Science of The Total Environment, Volume 737, October 2020, Article number 139508
DOI:10.1016/j.scitotenv.2020.139508

Identifying source of dust aerosol using a new framework based on remote sensing and modeling

Omid Rahmati[a,b], Farnoush Mohammadi[c], Seid Saeid Ghiasi[c], John Tiefenbacher[d], Davoud Davoudi Moghaddam[e], Frederic Coulon[f], Dieu Tien Bui[g*]

[a] Geographic Information Science Research Group, Ton Duc Thang University, Ho Chi Minh City, Viet Nam

[b] Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City, Viet Nam

[c] Faculty of Natural Resources Management, University of Tehran, Karaj, Iran

[d] Department of Geography, Texas State University, San Marcos, TX 78666, USA

[e] Department of Watershed Management, Agriculture and Natural Resources Faculty, Lorestan University, Khorramabad, Iran

[f] School of Water, Energy and Environment, Cranfield University, Cranfield MK43 0AL, UK

[g] Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

Corresponding author's email address: buitiendieu@duytan.edu.vn

Abstract

Dust particles are transported globally. They can adversely impact both human health and the environment, but they also impact transportation infrastructure, agriculture, and industry, occasionally severely. The identification of the locations that are the primary sources of dust, especially in arid and semi-arid environments, remains a challenge as these sites are often in remote or data-scarce regions. In this study, a new method that uses state-of-the-art machine-learning algorithms – random forest (RF), support vector machines (SVM), and multivariate adaptive regression splines (MARS) – was evaluated for its capacity to spatially model the distribution of dust-source potential in eastern Iran. To accomplish this, empirically identified dust-source

Published by Elsevier. This is the Author Accepted Manuscript issued with: Creative Commons Attribution Non-Commercial No Derivatives License (CC:BY:NC:ND 4.0).
The final published version (version of record) is available online at DOI:10.1016/j.scitotenv.2020.139508. Please refer to any applicable publisher terms of use.

locations were determined with the ozone monitoring instrument aerosol index and the Moderate-Resolution Imaging Spectroradiometer (MODIS) Deep Blue aerosol optical thickness methods. The identified areas were divided into training (70%) and validation (30%) sets. Measurements of the conditioning factors (lithology, wind speed, maximum air temperature, land use, slope angle, soil, rainfall, and land cover) were compiled for the study area and a predictive model was developed. To validate the prediction maps, the area-under-the-receiver operating characteristics (AUC) curve and true-skill statistics (TSS) were used. The results show that the RF algorithm performed best (AUC= 89.4% and TSS= 0.751), followed by the SVM (AUC= 87.5%, TSS= 0.73) and the MARS algorithm (AUC= 81%, TSS= 0.69). The results of the RF indicated that wind speed and land cover are the most important factors affecting dust generation. The highest dust-source potential region identified by the RF is in the eastern parts of the study region. This model can be applied to other arid and semi-arid environments that experience dust storms to promote management that prevents desertification and reduces dust production.

Keywords: Dust storm; Modelling; GIS; Remote sensing; Iran,

1. Introduction

As extreme air pollution events, dust storms significantly elevate particulate matter (PM) (of diameter of 10 and 2.5 microns, PM_{10} and $PM_{2.5}$) concentrations across affected areas. High PM_{10} and $PM_{2.5}$ concentrations can have adverse effects on human respiratory systems and promote bronchitis, chronic cough, and other chest illnesses (Goudie, 2009; Panikkath et al., 2013; Yu et al., 2013; Perini et al., 2017). The pathological mechanism through which dust induces human respiratory and immune systems responses is the increase of oxidative stress in respiratory

epithelial cells (Goudie, 2014; Zhang et al., 2016). Environmental epidemiology research has shown that dust storms transmit diseases by transporting heavy metals, bacteria, and fungi that impact exposed populations (Shotyk et al., 2016; Ardon-Dryer et al., 2020; Goudie, 2020). Beyond the health impacts, dust storms also damage transportation infrastructure (i.e., roads and railroads) (Yang et al., 2001; Lee, 2018), and agricultural and industrial lands. Additionally, dust storms can increase soil salinity, promote desertification, and diminish the quality of water supplies so much that great social and economic damages occur (Cao et al., 2015a; Middleton, 2017). Although dust storms typically happen in arid and semi-arid environments, they can spread great amounts of PM, pollutants, and allergens thousands of km from the source regions (Zoljoodi et al., 2013; Cao et al., 2015a). Dust storms also provide nutrients to the open oceans and affect the marine ecology (Griffin and Kellogg, 2004; Singh et al., 2008). Identifying potential dust-source areas is critical to discern the factors that determine the availability of dust and the mechanisms that generate dust storms, and to help anticipate likely impacts.

In general, different approaches have been used to identify dust-source areas: (1) horizontal visibility, (2) the mineralogy of dust samples, (3) Lagrangian back-trajectory, (4) satellite-based remote sensing, (5) inverse modeling, and (6) Goddard Earth Observing System chemistry (GEOS-Chem) adjoint modeling techniques (Schepanski et al., 2012). Horizontal visibility analyses employ the reports generated from synoptic stations. Although synoptic stations provide a long-term record of local visibility, the observation network is sparse, very localized, and sited close to cities rather than in natural environments or in dust-source areas (Mahowald et al., 2007; Schepanski et al., 2012). The horizontal visibility method is also limited in value because transported dust can affect measurements at the stations. Consequently, interpretation of visibility data to identify the sources of dust is difficult (Rashki et al., 2012; Camino et al., 2015). Given

these limitations, reports generated from synoptic stations do not allow the proper identification of source areas of dust. Therefore, the other 4 methodologies are more useful solutions to this problem. The mineralogical method assesses the transport pathway and source geomorphology of dust samples and links the mineral composition of dust to its potential source areas. However, this approach requires numerous field investigations and significant laboratory analysis, which are both time-consuming and expensive, especially when potential source regions cover large areas (Formenti et al., 2011). The Lagrangian back-trajectories method traces the travel path of dust backwards to the likely dust-source region (e.g., Alonso-Pérez et al., 2012). Measurements for this process require finite, but often large, volume of air. The main shortcomings of this tactic are that the Lagrangian model depends upon variables with significant uncertainties (i.e., parameterizations and meteorological data that are inputs to the model).

Previous studies have focused on the determination of dust-source areas using remote sensing data acquired on days with dust storms (Prospero et al., 2002; Schepanski et al., 2007, 2012; Ginoux et al., 2012; Baddock et al., 2016; Nabavi et al., 2016; Feuerstein and Schepanski, 2019). Some indicators have been commonly used to identify the sources: false color composition and brightness temperature difference (Bullard et al., 2008; Zhang et al., 2008; Lee et al., 2009; Sharma et al., 2009; Hahnenberger and Kathleen, 2014). Parajuli et al. (2014) used the Moderate Resolution Imaging Spectroradiometer (MODIS) Deep Blue aerosol optical depth, a land cover map, and meteorological and geomorphological factors to identify potential dust source regions. Their results reveal that the identification of the spatiotemporal extent of dust sources can be very complex and requires a clear understanding the properties of surface features like the geology, land cover, and physical and chemical compositions of soils. However, conducting field surveys and analyzing the mineral composition of dust samples to identify dust sources is difficult and costly

because deserts are usually spatially extensive areas (Wang et al., 2005; Ma et al., 2016). Furthermore, the relationships among the geomorphological, environmental, and meteorological factors further complicate the source-identification process. With the emergence of artificial-intelligence and machine-learning algorithms came the opportunity to address these challenges more effectively (Lary et al., 2016; Proietti et al., 2016). Machine-learning models are useful for extraction of relationships or patterns between a set of variables and a target factor, they can solve complex issues, and they can be used to extract information from available data (Olden et al. 2008; Naghibi et al., 2016). Recently, Rahmati et al. (2020) used an adaptive neuro-fuzzy inference system (ANFIS) for dust source modeling and prediction. Their results indicated that the hybridized ANFIS model had a good prediction performance for identifying dust source areas.

Previous works have identified western Asia (Syria, Iran, Iraq, and the Arabian Peninsula) as a major source region for dust using remote sensing techniques (Boloorani et al., 2013; WMO and UNEP, 2013). In fact, it is known that Asian dust storms are key influences in global biogeochemical cycles (Deng et al., 2007; Chan et al., 2008; Chen et al., 2018). Cao et al. (2015b) shows that the midwestern, southwestern, southern, and southeastern parts of Iran have very high potential for dust-storm occurrence. However, dust source potential and its driving factors have been rarely investigated (Rahmati et al., 2020). To the best of our knowledge, detailed information about dust-source areas in all these countries, especially Iran, is still limited.

This study has developed a novel methodology that assesses three state-of-the-art machine-learning algorithms – the random forest (RF) method, support vector machines (SVM), and multivariate and adaptive regression splines (MARS) technique, – and two remote-sensing techniques – ozone monitoring instrument aerosol index, and the MODIS Deep Blue aerosol optical thickness – to identify the areas with high map dust-source potential areas. The RF, SVM-

RBF, and MARS algorithms have achieved acceptable performance responding to several geo-environmental problems such as groundwater potential, landslides, gullying, floods, forest fires, and ecological studies (Leathwick et al., 2006; Conoscenti et al., 2015; Tehrany et al., 2014, 2015; Pourghasemi et al., 2017; Golkarian et al., 2018; Sachdeva et al., 2018). Due to the assortment of potential algorithms, we selected RF from tree-based models, radial basis function (RBF) kernel from SVMs, and MARS from non-parametric regression methods to provide perspective about the strengths and weaknesses of each approach in this field of study. To spatially model dust-source potential in the study area, several conditioning factors (wind speed, lithology, maximum air temperature, land cover, slope, soil, rainfall and land use) were considered potential predictors of dust-generation.

The main objectives of this research are to: (i) model dust-source potential by using RF, SVM, and MARS algorithms; (ii) compare the accuracy of the outputs using statistical evaluation criteria and the ground-truth; and (iii) determine the importance of each conditioning factor on dust-source potential. To the best of our knowledge, this is one of the first efforts to spatially model dust source potential using machine learning models. The results demonstrate how machine-learning models can be used to identify and predict dust-source areas in a representative dust-storm area.

2. Material and methods

2.1. Study area

The study area includes four cities (Mashhad, Birjand, Zabol, and Zahedan) from three provinces (South Khorasan, Razavi Khorasan, and Sistan & Balochestan) in Iran. The study area is in the eastern part of Iran. The neighbor states of Tajikestan, Afghanistan, and Pakistan establish

the eastern boundary of the region, which is roughly delimited by 25°06' to 37°41' N and 55°22' to 63°18' E. It encompasses an area of 444,904 km^2 (Fig. 1). The region is classified as either arid or semi-arid climatologically. According to Cao et al. (2015b), the southern parts of this area are significant sources of dust, as evidenced by their investigation of dust storms that occurred between 2000 and 2013. Dust storms in this region have generated numerous social problems and have led to emigration to other regions (Rahmati et al., 2020). This has obligated the Iranian government to define the conditions that are dangerous for sensitive people (Cao et al., 2015b).

