

Kritik an einer neuen Studie von Santer et al. (2019)

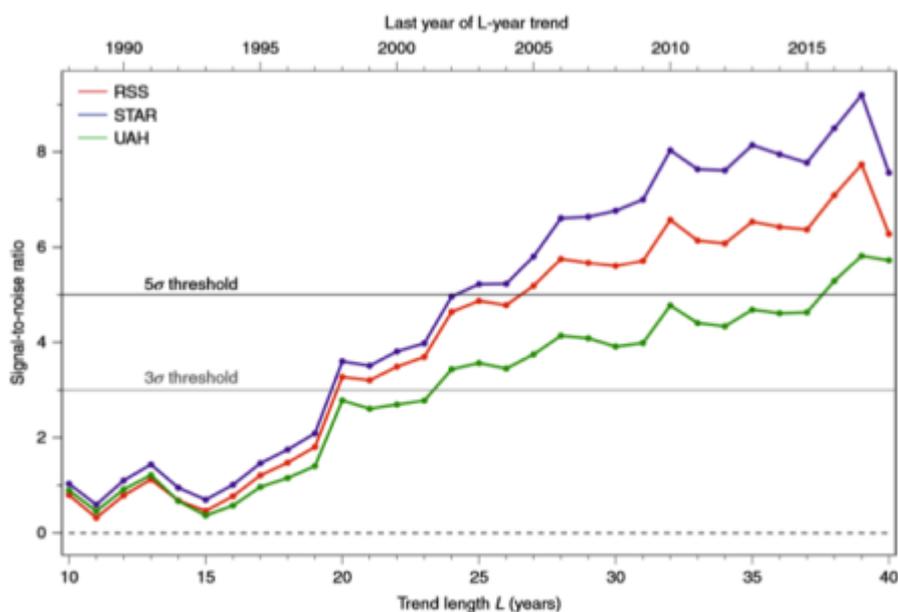


Abbildung 1: aus Santer et al. (2019)

Die Ergebnisse ihrer Studie zeigt Abbildung 1. Es ist keine Graphik der Temperatur, sondern eines geschätzten Verhältnisses zwischen „Signal und Rauschen“. Die horizontale Linie repräsentiert Sigma-Einheiten, die man – falls das zugrunde liegende statistische Modell korrekt ist – interpretieren kann als die Punkte, an welchen der Schweif der Verteilung sehr klein wird. Wenn also die Linie ein Sigma-Niveau passiert, ist das „Signal“ anthropogener Erwärmung aus dem „Rauschen“ der natürlichen Variabilität im passenden Umfang hervorgetreten. Sie berichten, dass die 3-Sigma-Grenze einen p-Wert von 1/741 aufweist, während die 5-Sigma-Grenze einen p-Wert von 1/3,5 Millionen hat. Da alle Linien die 5-Sigma-Marke bis 2015 überqueren, folgern sie, dass die anthropogene Auswirkung auf das Klima definitiv gefunden ist.

Ich möchte hier vier Aspekte dieser Studie ansprechen, welche die Schlussfolgerungen erheblich schwächen: (a) die Differenz zwischen der Existenz eines Signals und der Größenordnung der Auswirkung, (b) die verwirrende Natur ihres experimentellen Designs, (c) das invalide Design

ihres Nur-Natur-Vergleichsparameters und (d) Probleme bei der Relation von „Sigma“-Grenzen und Wahrscheinlichkeiten.

(a) Existenz eines Signals vs. Größenordnung der Auswirkung

Nehmen wir einmal an, wir stellen einen alten Analog-Radioempfänger auf ein schwaches Signal einer weit entfernten Radiostation ein. Mittels Verschieben des Senderknopfes bekommt man vielleicht ein Signal, das gut genug ist um zu erkennen, dass eine Komposition von Bach gespielt wird. Aber die Stärke des Signals sagt nichts über das Tempo der Musik; das ist eine andere Berechnung.

