



Machine learning for observation bias correction with application to dust storm data assimilation

Jianbing Jin¹, Hai Xiang Lin¹, Arjo Segers², Yu Xie¹, and Arnold Heemink¹

¹Delft Institute of Applied Mathematics, Delft University of Technology, Delft, the Netherlands

²TNO, Department of Climate, Air and Sustainability, Utrecht, The Netherlands

Correspondence: Jianbing Jin (J.Jin-2@tudelft.nl)

Abstract. Data assimilation algorithms rely on a basic assumption of an unbiased observation error. However, the presence of inconsistent measurements with nontrivial biases or inseparable baselines is unavoidable in practice. Assimilation analysis might diverge from reality, since the data assimilation itself cannot distinguish whether the differences between model simulations and observations are due to the biased observations or model deficiencies. Unfortunately, modeling of observation biases or baselines which show strong spatiotemporal variability is a challenging task. In this study, we report how data-driven machine learning can be used to perform observation bias correction for data assimilation through a real application, which is the dust emission inversion using PM₁₀ observations.

PM₁₀ observations are considered as unbiased, however, a bias correction is necessary if they are used as a proxy for dust during dust storms since they actually represent a sum of dust particles and *non-dust* aerosols. Two observation bias correction methods have been designed in order to use PM₁₀ measurements as proxy for the dust storm loads under severe dust conditions. The first one is the conventional chemical transport (CTM) model that simulates life cycles of *non-dust* aerosols. The other one is the machine learning model that describes the relations between the regular PM₁₀ and other air quality measurement. The latter is trained by learning two-year's historical samples. The machine learning based *non-dust* model is shown to be in better agreements with observations compared to the CTM. The dust emission inversion tests have been performed, either through assimilating the raw measurements, or the bias-corrected dust observations using either the CTM or machine learning model. The emission field, surface dust concentration and forecast skill are evaluated. The worst case is when we directly assimilate the original observations. The forecasts driven by the posterior emission in this case even results in larger errors than the reference prediction. This shows the necessities of bias correction in data assimilation. The best results are obtained when using a machine learning model for bias correction, with the existing measurements used more precisely and the resulting forecasts close to reality.

1 Introduction

For centuries, East Asia experienced regular dust storms in the spring time. Those dust events mainly originated from the dust source regions of the Gobi and Taklamakan deserts. Annually, thousands tons of "yellow sands" are blown



eastward over the densely populated areas in China, the Korean peninsula, and Japan by the prevailing winds. Dust storms can also carry irritating spores, bacteria, viruses and persistent organic pollutants (WMO, 2017). Next to the human health, the resulting low visibility can cause a severe disruption of transportation systems. For example, more than 1,100 flights have been delayed/canceled in Beijing after the city was struck by a choking dust storm in early May 2017.

A large number of dust simulations models has been developed over the past decades (Wang et al., 2000; Gong et al., 2003; Liu et al., 2003). These chemical transport models help to understand the life cycles of the dust storms, and are also used for dust forecasts and to aid early warning systems. Apart from advances in simulation of dust storms, progresses have also been made in the monitoring of dust or general aerosol loads. Field station networks are constructed to observe the in-situ particulate matter (PM) levels over densely populated regions (Li et al., 2017a). Ground-based sun photometers, e.g., the global Aerosol Robotic Network (AERONET) (Cesnulyte et al., 2014), are widely used to monitor column-integrated aerosol profiles. Satellite onboard instruments such as Moderate Resolution Imaging Spectroradiometer (MODIS) (Remer et al., 2005), Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) (Sekiya et al., 2010), and Advanced Himawari Imager/Himawari-8 (Yoshida et al., 2018) provide the measurements of airborne particles with further wide coverages. These measurements could be used to calibrate the parametrizations in the dust simulation models and to evaluate their ability to forecast dust concentrations. Moreover, the observations could be combined with a dust modeling system through data assimilation to improve the forecast skills.

A wide variety of data assimilation techniques has been used with dust simulation models, including variational methods (Yumimoto et al., 2008; Niu et al., 2008; Gong and Zhang, 2008; Jin et al., 2018) and ensemble-based sequential methods (Lin et al., 2008; Sekiya et al., 2010; Khade et al., 2013; Di Tomaso et al., 2017). In these systems, the available observations are either used to estimate the model states (dust concentrations) or to reduce uncertainties in the emissions and/or other model parameters. Challenges for dust assimilations include development of more and more accurate dust simulations, and use of new types of observations including vertical profiles from Lidars and latest satellite observations. A further challenge for any assimilation system is the proper definition of the observation and representation errors, as well as characterization of biases.

In general, the commonly used data assimilation schemes all rely on the basic assumption of an unbiased observation. In real applications, however, measurement biases are often unavoidable. In presence of biases, it is impossible to determine whether a difference between an *a priori* simulation and an observation are due to the biased observations or model deficiencies. The biases might lead to assimilations that diverge from reality (Lorente-Plazas and Hacker, 2017). A well known example of observation biases is in radiance observation assimilation systems in presence of clouds (Eyre, 2016; Berry and Harlim, 2017). To avoid problems with these biases, up to 99% of cloudy observed measurements are discarded although they may also contain valuable information. If dust storms are coincident with clouds, it is also possible that in satellite retrieval algorithms clouds are mistaken for dust, leading to strong biases in the data to be assimilated (Jin et al., 2019).



Another example where observation biases are important is when ground-based PM_{10} measurements are assimilated in dust simulation models. Due to the high temporal resolutions and the rather dense observation network, the ground-based air quality observing network has become a powerful source of measurements on dust aerosols. The records, mainly the PM_{10} feature, were widely used to calibrate, assess or estimate the dust model (Lin et al., 2008; Wang et al., 2008; Huneus et al., 2011; Yumimoto et al., 2016; Benedetti et al., 2018). However, the observed PM_{10} concentrations do not only consist of dusts, but are actually the sum of the dust and other regular particles. The latter one are emitted not only from anthropogenic activities such as industries, vehicles, and households, but also from natural sources such as wild fires and sea spray. In this paper we will simply refer to these particles as the *non-dust* fraction of the total PM_{10} . The concentrations of *non-dust* aerosols in urbanized areas could be substantial, reaching values up to $500 \mu\text{g}/\text{m}^3$ (Shao et al., 2018).

Although PM_{10} observations include a nontrivial bias, the wide spread availability makes them still useful in dust storm assimilation system. During dust storm events, extreme high peaks of more than $1000\text{--}2000 \mu\text{g}/\text{m}^3$ PM_{10} are recorded which can be attributed mainly to dust. If these would be assimilated directly in dust simulation model, ignoring the fact that at least some part represents *non-dust*, the assimilation system would diverge to states that overestimate the dust load. In case of less severe dust events, the dust analysis divergence would then become extremely critical.

However, modeling of observation biases is very challenging when they have strong spatial and temporal variabilities. Little progress has been made in bias correction of full-aerosol measurements for their use in dust storm data assimilation. Lin et al. (2008) selected only PM_{10} observations for assimilation when at least one occurrence of dust clouds was reported by the local stations. In Jin et al. (2018), it was found that on sites with both PM_{10} and $\text{PM}_{2.5}$ observations, only the PM_{10} concentration increased during a dust episode, while the $\text{PM}_{2.5}$ concentrations were not affected and remained at a constant level. Besides, Xu et al. (2017) and Jin et al. (2018) suggested a strong correlation between $\text{PM}_{2.5}$ and *non-dust* PM_{10} . Therefore, a very simple *non-dust* PM_{10} baseline removal (called observation bias correction) was proposed, in which the available $\text{PM}_{2.5}$ was used to approximate the *non-dust* PM_{10} (or baseline) during a dust event by:

$$\text{PM}_{10}^{\text{non-dust}} = b + r \times \text{PM}_{2.5} \quad (1)$$

where the b and $r > 1$ are linear regression parameters based on a 24-hour history of measurements before arrival of the dust storm.

The aforementioned methods either exclude a selection of the measurements, which may still contain useful information, or work under ideal circumstance only when a simple correlation between PM_{10} and $\text{PM}_{2.5}$ is validated. To fully exploit the dust information present in total PM observations, a more advanced method is needed. In this paper we proposed two methods, either using a conventional chemistry transport model, or a machine learning model.



A chemistry transport model (CTM) implements all available knowledge on emission, transport, deposition, and other physical processes in order to simulate concentrations of trace gases and, important here, aerosols. Daily air quality forecasts are often provided using such CTMs. A simulation model for dust storm events is usually just a CTM with all tracers removed except dust; by using the full CTM, an estimate of the *non-dust* part of the aerosol load could be made. In this study, the LOTOS-EUROS CTM is used to simulate the dust as well as the *non-dust* aerosol concentrations. If the non-dust model was perfect, the difference between simulation and observed PM₁₀ would be unbiased, and assimilation could be applied on the combined dust and *non-dust* concentrations. In case of a dust storm event, it remains necessary to distinguish between the dust and *non-dust* part of the simulations since the two parts will have very different error characteristics. The dust part is quickly varying and has a large uncertainty, while the *non-dust* part is more smooth but very persistent in time and has a relatively small uncertainty. An assimilation system on the combined simulations should be able to handle these differences. However, the error attribution to their proper sources (dust and *non-dust* error) then becomes extremely critical as explained in Section 2.4. Since this paper focuses on dust during a severe event only, we will not explore the error characteristics of the *non-dust* part of the model. Therefore we will not apply an assimilation on the combined aerosol (dust and *non-dust*) model. Instead, the *non-dust* simulations will solely be used to remove the *non-dust* baseline from PM₁₀ observations.

Similar to the air quality forecast, the accuracy of a CTM for *non-dust* aerosols is hampered by lack of accurate input data. For example, the timely update of anthropogenic emission inventories is always a key issue for air quality forecasts. With the ever-increasing complexity and resolution, the CTMs are now becoming highly nonlinear and time-consuming. However, they may still not be able to identify explicit representations of the non-dust aerosol dynamics, especially regarding fine-scale processes.

