Application of Artificial Intelligence in Agrometeorology

Homayoun Faghih
Urmia University, Urmia, Iran
Javad Behmanesh (j.behmanesh@urmia.ac.ir)
Urmia University
Hossein Rezaie
Urmia University, Urmia, Iran
Keivan Khalili
Urmia University, Urmia, Iran

Research Article

Keywords: ANN , GDD , GEP , Mann-Kendall (M-K) , Multivariate statistical analysis , Zadoks scale

Posted Date: July 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-565358/v1

Application of artificial intelligence in agrometeorology

Homayoun Faghih[1]. Javad Behmanesh[1*]. Hossein Rezaie[1]. Keivan Khalili[1]

Abstract

Replacing irrigated with rainfed crops and sustainable production of major rainfed plants (such as wheat) can be an efficient strategy to restore water resources that are drying up. Identifying plant response to climate is essential to advancing this strategy and planning for precision agriculture. Wheat is the main plant of Saqez in the Lake Urmia basin of Iran, whose yield is associated with severe fluctuations. This study was conducted to investigate the climate effect on wheat yield fluctuation. For this purpose, the method of growing degree days (GDDs) and the Zadoks scale were used to divide the wheat growth period into seven stages. Forty-seven climatic variables of the first six stages were used to do factor analysis and to develop the model for forecasting pre-harvest yield. Gene expression programming (GEP), artificial neural networks (ANNs), and multivariate linear regression (MLR) methods were applied to develop the model. The results showed that 90.7% of the total variance of 47 variables can be explained by 10 factors. Eighty-two percent of yield variations were modeled by these 10 factors (r =0.91). The mean absolute percentage error (MAPE) for the models developed by the GEP and ANN methods was 26%, and its amount for the MLR model was 35%. In this study, for the first time, the GEP method was used to model rainfed wheat yield. Comparison with MLR and ANN methods shows that GEP is suitable for modeling in this field.

[*] Corresponding Author: Javad Behmanesh (j.behmanesh@urmia.ac.ir)

Homayoun Faghih (hfkuir@gmail.com), Hossein Rezaie (h.rezaie@urmia.ac.ir) and Keivan Khalili (khalili2006@gmail.com)

[1] Department of Water Engineering, Urmia University, Urmia, Iran.

Keywords ANN . GDD . GEP . Mann-Kendall (M-K) . Multivariate statistical analysis . Zadoks scale

1 Introduction

Urmia Lake of Iran is the largest saltwater lake in the Middle East (Zoljoodi and Didevarasl 2014). Over 20 years (1995-2015), the lake water level has decreased by more than eight meters (Khazaei et al. 2019). Continuation of the drying process of Urmia Lake will destroy the basin ecosystem and will cause a lot of damage to its surroundings.

Climate change and low water use efficiency in the agricultural sector are the main reasons for reducing the volume of water entering Urmia Lake and its drying. In such circumstances, reducing water use in agriculture via rainfed farming can be one of the most important strategies to restore the lake. Success in this strategy economically requires that rainfed plant production be increased and sustainable. The climate has a random nature among the main factors affecting the yield of rainfed crops (including climate, field operations, plant, and soil characteristics) (Faghih et al. 2020). Therefore, rainfed crops' development and sustainable production need to correctly identify the plant's response to climatic factors and their changes. In this regard, the simulation of plant behavior in the rainfed cultivation system has always been considered and will be more important in the future in areas where water resources are facing tension and crisis.

The importance of accurately identifying and simulating the relationship between climatic variables and yield of (especially rainfed) plants has made agricultural and meteorological researchers interested in using other sciences to do this important work.

Bal et al. (2004) modeled and expressed 69% of the variations in rainfed wheat yield in the Punjab state of India using the variables of daily minimum air temperature and GDD (as

independent variables) and multivariate regression model. Lee et al. (2013) reported that wheat yield in Oklahoma, USA, is strongly influenced by climatic variables and that the yield is directly related to precipitation and inversely related to air temperature. Ceglar et al. (2016) stated that the occurrence time of the sensitive stages of wheat to climate is very different in diverse regions of France. Kheiri et al. (2017) identified the role of climate variables in spring and April and May in changing wheat yield in northwestern Iran as more important than that in other seasons and months of the year, respectively. Faghih et al. (2020) used multivariate statistical methods and climatic variables related to the wheat growth period to model rainfed wheat yield in Qorveh, Iran. The results showed that 85% of the yield changes in this region can be justified by climatic factors. They considered the climatic factor in the growth stage of jointing to flowering, which coincided with May, to be more effective than other factors on the rainfed wheat yield in Qorveh. They identified that the increasing trend of the rainfed wheat yield in Qorveh was due to a significant trend in climatic variables. Lamba and Dhaka (2014) reviewed the major models for wheat crop yield prediction in all areas. These forecasting models included models based on statistical, meteorological, simulation, agronomic, remote satellite sensed, synthetic, and mathematical methods. They concluded that ANN was more efficient than other methods because of its ease, low cost, and good accuracy. Chen and Jing (2017) used nonlinear (ANNs) and linear (partial least squares regression (PLSR)) methods to model and predict wheat yield. Comparison of methods showed the superiority of the nonlinear method. Dornelles et al. (2018) found the use of ANNs to be more effective than multiple regression to simulate the oat grain yield in Brazil. Khaki and Wang (2019) were able to accurately predict the yield in 2017 using the ANN method based on weather data and yield data of 2267 maize hybrids planted in 2,247 locations between 2008 and 2016 years. They identified this method as superior to the regression tree method for modeling maize yields. Adisa et al. (2019) stated that the ANN method for

modeling and predicting Maize production based on climate variables in South Africa is accurate.

Saqez is one of the main areas of rainfed wheat cultivation in the Urmia Lake basin and Iran. In a period of 51 years (1968-2018), the average yield of rainfed wheat in this region has fluctuated in the range of [175-1550] kg/ha (Faghih 2018). This large range of yield changes indicates the possibility of production with maximum yield in the region. Stability or increase in maximum yield requires proper knowledge of the factors affecting it. This study was conducted to investigate the effect of climate on the rainfed wheat yield in Saqez and to develop the models for estimating it based on climatic factors using three methods. Evaluation and comparison of the modeling performance of MLR and ANN methods, as common methods, and the GEP method, which has recently been considered, was another goal of this research. The results of this study will identify the limitations and capabilities that the climate creates in the sustainable production of wheat in Saqqez and will be useful for the sustainable use of resources and making appropriate decisions for precision agriculture.

2 Materials and methods

Multivariate statistical analysis was used to determine the main climatic factors affecting the yield of rainfed wheat in the Saqez region. To develop the models for predicting the yield of rainfed wheat based on the identified main factors, MLR, ANN, and GEP methods were used. Before performing factor analysis and developing the model structure, the data series were divided into three groups. Sixty-five percent of the data were used as training data for factor analysis and model structure development. Fifteen percent of data, as validation data, was used to evaluate the training phase results and make decisions about stopping model training. The accuracy of the model, or the model test, was determined using the remaining 20% of the data. To select testing data, a sample of every five years or an element of five elements from

the data series was taken. Again, one out of every five elements of the remaining data was sampled to select the validation data. SPSS 16, XLSTAT Premium v2016.02.28451, NeuroSolutions 5.05, and GeneXproTools 5.0 software were used to perform the calculations.

2.1 Study area

The study area is Saqqez County in the southern part of the Urmia Lake basin of Iran (Fig. 1). According to the Emberger method (Arkian et al. 2018), Saqqez has a cold semi-humid climate. The minimum and maximum air temperatures for Saqqez were recorded as -36°C and 43°C, respectively. The average precipitation in this area is approximately 450 mm/y. The average amount of arable land in Saqqez is 118,000 hectares, in 70% of which rainfed wheat is planted (Faghih 2018). Therefore, wheat can be considered the main crop in this region.

