United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research and
Applications
Division

SRB Staff Report
Number SRB-90-05

March 1990

AN OPERATIONAL TEST USING WEATHER DATA TO FORECAST CORN EAR WEIGHT, 1988

Fred Warren

AN OPERATIONAL TEST USING WEATHER DATA TO FORECAST CORN EAR WEIGHT, 1988, by Fred B. Warren and Gary Keough, Research and Applications Division, National Agricultural Statistics Service, U. S. Department of Agriculture, Washington, D.C. 20250, April 1990. Research Report No. SRB 90-05.

ABSTRACT

Final average weight of grain per corn ear was predicted from weather models for the 1988 August 1 Crop Report for each of the ten States in the National Agricultural Statistics Service's (NASS) corn objective yield program. The daily weather data required by these models was obtained by the NASS State Statistical Offices from the National Weather Service, National Oceanic and Atmospheric Administration. These forecasts were prepared using daily weather data from May 1 through about July 25. For most States, the models was developed from nineteen years (1967-86, less 1970) of final corn ear weights and weather data. Parameters were estimated from data through 1987. On the average, the weather models provided better predictions of final ear weight than did the regular objective yield procedures. However, when farmer reported crop conditions (non-probability sample and judgment type questions) are used to adjust the regular objective yield indications, the adjusted objective yield models outperformed the weather models. This was because the weather models did not fully consider the effects of extremely hot dry weather in late May and June of 1988. Further research into weather models is recommended. This would include the use of early season weather data in forecasting the average number of ears per stalk as well as average weight of grain per ear.

KEY WORDS

Weather modeling, corn ear weight modeling.

ACKNOWLEDGMENTS

The authors thank the staff members of the ten NASS State offices for their invaluable efforts in collecting the weather data required for this study.

Washington, D.C. April 1990

TABLE OF CONTENTS

An Operational Test Using Weather Data
To Forecast Corn Ear Weights

SUMMARY

Final average weight of grain per corn ear was predicted from weather models for the 1988 August 1 Crop Report for each of the ten States in the National Agricultural Statistics Service's (NASS) corn objective yield program. The daily weather data required by these models was obtained from the National Weather Service (NWS), National Oceanic and Atmospheric Administration by the NASS State Statistical Offices (SSO's).

For the 1988 crop season, predicted ear weights were higher than the final objective yield survey estimates for most States. However, the forecast errors were smaller than the regular objective yield survey forecast errors in all States except Nebraska. For the region as a whole, the average forecast error for weight of grain per ear from weather models for August 1, 1988 was almost 40 percent smaller than the average error from the regular August 1 objective yield survey procedures. However, when farmer reported crop conditions (non-probability sample and judgment type questions) are used to adjust the regular objective yield indications, the adjusted objective yield models outperformed the weather models.

The performance of weather models derived from data through 1986 and of models derived from data through 1983 only was evaluated for the following time periods: 1979-83, 1984-86, and 1987-88. Models derived from data through 1986 consistently had smaller Root Mean Square Forecast Errors (RMSFE) for both the 1979-83 and 1984-86 time periods. However, there was no appreciable difference in the performance of the two sets of models during 1987 and 1988.

At the end of the data collection period, the SSO's were asked to comment on the adequacy of the weather data collection, key entry, and editing instructions. Many suggested changes have been adopted.

Further research is recommended. This research should review the possible importance of the timing of critical weather events. If this review shows that the timing of critical weather is not as critical as indicated by the models used in this study, then further research should consider both:

a. the use of weather models to provide better August 1 estimates of the average number of ears per stalk; and
b. the use of weather models to provide better September 1 estimates of final weight of grain per ear.

INTRODUCTION

The original purpose of this study was to show that the weather data needed by the corn ear weight weather forecast models[1/] could be obtained for predicting corn ear weights for the 1988 August 1 corn objective yield summary in four selected States. The study was expanded to all ten corn objective yield States in late June, after an early-season drought had developed over most of the ten States in the Corn Belt.

Since the project was designed to be a quasi-operational test of the weather models, this document includes discussions of:

- (a) data collection procedures,
- (b) adequacy of the weather station network coverage,
- (c) data entry and editing procedures,
- (d) models,
- (e) summary program,
- (f) costs,
- (g) SSO comments, and
- (h) comparisons of predicted values, from the weather models and from the objective yield survey.

All data entry, edit, and summary programs for this project are written in the SAS programming language. These programs can execute on either a mainframe or a personal computer.

DATA COLLECTION

The models require daily reports of precipitation, and minimum and maximum temperatures from about May 1 until late July. The actual dates are determined by the median planting and silking dates in each State, and by reporting and central computer processing schedules. In many States, daily observations are taken from Saturday through Friday. They are mailed on the next day, Saturday, and received by the local NWS office on Monday or Tuesday. In 1988, weather data for any day during the week of July 23-29 was not available until August 1 or 2. The corn objective yield summary programs ran on August 3 or 4, so there was little time for entering and editing data from after July 23. Fortunately, the 1988 growing season was early enough that data for July 24-29 was not needed. In a later growing season, the flexibility inherent in processing on personal computers in the SSO's could permit using more data.

Instructions[2/] for collecting and entering daily weather data into the NASS computer system were sent to the four original SSO's in late April. These instructions:

- (1) assumed that the required weather data could be obtained from existing weather reporting networks[3/];
- (2) included guidelines as to the minimum number of stations desired for each Agricultural Statistics District (ASD);
- (3) provided data entry formats; and

1/ "*Forecasting Grain Weight per Corn Ear on August 1*", Fred B. Warren and Paul W. Cook, USDA/NASS Research Report Number SRB-88-03, February 1988.
2/ See Appendix A.
3/ Previous contacts had determined that such networks did exist.

(4) provided due dates for the completion of data entry and editing.

Data was requested for at least 25 to 30 weather stations in each State. The desired minimum number of weather stations in each ASD varied according to the importance of the ASD. The guidelines called for reports from at least two (2) weather stations in the least important ASD's, and at least five (5) stations in the more important ASD's.

Similar instructions were sent to the six additional SSO's when the project was expanded. Because these SSO's started late, they were instructed to collect and enter the data as quickly as possible.

