

Visuelle Analyse großer Daten in der Klimaforschung

Niklas Röber · Michael Böttinger

Das Klima und die Klimaforschung bekommen in den aktuellen Debatten, sowohl in der öffentlichen Wahrnehmung als auch in der Politik, endlich die Bedeutung, die dieses wichtige Thema verdient. Präzise Projektionen zur Entwicklung des Klimas unter veränderten Bedingungen sind erforderlich und schon vielfach berechnet und analysiert worden. Aber auch die Klimaforschung entwickelt sich weiter, die Daten werden komplexer, da die Modelle durch gestiegene Rechenleistung immer feiner aufgelöst rechnen können, sodass sich auch kleinräumige Phänomene, wie z. B. Wolken realistischer abbilden und simulieren lassen. Die dabei von Klimamodellen produzierte Datenmenge wächst seit Jahren kontinuierlich an. Aktuell sind am DKRZ etwa 120 PByte Klimasimulationsdaten archiviert, mit einem Zuwachs von 40 PByte allein im letzten Jahr. Neue Modelle können die Feinheiten des Klimasystems besser nachbilden und mehr Prozesse, Abhängigkeiten und Wechselwirkungen in der Simulation berücksichtigen. Zusätzlich können heute anhand von Ensemblesimulationen nicht nur die Robustheit der Modellergebnisse bestimmt, sondern auch statistisch unterlegte Aussagen zum Eintreffen bestimmter Ereignisse gemacht werden. Ensemblesimulationen werden schon seit einigen Jahren operationell in der Wettervorhersage eingesetzt und gewinnen nun auch in der Klimaforschung zunehmend an Bedeutung.

Klimamodelle

Zur Simulation des Klimas über längere Zeiträume werden gekoppelte Modelle eingesetzt, die unterschiedliche Aspekte bzw. Kompartimente des Klimasystems mit unterschiedlichen Teilmodellen abbilden. Das Modell MPI-ESM [11], welches auch für die Simulationen eingesetzt wurde, die im Hinblick auf den deutschen Beitrag zum aktuellen fünften Weltklimastatusbericht durchgeführt wurden, beinhaltet neben den klassischen Komponenten für Atmosphäre, Ozean (inkl. Meereis) und Landoberfläche auch ein Modell der marinen Biogeochemie sowie der Landbiosphäre. Ein Koppler sorgt für den Austausch der Informationen zwischen den einzelnen Teilmodellen, die oft auch unterschiedliche Rechengitter und -zeitschritte verwenden. Jedes Teilmodell berechnet für das entsprechende Rechengitter und für jeden Zeitpunkt eine Vielzahl prognostischer und diagnostischer Variablen, die es auszuwerten gilt.

ICON (ICOsahedral Non-hydrostatic) ist ein neues und für vielseitige Experimente verwendetes Modell, welches in Kooperation zwischen dem MPI-M, dem DWD und dem DKRZ entwickelt wird [12]. Als Basis dient ein Ikosaedergitter mit gleich großen Gitterzellen, bei dem die Daten in der Mitte der Dreiecke, auf den Eckpunkten und auf der Mitte der Kanten liegen. Das vertikale Gitter ist rektilinear und in der Nähe der Erd- bzw. Ozeanoberfläche feiner aufgelöst. Obwohl das ICON-Modellgitter unstrukturiert ist, hat es gegenüber anderen in der Klimamodellierung gebräuchlichen Gittern Vorteile: Die mathematischen Polstellen entfallen, das Gitter lässt sich in Regionen leicht verfeinern und die

> https://doi.org/10.1007/s00287-019-01222-w © The Author(s) 2019.

Niklas Röber - Michael Böttinger Deutsches Klimarechenzentrum GmbH (DKRZ),, Bundesstraße 45a, 20146 Hamburg E-Mail: roeber@dkrz.de Kopplung von Ozean und Atmosphäre ist einfacher und weniger rechenintensiv.

In den vergangenen Jahren wurde ICON im Rahmen des vom BMBF geförderten Projektes HD(CP)² ("High Definition Clouds and Precipitation for Advancing Climate Prediction") derart erweitert, dass es in hohen räumlichen Auflösungen Wolken und Niederschlag in einem regionalen Modell-Set-up explizit auflösen und berechnen kann [13]. In der feinsten regionalen Auflösung (156 m pro Zelle, simuliertes Gebiet: Deutschland) bestehen 2-D-Skalarfelder aus 22 Mio., 3-D-Skalarfelder aus knapp 3,3 Mrd. Zellen. Dies entspricht in etwa 90 MByte bzw. 13 GByte pro 2-D-/3-D-Variable und Zeitschritt. Obwohl diese Daten schon recht groß sind, können sie dennoch mit einem klassischen sog. Postvisualisierungsworkflow analysiert und ausgewertet werden.

