

Gavin Schmidt und seine „Trickserei“ mit Referenz-Perioden

geschrieben von Steve McIntyre | 11. Mai 2016

Schmidt erhob alle Vorwürfe, die er schon zu oft zuvor erhoben hatte:

@curryja use of Christy's misleading graph instead is the sign of partisan not a scientist. YMMV. tweet;

@curryja Hey, if you think it's fine to hide uncertainties, error bars & exaggerate differences to make political points, go right ahead. tweet.

[Übersetzung extra, weil ich nicht weiß, ob diese links erhalten bleiben: *Der Gebrauch von Christys irreführender Graphik ist ein Zeichen dafür, dass er ein Partisan und kein Wissenschaftler ist.*

Falls Sie glauben, dass es in Ordnung ist, Unsicherheiten und Fehlerbalken zu verstecken sowie Differenzen zu übertreiben, dann machen Sie einfach weiter.]

Als Folge hiervon hat sich Curry entschlossen, Christys Graphik nicht bei ihrem jüngsten Vortrag vor einem Komitee des Kongresses zu zeigen. Im heutigen Beitrag möchte ich die Validität von Schmidts Kritik untersuchen (oder das Fehlen derselben).

Soweit ich das verstehe, ging es bei Schmidts primärem Streit darum, dass Christy seine Modell- und Beobachtungsdaten so aufbereitet hatte, dass sie beide im Jahre 1979 ihren Ausgang nahmen, also dem Beginn der Satelliten-Ära. Dieses Verfahren zeigt (offensichtlich) eine größere Diskrepanz am Ende des Zeitraumes als wenn die Daten in der Mitte des Zeitraumes platziert wären. Ich werde die Unterstützung für Christys Verfahren durch seinen vieljährigen Kontrahenten Carl Mears unter die Lupe nehmen, dessen eigene Vergleiche zwischen Modell und Beobachtungen eine kurze Periode (1979 bis 1983) überdeckten, „damit die Änderungen mit der Zeit besser erkennbar waren“. Während sowohl Christy als auch Mears rationale Argumente für ihre Grundsatzentscheidung vorbrachten, war Schmidts Argumentation nur wenig mehr als Geschrei.

Hintergrund

Die ganze Geschichte der Kontroverse um die Diskrepanz zwischen Modellen und Beobachtungen in der tropischen Troposphäre ist literaturmäßig voluminös. Während die wesentlichen Protagonisten Christy, Douglass und Spencer auf der einen Seite sowie Santer, Schmidt, Thorne und andere auf der anderen Seite waren, haben auch Ross McKittrick und ich selbst zu diesem Thema bereits Einiges gesagt, und [die Kommentare von] McKittrick et al. (2010) wurden ausführlich im AR 5 diskutiert, allerdings unglücklicherweise wie so oft hinsichtlich der Schlüsselpunkte

betrügerisch.

Startzeitpunkte und Referenzperioden

Christy und Spencer haben viele Jahre lang stilistisch ähnliche Graphiken erstellt. Roy Spencer zeigte hier im Jahre 2014 eine ähnliche Graphik um den Zeitraum 1979 bis 1983 (unten). Tatsächlich war es diese frühere Version, die Bart Verheggen zu wüsten Kommentaren veranlasste, Kommentare, die offenbar ihren Ursprung in einigen der vorherrschenden alarmistischen Memes* haben.

[Wer wie ich mit dem Begriff „Meme“ nichts anfangen kann – hier steht eine gute Erklärung. A. d. Übers.].

