Aeolian Research 35 (2018) 69–84

Contents lists available at ScienceDirect

Aeolian Research

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

Prediction of aerosol optical depth in West Asia using deterministic models and machine learning algorithms

Seyed Omid Nabavi[a,*], Leopold Haimberger[a], Reyhaneh Abbasi[b], Cyrus Samimi[c,d]

[a] Department of Meteorology and Geophysics, University of Vienna, Faculty of Earth Sciences, Geography and Astronomy, UZA II, Althanstrasse 14, A-1010 Vienna, Austria
[b] The Austrian Academy of Sciences Acoustics Research Institute, Wohllebengasse 12-14, A-1040 Vienna, Austria
[c] Faculty of Biology, Chemistry and Earth Sciences, University of Bayreuth, Universitätsstr. 30, 95447 Bayreuth, Germany
[d] Bayreuth Center of Ecology and Environmental Research, BayCEER, Dr. Hans-Frisch-Straße 1-3, 95448 Bayreuth, Germany

ARTICLE INFO

Keywords:
Machine learning algorithms
Deterministic weather prediction models
Dust storms
West Asia
MODIS deep blue AOD

ABSTRACT

Because of the lack of ground-based observations in large parts of West Asia, Aerosol Optical Depth (AOD) is mainly monitored by using remote sensing techniques. AOD can also be predicted by short term forecasts with commonly called Deterministic weather prediction models (DMs). The skill of DMs in reproducing remotely sensed observations when averaged over monthly time scales over West Asia is rather limited due to significant uncertainties in inputs and complexity of dust, which is the dominant type of aerosols in the region.

Machine Learning Algorithms (MLAs), which require much less computational expenses than DMs, can be used. Using Moderate Resolution Imaging Spectroradiometer (MODIS) Deep Blue (DB) AOD as the representative of response variable, MLAs, especially Multivariate Adaptive Regression Splines (MARS) and Support Vector Machines (SVM), outperformed DMs on monthly time scale. MLAs have yielded lower prediction error (RMSE) and higher correlation with observations than DMs. In addition, findings disclosed that DMs, especially MACC, have failed to simulate observed AOD values over western Iran where the Zagros Mountains prevent advection of fine dust particles to the east of the study area. Prediction errors of MLAs and DMs along with major DB AOD peaks, over Iraq, can be traced back to the rough resolution of variable datasets, omission of some unknown influential predictors representing the life cycle of dust and/or other aerosols, and scarcity of extreme cases. It also remains to be tested in how far the results presented can be generalized to other regions and time scales.

1. Introduction

Aerosol Optical Depth (AOD) is the measure of the extinction of the solar beam by aerosols (e.g., urban pollutants, smoke, mineral dust, and sea salt) distributed in the vertical column of atmosphere (Choi et al., 2013). The attenuation of light by dust particles contributes to the major portion of AOD from the west coast of North Africa, through West Asia, defined here as the area between latitudes 29°–37° N and longitudes 39°–49° E (Fig. 2), well into Central Asia (Prospero et al., 2002). Because of this, one may consider AOD as an indicator of dust abundance over these regions. Deterministic weather prediction models (DMs) are commonly used for the prediction of aerosol concentration (Marticorena and Bergametti, 1995, Liu et al., 2007, Kumar et al., 2014). DMs simulate the atmospheric environment using a mathematical representation of physical and chemical mechanisms (Hoshyaripour et al., 2016). For short-range (up to three days) operational forecasts where there is nearly complete theoretical knowledge about the nature of the relationships between prognostic variables and boundary conditions, they are quite powerful prediction tools. The improvement of DMs' structure and computational power will assist their strong physical basis for more accurate short-term predictions in the forthcoming years (Taheri Shahraiyni and Sodoudi, 2016).

While aerosol surface concentrations are the most considered output of DMs, since the high concentration of aerosols affects human health, they are rarely observed over West Asia. Because of this, they have a limited use for model validation. Monthly mean AOD is another important output parameter which, besides being relevant for climate, is extensively observed by satellites and is, therefore, a valuable benchmark for any method trying to predict it. So far DMs still have difficulties reproducing AOD on a monthly time scale because of limitations in both the accuracy of the predicted aerosol concentrations and of the

* Corresponding author.
E-mail addresses: seyed.omid.nabavi@univie.ac.at (S.O. Nabavi), leopold.haimberger@univie.ac.at (L. Haimberger), rabbasi@kfs.oeaw.ac.at (R. Abbasi), cyrus.samimi@uni-bayreuth.de (C. Samimi).

https://doi.org/10.1016/j.aeolia.2018.10.002
Received 22 February 2018; Received in revised form 2 September 2018; Accepted 22 October 2018
Available online 30 October 2018
1875-9637/ © 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).

Acronyms

DMs	deterministic weather prediction models
MLAs	machine learning algorithms
DB	deep blue
FSC	feature selection criteria
MARS	multivariate adaptive regression splines
SVMs	support vector machines
MLR	multiple linear regression
ANN	artificial neural networks
RF	random forest
DUP	dust uplift potential
SM	soil moisture
ST	soil temperature
SF	source function
WASF	West Asia source function
SPEI	standardized precipitation-evapotranspiration index
PCC	Pearson correlation coefficient
SCC	Spearman correlation coefficient
MI	mutual information

observation operators, which calculate AOD from the concentrations (Liu et al., 2011a).

DMs are not the only way to predict monthly mean AOD. As any observable, AOD can be seen as a stochastic variable that depends on several potential predictors at least in a statistical sense. These dependencies can be estimated if there exists a significant number of observations of both predictors and predictands. Machine Learning Algorithms (MLAs) have shown promising performance inferring such relationships, particularly in engineering problems, for more than three decades (Carbonell et al., 1983; Cortes and Vapnik, 1995; Kotsiantis et al., 2007; Hempel et al., 2012; Abbasi et al., 2014; LeCun et al., 2015; Mayr et al., 2018). MLAs can identify the underlying behavior of a system from long-term observations at relatively low computational cost (Lary et al., 2016). MLAs are already used for the prediction of air quality in urban areas (Taheri Shahraiyni and Sodoudi, 2016) and have also been applied to the adjustment of satellite AOD (Hyer et al., 2011; Albayrak et al., 2013) to have a better fit to ground-based observations. A rather large class of algorithms may be referred to as Machine Learning Algorithms. Probably the best known way to estimate statistical relationships between predictors and predictands is Multiple Linear Regression (MLR). Klingmüller et al. (2016) modeled annual the Moderate Resolution Imaging Spectroradiometer (MODIS) Deep Blue (DB) AOD in West Asia in a coarse resolution of 2 degrees using different predictors and MLR. In order to evaluate the importance of predictors, they applied the Akaike Information Criterion (AIC). Results point to soil moisture as the dominant factor for AOD (dust) prediction in Saudi Arabia and Iraq.

Although their study could shed light on the applicability of MLAs in AOD prediction, the determination of an efficient MLA based prediction tool requires the inter-comparison of different MLAs over a long-term period. In addition, there is a lack of comparative studies evaluating DM and MLA performance in AOD prediction on finer spatial and temporal resolution.

We therefore aim to make a more comprehensive and objective comparison of AOD prediction methods for West Asia, using satellite-measured AOD. Output from five MLAs and two DMs has been compared at higher spatial and temporal resolutions than can be found in the literature. Specifically we try (i) to demonstrate the feasibility of MLAs for predicting monthly mean AOD, (ii) to detect the most influential predictors, (iii) to check whether the estimated dependencies between predictors and AOD can be found also in DMs. This might help improving DMs which are known to be deficient in predicting monthly mean AOD, at least over West Asia.

It is worth mentioning that the same potential predictors have been provided to all MLAs. From these the best predictors have been selected by considering the results of two filter-type methods (Chandrashekar and Sahin, 2014). The present study uses input variables and the predictand recorded at the same time, i.e. valid for the same month. The input variables for the MLAs could be observations, analyses or forecasts from weather/climate models. In this paper, we use only analysis data. The descriptions of research data and methods are presented in Section 2. Sections 3 and 4 are allotted to results and conclusions, respectively.

2. Materials and methods

The general process of MLA training and prediction is shown in Fig. 1.

Since MLAs require a significant number of observations for training, the first priority is to use observation datasets with relatively high spatial (less than 1 degree) and temporal resolution. In addition, dust activity, which mainly determines AOD over the region, is the result of complex interactions between the atmosphere and land surface, so its predictors should be also representatives of both environments. We stress that this study assumes dust aerosols as the main contributor of AOD measured over the region. This assumption is consistent with findings of Boloorani et al. (2013). They have run the trajectories of dust-laden air parcels and found that western Iran is the main receptor of dust storms formed in Iraq and Saudi Arabia. Therefore, potential features are selected in such a way to primarily represent

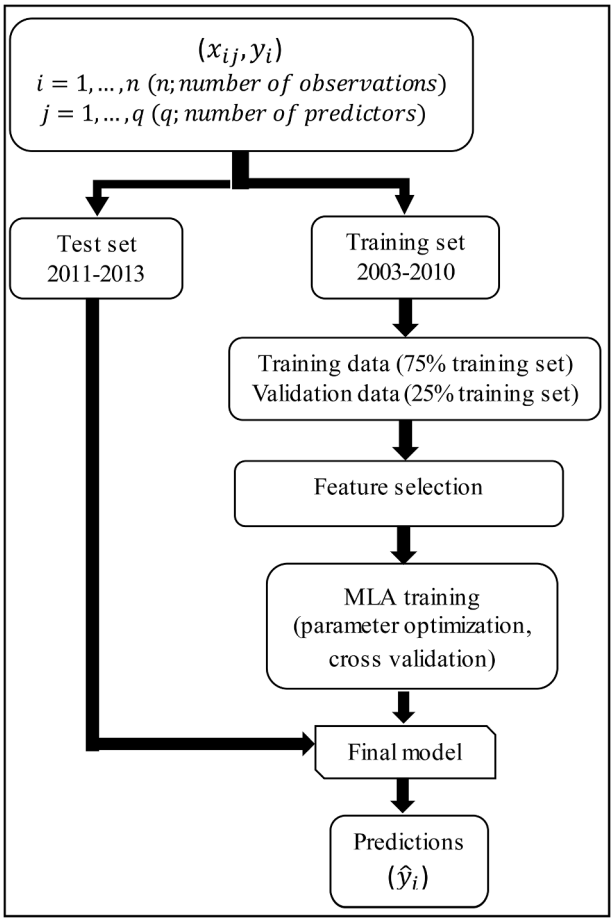


Fig. 1. Flow chart of MLA prediction. During the training process parameters are automatically optimized using cross validation until the final model is defined. This model is applied to the test set.

the dust cycle. However, most of them such as wind speed, soil moisture and temperature, and precipitation can explain the formation of biomass burning aerosols during sporadic wildfires of Zagros Mountains forest steppe (Jaafari et al., 2017), in western Iran, as well. In the following subsection, datasets and their sources used for MLA setup are described with the advisable brevity.

2.1. Input variables of MLAs

In this study, the collection 6 MODIS DB AOD (Hsu et al., 2004; Sayer et al., 2014) is chosen as long-term record of predictand. While Terra DB AOD is available since 2000, we used MODIS daily DB AOD from the Aqua platform, available since 2002, because it has much less missing data (Fig. 2). This dataset (MYD04_L2), with a resolution of 10 km, is pre and post-processed (satellite images are spatially combined) by the Level-1 and Atmosphere Archive & Distribution System (LAADS), https://ladsweb.modaps.eosdis.nasa.gov/search/. Following Tao et al. (2015) and Hsu et al. (2013), only DB retrievals with quality flags set to 2 or 3 (52% and 26% of retrievals (89,746 pixels), respectively) are examined in this study and the quality flag 1 is used for filtering out invalid retrievals. Quality flags of DB retrievals in collection 6 are dependent on the number (N) of retrieved AOD pixels at 550 nm and their standard deviation (σ) within 10 × 10 pixels. The minimum N = 40 and 60 out of 100 are considered for quality flags 2 and 3, respectively (Hsu et al., 2013). For each grid point daily data have been aggregated to monthly means if the number of days with valid retrievals per month is higher than 20.