Im kürzlich gestarteten und von der EU geförderten Projekt ESiWACE2 soll die Auflösung globaler Modelle weiter verfeinert werden. Bei der angestrebten horizontalen Auflösung von etwa 1,25 km werden dann 360 Mio. Zellen pro Schicht und - je nach Anzahl der Schichten - 30-60 Mrd. Zellen pro 3-D-Variable und Zeitschritt anfallen. Die dabei produzierte Datenmenge wird zu groß sein, um alle für eine sinnvolle Analyse benötigten Daten in voller räumlicher und akzeptabler zeitlicher Auflösung speichern zu können. 2-D-Daten können noch parallel verarbeitet und visualisiert werden, bei globalen 3-D-Daten ist aber auch dies nicht mehr möglich und alternative Strategien zur Bewältigung dieser Datenmenge sind erforderlich. Beispiele sind hier die In-situ-Visualisierung sowie die Datendekomposition (einschließlich Kompression), welche einen progressiven Datenzugriff erlaubt.

Am DKRZ sind die Rechner für die Datenvorund -nachbearbeitung, Analyse und Visualisierung integraler Teil des Hochleistungsrechners und ebenfalls mit dem parallelen Dateisystem verbunden. Der Datenzugriff erfolgt über das Netzwerk, entweder über die Konsole oder für die 3-D-Visualisierung über eine VNC-Schnittstelle, mit dem Vorteil, dass alle Daten im Rechenzentrum verbleiben und nicht extra kopiert oder heruntergeladen werden müssen. Die Daten sind üblicherweise in NetCDF-Containern gespeichert, teilweise aber auch als GRIB verlustbehaftet komprimiert. Zur Datenverarbeitung kommen dabei die Werkzeuge NCO (NetCDF Operators), CDO (Climate Data Operators), aber auch Python, matplotlib, MatLab, NCL und IDL zum Einsatz. Für die 3-D-Visualisierung werden ParaView, VAPOR und Met.3D genutzt. Der Artikel von Rautenhaus et al. [23] gibt einen ausführlichen Überblick über den Einsatz von Visualisierungswerkzeugen in der Meteorologie.

Analyse und Visualisierung von Ensemblesimulationen

Nicht alle Daten sind per se groß. Eine einzelne Realisation einer Ensemblesimulation ist durchaus handhabbar und leicht zu analysieren, alle Ensemblemitglieder aber - teilweise mehr als 100 - nicht. Die zunehmende Verwendung von Ensembletechniken in der Wetter- und Klimaforschung trägt so zur exponentiellen Zunahme des Datenvolumens bei. Um etwa die Unsicherheit in den Anfangsbedingungen einer Simulation zu berücksichtigen, wird diese entsprechend mehrfach mit leicht variierten Anfangsbedingungen durchgeführt. Der resultierende Ensembledatensatz umfasst unterschiedliche Verläufe des Wetter- bzw. Klimageschehens. Bei Wettervorhersagen wird so etwa die Unsicherheit in den Anfangsbedingungen modelliert. Eine statistische Analyse des Ensembles erlaubt Aussagen zu Wahrscheinlichkeiten für bestimmte Wetterentwicklungen. Bei Klimasimulationen repräsentiert die Bandbreite der Entwicklungen im Idealfall die natürliche Variabilität. Damit hat ein Ensemble gegenüber einer einzelnen Simulation also eine zusätzliche Dimension, die Unsicherheit oder - positiv ausgedrückt - die Robustheit. Eine weitere Stufe der Komplexität erreicht hier die Auswertung der Simulationen, die weltweit von etwa 25 Einrichtungen mit insgesamt 40 Modellen im Hinblick auf den fünften IPCC-Bericht durchgeführt wurden. Dieses Multimodellensemble ist die wesentliche Datenbasis für die im Bericht vorgestellten Klimaprojektionen, die anhand unterschiedlicher Szenarien den Raum möglicher Klimazukünfte für unterschiedliche sozioökonomische und technologische Entwicklungen ausleuchten.

