

The Multi-Run Simulation Environment SimEnv

or: Would you trust an orthopaedist who never x-rays?



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

Michael Flechsig
Uwe Böhm
Thomas Nocke
Claus Rachimow

PIK – Potsdam Institute for Climate Impact Research, Potsdam, Germany
DWD – German National Meteorological Service
PIK
PIK

SimEnv is . . .

- a multi-run simulation environment
- that addresses the evaluation and usage of simulation models
- mainly for uncertainty, sensitivity and scenario analyses
- applying pre-formed probabilistic, deterministic and Bayesian sampling strategies
- in high-dimensional (100+) parameter / initial value spaces of simulation models
- with large volume (GB) of multi-variate (100+) and multi-dimensional (≤ 9) model output

Motivation: Global Change Research

- Model development and application as key working techniques
- Complexity of the Earth system, intrinsic variability of processes, limits of knowledge
- Growing demand for uncertainty and sensitivity measures in Earth system models
- Need for communicating uncertainty to the community, stakeholders and the public

Nature, 486, 183-184 (2012)

Climate models at their limit?

Estimates of climate-change impacts will get less, rather than more, certain. But this should not excuse inaction, say Mark Maslin and Patrick Austin.

Nature Climate Change, 3, 769-771 (2013)

Uncertainty analysis in climate change assessments

Richard W. Katz, Peter F. Craigmile, Peter Guttorp, Murali Haran, Bruno Sansó and Michael L. Stein
Use of state-of-the-art statistical methods could substantially improve the quantification of uncertainty in assessments of climate change.

Nature Climate Change, 1, 198-200 (2011)

Opportunities from uncertainty

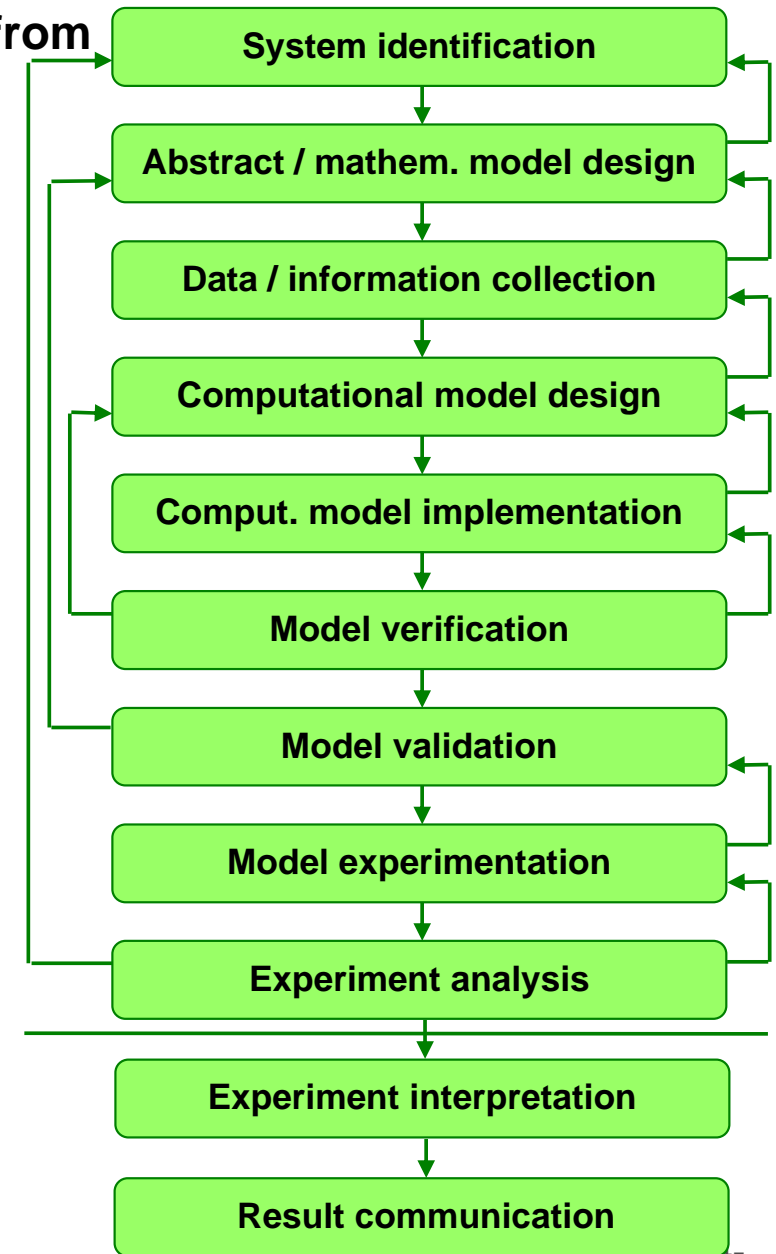
The inability to verify nations' reported progress towards emission-reduction commitments is a stumbling block in climate change negotiations. Narrowing uncertainties in the global carbon cycle could help overcome this obstacle.

Gary W. Yohe

Model Uncertainties

Uncertainties in modelling and experimentation arise from

- Framing / delimitation of the modelled system's part
- Inherent properties of the real system
- Incomplete knowledge about the real system
- Mathematical model
- Model implementation
(discretization schemes, numerical algorithms, ...)
- Model parametrization and parameter variability
- Machine accuracy (rounding errors, ...)
- Experiment settings (scenarios, drivers)
- Multi-model variability (e.g., CMIPs, ISI-MIP)
- Interpretation of model results
- Communication of experiment findings



Sensitivity & Uncertainty Analyses

Sensitivity Analysis SA

(A. Saltelli, JRC Ispra)

The study of how the uncertainty in the model output / states can be apportioned to different sources of uncertainty in the model input

Uncertainty Analysis UA

(P. Janssen, RIVM)

The study of the uncertain aspects of a model and of their influence on the (uncertainty of the) model output / states

- SA is more straightforward: it is on uncertainty **decomposition**
- UA is more general: it is on uncertainty **propagation & quantification**

