

Sind Klimamodelle „einfach Physik“?

geschrieben von Chris Frey | 4. Juni 2026

Häufig wird behauptet, man solle Klimamodellen Glauben schenken, weil sie lediglich auf physikalischen Gesetzen beruhen. In einem Artikel von mir von vor einigen Jahren habe ich dargelegt, dass Wissenschaft so nicht funktioniert. Selbst gültige wissenschaftliche Theorien liefern unter Umständen keine präzisen Vorhersagen, und zwar aus verschiedenen Gründen wie beispielsweise Heterogenität (z. B. bei Erdbeben). In diesem Beitrag fasse ich einige der wichtigsten Ergebnisse zusammen, etwa ein Drittel des gesamten Artikels. Dieser steht hier:

Loehle, C. 2018. Epistemological Status of General Circulation Models. *Climate Dynamics* 50:1719-1731. DOI 10.1007/s00382-017-3737-7.

Der erkenntnistheoretische Status allgemeiner Zirkulationsmodelle (General Circulation Models; GCMs)

Craig Loehle, Ph.D., National Council for Air and Stream Improvement, Inc. (NCASI)

Abstract: Prognosen sowohl zu den wahrscheinlichen anthropogenen Auswirkungen auf das Klima als auch zu den daraus resultierenden Auswirkungen auf Natur und Gesellschaft basieren auf umfangreichen, komplexen Softwaretools, den sogenannten Globalen Zirkulationsmodellen (GCMs). Von GCMs erstellte Prognosen wurden bei der politischen Entscheidungsfindung im Zusammenhang mit dem Klimawandel intensiv genutzt. Der Zusammenhang zwischen den zugrunde liegenden physikalischen Theorien und den von GCMs erzeugten Ergebnissen ist jedoch unklar. Im Falle von GCMs werden zahlreiche Diskretisierungen und Näherungen vorgenommen, und die Simulation von Prozessen des Erdsystems ist alles andere als einfach und führt derzeit zu einigen Ergebnissen mit unbekanntem Ausmaß der Auswirkungen auf die Energiebilanz. Statistische Tests der GCM-Prognosen hinsichtlich des Übereinstimmungsgrades mit den Daten würden die Beurteilung der Gebrauchstauglichkeit erleichtern. Wenn Modellergebnisse aufgrund von Modellverzerrungen auf Anomaliebasis gestellt werden müssen, hängen sowohl visuelle als auch quantitative Maße für die Modellanpassung stark vom für die Normalisierung verwendeten Referenzzeitraum ab, was die Prüfung problematisch macht. Die Erkenntnistheorie wird hier auf Probleme der statistischen Inferenz während der Prüfung, die Beziehung zwischen der zugrunde liegenden Physik und den Modellen, die epistemische Bedeutung von Ensemble-Statistiken, Probleme der räumlichen und zeitlichen Skala, das Vorhandensein oder Nichtvorhandensein einer ungezwungenen Nullhypothese für Klimaschwankungen, die Bedeutung bestehender Unsicherheitsschätzungen und andere Fragen angewendet. Eine strenge Argumentation erfordert eine sorgfältige Quantifizierung des Unsicherheitsgrades.

1 Introduction

Globale Zirkulationsmodelle (GCMs) versuchen, den aktuellen Wissensstand zur Klimadynamik mittels Prozessgleichungen abzubilden und diese Gleichungen numerisch zu lösen, um das Klima unter verschiedenen Szenarien menschlicher Einflüsse zu simulieren (Taylor et al. 2012). Diese Modelle sind komplex und werden seit den 1960er Jahren weiterentwickelt (Manabe und Wetherald 1967). Den Ergebnissen der GCMs kommt bei der Formulierung der öffentlichen Energiepolitik eine zentrale Rolle zu. Die Grundlage für diese zentrale politische Stellungnahme ist, dass die Modelle auf physikalischen Grundlagen beruhen (IPCC 2013), wobei vielen Attributions- und Prognoseergebnissen ein hohes Maß an Zuverlässigkeit (>95 %) beigemessen wird (IPCC 2013 SPM). Der IPCC berichtet zudem, dass GCMs historische Daten gut abbilden und dass die Übereinstimmung ohne Einbeziehung von Treibhausgasen nicht gut ist (IPCC 2013, Abb. SPM.6).

Es gibt eine umfangreiche Literatur, welche die Ergebnisse von GCMs mit verschiedenen Klimamerkmale vergleicht (siehe folgende Abschnitte). Solche Vergleiche werden durch den stochastischen Charakter sowohl des Klimas als auch der Modelle erschwert. Vergleiche zwischen GCMs und Daten werden je nach Variable und Studie als mangelhaft, angemessen, gut oder ausgezeichnet bewertet (McWilliams 2007). Diese Unklarheit resultiert aus einer Vielzahl von Kriterien für die Modellgüte sowie aus unterschiedlichen Ergebnissen.

Die Bewertung von Wissensaussagen (von denen es mehrere gibt), die auf GCMs basieren, kann durch eine Betrachtung der Erkenntnistheorie (siehe Williams 2001 für einen Überblick) unterstützt werden, die den logischen Rahmen für die Bewertung dessen bildet, wie wir wissen und was man wissen kann. Mit einer epistemologischen Analyse können wir den Status einer Theorie/eines Modells hinsichtlich ihrer/seiner logischen Grundlage, Zuverlässigkeit und Stringenz bewerten. Mit diesem Rahmen können wir sowohl die Tests der Modellgüte als auch die Übereinstimmung der aus GCMs abgeleiteten Ergebnisse mit der bekannten Physik bewerten. Ich werde diese Fragen zunächst anhand verschiedener Bereiche der Wissenschaft veranschaulichen und dann auf die Frage nach dem epistemologischen Status von Klimamodellen zurückkommen.

