RESEARCH ARTICLE

Introducing the coupled stepwise areal constraining and Mahalanobis distance: a promising MCDM-based probabilistic model for landfill site selection

Davoud Davoudi Moghaddam[1] · Ali Haghizadeh[1] · Naser Tahmasebipour[1] · Hossein Zeinivand[1]

Abstract

This study sets out to propose a new ensemble of probabilistic spatial modeling and multi-criteria decision-making comprised of stepwise areal constraining and Mahalanobis distance algorithms in order to assess areal suitability for landfilling. The Ardak watershed was selected as the study area due to encountering several cases of open garbage dumps and uncontrolled landfills which are one of the main sources of river water pollution in the upstream of the Ardak dam. The results revealed that the proposed algorithm successfully assists in inventory-irrespective probabilistic modeling of landfill siting which is mainly indebted to the role of areal constraining in providing training and validation samples for the Mahalanobis distance model. The latter also showed a robust pattern recognition results from which a discernible differentiation of the area was attained while the spatial dependencies between the environmental factors were taken into account. Mahalanobis distance also gave an outstanding performance in terms of goodness of fit (area under the success rate 89.367) and prediction power (area under the success rate 89.252). Based on a five-point scale classification scheme, about 2.7% and 2.6% of the study area, respectively, have high and very high suitability for landfilling, while the remaining area is shared between very low-to-moderate suitability classes. According to the current trail of literature regarding landfill site selection which mostly relies on mere areal filtering, a probabilistic model would give invaluable inferences regarding the pattern of suitability/susceptibility of the area of interest and causative role of the influential factors.

Keywords Multi-criteria decision-making · Probabilistic modeling · Success rate curve · Prediction rate curve · Goodness of fit

Introduction

Locating the landfill sites as the disposal location of waste materials that are of no use for recycling has been a pivotal issue for municipal and industrial waste management agencies (Ngoc and Schnitzer 2009; Khan et al. 2018). Despite the official reports concerning the considerable waste generation per capita in rural areas accounting for about 220–340 g per day in 2016 as well as the regulations designed to govern waste management in Iran (Khayamabshi 2016; Tehran

times 2017), the latter has come to a head and municipalities have begun to identify safe landfilling locations and propose modern disposal techniques such as digesters, incinerators, and biogas (del Valle-Zermeño et al. 2015; Khayamabshi 2016). Landfill site locating is substantial in developing countries where environmental restrictions and organizational benefits may occasionally stand on opposite sides. The countless poorly located landfill sites at or nearby the rural areas have been threatening the surrounding environment including its inhabitants by producing an unpleasant odor, diminishing esthetic values, and contaminating water resources (Rathi 2006; Mondelli et al. 2007; Li'ao et al. 2009; Khorram et al. 2015). Open garbage dumps and uncontrolled landfills have been significantly affecting the quality of the streams in the Ardak watershed in the north of Khorasan Razavi province, Iran. According to the official reports of Water Company and Forest, Range and Watershed Management Organization of Khorasan Razavi province, effluents from landfill leachates

✉ Ali Haghizadeh
  haghizadeh.a@lu.ac.ir

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

have contaminated nearby surface waters as well as groundwater resources by injecting different pollutants such as xenobiotic organic compounds, heavy metals, and nutrient salts (RWCKR[1] 2018). Hence, precise and effective landfill site selection techniques are essential for water quality management to address downstream pollution concerns.

Rating, weighting, and ranking have been the most common basis of landfill site selection so far with which many novelties have been offered in terms of model integration, data inclusion/configuration, and geographical toolboxes with setback spatial criteria go forth. Developing a GIS tool underpinned by the ensemble of Not-In-My-Backyard (NIMBY) syndrome and nuisance score (Simsek et al. 2014a, b) and incorporating different multi-criteria decision-making methods such as simple additive weighting (SAW), analytical hierarchy process (AHP), and TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) (Rahmat et al. 2017, 2018; Chabuk et al. 2017; Spigolon et al. 2018; Khan et al. 2018; Şener et al. 2010; Kamdar et al. 2019; Chabuk et al. 2019; Özkan et al. 2019), remote sensing studies with employing aerial photography, airborne and spaceborne sensors (Manzo et al. 2016; Mahmood et al. 2017), novel Thiessen polygon–based ranking schemes integrated with remotely examined environmental criteria (Richter et al. 2019), and electrical resistivity combined with a MCDM method (Goulart Coelho et al. 2017; Yousefi et al. 2018; Akintorinwa and Okoro 2019) are some of these endeavors.

Although such methodologies have expanded the knowledge on the ways to better concentrate on the desired landfill sites which are most commonly termed as areal constraining/filtering, the current trail of literature has not considered the element of data mining and pattern recognition nor acknowledged their merits. Here we proposed a novel data driven-probabilistic ensemble model that bears not only on-site selection but also recognition of similar patterns with different degrees of similarities to the ideal conditions. The latter, in particular, is a key principle when authorities require a highly differentiated spatial map that includes the transient zone which may be transferred into either suitability or unsuitability conditions over the course of time. Such a technique would simply improve traditional binary classifications. In this process, we attempted to employ the most easy-to-acquire environmental data in order to provide an all-inclusive application of the proposed methodology. Hence, the main contribution of this paper is the probabilistic site selection scheme which would serve the decision-makers with much more flexible decision-making maps—a theme that is beyond a binary presence-absence classification.

Study area

The Ardak watershed extends for an area of about 479 km^2 and is located in the north of Khorasan Razavi province, Iran.

It lies between latitudes 4,066,633 to 4,096,281 N and longitudes 690,339 to 725,597 E, UTM zone 40 N. Elevation ranges between 1235 and 2950 m.a.s.l (Fig. 1). A total of 15 residential centers are evident which are sporadically distributed across the study area. Roads at 55.5 km—paved an unpaved—are extended across the region. A 146-km stream network is sprawled over a diverse land cover types. Land covers were disintegrated in terms of type and the density of the vegetation cover and usage-wise to attain more realistic and precise results. Based on the latter, poor pastures cover the largest area (74.5% of the entire region), followed by moderate-condition pastures (8.5%), and very-low-density forests (8.1%) while the remaining area is shared between orchards, rainfed farming, low-density forests, and woodlands. The appearance of 238 springs across the region intuitively implies the high groundwater potential in the area. Moreover, there is an important dam reservoir in the southernmost part of the study area. Mashhad, one of the largest metropolitan cities of Iran, is located at the reservoir's downstream end and accommodates above 3 million residents. The latter indicates the pivotal role of this watershed in providing the water supply of the downstream regions. Many garbage dumps were evident mostly nearby the streams, roads, and residential centers during the filed surveys where an instant change in natural color and quality of surface water were discernible right after the contaminated area, endangering the natural habitats and human beings (Fig. 2).

Materials and methods

The workflow of this study is presented in Fig. 3. What follows is thorough explanations of the adopted methodological stages.

