An assessment of drought impact on barley yield using a county wide drought Severity and Coverage Index and Adaptive Network-based Fuzzy Inference System (ANFIS) Model

Halis Simsek[1], Bilal Cemek[2], Navaratnam Leelaruban[3], G. Padmanabhan[3]*
[1]Department of Agricultural and Biosystems Engineering, North Dakota State University, Fargo, ND, USA
[2]College of Agriculture, Ondokuz Mayis University, Samsun, Turkey
[3]Department of Civil and Environmental Engineering, North Dakota State University, Fargo, ND, USA

Abstract

Agriculture is one of the major sectors often affected by droughts. A clear understanding of drought impact on agriculture will help address the future drought mitigation and management strategies. North Dakota (ND) is the leading barley producing state in the United States. In this study, the impact of drought severity and areal coverage on barley (*Hordeum vulgare* L.) yield in ND is estimated using an artificial intelligence technique, the Adaptive Network-based Fuzzy Inference System (ANFIS). A refined county level drought Severity and Coverage Index (I_{SC}) based on U.S Drought Monitor (USDM) drought severity and coverage data, and USDA National Agricultural Statistics Service (NASS) county level yield data are used in this study. The results show 41% and 12% of variation in yield can be explained by combination of drought conditions and I_{SC}, respectively.

Key words: Barley, ANFIS, regression, drought impact, yield, drought severity and coverage index

1. Introduction

Drought is a part of climatic variability and a natural hazard. Impact of drought on various sectors has long been recognized. Agriculture is one of the major sectors that experience significant loss during drought events. Agriculture also is the first sector to be affected at the onset of drought because crops at various stages of its growth depend on water and soil moisture [1]. Impact of drought on agriculture has been studied by several investigators [2, 3, 4]. Significant loss in yields of major crops may occur in the future due to drought. The loss in crop production across the United States during the last three decades is approximately $145 billion [2]. A better understanding of the drought-yield relationship could help reduce future losses. Crop yield variability is mainly influenced by local weather and climate rather than by large scale climatic patterns [5]. So, it is important that we study the drought-yield relationship at county scale. The State of North Dakota (ND) is a leading producer of many crops in the USA, particularly barley accounting for 24% of nation's production. Since North Dakota is a drought prone state, we chose to study the drought-barley yield relationship in North Dakota [6, 7].

Statistical tools such as Multiple Linear Regression (MLR) and Adaptive Network-based Fuzzy Inference System (ANFIS) are widely used in many disciplines for assesment, prediction, and classification purposes. MLR is a well known traditional statistical technique, and it has an

*Corresponding author: Address: Department of Civil and Environmental Engineering, North Dakota State University, Fargo, ND 58102, USA. E-mail address: G.Padmanabhan@ndsu.edu, Phone: +1 (701) 231-7043

established methodology. ANFIS is a an artificial intelligence technique that relies on the fuzzy logic of a system that transforms human knowledge or experience into a fuzzy inference system (FIS). Thus ANFIS can be imbedded with an ability to learn the expertise in any discipline and consequently it can adaptively improve its performance to simulate and forecast. It has been used to solve non-linear and high dimensional problems. The Fuzzy if-then rules are defined and calibrated using learning data sets and the predictability of measured data is validated using validation data sets. ANFIS has been broadly used in many fields including agriculture, hydrology, water quality [8, 9], and drought [10, 11, 12].

The objective of this study is to assess the impact of drought conditions on barley (Hordeum vulgare L.) yield using the MLR and ANFIS models. Different drought conditions and a refined countywide drought Severity and Coverage Index (I_{SC}) are used in this study to investigate the impact.

2. Study Area, Data, and Methods

2.1. Study area

North Dakota State is one of the north-central states of U.S. comprising of 53 counties (see Figure 2). North Dakota has a North-South temperature gradient, and Southeast-Northwest precipitation gradient. Droughts [6] have significantly impacted North Dakota in the past. A severe drought occurred recently in years 2006, 2008, and 2102. Frequent droughts are a characteristic feature of North Dakota climate [6]. The region is well known as the bread basket of the world because of its large-scale agricultural production.