Durch Ensembletechniken lassen sich Vorhersagen qualitativ verbessern, aber für deren Auswertung und Erschließung des Mehrwertes sind aufgrund der gestiegenen Komplexität und Datenmenge weitergehende Methoden erforderlich. Dies betrifft sowohl den Einsatz der Statistik in der Datenbearbeitung als auch die Visualisierung. Eine aktuelle und umfassende Übersicht zur Visualisierung und



Abb. 1 Darstellung der für die Szenarien RCP2.6 und RCP8.5 projizierten Temperaturänderung für den Zeitraum 2081–2100 verglichen mit 1986–2005 aus der "Summary for Policymakers" des fünften IPCC-Sachstandsberichtes zu den physikalischen Grundlagen des Klimawandels [15, dort Figure SPM.8(a)]. Die überlagerten Punktmuster zeigen Gebiete an, wo die projizierte Klimaänderung statistisch robust ist und wo gleichzeitig mehr als 90 Prozent der IPCC-Modelle über das Vorzeichen der Temperaturänderung einig sind.

visuellen Analyse von Ensembledaten findet sich in dem Übersichtsartikel von Wang et al. [28].

Bei der Darstellung von Ensembleergebnissen beschränkt man sich oft auf das Ensemblemittel als "beste Vorhersage", die Information über die Robustheit der Vorhersage geht dabei aber verloren. Hier ist gefordert, die prognostizierte Größe gemeinsam mit der dazugehörigen Unsicherheit darzustellen. Man findet heute Darstellungen der Unsicherheit integraler Größen auch in der Kommunikation an die Öffentlichkeit, wie beispielsweise Kurven eines prognostizierten Temperaturverlaufs mit einem hinterlegten schattierten Bereich, der größer wird, je weiter die Vorhersage in die Zukunft geht. Die Schattierung zeigt mögliche, die Kurve die wahrscheinliche Entwicklung.

Die Visualisierung zwei- und dreidimensionaler zeitabhängiger Ensembledatensätze stellt zusätzliche Anforderungen. Im Fall einer zweidimensionalen Größe, die man als Isoflächenkarte darstellt, kann die zusätzliche statistische Information z. B. durch eine Überlagerung mit Isolinien oder gemusterten Isoflächen gezeigt werden. Diese Technik ("stippeling", "hatching") wird etwa im fünften IPCC-Bericht verwendet, um die projizierte zukünftige Erwärmung sowie gleichzeitig die Robustheit der Projektionen darzustellen. Abbildung 1 zeigt für die beiden extremen Szenarien RCP8.5 ("business as usual") und das RCP2.6 (Abkehr von fossilen Brennstoffen) die aus der Zusammenfassung für Entscheidungsträger des IPCC-Berichtes entnommene Darstellung der mit dem Multimodellensemble für das Ende dieses Jahrhunderts

projizierten Temperaturänderungen und deren Robustheit.

Obwohl sich diese Zusammenfassung an Nichtexperten richtet, ist die Abbildung schwierig zu lesen; die Musterung, die anzeigt, wo das Klimaänderungssignal statistisch robust ist, verdeckt teilweise die eigentlich relevantere Darstellung der Temperaturänderung. Am DKRZ haben wir hierfür eine alternative Visualisierung erstellt, in der die nichtrobusten Bereiche halbtransparent schattiert werden, während die Bereiche, in denen das Signal größer als das Rauschen ist, die projizierte Temperaturänderung in klaren Farben zeigen (Abb. 2).

Neben der 2-D-Überdeckung durch Isolinien oder Muster bei 2-D-Karten kann für die Visualisierung der zusätzlichen Dimension der Unsicherheit bzw. Robustheit in 3-D-Visualisierungssystemen auch die dritte Dimension genutzt werden, etwa durch eine Kombination von Isolinien und Überhöhung, die gleichzeitig den Vorhersage-Skill und die natürliche Variabilität zusätzlich zur Vorhersagegröße darstellen [8]. Bei der 3-D-Visualisierung zeitabhängiger 3-D-Ensembledaten ist es allerdings nicht mehr möglich, eine zusätzliche Dimension für die Darstellung zu nutzen. Hier müssen die in den raumfüllenden Rohdaten versteckten Informationen deutlich stärker reduziert werden, um wesentliche Prozesse oder Merkmale sichtbar zu machen. Am Beispiel des Jetstreams zeigen Kern et al. [18] beispielsweise eine Kombination aus Merkmalsextraktion und Unsicherheitsvisualisierung mit dem Visualisierungstool Met.3D [24], die eine sehr starke Reduktion der Daten erfordert.