2.2 Crop growth stages

The plant life cycle consists of interconnected stages. The plant needs a certain amount of heat to pass through each of its growth stages, which is expressed in terms of growing degree days (GDDs). The GDDs related to some wheat growth stages are presented in Table 1 (Bowden et al. 2008). Equation (1) was used to calculate GDDs.

$$GDD=\sum_{j=1}^{n}\left(\left(\frac{1}{m}\cdot\sum_{i=1}^{m}T_i\right)-T_b\right) \quad (1)$$

Where T_i is the hourly air temperature (°C), m is the number of hourly air temperature records, n is the number of growth-stage days, and T_b is the base temperature or physiological zero. At temperatures below T_b, plant growth stops. The value of the base temperature is different for each plant. T_b was considered 0 °C for wheat (Bowden et al. 2008).

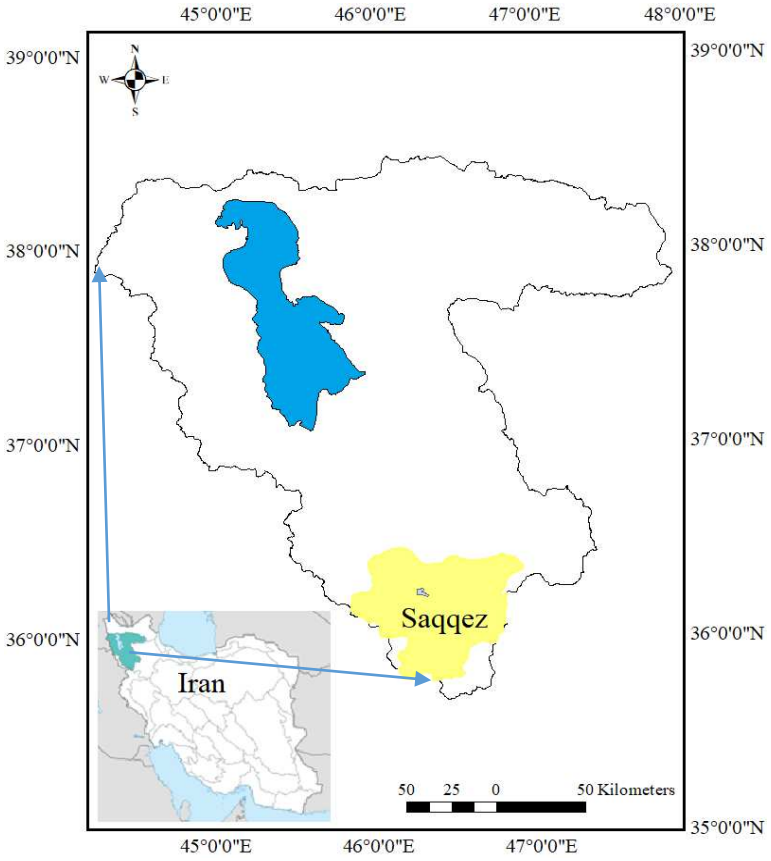

Fig. 1 Location of Urmia Lake basin and Saqqez County in Iran

Table 1 Generic GDDs related to some stages of wheat growth (Bowden et al. 2008)

Row	Growth Stage (code*) (abbreviation)	GDDs (°C)	Row	Growth Stage (code*) (abbreviation)	GDDs (°C)
1	Sowing (00) to emergence (SE)	100	5	Jointing (36) to anthesis or flowering (JA)	515
2	Emergence (09) to three leaves unfolded or trifoliate (ET)	200	6	Anthesis (69) to maturity (89) (AM)	800
3	Trifoliate (13) to end of tillering or double ridge (TT)	305	7	Maturity to harvest product (99) (MH)	300
4	Double Ridge (29) to jointing (TJ)	255	8	Total	2475

* The codes are derived from the famous cereal code developed by Zadoks et al. (1974).

2.3 Data

Climatic data of the study area for a period of 56 years (1961-2017) were obtained from Saqez station. This station (with a latitude of 36° 15′ N, a longitude of 46° 16′ E, and an altitude of 1523 m above sea level) belongs to the Kurdistan Meteorological Organization. The climatic data used and their statistical summary are presented in Table 2. Wheat yield data for 46 years in the region was recorded by the Kurdistan Agriculture-Jahad Organization, which was received and used.

Because the purpose of this study was to develop a model to estimate pre-harvest yield, so the climatic data of the last stage and those that were almost constant during the statistical period were not used in the analysis and development of models. Data whose values were almost constant are marked with empty cells in Table 2. The number of variables and indicators used was equal to 47.

2.4 Planting date

In rainfed cultivation, the water requirement of the crop is supplied with precipitation. Effective precipitation and the ideal temperature are necessary for seed germination after sowing. Faghih et al. (2021) based on climate and soil conditions introduced a new definition of effective precipitation and onset of germination dates that have been used in this study. In this method, the depth (threshold) of effective precipitation (D_e) is calculated from Eq. (2). The occurrence time of precipitation equal to De (in one or some consecutive days after sowing) coincides with the appropriate temperature conditions for wheat germination, was considered as the date of germination (Faghih et al. 2021). Germination can occur between 4 and 37 °C, but the ideal temperature for its occurrence is between 12 and 25 °C (Bowden et al. 2008), which was used in this study.

Table 2 Summary statistics of climatic variables and indices at different stages of wheat growth for the 56-year study period (1961-2017)

		Variables and indices (abbreviation) (Unit)								
Growth stage	Descriptive statistics	Mean relative humidity (RH) (%)	Precipitation (P) (mm/day)	Mean air temperature (T) (°C)	Number of growth-stage days (L) (day)	Days with a mean temperature equal to or less than 0 °C (T_0) (day)	Days with a mean temperature equal to or greater than 25 °C (T_{25}) (day)	Days with P equal to or greater than 1mm (P_1) (day)	Days with P equal to or greater than 5mm (P_5) (day)	Days with P equal to or greater than 10mm (P_{10}) (day)
	Mean	64.9	3.15	8.35	20.0	6.00		8.00	3.00	2.00
1-SE	St. dev.	10.6	2.27	4.70	25.0	18.0		10.0	3.00	2.00
	Skew.	-0.34	1.22	-1.10	2.78	3.15		2.66	2.46	1.51
	Mean	66.1	2.07	4.64	55.0	18.0		19.0	8.00	5.00
2-ET	St. dev.	9.06	1.36	4.26	37.0	25.0		16.0	7.00	6.00
	Skew.	-0.71	1.89	0.19	0.94	1.39		1.05	2.01	3.53
	Mean	64.5	2.12	4.52	77.0	23.0		26.0	10.0	5.00
3-TT	St. dev.	8.30	1.10	4.61	41.0	26.0		17.0	7.00	4.00
	Skew.	-0.87	0.67	0.26	0.20	0.96		0.69	0.82	0.72
	Mean	58.6	2.48	11.2	26.0	2.00		10.0	4.00	2.00
4-TJ	St. dev.	7.72	1.30	3.40	17.0	8.00		6.00	3.00	2.00
	Skew.	0.13	0.64	-1.34	3.97	5.11		2.00	2.28	2.20
	Mean	51.2	1.44	15.8	33.0			9.00	3.00	2.00
5-JA	St. dev.	8.89	1.49	2.62	7.00			6.00	3.00	2.00
	Skew.	0.04	2.21	0.02	0.06			0.74	1.77	2.06
	Mean	39.0	0.3	21.9	37.0		5.00	3.00	1.00	0.00
6-AM	St. dev.	6.45	0.38	1.60	3.00		4.00	3.00	1.00	1.00
	Skew.	0.40	1.89	0.05	0.29		0.86	1.39	1.69	2.60
	Mean	32.4	0.04	26.0	12.0		7.00			
7-MH	St. dev.	7.24	0.17	2.05	1.00		3.00			
	Skew.	0.95	6.62	0.57	0.08		-0.40			

$$D_e = (\theta_{FC} - 0.5\theta_{PWP}) \cdot d_e \quad (2)$$

Where D_e is effective precipitation depth to start seed germination (mm), d_e is average seed sowing depth (mm), θ_{FC} and θ_{PWP} are soil water content at field capacity and wilting point (m^3/m^3). Based on the texture of 94 soil samples in the study area, most of which were Clay Loam and Clay (Fig. 2), θ_{FC} and θ_{WP} were 0.32 and 0.16 (m^3/m^3), respectively. According to researchers at the Kurdistan Agricultural and Natural Resources Research and Education Center, the d_e in Saqqez was considered to be approximately 40 mm. Using these values in Eq. (2), D_e was estimated to be approximately 10 mm.

