Kelly
Kirkbride

FORECASTING CROP YIELDS

Bruce W. Kelly 1/ and John W. Kirkbride 2/
Statistical Reporting Service, USDA

There are those who hold that historians tell us about the past, economists tell us about the future, and so it is only the present that is confusing. The discussions of the seminar may substantiate this belief.

We have heard about the past -- weatherwise, crop production-wise, and research-wise. We can look forward to hearing about the future and what it holds fo: crop yields. That leaves us only to worry about what is happening at the present time. That task is generally left to the Statistical Reporting Service along with a handful of professional estimators and thousands of self-appointed prognosticators.

We can all agree that crop yields are the culmination of a wide variety of variables, most of which show varying degrees of relationship to one another--some positive and some negative in terms of crop output. One of the most controversial variables is weather but even here we can agree that crop yields are dependent upo: the weather--assuming weather in its broadest sense. Other variables that exert influence on yields are soil type, soil fertility, plant population, variety, insects, disease, and cultural practices.

What is the interaction of these items with weather--some of which have occurred during the growing season to date, some of which must still occur during the growing season? These offer interesting thoughts for speculation. Researcher can and do isolate one or more of these items and present evidence of their impact on yield. One of the problems to date has been the rather wide variation in evidence. You are aware of the various opinions relative to the effect of weather on the recent sharp uptrend in yields for certain crops--ranging from only minor effect to accounting for more than 80 percent of the increase. Similar difference are voiced relative to plant population, application of fertilizer, new varieties-all interesting items for speculation and helpful in the evaluation of a given set of conditions in relation to yield, but how well do such opinions or results measure the combined effects of the many factors that result in the amount of product removed from a given acre. These opinions and research results do illustrate the luxury enjoyed by some in speculating about the cause and effect of yields. We in the Statistical Reporting Service seldom enjoy such luxury--ours is the role of being expected to know what is happening to yield month by month.

The Statistical Reporting Service has the responsibility for making (1) forecasts of crop production from current crop conditions during the growing season and (2) annual estimates of crop production. These are two separate and distinct functions. We use "estimate" to indicate a measure of accomplished fact, such as at harvest time or later; the term "forecast" is used to refer to expectations of what is likely to be accomplished at some time in the future.

It should be clearly understood that a forecast is a statement of the most likely magnitude of yield or production on the basis of known facts on a given date, assuming weather conditions and damage from insects or other pests and disease during the remainder of the growing season to be about the same as the

1/ Chief, Research and Development Branch
2/ Head, Grain and Hay Section

Presented by John W. Kirkbride at a Seminar on Weather and Our Food Supply, sponsored by the Center for Agricultural and Economic Development, Iowa State University, Ames, Iowa, May 4-6, 1964.

and lesser or greater numbers representing conditions poorer or better than "aver age." However, it soon became evident that farmers had difficulty in visualizing an average condition. This was demonstrated by the fact that over a period of years, the average of all reports of condition was somewhat less than 10. To get away from the use of "average", the concept of "normal" condition became the stan ard by which reporters were asked to rate condition of crops.

A normal condition is not an average condition, but a condition above averag giving promise of more than an average crop. Furthermore, a normal condition doe not indicate a perfect crop, or a crop that is or promises to be the very largest in quantity that the area reported upon may be considered capable of producing. The normal indicates something less than this and thus comes between the average and the possible maximum. The normal can be described as a condition of perfect healthfulness, unimpaired by drought, hail, insects, or other injurious agency, and with such growth and development as may be reasonably looked for under these favorable conditions.

The concept of what constitutes a "normal" condition of a crop obviously varies from one locality to another with difference in soil and climate. It also changes slowly, over time, in the same locality because of change in varieties, cultural practices and soil fertility. Shifts in the acreage distribution of a crop within a State, from acres of low yields to acres of high yields, may mean that the same reported condition will indicate a different yield than it once did while a shift in the opposite direction may have the reverse effect. The relativ constancy of the aggregate of all the individual reporters' ideas of normal condi tion has greatly enhanced its usefulness.

