FORECASTING CORN EAR WEIGHTS FROM DAILY WEATHER DATA

Fred B. Warren
FORECASTING CORN EAR WEIGHTS FROM DAILY WEATHER DATA

by

Fred B. Warren
National Agricultural Statistics Service
U. S. Department of Agriculture
Washington, D. C.

ABSTRACT

Statistical models were developed to predict the State average grain weight per ear using daily temperature and precipitation data, recorded from May 1 through late July. The required daily weather data was successfully obtained in an operational test of these models for ten major corn producing States in 1988. Relative forecast errors of ear weight averaged almost one-third smaller than those from a regular survey. Additional refinements of the models to make them more responsive to abnormally early adverse weather, as in 1988, are underway.

Key words: modeling, weather, corn ear weight

INTRODUCTION

The National Agricultural Statistics Service (NASS) conducts objective surveys of corn yields for the ten major corn producing States. Fields for these surveys are selected systematically, with probabilities proportionate to their expanded sizes, from fields enumerated on the NASS June Enumerative Survey. Observations are taken in two randomly selected plots in each field. The first observations of each crop season are taken the last week of July. Additional observations are taken at monthly intervals, until harvest. The computational advantage of this sampling procedure is that the observations from each field have equal weight in the State summaries. The data from each monthly survey is used in developing the official predictions of yield published on about the 10th day of the next month.

Final pre-harvest observations are obtained in each sampled field as it is about to be harvested by the farmer. These observations include both counting and weighing ears in the sample plots, and shelling and weighing a sub-sample of the ears. These observations are used to estimate the average number of ears per acre in the field, and the average weight of grain per ear. Sample plots in a sub-sample of the sample fields are gleaned after harvest to provide an estimate of harvesting loss. The final average yield for the State is then estimated as
(1) \( \text{Yield} = \frac{\sum_{i=1}^{n} (E_i \times W_i - L_i)}{n}, \)

where,
- \( E_i \) is the estimated number of ears per acre for the \( i \)-th sample field,
- \( W_i \) is the estimated weight of grain per ear for the \( i \)-th sample field, and
- \( L_i \) is the estimated harvest loss for the \( i \)-th sample field. The current year average is used for fields not subsampled at harvest.

Before the field is harvested, the number of ears and weight of grain per ear at harvest are forecast from counts and measurements of ears from the most recent monthly survey. The harvest loss is assumed equal to the average harvest loss for the five previous years. If ears have not developed enough for meaningful ear measurements, as is usual in late July, a five year running average weight of grain per ear for the State is used as the forecast weight of grain per ear.

The difficulty with using the five year average for the July forecast is that the State average weight of grain per ear varies considerably from year to year (Table 1). Consequently, the forecast error of the August 1 predicted weight of grain per ear can be quite large. Since most of the annual variation in average weights of grain per ear results from fluctuations in weather, a desirable solution would be to develop models that consider the impact of weather before August 1 on the final average weight of grain per ear. (An obvious example of the effects of weather on ear weight is that of high temperatures on pollination. An often more serious effect of adverse weather is to retard growth of the embryo ear shoot. This may reduce the number of potential kernel sites, or even cause the plant to abort the primary ear shoot.)

**DAILY WEATHER DATA**

The National Weather Service (NWS), National Oceanic and Atmospheric Administration, collects daily observations of minimum and maximum temperature and precipitation from volunteer weather observers in each State. Traditionally, each observer submits a combined report to a State NWS office, on Saturday, for the preceding seven days. However, observers in several States now use Touch-tone telephones to report daily, to a regional computer. Eventually, all reports are assembled on the NWS National Service Center computers at Asheville, NC.

A major concern of NASS, and the principal reason for the 1988 study, was to determine if this stream of weather data could be intercepted and made available to NASS for its use in preparing predicted yields for the August 1 Crop Report (published on about August 10.)
MODEL DEVELOPMENT

Three somewhat different regression models were developed for each State. These models predicted the average weight of grain per corn ear for the entire State. Not more than three regressor variables were used for each model. These models were developed under the following constraints for evaluative purposes:

1. One model would use no soil moisture related data,
2. One model would use at least one soil moisture variable, and
3. One model would use only data for time periods that would be available even in a late growing season.

The decision to predict average weight of grain per corn ear at the State rather than at the field level was dictated by the following factors.

