### The best suitable areas for the cultivation of colza using multiple regression technique in GIS environment Case Study: Sabzevar Township, IRAN

Hamed adab<sup>1</sup>, Manuchehr Farajzadeh<sup>2</sup>, Parviz Zeaiean<sup>3</sup>, Esmail Filehkesh<sup>4</sup>, Reza Amiri<sup>5</sup>

 <sup>1</sup> Climatological Research Institute, Climate Change Group, Mashad, Iran adabgeo@gmail.com
 <sup>2</sup> Tarbiat Modares University, Department of Physical Geography, Tehran, Iran farajzam@modares.ac.ir

<sup>3</sup> Shahid Beheshti University, Department of remote sensing & GIS, Tehran, Iran

<sup>4</sup> Agricultural Researches & Natural Resources Sabzevar Station, Sabzevar, Iran

<sup>5</sup> Tarbiat Modares University, Tehran, Iran

# Abstract

Models were designed to simulate specific aspects of reality. We utilize the models to ease the complicated genuine problems. A model shows not only the whole fact of a phenomenon; but also, reveals the useful and comprehensible part of it. Regression technique is a statistical model which demonstrates the relation between variables.

In this research, collecting environmental data and statistical analysis are carried out to identify the yield of Colza farms by GPS set in the year 2004-2005 in Sabzevar region, in order to sample of this product. The result has been compared with ten environmental factors which are more effective on Colza growth. A multiple regression model has been utilized in SPSS software, and then by revealing the yield condition, backward model has been selected. Moreover, the aforementioned model has been tested with yield condition of selected farms to obtain the amount of error in the final model which concerning the actual performance of the selected farms.

The result of statistical experiments indicates backward method is up to 86% more capable than the Enter and Stepwise methods in estimation of yield.

### Keyword: Modeling, multiple regression, Agriculture, Brasica Napus

#### 1. Introduction

Oil-seeds compose the second world's largest food resources, after grains. These crops contain protein as well as rich resources of fatty acids. Colza has scientific name "Brassica napus" (in English named rapeseed and in French named colza) which is known as one of the most important oil-plants all over the world. By producing 1710 tons of colza in the year 2000, Iran had a portion of 0.04% of the whole world's production (Shariati & Ghazi, 2000). Considering the general suitable environmental conditions and the compatibility of colza with the climatic type of the Sabzevar region, many farmers and officials in Sabzevar have lately been attracted to colza cultivation. It should be taken into account that the development of colza cultivation in this region must be in accordance with the region's environmental capabilities, so a full advantage could be taken from the cultivation of colza in the given region and in achieving the desired success.

In order to develop the suitable areas for the cultivation of colza, the physical evaluation of lands is necessary. The problem of selecting the correct land for the cultivation of a certain agricultural product is a long-standing and mainly empirical issue (Kalogirou, 2002). Although many researchers, organizations, institutes and governments have tried to provide a framework for optimal agricultural land use, it is suspected that much agricultural land is used under its optimal capability. The increased need for food production and the shortage of resources stimulate a need for sophisticated methods of land evaluation to aid decision makers in their role to both preserve highly suitable lands and satisfy producers' demand for increased profits.

In general, the agricultural productivity of a geographic area is dependent on many factors including inherent soil and terrain characteristics, climatic constraints, human behavior and management. These factors are interdependent and constantly evolving in time and space. Unanticipated climatic events and human impacts may sometimes greatly affect them. Thus, decision-making in agricultural production is a challenging task. Therefore a desirable yield depends not only on the decrease of harmful agricultural agents such as hail, drought, frost and diseases, but also on the evaluations of area.

Consequently finding the relation between two or more variables could be done in different way such as

