Machine learning based bias correction for numerical chemical transport models

Xu, Min; Jin, Jianbing ; Wang, Guoqiang ; Segers, Arjo; Deng, Tuo; Lin, Hai Xiang

DOI
10.1016/j.atmosenv.2020.118022

Publication date
2021

Document Version
Final published version

Published in
Atmospheric Environment

Citation (APA)
Xu, M., Jin, J., Wang, G., Segers, A., Deng, T., & Lin, H. X. (2021). Machine learning based bias correction for numerical chemical transport models. *Atmospheric Environment*, *248*, 1 - 10. Article 118022. https://doi.org/10.1016/j.atmosenv.2020.118022

Atmospheric Environment 248 (2021) 118022

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: http://www.elsevier.com/locate/atmosenv

Machine learning based bias correction for numerical chemical transport models

Min Xu [a], Jianbing Jin [b,*], Guoqiang Wang [a], Arjo Segers [c], Tuo Deng [d], Hai Xiang Lin [d]

[a] School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai, 201620, China
[b] Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control, Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, School of Environmental Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, 210044, China
[c] TNO, Department of Climate, Air and Sustainability, Utrecht, the Netherlands
[d] Delft Institute of Applied Mathematics, Delft University of Technology, Delft, the Netherlands

HIGHLIGHTS

- An approach based on machine learning is applied to predict bias in CTMs, this is the first time that it is used this way.
- Machine learning is an effective tool to improve the accuracy of CTM prediction, especially in case of short term forecast.
- This approach requires less computing power than other model bias correction techniques like data assimilation.

ARTICLE INFO

Keywords:
Air quality
Neural network
$PM_{2.5}$
Short term forecast

ABSTRACT

Air quality warning and forecasting systems are usually based on numerical chemical transport models (CTMs). Those dynamic models perform predictions by simulating the life cycles of the atmospheric components, including emission, transport and removal. However, the accuracy of these CTMs are still limited because of many imperfections, e.g., uncertainties in the input sources such as emission inventories, wind fields, boundary conditions, as well as insufficient knowledge about the atmospheric dynamics themselves. All these will mislead the CTM prediction constantly, or in a systematic way.

In this paper, an approach based on machine learning is applied to predict model bias in the CTM. It is then combined with the CTM for formulating a hybrid forecast system. To our knowledge, it is the first time that machine learning methods are used in this way. The hybrid system is tested on the fine particular matter ($PM_{2.5}$) prediction in Shanghai, China. The results showed that machine learning can be an effective tool to improve the accuracy of CTM prediction. In case of short term $PM_{2.5}$ forecast (forecast length less than 12 h), statistical metrics of the root mean square error, mean absolute error, mean absolute percentage error as well as the air quality rank predicted accuracy all show the forecast skill is remarkably improved; while for long term prediction, improvement is not ensured.

1. Introduction

Air pollution has wide-ranging effects on environments, biodiversity and sustainable growth (Manders et al., 2017). Atmospheric pollutants, e.g., aerosol, were long recognized to pose great threats to human health and atmospheric environments since early 1950s (Haagen-Smit, 1952). At present, the air pollution has been one of the biggest environmental risks to health, carrying responsibility for about three million deaths each year (World Health Organization, 2016).

To reduce the threats by the air pollution, early warning and forecasting system is essential for the public health. In recent years, many efforts have been made to enrich approaches for predicting air pollutant concentrations (Wang et al., 2012; Chen et al., 2015; Hu et al., 2017). Those methods generally fall into two major categories, numerical chemical transport models (CTMs) and data based methods.

Those numerical CTMs adopt physical principles and statistical methods to simulate the life cycles of the atmospheric components, mainly including emission, advection, diffusion, chemistry reactions

* Corresponding author.
E-mail address: jianbing.jin@nuist.edu.cn (J. Jin).

https://doi.org/10.1016/j.atmosenv.2020.118022
Received 19 March 2020; Received in revised form 14 October 2020; Accepted 16 October 2020
Available online 22 January 2021
1352-2310/

and deposition. The implementation of the CTMs can be traced back to the early 1970's (Reynolds et al., 1973). Since then this approach is applied extensively in practice. A large number of regional/global CTMs have been developed and used for operational air quality prediction, e. g., GEOS-CHEM (Bey et al., 2001; Simpson et al., 2012), CUACE (Zhou et al., 2012), LOTOS-EUROS (Manders et al., 2017), and WRF-CHEM (Abdi-Oskouei et al., 2018). Those CTMs not only help health professionals to predict the formation and spread of ambient pollutants, but also provide valuable information to policy makers in making effective emission reduction strategies. Despite efforts and advances in atmospheric modeling during the past decades, the accuracy of those CTM forecasts is still limited.

The perceived inadequacies in CTM model have different origins, e. g., the mismatch in transport (Gilliam et al., 2015; Solazzo et al., 2017), error in removal parametrization (Croft et al., 2012; van der Graaf et al., 2018), uncertainty in nature and anthropogenic emission quantification (Bates et al., 2006; Schutgens et al., 2012), and impossibility for models to resolve the fine-scale and vertical variability. For instance, CTM requires an essential temporal profile for describing anthropogenic emission inventory which only defines the total emission. Those existing profiles usually remain relatively simple and assumed to be spatially constant (Song et al., 2019). However, the emission temporal variability is actually much more complicated and influenced by a group of factors, e.g., urbanization, weather, and weekday/weekend. Those factors are supposed to be considered for building a representative emission module in CTMs. Imperfections in those CTMs are factors that influence the forecasts constantly, that is, in a systematic way. Therefore, they are considered as systematic errors in atmospheric modeling.

Data assimilation has already been widely used to correct those imperfect parameters which would then drive the new model runs and result in better forecasts. It is a powerful tool to feed the available observations into the dynamic models and adjust the uncertain sources in order to best fit both the model and observations (Sekiyama et al., 2010; Jin et al., 2018, 2019b, a). Existing data assimilation methods can be described as either sequential or variational (Talagrand, 1997). Though remarkable improvements on the CTM forecast accuracy were achieved using both of these two data assimilation algorithms (Kalnay, 2003), in practice, the implementation of data assimilation always requires a huge amount of computing power. For instance, the typical sequential data assimilation method, EnKF (Evensen, 1994), requires to forward ensemble model runs simultaneously. Whereas the variation data assimilation algorithm, e.g., 4DVar (Lorenc et al., 2015), requires a huge effort to build and maintain the extra adjoint models for realistic models.

As for data based methods, they avoid those sophisticated CTMs and predict air quality purely based on data (Reddy et al., 2018). At present, data driven methods generally refer to machine learning methods. The advances in the electronic sensor technology have made large scale air quality measurements feasible. A combination with the increasing power of computing platforms has led to a new paradigm in the computational and statistical methods for processing and analyzing data (Hey et al., 2009). Commonly used machine learning models include Support Vector Regression (SVR) (Osowski and Garanty, 2007), Random Forest (Yang et al., 2016), Hidden Markov Model (HMM) (Sun et al., 2013), and artificial neural network (ANN) (Feng et al., 2015). Among these models, the ANN, which can perform nonlinear mapping, generally provides superior performance for complex systems. It has attained enormous success in image classification, natural language processing, prediction tasks, object detection and motion modeling (Li et al., 2017b).

Machine learning has also been investigated in air quality simulation with various structures (Antanasijevic et al., 2013; Zhao et al., 2019). The algorithm is powerful in extracting representative air quality features without prior knowledge, which may lead to superior performance for air quality prediction (Li et al., 2016; Biancofiore et al., 2017; Fan et al., 2017; Xie et al., 2018). Another adding value using machine learning in operational forecast system is that the most time-consuming model training can be performed off-line. Once the model is trained, it is relatively fast to forward the trained model in real time.

