https://doi.org/10.1016/j.atmosenv.2020.117320

Hybridized ANFIS Models for Dust-source Modelling

Omid Rahmati[a,b], Mahdi Panahi[c], Seid Saeid Ghiasi[d], Ravinesh C. Deo[e], John P. Tiefenbacher[f], Biswajeet Pradhan[g], Aiding Kornejady[h], Himan Shahabi[i], Ataollah Shirzadi[i], Mahmood Samadi[j], Hassan Khosravi[j], Farnoush Mohammadi[j], Arash Zare[k], Omid Asadi[k], Dieu Tien Bui[l,m,*]

[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] Department of Geophysics, Young Researchers and Elites Club, North Tehran Branch, Islamic Azad University, Tehran, Iran

[d] Young Researchers and Elites Club, Science and Research Branch, Islamic Azad University, Tehran, Iran

[e] School of Agricultural, Computational and Environmental Sciences, Centre for Sustainable Agricultural Systems & Centre for Applied Climate Sciences, University of Southern Queensland, Springfield, QLD 4300, Australia

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

[g] Center for Advanced Modeling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT, University of Technology Sydney, 2007 NSW, Australia

[h] Young Researchers and Elite Club, Gorgan Branch, Islamic Azad University, Gorgan, Iran

[i] Department of Geomorphology, Faculty of Natural Resources, University of Kurdistan, Sanandaj, Iran

[j] Department of Arid and Mountainous Regions Reclamation, Faculty of Natural Resources University of Tehran, Tehran, Iran

[k] Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran

[l] Institute of Research and Development, Duy Tan University, Da Nang 550000, Viet Nam

[m] Geographic Information System Group, Department of Business and IT, University of South-Eastern Norway, N-3800 Bø i Telemark, Norway

* Corresponding author's email address: dieu.t.bui@usn.no

Abstract

Dust storms are believed to play important roles in many climatological, geochemical, and environmental processes. This particular atmospheric phenomenon can have a significant negative impact on public health and significantly disturb natural ecosystems. Identifying dust source areas is thus a fundamental task necessary to control the effects of this hazard. In this study, a new methodology based on hybridized machine-learning algorithms is developed to identify dust source areas. Each hybridized model, designed as an intelligent system, consists of an adaptive neuro-fuzzy inference system (ANFIS), integrated with a combination of metaheuristic optimization algorithms: the bat algorithm (BA), cultural algorithm (CA), and the differential evolution (DE) algorithm. The data acquired from two key sources – the Moderate Resolution Imaging Spectroradiometer (MODIS) Deep Blue and the Ozone Monitoring Instrument (OMI) – are incorporated into the hybridized model, along with relevant data from field surveys and dust samples from the study region. Goodness-of-fit analyses are performed to evaluate the predictive capability of the hybridized models using different statistical criteria, including the true skill statistic (TSS) and the area under the receiver operating characteristic curve (AUC). The results demonstrate that the hybridized ANFIS-DE model (with AUC=84.1%, TSS=0.73) outperforms other hybridized models tailored for dust-storm prediction. The results suggest that the hybridized ANFIS-DE model can be adopted as a promising, cost-effective method for efficiently identifying the dust source areas, with benefits for both public health and natural environments where excess dust presents a significant challenge.

Introduction

Dust storms are natural atmospheric events that occur mainly in arid areas, reducing air quality and visibility[24]. Dust is comprised of large-grained particulate matter (PM) that is light enough to be entrained by horizontal atmospheric flows. But dust storms also carry minute and fine-grained solid matter that is small enough to be more easily elevated aloft and carried by prevailing winds. The occurrence of dust storms has increased in recent years, providing compelling evidence that dust particles are carried long distances[5,18].

Dust storms are integral to Earth's natural systems and have impacts that are numerous and wide-ranging. These include effects on air chemistry, soil characteristics, water quality, nutrient dynamics, and biogeochemical cycling in both oceanic and terrestrial environments[18,23]. Local and regional climates can be affected by dust storms for the scattering and absorption of solar radiation by dust particles, but the impacts can extend great distances from the sources of dust. Dust can modify the microphysical properties of clouds and change precipitation efficiency. In sum, dust storms can affect atmospheric conditions at many scales[41,47].

Airborne PM is a health-damaging airborne pollutant that can adversely affect the human cardiovascular system and can cause respiratory problems[5]. Inhalation of PM can also exacerbate various diseases and trigger health issues such as asthma in children and elderly ultimately increasing morbidity[17]. Pathogenic and non-pathogenic microorganisms (including Coxiella Burnetii, Mycobacterium, Aspergillus, Mycobacterium, Brucella, Cladosporium, Actinomycetes, Clostridium perfingens, and Bacillus), toxins, and influenza viruses can adhere to dust particles and can be transported to great distances[8,21,34]. Moreover, metallic elements are transported as inhalable dust particles, and these could potentially affect the respiratory tracts and can cause neurological and other physiological impacts[25,43]. In addition to health impacts, there are economic impacts from sand and dust storms. Agricultural crops and livestock have been destroyed by dust and sandstorms[36].

Dust particles emitted from different sources are likely to affect plant life in different ways (Supe and Gawande, 2015). The largest sources of dust in Earth's atmosphere are from the Sahara and Sahel regions of North Africa (so called "African dust"), the Gobi, Taklamakan, and Badain Juran deserts of Asia ("Asian dust"), and Australian desert environments ("Australian dust")[9,40]. Asian dust particles can also migrate globally, perhaps circumnavigating Earth in as little as 13 days, as recorded in the French Alps[11] and in ice and snow cores from Greenland[4]. Recent changes to regional climates have considerably increased the frequency of dust storm events in the Middle East[47].

In the view of the hazardous effects of dust storms, new measures are needed to proactively identify and control their genesis regionally. Furthermore, it is also important for all sectors to mitigate the catastrophic effects of dust storms.

Though dust has long been known to be important in weather processes and storms and can influence local weather, prediction of dust and sandstorms is rudimentary and not effective. Despite sophisticated weather models, it remains difficult to forecast the entrainment and transport of dust in the lower atmosphere. One reason for this a limited understanding of the distribution of the sources of dust and their behaviors with respect to their spatiotemporal volatility in response to various activities and processes[6]. For the analysis of dust sources, and the modelling of their impact on Earth's natural system, it is crucial to identify the spatial and temporal diffusion rates of sources[6]. In a number of previous studies, a diverse range of remotely operating methods have been used to identify dust source areas including, but not limited to: (1) remote sensing analysis, (2) horizontal visibility, (3) mineralogy of dust samples, and (4) Lagrangian back-trajectory[2]. The drawbacks of each have been discussed in Schepanski *et al.*[33]. Although these approaches provide useful information regarding the potential sources of dust, coupling and analysis of geo-environmental and weather conditions to recognize the dust sources over large areas is relatively difficult.

To cope with this problem, several artificial-intelligence models using machine-learning techniques have been developed in the context of geo-environmental research, however their capacity to deduce the presence of and to model the movements of dust sources in different regions has not yet been evaluated. To address this significant gap in dust-storm prediction methodologies, this study develops hybridized artificial intelligence models using an adaptive neuro-fuzzy inference system (ANFIS) where a metaheuristic optimization algorithm is used to improve the predictive model. Field investigations were conducted and statistical analyses were performed to identify dust-source areas in the eastern part of Iran, particularly in three arid provinces: Razavi Khorasan, Jonobi Khorasan, and Sisstan-Balochestan (Figure 1).

The study region covers an area of 444,904 km^2 and forms a homogenous geographical unit that shares certain characteristics: proximity to the eastern deserts of the Iran plateau, the variability and deficiency of precipitation, desertification, high evaporation rates, and lack of permanent surface water bodies. The climate of this region is relatively warm and dry. Wind is more frequent here than other parts of the country, with days with wind numbering 120; a significant feature of this region. To develop a predictive model for dust storms, the hybridization of the respective models was carried out with an ANFIS in combination with a metaheuristic optimization algorithm that includes the bat algorithm (BA), cultural algorithm (CA), and the differential evolution (DE) approaches. A hybridized modeling framework integrated several modeling approaches and thus achieved a superior model performance with an efficient computing time. The method presented here can be used to improve our understanding of dust sources in various areas in other arid and semi-arid regions.

