International Journal of Botany

ISSN: 1811-9700

science alert

ANSInet
an open access publisher
http://ansinet.com

International Journal of Botany 3 (4): 359-365, 2007
ISSN 1811-9700

The Preparation of the Colza (*Brassica napus*) Suitability Map Using Statistical Analysis and GIS; Case Study: Sabzevar Township, Iran

[1]Manuchehr Farajzadeh, [2]Hamed Adab and [3]Reza Amiri
[1]Department of Remote Sensing and GIS, Tarbiat Modares University, Tehran, Iran
[2]Department of Geography, Tarbiat Modares University, Tehran, Iran
[3]Department of Geography, Payame Noor University, Hashtrud, Iran

Abstract: In this study, 10 environmental data items including climatic and physical variables for colza cultivation collected from 24 sample farmlands in Sabzevar township within the period of agricultural year 2004-2005 and then required statistical analysis carried out for the selected farmlands in order to select the significant linear multiple regression procedures including enter, stepwise and backward in connection to colza yield. After selecting backward procedure as the best model, four environmental condition variables including degree-day, mean temperature, slope and EC of used groundwater extracted as necessary variables for zonation of suitable areas in the whole the study area. Then the spatial maps of the selected variables were prepared and the zonation map of suitable areas for cultivation of colza was prepared using GIS functions. The results show that the appropriate regions for cultivation of colza are located in the plain region in Sabzevar. Toward the border of the plain, suitability conditions of colza cultivation is degrading.

Key words: Colza, Sabzevar, linear multiple regression, Geographical Information System (GIS)

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 (Shahab and Poran, 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). FAO presented a procedure for evaluation of lands potential to cultivation of crops based on soil and environmental characteristics, by introducing five classes including highly suitable, moderately suitable, marginally suitable, currently not suitable and permanently not suitable (FAO, 1976, 1984, 1985). 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.Malczewski (2004) classified GIS-based land use suitability analysis into 3 main groups: 1) computer-assisted overlay mapping, 2) multicriteria decision making methods including multiobjective methods and multiattribute methods and 3) artificial intelligence methods including fuzzy logic techniques, neural networks, evolutionary (genetic) algorithms and cellular automata.

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

Corresponding Author: Manuchehr Farajzadeh, Department of Remote Sensing and GIS, Tarbiat Modares University, Tehran, Iran Tel: +98 912 172 3124 Fax: +98 21 88006544

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.

Prediction models for yield estimation of agricultural crops were first used for some selected countries in the south and southeast of Asia (Mavi, 1986). Avril (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. Using multiple regression models, Sawasawa (2003) did a survey in Nezam Abad region in Andhra Pradesh, India, in order to estimate the rice crop. Iqbal *et al*. (2005) studied the relationship between topographical factors and hydrology on one hand and the function of cotton crop on the other, using GIS techniques and MVI 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, Manuchehr and Azer (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. Ghasem and Daryush (2003) 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 study is the physical evaluation of lands based on the extraction of the most important variables influencing the colza yield based on linear multiple regression analysis. In order to perform the analysis, the selected variables of statistical analysis was imported into GIS and the suitability map of colza cultivation was produced.

MATERIALS AND METHODS

Sabzevar is located in Razavi Khorasan province, Iran extending from the east longitude of 56° 04' to 58° 15' and northern latitude of 35°30' to 36°58', with the area of 20502 km^2. The average altitude of the area is 977 m above sea level (Fig. 1).

Experimental field area: Twenty four experimental field areas selected randomly from 228 colza farmlands in different regions of the study area. Dimensions of each field area was determined one hectares and colza yield along with other climatic and agronomical conditions were measured and estimated separately for each plots in the period of agricultural year 2004-2005 (Fig. 1).

