Forecasting wheat yield in the Canadian Prairies using climatic and satellite data

V. Kumar, University of Manitoba
C. E. Haque, Brandon University

Abstract: Wheat yield forecasting is a significant component of planning and management of wheat exports from the Canadian Prairies. Commonly, regression models employing agroclimatic data pertaining to post-planting period are used as forecasting tools. However, the planning process of wheat export requires yield estimates even in advance of crop planting. Such advanced estimates are possible by using time series analysis. In this paper, we use both regression and time series approaches to develop models to forecast spring wheat yield using climate and satellite data pertaining to Swift Current, Saskatchewan.

In the case of regression, two categories of models were developed. In the first category, the climate data i.e., accumulated precipitation from April to July, and average temperature during the same period were used as independent variables. The second category included NOVAA/ AVHRR satellite data based Normalized Difference Vegetation Index (NDVI) in addition to the climate data used in the first category. While the climate data were available for 1975-1994 period, the satellite data were available only for 1987-1994 period. Hence, the time series models and the first category of regression models were developed using the 1975-1991 data and tested for the 1992-1994 period, while the second category of regression models were developed using 1987-1992 data and tested for the years 1993 and 1994. Based on the mean absolute percent error, the comparative performances of the models are presented and discussed

Introduction

Wheat is a major crop in the Prairies and its export contributes significantly to the Canadian economy. About 20 to 30 million tonnes of wheat is exported every year from the Prairies. The variation in exports is attributed largely to drought-prone climate, which, if not accurately predicted, can result in millions of dollars of loss. The export targets and prices that are set much in advance of crop harvests may not turn out to be accurate after a drought occurrence.

Since weather is the single most important factor affecting wheat yields in the Prairies, temperature and precipitation data over the crop growing period are used to forecast wheat yield. Most commonly, regression models are used for the purpose, and the weather data after the crop has been planted are employed. But the yield estimates are required even prior to crop-planting, for advanced export planning by a marketing agency such as the Canadian Wheat Board (CWB). Yield estimation prior to crop planting is possible by using time series analysis wherein a forecast of yield in the following year is made on the basis of yield data in the past. In this paper, we have developed models using regression as well as time series analysis and have provided comparative results that may be helpful in making optimum decision regarding wheat export from the Canadian Prairies.

Data Used

Spring wheat yield, climate, and satellite data pertaining to Swift Current, Saskatchewan, were used in the study. The yield and satellite data were originally obtained from Statistics Canada, and the climate data (temperature and precipitation) from Environment Canada. The satellite data refer to AVHRR (Advance Very High Radiometric Resolution) sensor of NOAA (National Oceanic Atmospheric Administration) satellite. The NOAA / AVHRR data is available in five wavelength channels, but only the first two channels (red : 0.58 - 0.68 m, and infrared : 0.725 - 1.10 m) have been commonly utilized to develop various indices for monitoring vegetation conditions (Goward et al., 1991). NDVI as defined below, has been found to be a globally useful index in crop yield estimation (Barnett and Thompson, 1982; Brown et al, 1982; Rudorff and Batista, 1990; Bullock, 1992).

$$\mathbf{NDVI=(R-IR)/(R+IR)} \qquad [1]$$

Table 1: *The dataset used in the study.*

Year	Yield (bu/ac)*	Total Precipitation (April-July), mm	Average Temperature (April-July), ^{0}C	NDVImax (week23-week30)	NDVIavg (week23-week30
1975	24.9	198.3	10.58		
1976	32.8	215	13.05		
1977	30.4	222.9	13.40		
1978	25.5	178.9	12.63		
1979	22.7	166.7	10.85		
1980	22.9	171.6	14.20		
1981	26.8	243	12.08		
1982	32.9	248.2	10.30		
1983	28	212.4	11.70		
1984	19	127.4	12.90		
1985	13.5	84.3	12.85		
1986	31	247.6	12.18		
1987	27.6	173	14.23	106.8	36.1
1988	7.4	155.4	15.13	125	117.2
1989	24.9	236.3	12.58	160.5	104.4
1990	30.9	227.1	11.75	169.6	140.6
1991	33.2	409.3	12.53	179.5	132.8
1992	29.1	204.4	11.55	189.6	131.3
1993	32.9	219.8	11.35	155.6	114.6
1994	27.2	186.8	12.45	171.3	123.1

* bu/ac (bushel per acre) which is equivalent to 0.06725 t/ha .

where **R** and **IR** are the reflectance in the red and the infra-red bands of the AVHRR sensor of the NOAA satellite.

