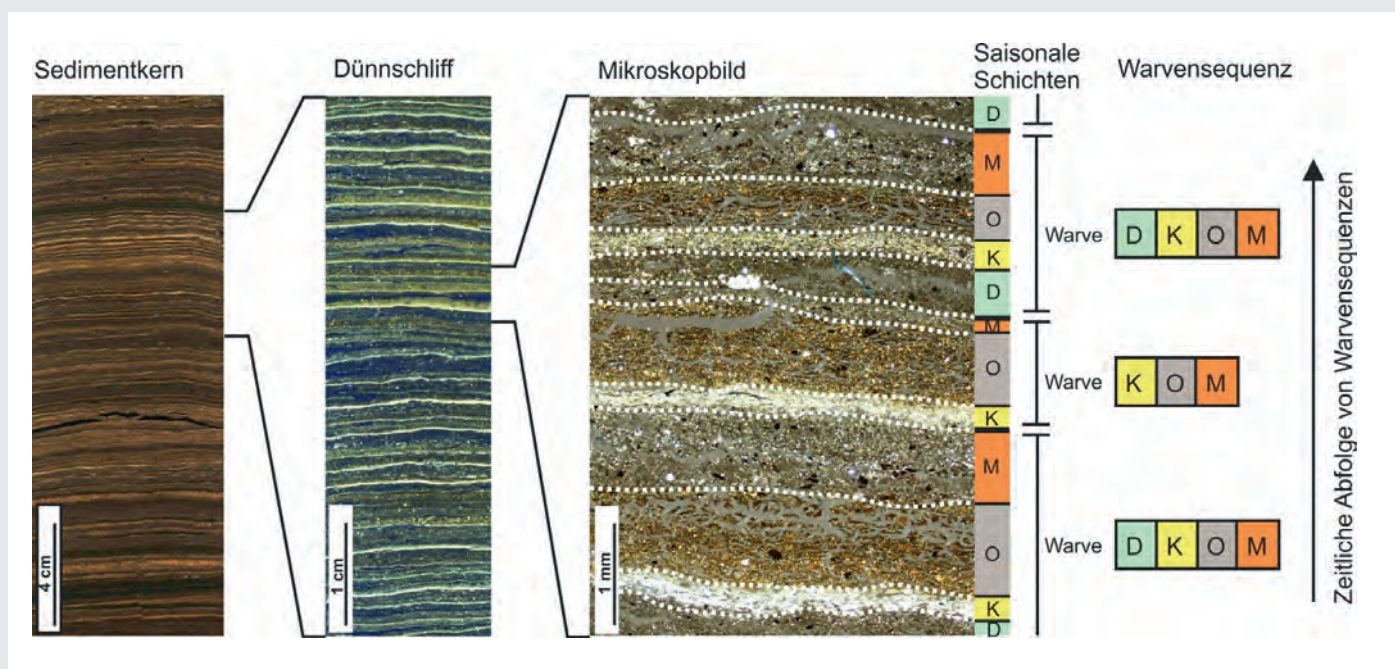


Visual Analytics – eine interaktive Methode für die Interpretation geowissenschaftlicher Daten

Andrea Unger, Doris Dransch, Janis Jatnieks, Mike Sips, Nadine Dräger, Achim Brauer, Volker Klemann, Thomas Kempka, Marco De Lucia, Michael Kühn
Deutsches GeoForschungsZentrum GFZ, Potsdam

Visual Analytics is a method that combines visualization and automated analysis in highly interactive interfaces. These systems enable geoscientists to explore their data sets, to steer the analysis process and assess results. In this article, we explain the potential of Visual Analytics with three examples that were developed at the GFZ in recent years in close cooperation between computer scientists and geoscientists. They include (1) the analysis of lake sediments to understand climate and landscape developments in the past, (2) the analysis of complex geochemical simulation models in fluid systems modeling, and (3) the validation of simulation models in Earth system modeling. The developed Visual Analytics concepts largely differ from each other due to the variety of analytical tasks and data characteristics across the applications. Nevertheless, the examples reveal common benefits of using Visual Analytics for data interpretation in geoscientific research. The method supports the full exploration of complex data sets rather than relying on small sample sets. Geoscientists can employ their expert knowledge to correctly interpret data where automatic methods alone are not sufficient. Yet, the close linking with automated methods enables the handling of large and complex data sets. In sum, the novel Visual Analytics concepts contribute to data interpretation and thereby to answering relevant geoscientific questions.