In gleicher Weise sagt uns obiges Diagramm nichts über die Größenordnung der Auswirkung von Treibhausgasen auf die Temperatur und das Klima. Es zeigt nur zweierlei: Eine Maßzahl der Rate, wie sich die Korrelation zwischen Beobachtungen und Modellergebnissen, getrieben durch natürliche und anthropogene Antriebe mit der Zeit, verbessert, dividiert durch eine Maßzahl der Standardabweichung der gleichen Maßzahl unter einer „Null-Hypothese“ einer (vermeintlich) rein natürlichen Variabilität. In dieser Hinsicht ist es wie eine t-Statistik [?], die ebenfalls in Sigma-Einheiten gemessen wird. Da es keine Verbesserung mit der Zeit geben kann im *Fit* zwischen den Beobachtungen und dem *natural-only comparator**, führt jede Verbesserung des Signals zu einer Erhöhung des Sigma-Niveaus.

[*Eine passende Übersetzung ist schwer. Weiter unten ist der Terminus übersetzt worden mit „nur-natürlich-Komparator“. Bessere Vorschläge seitens der Leserschaft sind willkommen. Anm. d. Übers.]

Selbst falls man Abbildung 1 als Nominalbeitrag akzeptiert, ist sie konsistent mit einer sehr hohen oder sehr geringen Sensitivität bzgl. Treibhausgasen – oder irgendwo dazwischen. Sie ist beispielsweise konsistent mit den Ergebnissen von [Christy and McNider](#), welche ebenfalls auf Satellitendaten basieren, dass nämlich die Sensitivität bei Verdoppelung der Treibhausgase viel geringer ist als typischerweise von den Modellen simuliert.

(b) Verwirrtes Signal-Design

Der Begleitinformation zufolge zogen Santer et al. jährlich gemittelte Klimamodell-Daten auf der Grundlage historischer Antriebe und auf dem (RCP8.5)-Szenario basierende natürliche und anthropogene Antriebe heran und konstruierten daraus eine Temperatur-Zeitreihe der mittleren Troposphäre. Diese enthielt auch eine Adjustierung für stratosphärische Abkühlung (d. h. eine „Korrektur“). Sie mittelten alle Modelle und Modellläufe, vernetzten die Daten in 10 Grad X 10 Grad-Gitterboxen (insgesamt 576, wobei die Polargebiete außen vor blieben) und extrahierten daraus 40 jährliche Temperatur-Anomalien für jede Gitterbox im Zeitraum 1979 bis 2018. Aus diesen extrahierten sie einen räumlichen „Fingerabdruck“ des vom Modell erzeugten Klima-Musters unter Anwendung einer wesentlichen Komponenten-Analyse aka empirischer Orthogonal-Funktionen. Man kann es sich vorstellen als ein gewichtetes zeitliches Mittel der Anomaliewerte für jede Gitterbox. Obwohl es in der Studie oder in der Begleitinformation nicht gezeigt wird, sieht das Muster so aus (die Graphik stammt aus einer separaten Studie):



Abbildung 2: Räumliches Muster eines Fingerabdrucks

Die grauen Bereiche in Abbildung 2 über den Polen repräsentieren außen vor gelassene Gitterboxen, da nicht alle Satelliten-Reihen auch die Pole abdecken. Die Farben markieren *PC „loadings“* [?] und nicht Temperaturen, aber da das erste *PC* etwa 98% der Varianz erklärt, kann man sie als mittlere Temperaturanomalien betrachten, ohne zu weit abzuschweifen. Folglich ist das Fingerabdruck-Muster ein solches von verstärkter Erwärmung in den Tropen mit lokalen Abweichungen hier und dort.

Dies ist das Muster, mit dem sie die Beobachtungen korrelieren wollen als ein Verfahren, um den anthropogenen „Fingerabdruck“ aufzuspüren. Aber in den Modellen ist es verbunden mit sowohl natürlichen als auch anthropogenen Antrieben zusammen im Zeitraum 1979 bis 2018. Sie bezeichnen dies als HIST+8.5-Daten, was heißt, dass die Modellläufe bis 2006 mit historischen Antrieben (sowohl natürlich als auch anthropogen) liefen und danach nur noch mit Antrieben gemäß dem RCP8.5-Szenario. Die Schlussfolgerung der Studie lautet, dass die Beobachtungen jetzt den in obiger Abbildung dargestellten Verhältnissen ähneln als eine Null-Hypothese („Nur natürliche Antriebe“) – ergo: anthropogener Fingerabdruck entdeckt. Aber HIST8.5 ist ein kombinierter Fingerabdruck, und tatsächlich haben sie den anthropogenen Anteil nicht herausgefiltert.