Although the use of MODIS DB AOD could provide the needed measurements over the study area, some caution has to be taken when assuming this product as a replacement for the ground truth. According to Sayer et al. (2013, 2014), the DB algorithm yields higher uncertainties for high-AOD (dust) cases than for clean conditions. In addition, the former study shows that the quality of retrievals may affect the agreement of observations and measurements. For example, the Pearson Correlation Coefficient (PCC) between MODIS DB AOD and observations from AErosol RObotic NETwork (AERONET) station Solar Village in Saudi Arabia decreased from 0.82 to 0.69 when the changing quality flag from 3 to 2. However, the comparison of monthly DB AOD with observations from four AERONET stations in the regions shows that this product is accurate enough (Fig. 3) to be chosen as the predictand on a monthly scale. These correlations are clearly higher than those with predicted AOD shown later.

For training and application of the MLAs, we apply the natural logarithm to the monthly mean DB AOD data to make the data distribution more symmetric (Feng et al., 2013; Benoit, 2011). This transformation reduces the negative effect of high extreme values on

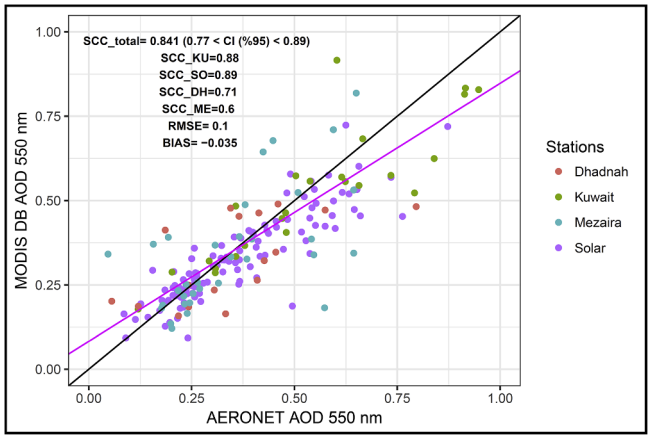


Fig. 3. Spearman correlation coefficient between MODIS DB AOD 550 nm and observations from four AERONET stations including Dhadnah, Kuwait University, Mezaira, and Solar Village during the study period. Please note that only station Kuwait is within the study window. Mezaira and Dhadnah are affected by sand dunes or are close to water, which causes somewhat lower correlations.

MLA predictions, as will be discussed later. This implies, however, that the MLAs also predict the logarithm of DB AOD, which has to be transformed back (with the exponential function) in order to compare it with observed AOD.

As discussed before, dust is the dominant type of AOD in West Asia. This is shown in Fig. 4. The MODIS DB Angstrom exponent is less than 1, a threshold which is used as a proxy to discriminate dust particles or sea salt from fine aerosols (Dubovik et al., 2002), over most of the study area except western Iran where the Zagros Mountains act as barrier against the entrance of dust particles to the west of Iran. Although we are aware of the criterion for the discrimination of dust from sea salt, proposed by Ginoux et al. (2012), single scattering albedo < 0.95 was considered to rule out the contribution of sea salt to AOD, we did not consider it. This is because sea salt has a small contribution within our study area where is relatively far away from large water bodies.

The successful prediction of AOD, by MLAs, in West Asia requires a general insight into the influential factors which are governing the dust cycle. In other words, we first need to roughly determine those factors which are of high importance in dust emission, transportation, and deposition. Some of these potential predictors may, however, be redundant and should be eliminated afterwards by Feature Selection Criteria (FSC, Subsection 2.2). Following the literature (Klingmüller et al., 2016; Yu et al., 2015; Kaboodvandpour et al., 2015) and the

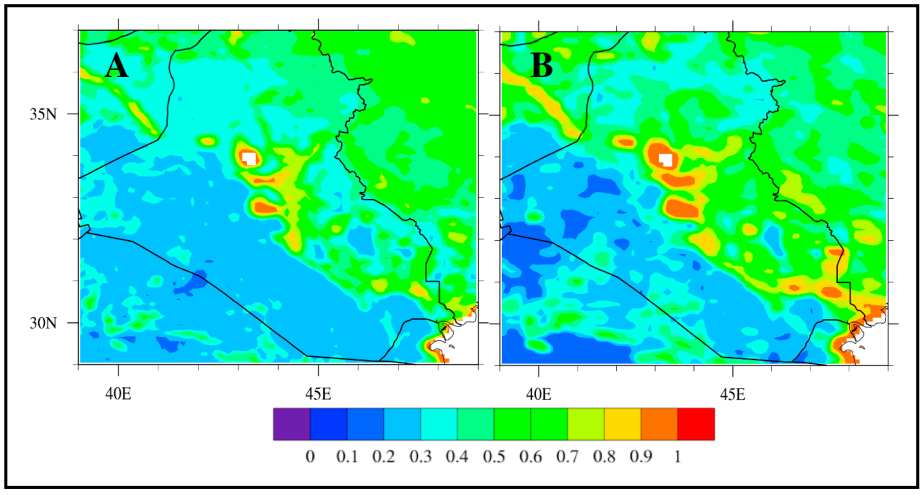


Fig. 2. Fraction of missing data (cases with no retrieval or cases flagged with 1) of Aqua (A) and Terra (B) DB AOD during the study period (Apr-Sep 2003–2013). The red spots in the middle of Iraq are related to water reservoirs behind major dams (e.g. Mossul dam). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

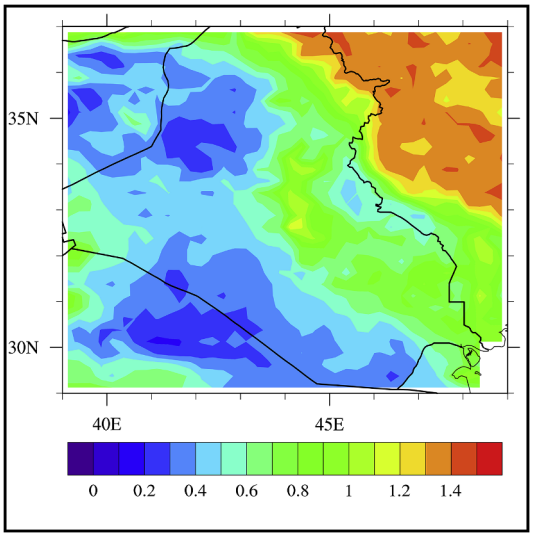


Fig. 4. Averaged DB angstrom exponent during warm months (Apr–Sep) of 2003–2013. As defined in Dubovik et al., (2002), angstrom exponents less than 1 indicate that the aerosol mass mostly consists of dust or other large aerosols like sea salt.

authors' experiences, nine environmental parameters are chosen as potential predictors for dust: (i) 10 m wind, (ii) vertical velocity (omega), (iii) soil temperature, (iv) albedo, (v) soil moisture, (vi) precipitation, (vii) vegetation cover, (viii) drought intensity, and (ix) susceptibility of dust emission. The first four parameters are acquired from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis dataset, accessible from 1979 to present, with a grid resolution of 0.75 degrees (Dee et al., 2011). The low-level horizontal erosive speed of air parcels is represented by Dust Uplift Potential (DUP) at 10 m (White, 1979). According to Cowie et al. (2015), DUP is calculated as follows:

$$\mathrm{DUP}=\begin{cases} U^3\left(1+\frac{U_t}{U}\right)\left(1-\frac{U_t^2}{U^2}\right), & \text{if } U>\ U_t \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where U is wind speed at 10 m and U_t is a threshold for dust emission. Here U_t is defined as the long term average of 10-meter wind speed simultaneous with DB AOD $>$ 0.7. We also expect that the uplift from near the surface to higher levels is related to omega at 850 hPa from ERA-Interim, therefore we include this field as a predictor as well. The top layer (1–7 cm) soil temperature is provided by the ERA-Interim "soil temperature level 1" (ST) parameter which is available every three hours. Surface albedo from ERA-Interim is used to feed the high surface reflectance of dust-prone areas into MLAs. All mentioned ECMWF datasets are reanalyzed at 6 am (around 9 am local time). This time of the day is chosen because the visual examination of MODIS images shows that the first dust plumes of study area mostly form a few hours after sunrise. Mid-morning is also found as the initial time of dust storms in previous studies Mbourou et al. (1997) and Schepanski et al. (2009).

The European Space Agency Climate Change Initiative (ESA-CCI) has provided daily surface soil moisture (SM) based on satellite mounted active and passive microwave sensors. In this study, the COMBINED data set, on the grid resolution of 0.25 degree, is used (Liu et al., 2011b, 2012; Wagner et al., 2012). Precipitation has been taken from the Global Precipitation Climatology Centre (GPCC) dataset (Schneider et al., 2011). It provides monthly 0.5° × 0.5° precipitation (Total Full V7) from quality controlled station data during 1901–2013. In order to incorporate the variations of vegetation cover in dust predictions, we have used the normalized difference vegetation index (NDVI) dataset from the Global Inventory Modeling and Mapping Studies (GIMMS), called NDVI3g. Tucker et al. (2004) have provided this refined product from 15-day maximum NDVI values. NDVI3g, derived from AVHRR sensor data of NOAA 7–18 satellites, has 1/12° spatial and bi-monthly temporal resolutions and it covers the time period from 1981 to 2015. The impact of successive droughts on dust outbreaks is considered by using the 9-month aggregated Standardized Precipitation-Evapotranspiration Index (SPEI). SPEI is a simple measure of the water surplus or deficit that is calculated based on the monthly (or weekly) difference between precipitation and potential evapotranspiration (Vicente-Serrano et al., 2010).

The uneven potential of dust emission in arid areas necessitates the use of a spatially varying dust Source Function (SF) in dust models. In fact, SF allocates a certain potential of dust release to each place. Following Nabavi et al. (2016), we have used the West Asia Source Function (WASF) which is determined from the long-term study of Aerosol Index (AI) and MODIS DB AOD (Fig. 5-A). They have analyzed AI for the large-scale, binary determination of dust sources. Subsequently, the potential of identified sources is quantitatively defined as the long-term fraction of dust occurrence determined by DB AOD $>$ 0.7.

In comparison with original WASF, we made two modifications. First, the threshold of dust occurrence within dust sources in the northwest of Iraq was decreased to 0.6 (instead of 0.7) to detect any active dust sources. Second, we considered the possibility of dust occurrence in other regions of the study area by applying a dust threshold DB AOD $>$ 0.8 (Fig. 5-B). The higher threshold (than major dust sources in the northwest of Iraq) ensures that high AOD over these areas (especially over the southeast of Iraq) is not because of transporting

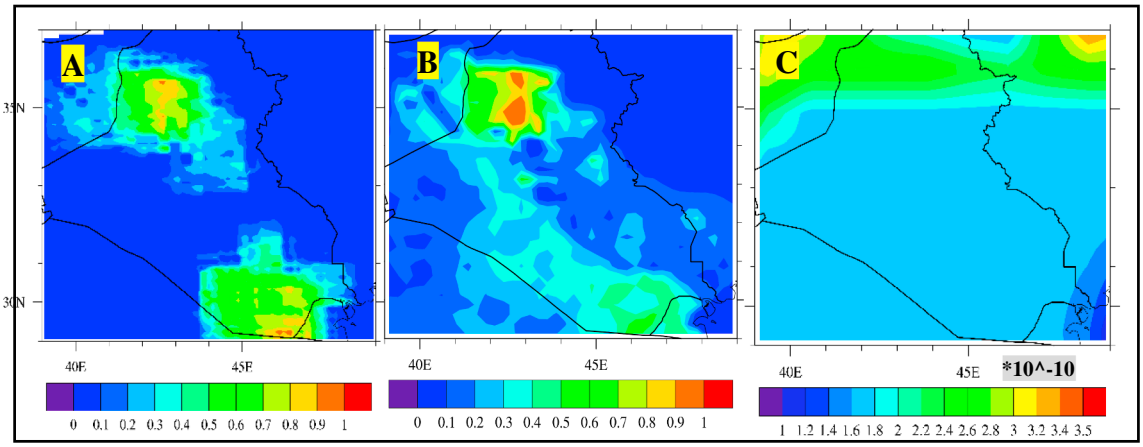


Fig. 5. A and B are respectively the original (Nabavi et al., 2017) and modified WASF, calculated based on the fraction of dust occurrence. C depicts the source function as used by MACC (in units kg s^2 m^{-5}).

dust originated from upstream sources. The original WASF assumes no dust emission out of dust sources, which is likely too stringent.