The weekly NDVI values (Julian week 23 to week 30 i.e., June and July) for the 1987-1994 were used in the analysis. This period was chosen as there is high level of canopy coverage during this period and yield is considered to be related to NDVI during this period. The complete data set used in the study is given in Table 1.

Wheat Yield Estimation

The yield can be estimated by radiative and temperature regimes, soil water available for plant growth, plant nutrients , interference of pests and disease, and farm management activities (Parry et al., 1988). Various models have been developed based on basic as well as derived parameters related to soil moisture, weather , and crop characteristics. For example, some of the parameters that have been used in wheat yield estimation are albedo i.e., ratio of reflected light to incident light on the crop during the growing season (Idso et al. ,1979), moisture anomaly index i.e., difference between monthly observed precipitation and climatically appropriate precipitation, ratio of evapotranspiration to potential evapotranspiration (Sakamoto, 1978), crop water use i.e. total evapotranspiration (Slabbers and Dunin , 1981), canopy temperature indices i.e., Stress Degree Day, Temperature Stress Day and, and Crop Water Stress Index (Diaz et al., 1983). All these studies were carried out on experimental basis requiring data that are otherwise difficult to measure and therefore not available for consistent period over a large area such as the Prairies. In western Canada, Raddatz et al. (1994) estimated average yield of Spring wheat using regression analysis. End-of-season ratios of water- use to water demand and the modelled days-to-maturity were used as variables. Up to 69 percent of the variation in the observed yields of spring wheat could be explained by the regression models.

Methodology

Two different approaches of forecasting: statistical regression, and time series analysis have been attempted in this paper as explained in the following.

Regression Analysis

Yield versus climatic variables:

Considering spring wheat yield as dependent variable, and average temperature (April to July) , T, and total precipitation (April to July), P, as independent variables, a regression analysis was conducted using 1975-1994 data. Three different regression models using 1975-1991, 1975-1992, and 1975-1993 data were developed to forecast yields in the year 1992, 1993, and 1994 respectively.

As can be seen from Table 1, the yields were very low in the year 1985, and 1988, due to drought. To examine the effect of

Table 2: *Regression analysis using climatic data.*

Year	Reported yield (bu/ac)	Estimated Yield (bu/ac)			
		Using continuous data range		Using discrete data range (excluding 1985,1988 data)	
		R^2	Yest.	R^2	Yest.
1992	29.1	0.57	26.92	0.49	26.95
1993	32.9	0.57	28.42	0.48	27.9
1994	27.2	0.57	24.74	0.45	26.47
	MAPE (%)		10.05		8.42

drought on forecasting, the 1985 and 1988 data were excluded from the dataset and regression analysis was repeated. Table 2 shows the results thus obtained. The mean absolute percent error (MAPE) as defined blow was used as a measure of accuracy of the models.

$$\text{MAPE} = \frac{\sum_{i=1}^{n} |Yest.i - Y| / Y * 100}{n} \qquad [2]$$

where ***Yest.*** is estimated yield, **Y** is reported yield and ***n*** is number of observations.

Yield versus climatic and satellite data:

The NDVI data were available only for 1987-1994 period. A common data set (1987-1994) was therefore used for regression analysis. The yield was predicted for the year 1993 and 1994, using 1987-1992, and 1987-1993 data respectively. Regression analysis was repeated for the dataset without 1988 data (the drought year). Regression results thus obtained are given in Table 3.

