Shahid Beheshti University

Sustainable Earth Trends

Journal homepage: http://sustainearth.sbu.ac.ir

Multivariate statistical analysis and machine learning methods to predict grain yield in barley (Hordeum vulgare L.) in dry regions of Fars Province

Ali Behpouri[a*], Ehsan Bijanzadeh[a]

[a] Department of Agroecology, College of Agriculture and Natural Resources of Darab, Shiraz University, Shiraz, Iran

A B S T R A C T

This study aimed to identify the importance of farm management variables that affect grain yield in barley. Barley is a significant cereal crop that farmers typically cultivate in poor, saline, and dryland regions around the world. Data corresponding to 15 agronomic variables and grain yield were collected from 104 farms in southern parts of Fars Province, Iran. Multivariate statistical analysis (stepwise linear regression, correlation, Principal component analysis (PCA)) and machine learning modeling techniques, such as support vector regression (SVR) models and partial least squares regression (PLSR), were applied to agronomic and farm management variables influencing barley grain yield under dry regions of south parts of Fars Province. The results of multivariate statistical analysis showed that barley grain yield had positive correlations with most of the studied variables except for pest damage, disease damage, number of weeds m-2, seeding depth, and salinity level. The highest positive correlation coefficients for grain yield in this study were obtained between grain yield and irrigation (0.860**). The results of stepwise regression analysis showed that irrigation (x4), salinity level (x11), Phosphorous fertilizer application (x14), and weeds infestation percentage (x8), justified the maximum grain yield in barley. The results of the 3 statistical modeling methods were close to each other and the highest R^2 (0.79) belonged to the stepwise linear regression method.

A R T I C L E I N F O

Keywords:
Drought condition
Regression analysis
SVR
PLSR
PCA

Article history:
Received: 05 Jan 2025
Accepted: 26 Jan 2025

*Corresponding author
E-mail address:
behpoori@shirazu.ac.ir
(A. Behpouri)

Citation:
Behpouri, A. & Bijanzadeh, E., (2025). Multivariate statistical analysis and machine learning methods to predict grain yield in barley (Hordeum vulgare L.) in dry regions of Fars Province, *Sustainable Earth Trends:* 5(4), (22-29).

DOI: 10.48308/set.2025.238252.1105

1. Introduction

Barley (Hordeum vulgare L.) is cultivated in nearly all parts of the world for human consumption and animal feed. Its popularity stems from the fact that this crop can be grown in dry regions where the cultivation of other crops is restricted due to environmental limitations. Barley is the fourth most cultivated crop in the world following wheat, maize, and rice in terms of growing area and production (Czembor et al., 2022). Yield forecasting has become an essential factor, particularly for crops whose yields are affected by many variables such as weather, soil, and agronomic management factors. (Farokhzadeh et al., 2021; Behpouri et al. 2023).

Barley is also a suitable candidate to grow in unfavorable conditions because this crop can tolerate moderately dry conditions (Saed-Moucheshi et al., 2021; Czembor et al. 2022).
Previous studies revealed that higher yields of barley have been mainly achieved when the plants were able to have a faster canopy development and consequently earlier flowering and maturity stages (Lopez-Castaneda and Richards, 1994). Crop productivity is the result of complex interactions of various factors including weather indices, soil properties, topography, and management practices (Godwin and Miller, 2003).

Hlisnikovsky et al., (2024) in a long-term study, investigated the effects of several factors such as climatic conditions, soil fertilization, and different varieties on the yield of barley grain. Three notable correlations were identified with barley grain yield including (1) June precipitations, (2) minimal temperature in July, and (3) sunshine duration in May. Researchers have used different parameters and different methods to predict the yield of barley which may be appropriate based on the conditions of studies. One of the most significant factors is weather parameters (Wang et al., 2018; Zhu et al., 2019).

