Annals of Arid Zone 46(2): 133-144, 2007

Predicting Drought Using Pattern Recognition

V.K. Boken[1*], C.E. Haque[2], and G. Hoogenboom[1]
[1] *Department of Biological and Agricultural Engineering, The University of Georgia, Griffin, GA 30223, USA*
[2] *Natural Resources Institute, University of Manitoba, Winnipeg, Canada R3T 2N2*

Abstract: Predicting droughts and their impacts upon overall agricultural production helps in drought management. Generally, statistical regression or time series techniques are employed to predict agricultural droughts quantitatively. Linear (error correction, linear discriminant) and nonlinear (k-Nearest Neighbor) techniques of pattern recognition were used for predicting agricultural droughts qualitatively. A total of five crop districts in the province of Saskatchewan in the Canadian prairies, were selected. Thirty two variables were derived for each district from the daily temperature and precipitation data for the period from 1975 to 2002 to develop pattern recognition models. The variables derived from the minimum or maximum temperatures were found to be more significant than the variables derived from the precipitation for predicting moderate-to-very severe agricultural droughts. The 1975-1997 data were used for model development while the 1998-2002 data were used for model testing. About 83% accuracy was achieved in predicting the non-drought category while 71% accuracy was achieved in predicting the drought category. It was concluded that pattern recognition techniques could be applied for predicting drought qualitatively, which would aid current methods of drought prediction.

Key words: Discriminant analysis, crop yield, prairies, wheat.

Drought is an important climatic phenomenon that occurs due to water scarcity and affects various economic sectors e.g., agricultural, industrial, municipal, and recreational (Schipper, 2003). At a global scale, more people are affected by drought than any other natural disaster (Hewitt, 1997). There is a wide variation in how people, across the globe, perceive drought. More than 150 drought definitions are available in the literature (Krishnan, 1979; Dracup *et al.*, 1980; Wilhite and Glantz, 1987). In this study, we developed a method of predicting an agricultural drought (hereafter referred to as drought) for the Canadian prairies, which are characterized with a semi-arid climate system. Droughts occur when crop yields are significantly lower than their long-term averages, usually due to adverse weather conditions.

Spring wheat (hereafter referred to as wheat), canola, and barley are the major crops of the prairies with the combined harvested area of about 20 Mha (Statistics Canada, 1998), half of which is occupied by wheat. About 75% of the total wheat produced on the prairies is exported, which contributes significantly to the Canadian economy (McKay, 1983; Walker, 1989). Drought is the most costly hydro-meteorological natural disaster of Canada (Dore, 2003) and has a direct impact on

* Department of Geography and Earth Science, The University of Nebraska at Kearney, Kearney, NE 68849, USA.

Canadian wheat export. A better marketing strategy for wheat export can be developed and higher profits achieved if a drought could be predicted accurately by predicting the decline in wheat production around the harvest time as explained in the following section.

Drought Prediction

Drought prediction can be categorized as quantitative or qualitative. In the quantitative case, one predicts the value of a drought-defining variable, while in the latter case the prediction is made as to whether or not a drought of a given severity would occur. In both cases, a drought-defining variable is required to begin the prediction process. Since the crop yield (production per unit area) is directly affected by drought (Rao *et al.*, 1984; Venkateswarlu, 1993), it can be considered as a drought-defining variable. In the present study, the wheat yield is considered as a quantitative variable to define drought, because wheat is the single most important crop of the prairies with the largest harvested area and has a great economic value (Kumar, 1999). Based on a negative deviation of the estimated yield from their long-term averages, drought severity levels (e.g., nil, mild, moderate, severe, or very severe) could be determined. For example, the deviation could be greater than -10% for nil drought, range from -10% to -20% for mild drought, from -20% to -40% for moderate drought, from -40% to -60% for severe drought, and be lower than -60% for very severe drought.