Data compilation

Although most of the spatial modeling efforts require inventory data for the modeling or model validation phases, the methodology proposed in this work has an inventory-irrespective scheme. The process consists of two main phases, namely, areal constraining and probabilistic modeling. In fact, the final product of areal constraining signifies the most suitable sites for landfills which, itself, provides the inventory for probabilistic modeling. This procedure follows the selection and classification of landfill-representative factors. The adopted factors were selected based on different criteria such as literature review, data availability, spatial heterogeneity of the factors' pattern across the study area (i.e., excluding homogenous factors with less unique information), and, most importantly, the multicollinearity issue. The latter can cause bias in models' results due to the existing strong correlation between the used thematic maps which should be obviated
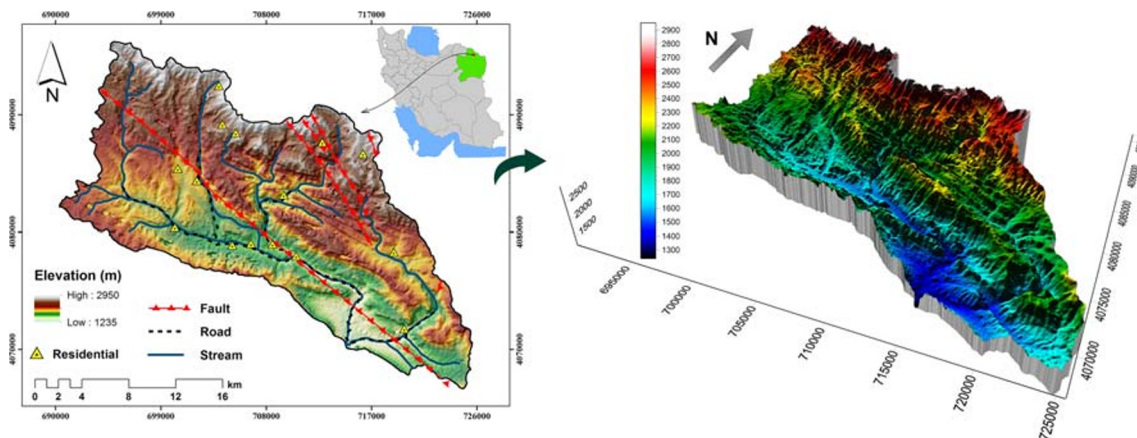
---
[1] Regional Water Company of Khorasan Razavi

Fig. 1 The geographical location of the study area in Khorasan Razavi province in Iran. Right, an oblique view of the region

prior to the modeling process. In this work, we used the variance inflation factor (VIF) and its reciprocal, tolerance index, to assess the strength of correlation between independent variables (i.e., thematic maps) and the multicollinearity issue. The VIF value higher than 5 and accordingly the tolerance index lower than 0.2 indicates critical multicollinearity (O'brien 2007). In this regard, Kutner et al. (2004) also suggested that the VIF values higher than 10 can represent critical multicollinearity (Kutner et al. 2004); however, the VIF value of 5 was considered the basis of factor exclusion procedure to ensure the latter issue (Table 1). The correlation matrix is another way to address the strong correlation between the factors in a pairwise manner which can be calculated with different methods such as principle comparison analysis

(PCA), yet the VIF test is comparatively more elaborative and distinctly points out the responsible factor for the multicollinearity issue, although both methods revolve around the same concept and selecting between them is the matter of choice. Additionally, the VIF method has been increasingly used and getting more common in the literature specifically for multicollinearity issue (Kornejady et al. 2018; Chen et al. 2017; Haghizadeh et al. 2017; Rahmati et al. 2019; Moghaddam et al. 2020a, b).

In light of these sieving procedures, eight predictors (i.e., drivers) were selected including slope degree, proximity to faults/reservoir/residential areas/roads/springs/streams, and land cover types. The thematic maps of the conditioning factors were prepared in ArcGIS 10.3 as presented in Fig. 4.

Fig. 2 Some photographs of open garbage dumps nearby streams (**a**, **c**, **e**) causing appearance of harmful green algae in water (**b**) and fish die-offs (**d**) due to the contaminations
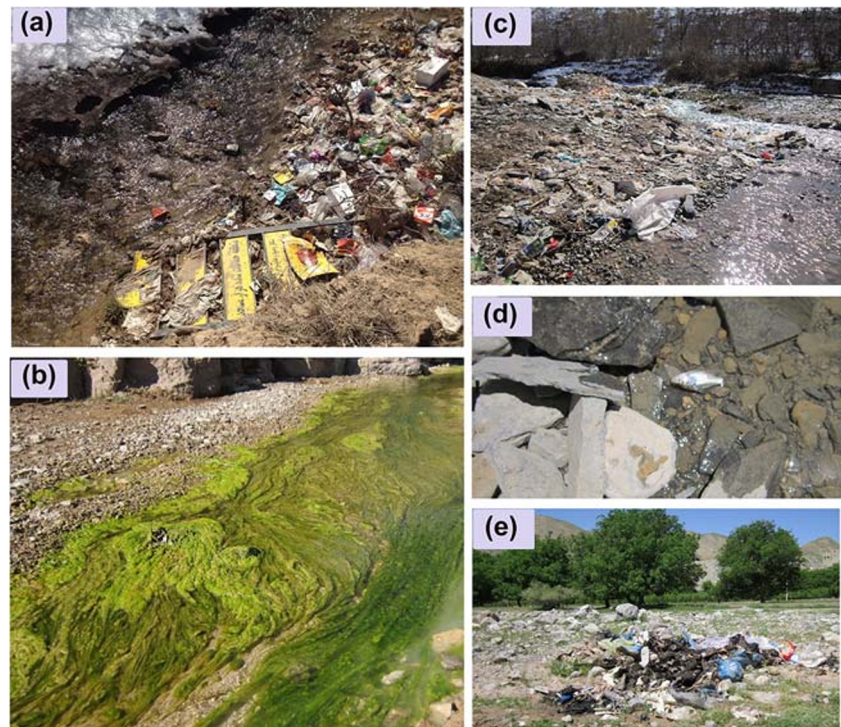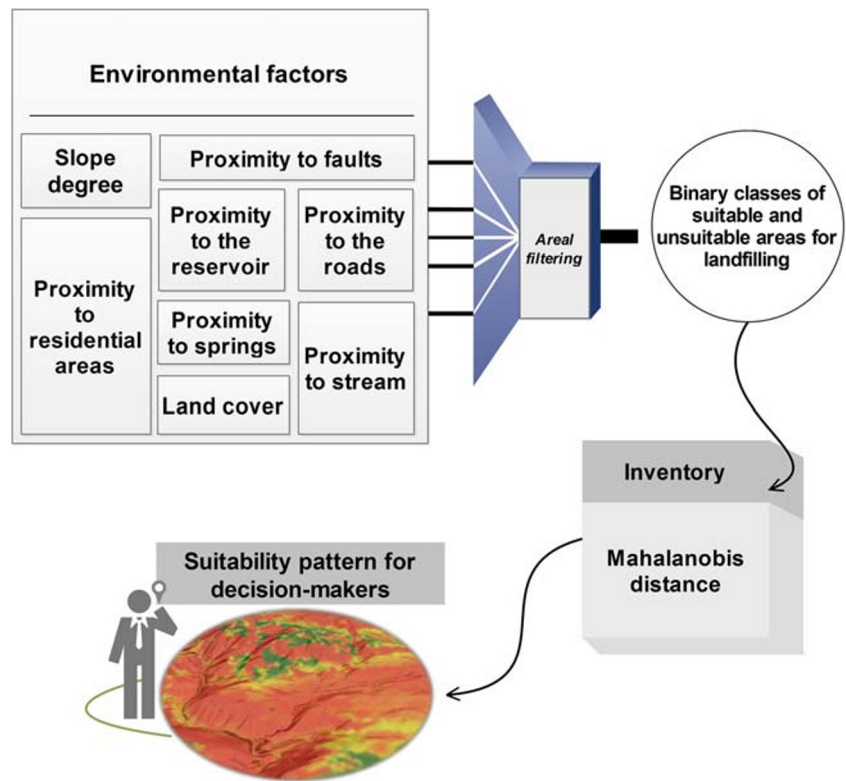
Fig. 3 Methodological flowchart adopted in this study

More detailed information regarding the factors' scale, their role, and the areal exclusion thresholds are provided in Tables 2 and 3. Constraining threshold values were adopted from similar works such as Şener et al. (2010) and Abd-El Monsef and Smith (2019). The final product of areal constraining would give us the proper place to extract training and validation points for probabilistic modeling. By doing so, two sets of 1000 random points were selected each respectively per training the Mahalanobis distance model and validating the results. The number of samples was arbitrarily defined just to ensure the data scarcity issue and to be certain of the adequacy of samples both for training the model and validation of its results. The latter, in fact, guarantees that the balanced samples would satisfy the training and validation tasks since the number of samples are abundant which ease the data noise issue derived from inadequate sample number.

