Current Pollution Reports (2020) 6:399–409
https://doi.org/10.1007/s40726-020-00159-z

AIR POLLUTION (H ZHANG AND Y SUN, SECTION EDITORS)

Deep Learning for Air Quality Forecasts: a Review

Qi Liao [1,2] · Mingming Zhu [1,2] · Lin Wu [1] · Xiaole Pan [1] · Xiao Tang [1] · Zifa Wang [1,2,3]

Published online: 3 September 2020

Abstract

Air pollution is one of major environmental issues in the twenty-first century due to global industrialization and urbanization. Its mitigation necessitates accurate air quality forecasts. However, current state-of-the-art air quality forecasts are limited from highly uncertain chemistry-transport models (CTMs), shallow statistical methods, and heterogeneous and incomplete observing networks. Recently, deep learning has emerged as a general-purpose technology to extract complex knowledge using massive amount of data and very large networks of neurons and thus has the potential to break the limits of air quality forecasts. Here, we provide a brief review of recent attempts on using deep learning techniques in air quality forecasts. We first introduce architectures of deep networks (e.g., convolutional neural networks, recurrent neural networks, long short-term memory neural networks, and spatiotemporal deep network) and their relevance to explore the nonlinear spatiotemporal features across multiple scales of air pollution. We then examine the potential of deep learning techniques for air quality forecasts in diverse aspects, namely, data gap filling, prediction algorithms, improvements of CTMs, estimations with satellite data, and source estimations for atmospheric dispersion forecasts. Finally, we point out some prospects and challenges for future attempts on improving air quality forecasts using deep learning techniques.

Keywords Air quality forecasts · Deep learning · Neural network

Introduction

Air pollution is a widespread environmental issue in the twenty-first century. Particularly in densely populated megacities, air pollution has been regarded as one of the largest environmental threads [1]. The primary contributors to air pollution are human-induced activities such as industrial productions, moving vehicle exhausts, and coal combustion. These human activities produce gaseous pollutants and particulate matter into the atmosphere, which cause acute and chronic effects on human health, especially for young and elderly [2]. The World Health Organization (WHO) estimated that air pollution has caused around seven million premature deaths across the world in 2012, of which atmospheric aerosol plays a significant contribution [3]. Among those aerosol particles, fine particulate matter with a diameter of less than 2.5 μm ($PM_{2.5}$), which is the major driver of the frequent "haze" weather in cities and suburban areas, is consistently listed as one of the main factors for the global burden of disease [4, 5]. For instance, the resulting premature deaths by $PM_{2.5}$ exposure are estimated to be 1.2 and 1.6 million in 2010 and 2014, respectively [6, 7]. For more comprehensive evaluations of the health impacts of air pollutions, government agencies monitor the Air Quality Index (AQI) which consists of measurements from an ensemble of major air contaminants including ground level ozone (O_3), particulates (e.g., $PM_{2.5}$ and PM_{10}), sulfur dioxide (SO_2), carbon monoxide (CO), and nitrogen dioxide (NO_2). When the AQI increases, an increasingly large percentage of the population is expected to experience severe health effects [8]. To control air pollution and protect humans from adverse health effects, it is of vital importance to have accurate and real-time air quality information [9].

This article is part of the Topical Collection on *Air Pollution*

Lin Wu
wlin@mail.iap.ac.cn

Xiaole Pan
panxiaole@mail.iap.ac.cn

1 State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

2 College of Earth Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

3 Center for Excellence in Urban Atmospheric Environment, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China

Accurate air quality forecasts are nevertheless challenging due to the fact that the needed data are usually of a wide range of sources and not necessarily dense and that the dynamics of the air pollutants is high dimensional and nonlinear. Air quality data can come from ground level stations, lidars, aircrafts, and satellites. They represent different spatiotemporal sampling of the evolving pollutants which are highly heterogeneous in space and time. It is difficult to aggregate diverse data to form a paramount and consistent picture of changes of the pollutants. Ground-based measurements provide more direct information as they are close to emissions, but their distribution can be uneven, with monitoring networks relatively dense in urban areas and sparse in rural areas. The fate of pollutants in the atmosphere involves complex chemical and physical processes. Chemistry-transport models (CTMs) numerically solve the mathematical equations of those chemical and physical processes and provide much more full and even coverage information than those from diverse monitoring networks. However, the predictability of CTMs suffers from uncertainties that generated by the deficiencies in characterizing the complex chemical mechanisms, the less known role of the driving meteorological fields, and the inaccurate emissions data. Most CTMs compute pollutant concentrations at regular grids, but cannot represent variations of pollutants finer than one or two grid cells. Moreover, connections from distant regions and across scales (e.g., synoptic and regional impacts on medium- and long-range air quality forecasts) are not directly accessible from CTM simulations.

Deep learning has recently emerged as a powerful technique to extract high-level complex abstraction and knowledge from very large and high-dimensional datasets [10]. Such knowledge and relationships hidden in big datasets are multivariate, nonlinear, multiscale, and usually beyond the scope of traditional modeling based on first mathematical or physical principles. Therefore, deep learning, as a general-purpose technique, has great potential to change the status quo in many domains such as Earth system science [11] and remote sensing [12]. Here, we briefly review the attempts on applying deep learning techniques to break the conventional limits in air quality forecasts. Our aim is not meant to be exhaustive, but rather review recent advances in a selective set of topics such as data gap filling, prediction algorithms, speedups of CTMs, satellite use, and source estimations. We first begin with a short introduction to traditional air quality forecast approaches and deep learning concepts and techniques. We then examine recent attempts on applying deep learning techniques for improving air quality forecasts, and finally, we discuss some future directions that we found challenging.

Traditional Approaches for Air Quality Forecasts

Current state-of-the-art air quality forecasts are based on CTM simulations [13] and statistical methods [14]. CTMs numerically solve the mathematical equations derived from first principles such as global conservation laws and local kinematic laws that describe the generation, dispersion, transformation, and removal processes of pollutants. They can be either driven offline by or coupled online with the meteorological fields and emissions [15]. Commonly used CTMs are, for instance, the Community Multiscale Air Quality (CMAQ) model [16, 17], the Weather Research and Forecasting model coupled to Chemistry (WRF-Chem) [18], and the Nested Air Quality Prediction Modeling System (NAQPMS) [19]. However, the numerical simulation of CTMs can be time-consuming due to the computationally expensive chemical mechanisms [20]. Many of the chemical reactions occur at a very short timescale, which require intensive computations [21]. Most CTM simulations work with the offline mode and are known to be subject to high uncertainties of chemical, meteorological, or emission-related origin [22].

The statistical methods pay less attention to the physical and chemical mechanism of pollutants and directly correlate the input data such as the meteorological variables and pollutant data of previous time steps with the future pollutant concentrations. Various classical statistical methods have been applied to air quality forecasts including the Auto Regression Moving Average (ARMA) model [23], the Auto Regression Integrated Moving Average (ARIMA) model, the Multiple Linear Regression (MLR) [24], and the Geographically Weighted Regression (GWR) [25]. However, these regression models are by principle rather for analysis purpose and have limited ability to capture the nonlinear dynamics in air quality data.

