The effect of sample size on different machine learning models for groundwater potential mapping in mountain bedrock aquifers

Davoud Davoudi Moghaddam[a], **Omid Rahmati**[b*], **Mahdi Panahi**[c,d], **John Tiefenbacher**[e], **Hamid Darabi**[f], **Ali Haghizadeh**[a], **Ali Torabi Haghighi**[f], **Omid Asadi Nalivan**[g], **Dieu Tien Bui**[h*]

[a] Department of Watershed Management, Agriculture and Natural Resources Faculty, Lorestan University, Iran
[b] Soil Conservation and Watershed Management Research Department, Kurdistan Agricultural and Natural Resources Research and Education Center, AREEO, Sanandaj, Iran
[c] Division of Science Education, Kangwon National University, Chuncheon-si, Gangwon-do 24341, Korea
[d] Geoscience Platform Research Division, Korea Institute of Geoscience and Mineral Resources (KIGAM), 124, Gwahak-ro Yuseong-gu, Daejeon 34132, Korea
[e] Department of Geography, Texas State University, San Marcos, TX 78666, USA
[f] Water, Energy and Environmental Engineering Research Unit, University of Oulu, Oulu, Finland
[g] Department of Watershed Management, Natural Resources Faculty, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran
[h] Institute of Research and Development, Duy Tan University, Da Nang 550000, Viet Nam

Corresponding authors' Email addresses: o.rahmati@areeo.ac.ir; buitiendieu@duytan.edu.vn

Abstract

Machine learning models have attracted much research attention for groundwater potential mapping. However, the accuracy of models for groundwater potential mapping is significantly influenced by sample size and this is still a challenge. This study evaluates the influence of sample size on the accuracy of different individual and hybrid models, adaptive neuro-fuzzy inference system (ANFIS), ANFIS-imperial competitive algorithm (ANFIS-ICA), alternating decision tree (ADT), and random forest (RF) to model groundwater potential, considering the number of springs from 177 to 714. A well-documented inventory of springs, as a natural representative of groundwater potential, was used to designate four sample data sets: 100% (D_1), 75% (D_2), 50% (D_3), and 25% (D_4) of the entire springs inventory. Each data set was randomly split into two groups of 30% (for training) and 70% (for validation). Fifteen diverse geo-environmental factors were employed as independent variables. The area under the operating receiver characteristic curve (AUROC) and the true skill statistic (TSS) as two cutoff-independent and cutoff-dependent performance metrics were used to assess the performance of models. Results showed that the sample size influenced the performance of four machine learning algorithms, but RF had a lower sensitivity to the reduction of sample size. In addition, validation results revealed that RF (AUROC=90.74–96.32%, TSS=0.79–0.85) had the best performance based on all four sample data sets, followed by ANFIS-ICA (AUROC=81.23–91.55%, TSS=0.74–0.81), ADT (AUROC=79.29–88.46%, TSS=0.59–0.74), and ANFIS (AUROC=73.11–88.43%, TSS=0.59–0.74). Further, the relative slope position, lithology, and distance from faults were the main spring-affecting factors contributing to groundwater potential modelling. This study can provide useful guidelines and valuable reference for selecting machine learning models when a complete spring inventory in a watershed is unavailable.

Keywords*:* Groundwater management; Geo-environmental factors; Sample size; Spatial modelling; Random Forest

1. Introduction

Long-lasting droughts and increasing rates consumption are threatening water, energy, and food security and they do not bode well for the future of arid and semi-arid environments. In many parts of the world, groundwater resources have been relatively inexpensive and often an abundant source of usable fresh water (Jackson et al., 2001; Chenini and Mammou, 2010; Parisi et al., 2018). Groundwater resources are less vulnerable to contamination than surface water sources in areas where adequate groundwater protection measures are implemented and, therefore, would be highly beneficial for public water supplies as well as for irrigation and industrial and domestic uses (Mukherjee et al., 2012; Choubin and Malekian, 2017; Shenga et al., 2018a). The Sustainable Development Goals (SDGs) of the United Nations stress this issue (Rasul, 2016; Velis et al., 2017). However, uncontrolled groundwater withdrawal has caused depletion of groundwater resources in many regions, especially in Iran (Motagh et al., 2008; Ravilious 2018). Overexploitation and severe droughts in recent years have caused increasing rates of groundwater withdrawals, because groundwater resources are viewed as dependable, especially where sustainable groundwater exploitation and protection measures have been implemented. Fundamental to achieving these measures is effective modeling of groundwater potential.

Fractured bedrock aquifers (also termed "mountain bedrock aquifers") are important sources of water in some parts of the world. These zones play a critical role in the hydrologic cycle as they serve as long-term stores of water and they initiate transport of water from the surface to local and regional aquifers (Viviroli et al., 2007). However, flow and transport processes in such aquifers are intrinsically complex, more than granular, porous-media aquifers (i.e., generally located in plains) due to the nature of their depths. They are discrete, exhibit anisotropy, have small pathways, and are less common (Coleman et al., 2015). The occurrence and movement of groundwater in fractured bedrock aquifers in a given area is very complicated and governed by many factors, such as lithology, landforms, topography, secondary porosity, geological structures, fracture density, aperture and connectivity, drainage pattern, groundwater recharge, groundwater table distribution, slope, land cover, climatic conditions, and their interrelationships (Levison et al., 2012; Rathay et al., 2018). In general, there is insufficient data regarding groundwater in fractured bedrock aquifers in mountainous environments worldwide due to a lack of piezometric wells in these high-elevation settings (Voeckler and Allen, 2012). Consequently, pumping tests are often not feasible. Therefore, water table data have not been gathered and a clear understanding of the hydrodynamic components of these areas have not been acquired.

Because of these challenges, numerical models cannot be created in mountainous areas to understand how groundwater moves in bedrock (Shenga et al., 2018b). Various types of machine-learning and data-mining models have been employed for groundwater potential modelling, however, and they include binary logistic regression (Ozdemir, 2011b), weights of evidence (Ozdemir, 2011a; Pourtaghi and Pourghasemi, 2014; Chen et al., 2018), frequency ratio (Oh et al., 2011; Manap et al., 2014), artificial neural networks (Lee et al., 2012, 2018), random forest (Naghibi et al., 2016; Rahmati et al., 2016; Zabihi et al., 2016; Naghibi et al., 2017b; Golkarian et al., 2018), support vector machine (Naghibi et al., 2017b), boosted regression trees (Mousavi et al., 2017; Kordestani et al., 2019), generalized linear and additive models (Falah et al., 2017), classification and regression trees (Naghibi et al., 2016; Choubin et al., 2019), multivariate adaptive regression spline (Zabihi et al., 2016; Golkarian et al., 2018), evidential belief function (Nampak et al., 2014; Pourghasemi and Beheshtirad, 2015), maximum entropy (Rahmati et al., 2016), decision trees (Lee and Lee, 2015; Naghibi et al., 2019), and logistic model tree (Rahmati et al., 2018). Recently, hybrid models and ensemble techniques have been proposed to improve the structure of data mining models (Naghibi et al., 2017a; Rahmati et al., 2018; Chen et al., 2019; Kordestani et al., 2019; Miraki et al., 2019; Naghibi et al., 2019). In these studies, inventories of springs have been used for dependent variables. The number of springs (i.e., sample size) is a very important factor in groundwater potential modelling. On the other hand, the greatest cost related to geospatial modelling of groundwater potential is the expense of collecting spring location data, given the significant resources and time requirements of field surveys and investigations (Hancock and Boulton, 2009; Kollat et al., 2011; Leach et al., 2016).

It has been known that gathering hydrological data in the field is a main challenge in developing countries. However, a systematic assessment of the effect of sample size (i.e. the number of springs used to model groundwater potential) is needed not only to evaluate groundwater potential, but also to evaluate the predictive capability of machine learning models. Owing to cost and time constraints in most countries, it is essential to identify robust machine learning models that are less sensitive to the sample size (i.e., the number of springs). Determining the dependence of a model's results on the sample size used is a research gap. Therefore, there are two focal questions of this research: "What is the response of machine-learning models to changing sample size?" and "How large must a spring data sample be to construct robust, high-resolution groundwater-potential maps?"

To answer these questions, four sample data sets were generated from a spring inventory set: data set 1 (D_1) (100% of the inventory), data set 2 (D_2) (75% of the inventory), data set 3 (D_3) (50% of the inventory), and data set 4 (D_4) (25% of the inventory). This sampling enabled examination of the effect of changing sample sizes on the predictive performance of machine learning models for groundwater potential mapping. In contrast to the above studies, the effects on the quality of predicted maps by changing the sample size was explored in terms of predictive performance. Another main difference between this study and others is that individual ANFIS, a hybrid model namely ANFIS-imperial competitive algorithm (ANFIS-ICA), and alternating decision tree (ADT) were used for groundwater potential mapping; their capabilities for groundwater-potential studies have not been gauged before. The RF algorithm was used as a benchmark model to compare to the other models. Selection of robust yet less data-demanding machine learning models, which do not require a large spring inventory for groundwater potential mapping, saves time and financial expenditures and enables more effective water resource management. The objectives of this study are to: 1) evaluate the influence of sample size on the performance of four machine learning models – RF, ANFIS, ANFIS-ICA, and ADT; 2) compare the predictive performances of these models by means of cutoff-dependent and cutoff-independent indices; and 3) assess the importance of geo-environmental spring-affecting factors in modelling. This study provides useful reference information regarding the capability and sensitivity of models to sample size for water resources and environmental modelling and for application of machine learning algorithms to groundwater potential mapping.

2. Study area

The Hableh-Roud basin extends from Tehran Province to Semnan Province in Iran. There are two morphologically distinct districts in the northern and southern parts of the basin. The current study is focused on the northern basin (northern Hableh-Roud) which is characterized by a mountainous morphology and a semi-humid, cold climate covering an area of approximately 5,203 km^2. Geographically, the basin lies between 35°19′ and 35°55′N and between 51°46′ and 53°08′E (Fig. 1). The elevation ranges from 968 to 4,036 m.a.s.l. The average annual precipitation amounts to 470 mm, and the average temperature is 11^{o}C. From a geological viewpoint, the region is in the Central Alborz structures and has a diverse lithology in which 38 rock units are evident. Rangelands, agricultural lands, forests, orchards, bare lands, and residential development are the primary land uses in the study area. The total population is 157,168 (86,947 reside in cities and 70,221 in rural areas). The region contains a notable number of springs, 83% of which are used for agriculture, 16% for potable water and for health, and a very small number are used by industries. The annual discharge of springs adds up to around 18 million cubic meters. The abundance of springs distributed over the region suggests that there is a considerable amount of groundwater in the region.

The host rocks have been either metamorphosed, deformed through tectonics, and/or is comprised of igneous intrusions. Consequently, the hydraulic conductivity in the study area is often controlled by a network of discrete fractures. A layer of thin surface deposits overlies the bedrock in most of the study area. These deposits are fluvial and colluvial sediments and they reach a maximum of 6.5 m thick. They were usually deformed by compression, erosion, and even uplift. The complexity of fracturing and the lithology can be observed at a range of scales (from lineaments to outcrops) (Table 1 and Fig. 2). Most of the rocks in the region belong to Ek (well bedded, green tuff and shale), Jl (Light grey, thin - bedded to massive limestone), OMql (Massive too thick-bedded limestone), and Plc (conglomerate and sandstone).

Fig. 1 here

Fig. 2 here

Table 1 here

3. Methodology

A methodology was conceptualized that involved several steps: data compilation, multicollinearity analysis, and spatial modelling (Fig. 3).

Fig. 3 here

3.1. Data compilation

Drawing on an extensive literature review, the most representative spring-affecting factors were chosen to model groundwater potential. The factors were selected based on a sieving process with transparent contribution to groundwater volume. Among these, topological, hydrological, lithological, environmental, and anthropological factors, as well as the ones with dual characteristics (e.g., topo-hydrological factors), are demonstrated contributors to groundwater potential modelling (e.g., Oh et al., 2011; Manap et al., 2013; Rahmati et al., 2016; Naghibi et al., 2019).

Based on data availability, 15 spring-affecting factors were chosen (Table 2): elevation, slope percentage, aspect, soil texture, land use, soil hydrological groups, lithological formation, topographic wetness index (TWI), relative slope position (RSP), plan and profile curvature, topographic position index (TPI), terrain ruggedness index (TRI), distance from faults, and drainage density (Fig. 4). These factors can be classified into a single category as each factor can be interpreted in a dual- or multilateral manner. Some factors, such as lithological formations, soil texture, formation, soil hydrological groups, and distance from faults with both geological and geo-hydrological fabrics, represent subsurface water infiltration and percolation that contribute to groundwater. Meanwhile, topological and topo-hydrological factors, as derivatives of the digital elevation model (DEM), primarily represent surface-runoff generation, accumulation, and transformation which, in conjunction with subsurface processes, can address groundwater potential. The identification of factors that contribute the most to the modelling of groundwater potential, the ways in which groundwater potential reacts to the changes of each factor, and the ways these factors interact will be evident from the modeling process and each will be discussed in detail below.

