

Predicting the dust events frequency around a degraded ecosystem and determining the contribution of their controlling factors using gradient boosting-based approaches and Game theory

Zohre Ebrahimi-Khusfi (✉ zohreebrahimi2018@ujiroft.ac.ir)

University of Jiroft <https://orcid.org/0000-0002-1036-6386>

Fatemeh Dargahian

Agricultural Research Education and Extension Organization

Ali Reza Nafarzadegan

Hormozgan University

Research Article

Keywords: Dust storm, land degradation, machine learning, remote sensing, extreme gradient boosting, international wetland

Posted Date: April 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-417390/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

In this study, for the first time, an attempt was made to evaluate the performance of Gradient Boosting Machine (GBM) and extreme gradient boosting (XGB) models with linear, tree, and Dart boosters to predict monthly DEF (MDEF) around a degraded wetland in southwestern Iran. The monthly required data were obtained through observational data recorded at ground stations and satellite imagery from 1988 to 2018. The best predictor variables were selected among the eighteen climatic, terrestrial, and hydrological variables based on the multi-collinearity test (MCT) and Boruta algorithm. The models' performance was evaluated using the Taylor diagram. Game theory (i.e., SHAP values: SHV) was then used to determine the contribution of factors controlling MDEF in different seasons. Mean wind speed, maximum wind speed, rainfall, standardized precipitation evapotranspiration index (SPEI), soil moisture, erosive winds frequency, vapor pressure, vegetation area, water body area, and dried bed area of the wetland were confirmed as the best predictive variables. The XGB-linear and XGB-tree showed a higher capability in predicting the MDEF variations in summer and spring seasons. However, the XGB-Dart yielded a better than other study models in forecasting the MDEF in the autumn and winter seasons. The results also showed that the rainfall (SHV = 1.6), surface water discharge (SHV = 2.4), mean wind speed (SHV = 10.1), and erosive winds frequency (SHV = 1.6) had the largest contribution in the variability of MDEF in winter, spring, summer, and autumn, respectively. The results can be useful to provide different scenarios for combating hazards caused by wind erosion events around degraded wetlands.

1. Introduction

Wetlands are considered as the transitional area between aquatic and terrestrial ecosystems (Gokce 2018). In general, these valuable ecosystems are divided into two categories of natural and man-made wetlands and provide a wide range of services to humanity, including water supply, climate regulation, and carbon sequestration (Jia et al. 2020; Martins et al. 2020). Considering the 1% discount rate, the economic loss of not protecting West Asian wetlands will reach \$ 7.2 billion over the next 30 years (Eppink et al. 2014). In recent years, the water levels of these valuable ecosystems have declined due to climate change, meteorological droughts, and overuse of surface and groundwater resources (Cao et al. 2012; Ebrahimi-Khusfi et al. 2020a). According to a report provided by Davidson (2014), more than two-thirds of the world's wetlands have been destroyed since 1900. Wetland degradation, especially in arid areas, has led to the development of desertification and increased the occurrence of wind erosion and the activity of dust storms (Poornazari et al. 2020). In addition to climatic conditions, the frequency of dust events around degraded wetlands is a function of changes in the moisture content of wetlands and the flow discharge variations in the rivers leading to wetlands (Khusfi et al. 2017). Using the spectral indices derived from remotely-sensed data, it was demonstrated that the water level of many wetlands has decreased in many countries, including China (Jiang et al. 2017; Song et al. 2014), Nigeria (Ayanlade and Proske 2016), Ethiopia (Gebreslassie et al. 2014), India (Chatterjee et al. 2015), and Iran (Ebrahimi-Khusfi et al. 2020a).

The most important wetlands in Iran are Hamoun-e-pouzak, Hamoun-e-Sabori, Hamoun-e-Hirmand, Jazmourian, Gavkhoni, and Hour-al-Azim wetland (Rashki et al. 2017; Salmabai and Saeedi 2018; Vali et al. 2016). The Shadegan wetland is also a destroyed wetland in the southwest of Iran that has been conducted numerous works on its various aspects such as the ecological risk assessment of heavy metals in its sediments and water (Ashayeri and Keshavarzi 2019), its economic valuation (Shamsudin et al. 2011), identify the atrazine sources (Almasi et al. 2020), the impact of dust on its vegetation variations (Bayat et al. 2016), and the prediction of the influencing factors on conservation behavior of rural operators (Ghanian et al. 2015). However, no attempt has been made to predict the monthly dust events frequency (MDEF) and to identify its controlling factors around the Shadegan wetland. The dried bed of destroyed wetlands has become areas prone to soil erosion and dust production has negatively affected the life quality of surrounding cities (Dahmardeh 2016; Rashki et al. 2021; Shahraki et al. 2021).

Ahvaz metropolis, as one of the dustiest cities in the world, is located near Shadegan wetland which has suffered greatly from the increase in air pollution caused by wind erosion and dust events in recent decades (Salmabadi et al. 2020). Recognizing the major influencing agents on dust event frequency in different seasons is an appropriate strategy to deal with the dangers of dust storms in various regions (Ebrahimi-Khusfi et al. 2020c). Although the impact of various climatic and terrestrial factors on dust events has been addressed in some previous studies, no attempt has been made to determine the agents affecting MDEF in the degraded wetlands using advanced modeling techniques. Accordingly, in this study, we tried to analyze this issue for the destroyed Shadegan wetland in the southwest of Iran using the best algorithm selected among the stochastic gradient boosting machine (GBM) and extreme gradient boosting (XGB) algorithm with different boosters. The ability to reproduce and the ability to quantitatively analyze the contribution of factors affecting natural hazards and their potential for continuous updating are among the capabilities of machine learning models (Youssef and Pourghasemi 2021). Although the good performance of various machine learning models in spatial predicting of dust-sensitive areas and their relatively good performance in temporal predicting of dust storm index (DSI) has been proven in some previous works (Chen et al. 2019; Ebrahimi-Khusfi et al. 2020b; Gholami et al. 2020), the efficiency of XGB model with different boosters and GBM model has not been evaluated to forecast the MDEF in different seasons around degraded wetlands in the world. Thus, this evaluation was performed in this study. Given that strong and problematic correlations between some influencing variables may lead to deviations in the results, it is necessary to identify the most important variables based on feature selection techniques before modeling procedure. This issue has received less attention in dust studies, while it has received more attention in the present study. In dust studies, the multi-collinearity (MC) test has often been used for choosing variables with the least correlation coefficient. However, feature selection algorithms have not been used for the prediction of MDEF in different seasons in previous dust studies. Boruta is a well-known algorithm that has been utilized in researches related to other natural hazards such as floods, landslides, forest fire, gully, and soil salinity (Amiri et al. 2019; Pourghasemi et al. 2020; Xu et al. 2020). Therefore, in this paper, in addition to the MC test, we used this algorithm to select the most input data for predicting the variability in MDEF around the Shadegan wetland.

Furthermore, determining the contribution of influencing agents on the target variable variations and improve the interpretability of the predictive models are important issues in the modeling process (Gilpin et al. 2018). One of the robust approaches to improve interpretability and to analyze the contribution of environmental drivers is the game theory (Gu et al. 2021). Up to now, this approach has not been utilized in modeling MDEF variations across various geographical regions. Thus, SHapley Additive exPlanation (SHAP) values obtained by the cooperative game concept were used to improve the interpretability of the models used in this study and increase our knowledge about the contribution of influential variables.

In total, the major objectives of this study are (i) to choose the optimal input data combination for predicting MDEF in different seasons using the MC test and Boruta algorithm, (ii) to evaluate the GBM and XGB models with different boosters- algorithms for predicting the MDEF variations around the Shadegan international wetland, and (iii) to determine the contribution of the influencing factors on the MDEF in different seasons over the study are using game theory (i.e., SHAP values).

2. Study Area

Shadegan international wetland is the 34th international wetland in the world (<http://www.ramsar.org>), located in southwestern Iran, near the northern shores of the Persian Gulf (Fig. 1a). Three rivers, the Maleh, the Gupal, and the Jarrahi, lead to this wetland (Fig. 1b), and the wetland is mainly fed by the Jarrahi River (Ghorbani et al. 2016). Disruption of wetland hydrological systems is a result of the construction of the Maroon Reservoir Dam on the Jarrahi river, the development of irrigation networks in the upstream of the catchment, and the addition of agricultural wastewaters into the wetland in recent years (Ashayeri and Keshavarzi 2019; Sima and Tajrishy 2006). The prevailing wind direction in the northern and southern regions of the study area is mainly from the west and northwest, respectively (Fig. 1c). Based on the long-term average of monthly data from 1988 to 2018, mean annual rainfall, temperature and wind speed in this area varied between 47.2 to 298.7 mm, 24.3 to 27.8 °C, and 4 to 5 m/s, respectively.

3. Material And Methods

The framework used to perform the present research is shown in Fig. 2, and more details are provided in the following sections.

3.1 Data collection

The data used in this study are divided into two general categories. The first group is the data related to the target variable. MDEF is considered as the target variable in the current research. For calculating the MDEF, three-hourly data on dust events which are observed and recorded in synoptic stations with codes of 6 to 9, 30 to 35, and 98 were used. These data were obtained from the Islamic Republic of Iran Meteorological Organization (IRIMO) for the nearest synoptic stations located around the Shadegan wetland from 1998 to 2018.

The second group of data used in this study is the influential variables which are divided into three main groups: i) meteorological data, ii) terrestrial data, and iii) hydrological data.

Some meteorological data such as the mean monthly values of wind speed, precipitation, relative humidity, air temperature, maximum values of wind speed, temperature, relative humidity, and minimum values of temperature and relative humidity were acquired from IRIMO for the synoptic stations around the wetland for the years 1988 to 2018. Some other meteorological data such as vapor pressure, deficit vapor pressure, and downward surface shortwave radiation were obtained from the TerraClimate data set that provided by the University of Idaho on a monthly time scale at 4 km resolution (Abatzoglou et al. 2018). Furthermore, hourly wind velocity data was utilized to compute the frequency of erosive winds ($>6\text{m/s}$) in all months of the study period. Drought and potential evapotranspiration (PET) are other climatic drivers that affect dust events (Al-Khalidi et al. 2021; Moghanlo et al. 2021). In this work, the standardized precipitation-evapotranspiration index (SPEI) and PET were respectively calculated using the equations proposed by Vicente-Serrano et al. (2010) and Thornthwaite (Thornthwaite and Mather 1957) in Matlab software for all study months from 1988 to 2018.