linear or non-linear relation such regression methods refers to a statistical modeling method which analyzes mutual relations. Making use of index, the relation between free and bound variables should be estimated or predicted. This is done by regression modeling method. The method which are specially designed to show the relative importance of free variable in regression equations are mostly enter, stepwise and backward methods. Prediction models for yield estimation of agricultural crops were first used for some selected countries in the south and southeast of Asia (Mavi, 1986). Using multiple regression models, Sawasawa (2003) did a survey in Nezam Abad region in Andhra Pradesh, India, in order to estimate the rice crop. Sabaramaya & Rupakumar (1980) used a model to predict the yield of sugar cane in Andhra Pradesh, India. Chang (1981) presented a model to estimate the yield of rice crops. His model was based on the effect of solar radiation parameters and temperature on rice crops in tropical regions. Price and Budgen(1990) presented an index which was strongly related to the crop's yield and could be used for prediction of the quantity of the desired crop. Iqbal et al. (2005) studied the relationship between topographical factors & hydrology on one hand and the function of cotton crop on the other, using GIS techniques and NDVI index. They presented Stepwise model in order to estimate the function of this crop. Leilah and Al-Khateeb (2005) studied wheat yield under drought conditions using statistical analysis in Saudi Arabia. In spite of the extensive research on some products like wheat, rice, corn, maize and so on, the study of colza was not considered. In Iran, Farajzadeh & Zarin (2001) have presented a model for the estimation of dry farming wheat yield in western Azerbaijan province of Iran, considering 11 climatic factors and using multiple regression method. Azizi & Yarahmadi (2002) representing a regression model, studied the relationship between climatic parameters and the increase or decrease of dry-farmed wheat's function in Silakhor area in Lorestan province, Iran. Again, the prediction of yield or recognition of suitable areas for colza cultivation was not studied. The main goal of this paper is the physical evaluation of lands based on the extraction of the most important variables influencing the colza vield based on linear multiple regression analysis. In order to perform the analysis, the selected variables of statistical analysis were imported into SPSS software.

# 2. Materials and methods

### 2.1. Study area

Sabzevar is located in Razavi Khorasan province, Iran extending from the east longitude of  $56^{\circ} 04'$  to  $58^{\circ} 15'$  and northern latitude of  $35^{\circ}30'$  to  $36^{\circ}58'$ , The average altitude of the area is 977 meters above sea level . (*Figure 1*)

# 2.2. Data used

To do this study, first sample farmlands were chosen based on the climatic map of Sabzevar Township. Colza farmlands are categorized in the 2 climatic types including semi-dry cold and arid cold climates. Then 10% samples were taken from the colza farmlands statistical society i.e. in the year 2004-2005, 24 out of 228 colza fields were chosen in Sabzevor Township as samples that regard the 2 mentioned climatic types, 12 samples have been taken from the society. Field have been divided into 5 categories regarding their green surface in a way that considering the frequency, some percentage is allocated to each category, and as the frequency of fields in each category grows, there will be more chosen fields. In the next step, the selected colza farmlands were positioned in a field - test by the GPS set. The required data about colza yield and other environmental factors for each field area was measured by Jahad Keshavarzi Organization of Sabzevar Township. The data items were: mean temperature (°C), degree-day growth (°C), mean absolute minimum temperature (°C), mean absolute maximum temperature (°C), mean temperature in sowing time, mean potential evapotranspiration (mm), slope (degree), EC of groundwater, PH of groundwater, mean relative humidity (%) and colza yield (kg/hectares). In this study, colza yield measures were considered as dependent variable and other 10 variables as independent in regression analysis. Table 1 indicates general features of the data used in the study area. The data about colza yield along with other related measures were analyzed by the following statistical procedures:

Simple models between colza crop's function and denary environmental factors to do this, the regression model of denary factors with 95% confidence and its index with the selected farmlands function was obtained which is represented in table 2.

Simple correlation: A matrix of simple correlation coefficients between colza yield and its components were computed in Table 3.

Calculation of multiple - variable models and selection of optimum model to do the modeling; the

multiple regression method has been used based on the data from 24 sampled fields. To do that, the gathered data from the selected colza fields from denary environmental were entered the statistical software environment (SPSS), then in order to equalize the data, standard Z mark was calculated and using 3 statistical modeling methods, stepwise, enter & backward (of this software), the given models were calculated.

1- Enter multiple linear regression: Multiple linear regression and partial coefficient of determination  $(R^2)$  was estimated for each yield component (Snedecor and Cochran, 1981) in order to evaluate the relative contribution and to develop the prediction model for yield (Y) according to the formula:

 $Y=a+b1X1+b2X2+b3X3+\ldots+bnXn.$ 

(1)

In the enter method all the variables entered the model.

2- Stepwise multiple linear regression: This procedure computes a sequence of multiple linear regression in a stepwise manner. One variable was added to the regression equation at each step. The added variables were the one, which induced to have the greatest reduction in the sum of, squares error. It was also the variable, which had the highest partial correlation with the dependent variable for fixed values of those variables already added. Moreover, it was the variable, which had the highest F value.

3- Backward multiple linear regression: This procedure computes elimination enters all of the variables in the block in a single step and then removes them one at a time based on removal criteria.

Appropriate statistical analysis was completed using SPSS (SPSS Inc, 2001) package.