2.5 Mann-Kendall trend test

In this method, the Z statistic, which almost follows the standard normal distribution, is calculated using Eqs. (2) to (5).

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{sgn}(X_j - X_i) \quad , \quad \mathrm{sgn}(X_j - X_i) = \begin{cases} -1 & \text{if } (X_j - X_i) < 0 \\ 0 & \text{if } (X_j - X_i) = 0 \\ 1 & \text{if } (X_j - X_i) > 0 \end{cases} \quad (3)$$

$$\mathrm{Var}(S) = \frac{1}{18} \cdot [n \cdot (n-1) \cdot (2n+5) - \sum_{i=1}^{m} t_i \cdot (t_i - 1) \cdot (2t_i + 5)] \quad (4)$$

$$Z = \begin{cases} \dfrac{S-1}{\sqrt{\mathrm{Var}(S)}} & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ \dfrac{S+1}{\sqrt{\mathrm{Var}(S)}} & \text{if } S < 0 \end{cases} \quad \mathbf{(5)}$$

Where the X_j is the data amount, n is the data number of time series, m is the number of tied groups (a tied group is a set of sample data having the same value), and t_i is the number of data points in the ith tied group. If $|Z|$ is greater than $Z\alpha$, the data series has a trend at the significant level α. Positive and negative Z values indicate an uptrend and a downtrend,

respectively (Yue and Wang 2004 and Pohlert 2020). The corresponding values of $Z_{0.05}$, $Z_{0.01,}$ and $Z_{0.001}$ in the standard normal distribution table are 1.960, 2.576, and 3.291, respectively.

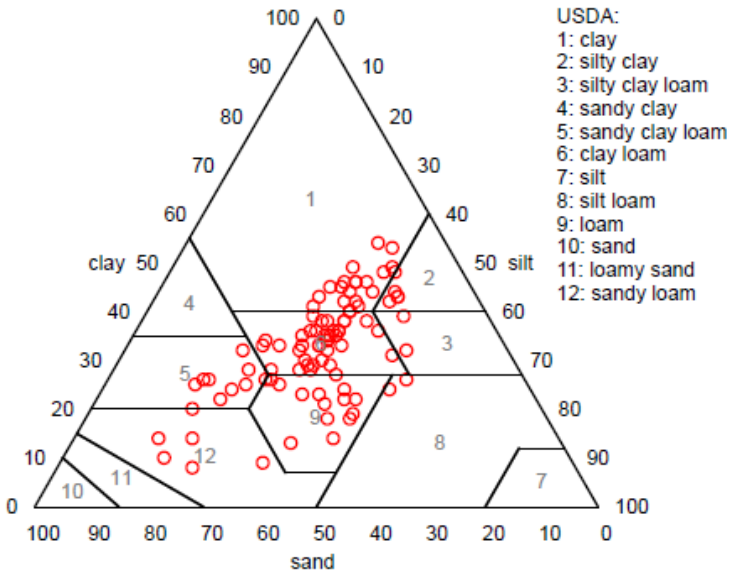

Fig. 2 Texture position of 94 soil samples of the study area in the USDA soil texture triangle

2.6 Factor analysis

In the factor analysis method, the random vector **X** is expressed as a linear combination of common (**F**) and specific (**ε**) factors (matrix Eq. (6)).

$$\underset{(p\times 1)}{\mathbf{X_i}} - \underset{(p\times 1)}{\boldsymbol{\mu_i}} = \underset{(p\times m)}{\mathbf{L_{ij}}} \times \underset{(m\times 1)}{\mathbf{F_j}} + \underset{(p\times 1)}{\boldsymbol{\varepsilon_i}} \tag{6}$$

Where **μ** and **Σ** are the mean and variance-covariance matrices of the vector **X**, *m* and *p* are the numbers of common and specific factors, and $\mathbf{L_{ij}}$ coefficient matrix is the factor loading that specifies the value of the commonalities of factor *j* to the variable *i*.

The following assumptions are considered to perform factor analysis calculations.

- Common factors are independent of each other and have a mean equal to zero and a variance equal to one.
- Specific factors are also independent of each other and have a mean equal to zero and a variance equal to $\mathbf{\Psi}$.
- The common and specific factors are independent of each other.

Based on Eq. (6) and these assumptions, Eq. (7) can be derived. If the dimensions (units) of the variables are different, standardized variables are used. In this case, the variance-covariance matrix is the same as the correlation matrix ($\mathbf{R}$), and the mean ($\boldsymbol{\mu}$) is equal to zero. According to Eq. (8), the initial variables were standardized and dimensionless to have a mean of zero and a standard deviation of one. Therefore, Eq. (6) and Eq. (7) can be written as Eq. (9) and Eq. (10), respectively, and the variance (var ($\mathbf{X}$)) can be obtained from Eq. (11). The portion of var($\mathbf{X}$) that is related to the factors is called commonalities (h^2), can be calculated from Eq. (12). Principal component analysis (PCA) and maximum likelihood are two common methods for estimating communality rates, the first of which was used here.

$$\underset{(p\times p)}{\mathbf{\Sigma}} = \underset{(p\times m)}{\mathbf{L}} \times \underset{(m\times p)}{\mathbf{L}'} + \underset{(p\times p)}{\mathbf{\Psi}} \tag{7}$$

$$Z_{ij} = \frac{X_{ij} - \bar{X}_i}{S_i} \tag{8}$$

$$\mathbf{X_i} = \mathbf{l_{i1}} \times \mathbf{F_1} + \mathbf{l_{i2}} \times \mathbf{F_2} + \mathbf{...} + \mathbf{l_{im}} \times \mathbf{F_m} + \varepsilon_i \tag{9}$$

$$\underset{(p\times p)}{\mathbf{R}} = \underset{(p\times m)}{\mathbf{L}} \times \underset{(m\times p)}{\mathbf{L}'} + \underset{(p\times p)}{\mathbf{\Psi}} \tag{10}$$

$$\mathrm{var}(\mathbf{X_i}) = \boldsymbol{\sigma_{ii}} = \mathbf{l_{i1}}^2 + \mathbf{l_{i2}}^2 + \ldots + \mathbf{l_{ij}}^2 + \ldots + \mathbf{l_{im}}^2 + \mathbf{\Psi_i} \tag{11}$$

$$\mathbf{h^2} = \boldsymbol{\sigma_{ii}} = \mathbf{l_{i1}}^2 + \mathbf{l_{i2}}^2 + \mathbf{...} + \mathbf{l_{ij}}^2 + \mathbf{...} + \mathbf{l_{im}}^2 \tag{12}$$

Where $\mathbf{L}'$ is the $\mathbf{L}$ matrix transpose, Z_{ij} is the standardized data, X_{ij} is the corresponding data of the variable i and the experimental unit j, $\bar{X}_i$ and S_i are the mean and the standard deviation of the variable i, respectively (Härdle and Simar 2015).

2.7 Principle component analysis

In this method, the initial variables, which are usually correlated, are converted to a smaller number of new uncorrelated variables, called principal components, through linear transformations. The principal components in the variance-covariance matrix ($\mathbf{\Sigma}$) can be estimated by the spectral decomposition method (Eq. (13)).

$$\mathbf{\Sigma} = \lambda_1 \mathbf{e_1} \mathbf{e_1}' + \lambda_2 \mathbf{e_2} \mathbf{e_2}' + \ldots + \lambda_p \mathbf{e_p} \mathbf{e_p}' \quad (13)$$

Where λ is eigenvalue (such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$), $\mathbf{e}$ is an eigenvector, $\mathbf{e}'$ is the transpose of $\mathbf{e}$ vector and p is the number of initial variables (Härdle and Simar 2015).