Table 2 here

Fig. 4 here

The locations of springs represents groundwater potential across the northern Hableh-Roud. Therefore, archival data, complemented with information from extensive field surveys that geolocated springs with a GPS unit, were used to document an inventory of spring locations. The inventory, a total of 714 springs in 2018, was mapped as point features in ArcGIS. To test the robustness of the models and the degree to which they are sensitive to the input data sample sizes, we created four data sets following a subordinate data sample reduction strategy: D_1 (100% of springs, n=714), D_2 (75% of springs, n=535), D_3 (50% of springs, n=358), and D_4 (25% of springs, n=177). Each data set was randomly split into two groups with a 70:30 (training:validation) ratio (Fig. 5).

Fig. 5 here

3.2. Multicollinearity analysis

To assess the multicollinearity of spring-affecting factors and to avoid bias, we used the variance inflation factor (VIF) and tolerance (TOL) indices as are customarily used to estimate multicollinearity of predictive factors in geospatial modelling (Bui et al., 2016). According to the literature (O'Brien 2007), a VIF >10 or tolerance <0.1 indicates a critical multicollinearity which itself suggests that the factor(s) has such strong correlation to others that it should be removed from modelling (Hair et al., 2009).

3.3. Spatial modelling of groundwater potential

3.3.1. Adaptive Neuro-Fuzzy Inference System (ANFIS)

Artificial neural networks are powerful memorizing machines but are not capable of generalizing and predicting patterns (Liška et al., 2018). The learning stage is prone to overfitting, since the algorithm is designed to find the best solution in the shortest time and, therefore, the function becomes trapped in a suboptimal equilibrium point called local minimum and is unable to reach to the global minimum (i.e., the best solution) (Jang et al., 1997). Hence, the fuzzy inference system was fused with ANNs to compensate for their prediction deficiencies. More concisely, the Takagi–Sugeno fuzzy inference system updates the information on the phenomenon under study by perpetually setting new rules (termed "fuzzy if-then rules") (Bui et al., 2018; Zare and Koch, 2018). As information rises through the updating procedure, the learned linear and nonlinear parameters are optimized using the gradient descent and recursive least-square algorithms (Premkumar and Manikandan, 2014). More details of this process can be found in Jang (1991) and Jang (1993).

3.3.2. ANFIS-Imperial Competitive Algorithm (ICA)

Finding the optimal values of learning parameters for a data-mining model usually takes considerable time. The state-of-the-art parameter-tuning algorithms have significantly solved this problem. In this study, the imperial competitive algorithm (ICA) was selected and then fused with the ANFIS model. As a derivative of evolutionary computation, the ICA imitates the socio-political evolution of nations (Atashpaz-Gargari and Lucas, 2007). Seeking an optimal solution, the algorithm starts with a random selection of a solution, and proceeds by setting cost functions (Hosseini and Al Khaled, 2014). Random solutions imitate a candidate country. Candidate countries with the least cost function values create imperialists and by taking control of other country colonies form an empire. This process proceeds by following three evolutionary operators called assimilation (i.e. incorporating similar colonies), revolution (i.e. random reposition of the countries by giving them chances to take over), and competition (i.e. possessing the weak empires by the powerful ones) until the algorithm attains the optimal solution which is the optimal values of the learning parameters (Nazari-Shirkouhi et al., 2010).

3.3.3. Alternating Decision Tree (ADT)

Decision trees generally follow a divide-and-conquer strategy in which the problem is dissected into many branches (i.e. nodes and leaves) and instances, and similar features stored in data are categorized into the same groups (Witten et al., 2016). This process continues until the model attains the optimal solution, in that tree-pruning and the prediction power are balanced. Building a simple tree structure linked by the boosting algorithm, alternating decision trees (ADTs) render the problem into rule sets (i.e. leaves) and outcomes (i.e. end node of each branch) (Hong et al., 2015). The ADT starts by setting a constant value (e.g. a predicate condition and prediction nodes) and proceeds to split the branches. An ADT contains a decision node and a prediction node (i.e. a single number). Each time a weight is assigned to the node which is proportional to the number of training instances that lead to a specific classification. The final prediction (in probability terms) results from the summation of all the weights contributed to the root (Freund and Mason, 1999). In contrast to other classifiers, such as classification and regression trees (CART), which follow only one pass through the tree, ADT seeks through all the passes with true decision nodes and predictions (Breiman et al., 1984). More details of these processes can be found in Freund and Mason (1999) and Pfahringer et al. (2001).

3.3.4. Random Forest (RF)

Random forest modeling is one of the most popular data mining techniques and has been widely applied to many environmental studies (Vorpahl et al., 2012). As an ensemble of the classification and regression tree models, the popularity of RF is indebted to the unique learning and prediction algorithm which is expedited by a recursive random data use at each tree node and an error minimization technique (Moghaddam et al., 2019). The recursive factor selection technique gives the spring-affecting factors many chances to contribute to the modelling process and, in parallel, the random aspect strengthens the model's robustness and improves the learning process by shifting from one randomly partitioned inventory subset to another and accordingly the pattern of phenomenon of interest will be attained (Prasad et al., 2006). The error minimization technique follows an out-of-bag (OOB) error estimation which is a measure of prediction error (Bachmair and Weiler, 2012; Were et al., 2015). Simply put, the OOB is a mean prediction error and is estimated during the bootstrapping aggregation of the data sub-sample x_i by using only those trees that did not have that subset in their bootstrap sample (James et al., 2013). The final prediction averages the prediction values at each node. The mathematics of the RF model is articulated by Ho (1995) and Breiman (2001).

3.4. Performance assessment

The assessment of the performance of data mining models consists of cutoff-dependent and cutoff–independent measures. The receiver operating characteristic (ROC) curve as a practical cutoff-independent measure was used to assess the learning and prediction power of the models (Pradhan, 2013; Umar et al., 2014; Chen et al., 2017; Kornejady et al., 2017; Pourghasemi et al., 2017). The ROC curve plots the false positive rate (i.e. incorrectly predicting event-free locations or so-called absences as presences) on the X-axis against the true positive rate (i.e. correctly predicted presence locations) on the Y-axis (Pontius and Schneider, 2001; Swets, 2014). The area under the ROC curve (AUROC) is a common performance criterion in spatial modelling of groundwater potential (Naghibi et al., 2016; Rahmati et al., 2018). Additionally, the true skill score (so-called Hanssen and Kuipers discriminant or Pierces skill score) was used to address the differentiation power between the presence and absence localities (Allouche et al., 2006; Frattini et al., 2010). The performance measures were calculated using the ArcGIS toolbox PMT. For more information on PMT and mathematics of the performance measure one can refer to Rahmati et al. (2019).

4. Results

4.1. Multicollinearity analysis of predictive factors

The results of multicollinearity analysis demonstrated that the highest VIF value was 3.723 and the lowest TOL value was 0.226 which indicates that there is no multicollinearity among the predictive factors (Table 3). Therefore, all the predictive factors were used in modelling.

Table 3 here

4.2. Assessing the accuracy of maps produced with sample data sets of different sizes

Considering the goodness-of-fit and predictive performance values (Tables 4 and 5), a discernible performance decline was evident while progressing through the data sets (i.e., from D_1 towards D_2). This is in line with the preliminary graphical check. A rather sharp decrease in the predictive ability of the ADT and ANFIS models was observed when the number of springs decreased from 75% to 50% and from 50% to 25%. However, RF and ANFIS-ICA were more stable than were ADT and ANFIS as the spring sample size decreased.

Table 4 here

Table 5 here

Attesting to the previous inference, the descending pattern of learning capability and predictive power of all the models strictly stemmed from sample data reduction reflected by the data sets (i.e., 100%, 75%, 50%, and 25%), whether for the locations of spring presence or spring absence. Providing large positive and negative samples for presence-absence data-mining models is crucial, since larger samples make a larger and more comprehensive information matrix for each model and accordingly facilitate the learning process, prediction power, and generalization capability. Despite the performance decline, all four models demonstrated good learning and predictive power in the training step (i.e., AUROC ranged between 0.8 and 0.9) and in some cases, they had reached excellent performance (i.e., AUROC higher than 0.9). Since there was no evidence of any drastic decline of performance as all AUROC values were within and above the acceptable range, it is safe to say that all the models showed rather stable results and the sensitivity of the results to input sample-size alteration was somewhat negligible.

4.3. Groundwater potential mapping

It is evident that by moving from D_1 to D_4, almost all four models lack spatial differentiation as they conservatively categorized most of the study area into the moderate, high, or very high groundwater potential classes (yellow, green, blue) (Figs. 6, 7, 8, and 9). This is particularly evident in the groundwater potential maps generated by the ANFIS model (Fig. 6) and less so in the case of the ADT model (Fig. 9).

Fig. 6 here

Fig. 7 here

Fig. 8 here

Fig. 9 here

Since the RF model performed best, the final groundwater potential map derived from the RF model (data set D_1) was categorized into five potentiality classes (i.e., very low to very high) using the equal interval classification scheme (Akgun, 2012; Rahmati et al., 2018) (Fig. 10). It appears that a high potential site is located in the westernmost part of the study area. Other patches are scattered across the northeastern, southeastern, and central parts of the region and are mostly associated with deep valleys, dense drainage networks, and fault traces. About 18.7% of the study was found to have high and very high groundwater potential (Table 6).

Table 6 here

Fig. 10 here

4.4. Importance analysis of predictive factors

Results of the variable importance (VI) analysis based on the premier model and replication strategy (i.e., the RF model built upon 100% sample size) signified that the RSP index had contributed strongly to the modelling of groundwater potential; visually the pattern of RSP is similar to the spatial pattern observed in the final groundwater potential map (Table 7).

Table 7 here

5. Discussion

5.1. Model validation and comparison: influence of sample size on the model performance

In spatial modelling studies, the sample size has proven to be a significant factor affecting the predictive abilities of the models (Guisan et al., 2007). Since the gathering of hydrogeological data in mountainous regions is often the most costly and challenging part of a study, and because additional information may be difficult to acquire, guidance for researchers concerning model sensitivity and the amount of input data needed would be extremely useful (Hjort and Marmion, 2008). As discussed by Kresic and Bonacci (2010), this is especially important in studies using traditional survey methods and *in situ* measurements in remote areas. This study investigated the influence of sample size on the predictive performance of different models for groundwater potential mapping in a mountain bedrock aquifer. On the other hand, since the main role of performance metrics in the validation step is to filter the models by ranking their generalization capacity and spatial transferability, even the slight differences in a models' performance can elevate one model over another (Osna et al., 2014). Therefore, considering that the AUROC and TSS values of ANFIS-ICA and RF models were higher compared to their counterparts and given that they are respectively the hybrid or advanced versions of ANFIS and ADT models, the promising role of hybrid and ensemble learning techniques seems apparent. Results indicate that coupling ANFIS with an optimization algorithm called ICA improved both learning and prediction capabilities by improving the efficiency (in terms of time and computational space) of ANFIS' search for the best parameter values; this has also been discussed by Bui et al. (2018). The latter also enables ANFIS to overcome overfitting by avoiding selection of an amalgamation of parameter values and by circumventing a highly iterative learning process and becoming accustomed to the training data set (Jaafari et al., 2019). This process also improves the prediction power of ANFIS. In parallel, RF, by bearing on an ensemble mechanism comprised of numerous decision trees and choosing the training data sets with many replacements, can outperform simple tree models like ADT. In other subfields related to groundwater, Rahmati et al. (2019) compared the performances of several tree-based machine learning algorithms for predicting land subsidence hazards and found that RF outperformed simple decision-tree models. They explained that decision tree models, such as ADT, are simple structures where non-terminal nodes represent tests on one or more attributes and terminal nodes reflect decision outcomes.

The TSS metric, viewed as the true success of a model, is in complete accordance with the AUROC values. The differences between the values of TSS and AUROC result from their different computational methods. However, apart from TSS being an all-inclusive metric by using all the elements of the confusion matrix in conjunction, AUROC is reportedly used as a more reliable performance index (Nampak et al., 2014; Rahmati et al., 2016) particularly because it is cutoff independent, while TSS was calculated under a predefined 50% cutoff and any other cutoff value would result in a different set of TSS values. The AUROC and TSS metrics agree on the performances of models in the validation step, upon which the RF model can be determined to be the premier model for all four data sets of different sample sizes.

