LARGE AREA YIELD ESTIMATION/FORECASTING USING PLANT PROCESS MODELS

by

Chapman P. Gleason
Supervisory Mathematical Statistician
Statistical Reporting Service, USDA
Washington, D. C.

SUMMARY:
This paper describes the history and purpose of the Statistical Reporting Service, our current operational yield forecasting and estimation methodology, and potential application of plant process models (PPM) to large area yield forecasting. Potential problems, our statistical research and the future of PPM for forecasting large area yields is also discussed.

1907-1982

Food, Fiber and the Future

American Society of Agricultural Engineers

St. Joseph, Michigan 49085

Large Area Yield Estimation/Forecasting Using Plant Process Models
Chapman P. Gleason*
Statistical Reporting Service, USDA

ABSTRACT

Plant process models (PPM) are complex weather and environmental data driven computer simulation models of plant growth. Most models use daily meteorological variables (maximum and minimum temperatures, precipitation, and solar radiation), management variables (plant density, cultivar, etc.), and soil variables (soil properties, initial soil water, etc.) as inputs to simulate the growth of a plant -using a daily timestep - and ultimately estimate gross grain yield of a typical (or "average") plant in a homogeneous plant community. These computer simulation models have become more refined and some are being designed to operate on micro computers so that individual farm operators can make management decisions based on local inputs to the model and derive a potential yield. In contrast to these uses of a PPM, for micro level decisions, several authors have also proposed using PPM's for large area yield forecasting and estimation (3, 13, 17). This paper describes the sampling techniques used currently by the Statistical Reporting Service (SRS) to estimate and forecast net grain yield for large areas and a method to incorporate PPM's into large area yield forecasting and estimating methodology of SRS. The incorporation of PPM into current USDA operational yield estimation procedures presupposes that a model can accurately mimic (or model) plant growth with minimal data requirements. Some user opinion of current PPM simulation methodology and refinements needed to apply these models to large area yield forecasting/estimation is also discussed.

INTRODUCTION

Many computer simulation models of plant growth have been documented in the literature (e.g. 2, 7, 8, 9, 19,21, 24, 29). These physiologically based models require extensive meteorological, management, and soil inputs with varying degrees of accuracy in order to model crop growth. These PPM's have been applied to -- managing water resources, diseases, pests, and several other management factors which influence and ultimately determine crop yield. While each model attempts to mimic plant growth, no PPM of crop growth has been extensively validated for a wide range of growing conditions. This is due to the fact that model development is still going on and also due to the substantial data requirements necessary for an extensive validation.

This paper briefly describes the functions of SRS, how we think we may be able to use a PPM, and the research activities we are developing and supporting to use these physiologically or process based models to forecast crop yields. Our ultimate purpose in using a PPM is to give decision makers in Government, in farm management, and in agribusiness a more timely, accurate appraisal of projected yield based on growing conditions during the current crop year. To present this approach, this paper is organized into the following sections:

1) History and purpose of SRS,
2) Current SRS yield estimation and forecasting methodology,
3) Applying a PPM on a large area basis,

---

* The Author wishes to acknowledge colleagues in the Yield Research Branch for their ideas, critical reviews, and help in preparing this paper.

4) Some problems and pitfalls in the application of PPM's,
5) SRS research with PPM and large area yield forecasting,
6) The future of PPM modeling and large area yield forecasting.

History and Purpose of SRS

The Statistical Reporting Service is the main data collection Agency of the U.S. Department of Agriculture. This organization for collecting and diseminating agricultural statistics has been in operation for more than a hundred years. During that time the name of the organization has changed several times and its job has become larger and more vital, but its main objective -- to report the basic statistical facts of the nation's agriculture -- has remained the same. That is, the Statistical Reporting Service measures the pulse rate, blood pressure, etc. of the nation's agriculture.

Some 275 reports from Washington are issued each year giving current national and state estimates of livestock inventories, crop production, stocks, and prices paid and received by farmers for nearly 200 farm products. Crop reports include estimates of acreage farmers intend to plant, actual acres planted and harvested, and production. Forecasts of production are made for each major crop during the growing season. These forecasts are based on relative crop condition and expected yield per acre as they are reported to the Service on the first of the month. The reports are provided voluntarily by individual growers and farmers. Objective Yield Surveys (OYS) are conducted for several crops to increase the accuracy of the forecasts and estimates.

