

Application of random forest time series, support vector regression and multivariate adaptive regression splines models in prediction of snowfall (a case study of Alvand in the middle Zagros, Iran)

Omid Hamidi¹ · Leili Tapak² · Hamed Abbasi³ · Zohreh Maryanaji⁴

Received: 7 June 2017 / Accepted: 11 October 2017 / Published online: 23 October 2017
© Springer-Verlag GmbH Austria 2017

Abstract We have conducted a case study to investigate the performance of support vector machine, multivariate adaptive regression splines, and random forest time series methods in snowfall modeling. These models were applied to a data set of monthly snowfall collected during six cold months at Hamadan Airport sample station located in the Zagros Mountain Range in Iran. We considered monthly data of snowfall from 1981 to 2008 during the period from October/November to April/May as the training set and the data from 2009 to 2015 as the testing set. The root mean square errors (RMSE), mean absolute errors (MAE), determination coefficient (R^2), coefficient of efficiency (E%), and intra-class correlation coefficient (ICC) statistics were used as evaluation criteria. Our results indicated that the random forest time series model outperformed the support vector machine and multivariate adaptive regression splines models in predicting monthly snowfall in terms of several criteria. The RMSE, MAE, R^2 , E,

and ICC for the testing set were 7.84, 5.52, 0.92, 0.89, and 0.93, respectively. The overall results indicated that the random forest time series model could be successfully used to estimate monthly snowfall values. Moreover, the support vector machine model showed substantial performance as well, suggesting it may also be applied to forecast snowfall in this area.

1 Introduction

Snowfall is one of the most important and sensitive components of a climate system that may be severely affected by climate change (Ke et al. 2009). This fascinating phenomenon, which is a highly reflective and emissive climate element, has low thermal conductivity and affects the global heat budget. This effect mainly works via triggering an increase in surface albedo along with outgoing longwave radiation that leads to a feedback to surface temperature which in turn causes climatic fluctuations (Robinson and Kukla 1985; Barnett et al. 1989; Ke et al. 2009). Furthermore, water resources are influenced by climate change through alteration of the snowfall distribution pattern and the intensity and the amount of precipitation and evaporation resulting from temperature and radiation changes as well as changes in vegetation response (Matondo and Msibi 2001). Several studies have indicated snowfall variation exerts an influence on climatic parameters such as temperature, precipitation, and circulation (Knowles et al. 2006; Walland and Simmonds 1997; Frei et al. 1999; Bednorz 2004).

Time series models have been widely used to predict the future behavior of climatic phenomena including snowfall which occur in cyclic patterns with spatial and temporal fluctuations. Long-term forecasting of snowfall is usually conducted through classical time series models like the

✉ Zohreh Maryanaji
z.maryanaji@gmail.com; maryanaji@sjau.ac.ir

Omid Hamidi
omid_hamidi@hut.ac.ir

Leili Tapak
l.tapak@umsha.ac.ir

¹ Department of Science, Hamadan University of Technology, Hamadan 65155, Iran

² Modeling of Noncommunicable Diseases Research Center, Department of Biostatistics and Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan 65175-4171, Iran

³ Department of Geography, Lorestan University, Khorramabad, Lorestan, Iran

⁴ Department of Climatology, Sayyed Jama'eddin Asadabadi University, Asadabad 6541835583, Iran

autoregressive integrated moving average (ARIMA) models. Despite several advantages including easy interpretation and automatic model selection, time series models suffer from some shortcomings as well. In particular, these models usually do not take the nonlinear characteristics of the data into account (Kisi and Parmar 2016) and attempt to remove high-frequency noises from the data in order to detect local trends based on linear dependence in observations (Kane et al. 2014). Moreover, ARIMA models assume that the standard deviation of the errors is constant over time. This issue can be addressed by utilizing another class of time series model known as Autoregressive Conditional Heteroskedasticity (GARCH). Nevertheless, this model also suffers from limitations. For example, optimization of the GARCH model for parameter estimation presents significant challenge (Kane et al. 2014). To address the issues related to traditional time series models, a new class of regression models has been developed, whose framework rests on machine learning methods. Examples of these models include support vector machine (SVM), random forest (RF), and multivariate adaptive regression splines (MARS) (Jalalkamali et al. 2015; Kane et al. 2014; Kisi and Cimen 2012; Kisi and Parmar 2016; Leathwick et al. 2006; Sedighi et al. 2016; Zapranis and Alexandridis 2011).