Consider

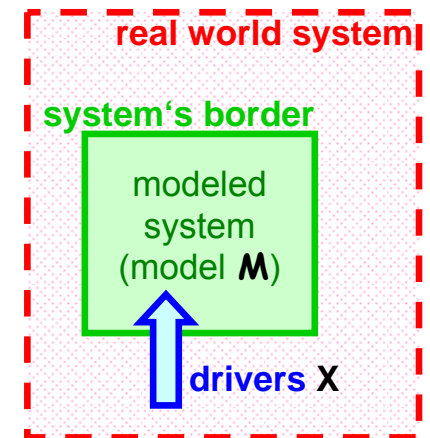
$$\left. \begin{aligned} \mathbf{M}: \quad \{Z_t\} &= F(\{Z_{t-\Delta t}\}, \{Z_0\}, \{P\}, \{X_t\}) \\ \{Y_t\} &= f(\{Z_t\}) \end{aligned} \right\} \text{(example)}$$

as

$$\mathbf{M}: \quad Y = g(\mathbf{x}) \quad \mathbf{x} = (x_1, x_2, \dots, x_k) \subset \{P, Z_0\}$$

= k **input factors**: subset of parameters $\{P\}$ and initial values $\{Z_0\}$

Y = model output / states Z



Model vs. SA Method Characteristics

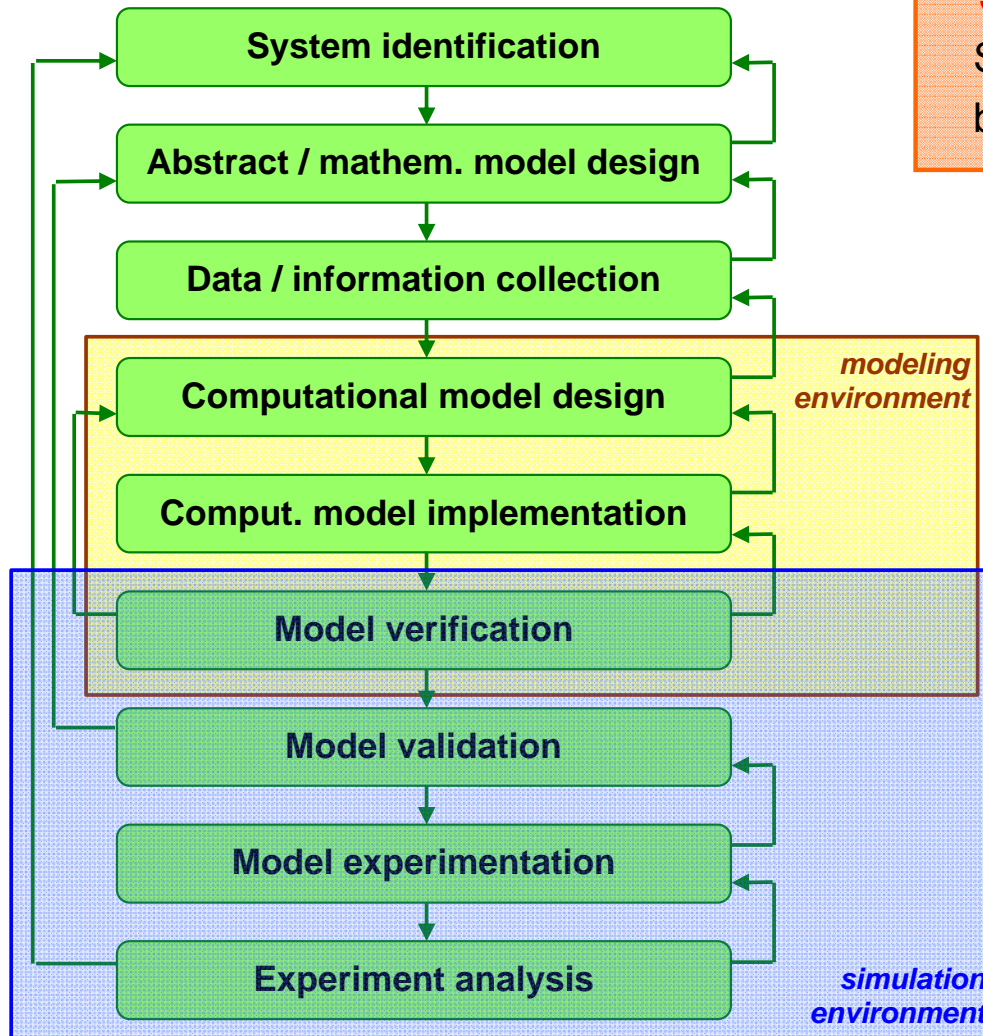
Models

- (M1) Complex phenomena (nonlinear, threshold effects, strong interactions)
- (M2) Large number of uncertain factors
- (M3) High computational costs (often) in terms of CPU time consumption
- (M4) Large volume of multi-dimensional model output of interest

SA Methods (A. Saltelli)

- (M1) → Model independency
 - ┆ Applicable to non-linear / non-additive / non-monotonic models
 - (M1) → Factor variability
 - ┆ Consider the whole range of factor variation e.g., by a pdf
 - (M1) → Multi-factor measures
 - ┆ Simultaneously take into account variation of all factors
 - (M2) → Factor grouping
 - ┆ Treat grouped factors as if they were single factors
 - (M3) → Cost efficiency
 - ┆ In terms of the number of single simulation runs
- model free
- } → global method
- computational costs C

Simulation Environments



Simulation environment =
 Software tool to support experimenting with models
 by a structured approach

Model \mathbf{M} : $y = f(\mathbf{x})$

- Typical questions (settings)
 - How does a factor x influence y ?
 - What are the most sensitive factors on y ?
 - What is the factor value that minimizes y ?