All SSO's were successful in obtaining daily weather data from the NWS and in entering and editing the daily data in time for the August 1 forecast summary. However, the NWS weather networks did not always have the "desired minimum" number of stations in all ASD's. Further, as seen in Table E-1, Appendix E, there were times when all stations in an ASD failed to report for a particular day. When this happened, that ASD received zero weight in the State average for that day. (NOTE: Weather stations located near ASD boundaries were used only in the ASD's in which they were located -- even though data from these stations would also reflect weather conditions in adjacent ASD's.)

A detailed appraisal of the existing NWS network's ability to meet the guidelines is contained in Appendix E.

DATA ENTRY AND EDITING

The instructions given the SSO's for data entry and running the weather edit program on the Martin-Marietta Data System (MMDS) are listed in Appendix B. Principal features of the edit program were the following:

1. Each record was checked against a master file of weather station names and ASD codes. All records for which the station name and ASD code did not match this master list were rejected.
2. Daily maximum temperatures were identified as the largest of the reported minimum and maximum temperatures.
3. Warning messages were issued if:
   (a) the spread between daily minimum and maximum temperatures was less than an arbitrary value,
   (b) the daily minimum and maximum temperatures were outside arbitrary ranges of values, and
   (c) if the precipitation was greater than an arbitrary value.
4. The edited file was converted to a SAS data set and merged into an existing data set for that State. Any corrections would be merged in, replacing the erroneous data, in a later edit run.

The following changes in data entry instructions were made during the 1988 season.

1. The order of entry for minimum and maximum temperatures was changed to enter the maximum temperatures first. This change corresponds to the way the temperature data usually is received from the weather stations. (The edit program will accept either order and will identify the larger of the two temperature values as the maximum temperature.)

2. Drop the requirement that precipitation be entered to the nearest hundredth of an inch, WITH the decimal entered. Precipitation is still entered to the hundredth of an inch but the decimal is implied.

COSTS

The cost of collecting, entering, and editing the weather data varied considerably by SSO (Table 1). Major factors influencing these costs were the number of weather stations and whether the data could be obtained in electronic or hard copy form. Data collection costs were lower for SSO's which had access to weather data in electronic form.

Table 1. Total costs of obtaining weather data, by SSO.

SSO	Total Cost
	dollars
Illinois	*/
Indiana	400
Iowa	1250
Michigan	1400
Minnesota	1100
Missouri	500
Nebraska	900
Ohio	200
South Dakota	1300
Wisconsin	350
Total	7400

*/ Included in Indiana SSO costs.

Locating sources of weather data and key entry were the biggest cost items for most SSO's. Weather data for Indiana, Illinois, and Ohio were available in electronic form from the NWS data base in West Lafayette, Indiana. The Indiana SSO collected and edited data for both Indiana and Illinois. The Ohio SSO electronically transmitted their weather data from the NWS data base and reformatted it using a BASIC program to eliminate key entry and associated costs. The remaining SSO's key entered data for some if not all weather stations.

Ranges of computer costs, key punching charges, clerical time costs, and statistician time costs are available. Computer edit costs were the least expensive item. Edit runs usually required about 4 Resource Units (RU).[4/] Total costs for edit runs were usually less than $25 per State. Costs for key punching and other clerical time ranged from just over $100 to almost $500. Statistician time costs ranged from almost $200 to almost $600. SSO's which keypunched most of their data had higher costs in all categories.

The summary program was run on the mainframe computer by the Estimates Research Section and cost about seven (7) RU per State.

SSO COMMENTS AND RECOMMENDATIONS

The six additional SSO's found the project lead time to be short and felt like they "were always playing catch-up." Accurate cost accounting was difficult because project codes were not included in the instructions. Most SSO's suggested an improvement to the awkward key entry format. Program errors and file access problems caused early edit runs to fail for a couple of SSO's. Lower bounds in the edit program for daily minimum temperatures were too high for some States.

Only one SSO reported having problems keypunching the objective yield data and the weather data on time (by close of business August 1). This SSO did complete keypunching the weather data on August 2.

Recommendations provided by the SSO's are summarized below:

- -- More lead time. Most of the work had to be done in late June and early July, the busiest time for some SSO's.
- -- More complete instructions. This includes accounting code, background, edit and summary timetable, data entry layouts, job streams, file access codes, error messages and contact persons.
- -- Headquarters should receive copies of edit runs.
- -- Headquarters should download weather data for all States where the data is in electronic format.
- -- Allow SSO's to input ASD numbers for weather stations from within the SAS program.
- -- Do not collect July data on a weekly basis. Wait until the final day is known, and then collect for the required number of days. This suggestion came from a SSO which obtained data directly from weather observers.

MODELS

Model development

Two versions of ear weight prediction models, three models for each version, were developed for each State. Variables in each model were selected in stepwise fashion, subject to the restriction that adding each variable resulted in a statistically significant decrease in the residual mean square error (MSE) for that model. Each model included not more than three independent variables. (See Appendix C for a detailed description of the models.)

4/ Resource Units are the billing units used by the mainframe computer.

The Version 1 models were those listed by Warren and Cook[5/]. In order to compare the relative effectiveness of models which included soil moisture variables, the Version 1 set was required to include:

1. at least one model per State which required some use of soil moisture,
2. at least one which did not require soil moisture, and
3. at least one model which used variables that should be available even in years with late median silking dates.

The purpose of these restrictions was to allow a comparative evaluation of the relative forecasting ability of the three types of models. Unfortunately, these restrictions sometimes forced the selection of models for which one variable did not result in a significant reduction in the MSE. The partial regression slope of such a variable was not significantly different from zero -- therefore its presence was both misleading and superfluous. Further, one of the "soil moisture" variables, computed as the product of the soil moisture index and a temperature variable, was highly correlated with the temperature variable. In addition, the Version 1 models were developed from data only from 1967 through 1983, excluding 1970. (Weather data was available only through 1983 when work on the Version 1 models began. Data for 1970 was excluded to eliminate any anomalous effects from the Southern corn leaf blight epidemic.) Consequently, models for most States were derived from only 16 observations. Also, one State (Minnesota) was unable to provide information on median planting and silking dates for years before 1973. Therefore the models for Minnesota were based on only 11 observations.

In response to the objections listed above, a second set (Version 2) of models was developed for use from weather data through 1986. These models were based upon an additional three years of data, did not require the use or non-use of soil moisture variables, and did not consider the product of temperature and soil moisture indices.

