

Simulationsumgebungen – Anforderungen, Entwicklung und Einsatz in der Klimafolgenforschung

Michael Flechsich und Thomas Nocke

Potsdam-Institut für Klimafolgenforschung
Telegrafenberg, 14473 Potsdam
flechsich@pik-potsdam.de, nocke@pik-potsdam.de

Zusammenfassung:

Simulationsumgebungen gestatten in einem strukturierten Ansatz den Entwurf von und/oder das Experimentieren mit Simulationsmodellen. In der Klima- und Klimafolgenforschung sind dabei insbesondere die Möglichkeiten von Simulationsumgebungen für Entwurf, Ausführung und Analyse von Experimenten mit Modellen von Interesse. Wir diskutieren Anforderungen an solche Simulationsumgebungen unter besonderer Beachtung des Experimententwurfs und stellen diese den Ergebnissen einer Umfrage unter Klimafolgenforschern zur Nutzung entsprechender Techniken und Tools gegenüber. Mit SimEnv präsentieren wir eine Multi-Run Umgebung mit den Schwerpunkten Modellsensitivität, Modellunsicherheit und Szenarioanalyse und zeigen daraus Beispiele für die explorative und visuelle Analyse des Experimentoutputs.

1 Begriffs- und Problemabgrenzung

Simulationsumgebungen sind Software-Tools, die den Lebenszyklus eines Modellierungs- und Simulationsprojektes (Abb. 1) in strukturierter Weise unterstützen. Die Definition des Begriffs ist nicht eindeutig. Während in einigen Umgebungen der Schwerpunkt auf die Modellentwicklung gelegt wird („Modellierungsumgebung“) liegt er in anderen auf dem Experiment („Experimentierumgebung“) und es werden außerhalb der Umgebung entwickelte Modelle unterstützt. Wir werden uns mit Experimentierumgebungen beschäftigen und den Begriff Simulationsumgebung synonym benutzen.

Formal betrachten wir ein Modell $M: y = F(x)$, das einen Inputvektor x mit k Faktoren (x_1, \dots, x_k) aus dem Faktorraum X_k (Box et al. 1978) durch einen Algorithmus F in einen Output (vektor) y überführt. x sind diejenigen Parameter, Anfangs- bzw. Randwerte vom M , die einer Variabilität unterliegen, da sie z.B. im Realsystem nicht genau bestimmbar sind, fluktuieren oder ganz bewusst von dessen aktuellem Zustand abweichen sollen und damit für Modellentwickler und -nutzer besonders interessant sind. Zum Experimentieren wird aus X_k eine Stichprobe (sample) $\{X_k\}$ so gezogen, dass mit dieser eine Fragestellung durch das Experiment beantwortet werden kann. Typische Fragestellungen sind z.B. die nach dem Einfluss eines Faktors auf y , nach den sensitivsten Faktoren bzgl. y oder nach demjenigen x , das y minimiert. Nach einer Multi-Run Ensemble Simulation von M über $\{X_k\}$ wird der Experimentoutput $\{y\} =$

$\{F(x)\}$ in Abhängigkeit vom gewählten Experimentdesign D so analysiert, dass mit Hilfe von Operatoren $G_D(\{y\})$ die Fragestellung beantwortet werden kann.

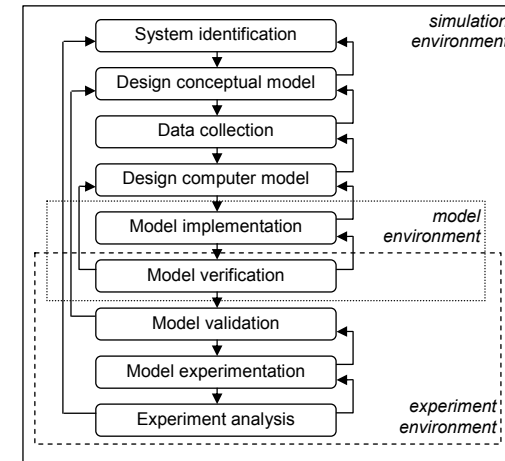


Abbildung 1. Zyklus eines Modellierungs- und Simulationsprojektes

2 Herausforderungen in der Klimafolgenforschung

In der Klima- und Klimafolgenforschung sind Modellentwicklung und -simulation neben der Datenanalyse und -integration die hauptsächlichsten methodischen Ansätze. Experimente mit dem Realsystem lassen sich – wenn überhaupt – nur für thematisch, räumlich und zeitlich sehr begrenzte Teilsysteme des gesamten Erdsystems durchführen. Eine interdisziplinäre Zusammenschau des Erdsystems und seiner Teilsysteme in ihrer Komplexität und die Einschätzung möglicher Entwicklungspfade sind nur modellgestützt möglich. Dies setzt hohe Maßstäbe an die Bewertung von Modellergebnissen (z.B. IPCC 2004) insbesondere vor dem Hintergrund weitreichender gesellschaftlichen Konsequenzen, die potenziell aus Modellanalysen gezogen werden könnten. In den letzten Jahren werden Unsicherheitsabschätzungen der Modellergebnisse immer mehr von Entscheidungsträgern und Politik nachgefragt (Webster 2003).