Furthermore, the other group of variables is soil properties such as soil pH, soil structure, organic matter, and soil nutrients (Kitchen et al., 2003; Cammarano et al., 2020). On the other hand, there are farm management factors as another group of variables that have crucial impacts on barley yield production. To the best of our knowledge, there is less information about the effect of farm parameters on barley yield. Ayoubi and Sahrawat (2011) used artificial neural networks (ANN) and multivariate regression analysis to assess the soil variables affecting barley yield in northern Iran. Results showed that total nitrogen, available phosphorous, soil electrical conductivity, sodium absorption ratio, pH, and organic matter consistently affected barley biomass and grain yield. Mokarram and Bijanzadeh (2016) studied the prediction of barley yield using multiple regression and artificial neural networks. They found that soil organic content, applied nitrogen, irrigation regime, and crop density are the most significant factors in models that predict the yield of barley. Identification of the most important factors that contribute most to yield can assist the farmers and governments to focus on more important variables and achieve higher yields. As mentioned above, researchers have identified the relationships of grain yield and various variables but to the best of our knowledge, less information is available about the effect of farm management practices on grain yield. The PCA is a statistical analysis that can classify the variables into a smaller number set of variables. Variables in each axis have a higher correlation together (Farokhzadeh et al., 2022).

Different linear and non-linear models have been identified between yield and other significant variables. These models can predict the yield based on the changes in independent variables. Partial least squares (PLSR) are another beneficial statistical modeling method that combines multiple linear regression and PCA to convert the data matrix and alleviate the collinearity issue of independent variables. This method has been used to predict grain yield in several agricultural studies. For instance, Shaibu and Adnan (2015) predicted the grain yield in maize using drought tolerance variables.

Zhang et al. (2020) used this method to determine the main factors affecting grain yield in wheat. Support vector machine (SVM) is another popular tool that was introduced by Vapnik et al. (1995). This method has also been used in data classification and support vector regression (SVR). The advantage of the SVR technique is the high flexibility in using variables and the control of penalty terms (Hu et al., 2018; Zhang et al., 2020).

In recent years, both PLSR and SVR have been used in agricultural research, particularly in drought prediction (Tian et al., 2018), yield prediction in pepper (Wilson et al., 2021), prediction of ber fruit mass (Abdel-Sattar et al. 2021) and prediction of bread wheat yield (Behpouri et al., 2023). So, this study focuses on the effect of manageable farm variables associated with the grain yield of barley. A comprehensive study of 15 key variables that contribute to barley yield was conducted in different parts of Fars Province. Data were collected from 104 different farms. The application of multivariate analysis including stepwise regression, machine learning, and principal component analysis (PCA) was used to ascertain the relationships between important manageable farm variables and grain yield. The novelty of this study compared to similar studies was that to the best of our knowledge, in the previous research regarding barley yield, the studied factors were a group of weather variables, soil variables, or a combination of them (Kitchen et al., 2003; Ayoubi and Sahrawat, 2011; Mokarram and Bijanzadeh, 2016; Czembor et al., 2022).

Yigit and Chmielewski (2024) in an investigation of winter and spring barley studied the effects of meteorological factors on grain yield. Additionally, they highlighted that farm management factors investigations will be needed in the future.

In the present study, we investigated a group of manageable farm variables to predict barley yield based on farm management factors.

2. Material and methods

2.1. Data collection

The data used in this study were collected from various regions of 104 farms in Fars Province, Iran, during 2022-2023. The sampling strategy was based on stratified sampling. The southern regions of Fars province were divided into 5 sub-regions and the data were collected proportionally. Grain yield (t ha^{-1}) was measured on every farm using a 1-m^2 quadrat based on random sampling with 3 replicates. Other variables include the application of animal manure, number of irrigation cycles, seed rate (kg ha^{-1}), use of herbicides (narrow leaf-herbicide and broad-leaf herbicide), time to plant maturity (month) and planting depth (cm), the nitrogen fertilizer (N, kg ha^{-1}), phosphorus fertilizer (P, kg ha^{-1}), potassium chloride fertilizer (K, kg ha^{-1}), were collected using a questionnaire on each farm. These variables were selected based on the literature and previous work of the authors (Mokarram and Bijanzadeh, 2016; Behpouri et al., 2023). The normal distribution of data was checked using the Kolmogorov-Smirnov test, the Shapiro-Wilk test, and the criteria of skewness and Kurtosis. To identify the relationships of variables, Pearson's correlation coefficients were calculated. Also, stepwise regression analysis was applied to recognize the most significant variables on grain yield. The multicollinearity test was conducted to measure TOL (tolerance) and VIF (variance inflation factor) using SPSS software. The VIF index was smaller than 10 and the TOL index was greater than 0.1 which indicated that there was no multicollinearity between variables. The PCA was performed to identify principal components and PCA bi-plot which contain variable combinations. PCA analysis and drawing of the Loading blot was performed using Minitab (18) software.