It is worth mentioning that all discussed datasets (Table 1) are interpolated or aggregated to a 0.25-degree grid and, if needed, averaged to get monthly means. The only exception is DUP that is summed monthly. It should be noted that all MLAs and the Weather Research and Forecasting Model coupled with chemistry (WRF-chem) applied here use modified WASF as source function, whereas the Monitoring Atmospheric Composition and Climate (MACC) product uses source functions shown in panel C of Fig. 5.

One challenging aspect of this study is how to set up a fair comparison between the estimates of AODs from MLAs with those from DMs and MACC. DMs are prediction models, i.e. they use instantaneous data as input and make short term forecasts of dust (and finally AOD), in our case up to 36 h ahead. On short time scales of a few days this is quite different from MLAs, which estimate AOD for the same point in time. On a monthly time scale, however, the difference almost disappears, since also for the DMs the input data for AOD are from the same month as the predicted AOD. The MACC AOD product is an assimilated product, it contains information AOD observations from the past as well as from the time for which the analysis is valid. As such it even has an information advantage compared to MLAs and WRF-Chem. For the comparison , we have used DM daily simulations at 9 UTC (12.30 at local time) for getting monthly averages of AOD. This time of day is chosen because MACC reanalysis is only available 6-hourly and it is the closest time to MODIS-Aqua overpass (1.30 PM).

The study period is the warm months of the year (Apr–Sep) between 2003 and 2013. In these months most of the region are cloud-free, dust storms are most frequent in West Asia (Boloorani et al., 2014) and all datasets are available. The two partitions of 2003–2010 and 2011–2013 are chosen as training and test sets, respectively.

2.2. Feature selection criteria (FSC)

Machine learning algorithms generally optimize the combination of potential predictors to get the best statistical estimations of a particular predictand, in our case DB AOD. The process of feature selection aims to identify the optimal set of predictors, from a (much) larger set of potential predictors, to be used as a same set of input variables for the development of all MLAs. Regardless of which MLAs are used for prediction, feature selection is a critical step, which has a direct effect on the level of accuracy and, at the same time, complexity of the model. It also regulates the generalizability/overfitting of MLAs. Filter-type methods are the most commonly used FSC which estimate the importance of explanatory variables regardless of the model performance. These methods are computationally effective and robust against overfitting. PCC and Mutual Information (MI) are two well-known filter-type methods. Correlation ranking simply considers the linear relationship between each predictor X_j and response variable Y:

$$\mathrm{PCC}(j) = \frac{\mathrm{cov}(X_j,\mathrm{Y})}{\sqrt{\mathrm{var}(X_j) \times \mathrm{var}(\mathrm{Y})}} \tag{2}$$

where *cov* and *var* represent the covariance and variance, respectively. MI is the measure of a relationship between two random variables that are sampled simultaneously (Paninski, 2003). In other words, MI measures how much information random variables have about each other (Eq. (3)). Zero MI means predictor $X_j$and response variable Y are independent whereas high MI indicates that there is a large amount of information shared. The MI of two continuous variables X_j and Y whose joint distribution is defined by P (x_j, y) is as follows;

$$\mathrm{MI}(\mathrm{X_j},\mathrm{Y}) = \int_{\mathrm{Y}} \int_{\mathrm{X_j}} \mathrm{P}(\mathrm{x_j},\mathrm{y}) \log \frac{\mathrm{P}(\mathrm{x_j},\mathrm{y})}{\mathrm{P}(\mathrm{x_j})\mathrm{P}(\mathrm{y})} \mathrm{dx_j dy} \tag{3}$$

$P(x_j)$and$P(y)$ are the marginal distributions of X_j and Y.

2.3. Machine learning algorithms

In this subsection, we provide the concise explanation of five MLAs used in this study including Random Forest (RF) and Multivariate Adaptive Regression Splines (MARS), Support Vector Machines (SVMs), Artificial Neural Network (ANN), and MLR. For the sake of brevity, the detailed descriptions of each algorithm is provided in the Appendix.

Breiman (2001) proposed RF as an ensemble of decision trees algorithm. The latter is to increase the predictability of output by splitting observations (root nodes) into new classes (sub-nodes). It evaluates the splits of all variables at each node and, then, it selects that split (variable) which results in less inhomogeneity. This is repeated recursively until data has been categorized into homogenous groups (Appendix, Eq. (A1)). However, the prediction of response value through a single tree mostly yields high bias and/or variance (over-fitting). To deal with this problem, RF constructs numerous trees using bootstrap samples (Efron and Tibshirani, 1994) of the data and random selection of a subset of predictors (not all predictors) for making sub-nodes.

MARS is a nonparametric statistical method that makes no assumptions about the functional relationship of the variables. In order to improve the prediction of a non-linear system, MARS splits the linear relationship between explanatory and response variables into separate piecewise linear segments (splines) of differing gradients (Zhang and Goh, 2016). This process continues until the model reaches a predetermined error level or/and a threshold number of splines. This usually results in a purposely complicated and overfitted model. Due to this, the backward phase is used to improve the model by pruning (removing) the less significant terms. At the end of the backward phase, the model with the lowest Generalized Cross-Validation value (GCV) (Appendix, Eq. (A3)) is selected as the final model.

SVMs, proposed by Vapnik (1995), was firstly introduced for classification and later also for regression (Smola and Schölkopf, 1998). An SVM uses a device called kernel, such as the Gaussian and polynomial kernels, to map data into a high-dimensional feature space in which the nonlinear problem becomes linearly separable (Zhang et al., 2004). The SVMs follows the same principles for classification and regression. It searches for the optimal hyperplanes which maximize the margin between classes of data and minimize unexpected errors.

Neural networks are multivariate nonlinear models. They consist of processing neurons nested in three layers including an input layer, one or more hidden layers, and an output layer (Konate et al., 2015). The number of input neurons is equal to the number of independent variables while the output neuron(s) represent the dependent variable(s).

Table 1
The list of datasets used in this study. Please note that temporal coverage of datasets belongs to the time of writing this paper.

Dataset	Spatial/temporal resolution	Temporal coverage	Data source
DB AOD	10 km/daily	2003 to present	MODIS-Aqua
wind components, vertical velocity, soil temperature, albedo	~79 km/6-hourly	1979 to preset	ECMWF ERA-Interim
soil moisture	0.25 degree/daily	1979 to present	ESA-CCI
Precipitation	0.5 degree /monthly	1901–2013	GPCC
NDVI	1/12 degrees/bi-monthly	1981–2015	GIMMS
SPEI	0.5 degree/monthly	1901–2015	Vicente-Serrano et al. (2010)
SF	10 km/time-invariant	–	Nabavi et al. (2016)

The neurons are interconnected by connection strengths called weights which are updated, during training process, in such a way that minimizes the cost function (for example MSE).

Linear models are the most simple and commonly used machine learning algorithms. They try to find a linear relationship, if any, between one or more predictors and a response variable by fitting a linear equation to observed data. The coefficients are estimated by minimizing the sum of the squares.

2.4. MLA configuration

Although one may discuss that MLAs are designed to make a machine system that automatically builds models from data without human involvement, the best performance of MLAs only occurs when their optimal parameters are obtained through tuning. In this study, the optimal parameters are tuned based on the automated evaluation of prediction errors (RMSE) resulting from K-fold cross-validation. This validation method divides randomly the data into K roughly equal parts (here 10 parts). At each loop iteration, one subsample of the k subsamples is retained as the validation subset for evaluating the model, and the $k-1$ subsamples are used for training the model. Optimal parameters are those with the least averaged prediction error (for example averaged RMSE of 10 folds) resulted from repeating validation process for k times. Both the tuning and training of MLAs are done by using the Caret package (Kuhn, 2008), implemented in R. In Table 2 tuned parameters of MLAs are presented. Names of parameters are replicated to be easily found in the help page.

2.5. Deterministic weather prediction models

DMs are the present standard tool for predicting important processes of aerosol life cycle. Physical or empirical laws are employed to parameterize those mostly sub-grid scale processes. In most cases aerosol models are deterministic, i.e. no stochastic forcing is present in the forecast equations. The initial state of the aerosol forecasts is computed with different degrees of sophistication as will be described below. From this state short term forecasts are performed to predict aerosol concentrations and AOD a few days ahead.

In this paper, we use hindcasts from WRF-chem (Grell et al., 2005; Fast et al., 2006; Skamarock et al., 2008) and analyses from MACC to get fields of monthly mean AOD:

- WRF-chem can be run with various aerosol species. It uses Goddard Chemistry Aerosol Radiation and Transport (GOCART) aerosol background fields as the initial and boundary conditions. In the present paper, WRF-chem 3.6.1 is executed for the study period using the configuration explained by Nabavi et al. (2017). It is run over the domain shown in Fig. 2 with the GOCART dust scheme modified to use modified WASF as source function. ERA-Interim analyses are used as lateral boundary conditions for forecasts of the warm months (Apr–Sep). Newtonian nudging toward ERA-Interim, with nudging time of 6 h, is used to keep the forecasts close to observed atmospheric state. This is a well-proven method especially for hindcasts that allows avoiding explicit and expensive analysis steps (Deng et al., 2007). Soil moisture is provided from National Centers for Environmental Prediction (NCEP) Final (FNL) Operational Global Analysis data and precipitation is a standard forecast product. Total column aerosol concentration is the primary forecast variable which is converted into AOD at 550 nm using a radiative transfer code (a so-called observation operator, (Chin et al., 2002). AOD is then averaged to yield a monthly mean.
- The MACC (2003–2012) project and its successor the Copernicus Atmospheric Monitoring Service (CAMS) (July 2012 to present) have been and are devoted to air quality monitoring. Many chemical species but also five aerosol species, including mineral dust, are monitored and forecasted. The basic meteorological forecast system used for data assimilation is the Integrated Forecasting System (IFS) of ECMWF. The formulation of the aerosol model has remained largely similar during the transition from MACC to CAMS, based on Morcrette et al. (2009). It is a global forecasting system and as such has limited spatial resolution. Contrary to WRF-chem as used here, a forecast model is used in MACC only to assimilate satellite observations (Benedetti et al., 2009) such that an optimal state of conventional meteorological as well as chemical and aerosol species is found. The MACC aerosol product is thus not a forecast but a reanalysis where MODIS AOD have been assimilated, which certainly strengthens this product (Cuevas et al., 2015). MACCs primary focus is on atmospheric chemical species, not so much mineral dust. As for WRF-Chem AOD is not part of the model state but the total column aerosol concentration analyses are converted into AOD using an observation operator and then are averaged to yield a monthly mean. Since AOD observations have entered the MACC product one would expect that it matches AOD measured by satellites best. The MACC products are freely available from apps.ecmwf.int.

Nabavi et al (2017) also included results from the Dust REgional Atmospheric Model (DREAM) modeling system (Basart et al., 2012), which yielded skill scores similar to MACC. In this paper, which is more stringent in the choice of predictands, we decided not to include DREAM since it provides only dust optical depth (DOD) but not AOD as output field. MACC AOD is available from 2002 to 2012. Therefore, this dataset is six months shorter than WRF-chem and MLA predictions, which are available during the whole study period.

3. Results and discussion

In this section, we first report about the determination of predictors of DB AOD using the feature selection criteria described in Subsection 2.2, and then about the prediction of DB AOD by MLAs and DMs (3.2).

3.1. Potential predictors of DB AOD

In spite of some differences between the importance levels determined by MI and PCC, for example, the assigned importance of precipitation, they give an approximately similar image from the influential features for AOD prediction. For simplicity, importance degrees are averaged for each feature shown by blue bars (Fig. 6). The overall outcome has respectively assigned the least and highest importance to SF and SPEI. In order to find out the optimum number of features for the final (pruned) model, we have performed a preliminary run in which all five MLAs were iteratively trained and validated while variables are eliminated one by one at each iteration.

Two evaluation metrics including averaged Spearman Correlation Coefficient (SCC) and averaged RMSE between MLA predictions and MODIS observations are used for the examination of MLA performance. According to Fig. 7, the elimination of SF has caused the highest

Table 2
Optimal parameters of MLAs as result of tuning. Parameter mtry is the number of variables randomly sampled as candidates at each split, ntree is the number of trees, nprune is the maximum number of terms (including intercept) in the pruned model, degree is the maximum degree of interaction, sigma is the width of radial kernel (also known as smoothing parameter), C is the constant of the regularization term in the Lagrange formulation, size is the number of units in the hidden layer, and the decay is parameter for weight decay (Kuhn, 2008).