Time Series Model:

The yield, Yt, was modelled as a function of time. In general such models can be written as (Abraham and Ladolter , 1983):

Table 3: *Yield prediction using regression models using climatic and satellite data.*

Variables used	Predicted yield (bu/ac)								MAPE(%)	
	1993 (reported yield : 32.9 bu/ac)				1994 (reported yield : 27.2 bu/ac)					
	using 1987-92		excluding 1988 data		using 1987-93		excluding 1988 data		using full data	excluding 1988 data
	R^2	Yest.	R^2	Yest.	R^2	Yest.	R^2	Yest.		
T,P	0.68	31.81	0.49	29.18	0.71	26.37	0.44	28.55	3.18	8.14
T,P, NDVIavg	0.94	32.82	0.54	24.17	0.94	24.33	0.54	26.24	5.40	15.03
T,P, NDVImax	0.94	41.78	0.58	32.25	0.88	22.13	0.67	25.99	22.82	3.21

$$\mathbf{Yt=f(t:\beta) + \varepsilon_t} \qquad [3]$$

where $\mathbf{f(t:\beta)}$ is a function of time **t** and an unknown coefficient $\boldsymbol{\beta}$**:** ε_t refers to uncorrelated errors.

The 1975-1994 yield data was used to construct the following time series. To test performance of each model, the yields in 1992, 1993, and 1994 were forecast using 1975-1991, 1975-1992, 1975-1993 data respectively.

Trend analysis:

Two types of trend models were attempted: linear, and quadratic. The linear model can be defined as

$$\mathbf{Yt= \beta_0 + \beta_1 t + \varepsilon_t} \qquad [4]$$

where β_1 represents the average change from one period to the next, and quadratic trend model accounting for curvature in the trend is defined as

$$\mathbf{Yt= \beta_0 + \beta_1 t + \beta_2 t^2 + e_t} \qquad [5]$$

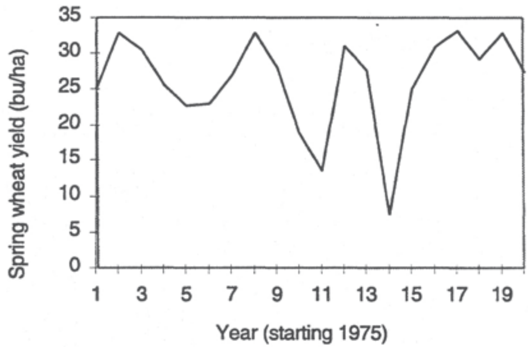

Figure 1: *Yield variation in Swift Current (1975-1994).*

Since 1985, 1988 data appear to have distorted the trend (Figure 1), the trend models were also fitted on yield series wherein the 1985, and 1988 observations were replaced with the mean yield. Table 4 gives the results thus obtained.

Moving average method:

This method uses moving averages to smooth out noise in a time series . Moving length was chosen as 2 in the present case as there is frequent fluctuation in yield series.

Exponential smoothing:

This method uses an exponentially weighted average of all past values of series to calculate smoothed value at each period. The nitial value was calculated by backcasting.

ARMA model

Here, the observation at time t, Yt, is modelled as a linear combination of previous observations,

$$\mathbf{Yt= \Sigma_{j\geq 1}\, \pi_j Yt\text{-}j + \varepsilon_t} \qquad [6]$$

Such a representation is called an autoregressive model , since the series at time **t** is regressed on itself at lagged time periods.

