

subsequent validations. Boote and Pickering describe the essential components of a row crop photosynthesis model. The available C supply and the demand by different plant C sinks are used by DeJong and Grossman to develop a model for annual vegetative and reproductive growth in deciduous fruit trees.

The CO₂ level in the atmosphere is expected to double in the next 50 to 100 years, resulting in a 2 to 3°C increase in Earth's temperature (Gates, 1983). How this will affect life processes, including photosynthesis, is of great concern. Reddy, Pachepsky, and Acock describe a model that can be used to simulate the effect of CO₂ and temperature on canopy photosynthesis.

Literature Cited

- Bonner, J. 1962. The upper limit of crop yield. *Science* 137:11–15.
- Gates, D.M. 1965. Radiant energy, its receipt and disposal, p. 1–26. In: P.E. Waggoner (ed.). *Agricultural meteorology*. Meteorol. Monogr. 6. Amer. Meteorol. Soc., Boston.
- Gates, D.M. 1983. An overview, p. 7–20. In: E.R. Lemon (ed.). *CO₂ and plants*. Westview Press, Boulder, Colo.
- Whisler, F.D., B. Acock, D.N. Baker, R.E. Fye, H.F. Hodges, J.R. Lambert, H.E. Lemmon, J.M. McKinion, and V.R. Reddy. 1986. Crop simulation models in agronomic systems. *Adv. Agron.* 40:141–208.

Mechanics of Model Building

James M. McKinion and Ronaldo A. Sequeira

Crop Simulation Research Unit, Agriculture Research Service, Mississippi State University, MS 39762-5367

In the late 1960s, with the advent of wide availability of the digital computer, it became possible by applying systems theory to build simulation models of biological entities. In production agriculture, scientists have built models for various field crops: cotton (*Gossypium hirsutum* L.), corn (*Zea mays* L.), wheat (*Triticum aestivum* L.), soybean [*Glycine max* (L.) Merr.], and alfalfa (*Medicago sativa* L.). In most of the models built to date, crop phenology is simulated directly with detailed aspects of crop physiology. Some models also rely on principles of soil physics to some degree. This paper addresses the methodology of constructing process-level, physiologically based crop simulation models that possess detailed descriptions of plant physiological and soil physical mechanisms and proposes a minimum set of design features.

THE DATABASE

Data availability largely determines the type of model of a physiological process, or an entire crop, that a modeler can develop. Unfortunately, most field-crop experiments are designed to determine either differences between means or the significance of specific treatments, not an understanding of the cause of treatment differences. As a result, one cannot determine from field data which external events caused responses in the biological–physical system. If a model is developed from such a data set, that model will be valid only for data conditions at that site for the years over which the data were collected and cannot reliably be extended to other sites or other years.

In a simulation model, detailed information must be available on all exogenous variables that can affect the process or crop being studied. For example, if one were modeling leaf growth and development, many variables would have to be known to build a comprehensive leaf model: date the leaf was initiated and the change in leaf size, N concentration, water potential, temperature, exposure to solar radiation, dry weight, stomatal conductance, and photosynthesis rate. Of course, most of these variables would have to be measured from leaf initiation through senescence. If significant information is missing from this data set, the type of model one can construct is limited. If temperature data were missing, the job would become very difficult. If leaf N concentration data are missing, a leaf model could be built, but the model would be confounded by the unknown leaf N status under which the data set was collected.

In applying systems methodology to modeling biological systems, experiments are designed to determine cause and effect and measure external variables pertinent to the system being studied. Even when one strives to be comprehensive in field research experiments, certain classes of experiments cannot be conducted in the field. For example,

determining the isolated effect of temperature on the rate of single leaf area expansion would be almost impossible to do in the field. The first problem is the range of temperatures the leaf would experience at a particular site. The second problem is the confounding effect of N variation under field conditions. A third problem is the confounding effect of water stress on a field crop. Any model built describing leaf area development from such a set of field-collected data would have these confounding effects built into the model.

If one had enough resources, land, and time, field experiments could be conducted across the ecological range of the crop under variable conditions of temperature, solar radiation, water regimes, and plant nutrition so that one could unravel cause and effect in field-collected data sets. These would have to be replicated over sites and years in hundreds of locations for each crop being modeled. However, there is a better way that is much more efficient and much more accurate for collecting data for building process-level models.