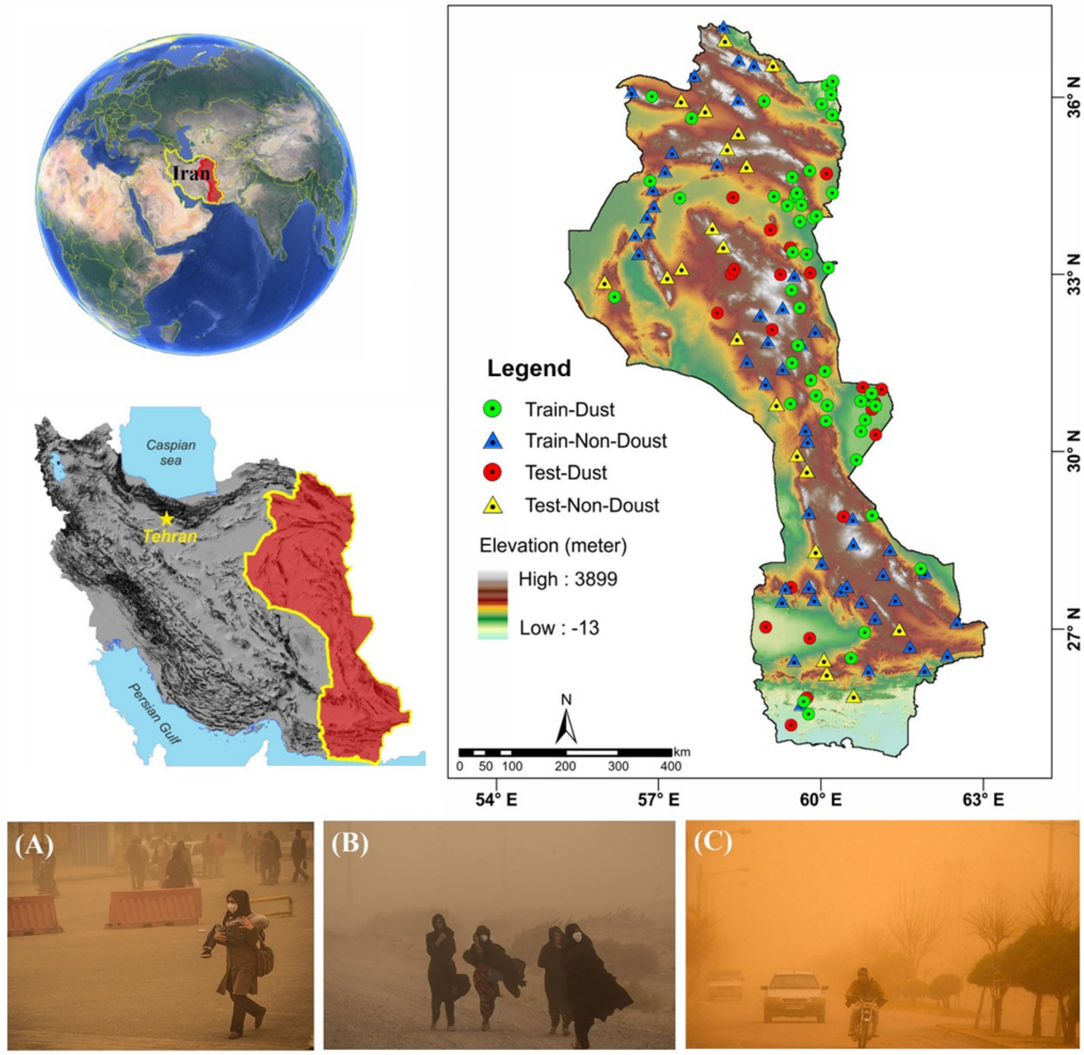


Figure 1. A map of the present study area and field photographs of some dust storms that have occurred in the study area at locations in (A) Zabol, (B) Zahedan, and (C) Iranshahr.

Results and discussion

Preparation of potential maps of dust-sources

The spatial distribution of potential dust sources derived from standalone ANFIS models, and from equivalent hybridized models in which optimization algorithms are used are illustrated in Figure 2. Upon initial inspection, the spatial distribution of potential dust sources seems to be clearly differentiated across the study area. Notably, all four predictive models (i.e., the standalone ANFIS, as well as the ANFIS-BA, ANFIS-CA, and ANFIS-DE hybridized models) reveal a relatively similar spatial pattern of dust potential across the study region. The northern, eastern, and southwestern parts of the region are highly active dust-production sources, while the central parts show significantly less dust-potential and is a rather low-dust zone.

Visual comparison of the enlarged insets clipped from the dust-potential maps reveals the less precise classifications dust-potential produced by the standalone ANFIS model (Fig. 1a inset), particularly in areas without original source-data. The hybridized models produce a clearer and more precise differentiation of localities with and without dust storms. This is discernible in the proportional distribution of the dust-potential classes each hybridized model generates (Table 1).

The ANFIS model has classified nearly 69% of the region as highly dust storm active, which contradicts empirical evidence of dust storms in this region. These predictions are of little practical value to guide pragmatic action to mitigate the impacts of dust storms. Conversely, the areas classified as ‘high’ and ‘very high’ by the hybridized models are smaller proportions of the whole; they present more realistic representations of dust storm occurrence. This attests to the enhancement that optimized ANFIS models provide for more differentiation between classes and therefore, perhaps, a more accurate solution.

Table 1 The area of dust-source potential classes in different models (in percent)

Model Type	Model Name	Very low	Low	Medium	High	Very high
Standalone	ANFIS	22.30	1.56	7.49	60.25	8.38
Hybridized Models	ANFIS-BA	5.60	17.95	29.97	33.60	12.85
	ANFIS-CA	3.28	18.47	31.10	24.41	22.71
	ANFIS-DE	2.95	15.82	29.40	36.78	15.1