2.8 Multivariate linear regression

The linear relationship between a dependent (response) variable and independent variables (predictors), called multivariate linear regression, is written as Eq. (14).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon \quad (14)$$

Where Y is the dependent variable, β_0, β_1, β_2, ... and β_p are constant and unknown coefficients, X_1, X_2, ... and X_p are independent variables, ε is the residual or random error (Härdle and Simar 2015). The assumptions about ε, on which the regression is based, should be evaluated as follows.

- Residuals are independent of each other;
- Their mean is zero and they have a common variance;
- Residuals follow the normal distribution.

The validity of the assumptions of normality and independence of the residuals were evaluated by Kolmogorov-Smirnov (Karamouz et al. 2003) and Durbin-Watson (Draper and Smith 2014) methods, respectively.

2.9 ANN method

In most ANN models, the mathematical model of the nerve cell, called the neuron or perceptron, is used. The neuron is the smallest unit of the neural network. Each network consists of an input layer, an output layer, and one or more intermediate (hidden) layers. Inside each of these layers are several neurons that are connected by weighted connections. During the network training process, these weights change sequentially to minimize errors. Linear, sigmoid, and hyperbolic tangent transfer functions are commonly used to transfer the output of each layer to subsequent layers. The topology of each network shows the relative status (number, grouping, and connections) of cells in the network. Topology is the system of connecting neurons to each other, which, together with the relevant software (i.e. the mathematical method of information flow and calculation of weights), determines the type of performance of the neural network. The simplest type of network has the feedforward topology, in which the flow of information is always from input to output. The topology of multilayer perceptron networks is complemented by the error backpropagation learning rule which has various training algorithms, such as conjugate gradient, momentum, and Levenberg-Marquart.

To develop the ANN model structure, the data are divided into three groups. The first group of data, as training data, is used to determine the weights and biases of the network. The second group of data, called validation data, is used to evaluate the training phase results and make decisions about stopping network training. The test of the model is determined

using the third group of data that was not used in compiling the model. The mathematical form of the ANN is expressed as Eq. (15):

$$O_k = S(\sum_{j=1}^{m} W_{jk} \cdot S(\sum_{i=1}^{n} W_{ij} \cdot X_i)) \quad (15)$$

Where, O_k and X_i are the output and input values of the network, W_{ij} is the communication weights between the input layer and the hidden layer, W_{jk} is the communication weights between the hidden layer and the output layer and S is the transfer function. Also, n, m, and p are the number of inputs, hidden layers, and outputs, respectively (Karamouz et al. 2003).

In this study, the ANN model with the multilayer feedforward topology was used to estimate the rainfed wheat yield in the Saqqez region. In designing the model structure, first, the number of neurons in the input and output layers was selected according to the number of input variables (10 factors) and model output (rainfed wheat yield), respectively. Then, by changing the adjustable parameters (including transfer function, learning rule, momentum value, number of hidden layers, number of hidden layer neurons, number of epochs, and input variables), a large number of ANN models with different structures were designed and evaluated. The accuracy of these models was evaluated by calculated statistical criteria. Finally, the model that had the closest results to the actual results was selected as the main model.

2.10 GEP Method

Genetic expression programming (GEP) uses the populations of individuals, selects them according to fitness, and introduces genetic variations using one or more genetic operators. Therefore, it such as genetic algorithms (GAs) and genetic programming (GP) is a genetic algorithm (Mitchell 1996). In GEP the individuals (chromosomes) are composed of genes that are structurally organized in a head and tail. The chromosomes function as a genome. They

are encoded as linear strings of fixed length which are afterward expressed as nonlinear entities of different sizes and shapes. They are subjected to modification using mutation, transposition, root transposition, gene transposition, gene recombination, and one- and two-point recombination. The creation of these separate entities (genome and expression trees (ET)) with distinct functions allows the algorithm to perform with high efficiency that greatly surpasses existing adaptive techniques (Ferreira 2001).

In this study, GeneXproTools 5.0 program (https://www.gene-expression-programming.com) was used to model rainfed wheat yield through GEP. The main steps and parameter settings used to design the GEP model are presented in Table 3. See more details about the GEP in Ferreira (2001).

2.11 Model evaluation criteria

Root mean square error (RMSE) (Eq. (16)), mean absolute percentage error (MAPE) (Eq. (17)), and correlation coefficient (r) (Eq. (18)) were used to evaluate the accuracy of the models. The lower the RMSE and MAPE values, and the closer they are to zero, the better the model performance. The coefficient of determination (i.e. r^2) is between zero and one, and the closer it is to one, the more accurate the model (Härdle and Simar 2015, Karamouz et al. 2003 and Kim and Kim 2016).

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (Y_{obs} - Y_{est})^2} \tag{16}$$

$$\mathrm{MAPE} = \frac{100}{n} \cdot \sum_{i=1}^{n} \left| \frac{Y_{obs} - Y_{est}}{Y_{obs}} \right| \tag{17}$$

$$r = \frac{\sum_{i=1}^{n} \left(\left(Y_{obs} - \bar{Y}_{obs} \right) \cdot \left(Y_{est} - \bar{Y}_{est} \right) \right)}{\sqrt{\sum_{i=1}^{n} \left(Y_{obs} - \bar{Y}_{obs} \right)^2 \cdot \sum_{i=1}^{n} \left(Y_{est} - \bar{Y}_{est} \right)^2}} \quad \text{and always} \quad -1 \le r \le 1 \tag{18}$$

Where Y_{obs} and $\bar{Y}_{obs}$ are the observed (actual) values and their mean, Y_{est} and $\bar{Y}_{est}$ are the estimated values (model output) and their mean, and n is the number of observations.

Table 3 Main steps and parameter settings used to design the GEP model

Step	Settings and parameters (used in this study)
Selection of the fitness function	RMSE
Selection of the terminals and functions	Regression default (+, -, *, /, exp(x), ln(x), 1/x, x^2, $x^{1/3}$, avg(x_1,x_2), arctan(x), tanh(x), (1-x))
Determine the structure of chromosomes	(Number of chromosomes: 30), (Head size: 10), (Number of genes: 4)
Determining the linking function	Addition function
Determine the characteristics of the operators	(Root insertion sequence transposition: 0.00546), (Insertion sequence transposition rate: 0.00546), (Gene transposition rate: 0.00277), (Two-point recombination rate: 0.00277), (One-point recombination rate: 0.00277), (Gene recombination rate: 0.00277), (Mutation rate: 0.00138), (Inversion rate: 0.00456)
Determining the stop condition of model training and iteration of modeling	(Generation number: 10000); (Runs: 150)
Evaluation of the models and selection of the most accurate model	RMSE and MAPE

3 Results and discussion

3.1 Correlation matrix

Forty-seven variables and indicators (see Table 1) were entered into SPSS and XLSTAT software to perform factor analysis calculations. The correlation matrix determinant was approximately zero. This indicates that there is a correlation between the used variables that is

suitable for factor analysis (Härdle and Simar 2015). Because the correlation matrix (47 x 47) was large, it was not presented here.

3.2 Eigenvalues and eigenvectors

Initial and rotated (by varimax method) eigenvalues were extracted from the correlation matrix, the results of which are presented in Fig. 3. Ten components had specific values greater than one that was selected as the principal factors. The first ten components, which had eigenvalues greater than one, were selected as the main factor (Härdle and Simar 2015). These 10 factors explained 90.7% of the total variance of the 47 variables used.

3.3 Factor loading

Preliminary calculation of the factors showed that some variables have a high correlation with several factors. This made it difficult to interpret the results. In such cases, the factors are rotated to make their structure simpler and more interpretable. To rotate the factors, the conventional Varimax method (Härdle and Simar 2015) was used. The pattern (structure) of the rotating factors is presented in Table 4. Based on Table 4 and the phenological stages of rainfed wheat, the factors (for use in modeling) were calculated and (for use in interpreting the results) were named. The results of naming the factors are presented in Table 5.

