EVALUATION OF CROP/WEATHER YIELD MODELS: A DEVELOPER'S VIEWPOINT

Arlin M. Feyerherm, Kansas State University

INTRODUCTION

It is very fitting that this session on methods for evaluating crop yield models be dedicated to the work and memory of H.O. Hartley. His contributions to the theory and applications of statistics included those devoted to the Large Area Crop Inventory Experiment (LACIE). This experiment, sponsored by the tri-agency group, NASA-NOAA-USDA, was designed to demonstrate how satellite, weather instrumentation, and computer technology could be combined to estimate wheat production on a country-wide basis. Dr. Hartley contributed to the sampling methodology employed and peer reviews of results. He posed questions, summarized progress, and defined remaining problems in his own inimitable manner. He will be missed.

It was during the LACIE program that a number of us had an opportunity to become involved in problems surrounding the development of weather/crop yield models. Seven years ago, at the beginning of LACIE, we asked many questions about such models not the least of which were:

(1) How much of the yield increases of the past 25 years should be attributed to weather and how much to technology and could they be separated out?

(2) If weather and technological effects could be separated, could we further subdivide technological effects into the contributions due to improved varieties, fertilizers, herbicides, pesticides, cropping practices, and others? What if one simply used "time" as a surrogate for their combined effect?

(3) Relative to the weather contribution would it be possible to determine a mathematical relation of yield to weather events that would hold universally for the major wheat growing areas of the world? --- or would it be necessary to build a mini-model for each region? If the latter were the case, what about sparse data sets for parameter estimation in foreign areas?

Seven years of work has produced insight and understanding but questions of approach, methodology, variables to use, mathematical form, size of observation unit for model development and prediction, and others still remain.

In this paper we discuss development and testing of a crop/weather model to predict spring wheat yields on a large-area basis with special application to the state of North Dakota. Following this, we make a few comments on evaluation.

SPRING WHEAT YIELD MODEL

Data Set for Development. We approached the problem of developing a spring wheat yield model from the viewpoint that it was possible to develop a universal model but to determine the relation would require:

(1) a unit of observation be selected to maximize the number of primary sources of variation (weather events, applied nitrogen, variety planted, cropping practice, soil factors, pests, hail and other episodal events) about which historical information is available,

(2) the data set to estimate parameters should contain values of weather variables with a wide range of variation,

(3) standardizing some of the variables so that they carry the same weight at all locations.

These conditions led us to use replicated plot yields of varietal performance trials as our response variable for model development. Such trials have been conducted at experiment stations in the United States for 50-60 years at some locations. The stations were located throughout the spring wheat area of Montana, North Dakota, South Dakota, and Minnesota. The breadth of data over many years and across climatic zones helped to insure a range of variation of values of weather variables not likely to be exceeded either in future years or in other spring wheat areas of the world.

Standardizing Variables. The following standardization procedures, though somewhat crude helped to reduce experimental error and biases.

(1) Yields from different varieties were adjusted to a "standard" or "base" variety.

(2) Weather variables were measured within simulated stages of plant development rather than within specified weeks or months.

(3) Yields were culled to remove those reduced by disease, hail, pests, and other nonmodeled factors. Further reduction in experimental error was accomplished through use of simulated evapotranspiration amounts rather than precipitation per se to measure effects of drought.

Form and Substance. Our yield/weather relation took on the form shown in Eq. 1. Coefficients were determined from n=249 vectors of observations.

(Eq. 1) Standardized Plot Yield = Location constant + 1.0 (CN-P) - 0.08 $(PR\text{-}PJ)^2$ + 3.8 (ET-JD) - 0.23 $(ET\text{-}JD)^2$ - 0.25 (TX-JF) - 0.36 (TX-FH) - 0.01 (TX-HM) * (PR-HM) - 0.003 $(TX\text{-}HM)^2$ - 7.0 (TX-MD) + 0.04 $(TX\text{-}MD)^2$ + 0.065 (NITROGEN).

where the letters after the hyphen designate stimulated phenological stages; that is, P = planting, J = jointing, F = flag leaf, H = heading, M = milk, D = dough.

The model shown in Eq. 1 for a standardized plot yield indicates the deleterious effect of high daytime temperatures (TX terms) from the jointing-to-dough stages (roughly through June and July for most of the spring wheat region). Anything less than a full soil moisture profile at planting (CN-P) and too much precipitation between planting and jointing (PR-PJ) also reduces yields. The convex quadratic function for evapotranspiration (ET) indicates too little or too much moisture can lead to less than optimum yields.

The effect of applied nitrogen is relatively small for spring wheat with an estimated 10 pounds required to provide an additional 0.65 bushel per acre.

We consider this part of the model to be universal and applicable to other areas of the globe where wheat is planted in the spring.

Large-Area Estimates. To estimate yields on a state-wide basis, we start with the model shown in Eq. 2.