Most of the landfill site selection studies wind up in a mere spatial constraining; however, the product of such procedure, despite its substantial merits, is only a binary classification (i.e., suitable or unsuitable) at a rather rudimentary level since areal constraining follows an arbitrary factor classification scheme. Hence, probabilistic models can enable the decision-makers to avoid such determinism in identifying the most suitable sites and not only define the latter in probability terms but also give a chance to other areas with somewhat similar suitability degrees. Such areas can be classified as the transient zone which can locate in a flexible zone somewhere in between the most and the least suitable zones. Such a view also gives the authorities the opportunity to prioritize different options rather than relying only on two distinct and restricted zones. As opposed to deterministic models that only consider a single outcome, probabilistic models consider the geo-environmental random variables (here predictors) and their probability distributions to consider different possibilities and are very useful in predicting future events and patterns.

Given that landfill sites are subjected to different natural factors that act as random variables, probabilistic modeling pays the way to identify different locations with similar conditions as well as avoids determinism in site selection which, itself, can be problematic when dealing with such environmental agents. Such probabilistic modeling was carried out

Table 1 The VIF values representing the multicollinearity between different opted factors

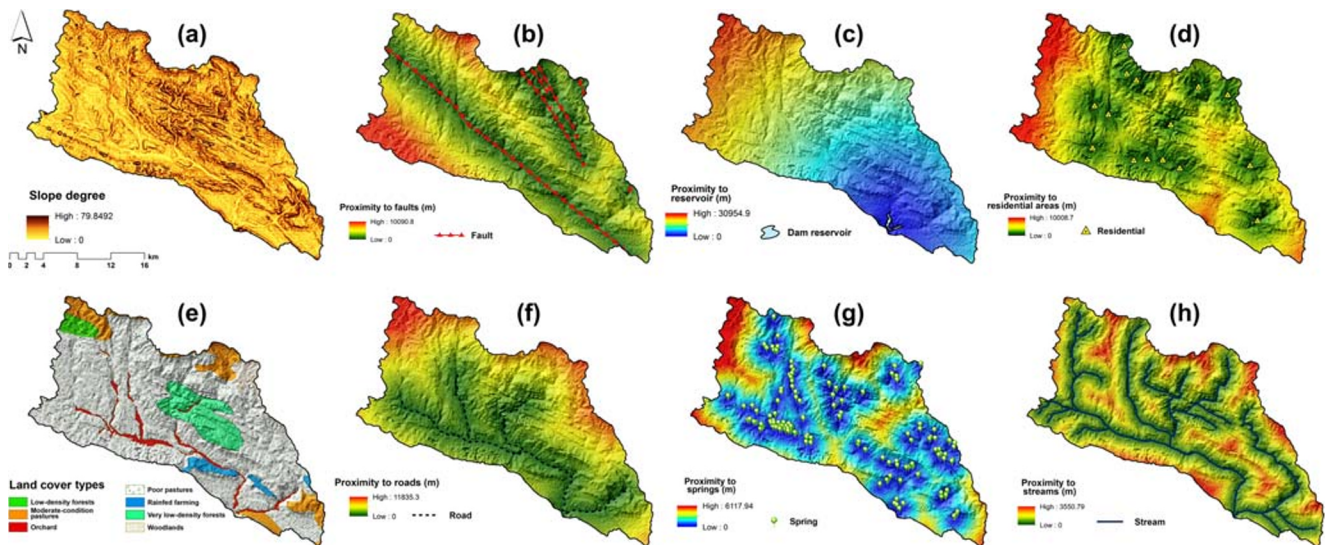| Factors	Collinearity statistics		Std. error
	Tolerance	VIF	
Proximity to faults (m)	0.624	1.603	0
Land cover	0.987	1.013	0.002
Proximity to residential areas (m)	0.337	2.969	0
Proximity to the reservoir (m)	0.445	2.245	0
Proximity to roads (m)	0.673	1.486	0
Slope angle (degree)	0.919	1.089	0

Fig. 4 Causative environmental factors used for areal constraining, **a** slope degree, **b** proximity to faults (m), **c** proximity to the reservoir (m), **d** proximity to residential areas (m), **e** land cover types, **f** proximity to roads (m), **g** proximity to springs (m), and **h** proximity to streams (m)

by incorporating the areal constraining scheme with a robust pattern recognition algorithm as explained in the following section. What follows is the detailed explanation of the proposed landfill suitability mapping framework.

Landfill site selection using the coupled MCDM-Mahalanobis distance

Determination of landfill sites merely based on areal constraining as a multi-criteria decision-making method would provide binary presence/absence locations at a rather rudimentary level which may manifest shortcomings in terms of the uncertainty emanated from the arbitrary nature of factor classification and combination techniques. Instead, feeding such localities into a powerful modeling scheme with the capability of pattern recognition would provide the final suitable sites in probabilistic terms which is in more favor of decision-makers (i.e., avoiding absolute certainty and determinism in their reports) and also has prediction merits. Such an

information retrieval process is attainable by the Mahalanobis distance approach. Mahalanobis distance was first expounded by Mahalanobis (Mahalanobis 1936) and ever since some application of this method has been reported such as data enrichment in landslide susceptibility modeling (e.g., Tsangaratos and Benardos 2014; Kornejady et al. 2018), outlier detection (Todeschini et al. 2013; Leys et al. 2018), and proposing novel validation techniques (Zhao et al. 2017). In a nutshell, it measures the similarity of different locations with different physical conditions in the study domain (i.e., different pixels with different environmental, geological, or anthropogenic factor values) to the target conditions (i.e., the conditions of the rudimental sites for landfill). In other words, locations (i.e., pixels) with the similar values of conditioning factors to those of landfill suitable sites would be equivalently considered suitable. The notion of suitability in Mahalanobis distance is translated into the average condition. For instance, if the average value of slope degrees at the rudimental landfill sites is equal to 5°, the latter would then signify the suitable

Table 2 Role of the opted factors in landfill site selection, their spatial scale, and established exclusion zones

Factors	Role	Scale	Exclusion zones
Proximity to faults (m)	Leakage pathways	1:50,000	500 m
Land cover	Environmental compatibility; esthetic aspects	1:100,000	See Table 2
Proximity to residential areas (m)	Avoiding diseases and unpleasant odor; esthetic aspects	1:50,000	5 km
Proximity to the reservoir (m)	Protecting natural and human habitats (environmental, social, and political aspects)	1:50,000	5 km
Proximity to roads (m)	Clean transportation routes	1:50,000	2 km
Slope angle (degree)	Implementation feasibility	1:50,000	Higher than 10 °
Proximity to springs (m)	Protecting groundwater resources	1:50,000	2 km
Proximity to streams (m)	Protecting surface water resources	1:50,000	1 km

Table 3 Assigned rates to different land covers based on their compatibility with a landfill site (adopted from Şener et al. 2010)

Land cover	Rate
Rainfed farming	7
Poor rangeland	6
Moderate-condition rangeland	5
Woodlands	4
Orchard	3
Very low-density forests	2
Low-density forests	1

slope for landfill sites. The same applies to the other environmental factors with the difference that qualitative factors such as land cover should be first quantified in order to become mathematically analyzable. Moreover, the notion of similarity to the ideal locations is Mahalanobis distance is rendered into a distance in which the closer distances to the ideal location indicates a higher probability of the overall environmental suitability. Such distance follows the expression (Fukunaga 1990):

$$Dm = \sqrt{(x-m)^T \Sigma^{-1} (x-m)} \qquad (1)$$

where x is the vector of conditioning factors, m is the factors' mean value, T is the transposed matrix of the factors' values, and $^{-1}$ is the inverse covariance matrix. In order to transform the distance values into probabilities, chi-square values of Mahalanobis distance values will then be transformed into P values based on the chi-Square distribution with $N-1$ degrees of freedom (number of variables 1). The latter was carried out in *Land Facet Corridor Tools* extension in ArcGIS 10.3.

