G Model
EURAGR-25465; No. of Pages 8
ARTICLE IN PRESS
Europ. J. Agronomy xxx (2015) xxx–xxx
Contents lists available at ScienceDirect
European Journal of Agronomy
journal homepage: www.elsevier.com/locate/eja

A simple Bayesian method for adjusting ensemble of crop model outputs to yield observations

David Makowski[a,*]

[a] UMR Agronomie, INRA, AgroParisTech, Université Paris-Saclay, 78850 Thiverval-Grignon, France

ARTICLE INFO

Article history:
Received 6 August 2015
Received in revised form
17 December 2015
Accepted 22 December 2015
Available online xxx

Keywords:
Bayesian method
Climate change
Ensemble modelling
Uncertainty
Yield

ABSTRACT

Multi-model forecasting has drawn some attention in crop science for evaluating effect of climate change on crop yields. The principle is to run several individual process-based crop models under several climate scenarios in order to generate ensembles of output values. This paper describes a simple Bayesian method – called Bayes linear method – for updating ensemble of crop model outputs using yield observations. The principle is to summarize the ensemble of crop model outputs by its mean and variance, and then to adjust these two quantities to yield observations in order to reduce uncertainty. The adjusted mean and variance combine two sources of information, i.e., the ensemble of crop model outputs and the observations. Interestingly, with this method, observations collected under a given climate scenario can be used to adjust mean and variance of the model ensemble under a different scenario. Another advantage of the proposed method is that it does not rely on a separate calibration of each individual crop model. The uncertainty reduction resulting from the adjustment of an ensemble of crop models to observations was assessed in a numerical application. The implementation of the Bayes linear method systematically reduced uncertainty, but the results showed the effectiveness of this method varied in function of several factors, especially the accuracy of the yield observation, and the covariance between the crop model output and the observation.

1. Introduction

Many studies have examined the effects of future climate change on crop yields using process-based dynamic crop models. These models use mathematical functions to describe the status of the crop and soil conditions through time. They predict the dynamic of plant growth and development in function of input variables describing soil characteristics, farmers' practices and climate variables. They are now widely used to assess the effect of climate change factors, such as temperature, CO_2 concentration and rainfall, on crop yield (Tubiello and Ewert, 2002).

Crop model predictions are subject to several sources of errors. Their parameter values are often poorly estimated and their input variables are not always accurately measured (Makowski et al., 2006; Wallach, 2011; Wallach et al., 2014). Model predictions and their accuracy also depend on the equations selected to describe key crop processes (Martre et al., 2015). For major crops such as wheat, maize and rice, many different process-based models are now in use, and recent studies showed that, for given sites, these models differed considerably in their yield simulations (Asseng et al., 2013; Bassu et al., 2014; Li et al., 2015).

Multi-model forecasting has drawn some attention in crop science. The use of multi-model ensembles has been proposed to analyse uncertainties in crop model outputs (Asseng et al., 2013; Rosenzweig et al., 2013). The principle is to run several individual crop models for one or several climate scenarios in order to generate ensembles of output values. This approach was recently used to analyse how much individual models differed in their yield outputs under various climate and soil conditions, and the results revealed the ranges of possible yield values were very large (Asseng et al., 2013; Bassu et al., 2014; Li et al., 2015). Several authors recently claimed that the use of mean or median of ensembles of crop models results in smaller prediction errors compared to individual models, and that multi-model ensembles could be used to improve the reliability of yield simulations (Asseng et al., 2015; Martre et al., 2015).

In most cases, ensembles of crop model outputs are analysed using simple graphical methods. Ranges of individual model outputs are usually described using box-plots showing minimum and maximum values, median, mean, and 1st and 3rd quartiles (Asseng et al., 2013; Bassu et al., 2014; Li et al., 2015). More extreme per-

* Correspondence address: UMR Agronomie, INRA, AgroParisTech, Université Paris-Saclay, 78850 Thiverval-Grignon, France. Fax: +33 1 30 81 54 25.
E-mail address: makowski@grignon.inra.fr

http://dx.doi.org/10.1016/j.eja.2015.12.012

centiles (e.g., 10th and 90th) and coefficients of variation are also sometimes presented (Martre et al., 2015; Asseng et al., 2015). Recently, Makowski et al. (2015) proposed a more sophisticated statistical method. The principle is to fit a statistical regression model to the original ensemble of simulated outputs. The fitted statistical model is then used as a meta-model to emulate the ensemble of process-based crop models, to estimate the response of yield to temperature and CO_2 concentration, and to analyze the between-crop model variability.

This paper describes a simple Bayesian method – called Bayes linear method (Goldstein, 1999; Goldstein and Bedford, 2007) – for updating ensemble of crop model yield outputs using yield observations. The principle is to summarize the ensemble of crop model outputs by its mean and variance, and then to adjust these two quantities using yield observations. The adjusted mean and variance combine two sources of information, i.e., the ensemble of crop models and the observations. Potentially, the proposed method can handle different types of observation, such as yield data collected in field experiment, or yield data collected in heating experiments. Observations and yield predictions do not need to correspond to the same climate scenario; observations collected under a given scenario (e.g., current climate condition) can be used to adjust mean and variance of yield under a different scenario (e.g., a climate change scenario). The ensemble of crop model outputs is summarized by its mean and variance and these two quantities are directly updated using yield observations. Bayes linear analysis follows the same principles as those considered in traditional Bayesian analysis: prior beliefs about a quantity of interest are specified and then updated using new observations. Contrary to many standard Bayesian methods, Bayes linear analysis does not require large number of simulations and makes no assumption about the form of the probability distributions of the uncertain quantities. An advantage of Bayes linear analysis is that it gives us a straightforward way for updating prior assessments on the basis of observations (Goldstein and Bedford, 2007).

