Environmental Modelling & Software 84 (2016) 529–539

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Estimating model prediction error: Should you treat predictions as fixed or random?

Daniel Wallach [a,*], Peter Thorburn [b], Senthold Asseng [c], Andrew J. Challinor [d,e], Frank Ewert [f], James W. Jones [c], Reimund Rotter [g], Alex Ruane [h]

[a] INRA, UMR 1248 Agrosystèmes et développement territorial (AGIR), 31326, Castanet-Tolosan Cedex, France
[b] CSIRO Agriculture Flagship, Dutton Park, QLD, 4102, Australia
[c] Agricultural & Biological Engineering Department, University of Florida, Gainesville, FL, 32611, USA
[d] Institute for Climate and Atmospheric Science, School of Earth and Environment, University of Leeds, Leeds, LS29JT, UK
[e] CGIAR-ESSP Program on Climate Change, Agriculture and Food Security, International Centre for Tropical Agriculture (CIAT), A.A. 6713, Cali, Colombia
[f] Institute of Crop Science and Resource Conservation INRES, University of Bonn, 53115, Germany
[g] Natural Resources Institute Finland (Luke), FI-50100, Mikkeli, Finland
[h] NASA Goddard Institute for Space Studies, New York, NY, 10025, USA

A R T I C L E I N F O

Article history:
Received 7 July 2015
Received in revised form
21 June 2016
Accepted 6 July 2016

Keywords:
Crop model
Uncertainty
Prediction error
Parameter uncertainty
Input uncertainty
Model structure uncertainty

A B S T R A C T

Crop models are important tools for impact assessment of climate change, as well as for exploring management options under current climate. It is essential to evaluate the uncertainty associated with predictions of these models. We compare two criteria of prediction error; $MSEP_{fixed}$, which evaluates mean squared error of prediction for a model with fixed structure, parameters and inputs, and $MSEP_{uncertain}(X)$, which evaluates mean squared error averaged over the distributions of model structure, inputs and parameters. Comparison of model outputs with data can be used to estimate the former. The latter has a squared bias term, which can be estimated using hindcasts, and a model variance term, which can be estimated from a simulation experiment. The separate contributions to $MSEP_{uncertain}(X)$ can be estimated using a random effects ANOVA. It is argued that $MSEP_{uncertain}(X)$ is the more informative uncertainty criterion, because it is specific to each prediction situation.

1. Introduction

Crop models are important tools in agriculture and environment, including applications in crop breeding and crop management (Boote et al., 2010). A recent major focus is in using crop models to evaluate the impact of climate change on crop production and other crop responses (Rosenzweig et al., 2013).

As for all models, it is essential to estimate the uncertainty in crop model predictions, i.e. the extent to which predicted values may differ from the true values. There is increasing recognition in the crop modeling community that more attention needs to be paid to uncertainty in the crop models (Rötter et al., 2011; Rosenzweig et al., 2013). Recently, studies have been done using both multiple climate models and multiple crop models, as a way of evaluating uncertainty arising from both types of model. Preliminary evidence indicates that in fact, the uncertainty due to the variation between crop models may be larger than that due to climate models (Asseng et al., 2013), which emphasizes the importance of estimating crop model prediction uncertainty (Koehler et al., 2013). Estimating uncertainty is of primary importance for all uses of crop models, for example for exploring crop management options under current climate (Baigorria et al., 2007). It is also of major importance for models in other fields, including climate modeling (Holzkämper et al., 2015; Tebaldi and Knutti, 2007), environmental studies (Uusitalo et al., 2015) or hydrologic modeling (Refsgaard et al., 2006; Renard et al., 2010).

Past crop model uncertainty studies can be grouped into three different approaches. The first is based on comparing model hindcasts to observed data (Fig. 1a). A common measure of discrepancy is mean squared error, but there are many other possible measures of discrepancy, and there have been several

* Corresponding author.
E-mail addresses: daniel.wallach@toulouse.inra.fr (D. Wallach), peter.thorburn@csiro.au (P. Thorburn), sasseng@ufl.edu (S. Asseng), a.j.challinor@leeds.ac.uk (A.J. Challinor), fewert@uni-bonn.de, eeyshire@uni-bonn.de (F. Ewert), jimj@ufl.edu (J.W. Jones), reimund.rotter@luke.fi (R. Rotter), alexander.c.ruane@nasa.gov (A. Ruane).

http://dx.doi.org/10.1016/j.envsoft.2016.07.010
1364-8152/

studies devoted to examining and comparing them (Bellocchi et al., 2010; Bennett et al., 2013; Yang et al., 2014; Wallach et al., 2014). The use of hindcasts is the standard method of evaluating crop models and there have been numerous studies of this type, aiming to evaluate various models for various applications (Basso et al., 2016; Coucheney et al., 2015). This is typically referred to by various terms, such as validation, verification and/or evaluation. The assumption is that the discrepancy between past observations and simulated values is an indication of the likely discrepancy in new predictions. That is, observed discrepancies are taken as a measure of uncertainty for predictions. There is no explicit treatment of the uncertainties in the model itself in this approach.