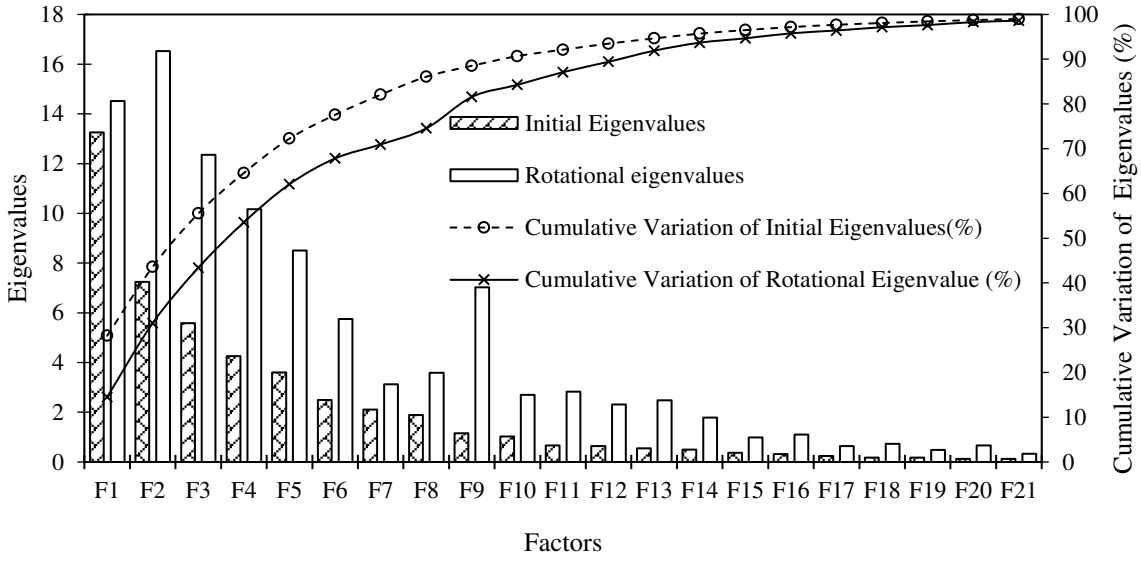

Fig. 3 Initial and rotational eigenvalues and their variation percentage from the total variance

Table 4 Rotational factor loading matrix

Variables-growth stage	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	h^2
RH-SE	0.252	0.006	0.567	-0.245	0.034	-0.001	-0.060	0.585	0.126	-0.102	0.818
RH-ET	0.241	-0.126	-0.184	0.116	0.043	0.141	0.467	0.432	0.489	0.205	0.829
RH-TT	0.380	0.777	-0.072	-0.153	0.089	0.081	0.072	0.011	0.297	0.177	0.916
RH-TJ	0.425	0.050	-0.017	-0.089	0.675	-0.052	0.203	0.163	0.238	0.343	0.891
RH-JA	0.449	-0.069	-0.191	0.593	0.218	-0.104	0.424	0.160	0.055	0.278	0.938
RH-AM	0.278	-0.165	-0.240	0.289	0.245	0.095	0.514	0.292	0.053	0.296	0.755
P-SE	0.196	-0.030	0.044	-0.042	0.262	-0.054	-0.002	0.830	-0.091	-0.113	0.824
P-ET	0.054	-0.018	-0.156	0.093	-0.035	-0.009	0.840	0.015	0.323	0.051	0.850
P-TT	-0.286	0.176	0.167	-0.202	-0.054	0.064	-0.063	-0.238	0.539	0.576	0.872
P-TJ	-0.199	0.147	0.051	0.166	0.821	0.123	-0.006	0.206	0.065	0.010	0.828
P-JA	-0.021	0.063	-0.113	0.960	0.059	0.149	0.079	-0.017	-0.093	0.023	0.979
P-AM	0.066	0.158	-0.051	0.304	0.221	0.867	0.053	-0.011	-0.114	0.014	0.941
T-SE	0.403	0.357	-0.724	0.230	0.231	0.002	0.073	0.025	-0.051	-0.022	0.929
T-ET	0.176	0.690	0.276	0.078	0.336	-0.077	0.036	-0.432	-0.122	-0.195	0.948
T-TT	-0.283	-0.808	0.351	-0.171	0.032	-0.135	-0.020	-0.102	0.217	0.022	0.963
T-TJ	-0.797	-0.129	0.302	-0.052	-0.249	-0.111	-0.095	0.019	0.280	0.127	0.925
T-JA	-0.770	-0.071	0.304	-0.367	-0.187	-0.079	-0.139	-0.100	0.233	0.062	0.953
T-AM	-0.878	-0.128	0.237	-0.128	-0.108	-0.153	-0.108	-0.124	-0.025	0.008	0.922
L-SE	-0.334	-0.186	0.859	-0.089	-0.071	-0.061	-0.047	-0.156	-0.189	0.035	0.965
L-ET	-0.154	-0.730	-0.369	-0.153	-0.211	0.015	0.022	0.280	0.178	0.319	0.974
L-TT	0.096	0.912	-0.305	-0.003	0.135	0.032	-0.017	-0.014	0.069	-0.146	0.980
L-TJ	0.790	0.035	-0.164	0.051	0.339	0.046	0.097	-0.169	-0.230	-0.190	0.898
L-JA	0.707	-0.015	-0.166	0.340	0.078	0.064	0.119	0.060	-0.541	0.026	0.965
L-AM	0.879	0.131	-0.224	0.095	0.115	0.134	0.166	0.091	-0.001	-0.066	0.921
T_0-SE	-0.301	-0.099	0.856	-0.071	-0.014	-0.113	-0.047	-0.214	-0.163	0.032	0.927
T_0-ET	-0.228	-0.716	-0.378	-0.225	-0.183	0.005	0.062	0.233	0.082	0.330	0.965
T_0-TT	-0.056	0.889	-0.275	-0.011	0.071	-0.002	0.045	0.011	-0.206	0.038	0.921
T_0-TJ	0.295	0.118	0.047	-0.122	-0.125	0.039	0.685	-0.199	-0.093	-0.198	0.692
T_{25}-AM	-0.788	-0.212	0.155	0.009	0.023	-0.169	0.061	-0.195	-0.125	-0.131	0.793

P_1-SE	-0.333	-0.169	0.883	-0.151	0.065	-0.015	-0.058	-0.011	-0.046	-0.080	0.958
P_1-ET	-0.232	-0.672	-0.283	-0.163	-0.024	-0.086	0.183	0.072	0.527	0.189	0.973
P_1-TT	0.010	0.893	-0.201	-0.227	-0.134	0.030	-0.017	0.144	-0.168	0.061	0.961
P_1-TJ	0.366	0.086	0.028	-0.194	0.833	0.156	0.107	0.087	0.035	-0.011	0.919
P_1-JA	0.531	-0.079	-0.078	0.755	0.040	0.137	0.185	-0.045	-0.088	-0.086	0.936
P_1-AM	0.395	0.070	0.026	0.191	-0.057	0.788	0.025	0.017	0.068	0.053	0.831
P_5-SE	-0.230	-0.167	0.855	-0.106	0.120	-0.060	-0.051	0.240	-0.016	-0.145	0.923
P_5-ET	-0.157	-0.418	-0.243	-0.125	0.075	-0.079	0.165	0.048	0.816	0.051	0.983
P_5-TT	-0.089	0.903	-0.212	0.005	-0.066	0.120	0.041	0.006	-0.094	0.260	0.965
P_5-TJ	0.235	0.025	0.003	0.157	0.887	0.027	-0.066	0.057	-0.100	-0.070	0.891
P_5-JA	0.178	0.050	-0.111	0.946	0.106	0.095	-0.037	-0.051	-0.045	-0.007	0.968
P_5-AM	0.169	0.047	-0.067	-0.071	0.183	0.930	-0.011	0.021	-0.096	0.074	0.955
P_{10}-SE	-0.090	-0.114	0.725	-0.092	0.148	0.066	-0.111	0.494	-0.093	-0.004	0.847
P_{10}-ET	-0.147	-0.241	-0.153	-0.119	0.120	-0.101	0.148	-0.046	0.901	-0.004	0.977
P_{10}-TT	-0.063	0.782	-0.257	0.108	-0.018	0.131	-0.047	-0.039	0.018	0.437	0.905
P_{10}-TJ	0.148	-0.005	0.030	0.261	0.841	0.154	-0.149	-0.077	0.068	0.020	0.856
P_{10}-JA	0.046	0.046	-0.161	0.939	0.085	0.079	-0.105	-0.079	-0.116	0.008	0.955
P_{10}-AM	-0.071	0.103	-0.180	0.187	0.137	0.499	0.026	-0.154	0.038	0.608	0.745

