Cold Regions Science and Technology 122 (2016) 26–35

Contents lists available at ScienceDirect

Cold Regions Science and Technology

journal homepage: www.elsevier.com/locate/coldregions

Comparison of artificial neural network and decision tree models in estimating spatial distribution of snow depth in a semi-arid region of Iran

Samaneh Gharaei-Manesh [a,1], Ali Fathzadeh [b,*], Ruhollah Taghizadeh-Mehrjardi [b,2]

[a] Yazd University, Natural Resources Faculty, P. O. Box: 8961719311, Iran
[b] Ardakan University, College of Agr. and Natural Resources, P. O. Box: 8951656767, Iran

ARTICLE INFO

Article history:
Received 19 December 2014
Received in revised form 13 September 2015
Accepted 2 November 2015
Available online 12 November 2015

Keywords:
Snow depth
Cubist
Decision tree
Terrain parameters
Artificial neural networks

ABSTRACT

There is no doubt that snow cover plays an important role in the hydrological cycle of mountainous basins. Therefore, it is essential to measure snow parameters such as snow depth and snow water equivalent in these areas. The aim of this study is to estimate the snow depth from terrain parameters in the Sakhvid Basin, Iran using artificial neural networks (ANNs) and M5 algorithm of decision tree. For this purpose, snow depths were measured in 206 sites based on systematic network. Furthermore, 30 terrain parameters were extracted from a digital elevation model (DEM) of the basin. The results indicated that the decision tree model is the most suitable method to estimate snow depth in the study area with a Nash–Sutcliffe Efficiency (E_{ns}) of 0.80, followed by ANNs with an E_{ns} of 0.73. Moreover, the most significant parameters in the M5 decision tree algorithm are: channel network base level, stream power, wetness index and height.

1. Introduction

Snowfall makes up a significant part of total annual precipitation in a variety of high latitude areas (Marofi et al., 2011). Measurement of the amount of water stored in the snowpack is essential for management of water supply (Shi and Dozier, 2000). In recent years, there is an urgent requirement to predict the snowpack. This is not only because of a rising demand for fresh water, but also due to the concerns about the effects of climate change (Gleick, 1993; López-Moreno et al., 2009). Climate change is likely to change the snow cover area and alter the water availability in the future making long term water management more challenging (Khadka et al., 2014). Simona et al. (2015) analyzed the snow depth and snowfall data in the western Italian Alps during the period of 1961–2010 and showed a significant decrease of snow depth in all the stations over seasonal time scale. The earth warming is ascribable to the threat posed by climate change and snow accumulations throughout the world. This requirement in arid and semi-arid regions such as Yazd Province in Iran with seasonal snowfall is apparent. In the Sakhvid basin of Yazd Province, although snow events may take place only once or twice a year, this small quantity of snowfall has a principle function to drinking water supplies of downstream regions (Yazd Regional Water Authority, 2015).

In order to analyze, quantify, and model the snowmelt runoff, it is necessary to account for spatial differences in snow accumulation (Luce et al., 1998; Seyfried and Wilcox, 1995). Generally, the spatial resolution of snow data and in situ observations of snow distribution are sparse and poor (Tarboton et al., 2000). Because of the enormous spatial variability of snow properties (i.e. snow depth and snow water equivalent), these few snow samples may not be illustrative of spatial patterns (Elder et al., 1991). Snow accumulation is heterogeneous, and once on the ground, the snow may be redistributed by some secondary agents such as wind, avalanching and sloughing (Blöschl et al., 1991; Elder et al., 1991; Kind, 1981).

Snow depth (SD) is an important variable in climate and hydrological model simulations (Dressler et al., 2006; Gong et al., 2007). Snow depth presents an extra dimension for snow cover studies by providing information relevant to water resources, soil processes, moisture and energy balance, and ecosystems (Dyer and Mote, 2006). As was previously mentioned, since the number of snow survey stations in mountainous areas is inadequate, employing a prediction technique to overcome this deficiency is necessary. In parallel with the research into snowpack, new modeling approaches, such as machine learning (ML) are emerging. ML approaches (i.e. artificial neural networks (ANNs), fuzzy logic, decision trees, support vector regression) are being employed in all fields of water resource sciences as an alternative to traditional methods (i.e. regression or auto-correlation-based statistical method such as ARIMA) to find a functional relationship between input (terrain attributes) and output (snow depth) variables. (Deka, 2014; Nourani et al., 2014).

* Corresponding author. Tel.: +98 35 35240910; fax: +98 35 32226767.
E-mail addresses: samaneh.gharaei@gmail.com (S. Gharaei-Manesh), fatzade@ardakan.ac.ir (A. Fathzadeh), rtaghizadeh@ardakan.ac.ir (R. Taghizadeh-Mehrjardi).
[1] Tel./fax: +98 35 38210312.
[2] Tel.: +98 35 32240911; fax: +98 35 32226767.

http://dx.doi.org/10.1016/j.coldregions.2015.11.004
0165-232X/

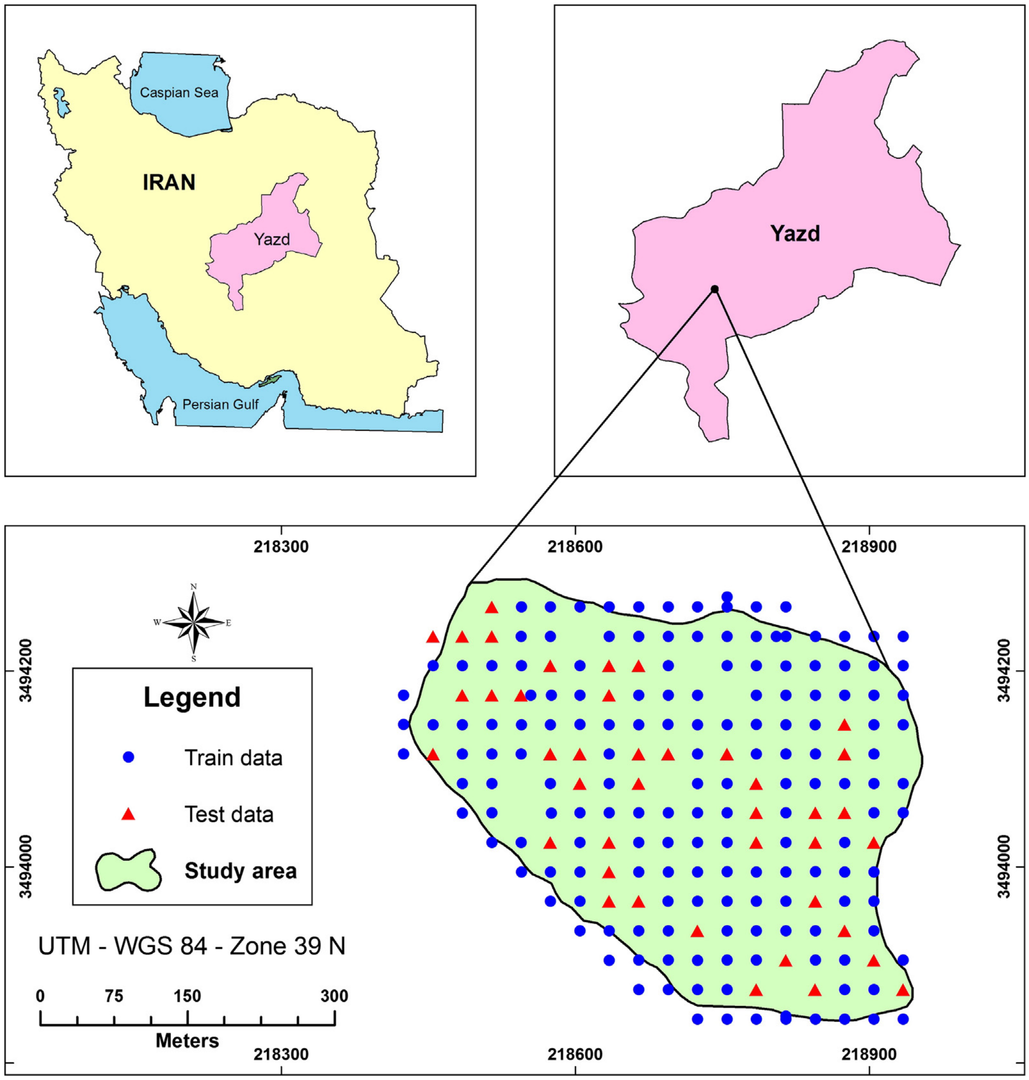

Fig. 1. The study area and location of measured data and spatial distribution of testing and training points.

