

Visual Analytics zur Unterstützung von Multi-Agenten-Simulationen

Janus Dybulla
janus-dybulla@haw-hamburg.de
Hamburg University of Applied Sciences,
Dept. Computer Science,
Berliner Tor 7
20099 Hamburg, Germany
<http://www.haw-hamburg.de>

Zusammenfassung. Wissenschaftler aus Bereichen wie der Ökologie, Medizin oder Biologie benutzen multivariate Multi-Agenten-Simulationen um komplexe Systeme zu beschreiben und aus ihnen neue Erkenntnisse zu gewinnen. Die resultierenden, stetig ansteigenden und heterogenen Simulationsdaten, sowie die erhöhten Erwartungen und Anforderungen der Modellierer an solche Systeme, erfordern neue Ansätze. Visual Analytics bietet als ein interdisziplinärer Ansatz die Kombination aus Datenanalysemethoden, interaktiven Visualisierungskonzepten und menschlichem Abstraktionsvermögen. Diese Arbeit gibt einen Einblick, wie durch Visual Analytics agentenbasierte Simulationen validiert, analysiert und explorativ visualisiert werden können.

Schlüsselwörter: visual analytics, model-agent based simulation, big data, data analysis, information visualisation

1 Einleitung

Neue Möglichkeiten in der Agentenmodellierung und zur Bewältigung von facettenreichen Problemen wurden durch Multi-Agenten-Simulationen (*MAS*) geschaffen. Während traditionelle Simulationssysteme auf bestimmte Phänomene beschränkt sind, bieten *MAS* effektive Konzepte der Emergenz, Adaption und Selbstorganisation. Durch sie lassen sich verschiedene Verhaltensmuster und -ergebnisse von beispielsweise Klima-, Landnutzungs- und Populationsänderungen erforschen, den benutzten Modellfaktoren verschiedenen Bedeutungen und Auswirkungen zuordnen, sowie potentielle Systemwechsel identifizieren [RAC⁺13]. Sie ermöglichen Wissenschaftlern während einer Simulation die „Was dann“-Szenarien bei Veränderungen zu erkunden. *MAS* repräsentieren eine große Vielzahl an (teil-)autonomen Entitäten oder Agenten, bei denen jeder Agent entsprechend einer Zusammenstellung an vorgeschriebenen Regeln und Zielen mit anderen Agenten interagiert und über die Zeit hinweg von nahestehenden Agenten und seiner Umgebung beeinflusst werden kann [Hel12a].

Durch die kontinuierliche Steigerung der Computer- und Speicherleistung können Modellierer von MAS noch weit aus anspruchsvollere Simulationen mit verteilten, skalierbaren Systemen verwenden, bei denen die resultierenden Simulationsdaten an Komplexität und Fülle drastisch zunehmen und die Voraussetzungen für die Big Data Herausforderung schaffen. Um diese zu meistern und Erkenntnisse aus den Daten zu gewinnen, gibt es verschiedene numerische und algorithmische Ansätze aus den Bereichen Data Mining, Knowledge Discovery und Machine Learning. Auch wenn diese Methoden in vielen Bereichen der Computerwissenschaften effektiv eingesetzt werden, stoßen sie besonders bei der automatischen Auswertung von multidimensionalen und heterogenen Daten an ihre Grenzen [SWLL13]. Iterative Anpassungen an Simulationen und Methoden werden hierdurch nicht in Betracht gezogen. Es ist schwer, interessante Muster auf eine intuitive und aussagekräftige Art zu verdeutlichen [KKP⁺11]. Es müssen daher effektive Wege gefunden werden, um unterschiedliche Attribute und Verhaltensweisen eines Modells sowie Simulationen zwischen verschiedenen Simulationsläufen und über verschiedene Simulationsläufe hinweg vergleichen zu können [MDSC03].

Zur Beantwortung dieser Probleme wurden neue Denkansätze und Methoden geschaffen, die die automatische Analyse mit interaktiven Visualisierungen zu einer visuell analytischen Kombination vermischen und den Menschen in den Vordergrund setzen [SRS⁺13]. Visual Analytics unterscheidet sich daher von der gewöhnlichen Visualisierung, da bei der Visual Analytics menschliche, kognitive Faktoren und Datenanalysemethoden miteinander verknüpft werden. Durch die aktive Rolle des Menschen in der Visual Analytics wird die Datenanalyse zu einem aktiven Prozess unter Einbeziehung der Echtzeitinteraktionen zwischen Computer und Benutzer. Interaktive, visuelle Abfragen und dynamische Datenübersichten verbinden den Wissenschaftler mit den Daten hinter der Visualisierung [TC05]. Unter anderem finden sich verschiedene Forschungsrichtungen in dieser neuartigen Disziplin wieder: Informationsvisualisierung, Data Mining, Datenmanagement, Statistik und menschliche Kognition und Wahrnehmung.

Im Rahmen der MARS Group (Multi Agent Research and Simulation) der HAW Hamburg¹ wird ein MAS-Framework entwickelt, welches mit einer aktuellen Forschungsarbeit über das Ökosystem der Savanne im Krüger-Nationalpark Lösungen aus der Visual Analytics fordert.