Performance metrics

The area under the success and prediction rate curves (AUPRC and AUSRC), as probability cutoff-independent metrics, will be the judge of the model's performance in terms of goodness of fit and prediction power, respectively. The success rate curve plots the cumulative percentage of suitable areas (from the highest values to the lowest) on the X-axis against the cumulative percentage of the training set on the Y-axis (Chung and Fabbri 1999; Kornejady et al. 2018). The same applies to the prediction rate curve with the difference that, instead of including the training set, the validation dataset is used. The AUSRC and AUPRC values close to 100 (i.e., 1 in the range of 0–1) indicate a perfect model (i.e., highly trained and highly predictive), while values close to 50 (i.e., 0.5 in the range of 0–1) signify a neutrally operating model with an unsatisfying performance. Plotting the SRC and PRC was executed in the recently developed performance analyses toolbox in ArcGIS, named PMT (Rahmati et al. 2019). It is noteworthy that the presence-absence points were balanced

(i.e., an equal number of inventory and noninventory locations) in order to calculate the performance metrics. The balanced sample partitioning technique was selected due mainly to avoid the variance inflation in the results (Petschko et al. 2014).

Additionally, three holdout-dependent performance assessment metrics including accuracy (efficiency), precision, and sensitivity was calculated both in the training and validation stages following Eqs. 2–4 (Rahmati et al. 2019).

$$Accuracy = \frac{TP + TN}{T} \qquad (2)$$

$$Precision = \frac{TP}{TP + FP} \qquad (3)$$

$$Sensitivity = \frac{TP}{TP + FN} \qquad (4)$$

where TP, TN, FP, and FN, respectively, denote true positives, true negatives, false positives, and false negatives.

Results and discussion

Table 1 provides in detail the multi-collinearity values between the controlling factors of landfilling sites based on which the highest VIF values correspond to proximity to residential areas, proximity to springs, and proximity to the reservoir, while the lowest values are mainly attributed to land cover and slope angle. The results indicated that there is no critical multi-collinearity between the factors since the VIF values are in an acceptable range (i.e., less than 5). The latter is due to the fact that the authors avoided involving different DEM-derivative indices that have been commonly used in spatial modeling, but rather factors that can decisively delineate the landfilling sites are engaged. Among these, gentle slopes are the first opted criterion for most landfill sites and other criteria are sieved afterward. Faults are also critical gateways in such a way that contaminated water from landfills can percolate into the deep strata and contaminate the groundwater resources. Hence, landfill sites should be simply far from the faults. Dam reservoir is also considered a critical source of water supply for the residential centers, particularly Mashhad, and agricultural uses for arable lands. Moreover, the inhabitants of residential areas in the region would be bothered by the unpleasant odor of landfills as well as other imminent diseases. Regarding the land covers, most national environmental laws are strictly against siting landfills accordingly in dense forests, woodlands, good-condition pastures, and orchards, while nonarable lands and poor pastures are more desirable for such a purpose (Department of Environment 2018) (Table 3). Although roads are necessary for transporting waste to landfilling sites, considering setback margins is necessary particularly for the local residents

commuting from adjacent areas. The presence of springs indicates high groundwater potential in the study area. Hence, the farther the landfills from the springs, the better. Such areal filtering resulted in narrowing down that study area to the final extent which is suitable for siting landfills. As presented in Fig. 5, suitable sites identified at the rudimentary level are mostly concentrated in westernmost parts of the area while some sporadic spots are evident in the north, center, and south which altogether account for about 2% of the study area (i.e., 915.12 ha).

These sites, however, resulted from arbitrary classifications, and one cannot decisively introduce the final areas as the most suitable sites. Additionally, such a constraining process would signify the remaining area as unsuitable due to the binary format of the final map, while areas with lower suitability yet somewhat similar to the ideal sites would be left undiscovered. Moreover, even the sites identified as ideal by areal constraining need further investigation and cannot be determinately selected for pragmatic actions. Such an issue was obviated by the context of Mahalanobis distance. Extracting training and validation points (Fig. 6) and feeding them, respectively, into the modeling process and model validation created a probabilistic map for siting landfilling areas. The mean values, correlation matrix, and covariance matrix are, respectively, presented in Tables 4, 5, and 6 which together made the main components of the Mahalanobis distance. The mean values revealed that preselected suitable landfilling sites are located at an average distance of 3.7 km from the faults, 4.3 km from residential areas, 16 km from the reservoir,

4.4 km from roads, 2.8 km from springs, 1.7 km from streams, and a slope degree of 7.7°. Covariance and correlation matrices give valuable information regarding the dependency between the factors. What sets them apart is that in addition to the covariance matrix that indicates the direction of the linear correlation between factors, the correlation signifies the strength and direction of such a relationship in a limited range of + 1 and − 1. The correlation value of 0.71 between the proximity to springs and residential centers is also in accordance with the VIF results and the positive sign of the values also signifies the positive direction of their correlation. Mahalanobis distance uses such spatial dependence and the interconnection between the factors instead of considering them completely independent.

Landfill suitability map derived from Mahalanobis distance is presented in Fig. 7. The excerpt area shows that based on the probabilistic pattern recognition algorithm of Mahalanobis distance, the target regions derived from areal constraining do not necessarily pinpoint the high suitability but spatially varying suitability pattern may exist within the final constraints. More interestingly, a suitable patch is also evident in the bottom right corner of the excerpt which has been previously missed by areal constraining. The latter indicates that the rigidity of areal constraining may ignore spatial similarities due to lacking a perfect condition (i.e., being entitled to all the constraining rules), while a flexible algorithm acknowledges the random variation. In other words, if fully deterministic analysis (i.e., areal constraining) is taken into account, considering the same constraining rules on environmental