In general, drought is predicted quantitatively by estimating crop yields using statistical regression or time series techniques (Walker, 1989; Kumar, 1998; Boken, 2000). Walker (1989) developed a model to estimate wheat yields for the prairies using monthly temperature and precipitation data. The crop yield also depends on additional variables relating to, for example, irrigation, fertigation, and crop disease. For the prairies, however, the irrigated area is negligible and the amount of fertigation can be considered stable over the years. Occurrence of the crop disease is spatially and temporally random and is not easy to model its impact on crop yields. As a result, only weather-based variables were used to predict wheat yields. Boken and Shaykewich (2002) modified the Walker's model by using daily temperature, precipitation, and satellite data. The objective of the present study was to test a newer approach called pattern recognition (PR) for predicting drought qualitatively. Drought being a very complex phenomenon, it is prudent to employ more than one technique to strengthen the prediction. Pattern recognition has been successfully applied in various fields, but its application to drought prediction is rather new.

Methodology

Study area and data collection

The prairies extend from 49°N (Canada-US border) to 54°N latitude, and between approximately 96°W and 114°W longitude encompassing western Manitoba, Saskatchewan, and eastern Alberta. Saskatchewan produces approximately 60% of the total wheat grown on the prairies. Five crop districts (districts-1b, 3bn, 4b, 6a, and 9a) of Saskatchewan were selected for the study. These districts lie in the zones that experience varying degrees of

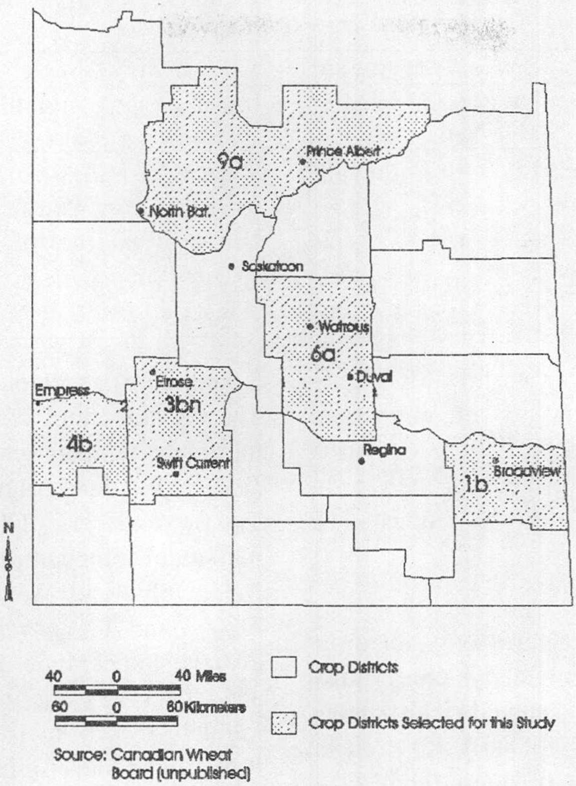

Fig. 1. The five crop districts in Saskatchewan, Canada, that were selected for the study.

drought proneness (Fig. 1). The climate of the region encompassing these districts is generally classified as semi-arid with long (and cold) winters and short (and warm) summers. Table 1 presents the variability in wheat yields and climate for the selected districts. While the district-4b experiences higher temperature and lower seasonal rainfall, the district-9a falls in a moisture-adequate region. Such a selection of districts was appropriate for the study, because it would permit model testing under non-uniform climatic conditions and therefore will help improve the model reliability.

For the selected districts, wheat yield and weather data from 1975 to 2002 were collected from the Canadian Wheat Board. The original source for the data, was Statistics Canada (www.statcan. ca). While the yield data were collected at district levels, weather data were collected for a few weather stations located across districts. Only those weather stations for which the complete data were available for the entire study period were selected. The selected weather stations included Broadview (district-1b), Elrose and Swift Current (district-3bn), Empress (district-4b), Dual and Watrous (district-6a), Prince Albert and

Table 1. The variability in spring wheat yields and short-term climate for the selected crop districts in Saskatchewan, Canada, based on the 1975-2002 data