Terrestrial data used in this work include the area of the vegetation cover, water body, and dried bed of the wetland. These data were obtained from the multi-temporal Landsat images for the study region (path: 165 and row: 39) from 1988 to 2018 via the Google Earth Engine (GEE) platform. In order to preprocess Landsat images, radiometric and atmospheric corrections were performed based on methods applied by (Ebrahimi-Khusfi et al. 2020a). In the image processing stage, the super vector machine (SVM) method was used to separate the classes of the vegetation cover, waterbody, and dried bed of the wetland. The training samples used in this classification were selected based on Google Earth images and field observations. Lastly, the areas of three classes of the water body and vegetation cover were calculated for all study months during 1988-2018 in GEE. Soil moisture is another ground-based variable used in this study, which was obtained from data provided by the University of Idaho. Due to the large number of images prepared on a monthly scale for the study period, only changes in the water body, vegetation cover, and dried bed of the Shadegan Wetland on an annual scale are shown in Fig. 3. Of note, in the modeling process, monthly data related to these parameters along with the hydro-climatic data have been used, which their monthly changes over the study period are presented in Fig.4.

The third group of data, the mean monthly values of surface water discharge to Shadegan wetland, was also obtained from Iran Water and Power Resources Development Company (IWPCO).

Temporal variations of all variables used in this study from 1988 to 2018 are indicated in Fig 4. It is worth noting that monthly data for each season were used to predict seasonal changes. In other words, for each season, data related to 93 months during the study period have been used.

3.2 Multi-collinearity test (MCT)

Selecting non-correlated attributes that affect the target attribute is one of the main steps before the modeling procedure that helps to reduce the prediction error (Adeboye et al. 2014). The tolerance coefficient (TC) is one of the most common techniques used to remove strongly correlated variables and to select non-correlated variables (Amare et al. 2021), which was applied in this study. The $TC > 0.1$ indicates a weak correlation between a variable with other predictor variables. For the current work, the MCT has been performed in SPSSv.26 software.

3.3 Boruta algorithm

The Boruta algorithm is a random forest-based algorithm and a good technique for selecting features. In this algorithm, the status of study features were determined as rejected, confirmed, or tentatively important (Hassanien et al. 2012). By growing randomness to the structure of this algorithm and collecting outcomes from the randomized samples, the oblique effect of random discord and relationships reduce (Pourghasemi et al. 2020). As a result, we get a clearer insight into the really important and influential features of the target variable. In the present work, The Boruta package was utilized for performing this algorithm in R statistical software.

3.4 Stochastic Gradient boosting machine(GBM)

The GBM model creates a set of decision trees in a sequential manner and the main goal is to reduce the difference between the measured and forecasted values by all its previous escapes (Natekin and Knoll 2013). Using this algorithm, decision trees are continuously trained and modified for the residuals of all their previous trees to make the best predictions. Generally, the GBM algorithm is trained by Equation (1): (Kong et al. 2020)

$$G_n(z) = G_{n-1}(z) + u\lambda^n L(z) \quad G_n(z) = G_{n-1}(z) + u\lambda^n L_n(z) \quad (1)$$

Here, $G_n(z)$ refers to the output model in which n reaches N . λ is computed in the procedure of reducing the loss function. $G_{n-1}(z)$ is utilized to illustrate the prevalent mood of the GBM model that has an $n-1$ decision tree. Hyper-parameters of the GBM model tuned by five-fold cross-validation are Boosting iterations and maximum tree depth. In this work, the R packages of caret and gbm were utilized to perform the GBM model.

3.5 Extreme gradient boosting (XGB) data mining model

The Extreme gradient boosting (XGB) algorithm was first introduced by Chen and Guestrin (2016). This algorithm uses a gradient boosting structure which has the benefits of parallel tree boosting. This can help increase modeling performance by overcoming the over-fitting problem and decrease the computation time by applying the parallel procedure (Bansal and Kaur 2018). This algorithm can work well for data sets that have missing values (Chen et al. 2019). Thus, three boosters of linear, tree, and Dart (dropout-introduced multiple additive regression tree) were used to predict the monthly DEF in our study area. Using the linear booster, linear relationships between influencing agents, and using two other boosters, more complex relationships between them are explored (Naghibi et al. 2020). For the XGB model, parameters controlling the tuning process (hyper-parameters) include learning rate, column sample rate, number of trees, and maximum tree depth.

In total, the XGBoost model is expressed by Equation (2): (Ebrahimi-Khusfi et al. 2020b)

$$Z_j = \sum_{k=1}^K f_k(x_j), \quad f_k \in F \quad (2)$$

In which, K denotes the decision trees. Equation (2) is used to evaluate the XGBoost model for reducing the overfitting problem:

$$S_j^{(t)} = \sum_{j=1}^k m(v_j, v_j) + \sum_{c=1}^c R(f_c), \quad \theta = (f_1, f_2, \dots, f_m) \quad (3)$$

In equation (3), m shows the loss function, k is the number of data, r refers to the regularization factor. It is defined as (Equation 4):

$$R = p_1 L + \frac{1}{2} p_2 \sum_{j=1}^L v_j^2 \quad (4)$$

where vs is the vector scores, L indicates the leaves number in a decision tree, p1 and p2 refer to the penalty factors.

The R packages of caret and xgboost were applied to tune and execute the XGB-based models with different boosters. In noting that the five-time five-fold cross-validation method was conducted with about 80% of the data set for the training of models and 20% of them (hold-out data) for testing.

3.6 Performance evaluation of predictive models

In this stage of the present study, the performance of the predictive models was assessed through a comparison between observed and predicted values of DEF in different seasons using the metrics provided in the Taylor diagram (equations 5 and 6) (Shabani et al. 2021):

$$r = \frac{n \sum_i ADEF_i PDEF_i - \sum_i ADEF_i \sum_i PDEF_i}{\sqrt{n \sum_i ADEF_i^2 - (\sum_i ADEF_i)^2} \sqrt{n \sum_i PDEF_i^2 - (\sum_i PDEF_i)^2}} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_i (ADEF_i - PDEF_i)^2} \quad (6)$$

Here, r and RMSE are correlation coefficient and root mean square of error in a study time series, respectively. The ADEF and PDEF indicate the actual and predicted values of DEF for the i th variable, respectively. Also, $\overline{ADEF_i}$ refers to the mean value of actual DEF data.

3.7 Interpretability and importance of controlling factors using Game theory

Improving the interpretability of machine learning models is one of the most important topics in studies associated with modeling (Kaur et al. 2020; Rice et al. 2020). Game theory, as one of the most important theories that investigate the strategic relationships between illustrative decision-makers, has engrossed the attention of some researchers to enhance the interpretability of data mining models (Kaur et al. 2020; Wu et al. 2017). SHAP measures the effect of attributes by considering the interaction with other attributes. Shapley values (SHV) obtained from the cooperative game concept is used to determine the share of the influencing agents on the target variable (Lundberg and Lee 2017), which is calculated by the following equation (7):

$$Phi_k = \frac{1}{|n|} \sum_{g \subseteq n \setminus \{k\}} [|g|!(n-|g|-1)!] [f(g \cup \{k\}) - f(g)] \quad (7)$$

In equation (7), Phi_k refers to the Shapley value and n denotes the total number of predictors. $f(g)$ is the result of the ML model that is explained by a set g of an attribute. In the present study, SHAP values were computed using R package SHAPforxgboost.

4. Results

4.1 Effective factors selected by Multi-collinearity test and Boruta algorithm

The outcomes obtained from the MC test to select the non-correlated variables for modeling MDEF in the study area are presented in Table 1. According to the MC test, the variables of Tmax, Tmin, RHmean, RHmax, RHmin, Vpd, and Dssr had a TC < 0.1 in all months of the study seasons. In addition to these factors, Autumn rainfall (Ra), winter Vap and summer Vap had a TC < 0.1, highlighting that a problematic correlation between these parameters in the studies seasons. Therefore, these parameters were excluded and non-correlated variables with TC < 0.1 were selected at this stage of the present study. They include DbA, WA, Wsmean, Wsmax, WS6, Ra (except for autumn), SPEI, SM, Q, VegA, and Vap (except for winter and summer). These variables were introduced to the Boruta algorithm for choosing the most important predictive variables. The results are summarized in Fig. 5.

Table 1
Tolerance coefficients of explanatory variables according to the MC test.

Definition (unit)	Abbreviation	Winter	Spring	Summer	Autumn
Dried bed area (Km ²)	DbA	0.89	0.68	0.54	0.37
Water body area (Km ²)	WA	0.36	0.36	0.35	0.52
Mean wind speed (m/s)	Wsmean	0.50	0.31	0.29	0.50
Max wind speed (m/s)	Wsmax	0.62	0.81	0.62	0.66
Winds speed > 6m/s	WS6	0.37	0.32	0.37	0.35
Rainfall (mm)	Ra	0.13	0.23	0.41	0.08
Max temperature (°C)	Tmax	0.01	0.01	0.00	0.01
Min temperature(°C)	Tmin	0.03	0.02	0.01	0.03
Mean relative humidity (%)	RHmean	0.00	0.00	0.01	0.01
Max relative humidity (%)	RHmax	0.01	0.01	0.02	0.01
Min relative humidity (%)	RHmin	0.01	0.01	0.06	0.01
Standardized Precipitation-Evapotranspiration Index (SPEI)	SPEI	0.13	0.24	0.45	0.22
Vapor pressure (Kpa)	Vap	0.08	0.28	0.03	0.13
Vapor pressure deficit (Kpa)	Vpd	0.02	0.01	0.00	0.01
Downward surface solar radiation (W/m ²)	Dssr	0.06	0.03	0.03	0.02
Discharge (m ³ /s)	Q	0.24	0.21	0.37	0.66
Vegetation area (Km ²)	VegA	0.15	0.61	0.24	0.60

According to the importance score obtained from the Boruta algorithm (Fig. 5), Wsmean (12.5), Rainfall (8.2), Wsmax (6.2), and discharge (3.8) were detected as the most influencing variables for the occurrence of the winter dust events in the study area. The highest score of influential variables for occurrence of the spring dust events was assigned to WS6 (7.5), Wsmean (7.1), Q (6.8), and SM (4.9). Furthermore, the highest value of Z-score was achieved for WS6 (12.1), WSmean (9.8), WA (5.2), and Wsmax (4.8) for the occurrence of summer dust events around the Shadegan wetland. Six variables of WS6 (8.2), Ra (7.8), Wsmean (4.2), WA (4.3), Wsmax (3.8), and Vap (2.8) were confirmed as the most important drivers affecting autumn dust events in the study area. Using the Boruta, some variables were detected as tentative variables. For winter dust events, WA, VegA, DbA, and SPEI, and for spring events, WA had a status of tentative (Z-score: 2.1–2.5). Similar status was observed for SM, VegA, and DbA in summer and for SM and SPEI in the autumn season. We used these features for modeling MDEF.