- Experiment steps to answer a question about \mathbf{M}
 - Design
 - Perform
 - Analyze

Requirements for Simul. Environments

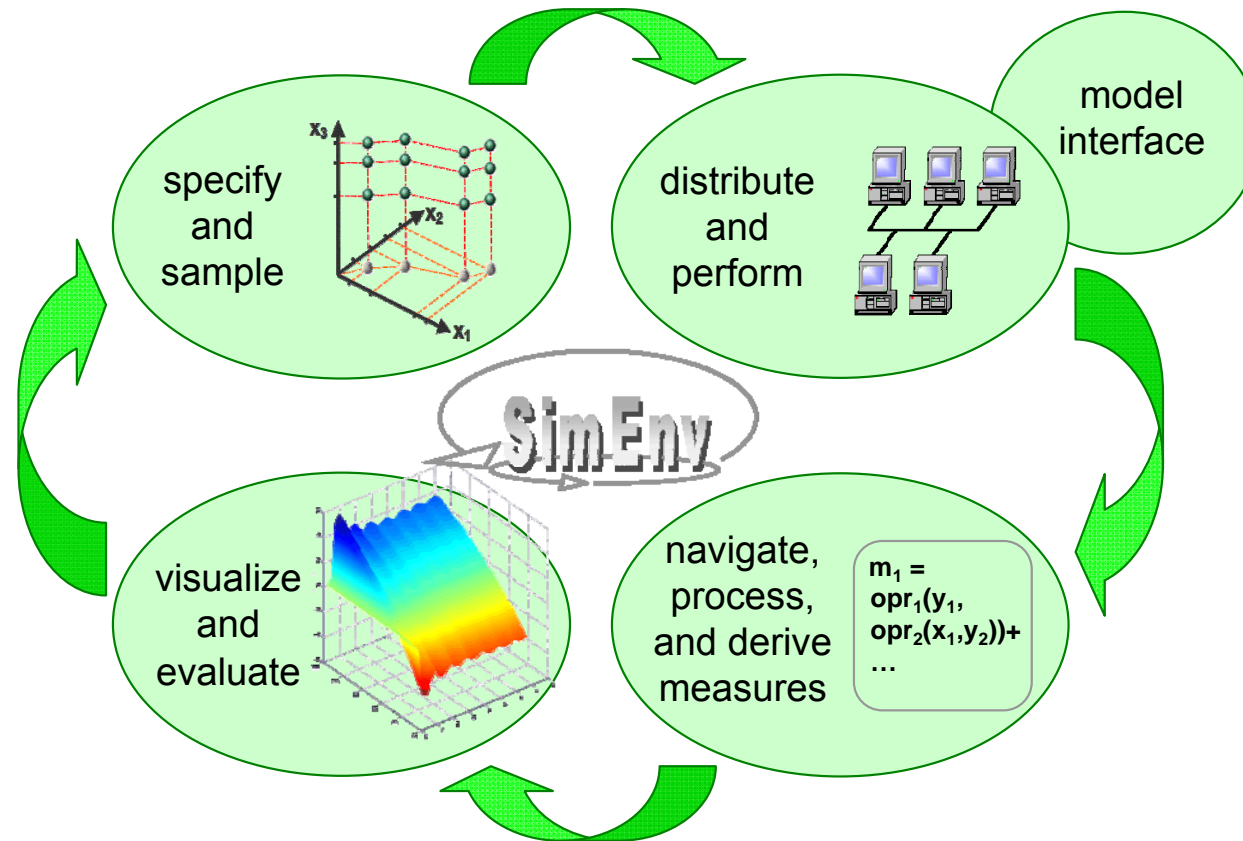
- Open architecture
 - Model interface
 - Import sampled factor values into model (*)
 - Export model output directly from the model to the environment or transform at least native model output to data structures of the simulation environment
- Load distribution (*)
 - Parallelize the experiment if sample can be drawn before the experiment
- Experiment analysis
 - (i) Single run related aggregation and transformation operators for the model state space
 - (ii) Method-specific operators to derive measures over the run ensemble / at the experiment factor space (*)
 - Apply explorative visualization techniques to experiment output and measures
- Data interface
 - Import reference data into simulation environment
 - Export data from experiment analysis

(*) = for multi-run environments

SimEnv: The Basics

- Focus on sample-based experiment techniques for uncertainty, sensitivity and scenario analyses
- Support of high-dimensional factor spaces and models with large volume of multi-variate / multi-dimensional output
- Simple model interface for factors and model output
 - **“include for each factor and for each model output field one SimEnv function call into model source code”**
 - at programming language level: C/C++ Fortran Java Python
 - at modeling / tool language level: Matlab Mathematica GAMS R (in prep.)
 - at shell script level and for native model ASCII file output
- Storage of model output of such interfaced models and of analysis results in self-describing NetCDF or IEEE compliant binary format
- Distribution of the single runs of an experiment at a compute cluster (using MPI) or at a multi-core machine
- Interactive experiment postprocessor with 100+ operators and coupled visualization framework SimEnvVis

Basic SimEnv Workflows



SimEnv Built-in Experiment Types

Experiment type	global	model-free	correlated factors	computational costs (k factors)
Qualitative Factor Ranking (Morris method)	✓	✓		$C \approx 10*(k+1)$
Variance Decomposition (Sobol' method)	✓	✓		$C = N_{MC}*(k+2)$
Monte Carlo Analysis	✓	✓	✓	$C = N_{MC}$
Local Sensitivity Analysis (linearity, symmetry ...)		✓		$C = 2*k$
Deterministic Factorial Design (OAT, full/fractional)	(✓)	✓	✓	$C =$
Bayesian Calibration (Metropolis algorithm)	✓	✓	✓	$C \gg N_{MC}$
Optimization (simulated annealing)	✓	✓	(✓)	$C =$

Example: Model Interface

model source code

```

...
real*8 y(180,360,100)
x1 = 5.

x2 = 4.

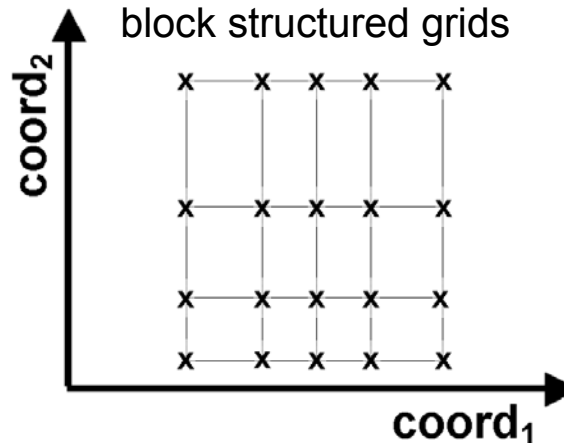
... (compute model output y)
istatus = simenv_put_f('y',y)
...
    
```

