United States
Department of
Agriculture

**Statistical
Reporting
Service**

**Statistical
Research
Division**

Forecasting Yields Using Weather-Related Indices

in Cooperation with Texas Agricultural Experiment Station, Blacklands Research Center, Temple, Texas

Stephan J. Maas

Forecasting Yields Using Weather-Related Indices. By Stephan J. Maas. Statistical Research Division, Statistical Reporting Service, U.S. Department of Agriculture, Washington, D.C. 20250. March 1982. SRS Staff Report No. AGES820317.

ABSTRACT

A technique for including the effects of weather in objective yield forecasts is developed and evaluated. The forms of the regression-type yield component models in current use are not changed. Rather, weather-related indices are developed to stratify historic objective yield data used to build the models. The objective is to stratify the data so that it will be more representative of growing conditions during the forecast year than the 3- or 5-year sequences of data selected using the current technique. Two stratification techniques are investigated--one in which conditions for individual years are compared to the forecast year and another in which sub-groups of data are selected from a pooling of all historic data. The proposed techniques are applied to seven years of winter wheat data from Kansas and the results of forecasts are compared to those made using the current technique.

Keywords: Wheat objective yield forecasts, stratification by historic weather, multiple regression.

FOREWORD

Stephen Maas conducted this research while with the Yield Research Branch as an Intergovernmental Personnel Act (IPA) scientist from Texas A&M University, Texas Agricultural Experiment Station. His academic and professional experience is in agricultural meteorology and related research.

The results of procedures reported in this publication will be subjected to further analysis to determine if there are statistically significant findings. Analysis performed and conclusions will be reported in a separate addendum publication. Those receiving this publication will also receive the addendum.

CONTENTS

FORECASTING YIELDS USING WEATHER-RELATED INDICES

Stephan J. Maas

SUMMARY

The performance of wheat objective yield component forecast models in Kansas was evaluated using three different methods of selecting the historic objective yield data used in constructing the models. The first, currently operational, procedure was to use all of the objective yield data obtained during the previous three years. The second procedure utilized between year stratification (BYS) of the previous years of data and selected the three years where the weather was most similar to the current year. The third procedure used within year stratification (WYS) to stratify all historic objective yield data into subgroups based on indices of weather values. The subgroups which were most similar to the current year were then chosen to construct the component forecast models.

Disregarding one anomalous year (1980), both methods of stratification resulted in a smaller average absolute forecast error for number of heads and for weight per head than did the current operational procedure of using the three previous years. However, in their present forms, both the BYS and WYS techniques produced yield forecasts with about the same degree of error. Although the WYS technique represents a more sophisticated and potentially more accurate method for forecasting, it suffers from an unavailability of soil characteristics and precipitation data at the sample locations. Until this information is collected, it is unlikely that the accuracy of the WYS technique will markedly surpass that of the BYS technique.

For those states with a sufficient historic data record, the BYS technique may provide an increase in accuracy over the present 3 or 5 previous year current method, and with a nominal expenditure of resources and effort. The ability to select years similar to the forecast year is enhanced by having many years to choose from. This also allows known anomalous years to be excluded from the analysis. The technique might not result in improved forecasts every time it is used. However, it does appear to reduce the probability of producing inaccurate forecasts when there are significant year-to-year variations in weather conditions that have an impact on crop yield.

INTRODUCTION

One of the major responsibilities of the Statistical Reporting Service, U.S. Department of Agriculture, is the preparation of crop yield forecasts prior to harvest. The objective yield forecasting technique currently utilized for the major field crops uses statistical models to relate plant characterisitcs observed during the growing season to yield at harvest. These models are constructed using data obtained during previous years. It is assumed that whatever universal factors that determined yield during the years used to build the model also are in effect during the forecast year.

It has been recognized that the year-to-year variability in weather might lead to significant inaccuracies in yield forecasts produced by the current technique. This is particularly true when the weather during the forecast year is markedly different from that experienced during the years used to construct the yield model. Although the use of weather data to forecast and estimate crop yields has been investigated, no generally-applicable technique for incorporating the effects of weather into the objective yield forecasting technique has been established.

The purpose of this report is to describe a method for stratifying objective yield data prior to constructing the statistical models used in the current yield forecasting technique. Stratification is performed based on the values of indices derived from weather-related environmental conditions. The applicability of such a technique is demonstrated using seven years of winter wheat data from Kansas.

YIELD FORECASTING TECHNIQUES

For the major field crops, the objective yield forecasting technique represents the accepted level of sophistication in predicting yields. This section provides a decription of the current technique and the proposed modification to allow for the effects of weather in the yield forecast.

Current Technique

In the current objective yield forecasting technique, a yield component (such as the number or weight of fruiting bodies) at harvest is related to some plant characteristic (such as the number or weight of vegetative organs) observed earlier in the growing season through a regression-type model. Yield component forecasts are made on a plot basis so that the final yield (the product of the yield components) can be expanded to a per-acre value. The regression coefficients in a forecast model are determined by relating observed yield components to observed plant characteristics obtained from historic data. Generally, data from the three years directly preceding the forecast year are pooled to construct the yield component models.

Some yield components are forecast early in the growing season before the appearance of any plant characteristics which can be used in a regression-type model. In these cases, the forecast model is reduced to a historic average value of the yield component. Generally, data from the five years directly preceding the forecast year are pooled to calculate this average value.

Technique Development

The technique developed to incorporate the effects of weather into the yield forecast does not affect the form of the yield component models. Rather, it involves the manner in which historic objective yield data are selected for constructing the models.

When objective yield data are plotted to show some component of yield versus an observed plant characteristic, one typically observes a dispersion of points across the graphical space. Many factors act to produce the scatter among the data, but analysis indicates that certain weather conditions have a more regular effect on distributing the points. Figure 1 shows a plot of winter wheat head numbers versus stalk numbers over a unit area for the years 1975 and 1977 in Kansas. Although there is some overlap of the point distributions for the two years, it is apparent that more of the 1977 data occupy the upper edge of the combined distribution and more of the 1975 data occupy the lower edge of the combined distribution. The growing season in 1975 was characterized by adequate rainfall, while conditions in 1977 were very dry. Thus, one might take advantage of this behavior of the data to improve yield forecasts based on weather conditions during the forecast year.

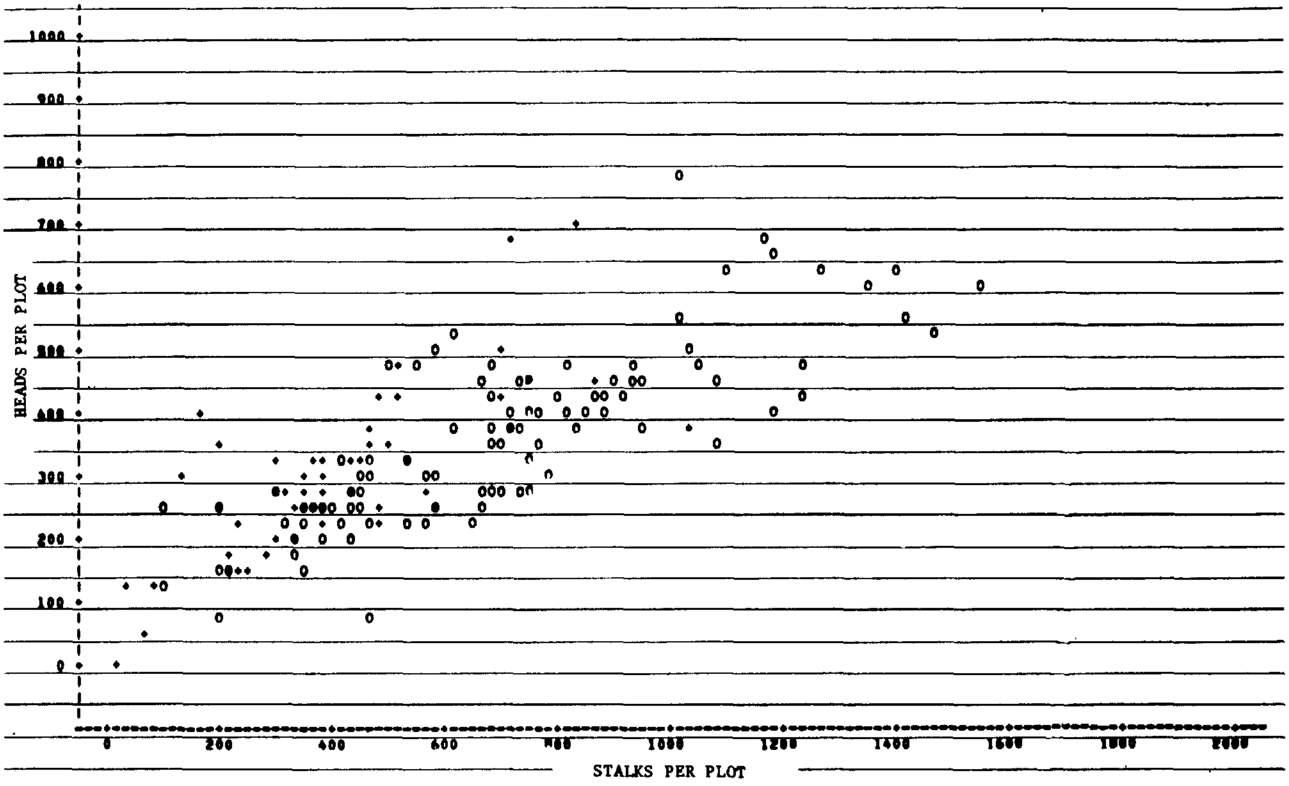

Figure 1. Plot of winter wheat head numbers versus stalk numbers for 1975 (open circles) and 1977 (crosses) in Kansas (solid circles represent data from both years).

