European Journal of Agronomy 94 (2018) 25–35

Contents lists available at ScienceDirect

European Journal of Agronomy

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

Towards improved calibration of crop models – Where are we now and where should we go?

S.J. Seidel[a,*], T. Palosuo[b], P. Thorburn[c], D. Wallach[d]

[a] Institute of Crop Science and Resource Conservation, University of Bonn, Katzenburgweg 5, D-53115 Bonn, Germany
[b] Natural Resources Institute Finland (Luke), 00790 Helsinki, Finland
[c] CSIRO Agriculture and Food, Brisbane, Australia
[d] INRA, UMR AGIR, Castanet Tolosan, France

ARTICLE INFO

Keywords:
Crop modelling
Model calibration
Parameter estimation
Survey
Simulation modelling
Uncertainty

ABSTRACT

Crop simulation models are increasingly used in agricultural decision making. Calibration is a demanding and critical step in developing and applying a model. Despite its importance little attention has been paid to documenting and analysing current calibration practices. This study reports the results from 211 responses to a web-based survey of calibration practices. The survey questions covered multiple choices that are required when doing a calibration exercise. Concerning data, most respondents used field data, but regional data and a combination of field and regional data were also used. Almost all respondents used multiple data types, the most common being phenology and yield data. The median number of estimated parameters was 6, and often this number was only slightly smaller than the number of environments that provided the data. Most respondents fit the data in multiple stages, starting in most cases with phenology data. Many respondents searched for parameter values that minimized a sum of squared errors, but substantial groups used an ad hoc measure of goodness-of-fit, the GLUE method, a weighted least squares method or a Bayesian approach. Nearly half the respondents simply used trial-and-error to search for the best-fit parameters. The other respondents were split more or less equally between those who used existing software and those who wrote new software. Slightly less than half the respondents obtained information on parameter uncertainty. Model evaluation was based on goodness-of-fit or data splitting or cross validation. The median time devoted to crop model calibration was 25 days. Based on these results, a list of topics that should be covered in guidelines for calibration is suggested.

1. Introduction

Nowadays, crop models are widely used in agronomic research and are proposed as possible tools for multiple applications including crop management, advice to policy makers, an aid in crop breeding or climate change impact assessment and adaptation to climate change (Ewert et al., 2015; Rosenzweig et al., 2013; Rötter et al., 2015). Recently several studies have involved multi-model ensembles of crop models (Asseng et al., 2013; Bassu et al., 2014; Palosuo et al., 2011). These studies have all shown that there is a wide variability between crop model predictions and even among the model predictions (see e.g. Martre et al. (2015)) by the same model when used by different users (Confalonieri et al., 2016). Clearly there is major uncertainty in crop model predictions and it is of fundamental importance to improve these models and how they are used in order to allow practical applications.

Uncertainty in crop model simulation can arise from three sources: (1) model structure, (2) input data and (3) model parameters (van Oijen and Ewert, 1999). (Note that Walker et al. (2003), in his list of possible locations of uncertainty, includes the above three sources plus context uncertainty, related to choice of boundaries of the modelled system, and outcome uncertainty, which is the accumulated uncertainty in model predictions and which results from the other sources.) Crop model improvement can target any of these, alone or in combination. Model structural uncertainty is partly due to the fact that models do not include all relevant explanatory variables. No model can include exhaustively all of the explanatory variables that determine crop growth and development. Various studies have proposed adding additional explanatory variables to existing crop models in order to capture specific sources of variability, for example row spacing (Flenet et al., 1996), CO_2 concentrations (Kersebaum et al., 2009; Tubiello et al., 2000) or ground level ozone concentrations (Cappelli et al., 2016). Structural uncertainty also arises from the fact that the model equations, which describe the effects of the explanatory variables on the output variables, may not have the correct or best functional form.

* Corresponding author.
E-mail addresses: taru.palosuo@luke.fi (T. Palosuo), Peter.Thorburn@csiro.au (P. Thorburn), Daniel.Wallach@inra.fr (D. Wallach).

https://doi.org/10.1016/j.eja.2018.01.006
Received 1 August 2017; Received in revised form 19 December 2017; Accepted 9 January 2018
1161-0301/

There have been many comparisons of different possible functional forms, for example development rate as a function of temperature (Alderman and Stanfill, 2017; Kumudini et al., 2014) or photosynthesis (Seidel et al., 2016).

Crop model inputs are affected by measurement or sampling errors of observed variables, limitations in data availability and high spatial and temporal variability of, e.g. weather variables. There have been several studies describing how to characterize data and discussing data requirements for crop model calibration (Boote et al., 2016; Kersebaum et al., 2015).

Finally, the values of the parameters in a model may not be those that minimize prediction error. Partly this is again a data problem; measurement errors and insufficient data both degrade parameter estimation. In addition, the calibration technique may not be capable of finding the best parameter values given the data. There have been studies in the literature proposing or comparing calibration methods for crop models (He et al., 2010; Makowski et al., 2002; Wallach et al., 2014). The authors show that multiple calibration approaches are possible, and that the choice has a significant effect.