Fig. 1 here

2.2. Method

A method was developed and run to generate dust-source potential maps using the selected three algorithms and validated their results. First, field surveys, weather station data, and geological information were used to identify examples of dust source areas. Then, the influencing factor data of the dust-source and non-dust-source locations were input into the RF, MARS, and SVM algorithms to map dust-source potential (Fig. 1). Finally, the maps were validated using the performance assessment indices of area under receiver operating characteristics curve (AUC) and true-skill statistics (TSS).

Fig. 2 here

2.2.1. Detecting dust source samples on satellite images

In our previous study (Rahmati et al., 2020), we determined dust sources using two indices Deep Blue aerosol optical thickness and the Ozone Monitoring Instrument aerosol index and field surveys. The two most widely used algorithms are the dark target (Levy et al., 2013), and deep blue (Hsu et al., 2013; Sayer et al., 2014) retrievals, which are used over dark (ocean and forest

areas) and bright (deserts) surfaces, respectively. The aerosol index assesses ultraviolet wavelengths (Torres et al., 2007). It is a semi-quantitative aerosol-absorbing index. There is very low surface reflectivity of ultraviolet wavelengths. According to the literature, the "frequency of observation (FOO)" (also termed frequency of occurrence (FOO)) of elevated dust signals can identify the persistent dust sources in large-scale regions (Sannazzaro et al., 2014; Moridnejad et al., 2015; Baddock et al., 2016). Therefore, we used the FOO approach to produce a verified dust-source inventory map. In fact, the frequency of occurrence refers to the number of days when aerosol optical thickness is greater than its threshold, and the "Angstrom exponent and single scattering albedo" fulfil the criteria for being deemed a dust source (Srivastava et al., 2014; Singh et al., 2016). Dust particles absorb these wavelengths and reduce their reflectance off the surface. This feature can be used to detect dust storms and their dust-storm sources by some indices and thresholds (Bullard et al., 2008). Dust storm thresholds of aerosol optical thickness and aerosol index were determined based on trial and error, as is suggested by Middelton and Goudie (2001), Prospero et al. (2002), and Ginoux et al. (2010). The current study investigated dust storms between April 2014 and May 2018. The details of remote sensing analysis and filed surveying can be find in Rahmati et al. (2020). The dust-source inventory map includes 85 locations (N = 85) were determined to have been dust-source areas for dust storms.

When running the models, the validation data should be different from those applied in the training step (Rahmati et al., 2019). Therefore, the dust-source inventory map (i.e., dependent dataset) is commonly partitioned into two groups: training (70% of dust-source locations) and validation (30% of dust-source locations). The machine learning models often need two sorts of samples to predict dust-source areas: the first group must be representative of the dust-source conditions (termed presence or positive cases) and the second group must be representative of

stability conditions (called absence or negative cases). In natural disasters modeling, it is generally recommended that the balances of positive/negative cases (i.e., dust source/non-dust source) should be equal to 1 (Nefeslioglu et al., 2008; Pham et al., 2017; Zêzere et al., 2017; Choubin et al., 2019). Therefore, similar to the dust-source dataset, 85 areas were randomly and away from the dust source areas selected to represent non-dust-source locations (i.e., negative cases). After random selection, those areas were checked to ensure that they were not close to dust-source areas. From the sets of dust-source and non-dust-source areas, 70% (i.e., 56 dust-source and 56 non-dust-source) were randomly extracted to train the models and the remaining 30% (i.e., 29 dust-source and 29 non-dust-source areas) were used to validate the results of the models (Fig. 3). The accuracy of the models in the training and validation phases termed as goodness-of-fit and predictive performance, respectively (Pourghasemi et al., 2017). Since the validation dataset is not used in model building/calibration, the predictive performance shows the validity and accuracy of final output of the models (Dou et al., 2015).

Fig. 3 here

2.2.2. Dust source conditioning factors

Wind speed

Dust storms are produced by wind erosion and the most important representation for this type of erosion is wind speed. According to the literature, summers correspond with the highest dust emission (Washington et al., 2003; Francis et al., 2017; Rashki et al., 2017). During the summer season, the region is under the influence of a low pressure system as a consequence of the south Asian monsoon pattern, resulting in developing of the “120-day wind” (about May to August). This wind type causes frequent dust emission in the study area (Rashki et al., 2013).

Therefore, mean wind speed (over the period May 2018 to August 2018) were calculated based on data from 27 synoptic stations collected by the Forests, Range and Watershed Management Organization (FRWMO). To map wind speed in the study region, wind speed data were obtained and interpolated using the inverse-distance weighting method. Wind speeds ranged between 2.2 and 11.09 m s^{-1} (Fig. 4a). The eastern parts of the study region experience higher wind speeds than other parts throughout the year.

Fig. 4 here

Lithology

Lithology classes are important influences as dust storms are generated in areas sensitive lithology rather than in areas with resistant units (Sissakian et al., 2013). The spatial data provided in a geographic information system layer of classified lithology was acquired from the Geology Survey of Iran (GSI) (1997) at a scale of 1:100,000 with a spatial resolution of 30×30 m. There are several lithological units from different geological era, periods, and epochs in the study region (Table 1; Fig. 4b).

Table 1 here

Maximum air temperature

Air temperature is an important factor that affects soil moisture and subsequently threshold wind velocity (Yang et al., 2012). The lower the soil moisture is, the lower the threshold wind velocity is. Decreasing cohesive force through reducing soil moisture lead to this fact that the soil surface became more erodible in summer (Li and Zhang, 2012). Therefore, similar to wind speed factor, maximum air temperature data were selected for the period May 2018 to August 2018 (i.e., when the highest dust emission is occurred). The 27 synoptic meteorological stations provided

maximum temperature data for the study period. These values were interpolated for the study area using inverse-distance weighting. Maximum air temperature ranged from 42.1° to 49°C; the eastern and southern parts of the study area had the highest temperatures (Fig. 4c).

Land use

A map of land use classification in the study region in 2018 was acquired from FRWMO (2018). There are nine land use categories: surface water, forest, agriculture, gardens, bare land, rangeland, salt land, urban, and airport. Salt land and rangeland cover the largest parts of the study area with 54.82% and 23.42% of the region, respectively (Fig. 4d).

Table 2 here

Slope

The threshold for wind erosion varies for different lithology and land use types, but it is also influenced by slope angle (represented as slope percentage). A digital elevation model at a spatial resolution of 30m×30m was obtained from Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) images. A slope map of the study area was created and classified from 0 to 185.3% (Fig. 4e).

Soil

The susceptibility of a land surface to wind erosion and dust emission is highly dependent on characteristics of soils (Grini et al., 2005; Webb and Strong, 2011). Erosive soils without vegetation are fields of high potential for dust production (Sissakian et al., 2013). The soil map of the study area at a scale of 1:100,000 was obtained from the FRWMO (2018). Thirteen soil types

are found in the study area. Rocky lands, aridisols, entisols/aridisols, and rock outcrops/Entisols are the major soil types (Fig. 4f; Table 3).

Table 3 here

Rainfall

Rainfall increases soil moisture content that results in increases the cohesive forces between soil particles (Kim and Choi, 2015). Rainfall also promotes vegetation that plays an important role in reducing wind effects. Therefore, rainfall characteristics in a given area provide vital information about dust source areas. Since the study are does not have considerable rainfall during the summer, mean annual rainfall data from 27 rain gauge stations were acquired from the FRWMO (2018). These data were interpolated using inverse-distance weighting. Annual rainfall in the study area ranged from 57.4 to 290.1 mm (Fig. 4g). Rainfall is highest in the northern part of the study area and diminishes southward.

Land cover

Lands that are devoid of vegetation usually yield significant amounts of dust. Therefore, land-cover classes are important for predicting the sources of dust. A land-cover map of the study area was obtained from the FRWMO (2018) (Fig. 4h). There are four land-cover classes represented on the map. These include bare lands, urban, vegetation, and water body.

2.2.3. Machine-learning algorithms

The RF algorithm has been used in several studies to predict locations that might be prone to several land-management or resource problems (e.g., landslides, flood susceptibility, and

groundwater potential) (Rahmati et al., 2016; Tehrany et al., 2015; Pourghasemi and Kerle, 2016). There have been many successful applications of this method. RF uses “bagger and random feature selection” techniques together to predict the target variable (Ho, 1998). The main advantage of the RF is that it mixes several “binary decision trees” to yield highly accurate projections of the patterns of the phenomena of importance (Breiman, 2001). RF models produce highly accurate predictions as these models diminish the instability that was seen in previous renditions of tree-based models. RF extracts random binary models that employ a portion of the cases in a “bootstrapping” analysis (Tehrany et al., 2014; Pourghasemi and Kerle, 2016). RF enhances diversity among the classification trees through regrouping the data and altering the predictive factors (Naghibi et al., 2016). The important parameters in RF could be regarded as “the number of trees and the number of factors” that are implemented to define the split “at each node” (Vorpahl et al. 2012). An advantage of the RF model is that it evaluates the relative importance of the conditioning factors based on mean-decrease accuracy (Archer and Kimes, 2008; Grömping, 2009).

MARS, by contrast, is a technique to fit adaptive non-linear regression. It implements piecewise- basis functions for modelling the response factor with corresponding input factors (Friedman, 1991). This algorithm conducts a non-parametric modelling technique that does not require normal distributions of the factors (Friedman, 1991; Balshi et al., 2009). MARS involves dividing the range of the interpretive factors to groups by creating a linear regression equation for the groups (Chen et al., 2017). A basic and more comprehensive description of MARS can be found in Friedman (1991).

SVM, initially built by Vapnik (1998) and Cherkassky and Mulier (2007), is a supervised classifier for binary issues. SVM operates under the process of minimizing structural risk, which

eliminates the shortcomings of the more common empirical risk-minimization methods when using limited datasets for training of the model (Betrie et al., 2013). SVM segregates hyper-plane formation from the selected dataset for training (Tehrany et al., 2014). It creates the “hyper-plane” in the middle of the highest margin (Pradhan, 2013). Accuracy of the findings of SVM are influenced by the kernel type. This research used the radial basis function (RBF) because of the advantages it offers compared to other kernel types, such as its higher interpolation ability (Tehrany et al., 2015).

In this study, the format of predictive files (i.e., independent variables) was converted to ASCII files. Both presence and absence points were merged and used as the dependent variable. The construction of the database and extraction of information were done using “raster”, “rgdal”, “maptools”, “modelmap”, “sp”, “dismo”, “biomod2”, “sdmtools”, and “sdmvspecies”. All models were implemented with the package "sdm" in R software that developed by Naimi and Araújo (2016).

