

Künstliche Intelligenz & Big Data in der Klimaforschung

– Versuch einer Einschätzung

Caveat

Ich habe keine klare Vorstellung weder was Künstliche Intelligenz noch was Big Data bedeutet. Ich assoziiere damit numerische Algorithmen, die sehr große Datenmengen verarbeiten, u.a., um statistische Zusammenhänge zu finden, etwa für Vorhersagen und zur Ableitung effizienter Approximationen. Dabei ist der Begriff „groß“ zeitabhängig – was vor 30 Jahren viele Daten waren, sind heute wenige – ganz analog zur Entwicklung der als „Supercomputer“ bezeichneten Rechner.

Meine eigenen Erfahrungen der letzten vier Jahrzehnte

Vorab – wenn ich hier von „Klimaforschung“ spreche, dann meine ich die Forschung über das Klima als physikalischem Gegenstand, wie ich sie seit 4 Jahrzehnten begleite. In diesen Tagen jährt sich der 40te Jahrestag meiner Promotion. Mein Vortrag hier spiegelt meine Erfahrungen und meine Nicht-Erfahrung wider: Numerische Modellierung, statistische Analyse, aber keine Fernerkundung, keine Prozeßstudien, keine Ableitung von Parametrisierungen.

In meinen mehr als vier Jahrzehnten ist die Datenmenge erheblich angewachsen – in meinen *frühen Tagen* simulierten wir das Wetter über 5, später 30 dann 330 Tagen auf der Nordhalbkugel. Drei vertikale Schichten, 2.8° Gitter Auflösung. Es ging um numerische Stabilität, um die Qualität der Simulation des Januar-Zustandes – die 330 Tage waren permanenter Januar - aber auch um Blocking. Der Modellbau lag bei Erich Roeckner, die Blocking-Fragen bei Günter Fischer. An Daten hatten wir insbesondere Bodenanalysen des Deutschen Wetterdienstes, die händisch hergestellt wurden. Aus den USA gab es die Oort & Rasmusson zonalen Klimastatistiken über 5 Jahre. Dazu dann lokale Datenreihen, etwa von Wetterschiffen im Nordatlantik.

Später dann, in den 1980er Jahren, am Max-Planck-Institut für Meteorologie wurde das Wettervorhersagemodell des EZMW übernommen und mit den Parametrisierungen des Nordhemisphärenmodell versehen. Wieder unter der Führung von Erich Roeckner. Jetzt wurden Jahregänge simuliert, schon viele. Zur Verifizierung nutzten wir etwa die Sammlung der digitalisierten Wetterkarten der Bodendruckverhältnisse des Britischen Wetterdienstes.

Dann ging es Schlag auf Schlag; das Atmosphärenmodell wurde mit Darstellungen des Ozeans verbunden, zunächst nur ein stationärer „Sumpf“, der Wärme aufnehmen und abgeben konnte, dann Deckschickmodelle, die einen vertikalen Austausch ermöglichten und schließlich 3-d dynamische Ozeanmodelle. Dabei nahm die Datenmengen stetig zu; neben

dem Versuch, El Nino vorherzusagen, begannen die Berechnungen von möglichen Änderungen aufgrund der erhöhten Treibhausgaskonzentrationen. Zunächst sogenannte Zeitscheibenexperimente, also 30 Jahre mit konstant erhöhten Treibhausgaskonzentrationen, dann einzelne „transiente“ Simulationen mit stetig erhöhten Konzentrationen über hundert Jahre; dann Verkleinerung der Gittergrößen, dann Ensembles von solchen Simulationen. Der zentrale Wissenschaftler war Ulrich Cubasch. Dazu kamen dann sehr lange Kontrollsimulationen, zunächst typischerweise 1000 Jahre; inzwischen auch Simulationen über tausende von Jahren mit vorgegebenen Antrieben wie stratosphärische Präsenz vulkanischen Materials und Variationen der Sonnenleistung. Hier war Eduardo Zorita ein Pionier.

Abbildung A

Aktuelle aufwändige Simulationen mit sehr hohem Speicherplatzbedarf am DKRZ waren eine historische Simulation am DKRZ über 2000 Jahre mit dem Modell MPI-ESM-P, das 60 Terrabyte in Beschlag legte; eine zweite Simulation über 7000 Jahre mit deutlich größerer Gitterauslösung beanspruchte nur 7 Tb.

Das DKRZ

Das Deutsche Klimarechenzentrum wurde 1988 gegründet als ein Instrument ausschließlich der Klimaforschung. Der Betrieb wird durch vier Gesellschafter finanziert, nämlich das Max-Planck-Institut für Meteorologie und die Universität Hamburg (jeweils ca. 40%) sowie das Helmholtz Zentrum Geesthacht (HZG) und das Alfred-Wegener Institut für Polar- und Meeresforschung (AWI (jeweils ca. 10%). Die regelmäßigen Investitionen in neue Rechnerarchitektur wurde vom zuständigen Bundesforschungsministerium (mit seinen wechselnden Namen) getragen. Die Hälfte der Rechenzeit steht den Gesellschaftern für Klimaforschungsprojekte zur Verfügung, die andere Zeit kann über entsprechende Anträge aus dem 50%igen Ministeriumsteil abgeholt werden.

Abbildung B

Das DKRZ unterscheidet sich von anderen Rechenzentren, dass es spezifisch für Klimaforschungsaufgaben reserviert ist – dazu werden dort Standardmodelle und „Konsortialrechnungen“ (also umfangreiche Rechnungen, die auf Wunsch der Community durchgeführt wurden) vorgehalten. Hardwareseitig besonders ist am DKRZ das Vorhalten sehr großer Datenspeicherung – beides „compute“ and „store“ sind die Aufgaben des DKRZ.

Abbildung C

Seit 1988 gab es bisher 4-mal einen Rechnerwechsel, der natürlich mit erhöhten Rechenleistungen und weiter erhöhten Speichermöglichkeiten verbunden war.

Gegenstand Klima

Abbildung D

Klima, das ist zu allererst die Statistik von Phänomenen im Erdsystem, die mit dem Wetter in Atmosphäre, Ozean und Kryosphäre verbunden sind. Klimaforschung hebt vor allem ab auf die Interaktion der Komponenten im Klimasystem, zu dem auch neben anderen der

Kohlstoffkreislauf gehört, auf die Abhängigkeit der Statistik der Phänomene von Antriebsfaktoren (anthropogene und natürliche) und der internen Dynamik, die regionale Ausprägung sowie die zeitliche Variabilität (incl. Vorhersagbarkeit).

