YRB 83-05

United States
Department of
Agriculture

Statistical
Reporting
Service

Statistical
Research
Division

November 1983
AGES831125

Evaluation of Haun Models for Predicting Spring Wheat Yields in North Dakota

Fatu G. Bigsby
and
Jeanne L. Sebaugh

EVALUATION OF HAUN MODELS FOR PREDICTING SPRING WHEAT YIELDS IN NORTH DAKOTA. By Fatu G. Bigsby and Jeanne L. Sebaugh; Statistical Research Division, Statistical Reporting Service, U.S. Department of Agriculture, SRS Staff Report No. AGES831125

ABSTRACT

Haun developed models for predicting spring wheat yields in North Dakota using stepwise regression procedures. Individual submodels, consisting of different sets of agronomic and weather variables, exist for making predictions using weather data up to 4, 8, 12, 16, or 20 weeks after the fifty percent planting date for each crop reporting district (CRD). The data base for each submodel consists of a pooling of the historic data from the nine CRDs. The 12 week model performed best among the submodels and was compared to another spring wheat model, the CEAS model. No significant differences in performance were found. However, since the Haun model is more complex and less timely, its use is not justified.

Key Words: Model evaluation, model comparison, yield regression model, spring wheat yield model, growth rate index.

ACKNOWLEDGEMENTS

The authors wish to thank Wendell W. Wilson and other AgRISTARS Yield Model Development Project personnel for their comments and assistance and Jean Sparks for typing this report.

Columbia, MO

November 1983

EVALUATION OF HAUN MODELS FOR PREDICTING SPRING WHEAT YIELDS IN NORTH DAKOTA

by

Fatu G. Bigsby
Mathematical Statistician

and

Jeanne L. Sebaugh, Ph.D.
Mathematical Statistician

This research was conducted as part of the AgRISTARS* Program. It is part of task 4 (subtask 1) in major project element number 1 as identified in the 1983 Yield Model Development Project Implementation Plan. As an internal project document, this report is identified as shown below.

---

*AgRISTARS is an acronym for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing. It is a multi-agency research program to meet some current and new information needs of the U. S. Department of Agriculture.

AgRISTARS
Yield Model Development
Project

YMD-1-4-1(83-11.1)

TABLE OF CONTENTS

TABLE OF CONTENTS: Continued

LIST OF TABLES

LIST OF FIGURES

EVALUATION OF HAUN MODELS FOR PREDICTING SPRING WHEAT YIELDS IN NORTH DAKOTA

by
Fatu G. Bigsby
and
Jeanne L. Sebaugh, Ph.D.

SUMMARY

Haun (1979) developed regression models to predict spring wheat yields at 4, 8, 12, 16 and 20 weeks after the 50 percent planting date by pooling historic data from the nine North Dakota crop reporting districts (CRDs). Daily weather data were used to construct weather related variables, including a morphologically defined growth rate index. The amount of nitrogen fertilizer applied was used to account for changes in yield levels due to technology. Linear, quadratic, cubic and interaction terms were submitted to stepwise regression analysis for model development.

Modified Haun type models were developed by the authors by respecifying the original Haun models so that Haun's basic independent variables could be chosen by stepwise regression procedures and so that indicator (0, 1) variables could represent the differences among the yield levels of the CRDs. Two types of models were developed, one excluding and one including outlier years. Both week 12 models performed better in bootstrap tests than the models developed to predict yield at 4, 8, 16 and 20 weeks after the planting date. The week 12 model with outliers included performed better (RMSE equal to 2.27 bu/acre) than the model with outliers excluded (RMSE equal to 2.82 bu/acre). The 12 week Haun model with outliers included was then compared with the CEAS Trend and Monthly Weather Data Model for spring wheat in North Dakota. The comparison showed no significant difference between the performance of the Haun model and the CEAS model (RMSE was 2.27 for the Haun model and 1.99 for the CEAS model). The Haun model was not as timely as the CEAS model.

Because of the difficulty in interpreting the independent variables in the Haun model, its consistency with scientific knowledge is difficult to judge. Its performance does not justify its increased complexity.

INTRODUCTION

Description of Models

Original Haun Model

The Haun spring wheat model for North Dakota (Haun, 1979) is a regression model based on historical data pooled from the nine North Dakota CRDs for 1965 through 1972. Except for the amount of nitrogen fertilizer applied, which is at the state level, the variables in the model are at the CRD level. Submodels were developed which use weather data for various periods of time, from 4, 8, 12, 16 up to 20 weeks after each CRD's 50 percent planting date. The yield prediction equation is the same for all CRDs for a given submodel. The authors obtained aggregated state level predicted yields by weighting the model predicted CRD yields by the number of acres of all spring wheat harvested in each CRD.

During the model development, Haun identified several "outlier years" for some CRDs. The data for these outlier years were excluded when the model was developed (Haun, personal communication, August 1982). Dr. Haun conducted testing of the model on regions of the U.S. and U.S.S.R.

Model Variables. The unit of measure used by Haun for yield, the dependent variable, is kilograms per harvested hectare. Haun used forward stepwise regression procedures to choose the independent variables in the model. Each independent variable was considered by Haun in linear, quadratic, and cubic form and in interaction with each of the other variables. The independent variables considered are briefly tabulated as follows:

Variable	Description
F	Average amount of nitrogen fertilizer applied annually in North Dakota to spring wheat (kilograms per hectare).
PD	Planting date (the Julian day on which 50 percent of the crop is planted in a given CRD).
PP	Sum of daily precipitation from September 1 to April 1 (cm).
PSUM	Sum of daily precipitation from the planting date to the date of prediction (cm).
S	Estimated soil moisture on the planting date, expressed as a percent of the estimated capacity.
GRI	Average daily growth rate of the crop (morphological unit).

Values of GRI and PSUM are time dependent and vary across the submodels. Values of the other independent variables are the same for all of the submodels.

F is defined as the average amount of nitrogen fertilizer (kilograms per hectare) applied annually in North Dakota to spring wheat. F is the same for all CRDs in a given year because the data are only available at the state level. The Statistical Reporting Service (SRS) of the USDA collects data on fertilizer use during the objective yield survey. The data published are not official SRS estimates; they are intended to provide indications of trends in the use of fertilizer in North Dakota. The indications are published for all wheat, but North Dakota is principally a spring wheat state. Two indications published by SRS are rate per receiving acre (acres on which nitrogen fertilizer was applied) and percent of acres that received nitrogen fertilizer. F is then calculated by multiplying the rate per receiving acre by the percent of acres receiving fertilizer and converting from pounds per acre to kilograms per hectare.

The planting date (PD) is defined as the date on which 50 percent of the crop is planted. It is important because it is not only used as an independent variable in the yield prediction equation but also because the values of three other independent variables (PSUM, GRI and S) depend on the planting date. As noted above, the first yield prediction for the state cannot be made until four weeks after the planting date has occurred in all nine CRDs. The planting dates used in the original model were obtained by using the planting date model developed by Haun (1976) to estimate the planting dates for 1965 to 1966 and by "graphical estimation" of planting dates from USDA records for 1967 to 1976 (Haun, personal communications, August 1982).

Precipitation enters the model directly through PP and PSUM. PP (preseason precipitation) is defined as the sum of daily precipitation, in centimeters, from September 1 to April 1. PSUM is the sum of daily precipitation, in centimeters, from the planting date to the date of prediction.

The effects of moisture stress on the crop are accounted for in the model directly by S and indirectly by GRI. S is defined as the estimated soil moisture percent (EO) on the planting date. EO is obtained by converting ESM (estimated soil moisture in centimeters) to percent of field capacity. The intermediary variables used to estimate ESM are field capacity (FC), day length (DL) and potential evapotranspiration (PE). FC is defined as centimeters of water which can be obtained in the upper 24 inches of the soil. Haun used 10.16 centimeters (4 inches) as the field capacity for North Dakota (Haun, personal communication, August 1982). The method developed by R. A. Stuff (1975) was used to estimate DL. The Thornthwaite (1948) method was used to estimate PE since the required input weather data can be limited to daily maximum and minimum temperature and daily precipitation.

Conceptually, GRI is the most important independent variable in the model. It was placed in the model because of Haun's assumption that yield is related to the growth rate of the crop. The manner in which the growth rate is estimated distinguishes the Haun models from other yield models that are based on plant growth. The growth rate is estimated daily and is based on morphological rather than phenological changes in the plant. (Webster's Dictionary defines morphology as features comprised in the form and structure of an organism or any of its parts. Phenology is defined as relations between climate and periodic biological phenomena of plants such as flowering or fruiting.) Yield models that use phenological changes to estimate the growth rate can do so only when distinct developmental stages like flowering or fruiting occur; whereas, morphological changes can be observed more frequently because they can involve any observable plant change.

Haun used morphological changes such as the appearance of plant leaves to divide the growth of the plant into daily growth units (see Figure 1 in Haun, 1979). The growth stage of a plant on a given day is the total number of growth units that the plant has accumulated. The daily growth rate is defined as the difference between the number of growth units accumulated on two successive days. The sum of the daily growth rates from planting to prediction (determined by the submodel used) defines GRI.

The growth rate model used by Haun (1979) was developed by stepwise regression of combinations and transformations of environmental variables on the average daily growth rate. The growth rate model is

$$\hat{GR} = 0.255-1.810\times10^{-3}(E1)+8.657\times10^{-8}(E33)+1.972\times10^{-3}(TX1)-8.404\times10^{-3}(P3)$$

where

$\hat{GR}$ = estimated daily growth rate in morphological units (2 day moving average)

E1 = estimated soil moisture percent, lagged 1 day

E33 = (estimated soil moisture percent)^3, lagged 3 days

TX1 = daily maximum (C°), lagged 1 day

P3 = daily precipitation (cm), lagged 3 days

Values of the dependent variable were obtained by pooling daily observations of plant growth at four locations in North Dakota. Haun described these locations as "important commercial production areas" with "diverse environmental conditions."

Modified Haun Models

As part of the evaluation of the Haun model, the data base used to develop it was reviewed by the authors to determine if the data values used were accurate and if the data values would be available for possible future use of the model (Working Paper, Bigsby and Sebaugh, 1982). Some incorrect data values were found and corrected. Regression coefficients were estimated from data for the base period 1965 to 1972. The results from estimating the model coefficients for the base period indicated changes from Haun's results in the significance level of the independent variables in all submodels. Although some of the model coefficients as estimated by Haun were not significantly different from zero at the .05 or the .10 level, the results obtained by the authors (with the corrected data) showed even more variables which were not significantly different from zero.

The authors did not feel that they could accurately reproduce Haun's decisions pertaining to the inclusion of variables in his model. However, statistical analyses were performed to determine if Haun's model could be respecified to keep his basic independent variables yet have all independent variables significant at the .05 level.

In respecifying the model, the authors attempted to make the model sensitive to the differences in the yielding abilities of the CRDs as well as differences due to weather. If an attempt is not made to account for CRD to CRD non-weather differences in yield levels, e.g., soil types or fallowing practices, then one is attempting to model these differences along with year to year weather differences by using only weather variables.

Indicator (0, 1) variables were created for the various CRDs. Through statistical analysis, the authors found that the yield levels are similar in CRDs 1, 2, and 5 and in CRDs 7 and 9, but that each of the other CRDs, 3, 4, 6, and 8, are different from each other. The indicator variable for CRD 8 was later excluded because excluding one indicator variable allows an intercept term to be estimated which makes it easier to interpret the estimated coefficients. CRD 8 was excluded because it produces the smallest percent of spring wheat in North Dakota.

Stepwise procedures were used to select a "best" set of submodels using linear, quadratic and interaction terms of the weather and nitrogen fertilizer variables which Haun had used. The variables used in the stepwise procedures are shown in the appendix. The indicator variables representing the CRDs were forced into each submodel. All other variables were required to be statistically significant at the .05 level. The submodels were then developed using the data for 1965 to 1972. Submodels of this form will subsequently be referred to as Haun Type I models.