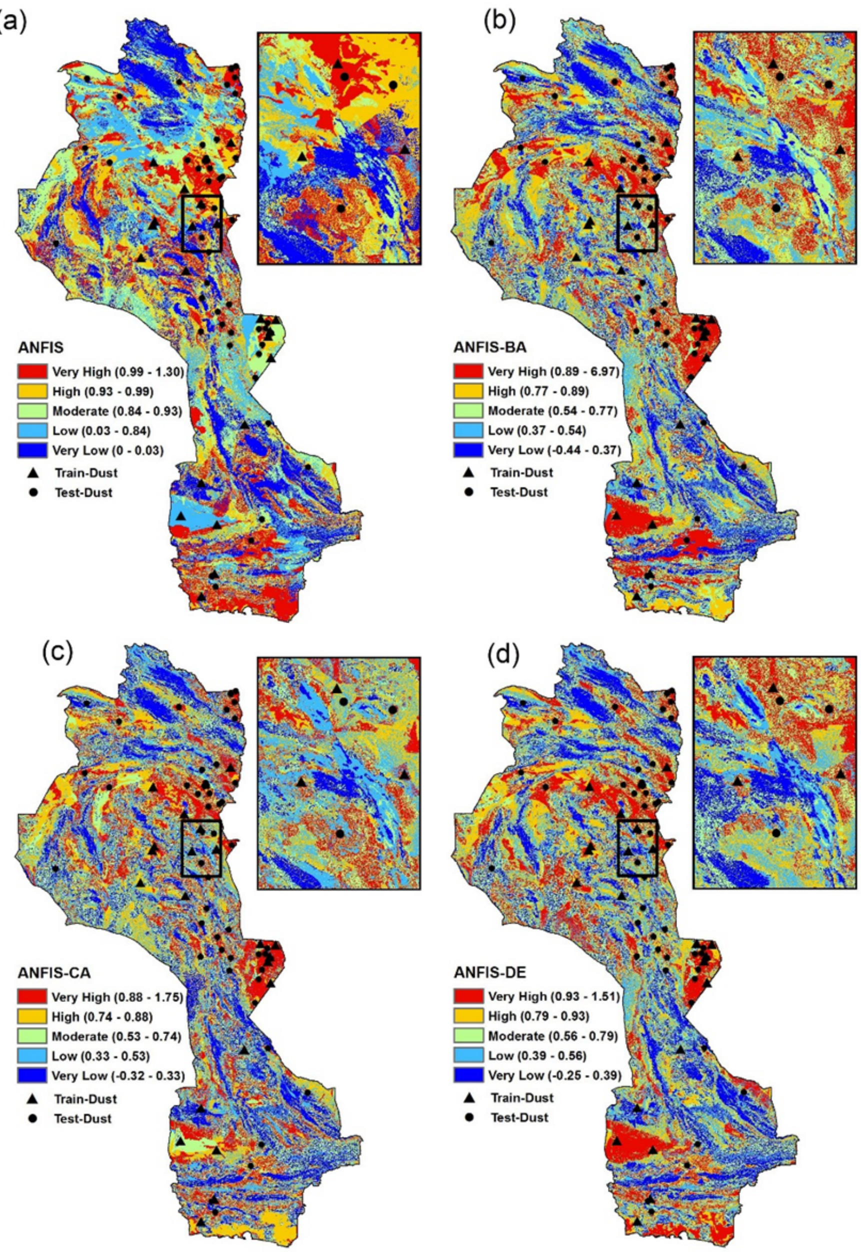


Figure 2. Dust-source potential mapping prepared by hybridized and standalone ANFIS models: a) ANFIS, b) ANFIS-BA, c) ANFIS-CA, and d) ANFIS-DE

Validation and comparison of the novel hybridized- and standalone-ANFIS models

To determine the accuracy of the ANFIS models, a goodness-of-fit test based on the training and validation datasets using three new hybridized models for spatial prediction of dust storms was assessed with the values of the mean-square error (MSE), root mean-square error (RMSE), mean, and standard deviation (StD) metrics from the observed and predicted data (Figures 3 – 6). All performance metrics (i.e., MSE, RMSE, & StD) produced by the ANFIS model with the training dataset were 0

(Figure 2b, c). However, the values generated by the validation dataset for the MSE, RMSE, mean, and StD were 0.072, 0.269, 0.018, and 0.271, respectively (Figure 3e, f), indicating that the model over-fit the training dataset during its learning stage. These results clearly demonstrate the tendency of the standalone ANFIS model to over-fit, as was shown in Tien Bui *et al.*[38]. By contrast, in the ANFIS-BA model, the values of 0.023, 0.153, 0.06, and 0.154 were obtained for the MSE, RMSE, mean, and StD, respectively, in the training phase (Figure 4b, c). The values for the same variables generated with validation data were 0.020, 0.143, 0.013, and 0.144, respectively (Figure 4e, f).

Similarly, the values for the MSE, RMSE, mean, and StD obtained with the training dataset as input into the hybridized ANFIS-CA model were 0.021, 0.146, 0.016, and 0.146, respectively (Figure 5b, c). And for the validation dataset, they were 0.022, 0.149, 0.010, and 0.150, respectively (Figure 5e, f). For the hybridized ANFIS-DE model, the training-data generated values for MSE, RMSE, mean, and StD were 0.016, 0.126, 0.005, and 0.127, respectively (Figure 6b, c) and the validation-data values were 0.020, 0.142, -0.016, and 0.143, respectively (Figure 6e, f). In this regards, as was determined by Bui *et al.*[37,38], we have demonstrated that a hybridized-ANFIS model is a more robust predictive model for dust-storm prediction, as it attained greater accuracy than with the standalone-ANFIS model.

Therefore, it is evident that as MSE and RMSE values diminish, goodness-of-fit increases, as does the overall performance for each optimized hybridized-ANFIS model. In terms of performance among these models, the ANFIS-DE model performed the best, and was followed by the ANFIS-BA, ANFIS-CA, and the ANFIS models. As discussed by Khazraee *et al.*[19], the use of the differential evolution (DE) algorithm is likely to generate a more robust and efficient optimization tool for any predictive model, given its ability to perform a direct search of data features without requiring any derivative estimation or assumptions. This explains the enhanced performance of the ANFIS-DE hybridized model.

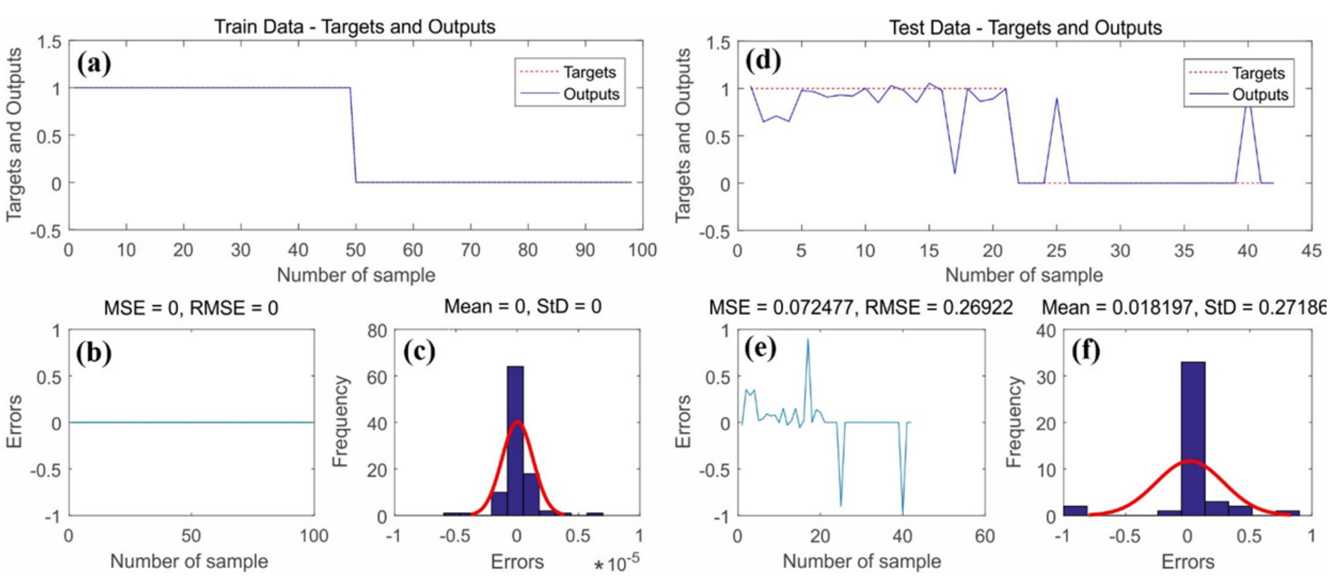


Figure 3. Schematic of the ANFIS model: a) target and output ANFIS value of training data samples, b) MSE and RMSE value of training data samples, c) frequency errors of training data samples, d) target and output ANFIS value of testing data samples, e) MSE and RMSE value of testing data samples, and f) frequency errors of testing data samples