To improve the CTM forecasting performance, data assimilation methods can be put forward to better combine the knowledge in CTM models and information in various pollutant observations for more accurate forecasts [26–29]. Nevertheless, in geoscience, the dimensions of data are bound to be as large as 10^8, which cause the so-called curse of dimensionality for the assimilation of observations and result in computational error [30]. Ensemble forecast technique helps in reducing error of prediction by weighting an ensemble of simulations [31–33], as leads to probabilistic forecasting and addresses uncertainty quantification [34, 35].

The main limitation of above-mentioned traditional forecasting approaches is that they have difficulties in probing complex high-dimensional relationships from massive datasets. There is a lack of flexible multiscale frameworks, which prevents the traditional forecasting approaches to solve the curse of dimensionality. While their complete solutions of the forecast problem are computationally prohibitive, approximate solutions lead to decreases in forecasting performances.

Deep Learning Concept

Recently, deep learning has emerged as a general-purpose technology to extract complex knowledge using massive amount of data and very large networks of neurons and thus has the potential to break the conventional limits of air quality forecasts using traditional methods.

Deep learning is a subfield of machine learning, which has rapidly developed after 2010. The breakthrough lies in the successful training of very large neural networks of multiple hidden layers to extract the intrinsic features and patterns in massive datasets. The deep networks succeed in learning representations of knowledge in very high-dimensional spaces in forms of connecting weights in between neurons of the networks. The organization of the networks, or the network architecture, is thus a mapping of the knowledge space of various domains. Successful applications of deep learning prevail, first in computer science such as computer vision, natural language processing and speech recognition, and then in fundamental sciences such as physics and chemistry. Such success can also be expected for air quality forecasts, as the pollutants evolution can be analogous to image sequences prediction [36, 37].

Over the past two decades, the amount of data created in air quality has been growing exponentially. For instance, satellite remote sensors constantly produce new global images, whereas improved computing power enables us to make routine simulations with numerical models. However, our ability to collect and create data far outpaces our ability to understand them [11]. To use these big data for air quality forecasts, we can model air quality systems based on deep learning. Although these systems are considerably dynamic, spatially expansive, and behaviorally heterogeneous [8], deep learning can automatically learn high-dimensional data, perform dimensionality reduction processing on the data, and extract low-dimensional features [38].

Deep Network Architectures for Air Quality Forecasts

The architectures of deep networks determine their ability to extract the complex nonlinear features across scales from the high-dimensional datasets. Deep networks differ from shallow artificial neural networks [39–41] in that they have multiple layers of neurons. The connections of these neurons form distinct structures of the deep networks which are suitable for different areas of applications. Deep networks with better ability to extract spatiotemporal features are more appropriate for air quality forecasts.

We briefly present some architectures of deep networks according to their spatiotemporal characteristics. The architectures adapting to temporal series predictions include the recurrent neural network (RNN), the long short-term memory (LSTM) network, and the gated recurrent unit (GRU) network. The architectures adapting to spatial feature extractions include the convolutional neural network (CNN), the stacked autoencoder (SAE), and the deep belief network (DBN). The spatiotemporal deep learning (STDL) architectures can be constructed by connections or coupling of temporal and spatial modules. The STDL architectures are supposed to better adapt to air quality forecasts thanks to their ability to represent complex spatiotemporal features across scales of the pollutant evolutions. Diverse architectures of deep networks and the framework of deep learning for air quality forecasts are shown in Fig. 1.

Architectures for Temporal Predictions

Recurrent Neural Networks The RNNs are variants of feed-forward neural networks (FNN). FNNs enable signals to travel only one way from input to output. They are straightforward networks without loops that associate inputs with outputs [42]. RNNs introduce self-connection of neurons cyclic structure into the network, which are based on FNN. Thus, input data can be memorized, and sequences of data can influence network outputs through self-connected neurons. Taking advantage of their memory characteristics, RNNs outperform FNNs in many applications. However, RNNs may fail to capture long time dependencies in input data, and it may face the problems of vanishing and exploding gradients when the time of training is too long [43, 44].

Long Short-Term Memory LSTM networks (LSTM) [43] are enhanced RNNs. They introduce memory blocks to overcome the vanishing gradient problem. The memory blocks consist of three types nonlinear multiplicative gates: 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 [10]. LSTM networks have the advantage to train long time sequences and perform better than traditional RNN in many applications [15, 45].

Gated Recurrent Unit The GRU networks [46] are simplified versions of the LSTM networks. They only consist of update and reset gates but can still balance the data flows inside the unit [10]. The update gate replaces the input and forgets gates in LSTM, which determines whether information needs to be remembered. The advantage of using GRU compared with LSTM is that GRU have fewer parameters and thus less computational loads for training. Nevertheless the GRU networks have shown similar performances on music and speech signals as LSTM or ever better performance on smaller datasets [47].

Architectures for Spatial Feature Extractions

Convolutional Neural Networks The CNNs [48] are deep feed-forward networks which consist of a series of convolutional layers. They are capable of analyzing multiscale shift invariant

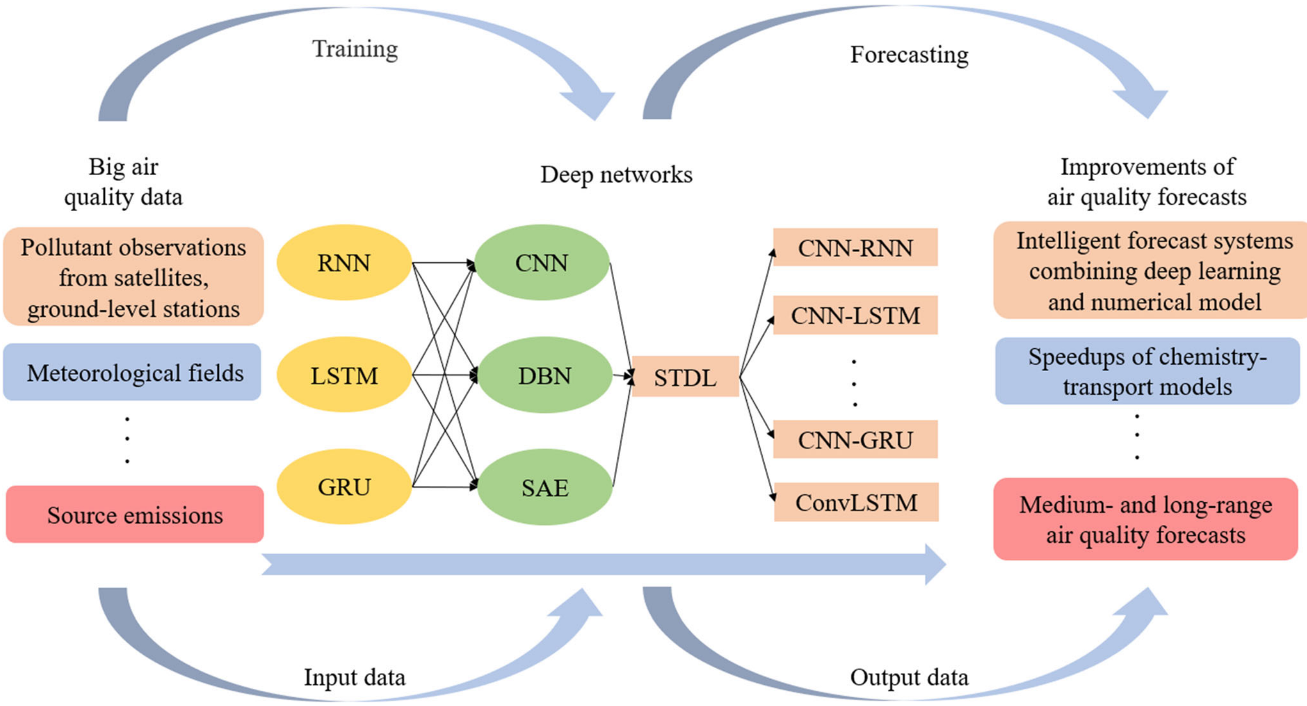

Fig. 1 Deep network architectures and deep learning framework for air quality forecasts