5.2. Assessment of variable importance

Assessing the relative importance of independent variables (i.e., groundwater conditioning factors) is of practical relevance to environmental managers and decision makers dealing with efficiency in the planning and allocation of limited resources (Testa et al., 2016). Although various numerical, statistical (e.g., logistic regression), and expert opinion-based (e.g., analytical hierarchy process (AHP)) models have been employed to spatially predict groundwater potential, the relative importance of geo-environmental factors is still debated (Naghibi et al., 2019). Machine learning algorithms have been helping increasing numbers of environmental management decision makers to achieve new insights, both in terms of the relationships between hydrogeological and geo-environmental factors and groundwater potential, and they are now regarded as practical components that efficiently contribute to improvement of water resources management (Deo and Şahin, 2016). On the other hand, the relative importance of factors is often affected by the model used (Bui et al., 2016). Therefore, to provide a fair judgment, a commonly used model should be selected to analyze the ranking of independent variables. Following the literature, the RF model was selected as it has performed well in groundwater-potential modeling in several regions (Rahmati et al., 2016; Zabihi et al., 2016; Naghibi et al., 2017b). Results indicate that RSP makes the highest contribution to the prediction of groundwater potential. The importance of RSP has been confirmed by Rahmati et al. (2018). The RSP factor is a geomorphometric index with values that range from 0 to 1 and it represents the relative position of a cell with respect to the valley floor and ridgetop (Krishnamurthy et al., 2016). Intuitively, areas with lower RSP values (i.e. closer to valley floor) were characterized as having higher groundwater potential. Conversely, the areas located at or near ridgetops correspond with fast subsurface flow (average elevations with steep slopes) or deep percolation mechanisms (high elevations with gentle slopes) where the water table was deep in the strata. Hence, the importance of RSP for groundwater potential mapping is sensible.

Lithological formations and their structures represent the subsurface mechanism, such as permeability, porosity, and ultimately the way some rocks manage to restrict the water table close to the surface. This is in line with previous studies (Ozdemir, 2011; Chen et al., 2019). The distance from faults and generally any fissure formed by soil surface tension can be a good way to convey surface water to and nourish the water table. As has been shown (Adiat et al., 2012), faults are geological features with secondary permeability that control the movement and/or storage of groundwater and can provide important information about subsurface characteristics. In addition, drainage density and TWI are the other important representative indices for surface mechanisms, in that, the highest values represent the areas contributing most to runoff generation that would contribute to the water table if the geo-topological conditions have been met.

5.3. Limitations of the research

The most important limitation is that hydrogeological data in the study area were unavailable. Hydrogeological parameters can improve model prediction, especially when models are applied at a large-scale where heterogeneity causes uncertainty (de Barros et al., 2012). As explained by Worthington (2015), groundwater modeling in bedrock aquifers is complex because there is often substantial flow through fractures, and the interconnectivity and magnitude of these fractures are usually uncertain. This challenge can be dealt with using multicomponent analysis of tritium data and stable isotope data, although they would be costly and time-consuming to implement (Chen et al., 2018). Geophysical techniques can provide useful information that enhance the predictive capacity of the model. Without the application of geophysical techniques, our knowledge about subsurface structures would remain extremely limited.

5.4. Wider landscape-scale implications of the findings

Numerical modeling of groundwater at a small-scale in mountain bedrock aquifers is often subject to substantial uncertainty because there is often a need for data regarding hydrogeological parameters that are not readily available (Wu and Zeng, 2013). Therefore, groundwater potential modeling using numerical models at small scales, especially in mountainous regions with bedrock aquifers, is difficult. However, machine learning models have overcome this issue and can spatially model groundwater potential over larger areas. For example, Ghorbani Nejad et al. (2017) successfully applied some data mining models for groundwater potential mapping in mountain bedrock aquifers of Lorestan province. Falah et al. (2017) assessed the applicability of generalized additive model for groundwater potential modelling in a large region and compared its performance with some bivariate statistical methods. Their results clearly indicated that machine-learning models can model groundwater potential in large areas and also in data-scarce regions. These powerful models can analyze spatial variation of the groundwater potential conditions and can identify the relationships between the groundwater conditioning factors (topo-hydrological, lithology, etc.) and groundwater potential (Naghibi et al., 2016).

One of the important points in applying machine learning and data-mining models for groundwater potential mapping over large regions is that sample size should be logical that models could analyze all parts of the region (Roy et al., 2013). This study investigated the sensitivity of models to groundwater sample size and distinguished their capabilities. Results clearly demonstrate that the RF model provides a better prediction of groundwater potential when the sample size limited. RF models can be applied in other large regions even if a spring inventory is not readily available. This conclusion is vital and useful for groundwater potential modeling in large and data-scarce regions.

6. Conclusions

This study was designed to determine the effect of the impact of sample size on the performance of different models for groundwater-potential mapping in a mountain bedrock aquifer. Four sample data sets were prepared from a spring-inventory: D_1 (100% of springs, n=714), D_2 (75% of springs, n=535), D_3 (50% of springs, n=358), and D_4 (25% of springs, n=177). The response of four machine learning algorithms including ANFIS, ANFIS-ICA, RF, and ADT to the change of sample size (i.e., springs number) was evaluated. The findings reported here lead to three main conclusions:

I. Reducing sample sizes affected groundwater potential modelling, although the model type can considerably influence the level of its response. The performance of ADT (AUROC=79.29–

88.46%, TSS=0.59–0.74), and ANFIS (AUROC=73.11–88.43%, TSS=0.59–0.74) sharply decreased when sample size decreased, whereas the hybrid ANFIS-ICA model and RF showed rather higher learning and prediction powers for all data sets. The ICA optimization algorithm enabled the ANFIS to find the best set of parameters and avoid any amalgamation of parameters upon which the model is to be built and trained.

II. Ultimately, the RF model (AUROC=90.74–96.32%, TSS=0.79–0.85) was identified as the premier model based on performance metrics in both the training and validation stages. If only a limited number of observations (springs) is available for predictive modelling, the RF is the best choice for geospatially modelling groundwater potential. Results of this study suggest that RF should continue to be used as the benchmark model for groundwater potential modelling.

III. Although the premier model is fed by all the spring-affecting factors, as with any spatial modelling effort, high groundwater potential can be rather more responsive to a few uniquely informative factors, such as RSP (VI=24.35%), lithology (VI=20.41%), and distance from faults (VI=15.29%).

Acknowledgements

We thank Iranian Department of Water Resources Management (IDWRM) and Department of Geological Survey of Iran (GSI) for providing necessary data and maps. This research was supported by the GIS research group, Ton Duc Thang university, Vietnam.

References

Akgun, A. 2012. A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: a case study at İzmir, Turkey. Landslides 9(1), 93-106.

Allouche, O., Tsoar, A., Kadmon, R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. 43(6), 1223-1232.

Atashpaz-Gargari, E., Lucas, C. 2007. Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition. In 2007 IEEE congress on evolutionary computation (pp. 4661-4667). IEEE.

Bachmair, S., Weiler, M. 2012. Hillslope characteristics as controls of subsurface flow variability. Hydrol. Earth Syst. Sc. 16(10), 3699-3715.

Breiman, L. 2001. Random forests. Mach. Learn. 45(1), 5-32.

Breiman, L., Friedman, J., Olshen, R., Stone, C. 1984. Classification and regression trees–crc press. Boca Raton, Florida.

Bui, D.T, Khosravi, K., Li, S., Shahabi, H., Panahi, M., Singh, V., et al. 2018. New hybrids of anfis with several optimization algorithms for flood susceptibility modeling. Water 10(9), 1210. https://doi.org/10.3390/w10091210.

Bui, D.T., Panahi, M., Shahabi, H., Singh, V.P., Shirzadi, A., Chapi, K., Khosravi, K., Chen, W., Panahi, S., Li, S., Ahmad, B.B., 2018. Novel hybrid evolutionary algorithms for spatial prediction of floods. Sci. Rep-Uk 8(1), p.15364. https://doi.org/10.1038/s41598-018-33755-7.

Bui, D.T., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I., 2016. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13(2), 361-378.

Chen, W., Li, H., Hou, E., Wang, S., Wang, G., Panahi, M., Li, T., Peng, T., Guo, C., Niu, C., Xiao, L., 2018. GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models. Sci. Total Environ. 634, 853-867.

Chen, W., Panahi, M., Khosravi, K., Pourghasemi, H.R., Rezaie, F., Parvinnezhad, D., 2019. Spatial prediction of groundwater potentiality using ANFIS ensembled with teaching-learning-based and biogeography-based optimization. J. Hydrol. 572, 435-448.

Chen, W., Pourghasemi, H. R., Kornejady, A., Zhang, N. 2017. Landslide spatial modeling: introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. Geoderma 305, 314-327.

Chen, W., Tsangaratos, P., Ilia, I., Duan, Z., & Chen, X. 2019. Groundwater spring potential mapping using population-based evolutionary algorithms and data mining methods. Science of The Total Environment, 684, 31-49.

Chen, X., Zhang, Z., Soulsby, C., Cheng, Q., Binley, A., Jiang, R. and Tao, M., 2018. Characterizing the heterogeneity of karst critical zone and its hydrological function: an integrated approach. Hydrological processes, 32(19), pp.2932-2946.

Chenini, I., Mammou, A.B., 2010. Groundwater recharge study in arid region: an approach using GIS techniques and numerical modeling. Comput. Geosci. 36(6), 801-817.

Choubin, B., Malekian, A. 2017. Combined gamma and M-test-based ANN and ARIMA models for groundwater fluctuation forecasting in semiarid regions. Environ. Earth Sci. 76(15), 538.

Choubin, B., Rahmati, O., Soleimani, F., Alilou, H., Moradi, E., Alamdari, N., 2019. Regional Groundwater Potential Analysis Using Classification and Regression Trees. In Spatial Modeling in GIS and R for Earth and Environmental Sciences (pp. 485-498). Elsevier.

Choubin, B., Rahmati, O., Soleimani, F., Alilou, H., Moradi, E., Alamdari, N. 2019. Regional Groundwater Potential Analysis Using Classification and Regression Trees. In Spatial Modeling in GIS and R for Earth and Environmental Sciences (pp. 485-498). Elsevier.

Coleman, T.I., Parker, B.L., Maldaner, C.H., Mondanos, M.J., 2015. Groundwater flow characterization in a fractured bedrock aquifer using active DTS tests in sealed boreholes. J. Hydrol. 528, 449-462.

de Barros, F.P., Ezzedine, S. and Rubin, Y., 2012. Impact of hydrogeological data on measures of uncertainty, site characterization and environmental performance metrics. Advances in Water Resources, 36, 51-63.

Deo, R.C., Şahin, M., 2016. An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland. Environ. Monit. Assess. 188 (2), 90.

Falah, F., Ghorbani Nejad, S., Rahmati, O., Daneshfar, M., Zeinivand, H., 2017. Applicability of generalized additive model in groundwater potential modelling and comparison its performance by bivariate statistical methods. Geocarto Int. 32(10), 1069-1089.

Frattini, P., Crosta, G., Carrara, A. 2010. Techniques for evaluating the performance of landslide susceptibility models. Eng. Geol. 111(1-4), 62-72.

Freund, Y., Mason, L. 1999. The alternating decision tree learning algorithm. In icml. 99, 124-133.

Friedman, J., Hastie, T., Tibshirani, R. 2001. The elements of statistical learning (Vol. 1, No. 10). New York: Springer series in statistics.

Ghorbani Nejad, S., Falah, F., Daneshfar, M., Haghizadeh, A. and Rahmati, O., 2017. Delineation of groundwater potential zones using remote sensing and GIS-based data-driven models. Geocarto international, 32(2), pp.167-187.

Golkarian, A., Naghibi, S.A., Kalantar, B., Pradhan, B., 2018. Groundwater potential mapping using C5. 0, random forest, and multivariate adaptive regression spline models in GIS. Environ. Monit. Assess. 190(3), p.149.

Guisan, A., Graham, C.H., Elith, J., Huettmann, F., the NCEAS Species Distribution Modelling Group, 2007. Sensitivity of predictive species distribution models to change in grain size. Diversity and Distributions 13, 332–340.

Hair JF, Black WC, Babin BJ, Anderson RE 2009. Multivariate data analysis. Prentice Hall, New York

Hancock, P.J., Boulton, A.J., 2009. Sampling groundwater fauna: efficiency of rapid assessment methods tested in bores in eastern Australia. Freshwater Biol. 54(4), 902-917.

Hjort, J. and Marmion, M., 2008. Effects of sample size on the accuracy of geomorphological models. Geomorphology, 102(3-4), 341-350.

Ho, T. K. 1995. Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition (Vol. 1, pp. 278-282). IEEE.

Hong, H., Pradhan, B., Xu, C., Bui, D. T. 2015. Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. Catena 133, 266-281.

Hosseini, S., Al Khaled, A. 2014. A survey on the imperialist competitive algorithm metaheuristic: implementation in engineering domain and directions for future research. Appl. Soft. Comput. 24, 1078-1094.

Jaafari, A., Zenner, E.K., Panahi, M. Shahabi, H., 2019. Hybrid artificial intelligence models based on a neuro-fuzzy system and metaheuristic optimization algorithms for spatial prediction of wildfire probability. Agr. Forest Meteorol. 266, 198-207.

Jackson, R. B., Carpenter, S. R., Dahm, C. N., McKnight, D. M., Naiman, R. J., Postel, S. L., Running, S. W. 2001. Water in a changing world. Ecol. Appl. 11(4), 1027-1045.

James, G., Witten, D., Hastie, T., Tibshirani, R. 2013. An introduction to statistical learning (Vol. 112, p. 18). New York: springer.

Jang, J. S. 1993. ANFIS: adaptive-network-based fuzzy inference system. IEEE transactions on systems, man, and cybernetics, 23(3), 665-685.

Jang, J. S. R. 1991. Fuzzy modeling using generalized neural networks and kalman filter algorithm. In AAAI 91, 762-767.