2.2. Machine learning methods

PLSR and SVR analysis as machine learning models were conducted using PYTHON (multi-paradigm programming language, v. 3.10.5) for prediction, SPSS (v.21) for PCA and correlation analysis and EXCEL software for statistical analyses and graphs drawing. To do this, 80% of the data was applied for the training stage, and the rest 20% of the data, for the testing stage. Mathematical background on PLSR and SVR is as follows:

SVR Method: The principles used in SVR are similar to those used in the SVM method. SVR is a modified form of SVM so that numerical data is used instead of categorized dependent variables.

PLSR Method: PLSR is a powerful, efficient regression method for multivariate analysis having a diverse range of data (Martens and Martens, 2000). This method decreases the number of independent variables to a smaller set of non-correlated elements for least square regression. This method is greatly useful particularly when the predictor variables have a high degree of collinearity.

2.3. Model performance

Four statistical indices were used to evaluate the model performance, including the coefficient of determination (R2) as Eq. 1, the root-mean-square error (RMSE) as Eq. 2, the mean squared error (MSE) as Eq. 3, and BIAS (Eq. 4) for the training and testing datasets. These statistical indices were calculated as:

$$R^2 = \frac{\left(n\sum_{i=1}^{n} X_{Pi} X_{mi} - \sum_{i=1}^{n} X_{Pi} \sum_{i=1}^{n} X_{mi}\right)^2}{\left[n\sum_{i=1}^{n} X_{pi}^2 - \left(\sum_{i=1}^{n} X_{Pi}\right)^2\right]\left[n\sum_{i=1}^{n} X_{mi}^2 - \left(\sum_{i=1}^{n} X_{mi}\right)^2\right]} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{pi} - X_{mi}\right)^2}{n}} \quad (2)$$

$$MSE = \frac{\sum_{i=1}^{n}\left(X_{pi} - X_{mi}\right)^2}{n} \quad (3)$$

$$BIAS = \frac{\sum_{i=1}^{n} X_{pi} - X_{mi}}{n} \quad (4)$$

Where n is the number of samples, Xmi is the measured dependent variable (yield value) in the field, and Xpi is the predicted yield value. A combination of these statistical criteria is sufficient for model evaluation. Khozani et al.

(2020) used these indices to examine the performance of models.

3. Results and discussion

To test the normality distribution of the data, both the Kolmogrove-Smirnove and Shapiro-Wilk tests were conducted. The results indicated that the variables are normally distributed. Descriptive statistics of the variables are presented in Table 1.

3.1. Correlation

The Pearson's correlation coefficient (rp) between 16 variables is listed in Table 2. The correlation coefficients ranged from 0.015 to 0.860. Grain yield indicated positive correlations with most of the studied variables except for pest damage, disease damage, number of weeds m^{-2}, seeding depth, and salinity level. The highest positive correlation coefficients for grain yield in this study were obtained between grain yield and irrigation (0.860**) and phosphorous fertilizer application (0.473**) respectively. Phosphorus is a fundamental plant nutrient that is associated with the construction of nucleic acids, phospholipids, adenosine triphosphates, and many coenzymes. Some studies have revealed that phosphorus and water have synergistic effects on plant growth. Optimization of water and phosphorus levels can effectively improve the absorption, transformation, and utilization of fertilizers by plants. Pertinent fertilization can reduce the negative effects of soil water deficiency on crop growth and development to a certain extent and can also increase the phosphorus concentration in plants (Gu et al., 2018). On the other hand, the highest negative correlation coefficients for grain yield in this study belonged to salinity level (-0.521**). Mokarram and Bijanzadeh (2016) in an investigation regarding factors affecting biological yield and grain yield in barley found that the highest positive correlation coefficients for biological yield obtained with 1000-kernel weight (r=0.700**), plant height (r=0.561**), nitrogen application (r=0.439**), irrigation regime (r=0.413**) and organic matter of the soil (r=401**). On the other hand, the number of the spike (m^{-2}) (r= -0.540**), water EC (r= -0.479**), and pH (r= -0.405**) had a negative correlation with biological yield. Also, they found that there was a positive and highly significant correlation between grain yield with 1000-kernel weight (r=0.635**), HI (r=0.622**), OC (r=0.544**), and spike/m2 (r=0.508**). However, water EC (r=- 0.535**), soil pH (r=-0.476**), phosphorous application (r=-0.324**), and potassium application (r=-0.178**) had a negative correlation with grain yield. Behpouri et al. (2022) in research studied the relationships of grain yield in bread wheat and 22 other agronomic variables. The results showed that there were high correlation coefficients between grain yield and a number of irrigation cycles and also farm manure application. Surprisingly, there was no significant correlation between grain yield and rainfall (precipitations). In the current study, the correlation coefficient between grain yield in barley and precipitations was positively significant. This is probably because when precipitation is not sufficient, wheat growers continue irrigation till the maturity of wheat plants in dry regions.

