

# Climatic and atmospheric indices teleconnection impact on the characteristics of frost season in western Iran

Zohreh Maryanaji, Leili Tapak and Omid Hamidi

## ABSTRACT

The large-scale variability of atmospheric and ocean circulation patterns cause seasonal climate changes in the Earth. In other words, climate elements are affected by phenomena like El Niño Southern Oscillation (ENSO), El Niño (NINO), and Northern Atlantic Oscillation (NAO). In this study the characteristics of the frost season over a 20-year period (1996–2015) from seven synoptic stations in western Iran were evaluated using support vector machine and random forest regression. Comparing determination coefficients obtained by these models between atmospheric and ocean circulation indices and the characteristics of the frost season showed a positive effect. Thus, the onset and the end of the frost season in this region were highly correlated with the Southern Oscillation Index (SOI) and NAO, respectively. In regions with lower correlations (central areas and some regions of Alvand Mountain), the role of the geographical factors, altitude and topography becomes more pronounced and the impact of the global indices is reduced. Cluster analysis was also conducted to detect patterns and to identify regions according to the effect of the atmospheric and oceanic indices on frost season, and three regions were identified. The largest correlations with global indices (in both models) belonged to the first and third classes, respectively. The results of this study could be applied for planning environmental and agricultural activities.

**Key words** | frost season, NAO, Niño, SOI, teleconnection, western Iran

### Zohreh Maryanaji

Department of Geography,  
Sayyed Jamaledin Asadabadi University,  
Asadabad 6541835583,  
Iran

### Leili Tapak

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

### Omid Hamidi (corresponding author)

Department of Science,  
Hamadan University of Technology,  
Hamadan, 65155,  
Iran  
E-mail: [omid\\_hamidi@hut.ac.ir](mailto:omid_hamidi@hut.ac.ir)

## INTRODUCTION

Climate change leads to the advent of global indices as well as the establishment of fluctuations in these indices. These indices may affect climate elements of farther locations (teleconnection). Among the climate elements, the most important is temperature, the fluctuation of which induces weather phenomena including extreme heat and cold as well as frost. Frost refers to the conditions in which air temperature reaches zero or below zero degrees Celsius at an altitude of about one to two meters from the ground. The spatio-temporal instability of frost can cause irreparable damage in many situations. Today, most of the environmental studies are attempting to understand climate change and reduce its effects. One of the most important aspects of climate change is displacement of frost seasons.

Frost and temperature drop play important roles in mountainous and hilly areas. Especially favorable precipitation in these areas has led to agriculture playing a critical role in the economies of these regions. Study of frost behavior and the probability of its occurrence are among the most practical fields of research in the agricultural economy of areas that can guide planners in the reduction of damage caused by cold and frost phenomena. Almost every part of the economy may be adversely affected by unexpected frosts. Teleconnection study of frost provides better understanding of its behavior.

El Niño (NINO), El Niño Southern Oscillation (ENSO), and Northern Atlantic Oscillation (NAO) are very important climate indices. The general circulation of the

atmosphere consists of three main cells, where the Hadley cell is its main axis. The tropical and subtropical regions of the Hadley cell are zone-shaped. This situation leads to the emergence of different cells depending on the distribution of subtropical high pressure and the tropical low pressure centers and oceanic currents. The pressure changes of these centers in the Equatorial Pacific are determined by SOI, which shows an oscillating movement of air pressure on a global scale between the eastern and western equatorial Pacific (Walker 1922; Katz 2002). The equatorial Pacific Ocean is divided into the four regions of Niño 3.4, Niño 3, Niño 1 + 2, and Niño 4 to facilitate the study and measurement of surface water temperatures in these areas. Among these areas, the Niño 3.4 area is of particular importance, therefore the surface water temperature in the area is continuously measured to detect the El Niño (hot phase) and the La Niña (cold phase). The NAO is also one of the most prominent teleconnection patterns active in the northern hemisphere throughout the year (Wallace & Gutzler 1981; Barnston & Livezey 1987). The negative phase of the NAO represents a negative geopotential height and pressure anomalies in the northern latitudes of the North Atlantic, and a positive geopotential height and pressure anomalies in the North mid-Atlantic, east of the United States and western Europe. In the positive phase, the reverse situation occurs. Both positive and negative phases of the NAO are associated with severity and position changes of the northern Atlantic jet stream as well as the precipitation paths throughout the Atlantic Ocean, and on a macro scale it adjusts the average patterns of the orbital shifting and meridian transmissions of heat and humidity (Hurrell 1995). In certain periods, where a positive or negative phase lasts for a long time, a pattern of geopotential height and temperature anomalies is drawn up to the center of Russia and northern-central Siberia. The NAO indicates the presence of significant seasonal and annual fluctuations.

Following the warming-up of the Earth's climate, the general air circulation pattern has changed and alteration rates of the global indices such as the SOI and NAO have affected the climate of other regions of the world. Several studies have investigated the teleconnection effect of global indices on the climate elements (McCabe & Dettinger 1999; Müller & Ambrizzi 2006a, 2006b, 2007). However, no study has investigated their potential effects on the

characteristics of frost season. The usual method to investigate the relationship between NINO, ENSO, and NAO and climatic parameters is classical regression. However, despite easy interpretation, it considers a linear relationship between response variable and predictors which is a restrictive and unrealistic assumption in the real world. This class of models usually does not account for the nonlinear relationship between variables (Kisi & Parmar 2016). To address these issues, a new class of regression model has been proposed. The framework of these methods relies on machine-learning techniques. Among them, support vector machines (SVMs) and random forest (RF) have been widely used and their promising performance compared to the others has been confirmed (Zapranis & Alexandridis 2011; Kisi & Cimen 2012; Kane *et al.* 2014; Jalalkamali *et al.* 2015; Kisi & Parmar 2016). SVMs work based on the structural risk minimization to prevent overfitting problems (Adnan *et al.* 2017) and the RF regression combines the results of several decision trees (e.g., 1,000 trees) by averaging. Each of these trees recursively partitions the predictor's space to produce homogeneous small subspaces (Breiman 2001; Kane *et al.* 2014). While RF and SVM have been used effectively in the prediction of several meteorological and climatological data (Kane *et al.* 2014), no study has utilized these two methods for the purpose of investigating the teleconnection effect of global indices on climate elements.