Sie haben also keineswegs einen eindeutigen anthropogenen Fingerabdruck identifiziert. Was sie entdeckt haben ist, dass Beobachtungen besser zu Modellen passen, wenn sie ein Erwärmungsmuster wie in Abbildung 2 enthalten, unabhängig von den Gründen, als jene, die das nicht enthalten. Es kann sein, dass ein Graph, welcher das ausschließlich anthropogene Signal darstellt, so aussehen würde wie in Abbildung 1, aber aus ihrer Analyse geht das in keiner Weise hervor.

(c) Invalider nur-natürlicher Komparator

Obiges Argument würde eine geringere Rolle spielen, falls der „nur-natürliche“ Komparator als Treiber der Erwärmung durch natürliche Antriebe fungieren würde. Aber das ist konstruktionsbedingt nicht der Fall.

Das Fingerabdruck-Verfahren beginnt damit, die beobachtete räumliche Verteilung von Temperaturanomalien heranzuziehen und diese mit dem Muster in Abbildung 2 oben zu korrelieren, wobei sich ein Korrelations-Koeffizient für jedes Jahr ergibt. Dann betrachten sie den Trend jener Korrelations-Koeffizienten als Maßzahl dafür, wie gut der *Fit* mit der Zeit wird. Die Korrelationen selbst werden weder in der Studie noch in den Begleitinformationen angesprochen.

Danach konstruieren die Autoren ein „Rauschen“-Muster, welches als „nur-natürlich“ dienen soll, entgegen den Fakten in obigem Diagramm. Sie beginnen mit der Auswahl von 200-Jahre-Kontrollläufen von 36 Modellen und teilen sie in das gleiche 10 X 10-Format ein. Vielleicht werden sie über alles mitteln, aber erst wird jede Gitterbox in jedem Modell trendbereinigt, was ich als fehlerhaften Schritt betrachte.

Alles hängt davon ab, wie valide der Komparator der natürlichen Variabilität ist. Man gibt uns keine Erklärung, warum die Autoren glauben, dass es eine glaubwürdige Analogie ist zu natürlichen Temperatur-Mustern, die assoziiert sind mit nicht-anthropogenen Antrieben nach 1979. Das ist fast mit Sicherheit nicht der Fall. Die Summe der Vulkan- und Solar-Reihe nach 1979 im AR 5 des IPCC sieht so aus:

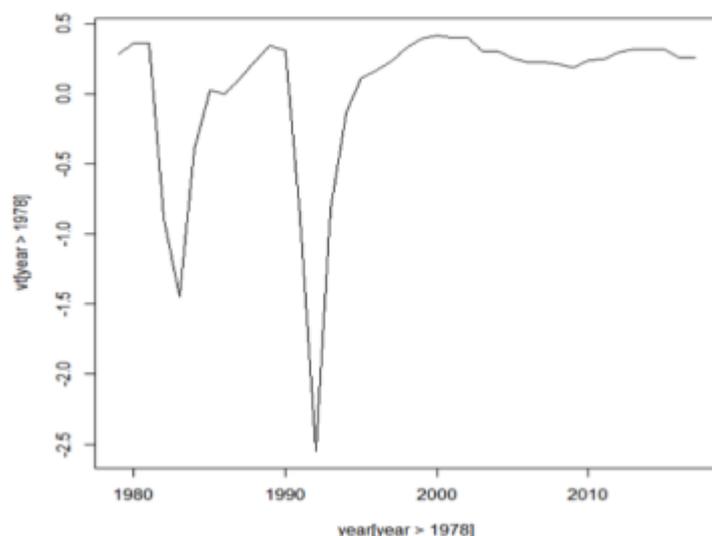


Abbildung 3: Natürliche Antriebe laut IPCC von 1979 bis 2017

Dies impliziert eindeutig, dass natürliche Antriebe eine insgesamt Erwärmung induziert hätten über das Stichproben-Intervall, und da die tropische Verstärkung unabhängig von der Art des Antriebs erfolgt, würde eine räumliche „nur-natürlich“-Verteilung wahrscheinlich so aussehen wie Abbildung 2. Aber mit der Trendbereinigung jeder Gitterbox haben Santer et al. derartige Muster eliminiert und den geschätzten natürlichen Komparator nach 1979 künstlich verschlechtert.