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),

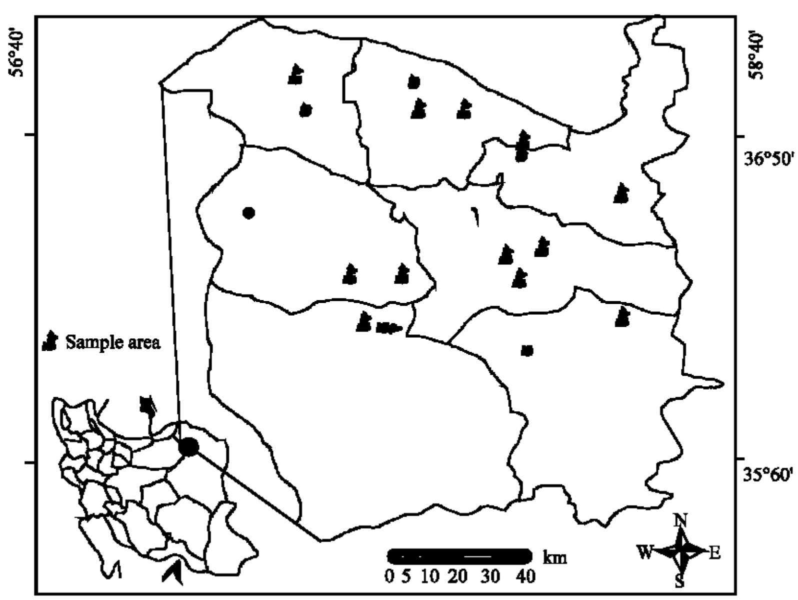

Fig. 1: The location of the study area on the map of Iran associated with colza sampling sites

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) (X_1)	15.7	22.5	18.53	2.03
Growth degree day (°C) (X_2)	3928.0	4723.0	4296.20	248.12
Mean of absolute minimum temperature (°C) (X_3)	2.7	4.2	3.26	0.46
Mean of absolute maximum temperature (°C) (X_4)	29.5	31.8	30.46	0.75
Temperature mean in sowing time (°C) (X_5)	3.5	5.1	4.14	0.53
Potential evapotranspiration mean (mm) (X_6)	624.0	730.0	677.16	30.35
Slope gradient (degree) (X_7)	0.0	1.7	0.56	0.54
EC of groundwater (X_8)	600.0	10254.0	3835.00	3318.17
pH of groundwater (X_9)	7.6	8.9	8.21	0.45
Relative humidity mean (%) (X_{10})	38.0	56.9	46.61	6.34
Yield (kg ha^{-1} $year^{-1}$) (Y)	464.0	4485.0	1760.50	953.28

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 ha^{-1}). In this study, colza yield measures were considered as dependent variable and other 10 variables as independent in regression analysis (Table 1).

Statistical techniques: The data about colza yield along with other related measures were analysed by the following statistical procedures:

Simple correlation: A matrix of simple correlation coefficients between colza yield and its components were computed (Senecore and Cochran, 1981).

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

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_nX_n \quad (1)$$

In the enter method all the variables entered the model.

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.

Backward multiple linear regressions: 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 (2001) package.

GIS analysis: In order to prepare zonation map of potential colza cultivation for the study area, after analysis of different statistical procedures, the backward method was used for zonation of suitable areas in GIS environment. Therefore considered data layers were prepared for the mentioned agricultural year according to the selected model. Then GIS database was used to recognize the suitable areas for colza cultivation.

IDRISI (Esteman, 1997) and ARCGIS (Anonymous, 2004) packages was used to access the required GIS analysis functions.

RESULTS AND DISCUSSION

Simple correlation analysis: Minimum and maximum values, arithmetic mean and standard deviation for all the estimated and measured variables. Simple correlation coefficients between variables are presented in this study. 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 (Table 2). 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.