Naturally sunlit-controlled-environment plant growth chambers have been used to collect data for constructing process-level crop models since the mid-1970s (McKinion, 1986; Phene et al., 1978). These soil–plant–atmosphere–research (SPAR) chambers are used to control all external independent variables except solar radiation.

Methodology has been developed with these SPAR chambers to run appropriate experiments and collect process data to parameterize process-level crop growth models that require only 3 years of experiments. Such models have all primary and secondary plant processes described and calibrated. Some tertiary effects may also be addressed, but additional controlled-environment experiments are required to address crop-specific processes to complete the model.

MODEL-BUILDING METHODOLOGY

The methodology we have developed for calculating photosynthesis in cotton, wheat, and soybean is based on plant canopy geometry. The experimental method for acquiring these data involves using SPAR chambers in which these crops have been grown in a row-crop configuration (McKinion, 1986; Parson et al., 1980; Phene et al., 1978). Solar radiation is allowed to vary naturally, but air temperature, CO₂ concentration, and soil conditions are controlled. Thus, we measure apparent photosynthesis under various air temperature, soil moisture, and soil N regimes (McKinion and Baker, 1982; Reddy et al., 1992). We also measure light, maintenance, and growth respiration rates so that, given crop biomass, we can calculate net photosynthesis. Statistically, the equations derived from this experimental approach and database have an R^2 of 0.92 or better for cotton and wheat.

The methodology we use for building dynamic simulation models of field crops has led us to propose a minimum set of design features:

- 1) The model will be a materials balance accounting for dry-matter production and organ senescence.
- 2) The model will calculate photosynthate production separately from respiration (since the forcing functions in the two processes are different and the ratios change).

A contribution of the U.S. Dept. of Agriculture–Agricultural Research Service Crop Simulation Research Unit in cooperation with the Agronomy Dept., Mississippi State Univ., and the Mississippi Agricultural and Forestry Experiment Station. The cost of publishing this paper was defrayed in part by the payment of page charges. Under postal regulations, this paper therefore must be hereby marked *advertisement* solely to indicate this fact.

- 3) Photosynthesis and transpiration will be based on canopy light interception.
- 4) Daily canopy light interception will be calculated on a plant or row element and on canopy density rather than on a leaf area index, *per se*.
- 5) Nutrient uptake will include the process of nutrient entrainment in the transpiration stream.
- 6) Physiological stress will be calculated in a function containing metabolite and nutrient supply and demand.
- 7) Decisions about the abortion of fruiting structures will be based on physiological stress.
- 8) Morphogenetic rates will be stated as temperature functions that are reduced by physiological stress.
- 9) Water stress will be considered an environmental resultant that determines physiological stress. Plant-water relations will determine the turgor levels of growing points and, thus, indirectly, the demand for nutrients and metabolites. Plant-water relations will also determine photosynthesis and transpiration rates. Thus, water stress will indirectly determine morphogenetic rates or the abortion of fruiting parts.
- 10) Plant-water relations will be determined partly by the relative location of roots and soil water in the soil profile. Root distribution will respond to soil penetration resistance and to O₂ and nutrient gradients. Soil water distribution will be considered a function of soil hydrological properties.

From the design features mentioned above used with the SPAR experimental facility and high-speed desktop computers, the methodology outlined in this paper for building causal, process-oriented, physiological crop models is based on a rational, scientific approach.

