MODEL BUILDING TO MEASURE IMPACT OF WEATHER ON CROP YIELDS

Arlin M. Feyerherm

Gary M. Paulsen

Recommended Citation

MODEL BUILDING TO MEASURE IMPACT OF WEATHER ON CROP YIELDS

Arlin M. Feyerherm   Gary M. Paulsen
Department of Statistics   Department of Agronomy
                    Kansas State University
                    Manhattan, KS 66506

ABSTRACT

The object of this research was to identify and evaluate alternatives when building mathematical models to measure the impact of weather on crop yields. Alternatives exist relative to selection of: (1) observational units with attention to size and coverage (areal and temporal), (2) observational periods for defining weather variables, and (3) mathematical forms and types of weather variables to measure impacts of moisture and temperature. The study involved an analysis of four weather-yield functions for winter wheat. The functions represented combinations of levels of two factors: (1) size and coverage of the observational units (plot yields from a multi-state area vs. average farm yields over Agricultural Statistics Districts in Kansas) and (2) weather variables used to represent moisture impacts (precipitation vs. evapotranspiration). From an eight-year test, using data from Kansas, we concluded that functions developed from a broad coverage (plot yields from a multi-state area) may have had a slight edge in precision.

KEYWORDS: weather, wheat

1. INTRODUCTION

Various approaches have been taken to address the problem of modeling crop yields as a function of soil, plant, and atmospheric factors. Many of these were referenced in Doraiswamy et al, (1979). The Large Area Crop Inventory Experiment (LACIE) (NASA, 1979), a joint effort of NASA, USDA, and NOAA to apply satellite and computer technology to global crop forecasts, hastened the search for weather-based yield models. Our efforts in this direction dated from the LACIE project and its follow-up program AgRISTARS (Agriculture and Resources Inventory Surveys Through Remote Sensing), and were reported in Feyerherm and Paulsen (1981, 1984).

In this paper, we take a systematic look at some major features of model development (sample selection, selection of weather variables and their observational periods, and application of statistical methods for variable selection) to estimate and forecast crop yields. We conclude with applications and tests of models to forecast winter wheat yields, within the state of Kansas. The models display grain yield as a function of indicator variables to measure differences in yield levels across the state, linear trend terms (as surrogates for technology), and a weather-yield function (WYF) to measure departure of a yield from its expected (trend) value.

2. MATERIALS AND METHODS

The yield (Y_{st}) of a crop at location (s) in year (t) may be expressed as a general linear model of the form:

$$Y_{st} = \sum_{i=1}^{R} \beta_i X_{ist} + \sum_{j=1}^{S} \gamma_j Z_j + \sum_{j=1}^{S} \delta_j Z_j t + \epsilon_{st} ; \qquad [1]$$

$$s = 1,2,\ldots,S; \; t = 1,2,\ldots,T;$$

where

X_{ist} = value of the _ith_ weather variable at location (s) in year (t);

Z_j = 1 if j = s; Z_j = 0 if j ≠ s,

ϵ_{st} = random error for location (s) and year (t);

β_i (i = 1,2,...R), γ_j, and δ_j (j = 1,2,...,S) are parameters to be estimated from a sample.

We confine our discussion to models of the form of Eq. [1] (linear in the parameters). The expression $\sum_{i=1}^{R} \beta_i X_{ist}$ will be called a weather-yield function, which takes on a value at location (s) and year (t). A sample of observations used to estimate the parameters will be called a developmental data set.

Structure of a Development Data Set

The data set to estimate parameters in Eq. [1] consists of vectors of observations of Y_{st}'s, X_{ist}'s, Z_j's and t; one vector for each location and year. In selecting a data set, choices must be made relative to: (1) size of the observational unit for yields [e.g., experimental plot, ASD (Agricultural Statistics District) within a state, state]; (2) spacial coverage of observational units (e.g., one ASD, multiple ASD's, one state, multiple states), (3) the observational periods for weather variables (e.g., weeks, months, stages of plant development), and (4) the weather variables to show the impact on yields of temperatures (e.g., average daily maximum, average daily minimum for air/soil/plants) and moisture (e.g., precipitation in various forms, soil moisture at various depths, evapotranspiration) over the different observational periods.

Previous studies involve various combinations of the above four choices. For example, some related yields for a single state (S = 1) for T years to average daily temperatures and accumulated precipitation for monthly periods over a state-wide unit (Thompson, 1969; Strommen et al, 1979). Others observed yields on experimental plots over multi-state locations and used average daily minimum and/or maximum temperatures, precipitation, simulated soil moisture, or simulated evapotranspiration as weather variables observed within different stages of plant development (Baier, 1973; Haun, 1974; Feyerherm and Paulsen, 1981, 1984).