ANNs are one of the ML algorithms that have been widely applied in hydrology science (Govindaraju, 2000). An ANN model can overcome large scale complex problems such as non-linear modeling, classification and association by learning and generalizing the knowledge from adequate pairs of data (Govindaraju, 2000). There is no need to have the knowledge about the physical process being modeled by the ANN technique (Nourani et al., 2011). Therefore, due to these features of ANNs, they are suitable methods for prediction in hydrologic science. Although, the ANN methods are applied widely in predicting hydrological variables, they have some difficulties. For instance, using trial and error method in order to detect the number of hidden layers and neurons is time consuming. Furthermore, the regression based models are black box models (Etemad-Shahidi and Mahjoobi, 2009).

Another type of ML algorithms is the model tree (Quinlan, 1992) that produces binary decision trees and is a spread of regression trees (Etemad-Shahidi and Mahjoobi, 2009). According to some researches, the advantages of decision tree against neural networks are that they demonstrate clear rules and can be trained faster. The rules are simple and they can be easily understood. In addition, the model tree does not require the optimization of geometry and internal network (Etemad-Shahidi and Mahjoobi, 2009; Solomatine and Xue, 2004).

Balk and Elder (2000) in their study modeled the spatial distribution of snow using binary decision tree and geostatistical techniques in Loch Vale Watershed (LVWS), Rocky Mountain National Park, Colorado. The

Table 1
The terrain parameter used for both of the ANN model and the model tree.

No.	Parameters	No.	Parameters
1	Longitude (X)	16	Strahler order
2	Latitude (Y)	17	Stream power
3	Slope	18	Flow accumulation
4	Slope length	19	Flow direction
5	Mid-slope position	20	Flow connectivity
6	Ls factor	21	Analytical hill shading
7	Catchment slope	22	Aspect
8	Slope height	23	Convergence index
9	Height	24	Catchment area
10	Normalized height	25	Modified catchments area
11	Curvature	26	Wind effect
12	Plan curvature	27	Multi resolution index of valley bottom flatness
13	Profile curvature	28	Multi resolution ridge top flatness index
14	Valley depth	29	Altitude above channel network
15	Wetness index	30	Channel network base level

result showed an improvement over previous approaches in predicting the distribution of snow water equivalent (SWE) in mountainous basins. Tedesco et al. (2004) in their study also used the ANNs, spectral polarization difference (SPD) algorithm, the Helsinki University of Technology (HUT) model-based iterative inversion and the Chang algorithm to predict the SWE and snow depth from Special Sensor Microwave Imager (SSM/I) data and observed series across Finland. It was concluded that the ANN model performance with the observed data produced the most accurate predictions. Furthermore, Bhattacharya and Solomatine (2005) in their study investigated the relationship between water level and discharge using ANNs and M5 model tree in Swarupgunj on the river Bhagirathi, India. The result indicated that the ANNs and the M5 model tree performed substantially better than the traditional models such as polynomial regression and auto-correlation-based statistical method (i.e. ARIMA).

Erickson et al. (2005) applied a geostatistical approach with a complex variable mean to model the spatial distribution of snow depth over a wind-dominated alpine basin, Green Lakes Valley watershed in Colorado. The terrain parameters including elevation, slope, and potential radiation, an index of wind sheltering and wind drifting were supposed. They concluded that all parameters affected the snow depth substantially when there is a non-linear interaction between the mentioned parameters.

Tabari et al. (2010) in their study compared the snow depth and snow water equivalent estimated by some geostatistical and artificial intelligence methods in the Samsami basin of Iran. The result showed that an artificial intelligence method i.e. neural network–genetic algorithm (NNGA) and ANNs provided the best results among the other models. Marofi et al. (2011) also studied SD and SWE to detect the amount of water stored in the snow in the Samsami basin, Iran. In this regard, they applied a multivariate non-linear regression (MNLR) method, and four types of ANNs and a neural network–genetic algorithm (NNGA) model. They concluded that the NNGA model gave the best performance in terms of estimating SWE in the studied area.

Although, the M5 model tree has shown good performances in several hydrologic studies such as rainfall–runoff modeling (Solomatine and Dulal, 2003), flood forecasting (Solomatine and Xue, 2004), modeling water level discharge relationship (Bhattacharya and Solomatine, 2005), and sediment transport (Bhattacharya et al., 2007), it has not been given enough attention in snow related investigations.

Wang and Witten (1997) in their investigation reconstructed and improved the M5 model in a system called M5′. The M5′ model tree is an effective learning method for predicting real values. Similarly, the Cubist model, an advanced version of M5, is a data mining algorithm which allows one to explore non-linear relationships in observed data (Minasny and McBratney, 2008). The primary objective of this research is to analyze how much the new ML algorithms can improve the prediction of spatial distribution of SD and secondary, according to the high quantity of terrain parameters, which one of them have the most influence on snow distribution.

2. Methodology

2.1. Study area and data

The Sakhvid Mountain is located in the southern part of Yazd province, Iran (53.84–53.93°E, 31.58–31.67°N) and covers the area of 92.5 km^2. The average annual temperature and rainfall account for 14 °C and 222.8 mm, respectively. Most parts of the Sakhvid area are mountainous and the elevation ranges from 2840 to 2990 m a.s.l. The snow accumulation in the Sakhvid Mountain is a considerable part of water resources in Yazd province. The area of 16 ha, as the representative area of the basin, was chosen in order to study the snow depth. The SD parameter can be obtained by sampling at different sites. The sites may be selected based on physical features of the catchment, executive limitations and funds (Marofi et al., 2011). In this study, the sample sites were selected based on topographic features of the area, existence of meteorological stations and accessibility. The SD measurements were made by Mt. Rose Snow Sampler in 206 points with 30 m distance based on systematic network. Fig. 1 shows the study area with the location of measuring points. Because of the quick change of SD data in the arid regions, the period of snow survey was short. In this study, the field work was carried out within an interval of 3 days in February, 2012.

Considering the impact of terrain parameters on SD, these parameters were utilized as the auxiliary data. Terrain parameters were calculated directly by the analysis of a digital elevation model (DEM) with a 20 m grid cell size. The DEM used in this study was originally prepared

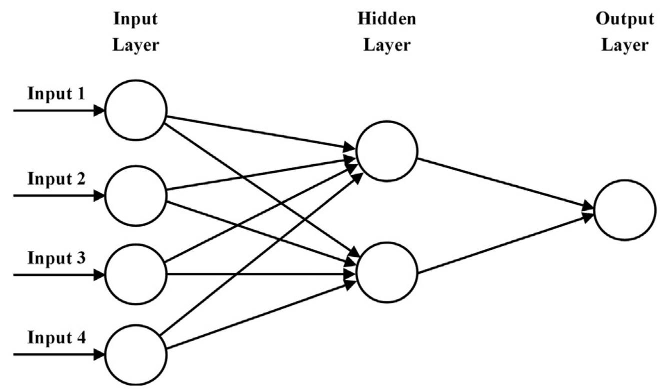

Fig. 2. The typical structure of ANNs model.