SYSTEMS ANALYSIS AND MODEL VALIDATION

By applying the principles outlined above and tenets of system analysis, one can arrive at a collection of many processes in which the individual process has been transformed from a nonlinear system into a linear, mathematically tractable system. The use of the Taylor series expansion of a nonlinear partial differential equation about a defined region is an example of transforming a system from nonlinear to linear. Many biological processes are nonlinear. The fact that we can describe many of these processes in terms of mechanistic models is good, but it is not enough to aggregate a collection of linearized mechanisms into a simulation model. Biological systems have many feedback paths between processes. The more closely we can identify and code these feedback paths into the simulation model, the more robust the model becomes (Swartzman and Kaluzny, 1987). At the same time, the simulation model becomes more nonlinear. Synergistic, agonistic, and n-way interactions are common. Individual parts of the system can be described in terms of mechanism and tested in terms of test data sets. The soil physics in GOSSYM (Baker et al., 1983) is an example. Currently there are >400 cotton-producing soils in our database. These soils are identified by soil series and name. The soil physical properties are stored in the database for each soil layer down to 2 m. The data for each layer are the soil water retention curve; bulk density; percent sand and clay content; and initial moisture, organic matter, ammonia, and nitrate content. The operation of the soil hydrology code using these data is tested for each new soil type before it is added to the database. However, the model is more than a linear combination or sum of its parts. This implies that another step beyond combining individually studied model components is needed. This additional step is whole-model calibration and validation.

The model should be calibrated with data sets that are independent from the data sets with which validation tests are made. In calibrating the GOSSYM (Boone et al., 1993) model, we typically use ≈25 calibration data sets. For validation we use ≈80 data sets, with another 100 data sets reserved for odd circumstances. Until now, the only way to deal with this problem has been iterative and heuristic.

Model calibration will continue to be necessary in the near future. It is therefore imperative to consider procedures that can automate the parameterization process. We have developed genetic algorithm (GA)

techniques that permit the implementation of self-correcting, self-maintaining simulation models.

MODEL CALIBRATION AND GENETIC ALGORITHMS

There are several reasons for parameterizing or recalibrating a model. Variable response under different environments and varietal idiosyncrasies are often noted as sources of variability, and much experimental research relies on the ability to control these factors (Fye et al., 1984; Reddy et al., 1985). System parameterization is a well-defined, straightforward procedure, given that a model whose processes have been validated exists (Curry and Feldman, 1987). However, biological systems are not perfectly controllable or observable, as defined by Schultz and Melsa (1967). Curry and Feldman (1987) noted that even a good model must be recalibrated for different situations, but there are no systems with an identified mechanism by which this recalibration can be automated. This constant recalibration requirement is due to missing aspects in the model that fail to capture the differential response across all environments or situations. Recalibration is a necessary trade-off between the expense of further model refinement and an iterative correcting procedure. Further, Shannon (1976) noted that, due to the accumulation of measurement errors, adding new variables or processes will not necessarily increase prediction ability. Vansteenkiste and Spriet (1982) summarize some of the special constraints presented by biological systems and the resulting need for parameterization.

GAs are search procedures based on the biological mechanisms of selection and the transmittal and evolution of genetic information. They were developed by J. Holland (1992) and constitute one of the fastest growing areas of emergent computation research. GAs use biological evolution metaphors to find optimal, naturally evolved solutions in artificial systems. Software that can learn and adapt to new conditions is a revolutionary development in modeling and in the broader area of software engineering and hardware design. Heretofore, static computer systems have required constant maintenance to respond to changes in the problem domain. The emulation by computer programs of evolutionary processes that lead to increased fitness will enable them to be resilient to changes in their targeted environments (Goldberg, 1989). A methodology has been established for implementing GAs that is problem-independent (Davis, 1991; Goldberg, 1989; Goldberg and Richardson, 1987). GAs use strings (chromosome analogs) to represent the structural elements whose overall performance is of interest. A string is constructed in a manner analogous to DNA encodings in cellular nuclei. Strings are composed of genes, each of which describes an attribute of the desired solution. A GA works by reproducing chromosomes iteratively over multiple generations while applying genetic operators, which modify the informational content of the strings. Strings reproduce at a rate that is proportional to the fitness of the solution they encode. Thus, a population of strings (each of which represents a solution) becomes gradually more fit (i.e., the average solution gets more optimal). Generally, the difference in the average fitness between consecutive generations decreases and the population tends toward high fitness. It is said the population converges toward a solution at this time. Because of string or chromosome reproduction, their rearrangement, and subsequent evaluation and selection, a GA makes it possible to explore many potential solutions. The ability of a GA to search a massive state-space of alternatives simultaneously to a defined problem is greater than for other search algorithms (Goldberg and Richardson, 1987; Liepins et al., 1987).

