Agricultural Systems 103 (2010) 256–264

Contents lists available at ScienceDirect

Agricultural Systems

journal homepage: www.elsevier.com/locate/agsy

Influence of likelihood function choice for estimating crop model parameters using the generalized likelihood uncertainty estimation method

Jianqiang He, James W. Jones *, Wendy D. Graham, Michael D. Dukes

University of Florida, Agricultural and Biological Engineering, Museum Road, Gainesville, FL 32611, United States

ARTICLE INFO

Article history:
Received 6 January 2009
Received in revised form 14 December 2009
Accepted 27 January 2010
Available online 7 March 2010

Keywords:
Parameter estimation
GLUE
Likelihood function
CERES-Maize
DSSAT
Sweet corn

ABSTRACT

Proper estimation of model parameters is required for ensuring accurate model predictions and good model-based decisions. The generalized likelihood uncertainty estimation (GLUE) method is a Bayesian Monte Carlo parameter estimation technique that makes use of a likelihood function to measure the closeness-of-fit of modeled and observed data. Various likelihood functions and methods of combining likelihood values have been used in previous studies. This research was conducted to determine the effects of using previously reported likelihood functions in a GLUE procedure for estimating parameters in a widely-used crop simulation model. A factorial computer experiment was conducted with synthetic measurement data to compare four likelihood functions and three methods of combining likelihood values using the CERES-Maize model of the Decision Support System for Agrotechnology Transfer (DSSAT). The procedure used an arbitrarily-selected parameter set as the known "true parameter set" and the CERES-Maize model to generate true output values. Then synthetic observations of crop variables were randomly generated (four replicates) by using the simulated true output values (dry yield, anthesis date, maturity date, leaf nitrogen concentration, soil nitrate concentration, and soil moisture) and adding a random observation error based on the variances of corresponding field measurements. The environmental conditions were obtained from a sweet corn (*Zea* mays L.) experiment conducted in 2005 in northern Florida. Results showed that the method of combining likelihood values had a strong influence on parameter estimates. The combination method based on the product of the likelihoods associated with each set of observations reduced the uncertainties in posterior distributions of parameter estimates most significantly. It was also found that the likelihood function based on Gaussian probability density function was the best among those tested. This combination accurately estimated the true parameter values, suggesting that it can be used when estimating CERES-Maize model parameters for real experiments.

1. Introduction

Proper estimation of model parameters is required for ensuring accurate model predictions (Makowski et al., 2002). Modeling of complex environmental systems generally involves the indirect identification of model components or parameters by posing an inverse problem. Often, such inverse problems involve multiple parameters and observations that are only indirectly related to the parameters of interest, or which may be at different scales to the variables and parameters used in distributed predictions. There are many methods for estimating parameters using inverse modeling methods. Bayesian approaches can be used to estimate parameters using two types of information, a sample of data and prior information about parameter values. Results from a Bayesian method are probability distributions of parameter values and predicted outputs (Makowski et al., 2006a).

Bayesian methods are becoming increasingly popular for estimating parameters of complex mathematical models (Campbell et al., 1999). The Generalized Likelihood Uncertainty Analysis (GLUE) methodology (Beven and Binley, 1992), one such Bayesian method, allows information from different types of observations to be combined to estimate probability distributions of parameter values and model predictions (Lamb et al., 1998). Many parameter sets are generated from specified prior distributions of parameters and then used to simulate outputs by Monte Carlo simulation. The performance of each parameter set in predicting observed model states is evaluated via a likelihood measure that is used to weight the predictions from the different parameter sets. The GLUE method transforms the problem of searching for an optimum parameter set into a search for sets of parameter values that would give reliable simulations for a range of model inputs (Candela et al., 2005).

Parameters estimated using any inverse modeling approach are uncertain and subject to equifinality (e.g. Beven and Binley, 1992; Beven and Freer, 2001; Beven, 2006). Equifinality refers to the situation where the likelihood values are equal for two or more

* Corresponding author. Tel.: +1 352 392 1864; fax: +1 352 392 4092.
E-mail address: jimj@ufl.edu (J.W. Jones).

0308-521X/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.agsy.2010.01.006

parameter values, and one cannot select a best one from these two or more values. It can be argued on grounds of physical theory that there may be sufficient interactions among the components of a system that, unless the detailed characteristics of these components are specified independently, many representations may be equally acceptable. This is particularly true of those parameters to which the model is not sensitive in a particular environment. For this reason, a sensitivity analysis is needed to select parameters to which the model is sensitive for the range of experiments being used before attempting to estimate them using inverse modeling methods. One implication of equifinality is that the uncertainty associated with the use of models might be wider than is usually considered.