Weather variables considered for inclusion in the Version 2 models included four types of temperature variables, a weekly soil moisture index, and a rate of change in the soil moisture index. The temperature variables were

(1) weekly averages of the average daily temperature for the State,
(2) weekly averages of the average difference between the minimum and maximum daily temperatures,
(3) weekly averages of the number of degrees (degree days) by which the maximum daily temperature exceeded 88°F, and
(4) weekly averages of the number of degrees by which the maximum daily temperature failed to reach 88°F.

A non-weather variable used was the median silking date.

As with the Version 1 models, the weekly variables were constructed from a period which began 28 days before the median (50 percent) silking date for each year. The last week considered depended upon the number of days between the first of August and the median silking date for the State. Generally, models for South Dakota, Minnesota, Wisconsin and Michigan could use weather data until the median silking date. However, models for Nebraska, Iowa,

----

5/ "Ibid.

Illinois, Indiana, and Ohio could use data until 7 days after the median silking date, and those for Missouri could use data up to 14 days after the median silking date. The soil moisture indices require weather data from May 1.

All possible one, two, three, and four variable combinations of the candidate variables were considered in selecting the Version 2 models. Acceptable combinations of variables were those where each variable accounted for a significant reduction in the MSE. A few four variable combinations met this criterion, but were rejected because the variables were highly related. Using weather data through 1986 resulted in identifying acceptable three variable combinations in all States.

The major factors in selecting three alternative models from the acceptable combinations were: a restriction to use not more than three independent weather variables in each model, and a significant reduction in the MSE for each added variable. Other factors considered included the collinearity between variables and a desire to have the alternative models use variables from different weather periods. Many of the alternative models differed by only a single variable.

Model evaluation

Both sets of models were evaluated over a nine year period, from 1979 through 1987 (Table 2). This evaluation used a quasi-operational procedure where the models were fixed but the parameters were calculated from weather data only from previous years. (Variables used by the Version 1 and Version 2 models are listed in Appendix C.) On the average, the Version 2 models had a 28 percent smaller root mean square forecast error (RMSFE) than did the Version 1 models.

It can be argued that this method of evaluation places the Version 1 models at a disadvantage. This is because the Version 1 models were developed using data only through 1983 whereas the Version 2 models were developed using data through 1986. Therefore the Version 2 models would be expected to perform better than the Version 1 models for the years of 1984 through 1986. However, the weighted average RMSFE and Relative Efficiencies (RE)[6/] for Version 2 versus Version 1 in 1979-83 and 1984-86 are comparable. Also, 1983 was used to fit models for both Versions 1 and 2, and the Version 2 models were much more effective than Version 1 in 1983.

The Version 2 models were slightly more effective, on the average, than the Version 1 models in 1987. However, on the average, they were slightly less effective in 1988. The Version 2 models were more effective in 1988 in the major States of Illinois, Iowa, and Indiana. They were less effective in Minnesota, Missouri, and Nebraska. It has been shown that the failure of the Version 2 models resulted because the critical weather in those States occurred earlier than at any time in the 1967-86 base period.[7/] The Version 2 predicted ear weights would have had much smaller errors if the weather data

---

6/ Relative Efficiency = RMSFE(numerator)/RMSFE(denominator).

7/ "*Forecasting Corn Ear Weights From Daily Weather Data*", Fred B. Warren, Proceedings of First Annual Conference on Applied Statistics, Kansas State University, 1989.

had been taken from the weeks of highest temperatures, rather than those specified by the models. The challenge to future research is to use this approach in an objective rather than a subjective procedure.

Table 2. Regional root mean square forecast errors and relative efficiencies of regular objective yield survey and Version 1 and Version 2 weather models, 1979-88.

	Root Mean Square Forecast Error			Relative Efficiency*		
Year	Objective Yield	Version 1	Version 2	Vers. 1 / Obj Yld	Vers. 2 / Obj Yld	Vers. 2 / Vers. 1
1979	0.035	0.024	0.020	0.67	0.57	0.86
1980	0.024	0.031	0.028	1.30	1.16	0.89
1981	0.019	0.017	0.016	0.89	0.80	0.90
1982	0.013	0.007	0.006	0.55	0.45	0.82
1983	0.032	0.019	0.009	0.59	0.29	0.49
Avg	0.0246	0.0196	0.0158	0.80	0.64	0.81
1984	0.017	0.014	0.009	0.83	0.51	0.61
1985	0.009	0.010	0.007	1.04	0.71	0.68
1986	0.009	0.008	0.007	0.89	0.77	0.86
Avg	0.0117	0.0103	0.0077	0.88	0.66	0.75
1987	0.008	0.006	0.006	0.78	0.72	0.92
1988	0.019	0.012	0.013	0.64	0.67	1.06
Avg	0.0135	0.0090	0.0095	0.67	0.70	1.06

* "Relative Efficiencies" are computed from unrounded data.

SUMMARY PROGRAM

The summary program is in two parts. Part 1 is a collection of SAS macro programs which can be used to summarize any number of States on a single run. The number of states summarized will depend upon the parameters listed in Part 2. Instructions for obtaining these parameters are in Appendix D.

Parameters

The "parameter list" has three sections. Section 1 contains three SAS data steps which input the following values for all ten States.

1. The State FIPS code and abbreviated name.
2. The historic average (1967-86) number of stalks per acre on the corn objective yield August 1 survey.
3. The acreage of corn harvested for grain in each ASD during the previous year.
4. The median emergence and silking dates for the current year.
5. A time-trend projection for the number of stalks per acre for the current year.

Section 2 reads in the minimum, maximum, long-term (since 1967) and short-term (last five years) averages of historic final weight of grain per ear for each State.

Section 3 of the parameter list contains the following for each State to be summarized.

1. A parameter to identify the State, and the data input file.
2. Statements to call the different program subroutines, as needed.
3. Parameter cards which contain:
   (a) the number of variables in each model, and
   (b) the names and regression coefficients of those variables.