Abb. 2 Visualisierung der für die Szenarien RCP2.6 und RCP8.5 projizierten Temperaturänderung für die Zeitpunkte 1980, 2031, 2082 und 2100 verglichen jeweils mit dem Zeitraum 1986–2005. Statistisch robuste Temperaturänderungen sind in klaren Farben dargestellt. Die graue halbtransparente Schattierung zeigt an, wo die simulierten Temperaturänderungen kleiner als zwei Standardabweichungen der internen Variabilität sind.

Eine weitere große Herausforderung liegt in der Analyse der räumlich-zeitlichen Entwicklung in Ensemblesimulationen. Das Klimasystem ist von einer großen Variabilität auf verschiedenen Raumund Zeitskalen geprägt. Je nach betrachteter Größe gibt es bestimmte Zustände, die gehäuft auftreten. Ein prominentes Beispiel hierfür ist El Niño, eine quasiperiodisch auftretende Oszillation in der Meeresoberflächentemperatur des tropischen Pazifiks. Das gehäufte Auftreten wiederkehrender Muster im zeitlichen Verlauf von Ensemblesimulationen kann z. B. mithilfe von Clustering-Verfahren oder topologischen Methoden klassifiziert und analysiert werden [16, 17].

In-situ-Visualisierung

Die Idee, Simulationsdaten in situ, also parallel zur Simulation, zu visualisieren, ist nicht neu, sondern reicht bis in die frühen 1990er-Jahre zurück, in die Ära des sog. Co-Processing [7, 29]. Sie hat sich aber dennoch bis heute nicht wirklich durchgesetzt und kommt primär dann zum Einsatz, wenn der bisher genutzte Postanalyseansatz versagt. Grund hierfür ist, dass nicht nur die Simulationen und Daten immer größer und komplexer wurden, auch die Leistungsfähigkeit der Hard- und Software, welche für die Visualisierung genutzt wird, hat sich stetig weiterentwickelt. Mittlerweile klafft aber eine immer größer werdende Lücke zwischen den Möglichkeiten, die Daten zu berechnen und diese dann auch auf einer Festplatte abzuspeichern. Die aktuellen Zuwachsraten von Dateisystemen skalieren schon lange nicht mehr mit den gestiegenen Anforderungen eines modernen HPC-Systems.

Dieser Umstand wird durch Leuchtturmprojekte wie ESiWACE2 noch beschleunigt. Bei einer globalen Simulation der Atmosphäre in 1,25 km Auflösung fallen für einen simulierten Tag – bei 30-minütigem Speicherintervall – in etwa 43 TByte an Daten an, für 20 2-D- und acht 3-D-Variablen mit je 75 Schichten. Aus diesem Grund wird am DKRZ an einem Prototyp gearbeitet, der die Möglichkeiten einer an das ICON-Modell gekoppelten In-situ-Visualisierung auslotet. Aufgrund von positiven Erfahrungen und öffentlich verfügbarem Code haben wir uns für ParaView und seine In-situ-Komponente Catalyst entschieden [1–3]. Andere In-situ-Frameworks sind z. B. Visit/libsim [9] so-



Abb. 3 In-situ-Workflow mit ParaView/Catalyst

wie die beiden Exascale-Visualisierungsinitiativen SENSEI [4] und ALPINE/Ascent [19].

ParaView wird genutzt, um mithilfe von Beispieldaten ein Python-Skript zu erzeugen, welches die spätere In-situ-Datenverarbeitung und -Visualisierung beschreibt. Dies kann das Erzeugen einer Visualisierung verbunden mit der Speicherung von Bildern im Abstand von 10 Modellzeitschritten sein, aber auch eine Speicherung reduzierter Daten. Als Einsatzgebiet für die In-situ-Datenverarbeitung sehen wir die folgenden Felder:

- Erstellung einer Vorschau (Preview),
- Datenreduktion und parallele Speicherung,
- Überwachung der Simulation zur Laufzeit (Livevisualisierung),
- Detektion, Verfolgung und Analyse bestimmter Features (z. B. Wolkentypen, Ozeanwirbel).

Die Modifikation des eigentlichen ICON-Modells beschränkt sich dabei auf wenige Zeilen Code. Die Daten, Gitterinformation und prognostische Variablen, werden per Zero-copy von FORTRAN über einen Adapter an C++ und das Visualisierungstoolkit VTK weitergereicht. VTK, welches auch ParaView zugrunde liegt, arbeitet dann entsprechend das oben erwähnte Python-Skript ab, generiert dabei Bilder und/oder speichert Daten [3]. Zusätzlich lassen sich die simulierten Daten gleichzeitig auch live mit ParaView visualisieren, sodass man der Simulation direkt zuschauen kann. Hierbei werden die visualisierten Daten aktualisiert, sobald von der Simulation neue Daten an Catalyst zur Verfügung gestellt werden. Abbildung 3 verdeutlicht den Workflow und die Einsatzgebiete.