2 Modelle und Epistemologie

Wissenschaft ist der Prozess der formalen Entdeckung von Gesetzmäßigkeiten in der Natur. Eine Erklärung oder ein formales Modell für eine solche Gesetzmäßigkeit wird als Theorie bezeichnet (oder als Gesetz, wenn sie gut belegt ist). Newtons Gravitationsgesetz ist ein klassisches und einfaches Beispiel. In diesem Fall gehorchen Objekte diesem Gesetz auf menschlicher Maßstabsebene offenbar exakt. Solche hochpräzisen Theorien werden gemeinhin als erklärend betrachtet.

Der Idealfall überprüfbarer Theorien findet sich in der klassischen

Physik. Newtons und Maxwells Gesetze liefern sehr spezifische Vorhersagen und schließen zudem bestimmte Ereignisse aus. Diese Gesetze wurden durch Experimente überzeugend bestätigt, doch ist zu beachten, dass selbst hier Störfaktoren wie Reibung kontrolliert werden müssen, um sie zu überprüfen. In diesen Fällen ist der Maßstab für die Gültigkeit einer Theorie sehr hoch. Experimentelle Daten stimmen oft fast perfekt mit der Theorie überein, und Ereignisse wie die Rückkehr eines Kometen lassen sich Jahrzehnte im Voraus vorhersagen. Die scheinbare Perfektion dieser Gesetze hat vielleicht zu der Überzeugung geführt, dass sie im absoluten, logischen Sinne „wahr“ sind, doch wie bereits erwähnt, weist selbst die Schwerkraft einige ungeklärte Merkmale auf.

Gültige und nützliche Theorien entstehen jedoch nicht fertig ausgearbeitet und perfekt, und sie sind auch nicht immer so präzise wie die Maxwell'schen Gleichungen. Als Alfred Wegener 1912 die Theorie der Kontinentalverschiebung vorstellte, konnte man keineswegs behaupten, dass seine Theorie ausgereift war. Es fehlte ein Auslöser für die Kontinentalbewegung (und vielen erschien es unmöglich, dass sich Kontinente bewegen könnten), ebenso wie ausreichende Belege. Als Daten gesammelt wurden, insbesondere zur Meeresbodenspreizung und zum Prozess der Subduktion, entstand ein schlüssiges Bild von Plattenbewegungen, der Entstehung von Gebirgszügen, dem Ursprung von Vulkanen und dem Grund für die Lage von Erdbebenzonen. Doch auch nach einem Jahrhundert der Reifung dieser Theorie bleibt sie eine qualitative Theorie, denn obwohl sie die allgemeinen Lageorte von Erdbeben- und Vulkanzonen erklären kann, lässt sie sich aufgrund der Heterogenität der Erdkruste und der Unmöglichkeit, detaillierte Daten zu erheben, weder die Stärke noch der genaue Ort oder der Zeitpunkt von Erdbeben oder Vulkanausbrüchen vorhersagen. Somit muss selbst eine mechanistische und gut erprobte Theorie nicht in der Lage sein, präzise Vorhersagen zu treffen – vielleicht niemals. Wenn eine Theorie reift, wird sie hoffentlich präziser, doch dies ist nicht garantiert (Loehle 1983).

In seinem berühmten Abgrenzungsprinzip weist Popper (1959, 1963) auf eine Asymmetrie hin: Es ist möglich, eine Theorie zuverlässig zu widerlegen, doch eine Theorie kann niemals bewiesen werden. Stattdessen stärken aufeinanderfolgende erfolgreiche Tests einer Theorie lediglich unser Vertrauen in sie. Das bedeutet nicht, dass wir nichts wissen, wie Wissensrelativisten behaupten mögen, sondern vielmehr, dass wissenschaftliches Wissen vorläufig, begrenzt (die Schwerkraft ist auf atomarer Ebene nicht eindeutig erklärbar) und eine Frage des Grades ist (Loehle 2011). In manchen Fällen kann dieses Wissen viele signifikante Stellen umfassen, in anderen Fällen ist es möglicherweise eher qualitativer Natur.

Entscheidend ist, dass die Überprüfung einer sich weiterentwickelnden Theorie nicht dem einfachen Modell der Hypothesenprüfung folgt, wie es in empirischen Experimenten verwendet wird, und dies auch nicht tun sollte. Bei der Prüfung eines Medikaments im Vergleich zu einem Placebo liefern statistische Tests oft eine einfache Antwort darauf, ob es

besser oder schlechter ist oder in welchem Ausmaß. Bei der Überprüfung einer Theorie gibt es hingegen mehrere Aspekte der Theorie, die zu einem bestimmten Zeitpunkt jeweils teilweise bestätigt werden können, sowie alternative Erklärungen, die möglicherweise ausgeschlossen werden müssen (Reiss 2015). Ein Netzwerk aus Bestätigung, Mathematik und kausaler Erklärung stützt den Glauben an eine Theorie zu jedem Zeitpunkt, nicht ein einfaches Ja oder Nein. Je ausgereifter eine Theorie wird und je strenger sie getestet wird, desto höher steigen wir auf der Skala der epistemischen Gewissheit. Es besteht jedoch eine Asymmetrie zwischen dem Beweisen einer Theorie und ihrer Anwendung für eine Berechnung. Die Tests, die zur Akzeptanz einer Theorie als „wahr“ führen, werden oft unter sorgfältig kontrollierten und idealen Bedingungen durchgeführt, wie beispielsweise im Vakuum. Bei jeder auf einer Theorie basierenden Berechnung wenden wir sie hingegen unter nicht-idealen Bedingungen an. Beispielsweise verhält sich eine fallende Feder im Vakuum anders als in der Luft. Die Brücke von der idealisierten Physik zu Anwendungen in der realen Welt bildet die Gesamtheit aus Näherungen, Vereinfachungen, Diskretisierungen, empirischen Beziehungen, geschätzten Anfangsbedingungen und numerischen Methoden, die zur Erstellung eines Berechnungswerkzeugs (Loehle 1983) verwendet werden, mit dem sich ein Ergebnis berechnen lässt. Diese Brückenbeziehungen verhindern, dass ein Berechnungswerkzeug eine perfekte Darstellung der zugrunde liegenden physikalischen (oder anderen) Theorie ist. Sind diese Störfaktoren nur schwer zu quantifizieren und zu modellieren, können wir möglicherweise keine Vorhersagen treffen (z. B. über die Flugbahn einer fallenden Feder). Die Richtigkeit eines Berechnungswerkzeugs ist somit eine empirische Frage danach, wie genau oder nützlich es ist, und nicht eine Frage von wahr oder falsch, wie wir es bei Theorien/Gesetzen annehmen.

