[COR: alpha]
[FN: Number]

Deep learning for Air Pollutant Concentration Prediction: A Review

Bo Zhang[a,*], Yi Rong[a,1], Ruihan Yong[a], Dongming Qin[c], Maozhen Li[a,d,*], Guojian Zou[b,*] and Jianguo Pan[a,*]

[a]College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234, PR China
[b]The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, PR China
[c]College of Electronic and Information Engineering, Tongji University, Shanghai, 201804 and now is with the 3Clear, Beijing 100029, PR China
[d]The Department of Electronic and Computer Engineering, Brunel University, London, Uxbridge UB8 3PH, UK

ARTICLE INFO

Keywords:
Air pollutant concentration prediction
Deep learning
Spatial correlation
Temporal correlation
Spatio-temporal correlations

ABSTRACT

Air pollution has become one of the critical environmental problem in the 21st century and has attracted worldwide attentions. To mitigate it, many researchers have investigated the issue and attempted to accurately predict air pollutant concentrations using various methods. Currently, deep learning methods are the most prevailing ones. In this paper, we extend a comprehensive review on deep learning methods specifically for air pollutant concentration prediction. We start from the analysis on non-deep learning methods applied in air pollutant concentration prediction in terms of expertise, applications and deficiencies. Then, we investigate current deep learning methods for air pollutant concentration prediction from the perspectives of temporal, spatial and spatio-temporal correlations these methods could model. Further, we list some public datasets and auxiliary features used in air pollutant prediction, and compare representative experiments on these datasets. From the comparison, we draw some conclusions. Finally, we identify current limitations and future research directions of deep learning methods for air pollutant concentration prediction. The review may inspire researchers and to a certain extent promote the development of deep learning in air pollutant concentration prediction.

1. Introduction

Air pollution, one of our eras greatest misfortunes [1], has progressively increased climate crisis, and exacerbated environmental degradation [2]. Air pollutants have posed great threats on public and individual health [3]: people exposed to contaminated air are more likely to get respiratory and cardiovascular diseases [4, 5]. Accurate prediction of air pollutant concentration are of great significance for strategies to control and limit air pollution, and for better human health and environmental management [6].

1.1. Challenges

In most studies, air pollutants refer to a mixture of particles and gaseous materials, including SO_2, $PM_{2.5}$, PM_{10}, NO_2, CO, and O_3 [7]. Prediction on their concentrations is challenging, and usually affected by the following factors.

1) Concentrations of pollutants contain complex correlations in both temporal and spatial dimensions, and the correlation of each dimension are dynamically changing.

Dynamic spatial correlation Researchers generally treat air pollutant concentration prediction as a single-site prediction problem. Existing studies have demonstrated that pollutant concentrations are affected by other factors and diffuses between sites, including geographic locations and meteorological conditions, like temperature, humidity, wind direction, etc. For example, there are two neighboring sites in a region with a north-south distribution. If the wind is blowing from the south, we can predict that the air pollutant concentration at the northern site will increase [8–10]. This property is defined as dynamic spatial correlation, i.e., the influence of the surroundings on the pollutant concentration at the predicted location varies over time. Therefore, the impact of dynamic spatial correlation must be incorporated into air pollutant concentrations prediction, especially to extract complex and non-linear spatial correlations in depth.

*Corresponding author.
zhangbo@shnu.edu.cn (B. Zhang); 1000496945@small.shnu.edu.cn (Y. Rong); yongruihan@163.com (R. Yong); qindm@3clear.com (D. Qin); maozhen.li@brunel.ac.uk (M. Li); 2010768@tongji.edu.cn (G. Zou); panjg@shnu.edu.cn (J. Pan)
ORCID(s):
[1]This author contributed equally to this work and should be considered co-first author.

Dynamic temporal correlation Studies have shown that the concentration of air pollutants at a specific site changes dynamically and continuously in the temporal dimension [11–13]. This property is defined as dynamic temporal correlation, i.e., historical observation timesteps have different effects on prediction timesteps. Therefore, we must consider the impact of each timestep in the history on the prediction timestep, and the weight of the impact is dynamically updated, which needs to be calculated in real-time.

2) Auxiliary factors

Air pollutant concentrations are also influenced by some other factors, like point of interests (POIs), which refer to the unique features of the target region. POIs features may include: factories, decoration and furniture markets, parks, subway stations, schools, etc. In addition to this, there are traffic, complex chemical mechanisms, meteorological conditions, regional geographic relationships, temporal features.

1.2. Related summaries on air pollutant concentration prediction

Some researchers have reviewed prediction approaches for air pollutant concentrations from different perspectives. Bai et al. reviewed the theory and application of statistical prediction methods, artificial intelligence methods, numerical prediction methods and hybrid models. In addition, based on a comparison of different prediction methods, the advantages and disadvantages of these models are also presented [14]. Masih presented a survey of machine learning methods for air pollutant concentration prediction from 2013 to 2018. This work aims to focus on the fundamentals of machine learning techniques and their role in improving predictive performance. The following four main summaries were conducted: (1) the role of input predictors in improving the accuracy of model predictions; (2) geographic locations where these studies were conducted; (3) major machine learning techniques applied to pollution concentration prediction or estimation; (4) whether these techniques were based on Linear Regression (LR), Neural Network (NN), Support Vector Machine (SVM) or Ensemble learning algorithms (EL) [15]. Cabaneros et al. mainly reviewed the use of Artificial Neural Networks (ANNs) for long-term prediction of outdoor $PM_{2.5}$, PM_{10}, oxides of nitrogen, and ozone. The vast majority of the identified works utilized meteorological and source emissions predictors almost exclusively. In conclusion, the results of the study demonstrate the need to formulate a systematic protocol for developing powerful ANNs models [16]. Liao et al. provided a brief review of recent attempts on using deep learning methods in air pollution prediction. Their main work is to introduce deep network architectures for air pollution prediction and their relevance to explore the non-linear spatio-temporal correlations across multiple scales of air pollution. The potential of deep learning techniques for air pollution prediction is then examined in terms of data gap filling, prediction algorithms, improvement of Chemical Transport Models (CTMs), satellite data estimation, and source estimation for atmospheric dispersion prediction [17]. Masood and Ahmad presented an overview of artificial intelligence-based methods commonly used for air pollution prediction between 2003 and 2021. In addition, some technological gaps in these literatures and the pros and cons associated with the different AI techniques, were discussed [18]. Despite the extensiveness of these reviews, they covered studies from a single or several aspects, and an overarching and comprehensive review for air pollutant concentration prediction is still on demand.

1.3. Classification of existing methods

Due to the progress of artificial intelligence, many algorithms have emerged to predict air pollutant concentrations. These approaches can be divided into two categories: non-deep learning methods and deep learning methods.

The non-deep learning methods roughly include two major models: deterministic ones and statistical ones [19]. The most representative deterministic methods are the Community Multiscale Air Quality (CMAQ) model [20–22], the Weather Research and Forecasting model coupled with Chemistry (WRF-Chem) [23, 24], Weather Research and Forecasting/ Chemistry-Madrid (WRF/Chem-MADRID) [25], the Nested Air Quality Prediction Modeling System (NAQPMS) [26, 27], Chemical Lagrangian Model of the Stratosphere (CLaMS) [28], Operational Street Pollution Models (OSPM) [29] and Comprehensive Air-quality Model with extension (CAMx) [30], LOTOS-EUROS [31], MOZART [32]. Due to some reasons, such as the use of ideal theory in the determination of model structure and the estimation of parameters by experience, the predictive performance of these models is limited [8, 33, 34]. Statistical methods are well known to researchers because they avoid sophisticated theoretical models and simply apply statistics-based models, gradually emerging in air pollution prediction [35, 36]. These methods can also be classified into two categories, classical statistical ones and traditional machine learning ones. Classic statistical methods refer to those based on Auto Regression Integrated Moving Average (ARIMA) [37–40], or Geographically Weighted Regression (GWR) [41]; while traditional machine learning methods usually use the Random Forest (RF) [42], SVM [40, 43–46], LR [47, 48] and ANNs [49–51]. These methods can capture non-linear features from raw data to a certain extent, but

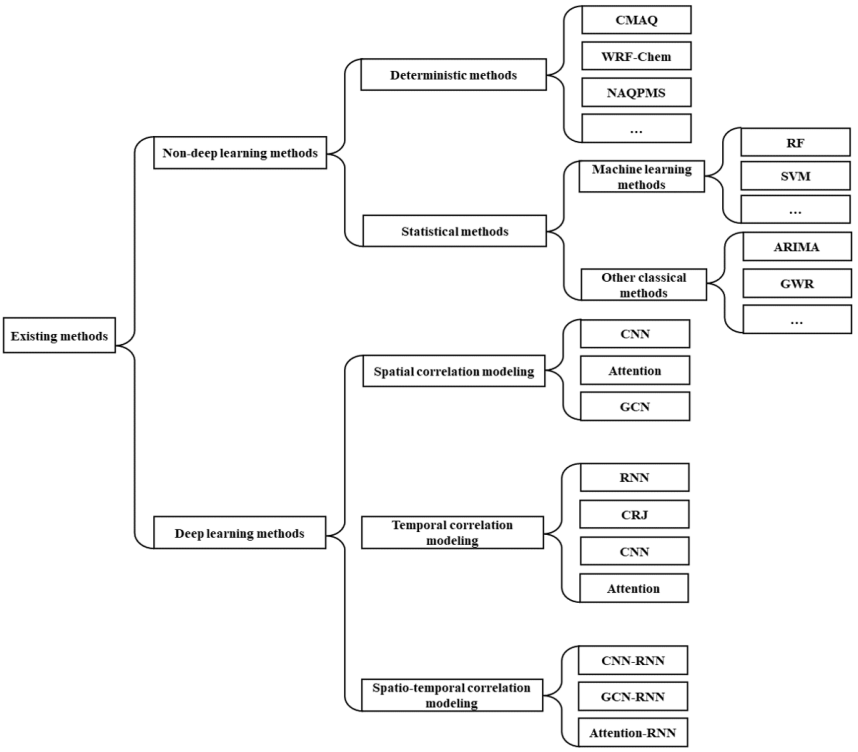

Figure 1: Classification of existing methods for air pollutant concentration prediction.

complex spatio-temporal correlations in the historical data cannot be fully extracted [52, 53].

With the rapid growth of artificial intelligence and big data techniques, deep learning (DL) technologies have been extremely used in big data analysis to solve various classification and regression problems, including: computer vision [54], image classification [55], speech recognition [56], policy analysis [57], time series prediction [58] and natural language processing [59], especially in the field of air pollution prediction [60–62]. They can not only predict specific values of pollutant concentrations, but also classify pollutant levels as well as achieve long-term prediction tasks for multiple days in the future [63, 64]. In addition to this, some researchers have shown that deep learning technologies determine the underlying reasons of meteorological conditions, seasonal variations, regional geographical relationships, chemical mechanisms, etc. that can change air pollutant concentrations [11, 65–67].

The advantages of deep learning technology in dealing with non-linear spatio-temporal correlations have been gradually discovered, which has dramatically improved the accuracy of pollutants concentration prediction [68, 69]. A typical deep learning method uses a hierarchical model to map the input to the final output [70–73]. Its modeling idea is to construct certain blocks or layers together to form an end-to-end deep network structure. Deep learning models are quite suitable for processing historical air pollution data with spatio-temporal correlations. For instance, Convolutional Neural Networks (CNNs) [74–77] are used to extract spatial correlation in regional multi-site data; Graph Convolutional Neural Networks (GCNs) [68, 78] apply convolution operations to graph structure data, thus they can well represent the network of air quality monitoring sites in non-Euclidean space. Recurrent Neural Networks (RNNs), like the Original Recurrent Neural Network (Original RNN) [79–82] and its variants, Gated Recurrent Units (GRUs) [82–85], Long Short-Term Memory (LSTM) [82, 86–91], Read-first LSTM (RLSTM) [11], Long Short-Term Memory Neural Network Extended (LSTME) [92], can be used to extract temporal correlation in raw data. Convolutional Neural Network- Long Short-Term Memory (CNN-LSTM) based on CNN and LSTM [8, 9, 52, 93–96] can be used to extract spatio-temporal correlations from historical data. Here we summarize the key techniques commonly used in existing air pollutant concentration prediction methods, as shown in Figure 1. The full names of the abbreviated methods mentioned in the paper are listed in Table 1, Table 2 and Table 3, where Table 1 corresponds to deterministic methods, Table 2 to statistical methods and Table 3 to deep learning methods.

Table 1
The description of deterministic methods

Model name	Full name
CAMx [30]	Comprehensive Air-quality Model with extension
CMAQ [20–22]	Community Multiscale Air Quality
CLaMS [28]	Chemical Lagrangian Model of the Stratosphere
CTMs [97]	Chemical Transport Models
NAQPMS [26, 27]	Nested Air Quality Prediction Modeling System
OSPM [29]	Operational Street Pollution Models
WRF [21, 23–25]	Weather Research and Forecasting
WRF-Chem [23, 24]	Weather Research and Forecasting/ Chemistry
WRF/Chem-MADRID [25]	Weather Research and Forecasting/ Chemistry-Madrid

Table 2
The description of statistical methods

Model name	Full name
ARIMA [37–40]	Autoregressive Integrated Moving Average Model
ANNs [49–51]	Artificial Neural Networks
BPNN [49]	Backpropagation Neural Network
DWT [98]	Discrete Wavelet Transform
EMD [99]	Empirical Mode Decomposition
GWR [41]	Geographically Weighted Regression
GRNN [50]	Generalized Regression Neural Network
LR [47, 48]	Linear Regression
LSSVM [46]	Least Squares Support Vector Machine
RF [42]	Random Forest
RBFNN [51]	Radial Basis Function Neural Network
SVM [40, 43–46]	Support Vector Machine
WT [98, 100–102]	Wavelet Transform

Table 3
The description of deep learning methods

Model name	Full name
Attention-LSTM [9, 10, 103]	Spatial Attention Mechanism-Long Short-Term Memory
Attention [9, 10, 103]	Spatial Attention Mechanism
Attention [104–107]	Attention Mechanism
BiLSTM [11]	Bidirectional Long Short-Term Memory Network
CRJ [108, 109]	Cycle Reservoir with Regular Jumps
CNNs [74–77, 110]	Convolutional Neural Networks
CNN-LSTM [8, 9, 52, 93–96]	Convolutional Neural Network-Long Short-Term Memory
C-LSTME [95]	Convolutional and Long Short-Term Memory Neural Network Extended
CNN-GRU [111]	Convolutional Neural Network-Gated Recurrent Unit
EEMD-LSTM [112]	Ensemble Empirical Mode Decomposition-Long Short-Term Memory
ESNs [113–115]	Echo State Networks
EMD-GRU [84]	Empirical Mode Decomposition-Gated Recurrent Unit
EDS Model [11]	Encoder-Decoder Stacked Model based on Read-first LSTM
FNNs [116]	Feedforward Neural Networks
GRUs [82–85]	Gated Recurrent Units
GCNs [78]	Graph Convolutional Neural Networks
GCN-LSTM [68]	Graph Convolutional Neural Network-Long Short-Term Memory
LSTM [82, 86–91]	Long Short-Term Memory
LSTME [92]	Long Short-Term Memory Neural Network Extended
Lag-LSTM [117]	Lag Layer-Long Short-Term Memory
MTCAN [63]	Multi-directional Temporal Convolutional Artificial Neural Network
Original RNN [79–82]	Original Recurrent Neural Network
RNNs [11, 80–82, 84, 85, 87–92]	Recurrent Neural Networks
RLSTM [11]	Read-first LSTM
Seq2Seq [104–106]	Sequence to Sequence
SA-LSTM [118]	Spatial Attention-based Long Short-Term Memory
TLS-BLSTM [119]	Transfer Stacked Bidirectional Long Short-Term Memory Network
TCN [63, 120–123]	Temporal Convolutional Network

1.4. Literature search and select strategy

The literature search focused on the highly indexed database Google Scholar and Web of Science journal publications to systematically study the relevant literature published in the last decade. They are one of the few compiling the most important scientific databases such as: IEEE Xplore, ScienceDirect and Springer, where major papers on air pollutant concentration prediction can be found.

