



University of Tehran Press

Environmental

Hazards

Management



Iranian Hazardology Association

Online ISSN: 2383-0530

Home Page: <https://jhsci.ut.ac.ir>

Hazardous Dust Source Susceptibility Mapping in Wet and Dry Periods of the Tigris-Euphrates Basin: A Meta-Heuristics and Machine Learning Algorithm

Azher Ibrahim Al-Taei¹ | Ali Asghar Alesheikh^{2*} | Ali Darvishi Bolorani³

1. PhD Candidate in GIS, Department of GIS, Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran. Email: azher.ataei@email.kntu.ac.ir
2. Corresponding Author, Professor in GIS, Center of Excellence for Geospatial Information Technology, Faculty of Geomatics Eng. , K. N. Toosi University of Technology, Tehran, Iran. Email: alesheikh@kntu.ac.ir
3. Associate Professor of Geoinformatics, Department of Remote Sensing and GIS, University of Tehran, Tehran, Iran. Email: ali.darvishi@ut.ac.ir

ARTICLE INFO

Article type:
Research Article

Article History:
Received 20 February 2024
Revised 15 March 2024
Accepted 16 March 2024
Published 18 March 2024

Keywords:
*Hotspot dust sources,
Random forest,
Meta-heuristics methods,
Tigris-Euphrates Basin,
Wet and Dry periods.*

ABSTRACT

Dust storms are a severe form of air pollution that poses significant threats to the environment and human health. To deal with this phenomenon, it is crucial to comprehend the mechanisms accountable for dust generation. This can be achieved by utilizing machine learning in dust source susceptibility mapping. Although dust activities vary spatiotemporally due to the constantly changing atmosphere, few papers have addressed dust source susceptibility mapping considering Earth system frameworks such as wet and dry periods. Also, while optimizing hyperparameters is crucial for improving machine learning performance, many studies have neglected this aspect in this particular application. To address this research gap, the objective of this study was to create a framework for mapping the susceptibility of hazardous hotspot dust sources (HDS) during wet and dry periods (based on the changes in water bodies) using a fine-tuned random forest (RF) model with teaching learning-based optimization (TLBO) and student psychology based optimization (SPBO) optimizers. To achieve this, the study analyzed 10,392 identified HDS, along with various environmental influential factors between 2000 and 2020 in the transnational shared Tigris-Euphrates Basin, which is a significant source of dust in the Middle East and globally. The results showed that RF-TLBO performed slightly better than RF-SPBO, with an average mean absolute error (MAE) of 0.146, average root mean squared error (RMSE) of 0.194, and average Willmott index (WI) of 0.761, compared to RF-SPBO's average MAE of 0.148, average RMSE of 0.195, and average WI of 0.757. The TLBO tuned RF with a lower number of trees and a lower maximum depth value, making it a simpler model. We utilized RF-TLBO and observed more areas that are more susceptible to hazardous dust sources during dry periods, confirming the meaningful relationship between wet and dry periods and hazardous dust susceptibility. Higher susceptibilities were found near water bodies and marshlands, indicating the significant impact of fluctuating water bodies on the generation of hazardous dust sources. The Gini index results also show that vegetation cover, elevation, wind speed, and soil texture have a high impact on land susceptibility to be a hazardous dust source.

Cite this article: Al-Taei, A. I.; Alesheikh, A. A. & Darvishi Bolorani, A. (2024). Hazardous Dust Source Susceptibility Mapping in Wet and Dry Periods of the Tigris-Euphrates Basin: A Meta-Heuristics and Machine Learning Algorithm. *Environmental Hazards Management*, 10 (4), 355-370. DOI: <http://doi.org/10.22059/jhsci.2024.373445.821>



© Azher Ibrahim Al-Taei, Ali Asghar Alesheikh, Ali Darvishi Bolorani.

Publisher: University of Tehran Press.

DOI: <http://doi.org/10.22059/jhsci.2024.373445.821>

1. Introduction

The world meteorological organization (WMO) defines dust storm among the most severe forms of air pollution, reducing horizontal visibility to less than one kilometer [60]. Dust storms have been identified as significant threats that cause considerable harm to climate [29], human health [31], quality of life [41], plant phenology [10], and infrastructure [40], ranging from local urban to continental, and from minutes to several years. Dust storms are prevalent in arid and semi-arid areas where precipitation is low and evaporation is high [23, 35]. The Middle East is a major source of dust storms globally, contributing over 20% of total worldwide dust emissions [8, 24]. Additionally, due to the Middle East's susceptibility to severe drought and climate change, occurrences of dust storms in the region have experienced higher frequencies and intensities in recent years [40]. Tigris-Euphrates Basin (TEB) has emerged as a source of dust in the Middle East because of recurring drought periods [40]. Additionally, a variety of natural and human factors, such as drought [5], heightened water consumption [26], deforestation [4], excessive farming [21], and water resources manipulations like dam construction [26] may amplified this issue in TEB. As a result, these new generated dust sources are expanding and impacting the neighboring nations. For example, Al-Taei et al. (2023) identified three primary regions in the eastern and southeastern parts of Iraq, northwestern Iraq, and eastern Syria where LULC change were severe and dust storm emissions were highly likely.

It is crucial to anticipate the susceptibility of the TEB to dust storm formation, taking into account spatiotemporal variability, and to gain understanding regarding the mechanisms responsible for dust generation [40]. Various factors including soil characteristics, meteorological conditions, wind regime, vegetation cover, hydrological condition, and morphological features influence dust source formation [40, 57]. Hence, dust source susceptibility mapping is a challenging procedure as it requires understanding the complex interrelationships among driving factors [6, 28].

Machine learning is a procedure for building phenome behavior models, since machine learns relationships between dependent and independent variables. Machine learning has proven effective for mapping dust storm sources' susceptibility through different conditioning factors. Various machine learning algorithms including support vector machine [11], multivariate adaptive regression spline [46], classification tree analysis [28], artificial neural network [13], random forest (RF) [11], Naïve Bayes [11], generalized linear model [28], flexible discriminant analysis [13], and logistic regression [46] have been applied for dust source susceptibility mapping, and the literature [28, 46, 58] acknowledges the RF as a more practical method in comparison to other algorithms. RF adopts ensemble learning techniques, improving interpretability and accuracy, making it an efficient data model for spatiotemporal studies.

However, one of the main limitations of RF is fine-tuning the hyperparameters [50]. To optimize machine learning algorithms, adopting meta-heuristic algorithms is an efficient approach [66]. Meta-heuristic algorithms are decentralized and self-organizing methods that utilize team intelligence to solve complex problems. They facilitate efficient exploration of the search space and provide near-optimal solutions. The integration of RF and meta-heuristic methods has been explored, and the results have shown that it can improve modeling outputs [34]. Compared to other meta-heuristic algorithms, literature [33, 64] has introduced two more practical methods for improving machine learning: student psychology-based optimization (SPBO) and teaching learning-based optimization (TLBO). SPBO does not require tuning parameters, simplifying computational experiments [45], while TLBO is simple, easy to describe and implement [37].

Few studies have focused on optimizing the hyperparameters of RF for mapping the susceptibility of hazardous dust sources. Meanwhile, though regional transitions to dry periods can significantly increase susceptibility to dust storms [40], there has been limited research on dust source susceptibility mapping considering wet and dry periods. To address this research gap, this study aimed to detect susceptible hazardous HDS in the TEB during wet and dry periods (based on the changes in water bodies) from 2000 to 2020. The study seeks to create susceptibility dust emission map utilizing the improved RF by meta-heuristic algorithms of TLBO and SPBO, and to link changes in susceptibility between the periods to various influential factors using the Gini index. These all can aid policymakers in pinpointing susceptible regions and enacting appropriate mitigation measures against dust formation in the TEB.