However, other agents were rejected and have not been used for modeling monthly variations of DEF around the Shadegan wetland.

4.2 Hyper-parameters tuning results

The optimal values of tuned parameters for different seasons are summarized in Fig. 6. The best-tuned parameters of the GBM model for winter, spring, summer, and autumn were respectively observed in max tree depth of 1, 2, 3, 1, and ntrees of 50 with RMSE values of 6.15, 10.42, 8.71, and 4.21, respectively. For the XGB-L model, the optimum values of lambda (alpha) were estimated at 0.1 (0.0001), 0.1 (0), 0.0001 (0.0001), and 0 (0.1) for, respectively, winter (RMSE = 6.22), spring (RMSE = 12.32), summer (RMSE = 10.32), and autumn (RMSE = 5.38). In the XGB-T model, the controlling parameters of the training procedure including maximum tree depth, colsample bytree, eta, and subsample had respectively values of 2, 0.8, 0.3, and 0.75 for winter, values of 1, 0.8, 0.3, and 1 for spring, 1, 0.8, 0.3, and 1 for summer, and 1, 0.8, 0.3, and 1 for autumn. The RMSE values were respectively estimated at 5.7, 12.3, 9.1, and 4.4 for the study seasons. The optimal values of maximum tree depth, colsample bytree, eta, and subsample in the XGB-D model were obtained 2, 0.8, 0.4, 0.75 and 2, 0.6, 0.4, 0.5, respectively, for winter and spring seasons. However, values of 1, 0.8, 0.3, and 1 with RMSE = 9.1, and values of 1, 0.8, 0.4, and 0.75 with RMSE = 1.3 were assigned to the above-mentioned hyper-parameters for summer and autumn, respectively.

4.3 Models' Performance for predicting MDEF

The performance of GBM and XGB with different boosters for predicting MDEF in different seasons using the Taylor diagram are shown in Fig. 7. The r and RMSE are important performance evaluation indicators included in this diagram.

According to the Taylor diagram related to training data sets, the GBM model resulted in $r = 0.72$ and RMSE = 4.9 while XGB-L, XGB-T, and XGB-D models resulted in $r = 0.99$ with RMSE of 0.0008, 0.51, and 0.44, respectively, for predicting MDEF in winter across the study area (Fig. 7a). In modeling MDEF related to spring, the (r and RMSE) metrics had respectively values of (0.74 and 7.8), (0.99 and 0.0008), (0.79, 0.8) and (0.79 and 7.3) based on the GBM, XGB-L, XGB-T, and XGB-D models (Fig. 7b).

In forecasting the MDEF in the summer season, the GBM model resulted in $r = 0.78$ and RMSE = 6.7. Furthermore, values of 0.99 and 0.0006, 0.88 and 5.2, 0.74 and 7.4 were obtained based on the XGB-L, XGB-T, and XGB-D models for two accuracy metrics of r and RMSE, respectively (Fig. 7c). In the MDEF prediction of autumn, the r values of 0.75, 0.99, 0.86, and 0.88 were respectively estimated using the GBM, XGB-L, XGB-T, and XGB-D models. However, the RMSE was estimated at 3.4, 0.02, 2.7, and 2.6, respectively, according to the above-mentioned models (Fig. 7d).

The study models were then performed for the prediction of MDEF using a hold out data set. According to the obtained results, it was found that for MDEF forecasting in summer months, the XGB-L and XGB-T resulted in $r = 0.86$ and 0.62 and RMSE = 9 and 11.8, respectively. Also, values of 0.63 and 0.72 as well as 12.5 and 11.7 were obtained after performing the XGB-D and GBM models on the holdout data sets of

summer months (Fig. 8c). For forecasting MDEF in spring season, the (r ; RMSE) had respectively values of (0.76; 2.6), (0.67; 5.2), (0.78; 2.6) and (0.82; 2.1) based on the GBM, XGB-L, XGB-T, and XGB-D models (Fig. 8b).

4.4 Driving forces of MDEF

In this study, SHAP values were calculated to identify and prioritize the contribution of the driving forces affecting MDEF changes in different seasons around the study area. The results are shown in Fig. 9.

Based on SHAP values (SHV), Ra (SHV = 1.5) and Wa (SHV = 1.4) followed by WSmean (SHV = 1.2), WSmax (SHV = 1.05), Q (SHV = 1.01), DbA (SHV = 0.9), VegA (SHV = 0.72), and SPEI (SHV = 0.61) were the most driving forces for winter dust events in our study area (Fig. 9a). In spring season, the priority of influencing various agents was as follows: Q (SHV = 2.3), WSmean (SHV = 1.7), SM (SHV = 1.5), WA (SHV = 1.3), and WS6 (SHV = 1.1) (Fig. 9b). For the summer season, WSmean (SHV = 10.1), WS6 (SHV = 4.9), DbA (SHV = 2.3), VegA (SHV = 1.9), WSmax (SHV = 1.5), SM (SHV = 0.8), and WA (SHV = 0.7) were respectively detected as the major causes of variability in summer DEF across our study area (Fig. 9c). In addition, we found that 8 important factors affect the fall events, which are in order of importance: WS6 (SHV = 1.5), WSmean (SHV = 0.88), Ra (SHV = 0.84), WA (SHV = 0.63), WSmax (SHV = 0.59), Vap (SHV = 0.32), SM (SHV = 0.21), and SPEI (SHV = 0.12) (Fig. 9d).

5. Discussion

In this study, the best combination of variables affecting MDEF in different seasons was selected from climatic, terrestrial, and hydrological variables using the MC test and Boruta algorithm. These techniques provide quantitative results and have the important advantage that permits the potential comparison of research in various parts of the world (Ebrahimi-Khusfi et al. 2020c; Pourghasemi et al. 2020). Based on these techniques, the higher importance of some climatic parameters such as wind speed, erosive winds frequency, and rainfall as well as some ground-based parameters such as the dried bed of the wetland was confirmed to predict MDEF in all seasons (Fig. 3). In addition to these factors, the importance of monthly changes in the surface water discharge to Shadegan wetland was also confirmed to predict and analyze the influential factors in winter and spring seasons. In the study conducted by Ebrahimi-Khusfi et al. (2020b), the most important parameters used to predict the seasonal dust storm index in arid regions of Iran were wind speed, air temperature, rainfall, evapotranspiration, and vegetation cover. Of note, these factors were only chosen by the MC test, while the factors selected in this work were selected after performing two techniques of Boruta and MC test. Ebrahimi-Khusfi et al. (2020c) stated that in addition to these factors, the meteorological drought index was also important to predict temporal changes in dust concentrations over semi-arid regions of Iran. At the spatial scale, some important confirmed factors for predicting dust-prone areas were soil texture, soil bulk density, Dem, vegetation, precipitation, organic matter, land use, and wind speed (Gholami et al. 2020). These findings indicate the importance of some variables such as rainfall and wind speed on dust events at both temporal and spatial scales. The best input data sets were utilized to predict MDEF in different around the Shadegan wetland. In the modeling

procedure, it was observed that the performance of study models was not similar for predicting MDEF around the Shadegan international wetland (Figs. 7 and 8).

Based on the trained data, the high performance of the XGB model with linear booster was observed in most seasons (Fig. 7). According to the results of performance evaluation in testing data sets (Fig. 8), the use of this model only increased the prediction accuracy of the MDEF in summer by at least about 15% compared to other study models (Fig. 8c). However, its higher performance than the other three models for predicting monthly changes in the target variable was not observed in other seasons (Fig. 8; a, b, and d). In winter, the lowest performance was observed for the XGB-tree model while the higher and almost similar performance was observed by the other three models, especially by the XGB-Dart (Fig. 8a). In the contrary, the highest performance of MDEF prediction was observed for the XGB-tree model in the spring months. The yield of predictive models was as follows for this season: XGB-T > XGB-D > XGB-L > GBM (Fig. 8b). The forecast performance of monthly changes in the target variable across the autumn was improved by this XGB-Dart by 4%, 7%, and 16% compared to the XGB-tree, GBM, and XGB-linear models, respectively (Fig. 8d).

In total, the results of performance assessment according to the hold out data (Fig. 8) showed a stronger capability of XGB-D for the prediction of the MDEF in cold seasons. The higher efficiency of XGB-L and XGB-T was also demonstrated for prediction of the DEF in the summer and spring months. In other words, these results indicate that the linear booster in the XGB model causes weak generalization capability in winter, spring and autumn. Although the capability of the XGB model is poorly evaluated to predict dust storms in arid regions of Iran (Ebrahimi-Khusfi et al. 2020b), its high efficiency has been proven in forecasting PM_{2.5} (Ma et al. 2020) and other natural hazards, including the gully erosion (Chen et al. 2021) and landslide studies (Sahin 2020). Differences in the results related to the performance of the XGB model with past dust studies may be due to differences in the criteria considered for measuring dust occurrences and the studied years. By comparing the outcomes of this study with the Ebrahimi-Khusfi et al. (2020b) findings, it may also be concluded that the XGB model has a greater ability to predict the frequency of dust events, while it has less ability to predict the dust storm index in the desert areas.

In addition to evaluating the performance of the models, we used SHAP values to improve the interpretability of selected models as well as to prioritize the drivers influencing the MDEF based on the optimal model for each season. Prioritizing influential factors the natural hazards can help policymakers provide better decisions and take soil conservation measures (Ebrahimi-Khusfi et al. 2020c; Pourghasemi et al. 2020). According to SHAP values, variations in MDEF were mostly affected by changes in rainfall and the dried surface of the wetland, followed by changes in wind speed and surface water discharge to Shadegan wetland during the winter season of the study period (1988 to 2018). The higher contribution of discharge, wind speed, and soil moisture parameters was observed for controlling the MDEF variations in the spring season. For summer and autumn, it was found that the changes in dust events were mainly occurred due to changes in the average surface wind speed and the frequency of erosive winds (Figs. 9c and 9d). The effect of changes in the dried surface of the wetland on dust storms around Shadegan wetland was mainly observed in the summer months (Fig. 9c). Meteorological drought is another

important climatic parameter, the impact of which is undeniable, especially in the winter and autumn months. Vegetation cover also plays an important role in intensifying or weakening wind erosion events, which in our study area, its effect was more observed on summer events. In addition, vapor pressure was identified as the sixth most influential factor in autumn dust events frequency around the study area. Generally, these results indicate that the priority of influencing climatic, terrestrial, and hydrological factors on the frequency of dust events during different seasons is different. According to Baltaci (2021), erosive wind speed (> 6.5 m/s) is one of the main parameters for the occurrence of dust storms. However, Ebrahimi-Khusfi et al. (2020b) reported that dust events in different seasons are influenced by different factors. For example, wind speed, vegetation cover, and maximum wind speed have been respectively identified as the most important agents for the activity of dust events in winter, spring, summer, and autumn over arid regions of Iran. It has been reported that wind speed and vegetation are two driving forces that a great impact on Middle East dust events (Li et al. 2020). As early mentioned, changes in the dried bed of Shadegan wetland have also been effective on dust events around the wetland. Some other researchers have also concluded that the drying up of Iranian wetlands, especially Hamoun, Meighan, and Jazmourian, had a great impact on intensifying the spread of dust in these areas (Arjmand et al. 2018; Ebrahimi-Khusfi et al. 2020a; Gholami et al. 2020). In addition, the reduction in inflow discharge to the Gavkhooni wetland in central Iran has led to the wetland drying up and degradation in air quality due to increased dust particle emissions (Khusfi et al. 2017), as we observed for our study area.