SVM uses structural risk minimization which alleviates the overfitting problem by attempting to find a global optimum instead of a local one (Adnan et al. 2017). Adequate performance of the SVM has been verified by several studies in the climatology field including precipitation (Kisi and Cimen 2012; Pour et al. 2016), Rainfall–Runoff (Sedighi et al. 2016), drought (Chi et al. 2013; Jalalkamali et al. 2015), and temperature (Zapranis and Alexandridis 2011). RF regression is an ensemble learning technique that creates several decision trees, with each tree recursively partitioning the input space until homogeneous small subspaces are created (Kane et al. 2014). A prediction rule is, then, created by calculating the average of the outcome dependent variable associated with the input variables in the subspace (Breiman 2001; Kane et al. 2014). While RF has been utilized effectively in prediction of time series data (Kane et al. 2014), few studies have utilized RF for the purpose of forecasting meteorological and climatological data (Pour et al. 2016). MARS is a data mining technique, which according to some studies offers adequate flexibility and precision along with the ability to rapidly forecast both continuous and binary output variables (Kisi and Parmar 2016). MARS models develop a functional relation using a set of coefficients and basic functions based on the regression data (Kisi and Parmar 2016). The main advantage of MARS model is that in these models, the relationships are considered additive and interactive, thereby leading to fewer variable interactions (Lee et al. 2006; Leathwick et al. 2006; Kisi and Parmar 2016).

Snowfall as a climatic element with high volatility has substantial positive and negative socioeconomic effects

particularly on agriculture as well as water resources. Seasonal forecasting of precipitation especially snowfall plays a key role in the planning and managing of water resources. Moreover, quantitative forecasting of precipitation in the form of snowfall in high-altitude regions during cold episodes of year from October/November to April/May (which is the prevailing period of middle-latitude westerlies in Iran, and the precipitation is often in the form of snowfall in altitudes) will be very helpful for government policymaking. In the present study, the snowfall data related to the six cold months of the year at Hamadan Airport sample station (located in the mountain ranges of the Zagros altitudes) were utilized for long-term forecasting of snowfall. The previous studies conducted in this area have confirmed the importance of detecting the likely impact of climate change on the water resources on the regional and local scale. To our knowledge, no study has investigated the performance of RF, SVM, and MARS in forecasting snowfall. Therefore, this study aimed compares the performance of RF, MARS, and SVM time series models for prediction of snowfall.

2 Material and methods

2.1 Study area and data description

The site of the present study, Hamadan, is situated in a mountainous area in the West of Iran as indicated on the map of Fig. 1. In this study, monthly snowfall data during six cold months from 1981 to 2015 collected from the Hamadan synoptic station also known as the airport station were used. The area is relatively mountainous with elevations ranging from 1730 to 3550 m (latitude: 35° 20' N; longitude: 48° 68' E). The mean precipitation including both rainfall and snowfall in this area is about 300 mm annually ranging from 280 in the central low lands of the region to 550 mm in the mountainous area (Maryanaji et al. 2017), and exhibits a strong temporal variability. Rainfall occurs at lower elevations (less than 2500 m) during autumn and winter. Snowfall occurs during winter, spring, and autumn above this elevation, but rain is the dominant precipitation of the region. Generally, during April and March, precipitation is often minimal throughout the area, but the rainfall events normally occur with high intensity. The mean annual air temperature of the area is 11.8 °C. The central and eastern parts of the area are characterized by low temperatures, while the southern parts are characterized by high temperatures.

In this study, the monthly snowfall data was extracted and registered. Before carrying out any calculation, we performed the Run test to check the accuracy as well as to examine homogeneity of the data. The homogeneity of the data was confirmed, and there was no gap. To prevent the problem of overfitting, the cross-validation method was applied. To this

Fig. 1 Geographical situation of the study area



end, the data was divided into two parts, namely the training part representing 80% of the data and the testing sets corresponding to the remaining 20% of the data. Specifically, monthly snowfall data from 1981 to 2008 corresponding to the period from October/November to April/May were considered as training set, and the remaining data were used as the testing set from 2009 to 2015. The monthly statistics of the data set including mean, standard deviation, and maximum and minimum of snowfall are given in Table 1. While for the data used, the distribution of the training and test sets was slightly different from each other; we considered them to be approximately similar.