Eine sinnvolle Klimaforschung ist nur möglich, wenn sie auf gewisse Voraussetzungen aufbauen kann, nämlich

1. Regelmäßige, homogene und belastbare Analysen des Zustandes des einschlägigen Erdsystems. Man spricht von **Datenanalyse**. Hier hat künstliche Intelligenz ihren festen und nicht mehr hinterfragten Platz gefunden.
2. Modellentwicklung – die dynamischen **Klimamodelle** (im Folgenden kurz: Klimamodelle) werden zusammengesetzt aus dynamischen Modellen der Komponenten, also insbesondere der Atmosphäre und des Ozeans. Modell. Da sie räumlich mit einer endlichen Auflösung operieren, können nicht alle Prozesse dargestellt werden. Einige dieser nicht berücksichtigten Prozesse werden parametrisiert (z.B. Konvektion in Ozean und Atmosphäre), d.h. die Nettowirkung dieser Prozesse wird abgeschätzt als konditionierter Erwartungswert der Wirkung gegeben die beschriebenen Zustandsgrößen. Bei der Entwicklung von **Parametrisierungen** zeichnet sich ab, dass KI an Bedeutung gewinnt. Auch die Qualitätssicherung sowohl der Komponenten der Modelle als auch der zu einem Klimamodell aggregierten Systeme wird effizienter gestaltet durch den Einsatz von Algorithmen.
3. Ein breites Prozeßverständnis.

Abbildung E

4. Methoden zur Mustererkennung, etwa zur automatischen Identifikation von Wirbeln in Modellsimulationen oder andere Merkmale. Hier werden geeignete automatisierte Algorithmen, also KI, entwickelt und eingesetzt.

Dies ist eine ad-hoc Liste, und sie ist unvollständig. Die Arbeiten zu diesem Themen stellen wissenschaftliche Herausforderungen dar, schaffen aber zunächst kein „Verstehen“ der Dynamik des Klimasystems, seiner Abhängigkeiten von externen Antrieben, und seiner Wirkung auf Gesellschaft und Ökosysteme.

Wissenschaftliche Kernfragen der Klimaforschung

Folgt man der öffentlichen Diskussion in diesen Tagen mit dem Dominanz des „Klimakrisen“-Themas, die Klimawissenschaftlern eine wichtige wenn nicht bestimmende Funktion einer zukünftigen Klimapolitik zuweist, so könnte man meinen, dass es eigentlich nur darum ginge, abzuschätzen, wie sich das Klima bei welchen Konzentrationen von Treibhausgasen entwickeln würde, und welche Folgen für Ökosysteme und Gesellschaft damit einher gehen.

Abbildung F

So verwundert es nicht, dass junge Wissenschaftler, Klimaforscher wie Nicht-Klimaforscher gleichermaßen, auf die Frage nach der Hauptaufgabe der Klimaforschung auf „Motivierung der Öffentlichkeit“ verweisen, und nicht auf die Klärung der Zusammenhänge.

Die Fragen nach Szenarien und Klimawirkung gehören zum Kanon der wissenschaftlichen Fragestellungen, und insbesondere vom UNO-Klimarat IPCC sind erfolgreiche Versuche unternommen worden, die einschlägigen Resultate zu sichten und in eine belastbare Form für Gesellschaft und Politik zu bringen.

Geophysikalisch gesehen, ist **das Besondere an der Klimaforschung** die Tatsache, dass das betrachtete System sowohl extrem hochdimensional aber dabei dynamisch skalenabhängig ist, und ubiquitär Nichtlinearitäten wirken. Als Folge greifen die schönen Konzepte niederdimensionaler Dynamik nichtlinearer Systeme nur in Ausnahmefällen idealisierter Set-Ups. Eine zweite wesentliche Folgerung ist, dass das mathematische Konstrukt der stochastischen Variabilität ein wirksames Werkzeug zur Beschreibung der Phänomene, ihrer Variabilität und ihrer Abhängigkeit von „Antrieben“ ist.

Anders als in anderen Gebieten gibt es keine Möglichkeit des Experimentierens und keine Möglichkeit beliebig weitere Stichproben zu gewinnen (außer durch zeitliches Zuwarten); diese Möglichkeiten können aber durch die Nutzung von quasirealistischen Modellen geschaffen werden, in der Hoffnung, dass diese Realität nachempfinden. Die „Beobachtungsdaten“, d.h. im Lauf der Zeit im Feld erhobenen Daten sind zeitlich variabel was Datendichte und -qualität angeht. Man spricht von „Inhomogenitäten“, die beim Einsatz automatisierter Systeme wesentlich werden können. Auch das Aufspüren und Beseitigen solcher Inhomogenitäten kann Gegenstand künstlicher Intelligenz sein. Allerdings gibt es immer wieder Beispiele, dass dieser Vorbehalt der Inhomogenität nicht beachtet wird, und merkwürdig-interessante Resultate veröffentlicht werden. Sachbezogene menschliche Intelligenz ist wesentlich, um mit diesem Problem der scheinbaren Links, Trends, Oszillationen etc. aufgrund veränderter Beobachtungs- und Analysepraxis umgehen zu können.

Mögliche Anwendungen von „big data“ bzw. „Künstliche Intelligenz“

Im Folgenden möchte ich drei Fälle diskutieren, in denen ich Einsatzmöglichkeiten für „künstliche Intelligenz“ bzw. „Big Data“ sehe. Es gibt weitere Fälle, aber diese scheinen mir die erfolgversprechendsten zu sein. Meine Einschätzung ist subjektiv und vermutlich auch meiner mangelnden Übersicht geschuldet. Ein erfolgreiches Feld ist sicher die Invertierung von Satellitendaten, für das ich keine Erfahrung habe.

Ich möchte aber betonen, dass in keinem Fall „Verstehen“ das Resultat ist, sondern nur eine bessere Beschreibung bzw. Charakterisierung ist. Letzteres ist eine Voraussetzung für „Verstehen“ – und der Zweck von Wissenschaft ist eben: „Verstehen“

a) Untersuchung des Phasenraumes

Wie erwähnt hat das Klimasystem praktisch unendlich viele Freiheitsgrade, die keinen einfachen Skalierungseigenschaften genügen. Tatsächlich gibt es nicht mal eindeutige

Differentialgleichungen, sondern nur Differenzgleichungen, die sich mit fortschreitender Gitterauflösung ändern, sowohl in der Formulierung als auch in den Zustandsgrößen. Bei einer horizontalen Gitterauflösung von 200 km beschreibt man keine Konvektion, weil die atmosphärischen Konvektionssysteme kleiner als 200 km sind; in dem Falle arbeitet man mit der hydrostatischen Approximation für die Vertikalbewegungen. Geht man zu 1 km horizontaler Gitterauflösung, dann verwendet man die Bewegungsgleichungen auch in der Vertikalen, und beschreibt die Konvektion explizit. Ebenso werden Wolken bei groben horizontalen Auflösungen parametrisiert; wenn man diese besser beschreiben will, braucht man nicht nur die Feuchte in der Gitterzelle, sondern auch das Spektrum der Wolkentropfen in der Zelle usw.