In addition to the conventional CTM, we propose a new method for removing the *non-dust* part of the PM₁₀ observations which is based on machine learning (ML). Data-driven methods have already been proved to be a powerful tool to provide air quality forecasts for horizons of a few days, (e.g., Li et al. (2016); Fan et al. (2017); Li et al. (2017b); Chen et al. (2018)). Different from the chemical transport models which simulate the aerosol physical processes, machine learning models describe mathematical relations of input-output and trained by learning a large number of samples from historical records. Our machine learning system used a neural network, namely *long short term memory* (LSTM). The input is formed by air quality indices for a number of relevant tracers (PM_{2.5}, SO₂, NO₂, CO, and O₃), as well as meteorology data. The output of the system is an estimate of the *non-dust* PM₁₀ concentration. The input features are to large extent independent of the dust storms, even the PM_{2.5} concentrations as shown in Jin et al. (2018); observations of PM₁₀ are excluded since excessive dust loads are visible mainly in this component. Recent development and the availability of open source machine learning tools provide a good opportunity to estimate the air quality indices using a data-driven machine learning models.

Whereas these are previous studies on dust storm data assimilation using various kinds of combined aerosol measurements, we are the first to investigate the necessities of bias correction for these full-aerosol observations in order to use them as 'real' dust measurements in a dust storm assimilation system. The adding values of observation



bias correction in dust emission inversion is explored through the ground-based PM_{10} measurement assimilation. It can be easily applied on others general applications, e.g., remote sensing data assimilation. Our contributions are threefold. Firstly, we present and examine the conventional CTM for removing the *non-dust* part from PM_{10} observations. Secondly, we design and examine a novel machine learning based bias correction which is data-driven and free of the time-consuming numerical CTMs. Thirdly, we evaluate the two *non-dust* aerosol model simulations by comparing to the PM_{10} measurements during regular periods (rare dust events involved); we evaluate dust emission fields, surface dust concentration simulation and forecast skills which are obtained by either assimilating the raw PM_{10} data, or bias-corrected measurements either using the CTM or machine learning model.

The paper is organized as follows. A brief description of our dust simulation model (LOTOS-EUROS/Dust) and the four dimensional variational data assimilation method for emission inversion are presented in Section 2. The biased observation representing error and its influence on the assimilation system are also explained. The two bias correction methods, the *non-dust* aerosol regional chemical transport model and a machine learning model, are discussed and the bias simulation is evaluated in Section 3. Section 4 reports the assimilation results using the two bias correction methods, and evaluates the forecast skills using independent measurements. Section 5 discusses the necessities of observation bias correction in assimilation works, highlights our key contributions, and proposes possible further improvements.

2 Dust storm data assimilation system

2.1 Dust model

The dust storm event studied in this paper took place in East Asia in April 2015, and has already been used as a test case for assimilation experiments in Jin et al. (2018). The LOTOS-EUROS/Dust simulation model is used with similar configurations to our previous studies, which is configured on a domain from 15°N to 50°N and 70°E to 140°E , but with a higher model resolution of 0.25° . The model is driven by European Center for Medium-Ranged Weather Forecast (ECMWF) operational forecasts for horizons of 3-12 hours. The dust load is described by 5 aerosol bins within a diameter range $0.01\text{ }\mu\text{m} < D_p < 10\text{ }\mu\text{m}$. Physical processes included are emission, advection, diffusion, dry and wet deposition, and sedimentation. The dust emission scheme implemented in LOTOS-EUROS is mainly based on the formulation of horizontal saltation flux (Marticorena and Bergametti, 1995) and sandblasting efficiency (Shao et al., 1996). A terrain preference parameter F_{ps} was used in the dust emission in Jin et al. (2018). This geographic dependent parameter was first introduced by Ginoux et al. (2001), and used to approximate the probability of having accumulated sediments that can be resuspended. In this work, F_{ps} is disabled since the preference factor was found to limit the emission rate in some regions where the fine-scale topographic feature is actually unknown. Snapshots of a reference simulation of the dust episode has been performed and is shown in Fig.8(a).



2.2 Observation network

The China Ministry of Environmental Protection (MEP) has commenced to release the hourly-average measurements of atmospheric constituents including PM_{2.5}, PM₁₀, CO, O₃, NO₂ and SO₂ since 2013. A huge number of ground stations measuring these air quality indices have been established in densely populated areas. Up to now, the monitoring network grows up to 1,500 field stations covering all over China as shown in Fig.1.

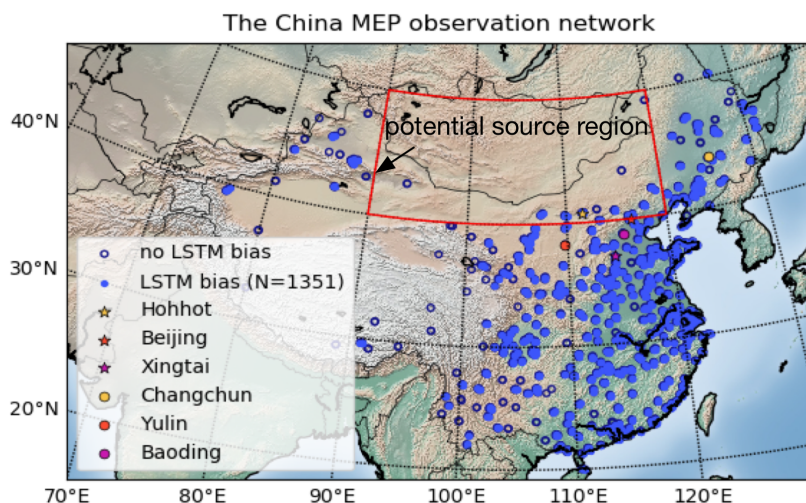


Figure 1. The China MEP air quality monitoring network and the potential dust storm source region. LSTM base *non-dust* PM₁₀ forecast are performed only in stations of blue dot, while ones of black circles are skipped.

2.3 Reduced tangent linearization 4DVar

The assimilation system, which will be used to combine bias-corrected PM₁₀ observations with simulations, is based on a reduced-tangent-linearization four dimensional variational (4DVar) data assimilation. The goal of a 4DVar technique is to find the maximum likelihood estimation of a state vector, which is here the dust emission field \mathbf{f} , given the available observations over a time window. A common approach is to use an incremental formulation, which aims to find the optimal emission deviation $\delta\mathbf{f}$ as the minimum of the cost function:

$$J(\delta\mathbf{f}) = \frac{1}{2} \delta\mathbf{f} \mathbf{B}^{-1} \delta\mathbf{f} + \frac{1}{2} \sum_{i=1}^k (\mathbf{H}_i \mathbf{M}_i \delta\mathbf{f} + \mathbf{d}_i)^T \mathbf{R}_i^{-1} (\mathbf{H}_i \mathbf{M}_i \delta\mathbf{f} + \mathbf{d}_i) \quad (2)$$

where k is the number of time steps within the assimilation window. The vector $\delta\mathbf{f}$ denotes a perturbation of the emissions with respect to the background one. For an observation time i , the innovation vector (length m_i) is defined as the difference between the simulations and observations:

$$\mathbf{d}_i = \mathcal{H}_i(\mathcal{M}_i(\mathbf{f})) - \mathbf{y}_i \quad (3)$$



where \mathcal{M}_i denotes the LOTOS-EUROS/Dust transport model, \mathcal{H}_i is the operator that converts state variables into observation space, and \mathbf{y}_i is the vector with dust observations at this time step i . The operators \mathbf{H}_i and \mathbf{M}_i denote linearizations of \mathcal{H}_i and \mathcal{M}_i around the reference emission vector \mathbf{f}_b . Following Jin et al. (2018), the errors in dust emission field were assumed to be only due to the uncertainty in the friction velocity threshold, and similar assumptions on the uncertainty are used to build an emission error covariance \mathbf{B} . The observation error term is weighted by an observation error covariance \mathbf{R} , for which the individual elements will be described in Section 4.1.

To release the efforts in updating the full *tangent linear model* \mathbf{M}_i , a reduced-tangent-linearized 4DVar (Jin et al., 2018, 2019) is used. The simplified method is based on proper orthogonal decomposition (POD) of the background covariance \mathbf{B} which efficiently carries out model reduction by identifying the few most energetic modes:

$$\mathbf{B} = \mathbf{U}\mathbf{U}^T \approx \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T \quad (4)$$

$$\delta\mathbf{f} \approx \tilde{\mathbf{U}} \delta\mathbf{w}$$

where $\mathbf{U} \in \mathbf{R}^{P \times P}$ is the background emission covariance square root, with P the size of the emission field of $O(10^4)$ elements. while $\tilde{\mathbf{U}} \in \mathbf{R}^{P \times p}$ is the truncation of \mathbf{U} based on POD, with p the reduced rank size of $O(10^2)$. The vector $\delta\mathbf{w} \in \mathbf{R}^p$ stores the transformed control variables.

The cost function of the reduced-tangent-linearization 4DVar is formulated as:

$$J(\delta\mathbf{w}) = \frac{1}{2} \delta\mathbf{w}^T \delta\mathbf{w} + \frac{1}{2} \sum_{i=1}^k (\mathbf{H}_i \tilde{\mathbf{M}}_i \tilde{\mathbf{U}} \delta\mathbf{w} + \mathbf{d}_i)^T \mathbf{R}_i^{-1} (\mathbf{H}_i \tilde{\mathbf{M}}_i \tilde{\mathbf{U}} \delta\mathbf{w} + \mathbf{d}_i) \quad (5)$$

where $\tilde{\mathbf{M}}_i$ denotes the reduced tangent linear model with a rank p , which is approximated using the perturbation method. More details about the reduced-tangent-linearization 4DVar algorithm can be found in Jin et al. (2018).

2.4 Biased observation representing error

In real applications, the observations inevitably have biases which cannot be attributed to the model simulation, as following:

$$\mathbf{y}_i = \mathcal{H}_i(\mathcal{M}_i(\mathbf{f})) + \mathbf{b}_i + \boldsymbol{\sigma}_i \quad (6)$$

where $\boldsymbol{\sigma}_i$ is the vector of Gaussian distributed observation errors which have zero means and a known covariance matrix \mathbf{R}_i , and \mathbf{b}_i denotes the vector of observation bias. In our application, the vector \mathbf{y}_i contains the observed PM₁₀ concentrations, while the aerosols released in the local anthropogenic activities and other *non-dust* related processes are referred as \mathbf{b}_i .

In the course of data assimilation, it is impossible to determine whether the departures (\mathbf{d}_i) of the prior simulations from the observations are due to the biased observations \mathbf{b}_i or emission errors $\delta\mathbf{f}$. Thus, the assimilation result will diverge from the true state when a bias is present. In complex dynamic models as the atmospheric transport model, the biases (*non-dust* aerosols) could have high spatial and temporal variabilities and is therefore difficult to quantify.



