



Land degradation risk mapping using topographic, human-induced, and geo-environmental variables and machine learning algorithms, for the Pole-Doab watershed, Iran

Ali Torabi Haghighi¹ · Hamid Darabi¹ · Zahra Karimidastenaie¹ · Ali Akbar Davudirad² · Sajad Rouzbeh³ · Omid Rahmati⁴ · Farzaneh Sajedi-Hosseini⁵ · Björn Klöve¹

Received: 20 April 2020 / Accepted: 23 November 2020 / Published online: 29 December 2020

© The Author(s) 2020

Abstract

Land degradation (LD) is a complex process affected by both anthropogenic and natural driving variables, and its prevention has become an essential task globally. The aim of the present study was to develop a new quantitative LD mapping approach using machine learning techniques, benchmark models, and human-induced and socio-environmental variables. We employed four machine learning algorithms [Support Vector Machine (SVM), Multivariate Adaptive Regression Splines (MARS), Generalized Linear Model (GLM), and Dragonfly Algorithm (DA)] for LD risk mapping, based on topographic ($n=7$), human-induced ($n=5$), and geo-environmental ($n=6$) variables, and field measurements of degradation in the Pole-Doab watershed, Iran. We assessed the performance of different algorithms using receiver operating characteristic, Kappa index, and Taylor diagram. The results revealed that the main topographic, geo-environmental, and human-induced variable was slope, geology, and land use change, respectively. Assessments of model performance indicated that DA had the highest accuracy and efficiency, with the greatest learning and prediction power in LD risk mapping. In LD risk maps produced using SVM, GLM, MARS, and DA, 19.16%, 19.29%, 21.76%, and 22.40%, respectively, of total area in the Pole-Doab watershed had a very high degradation risk. The results of this study demonstrate that in LD risk mapping for a region, topographic, and geological factors (static conditions) and human activities (dynamic conditions, e.g., residential and industrial area expansion) should be considered together, for best protection at watershed scale. These findings can help policymakers prioritize land and water conservation efforts.

Keywords Pole-Doab watershed · Dragonfly Algorithm · ROC–AUC · Kappa index · Taylor diagram

✉ Ali Torabi Haghighi
ali.torabihaghighi@oulu.fi

¹ Water, Energy and Environmental Engineering Research Unit, University of Oulu, P.O. Box 4300, 90014 Oulu, Finland

² Agricultural Research, Education and Extension Organization (AREEO), Agricultural and Natural Resources Research and Education Center of Markazi Province, P. O. Box: 38188-9-3811, Arak, Iran

³ Department of Watershed Management, Sari Agriculture Science and Natural Resources University, P.O. Box 737, Sari, Iran

⁴ Soil Conservation and Watershed Management Research Department, Kurdistan Agricultural and Natural Resources Research and Education Center, AREEO, Sanandaj, Iran

⁵ Department of Reclamation of Arid and Mountainous Regions, Faculty of Natural Resources, University of Tehran, Karaj, Iran

Introduction

Land degradation (LD) is now a critical environmental issue worldwide, posing a threat to food security and socio-economic development, and the problem will worsen without rapid remedial action (Jiang et al. 2019; Shao et al. 2020). Land degradation, defined as declining capability of the biological or economic productivity of land to provide ecosystem services, is closely connected to food security, human well-being, and development (Wieland et al. 2019; Crossland et al. 2018; Gichenje et al. 2019). It is caused by a combination of direct factors (land use/land cover changes (LULCC), climate change) and indirect factors (population pressure, socioeconomic, and social–ecological conditions, interactions between humans and nature, land management policy), and can vary in severity over time and with location (Riva et al. 2017; Okpara et al. 2018; Gichenje et al. 2019).

As a result of human activities, such as land use change, LD can alter hydrological conditions that are crucial for water resources and sustainable river basin management (Aladejana et al. 2018; Jiang et al. 2019; Haghghi et al. 2020). Therefore, efforts to prevent land degradation must be taken by agencies and governments worldwide (Keesstra et al. 2016; Solomun et al. 2018). To assess LD, it is necessary to consider both natural and human-induced factors, e.g., climate change, urbanization, and rising demand for food and fuel (AbdelRahman et al. 2018; Wunder and Bodle 2019; Liniger et al. 2019). Owing to major concerns about conserving land for ecosystem services and the impact of LD on societies and the environment, soil, and water protection has become an important issue for international organizations working with sustainable development (Solomun et al. 2018; Djanibekov et al. 2018). Identifying the causes of LD is essential for its prevention. Globally, LULCC (decline in rangeland area and conversion to farmland with low productivity) is recognized as a major driver of LD (Krkoška Lorencová et al. 2016). An increasing proportion of land with low productivity and a lack of financial resources for land managers in developing countries are exacerbating the risk of LD and lowering resilience within rangeland landscapes (Darabi et al. 2018; Pirnia et al. 2018; Cowie et al. 2019; Pirnia et al. 2019).

To our knowledge, most previous studies assessing LD conditions have used geographic information system (GIS) and remote sensing techniques in spatial assessments of LD risk based on the environmental conditioning variables (Právělie et al. 2017; Mariano et al. 2018; Cerretelli et al. 2018). Spatiotemporal patterns of land use change (anthropogenic factors) are the main factor in land degradation (Bewket and Sterk 2005; Gebremicael et al. 2018). Other researchers have reported that direct anthropogenic disturbances in environments and ecosystems can increase land degradation (Ahiablame and Shakya 2016; Davudirad et al. 2016; Aladejana et al. 2018; Schwieger and Mbidzo 2020; Shao et al. 2020). Jaquet et al. (2015) found that outmigration has led to land degradation in a western Nepal watershed. Wei et al. (2020) examined the impacts of land degradation on lake and reservoir water quality and showed a clear trend for degradation, with significant adverse impacts on lake/reservoir water quality. Yatheendradas et al. (2008) concluded that land degradation is the result of dynamic and complex interactions between LULCC, climate variables, and hydrological processes in a watershed.

The environmental problems associated with LD are particularly severe in dryland regions, which poses a threat to many people, especially in developing countries, such as Iran (Khosravi et al. 2015; Darabi et al. 2018). During the recent decades, land degradation in Iran (e.g., soil erosion, such as gully development) has accelerated in Iran due to many factors, such as increasing population, socio-economic

development, LULCC (demand for agricultural products has resulted in large-scale conversion of rangeland and forest to cropland), over-exploitation of water resources, geology and topography, and climate change (Pour et al. 2009; Seraji et al. 2009; Davudirad et al. 2016; Bakhshandeh et al. 2019).

Owing to the many interacting factors causing LD, machine learning techniques could be useful in LD risk mapping. In this study, we applied four novel machine learning algorithms, namely Support Vector Machine (SVM), Multivariate Adaptive Regression Splines (MARS), Generalized Linear Model (GLM), and Dragonfly Algorithm (DA). These have already been successfully applied in other fields, e.g., in flood risk and hazard mapping, fog-water harvesting, agricultural drought assessment, and groundwater risk assessment (Zhao et al. 2019; Darabi et al. 2020; Karimidashtenaee et al. 2020; Rahmati et al. 2020; Choubin et al. 2020).

Many studies have pointed out that knowledge about LD conditions, especially in arid and semi-arid regions with rapid industrialization and urbanization, is important for achieving the global aim of sustainable development in the long term (Gu et al. 2016; Tripathi et al. 2017; Cao et al. 2018; Van Haren et al. 2019; Giuliani et al. 2020). Hence, the aim of the present study was to develop a new quantitative LD mapping approach using machine learning techniques, benchmark models, and selected socio-environmental conditioning variables. Different types of data and information were used with the four different machine learning algorithms to develop distributed maps of LD risk for the case of a watershed in Iran. The novelty of the study lies in (1) comparing conventional algorithms (support vector machine (SVM), multivariate adaptive regression spline (MARS), and generalized linear model (GLM)) with new algorithms, including DA, for LD mapping applications; (2) developing a spatial framework for LD mapping by applying new conditioning factors; (3) considering and introducing important socio-environmental variables in land degradation; and (4) evaluating socio-environmental conditioning variables for creating useful LD maps based on the model results.

Materials and methods

Study area

The Pole-Doab watershed (49° 04' 15"–49° 52' 12" E, 33° 44' 42"–34° 12' 13" N) covers an area of 1740 km² in central Iran (Fig. 1). It lies within a semi-arid-moderate to semi-arid-cold region based on the Domartan climate index, with maximum temperature in July (42 °C) and minimum temperature in January (− 25.7 °C). The precipitation regime is rainfall–snow, with a mean annual total (1988–2017) of 430 mm, which mainly falls during November, December,

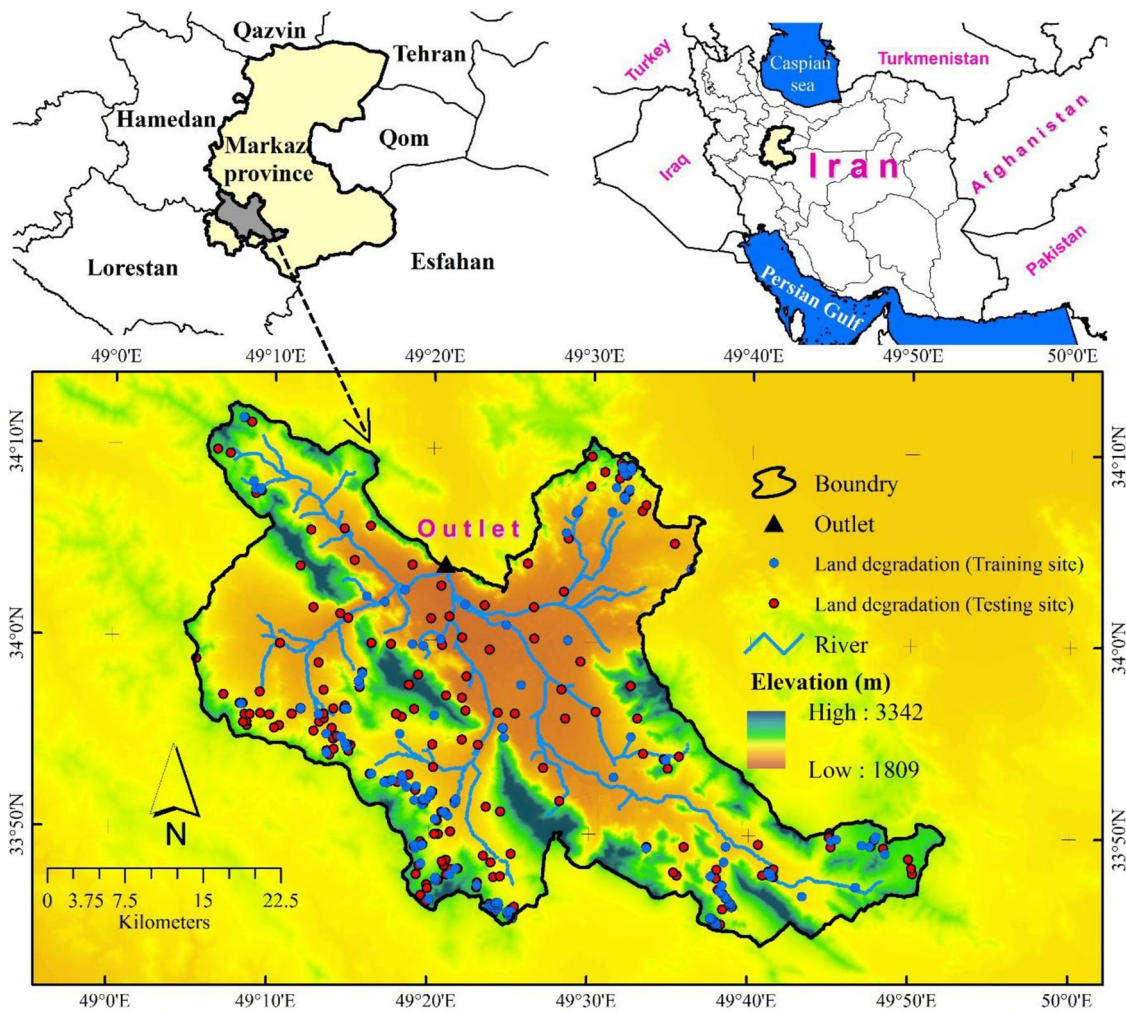


Fig. 1 Location of the study area, the Pole-Doab watershed in central Iran

and May. The topography of the Pole-Doab watershed consists of rugged and mountainous terrain surrounding plains, with elevation varying from around 1809 m above sea level (asl) on the plains to 3342 m asl in the mountains. This complex topography and steep gradients create a high risk of LD, particularly when combined with human-induced activities such as LULCC, urbanization, and industrialization in the watershed (Davudirad et al. 2016). The Pole-Doab watershed is one of the main sub-basins in headwaters of the Qareh–Chai river basin, which has been regulated by the Saveh reservoir since 1995. The Shazand plain, located in the center of the watershed, is used intensively for agriculture (Davudirad et al. 2016). In addition, considerable recent development of infrastructure, industries, and urban areas has altered lifestyles significantly. These rapid LULCC (increasing agriculture, urban expansion, industrial development) have led to extensive land degradation (Davudirad et al. 2016; Sadeghi et al. 2019; Hazbavi et al. 2020).