Fig. 5 Final delimited area derived from the stepwise areal constraining procedure
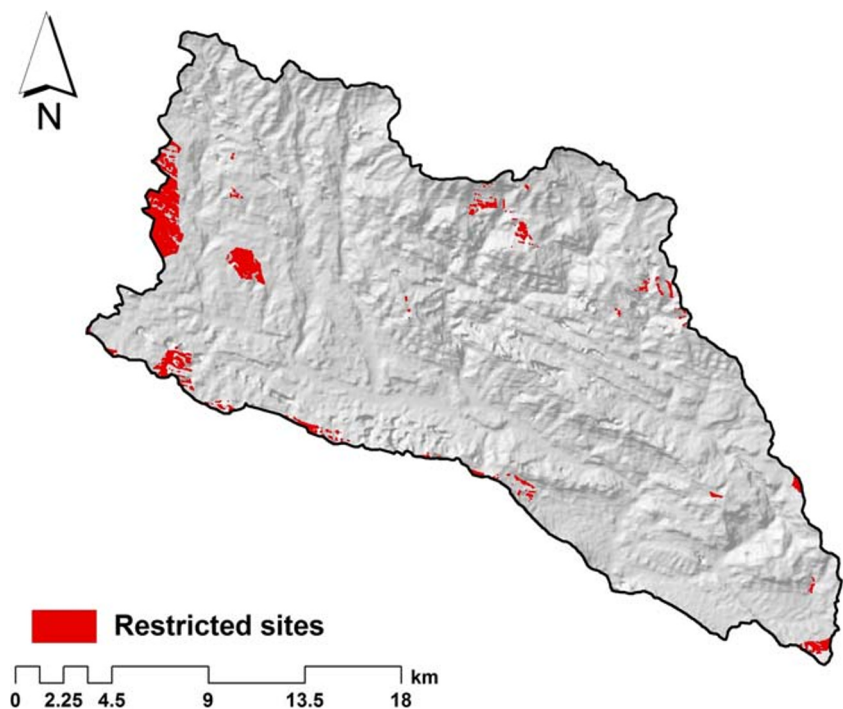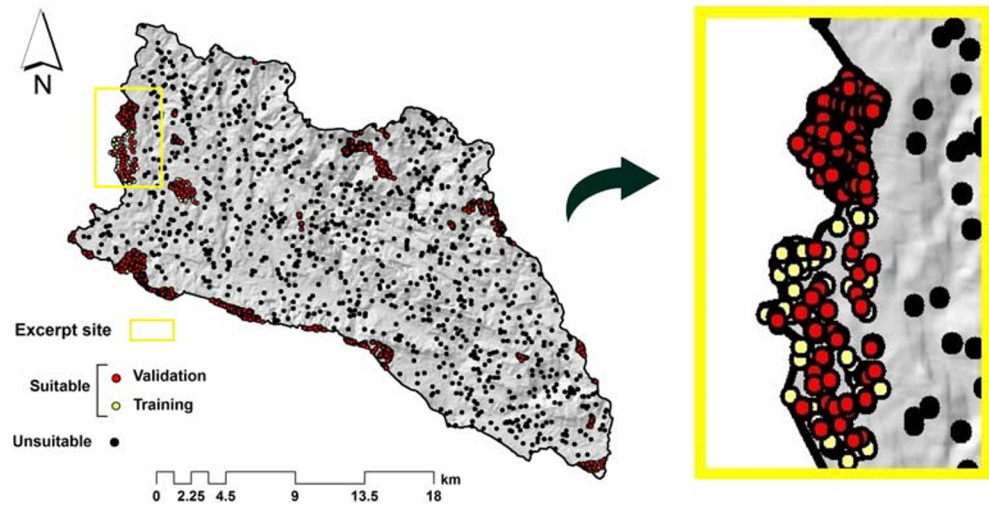
Fig. 6 Extracted training and validation (presence) samples from the constrained domain as well as the noninventory (absence) samples from the remaining area

factors in different areas would always give the same results. However, factors' randomness is highly appraised through many simulations which are mostly based on the distribution of each environmental factor. Since environmental factors often exhibit random variation, considering such matter would have a major influence on final results.

The coupled Mahalanobis distance and areal constraining techniques were found to have an AUSRC value of 89.36 (Fig. 8a). According to a classification range proposed by Hosmer Jr et al. (2008), the resulted value indicates that our ensemble model has almost shown an outstanding training performance (i.e., close to 0.9). Interestingly, the predictive power of the model is also very close to an excellent performance where the AUPRC value amounted to 89.25 (Fig. 8b). The latter highlights the merits of probabilistic spatial modeling especially when spatial dependencies between the covariates are taken into account. The highly identical values of AUSRC and AUPRC suggests two important features: (1) a model that has obviated the overfitting issue since the higher value of goodness of fit has not backfired into yielding a lower

predictive power, (2) a robust model where changes in the inputs do not much influence the model's performance, and (3) an inadequate segregation between the training and validation datasets due mainly to selection of the input data from a very small domain (i.e., 2% of the entire study area). The latter can be further assessed by spatially differentiating the training and validation dataset so that the former is selected from a separated area and fed into the model while another region is kept apart for validating the previous results.

In addition to the ROC curve, the holdout-dependent performance metrics were calculated (Table 7) based on which, in line with the results of ROC curve, the goodness of fit and prediction performances of the proposed suitability assessment framework found to be good which remains almost intact in both training and validation stages as a criterion of the robustness and reliability of model's results through the modeling and prediction stages.

The literature review showed a discernable pattern of adopting areal constraining as the basis of landfill suitability zonation. As such, Simsek et al. (2014a, b), Manzo et al. (2016), Mahmood et al. (2017), Rahmat et al. (2017a, b), Chabuk et al. (2017), Goulart Coelho et al. (2017), Yousefi et al. (2018), Spigolon et al. (2018), Khan et al. (2018), Şener et al. (2010), Kamdar et al. (2019), Chabuk et al. (2019), Özkan et al. (2019), Richter et al. (2019), and Akintorinwa and Okoro (2019) although attempted to incorporate the most novel methods to map suitable landfill sites, ultimately ended up in preparing a binary classification. As opposed to the traditional binary classification (i.e., suitable or unsuitable) derived from areal constraining methods, the probabilistic maps are highly differentiable in such way different transient zones can be located in between the suitable and unsuitable classes. Here, we chose the natural break classification scheme and classified the Mahalanobis distance-based landfill

Table 4 Factor's mean values at the training samples

Factors	Mean values
Proximity to faults (m)	3693.824
Land cover	4.013
Proximity to residential areas (m)	4290.372
Proximity to the reservoir (m)	15,913.415
Proximity to roads (m)	4386.635
Slope angle (degree)	7.699
Proximity to springs (m)	2766.084
Proximity to streams (m)	1712.438

Table 5 Correlation matrix representing the strength and direction of the normalized dependence between the factors

Factors	Proximity to faults (m)	Land cover	Proximity to residential areas (m)	Proximity to the reservoir (m)	Proximity to roads (m)	Slope angle (degree)	Proximity to springs (m)	Proximity to streams (m)
Proximity to faults (m)	1	−0.044	0.266	0.324	−0.405	−0.113	−0.105	−0.046
Land cover	−0.044	1	−0.017	−0.147	−0.137	−0.050	0.028	0.061
Proximity to residential areas (m)	0.266	−0.017	1	0.662	0.246	−0.089	0.716	−0.364
Proximity to the reservoir (m)	0.324	−0.147	0.662	1	0.169	0.001	0.529	−0.501
Proximity to roads (m)	−0.405	−0.137	0.246	0.169	1	−0.020	0.400	−0.329
Slope angle (degree)	−0.113	−0.050	−0.089	0.001	−0.020	1	−0.063	−0.004
Proximity to springs (m)	−0.105	0.028	0.716	0.529	0.400	−0.063	1	−0.180
Proximity to streams (m)	−0.046	0.061	−0.364	−0.501	−0.329	−0.004	−0.180	1

suitability map into two common classes of five-point and three-point scales (Fig. 9). A higher number of classes will, in turn, result in a reduced areal extent of each class. Selecting either of these two classifications would also result in different areal delineation, and hence different budget and time allocated to the area. Highly differentiated maps may be more in favor of on-site pragmatic actions, while less differentiated ones are of more general use in less detailed projects and less urgent circumstances. The areal extent of the suitability classes obtained from the five-point scale and three-point scale classification of the Mahalanobis distance map are also provided in Tables 8 and 9.