In this work, we proposed two methods to quantify the bias levels for the observation bias correction. The first one is the *non-dust* parts of LOTOS-EUROS chemical transport model (CTM) which simulates the aerosol life cycles including emission, transport and deposition. The second method is to describe the *non-dust* aerosol levels using a data-driven machine machine model. Details of these two methods are illustrated in Section 3.

5 In fact, both LOTOS-EUROS CTM and machine learning model are imperfect, and some biases might still exist after the correction. The former one is known to be limited by errors in the emission inventories, meteorological forecasts and all kinds of input sources. The latter is then hampered by the deficiency of the type model (e.g., insufficient to represent the complexity of the phenomenon), inadequate amount of training data. However, by combining the bias-corrected observation with the dust model, the assimilation will adapt to posteriors which are
 10 more close to reality.

There were a few studies that addressed both the model deficiency and uncertainty in observation bias simultaneously using either variational data assimilation (Dee and Uppala, 2009) or sequential filters (Dee, 2005; Lorente-Plazas and Hacker, 2017). Those assimilation schemes not only require a formulation of a model for the bias, but also need a quality-assured reference to describe the uncertainty of the bias model. The need to attribute errors to their proper
 15 sources is obviously a key part in any assimilation systems, but becomes especially critical when it involves bias correction. This is because a wrong error attribution will force the assimilation to be consistent with a biased source. If the source of a known bias is uncertain, assimilation without considering the uncertainty of bias model is the safest option (Dee, 2005). Therefore, these two *non-dust* models are solely set as references for the bias, and the uncertainties are not explored here.

20 2.5 Assimilation Window

Fig.2 shows a time line for the assimilation experiment around the April 2015 dust event, which is very similar to what was used in Jin et al. (2018). The dust event has a short duration, and therefore only a single assimilation window with a length of 36 hours is used. The dust emissions take place at the start of the window, while the observations become available at the end of the window since they are located downwind from the source region (see
 25 Fig.1). A long assimilation window is therefore necessary in order to estimate the correct emission parameters given the observations.

When we perform the assimilation analysis at April 15, 19:00, only the dust observations from April 15, 08:00 to 19:00 will be assimilated and they are calculated by subtracting the *non-dust* part (CTM based or ML based) from the PM₁₀ observations. After the analysis, the simulation model is used to perform a dust forecast for the next 12
 30 hours using the newly-estimated emission parameters. A full-aerosol PM₁₀ forecast will then be calculated by adding the dust forecast and *non-dust* aerosol forecast, where the later again originates from either the CTM and machine learning model.

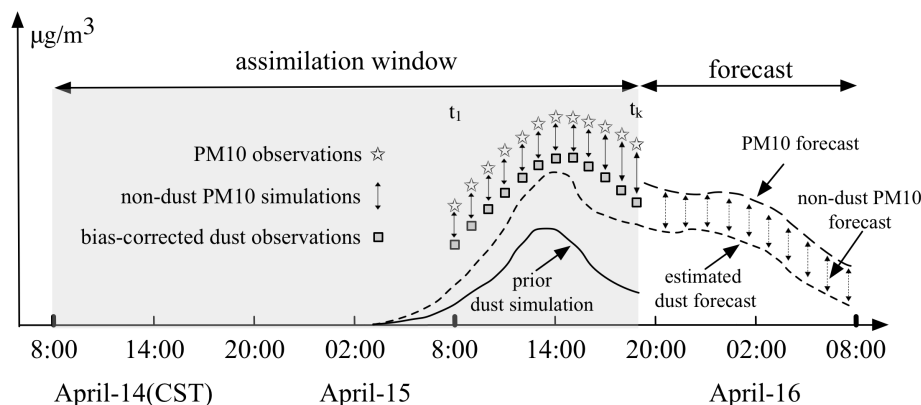


Figure 2. Timeline of observation availability, assimilation cycles and forecasts

3 Observation bias correction methods

Two systems are introduced to correct the *non-dust* bias when using PM_{10} observations in a dust assimilation. The first one is CTM LOTOS-EUROS/*non-dust* model that simulates the physical processes of the *non-dust* aerosols. The latter is the machine learning model that estimates the *non-dust* aerosol based on historical records. The following sections describe the two methods in more detail.

3.1 Chemistry transport model (LOTOS-EUROS/*non-dust*)

The regional CTM LOTOS-EUROS/*non-dust* is configured similar to the LOTOS-EUROS/Dust used in the assimilation, but now includes all trace gases and *non-dust* aerosols. The configuration is similar to what is used for daily air quality simulations over China as described in (Timmermans et al., 2017). Anthropogenic emissions are taken from the Multi-resolution Emission Inventory for China (MEIC) inventory (<http://www.meicmodel.org>). Natural emissions included are sea salts, biogenic emissions, and wild fires. The LOTOS-EUROS full aerosol operational forecast over this modeling domain are released via the MarcoPolo-Panda projects (www.marcopolo-panda.eu).

3.2 Machine learning for *non-dust* PM_{10} simulation

Given a set of training data, a machine learning algorithm attempts to find the relation between input and output. When a proper model is used, the machine learning algorithm can learn to reproduce the complex behaviors of a dynamic system. The description is purely based on the data, physical knowledge is not included. Machine learning algorithms are popular tools to forecast the air quality indices using the history records (Li et al., 2016; Fan et al., 2017; Chen et al., 2018; Lin et al., 2019). In this study, the machine learning algorithm used is the *long short term*



memory (LSTM) neural network, which has demonstrated its ability in predicting spatial and temporal varying time series problems (Li et al., 2017b).

The LSTM operator \mathcal{L} , which is configured with parameters θ , for predicting *non-dust* PM₁₀ can be described as:

$$\mathbf{b}^{t_0+t} = \mathcal{L}_{\theta}(\mathbf{x}^{t_0}, \mathbf{x}^{t_0-1}, \dots, \mathbf{x}^{t_0-m+1}) \quad (7)$$

where \mathbf{b}^{t_0+t} represents the predictor, which is in this study the *non-dust* PM₁₀ concentration forecast t hours in advance. The temporal correlation between the input features and the forecasts declines with the forecast step. In our system, the maximum forecast period t is 12 hours. The input vectors $\mathbf{x}^{t_0}, \mathbf{x}^{t_0-1}, \dots, \mathbf{x}^{t_0-m+1}$ are the observed data of the past m hours, which is set as 18 hours empirically. The input vectors consist of:

- observations of PM_{2.5}, SO₂, NO₂, O₃, and CO from the ground based air quality network described in Section 2.2;
- observations of PM_{2.5} at the nearby sites;
- local meteorological data (temperature and dew point at 2 m, wind speed at 10 m) which are taken from the LOTOS-EUROS model input and originate from the European Center for Medium-Ranged Weather Forecast (ECMWF).

The LSTM neural network parameters θ are determined by minimizing the objective function J_{θ} that represents the mean squared error of predictors \mathbf{b} with respect to the measured values \mathbf{y}^b :

$$J_{\theta} = \frac{1}{m} \sum_{i=1}^m (\mathbf{b}_i - \mathbf{y}_i^b)^2 \quad (8)$$

The LSTM model \mathcal{L} is trained using the records from January 2013 to March 2015. Since dust storms are very low frequent occurrence during the training period, all the PM₁₀ records \mathbf{y}^b are actually the *non-dust* PM₁₀ concentrations. The regression model \mathcal{L} is thus assumed to reflect only the relation between input features and the *non-dust* PM₁₀.

Note that including PM₁₀ observations in the series of input vectors will certainly improve the skill of the machine learning forecasts. However, the LSTM model would then lack the ability to discriminate between the dust and *non-dust* fractions in PM₁₀ during a dust event. Earlier studied showed that the input variables, including PM_{2.5}, are independent on the dust storm as illustrated in Jin et al. (2018).

The PM_{2.5} measurements at the nearby sites are considered as the essential input for eliminating errors caused by incorrect observations at the modeling site. In this study, a data instance will only be selected for training the LSTM model if there is at least one nearby site within a radius 0.8° (approx 80 km), and a maximum of 3 nearby sites will be randomly selected where observation stations are densely distributed. Learning in the presence of data missing is pervasive in machine learning, which could be due to human errors during data entry, incorrect sensor readings,



or software bugs in the data processing pipeline. In our study, data interpolations are performed using both a linear interpolation and a k-Nearest-Neighbor algorithm (Zhang, 2012) if the modeling site have a data missing rate of at most 30%.

Fig.1 presents the original field observation network established by the China Ministry of Environmental Protection (MEP) up to 2018, as well as the sites (N=1351) where LSTM based *non-dust* forecasts are performed. It is clear that the LSTM forecast cannot be performed in each monitoring site. A part of the sites is skipped due to the lack of nearby sites, the rest are caused by high data missing rate in the training period.

3.3 Evaluation of *non-dust* PM₁₀ bias corrections

Our two bias models, LOTOS-EUROS/*non-dust* and LSTM, could both be used for air quality forecast when there is no dust storm. The forecasts are expected to have a good performance when dust is not present, and to underestimate the PM₁₀ levels in case of dust storms.

Both the CTM LOTOS-EUROS and LSTM are tested to forecast *non-dust* PM₁₀ over April-May 2015. This period includes the 2 to 3 days dust event that is used as test case for the assimilation. Fig.3(a)~(c) show density plots comparing PM₁₀ observations with either LOTOS-EUROS/*non-dust* forecasts, or with LSTM forecast 1 hour and 12 hours in advance. The underestimation of PM₁₀ during the dust event is visible in the bottom right corners of these plots.

The CTM LOTOS-EUROS/*non-dust* in general underestimates the *non-dust* PM₁₀. This could be explained from the fact not all types of particulate matters, such as secondary organic aerosols, are included in the model, and some aerosol emissions are very difficult to estimate (e.g., wood burning by households).

The two LSTM forecasts show on average a good agreement with the observations. As expected, a smaller forecast period $t=1$ hour gives a better result than the forecast over 12 hours.

When we perform the assimilation analysis at April 15, 19:00, the short period of $t=1$ hour in advance forecast will be treated as the *non-dust* levels in the bias correction of the original PM₁₀ measurements. Subsequently, the bias-corrected data are used to estimate the dust emissions over the past 36-hour window. Obviously, one important aim of the assimilation is to make a better forecast, in this study, the forecast skills will be evaluated in the following 12 hours from April 15, 19:00. Besides, the forecast is assessed by comparing the combined PM₁₀ forecast to PM₁₀ observations. The LSTM forecast with $t=12$ hours in advance will be added to the dust storm forecast to build the combined aerosol forecast.

