Ecological Modelling 368 (2018) 344–356

Contents lists available at ScienceDirect

Ecological Modelling

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

A systematic approach to identifying key parameters and processes in agroecosystem models

Vasileios Myrgiotis[a,b,*], Robert M. Rees[a], Cairistiona F.E. Topp[a], Mathew Williams[b]

[a] SRUC, Edinburgh EH9 3JG, UK
[b] School of GeoSciences, University of Edinburgh, Edinburgh EH9 3JN, UK

ARTICLE INFO

Article history:
Received 26 June 2017
Received in revised form
11 December 2017
Accepted 12 December 2017

Keywords:
Soil biogeochemistry
Ecosystem modelling
Landscape-DNDC
Sensitivity analysis

ABSTRACT

Process-based agroecosystem biogeochemistry models are widely used to quantify the flow of water and nutrients in agricultural ecosystems and they have become important tools in the effort to address the twin challenges of reducing greenhouse gas emissions and improving agricultural sustainability. Model parameters require careful calibration, as they affect the simulated processes and outputs. Sensitivity analysis (SA) is commonly used to quantify the impacts of parameters on outputs, and guide the calibration process. Here we demonstrate a systematic approach for SA, which assures that (1) the role of time-dependency in the sensitivity indices is considered and (2) the SA is not biased by the edapho-climatic conditions at individual sites. Demonstrating this approach, we examine the parametric sensitivity of an advanced agroecosystem model (Landscape-DNDC) using a framework that is based on (1) the Sobol SA method, (2) model simulations at three UK arable sites and (3) the grouping of the model's parameters according to the processes they affect. The findings of this research identify the parameters and processes that should be carefully examined in order to minimise the impact of parametric uncertainty on model outputs. We show that a limited number of parameters are responsible for a large part of the sensitivity of model outputs. The description of soil microbial dynamics is identified as a key source of output sensitivity. Also, we show that individual management activities can significantly affect the time-dependency of the parametric sensitivity indices for certain model outputs.

1. Introduction

Agroecosystem biogeochemistry (BGC) models are computational tools that simulate the processes that drive the fluxes of nutrients through agricultural ecosystems, their interactions and their environmental sensitivity. They take measurable information on the drivers and initial state of the ecosystem (e.g. climate, vegetation type, soil properties, etc.) and feed them to a set of mathematically-described interacting processes that represent the system and its evolution. Measured input data typically contain uncertainties while the modelled processes can be highly customisable especially if they depend on several parameters. As a consequence, model outputs encapsulate the effects of data and model-related uncertainties. These uncertainties are caused by (1) the spatial and temporal variability of the measured input data (2) the model's structure/architecture and (3) the lack of "precise" quantification of the mathematical and/or statistical parameters that make up the model's formulation. These three uncertainty sources are also known as input, structural and parametric respectively, and they have a combined impact on the model's predictive quality (Campolongo et al., 2007; Norton, 2015; Baroni and Tarantola, 2014).

Analyses of the sensitivities of model outputs to input, structural and parametric uncertainties form an important part of model development and application (Della Peruta et al., 2014; Qin et al., 2016; Fan et al., 2016). SA can be used in model development as a way to simplify a model (i.e. identify less significant parameters/processes) and refine the prior ranges of its parameters (Heinen, 2006). Model users apply SA to identify which parameters to include in model calibration and to gain an understanding of the model's behaviour under the conditions that are specific to their work. Sensitivity analysis of model outputs to model inputs is used to derive estimates of the impacts that the spatiotemporal variability of measurable inputs (e.g. data on climate, soil properties) can have on a model's outputs (Van Oijen et al., 2005; Rafique et al., 2015; van Oijen et al., 2011). The existence of persistent bias in a model's outputs can be controlled by identifying how the architecture of a model's mechanisms, and the mechanisms them-

* Corresponding author at: SRUC, Edinburgh EH9 3JG, UK.
E-mail address: vasileios.myrgiotis@sruc.ac.uk (V. Myrgiotis).

https://doi.org/10.1016/j.ecolmodel.2017.12.009
0304-3800/

selves, affect the model's outputs. The quantification of structural uncertainty can be achieved by evaluating a model under different architectures and module combinations (Sándor et al., 2016; Ruane et al., 2016). Nevertheless, such an exercise requires models that can accommodate a set of conceptually different but interoperable modules and is, thus, more difficult to examine. On the other hand, the quantification of the sensitivities of different outputs to a model's parameters (i.e. parametric sensitivity) is mainly dependent on whether the model's format offers access to its parameters. In this study, we focus exclusively on parametric sensitivity analysis, to which we hereafter refer when using the term *sensitivity analysis* (SA).