When the spring wheat model was developed, Haun identified outlier years for some CRDs in each submodel using "empirical" methods (Haun, personal communication, January 1983). The outlier years were excluded when the parameters were estimated. In order to determine if there were outliers in the data values used by the authors, the procedures discussed in Regression Diagnostics by Belsley, Kuh and Welch (1980) and available in the Statistical Analysis System's REG Procedure (1982) were used. The model used in the analysis was the Haun Type I model. The outliers identified by the authors were generally not the same as Haun's. After the outliers were identified by the authors, they were excluded from the data set and submodels based on the reduced data set were estimated in the same manner as before. These submodels will be referred to as Haun Type II models. Dr. Haun (personal communication, January 1983) agreed to allow the evaluation to be carried out for the Haun Type I and Type II models, because the residual mean square errors of the modified models were substantially lower than the residual mean square errors of the original models (see Table 1). Therefore, evaluations and comparisons in this paper will refer to the modified Haun models and not the original models. Table 2 shows the independent variables and estimated coefficients of the original and modified Haun models (week 12).

Comparison of Haun Models with CEAS Model

The performance of the Haun models was compared to the performance of the state level CEAS Trend and Monthly Weather Data models for spring wheat yield in North Dakota. The CEAS models were developed by the Assessment and Information Services Center (AISC). AISC is a part of the National Oceanic and Atmospheric Administration (NOAA) of the U. S. Department of Commerce.

The CEAS models are a regression model based on historic values for yield and the independent variables. The years used in the model development were 1931 to 1978. The CEAS models can be used for predicting yields on the first of the month, for March through August. The model that predicts yield as of August 1 (based upon data through July) is considered the final model.

To account for the relationship between yield and technology, the CEAS model uses linear functions of the year number. The other independent variables in the model are either meteorological or agroclimatic. The criteria used to determine which meteorological or agroclimatic variables to place into the model were: (1) linear correlation between detrended yield and the independent variables or knowledge that the relationship between the variable and yield is of physical importance, (2) the selection of the variable by stepwise regression procedure and (3) the sign of the variable did not conflict with what prior knowledge indicated it should be.

Table 1. Residual mean square errors for the original Haun models and the modified Haun models (kilograms/hectare)^2
Model development period 1965 to 1972

Model week	Estimate of original Haun model1/	Haun Type I model	Haun Type II model
4	46295	19754	19239
8	39604	12036	8947
12	38033	10934	11007
16	42760	13137	11963
20	37763	18830	13984

1/ Estimate of model parameters by authors, using corrected data.

Table 2. Independent variables, estimated coefficients and residual mean square errors (kilograms/hectare)^2 of the week 12 submodel for the original and modified Haun models; model development period 1965-1972

Model	Residual MSE	Independent variables	Estimated coefficients
Original Haun Model estimated by authors	38033	FPSUM	0.64
		F	223.56
		FPP	2.83
		$(PP)^3$	-0.09
		PDPP	0.47
		$(F)^3$	-4.62
		GRIPSUM	14.16
		PSUM	248.38
		PD	21.55
		FPD	-1.33
Haun Type I	10934	CRD 1, 2 and 5	356.49
		CRD 3	672.10
		CRD 4	211.58
		CRD 6	617.47
		CRD 7 and 9	173.44
		$(GRI)^2$	-17.41
		$(PP)^2$	-1.09
		F	430.45
		SGRI	2.81
		PDGRI	2.60
		PDS	-0.37
		FPD	-1.56
		FPP	3.94
		FGRI	-13.86
Haun Type II	11007	CRD 1, 2 and 5	381.26
		CRD 3	739.12
		CRD 4	232.59
		CRD 6	539.92
		CRD 7 and 9	155.36
		PP	97.23
		GRIPP	-7.48
		FPD	-0.45
		FPP	4.10
		FPSUM	2.06

The meteorological variables considered for inclusion into the CEAS model were average monthly temperature, cumulative precipitation, monthly deviations from normal temperature and precipitation, and squared monthly deviations from normal precipitation. Monthly potential evapotranspiration (PET) and actual evapotranspiration (ET) were the agroclimatic variables considered for inclusion into the model. PET and ET were estimated using the Thornthwaite (1948) and Palmer (1965) methods. For a comprehensive description of the CEAS model see LeDuc (1981); for an evaluation, see Sebaugh (1981).

Although the focus of this paper is the evaluation of the Haun models, there is justification for comparing their performance with the performance of the CEAS model. The data collection requirements of the Haun models are greater than they are for the CEAS model because (1) the Haun models require daily weather values while the CEAS model requires monthly weather values, (2) the Haun models require annual values for nitrogen fertilizer applied (F) in order to account for technology while the CEAS model requires no data collection efforts to do this since linear functions of year number are used for this purpose, and (3) annual values of the 50 percent planting dates for nine North Dakota CRDs have to be estimated for the Haun models. The Haun planting date model has not been proven to be effective in predicting the 50 percent planting dates; therefore, it is likely that this information will have to be collected from other sources.

Given that the Haun model is more complex than the CEAS model, in order to justify the use of the Haun model over that of the CEAS model, the Haun model would have to be more timely with the same accuracy or at least as timely with better accuracy.

EVALUATION METHODOLOGY

Model Characteristics To Be Discussed

The document, Crop Yield Model Test and Evaluation Criteria, (Wilson, et al., 1980), states:

> "The model characteristics to be emphasized in the evaluation process are: yield indication reliability, objectivity, consistency with scientific knowledge, adequacy, timeliness, minimum costs, simplicity, and accurate current measure of modeled yield reliability."

This section describes the methodology by which indicators of yield reliability may be computed and used to evaluate the performance of a crop yield model (see appendix for formulas).

Bootstrap Technique Used to Generate Indicators of Yield Reliability

Indicators of yield reliability (reviewed below) require that the parameters of the regression model be computed for a set of data and that a yield prediction be made based on that data for a given "test" year. The values required to generate indicators of yield reliability include the predicted yield, $\hat{Y}$, the observed (reported) yield, Y, and the difference between them, $d=\hat{Y}-Y$, for each test year. It is desirable that the data used to generate the parameters for the model not include data from the test year.

To accomplish this, the "bootstrap" technique is used. Years from an earlier base period are used to fit the model and obtain a prediction equation. The values of the independent variables for the test year following the base period are inserted into the equation and a predicted yield is generated. Then, that test year is added to the base period and the process is repeated for the next sequential test year. Continuing in this way, seven (1973-1979) predictions of yield are obtained, each independent of the data used to fit the model. This process attempts to simulate how the model would have performed had it actually been used on a current basis for those years. For the Haun models data for 1965-1972 (8 years) are used to fit the prediction model for 1973. Data for 1965-1973 (9 years) are used to fit the prediction model for 1974, etc.

Review of Indicators of Yield Reliability

The Y, $\hat{Y}$ and d values for the seven-year test period may be summarized into various indicators of yield reliability. During the estimation of the Haun models, the units of measurements used by Haun were also used by the authors. However, before obtaining the indicators of yield reliability, predicted yields for the Haun models were converted from kilograms per hectare to bushels per acre. The indicators of yield reliability described below are considered to be descriptive statistics which are helpful in characterizing the performance of a model. Tests of hypotheses based on these indicators would have little statistical power due to the small sample size and are not of primary interest.

Indicators Based on the Differences Between $\hat{Y}$ and Y ($d = \hat{Y}-Y$) Demonstrate Accuracy, Precision and Bias

The d value provides estimates of the mean square error (root mean square error and relative root mean square error), the variance (standard deviation and relative standard deviation), and the bias (its square and the relative bias).

The root mean square error (RMSE) and the standard deviation (SD) indicate the accuracy and precision of the model and are expressed in bushels per acre. Assuming the d values are normally distributed, it is about 68% probable that the absolute value of d for a future year will be less than one RMSE and 95% probable that it will be less than twice the RMSE. So, accurate prediction capability is indicated by a small RMSE.

A non-zero bias means the model is, on the average, overestimating the yield (positive bias) or underestimating the yield (negative bias). The SD is smaller than the RMSE where there is non-zero bias and indicates what the RMSE would be if there were no bias. If the bias is near zero, the SD and the RMSE will be close in value. A model whose bias is close to zero is preferred.

Indicators Based on Relative Differences Between $\hat{Y}$ and Y ($rd = 100d/Y$) Demonstrate Worst and Best Performance

The relative difference, rd, is an especially useful indicator in years where a low observed yield is not predicted accurately. This is because years with small observed yields and large differences often have the largest rd values.

Several indicators are derived using relative differences. In order to calculate the proportion of years beyond a critical error limit, we count the number of years in which the absolute value of the relative difference exceeds the critical limit of 10 percent. Values between 5 and 25 percent were investigated and a critical limit of 10 percent was found most useful in describing model performance. The worst and next to worst performance during the test period are defined as the largest and next to largest absolute value of the relative difference.

Indicators Based on $\hat{Y}$ and Y Demonstrate Correspondence Between Observed and Predicted Yields

Another set of indicators demonstrates the correspondence between observed and predicted yields. It is desirable for increases in observed yield to be accompanied by increases in predicted yields. It is also desirable for large (small) predicted yields to correspond to large (small) observed yields.

Two indicators relate the change in direction of predicted yields to the corresponding change in observed yields. One looks at change from the previous year (six observations) and the other at change from the average of the previous three years (four observations). A base period of three years is used since a longer base period would further decrease the number of observations, while a shorter period would not be very different from the comparison to a single previous year.

Finally, the Pearson correlation coefficient, r, between the set of observed and predicted values for the test years is computed. It is desirable that r $(-1 \leq r \leq +1)$ be large and positive. A negative r indicates smaller predicted yields occurring with larger observed yields (and vice versa).

EVALUATION OF HAUN MODELS FOR YIELD INDICATION RELIABILITY

Indicators of Yield Reliability Show That the Week 12 Submodels of the Modified Haun Models Perform Best

Table 3 shows the indicators of yield reliability for the Haun Type I submodels. Among these submodels, week 12 shows the best indication of yield reliability. It performed just as well or better (in most cases) than the other submodels, except for one indicator. The indicator for which the week 12 submodel does not show superior indication of yield reliability is the number of years out of 4 in which the direction of change from the average of the previous three years in the predicted yields agrees with the observed yields. This number is 4 for the week 16 submodel and 3 for the week 12 submodel.

The indicators of yield reliability for the Haun Type II submodels are also shown in Table 3. Except in two cases, the performance of the week 12 submodel is just as good or better than the other Type II submodels. While the week 12 submodel has a larger estimated bias (2.23) than for both the week 8 (2.04) and 20 (1.70) submodels, it has a much smaller mean square error than both of these submodels. The number of years out of 6 in which the direction of change from the previous year in the predicted yield agrees with the observed yields is 4 for the week 12 submodel and all 6 for the week 20 model.