Jang, J. S. R., Sun, C. T., Mizutani, E. 1997. Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [Book Review]. IEEE T. Automat. Contr. 42(10), 1482-1484.

Kollat, J.B., Reed, P.M., Maxwell, R.M., 2011. Many-objective groundwater monitoring network design using bias-aware ensemble Kalman filtering, evolutionary optimization, and visual analytics. Water Resour. Res. 47(2). https://doi.org/10.1029/2010WR009194.

Kordestani, M. D., Naghibi, S. A., Hashemi, H., Ahmadi, K., Kalantar, B., Pradhan, B. 2019. Groundwater potential mapping using a novel data-mining ensemble model. Hydrogeol. J. 27(1), 211-224.

Kornejady, A., Ownegh, M., Bahremand, A. 2017. Landslide susceptibility assessment using maximum entropy model with two different data sampling methods. Catena 152, 144-162.

Kresic, N. and Bonacci, O., 2010. Spring discharge hydrograph. In Groundwater hydrology of springs (pp. 129-163). Butterworth-Heinemann.

Krishnamurthy, R., Cushman, S.A., Sarkar, M.S., Malviya, M., Naveen, M., Johnson, J.A., Sen, S., 2016. Multi-scale prediction of landscape resistance for tiger dispersal in central India. Landscape Ecol. 31(6), 1355-1368.

Leach, J.M., Coulibaly, P., Guo, Y., 2016. Entropy based groundwater monitoring network design considering spatial distribution of annual recharge. Adv. Water Resour. 96, 108-119.

Lee, S., Hong, S.M. and Jung, H.S., 2018. GIS-based groundwater potential mapping using artificial neural network and support vector machine models: the case of Boryeong city in Korea. Geocarto Int. 33(8), 847-861.

Lee, S., Lee, C.W. 2015. Application of decision-tree model to groundwater productivity-potential mapping. Sustainability 7(10), 13416-13432.

Lee, S., Song, K.Y., Kim, Y. and Park, I., 2012. Regional groundwater productivity potential mapping using a geographic information system (GIS) based artificial neural network model. Hydrogeol. J. 20(8), 1511-1527.

Levison, J., Novakowski, K., Reiner, E.J. and Kolic, T., 2012. Potential of groundwater contamination by polybrominated diphenyl ethers (PBDEs) in a sensitive bedrock aquifer (Canada). Hydrogeol. J. 20(2), 401-412.

Liška, A., Kruszewski, G., Baroni, M. 2018. Memorize or generalize? searching for a compositional RNN in a haystack. arXiv preprint arXiv:1802.06467.

Manap, M.A., Nampak, H., Pradhan, B., Lee, S., Sulaiman, W.N.A., Ramli, M.F., 2014. Application of probabilistic-based frequency ratio model in groundwater potential mapping using remote sensing data and GIS. Arab. J. Geosci. 7(2), 711-724.

Manap, M.A., Sulaiman, W.N.A., Ramli, M.F., Pradhan, B., Surip, N., 2013. A knowledge-driven GIS modeling technique for groundwater potential mapping at the Upper Langat Basin, Malaysia. Arab. J. Geosci. 6(5), 1621-1637.

Miraki, S., Zanganeh, S.H., Chapi, K., Singh, V.P., Shirzadi, A., Shahabi, H., Pham, B.T., 2019. Mapping Groundwater Potential Using a Novel Hybrid Intelligence Approach. Water Resour. Manag. 33(1), 281-302.

Moghaddam, D. D., Pourghasemi, H. R., Rahmati, O. 2019. Assessment of the Contribution of Geo-environmental Factors to Flood Inundation in a Semi-arid Region of SW Iran: Comparison of Different Advanced Modeling Approaches. In Natural Hazards GIS-Based Spatial Modeling Using Data Mining Techniques (pp. 59-78). Springer, Cham.

Motagh, M., Walter, T. R., Sharifi, M. A., Fielding, E., Schenk, A., Anderssohn, J., Zschau, J. 2008. Land subsidence in Iran caused by widespread water reservoir overexploitation. Geophys. Res. Lett. 35(16). https://doi.org/10.1029/2008GL033814.

Mousavi, S. M., Golkarian, A., Naghibi, S. A., Kalantar, B., Pradhan, B. 2017. GIS-based groundwater spring potential mapping using data mining boosted regression tree and probabilistic frequency ratio models in Iran. AIMS Geosci. 3(1), 91-115.

Mukherjee, P., Singh, C. K., Mukherjee, S. 2012. Delineation of groundwater potential zones in arid region of India—a remote sensing and GIS approach. Water Resour. Manag. 26(9), 2643-2672.

Naghibi, S. A., Moghaddam, D. D., Kalantar, B., Pradhan, B., Kisi, O. 2017a. A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping. J. Hydrol. 548, 471-483.

Naghibi, S.A., Ahmadi, K., Daneshi, A., 2017b. Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. Water Resour. Manag. 31(9), 2761-2775.

Naghibi, S.A., Dolatkordestani, M., Rezaei, A., Amouzegari, P., Heravi, M.T., Kalantar, B., Pradhan, B., 2019. Application of rotation forest with decision trees as base classifier and a novel ensemble model in spatial modeling of groundwater potential. Environ. Monit. Assess 191(4), p.248. https://doi.org/10.1007/s10661-019-7362-y.

Naghibi, S.A., Pourghasemi, H.R., Dixon, B., 2016. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. Environ. Monit. Assess. 188(1), p.44. https://doi.org/10.1007/s10661-015-5049-6.

Nampak, H., Pradhan, B., Manap, M.A., 2014. Application of GIS based data driven evidential belief function model to predict groundwater potential zonation. J. Hydrol. 513, 283-300.

Nazari-Shirkouhi, S., Eivazy, H., Ghodsi, R., Rezaie, K., Atashpaz-Gargari, E. 2010. Solving the integrated product mix-outsourcing problem using the imperialist competitive algorithm. Expert Syst. Appl. 37(12), 7615-7626.

O'Brien, R. M. 2007. A caution regarding rules of thumb for variance inflation factors. Qual. Quant. 41(5), 673-690.

Oh, H.J., Kim, Y.S., Choi, J.K., Park, E., Lee, S., 2011. GIS mapping of regional probabilistic groundwater potential in the area of Pohang City, Korea. J. Hydrol. 399(3-4), 158-172.

Osna, T., Sezer, E.A., Akgun, A., 2014. GeoFIS: an integrated tool for the assessment of landslide susceptibility. Comput. Geosci. 66, 20-30.

Ozdemir, A., 2011a. GIS-based groundwater spring potential mapping in the Sultan Mountains (Konya, Turkey) using frequency ratio, weights of evidence and logistic regression methods and their comparison. J. Hydrol. 411(3-4), 290-308.

Ozdemir, A., 2011b. Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey). J. Hydrol. 405(1-2), 123-136.

Parisi, A., Monno, V., Fidelibus, M.D., 2018. Cascading vulnerability scenarios in the management of groundwater depletion and salinization in semi-arid areas. Int. J. Disast. Risk Re. 30, 292-305.

Pfahringer, B., Holmes, G., Kirkby, R. 2001. Optimizing the induction of alternating decision trees. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 477-487). Springer, Berlin, Heidelberg.

Pham, B. T., Jaafari, A., Prakash, I., Singh, S. K., Quoc, N. K., & Bui, D. T. 2019. Hybrid computational intelligence models for groundwater potential mapping. Catena, 182, 104101.

Pontius Jr, R. G., Schneider, L. C. 2001. Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. Agr. Ecosyst. Environ. 85(1-3), 239-248.

Pourghasemi, H. R., Yousefi, S., Kornejady, A., Cerdà, A. 2017. Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. Sci. Total Environ. 609, 764-775.

Pourghasemi, H.R. and Beheshtirad, M., 2015. Assessment of a data-driven evidential belief function model and GIS for groundwater potential mapping in the Koohrang Watershed, Iran. Geocarto Int. 30(6), 662-685.

Pourtaghi, Z.S. and Pourghasemi, H.R., 2014. GIS-based groundwater spring potential assessment and mapping in the Birjand Township, southern Khorasan Province, Iran. Hydrogeol. J. 22(3), 643-662.

Pradhan, B. 2013. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Comput. Geosci. 51, 350-365.

Prasad, A. M., Iverson, L. R., Liaw, A. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9(2), 181-199.

Premkumar, K., Manikandan, B. V. 2014. Adaptive Neuro-Fuzzy Inference System based speed controller for brushless DC motor. Neurocomputing 138, 260-270.

Rahmati, O., Golkarian, A., Biggs, T., Keesstra, S., Mohammadi, F., Daliakopoulos, I. N. 2019. Land subsidence hazard modeling: Machine learning to identify predictors and the role of human activities. J. Environ. Manag. 236, 466-480.

Rahmati, O., Kornejady, A., Samadi, M., Deo, R. C., Conoscenti, C., Lombardo, L., et al. 2019. PMT: New analytical framework for automated evaluation of geo-environmental modelling approaches. Sci. Total Environ. 664, 296–311.

Rahmati, O., Naghibi, S.A., Shahabi, H., Bui, D.T., Pradhan, B., Azareh, A., Rafiei-Sardooi, E., Samani, A.N., Melesse, A.M., 2018. Groundwater spring potential modelling: Comprising the capability and robustness of three different modeling approaches. J. Hydrol. 565, 248-261.

Rahmati, O., Pourghasemi, H.R., Melesse, A.M., 2016. Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran Region, Iran. Catena 137, 360-372.

Rahmati, O., Tahmasebipour, N., Haghizadeh, A., Pourghasemi, H. R., Feizizadeh, B. 2017. Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. Geomorphology 298, 118-137.

Rasul, G., 2016. Managing the food, water, and energy nexus for achieving the Sustainable Development Goals in South Asia. Environ. Dev. 18, 14-25.

Rathay, S.Y., Allen, D.M., Kirste, D., 2018. Response of a fractured bedrock aquifer to recharge from heavy rainfall events. J. Hydrol. 561, 1048-1062.

Ravilious, K. 2018. Tehran's drastic sinking exposed by satellite data. Nature 564(7734), 17-18.

Roy, J.W., Zaitlin, B., Hayashi, M. and Watson, S.B., 2011. Influence of groundwater spring discharge on small-scale spatial variation of an alpine stream ecosystem. Ecohydrology, 4(5), pp.661-670.

Shenga, Z.D., Baroková, D., Šoltész, A., 2018a. Modeling of groundwater extraction from wells to control excessive water levels. Pollack Periodica, 13(1), 125-136.

Shenga, Z.D., Baroková, D., Šoltész, A., 2018b. Numerical modeling of groundwater to assess the impact of proposed railway construction on groundwater regime. Pollack Periodica, 13(3), 187-196.

Singh, L. K., Jha, M. K., & Chowdary, V. M. 2018. Assessing the accuracy of GIS-based Multi-Criteria Decision Analysis approaches for mapping groundwater potential. Ecological Indicators, 91, 24-37.

Swets, J. A. 2014. Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers. Psychology Press. https://doi.org/10.4324/9781315806167.

Testa, F., Gusmerottia, N.M., Corsini, F., Passetti, E. and Iraldo, F., 2016. Factors affecting environmental management by small and micro firms: The importance of entrepreneurs' attitudes and environmental investment. Corporate Social Responsibility and Environmental Management, 23(6), pp.373-385.

Umar, Z., Pradhan, B., Ahmad, A., Jebur, M. N., Tehrany, M. S. 2014. Earthquake induced landslide susceptibility mapping using an integrated ensemble frequency ratio and logistic regression models in West Sumatera Province, Indonesia. Catena 118, 124-135.

Velis, M., Conti, K.I., Biermann, F. 2017. Groundwater and human development: synergies and trade-offs within the context of the sustainable development goals. Sustain. Sci. 12(6), 1007-1017.

Viviroli, D., Dürr, H.H., Messerli, B., Meybeck, M., Weingartner, R., 2007. Mountains of the world, water towers for humanity: Typology, mapping, and global significance. Water resources research, 43(7). https://doi.org/10.1029/2006WR005653.

Voeckler, H., Allen, D.M., 2012. Estimating regional-scale fractured bedrock hydraulic conductivity using discrete fracture network (DFN) modeling. Hydrogeol. J. 20(6), 1081-1100.

Vorpahl, P., Elsenbeer, H., Märker, M., Schröder, B. 2012. How can statistical models help to determine driving factors of landslides?. Ecol. Model. 239, 27-39.

Were, K., Bui, D. T., Dick, Ø. B., Singh, B. R. 2015. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecol. Indic. 52, 394-403.

Witten, I.H., Frank, E., Hall, M.A., Pal, C.J. 2016. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, Burlington, USA.

Worthington, S.R., 2015. Diagnostic tests for conceptualizing transport in bedrock aquifers. Journal of Hydrology, 529, pp.365-372.

Wu, J. and Zeng, X., 2013. Review of the uncertainty analysis of groundwater numerical simulation. Chinese Science Bulletin, 58(25), pp.3044-3052.