In so einem großen Zustandsraum belastbare Zusammenhänge zu finden, ist schwierig. Daher hat Klaus Hasselmann das Konzept der „Principal Interaction Patterns“ konstruiert; dabei handelt es sich um niederdimensionale Unterräume, innerhalb dessen die Interaktionen in erster Näherung abgeschlossen ablaufen, und der Rest des Phasenraumes nur stochastisch, ggfs. konditioniert auf den Zustand des niederdimensionalen Unterraumes, wirken. In anderen Worten: der Unterraum ist dynamisch abgeschlossen, und das Geschehen außerhalb wird (stochastisch) parametrisiert. Die Koordinaten des Unterraumes sind dann die Principal Interaction Patterns. Diese sind durch ein geeignetes Datascreening empirisch zu bestimmen.

So schön dies Konzept in seiner Allgemeinheit ist, so gibt es doch keine wirklich guten Beispiele, die das volle Potential darstellen; stattdessen werden Ansätze gemacht, wie diese Unterräume aussehen können, und welche Dynamik dort herrschen soll. Die „Principal Oscillation Patterns“ sind die lineare Variante, die in der Tat mit empirisch bestimmten Räumen operieren: hier gibt es eine Reihe von Beispielen, etwa jenes von Mikolajewicz und Maier-Reimer zur niederfrequenten stochastischen Oszillationen in einem dynamischen Ozeanmodell im Südliche Ozean als Reaktion auf Anregung durch weißes Rauschen.

Abbildung G

Die Verfügbarkeit sehr umfangreicher (simulierten) realitätsnaher Datenmengen eröffnet neue Möglichkeiten, derartige Unterräume zu bestimmen.

In einem anderen Falle wurde versucht, die atmosphärische Zirkulation über dem Nordatlantik als so einen Unterraum zu verstehen, und diesen mit einer bimodalen Verteilung zu charakterisieren, also einen Zustand mit dominanter Westwinddrift und einen zweiten mit Blocking. Die Veröffentlichung im Jahr 1986 erzeugte viel Aufmerksamkeit, weil sie andeutete, dass der Traum der Anwendbarkeit niederdimensionaler nichtlinearen Analytik auch in der Analyse der Atmosphäre erfüllt würde.

Es stellte sich dann aber heraus, dass die Analyse eine Methodik nutzte, die auch bei Unimodalität oft Bimodalität anzeigt – und dass für eine deutliche Identifikation wesentlich mehr Daten erforderlich wären. Seitdem scheint das Thema zu schlafen.

Auch hier könnte „Big Data“ einen neuen Versuch anraten. Allerdings ist das Interesse nach der fast vollständigen Fixierung auf Fragen des Klimawandels gering, wenn nicht erloschen.

Andere Versuche heben ab auf die Atlantic Meridional Overturning (AMOC) etwa nach dem sogenannten Stommel-Modell. All diese interessanten niederdimensionalen Strukturen

treten in Modellsimulationen auf, oft in sehr vereinfachten Systemen, im Falle der AMOC etwa, wenn die Atmosphäre nur als Gleichgewichtspartner ohne Wetter mit dem Ozean interagiert. Es sind dies alles schöne Beispiele, deren Nachweis des Realismus aber noch aussteht. Heutzutage, nachdem die Rechner- und Speicherkapazität weiter deutlich angestiegen ist, können Simulationen über tausende oder gar Zehntausende von Jahren mit quasirealistischen Klimamodellen, die realitätsnah die diversen Prozesse einschließlich des Wetter berücksichtigen, durchgeführt werden. Diese bieten sich an, diese Vorschläge erneut auf Konsistenz mit den stochastischen Dimension des Klimasystems zu testen.

b) Zustandsraummodelle

Damals, in den 1970er Jahren, als ich Mathematik studierte, gab es Differentialgleichungen, die über Rand- und Anfangsbedingungen fixiert werden, und in numerischen Modellen integriert wurden. Inzwischen ist die Formulierung der Zustandsraum-Modelle hinzugetreten. Vermutlich gibt es verschiedene Begriffe für den gleichen Ansatz in verschiedenen Disziplinen. Vielleicht hilft es, hier auf Kalman-Filter zu verweisen, obwohl dies ein Spezialfall ist.

Abbildung 1

So ein Zustandsraummodell hat zwei Gleichungssysteme. Ein dynamisches Modell, das die Zustandsgrößen in der Zeit vorwärts rechnet, und dabei Fehler macht; und ein Beobachtungsmodell, das die vorhergesagten Zustandsgrößen in Relation zu „beobachteten“ und sonstwie irgendwie bekannten Größen stellt. Dies kann etwa die Temperaturverteilung längs der Flugbahn eines Flugzeuges sein, während die Zustandsgröße die Temperatur in Gitterboxen in einer gewissen Höhe sein. Auch die Beobachtungsmodell ist nicht fehlerfrei.

Dann werden diese beiden Modelle alternierend eingesetzt. Zunächst schätzt man den Zustand zum Zeitpunkt $t+1$ mit der dynamischen Gleichung, dann vergleicht man diese Vorhersage mit den „Beobachtungen“ wo denn möglich, und korrigiert die Vorhersage mit Termen, die die Differenz von Vorhersage und Beobachtung berücksichtigen. Durch diese Manipulation wird erzwungen, dass die abgeleitete Trajektorie der Zustandsgrößen zumindest konsistent bzw. „in der Nähe“ der Beobachtungen verläuft.

Diese algorithmischen Verfahren werden mit großem Erfolg eingesetzt zur Bestimmung des atmosphärischen Zustandes alle paar Stunden am Tag, die dann als Anfangszustände für die Wettervorhersage verwendet werden. Sie werden auch genutzt, um die Entwicklung über lange Zeit in konsistenter Weise näherungsweise zu beschreiben (Re-Analysen), bzw. um großskaligen Strukturen konsistente kleinskalige Details hinzuzufügen (Downscaling).

Für mich ist dies der überzeugendste, weil bis dato erfolgreichste Fall von Künstlicher Intelligenz und die Nutzung von Big Data in meinem Erfahrungsbereich.

c) Detektion und Attribution

Eine methodische Kernaufgabe in der Klimaforschung ist die Unterscheidung von Signal und intern erzeugten Variabilität, gerne auch als Noise oder Rauschen bezeichnet. Als Pionier ist hier Klaus Hasselmann zu nennen. Erst wenn sich Strukturen im hochdimensionalen Raum bilden, die nicht im Rahmen der internen Variabilität des Systems zu erwarten sind, schließt man auf die Gegenwart eines Signals. Dies nennt man Detektion –

die Feststellung der Gegenwart eines extern verursachten Signals. Dann stellt sich die Frage, welche externe Ursache denn als Erklärung greift. Dies nennt man Attribution.