SRS funds research on yield forecasting through cooperative agreements, grants, and contracts with university scientists, and the employment of 15 full time technical staff (mostly Mathematical Statisticians) engaged in yield research. The Yield Research Branch has primary responsibility to develop new and improved yield forecasting and estimation methodology, and recommend its adoption as part of our ongoing yield surveys.

Current SRS Yield Estimation and Forecasting Methodology

The methodology SRS uses to make production estimates is as follows. Production for any given State is estimated as acreage times yield.

$$\text{Production} = \text{Acreage} * \text{Yield} \quad (1)$$

To estimate acreage SRS conducts a survey in late May and early June called the June Enumerative Survey (JES) which is a stratified random sample of areas of land called segments. Segments are approximately a square mile in size. A similar survey in late November and early December called the December Enumerative Survey (DES) enumerates acreage of fall-sown small grains. The acreage of each crop is estimated from the JES or DES or both in the case of winter wheat. An Objective Yield Survey (OYS) is conducted for most major crops in major producing states. OYS fields are selected from those in the JES or DES samples depending on the crop. The OYS sample is selected randomly with probability proportional to planted acreage intended for grain harvest. The sample size varies widely for individual States and crops. For example, the winter wheat sample size is between 50 and 300 fields depending on the State. Two plots are randomly laid out in each selected field and various counts and measurements are made on the plants in each of these plots at specified times (usually monthly) during the

growing season. Since the primary sampling unit is a field, the two plots within each field are collectively referred to as samples. I will, in the discussion that follows, use wheat as the example, but similar definitions and procedures are followed operationally for corn, cotton, and soybeans, and experimentally for sorghum, rice, and sunflowers. There is also a yield estimation program for potatoes but no forecasting is done.

The yield component in (1) is <u>estimated</u> at harvest as follows:

$$\text{Gross Yield} = \text{Number of Heads} * \text{Grain Weight/Head} \quad (2)$$

Number of heads is estimated by counting the number of heads in each sample. A crop cutting procedure is employed to estimate the weight of grain. Designated sections of the plots are cut, threshed, dried, and weighed to calculate the weight of grain per head.

The gross yield from each sample (two plots) is calculated and expressed in bushels per acre with a standard bushel weight and moisture content equivalent. The average of the sample fields is computed for each State and this is the final estimate of gross yield per acre.

Harvest loss is estimated by selecting a random subsample of the fields selected in the OY sample, laying out post harvest gleaning plots, collecting and weighing the grain remaining in the selected plot after harvest, and expressing it on a per acre basis. Once again the average over all sample fields is the estimate of average harvest loss for a State. The harvest loss estimate is subtracted from gross biological yield to derive net yield.

$$\text{Net Yield} = \text{Gross Yield} - \text{Harvest Loss} \quad (3)$$

When <u>forecasting</u> the final net yield (i.e. prior to harvest), we do not have current counts and measurements of the components in (2). Historic averages and regression equations are used to forecast number of heads and grain weight per head at harvest. Prior to development of measurable plant characteristics, historic averages are used. Once plant measurements can be obtained, regression relationships are established between early season counts and final yield components. The parameters in the current season regression models are estimated from sample data collected in the three previous years.

For example, a linear regression model is developed which relates number of heads at harvest to stalk counts made at or before flag leaf emergence, such as:

$$\text{No. heads at harvest} = A + B * (\text{No. stalks prior to harvest}) \quad (4)$$

The grain weight per head at harvest is forecast using 5-year historic averages and/or regression models based on the number of fertile spikelets, grain number, or head weight depending on the maturity category. For example, after grain filling has begun, one of the regression models is:

$$\text{Grain weight at harvest} = A + B (\text{No. grain prior to harvest}) \quad (5)$$

Regression equations are developed for each wheat maturity category: pre flag, flag or early boot, late boot or flower, milk, soft dough, hard dough, and mature.

When the Crop Reporting Board of SRS convenes, the objective yield estimate is used in determining State yields along with farmer reported yields from mail surveys. Documentation of the methodology used by the Board in setting estimates is given in (25).