2.2 The random forest model

In a RF regression model, which is an ensemble tree method, a large number of trees, for example, 1000 are created (Grömping 2009). This method utilizes randomness in two ways: (i) each of the trees in the RF is created using a random subset of the observations (boot strapped sampling), and (ii) each split in a tree is created using a random subset of

candidate variables (Breiman 2001; Grömping 2009). As these trees are relatively unstable, this randomness leads to establishment of differences in individual predictions obtained from each tree (Barnett et al. 1989). Injecting the randomness into the base learning process improves the performance of this ensemble learning method (Barnett et al. 1989). To obtain an overall prediction for the final forest, the mean of the predictions obtained from the individual trees is calculated. This can significantly improve the performance of the learning process (Barnett et al. 1989). Random forest takes into account

Table 1 The statistical parameters of monthly snowfall data set (cm/month)

Parameter	Entire data	Training set	Test set
	1981–2015	1981–2008	2009–2015
Mean	18.86	19.92	16.86
Minimum	0.00	0.00	0.00
Maximum	112.50	112.5	108.00
Standard deviation	23.10	23.38	23.47

nonlinear effects or higher order interactions of predictors as well as complex relationships between them automatically (Barnett et al. 1989).

2.3 The support vector machine model

SVM is one of the most widely used machine learning methods for classification and regression problems that works based on structural risk minimization (Hamidi et al. 2015; Kisi and Cimen 2012; Yoon et al. 2011). Due to the fact that the SVM minimizes the experimental error and the complexity simultaneously, its generalization ability for prediction purposes is improved greatly (Yoon et al. 2011). This method uses the basic idea of mapping the input vector of x space into a space with higher dimensions using an appropriate nonlinear kernel function, $\phi(x)$. Therefore, a simple linear regression may address the complex nonlinear regression of the input space (Kisi and Parmar 2016; Hamidi et al. 2015). To explain the SVM problem, let (x, y) be a set of variables, where $x \in \mathbb{R}^m$ stands for an input vector with m components serving as predictors, and y stands for an output variable representing the outcome. An SVM estimator (f) for the regression problem can be mathematically represented by the following equation:

$$y = f(x) = w \cdot \phi(x) + b \quad (1)$$

where w shows a weight vector representing the regression coefficient, and b shows the bias term in the equation (Hamidi et al. 2015). The solution to this equation is obtained using a convex optimization method with an ε -insensitivity loss function (Yoon et al. 2011; Hamidi et al. 2015). To obtain the weight vector and the bias term, the objective function can be converted to the following expression:

$$\frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i + C \sum_{i=1}^N \xi_i^* \quad (2)$$

which is minimized with respect to the restrictions given the following constraints:

$$\begin{cases} w^T \phi(x_i) + b - y_i \leq \varepsilon + \xi_i \\ y_i - w^T \phi(x_i) - b \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, N \end{cases} \quad (3)$$

In expression (2), C is a tradeoff parameter that takes positive values and determines the extent of the empirical error in the optimization problem (Hamidi et al. 2015). In addition, ξ_i and ξ_i^* are slack variables that penalize training errors by the loss function over the error tolerance ε . Projecting the input space into high dimensional feature space is performed using common kernel functions (Çimen and Kisi 2009) including polynomial, Gaussian radial basis (GRBF), and exponential radial basis. The present study exploited the GRBF kernel function $k(x_i, x) = \exp(-\gamma|x_i - x|^2)$ (Hamidi et al. 2015).

2.4 The multivariate adaptive regression splines model

Another nonlinear regression model that is utilized for predicting continuous numeric outcomes is MARS. This model is a nonparametric technique that avoids the questionable linearity assumption of classical time series and regression models (Zhang and Singer 2010). The main advantage of the MARS model is that it explains the complex nonlinear relationship of the inputs and the outcome variable (Kisi and Parmar 2016). The MARS model has the following form (Zhang and Singer 2010):

$$y = \beta_0 + \sum \beta_{ij} (x_i - \tau_j)^* + \sum_{i \neq k} \beta_{ijkl} (x_i - \tau_j)^* (x_k - \tau_l)^* + \dots$$

In the above formula, $(x_i - \tau_j)^*$ is a positive (or negative) truncated function. By adopting a value for the variable (which is used to define the inflection point in the range of predictors representing input variables or two basic functions, the function maps from the predictor space (variable x) into the output space (new variable y) using $y = \text{maximum}(0, x - c)$ and $y = \text{maximum}(0, c - x)$, where c represents the threshold value. The intersection between two adjacent splines at a knot is used to maintain the continuity of the basic functions (Kisi and Parmar 2016). There are several research areas that MARS model can be applied. One of these situations is when there is time series data. The MARS model also performs variable selection, and using the backward stepwise procedure eliminates the unnecessary variables thereby improving forecasting accuracy (Kisi and Parmar 2016).