The search terms include different combinations of "air pollutant concentration", "air quality", "prediction", "deep learning", "machine learning", "forecast", "hybrid model". The main objective is to analyze the most authoritative

scientific research papers in related fields worldwide in recent years. In addition, a number of quality conference papers were selected for joint study. The search process was repeated until the relevant citation stopped. Afterwards, a total of 226 papers were screened through the literature search. The literature was classified according to the input variables, modeling process, and prediction task type. Besides the citations of original methodology papers, the application papers were published in 2000-2022, with 90 percent of them published in 2015-2022.

Further selection was based on the impact factor of these papers and the Journal Citation Reports (JCR) partition. The limation of 27 papers were to focus on the recent quartiles of the journals, according to JCR, which conclude the Q1-Q4 categories. Therefore, it is not for reference because of low impact factors. Finally, excluding the 41 papers on basic methods and principles introduction, a total of 158 application papers were included in the reference.

1.5. Our Contributions

We intend to present an overview of recent approaches for air pollutant concentration prediction that utilize deep learning technologies. To our knowledge, it is the comprehensive analysis on this scope from multiple perspectives, including methods, public datasets, experiment analysis, current challenges and future directions. The contributions of this review are as follows:

1) We categorize the existing methods and conduct a comprehensive review on current deep learning methods for air pollutant concentration prediction. We specifically analyze their characteristics and principles from the perspectives of temporal, spatial and spatio-temporal relations.

2) We focus on some advanced techniques, including data decomposition, transfer learning and statistical methods, that are fused into deep learning models for air pollutant concentration prediction. This demonstrates the feasibility of these techniques to help improve the prediction performance on certain air pollution prediction tasks.

3) We list major public datasets and commonly used auxiliary features for air pollutant concentration prediction. This would provide a good reference for subsequent research, and can help researchers in this field to better identify the foothold.

4) We compare the results of experiments conducted by deep learning and non-deep learning models on these listed datasets, and some valuable conclusions were drawn based on the comparison.

5) We outline the existing challenges and limitations of the deep learning methods for air pollutant concentration prediction and suggest some future development directions of this field.

The rest of the paper is organized as follows: Section 2 summarizes non-deep learning methods for air pollutant concentration prediction from the perspective of characteristic, application and limitation. Section 3 reviews deep learning methods for air pollutant concentration prediction, including commonly used techniques for modeling temporal and spatial correlations and some other new variants. Section 4 lists major public datasets and important auxiliary features for air pollutant concentration prediction. Section 5 generalizes representative experiments of the deep learning and non-deep learning models for different prediction tasks on the listed public datasets and draws some conclusions. Section 6 identifies existing limitations and the future research directions of this field. Some conclusions are drawn in Section 7.

2. AIR POLLUTANT CONCENTRATION PREDICTION BASED ON NON-DEEP LEARNING

Non-deep learning methods for air pollutant concentration prediction extend in two directions: deterministic methods and statistical methods. Table 4 lists current most representative ones.

Deterministic methods, adopt meteorological principles and mathematical equations to simulate the process of pollutant emission, transformation, diffusion and removal based on atmospheric physical and chemical reactions [92]. For example, CTMs are designed to describe chemical and meteorological processes in the atmosphere, focusing on the emission, transport and transformation of air pollutants to establish mathematical algorithms [97]. The Weather Research and Forecasting (WRF) models [21] are used for atmospheric research and application in prediction, such as WRF-Chem [23, 24] and WRF/Chem-MADRID [25]. In addition, other deterministic methods such as CMAQ [20–22], CAMx [30] are also applied to air pollution prediction [19]. Chemical dynamic conditions, reaction index, and chemical products need to be included in the reference range. Although developed theories provide valuable insights for understanding air pollution prediction, these deterministic models are associated with complex priori knowledges, unreliable and limited data, and various usage limitations. It is not enough to explain the nonlinearity of many factors

Table 4
Classification of non-deep learning methods for air pollutant concentration prediction

Category		Representative models	Literatures
Deterministic methods		CMAQ	[20–22]
		WRF	[21, 23–25]
		NAQPMS	[26, 27]
		MOZART	[32]
		CLaMS	[28]
		LOTOS-EUROS	[31]
		OSPM	[29]
		CAMx	[30]
Statistical methods	machine learning	LR	[47, 48]
		ANNs	[49–51]
		RF	[42]
		SVM	[40, 43–46]
	Classic statistical	ARIMA	[37–40]
		GWR	[41]

related to the formation of pollutants [8, 33, 34]. To circumvent these deficiencies, some researchers have started to apply statistical methods to cope with them.

Unlike the deterministic models, statistical methods unravel non-linear relationships between various variables that may affect air pollutant concentrations [37–41]. They usually deploy mathematical theories including statistics, probability, and stochastic processes, and can be divided into two categories, classical methods and traditional machine learning methods. A typical classical method is ARIMA [37–40]. ARIMA is a traditional time series prediction method which integrates moving average and autoregressive components to model historical time sequence data. Ni et al. used an experiment in Beijing to demonstrate that the ARIMA model can predict $PM_{2.5}$ concentration more accurately for a short-term of one hour [39]. However, these methods involve considerable amounts of statistical data, and rely on certain empirical rules to make predictions. And they cannot fully capture non-linear correlations in air pollutant concentration data. Therefore, some researchers have applied traditional machine learning methods for better solutions.

The main advantage of the traditional machine learning methods over the classical statistical methods is its ability to handle non-linear features, resulting in higher prediction accuracy [124]. In addition, the application does not require a deep understanding of the dynamic and chemical processes between air pollutants and other relative variables in the atmosphere [16]. A common machine learning predictor is ANNs, which simulates the system of the human brain and nervous to model non-linear sequence. ANNs have also improved in years of research and applications, bringing more evolved versions to air pollution prediction, including the Backpropagation Neural Network (BPNN) [49], the Generalized Regression Neural Network (GRNN) [50], the Radial Basis Function Neural Network (RBFNN) [51]. Other models, like SVM, have also received a lot of attention in air pollution prediction. Similar to ANNs, there are improved versions of SVMs that can improve the prediction performance over the original model, such as the Least Squares Support Vector Machine (LSSVM) [46].

Although statistical methods have elevated the prediction accuracy from previous models, they are adapt to learn shallow hidden features and hold limited ability of extracting complex and non-linear spatio-temporal correlations and they only work well on small-scale datasets [52, 53]. Therefore, it is not enough for complex and dynamic air pollutant concentration data on a large scale and with complex spatio-temporal correlations.

3. AIR POLLUTANT CONCENTRATION PREDICTION BASED ON DEEP LEARNING

The deep learning network architectures determine their ability to extract complex and non-linear spatio-temporal correlations from data. Unlike shallow ANNs, the deep learning networks have multiple layers of neurons, which are called deep network structures. The deep network structures formed by the connection of these neurons and the settings of the activation function in the networks make them suitable for various application fields. The deep learning models have powerful capabilities in the extraction of spatio-temporal correlations, so they are suitable for air pollutant concentration prediction.

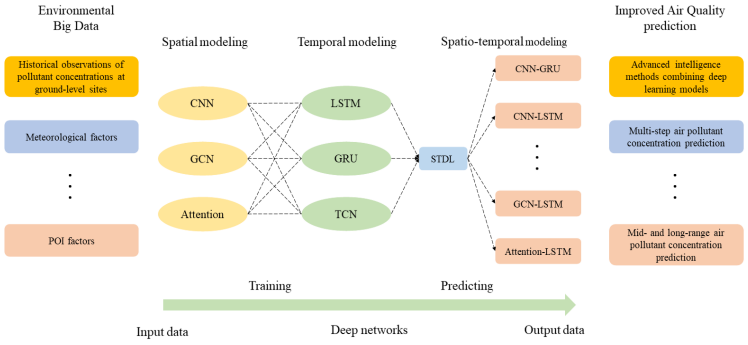

Figure 2: Air pollutant concentration prediction framework based on deep learning.

Table 5
Classification of Deep learning methods for air pollutant concentration prediction

Spatial modeling	Temporal modeling	Literatures
CNN	\	[74–77]
CNN	**LSTM**	[8, 9, 52, 93–96]
	GRU	[111]
	LSTM+Attention	[125]
GCN	**LSTM**	[68]
Attention	**LSTM**	[9, 10, 103]
\	\	[110]
CNN	**Dilated Causal Convolution**	[63, 120–123]
\	**+Auto-encoder**	[126]
Original RNN	\	[80–82]
\	**Seq2seq with GRU(+Attention)**	[104, 105]
GRU	\	[82, 84, 85]
\	\	[11, 82, 87–92]
LSTM	**+Auto-encoder**	[53, 127]
	Seq2seq with LSTM(+Attention)	[104, 106]
\	**CRJ**	[108, 109]

According to their ability to extract temporal and spatial correlations, we briefly introduce some commonly used deep learning models. The models adapting to temporal correlation predictions include LSTM, GRU, Temporal Convolutional Network (TCN). The models applied to spatially correlated feature extraction include CNN, GCN, Spatial Attention Mechanism (Attention). Spatio-temporal Deep Learning Architectures (STDL) can be built by coupling temporal and spatial modules. The STDL architectures support dealing with air pollution-related data due to their ability to represent complex and non-linear spatio-temporal correlations. Different architectures of deep networks applied to air pollutant concentration prediction are shown in Figure 2.

Table 5 presents the classification of current deep learning-based methods from the perspectives of spatial and temporal correlations the models can extract. Detailed explanation on their respect theories and applications are extended as follows.

3.1. Modeling spatial correlation

CNNs Many scholars have deployed CNNs to capture the spatial correlation in the data of air pollutant concentrations [8, 9, 52, 74–77, 93–96, 111, 125]. CNN [128] is a deep feedforward network composed of convolutional layers. They are capable of analyzing multiscale shift invariant features of data. A subsampling operation is performed between two consecutive convolutional layers. Two commonly used subsampling operations are max pooling and mean pooling. Pooling layers can be replaced by convolutional layers, simplifying the network structure [129]. The units in

the convolution layer are organized in the feature map, and each unit is connected to the local weights in the feature map of the previous layer through filters. The sum of the local weights is passed through an activation function that can take various forms, such as a Rectified Linear Units (ReLU) [130]. Some researchers integrated spatial data between different regions to a one-dimensional or two-dimensional tensor, thus the CNN was facilitated to extract the spatial correlation hidden in the tensor [52, 93]. However, the distributions of regions are not uniform, and encapsulating them into a one-dimensional or two-dimensional tensor destroys the original spatial correlation. Therefore, several studies attempted to convert these spatial data into standard grids. Each of them represents a region and the center one is the region to be predicted. For example, Soh et al. designed a matrix of 121 square sections where the center grid represents the observed location and others the regions surrounding it. In this study, CNN can extract useful spatial correlation information from the grids [94]. Some other researchers noticed CNNs great achievements in computer vision and utilized this expertise to extract spatial relations in images. Rijal et al. took images of the predicted area as the input and developed three CNN models with different structures to extract spatial correlation for $PM_{2.5}$ concentration assessment [74]. Zhang et al. also regarded images as the target for pollution prediction and applied a CNN model containing nine convolution layers, two pooling layers and two dropout layers. The convolutional layers are used to extract the spatial correlation in the raw images, and the dropout layers are used to prevent overfitting [75].

GCNs CNNs are born to operate on Euclidean spaces. Since spatial data among multiple regions usually form a non-Euclidean structure, some researchers steer to GCNs for air pollutant concentration prediction [68]. In this application, GCNs are usually based on the spectral method. They define the spectral graph convolution based on the graph Fourier Transform and the Laplacian matrix L. $x \in \mathbb{R}^N$ is set as a signal on the nodes of the undirected graph G where x_i represents a scalar signal on the i-th node. The expression of the spectral graph convolution is the multiplication of the convolution kernel g_θ and the signal x in the Fourier domain. The spectral graph convolution operator $*$ is defined as

$$y = g_\theta * x = g_\theta(L)x = g_\theta\left(U\Lambda U^T\right)x = U g_\theta(\Lambda)U^T x \tag{1}$$

where $\theta \in \mathbb{R}^N$ is the Fourier coefficient that needs to be trained. $L \in \mathbb{R}^{N\times N}$ is the normalized graph Laplacian, and described as

$$L = I_N - D^{-\frac{1}{2}}AD^{-\frac{1}{2}} = U\Lambda U^T \tag{2}$$

where $I_N \in \mathbb{R}^{N\times N}$ is the identity matrix. $D \in \mathbb{R}^{N\times N}$ is the diagonal matrix, $D_{ii} = \sum_j A_{i,j}$ [131]. $\Lambda \in \mathbb{R}^{N\times N}$ is the diagonal matrix of eigenvalues of L. $U \in \mathbb{R}^{N\times N}$ is the matrix of eigenvectors of L. Convolution kernel g_θ can be interpreted as a function of the eigenvalues of L, i.e. $g_\theta(\Lambda)$. $U^T x$ represents the Graph Fourier Transform of x.

Calculating spectral graph convolution costs too much due to $\mathcal{O}\left(n^2\right)$ multiplications with matrix U. In practice, the truncated expansion of $g_\theta(\Lambda)$ by Chebyshev polynomials is used to reduce the computational cost [132]. Equation (1) can be generalized by Chebyshev polynomials as:

$$H = \sum_{k=0}^{K} T_k(\tilde{L})X_t\Theta_k \tag{3}$$

where $X_t \in \mathbb{R}^{N\times M}$ is the signal matrix. $\Theta_k \in \mathbb{R}^{M\times W}$ is the matrix of convolution kernel parameters. $H \in \mathbb{R}^{N\times W}$ is the convolved spatial feature matrix. $T_k(\tilde{L}) \in \mathbb{R}^{N\times N}$ is the Chebyshev polynomial of the order k. The scaled Laplacian $\tilde{L} = \frac{2L}{\lambda_{\max}} - I_N$. $\lambda_{\max}$ represents the largest eigenvalue of L. For the input variable s, T_k is defined as $T_k(s) = 2sT_{k-1}(s) - T_{k-2}(s)$, which is a recurrence relation when $T_0 = 1$ and $T_1 = s$.