2.2.4. Performance assessment

Validation is an important step in any modelling exercise. In this study, ROC curve was plotted and TSS was calculated to determine the usefulness of the models (Naghibi et al., 2017, 2019; Rahmati et al., 2019). TSS was implemented in this research study because it considers “omission and commission errors” (Rahmati et al., 2019). ROC, on the other hand, was selected for this study because it reveals overall performance of a binary model and the impact of other indices such as sensitivity, specificity, true positive rate, true negative rate, false positive rate, and false negative rate. The ROC curve plots true positive rate (TPR) versus false positive rate (FPR) to describe the quality of a model (Tien Bui et al., 2016a). The area under the curve (AUC) ranges

from 0 to 1; the higher the value, the better the performance. Considering elements of a confusion matrix (TP, TN, FP, and FN) (Table 4), TPR and FPR are defined as:

$$TPR = TP/(TP + FN) \quad (1)$$

$$\mathrm{FPR} = FP/(FP + TN) \quad (2)$$

where TP and TN show true positive and true negative, respectively; and FP and FN depict false positive and false negative, respectively. TSS is calculated as (Fukuda et al., 2013; Rahmati et al., 2018):

$$TSS = TPR - FPR \quad (3)$$

In an analytic expression of the ROC curve, AUC metric can be calculated as (Eq. 4):

$$AUC = \int_0^1 f(FPR)\, dFPR = 1 - \int_0^1 f(TPR)\, dTPR \quad (4)$$

Table 4 here

In this study, the accuracy of the models in the training (i.e., goodness-of-fit) and validation (i.e., predictive performance) steps was assessed using TSS and AUC metrics. The goodness-of-fit just reflects how well the model fits the training data set (Dodangeh et al., 2012; Pourghasemi et al., 2017). Hence, the prediction power of the model cannot be judged by the goodness-of-fit, because it is determined by the presence data (here sampled dust-source locations) that were already used to calibrate the model (Frattini et al., 2010; Bui et al., 2013). Therefore, predictive performance is well-known as the main step for comparing the accuracy of the models (Rahmati et al., 2019). Importantly, as explained before (section 2.2.1), dust-source locations that identified by analyzing satellite images were assessed by field surveying.

3. Results and discussion

3.1. Dust storm potential maps

Using the training dataset as well as RF, SVM and MARS algorithms, three dust-source potential maps were prepared for the study area (Fig. 5a- c). These maps were classified into four categories (low, medium, high, and very high) using an equal-interval classification scheme to enable comparison of the model outputs (Figs. 6a-c). The four categories (low to very high) covered 17.6, 39.7, 35.3, and 7.4% of the study area, respectively, on the RF model map. The SVM algorithm, predicted (low to very high) 22.1, 34.8, 33.6, and 9.5%, respectively. The MARS model predicted 37.4, 13.4, 7.3, and 41.9%, respectively. Our results showed that Sistan and Zabol basins are core sources of dust storm which is in line with findings of Cao et al., (2015b). These areas include the areas surrounding Khaf, Sangan, Aliabad, Ghasem Abad, and Hoseynabad in the northeastern parts of the study region. In addition, the results of Prospero et al. (2002) indicated that the areas surrounding the city of Zabol – including Hamun-e Saberi, Hamun-e Puzak, and Daryacheh-ye-Hamun – are main sources of dust storm in eastern Iran. The results of this study also showed high dust source potential for these areas. These findings are consistent with our previous study in this study area (Rahmati et al., 2020) that applied hybridized adaptive neuro-fuzzy inference system (ANFIS) models.

Fig. 5 here

Fig. 6 here

3.2. Importance of the conditioning factors

The importance of the conditioning factors was evaluated with the results produced by the RF algorithm by calculating the mean decrease accuracy (Table 5). Wind speed (PC = 29.5) was

shown to make the greatest contribution to the model. This was followed by land cover type (PC = 25.3), soil type (PC = 14.8), and maximum air temperature (PC = 9.5). Whereas slope, lithology, and rainfall contributed least to the prediction of dust-source potential.

Table 5 here

The importance of wind speed could be related to its role in initiating dust storms, the main driving force behind this phenomenon (Liu et al., 2004). The results are consistent with Alizadeh-Choobari et al. (2014), which evaluated the relationship between wind speeds and dust-storm activity in a portion (the Sistan area) of this study area. They found that strong near-surface wind speeds generate substantial dust emissions from the Sistan area. Lee et al. (2009) proved that in flat areas, winds easily reach the erosion threshold, and thus initiate dust storms.

Land cover also influences both soil and wind speed. Soil properties often reflect the amount of vegetative cover and various types of land cover generate greater or less drag on wind due to roughness length and can create windbreaks that reduce wind speed. Two categories of land cover, bare lands and salt lands, are more likely to be sources of dust as they lack vegetation and the wind-erosion threshold is at its lowest (i.e., most easily surpassed). Lee et al. (2012) indicated that land cover management of anthropogenic activities strongly affects the spatial pattern of dust generation.

Soil is the third important factor in dust-source modelling. This is because the erosion threshold is directly tied to the soil characteristics and therefore dust-storm occurrence is heavily influenced by soil type. The majority of dust sources in the region are ephemeral lakes (sometimes known as playas) which have surfaces covered in fine particles (Goudie, 2018). This agrees with the conclusions of Hahnenberger and Nicoll (2014) which said that playas are the land surfaces

most susceptible to wind erosion and dust production. Erosion-sensitive soils have more potential to be dust source areas (Sissakian et al., 2013). Therefore, this soil quality may be an effective distinguishing factor to predict dust potential.

3.3. Performance of the models

Goodness-of-fit of models were shown in Table 6. The RF, SVM, and MARS algorithms had AUC values of 0.92, 0.898, and 0.86, respectively. Additionally, they had TSS values of 0.785, 0.758, and 0.721, respectively. Consideration of only goodness-of-fit is insufficient to judge the capability of models because it is determined based on the training dataset that used in model calibration (Tien Bui et al., 2012; Tehrany et al., 2015). The prediction performance must be assessed to investigate the predictive skill of models (Tien Bui et al., 2013; Pourghasemi et al., 2017). The predictive performance of the models was summarized in Table 7. The results showed that the RF (AUC = 0.894, TSS = 0.751) was the most effective model based on ground-truth data. SVM was the second best model (AUC = 0.875, TSS = 0.73), followed by MARS model (AUC = 0.81, TSS = 0.69). Therefore, the predictive performance analysis reveals RF to provide the best performance. The performance of models can be classified based on the AUC metric: < 70% (fair), 70 – 80% (good), 80 – 90% (very good), and > 90% (excellent). Since all three models showed an AUC value higher than 0.8 in the validation step, they classified as very good models for generating valid dust-source potential maps.

Table 6 here

Table 7 here

The direct comparison of the results of this study with our previous work in this study area (Rahmati et al., 2020) provides a suitable platform for judging the predictive skill of the models

employed in this study. The most accurate model in our previous study was the hybridized ANFIS-DE model (i.e., hybridized ANFIS and differential evolution metaheuristic algorithm) with an AUC value of 84.1%, whereas the best model in the current study was RF (AUC = 89.4%). Even the second best model was SVM (AUC = 87.5%) in this study had a better accuracy than ANFIS-DE model in our previous work.

The high-accuracy dust-source potential map produced by the RF algorithm could be the result of several advantages that the method possesses. The iterative nature of RF is unique among the models, as this effectively bootstraps the data for more robust and flexible predictions (Hudak et al., 2008). Moreover, the RF model differs fundamentally from other tree-based models (like the decision tree or the boosted regression tree) as it iteratively and randomly samples the data and variables to generate a forest (i.e., a large group) (Belgiu and Drăguţ, 2016). Furthermore, random subsets of predictor variables prevent problems produced by overfitting (Breiman, 2001). However, the RF and SVM algorithms are better choices for this purpose. Consequently, the dust-source potential maps of these models are similar, but both are very different from the map produced by the MARS model. Naghibi et al. (2018) states that the high efficacy of the RF algorithm in classification issues can be a function of its capacity to handle large datasets and to model larger numbers of input factors. Catani et al. (2013) indicated that RF is also capable of detecting non-linear relationships between the response and input factors, which makes it stronger than linear algorithms. This could be a very important trait for models to be used for land management, as the relationships between environmental phenomena (i.e., dust storms) and topo-hydrological and geo-environmental factors are not linear and are often complex. Breiman (2001) stated that RF does not have an overfitting issue and can manage the inclusion of outliers within the training data. Feature selection and pruning are two other techniques that can make the results

of tree-based models more accurate and may make them more useful more generally. Tien Bui et al. (2016b) stated that SVMs can model sophisticated, non-linear relationships. SVM has also robustness to noise.

Kamani et al. (2015) assessed heavy-metal contamination in street dusts in Zahedan (a part of this study area), Iran and provided evidence that people suffer from this problem. Public health implications influence migration from this region. Derakhshandeh et al. (2014) reported the results of a public survey about the influence of dust storms on people's desire to migrate from Ahvaz (a city in western Iran). They showed that residents tend to move from Ahvaz if they can overcome financial barriers preventing them from leaving. Based on our data, the population of 5,800 rural districts in the study region is currently decreasing, which can cause severe damage to local economies and to social conditions both in those rural areas and the cities to which they migrate.

4. Conclusion

Considering the health problems that are caused by dust storms, this study assessed a new method for assessing the areas of genesis for dust storms. Three advanced machine-learning algorithms (RF, SVM, and MARS) were assessed to identify likely dust source locations. The findings reported here lead to three main conclusions:

- The RF algorithm yielded the best dust-source potential map, followed by SVM and MARS. The results of this study highlight the benefits of tree-based and support vector machine algorithms for modelling complex, nonlinear phenomenon like dust-source potential. The current methodology takes advantages from remote sensing data and techniques in order to define dust source areas which reduces the costs and makes it possible to achieve acceptable results in a lower time. Implementation of the remote sensing data makes it possible to apply

the methodology even in data scarce regions which is another positive feature of the current methodology. Considering the advantages of the methodology, it is suggested to apply the methodology to assess the main dust-source areas in the Middle East, including parts of Iran, Iraq, Syria, and the Arabian Peninsula.

- Wind speed had the greatest contribution to the model. This was followed by land cover type, soil type, and maximum air temperature. However, slope, lithology, and rainfall contributed least to the prediction of dust-source potential.
- The dust-source potential map of the study region can therefore be used by land-use planners for dust storm prevention and dust-storm mitigation by focusing on land use management in their source areas. The planners are suggested to focus on high potential dust sources in the Eastern Iran such as Sistan and Zabol, Khaf, Sangan, Aliabad, Ghasem Abad, Hoseynabad, Hamun-e Saberi, Hamun-e Puzak, and Daryacheh-ye-Hamun. Mulching at high dust source potential areas could be a good option to control wind erosion and dust storm form the sources areas. Additionally, for the Hamun lakes which is dried now, the managers could have some plans to rehabilitate the lake by restoration of its feeding rivers.

Acknowledgements

The authors would like to thank the Geological Surveys of Iran (GSI) and the Iranian Department of Water Resources Management (IDWRM) for supplying required data, reports and useful maps. The authors also would like to thank Prof. Andrew S. Goudie (School of Geography and the Environment, University of Oxford) and Prof. Xavier Querol (Spanish Research Council) for constructive comments. In addition, this research was partially supported by the Geographic Information Science Research Group, Ton Duc Thang University, Ho Chi Minh City, Viet Nam.