Table 5 Results of factor naming

Factor	Variable with a large positive coefficient	Variable with a large negative coefficient	Factor name
F_1	(L-AM,0.88) (L-TJ,0.79) (L-JA,0.71)	(T-AM,0.88) (T-TJ,0.80) (T_{25}-AM,0.79) (T-JA,0.77)	Temperature effect in the TJ, JA, and AM stages
F_2	(L-TT,0.91) (P_5-TT,0.90) (P_1-TT,0.89) (T_0-TT,0.89) (P_{10}-TT,0.78) (R-TT,0.78) (T-ET,0.69)	(T-TT,0.81) (L-ET,0.73) (T_0-ET,0.72)	The effect of precipitation distribution and humidity in the TT stage and the temperature effect in TT and ET stages
F_3	(P_1-SE,0.88) (L-SE,0.86) (T_0-SE,0.86) (P_5-SE,0.85) (P_{10}-SE,0.72)	(T-SE,0.72)	The effect of precipitation distribution and temperature in the SE stage
F_4	(P-JA,0.96) (P_5-JA,0.95) (P_{10}-JA,0.94) (P_1-JA,0.75) (R-JA,0.59)		The effect of precipitation and humidity in the JA stage
F_5	(P_5-TJ,0.89) (P_{10}-TJ,0.84) (P_1-TJ,0.83) (P-TJ,0.82) (R-TJ,0.67)		The effect of precipitation and humidity in the TJ stage
F_6	(P_5-AM,0.93) (P-AM,0.87) (P_1-AM,0.79) (P_{10}-AM,0.50)		Precipitation effect in AM stage
F_7	(P-ET,0.84)		Precipitation amount effect during ET stage
F_8	(P-SE,0.83) (R-SE,0.59)		The effect of precipitation amount and humidity in the SE stage

F_9	(P_{10}-ET,0.90) (P_5-ET,0.82) (P_1-ET,0.53) (R-ET,0.49)	Precipitation distribution effect in ET stage
F_{10}	(P-TT,0.58)	Precipitation amount effect in tillering (TT) stage

3.4 MLR, ANN, and GEP models

At this step, the effect of ten factors (F_1 to F_{10}) as independent variables on the dependent variable, namely rainfed wheat yield (Y), was analyzed using MLR, ANN, and GEP methods. The models for estimating rainfed wheat yield based on MLR and GEP methods were obtained in the form of Eqs. (19) and (20), respectively. The ANN method is not able to provide an explicit equation that shows the relationships between variables. The best ANN model had a hyperbolic tangent transfer function, the conjugate gradient learning rule, and a hidden layer with four neurons.

$$Y_{MLR} = 0.273\times F_1 - 0.162\times F_2 - 0.294\times F_4 + 0.240\times F_5 + 0.240\times F_7 + 0.476\times F_9 \quad ; \ r^2 = 0.53 \tag{19}$$

$$\begin{aligned} Y_{GEP} &= ((1-(1-(((9.81+F_8)/2)/(F_{10}\times F_4))))\times((1-F_9)+(1-F_8))^{(1/3)}) \\ &+ ((F_6\times((-6.69\times F_8)-6.69))^2-(5.90\times F_4-F_1^2)^2) \\ &+ (1-(9.87^2\times(-8.04-F_7)-F_5)-(1/(1-(F_4+F_2)))) \quad ; \quad r^2 = 0.82 \\ &+ (((-9.43+F_4)^2+28.2)\times ATAN(12.2\times F_9)) \end{aligned} \tag{20}$$

Where Y and Fs are the yield of rainfed wheat and factors, respectively. The coefficient of determination of Eq. (20) shows that the relationship between the yield of rainfed wheat with climatic factors is significant and that approximately 82% of yield changes (in Saqqez) can be justified and modeled by these factors. This significant relationship between wheat yield and climate in other regions has been confirmed and based on it, other researchers have provided suitable models to predict the yield of this plant (Faghih et al. 2020; Mukherjee et al. 2019; Lee et al. 2013; Qian et al. 2009; Bal et al. 2004).

To evaluate the models, the yield values of rainfed wheat in Saqqez for three data sets, training (28 years), validation (8 years), and testing (10 years) were estimated using the models. Results from the evaluation of the developed models are presented in Table 6 and plotted in Fig. 4. According to the RMSE criterion for the total of three data sets, the accuracy of the models developed by the GEP and ANN methods is almost equal and is 21.5% higher than the MLR model. Also, based on this criterion, the accuracy of the GEP model in estimating validation data that was not used in model development was 28 and 37% higher than ANN and MLR models, respectively. The performance of all three models for estimating testing data, not used in model development, was almost the same. Since the MAPE value for the GEP and ANN models was 26% and for the MLR model was 35%, it can be concluded that all three models can predict wheat yield based on climatic factors. The ability of the MLR and ANN methods to model crop yield has also been confirmed in the studies of Bal et al. (2004), Lee et al. (2013), Ceglar et al. (2016), Chen and Jing (2017), Dornelles et al. (2018), Khaki and Wang (2019), Adisa et al. (2019) and Faghih et al. (2020).

In this study, the GEP method was used for the first time in this field, and comparing its results with two other methods indicates that it is suitable for modeling rainfed wheat yield.

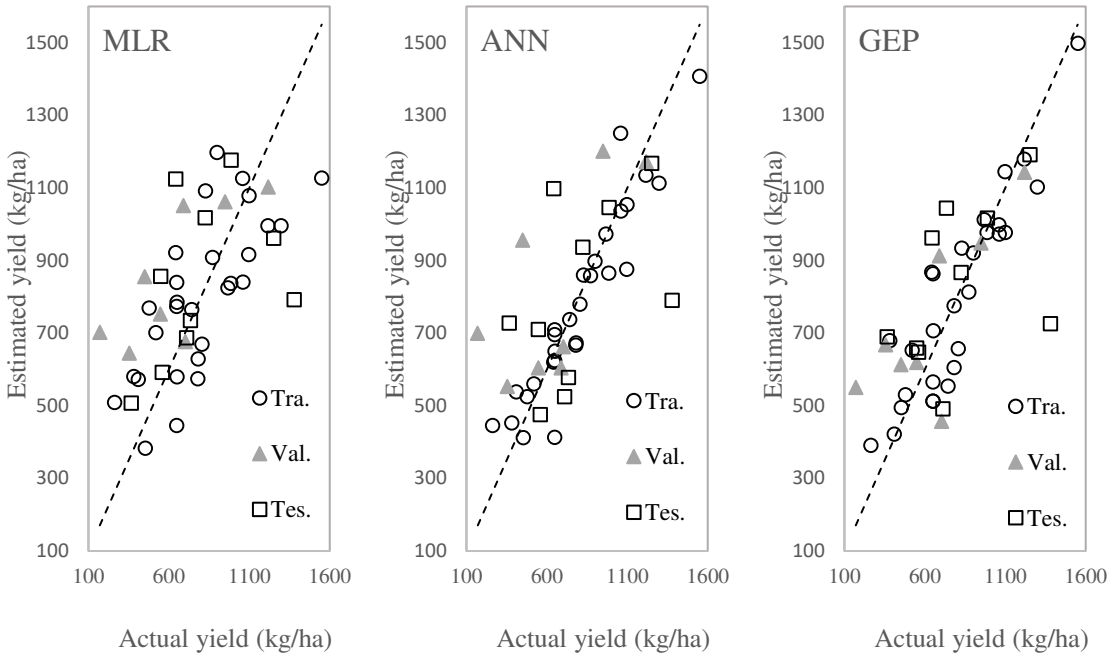

Fig. 4 The relationship between observed and estimated wheat yield using the models