Multiple linear regression analysis: The result of calculation of multiple linear regression in 3 different procedures are as follow:

Table 2: A matrix of simple correlation coefficients (r) for the estimated variables of colza

Variables	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
Temperature mean (°C) (X_1)										
Growth degree day (°C) (X_2)	0.81**									
Mean of absolute minimum temperature (°C) (X_3)	0.58**	0.76**								
Mean of absolute maximum temperature (°C) (X_4)	0.87**	0.7**	0.67**							
Temperature mean in sowing time (°C) (X_5)	0.79**	0.85**	0.91**	0.82**						
Potential evapotranspiration mean (mm) (X_6)	0.81**	0.89**	0.72**	0.62**	0.84**					
Slope gradient (degree) (X_7)	-0.48*	-0.33^{NS}	0.09^{NS}	-0.25^{NS}	-0.08^{NS}	-0.41*				
EC of groundwater (X_8)	0.54**	0.49*	0.58**	0.47*	-0.65**	0.54**	0.06^{NS}			
pH of groundwater (X_9)	-0.32^{NS}	-0.11^{NS}	-0.40^{NS}	-0.33^{NS}	-0.34^{NS}	-0.30^{NS}	0.08^{NS}	-0.47*		
Relative humidity average (%) (X_{10})	-0.62**	-0.55**	-0.67**	-0.69**	-0.77**	-0.69**	0.05**	-0.65**	0.38^{NS}	
Yield (kg ha^{-1} $year^{-1}$) (Y)	-0.62**	-0.21^{NS}	-0.15^{NS}	-0.58**	-0.35^{NS}	-0.40*	0.49*	-0.50*	0.53**	0.43*

* and **: Means that R is significant at 5, 1% level of probability. NS: Not Significant

Table 3: The regression coefficient (B), Standard Error (SE), t-value and probability of the estimated variables in predicting colza yield by the enter multiple regression analysis

Variables	df	Coefficients of regression (B)	Standard error (SE)	T	Prob>|T|
Temperature mean (°C) (X_1)	1	-39.87	186.23	-0.37	0.71^{NS}
Growth degree day (°C) (X_2)	1	3.87	1.57	2.46	0.02*
Mean of absolute minimum temperature (°C) (X_3)	1	512.97	810.80	0.63	0.53^{NS}
Mean of absolute maximum temperature (°C) (X_4)	1	-768.14	379.93	-2.02	0.06*
Temperature mean in sowing time (°C) (X_5)	1	-510.12	928.00	-0.55	0.59^{NS}
Potential evapotranspiration mean (mm) (X_6)	1	-11.80	10.46	-1.12	0.28^{NS}
Slope gradient (degree) (X_7)	1	758.22	309.23	2.45	0.02*
EC of groundwater (X_8)	1	-0.15	0.05	-3.03	0.009**
pH of groundwater (X_9)	1	169.67	345.54	0.49	0.63^{NS}
Relative humidity mean (%) (X_{10})	1	-35.63	31.64	-1.12	0.28^{NS}

* 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

Enter multiple linear regression analysis: The obtained results showed that the prediction equation for colza yield (Y) can be formulated using the colza plant variables as follows:

$$Y = 18664.39-69.87X_1 + 3.87X_2 + 512.97X_3-768.14X_4-510.12X_5-11.8X_6 + 758.22X_7-0.15X8 + 169.67X_9-35.63X_{10} \quad (2)$$

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 (Table 3).

Stepwise multiple linear regression analysis: 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 (Table 4).

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 5. The predicted equation for colza yield (Y) was:

Table 4: 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^2	SE of estimates
1	Temperature mean (°C) (X_1)	0.385	764.33
2	Growth degree day (°C) (X_2)	0.641	597.95

Table 5: 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) (X_1)	-628.12	106.68	-5.88	0.000**
Growth degree day (°C) (X_2)	3.38	0.87	3.86	0.001**

**: means that r is significant at 5, 1% level of probability. Y-intercept (a) = -1156.91, SE = 2429.91, R^2 = 0.6410, Adj, R^2 = 0.6070

$$Y = -1156.91-628.12X_1 + 3.38\ X_2 \quad (3)$$

Backward multiple linear regression analysis: Table 6 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 7. The predicted equation for colza yield (Y) was:

$$Y = 15368.78 + 2.90X_2-854.16X_4 + 1061.58X7-0.172X_8 \quad (4)$$

Table 6: Relative contribution (partial and model R^2) in predicting colza yield, F-value and probability by the backward procedure analysis