Zabihi, M., Pourghasemi, H.R., Pourtaghi, Z.S., Behzadfar, M. 2016. GIS-based multivariate adaptive regression spline and random forest models for groundwater potential mapping in Iran. Environ. Earth Sci. 75(8), p.665.

Zare, M., Koch, M. 2018. Groundwater level fluctuations simulation and prediction by ANFIS-and hybrid Wavelet-ANFIS/Fuzzy C-Means (FCM) clustering models: Application to the Miandarband plain. J. Hydro-Environ. Res. 18, 63-76.

Table 1 Lithology of the study area.

Unit	Description	Age	Area (ha)
Cb	Dolomite, limestone and variegated shale (Barut formation)	Cambrian	12.88
Cl	Dark red medium - grained arkoses to sub arkoses sandstone and siltstone (Lalun formation)	Cambrian	3536.6
Cm	Dark grey to black fossil limestone with subordinate black shale (Mobarak formation)	Carboniferous	2742.8
Czl	Undifferentiated unit, composed of dark red siltstone and sandstone	Cambrian	223.4
Db-sh	Undifferentiated limestone, shale and marl	Devonian	685.2
E1c	Pale-red, polygenic conglomerate and sandstone	Paleocene-Eocene	280.8
E1m	Marl, marl and limestone	Eocene	25447.2
E2s	Sandstone, marl and limestone	Eocene	13612.8
E3m	Marl, sandstone and limestone	Eocene	5856.9
Eav	Andesitic volcanic	Eocene	6156.6
Ek	Well bedded green tuff and shale (Karaj formation)	Eocene	83437.1
Ekgy	Gypsum	Eocene	592.9
EOgy	Gypsum (salt plug)	Eocene-Oligocene	13404.9
EOsa	Salt dome	Eocene-Oligocene	2694.5
Jd	Well - bedded to thin - bedded, greenish - grey argillaceous limestone with intercalations of calcareous shale (Dalichai formation)	Jurassic	353.3
Jl	Light grey, thin - bedded to massive limestone (Lar formation)	Jurassic-Cretaceous	44607.2
Juc	White, quartzes conglomerate	Middle. Jurassic	1940.6
K	Cretaceous rocks in general	Cretaceous	815.3
K2c	Conglomerate and sandstone	Cretaceous	3682.1
Kbv	Basaltic volcanic	Cretaceous	2378.4
Ktzl	Thick bedded to massive, white to pinkish orbitolina bearing limestone (Tizkuh formation)	Cretaceous	26267.3
Ku	Upper cretaceous, undifferentiated rocks	Cretaceous	16386.4
Mur	Red marl, marl, sandstone and conglomerate (Upper red formation)	Miocene	3078.4
Murm	Light - red to brown marl and marl with sandstone intercalations	Miocene	21920.8
Murmg	marl	Miocene	21669.5
Mursh	Variegated shale, marl and sandstone	Miocene	16824.2
Olgy	Gypsum	Oligocene	2049.2
OMql	Massive too thick - bedded limestone	Oligocene-Miocene	43079.1
pC-C	Late Proterozoic - early Cambrian undifferentiated rocks	Precambrian	3281.1
PeEz	Reef-type limestone and marl (Ziarat formation)	Paleocene-Eocene	2081.6
Pgkc	Light-red coarse grained, polygenic conglomerate with sandstone intercalations	Paleocene-Eocene	14807.9
Plc	conglomerate and sandstone	Pliocene	42065.3
PlQc	Fluvial conglomerate, Piedmont conglomerate and sandstone.	Pliocene-Quaternary	2192.7
Pr	Dark grey medium - bedded to massive limestone (Ruteh Limestone)	Permian	587.9
Qft1	High level piedmont fan and valley terrace deposits	Quaternary	34551.4
Qft2	Low level piedmont fan and valley terrace deposits	Quaternary	32296.8
TRe	Thick bedded grey o'olitic limestone; thin - platy, yellow to pinkish shale limestone with worm tracks and well to thick - bedded dolomite and dolomitic limestone (Elikah formation)	Triassic	9300.2
TRJs	Dark grey shale and sandstone (Shemshak formation)	Triassic-Jurassic	15402.8

Table 2 Description of controlling factors used for groundwater potential modelling.

Spring affecting-factor	Scale	Data source	Data type
Topographic wetness index (TWI)	1:50,000	DEM-derived	Grid (10 m ×10 m)
Relative slope position (RSP)	1:50,000	DEM-derived	Grid (10 m ×10 m)
Plan and profile curvatures	1:50,000	DEM-derived	Grid (10 m ×10 m)
Topographic position index (TPI)	1:50,000	DEM-derived	Grid (10 m ×10 m)
Terrain ruggedness index (TRI)	1:50,000	DEM-derived	Grid (10 m ×10 m)
Distance from fault	1:50,000	DEM-derived	Grid (10 m ×10 m)
Drainage density	1:50,000	DEM-derived	Grid (10 m ×10 m)
Slope	1:50,000	DEM-derived	Grid (10 m ×10 m)
Elevation	1:50,000	DEM-derived	Grid (10 m ×10 m)
Soil texture	1:50,000	IDWRM[a]	Polygon
Aspect	1:50,000	DEM-derived	Grid (10 m ×10 m)
Land use	1:50,000	IDWRM[a]	Polygon
Soil hydrological group	1:50,000	IDWRM[a]	Polygon
Lithology	1:50,000	GSDI[b]	Polygon

IDWRM[a]: Iranian Department of Water Resources Management; GSDI[b]: Geological Surveys Department of Iran

Table 3 Multicollinearity of predictive factors based on the VIF and TOL indices

Factor	VIF	TOL
Elevation	3.723	0.226
Topographic wetness index (TWI)	3.168	0.318
Aspect	1.632	0.896
Slope	1.520	0.919
Distance from faults	1.146	0.886
Drainage density	1.797	0.871
Relative slope position (RSP)	2.212	0.541
Topographic position index (TPI)	3.663	0.337
Lithology	1.455	0.769
Terrain roughness index (TRI)	2.934	0.394
Plan curvature	2.781	0.523
Profile curvature	2.626	0.518
Soil texture	1.354	0.869
Soil hydrological group	2.433	0.437
Land use	1.590	0.910

Table 4 Goodness-of-fit of the models for four data replicates (D_1-D_4) in the training step

Evaluation criteria	Data set	Models			
		ANFIS	ANFIS-ICA	RF	ADT
AUROC (%)	D1	**90.06**	**92.66**	**97.35**	**91.62**
	D2	88.71	90.32	97.08	90.03
	D3	87.33	88.25	94.51	87.17
	D4	86.52	85.13	92.44	85.91
TSS	D1	**0.79**	**0.81**	**0.87**	**0.80**
	D2	0.73	0.78	0.84	0.71
	D3	0.65	0.73	0.82	0.64
	D4	0.61	0.66	0.80	0.60

Table 5 Predictive performance of the models for four data replicates (D_1-D_4) in the validation step

Evaluation criteria	Data set	Models			
		ANFIS	ANFIS-ICA	RF	ADT
AUROC (%)	D1	**88.43**	**91.55**	**96.32**	**88.46**
	D2	84.32	88.98	95.41	85.12
	D3	78.59	84.45	92.86	81.33
	D4	73.11	81.23	90.74	79.29
TSS	D1	**0.74**	**0.81**	**0.85**	**0.74**
	D2	0.71	0.78	0.83	0.65
	D3	0.62	0.75	0.82	0.61
	D4	0.59	0.74	0.79	0.59

Table 6 Relative distributions of the groundwater potential classes based on the RF model (data set D_1)

Number	Class	Area (%)
1	Very low	17.98
2	Low	22.93
3	Medium	40.41
4	High	6.35
5	Very high	12.33

Table 7 Results of the factor importance analysis derived from the RF model

Factor	Variable importance (VI) (%)
Relative slope position (RSP)	24.35
Lithology	20.41
Distance from faults	15.29
Drainage density	7.44
Topographic wetness index (TWI)	5.32
Elevation	4.12
Topographic position index (TPI)	3.56
Aspect	3.24
Slope	3.11
Terrain roughness index (TRI)	2.97
Plan curvature	2.94
Profile curvature	2.86
Soil texture	2.43
Soil hydrological group	1.38
Land use	0.58

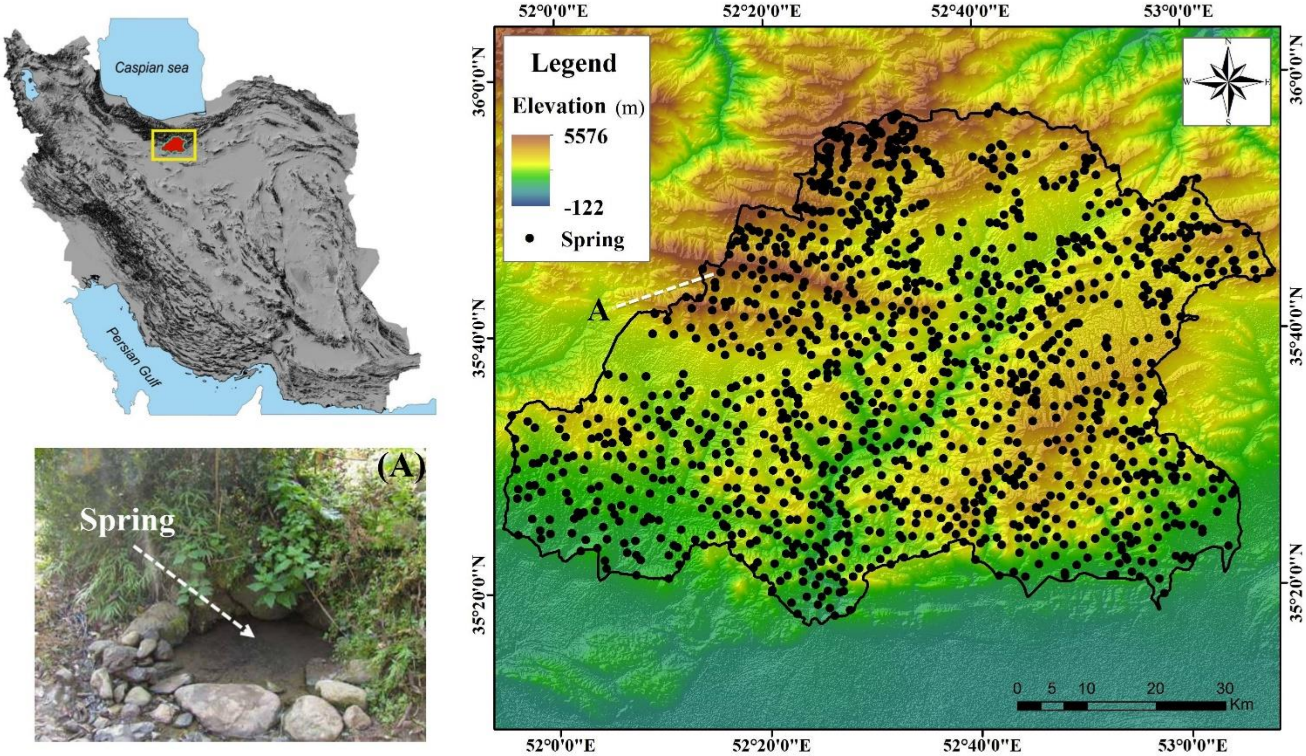


Fig. 1 Geographical location of the northern Hableh-Roud basin and the distribution of springs with inset (A) showing a photograph of a spring observed in the study area.