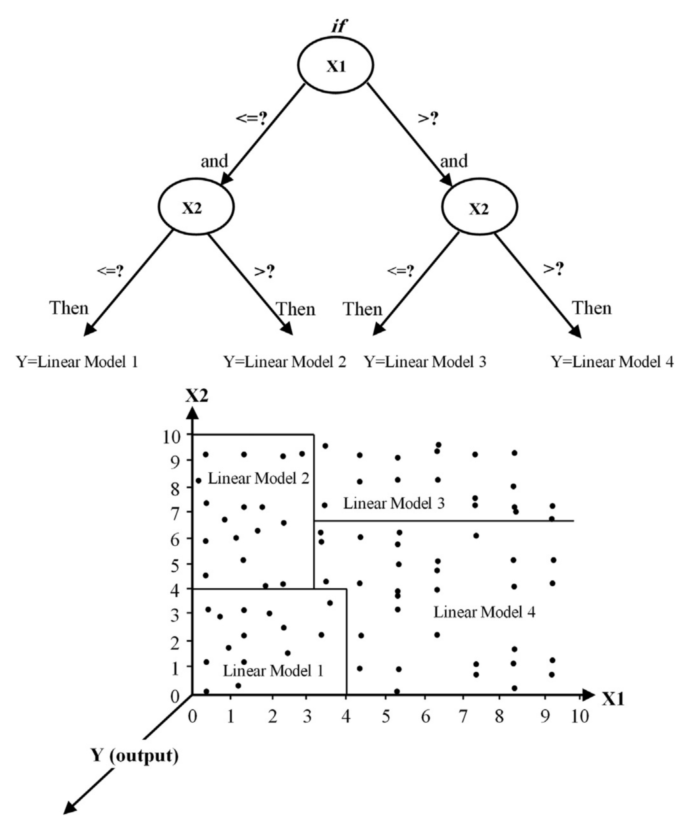

Fig. 3. An example of tree-building process in M5 model trees, dividing the domain of X1 × X2 into four sub-domains.

Table 2
Summary statistics of measured snow depth.

	Mean (cm)	Max (cm)	Min (cm)	Range (cm)	Standard deviation (cm)	Coefficient of variation (%)	Kolmogorov–Smirnov
Snow depth	53.54	114	17	97	20.9	38.5	0.614

from RADAR images. The raw DEM contains a number of anomalies – and by all accounts, is a 'noisy' dataset – as a result of the data collection and method of preparation. This in turn affected the derivation of the terrain attributes from the DEM. Therefore, in this study, the simple filtering method was used for removing noise from the terrain attributes. Then, the System for Automated Geo-scientific Analysis (SAGA) was applied to calculate terrain parameters. Table 1 shows the extracted parameters from the DEM map. The next step was to select the most appropriate auxiliary data to reduce the dimensionality but also allow learning algorithms to operate more effectively. Moreover, irrelevant and redundant information may decrease the prediction accuracy in common machine learning algorithms (Hall, 1997; Hall et al., 2009; Mollazade et al., 2012). Different techniques can be used to rank the relevance of auxiliary variables, including correlation-based feature selection (CFS), principal component analysis (Omid et al., 2010), factor analysis, and sensitivity analysis. Here, a correlation-based feature selection (CFS) was applied using the CfsSubsetEval algorithm available in the WEKA software package (Hall et al., 2009). Correlation-based feature selection is a fully automatic algorithm, not requiring predefined thresholds or number of features. The algorithm ranks auxiliary data according to a correlation based heuristic evaluation function, retaining relevant auxiliary data that are highly correlated and soil classes. Irrelevant data, with low correlations, were screened out. Correlation-based feature selection typically eliminated over half of the features. In our case, CFS algorithm reduced the size of covariates from 30 to 16 layers, including mid-slope position, slope, height, normalized height, curvature, plan curvature, profile curvature, wetness index, stream power, analytical hill shading, aspect, wind effect, multi resolution index of valley bottom flatness, multi resolution ridge top flatness index, altitude above channel network and channel network base level. 80% of the data was utilized to train both the ANNs and decision tree models and the remaining data employed for validating the models. Fig. 1 shows the location of testing and training points.

2.2. Artificial neural networks (ANNs)

ANNs are mathematical models, which try to copy the parallel local computing system of the human brain in the simplest way (Huo et al., 2012). The most common type of ANNs is a feed-forward back-propagation (FFBP) neural network. The network contains interconnecting nodes called neurons that are connected to each other through weighted synapses. In order to teach an ANN model, first, the random initial values of weights are given to the synapses. Then, these values are progressively corrected during a training phase. Afterward, the computed outputs of the network are compared with the real values. Finally, the errors are back-propagated to adjust the values of the weights in order to minimize the errors (Kisi, 2005). In this study, an FFBP neural network with one hidden layer was utilized. The network was trained through back-propagation Levenberg–Marquardt algorithm (Bernd et al., 1999). The number of neurons in the hidden layer was obtained by trial and error method. A typical structure of the ANN model is shown in Fig. 2. In the present study, the NeuroSolution 5 software (NeuroDimension, Inc.) was employed to establish the FFBP neural network.

2.3. M5 model tree algorithm

Fig. 3 illustrates the procedure of the tree build, where the space is split into four subspaces with each possessing a linear regression model. At the first step, the M5 model tree algorithm partitions the parameter space to make the basic tree based on the splitting criterion, which is the standard deviation reduction (SDR):

$$SDR = sd(T) - \sum_{i} \frac{T_i}{T} \times sd(T_i) \quad (1)$$

where T denotes the set of examples that reach the node, T_i are the sets that result from splitting the note according to the chosen attribute and Sd is the standard deviation (Wang and Witten, 1997).

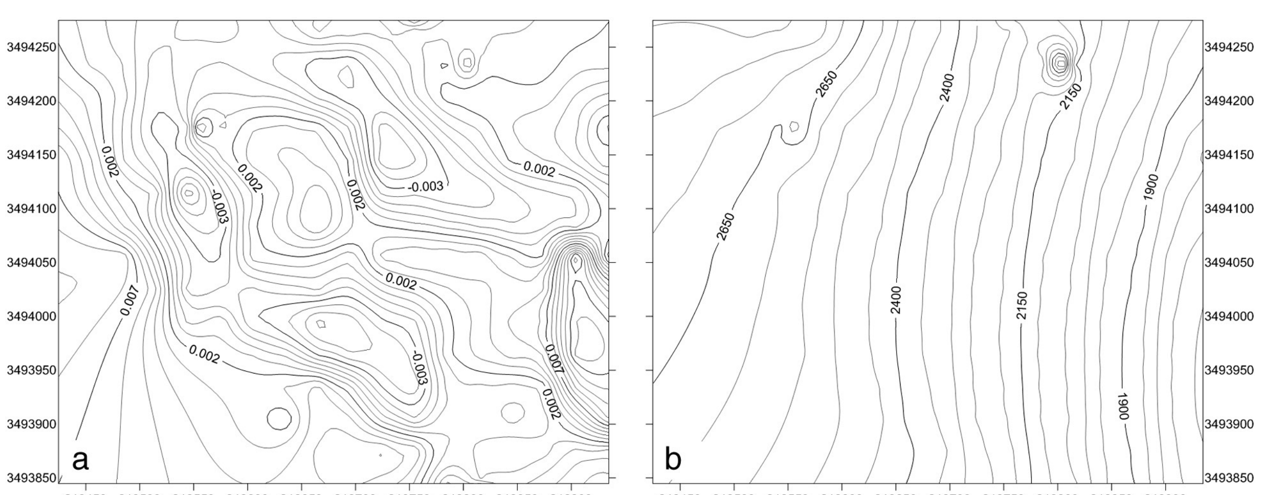

Fig. 4. Spatial distribution of two terrain parameters derived from DEM a) plane curvature and b) channel network based level.