Im Bereich der Klimafolgenforschung findet man Modellimplementierungen sowohl in Programmiersprachen (C, C++, Fortran, Python) als auch in anderen Systemen wie Matlab, Mathematica oder z.B. für ökonomische Probleme in GAMS (<http://www.gams.com>). Die Modelle sind nichtlinear, produzieren teilweise umfangreichen multidimensionalen und multivariaten Output, manchmal im Gigabyte-Bereich, und haben oft lange Laufzeiten von Stunden bis zu Monaten. Wie häufig in

den Umwelt- und Geowissenschaften sind Teilmodelle in der Regel sehr gut untersucht, ein daraus gekoppeltes Gesamtmodell erfordert aber umfangreiche Validierungsstudien. Die darauf aufsetzenden Simulationsexperimente sind vorwiegend Szenario- und Re-Analysen, in denen unterschiedliche Entwicklungsoptionen für das Erdsystem bzw. ein Teilsystem studiert werden.

Damit ergeben sich die methodischen und praktischen Herausforderungen beim Experimentieren mit Simulationsmodellen in der Klimafolgenforschung u.a. aus der Notwendigkeit, (i) Betrachtungen der Modellsensitivität und -unsicherheit über hochdimensionalen Faktorräumen in die Simulationsstudien mit einzubeziehen (z.B. Murphy et al. 2004), (ii) komplexe Szenariostudien zu bewerten und einzelne Szenarien miteinander zu vergleichen, (iii) nichtlineare Modelle in unterschiedlichen Sprachen zu bedienen und (iv) große raum-zeitliche Modelloutputfelder über X_k nach dem Experiment inner- und außerhalb einer Simulationsumgebung auch visuell zu analysieren.

3 Anforderungen an Simulationsumgebungen

3.1 Experimentdesign

Simulationsumgebungen für Modelle der Klimafolgenforschung müssen Experimentier-techniken / -designs bereitstellen, die es erlauben, (i) ein grundlegendes Verständnis des betrachteten Modells und damit Systems zu erhalten, (ii) unterschiedliche Systemzustände und -entwicklungen miteinander zu vergleichen und (iii) robuste (und nicht unbedingt optimale) Entscheidungen auf der Basis des Modells zu treffen (Kleijnen et al. 2005). Während Punkt (i) Beziehungen zu Modellkalibrierung und -validierung aufweist, steht (ii) in unmittelbarer Verbindung zu Szenarioanalysen. Punkt (iii) dagegen führt zu Unsicherheits- und Sensitivitätsbetrachtungen. Unsicherheitsanalysen sind dabei das umfassendere Konzept zur Untersuchung unsicherer Modellentitäten wie z.B. Faktoren, Prozesse und Strukturen und ihres Einflusses auf den Modelloutput. In einer Sensitivitätsanalyse dagegen werden Unsicherheitsmaße am Modelloutput bestimmt und in Zusammenhang gebracht mit Unsicherheiten der Faktoren.

Auf Grund der speziellen Situation in der Klimafolgenforschung konzentrieren wir uns hier auf Experimentdesigns, die die drei obigen Anforderungen unterstützen. Sie repräsentieren nur einen Teil der Möglichkeiten, die für Design und Analyse von Computereperimenten (DACE) (z.B. Santner et al. 2003) zur Verfügung stehen. Die Experimentdesigns sollten damit die folgenden Forderungen erfüllen:

Modellunabhängige („model free“) Verfahren: Anwendbarkeit der Techniken auf die vorherrschende Klasse der nicht-additiven, nichtlinearen, nicht-monotonen Modelle

Deterministische und probabilistische Verfahren: Deterministische Ansätze für Modellvalidierung, Faktorinspektion und Szenarioanalysen durch blockstrukturiert-gitterte Designs über X_1 (one at a time OAT) und $X_{k>1}$ ((fractional) factorial design). Probabilistische Verfahren mit random und Latin hypercube sampling verschiedener Verteilungen für Monte Carlo-Analysen und damit verwandte Verfahren. Möglichkeit von Inputkorrelationen anstelle der Orthogonalität der Faktoren.

Lokale und globale Verfahren: Lokale Verfahren zur Identifikation der Modellsensitivität in der Nähe des aktuellen Arbeitspunktes in X_k , globale Verfahren zu deren Identifikation über dem gesamten zulässigen Inputraum X_k . Aggregationstechniken, um von vielen lokalen statistisch auf globale Eigenschaften zu schlussfolgern.