2. Material and Methods

2.1. Study area

TEB is a transboundary river Basin covering an area of 879,000 square kilometers in the Middle East (Figure 1). It is located between 36° 41' and 52° 8' east longitude and 27° 13' and 40° 23' north latitude, with most of it situated in Iraq and smaller portions in Iran, Turkey, and Syria, as well as northern Saudi Arabia and Kuwait. Since the 1980s, TEB has been a significant source of dust [16]. Most of the dust originates from Iraq, the lowlands of southwest Iran, and the countries surrounding the Persian Gulf [29, 52].

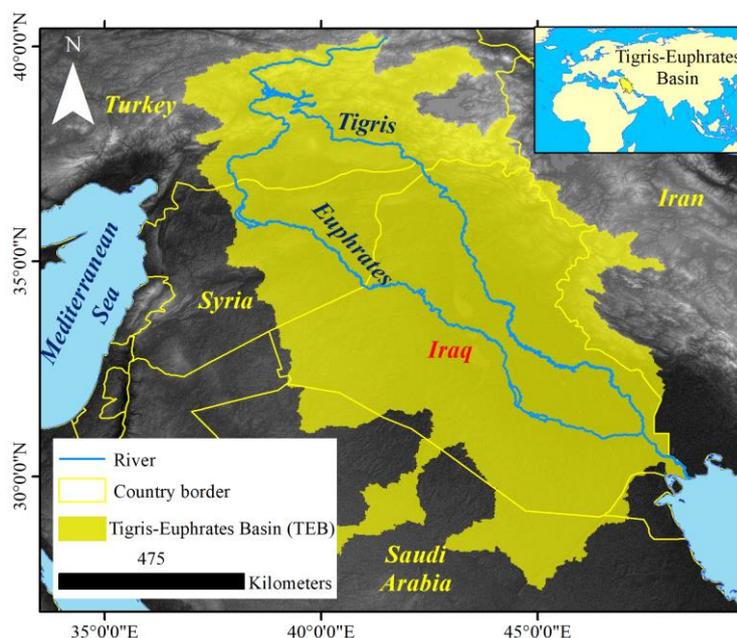


Fig. 1. The Tigris-Euphrates Basin study area

2.2. Hotspot dust sources

A total of 10,392 HDS were identified in the TEB region between 2000 and 2020 <https://data.mendeley.com/datasets/7937gn7g8c/1> [17]. The majority of these dust occurrences, approximately 70%, took place between March and August. The HDS were classified into very high (1), high (0.8), moderate (0.6), low (0.4), and very low (0.2) hazardousness levels by calculating the frequency, intensity, and continuity of dust events.

Changes in lakes and wetlands are influenced by both climatic conditions and human activities, such as irrigation, agriculture, and upstream water management strategies [40]. As water bodies decrease, previously moist or vegetated surfaces become exposed to wind erosion, which increases the likelihood of dust storm events [18]. Darvishi Bolorani et al. (2021) identified two wet periods (2005–2007 and 2013–2020) and two dry periods (2000–2004 and 2008–2012) in the TEB based on changes in the lakes and marshland areas. During the dry periods, the areas of water bodies decreased by 31% and 33%, respectively. These four periods were considered for modeling dust source susceptibility under different wet and dry conditions.

2.3. Conditioning factors for dust source formation

The initial step in dust source susceptibility mapping is to identify the conditioning factors that lead to increasing dust generation. The literature has indicated that the formation of dust sources is primarily influenced by a combination of meteorological factors, topographic features, and surface conditions [11, 15, 43, 54, 63, 67]. In this study, we selected several input factors that include hydro-climatological factors (wind speed and precipitation), morphological factors (elevation, slope angle, slope aspect, plan curvature, and profile curvature), soil characteristics (soil texture), and vegetation cover (vegetation index).

2.3.1. Hydro-climatological factors

Wind and precipitation are two primary factors in the formation of dust sources. Wind creates the necessary force to dislodge soil particles. Strong winds exert pressure on the ground, causing soil particles to be lifted into the air [39-40]. Additionally, the moisture in the soil, derived from precipitation in the active aeolian layer, impacts dust emission by increasing the cohesiveness of the soil, which can prevent dust emission [42]. To account for the influence of wind speed and precipitation, we utilized the National Centers for Environmental Prediction (NCEP) reanalysis data assimilation system (FLDAS) dataset [36] and calculated the median near surface wind speed and average total precipitation rate. FLDAS dataset has a spatial resolution of about 11 km.

2.3.2. Morphological factors

Morphological factors have a significant impact on climatic conditions, drainage networks, water distribution, vegetation cover, and soil characteristics [7, 44]. These factors, in turn, affect soil erodibility [25]. To analyze these factors, we used the NASA-DEM digital elevation model [2] with a spatial resolution of 30 m in the Google Earth Engine platform. Then, we processed the DEM using ArcGIS Pro to obtain morphological factors including slope angle, slope aspect, plan curvature, and profile curvature.

The slope angle is the main factor that affects groundwater feeding, infiltration, runoff, and water movement speed [49]. Additionally, the slope aspect significantly impacts soil water content as it controls various hydrological processes, including precipitation direction, physiographical trends, and snow melting [44]. Plan curvature is associated with flow convergence and divergence, while profile curvature represents the maximum slope aspect direction and primarily impacts surface flow speed [3]. These components collectively influence the susceptibility of soil to become a dust source by regulating the stability of soil particles and their capacity to be transported by wind. Figure 2 illustrates the location maps of morphological factors applied in modeling.

2.3.3. Soil characteristics

Soil characteristics significantly impact soil erodibility to wind and determine the extent of susceptibility to dust source generation. Soil texture is closely linked to the source and spread of dust storms [40]. Soil texture containing fine particles is a prerequisite for an area to become a dust source. Additionally, disturbed soil structure can lead to increased dust emissions due to the increased susceptibility of topsoil particles to wind erosion [56]. The soil texture map of TEB was obtained using the Open Land Map Soil Texture Class (USDA System) [27] and includes twelve classes with a spatial resolution of 250 m (Figure 2).

2.3.4. Vegetation index

Vegetation cover influences the generation of dust sources by directly reducing the exposed bare soil, thereby suppressing dust emissions, and by acting as a major roughness element, consuming some of the wind momentum and increasing surface drag, further reducing dust emissions [61]. To apply the influence of vegetation cover to dust source susceptibility, we used the normalized difference vegetation index (NDVI). NDVI is a widely used vegetation index ranging between -1 and +1. Negative NDVI refers to water bodies, values close to +1 represent dense vegetation, and values close to 0 refer to bare land or built areas [40].

We applied the median NDVI values from March to August obtained from the MOD13Q1.061 Terra Vegetation Indices for the studied periods (Figure 2). This dataset contains vegetation indices at 16-day intervals with a spatial resolution of 250 m.

2.4. Methodology

This study comprises five main steps (Figure 3): (1) identifying the location of HDS and non-HDS points (based on the previous work) and dividing them into wet and dry periods, (2) determining influential factors on dust source formation according to the literature, (3) identifying possible collinearity relationships between independent variables by variance inflation factor (VIF) and tolerance coefficient, (4) modeling dust source susceptibility by integrated meta-heuristics-RF model and determining the primary conditioning factors by Gini index, and (5) validating the results by root mean squared error (RMSE), mean absolute error (MAE), and Willmott index (WI) metrics.