„Sich ein Bild von etwas machen“ – diese Redewendung gilt für die Geowissenschaften oft wortwörtlich. So kann z. B. die Bedeutung von gemessenen oder simulierten Daten besser verstanden werden, indem die Daten grafisch dargestellt werden. Diagramme und Karten sind gängige Methoden, um Muster und Zusammenhänge in den Daten zu entdecken. Doch das Potenzial von visuellen Darstellungen in der Datenanalyse geht weit über statische Bilder hinaus. „Visual Analytics“ verbindet Visualisierungsmethoden und automatische Analyse in hochinteraktiven Systemen. Mit ihnen erhalten Forschende die Möglichkeit, den Analyseprozess interaktiv zu steuern, Ergebnisse direkt zu bewerten sowie neue Hypothesen zu generieren oder zu prüfen. Durch die Verknüpfung mit automatischen Methoden ist es möglich, große und komplexe Datenmengen zu handhaben.

In enger Kooperation zwischen der Informatik und den Geowissenschaften wurden am Deutschen GeoForschungsZentrum GFZ in den letzten Jahren neue Visual-Analytics-Konzepte entwickelt. Dieser Beitrag stellt drei Beispiele mit unterschiedlichen Einsatzbereichen in der Geoforschung vor: (1) die Analyse von Seesedimenten, um Klima- und Landschaftsentwicklungen der Vergangenheit zu verstehen, (2) die Analyse von komplexen geochemischen Simulationsmodellen in der Fluidsystemmodellierung und (3) die Validierung von Simulationsmodellen glazial-isostatischer Anpassungen in der Erdsystemmodellierung.

Mikrofazienexplorer: Analyse von Seesedimenten, um Klima- und Landschaftsentwicklungen zu verstehen

Das erste Beispiel ist aus der Zusammenarbeit zwischen den GFZ-Sektionen „Geoinformatik“ und „Klimadynamik und Landschaftsentwicklung“ entstanden (Unger *et al.*, 2018). Im Mittelpunkt steht

die Analyse von Sedimentablagerungen in Seen, um das Klima der Vergangenheit zu verstehen und damit Klimaprojektionen für die Zukunft zu verbessern. Die Informationen über das Klima in der Vergangenheit stammen aus geologischen Archiven. Seesedimente sind dabei besonders interessant, da sie Daten aus dem Lebensraum des Menschen liefern. Im See lagert sich Jahr um Jahr alles ab, was sich im Wasser bildet (Algen, Minerale) und was von außen in den See transportiert wird (Staub, Pollen). Die Zusammensetzung dieser Sedimente wird also stark von den bestehenden Umweltbedingungen beeinflusst. Werden Seesedimente aus der Vergangenheit untersucht, können Rückschlüsse auf die Klima- und Landschaftsgeschichte gezogen werden. Die Sektion „Klimadynamik und Landschaftsentwicklung“ untersucht daher vor allem jährlich geschichtete (d. h. varvierte) Seesedimente (Abb. links). Das Alter varvierter Seesedimente kann bestimmt werden, da sich in jedem Jahr eine Abfolge von charakteristischen Sedimentschichten am Seeboden ablagert – die sogenannte Warvensequenz. So werden saisonale Klimainformationen im Sediment abgespeichert. Diese Sedimentschichten können unter dem Mikroskop ausgemessen und beschrieben werden. Oft werden 10 bis 20 verschiedene Parameter zu jeder Warvensequenz erfasst, z. B. Klassifizierung, Mächtigkeit und Anteil an bestimmten Sedimentbestandteilen. In einigen Seen ändert sich zudem auch die Warvensequenz selbst – wie z. B. im Tiefer See in Norddeutschland. Hier können 600 verschiedene Warvensequenzen unterschieden werden, die über einen Zeitraum von 6000 Jahren auftreten. Mit herkömmlichen Methoden gestaltet es sich schwierig, solche umfangreichen und komplexen Datensätze zu untersuchen. Bisher können Warvensequenzen im Detail nur stichprobenartig für wenige Jahre betrachtet werden. Möchte man den gesamten Untersuchungszeitraum ansehen, so ist dies nur für einzelne Parameter möglich. Ziel ist es jedoch, alle Parameter und die Informationen der Warvensequenzen zusammenhängend zu betrachten und damit Phasen von Klimaveränderungen festzustellen.