Anzahl der Faktoren: Auch bei sehr vielen Faktoren (großem k) sind in der Regel nur einige Faktoren wirklich wichtig für das Systemverhalten. Sequentielle Verfahren mit Gruppierung der Faktoren und globale Screeningverfahren ermöglichen die Bestimmung der wichtigsten Faktoren. Folgende Experimente können dann über einem Teilraum von X_k arbeiten.

Effiziente Verfahren: Die Anzahl der durchführbaren Läufe („computational costs“) ist für viele Simulationsstudien wegen der Modellaufzeit bzw. der zur Verfügung stehenden Hardwareressourcen limitiert. Deshalb sind insbesondere solche Experimententwürfe von Interesse, die mit einem geringen Umfang der Stichprobe $\{X_k\}$ durchgeführt werden können. Solch effiziente Designs setzen aber oft Bedingungen an die Modellklasse (Modellunabhängigkeit) bzw. an den Faktorraum (Orthogonalität).

Qualitative und quantitative Verfahren: Aus Effizienzgründen sind häufig bei hochdimensionalen Faktorräumen qualitative Verfahren, mit denen lediglich eine Rangfolge der Faktor-Sensitivitäten bestimmbar ist, den aufwändigeren quantitativen Verfahren vorzuziehen.

Von den probabilistischen Verfahren sind varianzbasierte Techniken von besonderem Interesse. Sie sind modellunabhängig und ermöglichen die Bestimmung von Sensitivitätsmaßen auf der Basis der Variation aller Faktoren durch Dekomposition der Varianz am Modelloutput y auf alle Faktoren: Wieviel Prozent der Variabilität am Output sind auf die Variabilität jedes einzelnen Faktors zurückzuführen? In den letzten Jahren sind Designs entwickelt worden, die ohne Einschränkung der Anforderungen an das Modell die Anzahl der Läufe von $\sim k \cdot N$ auf N verringern, wobei N die Anzahl der Läufe eines Monte Carlo-Ensembles ist.

Zu den angeführten Verfahren und weiterführenden Literaturangaben siehe z.B. Saltelli et al. (2004).

3.2 Offene Architektur

Ebenso wichtig wie die Bereitstellung von Experimententwürfen sind für Simulationsumgebungen – wie für andere Software-Tools auch – deren offener Entwurf und dessen Implementierung. Insbesondere umfasst dies die folgenden Punkte:

Modellschnittstelle: Über die Modellschnittstelle müssen die numerischen Werte der Faktoren in das Modell importiert werden. Wenn Modelle außerhalb der Umgebung entwickelt werden, muss die Schnittstelle die verwendete Modellierungssprache unterstützen. Zusätzlich ist der native Modelloutput in solche Datenstrukturen zu transformieren, dass der Experimentoutput innerhalb der Umgebung analysiert werden kann. Oft wird in der Klima- und Klimafolgenforschung Modelloutput im selbstbeschreibenden NetCDF-Format ausgegeben.

Datenschnittstelle: Betrifft vorrangig die Unterstützung von Datenformaten, in denen zusätzliche Daten (z.B. Messwerte des Realsystems) für die Experimentanalyse importiert und Daten nach der Experimentanalyse exportiert (z.B. zu Statistikpaketen oder Visualisierungstools) werden können.

Lastverteilung: Die meisten Experimentdesigns ermöglichen die Ziehung der Stichprobe vor dem Start des Experimentes und erfordern wie z.B. bei Optimierungsverfahren nicht die Analyse des aktuellen Simulationslaufes, um einen nächsten Punkt in X_k bestimmen zu können. Damit sind alle Simulationsläufe vor Experimentbeginn bestimmt und unabhängig voneinander, und das Experiment an sich kann parallelisiert werden, wenn z.B. ein Compute-Cluster zur Verfügung steht.

Experimentanalyse: Viele der adressierten Modelle geben den Output granular aufgelöst aus und die eigentlichen Zielgrößen müssen nach dem Experiment abgeleitet werden. Die Experimentanalyse ist so flexibel anzulegen, dass durch Operatoren S vom Experimentoutput $\{y\}$, Referenzdaten r und Output e anderer Experimente ein sekundärer Output $\{S(y,r,e)\}$ abgeleitet werden kann und die eigentliche Analyse darauf angewendet wird: $G_D(\{S(y,r,e)\})$. Für die explorative Analyse des multidimensionalen Experimentoutputs ist dabei neben der statistischen Analyse die Visualisierung eine Schlüsseltechnologie.

