Techniques for Quality Assurance of Models in a Multi-Run Simulation Environment

M. Flechsig [a], U. Böhm [b], T. Nocke [c], C. Rachimow [a]

[a]Potsdam Institute for Climate Impact Research, 14412 Potsdam, Germany
[b]University of Potsdam, Institute for Physics, 14415 Potsdam, Germany
[c]University of Rostock, Institute of Computer Graphics, 18051 Rostock, Germany
E-mail: flechsig@pik-potsdam.de, boehm@pik-potsdam.de,
nocke@informatik.uni-rostock.de, rachimow@pik-potsdam.de

Abstract: In this paper we describe the multi-run simulation experiment environment SimEnv and its application in quality assurance matters for computer models. SimEnv has been developed to provide key working techniques for experimenting with complex models. This includes a wide range of simulation and model output evaluation methods in combination with corresponding visualization techniques. The SimEnv framework facilitates the easy execution of multi-run model simulation experiments for standardized, pre-formed experiment types which represent different sampling strategies of the model's input space. Further experiment types may easily be included, making SimEnv an open experimentation system. The coupling of models to the environment is supported by a simple interface, requiring only minimal model source code modifications. Uncertainty and sensitivity analyses are enabled in SimEnv by combining experiments available from the pool of pre-defined experiment types with interactive post-processing, applying sequences of related operators to both model output and reference data. Use of SimEnv as an experimental framework for models in global change research demonstrates the applicability of the approach to multi-input / multi-output problems with large amounts of spatio-temporal model output and emphasizes the importance of graphical result presentation and evaluation by appropriate visualization techniques.

Keywords: simulation environment, multi-run experiments, uncertainty analysis, sensitivity analysis

1. INTRODUCTION

Dealing with uncertainty and communicating it to decision makers and the general public is crucial in climate change research [1]. Recent papers address this issue for climate projections e.g., [2] on the basis of the findings of the Third Assessment Report of the Intergovernmental Panel of Climate Change. Identifying uncertainty in climate predictions requires comprehensive experiments for the diagnosis of the models used. The design of such models, simulation and evaluation are cornerstones in climate impact and global change research. In the past, chains of stand-alone model simulations were performed to derive from an input scenario (e.g., of greenhouse gas emission over time) of one model, outputs (e.g., climate change over time) then used as inputs to a succeeding model. The complete system can be studied and investigated this way. Nowadays, one of the challenges in global change research is the development of integrated models, which is being achieved mainly by the additional knowledge gained through feedbacks between the studied sub-systems on one hand, and through increasing computing power on the other.

Such complex simulation models are often based on legacy source code applications written in a programming language rather than in a model design language. They produce a large amount of (spatio-temporal) model output that has to be handled in the course of model validation, corroboration and/or scenario analyses. These aspects hamper the application of quality assurance techniques to this kind of models, 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 are often very high, which demands structured experimentation approaches.

2. GENERAL SIMENV APPROACH

SimEnv [3] has been developed to provide a toolbox-oriented simulation environment that enables the modeller and/or model user to deal with model-related quality assurance matters and scenario analyses for such models as described above. Both foci require flexible experiment design and model output evaluation to enable model inspection, validation / corroboration, uncertainty and sensitivity analyses without the necessity to change a complex model in general.

With respect to systems theory we consider a dynamic model **M** that can be formulated for the time dependent, time discrete, and state deterministic case - without limitation of generality - as

$$\mathbf{M}: \quad \mathbf{Z}(t) = \mathbf{ST}\ (\mathbf{Z}(t-\Delta t), \ldots, \mathbf{Z}(t-\mathbf{n}*\Delta t), \mathbf{X}(t), \mathbf{T})$$

$$\mathbf{Y} = \mathbf{OU}\ (\mathbf{Z}(t), \mathbf{T}),$$

with

ST	state transition description
OU	output function
Z	state vector
X	input vector
T	parameter, initial value $\mathbf{Z}(t_0)$, and/or boundary value vector
t	time
Δt	time increment
n	time delay
Y	output vector

In the following, **z** and **t** are components of the vectors **Z** and **T** respectively.