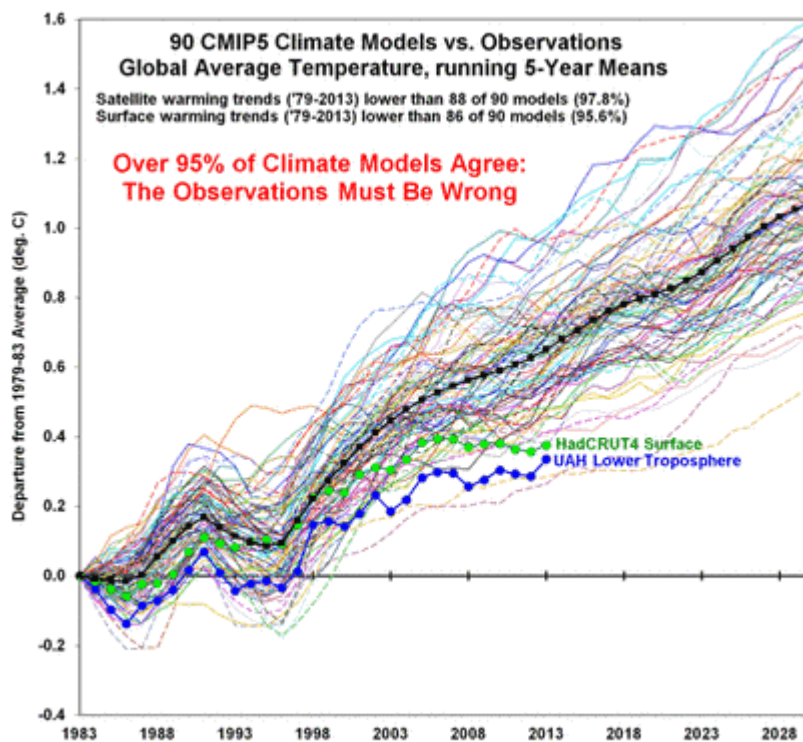


Abbildung 1: Version 2014 der Graphik von Christy, entnommen dem Blog von Roy Spencer hier. Sie verwendete den Referenzzeitraum 1979 bis 1983. Dies wurde später hier von Bart Verheggen kritisiert.

Christys Vortrag vom Februar 2016 erklärte diesen allgemeinen Ursprung als die geeignetste Referenzperiode, wobei der Start eines Rennens als Metapher angeführt wurde:

Hierzu sage ich im Gegenteil, dass wir die Daten auf bedeutsamste Art und Weise dargestellt haben. Es geht hier um die Rate der Erwärmung des Großteils der Atmosphäre, d. h. den Trend. Diese Metrik sagt uns, wie schnell sich Wärme in der Atmosphäre akkumuliert – die fundamentale Metrik der globalen Erwärmung. Um dies visuell zu verdeutlichen habe ich alle Datensätze adjustiert, so dass sie einen gemeinsamen Startzeitpunkt haben. Man denke an folgende Analogie: Ich habe während der letzten 25

Jahre an über 500 Rennen teilgenommen, und in jedem Falle starten alle Läufer an der gleichen Stelle zur gleichen Zeit. Damit soll einfach festgestellt werden, wer der Schnellste ist und um wie viel schneller an der Ziellinie. Offensichtlich wird die Gesamt-Relativgeschwindigkeit eindeutig bestimmt durch ihre Reihenfolge bei der Überquerung der Ziellinie – aber alle müssen zusammen starten.

Das in der Graphik von 2016 angewendete Verfahren variierte in gewissem Ausmaß vom früheren Stil: sie verwendete den Wert aus dem Jahr 1979 des Trends von 1975 bis 2005 als Referenz für die Zentrierung; also einen Wert, der sehr nahe dem Mittelwert von 1979 bis 1983 lag.

Carl Mears

Ironischerweise hat Christys immerwährender Gegner Carl Mears auf der RSS-Website des Vergleichs zwischen Modell und Beobachtung eine fast identische Referenzperiode herangezogen (1979 bis 1984), damit **„die Änderungen mit der Zeit einfacher erkennbar sind“**. Mears schrieb: „Falls die Modelle als Ganzes akzeptable Arbeit geleistet hätten bei der Simulation der Vergangenheit, dann würden die Beobachtungen zumeist innerhalb des gelben Bandes liegen“, aber „dies war nicht der Fall“.