Methods

Field measurements of land degradation

Several processes associated with LD, including water and wind erosion and soil fertility decline, were considered in LD risk mapping. Information on these processes in the Pole-Doab watershed was extracted from an inventory of LD sites in the region, based on field surveys and some documents from the Forest, Range, and Watershed Management Organization of Markazi Province, Iran. The LD sites, which represented different types of degradation (e.g., gully erosion, riverside erosion, surface erosion, and mining), were plotted in an LD inventory map (Fig. 2). In order to prepare an urban LD risk map, degraded areas and non-degraded areas were allocated a value of 1 and 0, respectively. Hence, the historical occurrence of LD was a source of essential information. In field surveys in the Pole-Doab watershed, 200 degraded sites (value = 1) and 200 non-degraded sites (value = 0) were chosen randomly



Fig. 2 Examples of land degradation in the Pole-Doab watershed: Riverside erosion (a, d, n), stream erosion (b, c, k, l), gully erosion (e, m, g), mining (f, i), badlands (j), and pollution and industrial causes (k)

for the analysis. For the purposes of the present study, the LD inventory map was randomly divided into two groups, training/learning and testing/validation datasets. The training dataset, which comprised 70% of the LD (140 points), was used for training/learning of the machine learning algorithms. The validation dataset, which comprises 30% the LD inventory

(60 points), was used for validation of the models. Non-land degraded locations were selected randomly at a distance from the land degraded areas, as suggested in the literature (Hong et al. 2018; Rahmati et al. 2020; Darabi et al. 2020). Therefore, 200 non-land degraded locations were selected, with 70% of the non-land degraded inventory (140 points) used for model

training and 30% (60 points) to validate the machine learning algorithms.

Dragonfly algorithm (DA)

A number of intelligence algorithms have been developed in the recent years and these have enormous potential to solve non-linear problems. Intelligence algorithms perform intelligent behavior by collecting conditioning factors to solve problems. The Dragonfly Algorithm (DA), one of the pioneer intelligence algorithms, has been extensively studied in the recent years (Mirjalili 2016; KS and Murugan 2017; Jafari and Chaleshtari 2017; Díaz-Cortés et al. 2018; Shilaja and Arunprasath 2019; Li et al. 2020). It is a meta-heuristic optimization algorithm that was developed using the particle swarm optimization technique with distinctive and extraordinary swarming behavior, which is intended to represent a tiny predator in nature, because of its simple and easy implementation. The main inspiration and purpose of the DA is to hunt and migrate through static and dynamic swarming, based on the unique and superior swarming behavior of dragonflies (KS and Murugan 2017; Shilaja and Arunprasath 2019). The DA starts the optimization procedure by generating a set of random solutions for a specific problem. The situation and stage vectors of dragonflies are bootied by random values defined within the minimum and maximum values of the variables (Mirjalili 2016). In this study, the DA was used as an artificial intelligence algorithm to prepare a LD risk map based on socio-environmental conditioning variables. DA can be described by the expression:

$$X_i = (X_i^1, X_i^d, \dots, X_i^N) \tag{1}$$

where N is size of the population of dragonflies, $i = 1, 2, 3, \dots, N$, and X_i^d refers to the position of the i th dragonfly in d th dimension of the search space.

Based on the initial position values (randomly produced between the lower and upper limits of the variables), the fitness function is evaluated. For updating the velocity and position of the separation, alignment, cohesion, food, and enemy coefficients are calculated as follows:

$$S_i = - \sum_{j=1}^N X - X_i \tag{2}$$

$$A_i = \frac{\sum_{i=1}^N V_i}{N} \tag{3}$$

$$C_i = \frac{\sum_{i=1}^N X_i}{N} - X \tag{4}$$

$$\begin{aligned} F_i &= X^+ - X \\ E_i &= X^- - X \end{aligned} \tag{5}$$

where $S_i, A_i, C_i, F_i,$ and E_i are the weights for separation, alignment, cohesion, food, and enemy factors for each dragonfly; V_i and X_i refer to the velocity and position of the i th individual; X refers to the position of the current individual; and N indicates the number of individuals (KS and Murugan 2017; Rahman and Rashid 2019; Debnath et al. 2020).

Support vector machine (SVM)

Support vector machine (SVM) model is a classification/regression method with a set of linear indicator functions based on non-parametric statistical learning theory (Mountrakis et al. 2011). It specifies the boundary of classes by an optimization algorithm (Sajedi-Hosseini et al. 2018). The particular attributes of decision level in SVM enable high extension capability of the learning machine, which makes it effective in handling non-separable training datasets (Drucker et al. 1996). The main difficulty in SVM modeling lies in selecting important modeling variables. Transformation of data in SVM is carried out using kernel mathematical functions, and there are numerous standard transformations which can be applied for specific purposes. The SVM kernel functions were used here to transform data into two classes, consisting of land-degraded and non-land-degraded locations (0, 1). The ability of SVM is reliant on choosing suitable kernel functions (e.g., sigmoid kernel, radial basis function (RBF), linear kernel, polynomial kernel). According to the previous studies (Tien Bui et al. 2012; Hong et al. 2018; Choubin et al. 2019), RBF provides the most accurate results. It was therefore used in the present study in R software ('e1071' package) (Meyer et al. 2019). The RBF kernel is commonly used in SVM classification in various kernelized learning algorithms, and is defined as (Vert et al. 2004; Cura 2020):

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2 \times \sigma^2}\right) \tag{6}$$

where x_i and x_j are two features for the RBF kernel ($K(x_i, x_j)$); $x_i - x_j$ is Euclidean distance between two features; and σ is a free parameter. The RBF kernel value decreases with distance and ranges between 0 and 1 ($x = x'$).

Multivariate adaptive regression splines (MARS)

The Multivariate Adaptive Regression Splines (MARS) approach is an adaptive modeling process of machine learning techniques that can be used for identifying relationships

between a set of independent variables and response variables with high-dimensional data (Friedman 1991). In MARS, relationship modeling between a response variable and independent conditioning factors is performed with simple functions (Darabi et al. 2020). In essence, it is a local regression procedure that utilizes a collection of foundation functions to model non-linear complex communications. The prognosticator space is splined into multiple overlapping places, called spline functions, which are appropriate. The MARS model uses the following equation (Xu et al. 2010; Zhang and Goh 2016; Serrano et al. 2020):

$$\hat{f}(x) = \sum_{i=1}^k c_i \times B_i(x), \quad (7)$$

where $B_i(x)$ shows the base function of the MARS model (which may be a sole spline function or a yield (interaction) of more than two spline functions); c_i is a constant coefficient; and i is the size of the base function contained in the model. The base function is defined as:

$$B_i(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}. \quad (8)$$

For each dataset in MARS, m explanatory and n individual variables are defined ($n \times m$ basic functions). To obtain and prune the definitive model in MARS, a progressive selection of basic functions is used, which leads to a much overfitted model. In the present study, the method was built in R software, using the “earth” package.

Generalized linear model (GLM)

Generalized linear model (GLM), an extension of the predictable linear regression model, was formulated by Nelder and Wedderburn (1972) to produce answers based on the Maximum Likelihood (ML) of the training variables. The GLM allows the dataset to be overfitted by exponential distribution (normal, binomial, or gamma distribution) (Nordin et al. 2020). Regression methods, including linear, logistic, and log-linear regression, have been widely used to obtain the best model to illustrate the communication between a dependent parameter and multiple independent parameters (Ozdemir and Altural 2013; Karimidastenaie et al. 2020). The GLM approach can be used to process data of different types, such as normal data, Bernoulli success/failure data, Poisson count data, and others (McCullagh and Nelder 1989). A detailed description of the GLM model is presented by Breslow (1996). In a GLM, each dependent variable (here Y) is assumed to be created from a distribution in an exponential family. The mean of the distribution (μ) depends on the independent variables (X). In the GLM, the linear predictor is given as (Nordin et al. 2020):

$$E(Y) = \mu \times g^{-1}(X\beta), \quad (9)$$

where $E(Y)$, $X\beta$, and g are the value of Y , linear predictor, and the link function, respectively. In this context, the variance (V) is typically a function of μ :

$$\text{Var}(Y) = V(\mu) = V(g^{-1}(X\beta)). \quad (10)$$

It is suitable if V tracks from an exponential distribution, but it may simplify matters if V is a function of the predicted value. The β parameter is naturally estimated with the maximum likelihood (ML) and maximum quasi-likelihood (MA-L), or Bayesian models. In this study, the GLM model was run in the R software environment.

Land degradation conditioning factors

There are many different types of LD worldwide and many different conditioning variables can be distinguished depending on the region and causes of LD. Thus in general, there is no universal definition of LD or of conditioning factors (Sklenicka 2016). In the present study, based on land degradation conditions in the Pole-Doab watershed, 18 biophysical conditioning variables were identified and categorized into three groups: topographic variables (elevation, slope, curvature, topographic wetness index, terrain ruggedness index, sky view factor, aspect); human-induced variables (land use, population density, population growth rate, residential and industrial expansion, distance to road); and geo-environmental variables (geology, soil type, precipitation, wind effect, distance to river, C-factor). The scale and resolution of these land degradation conditioning factors, classified into three groups, are presented in Table 1.

Topographic variables

Digital elevation model (DEM) We used a 30-m resolution digital elevation model (obtained from the Forest, Range, and Watershed Management Organization of Markazi province) which shows the 1809–3342 m asl altitude variation in the watershed (Fig. 3a).

Slope (%) We derived slope values from the 30-m DEM in ArcGIS 10.5 using the slope tool Spatial Analyst. The slope values in the watershed varied from 0% to more than 67.60% (Fig. 3b).

Curvature Curvature was derived from the DEM and categorized into three classes (Fig. 3c): concave (< -0.05 , upwardly concave surface), flat (-0.05 to 0.05), and convex (> 0.05 , upwardly convex surface) (Karimidastenaie et al. 2020; Tehrani et al. 2019).

Topographic wetness index (TWI) TWI, which indicates soil moisture content and spatial variability in surface saturation, was used to quantify local topographical impacts on LD

Table 1 Land degradation conditioning factors

Topographic factors	Scale	Spatial resolution (m)
Elevation	1:25,000	30
Slope	1:25,000	30
Aspect	1:25,000	30
Terrain ruggedness index	1:25,000	30
Topographic wetness index	1:25,000	30
Sky view factor	1:25,000	30
Curvature	1:25,000	30
Human-induced factors	Scale	Spatial resolution (m)
Distance to road	1:25,000	30
Land use	1:25,000	30
Residential and industrial area expansion	1:25,000	30
Population density	1:25,000	30
Population growth rate	1:25,000	30
Geo-environmental factors	Scale	Spatial resolution (m)
Geology	1:25,000	30
Land type	1:25,000	30
Distance to river	1:25,000	30
Precipitation	1:25,000	30
C-factor	1:25,000	30
Wind effect	1:25,000	30

conditions (Fig. 3d). It was calculated using ArcGIS 10.5 as (Zhu et al. 2018; Karimidastenaie et al. 2020):

$$TWI = Ln\left(\frac{A_s}{\tan \beta}\right), \tag{11}$$

where A_s is the local upslope drainage area for a certain grid cell and β is the local slope.