Table 1. Descriptive statistics of the variables.

Variable	Variable Symbol	Minimum	Maximum	Mean	Std. Deviation	Correlation coefficient of variables with yield
Yield (kg ha^{-1})	Y	1.80	5.80	3.77	0.83	1.00
Narrow-leaf herbicide (liter)	x1	0.00	2.00	1.34	0.49	0.240^{*}
Broad-leaf herbicide (liter)	X2	0.00	2.00	1.10	0.48	0.291^{**}
Maturity period (month)	X3	5.00	7.50	6.33	0.63	0.377^{**}
Irrigation (litre)	X4	20000.00	70000.00	37778.85	10636.95	0.860^{**}
Farm manure (kg ha^{-1})	X5	0.00	20000.00	682.70	2147.54	0.299^{**}
Pest damage (%)	X6	0.00	11.00	4.90	1.91	-0.107
Disease damage (%)	X7	0.00	13.00	3.76	3.20	-0.378^{**}
Number of weeds m^{-2}	X8	3.00	18.00	7.86	3.35	-0.218^{*}
Precipitations (mm)	X9	66.60	256.00	102.02	36.58	0.140
Seeding depth (mm)	X10	2.00	5.00	3.37	0.70	-0.249^{**}
Salinity Level (ds m^{-2})	X11	0.50	3.50	1.45	0.76	-0.521^{**}
Seeding rate (kg)	X12	220.00	350.00	283.90	30.74	0.142
Nitrogen fertilizer (n, kg ha^{-1})	X13	100.00	350.00	226.63	52.91	0.369^{**}
Phosphorous fertilizer (p, kg ha^{-1})	X14	0.00	150.00	77.79	45.78	0.473^{**}
Potassium fertilizer (k, kg ha^{-1})	X15	0.00	100.00	33.99	32.56	0.315^{**}

* and **: Significant (α= 5%), highly significant (α= 1%) respectively.

Table 2. Correlation matrix among 16 variables.

	Yield	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13N	X14P	X15K
Yield	1															
X1	.240*	1														
X2	.291**	.024	1													
X3	.377**	.173	.145	1												
X4	.860**	.213	.324	.450	1											
X5	.299**	.015	.071	.061	.246	1										
X6	-.107^{ns}	.084	-.036	-.037	-.083	-.029	1									
X7	-.378**	-.005	.031	-.139	-.375	-.133	.423	1								
X8	-.218*	.137	.039	-.327	-.237	-.020	.430	.441	1							
X9	.140^{ns}	.021	.046	-.063	.183	.525	-.021	-.099	-.112	1						
X10	-.249**	-.065	-.153	-.047	-.299	-.100	.134	.079	.094	-.050	1					
X11	-.521**	.041	-.186	.103	-.442	-.195	.195	.382	-.073	-.147	.199	1				
X12	.142^{ns}	-.009	.128	.270	.168	-.063	.197	.136	.155	-.237	.146	-.063	1			
X13N	.369**	.282	.187	-.073	.378	.193	-.163	-.236	.237	.150	-.247	-.522	.065	1		
X14P	.473**	.261	.272	.026	.396	.151	.008	.151	.231	.243	-.295	-.331	-.020	.520	1	
X15K	.315**	.316	.231	.009	.262	.081	.138	.179	.392	-.010	-.362	-.205	-.040	.314	.554	1

Pearson's correlation matrix among 15 variables: Y, yield (Kg ha^{-1}) x1, narrow-leaf herbicide; x2, broad-leaf herbicide; x3, Maturity period (months); x4, Number of Irrigation cycles; x5, Farm manure application (ton ha^{-1}) x6, pest damage; x7, Disease damage; x8, Number of total weeds (m^{-2}); X9, Precipitations (mm); x10, Seed depth (cm); x11, Soil Salinity (ds m^{-1}); X12, Seeding rate (kg ha^{-1}); x13, Nitrogen fertilizer (n; kg ha^{-1}); x14, P, Phosphorous fertilizer (p; kg ha^{-1}); x15, Potassium chloride fertilizer (k; kg ha^{-1}).