features of data. Subsampling operations are performed between two successive convolutional layers. Two commonly used subsampling operations are max pooling and mean pooling. Pooling layers can be replaced by convolutional layers, as simplifies the network structure [49]. Units in a convolutional layer are organized in feature maps, and each unit is connected to local weights in the feature maps of the previous layer through a filter bank [10]. The sum of local weights is passed through an activation function which can take various forms such as Rectified Linear Units (ReLU) [50] and Scaled Exponential Linear Units (SELU) [51]. The CNNs have produced outstanding results in processing multiple-dimensional array data with spatial structure [52, 53]. They are widely used in speech recognition and image recognition, as motivates researchers to estimate environmental exposures through digital images using CNNs [54]. CNN can effectively extract the spatial features of pollutants.

Stacked Autoencoder An autoencoder is a neural network that attempts to reconstruct its inputs [55]. This is done by minimizing the discrepancy between inputs and network outputs. It has the ability to extract the features in reduced spaces through the reconstructions of the inputs. Stacked autoencoder (SAE) is a deep model formed by stacking successive layers of autoencoders [56, 57]. For a SAE with M hidden layer, the autoencoders perform unsupervised pre-training from hidden layer k ($k < M$) to hidden layer ($k + 1$). Each hidden layer is a higher-level abstraction of the previous layer, and the final hidden layer contains high-level feature which is more effective for prediction [58].

Deep Belief Networks The DBNs [59, 60] are formed by stacking multiple energy-based Restricted Boltzmann Machine (RBM) [61].They have achieved excellent results in feature recognition and classification as well as prediction problems [62–64]. An RBM consists of a visible layer and a hidden layer, where the hidden layer of the prior RBM is the visible layer of the next RBM [65]. The RBMs performs unsupervised pre-training layer by layer from bottom to top to initialize the network parameters of each layer. After the pre-training process, a softmax classifier is set in the last layer of DBN to classify the features. Finally the entire network is tuned supervised tuned through the labeled network using back propagation algorithm. Furthermore, a tenfold cross-validation technique is commonly applied to evaluate the model performance and test the model overfitting [66].

Spatiotemporal Deep Learning

Spatiotemporal deep learning (STDL) architectures can be designed to inherently deal with the spatiotemporal features in high-dimensional data. Those STDL networks mainly include a spatial module, a deduction module, and a temporal prediction module. The STDL networks are hypernetworks whose modules are themselves networks [67]. For instance, the spatial features can be extracted by the spatial modules designed as CNN, SAE, or DBN. Those output from spatial modules can then be passed into the temporal modules (e.g., RNN, LSTM, GRU) through the deduction module to learn the temporal features. The deduction module bridges the spatial and temporal modules, which considers the intrinsic causality between spatial and temporal characteristics [68]. The deduction module can be a simple matrix multiplication (e.g., CNN-LSTM) or other kinds of operations such convolution. The STDL networks are supposed to achieve better forecasting performances than temporal architectures (e.g., LSTM) as they can better extract the spatiotemporal features in data.

Deep Learning for Air Quality Forecasts

In this section, we briefly review how deep learning can be applied to improve air quality forecasts. Table 1 summarizes how deep learning can improve air quality forecasts in diverse ways.

Data Interpolation and Gap Filling to Generate Datasets for Air Quality Forecasts

Air quality forecasts based on deep learning requires a large number of datasets to train the model. However, the ground observation data is usually uneven distributed and can be missing due to instrumental flaws. Hence, it would be desirable to fill missing data or interpolate data to generate bigger datasets so that the deep learning models can be better trained.

Simple interpolations can be performed by considering the similarity of different building and geographic locations on air pollution stations when interpolating station data into grid data. Such similarity can be calculated by the surrounding important geographic features according to their impacts on air quality for each station. For spatiotemporal data, the nearby observations tend to be more alike than those far apart [85, 86]. Dynamic time warping (DTW) [87, 88] is a commonly used method to calculate similarity between two time series. For instance, Soh et al. [89] used the DTW distance to compute k-nearest neighbor for exploring the similarities for time sequences at different locations.

Spatial images, which include different built environments around the stations, can be used to train the spatial deep learning models (e.g., CNN, SAE, and DBN) in order to obtain the spatial features of air pollution. For instance, Yi et al. [90] and Lin et al. [91] take the similarity of the built environment between the locations of air quality stations into consideration when constructing the STDL models. Deep learning techniques can be used to recovery missing data and interpolate station data into grids. Li et al. [69] proposed a deep learning method named M-BP algorithm to recover the missing data. Qi et al. [70] divided the Beijing city into 1 km × 1 km grids and labeled the instance if there is an air quality monitor station located in the grid. They further constructed a semi-supervised deep learning method to complete the site-to-grid interpolation and predict air pollutant concentrations.

Inversion of Ground-Level Data from Satellite Data to Generate Datasets for Air Quality Forecasts

With the improvement of imaging spectrometers and sensor spectral detection capabilities, the satellite data is increased rapidly in the past 20 years. More datasets can be generated by combining satellite data and ground observations to estimate ground-level pollution concentrations. The satellites data available for aerosol inversion include those from Moderate Resolution Imaging Spectroradiometer (MODIS) [92, 93], Multi-Angle Implementation of Atmospheric Correction (MAIAC) [94], Multi-angle Imaging Spectroradiometer (MISR) [95], and Landsat 8 Operational Land Imager (OLI). Some researchers have directly used the satellite-derived Aerosol Optical Depth (AOD) to estimate ground-level air

Table 1 Deep learning approaches for air quality forecast tasks

Scientific task	Conventional approaches	Deep learning approaches	Potential advantages
Data interpolation and filling to generate dataset for pollution forecasts	Inverse distance weighting (IDW), linear or cubic spline interpolation	Back propagation deep learning neural network [69], spatiotemporal semi-supervised deep learning neural network [70]	Suitable in regions without monitoring station
Inversion estimates of ground-level pollutants from satellite observations	Linear or nonlinear regression on aerosol optical depth (AOD)	CNN [71], DBN [72]	Avoid inherent retrieval errors and suitable in regions without stations
Spatiotemporal air quality forecasts	Statistics methods (e.g., ARIMA, MLR), CTM simulations (e.g., CMAQ, NAQPMS)	Temporal deep networks (e.g., RNN, LSTM, GRU [73–76]), STDL models (e.g., CNN-LSTM, CNN-GRU [77–79])	Extract nonlinear features from high-dimensional data
Improve CTM forecasts	Improve parameterization scheme	Deep learning model to replace and reproduce the chemical mechanism [80, 81], encoder-operator-decoder neural networks [82]	Improve, speedup, and build a more stable numerical model
Source estimation for air pollution dispersion forecasts	Gaussian dispersion model, computational fluid dynamics (CFD)	Back propagation network combined with Gaussian model [83], mask-RCNN (region convolution neural network) [84]	Satisfy the requirement of emergencies with high efficiency and accuracy

pollutant concentrations (e.g., $PM_{2.5}$ and PM_{10}) [96, 97]. The inversion methods can be linear or nonlinear regression [98].