We developed a GA (GA_{GOSSYM}) to automate the parameterization of GOSSYM (McKinion et al., 1989; Whisler et al., 1986). Data first used by Reddy et al. (1985) in the GOSSYM validation were used to develop our GA. A better overall fit was observed when compared to manual or heuristic parameterization (Sequeira et al., 1993). Phenological event timing, organ generation, and mass accretion were predicted better by GA_{GOSSYM} than by the model calibrated by Reddy et al. (1985). Notably, about a 20% improvement in the prediction of the number of green bolls (as measured by the sum of the absolute deviations) was obtained. Similarly improved predictions of peak leaf weight timing and, of particular interest to producers, a 40% improve-

ment in the prediction of overall fruit mass accretion were obtained by GA_{Gossym}. Running GA_{Gossym} with the full parameter set resulted in an overall fitness improvement of ≈25% over all output types normally used to assess fit.

Our discovery of the potential of GAs to derive improved parameter sets from high-level data will have profound implications on the speed at which we can develop locality-specific systems and on the accuracy of these systems to perform at that specific site. The GA was restricted to investigate parameter ranges that were within the realm of the field-observed variability. That is, parameters continued to maintain the original biological interpretations. We have demonstrated that we can develop a fine-tuned system that responds to local environments optimally given logical constraints on parameter ranges and functional forms.

GOSSYM is currently used throughout U.S. cotton-growing regions. A self-correcting feature was developed that allowed the model to rederive its parameters when field data indicated that forecasts were beyond an arbitrarily set threshold of acceptability. The process of finding the optimal configuration of parameter sets has been completely automated. The process of model correction thus depends only on the model's perceived goodness of fit. If it does not fit data well, it will invoke its associated GA and evolve a better set of parameters. This research implies that models may be applicable and have improved forecast efficacy over a wider range of conditions than is presently the case. This improvement in performance and applicability will affect natural resource management.

CONCLUSIONS

The level of detail used in a biological simulation model depends on its objectives and the accuracy or fidelity with which processes are simulated. Hierarchy theory states that if one is simulating at the whole-plant level, then details must be included at least two levels of abstraction down from the whole-plant level. It has been our experience that, as one increases the level of scientific mechanism in crop models, the system becomes more robust and less brittle. One must validate, validate, validate! Mechanisms that on the surface seem to be perfectly reasonable and plausible to put in a model may actually not function properly. The only way to find out is to integrate the mechanism, calibrate the model, and run validation tests. Verification tests at the component level do not always tell the story. New methodology is being developed to aid in model calibration. This new methodology uses GAs from the field of artificial intelligence. Our first efforts in this arena have given very promising results.

Literature Cited

Baker, D.N., J.R. Lambert, and J.M. McKinion. 1983. GOSSYM: A simulator of cotton growth and yield. South Carolina Agr. Expt. Sta. Tech. Bul. 1089.