Natürlich kann man die in situ zur Verfügung stehenden Daten auch für eine Verarbeitung jenseits der Bildgenerierung nutzen. So ist es aktuell schon möglich, anhand der ISCCP-Kategorien [26] eine automatische Klassifikation der verschiedenen Wolkentypen in wolkenauflösende Simulationen zu erstellen. Die einzelnen Wolkentypen können dabei recht einfach anhand der Variablen Wolkenwasser, Wolkeneis, Regenmenge, Höhe, Temperatur, Druck, Luftfeuchte und vertikale Windgeschwindigkeit bestimmt werden. Meteorologen interessieren sich hierbei besonders für die zeitliche Entwicklung der Wolken und den Übergang von einem Wolkentyp in einen anderen. Nur Zellen, welche auch Wolken beinhalten, werden am Ende der Detektion als Integer Array auf der Festplatte platzsparend gespeichert. Eine automatische Detektion von Ozeanwirbeln ist zurzeit noch in Arbeit. Hier sollen neben den eigentlichen Wirbeln zusätzlich noch quantitative Informationen, wie z. B. Größe und Volumen der Wirbel, Geschwindigkeit, Energiegehalt und Lebensdauer, mit abgespeichert werden.

Eine In-situ-Visualisierung ist entweder eng oder locker an die Simulation gekoppelt, d. h., die Visualisierung erfolgt direkt auf den Rechenknoten, auf denen schon die Simulation läuft, oder separat auf dedizierten Visualisierungsknoten [17]. In unserem Set-up ist eine enge Kopplung sehr viel performanter und effizienter. Das größte In-situ-Experiment am DKRZ benötigte 540 Knoten des HPC-Systems Mistral, d. h. 1/6 des gesamten Rechners. Dabei wurde die Atmosphäre mit ICON global in einer Auflösung von 2,5 km simuliert. Pro Knoten rechneten dabei 8 ICON-MPI-Prozesse und zusätzlich noch einmal 8 Catalyst-MPI-Prozesse, insgesamt also 8640 MPI-



Abb. 4 Visualisierung von in situ reduzierten Variablen Wolkenwasser und Wolkeneis einer globalen ICON-Atmosphärensimulation mit einer 2,5 km Auflösung.

Prozesse. In situ wurden dabei Daten visualisiert und als Bild gespeichert sowie die beiden 3-D-Variablen Wolkenwasser und Wolkeneis mit einem Schwellenwert von 1,0e-7 kg/kg versehen, reduziert und parallel gespeichert. Dabei konnten die 3-D-Daten pro Zeitschritt um den Faktor 40 auf knapp 150 Mio. Zellen pro 3-D-Variable reduziert werden, ohne für die Visualisierung relevante Daten zu verlieren; Abb. 4 zeigt einen Ausschnitt einer Visualisierung der Ergebnisse. Die Initialisierung von Modell und Visualisierung hat insgesamt 6,1 s (ca. 11%) länger gedauert als ohne In-situ-Visualisierung. Pro Zeitschritt, bei dem Catalyst tätig wird, verlängert sich die Zeit um maximal 1 s (ca. 3 %). Dies zeigt, dass die Zeit, die für eine In-situ-Visualisierung zusätzlich benötigt wird, im Vergleich mit der eigentlichen Modelllaufzeit fast vernachlässigbar ist.

Kompression und progressive Datenvisualisierung

Nachteil einer in-situ-basierten Verarbeitung von Simulationsdaten ist das erforderliche Wissen um die für die Darstellung benötigten Parameter, wie z. B. Schwellenwerte für die Datenreduktion oder Isoflächen, welche schon vor dem Simulationslauf gesetzt werden müssen. Ohne die richtigen Werte wird es schwierig, die gesuchten Merkmale korrekt zu extrahieren und zu visualisieren. Zudem können unbekannte bzw. unerwartete Strukturen in den Daten verborgen und somit unentdeckt bleiben. Daher zielt eine zweite Möglichkeit im Umgang mit sehr großen Datenmengen auf eine Umordnung der Daten und einen progressiven Visualisierungsansatz. Hierbei werden die Daten transformiert und in einer Level-of-Detail-(LoD-)Datenstruktur abgespeichert. So ist der Zugriff auf die Daten nicht nur schneller möglich, es werden auch nur die Daten geladen und visualisiert, die sich im aktuellen Sichtfeld befinden. Details können automatisch nachgeladen oder für eine schnellere Interaktion auch zurückgehalten werden. Wichtig bei einer solchen Zerlegung ist natürlich, dass die für die Wissenschaft interessanten Strukturen auch in niedrigeren Auflösungen erkennbar bleiben, sodass der Wissenschaftler weiß, welche Gebiete für ihn interessant sind und wo sich das Nachladen der Details lohnt.