However, weakness of machine learning applications on dynamic models is also obvious. For instance, the majority of the machine learning tools are purely data-driven and the knowledge about physical laws does not play any role, hence unrealistic results would be unavoidable (Lin et al., 2019). Machine learning to predict air quality can only be applied at a location where sufficient historical datasets are available, this in contrast to CTMs that are designed for predictions at millions or even billions grid location. Therefore, the data based method is constraint, and is not yet considered as a replacement of operational air quality forecast models. Satellite data, especially the geostationary one, with a large observing coverage is a potential new source for locations where no ground observing instrument is installed. However, the high data-missing rate problem needs to be solved before it can be used in real time air quality forecast.

With the perceived CTM systematic errors that are defined as model bias in this paper, we proposed a hybrid air quality forecast system by combining CTMs and machine learning. Machine learning is adopted to mathematically describe inherent relation between the targeted CTM bias and the features that are supposed to be related. They are air quality observations in the past several hours, meteorological and CTM forecast. The bias forecast would then be treated as a compensation of the existing air quality forecast system. To our knowledge, it is the first time that machine learning is applied to predict CTM bias in the operational air quality forecast system. Alternatively, the CTM bias correction could also be conducted by data assimilation. However, compared to data assimilation the machine learning method is free of model uncertainty analysis and require much less computing costs. The proposed method is capable of correcting bias in CTMs for all atmospheric components, e.g., $PM_{2.5}$, PM_{10}, NO_x and SO_2. However, it is our first work that explores the combination of machine learning and CTMs, the hybrid method is only tested on the short-term $PM_{2.5}$ prediction over several sites in Shanghai, China. Note that all CTM forecast is specifically referred as the $PM_{2.5}$ prediction in the following study. Within the same framework, we compared the predictive power of the hybrid system with forecast horizons from 1 to 12 h in advance. The results demonstrated that the hybrid system has achieved a significant performance improvement compared to the original CTM forecast.

The remaining paper is organized as the following: We start in Section 2 by briefly describing the operational CTM, LOTOS-EUROS, that is used in this study, as well as the main points of the data analysis procedure used. The hybrid system using a Multilayer Perceptron (MLP) for CTM bias correction is illustrated in Section 3. Section 4 evaluates the forecast skills of the hybrid system using independent measurements. Section 5 concludes the paper and further discusses future work.

2. CTM and data

2.1. Regional chemical transport model

The chemical transport model (CTM) and machine learning hybrid method is tested for $PM_{2.5}$ prediction in Shanghai. The LOTOS-EUROS regional CTM is used to simulate the atmospheric components over China. This CTM has been used for a wide range of applications supporting scientific research, regulatory programs and air quality forecasts all over the world. At present, its operational forecasts over China are released via the MarcoPolo-Panda projects (Timmermans et al., 2017; Brasseur et al., 2019) through the link http://www.marcopolo-panda.eu/forecast/(last access: Feb 2020).

To establish the air quality simulation system over China, the LOTOS-EUROS model is configured on a domain from 15° N to 50° N and 70° E to 140° E as shown in Fig. 1(a), with resolution 0.25° × 0.25°. Vertically, the model consists of 8 mixing layers with the top layer at 10 km. The model is driven by European Center for Medium-Ranged Weather Forecast (ECMWF) operational forecasts for forecast steps of

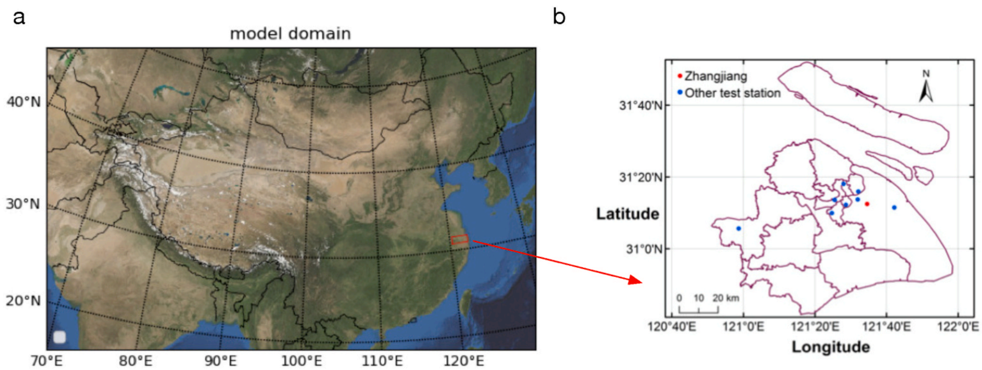

Fig. 1. (a): LOTOS-EUROS model domain; (b): China MEP air quality monitoring network over Shanghai.

3–12 h, starting from the 00:00 and 12:00 analyses. The data is extracted from the archive at regular longitude/latitude grid of about 7 km resolution. Physical processes included are advection, diffusion, dry and wet deposition, chemistry reaction and sedimentation. More details regarding the LOTOS-EUROS can be found in (Manders et al., 2017), and an open-source version of the model can be acquired through the website https://lotos-euros.tno.nl/.

2.2. CTM bias

As aforementioned, the CTMs forecast skills are limited due to the various challenges including uncertain meteorological input, boundary conditions, emission inventories as well as coarse model resolutions to represent fine-scale atmospheric component variability. Therefore, the discrepancies between CTM predictions against real observations are unavoidable, which is defined as CTM bias as shown in Fig. 2(a). The systematic bias itself can be treated as the state variable like the atmospheric tracer states. Note that the CTM bias is also caused by observation errors partially next to the perceived model imperfections in practice. Representation error is considered as the dominant source when CTM result is directly compared against the observation, since the former one represents the mean status over 25 km × 25 km grid cell while the in-situ measurement covers the atmosphere less than 5 km surrounding them (Schutgens et al., 2016). Therefore, the model bias calculated by subtracting the CTM forecasts from ground-based observations might only reflect the major part of bias levels, but not the reality.

This study aims to estimate the mathematical relations between the bias and available features based on training dataset off-line. As can be seen in Fig. 2(a), the trained model will be then applied to estimate the operational CTM forecast bias in real time. In practice, the sum of CTM and the machine learning based bias-estimation can be taken as the new air quality forecast, which is defined as hybrid forecast in this paper.

2.3. Air quality monitoring system

Since 2013, the China Ministry of Environmental Protection (MEP) has commenced to release the hourly-average measurements of atmospheric constituents including $PM_{2.5}$, PM_{10}, NO_2, CO, O_3 and SO_2 (Li et al., 2017a). A large number of ground stations measuring these air quality indices have been established in densely populated areas. Enhanced by the high temporal resolutions and the rather dense monitoring network, the rich dataset makes it possible to dig out the inherent temporal varying patterns of pollutants.

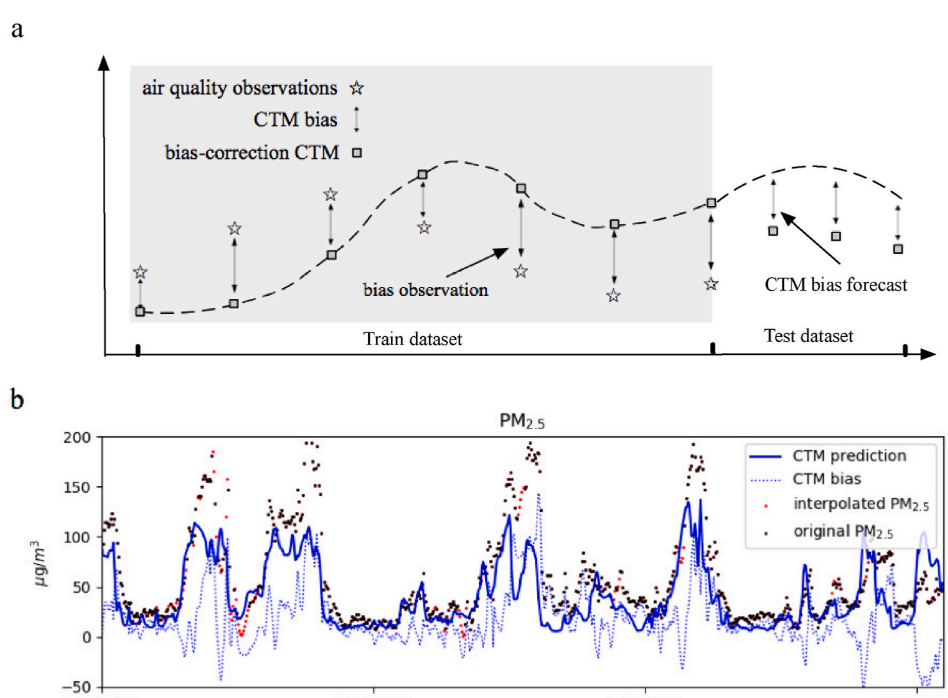

Fig. 2. (a): The CTM bias; (b): CTM prediction, CTM bias and $PM_{2.5}$ observations over the tested site Zhangjiang in Shanghai in December 2016.

