, B.R. Hughes, A review of artificial neural network models for ambient air pollution prediction, Environ. Model. Softw. 119 (2019) 285–304.
[39] M.M. Kamal, R. Jailani, R.L.A. Shauri, Prediction of ambient air quality based on neural network technique, in: 2006 4th Student Conference on Research and Development, IEEE, 2006, pp. 115–119.
[40] H. Wahid, Q.P. Ha, H.N. Duc, Computational intelligence estimation of natural background ozone level and its distribution for air quality modelling and emission control, in: Proceedings of the 28th International Symposium on Automation and Robotics in Construction, ISARC 2011, 2011.
[41] D.Z. Antanasijević, V.V. Pocajt, D.S. Povrenović, M.Đ. Ristić, A.A. Perić-Grujić, PM10 emission forecasting using artificial neural networks and genetic algorithm input variable optimization, Sci. Total Environ. 443 (2013) 511–519.

[42] S. Zhang, X. Li, Y. Li, J. Mei, Prediction of urban PM 2.5 concentration based on wavelet neural network, in: 2018 Chinese Control and Decision Conference, CCDC, IEEE, 2018, pp. 5514–5519.
[43] T. Li, X. Li, L. Wang, Y. Ren, T. Zhang, M. Yu, Multi-model ensemble forecast method of PM2. 5 concentration based on wavelet neural networks, in: 2018 1st International Cognitive Cities Conference, IC3, IEEE, 2018, pp. 81–86.
[44] Z. Li, J. Yang, PM-25 forecasting use reconstruct phase space LS-SVM, 2010 the 2nd Conference on Environmental Science and Information Application Technology, IEEE, 2010, pp. 143–146.
[45] Y. Yildirim, M. Bayramoglu, Adaptive neuro-fuzzy based modelling for prediction of air pollution daily levels in city of Zonguldak, Chemosphere 63 (2006) 1575–1582.
[46] T.E. Alhanafy, F. Zaghlool, A. Moustafa, Neuro fuzzy modeling scheme for the prediction of air pollution, J. Am. Sci. 6 (2010) 605–616.
[47] J. Xie, Deep neural network for PM2.5 pollution forecasting based on manifold learning, in: 2017 International Conference on Sensing, Diagnostics, Prognostics, and Control, SDPC, IEEE, 2017, pp. 236–240.
[48] E. Eslami, Y. Choi, Y. Lops, A. Sayeed, A real-time hourly ozone prediction system using deep convolutional neural network, Neural Comput. Appl. (2019) 1–15.
[49] B.S. Freeman, G. Taylor, B. Gharabaghi, J. Thé, Forecasting air quality time series using deep learning, J. Air Waste Manage. Assoc. 68 (2018) 866–886.
[50] A. Masih, Machine learning algorithms in air quality modeling, Glob. J. Environ. Sci. Manage. 5 (2019) 515–534.
[51] Y. Rybarczyk, R. Zalakeviciute, Machine learning approaches for outdoor air quality modelling: A systematic review, Appl. Sci. 8 (2018) 2570.
[52] L. Bai, J. Wang, X. Ma, H. Lu, Air pollution forecasts: An overview, Int. J. Environ. Res. Publ. Health 15 (2018) 780.
[53] H. Liu, K. Jin, Z. Duan, Air PM 2.5 concentration multi-step forecasting using a new hybrid modeling method: Comparing cases for four cities in China, Atmos. Pollut. Res. 10 (2019) 1588–1600.
[54] D. Mishra, P. Goyal, Development of artificial intelligence based NO2 forecasting models at Taj Mahal, Agra, Atmos. Pollut. Res. 6 (2015) 99–106.
[55] Y. Bai, Y. Li, X. Wang, J. Xie, C. Li, Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions, Atmos. Pollut. Res. 7 (2016) 557–566.
[56] R.C. Deo, M.K. Tiwari, J.F. Adamowski, J.M. Quilty, Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model, Stoch. Environ. Res. Risk Assess. 31 (2017) 1211–1240.
[57] S. Zhu, X. Qiu, Y. Yin, M. Fang, X. Liu, X. Zhao, Y. Shi, Two-step-hybrid model based on data preprocessing and intelligent optimization algorithms (CS and GWO) for NO2 and SO2 forecasting, Atmos. Pollut. Res. 10 (2019) 1326–1335.