We greatly appreciate the assistance of the associate editor, Prof. Jianmin Chen, and anonymous reviewers for their constructive comments that helped us to improve the paper.

References

Alizadeh-Choobari, O., Zawar-Reza, P., Sturman, A., 2014. The "wind of 120 days" and dust storm activity over the Sistan Basin. Atmospheric Research, 143, pp.328-341.

Alonso-Pérez, S., Cuevas, E., Querol, X., Guerra, J.C., Pérez, C., 2012. African dust source regions for observed dust outbreaks over the Subtropical Eastern North Atlantic region, above 25 N. Journal of Arid Environments 78, 100-109.

Archer, K.J., Kimes, R.V., 2008. Empirical characterization of random forest variable importance measures. Computational Statistics & Data Analysis 52(4), 2249-2260.

Ardon-Dryer, K., Mock, C., Reyes, J., Lahav, G., 2020. The effect of dust storm particles on single human lung cancer cells. Environmental Research 181, p.108891. https://doi.org/10.1016/j.envres.2019.108891

Baddock, M.C., Ginoux, P., Bullard, J.E., Gill, T.E., 2016. Do MODIS-defined dust sources have a geomorphological signature?. Geophysical Research Letters 43(6), 2606-2613.

Balshi MS, Mcguire AD, Duffy P, Flannigan M, Walsh J, Melillo J., 2009. Assessing the response of area burned to changing climate in western boreal North America using a Multivariate Adaptive Regression Splines (MARS) approach. Glob Change Biol 15, 578–600.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing 114, pp.24-31.

Betrie, G.D., Tesfamariam, S., Morin, K. A., Sadiq, R., 2013. Predicting copper concentrations in acid mine drainage: A comparative analysis of five machine learning techniques. Environ. Monit. Assess. 185, 4171–82. https://doi.org/10.1007/s10661-012-2859-7

Boloorani, A.D., Nabavi, S.O., Azizi, R., et al. 2013. Characterization of Dust Storm Sources in Western Iran Using a Synthetic Approach. Advances in Meteorology, Climatology and Atmospheric Physics. Berlin Heidelberg: Springer, 415–420.

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

Bui, D.T., Pradhan, B., Lofman, O., Revhaug, I., Dick, Ø.B., 2013. Regional prediction of landslide hazard using probability analysis of intense rainfall in the Hoa Binh province, Vietnam. Natural hazards 66(2), 707-730.

Bullard, J., Baddock, M., McTainsh, G., Leys, J., 2008. Sub-basin scale dust source geomorphology detected using MODIS. Geophysical Research Letters 35(15), 1-19.

Camino, C., Cuevas, E., Basart, S., Alonso-Pérez, S., Baldasano, J.M., Terradellas, E., Marticorena, B., Rodríguez, S. and Berjón, A., 2015. An empirical equation to estimate mineral dust concentrations from visibility observations in Northern Africa. Aeolian Research, 16, pp.55-68.

Cao, H., Amiraslani, F., Liu, J., Zhou, N., 2015a. Identification of dust storm source areas in West Asia using multiple environmental datasets. Sci. Total Environ. 502, 224–235. doi:10.1016/j.scitotenv.2014.09.025

Cao, H., Liu, J., Wang, G., Yang, G., Luo, L., 2015b. Identification of sand and dust storm source areas in Iran. J. Arid Land 7, 567–578. doi:10.1007/s40333-015-0127-8

Catani, F., Lagomarsino, D., Segoni S., Tofani, V., 2013. Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. Nat Hazards Earth Syst Sci 13, 2815–2831

Chan, C.C., Chuang, K.J., Chen, W.J., Chang, W.T., Lee, C.T., Peng, C.M., 2008. Increasing cardiopulmonary emergency visits by long-range transported Asian dust storms in Taiwan. Environmental Research 106(3), 393-400.

Chen, Q., Wang, M., Sun, H., Wang, X., Wang, Y., Li, Y., Zhang, L., Mu, Z., 2018. Enhanced health risks from exposure to environmentally persistent free radicals and the oxidative stress of PM2. 5 from Asian dust storms in Erenhot, Zhangbei and Jinan, China. Environment International 121, 260-268.

Chen, W., Pourghasemi, H.R., Naghibi, S.A., 2017. Prioritization of landslide conditioning factors and its spatial modeling in Shangnan County, China using GIS-based data mining algorithms. Bull. Eng. Geol. Environ. 1–19. doi:10.1007/s10064-017-1004-9

Cherkassky, V.S., Mulier, F. 2007. Learning from data: Concepts, theory, and methods. New Jersey: Wiley

Choubin, B., Rahmati, O., Tahmasebipour, N., Feizizadeh, B., Pourghasemi, H.R., 2019. Application of fuzzy analytical network process model for analyzing the gully erosion susceptibility. In Natural hazards gis-based spatial modeling using data mining techniques (pp. 105-125). Springer, Cham.

Conoscenti, C., Ciaccio, M., Caraballo-Arias, N.A., Gómez-Gutiérrez, Á., Rotigliano, E., Agnesi, V., 2015. Assessment of susceptibility to earth-flow landslide using logistic regression and multivariate adaptive regression splines: A case of the Belice River basin (western Sicily, Italy). Geomorphology 242, 49–64. doi:10.1016/j.geomorph.2014.09.020

Deng, F., Guo, X., Liu, H., Fang, X., Yang, M., Chen, W., 2007. Effects of dust storm PM2. 5 on cell proliferation and cell cycle in human lung fibroblasts. Toxicology in vitro, 21(4), 632-638.

Derakhshandeh, M., Rostami, M.H., Goudarzi, G., Rostami, M.Z., 2014. Evaluation of the tendency to migration in the case of ahvaz dust storm occurrence: A public survey. Advances in Civil and Environmental Engineering 2(1), 55-69.

Dodangeh, E., Shao, Y., Daghestani, M., 2012. L-Moments and fuzzy cluster analysis of dust storm frequencies in Iran. Aeolian Research 5, 91-99.

Dou, J., Yamagishi, H., Pourghasemi, H.R., Yunus, A.P., Song, X., Xu, Y., Zhu, Z., 2015. An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan. Natural Hazards 78(3), 1749-1776.

Feuerstein, S., Schepanski, K., 2019. Identification of Dust Sources in a Saharan Dust Hot-Spot and Their Implementation in a Dust-Emission Model. Remote Sensing 11(1), p.4.

Formenti, P., Schütz, L., Balkanski, Y., Desboeufs, K., Ebert, M., Kandler, K., Petzold, A., Scheuvens, D., Weinbruch, S., Zhang, D., 2011. Recent progress in understanding physical and chemical properties of African and Asian mineral dust. Atmospheric Chemistry and Physics 11(16), 8231-8256.

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 Research 24, 15-31.

Frattini, P., Crosta, G., Carrara, A., 2010. Techniques for evaluating the performance of landslide susceptibility models. Engineering Geology 111(1-4), 62-72.

Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Stat. 19, 1–141 (with discussion).

Fukuda, S., De Baets, B., Waegeman, W., Verwaeren, J., Mouton, A.M., 2013. Habitat prediction and knowledge extraction for spawning European grayling (Thymallus thymallus L.) using a broad range of species distribution models. Environ. Modell. Softw. 47, 1-6.

Geology Survey of Iran (GSI) 1997. Geology map of the Chaharmahal-e-Bakhtiari Province. http://www.gsi.ir/Main/ Lang_en/index.html. Accessed September 2000

Ginoux, P., Garbuzov, D., Hsu, N. C. 2010. Identification of anthropogenic and natural dust sources using moderate resolution imaging spectroradiometer (MODIS) deep blue level 2 data. Journal of Geophysical Research Atmospheres, 115(5), 1–10. doi:10.1029/2009JD012398

Golkarian, A., Naghibi, S.A., Kalantar, B., Pradhan, B., 2018. Groundwater potential mapping using C5.0, random forest, and multivariate adaptive regression spline models in GIS. Environ. Monit. Assess. 190. doi:http://dx.doi.org/10.1007/s10661-018-6507-8

Goudie A.S., 2009. Dust storms: Recent developments. J. Environ. Manage. 90, 89–94

Goudie, A.S., 2014. Desert dust and human health disorders. Environment Int. 63,101-113.

Goudie, A.S., 2018. Dust storms and ephemeral lakes. Desert, 23(1), pp.153-164.
Goudie, A.S., 2020. Dust Storms and Human Health. In Extreme Weather Events and Human Health (pp. 13-24). Springer, Cham.
Griffin, D.W., Kellogg, C.A., 2004. Dust storms and their impact on ocean and human health: dust in Earth's atmosphere. EcoHealth 1(3), 284-295.
Grini, A., Myhre, G., Zender, C.S. and Isaksen, I.S., 2005. Model simulations of dust sources and transport in the global atmosphere: Effects of soil erodibility and wind speed variability. Journal of Geophysical Research: Atmospheres 110(D2). https://doi.org/10.1029/2004JD005037
Grömping, U., 2009. Variable importance assessment in regression: Linear regression versus random forest. The American Statistician 63(4), 308-319.
Hahnenberger, M., Nicoll, K., 2014. Geomorphic and land cover identification of dust sources in the eastern Great Basin of Utah, USA. Geomorphology 204, 657-672.
Hahnenberger, M., Kathleen, N., 2014. Geomorphic and land cover identification of dust sources in the eastern Great Basin of Utah, U.S.A. Geomorphology 204(2), 657-672.
Hsu, N. C., Tsay, S. C., King, M. D., Herman, J. R. 2006. Deep Blue retrievals of Asian aerosol properties during ACE-Asia. IEEE Transactions on Geoscience and Remote Sensing 44(11), 3180–3195. doi:10.1109/TGRS.2006.879540
Hsu, N. C., Tsay, S., King, M. D., Kaufman, Y. J., Herman, J. R., Goddard, N., Flight, S. 2004. Aerosol Properties over "Bright-Reflecting. IEEE Transactions on Geoscience and Remote Sensing 42(3), 23. doi:10.1109/TGRS.2004.824067
Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, D.E., Falkowski, M.J., 2008. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. Remote Sensing of Environment 112(5), 2232-2245.
Kamani, H., Ashrafi, S.D., Isazadeh, S., Jaafari, J., Hoseini, M., Mostafapour, F.K., Bazrafshan, E., Nazmara, S., Mahvi, A.H., 2015. Heavy metal contamination in street dusts with various land uses in Zahedan, Iran. Bulletin of Environmental Contamination and Toxicology 94(3), 382-386.
Kim, H., Choi, M., 2015. Impact of soil moisture on dust outbreaks in East Asia: Using satellite and assimilation data. Geophysical Research Letters 42(8), 2789-2796.
Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. Geoscience Frontiers 7(1), 3-10.
Leathwick, J.R., Elith, J., Hastie, T., 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. Ecol. Modell. 199, 188–196. doi:10.1016/j.ecolmodel.2006.05.022
Lee, J.A., Baddock, M.C., Mbuh, M.J. and Gill, T.E., 2012. Geomorphic and land cover characteristics of aeolian dust sources in West Texas and eastern New Mexico, USA. Aeolian Research 3(4), 459-466.
Lee, J.A., Gill, T.E., Mulligan, K.R., Dominguez Acosta, M., Perez, A.E., 2009. Land use/land cover and point sources of the 15 December 2003 dust storm in southwestern North America. Geomorphology 105, 18–27. doi:10.1016/j.geomorph.2007.12.016
Lee, J.S., 2018. The environmental costs of Asian dust damages in Korea: applying a choice experiment. Environmental Economics and Policy Studies 20(3), 641-654.
Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., Hsu, N. C., 2013. The Collection 6 MODIS aerosol products over land and ocean. Atmospheric Measurement Techniques, 6(11), 2989.
Li, W., Liu, Z., 2011. A method of SVM with normalization in intrusion detection. Procedia Environ. Sci. 11, 256–262. doi:10.1016/j.proenv.2011.12.040
Li, X., Zhang, H., 2012. Seasonal variations in dust concentration and dust emission observed over Horqin Sandy Land area in China from December 2010 to November 2011. Atmospheric Environment 61, 56-65.
Liu, X., Yin, Z.Y., Zhang, X., Yang, X., 2004. Analyses of the spring dust storm frequency of northern China in relation to antecedent and concurrent wind, precipitation, vegetation, and soil moisture conditions. Journal of Geophysical Research: Atmospheres 109(D16). doi: 10.1029/2004JD004615.