Die Schlussfolgerungen der Autoren hängen entscheidend von der Hypothese ab, dass ihre Schätzung aufgrund der „natürlichen“ Modell-Variabilität plausibel repräsentiert, wie es von 1979 bis 2018 ohne Treibhausgase ausgesehen hätte. Die Autoren betonen die Bedeutung dieser Hypothese in ihren Begleitinformationen (Seite 10):

„Unsere Hypothese hinsichtlich der Angemessenheit unserer Schätzungen bzgl. der Modell-Variabilität ist entscheidend. Beobachtete Temperaturaufzeichnungen sind simultan beeinflusst sowohl von interner Variabilität als auch von multiplen externen Antrieben. Wir verfolgen nicht die „reine“ interne Variabilität, so dass es immer etwas nicht reduzierbare Unsicherheit bei der Unterteilung der beobachteten Temperaturaufzeichnungen in intern und extern getriebene Komponenten geben wird. Alle Vergleiche der beobachteten Variabilität mit derjenigen in den Modellen sind beeinflusst

durch diese Unsicherheit, vor allem in weniger gut aufgezeichneten Zeitmaßstäben vieler Jahrzehnte“.

Wie sie sagen, jede Studie zur Erkennung von Fingerabdruck und Signalen ist abhängig von der Qualität des „nur-natürlich“-Komparators. Unglücklicherweise haben sie durch die Trendbereinigung ihrer Kontrollläufe Gitterbox für Gitterbox so ziemlich sichergestellt, dass das Muster der natürlichen Variabilität als ein Komparator künstlich degradiert wird.

Das ist so, als ob man von einem Bankräuber wüsste, dass er ein 1,80 Meter großer Mann ist und die Polizei ihre bevorzugten Verdächtigen in eine Linie mit kleinen Frauen stellen würde. Man könnte eine vertrauenswürdige Zeugen-Identifikation bekommen, aber man wüsste nicht, ob diese valide ist.

Was alles noch viel schlimmer macht, die Treibhausgas-beeinflusste Erwärmung stammt aus Modellen, welche so frisiert worden sind, dass sie zu grundlegenden Aspekten der beobachteten Erwärmungstrends des 20. Jahrhunderts passen. Während dies in der mittleren Troposphäre weniger eine Rolle spielen würde als an der Oberfläche, gäbe es dennoch die teilweise Verstärkung der Gleichheit zwischen Modellsimulationen und Beobachtung infolge nachträglicher Manipulation.

...

[Anmerkung des Übersetzers: Es folgen noch mehrere detaillierte Darstellungen und Vergleiche zwischen den Parametern, die nur für Fachleute klar durchschaubar sind. Dazu beschreibt der Autor die Entwicklung eines eigenen Modells. Alles läuft darauf hinaus, dass Santer's Studie wieder einmal viel verschleiert und nichts enthüllt. Dies geht auch aus der Bewertung des Autors am Ende des Beitrags hervor:]

Schlussfolgerungen

Die Tatsache, dass die t-Statistik von *anthro* in meinem Beispiel auf ein tiefes Niveau fällt, „beweist“ nicht, dass anthropogene Antriebe keine Auswirkungen auf die Temperatur haben. Sie zeigt aber, dass im Rahmen meines Modells die Auswirkungen nicht signifikant sind. ...

Und das liegt in der Natur der Sache statistischer Modellierungen: sie beruhen auf Spezifikationen und Hypothesen. ...

Weil ich also genug Gründe habe, die Validität des Modells von Santer et al. zu bezweifeln, akzeptiere ich auch nicht ihre Schlussfolgerungen. Sie haben nicht das gezeigt, was sie sagen gezeigt zu haben. Im Einzelnen haben sie keinen eindeutigen anthropogenen Fingerabdruck identifiziert oder eine glaubwürdige Berücksichtigung der natürlichen Variabilität vorgelegt. Auch haben sie nicht den Gebrauch von Gauss'schen p-Werten gerechtfertigt. Ihre Behauptung, einen „goldenen Standard“ von Beweisen erhalten zu haben, ist unberechtigt, teils schon deswegen, weil statistische Modellierung das niemals ergeben kann, und teils wegen der spezifischen Probleme in ihrem Modell.

Link:

<https://judithcurry.com/2019/03/01/critique-of-the-new-santer-et-al-2019-paper/>

Übersetzt von [Chris Frey](#) EIKE