# 3. Results and discussion

#### 3.1. Simple correlation analysis

Table 1 shows the minimum and maximum values, arithmetic mean and standard deviation for all the estimated and measured variables. The highest  $R^2$  index is seen in the mean temperature with (0.39) and the lowest  $R^2$  index is seen in the mean absolute minimum temperature with (0.02) table 2, In this table, (Y) represents the estimation of colza function and (X) is the given variable by putting which in the model the quantity of function is obtained under the influence of the given factor. Simple correlation coefficients between variables are presented in table 3. Values close to 1 indicate that the two elements are behaving almost identically. Conversely, a value close to -1 indicates that the two elements are behaving in opposite manner, i.e. when one element is increased the other decreased. A value near 0 indicates that the two elements are independent of each other. Results revealed that all variables in the study have significant negative correlation with colza yield, except relative humidity. Colza yield per hectares was negatively correlated with mean temperature (-0.62), mean absolute maximum temperature (-0.59), mean potential evapotranspiration (-0.44), slope (-0.49), EC of groundwater (-0.5), PH of groundwater (-0.52), relative humidity mean (0.43). Therefore a strong relationship exists between temperature values and colza yield.

### 3.2. Multiple linear regression analysis

Enter multiple linear regression analysis, the result of calculation of multiple linear regression in 3 different procedures are as follow:

Data presented in table 4 shows regression coefficient and probability of the estimated variables in predicting colza yield. The obtained results showed that the prediction equation for colza yield (Y) could be formulated using the colza plant variables as follows:

Y = 18664.39 - 69.87X1 + 3.87X2 + 512.97X3 - 768.14X4 - 510.12X5 - 11.8X6 + 758.22X7 - 0.15X8 + 69.67X9 - 35.63X10. (2)

The equation explains 98% of the total variation within the colza yield components, while the remaining 2% may be due to residual effects. The t-test showed that degree-day growth, mean absolute maximum temperature, slope and EC of groundwater have contributed significantly to colza yield, while the other six variables did not. The overall results reflect the importance of mentioned four variables in colza yield in the study area.

Stepwise multiple linear regression analysis, Table 5 shows the data representing partial and cumulative

 $R^2$  as well as the probability for the acceptability of limiting colza variables in yield prediction into two items. These variables are: mean temperature (38.5 %) and degree-day growth (64.1 %). According to the results, 64.1 % of the total variation in colza yield could be attributed to these two mentioned variables.

The other variables were not included in the analysis due to their relatively low contributions. Regression

coefficients for the accepted variables are shown in Table 6. The predicted equation for colza yield (Y) was: Y=-1156.91-628.12X1+ 3.38 X2 (3)

Backward multiple linear regression analysis, Table 7 shows the data representing partial and cumulative

 $R^2$  as well as the probability for the acceptability of limiting colza variables in yield prediction into four items. These variables are: degree-day growth, mean temperature, slope and EC of groundwater. According to the results, 83 % of the total variation in colza yield could be attributed to these four mentioned variables.

The other variables were not included in the analysis due to their relatively low contributions. Regression coefficients for the accepted variables are shown in Table 8. The predicted equation for colza yield (Y) was:

(4)

Y=15368.78+2.90X2-854.16X4+1061.58X7-0.172X8

# 3.3 selected multiple regression procedures

The result of applying different multiple linear regression was presented in table 9. As table 9 indicates, all independent variables were used in enter model, but in Stepwise model, the main variables in the model were only 2 including mean temperature and, degree-day growth. Conversely, in the backward procedure, variables were: degree-day growth, mean absolute maximum temperature, slope and EC of groundwater. A comparison of used procedures indicates that the enter procedure is very demanding in terms of the data required for the prediction of colza yield, even though some of them have not significant role in the calculated formula (table 4). In other side, in the stepwise procedure, there is a lack of sufficient predictor's variables that is limited into two variables resulting in only 68 % coverage of variation of data, but it seems that the backward procedure can be used as a suitable model to predict colza yield in the study area because it covers main independent variables. (Figure 2) and other methods (Figure 3,4) In agricultural environmental modeling the last step is chosen to be testing the final model. In the modeling process the researcher is after similarities and differences, i.e., how closely the selected colza farmlands represent the actual effective environmental conditions on the function of autumn colza in the region. Do we observe any change in index rate of predicted autumn Colza function by the selected model on one hand and the actual functions of the selected farmland on the other, just by charging the number of samples from colza selected farmland? After and analyzing the models in the research and choosing back ward model as the research's final model, the model were tested. To do so, by recalling the data of estimation and function of the selected colza farmlands 20% of the whole selected colza farmlands (5 farmlands) were chosen for testing and 80% (19 farmland) for training, and then the comparison is made to the selected colza farmlands. The results show that in spite of differences in the rate of indices in the 3 information layers, the rate of index has increased non the opposition – in training back ward and testing backward. (Table 10) but this fact that the crop's planting is irrigated is a vary significant factor in limiting this environmental resources' usage which are sometimes full of economic and agricultural benefits, therefore, the current access borders to water resource with more than 25 liters per second (well, aqueduct, fountain) were determined and also a onekilometer area was made which con irrigate the potential regions for planting colza. Then using the lager of the mode water resources boundary, a clip was made of the lager of the regions of estimated colza's function, so at the end the regions for planting autumn colza which have access to the water resources would be determined, then in the next stage considering the layers of the time of the first autumnal frost (75% probability), a clip was made of the final layer of potential regions for planting autumn colza, so finally the potential regions for planting colza, so finally the potential regions for planting colza in sabzevar township were determined. (Figure 5)