Inhaltlich intuitiver Zugang, Nutzerinterface, Hilfesystem: sind wesentliche Komponenten nutzerfreundlicher Software-Tools

4 Ergebnisse einer Umfrage zum Experimentieren mit Modellen

Um einen Überblick über die Anforderungen für Datenspeicherung und -verwaltung, Experimentieren mit Modellen und der Visualisierung von Daten im Bereich der Klimafolgenforschung zu erhalten, wurden standardisierte Interviews mit 76 Wissenschaftlern (33 Doktoranden, 28 Postdocs, 15 Senior Researchers) am Potsdam-Institut für Klimafolgenforschung geführt. Sie vertreten hauptsächlich die Fächer Me-

teologie, Klimatologie, Ozeanographie, Hydrologie (24), Ökonomie und Sozialwissenschaften (30), Ökologie und Biologie (27), Physik (14) und Geographie und Geostatistik (14) (Mehrfachnennungen waren möglich). Als Antworten auf die Fragen waren „häufig“, „manchmal“, „selten“ und „nie“ vorgegeben. Die nachfolgenden Befragungsergebnisse beziehen sich auf die 64 Befragten, die mit Modellen arbeiten.

- 55% der Nutzer verwenden Modelle, die in Fortran implementiert sind, 36% in C, 31% in C++, 30% in Matlab, 23% in GAMS, 5% in Python, 3% in Java, 2% in Mathematica (Mehrfachnennungen waren möglich).
- 66% der Befragten nutzen manchmal oder häufig deterministische und 25% probabilistischen Sampling-Verfahren. Bei Letzteren sind zufällige und Latin hypercube Stichproben gleichermaßen vertreten.
- Bei den Experimentiertechniken werden gegitterte Entwürfe sowohl für den eindimensionalen Fall (OAT) als auch für mehrdimensionale Anwendungen ($k>1$) von jeweils 49% der Nutzer manchmal oder häufig verwendet. Unerwartet hoch war die Anwendung von lokalen Verfahren zur Modellsensitivität durch 50% der Umfrageteilnehmer. Entsprechende analytische Verfahren und Techniken zur Generierung adjungierter Modelle durch algorithmisches Differenzieren (Griewank 2000) kommen dagegen nur selten bzw. nie zum Einsatz. In der Rangfolge der eingesetzten Experimentiertechniken folgen Monte Carlo-Analysen bei 27% der Nutzer, gefolgt von 13%, die manchmal oder häufig Techniken für eine globale Sensitivitätsanalyse verwenden. 20% der Befragten nutzen ein- und multikriterielle Optimierungsverfahren über ein- oder mehrdimensionalen Faktorenräumen.
- Die Umfrage bestätigte erwartungsgemäß, dass viele Wissenschaftler mit Simulationsmodellen experimentieren. Simulationsumgebungen dagegen werden nur sehr zögerlich genutzt. Lediglich 24% der Befragten setzen solche Tools ein, davon einige SimLab und alle anderen SimEnv (siehe unten). 44% wenden im Umfeld von Simulationsexperimenten Matlab oder Statistikpakete an, 47% setzen eigene Software vorrangig zur Datenintegration und Experimentauswertung ein, 7% sehen Potenzial, diese Software breiter verfügbar zu machen.
- 84% der Befragten nutzen Modelle, bei denen der Output in ASCII-Files abgelegt wird, 50% in Binärfiles, 29% in selbstbeschreibenden Datenformaten (NetCDF, HDF) und 8% im Grib-Format für meteorologische und Klimadaten. 8% der Nutzer transformieren ihren Modelloutput in Formate von Microsoft-Office (z.B. Excel) und analysieren ihn dort (Mehrfachnennungen waren möglich). Eine Direktausgabe in Datenbanken oder in XML-Files ist in keinem der Modelle implementiert.
- Zu den Ergebnissen der Umfrage in Bezug auf den Einsatz von Visualisierungstechniken und -produkten siehe Nocke et al. (2008).

Die Umfrage zeigt klar, dass für eine Vielzahl von Modellen, die in unterschiedlichen Sprachen implementiert sind, diejenigen Experimentiertechniken Anwendung finden, die besonders gut für Modellvalidierung, Szenario-, Unsicherheits- und Sensitivitäts-

analysen geeignet sind. Bei Sensitivitätsuntersuchungen dominieren die lokalen Methoden, die Anwendung entsprechender qualitativer und quantitativer globaler Methoden ist dagegen unterrepräsentiert. Die Umfrage dokumentiert gleichzeitig aber auch die Situation, dass Simulationsumgebungen zum Experimentieren nur zögerlich eingesetzt werden, obwohl sie nicht nur adäquate Methoden bereitstellen sondern auch zu einer effektiven Arbeit im Simulationsumfeld beitragen können.