Program steps

The SAS programs do the following:

1. Compute ASD and State level averages of daily minimum and maximum temperatures, precipitation, and soil moisture indices. Daily ASD averages are straight (unweighted) averages of all stations for which reports were available. Daily State averages are weighted (by acres harvested the previous year) averages of those ASD's which had at least one valid report. ASD's for which there were no reports received no weight in the State average for that day.
2. Beginning 28 days before the median silking date, compute
   (a) weekly averages for the soil moisture indices, average temperatures, and daily differences between the minimum and maximum temperatures,
   (b) weekly averages for degree days greater than and less than 88°F, and
   (c) the daily rate of change in the weekly soil moisture indices.
3. Read, from the parameter list, the names and regression coefficients of the variables required by each model.
4. Compute a predicted final weight of grain per ear from each model and compare these predictions with the limits established by the largest and smallest final ear weights since 1966. Any prediction that exceeds either limit will be censored to equal the appropriate limit.
5. Prepare an output table which will contain:
   (a) the names, values, and regression coefficients of the variables in each model;
   (b) the predicted ear weight from each model;
   (c) if the predicted ear weight is larger (or smaller) than the largest (or smallest) final average ear weight since 1966, a statement that the model prediction has been censored to the largest (or smallest) historic ear weight;
   (d) the (censored) average of the three predictions;
   (e) the minimum, maximum, and average of all final ear weights since 1966; and
   (f) the average final ear weight for the previous five years.

PREDICTED EAR WEIGHTS

Predictions

Predicted State level (all maturity categories) ear weights were computed for each of three different models for Version 1 and for Version 2. Averages of the censored predictions for each version of the weather models are listed in Table 3. The average predicted ear weights from the regular objective yield procedures, regular objective yield adjusted for farmer reported crop condition[8/] (adjusted objective yield), and the estimated final weight of grain per ear (from the November objective yield survey) are also shown in Table 3. Relative forecast errors, computed as

$$\{(\text{predicted} - \text{final})/\text{final}\} * 100,$$

are listed in Table 4.

Table 3. August 1 weather model and regular objective yield survey predictions, and final objective yield survey estimate of average corn ear weights, 1988.

State	Weather Models Version 1 2/	Weather Models Version 2	Regular Objective Yield	Adjusted 1/ Objective Yield	Final Objective Yield
Illinois	.313	.300	.333	.244	.254
Indiana	.300	.273	.360	.276	.282
Iowa	.298	.281	.315	.249	.261
Michigan	.241	.238	.282	.238	.233
Minnesota	.292	.317	.307	.235	.249
Missouri	.368	.385	.382	.392	.312
Nebraska	.403	.418	.387	.374	.378
Ohio	.306	.341	.333	.259	.283
S. Dakota	.285	.274	.315	.285	.242
Wisconsin	.249	.270	.308	.251	.227
Region	.313	.312	.334	.275	.276

1/ Adjusted for models using farmer reported condition.
2/ Models and parameters based on weather data through 1983.

Final estimates of ear weight in the ten State region ranged from nearly normal in Nebraska to record (since 1967) lows in Illinois, Iowa, Michigan, and Wisconsin, and near-record lows in all other States except Missouri.

Table 4. Relative forecast errors, August 1 weather models, and regular and adjusted objective yield surveys, 1988.

State	Weather Models Version 1	Weather Models Version 2	Regular Objective Yield	Adjusted Objective Yield
	%	%	%	%
Illinois	23	18	31	- 4
Indiana	6	- 3	28	- 2
Iowa	14	8	21	- 5
Michigan	3	2	21	2
Minnesota	17	27	23	- 6
Missouri	18	23	22	26
Nebraska	7	11	2	- 1
Ohio	8	20	18	- 8
S. Dakota	18	13	30	18
Wisconsin	10	19	36	11
Region	13	13	21	- 1

Weather
The 1988 corn growing season was hotter and drier during May and June than for any year since 1966, but the weather did moderate during July. This scenario had not occurred since before 1967. Since the weather models were developed from data since 1966, the weather model predictions of ear weight were based upon the more moderate weather conditions during July. Therefore, except for Indiana and Michigan, and especially for Nebraska and Missouri, all weather model predictions of ear weight were above the final estimates.

Comparisons
The regional average relative forecast errors (Table 4) from the weather models were almost forty percent smaller than those from the regular August 1 objective yield survey procedures.

The Version 1 weather model predictions were closer to the final estimated ear weight than the regular objective yield survey in all States except Nebraska. (Nebraska was the one State where the final estimate of ear weight was about average.) The Version 2 weather model predictions were more accurate than the regular objective yield survey in six of the ten States. The Version 2 predictions were more accurate than the Version 1 predictions in five States, including the major States of Illinois, Iowa, and Indiana.

For samples in the "blister" and "pre-blister" stages of development, the "adjusted objective yield" model predicts final ear weight from a regression of State average farmer reported condition of corn against final ear weight. These predictions were combined with the regular objective yield survey predictions for more mature samples. The "condition" values are subjectively determined by a non-probability sample of "crop reporters." The "adjusted objective yield" predictions of final ear weight were more accurate than any of the other predictions in eight of the ten States.

Censoring
Weather model predicted ear weights were "censored" to remain within the bounds set by the minimum and maximum historic ear weights for each State. For the 1988 crop season, censoring of the Version 1 weather model predictions was required in three States, Illinois, Indiana and Michigan. Censoring resulted in substantial improvements in the accuracy of the Version 2 predictions for Indiana and Michigan but diminished it for Illinois. Censoring also resulted in substantial improvements in the accuracy of the Version 1 predictions in Indiana and Michigan.

Median silking date
The validity of using the median silking date in the weather models as a specific variable continues to be questionable. One rationale for using it is that an earlier silking date means that the crop is exposed to more sunlight

----

8/ For samples in the blister and pre-blister stages of development, this adjustment effectively consists of substituting a regression of farmer reported State-average condition against final ear weight for the historic average ear weight.

during its developmental stage, hence more photosynthesis occurs and the crop is more productive. This theory does not consider the situation, as this year, when abnormally high temperatures reduce the amount of photosynthetic activity. The median silking date was used in Version 2 weather models for only three States. In one of those States, Missouri, the median silking date contributed substantially to the overestimate in the Version 2 prediction.

CONCLUSIONS AND RECOMMENDATIONS

The daily weather data required by the corn ear weight weather models can be obtained in a timely manner from the National Weather Service networks in the ten corn objective yield States. However, the number and distribution (see Appendix E) of weather stations in these networks in some States, particularly Illinois, Michigan, and Ohio, should be improved. Also, edit programs to impute data for missing reports should be developed.