3.3.1 Spatial patterns at observation sites

To assess our two *non-dust* PM₁₀ models, Fig.4 shows the snapshots of the PM₁₀ measurements, LOTOS-EUROS/*non-dust* simulations, LSTM forecasts, and the corresponding bias-corrected dust observations at three timestamps: April 15 08:00, 19:00 and 22:00. These first two moments are the start and end of the observation interval in the assimilation window (only observations from the last 12 hours of the assimilation window are assimilated as shown in Fig.2),

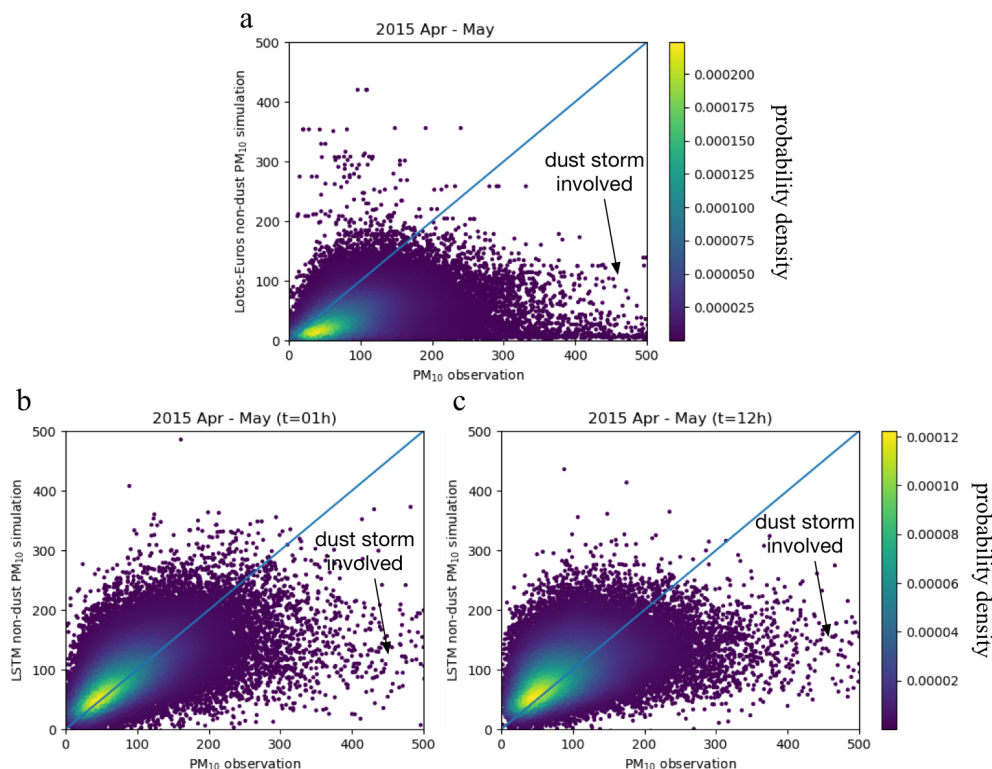


Figure 3. *Non-dust* PM_{10} simulation evaluations. (a): LOTOS-EUROS/*non-dust* forecast vs. PM_{10} measurements; (b): LSTM forecast 1 hour in advance vs. PM_{10} measurements; (c): LSTM forecast 12 hour in advance vs. PM_{10} measurements; (NOTE: the plots show 5% random samples over the period of April to May and is adapted from Fig.3 in Lin et al. (2019))

and observations at 22:00 is treated as independent data for cross-validation. At 08:00, actually only few stations close to the dust source area have already observed the dust storm. Some of the sites in central China observed high PM_{10} concentrations which are believed to be caused by presence of *non-dust* aerosols. Nearly all the stations in north China reported this dust storm at 19:00 and 22:00, as a band covering central and northeast China, see Fig.4 (a.2)~(a.3). Fig.4 (b.1)~(b.3) shows that the LOTOS-EUROS/*non-dust* model forecasts quite stable and constant *non-dust* PM_{10} levels, most of the simulated values are less than $100 \mu g/m^3$. Subsequently, the corresponding bias-corrected dust measurements (see Fig.4 (c.1)~(c.3)) are very similar to the original PM_{10} observations. This could be problematic when trying to measure the dust storm from the PM_{10} observations; for instance at 08:00 in Fig.4 (c.1), according to the bias-corrected observations the dust storm seems to have already reached central China which was in reality not the case. In comparison, the LSTM based bias-corrected dust observations (see Fig.4 (e.1)~(e.3)), which is calculated by subtracting the LSTM *non-dust* part (see Fig.4(d.1)~(d.3)) from the raw PM_{10} measurements, are close to our expectations. Only for sites that are very close to the source regions high dust concentrations are derived at 08:00, while for the other sites hardly any dust is derived. At 19:00, thus 11 hours later, at half of the



stations in the north of the domain high dust concentrations are derived. In the southeast of the domain, the derived dust concentrations remain almost zero since the dust plume did not arrive there yet. At 22:00, the plume is moved further south, and the dust load closer to the source region started to decrease.

3.3.2 Time series

- 5 To further evaluate the two bias correction methods, Fig.5 shows the time series at the following selected sites: Hohhot, Changchun, Beijing, Baoding, Xingtai and Yulin. The location of these sites can be found in Fig.1. These sites were selected because they illustrate the general performance of the LOTOS-EUROS/*non-dust* and LSTM methods.

The LOTOS-EUROS grid cells with the selected sites all include other observation sites as well, and to illustrate the spread in the observations the maximum and minimum observed values in the grid cell are added to the time series too. Similarly, the LSTM *non-dust* PM₁₀ simulation is given together with the spread within the grid cell.

Before the dust storm arrived at these sites, the LSTM model reproduces the variations in PM₁₀ rather well. Some errors are present, for example as can be seen on April 14 from 12:00 to 23:00 in Yulin. After the arrival of the dust storm, the PM₁₀ observations strongly increase, while the LSTM *non-dust* fraction remains at a low level since it is independent of the dust storm. The real dust measurement is then calculated by subtracting the *non-dust* part from the raw PM₁₀ observations.

The LOTOS-EUROS/*non-dust* simulations underestimate the *non-dust* PM₁₀ at all the six locations. Thus, the derived bias-corrected dust observations overestimate the actual dust load, and this will affect the dust assimilation results.

4 Data assimilation experiments

Three different sets of observations are now available for assimilation in the dust model: the original PM₁₀ observations, the PM₁₀ observations with LOTOS-EUROS bias correction, and the PM₁₀ observations with machine learning bias correction. The results have been compared in terms of the posterior dust emission fields and surface dust concentrations.

- 25 A practical use of assimilated concentrations is to use them as a start point for a forecast. This could be used to provide early information about the arrival of the dust plume and the expected dust level. The dust forecast after the end of the assimilation window at April 15 19:00 uses the newly estimated emissions. Apart from the dust concentrations, the forecast will also be evaluated in terms of skill scores for the total PM₁₀ concentrations in Section 4.3.

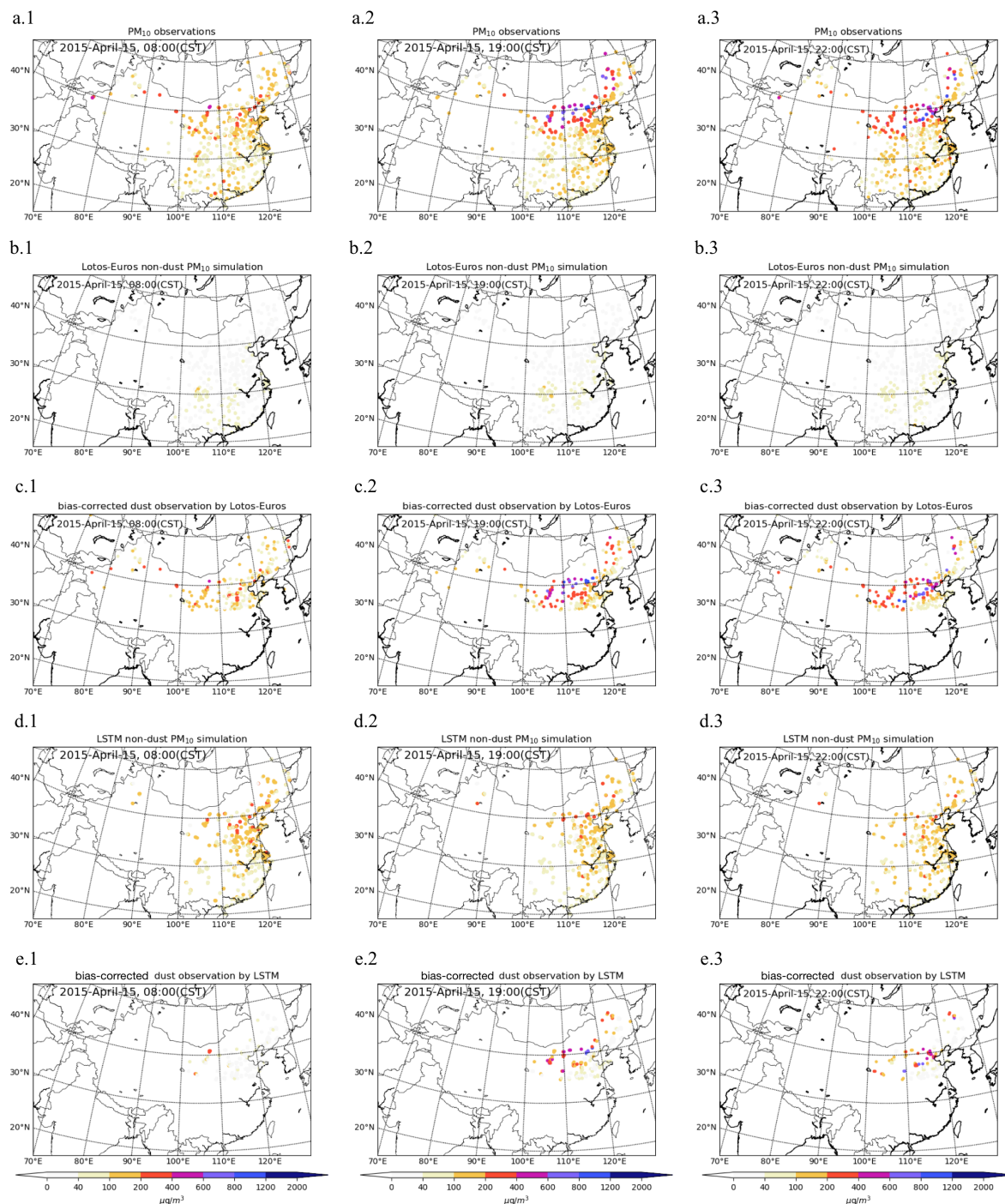


Figure 4. Original PM_{10} measurements (a.1~a.3), LOTOS-EUROS/*non-dust* simulated PM_{10} (b.1~b.3) and the corresponding bias-corrected dust observations (c.1~c.3), LSTM predicted *non-dust* PM_{10} (d.1~d.3) and the derived dust observations (e.1~e.3) at three time snapshots: April 15, 08:00 (a.1~e.1), 19:00 (a.2~e.2) and 22:00 (a.3~e.3)