Table 3. Comparison of indicators of yield reliability for submodels of Haun Type I and Haun Type II models based on independent bootstrap test years 1973 to 1979

Indicator of yield reliability (unit)	MODELS		
	Week	Haun Type I	Haun Type II
Bias = B (BU/A)	4	4.09	2.39
	8	2.21	2.04
	12	1.57	2.23
	16	3.36	3.00
	20	2.84	1.70
Mean Square Error = MSE $(BU/A)^2$	4	65.63	26.19
	8	9.05	13.46
	12	5.13	7.97
	16	23.45	20.53
	20	18.28	13.42
Variance = Var $(BU/A)^2$	4	48.94	20.50
	8	4.15	9.29
	12	2.66	3.00
	16	12.18	11.53
	20	10.20	10.53
Number of Years out of 7 in which $\|rd\| > 10\%$	4	4	4
	8	3	5
	12	2	3
	16	4	3
	20	4	4
rd of largest $\|rd\|$ (%)	4	65.5	42.9
	8	21.9	23.6
	12	19.7	23.6
	16	36.9	33.5
	20	36.0	24.9
rd of next to the largest $\|rd\|$ (%)	4	32.9	25.1
	8	21.2	19.4
	12	10.9	16.2
	16	26.5	30.5
	20	22.9	20.7
rd of smallest $\|rd\|$ (%)	4	-3.4	-6.2
	8	-0.4	5.5
	12	-0.4	-1.1
	16	-4.0	-1.5
	20	-1.3	-3.2

Table 3 (Contd.). Comparison of indicators of yield reliability for submodels of Haun Type I and Haun Type II models based on independent bootstrap test years 1973 to 1979

Indicator of yield reliability (unit)	Week	MODELS Haun Type I	 Haun Type II
Number of years out of 6 in which the direction of change from the previous year in the predicted yields agrees with the observed yields	4	3	3
	8	4	4
	12	4	4
	16	1	3
	20	3	6
Number of years out of 4 in which the direction of change from the average of the previous three years in the predicted yields agrees with the observed yields	4	2	3
	8	3	3
	12	3	4
	16	4	3
	20	3	3
Pearson correlation coefficient between observed and predicted yields	4	.19	-0.02
	8	.71	.50
	12	.83	.78
	16	.25	.25
	20	.30	.54

Since it is clear that the week 12 submodels of both the Haun Type I and the Haun Type II models show superior indications of yield reliability over the other submodels, the remainder of the evaluation of the Haun models will be carried out for the week 12 submodels only. For the sake of brevity, statements about the Haun I model and the Haun II model in the following sections will refer to the week 12 submodel of each type. Table 2 (page 7) showed the independent variables and estimated coefficients of the week 12 submodels for the original and modified Haun models.

Indicators of Yield Reliability Show That the Haun I Model Performs Somewhat Better Than the Haun II Model

The indicators of yield reliability for the Haun I model and the Haun II model are summarized in Table 4. The root mean square errors for both models are between 2 and 3 bushels per acre. The Haun I model has lower values for RMSE. These lower values indicate that the Haun I model is a more accurate predictor of yield.

Bias is 1.57 bushels per acre for the Haun I model and 2.23 bushels per acre for the Haun II model. The positive bias of each model shows that each model, on the average, overestimates the observed yields.

The absolute value of the relative difference is greater than 10 percent in 2 out of 7 years for the Haun I model and in 3 out of 7 years for the Haun II model. Both models tended to have the best performance and the worst performance during the same years. The year 1974 has the largest |rd| values for both models (19.7 percent for Haun I and 23.6 percent for Haun II). This was a year of below average yields in North Dakota. The smallest |rd| values for both the Haun I and the Haun II models occurred for the 1979 yield predictions.

One of the three indicators based on $\hat{Y}$ and Y is the number of years in which the direction of change from the previous year in the predicted yields agrees with the observed yields (see Figure 1). This occurred in 4 out of 6 years for both the Haun I model and the Haun II model. Both models failed to predict the direction of change correctly for the observed yields in 1976 and 1977. The direction of change from the average of the previous three years in the observed yields was predicted correctly in 3 out of 4 years for the Haun I model and in all 4 years for the Haun II model.

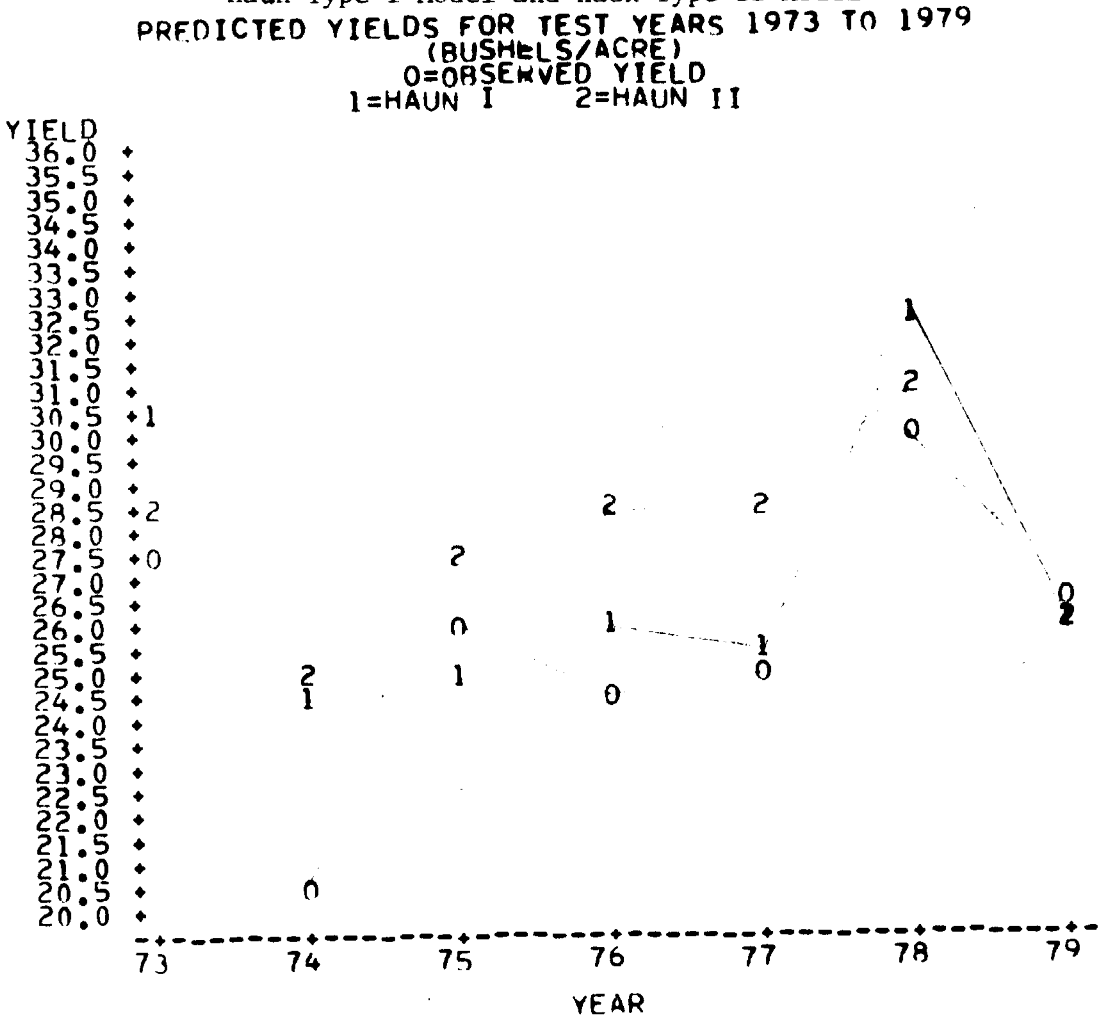
The Pearson correlation coefficient, r, is .83 for the Haun I model and .78 for the Haun II model. The values of r for both models indicate a linear relationship between the predicted yields and the observed yields.

The indicators of yield reliability based on the bootstrap test years 1973 to 1979 show that the performance of the Haun I model is somewhat better than the performance of the Haun II model. Therefore, the remainder of the evaluation of the Haun models will be carried out for the Haun I model only.

Table 4. Indicators of yield reliability for Haun Type I and Haun Type II models (week 12) based on independent bootstrap test years 1973 to 1979

Indicator of Yield Reliability (Unit)	Models Haun I	 Haun II
Bias = B (Bu/A)	1.57	2.3
Relative Bias = RB (%)	6.1	8.7
Mean Square Error = MSE $(\mathrm{Bu/A})^2$	5.13	7.97
Root Mean Square Error = RMSE (Bu/A)	2.27	2.82
Relative Root Mean Square Error = RRMSE (%)	8.8	11.0
Variance = Var $(\mathrm{Bu/A})^2$	2.66	3.00
Standard Deviation = SD (Bu/A)	1.63	1.73
Relative Standard Deviation = RSD (%)	6.0	6.2
Number of Years out of 7 in which $\|rd\| > 10\%$ (%)	2	3
rd of Largest $\|rd\|$ (%)	19.7	23.6
rd of Next Largest $\|rd\|$ (%)	10.9	16.2
rd of Smallest $\|rd\|$ (%)	-0.4	-1.1
Number of years out of 6 in which the direction of change from the previous year in the predicted yields agrees with the observed yields	4	4
Number of years out of 4 in which the direction of change from the average of the previous three years in the predicted yields agrees with the observed yields	3	4
Pearson correlation coefficient between observed and predicted yields	.83	.78

Figure 1
Haun Type I Model and Haun Type II Model



EVALUATION OF HAUN MODELS FOR CONSISTENCY WITH SCIENTIFIC KNOWLEDGE

More Evidence Is Needed to Demonstrate Consistency with Scientific Knowledge

The emphasis in the evaluation of the Haun models for consistency with scientific evidence is on the definitions and interpretation of the independent variables in the models. Although justification is given for the basic independent variables considered for use in the models, no agronomic meaning or interpretation was provided for the quadratic and interaction terms considered for inclusion by Haun. Justification is not provided for the terms in the models because the models are "data-driven." That is, except for the indicator variables representing the crop reporting districts, the choice of the variables retained in the models was determined exclusively by stepwise regression procedures and not by agronomic principles. Some of the interaction terms are particularly difficult to attach meaning to because they represent combinations of variables that are measured in different units. An example of one such variable is GRIPP in the Haun II model. GRI is measured in morphological units and PP is measured in centimeters.

GRI, considered for inclusion in the models because of Haun's assumption that it is related to yield, is defined as the sum of the average daily growth rate of the plant from planting to prediction date. If GRI values for two plants were compared, the plant that had the larger value for GRI would be said to be growing faster.

The growth rate prediction model which was used to estimate the daily average growth rate for the original and modified Haun models was not verified by the authors because the data needed (for both the dependent and independent variables) to make an independent estimate of the model were not available. The values of the dependent variable, average daily growth rate in morphological units, were obtained by pooling daily observations of plant growth at four locations in North Dakota for 1974 and 1975. Each location had a sample size of 25 plants. The criteria used to choose the locations were their commercial importance and the diversity of their environmental conditions. There is no justification given for the sample size (four locations times 25 plants) used to obtain the values of the dependent variable. How effective the growth rate prediction model is in predicting daily growth rates for the state of North Dakota and for other locations will depend partly on how "representative" the locations and years chosen for the sample are. As noted before, 1974 was a year in which yields were below average in North Dakota.

The contribution of GRI to yield is not clear. Haun (1979) alluded to this when he stated, "The nature of its (GRI) relationship to yield is not readily obtained. This is due in part to the various transformations used and to the extent of the intercorrelations among variables." GRI appears as a quadratic term and in interaction terms with other variables in the modified Haun models (see Table 2, p. 7). SGRI and GRIPP are examples of interaction terms that include GRI. SGRI (Haun I) has a positive sign which means, other factors being constant, that the faster the plant grows and the larger the percent of soil moisture on the planting date (S) is, the larger the predicted yield. GRIPP (Haun II) has a negative sign which means, other factors being constant, that the faster the plant grows and the greater the amount of preseason precipitation, the smaller the yield would be.

F, the average amount of nitrogen fertilizer applied annually in the state, is used to account for technological changes that affect yield. Figure 2 shows the observed yields and the average amounts of nitrogen fertilizer (F) applied in North Dakota for both the model development period (1965 to 1972) and the bootstrap test years (1973 to 1979). The plot shows that the amount of F applied annually showed a general increase from 1965 to 1979. The pattern of observed yields is not as smooth; and the relationship between F and the observed yields is not evident. Because changes in yield levels also result from other technological factors such as the adoption of new varieties and acreage programs, the relationship between yield and fertilizer usage is not easily defined.