The corn ear weight weather prediction models are restrictive in terms of the time periods which they consider. Attempts will be made to develop alternative models which are based upon specific weather events and crop maturity. For example, one variable could be the seven consecutive days, between tassel initiation and August 1, for which average maximum temperature was highest. If alternative models cannot be developed, then the Version 2 models developed from weather data through 1986 should be used in the major States. However, the original (Version 1) models may be preferable in some of the minor States.

APPENDIX A:

Notification and Data entry instructions

I. Notification:

April 21, 1988

SUBJECT: Operational testing of Weather/Corn August 1 Ear Weight Models

TO: Statisticians in Charge -- Indiana, Iowa, Missouri, and Nebraska

THRU: George Hanuschak, Chief, Survey Research Branch

The purpose of this memo is to expand upon information provided in T-15-88.

One recommendation of the 1988 Corn Specifications meeting was that real-time field testing of the weather driven models for August 1 forecasts of corn ear weight should be conducted at a semi-operational level in your four States in 1988. As discussed in our previous conversations, your primary involvement will be to obtain, key, and edit the required weather data. All data editing and processing will be on the Martin-Marietta Data System (MMDS).

Data Requirements: The principal data requirement for these models is for timely reports of daily minimum and maximum temperatures and precipitation from at least 30 weather stations per State, from May 1 through July 29. This will require obtaining data from at least a select group of the cooperative weather stations in each state. There should be at least two weather stations per Crop Reporting District (CRD), and at least four or five stations in each of the more important CRD's. Also, at least for July, this data must be obtained at least weekly.

Clean edited data for May should be on MMDS by June 30, June data by mid-July, and July data through July 15 by July 19. Data for each week (Saturday through Friday) of the remainder of July (through July 29) should be entered on MMDS by the following Tuesday.

Other data needs will be for 1987 CRD acreage weights for corn to be harvested for grain, and a list of the weather stations to be used together with the CRD's in which they are located. These are to be sent to Fred Warren, Estimates Research Section, by June 1. Median (50 percent completion) 1988 planting and silking dates will be derived by Fred Warren from data in the Weekly Crop Weather Report.

Data Entry: Data for up to five (5) consecutive days for a single station may be entered on a single data input record, as specified on the attachment. Data to be entered in the ID section of each input record will include (a) your State FIPS code, (b) the code (optional) for the CRD in which the station is located, (c) the station name/location or National Weather Service number, and (d) the month and day of the month for the first day of data on the record. Data to be entered for each of up to five (5) consecutive days for that station will be the daily minimum and maximum temperatures (Fahrenheit) and precipitation. Temperature values are to be entered as whole numbers (integers) but precipitation is to be entered to the nearest hundredth of an inch, punching the decimal. If there was no precipitation for that day, enter a zero (0). If any or all items are missing for a particular day, enter only a decimal (.) in that/those field(s). A sample format layout is attached for your use. Please notify Fred Warren (475-3490), Estimates Research Section, if this format is not satisfactory.

Data Processing: All data editing and processing, including the preparation of the August 1 forecast grain weights per ear, will be done on the Martin-Marietta Data System. SAS language edit programs and instructions will be supplied for your use in checking for probable key-entry data errors. The summary and forecast programs will be run by the Estimates Research staff.

Fred B. Warren
Mathematical Statistician

cc Fred Vogel, Director, State Statistical Division

II. Data entry format:

FORMAT FOR ENTERING DAILY WEATHER DATA
1988 pilot corn ear weight project

Card Column	Item
1-2	State Fips Code
3-4	CRD number
6-16	Station identifier (alpha)
18-19	Number of month (5=May, 6=June, 7=July)
20-21	Day of month -- for data in cc 23-33
	Enter data for day (in cc 20-21)
23-25	Minimum daily temperature (degrees Fahrenheit)
26-28	Maximum daily temperature (degrees Fahrenheit)
29-33	Precipitation, in hundredth of an inch -- punch decimals, enter zero if none.
	Enter daily data for up to four (4) more days
34-36	Minimum daily temperature for second day
37-39	Maximum daily temperature for second day
40-44	Precipitation for second day
45-47	Minimum daily temperature for third day
48-50	Maximum daily temperature for third day
51-55	Precipitation for third day
56-58	Minimum daily temperature for fourth day
59-61	Maximum daily temperature for fourth day
62-66	Precipitation for fourth day
67-69	Minimum daily temperature for fifth day
70-72	Maximum daily temperature for fifth day
73-77	Precipitation for fifth day

APPENDIX B:

Instructions for editing daily weather data

June 17, 1988

SUBJECT: EDITING 1988 CORN EAR WEIGHT WEATHER DATA

TO: Statisticians in Charge--Indiana, Iowa, Missouri, Nebraska

My memo of April 21 provided instructions for keying the daily weather data required for use in the 1988 Weather/Corn Ear Weight project, and suggested that the daily weather data for May be keyed into a Martin-Marietta Data Systems file, ready for editing, by June 30. This memo provides instructions for running the weather data edit program and for correcting any erroneous entries.

The job control cards needed to run the weather data edit program on Martin-Marietta follow:

```
//  EXEC SAS,REGION=2000K
//RAWDATA DD DSN=your-weather-data-file,DISP=SHR,UNIT=SYSSR
//STNCRD DD DSN=your-master-station-name-file,
//    DISP=SHR,UNIT=SYSSR
//SASPGMS DD DSN=SR770.Y11.EDIT88WX.DATA,DISP=SHR,UNIT=SYSSR
//GOODDATA DD DSN=SR770.Y11.xy88WX.SAS.DATA,DISP=OLD,
//        UNIT=SYSSR
//SYSIN DD *
    %INCLUDE SASPGMS;
```

where:

1. the string 'xy' is to be replaced by the 2 character abbreviation for your State (IN, IA, MO, or NE).
2. the DDNAME 'RAWDATA' defines the file of daily station weather reports to be edited.
3. the DDNAME 'STNCRD' identifies a file of station and Agricultural Statistics District (ASD) identifiers. I was to have constructed this file from a list which you were to have sent me. If this list has not been submitted, you may construct your own file as follows. For each station, enter the ASD number in columns 1-2, and the station identifier (as used in keying the weather data) starting in column 3.
4. the DDNAME 'SASPGMS' identifies the weather edit program. This program is called by the '%INCLUDE SASPGMS' statement.
5. the DDNAME 'GOODDATA' identifies a SAS data set which will include all of the 'good' weather data output by the edit program.

