

General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

ORIGINAL PAGE IS
OF POOR QUALITY

USERS APPRAISAL OF YIELD MODEL EVALUATION CRITERIA

by

Fred B. Warren

Yield Research Branch
Statistical Research Division
Statistical Reporting Service
U. S. Department of Agriculture

AgRISTARS
Yield Model Development
Project

YM-U2-00740

April, 1982

ORIGINAL PAGE IS
OF POOR QUALITY

USERS APPRAISAL OF YIELD MODEL EVALUATION CRITERIA

by

Fred B. Warren

ztype (File) repouT.DOC;5 (Starting at page)

Yield Research Branch
Statistical Research Division
Statistical Reporting Service
U. S. Department of Agriculture

AgRISTARS
Yield Model Development
Project

^C
@^C
@^C
@

ORIGINAL PAGE IS
OF POOR QUALITY

USERS APPRAISAL OF YIELD MODEL EVALUATION CRITERIA

SUMMARY

The five major potential USDA users of AgRISTARS crop yield forecast models rated the Yield Model Development (YMD) project "Test and Evaluation Criteria" by the importance they placed on them. These users were agreed that the "TIMELINESS" and "RELIABILITY" of the forecast yields would be of major importance to them in determining if a proposed yield model was worthy of adoption by their units. Although there was considerable difference of opinion as to the relative importance of the other criteria, "COST", "OBJECTIVITY", "ADEQUACY", and "MEASURES OF ACCURACY" generally were felt to be more important than "SIMPLICITY" and "CONSISTENCY WITH SCIENTIFIC KNOWLEDGE". However, some of the comments which accompanied the ratings did indicate that several of the definitions and descriptions of the criteria were confusing.

INTRODUCTION

The original "Crop Yield Model Test and Evaluation Criteria" document was prepared without specific inputs from potential USDA users of the crop yield models. There were several reasons for this apparent oversight. First, there was a feeling among the YMD authors that, at least at that time, the potential users were not sufficiently experienced in this field to provide knowledgeable input. Another, related, reason was that the initial document was never intended to be more than an interim procedure, and that the potential users could be approached after we had gained some experience with the criteria as originally established and could also present examples of how the initial criteria were used. Consequently, a specific subtask to identify and query potential users of YMD crop yield models was included in the FY82 YMD Program Implementation Plan. The objectives of this query were to (1) determine what value these potential users actually placed on the eight particular evaluation criteria as defined in the initial document, and (2) obtain whatever suggestions they might have which could be used in developing a more meaningful set of criteria. (The AgRISTARS Foreign Crop Production Forecast (FCPF) and Supporting Research (SR) projects were not considered to be "end-users" of the models, even though the FCPF does require YMD to produce a number of yield models for their use. This is because the FCPF procedures are intended for eventual USDA use.)

The potential USDA users identified, and queried, were the Estimates Division, Statistical Reporting Service (ED/SRS); the World Agricultural Outlook Board (WAOB); the Grain and Feed Division, Foreign Agricultural Service (GFD/FAS); the Crop Condition Assessment Division, Foreign Agricultural Service (CCAD/FAS); and the International Economics Division, Economics Research Service (IED/ERS).

Materials sent to each of the potential users included a copy of the "Crop Yield Model Test and Evaluation Criteria", a short description of each of the eight criteria, a copy of a YMD yield model evaluation report, and a rating sheet. (Copies of the short descriptions of the criteria and of the rating sheet are in the appendix to this report.)

RESPONSES

Responses were received from all five of the potential USDA crop yield model users, although one (FCCAD/FAS) chose not to complete the rating sheet. These ratings are summarized in Table 1. In addition, the following comments appeared to be pertinent and worth repeating.

"What we need are models that reduce RMSE over current procedures - are timely - data can be obtained within budget"

"... you might want to consider another draft of the "Crop Yield Model Test and Evaluation Criteria" paper. We found that we had to reread several sections before we understood the points being made. Any rewrite should emphasize communication rather than content."

"... comparing models with existing methods. One type of benchmark to use in such a comparison is to do a similar evaluation on the published USDA forecasts (from whatever method) of yields."

"...test to determine if the model can accurately estimate the extremes in periods of radical differences in production - low or high - and will be of universal use."

"Timeliness and reliability ...are the key criteria and the question of cost will be highly dependent upon the accuracy and timeliness of the model results."

"...the primary criteria we need are that the estimates be reliable, available in a timely manner, and be affordable. In addition, ... that the indication be consistent and available over time."

It will be seen from Table 1 that TIMELINESS and RELIABILITY were the two most uniformly highly regarded criteria. These were followed, not too closely, by COST, OBJECTIVITY, ADEQUACY, and "MEASURES OF ACCURACY (of the current prediction)". The criteria judged to be least important were SIMPLICITY and CONSISTENCY (with scientific knowledge). My interpretations of these ratings are:

Users Appraisal of Yield Model Evaluation Criteria

Table 1. Summary of users ratings of AgRISTARS yield model evaluation criteria.