By 2016, there were 9 monitoring stations over Shanghai, the CTM bias forecast tests are performed all over the 9 sites shown in Fig. 1(b). These measurements not only serve as parts of the input features in the machine learning model, but also involve in interpolation processing that are described in the Section 2.4.

2.4. Data interpolation

The essential step of the proposed methodology is dataset analysis and pre-processing, which is an important step in almost all machine learning problems. In our air quality time series application, the data missing is the major problem. Data missing errors occur due to hardware bugs in data acquisition, incorrect sensor readings or software bugs in the data processing pipeline. For instance, the $PM_{2.5}$ missing rate over one of the test air quality monitoring stations, Zhangjiang, over 2016 is about 7.5%.

To have the continuous time series for machine learning training and the continuous forecast test, data interpolations, specifically a k-Nearest-Neighbor algorithm (Zhang, 2012) and a cubic spline interpolation (Kincaid et al., 2009), of air quality measurements (PM_{10}, $PM_{2.5}$, SO_2, NO_2, O_3 and CO) are sequentially implemented. All interpolated measurements are constrained with a minimum of 0. A scenario of the original and interpolated $PM_{2.5}$ concentration over the test site Zhangjiang can be seen in Fig. 2(b).

2.5. Forecast evaluation

The Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) are used to evaluate the $PM_{2.5}$ prediction either using CTM or hybrid prediction system. Denote the $PM_{2.5}$ concentration and the simulated $PM_{2.5}$ concentration as y_t and $\hat{y}_t$, respectively, the statistical metrics, namely the RMSE, MAE, and MAPE, are calculated via:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\hat{y}_t - y_t\right)^2} \tag{1}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|\hat{y}_t - y_t\right| \tag{2}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|\hat{y}_t - y_t\right|}{y_t} \tag{3}$$

where n is the size of the dataset. The indicators of RMSE, MAE and MAPE will be used to measure the overall fitness of the results with the reality.

In practice, the air quality rank based on the National Technical Regulation of the Ambient Air Quality Index are released to public together with the airborne pollutant concentrations in China (Li et al., 2017b). As shown in Table 1, air quality is classified into 6 health-related ranks for different pollutant levels. These ranks are designed to help public better understand risks exposed in the situation of the certain air pollution. In this paper, we also generate the rank of the $PM_{2.5}$ predictions (CTM and hybrid method) and the measurements. The overall predicted rank true accuracy is calculated via:

$$\mathscr{A}_k = \frac{\sum_{i=k}^{6}\mathscr{N}_i}{\sum_{i=k}^{6} n_i} \tag{4}$$

where $\mathscr{A}_k$ represents the predicted accuracy for air quality situation in ranks [Lk, L6], N_i denotes the number of cases that prediction falls into the same rank i to the $PM_{2.5}$ observation; while n_i represents the sum of cases that the actual $PM_{2.5}$ level is in the rank i. When $k = 3$, $\mathscr{A}_3$ represents the air quality predicted rank accuracy in unhealthy situations.

Table 1
Ambient air quality index.

Rank	Range (μg/m^3)	Description
L1	<35	Good
L2	(35, 75)	Moderate
L3	(75, 115)	Unhealthy for sensitive groups
L4	(115, 150)	Unhealthy
L5	(150, 250)	Very unhealthy
L6	>250	Hazardous

3. Machine learning based CTM bias estimation systems

3.1. Multilayer Perceptron architecture

While many methods such as regressions and artificial neural networks could be applied to estimate the CTM bias, we adopt the open-loop Multilayer Perceptron neural network (MLP NN) in this research. Note that other machine learning models, e.g., long short term memory and random forest, are also tested and resulted comparable results. The focus of this paper is to investigate the applicability of machine learning to estimate CTM bias in a hybrid system. The question, which is the best machine learning method for CTM bias estimation is out the scope of this research, without loss of generality we choose the simple MLP NN that suffices to demonstrate performance of the hybrid system.

MLP is a specific structure of ANN, which consists of a set of input nodes, one or more sets of hidden nodes and a set of output nodes. It is a nonlinear and data-driven adaptive information processing system that models the way that the brain processes information (Gardner and Dorling, 1998). In this structure, each neuron is connected to several of its neighbors, with varying weights representing the relative influence of the different neuron inputs to the other neurons. The weighted simulation of the inputs is transmitted to the hidden neurons, where it is transformed using an activation function. In turn, the outputs of the hidden neurons serve as inputs to the output neuron where they undergo another transformation. The architecture of the MLP is illustrated in Fig. 3.

3.2. MLP based CTM bias forecast

As aforementioned, the CTM bias $\boldsymbol{b}$ is defined as the difference between CTM prediction $\boldsymbol{x}$ and observation $\boldsymbol{y}$. At any instant t, a bias prediction $\boldsymbol{b}^{t+t_0}$ t_0 hour in advance will be made using the trained MLP model. The hybrid air quality is calculated as

$$\boldsymbol{x}_h^{t+t_0} = \boldsymbol{x}^{t+t_0} + \boldsymbol{b}^{t+t_0} \tag{5}$$

where the $\boldsymbol{x}^{t+t_0}$ is extracted from the operational air quality model LOTOS-EUROS, and bias is provided by the trained MLP model $\mathscr{L}_\theta$ which is configured as:

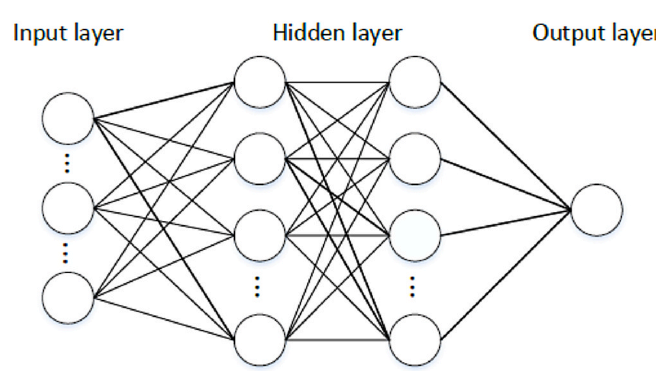

Fig. 3. The MLP network structure.

$$\boldsymbol{b}^{t+t_0} = \mathscr{L}_\theta\left(\boldsymbol{x}^{t+t_0}, \cdots, \boldsymbol{x}^{t+t_0-m+1}, \boldsymbol{w}^{t+t_0}, \cdots, \boldsymbol{w}^{t+t_0-m+1}, \boldsymbol{y}^t, \cdots, \boldsymbol{y}^{t-m+1}\right) \quad (6)$$

The transport of atmospheric components is driven by the atmospheric dynamic, hence the weather prediction is configured as parts of the inputs.