The basic idea for the system design of SimEnv is to study **M** in dependence on numerical changes of a subset **t** of the parameter, initial value, and/or boundary value vector **T**:

$$\mathbf{z} = \mathbf{M}\ (\mathbf{t}),$$

where **z** is normally associated with large-scale multi-dimensional state vectors, defined over time and (geographic) space.

Simulation studies in SimEnv are supported by introducing standardized, pre-formed experiment types. An experiment type represents a multi-run simulation experiment technique with a sequence of co-ordinated single runs. According to the strategy of a selected experiment type the experiment inputs **t** (so-called targets) are sampled in the target space **{t}.** For each realization from the sample, a single simulation run of the run ensemble is performed. After setting up an experiment by equipping an experiment type with related information about the sample in {**t**} all single runs from the run ensemble are performed independently of each other. Consequently, they can be performed sequentially or in distributed mode on a cluster of networking computers using the generic Message Passing Interface MPI [4].

Preparation of a model for coupling it to SimEnv involves minimal source code manipulations for a set of supported model programming languages. Experiment-specific model output post-processing enables navigation in the combined experiment - model output space **{t∪z}** spanned up from the considered targets **t** and the multi-dimensional state vectors **z**. Application of built-in and user-defined post-processing operator sequences enables interactive filtering of model output and of reference data. Visualization of post-processed model output with pre-formed visualization modules forms a major component within the result evaluation component. Fig. 1 shows the general pathway for experimenting within SimEnv.

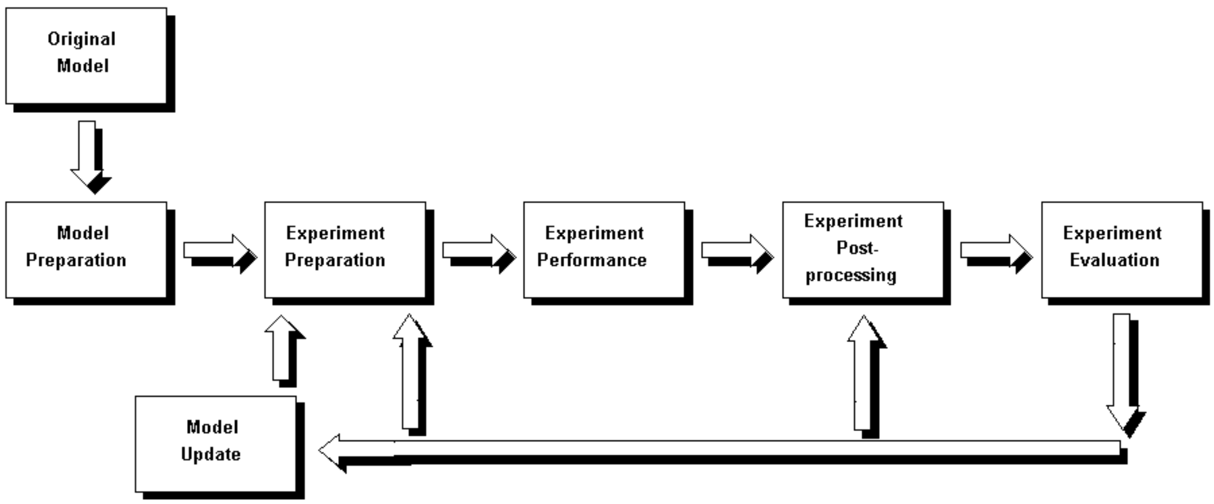

Figure 1. SimEnv System Design.

3. MODEL COUPLING

The SimEnv approach to plug in models to the simulation environment demands the availability of source code for minimal source code adaptations in order

- to map targets **t** with which the modeller wants to experiment and numerical adjustments of these from the simulation environment to the model **M,** and
- to store (n-dimensional) state variables **z** and targets **t** from **M** to SimEnv data structures for later post-processing

for each realization from the total sample on **{t}**.