Terrain ruggedness index (TRI) TRI, which was developed by Riley et al. (1999), was calculated using SAGA GIS to explain the elevation difference between a given point (cell) and the mean of surrounding points (eight-cell matrix cells). TRI quantifies surface roughness by including maximum elevation values in the surroundings of a given point or cell in a DEM (Riley et al. 1999; Karimidastenaie et al. 2020). In the Pole-Doab watershed, TRI values varied from highly rugged (46.00) to completely level surface (0 m) (Fig. 3e).

Sky view factor (SVF) SVF is the visible sky in a hemisphere centered visible from the ground at a given point (cell in the raster map). It varies significantly with the topography of different regions and is used to account for obstruction of the overlying sky hemisphere by surrounding land surface

as an adjustment factor, with regions with lower visibility related to lower LD risk (Zakšek et al. 2011; Bernard et al. 2018). It is defined as:

$$SVF = \frac{1}{N} \times \sum_{i=1}^N [\cos \beta \times \cos^2 \beta \varphi_i + \sin \beta \times \cos (\theta_i - \alpha) \times (90 - \varphi_i - \sin \varphi_i \times \cos \varphi_i)], \tag{12}$$

where N is the number of directions, φ_i and θ are horizon angle and azimuth in the i th direction, respectively, around each cell in an elevation map, and α and β are the slope aspect and angle, respectively. In the present study, SVF was calculated using SAGA GIS, and the value for the study watershed varied from absolutely horizontal surface (= 1) to absolutely obstructed land surface (= 0) (Fig. 3f).

Aspect Aspect affects solar radiation received in a mountainous watershed and plays an important role in environmental changes. As the Pole-Doab watershed is located in the northern hemisphere, its north-facing slopes are less exposed to sunlight than south-facing slopes and thus have a higher moisture content, which influences the temperature

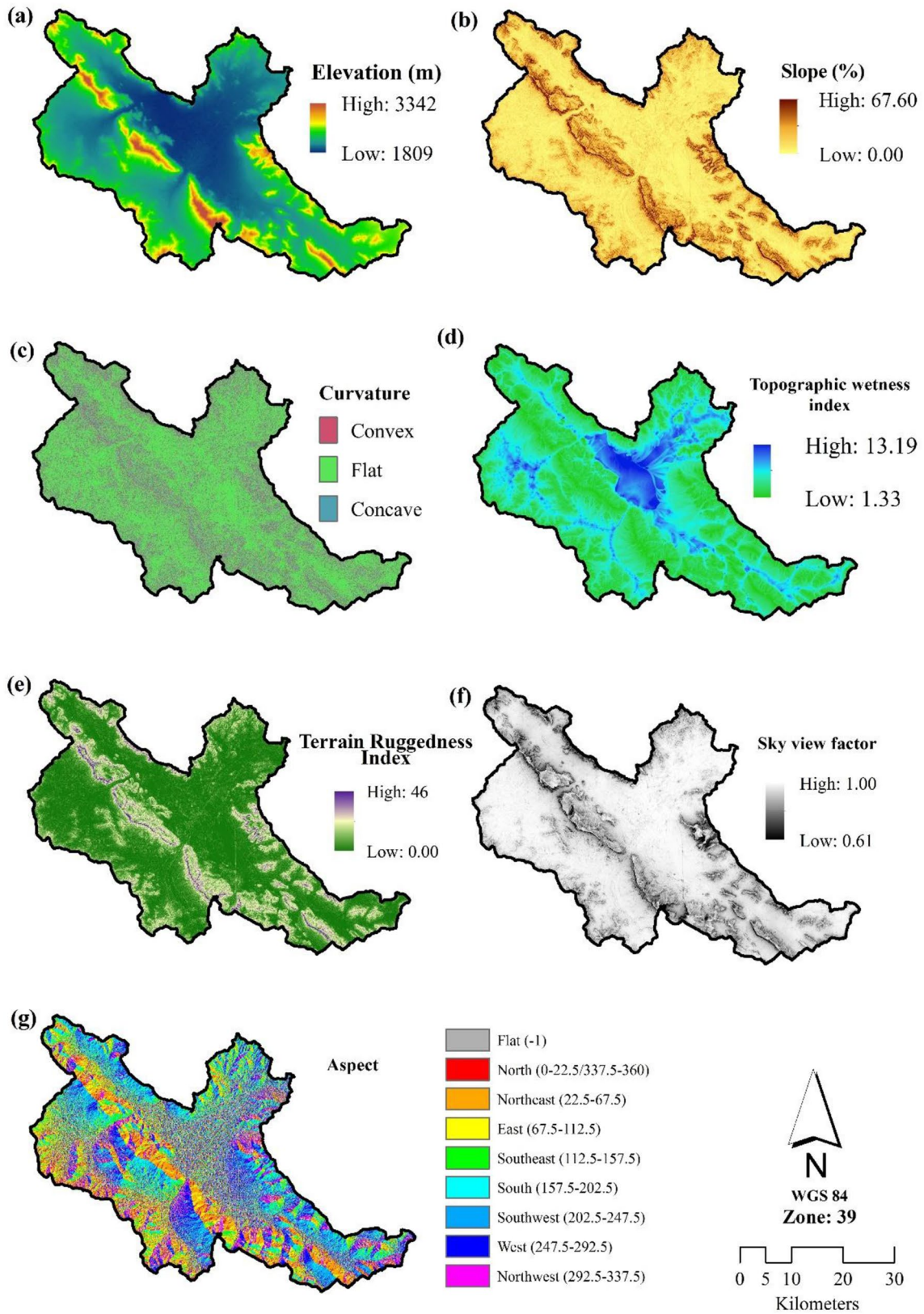


Fig. 3 Topographic variables used in land degradation risk mapping: **a** elevation, **b** slope aspect, **c** curvature, **d** topographic wetness index, **e** terrain ruggedness index, **f** sky view factor, and **g** aspect

gradient and surface warming and leads to differences in erosion pattern (Darabi et al. 2014, 2016) (Fig. 3g).

Human-induced variables

A number of human-induced conditioning factors of LD have been identified in previous studies (Huber-Sannwald et al. 2006; Lu et al. 2007; Právělie et al. 2017; Mekonnen et al. 2018; Speranza et al. 2019). We selected five of these for use in LD risk mapping in the study watershed.

Land use Land use information was prepared using Operational Land Imager (OLI) images from the Landsat 8 satellite, with path 165 and row 036-037. The images were acquired from the USGS dataset for 04 June 2019. In a pre-processing step, atmospheric correction of Landsat-OLI data was carried out using QUick Atmospheric Correction (QUAC) in ENVI 5.3 software. Using the maximum likelihood method (supervised classification), a land use map was then prepared in the ENVI 5.3 software (El-Khoury et al. 2015; Pullanikkatil et al. 2016; Torabi Haghghi et al. 2018). In the Pole-Doab watershed, there are seven land use types: Bare land, dry farming, irrigation farming, orchard, rangeland, residential, and rock zones, occupying an area of 59.86 km² (3.44%), 441.78 km² (25.39%), 205.49 km² (11.81%), 666.739 km² (38.32%), 89.85 km² (5.16%), and 140.80 km² (8.09%), respectively (Fig. 4a).

Population density The impact of population density on LD is unclear, but it is obvious that higher population density (population per unit area) would lead to more land degradation, with more serious degradation in areas with higher population density (Li et al. 2015). In this study, the impact of population density on LD risk in the Pole-Doab watershed was estimated based on human-induced changes in 10 counties within the watershed (Amiriyeh, Astaneh, Pole-Doab, Khorram dasht, Sadeh, Shamsabad, Gharehkahriz, Kazzaz, Koohsar, and Nahremian) (Fig. 4b).

Population growth rate Population growth rate is mainly responsible for population pressure on natural ecosystems (rangeland) and also conversion of rangeland to farmland and residential areas, which can affect flooding, sediment yield, and soil erosion, and consequently land degradation conditions. Population growth leads to increasing demand for housing and other facilities, which in turn leads to increased area of impervious surface as a result of urban development, infrastructure construction, and deforestation (Li et al. 2015). According to census data for Iran, the population growth rate in counties in the Pole-Doab watershed has increased rapidly in the recent decades (1976–2016) (Davudirad et al. 2016). We therefore assessed the impact of population pressure on LD risk in the Pole-Doab watershed by considering the population growth rate in the 10 counties in the watershed (Fig. 4c).

Residential and industrial area expansion Rapid urbanization and industrialization and conversion of neutral land to impervious land can affect LD conditions by increasing surface runoff and flooding conditions (Li et al. 2015). In this study, we used residential and industrial area expansion in the Pole-Doab watershed 1973–2016 (produced using TerrSet software) as a human-induced variable in LD risk mapping (Fig. 4d).

Distance to road Distance to road as impervious surface, and also as an indicator of development and infrastructure construction, is an important factor in LD risk mapping (Li et al. 2015). Here it was derived using the distance module in GIS 10.5 for each raster cell (Fig. 4e).

Geo-environmental variables

Geology The geology of a watershed can affect soil erosion and land degradation in two ways: (1) As an intrinsic effect related to the geological formation; and (2) as an effect of external and indirect factors such as climate (e.g., weathering). In this study, the geology of the watershed was divided into four formations: Quaternary, limestone, granite-granodiorite, and sandstone-shale (Fig. 5a).

Soil type Land type is typically defined by soil type and land form, which can affect soil erosion and land degradation (Nunes et al. 2011; Qiang et al. 2016). In this study, watershed soil types were divided into seven categories: alluvial fans, colluvium fans, hills, lowland, mountains, piedmont plains, and plateau and upper terraces (Fig. 5b).

Precipitation Annual precipitation data for 13 stations run by the Iranian Meteorological Organization (IRIMO) were used to produce a precipitation map for the Pole-Doab watershed. Analysis of the interpolation accuracy was carried out based on root mean square error (RMSE) in ArcGIS GIS 10.5, so the simple Kriging interpolation method was selected as it has the lowest RMSE (0.96) (Darabi et al. 2016). Mean annual precipitation varied from 461 mm in the west and southwest to 298 mm in the east and northeast of the study area (Fig. 5c).

Wind effect Land degradation by wind is one of the most serious environmental problems related to soil erosion, threatening environmental quality, ecosystem services, and land productivity (Chi et al. 2019). Wind effect assessments are relatively rare in the literature, due to poor data availability. Because the amount of evapotranspiration is greatly affected by high winds and high temperatures in summer (Fenta et al. 2020), wind effect was included as a biophysical variable for LD risk mapping in the present study (Fig. 5d). Information on wind effect in the study watershed was obtained based on the DEM in the SAGA GIS software.

