Agronomie 22 (2002) 115–117

EDITORIAL

The papers in this special issue were presented at a seminar on parameter estimation for crop models which was organized by the INRA modeling group and held in Toulouse in June, 2000. The motivation behind this seminar was very simple. The quality of crop model predictions clearly depends both on the structure of the model, that is the mathematical forms of the equations, and on the numerical values of the parameters. Research in agronomy has tended to emphasize model structure. Even though parameter estimation is a nearly universal problem for crop model developers and users alike, there has been little organized discussion of how to do it. No doubt one reason for this is that the subject is at the interface between agronomy and statistics. The problem of parameter estimation in non-linear models has been extensively studied in statistics, but the standard statistical methods cannot normally be applied to crop models. First of all, these models are simply too complex. Furthermore, the models are not simply empirical descriptions of the data. They are based on descriptions of the underlying processes, and there is information available about these processes which should be taken into account. Parameter estimation must thus take into account both statistical and agronomic considerations. The seminar had as its major objective to bring together different researchers who work on parameter estimation for crop models or on closely related topics, both agronomists and statisticians, in order to help identify the various facets of the problem and the approaches that have been proposed.

Before considering in detail the topics covered in the papers here, it is worthwhile to bring up two fundamental issues related to parameter estimation for crop models. First of all, what is the true meaning of the parameters? This obviously has importance for our attitude and our approach to parameter estimation. Certainly some of the parameters are physically measurable quantities, such as the extinction coefficient or the number of degree days between phenological stages. These parameters have a clear meaning without reference to any specific model. The meaning of prior information here is quite clear, it is based on measured values of these parameters. (However, one must keep in mind that the values of those parameters probably differ between different growth situations. Thus there is not a single “true” parameter value, but rather an average value for the growth situations of interest.) Other parameters only have meaning within the context of the model. For example, the effect of water stress on net photosynthesis depends on how water stress is calculated in the model. Hydrological model parameters depend on the grid resolution of the model. The notion of what is represented here by prior information is perhaps not too clear.

A second basic question is: what are the true objectives of parameter estimation? Several possibilities are or can be proposed. One is to recover the true parameter values, where the notion of “true” must be understood in the light of the previous paragraph. A second objective is to obtain accurate predictions. For example, the mean squared error of prediction could be the criterion in this case. Here, however, it is necessary to specify the range of growth conditions for which one wants to predict. The best method of prediction may depend on the growth conditions of interest. Wallach et al. illustrate this point. A third possible objective is related to decisions based on the model. It is thus not the predictive accuracy of the model that is of interest, but rather the value of the decisions calculated using the model. Makowski and Wallach specifically consider such an objective.

It is also of interest to observe that two different methodologies of investigation are employed in the papers here. In the first, an approach to parameter estimation is proposed and applied to a model, and the results compared to observed data. The advantage here is that one is testing crop models that are really in use and real data, thus one automatically takes into account all the complexities of such models and data. The disadvantage is that the data are rarely if ever sufficient to determine with precision how good the parameter estimation approach really is. In the second approach, a statistical simulation study is carried out. One assumes that the observations have some specific statistical distribution, and then one tests an approach to parameter estimation for a number of simulated data sets. Since the true parameter values underlying the data are known by assumption, one can calculate exactly the ability of different approaches to reproduce the correct parameter values. The disadvantage is that one assumes, and no doubt simplifies compared to reality, the relation between the model and the data. The two approaches are thus complementary, and progress will no doubt require a combination of the two. It should be noted that the simulation approach is very computer intensive, since it can require many thousands of model runs, and so in most cases the simulation approach is applied to simplified models. However, the advances in computing power will soon, if this is not yet the case, make statistical simulation studies on complex crop models feasible.

We turn now to a consideration of the different topics covered by the papers here. A first question is the extent to which parameter estimation is really necessary. Can the same parameter values be used over a very wide range of conditions? Gabrielle et al. investigate this basic question.