Ma, L., Li, M., Zhang, H., Li, L., Huang, Z., Gao, W., Chen, D., Fu, Z., Nian, H., Zou, L., Gao, J., 2016. Comparative analysis of chemical composition and sources of aerosol particles in urban Beijing during clear, hazy, and dusty days using single particle aerosol mass spectrometry. Journal of Cleaner Production 112, 1319-1329.

Mahowald, N.M., Ballantine, J.A., Feddema, J., Ramankutty, N., 2007. Global trends in visibility: Implications for dust sources. Atmospheric Chemistry and Physics 7(12), 3309-3339.

Middleton, N.J., 2017. Desert dust hazards: A global review. Aeolian Research 24, 53-63.

Middleton, N.J., Goudie, A.S., 2001. Saharan dust: Sources and trajectories. Transactions of the Institute of British Geographers, 26(2), 165-181.

Moridnejad, A., Karimi, N., Ariya, P.A. 2015. A new inventory for middle east dust source points. Environmental Monitoring and Assessment 187(9). doi:10.1007/s10661-015-4806-x

Nabavi, S.O., Haimberger, L., Samimi, C., 2016. Climatology of dust distribution over West Asia from homogenized remote sensing data. Aeolian Research 21, pp.93-107.

Naghibi, S.A., Dolatkordestani, M., Rezaei, A., Amouzegari, P., Heravi, M.T., Kalantar, B., Pradhan, B., 2019. Application of rotation forest with decision trees as base classifier and a novel ensemble model in spatial modeling of groundwater potential. Environ. Monit. Assess. 191, 248. https://doi.org/10.1007/s10661-019-7362-y

Naghibi, S.A., Moghaddam, D.D., Kalantar, B., Pradhan, B., Kisi, O., 2017. A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping. J. Hydrol. 548, 471-483.

Naghibi, S.A., Pourghasemi, H.R., Abbaspour, K., 2018. A comparison between ten advanced and soft computing models for groundwater qanat potential assessment in Iran using R and GIS. Theor. Appl. Climatol. 131, 967–984. doi:10.1007/s00704-016-2022-4

Naghibi, S.A., Pourghasemi, H.R., Dixon, B., 2016. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. Environ. Monit. Assess. 188, 44. doi:10.1007/s10661-015-5049-6

Naimi, B., Araújo, M.B., 2016. sdm: a reproducible and extensible R platform for species distribution modelling. Ecography 39(4), 368-375.

Nefeslioglu, H.A., Gokceoglu, C., Sonmez, H., 2008. An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. Engineering Geology 97(3-4), 171-191.

Olden, J. D., Lawler, J. J., Poff, N. L. 2008. Machine learning without tears: A primer for ecologists. The Quarterly Review of Biology 83(2), 171–193.

Panikkath, R., Jumper, C.A., Mulkey, Z., 2013. Multilobar Lung Infiltrates after Exposure to Dust Storm: The Haboob Lung Syndrome. Am. J. Med. 126, e5–e7. https://doi.org/10.1016/j.amjmed.2012.08.012

Parajuli, S.P., Yang, Z.L., Kocurek, G., 2014. Mapping erodibility in dust source regions based on geomorphology, meteorology, and remote sensing. Journal of Geophysical Research: Earth Surface, 119(9), 1977-1994.

Perini, K., Ottelé, M., Giulini, S., Magliocco, A., Roccotiello, E., 2017. Quantification of fine dust deposition on different plant species in a vertical greening system. Ecological engineering, 100, 268-276.

Pham, B.T., Bui, D.T., Prakash, I., Dholakia, M.B., 2017. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. Catena 149, 52-63.

Pourghasemi, H.R., Kerle, N., 2016. Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. Environ. Earth Sci. 75, 185. doi:10.1007/s12665-015-4950-1

Pourghasemi, H.R., Yousefi, S., Kornejady, A., Cerdà, A., 2017. Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. Sci. Total Environ. 609, 764–775. doi:10.1016/j.scitotenv.2017.07.198

Pourghasemi, H.R., Yousefi, S., Kornejady, A., Cerdà, A., 2017. Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. Science of the Total Environment, 609, 764-775.

Proietti, A., Liparulo, L., Leccese, F., Panella, M., 2016. Shapes classification of dust deposition using fuzzy kernel-based approaches. Measurement 77, 344-350.

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. Reviews of Geophysics 40(1), 1–31. doi:10.1029/2000RG000095

Rahmati, O., Falah, F., Naghibi, S. A., Biggs, T., Soltani, M., Deo, R. C., et al. 2019. Land subsidence modelling using tree-based machine learning algorithms. Science of The Total Environment 672, 239–252. doi:10.1016/j.scitotenv.2019.03.496

Rahmati, O., Falah, F., Naghibi, S.A., Biggs, T., Soltani, M., Deo, R.C., Cerdà, A., Mohammadi, F., Tien Bui, D., 2019. Land subsidence modelling using tree-based machine learning algorithms. Sci. Total Environ. 672, 239–252. https://doi.org/10.1016/j.scitotenv.2019.03.496

Rahmati, O., Kornejady, A., Samadi, M., Deo, R.C., Conoscenti, C., Lombardo, L., Dayal, K., Taghizadeh-Mehrjardi, R., Pourghasemi, H.R., Kumar, S., Bui, D.T., 2019. PMT: New analytical framework for automated evaluation of geo-environmental modelling approaches. Science of the Total Environment, 664, 296-311.

Rahmati, O., Naghibi, S.A., Shahabi, H., Bui, D.T., Pradhan, B., Azareh, A., Rafiei-Sardooi, E., Samani, A.N., Melesse, A.M., 2018. Groundwater spring potential modelling: Comprising the capability and robustness of three different modeling approaches. J. Hydrol. 565, 248–261. doi:10.1016/J.JHYDROL.2018.08.027

Rahmati, O., Panahi, M., Ghiasi, S.S., Deo, R.C., Tiefenbacher, J.P., Pradhan, B., Jahani, A., Goshtasb, H., Kornejady, A., Shahabi, H., Shirzadi, A., 2020. Hybridized neural fuzzy ensembles for dust source modeling and prediction. Atmospheric Environment 224, 117320.

Rahmati, O., Pourghasemi, H.R., Melesse, A.M., 2016. Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran. Catena 137, 360–372. doi:10.1016/j.catena.2015.10.010

Rashki, A., Eriksson, P.G., Rautenbach, C.D.W., Kaskaoutis, D.G., Grote, W., Dykstra, J., 2013. Assessment of chemical and mineralogical characteristics of airborne dust in the Sistan region, Iran. Chemosphere 90(2), 227-236.

Rashki, A., Kaskaoutis, D.G., Eriksson, P.G., Qiang, M., Gupta, P., 2012. Dust storms and their horizontal dust loading in the Sistan region, Iran. Aeolian Research 5, 51-62.

Rashki, A., Kaskaoutis, D.G., Francois, P., Kosmopoulos, P.G., Legrand, M., 2015. Dust-storm dynamics over Sistan region, Iran: Seasonality, transport characteristics and affected areas. Aeolian Research 16, 35-48.

Sachdeva, S., Bhatia, T., Verma, A.K., 2018. GIS-based evolutionary optimized Gradient Boosted Decision Trees for forest fire susceptibility mapping. Nat. Hazards 92, 1399–1418. doi:10.1007/s11069-018-3256-5

Sannazzaro, F., Filizzola, C., Marchese, F., Corrado, R., Paciello, R., Mazzeo, G., Pergola, N., Tramutoli, V., 2014. Identification of dust outbreaks on infrared MSG-SEVIRI data by using a Robust Satellite Technique (RST). Acta Astronautica, 93, 64-70.

Sayer, A.M., Munchak, L.A., Hsu, N.C., Levy, R.C., 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. Journal of Geophysical Research: Atmospheres 119(24), 13-965.

Schepanski, K., Tegen, I., Laurent, B., Heinold, B., Macke, A., 2007. A new Saharan dust source activation frequency map derived from MSG-SEVIRI IR-channels. Geophysical Research Letters 34(18). https://doi.org/10.1029/2007GL030168

Schepanski, K., Tegen, I., Macke, A., 2012. Comparison of satellite based observations of Saharan dust source areas. Remote Sensing of Environment 123, 90-97.

Sharma, A.R., Kharol, S.K., Badarinath. K.V.S., 2009. Satellite observations of unusual dust event over North-East India and its relation with meteorological conditions. Journal of Atmospheric and Solar-Terrestrial Physics 71(18), 2032-2039.

Shotyk, W., Bicalho, B., Cuss, C.W., Duke, M.J.M., Noernberg, T., Pelletier, R., Steinnes, E. and Zaccone, C., 2016. Dust is the dominant source of "heavy metals" to peat moss (Sphagnum fuscum) in the bogs of the Athabasca Bituminous Sands region of northern Alberta. Environment International 92, 494-506.

Singh, A., Tiwari, S., Sharma, D., Singh, D., Tiwari, S., Srivastava, A.K., Rastogi, N., Singh, A.K., 2016. Characterization and radiative impact of dust aerosols over northwestern part of India: a case study during a severe dust storm. Meteorology and Atmospheric Physics 128(6), 779-792.

Singh, R.P., Prasad, A.K., Kayetha, V.K., Kafatos, M., 2008. Enhancement of oceanic parameters associated with dust storms using satellite data. Journal of Geophysical Research: Oceans 113(C11). https://doi.org/10.1029/2008JC004815.

Sissakian, V., Al-Ansari, N., Knutsson, S., 2013. Sand and dust storm events in Iraq. Nat. Sci. 5, 1084e1094. http://dx.doi.org/10.4236/ns.2013.510133.