Savannen bestehen aus der Koexistenz von Holzgewächsen und Gräsern und bilden ein fragiles Gleichgewicht, welches vom Vorhandensein von Wasservorkommen, Feuer, Nährstoffen, Menschen und vor allem Tiere beeinflusst wird. Aktuelle Forschungen belegen unter den gegenwärtigen Klimabedingungen und der Elefantenpopulation die wesentlichen ökologischen Effekte von afrikanischen Elefanten auf die Verhältnisse in der Savanne [BG05]. Basierend auf diesen Erkenntnissen wollen Wissenschaftler wissen, inwieweit eine rasch wachsende Elefantenpopulation konditioniert werden muss und wie ein Naturschutzpark auf

¹ Weitere Information über MARS siehe [TC15], <http://www.mars-group.mars.haw-hamburg.de>

Naturereignisse und Verhaltensweisen von Menschen und Tieren angepasst werden kann [SH12]. Mit Hilfe von MAS und visuellen Analyseverfahren können diese Simulationsmodelle nachgestellt, bewertet, optimiert und mit empirisch erhobenen Daten verglichen werden.

Nachfolgend wird das Themengebiet und wichtige Errungenschaften der Visual Analytics vorgestellt. Anforderungen von Multi-Agenten-Simulationen werden in diesem Kontext beschrieben und es wird untersucht, wie die MAS mit dem neuen Ansatz von Visual Analytics erweitert werden können. Anhand aktueller Forschungsarbeiten werden exemplarisch einige Lösungen erörtert. Abschließend finden sich im Ausblick die Ziele für das Grundprojekt im Rahmen von MARS.

2 Visual Analytics

Visual Analytics verfolgt das Ziel die explorative Erforschung von Big Data zu unterstützen. Der Forschungsbereich der Visual Analytics motiviert sich wesentlich aus dem Problem der sogenannten Informationsüberlastung, d.h. der Mensch ist aufgrund von zu vielen Informationen manchmal nicht mehr in der Lage diese zu verarbeiten, zu verstehen und Entscheidungen zu treffen [LSFK12].

Der Lösungsansatz für dieses Problem liegt hier in der Verschmelzung von Big Data Analyseprozessen mit der menschlichen, intuitiven Wahrnehmung von interaktiven Visualisierung [Kei10]. Die Kombination beider liegt darin begründet, dass das Datenvolumen für eine direkte (nicht vereinfachte) Darstellung oft viel zu groß ist. Folglich bedient man sich an Methoden, wie beispielsweise des Dataminnings, um die Daten anschließend adäquat visualisieren zu können. Jedoch sind die erforderlichen Aufgaben oft zu komplex für rein analytische Methoden und benötigen eine menschliche Komponente in Form von Analysten, die mittels Visualisierungen mit den Daten interagieren können. Folglich bedarf dieser Lösungsansatz ebenfalls des Verständnisses menschlicher, kognitiver Denk- und Analyseprozesse. Damit erfordert das Fachgebiet der Visual Analytics einen interdisziplinären Ansatz [EHR⁺14].

Heute genießt die Visual Analytics als wachsendes Forschungsgebiet internationale Anerkennung und bringt zunehmend anspruchsvollere Technologien hervor. Ihre Wurzeln gehen auf die Community der Informationsvisualisierung zurück, die mindestens seit den 1980ern besteht [KPG12]. Im Laufe des folgenden Jahrzehnts haben Forscher der Computerwissenschaften, damit eingeschlossen Datenmanagement, Mensch-Computer-Interaktion, Informationsrückgewinnung und Computergrafik, die Gemeinsamkeiten zwischen ihren jeweiligen Forschungsfeldern und der Informationsvisualisierung entdeckt. Die verschiedenen Gemeinschaften fanden sich anschließend im *National Visualization and Analytics Center*-Projekt (NVAC) zusammen und legten die Grundlage zur Entstehung der Visual Analytics in der Agenda *Illuminating the Path* [TC05] fest. Das

von der Europäischen Union geförderte Projekt VisMaster² führte die Agenda fort und wurde hauptsächlich von den Universitäten Konstanz, Rostock und Stuttgart sowie den Institutionen der Fraunhofer IGD und IASIS betreut [Kei10]. Verschiedene Konferenzen haben sich dem Themenfeld der Visual Analytics gewidmet, so beispielsweise die seit 1994 bestehende *Visualization and Data Analysis* (VDA)³, die *Big Data Visual Analytics* (BDVA)⁴ sowie die größte Forschungskonferenz *IEEE VIS*⁵, welche aus den drei Bereichen der *Information Visualization* (InfoVis), *Scientific Visualization* (SciVis) und schließlich der *Visual Analytics Science and Technology* (VAST) besteht. Letztere beinhaltet zudem die populären *VAST Challenges*⁶. Bei diesem Wettbewerb nehmen Teams aus der Forschung und Industrie teil und widmen sich Aufgabenstellungen, die ohne den Einsatz von Visual Analytics-Methoden schwierig zu lösen sind.

Beispielsweise können automatisierte Algorithmen in die Irre führen, wenn die zu untersuchenden Hypothesen zu Beginn schlecht definiert oder die Algorithmen auf ein vage definiertes Probleme angewandt werden. Die Beziehung zwischen den Eingabe- und Ausgabedaten der Algorithmen sind für den Analytikern oft nicht schlüssig, so dass man der Analyzelösung blind vertrauen muss und die Nachvollziehbarkeit hinterfragt wird [KMT10]. Die Idee besteht also darin, solche Ansätze mit Visualisierungen zu verknüpfen, um beispielsweise die Festlegung von Parametern in verschiedenen Schritten eines Data-Mining-Algorithmus zu kontrollieren. Durch die interaktive und visuelle Untersuchung der Originaldaten sowie abgeleiteter Eigenschaften soll das Erwartete erkennbar und das Unerwartete entdeckbar sein [TC05]. Wissenschaftler sollen Schlussfolgerungen ziehen und Hypothesen auf der Grundlage der visuellen Informationen erzeugen können, sowie die Ergebnisse des analytischen Argumentationsprozesses kommunizieren und präsentieren können [KH13]. Keim et al. [KAF⁺08] schlägt eine entsprechende Abänderung und Erweiterung des „Informationssuche“-Mantras von Shneidermann⁷ für Visual Analytics vor: „Analyze first, show the Important, zoom, filter and analyze further, details-on-demand.“