Short-term climate forecasts are often evaluated on the basis of skill scores, which compare some criterion of model fit with that of a naïve predictor (Murphy, 1988; Reichler and Kim, 2008). This is comparable to the approach above. However, a major difference with crop models is that in general there is much more data available for testing climate models, although with remote sensing this may become less true. One result is that one can look at performance of climate models as a function of the prediction situation (geographical area, lead time), while evaluation of crop models is generally limited to estimating a single, average quality of prediction.

In a second approach (Fig. 1b), the uncertainties in the model inputs or parameters are of primary concern. It is well understood that the values of the parameters in crop models are only approximations, and may have fairly large uncertainties (Dzotsi et al., 2013). Similarly, many of the input variables in crop models are difficult to estimate or measure and may have large uncertainties due to high spatial or temporal variability (Aggarwal, 1995; Bouman, 1994; Roux et al., 2014). This approach propagates the uncertainty in parameters and/or inputs through the crop model, in order to evaluate the resulting uncertainty in predictions. While these studies clearly evaluate an aspect of prediction uncertainty, the major objective is often elsewhere, namely to identify those factors (inputs or parameters) that contribute most to prediction uncertainty, using sensitivity analysis.

The third, more recent approach is based on multi-model ensembles (MMEs) (Fig. 1c). For many crops multiple different crop models have been developed by different research teams. Models might for example differ in the way primary production or soil water or development rate is modeled. Model structure uncertainty is a major source of uncertainty in predictions, not only for crop models (Wintle et al., 2003) but for mechanistic models in general

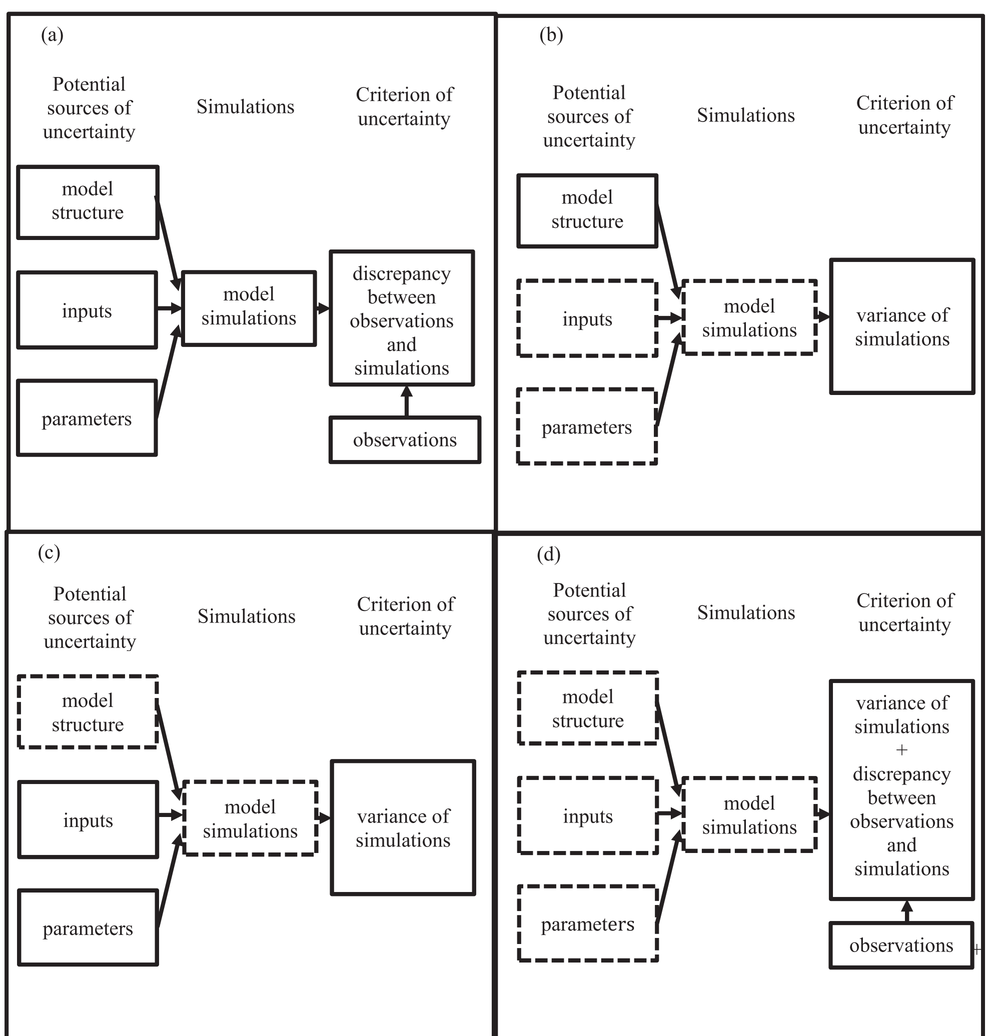

Fig. 1. Schematic diagrams of different approaches to estimation of prediction uncertainty. a) Based on comparison of hindcasts with observations. b. Based on propagating input and/or parameter uncertainty through the model. c. Based on multi-model ensemble studies. d. Based on simulations with multiple model structures, multiple input vectors and multiple parameter vectors for each model. Elements that are explicitly treated as random are within a dash enclosed box.