Srivastava, A.K., Soni, V.K., Singh, S., Kanawade, V.P., Singh, N., Tiwari, S., Attri, S.D., 2014. An early South Asian dust storm during March 2012 and its impacts on Indian Himalayan foothills: A case study. Science of the Total Environment 493, 526-534.

Stohl, A., Eckhardt, S., Forster, C., James, P., Spichtinger, N., Seibert, P., 2002. A replacement for simple back trajectory calculations in the interpretation of atmospheric trace substance measurements. Atmospheric Environment 36(29), 4635-4648.

Tehrany, M.S., Pradhan, B., Jebur, M.N., 2014. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. J. Hydrol. 512, 332-343.

Tehrany, M.S., Pradhan, B., Mansor, S., Ahmad, N., 2015. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. Catena 125, 91–101. doi:10.1016/j.catena.2014.10.017

Tien Bui D, PhamBT, Nguyen QP, Hoang N-D 2016b. Spatial prediction of rainfall-induced shallow landslides using hybrid integration ap- proach of least-squares support vector machines and differential evolution optimization: A case study in Central Vietnam. Int. J. Digital Earth 8947, 1–21. doi:10.1080/17538947.2016.1169561

Tien Bui, D., Le, K.-T., Nguyen, V., Le, H., Revhaug, I., 2016a. Tropical Forest Fire Susceptibility Mapping at the Cat Ba National Park Area, Hai Phong City, Vietnam, Using GIS-Based Kernel Logistic Regression. Remote Sens. 8, 347. doi:10.3390/rs8040347

Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., Dick, O.B., 2012. Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): A comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. Catena 96, 28–40. https://doi.org/10.1016/j.catena.2012.04.001

Tien Bui, D.T., Pradhan, B., Revhaug, I., Nguyen, D.B., Pham, H.V., Bui, Q.N., 2013. A novel hybrid evidential belief function-based fuzzy logic model in spatial prediction of rainfall-induced shallow landslides in the Lang Son city area (Vietnam). Geomatics, Nat. Hazards Risk 5705, 1–30. https://doi.org/10.1080/19475705.2013.843206

Torres, O., Tanskanen, A., Veihelmann, B., Ahn, C., Braak, R., Bhartia, P. K., et al. (2007). Aerosols and surface UV products form Ozone Monitoring Instrument observations: An overview. Journal of Geophysical Research Atmospheres, 112(24), 1–14. doi:10.1029/2007JD008809

Vapnik, V., 1998. Statistical learning theory. New York: Willey

Wang, Y.Q., Zhang, X.Y., Arimoto, R., Cao, J.J., Shen, Z.X., 2005. Characteristics of carbonate content and carbon and oxygen isotopic composition of northern China soil and dust aerosol and its application to tracing dust sources. Atmospheric Environment 39(14), 2631-2642.

Washington, R., Todd, M., Middleton, N.J., Goudie, A.S., 2003. Dust-storm source areas determined by the total ozone monitoring spectrometer and surface observations. Annals of the Association of American Geographers 93(2), 297-313.

Webb, N.P., Strong, C.L., 2011. Soil erodibility dynamics and its representation for wind erosion and dust emission models. Aeolian Research 3(2), 165-179.

WMO, UNEP. 2013. Establishing a WMO sand and dust storm warning advisory and assessment system regional node for West Asia: Current capabilities and needs. In: WMO Technical Report, 1121.

Yang Y, Squires V, Lu Q. Global alarm: Dust and sandstorms from the world's drylands. United Nations Convention to Combat Desertification, Bangkok; 2001. p. 345.

Yang, X., He, Q., Ali, M., Huo, W., Liu, X., Strake, M., 2012. A field experiment on dust emission by wind erosion in the Taklimakan Desert. Acta Meteorologica Sinica 26(2), 241-249.

Yu, H.L., Yang, C.H. and Chien, L.C., 2013. Spatial vulnerability under extreme events: A case of Asian dust storm's effects on children's respiratory health. Environment International 54, 35-44.

Zêzere, J.L., Pereira, S., Melo, R., Oliveira, S.C., Garcia, R.A., 2017. Mapping landslide susceptibility using data-driven methods. Science of the Total Environment 589, 250-267.

Zhang, B., Tsunekawa, A., Tsubo, M., 2008. Contributions of sandy lands and stony deserts to long-distance dust emission in China and Mongolia during 2000–2006. Global and Planetary Change 60(6), 487−504.

Zhang, X., Zhao, L., Tong, D., Wu, G., Dan, M., Teng, B., 2016. A systematic review of global desert dust and associated human health effects. Atmosphere 7(12), p.158.

Zoljoodi M, Didevarasl A, Saadatabadi AR. 2013. Dust events in thewestern parts ofIran and the relationship with drought expansion over the dust-source areas in Iraq and Syria. Atmos. Clim. Sci. 03, 321–36.

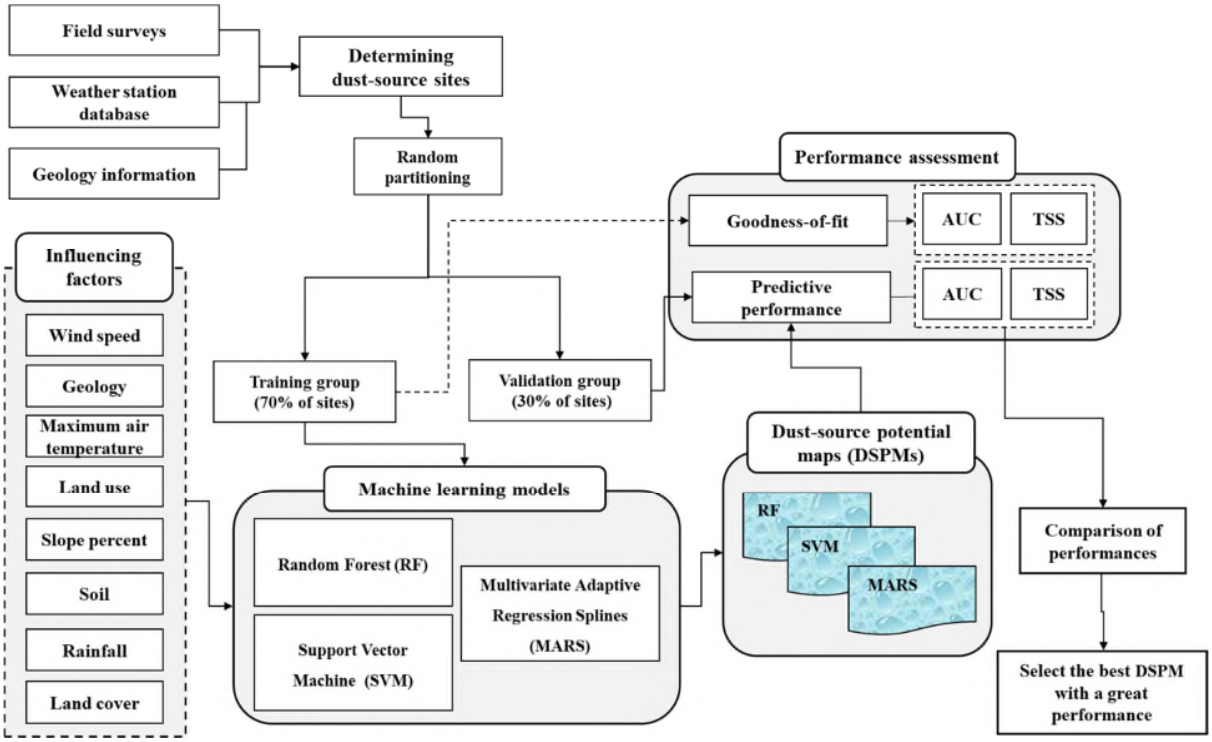

Fig. 1 Flowchart of the study.

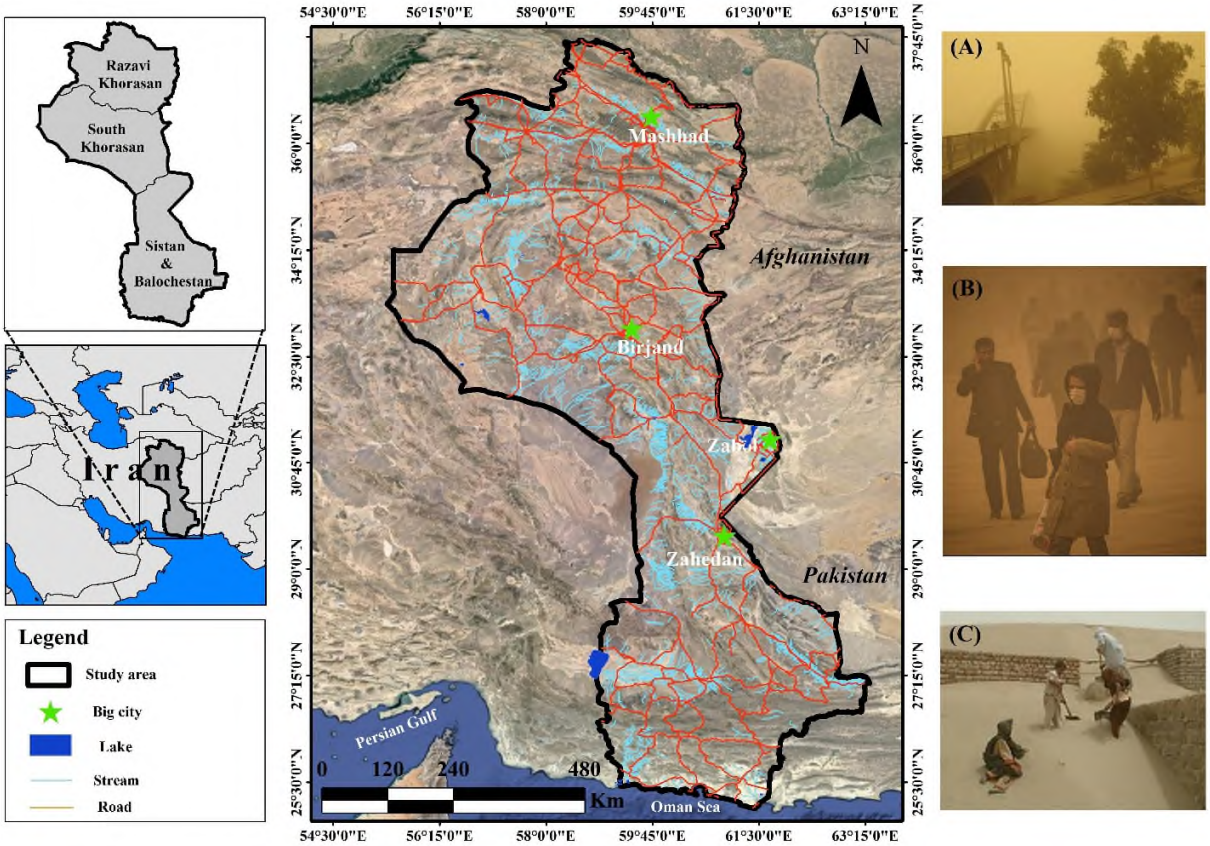

Fig. 2 Location of the study area in Iran. Photographs showing the severity of the dust storms in the study area: (A) Mashhad, (B) Zabol, and (C) Zahedan. (photographs were taken by the authors)

