Environmental Modelling & Software 81 (2016) 165–173

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Uncertainty in crop model predictions: What is the role of users?

Roberto Confalonieri [a,*], Francesca Orlando [b], Livia Paleari [b], Tommaso Stella [b], Carlo Gilardelli [b], Ermes Movedi [b], Valentina Pagani [b], Giovanni Cappelli [b], Andrea Vertemara [c], Luigi Alberti [c], Paolo Alberti [c], Samuel Atanassiu [c], Matteo Bonaiti [c], Giovanni Cappelletti [c], Matteo Ceruti [c], Andrea Confalonieri [c], Gabriele Corgatelli [c], Paolo Corti [c], Michele Dell'Oro [c], Alessandro Ghidoni [c], Angelo Lamarta [c], Alberto Maghini [c], Martino Mambretti [c], Agnese Manchia [c], Gianluca Massoni [c], Pierangelo Mutti [c], Stefano Pariani [c], Davide Pasini [c], Andrea Pesenti [c], Giovanni Pizzamiglio [c], Adriano Ravasio [c], Alessandro Rea [c], David Santorsola [c], Giulia Serafini [c], Marco Slavazza [c], Marco Acutis [b]

[a] *University of Milan, DEMM, Cassandra Lab, Italy*
[b] *University of Milan, DiSAA, Cassandra Lab, Italy*
[c] *University of Milan, Cropping Systems MS Course, Italy*

A R T I C L E I N F O

Article history:
Received 13 June 2015
Received in revised form
6 April 2016
Accepted 7 April 2016
Available online 14 April 2016

Keywords:
Calibration
Maize
Model ensemble
Parameter uncertainty
Rapeseed
Uncertainty in predictions

A B S T R A C T

Crop models are used to estimate crop productivity under future climate projections, and modellers manage uncertainty by considering different scenarios and GCMs, using a range of crop simulators. Five crop models and 20 users were arranged in a randomized block design with four replicates. Parameters for maize (well studied by modellers) and rapeseed (almost ignored) were calibrated. While all models were accurate for maize (RRMSE from 16.5% to 25.9%), they were, to some extent, unsuitable for rapeseed. Although differences between biomass simulated by the models were generally significant for rapeseed, they were significant only in 30% of the cases for maize. This could suggest that in case of models well suited to a crop, user subjectivity (which explained 14% of total variance in maize outputs) can hide differences in model algorithms and, consequently, the uncertainty due to parameterization should be better investigated.

1. Introduction

The increasing use of crop models for estimating the impact of climate change on crop productions (Tubiello et al., 2007) is pushing modellers towards the development of strategies to properly manage the uncertainty associated with model estimates (e.g., Martre et al., 2014; Li et al., 2015). Indeed, uncertainty in crop model predictions is, on the one hand, unavoidable because these mathematical analogies do not reproduce plant processes at the hierarchical level they actually originate in (i.e., cells, organelles; Acock and Acock, 1991). On the other hand, part of the uncertainty is not directly ascribable to the models themselves, but to the input data used to drive the models and to the way the models are parameterized (e.g., Wallach, 2011; Rötter et al., 2012; Confalonieri et al., 2016).

The international modelling community has recently launched collaborative initiatives to analyse and manage uncertainty due to different approaches used to formalize knowledge, such as the Agricultural Model Intercomparison and Improvement Project (AgMIP; Rosenzweig et al., 2013) and the Modelling European Agriculture with Climate Change for Food Security (MACSUR). These initiatives propose effective techniques aimed at managing uncertainty in predictions through the evaluation and use of ensembles of models (e.g., Bassu et al., 2014; Li et al., 2015).

However, when various models are evaluated using the same datasets, scientists implicitly consider models as autonomous entities. Statements such as “model A overestimated biomass

* Corresponding author.
E-mail address: roberto.confalonieri@unimi.it (R. Confalonieri).

http://dx.doi.org/10.1016/j.envsoft.2016.04.009
1364-8152/© 2016 Elsevier Ltd. All rights reserved.

accumulation during early stages" are frequently used to discuss model performance, whereas the uncertainty due to parameterization, which includes the impact of model users (Janssen and Heuberger, 1995), would suggest to better underline that we are at least in part evaluating and comparing the 'model + user', rather than solely the model. As a result, it might be more appropriate to state "model A parameterized by user B overestimated …" Of course, it is surely a matter of wording because everybody is aware that a model's predictions are affected by a *user effect*. The point is what is the weight of this effect?

Diekkrüger et al. (1995) used different models for soil water dynamics to demonstrate how two users can obtain significantly different results using the same computer code (thus, with the same model algorithms) and that simple models, when used by experts, can be calibrated to be more accurate than complex ones, at least in a curve-fitting sense. Other authors, conscious of the dependence of the calibration process on users' expertise and subjectivity, underlined the need for automatic calibration algorithms and well-structured and systematic calibration approaches (Janssen and Heuberger, 1995). However, these issues have been addressed mainly for hydrological models (e.g., Eckhardt et al., 2005), and few studies on the application of such techniques to crop models are available (e.g., Acutis and Confalonieri, 2006; Dumont et al., 2014). This is partly due to the wide range of approaches used for crop simulation, the presence of parameter sets that differ widely among models and to the model sensitivity to the parameters, which is known to markedly change across different environmental and management conditions (Confalonieri et al., 2012). Another reason is the uncertainty in the observations used during calibration (Confalonieri et al., 2016) and to the presence of inconsistencies in data patterns over time, which can be better interpreted by expert users rather than by optimization algorithms. Consequently, most calibrations of crop model parameters are still based on trial and error. Moreover, as discussed by a number of authors, most studies on model intercomparison do not provide details on the methods used for calibration (Porter et al., 1993; Rötter et al., 2012; Rosenzweig et al., 2013).