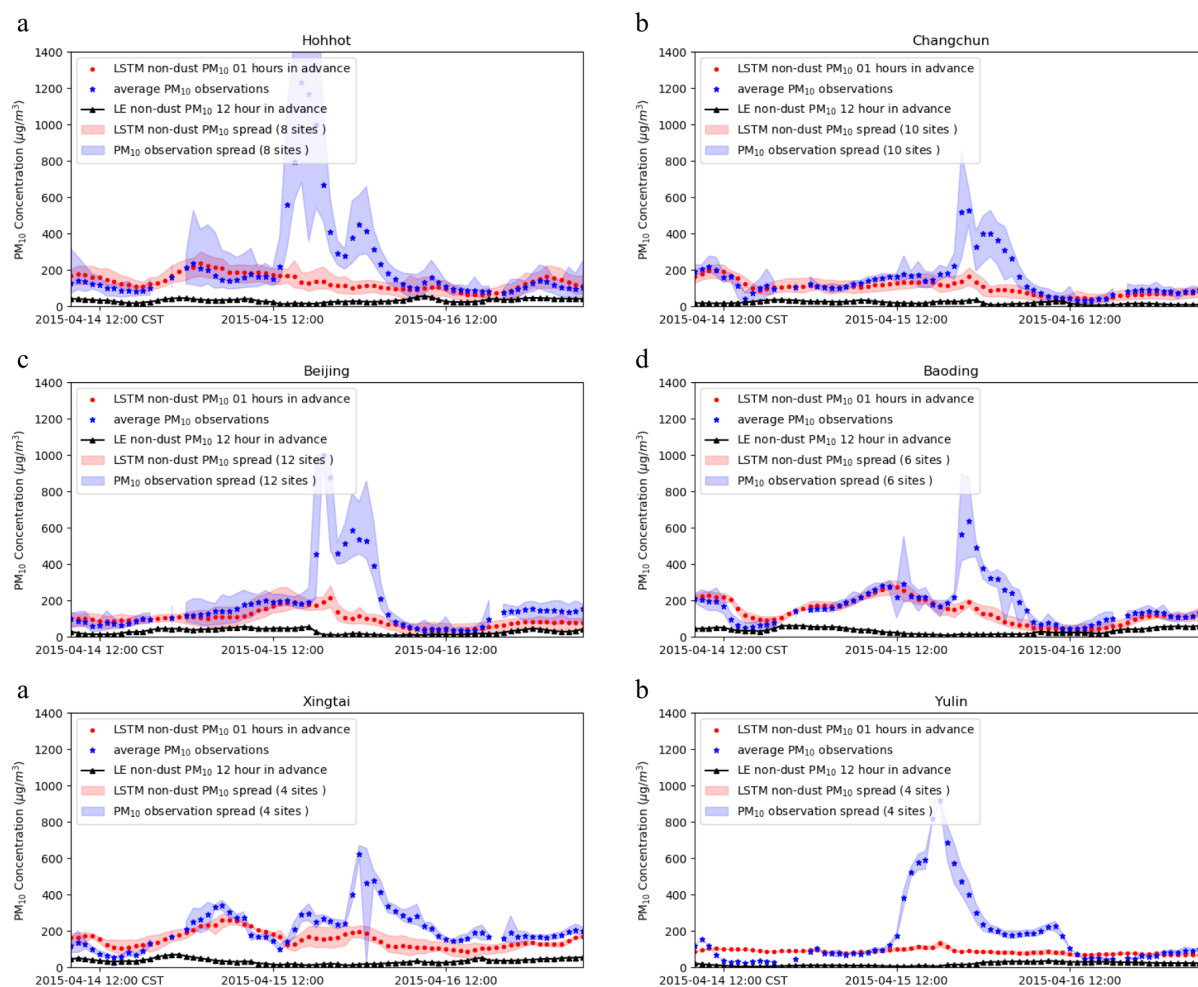


Figure 5. Time series of PM_{10} measurements, LOTOS-EUROS/*non-dust* and LSTM predicted PM_{10} levels at six cities: Hohhot, Changchun, Beijing, Baoding, Xingtai and Yulin. LE: LOTOS-EUROS



4.1 Observation error configuration

A key element of the data assimilation system is the observation error covariance matrix \mathbf{R} . This covariance quantifies the possible difference between simulations and observations. The observations with a smaller representation error have a higher weight in the assimilation process.

- 5 The dust observation error is usually empirically quantified. Lin et al. (2008) assumed that the observation error is proportional to the measurement with a constant factor of 10%. Jin et al. (2018) used a similar error setting but also assigned a higher representation error to low valued measurements since the model might easily results in relative large errors when simulating minor dust loads.

In this study, the availability of multiple measurement sites in a single model grid cell provides an alternative way to quantify the observation error. When multiple observations are present, the spread in the observed values reflects the observation uncertainty. An example is the grid cell covering the city of Beijing, where observations from 12 different field stations are available. Note that it is the grid cell which has the most monitoring stations. The spread of the hourly measurements is shown in Fig.5(c). For each hour, the standard deviation of the measured PM_{10} values is plotted against the mean in Fig.6, where the red markers represent 'regular' polluted conditions, and the blue markers the dust event. The result shows that the spread in the observations (and thus the representation uncertainty) indeed increases with the average pollution level. Based on this result, a simple linear regression is used to obtain a parametrization for the observation error:

$$\sigma = \max(a \cdot y + b, \sigma_{min}) \quad [\mu\text{g}/\text{m}^3] \quad (9)$$

where $a = 0.12$ and $b = 55.7$ are the linear regression parameters based on the dust event data (blue markers). It should be noted that the observation sites in Beijing truncate observations at a maximum of $1000 \mu\text{g}/\text{m}^3$, and therefore observations close to this number are not used since the true values might have been much higher. A minimum observation representation uncertainty of $\sigma_{min} = 100 \mu\text{g}/\text{m}^3$ is used for the 'dust' observations (PM_{10} with bias correction) to avoid a too strong impact of low valued observations (hardly dust) on the estimation of dust emissions. In case the simulation model estimates dust concentrations at the surface while in reality the plume is elevated, the low valued observations might lead to an unrealistic strong decrease of the dust emissions.

For most other grid cells the number of observations sites is simply one, which makes it difficult to parametrize a representation error in a similar way. Therefore, the representation error parametrized for Beijing is used for all other locations too.

4.2 Dust emission estimation

To evaluate the posterior dust emission field that is obtained by assimilation of the bias corrected 'dust' observations, an emission index \mathcal{F}_i (g/m^2) is defined as in (Jin et al., 2018). The index represents the accumulated dust emission in a cell i between April 14 08:00 and April 15 19:00. Fig.7 shows the emission index map of the *a priori* model, and

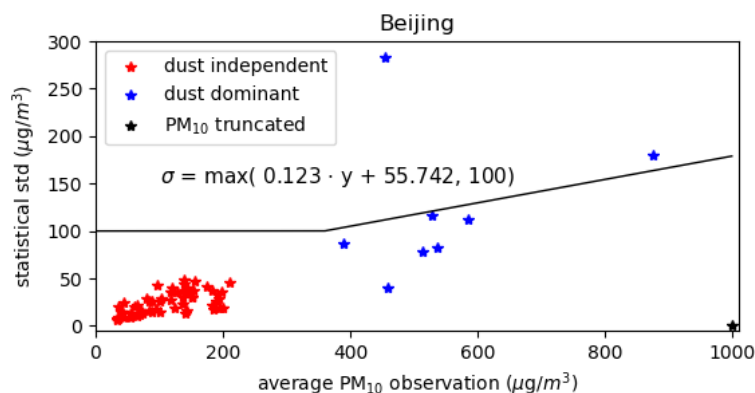


Figure 6. Average vs. Standard deviation of the hourly PM_{10} observations range from April 14 08:00 to April 17 07:00 in the grid cell of Beijing. See Fig.5(c) for the time series.

posteriori emissions obtained from assimilation of either the original PM_{10} observations, or the LOTOS-EUROS or LSTM based bias-corrected 'dust' measurements.

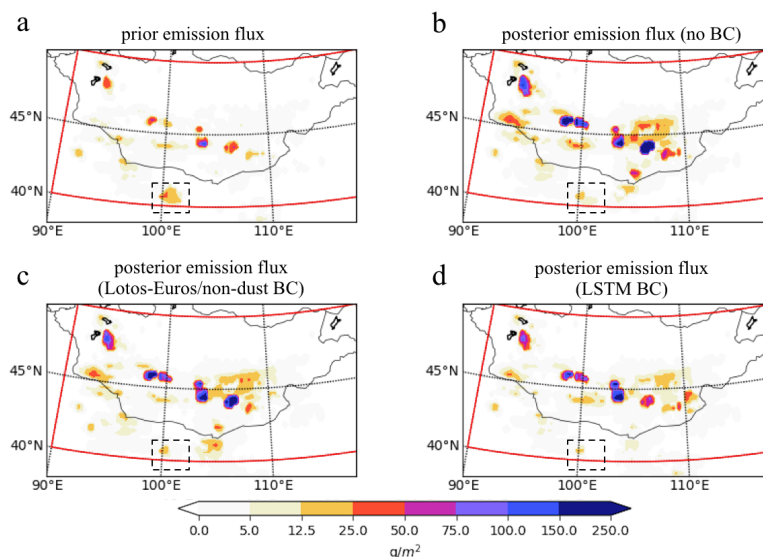


Figure 7. Accumulated dust emission map \mathcal{F} between April 14 08:00 and April 15 19:00 of a *priori* model (a), or (b) a *posteriori* estimates using the original PM_{10} observations, (c) LOTOS-EUROS or (d) LSTM based bias-corrected dust measurements. BC: bias correction

As shown in Fig.7(a), the *a priori* emission was in general rather weak, which resulted in an underestimated surface dust concentration simulation as can be seen for example in Fig.8(a.1)~(a.2). The *posteriori* emissions are almost everywhere higher than the *a priori*. An exception is the black marked region, where the *a priori* emissions are higher.



