Environmental Modelling & Software 87 (2017) 84–109

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Variance-based sensitivity analysis of a wind risk model - Model behaviour and lessons for forest modelling

Tommaso Locatelli [a, b, *], Stefano Tarantola [c], Barry Gardiner [b], Genevieve Patenaude [a]

[a] School of GeoSciences, University of Edinburgh, Institute of Geography, Drummond Street, EH8 9XP Edinburgh, Scotland, United Kingdom
[b] INRA Unité Ephyse, 33882 Villenave d'Ornon Cedex, France
[c] Joint Research Centre, Via Enrico Fermi 2749, I - 21027 Ispra, VA, Italy

ARTICLE INFO

Article history:
Received 12 March 2016
Received in revised form
27 October 2016
Accepted 30 October 2016

Keywords:
Method of Soboĺ
Assessment of model performance
Copula method
Correlated variables

ABSTRACT

We submitted the semi-empirical, process-based wind-risk model ForestGALES to a variance-based sensitivity analysis using the method of Soboĺ for correlated variables proposed by Kucherenko et al. (2012). Our results show that ForestGALES is able to simulate very effectively the dynamics of wind damage to forest stands, as the model architecture reflects the significant influence of tree height, stocking density, dbh, and size of an upwind gap, on the calculations of the critical wind speeds of damage. These results highlight the importance of accurate knowledge of the values of these variables when calculating the risk of wind damage with ForestGALES. Conversely, rooting depth and soil type, i.e. the model input variables on which the empirical component of ForestGALES that describes the resistance to overturning is based, contribute only marginally to the variation in the outputs. We show that these two variables can confidently be fixed at a nominal value without significantly affecting the model's predictions. The variance-based method used in this study is equally sensitive to the accurate description of the probability distribution functions of the scrutinised variables, as it is to their correlation structure.

Software availability

Name of software: ForestGALES
Developers: Forest Research, INRA, and the University of Edinburgh
Contact address: Forest Research, Northern Research Station, Roslin, Midlothian EH25 9SY, United Kingdom Email: forestgales.support@forestry.gsi.gov.uk
Availability and Online Documentation: The software along with supporting material is freely available. Go to http://www.forestresearch.gov.uk/forestgales to find out how to obtain the software or email forestgales.support@forestry.gsi.gov.uk
Year first available: 2000
Hardware required: IBM compatible PC
Software required: MS Windows
Programming language: Borland Delphi 5.0®. Versions have also been written in Python, Fortran, R and Java. Contact Prof. Barry Gardiner (barry.gardiner@bordeaux.inra.fr) for further details. Contact the corresponding author (tom.locatelli@forestry.gsi.gov.uk) for information on the R version
Program size: 10 MB. With all additional support files and manuals: 25 MB. For free professional tools for sensitivity analysis please visit the European Commission Joint Research Centre sensitivity analysis page at https://ec.europa.eu/jrc/en/samo/simlab Please contact Dr. Stefano Tarantola (stefano.tarantola@jrc.ec.europa.eu) for information on the Matlab scripts of the Soboĺ method for the case of correlated variables

1. Introduction

Environmental modelling has become a crucial part of the study of environmental phenomena. Significant advances in the fields of hardware and computing now allow for the creation of complex, computationally-demanding, process-based models, aimed at the investigation of natural systems (e.g. Nossent et al., 2011). These complex models are extensively adopted in support of decision-

* Corresponding author. Current address: Forest Research, Northern Research Station, Roslin EH25 9SY, Scotland, United Kingdom.
E-mail address: tom.locatelli@forestry.gsi.gov.uk (T. Locatelli).

http://dx.doi.org/10.1016/j.envsoft.2016.10.010

making and for environmental policy settings (e.g. Rahmstorf et al. (2007) on IPCC projections). While a large amount of time and resources are spent to formalise nature in mathematical terms, considerably less effort is often made to investigate the behaviour of mathematical models, which is often done as an "afterthought" (Saltelli and Funtowicz, 2014). As elegantly discussed by Oreskes et al. (1994), the same practices of model validation, evaluation, and confirmation, are philosophical and practical minefields. Modellers are confronted with these issues for a number of reasons: natural systems, which are inherently open in nature, are forced into closed systems to obtain mathematical solutions; scaling issues can arise when the scales at which some elements of a model are calculated differ from the scale of application of the model; nonuniqueness of modelling approaches might result in a faulty model providing "reasonable" outputs (Oreskes et al., 1994). Ultimately, however, the main issue with environmental modelling is the same reason why models are built: we can never exactly know all the data, and those that we do know, we do so with a degree of uncertainty. With regards to the modelling process, in our paper we refer to uncertainty as incomplete knowledge of parameter values (Gaber et al., 2009). Deterministic approaches to modelling require elimination of these uncertainties, thus effectively further removing a model from its intended representation of reality. The inadequacy of the attempts to eliminate at all costs the uncertainties of the parameters and variables of a model, in order to produce completely deterministic results, is nowadays generally accepted (e.g. Penman et al., 2003). The transparency of model predictions is an important requirement especially when models are applied for decision-making, and in policy frameworks (e.g. the US Environmental Protection Agency, see Gaber et al. (2009)). To this end, uncertainty analysis is normally applied to quantify the uncertainties of the input variables, parameters, and outputs of a model, thus providing some insight on the reliability and the applicability range of the model.