Instead of arbitrarily choosing a set of objective yield data to build the yield models, data may be selected that are associated with weather conditions similar to those experienced during the forecast year. This would require that the historic data be stratified according to the value of some weather-related criterion established prior to the point in the growing season at which the forecast is made. In its simplest form, the criterion could be represented by one or more index values associated with each datum in the historic set. These indices, computed from weather-related variables, would be used to stratify or select historic data for building the yield component models.

Two methods of stratifying data are examined in this study. The index values associated with objective yield samples within a state may be averaged on a yearly basis, thus providing an average indicator of sampled growing conditions within the state each year. These average values can then be compared to an average index value computed for the forecast year, and three years can be selected which are judged to be similar to the forecast year based on their average index values. The data from these three years can be pooled to construct yield component models. This method can be termed <u>between-year stratification</u> (BYS). A second method would initially allow historic data from all years to be pooled and then stratified into sub-groups based on values of indices associated with the individual objective yield samples. These subgroups of data would be used to construct yield component models for use in the forecast year. This method can be termed <u>within-year stratification</u> (WYS). A detailed explanation of how these stratification methods can be applied to actual forecasts is presented in the following sections of this report.

PROJECT DESIGN

To examine the applicability of the indexing and stratification techniques, it is necessary to design a project which meets two basic requirements. First, the geographic area of interest should exhibit significant areal and year-to-year variability in weather. Second, the record of objective yield data for the area of interest should be long enough to capture the effects of the various weather conditions.

After considering a number of states, it was decided that Kansas most adequately met these requirements. Being centrally located in the Great Plains, Kansas is swept by weather systems that span the range from cold to hot and moist to dry. As shown in Figure 2, Kansas contains three generalized climatic regions. The western third of the state is semiarid, with sporadic precipitation and high potential evaporation. The central portion of the state is subhumid, with precipitation generally balancing the evaporative demand of the atmosphere. The extreme eastern portion of Kansas is humid, with precipitation generally exceeding evaporation. These zonations refer primarily to long-term conditions--any individual year may witness extremes of moisture or drought over the state.

Geologically, Kansas has relatively low relief and arable soils over its entire width. This allows certain crops to be grown all over the state. Of these, winter wheat was selected as the subject of this effort. At the time of this study, seven years (1974 through 1980) of detailed objective yield data were available for Kansas. This time period was of sufficient length to contain some favorable and unfavorable years for winter wheat production.

The forecast models for winter wheat are of the types described previously. In the case of this crop, head number per sampled area and average head weight at harvest are the yield components. The accuracy of the models increases as the growing season progresses, since the plant characteristics upon which the forecasts are based are physiologically more closely related to the final yield components as harvest is approached. Thus, it was decided to concentrate on evaluating the application of the techniques to the early-season (prior to heading) forecasts. In this case, forecast head number (Y) is related to observed stalk number (X) through the regression equation

$$Y = a + bX \qquad (1)$$

where a and b are parameters evaluated from historic objective yield data. The forecast of head weight (W) can be expressed through the simple relation

$$W = c \tag{2}$$

where c is the average head weight obtained from historic objective yield data. It is the selection of the historic data for determining the parameters a, b and c that is addressed in this study.

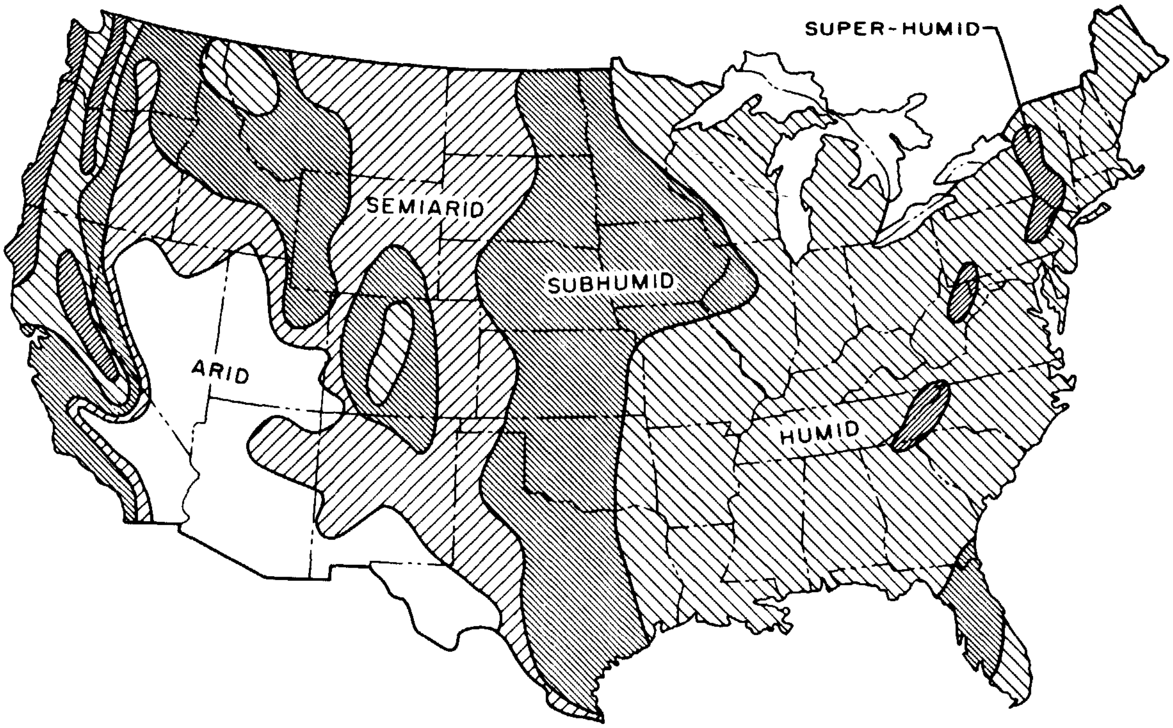

Figure 2. Generalized climatic regions of the contiguous United States (Bureau of Reclamation, 1977).

DATA ACQUISITION

As stated earlier, seven years (1974-1980) of detailed objective yield data were available for Kansas. For winter wheat, the important variables in this data record are head number, stalk number, head weight and sample number. The sample number was used to identify the county in which the sample was located. Exact location of the sample within a county was not possible based on the information in the archived data set. The number of objective yield samples in each county on a yearly basis is presented in the Appendix to this report.

There are over 400 stations in Kansas capable of providing some forms of weather information (Brown, 1978). Daily observations of precipitation were obtained from approximately 300 stations for the period up to 1979. For 1979 and 1980, precipitation data were obtained from approximately 150 stations across the state. Daily maximum and minimum temperature data for the entire period were obtained from approximately 150 stations. The sources of the data were records and summaries generated by the National Weather Service.

WEATHER INDEX DEVELOPMENT

The identification of indices is assisted by a knowledge of the response of plants to their environment. Plants need water, heat, light, air and nutrients to grow--this list of basic requirements provides a starting point in the search for indices for use in the proposed methodology. Some of these requirements, such as light and air, are generally not limiting to plant growth in the field. Nutrients, both what occurs naturally in the soil and what is applied by the farmer, can have a significant effect on growth and yield. Although water can be supplied by irrigation and heat modified by mulching, for a large number of agricultural situations the impact of these requirements can (at least qualitatively) be assessed from weather data. The remainder of this section describes the development of specific weather-related indices resulting from this general knowledge of plant response.

Soil Moisture

The availability of moisture places the major restriction on wheat yields in the dryland crop areas of the Great Plains. Baier and Robertson (1968) showed that wheat yields were more closely related to estimates of soil moisture than to climatic variables such as rainfall or temperature. Baier and Robertson (1967) demonstrated that not only total yield but also yield components (head number, grain number and grain weight) for wheat were related to soil moisture estimates obtained from standard weather data.