Specifically, the input feature vectors consist of:

- $\boldsymbol{x}^{t+t_0}, \cdots, \boldsymbol{x}^{t+t_0-m+1}$: LOTOS-EUROS $PM_{2.5}$ forecast over the t_0 hours at the test site;
- $\boldsymbol{w}^{t+t_0}, \cdots, \boldsymbol{w}^{t+t_0-m+1}$: Meteorological prediction over the t_0 hours, including surface dew point, temperature, horizontal wind speed at 10 m height over the test site, the data is extracted from ECMWF product that is used to drive the LOTOS-EUROS model;
- $\boldsymbol{y}^t, \cdots, \boldsymbol{y}^{t-m+1}$: air quality measurements from the last m hours, including $PM_{2.5}$, O_3, NO_2, SO_2 and CO hourly concentration.

In the data auto-correlation analysis, the $PM_{2.5}$ measurements are founded with a strong correlation $\mathscr{R} > 0.6$ with the time distance less than 9 h, hence the feature length m is set as 9 h.

3.3. Experiment setup

At any instant, the trained dataset could be rolling updated with the fresh data when the forecast system forwards. This on-line learning scheme might help to enrich the training dataset, but also require a large amount of computing power to train the models in real time. To save the efforts, off-line training strategy is adopted in this study.

The record covering the time period from 2016 January to August is used as training dataset, dataset extracted from September to October is used for validation. The trained model would then be used to forecast the CTM bias in the following two months, from 2016 November to December, which is used as test period.

The best choice of hyper-parameters, such as the number of hidden layers and neurons per layers, are selected to best fit the training dataset using the grid search method (Xie, 2018). Whether the number of epochs is sufficient for convergence is checked through comparing the training and validation loss in each epoch. The optimal hyper-parameters of MLP for modeling our LOTOS-EUROS $PM_{2.5}$ bias are listed in Table 2.

The choice of these hyper-parameters results in relatively consistent result in training, validating and testing dataset, respectively. A snapshot of statistic errors in term of RMSE over one of the test sites can be found in the Supplementary Material.

4. Results and discussion

4.1. Time series analysis

The $PM_{2.5}$ forecast over Shanghai by our CTM and by the hybrid prediction system are compared. First, the time series of LOTOS-EUROS, hybrid forecast ($t_0 = 3$ h), as well as the $PM_{2.5}$ observation at the test site Zhangjiang (location can be found in Fig. 1) over the whole December is shown in Fig. 4. It is selected to show the typical phenomenon that is observed in other sites all over Shanghai. In general, the LOTOS-EUROS results fit the observations very well in this test site, most of the increase and decline trends in $PM_{2.5}$ level are captured. Another advantage of the CTM compared to the data based model is that time and space information is more detailed, which helps to communicate the predictions to the public. But the difference is also obvious, specifically, CTM predictions agree well with the observations in case that slight $PM_{2.5}$ levels are present; but usually underestimate those high-value peaks. For instance, Fig. 4(c) shows there was an aerosol plume crossing over the test site from December 16 to 17, and LOTOS-EUROS almost reproduced the profile of the $PM_{2.5}$ growth and vanish. However, the simulated $PM_{2.5}$ indicated a peak of 100 μg/m^3 while the real maximum reached up to 210 μg/m^3. Besides, the CTM sometimes also mis-captured the $PM_{2.5}$ trend. For example, the simulated $PM_{2.5}$ seems to arrive 10 h earlier than reality around December 29. One possible reason for such underestimation or incorrect timing profile could be the usage of coarse emission timing inventory to capture the actual dynamic aerosol eruptions in CTMs. These imperfections are relatively difficult to be corrected through manually nudging the total aerosol emission rate in CTMs.

In comparison, the hybrid prediction curve can track the real $PM_{2.5}$ concentration values more accurately. The time series plotting results using the machine learning correction technique in Fig. 4 captures the trends most of the time, and the simulation minus observation gaps are reduced significantly. Not only those $PM_{2.5}$ peaks underestimated by LOTOS-EUROS (such as the one from Dec 16 to 17) are captured correctly, also the incorrect simulated $PM_{2.5}$ timing profile (e.g., the one around Dec 29) is fully reproduced.

4.2. Forecast skill evaluation

To evaluate the general performance of CTM and our bias correction method, the distribution of the original $PM_{2.5}$ observations (without interpolation) vs. CTM forecast, and vs. the hybrid prediction over all sites from the whole test period (November and December) are shown in Fig. 5. Panel (a) indicates there is a systematic underestimation in CTM prediction especially when high-value $PM_{2.5}$ levels are observed, and also a relatively wide spread (the narrower the better). The results of the hybrid system at different forecast lengths ($t_0 = 1, 3, 6, 9$ h) vs. $PM_{2.5}$ observations are plotted in Fig. 5(b)–(e). The application of the bias correction techniques effectively reduced the systematic underestimation, thereby produced an unbiased match that scatters around the 1:1 line (i.e., perfect prediction) for all the four cases. Compared to the distribution in Fig. 5(a), all these four scatter plots from hybrid predictions result in narrower spread, though they grows wider with the increase of the forecast length t_0. This is because the accuracy of hybrid system fundamentally relies on the correlations between the targeted CTM bias and input features, which inevitably vanishes with the increase of t_0.

To evaluate the forecast skills of the hybrid system over the CTM in $PM_{2.5}$ forecast, statistical metrics including RMSE, MAE, and MAPE of the prediction results with original $PM_{2.5}$ data are calculated. Fig. 6 shows the variation of these indicators either using CTM or using hybrid prediction methods with different forecast lengths (1~12 h in advance). Though our LOTOS-EUROS generally captured the $PM_{2.5}$ variation, the RMSE, MAE, MAPE of the CTM prediction is as high as 27.6 μg/m^3, 18.7 μg/m^3 and 70.6%, respectively. In comparison, the RMSE of hybrid prediction is steadily reduced to a lower level. For instance, the RMSE, MAE and MAPE of the hybrid forecast 3 h in advance stays around 16.1 μg/m^3, 11.9 μg/m^3 and 24.9%.

While the off-line machine learning based hybrid forecast could steadily improve the CTM forecast of $PM_{2.5}$ with only a low computing cost and high computing efficiency, it is found its accuracy reduced rapidly with increasing forecast time length. With a forecast length of 12 h, the improvement on $PM_{2.5}$ prediction are very limited in terms of RMSE and MAE, which are 22.6 μg/m^3 and 15.5 μg/m^3. Note that this might be further improved through either using more complex machine learning models which could resolve deeper relation between input features and the CTM bias, or using richer dataset. However, our hybrid method will be very limited in long-term (t_0) air quality prediction. Since the CTM bias that exhibits in the long run is weakly governed/related to those aforementioned input features.

Table 2
MLP hyper-parameters.