The coupling interface is available for models implemented in C, Fortran, Python and in the General Algebraic Modeling System GAMS [5] for mathematical programming problems. It supports all numerical data types. Plugging the model into SimEnv requires for the model source code additional implementation of

- one function call simenv_get for each target **t** to re-adjust its value numerically according to the current single run of the experiment and
- one function call simenv_put for each model output variable **z** to store it in SimEnv output files during the current single run for later post-processing.

Additionally, at the UNIX command shell level analogous scripts are available. Among other things, they enable manipulation of model control files or forwarding re-adjusted target values as arguments to the model before each model run without changing the model at all. SimEnv-related model output storage uses self-describing Network Common Data Form NetCDF format [6] or IEEE compliant binary format.

A model description file specifies in detail the model state variables **z** and the grid on which a state variable is defined. SimEnv supports usage of rectilinear (orthogonal with variable distance) grid definitions. Due to a flexible assignment of model variables to grids, model variables can exist on the same grid or on completely or partially disjointed grids.

4. EXPERIMENT TYPES

SimEnv aims at a well-tailored and co-ordinated simulation approach by performing run ensembles instead of single simulation runs. Co-ordination is achieved by use of pre-defined experiment types representing multi-run simulations. An experiment type scans a multi-dimensional target space **{t}** with a specific sampling strategy. Experiment types implemented so far are

- Behavioural analysis
  Deterministic inspection of the model's behaviour with a flexible sampling strategy in the target space
- Monte-Carlo analysis
  Probabilistic sampling of targets according to pre-defined distributions using different sampling methods
- Local sensitivity analysis
  Deterministic sampling in a local neighbourhood of the control scenario as the numerical nominal (default) target constellation of the model **M**.

Experiments are specified in an experiment description file by selecting an experiment type and defining the target space **{t}** and the sampling strategy.

SimEnv behavioural analysis is a generalization of the one-dimensional case, where the model behaviour is scanned in dependence on deterministic adjustments of one target **t**. The n-dimensional case demands a strategy for scanning multi-dimensional spaces in a flexible manner. On the basis of the SimEnv predecessors [7] and [8] subspaces of **{t}** can be scanned on the subspace diagonal (parallel on a one-dimensional hyperspace) or completely for all dimensions (combinatorial on a grid) and both techniques can be combined. Besides this regular sampling method an irregular, file-based technique is provided.

Fig. 2 describes the regular scanning technique by an example. In the left scheme the two-dimensional target space $\{\mathbf{t}\} = \{\mathbf{p}_1 , \mathbf{p}_2\}$ is scanned in a combinatorial manner, resulting in 4*4 = 16 model runs, while the middle scheme represents a parallel scanning pattern of the two targets at the diagonal by 1+1+1+1 = 4 model runs. The scheme on the right shows a combined scanning strategy of the 3-dimensional target space $\{\mathbf{t}\} = \{\mathbf{p}_1 , \mathbf{p}_2 , \mathbf{p}_3\}$ with (1+1+1+1)*3 = 12 model runs. Each filled dot represents a single model run.

In Monte-Carlo analysis pre-defined distributions can be used to generate a sample in the target space. Random and Latin hypercube sampling [9] is supported for uniform, normal, log-normal and exponential distributions. Currently, SimEnv only supports sampling of uncorrelated targets; as a workaround, there is an interface to import external samples.