Limitations and future works

The difference in spatial resolution of the thematic maps is a basic yet critical limitation of any spatial modeling work. This issue roots mainly from data scarcity especially where high-resolution maps are not provided due to being an exorbitant task for organizations. In such a predicament, the best getaway to govern the available data is in favor of the detailed data (i.e., here DEM-derivatives as the limiting

Table 6 Covariance matrix representing the direction of the un-normalized dependence between the factors

Factors	Proximity to faults (m)	Land cover	Proximity to residential areas (m)	Proximity to the reservoir (m)	Proximity to roads (m)	Slope angle (degree)	Proximity to springs (m)	Proximity to streams (m)
Proximity to faults (m)	9,154,353.762	−17.229	1,424,783.336	6,488,089.622	−2,287,417.083	−775.262	−260,428.706	−63,703.508
Land cover	−17.229	0.017	−4.029	−127.698	−33.465	−0.015	2.951	3.685
Proximity to residential areas (m)	1,424,783.336	−4.029	3,132,301.608	7,764,991.980	811,490.761	−355.317	1,039,104.529	−296,954.297
Proximity to the reservoir (m)	6,488,089.622	−127.698	7,764,991.980	43,906,564.038	2,089,744.910	12.886	2,876,997.871	−1,531,331.870
Proximity to roads (m)	−2,287,417.083	−33.465	811,490.761	2,089,744.910	3,485,930.227	−82.962	612,262.213	−283,290.288
Slope angle (degree)	−775.262	−0.015	−355.317	12.886	−82.962	5.123	−116.915	−4.356
Proximity to springs (m)	−260,428.706	2.951	1,039,104.529	2,876,997.871	612,262.213	−116.915	672,866.413	−67,912.021
Proximity to streams (m)	−63,703.508	3.685	−296,954.297	−1,531,331.870	−283,290.288	−4.356	−67,912.021	212,529.544

Fig. 7 Landfill suitability probability map obtained from the ensemble of stepwise areal constraining and Mahalanobis distance models
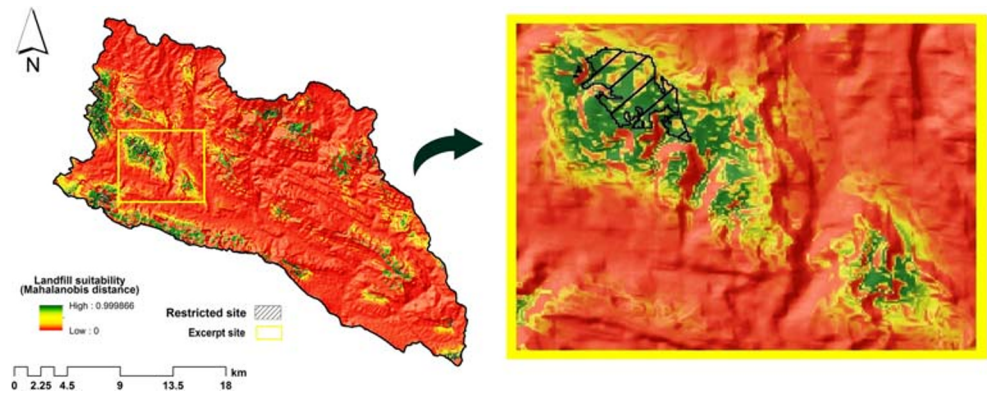
factors) rather than calculation space. Hence, all the spatial resolutions were transformed into the resolution of DEM-derived factors which, in this way, guarantees that no data have been lost. The other issue regards the sieving process and selection of the most influential factors. In this regard, soil permeability and wind velocity (i.e., speed and direction) are also among the most determinant factors. However, the former would require a detailed map consisting of soil physical properties including permeability yet such detailed data were not available for our study area. Although the Harmonized World Soil Database (FAO I, ISRIC I 2012) provides valuable information on soil properties on a global scale, using such data at a watershed scale would not be technically applicable. Hence, we avoided engaging less detailed data in the modeling process alongside other factors. Concerning wind velocity, it is noteworthy that, despite its invaluable information in identifying the exposed area to odor nuisance, generating the wind velocity map would require the synoptic data, while there is no synoptic station in the study area. Additionally, we strived to fill this gap by engaging another complementary factor, distance to residential areas, as potential sites were ultimately proposed to be located in the farthest areas possible from the residents. Nonetheless, the main objective of this study is to propose a globally

generalizable modeling framework for landfill suitability assessment and certainly engaging more factors with unique information would improve the final performance of the proposed model. As stated in different literature (e.g., Allen et al. 1997; Allen et al. 2003; Simsek et al. 2014a, b; Barakat et al. 2017; Rahmat et al. 2017; Alkhuzaie and Janna 2018; Deswal and Laura 2018), abovementioned maps are pivotal for landfill suitability assessment and should be used in future analyses.

The last issue corresponds to the validation stage of the model's results which stems from the fact that there might not be a precise and generalizable concept of a suitable landfill site in the first place. In other words, validation of the landfill suitability map can be carried out by the actual landfill sites that have been proved to be perfectly selected in terms of geo-topological criteria and residents' qualification metrics. However, we presented a pseudo-validation test in which validation points are selected from the sites that are the most ideal sites for landfills. For future studies, it is suggested to compare the proposed technique with other machine learning and deep learning techniques with flexible probabilistic themes in terms of their performance and robustness. Although almost intact values of performance metrics through the training and validation stages somewhat indicate the stability of the model, further

Fig. 8 Success rate curve (**a**) and prediction rate curve (**b**) generated in the PMT toolbox
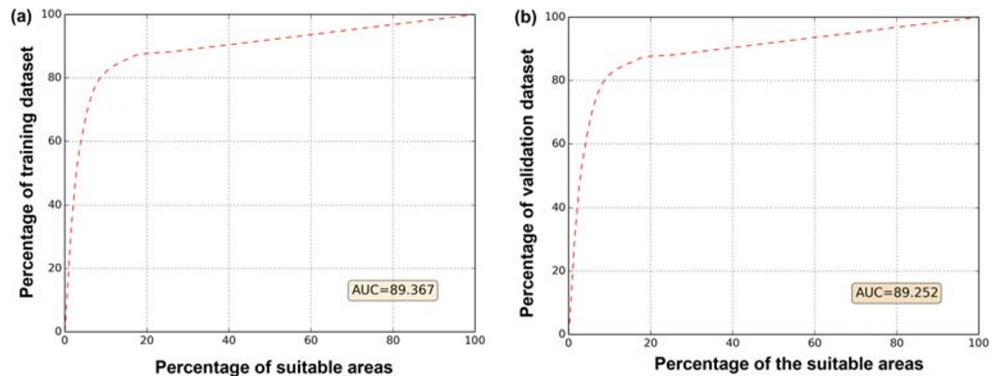
Table 7 Accuracy, precision, and sensitivity values calculated for the Mahalanobis distance model at the training and validation stages

Holdout-dependent performance metrics	Training stage	Validation stage
Accuracy	0.699	0.710
Precision	0.766	0.759
Sensitivity	0.998	0.997

analyses need to be conducted to assess the overall robustness which can be ascertained by *k*-fold cross-validation techniques.