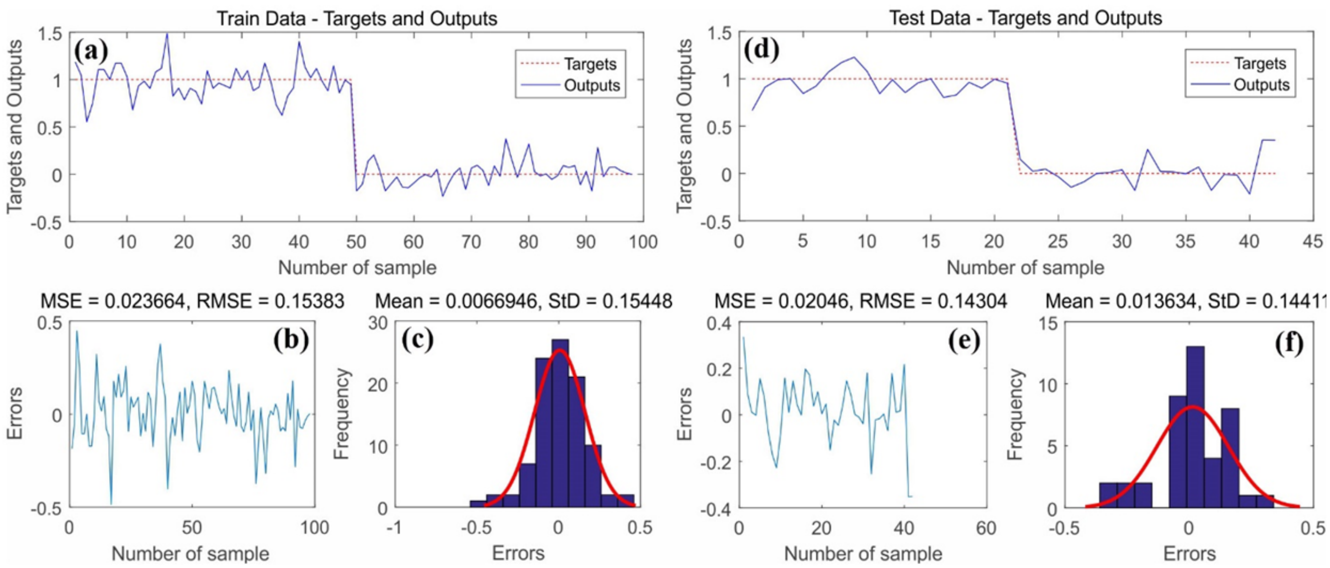


Figure 4. Schematic of the ANFIS-BA model: a) target and output ANFIS-BA value of training data samples, b) MSE and RMSE value of training data samples, c) frequency errors of training data samples, d) target and output ANFIS-BA value of testing data samples, e) MSE and RMSE value of testing data samples, and f) frequency errors of testing data samples

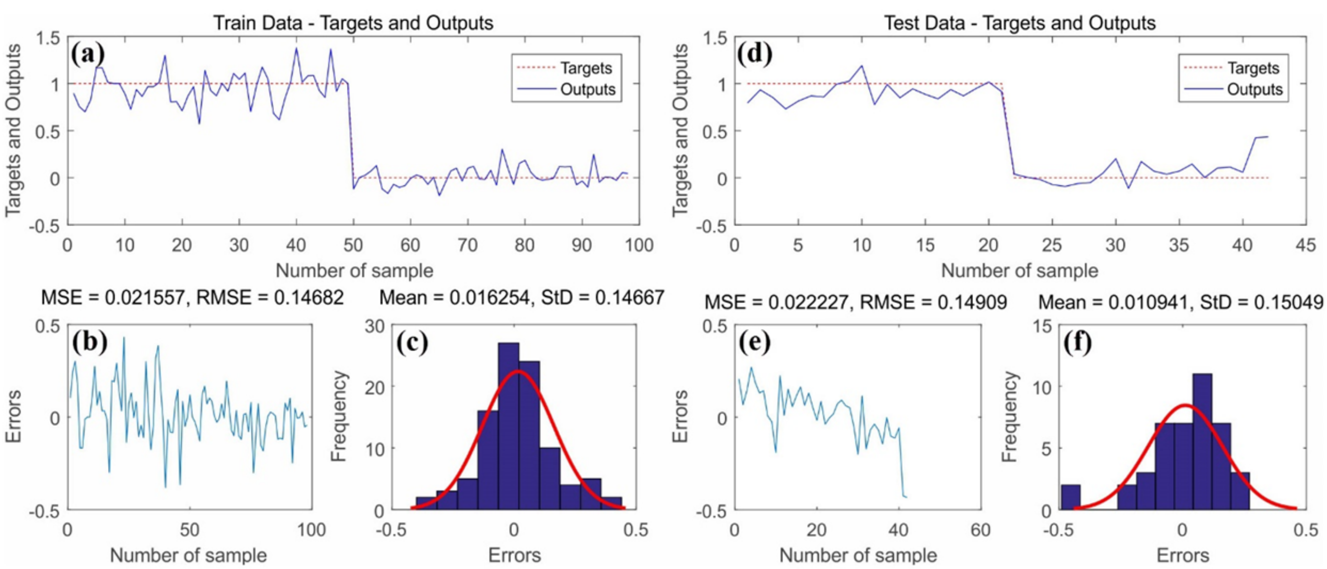


Figure 5. Schematic of the ANFIS-CA model: a) target and output ANFIS-CA value of training data samples, b) MSE and RMSE value of training data samples, c) frequency errors of training data samples, d) target and output ANFIS-CA value of testing data samples, e) MSE and RMSE value of testing data samples, and f) frequency errors of testing data samples

To evaluate the validity of the models for dust storm prediction, the resulting susceptibility maps were evaluated for validity spatially. We tested the accuracy of the prediction of dust storms that have occurred and those that are expected to occur using the training and validation datasets. The results show that the AUC in the training step (i.e., a measure of the goodness-of-fit) were 88.1%, 84.9%, 83.0%, and 85.4% for the ANFIS, ANFIS-BA, ANFIS-CA and ANFIS-DE models, respectively. These values in the validation step (i.e., predictive performance) were 63.7%, 83.4%, 80.3%, and 84.1%, respectively (Table 2).

Another statistical metric applied to validate the dust-susceptibility maps is the true skill statistic (TSS). Accordingly, the training TSS for ANFIS, ANFIS-BA, ANFIS-CA and ANFIS-DE models were found to be 0.78, 0.74, 0.73, and 0.75, respectively. Slightly lower values of 0.64, 0.72, 0.70 and 0.73 were produced with the validation dataset. Though the AUC and TSS metrics produced from the training data and the ANFIS model had the highest performance, ANFIS-BA's metrics using the validation dataset indicated the highest power of prediction. Therefore, the best hybridized models in order of performance are: ANFIS-DE, ANFIS-CA, and ANFIS.

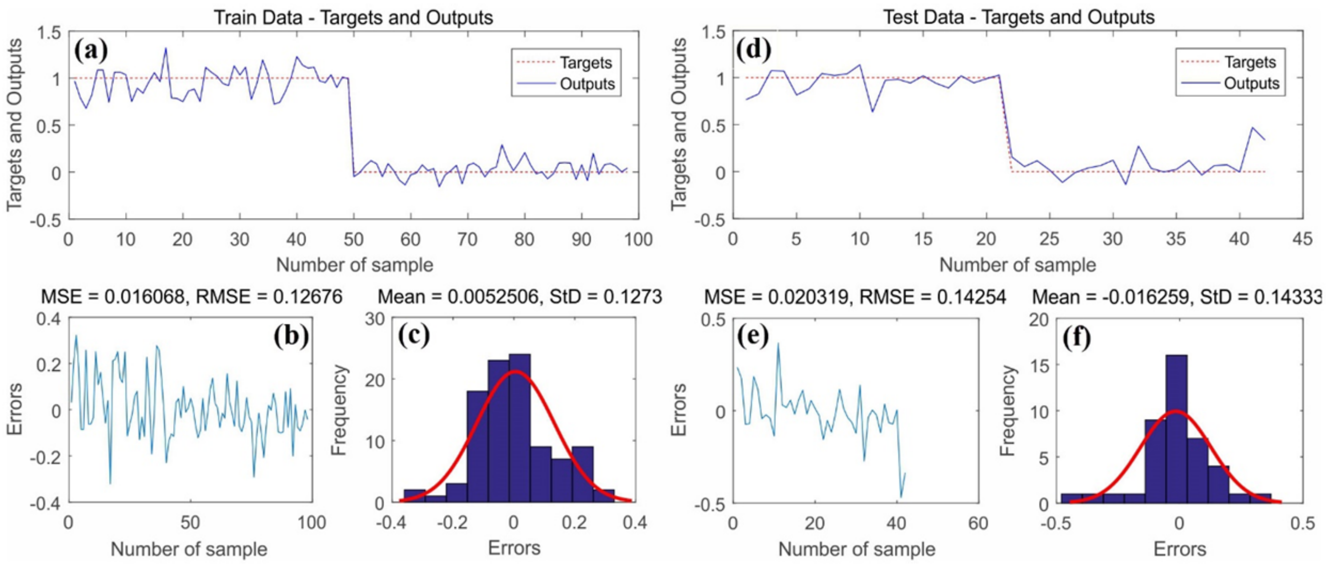


Figure 6. Schematic of the ANFIS-DE model: a) target and output ANFIS-DE value of training data samples, b) MSE and RMSE value of training data samples, c) frequency errors of training data samples, d) target and output ANFIS-DE value of testing data samples, e) MSE and RMSE value of testing data samples, and f) frequency errors of testing data samples

Table 2 Goodness-of-fit and predictive performance of hybrid and individual models based on AUC and TSS metrics.