3.2. Stepwise linear regression

The results of stepwise regression analysis (Table 3) showed that irrigation (x4), salinity level (x11), Phosphorous fertilizer application (x14), and weeds infestation percentage (x8), justified the maximum grain yield in barley. Mokarram and Bijanzadeh (2016) also, reported that 1000-seed weight (g), OC (%), soil pH, grain/spike, HI (%), plant height (cm), irrigation regime (according to FC), and plant density (plant/m2), were the most significant factors in the prediction of grain yield in barley (R^2 = 0.78).

Table 3. Stepwise regression analysis of barley grain yields as dependent variables and other variables as independent variables.

Variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	1.96	0.260	-	7.522	0.000	-	-
Irrigation (x4)	5.42	0.000	0.696	11.833	0.000	0.621	1.610
Salinity Level (x11)	-0.181	0.058	-0.167	-3.130	0.002	0.759	1.317
Phosphorous (x14)	0.003	0.001	0.166	3.043	0.003	0.719	1.390
Weeds (x8)	-0.026	0.013	-0.103	-1.992	0.049	0.800	1.250

* and **: Significant (α= 5%), highly significant (α= 1%), respectively; B: unstandardized coefficients; R^2 = 0.79, and adjusted R^2 = 0.78.

3.3. Principal Component Analysis (PCA)

The PCA analysis of data in this research showed six major principal components (with eigenvalues more than one) showing 73.422 % of the total variance among the 104 barley samples. The first principal component (PC1) explained 24.719 % of the total variance which contains grain yield (Y), irrigation (x4), nitrogen fertilizer (x13), phosphorous fertilizer (x14), and potassium fertilizer (x15). All of these variables have the greatest contribution in the first component. In the second component (PC2), pest damage (x6), disease damage (x7), and weeds (x8) have the most contribution. In PC3, maturity period (x3) and seeding rate (x12) have the most contribution. In PC4, farm manure application (x5) and the number of precipitations (x9) are the most effective variables. In PC5, seed depth (x10) and seeding rate (x12) have the most positive contribution. Eventually, narrow-leaf herbicide application (x1) and broad-leaf herbicide application (x2) had the most contribution to variation explained by PC6. In addition, the loading plot of PCA (Fig. 1.) demonstrates the first two principal components for barley grain yield and other variables. Loading the biplot reveals the relationships of the variables. The cosine of the angles between the vectors reveals the extent of correlation between variables. Grain yield (Y), as the most significant variable in this study, has the acutest angles with x4 (irrigation). The other variables that have acute angles with grain yield include x5 (farm manure application) x9 (precipitations), x2 (broad-leaf herbicide application) x13 (nitrogen fertilizer application). On the other hand, the angles between grain yield vector and salinity level (x11), seeding depth (x10), disease damage (x7), pest damage (x8) and a number of weeds m^{-2} vectors are obtuse indicating that these variables have a negative correlation with grain yield.

Yield estimation model results based on the linear regression, SVR, and PLSR models. In this study, three analyzing methods including stepwise linear regression, SVR, and PLSR were used on farm management variables obtained from 104 farms in Fars Province, Iran. In the current study, the results of 3 methods were close to each other (Table 4). The highest R2 (0.79) belongs to the stepwise linear regression method and R^2 in PLSR and SVR is 0.73 and 0.76 respectively. Other calculated indices such as MSE, RMSE, and bias were also almost the same in this study. There are many reports comparing multivariate regression analysis, SVR, and PLSR (Hu et al., 2018; Tian et al., 2018; Zhang et al., 2020; Kamboj et al., 2022; Behpouri et al., 2023). In the previous study regarding the use of SVR, PLSR, and multivariate linear regression models in predicting grain yield in wheat we showed that PLSR had better prediction capability (R^2=0.85, RMSE=0.32, MSE=0.01, and bias=-0.05) (Behpouri et al., 2023). However, in this study, the difference between these 3 models was not significant and all of the indices in these 3 models (R2, RMSE, MSE, and Bias) are close to each other (Table 4).