Um das Datenvolumen in der ersten Phase der Analyse zu reduzieren, kommen Analyseverfahren zum Einsatz, die sinnvolle Abstraktionen der Daten durch beispielsweise zeitliche oder strukturelle Cluster bilden. Nach einer ersten abstrahierten Visualisierung der Originaldaten können weitere Daten, die von Interesse sind, für eine gründlichere Überprüfung ausgesucht werden. Es besteht jedoch die Möglichkeit, dass eine einzelne Auswahl nicht genügt, um die Daten für die visuelle Darstellung ausreichend zu reduzieren, sodass weitere Abstraktionen berechnet werden müssen. Weitere Details können anschließend nach Bedarf durch mehrfache Iteration der zweiten eben beschriebenen Phase visualisiert und somit genauer analysiert werden [HSS11]. Weitere analytische Funktionen der Visual

² Ursprünglich <http://vismaster.eu>, aktuell <http://www.Visual-Analytics.eu>

³ <http://vda-conference.org>

⁴ <http://www.bdva.net>

⁵ <http://ieevis.org>

⁶ www.vacommunity.org/

⁷ „Overview first, zoom and filter, then details-on-demand.“ [Shn96]

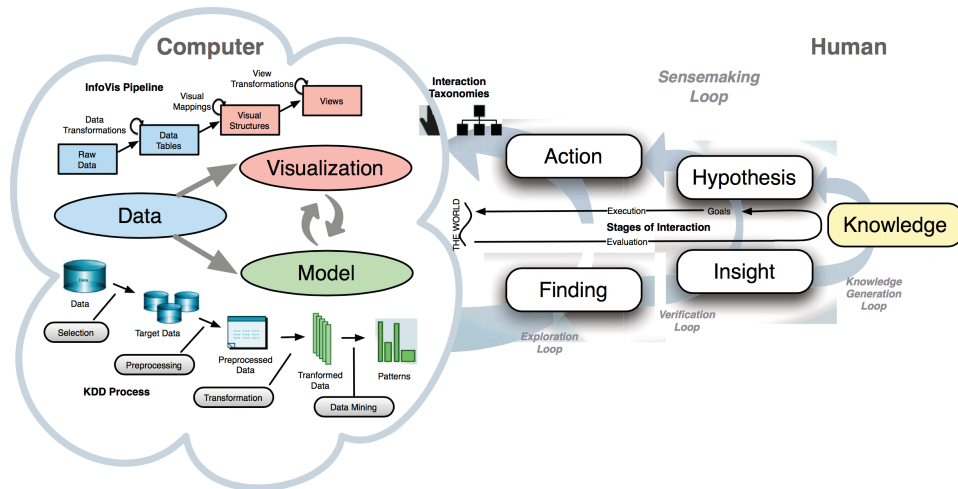


Abb. 1. Prozess-Modell zur Wissensbildung in Visual Analytics. Farben und Positionierungen im Bild verdeutlichen die Gemeinsamkeiten. Abbildung entnommen aus [SSS⁺14].

Analytics, wie beispielsweise der kognitive Wahrnehmung- oder Entscheidungswissenschaften spielen dabei eine zentrale Rolle [Kei10]. Aktuelle Arbeiten betonen, dass die Theorie der Visual Analytics über die Denkweise „Mensch in der Schleife“ hinausgehen und in „der Mensch ist die Schleife“-Denkweise übergehen sollte, um den menschlichen Arbeitsprozessen innerhalb der Analytik Rechnung zu tragen [EHR⁺14]. Diese Herangehensweise wurde von Keim et al. in einen **Visual Analytics-Prozess** umgesetzt⁸. In einer weiterführenden Publikation wurde die Wissensgenerierung eingehender beschrieben und der Prozess-Modell in zwei Teile geteilt (siehe Abb. 1). Der Prozess besteht zum einen aus dem Computersystem mit Daten, Visualisierungen und Analysemodellen und zum anderen aus der menschlichen Komponente, die den kognitiven Prozess einer Analyse abbildet. Das vorgestellte Prozess-Modell impliziert, dass jede Aktion eine Rückkopplung induziert und vom Menschen wahrgenommen und verarbeitet wird [SSS⁺14]. Die einzelnen durch Pfeile verdeutlichten Aktionen zeigen die relevanten Wechselwirkungen von visualisierungs- und modellzentrierten Analysen.

Im Hinblick auf das einleitende Beispiel der Savanne und das zu untersuchende Themenfeld der Simulationen ist es wichtig zu verstehen, wie und warum Tiere bestimmte Ressourcen nutzen, wie sie miteinander und mit ihrer Umwelt interagieren und wie sie konkurrieren und sich reproduzieren; dies bildet die wichtigsten Elemente evolutionärer Prozesse ab [DBC⁺15]. Das Verständnis der Prozesse auf der Grundlage von Bewegungen stellt beispielsweise eine Verbin-

⁸ Siehe [KAF⁺08]

dung zu Verteilungsdynamiken von Populationen dar. Hierbei handelt es sich um spatio-temporale Daten, welche in der Visual Analytics besonders allgegenwärtig sind [RLA⁺13]. Dabei können Raum und Zeit wie jedes andere Datenattribut behandelt werden. Jedoch haben in vielen Anwendungen die unabhängigen Dimensionen von Raum und Zeit eine semantische Bedeutung angenommen und werden besonders in der Geovisual Analytics betrachtet [KH13].