Global parametric SA (GSA) methods are commonly used in studies with agroecosystem models. In order to achieve their aim, the values of all the examined parameters are perturbed concurrently and the impact of each parameter (i.e. direct and indirect) on the output of interest is quantified (Pianosi et al., 2016; Norton, 2015; Cariboni et al., 2007). Morris and Sobol are two of the most widely used GSA methods with Sobol being more computationally expensive and detailed than Morris (Confalonieri et al., 2010; Sarrazin et al., 2016; Wainwright et al., 2014; Iooss and Lemaître, 2014; Campolongo et al., 2004). The parametric sensitivity of a model output can be quantified through SA by using (1) a single value (e.g. simulated soil CO_2 at day *d*); (2) the mean value during a defined period (e.g. annual or weekly mean) or (3) a cumulative amount during a defined period (e.g. cumulative soil CO_2 fluxes during one year). The use of a single simulated data point (e.g. CO_2 flux at day *d*) to quantify the parametric sensitivity of an output (e.g. CO_2) might not be appropriate for model outputs that behave in a highly dynamic manner (e.g. greenhouse gases). On the other hand, the use of cumulative values for a single time period (e.g. a year, week or month) might not capture all the possible effects of parametric uncertainty on a simulated variable if this variable is highly dependent on other actions. For example, soil N_2O fluxes might be strongly dependent on the timing of fertiliser application just like NO_3 loss through leaching might be dependent on the timing of heavy rainfall events (Gerber et al., 2016; Molina-Herrera et al., 2016; Castellano et al., 2010; Ma et al., 2010). In spite of that, the issue of time-dependency of the estimated sensitivity indices (SI) is rarely examined in SAs with ecosystem BGC models but has been considered in some studies with hydrological models (Song et al., 2013; Pianosi and Wagener, 2015; Guse et al., 2016).

Another important aspect, which is also rarely considered in relevant studies, is the heterogeneity of agroecosystems. Most studies on the parametric sensitivity of agroecosystem models use simulations at a single site to quantify the sensitivity of the model's outputs (Necpálová et al., 2015; Della Peruta et al., 2014; Qin et al., 2016, 2013). However, this approach does not account for the fact that the edapho-climatic conditions at the simulated site could be strongly influencing the estimated SIs and the SA overall (Li et al., 2004). In this respect, the performance of simulations at more than one site is a way to ensure the robustness of the SA. This is important particularly if the model's intended spatial scale of application is large (e.g. sub-national level). In general, a lack of studies using process-based agroecosystem BGC models and focusing on the parametric sensitivity of their outputs can be observed in the relevant literature. Most SA studies with agroecosystem BGC models focus on input uncertainty and only few studies have focused on the parametric sensitivity of the models (Qin et al., 2013; Wang and Chen, 2012; Del Grosso et al., 2010; Hastings et al., 2010; Zaehle et al., 2005; Klatt et al., 2016). The role of parametric uncertainty is more often considered in studies that deal with the calibration of model parameters and in which the results of parametric SAs are not always presented or discussed (van Oijen et al., 2011; Lehuger et al., 2009; Rafique et al., 2015; Li et al., 2015). Also, such studies tend to focus only on a single model output (e.g. soil N_2O emissions or soil C content) (Lehuger et al., 2009). In this context, the consideration of more than one model outputs in SAs can provide a more complete picture of how parameters affect model prediction.

In this study, we present a simple framework for the quantification of model parametric sensitivity that is tailored to agroecosystem models. The model that we use to demonstrate the framework is Landscape-DNDC, which is a typical process-based agroecosystem BGC model (Haas et al., 2012). Landscape-DNDC shares similarities with other agroecosystem models in terms of concept, mathematical formulation and parameterisation and more so with other DNDC-based models (Gilhespy et al., 2014; Abdalla et al., 2010; Smith et al., 2010). Therefore, we believe that the results of this study will be relevant to other agroecosystem models. The study focuses on the soil biogeochemistry aspect of the model and our SA examines the importance of the relevant parameters only (i.e. plant growth-related parameters not considered). We use the Sobol SA method (Campolongo et al., 2007) and collect model outputs for 10 key variables. Taking into account the aforementioned limitations of other SA studies, here, we consider the role of edapho-climatic conditions by performing simulations at three UK arable sites (representative of UK's soils and climate). In order to examine the time-dependency of the estimated sensitivity indices, we collect model outputs at eight different temporal resolutions (i.e. one annual value and seven weekly values). Also, we are interested in understanding the role of processes for model outputs since this can lead to observations that are of practical value in a broader sense. To examine this aspect, we sort the model's parameters into three groups according to the type and role of the processes that they affect, and examine the contribution of each group to output sensitivity. In summary, the main objectives of the study are to (1) quantify the parametric sensitivity of key outputs of the Landscape-DNDC model; (2) examine how parameter groups affect model outputs and (3) assess the impact of the temporal resolution of the SA on the estimated parametric sensitivities.

2. Materials and methods

2.1. The Sobol method

The Sobol SA method is a global, variance-based and model independent method that can be used to quantify the sensitivity of model outputs to inputs and parameters (Baroni and Tarantola, 2014). The method estimates the first order sensitivity index (S_i), which presents the direct contribution of a parameter (X_i) to an output (Y), and the total sensitivity index (S_T), which represents the direct and indirect contribution of parameter X_i to the sensitivity of Y. S_i and S_T are estimated using (1) and (2) respectively:

$$S_i = \frac{V\left[E\left(Y|X_i\right)\right]}{V\left(Y\right)} \quad (1)$$

$$S_T = \frac{E\left[V\left(Y|X_{-i}\right)\right]}{V\left(y\right)} \quad (2)$$

where X_{-i} denotes all inputs except X_i, V denotes the variance and E the expectation. The Sobol method also allows for the estimation of sensitivity indices of higher order (i.e. second, third, etc.). For example, the second order Sobol sensitivity index S_{ij} quantifies the variance caused to Y by the interaction between parameters X_i and X_j. The number of model simulations (R) that is required for the estimation of S_i and S_T is equal to $N(2D+2)$ where N is the sample size and D is the number of parameters (Nossent et al., 2011). The value of N is case-specific with values in the relevant literature ranging between a few hundred and tens of thousands (Nossent et al., 2011; Wainwright et al., 2014; Pianosi and Wagener, 2015). The examination of the convergence of the estimated SIs can be used to ensure that the chosen N was sufficiently large (Sarrazin