This study aimed to evaluate the potential impact of NINO, NAO, and SOI on the characteristics of the frost season in western Iran using two data mining techniques. In order to achieve an accurate attitude towards the influence mechanisms of these global indices on the frost season, the onset and the end dates as well as the length of the frost season in several synoptic stations of western Iran over a 20-year period (1996–2015) were evaluated. This area is considered to be one of the most appropriate regions in Iran to have great agricultural and tourism potential. Due to the fact that most parts of Iran are located in arid and semi-arid regions, sufficient precipitation has attracted special attention to this area. Therefore, this area plays an important role in the development and environmental planning of different fields such as agriculture and tourism. This area is also located in the main path of the arrival of Mediterranean systems and westerlies to Iran and is also important from a weather

clemency perspective. The results of this research may be used to forecast the date of frost, which in turn can be applied in agricultural planning.

## MATERIAL AND METHODS

### Study area

The area in the present study is located between 33° and 35° northern latitudes and 47° to 49° eastern longitudes. The study area is situated in western Iran between the Zagros Mountains, and due to its high altitude is considered one of the coldest regions of Iran. Because of its specific geographical situation, suitable soil and quality of air and water, this area has great agricultural production capabilities. Therefore, climatic changes may cause irreversible damage to the economy of the region. In this study, the data from seven different synoptic stations of the region (Arak, Hamadan Airport, Hamadan-Nojeh, Kangavar, Kermanshah, Nanhavand, and West Islamabad) were utilized. Table 1 shows the characteristics of these stations and their locations are shown in Figure 1.

### Data source and processing

To assess the teleconnection effect of the global indices of NAO, Niño 3.4, and SOI with the characteristics (start, stop, and length) of the frost season in western Iran, the data were extracted from the seven aforementioned synoptic stations of the region (Arak, Hamadan Airport, Hamadan-Nojeh, Kangavar, Kermanshah, Nanhavand, and West

Islamabad) (1996–2015). In this study, first, primary and raw data related to the minimum daily temperature were extracted. Then, to determine the start and end date of the frost season, the Julian calendar was used. Based on the desired threshold, the temperatures zero and less than zero were considered as frost. In this regard, the following steps were taken:

1. Data accuracy and quality control assessment: before performing any calculation, a homogeneity test was performed to verify the accuracy and homogeneity of the data.
2. Statistical indices assessment: after validation of the data, some primary indicators such as mean, variance, coefficient of variation, and standard deviation were calculated.

## STATISTICAL METHODS

### Support vector machine

Support vector machine (Cortes & Vapnik 1995) is one of the most widely used machine learning techniques for regression and classification problems that implements the structural risk minimization principle. Using this property, the SVM improves its generalizability based on only a limited number of learning patterns. Both the VC (Vapnik–Chervonenkis) dimension and empirical risk are involved in structural risk minimization. In the SVM (Vapnik *et al.* 1997), the training data is  $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{N} \times \mathbb{R}$ . Here, in the present study,  $x$  stands for a vector of the

**Table 1** | Characteristics of meteorological stations of the study area

Station	Altitude (m)	Longitude	Latitude	Type of climate	The date of occurrence of the first ice	The date of occurrence of the last ice	The length of the frost season
Arak	1,702	49°48'	34°1'	Cold semi-arid	319	90	136
Hamadan Nojeh	1,679	48°43'	35°15'	Cold semi-arid	301	113	177
Kermanshah	1,318	47°15'	34°35'	Semi-arid. Cool steppe	319	91	137
Hamadan Airport	1,741	48°32'	34°52'	Cold semi-arid	330	101	163
Nahavand	1,658	48°24'	34°09'	Cold semi-humid	320	95	139
Islam Abad	1,348	46°28'	34°07'	Cold semi-humid	330	85	120
Kangavar	1,468	34°30'	47°59'	Cold semi-arid	300	103	168