Ursprünglich wurde dies Konzept für die Fernwirkung des pazifischen „El Nino / Southern Oscillation“-Phänomens entwickelt; heutzutage ist es ein Standardansatz bei dem Nachweis der Wirkung erhöhter Treibhausgaskonzentrationen, ob beobachtete Trends eher als Teil des Spektrum der natürlichen Schwankungen angesehen werden sollte (wie etwa die Sturmtätigkeit in Nordeuropa) oder als systematische Veränderung (wie die Temperaturerhöhung in fast allen Teilen der Welt).

Ein Variante ist die sogenannte „Event Attribution“, in der gefragt wird, ob ein einzelnes Ereignis, etwa ein Sturm oder eine Hitzewelle, Ausmaße angenommen hat, wie es unter ungestörten klimatischen Verhältnissen extrem selten und damit auszuschließen sei, oder ob hier ein Einfluss des menschengemachten Klimawandels zu konstatieren ist. Derartige Analysensysteme werden zu diesem Zweck in automatisierte Algorithmen überführt, so dass künftig bei allen auffälligen Ereignissen zeitnah eine solche Einschätzung durchgeführt werden kann.

Vorbehalt

In der Geschichte der Erdsystemwissenschaft wurde schon lange nach wirksamen Methoden gesucht, die das System radikal vereinfachen, und so Vorhersagen einfach und automatisch ermöglichen. Der prominenteste Fall war die Erkenntnis, dass sich jede Reihe von Zahlen in Fourierkomponenten zerlegen lässt. Da die zugehörigen trigonometrischen Funktionen sich fortschreiben lassen in die Zukunft, erwartete man ungeahnte neue Vorhersagemöglichkeiten; es ging nur noch um die Identifikation der dominanten Periodizitäten. Es gab zahllose Anwendungen aller Art, etwa im „*Manual of Meteorology*“ von 1936, das auf viereinhalb Seiten derartige Periodizitäten auflistete. Alles Quatsch, außer dem Jahres- und Tagesgang.

1927 zeigte der russische Mathematiker E. Slutsky, dass auch weißes Rauschen sich als Fourierreihe darstellen lässt, und dass die Amplituden sich stark verändern, wenn die Reihe nur um einen Wert verlängert wird. Erst dann wurde der stochastische Charakter der Fourieranalyse verstanden und damit diese Praxis stark eingeschränkt. Was nicht bedeutet, dass schlichte Gemüter nicht immer noch auf die angeblich gute Perspektive abfahren.

Etwas Ähnliches geschah in den 1980er und 1990er Jahren, als neuronale Netze in Mode kamen, vor allem für Vorhersagezwecke etwa des „El Nino / Southern Oscillation“-Phänomens. Auch sie scheiterten, weil weder ein physikalisches Underpinning der Mechanismen noch eine Trennung von Signal zu Rauschen gelang.

Ein weiterer Vorbehalt ist, dass solche Systeme nur dann sinnvoll wirken können, wenn die Belastbarkeit der Daten, also deren Homogenität nicht kompromittiert ist durch Verunreinigung wegen veränderlicher Beobachtungs- und Analysepraktiken.

Abbildung H

Fazit

Künstliche Intelligenz ist im Routineeinsatz zur raum-zeitlich vollständigen und konsistente Bestimmung des Zustandsraums (Datenassimilation).

Sie eignet sich zur Analyse von dynamisch relevanten Teilräumen des Phasenraums, und der Trennung von Signal und Rauschen, und anderen Anwendungen.

Künstliche Intelligenz liefert aber **kein Verstehen der Dynamik**.

Beim Einsatz von automatischen Systemen sollte der vorhandene Datensatz immer so groß sein, dass er in mehrere Stücke aufgeteilt werden kann, wovon eines mit automatischen Methoden ausgewertet wird, und die anderen dazu dienen festzustellen, ob die gefundenen Resultate **stichprobenunabhängig** sind.

Künstliche Intelligenz & Big Data in der Klimaforschung – Versuch einer Einschätzung

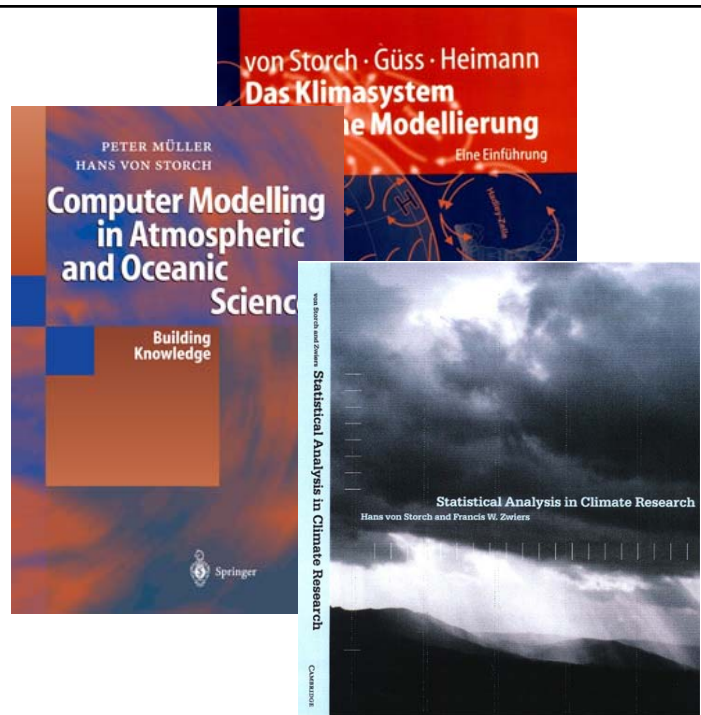
Hans von Storch
Institut für Küstenforschung, HZG, Geesthacht