For soil moisture to be used as an index, a value must be determined prior to the point in the growing season at which the forecast is made. Early-season forecasts for winter wheat in Kansas are made on May 1, so soil moisture values determined no later than the end of April may serve as indices for this forecast. The value of soil moisture early in the spring has somewhat of a predictive nature, in that high spring soil moisture usually indicates an adequate reserve for plant growth later in the season, while low values of soil moisture in the early spring indicate a likelihood of drought later in the season. Bole and Pittman (1980) showed that estimates of spring soil moisture were significantly related to final yields for barley.

If sufficient weather data are available, the simplest way to estimate soil moisture at a location is to model the balance over time between water entering and leaving the soil. This can be expressed by the equation

$$SW = SW' + P - ET - D - R \qquad (3)$$

in which SW is the current soil moisture, SW' is the soil moisture at some previous time, and P, ET, D and R are, respectively, the precipitation, evapotranspiration, sub-surface drainage, and surface runoff occurring between the current and previous times. Many such soil moisture balance models exist today, and they operate on a wide range of temporal and areal scales.

For this study, the inability to locate objective yield samples within their respective counties essentially dictates that the areal scale of such a soil moisture balance model be on the county level. As illustrated by Strand (1981), there is a general correspondence between the areal and temporal scales of models. A model representative of a county-size area should operate with a time scale increment of a week to a month. To reduce the computational burden, a monthly time scale increment was chosen for the model used in this study.

The choice of these areal and temporal scales allows the soil moisture balance model expressed by Equation 3 to be simplified to the form

$$SW = SW' + P - ET - D \tag{4}$$

Here, surface runoff has been assumed to be negligible. Also, when the sum of the first three terms (SW', P, -ET) on the right side of Equation 4 is less than field capacity (the maximum amount of plant-available water that the soil can hold), loss of water from the root-soil system through drainage is assumed to be zero. When the sum of the three terms exceeds field capacity, D is set equal to the difference between field capacity and the sum of the three terms. Thus, the value of SW never exceeds field capacity. The relative soil water (RSW) can be defined as the ratio of SW to field capacity. For the model described above, the value of RSW can range from 0 to 1, inclusive.

Of the terms in Equation 4 that determine the monthly change in soil moisture, only ET requires more than trivial computation. The value of ET is determined from the relationship

$$ET = r\,(PET) \tag{5}$$

where PET is the potential evapotranspiration and r is a parameter related to field conditions. Evaluation of PET follows a method described by Moe and Griffiths (1965), who found that monthly pan evaporation could be adequately estimated using a simple linear regression based on monthly average maximum air temperature. Thus, the expression for PET takes the form

$$PET = (\alpha + \beta T)\,K \tag{6}$$

in which T is the average maximum temperature, α and β are regression parameters, and K is a constant relating pan evaporation to free evaporation from a large water surface (assumed to represent potential evapotranspiration). The values of α and β are determined by regressing historic pan evaporation data onto temperature data. The value of K depends upon the type of pan used to obtain evaporation data.

As shown by Moe and Griffiths (1965), the values of α and β are not constant over a state. Four years (1976-1979) of data from 24 stations in Kansas which reported both evaporation and temperature were used to evaluate the slope and intercept parameters. Results of this analysis are presented in Figures 3 and 4, which show fields of isopleths constructed from the values of α and β determined at the 24 station locations. When used in Equation 6, these parameter values relate monthly average maximum temperature in degrees Fahrenheit to monthly PET in inches of water. Since standard Class A pans were used at these stations, a value of 0.77 suggested by Bloodgood et al. (1964) was used for the constant K in Equation 6.

The value of the parameter relating ET to PET in Equation 5 is not a constant but depends on the value of RSW in the portion of the soil occupied by roots (root zone). Based on data obtained by Meyer and Green (1980) for wheat, the functional form of r is illustrated in Figure 5. The depth of the root zone was allowed to vary in the model according to the time of year. Kmoch et al. (1957) showed that, in deep loamy soils in Nebraska, wheat roots generally did not penetrate below 5 feet prior to winter dormancy of the crop. However, after growth had resumed in the spring, roots could in some cases be found to a depth of 10 feet. Thus, the soil moisture balance model makes use of a two-level soil profile. During the period from August through March, water is allowed to be extacted through ET only from the upper 5 feet of the soil profile. The lower 5 feet of soil is used to store any moisture from precipitation that remains after the upper layer has been filled to field capacity. If this results in the lower layer also being filled, the excess is lost as drainage. During the second part of the growing season (April through July), the two layers are combined and ET is extracted from the entire 10 feet of soil. When the 10-foot layer is again divided into two layers at the end of the growing season, the remaining soil moisture is equally partitioned into the upper and lower layers.

The value of field capacity at a location is dependent on a number of soil characteristics, such as soil type and depth. Main (1979) provided estimates of field capacity in the upper 5 feet of the soil profile for Crop Reporting Districts in Kansas and other states (see Figure 6). Based on the simplicity of the soil moisture balance model used in this study, it was assumed that the value of field capacity in a Crop Reporting District (CRD) was representative of values in each of the counties constituting the CRD. It was also assumed that the field capacities of the upper and lower 5-foot layers of the soil profile were the same.

Values of monthly accumulated precipitation P and average maximum temperature T for use in Equations 4 and 6 were obtained from the climatic data described in the previous section of this report. Initially, these data were processed to produce monthly values from the daily observations contained in the historic record.

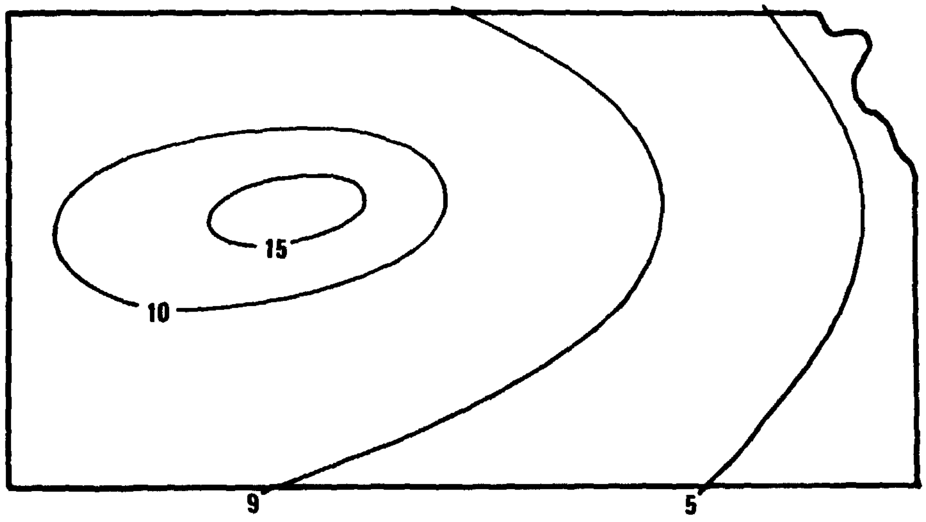

Figure 3. Intercept value in the regression between evaporation and maximum temperature in Kansas.

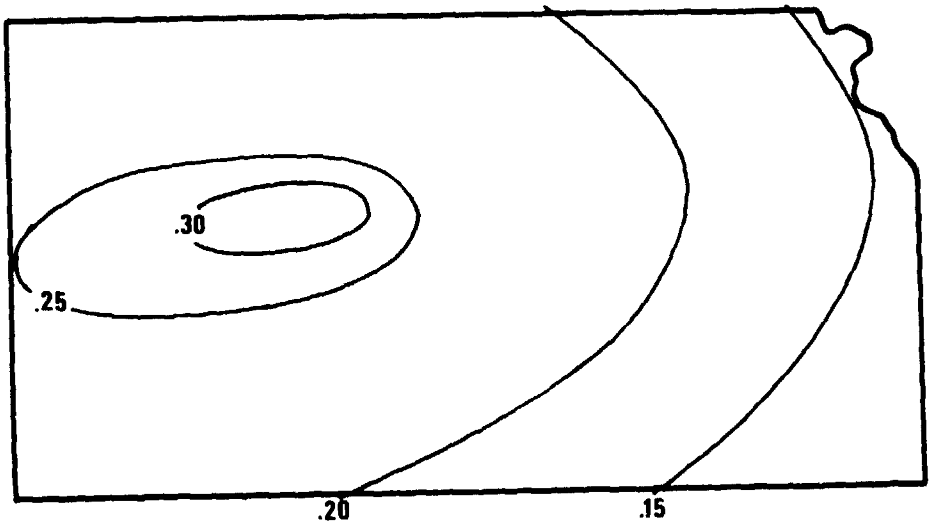

Figure 4. Slope value in the regression between evaporation and maximum temperature in Kansas.