If one does decide to estimate parameter values, there is the problem that it is in general impossible to estimate all the parameters in a crop model based on crop field data. There are simply too many parameters, compared to the amount of field data typically available. Furthermore, it is well known that overparameterization, that is, estimating too many parameters compared to the available data, often leads to a model that gives good agreement with the data used for parameter estimation but poor predictions. One possible approach then is to select a relatively small number of parameters to estimate. Two papers here illustrate two different ways of selecting parameters to estimate. Ruget el al. propose a sensitivity analysis approach that can be used to identify parameters to estimate. Wallach et al. employ a model selection procedure to identify a small number of parameters to estimate. Estimating a small number of parameters has, however, the disadvantage that the estimated parameters often have unrealistic values, because in fact they are compensating for other parameters not estimated. This is a problem if the parameters are considered as having intrinsic meaning. A possible remedy would be to estimate not single parameters but rather linear combinations of parameters, though again only a small number of linear combinations. This approach is explored in Wallach et al.

The data used for parameter estimation often have a rather complex structure, for example several different variables measured at various dates for each growth situation. There is then the problem of how to take this structure into account. This is a classic problem of multivariate regression in statistics, but it is rarely explicitly taken into account in crop model parameter estimation. Wallach et al. propose a user-defined weighting of different measurements. Makowski and Wallach, using a static model but with multiple measurements per individual, suggest using a random parameter model to take the data structure into account, and furthermore demonstrate that using this approach has important practical consequences.

There is in general prior information about many of the parameter values, as discussed above. This information is used, for example, in fixing the values of the unchanged parameters in procedures where only some of the parameter values are estimated. In this case the use of the prior information is all or nothing. A parameter either keeps its initial (prior) value, or is estimated from the data. A Bayesian approach provides a smoother blending of prior information and data. The result of a Bayesian approach is a posterior distribution for the parameter values, which takes into account both the prior information and the field data. Durand et al. discuss this approach. Makowski et al. compare two different algorithms for calculating the posterior distribution in a simple case. Combal et al. also combine prior information with measurement data, but in a somewhat special setting. In their case, the objective is to estimate canopy characteristics such as leaf area index rather than model parameters, and the data are remote sensing data.

It is of course of importance to provide a best estimate for the set of parameters which will be used, for example, in prediction. However, it is also of importance to estimate the uncertainty in the estimated parameter values, because this translates into uncertainty in the model predictions which in turn obviously affects how much confidence one would feel about the predictions. One source of uncertainty concerns the values of those parameters that are not estimated. Wallach et al. explore this uncertainty in the case of estimation of a few selected parameters by redoing the parameter estimation for different values of the unchanged parameters. A different approach is to accept that various sets of parameter values can give nearly equivalent agreement with the data, and to seek to identify those sets. Parameter estimation for hydrological models has long taken this into account, the emphasis being on identifying acceptable sets of parameter values, as explained in Durand et al. The posterior distribution calculated in a Bayesian approach of course automatically provides a description of the uncertainty in the parameter values.

There is clearly an intimate relation between the available data and parameter estimation in crop models. It is of interest and importance to explore this relationship further. This leads to the question of how model quality is related to the data used for parameter estimation. Monod et al. go one step further, and explore the calculation of optimal experimental designs. The underlying model in this paper is non-linear but otherwise very simplified compared to crop models. Nonetheless, much of the approach is no doubt relevant to crop models.

The emphasis in the papers here is, by design, on parameter estimation. We must not, however forget, and a number of the papers insist on this point, that parameter estimation cannot be separated from the structure of the model. Parameter estimation and verification of the model equations are equally necessary, and are mutually dependent.

There is no pretension here to having covered the field of parameter estimation for crop models. However, a number of the important issues are raised, and at least some of the promising approaches are described. As the papers in this collection show, there is as yet no standard approach to parameter estimation in crop models. The problem is still quite open, and there is need for substantial progress. I hope that this special issue, by focusing attention on the problem, will play a role in encouraging this progress.

D. Wallach
Toulouse (France)