GCNs can be an effective tool to mine the spatial information among different regions for air pollution prediction. Qi et al. presented a good example. They modeled spatial data as an undirected graph rather than in the way of modeling Euclidean data. In a graph, nodes represent the monitor sites and edges the connections between two sites. The spatial weight matrix is constructed based on distances between sites, which is reasonable given the quantitative spatial correlation among sites. The graph is fed into a GCN based on the spectral method to extract features in the spatial domain and Chebyshev polynomials are also introduced to simplify calculations [68].

Attention The Attention Mechanism is first applied to the field of Natural Language Processing (NLP) [107]. Air pollutant concentrations recorded by one site monitor can be affected by its surrounding sites. The Spatial Attention Mechanism is used to capture the spatial correlation between sites in air pollutant concentration prediction [9, 10, 103].

Its core theory is to assign different weights according to the spatial correlation degrees of different sites. Suppose a node in a graph represents a site. Given a set of node features $\hat{X}_t = \{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_N\}$, $\hat{x}_i \in R^M$, where N represents the number of nodes, and M represents the number of features of each node. The spatial correlation index between different sites can be expressed as

$$e_{i,j} = \sigma\left(W_\beta^T \left[W\hat{x}_i, W\hat{x}_j\right]\right) \tag{4}$$

where W represents the input feature weight matrix of the node. W_β represents the weight parameter vector of the fully connected layer. $e_{i,j}$ represents the spatial correlation strength of the feature of node j on node i. σ is the activation function. The weight coefficient $\gamma_{i,j}$ of node j relative to node i can be calculated as:

$$\gamma_{i,j} = \frac{\exp\left(e_{ij}\right)}{\sum_{k=1}^{n} \exp\left(e_{ik}\right)} \tag{5}$$

where $\gamma_{i,j}$ denotes the importance of node j to node i.

Huang et al. implemented a representative application using the Spatial Attention Mechanism. They introduced an improved version of the graph-based Spatial Attention Mechanism to leverage a self-loop normalized adjacency matrix and extract the spatial correlation between different sites [10].

3.2. Modeling temporal correlation

CNNs Gehring et al. applied fully convolutional neural networks for feature extraction and verified the feasibility of CNNs to process sequence structure data [133]. Some researchers appropriated this ability for air pollutant concentration prediction. For example, Sayeed et al. used a five-layer deep convolutional neural network to extract the temporal correlation from historical observation data and predict the ozone concentration in the next 24 hours [110].

Dilated Causal Convolution is another choice to deal temporal data. It is a special standard 1D-convolution that regulates the size of the receptive field by changing the value of the dilation rate. It is particularly suitable to capture long-term temporal correlation. Some studies [63, 120–123] used it as a time convolution layer to extract the temporal correlation of air pollutants. Compared with recurrent networks, convolutional networks have a fixed-size context representation. Expanding the depth of the network by stacking layers can make the size of the effective context larger. Calculations over all the elements in the sequence can be exerted in parallel, which can fully harness GPU resources and is easier to optimize. This is superior to RNNs that, maintain the entire hidden state in the past timesteps.

RNNs Original RNNs, are variants of Feedforward Neural Networks (FNNs) [116]. FNNs enable signals to travel only one way from input to output. They are straightforward network structures without recurrent connections associated with inputs and outputs [134]. On the basis of FNNs, Original RNNs introduces the self-connection of neuronal circulatory structure into the network [79]. Thus, the key information of the input data can be memorized. The sequence of data can have a certain impact on the output of the network. LSTM are enhanced versions of the Original RNNs [86]. They introduce memory blocks to overcome vanishing and exploding gradients. The memory block consists of three gated units: the input gate, output gate, and forget gate. The multiplicative gates control the memory block operation and determine whether the input information need to be remembered. The input gate controls the flow of cell activation from input into a memory cell, while output gate controls the flow of output from a memory cell into other nodes [72]. Compared with Original RNNs, LSTM networks can better handle long time sequences data. The GRUs are simplification of the LSTM models [83]. They are simple structure consisting of only update gates and reset gates [72]. The update gate replaces the input gate and forget gate in LSTM, and its function is to decide whether the information needs to be remembered. Compared with LSTM, the advantage of using GRUs is that they are more lightweight, requires fewer parameters to be trained, and the training speed is faster. And in some research tasks, GRUs show the same excellent performance as LSTM [135]. RNNs (including the original RNNs and their variants with GRUs and LSTM), have been widely used to process sequential data. To model the complex and nonlinear temporal correlation of historical pollutant data, RNNs were deployed to deal various air pollutant concentration prediction tasks [11, 80–82, 84, 85, 87–92]. However, given the way that RNNs processes data in the order of sequence data, when modeling long time sequence, these models are inept at remembering information they have learned many time-steps before [86, 136].

In the RNN-based air pollutant concentration prediction, a network structure called encoder-decoder is also used by some researchers [53, 104–106, 126, 127]. It is to encode historical observation data into a fixed-length vector and

then decode the final prediction from the information in the vector. The encoder-decoder network is described as:

$$h = \omega\left(\rho_t; \beta_1\right) \tag{6}$$

$$X_{t+1:t+s} = f\left(h; \beta_2\right) \tag{7}$$

where ω is the encoder and f is the decoder; ρ_t represents the input information at timestep t; h is represented by the semantic vector output by the encoder; $X_{t+1:t+s}$ is s-timestep-ahead prediction values; β_1 and β_2 are trained parameters.

The encoder-decoder network structure can also ignore important temporal information. Because however long the input and the output sequences are, the length of the semantic vector between the encoder and the decoder is always fixed. If the input sequence is too long, some important information may be lost.

Attention To avoid losing temporal information, researchers have added the Attention Mechanism to the encoder-decoder network. The Attention Mechanism can adaptively select important hidden state information as the input of the decoder, and discard redundant one. The selection is fulfilled by assigning different weights on different hidden states of the historical data. The Attention Mechanism can model the non-linear temporal correlation between the observed data, even if the input is too long.

Some studies on air pollutant prediction that utilize the Attention Mechanism have achieved satisfying results. Wang et al. integrated an Attention Mechanism into a sequence to sequence (seq2seq) model to extract the temporal correlation in the hidden state at different past timesteps [104]. Tu et al. proposed an improved Attention Mechanism that integrates time decay factor into the traditional Attention Mechanism. The time decay factor can alleviate the impact of the value observed from a longer time before while increasing the impact of the value from a closer time point. Their structure also incorporated the hidden states in the decoder to push the prediction spreading forward and thus to achieve long-term prediction for air pollutant concentration [106].

CRJ Cycle Reservoir with Regular Jumps (CRJ) is the improved version of the Echo State Networks (ESNs) [113]. In ESNs, only the output layer weights are trained and all the input and reservoir layers weights are randomly initialized and scaled with a scaling parameter. ESNs have been applied to time series modeling [114, 115]. However, the random connectivity and weight structure of ESNs limit their performance. In regard of this, the CRJ network was proposed by Rodan and Tiňo [137]. A CRJ network is composed of input layers, hidden layers and output layers. A hidden layer has a fixed regular topology: Nodes in the hidden layer are connected in an un-directional cycle with bi-directional shortcuts (jumps).

The CRJ network can exploit the sequence information by extracting temporal correlation and predict air pollutant concentrations. In [108], the CRJ network was used to predict the ground-level ozone pollution of two sites. In [109], CRJ was not directly used for prediction but to extract temporal correlation from historical data of ozone concentrations.

3.3. Deep learning modeling of joint spatio-temporal correlations

Although many methods utilize the learning ability of deep learning techniques to extract temporal or spatial correlation from air pollution data, they treat the temporal correlation and the spatial correlation separately and ignore possible connections between them. Some researchers have applied hybrid methods that consider the relations between the two types of correlation, as listed in Table 5, and obtain higher prediction accuracy. These hybrid methods consist of three modules: a spatial module, a deduction module, and a temporal prediction module. The spatial module, usually based on CNNs or GCNs, etc., extracts spatial correlation that will be sent into the temporal module, usually based on RNNs or CRJ, etc. The deduction module serves as the bridge between the spatial module and the temporal module and calculates the causal relationship between the spatial and the temporal through a matrix multiplication or other types of operations. For example, Pak et al. used a spatio-temporal model (CNN-LSTM) to predict the average PM2.5 concentration of Beijing in the next day. They introduced the mutual information (MI) estimator to analyze the temporal and spatial correlations and constructed the spatio-temporal feature vector to reflect the non-linear spatio-temporal correlations. Then, they deployed a CNN to extract important spatial correlation and LSTM to process the temporal correlation. The outputs of the two modules were regard as spatio-temporal correlations for final prediction [8]. Another similar hybrid deep learning model is for the non-Euclidean structure data. Qi et al. developed a model of Graph Convolutional Neural Network- Long Short-Term Memory (GCN-LSTM) to predict air pollutant concentration and obtained better results, compared to CNN-LSTM model [68].

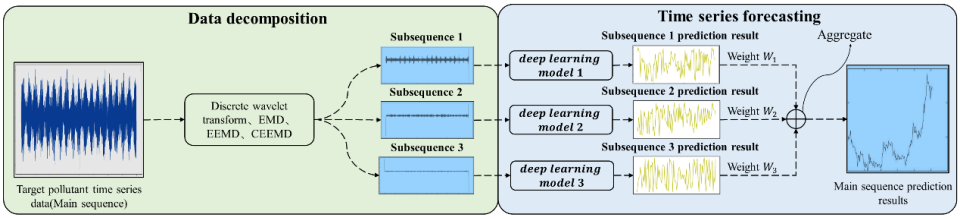

Figure 3: Framework for the hybrid method based on the data decomposition and deep learning methods

3.4. Deep learning plus other technologies

Deep learning approaches have outperformed non-deep learning methods in pollutant concentration prediction tasks and received a lot of attention, but is still in a developmental stage, and prediction accuracy is constantly improving. Since the deep learning approaches are constrained by their structure and training data, e.g., the problem of handling non-stationarity data [84, 98, 100, 102, 112, 121, 138], data shortage [119, 139, 140] and model structure constraints [85, 117, 141–144], it is not easy to fully reflect the advantages in the task of pollutant concentration prediction. Currently, many studies have improved methods based on deep learning, including improvements in input data, model structure and multi-type model mixing. For example, many researchers have integrated other technologies into the deep learning methods for air pollutant concentration prediction [84, 85, 98, 100, 102, 112, 117, 119, 121, 138–144] to fix these flaws.

Combining non-linear temporal correlation in historical pollutants concentration series, more and more deep learning methods are currently being applied to air pollution prediction. But most of them ignore the non-stationarity of time series, which leads to a lower accuracy of prediction [84]. A typical choice is to incorporate the data decomposition technology into the deep learning methods, as shown in Figure 3. It works in this way: the data decomposition module (e.g. Wavelet Transform (WT), Empirical Mode Decomposition (EMD) or their variants [99, 101]) decomposes the original time series into several sub-sequences and each sub-sequence is fed into its corresponding deep learning model to output its predicted result. Then all the predicted results of the sub-sequences are aggregated for obtaining the final prediction. Some studies have adopted this method and achieved excellent performance in air pollutant concentration prediction [84, 98, 100, 102, 112, 121, 138]. For example, Cabaneros et al. utilized a new hybrid method based on LSTM and the Discrete Wavelet Transform (DWT), and applied it to NO_2 prediction at six urban locations in Central London in the next hour [98]. Bai et al. took advantage of an Ensemble Empirical Mode Decomposition-Long Short-Term Memory (EEMD-LSTM) model for $PM_{2.5}$ concentration prediction. In this model, EEMD decomposes the original target pollutant data, and a single predictor LSTM outputs prediction on each sub-sequence. The final outcome is obtained by reconstructing the prediction results of sub-sequences [112].

In air pollutant prediction, lack of data in certain sites happens a lot and deep learning methods cannot perform well on insufficient training data. A good option is to integrate the transfer learning techniques into deep learning methods [119, 139, 140]. The transfer learning mechanism uses the similarity between two different but related datasets or models, and transfers the knowledge learned from the source domain to the target ones [145, 146]. It is particularly fits the situation when the number of samples is limited or the modeling process is complicated in the target domain. A deep learning model that has been well trained on sufficient data can be reused as the starting point of a model facing the similar task with insufficient data. Ma et al. exploited a model-based Transfer Stacked Bidirectional Long Short-Term Memory network (TLS-BLSTM) to predict air pollutants at new sites that lack data. A stacked BLSTM network had been pre-trained with large amounts of data from existing sites. Then some of its hidden layers were frozen, and the parameters in the remaining hidden layers were fine-tuned with a small amount of data from the new sites. In this way, the model retained the useful information learned from the existing sites, and boosted the prediction on the new sites [119].

Statistical methods can also be added to deep learning methods to improve the effectiveness of air pollutant concentration prediction [85, 117, 141–144]. For example, Kim et al. took advantage of the hierarchical clustering method to classify sites in the study region and then feature selection method was used to evaluate the importance of each input feature in deep learning model. Some redundant features irrelevant to prediction are eliminated. The addition of the clustering and feature selection methods improves the $PM_{2.5}$ prediction accuracy on 1-hour, 6-hour and 10-hour ahead [141]. Gu et al. integrated the Non-linear Auto Regressive Moving Average technique, automatic feature generation

Table 6
Literature classification for deep learning combined with other technologies for air pollutant concentration prediction

Other technology	Deep learning model	Literatures
WT and its variants	RNN	[98, 100, 102]
EMD and its variants	RNN, TCN	[84, 112, 121, 138]
Transfer learning	RNN	[119, 139, 140]
Statistical methods	Various deep learning methods	[85, 117, 141–144]

and feature selection methods into a deep learning model for $PM_{2.5}$ prediction. The experimental results demonstrate the superiority of proposed model over single deep learning model in prediction accuracy for peak values [142]. Table 6 lists typical literatures of deep learning combined with other technologies for air pollutant prediction.

4. PUBLIC DATASETS

High-quality datasets are important for accurate prediction of air pollutant concentration. In this section, we summarize public datasets for air pollutant prediction. The summary is presented in two parts: part A includes the published spatio-temporal sequential data and part B lists some auxiliary features that are often used to improve the accuracy of air pollutant prediction model. Due to different structures of models, these auxiliary features cannot be applied to all models.

4.1. Public datasets

We list currently well-maintained public datasets that store real data of air pollutant concentrations.

CNEMC The dataset of China National Environmental Monitoring Centre (CNEMC) belongs to the Ministry of Ecology and Environment of the People's Republic of China (http://www.mee.gov.cn/). CNEMC has collected air pollutant concentration data of 375 cities or regions and 2023 monitoring sites throughout China since Jan 1st, 2021. The source is available at: https://quotsoft.net/air/ or https://www.aqistudy.cn/historydata/.