During these early years there was much concern about the reliability of est mates. Efforts were made to improve the data by increasing the number of correspondents. This began a period of transition in the method of making annual produ tion estimates. Up to this time, estimates were based on reports by county reporters of the total crop production for the county as a percent of the previous year. Beginning about 1888, county indications were weighted to calculate State indications. During the season there were returns, first of area, then several consecutive returns of condition, then of yield per acre, and finally of productio compared with the previous year. These furnished data for three separate tests o amount of production, which were examined at the end of the season and harmonized for the final and only estimate. This was the beginning of the evolution that le to the current procedure of calculating crop production as a product of the two separate estimates of acreage and yield. During the late 1800's an increasing nu ber of reports were received from handlers and processors of agricultural product Their reports, which were used as supplementary indications became increasingly important, particularly as post-harvest check data on the amount of the crops.

As early as the 1880's some dealers began to interpret the reported conditio of each major crop in terms of actual bushels, tons or pounds of probable yield. The desirability of having such interpretations made by the Government and, therefore, available to all was recognized and in 1912 the Crop Reporting Board began to publish forecasts of yields.

The method used originally was the so-called "par method" which assumes a proportional relationship between reported condition and final yield over the entire range of reported condition values. The inflexibility of the "par method" necessitated subjective modification of the condition index or of the pars to eliminate the disturbing effect of highly atypical years and of trends in the data combination of variables on yield. Regressions of final yield using various combinations of rainfall, temperature, humidity, and other indices of weather were developed for most major crops by States.

During the late 1930's detailed special crop-weather projects were carried out for cotton, corn, and wheat. The projects involved special crop-weather plots at a number of Experiment Stations recording detailed plant and weather observations. Some exploratory work was also done at that time using complex equations. All of these studies added materially to the statistician's knowledge of crop yields in relation to weather. They showed the relative importance of weather by months, the effect of accumulated rainfall prior to the growing season and the general importance of factors other than weather.

While the correlations were significant and fairly high for some crops in certain States, the relationship when used in subsequent years would not be the same as for the years included in the study. For forecasting purposes, therefore, the previously observed relationships were misleading at times and generally much less reliable than estimates based on currently reported indices of yield per acre

While the so-called "direct" weather procedure in estimating crop yields per acre has not been abandoned, the emphasis has been shifted to what may be termed the "indirect" or supplemental weather approach.

In the present estimating program, considerable use is being made of multiple regressions in estimating yield with reported condition and/or yield, precipitation, or indices of weather as variables.

Multiple regression equations and charts using combinations of current prospects reported by crop correspondents and precipitation as variables are being used for winter, durum, other spring wheat, corn, and soybeans for some months and areas. In general, precipitation data contribute two factors to the equations: (1) accumulated precipitation for selected months before the forecast date, and (2) precipitation for the following month or combination of months. Precipitation after date has to be estimated from a knowledge of long-time trends, seasonal patterns in recent years and long-range weather forecasts. For most early season estimates, precipitation after date accounts for the major portion of the variance. The level of the indicated yield, therefore, is heavily influenced by the estimate of precipitation after the forecast date and the procedure becomes very subjective for current forecasting.

In appraising current prospects, crop reporters take into account seasonal progress, diseases, insects, quantity of fertilizer used, and other cultural practices. The reported condition or yield, therefore, reflects the composite effect of weather and cultural practices to date and reporters' evaluation of such factor on final outcome. When these measures of current prospects are used as variables along with actual precipitation to date and after date, the regression coefficient measure the contribution of the components used. Any persistent tendency for farmers to under or over-estimate for a given pattern of rainfall, therefore, is appropriately adjusted.

In this approach we are not necessarily limited to use of actual weather data as a variable. Other factors which are, in themselves, measures of weather or effects of weather are also used. Estimating procedures for cotton and tobacco are examples of such methods.

In the United States the Crop Reporting Service began experimenting with crop cutting just prior to 1940 with pre-harvest wheat surveys through the Plains States. These were discontinued after 1940 with no further work until about 10 years ago. At that time, an intensified program of objective counts was undertaken. One of the first steps taken was to approximately optimize plot sizes. Optimum sizes turned out to be rather small: two rows, 15 feet long for corn; two rows, 10 feet long for cotton; two rows, 3 feet long for soybeans; and a plot approximately 1/10,000th (0.00001th) of an acre for wheat. Experiments were conducted to find means of reducing the biases associated with these small plots. It was determined that bias could be controlled by making very precise measurements of the sample plots; by development of rules for handling border line plant and by careful training and supervision of the samplers.

A sample design has been worked out for field and plot selection. At presen an allocation of sample fields is made to States with consideration given to the precision of both State and regional estimates, and within States, a subsample is selected from the fields chosen in the spring general-purpose probability sample survey. The fields in the subsample are selected with probabilities proportional to acreage, and two plots per field are located by a random process. This procedure results in a self-weighting sample of plots. Incidentally, the optimum numb of plots per field appears to be something less than two, but one degree of freedom is desirable for analytical purposes, and the loss in efficiency is small.