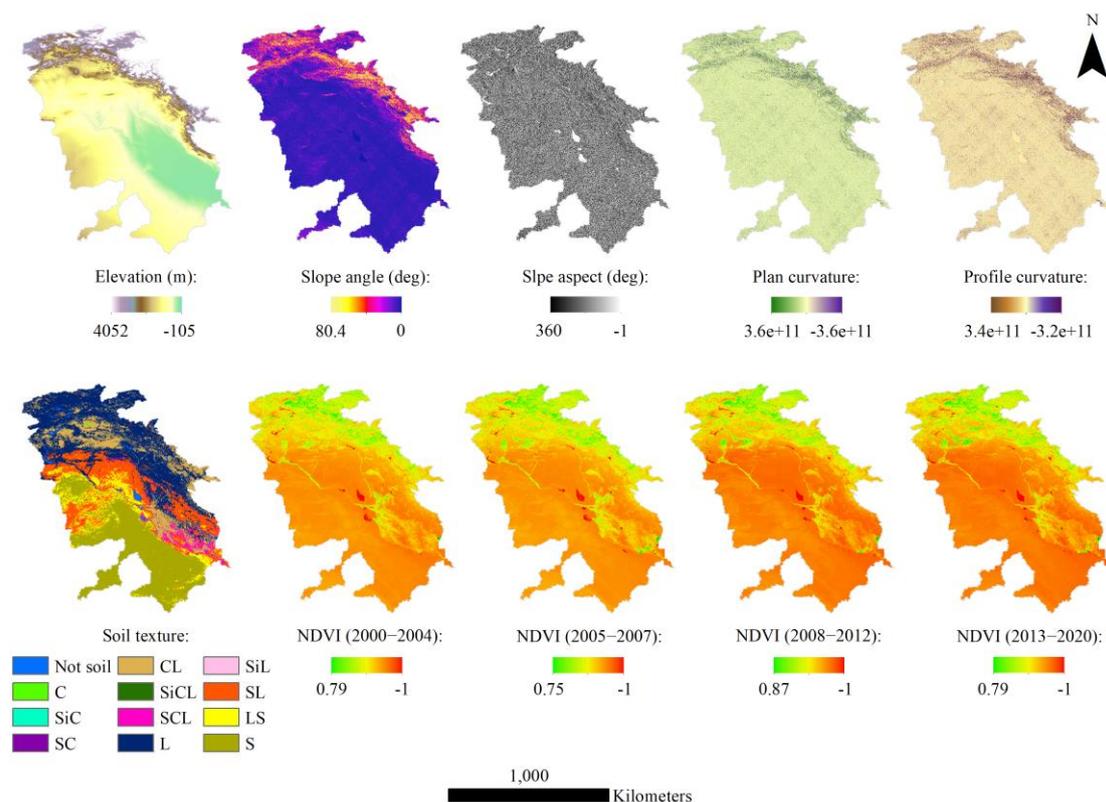


Fig. 2. Location maps of applied conditioning factors for dust source formation

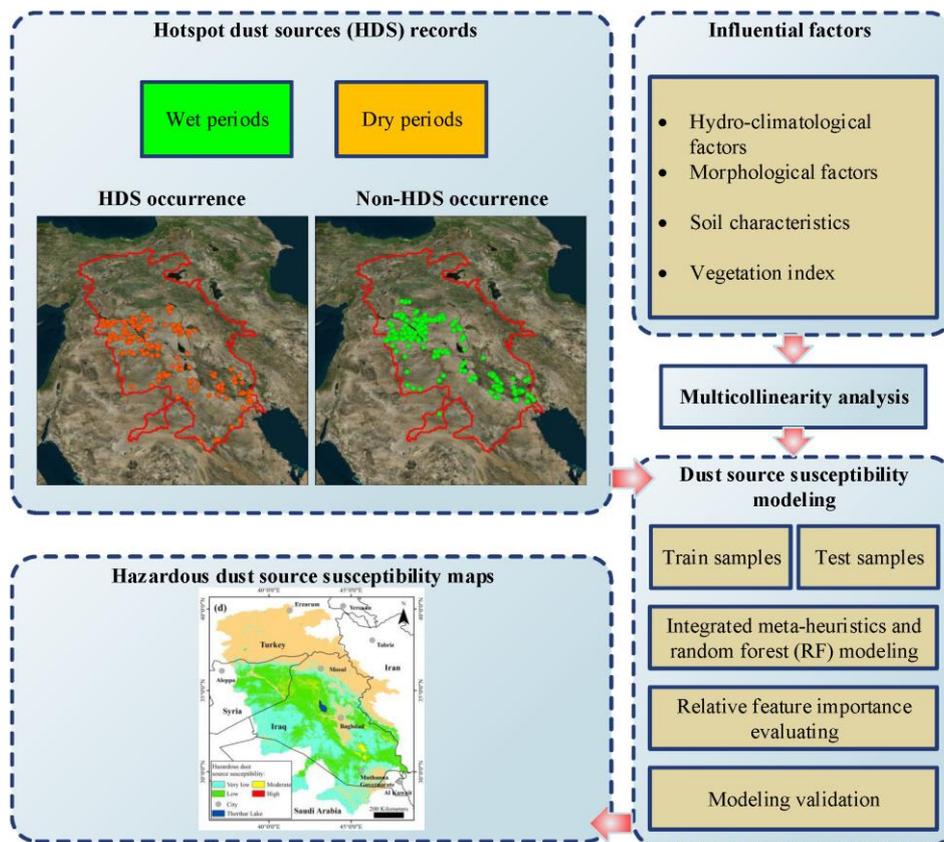


Fig. 3. Methodology of the study

2.5. Methods

2.5.1. Random forest modeling

RF is an ensemble machine learning algorithm that combines a set of decision trees for modeling. The decision tree, which serves as the base estimator for RF, is simple, practical, and fast, but it is prone to overfitting. A small change in the architecture of a decision tree can improve its outcomes greatly. Accordingly, RF is composed of a series of randomly generated decision trees. RF modeling starts with dividing samples into a random training group (the in-bag) and a random validation group (the out-of-bag). Next, base decision trees are prepared using 'bootstrap samples' taken randomly from the dataset. Finally, the final prediction is generated by averaging (in regression problems) or voting (in classification problems) between all trees. To improve RF performance, it is important to reduce the correlation between base decision trees and enhance their performance [22].

2.5.2. Teaching learning-based optimization

TLBO is a meta-heuristic algorithm introduced by Rao et al. (2011). It emulates the effect of a teacher on the students utilizing a population-based technique. The task consists the 'Teacher phase' and the 'Learner phase'. The first phase simulates a teaching process, while the second phase enhances learners' knowledge through their interaction. TLBO has gained popularity due to its fast convergence, lack of specific algorithm parameters, and simple implementation [48, 62]. The phases for the teacher and learner are defined as follows:

Teacher phase: The teacher improves the class level by transferring knowledge to learners. Assuming there are n students in the classroom, with an average grade of M_i for exam i , and the highest grade ($X_{T,i}$) belonging to the best learner who is acting as the teacher. The difference between the classroom's average grade and the highest grade is calculated as follows [51]:

$$Diff_i = r_i (X_{T,i} - T_F M_i) \quad (1)$$

The variable r_i represents a random number between 0 and 1. The teacher factor T_F , is associated with teaching quality and can be either 1 or 2. T_F is determined as follows [51]:

$$T_F = \text{round}[1 + \text{rand}(0,1)\{2-1\}] \quad (2)$$

Using the $Diff_i$, the new grade of student j in iteration i is expressed as follows [51]:

$$X'_{j,i} = X_{j,i} + Diff_i \quad (3)$$

The new and old grades of student j in iteration i are represented by $X'_{j,i}$ and $X_{j,i}$, respectively. The higher grade between the two will be used in the learner phase.