Crop district	Spring wheat yield (t ha^{-1})				Climate	
	Minimum	Maximum	Mean	Standard deviation	Average temperature	Average precipitation (mm)
1b	1.08	2.51	1.84	±0.36	15.4	252
3bn	0.50	2.24	1.77	±0.42	16.2	221
4b	0.44	2.42	1.66	±0.55	17.0	182
6a	0.61	2.36	1.78	±0.37	16.0	251
9a	0.65	2.62	1.98	±0.39	15.2	234

North Bat. (district-9a) as shown in Figure 1. The daily temperature (minimum and maximum, in °C) and the daily precipitation (in mm) data were collected for the typical cropping period (May 1 to August 31).

Deriving the variables

The data were collected only for three quantitative variables - daily temperature, daily precipitation, and annual yield. Annual yield was used only to label a particular year as drought year or non-drought year as required by the PR techniques used in the present study. Temperature and precipitation data were used for building a drought prediction model. These parameters, in their original forms, are not as effective indicators of drought as their derivates. Also, it is desirable to consider a higher number of variables from which to select a few variables to develop a prediction model. Therefore, variables were derived from the daily temperature and precipitation data.

For each selected district, 32 variables were derived. Sixteen variables were related to the monthly averages of temperature or monthly total of precipitation, and the remaining sixteen variables were related to the standard deviation in the daily temperature or precipitation for each month. The variables thus derived included: daily minimum temperature averaged for May, June, July and August (T_{n5avg}, T_{n6avg}, T_{n7avg}, and T_{n8avg}, respectively), daily maximum temperature averaged for May, June, July, and August (T_{x5avg}, T_{x6avg}, T_{x7avg}, and T_{x8avg}, respectively), the daily average temperature for May, June, July, and August (T_{5avg}, T_{6avg}, T_{7avg}, and T_{8avg}, respectively), the total precipitation for May, June, July, and August (P_5, P_6, P_7, and P_8, respectively), the standard deviation in the daily minimum temperature for May, June, July, and August (T_{n5sd}, T_{n6sd}, T_{n7sd}, and T_{n8sd}, respectively), the standard deviation in the daily maximum temperature for May, June, July, and August (T_{x5sd}, T_{x6sd}, T_{x7sd}, and T_{x8sd}, respectively), the standard deviation in the daily average temperature for May, June, July, and August (T_{5sd}, T_{6sd}, T_{7sd}, and T_{8sd}, respectively), and the standard deviation in the daily precipitation for May, June, July, and August (P_{5sd}, P_{6sd}, P_{7sd}, and P_{8sd}, respectively).

In addition, a variable with two categories (non-drought and drought) was included. A non-drought category was assigned to a year if the wheat yield for that year exceeded a threshold yield, Y_t (Table 2). Otherwise, a drought category was assigned to the year. Five values (-10, -20, -30, -40, and -50% deviation from the average yield) were examined to define Y_t. A value of -10% deviation meant that Y_t equaled 90% of the average yield, and so on. A value of -20% was considered appropriate, because a value above -20% will not really mean a drought of much concern and a value below -20% resulted in too few drought categories to attempt a satisfactory analysis. It should be noted here the drought definition adopted in this paper may not be considered a standard definition of drought. Table 3 presents an example of deriving quantitative variables only for the month of May for the district-3bn. These variables were derived for June, July, and August for each selected district. Using the 32 quantitative variables and a categorical variable for each district, various pattern recognition techniques were applied for predicting drought qualitatively.

