WORLD METEOROLOGICAL ORGANIZATION

AGRICULTURAL METEOROLOGY

CAgM REPORT NO. 14

MATHEMATICAL MODELS IN AGROMETEOROLOGY

J.R. Haun

CAgM Rapporteur on Mathematical
Simulation Modelling in Agrometeorology

1983

WORLD METEOROLOGICAL ORGANIZATION

Subject: CAgM Report No. 14

Distribution: (i) Members of CAgM

(ii) Members of CAgM Working Groups and Rapporteurs who are not members of CAgM

MATHEMATICAL MODELS IN AGROMETEOROLOGY

This report prepared by the CAgM Rapporteur on Mathematical Simulation Modelling in Agrometeorology is approved by the President for circulation to members of CAgM. A copy of the report is therefore attached for information.

Introduction

This report was developed from a review of approaches to the modeling of plant-environment relationships. The review was originally intended to cover the more specific area of mathematical simulation modeling in agrometeorology. However, it soon became apparent that the term simulation had several rather different connotations and interpretations, at least within the area of agrometeorology. A very general view is that any mathematical model relating plant responses to external and/or internal conditions "simulates" the relationship. From a mathematical or statistical point of view the term could be reserved for the process of generating data or information from a probability distribution. In addition, there are many other differences among investigators as to the terminology defining types of models and procedures for their development. In light of these differences a discussion of terminology was considered pertinent.

As in most areas of scientific investigation, the increase in publications of crop-weather or plant-environment models has been prodigious. Computer procedures have greatly increased the speed and scope of analyses that ultimately result in some type of model. Unfortunately the criteria for publication frequently do not include sufficient requirements for validation of results by authors. Comments such as,"predicted values, agree well with the observed, "are accepted by some editors, even though the "observed" values were used to formulate the model. Thus, the extensive volume of current literature is not an indication of its value.

Any review or characterization of the state of the art in any field presents questions of goals and priorities. The success of one project should not be confused with that of another with different objectives. Also, various levels of utility may be visualized. A model that accurately represents all possible effects of meteorological events on the development and yield of a plant or crop is obviously more useful than a model that is only applicable when one or more environmental parameters is held constant. Although it is sometimes difficult to distinguish between economic and academic values of research results it would be quite helpful if more workers in modeling would provide more discussion on this aspect of their work. Since this perspective is missing from many papers, their value, degree of success, or relation to other work may be somewhat different than their authors intended.