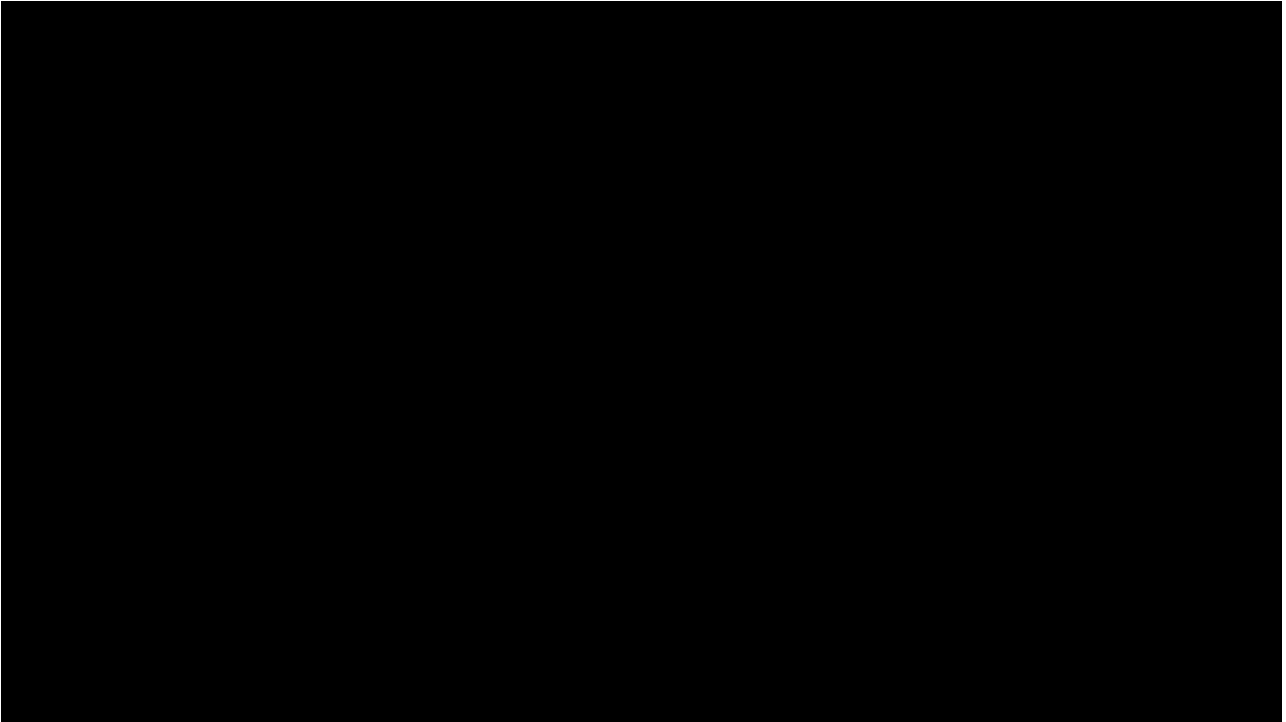


Digitale Arbeitswelten in Forschung und Entwicklung. Neue Möglichkeiten und Herausforderungen für die Wissenschaft, Europäische Akademie, Ahrweiler-Bad Neuenahr, 2. Juli 2019

Hans von Storch

- **Klimaforscher**
- **Spezialgebiet:** Klimastatistik, Klimamodellierung, Küstenklima, also Windstürme, Nord- und Ostsee, Nordatlantik
- **Emirierter Professor** am Meteorologischen Institut der Universität Hamburg
- Kooperation auch mit Sozialwissenschaftlern; Mitglied der WISO Fakultät der Universität Hamburg
- **Emeritierter Direktor** des Instituts für Küstenforschung des Helmholtz-Zentrums Geesthacht
- **Gastprofessor** an der Ocean University of China, Qingdao





Aufwändige Einzelrechnungen am DKRZ in der letzten Zeit:

Die Datenmässig umfangreichste Simulation ist eine historische Simulation mit geschätzten natürlichen und anthropogenen Antrieben über 2000 Jahre mit dem Modell MPI-ESM-P

atmosphärische Gitterauflösung T63: ca $1.8^\circ \times 1.8^\circ$ lat x lon (96x192), 40 Schichten

ozeanische Auflösung durchschnittlich $1^\circ \times 1^\circ$ lat x lon (180x360), 40 Schichten

Datenvolumen ist ca 30 Gb /Jahr = **60Tb** insgesamt

Die Datenmässig zweitumfangreichste war eine andere, längere historische Simulation OETZL Simulationen mit ECHO-G über 7000 Jahre.

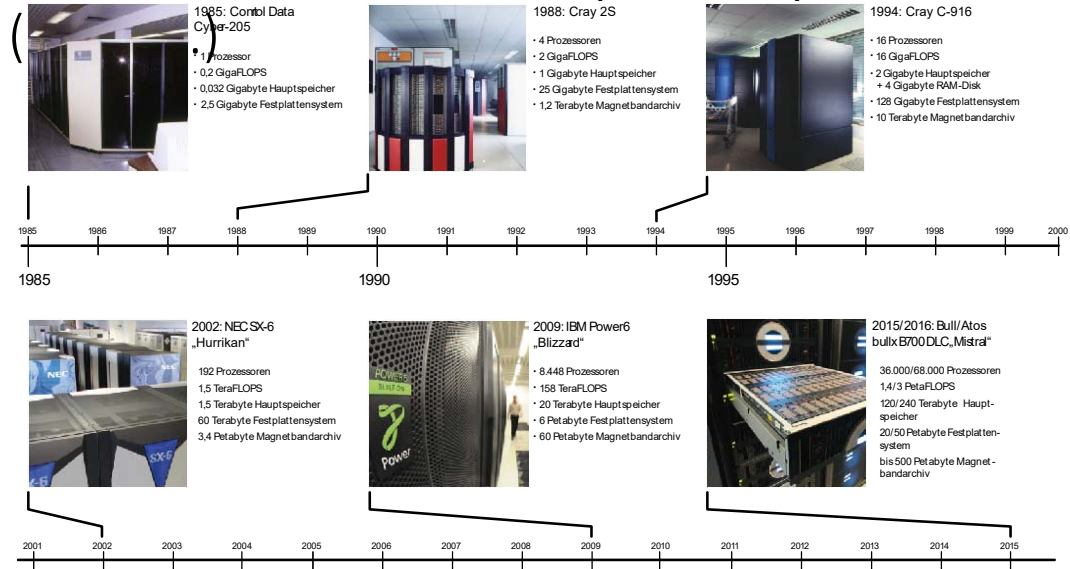
atmosphärische Gitterauflösung T30 , ca $3.75^\circ \times 3.75^\circ$ lat x lon, 19 Schichten

ozeanische Gitterauflösung, ca $2.8^\circ \times 2.8^\circ$, 20 Schichten

Datenvolumen ca 1 Gb/Jahr = **7Tb** insgesamt.