Meteorological drought occurs due to a long-term decrease in rainfall in an area, which in turn can make the soil of an area more prone to wind erosion (Teng et al. 2021). The impact of this phenomenon on wind erosion events over different parts of the world has been proven (Arcusa et al. 2020; Kandakji et al. 2021; Knapp et al. 2020), which supports our outcomes of the present research. Soil moisture was recognized as the third, sixth, and seventh dust control factors in summer, autumn, and winter, respectively. According to Han et al. (2021), this important terrestrial agent can shift the activity rate of dust events by affecting the soil erosion threshold. Some researchers have also concluded that reducing soil moisture content had remarkable impacts on dust emissions across arid and semi-arid areas (Meng et al. 2018; Wang et al. 2014). In general, the results of this study showed the combined effect of some hydro-climatic factors, including precipitation, wind speed, erosive winds velocity, discharge, and some terrestrial drivers, especially dried bed of wetland and soil moisture in the occurrence of dust events around the Shadegan wetland. In other words, these controlling factors had a great contribution in predicting the MDEF around the destroyed wetland of Shadegan, in southwestern Iran.

6. Conclusion

One of the appropriate strategies for sustainable development in areas prone to dust storms, such as degraded wetlands, is to predict the frequency of these events and to detect the parameters that control them in different seasons. In this work, GBM, XGB-linear, XGB-tree, and XGB-Dart models were used to predict the MDEF in various seasons around a destroyed wetland, namely Shadegan wetland, in southwestern Iran. The MC test and Boruta algorithm helped us to choose the best combination of input data to forecast the MDEF in different seasons. Rainfall, wind speed, erosive winds frequency, discharge,

metrological drought, vapor pressure, the v dried bed of wetland, and soil moisture confirmed by these techniques for predicting the target variable in the study area. The higher yield of the XGB-Dart model for predicting the variability of MDEF in winter and autumn and the XGB-L and XGB-tree for predicting MDEF in summer and spring, respectively, were proved by the Taylor diagram. The interpretability of optimal models using the SHAP values revealed that precipitation, followed by wind speed, discharge, dried bed area, vegetation, and drought had the highest contribution in predicting MDEF in the winter season. However, discharge, wind speed, soil moisture, water area, and frequency of erosive winds were the major causes of MDEF variations around the Shadegan wetland in the spring of during the study period (1988–2018). Surface winds velocity and frequency of strong winds had a great contribution in predicting MDEF in summer and autumn seasons across the study area. In the contrary, monthly changes in the water area of the wetland in summer and meteorological drought in autumn had the least share in the forecasting target variable. Researchers and stakeholders can use these results to mitigate the undesirable impacts of dust storm activity in areas around degraded wetlands, which can lead to a reduction in desertification and increase sustainable development in a susceptible ecosystem to dust events.

Declarations

Acknowledgments

The authors thank the Islamic Republic of Iran Meteorological Organization (IRIMO) for providing the meteorological data, Iran Water and Power Resources Development Company (IWPCO) for making the surface water discharge data to Shadgan wetland, and the Google Earth Engine team for allowing access to the OpenLandMap, TerraClimate datasets, and MODIS products.

Declarations

Ethical Approval:

Not applicable.

Consent to Participate:

Not applicable.

Consent to Publish:

Not applicable.

Authors' contributions:

ZE analyzed and interpreted the climate and satellite data, and was a major contributor in writing the manuscript. FD obtained climatic information from the Meteorological Organization of Iran and prepared some maps. AN performed modeling process. All authors read and approved the final manuscript.

Funding:

No funding was received for conducting this study.

Competing:

The authors declare that they have no competing interests.

Availability of data and materials:

The datasets used during the current study are available from the corresponding author on reasonable request.

References

- Abatzoglou JT, Dobrowski SZ, Parks SA, Hegewisch KC (2018) TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015 *Scientific data* 5:1-12
- Adeboye N, Fagoyinbo I, Olatayo T (2014) Estimation of the effect of multicollinearity on the standard error for regression coefficients *Journal of Mathematics* 10:16-20
- Al-Khalidi J, Bakr D, Abdullah AA (2021) Synoptic Analysis of Dust Storm in Iraq *EnvironmentAsia* 14
- Almasi H et al. (2020) Spatial distribution, ecological and health risk assessment and source identification of atrazine in Shadegan international wetland, Iran *Marine Pollution Bulletin* 160:111569
- Amare S, Langendoen E, Keesstra S, Ploeg Mvd, Gelagay H, Lemma H, van der Zee SE (2021) Susceptibility to Gully Erosion: Applying Random Forest (RF) and Frequency Ratio (FR) Approaches to a Small Catchment in Ethiopia *Water* 13:216
- Amiri M, Pourghasemi HR, Ghanbarian GA, Afzali SF (2019) Assessment of the importance of gully erosion effective factors using Boruta algorithm and its spatial modeling and mapping using three machine learning algorithms *Geoderma* 340:55-69
- Arcusa SH, McKay NP, Routson CC, Munoz SE (2020) Dust-drought interactions over the last 15,000 years: A network of lake sediment records from the San Juan Mountains, Colorado *The Holocene* 30:559-574
- Arjmand M, Rashki A, Sargazi H (2018) Monitoring of spatial and temporal variability of desert dust over the Hamoun e Jazmurian, Southeast of Iran based on the Satellite Data *Scientific-Research Quarterly of Geographical Data (SEPEHR)* 27:153-168
- Ashayeri NY, Keshavarzi B (2019) Geochemical characteristics, partitioning, quantitative source apportionment, and ecological and health risk of heavy metals in sediments and water: A case study in Shadegan Wetland, Iran *Marine pollution bulletin* 149:110495

- Ayanlade A, Proske U (2016) Assessing wetland degradation and loss of ecosystem services in the Niger Delta, Nigeria *Marine and Freshwater Research* 67:828-836
- Baltaci H (2021) Meteorological characteristics of dust storm events in Turkey *Aeolian Research* 50:100673
- Bansal A, Kaur S Extreme gradient boosting based tuning for classification in intrusion detection systems. In: *International Conference on Advances in Computing and Data Sciences*, 2018. Springer, pp 372-380
- Bayat R, Jafari S, Ghermezcheshmeh B, Charkhabi A (2016) Studying the effect of dust on vegetation changes (case study: Shadegan Wetland, Khuzestan)
- Cao C et al. (2012) Wetland changes and droughts in southwestern China *Geomatics, Natural Hazards and Risk* 3:79-95
- Chatterjee K, Bandyopadhyay A, Ghosh A, Kar S (2015) Assessment of environmental factors causing wetland degradation, using Fuzzy Analytic Network Process: A case study on Keoladeo National Park, India *Ecological Modelling* 316:1-13
- Chen T, Guestrin C Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016. pp 785-794
- Chen W, Lei X, Chakraborty R, Pal SC, Sahana M, Janizadeh S (2021) Evaluation of different boosting ensemble machine learning models and novel deep learning and boosting framework for head-cut gully erosion susceptibility *Journal of Environmental Management* 284:112015
- Chen Z-Y et al. (2019) Extreme gradient boosting model to estimate PM_{2.5} concentrations with missing-filled satellite data in China *Atmospheric Environment* 202:180-189
- Dahmardeh M (2016) Assessment of drought damage of Hamoun wetland on health condition of inhabitants of Sistan region *World Review of Science, Technology and Sustainable Development* 12:335-352
- Davidson NC (2014) How much wetland has the world lost? Long-term and recent trends in global wetland area *Marine and Freshwater Research* 65:934-941
- Ebrahimi-Khusfi Z, Ghazavi R, Zarei M (2020a) The Effect of Climate Changes on the Wetland Moisture Variations and Its Correlation with Sand-Dust Events in a Semiarid Environment, Northwestern Iran *Journal of the Indian Society of Remote Sensing* 48:1797-1808
- Ebrahimi-Khusfi Z, Taghizadeh-Mehrjardi R, Mirakbari M (2020b) Evaluation of machine learning models for predicting the temporal variations of dust storm index in arid regions of Iran *Atmospheric Pollution Research*

- Ebrahimi-Khusfi Z, Taghizadeh-Mehrjardi R, Nafarzadegan AR (2020c) Accuracy, uncertainty, and interpretability assessments of ANFIS models to predict dust concentration in semi-arid regions *Environmental Science and Pollution Research*:1-15
- Eppink FV, Brander LM, Wagtendonk AJ (2014) An initial assessment of the economic value of coastal and freshwater wetlands in West Asia *Land* 3:557-573
- Gebreslassie H, Gashaw T, Mehari A (2014) Wetland degradation in Ethiopia: causes, consequences and remedies *Journal of Environment and Earth Science* 4:40-48
- Ghanian M, Bakhshi A, YUSEFI HR, Hasheminejad A (2015) Neural network analysis to predict factors affecting conservation behavior of rural operators of Shadegan Wetland
- Gholami H, Mohamadifar A, Sorooshian A, Jansen JD (2020) Machine-learning algorithms for predicting land susceptibility to dust emissions: The case of the Jazmurian Basin, Iran *Atmospheric Pollution Research* 11:1303-1315
- Ghorbani R, Hosseini S, Hedayati S, Hashemi S, Abolhasani M (2016) Evaluation of effects of physico-chemical factors on chlorophyll-a in Shadegan International Wetland-Khouzestan Province-Iran *Iranian Journal of Fisheries Sciences* 15:360-368
- Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L Explaining explanations: An overview of interpretability of machine learning. In: 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA), 2018. IEEE, pp 80-89
- Gokce D (2018) Introductory Chapter: Wetland Importance and Management. In: *Wetlands Management- Assessing Risk and Sustainable Solutions*. IntechOpen,
- Gu J, Yang B, Brauer M, Zhang KM (2021) Enhancing the Evaluation and Interpretability of Data-Driven Air Quality Models *Atmospheric Environment* 246:118125
- Han T, Pan X, Wang X (2021) Evaluating and improving the sand storm numerical simulation performance in Northwestern China using WRF-Chem and remote sensing soil moisture data *Atmospheric Research* 251:105411
- Hassanien AE, Salem A-BM, Ramadan R, Kim T-h (2012) *Advanced Machine Learning Technologies and Applications: First International Conference, AMLTA 2012, Cairo, Egypt, December 8-10, 2012, Proceedings* vol 322. Springer,
- Jia M, Mao D, Wang Z, Ren C, Zhu Q, Li X, Zhang Y (2020) Tracking long-term floodplain wetland changes: A case study in the China side of the Amur River Basin *International Journal of Applied Earth Observation and Geoinformation* 92:102185