SimEnv model output description

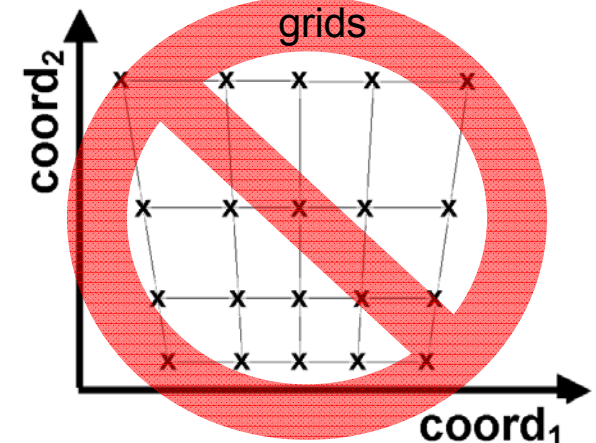
coordinate	lat	descr	geographic latitude
coordinate	lat	values	equidist_end 89.5 (-1) -89.5
coordinate	lon	values	equidist_nmb -179.5 (1) 360
coordinate	time	values	list 1,2,3, ..., 98,99,100
variable	y	descr	...
variable	y	type	real*8
variable	y	coords	lat , lon , time
variable	y	index_ext	1:180 , 1:360 , 1:100

generate grid

variables on rectilinear /
block structured grids



variables on curvilinear
grids



Example: Model Interface

Modified model source code

```

...
real*8 y(180,360,100)
x1 = 5.
x2 = 4.
... (compute model output y)
istatus = simenv_put_f(y,y)
...
    
```

SimEnv model output description

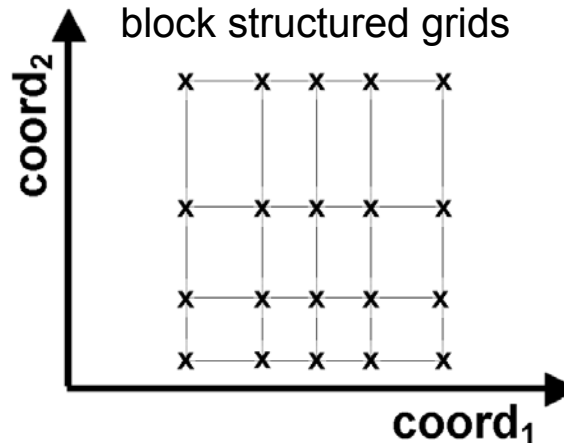
coordinate	lat	descr	geographic latitude
coordinate	lat	values	equidist_end 89.5 (-1) -89.5
coordinate	lon	values	equidist_nmb -179.5 (1) 360
coordinate	time	values	list 1,2,3, ..., 98,99,100
variable	y	descr	...
variable	y	type	real*8
variable	y	coords	lat , lon , time
variable	y	index_ext	1:180 , 1:360 , 1:100

— + — :

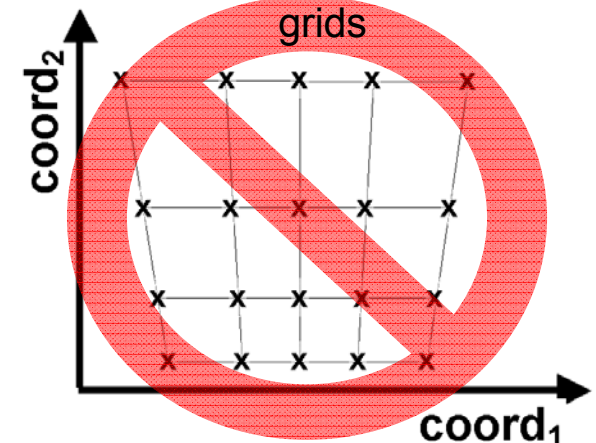
associate
 SimEnv model output
 description variable y
 with
 model source code variable y
 and
 output real*8 field y

generate grid

variables on rectilinear /
 block structured grids



variables on curvilinear
 grids



Example 1: Experiment Description

Modified model source code

```

...
real*8 y(180,360,100)
x1 = 5.
istatus = simenv_get_f('x1',x1,x1)
x2 = 4.
istatus = simenv_get_f('x2',x2,x2)
... (compute model output y)
istatus = simenv_put_f('y',y)
...
    
```

SimEnv model output description

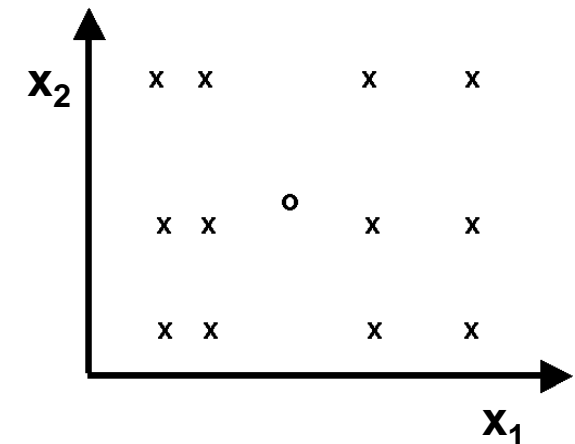
coordinate	lat	descr	geographic latitude
coordinate	lat	values	equidist_end 89.5 (-1) -89.5
coordinate	lon	values	equidist_nmb -179.5 (1) 360
coordinate	time	values	list 1,2,3, ... ,98,99,100
variable	y	descr	...
variable	y	type	real*8
variable	y	coords	lat , lon , time
variable	y	index_ext	1:180 , 1:360 , 1:100