2.2. Data

2.2.1. Drought Data

This study uses data from the U.S Drought Monitor (USDM) , a major source of drought data in the USA available to the public from the National Drought Mitigation Center (NDMC), University of Nebraska, Lincoln. The USDM is developed as a comprehensive tool to reflect the existing drought condition across the United States [13]. Several federal agencies including U.S. Department of Agriculture (USDA), and National Oceanic and Atmospheric Administration (NOAA), and NDMC contribute to produce drought monitor data products. The USDM releases its products (map and tabular data) every week which reflect the drought condition of the U.S (Figure 1). In this study, the USDM countywide weekly percent area coverage values were used as input for different drought intensity categories for the years 2000 to 2012. Drought intensities are categorized as follows; abnormally dry (D0), moderate drought (D1), severe drought (D2), extreme drought (D3), and exceptional drought (D4). Detailed description of the drought monitor data can be found in http://droughtmonitor.unl.edu/.

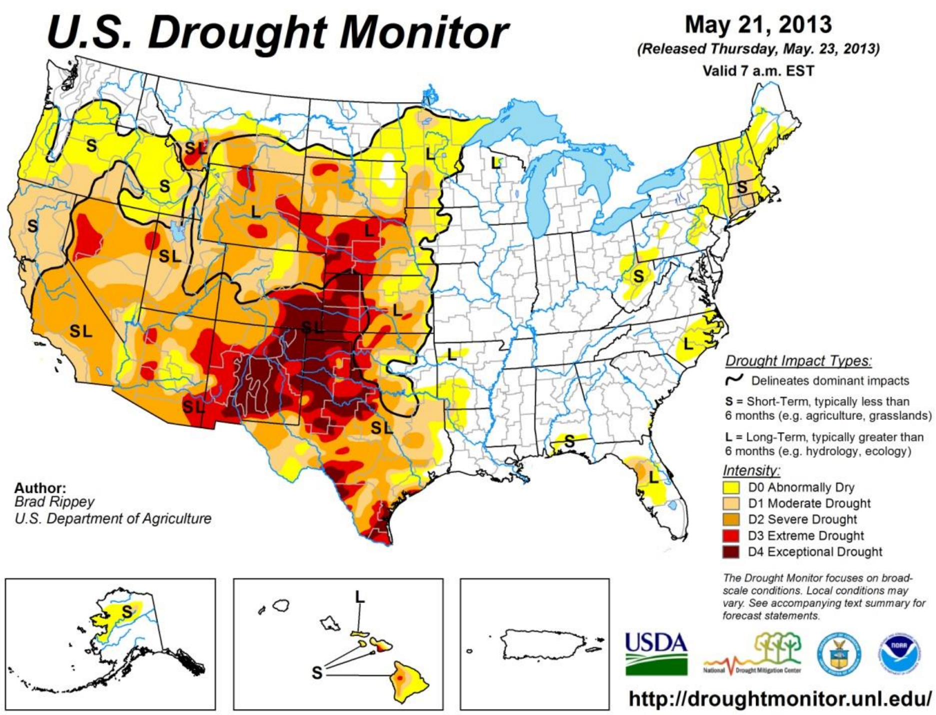

Figure 1. A Sample USDM weekly map

2.2.2 Crop Data

Barley is one of the major agricultural crops grown in North Dakota. County-by-county yield data of barley is derived from USDA National Agricultural Statistics Service (NASS) web portal for the study period (2000 – 2012). Generally, Barley planting starts in later part of April, and harvesting ends in early part of September in North Dakota. Figure 2 shows the North Dakota counties and barley yield in 2010.