2.5 Performance criteria

Several evaluation criteria were used to assess the performance of the methods. In particular, the root mean square error (RMSE), the coefficient of efficiency (E), the determination coefficient (R^2), and the mean absolute error (MAE) were used to evaluate the prediction accuracy of the three used methods of SVM, RF, and MARS. R^2 was utilized as a measure of the linear relation between the observed and estimated snowfall values. Higher values of R^2 indicate better prediction with $R^2 = 1$ representing a perfect prediction. The RMSE was used as a measure of the goodness of fit relevant to high snowfall values, and MAE was employed as a measure yielding a more balanced perspective of the goodness-of-fit at moderate snowfall values (Çimen and Kisi 2009; Hamidi et al. 2015). The smaller values of RMSE and MAE indicate better prediction with zero values for these two criteria indicating a perfect prediction. The coefficient of efficiency was also applied to measure the differences between the observed and estimated snowfall values relative to the variability in the observed snowfall values. The values of E that are greater than 90%

show very satisfactory performances (Çimen and Kisi 2009). The RMSE, E, and MAE are calculated as follows:

$$\begin{aligned} RMSE &= \sqrt{\frac{1}{n} \sum (Snowfall_{observed} - Snowfall_{predicted})^2} \\ MAE &= \frac{1}{n} \sum |Snowfall_{observed} - Snowfall_{predicted}| \\ E (\%) &= 1 - \frac{\sum (Snowfall_{observed} - Snowfall_{predicted})^2}{\sum (Snowfall_{observed} - Snowfall_{mean})^2} \end{aligned} \quad (4)$$

where n is the number of observations, and $Snowfall_{mean}$ is the average snowfall amount. We also applied intra-class correlation coefficient (ICC) to investigate the agreement between predicted and observed snowfall values.

3 Results

In the present study, monthly snowfall level in Hamadan, Iran was modeled using three data mining techniques. All analyses were conducted in R version 3.4.0 using random Forest (Liaw and Wiener 2002), e1071 (Dimitriadou et al. 2006), and Earth (Milborrow 2011) packages. The accuracy of the RF, SVM, and MARS models were calculated using the evaluation criteria described above. Specifically, the cross-validation technique was applied to investigate the performance of the models employed by dividing the data into training data and testing data subsets. The training and testing sets used for each model are given in Table 1. For the SVM, three parameters, namely C , γ , and ε were tuned. We determined the optimum values for C , γ , and ε using the trial and error method to be 1, 0.2, and 0.001, respectively.

For the three different methods employed, the RMSE, MAE, E, and R^2 statistics were calculated based on the training and testing data sets. The results are provided in Table 2. As is evident, the RMSE and MAE values for the RF model are smaller than those of the other two models in both the training and testing sets. In the RF model for snowfall prediction, these values were determined to be RMSE = 4.37 and

MAE = 2.47 based on the training data set and RMSE = 7.84 and MAE = 5.52 based on the testing data set. Moreover, the efficiency and the R^2 values in the RF model were greater than those of the other two models in both training and testing data sets ($E = 0.96$, $R^2 = 0.98$, and $ICC = 0.99$ for training set and $E = 0.89$, $R^2 = 0.99$, and $ICC = 0.93$ for the test set). This implies that the RF performance was better than the other two models for the given data. However, the SVM showed similar performance to that of the RF model with the values of the evaluation parameters being very close to those obtained based on the RF model.

The temporal variation of the observed monthly snowfall values and their estimated values obtained from the RF, SVM, and MARS models for the test period are plotted in Fig. 2a, b, c, respectively. It is clear from these graphs that the estimated snowfall values obtained from the SVM and the RF models are in good agreement with the observed values indicating that the models employed predicting snowfall fluctuations accurately. Moreover, the RF model predicted the best estimates for the observed values of snowfall followed by the SVM model. A residual plot is also illustrated for the three methods (Fig. 3). As is evident, the performance of the RF model was superior compared with the SVM and the MARS models.