Table 2. Categorization of a year as drought (D) or non-drought (ND) year on the basis of the deviation from the average wheat yield

Year	Spring wheat yield (t ha^{-1})	Yield deviation (%)[a]	Categoryb
1975	1.67	-5.6	ND
1976	2.21	24.9	ND
1977	2.04	15.3	ND
1978	1.71	-3.4	ND
1979	1.53	-13.5	ND
1980	1.54	-13.0	ND
1981	1.80	1.7	ND
1982	2.21	24.9	ND
1983	1.88	6.2	ND
1984	1.28	-27.7	D
1985	0.91	-48.6	D
1986	2.08	17.5	ND
1987	1.86	5.1	ND
1988	0.50	-71.7	D
1989	1.67	-5.6	ND
1990	2.08	17.5	ND
1991	2.23	26.0	ND
1992	1.96	10.7	ND
1993	2.21	24.9	ND
1994	1.83	3.4	ND
1995	1.88	6.2	ND
1996	1.74	-1.7	ND
1997	2.07	16.8	ND
1998	1.80	1.6	ND
1999	2.24	26.3	ND
2000	2.02	14.3	ND
2001	1.25	-29.2	D
2002	1.36	-23.2	D

Note: [a] Yield deviation is from 1.77 t ha^{-1}, the average yield for the 1975-2002 period, [b] Yield deviation below -20% results in non-drought (ND), and above -20% results in drought (D) category.

Pattern recognition techniques

Pattern recognition is a process that can be used to classify an object by analyzing the numerical data that characterize the object. Various academic fields, such as image processing, medical engineering, criminology, speech recognition, and signature identification have applied pattern recognition to classify objects of interest (Duda *et al.*, 2001). However, pattern recognition techniques have hardly been explored for drought prediction. Various pattern recognition techniques are available in the literature (Jain and Flynn, 1993; Duda *et al.*, 2001), but only a few techniques were selected for the present study. Both two-variable and multiple-variable cases were considered. In the two-variable case, an error-correction (EC) procedure (Duda

Table 3. A few of the derived variables that were considered for developing pattern recognition models to predict a drought for district 3bn, Saskatchewan, Canada

Year	Categorical variable	Quantitative variables (for May[a])							
		Average of daily data				Standard deviation in daily data			
		T_{x5avg}	T_{n5avg}	T_{5avg}	P_5	T_{x5sd}	T_{n5sd}	T_{5sd}	P_{5sd}
Training									
1975	ND	15.9	3.5	9.7	61.6	4.9	3.4	3.7	4.7
1976	ND	21.9	4.5	13.2	20.4	4.9	4.7	4.3	2.2
1977	ND	20.1	6.7	13.4	126.8	6.2	3.9	4.5	8.1
1978	ND	18.6	5.7	12.1	53.9	4.9	3.2	3.7	5.2
1979	ND	15.1	2.8	9.0	37.1	7.4	4.8	5.8	2.2
1980	ND	22.6	6.0	14.3	10.0	6.1	6.2	5.9	1.5
1981	ND	19.7	5.0	12.4	24.3	5.4	4.7	4.6	1.7
1982	ND	14.8	3.3	9.01	10.8	7.3	4.5	5.3	9.4
1983	ND	16.2	2.6	9.4	53.0	6.8	4.8	5.6	6.2
1984	D	18.0	3.0	10.5	29.4	6.0	4.8	5.0	3.6
1985	D	20.9	6.7	13.9	32.6	5.5	3.2	4.0	2.5
1986	ND	18.5	5.6	12.1	106.1	8.6	5.7	6.8	7.5
1987	ND	21.4	6.4	13.9	21.5	5.4	3.6	3.6	1.7
1988	D	23.7	7.4	15.6	28.2	5.6	4.3	4.5	4.5
1989	ND	17.9	4.7	11.4	76.6	5.2	3.8	3.9	5.9
1990	ND	17.4	4.1	10.8	60.4	5.9	5.0	5.0	4.8
1991	ND	17.9	4.8	11.4	76.0	6.3	4.6	5.1	5.5
1992	ND	18.1	4.2	11.2	40.4	6.8	4.1	4.9	2.9
1993	ND	20.3	4.5	12.4	15.5	5.0	4.4	4.1	1.4
1994	ND	18.7	5.4	12.1	67.3	5.3	3.6	3.8	4.6
1995	ND	17.5	3.2	10.4	53.5	5.7	4.5	4.6	4.2
1996	ND	14.0	3.1	8.6	72.5	5.2	4.3	4.4	4.6
1997	ND	17.6	4.2	10.9	54.0	6.2	4.3	4.7	5.1
Testing									
1998	ND	21.7	5.3	13.5	37.2	4.4	4.7	4.0	4.5
1999	ND	16.7	4.7	10.7	81.3	6.5	3.9	4.7	5.6
2000	ND	18.9	4.3	11.6	19.8	4.7	4.0	3.9	2.1
2001	D	21.0	4.8	12.9	19.2	4.6	4.2	4.0	2.6
2002	D	17.6	1.5	9.6	13.0	7.6	6.4	6.7	1.3

