

Ein verstecktes Universum der Unsicherheit

geschrieben von Chris Frey | 25. Oktober 2022

[Kip Hansen](#)

Jedes Mal, wenn jemand in unserer Gemeinschaft, der Gemeinschaft der Wissenschaftsskeptiker oder Realisten®, über Unsicherheit und deren Auswirkungen auf von Experten begutachtete wissenschaftliche Ergebnisse spricht, wird er sofort beschuldigt, Wissenschaftsleugner zu sein oder zu versuchen, den gesamten Bereich der Wissenschaft zu untergraben.

Ich habe hier immer wieder darüber geschrieben, dass die Ergebnisse der meisten Studien in der Klimawissenschaft die Unsicherheit ihrer Ergebnisse bei weitem unterschätzen. Lassen Sie mich dies so deutlich wie möglich sagen: **Jedes Ergebnis, das keine ehrliche Diskussion der mit der Studie verbundenen Unsicherheiten enthält, angefangen bei den Unsicherheiten der Rohdaten bis hin zu den Unsicherheiten, die durch jeden Schritt der Datenverarbeitung *hinzugefügt* werden, ist die digitale Tinte nicht wert, mit der es veröffentlicht wird.**

[Hervorhebung vom Übersetzer]

Eine neue große, von mehreren Forschungsgruppen durchgeführte und in den [Proceedings](#) of the National Academy of Sciences veröffentlichte Studie wird die Forschungswelt aufrütteln. Sie stammt ausnahmsweise nicht von John P.A. [Ioannidis](#), der für sein [Buch](#) „Why Most Published Research Findings Are False“ bekannt ist.

Der Titel der [Studie](#) lautet [übersetzt]: „Die Beobachtung vieler Forscher, die dieselben Daten und Hypothesen verwenden, enthüllt ein verborgenes Universum idiosynkratischer Ungewissheit“ [oder als [.pdf hier](#)].

Das ist gute Wissenschaft. So sollte Wissenschaft gemacht werden. Und so sollte Wissenschaft veröffentlicht werden.

Erstens, wer ist der Verfasser dieser Studie?

Das waren Nate Breznau und viele viele andere. Breznau ist an der Universität Bremen tätig. Bei den Koautoren gibt es eine Liste von 165 Koautoren aus 94 verschiedenen akademischen Einrichtungen. Das bedeutet, dass es sich nicht um die Arbeit einer einzelnen Person oder einer einzelnen verärgerten Forschungsgruppe handelt.

Was haben sie getan?

Die Forschungsfrage lautet wie folgt: „*Werden verschiedene Forscher zu*

ähnlichen Ergebnissen kommen, wenn sie dieselben Daten analysieren?„

Sie taten dies:

„Dreiundsiebzig unabhängige Forschungsteams haben identische länderübergreifende Umfragedaten verwendet, um eine etablierte sozialwissenschaftliche Hypothese zu testen: dass mehr Einwanderung die öffentliche Unterstützung für die Bereitstellung von Sozialmaßnahmen durch die Regierung verringert.“

Was haben sie herausgefunden?

„Anstelle einer Konvergenz variierten die numerischen Ergebnisse der Teams stark und reichten von großen negativen bis zu großen positiven Auswirkungen der Einwanderung auf die öffentliche Unterstützung.“

Eine andere Möglichkeit, dies zu betrachten, besteht darin, die tatsächlichen numerischen Ergebnisse der verschiedenen Gruppen zu betrachten, die dieselbe Frage unter Verwendung identischer Daten stellten:

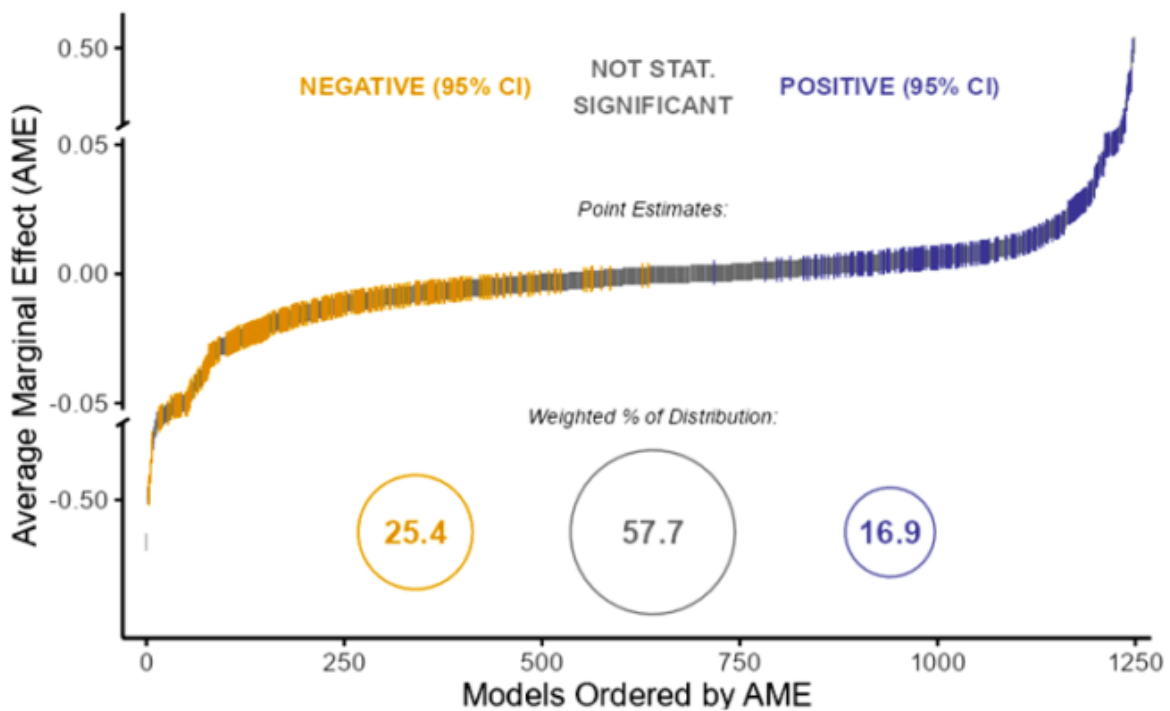


Fig. 1 Broad variation in findings from 73 teams testing the same hypothesis with the same data. The distribution of estimated average marginal effects (AME) across all converged models (N = 1,253) includes results that are negative (yellow, and in the direction predicted by the given hypothesis the teams were testing), not different from zero (grey) or positive (blue), using a 95% confidence interval. AME are XY-standardized. Y-axis contains two breaks at +/-0.05. Numbers inside circles represent the percentage of the distribution of each outcome inversely weighted by the number of models per team (see Interactive Results).

Der Diskussionsteil beginnt mit folgendem Satz:

„Diskussion: Die Ergebnisse unseres kontrollierten Forschungsdesigns in einem groß angelegten Crowdsourcing-Forschungsprojekt, an dem 73 Teams beteiligt waren, zeigen, dass die Analyse derselben Hypothese mit denselben Daten zu erheblichen Unterschieden bei den statistischen Schätzungen und inhaltlichen Schlussfolgerungen führen kann. Tatsächlich kamen keine zwei Teams zu den gleichen numerischen Ergebnissen oder trafen während der Datenanalyse die gleichen wichtigen Entscheidungen.“

Will jemand noch mehr wissen?

Wenn Sie wirklich wissen wollen, warum Forscher, die dieselbe Frage stellen und dieselben Daten verwenden, zu völlig unterschiedlichen und widersprüchlichen Antworten kommen, müssen Sie die Studie lesen.

Was hat dies mit dem *Many-Analysts-Ansatz* zu tun?

Im Juni letzten Jahres habe ich über einen [Ansatz](#) für wissenschaftliche Fragen geschrieben, der sich „*The Many-Analysts Approach*“ nennt.

Der *Many-Analysts-Ansatz* wurde wie folgt angepriesen:

„Wir argumentieren, dass die derzeitige Art der wissenschaftlichen Veröffentlichung – die sich mit einer einzigen Analyse begnügt – die „Modell-Myopie“, eine begrenzte Berücksichtigung statistischer Annahmen, verfestigt. Dies führt zu übermäßigem Selbstvertrauen und schlechten Vorhersagen. ... Um die Robustheit ihrer Schlussfolgerungen zu beurteilen, sollten die Forscher die Daten mehreren Analysen unterziehen, die idealerweise von einem oder mehreren unabhängigen Teams durchgeführt werden.“

In der neuen Studie, die heute diskutiert wird, heißt es dazu:

„Selbst hoch qualifizierte Wissenschaftler, die motiviert sind, zu genauen Ergebnissen zu kommen, unterscheiden sich enorm in dem, was sie herausfinden, wenn ihnen dieselben Daten und Hypothesen zur Prüfung vorgelegt werden. Die übliche Präsentation und der Konsum wissenschaftlicher Ergebnisse legten nicht die Gesamtheit der Forschungsentscheidungen im Forschungsprozess offen. Unsere Schlussfolgerung ist, dass wir ein verborgenes Universum idiosynkratischer Forschervariabilität erschlossen haben.“