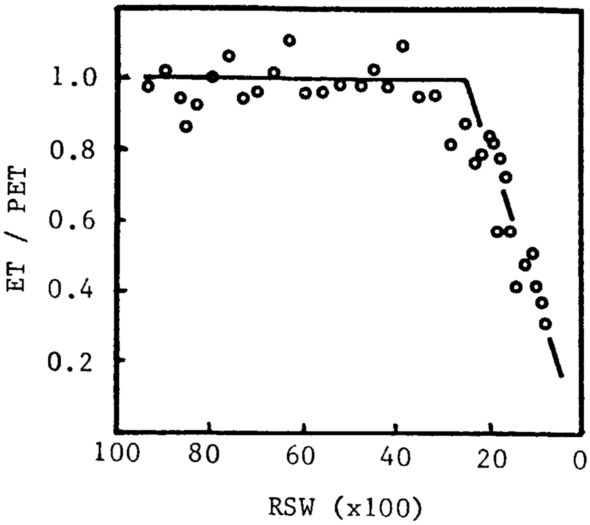

Figure 5. Ratio of evapotranspiration to potential evapotranspiration as influenced by relative soil water (Meyer and Green, 1980).

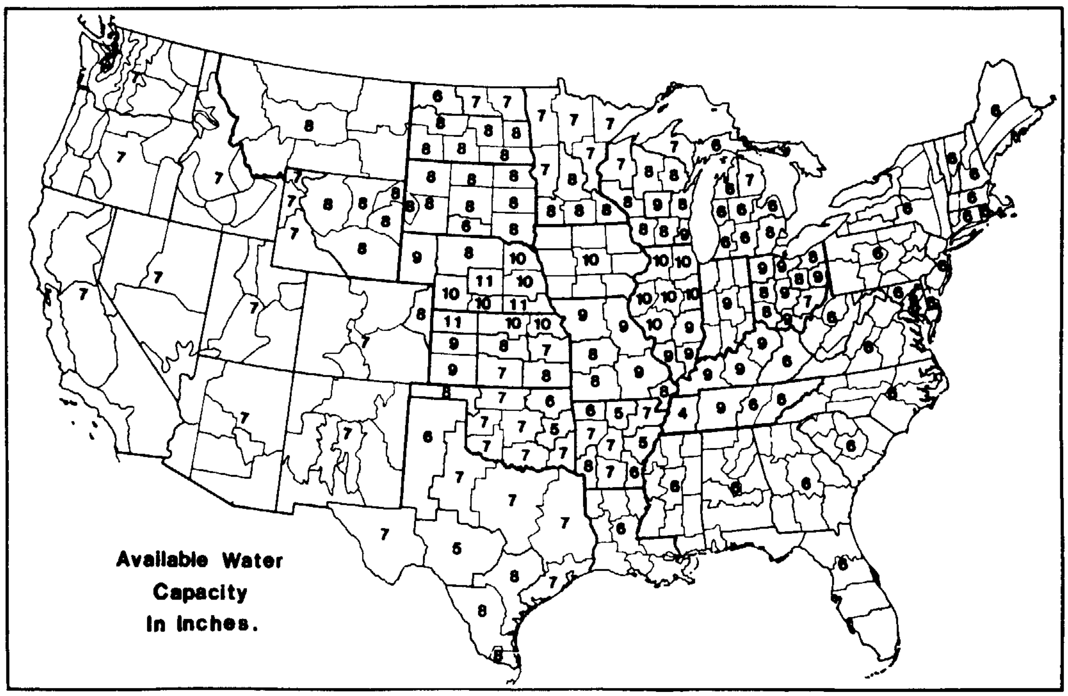

Figure 6. Estimates of field capacity in the upper 5 feet of soil for Crop Reporting Districts in the contiguous United States (Main, 1979).

Monthly values were then interpolated from their respective observing station locations to points centrally located in each of the 104 counties in Kansas. A distance-weighted interpolation technique described by Barnes (1964) was used in this computation. These interpolated values were assumed to be representative of weather conditions over the counties for which they were computed.

Modeling the monthly soil moisture balance for each county in Kansas commenced with weather data for January of 1971. At this time, a value of 0.7 was assumed for RSW in all counties based on general moisture conditions reported for the state. This allowed initial values of SW' to be determined to start the moisture balance computations. It was later found that assuming initial RSW values of 1.0 or 0.5 (full or half-full profiles) still allowed monthly soil moisture estimates to converge in less than 2 years to those values produced by assuming a value of 0.7 for RSW. Monthly soil moisture estimates were computed through July of 1980. The accuracy of these estimates could be qualitatively compared with the weekly Crop Mositure Index analyses published during the growing season by the NOAA/USDA Joint Agricultural Weather Facility.

A plot of the modeled monthly relative soil water (RSW) averaged over those counties having objective yield samples is presented in Figure 7 for the years 1974 through 1980. Also shown in the figure are corresponding monthly maximum temperature values used in the soil moisture balance computations. The RSW values exhibit a great deal of variability both within and between years. Three consecutive years (1976, 1977 and 1978) are revealed to be relatively dry, while the remaining years show substantial storage of moisture during the spring. The beginning of the severe drought of 1980 is intimated by the rapidly decreasing RSW values and soaring temperatures as summer approached.

As indicated earlier, spring soil moisture may be useful as an index to stratify historic objective yield data for building forecast models. April RSW values from Figure 7 are shown in Figure 8, along with averages for the years 1974-1980 of wheat head weights and the ratios of stalk to head numbers. Comparison of these averages with the April RSW values indicates to what degree the objective yield data reflects soil moisture conditions. The curve in Figure 8 for the average ratio of stalks to heads resembles the RSW curve, with 1977 marking the minima in each of the curves. There is a negative correlation between the average head weight curve and the RSW and stalk/head ratio curves. This type of response has often been observed in the field and results from the compensatory nature of the wheat plant--the crop makes up for fewer heads being produced by forming more and larger grains per head.

The curves in Figure 8 indicate that early-season soil moisture may be useful in stratifying historic objective yield data. The results of applying April RSW alone and with other indices to the yield forecasting technique is presented later in this report.

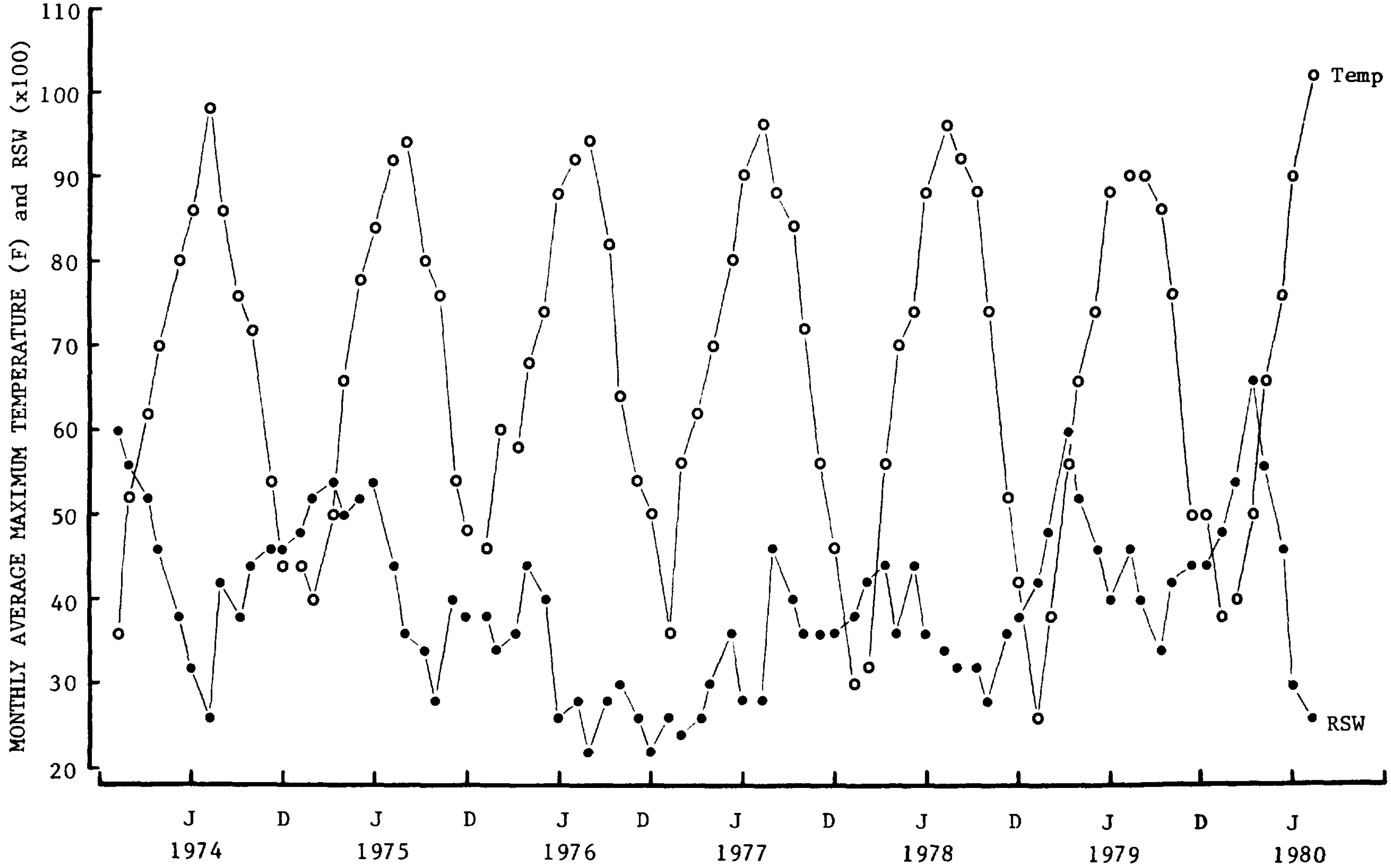

Figure 7. Monthly relative soil water (RSW) and average maximum temperature (Temp) for 1974 through 1980 in Kansas (J = June, D = December).