An dieser Stelle setzt ein neues, am GFZ entwickeltes Konzept zur visuellen Analyse von saisonalen Sedimentdaten an (Abb. 1). Es ermöglicht einen Überblick über den gesamten Datensatz. Dazu werden zwei Sichten verknüpft: Eine Sicht auf die zeitliche Abfolge der Warvensequenzen und eine Sicht auf ihre Häufigkeit im Datensatz. Ein großer Vorteil des Konzepts ist die Möglichkeit, Zusammenhänge zwischen den Jahren über den gesamten Zeitraum zu finden. Mit Hilfe eines semiautomatischen Clusteringverfahrens können Warvensequenzen klassifiziert werden, die auf ähnliche Umweltbedingungen hinweisen. Dafür schlägt das Tool mit einem passenden Algorithmus ähnliche Warvensequenzen vor. Mittels fachlicher Expertise kann anschließend die Klassifizierung interaktiv bewertet und gegebenenfalls die Auswahl manuell angepasst werden. Das Ergebnis ist ein zeitlicher Verlauf der Klassenhäufigkeit. So können Abschnitte ähnlicher Umweltbedingungen abgegrenzt und der Zeitpunkt von Veränderungen bestimmt werden.

Links: Sedimentablagerungen in Seen geben Aufschluss über die Klima- und Landschaftsgeschichte. Unter dem Mikroskop werden jährlich geschichtete Seesedimente in sogenannten Warvensequenzen zusammengefasst. Enthält ein Datensatz viele verschiedene Warvensequenzen, wird die Analyse zur Herausforderung. Am GFZ wurde deshalb ein neuer Visual-Analytics-Ansatz entwickelt, der in diesem Beitrag vorgestellt wird.

Left: Sediments in lakes reveal the history of climate and landscape development. Under the microscope, annually laminated lake sediments are grouped into so-called varve sequences. Analyzing data sets with many different varve sequences is a challenge. To this end, a novel Visual Analytics approach has been developed at GFZ, which is introduced in this article.



Kontakt: D. Dransch
(dransch@gfz-potsdam.de)

Mit dem neuen Visual-Analytics-Konzept ist es möglich, den Datensatz saisonaler Sedimentdaten in seiner gesamten Komplexität zu untersuchen und nicht nur Stichproben zu betrachten. Der große Mehrwert liegt im Gewinnen neuer Indikatoren, die vergangene Klima- und Umweltveränderungen beschreiben. Da die Interpretation der Sedimentdaten auf dem gesamten Datensatz basiert und nicht nur auf einzelnen Parametern, können die Daten umfassender interpretiert werden.

GCex: Interaktive Exploration von multivariaten geochemischen Simulationen

Das zweite Beispiel ist in Kooperation zwischen den GFZ-Sektionen „Geoinformatik“ und „Fluidsystemmodellierung“ entstanden (De Lucia et al., 2015). Das Konzept erleichtert die Analyse komplexer geochemischer Simulationen, bei denen zahlreiche Parameter miteinander interagieren. Geochemische Simulationen sind in

geowissenschaftlichen Anwendungen insbesondere dann von Bedeutung, wenn Prognosen über chemische Fluid-Fluid- oder Fluid-Gestein-Wechselwirkungen benötigt werden. Typische Anwendungen umfassen die geologische Speicherung von CO₂ und stofflichen Energieträgern, die Endlagerung radioaktiver Abfälle oder die Nutzung geothermaler Ressourcen. Bei der Simulation geht es vor allem um die möglichen chemischen Prozesse, die zur Lösung oder Ausfällung von Mineralen und zur Änderung der Zusammensetzung von Fluiden beitragen. Diese Prozesse können auf unterschiedlichen Zeitskalen ablaufen. Geochemische Simulationen liefern wichtige Beiträge zur Prognose der Langzeitintegrität von Reservoiren und ihren Deckgesteinen, zur Interpretation natürlicher Sedimentations- und Diagenesevorgänge sowie zur Optimierung von Grundwasser-Sanierungsmaßnahmen.

Geochemische Modelle sind in der Regel multivariat, umfassen also umfangreiche Eingangs- und Ausgangsparameter und basie-