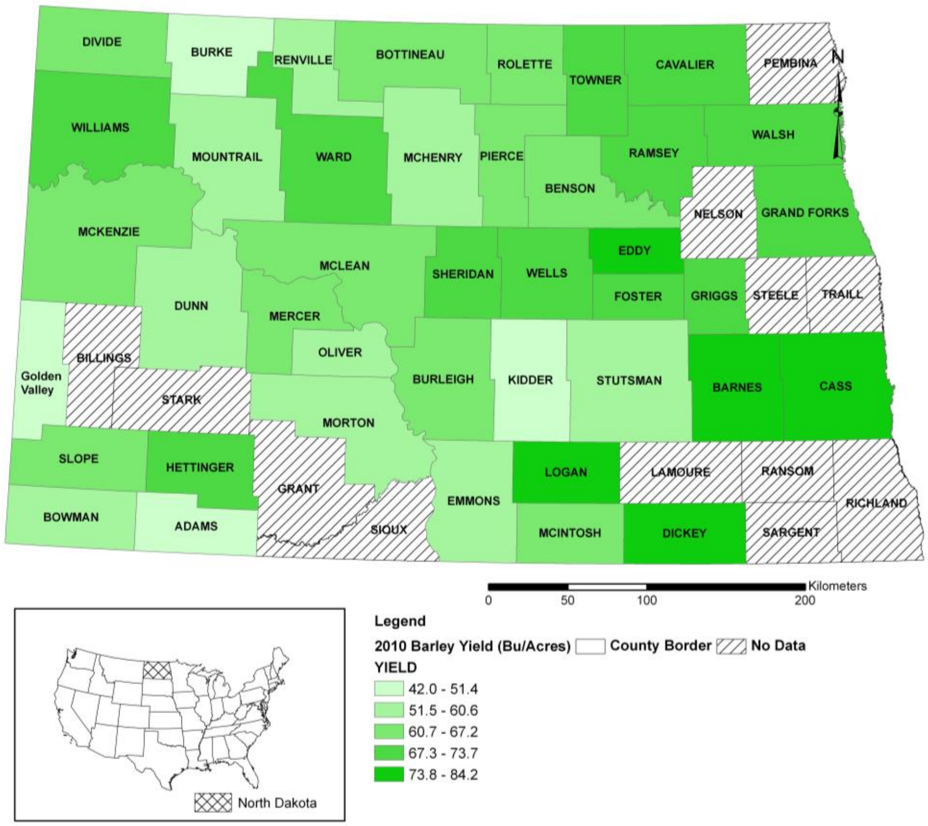

Figure 2. The North Dakota counties and barley yield in Bushel/Acres (1 Bushel = 0.03524 m^3; 1 Acre = 4046.86 m^2) for year 2010 (barley yield data is derived from USDA NASS web portal).

2.3. Drought Severity-Coverage Indices (I_{SC})

I_{SC} ($I_{Severity,Coverage}$) values were calculated from weekly percentage of areal coverage values of drought intensity. In this simple approach, areal coverage variables of higher severity conditions are assigned higher multiplying factors as shown in the equation below:

$$I_{SC} = 1\times(A_{D0}) + 2\times(A_{D1}) + 3\times(A_{D2}) + 4\times(A_{D3}) + 5\times(A_{D4}) \quad (1)$$

where A_{D0}, A_{D1} , A_{D2} , A_{D3} , and A_{D4} are percentage area coverage values for D0, D1, D2, D3, and D4 respectively. From Eq. (1), a numeric value of 500 can be regarded as the worst possible drought scenario implying that 100% of the county would be deemed under exceptional drought. A value of zero would therefore imply that 0% of the county is facing drought. A detailed description and application of the Drought Severity-Coverage Indices (I_{SC}) can be found in Leelaruban, et al [7].

2.4. Regression approach

A set of regression models were developed to ascertain the dependency of yield on drought conditions (model 2-8). The regression model parameters were estimated, and usefulness of each model in barley yield prediction was tested for significant level of 0.05. Average values of A_{D0}, A_{D1}, A_{D2}, and A_{D3} were calculated between planting and harvesting period from collected data for different drought intensity categories of areal coverage values, where A_{D0}, A_{D1}, A_{D2}, and A_{D3} are percentage area coverage values for D0, D1, D2, and D3 respectively. Then panel data set was constructed using barley yield, Avg(A_{D0}), Avg(A_{D1}), Avg(A_{D2}), Avg(A_{D3}), and Avg(I_{SC}). For i=1, 2, … 53 counties and t=1, 2, … 13 years (2000-2012) of observation.

$$Yield_{it} = \beta_{0x} + \beta_{1x} \times Avg(A_{D0})_{it} + \varepsilon \quad (2)$$

$$Yield_{it} = \beta_{0x} + \beta_{1x} \times Avg(A_{D1})_{it} + \varepsilon \quad (3)$$

$$Yield_{it} = \beta_{0x} + \beta_{1x} \times Avg(A_{D2})_{it} + \varepsilon \quad (4)$$

$$Yield_{it} = \beta_{0x} + \beta_{1x} \times Avg(A_{D3})_{it} + \varepsilon \quad (5)$$

$$Yield_{it} = \beta_{0x} + \beta_{1x} \times Avg(A_{D0})_{it} + \beta_{2x} \times Avg(A_{D1})_{it} + \beta_{3x} \times Avg(A_{D2})_{it} + \beta_{4x} \times Avg(A_{D3})_{it} + \varepsilon \quad (6)$$

$$Yield_{it} = \beta_{0x} + \beta_{1x} \times Avg(I_{SC})_{it} + \varepsilon \quad (7)$$

D4 category drought condition is not considered in this study because North Dakota experienced D4 category drought rarely during the study period. Regression model 2-5 were considered to investigate the capability of explaining the influence of the different drought condition individually on barley yield while model 6 was considered to investigate the effect of the combination of drought conditions on barley yield. Model 7 is used to assess the impact of drought on barley yield using the composite Drought Severity-Coverage Index(I_{SC}).