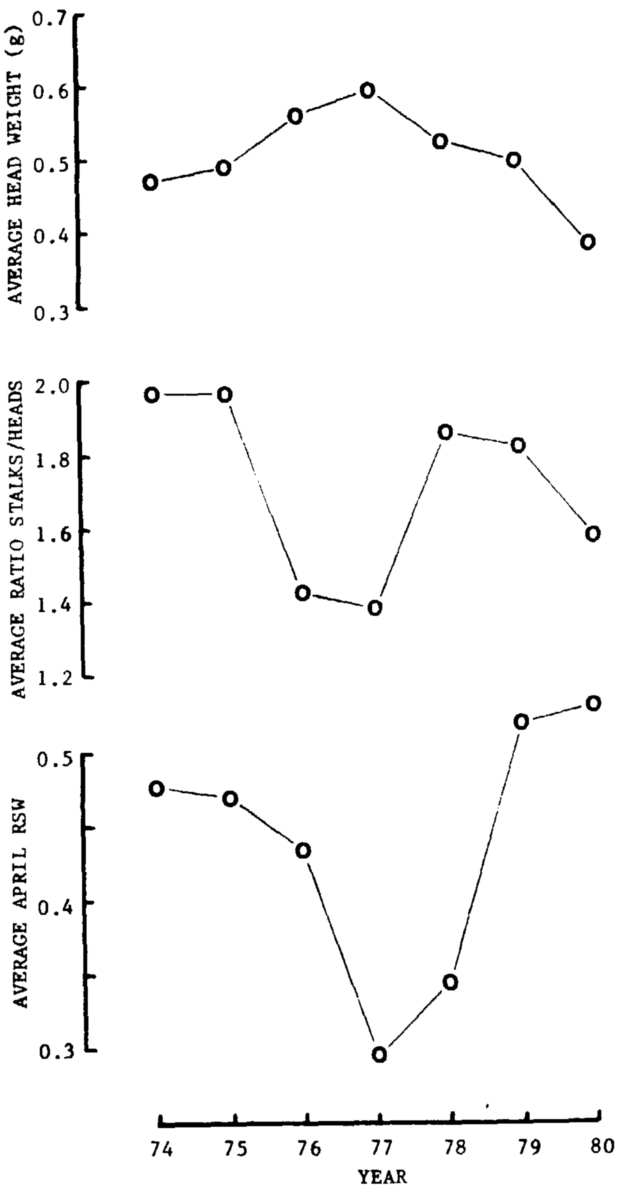

Figure 8. Average April RSW, wheat head weights and ratios of stalk to head numbers for the years 1974-1980 in Kansas.

Temperature

Temperature affects crops by regulating both the growth and development rates of plants. Unlike soil moisture, early-season temperatures are not 'stored' and thus do not have as much of a predictive nature for crop conditions later in the growing season. However, early-season temperatures may have an important effect on determining head numbers. One part of this effect is the formation of a greater number of stalks under warm conditions. Warm temperatures may have an additional effect, namely, the plants that are observed on a particular sample date may be more mature and have a higher percentage of fertile stalks (stalks in which a head has been initiated). Thus, warm temperatures may result in a value closer to 1.0 for the ratio of stalks to heads.

To quantify the effects of temperature, the average maximum temperatures shown in Figure 7 were summed for the months of January through April for each year of that period (1974-1980). These values are presented in Figure 9 along with the average head weight and stalk/head ratio values illustrated earlier. There is a general negative correlation between the temperature summations and the ratios of stalks to heads, possibly indicating more mature stalks being observed following the warmer 4-month periods in 1976 and 1977. However, warm conditions in 1974 result in a relatively high value of the stalk/head ratio. One should be aware that the individual effects of moisture and temperature may be combined or negated in producing the observed character of the objective yield data. The results of using these temperature summations alone and with other indices in making yield forecasts are presented later in this report.

Stalk Number

During the course of this study, it was noticed that the average observed stalk numbers per sample were correlated with the average ratios of stalks to heads. Thus, the independent variable in the yield component model given by Equation 1 might also serve as an index for stratifying the historic data used to evaluate the regression parameters. The average stalk number values for the years 1974-1980 are presented in Figure 10 in the same manner as the previous indices.

There is general correspondence between the three sets of values in Figure 10. A number of possible explanations for this phenomenon can be offered. Adverse weather conditions (such as low soil moisture) can cause reduced stalk numbers--this lower level of competition can allow more stalks to bear heads, thus lowering the stalk/head ratio. Also, observed stalk number takes into account plant population, which also affects the level of competition in the field. Thus, number of stalks in a plot effectively integrates environmental factors affecting growth up to the time of sampling. The compensation that occurs following the establishment of heads could result in a negative correlation between stalk number and head weight. With these considerations in mind, the use of stalk number as an index was investigated. The results of this application are presented in the following section.

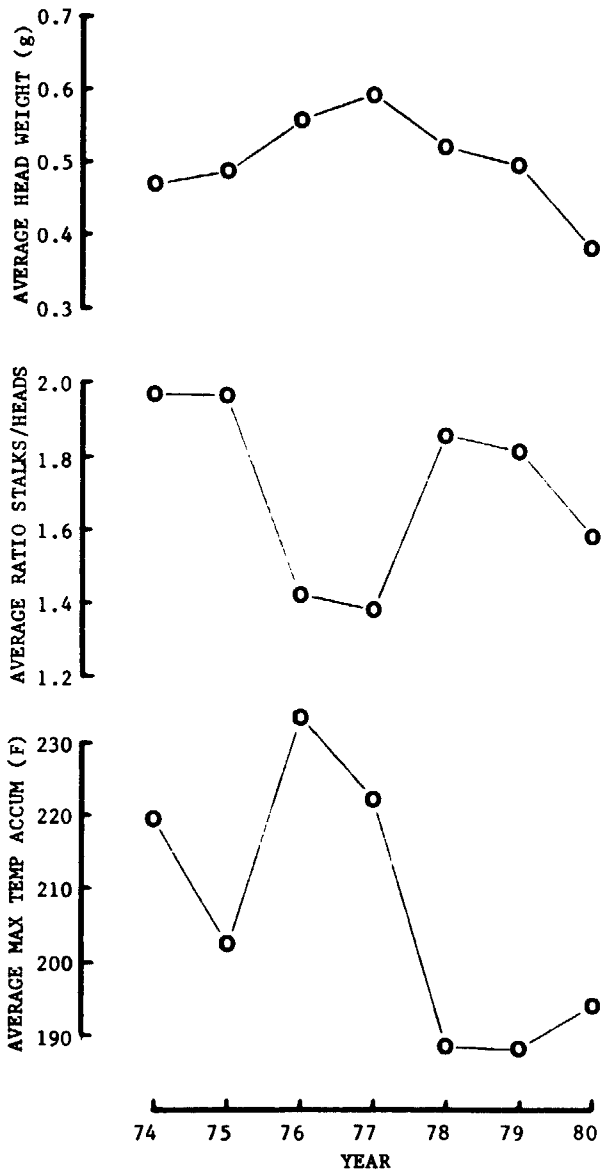

Figure 9. Average maximum temperature accumulations (for the months January through April), wheat head weights and ratios of stalk to head numbers for the years 1974-1980 in Kansas.