The AOD products have inherent retrieval errors relative to the satellite instrument, retrieval algorithms, and even the versions of software. Such retrieval errors in the AOD products lead to accumulation of errors in pollutant concentration estimates. Better estimates can be obtained by calibrating the retrieval errors using machine learning methods for regions without ground monitoring stations [99]. Li et al. [100] used a feedback Redial Basic Function (RBF) neural network to retrieve high-precision Aerosol Extinction Coefficient (AEC) profiles in order to obtain better optical properties of aerosols. Zhang et al. [101] used the multilayer perceptron neural networks to estimate $PM_{2.5}$ and PM_{10} concentrations directly from the band reflectance data of Landsat 8 OLI remote sensing images.

Deep learning has been proven to be an effective approach to retrieve features in satellite images for the inversion estimates of ground-level pollutant concentrations. Li et al. [72] used deep learning to establish the relationship between $PM_{2.5}$ and satellite Top-Of-Atmosphere (TOA) reflectance for the estimation of ground-level $PM_{2.5}$ concentrations. Furthermore, they incorporated geographical correlations into the DBN model for better estimation of $PM_{2.5}$ concentrations. Hong et al. [71] trained CNN with a large database of satellite images and ground level $PM_{2.5}$ measurements to estimate global $PM_{2.5}$ concentrations especially for regions without ground monitoring data or detailed emissions inventories.

Speedups of Chemistry-Transport Models

Numerical models, especially CTMs, which simulate the generation, transformation and removal processes of pollutants, are computationally expensive due to the complex physical and chemical mechanisms. Deep learning techniques have the potential to accelerate atmospheric chemistry computations of CTMs by emulating the chemical mechanism and to improve the accuracy of numerical simulation results by introducing deep learning modules into the numerical model.

Keller et al. [80] used random forest regressions to train a machine learning replacement for the gas-phase chemistry in the GEOS-Chem chemistry model on the timescales of days to weeks. They separated long-lived from short-lived species and made significant improvement in the model performance by imposing conservation of atoms for the NOx family. Kelp et al. [81] attempted to investigate the potential for machine learning to reproduce the behavior of chemical mechanism with reduced computational expense. They created a 17-layer residual multi-target regression neural network to match the hourly concentration predictions of 77 species of a gas-phase chemical mechanism. By this way, they achieved orders-of-magnitude speedups (250×). However, the neural network could occasionally make a very different prediction related to the target model. Such errors can propagate and grow rapidly and eventually yield meaningless prediction. In a follow-up study, Kelp et al. [82] used a recurrent training regime to train the multiple timescales chemical reactions with an encoder-operator-decoder framework to reduce the dimensionality of the chemical system. They successfully reduced error accumulation and created a more stable machine-learned model.

Spatiotemporal Air Quality Forecasts

Hitherto deep learning techniques for air quality forecasts are mostly tested with STDL networks, as they are known to have better ability to represent complex spatiotemporal features across scales of the pollutant evolutions. Bui et al. [73] and Kim et al. [76] used RNN and LSTM models to predict air pollution from historical time series pollutant data and meteorological data. Freeman et al. [74] used LSTM to predict the 8-h moving average concentrations of ozone, and they obtained forecasts with low errors up to 72 h. Athira et al. [102] found that the performance of GRU model is slightly better than RNN and LSTM models for the prediction of PM_{10} concentration. Xayasouk et al. [103] and Li et al. [57] used SAE model to train pollutant data and further constructed a regression model for air quality prediction. Li et al. [75] tested a LSTM model that can automatically extract inherent useful features from historical air pollutant data and auxiliary meteorological data. They showed that the use of auxiliary data is helpful to improve forecasting performance.

Many researchers attempt to use STDL models for spatiotemporal prediction of air quality. Soh et al. [78] tested CNN-LSTM model for air quality forecast up to 48 h, and extracted the correlations between adjacent locations and among similar locations in the temporal domain. Huang et al. [77] compared CNN-LSTM model with traditional machine learning models on the ability for $PM_{2.5}$ concentration forecast. They found that the forecasting performance of the CNN-LSTM model was the best with root mean square error (RMSE) and mean absolute error (MAE) both being the lowest. Wang et al. [79] succeeded in applying a CNN-GRU model to forecast three air pollutants ($PM_{2.5}$, PM_{10}, and O_3) of a city with limited monitoring stations over 48 h.

Source Estimation for Dispersion Forecasts

Dispersion forecast is important to avoid emergencies and protect people from air pollution. Current dispersion models mainly include Gaussian dispersion model [104, 105], Lagrangian stochastic model [106, 107], and Computational Fluid Dynamics (CFD) model [108, 109]. Accurate dispersion forecasts necessitate accurate location of the sources of pollution. To this end, a parameter estimation algorithm is needed.

In cases of emergencies, the source estimation and dispersion forecasts have to be fast and accurate.

Essentially fast forward models need to be established to estimate the source of pollution and perform dispersion forecasts. Neural or deep networks are candidates for such forward models. For instance, So et al. [110] used neural network to estimate the hazardous gas release rate with the optical sensor data. Wang et al. [111] proposed an approach based on the integration of gas detectors, neural network, and gas dispersion to predict the release gas concentration at certain off-site locations. However, short of atmospheric dispersion mechanism, source estimation was unable to be performed in these studies.

Ma et al. [83] performed source estimation by combining classic Gaussian model with machine learning algorithms for gas dispersion prediction. The emission source parameters were then identified with particle swarm optimization (PSO) method. Qiu et al. [112] followed this approach and also tested the expectation maximization (EM) for source parameter estimation. When datasets of remote sensing images are available, it is possible to leverage deep learning techniques for source estimation. For instance, Kumar et al. [84] proposed hyperspectral Mask-Region Convolution Neural Network (Mask-RCNN) to analyze large area hyperspectral imageries for representing and detecting the releases of methane plumes.

Prospects

Despite the recent advances in applying deep learning for air quality forecasts, a considerable amount of efforts is still needed to make full use of the strength of deep learning to break the conventional limits of forecasting. To date, there exist no studies that deal with comprehensive datasets of sufficient long time intervals including pollutant observations from all sources, CTM simulations, data assimilation products, forcing meteorological fields, and sources of emissions. Hence, to move forwards, it would be necessary to first build such comprehensive benchmark datasets for the test of learning algorithms and for the design of deep network structures. It is still not clear what network architecture is best suited for extracting the spatiotemporal features of the evolving pollutants hidden in the large and high-dimensional datasets, though advances in meteorological studies may provide some clues, suggesting using STDL networks like ConvLSTM and its variants [113, 114] that have the advantage of fusion across scales among diverse sources of information in the comprehensive datasets.