5 Multi-Run Simulationsumgebung SimEnv

SimEnv (Flechsig et al. 2005 und 2008) ist eine Umgebung für Multi-Run Ensemble Simulationsexperimente, die vorrangig für Unsicherheits-, Sensitivitäts- und Szenarioanalysen mit Multiinput - Multioutput - Simulationsmodellen entwickelt wurde. Mit SimEnv werden hochdimensionale Faktorräume und Modelle mit großem Output bis in den Gigabyte-Bereich unterstützt. Im Vergleich mit anderen Umgebungen, die im wissenschaftlichen Umfeld entwickelt wurden und dort auch angewendet werden (Tab. 1), zeichnet sich SimEnv durch folgende Besonderheiten aus, die mit dazu beitragen, dass die Simulationsumgebung für typische Aufgaben von Modellentwicklern und -nutzern in der Klima- und Klimafolgenforschung flexibel eingesetzt werden kann:

Tabelle 1. Simulationsumgebungen für Unsicherheits- und Sensitivitätsanalysen

Simulationsumgebung	URL
SimLab	http://simlab.jrc.ec.europa.eu/
DAKOTA	http://www.cs.sandia.gov/DAKOTA/
UNCSAM	http://www.rivm.nl/bibliotheek/

- **Modellschnittstelle:** Minimale Änderungen im Quellcode von C/C++, Fortran, Python, GAMS, Matlab und Mathematica Modellen und für Shellskripte gestatten die direkte Kopplung dieser Modelle an SimEnv: Pro Faktor, mit dem experimentiert werden soll, wird im Quellcode ein SimEnv-Funktionsaufruf eingefügt, um den aktuellen Wert des Faktors aus dem Sample im Modell verfügbar zu machen. Dies gestattet, das Modell direkt in SimEnv zu betreiben. Die Umgebungen aus Tab. 1 (SimLab in Vers. 2) stellen einen ASCII Sample-File bereit und das Experiment muss außerhalb der Umgebung durchgeführt werden und auf diesen zugreifen. SimLab Vers. 3 dagegen ist ein Software Development Kit für C/C++, Fortran und Matlab und gestattet, SimLab-Module direkt in Nutzersoftware einzubetten.
- **Experimentdesign:** SimEnv stellt grundlegende Experimentdesigns zur Verfügung (Tab. 2). Damit sind wesentliche Aufgaben für Unsicherheits- und Sensitivitätsanalysen durchführbar. Insbesondere für probabilistische Verfahren können

Samples aus anderen Umgebungen nach SimEnv importiert und so die Stärken z.B. von SimLab und DAKOTA in diesem Bereich genutzt werden. SimEnv-Experimente werden strukturiert in ASCII Files beschrieben (Tab. 3). SimEnv ist so offen konzipiert, dass zusätzliche Experimenttypen / -designs leicht implementiert werden können.

- **Lastverteilung:** SimEnv stellt Dienste bereit, um das Experiment auf einem Compute-Cluster mit dem Message Passing Interface MPI oder auf einer Multicore-Prozessor Maschine nach verschiedenen Strategien (dynamische Lauf-Verteilung in Abhängigkeit von der Maschinenauslastung oder Vorab-Allokation von Rechenkernen) zu parallelisieren. Ein automatischer und manueller Restart von Experimenten und die Abarbeitung von Teilerperimenten sind möglich.

Tabelle 2. Experimentdesigns in SimEnv

Design	Methoden
Deterministisches Faktorscreening	Blockstrukturiert-gegridete Experimentierpläne für mehrdimensionale Faktorräume X_k und deren Hyperräumen mit flexiblen Durchmusterungs- und Analysestrategien
Lokale Sensitivitätsanalyse	Faktorbezogenes lokales Sampling in der Umgebung des aktuellen Arbeitspunktes zur Bestimmung lokaler Sensitivitätsmaße erster Ordnung
Globale Sensitivitätsanalyse	Qualitatives Verfahren nach Morris zur Bestimmung der sensitivsten Faktoren. Varianzdekomposition am Modelloutput nach Saltelli et al. (2004) zur Bestimmung von Sensitivitätsmaßen erster Ordnung und der totalen Effekte
Unsicherheitsanalyse	Monte Carlo-Analysen für multidimensionale Faktorräume, Bestimmung von statistischen Maßen über dem Run-Ensemble
Optimierung	Stochastische Samplingstrategie zur Bestimmung des globalen Minimums einer Kostenfunktion über X_k durch einen Simulated-Annealing-Algorithmus

- **Experimentanalyse:** SimEnv unterstützt multivariaten und multidimensionalen Modelloutput. Für jedes Outputfeld y , das nach dem Experiment analysiert werden soll, wird im Quellcode ein SimEnv-Funktionsaufruf eingefügt und damit die Felder eines Modelllaufs in einem NetCDF-File abgespeichert. Während der Experimentanalyse wird auf diese Files zugegriffen. Daraus kann interaktiv mit SimEnv-Operatoren sekundärer Experimentoutput $\{S(y,r,e)\}$ und schließlich mit Design-spezifischen Operatoren G_D Maße $G_D(\{S(y,r,e)\})$ abgeleitet werden. SimEnv stellt über 100 Standardoperatoren bereit, zusätzliche Nutzer-definierte Operatoren können integriert werden. Die Operatoren ermöglichen u.a. die Navigation über dem gekoppelten Faktor-Output-Raum, den Vergleich mit externen Daten und anderen SimEnv-Experimenten sowie Aggregation und Transformation des Experimentoutputs. In den Simulationsumgebungen aus Tab. 1 (mit SimLab

in Vers. 2) hat der Nutzer dafür zu sorgen, dass (meist skalarer) Modelloutput in einem bestimmten Format dem Analysesystem zur Verfügung gestellt wird.

