The Multi-Run Simulation Environment SimEnv

M. Flechsig, U. Böhm (*), T. Nocke, C. Rachimow PIK – Potsdam Institute for Climate Impact Research (*) DWD Berlin

General SimEnv Approach

SimEnv is a sampling-based simulation environment for sensitivity, uncertainty and scenario analyses. It aims to quantify uncertainty and identify sensitive factors or processes of a simulation model **M**.

M: $y = F(X_k)$
$X_k = (x1,...,x_k)$ input factor vector (parameters, initial values)
y large volume multi-variate / -dimensional model output

Experiment Types

Pre-formed experiment type templates represent probabilistic, deterministic or Bayesian sampling strategies in X_k. They are equipped with numerical information to generate a sample. The result is a multi-run simulation experiment with the model for the sample.

{x} sample set ○ default factor value of **M** in X_2

Factor ranking	Qualitative ranking of a large number of factors with respect to their sensitivity on model output at random trajectories in X_k For determination of most important factors to focus on afterwards
Variance Decomposition	Orthogonal variance decomposition of model output to first order and total effects of factors by a Monte Carlo re-sampling {+} For - uncertainty analysis - model validation
Monte Carlo Analysis	Probabilistic (pseudo, quasi, stratified random) marginal factor sampling and determination of statistical measures For - error analysis - model validation
Local Sensitivity Analysis	Sampling in a local neighbourhood of the default factor values For local first order sensitivity measures by investigating finite difference approximations of derivatives
Determ. Factorial Design	Deterministic inspection with a flexible screening strategy in X_k For - one-factor-at-a-time experiments - (fractional) factorial experiments - response surface methodology
Bayesian Calibration	Qualified posterior distribution of X_k in terms of a representative sample by additional data from the system For - model parametrization / calibration - Bayesian model calibration
Optimization	Stochastic sampling to find the global minimum of a cost function on X_k applying simulated annealing (ASA) For - model validation - control design

Model Interface

It is based on minimal source code modifications for C/C++, Java, Fortran, Python, Matlab, Mathematica, and GAMS models, at shell level and for ASCII files by implementing a SimEnv function call

- simenv_get to get in **M** a sampled factor value x_i from SimEnv
- simenv_put to put model output field y from **M** to SimEnv

SimEnv experiment output is stored in self-describing Network Common Data Form NetCDF or IEEE compliant binary format.

Experiment Load Management

- Sequential on a local machine
- Distributed on a multi-core machine
- Parallel on a compute cluster, using MPI
- Optional partial experiments, experiment restart

Post-Processing and Visual Evaluation

Interactive post-processing allows

- to compute secondary output functions from model output y, reference data and other SimEnv experiments by applying chains of elemental/ selective/ analytical/ and statistical operators
- to navigate X_k and derive uncertainty and sensitivity measures for output functions over the run ensemble space by applying experiment type-specific operators

Currently, 100+ built-in operators are available. There is an interface to plug user-defined operators into the environment.

Analysis and evaluation of post-processed output and derived measures from experiment output benefit from the coupled visual analytics framework SimEnvVis.

Prospects

- Single factor experiments as multi-factor experiment types
- Multi-file support for very large experiment output
- Model interface for R

References

SimEnvVis	http://www.pik-potsdam.de/software/simenv	MPI	http://www.mpi-forum.org
CLM	http://clm-community.eu	GAMS	http://www.gams.com
NetCDF	http://www.unidata.ucar.edu/packages/netcdf	ASA	http://ingber.com/#ASA

Example

CCLM – regional climate model
Baltic Sea and Northern / Central Europe
Space: 0.5° lat x 0.5° lon x 20 vertical layers
Time: 6 hourly model output
Parametrization of the soil submodel:
Dependency of latent and sensible heat fluxes lhf and shf from soil in a X_2 = (crsmin , Tend)

Model output variable description file

coordinate	lat	values	35 (0.5) 67	# defines coord. latitude with ½° resolution
coordinate	lon	values	-25 (0.5) 40	# defines coord. longitude
coordinate	time	values	1 (1) 28	# defines coord. time (6 hourly time steps)
variable	lhf	coords	lat , lon , time	# defines lhf as lhf(lat,lon,time)
variable	shf	coords	lat , lon , time	# defines shf as shf(lat,lon,time)

Experiment Determ. Factorial Design: Experiment description file

factor	crsmin	sample	30 (5) 120	# specifies 19 sampled values for crsmin
factor	crsmin	default	60.	# default model value of crsmin
factor	crsmin	type	set	# directly use sampled values
factor	Tend	sample	273.1 (5) 333.1	
factor	Tend	default	313.15	
specific		comb	crsmin*Tend	# factorial screening: 19*13+1=248 runs