In addition, the estimated values of snowfall obtained from the RF, the SVM, and the MARS models along with their corresponding observed values of snowfall are illustrated in the form of scatter plots in Fig. 4. As indicated by the fitted line equations of the form $y = a_0 x + a_1$ in the scatter plots of Fig. 4, compared with the other two models the a_0 and a_1 coefficients associated with the RF model are closer to 1 and 0, respectively.

Based on these results, the RF and SVM models showed promising performances in predicting the given snowfall fluctuations. The methodology based on the RF model was found to be better than those based on the SVM and the MARS models for modeling snowfall fluctuations based on the used data set.

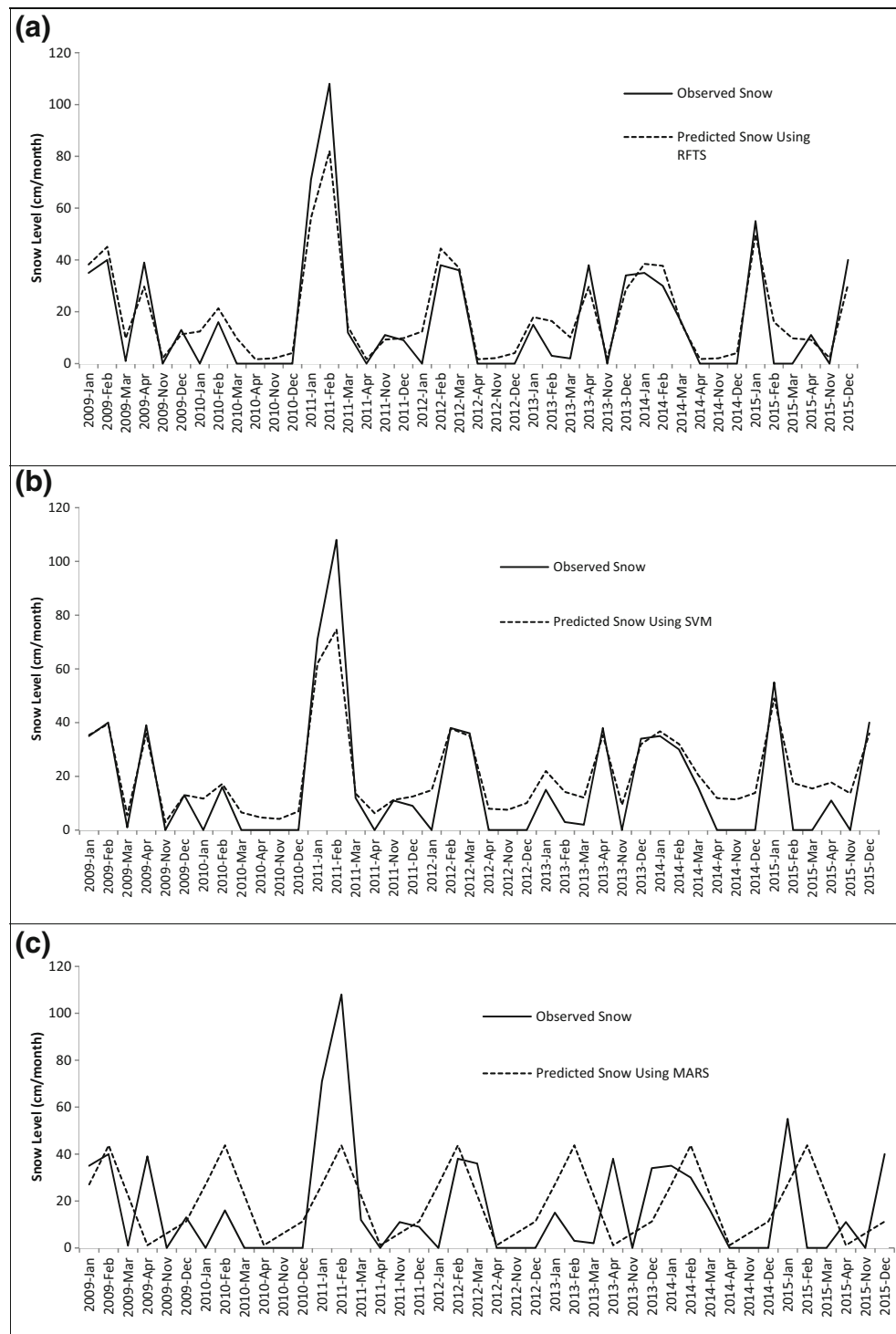
4 Discussion

Water resources management requires a comprehensive understanding of precipitation behavior. In mountainous regions and in the middle latitudes of Iran, winter precipitation regime is often in the form of snowfall. Therefore, forecasting the snowfall behavior as an important climatic element is beneficial for environmental planning and policymaking. This can be achieved through analysis of the hidden features of the snowfall. In Iran, snowfall starts from November to April as a consequence of the cold weather at high latitudes as a result of Mediterranean cyclones. Using statistical methods with minimal error and high performance plays an important role in providing prospects for understanding future climate