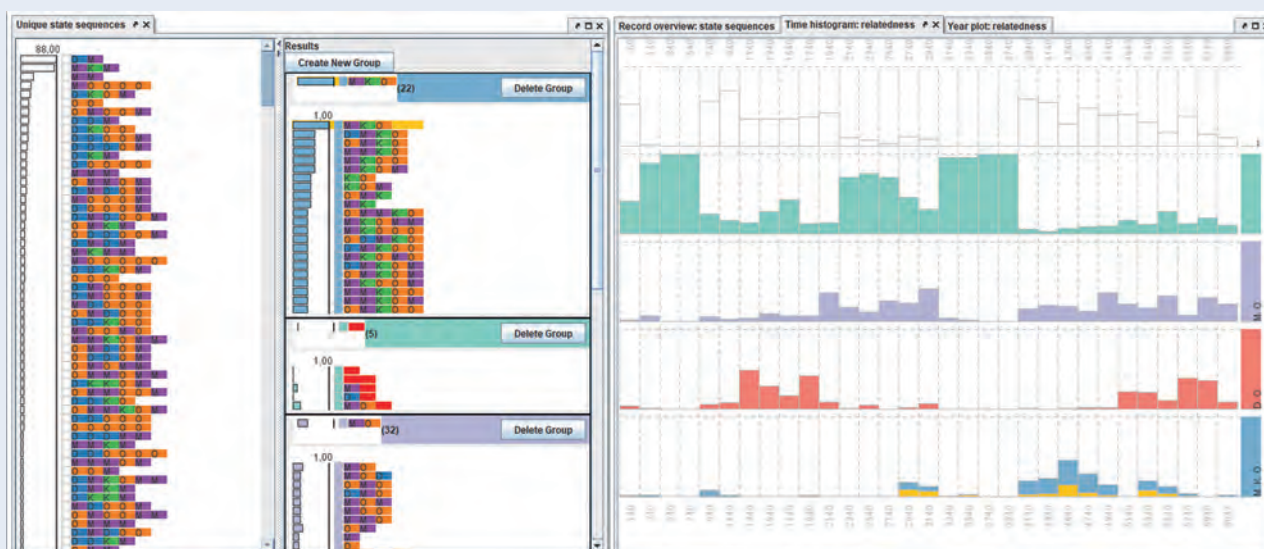


Abb. 1: Am GFZ entwickelter Visual-Analytics-Ansatz für die Analyse von Seesedimenten (Unger et al., 2018). Betrachtet werden Warvensequenzen, die zuvor in einer mikroskopischen Analyse der Seesedimente extrahiert wurden. Jedes Jahr ist durch eine Warvensequenz beschrieben. Die visuelle Oberfläche zeigt links die Liste der im Datensatz auftretenden Warvensequenzen, sortiert nach ihrer Häufigkeit. Rechts daneben sind die Gruppen von Warvensequenzen zu sehen, die die Geowissenschaftlerinnen und Geowissenschaftler interaktiv festgelegt haben, nachdem das System automatisch einander ähnliche Warvensequenzen vorgeschlagen hat. Die Histogramme rechts zeigen für jede Gruppe die zeitliche Verteilung. Es ist zu sehen, dass die einzelnen Gruppen in unterschiedlichen Zeitabschnitten auftreten. Auf diese Weise können zeitliche Abschnitte in der vergangenen Landschafts- und Klimaentwicklung voneinander abgegrenzt und anhand der auftretenden Warvensequenzen charakterisiert werden.

Fig. 1: Visual Analytics approach for the analysis of lake sediments (Unger et al., 2018) considering warve sequences, which have been extracted in a microscopic analysis of lake sediments. Each year is described by one warve sequence. On the left, the visual interface depicts the list of warve sequences that appear in the data set, sorted by frequency. Alongside, the groups of warve sequences are shown that the geoscientists have defined interactively, based on automatic recommendations of similar warve sequences. The histograms on the right show the temporal distribution for every group. It is apparent that the individual groups appear in separate time periods. This provides the basis to delineate temporal periods in the past climate and landscape development. These temporal periods can be characterized based on the appearing warve sequences.

ren auf einer großen Anzahl indirekter Parameter. Dazu gehören z. B. Reaktionsraten und Löslichkeitsprodukte von Mineralen, die in thermodynamischen Datenbanken vorgehalten werden. Allerdings sind fast alle Variablen und Parameter mit erheblichen Unsicherheiten behaftet. Geowissenschaftlerinnen und Geowissenschaftler müssen also umfangreiche Sensitivitäts- und Unsicherheitsanalysen durchführen, um die Relevanz einzelner Parameter und die Zusammenhänge zwischen bis zu 30 Eingangs- und 150 Ausgangsparametern zu verstehen. Solche Analysen basieren auf der Berechnung von Simulationsensembles – Gruppen von Simulationen, die zahlreiche mögliche Parameterkombinationen in einem vom Modellierer festgelegten Parameterraum abdecken. Zu diesem Zweck wurde das Visual-Analytics-Werkzeug GCex entwickelt (Abb. 2). Es erleichtert den Umgang mit großen Ensembles multivariater Simulationen und ermöglicht die interaktive Exploration geochemischer Modelle in Kombination mit Expertenwissen. Als Eingangspunkt dient eine sogenannte lineare varianzbasierte

Sensitivitätsanalyse, die von GCex automatisch berechnet und visualisiert wird. GCex bietet eine graphische Darstellung, die die Abhängigkeit mehrerer Ausgaben auf bis zu sechs Eingaben integriert (stacked parameter relation). Hierzu werden verschachtelte x-Achsen verwendet, wobei die Visualisierung interaktiv angepasst werden kann, um eine interpretierbare Datenvisualisierung zu erzielen. Rein automatische Methoden können die erheblichen Unterschiede in den Größenordnungen der Eingangs- und Ausgangsparameter, die geochemische Modelle charakterisieren, nicht handhaben. Dies gilt besonders hinsichtlich der physikalischen Relevanz der Varianz im Vergleich zu einem durch die Normierung der Variablen entstandenen numerischen Artefakt.