Distance to river According to data on riverside and riverbed erosion obtained from local authorities and in field surveys, distance to river plays an important role in LD in the

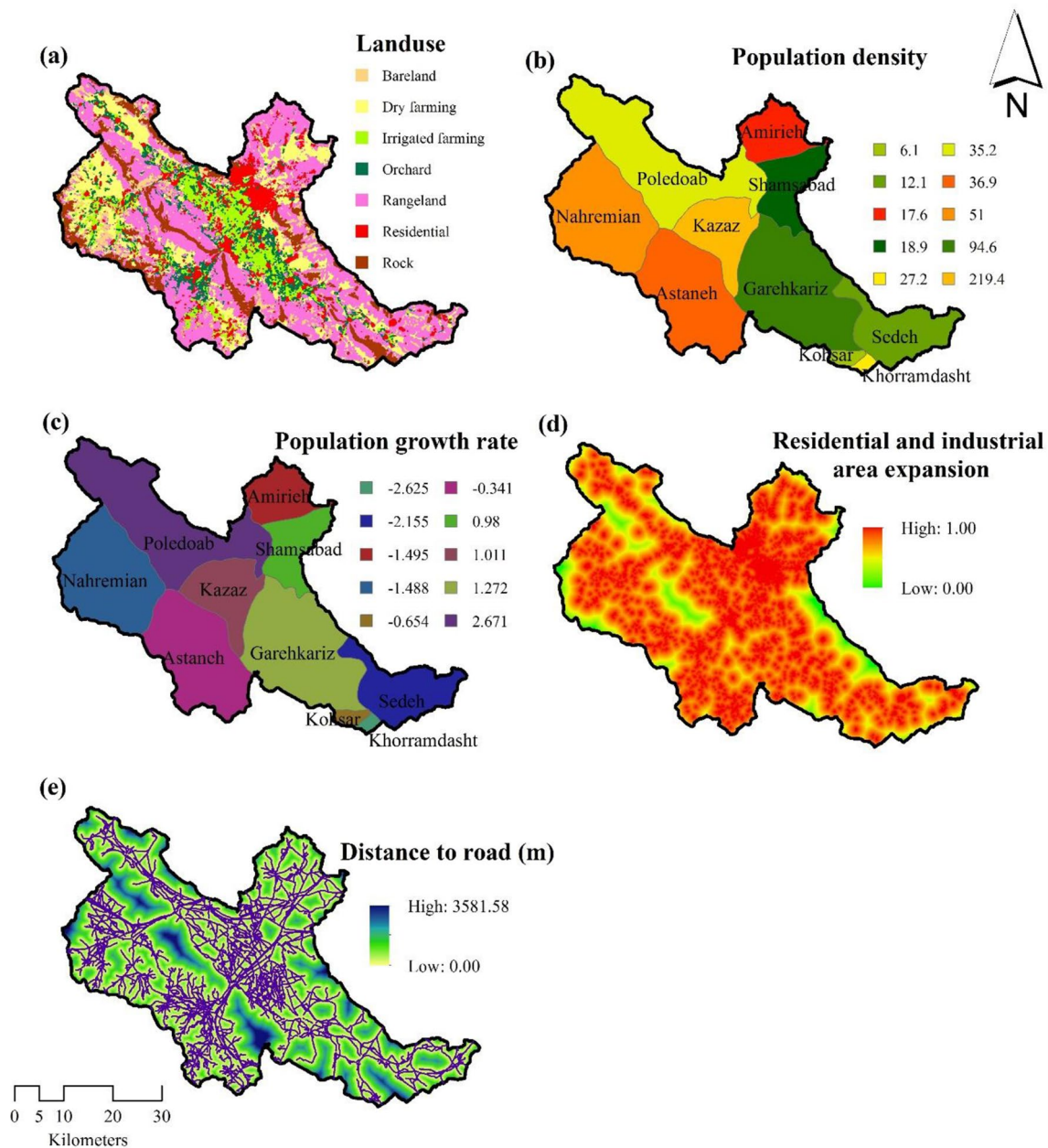


Fig. 4 Human-induced variables used in land degradation risk mapping: **a** land use, **b** population density in the 10 counties in the Pole-Doab watershed, **c** population growth rate in the different counties, **d** residential and industrial area expansion, and **e** distance to road

Pole-Doab watershed. The Euclidean distance to the river was calculated using the distance module in GIS 10.5 (Fig. 5e).

C-factor C-factor, a surface cover and roughness factor considered to show the effect of cropping and management practices on erosion conditions, is a critical indicator characterizing LD. C-factor mapping can provide suitable information for improving spatial and temporal modeling of land degradation and soil erosion. It is one of the most sensitive spatiotemporal factors, as it follows plant growth dynamics (Berendse et al. 2015; Vaverková et al. 2019). In this study,

C-factor used to consider the impact of soil and vegetation in LD risk mapping. It was derived using Landsat OLI (165-036 and 165-037) images for 04 June 2019 (Fig. 5f), which were obtained from the USGS website (Almagro et al. 2019). In a first step, Normalized Difference Vegetation Index (NDVI), which has a direct linear correlation to C-factor, was computed using Landsat data:

$$NDVI_{OLI} = \frac{\rho_{band5_{OLI}} - \rho_{band4_{OLI}}}{\rho_{band5_{OLI}} + \rho_{band4_{OLI}}} \tag{13}$$

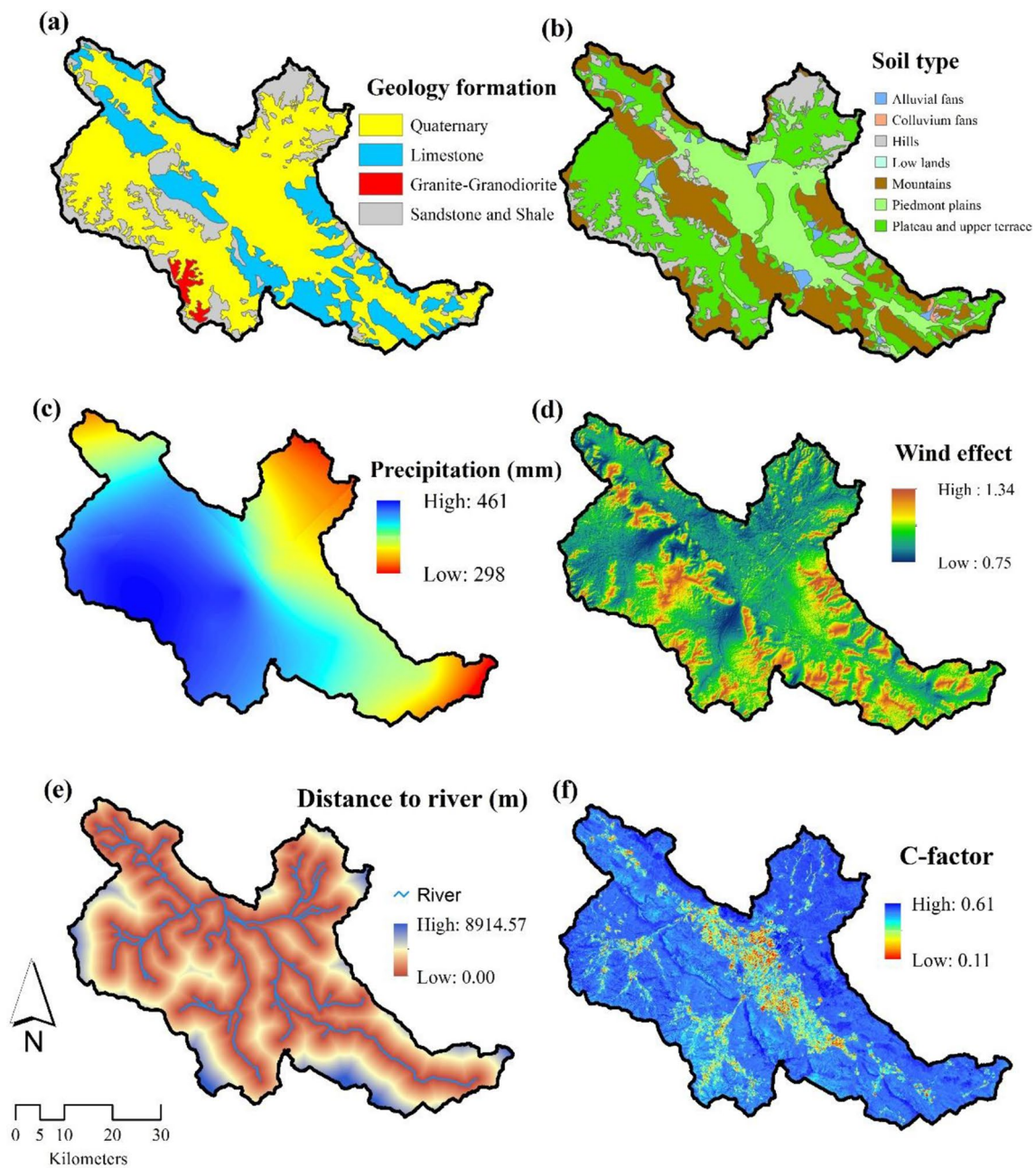


Fig. 5 Geo-environmental variables used in land degradation risk mapping: **a** geology, **b** soil type, **c** annual precipitation, **d** wind effect, **e** distance to river, and **f** C-factor

C-factor was then calculated as:

$$C = ((1 - NDVI)/2), \tag{14}$$

where ρ is the reflectance value of spectral bands for Landsat-OLI image: band 4_{OLI}: Red, band 5_{OLI}: NIR. C-factor varies in value from 0 to +1, representing good to bad conditions for soil erosion.

Calculation of land degradation index

Machine learning methods automate analytical model building, based on the idea that the model can learn from data, identify patterns, and make decisions (here prediction of LD index) with minimal human intervention. In this study, calculations of LD index were carried out using GIS layers (with ascii format), which were categorized into three groups

(topographic, human-induced, and geo-environmental variables), prepared in the same way in Arc map (with the same resolution, scale, and coordinate system), and considered as independent variables. Information related to LD (as point data) was considered as other input for the machine learning algorithms. Hence, after learning based on the above inputs, the models used in this study proceeded to predict the LD index as a final map with ascii format. All 18 conditioning variables, together with the 200 points selected as LD locations, were used in the R program to produce LD risk maps by the machine learning models. Using the natural break method (Tehrany et al. 2015; Choubin et al. 2019; Darabi et al. 2020) in ArcGIS 10.5, the LD risk was then classified into five classes: very low, low, moderate, high, and very high.

Model assessment

All machine learning models used in this study were assessed using the receiver-operator characteristic-area under the curve (ROC–AUC), which has been widely used for evaluating model performance (Frattini et al. 2010; Choubin et al. 2018; Darabi et al. 2020). The ROC–AUC value ranges from 0 to 1, with a value of 0.5–0.6, 0.6–0.7, 0.7–0.8, 0.8–0.9, and 0.9–1 indicating weak, average, good, very good, and excellent model performance, respectively (Choubin et al. 2018). The Kappa index, which employs model classification probabilities based on the null hypothesis to calculate the agreement by chance, was also used in model assessment. According to Monserud and Lee-mans (1992), the Kappa index is divided into five classes, with values of $k < 0.4$, $0.4 < k < 0.55$, $0.55 < k < 0.85$, $0.85 < k < 0.99$, and $0.99 < k < 1.00$ indicating poor, moderate, good, excellent, and perfect model performance, respectively. All model assessments were carried out in R software.

The importance of the 18 selected conditioning variables was evaluated for models showing high accuracy and precision. Visual assessment of model performance was carried out using a Taylor diagram (Taylor 2001) and three statistics: correlation coefficient, normalized standard deviation, and root mean square error (RMSE). In the Taylor diagram, models with high accuracy are close to the observations (Choubin et al. 2018).

Importance of variables

The importance of the conditioning variables was calculated from the results obtained through applying the selected model based on the ROC–AUC and Kappa index. The importance of independent variables (here topographic, human-induced, and geo-environmental variables) was calculated based on the frequency of dependent variables (here degraded locations) and spatial variation in the independent

variables, using the instructions of the selected model. In the selected model, importance of variables was considered as the reduction in node impurity weighted by the probability of reaching that node. The probability of the node was considered as the number of points influencing the node divided by the sum for the points (the more important the variable, the higher the value).

Results

Spatial distribution of land degradation

The spatial distribution maps of land degradation, obtained using the SVM, GLM, MARS, and DA algorithms, indicated that most parts of the Pole-Doab watershed were affected by LD, with high and low degradation conditions (Fig. 6a–d). All models showed the same overall spatial pattern, with high degradation in the south and southwest of the watershed. However, the spatial resolution at local scale derived from the different models varied. Regions with the highest (1.00) and lowest (0.00) risk of land degradation were successfully recognized by the DA, SVM, GLM, and MARS algorithms. In the spatial distribution of LD, the risk value ranged from 0.00 to 1.00. Using the natural break method in ArcGIS 10.5, the LD risk was divided into five classes: very low, low, moderate, high, and very high, the spatial distribution zones for which are presented in Fig. 6e–h. The LD risk maps obtained with all four algorithms indicated that the south of the Pole-Doab watershed is most exposed to degradation conditions. Based on degraded area obtained from SVM, GLM, MARS, and DA (Fig. 6e–h), the land area with a very high degradation risk represented 19.16%, 19.29%, 21.76%, and 22.40%, respectively, of the total area of the Pole-Doab watershed (Table 2). The LD risk maps also showed that most of the watershed was affected by some type of degradation, with more than 40% of the area falling into high and very high zones according to all algorithms. It is worth mentioning that some of the predictive variables used in the analysis may vary over time, leading to uncertainty in the results. Precipitation is one such variable, but since long-term precipitation data from 13 meteorological stations were used in the present analysis, the associated uncertainty was considered to be minimized.