Table 2 The evaluation criteria of RMSE, MAE, E, R^2 , and ICC statistics of SVM, RF, and MARS for prediction of snowfall

Criterion		RMSE	MAE	E (%)	R^2	ICC
RF	Train	4.37	2.47	0.96	0.98	0.99
	Test	7.84	5.52	0.89	0.92	0.93
SVM	Train	2.12	1.49	0.99	0.99	0.99
	Test	9.18	6.76	0.84	0.95	0.91
MARS	Train	18.08	12.49	0.40	0.40	0.73
	Test	20.95	16.46	0.10	0.16	0.54

Fig. 2 Snowfall prediction values (cm/month) obtained using **a** random forest (RF) time series, **b** support vector machine (SVM), and **c** MARS models along with the observed values (cm/month)



changes in different regions. In this context, comparing the performance of different models gives an insight for identification of better models for forecasting purposes. There are several regression methods for analyzing snowfall fluctuations. Among them, those methods that are based on statistical learning theory have shown promising performances in different areas of study including time series data analysis. This

study compared the accuracy of the RF, SVM, and MARS models in modeling monthly snowfall data. A cross-validation method was utilized for evaluating the performance of the models.

The performance of the models revealed that the RF model exhibited the highest potential in forecasting snowfall in the given mountainous area followed by the SVM model. Several

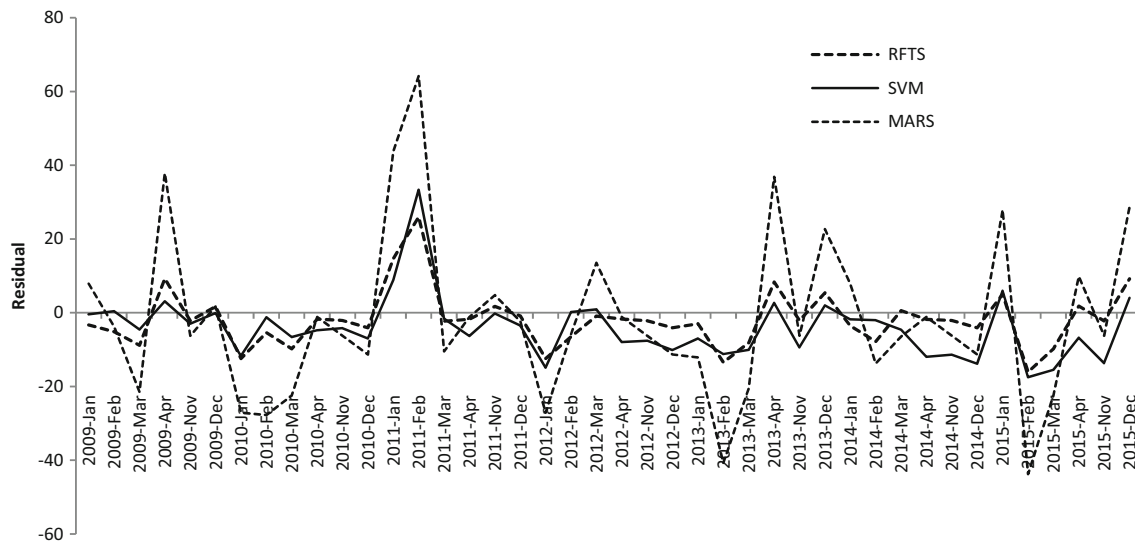


Fig. 3 Residuals for snowfall predictions (cm/month) obtained using random forest (RF) time series, support vector machine (SVM), and MARS models