#### 4. Conclusion

In this study we found a significant relationship between colza yield and environmental variables which two of them were related to climatic conditions of the area including degree-day and mean temperature and the rest were related to topography that create a background to cultivate crop in plain areas. The forth variable is EC of groundwater that was used in order to irrigate colza farmlands. The results of this study indicate that in spite of the effects of different variables on colza yield, the selected variables have the most important effects. Also the backward linear multiple regression was proved to be a valuable tool to select the most important variables among multiple environmental conditions factors.

#### References

-Azizy Ghasem, Yarahmady Daryush, (2003), the study of climatic factors and wheat yield suing regression models, journal of geographical research, No. 44, P 23-32. In Persian.

-Farajzadeh manuchehr, Zarin Azer, 2001, *the modeling of drying wheat yield regarding to agroclimatoogy factors in western Azarbayjan province-* Iran, Modares of humaunities journal, No. 25, P 77-97. In Persian.

- Iqbal, J. Read, J. Thomasson, A. Jenkins, J,(2005), *Relationships between Soil-Landscape and Dryland Cotton Lint Yield*. ". Soil Scl. Soc.Am. Vol. 69. P12-26.

-Kalogirou S., (2002). *Experts systems and GIS: an application of land suitability evaluation*, Computer, Environment and Urban Systems, 26, P 89-112.

-Leilah A. A., Al-Khateeb S. A., (2005), *Statistical analysis of wheat yield under drought conditions*, Journal of Arid Environment, 61, P 483-496.

- Mavi, H. S., (1986). Introduction to Agrometeorology, Oxford & IBH Publishing Co.

Price-Budgen, Avril, l. (1990), Using meteorological Information and Products, Horwood.

-Shariyaty Shahab, ghazi shahnyzadeh Poran, (2000), Colza, publication of the statistical organization department of Jihad Keshavarzy ministry of Iran. In Persian.

- Sawasawa, H., (2003), Crop *Yield Estimation Integrating RS, GIS and Management Factors, A case study of Birkoor and Kortigiri Mandals* – Nizamabad District India. International Institute for Geo-Information Science and Earth Observation Enschde, the Nethrlanda.

-Senecore, G. W., Cochran, W. G., (1981), *Statistical Methods*, seventh ed., Iowa Stat University Press, Iowa, USA.

-SPSS Inc., (2001), SPSS 11.0 for Windows, USA, Inc. (http://www.spss.com).

Table 1. Basics statistics (minimum, maximum, mean and standard deviation (SD) for the 24-studied experimental field area

Variables	Minimum	Maximum	Mean	SD
Temperature mean (°C) (X1)	15.7	22.5	18.53	2.03
Growth degree day (°C) (X2)	3928	4723	4296.2	248.12
Mean of absolute minimum temperature (°C) (X3)	2.7	4.2	3.26	0.46
Mean of absolute maximum temperature (°C) (X4)	29.5	31.8	30.46	0.75
Temperature mean in sowing time (°C) (X5)	3.5	5.1	4.14	0.53
Potential evapotranspiration mean (mm) (X6)	624	730	677.16	30.35
Slope gradient (degree) (X7)	0	1.7	0.56	0.54
EC of groundwater (X8)	600	10254	3835	3318.17
PH of groundwater (X9)	7.6	8.9	8.21	0.45
Relative humidity mean (%)(X10)	38	56.9	46.61	6.34
Yield (kg/hectares/year) (Y)	464	4485	1760.5	953.28

