

Multi-Run Simulation Environments – Requirements, Development & Application

Michael Flechsig, Thomas Nocke, Claus Rachimow
Potsdam Institute for Climate Impact Research, 14473 Potsdam, Germany
flechsig@pik-potsdam.de



ABSTRACT: Model development and application as well as data analysis and integration are main methodologies in climate impact research. In recent years, increasing computing capacity has advanced the evaluation of simulation models by uncertainty and sensitivity quantification. To bring related techniques at the desk of modelers and analysts easy to apply simulation environments are an essential prerequisite. We present with SimEnv such an environment that is used for sensitivity, uncertainty and scenario analyses.

REQUIREMENTS

Simulation environments (SE) [1] (Fig. 1) are software tools to support experimenting with simulation models, normally developed outside the environment.

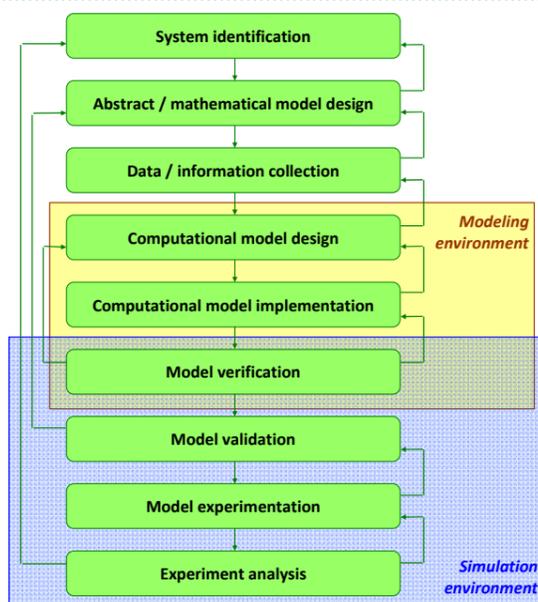


Fig. 1 Model life cycle, modeling and simulation environment

Important features of a SE for sensitivity and uncertainty analyses (SUA) are

- Corresponding experiment designs / sampling methods in factor spaces,
- Model interface to import factor values and to export model output,
- Experiment load distribution on computer networks,
- Experiment analysis to derive measures, and
- Coupled visualization systems for explorative visual analysis.

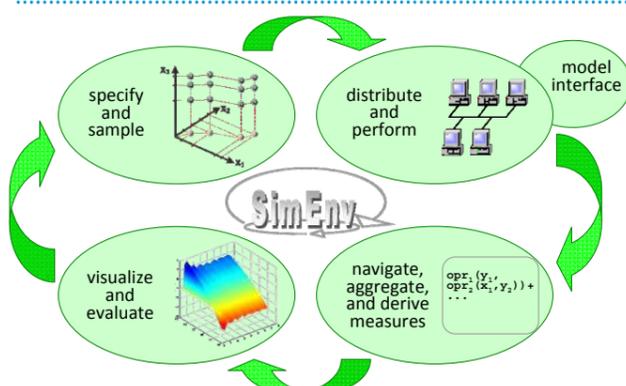


Fig. 2 SimEnv experiment workflow

DEVELOPMENT

SimEnv [2] (Fig. 2) is a multi-run SE mainly for SUA. Pre-defined experiment types (Tab. 1, Box 1) representing model-free SUA methods are based on probabilistic, deterministic or Bayesian sampling schemes in high-dimensional (100+) factor spaces and allow for flexible experimentation and analyses afterwards.

experiment type	sampling scheme	global	correl. factors	comput. costs (k factors)
Factor Ranking (modif. Morris method)		✓		$\sim 10 \cdot (k+1)$
Variance Decomposition (Sobol' method)		✓		$N_{MC} \cdot (k+2)$
Monte Carlo Analysis		✓	✓	N_{MC}
Local Sensitivity Analysis (linearity, symmetry ...)			(✓)	$2 \cdot k$
Determ. Factorial Design (full/fractional)		✓	✓	
Bayesian Calibration (Metropolis algorithm)		✓	✓	$\gg N_{MC}$
Optimization (simulated annealing)		✓	(✓)	

Tab. 1 SimEnv experiment types

Features:

- Large-volume (GB range), multi-dimensional (≤ 9) / multi-variate (100+) model output.
- Simple interface for C/C++, Fortran, Python, Java, Matlab, Mathematica, GAMS models and shell scripts:
Include for each factor and for each model output field one SimEnv function call into model source code.
- Flexible load distribution strategies for the single runs of the experiment on a computer network or on a multi-core machine.
- Post-processor to derive interactively output functions / measures in NetCDF or IEEE compliant format by chains of built-in / user-defined operators.
- Coupled visualization framework SimEnvVis [3] (Figs. 3 and 4).

```

general      type      UNC_MC      # experiment type Monte Carlo
factor p1   descr    this is ...  # factor description
factor p1   unit     kg/m**2     # factor unit
factor p1   default  1           # factor default value
factor p1   type     add         # add sampled values to default
factor p1   sample  distr N(1,0.04) # p1 ~ N(1,0.04)
factor p1   sample_method pseudo      # use pseudo RN generator
factor p1   sample_include 0.4:1.6    # sample only within 3σ interval

factor x2   ...
factor x2   sample  distr U(-5,5) # x2 ~ U(-5,5)
factor x2   sample_method stratified # use Latin hypercube sampling

factor par3 ...
factor par3 sample  file MySample.dat # import sample from file
...

specific    runs      2000      # number of Monte Carlo runs
specific    function sum(sin(y1+3*p1)-avg(y2)) # model output function
                                                    # for optional stopping rule

blue descriptors = mandatory elements
    
```

Box 1 Example of a SimEnv experiment description file

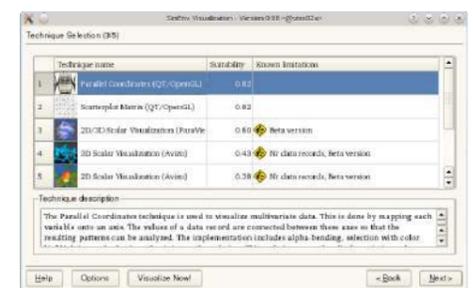


Fig. 3 The visualization wizard SimEnvVis: technique selection

APPLICATION

SimEnv has been applied in climate impact research for many problem settings. Studies with about 20 models range from scenario analyses of a global economy – energy model [4], uncertainty and sensitivity analyses of a model of the Indian monsoon [5] to Bayesian calibration of a forest growth model [6].

Preferred experiment types of 10.600 experiments performed so far by more than 30 users are multi-factor deterministic design (67%), Monte Carlo analysis (19%) and factor ranking (5%). On average per experiment (i) 8 factors were used resulting in 290 single runs, (ii) 11 variables were output with 51.000 values per single run and (iii) an experiment was distributed on 56 processor cores.

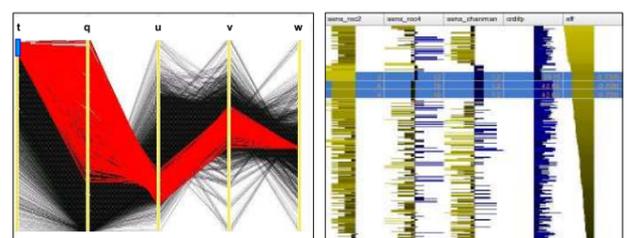


Fig. 4 Ensemble data visualization examples from SimEnvVis left: Parallel coordinates for a Monte Carlo analysis – temperature, humidity and wind components of an Earth system model of intermediate complexity right: Graphical table for a deterministic design – three factors (sens_*) and two aggregated outputs of a hydrological model

REFERENCES

- [1] J.P.C. Kleijnen, S.M. Sanchez, T.W. Lucas, T.M. Cioppa: State-of-the-art review: A user's guide to the brave new world of designing simulation experiments. INFORMS Journal on Computing, 17(3): 263-289, 2005.
- [2] M. Flechsig, U. Böhm, T. Nocke, C. Rachimow: The multi-run simulation environment SimEnv. User's Guide V3.1. PIK Potsdam, 2013.
- [3] T. Nocke, M. Flechsig, U. Böhm: Visual exploration and evaluation of climate-related simulation data. Winter Simulation Conference WSC'07 Washington D.C., 2007.
- [4] G. Luderer et al.: Description of the Remind model V1.5. PIK Potsdam, 2013.
- [5] B. Knopf, M. Flechsig, K. Zickfeld: Multi-parameter uncertainty analysis of a bifurcation point. Nonlin. Processes Geophys., 13: 531-540, 2006.
- [6] M. van Oijen et al.: Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe. Forest Ecol. and Managem., 289: 255-268, 2013.