Geovisual Analytics oder auch GeoSpatial Visual Analytics basiert auf den Prinzipien der Visual Analytics und kann als eine Unterkategorie dieser verstanden werden. Der Schwerpunkt liegt hier jedoch mehr auf den geographischen Aspekten, wie beispielsweise geospatialen Räumen und Ereignissen sowie der Kartographie. Hier ist der zeitliche Charakter geospatialer Daten von großer Bedeutung, wie auch die Charakteristika und Struktur der Zeit selbst [AAD⁺10]. Die beiden Professoren der City University London, Gennady Andrienko und Natalia Andrienko⁹, die zugleich auch verantwortlich für die Visual Analytics-Entwicklung bei der Fraunhofer IAIS sind, sind die wohl bekanntesten Akteure der GeoVisual Analytics. Sie veröffentlichten unter anderem die Monographie *Visual Analytics of Movement* [AA13].

Innerhalb der Geovisual Analytics werden Techniken und Werkzeuge zur interaktiven visuellen Analyse geospatialer und spatio-temporaler Daten entwickelt und evaluiert [AAK⁺11]. Die Animation von Karten hat sich mittlerweile als Standardwerkzeug zur Darstellung zeitabhängiger Daten und dynamischer Phänomene etabliert. Eine weitere weitverbreitete Technik stellt der „Space-Time Cube“ (siehe Abb. 2) dar, welcher aufschlussreiche Informationen über die zeitliche Abfolge von geospatialen Ereignissen und Bewegungen von Entitäten geben kann [AA13]. Er basiert auf einer der Ideen der zeitlichen Geographie, in der Raum und Zeit als untrennbar gelten und somit eine drei-dimensionale Darstellung erfordern. Der Raum wird in zwei Dimensionen und die Zeit in der dritten Dimension dargestellt.

Ein gemeinsamer Designansatz visueller, analytischer Systeme für hoch-dimensionale, komplexe Daten ist die mehrfach verbundene oder koordinierte Darstellung (Multiple Linked/Coordinated Views) der Daten [DBC⁺15]. Die Daten werden in interaktiv miteinander verbundenen Visualisierungen dargestellt, sodass jede Darstellung eine andere Perspektive und Beschreibung auf die Daten bietet. Wird ein dargestellter Datensatz durch eine Selektion in einer Visualisierung verändert, werden die verbleibenden Darstellungen ebenfalls entsprechend modifiziert. Die unterschiedlichen Datenvariablen können so in mehreren verlinkten Darstellungen parallel untersucht und analysiert werden. Diese Darstellungen bestehen zumeist aus Histogrammen, Scatterplot-Matrizen, Parallelkoordinaten oder Funktionsgraphen [KH13] [SWLL13], welche auch in Multi-Agenten-Simulationen angewendet werden, wie nachfolgend beschrieben wird.

⁹ Weiterführende Informationen auf <http://geoanalytics.net/and/>, zuletzt aufgerufen am 05.12.2015

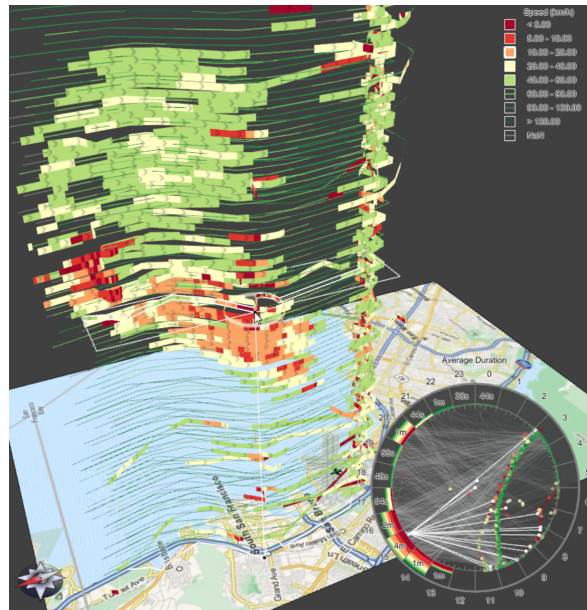


Abb. 2. Der Space-Time Cube zur Darstellung der Entwicklung eines Verkehrsstaus. Abbildung entnommen aus [TSAA12].

3 Visual Analytics in Multi-Agenten-Simulationen

Multi-Agenten-Simulationen (MAS) stellen ein hilfreiches Werkzeug zum besseren Verständnis sowie zur Gestaltung von jeglichen Gesellschaften dar, die vor allem (nur) mit sehr vielen Agenten beschrieben werden können [FKP05]. Das Besondere an agentenbasierten Simulationen ist, dass die aktiven Komponenten des zu untersuchenden Systems, also die Agenten, explizit mit ihrer Autonomie, ihren Zielen etc. individuell dargestellt werden können. Damit verfolgt die MAS das Ziel, die Wechselwirkungen zwischen Agenten und der Umwelt zu modellieren. Dabei gibt es eine Reihe von Anwendungsgebieten, beispielsweise kann die Umwelt eine geographische Region oder ein Wissensnetzwerk zwischen Wissenschaftlern darstellen. Obwohl das Verhalten der einzelnen Agenten im System voneinander unabhängig ist, kann dennoch kollektives Verhalten als Gruppe entstehen [RBU⁺14]. Auf diese Weise kann die Heterogenität realer Gesellschaften abgebildet und ihre Auswirkungen auf das Gesamtsystem analysiert werden.