[58] Q. Wu, H. Lin, Daily urban air quality index forecasting based on variational mode decomposition, sample entropy and LSTM neural network, Sustain. Cities Soc. 50 (2019) 101657.
[59] S. Chen, J. Wang, H. Zhang, A hybrid PSO-SVM model based on clustering algorithm for short-term atmospheric pollutant concentration forecasting, Technol. Forecast. Soc. Change 146 (2019) 41–54.
[60] J. Murillo-Escobar, J.P. Sepulveda-Suescun, M.A. Correa, D. Orrego-Metaute, Forecasting concentrations of air pollutants using support vector regression improved with particle swarm optimization: Case study in Aburrá Valley, Colombia, Urban Clim. 29 (2019) 100473.
[61] Rubal, D. Kumar, Evolving differential evolution method with random forest for prediction of air pollution, Procedia Comput. Sci. 132 (2018) 824–833.
[62] S. Qin, F. Liu, J. Wang, B. Sun, Analysis and forecasting of the particulate matter (PM) concentration levels over four major cities of China using hybrid models, Atmos. Environ. 98 (2014) 665–675.
[63] J.A. Kamińska, A random forest partition model for predicting NO2 concentrations from traffic flow and meteorological conditions, Sci. Total Environ. 651 (2019) 475–483.
[64] S.-Q. Dotse, M.I. Petra, L. Dagar, L.C. De Silva, Application of computational intelligence techniques to forecast daily PM10 exceedances in Brunei Darussalam, Atmos. Pollut. Res. 9 (2018) 358–368.
[65] B. Zhai, J. Chen, Development of a stacked ensemble model for forecasting and analyzing daily average PM2. 5 concentrations in Beijing, China, Sci. Total Environ. 635 (2018) 644–658.
[66] Y.A. Awad, P. Koutrakis, B.A. Coull, J. Schwartz, A spatiotemporal prediction model based on support vector machine regression: Ambient Black Carbon in three New England States, Environ. Res. 159 (2017) 427–434.
[67] A. Kurt, A.B. Oktay, Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks, Expert Syst. Appl. 37 (2010) 7986–7992.
[68] Y. Zhan, Y. Luo, X. Deng, M.L. Grieneisen, M. Zhang, B. Di, Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment, Environ. Pollut. 233 (2018) 464–473.
[69] W. Yang, M. Deng, F. Xu, H. Wang, Prediction of hourly PM2. 5 using a space–time support vector regression model, Atmos. Environ. 181 (2018) 12–19.
[70] J. Ma, Y. Ding, J.C.P. Cheng, F. Jiang, Z. Wan, A temporal-spatial interpolation and extrapolation method based on geographic Long Short-Term Memory neural network for PM2.5, J. Cleaner Prod. 237 (2019) 117729.
[71] U. Pak, J. Ma, U. Ryu, K. Ryom, U. Juhyok, K. Pak, C. Pak, Deep learning-based PM2.5 prediction considering the spatiotemporal correlations: A case study of Beijing, China, Sci. Total Environ. 669 (2019) 133561.
[72] D. Qin, J. Yu, G. Zou, R. Yong, Q. Zhao, B. Zhang, A novel combined prediction scheme based on CNN and LSTM for urban PM 2.5 concentration, IEEE Access 7 (2019) 20050–20059.
[73] H. Taheri Shahraiyni, S. Sodoudi, Statistical modeling approaches for PM10 prediction in urban areas; A review of 21st-century studies, Atmosphere 7 (2016) 15.
[74] X. Wang, J.J. Klemeš, X. Dong, W. Fan, Z. Xu, Y. Wang, P.S. Varbanov, Air pollution terrain nexus: A review considering energy generation and consumption, Renew. Sustain. Energy Rev. 105 (2019) 71–85.
[75] S. Gulia, S.S. Nagendra, M. Khare, I. Khanna, Urban air quality management-A review, Atmos. Pollut. Res. 6 (2015) 286–304.
[76] M. Casazza, M. Lega, E. Jannelli, M. Minutillo, D. Jaffe, V. Severino, S. Ulgiati, 3D monitoring and modelling of air quality for sustainable urban port planning: Review and perspectives, J. Cleaner Prod. 231 (2019) 1342–1352.
[77] W. Sun, J. Sun, Daily PM 2.5 concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm, J. Environ. Manage. 188 (2017) 144–152.