Table 4: *Yield prediction using trend analysis.*

Year	Reported yield (bu/ac)	Predicted yield (bu/ac)			
		linear trend		quadratic trend	
		no-change in data	drought yields replaced with mean yield	no-change in data	drought yields replaced with mean yield
1992	29.1	24.08	28.38	29.45	32.12
1993	32.9	25.03	28.71	30.9	31.98
1994	27.2	26.61	29.8	33.63	33.57
	MAPE (%)	14.45	8.26	10.31	12.20

Such an autoregressive representation can lead to models with many parameters that may be difficult to interpret. However, autoregressive models can be approximated by autoregressive moving average (ARMA) models of the form :

$$Yt= \phi_1 Y_{t-1} +...+ \phi_p Y_{t-p} + \varepsilon_t - \phi_1 \varepsilon_{t-1} -...- \phi_q \varepsilon_{t-q} \qquad [7]$$

In these models the observation, Yt , is written as a linear combination of past observations and past errors, p and q are the orders of the autoregression and of the moving averages. An appropriate model can be built from the past data following an approach developed by Box and Jenkins (1970). On observing the respective graphs, values of both p and q were found to be 2, and as a result ARMA (2, 2) was used to forecast yield. The results of moving average, exponential smoothing and ARMA model are given in Table 5.

Results and Discussion

Regression results are shown in Table 2 and Table 3. When regression models were tested to forecast yield, the mean absolute percent error (MAPE) was minimal (3.18%) for the models using

Table 5: *Yield prediction using moving average, exponential smoothing and ARMA models.*

Year	Reported yield (bu/ac)	Predicted yield (bu/ac)				
		Moving average	Exponential smoothing		ARMA	
			no change in data	drought yields replaced with mean yield	no change in data	drought yields replaced with mean yield
1992	29.1	32.05	25.9	30.3	30.01	25.02
1993	32.9	31.15	26.15	28	19.4	24.7
1994	27.2	31	26.7	28.2	26.7	31.5
	MAPE (%)	9.81	11.12	7.56	15.33	18.25

only temperature and precipitation data. These models were developed using contracted datasets (1987-1992 and 1987-1993). Contrary to a pre-analysis assumption, the NDVI did not improve model accuracy. It can be attributed to distortion in NDVI values because of clouds or other unknown factors. In the presence of clouds, which are frequent weather feature of the Prairies, the NOAA/AVHRR image does not show ground features but clouds, and therefore does not reflect crop conditions. Besides, presence of fallow fields may also affect the NDVI values. A typical NDVI profile distorted under cloudy conditions is shown in Figure 2.

Under normal conditions, NDVI profile looks like a normal curve. If the NDVI values are adjusted for clouds, the data may help improve the model's accuracy. Further, the average NDVI was found to be a better indicator of yield compared to the maximum NDVI. But again, more reliable findings can be achieved only with appropriately-adjusted NDVI data.

In order to examine the effect of droughts, the analysis was repeated on the dataset excluding drought data (1985, 1988). In the larger dataset, the prediction errors reduced from 10.05 to 8.4 %

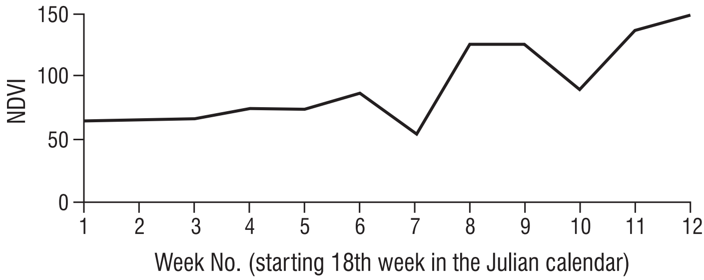

Figure 2: *Drop in NDVI values due to clouds.*

(Table 2). However, a similar trend was not observed in the case of the shorter dataset (Table 3). In regression models using, T and P, using T, P, and NDVIavg, the error increased . But, when the NDVImax was used in place of NDVIavg, exclusion of the 1988 data improved the model accuracy drastically (percent error reduced from 22.82 to 3.21). It can be seen from the Table 1 that the NDVI value for the minimum yield (7.4 bu/ac) in 1988 was very high (125), and even higher than in 1987 when the yield was rather normal (27.6 bu/ac). This tends to reveal that the NDVImax is not a good indicator of yield under drought conditions.