Die wohl populärste Konferenz im Bereich von Multi-Agenten-Simulationen ist die jährlich stattfindende AAMAS¹⁰, bei der der bekannte Multi-Agent-Based Si-

¹⁰ *International Conference on Autonomous Agents and Multiagent Systems* - s. <http://www.aamas-conference.org/index.html>, zuletzt aufgerufen am 05.12.2015

mulation (MABS) Workshop¹¹ abgehalten wird. Dieser bietet ein internationales Forum für Publikationen und Präsentationen aus den praktischen und theoretischen Feldern der MAS. Weitere wichtige Konferenzen, Workshops und Journals werden von den beiden großen Institutionen der *Society for Modeling & Simulation International (SCS)* und der *Special Interest Group on Simulation and Modeling (SIGSIM)* präsentiert, wie beispielsweise das *Transactions of The Society for Modeling and Simulation International (SIMULATION)*¹² oder die *Summer Simulation Multi-Conferenz (SummerSim)*¹³.

Heutzutage wird weitgehend anerkannt, dass ein umfassendes Verständnis komplexer agentenbasierter Modelle durch einen explorativen Ansatz vier Prinzipien beachten muss [GDZ13]. Die Darstellungen müssen die Entwicklungen unter verschiedenen Konfigurationen dynamisch, realistisch und klar visualisieren (1) und sich bezüglich räumlicher und zeitlicher Parameter oder durch alternative Perspektiven einfach und idealerweise interaktiv anpassen lassen (2). Abstrakte Eigenschaften sowie Informationen bezüglich der Dynamik, die durch die Simulation generiert werden, müssen abgeleitet und in Echtzeit visualisiert werden können (3). Dabei ist es wichtig mit den Visualisierungen auf natürliche Weise interagieren zu können, um das Modell selbst in einem interaktiven Designansatz modifizieren zu können (4).

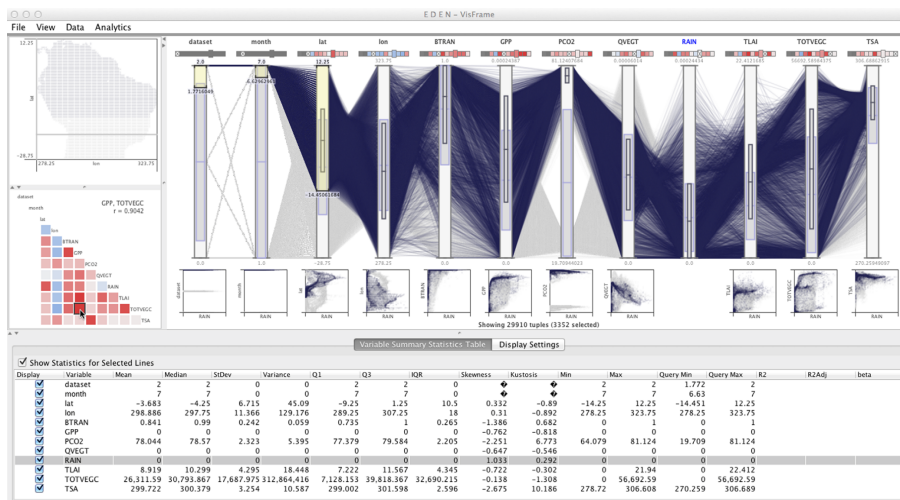


Abb. 3. EDEN: Selektierte Polygonzüge der Parallelkoordinaten zeigen in Verbindung mit Streudiagrammen die gestiegenen CO₂-Parameter einer Simulation für ein Szenario im Amazonas. Abbildung entnommen aus [SRS⁺13]

¹¹ <http://www.pcs.usp.br/~mabs/>

¹² <http://sim.sagepub.com>

¹³ <http://www.scs.org/summersim>

Ein Visual Analytics-Ansatz aus der Forschung zur Analyse von Simulationen ist das sogenannte *Exploratory Data analysis ENvironment* (EDEN) Framework, welches interaktive, visuelle Abfragen auf Simulationsdaten ermöglicht. Neben vielseitigen Filterfunktionen bietet es die visuelle Datenanalyse mit Parallelkoordinaten, welche über Multiple Linked Views mit Streudiagrammen und Korrelationsmatrizen verbunden sind (siehe Abb. 3). Jede vertikale Achse stellt in der Parallelkoordinaten-Darstellung eine der Modellvariablen dar. Jede Linie bzw. jeder Polygonzug, oft zur besseren Ansicht in Clustern zusammengefasst, charakterisiert wiederum multivariate Tupel der Daten. Visuelle Hinweise, wie dynamische Histogramme, schattierte Balkendiagramme und Box-Whisker-Plots, sind in die Achsen eingebettet. Visuelle Abfragen können unter anderem durch das Selektieren von Bereichen auf den separaten Achsen getätigt werden. Der Ansatz von Multiple Linked Views führt auch hier dazu, dass die weiteren Ansichten, wie die Streudiagramme unter den Achsen, mit angepasst werden. Diese Streudiagramme bieten zusätzliche Details, wie etwa Grenzwerte, nichtlineare Tendenzen und Cluster [SRS⁺13].

Ein Problem der Parallelkoordinaten ist jedoch die mögliche Anzahl an verschiedenen Anordnungen und Skalierungen der Achsen, welche bei ungünstiger Konstellation wichtige Details verbergen und nicht zur Übersichtlichkeit der zu erforschenden Daten beitragen können [TAE⁺09]. Tatu et al. [TAE⁺09] stellen ein automatisiertes Vorgehen vor, welches ein Set von potenziell aufschlussreichen Darstellungen und visuellen Strukturen anhand von anwenderspezifischen, festgelegten Metriken generiert. Der Analyst kann so die Darstellung mit der besten Relevanz als Ausgangspunkt der interaktiven Analyse nutzen.