1. Weather data was never collected for the sample objective survey fields.
2. The dependent variable, average weight of grain per ear at harvest, was available only for individual samples since 1979. However, State average estimated weights of grain per ear were available from 1967. (Although the corn objective yield surveys began earlier, there were major procedural changes in the survey in 1966. Therefore, data from before 1967 was not used in the development of the models. Data for 1970 was also excluded, to delete any anomalous effects of the Southern corn leaf blight epidemic in that year.)
3. The daily soil moisture procedure for corn, from Blaney and Criddle, requires knowledge of the planting and silking dates for the crop. Again, this information had not been collected for the fields sampled by the objective yield survey. However, information as to the progress of planting and development of corn within the State is compiled by the NASS State Statistical Offices (SSO) as part of their Weekly Crop Reports. This information was available for most, but not all, States from 1967 until now, and could be used to construct median (50 percent completion) planting and silking dates for each State and crop year. However, there were two States where data for years before 1973 had been lost.
4. The most recent year used in developing the models was 1986. Therefore, the number of years (observations) for any one State varied from 13 to 19 years. To avoid over-fitting the data, the maximum number of variables allowed in any model was three.

The locations of weather stations within a particular State usually are not optimized with respect to the distribution of corn. To reduce the effect of this maldistribution, straight averages of daily observations of minimum and maximum temperatures and precipitation were first aggregated.
for each Agricultural Statistics District (ASD) within the State. NASS estimates of corn planted in each ASD were then used to compute weighted State averages. The weighted State average data was then used to compute daily moisture consumption and soil moisture balances.

The daily estimates of soil moisture and temperature data were averaged over seven-day intervals, from 28 days before the median silking date and through August 1. The types of temperature variables used were:

1. Maximum daily temperatures in excess of 88°F,
2. Maximum daily temperatures below 88°F,
3. Mean daily temperature, and
4. Difference between minimum and maximum daily temperatures.

The significance of these variables lies in the following relationships:

1. Photosynthesis is greatly reduced after in-field temperatures reach a level of 85 to 90°F (30 to 32°C),
2. Wide differences between daily minimum and maximum temperatures are associated with low levels of soil moisture, and
3. Even in States where daily maximum temperatures seldom reached 88°F during this period, there was a negative relationship between maximum daily temperatures and grain weight per ear.

The effect of soil moisture is not easily defined, but appears to be confounded with temperature. In particular, high levels of soil moisture in conjunction with relatively low temperatures are detrimental. This is also true of low levels of soil moisture in conjunction with high daily temperatures. However, there was a wide intermediate range where the computed level of soil moisture appeared to have little effect on the grain weight per ear. The rate of change in soil moisture, from one seven-day period to the next, often appeared to affect weight of grain per ear more than did the actual level of soil moisture.

Variables for these models were selected from all possible 1, 2, and 3 variable combinations of the seven-day averages of the temperature and soil moisture variables, and rates of change in soil moisture from one period to the next. This selection was made in the following manner.

1. The variables were ranked by their linear correlations with the State average estimated weight of grain per ear.
2. 'Acceptable' two variable models for predicting final weight of grain per ear were identified by testing each variable in combination with every lower ranking variable. The acceptable models were those combinations where the partial Type I sum of squares from the additional variable was significantly larger than zero ($\alpha = 0.05$).
3. Acceptable three variable models for predicting final weight of grain per ear were identified by testing each acceptable two variable model in combination with every other variable. Acceptable three-variable models were those where (a) the partial Type I sums of squares from the additional variable was significantly larger than zero, (b) the partial type III sums of squares from each variable also were significantly different from zero, and (c) excessive collinearity did not appear to be a problem.

4. The acceptable two or three variable combinations were then ranked by their multiple coefficients of correlation (adjusted) and classified as to whether or not they included any soil moisture variables, and if they included only variables from time periods that should be available by late July, even in a late crop year.

5. The intended strategy was to select the highest ranking combination of each type. In practice, there were States (Michigan, Nebraska, Wisconsin) where all acceptable combinations included soil moisture variables, and other States (Missouri and Ohio) where no acceptable combinations included soil moisture variables. In such cases, the set of three models was filled out with the best remaining combinations.