- **Visuelle Experimentanalyse und -bewertung:** Um die Heterogenität an Daten, die in Simulation und Experimentanalyse erzeugt werden können, handhabbar zu machen, wurde für SimEnv eine speziell hierfür zugeschnittene Visualisierungskomponente SimEnvVis entworfen und umgesetzt. Diese Komponente unterstützt den Experimentator, basierend auf Metadaten aus einer Menge von interaktiven Visualisierungstechniken eine für die aktuelle Aufgabe geeignete Technik auszuwählen (siehe auch Nocke 2007). So lassen sich sowohl einzelne Modellläufe als auch Simulationsensembles analysieren. Abb. 2 zeigt Screenshots des SimEnv-Vis-Wizards, welche die Nutzerunterstützung bei der Auswahl und Parametrisierung von Visualisierungstechniken verdeutlichen. Darüber hinaus wurden Techniken aus dem Bereich der Informationsvisualisierung eingebunden und für die speziellen Aufgaben bei der Darstellung von Simulation-Ensembles angepasst. Die Herausforderung hierbei besteht darin, die Abhängigkeiten von multidimensionalem Modellinput zu multivariatem Modelloutput geeignet bei einer Vielzahl von Modellläufen für den Experimentator visuell erfassbar sowie die zum Teil großen Datenmengen interaktiv verarbeiten zu können. Insbesondere werden hierzu die Techniken „parallele Koordinaten“ sowie „graphische Tabelle“ (Abb. 3) eingesetzt.

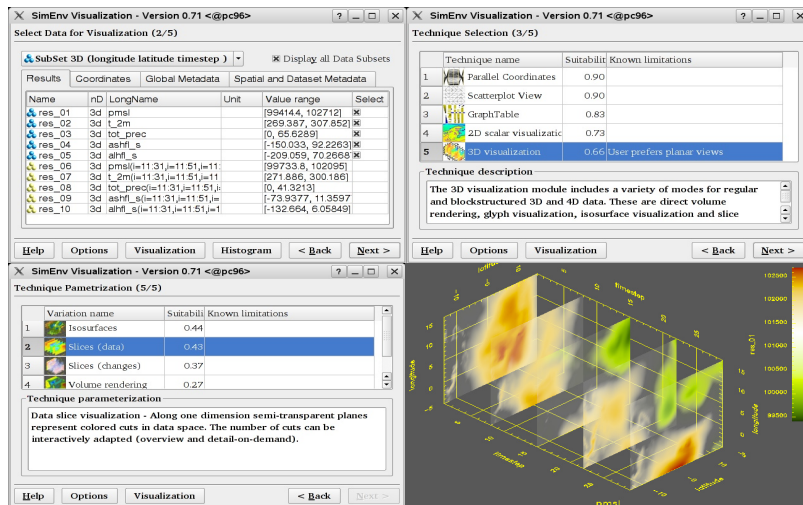


Abbildung 2. SimEnvVis-Wizard: Darstellung von NetCDF-Metadaten sowie Variablenauswahl (links oben); Auswahl von Visualisierungstechniken (rechts oben); Auswahl von Visualisierungstechnikvarianten (links unten); resultierende 3D-Schnittdarstellung (zeitliche Veränderung einer Variablen, Hervorhebung von Extremwerten unter Einsatz von Transparenz) (rechts unten)