Table 3
Evaluation of model performances using error criteria.

Efficiency coefficient	r (%)	RMSE (cm)	Bias (cm)	MAE (cm)	E_{ns}
M5 algorithm	89	8.7	0.53	7	0.80
Artificial neural network	85	12.1	−2.9	10	0.73

The M5 algorithm employed standard deviation as the amount of error of the values that reach a node. All attributes were examined at a node and the expected reduction in error was computed. Then, the attribute that maximized the SDR was selected. The process of splitting ceases when a slight variation is observed in the value of all instances that reach a node or only a few instances stand (Etemad-Shahidi and Mahjoobi, 2009). Next, a linear regression model was introduced for each subspace. The regression model uses only the related data to the branches examined in that subspace. The growing of tree raises the performance of the model, while it could lead to overfitting of the problem. This problem can be solved by pruning the tree, which merges a few lower subspaces. The tree is pruned if this process results in a lower expected estimated error.

The mean of absolute differences between the predicted value and the observed value for each of the training data points is the expected error (Wang and Witten, 1997). As the expected error may be underestimated, this value is multiplied by the following equation:

$$(n+\vartheta)/(n-\vartheta) \tag{2}$$

in which n is the number of training data points that reach that node and ϑ is the number of parameters in the model that represents the value at that node (Wang and Witten, 1997). On applying the pruning process, there will be a substantial discontinuity in the neighbor linear model at the leaves of the pruned tree. The M5 algorithm gets the final model at the leaf by combining the obtained model at that leaf with the existing model on the path to the root (Wang and Witten, 1997). In this process, first the estimated value of the linear model at the leaf is filtered along the path back to the root. Then, this value is smoothed at each node by combining it with the estimated value by linear model at that node as follows:

$$P' = \frac{np+kq}{n+k} \tag{3}$$

where P' is the prediction passed up to the next higher node, p is the prediction passed to this node from below, q is the value predicted by the node at this node, n is the number of training instances that reached the node below, and k is a constant (default value is 15).

In this article, the decision tree method was performed in the Cubist data mining software. Cubist is an advanced version of regression tree. This model is based on the M5 algorithm of Quinlan that generates different models, namely; rule-based models, composite and committee models (Quinlan, 1992). In rule based models, Cubist produces a model from training data and the model consists of several rules. In some applications, the performance of rule-based models can be improved by applying the composite models. The composite models are of the most benefit when there are a small number of input parameters and all the parameters participate in the prediction of the target value. Beside the composite rule-based models are committee models, which constitute some rule-based models and can be produced by Cubist. When the initial model is sufficiently accurate, the committee model becomes most effective (Quinlan, 2001). In this study, the M5 algorithm of the decision tree was constructed using a conjugation of committee models with composite models to produce the most effective results.

In order to evaluate the performance of the applied models, the correlation coefficient (r), root mean square error (RMSE), bias, mean absolute error (MAE) and Nash–Sutcliffe efficiency (E_{ns})were computed (Eqs. (4)–(8)).

$$r = \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2(Y_i-\overline{Y})^2}} \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[Y_i-X_i]^2} \tag{5}$$

$$Bias = \overline{Y}-\overline{X} \tag{6}$$

$$MAE = \frac{\sum_{i=1}^{n}|X_i-Y_i|}{n} \tag{7}$$

$$E_{ns} = 1-\frac{\sum_{i=1}^{n}(X_i-Y_i)}{\sum_{i=1}^{n}(X_i-\overline{X})} \tag{8}$$

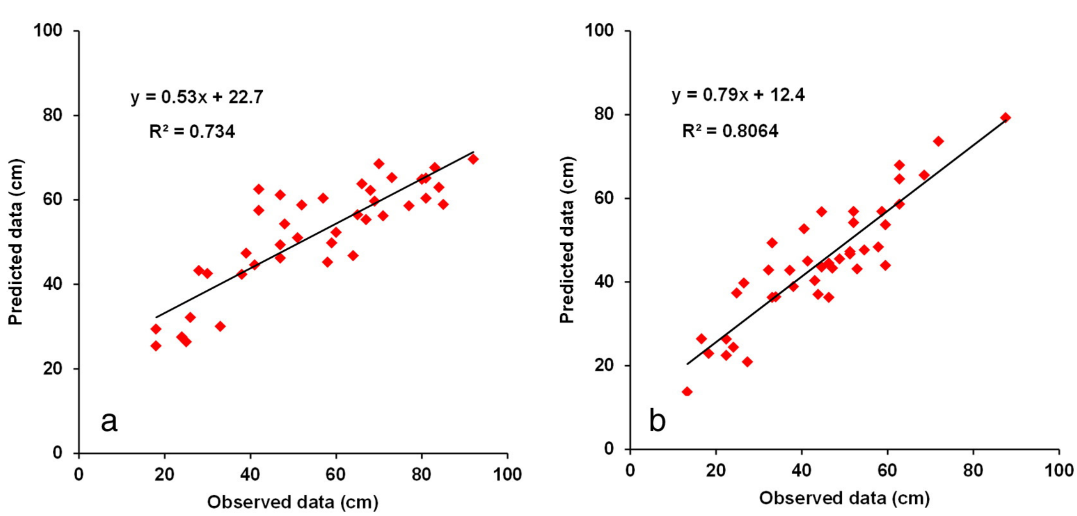

Fig. 5. Scatter plot of measured and predicted SD a) ANN and b) M5.