Ein klassisches Verfahren zur Erzeugung solcher LoD-Strukturen sind Wavelets [25, 29]. Für irreguläre Gitter – wie ICON – ist dies ein wenig komplizierter, aber dennoch möglich. Das ICON-Gitter lässt sich in einzelne Ikosaedergitter – sog. Icosahedral Maps – zerlegen, um einen einfacheren Zugriff auf die Daten zu gewährleisten [14], Abb. 5. In dieser Repräsentation lässt sich sehr leicht eine Wavelet-Transformation auf die Daten anwenden und eine mehrstufige Wavelet-Analyse durchführen, bei welcher die Daten in eine LoD-Datenstruktur überführt werden. Die dabei extrahierten Detailko-



Abb. 5 Auffaltung eines globalen ICON-Gitters in 10 Diamanten. Die Gitterkoordinaten lassen sich nun wie in einem rektilinearen 2-D-Gitter ansprechen.



Abb. 6 Beispielhafte Anwendung einer Wavelet-Transformation auf niedrig aufgelöste ICON-Ozeandaten: links Originaldaten, Icosahedral Maps, Wavelet-Analyse sowie Rekonstruktion mit 10 %- bzw. 95 %-Datenreduktion.

effizienten lassen sich zudem noch (verlustbehaftet) komprimieren, um zusätzlichen Platz zu sparen und um die spätere Rekonstruktion effizienter zu machen [14].

Abbildung 6 verdeutlicht noch einmal das Prinzip der Analyse und Rekonstruktion mithilfe von Wavelets für alle drei ICON-Gitter am Beispiel einer niedrig aufgelösten Ozeansimulation. Links sind die Originaldaten zu sehen, welche dann auf das Ikosaedergitter (Icosahedral Maps) abgebildet werden. Dann erfolgt die Wavelet-Analyse und zeigt eine niedriger aufgelöste Version der Daten sowie zwei daraus erfolgte Rekonstruktionen mit einer 10- bzw. 95-prozentigen Datenreduktion. Dabei ist erkennbar, dass die Daten sehr gut auf Kompression ansprechen; mehr Details dazu finden sich in [14]. Wissenschaftler am NCAR untersuchten bereits verschiedene verlustbehaftete Kompressionstechniken und konnten nachweisen, dass selbst eine starke Kompression nicht unbedingt das statistische Signal der Daten nachhaltig beeinträchtigt [5, 6].

Die hier vorgestellten Möglichkeiten der Wavelet-Zerlegung von ICON-Daten sind momentan noch in der Entwicklung, aber bereits prototypisch implementiert und getestet. Implementiert werden sie nun in C++ als zusätzliche In-situ-Methode (s. auch Abb. 3). VAPOR, eine interaktive 3-D-Visualisierungssoftware, welche am NCAR entwickelt wird, wird später für die progressive Visualisierung und Analyse dieser LoD-Daten eingesetzt [10]. VAPOR unterstützt bereits eine progressive Datenvisualisierung, jedoch bisher nur für rektilineare Gitter. In einer Kooperation mit dem NCAR wurde VAPOR bereits erweitert, sodass VAPOR nun auch irreguläre Gitter nach dem UGrid-CF-Standard unterstützt. Nun fehlt nur noch eine Erweiterung, die auch eine progressive Visualisierung für irreguläre Daten ermöglicht, woran zurzeit aber noch gearbeitet wird.