The rationale behind use of time series analysis for predicting the yield can be verified from the fact that yield affecting variables such as temperature, and precipitation have well- known time dependent cyclic variations. In the Prairies, precipitation is the main factor affecting yield and is part of a time dependent global hydrological cycle. Therefore, the yield can be indirectly considered as a time dependent parameter assuming that all other yield affecting factors are stable. As can be seen from Table 4 and 5, time series models were developed using data with no change in their values and using data in which the observations in 1985 and 1988 were replaced with the mean values. Fitting a trend line in the yield series, the quadratic trend showed better results (error 10.31 %) than the linear trend (14.45%). When the yields in drought years were replaced with the mean yields, linear trend provided an improved performance (error reduced to 8.26%) while the quadratic trend showed a decline in accuracy (error 12.20 %).

From Table 5, it is evident that the ARMA model gave the poorest result. The best results (error only 7.56 %) were achieved in case of exponential smoothing for the case where the drought yields were replaced with the mean yields. However, when no change in the data was made, the moving averages method gave a better result (error 9.81%). Hence, in the case of drought, the moving average technique may be a better option, while in the case of normal yield or moderate drought conditions, an exponential smoothing provides a more reliable forecasting technique.

Conclusion

1. A regression model based on average temperature (April o July) T, and total precipitation (April to July) was found to be the best among all the models developed in the study.
2. The average weekly NDVI for the period from June to July was a better indicator of wheat yield than the maximum NDVI over the same period.
3. In time series analysis, exponential smoothing of the yield series provided best estimates under normal conditions. However, the moving average method showed better performance under drought conditions.

References

ABRAHAM, B. and LEDOLTER, J. 1983 *Statistical Methods for Forecasting* New York : John Wiley and Sons

BARNETT, T.L. and THOMPSON, D.R. 1982 'The use of large-area spectral data in wheat yield estimation' *Remote Sensing of Environment* 12:509-518

BOX, G.E.P. and JENKINS, G.M. 1976 *Time Series Analysis: Forecasting and Control* San Fransisco :Holden-Day

BROWN, R.J., BERNIER, M., FEDOSEJEVS, G. and SKRETKOWICZ, L. 1982 'NOAA /AVHRR crop condition monitoring' *Symposium on Photogrammetric and Remote Sensing* Sept. 13-17 Toulouse, France

BULLOCK, P.R. 1992 'Operational estimates of western Canadian grain production using NOAA/AVHRR LAC data' *Canadian Journal of Remote Sensing* 8(1):23-28

DIAZ, R.A., MATTHIAS, A.D. and HANKS, R.J. 1983 'Evapotranspiration and yield estimation of spring wheat from canopy temperature' *Agronomy Journal* 75: 805-810

GOWARD, S. N., MARKHAM, B., DYE, D.G., DULANEY, W. and YANG, J. 1991 'Normalized difference vegetation index measurements from the Advanced Very High Resolution Radiometer' *Remote Sensing of Environment* 35:257 -277

IDSO, S.B., HATFIELD, J.L., JACKSON, R.D. and REGINATO, R.J. 1979 'Grain yield prediction: extending the stress-degree-day approach to accommodate climatic variability' *Remote Sensing of Environment* 8:267 - 272

PARRY, M.L., CARTER, T.R. and KONIJN, N.T. (Ed) 1988 *Impact Of Climatic Variations On Agriculture Vol. 1, Assessment In Cool Temperature And Cold Regions* London : Kluwer Academic Publishers

RADDATZ, R.L., SHAYKEWICH, C.F. and BULLOCK, P.R. 1994 'Prairies crop yield estimates from modelled phenological development and water use' *Canadian Journal of Plant Science* 74 : 429-436

RUDORFF, B. F. T. and BATISTA, G.T. 1990 'Spectral response of wheat and its relationship to agronomic variables in the tropical region' *Remote Sensing of Environment* 31:53 - 63

SLABBERS, P. J. and DUNIN, F.X. 1981 'Wheat yield estimation in northwest Iran' *Agricultural Water Management* 3 :291-304

SAKAMOTO, C.M. 1978 'The Z-index as a variable for crop yield estimation' *Agricultural Meteorology* 19: 305-313

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