RESEARCH IN YIELD FORECASTING

by

Harold F. Huddleston

November 1975

RESEARCH IN YIELD FORECASTING

Harold F. Huddleston
Statistical Reporting Service
U. S. Department of Agriculture

Introduction

During the past three years, the major emphasis has been given to developing yield models in which the parameters are derived from the current year for use prior to harvest. These models are referred to as "within-year models" and are considered more desirable than between-year models if each year is different than the preceding years. These models do not require a historical series of 3-5 years of similar information before yield forecasts can be made. This last consideration has been considered quite important when starting work on a new crop or developing a system for a country without a crop forecasting system. This type of model has been considered for yield forecasts based on both grower subjective yield forecasts and objective yield methods. In the future, there will be opportunities to introduce new grower yield forecast models based on probability samples as well as consideration of crops of considerable economic importance for inclusion in yield forecasting programs or estimating production at harvest.

It is helpful to start with a look at grower yield appraisals (or probable yields) which are used for most crops. The impetus for this effort came from the recognition that relatively few crops were in an objective yield program and the technical assistance work for foreign governments wishing to start current statistical programs for forecasting crop production.

Grower Subjective Yield Forecasts

The most common approach used by SRS is the charting of grower probable yields against the Board final yields. This approach is based on the relationships over years being the same for a period of 5-10 years and is normally considered usable after yields have been collected for 3-5 years. In most cases, these charts are based on voluntary reports returned by mail. Consequently, the reported probable yields may not be representative and/or the grower may not be able to forecast his crop accurately. In either case, the probable yields require adjustment or correction for various kinds of biases. Frequently, there appear to be different relationships indicated for different periods of years. The dashed lines in Chart 1 indicate approximately the nature of two different regressions and the solid line the least squares regression line over both periods. This chart illustrates a common problem associated with between-year regression lines. Neither the representativeness of the sample nor the ability of the growers to forecast their yields are measured or known. Consequently, a somewhat different approach is needed in order to overcome these shortcomings.