*Das gelbe Band zeigt die 5% bis 95%-Einhüllende des Ergebnisses von 33 CMIP5-Modellsimulationen (19 verschiedene Modelle, viele davon mit multiplen Erkenntnissen). Sie sollen das Erdklima während des 20. Jahrhunderts simuliert haben. Für die Zeit vor 2005 wurden die Modelle mit historischen Werten von Treibhausgasen gefüttert, mit vulkanischen Aerosolen und solarem Output. Nach 2005 wurden hierfür geschätzte Projektionen dieser Antriebe verwendet. **Falls die Modelle als Ganzes einen akzeptablen Job hinsichtlich der Simulation der Vergangenheit geleistet hätten, dann würden die Beobachtungen zumeist im Bereich des gelben Bandes liegen. Für die ersten beiden Plots (Abbildung 1 und 2), welche globale und tropische Mittelwerte zeigen, ist dies nicht der Fall.***

Mears illustrierte den Vergleich mit der folgenden Graphik, deren Legende den Referenzzeitraum 1979 bis 1984 und die damit verbundene Erklärung liefert:

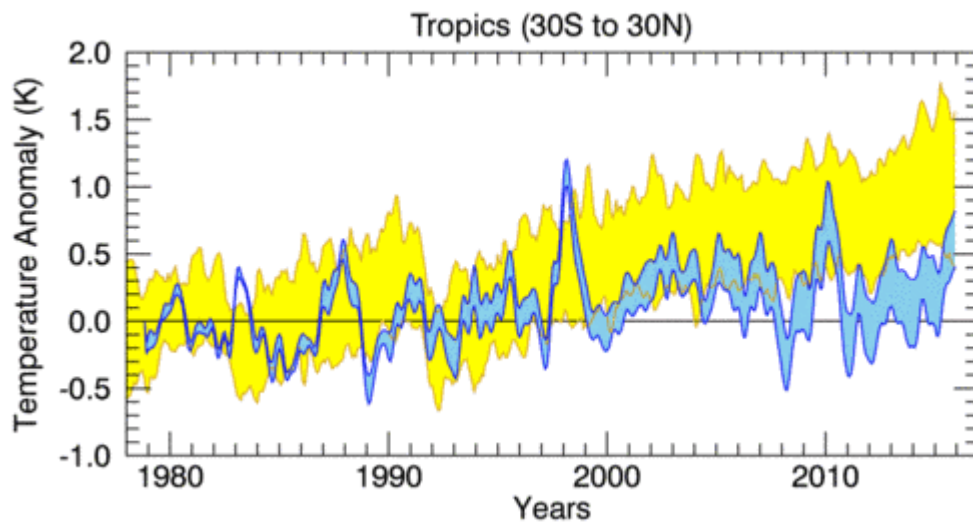


Abbildung 2, entnommen von RSS hier: **Original-Legende:** tropische (30S bis 30N) mittlere TLT-Anomalie, dargestellt als eine Funktion der Zeit. Das blaue Band ist die 5% bis 95%-Einhüllende für das RSS V3.3 MSU/AMSU-Temperatur-Unsicherheitsensemble. Das gelbe Band ist die Bandbreite von 5% bis 95% des Outputs von CMIP5-Klimasimulationen. **Der mittlere Wert jeder Zeitreihe von 1979 bis 1984 wurde auf Null gesetzt, so dass die Änderungen mit der Zeit deutlicher erkennbar sind.** Erneut liegen die Beobachtungen nach 1998 wahrscheinlich unter den simulierten Werten. Dies zeigt, dass die Simulation als Ganzes mehr Erwärmung prophezeit als von den Satelliten beobachtet.

Die sehr geringe Überlappung zwischen der Einhüllenden der Modelle und derjenigen der Beobachtungen ist ein klarer Beweis – für jeden geübten Beobachter – dass es eine statistisch signifikante Differenz gibt zwischen dem Ensemble-Mittel und den Beobachtungen, und zwar bei Verwendung der T-Statistik von Santer et al. 2008.