The edit program will:

1. Reject and print an error message for any record for which the station identifier and corresponding ASD code does not match

one of the stations listed in the 'STNCRD' file. To be included in the 'GOODDATA', any such weather data must be corrected and resubmitted in a future run. The program will also reject any record for which the month code is not '5', '6', or '7', or which has an invalid 'day' code.

2. Accept, but print a warning message, whenever the entered temperature and/or precipitation exceeds certain arbitrary limits. If your review shows that the keyed data was in error, submit a corrected record for that day(s) on the next edit run. Otherwise, no additional action is required.

3. Accept all records which pass the various edit limits, and merge them into the 'GOODDATA' file.

This data will be summarized and predicted ear weights computed for your use in preparing the August 1 Crop Report. Therefore, weather data for May and June should be edited at least monthly. Data for July, particularly for the last half of July, **must** be entered and edited weekly.

Any problems are to be referred to me, FTS 475-3490.

Fred Warren
Mathematical Statistician
Estimates Research Section

NOTE: The PC file **L:EDIT88WX.PGM** will, when submitted to MMDS, load the current version of the edit program into the MMDS file **SR770.Y11.EDIT88WX.DATA**, and will also give READ access to the ten corn objective yield states.

NOTE: The PC file **L:SETUPWX.SAS** will, when submitted to MMDS, initialize SAS libraries (for the edited weather station data) for each of the ten corn objective yield states. It will also give each state ALTER access to the library for its own data.

APPENDIX C:

Variables used by Version 1 (original) and Version 2 (revised) weather models.

Variable Definitions:

$AVTEMP_j$ -- The average daily temperature (average of minimum and maximum temperatures) during week 'j'.

$DDG88F_j$ -- The average amount by which the daily maximum temperatures exceeded 88°F during week 'j'.

$DDGSMB_j$ -- The product of $DDG88F_j$ and $WKLYSMB_j$.

$DDL88F_j$ -- The average amount by which the daily maximum temperatures failed to reach 88°F during week 'j'.

$DIFTMP_j$ -- The average difference between daily minimum and maximum temperatures during week 'j'.

$SLOPE_{ij}$ -- the average daily rate of change in soil moisture from week 'i' to week 'j'.

$WKLYSMB_j$ -- The average soil moisture index during week 'j'.

NOTE: Week '1' is 28 to 22 days before the median silking date (SILK_DAY), week '2' is 21 to 15 days, and so forth.

STATE	Model	Variables		RMSFE[1]	
		Version 1[2]	Version 2[3]	Version 1	Version 2
Illinois	1	DDG88F3 AVTEMP5	DDG88F3 DIFTMP3 DDL88F3	.0297	.0217
	2	DDGSMB3 DDGSMB4	DDG88F5 DIFTMP3 SLOPE15	.0325	.0253
	3	DDG88F3 AVTEMP4	DDG88F5 DIFTMP3 DIFTMP4	.0325	.0197
Indiana	1	DIFTMP2 DDG88F3 AVTEMP3	DIFTMP2 DDG88F3 AVTEMP3	.0113	.0113
	2	WKLYSMB3 DDLSMB4	WKLYSMB1 DIFTMP2 DDG88F3	.0267	.0114
	3	DDG88F3 AVTEMP3 WKLYSMB3	DDG88F3 AVTEMP3 SILK_DAY	.0218	.0125

1/ Root Mean Square Forecast Error, from 1979 through 1987.
2/ Models selected using data only through 1983.
3/ Models selected using data through 1986.

STATE	Model	Variables Version 1	Variables Version 2	RMSFE Version 1	RMSFE Version 2
Iowa	1	DDL88F5 DIFTMP2 DDG88F2	DDG88F5 DDG88F4 DIFTMP4	.0518	.0325
	2	SLOPE13 DIFTMP5	DDG88F5 DDG88F4 WKLYSMB2	.0419	.0318
	3	DIFTMP2 DDG88F2	DDG88F5 DIFTMP4 AVTEMP4	.0562	.0315
Michigan	1	SILK_DAY	DIFTMP4 SLOPE23 DDG88F3	.0251	.0169
	2	SLOPE13 DDLSMB3	DIFTMP4 SLOPE23 AVTEMP3	.0207	.0157
	3	SLOPE13 AVTEMP3	WKLYSMB4 WKLYSMB2 DDG88F3	.0242	.0242
Minnesota	1	DDG88F2 DIFTMP4	DDG88F2 DIFTMP4 SLOPE22	.0248	.0186
	2	DDGSMB2 DIFTMP4	DDG88F2 DIFTMP4 DIFTMP1	.0257	.0211
	3	DIFTMP4 AVTEMP2	DDL88F4 DDG88F2 AVTEMP4	.0260	.0258
Missouri	1	DDG88F5 AVTEMP5 DDG88F6	DDG88F5 AVTEMP5 DDG88F6	.0411	.0411
	2	DDG88F5 DDLSMB5 DDG88F6	DDG88F5 DIFTMP5 SILK_DAY	.0440	.0470
	3	DDG88F5 AVTEMP5	DDG88F5 DIFTMP6 DDL88F5	.0498	.0487

STATE	Model	Variables Version 1	Variables Version 2	RMSFE Version 1	RMSFE Version 2
Nebraska	1	DDG88F5 DIFTMP2	DIFTMP2 SLOPE22 WKLYSMB5	.0261	.0201
	2	WKLYSMB4 DIFTMP2 WKLYSMB2	DIFTMP2 SLOPE22 WKLYSMB4	.0193	.0193
	3	SLOPE12 DIFTMP2	DIFTMP2 SLOPE12 WKLYSMB2	.0317	.0158
Ohio	1	SILK_DAY DIFTMP2	DDL88F4 SILK_DAY DDL88F1	.0295	.0289
	2	DIFTMP2 SILK_DAY WKLYSMB5	DDL88F3 DDL88F4 DDG88F4	.0370	.0288
	3	SLOPE12 SLOPE13 SILK_DAY	DDL88F1 SILK_DAY AVTEMP4	.0361	.0280
S. Dakota	1	DDG88F3	DDG88F1 DIFTMP4 DIFTMP2	.0276	.0281
	2	SLOPE14 DDG88F3	DDG88F3 DIFTMP2 SLOPE14	.0323	.0396
	3	DDGSMB3	DDG88F1 SLOPE14 DIFTMP2	.0278	.0401
Wisconsin	1	DDG88F1 DDG88F2 AVTEMP1	DDG88F1 WKLYSMB2 SLOPE13	.0246	.0199
	2	DDGSMB2 DDG88F1	DDG88F1 DDG88F2 WKLYSMB2	.0204	.0215
	3	DDG88F1 DDG88F2	DDG88F1 DIFTMP4 SLOPE14	.0209	.0185

APPENDIX D:

Parameters required by summary program

Data set maintenance: Historic objective yield survey data is contained in the PC-SAS file **H:OYSURVEY.SSD**. This file contains monthly expansions and predicted state level values from the corn objective yield survey and must be updated at least annually. Sources of data for this update are, alternatively, the microfiche copies of monthly corn objective yield summary listings kept by the Methods staff, or the hard copy listings kept by Florence Moreland, Crops Branch.