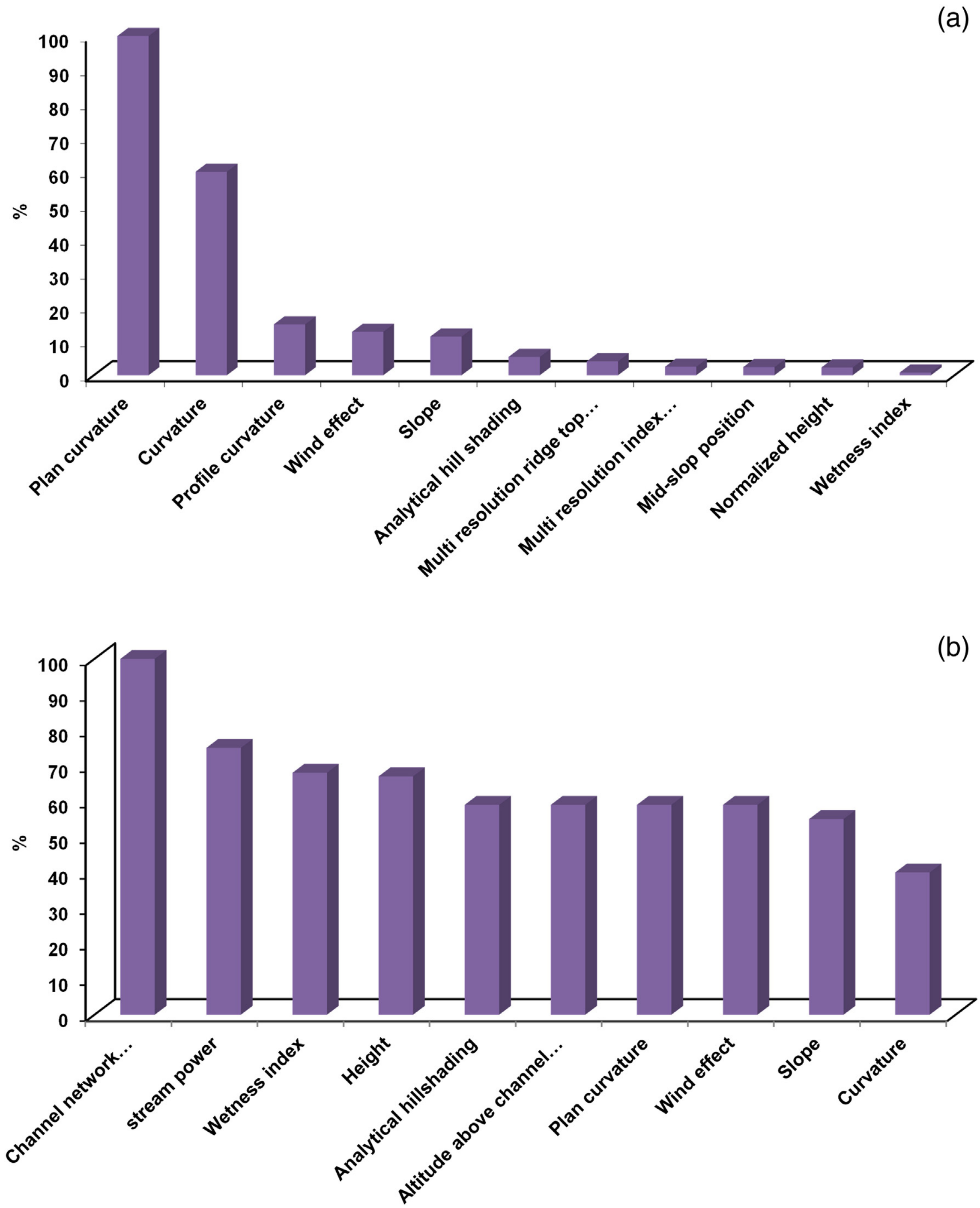

Fig. 6. The most important auxiliary data detected by a) ANN b) M5 models.

where X_i is the recorded value, Y_i is the predicted value, $\overline{X}$ denotes the average of the recorded values, $\overline{Y}$ is the average of the predicted value, and n is the length of dataset.

3. Results and discussion

3.1. Summary statistics

Relevant statistics of the SD is given in Table 2. The SD ranged from 17 to 114 cm in the study region. The average and range values of SD are 53.54 and 97 cm, respectively. The coefficient of variation for snow depth is relatively high (38.50), which indicates a wide range of values across the study area. The Kolmogorov–Smirnov test confirmed a normal distribution of the raw data.

3.2. Selection of auxiliary data

There are different techniques to rank the relevance of auxiliary variables. For the ANN model, a sensitivity analysis was applied and the decision tree auxiliary data were ordered according to their effectiveness. For example, Fig. 4a shows the changes of plan curvature on the study area which resulted from irregular changes of plan curvature over several parts of the study area while channel

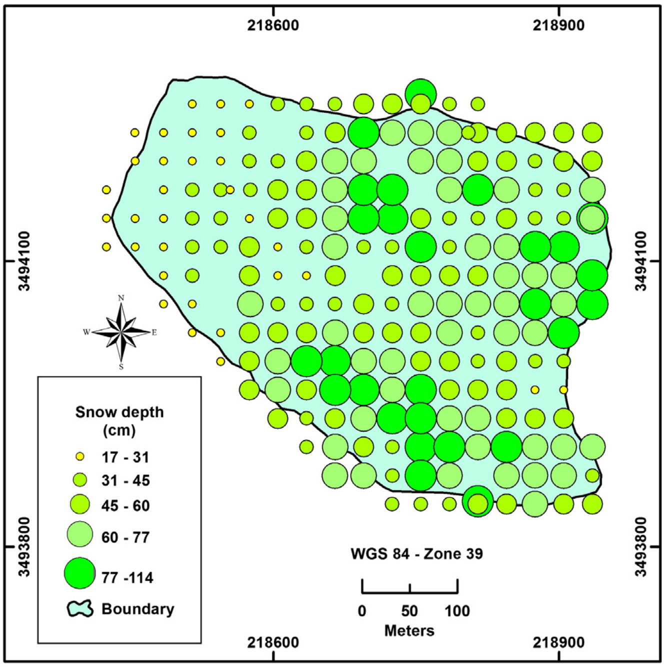

Fig. 7. Spatial distribution map of SD for observed data.

network base level has nearly a moderate and uniform change throughout the basin (Fig. 4b).

3.3. Spatial models

The results of multi-layer perceptron (MLP) to predict snow depth from 30 terrain parameters are shown in Table 3. The number of neurons in the hidden layer from 2 to 20 was tested with 80% of the sampled data and it was found that the network with 11 neurons produced the lowest RMSE, which is 12.1 cm. Therefore, the 30–11–1 structure was selected as the best architecture to forecast the snow depth. As shown in Table 3, the ANN models estimated snow depth at an acceptable level of precision with the correlation, coefficient of determination, RMSE, Bias, MAE and E_{ns} values of 85%, 0.73, 12.1 cm, −2.9 cm, 10 cm and 0.73 respectively. According to Fig. 5a, the strength of the relationship between the predicted values and the observed data is high.

Committee–composite models, which constitute several rule-based models, can be produced by Cubist. Each rule-based model (regression tree model) predicts the target value for a case and then the final predictions are obtained by averaging the all rule-based models' predictions (Quinlan, 2001). The M5 decision tree model performance with 80% of the sampled data revealed the committee models and corresponding rules as Appendix A.

The first model of a committee–composite model is always exactly the same as the model generated without the committee option. The second model is generated to correct the first model's predictions; if the predictions of the first model are too low for a case, the second model by forecasting a higher value will expiate. The third model attempts to correct the predictions of the second model, and so on. The suggested number of models is five.

The obtained models based on 80% of the sampled data were applied to estimate the snow depth for the remaining (20%) data sets. As shown in Table 3, the M5-decision tree algorithm performed better than the ANN model in terms of correlation, RMSE, Bias, MAE and E_{ns}. Fig. 5b also shows that the modeled snow depth by the decision tree fitted the observed data at a good level of accuracy.

According to the weight values given to each input parameter by the ANN model, it was found that the plan curvature was given the greatest weight and was the most effective parameter that affected the snow depth in the area. The important parameters were ordered as follows: plan curvature, curvature, profile curvature, wind effect, slope, and analytical hill shading (Fig. 6a).

One advantage of the decision tree is that it uses only the effective parameters. According to Fig. 6b, channel network base level is the most important parameter affecting snow depth by decision tree model. The significant parameters in M5 decision tree algorithm are ordered as