Ayoubi et al. (2011) in an investigation regarding the relationships of barley yield and soil characteristics demonstrated that soil electrical conductivity, sodium absorption ratio, pH, total nitrogen, available phosphorus, and organic matter consistently affected barley biomass and grain yield. Overall, many researchers using different statistical analyses conducting grain yield in wheat and barley highlighted the significance of organic matter, nitrogen, and phosphorous fertilizer application and irrigation as the key elements in the prediction of barley and wheat grain yield and biomass (Ayoubi et al. 2011; Mokarram and Bijanzadeh, 2016; Tian et al., 2018; Zhang et al., 2020; Cammarano et al., 2020; Czembor et al., 2022; Behpouri et al., 2023).

Yigit and Chmielewski (2024) recently studied the effects of agrometeorological data on spring and winter barley during 2009-2022. They realized that air temperature adversely affects barley yield in both ear formation and anthesis phases. Moreover, they emphasized that some farm management factors such as adjusting the sowing date and soil moisture content strategies will be needed in future investigations to obtain higher yields in barley. In Table 5, the principal component analysis (PCA) of 15 variables affecting grain yield in barley is shown.

Table 4. Barley grain yield estimation results based on stepwise linear regression, SVR (support vector regression), and PLSR (partial least square regression).

Models	R^2	RMSE	MSE	Bias
Stepwise linear regression	0.79	0.39	0.152	0.00
PLSR	0.73	0.52	0.27	-0.01
SVR	0.76	0.45	0.20	0.00

R^2: coefficient of determination, RMSE: root-mean-square error, MSE: mean-square error.

Table 5. Principal component analysis (PCA) of 15 variables affecting grain yield in barley.

Variables	Components					
	PC1	PC2	PC3	PC4	PC5	PC6
Yield (kg ha^{-1})	0.856	-0.169	0.243	0.063	0.049	0.038
Narrow-leaf herbicide (liter)	0.322	0.279	0.178	0.076	-0.389	0.658
Broad-leaf herbicide (liter)	0.412	0.134	0.206	-0.011	0.061	-0.569
Maturity period (month)	0.285	-0.301	0.689	0.270	-0.204	0.028
Irrigation (liter)	0.833	-0.204	0.301	0.096	0.022	-0.008
Farm manure (kg ha^{-1})	0.393	-0.100	-0.340	0.624	0.192	-0.019
Pest damage (%)	-0.193	0.566	0.258	0.401	0.150	0.049
Disease damage (%)	-0.371	0.688	0.126	0.236	-0.113	-0.270
Number of weeds m^{-2}	-0.063	0.841	-0.102	-0.073	0.278	0.110
Precipitations (mm)	0.302	-0.112	-0.518	0.645	0.058	-0.018
Seeding depth (mm)	-0.451	-0.050	0.142	0.197	0.473	0.355
Salinity Level (ds m^{-2})	-0.644	0.051	0.240	0.321	-0.455	0.016
Seeding rate (kg)	0.060	0.150	0.650	-0.009	0.541	-0.019
Nitrogen fertilizer (n, kg ha^{-1})	0.662	0.222	-0.239	-0.277	0.207	0.251
Phosphorous fertilizer (p, kg ha^{-1})	0.665	0.453	-0.116	0.020	-0.131	-0.068
Potassium fertilizer (k, kg ha^{-1})	0.497	0.601	-0.006	-0.089	-0.309	-0.063
Eigenvalues	3.950	2.424	1.747	1.347	1.231	1.045
Proportional variance (%)	24.719	15.144	10.916	8.418	7.694	6.532
Cumulative variance (%)	24.719	39.863	50.799	59.196	66.890	73.422