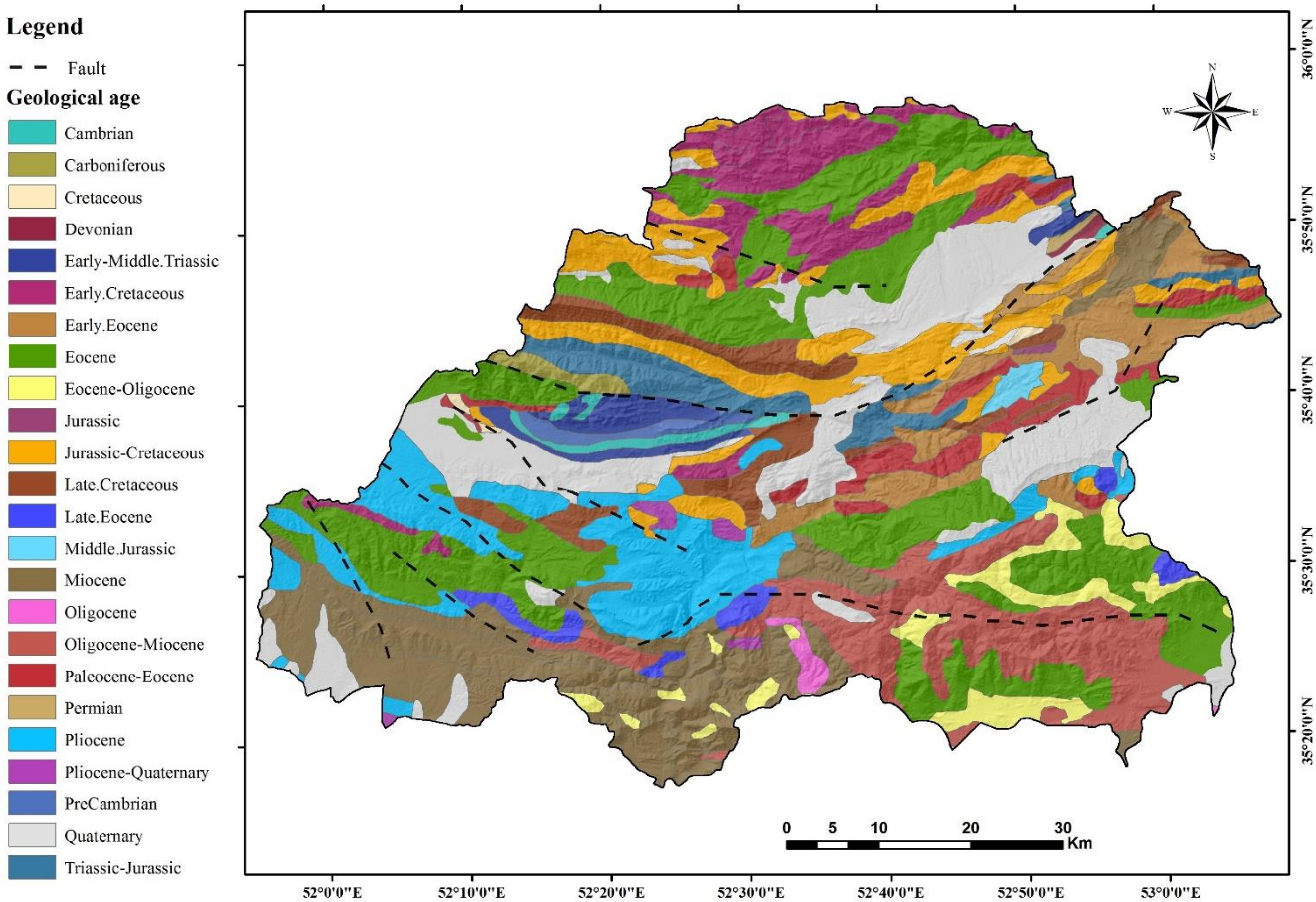


Fig. 2 Geological map of the study area showing faults and the age of geological units.

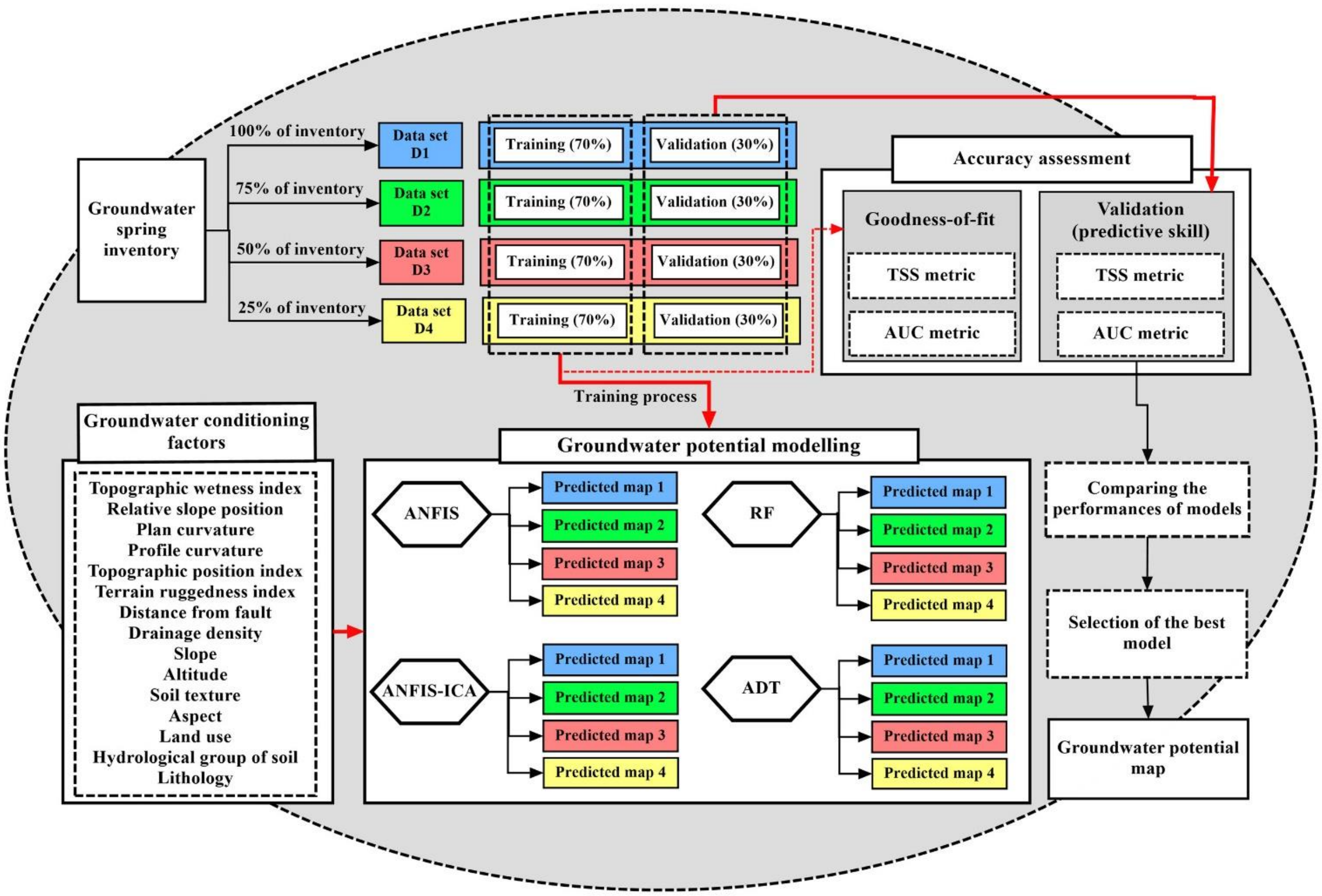
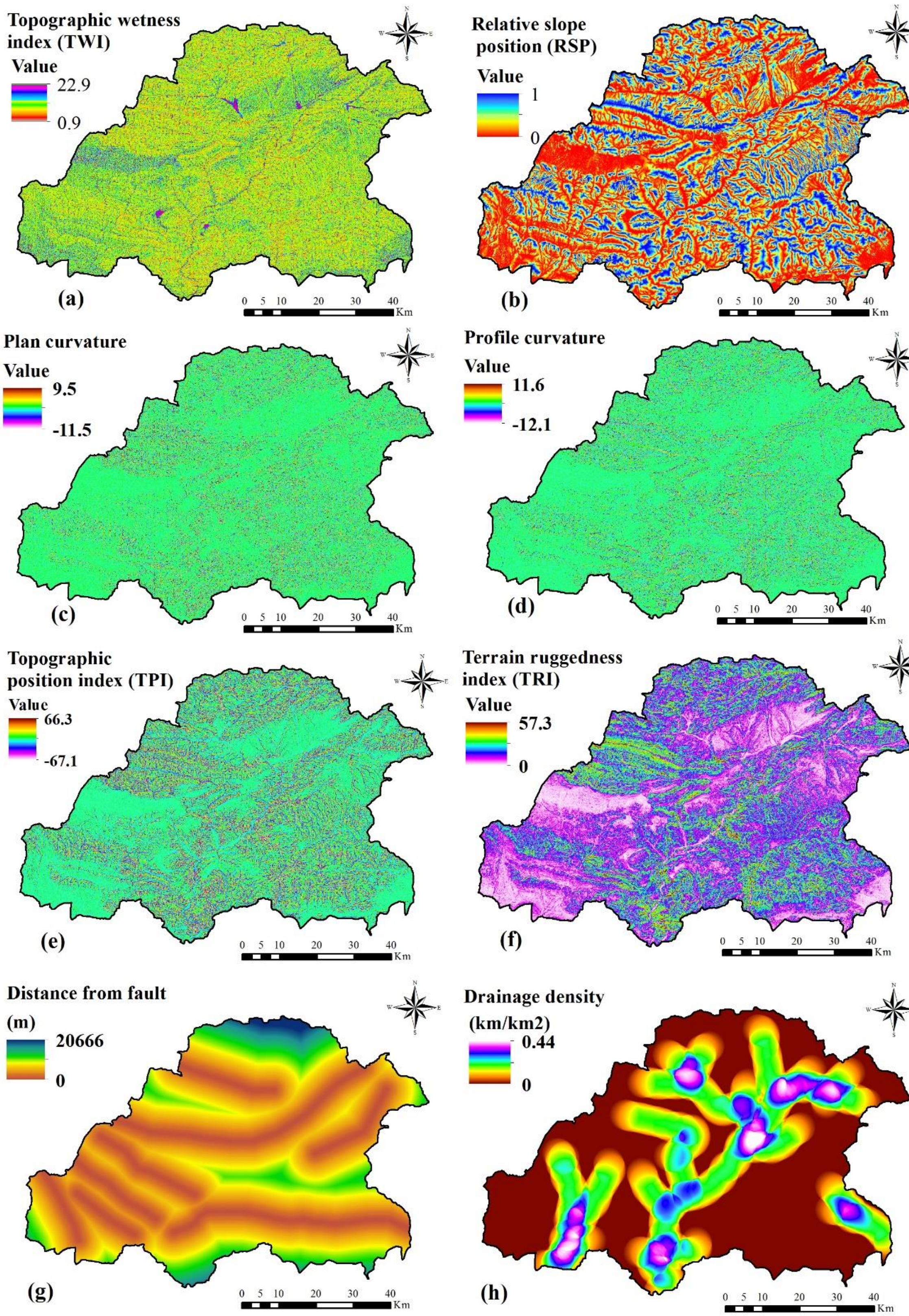
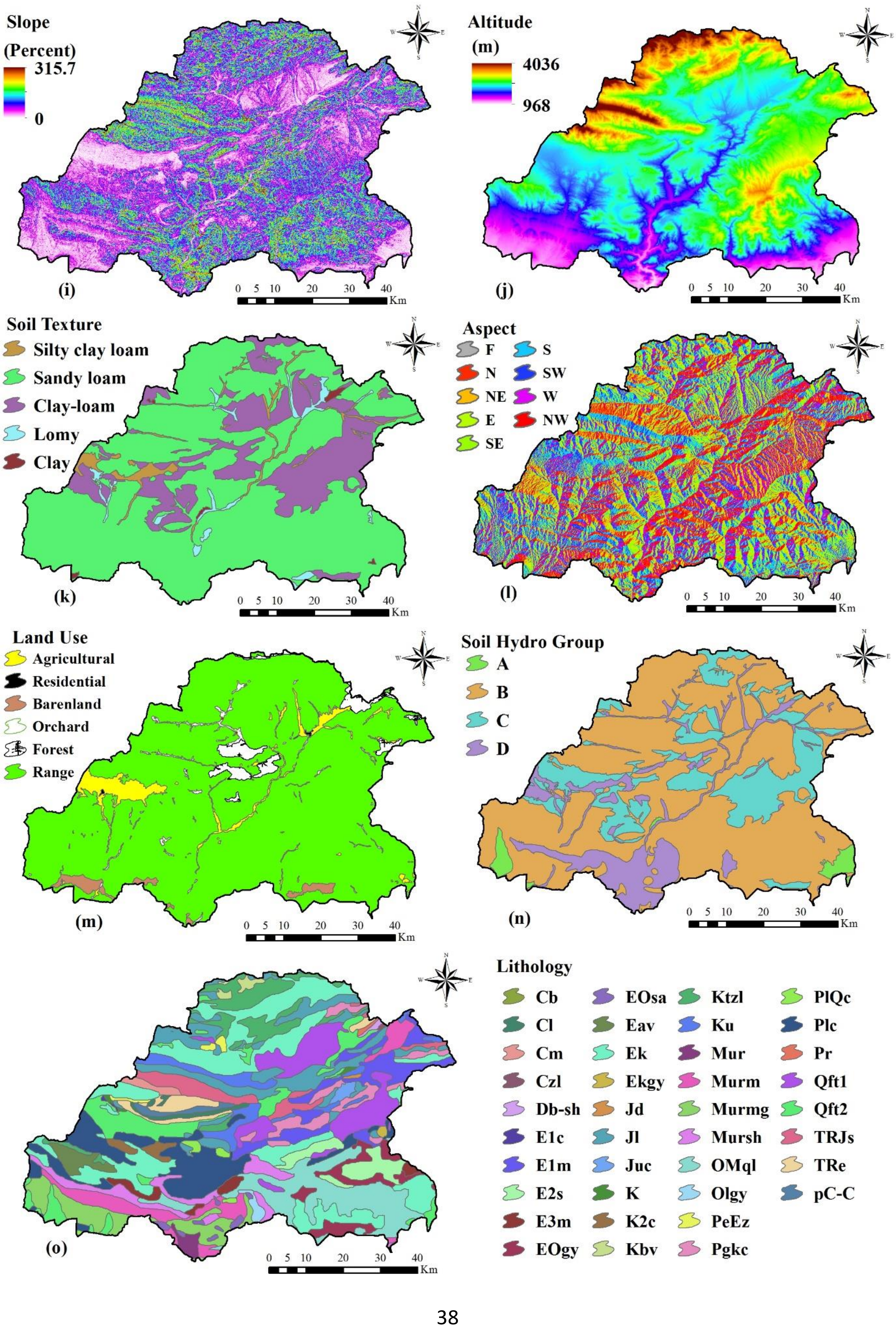


Fig. 3 Methodological flowchart adopted in this study


Topographic wetness index (TWI)
Value
22.9
0.9
(a)
Relative slope position (RSP)
Value
1
0
(b)
Plan curvature
Value
9.5
-11.5
(c)
Profile curvature
Value
11.6
-12.1
(d)
Topographic position index (TPI)
Value
66.3
-67.1
(e)
Terrain ruggedness index (TRI)
Value
57.3
0
(f)
Distance from fault
(m)
20666
0
(g)
Drainage density
(km/km2)
0.44
0
(h)
0 5 10 20 30 40 Km

790

Fig. 4 Controlling factors: a) topographic wetness index, b) relative slope position, c) plan curvature, d) profile curvature, e) topographic position index, f) terrain ruggedness index, g) distance from fault, h) drainage density, i) slope, j) elevation, k) soil texture, l) aspect, m) land use, n) hydrological group of soil, and o) lithology.