Learner phase: Successful students help others to improve. This obtains through collaborative efforts such as group assignments, which facilitate a better understanding of course materials beyond what is taught by the instructor. Assuming two randomly chosen students, A and B, the manner in which they assist each other can be described as follows [47, 51]:

$$X''_{A,i} = \begin{cases} X'_{A,i} + r_i (X'_{A,i} - X'_{B,i}) & \text{if } X'_{A,i} > X'_{B,i} \\ X'_{A,i} + r_i (X'_{B,i} - X'_{A,i}) & \text{if } X'_{B,i} > X'_{A,i} \end{cases} \quad (4)$$

If $X''_{A,i}$ is superior to $X'_{A,i}$, it will proceed to the next iteration. Otherwise, $X'_{A,i}$ will proceed to the next iteration.

2.5.3. Student psychology based optimization

SPBO is a meta-heuristic optimization algorithm developed by Das et al. (2020). This method is inspired by human psychology and helps students achieve the highest scores to excel in their class. This requires obtaining higher grades than others, which can be accomplished by putting in more effort across all subjects. However, students' abilities, efficiency, and interests vary, resulting in differing levels of performance. Typically, students can be classified into four groups based on their

subject-specific performance: Best students, Good students, Average students, and those who randomly attempt to improve [19].

Best student: The student with the highest overall scores is considered the best. To maintain this rank, best student strives to achieve the best scores in all subjects. This requires more effort than others [19]. The progress of the best student can be measured as follows:

$$X_{bestnew} = X_{best} + (-1)^k \text{rand} \times (X_{best} - X_j) \quad (5)$$

X_{best} represents the highest grade, X_j is the grade of the j^{th} student in a random subject, k is a random parameter (either 1 or 2), and rand equals to a random number between 0 and 1.

Good students: Good students strive to excel in their preferred subjects, thereby enhancing their overall performance. Due to the diversity of psychological traits, good students are chosen at random. To become the best student, some students try to exceed or meet the efforts of the best student [19]. This category of students is described as follows:

$$X_{newi} = X_{best} + \text{rand} \times (X_{best} - X_i) \quad (6)$$

Simultaneously, some students aim to exert more effort in their studies than the average students and emulate the efforts of the best student. This can be explained as follows:

$$X_{newi} = X + \text{rand} \times (X_{best} - X_i) + \text{rand} \quad (7)$$

In this equation, X represents the average class grade for a specific subject, while i represents the score obtained by the i^{th} student in that same subject.

Average students: If students are not particularly enthusiastic about a subject, they may only put in average effort. This average effort in a subject may prompt them to put more effort into other subjects. This group of students the average students in each subject [19]. The selection of these students is random and depends on their psychology as follows:

$$X_{newi} = X_i + \text{rand} \times (X_{mean} - X_i) \quad (8)$$

where X_{mean} represents the class average grade for a subject, and i represents the grade of the i^{th} student.

Students who try to improve randomly: Some students improve their grades by putting in varying levels of effort across different subjects. This approach involves putting a random amount of effort into each subject to increase their overall scores, as described as follows [19]:

$$X_{newi} = X_{min} + \text{rand} \times (X_{max} - X_{min}) \quad (9)$$

where X_{max} and X_{min} represent the highest and lowest possible grades for the subject, respectively.

2.5.4. Multicollinearity analysis

High correlation between two or more modeling features can invalidate a regression model. Multicollinearity is a criterion that helps detect collinearity relationships in a dataset. The VIF and tolerance coefficient measure multicollinearity for regression models. The tolerance coefficient ranges from 0 to 1; the higher the value, the lower the collinearity. If a dataset has a VIF value greater than 5 and a tolerance value less than 0.1, it may indicate the presence of multicollinearity [53].

2.5.6. Validation

The study employed RMSE, MAE, and WI as validation methods. RMSE and MAE are commonly used metrics for prediction error, with RMSE being sensitive to large errors and outliers, and MAE indicating the average prediction error. Both measures range from 0 to 1, with higher values indicating higher prediction error [30]. The calculation of RMSE and MAE is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2} \quad (10)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i| \quad (11)$$

where n is the number of samples, x_i represents the actual value, \hat{x}_i is the predicted value, and \bar{x} represents the mean of all actual values.

WI is a standardized measure that indicates the degree of model prediction error, ranging between 0 and 1. It shows the ratio of the mean square error to the potential error, with higher values representing better agreement between model predictions and observations [59] as follows:

$$\text{WI} = 1 - \frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{\sum_{i=1}^n (|\hat{x}_i - \bar{x}| + |x_i - \bar{x}|)^2} \quad (12)$$

3. Results and Discussion

Table 1 summarizes the results of the multicollinearity analysis. The VIF values for all independent factors, except precipitation and wind speed (2000-2004), were less than 5. This indicates that precipitation could contribute to multicollinearity and removing it could help improve the accuracy and stability of the regression model. Therefore, we removed the precipitation variable, and all other independent factors were used in the modeling process.

Table 1. Multicollinearity analysis

Independent variables	2000–2004	2005–2007	2008–2012	2013–2020
Wind speed	6.3	2.8	2.5	2.7
Precipitation	7.8	8.0	7.7	7.2
Elevation	3.0	2.8	3.0	2.9
Slope angle	2.1	1.9	1.9	2.0
Slope aspect	1.0	1.0	1.0	1.0
Profile curvature	1.3	1.3	1.3	1.6
Plan curvature	1.3	1.3	1.3	1.5
Soil texture	1.5	1.6	1.4	1.5
NDVI	2.4	2.3	2.4	2.7

The database was constructed with 10,392 HDS points and an additional 10,392 non-HDS points (absence locations) were selected and added at the regional level with a hazardousness level of 0. The database was then divided into 80% training samples and 20% validation samples. Each sample consisted of nine independent variables, which were conditioning factors for dust source formation, and one dependent variable, which was the hazardousness level.

Python's scikit-learn library was used to model dust source susceptibility. The hyperparameters that have the most influence on the performance of RF were selected based on expert. The RF model optimization parameters included the number of trees in the forest, the number of features considered for the best split, the maximum tree depth, the minimum number of samples required to split an internal node, the minimum number of samples required at a leaf node, and the number of samples to draw from X to train each base estimator. The Python pyMetaheuristic library was used to implement the TLBO and SPBO optimizers. The optimizers were utilized to minimize the objective function of RMSE value in 1000 iterations and 100 population size. The results indicate that both algorithms performed similarly in minimizing the objective function. However, TLBO tuned RF with a lower number of base estimators and lower maximum depth values, making RF a simpler model that can be trained and applied more easily and requires less disk space.

It is important to note that metaheuristic methods have been found to perform differently across various applications. While one method may excel in a particular issue, it may underperform in others. Despite TLBO's similarity to other metaheuristic techniques in terms of utilizing a set of solutions and its stochastic nature, its algorithmic inspiration is unique due to its less complex approach in tuning with the algorithm and other common parameters such as population size, number of iterations, and stopping criteria [65].

Table 2 summarizes the findings of the RMSE, MAE and WI values for the RF-SPBO and RF-TLBO models. There was no significant difference between the performance of RF-SPBO and RF-TLBO. However, the weakest performance with RMSE of 0.208, MAE of 0.156, and WI of 0.749 are related to RF-SPBO for modeling susceptibility of 2008 – 2012 period. In contrast, the most precise performance belongs to RF-TLBO (2000–2004) where RMSE is 0.193, MAE is 0.145, and WI is 0.767. Hence, we utilized the RF-TLBO model for further analysis and creating susceptible maps.