3 Physikalische Grundlagen von Klimamodellen

Welchen erkenntnistheoretischen Status haben GCMs hinsichtlich ihrer physikalischen Grundlage? GCMs sind eine Mischung aus simulierten Prozessen, die als gut verstandene Physik gelten (z. B. Strahlungstransport), und solchen, die nur unzureichend verstanden werden (z. B. Wolkenmikrophysik, IPCC 2013, S. 599). Inwieweit lassen sich die verwendeten Algorithmen direkt auf bekannte physikalische Erkenntnisse zurückführen? Inwieweit belegt die physikalische Grundlage ihren Wahrheitswert, ihre Erklärungskraft oder ihre Zuverlässigkeit? Wie wir oben gesehen haben, sind physikalische Theorien, die unseren gängigen Vorstellungen von „Wahrheit“ nahekommen, zumindest unter idealisierten Bedingungen (z. B. in reibungslosen Vakuums) in der Lage, sehr präzise Vorhersagen für die reale Welt zu treffen. Können GCMs solche sauberen physikalischen Theorien wie Newtons Bewegungsgesetze im Vakuum annähern? Wenn ja, dann ist ein hohes Maß an Vertrauen in ihre Ergebnisse gerechtfertigt. Doch selbst bei einem einfachen Problem wie dem Werfen eines Würfels oder einer Münze bedeutet die Sensitivität gegenüber den Anfangsbedingungen, dass das Ergebnis nicht vorhergesagt werden kann, obwohl es auf bekannter Physik basiert. Im Zusammenhang mit Klimamodellen stellen Rougier und Goldstein (2014) fest, dass die

Gesetze des Klimasystems der Erde nicht vollständig bekannt sind und sich bei ausreichender Auflösung nicht explizit lösen lassen. Katzav et al. (2012) weisen darauf hin, dass die Vollständigkeit und strukturelle Stabilität der Modelle unbekannt sind. Dies gilt insbesondere für die Navier-Stokes-Gleichungen (N-S-Gleichungen) der Strömungsmechanik, für die keine analytischen Lösungen bekannt sind. Diese Unfähigkeit, die Gleichungen explizit zu lösen, ist der Grund dafür, dass numerische Simulationen verwendet werden. Die korrekte Simulation der Gleichungen der Strömungsdynamik ist jedoch alles andere als einfach (Thuburn 2008). Ein besonderes Problem besteht darin, dass die korrekte Lösung dieser Gleichungen zwar die kontinuierliche Erhaltung von Masse, Energie, Impuls und anderen Eigenschaften (auf unendlich vielen Skalen) erfordert, da es sich um partielle Differentialgleichungen handelt, die Modelle jedoch diskret sind. Prozesse wie die Dissipation von Energie und die Ausbreitung von Wirbeln finden unterhalb der Gitterstufe statt, und es gibt keine Theorie, die garantiert, dass das Gittermodell diese korrekt behandelt (McWilliams 2007; Marston et al. 2016). Simulierte Prozesse innerhalb eines Gitters breiten sich möglicherweise nicht reibungslos auf benachbarte Zellen aus, was zu Ringing, der Anhäufung numerischer Lösungsfehler im Laufe der Zeit, oder zu Fehlern bei Winden oder der korrekten Modellierung von Phänomenen wie der Quasi-Biennial Oscillation führen kann (Thuburn 2008). Diese Probleme wurden noch nicht angemessen gelöst (z. B. Katzav et al. 2012), und tatsächlich bleibt die Lösung der N-S-Gleichungen ein Millennium-Problem (siehe <http://www.claymath.org/millennium-problems/navier-stokes-equation>). Daher können die Modelle gegen Erhaltungssätze verstoßen und numerische Lösungsartefakte aufweisen. Stevens und Bony (2013) zeigten beispielsweise, dass selbst in einem idealisierten Modell eines Wasserplaneten mit vorgegebenen Oberflächentemperaturen die räumlichen Reaktionen von Wolken und Niederschlag auf die Erwärmung je nach Modell sehr unterschiedlich ausfallen. Dies verdeutlicht, dass noch keine Einigkeit darüber besteht, wie diese Prozesse auf einem Gitter dargestellt oder berechnet werden sollen. Zhou et al. (2015) dokumentieren Fehler bei der zonalen Mittelung der Sonneneinstrahlung in einigen Modellen. Staniforth und Thuburn (2012) dokumentieren, dass alle bestehenden numerischen Gitterlösungsschemata bekannte Probleme aufweisen, darunter Gitterabdrücke und die Anregung von Berechnungszuständen. Die Unzulänglichkeit aktueller Gitterverfahren zeigt sich daran, dass ein Modell mit höherer Auflösung oft viele Abweichungen im Vergleich zu aktuellen Modellen aufweist (Sakamoto et al. 2012). Es werden weiterhin verbesserte numerische Methoden eingeführt, um die bekannten Probleme bei der Lösung von N-S-PDEs zu beheben (z. B. Marston et al. 2016). Darüber hinaus gibt es in allen Modellen Subgitter-Parametrisierungen (McWilliams 2007; Katzav et al. 2012; Hourdin et al. 2016), welche die Unsicherheit erhöhen. McWilliams (2007) stellt fest, dass kleine strukturelle (gleichungsformale) Unterschiede in Subgitter-Parametrisierungen zu unterschiedlichen dynamischen Attraktoren in solchen Strömungsdynamik-Systemen führen können.