The identification and ranking of acceptable two and three variable combinations (steps 1-4) was done by a special SAS (Statistical Analysis System) program. SAS PROC REG was used in identifying combinations with excessive collinearity.

THE 1988 STUDY

The original purpose for the 1988 study was to show that the daily weather data required by the models could be gathered, in four States, in time for the NASS August 1 Crop Report. However, as the 1988 drought progressed, the study was expanded to all ten of the corn objective yield States. This was with the expectation that the weather models would provide more accurate forecasts of the final average weight of grain per ear than would the conventional procedures.

The NASS SSO's were successful in obtaining daily minimum and maximum temperatures and precipitation from cooperating weather stations through Friday, July 22. Because the models are linked to the maturity of the crop and the growing season was early, this provided enough data for the models. Summary programs were executed on August 2. One feature of the summary program was that any model predictions that fell outside (below) the historic range of objective survey final ear weights for that State were censored. The censoring procedure was to substitute the upper or lower limit, as appropriate, of the historic range for the model predictions. The model predictions of average weight of grain per ear were reviewed in the NASS Washington headquarter. Objective survey predicted yields were
then adjusted by the ratio [average of the model predicted (censored) ear weight] / [objective survey predicted ear weight]. The adjusted objective yield predictions were sent to the SSO's by August 4 and were used by the NASS Crop Reporting Board in preparing the August 1 Crop Report.

The regular corn objective yield survey procedures were followed throughout the season, culminating with 'final' State estimates of average weight of grain per ear. Record low average weights of grain per ear were observed in the States of Illinois, Iowa, Michigan, and Wisconsin. Also, near-record low average weights of grain per ear were observed in all other States except Nebraska and Missouri.

The daily weather models produced more accurate predictions of grain weight per ear than did the regular survey procedure in six States, and less accurate predictions in four (Figure 1). Even with the less accurate prediction in four States, the forecast error of ear weight from the weather models for the entire region was 38 percent smaller than from the regular survey. The daily weather model predictions were very close (within five percent) to the final estimate in Indiana and Michigan. However, the model average forecast errors were greater than 15 percent in Illinois, Minnesota, Missouri, Ohio, and Wisconsin. (The "*'s" after Illinois, Indiana, and Michigan show that the actual prediction from at least one model was less than the previous low average weight of grain per ear. In these cases, the prediction was censored to the previous low estimate for that State.)

QUESTIONS!! -- AND ANSWERS??

A possible explanation for the failure of the weather models to do better in the remaining States hinges upon the timing of the 1988 drought and the way in which variables were selected for the weather models. Although weather from as early as 28 days before the median silking date was considered in developing the models, no variables were selected for time periods earlier than 14 days before the median silking date for most States. This could have been because historically, from 1967 through 1986, the most extreme weather conditions occurred after that point of time. However, the most extreme temperatures in 1988 in several States occurred slightly earlier in the season, 28 to 15 days before silking. (This would have been early in the development of the embryo ear shoot!) Therefore, then one can ask "Does this set of weather models relate to the real world?" Or, "Do they only fit the observations, from 1967 through 1986?" Or "If the relationships between final grain weight and the designated weather characteristics are valid, are they consistent over a slightly earlier period of time?"

If the answer to the last question is "Yes!", then a modified model could use the weather data from the week(s) of most extreme temperatures in 1988 with the regression coefficients from the original model. This was done to create modified forecasts of grain weight in six States -- Iowa, Minnesota, Missouri, Nebraska, South Dakota, and Wisconsin (Figure 2). If these modified forecasts were substituted for the original weather model predictions, the relative reduction in the forecast error for the entire
10-State region would increase from 33 to 69 percent. This reduction would have been even greater if the model prediction for Illinois was not censored to equal the previous low ear weight for that State. Therefore this procedure could have great potential for improving the forecast ear weight models. The problem lies in developing a computerized procedure which could produce equivalent results without subjective human intervention.

There is still a problem in developing weather models for ear weight for Ohio. During the base period (1967-86), there was relatively little year to year variation in average ear weight in that State. Also, the variation which did appear seems to have resulted from adverse weather after, not before, August 1.

**SUMMARY**

Regression models that use accumulations of daily weather data to predict final average ear weight have been developed for each of the ten major corn producing States. These models were tested during the July 1988 survey period.

1. Daily weather data needed by the weather models and reported by the NWS volunteer weather observers can be obtained as needed.