Conclusion

Uncontrolled landfills and open garbage dumps have been significantly contributing to surface and groundwater resource pollution and become a major concern for municipalities of the Ardak watershed and adjacent cities. The watershed exhibits high environmental and socio-economic values. The take-home messages of this work are as follows. Despite the merits of areal constraining techniques in narrowing down the domain to delineate

the optimal landfill sites, it is underpinned by determinism while a probabilistic model is required to consider both the inherent randomness within the factors' nature and spatial dependence between them. Mahalanobis distance supports such objectives and based on our results, gives a satisfying performance. The latter is apparent from the high goodness of fit and prediction power of the adopted model. Coupling Mahalanobis distance to areal constraining builds a powerful inventory-irrespective ensemble model. Areal constraining showed that almost 98% of the study area is unsuitable for landfilling, while the suitability map derived from the coupled areal constraining and Mahalanobis distance avoids such a rigid and rather inflexible response and appoints the end user for such a deduction. In other words, depending on the differentiation degree of interest,
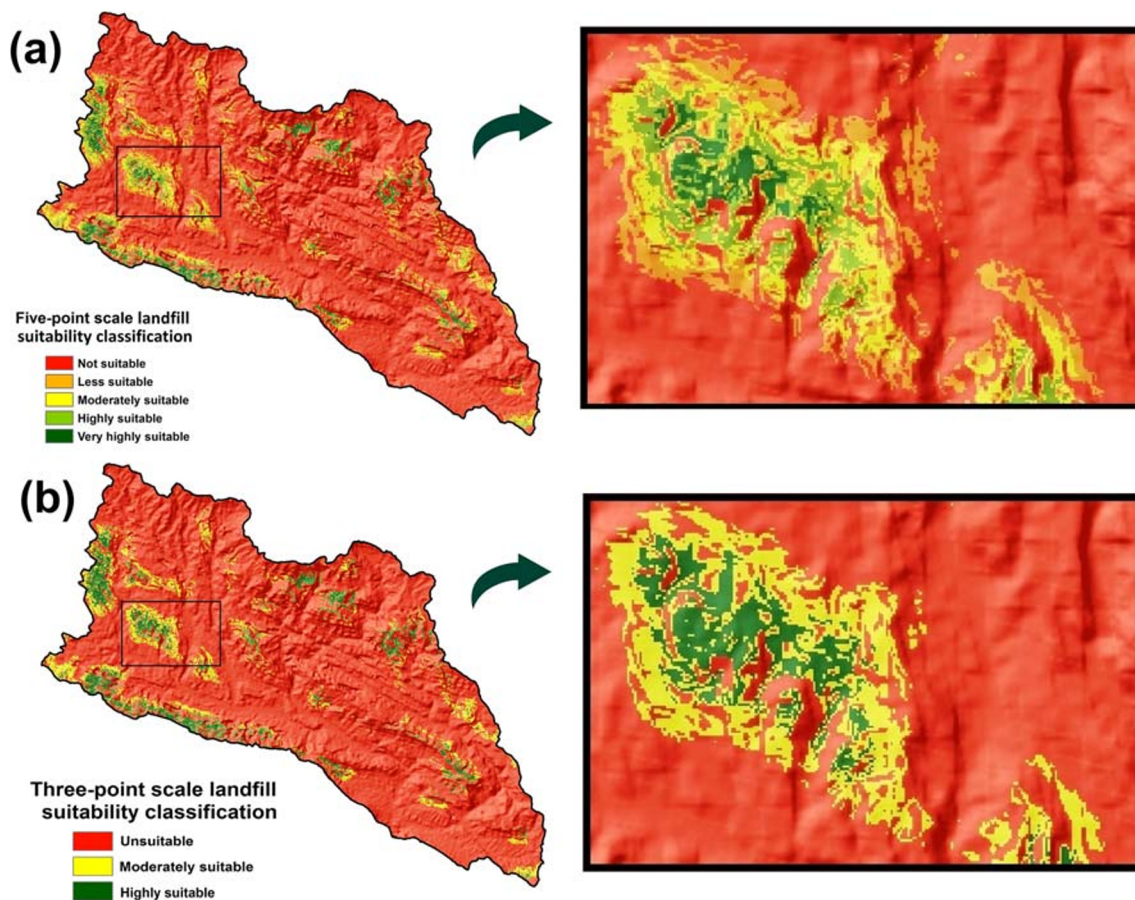

Fig. 9 Five-point scale (**a**) and three-point scale (**b**) classification of landfill suitability

Table 8 The areal extent of the suitability classes obtained from a five-point scale classification of Mahalanobis distance map

Landfilling suitability classes	Area (%)
Very low	83.848
Low	6.831
Moderate	3.919
High	2.793
Very high	2.609

different suitability percentages can be obtained and used for budget allocation. Higher differentiation (i.e., a five-point scale classification) would normally result in more strict areal concentration based on which 2.6% of the study area has very high suitability for landfilling, while 2.7% of the entire area is highly suitable. Additionally, a less strict classification scheme (i.e., three-point scale) suggests that up to 4.6% of the entire region can be highly suitable for landfilling, while the remaining area has low-to-moderate suitability. By using the coupled Mahalanobis distance and stepwise areal constraining method, we have provided a framework of landfill management for generating advice to managers and policymakers concerned with downstream water quality. The conventional suitability assessment techniques provide the authorities with a binary classification which is limited by a mere presence-absence scheme and accordingly distorts the reality of the area. Conversely, the proposed technique provides a flexible map that can be classified into different suitability ranges according to the priority of decision-makers. In such a way, different areas will be treated relying on their true potentials while no place is left alone or forced to be categorized into only two specific classes.

Table 9 The areal extent of the suitability classes obtained from a three-point scale classification of Mahalanobis distance map

Landfilling suitability classes	Area (%)
Unsuitable	88.230
Moderately suitable	7.205
Highly suitable	4.565

✓ Landfill siting using areal constraining results in a mere binary classification.

✓ Mahalanobis distance considers spatial dependencies of environmental factors.

✓ Ensemble of two abovementioned methods provides an inventory-irrespective analysis.

References

Abd-El Monsef H, Smith SE (2019) Integrating remote sensing, geographic information system, and analytical hierarchy process for hazardous waste landfill site selection. Arab J Geosci 12(5):155

Akintorinwa OJ, Okoro OV (2019) Combine electrical resistivity method and multi-criteria GIS-based modeling for landfill site selection in the southwestern Nigeria. Environ Earth Sci 78(5):162

Alkhuzaie MM, Janna H (2018) Optimum location for landfills sites based on GIS modeling for Al-Diwaniyah City, Iraq. International Journal of Civil Engineering and Technology 9(8):941–951

Allen AR, Dillon A, O'Brien M (1997) Approaches to landfill site selection in Ireland. Engineering geology and the Environment 2:1569–1574

Allen BG, Caetano P, Costa C, Cummins V, Donnelly J, Koukoulas S, Vendas D (2003) A landfill site selection process incorporating GIS modelling. In Proceedings of Sardinia 2003, Ninth International Waste Management and Landfill Symposium, S. Margherita di Pula, Cagliari

Barakat A, Hilali A, El Baghdadi M, Touhami F (2017) Landfill site selection with GIS-based multi-criteria evaluation technique. A case study in Béni Mellal-Khouribga Region, Morocco. Environmental Earth Sciences 76(12):413

Chabuk AJ, Al-Ansari N, Hussain HM, Knutsson S, Pusch R (2017) GIS-based assessment of combined AHP and SAW methods for selecting suitable sites for landfill in Al-Musayiab Qadhaa, Babylon, Iraq. Environ Earth Sci 76(5):209

Chabuk A, Al-Ansari N, Hussain HM, Laue J, Hazim A, Knutsson S, Pusch R (2019) Landfill sites selection using MCDM and comparing method of change detection for Babylon Governorate, Iraq Environ Sci Pollut Res 1–15

Chen W, Pourghasemi HR, Panahi M, Kornejady A, Wang J, Xie X, Cao S (2017) Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. Geomorphology 297:69–85

Chung CJF, Fabbri AG (1999) Probabilistic prediction models for landslide hazard mapping. Photogramm Eng Remote Sens 65(12):1389–1399

del Valle-Zermeño R, Romero-Güiza MS, Chimenos JM, Formosa J, Mata-Alvarez J, Astals S (2015) Biogas upgrading using MSWI bottom ash: an integrated municipal solid waste management. Renew Energy 80:184–189