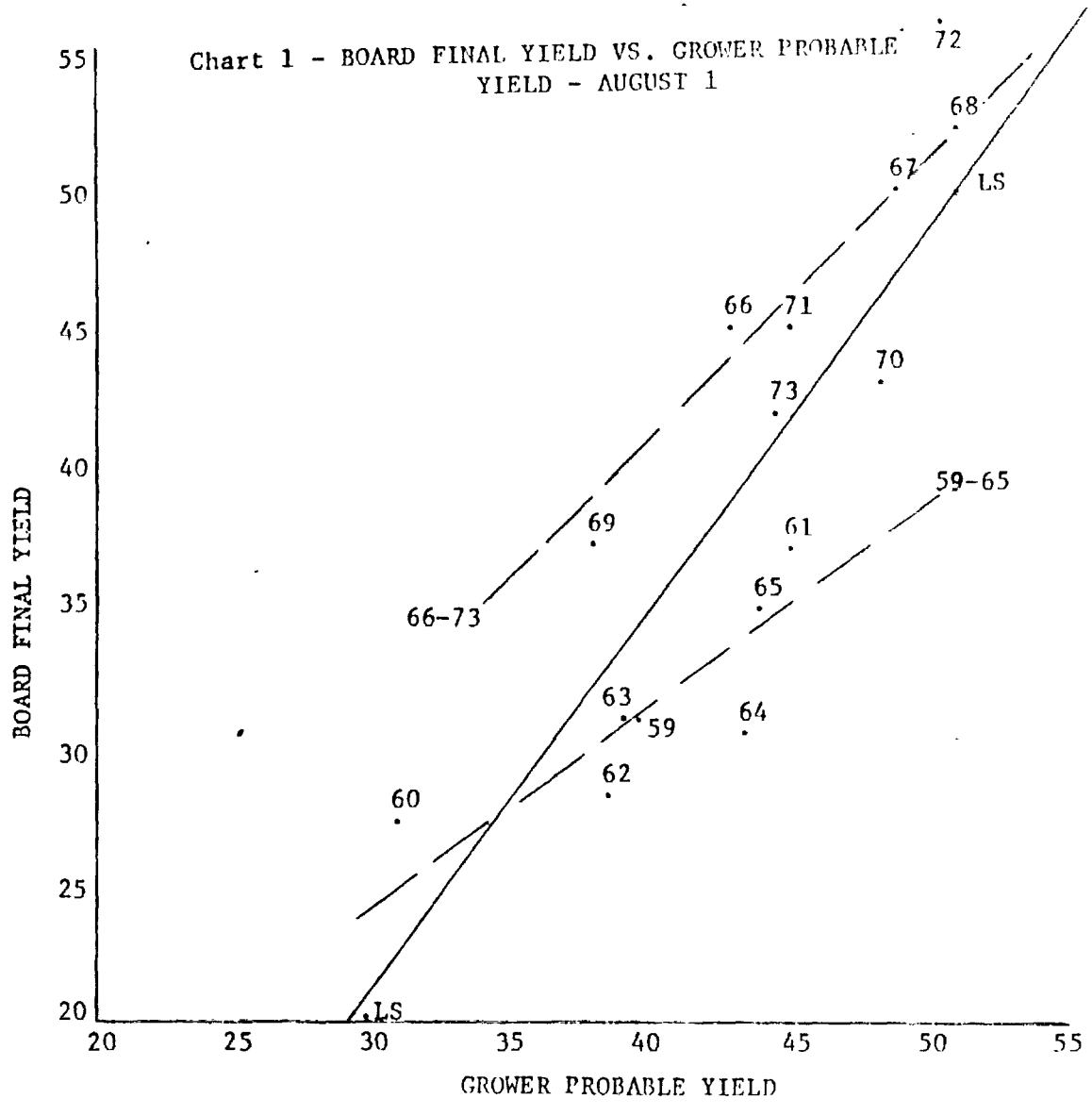


Chart 1 - BOARD FINAL YIELD VS. GROWER PROBABLE YIELD - AUGUST 1
BOARD FINAL YIELD
GROWER PROBABLE YIELD
55
50
45
40
35
30
25
20
20
25
30
35
40
45
50
55
72
68
67
LS
66
71
70
73
59-65
69
61
65
66-73
63
59
64
62
60
LS
. INDIVIDUAL YEAR DATA
- - REGRESSION LINE FOR YEAR

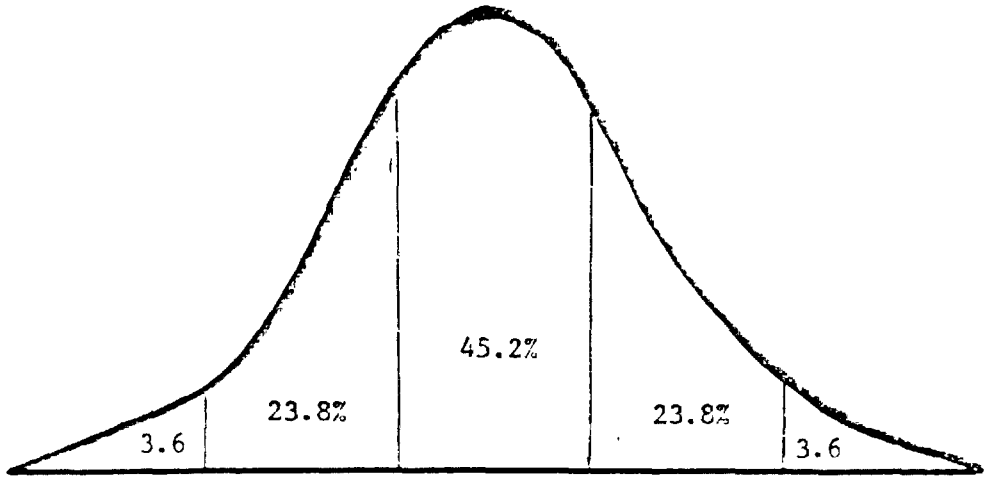
Several new approaches will be discussed which should provide answers to some of these problems, but obviously require evaluation as to their utility for SRS. The first method is referred to as the grower-graded-yield-appraisal. The method seeks to determine the following: (1) What does the grower expect the yield for a specific planting of a crop will be? (2) How does the grower rate (or evaluate) the expected yield of this planting of the crop based on five descriptive categories? The acreages (or areas) planted are then summarized by the five categories and the weighted average expected yield (or expected production) is derived based on the acreages reported.

The descriptive ratings provided by the growers are assumed to be distributed normally according to the grading system suggested by some teachers when a large number of students are to be graded. Thus, the name, grower-graded-yield-appraisal is given to the method since the grower "grades" his own yield appraisal. This grading scheme and its relation to the normal distribution is illustrated by Chart 2.