Slope
(Percent)
315.7
0
(i)
0 5 10 20 30 40
Km
Altitude
(m)
4036
968
(j)
0 5 10 20 30 40
Km
Soil Texture
Silty clay loam
Sandy loam
Clay-loam
Lomy
Clay
(k)
0 5 10 20 30 40
Km
Aspect
F
S
N
SW
NE
W
E
NW
SE
(l)
0 5 10 20 30 40
Km
Land Use
Agricultural
Residential
Barenland
Orchard
Forest
Range
(m)
0 5 10 20 30 40
Km
Soil Hydro Group
A
B
C
D
(n)
0 5 10 20 30 40
Km
Lithology
Cb
Cl
Cm
Czl
Db-sh
E1c
E1m
E2s
E3m
EOgy
EOsa
Eav
Ek
Ekgy
Jd
Jl
Juc
K
K2c
Kbv
Ktzl
Ku
Mur
Murm
Murmg
Mursh
OMql
Olgy
PeEz
Pgkc
PlQc
Plc
Pr
Qft1
Qft2
TRJs
TRe
pC-C
(o)
0 5 10 20 30 40
Km

Fig. 4 (continued)

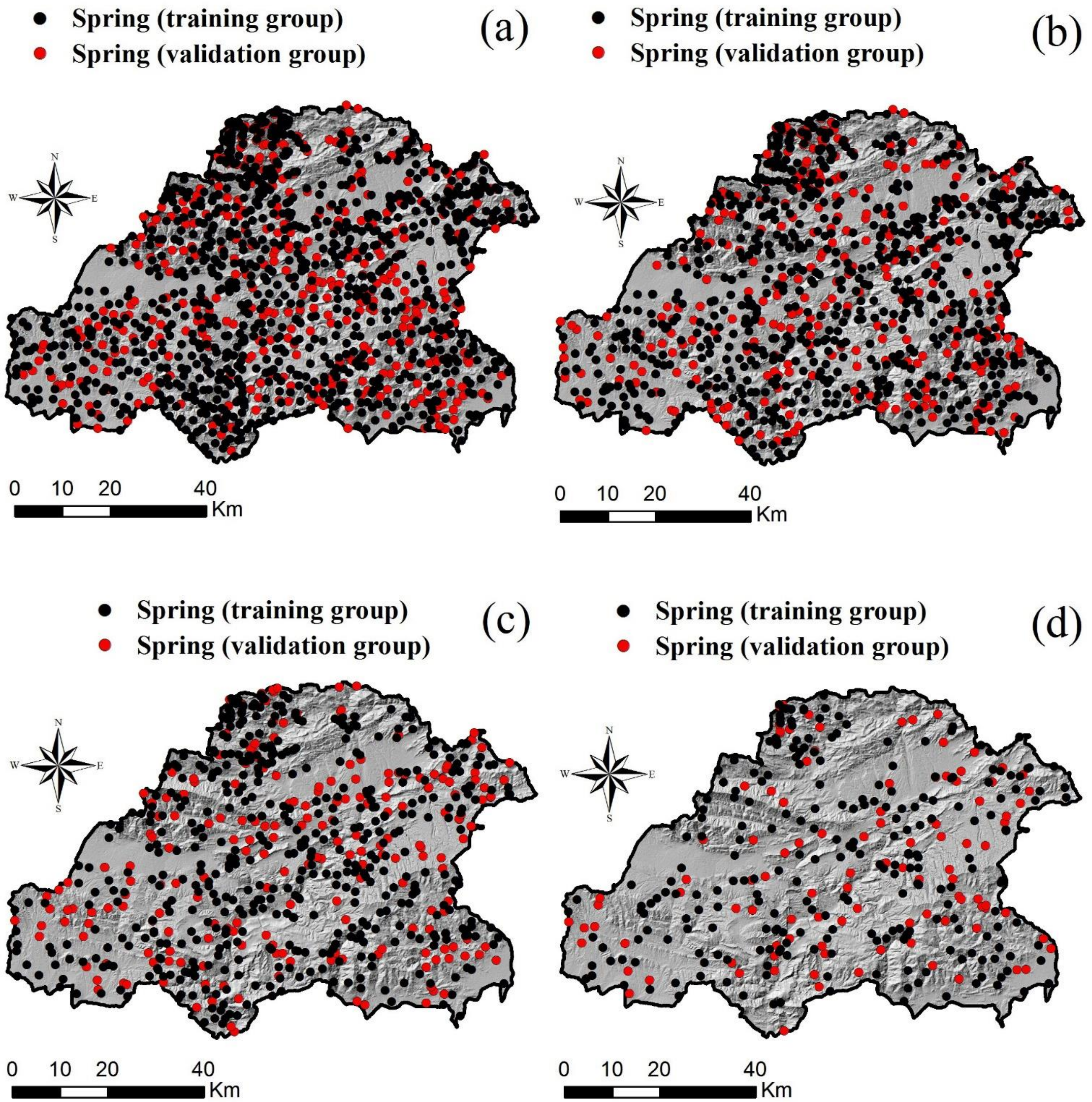
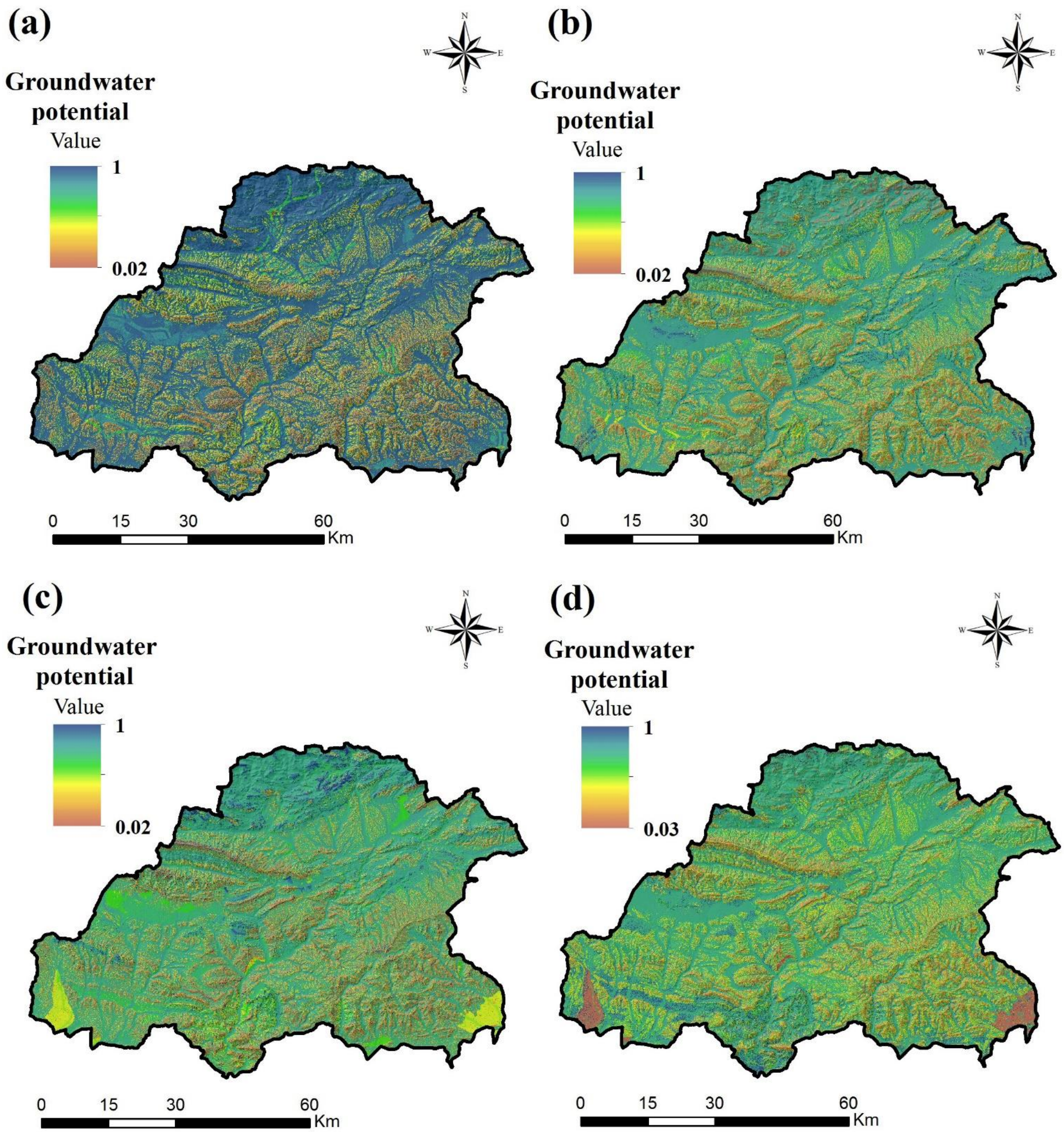


Fig. 5 Four spring sample data sets: a) data set D_1 (100% of springs), b) data set D_2 (75% of springs), c) data set D_3 (50% of springs), and d) data set D_4 (25% of springs).



(a)
Groundwater potential
Value
1
0.02
0 15 30 60 Km
(b)
Groundwater potential
Value
1
0.02
0 15 30 60 Km
(c)
Groundwater potential
Value
1
0.02
0 15 30 60 Km
(d)
Groundwater potential
Value
1
0.03
0 15 30 60 Km

Fig. 6 Groundwater potential maps generated by the ANFIS model built using: a) data set D_1, b) data set D_2, c) data set D_3, and d) data set D_4.

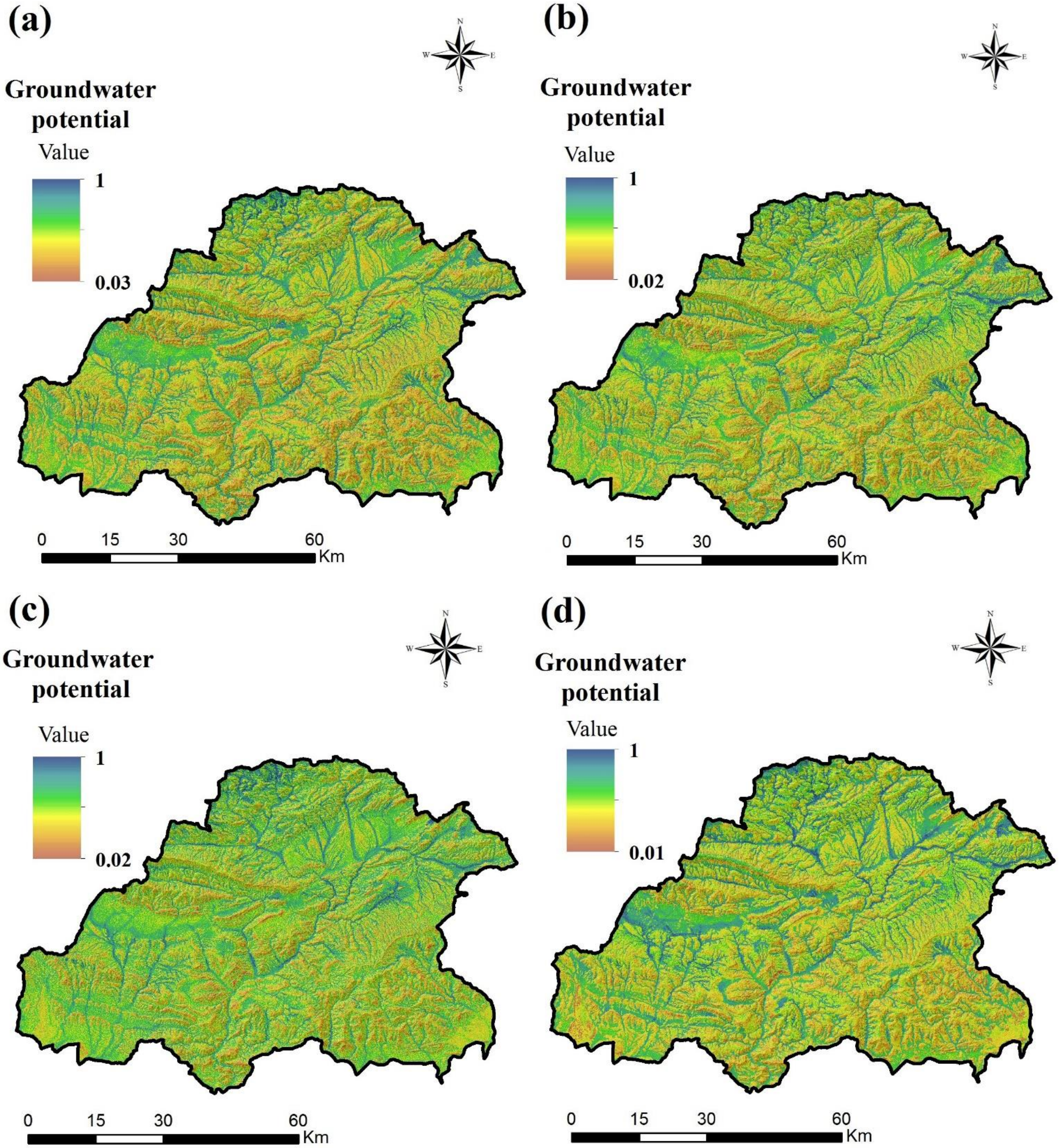


Fig. 7 Groundwater potential maps generated by the ANFIS-ICA model built using: a) data set D_1, b) data set D_2, c) data set D_3, and d) data set D_4.