Table 2: Simple regression models and R2 indices							
Variables	Number colza farmland with 95% confidence	R2	simple models				
mean absolute maximum temperature (°C)	12	0.35	<i>Y=</i> 24784.5 + (-754.5) <i>X</i>				
mean absolute minimum temperature (°C)	10	0.02	Y = 2702.45 + (-285.98)X				
mean temperature in sowing time (°C)	12	0.12	<i>Y=4338.74 + (-615.459) X</i>				
mean temperature (°C)	13	0.39	Y = 7174.92 + (-291.722)X				
mean potential evapotranspiration (mm)	13	0.2	Y = 11675 + (-14.6261)X				
mean relative humidity (%)	15	0.19	<i>Y</i> =- <i>13001</i> + <i>65.5791 X</i>				
degree-day growth (°C)	12	0.04	Y = 5393.3 + 862.496 X				
slope (degree)	8	0.24	<i>Y</i> =1274.55 +245.415 <i>X</i>				
PH of groundwater	13	0.27	Y = -7325.2 + 1106.23 X				

EC of groundwater

Table	Table 3. A matrix of simple correlation coefficients $(R)$ for the estimated variables of colza									
Variables	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Temperature mean (°C) (X1)										
Growth degree day (°C) (X2)	** 0.81									
Mean of absolute minimum temperature	** 0.58	0.76 **								
(°C) (X3)	0.56	0.70								
Mean of absolute maximum temperature	**	0.7	**							
(°C) (X4)	0.07	0.7	0.07							
Temperature mean in sowing time (°C)	**	**	**	**						
(X5)	0.77	0.05	0.91	0.02						
Potential evapotranspiration mean (mm)	**	**	0 72 **	**	**					
(X6)	0.01	0.09	0.72	0.02	0.04					
Slope gradient (degree) (X7)	-0.48 *	-0.33 <sup>NS</sup>	0.09 <sup>NS</sup>	-0.25 <sup>NS</sup>	-0.08 <sup>NS</sup>	-0.41 *				
EC of groundwater(X8)	0.54 **	0.49*	0.58 **	0.47 *	-0.65 **	0.54 **	0.06 <sup>NS</sup>			
PH of groundwater(X9)	-0.32 <sup>NS</sup>	-0.11 <sup>NS</sup>	-0.4 <sup>NS</sup>	-0.33 <sup>NS</sup>	-0.34 <sup>NS</sup>	-0.30 <sup>NS</sup>	0.08 <sup>NS</sup>	-0.47 *		
Relative humidity average (%)(X10)	-0.62 **	-0.55 **	-0.67	-0.69	-0.77	-0.69 **	0.05 <sup>NS</sup>	-0.65 **	0.38 <sup>NS</sup>	
Yield (kg/hectares/year) (Y)	-0.62 **	-0.21 NS	-0.15 <sup>NS</sup>	-0.58 **	-0.35 <sup>NS</sup>	-0.40 *	0.49 *	-0.50 *	** 0.53	0.43 *

 Table 4. The regression coefficient (b), standard error (SE), T- value and probability of the estimated variables in redacting colza yield by the enter multiple regression analysis

Variables	DF	Coefficients of regression (B)	Standard error (SE)	Т	Prob>   T
Temperature mean (°C) (X1)	1	-39.87	186.23	-0.37	0.71 <sup>NS</sup>
Growth degree day (°C) (X2)	1	3.87	1.57	2.46	0.02 *
Mean of absolute minimum temperature (°C) (X3)	1	512.97	810.80	0.63	0.53 <sup>NS</sup>
Mean of absolute maximum temperature (°C) (X4)	1	-768.14	379.93	-2.02	0.06*
Temperature mean in sowing time (°C) (X5)	1	-510.12	928	-0.55	0.59 <sup>NS</sup>
Potential evapotranspiration mean (mm) (X6)	1	-11.80	10.46	-1.12	0.28 <sup>NS</sup>
Slope gradient (degree) (X7)	1	758.22	309.23	2.45	0.02 *
EC of groundwater (X8)	1	-0.15	0.05	-3.03	0.009 **
PH of groundwater (X9)	1	169.67	345.54	0.49	0.63 <sup>NS</sup>
Relative humidity mean (%)(X10)	1	-35.63	31.64	-1.12	0.28 <sup>NS</sup>

\* And \*\* : means that r is significant at 5%, 1% level of probability. NS: Not significant.