Nichtsdestotrotz war Mears nicht der Ansicht, dass die Schuld bei den Modellen zu suchen war. Sondern er argumentierte stattdessen zusammen mit Santer, dass das Phänomen Fehlern der Antriebe geschuldet war, Fehler in den Beobachtungen und der internen Variabilität (siehe hier). Trotz dieser Differenzen bei der Diagnose stimmte Mears mit Christy überein hinsichtlich der Eignung eines gemeinsamen Ursprungs für diese Art von Vergleich.

IPCC AR5

Das IPCC, das sich nach Schmidts Worten nicht scheut, „Differenzen zu übertreiben oder zu minimieren, um politisch zu punkten“, hat eine Referenzperiode inmitten des Satelliten-Intervalls ausgewählt (1986 bis 2005) für seine Graphik im AR 5, Kapitel 11, Abbildung 11.25, in der ein globaler Vergleich von CMIP5-Modellen mit dem Mittel von 4 Datensätzen mit Messungen gezeigt wird.

Global mean temperature near-term projections relative to 1986–2005

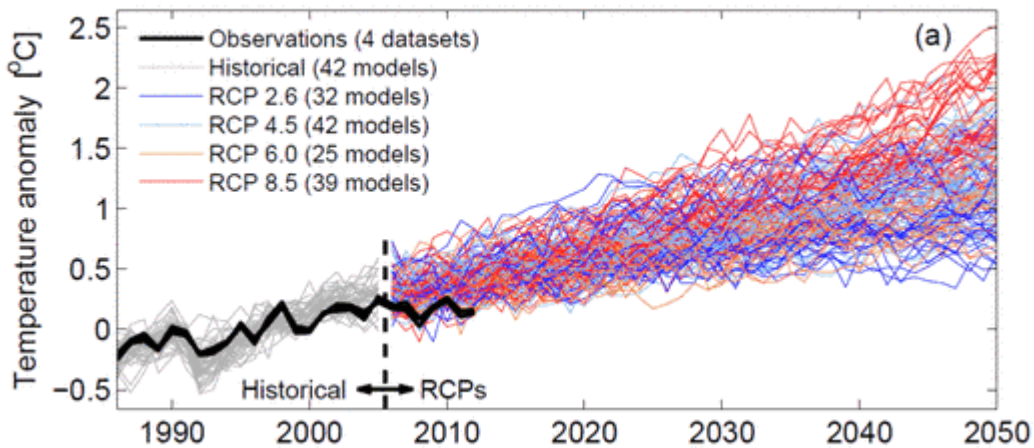


Abbildung 3: IPCC, AR5, WG1, Abbildung 11.25a

Der effektive Ursprung in dieser Graphik lag daher im Jahre **1995**, was die Divergenz zwischen Modellen und Beobachtungen auf etwa die Hälfte der gesamten Divergenz während der Satelliten-Ära hat schrumpfen lassen. Roy Spencer präsentierte das folgende Diagramm, welches die Auswirkung illustriert, wenn man zwei Reihen mit unterschiedlichen Trends um die Mitte des Zeitraumes (oben) und am Beginn des Zeitraumes (unten) einbettet. Falls man die beiden Trendreihen in der Mitte des Zeitraumes einbettet, dann ist die Lücke am Ende des Zeitraumes auf die Hälfte reduziert im Vergleich zur Einbettung am Startzeitpunkt, wenn beide Reihen den gleichen Ursprung haben (wie im Christy-Diagramm).