layers	neurons per layer	epochs	batch size
2	200	50	128

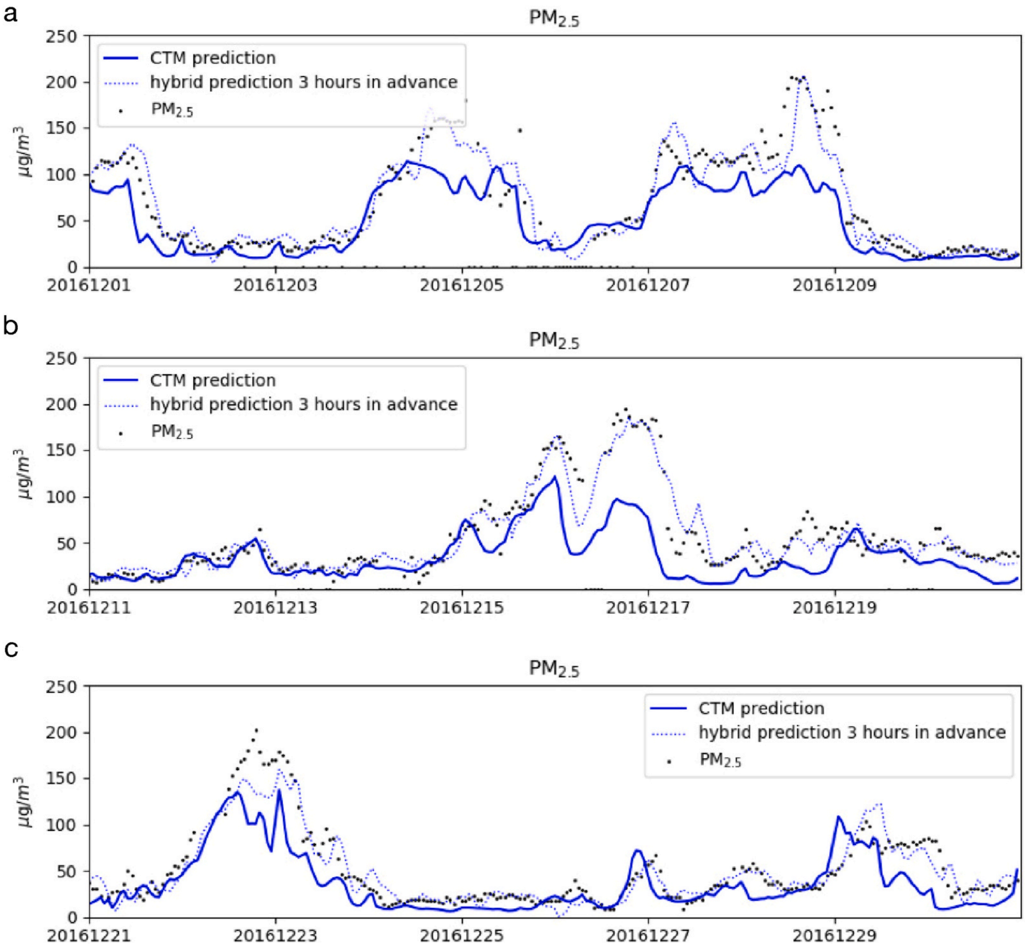

Fig. 4. Time series of the $PM_{2.5}$ observation, CTM prediction and hybrid forecasts with 3 h in advance in the test site of Zhangjiang. (a): Dec 01–10; (b): Dec 11–20; (c): Dec 21–30.

4.3. Air quality rank prediction

The hybrid method is further evaluated in terms of the air quality rank prediction. Fig. 7 shows the confusion matrix of CTM result, and our hybrid prediction over all test sites in Shanghai at different forecast length. For a confusion matrix, values in the main diagonal correspond to the times when the recorded air quality ranks are identical to the predicted air quality ranks, which indicate the ideal prediction, and an ideal confusion matrix is a diagonal matrix.

In general, our CTM performs as good as the hybrid method in case of health air quality condition (L1, L2); in both of the methods, over 70% cases are accurately predicted. In unhealthy situations (L3, L4, L5, L6), however, CTM accurately predicted numbers in the diagonal are much smaller than those off (over) the diagonal. This is consistent with what we observed in the Section. 4.1 that our LOTOS-EUROS sometimes cannot reproduce the high-value $PM_{2.5}$. For CTM, only 30% of the unhealthy air quality rank can be predicted. While for hybrid methods, the unhealthy air quality rank prediction accuracy is raised up to 73%, 54%, 45% and 43% with a forecast length $t_0 = 1, 3, 6, 9$ h. Therefore, the hybrid method effectively helps the public to foresee the air quality condition in practice.

5. Conclusion and future work

Despite the advances in chemical transport models (CTMs) for air quality forecasting, the accuracy is still not always adequate. There is still room for improvements, and the systematic error or bias is an important factor that needs to be reduced. Usually, data assimilation is the first option for estimating uncertain model parameters, which requires lots of computational power however. As an alternative, we explored a machine learning method for modeling the CTM bias with a high computational efficiency in this paper. The machine learning based bias forecasting method was then combined with the CTM for a hybrid prediction system with improved performance.

The proposed hybrid methodology could be implemented for prediction of all kinds of atmospheric components, e.g., PM_{10}, NO_x and SO_2, and is tested for $PM_{2.5}$ concentration predictions over 9 sites in Shanghai in this study. Performance of the hybrid system and the CTM, LOTOS-EUROS, is compared. The results demonstrate that machine learning, specifically Multilayer Perceptron neural network (MLP NN), is able to estimate the systematic mismatch that LOTOS-EUROS encounters. The hybrid system has a better performance than LOTOS-EUROS in terms of RMSE, MAE, MAPE as well as the predicted air quality rank accuracy $\mathscr{A}$. For instance, the predicted accuracy $\mathscr{A}_3$ in unhealthy air quality situations is raised from 30% (CTM only) to 54% (hybrid) with a forecast horizon of 3 h.

Although our results suggest that the hybrid method can reduce the CTM bias significantly, the relative improvement for a longer forecast length ($t_0 > 12$ h) is limited since the model exhibits different bias characteristics at different interval ranges. Since our machine learning bias-correction techniques can only effectively reduce systematic errors in the model within a shorter time interval, additional research is needed to develop methods to reduce systematic error for longer time horizon to further improve the forecast accuracy.

In future work, comparisons of using different machine learning models, and dataset covering a larger model region, e.g., the whole East