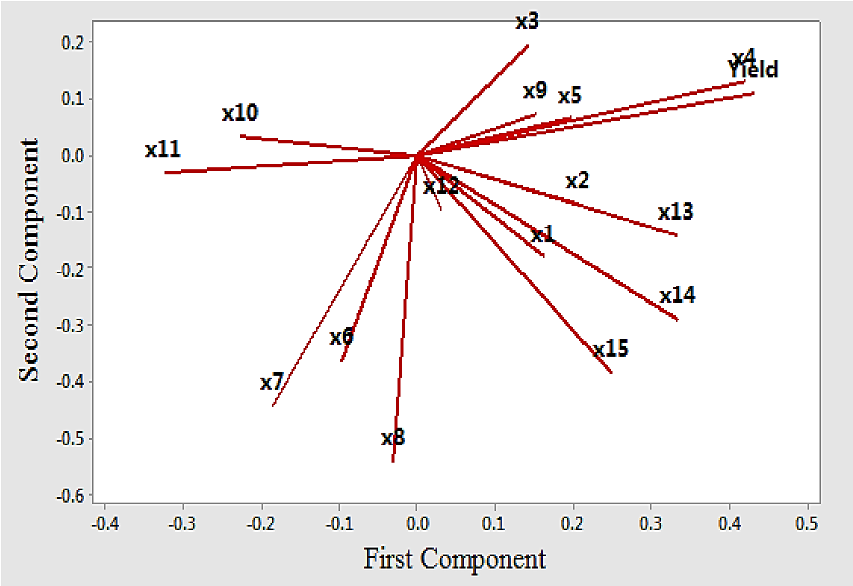

Fig. 1. The loading plot of the first two principal components for barley grain yield and other variables: Y, yield (Kg Y, yield (Kg ha^{-1}) x1, narrow-leaf herbicide; x2, broad-leaf herbicide; x3, Maturity period (months); x4, Number of Irrigation cycles; x5, Farm manure application (ton ha^{-1}) x6, pest damage; x7, Disease damage; x8, Number of total weeds (m^{-2}); X9, Precipitations (mm); x10, Seed depth (cm); x11, Soil Salinity (ds m^{-1}); X12, Seeding rate (kg ha^{-1}); x13, Nitrogen fertilizer (n; kg ha^{-1}); x14, P, Phosphorous fertilizer (p; kg ha^{-1}); x15, Potassium chloride fertilizer (k; kg ha^{-1}).

4. Conclusion

Many factors affect grain yield in barley particularly in dry regions when uncertainty about factors such as precipitation exists. Identification and prioritization of these variables and their relationships with grain yield can be essential in management programs to increase yield. In this study, we used multivariate statistical analyses including, correlation, stepwise linear regression, and principal component analysis (PCA) to select an effective subset of variables affecting grain yield. Additionally, SVR and PLSR methods were applied to evaluate the suitability of each method. The variables studied in this research are manageable including the use of different pesticides, weed control in farms, irrigation, and the use of various fertilizers. For instance, results showed that irrigation is one of the most significant factors in predicting grain yield, however, in the dry regions where precipitation and irrigation are limiting factors, other variables such as farm manure application, optimization of seeding depth, and application of N, P and K fertilizers are of great importance. The results of this research help the farmers to understand the significant factors that affect barley grain yield in the studied regions as well as in similar barley cultivating areas. For instance, farmers should take note that adjusting seeding rates, applying farm manure, controlling narrow and broad-leaf herbicides, and applying nitrogen and phosphorous fertilizers, are crucial for achieving higher barley yields. Furthermore, the results of this study demonstrate that stepwise linear regression is a convincing method to predict barley yield.

Acknowledgments

The authors are very thankful to the farmers of Fars Province for their collaboration in data collection, and to Abdollah Setoodeh, the technician of the Department of Agroecology for his help in typing the manuscript. Also, this research did not receive any specific grant from funding agencies.

References

Abdel-Sattar M., Aboukarima A.M., & Alnahdi B.M., 2021. Application of artificial neural network and support vector regression in predicting the mass of ber fruits (*Ziziphus mauritiana* Lamk.) based on fruit axial dimensions. *Plos One*. 16, e0245228.

Ayoubi, S. & Sahrawat, K.L., 2011. Comparing multivariate regression and artificial neural network to predict barley production from soil characteristics in northern Iran. *Archives of Agronomy and Soil Science*, 57, 549–565.

Cammarano, D., Holland, J. & Ronga, D., 2020. Spatial and Temporal Variability of Spring Barley Yield and Quality Quantified by Crop Simulation Model. *Agronomy*, 10, 393-398.

Czembor, E., Kaczmarek, Z., Pilarczyk, W., Mankowski, D. & Czembor, J.H., 2022. Simulating Spring

Barley Yield under Moderate Input Management System in Poland. *Agriculture*, *12*, 1091.

Farokhzadeh, S., Fakheri, B.A., Zinati, Z. & Tahmasebi, S., 2021. New selection strategies for determining the traits contributing to increased grain yield in wheat (*Triticum aestivum* L.) under aluminum stress. *Genet Resource Crop Evolution,* 68, 2061-2073.