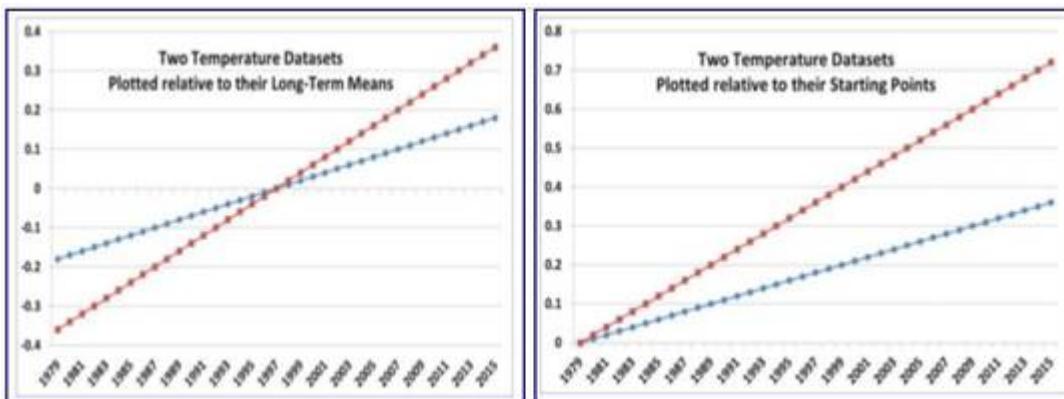


Abbildung 4: Das Diagramm von Roy Spencer, welches die Differenz zeigt zwischen Einbettung am Beginn und in der Mitte.

Bart Verheggen

Das alarmistische Narrativ über vermeintlich ungeeignete Grundlinien in der Abbildung von Christy scheint seinen Ursprung in einem Blogbeitrag von Bart Verheggen zu haben, in welchem eine oben gezeigte frühere Version der Graphik von Roy Spencers Blog (hier) mit der Einbettung von 1979 bis 1983 verunglimpft wurde, obwohl diese Auswahl fast exakt identisch war mit der Einbettung 1979 bis 1984, die später von RSS/Carl Mears verwendet worden ist.

Verheggen beschimpfte derartige Grundlinien als „teilweise betrügerisch“ und warf Christy und Spencer vor, die Modellläufe nach oben zu „verschieben“, um „die Diskrepanz zu vergrößern“:

Sie verschoben die modellierte Temperaturanomalie nach oben, um die Diskrepanz mit den Beobachtungen um etwa 50% zu vergrößern.

Verheggen behauptete, dass die Graphik mit einem Referenzzeitraum von 19867 bis 2005 begonnen hatte (also des im AR5 des IPCC herangezogenen Zeitraumes) und dass Christy und Spencer diesen Zeitraum auf die kürzere Periode von 1979 bis 1983 „umfunktioniert“ hätten, um „die visuelle Erscheinung einer Diskrepanz zu maximieren“:

Der nächste Schritt ist die Umfunktionierung der Abbildung, um die visuelle Präsenz einer Diskrepanz zu maximieren: Man stelle alles auf die Grundlinie des Mittels von 1979 bis 1983 (eine viel zu kurze Periode und, wie es scheint, sehr taktisch ausgewählt) ... Dies sieht überraschend ähnlich aus wie Spencers Trickserie-Graphik.

Verheggen hat auch nicht die Spur eines Beweises geliefert, der zeigte, dass Christy und Spencer die Graphik zuerst mit der in der Mitte des Intervall liegenden Referenzperiode erstellt und die Graphik dann „umfunktioniert“ hatten, um die Menschen zu „täuschen“. Angesichts des Umstandes, dass die Referenzperiode von „**1979 bis 1983**“ klar auf der Y-Achse aufgetragen war, brauchte man auch kaum *reverse engineering* um zu erkennen, dass Christy und Spencer eine Referenzperiode von 1979 bis 1983 herangezogen hatten. Auch sollte es nicht „überraschend“ gekommen sein, dass eine Nachbildung mittels einer Referenzperiode von 1979 bis 1983 ähnlich aussehen würde. Außerdem hat Verheggen die Mears'sche Verwendung eines Referenzzeitraumes von 1979 bis 1984 verdammt, damit die Änderungen „klarer hervortreten“.

Verheggens Vorwürfe finden in der alarmistischen Blog-Gemeinde immer noch Resonanz. Wenige Tage, nachdem Gavin Schmidt Judy Curry herausgefordert hatte, wurde Verheggens Beitrag bei Climate Cocks als die „bislang beste Analyse von John Christys Zaubergrafik“ bezeichnet, die „in der Leugner-Sphäre [deniosphere] so attraktiv ist“.

Die Trickserie geht jedoch vollständig in die umgekehrte Richtung. Graphische Verfahren, die aus einem Ursprung in der Mitte eines Zeitraumes (~1995) resultieren anstatt zum Startzeitpunkt (1979) reduzieren die Diskrepanz zum Ende um etwa 50%, verschleiern also sozusagen die Divergenz.