Model performance

Validation is an important phase in evaluation of model accuracy. For quantitative comparison of the models, ROC–AUC and Kappa index were used. The maps obtained for LD risk were compared with the validation data, to assess the performance of each model. ROC–AUC

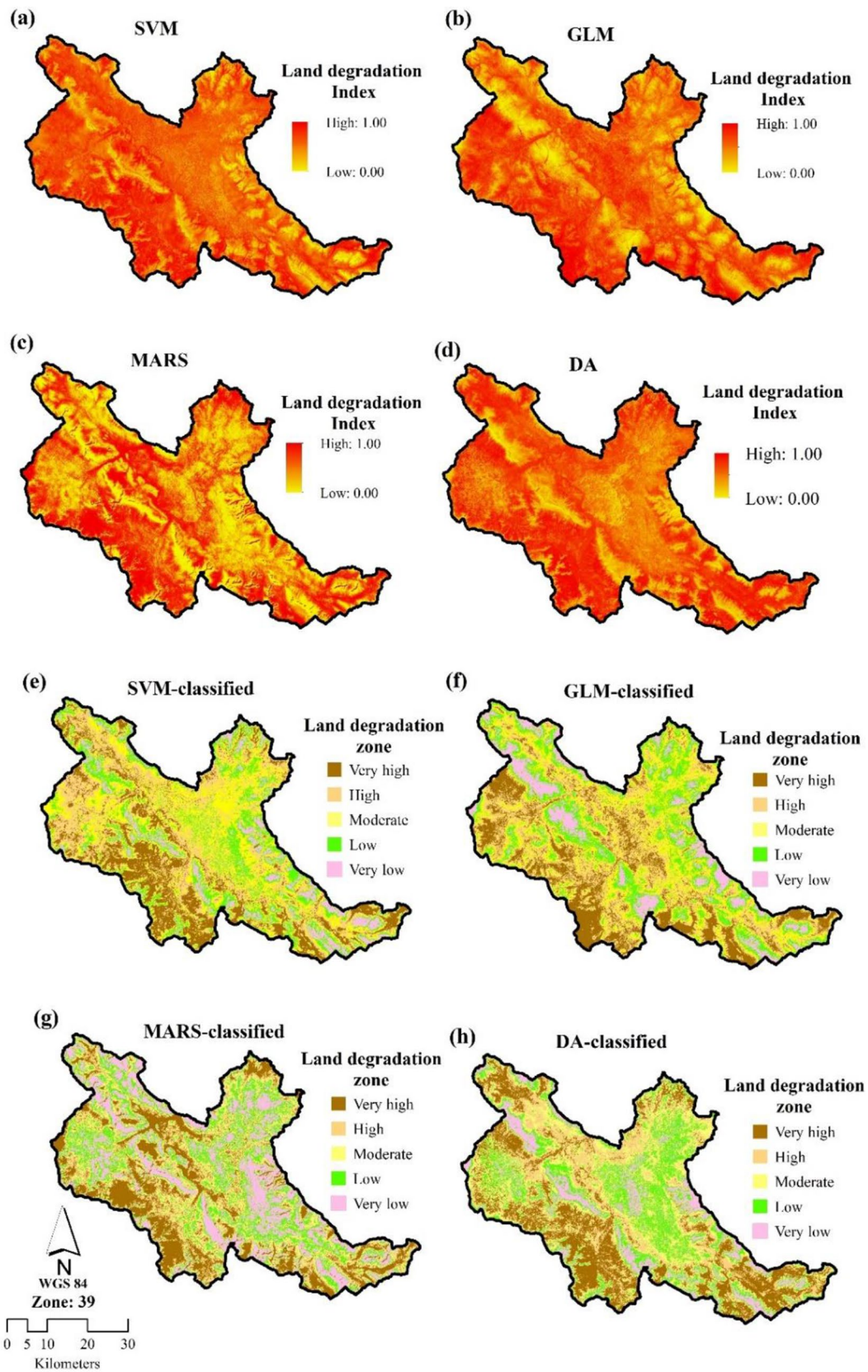


Fig. 6 Land degradation maps based on the benchmark algorithms: **a** SVM, **b** GLM, **c** MARS, and **d** DA, and **e–h** the respective risk zone classification

Table 2 Area of the study watershed falling within different land degradation zones according to the SVM, GLM, MARS, and DA models

	SVM		GLM		MARS		DA	
	Area (km ²)	Area (%)	Area (km ²)	Area (%)	Area (km ²)	Area (%)	Area (km ²)	Area (%)
Very low	157.35	9.06	211.64	12.19	423.14	24.37	213.90	12.32
Low	293.60	16.91	313.44	18.05	327.06	18.84	343.01	19.76
Moderate	455.34	26.23	409.55	23.59	300.49	17.31	387.14	22.30
High	497.20	28.64	466.71	26.88	307.76	17.72	403.36	23.23
Very high	332.75	19.16	334.97	19.29	377.86	21.76	388.83	22.40

Table 3 Performance of the SVM, GLM, MARS, and DA models, based on ROC–AUC and Kappa index (higher values indicate greater model accuracy)

Models	ROC–AUC	Kappa index
SVM	0.864	0.866
GLM	0.829	0.823
MARS	0.825	0.812
DA	0.880	0.892

determines the probability of correctly and incorrectly labeled pixels, with values close to 1 indicating a perfect model with maximum precision and values ≤ 0.5 indicating that the model is not suitable for the analysis. The accuracy and efficiency of the SVM, GLM, MARS, and DA models, based on ROC–AUC and Kappa index, are shown in Table 3. The highest ROC–AUC values were obtained for DA (0.880), followed by SVM (0.864), GLM (0.829), and MARS (0.825) (Table 3). The ROC–AUC curves of

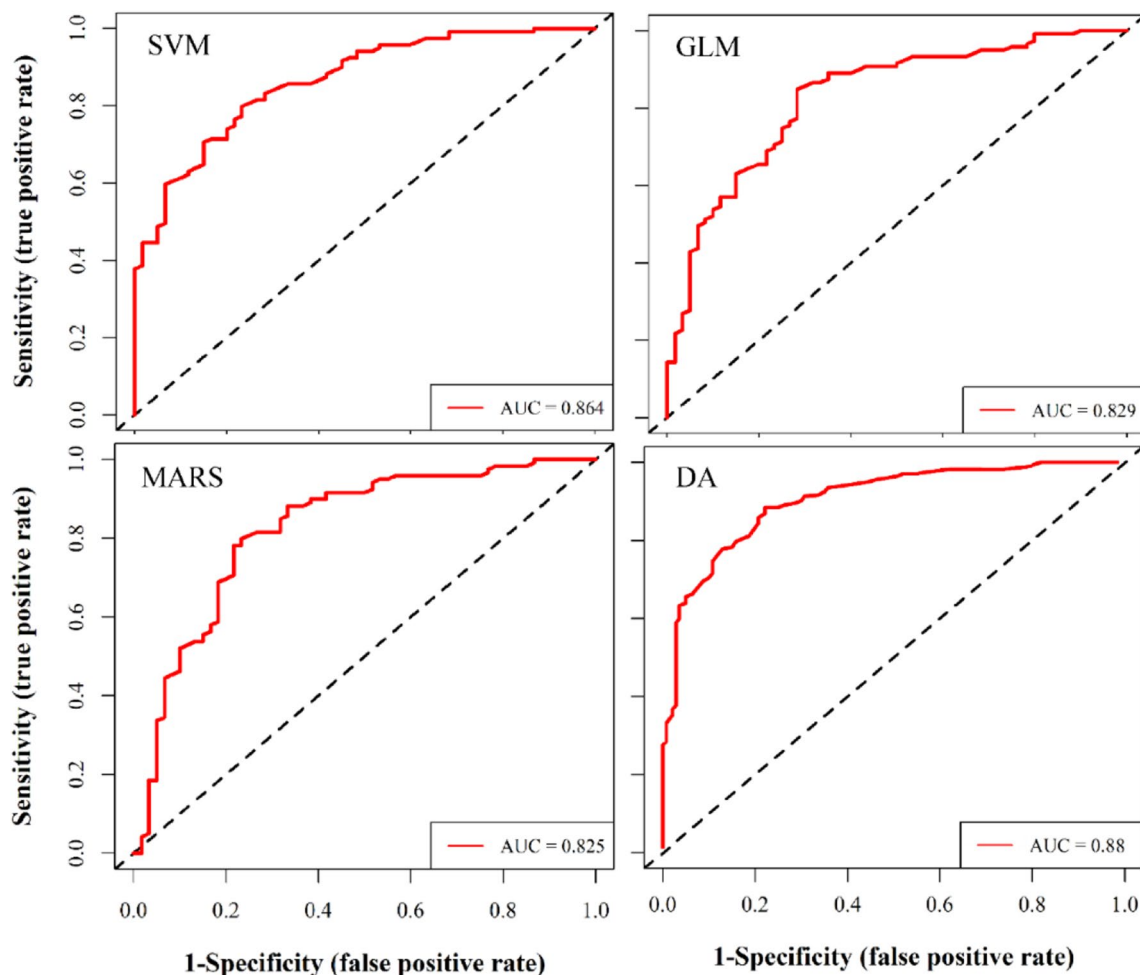


Fig. 7 Receiver-operator characteristic-area under the curve (ROC–AUC) for the SVM, GLM, MARS, and DA models, and the validation data-set

the SVM, GLM, MARS, and DA models for the validation dataset are presented in Fig. 7. In terms of Kappa index, SVM, GLM, MARS, and DA achieved a value of 0.886, 0.823, 0.812, and 0.892 rates, respectively, indicating excellent performance in all cases (Fig. 7). The results obtained in this study cannot be directly compared with those reported in previous studies, because the models we used have not been employed previously in LD risk mapping. An advantage of the machine learning algorithms used in this study was that interactions between natural hazards and biophysical factors causing LD were uncovered.

The Taylor diagram of model performance in producing land degradation risk maps indicated that the DA algorithm had a lower RMSE and higher correlation than the other algorithms (Fig. 8), which were approximately equal in this regard. Comparing the standard deviation of the models revealed that the DA and SVM algorithms were closer to observed values and more in agreement than the others. The standard deviation of the algorithms ranked the models in the order: DA, SVM, GLM, and then MARS. This indicates that DA had the highest accuracy and the other models (SVM, GLM, MARS) could not satisfactorily predict the LD risk map.

Rank variables

According to the aims of the study, the variables were classified into two types, (1) independent variables and (2) dependent variables. The importance of the different variables in LD risk mapping was assessed based on the results obtained with the DA model (selected model), due to its high efficiency and precision. Among the topographic variables, slope (first rank) had the highest importance value (5.842), followed by elevation (3.960), aspect (1.363), terrain ruggedness index (1.294), topographic wetness index (1.215), sky view factor (1.095), and curvature (0.963) (Table 4). Among the human-induced variables, land use (second rank) was the most important (2.974), followed by residential and industrial area expansion (1.539), population density (1.287), population growth rate (1.254), and distance to road (1.222) (Table 3). Among the geo-environmental variables, geology (third rank) had the highest importance value (2.168), followed by C-factor (2.020), precipitation (1.998), land type (1.723), distance to river (1.341), and wind effect (1.086) (Table 4).