Step	Variables removed	Partial R^2	SE of estimates
1	-	0.867	463.20
2	(X_1)	0.865	448.76
3	(X_1)-(x_9)	0.861	440.02
4	(X_1)-(X_9)-(X_5)	0.858	431.40
5	(X_1)-(X_9)-(X_5)-(X_3)	0.855	421.85
6	(X_1)-(X_9)-(X_5)-(X_3)-(X_{10})	0.838	433.56
7	(X_1)-(X_9)-(X_5)-(X_3)-(X_{10})-(X_6)	0.830	432.18

Table 7: 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 backward procedure

Variables	Coefficients of regression (B)	Standard error (SE)	t	Sig.
Growth degree day (°C) (X_2)	2.900	0.55	5.24	0.000**
Mean of absolute maximum temperature (°C) (X_4)	-854.160	-0.67	-4.97	0.000**
Slope gradient (degree) (X_7)	1061.580	0.60	5.79	0.000**
EC of groundwater (X_8)	-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

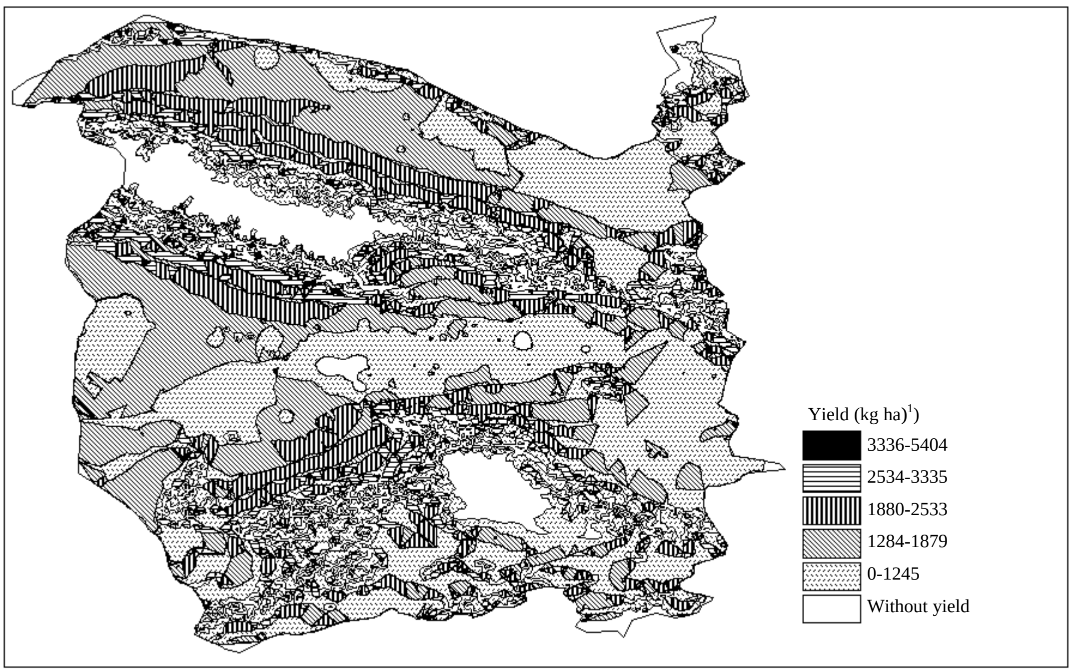

Fig. 2: The zonation of suitability for colza cultivation based on backward method

The zonation of the study area based on selected multiple regression procedures: 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 (Table 8). 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 3). In other side, in the stepwise procedure, there is a lack of sufficient predictor's variables which 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. Therefore the backward procedure was selected for zonation of colza cultivation in the study area. The distribution map of suitable areas for colza cultivation which was prepared based on linear multiple coefficients in GIS environment (Fig. 2). In order to accomplish the zonation, required data layers for applying backward