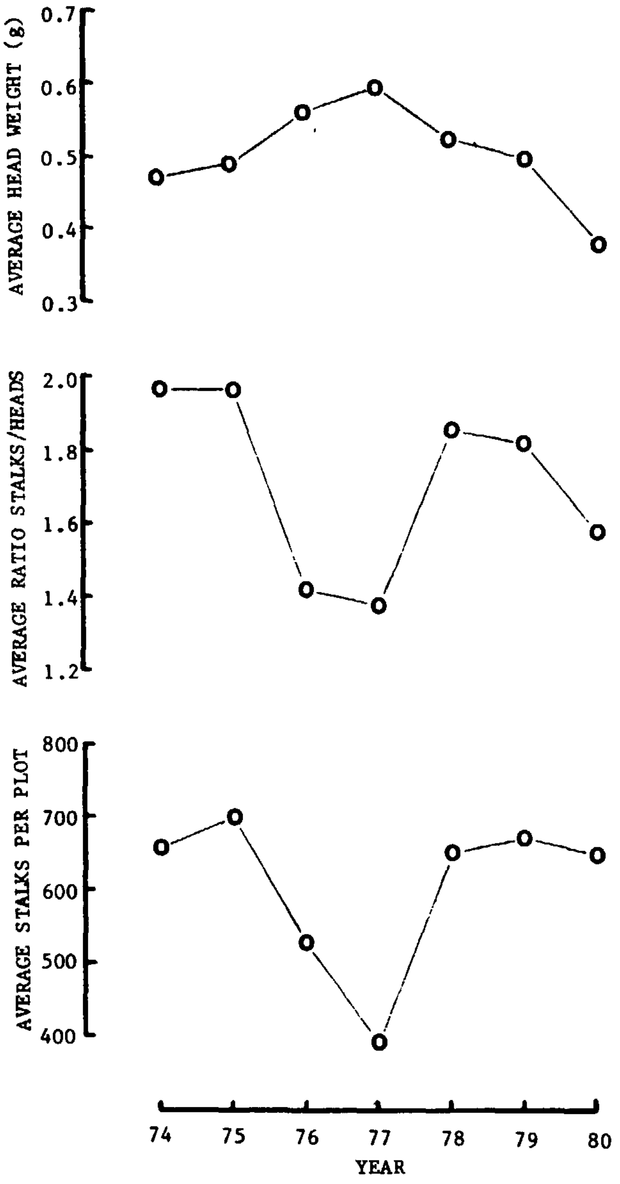

Figure 10. Average wheat stalk numbers, head weights and ratios of stalk to head numbers for the years 1974-1980 in Kansas.

YIELD FORECASTING RESULTS

The ability of the proposed technique to improve yield component forecasts can be examined by comparing the magnitudes of errors associated with the current and proposed techniques. The statistic selected for this comparison was the average absolute error (AAE),

$$AAE = \frac{1}{n} \sum_{i=1}^{n} | C'_i - C_i | \tag{7}$$

where n is the number of samples and C'_i and C_i are, respectively, the observed and forecast values of the yield component. For this study, head number (Y) and head weight (W) from Equations 1 and 2, respectively, would represent the yield components in Equation 7. This statistic was chosen because of its simplicity of computation and its ability to reflect the overall degree of accuracy of the forecast. It is superior to the average error in that it does not allow cancellation of error due to over- and under-prediction among a group of samples. The results of evaluating the between-year and within-year stratification techniques using this statistic are presented in the following paragraphs.

Between-Year Stratification

Three simple and three combined indices were selected for use in the proposed BYS technique, namely,

April relative soil water (RSWA) from Figure 8
Cumulative maximum temperature (TEMP) from Figure 9
Average stalk number per sample (STKN) from Figure 10
Combined effects of RSWA and TEMP (RSWA+TEMP)
Combined effects of TEMP and STKN (TEMP+STKN)
Combined effects of RSWA and STKN (RSWA+STKN).

For each of the years from 1974 through 1980, these indices were used to identify three similar years among those remaining in the sequence. For the simple indices, these 3 years were determined by the following procedure. First, differences were calculated between the index values for the year under consideration (the "forecast year") and each of the remaining years. Then, these differences were compared and the years associated with the three smallest difference values were considered to be similar to the forecast year. A somewhat different procedure was used for the combined indices. In these cases, values of the two simple indices that comprise the combined index were plotted against each other for each year, as demonstrated in Figure 11. Then, distances on the plot were calculated between

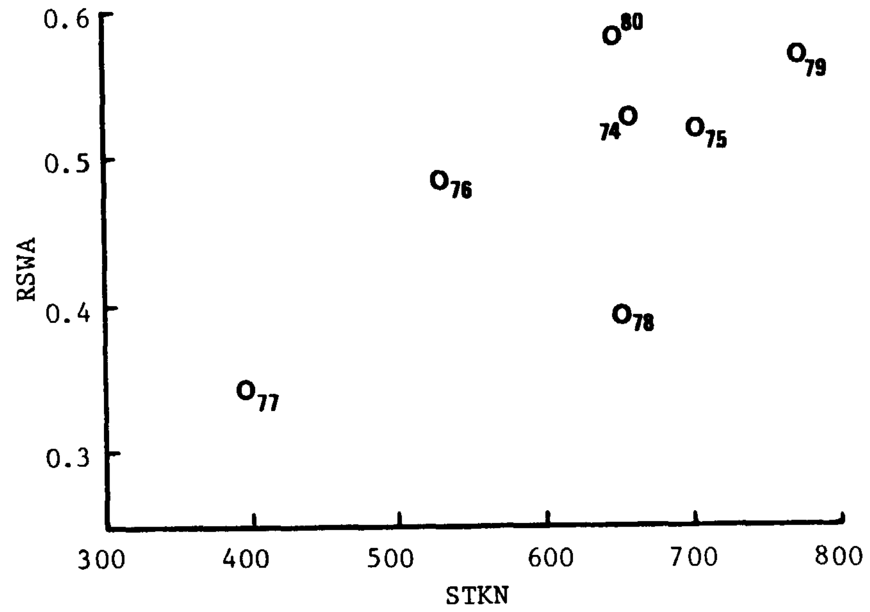

Figure 11. Plot of the simple index RSWA versus STKN for the years 1974-1980 (numbers associated with each point are the last two digits of the year).

the point associated with the forecast year and those associated with the remaining years. Those years associated with the three smallest distances were considered to be similar to the forecast year. The years selected by these means for each of the years 1974-1980 are presented in Table 1.

The objective yield data associated with the 3-year groupings shown in Table 1 were used to compute the values of the a, b and c parameters in Equations 1 and 2. Based on plant characteristics observed during the forecast years, head number and head weight forecasts were made for each objective yield sample. Corresponding forecast and observed yield component values were used in Equation 7 to compute AAE values for each forecast year. The results of these calculations are presented in Table 2.

The actual sequence of years 1974-1980 allows only four consecutive 3-year groupings (1974-1976, 1975-1977, 1976-1978 and 1977-1979) of objective yield data for forecasting head number using the current technique. However, by treating each of the seven years as independent of the others, 35 unique 3-year combinations can be formed. Using these hypothetical sequences, regression parameters can be computed to make 20 separate forecasts of head number for

Table 1--Years selected by indices for between-year stratification

Index*	Forecast Year						
	1974	1975	1976	1977	1978	1979	1980
RSWA	1975	1974	1974	1975	1975	1974	1974
	1976	1976	1975	1976	1976	1975	1975
	1979	1979	1979	1978	1977	1980	1979
STKN	1978	1974	1974	1976	1974	1974	1974
	1979	1978	1978	1978	1979	1978	1978
	1980	1979	1980	1980	1980	1980	1979
TEMP	1975	1974	1974	1974	1975	1975	1975
	1976	1978	1975	1975	1979	1978	1978
	1977	1979	1977	1976	1980	1980	1979
RSWA+TEMP	1975	1974	1974	1974	1975	1974	1974
	1976	1979	1975	1976	1977	1975	1975
	1980	1980	1977	1978	1979	1980	1979
TEMP+STKN	1975	1974	1974	1974	1975	1975	1974
	1976	1978	1977	1976	1979	1978	1975
	1980	1980	1980	1980	1980	1980	1978
RSWA+STKN	1975	1974	1974	1974	1974	1974	1974
	1979	1979	1978	1976	1975	1975	1975
	1980	1980	1980	1978	1976	1980	1979

*See text for identification of indices.

each forecast year. Likewise, the actual sequence of years allows only two consecutive 5-year groupings (1974-1978 and 1975-1979) of objective yield data for forecasting head weight using the current technique. However, 21 unique hypothetical 5-year combinations can be formed, allowing average head weights to be computed to make six separate forecasts of head weight for each forecast year. The AAE values associated with head number and head weight forecasts obtained from these hypothetical sequences are presented in Tables 3 and 4, ranked for each forecast year from smallest to largest.