Fig. 8 Taylor diagram comparing the performance of the SVM, GLM, MARS, and DA models in land degradation risk mapping

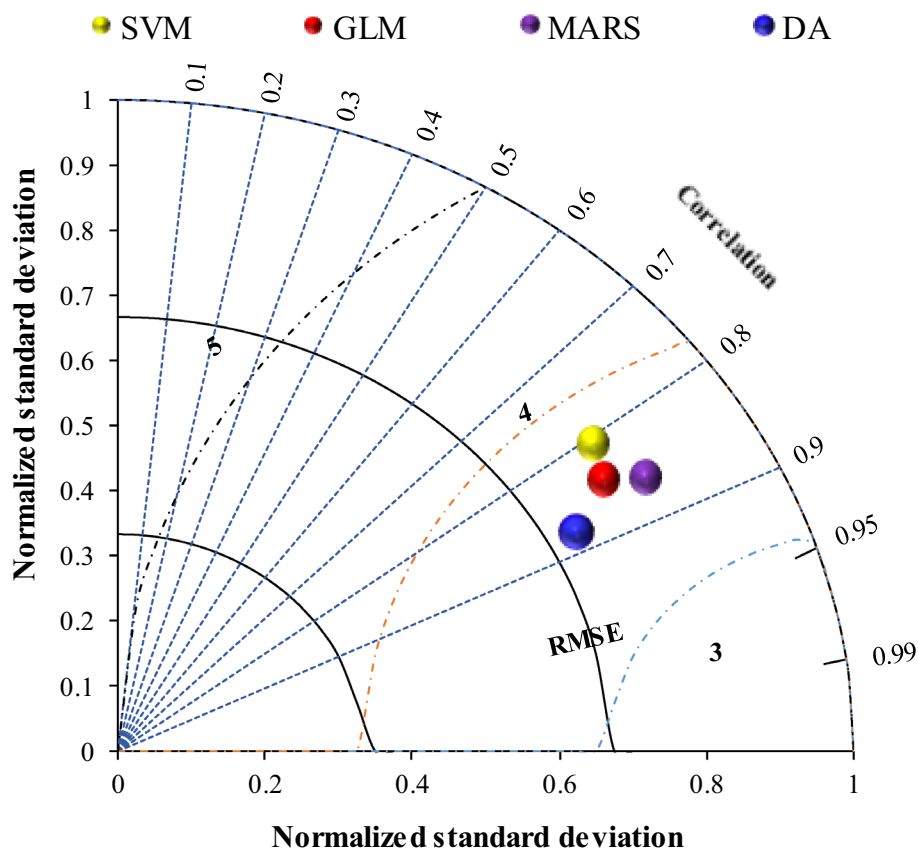


Table 4 Relative importance of different topographic, human-induced, and geo-environmental variables as land degradation conditioning factors

Topographic variables	Importance
Slope	5.842
Elevation	3.960
Aspect	1.363
Terrain ruggedness index	1.294
Topographic wetness index	1.215
Sky view factor	1.095
Curvature	0.963
Human-induced variables	Importance
Land use	2.974
Residential and industrial area expansion	1.539
Population density	1.287
Population growth rate	1.254
Distance to road	1.222
Geo-environmental variables	Importance
Geology	2.168
C-factor	2.020
Precipitation	1.998
Land type	1.723
Distance to river	1.341
Wind effect	1.086

Discussion

Assessment of land degradation status is important for watershed planning and management to protect water quality in lakes and reservoirs. Since LD has accelerated during recent years, precise spatial LD risk mapping is needed to assist authorities in making reliable and reasonable decisions on rehabilitation or restoration of ecosystems and in prioritizing investments. Land degradation problems can arise at three levels: (1) local (field) level, which leads to decreased land productivity and impacts on local businesses, (2) regional level, which causes many

problems for downstream infrastructure regarding decreasing water quality, changes in the hydrological process and flood damage, and also reduced dam capacity by sedimentation, and (3) global level, increasing emissions of greenhouse gases and global warming in the long term (Kust et al. 2018; Chasek et al. 2019; Smetanova et al. 2019). The importance and causes of LD problems vary at each level and differ from case to case. Each individual case at each level involves different types of land use and land cover (e.g., riversides, hills, agricultural areas, steep slopes, deforested areas). Therefore, planners and managers must know the capability and potential of different land uses in interacting with different conditioning factors (such as social and environmental variables) in order to prevent or reduce LD. In previous studies, LD assessments have been performed using different methodologies and scales of analysis. However, machine learning algorithms have not been used previously for this purpose, although they are widely used in assessments of other environmental issues, e.g., flood risk, groundwater pollution, and landslide risk (Tehrany et al. 2015; Termeh et al. 2018; Choubin et al. 2018; Moghaddam et al. 2020; Pourghasemi et al. 2020; Bozdağ et al. 2020). In the present study, we employed four different machine learning algorithms (SVM, MARS, GLM, and DA, advantages and disadvantages of models has been provided in appendix) to generate high quality and accurate LD risk maps for the Pole-Doab watershed in central Iran. Assessments of model performance indicated that DA had the highest accuracy and efficiency, with the greatest learning and prediction power in LD risk mapping. The analysis also clearly revealed the role of different conditioning factors in the LD process. Overall, the models and selected biophysical variables applied in this study provided excellent results and can be recommended for studies in other regions with different conditions and types of land degradation. Land degradation is caused by multiple forces (Cowie et al. 2019), but in this study the main conditioning variables in different categories were found to be slope (topographic variable), land use (human-induced variable), and geology (geo-environmental variable).

Conclusions

Land degradation is an important environmental issue that threatens the sustainability of economic growth and the welfare of the many people, especially rural societies depending on agriculture for their livelihoods. It can also have significant harmful effects globally, e.g., on biodiversity, climate change, and water resources. Good knowledge about the rate of LD would thus be helpful at local, national, and global scale. To this end, we applied machine learning algorithms in LD risk mapping for the Pole-Doab watershed, Iran. We charted the existing conditions for LD and assessed the future trajectory of LD status, as decision support for soil and water resources conservation. Using a novel framework employing socio-environmental conditioning variables in LD risk mapping, based on field measurements and documents describing LD conditions, we showed that serious LD is occurring in the study area. We also showed that this increase in LD is the result of unplanned urbanization with population explosion, development of multiple industries, and agricultural expansion involving conversion of natural rangeland to agricultural land, leading to more frequent flood events. These results

demonstrate the significant role of unsustainable development in LD in the study area. Additional long-term monitoring, considering climatic change and anthropogenic disturbances, is recommended to provide accurate decision support for future LD prevention efforts.

Acknowledgements Our thanks to the Amol authority for supplying the necessary data (flooded locations and thematic layers) and reports, and to the OLVI Foundation for great financial support for this project.

Funding Open access funding provided by University of Oulu including Oulu University Hospital.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix: Advantages and disadvantages of model used

Advantages and disadvantages of the four used models

SVM	GLM	MARS	DA
<i>Advantages</i>			
Works well with clear margin between classes	Easy to understand	Works with both categorical and continuous data	Possesses static and dynamic behaviors
More effective for high dimensional dataset	Easy to organize for any database formats	Automatic and flexible predictive variable selection	Works with few parameters for tuning
Effective for number of dimensions dataset which is greater than the number of samples	Manage different distributions of response	Suitable for large datasets	Contributes in different applications and suitable for large datasets
Memory efficient.	Despite complexity it is fast algorithm	Very fast in predictions	Reasonable time for processing
<i>Disadvantages</i>			
Takes long training time for large dataset	Unable to detect non-linearity directly	Sensitive to overfitting	Sensitive to overflowing
Overlapping in target classes by has noise in dataset	Long processing time and complex algorithm	Low performance with missing data	Premature convergence for the local optimum due to lack of internal memory
SVM will underperform, when number of features exceeds the number of training data	Low predictive authority and needs computing hardware with high power	Difficult to understand	Due to high exploitation rate easily stuck into local optima
Selection of a suitable kernel is challenging	Large number of training and testing runs	Not suitable for missing dataset	

References

- AbdelRahman MA, Natarajan A, Hegde R, Prakash SS (2018) Assessment of land degradation using comprehensive geostatistical approach and remote sensing data in GIS-model builder. *Egypt J Remote Sens Sp Sci*. <https://doi.org/10.1016/j.ejrs.2018.03.002>
- Ahiablame L, Shakya R (2016) Modeling flood reduction effects of low impact development at a watershed scale. *J Environ Manage* 171:81–91. <https://doi.org/10.1016/j.jenvman.2016.01.036>
- Aladejana OO, Salami AT, Adetoro OIO (2018) Hydrological responses to land degradation in the Northwest Benin Owena River Basin, Nigeria. *J Environ Manage* 225:300–312. <https://doi.org/10.1016/j.jenvman.2018.07.095>
- Almagro A, Thomé TC, Colman CB, Pereira RB, Junior JM, Rodrigues DBB, Oliveira PTS (2019) Improving cover and management factor (C-factor) estimation using remote sensing approaches for tropical regions. *Int Soil Water Conserv Res* 7(4):325–334. <https://doi.org/10.1016/j.iswcr.2019.08.005>
- Bakhshandeh E, Hossieni M, Zeraatpisheh M, Francaviglia R (2019) Land use change effects on soil quality and biological fertility: a case study in northern Iran. *Eur J Soil Biol* 95:103119. <https://doi.org/10.1016/j.ejsobi.2019.103119>
- Berendse F, van Ruijven J, Jongejans E, Keesstra S (2015) Loss of plant species diversity reduces soil erosion resistance. *Ecosystems* 18(5):881–888. <https://doi.org/10.1007/s10021-015-9869-6>
- Bernard J, Bocher E, Petit G, Palominos S (2018) Sky view factor calculation in urban context: computational performance and accuracy analysis of two open and free GIS tools. *Climate* 6(3):60. <https://doi.org/10.3390/cli6030060>
- Bewket W, Sterk G (2005) Dynamics in land cover and its effect on stream flow in the Chemoga watershed, Blue Nile basin, Ethiopia. *Hydrol Process* 19(2):445–458. <https://doi.org/10.1002/hyp.5542>
- Bozdağ A, Dokuz Y, Gökçek ÖB (2020) Spatial prediction of PM10 concentration using machine learning algorithms in Ankara, Turkey. *Environ Pollut*. <https://doi.org/10.1016/j.envpol.2020.114635>
- Breslow NE (1996) Generalized linear models: checking assumptions and strengthening conclusions. *Statist Appl* 8(1):23–41. <https://doi.org/10.1.1.50.6105>
- Cao JJ, Holden NM, Adamowski JF, Deo RC, Xu XY, Feng Q (2018) Can individual land ownership reduce grassland degradation and favor socioeconomic sustainability on the Qinghai-Tibetan Plateau? *Environ Sci Policy* 89:192–197. <https://doi.org/10.1016/j.envsci.2018.08.003>
- Cerretelli S, Poggio L, Gimona A, Yakob G, Boke S, Habte M, Black H (2018) Spatial assessment of land degradation through key ecosystem services: the role of globally available data. *Sci Total Environ* 628:539–555. <https://doi.org/10.1016/j.scitotenv.2018.02.085>
- Chasek P, Akhtar-Schuster M, Orr BJ, Luise A, Ratsimba HR, Safriel U (2019) Land degradation neutrality: the science–policy interface from the UNCCD to national implementation. *Environ Sci Policy* 92:182–190. <https://doi.org/10.1016/j.envsci.2018.11.017>
- Chi W, Zhao Y, Kuang W, He H (2019) Impacts of anthropogenic land use/cover changes on soil wind erosion in China. *Sci Total Environ* 668:204–215. <https://doi.org/10.1016/j.scitotenv.2019.03.015>
- Choubin B, Darabi H, Rahmati O, Sajedi-Hosseini F, Kløve B (2018) River suspended sediment modelling using the CART model: a comparative study of machine learning techniques. *Sci Total Environ* 615:272–281. <https://doi.org/10.1016/j.scitotenv.2017.09.293>
- Choubin B, Moradi E, Golshan M, Adamowski J, Sajedi-Hosseini F, Mosavi A (2019) An Ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Sci Total Environ* 651:2087–2096. <https://doi.org/10.1016/j.scitotenv.2018.10.064>
- Choubin B, Abdolshahnejad M, Moradi E, Querol X, Mosavi A, Shamshirband S, Ghamisi P (2020) Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. *Sci Total Environ* 701:134474. <https://doi.org/10.1016/j.scitotenv.2019.134474>
- Cowie AL, Waters CM, Garland F, Orgill SE, Baumber A, Cross R, Metternicht G (2019) Assessing resilience to underpin implementation of Land Degradation Neutrality: a case study in the rangelands of western New South Wales, Australia. *Environ Sci Policy* 100:37–46. <https://doi.org/10.1016/j.envsci.2019.06.002>
- Crossland M, Winowiecki LA, Pagella T, Hadgu K, Sinclair F (2018) Implications of variation in local perception of degradation and restoration processes for implementing land degradation neutrality. *Environ Dev* 28:42–54. <https://doi.org/10.1016/j.envdev.2018.09.005>
- Cura T (2020) Use of support vector machines with a parallel local search algorithm for data classification and feature selection. *Expert Syst Appl* 145:113133. <https://doi.org/10.1016/j.eswa.2019.113133>
- Darabi H, Shahedi K, Solaimani K, Miryaghoubzadeh M (2014) Prioritization of subwatersheds based on flooding conditions using hydrological model, multivariate analysis and remote sensing technique. *Water Environ J* 28(3):382–392. <https://doi.org/10.1111/wej.12047>
- Darabi H, Shahedi K, Mardian M (2016) Flood susceptibility and probability mapping using frequency ratio method in Pol-Doab Shazand Watershed. *Watershed Eng Manag* 8(1):68–79. <https://doi.org/10.22092/IJWMSE.2016.105977>
- Darabi H, Shahedi K, Solaimani K, Kløve B (2018) Hydrological indices variability based on land use change scenarios. *Iran J Watershed Manag Sci* 12(40):81–95. <http://jwmsei.ir/article-1-706-fa.html>
- Darabi H, Haghghi AT, Mohamadi MA, Rashidpour M, Ziegler AD, Hekmatzadeh AA, Kløve B (2020) Urban flood risk mapping using data-driven geospatial techniques for a flood-prone case area in Iran. *Hydrol Res* 51(1):127–142. <https://doi.org/10.2166/nh.2019.090>
- Davudirad AA, Sadeghi SH, Sadoddin A (2016) The impact of development plans on hydrological changes in the Shazand Watershed, Iran. *Land Degrad Dev* 27(4):1236–1244. <https://doi.org/10.1002/ldr.2523>
- Debnath S, Baishy S, Sen D, Arif W (2020) A hybrid memory-based dragonfly algorithm with differential evolution for engineering application. *Eng Comput*. <https://doi.org/10.1007/s00366-020-00958-4>
- Díaz-Cortés MA, Ortega-Sánchez N, Hinojosa S, Oliva D, Cuevas E, Rojas R, Demin A (2018) A multi-level thresholding method for breast thermograms analysis using Dragonfly algorithm. *Infrared Phys Technol* 93:346–361. <https://doi.org/10.1016/j.infrared.2018.08.007>
- Djanibekov U, Van Assche K, Boezeman D, Villamor GB, Djanibekov N (2018) A coevolutionary perspective on the adoption of sustainable land use practices: the case of afforestation on degraded croplands in Uzbekistan. *J Rural Stud* 59:1–9. <https://doi.org/10.1016/j.jrurstud.2018.01.007>
- Drucker H, Burges CJ, Kaufman L, Smola A, Vapnik V (1996) Support vector regression machines. *Adv Neural Inf Proc Syst* 9:155–161
- El-Khoury A, Seidou O, Lapen DR, Que Z, Mohammadian M, Sunohara M, Bahram D (2015) Combined impacts of future climate and land use changes on discharge, nitrogen and phosphorus