[78] W.-Z. Lu, W.-J. Wang, X.-K. Wang, S.-H. Yan, J.C. Lam, Potential assessment of a neural network model with PCA/RBF approach for forecasting pollutant trends in Mong Kok urban air, Hong Kong, Environ. Res. 96 (2004) 79–87.
[79] Y. Cheng, H. Zhang, Z. Liu, L. Chen, P. Wang, Hybrid algorithm for short-term forecasting of PM2.5 in China, Atmos. Environ. 200 (2019) 264–279.
[80] A. Prakash, U. Kumar, K. Kumar, V.K. Jain, A wavelet-based neural network model to predict ambient air pollutants' concentration, Environ. Model. Assess. 16 (2011) 503–517.
[81] F. Jiang, J. He, T. Tian, A clustering-based ensemble approach with improved pigeon-inspired optimization and extreme learning machine for air quality prediction, Appl. Soft Comput. 85 (2019) 105827.
[82] H. Liu, Z. Duan, C. Chen, A hybrid framework for forecasting PM2.5 concentrations using multi-step deterministic and probabilistic strategy, Air Qual. Atmos. Health 12 (2019) 785–795.
[83] S. Zhu, X. Lian, H. Liu, J. Hu, Y. Wang, J. Che, Daily air quality index forecasting with hybrid models: A case in China, Environ. Pollut. 231 (2017) 1232–1244.
[84] M. Niu, K. Gan, S. Sun, F. Li, Application of decomposition-ensemble learning paradigm with phase space reconstruction for day-ahead PM2.5 concentration forecasting, J. Environ. Manage. 196 (2017) 110–118.
[85] Y. Bai, B. Zeng, C. Li, J. Zhang, An ensemble long short-term memory neural network for hourly PM2.5 concentration forecasting, Chemosphere 222 (2019) 286–294.
[86] S. Zhu, X. Lian, L. Wei, J. Che, X. Shen, L. Yang, X. Qiu, X. Liu, W. Gao, X. Ren, PM2. 5 forecasting using SVR with PSOGSA algorithm based on CEEMD, GRNN and GCA considering meteorological factors, Atmos. Environ. 183 (2018) 20–32.
[87] J. Wang, H. Li, H. Lu, Application of a novel early warning system based on fuzzy time series in urban air quality forecasting in China, Appl. Soft Comput. 71 (2018) 783–799.
[88] Y. Xu, H. Liu, Z. Duan, A novel hybrid model for multi-step daily AQI forecasting driven by air pollution big data, Air Qual. Atmos. Health 13 (2020) 197–207.
[89] J. Wang, L. Bai, S. Wang, C. Wang, Research and application of the hybrid forecasting model based on secondary denoising and multi-objective optimization for air pollution early warning system, J. Cleaner Prod. 234 (2019) 54–70.
[90] Y. Xu, P. Du, J. Wang, Research and application of a hybrid model based on dynamic fuzzy synthetic evaluation for establishing air quality forecasting and early warning system: A case study in China, Environ. Pollut. 223 (2017) 435–448.
[91] M. Deng, W. Yang, Q. Liu, R. Jin, F. Xu, Y. Zhang, Heterogeneous space–time artificial neural networks for space–time series prediction, Trans. GIS 22 (2018) 183–201.
[92] M.A. Elangasinghe, N. Singhal, K.N. Dirks, J.A. Salmond, Development of an ANN–based air pollution forecasting system with explicit knowledge through sensitivity analysis, Atmos. Pollut. Res. 5 (2014) 696–708.
[93] G. de Gennaro, L. Trizio, A. Di Gilio, J. Pey, N. Pérez, M. Cusack, A. Alastuey, X. Querol, Neural network model for the prediction of PM10 daily concentrations in two sites in the Western Mediterranean, Sci. Total Environ. 463 (2013) 875–883.

[94] Z. Jiang, X. Meng, C. Yang, G. Li, A BP neural network prediction model of the urban air quality based on rough set, in: 2008 Fourth International Conference on Natural Computation, IEEE, 2008, pp. 362–370.
[95] H. Zhang, Y. Liu, R. Shi, Q. Yao, Evaluation of PM10 forecasting based on the artificial neural network model and intake fraction in an urban area: A case study in Taiyuan City, China, J. Air Waste Manage. Assoc. 63 (2013) 755–763.
[96] H. Zheng, X. Shang, Study on prediction of atmospheric PM2.5 based on RBF neural network, in: 2013 Fourth International Conference on Digital Manufacturing & Automation, IEEE, 2013, pp. 1287–1289.