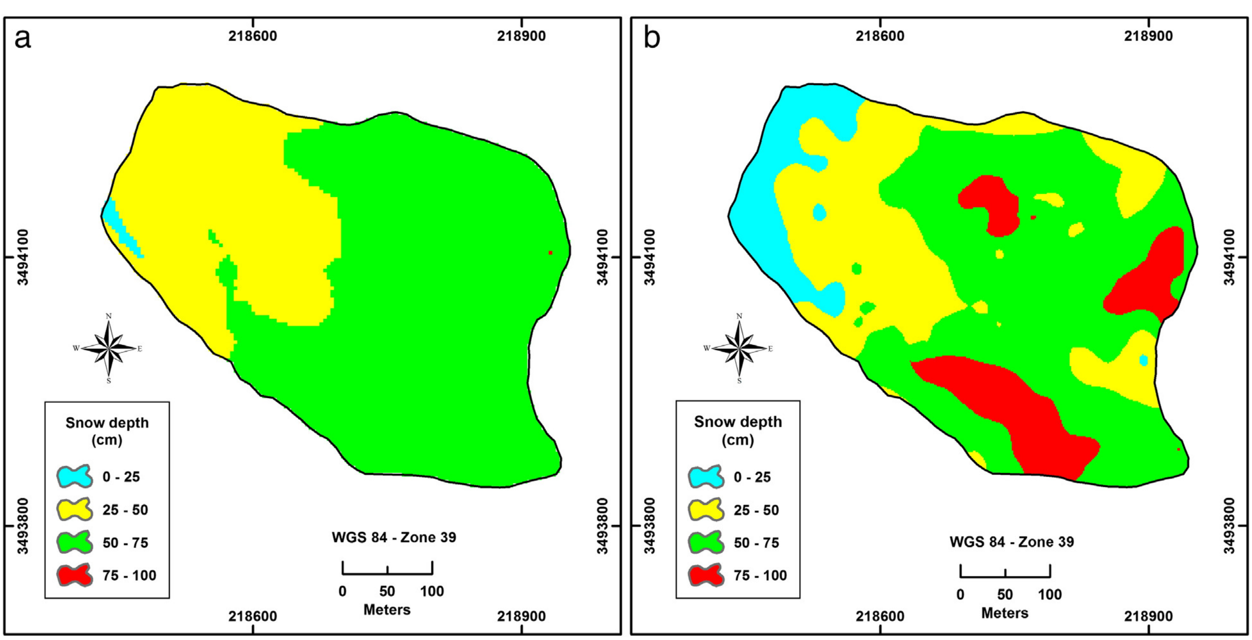

Fig. 8. Spatial distribution map of SD generated by a) ANN and b) M5 models.

follows: channel network base level, stream power, wetness index, height, analytical hill shading, altitude above channel network, plan curvature, wind effect, slope and curvature.

General findings revealed that the M5 decision tree algorithm was more appropriate for snow depth estimation in the basin. This method was able to simulate 89% of changes in snow depth, whereas, ANN model performed at correlation coefficient values of 85%. In term of E_{ns} value, the M5 decision tree with $E_{ns} = 0.80$ peformed better than the ANN model with $E_{ns} = 0.73$. Therefore, we could recommend M5 as the best model for the prediction of SD. This result is similar with the finding of other researchers who demonstrated the performances of decision tree model (Balk and Elder, 2000; Elder et al., 1995). From a statistical point of view, both algorithms M5 and ANN performed well in terms of prediction ability, and hence, can be recommended as models for spatial modeling. However, M5 is easier to interpret in comparison with neural networks (Clark and Pregibon, 1992). Another advantage of M5 is their ease of interpretation and their ability to incorporate both continuous and categorical auxiliary data (Grinand et al., 2008); which would make using this type of modeling approach suitable in places such as Iran. Furthermore, there is no concern about the number of predictors and variable selection (Grinand et al., 2008; Taghizadeh-Mehrjardi et al., 2014).

3.4. Spatial distribution

Fig. 7 illustrates the spatial distribution of observed snow depth in the area and it is clear that the highest snow depth is seen in the south, northern and eastern parts of the basin, while west of the watershed had the lowest snow depth. Similarly, Fig. 8 shows the spatial distribution of snow depth for the area obtained by ANNs and M5 decision tree algorithm. These distribution maps have a fine similarity to the distribution map produced from the recorded data (Fig. 7). However, the M5-decision tree produced a more accurate map.

As an additional visual analysis of the validation, the differences between actual and predicted SD values are shown in Fig. 9. From

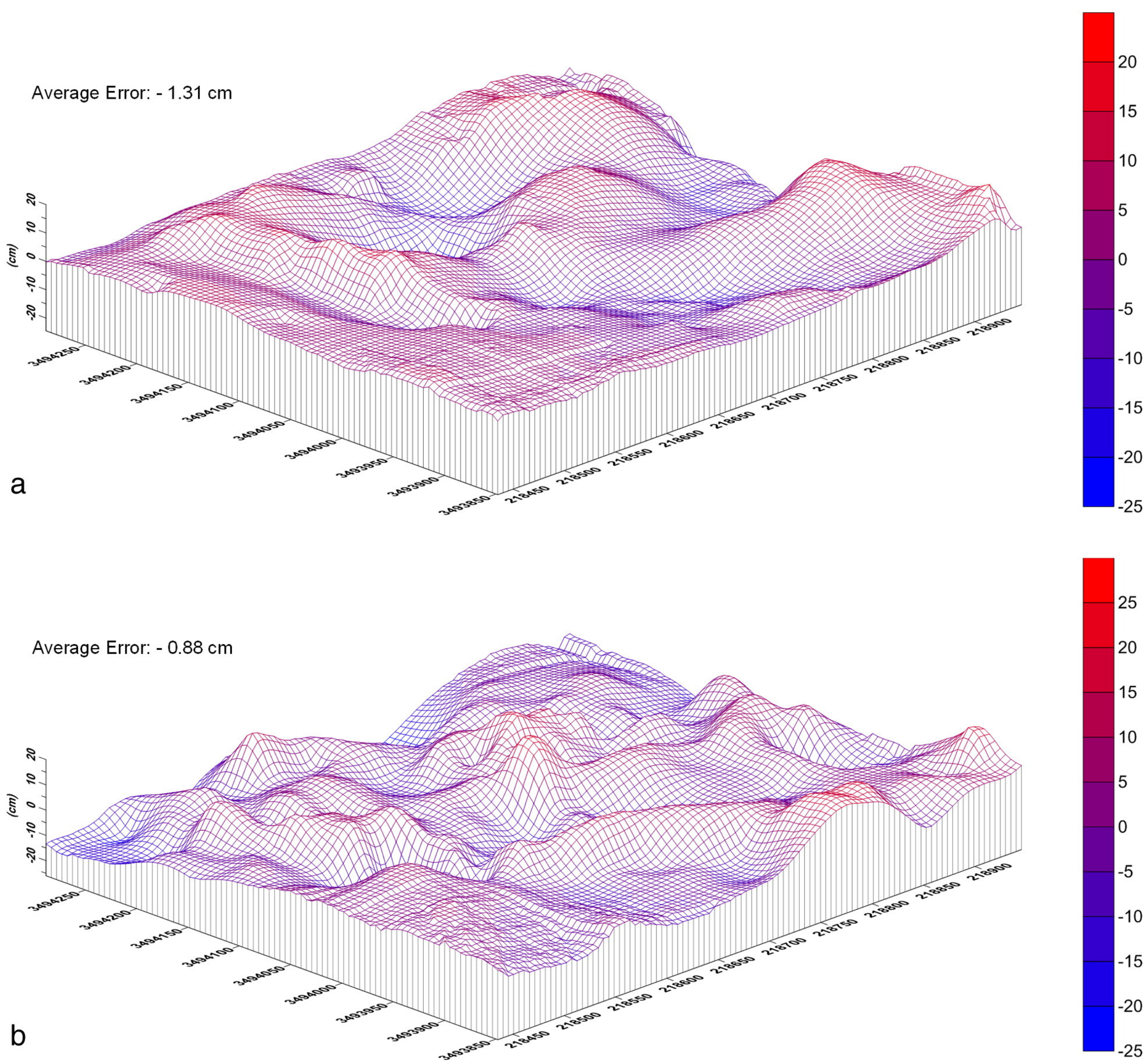

Fig. 9. Spatial distribution map of error resulted from the difference between observation data and a) ANN and b) M5 models.

these plots, it is possible to get a sense that the average error in the ANN and M5 models is −1.31 and −0.88 cm, respectively. In both models, the greatest error is in the northeast of the study area where the elevation was higher than the other parts.

4. Conclusions

In this study, the potential of using terrain-based parameters in evaluating the spatial distribution of snow has been inquired through the algorithms of artificial intelligence. Terrain parameters are the variables which can be derived from DEM simply. In spite of the multiplicity of these parameters, the most effective of them were: channel network base level, stream power, wetness index, height, analytical hill shading, altitude above channel network, plan curvature, wind effect, slope and curvature, respectively.