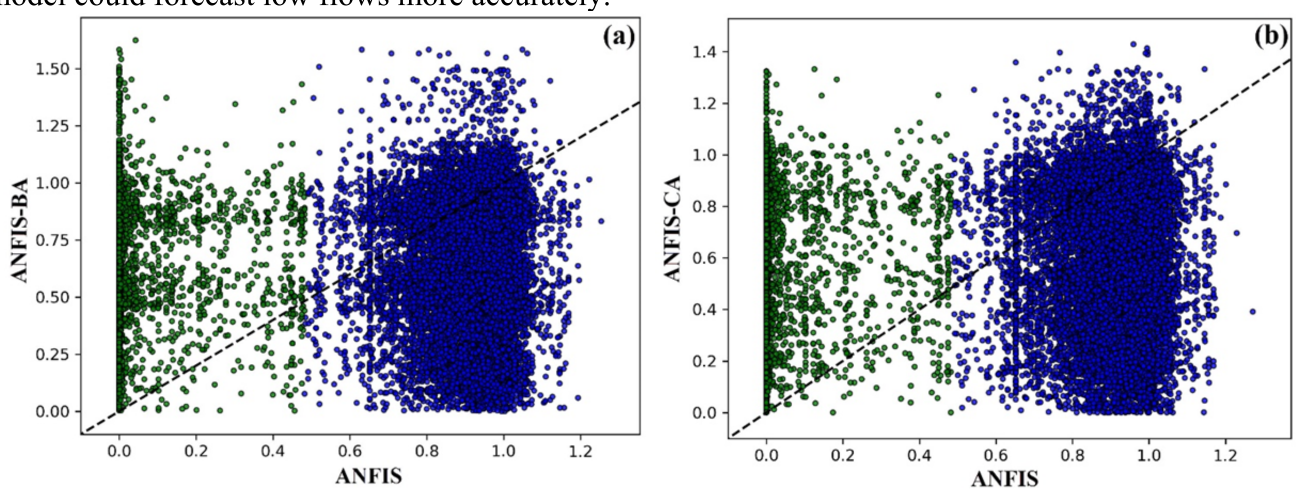
Model Type	Model Name	AUC (%)		TSS	
		Training	Validation	Training	Validation
Standalone	ANFIS	88.1	63.7	0.78	0.64
Hybridized Models	ANFIS-BA	84.9	83.4	0.74	0.72
	ANFIS-CA	83.0	80.3	0.73	0.7
	ANFIS-DE	85.4	84.1	0.75	0.73

Comparison of the models' predictions

A set of scatter plots of the standalone ANFIS model predictions versus those of the hybridized ANFIS model predictions together with the best (1:1) line is presented in Figure 7. The distribution of points very close to and evenly at both sides of the 1:1 line implies a high degree of agreement between the two data series (i.e., the predictions of ANFIS and the hybridized models are shown accordingly). Such a pattern is not discernable in the plots above, indicating that there is almost no agreement between the predictions of the ANFIS and hybridized ANFIS models. However, two distinct point-patterns are visually discernable on the plots, which are grouped as two clusters of points using the cluster analysis. Most of the high values predicted by the ANFIS model (roughly higher than 0.5 on the *x*-axis) lie below the 1:1 line, which means that they are under-predicted by the hybrid models. In contrast, most of the low values produced by the ANFIS model (values lower than 0.5 on the x-axis) are over-predicted by the hybrid models.

Overall, the ANFIS model tends to generate results with more extreme outliers, while the hybridized models seem to produce predictions with outliers that have more moderate values. Although this does not prove that the hybridized ANFIS models perform significantly better than the standalone ANFIS model, there is a significant difference between the prediction patterns of the standalone ANFIS and hybridized ANFIS models. Since ANFIS by itself has not yet applied to topics in this field of study, a comparison to results from previous studies is not possible.

However, several other studies in environmental and hydrological fields have demonstrated that hybridized ANFIS models can improve prediction of extreme observed values compared to a standalone ANFIS model. For example, Yaseen *et al.*[45,46] found that the standalone ANFIS model integrated with the firefly optimization algorithm (ANFIS-FFA) was able to capture heavy to extreme rainfall events more accurately than did a standard, non-optimized ANFIS model. In a study of streamflow forecasting, the same authors demonstrated that although both standalone- and hybridized-ANFIS models were able to forecast peak stream flow data points quite successfully, the hybrid ANFIS model could forecast low flows more accurately.



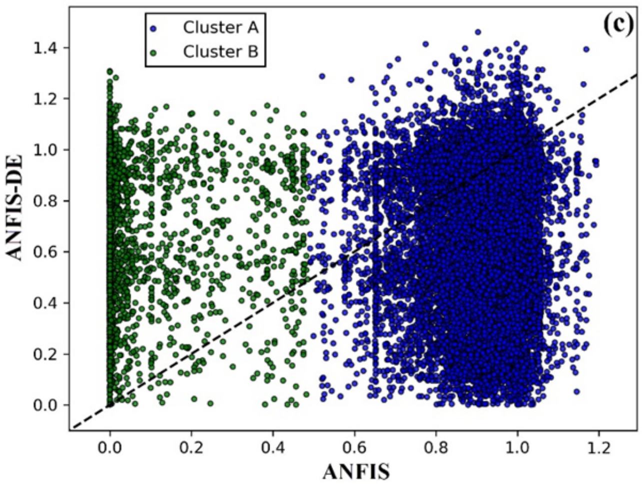


Figure 7. Results of cluster analysis: a) ANFIS-BA versus ANFIS, b) ANFIS-CA versus ANFIS, and c) ANFIS-DE versus ANFIS

Conclusion

In this paper, an ANFIS model was developed and hybridized with model-optimization algorithms to perform a comparative analysis for the identification of dust source areas. The four state-of-the-art models tested are: a) standalone ANFIS; and three equivalent hybridized models – b) ANFIS-BA, c) ANFIS-CA, and d) ANFIS-DE. The resulting dust-source maps were validated using actual field data and statistical metrics comparing predicted and observed dust-source datasets divided into training and validation subsets. Various model parameters – historical dust-storm data, high-speed wind event data, soil types, air temperatures, geomorphic units, slope, land use, and rainfall – were used as predictive factors to map the potential source areas of dust. The results show that sedimentary rock deposits were the most frequent generators of dust, primarily due to their dominant spatial extent and the presence of high-speed winds to generate erosion. A number of factors contribute to dust generation in the study area, so the study of these factors is of prime importance in the region. Aeolian abrasion is the primary process that produces dust particles in the high-wind areas. Based upon the models developed, we demonstrate that there is significant potential for increased dust mainly because of the interaction of the contributing factors that initiate and fuel dust production.

These results show that the proposed ANFIS hybridized models can be used to map the source areas of dust on a regional scale, creating pathways for assessments of dust-storm potential and for examining the effects of these storms on human health and the environment. Notably, the four ANFIS models generated discernibly strong predictive performances as indicated by the AUC and TSS statistical tests: standalone ANFIS (AUC=63.7%, TSS=0.64), followed by hybridized ANFIS-BA (AUC=83.4%, TSS=0.72), ANFIS-CA (AUC=80.3%, TSS=0.7), and ANFIS-DE (AUC=84.1%, TSS=0.73). These accuracy assessments demonstrate the enhanced effectiveness of hybridized algorithms (relative to a standalone ANFIS model) to identify dust-source locations. The results of this study are likely to attract the attention of local environmental and health agencies and national governmental bodies to identify and mitigate dust sources and to transfer the methods for examination of other regions that may be experiencing similar issues. This new dust-storm potential modeling approach that considers geo-environmental factors at high resolution can be replicated in other areas to identify current and future dust sources.