As with any model parameter estimation method, the GLUE method requires the definition of some measure of goodness-of-fit or likelihood. Beven and Binley (1992) pointed out that various likelihood measures might be appropriate in a given application. For example, Romanowicz et al. (1994, 1996) used a likelihood measure based on an autocorrelated Gaussian error model; Beven and Binley (1992) used a likelihood measure based on inverse error variance with a shaping factor *N*; Freer et al. (1996) suggested a likelihood measure based on Nash and Sutcliffe efficiency criterion; and Keesman and Van Straten (1989, 1990) used a likelihood measure based on scaled maximum absolute residuals. However, Stedinger et al. (2008) criticized the use of an arbitrary likelihood function. The choice of a likelihood function is critical and needs to be a reasonable description of the distribution of model errors for the statistical inference and resulting uncertainty and prediction intervals to be valid. If an arbitrary likelihood measure is adopted that does not reasonably reflect the distribution of model errors, then GLUE may generate arbitrary results without statistical validity that should not be used in scientific work.

With multiple observations and multiple types of observations, likelihood values for each observation must be combined into an overall value for each candidate parameter set (Beven and Binley, 1992). Available methods of combining likelihood values include multiplication (e.g. Beven and Binley, 1992), weighted addition (Zak et al., 1997), pseudomaximum likelihood measure (Van Straten, 1983), fuzzy union, fuzzy interaction, and weighted fuzzy combination (Aronica et al., 1998). Beven and Freer (2001) and Beven and Binley (1992) suggested that when a likelihood approach is being considered, the choice of method of combining likelihood values is subjective. However, when the GLUE method is used with a model for the first time, it is important to make sure that the choice of likelihood measure and combination method can produce reliable model parameters.

The CERES-Maize model (Jones and Kiniry, 1986; Ritchie, 1998; Hoogenboom et al., 2003) is a maize (*Zea mays* L.) crop growth model in the cropping system model (CSM) that is in the Decision Support System for Agrotechnology Transfer (DSSAT) (Jones et al., 2003; Tsuji et al., 1998). The DSSAT-CSM incorporates all crops as modules using a single soil model. Hereafter, CERES-Maize will be used to refer to the model used in this study. This model has many parameters that characterize crop and soil processes, a number of which usually need to be estimated using field experiments. Over the years, a number of methods have been used to estimate parameters for the DSSAT models, including the simplex method (Grimm et al., 1993), simulated annealing (Mavromatis et al., 2002), sequential search software (Hunt et al., 1993), and even visual methods. Each of these methods has its own advantages and limitations. Our main reasons for selecting the GLUE method for this study were that it can help us understand uncertainties in the parameters and how those uncertainties affect predictions and it is relatively simple and straightforward to implement.

In using this method for the first time with the CERES-Maize model, the question arises as to how much the different likelihood measures and combination methods influence the results of parameter estimation. The objective of the study was to answer this question for this widely-used crop model. We evaluated the influence of four different likelihood functions and three combination methods in GLUE on the parameter estimates for the CERES-Maize model.

2. Materials and methods

2.1. CERES-Maize model

Crop growth and development are simulated by the CERES-Maize model in DSSAT V4.0 (Hoogenboom et al., 2003) with a daily time step from planting to maturity using physiological process relationships that describe the responses of maize to soil and environmental conditions. Potential growth is dependent on photosynthetically active radiation and its interception, whereas actual biomass production on any day is constrained by suboptimal temperatures, soil water deficits, and nitrogen deficiencies (Ritchie and Godwin, 1989; Ritchie, 1998).

There are four types of input data to the model: weather, plant, soil, and management. The weather input data are daily sum of global radiation (MJ m^{-2}), daily minimum and maximum air temperatures (°C), and daily sum of precipitation (mm). Plant parameters and physiological characteristics are given in the form of genetic coefficients, which describe physiological processes such as development, photosynthesis, and growth for individual crop varieties in response to soil, weather, and management during a season. Soil inputs describe the physical, chemical, and morphological properties of the soil surface and each soil layer within the root zone. The management information includes planting density, row spacing, planting depth, irrigation, application of fertilizer, etc. (Ritchie, 1998).

2.2. Soil parameters and genetic coefficients

There are usually many parameters and inputs in complex crop simulation models. Each of these parameters and inputs is subject to errors. Ideally, one would directly measure all inputs and parameters, but this is not possible in many cases (Bechini et al., 2006). Furthermore, uncertainties in some parameters are likely to cause more variations in simulated results than others. Thus, a common strategy is to select a subset of parameters to estimate using sensitivity analysis, and fixing the others to their nominal values (Makowski et al., 2006a,b; Monod et al., 2006; Wallach et al., 2001). Through a global sensitivity analysis with one-at-a-time (OAT) method (Morris, 1991), He (2008) selected the most sensitive genetic and soil parameters (Table 1) relative to their influence on CERES-Maize model predictions of dry matter yield and cumulative nitrogen leaching for the growing conditions in this study. Other parameters or inputs may be important for different environmental and management conditions, and the sensitivity analysis would need to be repeated for other experiments.

The selected soil parameters (SLLL, SDUL, and SSAT) define soil water holding capacity and influence the amount of available water in the soil profile on a day to day basis. Parameters SLRO and SLDR influence the amount of soil water runoff and water drained from the soil profile. Parameter SLPF represents the effect of other limiting soil factors that reduce crop growth. Genetic coefficients P1 and P5 control the phenological development of the crop through their effects on anthesis and maturity dates. Coefficient PHINT influences both phenological development and yield. See Jones and Kiniry (1986) and Ritchie (1998) for more details regarding these parameters in the CERES-Maize model.