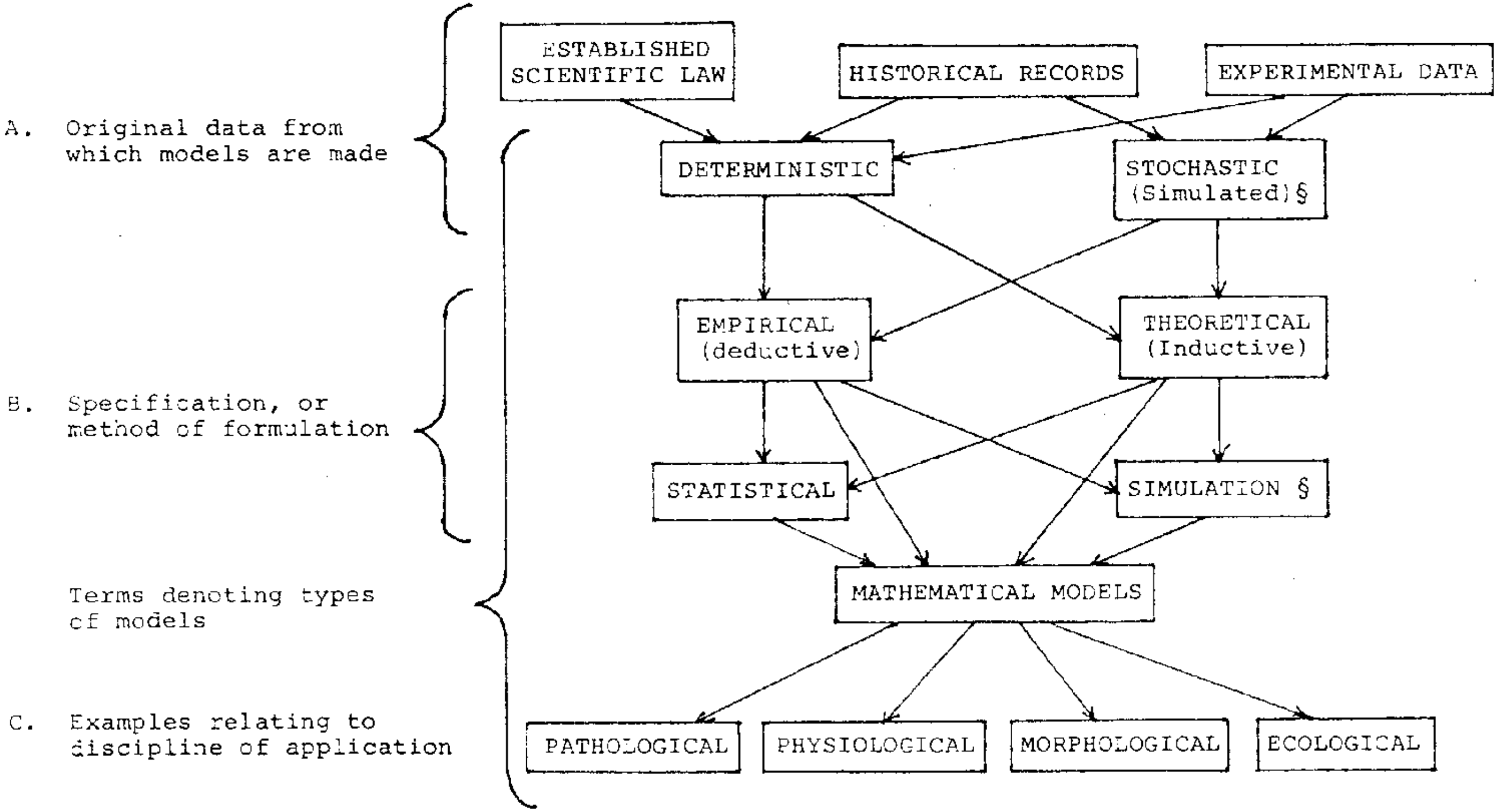
Terminology

Diversity in classifications of models among various authors has been discussed by Baier (1979). Terms in use suggest that there is a search for additional dimensions to describe types of models. Frequently the terminology tends to restrict or polarize model types to only two or three alternatives. Clearcut distinctions are difficult in many cases. Further effort to structure the nomenclature of models is given in Fig. I. If one considers that model names should reflect the types of data from which they are developed and also the types of procedures used in their computation it is evident that many combinations of terms could be used for types of models in existence.

The arrows in Fig. I show potential connections or associations of terms. Although the term "simulation", when used in model development, is considered a statistical procedure, it is contrasted in this figure with the term "statistical" which refers to all other statistical procedures, such as regression analysis. Simulation can also refer to the output of a model under hypothetical conditions. Data for development of a model may be used in its original form (Deterministic) or in a simulated form (Stochastic) generated from a probability distribution. Specification of the model is then accomplished by some preconceived or "established" plan (Theoretical) or some arbitrary "fitting" (Empirical) procedure. Frequently there is a mixture of the two. Even when a theoretical basis for a model is used, the same statistical process, such as least squares analysis, may be used to determine coeffieients. Various dual meanings to terms in use, plus the frequent mixture of modeling procedures with seemingly opposite implications of their use, have contributed to some confusion in this area of research. This problem has been emphasized because workers in modeling have frequently been preoccupied with methodology when the focus should have been on the validation of model output, i.e., if a model will predict accurately under conditions different from those on which it was based, then it has potentially incorporated fundamentally sound functional relationships and procedures of analysis. On the other hand, if it does not predict accurately, then there is little justification for extensive documentation in the literature.

FIGURE I

PLANT-ENVIRONMENT MODEL TERMINOLOGY



§ Generated from a probability distribution

Many models have been named with letters of abbreviation or acronyms which have meaning only to the authors and others who may have obtained their particular computer program (Hesketh and Jones 1976). Since many of these models are composed of numerous submodels formulated in a variety of ways, it would be difficult to devise names that would simultaneously represent data origins and analytical procedures.