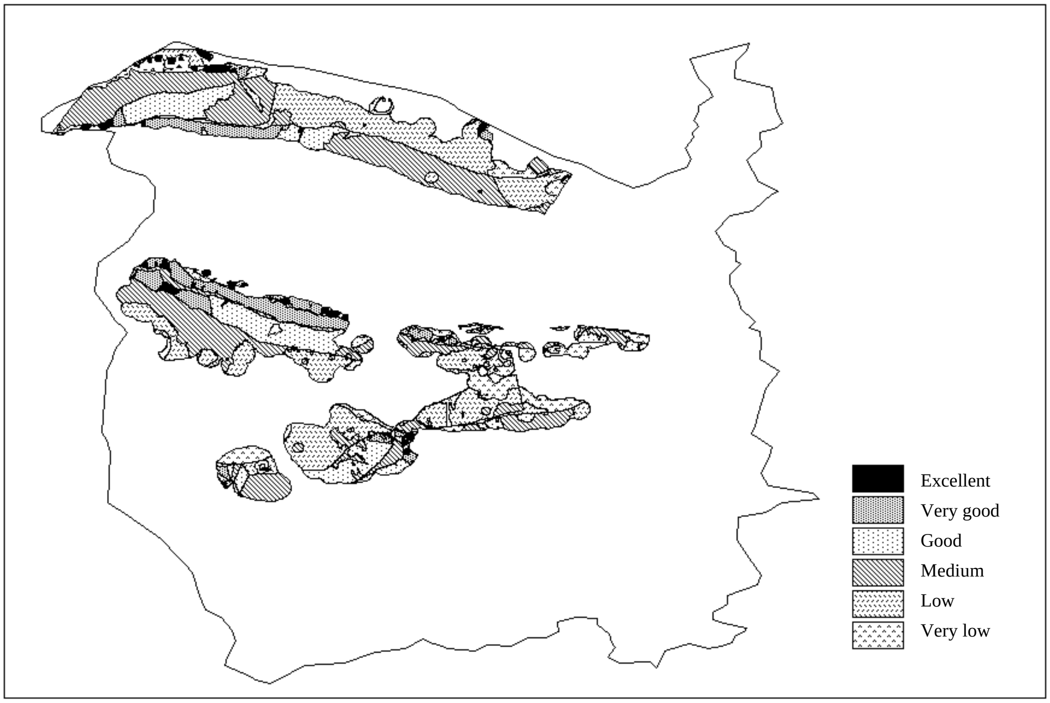

Fig. 3: The best suitable areas for the cultivation of colza

Table 8: The results of applying different multiple linear regression

Variables	1+	2	3	4
Temperature mean (°C) (X_1)	√	√	√	
Growth degree day (°C) (X_2)		√	√	√
Mean of absolute minimum temperature (°C) (X_3)		√		
Mean of absolute maximum temperature (°C) (X_4)	√	√		√
Temperature mean in sowing time (°C) (X_5)		√		
Potential evapotranspiration mean (mm) (X_6)	√	√		
Slope gradient (degree) (X_7)	√	√		√
EC of groundwater (X_8)	√	√		√
pH of groundwater (X_9)	√	√		
Relative humidity average (%) (X_{10})	√	√		

1^+ = Simple correlation, 2 = Enter multiple regression, 3 = Stepwise multiple regression, 4 = Backward multiple regression

Table 9: The suitable area for colza cultivation based on added economical value

Economical yield of colza (kg ha^{-1} year^{-1})	Descriptive	Profit against expenses (%)	Area (ha)	Area (%)
Y = 1443	Excellent	20	29175	8
1443-1869	Very good	60	100555	29
1869-2405	Good	100	116715	34
2405-2887	Medium	140	53275	15
2887-3908	Low	200	33442	10
3908-5403	Very low	300	6990	2

model were prepared based on Eq. 4 for the whole study area and then required calculations was performed in GIS environments.