DFD = Deterministic Factorial Design

SimEnv experiment description

general		type	DFD
factor	x1	default	5
factor	x1	sample	list 2, 3, 7, 10
factor	x2	default	4
factor	x2	sample	list 1, 3.5, 7
specific		comb	x1*x2

generate sample



Example 1: Experiment Description

Modified model source code

```

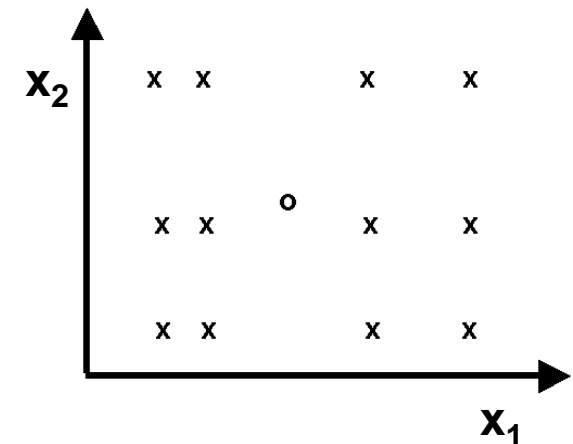
...
real*8 y(180,360,100)
x1 = 5.
istatus = simenv_get_f(x1,x1,x1)
x2 = 4.
istatus = simenv_get_f('x2',x2,x2)
... (compute model output y)
istatus = simenv_put_f('y',y)
...
    
```

— : associate SimEnv experiment description factor x_1 with model source code parameter x_1
— : get adjusted model source code parameter

SimEnv experiment description

general		type	DFD
factor	x1	default	5
factor	x1	sample	list 2, 3, 7, 10
factor	x2	default	4
factor	x2	sample	list 1, 3.5, 7
specific		comb	x1*x2

generate sample



Example 2: Monte Carlo Analysis

Modified model source code

```

...
real*8 y(180,360,100)
x1 = 5.
istatus = simenv_get_f('x1',x1,x1)
x2 = 4.
istatus = simenv_get_f('x2',x2,x2)
... (compute model output y)
istatus = simenv_put_f('y',y)
...
    
```

SimEnv model output description

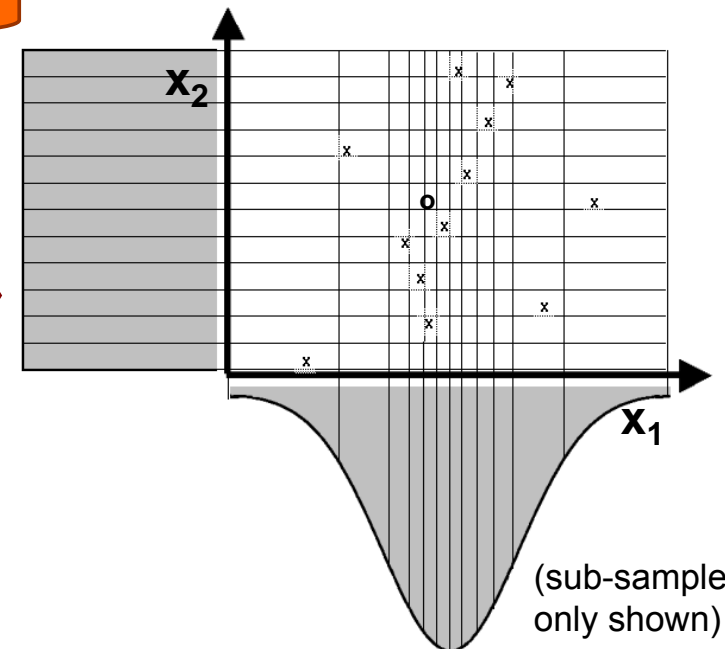
coordinate	lat	descr	geographic latitude
coordinate	lat	values	equidist_end 89.5 (-1) -89.5
coordinate	lon	values	equidist_nmb -179.5 (1) 360
coordinate	time	values	list 1,2,3, ... ,98,99,100
variable	y	descr	...
variable	y	type	real*8
variable	y	coords	lat , lon , time
variable	y	index_ext	1:180 , 1:360 , 1:100

UNC_MC = Monte Carlo Analysis

SimEnv experiment description

general		type	UNC_MC
factor	x1	default	5.
factor	x1	sample	distr N(5.,0.5)
factor	x1	s_method	stratified
factor	x2	default	4.
factor	x2	sample	distr U(3.5,4.5)
factor	x2	s_method	stratified
specific		runs	250