Table 2--Average absolute errors associated with forecasts made using between-year stratification

Index*	Yield Component	Forecast Year 1974	1975	1976	1977	1978	1979	1980
RSWA	Head Number	70.8	63.6	70.8	61.7	78.6	66.6	97.9
	Head Weight(g)	0.1018	0.1181	0.1437	0.1408	0.1068	0.1051	0.1304
STKN	Head Number	74.5	61.0	63.0	60.9	79.7	70.1	94.8
	Head Weight(g)	0.1037	0.1169	0.1677	0.1669	0.1280	0.1016	0.1370
TEMP	Head Number	69.2	61.0	68.6	63.1	79.6	60.3	95.7
	Head Weight(g)	0.1135	0.1169	0.1326	0.1458	0.1252	0.0994	0.1390
RSWA+TEMP	Head Number	73.4	67.2	68.6	61.3	77.6	66.6	97.9
	Head Weight(g)	0.1037	0.1212	0.1326	0.1413	0.1068	0.1051	0.1304
TEMP+STKN	Head Number	73.4	69.7	66.8	59.6	79.6	60.3	97.5
	Head Weight(g)	0.1037	0.1218	0.1566	0.1769	0.1083	0.0994	0.1352
RSWA+STKN	Head Number	71.7	67.2	63.0	61.3	76.6	66.6	97.9
	Head Weight(g)	0.1050	0.1212	0.1677	0.1413	0.1083	0.1051	0.1304

*See text for identification of indices.

If the weather-related condition embodied by an index did not have an observable effect in the objective yield data, then one might expect the AAE values associated with yearly stratified forecasts to be more-or-less randomly ranked among the AAE values obtained using the current technique. If a large number of forecast years were used, then about as many index-related AAE values would be found in the lower half of the ranking as in the upper half. For an index to be potentially useful, it should consistently produce forecasts with AAE values in the lower half of the ranking (rank 1-10 for head number and 1-3 for head weight).

It should not be expected that the technique will always produce forecasts with the smallest AAE values--the indices chosen for this study can account for only the grossest impact of weather on the crop. Many other factors that can have significant effects on yield, such as diseases, insects and fertilizers, are ignored in this analysis. In fact, it is possible that the use of weather-related indices might

Table 3--Average absolute errors associated with head number forecasts made using the current technique

Rank	Forecast Year 1974	1975	1976	1977	1978	1979	1980
1	66.2	61.0	60.7	59.6	76.4	60.3	86.8
2	67.4	62.0	61.5	60.2	76.4	62.8	87.8
3	68.1	63.6	61.9	60.9	76.6	63.0	90.4
4	68.1	64.2	62.1	60.9	77.6	63.0	90.6
5	69.2	65.0	62.8	61.3	77.6	64.5	90.7
6	69.2	65.7	63.0	61.3	77.7	66.1	91.1
7	70.8	66.5	63.2	61.4	78.5	66.2	91.2
8	71.1	67.2	64.3	61.7	78.5	66.6	92.1
9	71.2	67.5	64.7	62.1	78.6*	67.1	92.4*
10	71.7	69.6	65.4	62.2	79.6	68.8	92.5
11	72.6	69.7	66.1	62.5	79.7	68.9	93.2
12	73.4	70.7	66.4	62.6	79.8	69.1*	93.7
13	73.6	71.7	66.8	62.7	80.1	69.9	94.1
14	74.2	73.0	67.7	62.7	80.6	70.1	94.4
15	74.5	74.1	68.1	62.9	81.1	71.3	94.8
16	74.8	74.9	68.3	63.1*	81.6	71.6	95.3
17	75.8	76.0	68.3	63.7	82.1	71.7	95.4
18	77.4	76.9	68.6	64.1	82.5	77.5	95.7
19	79.1	78.4	70.8	64.2	83.9	84.1	97.5
20	82.5	86.8	71.7	66.4	89.1	93.2	97.9

*Result using actual previous 3-year sequence.

Table 4--Average absolute errors associated with head weight forecasts made using the current technique

	Forecast Year						
Rank	1974	1975	1976	1977	1978	1979	1980
1	0.1005	0.1161	0.1345	0.1462	0.1068	0.0959	0.1433*
2	0.1005	0.1162	0.1461	0.1602	0.1126	0.0960	0.1485
3	0.1005	0.1162	0.1472	0.1617	0.1132	0.0961	0.1535
4	0.1008	0.1165	0.1487	0.1641	0.1139	0.0963	0.1584
5	0.1012	0.1165	0.1497	0.1649	0.1163	0.0964	0.1587
6	0.1060	0.1217	0.1554	0.1682	0.1182	0.1041*	0.1603

*Result using actual previous 5-year sequence.

result in a forecast that is less accurate than the forecast obtained using the current technique. This situation could occur when early-season weather conditions are very different from late-season conditions. Since, at the time of the forecast, there is no indication of an extreme change in the coming weather, an inappropriate set of years may be selected from the historic data record to build the yield component models. In these cases, the use of early-season indices will probably fail to improve forecasts. However, the occasional failure of the proposed technique due to anomalous years can be compensated by consistently improved forecasts for the balance of years. This is particularly true in light of the fact that the index-associated forecasts for an anomalous year should improve later in the season as the changeable nature of the weather becomes observable.

In comparing the results presented in Tables 2 and 3, one observes that, for 1980, none of the indices produced head number forecasts with AAE values ranked lower than 15th among the values obtained using the current technique. This exemplifies the point that the technique might not be effective in improving early-season forecasts for anomalous years. Referring to Figure 7, 1980 could be considered anomalous due to the change from cool, wet conditions early in the year to hot, dry conditions later in the growing season. In fact, the three years that are associated with the best head number forecast for 1980 are the dry years 1976, 1977 and 1978. For the remaining years (1974-1979), AAE values associated with the use of the simple index RSWA were ranked within the ten smallest for five out of six years, while the combined index RSWA+STKN produced error values ranked within the lower half for all six years. The use of TEMP alone or in combination with another index did not appear to consistently improve head number forecasts.

The failure of STKN alone to improve head number forecasts can be explained by referring to Table 1, which shows that data from the anomalous year 1980 were selected to build the forecast models for all years in the sequence 1974-1979 except 1975. The effects of the anomalous year may be removed from the yield component forecasts by excluding 1980 from selection by the indices. The years selected by the various indices excluding 1980 are shown in Table 5, while Table 6 presents the error values associated with head number and head weight forecasts based on the data sets selected in Table 5. Comparing the results presented in Tables 3 and 6, one now observes that the error values associated with the use of STKN are ranked within the ten smallest for all years in the sequence 1974-1979. The head number forecasts using the combined index RSWA+STKN also have AAE values ranked in the ten smallest for this period. The forecast capability associated with other indices is not markedly improved by the exclusion of data from 1980. Thus, data from the anomalous year had a deleterious effect on the ability of the index STKN to improve head number forecasts. In general, it would be advantageous to exclude known anomalous years from the set available for building the yield component models, unless it is anticipated that the forecast year will also be anomalous. Unfortunately, excluding years from a data set of limited size increases the probability that a year will be selected that is not as representative of the forecast year as one might desire. The best solution to this problem is to have a large number of years of objective yield data to select from.

As indicated by the results presented in Tables 2, 3 and 6, head number forecasts based on the indices RSWA+STKN and STKN appear considerably more accurate than those obtained using the actual preceding 3-year sequences for 1977, 1978 and 1979. Head number forecasts based on indices resulted in larger errors for the anomalous year 1980 than that produced by using the actual preceding 3-year sequence (1977-1979) for that year.

It is more difficult to assess improvements to head weight forecasts due to application of the BYS technique. All indices appear to have improved the accuracy of head weight forecasts for the anomalous year 1980. This occurred because none of the indices allowed data from 1976 or 1977 to be used to calculate the average head weight parameter in the yield component model (Equation 2). As shown in Figure 8, average head weights for these two years are much larger than for 1980. Thus, exclusion of the 1976 and/or 1977 data from building the model considerably improves its ability to forecast for 1980. The use of five years of data in the current technique allows at least one of these years (1976 or 1977) to influence the 1980 head weight forecast.

Table 5--Years selected by indices for between-year stratification excluding 1980 data

Index*	Forecast Year						
	1974	1975	1976	1977	1978	1979	1980
RSWA	1975	1974	1974	1975	1975	1974	1974
	1976	1976	1975	1976	1976	1975	1975
	1979	1979	1979	1978	1977	1976	1979
STKN	1975	1974	1974	1974	1974	1974	1974
	1978	1978	1977	1976	1975	1975	1978
	1979	1979	1978	1978	1979	1978	1979
TEMP	1975	1974	1974	1974	1974	1974	1975
	1976	1978	1975	1975	1975	1975	1978
	1977	1979	1977	1976	1979	1978	1979
RSWA+TEMP	1975	1974	1974	1974	1975	1974	1974
	1976	1978	1975	1976	1977	1975	1975
	1979	1979	1977	1978	1979	1978	1979
TEMP+STKN	1975	1974	1974	1974	1974	1974	1974
	1976	1978	1975	1976	1975	1975	1975
	1978	1979	1977	1978	1979	1978	1978
RSWA+STKN	1975	1974	1974	1974	1974	1974	1974
	1978	1978	1975	1976	1975	1975	1975
	1979	1979	1978	1978	1976	1978	1979

*See text for identification of indices.