UCI Machine Learning Repository The UCI Machine Learning Repository is a collection of databases that are used by research on machine learning algorithms. Three datasets of it, Beijing $PM_{2.5}$ Data, Beijing Multi-Site Air-Quality Data and $PM_{2.5}$ Data of Five Chinese Cities, record air pollutant concentrations. The source is available at: https://archive.ics.uci.edu/ml/datasets.php.

Beijing $PM_{2.5}$ **Data** It contains the hourly $PM_{2.5}$ data monitored at the US Embassy in Beijing, and meteorological data monitored at Beijing Capital International Airport from Jan 1st, 2010 to Dec 31st, 2014.

Beijing Multi-Site Air-Quality Data It records six air pollutants and six relevant meteorological variables that were hourly measured by 12 national monitoring sites in Beijing from Mar 1st, 2013 to Feb 28th, 2017.

$PM_{2.5}$ **Data of Five Chinese Cities** It gleaned $PM_{2.5}$ and meteorological data from Beijing, Shanghai, Guangzhou, Chengdu and Shenyang. from Jan 1st, 2010 to Dec 31st, 2015.

Urban Air Urban Air dataset is constructed by Microsoft Research. It contains data from four cities, Beijing, Tianjin, Guangzhou and Shenzhen, and 39 neighboring cities within 300km from them from May 1st, 2014 to Apr 30th, 2015. In total, there are 2,891,393 air quality records measured hourly by 437 air quality monitoring sites, 1,898,453 meteorology records collected hourly at a district (or city) level, and 910,576 weather prediction records of two coming days with a district (or city)- level and a temporal granularity of 3 hour, 6 hour, or 12 hour. The source is available at http://research.microsoft.com/apps/pubs/?id=246398.

AirNet AirNet dataset consists of data of SO_2, $PM_{2.5}$, PM_{10}, NO_2, CO, O_3 and AQI collected from 1498 sites as well as meteorological data in China. The data in AirNet are four-dimensional containing information of latitude, longitude, timesteps, features, ranging from Apr 1st, 2015 to Sep 1st, 2017. For each timestep, there are six meteorological features and seven air quality indices. A processed version is available at: http://airnet.caiyunapp.com.

G.Zou This dataset contains historical pollutant concentration and meteorological data from monitoring sites in 10 cities collected from May 13th, 2014 to May 30th, 2018 (http://github.com/zouguojian/data). The data in it is city level, that is, the sample data of each city of every hour is a one-dimensional feature vector composed of seven pollutants, including SO_2, $PM_{2.5}$, PM_{10}, NO_2, CO, O_3 and AQI, and nine meteorological factors. Ten cities throughout China are selected, Shanghai, Nanjing, Hangzhou, Wuhan, Beijing, Shenyang, Harbin, Chengdu, Wulumuqi, and Lasa.

DEFRA It depicts the whole air quality picture of the United Kingdom, containing pollutant concentration (SO_2, $PM_{2.5}$, PM_{10}, NO_2, O_3) data collected from more than 1500 sites. There are two major types: automatic monitoring sites and non-automatic monitoring sites. The oldest data from the automatic monitoring sites go back to 1972, and the oldest from the non-automatic monitoring sites trace back to 1961. The source is available at: https://uk-air.defra.gov.uk.

EPD EPD dataset contains data collected from 18 air quality monitoring sites in Hong Kong. These sites can be divided into two categories, general sites and roadside sites. Each monitoring site records the hourly, daily mean, monthly average and yearly average air pollutant concentration (SO_2, $PM_{2.5}$, PM_{10}, NO_2, CO, O_3 and NO_x) data, from 1990 to 2021. The data are available at: https://cd.epic.epd.gov.hk/EPICDI/air/station/?lang=en.

US EPA It stores daily air pollutant concentration data (SO_2, $PM_{2.5}$, PM_{10}, NO_2, CO, O_3 and Pb) from outdoor monitors across the United States, Puerto Rico, and the U. S. Virgin Islands from 1980-2021. The source is available at: https://www.epa.gov/.

TW EPA This dataset contains hourly data of SO_2, $PM_{2.5}$, PM_{10}, NO_2, CO and O_3 from more than 75 monitoring sites in the Taiwan Main Island, ranging from 2018 to 2022. The source is available at: https://www.epa.gov.tw/.

Air Korea Air Korea dataset collects air pollutant concentration data from monitoring points in various cities in South Korea, including major cities like Seoul and Busan. The web address of the paper providing the relevant data is: https://www.mdpi.com/2071-1050/12/19/8014.

CPCB CPCB dataset collects data from 703 air quality monitoring sites. These sites are located in 307 cities/towns of 29 states and 6 Union Territories in India. The time interval for the collected data is 15 minutes. The The URL of the paper providing the data is: https://www.sciencedirect.com/science/article/pii/S2212095521000304.

4.2. Auxiliary features

Prediction on air pollutant concentrations can be affected by some complicated factors, usually called auxiliary features. We list some auxiliary features considered by most researchers.

Meteorological factors: Studies have shown that air pollutants vary under different meteorological conditions, e.g., the temperature affects the atmospheric and ventilation conditions; humidity and precipitation can change the deposition characteristics of particulate matter; wind speed promotes the diffusion and spread of pollutants [147, 148]. In addition, the following meteorological factors can also cause different degrees of impact: dew point, pressure, wind direction, rainfall, snowfall, solar radiation, net radiation, clouds, conds, visibility, weather [11].

Chemical mechanism factors: The formation process of air pollutants is quite complex and variable, containing several chemical reactions. For example, Nitrogen oxides and volatile organic compounds can interact photochemically under favorable meteorological conditions to produce surface ozone [149]. Therefore, chemical reactions between pollutants and between pollutants and other substances can change pollutant concentrations. Some chemical materials can be taken into account when making air pollution predictions, e.g., Nitrous oxide, Ammonia, Ozone, Sulfur dioxide, Carbon dioxide, Carbon monoxide, Nitrogen dioxide, $PM_{2.5}$, PM_{10}, Organic compound [67, 89, 109].

Traffic factors: Traffic emissions have been one of the contributors to air pollution in many cities around the world and can deteriorate ambient air quality on a large spatial scale, especially during the morning and evening rush hours in urban regions [150]. With this in mind, the researchers included traffic-related data in their air pollution predictions [151, 152].

POI factors: point of interests (POIs) refer to the unique features of a particular region. The category and density of POIs in the region indicates the land use and function of the region, thus contributing to the prediction of air pollution in this region. Some POIs are even directly related to changes in air pollution. For example, if an region is industrial, its air quality is usually bad; if a park exists in an region, its air quality will become good [6]. Some of the POIs categories that are frequently used by researchers are vehicle services (gas stations, repair), factories, decoration and furniture markets, food and beverage, shopping malls and supermarkets, parks, etc. [6].

Regional geographic factors: Studies have shown that air pollution in sites is highly correlated. when the right neighboring districts are selected, or when the distance-based selection of neighboring regions is applied, the geographic prediction models are superior to non-geographic models [65]. These geographical factors relate to: the longitude and the latitude of air quality ground monitoring sites or cities, the distance between monitoring sites or cities [68].

Temporal factors: The purpose of introducing temporal factors is to allow the model to better identify the patterns of change in the temporal dimension of historical data and thus accurately capture the dynamic temporal correlation. The main temporal factors include: seasons, days of a week, time of a day [66, 121]:

1) The monitoring season is mapped using one to four.

2) Days of a week includes weekdays (Monday to Friday) and weekends (Saturday and Sunday) according to their different characteristics.

3) Time of the day is represented by 24 time points indicating 24 hours of a day.

5. EXPERIMENTAL ANALYSIS

In this section, we compare the experiments that the deep learning and non-deep learning methods have conducted on the public datasets and draw some conclusions from the comparison.

5.1. Model performance evaluation indicator

Two indicators are usually used to evaluate prediction performance of the models: Rooted Mean Squared Error (RMSE) and Mean Absolute Error (MAE). They are defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \tag{8}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| \tag{9}$$

where y_i is the observed value of the air pollutant concentration, while $\hat{y}_i$ is the predicted value; and n is the number of samples in the test set. RMSE measures models stability: smaller value implies higher stability. MAE indicates the total error between the predicted value and the observed one. The smaller value of MAE reflects the higher prediction accuracy of the model.

5.2. Experimental analysis on public datasets

In the pollutant concentration prediction task, the dataset is generally divided into the training set, validation set, and test set, or training set and test set. The training set is used for model learning, the validation set is used to determine hyperparameters, and the test set is used to verify the model's performance or to determine hyperparameters. For the model parameter selection process, non-deep learning methods are initialized by prior knowledge and experience, while deep learning methods are trained by continuously trying the settings of hyperparameters to select the optimal hyperparameter combination. There will be differences in hyperparameters between models for the same dataset, but they are all close to optimal for each study. Therefore, there are inevitably differences in hyperparameter configurations, but this does not affect comparing the best prediction performance between these methods. We summarized deep learning and non-deep learning models that have been tested on the 11 public datasets. Non-deep learning models include two main categories: statistical and deterministic methods. The final results are listed in Table 7 and Table 8, where the prediction granularity in Table 7 is hours and in Table 8 is days.

According to the different pollutant concentration prediction tasks, we divide the prediction tasks into single- and multi-step prediction, as shown in Figure 4. Where x_{t-r+1} denotes the historical observations of the past $t - r + 1$ timestep, used as inputs to the model. $\hat{y}_{t+n}$ and $\hat{y}_{t+m}$ are the model's predicted values for future timestep $t + n$ and $t + m$, respectively. Some models output single value for predicting air pollution concentration at one timestep in the future, which is called single-step prediction. Multi-step prediction means that models outputs multiple values that are used to predict target pollutant concentrations at multiple timesteps ahead simultaneously. Furthermore, we unified single- and multi-step prediction tasks for target pollutants under the public datasets to provide a more valuable comparison. These selected prediction tasks are the most studied by researchers and representative. Based on the experimental results on the public datasets [8, 10, 11, 24, 60–63, 81, 84, 93, 95, 105, 117–119, 121, 153–184], some interesting findings can be drawn from Tables 7 and 8:

1) For the prediction of the same target pollutant concentration on the same dataset, the single-step prediction performed better than the multi-step prediction. Tables 7 and 8 show that whether deep learning, statistical or deterministic models, their single-step predictions for target pollutants in all datasets almost outperformed multi-step predictions. Especially for the $PM_{2.5}$ prediction under the Urban Air dataset, the RMSE value of deep learning, statistical, and deterministic models in single-step prediction of the next hour was reduced by 70.08%, 68.76% and 80.46% and the MAE

Table 7
Prediction performance of different models on public datasets at hourly granularity

Dataset	Air pollutants	Prediction task		RMSE			MAE		
		Single-prediction	Multi-prediction	DL Model	Statistic Model	Deterministic Model	DL Model	Statistic Model	Deterministic Model
CNEMC	$PM_{2.5}$	T+1h	T+(1-24) h	1.11[121]/17.83[95]	5.26[121]/80.16[157]	14.75[181]/35.80[24]	0.66[121]/9.68[95]	3.49[121]/55.11[157]	11.42[181]/29.82[160]
	PM_{10}	T+1h	T+(1-5) h	7.13[155]/7.98[155]	20.70[156]/57.50[156]	71.50[154]/N	16.55[174]/23.72[174]	15.50[156]/35.00[156]	N
	O_3	T+1h	T+(1-24) h	10.04[119]/22.08[161]	14.56[119]/N	26.11[105]/N	8.09[119]/16.11[161]	21.84[162]/N	22.61[105]/28.17[160]
	NO_2	T+1h	T+(1-12) h	6.65[62]/19.55[172]	7.71[62]/16.00[173]	N/17.13[159]	4.56[62]/13.82[172]	5.46[62]/15.00[173]	N/12.86[159]
Air Net	NO_2	T+1h	T+(1-24) h	1.63[60]/9.72[164]	2.12[60]/15.14[164]	N	1.30[60]/7.04[164]	1.63[60]/10.60[164]	N
US EPA	$PM_{2.5}$	T+1h	\	3.48[117]	4.67[117]	7.36[177]	1.85[117]	2.68[117]	9.80[176]
UCI	$PM_{2.5}$	T+1h	T+(1-2) h	11.23[84]/11.83[84]	20.84[84]/27.64[165]	N	6.46[84]/7.70[84]	11.44[84]/16.72[165]	N
Urban Air	$PM_{2.5}$	T+1h	T+(1-25) h	11.31[10]/37.80[10]	15.50[10]/49.61[10]	23.21[167]/118.77[167]	6.52[10]/26.04[10]	10.62[10]/30.05[10]	8.15[167]/57.45[166]
G .Zou	$PM_{2.5}$	T+1h	T+(1-24) h	5.60[11]/22.30[11]	19.02[11]/41.80[11]	34.09[11]/36.3[11]	3.20[11]/15.50[11]	13.28[11]/30.00[11]	N
DEFRA	$PM_{2.5}$	T+1h	\	3.45[175]	7.31[175]	N	2.35[175]	5.48[175]	N
Air Korea	$PM_{2.5}$	T+24h	\	6.93[93]	13.49[93]	21.16[169]	5.07[93]	9.03[93]	7.38[169]
CPCB	$PM_{2.5}$	T+1h	T+(1-24) h	6.67[153]/12.00[61]	12.98[81]/33.00[61]	10.99[170]/50.00[171]	3.94[153]/12.00[61]	10.31[81]/31.00[61]	9.79[170]/N

Table 8
Prediction performance of different models on public datasets at daily granularity

Dataset	Air pollutants	Prediction task		RMSE			MAE		
		Single-prediction	Multi-prediction	DL Model	Statistic Model	Deterministic Model	DL Model	Statistic Model	Deterministic Model
CNEMC	$PM_{2.5}$	T+1day	T+(1-8) days	3.00[8]/43.14[178]	20.28[8]/48.10[178]	46.50[179]/N	2.21[8]/29.15[178]	14.98[8]/30.77[178]	N
	O_3	T+1day	T+(1-3) days	17.43[158]/18.13[158]	26.74[158]/30.91[158]	27.90[24]/27.03[24]	12.35[158]/13.02[158]	20.38[158]/23.29[158]	21.90[24]/21.10[24]
	O_3	T+1day	\	16.47[118]	18.32[118]	19.30[182]	12.75[118]	13.67[118]	N
EPD	NO_2	T+1day	\	14.10[118]	15.96[118]	16.70[182]	10.35[118]	11.32[118]	N
	$PM_{2.5}$	T+1day	\	6.74[118]	7.46[118]	8.60[182]	4.92[118]	5.19[118]	N
UCI	$PM_{2.5}$	T+1day	T+(1-9) days	17.20[161]/12.00[63]	19.90[161]/57.00[63]	N	11.90[161]/10.00[63]	12.30[161]/52.00[63]	N
CPCB	$PM_{2.5}$	T+1day	T+(1-9) days	10.60[163]/15.00[63]	12.35[168]/68.00[63]	131.13[180]/N	5.45[163]/12.00[63]	8.52[168]/58.00[63]	N
TW EPA	$PM_{2.5}$	T+1day	T+(1-3) days	7.12[183]/7.97[183]	9.72[184]/N	10.03[183]/10.15[183]	4.59[183]/5.27[183]	6.09[184]/N	7.02[183]/7.14[183]

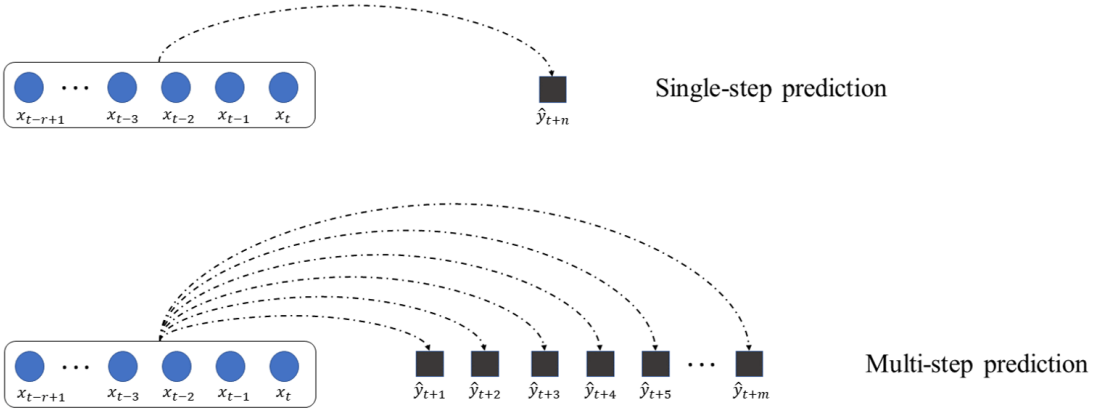

Figure 4: Single-step and multi-step prediction.

value was reduced by 75.02%, 64.66% and 85.81% respectively, in contrast with their multi-step prediction of the next 1 to 25 hours. The performance of multi-step prediction is lower than that of single-step prediction, and the reasons can be summarized in two aspects: on the one hand, since the period of multi-step prediction is much larger than that of single-step prediction, the historical data that needs to be referenced is more and broader. On the other hand, there is an error propagation situation in multi-step prediction, and as the prediction timestep increases, the prediction error will gradually accumulate.