Parameters: Three different sets of parameters are required. These are to be inserted into the program file **L:EWT88SUM.PGM** as indicated below.

Weights: A card image file, to be read as input to the Data Step WEIGHTS. This file will contain the following data for each state, in this sequence.

1. The State FIPS code and alpha abbreviations, for example "17 IL".
2. The average number of plants per acre, from the first year of usable weather data through 1986.
3. The estimated number of acres harvested for corn last year, by ASD (Agricultural Statistical District). Note: enter a " . " (missing) for Nebraska ASD #4.
4. The Julian dates for median (50 percent completion) planting and silking dates for the current year.
5. A "best" estimate of the average number of stalks per acre for the current year. If current objective survey data is not available, use a trend projection from previous years.

Historic Ear Weights: Four card image records to be read as input to the Data Step MINMAX. These card images will contain the following information.

1. The smallest historic (since 1966) final average ear weight, for each of the 10 objective yield states, ordered alphabetically from Illinois through Wisconsin.
2. The largest historic final average ear weight, for each of the 10 objective yield states, ordered alphabetically from Illinois through Wisconsin.
3. The long term mean (since 1966) final ear weight, for each of the 10 objective yield states, ordered alphabetically from Illinois through Wisconsin.
4. The average ear weight for the five year period just before the current year, for each of the 10 objective yield states, ordered alphabetically from Illinois through Wisconsin.

Job Stream: The job stream for each State will consist of thirteen (13) cards.

* The first card will always be a "%LET SN = 'state FIPS'" statement.
* The second card will always be a "%MAIN" statement.
* The third card will be a "CARDS;" statement.
* The fourth card will contain the number of variables for each of the three weather models in positions (cc) 1, 3, 5.
* The fifth card will contain a "RUN; %PARAMS" statement.
* The sixth card will be a "CARDS;" statement.

* The seventh card will contain the names of the weather variables used by the first model. These must be contained within cc 1-9, 10-18, 19-27, etc..
* The eighth card will contain the regression coefficients for the first model. The intercept term will be entered first, followed by the partial regression coefficients for each of the specified weather variables. These values are to be separated by at least one blank space.
* Cards nine and ten will contain the same information for the second model.
* Cards eleven and twelve will contain the same information for the third model.
* Card thirteen will contain a "RUN; %FORECAST" statement.

APPENDIX E:

Appraisal of National Weather Service Networks

The following appraisal of the distribution and reporting frequency of existing networks of weather stations within individual states (Table 1) identifies areas where improvements would be desirable.

Illinois -- At least one additional station in District 3 and in District 6 would be desirable.

Indiana -- Only about half of the stations reported for the last three weekends in July. This included a complete loss of data for Districts 1 and 9.

Iowa -- The overall number and distribution of weather stations was very good. Reports were obtained from at least 69 stations every day from May 1 through July 22. Because of the early median silking date, weather data after July 22 was not needed by the ear weight models.

Michigan -- The overall number of weather stations is good but at least two additional stations are needed in District 5. Also, the number of weather stations in Districts 1 and 2 is unnecessarily large.

Minnesota -- The number and distribution of weather stations are at least adequate.

Missouri -- The number and distribution of weather stations are at least adequate.

Nebraska -- The overall number and distribution of weather stations are very good.

Ohio -- At least one more station each in District 4 and in District 6 would be desirable. Alternatively, since these districts are quite rectangular, devise a method of incorporating data from nearby weather stations.

South Dakota -- The overall number and distribution of weather stations are at least adequate. However, the number of stations in Districts 1 and 4 is unnecessarily large.

Wisconsin -- The overall number and distribution of weather stations are adequate, but no reports were received from District 3 during July. Also, the number of stations reporting on weekends is much smaller than the number reporting on weekdays.

In conclusion, the total number of weather stations in each state varied from barely adequate to quite sufficient. However, additional stations are needed in some districts. Also, there are districts where most of the existing stations do not report on weekends. Until existing networks are improved, effects of these problems could be reduced through (a) counting certain borderline stations in more than one district, and (b) more sophisticated edit programs which would test for missing data and impute values as needed.

Table 1: Minimum and maximum numbers of weather stations reporting, by districts and States, 1988.

State	District									
	1	2	3	4	5	6	7	8	9	All
Illinois	3	4	2	3	4	2	4	3	3	28
Indiana	0-4	2-4	3-4	2-5	4-5	1-3	2-3	1-3	0-1	15-32
Iowa	7-9	4-6	9-11	6-9	8-10	7-10	3-7	4-6	4-5	52-73
Michigan	7	8	4	5	1	7	9-10	5	7	53-54
Minnesota	4	3	3	5-6	6	4	4	4	5	38-39
Missouri	4-5	3-4	3-5	5-6	6	4-5	3-4	3-4	5	39-41
Nebraska	5-6	4-6	8-10	*/	7-8	13-15	9	7	9-10	63-71
Ohio	3	3	5	2	5	1	4	3	3	29
S. Dakota	9-10	7	7	6	5	5-6	4	4	6	53-55
Wisconsin	2-4	3	0-3	2-4	2-4	4-5	1-4	3-5	2-4	21-36

*/ Nebraska does not have a District 4.

APPENDIX F:

A Pilot Test of the Weather/Corn Ear Weight Forecast Procedure, Indiana and Iowa, 1987

SUMMARY

Collection and summarization of daily weather data was accomplished in time for the August 1, 1987 forecast of corn yields. This was done by NASS[1/] field office staff in the States of Indiana and Iowa. The predicted ear weight from the weather models for Iowa was closer to the final objective yield survey estimate of grain per ear than was the regular objective yield survey forecast. This was not true for Indiana.