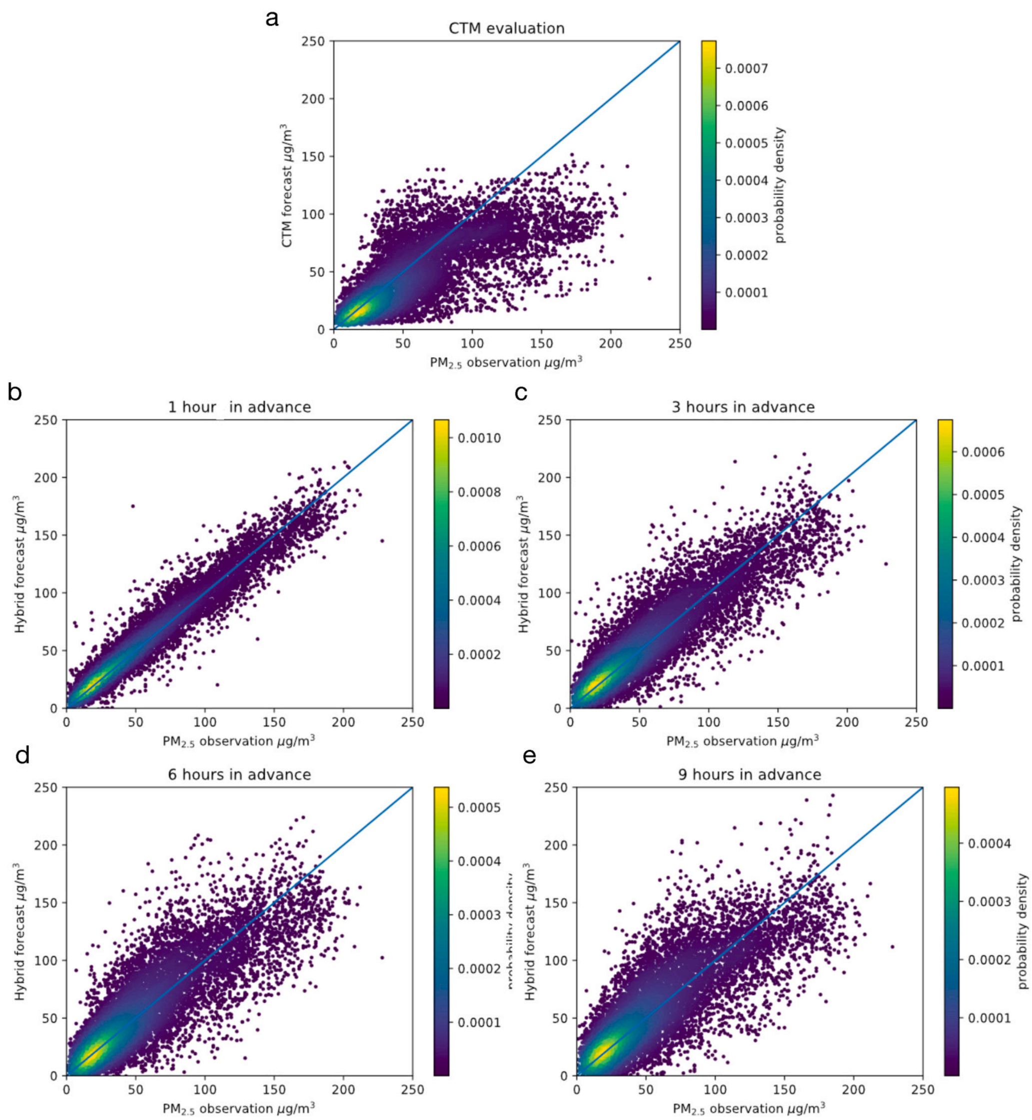

Fig. 5. (a): CTM prediction vs. $PM_{2.5}$ observations; (b): Hybrid prediction 1 h in advance; (c): Hybrid prediction 3 h in advance; (d): Hybrid prediction 6 h in advance; (e): Hybrid prediction 9 h in advance vs. $PM_{2.5}$ measurements.

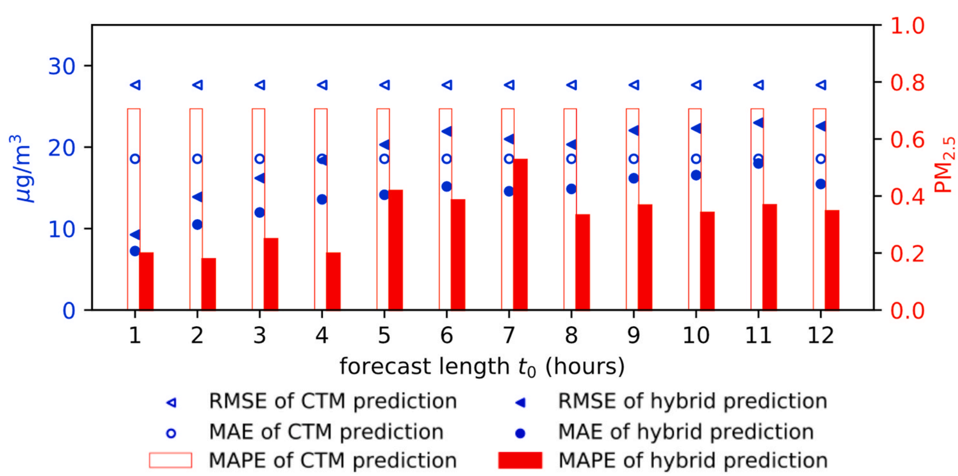

Fig. 6. RMSE, MAE, and MAPE at different forecast length.

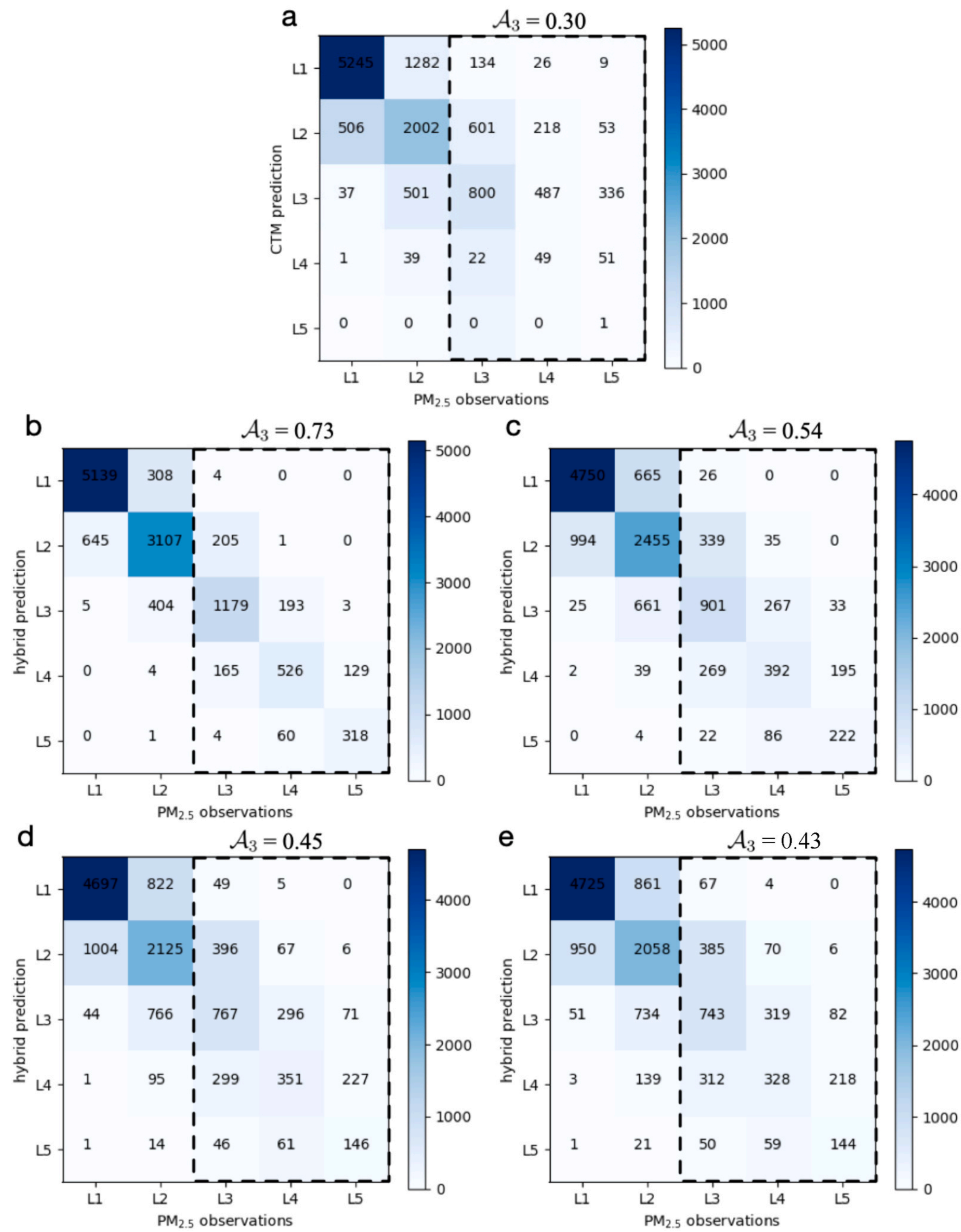

Fig. 7. Confusion matrix of CTM prediction (a); hybrid prediction 1 h in advance (b); hybrid prediction 3 h in advance (c); hybrid prediction 6 h in advance (d); hybrid prediction 9 h in advance (e). Note that no observation or simulation falls into the rank L6, therefore coordinate L6 is not listed.

Asia, or new type of measurements (such as satellite data) will be performed.

Declaration of competing interest

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

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 11971302) and the Startup Foundation for Introducing Talent of NUIST.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2020.118022.

References

Abdi-Oskouei, M., Pfister, G., Flocke, F., Sobhani, N., Saide, P., Fried, A., Richter, D., Weibring, P., Walega, J., Carmichael, G., 2018. Impacts of physical parameterization on prediction of ethane concentrations for oil and gas emissions in wrf-chem. Atmos. Chem. Phys. 18 (23), 16863–16883. URL. https://acp.copernicus.org/articles/18/16863/2018/.

Antanasijevic, D.Z., Pocajt, V.V., Povrenovic, D.S., Ristic, M.D., Peric-Grujic, A.A., 2013. PM10 emission forecasting using artificial neural networks and genetic algorithm input variable optimization. Sci. Total Environ. 443, 511–519. https://doi.org/10.1016/j.scitotenv.2012.10.110. URL.