2) Deep learning methods perform better than non-deep learning ones on single- or multi-step predictions. As can be seen from Table 7, when the prediction granularity is hours, the RMSE and MAE values of the deep learning model decrease in different degrees compared to the statistical and deterministic models for both single-step and multi-step prediction. For example, in the 1-hour ahead prediction of $PM_{2.5}$ on the US EPA dataset, the RMSE values of the deep learning models were reduced by 25.48% and 52.72% compared to the statistical and deterministic models, respectively. The MAE values decreased by 30.97% and 81.12%, respectively, compared to the other two types of models. Similarly, it can be seen from Table 8 that, in the 1-3 days ahead prediction of O_3 on the CNEMC dataset, the RMSE values of the deep learning models decreased by 41.35% and 32.93%, and the MAE values reduced by 44.10% and 38.29%, compared to the statistical and deterministic models. The main reasons are the following: First, non-deep learning methodsare inept at determining the complex, non-linear relationships in data due to their limited

Table 9
Predictive performance of different models on commonly used public datasets under the influence of seasonality

Method	Dataset	Target pollutants	Prediction task	Performance Index (RMSE/MAE)			
				Spring	Summer	Autumn	Winter
DL Model			T+1 h	10.28[66]/8.01[66]	6.70[66]/5.60[66]	23.97[66]/17.65[66]	34.56[66]/24.76[66]
Statistic Model	**CNEMC**	$PM_{2.5}$	T+1 h	2.14[185]/1.51[185]	3.20[185]/2.59[185]	5.45[185]/4.54[185]	9.61[185]/7.61[185]
Deterministic Model			T+(1-24) h	17.05[186]/N	17.04[186]/N	18.18[186]/N	13.67[186]/N
DL Model			T+1 day	9.83[187]/N	11.91[187]/N	10.13[187]/N	8.79[187]/N
Statistic Model	**EPD**	O_3	T+1 day	14.10[188]/10.64[188]	N	21.38[188]/14.23[188]	8.48[188]/6.78[188]
Deterministic Model			\	N	N	N	N
DL Model			T+1 h	9.17[189]/N	7.14[189]/N	7.15[189]/N	9.06[189]/N
Statistic Model	**UCI**	$PM_{2.5}$	T+(1-2) h	20.58[190]/16.11[190]	17.61[190]/7.83[190]	28.42[190]/25.56[190]	27.65[190]/22.99[190]
Deterministic Model			\	N	N	N	N
DL Model			T+1 h	3.800-3.825[153]/N	3.825-3.850[153]/N	3.775-3.800[153]/N	3.850-3.875[153]/N
Statistic Model	**CPCB**	$PM_{2.5}$	T+1 day	18.26[191]/15.33[191]	18.26[191]/15.33[191]	22.97[191]/17.28[191]	17.29[191]/13.56[191]
Deterministic Model			T+1 day	40.80[192]/14.57[192]	49.90[192]/11.80[192]	46.94[192]/14.43[192]	38.57[192]/14.27[192]

Table 10
Performance comparison of hybrid models and single models with commonly used public datasets among deep learning methods

Dataset	Target pollutant	Prediction task	Model	RMSE	MAE
CNEMC	$PM_{2.5}$	T+(1-24) h	C-LSTME [95]	17.83	9.68
			LSTME [95]	29.61	14.20
EPD	O_3	T+1 day	SA-LSTM [118]	16.47	12.75
			LSTM, GRU, Original RNN [118]	17.85	13.55
G.Zou	$PM_{2.5}$	T+(1-24) h	EDS Model [11]	22.30	15.50
			GRU, RNN, LSTM, Bi-LSTM [11]	36.25	25.43
UCI	$PM_{2.5}$	T+1 h	EMD-GRU [84]	11.23	6.46
			RNN, LSTM, GRU [84]	20.76	11.14
US EPA	$PM_{2.5}$	T+1 h	Lag-LSTM [117]	3.48	1.85
			RNN, LSTM [117]	4.59	2.52

structure; second, non-deep learning methods are not suitable for large-scale datasets because they are restricted by model complexity. For deep learning methods, their flexible and complex structures help them break through the above constraints of non-deep learning methods. They perform better both on single-step and multi-step prediction for air pollutant concentration.

In addition, we have done research on two other perspectives. We investigated the impact of seasonal changes in pollutants on the prediction performance of these models. The performance indicators of the models which act on the target pollutants of the commonly used public datasets are summarized in four seasons: spring, summer, autumn and winter. The final results are presented in Table 9. In the category of deep learning models, we also explored the performance of hybrid models compared to single models on the target pollutants of the commonly used public datasets. The results are shown in Table 10.

Studies have shown that the change of seasons can have an impact on air pollution [66, 193]. Different seasons are accompanied by different climatic conditions and human activities such as low temperature and low wind speed winter atmospheric conditions, coal burning and firecrackers in winter, which lead to changes in air pollution [194]. Therefore, researchers usually consider seasonal influence factors when conducting air pollution prediction studies. This is done by dividing the dataset into four datasets: spring, summer, autumn, and winter, and then exploring the predictive performance of deep learning and non-deep learning models on each of the four datasets, as shown in Table 9 [66, 153, 185–192]. As a result, the values of RMSE and MAE show that some researchers have improved the prediction accuracy by including seasonal effects as a reference whether using deep learning, statistical or deterministic models. For example, combining Tables 7 and 9, in the prediction of $PM_{2.5}$ for the next hour based on the UCI dataset, the best prediction performance of the deep learning model occurs in the summer prediction [189]; compared to the deep learning model without considering the seasonal effect [84], the RMSE value decreased by 36.42%. This proves that seasonal variation has some effect on the prediction performance of the three models.

Among deep-learning methods, the hybrid methods formed by multi-deep learning models often perform better than those with a single deep learning model for air pollutant concentration prediction on single- and multi-step tasks on commonly used public datasets [11, 95, 118]. For example, as shown in Table 10, the Convolutional and Long Short-

Term Memory Neural Network Extended model (C-LSTME) [95] consisting of 3D-CNN and LSTME outperformed single LSTME model on CNEMC dataset for 1-24 hours ahead prediction task facing $PM_{2.5}$, with RMSE and MAE reduction of 39.78% and 31.83%, respectively. Methods with a single deep learning model generally consider only one aspect of factors, temporal or spatial, for prediction; while the hybrid models can often extend the strengths of single deep learning models and simultaneously extract complex and non-linear spatio-temporal correlations. In addition, hybrid models integrating deep-learning models and other technologies can also surpass methods with a single deep learning model in some prediction tasks [84, 117]. From Table 10, a hybrid model based on EMD and GRU defeated a single RNN, LSTM, and GRU network when predicting $PM_{2.5}$ concentrations at sites in Beijing in the next hour with RMSE and MAE values reduced by 45.91% and 42.10%, respectively [84]. This is because deep learning models are not foolproof due to the limitations of input data, parameter optimization, model structure. In some cases they don't work very well [84, 85, 98, 100, 102, 112, 117, 119, 121, 138–144]. Thus, in the prediction process, we can consider introducing other techniques to make up for the deficiencies of deep learning models in these aspects, improving the prediction performance.

6. LIMITATIONS AND FUTURE DIRECTIONS

Although deep learning technology has made great progress in air pollutant concentration prediction, there are still some challenges worth exploring. In this Section, we discuss the limitations of current studies and identify some directions for future research.

6.1. Limitations

Data shortage Models based on deep learning require a large amount of data to achieve excellent prediction results, but data for air pollution prediction are not always sufficient. For example, extreme situations, like inclement weathers or equipment maintenance, often interrupt data collection; and air-related records in some areas are gleaned intermittently due to their under-developed air-quality monitoring systems. Insufficient training data have weakened the performance of deep-learning models. Some researchers have utilized the transfer learning technology to tackle this problem [119, 139, 140]. This technique aims to transfer knowledge from regions with rich data sources to target regions with scarce data. Despite its effectiveness, it still needs further exploration, such as, how to design a high-quality mathematical model to match two regions, etc., still need to be explored.

Long-term prediction Most deep learning methods focus on short- or medium- term prediction of air quality, and a few of them delve into long-term tasks [106]. Long-term prediction involves more complex and flexible spatio-temporal correlations and more uncertain factors. The principle of conducting it also differ from the short or medium tasks: the effect of historical data may decrease and multiple, still-unclear features may be encompassed.

High-dimensional data Most raw data that current deep-learning methods use for air-quality assessment are in a single dimension, and other impact factors, like, land usage, forest coverage rate, populations, etc., have not been sufficiently considered. Although higher-dimensional data that fuse multi-factors can elevate air pollution prediction, the costs are huge. Current algorithms cannot afford "curse of dimensionality" and expensive computational consumption, and how to choose appropriate number of suitable factors is a problem to be solved.

Baseline for performance evaluation Given the flexible but similar architecture, most deep-learning methods work and construct in a similar way for air-quality assessment. It is quite difficult to compare these models in a more comprehensive and detailed way without a standardized baseline that consists of definitive experiment settings and datasets. Meanwhile, as the complexity of the models is increasing, the efficacy of each component has become more ambiguous, although ablation studies can be used for discussion.

Interpretability problem Models based on deep learning have been suffering the black box problem from the day their were born. Their working procedures are uninterpretable due to their complex structure and a large number of parameters. Although they output more accurate prediction results than non-deep learning models, the reason for this and the procedures for determining parameters are still unclear. How to construct a deep learning model with high interpretability is a problem urgently to be solved.

Expensive computing costs Models based on deep learning usually require GPU or TPU to fulfill their computation instead of CPUs. In addition, with the rising number of parameters, they also cost much more memory. How to ensure that deep learning models do not consume too much computational resources with the same training and testing results deserves to be further explored.

6.2. Future directions

Selection of optimal model As various methods can be a choice for one prediction task, a set of well-accepted criteria for the selection is still required. At present, RNNs are usually used to deal with one-location historical data, and CNNs, GCNs and Spatial Attention Mechanism are commonly developed for multi-location data. But how to choose an appropriate one for a new task to obtain the possible max prediction performance is still an open question.

Prediction under noises Some raw data of air quality recorded by monitoring sites usually deviate from true vales due equipment failures or environmental impact. These noise-contaminated data may influence the final prediction of deep-learning methods. Currently, data processing and model prediction are separated as two independent procedures to avoid noise impact. It is of great significance to build a robust model that could effectively tackle this problem.

Intergradation into the Internet of Things in smart cities Smart city is an intelligent system that connects all independent units in a city together through the technology of Internet of Things (IoTs) [195]. This system can collect and process large amounts of real-time data and help decision makers present better strategies [196]. Integrating air quality prediction function into a smart city is a promising direction for the research of deep learning-based air pollution prediction.

Diversified data collection National monitoring sites are major source of air quality data. It is not enough and some researchers have proposed new devices for collecting data, such as smart mobile pollution monitoring devices and low-cost air pollution sensors [197]. These devices are promising, and encompassing their data into the historical information will improve the accuracy of deep-learning prediction. The diversified data can also compose a massive knowledge database for further exploitation.

Application of distributed computing Although the deep learning model has made great achievements in air pollutant concentration prediction, its slow training speed is still a problem. Distributed computing has been verified an effective alternative to improve the training speed of these models [198]. For example, Apache Spark framework [199] is a popular tool that can process large-scale data in a distributed way. It might be promising to combine distributed computing with deep learning technology for air pollutant concentration prediction.

7. CONCLUSION

We conduct a comprehensive review on deep learning methods for air pollutant concentration prediction. We classify the existing methods and review the air pollutant concentration prediction based on non-deep learning from the aspects of characteristic, application and limitation of prediction method. Furthermore, we summarize the methods based on deep learning from perspectives of temporal, spatial, spatio-temporal and its variants. We enumerate the major public datasets and representative experimental results in different prediction tasks based on listed datasets to compare the prediction performance of non-deep learning and deep learning methods and draw some conclusions. Finally, we put forward some limitations and future research directions that are worthy of reference. This paper provides an overview on the research progress of air pollutant concentration prediction, and could be a useful reference for researchers in this field.

Acknowledgement

This work is funded by National Natural Science Foundation of China (61572326, 61802258, 61702333), Natural Science Foundation of Shanghai (18ZR1428300), the Shanghai Committee of Science and Technology (17070502800).

References

[1] Jun Song and Marc EJ Stettler. A novel multi-pollutant space-time learning network for air pollution inference. *Science of The Total Environment*, 811:152254, 2022.

[2] Yousef SH Najjar et al. Gaseous pollutants formation and their harmful effects on health and environment. *Innovative energy policies*, 1:1–9, 2011.

[3] Fanglin Chen and Zhongfei Chen. Cost of economic growth: Air pollution and health expenditure. *Science of the Total Environment*, 755:142543, 2021.

[4] Marilena Kampa and Elias Castanas. Human health effects of air pollution. *Environmental pollution*, 151(2):362–367, 2008.