Department of Environment (2018) Article 2 and Article 5 of the Executive Regulation on Waste Management

Deswal M, Laura JS (2018) GIS based modeling using analytic hierarchy process (AHP) for optimization of landfill site selection of Rohtak city, Haryana (India). Journal of Applied and Natural Science 10(2):633–642

FAO I, ISRIC I (2012) JRC: harmonized world soil database (version 1.2). FAO, Rome, Italy and IIASA, Laxenburg, Austria

Fukunaga K (1990) Introduction to statistical pattern recognition. Academic Press, Boston

Goulart Coelho LM, Lange LC, Coelho HM (2017) Multi-criteria decision making to support waste management: a critical review of current practices and methods. Waste Manag Res 35(1):3–28

Haghizadeh A, Moghaddam DD, Pourghasemi HR (2017) GIS-based bivariate statistical techniques for groundwater potential analysis (an example of Iran). J Earth Syst Sci 126(8):109

Hosmer DW Jr, Lemeshow S, May S (2008) Applied survival analysis: regression modeling of time-to-event data (Vol. 618). Wiley-InterScience

Kamdar I, Ali S, Bennui A, Techato K, Jutidamrongphan W (2019) Municipal solid waste landfill siting using an integrated GIS-AHP

approach: a case study from Songkhla, Thailand. Resour Conserv Recycl 149:220–235

Khan MMUH, Vaezi M, Kumar A (2018) Optimal siting of solid waste-to-value-added facilities through a GIS-based assessment. Sci Total Environ 610:1065–1075

Khayamabshi E (2016) Current status of waste management in Iran and business opportunities. 5th Green Industry Conference for Sustainable Development, Waste Management on Occasion of Smart Engineering, Tokyo

Khorram A, Yousefi M, Alavi SA, Farsi J (2015) Convenient landfill site selection by using fuzzy logic and geographic information systems: a case study in Bardaskan, East of Iran. Population 6:10

Kornejady A, Ownegh M, Rahmati O, Bahremand A (2018) Landslide susceptibility assessment using three bivariate models considering the new topo-hydrological factor: HAND. Geocarto International 33(11):1155–1185

Kutner MH, Nachtsheim CJ, Neter J (2004) Simultaneous inferences and other topics in regression analysis. Applied linear regression models, 4th edn. McGraw-Hill Irwin, New York, pp 168–170

Leys C, Klein O, Dominicy Y, Ley C (2018) Detecting multivariate outliers: use a robust variant of the Mahalanobis distance. J Exp Soc Psychol 74:150–156

Li'ao W, Ting'quan P, Chuan H, Hui Y (2009) Management of municipal solid waste in the Three Gorges region. Waste Manag 29(7):2203–2208

Mahalanobis PC (1936) On the generalized distance in statistics. Proceedings of the National Institute of Sciences (Calcutta) 2:49–55

Mahmood K, Batool A, Faizi F, Chaudhry MN, Ul-Haq Z, Rana AD, Tariq S (2017) Bio-thermal effects of open dumps on surroundings detected by remote sensing—influence of geographical conditions. Ecol Indic 82:131–142

Manzo C, Mei A, Fontinovo G, Allegrini A, Bassani C (2016) Integrated remote sensing for multi-temporal analysis of anthropic activities in the south-east of Mt. Vesuvius National Park. Journal of African Earth Sciences 122:63–78

Moghaddam DD, Rahmati O, Haghizadeh A, Kalantari Z (2020a) A modeling comparison of groundwater potential mapping in a mountain bedrock aquifer: QUEST, GARP, and RF models. Water 12:679

Moghaddam DD, Rahmati O, Panahi M, Tiefenbacher J, Darabi H, Haghizadeh A, ... Bui DT (2020b) The effect of sample size on different machine learning models for groundwater potential mapping in mountain bedrock aquifers. Catena 187:104421

Mondelli G, Giacheti HL, Boscov MEG, Elis VR, Hamada J (2007) Geoenvironmental site investigation using different techniques in a municipal solid waste disposal site in Brazil. Environ Geol 52(5):871–887

Ngoc UN, Schnitzer H (2009) Sustainable solutions for solid waste management in Southeast Asian countries. Waste Manag 29(6):1982–1995

O'brien RM (2007) A caution regarding rules of thumb for variance inflation factors. Qual Quant 41(5):673–690

Özkan B, Özceylan E, Sarıçiçek İ (2019) GIS-based MCDM modeling for landfill site suitability analysis: a comprehensive review of the literature. Environ Sci Pollut Res 1–20

Petschko H, Brenning A, Bell R, Goetz J, Glade T (2014) Assessing the quality of landslide susceptibility maps–case study Lower Austria. Natural Hazards & Earth System Sciences 14(1):95–118

Rahmat ZG, Niri MV, Alavi N, Goudarzi G, Babaei AA, Baboli Z, Hosseinzadeh M (2017) Landfill site selection using GIS and AHP: a case study: Behbahan, Iran. KSCE J Civ Eng 21(1):111–118

Rahmati O, Kornejady A, Samadi M, Nobre AD, Melesse AM (2018) Development of an automated GIS tool for reproducing the HAND terrain model. Environ Modell Softw 102:1–12

Rahmati O, Kornejady A, Samadi M, Deo RC, Conoscenti C, Lombardo L, Dayal K, Taghizadeh-Mehrjardi R, Pourghasemi HR, Kumar S, Bui DT (2019) PMT: new analytical framework for automated evaluation of geo-environmental modelling approaches. Sci Total Environ 664:296–311

Rathi S (2006) Alternative approaches for better municipal solid waste management in Mumbai, India. Waste Manag 26(10):1192–1200

Regional Water Company of Khorasan Razavi [RWCKR] (2018) Detailed Action Plan of Ardak Watershed: quantitative and qualitative assessment of water resources

Richter A, Ng KTW, Fallah B (2019) Bibliometric and text mining approaches to evaluate landfill design standards. Scientometrics 118(3):1027–1049

Şener Ş, Şener E, Nas B, Karagüzel R (2010) Combining AHP with GIS for landfill site selection: a case study in the Lake Beyşehir catchment area (Konya, Turkey). Waste Manag 30(11):2037–2046

Simsek C, Elci A, Gunduz O, Taskin N (2014a) An improved landfill site screening procedure under NIMBY syndrome constraints. Landsc Urban Plan 132:1–15

Simsek C, Elci A, Gunduz O, Taskin N (2014b) An improved landfill site screening procedure under NIMBY syndrome constraints. Landsc Urban Plan 132:1–15

Spigolon LM, Giannotti M, Larocca AP, Russo MA, Souza NDC (2018) Landfill siting based on optimisation, multiple decision analysis, and geographic information system analyses. Waste Manag Res 36(7):606–615

Tehran Times (2017) Waste management has come to a head. **December 11**, https://www.tehrantimes.com/news/419209/Waste-management-has-come-to-a-head

Todeschini R, Ballabio D, Consonni V, Sahigara F, Filzmoser P (2013) Locally centred Mahalanobis distance: a new distance measure with salient features towards outlier detection. Anal Chim Acta 787:1–9

Tsangaratos P, Benardos A (2014) Estimating landslide susceptibility through an artificial neural network classifier. Nat Hazards 74(3):1489–1516

Yousefi H, Javadzadeh Z, Noorollahi Y, Yousefi-Sahzabi A (2018) Landfill site selection using a multi-criteria decision-making method: a case study of the Salafcheghan special economic zone, Iran. Sustainability 10(4):1107

Zhao L, Lu Z, Yun W, Wang W (2017) Validation metric based on Mahalanobis distance for models with multiple correlated responses. Reliability Engineering & System Safety 159:80–89