Betrachten wir die grundlegendsten physikalischen Aspekte von Klimamodellen: die Strahlungseigenschaften von CO₂ in der Atmosphäre. Zwar gibt es eine grundlegende Theorie für diesen Prozess, doch gibt es zahlreiche Software-Tools zur Strahlungstransportberechnung (Oreopoulos und Mlawer 2010), da die Berechnung des Strahlungstransports auf einem Globus mit heterogener Atmosphäre ein schwieriges numerisches Problem darstellt – im Gegensatz zur Beschleunigung eines fallenden Körpers im Vakuum. Das Spektrum wird in jedem dieser Programme mit unterschiedlichen Auflösungen unter Verwendung verschiedener geometrischer Annahmen und Verfahren bewertet. Noch gravierender ist, wie Oreopoulos und Mlawer (2010) dokumentieren, dass 1) sich die grundlegende Theorie selbst ständig weiterentwickelt; 2) die in GCMs verwendeten Algorithmen aus rechnerischen Gründen stark vereinfacht sind; und 3) verschiedene GCMs nicht die gleichen Strahlungstransfer-Algorithmen verwenden. Es ist somit klar, dass auch hier eine Lücke zwischen der grundlegenden Theorie und den Berechnungen besteht, mit unklaren Konsequenzen.

Ebenso trifft jedes GCM unterschiedliche Annahmen hinsichtlich des Antriebs, der Wolken, der Landoberflächen, der räumlichen Rasterung usw. und verwendet unterschiedliche numerische Verfahren zur Lösung. Die geschätzten Antriebe haben sich zwischen den IPCC-Berichten AR4 und AR5 erheblich verändert, und der Einfluss von Aerosolen wird derzeit noch überarbeitet (z. B. Stevens 2015), wobei es erhebliche Unterschiede in der Darstellung zwischen den Modellen gibt (Wilcox et al. 2013). Für Prozesse, die unterhalb der Gitterauflösung ablaufen, wie z. B. Wolkenverhalten und Niederschlag, werden Parametrisierungen (d. h. empirische Beziehungen) verwendet (McWilliams 2007). Diese empirischen Beziehungen enthalten freie Parameter, die abgestimmt werden müssen (Lahsen 2005; McWilliams 2007; Mauritsen et al. 2012; Schmidt und Sherwood 2015; Hargreaves 2010; Hourdin et al. 2016), und diese Anpassungen können willkürlich sein (z. B. Soon et al. 2001, deren Abb. 4). Fehler in diesen Näherungen sind schwer zu quantifizieren, entfernen die Modelle jedoch zweifellos weit vom Bereich der reinen Darstellung idealer physikalischer Gesetze, wie etwa der Schwarzkörperstrahlung von einer gleichmäßigen Oberfläche bekannter Temperatur, wie auch von Katzav et al. (2012) argumentiert. Es lässt sich auch argumentieren, dass wesentliche physikalische Prozesse in den Modellen unberücksichtigt bleiben, wie beispielsweise die Auswirkungen des elektrischen Feldes der Erde (Andersson et al. 2014).

Wenn globale Klimamodelle nicht als präzise Abbildungen der Theorie betrachtet werden können, die auf der Ableitung einiger Komponenten aus fundierten physikalischen Erkenntnissen beruhen (wie oben dargelegt), welchen epistemologischen Status haben sie dann? Ein Ansatz zur Beurteilung ihres Wahrheitswerts besteht darin, nicht ausgehend von der zugrunde liegenden Physik zu argumentieren, sondern rückwärts von der Qualität ihrer Ergebnisse. Es lässt sich erfolgreich argumentieren, dass sie Aspekte des aktuellen Verständnisses des Erdklimasystems verkörpern, da sie sonst überhaupt nicht funktionieren würden. Katzav (2014) sowie

Schmidt und Sherwood (2015) argumentieren beispielsweise, dass diese Verkörperung von Wissen durch die Überlegenheit aktueller Modelle gegenüber einem naiven Modell oder gegenüber Klimamodellen der vorherigen Generation belegt wird. Smith (2002) und Oreskes et al. (1994) schlagen vor, dass die Modelle eine nützliche Analogie oder Heuristik darstellen. McWilliams (2007) argumentiert, dass aufgrund der irreduziblen Unsicherheit in den Modellausgaben, die auf chaotische Dynamiken zurückzuführen ist, GCMs eher anhand ihrer Plausibilität beurteilt werden sollten als danach, ob sie korrekt oder die besten sind. Er argumentiert, dass die Modelle „Raum-Zeit-Muster liefern, die an die Natur erinnern ... und damit eine sinnvolle Art von Turing-Test zwischen dem Künstlichen und dem Tatsächlichen bestehen“. Der IPCC (2013, S. 145) stellt fest, dass diese Modelle als Werkzeuge zum Verständnis des Klimasystems betrachtet werden können. Viele Ergebnisse (insbesondere die Temperatur) zeigen eine gute Übereinstimmung zwischen den Modellen, was auf einen gewissen Wahrheitswert der Modelle hindeutet (Räisänen 2007). Die Übereinstimmung zwischen den Modellen kann jedoch auf gemeinsamen Annahmen, gemeinsamen Algorithmen und ähnlichen Daten beruhen, die zur Kalibrierung verwendet werden. Parker (2011) argumentiert, dass die Übereinstimmung der Vorhersagen zwischen den Modellen zwar gewisse Belege liefert, jedoch nicht ausreicht, um eine epistemische Gewissheit hinsichtlich ihres Wahrheitswerts zu begründen. Aus diesen Gründen verfehlen Bemühungen zur Bestätigung (Verifizierung) von Klimamodellen (z. B. Lloyd 2010, Diskussion in Katzav et al. 2012) ihr Ziel. Auch wenn diese Modelle plausibel sein können, eine Art Turing-Test bestehen und miteinander übereinstimmen, bedeuten die Probleme der irreduziblen Dynamik und der numerischen Unsicherheit (z. B. McWilliams 2007) sowie andere Aspekte, dass die theoretische Grundlage der Modelle nicht als Garant für die Gültigkeit nützlicher Vorhersagen angesehen werden kann. Dies wirft die Frage nach ihrer Nützlichkeit als Vorhersageinstrumente auf, die im Folgenden erörtert wird.