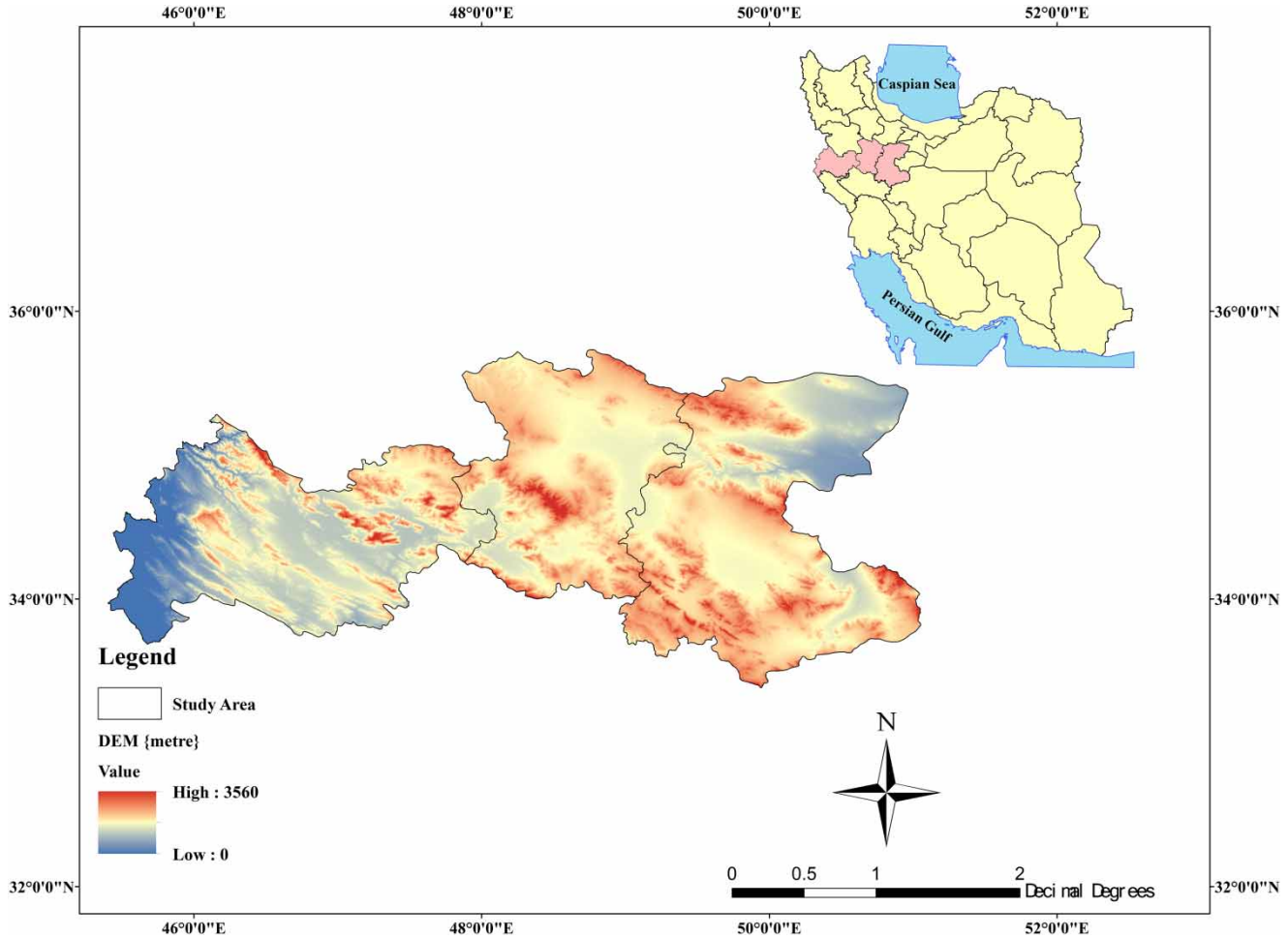


Figure 1 | Location of study area.

values of the global indices of NAO, Niño 3.4, and SOI and  $y$  stands for one of the characteristics of the frost season (start, stop, and length). In the regression version of the SVM, a function  $f(x)$  is estimated having at a maximum  $\epsilon$  deviation from the actual obtained output variable (start, stop, and length) ( $y_i$ ) for all the training data and being as flat as possible (Basak et al. 2007). Therefore, the objective function is  $y = f(x) = w \cdot \phi(x) + b$ . Formally, to find the estimated values of  $w$  and  $b$  (which shows the regression coefficients related to each predictor of NAO, Niño 3.4, and SOI and the intercept term (bias term)), the following convex optimization problem is written as:

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

given the following conditions:

$$\begin{cases} w^T \phi(x_i) + b - y_i \leq \epsilon + \xi_i \\ y_i - w^T \phi(x_i) - b \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, N \end{cases} \tag{2}$$

In the above equations,  $C > 0$  stands for a trade-off parameter (to determine the degree of the empirical error) and  $\xi_i$  and  $\xi_i^*$  stand for slack variables to penalize training errors (Tapak et al. 2013; Zhang et al. 2014).

The SVM utilizes some kernel functions to project the input space into a feature space with higher dimension (Çimen & Kisi 2009). There also exist different kernels used by the SVM, including Gaussian radial basis (GRBF), polynomial, exponential radial basis, and so on. (Friedman et al. 2001). In this study we used several kernels and

chose the GRBF kernel ( $k(x_i, x) = \exp(-\gamma|x_i - x|^2)$ ) as it returned the best results according to some performance criteria.

### Random forest

The fundamental idea of random forest is based on producing recursively regression trees, say 1,000, and improving trees' performance by introducing randomness in a subset selection of samples (boot strapped sampling) as well as variables ( $m$  out of  $p$  original variables) in each tree (Breiman 2001). Then the final prediction for these values is calculated using averaging over all trees. To achieve an overall prediction for the outcome variable (here, the quantitative outcome variables (outputs) were the characteristics of the frost season, i.e., start, stop, and length) from the forest, the average of several individual predictions computed from the individual trees is calculated (Liaw & Wiener 2002). The RF algorithm was used in a number of steps: (1) a predefined number of bootstrap samples (from the original data) was drawn; (2) an unpruned regression tree for each of the bootstrapped samples (at each node instead of choosing the best split from all inputs (the global indices of NAO, Niño 3.4, and SOI)) was grown up, a random sample of variables was selected and the best split from those variables was chosen; and (3) the outcome value (NAO, Niño 3.4, and SOI) for a new data set (called out of bag) by accumulating (averaging) the predictions related to all trees was predicted (Liaw & Wiener 2002). Here, we used 1,000 trees to create the forest. RF can deal with the issues of collinearity of inputs and high dimensionality of the data and takes into account their nonlinear effects or higher order interactions automatically.