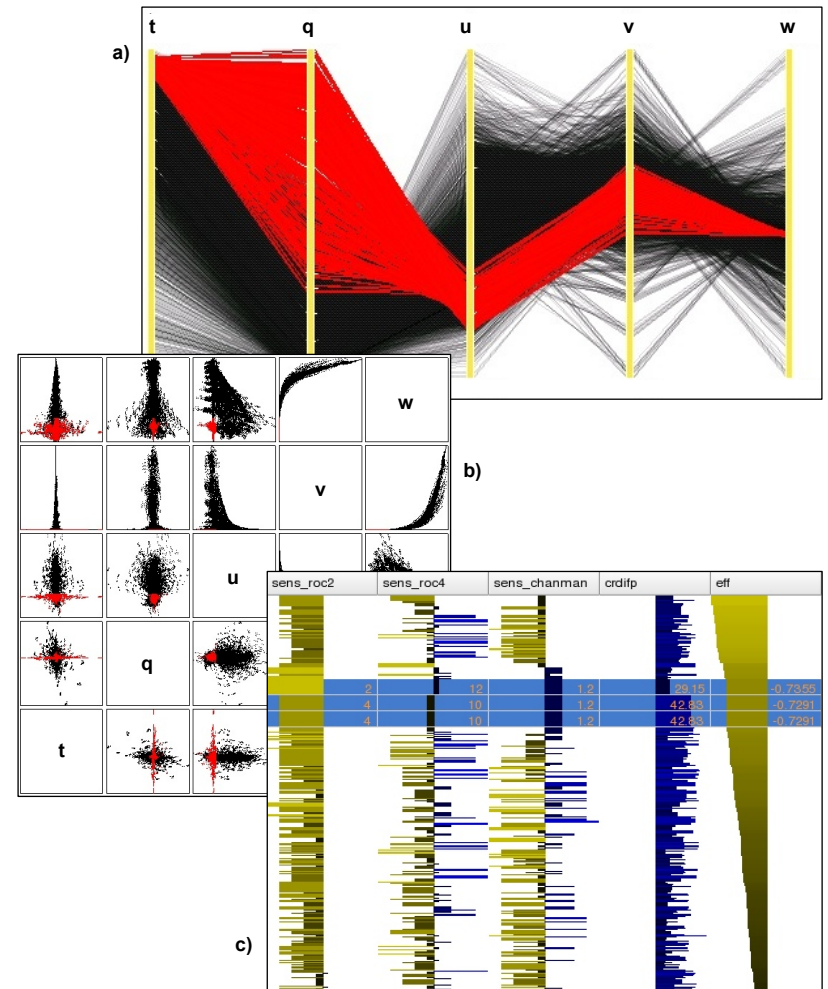


Abbildung 3. Beispiele für die Visualisierung von multivariatem und -dimensionalem Experimentoutput. a) und b): fünf 3-dimensionale Variablen eines globalen Atmosphärenmodells; Repräsentation der Variablen mit paralleler Koordinatentechnik (a) und Scatterplot-Matrix-Technik (b) c): Modelloutputs crdifp und eff eines regionalen hydrologischen Modells in Abhängigkeit von drei Faktoren sens_* einer globalen Sensitivitätsanalyse mit 840 Single Runs in einer graphischen Tabelle, sortiert nach der Variablen eff

SimEnv wird von Modellentwicklern und -nutzern zur Modellvalidierung, für Unsicherheitsanalysen (Bauer et al. 2005, Knopf et al. 2006), Sensitivitätsanalysen (Knopf et al. 2008) und für Szenarioanalysen eingesetzt. Insbesondere bei Monte Carlo-Analysen mit Erdsystemmodellen mittlerer Komplexität (Schneider von Deimling et al. 2006) zur Bestimmung der Klimasensitivität unter Änderung der globalen CO₂-Konzentration (~5000 Modellläufe, Modelllaufzeit im Stundenbereich, Modelloutput ~30 MByte) hat sich die dynamische Lastverteilung des Experiments auf Compute-Clustern als hilfreich erwiesen.

Tabelle 3. Definition eines Experiments in SimEnv am Beispiel eines deterministischen Faktorscreenings für drei Faktoren. Eine alternative Durchmusterungsstrategie auf der x1-x2-Diagonalen zusammen mit x3 (fractional factorial design) wird durch `comb (x1,x2)*x3` für 10*20 Läufe definiert.

general	type	behaviour	# factorial design
general	descr	...	# optional experiment description
factor	x1	descr	...
factor	x1	unit	...
factor	x1	default	1.
factor	x1	sample	list 1,3,4,5,7,8,9,10,11,12
factor	x2	default	2.
factor	x2	sample	equidist_end 1(1)10
factor	x3	default	3.
factor	x3	sample	file 20values.dat
specific	comb	x1*x2*x3	# full factorial design in X ₃ for # 10*10*20 runs

6 Diskussion und Schlussfolgerungen

Simulationsumgebungen zum Experimentieren mit Simulationsmodellen, die außerhalb der Umgebung entwickelt wurden, unterliegen neben den methodischen Anforderungen zusätzlichen Akzeptanzkriterien. Dabei erweist sich die Modellschnittstelle zum Import der Faktorenwerte und deren direkter Adressierung im Modellcode als eine wichtige Eigenschaft solcher Systeme. Hinzu kommt, dass der native Modelloutput in solche Datenstrukturen überführt werden muss, die vom Analyseteil der Experimentierumgebung weiterverarbeitet werden können bzw. dass solche Datenstrukturen direkt vom Modell generiert werden müssen. Typische Modelle in der Klima- und Klimafolgenforschung liegen in unterschiedlichen Sprachen vor, haben oft lange Laufzeiten und produzieren umfangreichen multivariaten und multidimensionalen Modelloutput. Schwerpunkte der Modellnutzung sind dabei Unsicherheits-, Sensitivitäts- und Szenarioanalysen über hochdimensionalen Faktorräumen.