Table 2. Validation of RF model with RMSE and MAE

	RF-SPBO			RF-TLBO		
	RMSE	MAE	WI	RMSE	MAE	WI
2000–2004	0.195	0.147	0.772	0.193	0.145	0.767
2005–2007	0.196	0.150	0.739	0.194	0.145	0.753
2008–2012	0.208	0.156	0.749	0.208	0.153	0.756
2013–2020	0.182	0.140	0.769	0.181	0.140	0.768

The importance of modeling features was evaluated using the Gini index (Table 3). The calculated importance weights varied across the four study periods. However, in both wet and dry periods, NDVI was the most significant influential factor, followed by elevation, wind speed, and soil texture. On average, approximately 80% of the total importance weights were attributed to NDVI (0.371), elevation (0.203), wind speed (0.144), and soil texture (0.113), respectively.

Table 3. Evaluating the importance of modeling features by Gini index

	Wind speed	Elevation	Slope angle	Slope aspect	Plan curvature	Profile curvature	Soil texture	NDVI
2000–2004	0.118	0.238	0.064	0.063	0.045	0.044	0.094	0.334
2005–2007	0.147	0.208	0.049	0.043	0.031	0.033	0.118	0.371
2008–2012	0.128	0.175	0.050	0.044	0.031	0.033	0.119	0.420
2013–2020	0.183	0.190	0.051	0.040	0.029	0.027	0.121	0.358

The RF-TLBO model was used to map dust source susceptibility in four wet and dry periods across the entire study area (Figure 4). Although we used HDS records with hazardous levels ranging from 0.2 to 1 (very low to very high), the results show that no location was identified as susceptible to the formation of very high hazardous HDS. This observation may be due to the limited number of samples with a very high hazardous level, resulting in imbalanced data. Imbalanced datasets pose a primary concern as predictive models may become biased towards majority target values, leading to poor performance on less frequent samples, which are often of greater interest [14]. However, ensemble methods such as RF can improve predictive ability towards under-represented values, which in turn affects the predictive ability of models concerning the average behavior of the data [38]. Therefore, it is also possible that the selected influential factors on dust source generation did not fully explain the behavior of HDSs with a very high hazardous level.

Although dust susceptibility in TEB is not limited to considered periods and also depends on other Earth system frameworks, including climate oscillations [1, 32], landscape dynamics [28], atmosphere-ocean interaction [55], and droughts [12], this study emphasized the high importance of considering wet and dry periods in the formation of hazardous dust sources. According to Figure 4, the modeling results for wet and dry periods in the TEB differed and there were more areas susceptible to the formation of moderate and high hazardous dust sources during dry periods. From 2000 to 2004, 21.52% of the TEB was susceptible to very low hazardous HDS, 24.33% was susceptible to low hazardous HDS, 5.66% was susceptible to moderate hazardous HDS, and 0.11% was susceptible to high hazardous HDS. The similar proportions in 2005–2007 were 27.37%, 28.65%, 3.29%, and 0.01%, in 2008–2012 were 30.96%, 21.67%, 7.09%, and 0.03%, and in 2013–2020 were 32.80%, 25.83%, 1.83%, and 0.00%. Additionally, the northern areas of Lake Tharthar were susceptible to the formation of moderate to high hazardous dust sources during dry periods, while the same areas were susceptible to the formation of low hazardous dust sources during wet periods.

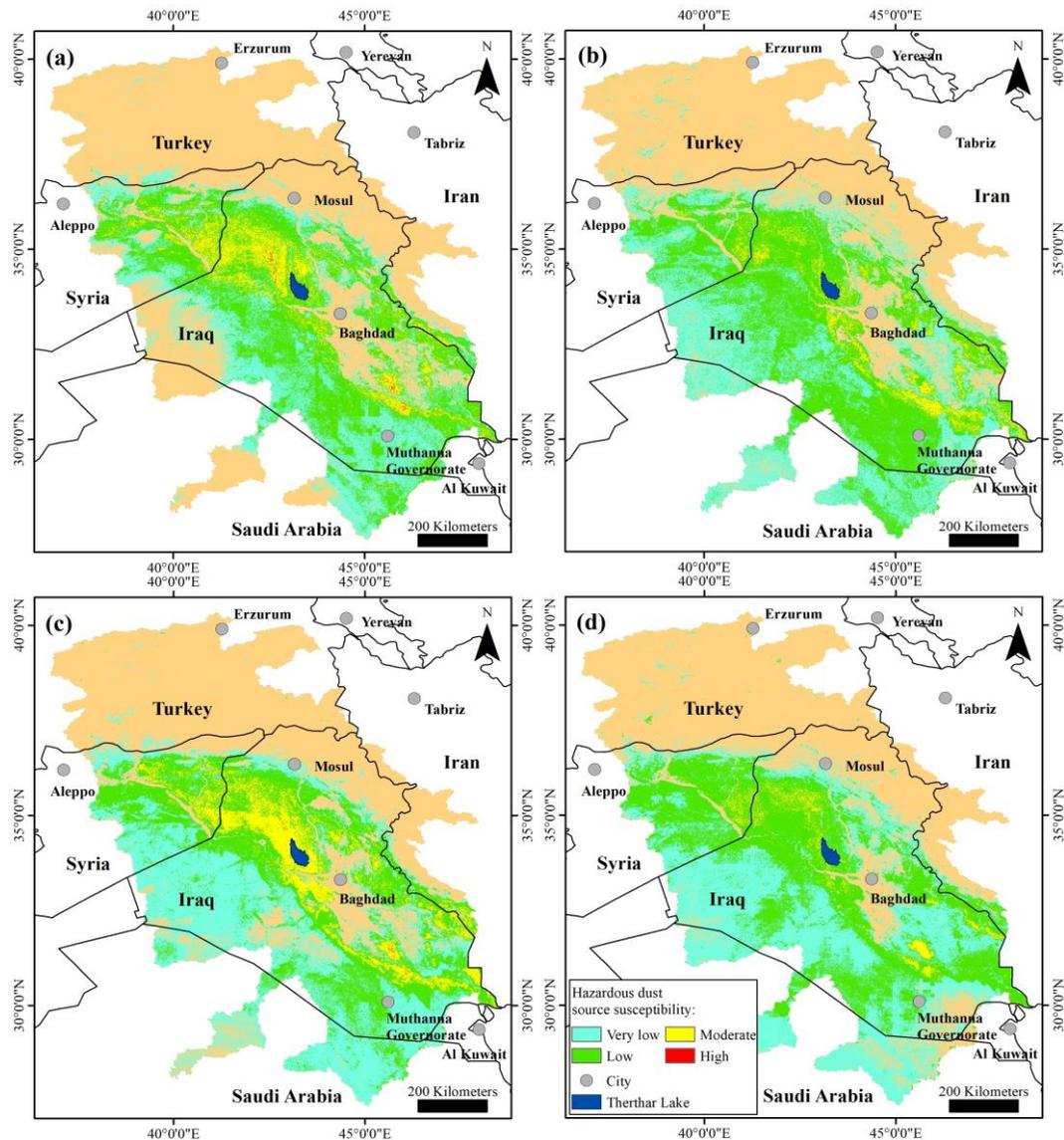


Fig. 4. Hazardous dust source susceptibility maps by RF-TLBO. a) 2000 – 2004, b) 2005–2007, c) 2008 – 2012, and d) 2013 – 2021.