generate sample



Experiment Description File

- To generate a sample in the factor space under study an experiment type has to be equipped with numerical information

```

general          type          UNC_MC          # (M) experiment type Monte Carlo Analysis
factor  x1      descr          parameter influencing ... # description
factor  x1      unit           kg/m**2                # unit
factor  x1      default        1.                    # (M) factor nominal value
factor  x1      type           add                    # (M) add sampled values to nominal value
factor  x1      sample         distr N(1,0.04)        # (M+E) X1 ~ N(1,0.04) normal distribution
factor  x1      sample_method  pseudo                # use "standard" random number generator
factor  x1      sample_include  0.4:1.6              # sample only within 3σ interval

factor  x2      ...
factor  x2      sample         distr U(-5,5)         # X2 ~ U(-5,5) uniform distribution
factor  x2      sample_method  stratified            # use Latin hypercube sampling

factor  x3      ...
factor  x3      sample         file MyOwnSample.dat  # import sample for X3 from file

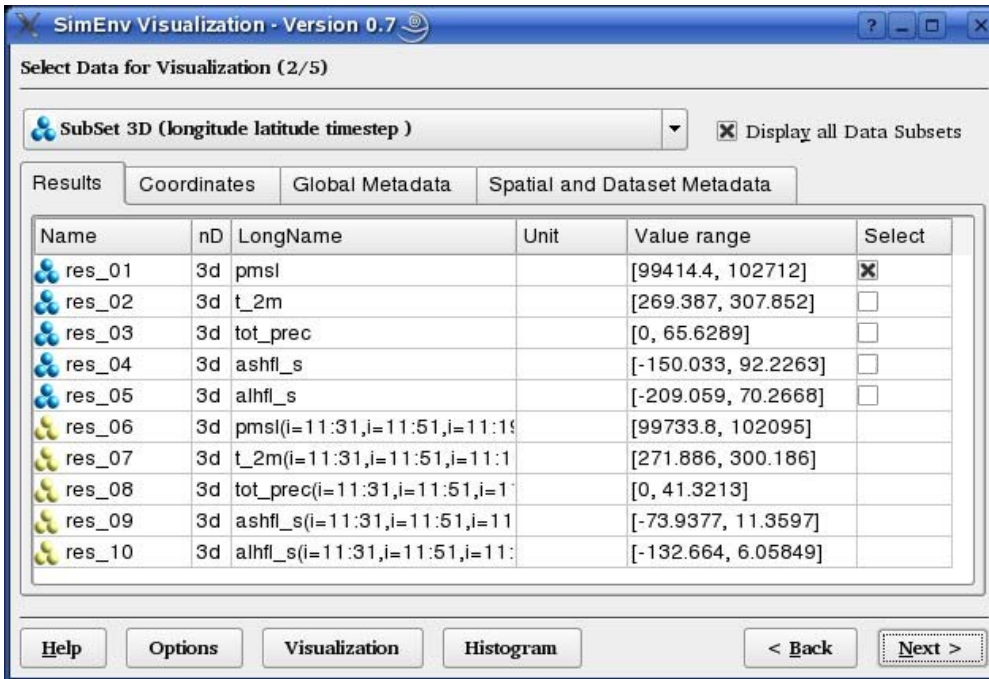
specific         runs          5000                # (M+E) max. number of Monte Carlo runs
specific         function       sum(sin(y+3*x1)-avg(y)) # model output function
                                                           for optional UNC_MC stopping rule
    
```

M = mandatory
E = experiment type specific

Experiment Analysis

- Process (aggregate, transform, ...) experiment output and derive measures
- By navigating the model output (state variable) space $\{y\}$ and the experiment (factor) space $\{X_k\}$
- Compute secondary output from model output and derive experiment related measures over $\{X_k\}$ by application of operator chains to experiment output, reference data and other SimEnv experiments
 - > 100 built-in operators
 - Interface to plug in user-defined operators
- Visual experiment analysis and evaluation by the visualization framework SimEnvVis

SimEnvVis Visualization Framework



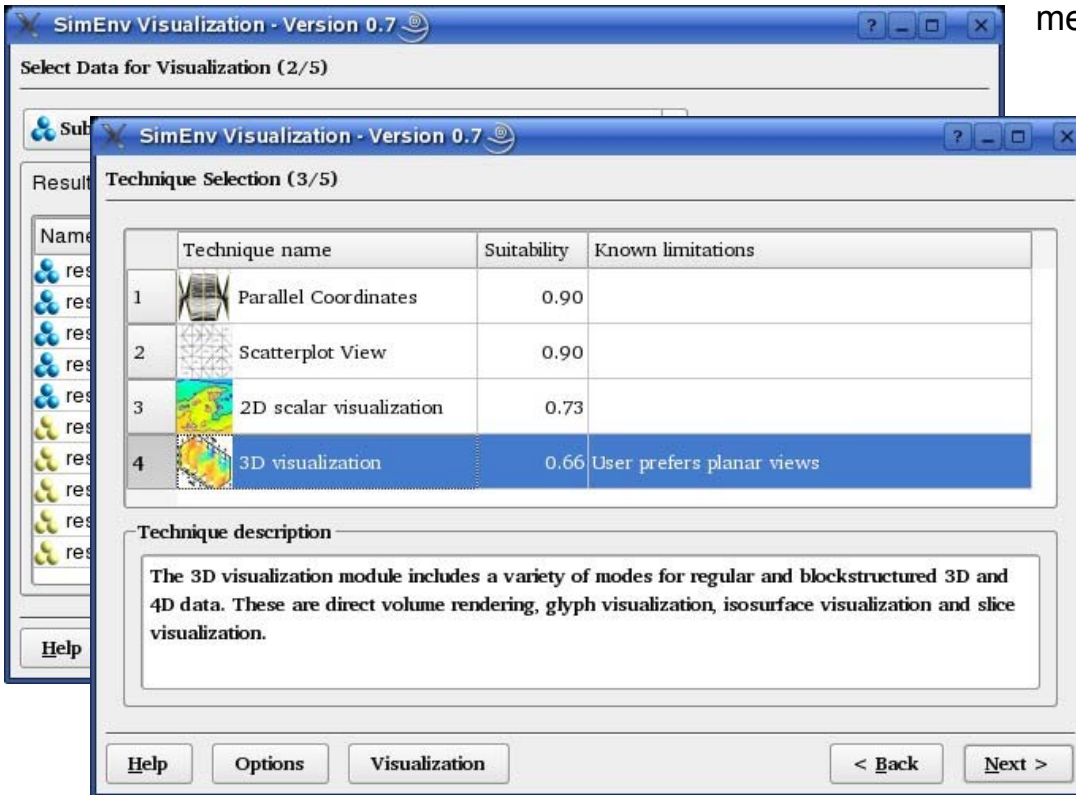
metadata display, adaptation, filtering

SimEnvVis (T. Nocke, PIK and M. Heitzler, ETH Zurich):

A visualization framework consisting of

- a wizard to analyze data, select a technique and parametrize it and
- coupled visualization systems / maps