Expertinnen und Experten verfügen mit GCex erstmalig über ein graphisches Werkzeug, mit dem sie systematisch das Verhalten multivariater geochemischer Modelle bis ins Detail für einen großen Parameterraum kontrollieren, verstehen und visualisieren

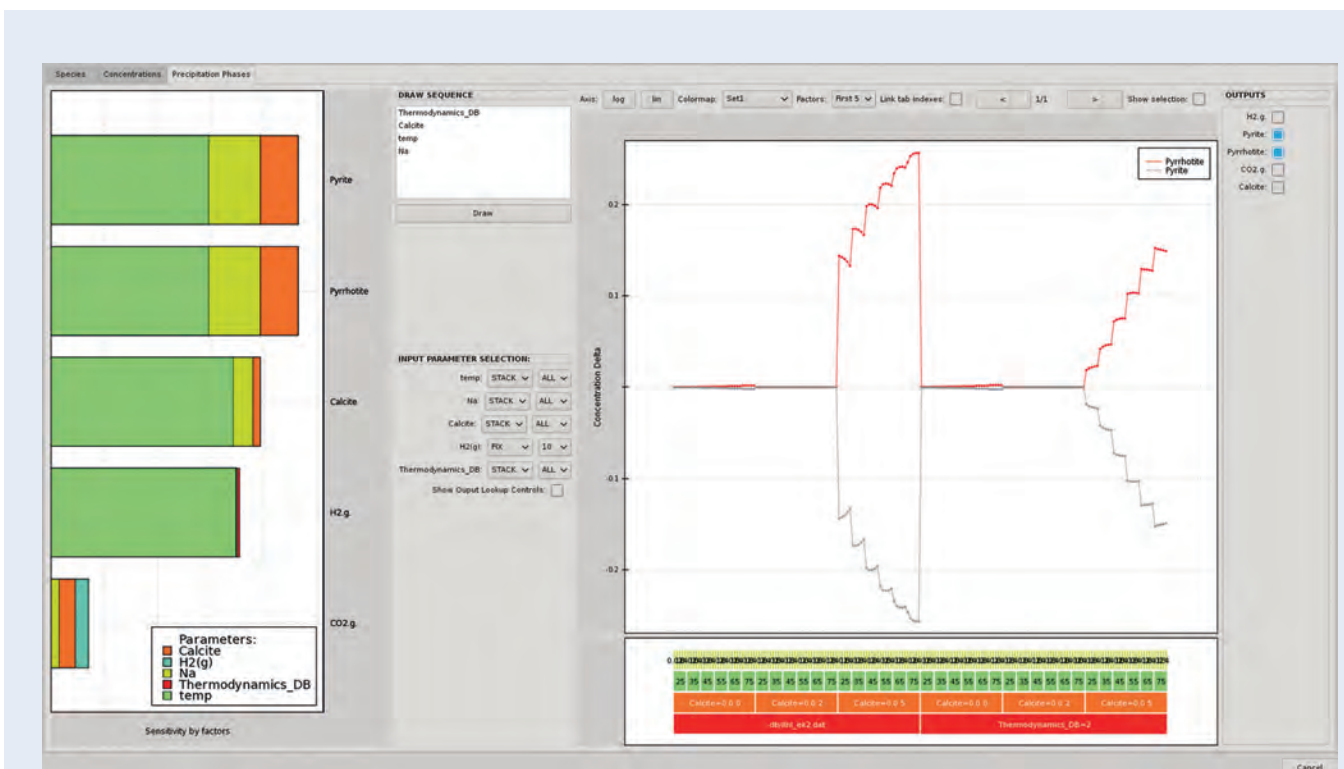


Abb. 2: Kernkomponente von GCex zur Exploration von komplexen geochemischen Simulationsmodellen: Links wird die varianzbasierte Sensitivität für mehrere Eingangsparameter gezeigt. In der Mitte erfolgt die Kontrolle über die Eingaben. Rechts ist das Explorationsfenster mit „stacked parameter relationship“-Visualisierung zu sehen. In diesem Beispiel (De Lucia et al., 2015) wird der Zusammenhang zwischen zwei Ausgaben für vier unterschiedliche Eingaben visualisiert. Mit Expertenwissen kann die Visualisierung interaktiv angepasst werden, insbesondere die Reihenfolge der x-Achsen, so dass Zusammenhänge zwischen den Parametern sichtbar werden.

