Environmental Modelling & Software 51 (2014) 136–148

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Relative yield decomposition: A method for understanding the behaviour of complex crop models

Grégoire Talbot [a], Sébastien Roux [a,*], Anil Graves [b], Christian Dupraz [a], Hélène Marrou [a], Jacques Wery [c]

[a] INRA, UMR SYSTEM, Bâtiment 27, 2 Place Viala, 34060 Montpellier Cedex 1, France
[b] Natural Resources Management Centre, School of Applied Sciences, Cranfield University, Cranfield, Bedford MK43 0AL, UK
[c] SUPAGRO, UMR SYSTEM, Bâtiment 27, 2 Place Viala, 34060 Montpellier Cedex 1, France

ARTICLE INFO

Article history:
Received 21 January 2013
Received in revised form
13 September 2013
Accepted 16 September 2013
Available online 22 October 2013

Keywords:
Process-based model
Model behaviour
Virtual experiment
Crop yield

ABSTRACT

Dynamic crop simulation models are widely used to investigate, through virtual experiments, the response of crop yield to changes in climate, management or crop genetic traits. In a search for widespread applicability, crop models include a large number of processes, sometimes to the detriment of their mathematical transparency.

Simulated crop yield responses to variation in model inputs result from the integration over a long period (one or several years) of many different crop processes interacting at the model time-step, typically the day. Thus, by definition, yield explanatory factors are intricate and difficult to link efficiently to the crop processes. Ranking their relative contributions to the final yield output is for example almost impossible.

In this work, we introduce a new approach to understand the response of crop yield Y by comparing two simulation runs (computing two yields Y_1 and Y_2) of the same model and by focussing on the relative yield: $y = Y_1/Y_2$. Providing that the mathematical formulation of the dynamic crop model verifies simple hypotheses held by most crop models, we show that it is possible to factorise the relative yield y into several terms. These terms can be (i) interpreted as the specific effects of the modelled crop processes on the crop yield, (ii) compared to rank the effects of the crop processes on the crop yield. Their definition involves using state variables of the model computed during the simulation runs. The method does not involve running the model numerous times, neither changing its formulation. It may require to output new variables that are not in the set of variables proposed by the released version of the model. We call our method the relative yield decomposition (RYD) method.

We illustrate how the RYD provides insight in the analysis of complex crop models by applying it to two models: Yield-SAFE (agroforestry model) and STICS (crop model). The method allows to identify and quantify the importance of the main processes responsible for crop yield variations for different simulation configurations in the two models.

The relative yield decomposition method is complementary to other model analysis methods like sensitivity analysis or multiple model simulations. We show that it could be applied to some widely used crop models (e.g. AQUACROP, CERES, CROPGRO, CROPSYST, EPIC, SIRIUS, SUCROS). The relative yield decomposition method appears as a powerful and generic tool to analyse the behaviour of complex crop models that can help to improve the formulation of the models, or even to study specific plant traits or crop processes when applied to a model accurate enough.

1. Introduction

Since the development of early crop simulation models on mainframe computers in the 1960s (Loomis and Williams, 1962; Bouman et al., 1996; Graves et al., 2005) the use of dynamic simulation models of tree and crop growth, either in pure or mixed systems has expanded, facilitated by advances in modern computing technology. Such crop simulation models (we use this term broadly to refer to all models that simulate the growth of crops and trees in pure or mixed systems) have been used in research, education, and decision-making (Matthews and Stephens,

* Corresponding author.
E-mail address: rouxs@supagro.inra.fr (S. Roux).

1364-8152/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.envsoft.2013.09.017

Notations

Upper case letters refer to absolute variables, while lower case letters refer to relative variables. Relative variables are defined as the ratio of the absolute values for the scenario under analysis compared to the reference scenario $x = X/X^*$ where superscript * refers to the reference scenario.

Letters with superscript t refer to values at the model time-step t while letters without superscript refer either to values independent from time or values averaged or accumulated with time.