The first recommendation is to continue the testing of these models in 1988, in two more states. Second, the models should be revised, considering data through 1986.

INTRODUCTION

Regression models to predict the final average ear weight at the State level using weather data from planting to about August 1 have been developed. These models were evaluated using historic weather and objective yield (OY) survey data.[2/] These evaluations did not explicitly address the question of whether individual State Statistical Offices (SSO's) could obtain and process the required weather data in time for the forecast ear weights to be used for NASS's "August 1" forecast of corn yield and production. (This forecast, relating to conditions at about August 1, is usually summarized and released to the public on August 10.) Therefore this study was conducted in the States of Iowa and Indiana to gain experience with the ability of SSO's to acquire, edit, and summarize the required weather data in the requisite time frame.

DATA

Types of data to be obtained/derived by the SSO's were:

a. Estimates of median planting and silking dates for the individual State.
b. Estimates of the acreage of corn to be harvested for grain in each Agricultural Statistics District (ASD) of each State.
c. Daily precipitation and minimum and maximum temperatures from May 1 until seven days after the median silking date. Reports were to be obtained from as many weather stations as possible.
d. Estimates of the State average number of corn plants per acre.

1/ National Agricultural Statistics Service, United States Department of Agriculture.
2/ "Forecasting Grain Weight Per Corn Ear On August 1", Warren, Fred B., and Cook, Paul W., NASS/USDA Research Report Number SRB-88-03, February 1988.

This data was obtained as follows.

a. Median planting and silking dates were interpolated from weekly crop progress percentages in the Weekly Crop-Weather Bulletin, published jointly by NASS and the National Weather Service, NOAA, Washington, DC.

b. Estimated acreages of corn harvested in 1986 were used for the 1987 weights.

c. Daily weather:
Indiana -- The National Weather Service has a Mid-West Agricultural Weather Service Center at Purdue University. Jim McIntyre, 317-494-8900, was the meteorologist in charge of that facility. Our contact was Ken Scheeringa, 317-494-8105. They receive daily reports from an average of 25 weather stations per State (34 for Indiana) for a five state area -- Illinois, Indiana, Michigan, Ohio, and Kentucky, plus part of Missouri. The Center provided the Indiana SSO with computer printouts of the daily reports for Indiana as requested. (There is also a ROSA network which receives, intermittently, daily reports from volunteers in Iowa, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota.)
Iowa -- Weekly (Saturday through Friday) reports of daily precipitation and maximum and minimum temperatures are received by the State Climatologist (at the Des Moines airport) on Monday. These reports were picked up and keypunched by the Iowa SSO.

d. Plants per acre: this statistic came from the 1987 August 1 corn objective yield survey.

DATA PROCESSING

General: Data for Indiana was processed on a main-frame computer while that for Iowa was processed on a 'personal computer'. All programs were written in the SAS (Statistical Analysis System) language.

Daily weather data was keyed and edited in batches. These were monthly for May and June, and weekly thereafter. The edit program was essentially a check for obvious data entry errors. An error listing was produced in the following instances.

1. The recorded date was before May 1,
2. The Agricultural Statistics District code identifier was incorrect.
3. the recorded minimum temperature was less than 50°F, more than 90°F, or more than the recorded maximum temperature.
4. the recorded maximum temperatures was less than 70°F or greater than 100°F.

Data which failed (3) and/or (4) above was allowed to pass through into the 'clean data' file. All records appearing in the error listing were reviewed against the original records. Where indicated, corrected records were resubmitted and used to update the previous 'clean data' file.

Daily weather data was collected until one week after the median silking date in each State. (Because of the early planting season in 1987, the median silking dates were much earlier than usual -- July 9 in Indiana and July 14 in

Iowa. More usual median silking dates for these states are July 22 to July 29.)

The summary program included the following procedures.

1. Computation of ASD daily average minimum and maximum temperatures and precipitation.
2. Computation of State weighted (by 1986 acres) daily average minimum and maximum temperatures, precipitation, and soil moisture indices.
3. Computation of the weekly weather variables required for the forecast program.
4. Computation of three separate forecast ear weights, and a simple average of the three forecasts, for each state.

August 1 Forecasts: Because of the very early growing season, there was no difficulty in obtaining all the weather data required by the weather models. Also, ear measurements could be taken in most of the corn objective yield samples in these two states. Predicted average weights of grain per ear were computed from each of three different models for each state. The individual model predictions were then averaged to produce a single indication for each state.

The objective yield survey measurements led to predictions of a record high ear weight in Indiana, and a far above average ear weight in Iowa. The averages (composites) of the weather model predictions were for about an average ear weight in Indiana and slightly above average in Iowa. However, the individual model predictions for Indiana did vary from .374 to .333 pounds per ear. In this case, the extreme errors resulted from models which used a soil moisture index variable. This suggests that the models should be reviewed, including data at least through 1986 rather than only through 1983, to determine if the models should be changed for 1988.

The objective yield survey end of season estimates of ear weight showed a record high ear weight in Indiana, and a slightly above average ear weight in Iowa. The final estimate for Iowa was slightly lower than the composite weather model prediction.

August 1, 1987 predictions of corn ear weight.

Indiana		Iowa	
Model Variables	Forecast (pounds/ear)	Model Variables	Forecast (pounds/ear)
DIFTMP2 DDG88F3 AVTEMP3	0.374	DDL88F5 DIFTMP2 DDG88F2	0.381
WKLYSMB3 DDLSMB4	0.337	SLOPE13 DIFTMP5	0.374
WKLYSMB3 DDG88F3 AVTEMP3	0.333	DIFTMP2 DDG88F2	0.363
Composite	0.348	Composite	0.373
Running (1982-86) average	0.356		0.341
Objective survey prediction	0.393		0.382
Actual weight of grain per ear	0.392		0.369

RECOMMENDATIONS

1. Operational testing of the weather models will be continued for another year in a larger number (four) of states. This would expand the base of experience in collecting weather data. This testing will also provide a set of documented procedures for future use.

2. The weather models developed from weather and other data from (at the earliest) 1967 through 1983 should be reworked, to include weather data through at least 1986.

[illegible]