4 Klimamodelle als Berechnungs-Grundlagen

Da sich globale Klimamodelle ständig weiterentwickeln und es einigen Aspekten möglicherweise an einer strengen und engen Verbindung zur zugrunde liegenden Physik mangelt, sind sie nach Poppers Kriterien nicht widerlegbar (siehe Curry und Webster 2011) und müssen als Berechnungsinstrumente betrachtet werden. Daher ist es notwendig, die Modelle vor ihrer Anwendung in irgendeiner Form zu überprüfen.

Das Testen komplexer Simulationsmodelle ist schwierig. Die große Anzahl von (anhand von Daten geschätzten) Parametern in diesen Modellen (Murphy et al. 2004; Hargreaves 2010; Schmidt und Sherwood 2015; Hourdin et al. 2016) lässt vermuten, dass die parametrische Unsicherheit der Modelle hoch sein könnte, doch wurde dies bislang nur unzureichend untersucht (Guttorp 2014). Es gibt potenzielle Probleme hinsichtlich der Struktur (Gleichungsform), der Parameter und der Datenfehler (Loehle 1987, 1988; Hourdin et al. 2016), die bisher kaum untersucht worden sind. Es gibt

viele spezifische Arten von Sensitivitäts- und Fehleranalysen, die durchgeführt werden können (z. B. Falloon et al. 2014; Guttorp 2014; Rougier und Goldstein 2014), um die Zuverlässigkeit von Modellausgaben zu bewerten, doch diese Methoden wurden aufgrund ihres hohen Rechenaufwands fast nie auf GCMs angewendet (Falloon et al. 2014). Allen und Ingram (2002) sowie McWilliams (2007) argumentieren, dass „Ensembles of Opportunity“ (eine Sammlung von Modellen) die Modellunsicherheit nicht angemessen erfassen, und empfehlen eine umfassende Unsicherheitsanalyse (Anfangsbedingungen, Parameter, Funktionsform der Gleichungen, numerische Methode usw.), um mögliche Prognosen einzugrenzen – eine Analyse, die für GCMs bislang noch nicht durchgeführt worden ist. Daher stehen Entscheidungsträgern für GCMs keine entscheidenden Informationen zur Modellunsicherheit zur Verfügung.

Modelle der turbulenten Dynamik zeigen eine Empfindlichkeit gegenüber den Anfangsbedingungen (Frigg et al. 2013). Bei einem strukturell perfekten Modell (d. h. alle Gleichungen und Parameter sind korrekt; numerische Methoden funktionieren korrekt) lässt sich der Einfluss der Unsicherheit der Anfangsbedingungen abschätzen, indem mehrere Durchläufe mit gestörten Anfangsbedingungen durchgeführt werden, was eine Wahrscheinlichkeitsverteilung für die Ergebnisse liefert. Dies setzt voraus, dass die Fehler in den Anfangsbedingungen charakterisiert werden können und dass eine ausreichende Anzahl von Durchläufen durchgeführt werden kann, was bei Klimamodellen in der Regel nicht der Fall ist (McWilliams 2007). In einer einzigartigen Fallstudie störten Deser et al. (2016) einen Basislauf durch Rauschen auf Maschinenfehler-Niveau (d. h. Rundungsfehler), das auf das anfängliche Temperaturfeld angewendet wurde. Sie fanden bei 30 Durchläufen sehr große Unterschiede von mehreren °C in den 50-Jahres-Trends im Winter für Regionen Nordamerikas. Sie stellten fest, dass ein Ensemble-Ansatz die interne Variabilität vom erzwungenen Signal trennen könnte, um eine bessere Übereinstimmung mit historischen Daten zu erzielen. Dies basiert jedoch auf einer infinitesimalen Störung der Anfangsbedingungen. Die tatsächlichen Unsicherheiten der Anfangsbedingungen sind um viele Größenordnungen größer. Noch wichtiger ist, dass diese stochastische Störung der Anfangsbedingungen bei Vorliegen struktureller Fehler (falsche Gleichungsform zur Darstellung eines Prozesses) nicht nur uninformativ, sondern sogar irreführend sein kann. (Smith 2002; Frigg et al. 2014; Hourdin et al. 2016).

Es könnte aufschlussreicher sein, die Ergebnisse der globalen Klimamodelle (GCM) nicht als Ganzes, sondern genauer zu betrachten, um zu erkennen, was mit ausreichender Genauigkeit vorhergesagt werden kann. Der IPCC (2013) stellt die GCM-Ergebnisse der globalen Mitteltemperatur seit 1850 als Abweichungen vom Mittelwert dar; werden sie jedoch als absolute Temperaturen dargestellt, weichen die Zeitreihen bis zu 4 °C voneinander ab (SI Abb. 2). Ein ähnliches Ergebnis (Abweichungen 4 °C) wurde für die kontinentalen USA festgestellt (Anagnostopoulos et al. 2010). Dies ist kein trivialer Unterschied, da die Langwellenstrahlung eines Objekts gemäß der Stefan-Boltzmann-Relation proportional zur