Bates, T.S., Anderson, T.L., Baynard, T., Bond, T., Boucher, O., Carmichael, G., Clarke, A., Erlick, C., Guo, H., Horowitz, L., Howell, S., Kulkarni, S., Maring, H., McComiskey, A., Middlebrook, A., Noone, K., O'Dowd, C.D., Ogren, J., Penner, J., Quinn, P.K., Ravishankara, A.R., Savoie, D.L., Schwartz, S.E., Shinozuka, Y., Tang, Y., Weber, R.J., Wu, Y., 2006. Aerosol direct radiative effects over the northwest atlantic, northwest pacific, and north indian oceans: estimates based on in-situ chemical and optical measurements and chemical transport modeling. Atmos. Chem. Phys. 6 (6), 1657–1732. URL. https://www.atmos-chem-phys.net/6/1657/2006/.

Bey, I., Jacob, D.J., Yantosca, R.M., Logan, J.A., Field, B.D., Fiore, A.M., Li, Q., Liu, H.Y., Mickley, L.J., Schultz, M.G., 2001. Global modeling of tropospheric chemistry with assimilated meteorology: model description and evaluation. J. Geophys. Res.: Atmosphere 106 (D19), 23073–23095. URL. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2001JD000807.

Biancofiore, F., Busilacchio, M., Verdecchia, M., Tomassetti, B., Aruffo, E., Bianco, S., Di Tommaso, S., Colangeli, C., Rosatelli, G., Di Carlo, P., 2017. Recursive neural network model for analysis and forecast of PM10 and PM2. 5. Atmos. Pollut. Res. 8 (4), 652–659. URL. http://www.sciencedirect.com/science/article/pii/S1309104216304056.

Brasseur, G.P., Xie, Y., Petersen, A.K., Bouarar, I., Flemming, J., Gauss, M., Jiang, F., Kouznetsov, R., Kranenburg, R., Mijling, B., Peuch, V.-H., Pommier, M., Segers, A., Sofiev, M., Timmermans, R., van der A, R., Walters, S., Xu, J., Zhou, G., 2019. Ensemble forecasts of air quality in eastern China – Part 1: model description and implementation of the MarcoPolo–Panda prediction system, version 1. Geosci. Model Dev. (GMD) 12 (1), 33–67. URL. https://www.geosci-model-dev.net/12/33/2019/.

Chen, D., Xu, T., Li, Y., Zhou, Y., Lang, J., Liu, X., Shi, H., et al., 2015. A hybrid approach to forecast air quality during high-pm concentration pollution period. Aerosol Air Qual. Res. 15 (4), 1325–1337. https://doi.org/10.4209/aaqr.2014.10.0253. URL.

Croft, B., Pierce, J.R., Martin, R.V., Hoose, C., Lohmann, U., 2012. Uncertainty associated with convective wet removal of entrained aerosols in a global climate model. Atmos. Chem. Phys. 12 (22), 10725–10748. URL. https://www.atmos-chem-phys.net/12/10725/2012/.

Evensen, G., May, 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. 99 (C5), 10143–10162. https://doi.org/10.1029/94jc00572. URL.

Fan, J., Li, Q., Hou, J., Feng, X., Karimian, H., Lin, S., 2017. A spatiotemporal prediction framework for air pollution based on deep RNN. ISPRS Ann. Photogram. Rem. Sens. Spatial Inf. Sci. 4, 15. https://doi.org/10.5194/isprs-annals-IV-4-W2-15-2017. URL.

Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., Wang, J., 2015. Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation. Atmos. Environ. 107, 118–128. https://doi.org/10.1016/j.atmosenv.2015.02.030. URL.

Gardner, M.W., Dorling, S., 1998. Artificial neural networks (the multilayer perceptron)–a review of applications in the atmospheric sciences. Atmos. Environ. 32 (14–15), 2627–2636. URL. http://www.sciencedirect.com/science/article/pii/S1352231097004470.

Gilliam, R.C., Hogrefe, C., Godowitch, J.M., Napelenok, S., Mathur, R., Rao, S.T., 2015. Impact of inherent meteorology uncertainty on air quality model predictions. J. Geophys. Res.: Atmosphere 120 (23), 12259–12280. URL. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2 015JD023674.

Haagen-Smit, A.J., 1952. Chemistry and physiology of los angelos smog. Ind. Eng. Chem. 44, 1342–1346. https://doi.org/10.1021/ie50510a045. URL.

Hey, T., Tansley, S., Tolle, K., 2009. The fourth paradigm: data-intensive scientific discovery. Microsoft Research. URL. https://www.microsoft.com/en-us/research/publication/fourt h-paradigm-data-intensive-scientific-discovery.

Hu, J., Li, X., Huang, L., Ying, Q., Zhang, Q., Zhao, B., Wang, S., Zhang, H., 2017. Ensemble prediction of air quality using the wrf/cmaq model system for health effect studies in China. Atmos. Chem. Phys. 17 (21), 13103–13118. URL. https://acp.copernicus.org/articles/17/13103/2017/.

Jin, J., Lin, H.X., Heemink, A., Segers, A., 2018. Spatially varying parameter estimation for dust emissions using reduced-tangent-linearization 4DVar. Atmos. Environ. 187, 358–373. URL. https://www.sciencedirect.com/science/article/pii/S135223 1018303704.

Jin, J., Lin, H.X., Segers, A., Xie, Y., Heemink, A., 2019a. Machine learning for observation bias correction with application to dust storm data assimilation. Atmos. Chem. Phys. 19 (15), 10009–10026. URL. https://www.atmos-chem-phys.net/19/10009/2019/.

Jin, J., Segers, A., Heemink, A., Yoshida, M., Han, W., Lin, H.-X., 2019b. Dust emission inversion using Himawari-8 AODs over East Asia: an extreme dust event in may 2017. J. Adv. Model. Earth Syst. 11 (2), 446–467. URL. https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018MS001491.

Kalnay, E., 2003. Atmospheric Modeling, Data Assimilation and Predictability. Cambridge university press. URL. http://www.cambridge.org/9780521791793.

Kincaid, D., Kincaid, D.R., Cheney, E.W., 2009. Numerical analysis: mathematics of scientific computing. American Mathematical Soc. URL. http://www.ams.org/bookpages/amstext-2.

Li, G., Bei, N., Cao, J., Wu, J., Long, X., Feng, T., Dai, W., Liu, S., Zhang, Q., Tie, X., 2017a. Widespread and persistent ozone pollution in eastern China during thenon-winter season of 2015: observations and source attributions. Atmos. Chem. Phys. 17, 2759–2774. URL. http://www.atmos-chem-phys.net/17/2759/2017/.

Li, X., Peng, L., Hu, Y., Shao, J., Chi, T., 2016. Deep learning architecture for air quality predictions. Environ. Sci. Pollut. Control Ser. 23 (22), 22408–22417. https://doi.org/10.1007/s11356-016-7812-9. URL.

Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C., Chi, T., 2017b. Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation. Environ. Pollut. 231, 997–1004. https://doi.org/10.1016/j.envpol.2017.08.114. URL.

Lin, H.X., Jin, J., van den Herik, J., 2019. Air quality forecast through integrated data assimilation and machine learning. 11th Conference on Agents and Artificial Intelligence, pp. 787–793. URL. https://www.scitepress.org/Papers/2019/75552.

Lorenc, A.C., Bowler, N.E., Clayton, A.M., Pring, S.R., Fairbairn, D., 2015. Comparison of hybrid-4denvar and hybrid-4dvar data assimilation methods for global nwp. Mon. Weather Rev. 143 (1), 212–229. https://doi.org/10.1175/MWR-D-14-00195.1. URL.