Unabhängig von der spezifischen Zielsetzung einer Simulation möchte der Modellierer ein simuliertes System erzeugen, welches ein bestimmtes Verhalten oder gewisse Bedingungen in einem Kontext ganz oder teilweise erfüllt. Durch die Validierung der Simulation kann diese zumindest auf bestimmte Zielsetzungen geprüft und eine angemessene Abbildung hinsichtlich der zu erwartenden Resultate der Simulation gefunden werden [GBDD13].

Der Validierungsansatz von Unger et al. [USKD12] verfolgt einen Visual Analytics-Ansatz und bietet sich an, um Simulationsdaten mit gemessenen, empirischen Daten vergleichen zu können. Es wird ein Maß bezüglich der Übereinstimmung der Simulations- und empirischen Werte ermittelt, welches als Anpassungsgüte (goodness of fit) bezeichnet wird. In Abbildung 4 werden vier Faktoren der Anpassungsgüte eines geowissenschaftlichen Simulationsmodells anhand von interaktiven Visualisierungen dargestellt (*VC1-VC4*). Durch die direkte Visualisierung der Werte und dazugehörigen Anpassungsgüte, kann der mehrdeutige Zusammenhang zwischen Anpassungsgüte und den zugrundeliegenden empirischen sowie modellierten Werten aufgezeigt werden. So wird eine detaillierte Analyse der Variationen in Raum und Zeit unterstützt (*VC2*) und die Genauigkeit der empirischen Daten aufgezeigt. Erfüllen die Werte inklusive der Anpassungsgüte nicht die Erwartungen, kann der Modellierer beispielsweise schon frühzeitig in die laufende Simulation eingreifen und das Modell kalibrieren.

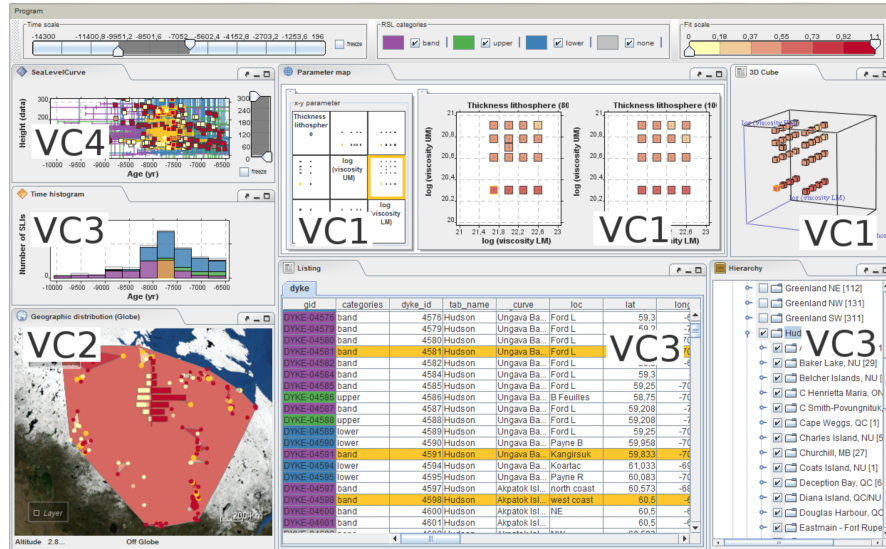


Abb. 4. Visual Analytics-Konzept zur Validierung und Exploration von Simulationen anhand der „goodness of fit“ im Parameterraum (*VC1*), der spatio-temporalen Varianten der Anpassungsgüte (*VC2*), der observierten Daten (*VC3*) und der Relationen zwischen der Anpassungsgüte und den Datenwerten (*VC4*). Abbildung entnommen aus [SRS⁺13].

Im Idealfall lassen sich die Modellparameter unabhängig voneinander messen. Bei Parametern mit konkreter Bedeutung lässt sich ihr Wertebereich oft auf ihr realistisches Auftreten reduzieren; ohne konkrete Bedeutung der Parameter ist dies jedoch kaum möglich [Hel12b]. Eine gestützte, automatisierte Kalibrierung der Modellparameter bedarf mehrerer Simulationsdurchgänge verschiedener Parameterkombinationen. Auf diese Weise können Parametersets bestimmt werden, die entweder die höchste Anpassungsgüte oder den größten Einfluss auf das Simulationsergebnis haben. In einer solchen Sensitivitätsanalyse werden mehrere Simulationsdurchgänge durchlaufen, indem die Werte bestimmter Modellparameter systematisch variiert werden [Ham]. Als Ergebnis wird die Verteilung der Werte pro Simulationslauf für jedes Datenattribut an jeder Position in Raum und Zeit wiedergegeben. Die Visualisierung mehrerer Simulationsdurchgänge ist somit von besonderem Interesse, da sie darüber hinaus einen weiteren Ansatz zur Darstellung von Unsicherheit (engl. uncertainty) zwischen Simulationen abbildet [KH13].

Eine hohe Übereinstimmung, d.h. ein hoher „goodness of fit“ während der Kalibrierung bedeutet jedoch nicht zwangsläufig, dass mit dem Modell genauere Vorhersagen neuer Datensätze getroffen werden können. Man sieht sich im Gegenteil oft mit dem Problem einer „Überanpassung“ konfrontiert, d.h. es besteht das Risiko, Rauschen oder irrelevante Details des Datensatzes mit anzupassen und

bestimmte Werte soweit zu kalibrieren, dass sie das ursprüngliche Modell nicht mehr repräsentieren [Hel12b]. Da man mit einer inadäquat gewählten Funktion zur Berechnung des „goodness of fit“ nur schlecht zwischen guten und schlechten Modellen unterscheiden kann, bedarf es einer entsprechenden, manuellen Validierung, um die Vorhersagekraft des Modells bestimmen zu können. Durch die Integration von Visual Analytics-Methoden, welche vor allem den Menschen einbeziehen, können zumindest Einschätzungen der Experten mehr als je zuvor einfließen.