### Clustering

To achieve spatial similarity of the areas in terms of determination coefficients and teleconnection characteristics, cluster analysis was conducted based on the average Euclidean distance and the Ward method (Johnson & Wichern 2014). In the Ward method, each observation belongs to a group that minimizes the total intra-group deviations. As a result, individuals in a cluster are located spatially on the map in the neighborhood of each other and the

geographic continuity of the regions is maintained. The Euclidean distance is calculated using the following formula:

$$e_{jk} = \sqrt{\sum_{i=1}^n (x_{ij} - x_{ik})^2} \quad (3)$$

i.e., to calculate  $e_{jk}$  for two subjects  $j$  and  $k$ , their related columns from the standardized original data matrix are used and therefore all of the variables related to these two subjects are taken into account. Then, the sum of the squared differences is calculated.

In the Ward method, the sum of square error is used to group the individuals, and error sum of square related to a group in each class is identified. Subjects who have the smallest error sum of squares in a pair of groups are put in a class. The error sum of squares in the Ward method is calculated using Equation (2).

$$E.S.S = \sum_{j=1}^k \left[ \sum_{i=1}^{n_j} X_{ij}^2 - \frac{1}{n_j} \left[ \sum_{i=1}^{n_j} X_{ij} \right]^2 \right] \quad (4)$$

In this equation,  $X_{ij}$  is the score of the  $i$ th subject in the  $j$  class and the total subjects in a class in each step is  $n_j$ . This error sum of squares is the index of the sum of square or variance. In the present study, interpolated data related to the determination coefficients obtained from the SVM and the RF for the regressions of the variables of start, stop, and the length of the frost season on NAO, Niño 3.4, and SOI indices in a 20-year period, as well as the standardized score of the variables, were used. Therefore, in Equations (3) and (4),  $X$  stands for the determination coefficients obtained from nonlinear regressions.

## RESULTS AND DISCUSSION

In the present study, to determine the general characteristics of the frost season, first the values of the start and stop as well as the length of the frost season at the mentioned stations were extracted from the minimum daily temperature data. The determination coefficients related to the



regressions of these variables on the global indices (Niño 3.4, NAO, and SOI) were extracted using the SVM and the RF. Then, the related  $R^2$  values were zoned by Surfer software using the Kriging method. Results are illustrated in Figures 2–6. Based on the maps, the following results were obtained.

### General characteristics of the frost season

Figure 2 shows the date of the first frost in autumn and the last one in spring based on the Julian calendar. According to Figure 2, the onset of the frost season in the north-central

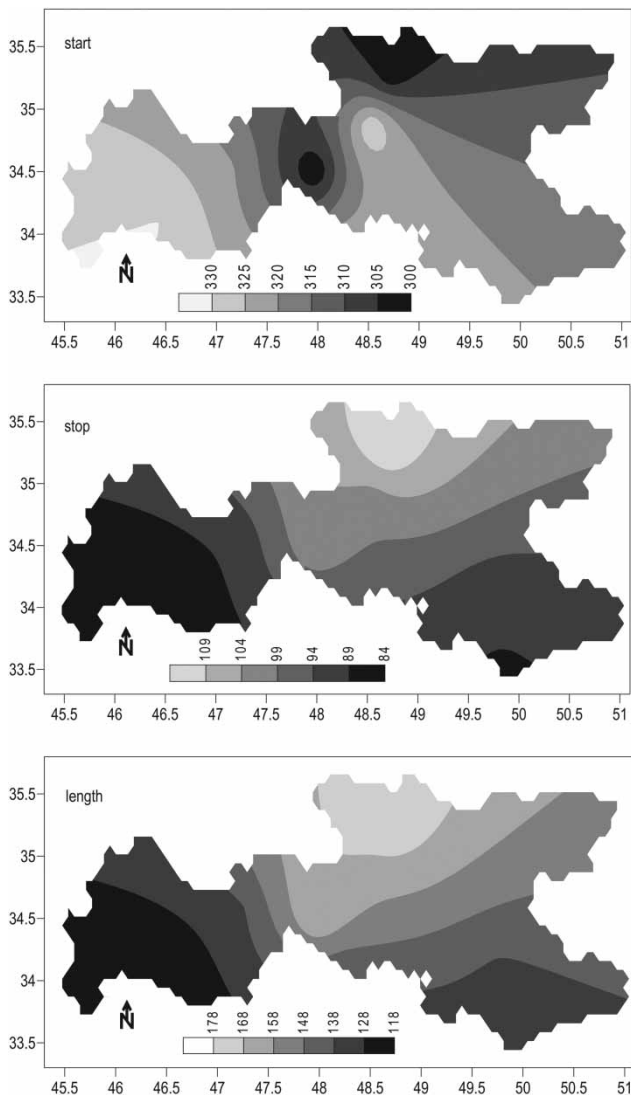


Figure 2 | Maps related to frost start, stop, and length.