Farokhzadeh, S., Shahsavand-Hassani, H., Zinati, Z. & Rajaei, M., 2022. Evaluation of triticale lines compared to wheat cultivars in terms of agronomic traits using supervised learning methods and multivariate statistics. Philippinian Agricultural Science, 105, 369-389.

Godwin, R.J. & Miller, P.C.H., 2003. A review of the technologies for mapping within-field variability. *Biosystem Engineering,* 84, 93–407.

Gu, Y.J., Han, C.L., Fan, J.W., Shi, X.P., Kong, M., Shi, X.Y., … Li, F.M., 2018. Alfalfa forage yield, soil water and P availability in response to plastic film mulch and P fertilization in a semiarid environment. *Field Crops Research,* 215, 94–103.

Hlisnikovsky, L., Zemanova, V., Roman, M., Mensik, L. & Kunzova, E., 2024. Long-Term Study of the Effects of Environment, Variety, and Fertilisation on Yield and Stability of Spring Barley Grain. *Plants*, 13(19), 2745.

Hu, Y., Wei, X., Hao, M., Fu, W., Zhao, J. & Wang, Z., 2018. Partial least squares regression for determining factors controlling winter wheat yield. *Agronomy Journal*, 110, 281-292.

Kamboj, U., Guha, P. & Mishra, S., 2022. Comparison of PLSR, MLR, SVM regression methods for determination of crude protein and carbohydrate content in stored wheat using near Infrared spectroscopy, Materials. Today: Proceedings, Volume 48, Part 3, 576-582.

Khozani, Z.S., Safari, M.J.S., Mehr, A.D. & Mohtar, W.H.M.W., 2020. An ensemble genetic programming approach to develop incipient sediment motion models in rectangular channels. *Journal of Hydrology*, 584, 124753.

Kitchen, N.R., Drummond, S.T., Lund, E.D., Sudduth, K.A. & Buchleiter, G.W., 2003. Soil Electrical Conductivity and Topography Related to Yield for Three Contrasting Soil–Crop Systems. *Agronomy Journal,* 95: 483-495.

Lopez-Castaneda, C. & Richards, R., 1994. Variation in temperate cereals in rainfed environments I. Grain yield, biomass and agronomic characteristics. *Field Crops Research*, 37, 51–62.

Mokarram, M. & Bijanzadeh, E., 2016. Prediction of biological and grain yield of barley using multiple regression and artificial neural network models. *Australian Journal of Crop Science, 10,* 895-903.

Saed-Moucheshi, A., Pessarakli, M., Mozafari, A.A., Sohrabi, F., Moradi, M. & Marvasti, F. B., 2021. Screening barley varieties tolerant to drought stress based on tolerant indices. *Journal of Plant Nutrition*, *45*, 739–750.

Shaibu, A.S. & Adnan, A.A., 2015. Predicting grain yield of maize using drought tolerance traits. *African Journal of Agricultural Research,* 10, 3332-3337.

Tian, Y., Xu, Y.P. & Wang, G., 2018. Agricultural drought prediction using climate indices based on support vector regression in Xiangjiang River basin. Science of the Total Environment, 622, 710-720.

Vapnik, V., 1995. The nature of statistical learning theory. Springer, NY.

Wang, J., Vanga, S. K., Saxena, R., Orsat, V. & Raghavan, V., 2018. Effect of Climate Change on the Yield of Cereal Crops: A Review. *Climate*, *6*, 41-48.

Wilson, A., Hemalatha, N. & Sukumar, R., 2021. Computational prediction model for pepper yield prediction using support vector regression. AgriRxiv Publication.

Yigit, A. & Chmielewski, F.-M., 2024. A Deeper Insight into the Yield Formation of Winter and Spring Barley in Relation to Weather and Climate Variability. Agronomy, 14(7), 1503.

Zhang, P.P., Zhou, X.X., Wang, Z.X., Mao, W., Li, W.X. &Yun, F., 2020. Using HJ-CCD image and PLS algorithm to estimate the yield of field-grown winter wheat. *Scientific Reports,* 10, 51-73.

Zhu, X., Troy, T.J. & Devineni, N., 2019. Stochastically modeling the projected impacts of climate change on rainfed and irrigated US crop yields. *Environmental Research Letters*, *14*, 74021.