Ausblick

Die Datenmengen in der Klimaforschung werden in der Zukunft eher zu- als abnehmen. Nicht alle Simulationen in der Zukunft werden hochauflösend rechnen, dennoch treiben gerade solche Simulationen die Entwicklung nicht nur im Simulations- und HPC-Bereich, sondern gerade auch in der Visualisierung und Datenanalyse voran. Um die Ergebnisse solcher Simulationen aber effektiv und effizient auswerten zu können, braucht es konzertierte Anstrengungen in unterschiedlichen Richtungen sowie einen Paradigmenwechsel im Workflow: Die Auswertung und Visualisierung werden mit der Simulation in Zukunft stärker verzahnt werden, die oben beschriebene In-situ-Visualisierung ist hiervon nur eine erste Ausprägung. Die Wissenschaftler werden sich schon im Vorfeld genauer überlegen müssen, welche Daten sie interessieren und wie sie diese Daten analysieren wollen. Die Speicherung der Simulationsdaten wird zunehmend schwieriger, gleichzeitig stehen aber mehr Ressourcen für die Simulation selbst zur Verfügung. So könnte es in naher Zukunft eine Simulation on demand geben, bei welcher der Nutzer die Simulation via Computational Steering sogar direkt aus der Visualisierungsumgebung startet, steuert und kontrollieren kann. Die Analyse und Auswertung der Daten stehen im Vordergrund und nicht mehr die Simulation. Die Daten werden live und in situ visualisiert und nur das Wichtigste auf der Platte gespeichert. Von der Visualisierungsumgebung hätte der Nutzer direkte Kontrolle über das Modell und könnte Parameter sogar zur Laufzeit verändern. Auch dieser Ansatz des Computational Steering geriet, ähnlich wie Insitu selbst, etwas in Vergessenheit, wird aber gerade im Rahmen der Exascale-Visualisierungsinitiative SENSEI neu belebt [4].

Zu der enorm großen Datenmenge kommt zusätzlich noch eine Fülle an Information, die in den Daten enthalten ist und die es auszuwerten gilt. Hier könnte es schwierig werden, die wirklich wichtigen Informationen zu finden bzw. so hervorzuheben, dass der Wissenschaftler sie auch findet.

Analysetechniken, die auf eine Extraktion und Visualisierung der Topologie setzen, werden sich hier als hilfreich erweisen, da sie in der Lage sind, große Daten auf ihre Kerninformation zu reduzieren [17, 27]. Zudem könnte die künstliche Intelligenz (KI) als eines der neueren Werkzeuge in der Datenanalyse helfen und den Wissenschaftler bei seiner Arbeit unterstützen. Zum einen kann eine KI die Simulation dahin gehend überwachen, dass zum einen das, was simuliert wird, meteorologisch plausibel ist, z. B. das Vorhandensein bestimmter Meeresströmungen oder Niederschlagsmuster [21], zum anderen den Wissenschaftler aber auch auf außergewöhnliche Ereignisse in diesen Daten hinweisen, z. B. die El-Niño- oder La-Niña-Phänomene. Matsuoka et al. nutzen beispielsweise Deep Learning, um tropische Wirbelstürme und deren Vorläufer zu detektieren, wodurch die Frühwarnung vor diesen Ereignissen verbessert werden kann [20]. Das System ist dabei sogar in der Lage, mit einer 75 %igen Genauigkeit die Vorläufer solcher tropischen Wirbelstürme bis zu 7 Tage vorher korrekt zu detektieren. Deep Learning kann aber auch schon bei der Simulation selbst eingesetzt werden, etwa um mit weniger Ressourcen und einem gut trainierten neuronalen Netz atmosphärische Prozesse noch auflösen und simulieren zu können, die bei der gewählten Gittergröße eigentlich nur über eine Parametrisierung zu beschreiben sind [22]. Auch an diesen Möglichkeiten wird am DKRZ geforscht und gearbeitet.

Open Access. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Literatur

- 1. Ahrens J, Geveci B, Law C (2005) ParaView: an end-user tool for large data visualization. Visualization Handbook. Elsevier
- 2. Ayachit U (2015) The ParaView Guide: a parallel visualization application. Kitware
- Ayachit U et al (2015) ParaView catalyst: Enabling in situ data analysis and visualization. In: Proceedings of the First Workshop on In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization (ISAV2015). ACM, New York, pp 25–29
- Ayachit U et al (2016) The SENSEI generic in situ interface. In: Second Workshop on In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization (ISAV), pp 40–44