2.5. Adaptive Neuro-Fuzzy Interference System (ANFIS)

ANFIS algorithm serves as a basis to build a fuzzy model of a system by constructing a set of fuzzy "if-then" rules with sufficient Membership Functions (MF). An ANFIS model for two-input parameters (two fuzzy "if-then" rules) is briefly described as follows [14]

R_1: If x is A_1 and y is B_1, then $f_1 = p_1x + q_1y + r_1$
R_2: If x is A_2 and y is B_2, then $f_2 = p_2x + q_2y + r_2$

R_1 and R_2 denote the fuzzy rules 1 and 2. A and B are the fuzzy sets in fuzzy rule 1 and 2, respectively. The x and y are two different input parameters and the p, q, and r are the fuzzy consequent parameters [14, 15].

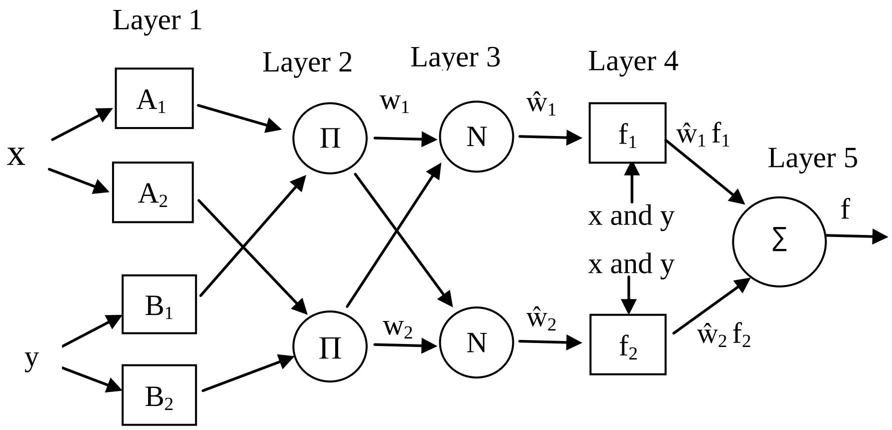

Figure 3. A typical architecture of two-input one-output ANFIS model [14].

Selection of input variables is crucial to develop a satisfactory estimation of ANFIS model since those variables determine the structure of the model corresponding weight coefficient. Different architecture of the ANFIS model was constructed for each phase of the model using all the data. About 70% of the data were selected randomly for training procedure while remaining data used for testing procedure. The performance of ANFIS model for training and testing data sets were evaluated for individual drought condition, which are Avg(A_{D0}), Avg(A_{D1}), Avg(A_{D2}) and Avg(A_{D3}) ,and combination of these four drought conditions to assess barley yield. Furthermore, AvgI_{SC} was used as one input parameter and barley yield is estimated. Three common statistical indicators of accuracy of estimation, which are root mean square error (RMSE), mean absolute error (MAE), mean bias error (MBE) are used along with coefficient of determination (R^2) as comparing criteria for the evaluation of the models' performances.

3. Results and Discussion

3.1. Influence of individual drought condition on barley yield estimation

Table 1 shows the estimated parameters for the developed regression model (model 2-5). The p-values suggest that the models are useful, and estimated parameters explain that drought has a negative influence on the barley yield. However, very low coefficients of determination (R^2) are obtained when the individual drought condition is used alone in the model.