Applying PPM on a Large Area Basis

SRS's basic forecasting premise is based on the assumption that the historic regression relationships are appropriate for data from the current year. That is, current plant counts and measurements and their relationship to eventual yield component are properly represented by relationship of the previous three years. However, if environmental conditions for a given crop year are unlike those during the previous three years or even if they are similar but the regression fit is not particularly good, then the forecast may be quite inaccurate. In statistical jargon, we assume that the counts made on a plot in a field (sample unit) came from the range of the X (independent variable) space. Therefore, if the observation (for which one is forecasting) is based on weather, biological or environmental conditions not represented in the data from which the model was developed, one is extrapolating beyond the range of the data, and a potentially substantial but unknown forecast error may result. Since plant process models use meteorological and other environmental data from the current year, and do not require us to base our forecasts on the premise above, we believe that forecasts using a PPM have the potential to be more accurate and reliable. The following example illustrates how SRS intends to use PPMs.

Assume that a plant process model is able to mimic the growth of a plant and can accurately "converge" on gross yield at the end of a growing season. The question becomes, "How do we use PPM and integrate the technology into SRS's current program of estimates?" The following procedure is suggested. A plant process model is run for each sample unit selected in the OY survey. Weather variables, soils, variety, management factors, and other inputs the model requires to simulate weight of grain per plant are collected for the plant population at that particular site. Depending on the climate, it may be possible to interpolate the daily meteorological inputs from the nearest weather stations. In dry climates it may be necessary to actually measure the rainfall at the sample field. The yield is simulated for that particular site via the PPM and is then averaged over all samples to give a forecast/estimate based on the plant process methodology. Note that there is no conceptual difference between using a PPM and a regression model, only the data base required to execute the model is different -- the regression model requires historic data to estimate the coefficients and current plant counts and measurements for the regression variables and the PPM requires current year initial inputs and weather data for the entire growing season. Since a PPM must have weather data for the entire growing season, yield forecasts necessitate us to simulate future weather from the forecast date to harvest. A stochastic weather simulator (6, 16) will be used to generate future weather inputs for each location.

Several forecasted yields could be generated for each given site. For example, one could forecast a yield based on simulated future "unfavorable" or stressed weather conditions, a yield based on simulated future "normal" weather conditions, and a yield based on simulated future "favorable" weather conditions. The terms "unfavorable", "normal", and "favorable" would have to be defined and quantified. The yields for each of these classes or conditions would be averaged to give a State level forecast for "stressed" conditions, for "normal" conditions, and another for "favorable" conditions, i.e.:

$$\hat{Y}_{stressed} \pm S. E. (\hat{Y}_{stressed})$$

$$\hat{Y}_{normal} \pm S. E. (\hat{Y}_{normal})$$

$$\hat{Y}_{favorable} \pm S. E. (\hat{Y}_{favorable})$$

$\hat{Y}_{stressed}$ is the forecast of yield made under stressed conditions, $S.E.(\hat{Y}_{stressed})$ is the standard error of $\hat{Y}_{stressed}$. Decision makers at the federal, state, and local level would have several forecasts which could help to answer questions related to "What if" analyses. Probable future weather conditions could possibly be obtained from long range projections from the National Oceanographic and Atmospheric Administration (NOAA) for different regions of the country to determine which of the three conditions is most likely to occur. In essence SRS would be moving from a point forecast to a confidence interval approach in yield forecasting with PPM models.

Some Problems and Pitfalls in the Application of PPM's

There are several problems in using plant process models for large area yield forecasting. They are:

(1)   The PPM models have not been adequately researched, documented, and validated on a local area (plot or site) basis.

(2)   The models require detailed site specific data input which is expensive, and prohibitive to collect on a large-area basis. Examples are models that require calculations of leaf area to run the model (7). Analyses need to be conducted to determine the effect of using estimated rather than actual measured inputs.

(3)   A great deal of computer expertise is required to execute the model and manage the data.

(4)   Extensive training and background in plant physiology and statistical modeling techniques are required to fully understand the application of the model.

(5)   Questions and problems related to generating and using simulated future weather have yet to be defined and worked out.

SRS Research With PPM's and Large Area Yield Forecasting

Addressing only the first problem listed above, SRS is doing extensive sensitivity analysis research on PPM models (4, 5, 14, 15, 30). In this context, sensitivity analysis is the application of statistical techniques to a computer simulation model to determine the relative importance of input variables, internal subroutines, parameters, and functions of the model.