Table 6 Results from the evaluation of developed models based on rainfed wheat yield and effective climatic factors in its growth stages for Saqqez using MLR, ANN, and GEP methods

Model	RMSE (kg/ha)				MAPE (%)			
	Tra.	Val.	Tes.	Total	Tra.	Val.	Tes.	Total
MLR	201	302	292	243	26.1	74.8	28.7	35.1
ANN	106	286	283	196	11.8	67.2	33.0	26.1
GEP	126	221	284	189	16.1	54.3	30.7	25.9

The comparison of models showed that GEP and ANN models are more accurate than the MLR model. But the structure of the MLR model (Eq. 19) is much simpler than the other two, and therefore it is easier to use to interpret the results. It can be inferred from Eq. (19) that the role of the ninth factor (i.e. precipitation distribution effect in ET stage) was more than other factors. The other five factors that had an important role in the rainfed wheat yield included the fourth factor (i.e. the effect of precipitation and humidity in the JA stage), the first factor

(i.e. temperature effect in the TJ, JA, and AM stages), the fifth factor (i.e. the effect of precipitation and humidity in the TJ stage), the seventh factor (i.e. precipitation amount effect during ET stage), and the second factor (i.e. the effect of precipitation distribution and humidity in the TT stage and the temperature effect in TT and ET stages), respectively.

To estimate the rainfed wheat yield (Y) in each year, the factors (F_1 to F_{10}) were calculated by Eq. (6) and placed in Eqs. (19) and (20). These factors were used as input variables of the input layer in designing the structure of the ANN model. Also, based on Eq. (19) and the factor loading, the following important results can be deduced.

- In the tillering stage (TT), which coincides almost with the cold season (winter), the rainfed wheat yield was directly related to air temperature and inversely related to growth stage duration. But in other stages, these relationships were the opposite. Faghih et al. (2020), Mukherjee et al (2019), and Lee et al. (2013) reported the inverse relationship of rainfed wheat yield with air temperature and its direct relationship with growth stage duration.
- In the JA (flowering) and TT stages, the rainfed wheat yield had an inverse relationship with precipitation, its temporal distribution, and relative humidity. But in other stages, the relationship is direct. The direct relationship of rainfed wheat yield with precipitation and relative humidity was reported by Faghih et al. (2020), Mukherjee et al. (2019), and Lee et al. (2013).
- In both stages, JA and TJ, all variables and indicators had a strong effect on rainfed wheat yield. In the ET stage, the distribution and the amount of precipitation and relative humidity had a greater effect on performance than other variables. At the AM stage, the effect of air temperature on yield was greater than that of other variables. In the TT stage, the effect intensity of all variables was almost the same and moderate. According to this,

the sensitive stages of wheat to climate variables were JA, TJ, ET, TT, and AM respectively.

- The stages of TJ and JA in Saqqez are approximately simultaneous with April and May. Climatic variables of these months played an important role in changing the yield of rainfed wheat compared to other months. Therefore, the yield of rainfed wheat is significantly more affected by spring climatic variables than other seasons. . Kheiri et al. (2017) and Faghih et al. (2020) reported similar results to the results of this study.
- Wheat yield sensitivity to climatic variables included precipitation, air temperature, and relative humidity, respectively.

3.5 Trend

The trend test results of the rainfed wheat yield and the climatic factors are presented in Table 7. It is observed that wheat yield has an upward trend at a significant level of lower than 0.1%. Of the six factors affecting rainfed wheat yield, four factors F_1, F_4, F_5, and $F9$ had significant trends.

According to the results (Table 7), it can be stated that the increasing trend of F_9 was due to the increasing trend of humidity and precipitation temporal distribution in the ET stage. Based on Eq. (19), the relationship between F_9 and yield is direct, so the increasing trend of this factor is influential on the incremental trend of yield. F_1 also had a direct relationship with yield. The upward trend of this factor was due to the downward trend of air temperature in the AM and TJ stages, which caused an upward trend of yield.

F_4 and F_5 had a significant downward trend. The decreasing trend in precipitation and its distribution in two stages, JA and TJ, caused a decreasing trend in these two factors. F_4 was inversely related to yield, and its downward trend could affect the yield increase. But the F_5 was directly related to yield. The downward trend in this factor should decrease yield, but its

inhibitory effect on yield growth has been compensated by lowering the air temperature at the TJ stage (i.e. F_1).

Finally, it can be concluded that much of the increasing trend of rainfed wheat yield in the Saqqez region has been due to the decreasing trend of temperature in the AM and TJ stages, the increasing trend of relative humidity and precipitation in the ET stage, and the decreasing trend of precipitation in the JA stage (flowering). Nassiri et al. (2006), Pishbahar and Darparnian (2016), Saei et al. (2019), and Faghih et al. (2020) also reported a significant trend in wheat yield due to significant trends in climate variables for different regions of Iran.

Table 7 The trend test results of the rainfed wheat yield and the climatic factors in Saqqez

	Yield	Effective factors					
		F_1	F_2	F_4	F_5	F_7	F_9
Z-MK	16.5^{***}	10.3^{***}	-0.56^{ns}	-4.94^{***}	-9.80^{***}	-0.93^{ns}	6.52^{***}
(Climatic variable - Stage, Z-MK)		(T-TJ, -2.33^{*}) (T-JA, -0.95^{ns}) (T-AM, -3.19^{**}) (T_{25}-AM, -3.54^{***}) (L-TJ, 1.78^{ns}) (L-JA, 0.84^{ns}) (L-AM, 4.31^{***})		(P-JA, -4.43^{***}) (P_1-JA, -2.66^{**}) (P_5-JA, -5.43^{***}) (P_{10}-JA, -4.74^{***}) (R-JA, 0.61^{ns})	(P-TJ, -10.3^{***}) (P_1-TJ, 0.16^{ns}) (P_5-TJ, -5.15^{***}) (P_{10}-TJ, -1.61^{ns}) (R-TJ, 2.48^{*})		(P_1-ET, 3.65^{***}) (P_5-ET, 3.79^{***}) (P_{10}-ET, 4.67^{***}) (R-ET, 6.24^{***})

Signs *, ** and *** indicate a trend at the significance level of 5%, 1%, and 0.1%, respectively. Also, *ns* indicates that there is no significant trend in the data series.

4 Conclusion

In this research, the effect of climatic variables on the yield of rainfed wheat in Saqez, Iran, was determined and modeled. Its important results can be presented as follows:

- Climatic variables always interact with each other. The impact of each climatic variable on agriculture cannot be considered separately.
- Rainfed wheat yield in the Saqqez region is highly correlated with climate conditions each year. Based on this high correlation, appropriate models have been developed to predict the

wheat yield by the MLR, ANN, and GEP methods. These models can acceptably predict rainfed wheat yield before harvesting.

- In this study, the GEP method was used for the first time for modeling rainfed wheat yield. Comparing its results with the MLR and ANN methods, which have proven their ability to model crop yield, indicates that it is suitable for modeling in this field.
- The comparison of models showed that GEP and ANN models are more accurate than the MLR model. But the structure of the MLR model is much simpler than the other two, and therefore it is easier to use to interpret the results.
- The yield of rainfed wheat during its growing season, except for the cold period, shows a direct relationship with precipitation and relative humidity and an inverse relationship with air temperature. Furthermore, the relationship between yield and precipitation in the flowering stage was inverse.
- rainfed wheat yield is significantly more affected by spring climatic variables than other seasons. Also, the climatic variables of April and May played an important role in changing the yield of wheat compared to other months.
- The impact intensity of climatic variables and indices on wheat yield, respectively, included precipitation and its distribution, air temperature, growth-stage length, and relative humidity.
- The significant trend of climatic variables is the main factor in the upward trend of rainfed wheat yield in the Saqqez region.