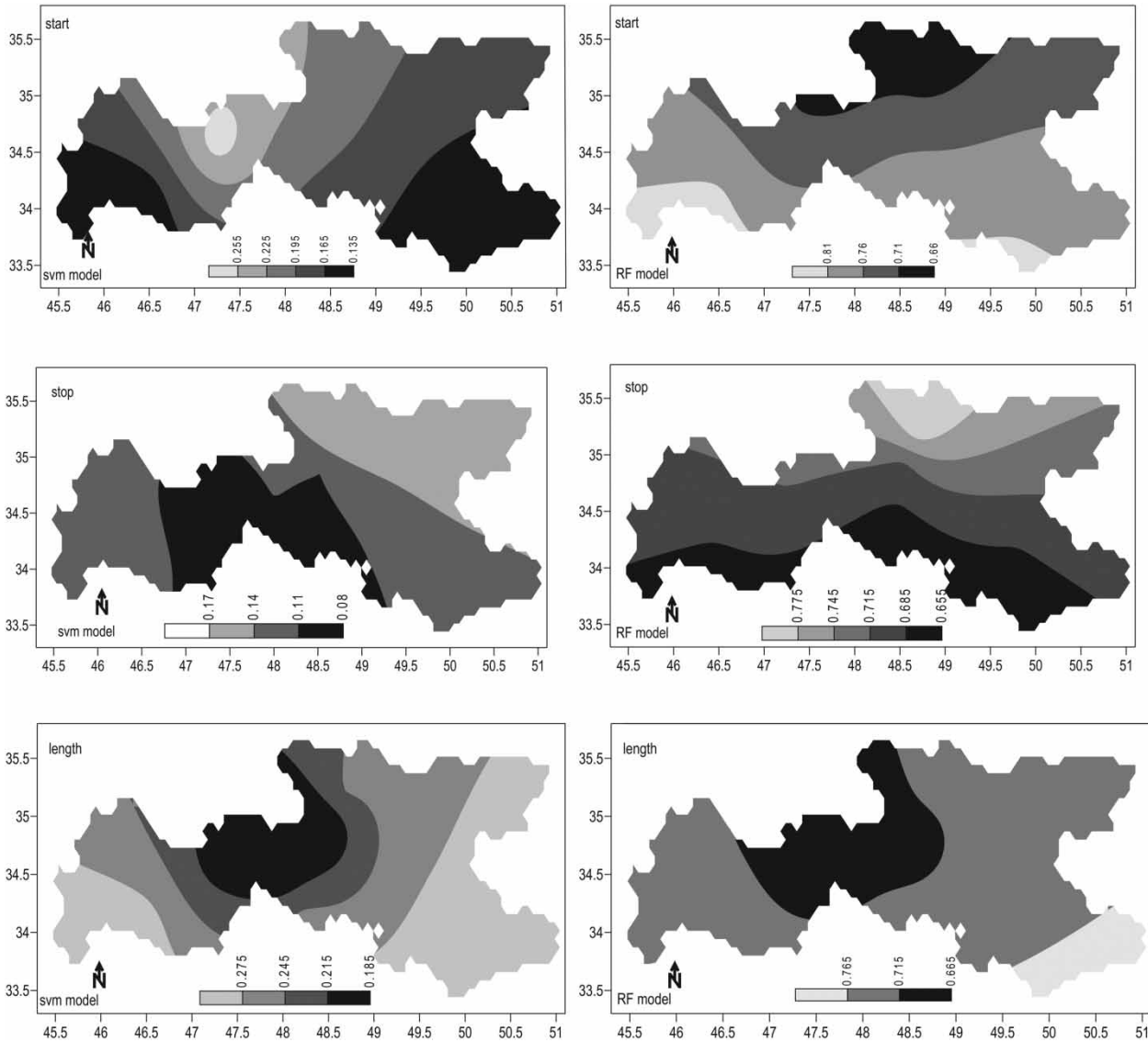
regions of the area under study starts earlier compared with other regions, while in the southern and western regions, the frost season occurs later compared with other areas.

In the map related to the end of frost season, provided based on the Julian calendar, the end of the frost season in the central and northern parts of the study area occurs later in comparison with other regions, indicating occurrence of a longer frost season in these areas. In the eastern and western parts, however, the end of the frost season occurs earlier than the other parts, implying that there is a shorter frost season in these areas. In the map of Figure 2, which identifies the length of the frost season, the central and northern regions have longer frost periods compared with the western and eastern regions.

### Analysis of maps corresponds to nonlinear models of global indicators

Figures 3–5 show the maps that were created based on the determination coefficients obtained from the SVM and the RF for the characteristics of the frost season (the start, stop, and the length of the frost season) on the global indices of SOI, Niño 3.4, and NAO, respectively. The values of the determination showed that the coefficients corresponding to the western, eastern, and central regions of the study region did not have a constant pattern (Figure 3), so that the smallest determination coefficient was related to the start of the frost season in the eastern and western regions. The largest determination coefficient was related to the length of the frost season in the same regions.

In the maps related to the determination coefficients between the characteristics of the frost season and the global indices of Niño 3.4 (Figure 4) and NAO (Figure 5), it is evident that in central regions of the studied area, which are parts of Alvand Mountain situated in Hamadan region, the influence (determination coefficient) of these global indices obtained from both models (the SVM and RF) was minimized. This is due to the topography and high altitude of the region as well as the influence of some local and geographical factors on the frost phenomenon. In other words, in this region, frosts are more affected by local factors.