Guidelines for Data Set Selection

With estimation of yields over large areas as a goal, the selection of weather variables (X_i's in Eq. [1]) and their observational periods and the choice of observational units (size and coverage) should be such that the β_i's (coefficients of X_i's in Eq. [1]) may be assumed stable for populations that include observations for seasons and locations not included in the developmental data set. Ideally, the weather-yield function portion of Eq. [1] would be applicable on a global basis and for decades to come. The γ_j's and δ_j's are expected to be location-specific, and the δ_j's may have to be reestimated in future years.

The number of parameters in Eq. [1] may be quite large. Accordingly, a large sample size (probably in the hundreds) is needed to detect certain β_i's $\neq 0$ with reasonably small standard errors of their estimators. Both locations and years may be used to increase sample size, but values of some pairs of weather variables may have higher correlations among locations than among years. This emphasized the importance of a large T (number of years) in the developmental data set.

To enhance chances that a weather-yield function will have stable parameters across populations, a sample needs to be selected so that weather variables range over a set of values similar to what is encountered or will be encountered for locations and years not included in the developmental data set. This will help avoid common problems that occur during extrapolation beyond values in data sets used for parameter estimation.

A final guideline for set selection is that the observational periods for defining weather variables should coincide with the different stages in plant development. Effects of temperature and moisture, as measured by β_i's, are expected to vary among stages. There must be a balance between selecting periods that are too short (e.g., a week), so that detection of $\beta_i \neq 0$ may be difficult, or too long, so that significant variation of a β_i within a period is missed. Equating stages of development at different locations is possible because the average stages of development can be equated to average temperatures, which then can be equated to a calendar period at each location (Feyerherm and Paulsen, 1984).

Procedures for Variable Selection

Available computer software provides a variety of procedures for selecting variables. It is tempting to begin an analysis by putting in every conceivable weather variable for which values are known. However, almost surely, it would be difficult to give meaningful interpretations to some of the β_i's because of multicollinearity and implicit relationships among variables. A better approach is to proceed through a sequence of trials in which new trials may build on what is learned from prior runs. The set of weather variables used as candidate variables on each trial should give rise to interpretable estimates of the β_i's and fundamentally sound relationships between yield and weather elements.

As a starting point, assume four stages of plant development in a season (e.g., establishment, vegetative, differentiation, and reproductive). For each stage, one might include: (a) both average daily minimum and average daily maximum temperature in linear form, and (b) variables or linear combinations of variables (e.g., $\beta_i X_i + \beta_2 X_i^2$) that are reasonable expressions of the contributions of moisture to yield. This will spotlight highly significant variables. For temperatures, it may give hints whether minimum or maximum temperature are preferred and whether average daily range should be considered (coefficients of both maximum and minimum temperatures significant and opposite in sign). For moisture, it will give the first indication of how to express moisture effects. The initial trial should include all indicator variables (Z_j's in Eq. [1] and trend terms ($Z_j t$); however, only those with significant effects need be retained for follow-up trials.

In succeeding trials, one should: (1) test alternative expressions of moisture and/or temperature for statistical significance, (2) include squares of X_i's as candidate variables to look for curvilinear relations between yield and X_i's, and (3) include products of X_i's from different periods to see if the influence of a variable is dependent on some antecedent weather condition. Relative to the last suggestion one may find that the effect of precipitation, during a given stage, is dependent on the accumulated precipitation up to that stage.

3. APPLICATION

Developing Weather-Yield Functions

Data were available to develop four different weather-yield functions (WYF) and test them by application to winter wheat yields in Kansas for the period 1981 through 1988. The four functions represented a 2x2 factorial experiment in which size of the observational unit and coverage (experimental plots over the Great Plains vs. ASD's over the state of Kansas) was one factor and type of variables to measure moisture effects (precipitation vs. simulated evapotranspiration) was the other factor. We refer to these four weather-yield functions as PLOT-PR, ASD-PR, PLOT-ET, and ASD-ET.

The data set for the PLOT WYF's consisted of yields and weather data from 966 location-years in the Great Plains beginning as early as 1920 and running through 1980. The ASD WYF's were developed from 279 USDA-estimated yields for the nine Kansas ASD's over the period 1950 through 1980 using weather data from two stations per ASD. In choosing weather variables (X_i's in Eq. [1]), we assumed that yields were reduced by weather conditions that were: (1) too dry, especially in arid climates, (2) too wet, especially in humid climates, (3) too hot or too cold in temperate zones. Accordingly, we used combinations of variables, in the WYF's in Table 1, that would express these meteorological conditions. For example, the PR-type WYF's have concave downward quadratic functions of precipitation in the establishment stage. These functions produce peak yields at

10.5 inches and 8.1 inches, for the PLOT-PR and ASD-PR models, respectively. For ET-type WYF's both ET (evapotranspiration) and XPR (excess precipitation) variables were needed to effect decreased yields during dry periods or during excessively wet periods. The linear temperature terms (TN and TX) carried negative signs indicating reduced yields with increasing temperatures. The AP*TR term in the vegetative stage indicated that yields were reduced in the drier western Kansas ASD's if the daily temperature range was high (associated with clear weather) and in the more humid eastern ASD's if the range was low (associated with cloudy weather). There were no terms to show decreased yields under extremely cold conditions which cause crown injury in the winter or head injury later in the season, especially if close to anthesis. These weather events can do damage within short periods of time (one or two days). Our attempts to define temperature thresholds below which damage occurs have not been successful.