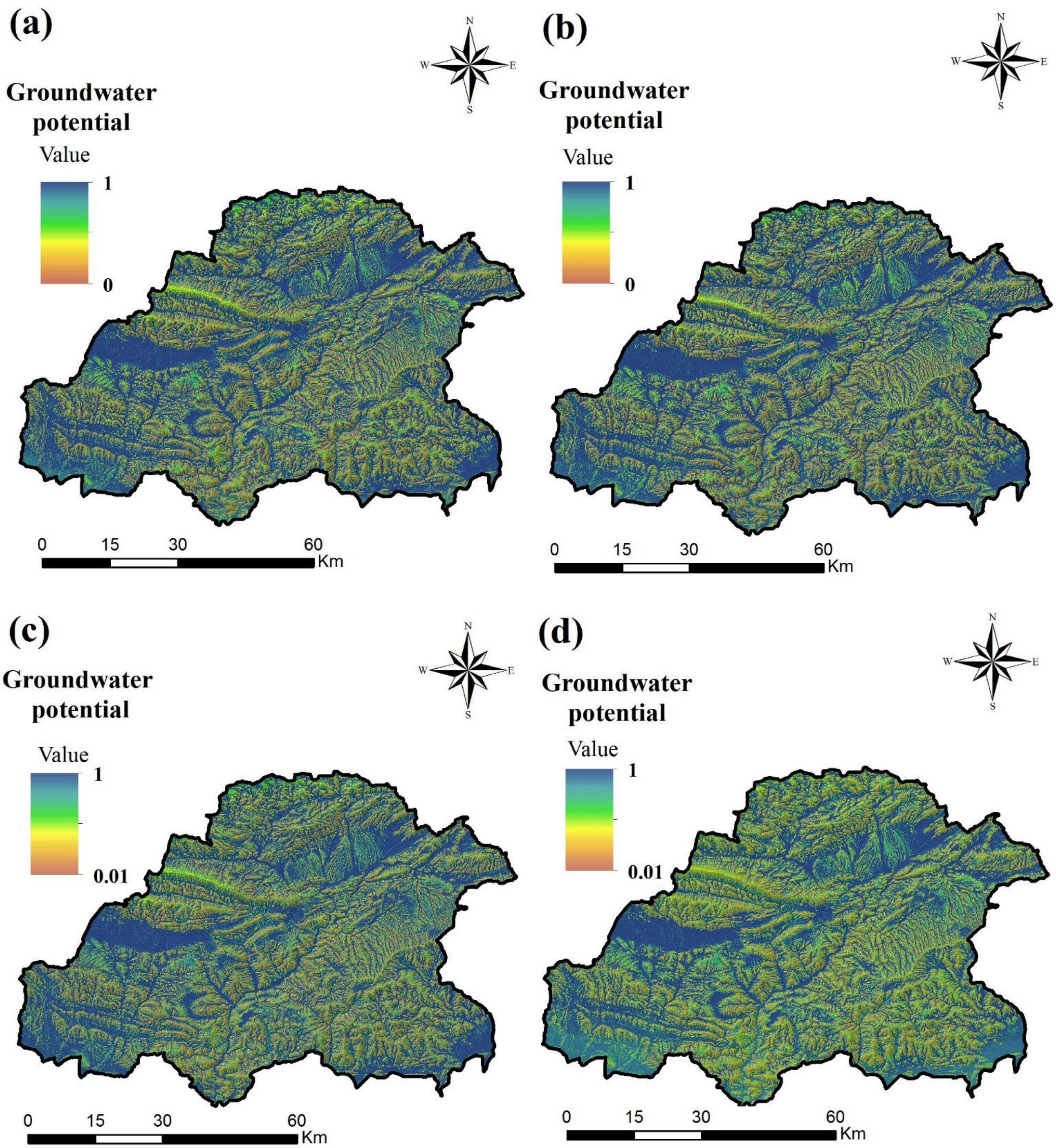

(a)
Groundwater potential
Value
1
0
0
15
30
60
Km
(b)
Groundwater potential
Value
1
0
0
15
30
60
Km
(c)
Groundwater potential
Value
1
0.01
0
15
30
60
Km
(d)
Groundwater potential
Value
1
0.01
0
15
30
60
Km

Fig. 8 Groundwater potential maps generated by the RF model built using: a) data set D_1, b) data set D_2, c) data set D_3, and d) data set D_4.

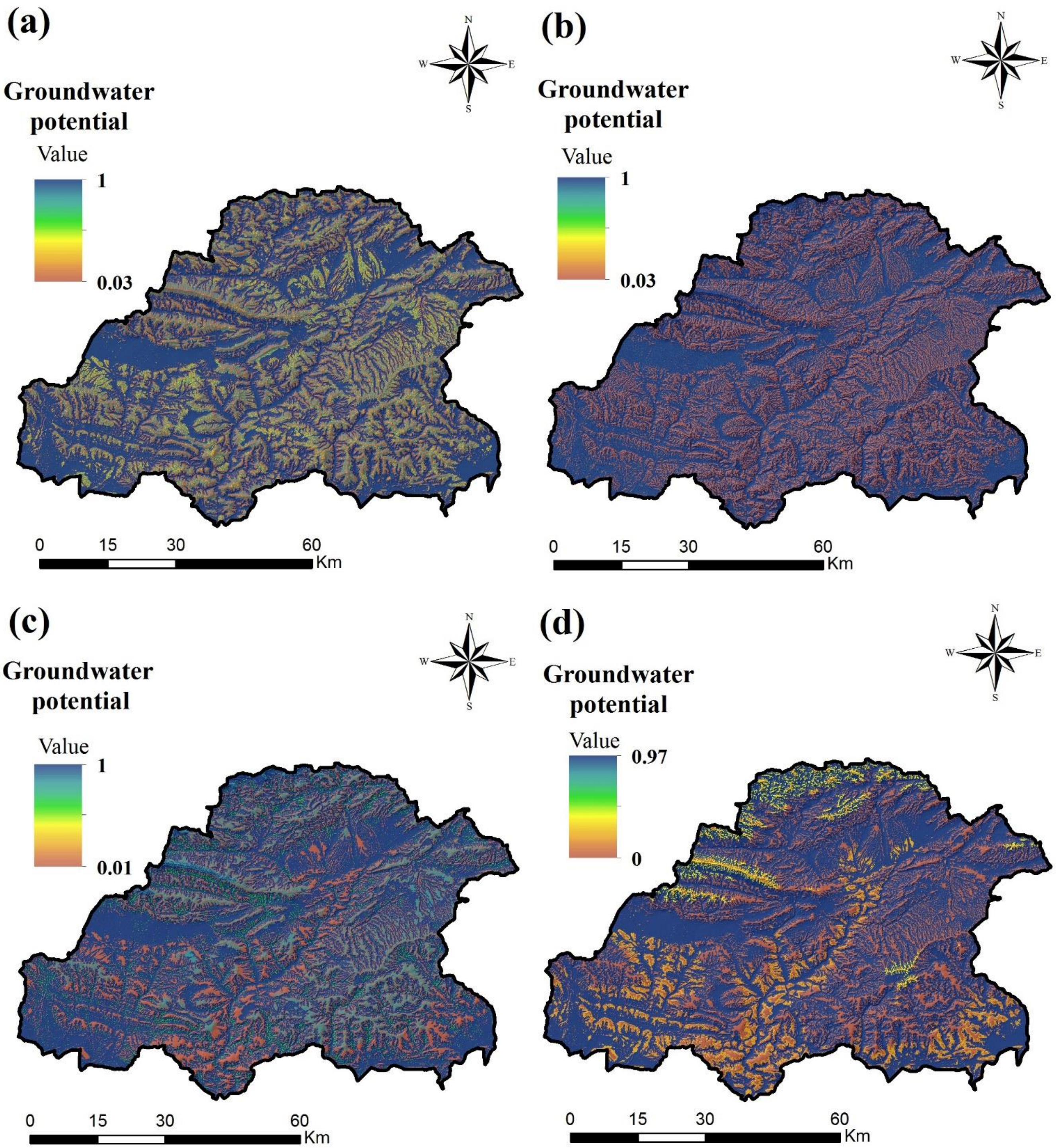


Fig. 9 Groundwater potential maps generated by the ADT model built using: a) data set D_1, b) data set D_2, c) data set D_3, and d) data set D_4.

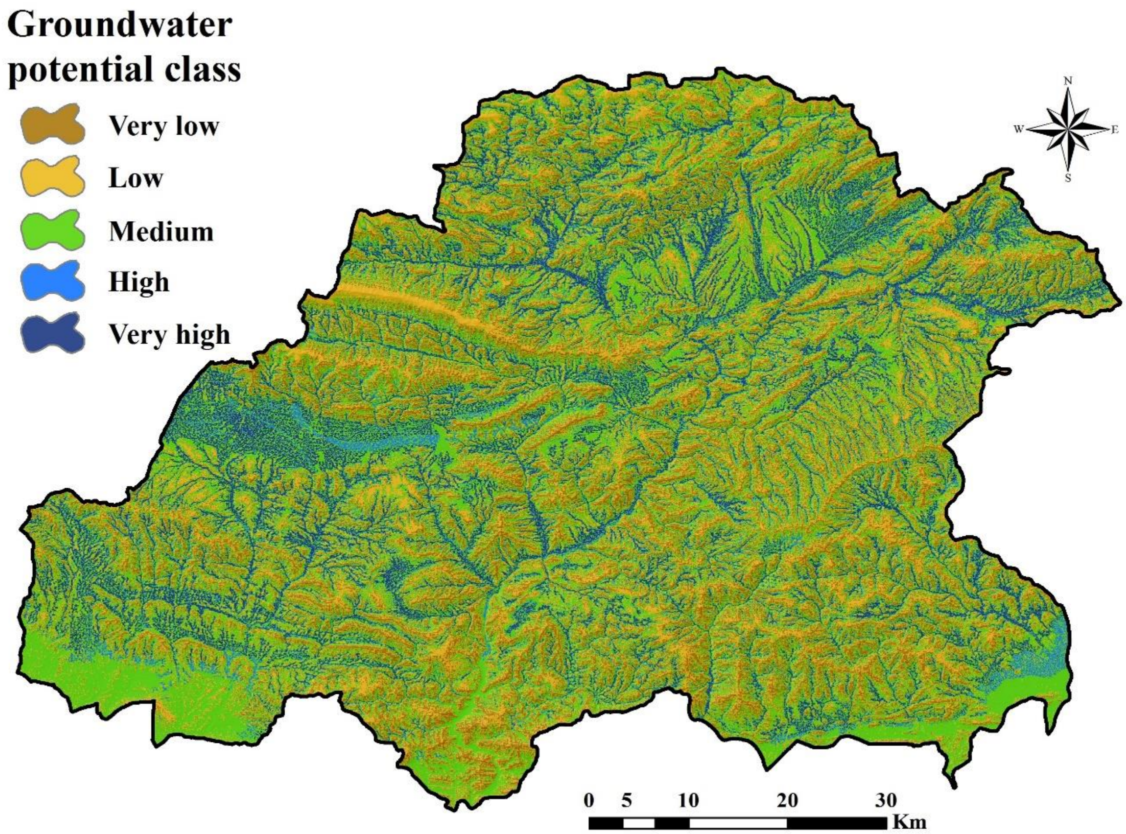


Fig. 10 Groundwater potential classes produced by the RF model (data set D_1).