Europ. J. Agronomy 52 (2014) 191–197

Contents lists available at ScienceDirect

European Journal of Agronomy

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

Combining input uncertainty and residual error in crop model predictions: A case study on vineyards

Sébastien Roux[a,*], François Brun[b], Daniel Wallach[c]

[a] UMR 1230 SYSTEM, INRA Montpellier 2, Place Viala, 34060 Montpellier cedex 1, France
[b] ACTA, RMT modélisation, UMR 1248 AGIR, INRA, 24 Chemin de Borde Rouge—Auzeville, 31326 Castanet Tolosan cedex, France
[c] UMR 1248 AGIR, INRA Toulouse, INRA, 24 Chemin de Borde Rouge—Auzeville, 31326 Castanet Tolosan cedex, France

ARTICLE INFO

Article history:
Received 25 June 2013
Received in revised form 5 September 2013
Accepted 18 September 2013

Keywords:
Error evaluation
Uncertainty analysis
Sensitivity analysis
Water balance crop model

ABSTRACT

As crop modelling has matured and been proposed as a tool for many practical applications, there is increased need to evaluate the uncertainty in model predictions. A particular case of interest that has not been treated before is that where one takes into account both uncertainty in the model explanatory variables and model residual error (the uncertainty in model predictions even when the explanatory variables are perfectly known). The specific case we consider is that of a model for predicting water stress of a vineyard. For many of the model explanatory variables, the vine grower (or the farmer advisor) has a choice between approximate values which are easily obtainable and more precise values that are more difficult (and more expensive) to obtain. We specifically discuss the explanatory variable “initial water stress” which is directly based on the initial soil water content and can be estimated or measured (precise but expensive). The vine grower is interested in the decrease in uncertainty that would result from measuring initial water stress, but it is the decrease in total uncertainty, including model residual error, that is of importance.

We propose using accurate measurements of water stress over time in multiple vineyards, to estimate model residual error. The uncertainty in initial water stress can be estimated if one has approximate and precise values of initial water stress in several vineyards. We then combine the two sources of error by simulation thanks to an independence hypothesis; the model is run multiple times with a distribution of values for initial water stress, and on each day a distribution of model residual errors is added to the result.

The results show that the resulting uncertainty is quite different in different fields. In some cases, uncertainty in initial water stress becomes negligible a short time after the start of simulations, in other cases that uncertainty remains important, compared to model residual error, throughout the growing season. In all cases, residual error is a substantial percentage of overall error and thus should be taken into account.

1. Introduction

As crop modelling has matured and been proposed as a tool for many practical applications, there is increased interest in evaluating the uncertainty in model predictions (Aggarwal and Kalra, 1994; Aggarwal and Mall, 2002; Lamboni et al., 2009; Wallach et al., 2008; Wang et al., 2005). If crop models are to be used by farmers, farm advisors or policy makers as the basis for decisions, it is essential to have a measure of the reliability of model results.

To date, there have been two major approaches to evaluating the uncertainty of model predictions. In the first that we will denote here as “error evaluation studies”, one compares observed and simulated values and calculates a criterion of model error, which is often mean squared error (MSE) (Abedinpoura et al., 2012; Palosuo et al., 2011; Rötter et al., 2012). If the data used for evaluating error have not been used for model calibration, then MSE can be used as an estimate of mean squared error of prediction (MSEP) (Wallach, 2006). To simplify the following discussion, we will assume that MSE is the criterion used for evaluation, and that it is an unbiased estimator of MSEP. This is an uncertainty calculation in the sense that it gives information about how uncertain our predictions are; specifically, MSE estimates the average squared distance between future observed values and simulated values. Here one is evaluating a model with some fixed set of parameter values and some fixed way of obtaining the values of the explanatory variables. The fact that other parameters could have been chosen, or that the explanatory variables may have some error, is not explicitly considered. The result is an estimate of MSEP for the model that one has.

* Corresponding author. Tel.: +33 4 99 61 21 54; fax: +33 4 99 61 30 34.
E-mail address: rouxs@supagro.inra.fr (S. Roux).

1161-0301/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.eja.2013.09.008

The error here, since it is based on a comparison of simulated and observed values, automatically includes all sources of error, including possible errors in parameters, in input variables or in model equations (and also errors in the observed values). The MSEP that one estimates is a single number of average squared error, where the average is over the range of situations represented by the data. This type of study does not give information about how error varies between situations.

The second approach, uncertainty propagations studies, focuses on the uncertainty in model parameters or input variables, and explores how those uncertainties propagate through to model responses, for example yield. There have been numerous studies of this sort, mainly but not only concerning the uncertainty engendered by parameter uncertainty (Confalonieri et al., 2006; Iizumi et al., 2009; Post et al., 2008; Wang et al., 2005). Many of these studies are mainly concerned with ranking the parameters, in order to identify which parameter uncertainties are mainly responsible for the uncertainty in model response (sensitivity analysis).