vierten Potenz der absoluten Temperatur ist (Anagnostopoulos et al. 2010). Wenn sich Modelle in der mittleren Temperatur so stark unterscheiden, behandeln sie dann die grundlegende Physik auf die gleiche Weise oder setzen sie die Physik mit korrekten Algorithmen um? Dies wirft epistemische Fragen hinsichtlich der von GCMs erstellten Vorhersagen auf. Hawkins und Sutton (2016) weisen darauf hin, dass bei einer linearen Reaktion auf einen erhöhten Antrieb die absolute Temperatur für die Abschätzung der Reaktion auf einen erhöhten Antrieb keine große Rolle spielt. Liegt jedoch eine starke positive Rückkopplung vor, ist die Reaktion auf einen erhöhten Antrieb bei höheren Temperaturen stärker (Bloch-Johnson et al. 2015, Gregory et al. 2015). Wenn hingegen eine negative Rückkopplung den CO₂-Antrieb dämpft (z. B. Spencer und Braswell 2011), würde dies ebenfalls von der tatsächlichen Temperatur abhängen. In beiden Fällen wäre die absolute Temperatur von Bedeutung (d. h., die Reaktion ist nichtlinear) und die Verwendung von Anomalien lässt sich nicht rechtfertigen. Anomalien, manchmal auch als „Bias-Korrektur“ bezeichnet, werden auch zum Vergleich anderer Klimadaten verwendet. Allerdings reagieren Nutzpflanzen, Biodiversität, Meeresspiegel und Eisschilde alle auf tatsächliche Niederschläge und Temperaturen, und somit würden die verschiedenen Modelle sehr unterschiedliche Auswirkungen prognostizieren, selbst wenn ihre Anomalietrends übereinstimmen würden, wie von Hawkins und Sutton (2016) angemerkt. Der Nettoeffekt der Bias-Korrektur oder der Verwendung von Anomalien besteht darin, den epistemologischen Status der Modelle zu verschleiern, indem die Streuung der Modellausgaben zueinander verringert wird und Unstimmigkeiten mit den Daten schwer zu bestimmen sind.

Die Anwendung einer Bias-Korrektur kann zu weiteren Schwierigkeiten bei der Überprüfung führen. Betrachten wir den Fall, dass globale Temperaturverläufe mit Modellausgaben verglichen werden. Wenn die Daten in tatsächlichen °C vorliegen oder über einen bestimmten Zeitraum auf eine gemeinsame Basislinie verschoben werden, wird die Korrelationsstatistik nicht beeinflusst, da der konstante Term aus der Berechnung herausfällt. Bei anderen Messgrößen kann die Basislinie jedoch einen Einfluss haben. Beispielsweise unterscheidet sich die R²-Statistik für die Modellanpassung bei tatsächlichen Reihen von der bei Anomalie-Reihen und kann bei nicht verschobenen Reihen sogar negativ sein (d. h., die Anpassung an die Daten ist schlechter als an einen einfachen Mittelwert der Daten). Hawkins und Sutton (2016) weisen darauf hin, dass die Normalisierung (Basislinienverschiebung) einer Klimareihe auf einem Referenzzeitraum basiert, typischerweise 30 Jahre, aber auch den gesamten Aufzeichnungszeitraum umfassen kann. Sowohl die Daten als auch die Modellausgaben werden nach oben oder unten verschoben, sodass ihre jeweiligen Mittelwerte über den Referenzzeitraum Null betragen. Beim Vergleich mehrerer Durchläufe eines einzelnen Modells oder mehrerer Modelle mit den Daten stimmen diese während des Referenzzeitraums am ehesten überein. Das bedeutet, dass der visuelle Eindruck der Modellanpassung oder der Zeitpunkt, zu dem ein Modell gut oder schlecht abschneidet, vollständig vom gewählten Referenzzeitraum abhängen kann

(Beispiele finden sich bei Hawkins und Sutton 2016). Dies wirkt sich beispielsweise auf die Frage aus, ob die Modelle derzeit höhere Werte vorhersagen als die Daten zeigen. Je näher der gewählte Referenzzeitraum an der Gegenwart liegt, desto größer erscheint die Übereinstimmung zwischen den Modellen und den Daten der letzten Jahre. Bei Anpassungsstatistiken wie R^2 kann die Wahl des Referenzzeitraums ebenfalls das Ergebnis und damit die implizierte Modellanpassung beeinflussen. In Abbildung 2 wird beispielsweise ein künstliches Beispiel gezeigt. In Abbildung 2a werden sowohl die Daten als auch das Modell auf den 100-jährigen Referenzzeitraum (Mittelwert 0) verschoben. Die Anpassung erscheint visuell recht gut, und $R^2 = 0,79$. In Abbildung 2b werden jedoch die letzten 30 Jahre als Referenzzeitraum verwendet. Nun scheint das Modell in der Vergangenheit schlechter und in den letzten Jahrzehnten besser (fast perfekt) zu passen, aber R^2 beträgt nun 0,54, was eine erhebliche Verschlechterung darstellt. Dies wirft ein epistemisches Dilemma auf. Wenn die Korrelation als Maß für gemeinsame Trends und Muster (z. B. Temperaturanstiege und -abfälle) verwendet wird, berücksichtigt dies nicht die Verzerrung (den Offset) in den Modellausgaben. Werden Modelle und Daten auf Anomaliebasis betrachtet, so setzt dies für Temperatur und Niederschlag voraus, dass die tatsächlichen Werte keine Rolle spielen, sondern nur der Trend, doch dies ist nach wie vor umstritten. Darüber hinaus beeinflusst der gewählte Referenzzeitraum sowohl den visuellen Eindruck der Modellanpassungsgüte (sowohl hinsichtlich der Ensemble-Streuung als auch des Anpassungsmusters über die Zeit) als auch alle Anpassungsstatistiken mit Ausnahme der einfachen Korrelation. Fragen wie diese haben Auswirkungen auf die epistemische Gewissheit.