There are several interaction terms in both models which include F (see Table 2). Examples of these terms are FPD (Haun I and II) and FPSUM (Haun II). FPSUM has a positive sign which means that larger amounts of fertilizer applied along with a larger amount of precipitation (from the planting date to the prediction date) are associated with larger yields. FPD has a negative sign which means that larger amounts of fertilizer applied and a late planting date are associated with lower yields.

The relationship of changes in yield to changes in technology is complex and involves a number of factors. It is unlikely that the use of F as the only trend variable in the model is sufficient to account for technology.

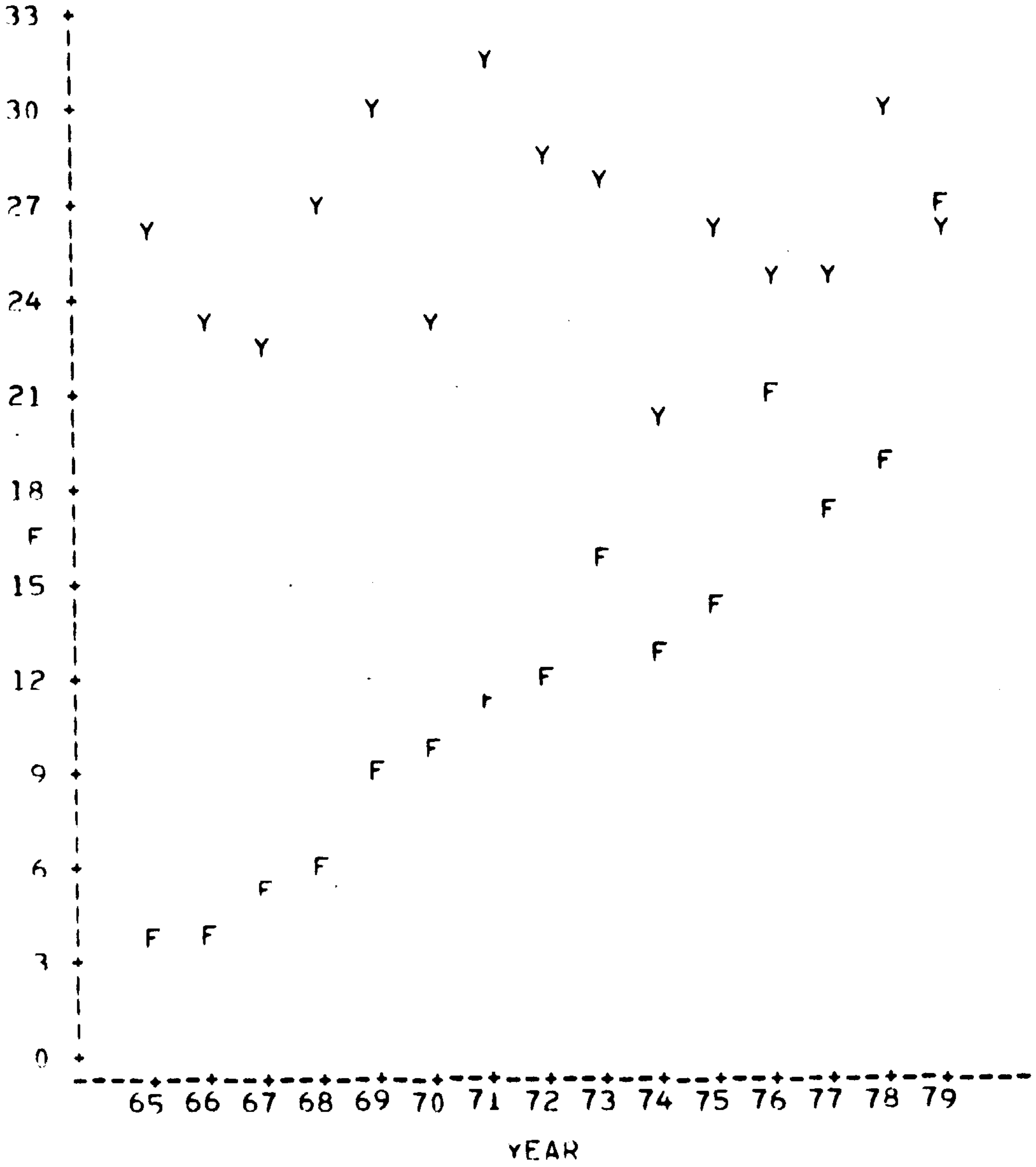
Diagnostic procedures from Regression Diagnostics (Belsley, et al., 1980) indicate that all forms of the Haun model as estimated by the authors have independent variables that are highly correlated. High collinearity among the independent variables makes the variances of the estimated regression coefficients large (although this condition could exist without high collinearity among the independent variables) and therefore reduces the precision of the estimated coefficients. When the model is used for prediction, high collinearity is not a problem if the range and the relationships among the independent variables in the period for which the prediction is made is similar to the model development period. Although collinearity may not be a problem for prediction, it makes interpretation of the estimated regression coefficients difficult since one cannot separate the effect of one variable from another.

When developing the original Haun model, "outlier years" for some crop reporting districts were removed by Haun. The comparison of the performance of the Haun I model (including outliers) and the Haun II model (excluding outliers) showed that the performance of the Haun I model is superior to the performance of the Haun II model. This raises the question of whether or not the removal of outlier years resulted in the loss of information about years that are not unlike years in the period for which predictions were made.

The difficulty in justifying the Haun models for spring wheat in terms of their consistency with scientific knowledge exists because the models are so data-dependent. The quadratic and interaction terms in the models were chosen because of the specific data used in the stepwise regression procedures and not because

Figure 2

NORTH DAKOTA OBSERVED SPRING WHEAT YIELDS(Y)
AND AVERAGE AMOUNT OF FERTILIZER NITOGEN APPLIED(F) BY YEAR
YIELDS IN BUSHELS PER ACRE F IN POUNDS PER ACRE



agronomic science establishes a relationship between them and spring wheat yield. The coefficients of a data-dependent model are difficult to interpret because the variables are not supported by scientific evidence and do not provide the user with an understanding of the basis of the predictions made by the model. For example, if preseason precipitation increases or decreases, it is not clear from the Haun models what the effect on yield would be.

COMPARISON METHODOLOGY

Model Characteristics To Be Compared

The model characteristics used to evaluate crop yield models may also be used to compare the performance of yield models. This section describes the methodology by which the reliability of the yield indications from two models may be compared (see appendix for formulas used to compute indicators). Of course, the previously described indicators of yield reliability may be computed for each model using the bootstrap technique over the same set of test years and the results compared. In addition, where differences in results may be anticipated because of differences in the philosophy or approach used by the model developer, such as the case with the Haun and CEAS models, it is desirable to perform some formal statistical hypothesis testing. Two such tests (parametric and nonparametric) will be described.

Models Are Compared Using Statistical Tests Based on $d = \hat{Y}-Y$

It is desirable to run a statistical test comparing the reliability of competing models. A formal statistical test considers the variability of model performance over time and allows the user to specify an upper limit on the probability of incorrectly declaring one model better than another. This probability is known as α, the level of significance, or the Type I error.

However, although desirable, it is challenging to construct meaningful statistical tests comparing the reliability of two yield models. Only models with some acknowledged degree of success usually reach the stage of formal comparison with other competing models. Therefore, *a priori*, great differences between the reliability of the models are not expected. A powerful statistical procedure is needed which is able to detect small, although important, differences in reliability. Also, the test should be able to function well with relatively small samples of data for each model.

It would appear that an F test could be useful in comparing the mean square errors of two models. However, if the mean square errors are based on seven years of test data and α = .05, then one model's mean square error must be about six times larger than another's before the models can be declared different. This is an unreasonable requirement since models which are in the evaluation process will almost always be more competitive than this.

Another approach could be to consider that one model is considered more reliable than another model if its predicted yields, $\hat{Y}$'s, are closer (on the average) to the observed yields, Y's. No difference in the reliability of two models for a particular year means that the absolute value of the difference between their predicted yields and the observed yield is the same. The absolute value of the difference is used because in assessing yield indication reliability one is equally concerned with overestimates and underestimates. The reliability of a model for that year is related to the amount of the discrepancy, not its direction. We may define $|d_1| = |\hat{Y}_1 - Y|$, $|d_2| = |\hat{Y}_2 - Y|$, and $D = |d_1| - |d_2|$. Then the models are equally reliable in a year for which D equals zero. If D is not equal to zero, one model is more reliable than the other for that year. In formal terms, we want to test the null hypothesis that there is no difference in the reliability of the models over all years. To do so the values of D from the ten test years may be used to compute a test statistic and a decision made whether or not to reject the null hypothesis. Since the results for the models are paired each year, paired-sample statistical tests are used.

Two types of paired-sample statistical tests are used: a parametric test using the student "t" test statistic and a nonparametric test using the Wilcoxon signed rank test statistic. One reason for applying both tests is that they require different assumptions. The parametric t-test assumes the D values are normally distributed while the nonparametric test does not. The d values may be considered to be approximately normally distributed. The $|d|$ values would then be folded normals rather than normally distributed. Although both models are folded at $|d| = 0$, their means may be different and the distribution of D has a possibility of not being normally distributed. The t-test is robust with respect to the normality assumption; however, this possible violation of the assumption is one reason for also running the nonparametric test.

The other reason for running both tests concerns the conditions under which the null hypothesis is rejected by each test. Using the parametric t-test, the basis for rejecting the null hypothesis is the average size of the D values as compared to their variability since the test statistic is the average of the sample D's divided by the sample standard error of the D's. The hypothesis will be rejected and the model with the smaller $|d|$ values declared more reliable if t is large (either positive or negative). However, it is possible that one model could have a smaller $|d|$ value for each of the test years, in other words be very consistent in outperforming the other model, and still the null hypothesis may not be rejected by the parametric test unless the average value of D is large enough. The parametric test implicitly requires that one model have more years with smaller $|d|$ values than the other model and explicitly requires that, on the average, the $|d|$ values be smaller by a sufficient amount before that model may be declared more reliable.

Using the nonparametric test, the null hypothesis will always be rejected if one model has smaller $|d|$ values for each of the test years, regardless of the magnitude of the D values. Therefore, if the models are very competitive in terms of the $|d|$ values each year, but one model consistently, although slightly, outperforms the other model, the nonparametric test will still declare the consistent model to be more reliable.

The hypothesis of equal model performance will only be rejected by the nonparametric test if one model has more years with smaller |d| values than the other model. The model with more smaller |d| values is considered the more reliable model in terms of consistency of performance. However, to reject the null hypothesis and declare one model clearly better than another, consistency of performance is not a sufficient requirement (although it is necessary). Consider the situation in which one model is more consistent than the other but the largest D values occur when the less consistent model performs better. In the few years the less consistent model performs better, it performs much better. A dilemma exists since one model is more consistent than the other but the biggest differences between the models occur when the consistent model performs worst. The null hypothesis will not be rejected and the consistent model will not be declared better if this situation occurs. The null hypothesis will be rejected only if one model is more consistent and the biggest differences between the models occur when the consistent model performs better.

COMPARISON OF HAUN I AND CEAS MODELS FOR YIELD INDICATION RELIABILITY

The Week 12 Haun I Model Is Compared to the August CEAS Model

The Haun I, week 12 model was found to be the best performing Haun model. The earliest state-level prediction date (twelve weeks after the latest CRD's fifty percent planting day) for each of the bootstrap test years is shown in Table 5. Since most of the dates occur around or after the first of August, the August CEAS model, which uses weather data through the end of July, was determined to be the most reasonable CEAS model to use for purposes of comparison (there is no CEAS model which uses August weather data). The predicted yields derived from the bootstrap tests of both models for the test years 1973 to 1979 will serve as the basis for comparison. Although the CEAS model was originally developed using yield expressed in quintals per hectare, for comparative purposes, the predicted yields were converted to bushels per acre.