In summary, it may be said that models are generally poorly indentified or described by a simple set of terms. Methodology is so diverse within individual models and among research workers that a successful unification of terminology is unlikely. It is predicted that more specific and useful terms will come into usage when successful models are finally recognized by a majority of workers in the field.

Comments on Current Methodologies in Modeling

Considering the level of interest and research effort concentrated on modeling there should be some similar thread of purpose or overall objective that is common to all mathematical agrometeorological models. For the purposes of this review it is assumed that all models on this subject are intended to provide some numerical relationship(s) between plant (or crop) development (or yield) and environmental conditions. Environmental conditions include internal and external conditions of plants that may be controlled and/or result from weather variations in the air or soil. This, of course, includes a very wide range of subject matter; e.g., the response of suspensions of chloroplasts to light and temperature variations, as well as yield responses of a crop to variations in photoperiod. In fact, there are few investigations within the many fields of plant science (where treatments are influenced by environment) that could not be described by some sort of mathematical model.

It is possible that current interest in modeling has been stimulated by the hope that modern computers and newer analytical techniques can finally make possible more useful predictions and profound explanations of plant processes by more accurate evaluation and/or by simultaneous analysis of many related processes influenced by environment. Unfortunately, these ultimate goals have not been reached.

Much of the plant- and crop-modeling literature deals with results of experiments under controlled environments. Over several decades since the first"phytotrons" were built many important discoveries, such as photoperiodic responses of plants, have been discovered by these procedures. Still, there has been very limited success in operational yield prediction in the field using large models aggregated from diverse sources, including controlled environment and laboratory experiments. Since many plant processes are known to be influenced by environmental factors it is not surprising that attempts would be made to construct large models with such sub-models. It is logical to break any large problem up into smaller portions that can be studied more intensively and with more control. However, when submodels are later joined, their structure may automatically introduce errors because of: (1) selection of experimental conditions that are not representative of potential conditions for which a "large" model may be used, (2) use of improper functional relationships or incomplete representation of total response curves, and (3) incorrect interactions among various submodels and interactions among variables. Of course, the more submodels the more chance for introduction of error. Thus, the problem with many large models is that they combine fragmentary bits of data that may seem adequate from the standpoint of physical or engineering systems but are quite inadequate to detail the complexity and variability of biological systems.

An alternative to the controlled environment approach is the use of data from plants grown under natural field conditions. Assuming there is adequate representation of conditions other than environmental, and that sampling procedures are appropriate, it is possible to record plant responses to a wide range of environmental conditions in a relatively short time. Within an entire season there are seldom any two days with the same values for all of the usual environmental parameters recorded. Two prerequsites are necessary for success and efficiency of this approach: (1) accurate measurement of some appropriate dependent variable, i.e. plant responses, such as size or chemical composition, and (2) appropriate computer procedures. Under such field conditions the range of values for independent variables is not only much more likely to be representative of future situations where model operations may be tested, but more importantly they accurately represent the interactions of variables.

Statistical and computer analyses of plant-environment relationships have involved various procedures, many of which are adapted from other disciplines where they were used earlier and/or more extensively, e.g. CSMP in engineering; time-series forecasting in economics. Considerable use has been made of more elementary means of establishing statistical significance of relationships (correlation) and the nature of relationships (regression). Unfortunately, misuse and misinterpretation of regression techniques have led to many useless models and, also, unjustified criticisms of successful applications. A most frequent and obvious misuse of regression statistics is the notation of values of R^2 (coefficient of determination) with no mention of number of observations or number of variables used in the analysis. This, of course, neglects the fact that the R^2 automatically increases with decreases in number of observations and/or with increases in the number of independent variables, without compensatory increases in accuracy of predictions when the model is applied to strange data. On the other hand, rigid adherence to statistical prerequisites for regression could prevent its use in many plant-environment analyses. The requirements for variables with "normal distribution" and freedom from intercorrelation among variables is very difficult to satisfy with most experimental data. Analysis of data from field sites frequently requires the use of temperature, moisture and solar radiation data that are strongly intercorrelated. Resort to controlled environment conditions to correct this problem obscures important interactions of environmental factors on plant responses. On balance, it should be said that statistical regression techniques represent, in many cases, the only available means for mathematically describing and evaluating the relationship of a given plant response to one or more environmental variables, but despite their inherent limitations, successful models can be obtained by their use.