MLA	Parameter(s)	RMSE
RF	mtry = 3, ntree = 100	0.16
MARS	nprune = 16, degree = 1	0.315
SVM	sigma = 0.237, C = 1	0.273
ANN	size = 15, decay = 0.01	0.059

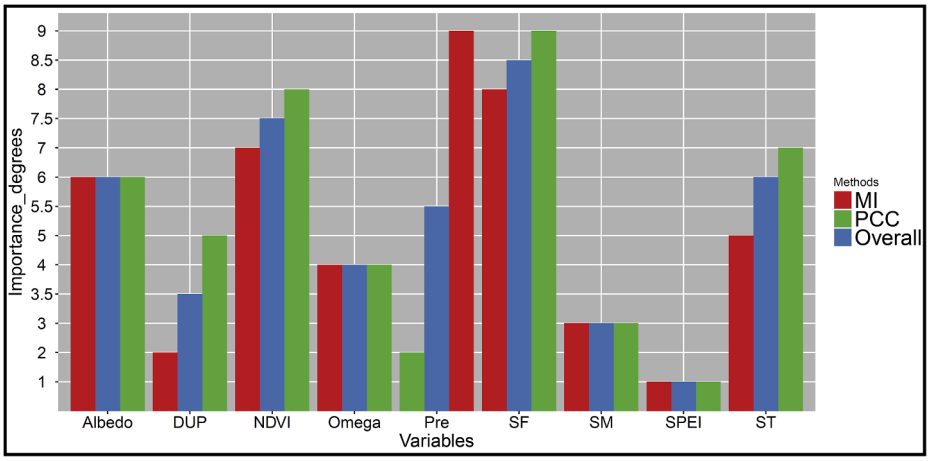


Fig. 6. The importance of variables determined by MI and PCC. The overall outcome is prepared by the average of variable importance assigned by mentioned methods. High numbers mean higher importance.

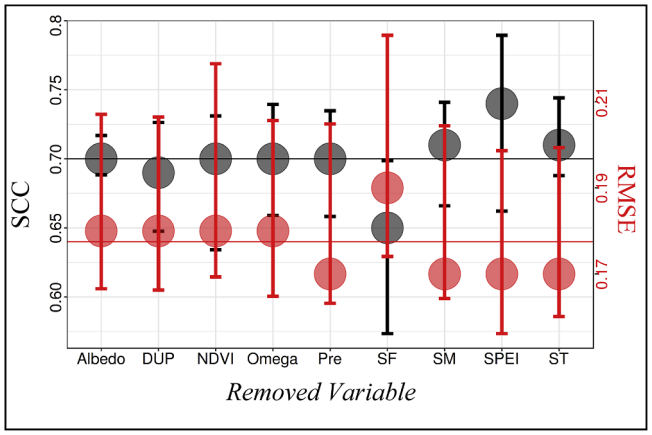


Fig. 7. The averaged SCC (black filled circle) and RMSE (red filled circle) between MLA predictions and observations during train period. Error bars are limited between the maximum and minimum values of SCC (black bars) and RMSE (red bars) resulted from a separate comparison of five MLA predictions and observations. The average of averaged SCC and RMSE are shown by the black and red lines, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

decrease in SCC. NDVI is found to be the most important variables among time-varying inputs. This again implies the importance of dust source-related variables in the estimation of AOD. In fact, it seems that those source points which are omitted in time-invariant variable SF are considered by NDVI. Kim et al. (2013) also found that NDVI represents seasonal changes in the extent of dust sources and improves the prediction of dust emission. Conversely, the averaged SCC increased when SPEI and SM are eliminated. Considering RMSE as a measure of prediction error leads to the same conclusion. It increases most significantly if SF is removed but actually decreases when SPEI and SM are pruned. Although the elimination of ST also improves the average performance of MLAs, it is not removed for the sake of the objectiveness of the feature selection process. Conclusively, seven features including Albedo, DUP, Omega, NDVI, ST, precipitation, and SF are selected to train MLAs during 2003 and 2010.

3.2. The prediction of AOD by MLAs and DMs

Following the notion of Efron and Hastie (2016), we have so far described mostly the *algorithms* (both MLAs and DMs) that provide AOD estimates. Now comes the *inference* part, which intends to compare the performance of the algorithms in a statistically sound way. This means to compare the same departure statistics for the same period.

While there are data from 2002 to 2013, the performance is discussed for the test period (2011–2013), since the rest of the data has been used for training the MLAs. As the first step, the temporal averages of predictions and MODIS observations (11-AA) were compared visually. This showed that MLAs (for example MARS, 8-A) have underestimated DB AOD especially over dust sources and paths in Iraq, characterized by Nabavi et al. (2016), that leads to the decrease of prediction accuracy in central Iraq. However, they still have yielded better performance than two DMs (8-C). All MLAs, except ANN, have produced higher correlation coefficients, less centered RMSE, and more similar standard deviations to that of observations (0.23) than DMs. Besides the fact that MLA underestimation may be partly attributed to non-linearities between the abundance of aerosols and DB AOD over the study area (Sayer et al., 2013; Nabavi et al., 2016) and/or retrieval error (Albayrak et al., 2013), three main reasons could cause underpredicting high extreme values. The lack of variables which thoroughly explain the variance of DB AOD can be seen as the primary reason. In other words, selected features for training MLAs and numerical solutions of DMs, especially WRF-chem in this case, do not perfectly represent the life cycle of dust or other types of aerosols. This requires further studies for the improvement of our knowledge about the mechanism of aerosol formation which is beyond the scope of this paper. Secondly, the scarcity of extremely high values results in the poor performance of both MLAs (Zhang et al., 2015) and DMs (Kumar et al., 2014) in the prediction of extraordinary cases. In fact, models are trained or formulated so that they yield the least overall bias with observations. That is, achieving the highest level of prediction accuracy does not necessarily mean that models accurately estimate the whole range of measured quantities but it means they are successful in the simulation of more frequent cases, inevitably at the cost of forecast sharpness (Wilks, 2011). For the prediction of phenomena which have a positively skewed frequency distribution, like dust abundance, this problem is known to be particularly serious. Using Synthetic Minority Over-sampling Technique (SMOTE), Torgo et al. (2013) tried to deal with this issue by under-sampling of frequent cases (irrelevant cases) and over-sampling of rare quantities (relevant cases) used for training MLAs. However, the extent of over-sampling/under-sampling requires researcher intervention and it changes case by case. Therefore, we decided not to manipulate the original distribution and to concede part of uncertainties related to the unbalanced distribution of DB AOD.

The third reason of differences between predicted and observed peaks is that a significant portion of AOD values at each pixel, especially over surrounding areas of dust sources, is related to the number of dust particles advected from upstream sources. Using WRF-chem for dust prediction, Nabavi et al. (2017) showed that shortcomings in the

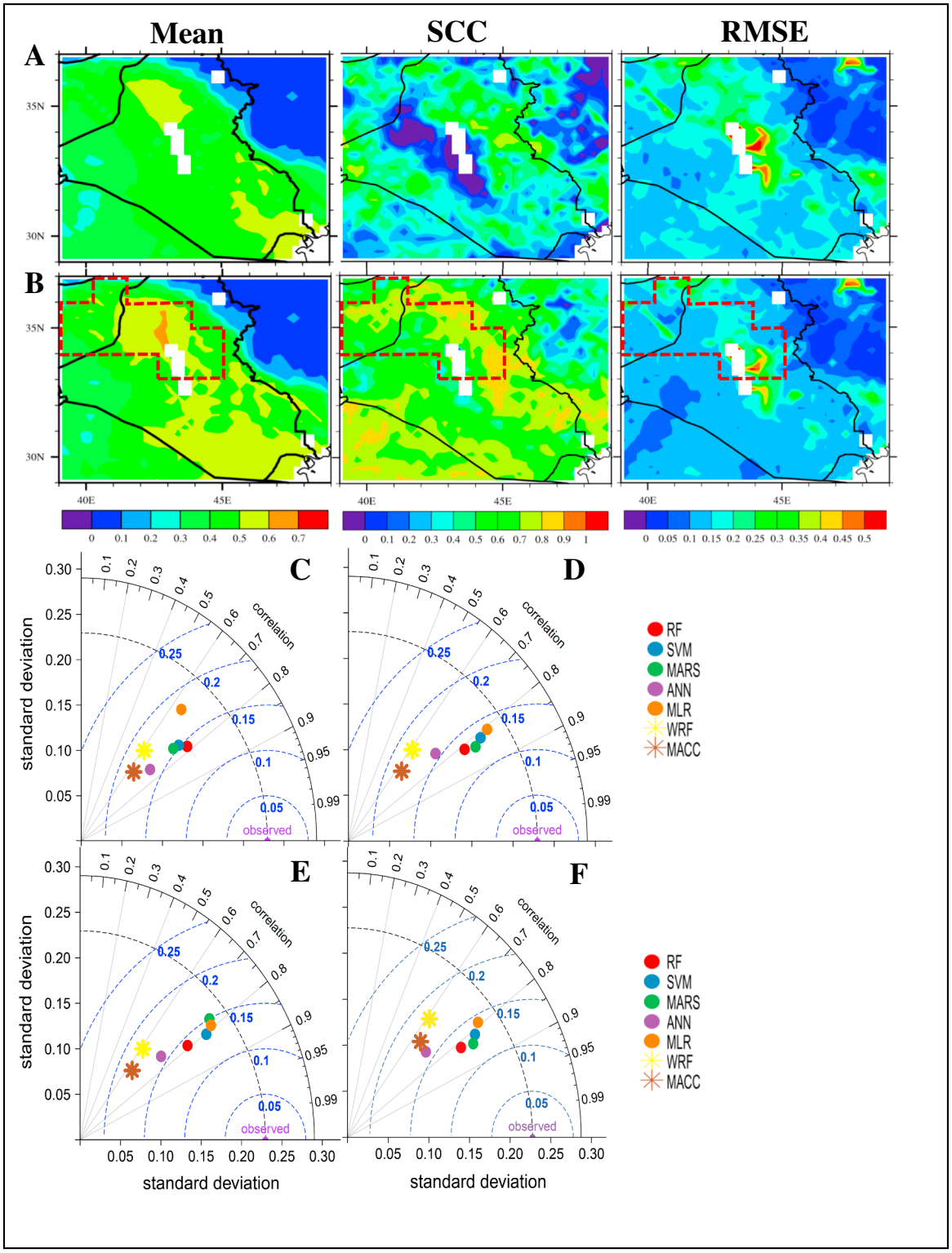


Fig. 8. A: the temporal statistics of MARS simulations before the inclusion of six area-averaged predictors during test period. B: same as A but after the inclusion of six area-averaged predictors. The red polygon shows the boundaries of dust sources in the east of Syria and northwest of Iraq where area averages are computed. The temporal statistics of other MLAs and DMs acquired from the second run are presented in Fig. 11. The spatial statistics, calculated between prediction and observation vectors within the study area, including centered RMSE (dashed blue line), standard deviation (dashed black line) and SCC correlation (solid black line) are presented in C (before the inclusion of area-averaged predictors) and D (after the inclusion of area-averaged predictors). E: same as D but after replacing the modified WASF with the original WASF for MLAs. F: same as D but after performing 4x4 point gridbox averaging for all MLA predictors except SF and NDVI. Note that small changes in the statistics of DMs is only because there are fewer data gaps after averaging. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