The indicatiors of yield reliability that will be used to compare the performance of the Haun I model with that of the CEAS model are shown in Table 6. As stated earlier, the indicators based on the differences between $\hat{Y}$ and Y (d = $\hat{Y}$-Y) demonstrate accuracy, precision and bias. The root mean square error for the CEAS model is smaller than the root mean square error for the Haun I model. The smaller value of RMSE for the CEAS model shows that its prediction of yield is more accurate.

The Haun I model has a smaller standard deviation than the CEAS model. The smaller SD values for the Haun I model show that it has higher precision (errors in the predicted yields are closer to their average value than for the CEAS model). The absolute value of the bias of the Haun I model is about twice the absolute value of the bias of the CEAS model. The negative bias of the CEAS model (-0.77) indicates that it tended to underestimate the observed yields during the bootstrap test years while the positive bias of the Haun I model (1.57) indicates that it tended to overestimate the observed yields. The indicators of yield reliability based on d = $\hat{Y}$-Y show that the CEAS model performed slightly better than the Haun I model.

Table 5. Prediction dates (Julian and calendar) for the Haun 1 Model (week 12) for bootstrap test years 1973 to 1979

Year	Julian Date	Calendar Date
1973	216	August 4
1974	243	August 31
1975	233	August 21
1976	219	August 7
1977	210	July 29
1978	227	August 15
1979	235	August 23

Table 6. Indicators of yield reliability for the Haun I model (week 12) and the CEAS model based on independent bootstrap test years 1973 to 1979

Indicator of Yield Reliability (unit)	Models	
	Haun I	CEAS
Bias = B (Bu/A)	1.57	-0.77
Relative Bias = RB (%)	6.1	-3.0
Mean Square Error = MSE $(\mathrm{Bu/A})^2$	5.13	3.97
Root Mean Square Error = RMSE (Bu/A)	2.27	1.99
Relative Root Mean Square Error = RMSE (Bu/A)	8.8	7.8
Variance = Var $(\mathrm{Bu/A})^2$	2.66	3.38
Standard Deviation = SD (Bu/A)	1.63	1.84
Relative Standard Deviation = RSD (%)	6.0	7.4
Number of years out of 7 in which $\lvert rd \rvert > 10\%$ (%)	2	2
rd of Largest $\lvert rd \rvert$ (%)	19.7	14.8
rd of Next Largest $\lvert rd \rvert$ (%)	10.9	-10.0
rd of Smallest $\lvert rd \rvert$ (%)	-0.4	1.6
Number of years out of 6 in which the direction of change from the previous year in the predicted yields agrees with the observed yields	4	4
Number of years out of 4 in which the direction of change from the average of the previous three years in the predicted yields agrees with the observed yields	3	2
Pearson correlation coefficient between observed and predicted yields	.83	.82

Indicators based on relative differences between $\hat{Y}$ and Y (rd=100/Y) demonstrate worst and best performance. The number of years in which the absolute value of the relative difference is greater than 10 percent was 2 out of 7 for both the Haun I model and the CEAS model. Both models performed worst for the same year, 1974, a year in which yields were below average in North Dakota. Indicators of yield reliability based on $|rd|$ show that the Haun I model performed about the same as the CEAS model.

Correspondence between observed and predicted yields is demonstrated by indicators based on $\hat{Y}$ and Y. The number of years in which the direction of change from the previous year in the observed yields is predicted correctly was 4 out of 6 for both the Haun I model and the CEAS model. Both models failed to give a correct prediction of the change in the direction of the observed yields in 1976 and 1977. The Haun I model does slightly better in predicting the direction of change from the average of the previous three years in the observed yields than the CEAS model does. The number of years in which the direction of change from the average of the previous three years in the observed yields is correctly predicted is 3 out of 4 for the Haun I model and 2 out of 4 for the CEAS model. Figure 3 illustrates the direction in which the predicted yields changed (for both models) in comparison to the changes in the observed yields.

The Pearson correlation coefficient, r, which gives the correspondence between the predicted yields and the observed yields is about the same for both models (.83 for Haun I and .82 for CEAS). These values of r, which are close to +1 for both models, indicate that the predicted yields from both models have a definite linear statistical relationship with the observed yields. Indicators of yield reliability based on $\hat{Y}$ and Y show that the Haun I model performed about the same as the CEAS model.

The parametric student t-test and the nonparametric Wilcoxon signed rank test, both based on $d = \hat{Y}-Y$, showed no significant difference between the performance of the Haun I model and the performance of the CEAS model. The student t-test had a P value (estimated α value) greater than .50 and the Wilcoxon signed rank test had a P value greater than .25.

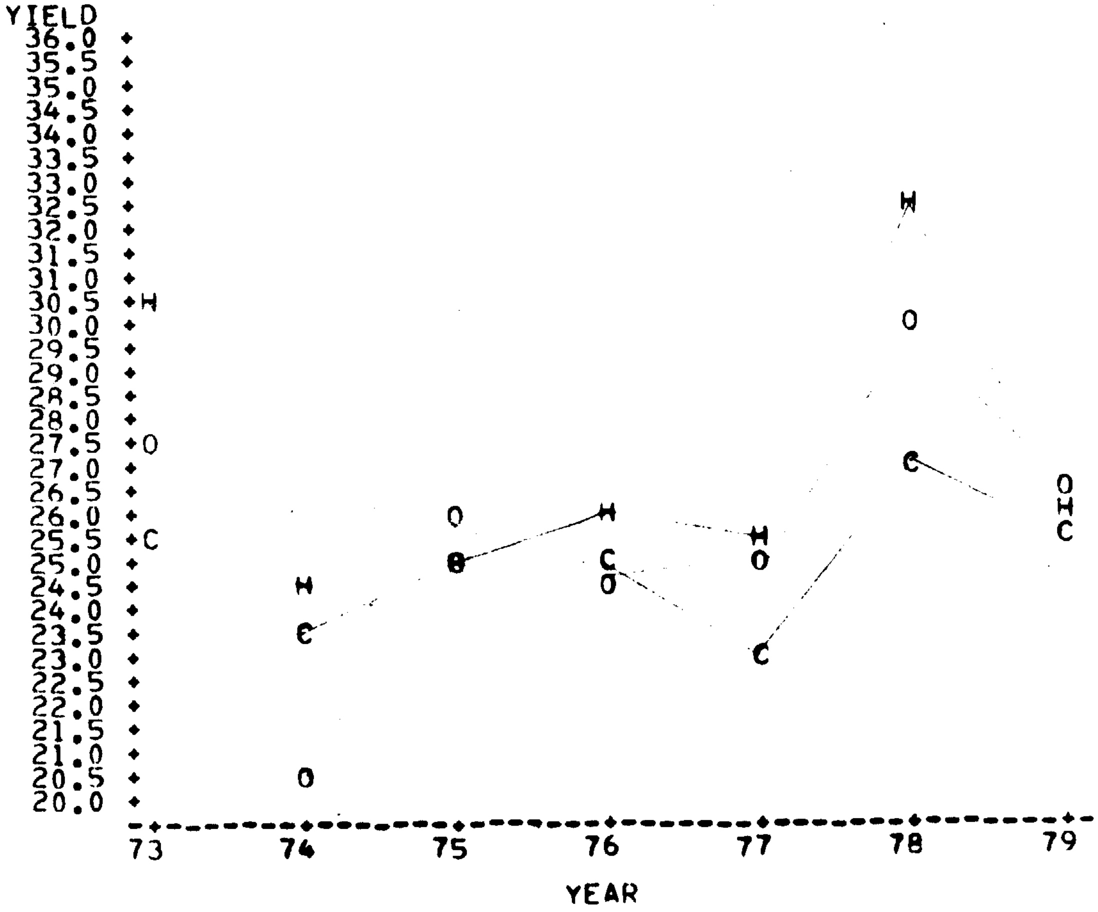
COMPARISON OF FORECASTS MADE BY THE HAUN TYPE I SUBMODELS AND THE CEAS MODELS

The Haun and CEAS models were both developed to predict yield before the end of the growing season of the crop. The date on which the first prediction can be made by the Haun model is four weeks after the occurrence of the latest 50 percent planting date in the nine North Dakota CRDs. For the bootstrap test years 1973 to 1979, the average latest CRD 50 percent planting date for North Dakota is Julian day 143 or May 23. For a given year the prediction dates for the week 16 and 20 Haun submodels may not occur until after the crop is harvested. The first prediction for the CEAS model can be made on March 1. In this section, the Haun Type I submodels and the CEAS models will be reviewed to determine how they compare with the USDA end-of-season observed yield as the growing season progressed for each bootstrap test year.

Plots of the yields predicted by both models for each of the bootstrap test years 1973 to 1979 are in the appendix. The predictions for each model are plotted at

Figure 3

HAUN I MODEL AND CEAS AUGUST 1 MODEL
PREDICTED YIELDS FOR TEST YEARS 1973 TO 1979
(BUSHELS/ACRE)
O=OBSERVED YIELD
H=HAUN 1 C=CEAS



the date for which they were made. The plots show the predictions made by the Haun models for which the prediction date did not occur after September 30 and the predictions made by the CEAS models beginning with April 1. To illustrate the pattern of yield predictions, the 50 percent state planting and combining dates are used as reference points and are plotted as P and C respectively at the level of the USDA observed yield.

Table 7 shows the d values for the Haun and CEAS submodels. These values can be used to objectively determine whether the predicted yields converged in a given year. Convergence is defined here as the continuing movement of the predicted yields toward the observed yield. Convergence implies that $|d|$ stays the same or becomes smaller as the season progresses. If we consider the week 12 Haun model and the August 1 CEAS model to provide the final season estimates, then in order to keep the time periods equivalent we may study the $|d|$ values for the 4, 8, and 12 week Haun models and the June, July, and August 1 CEAS models to determine whether convergence occurred. Results are given in Table 8. The predicted yields converged towards the observed yield in 4 out of 7 bootstrap test years for both models.

Table 8 also shows the $|d|$ value for each model's end-of-season estimate. This allows one to investigate the question of whether the property of convergence is associated with accuracy of yield prediction. The average value of $|d|$ for the years in which the predicted yields converged are 1.98 for the Haun Type I model and 1.85 for the CEAS model. These average $|d|$ values are higher than they are for the years in which the predicted yields did not converge (1.63 for Haun and 1.60 for CEAS).

In summary, the comparison of the forecasts made by the Haun Type I model and the CEAS models have shown that both models had predicted yields that got closer to the observed yield as the growing season progressed in 4 out of 7 of the bootstrap test years. Although the predicted yields converged in these four years, they were not more accurate, on the average, in predicting the end-of-season observed yield than the predicted yields were for the years in which convergence did not occur.

COMPARISON OF THE HAUN I MODEL AND THE CEAS AUGUST 1 MODEL FOR TIMELINESS

The use of the Haun model on a real-time basis requires that the daily weather data and the 50 percent planting dates for the nine North Dakota crop reporting districts be available. To use the CEAS model on a real-time basis the appropriate monthly weather values are required.

Data on the estimated percentage of wheat planted in North Dakota is available during the planting season from the North Dakota State Statistical Office of the USDA. This information is only published at the state level, but can be obtained at the CRD level for use in the Haun model.