The emissions from this black-dashed region contributed to a too-early arrival of the dust peak in the model cells over Hohhot and Xingtai as shown in Fig.9(a) and (c).

Fig.7 (b) shows the emission index \mathcal{F} that results from directly assimilating the original PM₁₀ measurements. As expected the estimated emissions are higher than those obtained by assimilating the bias-corrected observations, since all airborne aerosols observed are attributed to be dust. In comparison, the assimilation with LSTM baseline removed data results in a modest emission level as shown in Fig.7(d). The emissions estimated with LOTOS-EUROS based bias-corrected observations are in between, since the resulting 'dust' observations also overestimate the actual dust loads compared to the LSTM based bias-corrected dust measurements.

4.3 Dust simulation and forecast skill

Fig.8 (a)~(d) show the dust simulations at the surface layer at the end of the assimilation window (April 15 19:00, left column) and the forecast 3 hours later (22:00, right column) using the newly estimated emission field. Note that the average dust concentration over the affected downwind regions reached a peak around 22:00. Compared to background simulations in Jin et al. (2018), the *a priori* model simulations have been improved by disabling the topography-based preference factor as mentioned in Section 2.1; however, a large difference from the bias-corrected PM₁₀ observations in Fig. 4(e) is still present.

The *posteriori* concentrations in Fig.8(b.1)~(b.2) are the result of assimilating the original measurements PM₁₀ observations shown in Fig.4(a.1)~(a.2). As expected, these lead to the highest simulated dust concentrations since all the aerosols observed are assumed to represent dust. Especially in the center of the plume, the dust concentration can be as large as 2000 $\mu\text{g}/\text{m}^3$. Fig.8(c.1)~(c.2) show the results when using the LOTOS-EUROS/*non-dust* bias-corrected PM₁₀ observations as 'dust', and although concentrations are lower, they still overestimate the observations. The *posteriori* results using the LSTM bias-corrected measurements provide the lowest dust concentrations as shown in Fig.8(d). Only in the grid cells that are close to the source region, the surface dust concentration reach values as large as 2000 $\mu\text{g}/\text{m}^3$, while in the downwind areas the maximum dust concentrations are usually below 1200 $\mu\text{g}/\text{m}^3$.

To illustrate the improvements of assimilating bias-corrected measurements, Fig. 9 shows the observed and simulated PM₁₀ concentrations in the aforementioned grid cells covering Hohhot, Beijing, and Xingtai. These locations are neither the best nor the worst examples, but illustrate typical results and challenges to be solved in future. For a fair comparison with the PM₁₀ observations, the non-dust aerosol concentrations obtained from either LOTOS-EUROS/*non-dust* or LSTM were added to the dust simulations from the inversion system.

Site Hohhot is close to the main dust source region. The *a priori* model simulated the arrival of the dust plume 8 hours before it was actually visible in the PM₁₀ observations. The assimilation of the observations is able to produce simulations in which the dust plume arrives at the correct time. The assimilation with LSTM bias-corrected data has the best performance, with the peak of the simulated concentrations (dust plus bias) most close to the observed PM₁₀. During the forecast period ($t > \text{April } 15, 19:00$), all three assimilation based forecasts show a decline in concentrations, which slightly overestimate the observations. This can be explained from the fact that the dust

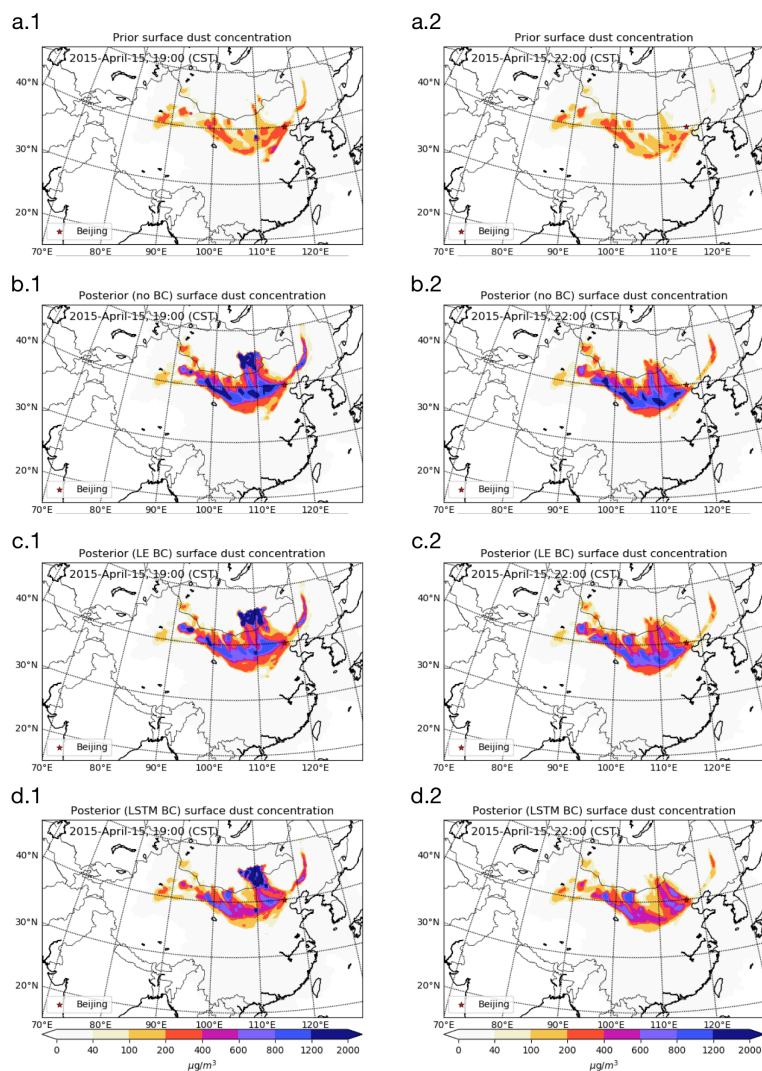


Figure 8. Surface dust concentration of *a* prior (a.1~a.2), *posterior* using no bias-corrected (no BC) data (b.1~b.2), *posterior* using LOTOS-EUROS/*non-dust* bias-corrected (LE BC) data (c.1~c.2), *posterior* using no bias-corrected (LSTM BC) data (d.1~d.2) at April 15, 19:00 (a.1~d.1) and 22:00 (a.2~d.2)



storm is a strong flow-dependent phenomenon in which concentrations at a certain location are strongly correlated to earlier concentrations at upwind locations. For Hohhot, only a limited number of observation sites is located upwind, and therefore hardly any data is available to constrain the concentrations at this location. To improve the forecast at Hohhot it will be necessary to have additional observation data, for example from sites actually within the source region, or from satellites observing the aerosol load over the source region (Jin et al., 2019).

For the grid cell Beijing, which is located further downwind from the dust source region, the arrival of the dust peak is correctly simulated. However, the amplitude of the concentration peak is underestimated compared to the average PM₁₀ observations. As can be seen in Fig.8, the dust plume forms a rather small band over central and northeast China. In each of the three assimilations, the dust concentrations in the band are rather low around Beijing. This suggests that the simulation model simply is not able to increase the dust concentrations here, for example because of uncertainties in the meteorological data, a removal of dust that is too efficient, or because some local sources of dust are absent.

The grid cell Xingtai is located more to the south, and the model is able to simulate high dust concentrations here. The *a priori* model simulates the arrival of a first dust peak already at 13:00, which is however not visible in the PM₁₀ data. The assimilation postpones the arrival of the main dust, which according to the measurements takes place around 22:00 and is already in the forecast period. The forecast simulations all overestimate the amplitude of the peak, especially when using the original PM₁₀ data as proxy for dust. The assimilation with the LSTM based baseline removal shows the best agreement with the observations.

4.4 Evaluation of forecast skill

To evaluate the forecast skill of the assimilation(s), the root mean square error (RMSE) of the reference and three posterior dust simulations (plus *non-dust* predictions) with respect to the observed PM₁₀ over the whole observation sites has been computed for each hour. A time series of this RMSE is shown in Fig.10; after the assimilation window (marked period), the results are based on the forecast simulations. The *a priori* RMSE values at the end of the assimilation window and during the forecast are about 200-250 $\mu\text{g}/\text{m}^3$. Direct assimilation of the original PM₁₀ measurement actually increases these values to above 300 $\mu\text{g}/\text{m}^3$ during the forecast, since dust concentrations become strongly overestimated. Assimilation of the LOTOS-EUROS/*non-dust* baseline removed observations nonetheless reduces the RMSE, in particular within the assimilation window. Strongest decrease in RMSE is obtained using the LSTM based baseline removal, with values of 150-200 $\mu\text{g}/\text{m}^3$ during the forecast.