The objective of sensitivity analysis is to determine if all variables and parameters are required in a PPM and to determine how accurately these variables need to be estimated. A variable which is necessary to run a PPM but does not substantially affect yield does

not need to be collected with a great deal of precision, so less costly methods of data collection could be used. For example, soil moisture is a very costly variable to collect if core samples are required. However, if a model could run on a "wet", "dry", "adequate" categorization from a visual inspection using a shovel, this would be less costly to collect.

Several different statistical techniques have been used to do sensitivity analysis:

(1) Fractional and full factorial experimental designs, and variations of these experimental design techniques (26, 27, 28),
(2) Response surface methods (4, 5, 10, 11, 12),
(3) Time Series and Kalman Filtering techniques (22, 23).

Our sensitivity analysis approach has been from a research point of view, with the ultimate objective of trying to apply PPM's to the large area yield forecasting problem. We have done extensive literature review in the area, funded research on development of PPM models (1, 20), and funded research on evaluation of these models (18). We are actively discussing a plan for a pilot survey to test a PPM for an entire State.

The Future of PPM's and Large Area Yield Forecasting

Currently SRS feels that most models require too many detailed data inputs to be operationally feasible as a large area forecasting methodology. However, this could be overcome with newer telecommunications and microprocessor technologies which could record local weather and possibly plant water and nutrient stresses for every one of our OY sample fields. A central computer could then poll the sites for site specific weather information. This approach requires extensive hardware, management, and large scale planning and development. We will have to feel confident that a PPM is an excellent candidate for large area yield estimates to be able to assume the capital risk to develop a system to implement a PPM based estimation system.

Another approach would be to use environmental satellite inputs for temperature, solar radiation, and moisture (or thermal IR/ET relationship). Research in this area is underway and both temperature and solar radiation test models are available for evaluation. Of course we would like to develop a PPM which is sensitive to climatic, environmental, pest, and management stresses and have minimal input data requirements. We believe some model in the "middle ground" between a very detailed PPM and the grosser level modeling such as a regression of county yields on average July precipitation and temperature could possibly meet our needs.

Finally, SRS feels that for a PPM to be useful in our operational program, that the model should adjust itself during the simulation run to use actual counts and measurements collected in the field for updating the model's internal "state" variables. This updating of a simulation model has been termed "feedback" by some modelers (20). The incorporation of feedback data should help a simulation model to converge at harvest to gross grain yield.

References

(1) Arkin, G. F., S. J. Maas, and C. W. Richardson. "Forecasting Grain Sorghum Yields Using Simulated Weather Data and Updating Techniques". Transactions of ASAE 23(3), 676-680. 1980.

(2) Arkin, G. F., R. L. Vanderlip, and J. T. Ritchie. "A Dynamic Grain Sorghum Growth Model". Transactions of the ASAE 19(4):622-630. 1976.

(3) Arkin, G. F., C. L. Wiegand, and H. F. Huddleston. "The Future Role of A Crop Model in Large Area Yield Estimating". Proceedings of Crop Modeling Workshop. Columbia, Missouri, October 3-5, 1977.

(4) Baker, F. D. and R. F. Bargmann. "Analysis of Simulation Models by Response Surface Techniques". Proceedings of American Statistical Association Annual Meeting. Detroit, Michigan. August 1981.

(5) Baker, F. D. and R. F. Bargmann. "Orthogonal Central Composite Designs of the Third Order in the Evaluation of Sensitivity and Plant Growth Simulation Models". Proceedings of The American Statistical Association Annual Meeting. Cincinnati, Ohio. August 1982.

(6) Crank, K. "Simulating Daily Weather Variables". Research and Development Branch, Research Division, Economics, Statistics, and Cooperatives Services, USDA, Washington, D. C. 20250. 1977.

(7) Feyerherm, A. M., E. T. Kanemesu and G. M. Paulsen. "Response of Winter Wheat Grain Yields to Meteorological Variation". NASA Final Report. NASA/JSC, 14282. 1977.

(8) Golley, F., O. L. Lange, and J. S. Olson. "Grassland Simulation Model". Volume 26 Springer-Verlag Inc., New York. 1978.

(9) Hill, R. W., D. R. Johnson, and K. H. Ryan. "A Model for Predicting Soybean Yields from Climatic Data". Agron. J. 71:251-256. 1979.

(10) Iman, Ronald L. "Sensitivity Analysis to Reduce the Complexity of Computer Models in the Face of Limited Data", Proceedings of American Statistical Association Annual Meeting. Cincinnati, Ohio. August 1982.