On the other hand, artificial intelligence methods have many applications in water resource sciences. The approach used employs two relatively new models of artificial intelligence. The spatial distribution of snow depth is a function of several conditions where artificial intelligence can desirably simulate these conditions and evaluate them. Although, this research was done in a limited area and time, it can draw a view on the amount of water in this region which has an important role in the water budget of Yazd. Generally, as there is no accessibility to many parts of this region, snow sapling is done in limited points and spatial distribution of snow carried out by some approximate methods such as Thiessen polygons. Because of some mentioned limitations in this research, it is recommended to carry out some similar investigations in a larger scale of temporal and spatial pattern, so as to improve the accuracy of the applied parameters and models.

General findings revealed that the M5 decision tree algorithm was more appropriate for snow depth estimation in the basin. This method was able to simulate 89% of changes in snow depth, whereas the ANN model performed at correlation coefficient values of 85%. In terms of E_{ns} value, the M5 decision tree with $E_{ns} = 0.80$ peformed better than the ANN model with $E_{ns} = 0.73$. The M5 decision tree algorithm not only had a higher performance, but also had more accurate rules and easier interpretation.

The results of the ANN application showed that the most effective parameters on snow depth are plan curvature, curvature, profile curvature, wind effect, slope, and analytical hill shading, respectively. However, for the M5 decision tree algorithm, the most effective parameters are channel network base level, stream power, wetness index, height, analytical hill shading, altitude above channel network, plan curvature, wind effect, slope and curvature, respectively. Plan curvature, curvature, wind effect, slope and analytical hill shading are common influential parameters on SD detected utilizing both the models.

The error map derived from the observations and simulation snow depth showed that both of the models had more errors in higher elevations. In spite of the dominant role of wind effect on redistribution of snow (Winstral et al., 2009), and because of the lack of weather station in the study area, it was not considered in this study. In addition, the study area was a rocky mountain and there were no vegetation types on it. Solar radiation was assumed constant over the basin due to very small area. Thus, the effect of these ground data was not considered in this research.

Acknowledgments

The authors would like to thank Eng. S. Ebdam (University of Yazd) for her assistance with the collection of our data. We would like to express our very great appreciation to Dr. Y. Falamarzi and Eng. N. Palizdan (University of Putra, Malaysia) for their valuable scientific assistance, too. Authors are also grateful to the editor and anonymous reviewers for their sincere valuable suggestions to improve the quality of the paper.

Appendix A

M5 algorithm Model 1:
Rule 1/1
If: *Channel Network Base level* > 0.63349
Then: *Predict*
= 3.2750122 − 3.29 *Channel Network Base level*
− 2.749 *Altitude Above Channel Network* + 0.484 *Wetness Index*
− 0.35 *Stream Power* − 0.36 *MRVBF* + 0.26 *High* + 0.151 *Wind Effect*
+ 0.17 *Flow Accumulation* − 0.12 *Plan Curvature*
− 0.072 *Analytical Hill shading* − 0.058 *Mid Slope Position*
Rule 1/2
If: *Channel Network Base level* <= 0.63349
Then: *Predict*
= −0.76906 + 1.05 *Wetness Index* + 0.653 *Altitude Above Channel Network*
+ 0.601 *Channel Network Base level* + 0.539 *Wind Effect*
− 0.3 *Stream Power* − 0.3 *MRVBF* + 0.22 *High* + 0.14 *Flow Accumulation*
− 0.11 *Plan Curvature* − 0.061 *Analytical Hill shading*
− 0.049 *Mid Slope Position*
Model 2:
Rule 2/1
If: *Channel Network Base level* > 0.7238551
Then: *Predict*
= −0.0968339 + 0.518 *Altitude Above Channel Network* + 0.68 *Curvature*
+ 0.484 *Channel Network Base level* − 0.46 *Profile Curvature*
− 0.51 *Plan Curvature* − 0.136 *Normalized Height* − 0.069 *slope*
+ 0.062 *LS Factor* − 0.04 *Stream Power* − 0.04 *Convergence Index*
Rule 2/2
If: *Channel Network Base level* <= 0.7238551
Then: *Predict*
= −4.6401841 + 15.97 *Curvature* + 8.718 *Altitude Above Channel Network*
+ 8.361 *Channel Network Base level* − 12.33 *Plan Curvature*
− 10.87 *Profile Curvature* − 0.613 *Slope* + 0.38 *LS Factor*
− 0.26 *Stream Power* − 0.22 *Convergence Index*
Rule 2/3
If: *Channel Network Base level* <= 0.7238551
LS Factor <= 0.1746937
Then: *Predict* = 2.6713503 − 15.221 *LS Factor*
Model 3:
Rule 3/1
If: *Channel Network Base level* > 0.63349
Then: *Predict*
= 2.6579609 + 2.57 *Curvature* − 1.485 *Channel Network Base level*
− 1.84 *Profile Curvature* − 2.01 *Plan Curvature* − 1.55 *High*
+ 0.124 *Wind Effect* − 0.116 *Slope* − 0.106 *Analytical Hill shading*
Rule 3/2
If: *Channel Network Base level* <= 0.63349
Then: *Predict*
= 0.3799909 + 3.43 *Curvature* − 2.4 *Profile Curvature*
− 2.62 *Plan Curvature* + 1.125 *Wetness Index* + 0.741 *Wind Effect*
+ 0.11 *High* − 0.061 *Slope* − 0.056 *Analytical Hill shading*
Model 4:
Rule 4/1
If: *Channel Network Base level* > 0.7238551

Then: *Predict* = −0.2077039 + 1.09 *High* − 0.183 *Normalized Height*
Rule 4/2
If: *Channel Network Base level* <= 0.7238551
Then: *Predict*
= −0.184755 + 0.651 *Wetness Index* + 0.7 *High* + 0.106 *LS Factor*
− 0.11 *Stream Power* − 0.08 *Slope*
Model 5:
Rule 5/1
If: *Channel Network Base level* > 0.63349
Then: *Predict*
= −0.3470167 + 2.53 *Altitude Above Channel Network* − 1.68 *High*
+ 1.08 *Channel Network Base level* − 0.291 *Analytical Hill shading*
+ 0.076 *Wetness Index* − 0.04 *Stream Power* − 0.05 *MRVBF*
+ 0.022 *Wind Effect*
Rule 5/2
If: *Channel Network Base level* <= 0.63349
Then: *Predict*
= −0.6330019 + 2.137 *Wetness Index* − 5.02 *Modified Catchment Area*
+ 1.474 *Wind Effect* + 0.665 *Channel Network Base level*
+ 0.444 *Altitude Above Channel Network* − 0.67 *Convergence Index*
− 0.18 *MRVBF* − 0.12 *Stream Power* − 0.059 *Analytical Hill shading*

References

Balk, B., Elder, K., 2000. Combining binary decision tree and geostatistical methods to estimate snow distribution in a mountain watershed. Water Resour. Res. 36, 13–26.
Bernd, T., Kleutges, M., Kroll, A., 1999. Nonlinear black box modelling–fuzzy networks versus neural networks. Neural Comput. & Applic. 8, 151–162.
Bhattacharya, B., Price, R.K., Solomatine, D.P., 2007. Machine learning approach to modeling sediment transport. J. Hydraul. Eng. 133, 440–450.
Bhattacharya, B., Solomatine, D.P., 2005. Neural networks and M5 model trees in modelling water level–discharge relationship. Neurocomputing 63, 381–396.
Blöschl, G., Kirnbauer, R., Gutknecht, D., 1991. Distributed snowmelt simulations in an alpine catchment: 1. Model evaluation on the basis of snow cover patterns. Water Resour. Res. 27 (12), 3171–3179.
Clark, L.A., Pregibon, D., 1992. Tree-based models. In: Chambers, J.M., Hastie, T.J. (Eds.), Statistical Models in S. Wadsworth & Brooks/Cole, Pacific Grove, CA.