Und das bedeutet für Sie und mich, dass weder der Ansatz mit vielen Analysten noch der Ansatz mit vielen Analyseteams das Real World™-Problem lösen wird, das sich aus den inhärenten Unsicherheiten des modernen wissenschaftlichen Forschungsprozesses ergibt – viele Analysten/Teams werden leicht unterschiedliche Ansätze, unterschiedliche statistische Verfahren und leicht unterschiedliche Versionen der verfügbaren Daten verwenden. Die Teams treffen Hunderte von winzigen Annahmen, die sie meist als „beste Praktiken“ betrachten. Und aufgrund dieser winzigen Unterschiede kommt jedes Team zu einem absolut

vertretbaren Ergebnis, das mit Sicherheit einer Peer-Review standhält, aber jedes Team kommt zu unterschiedlichen, ja sogar widersprüchlichen Antworten auf dieselbe Frage, die an dieselben Daten gestellt worden ist.

Das ist genau das Problem, das wir in CliSci jeden Tag sehen. Wir sehen dieses Problem in der Covid-Statistik, der Ernährungswissenschaft, der Epidemiologie aller Art und vielen anderen Bereichen. Dies ist ein anderes Problem als die unterschiedlichen Voreingenommenheiten bei politisch und ideologisch sensiblen Themen, der Druck in der Wissenschaft, Ergebnisse zu finden, die mit dem aktuellen Konsens im eigenen Fachgebiet übereinstimmen, und die schleichende Krankheit der Kumpel-Begutachtung (Pal-Review).

In der Klimawissenschaft herrscht der Irrglaube vor, dass mehr Verarbeitung – Mittelwertbildung, Anomalien, Kriging, Glättung usw. – die Unsicherheit verringert. Das Gegenteil ist der Fall: mehr Verarbeitung erhöht die Unsicherheiten. Die Klimawissenschaft erkennt nicht einmal die einfachste Art der Unsicherheit an – die ursprüngliche Messunsicherheit – sondern wünscht sie sich weg.*

[***Einschub des Übersetzers:** In seiner über 40-jährigen Praxis im Bereich Wetteranalyse und -vorhersage konnte der Übersetzer die Erfahrung machen, dass numerische Modelle im Kurzfristbereich zwar eine enorme Verbesserung der Vorhersagegüte erzielt werden konnte, während mittel- und langfristig (4 bis 10 Tage) im Voraus die Unsicherheiten direkt proportional zur Anzahl der Modellrechnungen stiegen.

Um die Güte numerischer Modellrechnungen zu testen, wird der Anfangszustand in ein und demselben Modell künstlich geringfügig verändert. Denn auch die Numerik kann nur auf der Grundlage aktueller Meldungen funktionieren. Nun gibt es aber bekanntlich riesige Meldelücken, die zwar mit High Tech überbrückt werden können (z. B. Fernerkundung mittels Satelliten), aber Extrapolationen müssen trotzdem gemacht werden. Die winzigen Änderungen des Ausgangszustandes sind mit bloßem Auge gar nicht erkennbar.

Lässt man nun aber die Modelle mit diesen geringfügigsten Änderungen immer wieder laufen, zeigt sich Folgendes:

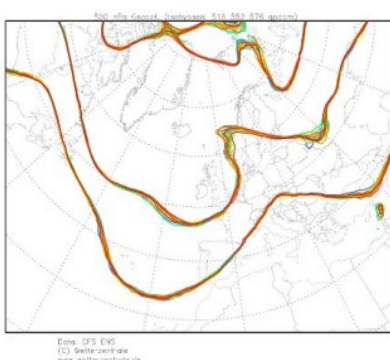


Abbildung 2: Numerische Simulation der Änderung im 500-hPa-Niveau.