5 Schlussfolgerungen

Wie verhält es sich nun mit der von den GCMs aufgeworfenen Wissensfrage? Als parametrisierte Simulatoren, die Klimaverläufe generieren, müssen diese Instrumente grundsätzlich statistisch und quantitativ bewertet werden. Qualitative Einschätzungen geben keine Antwort auf die entscheidenden politikrelevanten Fragen, wie stark, wann und wo die Erwärmung eintreten wird. Held (2005) argumentiert, dass die Erlangung eines besseren Klimaverständnisses die Entwicklung vereinfachter, idealisierter „Welten“ (siehe z. B. SI Abb. 1) erfordert, um die Prozesse großräumiger Turbulenzen, der Wärmeübertragung zu den Polen, der Ozeanzirkulation und insbesondere die Frage zu untersuchen, wie große Klimaphänomene wie ENSO bestehen bleiben können. Ohne diese Erforschung der Prozesse, so Held, sei es nicht möglich zu erklären, warum verschiedene GCMs unterschiedliche Ergebnisse liefern, warum sie von den Daten abweichen und wie sie verbessert werden können. Dies liege daran, dass die Komplexität der Modelle zu epistemischer Undurchsichtigkeit führe. Angemessene Erklärungen für das Verhalten komplexer hierarchischer Systeme wie des Klimas müssten in der Regel mehrstufig sein und Faktoren wie Meeresströmungen, Kontinente und Wolken berücksichtigen. Ein auf diese Weise gewonnenes besseres Verständnis könnte zu besseren Subgitter-Parametrisierungen führen. Ein Beispiel

hierfür ist die aktuelle Arbeit von Moncrieff et al. (2017), die einen multiskaligen Ansatz zum Verständnis organisierter tropischer Konvektion ableitet, der zur Entwicklung von Subgitter-Parametrisierungen genutzt werden kann.

Wenn Klimamodelle dem realen Erdsystem lediglich „ähneln“ und eher als Analogie (Oreskes et al. 1994) oder als explorative Werkzeuge dienen, sind sie am nützlichsten als Grundlage für qualitative Vorhersagen, etwa dass eine gewisse Erwärmung wahrscheinlich ist. Wenn die Modelle bestimmte Vorhersagen (z. B. zur globalen Temperatur) mit akzeptabler Genauigkeit treffen können, ist es wichtig zu ermitteln, welche Variablen auf diese Weise vorhergesagt werden können. Wenn Modelle eine gemeinsame Verzerrung aufweisen, kann diese Verzerrung möglicherweise bei politischen Entscheidungen berücksichtigt werden. Erklärungen für Unterschiede in der Modellleistung sollten gesucht werden, insbesondere für die große Bandbreite zukünftiger Entwicklungspfade. Angesichts der Komplexität des Erdklimasystems verdient die Grundlage für die auf GCMs basierenden Wissensaussagen größere Aufmerksamkeit. Richtig angewandte Erkenntnistheorie kann helfen zu klären, was wir wissen, wie wir es wissen und wo die Grenzen einer strengen, gerechtfertigten Argumentation liegen.

Der Klimawandel stellt ein komplexes politisches Problem dar. Sowohl Handeln als auch Untätigkeit bergen ein hohes Risiko. Diese Abhandlung führt zu keiner bestimmten politischen Schlussfolgerung. Vielmehr konzentriert sie sich auf die Verfahren, die zu einer stringenten Argumentation führen. Politische Entscheidungen beinhalten zwangsläufig auch Risikowahrnehmung, Risikotoleranz, kulturelle Werte, wirtschaftliche Aspekte und andere Faktoren, die über den Rahmen dieser Analyse hinausgehen. Jede Politik kann jedoch nur davon profitieren, wenn besser verstanden wird, wie Klimamodelle aufgebaut sind, auf welchen physikalischen Grundlagen sie beruhen, wie sie getestet werden können und wie ihre Ergebnisse zu bewerten sind.

Referenzen

Allen MR, Ingram WJ (2002) Constraints on future changes in climate and the hydrological cycle. *Nature* 419:224-232.

Anagnostopoulos GG, Koutsoyiannis D, Christofides A, Efstratiadis A, Mamassis N (2010) A comparison of local and aggregated climate model outputs with observed data. *Hydrol Sci J* 55:1094-1110.

Andersson ME, Verronen PT, Rodger CJ, Clilverd MA, Seppälä A (2014) Missing driver in the sun-earth connection from energetic electron precipitation impacts mesospheric zone. *Nat Comm* 5:5197.

Bloch-Johnson J, Pierrehumbert RT, Abbot DS (2015) Feedback temperature dependence determines the risk of high warming. *Geophys Res Lett* 42:4973-4980.

Curry JA, Webster PJ 2011. Climate science and the uncertainty monster. *Bull Am Meteorol Soc* 92:1667-1682.

Deser C, Terray L, Phillips AS (2016) Forced and internal components of winter air temperature trends over North America during the past 50 years: mechanisms and implications. *J Climate* 29:223-2258.

Falloon P, Challinor A, Dessai S, Hoang L, Johnson J, Koehler A-K (2014) Ensembles and uncertainty in climate change impacts. *Front Environ Sci* 2:33.

Frigg R, Bradley S, Du H, Smith LA (2014) Laplace's demon and the adventures of his apprentices. *Philos Sci* 81:31-59.

Gregory JM, Andrews T, Good P (2015) The inconstancy of the transient climate response parameter under increasing CO₂. *Philos Trans Roy Soc A* 373:20140417.

Guttorp P (2014) Statistics and climate. *Ann Rev Stat Appl* 1:87-101.

Hargreaves JC (2010) Skill and uncertainty in climate models. *Wiley Interdisciplinary Reviews: Climate Change* 1:556-564.

Hawkins E, Sutton R (2016) Connecting climate model projections of global temperature change with the real world. *Bull Am Meteorol Soc* 2016:963-980.

Held IM (2005) The gap between simulation and understanding in climate modeling. *Bull Am Meteorol Soc* 86:1609-1614.

Hourdin F, Mauritsen T, Gettelman A, Golaz J-C, Balaji V, Duan Q, Folini D, Ji D, Klocke D, Qian Y, Rauser F, Rio C, Tomassini L, Watanabe M, Williamson D (2016) The art and science of climate model tuning." *Bull Am Meteorol Soc* in press.

IPCC (2013) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

Katzav J (2014) The epistemology of climate models and some of its implications for climate science and the philosophy of science. *Studies in History and Philosophy of Modern Physics* 46:228-238.

Katzav J, Dijkstra HA, de Laat ATJ (2012) Assessing climate model projections: state of the art and philosophical reflections. *Studies in History and Philosophy of Modern Physics* 43:258-276.