- Jiang W, Lv J, Wang C, Chen Z, Liu Y (2017) Marsh wetland degradation risk assessment and change analysis: A case study in the Zoige Plateau, China *Ecological Indicators* 82:316-326
- Kandakji T, Gill TE, Lee JA (2021) Drought and land use/land cover impact on dust sources in Southern Great Plains and Chihuahuan Desert of the US: Inferring anthropogenic effect *Science of The Total Environment* 755:142461
- Kaur H, Nori H, Jenkins S, Caruana R, Wallach H, Wortman Vaughan J Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020. pp 1-14
- Khusfi Z, Vali A, Khosroshahi M, Ghazavi R (2017) The role of dried bed of Gavkhooni wetland on the production of the internal dust using remote sensing and storm roses (case study: Isfahan province) *Iranian Journal of Range and Desert Research* 24
- Knapp AK et al. (2020) Resolving the Dust Bowl paradox of grassland responses to extreme drought *Proceedings of the National Academy of Sciences* 117:22249-22255
- Kong G, Lin K, Hu Y (2020) Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU *BMC Medical Informatics and Decision Making* 20:1-10
- Li J, Garshick E, Al-Hemoud A, Huang S, Koutrakis P (2020) Impacts of meteorology and vegetation on surface dust concentrations in Middle Eastern countries *Science of The Total Environment* 712:136597
- Lundberg S, Lee S-I (2017) A unified approach to interpreting model predictions *arXiv preprint arXiv:170507874*
- Ma J, Yu Z, Qu Y, Xu J, Cao Y (2020) Application of the XGBoost machine learning method in PM_{2.5} prediction: A case study of Shanghai Aerosol and Air Quality Research 20:128-138
- Martins VS, Kaleita AL, Gelder BK, Nagel GW, Maciel DA (2020) Deep neural network for complex open-water wetland mapping using high-resolution WorldView-3 and airborne LiDAR data *International Journal of Applied Earth Observation and Geoinformation* 93:102215
- Meng Z, Dang X, Gao Y, Ren X, Ding Y, Wang M (2018) Interactive effects of wind speed, vegetation coverage and soil moisture in controlling wind erosion in a temperate desert steppe, Inner Mongolia of China *Journal of Arid Land* 10:534-547
- Moghanlo S, Alavinejad M, Oskoei V, Saleh HN, Mohammadi AA, Mohammadi H, DerakhshanNejad Z (2021) Using artificial neural networks to model the impacts of climate change on dust phenomenon in the Zanjan region, north-west Iran *Urban Climate* 35:100750
- Naghbi SA, Hashemi H, Berndtsson R, Lee S (2020) Application of extreme gradient boosting and parallel random forest algorithms for assessing groundwater spring potential using DEM-derived factors *Journal*

of Hydrology 589:125197

Natekin A, Knoll A (2013) Gradient boosting machines, a tutorial *Frontiers in neurorobotics* 7:21

Poornazari N, Khalilimoghadam B, Hazbavi Z, Bagheri Bodaghabadi M (2020) Land degradation assessment in the dust hotspot of southeastern Ahvaz, Iran *Land Degradation & Development*

Pourghasemi HR, Kariminejad N, Amiri M, Edalat M, Zarafshar M, Blaschke T, Cerda A (2020) Assessing and mapping multi-hazard risk susceptibility using a machine learning technique *Scientific reports* 10:1-11

Rashki A, Arjmand M, Kaskaoutis D (2017) Assessment of dust activity and dust-plume pathways over Jazmurian Basin, southeast Iran *Aeolian Research* 24:145-160

Rashki A, Middleton N, Goudie A (2021) Dust storms in Iran–Distribution, causes, frequencies and impacts *Aeolian Research* 48:100655

Rice JS, Saia SM, Emanuel RE (2020) Improved Accuracy of Watershed-Scale General Circulation Model Runoff Using Deep Neural Networks

Sahin EK (2020) Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest *SN Applied Sciences* 2:1-17

Salmabadi H, Khalidy R, Saeedi M (2020) Transport routes and potential source regions of the Middle Eastern dust over Ahvaz during 2005–2017 *Atmospheric Research* 241:104947

Salmabai H, Saeedi M Areal fluctuations monitoring of Al-Azim/Al-Havizeh wetland during the 1986–2017 period, using time-series Landsat data. In: *The 2 nd international conference on strategic ideas for architecture urbanism, geography, and the environment, Mashhad, Iran, 2018.*

Shabani E, Hayati B, Pishbahar E, Ghorbani MA, Ghahremanzadeh M (2021) A novel approach to predict CO2 emission in the agriculture sector of Iran based on Inclusive Multiple Model *Journal of Cleaner Production* 279:123708

Shahraki AS, Shahraki J, Monfared SAH (2021) An integrated model for economic assessment of environmental scenarios for dust stabilization and sustainable flora–fauna ecosystem in international Hamoun wetland *Environment, Development and Sustainability* 23:947-967

Shamsudin MN, Radam A, Rahim KA, Yacob MR, Muda A, Yazid M (2011) Economic valuation of Shadegan International Wetland, Iran: notes for conservation *Regional Environmental Change* 11:925-934

Sima S, Tajrishy M Water allocation for wetland environmental water requirements: the case of Shadegan Wetland, Jarrahi Catchment, Iran. In: *World Environmental and Water Resource Congress 2006: Examining the Confluence of Environmental and Water Concerns, 2006.* pp 1-10

Song K, Wang Z, Du J, Liu L, Zeng L, Ren C (2014) Wetland degradation: its driving forces and environmental impacts in the Sanjiang Plain, China *Environmental Management* 54:255-271

Teng Y, Zhan J, Liu W, Sun Y, Agyemang FB, Liang L, Li Z (2021) Spatiotemporal dynamics and drivers of wind erosion on the Qinghai-Tibet Plateau, China *Ecological Indicators* 123:107340

Thornthwaite CW, Mather JR (1957) Instructions and tables for computing potential evapotranspiration and the water balance. Centerton,

Vali A, Ebrahimi Z, Khosroshahi M, Ghazavi R (2016) Determination of the importance of hydro-climate parameters on drying in Gavkhooni wetland using artificial neural network and remote sensing data *Desert Ecosystem Engineering Journal* 5:79-94

Vicente-Serrano SM, Beguería S, López-Moreno JI (2010) A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index *Journal of climate* 23:1696-1718

Wang L, Shi Z, Wu G, Fang N (2014) Freeze/thaw and soil moisture effects on wind erosion *Geomorphology* 207:141-148

Wu Q, Ren H, Gao W, Ren J (2017) Benefit allocation for distributed energy network participants applying game theory based solutions *Energy* 119:384-391

Xu L, Rossel RAV, Lee J, Wang Z, Ma H (2020) A simple approach to estimate coastal soil salinity using digital camera images *Soil Research* 58:737-747

Youssef AM, Pourghasemi HR (2021) Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia *Geoscience Frontiers* 12:639-655

Figures

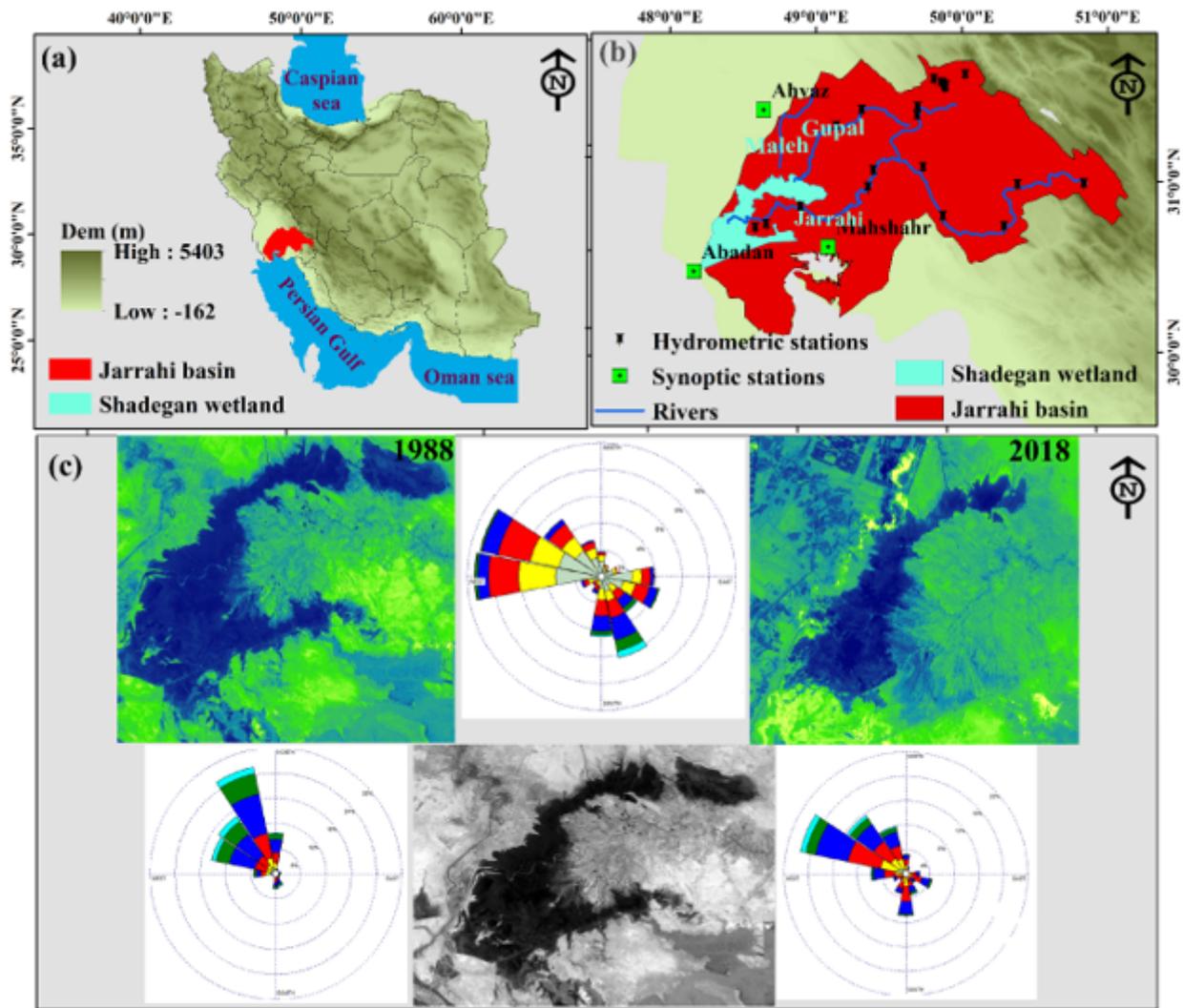


Figure 1

Geographical location of (a) Shadegan international wetland in Iran, (b) synoptic and hydrometric stations as well as main rivers in Jarrahi basin, and (c) prevailing winds direction around the wetland. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