Fig. 2: The core component of GCex, an approach for the exploration of complex geochemical simulation models. On the left, the variance-based sensitivity is shown for several input parameters. Input controls are provided in the middle. On the right, the exploration window with the stacked parameter relationship visualization is shown. In this example (De Lucia et al., 2015), the relation between two output variables and four different input variables is visualized. The expert can adapt the visualization interactively, especially the order of the x-axes, so that relations among variables becomes apparent.

können. Zusammenhänge zweiter und höherer Ordnung zwischen Parametern können intuitiv ergründet und quantitativ dargestellt werden, was wiederum zur Verfeinerung der Stichprobennahme oder zur Identifikation vernachlässigbarer Parameter führen kann.

SLIVISU: Validierung von Simulationsmodellen glazial-isostatischer Anpassungen

Die dritte Visual-Analytics-Anwendung unterstützt die Validierung von Simulationsmodellen, die glazial-isostatische Anpassungen, also die Deformation des Erdkörpers in Folge von Eislasten, abbilden (Unger et al., 2012). Um zu prüfen, ob ein Simulationsmodell die Realität ausreichend genau zeigt, werden Daten aus dem Simulationsmodell mit realen Daten verglichen. Diese Methode ist Ergebnis einer Kooperation der GFZ-Sektionen „Geoinformatik“ und „Erdsystemmodellierung“.

Pleistozäne Vereisungen in Nordeuropa und -amerika haben zur Deformation des Erdkörpers geführt, zur sogenannten glazial-isostatischen Anpassung (GIA). Die Ausgleichsphase nach dem Abschmelzen der Eismassen dauert bis heute an. Dieser Prozess ist für eine Vielzahl geowissenschaftlicher Prozesse von Bedeutung. So trägt die glazial-isostatische Anpassung zur heutigen Landhebung vor allem in den vormals vereisten Gebieten, zum Meeresspiegelanstieg in angrenzenden Regionen wie den Niederlanden sowie zur globalen Meeresspiegeländerung mit etwa -0,2 mm pro Jahr bei. GIA liefert ein räumlich verteiltes Feld linearer Änderung, um das eine Vielzahl geodätischer Observablen korrigiert werden muss: das zeitlich variable globale GRACE-Schwerefeld, die Polbewegung, die Bewegung des Massen-zentrums, der über Altimetrie und Pegel bestimmte Meeresspiegel sowie die Oberflächendeformation aus GNSS (globales Navigationssatellitensystem)-Daten. GIA ist auch für Klimarekonstruktionen von Bedeutung, da die Interaktion von Meeresspiegel- und Eismassenänderungen einen erheblichen Einfluss auf die Eisdynamik hat und in wesentlichen Prozesssimulationen, wie in der Paläomodellierungsinitiative PalMod des BMBF, berücksichtigt wird.

Zur Quantifizierung der GIA, die auf einer Zeitskala von 10 000 Jahren abläuft, kann die Vertikalbewegung für weiter zurückliegende Epochen nur über Meeresspiegeldaten rekonstruiert werden. Dazu dienen geologische Funde, so genannte Meeresspiegelindikatoren (SLI), über deren Ablagerungsbedingungen Aussagen hinsichtlich des damaligen Meeresspiegels möglich sind. Diese Daten weisen einige Charakteristika auf, die eine systematische Anwendung erschweren. Sie sind punktförmig und ungleichmäßig in Raum und Zeit verteilt und von unterschiedlicher Qualität. Unsicherheiten bezüglich Lokation, Datierung und Aussagekraft müssen für jeden Datenpunkt individuell untersucht werden.