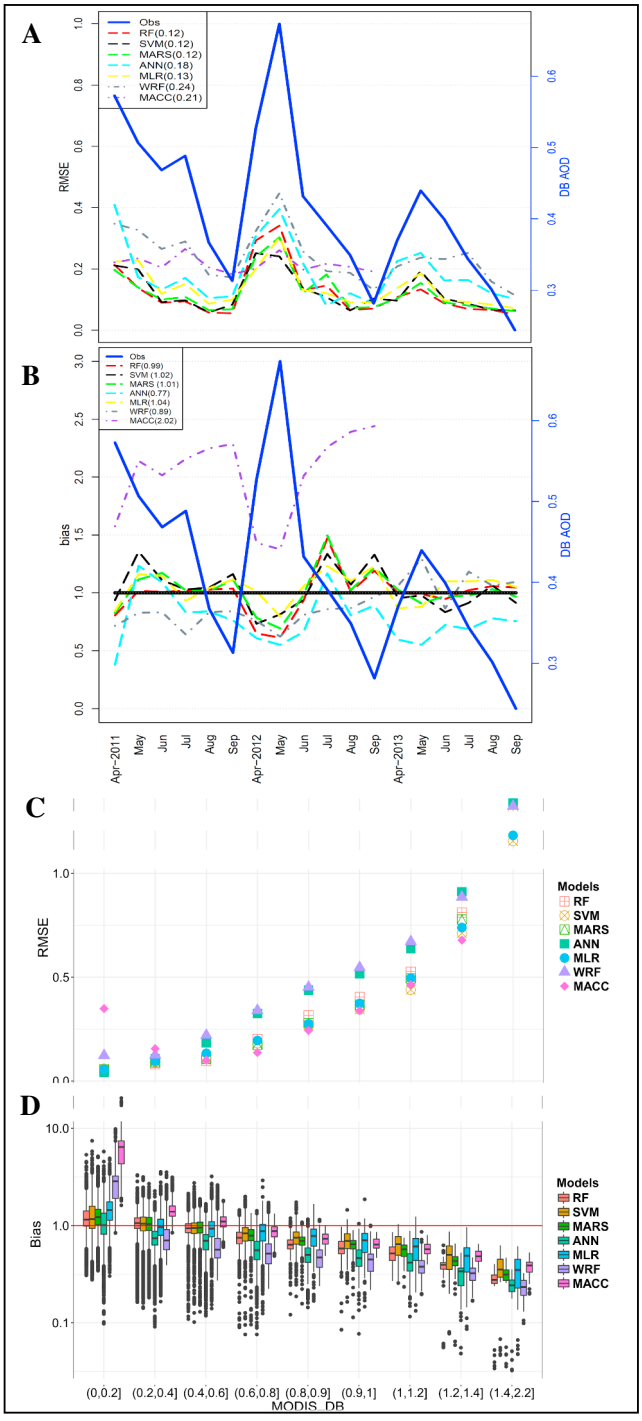


Fig. 9. A and B are showing RMSE and bias over the test period. Blue lines in A and B are area-averaged monthly observed DB AOD. Panels C and D show RMSE and bias (the ratios of observations and predictions shown on a logarithmic scale) for different classes of DB AOD. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

prediction of dust transport and deposition result in significant overestimation/underestimation of dust concentration over affected areas. They linked a great deal of AOD over West Asia to dust plumes originating from source points in the northwest of Iraq, not to the local potential of dust emission in downstream regions. Therefore, we have conducted a sensitivity experiment taking the area-averaged observed DB AOD over the main dust source of the region (red polygon in Fig. 8-B) in the northwest of Iraq and the east of Syria (Nabavi et al., 2016) as additional input for MLAs throughout the study area. This input can be also be seen as alternatives of transportation and deposition schemes, used in DMs, to regulate the amount of advected dust. As expected, it caused a significant improvement in the prediction of extreme values by MLAs (not shown here). In order to keep the MLAs independent of the response variable, this predictor (area-averaged DB AOD) is replaced by the area average of six predictors (except SF) over the mentioned region and, then, MLAs are retrained using 13 predictors (7 normal datasets + 6 corresponding area-averaged datasets). The effect of this modification on MARS performance, for instance, and other MLAs can be seen in Fig. 8-B and 8-D, respectively. The amendment of the advection component has significantly improved the MLA performance in the eastern half of the study area which was repeatedly affected by upstream dust sources. Interestingly, all MLAs have yielded high agreement with observations in the second run as the least and highest SCCs are 0.76 and 0.83 belonging to ANN and MARS, respectively. In contrast, the SCC of DM predictions, which are produced under a same setting in both runs, at most reaches 0.65 for MACC simulations. The analogy of standard deviation of MLA predictions (ranging from 0.14 to 0.2) and of observations (0.23) in the second run shows the importance of the newly added predictors in the more accurate simulation of DB AOD amplitude, compared to MACC and WRF-chem (0.06 and 0.07). Similarly, the centered RMSE between MLA predictions and observations decrease to less than 0.15 in the second run whereas DMs have yielded centered RMSE of 0.18. It should be noted here that we did not consider the effect of dust advection from northern Africa and Saudi Arabia. Shao et al. (2011) showed that only 5 percent of Saharan dust reaches to Middle East. Due to dominant atmospheric circulations of Middle East, in very exceptional cases dust particles formed in Saudi Arabia take northern direction. Arabian dust storms are normally transported to the Arabian Sea and Indian peninsula (Shao et al., 2011; Hamidi et al., 2013). We have even tested area averages of predictors over the southeast of Iraq (northeast of Saudi Arabia), but it caused no significant improvement in predictions. In addition to the sensitivity of predictions to the dust advection, we have also examined the effect of original (Fig. 5-A) and modified (Fig. 5-B) WASF specification on prediction accuracy since they are designated through subjective thresholds. In fact, we examined if modifications on WASF could improve predictions. Although the results summary of runs with original WASF (8-E) shows that there is only a small difference between these two runs, it could be also seen that the modified WASF had the positive effect on the prediction accuracy of MLAs as already shown in Fig. 8-D.

As discussed in the Subsection 2.1, we have used only reanalysis or quality-assured datasets as predictors for the hindcast of monthly AOD during 2011–2013. In cases where forecast is of interest, the predictors must be taken from seasonal predictions or from climate projections of general circulation models (GCMs). Apart from existing uncertainties of GCMs that may degrade the performance of MLAs and/or DMs, the coarser spatial resolution of GCM forecasts (Miao et al., 2014), may put the MLA capability for performing high-resolution forecast into doubt. We therefore repeated the experiments using aggregated means over 4×4 pixels ($\sim$100 km) of most predictors (except SF and NDVI, whose detailed spatial information is expected to vary only slowly even in a climate change scenario) as input for MLAs. Fig. 8-F shows a tangible degradation in the performance of ANN and MLR, but the predictive skill of other MLAs remained almost unchanged, compared to Fig. 8-D. Thus one can expect good performance of MLAs even with the use of GCM simulations for the temporally strongly varying input. Because MLA performance did not change significantly during the third (using original WASF) and fourth (using inputs with lower spatial resolution) runs, only the performance of MLAs in the second run (using area-averaged inputs as complementary features for MLAs) is discussed as follows.

In order to examine the performance of models in the course of time, monthly RMSE and area-averaged bias (prediction/observation) are plotted in relation to corresponding observed DB AOD (blue line)

(Fig. 9-A and B) over the test period. The noticeable point is that RMSE did not follow any positive trend by approaching the end of test period. This is important since it would be a sign of overfitting over the training period, if the level of prediction errors increased with distance from the start point of the test period. This is not the case in our examinations. The juxtaposition of DB AOD with simulations show that prediction error increase when DB AOD reach a peak, indicating underprediction of extremes. Apart from this period, all MLAs except ANN, have no significant bias against observations. In contrast, WRF-chem has left two different periods of underestimation and overestimation before and after June 2012, respectively. MACC has yielded very significant overestimation during the entire test period (until the last available data point in Sep 2012). Further examinations show that this model has left high prediction error (Fig. 9-C) and bias (Fig. 9-D) mainly in low

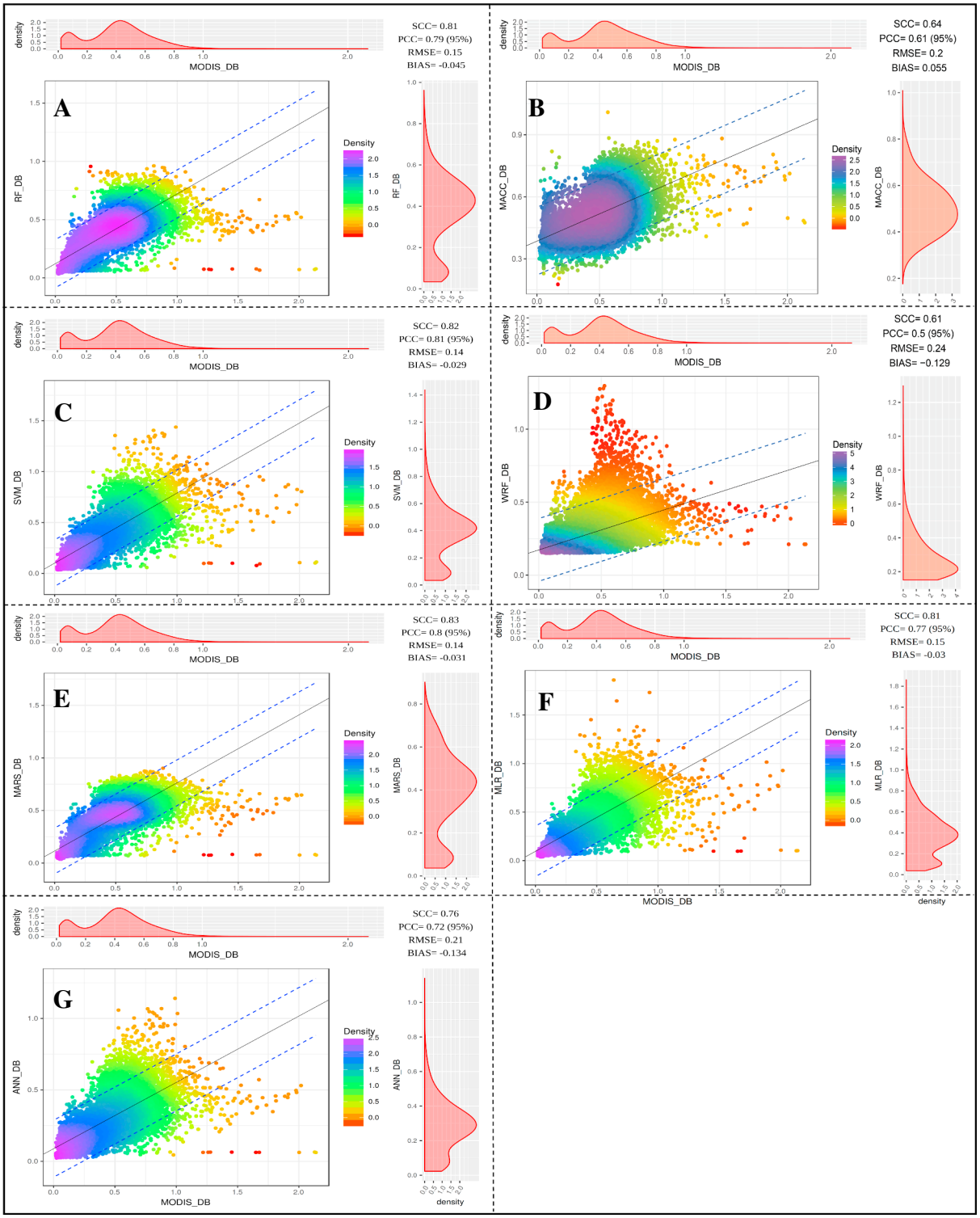


Fig. 10. scatter plots between MODIS DB AOD values (all 25x25km pixels of the study area, all months of test period) and predictions of MLAs; RF (A), SVM (C), MARS (E), ANN (G) and MLR (F) and DMs; MACC (B) and WRF-chem (D). Colors represent the estimate of the density function. The histograms of observations and predictions are at top and right margins of each plot, respectively.