Y-intercept (a) = 18664.39, SE = 12865.74,  $R^2 = 0.8670$ , Adj,  $R^2 = 0.7640$ 

 Table 5. Relative contribution (partial and model  $R^2$ ) in predicting colza yield, F-value and probability by the stepwise procedure analysis

Step	Variables entered	Partial R <sup>2</sup>	SE of estimates
1	Temperature mean (°C) (X1)	0.385	764.33
2	Growth degree day (°C) (X2)	0.641	597.95

 Table 6. Regression coefficient (b), standard error (SE), F value and probability (sig.) of the accepted variables that can be used to predict colza yield by the stepwise procedure

Variables	Coefficients of regression (B)	Standard error (SE)	t	Sig.
Temperature mean (°C) (X1)	-628.12	106.68	-5.88	0.000
Growth degree day (°C) (X2)	3.38	0.87	3.86	0.001 ***

\*\* Heans that r is significant at 5%, 1% level of probability. Y-intercept (a) = -1156.91, SE = 2429.91, R<sup>2</sup> = 0.6410,, Adj, R<sup>2</sup> = 0.6070

Table 7. Relative contribution (partial and model  $R^2$ ) in predicting colza yield, *F*-value and probability by the stepwise procedure analysis

Step	Variables removed	Partial R <sup>2</sup>	SE of estimates
1	-	0.867	463.20
2	(X1)	0.865	448.76
3	(X1)- (x9)	0.861	440.02
4	(X1)- (X9)-(X5)	0.858	431.40
5	(X1)- (X9)-(X5)-(X3)	0.855	421.85
6	(X1)- (X9)-(X5)-(X3)-(X10)	0.838	433.56
7	(X1)- (X9)-(X5)-(X3)-(X10)-(X6)	0.830	432.18

 Table 8. Regression coefficient (b), standard error (SE), F value and probability (sig.) of the accepted variables that can be used to predict colza yield by the stepwise procedure

Variables	Coefficients of	Standard error	t	Sig.
	regression (B)	(SE)		
Growth degree day (°C) (X2)	2.90	0.55	5.24	0.000 **
Mean of absolute maximum temperature (°C) (X4)	-854.16	-0.67	-4.97	** 0.000
Slope gradient (degree) (X7)	1061.58	0.60	5.79	** 0.000
EC of groundwater (X8)	-0.172	-0.59	-5.14	** 0.000

\*\* : Means that r is significant at 5%, 1% level of probability. Y-intercept (a) = 15368.78, SE = 4363.30,  $R^2 = 0.830$ , Adj,  $R^2 = 0.794$ 

Table 9. The results of applying different multiple linear regression

5 11 5 8 55		1	0		
Variables		1+	2	3	4
Temperature mean (°C) (X1)			١	1	
Growth degree day (°C) (X2)		$\checkmark$	١	1	
Mean of absolute minimum temperature (°C) (X3)					
Mean of absolute maximum temperature (°C) (X4)		$\checkmark$			
Temperature mean in sowing time (°C) (X5)		$\checkmark$			
Potential evapotranspiration mean (mm) (X6)		$\checkmark$			
Slope gradient (degree) (X7)	$\checkmark$	$\checkmark$			
EC of groundwater (X8)	$\checkmark$	$\checkmark$			
PH of groundwater (X9)	$\checkmark$	$\checkmark$			
Relative humidity average (%)(X10)		$\checkmark$			

 $1^+$  = simple correlation, 2 = enter multiple regression, 3 = stepwise multiple regression, 4 = backward multiple regression.

Table 10. Statistical analysis between selected Colza farmland'Testing'Training whit Main sample

Variables	R2	SD Kg/hectare	Minimum Kg/hectare	Maximum Kg/hectare	Mean Kg/hectare
Testing (Backward)	0.94	1052	292	3177	1618
Testing (yield farmland)	***	1008	676	3464	1840
Training (Backward) yield farmland Training	0.87 ***	815 967	632 464	4259 4485	1820 1738
Backward	0.86	876	292	4259	1778
Total selected yield farmland	***	954	464	4485	1759



Fig. 1. The location of the study area on the map of Iran associated with colza sampling sites



Fig. 2 – The zonation of suitability for colza cultivation based on backward method



Fig3 – The zonation of suitability for colza cultivation based on Stepwise method



Fig. 4 – The zonation of suitability for colza cultivation based on Enter method



Fig 5. The best suitable areas for the cultivation of colza