Note: [a] Interpreting the variables: T and P refer to the daily temperature and precipitation, respectively; subscript *n* is for minimum and x is for maximum; numbers 5, 6, 7, and 8 indicate May, June, July, and August, respectively; subscript *avg* is for average and *sd* is for standard deviation. For example, T_{x5avg} is the maximum daily temperature averaged for May; P_5 is the total precipitation for May.

et al., 2001; Kumar *et al.*, 1998) was applied and, in the multiple-variable case, both linear (linear discriminant analysis) and nonlinear (Nearest Neighbor) techniques were applied.

Prior to applying the EC procedure, an yield vector is formed with two elements (in the present case, temperature and precipitation data based variables). Subsequently, three steps are taken in sequence: (i) an additional element of 1 is included into the elements of all of the yield vectors, (ii) all of the elements in every vector of the second category (i.e., drought vector) are multiplied by -1 and (iii) a solution vector, W, is determined such that the product of W with any yield vector, Y_i, exceeds zero. That is,

$$Y_i W > 0 \quad \text{for all } i \qquad \text{...[1]}$$

where,
i, which is used to identify a vector, ranges from 1 to the total number of vectors (Duda *et al.*, 2001).

To begin with, W is assumed to be a unit vector i.e., every element is equal to 1. Then the product of W with the individual yield vector is computed. If the condition (Equation 1) is not satisfied, the W is corrected as follows:

$$W_{k+1} = W_k + c/k * Y_i \qquad \text{...[2]}$$

where,
c is greater than zero (usually chosen as 1); *k*, whose initial value is zero for unit vector, W, is increased by 1 every time a correction in the W is required. Y_i is the yield vector whose product with the W does not exceed zero, and as a result, a correction in the W is sought. This process of correcting the W continues until equation 2 is satisfied.

In the present case, two variables, out of 32 quantitative variables, were selected at a time as elements of yield vector to examine if a solution vector, W, existed that could linearly separate drought from non-drought events. An iterative procedure was applied using a computer program, but no such solution vector was found to exist. This reiterates the complexity involved in the analysis of agricultural drought. To proceed further, the multiple-variable case was considered for analysis.

Pattern recognition for a multiple variable case begins with determining a subset of the variables whose linear combination best reveals the differences among classes (drought and non-drought). For this purpose, a stepwise discriminant procedure of SAS software was applied. This procedure eliminates highly inter-correlated variables. Table 4 lists the variables that were found to be significant for developing a prediction model for each district. Using these variables the linear and nonlinear techniques of discriminant

Table 4. Variables that were found to be significant, as a result of the stepwise discriminant analysis for developing drought prediction models for the selected districts of Saskatchewan, Canada

Crop district	Significant variables	Average squared canonical correlation
1b	T_{n8sd}, T_{n5sd}, T_{x7sd}	0.67
3bn	T_{x7avg}, T_{n7avg},T_{n6sd}, P_{7sd}, P_{8sd}, T_{x5sd}, T_{n6avg}, T_{8sd}	0.94
4b	T_{x7avg}, T_{6sd}, T_{n7sd}, T_{5avg}, T_{n8avg}, T_{x6sd}, T_{n6sd}, T_{8avg}	0.90
6a	T_{n5sd}, T_{x7avg}, T_{x6sd}, T_{n7avg}, T_{7avg}	0.73
9a	T_{x6avg}, T_{n7sd}, T_{n5sd}, P_6	0.65

analysis were applied to predict drought for each district.