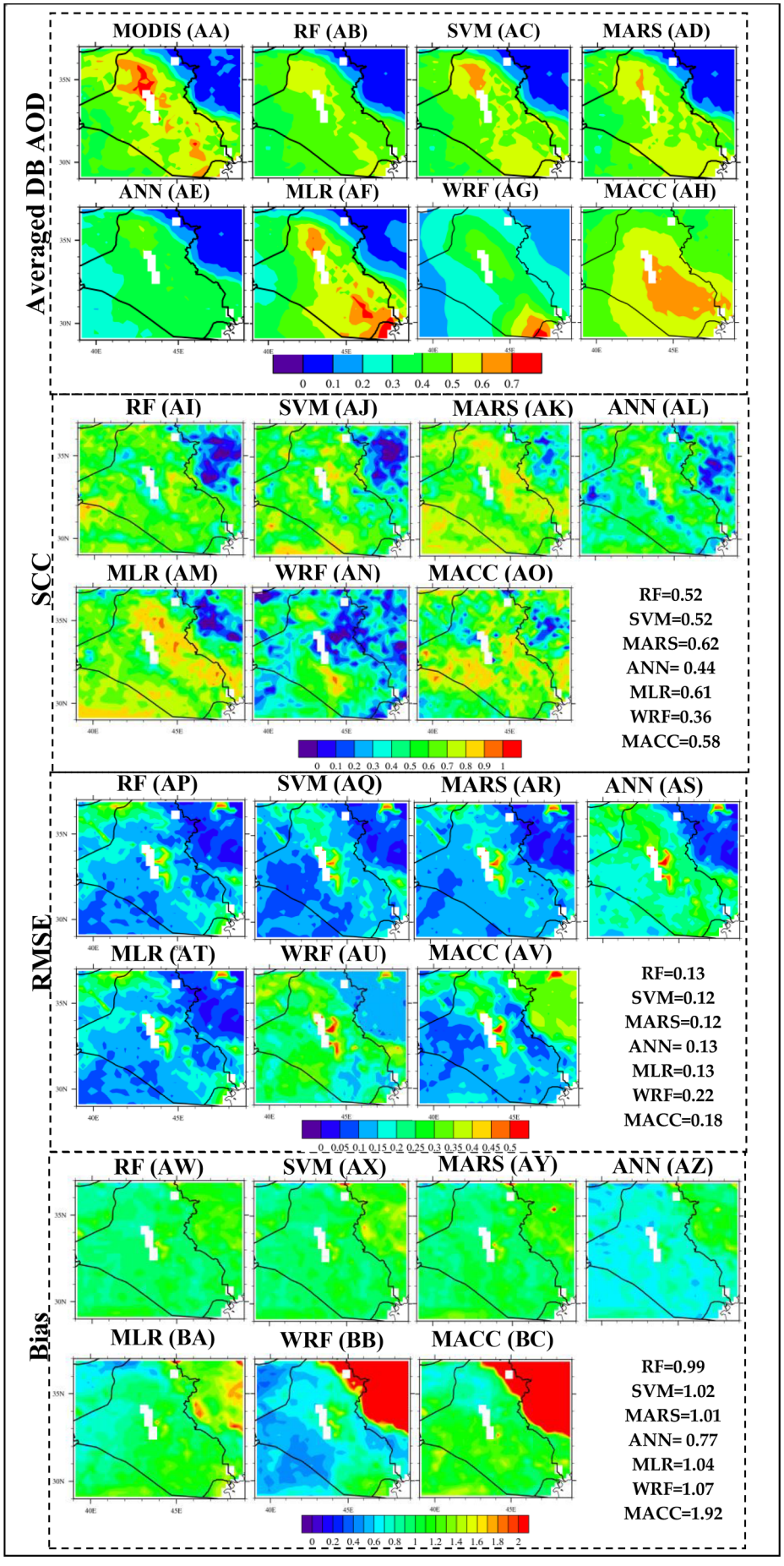
values of DB AOD. WRF-chem has also yielded high bias for quantities of DB AOD < 0.2, but because it is cancelled by significant underestimation of higher values, unlike MACC, it is not reflected in the spatially-averaged bias in Fig. 9-B. Besides this, all DMs and MLAs, especially WRF-chem and ANN, have yielded higher RMSE and significant underestimation by approaching higher values of DB AOD, again indicating having difficulties for the prediction of intense dust storms. The point-to-point comparison of simulations and observations (Fig. 10) also shows the better performance of MLAs as their predictions are mainly congested around regression lines, which can be interpreted as high correlation between observations and predictions. The lower bias of MLAs, except ANN, shown in Fig. 9-B, is also represented by the proximity of their regression and identity lines. The juxtaposition of observed and predicted histograms indicates that the frequency distribution of MLA predictions, especially SVM, is very similar to that of observations (Fig. 10-A, C, E, F, and G). In contrast, the larger bias of DM predictions and their low agreement with observations can be diagnosed from the significant tilt between the regression line and the identity line and higher spread of simulations, respectively (Fig. 10-B and D). The significant bias of simulated AOD by MACC against observations can be also seen in Fig. 10-B where simulated AOD start from 0.3. In the following, the spatial analysis of the discussed statistics characterizes more clearly the performance of DMs and MLAs over the study area and it also discloses the reason of overestimation of low DB AOD by DMs.

Fig. 11 presents the temporal average of observed and predicted DB AOD (Fig. 11AA–AH), and temporal SCC (Fig. 11AI–AO), RMSE (Fig. 11AP–AV), and bias (Fig. 11AW–BC), during the test period. It should be noted that three first statistics have been already presented for MARS in Fig. 8-B to show the effect of feeding advection component into MLAs. The distribution of observed DB AOD clearly depicts a dust hotspot in northwest of Iraq and a dust path with northwest-southeast direction. This pattern is almost reflected in the simulations of all models, which follows the prevailing wind of the region during summertime, called Shamal (Yu et al., 2016; Francis et al., 2017). However, DMs fail to estimate accurately the absolute quantities of DB AOD particularly over main dust source of the study area (northwest of Iraq). In addition, MACC and, to some extent, WRF-chem do not resolve Zagros Mountains, in western Iran, as a barrier against transportation of fine dust particles from plains in Iraq to the east of the study area which is represented by low DB AOD in observations and simulations of other models. The inter-comparison of MLAs shows that ANN is the only algorithm which could not well simulate AOD quantities, while other MLAs have provided a realistic distribution of DB AOD over the study area. It should be also noted that the agreement of MLR and MARS predictions with observations over the dust path is higher than other algorithms. Although RMSEs increase over dusty areas of Iraq, because of underestimation high dust concentrations in all algorithms, MLAs have generally produced much less RMSE than DMs. Similarly, the bias of MLA predictions, except ANN, is mostly around 1 (no bias) whereas DMs have yielded very high overestimation over Western Iran, as discussed above, and moderate underestimation over dust sources and dust path. In fact, dust schemes of DMs, used for the estimation of dust emission, deposition, and transportation, underestimated emitted dust and overestimated dust advection into Iran, delineated with high RMSE and bias in Fig. 11-BB and 11-BC. The analysis of SCC shows that the eastern half of study area, more or less, has received lower SCC in all predictions. As discussed before, these uncertainties can be attributed to suboptimal dust deposition by DMs and lower capability of MLA predictors in representing dust transportation and deposition (affecting the west of Iran) than representing dust emission (over Iraq). However, the examination of this claim requires access to transportation and deposition measurements over Iran. In addition, selected features may not be completely able to explain the life cycle of other types of aerosols such as biomass burning aerosols resulted from infrequent wildfires of Zagros Mountains forest steppe and anthropogenic aerosols formed in the major cities of western Iran. However, discussed findings showed that selected features for MLA setup could successfully represent the main environmental drivers of AOD (dust cycle) in the study area. Klingmüller et al. (2016) also found that AOD variations in West Asia are more relevant to changes in dust quantity rather than any other types of aerosols. Comparing to this study, we also managed to increase the spatial and temporal resolution of AOD predictions from 2 degrees to 0.25 degree and from annual to monthly scale, respectively.

4. Conclusions

In this study we demonstrated that it is possible to produce and validate AOD (dust) prediction even over regions with extremely sparse surface measurements, substituting them with satellite products with a resolution of 25 × 25 km for both input and validation. The second purpose of the paper was to demonstrate the applicability of Machine Learning Algorithms (MLAs) for dust prediction on monthly time scales. Satellite measurements and reanalysis data, simultaneous with AOD, were used to provide input variables, such as Normalized Difference Vegetation Index (NDVI) and soil moisture (SM) and temperature (ST), and the predictand (MODIS Deep Blue (DB) AOD) for MLA setup. Assuming dust as the dominant type of aerosols in the region, eight monthly datasets including ST, SM, The Standardized Precipitation-Evapotranspiration Index (SPEI), albedo, NDVI, Precipitation, Omega at 850 hPa, and Dust Uplift Potential (DUP) at 10 m were used as potential predictors. In addition, we made use of a time-invariant variable dust source function (SF) to regulate the potential of dust emission at each 25 × 25 km pixel. Using filter-type methods, SF was determined as the most important factor of AOD estimation. NDVI was found to be the most important variable among time-varying features selected for AOD prediction. It seems that seasonal variations of dust sources omitted in SF could be fed by using NDVI into MLAs. On the contrary, SPEI and SM were designated as the least important variables and they were eliminated. Because advected dust has a large influence on AOD of affected areas, the area average of time-variant variables over main dust source of study area, the east of Syria and northwest of Iraq, were taken as complementary predictors. Five MLAs, including Multiple Linear Regression (MLR), Random Forest (RF), Support Vector Machines (SVM), Multivariate Adaptive Regression Splines (MARS), and Artificial Neural Network (ANN) were trained and compared with two Deterministic weather prediction Models (DMs), including WRF-chem and MACC. In a nutshell, predictions of AOD by MLAs, especially SVM and MARS, outperformed DMs on the time scales considered in the present work. The analysis of statistics shows that MLAs have high agreement with observations and they, except ANN, yielded the smallest prediction errors over dust sources. Since the Source Function gained high weights in the feature selection it is quite likely that its careful specification as WASF has been a big advantage for MLAs. It helped their predictions to be more accurate in the representation of dust source distribution in West Asia than those of DMs, which mostly put less emphasis on specification of the SF. Although both MLAs and DMs were relatively successful in the simulation of general variations of aerosol concentration, they still all underestimated major DB AOD peaks. The rough resolution of used datasets, the scarcity of extreme values, and the omission of some unknown influential predictors representing the life cycle of dust and other aerosols are likely the main reasons. In addition, DMs, especially MACC, have failed to resolve the effect of Zagros Mountains as a natural barrier preventing the transportation of dust to the west of Iran which is reflected as a very high overestimation of low DB AOD. This is the main reason why the overall performance scores for MACC are relatively low.

We emphasize that the better performance of MLAs is demonstrated only on the space and time scales considered here. For short term forecasts on the daily scale with good knowledge of the initial atmospheric state and of initial surface properties such as soil moisture, DMs may well yield better performance than MLAs. This is because the



(caption on next page)

Fig. 11. Maps of temporal averages of observed (AA) and predicted DB AOD (AB-AH). Averaging period: warm season (April-September) of years 2011–2013. Following couple rows are maps of temporal statistics (SCC, RMSE, and Bias. Values in white box are the area average of corresponding statistics. Acronyms are explained in the text.

physical laws for short term predictions are well known and well implemented in such models. It should be also taken into account that the output variable (AOD) is the measure of optical depth induced by different types of aerosols (not only dust). That is, the part of error by DMs may be caused by failures in their representation of life cycle of aerosols other than dust. In addition, because this study has done the hindcast (not forecast) of AOD, it is possible that MLAs are better predictors of AOD than DMs, when the predictors are known simultaneously. We also put the caveat here that the validation has been done with AOD, which is not a state variable of DMs but has to be calculated with an observation operator. It is likely that the advantage of MLAs would be smaller if validation were done against in situ measurements of aerosol concentrations or even spectra. In the study area , those do barely exist and typically do not separate between mineral and other types of aerosol, particularly the black carbon in urban areas, which makes it difficult to prove this conjecture. Finally yet importantly, Nabavi et al. (2017) have constrained the amount of emitted dust through the adjustment of dust emission factor comparing satellite and WRF-chem AOD. It led to the preparation of a local dust source map called WASF used here as an input for all MLAs and as the emission factor for WRF-chem dust scheme. But, due to the lack of observations, they did not attempt to adjust the sensitivity of dust emission to wind speed (Cakmur et al., 2006). These kinds of optimizations can be the subject of future studies which may assist DMs to outperform MLAs.

However, the performance of MLAs, even after the aggregation of inputs to a coarser spatial resolution, is quite satisfactory and promising enough to put more effort into the identification of potential missing factors of AOD prediction and to optimize MLA parameters. We also plan to compare the performance of MLAs with DMs in the prediction of AOD on a finer spatial and temporal resolution in West Asia, using the newly released satellite product AVHRR DB AOD (Sayer et al., 2017). We believe these types of studies will help to identify influential factors reducing uncertainties of aerosol predictions over sources and receptors.

Acknowledgements

This work has been financially supported by EU 7th framework program ERA-CLIM (No. 265229) and the Austrian Science Funds FWF (Projects P25260-N29). We acknowledge scientists involved in the production of the MODIS, AERONET, ECMWF, ERA-Interim, and MACC datasets used in this research work.

Conflicts of interest

The authors declare no conflict of interest.

Appendix

This appendix describes the MLAs in a bit more detail than was possible in the main text. While it is still far from exhaustive it gives more references and describes the main ideas behind the algorithms.