Extraction of suitable areas: The zonation of colza yield in the study areas in Fig. 2 based on selected procedure present a map of suitable cultivation regions, but it is clear that the whole area can not be converted to colza cultivation. In order to prioritize the lands based on the increase of net profit value of more than 20%, the specific value of threshold was determined 1433 kg ha^{-1}. Based on this threshold and accessibility to agricultural water with more than 25 L sec^{-1} from surface currents, well and spring, the areas that has conversion capability to colza crop cultivation were determined. The areas in this map have been classified into 6 categories based on the threshold value. It means that the areas with the yield of more than about 3900 kg in hectares are the best areas for development colza agricultural lands in the area. The most suitable areas for colza cultivation are selected based on 34% of the area, 1869 to 2405 kg ha^{-1} and with a profit of 100% (considering the expense) (Fig. 2, 3 and Table 9).

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.

The suitability maps of colza cultivation in selected backward linear multiple regression in the study area indicates that the suitable areas for colza cultivation in the final map are mostly concentrated in Sabzevar, Davarzan and Jovin Jogtai plains. Therefore the spatial pattern for cultivation of colza follows the spatial pattern of the plains due to the effects of required environmental factors for cultivation of colza. On the other hand, among the investigated environmental factors, slope has intense effects on spatial distribution such that the increase of slope toward the margins of the plains resulted in a considerable reduction in the suitability conditions.

REFERENCES

Anonymous, 2004. ARCGIS 9, using ARCMAP. Environmental Systems Research Institute (ESRI) (http://www.esri.com).

Avril, P.B., 1990. Using Meteorological Information and Products. Horwood.

Esteman, J.R., 1997. IDRISI for windows user's guide, Version 2.0. Clark Laboratories for Cartographic Technologies and Geographic Analysis. Clark University, Worcester, MA.

FAO, 1976. A framework for land evaluation. Food and Agriculture Organization of the United Nations. Soils Bulletin No. 32, Rome: FAO.

FAO, 1984. Guidelines for land evaluation for rainfed agriculture. Food and Agriculture Organization of the United Nations. Soils Bulletin No. 52, Rome: FAO.

FAO, 1985. Guidelines for land evaluation for irrigated agriculture. Food and Agriculture Organization of the United Nations. Soils Bulletin No. 55, Rome: FAO.

Ghasem, A. and Y. Daryush, 2003. the study of climatic factors and wheat yield suing regression models. J. Geogr. Res. (In Persian), pp: 23-32.

Iqbal, J., J. Read, A. Thomasson and J. Jenkins, 2005. Relationships between soil-landscape and dryland cotton lint yield. Soil Sci. Soc. Am., 69: 12-26.

Kalogirou, S., 2002. Experts systems and GIS: An application of land suitability evaluation. Comput. Environ. Urban Syst., 26: 89-112.

Leilah, A.A. and S.A. Al-Khateeb, 2005. Statistical analysis of wheat yield under drought conditions. J. Arid Environ., 61: 483-496.

Malczewski, J., 2004. GIS-based land-use suitability analysis: A critical overview. Progr. Plann., 62: 3-65.

Manuchehr, F. and Z. Azer, 2001. The modeling of drying wheat yield regarding to agroclimatoogy factors in western Azerbaijan province-Iran. Modares Hum. J. (In Persian), 25: 77-97.

Mavi, H.S., 1986. Introduction to Agrometeorology. Oxford and IBH Publishing Co.

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 Netherlands.

Senecore, G.W. and W.G. Cochran, 1981. Statistical Methods. 7th Edn., Iowa Stat University Press, Iowa, USA.

Shahab, S. and G.S. Poran, 2000. Colza. Publication of the Statistical Organization Department of Jihad Keshavarzy Ministry of Iran (In Persian).

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