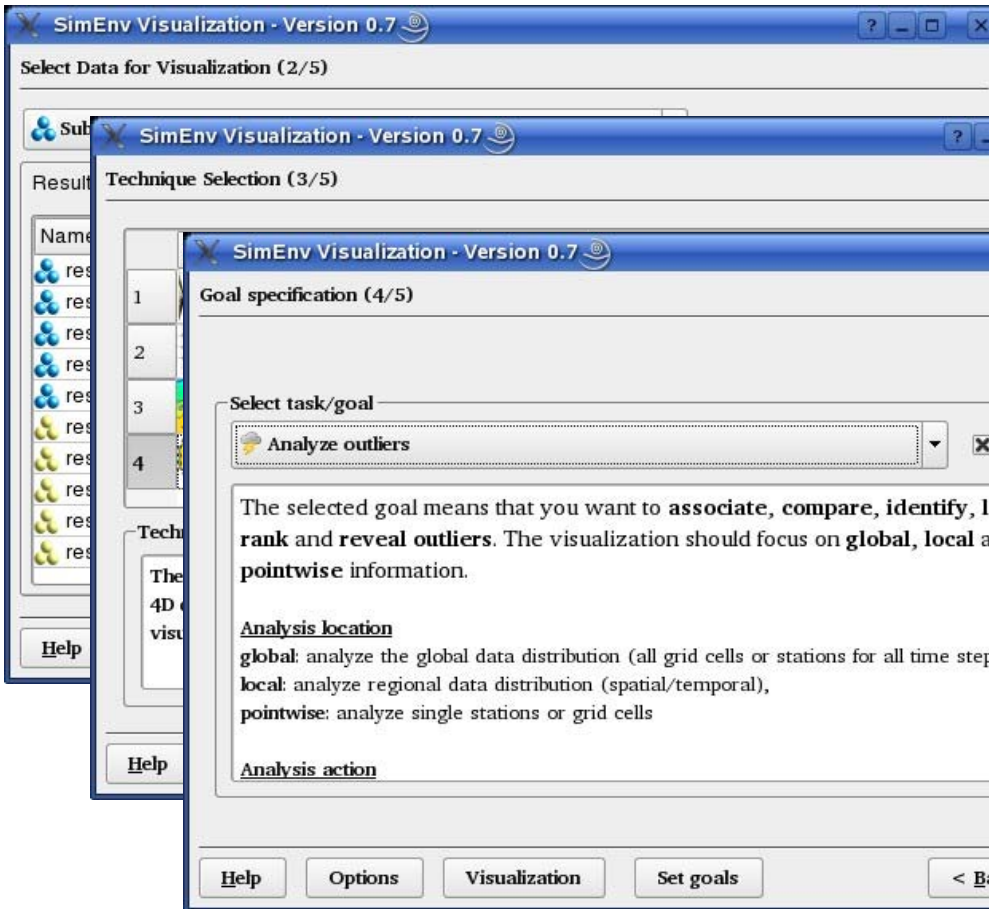
SimEnvVis Visualization Framework



metadata display, adaptation, filtering

selection of suited visualization techniques

SimEnvVis Visualization Framework



metadata display, adaptation, filtering

selection of suited visualization techniques

specification of goals




SimEnvVis Visualization Framework

The image shows four overlapping windows from the SimEnv Visualization software, illustrating the workflow for visualization setup. The windows are titled 'SimEnv Visualization - Version 0.7' and are numbered 2/5, 3/5, 4/5, and 5/5.

- Window 2/5: Select Data for Visualization** - Shows a list of data sources for selection.
- Window 3/5: Technique Selection** - Shows a list of visualization techniques for selection.
- Window 4/5: Goal specification** - Shows a list of goals for specification.
- Window 5/5: Technique Pametrization** - Shows a table of visualization techniques with their suitability and known limitations.

Annotations on the right side of the image describe the steps:

- metadata display, adaptation, filtering
- selection of suited visualization techniques
- specification of goals
- parametrization

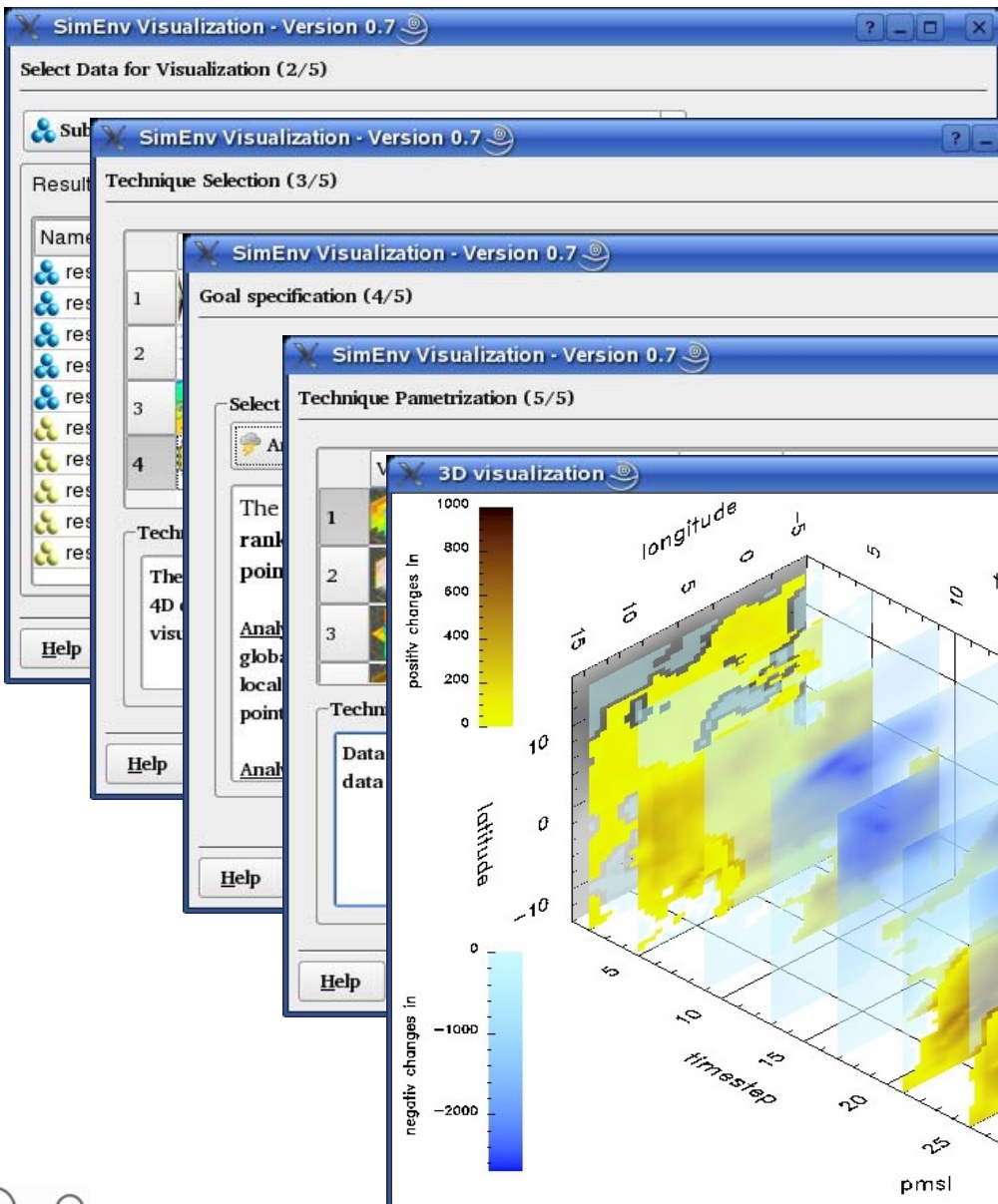
	Variation name	Suitability	Known limitations
1	 Slices (data)	0.61	
2	 Slices (changes)	0.58	
3	 Cutting planes	0.45	