Uncertainty and sensitivity analysis focus on the uncertainty due to some specific source of uncertainty. If for example the study specifically examines uncertainty in input values, then the effects of errors in parameters or in model equations is not taken into account. In compensation, such studies have two important advantages. First, they allow one to test hypothetical errors in parameters or inputs. Second, these studies give more detailed information than evaluation studies. The results are not an average level of error, but rather an uncertainty which is specific to the conditions investigated. For example, uncertainty in a model parameter that enters into the calculation of the effect of water stress may lead to large uncertainty in situations where there is water stress, but to no uncertainty in well watered conditions.

There are situations where it is important to combine the advantages of both techniques; one wants an overall measure of error, but one also wants to explore hypothetical situations and one wants information on how error varies across situations. An example would be using a model to compare management strategies. Suppose that it is reasonable to assume that input variables are measured correctly, but some of the parameters are obtained by calibration and thus have uncertainty due to the calibration process. One would want both overall error, and also to take into account the fact that the error related to parameter estimation is different for each management strategy. This can be achieved by using a Bayesian approach that calculates a posterior distribution both for the model parameters and for model residual error. Wallach et al. (2008) used such a Bayesian approach to assign uncertainties to a comparison of different irrigation strategies. (One could also take into account both parameter uncertainty and residual error using a frequentist approach to parameter estimation (Seber and Wild, 1989) but this does not seem to have been done for crop models).

A different situation is that where one wants to take into account both overall error and error arising from incomplete knowledge of model inputs. This would often be the case for a model to be used by growers; often some model inputs are difficult to obtain, and so growers might choose to use approximate values that are easier to obtain. It is of importance then not only to evaluate the uncertainty associated with using uncertain input values, but also to compare that with overall error. There do not seem to have been studies which combine both estimation of overall error and estimation of the error specifically due to uncertain input values.

The purpose of this paper is to show how to combine the advantages of model evaluation and uncertainty analysis, in order to obtain uncertainty estimates for crop models that on the one hand include all types of error, but on the other still allow one to get detailed information about the consequences of input uncertainty.

The example we treat concerns a model that is used by growers to evaluate the water stress dynamics of an intercropped vineyard (Celette et al., 2010). The model calculates daily the fraction of transpirable soil water (FTSW), which is a water stress index that is used by growers to assess the dynamics of the water status experienced by a vineyard (Pellegrino et al., 2006). Managing this water deficit is an important issue in optimizing grape quality.

The model requires a certain number of input variables, which can be obtained in different ways. In general there is a choice between an approximate procedure and a more accurate procedure, often direct measurement. The specific input variable we treat is FTSW at budburst (FTSW0), when the model calculations begin. This can be estimated by using model calculations (described in more detail in the next section), or can be measured. The grower must choose between the two possibilities. The model-based estimation is less expensive, but direct measurement is more accurate. The choice will depend on how much uncertainty in FTSW is introduced by uncertainty in FTSW0. However, the choice will also depend on the overall uncertainty in FTSW. This example comes from a larger study that considers the effect of using approximate values for many different input variables.

It will be clear that the approach used here could be generalized to other studies where one wants to combine a consideration of overall error with consideration of the effect of specific uncertainties in input variables.

2. Materials and methods

2.1. The model and its uncertain inputs

The model here is a model for FTSW in a vineyard. When a grass cover in present in the intercrop, the water balance is modified (more transpiration but better infiltration), which requires the use of a specifically adapted water balance model when assessing management strategies for grass cover introduction (Ripoche et al., 2011). Therefore the model here includes both the cases where there is and where there is not grass cover between the rows (Celette et al., 2010). This model has been specifically designed for use by growers, to enable them to estimate the vineyard FTSW during the growing season as a basis for making decisions on the introduction of grass cover.

The model inputs that can be obtained more or less accurately include budburst date, which can be estimated or observed, FTSW at budburst (FTSW0), which can be estimated or measured, daily weather, which can be obtained from a weather station at the site of the field or approximated using a more distant weather station, runoff curve numbers for bare and covered soil, which can be estimated from runoff measurements or based on plot characteristics, maximum available water, which can be estimated from soil texture or measured, etc. We refer to the inputs obtained using the more accurate approach as the “accurate” values and the values obtained using the less accurate approach as the “approximate” values.

Here we consider just the case where we want to study the difference between using accurate and approximate values of FTSW0 with all the other inputs fixed to their accurate value. The accurate value of FTSW0 is obtained by measuring FTSW at budburst. The approximate value for year y is obtained by setting FTSW = 0 at the end of summer of year $y-1$ and then running the model to budburst in year y. The rationale is that it is very common to have dry soil at the end of summer in vineyards in a Mediterranean environment. However, even if the assumption about soil water at the end of summer is correct, the approximate value for FTSW0 will still be subject to error related to model error. Since soil water on any day is equal to the value on the previous day plus some change, errors in initial soil water continue to affect soil water on subsequent days.