5 Summary and conclusion

In this study, a dust storm data assimilation experiment has been performed for an event over East Asia in the spring of 2015. PM₁₀ observation data from the China Ministry of Environmental Protection observing network were assimilated into a dust simulation model to estimate the dust emissions. The PM₁₀ measurements themselves

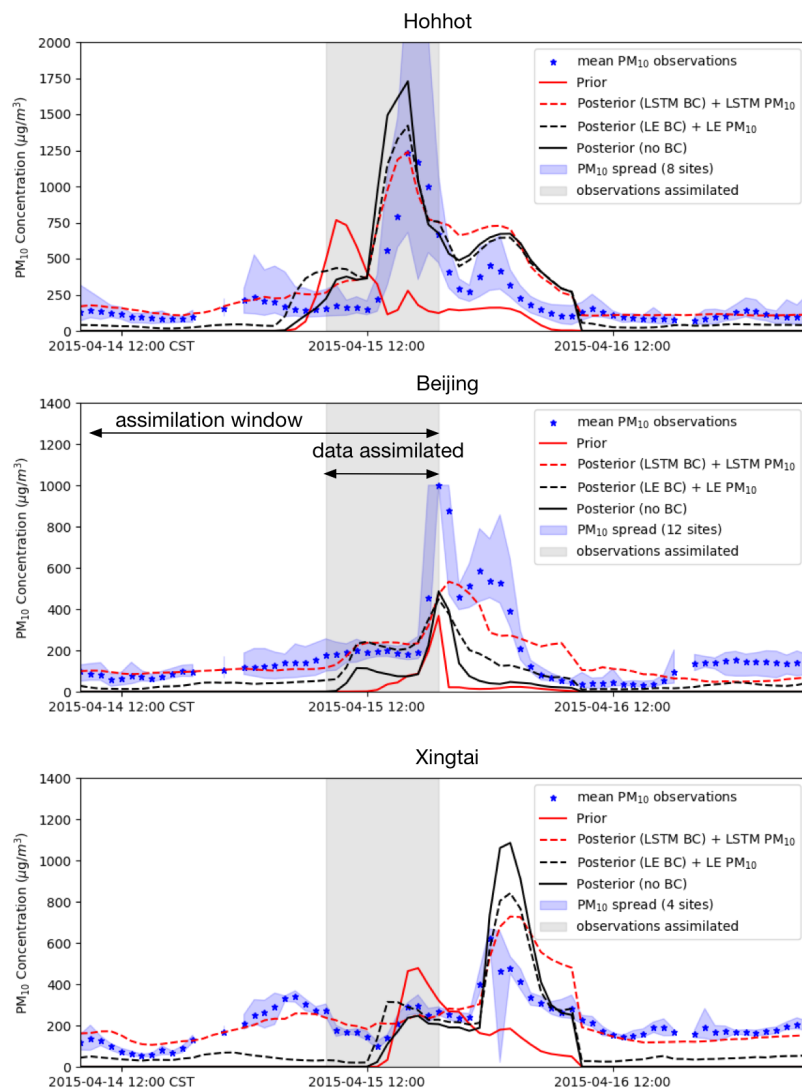


Figure 9. Time series of posterior dust concentration and PM₁₀ observations in three cities: Hohhot, Beijing, Xingtai (observations in the gray shadow are assimilated)

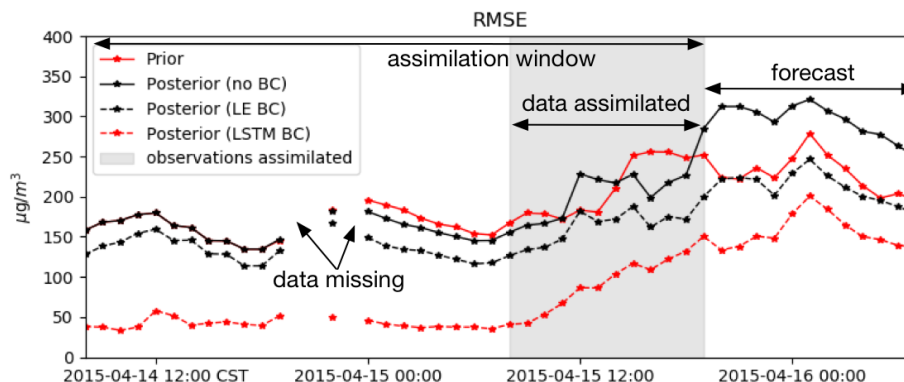


Figure 10. Time series of root mean square error compared to the ground PM_{10} . The assimilation window is set from April 14 08:00 to April 15 19:00, and PM_{10} observation in the gray shadow are assimilated.

are considered as unbiased. They clearly show the arrival of a dust plume throughout the region due to the high spatiotemporal resolution. However, the data cannot be compared directly to dust simulations since they actually represent a sum of the dust particles and other *non-dust* aerosols. Direct assimilation of these measurements would introduce a bias in the assimilation system, since it cannot distinguish between model and observation errors.

- 5 Two methods have been implemented to remove the *non-dust* part the PM_{10} observations during the dust event in order to use them as 'dust' proxy in a dust assimilation system. The first method uses a conventional regional chemical transport model, LOTOS-EUROS/*non-dust*, which simulates the emission, transport, chemistry, and deposition of aerosols mainly related to anthropogenic activities. The second method uses a machine learning model that statistically describes the relations between regular PM_{10} concentrations (outside dust events), and available air
- 10 quality and meteorological data.

The two methods to estimate the *non-dust* part of the PM_{10} load have been validated. The simulations by the LOTOS-EUROS/*non-dust* model in general underestimate the PM_{10} concentrations, which is caused by missing emissions and aerosol components such as secondary organic matter. In comparison, the data-driven machine learning model provided a better agreement with the measurements.

- 15 A variational data assimilation system has been used to estimate the dust emissions that lead to a severe dust storm in April 2015. The system either assimilated the original PM_{10} observations, or the bias-corrected 'dust' observations based on either LOTOS-EUROS/*non-dust* or LSTM model. The posterior simulations using the original observations resulted in a strong overestimation of the dust concentrations, since all PM_{10} are simply attributed to dust. Using the LOTOS-EUROS/*non-dust* bias-corrected observations, a clear improvement on the dust simulation
- 20 has been obtained, but overestimation of dust concentrations is still present. The best results are obtained when using a LSTM model to remove the *non-dust* part of the PM_{10} observations, with *posterior* concentrations in good agreement with the measurements.



The dust emissions estimated using the assimilation can be used to drive a dust forecast. When the original PM_{10} observations were used in the assimilation, the forecast skill of the system actually decreased due to the strong overestimation of dust concentrations. Better forecasts are obtained when using the model-based and especially the LSTM based bias-corrected observations.

5 The computational complexity of learning algorithm used to critically limits the implementations of machine learning especially when it envisions a very large dataset. However, the ever increasing power of computing platforms and massively scalable storage, as well as the progress in computational and optimization methods have led to a matured and computationally efficient paradigm for processing these large-scale problems.

Both our CTM and machine learning based bias correction methods can be improved. It might be useful to improve
10 the CTM simulations by assimilating PM_{10} observations during the hours where no dust storms are present, and use these improved simulations to remove the *non-dust* part of the observations during an event. These additional assimilations would then involve repeated forward ensemble bias-model runs which would be computationally expensive. The machine learning model in our *non-dust* PM_{10} simulation can also be further optimized, such as using a different configuration or deeper neural network, including extra input features like *non-dust* PM_{10} simulation from
15 CTMs (Lin et al., 2019) and other related records.

Data availability

The datasets including measurements and model simulations can be accessed from websites listed in the references or by contacting the corresponding author

20 Author contribution

JJ and HXL conceived the study and designed the experiments. JJ and YX performed the machine learning based non-dust simulation. AS performed the CTM based non-dust simulation. JJ and AS performed the assimilation tests and carried out the data analysis. AS, HXL, AH provided useful comments on the paper. JJ prepared the manuscript with contributions from all co-authors.

25 Acknowledgments

The real-time PM_{10} data are from the network established by the China Ministry of Environmental Protection and accessible to the public at <http://106.37.208.233:20035/>. One can also access the historical profile by visiting <http://www.aqistudy.cn/>.



References

- Benedetti, A., Di Giuseppe, F., Jones, L., Peuch, V. H., Remy, S., and Zhang, X.: The impact of data assimilation on the prediction of Asian desert dust using an operational 4D-Var system, *Atmospheric Chemistry and Physics Discussions*, pp. 1–17, <https://doi.org/10.5194/acp-2018-78>, https://editor.copernicus.org/index.php/acp-2018-78-RC1.pdf?_mdl=msover_md&_jrl=10&_lcm=oc108lcm109w&_acm=get_comm_file&_ms=66357&c=144470&salt=614587343304971732, 2018.
- Berry, T. and Harlim, J.: Correcting Biased Observation Model Error in Data Assimilation, *Monthly Weather Review*, 145, 2833–2853, <https://doi.org/10.1175/MWR-D-16-0428.1>, <http://dx.doi.org/10.1175/MWR-D-16-0428.1>, 2017.
- Cesnulyte, V., Lindfors, A. V., Pitkänen, M. R. A., Lehtinen, K. E. J., Morcrette, J. J., and Arola, A.: Comparing ECMWF AOD with AERONET observations at visible and UV wavelengths, *Atmospheric Chemistry and Physics*, 14, 593–608, <https://doi.org/10.5194/acp-14-593-2014>, https://www.researchgate.net/publication/290315319_1-s20-S2352938515000373-main/figures?lo=1, 2014.
- Chen, G., Li, S., Knibbs, L. D., Hamm, N. A. S., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M. J., and Guo, Y.: A machine learning method to estimate PM_{2.5} concentrations across China with remote sensing, meteorological and land use information, *Science of The Total Environment*, 636, 52–60, <https://doi.org/10.1016/j.scitotenv.2018.04.251>, <http://dx.doi.org/10.1016/j.scitotenv.2018.04.251>, 2018.
- Dee, D. P.: Bias and data assimilation, *Quarterly Journal of the Royal Meteorological Society*, 131, 3323–3343, <https://doi.org/10.1256/qj.05.137>, <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/qj.05.137>, 2005.
- Dee, D. P. and Uppala, S.: Variational bias correction of satellite radiance data in the ERA-Interim reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 135, 1830–1841, <https://doi.org/10.1002/qj.493>, <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.493>, 2009.
- Di Tomaso, E., Nick, Jorba, O., and Garcia-Pando, C. P.: Assimilation of MODIS Dark Target and Deep Blue observations in the dust aerosol component of NMMB-MONARCH version 1.0, *Geoscientific Model Development*, 10, 1107–1129, <https://doi.org/10.5194/gmd-10-1107-2017>, <https://www.geosci-model-dev.net/10/1107/2017/>, 2017.
- Eyre, J. R.: Observation bias correction schemes in data assimilation systems: a theoretical study of some of their properties, *Quarterly Journal of the Royal Meteorological Society*, 142, 2284–2291, <https://doi.org/10.1002/qj.2819>, <https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.2819>, 2016.
- Fan, J., Li, Q., Hou, J., Feng, X., Karimian, H., and Lin, S.: A Spatiotemporal Prediction Framework for Air Pollution Based on Deep RNN, *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-4/W2, 15–22, <https://doi.org/10.5194/isprs-annals-IV-4-W2-15-2017%7D>, https://www.google.nl/_/chrome/newtab?ie=UTF-8, 2017.
- Ginoux, P., Chin, M., Tegen, I., Prospero, J. M., Holben, B., Dubovik, O., and Lin, S.-J.: Sources and distributions of dust aerosols simulated with the GOCART model, *J. Geophys. Res.*, 106, 20 255–20 273, <https://doi.org/10.1029/2000jd000053>, <http://dx.doi.org/10.1029/2000jd000053>, 2001.
- Gong, S. L. and Zhang, X. Y.: CUACE/Dust – an integrated system of observation and modeling systems for operational dust forecasting in Asia, *Atmospheric Chemistry and Physics*, 8, 2333–2340, <https://doi.org/10.5194/acp-8-2333-2008>, <http://dx.doi.org/10.5194/acp-8-2333-2008>, 2008.