Random forest

Breiman (2001) proposed RF as an ensemble of decision trees algorithm. The latter is to increase the predictability of output by splitting observations (root nodes) into new classes (sub-nodes). It evaluates the splits of all available variables at each node and, then, it selects the split (variable) which results in less inhomogeneity. This is repeated recursively until data has been categorized into homogenous groups. However, the prediction of response value through a single tree mostly yields high bias and/or variance (over-fitting). To deal with this problem, RF constructs numerous trees using bootstrap (Efron and Tibshirani, 1994) samples of the data. In addition, it adds an additional layer of randomness by choosing the potential variable among a subset of predictors. In regression problems, the feature with the least residual sum of squares (RSS) of sub-nodes is selected (Eq. (A1)).

$$\mathrm{RSS} = \sum \mathrm{left}(y_i - y_L)^2 + \sum right\ (y_i - y_R)^2 \tag{A1}$$

where y_L and y_R are the mean y-value for right and left nodes, respectively. The y_i are the observed values. In the end, the predicted value is the average of observations already grouped in the leaf nodes of trees.

MARS

MARS (Multivariate Adaptive Regression Splines) is a nonparametric statistical method that makes no assumptions about the functional relationship of the variables. In order to improve the prediction of a non-linear system, MARS splits the linear relationship between explanatory and response variables into separate piecewise linear segments (splines) of differing gradients (Zhang and Goh, 2016).

$$\mathrm{f(X)} = \beta_0 + \sum_{\mathrm{m}=1}^{\mathrm{M}} \beta_{\mathrm{m}} \lambda_{\mathrm{m}}(\mathrm{X}) \tag{A2}$$

where each λ_{m} is a basis function (BF), β_{m} is a coefficient of parameter, and $\mathrm{X} = (\mathrm{X_1}, \cdots, \mathrm{X_q})$ is a matrix of q input variables. The term β_0 is a constant coefficient, estimated using the least-squares method. BF can be one spline function or the interaction of two or more BFs, depending on the order of f(X). BFs are connected through the connection/interface points called knots. During forward phase of MARS, candidate knots are placed at random positions to define a mirrored pair of BFs (Eq. (A3)).

$$\begin{aligned} \mathrm{Direct{:}Max}(0, x - c) &= \begin{cases} x - c, & \text{if } x \geq c \\ 0, & \text{otherwise} \end{cases} \\ \mathrm{Mirror{:}Max}(0, c - x) &= \begin{cases} c - x, & \text{if } c \geq x \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{A3}$$

Eq. (A3), also known as hinge function, shows how continuously the variable x is transformed using a constant "c" as a knot. At each step, the model picks up that knot and its corresponding pair of BFs (direct and mirror) which yield the minimum error. This process continues until the model reaches a predetermined error level or/and a threshold number of BFs, which usually results in a purposely complicated and overfitted model. Due to this, the backward phase is used to improve the model by pruning the less significant terms. At the end of the backward phase, the model with the lowest Generalized Cross-Validation (GCV) value is selected as the final model. The GCV criterion trades off goodness-of-fit against model complexity (Zarandi et al., 2013). For the training data with n observations, the GCV is calculated as (Eq. (A4)):

$$\mathrm{GCV} = \frac{\frac{1}{n}\sum_{i=1}^{n}[y_i - \hat{y}_i]^2}{\left[1 - \frac{M + d\times(M-1)/2}{n}\right]^2}, \quad d = \begin{cases} 3, & \text{if degree}>1 \\ 2, & \text{otherwise} \end{cases} \tag{A4}$$

Here M is the number of BFs, d is a penalty for each BF, and $\hat{y}_i$ represents the i_{th} predicted value. Thus, the numerator and denominator are, respectively, the MSE of the model and penalty for the prediction variance because of model complexity (Zhang and Goh, 2016).

Support vector machine

SVM, proposed by Vapnik (1995), have been applied successfully to both pattern recognition and more recently also to regression (Parrella, 2007) problems. For linear regression, SVM is formulated as follows (Eq. (A5)):

$$f(x) = < w, x> + b \quad w, x \in R^q, \quad b \in R \tag{A5}$$

where $< , >$ denotes the inner vector product, b is bias, w represents weight for each variable, q is the number of variables, and x represents input variables. In case of nonlinearity between response and explanatory variables, SVM kernels (Φ), like Gaussian (radial), are used to map the data into a feature space in which the problem becomes linearly separable (Eq. (A6)).

$$y = f(x) = < w, \Phi(x) > + b \tag{A6}$$

Vapnik (1995) suggested the following regularized cost function (Eq. (A7)) to estimate optimal w and b:

$$\frac{1}{2}\ ||w||^2 + \frac{C}{n}\sum_{i=1}^{n} |y_i - f(x_i)|_\varepsilon \tag{A7}$$

The factor C trades off training error against the complexity of the model. A large (small) value for C will decrease (increase) the number of training errors. However, a large C can also lead to overfitting and high variance of prediction error. The bigger (smaller) ε results in the wider (narrower) ε-insensitive zone, which is used to fit the fewer (more) support vectors and, on the other hand, more flat (overfitted) estimates (Cortes and Vapnik, 1995). In fact, both C and ε values control model complexity (but in a different way). The second term of Eq. (A7) can be defined as:

$$\left|y - f(x_i)\right|_\varepsilon = \begin{cases} 0, & if |y - f(x)| < \varepsilon \\ |y - f(x)| - \varepsilon, & \text{otherwise} \end{cases} \tag{A8}$$

The optimal weights are found by conversion of Eq. (A7) and the corresponding constraints (Eq. (A8)) to a Lagrange function by introducing a dual set of variables. By some manipulations with Lagrange multiplier and dual optimization, one obtains:

$$w = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)\Phi(x_i) \tag{A9}$$

α_i and α_i^* are Lagrange multipliers. Feeding Eq. (A9) into the Eq. (A6), SVM predictions f(X) are obtained for a test data point X. According to Mercer's condition (Burges, 1998), the inner product $\Phi(X)$ and $\Phi(x_i)$ can be defined through a kernel $K(X, x_i)$.

$$f(X) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)K(X, x_i) + b \tag{A10}$$

with

$$K(X, x_i) = < \Phi(x_i), \Phi(X) > \tag{A11}$$

Here we chose the radial (Gaussian) basis function as SVM kernel, while other choices (Polynomial and Sigmoid) would also have been possible.

Artificial Neural network (ANN)

Neural networks are multivariate nonlinear models. Feed-forward back propagation neural network (FFBP) is one of the most commonly used ANN models which consists of three layers; an input layer, one or more hidden layers, and an output layer (Konate et al., 2015). Each layer has processing units known as neurons or nodes. The neurons are interconnected by connection strengths called weights. In addition, there is a bias neuron with input 1 and corresponding weight connected to each processing unit in the hidden and output layers. The number of input neurons is equal to the number of independent variables while the output neuron(s) represent the dependent variable(s). The ANN model with one hidden layer can be written as

$$f(X) = \alpha_0 + \sum_{k=1}^{h} \alpha_k f\left(\sum_{j=1}^{q} \beta_{jk} X_j + \beta_{0k}\right) + \varepsilon \tag{A12}$$

where q is the number of input variables, h is the number of hidden neurons, β_{0k} (α_0) and β_{jk} (α_k) represent bias of the hidden layer (output layer) and weights of connections from input (hidden) neurons to hidden (output) neurons, respectively. The sigmoid transfer function (Eq. (A13)) is most commonly used forf:

$$\mathrm{sgm}(x) = \frac{1}{1 + e^{-x}} \tag{A13}$$

It is worth mentioning that Eq. (A12) assumes a linear transfer function in the output node for forecasting problems. In order to obtain the best weights for training a neural network, the back propagation algorithm uses MSE as the cost function.

Multiple linear regression

Linear models are the most simple and commonly used machine learning algorithms. They try to find a linear relationship, if any, between one or more predictors and a response variable by fitting a linear equation to observed data. MLR is used for the cases that the number of predictors is more than one variable (Eq. (A14))

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \ldots + \beta_q x_{i,q} + \varepsilon_i \tag{A14}$$

where q is the number of coefficients β and β_0 is the intercept. ε_i is the prediction error of the model. The coefficients are estimated by minimizing the sum of the squares.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aeolia.2018.10.002.

References

Abbasi, R., Moradi, M.H., Molaeezadeh, S.F., 2014. Long-term prediction of blood pressure time series using multiple fuzzy functions. In: Biomedical Engineering (ICBME), 2014 21th Iranian Conference on. IEEE, pp. 124–127.

Albayrak, A., Wei, J., Petrenko, M., Lynnes, C., Levy, R.C., 2013. Global bias adjustment for MODIS aerosol optical thickness using neural network. J. Appl. Remote Sens. 7 073514 073514.

Basart, S., Pérez, C., Nickovic, S., Cuevas, E., Baldasano, J., 2012. Development and evaluation of the BSC-DREAM8b dust regional model over Northern Africa, the Mediterranean and the Middle East. Tellus B 64, 18539.

Benedetti, A., Morcrette, J.J., Boucher, O., Dethof, A., Engelen, R., Fisher, M., Flentje, H., Huneeus, N., Jones, L., Kaiser, J., 2009. Aerosol analysis and forecast in the European centre for medium-range weather forecasts integrated forecast system: 2. Data assimilation. J. Geophys. Res. Atmos. 114.

Benoit, K., 2011. Linear regression models with logarithmic transformations. London Sch. Econ., London 22, 23–36.

Boloorani, A.D., Nabavi, S., Azizi, R., Bahrami, H., 2013. Characterization of dust storm sources in western Iran using a synthetic approach. Advances in Meteorology, Climatology and Atmospheric Physics. Springer.

Boloorani, A.D., Nabavi, S.O., Bahrami, H.A., Mirzapour, F., Kavosi, M., Abasi, E., Azizi, R., 2014. Investigation of dust storms entering Western Iran using remotely sensed data and synoptic analysis. J. Environ. Health Sci. Eng. 12, 124.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Burges, C.J., 1998. A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Disc. 2, 121–167.

Cakmur, R., Miller, R., Perlwitz, J., Geogdzhayev, I., Ginoux, P., Koch, D., Kohfeld, K., Tegen, I., Zender, C., 2006. Constraining the magnitude of the global dust cycle by minimizing the difference between a model and observations. J. Geophys. Res. Atmos. 111.

Carbonell, J.G., Michalski, R.S., Mitchell, T.M., 1983. An overview of machine learning. Mach. Learn. Springer.

Chandrashekar, G., Sahin, F., 2014. A survey on feature selection methods. Comput. Electr. Eng. 40, 16–28.

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B.N., Duncan, B.N., Martin, R.V., Logan, J.A., Higurashi, A., Nakajima, T., 2002. Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and Sun photometer measurements. J. Atmos. Sci. 59, 461–483.

Choi, Y., Ghim, Y., Holben, B., 2013. Identification of column-integrated dominant aerosols using the archive of AERONET data set. Atmos. Chem. Phys. Discuss. 26627–26656.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.

Cowie, S.M., Marsham, J.H., Knippertz, P., 2015. The importance of rare, high-wind events for dust uplift in northern Africa. Geophys. Res. Lett. 42, 8208–8215.

Cuevas, E., Basart, S., Baldasano Recio, J.M., Berjon, A., 2015. The MACC-II 2007–2008 reanalysis: atmospheric dust evaluation and characterization over northern Africa and the Middle East. Atmos. Chem. Phys. 15, 3991–4024.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137, 553–597.

Deng, A., Stauffer, D.R., Dudhia, J., Otte, T., Hunter, G.K., 2007. Update on analysis nudging FDDA in WRF-ARW. In: 8th Annual WRF User's Workshop. National Center for Atmospheric Research, Boulder, Colorado, pp. 11–15.

Dubovik, O., Holben, B., Eck, T.F., Smirnov, A., Kaufman, Y.J., King, M.D., Tanré, D., Slutsker, I., 2002. Variability of absorption and optical properties of key aerosol types observed in worldwide locations. J. Atmos. Sci. 59, 590–608.

Efron, B., Hastie, T., 2016. Computer Age Statistical Inference: Algorithms. Evidence and Data Science, Institute of Mathematical Statistics Monographs.

Efron, B., Tibshirani, R.J., 1994. An Introduction to the Bootstrap. CRC Press.

Fast, J.D., Gustafson, W.I., Easter, R.C., Zaveri, R.A., Barnard, J.C., Chapman, E.G., Grell, G.A., Peckham, S.E., 2006. Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of Houston using a fully coupled meteorology-chemistry-aerosol model. J. Geophys. Res. Atmos. 111.