On the other hand, the issue of sensitivity of model predictions to variation in model parameters and variables is still relatively underestimated. Quoting Saltelli et al. (2004), a sensitivity analysis is "*The study of how uncertainty in the output of a model (…) can be apportioned to different sources of uncertainty in the model input*". However, when performed appropriately (Saltelli and Annoni, 2010), sensitivity analysis (SA) of mathematical models is a tool that can help with fundamental issues about the robustness and the behaviour of a model (Tarantola et al., 2002; Norton, 2015). A number of techniques exist to perform sensitivity analysis (see https://ec.europa.eu/jrc/en/samo/methods). These can be broadly divided in two groups, typically referred to as "local" and "global", on the basis of the region of the input space that is scrutinised in the analysis. Local SA are normally based on derivatives of the output Y with respect to one factor X_i (e.g. $\delta Y/_{\delta X_i}$), where by factor here we denote either a model parameter or an input variable. These derivatives are often normalised by the input-output standard deviations (they are said to be sigma-normalised) to produce more robust sensitivity indices, as recommended by the Intergovernmental Panel on Climate Change in their guidelines on the inventories of greenhouse gases (IPCC, 1999; IPCC, 2000). However, with this approach only the base point where the derivatives are computed is investigated, which is an issue when the model is of unknown linearity (Saltelli et al., 2008). Local derivatives-based methods are mostly adopted within the context of one-at-a-time (OAT) approaches, where only one factor is perturbed while all the others are fixed at a nominal value (usually the mean). Therefore, the effects of factors interactions on the output variance are neglected with OAT methods, which are therefore only applicable for strictly additive models (Campolongo and Saltelli, 1997). Global SA (GSA) methods, on the other hand, allow for the exploration of the entire range of the factors, and for simultaneous perturbation of all the factors. The most powerful GSA methods are variance-based techniques that decompose the total variance of the output into conditional variances for single factors and for sets of factors. These techniques include the importance measures of Iman and Hora (1990) and of Sacks et al. (1989), the FAST (Fourier Amplitude Sensitivity Test) method (Cukier et al., 1973, 1978) and the extended FAST (Saltelli et al., 1999), and the method of Soboĺ (Soboĺ, 2001). The last two approaches can be solved numerically with Monte Carlo methods. Derivatives-based methods have been developed for global sensitivity measures (DGSM, e.g. Kucherenko et al., 2009; Soboĺ and Kucherenko, 2009). The values of DGSM is exactly equal to that of total sensitivity indices calculated with the Sobol' method (see section 2.2.1) in a number of cases, e.g. for linear models, while in a general case they correspond to the upper bound of the total Sobol' indices, with the advantage of a much shorter computational time. Variance-based GSA methods have a number of advantages: they are model-independent; they can capture the influence of the full range of variation of each input variable; they allow for the investigation of interaction effects amongst variables; and they provide the possibility of grouping factors (Saltelli et al., 2008). Their drawback is the high computational cost required for performing such techniques, due to the large number of model executions required for the convergence of the values of the sensitivity indices (Kucherenko et al., 2012). For this reason, a large body of research has been devoted to devise efficient algorithms for their computation (e.g. Kucherenko et al., 2012; Mara and Tarantola, 2012; Most, 2012; Saltelli, 2002).

Of the aforementioned variance-based GSA techniques, the method of Soboĺ has found favour with modellers in the environmental sciences, because of the relatively straightforward interpretation of the sensitivity indices calculated with this method, and because it very efficiently samples the factors space (Soboĺ, 1990; Yang, 2011; Kucherenko et al., 2015). The Soboĺ method is often used as a benchmark against which to compare the results of other SA techniques (Confalonieri et al., 2010). In a previous issue of this journal, Nossent et al. (2011) successfully applied the Soboĺ method to the identification of the most, and the least, important factors in a SWAT model (Soil and Water Assessment Tool). The authors also provided an exhaustive description of the Monte Carlo procedures required for the calculation of the Soboĺ sensitivity indices. Song et al. (2012) used the method of Soboĺ for the SA of the 3-PG2 forest growth model, aimed at model calibration. A known issue with variance-based GSA techniques is how to account for correlation between factors when calculating the conditional variances. Indeed, correlation amongst factors in environmental models is typical. A number of studies propose methods to obviate the issue of dependent factors in GSA (e.g. Mara and Tarantola, 2012; Most, 2012).

In this paper, we submit ForestGALES, a forest wind-risk model, to a variance-based GSA using the method of Kucherenko et al. (2012), a generalisation of the method of Soboĺ for correlated factors. The rationale of ForestGALES, together with the most important model calculations for the context of our GSA, is discussed in the Methods section. For a thorough description of the model, the interested reader is referred to Hale et al. (2015), published in a previous issue of this journal. Variance-based GSA are normally applied to complex models composed of a large number of factors, sometimes in excess of one hundred, mostly for the direct benefit of the modelling community. In this paper, we limit our GSA to the inputs of ForestGALES that are controllable by the end-users. Focussing on those input variables that are user-modifiable extends the benefits of a GSA to the end-user base of an environmental model, and facilitates the interpretation of the results of the SA in a practical setting. To extend the results of our GSA to a large