Breakthroughs brought by deep learning should be targeted to abrupt performance improvements on forecast skills as well as to enlarged range of forecasting such as the medium- and long-range forecasts over 7–15 days and the pollution predictions at local scale with complex topography at urban or site areas. These can hopefully be fulfilled by successful fusion of diverse heterogeneous information and by intelligent modeling of the chemical and physical processes across scales. To this end, frameworks of STDL models need to be established to assimilate various patterns in CTM simulations, driving factors of meteorological fields and emissions, and diverse pollutant observations. Under such frameworks, it is possible to build hybrid models that combine deep networks and the CTM dynamics. While CTM simulations serve as backbone of forecasting, a deep learning layer can play various roles of bias detection, uncertainty reduction, sub-grid downscaling or parameterization, source estimation, and surrogate model learning that replaces complex and occasionally unknown chemical processes to speed up the CTM simulations. In addition, deep learning techniques can be tested for optimal nonlinear combinations in ensemble forecasts, incorporation of various proxy data (e.g., traffic flows, POI information, and land use/cover images) for better forecasts, and construction of reanalysis products from extensive datasets including satellite data [115] and visibility data [116].

Conclusion

We have briefly reviewed the attempts on applying deep learning techniques to break the conventional limits inherent in traditional approaches using chemistry-transport models (CTMs) or shallow statistical methods for air quality forecasts. Deep learning is appealing with its ability to extract complex knowledge hidden in massive datasets. It is essential to design deep network architectures that best represent the nonlinear spatiotemporal features of the pollutants evolution across multiple scales. Deep learning has great potential to improve air quality forecasts in diverse aspects including data gap filling, estimations with satellite data, speedups of CTMs, prediction algorithms, and source estimations.

Deep learning is a general-purpose technique which impacts many areas of researches. Its applications in air quality forecasts are still in the early days and face a number of challenges. First, there is a data gap for comprehensive studies that lead to breakthroughs in disruptive forecasting performance improvements. To date, no common datasets exist to test deep learning algorithms so that benchmarking is possible to accumulate knowledge for better forecasts, as what experienced by ImageNet for image classification. Second, provided comprehensive datasets, diverse deep networks and learning algorithms need to be compared and innovated for best extracting complex nonlinear features across scales hidden in those comprehensive datasets. By far most, deep learning studies for air quality forecasts are limited in dealing with meteorological variables and pollutant observations from ground-level monitoring networks. The vast patterns in CTM simulations have yet to be assimilated by deep networks to justify their generalization performances. Third, open problems of both theoretical and practical values need to be proposed to stimulate deep learning studies for air quality forecasts. These problems should be targeted to break the conventional limits by traditional forecasting approaches. It is at those challenging problems that the strength of deep learning can be best illustrated, for instance, the skillful long-range forecast over 7–15 days which is vital for successful prevention and control of heavy air pollution events. Last but not least, deep networks are often criticized to be black boxes lack of interpretability, but by learning the dynamics of pollutants, the underlying deep networks would certainly capture some cause-and-effect relationships as forecasting performances increase. It requires joint efforts from scientists in both atmospheric chemistry and theoretical deep learning to make deep networks more transparent for air quality forecasts.

Funding This work is supported by the Pioneer Hundred Talents Program of Chinese Academy of Sciences, the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDC01040100) and the Key Program of National Natural Science Foundation of China (Grant No. 91644216).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

1. Chen LJ, Ho YH, Lee HC, Wu HC, Liu HM, Hsieh HH, et al. An open framework for participatory pm2.5 monitoring in smart cities. IEEE Access. 2017;5:14441–54. https://doi.org/10.1109/ACCESS.2017.2723919.
2. Kampa M, Castanas E. Human health effects of air pollution. Environ Pollut. 2008;151(2):362–7. https://doi.org/10.1016/j.envpol.2007.06.012.
3. Released, In New Estimates . 7 million premature deaths annually linked to air pollution. Air Quality and Climate Change. 2014. http://www.who.int/mediacentre/news/releases/2014/air-pollution/en/#.UzDo2BhK4xo.facebook.
4. Geng G, Zhang Q, Martin RV, Donkelaar AV, Huo H, Che H, et al. Estimating long-term PM 2.5 concentrations in China using satellite-based aerosol optical depth and a chemical transport model. Remote Sens Environ. 2015;166:262–70. https://doi.org/10.1016/j.rse.2015.05.016.
5. Jeffrey DS, Ashkan A, Emmanuela G, Stephen SL, Degu A, Kalkidan HA, et al. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. Lancet. 2018;392(10159):1923–94. https://doi.org/10.1016/S0140-6736(18)32225-6.
6. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the global burden of disease study 2010. Lancet. 2012;380(9859):2224–60. https://doi.org/10.1016/S0140-6736(12)61766-8.
7. Rohde RA, Muller RA. Air pollution in China: mapping of concentrations and sources. PLoS One. 2015;10(8):e0135749. https://doi.org/10.1371/journal.Pone.0135749.
8. Kang GK, Gao JZ, Chiao S, Lu S, Xie G. Air quality prediction: big data and machine learning approaches. Int J Environ Sci Dev. 2018;9(1):8–16. https://doi.org/10.18178/ijesd.2018.9.1.1066.
9. Zheng Y, Liu F, Hsieh HP. U-Air: when urban air quality inference meets big data. Proceedings of the 19th SIGKDD conference on Knowledge Discovery and Data Mining, 2013;1436–1444. https://www.microsoft.com/en-us/research/publication/u-air-when-urban-air-quality-inference-meets-big-data/.
10. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–44. https://doi.org/10.1038/nature14539.
11. Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Nuno Carvalhais N, et al. Deep learning and process understanding for data-driven earth system science. Nature. 2019;566(7743):195–204. https://doi.org/10.1038/s41586-019-0912-1.
12. Zhu XX, Tuia D, Mou L, Xia GS, Fraundorfer F. Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geosci Remote Sens Mag. 2017;5(4):8–36. https://doi.org/10.1109/MGRS.2017.2762307.
13. Zhang Y, Bocquet M, Mallet V, Seigneur C, Baklanov A. Real-time air quality forecasting, part I: history, techniques, and current status. Atmos Environ. 2012;60:632–55. https://doi.org/10.1016/j.atmosenv.2012.06.031.
14. Debry E, Mallet V. Ensemble forecasting with machine learning algorithms for ozone, nitrogen dioxide and PM10 on the Prev'Air platform. Atmos Environ. 2014;91:71–84. https://doi.org/10.1016/j.atmosenv.2014.03.049.
15. Fan J, Li Q, Hou J, Feng X, Karimian H, Lin S. A spatiotemporal prediction framework for air pollution based on deep RNN. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2017;4:15. https://doi.org/10.5194/isprs-annals-IV-4-W2-15-2017.
16. Byun QW, Ching JK. Science Algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System. Washington, DC, USA: Environmental Protection Agency; 1999. http://www.epa.gov/asmdnerl/models3/doc /scence/science.html
17. Chen J, Lu J, Avise JC, DaMassa JA, Kleeman MJ, Kaduwela AP. Seasonal modeling of PM2.5 in California's San Joaquin Valley. Atmos Environ. 2014;92:182–90. https://doi.org/10.1016/j.atmosenv.2014.04.030.
18. Grell GA, Peckham SE, Schmitz R, McKeen SA, Frost G, Skamarock WC, et al. Fully coupled “online” chemistry within the WRF model. Atmos Environ. 2005;39(37):6957–75. https://doi.org/10.1016/j.atmosenv.2005.04.027.
19. Wang Z, Maeda T, Hayashi M, Hsiao L, Liu K. A nested air quality prediction modeling system for urban and regional scales: application for high-ozone episode in Taiwan. Water Air Soil Pollut. 2001;130(1–4):391–6. https://doi.org/10.1023/a:1013833217916.
20. Zhang H, Linford JC, Sandu A, Sander R. Chemical mechanism solvers in air quality models. Atmosphere. 2011;2(3):510–32. https://doi.org/10.3390/atmos2030510.
21. Kelp MM, Tessum CW, Marshall JD. Orders-of-magnitude speedup in atmospheric chemistry modeling through neural network-based emulation. 2018.
22. Vautard R, Builtjes P, Thunis P, Cuvelier C, Bedogni M, Bessagnet B, et al. Evaluation and intercomparison of Ozone and PM10 simulations by several chemistry transport models over