- loads for a Canadian river basin. *J Environ Manage* 151:76–86. <https://doi.org/10.1016/j.jenvman.2014.12.012>
- Feng Q, Zhao W, Jun Wang X, Zhang MZ, Zhong L, Fang X (2016) Effects of different land-use types on soil erosion under natural rainfall in the Loess Plateau, China. *Pedosphere* 26(2):243–256. [https://doi.org/10.1016/S1002-0160\(15\)60039-X](https://doi.org/10.1016/S1002-0160(15)60039-X)
- Fenta AA, Tsunekawa A, Haregeweyn N, Poesen J, Tsubo M, Borrelli P, Kawai T (2020) Land susceptibility to water and wind erosion risks in the East Africa region. *Sci Total Environ* 703:135016. <https://doi.org/10.1016/j.scitotenv.2019.135016>
- Fratini P, Crosta G, Carrara A (2010) Techniques for evaluating the performance of landslide susceptibility models. *Eng Geol* 111(1–4):62–72. <https://doi.org/10.1016/j.enggeo.2009.12.004>
- Friedman JH (1991) Multivariate adaptive regression splines. *Ann Statist* 19(1):1–67. <https://doi.org/10.1214/aos/1176347963>
- Gebremicael TG, Mohamed YA, van Der Zaag P, Hagos EY (2018) Quantifying longitudinal land use change from land degradation to rehabilitation in the headwaters of Tekeze-Atbara Basin, Ethiopia. *Sci Total Environ* 622:1581–1589. <https://doi.org/10.1016/j.scitotenv.2017.10.034>
- Gichenje H, Pinto-Correia T, Godinho S (2019) An analysis of the drivers that affect greening and browning trends in the context of pursuing land degradation-neutrality. *Remote Sens Appl Soc Environ* 15:100251. <https://doi.org/10.1016/j.rsase.2019.100251>
- Giuliani G, Mazzetti P, Santoro M, Nativi S, Van Bemmelen J, Colangeli G, Lehmann A (2020) Knowledge generation using satellite earth observations to support sustainable development goals (SDG): a use case on Land degradation. *Int J Appl Earth Obs Geoinf* 88:102068. <https://doi.org/10.1016/j.jag.2020.102068>
- Gu W, Guo J, Fan K, Chan EH (2016) Dynamic land use Change and sustainable urban development in a third-tier city within Yangtze Delta. *Procedia Environ Sci* 36:98–105. <https://doi.org/10.1016/j.proenv.2016.09.019>
- Haghighi AT, Sadegh M, Bhattacharjee J, Sönmez ME, Noury M, Yilmaz N, Kløve B (2020) The impact of river regulation in the Tigris and Euphrates on the Arvandroud Estuary. *Progress Phys Geogr Earth Environ*. <https://doi.org/10.1177/0309133320938676>
- Hazbavi Z, Sadeghi SH, Gholamalifard M, Davudirad AA (2020) Watershed health assessment using the pressure–state–response (PSR) framework. *Land Degrad Dev* 31(1):3–19. <https://doi.org/10.1002/ldr.3420>
- Hong H, Panahi M, Shirzadi A, Ma T, Liu J, Zhu AX, Kazakis N (2018) Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. *Sci Total Environ* 621:1124–1141. <https://doi.org/10.1016/j.scitotenv.2017.10.114>
- Huber-Sannwald E, Maestre FT, Herrick JE, Reynolds JF (2006) Ecohydrological feedbacks and linkages associated with land degradation: a case study from Mexico. *Hydrol Process* 20(15):3395–3411. <https://doi.org/10.1002/hyp.6337>
- Jafari M, Chaleshtari MHB (2017) Using dragonfly algorithm for optimization of orthotropic infinite plates with a quasi-triangular cut-out. *Eur J Mech A/Solids* 66:1–14. <https://doi.org/10.1016/j.euromechsol.2017.06.003>
- Jaquet S, Schwilch G, Hartung-Hofmann F, Adhikari A, Sudmeier-Rieux K, Shrestha G et al (2015) Does outmigration lead to land degradation? Labour shortage and land management in a western Nepal watershed. *Appl Geogr* 62:157–170. <https://doi.org/10.1016/j.apgeog.2015.04.013>
- Jiang L, Jiapaer G, Bao A, Li Y, Guo H, Zheng G, De Maeyer P (2019) Assessing land degradation and quantifying its drivers in the Amudarya River delta. *Ecol Ind* 107:105595. <https://doi.org/10.1016/j.ecolind.2019.105595>
- Karimidastenaei Z, Haghighi AT, Rahmati O, Rasouli K, Rozbeh S, Pirnia A, Kløve B (2020) Fog-water harvesting Capability Index (FCI) mapping for a semi-humid catchment based on socio-environmental variables and using artificial intelligence algorithms. *Sci Total Environ* 708:135115. <https://doi.org/10.1016/j.scitotenv.2019.135115>
- Keesstra SD, Bouma J, Wallinga J, Tiftonell P, Smith P, Cerdà A, Montanarella L, Quinton JN, Pachepsky Y, Van Der Putten WH, Bardgett RD, Moolenaar S, Mol G, Jansen B, Fresco LO (2016) The significance of soils and soil science towards realization of the United Nations sustainable development goals. *Soil* 2:111–128. <https://doi.org/10.5194/soil-2-111-2016>
- Khosravi H, Moradi E, Darabi H (2015). Identification of homogeneous groundwater quality regions using factor and cluster analysis; a case study Ghir plain of Fars province. *J Irrig Water Eng* 6(21):119–133. http://www.waterjournal.ir/article_73846.html
- Krkoška Lorencová E, Harmáčková ZV, Landová L, Pártl A, Vačkář D (2016) Assessing impact of land use and climate change on regulating ecosystem services in the Czech Republic. *Ecosyst Health Sustain* 2(3):e01210. <https://doi.org/10.1002/ehs2.1210>
- Ks SR, Murugan S (2017) Memory based hybrid dragonfly algorithm for numerical optimization problems. *Expert Syst Appl* 83:63–78. <https://doi.org/10.1016/j.eswa.2017.04.033>
- Kust G, Andreeva O, Lobkovskiy V, Telnova N (2018) Uncertainties and policy challenges in implementing Land Degradation Neutrality in Russia. *Environ Sci Policy* 89:348–356. <https://doi.org/10.1016/j.envsci.2018.08.010>
- Li Z, Deng X, Yin F, Yang C (2015) Analysis of climate and land use changes impacts on land degradation in the North China Plain. *Adv Meteorol*. <https://doi.org/10.1155/2015/976370>
- Li LL, Zhao X, Tseng ML, Tan RR (2020) Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm. *J Clean Prod* 242:118447. <https://doi.org/10.1016/j.jclepro.2019.118447>
- Liniger H, Harari N, van Lynden G, Fleiner R, de Leeuw J, Bai Z, Critchley W (2019) Achieving land degradation neutrality: the role of SLM knowledge in evidence-based decision-making. *Environ Sci Policy* 94:123–134. <https://doi.org/10.1016/j.envsci.2019.01.001>
- Lu D, Batistella M, Mausel P, Moran E (2007) Mapping and monitoring land degradation risks in the Western Brazilian Amazon using multitemporal Landsat TM/ETM+ images. *Land Degrad Dev* 18(1):41–54. <https://doi.org/10.1002/ldr.762>
- Mariano DA, dos Santos CA, Wardlow BD, Anderson MC, Schiltmeyer AV, Tadesse T, Svoboda MD (2018) Use of remote sensing indicators to assess effects of drought and human-induced land degradation on ecosystem health in Northeastern Brazil. *Remote Sens Environ* 213:129–143. <https://doi.org/10.1016/j.rse.2018.04.048>
- McCullagh P, Nelder JA (1989) Monographs on statistics and applied probability. In: Generalized linear models, vol 37
- Mekonnen Z, Berie HT, Woldeamanuel T, Asfaw Z, Kassa H (2018) Land use and land cover changes and the link to land degradation in Arsi Negele district, Central Rift Valley, Ethiopia. *Remote Sens Appl Soc Environ* 12:1–9. <https://doi.org/10.1016/j.rsase.2018.07.012>
- Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F, Chang CC, Lin CC, Meyer MD (2019) Package ‘e1071’. The R Journal
- Mirjalili S (2016) Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. *Neural Comput Appl* 27(4):1053–1073. <https://doi.org/10.1007/s00521-015-1920-1>
- Moghaddam DD, Rahmati O, Panahi M, Tiefenbacher J, Darabi H, Haghizadeh A, Bui DT (2020) The effect of sample size on different machine learning models for groundwater potential mapping in mountain bedrock aquifers. *CATENA* 187:104421. <https://doi.org/10.1016/j.catena.2019.104421>