Currently, there is a three-month delay to get the published daily weather data for the cooperative weather stations from the National Climatic Data Center in Asheville, North Carolina. This delay could constitute a major handicap in using

Table 7. Differences ($d=\hat{Y}-Y$) between the observed yields and predictions made by the Haun Type I submodels and the CEAS models for test years 1973 to 1979 (bushels/acre)

Year	Model	Submodel				
	Haun Type I	4	8	12	16	20
1973		18.0	3.5	3.0	7.3	6.3
1974		6.6	4.3	4.0	7.5	7.3
1975		-1.9	1.4	-0.8	-1.9	0.6
1976		2.4	5.4	1.4	5.1	4.1
1977		8.2	-0.1	0.7	4.1	3.8
1978		-3.8	1.2	2.8	-1.2	-0.4
1979		-0.9	-0.2	-0.1	2.6	-1.8
	CEAS	April 1	May 1	June 1	July 1	Aug. 1
1973		-1.6	-2.2	-3.0	-2.8	-2.0
1974		10.2	10.6	9.5	5.9	3.0
1975		-0.7	0.9	0.9	2.3	-1.0
1976		2.3	2.3	1.4	0.8	0.4
1977		-0.9	-1.4	3.0	-2.2	-2.0
1978		-0.6	-1.9	1.2	-3.7	-3.0
1979		1.2	1.2	-1.1	-0.2	-0.9

Table 8. Comparison of convergence of the yields predicted by the Haun Type I model and the CEAS models for test years 1973 to 1979

	MODELS			
	Haun Type I		CEAS	
Year	Week 4, 8, 12 models converged	|d| for week 12	June 1, July 1, Aug. 1 models converged	|d| for Aug. 1
1973	Yes	3.0	Yes	2.0
1974	Yes	4.0	Yes	3.0
1975	Yes	0.8	No	1.0
1976	No	1.4	Yes	0.4
1977	No	0.7	Yes	2.0
1978	No	2.8	No	3.0
1979	Yes	0.1	No	0.8

either model on a real-time basis unless the values of the weather variables can be estimated or unless the cooperative weather station data could be obtained in a more timely manner.

Table 5 (page 22) shows the prediction dates for the Haun I model for the test years 1973 through 1979. Assuming that the weather data for both the Haun and CEAS models were available on the date of prediction, the yield could be predicted by August 1 in only 1 out of the 7 test years for the Haun model. In 4 out of 7 test years, the Haun models forecast would be at least 3 weeks later than the CEAS forecast. Although the Haun I week 8 model could provide an earlier indication, it is even less reliable than the week 12 model (see Table 3, page 11).

CONCLUSIONS

Among the (4, 8, 12, 16 and 20 week) submodels of both the Haun Type I model and the Haun Type II model, the week 12 submodel performed best in terms of yield indication reliability. A comparison of the Haun I model (week 12 with outliers included) and the Haun II model (week 12 without outliers) showed that the Haun I model had a superior indication of yield reliability.

Because of its superior performance over the Haun II model, the Haun I model was chosen for comparison with the CEAS model. The results for the comparison of the Haun I model and the CEAS model were mixed. The indicator of yield reliability based on $d=\hat{Y}-Y$ showed that the CEAS model performed slightly better than the Haun I model. The indicators based on $rd=100d/Y$ and based on the comparison of $\hat{Y}$ and Y showed about the same level of performance by both models. Tests of statistical hypotheses (parametric and nonparametric) indicated no significant difference between the performance of the two models. Both the Haun Type I model and the CEAS model did not provide reliable preseason forecasts of the observed yields for the bootstrap test years. In terms of timeliness, the Haun model was less timely in 6 out of the 7 years tested. Assuming that the weather data had been available on the date of prediction for both models, the Haun model forecast would not have been available for the USDA's August 1 Crop Report in 6 out of 7 years.

It is difficult to justify the Haun models in terms of their consistency with scientific knowledge. The independent variables in the model (except for the indicator variables representing the CRDs) were chosen exclusively by stepwise regression procedures, and it is difficult to attach a clear agronomic meaning to them.

As stated earlier, the Haun model is less timely, more complex to develop, interpret and operate and requires more data collection than the CEAS model. Therefore, in order to justify the use of the Haun model, its performance would need to be at a higher level. The Haun model does not provide a significantly better forecast, considering its complexity and lack of timeliness. Therefore, use of the Haun model is not justified.

REFERENCES

Bigsby, Fatu G. and Sebaugh, Jeanne L., 1982. A Review of the Model Specifications and the Data Required to Run Haun's Spring Wheat Model in North Dakota. Working Paper, Yield Evaluation Section of SRS/USDA, Columbia, Missouri

Haun, J. R., 1973a. Visual Quantification of Wheat Development. Agronomy Journal 65: 116-120

Haun, J. R., 1973b. Determination of Wheat Growth-Environment Relationships. Agronomy Journal 65: 813-816

Haun, J. R., 1974. Prediction of Spring Wheat Yields from Temperature and Precipitation Data. Agronomy Journal 66: 405-409

Haun, J. R., 1976. Development of Models for Specific Crop Calendar Events. Int. J. Biometeorology 20(3): 261-266

Haun, J. R., 1979. Wheat Models Based on Daily Plant-Environment Relationships. Proceedings of the Crop Modeling Workshop, Columbia, Missouri

LeDuc, Sharon, 1981. Spring Wheat Models for North Dakota and Minnesota. NOAA/ASIC/Models Branch, Columbia, Missouri

SAS Institute Inc. SAS User's Guide: Statistics, 1982 Edition. Cary, N.C.: SAS Institute Inc., 1982. 584 pp.

Sebaugh, Jeanne, L., 1981. Evaluation of the CEAS Trend and Monthly Weather Data Models for Spring Wheat Yields in North Dakota and Minnesota. AgRISTARS Yield Model Development Project, Document YMD-1-4-1(81-11.1).

Thornthwaite, C. W., 1948. An Approach Towards a Rational Classification of Climate. Geographical Review 38: 55-94.

Wilson, Wendell W., Barnett, Thomas L., LeDuc, Sharon K., Warren, Fred B., 1980. Crop Yield Model Test and Evaluation Criteria. AgRISTARS Yield Model Development Project, Document YMD-1-1-2(80-2.1).

APPENDIX

Independent Variables Considered for Inclusion in the Stepwise Regression Procedures for the Modified Haun Models

Linear terms	Quadratic terms	Interaction terms
GRI	$(GRI)^2$	PSUM*PP
PP	$(PP)^2$	PSUM*GRI
PD	$(PD)^2$	GRI*PP
PSUM	$(PSUM)^2$	PD*PP
S	$(S)^2$	PD*PSUM
F	$(F)^2$	S*PSUM
		PD*GRI
		PD*S
		F*PD
		F*PP
		F*PSUM
		F*GRI
		F*S
		S*PP
		S*GRI

APPENDIX - STATISTICAL FORMULAS

Measures of Model Performance

Definition of Terms:

Y_i = Yield as reported by U.S.D.A. for year i (observed yield)

$\hat{Y}_i$ = Yield as predicted by a model for year i.

$d_i = \hat{Y}_i - Y_i$ = difference between predicted and observed yield for year i.

$rd_i = 100\ d_i/Y_i$ = relative difference for year i.

$i = 1, \ldots, n$ = number of test years and $\Sigma = \sum_{i=1}^{n}$ = summation over the test years.

$\overline{Y} = 1/n\ \Sigma\ Y_i$ = average observed yield.

Measures:

Bias = B = $1/n\ \Sigma\ d_i = \bar{d}$.

Relative Bias = RB = $100\ B/\overline{Y}$.

Mean Square Error = MSE = $1/n\ \Sigma\ d_i^2$.

Root Mean Square Error = RMSE = $(MSE)^{\frac{1}{2}}$.

Relative Root Mean Square Error = RRMSE = $100\ RMSE/\overline{Y}$.

Variance = Var = $1/n\ \Sigma\ (d_i - \bar{d})^2$.

Standard Deviation = SD = $(Var)^{\frac{1}{2}}$.

Relative Standard Deviation = RSD = $100\ SD/(\overline{Y} + \bar{d})$.

Mean Square Error = Variance + $(\text{Bias})^2$,

or

Accuracy = Precision + $(\text{Bias})^2$.

Pearson r between $\hat{Y}_i$ and Y_i:

$$r = \left[\Sigma \hat{Y}_i Y_i - \frac{(\Sigma \hat{Y}_i)\ (\Sigma Y_i)}{n} \right] \left[\left(\Sigma \hat{Y}_i^{\,2} - \frac{(\Sigma \hat{Y}_i)^2}{n} \right) \left(\Sigma Y_i^{\,2} - \frac{(\Sigma Y_i)^2}{n} \right) \right]^{-\frac{1}{2}}$$

Paired-Sample Statistical Tests Comparing the Performance of Two Crop Yield Models

Definition of Terms:

$\hat{Y}_{1_i}$ = Yield as predicted by model 1 for year i.

$\hat{Y}_{2_i}$ = Yield is predicted by model 2 for year i.

$|d_{1_i}| = |\hat{Y}_{1_i} - Y_i|$ = Absolute value of the difference between model 1 predicted and observed yield for year i.

$|d_{2_i}| = |\hat{Y}_{2_i} - Y_i|$ = Absolute value of the difference between model 2 predicted and observed yield for year i.

$D_i = |d_{1_i}| - |d_{2_i}|$.

Rank $(|D_i|)$ = Ranks of the absolute values of D_i assigned in ascending order (smallest value of $|D_i|$ = rank 1, ..., largest value of $|D_i|$ = rank n). If two or more years have the same value for $|D_i|$, assign each year the average of the ranks.

Parametric Test - Student t:

H_0: $\mu_D = 0$

H_a: $\mu_D \neq 0$

Test Statistic = $t = \frac{\bar{D}}{s_{\bar{D}}}$, where

$\bar{D} = 1/n \, \Sigma D_i$,

$s_{\bar{D}} = (s_D{}^2/n)^{\frac{1}{2}}$, and

$s_D{}^2 = [\Sigma D_i{}^2 - 1/n(\Sigma D_i)^2]/(n-1)$.

Reject H_0 if $|t| > t_{\alpha,\ (n-1)}$.

Nonparametric Test - Wilcoxon Signed Rank:

H_0: There is no difference in the performance of the models.

H_a: There is a difference in the performance of the models.

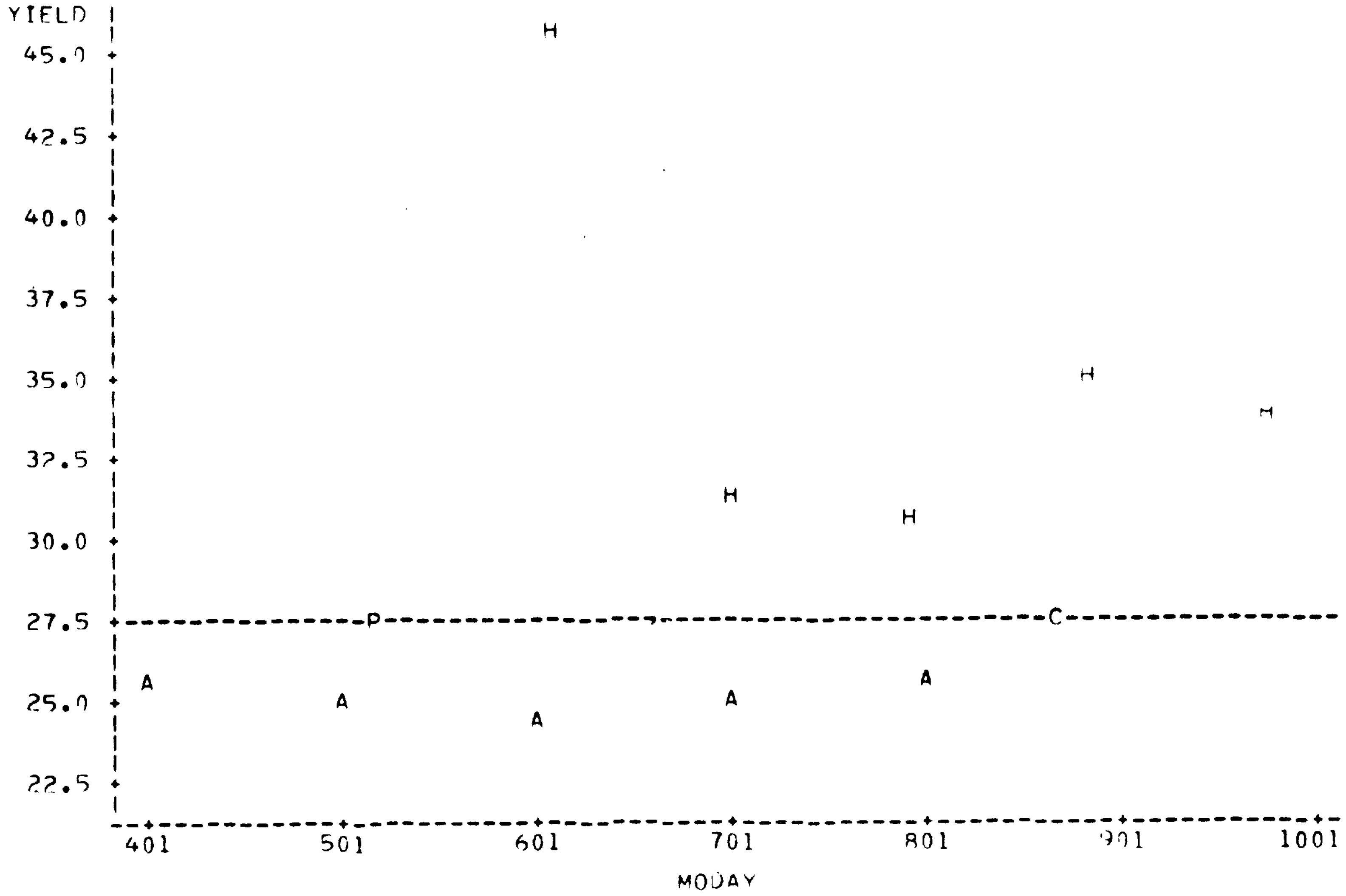
Procedure to compute test statistic, T:

1. Compute the D_i.
2. Assign ranks to $|D_i|$.
3. Assign signs to Rank ($|D_i|$) corresponding to the signs of D_i.
4. Let T = the absolute value of the sum of the ranks with the less frequent sign.