four European cities within the CityDelta project. Atmos Environ. 2007;41(1):173–88. https://doi.org/10.1016/j.atmosenv.2006.07.039.
23. Ziegel ER, Box GEP, Jenkins GM, Reinsel GC. Time series analysis, forecasting, and control. Technometrics. 1995;37(2):238. https://doi.org/10.2307/1269640.
24. Li C, Hsu NC, Tsay SC. A study on the potential applications of satellite data in air quality monitoring and forecasting. Atmos Environ. 2011;45(22):3663–75. https://doi.org/10.1016/j.atmosenv.2011.04.032.
25. Ma Z, Hu X, Huang L, Bi J, Liu Y. Estimating ground-level pm2.5 in China using satellite remote sensing. Environ Sci Technol. 2014;48(13):7436–44. https://doi.org/10.1021/es5009399.
26. Carrassi A, Bocquet M, Bertino L, Evensen G. Data assimilation in the geosciences: an overview of methods, issues, and perspectives. Wiley Interdiscip Rev Clim Chang. 2018;9(5):e535. https://doi.org/10.1002/wcc.535.
27. Wu L, Mallet V, Bocquet M, Sportisse B. A comparison study of data assimilation algorithms for ozone forecasts. J Geophys Res Atmos. 2008;113(D20). https://doi.org/10.1029/2008JD009991.
28. Elbern H, Schmidt H, Ebel A. Variational data assimilation for tropospheric chemistry modeling. J Geophys Res Atmos. 1997;102:15967–85. https://doi.org/10.1029/97JD01213.
29. Eibern H, Schmidt H. A four-dimensional variational chemistry data assimilation scheme for Eulerian chemistry transport modeling. J Geophys Res Atmos. 1999;104(D15):18583–98. https://doi.org/10.1029/1999JD900280.
30. Donoho DL. High-dimensional data analysis: The curses and blessings of dimensionality. In: AMS Conference on Math Challenges of the 21st Century. 2000. https://doi.org/10.1111/j.1751-0813.1937.tb04127.x.
31. Zhu Y. Ensemble forecast: a new approach to uncertainty and predictability. Adv Atmos Sci. 2005;22(6):781–8. https://doi.org/10.1007/BF02918678.
32. Monache LD, Deng X, Zhou Y, Stull R. Ozone ensemble forecasts: 1. A new ensemble design. J Geophys Res Atmos. 2006;111(D5). https://doi.org/10.1029/2005JD006310.
33. Mckeen SA, Chung SH, Wilczak J, Grell G, Yu S. Evaluation of several $PM_{2.5}$ forecast models using data collected during the icartt/neaqs 2004 field study. J Geophys Res Atmos. 2007;112(D10). https://doi.org/10.1029/2006JD007608.
34. Monache DL, Nipen T, Deng X, Zhou Y, Stull R. Ozone ensemble forecasts: 2. A Kalman filter predictor bias correction. J Geophys Res Atmos. 2006;111(D5). https://doi.org/10.1029/2005JD006311.
35. Garaud D, Mallet V. Automatic calibration of an ensemble for uncertainty estimation and probabilistic forecast: application to air quality. J Geophys Res Atmos. 2011;116(D19). https://doi.org/10.1029/2011JD015780.
36. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542:115–8. https://doi.org/10.1038/nature21056.
37. Donnelly A, Misstear B, Broderick B. Real time air quality forecasting using integrated parametric and non-parametric regression techniques. Atmos Environ. 2015;103(103):53–65. https://doi.org/10.1016/j.atmosenv.2014.12.011.
38. Chin C, Brown DE. Learning in science: A comparison of deep and surface approaches. Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching. 2000;37(2):109–38. https://doi.org/10.1002/(SICI)1098-2736(200002)37:23.0.CO;2-7.
39. Gardner M, Dorling S. Artificial neural network (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos Environ. 1998;6(32):2627–36 https://doi-org-443.webvpn.las.ac.cn/10.1016/S1352-2310(97)00447-0.
40. Prybutok VR, Yi J, Mitchell D. Comparison of neural network models with ARIMA and regression models for prediction of Houston's daily maximum ozone concentration. Eur J Oper Res. 2000;122(1):31–40. https://doi.org/10.1016/S0377-2217(99)00069-7.
41. Perez P, Reyes J. An integrated neural network model for PM10 forecasting. Atmos Environ. 2006;40(16):2845–51. https://doi.org/10.1016/j.atmosenv.2006.01.010.
42. Aiswarya B, Aneena AA. A review on various techniques used in predicting pollutants. IOP Conf Ser Mater Sci Eng. 2018;396:012016. https://doi.org/10.1088/1757-899X/396/1/012016.
43. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.
44. Kolen JF, Kremer SC. Gradient flow in recurrent nets: the difficulty of learning LongTerm dependencies. A Field Guide to Dynamical Recurrent Networks: IEEE; 2001. p. 237–43. https://doi.org/10.1109/9780470544037.ch14.
45. Zaytar MA, Amrani CE. 2016. Sequence to sequence weather forecasting with long short-term memory recurrent neural networks. Int J Comput Appl. 2016;143(11):7–11. https://doi.org/10.5120/ijca2016910497.
46. Cho K, Van Merriënboer B, Bahdanau D, Bengio Y. On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259. https://arxiv.org/abs/1409.1259.
47. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. Comput Sci. 2014; http://arxiv.org/abs/1412.3555v1.
48. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 1998;86(11):2278–324. https://doi.org/10.1109/5.726791.
49. Springenberg JT, Dosovitskiy A, Brox T, Riedmiller M. Striving for simplicity: the all convolutional net. In International Conference on Learning Representation (ICLR), 2015. https://arxiv.org/abs/1412.6806v3.
50. Zhang C, Yan Z, Li C, Rui X, Liu L, Bie R. On estimating air pollution from photos using convolutional neural network. Proceedings of the 2016 ACM on Multimedia Conference; 2016. p. 297–301. https://doi.org/10.1145/2964284.2967230.
51. Klambauer G, Unterthiner T, Mayr A, Hochreiter S. Self-normalizing neural networks. In Proceedings of the Advances in Neural Information Processing Systems 30, 2017. https://arxiv.org/abs/1706.02515v4.
52. Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks. Adv Neural Inf Proces Syst. 2012;25(2):1097–105. https://doi.org/10.1145/3065386.
53. Oquab M, Bottou L, Laptev I, Sivic J. Learning and transferring mid-level image representations using convolutional neural networks. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR); 2014. p. 1717–24. https://doi.org/10.1109/CVPR.2014.222.
54. Maharana A, Nsoesie EO. Use of deep learning to examine the association of the built environment with prevalence of neighborhood adult obesity. JAMA Netw Open. 2018;1(4):e181535. https://doi.org/10.1001/jamanetworkopen.2018.1535.
55. Lv Y, Duan Y, Kang W, Li Z, Wang FY. Traffic flow prediction with big data: a deep learning approach. IEEE Trans Intell Transp Syst. 2015;16(2):865–73. https://doi.org/10.1109/TITS.2014.2345663.
56. Bengio Y, Lamblin P, Popovici D, Larochelle H. Advances in neural information processing systems, 2007;19:153–160. https://ieeexplore.ieee.org/ document/6287632.
57. Li X, Peng L, Hu Y, Shao J, Chi T. Deep learning architecture for air quality predictions. Environ Sci Pollut Res. 2016;23(22):22408–17. https://doi.org/10.1007/s11356-016-7812-9.