The precision of the pre-harvest estimate of yield is of interest. A sample of 3,100 corn fields allocated to 24 North Central and Southern States gives a regional yield estimate with a standard error of about three-quarters of a bushel and a sample of 2,150 cotton fields allocated to 10 Southern States gives yield estimates for individual States and for the Region with a coefficient of variatic of about 5 percent and one and three-fourths percent, respectively. The bias in the procedure for estimating corn yield has been measured by comparing sample estimates made by harvesting plots with total production from the field, and has been found to be positive but less than 2 percent.

The timing of the objective yield surveys is geared to the forecasts and estimates published by the Statistical Reporting Service. During the growing season, forecasts of yield are made at monthly intervals beginning about 2 months before harvest. The surveys upon which the objective forecasts of yield are based are likewise made at monthly intervals. For corn, soybeans, and cotton the first survey is made about August 1; and for winter wheat, about May 1.

At the first visit to the sample fields the plots are carefully measured off and marked so that they may be found readily. At this and subsequent visits, the number of plants and the number of fruit by maturity classes are counted, and a sample of fruit sent in to a laboratory for weighing and determining moisture content. Then, at the last visit before harvest, the plots are completely harvested and their yield determined. Following harvest, gleanings are collected in simila sized plots for measuring harvesting losses.

Forecasting yield is more difficult than estimating it. Direct measurements of yield can be made only when a crop is mature. When plants are immature, yield as such does not exist and hence cannot be observed directly. But, components of yield such as plant numbers, numbers of fruit, and size or weight of fruit can be counted or measured, physiological observations of plant characteristics can be made, and the components of yield projected to harvest rather well.

growing conditions. Last season (1963) the August 1 prediction of yield was witl about 5.5 percent of that actually produced as estimated by the pre-harvest surve The corn forecasting model is also based upon simple linear relationships which were derived from experimental observations.

At the time of the August 1 survey, the corn in some of the more northerly States has not begun to form ears. When this is the case, the number of ears to be produced is predicted by a linear regression between stalk numbers and ears pr duced, derived from historical data, and a historical average ear weight is also used.

When ears are present, the problem is that of predicting ear weight. Fortunately, ears attain their maximum size by the time they reach the milk stage, anc equally fortunate there is a linear relationship between length of ear and weight of grain. By means of this relationship, the length of the cob, measured over th husk, has proven a good predictor of ear weight, provided adjustments for frost damage and early harvesting are made.

Studies have shown that dry matter is laid down in the ear until the moistur content of the grain is below 30 percent. Where early harvesting occurs, it is necessary to adjust the weight per ear for loss of dry matter as well as for moisture content.

To adjust the forecast for possible early frost, the August 1 stage of maturity is used to estimate the number of days to maturity, and by comparing this dat with a historical average of first frost dates for the locality, an adjustment fc the likelihood of frost damage is made.

Last season, for 11 North Central States, the August 1 corn yield survey predicted averages of 58.0 ears per plot weighing .413 pounds per ear. The pre-harv survey found 58.7 ears per plot and an average ear weight of .438 pounds. Consequently the August 1 forecast of corn yield turned out to be 4.5 percent below th pre-harvest indications.

Objective yield techniques have been developed for tree crops as well. Thes include oranges, lemons, peaches, pears, walnuts, filberts, and sour cherries. These techniques are based upon concepts similar to field crops. The essential differences are that the sampling unit is a tree and that the crop of fruit is se before the time of the first forecast so that it is not necessary to predict the number of fruit yet to come.

On the whole, the objective forecasting procedures in their present state of development are performing reasonably well. However, further refinements are needed in the form of more sensitive relationships that are clearly defined and which incorporate the effects of environment upon plant production.

Work still remains to be done in the area of improvement in forecasting crop yields as well as the true yield levels. There is need for more intensive studie relating crop yields to weather factors and to early season plant characteristics Detailed phenological and environmental observations are needed -- the relationsh of dry matter accumulation to weather factors over the entire growth period and the use of such relationships in predicting crop yields should be explored. Then special studies need to be separated into several areas of interest: (1) phenolo ical events such as emergence of plants, fruit emergence, and fruit counts by maturity category, and (2) the mechanism of growth and development over time as