To identify the main areas that were susceptible to hazardous HDS from 2000 to 2020, we created an average susceptibility map of hazardous HDS in the TEB by combining the four maps. We then aggregated this map with a cell factor of 10 and identified areas prone to the formation of moderate to high hazardous HDS (Figure 5). As shown in Figure 5, these areas were primarily situated near water bodies or marshes water bodies and seasonal marshes including Haditha Dam Lake, Tharthar Lake, Habbaniyah Lake, Hammar Lake, and Hammar Marsh. It indicates that the proximity of water bodies and marshes can create potential sources of severe dusts due to fluctuating water regimes.

Fluctuating water regimes, which can result from natural climatic variations or direct human intervention, significantly impact the exposure of fragile substrate materials. These materials are often found in areas with variable water levels and can become increasingly vulnerable when subjected to changes in their environment [20]. Climatic variations, such as droughts or heavy rainfall, can significantly affect water levels in these areas, which can lead to the exposure of substrate materials. Additionally, human activities, such as dam construction, irrigation, and other forms of water management, can also contribute to these fluctuations in water regimes [18].

When exposed, substrate materials are vulnerable to erosion by strong winds. These winds, capable of reaching high speeds, can lift large amounts of particulate matter from the substrate materials into the air. This process is a significant contributor to the formation of dust storms and is particularly

relevant in the TEB [18], the Middle East where water availability is shifting [9], and globally distributed 'playas' that contribute to approximately 34% of global dust emissions [68].

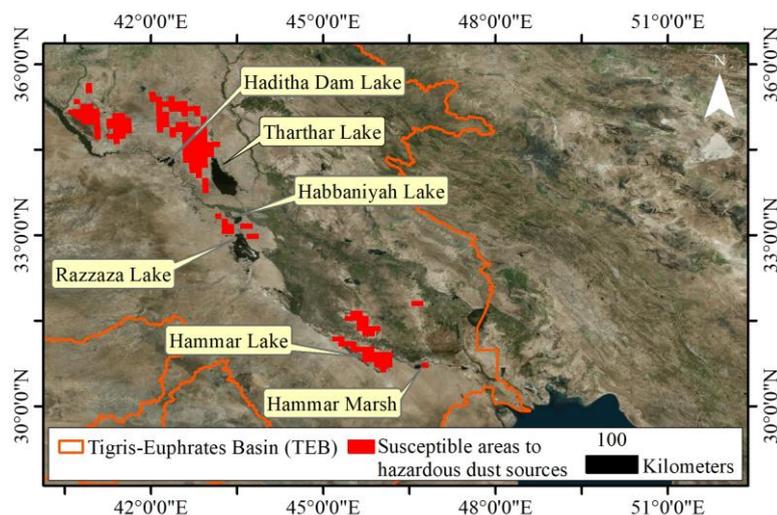


Fig. 5. Identified susceptible areas to hazardous dust sources from 2000 to 2020

The susceptibility maps created in TEB were compared to those in previous studies [11, 40]. The primary distinction between this study and prior research is the utilization of the hazardous level of dust sources in modeling to pinpoint regions that produce severe dust storms. Previous studies have often modeled susceptibility based solely on the frequency of dust source, without taking into account the severity of the resulting dust storm. It is important to consider both factors when assessing susceptibility to dust storms. However, following Naghibi et al. (2024) consideration of wet and dry periods for dust source susceptibility mapping in TEB, we identified higher susceptibilities near water bodies and marshlands. Similarly to Darvishi Bolorani et al. (2022), we found no significant susceptibility to dust sources in the northern and northeastern parts of Iraq and Turkey. However, there were differences in predicted susceptibility in southern and western parts of Iraq, where we observed susceptible areas to the formation of very low hazardous HDS. These inconsistencies in susceptibility maps may be due to the different predictive models applied and their configurations.

Acknowledgements

We thank the editor and the anonymous reviewers for their excellent suggestions to improve the original draft of the manuscript.

Conclusion

Dust storms are a significant form of air pollution that pose serious threats to both the environment and human health. They can lead to reduced air quality, visibility impairment, and adverse respiratory health outcomes. Therefore, it is of paramount importance to understand the mechanisms underlying dust generation. Modeling the susceptibility of hazardous dust sources is a critical step towards achieving this goal. This task involves identifying and characterizing areas with a high propensity for dust emission. However, the dynamic nature of dust activities presents a challenge, as they can vary significantly between wet and dry periods. Despite this challenge, few studies have considered the temporal variability of dust source susceptibility, leaving a gap in our understanding of dust storm generation and propagation. Our study aimed to map the susceptibility of hazardous dust sources in the TEB during wet and dry periods using the RF model. Although the performance of the RF model can be significantly influenced by its hyperparameters, there is limited literature on the optimization of these models for dust source susceptibility mapping. Our study thus contributed to this emerging field through fine-tuning the RF using TLBO and SPBO optimizers.

The study found that the RF-SPBO and RF-TLBO had statistically indistinguishable performance, indicating that both optimization algorithms are equally effective in tuning the RF model for dust

source susceptibility mapping. Notably, the TLBO algorithm produced a simpler RF model by tuning it with fewer base estimators and depths. Simpler models are often more interpretable and computationally efficient, without compromising predictive accuracy. The susceptibility maps generated by the RF-TLBO model revealed areas prone to hazardous dust sources during dry periods. Higher susceptibilities of dust formation were observed near water bodies and marshlands, emphasizing the crucial role of fluctuating water regimes in the generation of hazardous dust sources. The drying and wetting cycles of these areas can release significant amounts of dust into the atmosphere. Additionally, the Gini index analysis revealed that NDVI, wind speed, elevation, and soil texture are the most influential factors in dust generation. NDVI can indicate land cover conditions that are inversely related to dust emission potential. Dust emission is directly driven by wind speed, while elevation and soil texture can affect the spatial distribution of dust sources. These findings offer valuable insights into the environmental factors that control dust emission. This study can aid in the development of more effective strategies for predicting and mitigating dust storms.