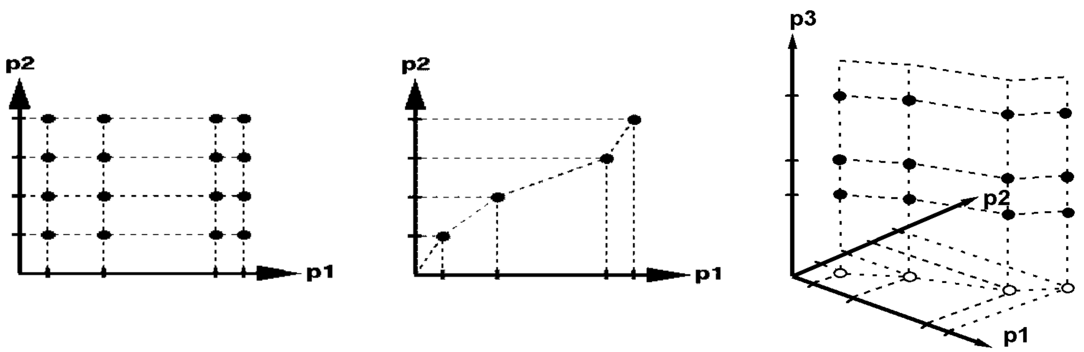

Figure 2. Behavioural analysis: Deterministic sampling of multi-dimensional target spaces.

For local sensitivity analysis the experiment is set up by single model runs in ε-neighbourhoods of the control scenario in the target space **{t}**. For each target $\mathbf{t}_i$ from the control scenario $\mathbf{t} = (t_1, \ldots, t_n)$ and each ε_j from $\varepsilon_j = (\varepsilon_1, \ldots, \varepsilon_m)$ two runs are performed for the both target constellations $(t_1, \ldots, t_{i-1}, t_i \pm \varepsilon_j, t_{i+1}, \ldots, t_n)$.

5. EXPERIMENT POST-PROCESSING AND VISUAL EVALUATION

Interactive post-processing is applied to compute output functions **y** from the model's outputs **z** by state space transformation operators and to derive uncertainty and sensitivity measures from these output functions by experiment type-specific operators. For this purpose, the SimEnv post-processor enables application of operator sequences to both model output and reference data. Currently, about 100 built-in operators are available. An interface enables users easily to declare their own operators and plug them into the environment. Each operator assigns to its result output a unique grid definition, derived from the operator definition and the grids of its operands. SimEnv post-processor output can be stored in NetCDF, IEEE compliant binary or ASCII format.

State space transformation operators cover elemental, selective, analytical, and statistical techniques, among others. The main focus is reduction of and aggregation in the output model state space to cope with its potentially high dimensionality and extent. Selective operators provide methods to access to a selected single run, to external data and other SimEnv experiments and to clip the extents or to reduce the dimensionality of an operand on its assigned grid. Statistical operators supply basic statistical information from operands on the whole grid or on grid layers for single grid dimensions.

Analysis and evaluation of post-processed data derived from large amounts of relevant model output benefit from visualization techniques. Based on metadata information about the post-processed experiment type, the applied operator sequence, and the dimensionality of the post-processor output, pre-formed visualization modules are evaluated by a suitability coefficient to determine how they can map post-processor output in an appropriate manner.

The visualization modules offer a high degree of user support and interactivity to cope with multi-dimensional data structures. Among others, they cover standard techniques such as scatter and parallel coordinate plots (the latter for abstract data visualization), and isolines, isosurfaces, direct volume rendering and 3D difference visualization techniques. Furthermore, approaches to navigate intuitively through large multi-dimensional data sets have been applied, including details on demand, interactive filtering and animation [10]. Using the open source visualization platform OpenDX [11] based on IBM's Data Explorer, extended

OpenDX techniques have been designed and implemented, suited to the context of analysis and evaluation of simulated multi-run output functions.

6. UNCERTAINTY AND SENSITIVITY ANALYSES

The key methodological approach for uncertainty and sensitivity analyses in SimEnv is the combination of experiments from the set of pre-defined experiment types with interactive exploration of the model output variables' set from the run ensemble in experiment post-processing, applying sequences of experiment-specific operators to both state space model output functions and reference data. Derived from the general experiment layout, SimEnv experiment types are associated with uncertainty and sensitivity analyses techniques in the following way:

- Behavioural analysis
  Can be used for uncertainty analysis, factorial screening, general one-factor-at-a-time approach, (fractional) factorial experiments and response surface methodology. All methods benefit from the flexible screening strategy of multi-dimensional target spaces in SimEnv.
- Monte-Carlo analysis
  Can be used for uncertainty analysis and global sensitivity analysis.
- Local sensitivity analysis
  Can be used for local first order sensitivity analysis by investigating finite difference approximations of derivatives.