58. Wang Q, Lin J, Yuan Y. Salient band selection for hyperspectral image classification via manifold ranking. IEEE transactions on neural networks and learning systems 2016;27(6):1279–89. https://doi.org/10.1109/TNNLS.2015.2477537.
59. Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural Comput. 2006;18(7):1527–54. https://doi.org/10.1162/neco.2006.18.7.1527.
60. Hinton GE. Learning multiple layers of representation. Trends Cogn Sci. 2007;11(10):428–34. https://doi.org/10.1016/j.tics.2007.09.004.
61. Rumelhart DE, McClelland JL. Information processing in dynamical systems: foundations of harmony theory. in parallel distributed processing: explorations in the microstructure of cognition: Foundations, 1987;194–281. https://ieeexplore.ieee.org/document/6302931?T p=&arnumber=6302931.
62. Mohamed AR, Sainath TN, Dahl GE, Ramabhadran B, Hinton GE, Picheny MA. Deep belief networks using discriminative features for phone recognition. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing; 2011. https://doi.org/10.1109/ICASSP.2011.5947494.
63. Tolstikov A, Biswas J, Nugent C, Parente G. Comparison of fusion methods based on DST and DBN in human activity recognition. Control Theory Technol. 2011;9(1):18–27. https://doi.org/10.1007/s11768-011-0260-7.
64. Srivastava N, Salakhutdinov R. Multimodal learning with deep Boltzmann machines. Journal of Machine Learning Research 2012;15(8):1967–2006. https://doi.org/10.1162/NECO_a_00311.
65. Li T, Shen H, Yuan Q, Zhang X, Zhang L. Estimating ground-level PM2.5 by fusing satellite and station observations: A geo-intelligent deep learning approach. Atmos Ocean Phys. 2017; https://arxiv.org/abs/1707.03558v1.
66. Rodriguez JD, Perez A, Lozano JA. Sensitivity analysis of k-fold cross validation in prediction error estimation. IEEE Trans Pattern Anal Mach Intell. 2010;32(3):569–75. https://doi.org/10.1109/TPAMI.2009.187.
67. Stanley KO, D'Ambrosio DB, Gauci J. A hypercube-based encoding for evolving large-scale neural networks. Artif Life. 2009;15(2):185–212. https://doi.org/10.1162/artl.2009.15.2.15202.
68. Pan Z, Liang Y, Zhang J, Yi X, Yu Y, Zheng Y. HyperST-Net: Hypernetworks for Spatio-Temporal Forecasting. Mach Learn. 2018; https://arxiv.org/abs/1809.10889.
69. Li V, Lam J, Chen Y, Gu J. Deep learning model to estimate air pollution using M-BP to fill in missing proxy urban data. GLOBECOM 2017–2017 IEEE Global Communications Conference, 2017;1–6. https://doi.org/10.1109/GLOCOM.2017.8255004.
70. Qi Z, Wang T, Song G, Hu W, Li X, Zhang Z. Deep air learning: interpolation, prediction, and feature analysis of fine-grained air quality. IEEE Transactions on Knowledge and Data Engineering 2018;30(12):2285–97. https://doi.org/10.1109/TKDE.2018.2823740.
71. Hong KY, Pinheiro PO, Weichenthal S. Predicting global variations in outdoor PM2:5 concentrations using satellite images and deep convolutional neural networks: Image and Video Processing; 2019. https://arxiv.org/abs/1906.03975
72. Li T, Shen H, Yuan Q, Zhang L. A novel solution for remote sensing of air quality: from satellite reflectance to ground pm2.5. Atmos Ocean Phys. 2017.
73. Bui TC, Le VD, Cha SK. A Deep Learning Approach for Forecasting Air Pollution in South Korea Using LSTM. Mach Learn. 2018; https://arxiv.org/abs/1804.07891v3.
74. Freeman BS, Taylor G, Gharabaghi B, Thé J. Forecasting air quality time series using deep learning. J Air Waste Manage Assoc. 2018;68(8):866–86. https://doi.org/10.1080/10962247.2018.1459956.
75. Li X, Peng L, Yao X, Cui S, Hu Y, You C, et al. Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation. Environ Pollut. 2017;231:997–1004. https://doi.org/10.1016/j.envpol.2017.08.114.
76. Kim S, Lee JM, Lee J, Seo J. Deep-dust: predicting concentrations of fine dust in Seoul using LSTM. Clim Inform. 2019; https://arxiv.org/abs/1901.10106.
77. Huang CJ, Kuo PH. A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities. Sensors. 2018;18(7):2220. https://doi.org/10.3390/s18072220.
78. Soh PW, Chang JW, Huang JW. Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations. IEEE Access, 2018;6:38186–99.
79. Wang H, Zhuang B, Chen Y, Li N, Wei D. Deep inferential spatial-temporal network for forecasting air pollution concentrations. Mach Learn. 2018; https://arxiv.org/abs/18 09.03964v1.
80. Keller CA, Evans MJ. Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10. Geosci Model Dev. 2019;12(3):1209–25. https://doi.org/10.5194/gmd-12-1209-2019.
81. Kelp MM, Tessum CW, Marshall JD. Orders-of-magnitude speedup in atmospheric chemistry modeling through neural network-based emulation. Atmos Ocean Phys. 2018; https://arxiv.org/abs/1808.03874.
82. Kelp M, Jacob DJ, Kutz JN, Marshall JD, Tessum C. Toward stable, general machine learned models of the atmospheric chemical system. Submitted to JGR: Atmospheres. 2020.
83. Ma D, Zhang Z. Contaminant dispersion prediction and source estimation with integrated Gaussian-machine learning network model for point source emission in atmosphere. J Hazard Mater. 2016;311:237–45. https://doi.org/10.1016/j.jhazmat.2016.03.022.
84. Kumar S, Torres C, Ulutan O, Ayasse, A, Roberts D, Manjunath BS. Deep remote sensing methods for methane detection in overhead Hyperspectral imagery. In the IEEE Winter Conference on Applications of Computer Vision, 2020;1776–1785. https://doi.org/10.1109/WACV45572.2020.9093600.
85. Tobler WR. A computer movie simulating urban growth in the Detroit region. Econ Geogr. 1970;46(2):234–40. https://doi.org/10.2307/143141.
86. Rao TS. Statistics for spatio-temporal data. J Time. 2012;33(4):699–700. https://doi.org/10.1111/j.1467-9892.2011.00765.x.
87. Rakthanmanon T, Campana B, Mueen A, Batista G, Westover B, Zhu Q, et al. Searching and mining trillions of time series subsequences under dynamic time warping. Proc ACM SIGKDD Int Conf Knowl Discov Data Min. 2012;2012:262–70. https://doi.org/10.1145/2339530.2339576.
88. Keogh E, Ratanamahatana CA. Exact indexing of dynamic time warping. Knowl Inf Syst. 2005;7(3):358–86. https://doi.org/10.1007/s10115-004-0154-9.
89. Soh PW, Chen KH, Huang JW, Chu HJ. Spatial-temporal pattern analysis and prediction of air quality in Taiwan. 10th International Conference on Ubi-media Computing and Workshops (Ubi-Media); 2017. p. 1–6. https://doi.org/10.1109/umedia.2017.8074094.
90. Yi X, Zhang J, Wang Z, Li T, Zheng Y. Deep distributed fusion network for air quality prediction. the 24th ACM SIGKDD International Conference: ACM; 2018. https://doi.org/10.1145/3219819.3219822.
91. Lin Y, Mago N, Gao Y, Li Y, Chiang YY, Shahabi C, et al. Exploiting spatiotemporal patterns for accurate air quality forecasting using deep learning. In: The 26th ACM SIGSPATIAL International Conference; 2018. p. 359–68. https://doi.org/10.1145/3274895.3274907.
92. Afonin SV. An appraisal of the method of AOD retrieval over land according to MODIS satellite measurements in IR spectral range.