Table 1. Estimated parameters for the regression model 2-5

MODEL	Constant	Yield predictors				P- value	R^2 (%)
		Avg(A_{D0})	Avg(A_{D1})	Avg(A_{D2})	Avg(A_{D3})		
2	56.0	-0.0627				0.017	0.01
3	56.8		-0.2060			0.000	0.07
4	56.3			-0.293		0.000	0.09
5	55.2				-0.354	0.000	0.05

Drought effect on barley yield was estimated using ANFIS model for individual drought conditions (Table 2). Table 2 shows the used MF and number of MF, and estimated R^2, RMSE, MBE, and MAE for both training and testing data. For every individual input, Trim, Tram and Gauss MFs were used with 3 rules for each MF. These rules were determined as low, medium and high drought conditions. The best ANFIS architecture was selected and evaluated among various architectures with different MFs in Table 2. As an individual drought condition, D3 drought condition provided the highest R^2 value (0.12) on training data set using trim MF while R^2 value for testing data sets is calculated as 0.09. RMSE, MBE and MAE values are low in this data set (ANFIS 10). The RMSE statistic estimate a value away from the mean value and it is higher or equal to MAE.

Table 2. Performance of ANFIS in estimating yield using individual drought conditions

Model	Input	MF	Number of MF	Training				Testing			
				R^2	RMSE	MBE	MAE	R^2	RMSE	MBE	MAE
ANFİS1	$Avg(A_{D0})$	Trim	3	0.03	18.58	8.78	15.35	0.08	19.47	9.29	16.00
ANFİS2		Tramp		0.02	25.63	13.77	20.51	0.06	25.85	13.21	20.04
ANFİS3		Gauss		0.02	29.40	16.13	22.95	0.06	29.35	15.26	22.16
ANFİS4	$Avg(A_{D1})$	Trim	3	0.05	12.02	0.00	9.27	0.14	11.97	2.51	9.44
ANFİS5		Tramp		0.06	11.97	0.00	9.25	0.08	12.33	2.62	9.83
ANFİS6		Gauss		0.05	11.98	0.00	9.26	0.09	12.24	2.60	9.76
ANFİS7	$Avg(A_{D2})$	Trim	3	0.05	12.02	-0.87	9.13	0.08	12.33	2.62	9.83
ANFİS8		Tramp		0.08	11.82	0.13	9.08	0.17	11.83	2.86	9.52
ANFİS9		Gauss		0.08	11.81	0.13	9.07	0.17	11.82	2.86	9.49
ANFİS10	$Avg(A_{D3})$	Trim	3	0.12	11.58	0.13	8.79	0.09	12.34	2.73	9.78
ANFİS11		Tramp		0.11	11.61	0.13	8.79	0.10	12.24	2.71	9.71
ANFİS12		Gauss		0.11	11.59	0.13	8.78	0.10	12.27	2.74	9.73

The estimated coefficient of determination (R^2) for the yield prediction using both regression and ANFIS model are low. That implies that only few percentage of variation of barley yield can be explained by an individual drought condition.

3.2. Influence of multiple drought conditions on barley yield estimation

Multiple Linear Regression model (model 6) can explain the influence of drought conditions on the variability of barley yield in North Dakota. In the multiple linear regression analysis, the yield of barley is used as the dependent variable and drought conditions are used as the independent variables (i.e., $Avg(A_{D0})$, $Avg(A_{D1})$, $Avg(A_{D2})$, and $Avg(A_{D3})$). The regression parameters were estimated for the model 6.

The obtained regression equation is;

$$Yield = (58.6) - 0.0699 \times Avg(A_{D0}) - 0.0976 \times Avg(A_{D1}) - 0.192 \times Avg(A_{D2}) - 0.249 \times Avg(A_{D3}) \quad (8)$$

The significance level for F statistic (p-value) for the model is zero, and that implies there is strong evidence that at least one of the model coefficient is nonzero, and overall model is useful to predict yield. Multiple coefficient of determination (R^2) was estimated as 0.133. The predicted results (R^2) suggest that 13.3% variation of barley yield can be explained by different drought conditions. In addition, model 8 suggests that influence of the drought is increasing with increasing drought severity. The model predictors for model 8 are tested for multicollinearity, and there is no serious multicollinearity effect found.

Drought assessment on barley yield using combination of drought conditions was determined using ANFIS and presented in Table 3. For four input parameters, Trim, Tram and Gauss models were used with 81 rules for each MF. These rules were determined as low, medium and high drought conditions. The best ANFIS architecture was chosen from the Table 3. The correlation between measured versus estimated yield was high in combination of drought conditions compared to using individual drought condition. The R^2 values were calculated as 0.41 and 0.52 for training and testing data sets, respectively for Gaussian MF (ANFIS3). The figure 3 shows the relationship between actual and estimated yield.