Multiple regression, principal component and other related analyses have been useful, particularly with the use of computers, in the comparison of variables. When a plant response is influenced by a large number of variables is of interest to have at least a suggestion of rank in significance among variables. This information may be useful in the explanation of historical events or in formulation of models to use in subsequent situations. Reasons for questioning results, as mentioned earlier, relate to the scope of variation of individual variables and to interrelationships among variables. The regression of yield on averaged monthly weather data for periods as short as 20 to 30 years exemplifies the limitations of this method of analysis. In this example adverse levels of specific monthly averages can be associated with poor yields, thus "explaining" certain historical events, yet the developed model will not predict accurately on an operational basis. This inaccuracy is considered to be due to the fact that the time series of years is not long enough (abnormal distribution) to represent sufficient variation in monthly-averaged variables. It must be emphasized that an essential requirement in formulation of a model is that it be tested on data that is different (i.e. from another period or set of conditions) from that on which it was based. This step is the most neglected aspect of current research in modeling. It is a laborious task to test many models with actual field conditions, but unless this link between theory and practice can be made the literature will continue to be glutted with unconfirmed and frequently misleading papers. Unfortunately, papers of this type are produced rather easily with the aid of computer simulation procedures.

Proceeding to commentary on more complicated simulation models composed of many submodels with the same inherent limitations mentioned above, it may be said that the problems in their construction are quite formidable. Passioura (1973) has compared results of this approach with those from other disciplines:

> '.... there have been highly successful simulations of physical systems. Rocket technology, for example, would not have reached the advanced stage it has without simulation models. It is tempting to extrapolate such success to biology. But there is a highly important difference between a rocket and an organism. The underlying physics on which the design of a rocket is based has been thoroughly worked out. The processes involved are, in general, well known. The complexity arises mainly because of the complicated interfaces between the various parts of the system. By contrast, the processes occurring in an organism are poorly understood. With a rocket we can use simulation as a sophisticated cement to bind well-fitting pieces of (the) jigsaw together. With an organism many of the

> jigsaw pieces are missing, and to pretend that we have them is to delude ourselves.'

Summary

The literature on plant-environment modeling has expanded greatly in recent years. Unfortunately, most of the models are academic exercises that have no immediate "practical" use. Many of them are in the research stage and are limited in value because of incomplete theoretical basis and/or insufficient data base. Due to the inherent limitations of most plant-environment data (scope, distribution, intercorrelation, etc.), improper statistical procedures have been used and their results interpreted incorrectly in many cases. Much of this difficulty could be corrected or avoided by the use of sufficient testing of models on "strange" data not used in the original formulation of the model. Simulation procedures have many potential uses in modeling. However, current models of this type are, like chains, no better than their weakest links. Efforts to aggregate as many separate plant processes as possible into large models are limited by the tentative or "approximate" nature of many submodels and insufficient knowledge of interactions among submodels.

Selected Bibliography

Baier, W., 1979. Note on the terminology of crop-weather models. Agric. Meterol., 20: 137-145.

Baker, D. N., Hesketh J. D. and Duncan, W. G., 1972. Simulation of growth and yield in cotton: I. Gross photosynthesis, respiration and growth. Crop Sci., 12: 431-435.

Bauer, A., 1972. Effect of water supply and seasonal distribution on spring wheat yields. N. Dakota Agr. Exp. Sta., Bulletin 490, N. Dak. Univ., Fargo, N. Dak.

Benci, J. F., Runge, E. C. A., Dale, R. F., Duncan, W. G., Curry, R. B. and School, L. A., 1975. Effects of hypothetical climate changes on production and yield of corn. Dept. of Transp., DOT-TST-75-55 (Part 2), 4-3--4-37.

Biswas, A. D., 1978. Current state of model use in policy formulation. Ecol. Modelling, 4: 71-73.

Box, G. E. P., Jenkins, G. M., 1970. Time-series Forecasting and Control. Holden-Day: San Francisco.

Clements, H. F., 1964. Interaction of factors affecting yield. Ann. Rev. Plant Physiol., 15: 409-442.