Experience with this approach in the Dominican Republic indicates that the growers do grade their yields in approximately this manner. That is, 40 to 50 percent of the acreage is reported by growers to have an expected yield which is "average" early in the crop season. The remaining expected yields are either one category above or below the average. This result suggests most growers merely report an average yield early in the crop season. The interpretation of the expected yield as being related to a harvested yield may be in serious error in any year that is not average or normal. Stated another way, most growers may either not be skillful forecasters of crop yields or do not wish to forecast a yield different from their average for purposes of reporting to public agencies. It may be that the most useful information comes from those growers who report a yield which is not average.

The procedure for reporting yield prospects to user agencies, private or public, for the coming harvest is as follows: (1) From sample surveys, report the actual acreage percentages reported by the growers for the grade categories used, (2) Report the growers expected yield, and (3) Derive a within-year average yield for the current year based on (1) and (2). The rationale behind this approach is to provide the grower's expected yield, the descriptive appraisals, and the derived within-year average yield so the data users may agree or disagree with this information as they see fit. Expected production can also be reported to the user in place of yield if this is considered preferable or in case probable production was reported. If the within-year derived average yield differs from the grower's last year's average yield (or a five year average), the user is aware of this difference and may wish to place a somewhat different interpretation or evaluation on crop prospects. For application to specific crops, the normal distribution may be skewed slightly by altering the tail probabilities and X scale values of the model. For example, in the Dominican Republic, coffee and rice are expected to have crop failures less frequently and outstanding crops more frequently than shown in Chart 2 because of increased management inputs and availability of water for rice. Corn and beans are two crops which would be expected to have their distributions skewed in an opposite manner.

Chart 2: Grower Graded Yield Appraisal Curve for a Large Number of Fields



Grade Scale	F	D	C	B	A
Possible crop description corresponding to grades	Crop failure Very poor crop No harvest	Below average Poor crop	Average Normal crop	Above average Good crop	Much above average Very good crop Outstanding Excellent
Uniform yield scale	0 .4 .8 1.2 1.6 2.0				
Midpoint of interval - X_i	.2	.6	1.0	1.4	1.8

Where the range of the yield scale is 0 to 2.0 and each of the 5 grades covers one-fifth of the X-axis (uniform scale).

$$E(X) = \sum_{1}^{5} p_i X_i = 1.00 \text{ (Average yield)}.$$

The scale values developed for use in the Dominican Republic were based on the approximate center of the probability assigned to the interval rather than a uniform scale. However, the merits of alternative scales for crops are still under study.

Center of interval - Z_i	.08	.32	1.00	1.68	1.92

The calculation of the expected value using the normal distribution frequencies and a set of reported data for beans are shown below. E(Z) = 1.20 = (1.92)·(.000) + (1.68)·(.427) + (1.00)·(.443) + (.32)·(.130) + (.08)·(.000). When E(Z) = 1.00 a normal or average yield is indicated. E(Z) = 1.20 in this example shows that the growers expect a yield better than what they consider would be an average or normal for this year.

Crop Condition	Center of Interval	Acreage Percentages for Normal Distribution	Acreage Percentages Reported for Sample Surveys
Much Above Average	1.92	.036	.000
Above Average	1.68	.238	.427
Average	1.00	.452	.443
Below Average	.32	.238	.130
Crop Failure	.08	.036	.000
E(Z)		1.00	1.20
Growers Reported Expected Yield			1.23 cwt./Tarea
Derived Growers Expected Average Yield			1.23 : 1.20 = 1.03 cwt./Tarea
Growers Reported Harvested Yield			1.05 cwt./Tarea

A second method is available which leads to essentially the same information. It can be referred to as the grower's-average-yield-and-appraisal method. For each planting of their crop early in the season the grower is asked for the expected yield followed by a question to determine what the grower considers an average yield to be for the crop planted in the same field. The grower's expected yield (or production) and the average yield for the same acreage are reported for the data user's evaluation. The grower's within-year average yield permits the user to judge whether this figure is consistent with the reported yield of the previous year or years.

An equally important phase of the yield information is to obtain similar information from the same growers after harvest. This second survey provides annual harvested acreage and crop production as well as a grower evaluation of the crop just harvested by five categories. That is, the grower is asked to grade the harvested yield (or production) by the categories given. This information provides a basis for evaluating how good the growers are at forecasting their crop yields early in the season and whether they evaluate the harvested crop in a manner consistent with the model. Early in the season, there appears to be a tendency for the growers to be somewhat pessimistic and after harvest to have a brighter evaluation with regard to the past season.

Copies available from:

International Programs Office
Statistical Reporting Service
Washington, D. C. 20250