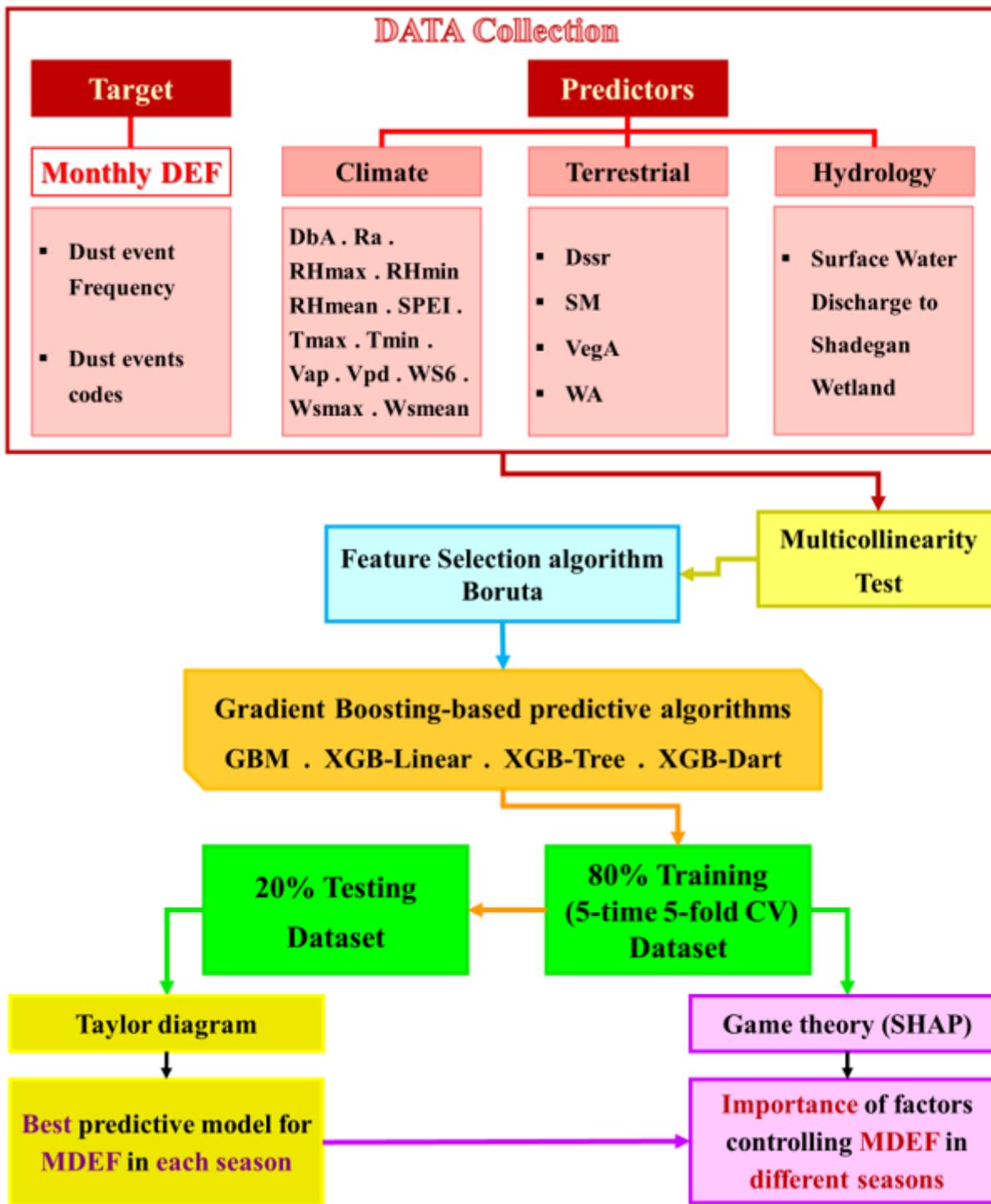


Figure 2

The methodology used for predicting the monthly dust events frequency and determining its controlling factors around the Shadegan wetland.

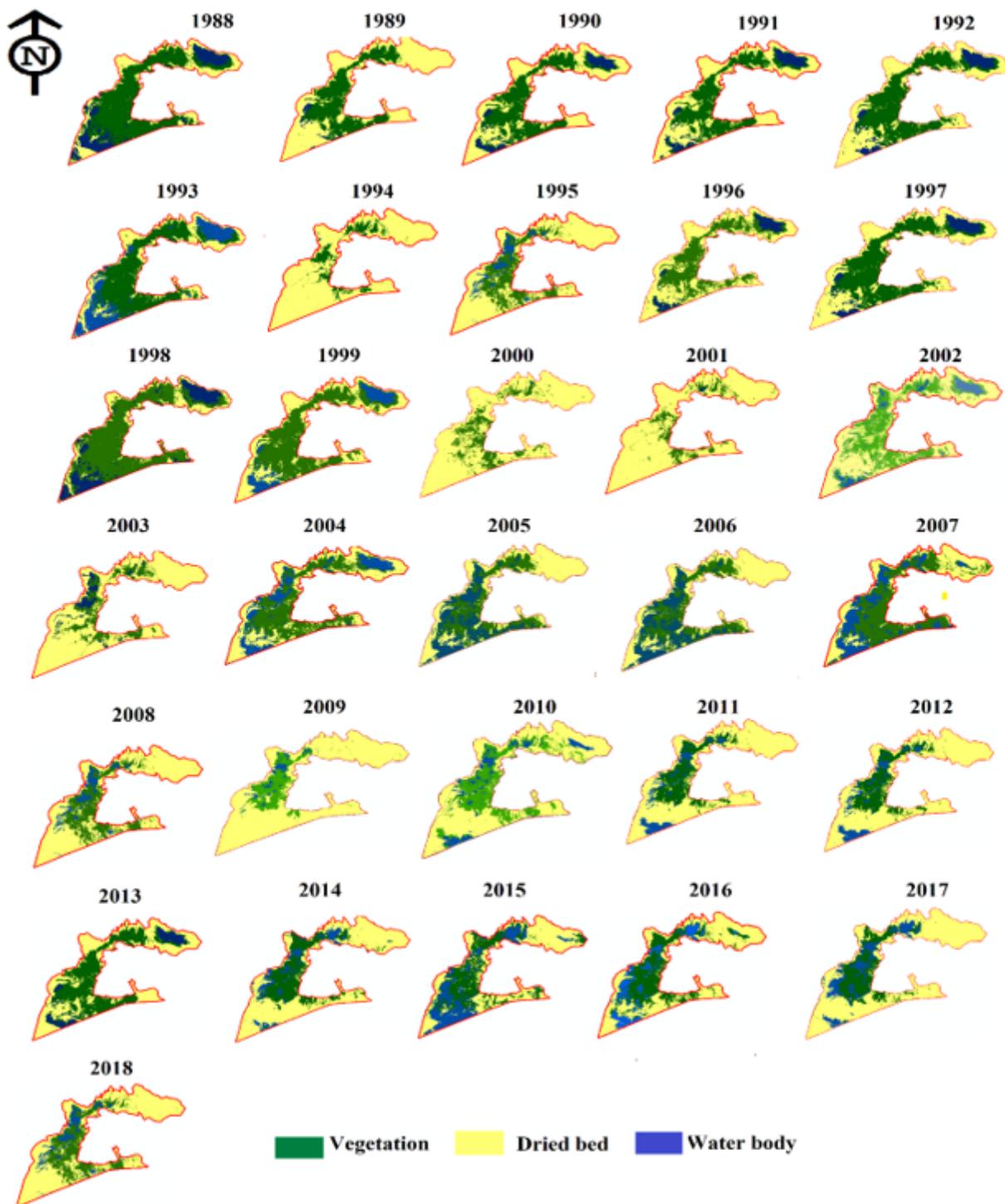


Figure 3

Annual changes in water body, vegetation cover and dried bed of Shadegan wetland during 1988-2018. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