[5] Xiao Han, Yuqin Liu, Hong Gao, Jianmin Ma, Xiaoxuan Mao, Yuting Wang, and Xudong Ma. Forecasting pm2. 5 induced male lung cancer morbidity in china using satellite retrieved pm2. 5 and spatial analysis. *Science of the Total Environment*, 607:1009–1017, 2017.

[6] Yu Zheng, Furui Liu, and Hsun-Ping Hsieh. U-air: When urban air quality inference meets big data. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1436–1444, 2013.

[7] Qunli Wu and Huaxing Lin. A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors. *Science of the Total Environment*, 683:808–821, 2019.

[8] Unjin Pak, Jun Ma, Unsok Ryu, Kwangchol Ryom, U Juhyok, Kyongsok Pak, and Chanil Pak. Deep learning-based pm2. 5 prediction considering the spatiotemporal correlations: A case study of beijing, china. *Science of The Total Environment*, 699:133561, 2020.

[9] Jiaqi Zhu, Fang Deng, Jiachen Zhao, and Hao Zheng. Attention-based parallel networks (apnet) for pm2. 5 spatiotemporal prediction. *Science of The Total Environment*, 769:145082, 2021.

[10] Yu Huang, Josh Jia-Ching Ying, and Vincent S Tseng. Spatio-attention embedded recurrent neural network for air quality prediction. *Knowledge-Based Systems*, 233:107416, 2021.

[11] Bo Zhang, Guojian Zou, Dongming Qin, Yunjie Lu, Yupeng Jin, and Hui Wang. A novel encoder-decoder model based on read-first lstm for air pollutant prediction. *Science of The Total Environment*, 765:144507, 2021.

[12] Mengfan Teng, Siwei Li, Jia Xing, Ge Song, Jie Yang, Jiaxin Dong, Xiaoyue Zeng, and Yaming Qin. 24-hour prediction of pm2. 5 concentrations by combining empirical mode decomposition and bidirectional long short-term memory neural network. *Science of The Total Environment*, 821:153276, 2022.

[13] Junbo Zhang, Yu Zheng, Dekang Qi, Ruiyuan Li, and Xiuwen Yi. Dnn-based prediction model for spatio-temporal data. In *Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems*, pages 1–4, 2016.

[14] Lu Bai, Jianzhou Wang, Xuejiao Ma, and Haiyan Lu. Air pollution forecasts: An overview. *International journal of environmental research and public health*, 15(4):780, 2018.

[15] A Masih. Machine learning algorithms in air quality modeling. *Global Journal of Environmental Science and Management*, 5(4):515–534, 2019.

[16] Sheen Mclean Cabaneros, John Kaiser Calautit, and Ben Richard Hughes. A review of artificial neural network models for ambient air pollution prediction. *Environmental Modelling & Software*, 119:285–304, 2019.

[17] Qi Liao, Mingming Zhu, Lin Wu, Xiaole Pan, Xiao Tang, and Zifa Wang. Deep learning for air quality forecasts: a review. *Current Pollution Reports*, 6(4):399–409, 2020.

[18] Adil Masood and Kafeel Ahmad. A review on emerging artificial intelligence (ai) techniques for air pollution forecasting: Fundamentals, application and performance. *Journal of Cleaner Production*, 322:129072, 2021.

[19] Hui Liu, Guangxi Yan, Zhu Duan, and Chao Chen. Intelligent modeling strategies for forecasting air quality time series: A review. *Applied Soft Computing*, 102:106957, 2021.

[20] Stephen F Mueller and Jonathan W Mallard. Contributions of natural emissions to ozone and pm2. 5 as simulated by the community multiscale air quality (cmaq) model. *Environmental science & technology*, 45(11):4817–4823, 2011.

[21] Tin Thongthammachart, Shin Araki, Hikari Shimadera, Shinnosuke Eto, Tomohito Matsuo, and Akira Kondo. An integrated model combining random forests and wrf/cmaq model for high accuracy spatiotemporal pm2. 5 predictions in the kansai region of japan. *Atmospheric Environment*, 262:118620, 2021.

[22] Yasmin Kaore Lago Kitagawa, Rizzieri Pedruzzi, Elson Silva Galvão, Igor Baptista de Araújo, Taciana Toledo de Almeida Alburquerque, Prashant Kumar, Erick Giovani Sperandio Nascimento, and Davidson Martins Moreira. Source apportionment modelling of pm2. 5 using cmaq-isam over a tropical coastal-urban area. *Atmospheric Pollution Research*, 12(12):101250, 2021.

[23] Pengfei Wang, Peng Wang, Kaiyu Chen, Jun Du, and Hongliang Zhang. Ground-level ozone simulation using ensemble wrf/chem predictions over the southeast united states. *Chemosphere*, 287:132428, 2022.

[24] Guangqiang Zhou, Jianming Xu, Ying Xie, Luyu Chang, Wei Gao, Yixuan Gu, and Ji Zhou. Numerical air quality forecasting over eastern china: An operational application of wrf-chem. *Atmospheric Environment*, 153:94–108, 2017.

[25] Ming-Tung Chuang, Yang Zhang, and Daiwen Kang. Application of wrf/chem-madrid for real-time air quality forecasting over the southeastern united states. *Atmospheric environment*, 45(34):6241–6250, 2011.

[26] ZiFa Wang, Jie Li, Zhe Wang, WenYi Yang, Xiao Tang, BaoZhu Ge, PinZhong Yan, LiLi Zhu, XueShun Chen, HuanSheng Chen, et al. Modeling study of regional severe hazes over mid-eastern china in january 2013 and its implications on pollution prevention and control. *Science China Earth Sciences*, 57(1):3–13, 2014.

[27] Z Wang, T Maeda, M Hayashi, L-F Hsiao, and K-Y Liu. A nested air quality prediction modeling system for urban and regional scales: Application for high-ozone episode in taiwan. *Water, Air, and Soil Pollution*, 130(1):391–396, 2001.

[28] P Konopka, J-U Grooß, G Günther, F Ploeger, R Pommrich, R Müller, and N Livesey. Annual cycle of ozone at and above the tropical tropopause: observations versus simulations with the chemical lagrangian model of the stratosphere (clams). *Atmospheric Chemistry and Physics*, 10(1):121–132, 2010.

[29] MJ Assael, M Delaki, and KE Kakosimos. Applying the ospm model to the calculation of pm10 concentration levels in the historical centre of the city of thessaloniki. *Atmospheric Environment*, 42(1):65–77, 2008.

[30] Youn-Seo Koo, Dae-Ryun Choi, Hi-Yong Kwon, Young-Kee Jang, and Jin-Seok Han. Improvement of pm10 prediction in east asia using inverse modeling. *Atmospheric Environment*, 106:318–328, 2015.

[31] AMM Manders, M Schaap, and R Hoogerbrugge. Testing the capability of the chemistry transport model lotos-euros to forecast pm10 levels in the netherlands. *Atmospheric Environment*, 43(26):4050–4059, 2009.

[32] Xuexi Tie, Guy P Brasseur, ChunSheng Zhao, Claire Granier, Steven Massie, Yu Qin, PuCai Wang, Geli Wang, PeiCai Yang, and Andreas Richter. Chemical characterization of air pollution in eastern china and the eastern united states. *Atmospheric Environment*, 40(14):2607–2625, 2006.

[33] Robert Vautard, Peter HJ Builtjes, Phillipe Thunis, C Cuvelier, Marco Bedogni, Bertrand Bessagnet, Cécile Honore, Nicolas Moussiopoulos, Guido Pirovano, Martijn Schaap, et al. Evaluation and intercomparison of ozone and pm10 simulations by several chemistry transport models over four european cities within the citydelta project. *Atmospheric environment*, 41(1):173–188, 2007.

[34] R Stern, P Builtjes, M Schaap, R Timmermans, R Vautard, A Hodzic, M Memmesheimer, H Feldmann, E Renner, R Wolke, et al. A model inter-comparison study focussing on episodes with elevated pm10 concentrations. *Atmospheric Environment*, 42(19):4567–4588, 2008.

[35] Zongwei Ma, Sagnik Dey, Sundar Christopher, Riyang Liu, Jun Bi, Palak Balyan, and Yang Liu. A review of statistical methods used for developing large-scale and long-term pm2. 5 models from satellite data. *Remote Sensing of Environment*, 269:112827, 2022.

[36] Yanzhao Li, Ju-e Guo, Shaolong Sun, Jianing Li, Shouyang Wang, and Chengyuan Zhang. Air quality forecasting with artificial intelligence techniques: A scientometric and content analysis. *Environmental Modelling & Software*, page 105329, 2022.

[37] Lanyi Zhang, Jane Lin, Rongzu Qiu, Xisheng Hu, Huihui Zhang, Qingyao Chen, Huamei Tan, Danting Lin, and Jiankai Wang. Trend analysis and forecast of pm2. 5 in fuzhou, china using the arima model. *Ecological indicators*, 95:702–710, 2018.

[38] Sivaraman Balachandran, Howard H Chang, Jorge E Pachon, Heather A Holmes, James A Mulholland, and Armistead G Russell. Bayesian-based ensemble source apportionment of pm2. 5. *Environmental science & technology*, 47(23):13511–13518, 2013.

[39] XY Ni, Hong Huang, and WP Du. Relevance analysis and short-term prediction of pm2. 5 concentrations in beijing based on multi-source data. *Atmospheric environment*, 150:146–161, 2017.

[40] PJ García Nieto, F Sánchez Lasheras, E García-Gonzalo, and FJ de Cos Juez. Pm10 concentration forecasting in the metropolitan area of oviedo (northern spain) using models based on svm, mlp, varma and arima: A case study. *Science of the Total Environment*, 621:753–761, 2018.

[41] Zongwei Ma, Xuefei Hu, Lei Huang, Jun Bi, and Yang Liu. Estimating ground-level pm2. 5 in china using satellite remote sensing. *Environmental science & technology*, 48(13):7436–7444, 2014.

[42] Sahar Masmoudi, Haytham Elghazel, Dalila Taieb, Orhan Yazar, and Amjad Kallel. A machine-learning framework for predicting multiple air pollutants' concentrations via multi-target regression and feature selection. *Science of the Total Environment*, 715:136991, 2020.

[43] WC Leong, RO Kelani, and Z Ahmad. Prediction of air pollution index (api) using support vector machine (svm). *Journal of Environmental Chemical Engineering*, 8(3):103208, 2020.

[44] Yanlai Zhou, Fi-John Chang, Li-Chiu Chang, I-Feng Kao, Yi-Shin Wang, and Che-Chia Kang. Multi-output support vector machine for regional multi-step-ahead pm2. 5 forecasting. *Science of the Total Environment*, 651:230–240, 2019.

[45] Min Li, Wei-li Wang, Zhan-yong Wang, Yu Xue, et al. Prediction of pm2. 5 concentration based on the similarity in air quality monitoring network. *Building and Environment*, 137:11–17, 2018.

[46] Zhong-hua Li and Jun Yang. Pm-25 forecasting use reconstruct phase space ls-svm. In *2010 the 2nd conference on environmental science and information application technology*, volume 1, pages 143–146. IEEE, 2010.

[47] Bernardo S Beckerman, Michael Jerrett, Randall V Martin, Aaron van Donkelaar, Zev Ross, and Richard T Burnett. Application of the deletion/substitution/addition algorithm to selecting land use regression models for interpolating air pollution measurements in california. *Atmospheric environment*, 77:172–177, 2013.

[48] Zhouming He, Pengfei Liu, Xiaoxi Zhao, Xiaowei He, Junfeng Liu, and Yujing Mu. Responses of surface o3 and pm2. 5 trends to changes of anthropogenic emissions in summer over beijing during 2014–2019: A study based on multiple linear regression and wrf-chem. *Science of The Total Environment*, 807:150792, 2022.

[49] Mahanijah Md Kamal, Rozita Jailani, and Ruhizan Liza Ahmad Shauri. Prediction of ambient air quality based on neural network technique. In *2006 4th Student Conference on Research and Development*, pages 115–119. IEEE, 2006.

[50] Davor Z Antanasijević, Viktor V Pocajt, Dragan S Povrenović, Mirjana Đ Ristić, and Aleksandra A Perić-Grujić. Pm10 emission forecasting using artificial neural networks and genetic algorithm input variable optimization. *Science of the Total Environment*, 443:511–519, 2013.

[51] Herman Wahid, QP Ha, and H Nguyen 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*, volume 2011, 2011.

[52] Rui Yan, Jiaqiang Liao, Jie Yang, Wei Sun, Mingyue Nong, and Feipeng Li. Multi-hour and multi-site air quality index forecasting in beijing using cnn, lstm, cnn-lstm, and spatiotemporal clustering. *Expert Systems with Applications*, 169:114513, 2021.

[53] Bo Zhang, Hanwen Zhang, Gengming Zhao, and Jie Lian. Constructing a pm2. 5 concentration prediction model by combining auto-encoder with bi-lstm neural networks. *Environmental Modelling & Software*, 124:104600, 2020.

[54] Yoshua Bengio et al. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

[55] Tsung-Han Chan, Kui Jia, Shenghua Gao, Jiwen Lu, Zinan Zeng, and Yi Ma. Pcanet: A simple deep learning baseline for image classification? *IEEE transactions on image processing*, 24(12):5017–5032, 2015.

[56] Abdel-rahman Mohamed, Tara N Sainath, George Dahl, Bhuvana Ramabhadran, Geoffrey E Hinton, and Michael A Picheny. Deep belief networks using discriminative features for phone recognition. In *2011 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 5060–5063. IEEE, 2011.

[57] Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for modern deep learning research. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13693–13696, 2020.

[58] Chun-Yang Zhang, CL Philip Chen, Min Gan, and Long Chen. Predictive deep boltzmann machine for multiperiod wind speed forecasting. *IEEE Transactions on Sustainable Energy*, 6(4):1416–1425, 2015.

[59] Ronan Collobert and Jason Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, pages 160–167, 2008.

[60] Zhili Zhao, Jian Qin, Zhaoshuang He, Huan Li, Yi Yang, and Ruisheng Zhang. Combining forward with recurrent neural networks for hourly air quality prediction in northwest of china. *Environmental Science and Pollution Research*, 27(23):28931–28948, 2020.

[61] K Krishna Rani Samal, Ankit Kumar Panda, Korra Sathya Babu, and Santos Kumar Das. An improved pollution forecasting model with meteorological impact using multiple imputation and fine-tuning approach. *Sustainable Cities and Society*, 70:102923, 2021.

[62] Chengxin Zhang, Cheng Liu, Bo Li, Fei Zhao, and Chunhui Zhao. Spatiotemporal neural network for estimating surface no2 concentrations over north china and their human health impact. *Environmental Pollution*, page 119510, 2022.

[63] K Krishna Rani Samal, Korra Sathya Babu, and Santos Kumar Das. Multi-directional temporal convolutional artificial neural network for pm2. 5 forecasting with missing values: a deep learning approach. *Urban Climate*, 36:100800, 2021.

