Environmental Modelling and Software 107 (2018) 281–297

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

An empirical workflow to integrate uncertainty and sensitivity analysis to evaluate agent-based simulation outputs

Carolina G. Abreu*, Celia G. Ralha

Computer Science Department, Institute of Exact Sciences, University of Brasília, P.O.Box 4466, Zip Code 70.904-970, Brasília, Brazil

ARTICLE INFO

Keywords:
Uncertainty analysis
Sensitivity analysis
Agent-based model
Spatial simulation
Land use cover change

ABSTRACT

This paper presents an empirical study comparing different uncertainty analysis (UA) and sensitivity analysis (SA) methods, focussing their usefulness for the output analysis of land use/land cover change (LUCC) agent-based models (ABMs). As a result, a workflow to integrate UA and SA is presented to evaluate ABMs outputs. We developed a baseline scenario and performed a comprehensive investigation of the impacts that differences in sample sizes, sample techniques, and SA methods may have on the model output. The analysis is done in the context of a particular agent-based simulator with a LUCC model in a Brazilian Cerrado case study. The experiments indicate that there are known challenges to be overcome by the use of statistical methods. Even though the presented analysis was done over a particular simulator, we intend to contribute to the community that understands the importance of statistical validation techniques to improve the level of confidence in agent-based simulation outputs.

1. Introduction

As cited in the literature, the land use/land cover change (LUCC) systems are dynamic, stochastic, and characterized by nonlinear and non-monotonic relationships between constant changing entities (Parker et al., 2003; Verburg, 2006; Rindfuss et al., 2008). Besides, agent-based models (ABMs) have been used as a natural metaphor to model LUCC dynamics, since they capture emergent phenomena and provide an original description of the modeled system (Schreinemachers and Berger, 2011; Murray-Rust et al., 2013; Ralha et al., 2013). However, ABMs are prone to uncertainty because they reflect the intrinsic randomness of environmental, physical, and social events. The uncertainty may also arise because of insufficient knowledge, lack of data, observation errors, measurements used to parametrize the model, or from vague premises of the model (Ligmann-Zielinska et al., oct 2014; Lilburne and Tarantola, 2009). As a result, one could argue whether there is any quality in model predictions due to high uncertainty and the considerable number of assumptions imposed by ABMs models.

In this scenario, uncertainty analysis (UA) and sensitivity analysis (SA) are currently popular topics in ABMs as well as for many other complex systems (Pappenberger et al., 2008). They are valuable tools in understanding LUCC models and deriving decisions on strategies to reduce model uncertainty. UA provides the variability of model results. SA presents which factors are responsible for this variability. This variability may be expressed quantitatively in terms of "elasticity" of performance concerning parameter levels. High sensitivities (elasticities) give cause for concern about the reliability of a model (Dayananda et al., 2002). A factor is any source of uncertainty in the modeling process, including model structure, initial conditions, and input parameters. Using the terminology proposed by the National Research Council (2012), uncertainty quantification (UQ) is the process of quantifying uncertainties in a computed quantity of interest (QOI), with the goals of accounting for all sources of uncertainty and quantifying the contributions of specific sources to the overall uncertainty, i.e., UA and SA applied in tandem.

Although UA and SA applications are rising, most ABMs struggle with a shortage of testing in general, mainly due to time and other resource constraints (Kelly (Letcher) et al., 2013). Lee et al. (2015) argue that while a modeler invests a lot of time and effort in the development of ABMs, the output analysis is not always considered as deserving the same resource-intensive attention. According to a survey carried out by Heath et al. (2009), less than 5% of ABM publications present any statistical validation techniques. Angus and Hassani-Mahmooei (2015) argue that one possible cause for this "methodological anarchy" derives from the fact that, with so many possible degrees of freedom within an ABM, the responsibility to ensure and to demonstrate that a model is structurally sound and the prediction is reliable falls into each modeler.

We present a UQ workflow to integrate UA and SA in the evaluation

* Corresponding author.
E-mail addresses: carolabreu@unb.br (C.G. Abreu), ghedini@unb.br (C.G. Ralha).

https://doi.org/10.1016/j.envsoft.2018.06.013
Received 10 January 2018; Received in revised form 8 June 2018; Accepted 13 June 2018
Available online 18 June 2018
1364-8152/

Table 1
Selected applications of sensitivity analysis approaches.

Reference	Research Field	No. factors	Sampling	SA method	No. runs
1	Urban drainage	17	MOAT	MOAT	3000
			FAST	E-FAST	3000
			LH	SRC	2800
2	Watershed	13	MC	SPEA	3000
			MC	SRC	3000
			MOAT	MOAT	3000
			METIS	MARS	3000
			METIS	SOT	3000
			MC	DT	400
			LH	DT	400
			OA	DT	529
			OALH	DT	529
			LPTAU	DT	3000
			METIS	DT	3000
			METIS	GP	3000
			FAST	FAST	2777
			rLH	McKey	2890
			SOBOL-QR	SOBOL	3000
3	Crop growth	47	FAST	E-FAST	2049
4	Watershed	5	SOBOL-QR	SOBOL	18000
			MC	MOAT	3000
			MC	LR	3000
			MC	RSA	3000
			SOBOL-QR	SDP	500
5	Flood inundation	6	rLH	SOBOL	8192
			rLH	MOAT	12000
			rLH	Entropy-based	3000
			rLH	RSA	5000
6	Watershed	18	SOBOL-QR	SOBOL	8192
			IFFD	ANOVA	1000
			LH	RSA	10000
			LP	PEST	10000