The three sources of error lead to very different approaches to crop model improvement. The first source of error is related to the behavior of the system. Reducing this source of error is mainly through experimentation targeted to obtaining new information about specific processes or conditions, followed by mathematically describing the experimental results. Reducing the second source of error involves obtaining more, and better, data. Reducing the third source of error on the other hand is mainly about making the best possible use of existing data. It is this last pathway to model improvement, namely improved model calibration, that is targeted here.

Calibration, or parameter estimation, has been very extensively studied in statistics. However, it is often difficult to apply standard statistical methods and software to crop models. Reasons for this include the fact that these models have complex mathematical structure generally involving non-linearities and multiple discontinuities, that these models have multiple outputs which lead to complex correlation structures of errors, and that the software for these models is often difficult to couple with existing calibration software (Wallach et al., 2014). The result is that there is no standard approach to crop model calibration. Different approaches have been taken including least squares (Ramirez-Villegas et al., 2017; Zeng et al., 2016), Bayesian parameter estimation (Wallach et al., 2012), generalized likelihood uncertainty estimation (GLUE) (Chisanga et al., 2015) and various ad hoc, often trial and error, approaches (Li et al., 2015).

In addition, the details of how a method is applied can have important repercussions on the estimated parameters. For example, in a study of the GLUE method, He et al. (2010) found that the choice of likelihood function and method of combining likelihood values had a strong influence on parameter estimation results for the CERES-Maize model. In a comparison of GLUE and Markov Chain Monte Carlo (MCMC) parameter estimation, Makowski et al. (2002) found that both approaches are sensitive to the prior assumptions made about parameter values. The choice of data to use for calibration, and the way that data is used, for example in multiple calibration steps, also strongly affects the results of calibration (Angulo et al., 2013; Guillaume et al., 2011). Strategies for calibration of several individual models have been proposed (Ahuja and Ma, 2011a). Often the suggested strategies focus on a minimum data set for calibration, and the specific parameters to adjust to fit each output variable (for example Ma et al. (2011)). In the conclusion to a book with chapters on calibration for several different models, Ahuja and Ma (2011b) state "… due to the complexity of system models, there has not been a standard method to parameterize a system model. Methods reported in this book and elsewhere are often model and user dependent." A highlighted message is "The development of a systematic and hopefully common protocol is needed to help users".

Recommendations for crop model calibration generally are proposed by model developers, as in Ahuja and Ma (2011a). It is not clear to what extent these methods are applied by modelers. There is in fact very little information on the actual calibration practices of modelers. What is the diversity of calibration situations that are encountered (models, data, objectives), what are the choices faced during crop model calibration, to what extent is there a consensus in the community of crop modeling practitioners on calibration, or is there a wide diversity of practices? It is important to answer these questions, as a first step towards improving crop model calibration practices. Indeed, recommendations for improved calibration must take into account current practices, and try to understand the reasons behind those practices.

The overall goal of this study was to obtain a detailed picture of current practices of crop model calibration, from a broad cross section of practitioners who have been involved in calibration activities. Information on current practices was obtained by conducting a web-survey among a large group of researchers involved in crop modeling. A survey was preferred to a literature search, because information on calibration is often described very succinctly, if at all, in publications, whereas the survey format allowed us to examine multiple aspects of calibration in detail. All together 211 usable questionnaires were returned and analyzed. The survey results were used to translate the information on current practices into an outline for guidelines for crop model calibration. Based on the survey we identified a number of questions that arise in practice, and the range of responses of modelers. This was the basis for identifying the problems that need to be addressed in future guidelines.

2. Materials and methods

2.1. Participation

The survey was targeted widely for the global crop modelling community. The request to participate defined the target audience as anybody who has "been responsible for a crop model calibration activity" (quotation from the survey questionnaire). To avoid multiple responses by the same researcher, each respondent was requested to fill out the survey only once, "for just one specific study (one data set, one model, one set of calibrated parameters), the one that you feel could be most useful to others."

Information about the survey, the request to participate and the link to the web-survey (Fig. 1 (top) shows the first page) were circulated via email and sent to the mailing lists of the Agricultural Model Intercomparison and Improvement project (AgMIP), the MACSUR project (Modelling European Agriculture with Climate Change for Food Security; a knowledge-hub consisting of agricultural modellers from 18 European countries) and the International Soil Modeling Consortium (ISMC) and to the mailing lists of the models APSIM, DAISY, DSSAT, and STICS. It was also posted on the homepages of several of those groups. It was clearly stated that the survey was open to anyone who wanted to participate. The survey was open altogether for about six weeks, from October 23 to November 30, 2016.

Overall 318 responses were received. Many of those had responses to only a few questions, and thus contributed few data to the survey. It was decided to disregard responses that answered none of four questions considered to be of fundamental importance, namely Q16 concerning the number of estimated parameters, Q20 concerning the number of stages of calibration, Q23 concerning the approach to calibration and Q34 concerning the major difficulty in calibration (the full text for each question is given in the Appendix). All the results in this paper refer to the remaining 211 submissions that answered at least one (in 144 cases all) of the fundamental questions.

2.2. The survey

The survey was divided into 11 sections, each concerned with a