[64] Lloyd H Macatangay and Rowell M Hernandez. A deep learning-based prediction and simulator of harmful air pollutants: A case from the

philippines. In *2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC)*, pages 381–386. IEEE, 2020.

[65] Atakan Kurt and Ayşe Betül Oktay. Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. *Expert Systems with Applications*, 37(12):7986–7992, 2010.

[66] Yun Bai, Yong Li, Bo Zeng, Chuan Li, and Jin Zhang. Hourly pm2. 5 concentration forecast using stacked autoencoder model with emphasis on seasonality. *Journal of Cleaner Production*, 224:739–750, 2019.

[67] SA Abdul-Wahab. Ier photochemical smog evaluation and forecasting of short-term ozone pollution levels with artificial neural networks. *Process Safety and Environmental Protection*, 79(2):117–128, 2001.

[68] Yanlin Qi, Qi Li, Hamed Karimian, and Di Liu. A hybrid model for spatiotemporal forecasting of pm2. 5 based on graph convolutional neural network and long short-term memory. *Science of the Total Environment*, 664:1–10, 2019.

[69] Hongye Zhou, Feng Zhang, Zhenhong Du, and Renyi Liu. A theory-guided graph networks based pm2. 5 forecasting method. *Environmental Pollution*, 293:118569, 2022.

[70] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[71] Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, et al. Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743):195–204, 2019.

[72] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[73] Peter L Bartlett, Andrea Montanari, and Alexander Rakhlin. Deep learning: a statistical viewpoint. *Acta numerica*, 30:87–201, 2021.

[74] Nabin Rijal, Ravi Teja Gutta, Tingting Cao, Jerry Lin, Qirong Bo, and Jing Zhang. Ensemble of deep neural networks for estimating particulate matter from images. In *2018 IEEE 3rd international conference on image, Vision and Computing (ICIVC)*, pages 733–738. IEEE, 2018.

[75] Chao Zhang, Junchi Yan, Changsheng Li, Xiaoguang Rui, Liang Liu, and Rongfang Bie. On estimating air pollution from photos using convolutional neural network. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 297–301, 2016.

[76] Qirong Bo, Wenwen Yang, Nabin Rijal, Yilin Xie, Jun Feng, and Jing Zhang. Particle pollution estimation from images using convolutional neural network and weather features. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3433–3437. IEEE, 2018.

[77] Avijoy Chakma, Ben Vizena, Tingting Cao, Jerry Lin, and Jing Zhang. Image-based air quality analysis using deep convolutional neural network. In *2017 IEEE international conference on image processing (ICIP)*, pages 3949–3952. IEEE, 2017.

[78] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203*, 2013.

[79] Jeffrey L Elman. Distributed representations, simple recurrent networks, and grammatical structure. *Machine learning*, 7(2):195–225, 1991.

[80] V Athira, P Geetha, Rab Vinayakumar, and KP Soman. Deepairnet: Applying recurrent networks for air quality prediction. *Procedia computer science*, 132:1394–1403, 2018.

[81] Kunwar P Singh, Shikha Gupta, Atulesh Kumar, and Sheo Prasad Shukla. Linear and nonlinear modeling approaches for urban air quality prediction. *Science of the Total Environment*, 426:244–255, 2012.

[82] Jorge Loy-Benitez, Paulina Vilela, Qian Li, and ChangKyoo Yoo. Sequential prediction of quantitative health risk assessment for the fine particulate matter in an underground facility using deep recurrent neural networks. *Ecotoxicology and environmental safety*, 169:316–324, 2019.

[83] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[84] Guoyan Huang, Xinyi Li, Bing Zhang, and Jiadong Ren. Pm2. 5 concentration forecasting at surface monitoring sites using gru neural network based on empirical mode decomposition. *Science of The Total Environment*, 768:144516, 2021.

[85] Rui Xu, Xiaoling Deng, Hang Wan, Yanpeng Cai, and Xipeng Pan. A deep learning method to repair atmospheric environmental quality data based on gaussian diffusion. *Journal of Cleaner Production*, 308:127446, 2021.

[86] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[87] Xingcheng Lu, Yu Hin Sha, Zhenning Li, Yeqi Huang, Wanying Chen, Duohong Chen, Jin Shen, Yiang Chen, and Jimmy CH Fung. Development and application of a hybrid long-short term memory–three dimensional variational technique for the improvement of pm2. 5 forecasting. *Science of The Total Environment*, 770:144221, 2021.

[88] Giulia Ulpiani, Patrick Nzivugira Duhirwe, Geun Young Yun, and Mathew J Lipson. Meteorological influence on forecasting urban pollutants: Long-term predictability versus extreme events in a spatially heterogeneous urban ecosystem. *Science of The Total Environment*, 814:152537, 2022.

[89] Jiachen Zhao, Fang Deng, Yeyun Cai, and Jie Chen. Long short-term memory-fully connected (lstm-fc) neural network for pm2. 5 concentration prediction. *Chemosphere*, 220:486–492, 2019.

[90] Ricardo Navares and José L Aznarte. Predicting air quality with deep learning lstm: Towards comprehensive models. *Ecological Informatics*, 55:101019, 2020.

[91] Cui-lin Wu, Rui-feng Song, Zhong-ren Peng, et al. Prediction of air pollutants on roadside of the elevated roads with combination of pollutants periodicity and deep learning method. *Building and Environment*, 207:108436, 2022.

[92] Xiang Li, Ling Peng, Xiaojing Yao, Shaolong Cui, Yuan Hu, Chengzeng You, and Tianhe Chi. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental pollution*, 231:997–1004, 2017.

[93] Axel Gedeon Mengara Mengara, Younghak Kim, Younghwan Yoo, and Jaehun Ahn. Distributed deep features extraction model for air quality forecasting. *Sustainability*, 12(19):8014, 2020.

[94] Ping-Wei Soh, Jia-Wei Chang, and Jen-Wei Huang. Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations. *Ieee Access*, 6:38186–38199, 2018.

[95] Congcong Wen, Shufu Liu, Xiaojing Yao, Ling Peng, Xiang Li, Yuan Hu, and Tianhe Chi. A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. *Science of the total environment*, 654:1091–1099, 2019.

[96] Unjin Pak, Chungsong Kim, Unsok Ryu, Kyongjin Sok, and Sungnam Pak. A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction. *Air Quality, Atmosphere & Health*, 11(8):883–895, 2018.

[97] Suling Zhu, Xiuyuan Lian, Haixia Liu, Jianming Hu, Yuanyuan Wang, and Jinxing Che. Daily air quality index forecasting with hybrid models: A case in china. *Environmental pollution*, 231:1232–1244, 2017.

[98] Sheen Mclean Cabaneros, John Kaiser Calautit, and Ben Hughes. Spatial estimation of outdoor no2 levels in central london using deep neural networks and a wavelet decomposition technique. *Ecological Modelling*, 424:109017, 2020.

[99] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, 454(1971):903–995, 1998.

[100] Yongkang Zeng, Jingjing Chen, Ning Jin, Xiaoping Jin, and Yang Du. Air quality forecasting with hybrid lstm and extended stationary wavelet transform. *Building and Environment*, 213:108822, 2022.

[101] Ingrid Daubechies. *Ten lectures on wavelets*. SIAM, 1992.

[102] Yong Cheng, Ling-Yan He, and Xiao-Feng Huang. Development of a high-performance machine learning model to predict ground ozone pollution in typical cities of china. *Journal of Environmental Management*, 299:113670, 2021.

[103] Xi Gao and Weide Li. A graph-based lstm model for pm2. 5 forecasting. *Atmospheric Pollution Research*, 12(9):101150, 2021.

[104] Hong-Wei Wang, Xiao-Bing Li, Dongsheng Wang, Juanhao Zhao, Zhong-Ren Peng, et al. Regional prediction of ground-level ozone using a hybrid sequence-to-sequence deep learning approach. *Journal of Cleaner Production*, 253:119841, 2020.

[105] Pengcheng Jia, Nianwen Cao, and Shaobo Yang. Real-time hourly ozone prediction system for yangtze river delta area using attention based on a sequence to sequence model. *Atmospheric Environment*, 244:117917, 2021.

[106] Xin-Yu Tu, Bo Zhang, Yu-Peng Jin, Guo-Jian Zou, Jian-Guo Pan, and Mao-Zhen Li. Longer time span air pollution prediction: The attention and autoencoder hybrid learning model. *Mathematical Problems in Engineering*, 2021, 2021.

[107] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[108] Alaa Sheta, Hossam Faris, Ali Rodan, Elvira Kovač-Andrić, and AlaM Al-Zoubi. Cycle reservoir with regular jumps for forecasting ozone concentrations: two real cases from the east of croatia. *Air Quality, Atmosphere & Health*, 11(5):559–569, 2018.

[109] Yuqin Mo, Qi Li, Hamed Karimian, Shuwei Fang, Boyuan Tang, Gong Chen, and Sonali Sachdeva. A novel framework for daily forecasting of ozone mass concentrations based on cycle reservoir with regular jumps neural networks. *Atmospheric environment*, 220:117072, 2020.

[110] Alqamah Sayeed, Yunsoo Choi, Ebrahim Eslami, Yannic Lops, Anirban Roy, and Jia Jung. Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. *Neural Networks*, 121:396–408, 2020.

[111] Qiang Zhang, Shun Wu, Xiangwen Wang, Binzhen Sun, and Haimeng Liu. A pm2. 5 concentration prediction model based on multi-task deep learning for intensive air quality monitoring stations. *Journal of Cleaner Production*, 275:122722, 2020.

[112] Yun Bai, Bo Zeng, Chuan Li, and Jin Zhang. An ensemble long short-term memory neural network for hourly pm2. 5 concentration forecasting. *Chemosphere*, 222:286–294, 2019.

[113] Herbert Jaeger and Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *science*, 304(5667):78–80, 2004.

[114] Hao Zhang, Mingwen Zheng, Yanping Zhang, Xiao Yu, Wenchao Li, and Hui Gao. Application of esn prediction model based on compressed sensing in stock market. *Communications in Nonlinear Science and Numerical Simulation*, 101:105857, 2021.

[115] Ziqiang Li and Gouhei Tanaka. Multi-reservoir echo state networks with sequence resampling for nonlinear time-series prediction. *Neurocomputing*, 467:115–129, 2022.

[116] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

[117] Jun Ma, Yuexiong Ding, Jack CP Cheng, Feifeng Jiang, Vincent JL Gan, and Zherui Xu. A lag-flstm deep learning network based on bayesian optimization for multi-sequential-variant pm2. 5 prediction. *Sustainable Cities and Society*, 60:102237, 2020.

[118] Guang Shi, Yee Leung, Jiang She Zhang, Tung Fung, Fang Du, and Yu Zhou. A novel method for identifying hotspots and forecasting air quality through an adaptive utilization of spatio-temporal information of multiple factors. *Science of The Total Environment*, 759:143513, 2021.

[119] Jun Ma, Zheng Li, Jack CP Cheng, Yuexiong Ding, Changqing Lin, and Zherui Xu. Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network. *Science of The Total Environment*, 705:135771, 2020.

[120] K Krishna Rani Samal, Korra Sathya Babu, and Santos Kumar Das. Temporal convolutional denoising autoencoder network for air pollution prediction with missing values. *Urban Climate*, 38:100872, 2021.

[121] Fuxin Jiang, Chengyuan Zhang, Shaolong Sun, and Jingyun Sun. Forecasting hourly pm2. 5 based on deep temporal convolutional neural network and decomposition method. *Applied Soft Computing*, 113:107988, 2021.

[122] Lei Zhang, Jiaming Na, Jie Zhu, Zhikuan Shi, Changxin Zou, and Lin Yang. Spatiotemporal causal convolutional network for forecasting hourly pm2. 5 concentrations in beijing, china. *Computers & Geosciences*, 155:104869, 2021.

[123] K Krishna Rani Samal, Ankit Kumar Panda, Korra Sathya Babu, and Santos Kumar Das. Multi-output tcn autoencoder for long-term pollution forecasting for multiple sites. *Urban Climate*, 39:100943, 2021.

[124] Yang Zhang, Marc Bocquet, Vivien Mallet, Christian Seigneur, and Alexander Baklanov. Real-time air quality forecasting, part i: History, techniques, and current status. *Atmospheric Environment*, 60:632–655, 2012.

[125] Bo Zhang, Ziyao Geng, Hanwen Zhang, and Jianguo Pan. Densely connected convolutional networks with attention long short-term memory for estimating pm2. 5 values from images. *Journal of Cleaner Production*, 333:130101, 2022.

[126] Bun Theang Ong, Komei Sugiura, and Koji Zettsu. Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting pm2. 5. *Neural Computing and Applications*, 27(6):1553–1566, 2016.

[127] D Saravanan and K Santhosh Kumar. Improving air pollution detection accuracy and quality monitoring based on bidirectional rnn and the internet of things. *Materials Today: Proceedings*, 2021.

[128] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[129] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*, 2014.

[130] Ülkü Alver Şahin, Cuma Bayat, and Osman N Uçan. Application of cellular neural network (cnn) to the prediction of missing air pollutant data. *Atmospheric Research*, 101(1-2):314–326, 2011.

[131] FFRK Chung and F Graham. Spectral graph theory (american mathematical soc.). 1997.

[132] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29, 2016.

[133] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin. Convolutional sequence to sequence learning. In *International conference on machine learning*, pages 1243–1252. PMLR, 2017.

[134] Aiswarya Baby and Aneena Ann Alexander. A review on various techniques used in predicting pollutants. In *IOP Conference Series: Materials Science and Engineering*, volume 396, page 012016. IOP Publishing, 2018.

[135] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

[136] Sepp Hochreiter, Yoshua Bengio, Paolo Frasconi, Jürgen Schmidhuber, et al. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies, 2001.

[137] Ali Rodan and Peter Tiňo. Simple deterministically constructed cycle reservoirs with regular jumps. *Neural computation*, 24(7):1822–1852, 2012.

[138] Xiaowei Wang, Wenjie Liu, Yingnan Wang, and Guotian Yang. A hybrid nox emission prediction model based on ceemdan and am-lstm. *Fuel*, 310:122486, 2022.

[139] Jun Ma, Jack CP Cheng, Changqing Lin, Yi Tan, and Jingcheng Zhang. Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques. *Atmospheric Environment*, 214:116885, 2019.

[140] Iat Hang Fong, Tengyue Li, Simon Fong, Raymond K Wong, and Antonio J Tallon-Ballesteros. Predicting concentration levels of air pollutants by transfer learning and recurrent neural network. *Knowledge-Based Systems*, 192:105622, 2020.

[141] Jusong Kim, Xiaoli Wang, Chollyong Kang, Jinwon Yu, and Penghui Li. Forecasting air pollutant concentration using a novel spatiotemporal deep learning model based on clustering, feature selection and empirical wavelet transform. *Science of The Total Environment*, 801:149654, 2021.

[142] Yuanlin Gu, Baihua Li, and Qinggang Meng. Hybrid interpretable predictive machine learning model for air pollution prediction. *Neurocomputing*, 468:123–136, 2022.