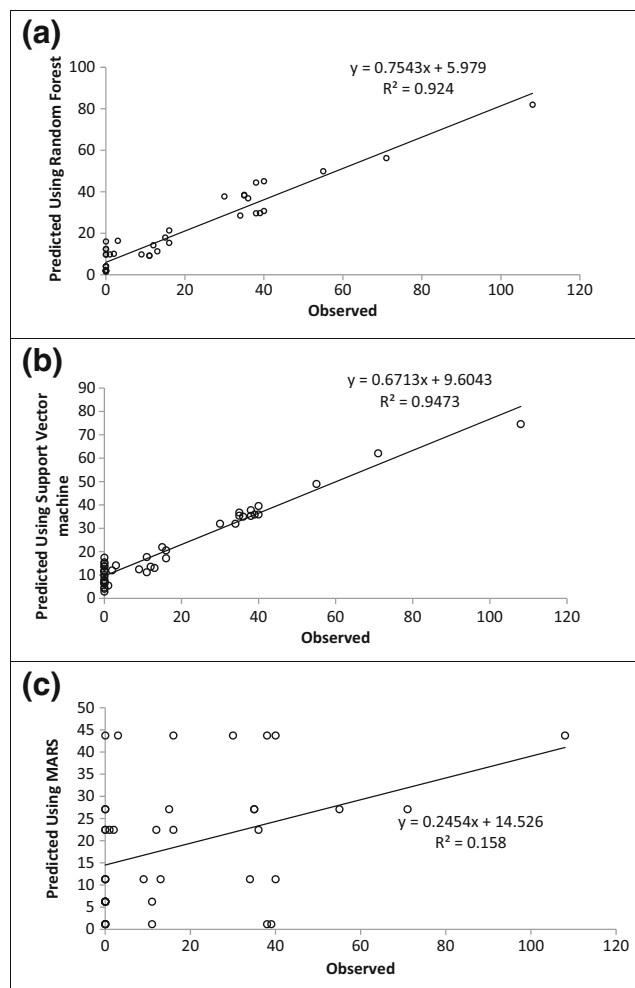


Fig. 4 Comparison of snowfall predictions (mm/month) using **a** random forest (RF) time series, **b** support vector machine (SVM with the observed values (mm/month) (minimum and maximum weather temperature was used as covariates)), and **c** multivariate adaptive regression splines (MARS)

criteria clearly demonstrated that the RF and SVM models are more capable than the MARS model in estimating snowfall values.

Other studies have confirmed that the performance of the SVM model is better as compared with relevant data mining techniques using artificial neural networks (Adnan et al. 2017; Hamidi et al. 2015; Yoon et al. 2011; Sedighi et al. 2016). In a study conducted by Kisi and Parmar, the performance of the SVM and the MARS model was similar in predicting monthly river water pollution (Kisi and Parmar 2016). Moreover, the RF performance was better compared with that of the ARIMA model in predicting avian influenza H5N1 outbreaks (Kane et al. 2014). Although we used significant lags as input variables to increase the performance of the models, their addition did not result in a significant increase in the accuracy of snowfall modeling.

Our results revealed that the RF model could be successfully used in estimating monthly snowfall. The results presented are related to long-term prediction of snowfall and are useful for management of the water resources. Consistency and agreement between observed and predicted data demonstrated the high capability of these techniques in modeling and estimating snowfall variations. In addition, these models are capable of displaying the periodic and non-periodic snowfall data over time. One of the most important advantages of applying data mining techniques compared to the classical time series models is that models such as RF and SVM do not rely on any distributional assumptions regarding the structure of the input and output variables. When one applies for example an ARIMA model, there is a need to evaluate the model assumptions such as linearity using residuals. However, it should be noted that the performance of the methods employed is data-dependent. Therefore, the performance of

these models should be investigated using other data sets. It is also worthwhile to assess the performance of other data mining techniques based on water resources data in the future.

5 Conclusion

In the present study, the performance of the RF, SVM, and MARS models was compared for prediction of monthly snowfall, and the potential of these techniques for modeling monthly snowfall was investigated. The results indicated that the RF model provided better results compared with the SVM and MARS models for prediction of monthly snowfall. Moreover, the performance of the SVM was similar, though slightly inferior, to that of the RF model. The performance of the MARS model was, however, deemed unsatisfactory based on the data used in the present study. The RF model uses randomization to improve its performance. The SVM works based on structural minimization, which is helpful in finding a global minimum, and leads to successful predictions. Therefore, the SVM and RF models are useful for prediction of monthly snowfall in the region considered in this study.