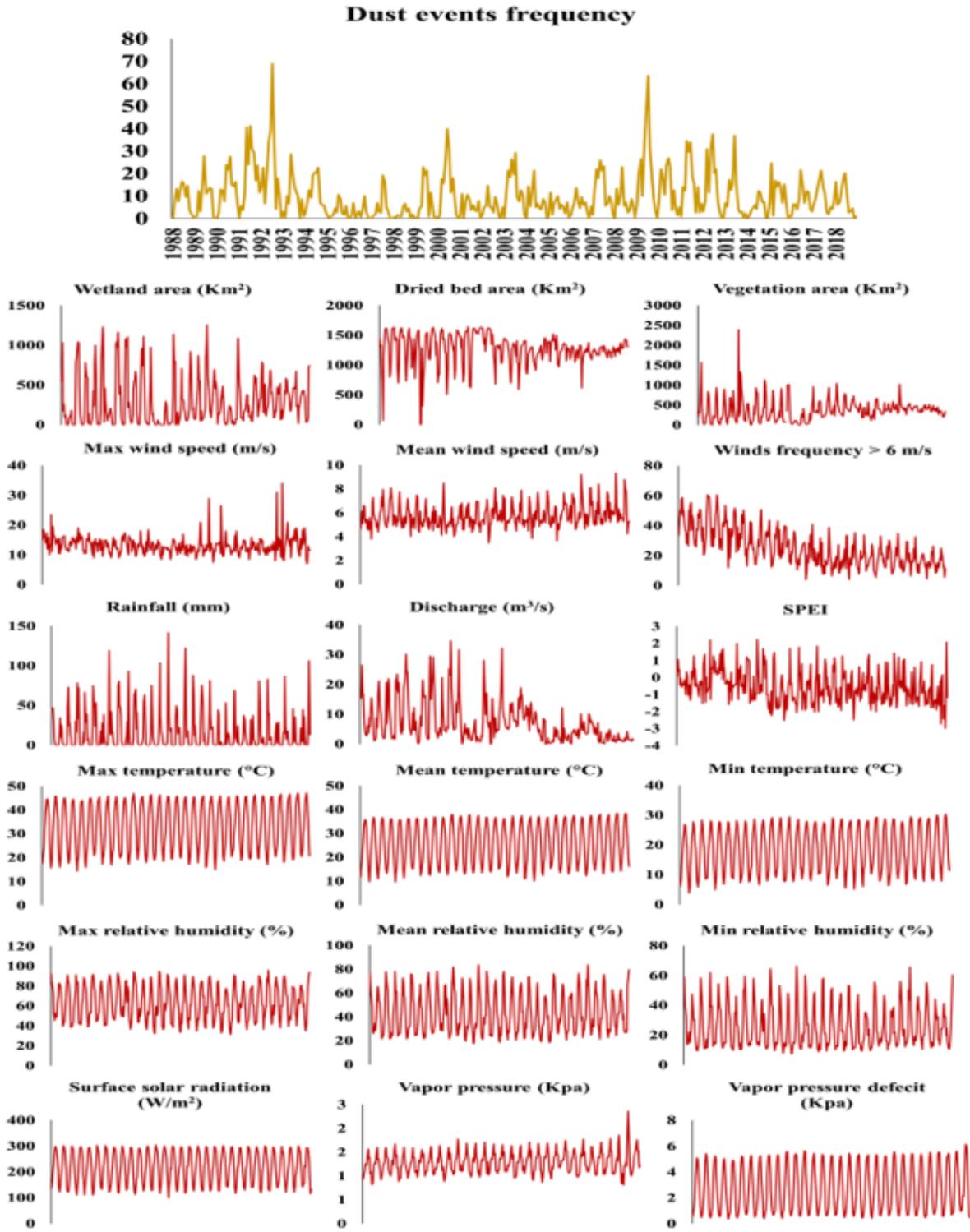


Figure 4

Mean monthly variations of dust events frequency and predictive variables in the study area from 1988 to 2018

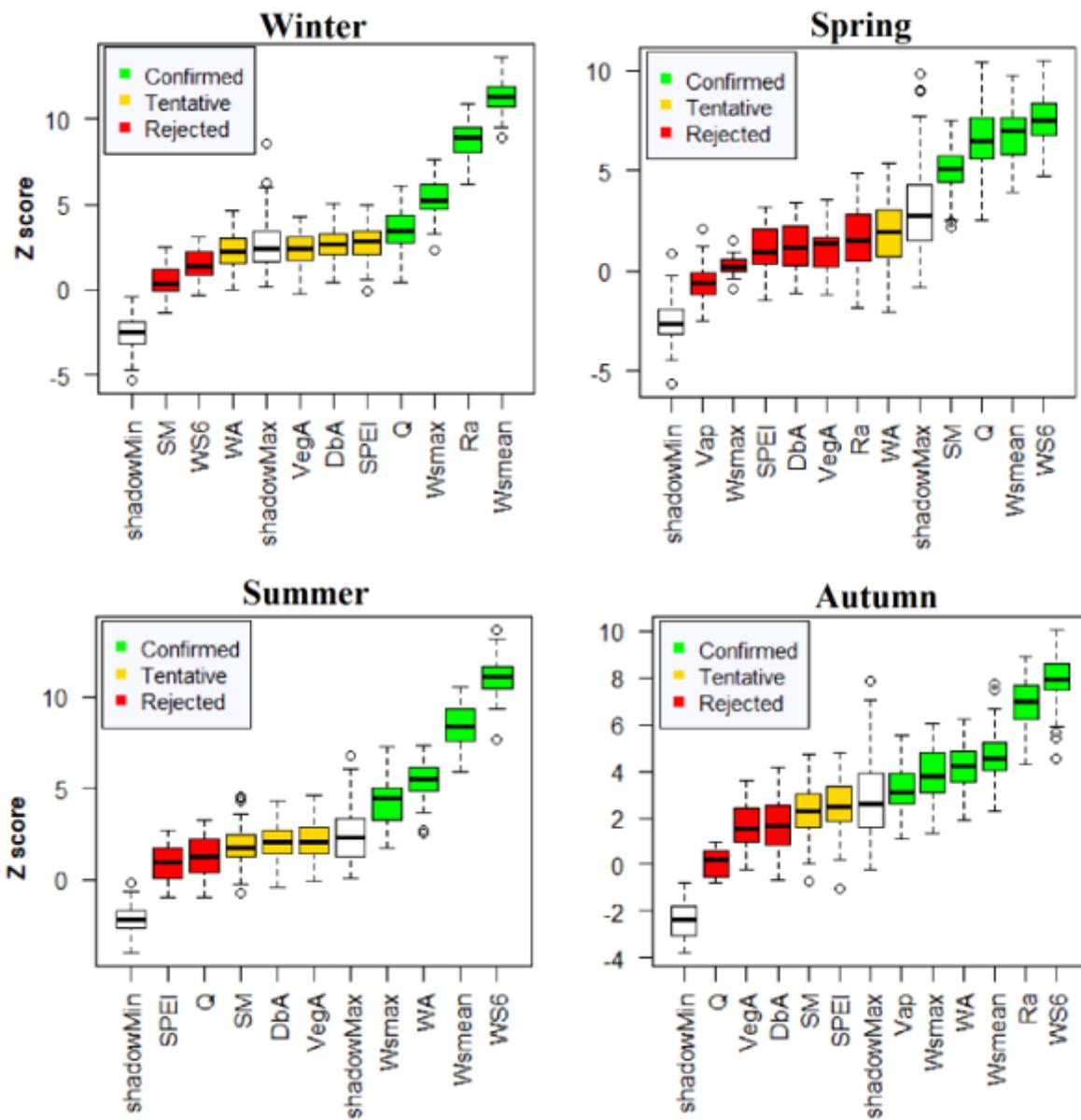


Figure 5

Importance score of explanatory variables in different seasons based on Boruta algorithm.

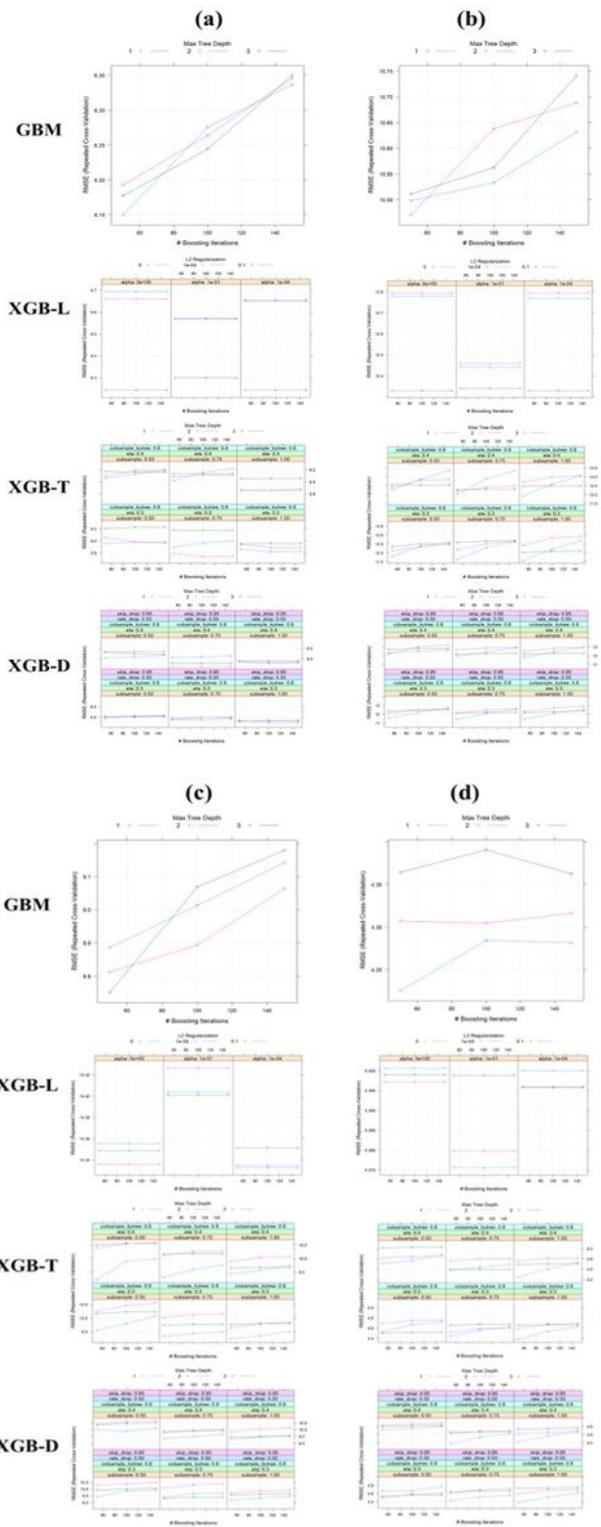


Figure 6

Hyper-parameters values for models tuned by five-fold cross-validation in the months of the (a) winter, and (b) spring seasons. Hyper-parameters values for models tuned by five-fold cross-validation in the months of (c) summer, and (d) autumn seasons.