Atmos Ocean Optics. 2011;24(6):584–6. https://doi.org/10.1134/S1024856011060029.
93. Li T, Shen H, Zeng C, Yuan Q, Zhang L. Point-surface fusion of station measurements and satellite observations for mapping pm2.5 distribution in china: methods and assessment. Atmos Environ. 2016. https://doi.org/10.1016/j.atmosenv.2017.01.004.
94. Xiao Q, Wang Y, Chang HH, Meng X, Geng G, Lyapustin A, et al. Full-coverage high-resolution daily PM 2.5 estimation using MAIAC AOD in the Yangtze River Delta of China. Remote Sens Environ. 2017;199:437–46. https://doi.org/10.1016/j.rse.2017.07.023.
95. Moon T, Wang Y, Liu Y, Yu B. Evaluation of a MISR-based high-resolution aerosol retrieval method using AERONET DRAGON campaign data. IEEE Trans Geosci Remote Sens. 2015;53(8):4328–39. https://doi.org/10.1109/tgrs.2015.2395722.
96. Brauer M, Amann M, Burnett RT, Cohen A, Dentener F, Ezzati M, et al. Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution. Environ Sci Technol. 2012;46(2):652–60. https://doi.org/10.1021/es2025752.
97. Di Q, Koutrakis P, Schwartz J. A hybrid prediction model for PM2.5 mass and components using a chemical transport model and land use regression. Atmos Environ. 2016;131:390–9. https://doi.org/10.1016/j.atmosenv.2016.02.002.
98. Gupta P, Christopher SA. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: multiple regression approach. J Geophys Res. 2009;114:D14205. https://doi.org/10.1029/2008JD011496.
99. Lary DJ, Alavi AH, Gandomi AH, Walker AL. Machine learning in geosciences and remote sensing. Geosci Front. 2015;7(1):3–10. https://doi.org/10.1016/j.gsf.2015.07.003.
100. Li H, Chang J, Xu F, Liu B, Liu Z, Zhu L, Yang Z. An RBF neural network approach for retrieving atmospheric extinction coefficients based on lidar measurements. Applied Physics B. 2018;124(9):184. https://doi.org/10.1007/s00340-018-7055-1.
101. Zhang B, Zhang M, Kang J, Hong D, Jian Xu J, Zhu X. Estimation of PMx concentrations from Landsat 8 OLI images based on a multilayer perceptron neural network. Remote Sens. 2019;11(6):646. https://doi.org/10.3390/rs11060646.
102. Athira V, Geetha P, Vinayakumar R, Soman KP. DeepAirNet: applying recurrent networks for air quality prediction. Procedia Comput Sci. 2018;132:1394–403. https://doi.org/10.1016/j.procs.2018.05.068.
103. Xayasouk T, Lee H. Air pollution prediction system using deep learning. WIT Trans Ecol Environ. 2018;230:71–9. https://doi.org/10.2495/AIR180071.
104. Lodge J. Handbook on atmospheric diffusion. Atmos Environ. 1983;17(3):673–5. https://doi.org/10.1016/0004-6981(83)90164-6.
105. Seaman NL. Meteorological modeling for air-quality assessments. Atmos Environ. 2000;34(12–14):2231–59. https://doi.org/10.1016/S1352-2310(99)00466-5.
106. Flesch TK, Wilson JD, Harper LA, Crenna BP, Sharpe RR. Deducing ground-to-air emissions from observed trace gas concentrations: a field trial. J Appl Meteorol. 2004;43(3):487–502. https://doi.org/10.1175/1520-0450(2004)043<0487:DGEFOT>2.0.CO;2.
107. Wilson JD, Sawford BL. Review of lagrangian stochastic models for trajectories in the turbulent atmosphere. Bound-Layer Meteorol. 1996;78(1–2):191–210. https://doi.org/10.1007/BF00122492
108. Pontiggia M, Derudi M, Busini V, Rota R. Hazardous gas dispersion: a CFD model accounting for atmospheric stability classes. J Hazard Mater. 2009;171(1–3):739–47. https://doi.org/10.1016/j.jhazmat.2009.06.064.
109. Mazzoldi A, Hill T, Colls JJ. CFD and Gaussian atmospheric dispersion models: a comparison for leak from carbon dioxide transportation and storage facilities. Atmos Environ. 2008;42(34):8046–54. https://doi.org/10.1016/j.atmosenv.2008.06.038.
110. So W, Koo J, Shin D, Yoon ES. The estimation of hazardous gas release rate using optical sensor and neural network. Computer Aided Chemical Engineering. Elsevier. 2010;28(10):199–204. https://doi.org/10.1016/S1570-7946(10)28034-3.
111. Wang B, Chen B, Zhao J. The real-time estimation of hazardous gas dispersion by the integration of gas detectors, neural network and gas dispersion models. J Hazard Mater. 2015;300:433–42. https://doi.org/10.1016/j.jhazmat.2015.07.028.
112. Qiu S, Chen B, Wang R, Zhu Z, Wang Y, Qiu X. Atmospheric dispersion prediction and source estimation of hazardous gas using artificial neural network, particle swarm optimization and expectation maximization. Atmos Environ. 2018;178:58–163. https://doi.org/10.1016/j.atmosenv.2018.01.056.
113. Shi X, Chen Z, Wang H, Yeung DY, Wong WK, Woo WC. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. Advances in neural information processing systems. 2015;802–810. https://arxiv.org/abs/1506.04214v1.
114. Shi X, Gao Z, Lausen L, Wang H, Yeung DY, Wong WK, Woo WC. Deep learning for precipitation nowcasting: A benchmark and a new model. In Advances in neural information processing systems. 2017;5617–27. https://arxiv.org/abs/1706.03458.
115. Xue T, Zheng Y, Tong D, Zheng B, Li X, Zhu T, et al. Spatiotemporal continuous estimates of PM2. 5 concentrations in China, 2000–2016: a machine learning method with inputs from satellites, chemical transport model, and ground observations. Environ Int. 2019;123:345–57. https://doi.org/10.1016/j.envint.2018.11.075.
116. Liu M, Bi J, Ma Z. Visibility-based PM2. 5 concentrations in China: 1957–1964 and 1973–2014. Environ Sci Technol. 2017;51(22):13161–9. https://doi.org/10.1021/acs.est.7b03468.