Estimates of the β_i's (Eq. [1]) for the four WYF's are shown in Table 1. All coefficients were significant at the 10% level, with most significant at a 1% level or less. The coefficients of temperatures were consistent across all four WYF's and were more alike for similar data source (Plot or ASD) than type of moisture terms (PR or ET). For PR-type WYF's, the coefficients were enough alike for Plot and ASD data sources to give added credence to their mathematical forms as expressions of moisture impacts. For ET-type WYF, there was consistency in the coefficients of first-order and second-order terms during the differentiation stage. For the remaining stages, the impact of ET and XPR terms varied by stages when PLOT and ASD-type models were compared.

Testing Yield Models

For a complete yield model, the γ_j's and δ_j's in Eq. [1] were estimated when the WYF's were developed using Kansas ASD's. To apply PLOT-type WYF's to the same data, it was necessary to estimate the γ_j's and δ_j's in Eq. [1] by regressing (Yields - WYF) values on the Z_j and $Z_j t$ quantities for the 1950-80 period. The result was four yield models for predicting ASD yields for 1981 through 1988. The means of residuals over the test years for each ASD are shown in Table 2 and the standard deviations in Table 3.

Results in Tables 2 and 3 are not consistent enough to favor one type of WYF over another to explain weather variation in wheat yields. Means over eight residuals, shown in Table 2, tend to be positive, but this is mostly due to underestimates of the δ_j's. Had we refitted the models each year using data up to a test year, the positive bias could have been reduced. The size of the standard deviations (Table 3) was partially due to some weather events not included in the WYF models. A late freeze on May 11, 1981 reduced the weight of grain per head and late-fall planting in 1985, followed by a very early onset of dormancy, lead to low tiller counts for the 1986 harvest. The remaining unexplained variation was due to factors not included in the model, such as losses from diseases, weeds, and insects.

4. SUMMARY AND CONCLUSIONS

We have discussed elements of model building to measure the impact of weather on crop yields. These elements concerned selection of observational units, observational periods within a season, and mathematical forms of weather variables. When applied to winter wheat, the mathematical forms were chosen to produce yield reductions when conditions were too dry, too wet, or too hot. We were not able to find variables that showed yield reductions from cold injury. That problem is complicated by a number of factors. One, there is a need for precise definition of temperature thresholds, and length of time when temperatures are below thresholds, for injury to occur. Two, injury can appear within a short time interval (e.g., overnight) and the threshold varies with maturity. Three, below normal temperatures, down to the point of injury, are generally favorable to yields.

We tested four different weather-yield functions (WYF's), as yield predictors, over an eight-year period in Kansas. The four WYF's represented a 2x2 factorial arrangement of two types of data sets used for model development (experimental plot yields over a multi-state region vs. farm yields over Agricultural Statistics Districts in Kansas) and two types of variables to represent moisture conditions (quadratic functions of precipitation, some of whose coefficients were functions of antecedent precipitation vs. a combination of variables measuring simulated evapotranspiration and excess moisture). The results from the test for Kansas suggest there was no significant gain in precision of large-area yield estimates when simulated evapotranspiration amounts were used in place of precipitation-type variables to express moisture conditions. Thus, for yield forecasting, precipitation-type variables may be preferable to simulated-evapotranspiration amounts because of their mathematical simplicity. We also found that functions developed from the broader coverage (plot yields from a multi-state area) may have had a slight edge in precision.

REFERENCES

1. Baier, W. 1973. Crop-weather analysis model: Review and Model Development. J. Appl. Meteorol. 12:937-947.

2. Doraiswamy, P. C., T. Hodges, and D. E. Phinney. 1979. Crop yield literature review for AgRISTARS crops corn, soybeans, wheat, barley, sorghum, rice, cotton, and sunflowers. Lockheed Electronics Company, Inc. Tech. Report. 1830 NASA Road 1, Houston, TX 77058.

3. Feyerherm, A. M. and G. M. Paulsen. 1981. Development of a wheat yield prediction model. Agron. J. 62:27-29.

4. _____ 1984. Development of a weather-yield function for winter wheat. Agron. J. 78:1012-1017.