The use of five years of data in the current technique assures a conservative head weight forecast which exhibits considerably less variation from year to year than the average head weights shown in Figure 8. For some of the forecast years (such as 1974, 1975, and 1979), many of the AAE values within a ranking are essentially the same. As shown in Tables 4 and 6, forecasts associated with the indices STKN and RSWA+STKN are generally as good or better than those produced using the previous five years, assuming that data from the anomalous year 1980 are excluded from building the model. The simple index RSWA also performed relatively well with or without the exclusion of 1980 data.

Table 6--Average absolute errors associated with forecasts made using between-year stratification excluding 1980 data

		Forecast Year						
Index*	Yield Component	1974	1975	1976	1977	1978	1979	1980
RSWA	Head Number	70.8	63.6	70.8	61.7	78.6	93.2	97.9
	Head Weight(g)	0.1018	0.1181	0.1437	0.1408	0.1068	0.1035	0.1304
STKN	Head Number	68.1	61.0	65.4	61.3	76.4	64.5	94.8
	Head Weight(g)	0.1010	0.1169	0.1230	0.1413	0.1176	0.1020	0.1370
TEMP	Head Number	69.2	61.0	68.6	63.1	76.4	64.5	95.7
	Head Weight(g)	0.1135	0.1169	0.1326	0.1458	0.1176	0.1020	0.1390
RSWA+TEMP	Head Number	70.8	61.0	68.6	61.3	77.6	64.5	97.9
	Head Weight(g)	0.1018	0.1169	0.1326	0.1413	0.1068	0.1020	0.1304
TEMP+STKN	Head Number	67.4	61.0	68.6	61.3	76.4	64.5	97.5
	Head Weight(g)	0.1046	0.1169	0.1326	0.1413	0.1176	0.1020	0.1352
FSWA+STKN	Head Number	68.1	61.0	65.4	61.3	76.6	64.5	97.9
	Head Weight(g)	0.1010	0.1169	0.1230	0.1413	0.1083	0.1020	0.1304

*See text for identification of indices.

Assuming that data from anomalous years can be excluded from building the yield component models, the simple index STKN appears to consistently produce head number and head weight forecasts as good or better than those produced using the previous five years. The combination of stalk number with RSWA, or RSWA alone, also appears to consistently improve forecasts. The use of the index TEMP alone or combined with other simple indices did not appear to be effective in consistently improving head number or weight forecasts.

Within-Year Stratification

The WYS technique ignores grouping the historic objective yield data by years but rather attempts to stratify the data into groups experiencing similar weather conditions over a number of years. Since indices for each individual sample are examined, the WYS technique could potentially be more accurate than the BYS technique because it could separate data from regions within a state experiencing different weather conditions during a year. In making forecasts for a particular year, all the historic data are initially pooled. Then this multi-year data set is stratified into subgroups based on previously-determined critical values of an index. These subgroups of data are used to determine the regression parameters in the yield component models. Finally, in making the forecast, the samples observed for the forecast year are similarly stratified and the appropriate yield models are applied to each sample based on its respective grouping.

To examine the applicability of the WYS technique in this study, the 7-year data set (1974-1980) was separated into the forecast year and the remaining years. Thus, seven forecasts could be made using the technique. The results of using the proposed stratification technique for a given forecast year were compared to the results obtained when the data from the remaining years were pooled but not stratified.

Based on its apparent success as an index in the earlier analysis, relative soil water at the end of April was initially selected as an index for the WYS technique. In this case, however, values of the index associated with individual samples (RSWA) were used in the stratification. In this context, RSWA used in the BYS technique represents the average of all the RSWA values for a given year.

Samples from the historic data are stratified by comparison of their respective RSWA values to critical values of the index. These critical values are determined so that they separate the objective yield data into groupings with similar yield-producing characteristics. To assist in determining these critical values of RSWA, statistical analysis was performed on the pooled objective yield data from the years 1974-1979. This analysis initially involved separating the data into five subsets containing data with RSWA values within the ranges 0-0.2, 0.2-0.4, 0.4-0.6, 0.6-0.8 and 0.8-1. Then, regressions based on these subsets and combinations of these subsets were tested using a technique described by Neter and Wasserman (1974, pp.160-165) to determine whether the regressions were similar or different. Based on the results of this analysis, the entire range of the index RSWA was split into three segments, namely, 0-0.2, 0.2-0.6 and 0.6-1. This subsetting is in agreement with the response of wheat ET to soil moisture observed by Meyer and Green (1980) and presented in Figure 5. For values of RSW above 0.6, ET approaches PET and growth of the crop is not hindered by soil moisture conditions. For values of RSW below 0.2, ET is severely reduced below PET and growth becomes limited by the lack of moisture. The gap between these two critical RSW values represents a transition zone in which crop growth becomes affected by limited soil moisture.

The results of forecasting head number and head weight using this technique are presented in Table 7. Stratifying by RSWA improved head number forecasts for 5 out of 7 years, while head weight forecasts were improved 4 out of 7 years. Neither head number nor head weight forecasts were improved for the anomalous year 1980.

Table 7--Average absolute errors associated with forecasts made using within-year stratification

Technique	Yield Component	Forecast Year 1974	1975	1976	1977	1978	1979	1980
Stratified	Head Number	71.5	66.0	68.0	57.7	78.8	60.7	94.2
	Head Weight(g)	0.0951	0.1132	0.1499	0.1509	0.1191	0.1008	0.1472
Non-Stratified	Head Number	71.8	68.6	64.3	61.1	79.1	66.7	92.9
	Head Weight(g)	0.1011	0.1186	0.1446	0.1626	0.1150	0.1027	0.1457

Although more sophisticated in its forecasting approach, the WYS technique performed no better than the BYS technique in this example. The main difficiency in this case probably arises from the lack of precipitation data and soil characteristics at the objective yield sample locations. The inability to locate samples within a county, along with the need to interpolate precipitation data from surrounding observing stations, results in the stratification of yield data by an index representative of conditions generalized to a county-size scale. Since the scales of soil characteristics and precipitation occurrences may be considerably smaller than county-size, a considerable amount of scatter may be introduced into the data used to build the yield component models. This scatter is generated by the method of analysis and is in addition to that already in the data due to factors ignored in the stratification analysis (fertilizers, diseases, insects, etc.).

Exclusion of data for the anomalous year 1980 from the pooled data used to build the yield models did not improve the overall accuracy of the technique. The WYS technique may be less sensitive to the effects of anomalous years than the BYS technique, although the nature of the data used in this study makes this difficult to determine.

REFERENCES

Baier, W. and G. W. Robertson (1967) Estimating Yield Components of Wheat from Calculated Soil Moisture. Can. J. Plant Sci., 47(6): 617-630.

______________(1968) The Performance of Soil Moisture Estimates as Compared with the Direct Use of Climatological Data for Estimating Crop Yields. Agr. Meteorol., 5(1): 17-31.

Barnes, S. L. (1964) A Technique for Maximizing Details in Numerical Weather Map Analysis. J. Appl. Meteorol., 3: 396-409.

Bloodgood, D. W., R. E. Patterson and E. L. Smith, Jr. (1964) Water Evaporation Studies in Texas. Texas Agricultural Experiment Station, November 1964.

Bole, J. B. and U. J. Pittman (1980) Spring Soil Water, Precipitation and Nitrogen Fertilizer: Effect on Barley Yield. Can. J. Soil Sci., 60: 461-469.

Brown, M. J. (1978) Sources of Weather and Climatic Information for Kansas. Bulletin 614, Kansas Agricultural Experiment Station, April 1978.

Bureau of Reclamation (1977) Final Environmental Statement for Project Skywater, Vol. 1. U.S. Department of the Interior, Bureau of Reclamation, Engineering and Research Center.

Kmoch, H. G., R. E. Ramig, R. L. Fox and F. E. Koehler (1957) Root Development of Winter Wheat as Influenced by Soil Moisture and Nitrogen Fertilization. Agron. J., 49: 20-25.

Main, W. A. (1979) Palmer Index Computations for Small Land Areas. Extended Abstracts, 14th Conference on Agricultue and Forest Meteorology, 2-6 April 1979, Minneapolis, Minnesota: 150-153.

Meyer, W. S. and G. C. Green (1980) Water Use by Wheat and Plant Indicators of Available Soil Water. Agron. J., 72: 253-257.

Moe, R. D. and J. F. Griffiths (1965) A Simple Evaporation Formula for Texas. MP-795, Texas Agricultural Experiment Station, December 1965.

Neter, J. and W. Wasserman (1974) Applied Linear Statistical Models. Richard D. Irwin, Inc., Homewood, Illinois.

Strand, B. W. (1981) Spatial Scale of Crop-Yield Models. Staff Report AGESS810320, U.S. Dept. of Agriculture, Statistical Reporting Service, Statistical Research Division, March 1981.

APPENDIX

The following seven figures show the number of objective yield samples in each of the 104 counties in Kansas that provided data for analysis in this study.

U.S. GOVERNMENT PRINTING OFFICE: 1982-360-929:SRS-626