Gavin Schmidt

Während sich Schmidt darüber beklagte, dass das Christy-Diagramm keine „vernünftige Grundlinie“ habe, hat er keine Kriterien dafür genannt, aufgrund derer eine Grundlinie „vernünftig“ sein sollte und eine andere nicht. Oder dafür, was falsch daran war, einen gemeinsamen Ursprung zu verwenden (oder einen Referenzzeitraum zum Beginn der Satelliten-Ära),

„damit die Änderungen mit der Zeit klarer hervortreten“, wie Mears es getan hatte.

Im März 2016 erzeugte Schmidt seine eigenen Graphiken unter Verwendung zweier unterschiedlicher Grundlinien, um Modelle und Beobachtungen zu vergleichen. Schmidt nahm andere ikonographische Variationen der Graphik vor, aber für diese Analyse sind nur die Referenzperioden von Interesse.

Schmidts erste Graphik (linker Teil der Abbildung unten – leider links und rechts verstümmelt in der Twitter-Version) wurde mit dem folgenden Kommentar vorgestellt:

Hoffentlich endgültige Versionen des Vergleiches von Zeitreihen und Trends zwischen Modellen und Messungen in der tropischen mittleren Troposphäre (bis 2016!)

In dieser Version wurde die Einbettung 1979 bis 1988 verwendet, eine Wahl, die relativ geringe Differenzen im Vergleich zu Christys Einbettung zeigt. Victor Venema hänselte Schmidt wegen seiner Erzeugung von Anomalien, die denen von Christy so ähnlich waren, und fragte sich bzgl. der Referenzperiode:

Haben diese Christy-Anomalien den Bezugszeitraum 1983? Oder ist es Zufall, dass die Beobachtungen zu Beginn so gut passen?
(@ClimateOfGavin)

Schmidt hat seine Graphik sofort neu erstellt, diesmal mit der Einbettung 1979 bis 1998, was die Ähnlichkeit zu den „Christy-Anomalien“ verringerte. Diese Überarbeitung (rechter Teil der Abbildung unten) kündigte er folgendermaßen an:

Die Änderung ist ganz einfach vorzunehmen. Hier folgt das Gleiche für den Zeitraum 1979 bis 1998. Vielleicht ist das besser ...
(@VariabilityBlog)

Nach dieser „Umfunktionierung“ der Graphik, zeigten sich die Beobachtungen jetzt während des gesamten Zeitraumes als innerhalb des Vertrauensintervalls liegend. Es war diese zweite Version, die Schmidt später Curry anbot als das Ergebnis der Auswahl einer „vernünftigeren“ Grundlinie.

