

SIMENV - A FLEXIBLE FRAMEWORK FOR SENSITIVITY AND UNCERTAINTY ANALYSES OF LARGE-VOLUME MODEL OUTPUT

M. Flechsig*, U. Böhm, T. Nocke, C. Rachimow
Potsdam Institute for Climate Impact Research, Germany
flechsig@pik-potsdam.de

Modelling and simulation are essential techniques in global change research. One of the challenges in this research field is the development of integrated models which is being achieved mainly by combining models for individual systems and fostered by the availability of increased computing power. Such integrated simulation models are often based on legacy source code applications written in a programming language rather than in a model design language. They typically produce large amount of multi-dimensional model output that has to be handled in the course of a model analysis. These aspects hamper the application of quality assurance techniques since source code is not always well known by model users and intensive code manipulations are normally beyond the scope of the work. Additionally, the computational costs for models in global change research are often very high which demands structured experimentation approaches. Dealing with uncertainty and communicating it to decision makers and the general public is crucial in climate change research [1], [2]. For instance, increasing computing power enabled sensitivity and uncertainty studies of climate projections [3], [4].

For these purposes, a structured and integrated approach for simulation model verification, validation as well as sensitivity and uncertainty analyses is required. This comprises planning experiments, distributing simulation load, post-processing experiment output, and visualizing analysis results. Typically, existing tools supply these functionalities only partially. Furthermore, their interfaces to the model potentially imply limitations concerning usability and flexibility. There is a demand from the modeller's perspective to manipulate a model as less as necessary. Finally, modellers and model users often need guidance in sensitivity and uncertainty related aspects of modelling and simulation.

To handle these problems we developed the simulation environment SimEnv [5] as a flexible framework for sample-based model evaluation that integrates most of the standard techniques normally used by non-expert users. It focuses on sensitivity and uncertainty analyses. SimEnv offers a well-structured and user-friendly approach to interface models, to define and run experiments, to post-process them, to derive quantitative measures, and to qualitatively evaluate them by interactive visualization techniques. It is designed to cope with multi-dimensional, large-volume data in all of these processing steps. Access to the system is facilitated by easy-to-use interfaces.

Experiment definition: The challenge for experiment definition from the developer's point of view is to cover a broad range of tasks for sensitivity, uncertainty and error analyses, numerical verification and validation as well as control design. As a solution we designed pre-formed experiment type templates representing different sampling plans for a subset of model factors (formed from parameters, initial and/or boundary values of the model). SimEnv supplies the following six experiment types: (1) a flexible deterministic screening of the model's behaviour in high-dimensional factor spaces, (2) a global sensitivity analysis based on the Morris method [6], (3) a local sensitivity analysis in the vicinity of the default factor values, (4) Monte Carlo experiments with built-in probability density functions for random and Latin hypercube sampling, (5) an uncertainty analysis to determine total and first order effects of factors on model output by the Sobol' method [7], [8], and (6) an optimization technique based on simulated annealing [9] to minimize a cost function derived from model output over the factor space.

Model interface: Crucial for the applicability of a model experiment framework is a flexible interface to plug in models easily. Thus we designed a model interface that allows with only minimal model modifications (1) to transfer the sample factor values into the model, and (2) to store model variables as SimEnv experiment output for later post-processing. Model interfacing is done at the model language level for C/C++, Fortran and Python by incorporating SimEnv interface function calls into model source code. The basic approach is to include per experiment factor and per model output variable field a SimEnv function call. Additionally, there are special interfaces for Matlab, Mathematica and GAMS models, for shell script models, and an interface to directly transform native ASCII model output to SimEnv experiment output.

Load distribution: Hardware resources are the main bottleneck for studies of large ensemble simulations. Factor values are sampled during experiment preparation which allows performing the single runs of an experiment in parallel. This is supported by SimEnv on workstation clusters by a single run distribution mechanism, applying the Message Passing Interface.