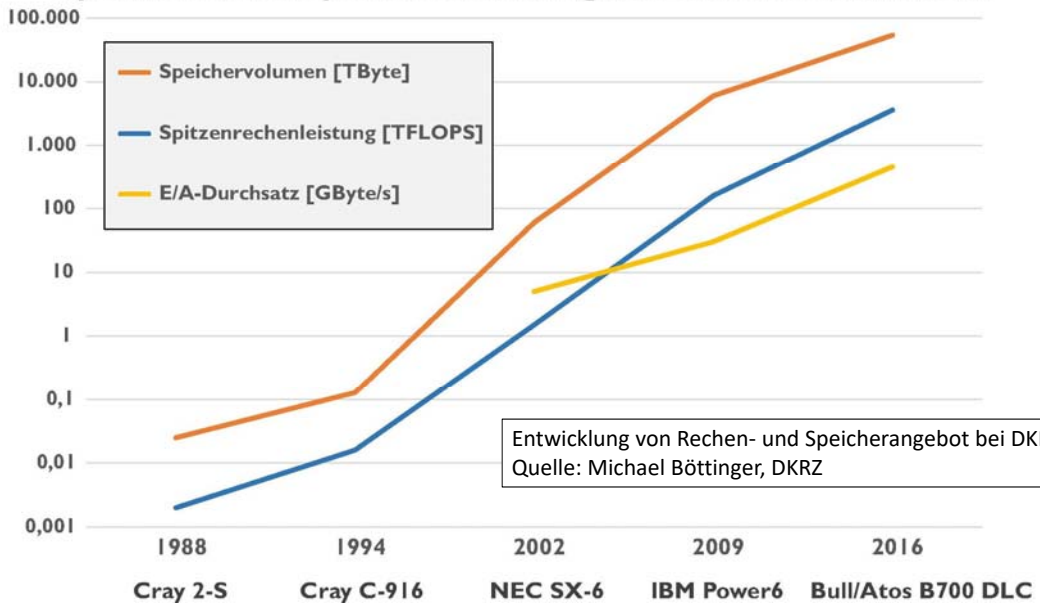
Auskunft: Eduardo Zorita, HZG

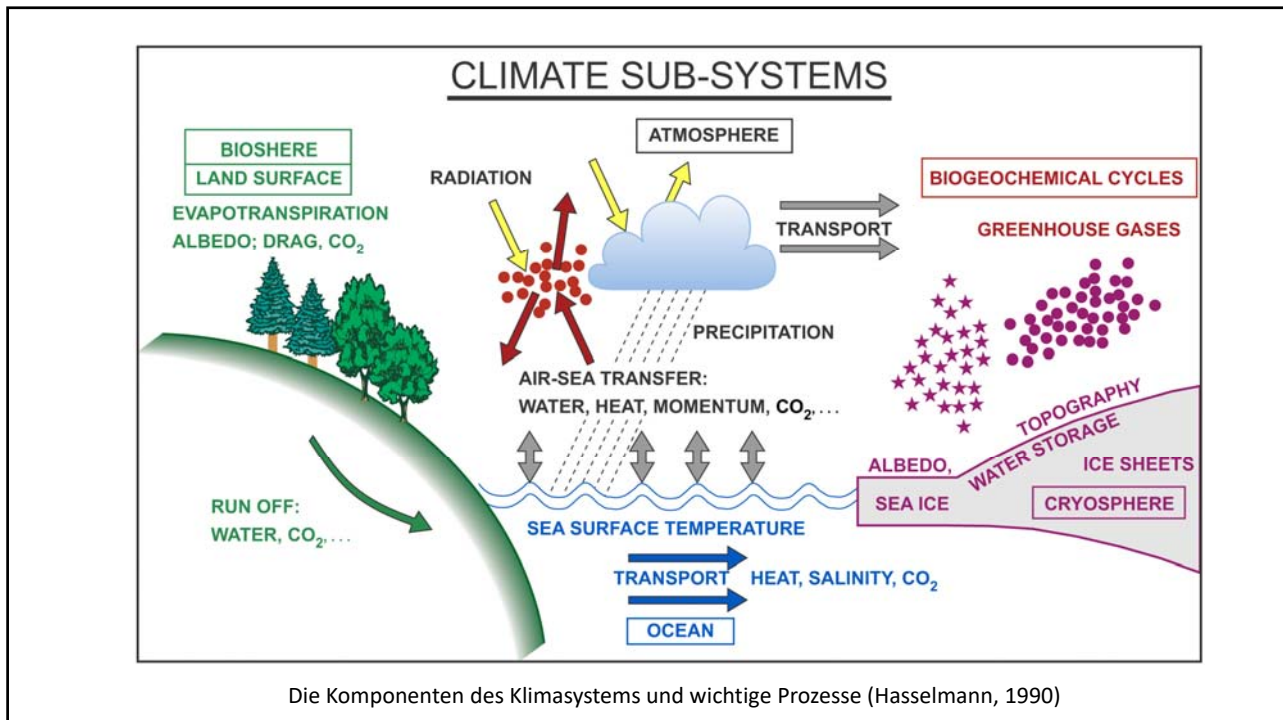
25+ Jahre Rechner- und Speichersysteme



5

Speichervolumen, Spitzenrechenleistung und E/A-Durchsatz am DKRZ



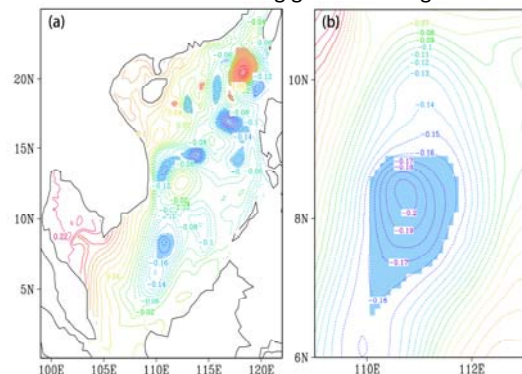


Mustererkennung: Wirbel im Südchinesischen Meer

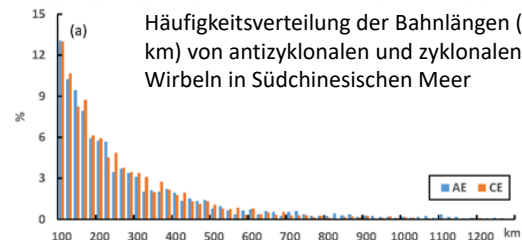
- In einer Simulation über 60 Jahre mit einem von realistischen atmosphärischen Bedingungen angetriebenen Ozeanmodell auf einem Gitter von ca. 10 km.
- 27,371 aktive Gitterpunkt
- Tägliche Daten.
- Gesucht: Statistik der jährlichen Häufigkeit, Intensität und Größe von wandernden Wirbeln im Südchinesischen Meer.
- Algorithmus, der anhand des täglichen Meeresspiegels im Mittel pro Jahr 28 antizyklonische Wirbel und 54 zyklonische Wirbel detektiert und charakterisiert.

Zhang M., et al., 2019: Statistics of travelling eddy variability in the South China Sea, 1950-2010. *Ocean Dyn.*, in press

Simulierter Meeresspiegel im Südchinesischen Meer an einem beliebig gewählten Tag.