Die Validierung von Erdmodellen muss deshalb interaktiv erfolgen. Nur so können die Geowissenschaftlerinnen und Geowissen-

schaftler bewerten, ob die automatisch berechnete Modellgüte angesichts der Qualität der Meeresspiegelindikatoren aussagekräftig ist. Mit SLIVISU wurde ein Visual-Analytics-Werkzeug entwickelt, das genau diese Aufgabe unterstützt (Abb. 3). Zehntausende Meeresspiegelindikatoren können umfassend mit Hilfe verschiedener, eng verknüpfter Sichten (Karten, Histogramme, Meeresspiegelkurven) über die Zeit und in Datentabellen bewertet werden. Gleichzeitig setzt SLIVISU die Meeresspiegelindikatoren in Bezug zu den Modelldaten und der berechneten Modellgüte, so dass die Aussagekraft der Modellgüte beurteilt werden kann. Die Funktionalität von SLIVISU geht über die Bewertung eines einzelnen Modells hinaus. Ähnlich wie bei den vorher beschriebenen geochemischen Modellen sind auch beim GIA-Modell bestimmte Eingangsparameter nicht sicher bestimmbar. Deshalb erlaubt SLIVISU die Exploration der Modellgüte innerhalb eines Modellensembles, das zahlreiche mögliche Parameterkombinationen abdeckt. So hilft SLIVISU dabei, die Variation der Modellgüte in Abhängigkeit von den Eingangsparametern zu verstehen und so die Werte der Eingangsparameter einzugrenzen.

SLIVISU unterstützt eine umfassende Validierung für Erdmodelle auch dann, wenn nur unsichere und spärlich verteilte terrestrische Observationen zur Verfügung stehen. Geowissenschaftlerinnen und Geowissenschaftler müssen die Validierung nicht mehr auf eine kleine, gut verstandene Auswahl von 30 bis 40 Datenpunkten einschränken, sondern können nun alle zur Verfügung stehenden Datenpunkte einbeziehen und interpretieren. Die Betrachtung eines Modellensembles statt einzelner Modelle vereinfacht zudem substanziell die Identifikation geeigneter Parameterwerte.

Fazit

So unterschiedlich die vorgestellten geowissenschaftlichen Fragestellungen sind, so unterschiedlich sind auch die Visual-Analytics-Konzepte, die am GFZ entwickelt werden. Dennoch hat sich gezeigt, dass diese neuen Methoden allgemeingültige Vorteile haben. Die neuen Visual-Analytics-Methoden erlauben die vollständige Exploration der komplexen Datenmengen. Zuvor haben die vorhandenen Methoden nur eine stichprobenartige Betrachtung von Datenmengen zugelassen. Geowissenschaftlerinnen und Geowissenschaftler können anhand ihres Fachwissens die Daten selbst in solchen Fällen richtig interpretieren, bei denen rein automatische Verfahren nicht zum Ziel führen. Die visuellen Methoden helfen, die Ausgangsdaten sowie die Vorschläge der Algorithmen zu bewerten. Interaktionsmethoden erlauben die Steuerung des Analyseprozesses und die Korrektur der automatischen Vorschläge. Auf diese Weise leisten alle drei Visual-Analytics-Konzepte einen wichtigen Beitrag zur Dateninterpretation und helfen somit bei der Beantwortung sehr unterschiedlicher geowissenschaftlicher Forschungsfragen.