Technique parameterization

Data slice visualization - Along one dimension semi-transparent planes represent colored cuts in data space. The number of cuts can be interactively adapted (overview and detail-on-demand).

Buttons: Help, Options, Visualization, < Back, Next >

SimEnvVis Visualization Framework



metadata display, adaptation, filtering

selection of suited visualization techniques

specification of goals

parametrization

start of interactive visualization

**regional climate
model CCLM**

Some SimEnv Applications at PIK

PIK

outside PIK

<i>Model</i>	<i>SimEnv Application</i>	<i>Model Interface</i>
4C	regional model applications, factorial designs, Bayesian calibration	Fortran
Aeolus	parameter sensitivity of the cloud module	C++
Climber-2	parameter screening / sensitivity, MCA, comparison of multi-parameter settings	Fortran
CLM	parameter screening	script level, Fortran
LPJmL	global / regional model application, parameter sensitivity	Fortran, C, script level
MICA	parameter screening for stabilization of climate coalitions games	GAMS
Monsoon	global sensitivity, uncertainty	Fortran, C, script level
REMIND	parameter screening uncertainty analyses	GAMS
Lagom	model inspection, global sensitiv. analyses, initial value and parameter sensitivity	Java / Scala, Matlab, Mathematica
Island	model sensitivity: Monte Carlo, Morris method, variance decomposition	GAMS

SimEnv Usage Statistics (PIK, 10 yrs.)

Total

	<i>sum</i>	<i>~ per day</i>
Users	40+	
Models	30+	
Services	65.000	18
Experiments	15.000	4
Single runs	4.000.000	1.100
Post-processor sessions	28.000	8

Experiment types

67%	Deterministic factorial design
19%	Monte Carlo analysis
5%	Factor ranking / Morris
9%	other types

Per experiment

	<i>min</i>	<i>mean</i>	<i>max</i>
Single runs	2	288	50.000
Factors	1	8	50
Output variables	1	11	97
Output values / single run	1	$50 \cdot 10^3$	$30 \cdot 10^6$
Output bytes / single run	4	$205 \cdot 10^3$	$125 \cdot 10^6$
Output bytes / experiment	40	$51 \cdot 10^6$	$18 \cdot 10^9$
Used processor cores	3	56	255

Summary / Conclusions

SimEnv ...

- ... is an experimentation tool to cope with model sensitivity and uncertainty by multi-run simulations from different perspectives
- ... enables a quick and easy approach for standard tasks of simulation models' quality assurance
- ... copes with high-dimensional factor spaces and large volume multi-dimensional model output
- ... supplies simple model and operator interfaces
- ... easily generates NetCDF model and post-processor output
- ... incorporates high-end visual evaluation of post-processed experiment output

Further Reading

Sensitivity & uncertainty analyses

- Eykhoff, P. (1974) System identification: Parameter and state estimation. John Wiley & Sons
- Sobol', I.M. (1993) Sensitivity estimates for non-linear mathematical models. Math. Model. Comput. Exp., 1, 407-414
- McKay, D.J. (2004) Information theory, inference, and learning algorithms. Cambridge
- Saltelli, A. et al. (2008) Global sensitivity analysis. The primer. John Wiley & Sons

SimEnv and applications

- Flechsig, M., Böhm, U., Nocke, T., Rachimow, C. (2005): Techniques for quality assurance of models in a multi-run simulation environment. In: Hanson, K.M. and Hemez, F.M. (eds.): Sensitivity Analysis of Model Output. Los Alamos National Laboratory, Los Alamos, NM, USA, 297-306
- Böhm, U., Kücken, M., Hauffe, D., Gerstengrabe, F.-W., Werner, P.C., Flechsig, M., Keuler, K., Block, A., Ahrens, W., Nocke, T. (2004): Reliability of regional climate model simulations of extremes and of long-term climatic characteristics. Natural Hazards and Earth System Sciences (NHES), 4: 417 – 431
- Knopf, B., Flechsig, M., Zickfeld, K. (2006): Multi-parameter uncertainty analysis of a bifurcation point. Nonlin. Processes Geophys., 13, 531-540
- Knopf, B., Zickfeld, K., Flechsig, M., Petoukhov, V. (2008): Sensitivity of the Indian monsoon to human activities. Advances in Atmospheric Sciences, 25/6, 932-945
- Van Oijen, M. et al. (2013): Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe. Forest Ecology and Management, 289, 255-268
- Reyer, C., Flechsig, M., Lasch, P., van Oijen, M.: Integrating parameter uncertainty of a process-based model in assessments of climate change effects on forest productivity. Annals of Forest Science, in prep.

Thank you for your Attention



PIK's Main Building (Michelson House)

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