Table 3. Performance of ANFIS in estimating yield using combination of drought conditions

Model	Input	MF	Number of MF	Training				Testing			
				R^2	RMSE	MBE	MAE	R^2	RMSE	MBE	MAE
ANFİS1	Avg(A_{D0})	Trim	3*3*3*3	0.39	9.64	-0.80	7.09	0.50	9.11	2.14	6.47
ANFİS2	Avg(A_{D1}) Avg(A_{D2})	Tram		0.35	9.94	-0.77	7.48	0.45	9.60	2.07	7.16
ANFİS3	Avg(A_{D3})	Gauss		0.41	9.47	-0.83	6.93	0.52	8.97	2.23	6.35

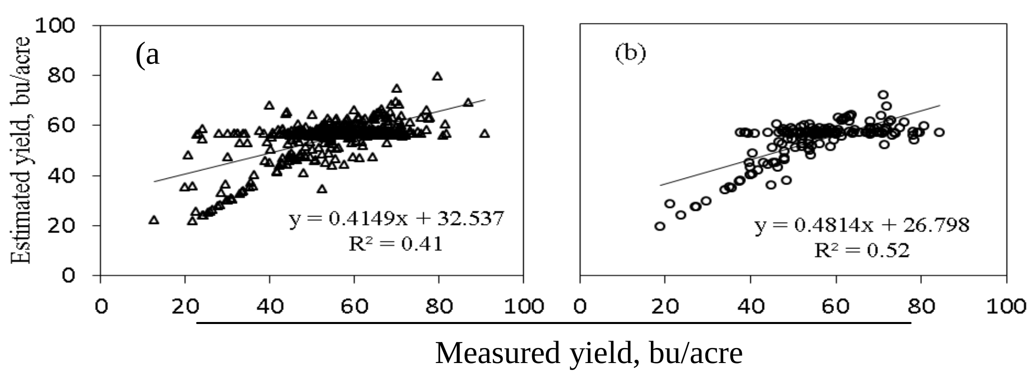

Figure 4. The measured versus estimated barley yield (bu/acre) using different drought conditions (i.e., Avg(A_{D0}), Avg(A_{D1}), Avg(A_{D2}), and (A_{D3})). (a) training data set (b) testing data set.

3.3. Influence of I_{SC} on barley yield estimation

The regression(model 7) and ANFIS models were used to assess the influence of I_{SC} on barley yield estimation. Based on the estimated parameters regression equation can be written as;

$$Yield = 58.5 - 0.0595\ AvgIsc \qquad (9)$$

The significance level for F statistic (p-value) for the model is zero, and that implies model is useful. Coefficient of determination (R^2) was estimated as 0.13, which is very similar to model 8. The model (9) suggests that with increasing I_{SC} barley yield will decrease as expected. Drought assessment on barley yield was modeled in ANFIS using I_{SC} and presented in Table 4. In this study three different architectures of ANFIS with 3, 4 and 5 MFs were used. The best ANFIS architecture was observed on ANFIS8 model in Table 4. The coefficient of determination in ANFIS8 was estimated 0.12 for training and 0.32 for testing data sets. Results from Table 4 expressed that by increasing the MF number from 3 through 5, the value of statistical parameters did not change in training data while slightly increased in testing data. In addition, a R^2 value of 0.64 was obtained by using years along with I_{SC} in ANFIS model. However, further investigation is needed to explain such a high influence of time (year) in predicting barley yield.

Table 4. Performance of ANFIS in estimating yield using Drought Severity and Coverage Index (I_{SC})

Model	Input	MF	Number of MF	Training				Testing			
				R^2	RMSE	MBE	MAE	R^2	RMSE	MBE	MAE
ANFİS1		Trim		0.12	11.58	0.87	9.01	0.27	2.33	8.82	10.99
ANFİS2	I_{sc}	Tramp	3	0.12	11.63	0.90	9.04	0.28	2.39	8.84	10.95
ANFİS3		Gauss		0.12	11.62	0.89	9.03	0.28	2.37	8.81	10.94
ANFİS4		Trim		0.12	11.61	0.89	9.02	0.29	2.37	8.77	10.91
ANFİS5	I_{sc}	Tramp	4	0.12	11.60	0.89	9.02	0.31	2.37	8.74	10.77
ANFİS6		Gauss		0.12	11.64	0.90	9.05	0.31	2.40	8.76	10.80
ANFİS7		Trim		0.12	11.59	0.84	9.02	0.31	2.24	8.61	10.76
ANFİS8	I_{sc}	Tramp	5	0.12	11.63	0.86	9.06	0.32	2.29	8.56	10.68
ANFİS9		Gauss		0.12	11.59	0.84	9.02	0.31	2.24	8.61	10.76