During post-processing uncertainty and/or sensitivity measures are provided by experiment-specific operators. A general behavioural analysis operator enables the modeller/user to navigate in the target space $\{\mathbf{t}\}$ and to derive aggregations and moments in its sub-spaces in a flexible manner. Monte-Carlo analysis operators support (among other things) computation of extremes, moments, quantiles and heuristic probability density functions from targets and output functions as well as regression, correlation, and covariance measures from targets, model output, or both of these together. For local sensitivity analysis a set of sensitivity operators (linear, squared, absolute, relative, symmetric) are available as finite approximations of the classical local sensitivity measure $\partial\mathbf{z}/\partial\mathbf{t}$.

7. EXAMPLE

We show from an ongoing study sensitivity results for CLM, a regional meteorological model CLM [12] in climate mode [13] where parameters controlling both the dynamic forecast part and the parametrization part for subgrid-scale diabatic source and sink processes in their relation to diagnostic and prognostic model output variables have been under investigation.

CLM is used with a horizontal resolution of 0.5° x 0.5° latitude/longitude and with 20 layers in the vertical for a region covering the Baltic Sea and most of Northern and Central Europe. The model time step is 90 seconds and output is stored every six hours. The model is based on the non-hydrostatic, fully compressible primitive equations of the atmospheric motion without scale approximations. The model uses a generalized terrain-following vertical coordinate and rotated geographical coordinates. It is subdivided into a so-called dynamic part, where the basic equations, spatially discretized by use of second-order finite differences, are solved for the prognostic variables wind velocity in x- and y- direction of the orthogonal z-system, perturbation pressure, to the hydrostatic basic stage, temperature, specific humidity, cloud water content and (optionally) cloud ice content. Sub-grid scale source- and sink-processes have to be parametrized and are computed before the dynamic part. Among others, also soil hydrological and thermal processes are described by such a parametrization.

In our investigations, we consider the hydrological section of the soil parametrization in CLM. One of the components of the near-surface water balance is transpiration by plants from two soil layers with a depth of 10 cm and 90 cm. This process is described by a Biosphere-Atmosphere Transfer Scheme [14]. The basic idea is to apply a resistance concept as in electricity to compute plant transpiration affected by atmospheric and stomatal factors. One of the used transpiration reduction factors accounts for the reduction of transpiration by the stomatal resistance r_s.

$$\frac{1}{r_s} = \frac{1}{crs_{max}} + \left(\frac{1}{crs_{min}} - \frac{1}{crs_{max}} \right) \cdot F_{rad} \cdot F_{wat} \cdot F_{temp} \cdot F_{hum}$$

r_s is described by the two parameters crs_{min} and crs_{max} and various influence functions F. For $crs_{min} = crs_{max}$ transpiration is not reduced by any of the influence functions. In the function

$$F_{temp} = \max \left[0 \text{ , } \min \left\{ 1 \text{ , } 4 \cdot \frac{(T_s - T_0) \cdot (T_{end} - T_s)}{(T_{end} - T_0)^2} \right\} \right]$$

for the influence of the surface temperature T_s the empirical constant T_{end} describes optimal conditions for plant transpiration. F_{temp} reaches its maximum for $T_s \approx T_{end}/2$.

We apply a behavioural analysis to assess the effect of the empirical parameters crs_{min} and T_{end} on latent and sensible heat fluxes lhf and shf from soil in a deterministic manner for crs_{max} = 1000 s/m. Both fluxes are defined on a grid spanned up from latitude, longitude and time. In Box 1 the experiment description file to scan the 2-dimensional parameter space $\{crs_{min}\text{ , }T_{end}\}$ combinatorially is shown. Additionally, in the model source code crs_{min} and T_{end} have to be re-adjusted by a simenv_get-call for each of both parameters.

```
target    crsmin   adjusts   30.(5.)120.        # specifies 19 adjustments for crs_min
target    crsmin   default   60.                # default model value of crs_min
target    crsmin   type      set                # do not modify adjustments by default
target    Tend     adjusts   273.15(5.)333.15   # specifies 13 adjustments for T_end
target    Tend     default   313.15
target    Tend     type      set
specific           comb      crsmin*tend        # factorial screening: 19*13+1=248 runs
```

Box 1. Experiment description file for a behavioural analysis.