Linear discriminant analysis: By performing this analysis on quantitative data separated by categories (non-drought and drought, in the present case), a linear discriminant function (LDF) is obtained to linearly discriminate one category from the other. To develop such a function, the whole data set for 28 years (1975-2002) was divided into two sets: the training set (1975-1997) and the testing set (1998-2002). While the training set was used for developing the LDF, the testing set was used for testing the prediction performance of the LDF. Table 5 shows the constants and coefficients for the LDFs that were obtained for all of the selected districts.

For applying linear discriminant analysis, the within-category data are required to be normally distributed. In the present case, however, the maximum number of drought years in the training period was too low (only 5) to satisfactorily test and identify the distribution. Hence, a nonparametric technique (Nearest Neighbor), which does not require a normally distributed data was also undertaken.

K-Nearest Neighbor analysis: This rule classifies an unknown subset of variables to the category of majority of its *k* Nearest Neighbors. Let $x'_n \in \{x_1, x_2, ..., x_n\}$, x'_n will be a Nearest Neighbor to x if

$$\min d(x_i, x) = d(x'_n, x) \qquad ...[3]$$

where, $i = 1, 2, ..., n$.

The value of *k* used for this technique, was 2. Table 5 presents the classification (or prediction) accuracy achieved using this technique. The Nearest Neighbor analysis was repeated with *k*=3, but the accuracy was found to be the same as in the case of *k*=2.

Results and Discussion

Significant variables

Three variables for the district-1b, eight variables each for the districts-3bn and 4b, five variables for the district-6a, and four variables for the district-9a were found to be most significant for drought prediction (Table 4). Some of the variables were found to be significant for more than one crop district. Two variables, T_{n5sd} and T_{x7avg}, were significant for the highest number (3) of crop districts. T_{n5sd} was significant for the districts-1b, 6a, and 9a, while T_{x7avg} was significant for the districts-3bn, 4b, and 6a. T_{n7avg} was significant for two crop districts (4b and 6a). In parallel, T_{x6sd} was found to be significant for two crop districts (4b and 6a), and T_{n5sd} too was significant for two districts (1b and 6a).

Based on the above observations, it can be stated that the maximum and minimum temperatures in May and July are the most critical variables to predict drought for the Canadian prairies. Wheat sowing takes place in May while the heading phenological phase occurs in July. Weather conditions during the sowing and heading phases affect crop yield significantly. Out of the 20 variables that were found significant, only three variables (P_6, P_{7sd} and P_{8sd}) were derived from the precipitation data while the remaining 17 variables were derived from the temperature data. Out of these 17 variables, only five variables (T_{5avg}, T_{6sd}, T_{7avg}, T_{8sd}, and T_{8avg}) pertained to the daily average temperature while the remaining variables were related to either

Table 5. Constants and coefficients for linear discriminant functions that were developed for predicting drought and non-drought events for five crop districts in Saskatchewan, Canada

District 1b			District 3bn			District 4b			District 6a			District 9a		
Variable or constant	Coeff1[a]	Coeff2[b]	Variable or constant	Coeff1	Coeff2	Variable or constant	Coeff1	Coeff2	Variable or constant	Coeff1	Coeff2	Variable or constant	Coeff1	Coeff2
Constant	-110.7	-164.3	Constant	-1482.0	2081.0	Constant	-770.5	-1003	Constant	-359.2	-467.6	Constant	-85.1	-154.8
Tn8sd	22.8	28.1	Tx7avg	212.6	258.7	Tx7avg	34.5	41.9	Tn5sd	24.2	29.3	Tx6avg	10.6	14.3
Tn5sd	16.0	19.7	Tn7avg	-281.3	-347.5	T6sd	129.6	199.2	Tx7avg	-665.2	-784.0			
Tx7sd	21.6	25.7	Tn6sd	154.3	186.1	Tn7sd	-61.8	-78.5	Tx6sd	31.9	36.0			
			P7sd	36.9	44.2	T5avg	23.4	27.7	Tn7avg	-692.4	-817.9			
			P8sd	44.4	52.7	Tn8avg	10.7	21.0	T7avg	1372.0	1617.0			
			Tx5sd	63.5	74.9	Tx6sd	-35.0	-64.4						
			Tn6avg	-24.9	-30.9	Tn6sd	2.8	-25.1						
			T8sd	-41.8	-53.9	T8avg	7.4	1.8						

Note: [a] Coeff.1 refers to the coefficients for category 1 i.e., nondrought, and [b] coeff.2 refers to the coefficients for category 2 i.e., drought.