4. Conclusion

The MLR and ANFIS models were used to assess the influence of drought on barley yield. Individual drought conditions, four different drought conditions together, and refined index I_{SC} were separately used as input to study the impact. Regression and ANFIS performances were similar in estimating barley yield from the individual drought conditions. ANFIS outperformed MLR in estimating barley yield when different drought conditions were used as multiple input parameters. In this study, The ANFIS model coefficient of determination (R^2) indicates that 41 percent of the variation in yield can be explained by multiple drought conditions input whereas only 13.3 percent by multiple regression. The refined index I_{SC} is better suited for the purposes of resource allocation for drought management and mitigation. However, I_{SC} does not have any advantage in estimating barley yield. Barley yield also greatly depend on other parameters such as soil characteristics, and management practices. Focus of this study was to assess the drought impact on barley yield, and not to develop a comprehensive prediction model for barley yield. However, quantification of drought impact on yield is vital in order to develop more powerful predictive models for drought management.

5. References

[1] Narasimhan B, Srinivasan R. Development and evaluation of Soil Moisture Deficit Index (SMDI) and Evapotranspiration Deficit Index (ETDI) for agricultural drought monitoring. Agricultural and Forest Meteorology. 2005;133(1-4):69-88.
[2] Lott N, Ross T. Tracking and Evaluating US Billion Dollar Weather Disasters, 1980-2005. 2006.
[3] Li Y, Ye W, Wang M, Yan X. Climate change and drought: a risk assessment of crop-yield impacts. Climate Research. 2009;39(1):31-46.
[4] Mishra V, Cherkauer KA. Retrospective droughts in the crop growing season: Implications to corn and soybean yield in the Midwestern United States. Agricultural and Forest Meteorology.2010;150: 1030–1045.
[5] Thomson AM, Brown RA, Rosenberg NJ, Izaurralde RC, Benson V. Climate change impacts for the conterminous USA: An integrated assessment - Part 3. Dryland production of grain and forage crops. Climatic Change. 2005;69(1):43-65.
[6] Karetinkov D, Parra N, Bell B, Ruth M, Ross K, Irani D. Economic impacts of climate change on North Dakota. A review and assessment conducted by the Center for Integrative Environmental Research (CIER), University of Maryland. 2008.
[7] Leelaruban N, Oduor PG, Akyuz A, Shaik S, Padmanabhan G. Leveraging a spatio–temporal drought severity and coverage index with crop yield modelled as a stochastic process. International Journal of Hydrology Science and Technology. 2012;2(3):219-236.
[8] Alavi N., Nozari V, Mazloumzadeh SM, Nezamabadi-pour H. Irrigation water quality evaluation using adaptive network-based fuzzy inference system. Paddy and Water Environment. 2010; 8(3): 259-266.
[9] Boskidis I, Gikas GD, Sylaios GK, Tsihrintzis VA. Hydrologic and water quality modeling of Lower Nestos River Basin. Water resources management. 2012; 26(10): 3023-3051.
[10] Bacanli, UG, Mahmut F, Fatih D. Adaptive neuro-fuzzy inference system for drought forecasting. Stochastic Environmental Research and Risk Assessment. 2009; 23(8):1143-1154.
[11] Chang, FJ, Kuo WW. A systematical water allocation scheme for drought mitigation. Journal of Hydrology. 2013; 507: 124-133.
[12] Chou, FNF, and Chen BPT. Development of drought early warning index: using neuro-fuzzy computing technique. In 8th international symposium on advanced intelligence systems. 2007.
[13] Hayes MJ, Svoboda M, Le Comte D, Redmond KT, Pasteris P. Drought monitoring: New tools for the 21st century. Drought and Water Crises: Science, Technology, and Management Issues. 2005;86:53-69.
[14] Jang JS. ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans. Syst., Man. Cybern. 1993;2: 665–685.
[15] Jang JS, Sun CT, Mizutani E. Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence. Upper Saddle River, NJ: Prentice-Hall.1997;5 - 15.