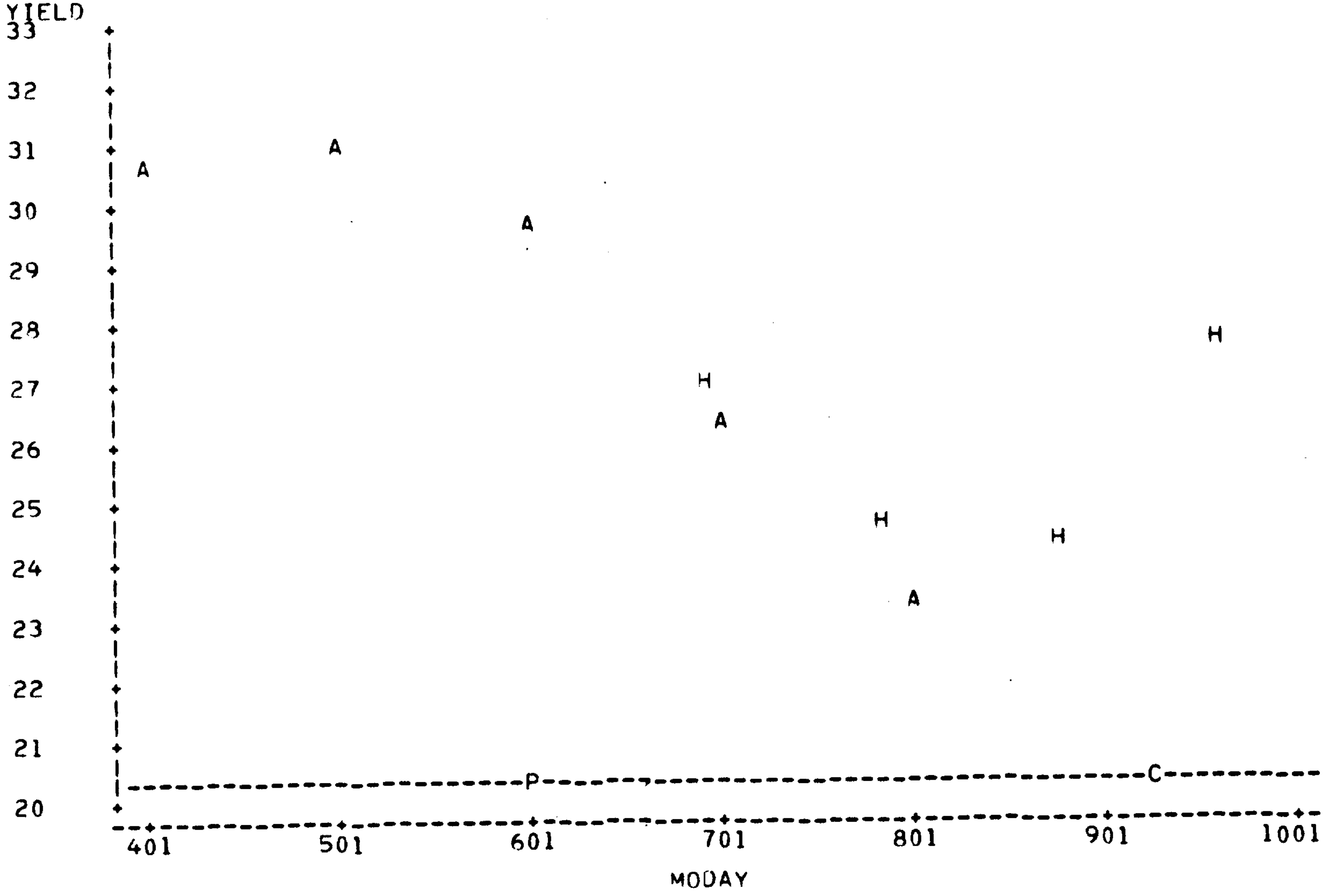
Reject H_0 if $T \leq T_{\alpha(1\text{ tailed}),\ n}$.

APPENDIX

OBSERVED AND PREDICTED SPRING WHEAT YIELDS,
NORTH DAKOTA, TEST YEAR 1973 (BUSHELS/ACRE)
DATE OF PLANTING IS P AND DATE OF COMBINING IS C
LINE REPRESENTS USDA YIELD ESTIMATE
H=HAUN I MODEL PREDICTED YIELD A=CEAS MODEL PREDICTED YIELD

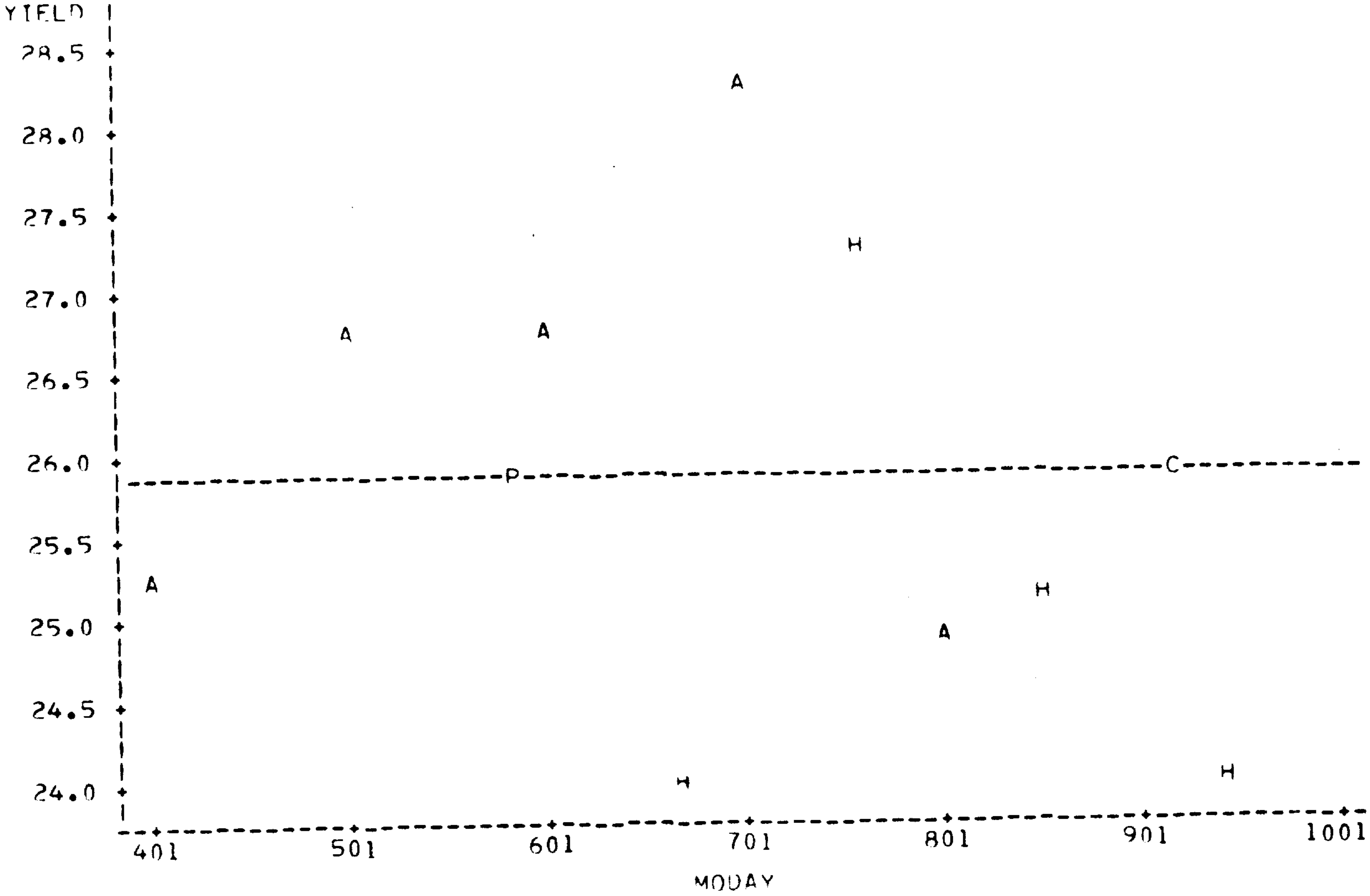


APPENDIX
OBSERVED AND PREDICTED SPRING WHEAT YIELDS
NORTH DAKOTA, TEST YEAR 1974 (BUSHELS/ACRE)
DATE OF PLANTING IS P AND DATE OF COMBINING IS C
LINE REPRESENTS USDA YIELD ESTIMATE
H=HAUN I MODEL PREDICTED YIELD A=CEAS MODEL PREDICTED YIELD