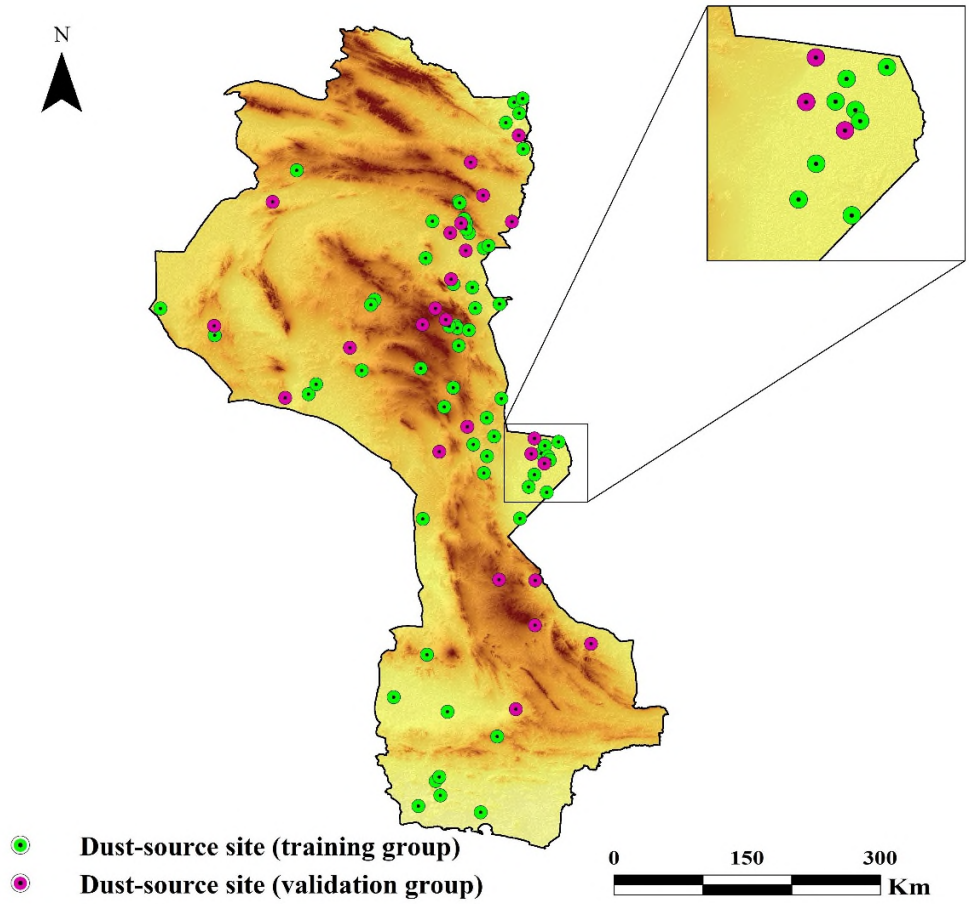

Fig. 3 Training and validation groups of dust source sites

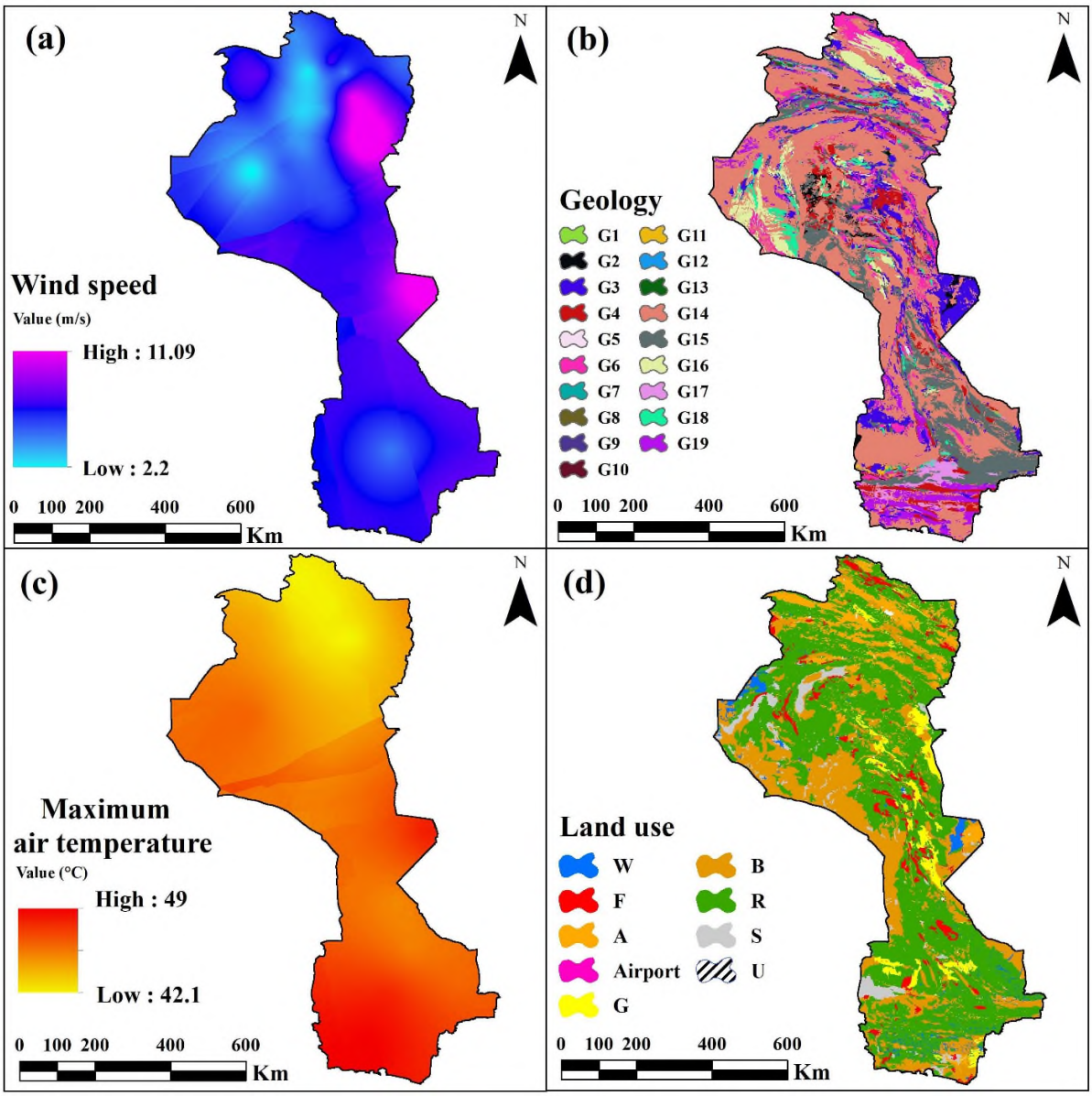

Fig. 4 Distributions of dust-production factors: (a) wind speed, (b) geology, (c) maximum air temperature, (d) land use, (e) slope, (f) soil, (g) rainfall, and (h) land cover

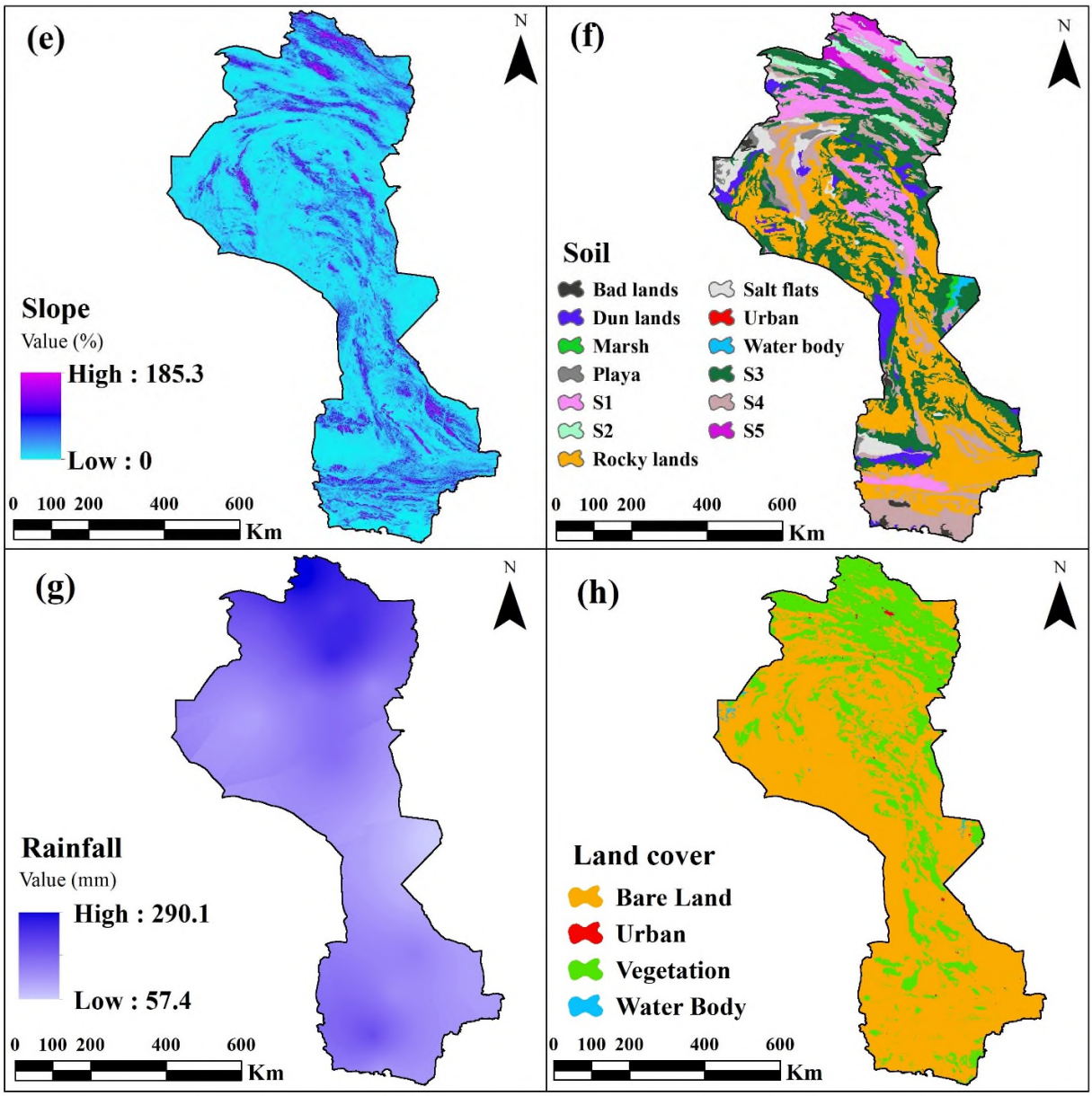

Fig. 4 (continued)