- Gong, S. L., Zhang, X. Y., Zhao, T. L., McKendry, I. G., Jaffe, D. A., and Lu, N. M.: Characterization of soil dust aerosol in China and its transport and distribution during 2001 ACE-Asia: 2. Model simulation and validation, *J. Geophys. Res.*, 108, 4262+, <https://doi.org/10.1029/2002jd002633>, <http://dx.doi.org/10.1029/2002jd002633>, 2003.
- Huneus, N., Schulz, M., Balkanski, Y., Griesfeller, J., Prospero, J., Kinne, S., Bauer, S., Boucher, O., Chin, M., Dentener, F., Diehl, T., Easter, R., Fillmore, D., Ghan, S., Ginoux, P., Grini, A., Horowitz, L., Koch, D., Krol, M. C., Landing, W., Liu, X., Mahowald, N., Miller, R., Morcrette, J. J., Myhre, G., Penner, J., Perlwitz, J., Stier, P., Takemura, T., and Zender, C. S.: Global dust model intercomparison in AeroCom phase I, *Atmospheric Chemistry and Physics*, 11, 7781–7816, <https://doi.org/10.5194/acp-11-7781-2011>, <http://dx.doi.org/10.5194/acp-11-7781-2011>, 2011.
- Jin, J., Lin, H. X., Heemink, A., and Segers, A.: Spatially varying parameter estimation for dust emissions using reduced-tangent-linearization 4DVar, *Atmospheric Environment*, 187, 358–373, <https://doi.org/https://doi.org/10.1016/j.atmosenv.2018.05.060>, <https://www.sciencedirect.com/science/article/pii/S1352231018303704>, 2018.
- Jin, J., Segers, A., Heemink, A., Yoshida, M., Han, W., and Lin, H.-X.: Dust Emission Inversion Using Himawari-8 AODs Over East Asia: An Extreme Dust Event in May 2017, *Journal of Advances in Modeling Earth Systems*, 11, 446–467, <https://doi.org/10.1029/2018MS001491>, <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018MS001491>, 2019.
- Khade, V. M., Hansen, J. A., Reid, J. S., and Westphal, D. L.: Ensemble filter based estimation of spatially distributed parameters in a mesoscale dust model: experiments with simulated and real data, *Atmospheric Chemistry and Physics*, 13, 3481–3500, <https://www.atmos-chem-phys.net/13/3481/2013/>, 2013.
- Li, G., Bei, N., Cao, J., Wu, J., Long, X., Feng, T., Dai, W., Liu, S., Zhang, Q., and Tie, X.: Widespread and persistent ozone pollution in eastern China during thenon-winter season of 2015: observations and source attributions, *Atmospheric Chemistry and Physics*, 17, 2759–2774, <http://www.atmos-chem-phys.net/17/2759/2017/>, 2017a.
- Li, X., Peng, L., Hu, Y., Shao, J., and Chi, T.: Deep learning architecture for air quality predictions, *Environmental Science and Pollution Research*, 23, 22 408–22 417, <https://doi.org/10.1007/s11356-016-7812-9>, <http://dx.doi.org/10.1007/s11356-016-7812-9>, 2016.
- Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C., and Chi, T.: Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation - ScienceDirect, *Environmental Pollution*, 231, 997–1004, <https://doi.org/https://doi.org/10.1016/j.envpol.2017.08.114>, <https://www.sciencedirect.com/science/article/pii/S0269749117307534>, 2017b.
- Lin, C., Wang, Z., and Zhu, J.: An Ensemble Kalman Filter for severe dust storm data assimilation over China, *Atmospheric Chemistry & Physics*, 8, 2975–2983, <https://www.atmos-chem-phys.net/8/2975/2008/>, 2008.
- Lin, H. X., Jin, J., and van den Herik, J.: Air Quality Forecast through Integrated Data Assimilation and Machine Learning, <http://insticc.org/node/TechnicalProgram/icaart/presentationDetails/75552>, 2019.
- Liu, M., Westphal, D. L., Wang, S., Shimizu, A., Sugimoto, N., Zhou, J., and Chen, Y.: A high-resolution numerical study of the Asian dust storms of April 2001, *J. Geophys. Res.*, 108, 8653+, <https://doi.org/10.1029/2002jd003178>, <http://dx.doi.org/10.1029/2002jd003178>, 2003.
- Lorente-Plazas, R. and Hacker, J. P.: Observation and Model Bias Estimation in the Presence of Either or Both Sources of Error, *Monthly Weather Review*, 145, 2683–2696, <https://doi.org/10.1175/MWR-D-16-0273.1>, <http://dx.doi.org/10.1175/MWR-D-16-0273.1>, 2017.



- Marticorena, B. and Bergametti, G.: Modeling the atmospheric dust cycle: 1. Design of a soil-derived dust emission scheme, *J. Geophys. Res.*, 100, 16 415–16 430, <https://doi.org/10.1029/95JD00690>, <http://dx.doi.org/10.1029/95JD00690>, 1995.
- Niu, T., Gong, S. L., Zhu, G. F., Liu, H. L., Hu, X. Q., Zhou, C. H., and Wang, Y. Q.: Data assimilation of dust aerosol observations for the CUACE/dust forecasting system, *Atmospheric Chemistry and Physics*, 8, 3473–3482, <https://doi.org/10.5194/acp-8-3473-2008>, <http://dx.doi.org/10.5194/acp-8-3473-2008>, 2008.
- 5 Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A., Martins, J. V., Li, R. R., Ichoku, C., Levy, R. C., Kleidman, R. G., Eck, T. F., Vermote, E., and Holben, B. N.: The MODIS Aerosol Algorithm, Products, and Validation, *Journal of the Atmospheric Sciences*, 62, 947–973, 2005.
- Sekiyama, T. T., Tanaka, T. Y., Shimizu, A., and Miyoshi, T.: Data assimilation of CALIPSO aerosol observations, *Atmospheric Chemistry and Physics*, 10, 39–49, <https://www.atmos-chem-phys.net/10/39/2010/>, 2010.
- 10 Shao, P., Tian, H., Sun, Y., Liu, H., Wu, B., Liu, S., Liu, X., Wu, Y., Liang, W., Wang, Y., Gao, J., Xue, Y., Bai, X., Liu, W., Lin, S., and Hu, G.: Characterizing remarkable changes of severe haze events and chemical compositions in multi-size airborne particles (PM₁, PM_{2.5} and PM₁₀) from January 2013 to 2016–2017 winter in Beijing, China, *Atmospheric Environment*, 189, 133–144, <https://www.sciencedirect.com/science/article/pii/S1352231018304291>, 2018.
- 15 Shao, Y. P., Raupach, M. R., and Leys, J. F.: A model for predicting aeolian sand drift and dust entrainment on scales from paddock to region, *Australian Journal of Soil Research*, 34, 309+, <https://doi.org/10.1071/sr9960309>, <http://dx.doi.org/10.1071/sr9960309>, 1996.
- 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., and Schaap, M.: Source apportionment of PM_{2.5} across China using LOTOS-EUROS, *Atmospheric Environment*, <https://doi.org/10.1016/j.atmosenv.2017.06.003>, <http://dx.doi.org/10.1016/j.atmosenv.2017.06.003>, 2017.
- 20 Wang, Y. Q., Zhang, X. Y., Gong, S. L., Zhou, C. H., Hu, X. Q., Liu, H. L., Niu, T., and Yang, Y. Q.: Surface observation of sand and dust storm in East Asia and its application in CUACE/Dust, *Atmospheric Chemistry and Physics*, 8, 545–553, <https://doi.org/10.5194/acp-8-545-2008>, <http://dx.doi.org/10.5194/acp-8-545-2008>, 2008.
- Wang, Z., Ueda, H., and Huang, M.: A deflation module for use in modeling long-range transport of yellow sand over East Asia, *J. Geophys. Res.*, 105, 26 947–26 959, <https://doi.org/10.1029/2000jd900370>, <http://dx.doi.org/10.1029/2000jd900370>, 2000.
- 25 WMO: WMO AIRBORNE DUST BULLETIN: Sand and Dust Storm Warning Advisory and Assessment System, https://library.wmo.int/doc_num.php?explnum_id=3416, 2017.
- Xu, L., Batterman, S., Chen, F., Li, J., Zhong, X., Feng, Y., Rao, Q., and Chen, F.: Spatiotemporal characteristics of PM_{2.5} and PM₁₀ at urban and corresponding background sites in 23 cities in China, *Science of The Total Environment*, 599600, 2074–2084, <http://www.sciencedirect.com/science/article/pii/S0048969717311488>, 2017.
- 30 Yoshida, M., Kikuchi, M., Nagao, T. M., Murakami, H., Nomaki, T., and Higurashi, A.: Common Retrieval of Aerosol Properties for Imaging Satellite Sensors, *Journal of the Meteorological Society of Japan. Ser. II, advpub*, <https://doi.org/10.2151/jmsj.2018-039>, https://www.jstage.jst.go.jp/article/jmsj/advpub/0/advpub_2018-039/_article/-char/en, 2018.
- 35 Yumimoto, K., Uno, I., Sugimoto, N., Shimizu, A., Liu, Z., and Winker, D. M.: Adjoint inversion modeling of Asian dust emission using lidar observations, *Atmospheric Chemistry and Physics*, 8, 2869–2884, <https://doi.org/10.5194/acp-8-2869-2008>, <http://dx.doi.org/10.5194/acp-8-2869-2008>, 2008.



Yumimoto, K., Murakami, H., Tanaka, T. Y., Sekiyama, T. T., Ogi, A., and Maki, T.: Forecasting of Asian dust storm that occurred on May 10–13, 2011, using an ensemble-based data assimilation system, *Particuology*, 28, 121–130, <https://doi.org/10.1016/j.partic.2015.09.001>, <http://dx.doi.org/10.1016/j.partic.2015.09.001>, 2016.

Zhang, S.: Nearest neighbor selection for iteratively kNN imputation, *Journal of Systems and Software*, 85, 2541–2552,

5 <https://www.sciencedirect.com/science/article/pii/S0164121212001586?via%3Dihub>, 2012.