Post-processed results for a simulated period of seven days are shown in Fig.3. The influence of the variation of crs_{min} and T_{end} on lhf and shf anomalies from the model nominal constellation is shown on the left. To produce during SimEnv post-processing this result from model output the applied operator sequences are

behav(‘ ‘, avg(shf)) - run(‘default’, avg(shf)) and
behav(‘ ‘, avg(lhf)) - run(‘default’, avg(lhf)),

where behav is the general behavioural operator to navigate in the experiment space, the operator run addresses one single run from the whole run ensemble, and the operator avg supplies the total average from a multi-dimensional model output variable. To get area-averaged flux anomalies for each time step time dependent on crs_{min} and for the default value of T_{end} we have to apply in post-processing

behav('sel_t(Tend=313.15)', avg_l('time', shf)) - run('default', avg_l('time', shf))

where avg_l supplies area averages for each time step. Fig. 3 on the right is the corresponding graphical representation.

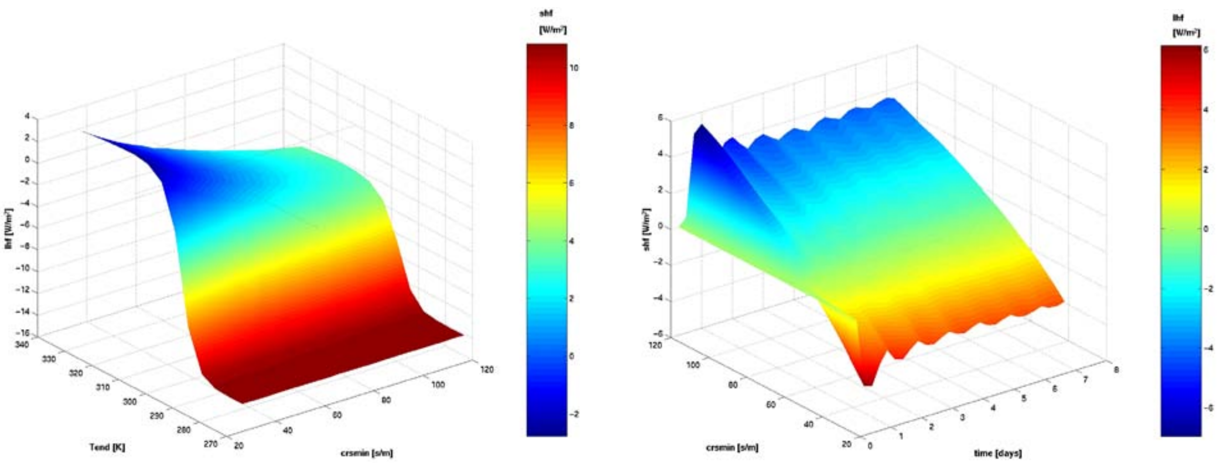

Figure 3. Surface heat flux anomalies from soil. Dynamic was compiled in SimEnv post-processing from the results of the 248 single runs of the experiment.
Left: Area and temporal mean dependent on T_{end} and crs_{min}.
Right: Area mean for each time step dependent on crs_{min}.

First it becomes visible from the right panel of Fig. 3 that both heat fluxes behave inversely to the changes in crs_{min}. As to be expected, the latent heat flux lhf decreases with increased resistance values, whereas the sensible heat flux shf increases to transport heat back from the surface to the atmosphere in this case and to maintain the surface energy balance. Secondly, the reaction of both heat fluxes is rather linear for the entire parameter space. Together with changes in T_{end}, however, the behaviour of the heat fluxes is significantly different: As shown on the left, only for rather high values of T_{end} the heat fluxes change with crs_{min} as for the default of T_{end} on the right. For T_{end} below about 273.15 K, this parameter dominates the reaction of the sensible and latent heat fluxes and nearly no modifications in the results due to crs_{min} can be identified.