Die Umfrage unter Modellierern und Modellnutzern aus diesem Wissenschaftszweig unterstreicht den Bedarf an und die Anwendung von deterministischen und probabilistischen Techniken für Experimententwurf und -analyse. Deutlich wird aber auch, dass in nur geringem Maße direkt Simulationsumgebungen eingesetzt werden, sondern häufig auf eigene Software, Matlab oder Statistikpakete für Experimententwurf, Datenintegration und Experimentanalyse zurückgegriffen wird. Vor diesem Hintergrund sind Simulationsumgebungen wichtig, die grundlegende Techniken für die o.a. Analyseziele bereitstellen und gleichzeitig so offen sind, dass leicht zu implementierende Schnittstellen für Modelle, Referenzdaten und Output anderer Experimente verfügbar sind. Hinzu kommt auf Grund der Komplexität des Modelloutputs die direkte Einbindung von Visualisierungsmethoden in den Analyseprozess zur visuellen Bewertung von Experimentoutput und davon abgeleiteten Maßen.

Mit SimEnv steht eine Simulationsumgebung zur Verfügung, die diese Anforderungen erfüllt. Im Vergleich mit ähnlichen Tools liegen ihre Vorteile auch in der Unterstützung selbstbeschreibender Datenformate, der Experimentdurchführung direkt in der Umgebung, deren Parallelisierung sowie generell in der offenen Architektur des Systems. Die angeführten Beispiele machen deutlich, dass für die Auswertung von Modellsimulationen in der Klima- und Klimafolgenforschung Visualisierungstechniken ein unabdingbarer Dienst von Simulationsumgebungen sind.

Literaturangaben

- Bauer, N., Edenhofer, O., Held, H., Kriegler, E. (2005): Uncertainty of the role of carbon capture and sequestration within climate change mitigation strategies. In: Rubin, E.S., Keith, D.W., Gilboy, C.F. (eds.): Proceedings of the 7th International Conference on Greenhouse Gas Control Technologies (5-9 September 2004, Vancouver, Canada), 931-939. Elsevier, Amsterdam
- Box, G.E.P., Hunter, W.G., Hunter, J.S. (1978): Statistics for experimenters: An introduction to design, data analysis and model building. Wiley, New York
- 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.): Proceedings of the 4th International Conference on Sensitivity Analysis of Model Output (SAMO 2004). Los Alamos National Laboratory, Los Alamos, NM, 297-306
- Flechsig, M., Böhm, U., Nocke, T., Rachimow, C. (2008): The multi-run simulation environment SimEnv. User's Guide. Potsdam Institute for Climate Impact Research, Potsdam, Germany
<http://www.pik-potsdam.de/software/simenv/>
- Griewank, A. (2000): Evaluating derivatives: Principles and techniques of algorithmic differentiation. Series Frontiers in Applied Mathematics, SIAM, Philadelphia
- IPCC (2004): Guidance notes for lead authors of the IPCC Fourth Assessment Report on addressing uncertainties. IPCC, Geneva
- Kleijnen, J.P.C., Sanchez, S.M., Lucas, T.W., Cioppa, T.M. (2005): A user's guide to the brave new world of designing simulation experiments. *Inform Journal of Computing*, 17(3), 263-289
- Knopf, B., Flechsig, M., Zickfeld, K. (2006): Multi-parameter uncertainty analysis of a bifurcation point. *Nonlinear 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

- Murphy, J.M., Sexton, D.M.H., Barnett, D.N., Jones, G.S., Webb, M.J., Collins, M., Stainforth, D.A. (2004): Quantification of modeling uncertainties in a large ensemble of climate change simulations. *Nature*, 430, 768-772
- Nocke, T. (2007): *Visuelles Data Mining und Visualisierungsdesign für die Klimaforschung*, Dissertation, Universität Rostock
- Nocke, T., Sterzel, T., Böttiger, M., Wrobel, M. (2008): Visualization of climate and climate change data: An overview. In Ehlers, M. et al. (eds.): *Digital Earth Summit on Geoinformatics 2008: Tools for Global Change Research (ISDE'08)*, Wichmann, Heidelberg, 226-232
- Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M. (2004): *Sensitivity in practice: A guide to assessing scientific models*. Wiley & Sons
- Santner, T.J., Williams, B.J., Notz, W.I. (2003): *Design and analysis of computer experiments*. Springer, New York, Berlin, Heidelberg
- Schneider von Deimling, T., Held, H., Ganopolski, A., Rahmstorf, S. (2006) Climate sensitivity estimated from ensemble simulations of glacial climate. *Climate Dynamics* 27, 149-163
- Webster, M. (2003): Communicating climate change uncertainty to policy-makers and the public. *Climatic Change*, 61(1-2), 1-8