Post-processing: Analysis of large-volume experiment output requires efficient strategies for navigating the high-dimensional factor space and for processing / reducing the model output data. Therefore, SimEnv provides an interactive post-processor with more than 100 filters / operators. Secondary model output functions can be derived by applying chains of general-purpose operators to model output and reference data. Run ensemble

related aggregated measures are calculated by experiment-specific operators. SimEnv supplies built-in operators and an interface for user-defined operators.

Visual evaluation: A further challenge is to analyse and communicate multi-dimensional experiment output and measures. Therefore, a tailored interactive visualization module has been developed and integrated. To handle the diversity of experiment data, this module provides both methods for spatial data sets (graphs, maps and 3D displays) and for data sets with abstract experiment dimensions (scatter plot matrices, parallel coordinates, histogram displays). This diversity of techniques with a variety of interaction mechanisms enables model evaluators to get deeper insight into the behaviour of their models and to communicate the results. However, a problem comes along with the number of visualization techniques and their numerous parameters. Thus, to help non-expert users to generate expressive and effective graphical representations, a mechanism to design appropriate visualizations that is based on data characteristics and user goals has been integrated (see [10] for an overview).

For most of its components, SimEnv offers data interfaces to NetCDF, IEEE compliant binary, and ASCII format. Its strength is to efficiently handle multi-dimensional, large-volume output fields. Due to the flexible and modular framework design, additional sampling strategies / experiment types can be integrated easily into the environment by implementing the corresponding modules for experiment preparation, post-processing, and visualization. The system has proved its applicability in a number of sensitivity and uncertainty case studies, ranging from conceptual qualitative models [11] to large scale climate models [4]. In the paper we will illustrate the capabilities of the system by a practical study.

References

- [1] Willows RI, Connell RK (Eds.): Climate adaptations: Risk, uncertainty and decision making. UKCIP Technical Report, UKCIP, Oxford, 2003.
- [2] Webster M: Communicating climate change uncertainty to policy-makers and the public. *Climatic Change* **61**, 1-8 (2003)
- [3] Stainforth DA, Aina T, Christensen C, Collins M, Faull N, Frame DJ, Kettleborough JA, Knight S, Martin A, Murphy JM, Piani C, Sexton D, Smith LA, Spicer RA, Thorpe AJ, Allen MR: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* **433**, 403-406 (2005), doi: 10.1038/nature03301
- [4] Schneider von Deimling T, Held H, Ganopolski A, Rahmstorf S: Climate sensitivity estimated from ensemble simulations of glacial climate. *Climate Dynamics* **27**, 149-163 (2006), doi: 10.1007/s00382-006-0126-8
- [5] Flechsig M, Böhm U, Nocke T, Rachimow C: The multi-run simulation environment SimEnv. User's Guide. Potsdam Institute for Climate Impact Research, Potsdam, Germany, 2006, <http://www.pik-potsdam.de/software/simenv>
- [6] Morris, MD: Factorial plans for preliminary computational experiments. *Technometrics* **33(2)**, 161-174 (1991)
- [7] Sobol' IM: Sensitivity estimates for nonlinear mathematical models. *Math. Modelling & Comp. Exp.* **1**, 407-414 (1990)
- [8] Saltelli A, Tarantola S, Campolongo F, Ratto M: Sensitivity analysis in practice: A guide to assessing scientific models. J. Wiley & Sons, Chichester, 2004
- [9] Ingber L: Adaptive simulated annealing (ASA): Lessons learned. *Control and Cybernetics* **25(1)**, 33-54 (1996)
- [10] Lange S, Nocke T, Schumann H: Visualisierungsdesign - ein systematischer Überblick (in German). Proc. Simulation and Visualization (SimVis'06), Magdeburg, Germany, 2006
- [11] Knopf B, Flechsig M, Zickfeld K: (2006): Multi-parameter uncertainty analysis of a bifurcation point. *Nonlin. Processes Geophys.* **13**, 531-540 (2006)