The results of a Monte-Carlo study on T_{end} and crs_{min} for a simulated period of seven days are shown in Fig. 4. Both parameters are drawn from a normal distribution with a Latin hypercube sampling technique where the mean is the nominal parameter value and variance is set to 20. Sample size is 150 runs. For the left panel of Fig. 4 the applied operator sequence is

hgr_e(15, avg_l('time', shf) – run('default', avg_l('time', shf))),

where the operator hgr_e supplies for each element of its second argument a heuristic probability density function over the whole run ensemble with 15 bins.

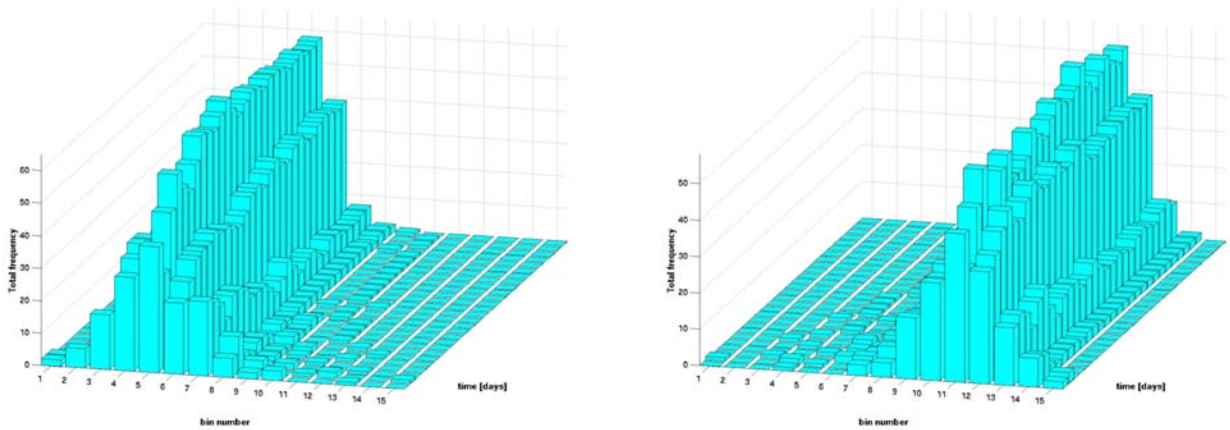

Figure 4. Probability density functions of surface heat flux anomalies from soil for time steps 2 - 28. Left: lhf, right: shf. lhf anomaly values range from -5.02 (bin # 1) to 2.03 (bin # 15), shf anomaly values from -1.51 to 3.87.

8. RESULTS AND CONCLUSIONS

The methodology presented here and its implementation have been proven to support the process of model evaluation from various perspectives of both model developers and users. In contrast to other simulation environments (e.g., SimLab [15], SCIRun [16], and Pingo [17]) that also focus on uncertainty and sensitivity matters, with SimEnv we try to support all steps in experimenting with models from easy-to-use model coupling to the system via experiment design, experiment load distribution, and model output post-processing to visual evaluation. The supported languages cover most of the model sources codes used in global change research. The concept of pre-defined experiment types seems to be an appropriate way to guide model developers and/or users in the process of experimenting with models and frees them from expensive workload. The plug-in interface for user-defined operators opens the post-processor to permit coupling to special-purpose applications or libraries on user demand. Additionally, output formats from the post-processor can be used to export model results to other applications, e.g. as statistical diagnosis and analysis tools, for in-depth investigations of specific research goals. One of the strengths of SimEnv is its support of multi-dimensional model output data on rectilinear grids in a persistent manner for model plug-in, post-processing, and visualization.

On the other hand, this holistic approach is at the same time one of the weaknesses of SimEnv. With SimEnv, we provide a general simulation environment for a broad spectrum of tasks without supporting special features in detail. For example, sampling strategies and built-in operators especially for uncertainty and sensitivity analyses techniques are limited.