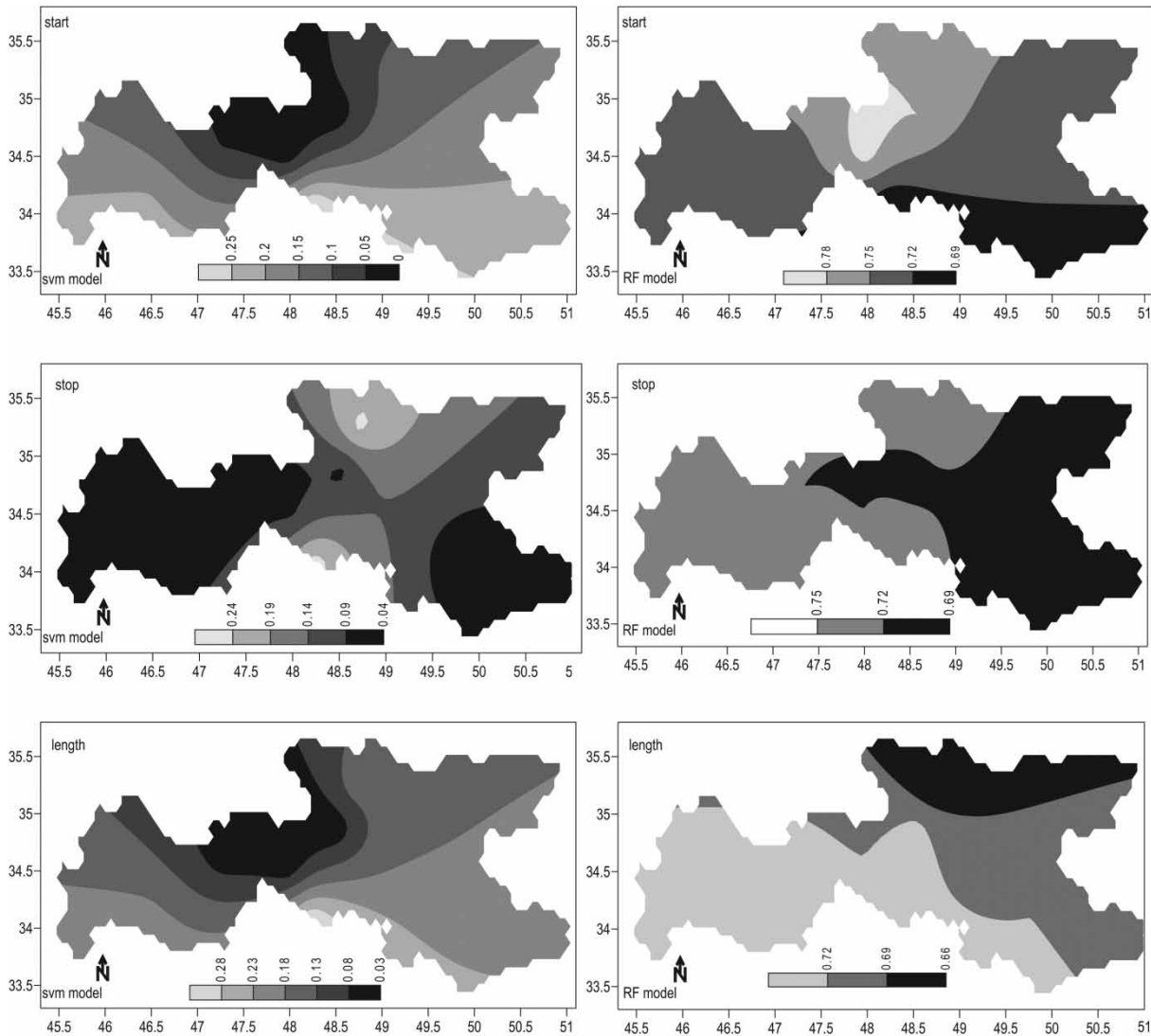


**Figure 3** | Maps related to two models, SVM and RF regression models, of frost start, stop, and length on yearly SOI.

### Classification of the impact of global indices on frost season

Clustering was conducted using the interpolated data (in pixels) related to the variables, such as the occurrence dates of the first frost of autumn and the last frost of spring, the length of the frost season, and the standardized score of the variables, as well as the determination coefficients obtained from the SVM and the RF regressions of the characteristics of the frost season on the Niño 3.4,

NAO, and SOI indices. For each cluster, a numerical measure was determined and then classified and zoned using the Kriging method (Figure 6). Based on the obtained results, three regions were identified in terms of the impact of global indices of NAO, Niño 3.4, and SOI on the frost season in the region. The characteristics of the identified regions are as follows. (1) Region 1 consisted of the eastern part of the region with a relatively long period of frost season, early onset and late ending of the frost season. In this region, the magnitudes of the correlations between



**Figure 4** | Maps related to two models, SVM and RF regression models, of frost start, stop, and length on yearly Niño 3.4.