[97] D.F. Specht, A general regression neural network, IEEE Trans. Neural Netw. 2 (1991) 568–576.
[98] L. Zheng, S. Yu, M. Yu, Monitoring NOx emissions from coal fired boilers using generalized regression neural network, in: 2008 2nd International Conference on Bioinformatics and Biomedical Engineering, 2008.
[99] D.Z. Antanasijević, M.Đ. Ristić, A.A. Perić-Grujić, V.V. Pocajt, Forecasting human exposure to PM10 at the national level using an artificial neural network approach, J. Chemometr. 27 (2013) 170–177.
[100] G. Sun, S.J. Hoff, B.C. Zelle, M.A. Smith, Development and comparison of backpropagation and generalized regression neural network models to predict diurnal and seasonal gas and PM10 concentrations and emissions from swine buildings, in: 2008 Providence, Rhode Island, June 29–July 2, 2008, American Society of Agricultural and Biological Engineers, 2008, p. 1.
[101] Z. Zainuddin, O. Pauline, Modified wavelet neural network in function approximation and its application in prediction of time-series pollution data, Appl. Soft Comput. 11 (2011) 4866–4874.
[102] J.L. Elman, Finding structure in time, Cogn. Sci. 14 (1990) 179–211.
[103] U. Brunelli, V. Piazza, L. Pignato, F. Sorbello, S. Vitabile, Two-days ahead prediction of daily maximum concentrations of SO2, O3, PM10, NO2, CO in the urban area of Palermo, Italy, Atmos. Environ. 41 (2007) 2967–2995.
[104] U. Brunelli, V. Piazza, L. Pignato, F. Sorbello, S. Vitabile, Three hours ahead prevision of SO2 pollutant concentration using an elman neural based forecaster, Build. Environ. 43 (2008) 304–314.
[105] S. Wu, Q. Feng, Y. Du, X. Li, Artificial neural network models for daily PM10 air pollution index prediction in the urban area of Wuhan, China, Environ. Eng. Sci. 28 (2011) 357–363.
[106] P.J.G. Nieto, E.F. Combarro, J.J.D.C. Díaz, E. Montañés, A SVM-based regression model to study the air quality at local scale in Oviedo urban area (Northern Spain): A case study, Appl. Math. Comput. 219 (2013) 8923–8937.
[107] B.C. Liu, A. Binaykia, P.C. Chang, M.K. Tiwari, C.C. Tsao, Urban air quality forecasting based on multi-dimensional collaborative Support Vector Regression (SVR): A case study of Beijing-Tianjin-Shijiazhuang, PLoS One 12 (2017).
[108] A.S. Sánchez, P.J.G. Nieto, P.R. Fernández, J.J. del Coz Díaz, F.J. Iglesias-Rodríguez, Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain), Math. Comput. Modelling 54 (2011) 1453–1466.
[109] L. Tang, C. Zhou, L. He, S. Zhang, Research on air quality of Beijing-Tianjin-Hebei region based on SVM and regression analysis, in: 2017 International Conference on Education, Economics and Management Research, ICEEMR 2017, Atlantis Press, 2017.
[110] W. Lu, W. Wang, Potential assessment of the support vector machine method in forecasting ambient air pollutant trends, Chemosphere 59 (2005) 693–701.
[111] A. Saxena, S. Shekhawat, Ambient air quality classification by grey wolf optimizer based support vector machine, J. Environ. Publ. Health 2017 (2017).
[112] J.A.K. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neural Process. Lett. 9 (1999) 293–300.
[113] Y. Zhai, X. Ding, X. Jin, L. Zhao, Adaptive LSSVM based iterative prediction method for NOx concentration prediction in coal-fired power plant considering system delay, Appl. Soft Comput. 89 (2020) 106070.
[114] Q. Song, B.S. Chissom, Fuzzy time series and its models, Fuzzy Sets Systems 54 (1993) 269–277.
[115] N.G. Dincer, Ö. Akkuş, A new fuzzy time series model based on robust clustering for forecasting of air pollution, Ecol. Inform. 43 (2018) 157–164.
[116] D. Domańska, M. Wojtylak, Application of fuzzy time series models for forecasting pollution concentrations, Expert Syst. Appl. 39 (2012) 7673–7679.