4 Fazit und Ausblick

Die Visualisierung großer Datenmengen ist ein herausforderndes Forschungsthema, welches mit Methoden und Algorithmen der interdisziplinären Visual Analytics angegangen werden kann. Die Herangehensweisen und die Kombination der Visual Analytics aus Bereichen der Informationsvisualisierung, Data-Mining, geospatialen und temporalen Analytik sowie Wissensrepräsentation wurden beleuchtet. Der Visual Analytics-Prozess von Keim bietet ein Vorgehensmodell, um effektive Techniken und Systeme der Visual Analytics zu erarbeiten. Das Themenfeld wurde tiefer durch die Geovisual Analytics und mögliche Optionen zur Unterstützung von Multi-Agenten-Simulationen vorgestellt. Bewegungen von zig Agenten und Ereignissen mit Raum-Zeit-Bezug lassen sich mit den exemplarisch vorgestellten Visualisierungs- und Analysetechniken verdeutlichen. Diese lassen sich auch zur Validierung und Kalibrierung von Multi-Agenten-Simulationen heranziehen und bringen dem Modellierer oder Domänenexperten die explorative, visuelle Analyse näher.

Im Rahmen des Grundprojekts soll eine solche Visual Analytics-Komponente für das MARS-Framework¹⁴ geschaffen werden, um beispielsweise dem Domänenexperten bei einem Entscheidungsprozess einen besseren Einblick in die komplexen Vorgänge und Parametrisierungen des simulierten Modells zu gewähren. Als erster Schritt bietet sich die kritische Evaluierung von verschiedenen hier vorgestellten und unzähligen weiteren Techniken der Visual Analytics an. Eine Integration des Space-Time Cubes in diese Komponente ist aufgrund der Relevanz des Space-Time Cubes wünschenswert. Weiterhin wird die Auswertung und Visualisierung von Korrelationen der Modellparameter über viele Simulationen hinweg und der Vergleich von Simulationsdaten mit den empirisch gemessenen Daten wichtig sein. Die Anwendbarkeit einer Validierung und Kalibrierung einer Simulation sowie die Untersuchung des Problems der Überanpassung, könnte vor allem an unzureichenden, empirischen Daten oder fehlerhaften Modellimplementationen scheitern. Ein weiteres potentiell Risiko bei der Entwicklung der Komponente könnte der *Modeling and Simulation as a Service*-Ansatz [HATCea16] von MARS sein, da bisher noch wenige Forschungsarbeiten zu Visual Analytics in solchen Cloudumgebungen zu finden sind.

¹⁴ Multi Agent Research and Simulation [TC15]-Framework der HAW Hamburg, <http://www.mars-group.mars.haw-hamburg.de>

Literatur

- [AA13] Natalia Andrienko and Gennady Andrienko. Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, 12(1):3–24, January 2013.
- [AAD⁺10] Gennady Andrienko, Natalia Andrienko, Urska Demsar, Doris Dransch, Jason Dykes, Sara Irina Fabrikant, Mikael Jern, Menno-Jan Kraak, Heidrun Schumann, and Christian Tominski. Space, time and visual analytics. *International Journal of Geographical Information Science*, 24(10):1577–1600, 2010.
- [AAK⁺11] Gennady Andrienko, Natalia Andrienko, Daniel Keim, Alan M MacEachren, and Stefan Wrobel. Challenging problems of geospatial visual analytics. *Journal of Visual Languages & Computing*, 22(4):251–256, 2011.
- [BG05] Peter WJ Baxter and Wayne M Getz. A model-framed evaluation of elephant effects on tree and fire dynamics in african savannas. *Ecological Applications*, 15(4):1331–1341, 2005.
- [DBC⁺15] Urška Demšar, Kevin Buchin, Francesca Cagnacci, Kamran Safi, Bettina Speckmann, Nico Van de Weghe, Daniel Weiskopf, and Robert Weibel. Analysis and visualisation of movement: an interdisciplinary review. *Movement Ecology*, 3(1):1–24, 2015.
- [EHR⁺14] Alex Endert, M. Shahriar Hossain, Naren Ramakrishnan, Chris North, Patrick Fiaux, and Christopher Andrews. The human is the loop: new directions for visual analytics. *Journal of Intelligent Information Systems*, 43(3):411–435, 2014.
- [FKP05] Manuel Fehler, Franziska Klügl, and Frank Puppe. *Engineering Societies in the Agents World V: 5th International Workshop, ESAW 2004, Toulouse, France, October 20-22, 2004. Revised Selected and Invited Papers*, chapter Techniques for Analysis and Calibration of Multi-agent Simulations, pages 305–321. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- [GBDD13] A Ghorbani, PWG Bots, V Dignum, and GPJ Dijkema. Maia: A framework for developing agent-based social simulations. *Journal of Artificial Societies and Social Simulation*, 16 (2), 2013, 2013.
- [GDZ13] Arnaud Grignard, Alexis Drogoul, and Jean-Daniel Zucker. *Computational Science and Its Applications – ICCSA 2013: 13th International Conference, Ho Chi Minh City, Vietnam, June 24-27, 2013, Proceedings, Part I*, chapter Online Analysis and Visualization of Agent Based Models, pages 662–672. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [Ham] D. M. Hamby. A review of techniques for parameter sensitivity analysis of environmental models. *Environmental Monitoring and Assessment*, 32(2):135–154.
- [HATCea16] Christian Huening, Mitja Adebahr, Thomas Thiel-Clemen, and Janus Dybulla et al. Modeling and simulation as a service with the massive multi-agent system mars. To appear in Proceedings of the 2016 Spring Simulation Multiconference, 2016.
- [Hel12a] Dirk Helbing. *Social Self-Organization: Agent-Based Simulations and Experiments to Study Emergent Social Behavior*, chapter Agent-Based Modeling, pages 25–70. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
- [Hel12b] Dirk Helbing. *Social Self-Organization: Agent-Based Simulations and Experiments to Study Emergent Social Behavior*, chapter Agent-Based Modeling, pages 25–70. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