Manders, A.M.M., Builtjes, P.J.H., Curier, L., Denier van der Gon, H.A.C., Hendriks, C., Jonkers, S., Kranenburg, R., Kuenen, J., Segers, A.J., Timmermans, R.M.A., Visschedijk, A., Wichink Kruit, R.J., Van Pul, W.A.J., Sauter, F.J., van der Swaluw, E., Swart, D.P.J., Douros, J., Eskes, H., van Meijgaard, E., van Ulft, B., van Velthoven, P., Banzhaf, S., Mues, A., Stern, R., Fu, G., Lu, S., Heemink, A., van Velzen, N., Schaap, M., 2017. Curriculum vitae of the LOTOS-EUROS (v2.0) chemistry transport model. Geosci. Model Dev. (GMD) 10 (11), 4145–4173. URL. http://www.geosci-model-dev-discuss.net/gmd-2017-88/.

Osowski, S., Garanty, K., 2007. Forecasting of the daily meteorological pollution using wavelets and support vector machine. Eng. Appl. Artif. Intell. 20 (6), 745–755. URL. http://www.sciencedirect.com/science/article/pii/S0952197 606001904.

Reddy, V., Yedavalli, P., Mohanty, S., Nakhat, U., 2018. Deep air: forecasting air pollution in Beijing, China. URL. https://www.ischool.berkeley.edu/projects/2017/deep-air.

Reynolds, S.D., Roth, P.M., Seinfeld, J.H., 1973. Mathematical modeling of photochemical air pollutionâ€"i: formulation of the model. Atmos. Environ. 7 (11), 1033–1061. URL. http://www.sciencedirect.com/science/article/pii/000469817390214X.

Schutgens, N., Nakata, M., Nakajima, T., 2012. Estimating aerosol emissions by assimilating remote sensing observations into a global transport model. Rem. Sens. 4, 3528–3543. https://doi.org/10.3390/rs4113528. URL.

Schutgens, N.A.J., Gryspeerdt, E., Weigum, N., Tsyro, S., Goto, D., Schulz, M., Stier, P., 2016. Will a perfect model agree with perfect observations? the impact of spatial sampling. Atmos. Chem. Phys. 16 (10), 6335–6353. URL. https://acp.copernicus.org/articles/16/6335/2016/.

Sekiyama, T.T., Tanaka, T.Y., Shimizu, A., Miyoshi, T., 2010. Data assimilation of CALIPSO aerosol observations. Atmos. Chem. Phys. 10 (1), 39–49. https://doi.org/10.5194/acp-10-39-2010. URL.

Simpson, D., Benedictow, A., Berge, H., Bergstrom, R., Emberson, L.D., Fagerli, H., Flechard, C.R., Hayman, G.D., Gauss, M., Jonson, J.E., Jenkin, M.E., Nyiri, A., Richter, C., Semeena, V.S., Tsyro, S., Tuovinen, J.-P., Valdebenito, A., Wind, P., 2012. The emep msc-w chemical transport model – technical description. Atmos. Chem. Phys. 12 (16), 7825–7865. URL. https://www.atmos-chem-phys.net/12/7825/2012/.

Solazzo, E., Bianconi, R., Hogrefe, C., Curci, G., Tuccella, P., Alyuz, U., Balzarini, A., Baró, R., Bellasio, R., Bieser, J., Brandt, J., Christensen, J.H., Colette, A., Francis, X., Fraser, A., Vivanco, M.G., Jiménez-Guerrero, P., Im, U., Manders, A., Nopmongcol, U., Kitwiroon, N., Pirovano, G., Pozzoli, L., Prank, M., Sokhi, R.S., Unal, A., Yarwood, G., Galmarini, S., 2017. Evaluation and error apportionment of an ensemble of atmospheric chemistry transport modeling systems: multivariable temporal and spatial breakdown. Atmos. Chem. Phys. 17 (4), 3001–3054. URL. https://acp.copernicus.org/articles/17/3001/2017/.

Song, J., Zhao, C., Lin, T., Li, X., Prishchepov, A.V., 2019. Spatio-temporal patterns of traffic-related air pollutant emissions in different urban functional zones estimated by real-time video and deep learning technique. J. Clean. Prod. 238, 117881. URL. http://www.sciencedirect.com/science/article/pii/S0959652 619327519.

Sun, W., Zhang, H., Palazoglu, A., Singh, A., Zhang, W., Liu, S., 2013. Prediction of 24-hour-average pm2. 5 concentrations using a hidden markov model with different emission distributions in northern California. Sci. Total Environ. 443, 93–103. URL. http://www.sciencedirect.com/science/article/pii/S0048969 712013617.

Talagrand, O., 1997. Assimilation of observations, an introduction (gtspecial issueltdata assimilation in meteology and oceanography: theory and practice). J. Meteorol. Soc. Japan. Ser. II 75 (1B), 191–209.

Timmermans, R., Kranenburg, R., Manders, A., Hendriks, C., Segers, A., Dammers, E., Zhang, Q., Wang, L., Liu, Z., Zeng, L., Denier van der Gon, H., Schaap, M., 2017. Source apportionment of PM2.5 across China using LOTOS-EUROS. Atmos. Environ. 164, 370–386. https://doi.org/10.1016/j.atmosenv.2017.06.003. URL.

van der Graaf, S.C., Dammers, E., Schaap, M., Erisman, J.W., 2018. Technical note: how are NH3 dry deposition estimates affected by combining the LOTOS-EUROS model with IASI-NH3 satellite observations? Atmos. Chem. Phys. 18 (17), 13173–13196. URL. https://acp.copernicus.org/articles/18/13173/2018/.

Wang, T., Jiang, F., Deng, J., Shen, Y., Fu, Q., Wang, Q., Fu, Y., Xu, J., Zhang, D., 2012. Urban air quality and regional haze weather forecast for yangtze river delta region. Atmos. Environ. 58, 70–83 air Quality Modeling in Asia (AQMA) 2011. URL. http://www.sciencedirect.com/science/article/pii/S135223 1012000349.

World Health Organization, 2016. Ambient air pollution: a global assessment of exposure and burden of disease. Tech. rep. https://apps.who.int/iris/handle/10665/250141.

Xie, J., Wang, X., Liu, Y., Bai, Y., 2018. Autoencoder-based deep belief regression network for air particulate matter concentration forecasting. J. Intell. Fuzzy Syst. 34 (6), 3475–3486. URL. https://content.iospress.com/articles/journal-of-intellig ent-and-fuzzy-systems/ifs169527.

Xie, Y., 2018. Deep Learning Architectures for pm2.5 and Visibility Predictions. Master's thesis. Delft University of Technology, Delft, the Netherlands. URL. http://resolver.tudelft.nl/uuid:c7ea9e3e-d6c2-426c-b4d5-8 c7143052257.

Yang, R., Nan, Z., Feng, Y., 2016. A novel approach based on an improved random forest to forecasting the air quality of second-hand housing. 2016 9th International Symposium on Computational Intelligence and Design (ISCID). URL. https://ieeexplore.ieee.org/abstract/document/7830343.

Zhang, S., 2012. Nearest neighbor selection for iteratively kNN imputation. J. Syst. Software 85, 2541–2552. URL. https://www.sciencedirect.com/science/article/pii/S016412 1212001586?via%3Dihub.

Zhao, J., Deng, F., Cai, Y., Chen, J., 2019. Long short-term memory - fully connected (LSTM-FC) neural network for PM2.5 concentration prediction. Chemosphere 220, 486–492. URL. http://www.sciencedirect.com/science/article/pii/S0045653 518324639.

Zhou, C.-H., Gong, S., Zhang, X.-Y., Liu, H.-L., Xue, M., Cao, G.-L., An, X.-Q., Che, H.-Z., Zhang, Y.-M., Niu, T., 2012. Towards the improvements of simulating the chemical and optical properties of Chinese aerosols using an online coupled model – CUACE/Aero. Tellus B 64 (1), 18965. https://doi.org/10.3402/tellusb.v64i0.18965. URL.