5. Haun, J. R. 1974. Prediction of spring wheat yields from temperature and precipitation data. Agron. J. 66:405-409.

6. NASA. 1979. Proceedings of Technical Sessions. Vol. I and II. The
    LACIE Symposium. National Aeronautics and Space Administration.
    Houston, TX.

7. Strommen, N.D., C. M. Sakamoto, S.K. LeDuc, and D. E. Umberger. 1979.
    Development of LACIE CCEA-I weather/wheat yield models. Proc. of
    Technical Sessions. Vol. I. The LACIE Symposium. NASA Johnson
    Space Center. Houston, TX.

8. Thompson, L.M. 1969. Weather and technology in the production of
    wheat in the United States. J. Soil Water Cons. 24:219-224.

Table 1. Comparison of estimates of β_i's (Eq. [1]) for four WYF's using two different data sources (Plots vs. ASD) and two different expressions of moisture (precipitation vs. evapotranspiration).

Stage of Development (Assigned length)	Weather Variables[†]	Estimates of β_i's			
		Plot-ET	ASD-ET	Plot-PR	ASD-PR
		Bushels/Acre/Unit of weather variable			
Establishment (60 days in Fall)	TX	-0.359	-0.585	-0.341	-0.414
	ET	0.94	0	----	----
	XPR	-0.41	0	----	----
	PR	----	----	2.23	3.02
	PR²	----	----	-0.106	-0.187
Dormancy (Variable length)	TN	-0.270	-0.203	-0.255	-0.215
	ET	0	1.32	----	----
	PR	----	----	0.90	1.15
Vegetative (40 days in early spring)	TN	-0.275	-0.167	-0.328	-0.228
	AP*TR	0.013	0.021	0.020	0.027
	CP*PR	----	----	0	-0.139
	PR²	----	----	0.120	0.204
Differentiation (40 days in mid-spring)	TN	-0.414	-0.167	-0.427	-0.173
	ET	5.70	5.11	----	----
	ET²	-0.482	-0.441	----	----
	XPR	-0.68	0	----	----
	PR	----	----	4.77	2.62
	CP*PR	----	----	-0.355	-0.195
	PR²	----	----	-0.374	-0.243
	CP*PR²	----	----	0.0257	0.0181
Reproductive (20 days in late spring)	TN	-0.540	-0.322	-0.495	0
	ET	1.68	0	----	----
	XPR3	-0.69	-0.48	----	----
	PR	----	----	1.44	0
	PR²	----	----	-0.197	0
	CP*PR²	----	----	0	-0.0034

[†] TX = average daily maximum temperature (°F),

TN = average daily minimum temperature (°F),

TR = average daily range of temperature (°F),
PR = total precipitation (inches),
CP = cumulative precipitation (inches) to beginning of stage,
AP = average annual long-term precipitation minus 30 (inches) for locations associated with observational unit,
ET = total simulated evapotranspiration (inches) using Baier-Robertson soil moisture budget (Baier, 1973),
XPR = precipitation (inches) in excess of that needed for a full 10-inch soil moisture budget,
XPR3 = precipitation (inches) in excess of three inches.

Table 2. Comparison of mean residuals for four models over eight test
years in Kansas.

	Type of WYF			
ASD	Plot-ET	ASD-ET	Plot-PR	ASD-PR
	----------------Bushels/Acre----------------			
Northwest	1.3	1.5	-1.3	0.8
West Central	1.3	1.9	0.6	2.7
Southwest	-6.3	-3.7	-4.9	-2.3
North Central	4.1	4.6	4.8	6.2
Central	1.6	1.7	0.9	1.3
South Central	2.6	3.9	1.5	2.8
Northeast	4.5	4.6	3.7	2.8
East Central	4.4	5.7	2.1	3.5
Southeast	3.7	5.3	2.8	4.3
Weighted Avg.[i]	1.0	2.1	0.5	2.0

[i] Weights based on harvested acreage from 1971 to 1975.

Table 3. Comparison of standard deviations of residuals for four models
over eight test years in Kansas.

	Type of WYF			
ASD	Plot-ET	ASD-ET	Plot-PR	ASD-PR
	----------------Bushels/Acre----------------			
Northwest	8.0	6.7	5.7	5.8
West Central	6.8	6.8	6.9	7.0
Southwest	5.5	4.9	5.2	4.8
North Central	5.6	5.2	5.4	5.9
Central	2.6	2.9	2.4	4.2
South Central	1.7	2.8	2.0	3.4
Northeast	6.9	8.0	7.7	7.3
East Central	4.0	5.1	3.1	2.8
Southeast	2.3	3.7	2.8	4.5
Weighted Avg.[i]	4.4	4.6	4.2	4.8

[i] Weights based on harvested acreage from 1971 to 1975.