[117] J.S.R. Jang, ANFIS: adaptive-network-based fuzzy inference system, IEEE Trans. Syst. Man Cybern. 23 (1993) 665–685.
[118] S. Jain, M. Khare, Adaptive neuro-fuzzy modeling for prediction of ambient CO concentration at urban intersections and roadways, Air Qual. Atmos. Health 3 (2010) 203–212.
[119] M. Oprea, S.F. Mihalache, M. Popescu, A comparative study of computational intelligence techniques applied to PM2.5 air pollution forecasting, in: 2016 6th International Conference on Computers Communications and Control, ICCCC, IEEE, 2016, pp. 103–108.
[120] Y.-C. Lin, S.-J. Lee, C.-S. Ouyang, C.-H. Wu, Air quality prediction by neuro-fuzzy modeling approach, Appl. Soft Comput. 86 (2020) 105898.
[121] G. Huang, Q. Zhu, C.K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (2006) 489–501.
[122] C. Zhao, M. van Heeswijk, J. Karhunen, Air quality forecasting using neural networks, in: 2016 IEEE Symposium Series on Computational Intelligence, SSCI, IEEE, 2016, pp. 1–7.
[123] J. Zhang, W. Ding, Prediction of air pollutants concentration based on an extreme learning machine: the case of Hong Kong, Int. J. Environ. Res. Publ. Health 14 (2017) 114.
[124] N. Liang, G. Huang, P. Saratchandran, N. Sundararajan, A fast and accurate online sequential learning algorithm for feedforward networks, IEEE Trans. Neural Netw. 17 (2006) 1411–1423.
[125] A. Bueno, G.P. Coelho, J.R. Bertini, Online sequential learning based on extreme learning machines for particulate matter forecasting, in: 2017 Brazilian Conference on Intelligent Systems, BRACIS, IEEE, 2017, pp. 169–174.
[126] H. Peng, A.R. Lima, A. Teakles, J. Jin, A.J. Cannon, W.W. Hsieh, Evaluating hourly air quality forecasting in Canada with nonlinear updatable machine learning methods, Air Qual. Atmos. Health 10 (2017) 195–211.
[127] W. Deng, Q. Zheng, L. Chen, Regularized extreme learning machine, in: 2009 IEEE Symposium on Computational Intelligence and Data Mining, IEEE, 2009, pp. 389–395.
[128] H. Liu, C. Chen, Prediction of outdoor PM2.5 concentrations based on a three-stage hybrid neural network model, Atmos. Pollut. Res. 11 (2020) 469–481.
[129] M. Krishan, S. Jha, J. Das, A. Singh, M.K. Goyal, C. Sekar, Air quality modelling using long short-term memory (LSTM) over NCT-delhi, India, Air Qual. Atmos. Health 12 (2019) 899–908.
[130] J.-H. Park, S.-J. Yoo, K.-J. Kim, Y.-H. Gu, K.-H. Lee, U.-H. Son, PM10 density forecast model using long short term memory, in: 2017 Ninth International Conference on Ubiquitous and Future Networks, ICUFN, IEEE, 2017, pp. 576–581.
[131] J. Li, X. Shao, R. Sun, A DBN-based deep neural network model with multitask learning for online air quality prediction, J. Control Sci. Eng. 2019 (2019) 9.
[132] P. Smolensky, Information Processing in Dynamical Systems: Foundations of Harmony Theory, Colorado Univ at Boulder Dept of Computer Science, 1986.
[133] R. Salakhutdinov, G. Hinton, Deep boltzmann machines, in: Artificial Intelligence and Statistics, 2009, pp. 448–455.
[134] C. Zhang, J. Yan, C. Li, X. Rui, L. Liu, R. Bie, On estimating air pollution from photos using convolutional neural network, in: Proceedings of the 24th ACM International Conference on Multimedia, ACM, 2016, pp. 297–301.
[135] N. Rijal, R.T. Gutta, T. Cao, J. Lin, Q. Bo, J. Zhang, Ensemble of deep neural networks for estimating particulate matter from images, in: 2018 IEEE 3rd International Conference on Image, Vision and Computing, ICIVC, IEEE, 2018, pp. 733–738.
[136] Q. Wang, T. Xie, S. Wang, Research on air pollution gases recognition method based on LSTM recurrent neural network and gas sensors array, in: 2018 Chinese Automation Congress, CAC, IEEE, 2018, pp. 3486–3491.