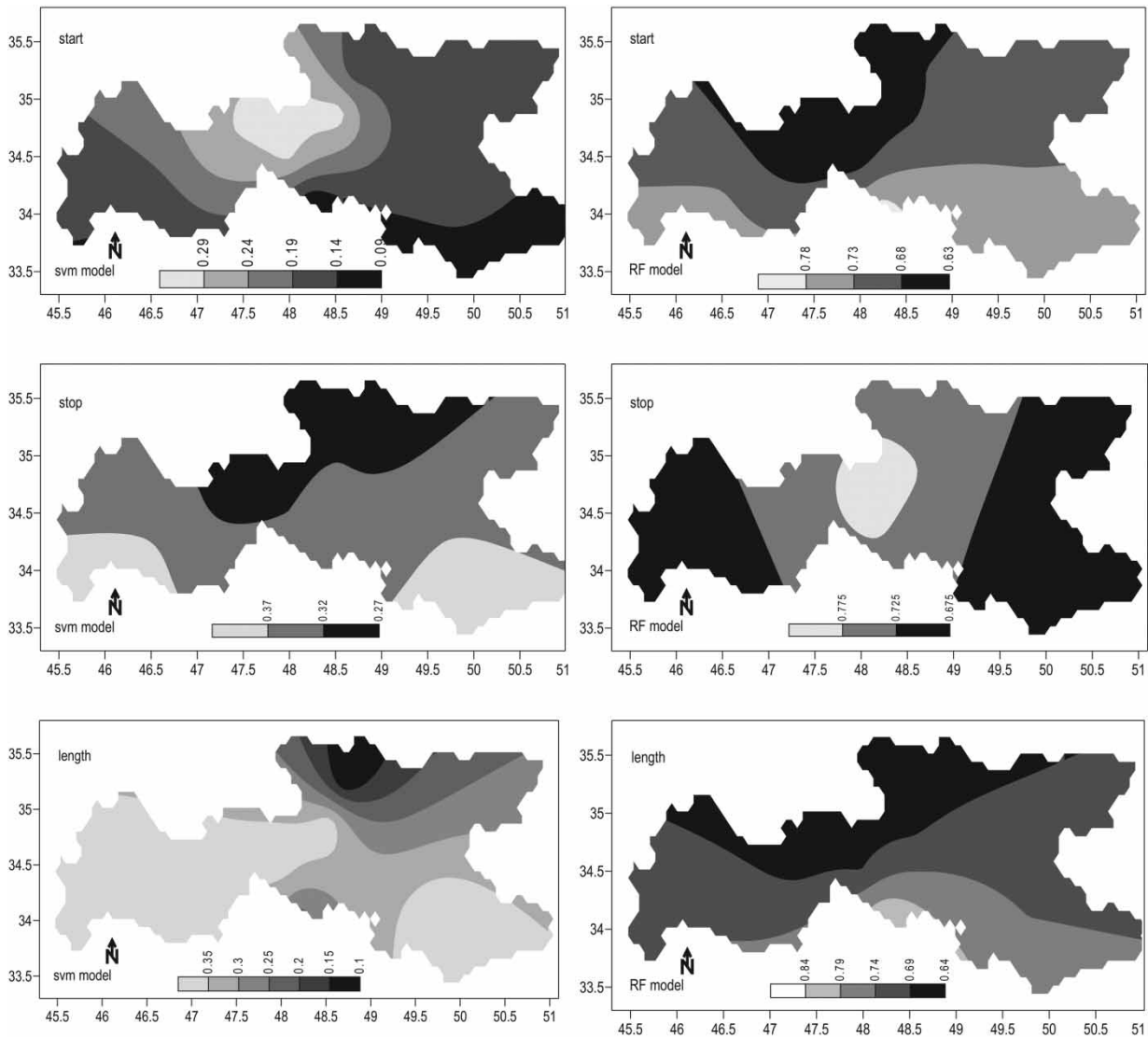
the characteristics of the frost season and the global indices using the SVM model were lower than those of the other regions. (2) Region 2 was located on the western border of region 1 with a lower altitude. The magnitude of the determination coefficients related to the Niño 3.4 and SOI indices (obtained using both the SVM and RF) were lower than those of the NAO index. (3) Region 3 had lower altitude compared with region 1 and region 2. On the other hand, the magnitude of the determination coefficients related to the global indices (obtained by both the SVM and RF) were greater than those of region 1 and region 2.

## CONCLUSION

Due to the importance of changes in the frost phenomenon and its widespread applications in the economic and developmental plans of the region as well as its impressibility from the global indices, in this study we compared and classified the changes and the characteristics of the frost season in western Iran over a 20-year period.

The results corresponding to the influence of the NAO, SOI, and Niño 3.4 indices on the characteristics of the frost season in western Iran showed that the onset (starting date) of the frost season had a significant and relatively strong



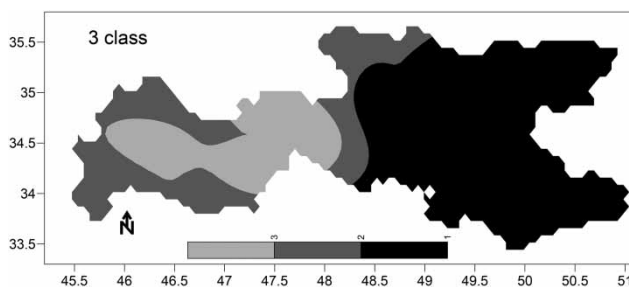


**Figure 5** | Maps related to two models, SVM and RF regression models, of frost start, stop, and length on yearly NAO.

relationship with the SOI index. The influence of this index on the other characteristics of the frost season (the end and

the length of the frost season) was also significant and the determination coefficients were relatively large. In this study, the two data mining regression techniques SVM and RF were utilized. The determination coefficients obtained by using the RF model were relatively large ( $>0.7$ ) which is due to the fact that this method considers the interaction between variables as well as their nonlinear effects (Breiman 2001).

The interesting point is that the start of the frost season was positively and highly associated with the SOI index among the three assessed indices of NAO, SOI, and Niño 3.4, while the end and the length of the frost season were positively and highly associated with the NAO index. It seems that the



**Figure 6** | Classification of frost season characteristics in the study area.

increase in the NINO index (the El-Niño phase), in turn, causes transmission of subtropical jet streams in the northern hemisphere to the lower latitudes. This transmission causes the north side of the Hadley cells to get closer to the equator. In this situation, high pressure and subtropical anti-cyclones are transferred to the lower latitudes leading to penetration of the waves of the westerlies and polar cold air masses as well as polar front into western Iran.

The magnitudes of the determination coefficients showed that the starting dates of the frost season increase (start later) as the values of the SOI and the NAO indices increase. This means that the frost season starts and finishes later in the El-Niño phase and NAO.