Post-processor

Fig. 1:	dfd(' ', avg(shf)) - run('default', avg(shf))	# area and temporal averaged shf bias
Fig. 2:	dfd('sel_s(Tend=313.15)', avg_l('time', shf)) - run('default', avg_l('time', shf))	# area averaged shf bias for each time step, # all crsmin, and the default value of Tend

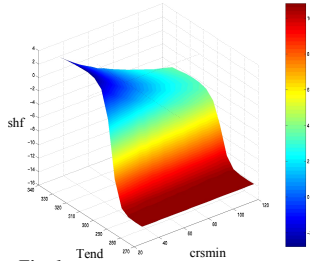


Fig. 1

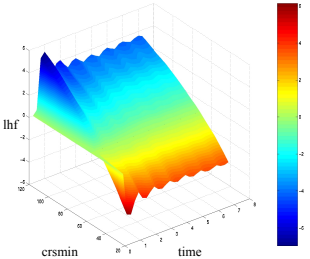


Fig. 2

Experiment Monte Carlo analysis

crsmin , Tend ~ N(def_val,20.)	Latin hypercube sampling	150 runs

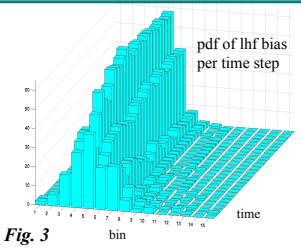


Fig. 3

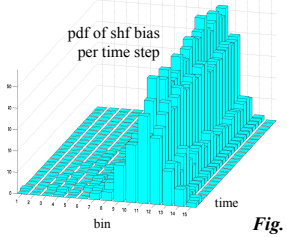
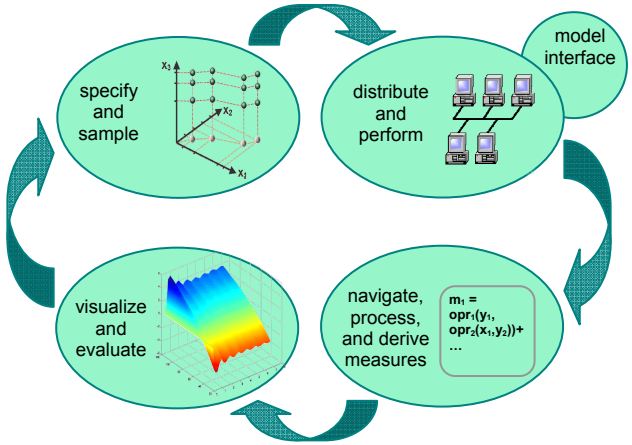


Fig. 4

SimEnv Applications at PIK

Model	Aim	Model interface
4C	Factorial designs, Bayesian calibration	Fortran
Aeolus	Factorial designs	C++
Climber-2	Monte Carlo and uncertainty analyses	Fortran
CLM	parameter screening, Monte Carlo analyses	Fortran, script level
Lagom (at GCF)	model validation, global sensitivity analysis	Java, Matlab, Mathematica,
LPJ	global / regional model applications	C, script level
REMIND / MAgPIE	parameter screening, uncertainty analyses	GAMS
Monsoon	sensitivity and Monte Carlo analyses, optimization	C, Fortran

SimEnv General Workflow



System Requirements

Component	Minimal specification
hardware	Intel-based systems and compatibles with a 32-bit processor i386
operating system	SUSE V 9.0
shell	Bourne shell
C/C++ compiler	gcc V 3.3
Fortran compiler	ifort V 10.0 or gfortran V 4.2
Python	V 2.3
NetCDF	V 3.6.0
OpenDX	V 4.4.4
Qt	V 3.3.5
MPI	V 1.0
Java (*)	V 1.4
Matlab (*)	V 7.7
Mathematica (*)	V 4.1
GAMS (*)	Distr. 20

(*): only for running a corresponding interfaced model

Contact

Potsdam Institute for Climate Impact Research
Telegrafenberg
D-14473 Potsdam
http://www.pik-potsdam.de

Michael Flechsig
Phone: +49 – 331 – 288-2604
Mail: flechsig@pik-potsdam.de
Web: http://www.pik-potsdam.de/software/simenv

Multi-Run Simulation Environment

http://www.pik-potsdam.de/software/simenv

SimEnv is
a multi-run simulation environment
that addresses
the evaluation and usage of models
with large and multi-dimensional output
mainly for uncertainty, sensitivity
and scenario analyses
applying pre-formed sampling strategies
in model parameter / initial value spaces.

Potsdam Institute
for Climate Impact Research