DeWit, C. T., 1971. Dynamic concepts in biology. The Use of Models in Agricultural and Biological Research. In Proc. of A.R.C. Symp., Ed. J.G.W. Jones.

Duncan, W. G., 1972. SIMCOT: A simulator of cotton growth and yield. In Workshop Modeling Tree Growth, p. 115-118., Proc. C. M. Murphy et al. (eds.) Duke Univ.

Fisher, R. A., 1924. The infulence of rainfall on the yield of wheat at Rothamsted. Phil. Trans. Roy. Soc. London. Ser. B, 213: 89-142.

Gross, E. R., Rust, R. H., 1972. Estimation of corn and soybean yields utilizing multiple curvilinear regression methods. Proc. Soil Sci. Soc. Amer., Vol. 36, No. 2.

Haun, J. R., 1974. Prediction of spring wheat yields from temperature and precipitation data. Agron. J., 66: 405-409.

Haun, J. R., 1975. Potato growth-environment relationships. Agric. Meteorol., 15: 325-332.

Hesketh, J. D., Baker, D. N., and Duncan, W. G., 1972. The simulation of growth and yield in cotton: II. Environmental control of morphogenesis. Crop Sci., 12: 436-439.

Hesketh, J. D. and Jones, J. W., 1976. Some comments on computer simulators for plant growth. Ecol. Modelling, 2: 235-247.

Holcomb Research Institute for the Scientific Committee on Problems of the Environment (The), 1976. Environmental Modelling and Decision Making. Praeger Publishers, Inc., New York, N. Y., 152 pp.

Holt, D. A., Bula, R. J., Miles, G. E., Schreiber, M. M. and Peart, R. M., 1975. Environmental physiology, modeling and simulation of alfalfa growth, 1. Conceptual Development of SIMED. Research Bulletin 907. Agri. Exper. Sta. Purdue University, West Lafayette, Ind., in cooperation with the Agri. Res. Sev., U. S. Dept. of Agri.

Jeffers, J. N. R. (Ed), 1971. Mathematical Models in Ecology. Blackwell Scientific Publications Oxford 398p. 12th Symposium of British Ecol. Soc. March 1971.

Jordan, K. A., Takakura, T., Buffington, D. E. and Boyd, L. L., 1975. Modeling air temperature for plant and animal simulation. J. Agr. Met., 31: 125-134.

Leeper, R. A., Runge, E. C. A. and Walker, W. M., 1974. Effect of plant available stored moisture on corn yields. Agron. J., 66:723-73

McQuigg, J. D., 1975. Economic Impacts of Weather Variability. University of Missouri-Columbia, Department of Atmospheric Science, 256 pp.

Passioura, J. B., 1973. Sense and nonsense in crop simulation. J. Australian Inst. Agric. Sci., 39: 181-183.

Robertson, G. W., 1973. Development of simplified agroclimatic procedures for assessing temperature effects on crop development. In: R. O. Slatyer (Editor). Plant Response to Climatic Factors. Unesco, Paris, 1973. Proceedings of the Uppsala Symposium, pp. 327-34

Runge, E. C. A. and Benci, J. R., 1975. Modeling corn production - estimating production under variable soil and climatic conditions. Proceedings of the Thirtieth Annual Corn and Sorghum Research Conference, 1975, pp. 195-214.

Shaw, R. H., 1977. Use of a moisture-stress index for examining climate trends and corn yields in Iowa. Iowa St. J. Res., 51: 249-254.

Shawcroft, R. W., Lemon, E. R., Allen, L. H., Stewart, D. W. and Jensen, S. E., 1974. The soil-plant-atmosphere model and some of its predictions. Agric. Meteorol., 14: 287-307.

Steyaert, L. T., LeDuc, S. K. and McQuigg, J. D., 1978. Atmospheric pressure and wheat yield modeling. Agric. Meteorol., 19: 23-34.

Thompson, L. M., 1969. Weather and technology in the production of wheat in the United States. J. Soil Water Conserv., 24: 219-224.

Thompson, L. M., 1969. Weather and technology in the production of corn in the U. S. corn belt. Agron. J., 61: 453-456.

Thompson, L. M., 1970. Weather and technology in the production of soybeans in the Central United States. Agron. J., 62: 232-236.