Table 6. Accuracy obtained in classifying (or predicting) a non-drought event using Nearest Neighbor technique, for training and testing datasets, for the selected crops districts of Saskatchewan, Canada

Crop district	Training set				Testing set			
	ND[a]	D[b]	Total	Accuracy (%)	ND	D	Total	Accuracy (%)
1b	17	1	18	94	4	1	5	80
3bn	20	0	20	100	3	0	3	100
4b	18	0	18	100	1	2	3	33
6a	18	1	19	95	3	0	3	100
9a	22	0	22	100	4	0	4	100
Total/ Average	95	2	97	95/97*100=98	15	3	18	15/18*100=83

Note: [a]ND refers to a non-drought, and [b]D refers to a drought event.

minimum or maximum temperature. Therefore, it would be appropriate to conclude that the discrimination of a drought event from a non-drought event is related more to the daily extreme temperatures than to the daily average temperatures. Nevertheless, it should be noted here that the drought category as defined in this paper refers to a combined severity level that ranges from moderate to very severe. It is possible that the influence of temperature or precipitation on drought prediction will change if a drought of more specific severity level is examined. Such an examination could not be possible as explained in the later part of the following section.

Prediction accuracy

For the present analysis, accuracy is defined as the percentage of events (drought or non-drought) correctly classified (or predicted) in a training or testing dataset using a PR model which was developed using the training dataset. The accuracy achieved by using the Nearest Neighbor technique for qualitative prediction of non-drought and drought events is presented in Tables 6 and 7, respectively. Accuracy is shown for both training and testing datasets. It can be observed that overall (or average) accuracy was higher for classifying non-drought events (Table 6) than drought events (Table 7). In addition, the average accuracy was higher in the case of the training set as compared to the testing set. Training set accuracy was 98% for predicting non-drought events and 100% for predicting drought events.

The average accuracy for the testing set was 83% for predicting the non-drought events and 71% for predicting the drought events. Since the number of drought events during the study period was low, droughts of moderate, severe, and very severe levels were combined to define a single category to obtain a reasonable number of drought events in the dataset to perform a satisfactory analysis. The threshold yield for a district was assumed to be 80% of the long-term average yield of the district.

In the present dataset, the wheat yield for the drought categories ranged widely

Table 7. Accuracy in classifying (or predicting) a drought event using a Nearest Neighbor technique, for training and testing datasets for the selected crop districts of Saskatchewan, Canada

Crop district	Training set (1975-1997)				Testing set (1975-1997)			
	ND[a]	D[b]	Total	Accuracy (%)	ND	D	Total	Accuracy (%)
1b	0	5	5	100	0	0	0	NA
3bn	0	3	3	100	0	2	2	100
4b	0	5	5	100	0	2	2	100
6a	0	4	4	100	1	1	2	50
9a	0	1	1	100	1	0	1	0
Total/ Average	0	18	18	100	2	5	7	5/7*100=71

Note: [a.]ND refers to a non-drought, and [b.]D refers to a drought event.

- from 20% to 75% of the long-term averages. If this range were narrower (e.g., 20% to 40%, 40% to 60%, or 60% to 80%) which would refer to a drought of more specific severity level, the behavior of temperature and precipitation based variables could be better judged. But in the present case, an adequate number of drought events did not exist within the narrower ranges and hence the analysis with varying ranges, though desired, could not be attempted.