- [HSS11] S. Hadlak, H. J. Schulz, and H. Schumann. In situ exploration of large dynamic networks. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2334–2343, Dec 2011.
- [KAF⁺08] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. *Information Visualization: Human-Centered Issues and Perspectives*, chapter Visual Analytics: Definition, Process, and Challenges, pages 154–175. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
- [Kei10] Daniel Keim, editor. *Mastering the information age : solving problems with visual analytics*. Goslar : Eurographics Association, 2010.
- [KH13] J. Kehrer and H. Hauser. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 19(3):495–513, March 2013.
- [KKP⁺11] Joern Kohlhammer, Daniel Keim, Margit Pohl, Giuseppe Santucci, and Gennady Andrienko. Solving problems with visual analytics. *Procedia Computer Science*, 7:117 – 120, 2011. Proceedings of the 2nd European Future Technologies Conference and Exhibition 2011 (FET 11).
- [KMT10] Daniel A. Keim, Florian Mansmann, and Jim Thomas. Visual analytics: How much visualization and how much analytics? *SIGKDD Explor. Newsl.*, 11(2):5–8, May 2010.
- [KPG12] Mike Kluse, Anthony Peurrung, and Deborah Gracio. *Expanding the Frontiers of Visual Analytics and Visualization*, chapter The Evolving Leadership Path of Visual Analytics, pages 31–42. Springer London, London, 2012.
- [LSFK12] T. Landesberger, T. Schreck, D. W. Fellner, and J. Kohlhammer. *Expanding the Frontiers of Visual Analytics and Visualization*, chapter Visual Search and Analysis in Complex Information Spaces—Approaches and Research Challenges, pages 45–67. Springer London, London, 2012.
- [MDSC03] Maria Bruno Marietto, Nuno David, Jaime Simão Sichman, and Helder Coelho. *Multi-Agent-Based Simulation II: Third International Workshop, MABS 2002 Bologna, Italy, July 15–16, 2002 Revised Papers*, chapter Requirements Analysis of Agent-Based Simulation Platforms: State of the Art and New Prospects, pages 125–141. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.
- [RAC⁺13] Celia G. Ralha, Carolina G. Abreu, Cassio G.C. Coelho, Alexandre Zaghetto, Bruno Macchiavello, and Ricardo B. Machado. A multi-agent model system for land-use change simulation. *Environmental Modelling and Software*, 42:30 – 46, 2013.
- [RBU⁺14] Tobias Ruppert, Jürgen Bernard, Alex Ulmer, Hendrik Lücke-Tieke, and Jörn Kohlhammer. Visual access to an agent-based simulation model to support political decision making. In *Proceedings of the 14th International Conference on Knowledge Technologies and Data-driven Business, i-KNOW '14*, pages 16:1–16:8, New York, NY, USA, 2014. ACM.
- [RLA⁺13] A. Rind, T. Lammarsch, W. Aigner, B. Alsallakh, and S. Miksch. Timebench: A data model and software library for visual analytics of time-oriented data. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2247–2256, Dec 2013.
- [SH12] Simon Scheiter and Steven I. Higgins. How many elephants can you fit into a conservation area. *Conservation Letters*, 5(3):176–185, 2012.

- [Shn96] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343, Sep 1996.
- [SRS⁺13] Chad A. Steed, Daniel M. Ricciuto, Galen Shipman, Brian Smith, Peter E. Thornton, Dali Wang, Xiaoying Shi, and Dean N. Williams. Big data visual analytics for exploratory earth system simulation analysis. *Computers and Geosciences*, 61:71 – 82, 2013.
- [SSS⁺14] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim. Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1604–1613, Dec 2014.
- [SWLL13] Guo-Dao Sun, Ying-Cai Wu, Rong-Hua Liang, and Shi-Xia Liu. A survey of visual analytics techniques and applications: State-of-the-art research and future challenges. *Journal of Computer Science and Technology*, 28(5):852–867, 2013.
- [TAE⁺09] A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnork, and D. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 59–66, Oct 2009.
- [TC05] J.J. Thomas and K.A Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Ctr (2005), Los Alamitos, CA., 2005.
- [TC15] Ulfia; Huening Christian; Schnoor Alexander; Kiker Greg Thiel-Clemen, Thomas; Lenfers. A benefit for knp-ecologists „Äi massive multi-agent simulation with mars. *Poster at Savanna Science Network Meeting, Kruger NP, South Africa*, 2015.
- [TSAA12] C. Tominski, H. Schumann, G. Andrienko, and N. Andrienko. Stacking-based visualization of trajectory attribute data. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2565–2574, Dec 2012.
- [USKD12] A. Unger, S. Schulte, V. Klemann, and D. Dransch. A visual analysis concept for the validation of geoscientific simulation models. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2216–2225, Dec 2012.