Where: MOAT = Morris screening One-at-A-Time; (E−)FAST = (Extended) Fourier Amplitude Sensitivity Testing; (r)LH = (replicated) Latin Hypercube; SRC= Standardized Regression Coefficient; MC = Monte-Carlo; LR = Linear Regression; SPEA = Spearman Correlation Coefficient; MARS = Multivariate Adaptive Regression Splines; SOT = Sum-of-Trees; DT = Delta δ Test; OA = Orthogonal Array; OALH = Orthogonal Array-based Latin Hypercube; IFFD = Iterated Fractional Factorial Design; SOBOL-QR = Sobol quasi-random; RSA = Regionalized Sensitivity Analysis; LP = Local Perturbation; PEST = Parameter Estimation Software.

of agent-based simulation outputs. We illustrate the use of this workflow in a particular spatial explicit LUCC case study in the framework Multi-Agent System for Environmental simulation, MASE-BDI. We apply general practices that should be a routine, to improve the level of confidence in results and to promote more rational and efficient use of ABMs. We may cite that broader and more complete workflows for the application of SA were already proposed, such as Pianosi et al. (2016) and Norton (2015). The UA-SA integrated proposal is what set our manuscript apart. We argue that UA should be used as an input to SA, in a broader process of UQ. Also, we noticed some conflicting results when we compared relevant studies on SA, mainly regarding the experimental setup. Table 1 summarizes the studies found in the literature (Vanrolleghem et al., 2015) (1), (Gan et al., 2014) (2), (Wang et al., 2013) (3), (Yang, 2011) (4), (Pappenberger et al., 2008) (5), (Tang et al., 2007) (6). Some authors have compared different SA methods and experimental setup, which are presented in the different lines of the table.

Table 1 illustrates a glimpse of the myriad of possible combinations of strategies for sampling the model parameter space and SA methods, to quantify the impacts of sampled parameters on the model QOI. We understand that there is no combination of sampling and SA method that fits all applications. Thus, the work of Gan et al. (2014) shows that different sample strategies can even produce different outputs regarding the same SA method. Also, it seems that there isn't a clear relationship between the number of factors and the number of necessary runs to compute SA. Furthermore, in some cases, the number of runs used in the same sampling and SA method is not even in the same order of magnitude. For example, Pianosi et al. (2016) recommend $>1000 \times M$ model runs to calculate variance-based SA, such as FAST, where M is the number of input factors subject to SA. Neither Wang et al. (2013) nor Vanrolleghem et al. (2015) nor Gan et al. (2014) executed this many number of runs. The first used a sample of size 2049 for a 47-factor problem (instead of >47,000), while the second used a sample size of 3000 for a 17-factor problem (instead of >17,000). The third used a sample size of 2777 for a 13-factor problem (instead of >13,000). One could ask whether the number of runs should be based on something more than M.

In this manuscript, we will test different experimental strategies for a UQ workflow and discuss their relative benefits and limitations. A baseline scenario was developed, and we performed a comprehensive investigation of the impacts that differences in sample sizes, sample techniques, and SA methods may have on the QOI. In this work, we address the research question: how UA and SA may be applied to improve users' understanding of the uncertainty and relations among input and output responses in LUCC agent-based simulations? We are interested in finding which parameters are responsible for most of the results' variability; if there is convergence when different SA techniques are applied; and finally, if there is a minimum sample size to achieve it. Although the statistical techniques are applied in a specific agent-based simulator, the methods described are quite general and may illustrate their application in another research.

In Section 2, we provide an overview of the different methods regarding variance stability, parameter space exploration, UA, and SA. We also present the proposed UQ workflow in Section 2. In Section 3, we describe the MASE-BDI framework and LUCC model used as a case study, followed by the experimental design. We present the results compared to related work. We discuss challenges and provide some assessment to extrapolate our finding into more general conclusions, to produce more robust or parsimonious models, as well as to make models more defensible in the face of scientific or technical controversy (Section 4). Finally, in Section 5 , we summarize our findings and outline future work.

2. Materials and methods

The methods we applied in the case study are presented in this section alongside their experimental design. The UQ experiments have the objective to perform an output analysis on spatial stochastic models, to measure uncertainty and to reduce it. Ultimately, we want to understand better how the model behaves and expand our confidence in the response of a LUCC model.

2.1. Variance stability

Agent-based simulations are often stochastic, and therefore any analytical exercise requires an outcome pool drawn from a sufficient number of samples. It is only possible to draw conclusions if the output mean and variance reaches relative stability. Otherwise, the statistics could harbor too much uncertainty to be reliable (Lee et al., 2015). Moreover, some ABM simulations (MASE-BDI included) can take longer run times, which makes the execution of large samples prohibitive. Hence, knowing the minimum sample size to reach variance stability can be more compelling to modelers.

There are many methods to assess variance stability (Law and Kelton, 2000; Lee et al., 2015). We chose to apply the method proposed by Lorscheid et al. (2012), whose strategy is to assess stability from metrics on an outcome for a sequence of sample sizes. The proposed metric relies on the functional ratio between the variance and the sampled mean. The coefficient of variation c_V is a dimensionless and normalized metric used to measure the uncertainty surrounding the variance, i.e., used for the analysis of experimental error variance. It is