- Baker AH, Hammerling DM, Mickelson SA et al (2016) Evaluating lossy data compression on climate simulation data within a large ensemble. Geosci Model Dev 9:4381–4403
- Baker AH, Xu H, Hammerling DM, Li S, Clyne J (2017) Toward a multi-method approach: lossy data compression for climate simulation data. In: ISC High Performance Workshops. LNCS 10524, pp 30–42
- Bauer A et al (2016) In situ methods, infrastructures, and applications on HPC platforms – a state-of-the-Art (STAR) Report. Comput Graphics Forum 35(3):577– 597
- Böttinger M, Pohlmann H, Röber N, Meier-Fleischer K, Spickermann D (2015) Visualization of 2D Uncertainty in Decadal Climate Predictions. EuroVis Workshop on Visualisation in Environmental Sciences (EnvirVis), Cagliari
- Childs H, Brugger E, Whitlock B, Meredith J, Ahern S, Pugmire D, Biagas K et al (2012) Vislt: an end-user tool for visualizing and analyzing very large data. In: High Performance Visualization–Enabling Extreme-Scale Scientific Insight, pp 357–372
- Clyne J, Mininni P, Norton A, Rast M (2007) Interactive desktop analysis of high resolution simulations: Application to turbulent Plume Dynamics and current Sheet Formation. New J Phys 9:301
- Giorgetta M et al (2013) Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. J Adv Mod Earth Syst, https://doi.org/10.1002/jame.20038, pp 572–597
- Giorgetta M et al (2018) ICON-A: the atmospheric component of the ICON Earth System Model. Part I: Model description. J Adv Model Earth Syst 10:1613–1637
- 13. Heinze R, Dipankar A et al (2017) Large-eddy simulations over Germany using ICON: a comprehensive evaluation. QJR Meteorol Soc 143:69–100
- Imrul Jubair M, Alim U, Röber N, Clyne J, Mahdavi-Amiri A (2016) lcosahedral Maps for a Multiresolution Representation of Earth Data. VMV 2016 – Vision, Modeling and Visualization. Bayreuth
- IPCC (2013) Summary for Policymakers, Book Section SPM. Cambridge University Press, Cambridge, United Kingdom New York, pp 1–30
- Kappe CP, Böttinger M, Leitte H (2018) Exploring variability within ensembles of decadal climate predictions. IEEE Trans Vis Comput Graphics 25(3):1499–1512
- 17. Kappe CP, Böttinger M, Leitte H (2019) Topology-based feature detection in climate data. Workshop on Visualisation in Environmental Sciences (EnvirVis)

- Kern M, Hewson T, Sadlo F, Westermann R, Rautenhaus M (2017) Robust detection and visualization of jet-stream core lines in atmospheric flow. IEEE Trans Vis Comput Graphics 24(1):893–902
- Larsen M et al (2017) The ALPINE in situ infrastructure: ascending from the ashes of strawman. In: Proceedings of the In Situ Infrastructures on Enabling Extreme-Scale Analysis and Visualization (ISAV2017). ACM, New York, pp 42–46, https://doi.org/10.1145/3144769.3144778
- Matsuoka D, Nakano M, Sugiyama D, Uchida S (2018) Deep learning approach for detecting tropical cyclones and their precursors in the simulation by a cloudresolving global nonhydrostatic atmospheric model. J Prog Earth Planet Sci 80(5):1
- 21. Matsuoka D et al (2019) Automatic detection of stationary fronts around Japan using a deep convolutional neural network. SOLA:2019-028
- Rasp S, Pritchard MS, Gentine P (2018) Deep learning to represent subgrid processes in climate models. Proc Natl Acad Sci United States Am 115(39): 9684–9689
- Rautenhaus M, Böttinger M, Siemen S, Hoffman R, Kirby RM, Mirzargar M, Röber N, Westermann R (2017) Visualization in meteorology – a survey of techniques and tools for data analysis tasks. IEEE Trans Vis Comput Graphics 24(12):3268–3296
- Rautenhaus M, Kern M, Schäfler A, Westermann R (2015) Three-dimensional visualization of ensemble weather forecasts – Part 1: The visualization tool Met. 3D (version 1.0). Geosci Model Dev 8:23–29
- Rodriguez B, Gobbetti M, Iglesias M, Guitian J, Makhinya M, Marton F, Pajarola R, Suter S (2014) State-of-the-Art in compressed GPU-based direct volume rendering. Comput Graphics Forum 33:77–100
- Schiffer RA, Rossow WB (1983) The international satellite cloud climatology project (ISCCP): The first project of the World Climate Research Programme. Bull Amer Meteorol Soc 64:779–784
- Tierny J (2017) Topological Data Analysis for Scientific Visualization. Mathematics and Visualization. Springer International, ISBN: 3319715062
- Wang J, Hazarika S, Li C, Shen H (2019) Visualization and visual analysis of ensemble data: a survey. IEEE Trans Vis Comput Graphics 25(9):2853–2872
- 29. Wes Bethel E, Childs H, Hansen C et al (2016) High performance visualization: enabling extreme-scale scientific insight. Chapman and Hall, CRC Press