- Boone, M.Y.L., D.O. Porter, and J.M. McKinion. 1993. Calibration of GOSSYM: Theory and practice. *Computers Electronics Agr.* 9:193-203.
- Curry, G.L. and R.M. Feldman. 1987. *Mathematical foundations of population dynamics*. Texas A&M Univ. Press, College Station.
- Davis, L. 1991. *Handbook of genetic algorithms*. Van Nostrand Reinhold, New York.
- Fye, R.E., V.R. Reddy, and D.N. Baker. 1984. Validation of GOSSYM: Part I. Arizona conditions. *Agr. Systems* 14:85-105.
- Goldberg, D.E. 1989. *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley, Reading, Mass.
- Goldberg, D.E. and J. Richardson. 1987. Genetic algorithm with sharing for multimodal function optimization, p. 41-49. In: J.J. Grefenstette (ed.). *Genetic algorithms and their applications*. Proc. 2nd Intl. Conf. Genet. Algorithms. L. Erlbaum Assoc. Publishers, Hillsdale, N.J.
- Holland, J.H. 1992. *Adaptation in natural and artificial systems*. MIT Press, Cambridge, Mass.
- Liepins, G.E., M.R. Hilliard, M. Palmer, and M. Morrow. 1987. Greedy genetics, p. 90-99. In: J.J. Grefenstette (ed.). *Genetic algorithms and their applications*. Proc. 2nd Intl. Conf. Genet. Algorithms. L. Erlbaum Assoc. Publishers, Hillsdale, N.J.
- McKinion, J.M. 1986. SPARNET: A data acquisition, analysis computer network. *Computers Electronics Agr.* 1:31-40.
- McKinion, J.M. and D.N. Baker. 1982. Modeling experimentation, verification, and validation: Closing the feedback loop. *Trans. Amer. Soc. Agr. Eng.* 25(3):647-653.
- McKinion, J.M., D.N. Baker, F.D. Whisler, and J.R. Lambert. 1989. Application of the GOSSYM/COMAX system to cotton crop management. *Agr. Systems* 31:55-65.
- Parsons, J.E., J.L. Dunlap, J.M. McKinion, C.J. Phene, and D.N. Baker. 1980. Microprocessor-based data acquisition and control software for plant growth chambers (SPAR system). *Trans. Amer. Soc. Agr. Eng.* 23(3):589-595.
- Phene, C.T., D.N. Baker, J.R. Lambert, and J.M. McKinion. 1978. SPAR—A soil-plant-atmosphere-research system. *Trans. Amer. Soc. Agr. Eng.* 21:924-930.
- Reddy, K.R., H.F. Hodges, and J.M. McKinion. 1992. Cotton response to global climate change. *World Resources Rev.* 4(3):348-365.
- Reddy, K.R., H.F. Hodges, and J.M. McKinion. 1993. A temperature model for cotton phenology. *Biotronics* 2:47-59.
- Reddy, V.R., D.N. Baker, and J.N. Jenkins. 1985. Validation of GOSSYM: Part II. Mississippi conditions. *Agr. Systems* 17:133-154.
- Schultz, D.G. and J.L. Melsa. 1967. *State functions and linear control systems*. McGraw-Hill, New York.
- Sequeira, R.A., R. Olson, and J. McKinion. 1993. Self-correction of simulation models using genetic algorithms. Intl. Joint Conf. Artificial Intelligence (IJCAI'93), AI Applications in Natural Resources, Chambéry, France. p. 77-87.
- Shannon, R. 1976. *Simulation: The art and the science*. Prentice Hall, New York.
- Swartzman, G.L. and S.P. Kaluzny. 1987. *Ecological simulation*. MacMillan, New York.
- Vansteenkiste, G.C. and J.A. Spriet. 1982. Modelling ill-defined systems. In: F.E. Cellier (ed.). 1982. *Progress in modelling and simulation*. Academic, New York.
- Whisler, F.D., B. Acock, D.N. Baker, R.E. Fye, H.F. Hodges, J.R. Lambert, H.E. Lemmon, J.M. McKinion, and V.R. Reddy. 1986. Crop simulation models in agronomic systems. *Adv. Agron.* 40:142-208.

Response of Crop Photosynthesis to Carbon Dioxide, Temperature, and Light: Experimentation and Modeling

V.R. Reddy, L.B. Pachepsky, and B. Acock

U.S. Department of Agriculture, Agricultural Research Service, Natural Resources Institute, Systems Research Laboratory, Building 007, Room 008, Beltsville Agricultural Research Center—West, Beltsville, MD 20705

Interest in the response of crop photosynthesis to atmospheric CO₂ concentration ([CO₂]), temperature, and light is growing in ecophysiological research and practical agriculture. Horticulturists are interested because greenhouses often have elevated CO₂. Agronomists and ecologists have to take into account in their crop development and productivity predictions that ambient [CO₂] is increasing, which is expected to increase global air temperature.

Canopy photosynthesis models have been proposed, but a method to validate, compare, and choose the most suitable model for a particular application is needed. Such a method is presented here and is used to show that a hyperbolic model is adequate for describing the effects of light, [CO₂], and air temperature on canopy photosynthesis.

The hyperbolic model was fitted to experimental data on cotton (*Gossypium hirsutum* L.) canopy CO₂ exchange obtained in daylight controlled-environment chambers in a range of [CO₂] and air temperatures. Canopy light-use efficiency depended on temperature, while canopy conductance to CO₂ transfer was a function of air temperature and [CO₂]. The complete model of cotton canopy photosynthesis

The cost of publishing this paper was defrayed in part by the payment of page charges. Under postal regulations, this paper therefore must be hereby marked advertisement solely to indicate this fact.