(Eq. 2) Yield (state-wide) = $\alpha + \beta * WX + IT + \gamma * UIT$ + Error

where

α = location constant for given state,

WX = weather effect = weather related terms in Eq. 1,

IT = identified technology = DYA + .065 * NI + FALINC,

UIT = unidentified technology
= (t - 55) if $55 \leq t \leq 64$,
= 9 if $t > 64$,

where

DYA = average differential yielding ability of a set of varieties,

NI = amount of applied nitrogen,

FALINC = yield increment due to fallowing.

We assume that a yield is the sum of a constant peculiar to a given state, weather effects, technological components identified with agronomic factors, non-identified components for which we are forced to use "time" as a surrogate, and random error.

We subtracted the identified technology effects from USDA yield estimates and regressed the differences on the weather and "time" variables. The model, generated by fitting n=15 (1955-69) vectors of observations for North Dakota, is shown as Eq. 3.

(Eq. 3) Yield = -7.15 + 0.75 (WX) + [DYA + 0.65 (NI) + FALINC] + 0.77 $(t - 55)^+$

where

$(t - 55)^+ = t - 55$ if $55 \leq t \leq 64$,
$= 9$ if $t > 64$.

The coefficient of our weather function (WX) is less than unity and represents a scaling down of the contribution of weather to yield variation on a plot basis relative to that found on a large-area basis. The other point of interest is that the unexplained technology contribution to yield appears to extend till about 1964 and then is not significantly different from zero. We suspect that the yield change represented by a 6.9 bushel/acre increase between 1955 and 1964 is largely due to increased use of herbicides since competition from weeds can be a major contributor to yield loss.

APPLICATION OF SOME TEST CRITERIA

A Look at Some Test Data. The seven seasons from 1970-76 were chosen as a test period. Now a sample of size 7 is certainly not large enough to accept a model as being a superb "performer". However, it may be large enough to detect one or more flaws if the "right situation" occurs to produce outliers. Due to the small sample size it is especially important to scrutinize the test data set.

Table 1. Comparison of Ranges of Values for Development and Test Data Sets.

Variables	Development Set	Test Set
CN-P (inches)	0.2 - 9.7	0.9 - 10.0
PR-PJ (inches)	0.0 - 8.3	0.2 - 4.4
ET-JD (inches)	1.3 - 9.9	1.0 - 8.1
TX-JF (degrees)	64 - 90	73 - 90
TX-FH (degrees)	68 - 90	71 - 89
TX-HM (degrees)	70 - 94	73 - 92
TX-MD (degrees)	71 - 96	75 - 96
PR-HM (inches)	0.0 - 5.0	0.0 - 4.3
WX (bu./acre)	14.7 - 32.2	16.2 - 31.4
Yield (bu./acre)	15 - 30	20 - 32

In Table 1 we show a comparison of ranges of values for weather variables between the developmental and test sets. With the exception of CN-P and ET-JD, where a small overextension exists, all values in the test set are inside those of the development set. For the response variable we have a range of 12 bushels/acre versus that of 15 for the development test. While a wider range than 12 bushels/acre would be desirable to test a model's ability to respond to large year-to-year variation in yield, it may be large enough to detect if a model is little more than a constant plus random error.

Table 2. Model and USDA Yields for North Dakota.

YEAR	MODEL	USDA ESTIMATE	DIFFERENCE	
1970	24.5	23.6	+ 0.9	
1971	29.8	31.8	- 2.0	
1972	31.7	28.9	+ 2.8	
1973	24.6	27.5	- 2.9	
1974	20.8	20.4	+ 0.4	
1975	27.9	25.9	+ 2.0	
1976	28.6	24.7	+ 3.9	
Ave.	26.8	26.1	+ 0.73	= Bias
			2.40	= RMSE

Bias and RMSE. In Table 2 we show a year-by-year comparison of model generated values to estimates made by the USDA. The estimate of the bias is small and nonsignificant while the root-mean-square error (RMSE) of 2.4 bushels/acre is, coincidently, equal to the standard error of estimate for the n=15 observations used for the development set.

Analysis of Technological Contributions. As each new year of test data was considered, a test was made to determine if a second line segment starting in 1964 in the unidentified technology component might have a non-zero slope. No evidence for a non-zero slope was found so the factors of improved varieties, amount of applied nitrogen and changes in cropping practices (more wheat planted on fallowed ground) explained technological yield increases after 1964. We think that such an analysis of technological gains is an important part of model building and reduces dependence on use of "time" as a surrogate.

In conclusion we have no quarrel with the criteria that Wendell Wilson and Jeanne Sebaugh set forth in the preceding paper. In fact, we think they are very reasonable and helpful both to make initial judgements about individual models and to compare competing models. However, we think additional attention should be given to two concepts in yield/weather models, namely: (1) universality of a model and (2) extent to which terms in a model explain technological gains without reliance on "time" as a surrogate.