Häufigkeitsverteilung der Bahnlängen (in km) von antizyklonalen und zyklonalen Wirbeln im Südchinesischen Meer



Umfrage unter jungen Wissenschaftlern und Studenten: Was ist die Hauptaufgabe der Klimaforschung heute:

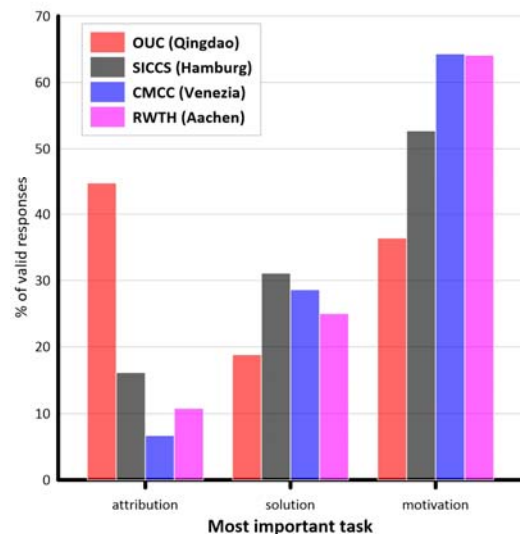
- **OUC** - Ocean University of China, College of ...
Response rate: 87 out of about 200, ca. 43%
May 1 to December 31, 2015, and September 1 to 30, 2016
- **SICCS** - PhD/Master School of Cluster of Excellence "Integrated Climate System Analysis and Prediction" (CiSAP) at the University of Hamburg
Response rate: 62 out of 100, 62%
June 1 to June 26, 2017
- **CMCC** - Centro Euro-Mediterraneo sui Cambiamenti Climatici, Venezia & Bologna
Response rate: 30 out of 200, 15%
October 2018
- **RWTH** - Institut für Wasserbau, RWTH Aachen,
January 2019

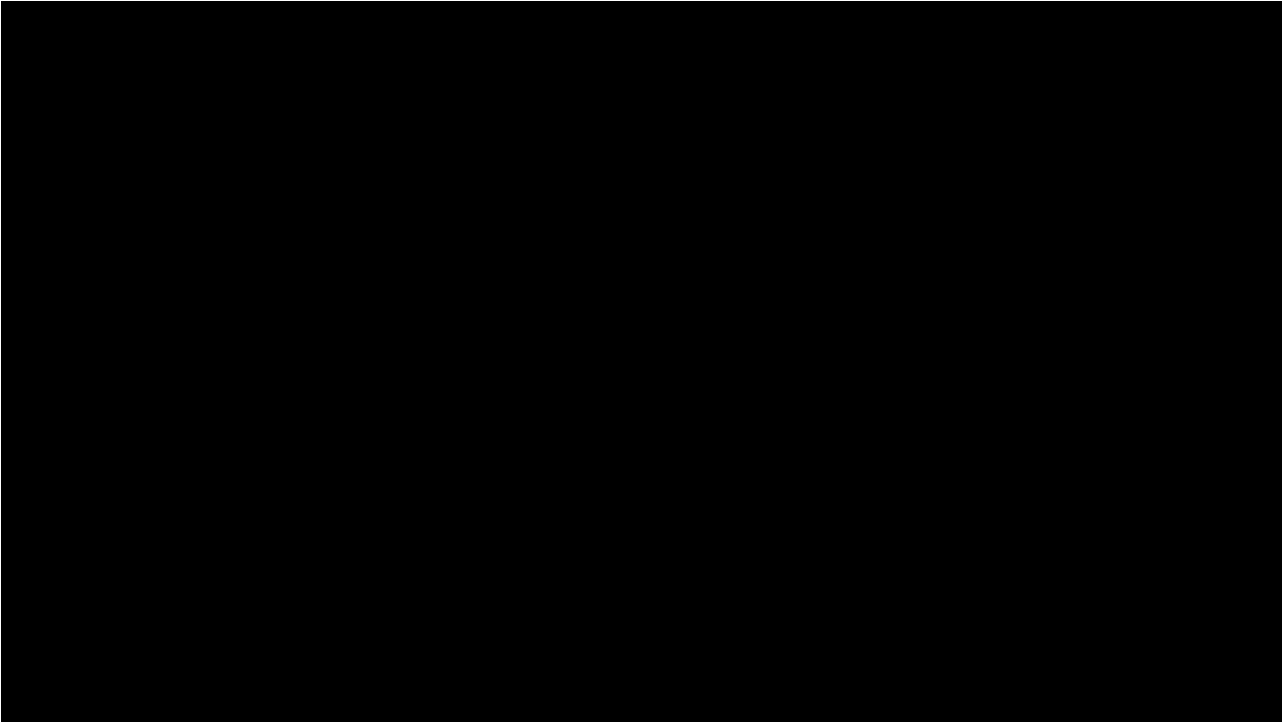
Surveys

von Storch, H., Chen X, B.Pfau-Effinger, D. Bray and A. Ullmann, 2019: Attitudes of young scholars in Qingdao and Hamburg about climate change and climate policy – the role of culture for the explanation of differences. *Advances in Climate Change Research*, online: 10.1016/j.accre.2019.04.001

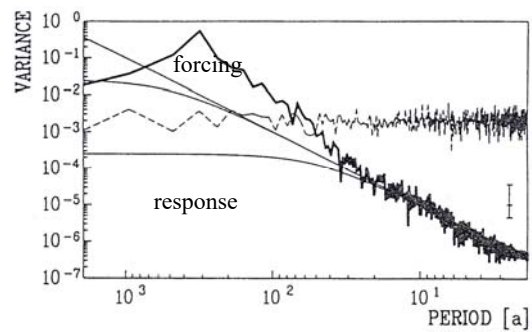
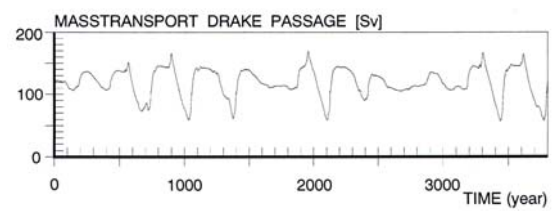
Today, what would you rate as the most important task facing the climate science community?