[137] X. Li, L. Peng, X. Yao, S. Cui, Y. Hu, C. You, T. Chi, Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation, Environ. Pollut. 231 (2017) 997–1004.
[138] K.S. Rao, G.L. Devi, N. Ramesh, Air quality prediction in Visakhapatnam with LSTM based recurrent neural networks, Int. J. Intell. Syst. Appl. 11 (2019) 18–24.
[139] B. Liu, S. Yan, J. Li, G. Qu, Y. Li, J. Lang, R. Gu, An attention-based air quality forecasting method, in: 2018 17th IEEE International Conference on Machine Learning and Applications, ICMLA, IEEE, 2018, pp. 728–733.
[140] X. Zhou, J. Xu, P. Zeng, X. Meng, Air pollutant concentration prediction based on GRU method, J. Phys.: Conf. Ser. (2019) 032058.
[141] Y. Xing, J. Yue, C. Chen, Y. Xiang, Y. Chen, M. Shi, A deep belief network combined with modified grey wolf optimization algorithm for PM2.5 concentration prediction, Appl. Sci. 9 (2019) 3765.
[142] M. Ashish, B. Rashmi, Prediction of daily air pollution using wavelet decomposition and adaptive-network-based fuzzy inference system, Int. J. Environ. Sci. 2 (2011) 185.
[143] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 454 (1998) 903–995.
[144] M. Niu, Y. Wang, S. Sun, Y. Li, A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term PM2.5 concentration forecasting, Atmos. Environ. 134 (2016) 168–180.
[145] C. Li, Z. Zhu, Research and application of a novel hybrid air quality early-warning system: A case study in China, Sci. Total Environ. 626 (2018) 1421–1438.
[146] P. Jiang, C. Li, R. Li, H. Yang, An innovative hybrid air pollution early-warning system based on pollutants forecasting and extenics evaluation, Knowl.-Based Syst. 164 (2019) 174–192.

[147] E. Sharma, R.C. Deo, R. Prasad, A.V. Parisi, A hybrid air quality early-warning framework: An hourly forecasting model with online sequential extreme learning machines and empirical mode decomposition algorithms, Sci. Total Environ. 709 (2020) 135934.
[148] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, IEEE Trans. Signal Process. (2014) 531–544.
[149] Y. Hao, C. Tian, The study and application of a novel hybrid system for air quality early-warning, Appl. Soft Comput. 74 (2019) 729–746.
[150] J. Gilles, Empirical wavelet transform, in: IEEE Trans. Signal Process., 2013, pp. 3999–4010.
[151] H. Liu, H. Wu, X. Lv, Z. Ren, M. Liu, Y. Li, H. Shi, An intelligent hybrid model for air pollutant concentrations forecasting: Case of Beijing in China, Sustain. Cities Soc. 47 (2019) 101471.
[152] H. Liu, Y. Xu, C. Chen, Improved pollution forecasting hybrid algorithms based on the ensemble method, Appl. Math. Model. 73 (2019) 473–486.
[153] K. Gan, S. Sun, S. Wang, Y. Wei, A secondary-decomposition-ensemble learning paradigm for forecasting PM2.5 concentration, Atmos. Pollut. Res. 9 (2018) 989–999.
[154] D. Wang, S. Wei, H. Luo, C. Yue, O. Grunder, A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine, Sci. Total Environ. 580 (2017) 719–733.
[155] D. Wang, Y. Liu, H. Luo, C. Yue, S. Cheng, Day-ahead PM 2.5 concentration forecasting using WT-VMD based decomposition method and back propagation neural network improved by differential evolution, Int. J. Environ. Res. Publ. Health 14 (2017) 764.
[156] K. Siwek, S. Osowski, Improving the accuracy of prediction of PM10 pollution by the wavelet transformation and an ensemble of neural predictors, Eng. Appl. Artif. Intell. 25 (2012) 1246–1258.
[157] K.S. Lee, Z.W. Geem, A new meta-heuristic algorithm for continuous engineering optimization: harmony search theory and practice, Comput. Methods Appl. Mech. Eng. 194 (2005) 3902–3933.
[158] P. Chen, Z. Zeng, Developing two heuristic algorithms with metaheuristic algorithms to improve solutions of optimization problems with soft and hard constraints: An application to nurse rostering problems, Appl. Soft Comput. 93 (2020) 106336.