Conclusion

The Nearest Neighbor technique using temperature and precipitation data was applied to classify or predict agricultural drought events for five crop districts in Saskatchewan, Canada. Within the data period (1975-2002), the moderate to very severe droughts were grouped into a single drought category. The overall accuracy was 83% for predicting non-drought events and 71% for predicting drought events. It can be concluded that pattern recognition techniques could also be used for drought prediction and their potential should be explored further. Analysis of agricultural drought is a very complex process and it would always be wise to attempt drought prediction using multiple techniques, quantitative as well as qualitative, to strengthen the prediction and improve the decision making for drought management.

References

Boken, V.K. 2000. Forecasting spring wheat yield using time series analysis: a case study for the Canadian Prairies. *Agronomy Journal* 92(6): 1047-1053.

Boken, V.K. and Shaykewich, C.F. 2002. Improving an operational wheat yield model for the Canadian Prairies using phenological-stage-based normalized difference vegetation index. *International Journal of Remote Sensing* 23 (20): 4157-4170.

Dore, M.H.I. 2003. Forecasting the conditional probabilities of natural disasters in Canada as a guide for disaster preparedness. *Natural Hazards* 28(2): 249-269.

Dracup, J.A., Lee, K.S. and Paulson, Jr., E.G. 1980. On definitions of droughts. *Water Resources Research* 16(2): 297-302.

Duda, R.O., Hart, P.E. and Stork, D.G. 2001. *Pattern Classification* (2^{nd} Ed.), John Wiley and Sons, New York.

Hewitt, K. 1997. *Regions at Risk. A Geographical Introduction to Disasters.* Addison Wesley Longman, UK.

Jain, A.K. and Flynn, P.J. 1993. *Three-dimensional Object Recognition Systems*. Elsevier, New York.

Krishnan, A. 1979. Definitions of drought and factors relevant to specifications of agricultural and hydrological droughts. In *Proceedings of the International Symposium on Hydrological Aspects of Droughts*, Vol. I, pp. 67-102. Indian Institute of Technology, New Delhi, India.

Kumar, V. 1998. An early warning system for agricultural drought in an arid region using limited data. *Journal of Arid Environments* 40(2): 199-209.

Kumar, V. 1999. Prediction of agricultural drought for the Canadian Prairies using climatic and satellite data. *Ph.D. Dissertation*, Department of Geography, University of Manitoba, Winnipeg, Canada. (Unpublished)

Kumar, V., Haque, C.E. and Pawlak, M. 1998. Determining agricultural drought using pattern recognition. In *Prairie Perspectives – Geographical Essays* (Eds. J.C. Lehr and H.J. Selwood), pp. 74-80. Department of Geography, University of Winnipeg, Winnipeg,

McKay, G.A. 1983. Climatic hazards and Canadian wheat trade. In *Interpretation of Calamity from the View Point of Human Ecology* (Ed. K. Hewitt), pp. 220-228. Allen and Unwin Inc., Boston.

Rao, G.G.S.N., Ramakrishna, Y.S., Ramana Rao, B.V. and Purohit, R.S. 1984. Incidence of droughts and its impact on food production in Rajasthan. *Annals of Arid Zone* 23(4): 327-334.

Schipper, O. 2003. Declining water resources raise food concerns. *Environmental Science and Technology* 37(15): 273A-274A.

Statistics Canada, 1998. *Estimates of Production of Principal Crops*, Field Crop Reporting Series No. 8, Statistics Canada, Ottawa.

Venkateswarlu, J. 1993. Effect of drought on kharif foodgrains production: A Retrospect and Prospect. *Annals of Arid Zone* 32(1): 1-12.

Walker, G.K. 1989. Model for operational forecasting of western Canada wheat yield. *Agricultural and Forest Meteorology* 44: 339-351.

Wilhite, D.A. and Glantz, M.H. 1987. Understanding the drought phenomenon. In *Planning for Drought: Toward a Reduction of Societal Vulnerability* (Eds. D.A. Wilhite and W. Easterling), pp. 11-27.Westview Press, Boulder.