- define the climate problems and attribute cause of climate change
- determine solutions to climate change
- motivate people to act on climate change

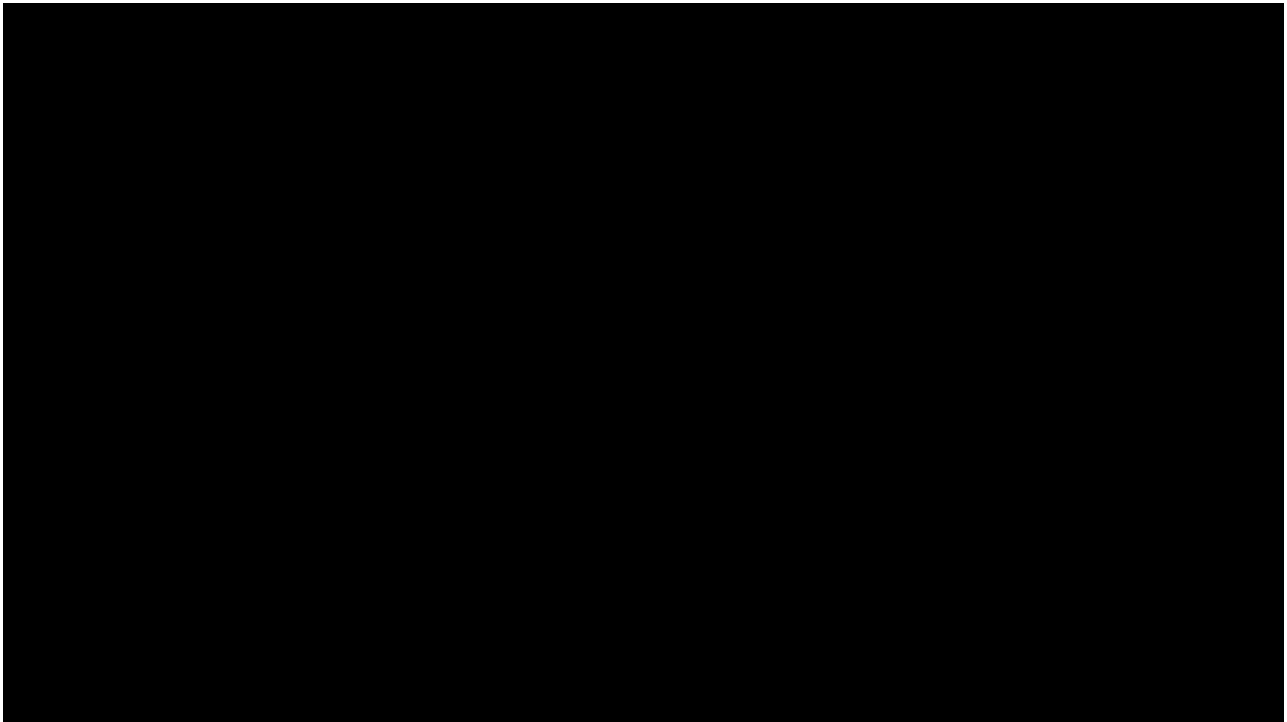




Numerical experiment with ocean model: standard simulation with steady forcing (wind, heat and fresh water fluxes) plus random zero-mean precipitation overlaid.



Mikolajewicz, U. and E. Maier-Reimer, 1990



State space equation $\Psi_{t+1} = F(\Psi_t, \eta_t) + \varepsilon_t$

Observation equation $d_t = G(\Psi_t) + \delta_t$

with ε_t, δ_t = model and observation errors

F = dynamical model

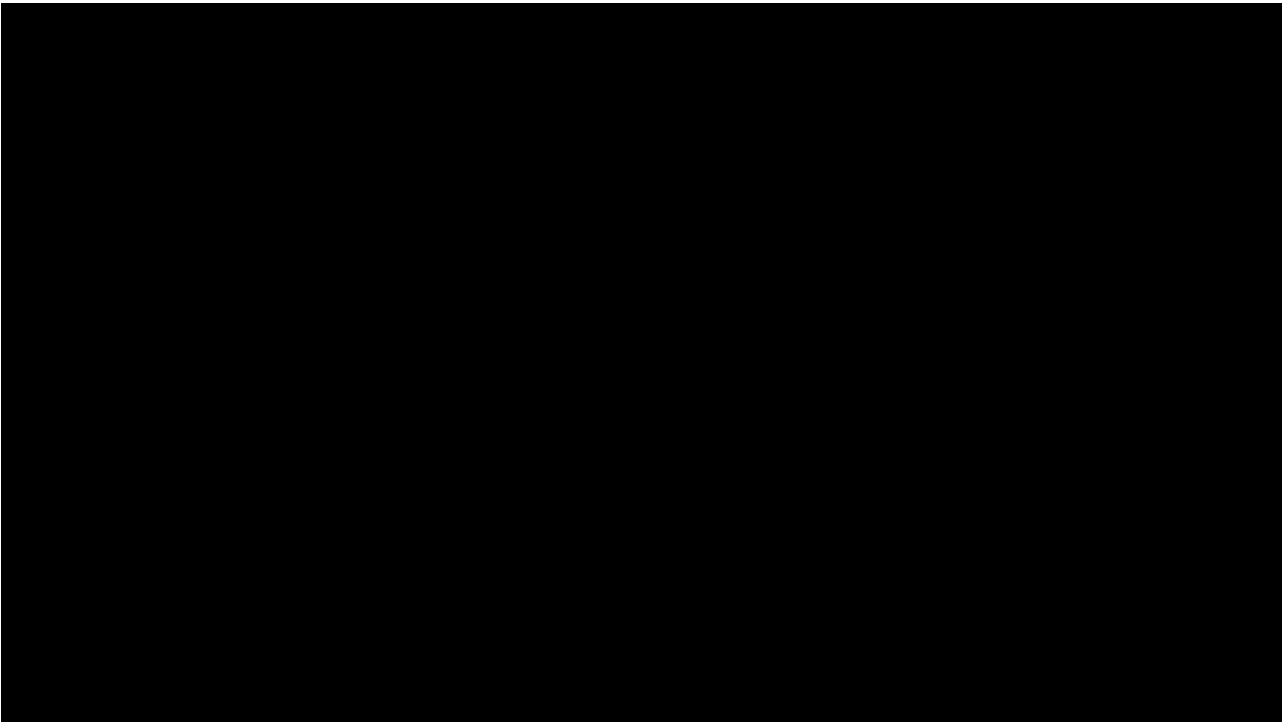
G = observation model

Forward integration : $\Psi_{t+1}^* = F(\Psi_t, \eta_t)$

$d_{t+1}^* = G(\Psi_{t+1}^*)$

$\Rightarrow \Psi_{t+1} = \Psi_{t+1}^* + K(d_{t+1}^* - d_{t+1})$

with a suitable operator K .



Fazit

- Künstliche Intelligenz ist im Routineeinsatz zur raum-zeitlich vollständigen und konsistente Bestimmung des Zustandsraums (Datenassimilation).
- KI eignet sich zur Analyse von dynamisch relevanten Teilräumen des Phasenraums, und der Trennung von Signal und Rauschen, und anderen Anwendungen.
- KI liefert aber **kein Verstehen der Dynamik**.
- Beim Einsatz von automatischen Systemen sollte der vorhandene Datensatz immer so groß sein, dass er in mehrere Stücke aufgeteilt werden kann, wovon eines mit automatischen Methoden ausgewertet wird, und die anderen dazu dienen festzustellen, ob die gefundenen Resultate **stichprobenunabhängig** sind.