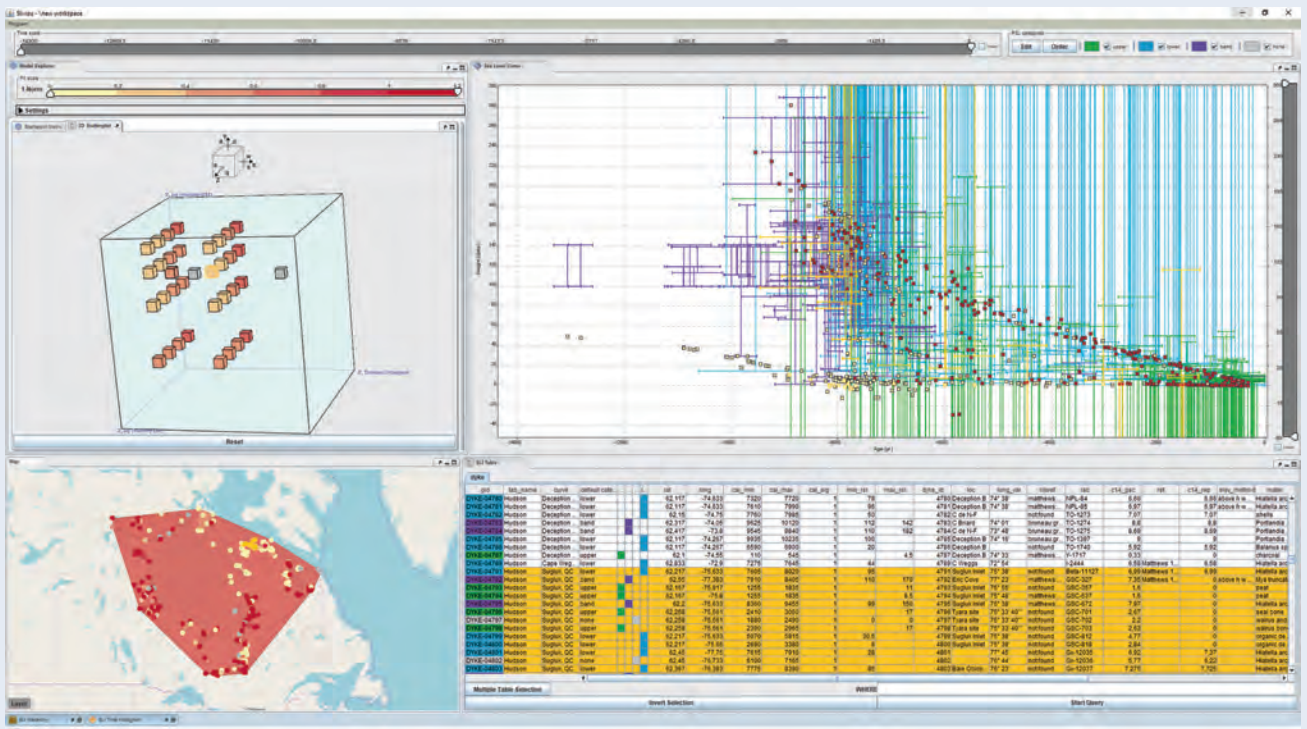


Abb. 3: SLIVISU – Visual-Analytics-Methode für den Vergleich von Simulationsmodellen mit unregelmäßig verteilten und unsicheren Beobachtungsdaten (Unger et al., 2012). Oben links wird der Parameterraum des Modellensembles dargestellt. Im Beispiel wird das Koordinatensystem von drei verschiedenen Modellparametern aufgespannt. Jeder Würfel steht für ein Modell, die Farbe zeigt die Modellgüte (rot: hohe Güte, gelb: geringe Güte). Die Güte wird automatisch für die aktuelle Datenauswahl berechnet. Die Datenauswahl wird interaktiv von den Geowissenschaftlerinnen und Geowissenschaftlern in den übrigen Sichten bestimmt. Die Meeresspiegelkurve oben rechts zeigt die Meeresspiegelhöhen der einzelnen Meeresspiegelindikatoren über die Zeit im Vergleich mit den simulierten Daten des ausgewählten Modells. Die Karte unten links zeigt die räumliche Variabilität der Modellgüte für die Datenauswahl. Die Datentabelle unten rechts bietet detaillierte Angaben zu den einzelnen Meeresspiegelindikatoren. Die enge Verknüpfung von Beobachtungs- und Modelldaten sowie der Modellgüte erlaubt es, die Aussagekraft der Modellgüte anhand der Datenqualität zu bewerten.

Fig. 3: SLIVISU – Visual Analytics method to compare simulation models with sparse and uncertain observation data (Unger et al., 2012). Top left, the parameter space of the model ensemble is shown. In the example, the coordinate system is spanned by three different model parameters. Every cube represents one model, the color depicts the goodness of fit of the model (red – high, yellow – low). The goodness of fit is automatically computed for the current data selection. The data selection is interactively defined by the geoscientists in the other views. On the top right, the sea level curve shows the sea level heights over time for the individual sea level indicators, in comparison with the simulated data of the currently selected model. The map on the bottom left shows the spatial variability of the goodness of the fit for the data selection. The data table on the bottom right provides detailed information about each sea level indicator. The tight linking of observation and model data as well as the goodness of fit enables the geoscientists to assess the reliability of the goodness of fit based on the data quality

Literatur

- De Lucia, M., Jatnieks, J., Sips, M. (2015): GCex: A Visual Analytics Approach for Interactive Exploration of Geochemical Models. - *Energy Procedia*, 76, pp. 616–622. DOI: <https://doi.org/10.1016/j.egypro.2015.07.881>
- Unger, A., Dräger, N., Sips, M., Lehmann, D. J. (2018): Understanding a sequence of sequences: Visual exploration of categorical states in lake sediment cores. - *IEEE Transactions on Visualization and Computer Graphics*, 24, 1, pp. 66–76. DOI: <https://doi.org/10.1109/TVCG.2017.2744686>

- Unger, A., Schulte, S., Klemann, V., Dransch, D. (2012): A Visual Analytics Concept for the Validation of Geoscientific Simulation Models. - *IEEE Transactions on Visualization and Computer Graphics*, 18, 12, pp. 2216–2225. DOI: <https://doi.org/10.1109/TVCG.2012.190>

Videos