A	light capture efficiency
B	biomass
C	incident light
D	density for agroforestry systems: proportion of cropped area for crops, tree density for trees.
E_i	generic notation for the effect of yield reducing factor i
$E_{i\|1:(i-1)}$	effect of factor i knowing the effects of all yield reducing factors from the first to the $(i-1)^{\text{th}}$. The corresponding mathematical definition (see Section 2) is: $E_{i\|1:(i-1)} = \sum_t \frac{E_1^t \cdot \ldots \cdot E_{i-1}^t}{\sum_{t'} E_1^{t'} \cdot \ldots \cdot E_{i-1}^{t'}} E_i^t$
F	effect of vigour at flowering stage on harvest index (model STICS)
G	effect of the length of grain filling stage on harvest index (model STICS)
H	harvest index
I	light competition index
L	light intercepted
M	effect of maintenance cost on harvest index for trees (model Yield-SAFE)
N	nitrogen stress effect on light use efficiency (model STICS)
P	effect of crop phenology on light use efficiency (model STICS)
R	light use efficiency $R = B/L$
S	effect of light saturation on light use efficiency (model STICS)
T	effect of extreme temperatures on light use efficiency (model STICS)
U	effect of extreme temperatures on harvest index (model STICS)
W	water stress effect on light use efficiency
$w^t_{i\|1:(i-1)}$	weights of the effects $E^t_{i\|1:(i-1)}$ at the model time-step in the calculation of the average effect $E_{i\|1:(i-1)}$ at the analysis time-step.
Y	yield, referring to grain biomass for crops, and timber biomass for trees.

2002). Within these arenas, they are widely used to investigate the behaviour of crop, tree, and crop-tree systems through "virtual" experimentation, especially when analysing the gap between actual and potential yields (Lobell et al., 2009). They have also been used to explore management options (e.g. DeJonge et al., 2007; Tojo Soler et al., 2007; Shili-Touzi et al., 2010), to assess the potential of cultivars with different phenotypic or genetic traits (e.g. Asseng et al., 2002; Debaeke, 2004; Chenu et al., 2009; Semenov et al., 2009), and to predict the behaviour of crop systems under climate change (e.g. Asseng et al., 2004; Meza and Silva, 2009; Lhomme et al., 2009; Liu et al., 2010).

Dynamic crop simulation models generally work on a time-step that is suited to the mechanisms involved in crop growth (sub-daily to daily time-step), which we refer to here as the "model time-step". The causal links between the input variables (for climate, soil, and management), and the state variables of the system (including those for crop growth) are defined at this time-step by the equations of the model. However, the evaluation of cropping systems is achieved through the use of outputs calculated at a longer time-step (one year to several decades), which we refer to as the "analysis time-step". At this time-step, the crop yield Y is a very complex function of the inputs and parameters of the model. Understanding how the final yield is achieved, what the key processes and influences are, and when the critical phases occur, is essential for both model users and developers. However, there are few analytical methods that can provide a means of achieving such insight.

The most frequently used approaches consist in analysing the response of the crop yield Y to variations in parameters and input variables. This can be achieved either through sensitivity analysis (on parameters or input variables) or by comparing multiple simulations, obtained after "switching off" selected crop processes within the model. Sensitivity analysis is widely used to understand the effect of parameter variation on model outputs. This has various critical uses, such as verification and debugging of models, as well as identification of sensitive parameters that, for example, need extra care in measurement or assessment, because they dominate the response of the model (Keesman et al., 2011). In the case of crop growth models, sensitivity analysis has been used to provide statistical data on the relationship between simulated yield and model inputs in order to understand the behaviour of simulated crop growth (Wang et al., 2013; Carpani et al., 2012; Pogson et al., 2012; Confalonieri et al., 2010; Makowski et al., 2006). This information can then be directly related to the crop processes, provided the studied parameters have a biophysical meaning. A second approach consists in comparing different model formulations so that the effect of a given process in determining crop yield can then be identified by comparing the model output with and without that process. Cox et al. (2006) and Crout et al. (2009) proposed such a methodology to simplify mechanistic models. Affholder et al. (2003) ran simulations with and without water and/or nitrogen stress with the STICS crop model (Brisson et al., 2009) to conduct a yield gap analysis in a farmer's field network. This approach actually considers different models to generate and compare different simulated yields. If this approach provides useful information on the yield factors, it does not give direct insight on how one simulated yield produced by one simulation run is explained by the different crop processes. For example, comparing the yield with or without water stress is different to directly quantifying how much water stress has impacted the simulated yield in comparison with other abiotic stresses.

In order to overcome these drawbacks, we developed a new approach to understand the response of crop yield Y by comparing two simulation runs (computing two yields Y_1 and Y_2) of the same model and by focussing on the relative yield: $y = Y_1/Y_2$. Our method applies to models verifying two hypotheses used by most crop models: firstly, the final yield value must be defined as a sum of positive yield increments at the various model time-steps; secondly, for all time-steps of the model, the yield increments must be written as a product of positive factors. We show that it is then possible to factorise the relative yield y into several factors which (i) can be interpreted as the specific effects of the modelled crop processes on the crop yield, (ii) can be compared to rank the effects