- Monserud RA, Leemans R (1992) Comparing global vegetation maps with the Kappa statistic. *Ecol Model* 62(4):275–293. [https://doi.org/10.1016/0304-3800\(92\)90003-W](https://doi.org/10.1016/0304-3800(92)90003-W)
- Mountrakis G, Im J, Ogole C (2011) Support vector machines in remote sensing: a review. *ISPRS J Photogramm Remote Sens* 66(3):247–259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>
- Nelder JA, Wedderburn RW (1972) Generalized linear models. *J R Stat Soc Ser A (General)* 135(3):370–384. <https://doi.org/10.2307/2344614>
- Nordin ND, Zan MSD, Abdullah F (2020) Generalized linear model for enhancing the temperature measurement performance in Brillouin optical time domain analysis fiber sensor. *Opt Fiber Technol* 58:102298. <https://doi.org/10.1016/j.yofte.2020.102298>
- Nunes AN, De Almeida AC, Coelho CO (2011) Impacts of land use and cover type on runoff and soil erosion in a marginal area of Portugal. *Appl Geogr* 31(2):687–699. <https://doi.org/10.1016/j.apgeog.2010.12.006>
- Okpara UT, Stringer LC, Akhtar-Schuster M, Metternicht GI, Dallimer M, Requier-Desjardins M (2018) A social-ecological systems approach is necessary to achieve land degradation neutrality. *Environ Sci Policy* 89:59–66. <https://doi.org/10.1016/j.envsci.2018.07.003>
- Ozdemir A, Altural T (2013) A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping: Sultan Mountains, SW Turkey. *J Asian Earth Sci* 64:180–197. <https://doi.org/10.1016/j.jseae.2012.12.014>
- Pirnia A, Golshan M, Darabi H, Adamowski J, Rozbeh S (2018) Using the Mann-Kendall test and double mass curve method to explore stream flow changes in response to climate and human activities. *J Water Clim Change*. <https://doi.org/10.2166/wcc.2018.162>
- Pirnia A, Darabi H, Choubin B, Omidvar E, Onyutha C, Haghghi AT (2019) Contribution of climatic variability and human activities to stream flow changes in the Haraz River basin, northern Iran. *J Hydro-environ Res* 25:12–24. <https://doi.org/10.1016/j.jher.2019.05.001>
- Pour RM, Haghghi AT, Sarmi H, Keshtkaran P (2009) Watershed management and its effect on sedimentation in Doroudzan dam. *Sichuan Daxue Xuebao (Ziran Kexueban)* 41:242–248
- Pourghasemi HR, Kornejady A, Kerle N, Shabani F (2020) Investigating the effects of different landslide positioning techniques, landslide partitioning approaches, and presence–absence balances on landslide susceptibility mapping. *CATENA* 187:104364. <https://doi.org/10.1016/j.catena.2019.104364>
- Prăvălie R, Săvulescu I, Patriche C, Dumitrașcu M, Bandoc G (2017) Spatial assessment of land degradation sensitive areas in southwestern Romania using modified MEDALUS method. *CATENA* 153:114–130. <https://doi.org/10.1016/j.catena.2017.02.011>
- Pullanikkatil D, Palamuleni L, Ruhiiiga T (2016) Assessment of land use change in Likangala River catchment, Malawi: a remote sensing and DPSIR approach. *Appl Geogr* 71:9–23. <https://doi.org/10.1016/j.apgeog.2016.04.005>
- Rahman CM, Rashid TA (2019) Dragonfly algorithm and its applications in applied science survey. *Comput Intell Neurosci*. <https://doi.org/10.1155/2019/9293617>
- Rahmati O, Falah F, Dayal KS, Deo RC, Mohammadi F, Biggs T, Bui DT (2020) Machine learning approaches for spatial modeling of agricultural droughts in the south-east region of Queensland Australia. *Sci Total Environ* 699:134230. <https://doi.org/10.1016/j.scitotenv.2019.134230>
- Riley SJ, DeGloria SD, Elliot R (1999) Index that quantifies topographic heterogeneity. *Intermt J Sci* 5(1–4):23–27. <https://doi.org/10.1371/journal.pone.0001298>
- Riva MJ, Daliakopoulos IN, Eckert S, Hodel E, Liniger H (2017) Assessment of land degradation in Mediterranean forests and grazing lands using a landscape unit approach and the normalized difference vegetation index. *Appl Geogr* 86:8–21. <https://doi.org/10.1016/j.apgeog.2017.06.017>
- Sadeghi SH, Hazbavi Z, Gholamalifard M (2019) Interactive impacts of climatic, hydrologic and anthropogenic activities on watershed health. *Sci Total Environ* 648:880–893. <https://doi.org/10.1016/j.scitotenv.2018.08.004>
- Sajedi-Hosseini F, Malekian A, Choubin B, Rahmati O, Cipullo S, Coulon F, Pradhan B (2018) A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. *Sci Total Environ* 644:954–962. <https://doi.org/10.1016/j.scitotenv.2018.07.054>
- Schwieger DAM, Mbidzo M (2020) Socio-historical and structural factors linked to land degradation and desertification in Namibia's former Herero 'homelands'. *J Arid Environ* 178:104151. <https://doi.org/10.1016/j.jaridenv.2020.104151>
- Seraji MHS, Haghghi AT, Keshtkaran P (2009) Comparing the real value of sediment load with the results of erosion models in Kor River. In: Special issue on international symposium of iahs-pub and the 2~(nd) international symposium of China-Pub–hydrological modeling and integrated water resources management in ungauged mountainous watershed. *Sichuan Daxue Xuebao (Gongcheng Kexue Ban)/J. Sichuan University (Eng. Sci. Edition)*, 41, pp 319–324
- Serrano NB, Sánchez AS, Lasheras FS, Iglesias-Rodríguez FJ, Valverde GF (2020) Identification of gender differences in the factors influencing shoulders, neck and upper limb MSD by means of multivariate adaptive regression splines (MARS). *Appl Ergonom* 82:102981. <https://doi.org/10.1016/j.apergo.2019.102981>
- Shao Y, Jiang QO, Wang C, Wang M, Xiao L, Qi Y (2020) Analysis of critical land degradation and development processes and their driving mechanism in the Heihe River Basin. *Sci Total Environ* 716:137082. <https://doi.org/10.1016/j.scitotenv.2020.137082>
- Shilaja C, Arunprasath T (2019) Internet of medical things-load optimization of power flow based on hybrid enhanced grey wolf optimization and dragonfly algorithm. *Future Gen Comput Syst* 98:319–330. <https://doi.org/10.1016/j.future.2018.12.070>
- Sklenicka P (2016) Classification of farmland ownership fragmentation as a cause of land degradation: a review on typology, consequences, and remedies. *Land Use Policy* 57:694–701. <https://doi.org/10.1016/j.landusepol.2016.06.032>
- Smetanova A, Follain S, David M, Ciampalini R, Raclot D, Crabit A, Le Bissonnais Y (2019) Landscaping compromises for land degradation neutrality: the case of soil erosion in a Mediterranean agricultural landscape. *J Environ Manage* 235:282–292. <https://doi.org/10.1016/j.jenvman.2019.01.063>
- Solomon MK, Barger N, Cerda A, Keesstra S, Marković M (2018) Assessing land condition as a first step to achieving land degradation neutrality: a case study of the Republic of Srpska. *Environ Sci Policy* 90:19–27. <https://doi.org/10.1016/j.envsci.2018.09.014>
- Speranza CI, Adenle A, Boillat S (2019) Land Degradation Neutrality-Potentials for its operationalisation at multi-levels in Nigeria. *Environ Sci Policy* 94:63–71. <https://doi.org/10.1016/j.envsci.2018.12.018>
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res Atmos* 106(D7):7183–7192. <https://doi.org/10.1029/2000JD900719>
- Tehrany MS, Pradhan B, Mansor S, Ahmad N (2015) Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *CATENA* 125:91–101. <https://doi.org/10.1016/j.catena.2014.10.017>
- Tehrany MS, Jones S, Shabani F (2019) Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. *CATENA* 175:174–192. <https://doi.org/10.1016/j.catena.2018.12.011>

- Termeh SVR, Kornejady A, Pourghasemi HR, Keesstra S (2018) Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. *Sci Total Environ* 615:438–451. <https://doi.org/10.1016/j.scitotenv.2017.09.262>
- Tien Bui D, Pradhan B, Lofman O, Revhaug I (2012) Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and Naive Bayes Models. *Math Probl Eng*. <https://doi.org/10.1155/2012/974638>
- Torabi Haghighi A, Menberu MW, Darabi H, Akanegbu J, Kløve B (2018) Use of remote sensing to analyse peatland changes after drainage for peat extraction. *Land Degrad Dev* 29(10):3479–3488. <https://doi.org/10.1002/ldr.3122>
- Tripathi V, Edrisi SA, Chen B, Gupta VK, Vilu R, Gathergood N, Abhilash PC (2017) Biotechnological advances for restoring degraded land for sustainable development. *Trends Biotechnol* 35(9):847–859. <https://doi.org/10.1016/j.tibtech.2017.05.001>
- Van Haren N, Fleiner R, Liniger H, Harari N (2019) Contribution of community-based initiatives to the sustainable development goal of Land Degradation Neutrality. *Environ Sci Policy* 94:211–219. <https://doi.org/10.1016/j.envsci.2018.12.017>
- Vaverková MD, Maxianová A, Winkler J, Adamcová D, Podlasek A (2019) Environmental consequences and the role of illegal waste dumps and their impact on land degradation. *Land Use Policy* 89:104234. <https://doi.org/10.1016/j.landusepol.2019.104234>
- Vert JP, Tsuda K, Schölkopf B (2004) A primer on kernel methods. *Kernel Methods Comput Biol* 47:35–70. <https://doi.org/10.7551/mitpress/4057.003.0004>
- Wei W, Gao Y, Huang J, Gao J (2020) Exploring the effect of basin land degradation on lake and reservoir water quality in China. *J Clean Prod*. <https://doi.org/10.1016/j.jclepro.2020.122249>
- Wieland R, Lakes T, Yunfeng H, Nendel C (2019) Identifying drivers of land degradation in Xilingol, China, between 1975 and 2015. *Land Use Policy* 83:543–559. <https://doi.org/10.1016/j.landusepol.2019.02.013>
- Wunder S, Bodle R (2019) Achieving land degradation neutrality in Germany: implementation process and design of a land use change based indicator. *Environ Sci Policy* 92:46–55. <https://doi.org/10.1016/j.envsci.2018.09.022>
- Xu X, Hoang S, Mayo MW, Bekiranov S (2010) Application of machine learning methods to histone methylation ChIP-Seq data reveals H4R3me2 globally represses gene expression. *BMC Bioinform* 11(1):396. <https://doi.org/10.1186/1471-2105-11-396>
- Yatheendradas S, Wagener T, Gupta H, Unkrich C, Goodrich D, Schaffner M, Stewart A (2008) Understanding uncertainty in distributed flash flood forecasting for semiarid regions. *Water Resour Res*. <https://doi.org/10.1029/2007WR005940>
- Zakšek K, Oštir K, Kokalj Ž (2011) Sky-view factor as a relief visualization technique. *Remote Sens* 3(2):398–415. <https://doi.org/10.3390/rs3020398>
- Zhang W, Goh AT (2016) Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geosci Front* 7(1):45–52. <https://doi.org/10.1016/j.gsf.2014.10.003>
- Zhao G, Pang B, Xu Z, Peng D, Xu L (2019) Assessment of urban flood susceptibility using semi-supervised machine learning model. *Sci Total Environ* 659:940–949. <https://doi.org/10.1016/j.scitotenv.2018.12.217>
- Zhu J, Wu W, Liu HB (2018) Environmental variables controlling soil organic carbon in top-and sub-soils in karst region of southwestern China. *Ecol Ind* 90:624–632. <https://doi.org/10.1016/j.ecoli.2018.03.073>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.