Authors' contributions All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by H. F. The first draft of the

manuscript was written by H. F. and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding No funding was received for conducting this study.

Availability of data

All data used for this study are available from public institutions (Ministry of Agriculture-Jahad and Meteorological Organization of Iran).

Code availability

GeneXproTools 5.0 is commercial software and can be accessed at:
https://www.gene-expression-programming.com
Other software used in this research is free.

Declarations

Ethics approval All authors comply with the guidelines of the journal Theoretical and Applied Climatology.

Consent to participate All authors agreed to participate in this study.

Consent for publication All authors agreed to the publication of this study.

Conflict of interest None.

References

Adisa, O. M., Botai, J. O., Adeola, A. M., Hassen, A., Botai, C. M., Darkey, D., Tesfamariam, E. 2019. Application of artificial neural network for predicting maize production in South Africa. *Sustainability*, 11(4), 1145. https://doi.org/10.3390/su11041145

Allen, R. G., Pereira, L.S., Smith, D. R. 1998. *Crop evapotranspiration (guidelines for computing crop water requirements).* Rome, Italy: FAO Irrigation and Drainage Paper No. 56.

Arkian, F., Nicholson, S. E., Ziaie, B. 2018. Meteorological factors affecting the sudden decline in Lake Urmia's water level. *Theoretical and Applied Climatology*, 131:641–651. https://doi.org/10.1007/s00704-016-1992-6

Bal, S., Mukherjee, J., Mallick, K., Hundal, S. S. 2004. Wheat yield forecasting models for Ludhiana district of Punjab state. *Journal of Agrometeorology*, 6:161-165.

Bazgeer, S., Kamali, G., Sedaghatkerdar, A., Moradi, A. 2008. Pre-harvest wheat yield prediction using agrometeorological indices for different regions of Kurdistan province, Iran. *Research Journal of Environmental Sciences*, 2(4):275-280.

Bowden, P., Edwards, J., Ferguson, N., McNee, T., Manning, B., Roberts, K. et al. 2008. *Wheat growth and development.* Department of Primary Industries, New South Wales.

Ceglar, A., Toreti, A., Lecerf, R., Van der Velde, M., Dentener, F. 2016. Impact of meteorological drivers on regional inter-annual crop yield variability in France. *Agricultural and Forest Meteorology*, 216:58-67

Chen, P., Jing, Q. 2017. A comparison of two adaptive multivariate analysis methods (PLSR and ANN) for winter wheat yield forecasting using Landsat-8 OLI images. *Advances in Space Research*, 59:987-995.

Dornelles, E. F., Kraisig, A. R., Da Silva, J. A. G., Sawicki, S., Roos-Frantz, F., Carbonera, R. 2018. Artificial intelligence in seeding density optimization and yield simulation

for oat. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 22(3):183-188. http://dx.doi.org/10.1590/1807-1929/agriambi.v22n3p183-188

Draper, N. R., Smith, H. 2014. *Applied regression analysis* (Third ed.) (Draper, N. R., Smith, H., Eds.) John Wiley and Sons Inc, Hoboken. https://doi.org/10.1002/9781118625590.ch7

Esfandiary, F., Aghaie, G., Dolati Mehr, A. 2009. Wheat yield prediction through agrometeorological indices for Ardebil district. *World Academy of Science Engineering and Technology*, 49:32-35

Faghih, H., Behmanesh, J., Rezaie, H., Khalili, K. 2021. Climate and rainfed wheat yield. *Theoretical and Applied Climatology*. https://dx.doi.org/10.1007/s00704-020-03478-9

Faghih, H. 2018. *Water requirement estimation of main field crops in Kurdistan*. Soil and Water Research Institute, Karaj Iran. (In Persian).

Ferreira, C. 2001. Gene expression programming: a new adaptive algorithm for solving problems. *Complex Systems*, 13(2):87-129.

Härdle, W. K., Simar, L. 2015. *Applied multivariate statistical analysis*. Springer, Berlin. https://doi.org/10.1007/978-3-662-45171-7

Karamouz, M., Szidarovszky, F., Zahraie, B. 2003. *Water resources system analysis*. Lewis Publishers, Washington

Khaki, S., Wang, L. 2019. Crop yield prediction using deep neural networks. *Frontiers in Plant Science,* 10:621. https://doi.org/10.3389/fpls.2019.00621

Khazaei, B., Khatami, S., Alemohammad, S. H., Rashidi, L., Wu, C., Madani, K., Kalantari, Z., Destouni, G., Aghakouchak, A. 2019. Climatic or regionally induced by humans? Tracing hydro-climatic and land-use changes to better understand the Lake Urmia tragedy. *Journal of Hydrology*, 569**:**203–217. https://doi.org/10.1016/j.jhydrol.2018.12.004

Kheiri, M., Soufizadeh, S., Ghaffari, A., AghaAlikhani, M., Eskandari, A. 2017. Association between temperature and precipitation with dryland wheat yield in the northwest of Iran. *Climatic Change*, 141(4):703-717. https://doi.org/10.1007/s10584-017-1904-5

Kim, S., Kim, H. 2016. A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32:669-679

Lamba, V., Dhaka, V. S. 2014. Wheat yield prediction using artificial neural network and crop prediction techniques (a survey). *International Journal for Research in Applied Science and Engineering Technology*, 2(IX):330-341.

Lee, B., Kenkel, P., Wade Brorsen, B. 2013. Pre-harvest forecasting of county wheat yield and wheat quality using weather information. *Agricultural and Forest Meteorology*, 168:26-35.

Mitchell, M. 1996. *An introduction to genetic algorithms*. MIT Press, Cambridge.

Mukherjee, A., Wang, S., Promchote, P. 2019. Examination of the climate factors that reduced wheat yield in Northwest India during the 2000s. *Water*, 11(2):343. https://doi.org/10.3390/w11020343

Nassiri, M., Koochechi, A., Kamali, G., Shahandeh, H. 2006. Potential impact of climate change on rainfed wheat production in Iran. *Archives of Agronomy and Soil Science*, 52(0):1-12.

Pishbahar, E., Darparnian, S. 2016. Factors creating systematic risk for rainfed wheat production in Iran, using the spatial econometric approach. *Journal of Agricultural Science and Technology*, 18(4):895-909.

Pohlert, T. 2020. Non-parametric trend tests and change-point detection. Available at: https://cran.r-project.org/web/packages/trend/vignettes/trend.pdf. Accessed 25 November 2020

Qian, B., Jong, R., Warren, R., Chipanshi, A., Hill, H. 2009. Statistical spring wheat yield forecasting for the Canadian Prairie provinces. *Agricultural and Forest Meteorology*, 149:1022-1031.

Saei, M., Mohammadi, H., Ziaee, S., Barkhordari, S. 2019. The impact of climate change on grain yield and yield variability in Iran. *Iranian Economic Review*, 23(2):509-531.

Smith, R., Adams, J., Stephens, D., Hick, P. 1995. Forecasting wheat yield in a Mediterranean-type environment from the NOAA satellite. *Australian Journal of Agriculture Research*, 46(1):113-125

Yue, S., Wang, C. 2004. The Mann-Kendall test modified by effective sample size to detect trend in serially correlated hydrological series. *Water Resources Management*, 18:201–218.

Zadoks, J. C., Chang, T. T., Konzak, C. F. 1974. A decimal code for the growth stages of cereals. *Weed Research*, 14:415-421.

Zoljoodi, M., Didevarasl, A. 2014. Water-level fluctuations of Urmia Lake: Relationship with the long-term changes of meteorological variables (solutions for water-crisis management in Urmia Lake basin). *Atmospheric and Climate Sciences*, 4(3):358-368. https://www.scirp.org/journal/paperinformation.aspx?paperid=47665