Criteria	R A T I N G S					
	Range	Mean	Rank			
RELIABILITY	8 to 9	8.7	2			
OBJECTIVITY	5 to 8	6.3	4			
CONSISTENCY	2 to 9	5.0	8			
ADEQUACY	0 to 9	6.2	5			
TIMELINESS	9 to 9	9.0	1			
COST	5 to 9	7.2	3			
SIMPLICITY	0 to 9	5.2	7			
MEASURES OF ACCURACY	5 to 8	6.0	6			

TIMELINESS -- To be of any value, the crop yield model prediction must be timely, and the earlier in the season, the better. Conversely, a prediction which is not available until after the crop is in the bin generally will not be of much value.

RELIABILITY -- Given two procedures of equal timeliness, the most "reliable" one would be preferred. Also, as indicated by FCCAD/FAS, the measures of "reliability" must include an appraisal of how well the model predictions indicate the full extent of annual changes, particularly the catastrophic changes, in yield/production. Conversely, a model of lesser reliability but greater timeliness may be used as an interim indicator.

COST -- The respondees generally placed a relatively high value on the cost of maintaining and operating the model, but a loose interpretation of one comment might be that "if a procedure is sufficiently reliable, then the cost is relatively unimportant".

OBJECTIVITY -- This is one of several criteria which the respondees had trouble understanding. My own understanding is that an objective model is one where (1) the coefficients for the model have been obtained in some documented manner so that they can be reproduced, and (2) the input variables are measurable. In informal communication, one respondee did question as to whether or not a model could be reliable if it was not also objective.

ADEQUACY -- This is another term which was not well understood, as evidenced by the wide range of values given it. And, frankly, I am not

particularly happy with the present definition! Possibly, a more descriptive label might be UNIVERSALITY, defined as the ease with which the model could be applied to different areas.

MEASURES OF ACCURACY (of the current prediction) -- Again, another confusing term (justifiably so). My own understanding is that the procedure should be able to provide an estimate (confidence interval, if you will) of the reliability of each individual estimate. In this respect, it differs from RELIABILITY in that it refers to the current prediction rather than to the overall historic reliability of the procedure.

SIMPLICITY -- Again, a wide range of opinions as to the relative importance of simplicity. However, with one exception, the respondees are agreed that simplicity should not be a major concern in developing or evaluating a model. Rather, the major concern should be with timeliness and reliability, and like cost, the concern for simplicity should be relegated to instances where the more important criteria are comparatively equal.

CONSISTENCY (with scientific knowledge) -- Again a wide range of opinions, possibly resulting from misunderstandings as to what is meant. (The problem may be one of selecting labels which have been abbreviated too much!) Possibly the definition should be amended to indicate that any changes in the model predictions should have resulted from "believable" impacts of identifiable inputs. (Actually, this criterion was inserted mostly to eliminate "black box" models.)

APPENDIX

ORIGINAL PAGE IS
OF POOR QUALITY

Evaluation Criteria for AgRISTARS Yield Models

RATING SHEET FOR AgRISTARS YIELD MODEL EVALUATION CRITERIA

Please rate each of the criteria listed below in terms of the importance you would place on them in evaluating a new procedure for predicting yields. Score each item on a 0 to 9 scale (0 = no value, 9 = very important). If you would prefer to rate the items subject to certain changes in the criteria, or in their application, your comments will be greatly appreciated.

<u>Criteria</u>	<u>Rating</u>
1. Reliability	_____
2. Objectivity	_____
3. Consistency with scientific knowledge	_____
4. Adequacy	_____
5. Timeliness	_____
6. Cost	_____
7. Simplicity	_____
8. Measures of accuracy of current prediction	_____

ORIGINAL PAGE IS
OF POOR QUALITY

Evaluation Criteria for AgRISTARS Yield Models

YMD YIELD MODEL EVALUATION CRITERIA

1. Yield indication reliability: Reliability will be measured in terms of the historic accuracy and precision of the model. Test statistics may include the Root Mean Square Error (RMSE) and its components, the standard deviation (SD) and the mean (b) of the errors, the proportion of errors which exceeded some arbitrary limit, the correlation between actual and predicted yields, and the ability of the model to predict the actual direction of change in yield from a previous year.
2. Objectivity: No subjective judgments are required either in deriving the model parameters or in obtaining the current input variables. All input values are "measurable" and the methods for estimating (deriving) parameters are fully documented.
3. Consistency with scientific knowledge: Agreement or consistency of a model's form and parameter signs and values with experimental data and scientific knowledge.
4. Adequacy: Adequacy can be assessed in terms of the extent of coverage of a crop, the level of detail provided through the model, and in the appropriateness of the model for intended applications.
5. Timeliness: Timeliness constitutes availability of sufficiently precise or accurate information at the time when the information is needed and can still be utilized.
6. Minimum costs: The primary consideration will be the cost of operating the model after it has been set up. However, the cost of acquiring the models, of establishing any historic data bases needed, and the cost of acquiring the current input data will also be considered.
7. Simplicity: Simplicity in model form and use of input data are often associated with cost. Another advantage of simplicity lies in the often improved ability of the user to understand the concept, capabilities and limitations of the model.
8. Accurate current measures of modeled yield reliability: Model will provide reliable measures of the reliability of current model indications. This characteristic would be evaluated by comparing the model generated measures of reliability with subsequently determining deviations between the modeled and "true" yields.