Acknowledgements We would like to express our appreciation to the Vice-chancellor of Education for technical support and the Vice-chancellor of Research and Technology of Hamadan University of Technology for their approval and support of this work.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Adnan RM, Yuan X, Kisi O, Yuan Y (2017) Streamflow forecasting using artificial neural network and support vector machine models. *Am Sci Res J Eng Technol Sci (ASRJETS)* 29:286–294
- Barnett T, Dümenil L, Schlese U, Roeckner E, Latif M (1989) The effect of Eurasian snow cover on regional and global climate variations. *J Atmos Sci* 46:661–686
- Bednorz E (2004) Snow cover in eastern Europe in relation to temperature, precipitation and circulation. *Int J Climatol* 24:591–601
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Chi D-C, Zhang L-F, Li X, Wang K, Wu X-M, Zhang T-N (2013) Drought prediction model based on genetic algorithm optimization support vector machine (SVM). *J Shenyang Agric Univ* 2:013
- Çimen M, Kisi O (2009) Comparison of two different data-driven techniques in modeling lake level fluctuations in Turkey. *J Hydrol* 378: 253–262
- Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., Weingessel, A. & Leisch, M. F. 2006. The e1071 package. *Misc Functions of Department of Statistics (e1071)*, TU Wien
- Frei A, Robinson DA, Hughes MG (1999) North American snow extent: 1900–1994. *Int J Climatol* 19:1517–1534
- Grömping U (2009) Variable importance assessment in regression: linear regression versus random forest. *Am Stat* 63:308–319
- Hamidi O, Poorolajal J, Sadeghifar M, Abbasi H, Maryanaji Z, Faridi HR, Tapak L (2015) A comparative study of support vector machines and artificial neural networks for predicting precipitation in Iran. *Theor Appl Climatol* 119:723–731
- Jalalkamali A, Moradi M, Moradi N (2015) Application of several artificial intelligence models and ARIMAX model for forecasting drought using the Standardized Precipitation Index. *Int J Environ Sci Technol* 12:1201–1210
- Kane MJ, Price N, Scotch M, Rabinowitz P (2014) Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. *BMC Bioinf* 15:276
- Ke C-Q, Yu T, Yu K, Tang G-D, King L (2009) Snowfall trends and variability in Qinghai, China. *Theor Appl Climatol* 98:251–258
- Kisi O, Cimen M (2012) Precipitation forecasting by using wavelet-support vector machine conjunction model. *Eng Appl Artif Intell* 25:783–792
- Kisi O, Pamarr KS (2016) Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. *J Hydrol* 534:104–112
- Knowles N, Dettinger MD, Cayan DR (2006) Trends in snowfall versus rainfall in the western United States. *J Clim* 19:4545–4559
- Leathwick J, Elith J, Hastie T (2006) Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecol Model* 199:188–196
- LEE T-S, CHIU C-C, CHOU Y-C, LU C-J (2006) Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Comput Stat Data Anal* 50:1113–1130
- Liaw A, Wiener M (2002) Classification and regression by randomForest. *R News* 2:18–22
- Maryanaji Z, Merrikhpour H, Abbasi H (2017) Predicting soil temperature by applying atmosphere general circulation data in west Iran. *J Water Clim Chang* 8:203–218
- Matondo JI, Msibi K (2001) Water resources development in the Usutu catchment: Swaziland under climate change. *Uniswa J Agric Sci Technol* 4:135–146
- Milborrow, S. 2011. Derived from mda: mars by T. Hastie and R. Tibshirani. *Earth: multivariate adaptive regression splines. R package*
- Pour SH, Shahid S, Chung E-S (2016) A hybrid model for statistical downscaling of daily rainfall. *Procedia Eng* 154:1424–1430
- Robinson DA, Kukla G (1985) Maximum surface albedo of seasonally snow-covered lands in the Northern Hemisphere. *J Clim Appl Meteorol* 24:402–411
- Sedighi F, Vafakhah M, Javadi MR (2016) Rainfall–runoff modeling using support vector machine in snow-affected watershed. *Arab J Sci Eng* 41:4065–4076
- Walland DJ, Simmonds I (1997) North American and Eurasian snow cover co-variability. *Tellus A* 49:503–512
- Yoon H, Jun S-C, Hyun Y, Bae G-O, Lee K-K (2011) A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J Hydrol* 396: 128–138
- Zapranis A, Alexandridis A (2011) Modeling and forecasting cumulative average temperature and heating degree day indices for weather derivative pricing. *Neural Comput Applic* 20:787–801
- Zhang, H. & Singer, B. 2010. *Recursive partitioning and applications*, Springer Science & Business Media