Feng, C., Wang, H., Lu, N., Tu, X.M., 2013. Log transformation: application and interpretation in biomedical research. Stat. Med. 32, 230–239.

Francis, D.B.K., Flamant, C., Chaboureau, J.-P., Banks, J., Cuesta, J., Brindley, H., Oolman, L., 2017. Dust emission and transport over Iraq associated with the summer Shamal winds. Aeolian Res. 24, 15–31.

Ginoux, P., Prospero, J.M., Gill, T.E., Hsu, N.C., Zhao, M., 2012. Global-scale attribution of anthropogenic and natural dust sources and their emission rates based on MODIS Deep Blue aerosol products. Rev. Geophys. 50 (3).

Grell, G.A., Peckham, S.E., Schmitz, R., McKeen, S.A., Frost, G., Skamarock, W.C., Eder, B., 2005. Fully coupled "online" chemistry within the WRF model. Atmos. Environ. 39, 6957–6975.

Hamidi, M., Kavianpour, M.R., Shao, Y., 2013. Synoptic analysis of dust storms in the Middle East. Asia-Pac. J. Atmos. Sci. 49, 279–286.

Hempel, S., Shetty, K.D., Shekelle, P.G., Rubenstein, L.V., Danz, M.S., Johnsen, B., Dalal, S.R., 2012. Machine learning methods in systematic reviews: identifying quality improvement intervention evaluations.

Hoshyaripour, G., Brasseur, G., Andrade, M., Gavidia-Calderón, M., Bouarar, I., Ynoue, R., 2016. Prediction of ground-level ozone concentration in São Paulo, Brazil: deterministic versus statistic models. Atmos. Environ. 145, 365–375.

Hsu, N., Jeong, M.J., Bettenhausen, C., Sayer, A., Hansell, R., Seftor, C., Huang, J., Tsay, S.C., 2013. Enhanced Deep Blue aerosol retrieval algorithm: the second generation. J. Geophys. Res. Atmos. 118, 9296–9315.

Hsu, N.C., Tsay, S.-C., King, M.D., Herman, J.R., 2004. Aerosol properties over bright-reflecting source regions. IEEE Trans. Geosci. Remote Sens. 42, 557–569.

Hyer, E., Reid, J., Zhang, J., 2011. An over-land aerosol optical depth data set for data assimilation by filtering, correction, and aggregation of MODIS Collection 5 optical depth retrievals. Atmos. Meas. Tech. 4, 379–408.

Jaafari, A., Gholami, D.M., Zenner, E.K., 2017. A Bayesian modeling of wildfire probability in the Zagros Mountains, Iran. Ecol. Inf. 39, 32–44.

Kaboodvandpour, S., Amanollahi, J., Qhavami, S., Mohammadi, B., 2015. Assessing the accuracy of multiple regressions, ANFIS, and ANN models in predicting dust storm occurrences in Sanandaj, Iran. Nat. Hazards 78, 879–893.

Kim, D., Chin, M., Bian, H., Tan, Q., Brown, M.E., Zheng, T., You, R., Diehl, T., Ginoux, P., Kucsera, T., 2013. The effect of the dynamic surface bareness on dust source function, emission, and distribution. J. Geophys. Res. Atmos. 118, 871–886.

Klingmüller, K., Pozzer, A., Metzger, S., Stenchikov, G.L., Lelieveld, J., 2016. Aerosol optical depth trend over the Middle East. Atmos. Chem. Phys. 16, 5063–5073.

Konate, A.A., Pan, H., Khan, N., Yang, J.H., 2015. Generalized regression and feed-forward back propagation neural networks in modelling porosity from geophysical well logs. J. Pet. Explor. Prod. Technol. 5, 157–166.

Kotsiantis, S.B., Zaharakis, I., Pintelas, P., 2007. Supervised machine learning: A review of classification techniques.

Kuhn, M., 2008. Caret package. J. Stat. Software 28, 1–26.

Kumar, R., Barth, M., Pfister, G., Naja, M., Brasseur, G., 2014. WRF-Chem simulations of a typical pre-monsoon dust storm in northern India: influences on aerosol optical properties and radiation budget. Atmos. Chem. Phys. 14, 2431–2446.

Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. Geosci. Front. 7, 3–10.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Liu, M., Westphal, D.L., Walker, A.L., Holt, T.R., Richardson, K.A., Miller, S.D., 2007. COAMPS real-time dust storm forecasting during Operation Iraqi Freedom. Weather Forecasting 22, 192–206.

Liu, X., Huneeus, N., Schulz, M., Balkanski, Y., Griesfeller, J., Prospero, J., Kinne, S.,

Bauer, S., Boucher, O., Chin, M., 2011a. Global dust model intercomparison in AeroCom phase I. Atmos. Chem. Phys. 11, 7781.
Liu, Y., Dorigo, W.A., Parinussa, R., de Jeu, R.A., Wagner, W., McCabe, M.F., Evans, J., van Dijk, A., 2012. Trend-preserving blending of passive and active microwave soil moisture retrievals. Remote Sens. Environ. 123, 280–297.
Liu, Y.Y., Parinussa, R., Dorigo, W.A., de Jeu, R.A., Wagner, W., van Dijk, A., McCabe, M.F., Evans, J., 2011b. Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. Hydrol. Earth Syst. Sci. 15, 425.
Marticorena, B., Bergametti, G., 1995. Modeling the atmospheric dust cycle: 1. Design of a soil-derived dust emission scheme. J. Geophys. Res. Atmos. 100, 16415–16430.
Mayr, M., Vanselow, K., Samimi, C., 2018. Fire regimes at the arid fringe: A 16-year remote sensing perspective (2000–2016) on the controls of fire activity in Namibia from spatial predictive models. Ecol. Indic. 91, 324–337.
Mbourou, G.N.T., Bertrand, J., Nicholson, S., 1997. The diurnal and seasonal cycles of wind-borne dust over Africa north of the equator. J. Appl. Meteorol. 36, 868–882.
Miao, C., Duan, Q., Sun, Q., Huang, Y., Kong, D., Yang, T., Ye, A., Di, Z., Gong, W., 2014. Assessment of CMIP5 climate models and projected temperature changes over Northern Eurasia. Environ. Res. Lett. 9, 055007.
Morcrette, J.J., Boucher, O., Jones, L., Salmond, D., Bechtold, P., Beljaars, A., Benedetti, A., Bonet, A., Kaiser, J., Razinger, M., 2009. Aerosol analysis and forecast in the European Centre for medium-range weather forecasts integrated forecast system: forward modeling. J. Geophys. Res. Atmos. 114.
Nabavi, S.O., Haimberger, L., Samimi, C., 2016. Climatology of dust distribution over West Asia from homogenized remote sensing data. Aeolian Res. 21, 93–107.
Nabavi, S.O., Haimberger, L., Samimi, C., 2017. Sensitivity of WRF-chem predictions to dust source function specification in West Asia. Aeolian Res. 24, 115–131.
Paninski, L., 2003. Estimation of entropy and mutual information. Neural Comput. 15, 1191–1253.
Parrella, F., 2007. Online support vector regression. Master's Thesis. Department of Information Science, University of Genoa, Italy.
Prospero, J.M., Ginoux, P., Torres, O., Nicholson, S.E., Gill, T.E., 2002. Environmental characterization of global sources of atmospheric soil dust identified with the Nimbus 7 Total Ozone Mapping Spectrometer (TOMS) absorbing aerosol product. Rev. Geophys. 40, 40.
Sayer, A., Hsu, N., Bettenhausen, C., Jeong, M.J., 2013. Validation and uncertainty estimates for MODIS Collection 6 “Deep Blue” aerosol data. J. Geophys. Res. Atmos. 118, 7864–7872.
Sayer, A., Hsu, N., Lee, J., Carletta, N., Chen, S.H., Smirnov, A., 2017. Evaluation of NASA Deep Blue/SOAR aerosol retrieval algorithms applied to AVHRR measurements. J. Geophys. Res. Atmos.
Sayer, A., Munchak, L., Hsu, N., Levy, R., Bettenhausen, C., Jeong, M.J., 2014. MODIS Collection 6 aerosol products: Comparison between Aqua's e-Deep Blue, Dark Target, and “merged” data sets, and usage recommendations. J. Geophys. Res. Atmos. 119.
Schepanski, K., Tegen, I., Todd, M., Heinold, B., Bönisch, G., Laurent, B., Macke, A., 2009. Meteorological processes forcing Saharan dust emission inferred from MSG-SEVIRI observations of subdaily dust source activation and numerical models. J. Geophys. Res. Atmos. 114.
Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Ziese, M., 2011. GPCC full data reanalysis version 6.0 at 0.5: monthly land-surface precipitation from rain-gauges built on GTS-based and historic data. doi: 10.5676/DWD_GPCC. FD_M_V6_050.
Shao, Y., Wyrwoll, K.-H., Chappell, A., Huang, J., Lin, Z., McTainsh, G.H., Mikami, M., Tanaka, T.Y., Wang, X., Yoon, S., 2011. Dust cycle: an emerging core theme in Earth system science. Aeolian Res. 2, 181–204.
Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Duda, M., Huang, X., Wang, W., Powers, J., 2008. A description of the Advanced Research WRF Version 3, NCAR technical note, Mesoscale and Microscale Meteorology Division. National Center for Atmospheric Research, Boulder, Colorado, USA.
Smola, A., Schölkopf, B., 1998. A tutorial on support vector regression. NeuroCOLT Tech. Rep. NC-TR-98-030. Royal Holloway Coll., Univ. London, UK.
Taheri Shahraiyni, H., Sodoudi, S., 2016. Statistical modeling approaches for PM10 prediction in urban areas; A review of 21st-century studies. Atmosphere 7, 15.
Tao, M., Chen, L., Wang, Z., Tao, J., Che, H., Wang, X., Wang, Y., 2015. Comparison and evaluation of the MODIS Collection 6 aerosol data in China. J. Geophys. Res. Atmos. 120, 6992–7005.
Torgo, L., Ribeiro, R.P., Pfahringer, B., Branco, P., 2013. Smote for regression. In: Portuguese Conference on Artificial Intelligence. Springer, pp. 378–389.
Tucker, C., Pinzon, J., Brown, M., 2004. Global Inventory Modeling and Mapping Studies. Global Land Cover Facility, University of Maryland, College Park, Maryland.
Vapnik, V.N., 1995. Introduction: four periods in the research of the learning problem. The Nature of Statistical Learning Theory. Springer.
Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. J. Clim. 23, 1696–1718.
Wagner, W., Dorigo, W., de Jeu, R., Fernandez, D., Benveniste, J., Haas, E., Ertl, M., 2012. Fusion of active and passive microwave observations to create an essential climate variable data record on soil moisture. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. (ISPRS Ann.) 7, 315–321.
White, B.R., 1979. Soil transport by winds on Mars. J. Geophys. Res. Solid Earth 84, 4643–4651.
Wilks, D., 2011. Statistical Methods in the Atmospheric Sciences, 2011. Academic Press.
Yu, Y., Notaro, M., Kalashnikova, O.V., Garay, M.J., 2016. Climatology of summer Shamal wind in the Middle East. J. Geophys. Res. Atmos. 121, 289–305.
Yu, Y., Notaro, M., Liu, Z., Wang, F., Alkolibi, F., Fadda, E., Bakhrjy, F., 2015. Climatic controls on the interannual to decadal variability in Saudi Arabian dust activity: toward the development of a seasonal dust prediction model. J. Geophys. Res. Atmos. 120, 1739–1758.
Zarandi, M.F., Zarinbal, M., Ghanbari, N., Turksen, I., 2013. A new fuzzy functions model tuned by hybridizing imperialist competitive algorithm and simulated annealing. Application: stock price prediction. Inf. Sci. 222, 213–228.
Zhang, L., Zhou, W., Jiao, L., 2004. Wavelet support vector machine. IEEE Trans. Syst. Man Cybern. Part B (Cybernetics) 34, 34–39.
Zhang, W., Goh, A.T., 2016. Multivariate adaptive regression splines and neural network models for prediction of pile drivability. Geosci. Front. 7, 45–52.
Zhang, Z., Ma, C., Xu, J., Huang, J., Li, L., 2015. A novel combinational forecasting model of dust storms based on rare classes classification algorithm. Geo-Informatics in Resource Management and Sustainable Ecosystem. Springer.