This study was aimed at (i) developing a procedure for testing the significance of the differences in the predictions from different models in light of the uncertainty due to different users; (ii) analysing model and user contributions to the uncertainty in the outputs from five widely used crop models for maize and rapeseed; (iii) encouraging scientists to consider more explicitly – within climate change studies – the uncertainty due to users, as is frequently done when addressing emission scenarios, general circulation models and crop simulators.

2. Materials and methods

2.1. The crop models

The five models used in this study are AquaCrop, DSSAT, CropSyst, STICS and WOFOST, which are used within a variety of international research contexts (e.g., Bassu et al., 2014; Martre et al., 2014). These models represent different ways to formalize knowledge, which reflected in their complexity and in the level of detail used to reproduce biophysical processes. DSSAT (Jones et al., 2003), CropSyst (Stöckle et al., 2003) and STICS (Brisson et al., 2002) were developed to study the effect of climate, soil and management on cropping systems. They are based on the concept of net photosynthesis and are suitable to reproduce the impact of a variety of agromanagement practices. WOFOST (van Keulen and Wolf, 1986) owes its broad use to the soundness of the approaches used for crop growth (e.g., gross photosynthesis, growth and maintenance respiration are explicitly simulated) and to the high level of detail in reproducing the interaction between plants and weather drivers. AquaCrop (Steduto et al., 2009) was recently proposed by the FAO to provide users with a simple but effective tool, and it is likely the most peculiar in the way processes are modelled (canopy ground cover is used instead of leaf area index, and biomass accumulation is based only on water productivity). STICS and CropSyst have algorithms for the reproduction of different categories of species, whereas DSSAT is actually a suite of models and sub-models that are used according to the specific crop/cropping system. The CERES-Maize and CROPGRO models contained within DSSAT were used for maize and rapeseed, respectively. Table 1 summarizes – in a comparative fashion – the different approaches implemented in the five models for key processes involved in crop growth and development.

2.2. The simulation experiment

The five models and 20 students of the Cropping Systems MS Course of the University of Milan were arranged in a randomized block design with four replicates, where model was the factor, levels were the five models, and blocks were four 5-student groups. The four students in charge of each of the five models were in different blocks (groups) and they were considered as performing independent calibration replicates, since they were recommended to avoid discussion with others working on the same model in other blocks.

All the students had completed a crop modelling course that focused on approaches to reproduce crop growth and development, as well as soil water and nitrogen dynamics in the plant-soil system. The four students that were placed in charge of using each of the five models received in-depth training on the algorithms implemented in that model, as well as on model usage and calibration techniques (see the next section for further details on how the calibrations were performed). The simulation experiment was carried out for two crops: maize represents a crop that is well studied by the modelling community, and rapeseed represents a crop that has largely been ignored by the crop modelling community. For each crop, data from different sites and years were used after randomly splitting the data into calibration and validation datasets (Table 2).

US (exps. 1 to 5) and Italian (exps. 11 to 13) maize data were collected under conditions where ET0 markedly exceeded rainfall during the growing season, with mean values for a normalized aridity indicator (SAM, unitless, −1 to +1; (rainfall − ET0)/(rainfall + ET0) during the growing season; negative values indicate aridity; Confalonieri et al., 2010) equal to −0.39 and −0.35, respectively. Conditions in France (exps. 6 to 10) were less continental, with milder thermal conditions during summer and a mean SAM value equal to −0.18. For rapeseed (Exps. 14 to 19), conditions at the experimental sites ranged from warm and moderately arid (Noto 2011/12; SAM = −0.23) to mild (Buscate 2012/13; SAM = 0.35).

For maize, medium- and medium long-cycle hybrids were grown in the experiments carried out in the US and Italy, respectively, whereas medium long-cycle cultivars were grown in the French experiments. For exps. 11 to 13, which were conducted in northern Italy in 2013 under drip-irrigated treatments with three replicates, the hybrid DKC6815 (medium long-cycle) was sown on April 25 in Anzola (exp. 11), and on April 12 in Gonzaga and Virgilio (exps. 12 and 13). According to soil analyses, different amounts of N were distributed in the three experiments: 65 kg N ha^{-1} in pre-sowing (19-9-27) and 184 kg N ha^{-1} top-dressed (urea) for exp. 11, 55 kg N ha^{-1} in pre-sowing (18–46) and 207 kg N ha^{-1} top-dressed (urea) for exp. 12, and 30 kg ha^{-1} in pre-sowing (14-6-6) and 160 kg N ha^{-1} distributed with the drip system (fertigation; 26-