References

- [1] Achakulwisut, P., Anenberg, S. C., Neumann, J. E., et al. Effects of increasing aridity on ambient dust and public health in the US Southwest under climate change. *GeoHealth*, 3(5), 127-144. (2019)
- [2] Aeronautics, N., & Laboratory, S. A. J. P. (2020). Nasadem merged dem global 1 arc second v001 [data set]. *NASA EOSDIS Land Processes DAAC*.
- [3] Al-Abadi, A. M., Al-Temmeme, A. A., & Al-Ghanimy, M. A. (2016). A GIS-based combining of frequency ratio and index of entropy approaches for mapping groundwater availability zones at Badra–Al Al-Gharbi–Teeb areas, Iraq. *Sustainable Water Resources Management*, 2, 265-283.
- [4] Al-Taei, A. I., Alesheikh, A. A., & Darvishi Bolorani, A. (2023). Land Use/Land Cover Change Analysis Using Multi-Temporal Remote Sensing Data: A Case Study of Tigris and Euphrates Rivers Basin. *Land*, 12(5), 1101.
- [5] Al Ameri, I. D., Briant, R. M., & Engels, S. (2019). Drought severity and increased dust storm frequency in the Middle East: a case study from the Tigris–Euphrates alluvial plain, central Iraq. *Weather*, 74(12), 416-426.
- [6] Alsubhi, Y., Qureshi, S., & Siddiqui, M. H. (2023). A New Risk-Based Method in Decision Making to Create Dust Sources Maps: A Case Study of Saudi Arabia. *Remote Sensing*, 15(21), 5193.
- [7] Aniya, M. (1985). Landslide-susceptibility mapping in the Amahata river basin, Japan. *Annals of the Association of American Geographers*, 75(1), 102-114.
- [8] Bank, W. (2019). Sand and Dust Storms in the Middle East and North Africa Region: Sources, Costs, and Solutions: World Bank.
- [9] Beyranvand, A., Azizi, G., Alizadeh, O., & Darvishi Bolorani, A. (2023). Dust in Western Iran: the emergence of new sources in response to shrinking water bodies. *Scientific reports*, 13(1), 16158.
- [10] Bolorani, A. D., Ranjbareslamloo, S., Mirzaie, S., Bahrami, H. A., Mirzapour, F., & Tehrani, N. A. (2020). Spectral behavior of Persian oak under compound stress of water deficit and dust storm. *International journal of Applied earth Observation and Geoinformation*, 88, 102082.
- [11] Bolorani, A. D., Samany, N. N., Papi, R., & Soleimani, M. (2022). Dust source susceptibility mapping in Tigris and Euphrates basin using remotely sensed imagery. *Catena*, 209, 105795.
- [12] Bolorani, A. D., Soleimani, M., Papi, R., et al. (2024). Assessing the role of drought in dust storm formation in the Tigris and Euphrates basin. *Science of The Total Environment*, 171193.
- [13] Boroughani, M., Mirchooli, F., Hadavifar, M., & Fiedler, S. (2023). Mapping land degradation risk due to land susceptibility to dust emission and water erosion. *Soil*, 9(2), 411-423.
- [14] Branco, P., Torgo, L., & Ribeiro, R. P. (2017). *SMOBN: a pre-processing approach for imbalanced regression*. Paper presented at the First international workshop on learning with imbalanced domains: Theory and applications.
- [15] Briceño-Zuluaga, F., Castagna, A., Rutllant, J. A., et al. (2017). Paracas dust storms: Sources, trajectories and associated meteorological conditions. *Atmospheric Environment*, 165, 99-110.
- [16] Darvishi Bolorani, A., Kazemi, Y., Sadeghi, A., Shorabeh, S. N., & Argany, M. (2020). Identification of dust sources using long term satellite and climatic data: A case study of Tigris and Euphrates basin. *Atmospheric Environment*, 224, 117299.
- [17] Darvishi Bolorani, A., Papi, R., Soleimani, M., et al. (2023). Visual interpretation of satellite imagery for hotspot dust sources identification. *Remote Sensing Applications: Society and Environment*, 29, 100888.
- [18] Darvishi Bolorani, A., Papi, R., Soleimani, M., Karami, L., Amiri, F., & Samany, N. N. (2021). Water bodies changes in Tigris and Euphrates basin has impacted dust storms phenomena. *Aeolian Research*, 50, 100698.
- [19] Das, B., Mukherjee, V., & Das, D. (2020). Student psychology based optimization algorithm: A new population based optimization algorithm for solving optimization problems. *Advances in Engineering software*, 146, 102804.
- [20] Ding, Q., Liu, X., & Zhang, X. (2014). Impacts of water level fluctuations on substrate environments of lakeshore zone of the lakes in the middle and lower reaches of the Yangtze River. *Journal of Lake Science*, 26(3), 340-348.
- [21] Eleftheriou, A., Mouzourides, P., Biskos, G., Yiallourous, P., Kumar, P., & Neophytou, M. K.-A. (2023). The challenge of adopting mitigation and adaptation measures for the impacts of sand and dust storms in Eastern Mediterranean Region: a critical review. *Mitigation and Adaptation Strategies for Global Change*, 28(6), 33.
- [22] Farhangi, F., Sadegh-Niaraki, A., Razavi-Termeh, S. V., & Nahvi, A. (2023). Driver drowsiness modeling based on spatial factors and electroencephalography using machine learning methods: A simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour*, 98, 123-140.
- [23] Feng, Y., Long, H., Yang, F., Yang, F., Cheng, H., & Zhang, G. (2023). Warmth Favored Dust Activities on the Northeastern Qinghai-Tibet Plateau. *Geophysical research letters*, 50(11), e2023GL103781.

- [24] Furman, H. K. H. (2003). Dust storms in the Middle East: sources of origin and their temporal characteristics. *Indoor and Built Environment*, 12(6), 419-426.
- [25] Gholami, H., Mohamadifar, A., Sorooshian, A., & Jansen, J. D. (2020). Machine-learning algorithms for predicting land susceptibility to dust emissions: The case of the Jazmurian Basin, Iran. *Atmospheric pollution research*, 11(8), 1303-1315.
- [26] Hamidi, M. (2020). The key role of water resources management in the Middle East dust events. *Catena*, 187, 104337.
- [27] Hengl, T. (2018). Soil texture classes (USDA system) for 6 soil depths (0, 10, 30, 60, 100 and 200 cm) at 250 m (Version v02)[Data set]. Zenodo.
- [28] Jafari, R., Amiri, M., Asgari, F., & Tarkesh, M. (2022). Dust source susceptibility mapping based on remote sensing and machine learning techniques. *Ecological Informatics*, 72, 101872.
- [29] Jish Prakash, P., Stenichikov, G., Kalenderski, S., Osipov, S., & Bangalath, H. (2015). The impact of dust storms on the Arabian Peninsula and the Red Sea. *Atmospheric Chemistry and Physics*, 15(1), 199-222.
- [30] Karunasingha, D. S. K. (2022). Root mean square error or mean absolute error? Use their ratio as well. *Information Sciences*, 585, 609-629.
- [31] Khaniabadi, Y. O., Daryanoosh, S. M., Amrane, A., et al. (2017). Impact of Middle Eastern Dust storms on human health. *Atmospheric pollution research*, 8(4), 606-613.
- [32] Li, J., Garshick, E., Huang, S., & Koutrakis, P. (2021). Impacts of El Niño-Southern Oscillation on surface dust levels across the world during 1982–2019. *Science of The Total Environment*, 769, 144566.
- [33] Li, J., Li, C., & Zhang, S. (2022). Application of Six Metaheuristic Optimization Algorithms and Random Forest in the uniaxial compressive strength of rock prediction. *Applied Soft Computing*, 131, 109729.
- [34] Liu, R., Li, G., Wei, L., et al. (2022). Spatial prediction of groundwater potentiality using machine learning methods with Grey Wolf and Sparrow Search Algorithms. *Journal of Hydrology*, 610, 127977.
- [35] Liu, Y., Wang, G., Hu, Z., et al. (2020). Dust storm susceptibility on different land surface types in arid and semiarid regions of northern China. *Atmospheric Research*, 243, 105031.
- [36] McNally, A. (2018). FLDAS Noah Land Surface Model L4 Global Monthly 0.1× 0.1 degree (MERRA-2 and CHIRPS), Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC).
- [37] Mengmeng, Z., & Yian, L. (2018). *Signal sorting using teaching-learning-based optimization and random forest*. Paper presented at the 2018 17th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES).
- [38] Moniz, N., Branco, P., & Torgo, L. (2017). *Evaluation of ensemble methods in imbalanced regression tasks*. Paper presented at the First International Workshop on Learning with Imbalanced Domains: Theory and Applications.
- [39] Munkhtsetseg, E., Shinoda, M., Gillies, J. A., Kimura, R., King, J., & Nikolich, G. (2016). Relationships between soil moisture and dust emissions in a bare sandy soil of Mongolia. *Particuology*, 28, 131-137.
- [40] Naghibi, A., Hashemi, H., Zhao, P., et al. (2024). Spatiotemporal variability of dust storm source susceptibility during wet and dry periods: The Tigris-Euphrates River Basin. *Atmospheric pollution research*, 15(1), 101953.
- [41] Neelamani, S., & Al-Dousari, A. (2016). A study on the annual fallout of the dust and the associated elements into the Kuwait Bay, Kuwait. *Arabian Journal of Geosciences*, 9, 1-11.
- [42] Okin, G. S. (2022). Where and how often does rain prevent dust emission? *Geophysical research letters*, 49(4), e2021GL095501.
- [43] 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.
- [44] Park, S., Hamm, S.-Y., Jeon, H.-T., & Kim, J. (2017). Evaluation of logistic regression and multivariate adaptive regression spline models for groundwater potential mapping using R and GIS. *Sustainability*, 9(7), 1157.
- [45] Peng, Y. (2022). Application of Educational Psychology Based on Improved SPBO Optimization Algorithm in English Learning. *Frontiers in Psychology*, 13, 949568.
- [46] Pourhashemi, S., Asadi, M. A. Z., Boroughani, M., & Azadi, H. (2023). Mapping of dust source susceptibility by remote sensing and machine learning techniques (case study: Iran-Iraq border). *Environmental Science and Pollution Research*, 30(10), 27965-27979.
- [47] Rao, R. V., & Patel, V. (2013). An improved teaching-learning-based optimization algorithm for solving unconstrained optimization problems. *Scientia Iranica*, 20(3), 710-720.
- [48] Rao, R. V., Savsani, V. J., & Vakharia, D. (2011). Teaching-learning-based optimization: a novel method for constrained mechanical design optimization problems. *Computer-aided design*, 43(3), 303-315.