Methodology

Dust source inventory

This study used two common datasets to identify dust sources in the study region: the "Moderate Resolution Imaging Spectroradiometer (MODIS)" Deep Blue and the Ozone Monitoring Instrument (OMI). These have been widely applied in previous research as they are cost-effective and robust sources of data[7,29]. We investigated dust storms using the previously described indices between April 2014 and May 2018. After May 2018, several field surveys were conducted and geo-environmental and terrain characteristics were identified and investigated. A total of 85 dust source areas were detected and geolocated with a GPS receiver. Those locations were randomly divided into two groups for training (n=56 or 70%) and for validation (n=29 or 30%) of the models (Figure 1).

Factors that influential dust generation

There is no predetermined set of geo-environmental and topographical factors known to be linked to dust source areas. According to field investigations and previous studies, eight factors – wind speeds, geology, maximum air temperatures, land uses, slopes, soils, precipitation amounts, and land cover were considered to be possible predictive factors for modeling locations of dust generation (Figure 8).

Wind speed. Wind is the primary factor for aeolian erosion[3]. Generally, winds at various altitudes can transport sands and dust and this is dictated by wind speed. In the study area, wind speed averages between 10 to 17 m/s at the surface (Figure 8a). Therefore, wind speed is an important factor for mapping dust-source potential because it increases the probability of dust production. The wind-speed map demonstrates that speeds are high in the eastern part of the region and are moderate in the western part. Winds tend to be lower in the northern portion of the study area.

Geology. The study area is geologically composed of alluvium, ophiolites, conglomerates, sandstones, acidic and basic igneous, and volcanic rocks (Figure 8b). In addition, dolomites, limestones, mud volcanics, recent volcanics and some coloured series are found in the study area. Some areas have not been geologically surveyed. Jaz Murian basin is the largest basin in the study area. However, rocks from the Cambrian to the Triassic period are found in this region. Pyroclasts, alluvium, limestone, sandstone, basic and ultra-basic stones, and ophiolites are easily eroded by wind and provide the for abundant sources of dust production.

Air temperature. Air temperature plays a key role for dust production. Higher air temperatures increase rock decomposition to rapidly generate significant quantities of dust particles[20]. The maximum air temperatures in the study region range from 49°C to 42.1°C (Figure 8c).

Land use. Land use is also an indicator used to map dust potential[20]. Land use reflects the intensity of human activities and the potential for environmental degradation and disturbance of the surface. This study used a land-use map derived from a Landsat OLI image (2016) employing an object-based image-classification technique (Figure 8d). The image was radiometrically corrected with a pre-processing technique by converting the detected radiometrics into reflectance values.

Slope. Slope is crucial to producing dust and it is incorporated into dust emission and transport models. The dust sources are widely distributed in areas of lower slopes and can be identified and assessed with remote sensed time-series data[12]. The slope value is represented as a percentage; the highest slope value was 185.3 (Figure 8e).

Soil. The characteristics of soils directly and indirectly affect the dust-storm initiation[12,20]. Eroded particles vary in size (i.e. from dust particle to boulder). Heavier materials cannot be moved very far by wind, but dust particles can be transported long distances and are deposited when they collide with obstacles in their paths or when wind speed diminishes and loses its capacity to move them. Soil type is also a primary influence on plant growth. Figure 8f shows the distribution of the dominant soil types in the region.

Rainfall. Rainfall influences soil moisture, significantly impacting the strength of some soils against erosion and consequently particulate production. If rainfall and or soil moisture decreases, dust increases. It therefore has a very important influence on the spatial distribution of dust potential. The study area is dominated by landscapes of sparse shrubs and annual plants that reflect the arid climate with low precipitation; the northern and southern parts receive more precipitation than the central region of the study area (Figure 8g).

Land cover. Land cover is relevant to discerning dust-source potential. Land cover influences the susceptibility of the soil to erosion. Compared to forests, land degradation is more severe on land with scant vegetation. A land cover map of the study area was obtained from the Forest, Range and Watershed Organization (FRWO) of Iran (Figure 8h).

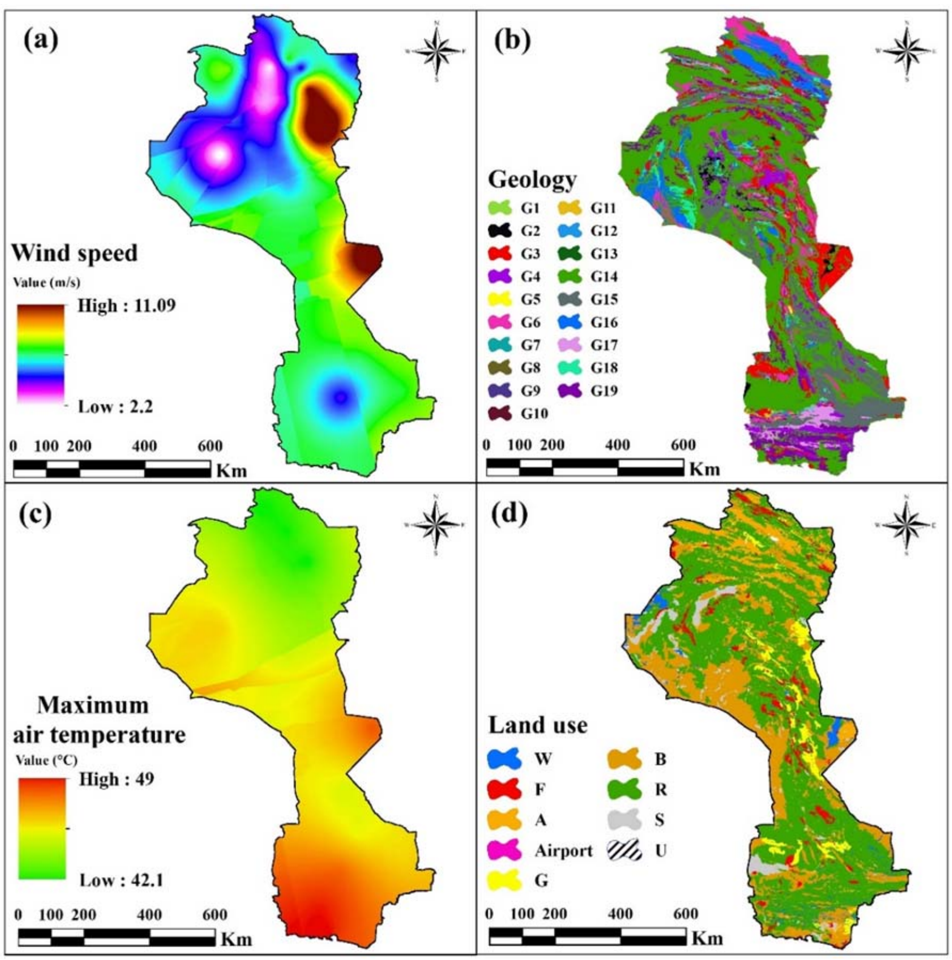


Figure 8. Dust influencing factors: a) wind speed, b) geology, c) maximum air temperature, d) land use, e) slope, f) soil, g) rainfall, h) land cover.