[143] Binjie Chen, Shixue You, Yang Ye, Yongyong Fu, Ziran Ye, Jinsong Deng, Ke Wang, and Yang Hong. An interpretable self-adaptive deep neural network for estimating daily spatially-continuous pm2. 5 concentrations across china. *Science of The Total Environment*, 768:144724, 2021.

[144] Zhezhe Han, Jian Li, Md Moinul Hossain, Qi Qi, Biao Zhang, and Chuanlong Xu. An ensemble deep learning model for exhaust emissions prediction of heavy oil-fired boiler combustion. *Fuel*, 308:121975, 2022.

[145] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[146] Rui Ye and Qun Dai. A novel transfer learning framework for time series forecasting. *Knowledge-Based Systems*, 156:74–99, 2018.

[147] Xiao Feng, Qi Li, Yajie Zhu, Junxiong Hou, Lingyan Jin, and Jingjie Wang. Artificial neural networks forecasting of pm2. 5 pollution using air mass trajectory based geographic model and wavelet transformation. *Atmospheric Environment*, 107:118–128, 2015.

[148] Pablo E Saide, Marcelo Mena-Carrasco, Sebastian Tolvett, Pablo Hernandez, and Gregory R Carmichael. Air quality forecasting for winter-time pm2. 5 episodes occurring in multiple cities in central and southern chile. *Journal of Geophysical Research: Atmospheres*, 121(1):558–575, 2016.

[149] SIV Sousa, Fernando Gomes Martins, Maria CM Alvim-Ferraz, and Maria C Pereira. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software*, 22(1):97–103, 2007.

[150] Ye Wu, Shaojun Zhang, Jiming Hao, Huan Liu, Xiaomeng Wu, Jingnan Hu, Michael P Walsh, Timothy J Wallington, K Max Zhang, and Svetlana Stevanovic. On-road vehicle emissions and their control in china: A review and outlook. *Science of the Total Environment*, 574:332–349, 2017.

[151] Kıymet Kaya and Şule Gündüz Öğüdücü. Deep flexible sequential (dfs) model for air pollution forecasting. *Scientific reports*, 10(1):1–12, 2020.

[152] Dan L Crouse, Mark S Goldberg, and Nancy A Ross. A prediction-based approach to modelling temporal and spatial variability of traffic-related air pollution in montreal, canada. *Atmospheric environment*, 43(32):5075–5084, 2009.

[153] S Abirami and P Chitra. Regional air quality forecasting using spatiotemporal deep learning. *Journal of Cleaner Production*, 283:125341, 2021.

[154] Jingbiao Liao, Tijian Wang, Ziqiang Jiang, Bingliang Zhuang, Min Xie, Changqin Yin, Xuemei Wang, Jialei Zhu, Yu Fu, and Ying Zhang. Wrf/chem modeling of the impacts of urban expansion on regional climate and air pollutants in yangtze river delta, china. *Atmospheric Environment*, 106:204–214, 2015.

[155] Weibiao Qiao, Yining Wang, Jianzhuang Zhang, Wencai Tian, Yu Tian, and Quan Yang. An innovative coupled model in view of wavelet transform for predicting short-term pm10 concentration. *Journal of Environmental Management*, 289:112438, 2021.

[156] Ming Cai, Yafeng Yin, and Min Xie. Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach. *Transportation Research Part D: Transport and Environment*, 14(1):32–41, 2009.

[157] Xuebo Jin, Nianxiang Yang, Xiaoyi Wang, Yuting Bai, Tingli Su, and Jianlei Kong. Integrated predictor based on decomposition mechanism for pm2. 5 long-term prediction. *Applied Sciences*, 9(21):4533, 2019.

[158] Wenjing Mao, Limin Jiao, and Weilin Wang. Long time series ozone prediction in china: A novel dynamic spatiotemporal deep learning approach. *Building and Environment*, 218:109087, 2022.

[159] Han Xiao, Qizhong Wu, Xiaochun Yang, Lanning Wang, and Huaqiong Cheng. Numerical study of the effects of initial conditions and emissions on pm 2.5 concentration simulations with camx v6. 1: a xi'an case study. *Geoscientific Model Development*, 14(1):223–238, 2021.

[160] Wei Wang, Xingqin An, Qingyong Li, Yangli-ao Geng, Haomin Yu, and Xinyuan Zhou. Optimization research on air quality numerical model forecasting effects based on deep learning methods. *Atmospheric Research*, 271:106082, 2022.

[161] Satvik Garg and Himanshu Jindal. Evaluation of time series forecasting models for estimation of pm2. 5 levels in air. In *2021 6th International Conference for Convergence in Technology (I2CT)*, pages 1–8. IEEE, 2021.

[162] Hua Lu, Min Xie, Xiaoran Liu, Bojun Liu, Minzhi Jiang, Yanghua Gao, and Xiaoli Zhao. Adjusting prediction of ozone concentration based on cmaq model and machine learning methods in sichuan-chongqing region, china. *Atmospheric Pollution Research*, 12(6):101066, 2021.

[163] Apeksha Aggarwal and Durga Toshniwal. A hybrid deep learning framework for urban air quality forecasting. *Journal of Cleaner Production*, 329:129660, 2021.

[164] Raquel Espinosa, José Palma, Fernando Jiménez, Joanna Kamińska, Guido Sciavicco, and Estrella Lucena-Sánchez. A time series forecasting based multi-criteria methodology for air quality prediction. *Applied Soft Computing*, 113:107850, 2021.

[165] Qing Tao, Fang Liu, Yong Li, and Denis Sidorov. Air pollution forecasting using a deep learning model based on 1d convnets and bidirectional gru. *IEEE access*, 7:76690–76698, 2019.

[166] Yu Zheng, Xiuwen Yi, Ming Li, Ruiyuan Li, Zhangqing Shan, Eric Chang, and Tianrui Li. Forecasting fine-grained air quality based on big data. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 2267–2276, 2015.

[167] Yinglong Zhang, Bin Zhu, Jinhui Gao, Hanqing Kang, Peng Yang, Lili Wang, Junke Zhang, et al. The source apportionment of primary pm2. 5 in an aerosol pollution event over beijing-tianjin-hebei region using wrf-chem, china. *Aerosol and Air Quality Research*, 17(12):2966–2980, 2017.

[168] Jinal Bhagat and Geetali Saha. Pollutant pm2. 5 multi step prediction under seasonal influences across 13 indian cities. In *2021 Innovations in Power and Advanced Computing Technologies (i-PACT)*, pages 1–8. IEEE, 2021.

[169] Sojin Lee, Chul Han Song, Kyung Man Han, Daven K Henze, Kyunghwa Lee, Jinhyeok Yu, Jung-Hun Woo, Jia Jung, Yunsoo Choi, Pablo E Saide, et al. Impacts of uncertainties in emissions on aerosol data assimilation and short-term pm2. 5 predictions over northeast asia. *Atmospheric Environment*, 271:118921, 2022.

[170] Shovan Kumar Sahu, Shubham Sharma, Hongliang Zhang, Venkatesh Chejarla, Hao Guo, Jianlin Hu, Qi Ying, Jia Xing, and Sri Harsha Kota. Estimating ground level pm2. 5 concentrations and associated health risk in india using satellite based aod and wrf predicted meteorological parameters. *Chemosphere*, 255:126969, 2020.

[171] Rajmal Jat, Bhola Ram Gurjar, and Douglas Lowe. Regional pollution loading in winter months over india using high resolution wrf-chem simulation. *Atmospheric Research*, 249:105326, 2021.

[172] Jiangeng Li, Xingyang Shao, and Rihui Sun. A dbn-based deep neural network model with multitask learning for online air quality prediction. *Journal of Control Science and Engineering*, 2019, 2019.

[173] Suling Zhu, Xuanlin Qiu, Yanru Yin, Min Fang, Xingrong Liu, Xuejing Zhao, and Yanjun Shi. Two-step-hybrid model based on data preprocessing and intelligent optimization algorithms (cs and gwo) for no2 and so2 forecasting. *Atmospheric Pollution Research*, 10(4):1326–1335, 2019.

[174] Junxiang Fan, Qi Li, Junxiong Hou, Xiao Feng, Hamed Karimian, and Shaofu Lin. A spatiotemporal prediction framework for air pollution based on deep rnn. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4:15, 2017.

[175] David A Wood. Trend decomposition aids forecasts of air particulate matter (pm2. 5) assisted by machine and deep learning without recourse to exogenous data. *Atmospheric Pollution Research*, 13(3):101352, 2022.

[176] Steven C Smyth, Weimin Jiang, Helmut Roth, Michael D Moran, Paul A Makar, Fuquan Yang, Véronique S Bouchet, and Hugo Landry. A comparative performance evaluation of the aurams and cmaq air-quality modelling systems. *Atmospheric Environment*, 43(5):1059–1070, 2009.

[177] James T Kelly, Shannon N Koplitz, Kirk R Baker, Amara L Holder, Havala OT Pye, Benjamin N Murphy, Jesse O Bash, Barron H Henderson, Norman C Possiel, Heather Simon, et al. Assessing pm2. 5 model performance for the conterminous us with comparison to model performance statistics from 2007-2015. *Atmospheric Environment*, 214:116872, 2019.

[178] Fei Xiao, Mei Yang, Hong Fan, Guanghui Fan, and Mohammed AA Al-Qaness. An improved deep learning model for predicting daily pm2. 5 concentration. *Scientific reports*, 10(1):1–11, 2020.

[179] Qingyang Xiao, Guannan Geng, Jing Cheng, Fengchao Liang, Rui Li, Xia Meng, Tao Xue, Xiaomeng Huang, Haidong Kan, Qiang Zhang, et al. Evaluation of gap-filling approaches in satellite-based daily pm2. 5 prediction models. *Atmospheric Environment*, 244:117921, 2021.

[180] Chinmay Jena, Sachin D Ghude, Rajesh Kumar, Sreyashi Debnath, Gaurav Govardhan, Vijay K Soni, Santosh H Kulkarni, G Beig, Ravi S Nanjundiah, and M Rajeevan. Performance of high resolution (400 m) pm2. 5 forecast over delhi. *Scientific reports*, 11(1):1–9, 2021.

[181] Jia Hong, Feiyue Mao, Wei Gong, Yuan Gan, Lin Zang, Jihong Quan, and Jiangping Chen. Assimilating fengyun-4a observations to improve wrf-chem pm2. 5 predictions in china. *Atmospheric Research*, 265:105878, 2022.

[182] Tong Liu, Alexis KH Lau, Kai Sandbrink, and Jimmy CH Fung. Time series forecasting of air quality based on regional numerical modeling in hong kong. *Journal of Geophysical Research: Atmospheres*, 123(8):4175–4196, 2018.

[183] Pu-Yun Kow, Li-Chiu Chang, Chuan-Yao Lin, Charles C-K Chou, and Fi-John Chang. Deep neural networks for spatiotemporal pm2. 5 forecasts based on atmospheric chemical transport model output and monitoring data. *Environmental Pollution*, 306:119348, 2022.

[184] Pei-Yi Wong, Hsiao-Yun Lee, Yu-Cheng Chen, Yu-Ting Zeng, Yinq-Rong Chern, Nai-Tzu Chen, Shih-Chun Candice Lung, Huey-Jen Su, and Chih-Da Wu. Using a land use regression model with machine learning to estimate ground level pm2. 5. *Environmental Pollution*, 277:116846, 2021.

[185] Runkui Li, Tianxiao Ma, Qun Xu, and Xianfeng Song. Using maiac aod to verify the pm2. 5 spatial patterns of a land use regression model. *Environmental Pollution*, 243:501–509, 2018.

[186] Jia Hong, Feiyue Mao, Qilong Min, Zengxin Pan, Wei Wang, Tianhao Zhang, and Wei Gong. Improved pm2. 5 predictions of wrf-chem via the integration of himawari-8 satellite data and ground observations. *Environmental Pollution*, 263:114451, 2020.

[187] Wei Ma, Zibing Yuan, Alexis KH Lau, Long Wang, Chenghao Liao, and Yongbo Zhang. Optimized neural network for daily-scale ozone prediction based on transfer learning. *Science of The Total Environment*, 827:154279, 2022.

[188] Dong Wang and Wei-Zhen Lu. Forecasting of ozone level in time series using mlp model with a novel hybrid training algorithm. *Atmospheric Environment*, 40(5):913–924, 2006.

[189] Yinan Chen, Shichong Yang, and Qiyuan Wang. Prediction of pm2. 5 concentration in guangzhou based on lstm neural network. In *2021 2nd International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI)*, pages 8–12. IEEE, 2021.

[190] Meizi Li, Yuqi Zhang, Yunjie Lu, Mao-Zhen Li, Yunwen Chen, Jianguo Pan, and Bo Zhang. An end-to-end air pollutant concentrations prediction algorithm based on attention mechanism for targeted season: A case study in north china. *Atmospheric Pollution Research*, 13(5):101396, 2022.

[191] Anurag Barthwal, Debopam Acharya, and Divya Lohani. Prediction and analysis of particulate matter (pm2. 5 and pm10) concentrations using machine learning techniques. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–16, 2021.

[192] Sri Harsha Kota, Hao Guo, Lauri Myllyvirta, Jianlin Hu, Shovan Kumar Sahu, Rajyalakshmi Garaga, Qi Ying, Aifang Gao, Sunil Dahiya, Yuan Wang, et al. Year-long simulation of gaseous and particulate air pollutants in india. *Atmospheric Environment*, 180:244–255, 2018.

[193] Paulo SG de Mattos Neto, Francisco Madeiro, Tiago AE Ferreira, and George DC Cavalcanti. Hybrid intelligent system for air quality forecasting using phase adjustment. *Engineering Applications of Artificial Intelligence*, 32:185–191, 2014.

[194] Chao Ye, Ruishan Chen, and Mingxing Chen. The impacts of chinese nian culture on air pollution. *Journal of Cleaner Production*, 112:1740–1745, 2016.

[195] Saurabh Singh, Pradip Kumar Sharma, Byungun Yoon, Mohammad Shojafar, Gi Hwan Cho, and In-Ho Ra. Convergence of blockchain and artificial intelligence in iot network for the sustainable smart city. *Sustainable Cities and Society*, 63:102364, 2020.

[196] M Mazhar Rathore, Anand Paul, Won-Hwa Hong, HyunCheol Seo, Imtiaz Awan, and Sharjil Saeed. Exploiting iot and big data analytics: Defining smart digital city using real-time urban data. *Sustainable cities and society*, 40:600–610, 2018.

[197] Adriana Simona Mihăiţă, Laurent Dupont, Olivier Chery, Mauricio Camargo, and Chen Cai. Evaluating air quality by combining stationary, smart mobile pollution monitoring and data-driven modelling. *Journal of cleaner production*, 221:398–418, 2019.

[198] Cristian Ramon-Cortes, Pol Alvarez, Francesc Lordan, Javier Alvarez, Jorge Ejarque, and Rosa M Badia. A survey on the distributed computing stack. *Computer Science Review*, 42:100422, 2021.

[199] Marjan Asgari, Mahdi Farnaghi, and Zeinab 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*, pages 89–93, 2017.