- [49] Razavi-Termeh, S. V., Sadeghi-Niaraki, A., & Choi, S.-M. (2019). Groundwater potential mapping using an integrated ensemble of three bivariate statistical models with random forest and logistic model tree models. *Water*, 11(8), 1596.
- [50] Razavi-Termeh, S. V., Sadeghi-Niaraki, A., Seo, M., & Choi, S.-M. (2023). Application of genetic algorithm in optimization parallel ensemble-based machine learning algorithms to flood susceptibility mapping using radar satellite imagery. *Science of The Total Environment*, 873, 162285.
- [51] Sarzaeim, P., Bozorg-Haddad, O., & Chu, X. (2018). Teaching-learning-based optimization (TLBO) algorithm. *Advanced optimization by nature-inspired algorithms*, 51-58.
- [52] Shao, Y. (2008). *Physics and modelling of wind erosion*: Springer.
- [53] Shogrkhodaei, S. Z., Razavi-Termeh, S. V., & Fathnia, A. (2021). Spatio-temporal modeling of PM_{2.5} risk mapping using three machine learning algorithms. *Environmental Pollution*, 289, 117859.
- [54] Speer, M. S. (2013). Dust storm frequency and impact over Eastern Australia determined by state of Pacific climate system. *Weather and Climate Extremes*, 2, 16-21.
- [55] Strong, J. D., Vecchi, G. A., & Ginoux, P. (2015). The response of the tropical Atlantic and West African climate to Saharan dust in a fully coupled GCM. *Journal of Climate*, 28(18), 7071-7092.
- [56] UNCCD. (2022). Sand and dust storms compendium: Information and guidance on assessing and addressing the risks: United Nations Convention to Combat Desertification.
- [57] Wang, N., Chen, J., Zhang, Y., Xu, Y., & Yu, W. (2023). The Spatiotemporal Characteristics and Driving Factors of Dust Emissions in East Asia (2000–2021). *Remote Sensing*, 15(2), 410.
- [58] Wang, W., Samat, A., Abuduwaili, J., De Maeyer, P., & Van de Voorde, T. (2023). Machine learning-based prediction of sand and dust storm sources in arid Central Asia. *International Journal of Digital Earth*, 16(1), 1530-1550.
- [59] Willmott, C. J., Ackleson, S. G., Davis, R. E., et al. (1985). Statistics for the evaluation and comparison of models. *Journal of Geophysical Research: Oceans*, 90(C5), 8995-9005.
- [60] WMO. (1995). Manual on Codes – International Codes. *WMO Report No.306*, Geneva, Switzerland.
- [61] Wu, C., Lin, Z., Shao, Y., Liu, X., & Li, Y. (2022). Drivers of recent decline in dust activity over East Asia. *Nature Communications*, 13(1), 7105.
- [62] Wu, D., Wang, S., Liu, Q., Abualigah, L., & Jia, H. (2022). An improved teaching-learning-based optimization algorithm with reinforcement learning strategy for solving optimization problems. *Computational Intelligence and Neuroscience*, 2022.
- [63] Xu, C., Guan, Q., Lin, J., et al. (2020). Spatiotemporal variations and driving factors of dust storm events in northern China based on high-temporal-resolution analysis of meteorological data (1960–2007). *Environmental Pollution*, 260, 114084.
- [64] Xue, R., & Wu, Z. (2019). A survey of application and classification on teaching-learning-based optimization algorithm. *IEEE Access*, 8, 1062-1079.
- [65] Yadav, R., & Kaur, M. (2024). Teaching learning based optimization-a review on background and development. *AIP Conference Proceedings*, 2986(1).
- [66] Zhang, J., Huang, Y., Wang, Y., & Ma, G. (2020). Multi-objective optimization of concrete mixture proportions using machine learning and metaheuristic algorithms. *Construction and Building Materials*, 253, 119208.
- [67] Zhao, C., Dabu, X., & Li, Y. (2004). Relationship between climatic factors and dust storm frequency in Inner Mongolia of China. *Geophysical research letters*, 31(1).
- [68] Zucca, C., Middleton, N., Kang, U., & Liniger, H. (2021). Shrinking water bodies as hotspots of sand and dust storms: The role of land degradation and sustainable soil and water management. *Catena*, 207, 105669.

Appendices

Parameters calculated to optimize RF

	By SPBO		By TLBO	
2000–2004	n_estimators	96	n_estimators	35
	max_features	4	max_features	4
	max_depth	68	max_depth	74
	min_samples_split	9	min_samples_split	4
	min_samples_leaf	3	min_samples_leaf	3
	max_samples	0.518	max_samples	0.905
	Minimized objective function value	0.195	Minimized objective function value	0.193
2005–2007	n_estimators	40	n_estimators	28
	max_features	2	max_features	3
	max_depth	52	max_depth	10
	min_samples_split	2	min_samples_split	2
	min_samples_leaf	6	min_samples_leaf	1
	max_samples	0.794	max_samples	0.741
	Minimized objective function value	0.196	Minimized objective function value	0.194
2008–2012	n_estimators	117	n_estimators	136
	max_features	2	max_features	3
	max_depth	145	max_depth	118
	min_samples_split	6	min_samples_split	5
	min_samples_leaf	5	min_samples_leaf	9
	max_samples	0.333	max_samples	0.899
	Minimized objective function value	0.208	Minimized objective function value	0.208
2013–2020	n_estimators	87	n_estimators	85
	max_features	3	max_features	3
	max_depth	117	max_depth	80
	min_samples_split	5	min_samples_split	6
	min_samples_leaf	8	min_samples_leaf	5
	max_samples	0.825	max_samples	0.580
	Minimized objective function value	0.182	Minimized objective function value	0.181

The number of trees in the forest (n_estimators), the number of features considered for the best split (max_features), the maximum tree depth (max_depth), the minimum number of samples required to split an internal node (min_samples_split), the minimum number of samples required at a leaf node (min_samples_leaf), and the number of samples to draw from X to train each base estimator (max_samples)