9. PROSPECTS

The following work packages are planned for further development of SimEnv:

- Special-purpose sampling designs: Support of special uncertainty and sensitivity experiments, e.g., the Fourier amplitude sensitivity test FAST and/or the method of Sobol [18] and implementation of corresponding post-processing operators and visualization techniques.
- Simulation-based optimization: Application of gradient-free methods for (mono- and) multi-criterial optimization of cost functions $f_i(z)$ in the target space **{t}.**

- Support of distributed models across computer networks or the Internet: Setting up a SimEnv experiment server to handle target dissemination and model output collection.

ACKNOWLEDGEMENT

The authors thank Martin Kücken and Detlef Hauffe for performing the corresponding experiments with the CLM model.

REFERENCES

1. M. Webster. Communicating Climate Change Uncertainty to Policy-Makers and the Public. *Climatic Change*, 61:1-8, 2003.
2. M. Webster, C. Forst, J. Reilly, M. Babiker, D. Kickligther, M. Mayer, R. Prinn, M. Sarofim, A. Sokolov, P. Stone, and C. Wang. Uncertainty Analysis of Climate Change and Policy Response. *Climatic Change*, 61:295-320, 2003.
3. M. Flechsig, U. Böhm, T. Nocke, and C. Rachimow. *The Multi-Run Simulation Environment SimEnv: User's Guide.* Potsdam Institute for Climate Impact Research, Potsdam. http://www.pik-potsdam.de/topik/pikuliar/simenv/home/simenv.pdf, 2004.
4. MPI. http://mpi-forum.org
5. GAMS: http://www.gams.com
6. NetCDF. http://www.unidata.ucar.edu/packages/netcdf/
7. V. Wenzel, E. Matthäus, and M. Flechsig. One Decade of SONCHES. *Syst. Anal. Mod. & Sim.*, 7:411-428, 1990.
8. M. Flechsig: *SPRINT-S: A Parallelization Tool for Experiments with Simulation Models.* PIK-Report No. 47. Potsdam Institute for Climate Impact Research, Potsdam, 1998 http://www.pik-potsdam.de/reports/pr-47/pr47.pdf
9. R.L. Imam and J.C. Helton. An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models. *Risk Anal.*, 8(1):71-90, 1998.
10. T. Nocke, H. Schumann, U. Böhm, and M. Flechsig. Information Visualization Supporting Modelling and Evaluation Tasks for Climate Models. In: S. Chick, P.J. Sanchez, D. Ferrin, D.J. Morrice, editors, *Proceedings of the 2003 Winter Simulation Conference*, December 2003, New Orleans, 2003.
11. OpenDX. http://www.opendx.org
12. G. Doms and U. Schättler. *The Nonhydrostatic Limited-Area Model LM of DWD – Part I: Scientific Documentation.* Technical Report, German Meteorological Office DWD, Offenbach/M., 1997.
13. M. Kücken and D. Hauffe. *The Nonhydrostatic Limited-Area Model LM of DWD with PIK Extensions. Part III: Extensions User Guide.*2003. http://w3.gkss.de/CLM/clm_home.html
14. R. Dickinson. Modeling Evapotranspiration for Three-dimensional Global Climate Models. In *Climate Processes and Climate Sensitivity.* Geophysical Monograph 29, Maurice Ewing Volume 5, American Geophysical Union, Washington, D.C., 58-72, 1984.
15. SimLab. *Software for Uncertainty and Sensitivity Analysis. POLIS-JRC-ISIS* http://sensitivity-analysis.jrc.cec.eu.int/default2.asp?page=forum
16. SCIRun Scientific Computing and Imaging http://software/sci/utah.edu
17. J. Waszkewitz, P. Lenzen, and N. Gillet (2001) *The PINGO Package* http://www.mad.zmaw.de/Pingo/pingohome.html, 2001.
18. K. Chan, K., S. Tarantola, A. Saltelli, and I.M. Sobol. Variance-Based Methods. In: Saltelli, A., Chan, K., Scott, E.M., editors, *Sensitivity Analysis.* J. Wiley & Sons, Chichester, 2000.