Potentially, Bayes linear analysis can reduce uncertainty in yield predictions. However, the usefulness of this approach may depend on several factors, specifically on the level of accuracy of the observations used to update the ensemble of crop models, on the correlation between crop model outputs and observations, and on the distribution of crop model outputs. In this paper, I describe the method and present a simple numerical application in which I assess the uncertainty reduction resulting from the adjustment of an ensemble of crop models to observations.

2. Method

2.1. Bayes linear analysis

Bayes linear analysis is a simple quantitative method to express subjective beliefs about a quantity of interest and to update these beliefs using observations (Goldstein, 1999; Goldstein and Bedford, 2007). Here, the quantity of interest is the yield (averaged over years) of a crop in a given site under a given climate scenario. This quantity is further noted Z.

The prior beliefs about Z are expressed through a set of yield outputs simulated by an ensemble of crop models for a given site and a given climate scenario, and are summarized by a prior mean $E(Z)$ and a prior variance $V(Z)$. The prior mean represents the mean value of Z across the ensemble of crop models, and the prior variance reflects the between-crop model variability. The Bayes linear method consists in updating $E(Z)$ and $V(Z)$ on the basis of an observed yield data D as follows (Goldstein, 1999; Goldstein and Bedford, 2007; Gosling et al., 2013):

$$E_D(Z) = E(Z) + \mathrm{Cov}(Z, D)\,[V(D)]^{-1}\,(D - E(D)) \quad (1)$$

$$V_D(Z) = V(Z) - \mathrm{Cov}(Z, D)^2\,[V(D)]^{-1} \quad (2)$$

$E_D(Z)$ and $V_D(Z)$ are the adjusted expectation and variance of Z given the observations D. They depend on the mean and variance of Z, $E(Z)$ and $V(Z)$, on the mean and variance of the observations, $E(D)$ and $V(D)$, and on the covariance between D and Z, $\mathrm{Cov}(Z, D)$. Note that Z and D may correspond to two different climate scenarios. For example, D may be collected in a field experiment exposed to current climate condition, while Z may correspond to the yield value under a climate change scenario.

Eqs. (1) and (2) can be easily generalised to handle multiple observations and multiple variables Z (Goldstein, 1999; Goldstein and Bedford, 2007; Gosling et al., 2013). In this case, $\mathrm{Cov}(Z, D)$ is defined as a matrix including the covariances between all observations and all variables, $E(Z)$, $E(D)$, and $E_D(Z)$ are vectors, $V(Z)$, $V(D)$, and $V_D(Z)$are variance-covariance matrices, and Eq. (2) becomes $V_D(Z) = V(Z) - \mathrm{Cov}(Z, D)\,[V(D)]^{-1}\mathrm{Cov}(Z, D)$.

The adjusted mean $E_D(Z)$ may be viewed as an estimator of Z minimizing the mean squared error (Goldstein, 1999; Goldstein and Bedford, 2007). $E_D(Z)$ and $V_D(Z)$ offer simple approximations of the posterior expected value and variance of Z. If the joint distribution of Z and D is multivariate Gaussian, these approximations are exact. In this case, the joint distribution is defined by

$$\begin{pmatrix} D \\ Z \end{pmatrix} \sim N\left(\begin{pmatrix} E(D) \\ E(Z) \end{pmatrix}, \begin{pmatrix} V(D) & \mathrm{Cov}(D,Z) \\ \mathrm{Cov}(D,Z) & V(Z) \end{pmatrix}\right)$$

and the conditional distribution of Z (i.e., the posterior distribution) is expressed as

$$Z|D \sim N\,(E_D(Z), V_D(Z))\,. \quad (3)$$

The implementation of the method described above requires the estimation of $E(Z)$, $E(D)$, $V(Z)$, $V(D)$, and $\mathrm{Cov}(Z, D)$. Estimation procedures are described in the next two sections.

2.2. Estimation of prior mean and variance with an ensemble of crop models

I present two different approaches for calculating $E(Z)$ and $V(Z)$ from an ensemble of simulated outputs. In the first approach, the prior mean and variance are directly calculated from the crop model simulations. $E(Z)$ and $V(Z)$ are therefore calculated as follows:

$$E(Z) = \tfrac{1}{N}\sum_{i=1}^{N} Z_i$$

and

$V(Z) = \tfrac{1}{N-1}\sum_{i=1}^{N}\left(Z_i - \tfrac{1}{N}\sum_{i=1}^{N} Z_i\right)^2$, where Z_i, $i = 1, \ldots, N$, are the yield values simulated by an ensemble of N different crop models for a given site and under a given climate scenario of interest.

The second approach does not estimate the prior mean and variance from the original simulated yield outputs, but from the outputs of a meta-model, i.e., a statistical model fitted to the original simulated outputs. Here, I use the meta-model developed by Makowski et al. (2015) that relates simulated yield output to climate variables as follows:

$$Z_i = \alpha_{0i} + \sum_{k=1}^{P} \alpha_{ki} X_k + \varepsilon_i \quad (4)$$

where Z_i is the yield output obtained with the ith crop model, $i = 1, \ldots, N$, for a given site and a given climate scenario of interest, X_k, $k = 1, \ldots, P$, are P climate variables characterizing the considered climate scenario, ε_i is a residual term assumed independently and identically distributed, $\varepsilon_i \sim N\left(0, \tau^2\right)$, α_{ki}, $k = 0, \ldots, P$, are random