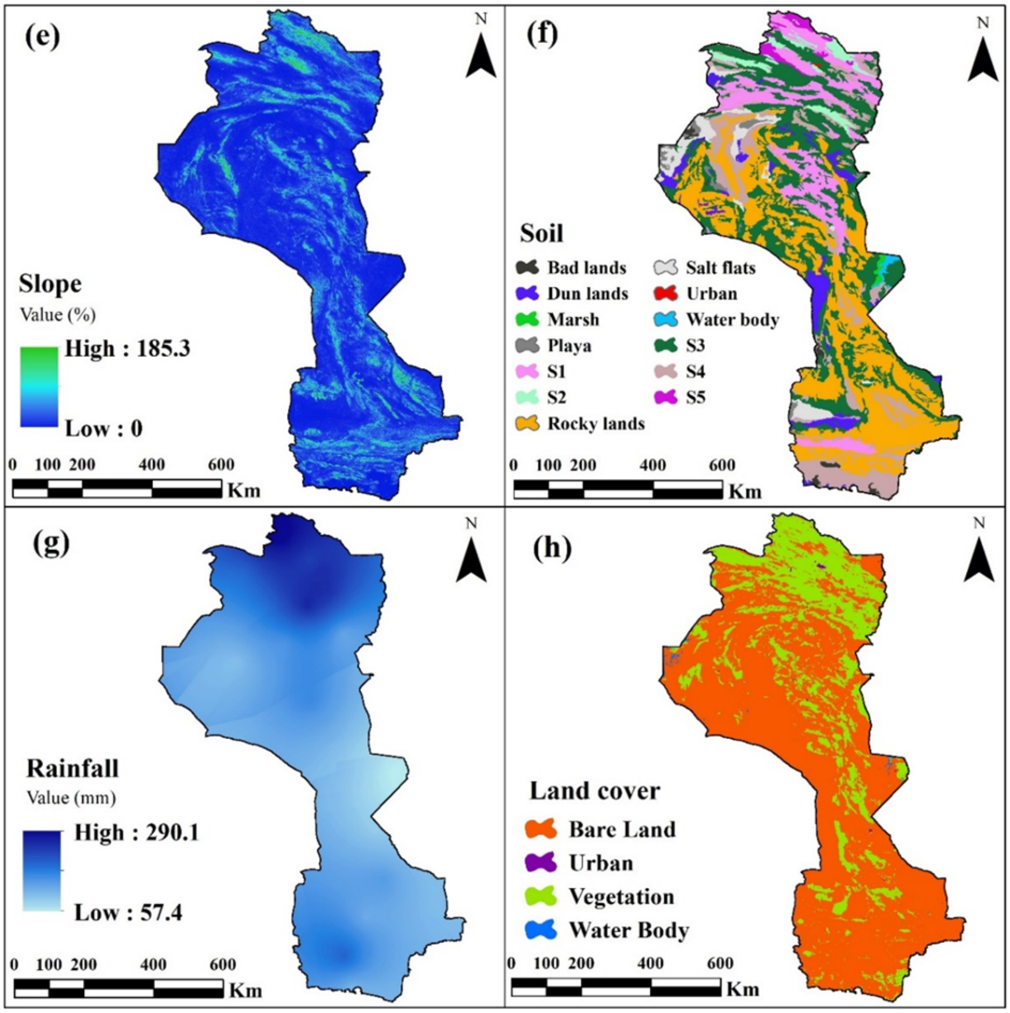


Figure 8 (continued)

Application of models

ANFIS, also known as the universal estimator, is the combination of artificial neural networks (ANNs) and the Takagi–Sugeno fuzzy-inference system which was first developed in the early 1990s[14,15,16]. ANNs are powerful learning-capable machines, yet they are unable to generalize or predict patterns. In particular, they cannot calculate non-linear functions of system components or phenomena of interest, mainly because ANNs do not learn in a compositional manner[22]. Hence, fuzzy if-then rules serve as an inference-engine that enables ANFIS to approximate non-linear patterns by: perpetually updating the knowledge of that system based on newly defined rules, and concurrently updating the linear and nonlinear parameters based on gradient descent and recursive least-square algorithms[10,27,28,39,42]. Tuning the learning parameters takes considerable time and requires a significant amount of input data. Therefore, many optimization algorithms are developed to automatically optimize these learning-parameters. Among these, three novel optimization methods – bat, cultural, and differential evolution algorithms – are adopted and fused into the ANFIS model as ANFIS-BA, ANFIS-CA, and ANFIS-DE. In summary, the bat algorithm, as the name implies, imitates the echolocation behavior of bats (i.e.

sound pulses) and was first developed by Yang[44]. It entails three main components: frequency, loudness, and pulse emission rate (See Yang[44] for details).

Flying with random velocity in a random space (i.e., randomly moving through the parameters' space) and analyzing the three aforementioned variables, bats distinguish an object from obstacles and obstacles from open space (i.e., the presence and absence of localities)[1,32]. With this information, the bat optimizer can tune the learning parameters of ANFIS.

The CE algorithm, on the other hand, develops with evolutionary computations. It is a mathematical representation of how societies evolve or adapt to their environments. First expounded by Reynolds[31], the algorithm is underpinned by a two-level computational process, termed a dual-inheritance[30,35]. The first level focuses on a population that shares a set of behavioral traits that is continuously handed down through the generations and is possibly spread to others in society by social motivators. The second level focuses self-experiences and self-forecasts that can be generalized and merged into a global belief. Thus, the circulation between the population, a belief and subcomponents therein provide an outline for a cultural-evolution framework that can be mathematically represented by various models, such as genetic algorithms[38].

DE, as a stochastic global-optimization method, was first introduced by Storn and Price (1997) to optimize the properties of a non-linear and non-differentiable problem in a continuous space. The DE targets an objective function (e.g., a cost function) and minimizes it under certain constraining functions with an easy-to-operate implementation process[39] (Liu and Lampinen 2002). Using a vector (or parameter) population and reliable handling of stochastic perturbations in the population enables DE to fairly quickly provide practical results. The DE has been used to contribute to evolutionary optimization and is one of the fastest and most practical optimization methods, particularly in comparison to other prominent minimization methods such as annealing and genetic algorithms (See Storn and Price 1997 for more details). In this research, all individual and hybrid models (i.e., ANFIS, ANFIS-BA, ANFIS-CA, and ANFIS-DE) were executed in MATLAB software.

Accuracy assessment

To suggest or reject a developed model for other susceptible areas, the reliability and performance of it should be evaluated using training and validation datasets[13]. In this study some common statistical metrics including mean squared error (MSE), root mean square error (RMSE), area under the receiver operating characteristic curve (AUC), and true skill statistic (TSS) were used. The MSE and RMSE are formulated as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(X_{est} - X_{obs})^2 \qquad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{est} - X_{obs})^2}{n}} \qquad (2)$$

where X_{est} and X_{obs} are defined as the dust estimated and observed (actual), respectively, and n is the number of dust observations. AUC is used to assess performance and measures how well a model generally performs[26]. The AUC is formulated as:

$$AUC = \frac{\sum TP + \sum TN}{P + N} \qquad (3)$$

where TP is true positive (dust correctly classified), TN is true negative (non-dust correctly classified), P and N are total number of dust and non-dust locations. TSS also is the other metrics to check the model performance based on the sensitivity and specificity statistical measures. It can be expressed as follows:

$$\mathrm{TSS} = \mathrm{Sensitivity} + \mathrm{Specificy} - 1 \quad (4)$$

Acknowledgements

The authors would like to thank the Iranian Department of Geological Surveys (IDGS) and the Forest, Range and Watershed Organization (FRWO) for supplying required data, reports, useful maps, and their nationwide geodatabase. This research was partially supported by the Geographic Information Science Research Group, Ton Duc Thang University, Ho Chi Minh City, Viet Nam.

Author Contributions

OR, SSG, MS, FM, OA, AZ, and HK conducted field investigations, collected field data and prepared maps. OR, BP, AK, HS, AS, HK, AZ, FM, and OA wrote the manuscript. MP, MS, and OR performed models and statistical analysis. DTB, RCD, AG, and JPT provided critical comments in planning this paper and edited the manuscript. All the authors discussed the results and revised the manuscript.

Competing Interests: authors declare that they have no competing interests.