(11) Iman, Ronald L., Jon C. Helton, James E. Campbell. "An Approach to Sensitivity Analysis of Computer Models: Part I -Introduction, Input Variable Selection and Preliminary Variable Assessment". Journal of Quality Technology. Vol 13, No. 3. July 1981.

(12) Iman, Ronald L., Jon C. Helton, James E. Campbell. "An Approach to Sensitivity Analysis of Computer Models: Part II - Ranking of Input Variables, Response Surface Validation, Distribution Effect and Technique Synopsis". Journal of Quality Technology. Vol 13, No. 4. October 1981.

(13) Keener, M. E., E. C. A. Runge, and B. F. Klugh. "The Testing of a Limited Corn Yield Model for Large-Area Corn Yield Prediction". Journal of Applied Meteorology. Vol 19, No. 11. November 1980.

(14) Larsen, Greg A. "Application of Path Coefficient Analysis to a Plant Simulation Model". Statistical Research Division, Statistical Reporting Service, Washington, D. C. 20250. December 1982.

(15) Larsen, Greg A. "Progress Report on the Evaluation of Plant Growth Models". Statistical Research Division, Statistical Reporting Service, Washington, D. C. 20250. Staff Report No. AGES810716. 1981.

(16) Larsen, G. A., and R. Pense. "Stochastic Simulation of Daily Climatic Data for Agronomic Models". Agronomy Journal, 74:510-514. 1982.

(17) LeDuc, S. K., T. Hodges, and V. French. "Large Area Yield From Plant Simulation Models". Proceedings of American Statistical Association Annual Meeting. Cincinnati, Ohio. 1982.

(18) Maas, S. J., and G. F. Arkin. "Sensitivity Analysis of SORGF A Grain Sorghum Model". Transaction of ASAE. Volume 23, No. 3, pp 671-675. 1980.

(19) Maas, S. J. and G. F. Arkin. "TAMW: A Wheat Growth and Development Simulation Model". Report No. 80-3. Texas Agr. Exp. Station, Blackland Research Center, Temple, Texas. 1980.

(20) Maas, S. J. and G. F. Arkin. "User's Guide to SORGF: A Dynamic Grain Sorghum Growth Model With Feedback Capacity". Research Center Program and Model Documentation No. 78-1. Texas Agr. Exp. Station, Blackland Research Center, Temple, Texas. 1978.

(21) Meyer, G. E., R. B. Curry, J. G. Streeter, and C. H. Baker. "SOYMOD as a Management Tool for Soybean Production". Paper No. 81-4012, 1981 Summer Meeting, American Society of Agricultural Engineers. 1981.

(22) Montgomery, Douglas C., Robert G. Conard. "Comparison of Simulation and Flight-Test Data for Missile Systems". Simulation. February 1980.

(23) Montgomery, Douglas C., and F. D. Baker. "Time Series Methods Applied to a Computer Simulation Model of Soybean Growth". Draft to be published.

(24) Ritchie, J. T. "Wheat Growth and Yield Model: Fortran Program". Unpublished.

(25) Scope and Methods of the Statistical Reporting Service, Misc. Publication No. 1308, U. S. Department of Agriculture, Statistical Reporting Service, Washington, D. C. 20250.

(26) Smith, Dennis E., and Carl A. Mauro. "Factor Screening in Computer Simulation". Simulation. February 1982.

(27)   Steinhorst, R. K. "Parameter Identifiability Validation and Sensitivity Analysis of Large System Models". pp 35-58 In: Systems Analysis of Ecosystems, International Cooperative Publishing House, Fairland, Maryland, 1979.

(28)   Steinhorst, R. K., G. Innis, and R. V. O'Neil. "Stochastic Difference Equation Models of Biological Systems". pp 1-22 In: Systems Analysis of Ecosystems, International Cooperative Publishing House, Fairland, Maryland. 1979.

(29)   Watts, D. G., and R. J. Hanks. "A Soil-Water-Nitrogen Model for Irrigated Corn on Sandy Soil". Soil Science Soc. Am. J. 42:492-499. 1980.

(30)   Wilson, W. W., T. L. Barnett, S. K. LeDuc, and F. B. Warren. "Crop Yield Model Test and Evaluation Criteria", AgRISTARS Yield Model Development Project, Report No. YMC-1-1-2 (80-2.1). 1980.