Comparing the determination coefficients of the RF and the SVM models implies that there is a direct (positive) simultaneous relationship between changes in large-scale atmospheric–oceanic circulation index and the characteristics of the frost season in western Iran. These changes, which have been monitored through sea surface water temperature indicators, cause changes in the frost season in this area. Moreover, comparing the determination coefficients of the SVM and the RF indicated that the determination coefficients obtained by the RF were greater than those of the SVM. Therefore, it is recommended that in remote sensing studies sufficient attention be paid to the nonlinear relationship between climatic indices as well. Given the fact that the economy of the study area relies on agriculture, recognizing the effects of the global indices on the characteristics of the frost season can be very beneficial in environmental planning.

## REFERENCES

- Adnan, R. M., Yuan, X., Kisi, O. & Yuan, Y. 2017 Streamflow forecasting using artificial neural network and support vector machine models. *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)* **29**, 286–294.
- Barnston, A. G. & Livezey, R. E. 1987 Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly Weather Review* **115**, 1083–1126.
- Basak, D., Pal, S. & Patranabis, D. C. 2007 Support vector regression. *Neural Information Processing-Letters and Reviews* **11**, 203–224.
- Breiman, L. 2001 Random forests. *Machine Learning* **45**, 5–32.
- Çimen, M. & Kisi, O. 2009 Comparison of two different data-driven techniques in modeling lake level fluctuations in Turkey. *Journal of Hydrology* **378**, 253–262.
- Cortes, C. & Vapnik, V. 1995 Support vector machine. *Machine Learning* **20**, 273–297.
- Friedman, J., Hastie, T. & Tibshirani, R. 2001 *The Elements of Statistical Learning, Springer Series in Statistics*. Springer, Berlin, Germany.
- Hurrell, J. W. 1995 Decadal trends in the north Atlantic oscillation: regional temperatures and precipitation. *Science-AAAS-Weekly Paper Edition* **269**, 676–678.
- Jalalkamali, A., Moradi, M. & Moradi, N. 2015 Application of several artificial intelligence models and ARIMAX model for forecasting drought using the Standardized Precipitation Index. *International Journal of Environmental Science and Technology* **12**, 1201–1210.
- Johnson, R. A. & Wichern, D. W. 2014 *Applied Multivariate Statistical Analysis*. Prentice-Hall, Upper Saddle River, NJ, USA.
- Kane, M. J., Price, N., Scotch, M. & Rabinowitz, P. 2014 Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. *BMC Bioinformatics* **15** (1), 276.
- Katz, R. W. 2002 Sir gilbert walker and a connection between el nino and statistics. *Statistical Science* **17**, 97–112.
- Kisi, O. & Çimen, M. 2012 Precipitation forecasting by using wavelet-support vector machine conjunction model. *Engineering Applications of Artificial Intelligence* **25**, 783–792.
- Kisi, O. & Parmar, K. S. 2016 Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. *Journal of Hydrology* **534**, 104–112.
- Liaw, A. & Wiener, M. 2002 Classification and regression by randomForest. *R News* **2**, 18–22.
- McCabe, G. J. & Dettinger, M. D. 1999 Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States. *International Journal of Climatology* **19**, 1399–1410.
- Müller, G. V. & Ambrizzi, T. 2006a Teleconnection patterns associated with extreme frequency of Generalized Frosts. Part I: Rossby waves propagation regions in the Austral Hemisphere. In: *8th International Conference on Southern Hemisphere Meteorology and Oceanography*, Foz do Iguazu, Brazil.
- Müller, G. V. & Ambrizzi, T. 2006b Teleconnection patterns associated with extreme frequency of Generalized Frosts. Part II: Origin and evolution of the Rossby waves propagation patterns in the Austral Hemisphere. In: *8th International Conference on Southern Hemisphere Meteorology and Oceanography*. Foz do Iguazu, Brazil.
- Müller, G. V. & Ambrizzi, T. 2007 Teleconnection patterns and rossby wave propagation associated to generalized frosts over Southern South America. *Climate Dynamics* **29**, 633–645.

- Tapak, L., Mahjub, H., Hamidi, O. & Poorolajal, J. 2013 [Real-data comparison of data mining methods in prediction of diabetes in Iran](#). *Healthcare Informatics Research* **19**, 177–185.
- Vapnik, V., Golowich, S. E. & Smola, A. J. 1997 Support vector method for function approximation, regression estimation and signal processing. In: *Advances in Neural Information Processing Systems* (M. Mozer, M. Jordan & T. Petsche, eds). MIT Press, Cambridge, MA, USA, pp. 281–287.
- Walker, S. G. T. 1922 *Correlation in Seasonal Variations of Weather, VII: The Local Distribution of Monsoon Rainfall*, Indian Meteorological Office. Mem. Ind. Met. Dep., 21: 23–29. (VIII) [Google Scholar]
- Wallace, J. M. & Gutzler, D. S. 1981 [Teleconnections in the geopotential height field during the northern hemisphere winter](#). *Monthly Weather Review* **109**, 784–812.
- Zapranis, A. & Alexandridis, A. 2011 [Modeling and forecasting cumulative average temperature and heating degree day indices for weather derivative pricing](#). *Neural Computing and Applications* **20**, 787–801.
- Zhang, X., Zhang, T., Young, A. A. & Li, X. 2014 [Applications and comparisons of four time series models in epidemiological surveillance data](#). *PLoS One* **9**, e88075.

First received 7 August 2017; accepted in revised form 28 November 2017. Available online 18 December 2017