Deka, P.C., 2014. Support vector machine applications in the field of hydrology: a review. Appl. Soft Comput. 19, 372–386.
Dressler, K.a., Leavesley, G.H., Bales, R.C., Fassnacht, S.R., 2006. Evaluation of gridded snow water equivalent and satellite snow cover products for mountain basins in a hydrologic model. Hydrol. Process. 20, 673–688.
Dyer, J.L., Mote, T.L., 2006. Spatial variability and trends in observed snow depth over North America. Geophys. Res. Lett. 33 (16).
Elder, K., Dozier, J., Michaelsen, J., 1991. Snow accumulation and distribution in an Alpine watershed. Water Resour. Res. 27 (7), 1541–1552.
Elder, K., Michaelsen, J., Dozier, J., Elder, K., Michaelsen, J., Dozier, J., 1995. Small basin modeling of snow water equivalence using binary regression tree methods. IAHS Publications – Series of Proceedings and Reports – Intern Assoc Hydrological Sciences 228, pp. 129–140.
Erickson, T.A., Williams, M.W., Winstral, A., 2005. Persistence of topographic controls on the spatial distribution of snow in rugged mountain terrain, Colorado United States. Water Resour. Res. 41.
Etemad-Shahidi, a., Mahjoobi, J., 2009. Comparison between M5′ model tree and neural networks for prediction of significant wave height in Lake Superior. Ocean Eng. 36, 1175–1181.
Gleick, P.H., 1993. Water and Conflict: Fresh Water Resources and International Security. International security, pp. 79–112.
Gong, G., Cohen, J., Entekhabi, D., Ge, Y., 2007. Hemispheric-scale climate response to Northern Eurasia land surface characteristics and snow anomalies. Glob. Planet. Chang. 56, 359–370.
Govindaraju, R.S., 2000. Artificial neural networks in hydrology. J. Hydrol. Eng. 124–137.
Grinand, C., Dominique, A., Bertrand, L., Manuel, P.M., 2008. Extrapolating regional soil landscapes from an existing soil map: sampling intensity, validation procedures, and integration of spatial context. Geoderma 143, 180–190.
Hall, M., 1997. Correlation-Based Feature Selection for Machine Learning. The University of Waikato (198 pp.).
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. ACM SIGKDD Explor. Newsl. 11 pp. 10–18.
Huo, Z., Feng, S., Kang, S., Huang, G., Wang, F., Guo, P., 2012. Integrated neural networks for monthly river flow estimation in arid inland basin of Northwest China. J. Hydrol. 420–421, 159–170.
Khadka, D., Babel, M.S., Shrestha, S., Tripathi, N.K., 2014. Climate change impact on glacier and snow melt and runoff in Tamakoshi basin in the Hindu Kush Himalayan (HKH) region. J. Hydrol. 511, 49–60.
Kind, R.J., 1981. Snow drifting. In: Gray, D., Male, D. (Eds.), Handbook of Snow: Principles, Processes, Management and Use. Elsevier, New York, pp. 338–359.
Kisi, O., 2005. Suspended sediment estimation using neuro-fuzzy and neural network approaches/Estimation des matières en suspension par des approches neurofloues et à base de réseau de neurones. Hydrol. Sci. J. 50 (null–696).
López-Moreno, J.I., Goyette, S., Beniston, M., 2009. Impact of climate change on snowpack in the Pyrenees: horizontal spatial variability and vertical gradients. J. Hydrol. 374 (3), 384–396.
Luce, C.H., Tarboton, D.G., Cooley, K.R., 1998. The influence of the spatial distribution of snow on basin-averaged snowmelt. Hydrol. Process. 12 (1011), 1671–1683.
Marofi, S., Tabari, H., Zare Abyaneh, H., 2011. Predicting spatial distribution of snow water equivalent using multivariate non-linear regression and computational intelligence methods. Water Resour. Manag. 25, 1417–1435.
Minasny, B., McBratney, A.B., 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. Chemom. Intell. Lab. Syst. 94 (1), 72–79.
Mollazade, K., Omid, M., Arefi, A., 2012. Comparing data mining classifiers for grading raisins based on visual features. Comput. Electron. Agric. 84, 124–131.
Nourani, V., Kisi, Ö., Komasi, M., 2011. Two hybrid artificial intelligence approaches for modeling rainfall–runoff process. J. Hydrol. 402, 41–59.
Nourani, V., Baghanam, A.H., Adamowski, J., Kisi, O., 2014. Applications of hybrid wavelet–artificial intelligence models in hydrology: a review. J. Hydrol. 514, 358–377.
Omid, M., Mahmoudi, A., Omid, M.H., 2010. Development of pistachio sorting system using principal component analysis (PCA) assisted artificial neural network (ANN) of impact acoustics. Expert Syst. Appl. 37, 7205–7212.
Quinlan, J.R., 1992. Learning with continuous classes 2. Constructing model trees. World Sci. 92, 343–348.
Quinlan, J.R., 2001. Cubist: An Informal Tutorial. http://www.rulequest.com.
Seyfried, M.S., Wilcox, B.P., 1995. Scale and the nature of spatial variability: field examples having implications for hydrologic modeling. Water Resour. Res. 31 (1), 173–184.
Shi, J., Dozier, J., 2000. Estimation of snow water equivalence using SIR-C/X-SAR. I. Inferring snow density and subsurface properties. IEEE Trans. Geosci. Remote Sens. 38 (6), 2465–2474.
Simona, F., Silvia, T., Fiorella, A., Mattia, F., Diego, G., Cristina, P.M., Secondo, B., 2015. How snow and its physical properties change in a changing climate Alpine context? Engineering Geology for Society and Territory. vol 1, pp. 57–60
Solomatine, D.P., Dulal, K.N., 2003. Model trees as an alternative to neural networks in rainfall–runoff modelling. Hydrol. Sci. J. 48, 399–411.
Solomatine, D.P., Xue, Y., 2004. M5 model trees and neural networks: application to flood forecasting in the upper reach of the Huai River in China. J. Hydrol. Eng. 9, 491–501.
Tabari, H., Marofi, S., Zare Abyaneh, H., Sharifi, M.R., 2010. Comparison of artificial neural networks and combined models in estimating spatial distribution of snow depth and snow water equivalent in Samsami basin of Iran. Neural Comput. & Applic. 19, 625–635.
Tarboton, D.G., Bölochl, G., Cooley, K., Kirnbauer, R., Luce, C., 2000. Spatial snow cover processes at Kühtai and Reynolds Creek. In: Grayson, R., Blo¨schl, G. (Eds.), Spatial Patterns in Catchment Hydrology: Observations and Modelling. Cambridge Univ. Press, New York, pp. 158–186.
Taghizadeh-Mehrjardi, R., Minasny, B., Sarmadian, F., Malone, B.P., 2014. Digital mapping of soil salinity in Ardakan region, central Iran. Geoderma 213, 15–28.
Tedesco, M., Pulliainen, J., Takala, M., Hallikainen, M., Pampaloni, P., 2004. Artificial neural networks-based techniques for the retrieval of SWE and snow depth from SSM/I data. Remote Sens. Environ. 90, 76–85.
Wang, Y., Witten, I.I., 1997. Inducing model trees for continuous classes. Proceedings of the Ninth European Conference on Machine Learning.
Winstral, A., Marks, D., Gurney, R., 2009. An efficient method for distributing wind speeds over heterogeneous terrain. Hydrol. Process. 23 (17), 2526–2535.
Regional Water Authority, Yazd, 2015. Available online www.yzrw.ir.