2. The models did quite well in three States, but their usefulness in other States was hindered by their orientation towards pre-determined time periods.

3. The weather models should be reviewed to determine if it is possible to objectively identify and use periods of 'most critical' weather, regardless of maturity.

<table>
<thead>
<tr>
<th>STATE</th>
<th>Median</th>
<th>Range</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>C.V. %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illinois</td>
<td>0.370</td>
<td>0.135</td>
<td>0.366</td>
<td>0.035</td>
<td>9.9</td>
</tr>
<tr>
<td>Indiana</td>
<td>0.362</td>
<td>0.119</td>
<td>0.356</td>
<td>0.032</td>
<td>9.1</td>
</tr>
<tr>
<td>Iowa</td>
<td>0.368</td>
<td>0.136</td>
<td>0.359</td>
<td>0.033</td>
<td>9.1</td>
</tr>
<tr>
<td>Michigan</td>
<td>0.293</td>
<td>0.087</td>
<td>0.293</td>
<td>0.021</td>
<td>7.3</td>
</tr>
<tr>
<td>Minnesota</td>
<td>0.309</td>
<td>0.120</td>
<td>0.311</td>
<td>0.030</td>
<td>9.7</td>
</tr>
<tr>
<td>Missouri</td>
<td>0.374</td>
<td>0.225</td>
<td>0.363</td>
<td>0.052</td>
<td>14.2</td>
</tr>
<tr>
<td>Nebraska</td>
<td>0.369</td>
<td>0.102</td>
<td>0.375</td>
<td>0.031</td>
<td>8.3</td>
</tr>
<tr>
<td>Ohio</td>
<td>0.348</td>
<td>0.103</td>
<td>0.338</td>
<td>0.029</td>
<td>8.6</td>
</tr>
<tr>
<td>S. Dakota</td>
<td>0.318</td>
<td>0.149</td>
<td>0.305</td>
<td>0.043</td>
<td>14.0</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>0.312</td>
<td>0.111</td>
<td>0.312</td>
<td>0.026</td>
<td>8.3</td>
</tr>
</tbody>
</table>
Figure 1. Relative Forecast Errors, Weather Model Predictions, August 1, 1988

Weather model. Symbol used is 'w'.
Regular objective survey. Symbol used is 'r'.

State

<table>
<thead>
<tr>
<th>Illinois*</th>
<th>+</th>
<th>w</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indiana*</td>
<td>+</td>
<td>w</td>
<td>r</td>
</tr>
<tr>
<td>Iowa</td>
<td>+</td>
<td>w</td>
<td>r</td>
</tr>
<tr>
<td>Michigan*</td>
<td>+</td>
<td>w</td>
<td>r</td>
</tr>
<tr>
<td>Minnesota</td>
<td>+</td>
<td>r</td>
<td>w</td>
</tr>
<tr>
<td>Missouri</td>
<td>+</td>
<td></td>
<td>rw</td>
</tr>
<tr>
<td>Nebraska</td>
<td>+</td>
<td>r</td>
<td>w</td>
</tr>
<tr>
<td>Ohio</td>
<td>+</td>
<td>r</td>
<td>w</td>
</tr>
<tr>
<td>S. Dakota</td>
<td>+</td>
<td>w</td>
<td>r</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>+</td>
<td>w</td>
<td>r</td>
</tr>
</tbody>
</table>

Relative Percent Forecast Error (%)

* - Censored weather model predictions
Figure 2. Relative Forecast Errors, Modified and Regular Weather Model August 1 Predictions, 1988

Weather model. Modified weather model.
Symbol used is 'w'. Symbol used is 'm'.

<table>
<thead>
<tr>
<th>State</th>
<th>Illinois</th>
<th>Indiana</th>
<th>Iowa</th>
<th>Michigan</th>
<th>Minnesota</th>
<th>Missouri</th>
<th>Nebraska</th>
<th>Ohio</th>
<th>S. Dakota</th>
<th>Wisconsin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>w</td>
<td>m</td>
<td>m w</td>
<td>w</td>
<td>m w</td>
<td>m</td>
<td>m w</td>
<td>m</td>
<td>m w</td>
<td>m w</td>
</tr>
</tbody>
</table>

Relative Forecast Error (%)

BIBLIOGRAPHY