Links: nach 24 Stunden (1 Tag; [Quelle](#)), **Mitte:** nach 168 Stunden (1 Woche; [Quelle](#)), **Rechts:** nach 384 Stunden (15 Tage; [Quelle](#))

Fazit: Nach einer Woche ist immerhin noch tendenziell ein Strömungsmuster erkennbar, nach 15 Tagen aber nicht mehr. **Alle Lösungen beruhen auf der gleichen Ausgangslage und sind gleich wahrscheinlich!**

Nun ja, unsere werten Alarmisten, Politiker und MSM wollen uns ja einreden, dass sie derartige Simulationen für **viele Jahrzehnte im Voraus** berechnen können – **mit einer Zuverlässigkeit, die angeblich schon jetzt drastische Maßnahmen mit in jeder Hinsicht verheerenden Folgen erfordert!**

Ende Einschub]

Ein anderer Ansatz, der sich sicher anbietet ist, die Ergebnisse der divergierenden Befunde nun einer Mittelwertbildung zu unterziehen oder den Mittelwert – eine Art Konsens – der Vielzahl von Befunden zu finden. Die Abbildung der Ergebnisse zeigt diesen Ansatz als den Kreis mit 57,7% der gewichteten Verteilung. Diese Idee ist nicht valider als die Mittelwertbildung von chaotischen Modellergebnissen, wie sie in der Klimawissenschaft praktiziert wird – mit anderen Worten: wertlos.

Pielke Jr. schlägt in einer kürzlich gehaltenen [Präsentation](#) und einer anschließenden Frage- und Antwortrunde mit der National Association of Scholars vor, dass es wahrscheinlich der beste Ansatz ist, die besten echten Experten in einem Raum zusammenzubringen und diese Kontroversen zu klären. Pielke Jr. ist ein anerkannter Fan des vom IPCC verwendeten Ansatzes – allerdings nur, solange dessen Ergebnisse nicht von Politikern beeinflusst werden. Trotzdem neige ich dazu zuzustimmen, dass es zu besseren Ergebnissen führen könnte, wenn man die besten und ehrlichsten (nicht kämpfenden) Wissenschaftler eines Fachgebiets zusammen mit Spezialisten für Statistik und die Bewertung von Programmmathematik in einem virtuellen Raum versammelt, mit dem Auftrag, die größten Unterschiede in den Ergebnissen zu überprüfen und zu beseitigen.

Man frage nicht mich

Ich bin kein aktiver Forscher. Ich habe keine spontane Lösung für die „Drei Ks“ – die Tatsache, dass die Welt 1) kompliziert, 2) komplex und 3) chaotisch ist. Diese drei Faktoren addieren sich und schaffen die Unsicherheit, die jedem Problem innewohnt. Diese neue Studie fügt eine weitere Ebene hinzu – die Ungewissheit, die durch die vielen kleinen Entscheidungen verursacht wird, die die Forscher bei der Analyse einer Forschungsfrage treffen.

Es hat den Anschein, dass die Hoffnung, die Ansätze mit vielen Analysten und vielen Analyseteams würden zur Lösung einiger der kniffligen

wissenschaftlichen Fragen unserer Zeit beitragen, enttäuscht wurde. **Es hat auch den Anschein, dass wir eher misstrauisch als beruhigt sein sollten, wenn Forschungsteams, die behaupten, unabhängig zu sein, zu Antworten kommen, die den Anschein einer zu engen Übereinstimmung haben.**

[Hervorhebung vom Übersetzer]

Kommentar des Autors:

Wenn Sie sich dafür interessieren, warum sich Wissenschaftler selbst bei einfachen Fragen nicht einig sind, dann müssen Sie diese Studie unbedingt lesen, und zwar jetzt. Pre-print .pdf [hier](#).

Wenn es Ihr Verständnis für die Schwierigkeiten, gute und ehrliche Wissenschaft zu betreiben, nicht ändert, brauchen Sie wahrscheinlich eine Gehirntransplantation. ... Oder zumindest einen neuen Kurs über kritisches Denken für Fortgeschrittene.

Wie immer sollten Sie sich nicht auf mein Wort verlassen. Lesen Sie den Artikel, und gehen Sie vielleicht zurück und lesen Sie meinen früheren Artikel über [Many Analysts](#).

Gute Wissenschaft ist nicht einfach. Und da wir immer schwierigere Fragen stellen, wird es auch nicht einfacher werden.

Die einfachste Sache der Welt ist es, neue Hypothesen aufzustellen, die vernünftig erscheinen, oder unrealistische Vorhersagen für die Zukunft zu machen, die weit über unsere eigene Lebenszeit hinausgehen. Die Zeitschrift *Popular Science* hat mit solchen Dingen einen Geschäftsplan gemacht. Die heutige „theoretische Physik“ scheint ein Spiel daraus zu machen – wer kann die verrückteste und dennoch glaubwürdige Idee darüber entwickeln, „wie die Dinge wirklich sind“.

Link:

<https://wattsupwiththat.com/2022/10/17/a-hidden-universe-of-uncertainty/>

Übersetzt von [Christian Freuer](#) für das EIKE