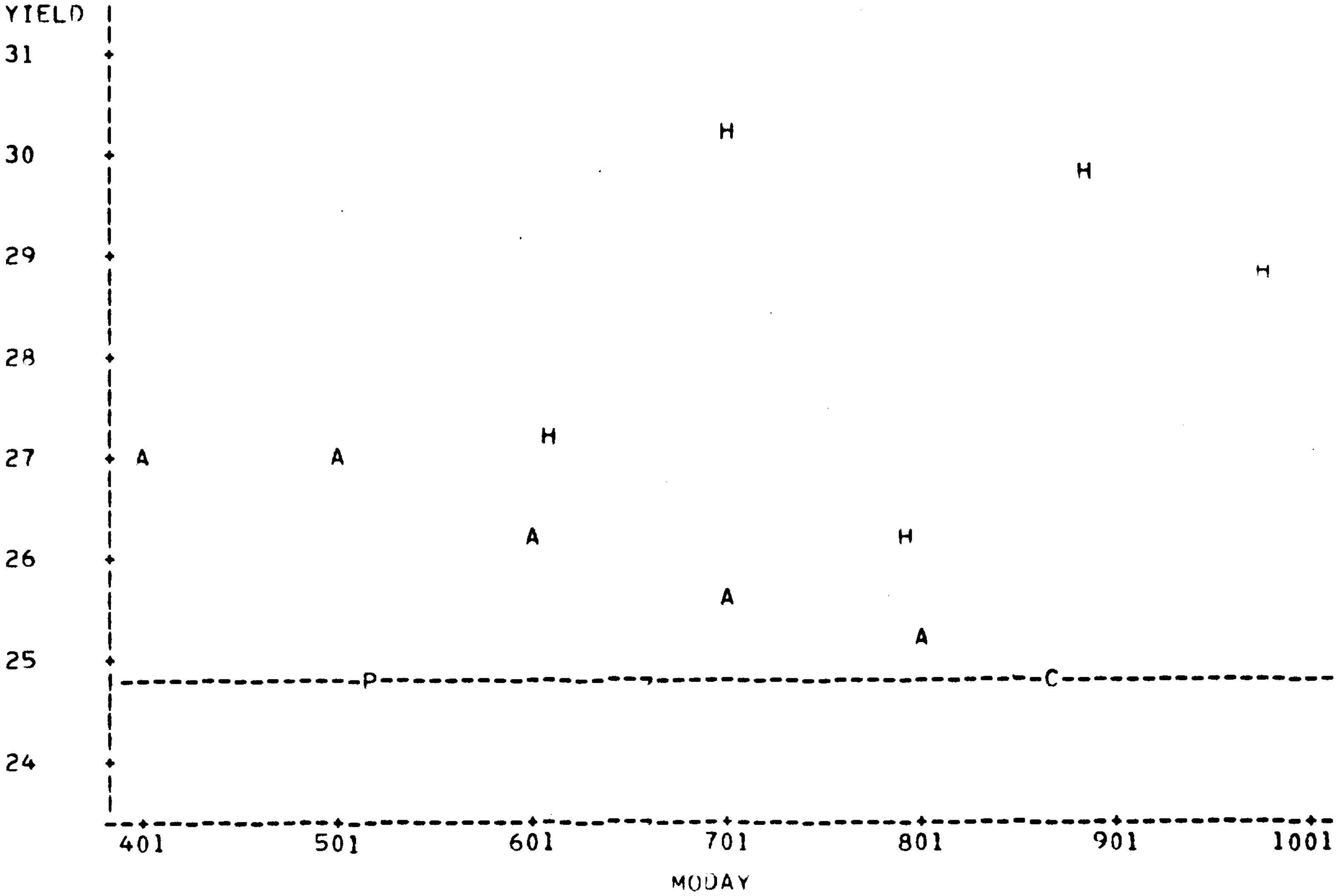
APPENDIX

OBSERVED AND PREDICTED SPRING WHEAT YIELDS
NORTH DAKOTA, TEST YEAR 1975 (BUSHELS/ACRE)
DATE OF PLANTING IS P AND DATE OF COMBINING IS C
LINE REPRESENTS USDA YIELD ESTIMATE
H=HAUN I MODEL PREDICTED YIELD A=CEAS MODEL PREDICTED YIELD



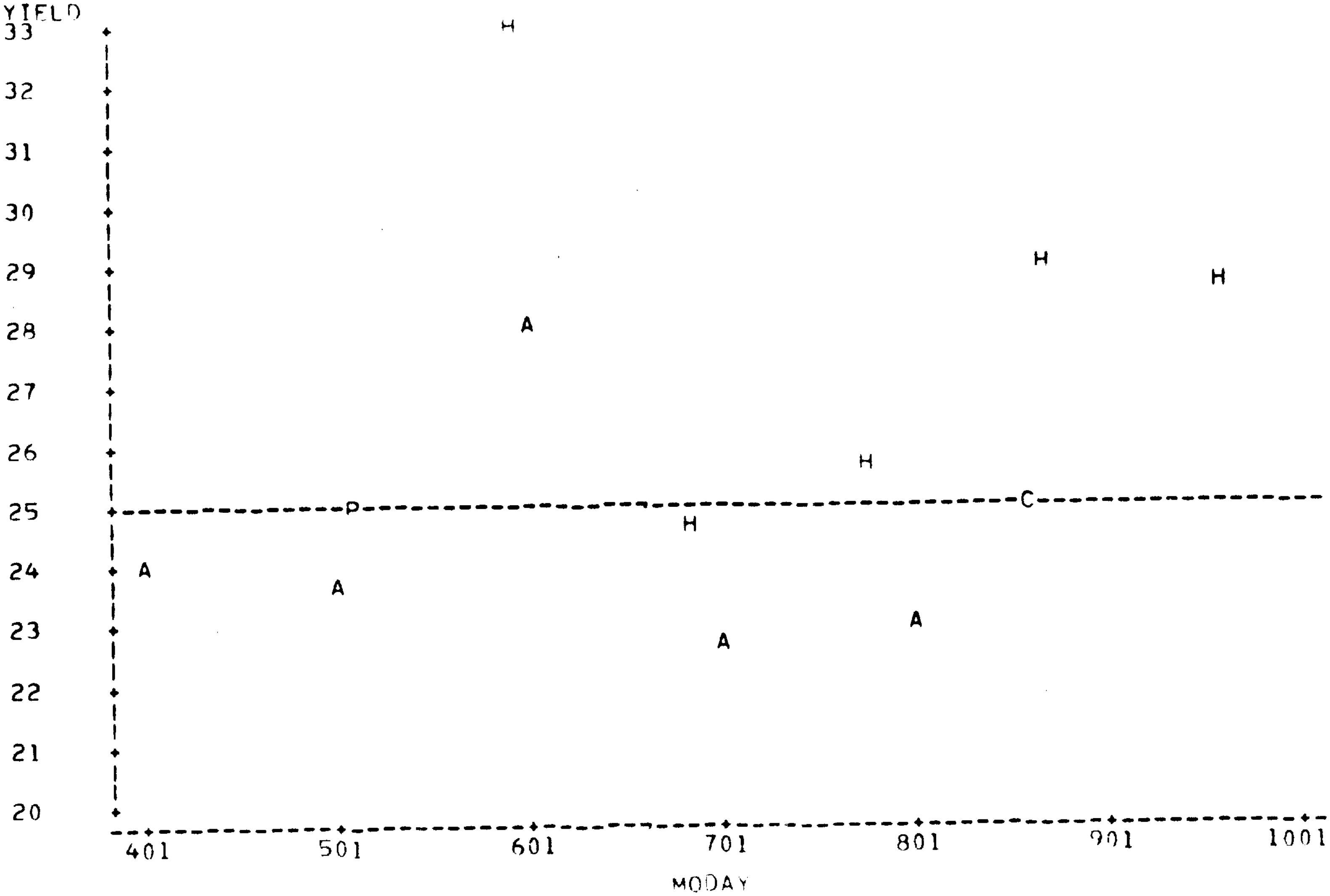
APPENDIX

OBSERVED AND PREDICTED SPRING WHEAT YIELDS
NORTH DAKOTA, TEST YEAR 1976 (BUSHELS/ACRE)
DATE OF PLANTING IS P AND DATE OF COMBINING IS C
LINE REPRESENTS USDA YIELD ESTIMATE
H=HAUN I MODEL PREDICTED YIELD A=CEAS MODEL PREDICTED YIELD

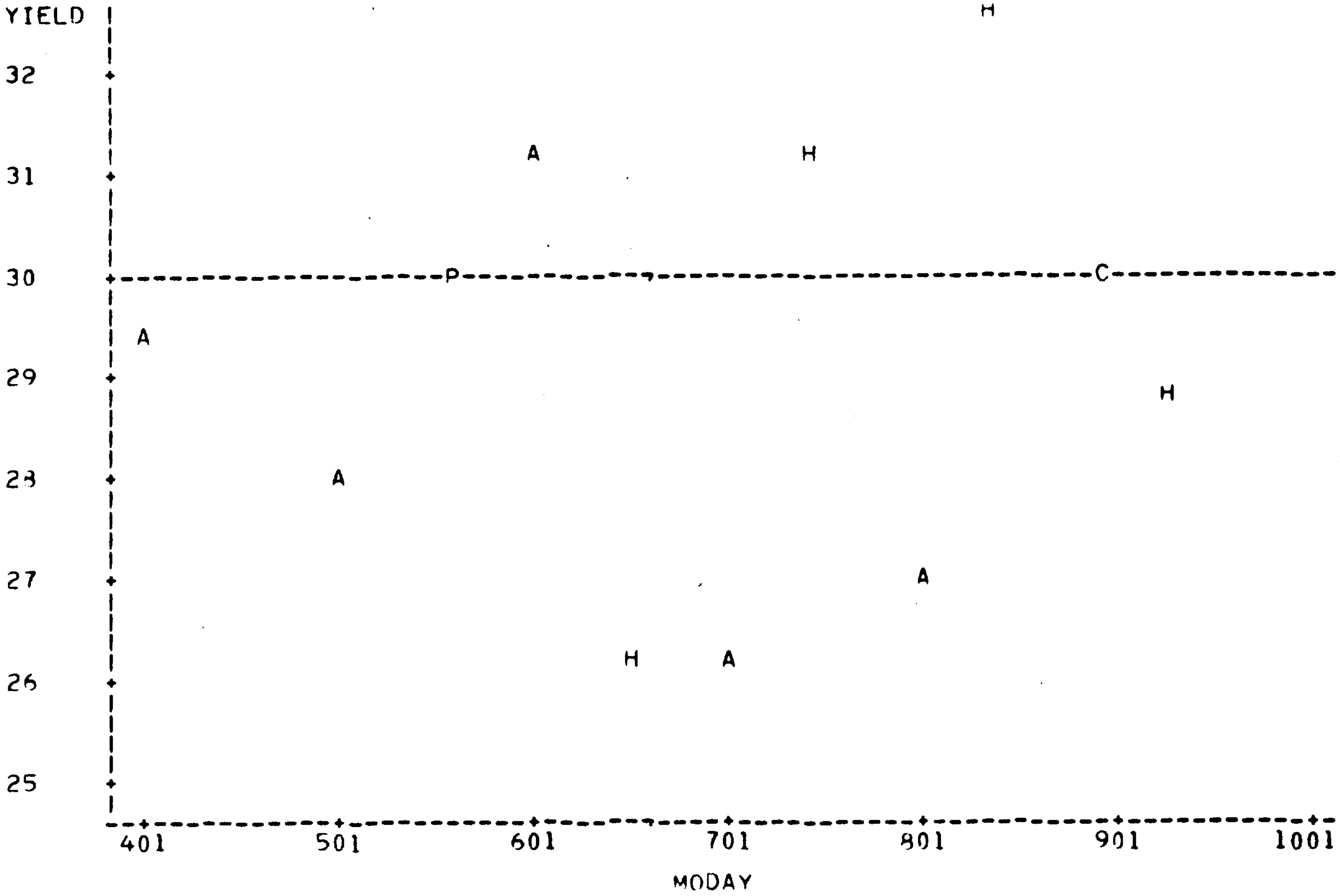


APPENDIX

OBSERVED AND PREDICTED SPRING WHEAT YIELDS
NORTH DAKOTA, TEST YEAR 1977 (BUSHELS/ACRE)
DATE OF PLANTING IS P AND DATE OF COMBINING IS C
LINE REPRESENTS USDA YIELD ESTIMATE
H=HAUN I MODEL PREDICTED YIELD A=CEAS MODEL PREDICTED YIELD



APPENDIX
OBSERVED AND PREDICTED SPRING WHEAT YIELDS
NORTH DAKOTA, TEST YEAR 1978 (BUSHELS/ACRE)
DATE OF PLANTING IS P AND DATE OF COMBINING IS C
LINE REPRESENTS USDA YIELD ESTIMATE
H=HAUN I MODEL PREDICTED YIELD A=CEAS MODEL PREDICTED YIELD



APPENDIX

OBSERVED AND PREDICTED SPRING WHEAT YIELDS,
NORTH DAKOTA, TEST YEAR 1979 (BUSHELS/ACRE)
DATE OF PLANTING IS P AND DATE OF COMBINING IS C
LINE REPRESENTS USDA YIELD ESTIMATE
H=HAUN I MODEL PREDICTED YIELD A=CEAS MODEL PREDICTED YIELD