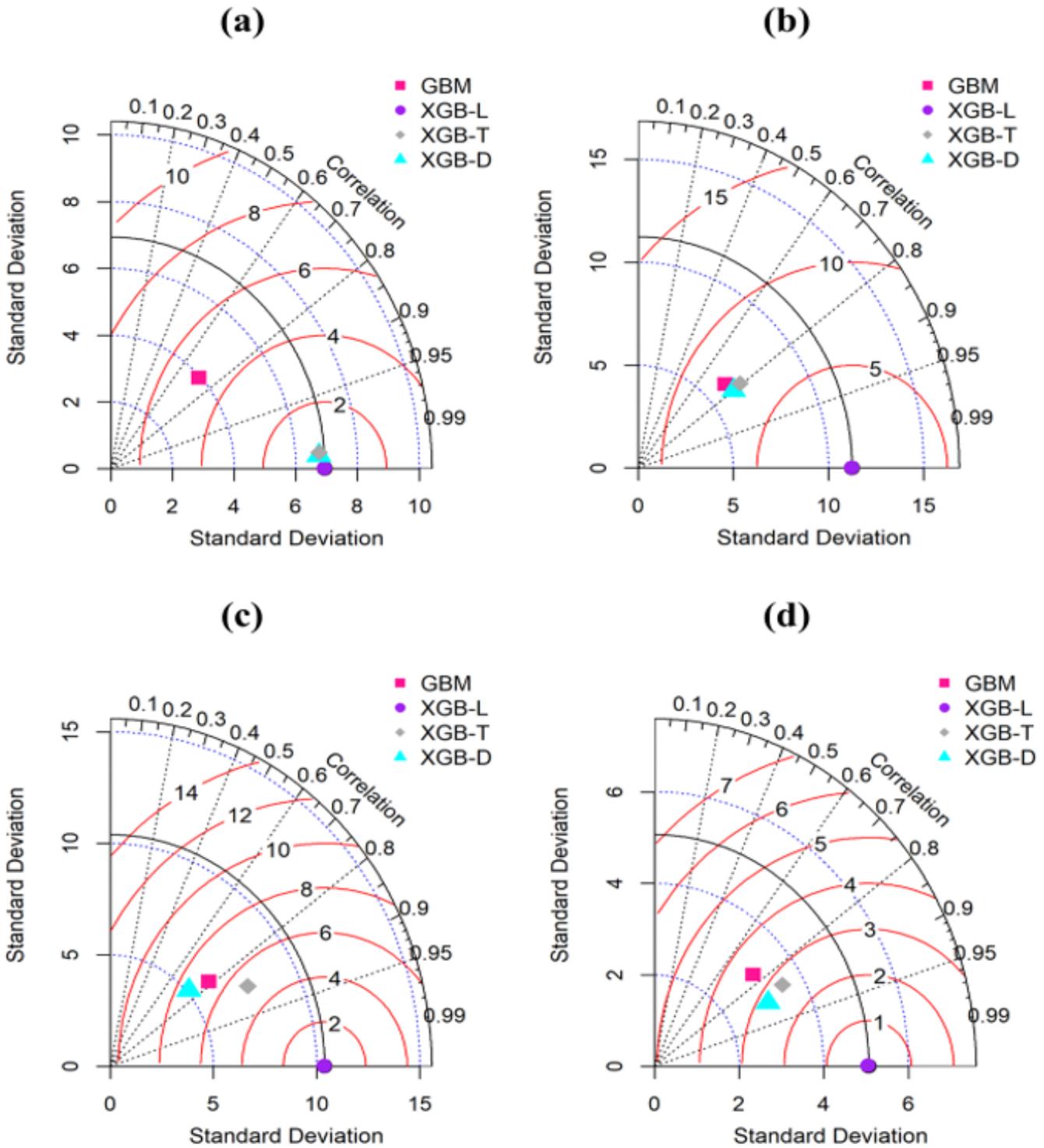


Figure 7

Performance of study models for predicting monthly dust events frequency in (a) winter, (b) spring, (c) summer, and (d) autumn around the Shadegan wetland according to the training data sets.

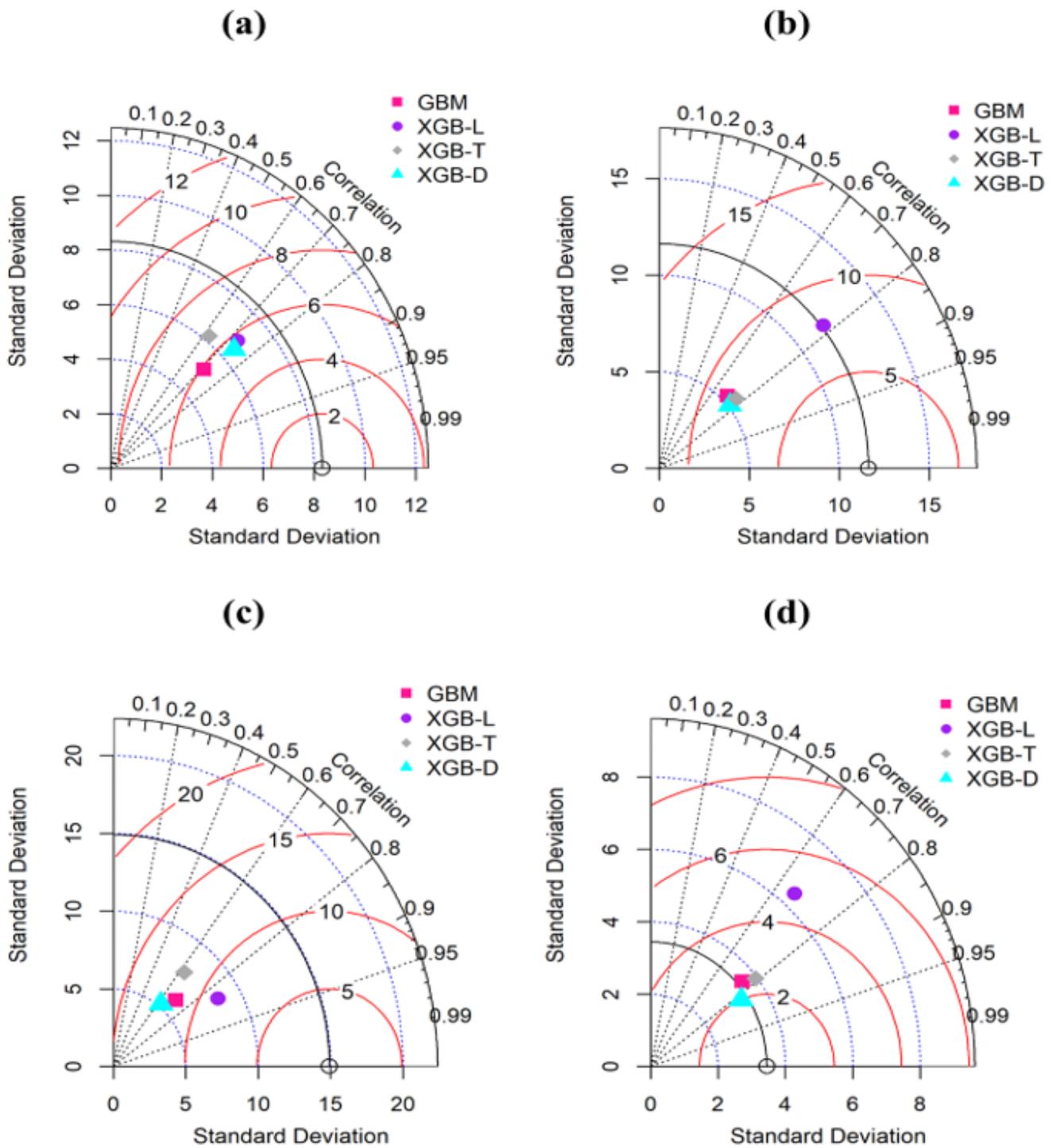


Figure 8

Performance of study models for predicting monthly dust events frequency in (a) winter, (b) spring, (c) summer, and (d) autumn around the Shadegan wetland according to the testing data sets.

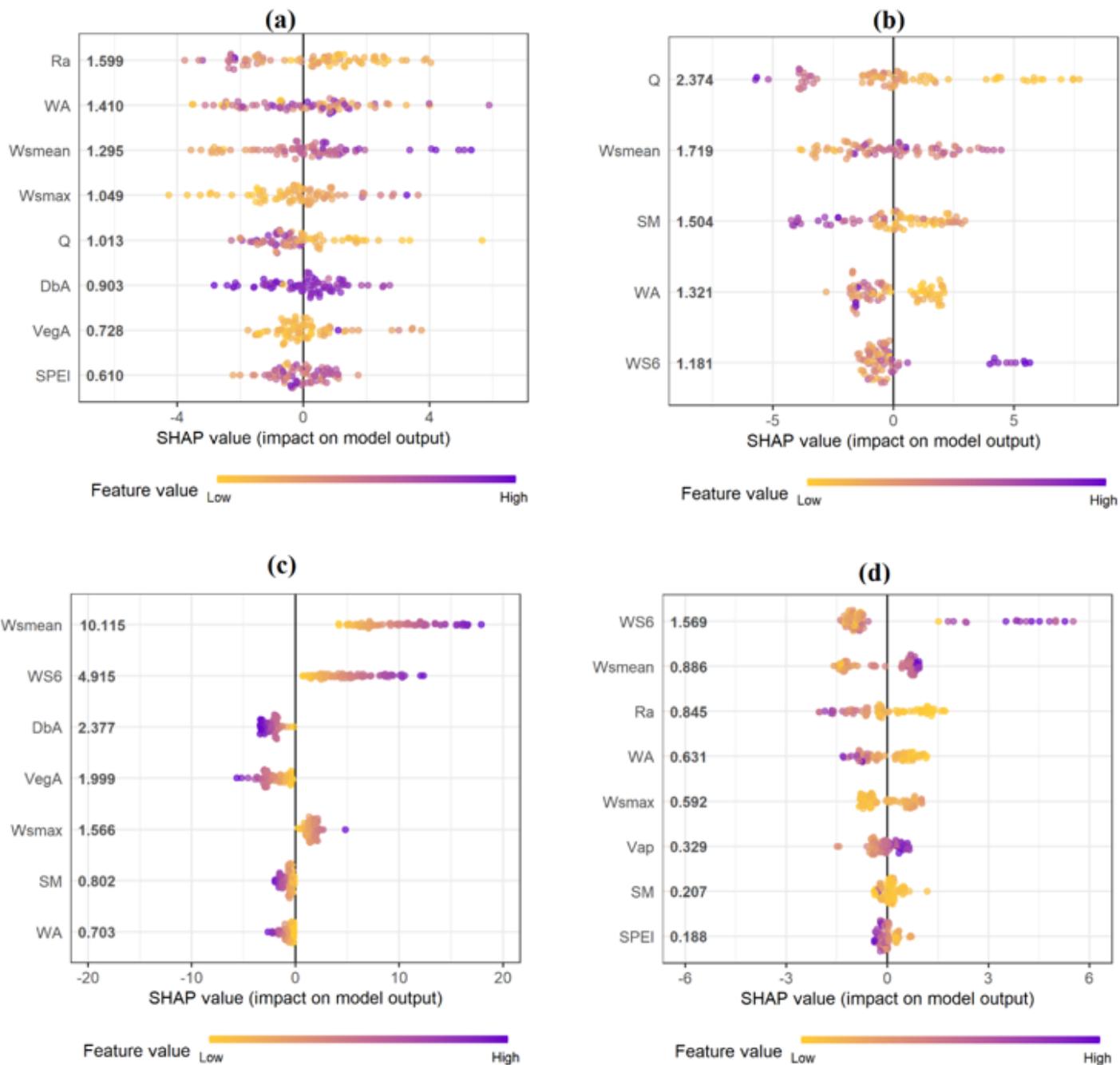


Figure 9

Prioritization of controlling factors the monthly frequency of dust events around Shadegan international wetland in (a) winter, (b) spring, (c) summer, and (4) autumn based on SHAP values.