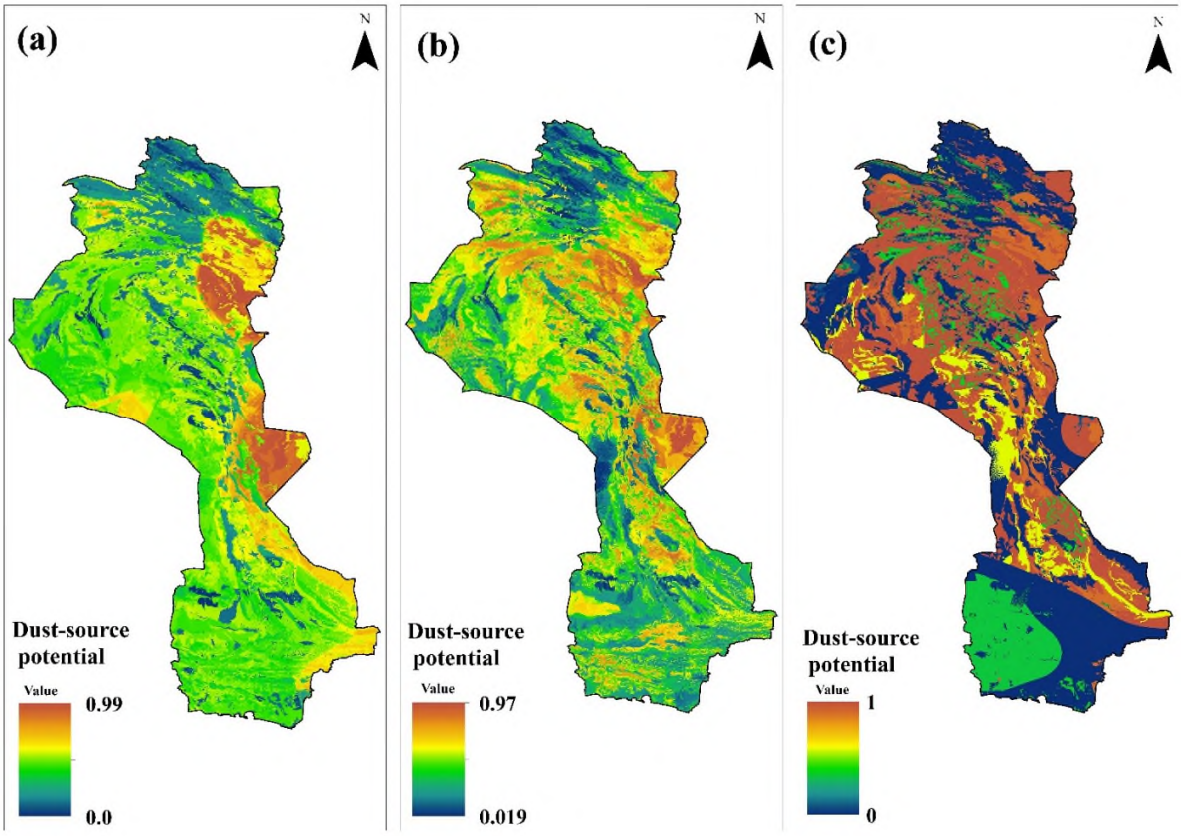

Fig. 5 The maps of dust source potential using a) RF, and b) SVM, and c) MARS

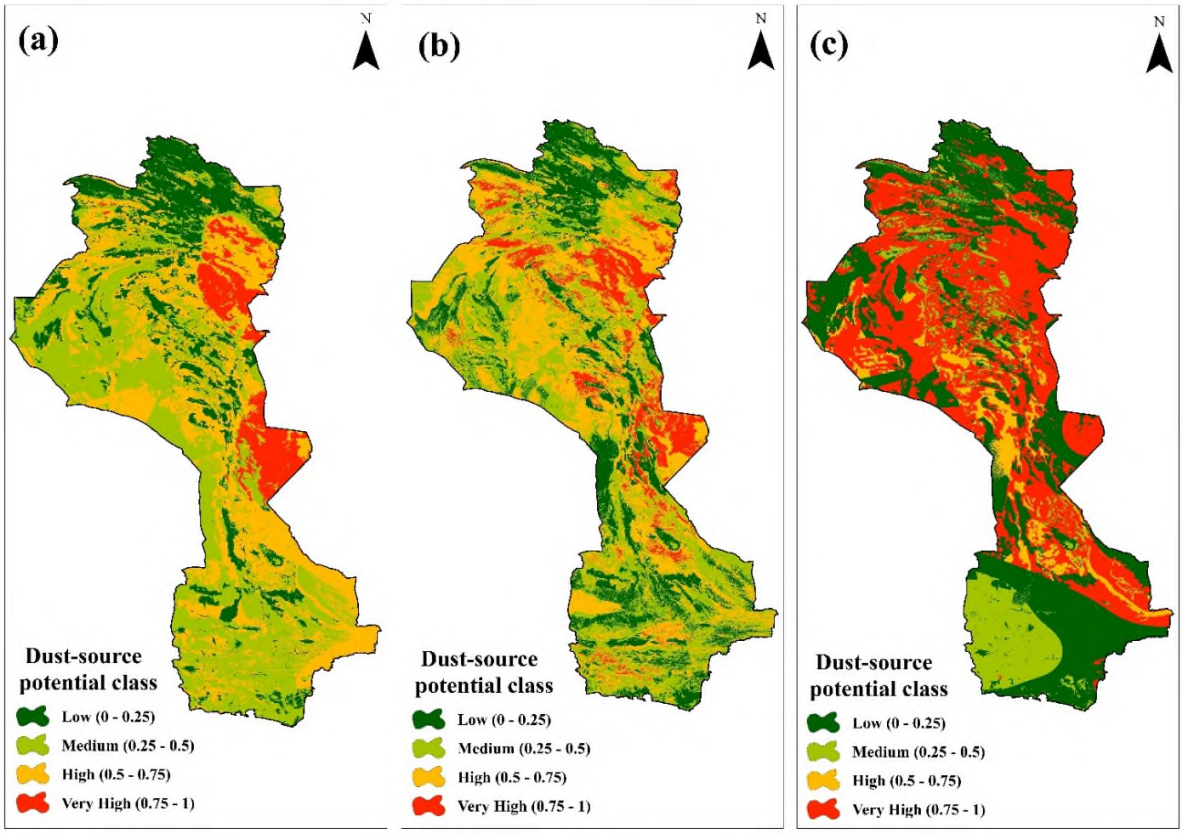

Fig. 6 The maps of dust source potential class using a) RF, and b) SVM, and c) MARS

Table 1 Lithology of the study area.

Age					Code	Geo-unit	Description	Area (%)
Eon	Era	Period		Epoch				
Phanerozoic	Cenozoic	Quaternary		Holocene	G14	Qs	Sand dunes and sand sheet	46.39
		Tertiary	Neogene	Pliocene	G3	Plms	Marl, shale, sandstone and conglomerate	9.75
				Miocene	G19	Murmg	Gypsiferous marl	6.83
				Oligocene	G4	Olm,s,c	Red and green silty, gypsiferous marl, sandstone and gypsum	4.87
					G2	L.E-Odi	Diorite	0.98
			Paleogene	Eocene	G15	Ef	Eocene flysch in general, composed of shale, marl, sandstone, conglomerate and limestone	15.27
	Mesozoic	Cretaceous			G5	ba	Basalt and basaltic andesite pillow lavas	0.04
					G7	db	Diabase	0.18
					G6	Kab	Blue - grey marl and shale	5.14
					G1	am	Amphibolite	0.0004
					G11	gsch	Glaucophane schist	0.26
		Jurassic			G16	Jks	Alternation of sandstone and shale	5.25
					G10	di-gb	Gabbro to diorite, diorite and trondhjemite	0.06
					G13	hz	Harzburgite	0.14
		Triassic			G18	TRmi	Shale and sandstone with coal seams	2.44
	Paleozoic	Permian			G17	sp1	Spilitic andesite and diabasic tuff	1.97
		Carboniferous			G9	Cs	Light olive-green shale with intercalations of quartzarenite and fossiliferous limestone	0.16
		Devonian			G8	Dp	Light red to white, thick bedded quartzarenite with dolomite intercalations and gypsum	0.08
					G12	Dsh	Alternation of shale, marl, and fossiliferous limestone, locally with intercalations of quartz arenite	0.11

Table 2 Land use of the study area

Number	Code	Land use type	Area (%)
1	W	Water body	1.88
2	F	Forest	3.79
3	A	Agriculture	7.80
4	G	Garden	0.005
5	B	Bare land	4.39
6	R	Rangeland	23.42
7	S	Salt land	54.82
8	U	Urban	3.67
9	Airport	Airport	0.18

Table 3 Soil type of the study area

Number	Code	Soil	Area (%)
1	Badlands	Badlands	0.75
2	Dune lands	Dune lands	4.15
3	Marsh	Marsh	0.26
4	Playa	Playa	1.12
5	S1	Rock Outcrops/Entisols	11.57
6	S2	Rock Outcrops/Inceptisols	2.26
7	Rocky lands	Rocky lands	31.71
8	Salt flats	Salt flats	3.58
9	Urban	Urban	0.02
10	Water body	Water body	0.35
11	S3	Aridisols	29.86
12	S4	Entisols/Aridisols	13.20
13	S5	Inceptisols	1.10

Table 4 Confusion matrix elements.

Observed	Predicted	
	Non-dust source (−)	Dust source (+)
Non-dust source (−)	(−|−) True negative (TN)	(+|−) False positive (FP; Error Type I)
Dust source (+)	(−|+) False negative (FN; Error Type II)	(+|+) True positive (TP)

Table 5 Results of variable contributions using the RF.

Variable	Percent contribution (PC)
Wind speed	29.5
Land cover	25.3
Soil type	14.8
Maximum air temperature	9.5
Land use	7.8
Rainfall	6.3
Geology	4.7
Slope	2.1

Table 6 Goodness-of-fit of models.

Evaluation criteria	Models		
	RF	SVM	MARS
AUC	0.92	0.898	0.86
TSS	0.785	0.758	0.721

Table 7 Predictive performance of models.

Evaluation criteria	Models		
	RF	SVM	MARS
AUC	0.894	0.875	0.81
TSS	0.751	0.73	0.69

Cranfield University

CERES Research Repository **https://dspace.lib.cranfield.ac.uk/**

School of Water, Energy and Environment (SWEE) Staff publications (SWEE)

Identifying source of dust aerosol using a new framework based on remote sensing and modelling

Rahmati, Omid

2020-05-19

Attribution-NonCommercial-NoDerivatives 4.0 International

Rahmati O, Mohammadi F, Ghiasi SS, et al., (2020) Identifying source of dust aerosol using a new framework based on remote sensing and modelling. Science of the Total Environment, Volume 737, October 2020, Article number 139508

https://doi.org/10.1016/j.scitotenv.2020.139508

Downloaded from CERES Research Repository, Cranfield University