[159] W. Liu, G. Guo, F. Chen, Y. Chen, Meteorological pattern analysis assisted daily PM2. 5 grades prediction using SVM optimized by PSO algorithm, Atmos. Pollut. Res. 10 (2019) 1482–1491.
[160] Y. Teng, X. Huang, S. Ye, Y. Li, Prediction of particulate matter concentration in Chengdu based on improved differential evolution algorithm and BP neural network model, in: 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis, ICCCBDA, 2018.
[161] H. Yang, Z. Zhu, C. Li, R. Li, A novel combined forecasting system for air pollutants concentration based on fuzzy theory and optimization of aggregation weight, Appl. Soft Comput. 87 (2020) 105972.
[162] Q. Wu, H. Lin, A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors, Sci. Total Environ. 683 (2019) 808–821.
[163] P. Du, J. Wang, Y. Hao, T. Niu, W. Yang, A novel hybrid model based on multi-objective harris hawks optimization algorithm for daily PM2.5 and PM10 forecasting, Appl. Soft Comput. (2020) 106620.
[164] B. Wang, Z. Tian, Prediction of air quality index based on improved neural network, in: 2017 International Conference on Computer Systems, Electronics and Control, ICCSEC, IEEE, 2017, pp. 200–204.
[165] W. Sun, C. Wang, C. Zhang, Factor analysis and forecasting of CO2 emissions in Hebei, using extreme learning machine based on particle swarm optimization, J. Cleaner Prod. 162 (2017) 1095–1101.
[166] F. Zhao, W. Li, A combined model based on feature selection and woa for pm2.5 concentration forecasting, Atmosphere 10 (2019) 223.
[167] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[168] R. Martınez-Espana, A. Bueno-Crespo, I. Timón, J. Soto, A. Munoz, J.M. Cecilia, Air-pollution prediction in smart cities through machine learning methods: A case of study in Murcia, Spain, J. UCS 24 (2018) 261–276.
[169] A. Philibert, C. Loyce, D. Makowski, Prediction of N2O emission from local information with Random Forest, Environ. Pollut. 177 (2013) 156–163.
[170] H. Liu, C. Chen, Spatial air quality index prediction model based on decomposition, adaptive boosting, and three-stage feature selection: A case study in China, J. Cleaner Prod. (2020) 121777.
[171] W. Sun, Z. Li, Hourly PM 2.5 concentration forecasting based on mode decomposition-recombination technique and ensemble learning approach in severe haze episodes of China, J. Cleaner Prod. (2020) 121442.
[172] C. Wen, S. Liu, X. Yao, L. Peng, X. Li, Y. Hu, T. Chi, A novel spatiotemporal convolutional long short-term neural network for air pollution prediction, Sci. Total Environ. 654 (2019) 1091–1099.
[173] J. Wang, G. Song, A deep spatial–temporal ensemble model for air quality prediction, Neurocomputing 314 (2018) 198–206.
[174] Y. Qi, Q. Li, H. Karimian, D. Liu, A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory, Sci. Total Environ. 664 (2019) 1–10.
[175] W. Sun, J. Sun, Daily PM 2.5 concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm, J. Environ. Manag. 188 (2016) 144–152.
[176] S. Singh, P.K. Sharma, B. Yoon, M. Shojafar, G.H. Cho, I.-H. Ra, Convergence of blockchain and artificial intelligence in IoT network for the sustainable smart city, Sustain. Cities Soc. 63 (2020) 102364.
[177] M.M. Rathore, A. Paul, W.-H. Hong, H. Seo, I. Awan, S. Saeed, Exploiting IoT and big data analytics: Defining smart digital city using real-time urban data, Sustain. Cities Soc. 40 (2018) 600–610.
[178] S. De Vito, G. Di Francia, E. Esposito, S. Ferlito, F. Formisano, E. Massera, Adaptive machine learning strategies for network calibration of IoT smart air quality monitoring devices, Pattern Recognit. Lett. 136 (2020) 264–271.
[179] A.S. Mihăiţă, L. Dupont, O. Chery, M. Camargo, C. Cai, Evaluating air quality by combining stationary, smart mobile pollution monitoring and data-driven modelling, J. Cleaner Prod. 221 (2019) 398–418.
[180] M. Asgari, M. Farnaghi, Z. Ghaemi, Predictive mapping of urban air pollution using Apache Spark on a Hadoop cluster, in: Proceedings of the 2017 International Conference on Cloud and Big Data Computing, 2017, pp. 89–93.