Lahsen M (2005) Seductive simulations? Uncertainty distribution around

climate models. *Social Studies of Science* 35:895-922.

Lloyd EA (2010) Confirmation and robustness of climate models. *Philos Sci* 77:971-984.

Loehle C (1983) Evaluation of theories and calculation tools in ecology. *Ecol Modell* 19:239-247.

Loehle C (1987) Errors of construction, evaluation, and inference: a classification of sources of error in ecological models. *Ecol Modell* 36:297-314.

Loehle C (1988) Philosophical tools: potential contributions to ecology. *Oikos* 51:97-104.

Loehle C (2011) The logic of scientific discovery. *Current Trends in Ecology* 2:75-81.

Marston JB, Chini GP, Tobias SM (2016) Generalized quasilinear approximation: application to zonal jets." *Physical Rev Lett* 116:21450.

Mauritsen T, Stevens B, Roeckner E, Crueger T, Esch M, Giorgetta M, Haak H, Jungclaus J, Klocke D, Matei D, Mikolajewicz U, Notz D, Pincus R, Schmidt H, Tomassini L (2012) Tuning the climate of a global model. *J Adv Model Earth Sys* 4:M00A01.

McWilliams JC (2007) Irreducible imprecision in atmospheric and oceanic simulations. *Proc Natl Acad Sci* 104:8709-8713.

Moncrieff, M.W., Liu, C., and Bogenschutz, Peter. 2017. Simulation, modeling, and dynamically based parameterization of organized tropical convection for global climate models. doi:10.1175/JAS-D-16-0166.1.

Murphy JM, Sexton DMH, Barnett DN, Jones GS, Webb MJ, Collins M, Stainforth DA (2004) Quantification of modelling uncertainties in a large ensemble of climate change situations. *Nature* 430:768-772.

Oreopoulos L, Mlawer E (2010) The Continual Intercomparison of Radiation Codes (CIRC): assessing anew the quality of GCM radiation algorithms. *Bull Am Meteorol Soc* 91:305-310.

Oreskes N, Shrader-Frechette K, Belitz K (1994) Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 263:641-646.

Parker WS (2011) When climate models agree: the significance of robust model predictions. *Philos Sci* 78:579-600.

Popper KR (1959) *The logic of scientific discovery*. Hutchinson, London.

Popper KR (1963) *Conjectures and refutations: the growth of scientific knowledge*. Harper & Row New York.

Reiss J (2015) A pragmatist theory of evidence. *Philos Sci* 82:341-362.

Rougier J, Goldstein M (2014) Climate simulators and climate projections. *Ann Rev Stat Appl* 1:103-123.

Sakamoto TT, Komuro Y, Nishimura T, Ishii M, Tatebe H, Shiogama H, Hasegawa A, Toyoda T, Mori M, Suzuki T, Imada Y, Nazawa T, Takata K, Mochizuki T, Ogochi K, Emori S, Hasumi H, Kimoto M (2012) MICR04h – a new high resolution atmosphere-ocean coupled general circulation model. *J Meteorol Soc Japan* 90:325-359.

Schmidt GA, Sherwood S (2015) A practical philosophy of complex climate modelling. *Eur J Philos Sci* 5:149-169.

Smith LA (2002) What might we learn from climate forecasts? *Proc Natl Acad Sci* 99:2487-2492.

Soon W, Baliunas S, Idso SB, Kondratyev KY, Posmentier ES (2001) Modeling climatic effects of anthropogenic carbon dioxide emissions: unknowns and uncertainties. *Climate Research* 18:259-275.

Spencer RW, Braswell WD (2011) On the misdiagnosis of surface temperature feedbacks from variations in Earth's radiant energy balance. *Remote Sensing* 3:1603-1613.

Staniforth A, Thuburn J (2012) Horizontal grids for global weather and climate prediction models: a review. *Quart J Royal Meteorol Soc* 138:1-26.

Stevens B (2015) Rethinking the lower bound on aerosol radiative forcing. *J Climate* 28:4794-4819.

Stevens B, Bony S (2013) What are climate models missing? *Science* 340:1053.

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc* 93:485-498.

Thuburn J (2008) Some conservation issues for the dynamical cores of NWP and climate models. *J Comput Phys* 227:3715-3730.

Wegener A (1966) *The origin of continents and oceans* (Biram J, trans.). Courier Dover. p 246.

Wilcox LJ, Highwood EJ, Dunstone NJ (2013) The influence of anthropogenic aerosol on multi-decadal variations of historical global climate. *Environ Res Lett* 8:024033.

Williams M (2001) *Problems of knowledge: a critical introduction to epistemology*. Oxford University Press.

Zhou L, Zhang M, Bao Q, Liu Y (2015) On the incident solar radiation in

CMIP5 models. *Geophys Res Lett* 42:1930-1935.

Link:

<https://wattsupwiththat.com/2026/06/02/are-climate-models-just-physics/>

Übersetzt von Christian Freuer für das EIKE

Kommentar des Übersetzers zu diesem Beitrag:

Antwort auf die Frage in der Überschrift: NEIN!

Ein Wettermodell erfasst die gegenwärtige Lage mittels aller verfügbaren Messungen. Weil die physikalischen Gesetze alle bekannt sind, nach welchen sich meteorologische Parameter entwickeln, können Wettermodelle auf einige Tage recht gut die großräumigen Wetterstrukturen abbilden (Planetarische Wellen der Höhenströmung z. B.). Schon bei der modelleigenen Interpretation der großräumigen Entwicklung versagen sie jedoch recht schnell (z. B. wenn man die Niederschlagsmenge für ein bestimmtes Gebiet simulieren will. Das hat sich in der Kältewelle Mitte Mai eindrucksvoll gezeigt).

Fazit: Wettermodelle basieren AUSSCHLIEßLICH auf Messungen, auf welche dann bekannte physikalische Gesetze angewendet werden. DAS ist Physik! Aber welche Messwerte gehen in Klimamodelle ein?