References

1. Ali, E. Optimization of power system stabilizers using BAT search algorithm. *International Journal of Electrical Power & Energy Systems* **61**, 683-690 (2014).
2. Baddock, M. C., Bullard, J. E. & Bryant, R. G. Dust source identification using MODIS: a comparison of techniques applied to the Lake Eyre Basin, Australia. *Remote Sensing of Environment* **113**, 1511-1528 (2009).
3. Borrelli, P., Ballabio, C., Panagos, P. & Montanarella, L. Wind erosion susceptibility of European soils. *Geoderma* **232**, 471-478 (2014).
4. Bory, A. J. M., Biscaye, P. E. & Grousset, F. E. Two distinct seasonal Asian source regions for mineral dust deposited in Greenland (NorthGRIP). *Geophysical research letters* **30** (2003).
5. Crooks, J. L. *et al.* The association between dust storms and daily non-accidental mortality in the United States, 1993–2005. *Environmental health perspectives* **124**, 1735-1743 (2016).
6. Feuerstein, S. & Schepanski, K. Identification of Dust Sources in a Saharan Dust Hot-Spot and Their Implementation in a Dust-Emission Model. *Remote Sensing* **11**, 4 (2019).
7. Ginoux, P., Garbuzov, D. & Hsu, N. C. Identification of anthropogenic and natural dust sources using Moderate Resolution Imaging Spectroradiometer (MODIS) Deep Blue level 2 data. *Journal of Geophysical Research: Atmospheres* **115** (2010).
8. Goudie, A. S. Desert dust and human health disorders. *Environment international* **63**, 101-113 (2014).
9. Griffin, D. W. & Kellogg, C. A. Dust storms and their impact on ocean and human health: dust in Earth's atmosphere. *EcoHealth* **1**, 284-295 (2004).
10. Griffin, N. & Lewis, F. in *[Proceedings 1989] IEEE International Workshop on Tools for Artificial Intelligence.* 246-251 (IEEE).
11. Grousset, F. E., Ginoux, P., Bory, A. & Biscaye, P. E. Case study of a Chinese dust plume reaching the French Alps. *Geophysical Research Letters* **30** (2003).

12. Hahnenberger, M. & Nicoll, K. Meteorological characteristics of dust storm events in the eastern Great Basin of Utah, USA. *Atmospheric environment* **60**, 601-612 (2012).
13. He, Q. *et al.* Landslide spatial modelling using novel bivariate statistical based Naïve Bayes, RBF Classifier, and RBF Network machine learning algorithms. *Science of the Total Environment* **663**, 1-15 (2019).
14. Jang, J.-S. ANFIS: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics* **23**, 665-685 (1993).
15. Jang, J.-S. Fuzzy modeling using generalized neural networks and kalman filter algorithm. *In AAAI* **91**, 762-767 (1991).
16. Jang, J.-S. R., Sun, C.-T. & Mizutani, E. Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [Book Review]. *IEEE Transactions on automatic control* **42**, 1482-1484 (1997).
17. Kanatani, K. T. *et al.* Desert dust exposure is associated with increased risk of asthma hospitalization in children. *American journal of respiratory and critical care medicine* **182**, 1475-1481 (2010).
18. Khaniabadi, Y. O. *et al.* Impact of Middle Eastern Dust storms on human health. *Atmospheric pollution research* **8**, 606-613 (2017).
19. Khazraee, S., Jahanmiri, A. & Ghorayshi, S. Model reduction and optimization of reactive batch distillation based on the adaptive neuro-fuzzy inference system and differential evolution. *Neural Computing and Applications* **20**, 239-248 (2011).
20. Kimura, R. Factors contributing to dust storms in source regions producing the yellow-sand phenomena observed in Japan from 1993 to 2002. *Journal of Arid Environments* **80**, 40-44 (2012).
21. Leski, T. A., Malanoski, A. P., Gregory, M. J., Lin, B. & Stenger, D. A. Application of a broad-range resequencing array for detection of pathogens in desert dust samples from Kuwait and Iraq. *Appl. Environ. Microbiol.* **77**, 4285-4292 (2011).
22. Liška, A., Kruszewski, G. & Baroni, M. Memorize or generalize? searching for a compositional RNN in a haystack. *arXiv preprint arXiv:1802.06467* (2018).
23. Middleton, N. & Kang, U. Sand and dust storms: impact mitigation. *Sustainability* **9**, 1053 (2017).
24. Nazari, S., Kermani, M., Fazlzadeh, M., Matboo, S. A. & Yari, A. R. The origins and sources of dust particles, their effects on environment and health, and control strategies: a review. *Journal of Air pollution and Health* **1**, 137-152 (2016).
25. Neisi, A. *et al.* Study of heavy metal levels in indoor dust and their health risk assessment in children of Ahvaz city, Iran. *Toxin reviews* **35**, 16-23 (2016).
26. Pham, B. T. *et al.* Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *Catena* **175**, 203-218 (2019).
27. Polat, K. & Güneş, S. A hybrid medical decision making system based on principles component analysis, k-NN based weighted pre-processing and adaptive neuro-fuzzy inference system. *Digital Signal Processing* **16**, 913-921 (2006).
28. Premkumar, K. & Manikandan, B. Adaptive neuro-fuzzy inference system based speed controller for brushless DC motor. *Neurocomputing* **138**, 260-270 (2014).
29. Prospero, J. M., Ginoux, P., Torres, O., Nicholson, S. E. & Gill, T. E. 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**, 2-1-2-31 (2002).
30. Reynolds, R. G. in *New ideas in optimization.* 367-378 (McGraw-Hill Ltd., UK).
31. Reynolds, R. G., Ali, M. & Jayyousi, T. Mining the social fabric of archaic urban centers with cultural algorithms. *Computer* **41**, 64-72 (2008).
32. Sambariya, D. & Prasad, R. Robust tuning of power system stabilizer for small signal stability enhancement using metaheuristic bat algorithm. *International Journal of Electrical Power & Energy Systems* **61**, 229-238 (2014).
33. Schepanski, K., Tegen, I. & Macke, A. Comparison of satellite based observations of Saharan dust source areas. *Remote Sensing of Environment* **123**, 90-97 (2012).
34. Soleimani, Z., Goudarzi, G., Sorooshian, A., Marzouni, M. B. & Maleki, H. Impact of Middle Eastern dust storms on indoor and outdoor composition of bioaerosol. *Atmospheric environment* **138**, 135-143 (2016).

35. Soza, C., Becerra, R. L., Riff, M. C. & Coello, C. A. C. Solving timetabling problems using a cultural algorithm. *Applied Soft Computing* **11**, 337-344 (2011).
36. Stefanski, R. & Sivakumar, M. in *IOP Conference Series: Earth and Environmental Science.* 012016 (IOP Publishing).
37. Tien Bui, D. *et al.* New hybrids of anfis with several optimization algorithms for flood susceptibility modeling. *Water* **10**, 1210 (2018).
38. Tien Bui, D. *et al.* Novel hybrid evolutionary algorithms for spatial prediction of floods. *Scientific reports* **8**, 15364 (2018).
39. Bui, D. T., Pradhan, B., Lofman, O., Revhaug, I. & Dick, O. B. Landslide susceptibility mapping at Hoa Binh province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS. *Computers & Geosciences* **45**, 199-211 (2012).
40. Uno, I. *et al.* Asian dust transported one full circuit around the globe. *Nature Geoscience* **2**, 557 (2009).
41. Wang, X., Liu, J., Che, H., Ji, F. & Liu, J. Spatial and temporal evolution of natural and anthropogenic dust events over northern China. *Scientific reports* **8**, 2141 (2018).
42. Wang, Y.-M. & Elhag, T. M. An adaptive neuro-fuzzy inference system for bridge risk assessment. *Expert systems with applications* **34**, 3099-3106 (2008).
43. Yamaguchi, N., Ichijo, T., Sakotani, A., Baba, T. & Nasu, M. Global dispersion of bacterial cells on Asian dust. *Scientific Reports* **2**, 525 (2012).
44. Yang, X.-S. in *Nature inspired cooperative strategies for optimization (NICSO 2010)* 65-74 (Springer, 2010).
45. Yaseen, Z. M. *et al.* Novel approach for streamflow forecasting using a hybrid ANFIS-FFA model. *Journal of Hydrology* **554**, 263-276 (2017).
46. Yaseen, Z. M. *et al.* Rainfall pattern forecasting using novel hybrid intelligent model based ANFIS-FFA. *Water resources management* **32**, 105-122 (2018).
47. Yilbas, B. S. *et al.* Influence of dust and mud on the optical, chemical, and mechanical properties of a PV protective glass. *Scientific reports* **5**, 15833 (2015).