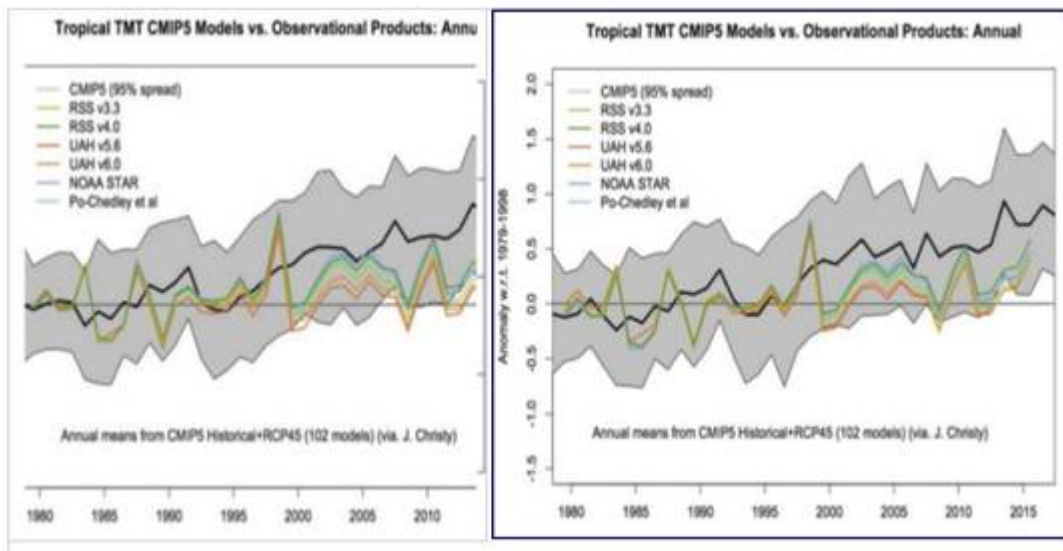


Abbildung 5: Zwei Abbildungen aus den Tweets von Gavin Schmidt vom 4. März 2016. Links: aus dem Tweet vom 4. März mit der Einbettung 1979 bis 1988. Man beachte, dass Teile der Graphik am linken und rechten Rand abgeschnitten worden sind, so dass die Graphik nicht bis zum Jahr 2015 reicht. Rechts: zweite Version mit der Einbettung 1979 bis 1998, was den Modellrahmen relativ zu den Beobachtungen verringerte.

Der Vorgang ist mehr als nur etwas ironisch im Zusammenhang mit Verheggen's früheren Vorwürfen. Er zeigte eine Sequenz von Graphiken einer Grundlinie von 1986 bis 2005 bis zu einer Grundlinie von 1979 bis 1983. Gleichzeitig warf er Spencer und Christy vor, die Grundlinien „umzufunktionieren“, um das „sichtbare Auftauchen einer Diskrepanz zu maximieren“ – was Verheggen „Trickserei“ nannte. Verheggen stellte diese Behauptungen ohne auch nur die Spur eines Beweises auf, dass Christy und Spencer von einer Referenzperiode 1986 bis 2005 ausgegangen waren – was zunächst einmal ein sehr fragwürdiger Zeitraum ist, wenn man Differenzen zeigen will über die Periode 1979 bis 2012, wie Mears erkannt hat. Andererseits hat Schmidt auf Veranlassung von Venema tatsächlich seine Graphik „umfunktioniert“, um das „sichtbare Auftauchen einer Diskrepanz zu reduzieren“.

Noch einmal zur Christy-Graphik

Judy Curry war sehr reserviert darüber, ob Schmidts „Umfunktionieren“ ausreichend war, um den Änderungen aus der Abbildung von Christy Rechnung zu tragen. Sie beobachtete:

Meine Reaktion war, dass diese Plots in keiner Weise wie der Plot von Christy aussahen, und das ist nicht nur eine Sache der Grundlinie.

Zusätzlich zur Änderung der Referenzperiode nahm Schmidt in seiner Graphik noch viele weitere Änderungen vor:

- Schmidt verwendete Jahresdaten anstatt eines 5-Jahre-Mittelwertes.

- Schmidt zeigte eine graue Einhüllende, welche das Vertrauensintervall 5% bis 95% repräsentieren sollte, anstatt die individuellen Spaghetti-Stränge zu zeigen.
- Anstatt 102 Läufe individuell zu zeigen, zeigte Christy Mittel für 32 Modelle. Schmidt scheint 102 Läufe individuell zu zeigen, basierend auf seiner nicht korrekten Referenz auf 102 Modelle (!) in seiner Legende.

Ich bin derzeit dabei, Schmidts Graphik zu reproduzieren. Um den Effekt von Schmidts Umfunktionieren der Christy-Graphik zu isolieren, habe ich die Christy-Graphik so genau wie möglich nachvollzogen (zweites Paneel), wobei ich das meiner Meinung nach Wesentliche berücksichtigt habe, und habe dann die Graphik mit Schmidts Einbettung reproduziert.

Das dritte Paneel isoliert den Effekt von Schmidts Einbettungs-Periode von 1979 bis 1998. Dies bringt sowohl Modelle als auch Beobachtungen nach unten, die Modelle etwas mehr als die Beobachtungen. Allerdings hat sich m. E. der visuelle Effekt gegenüber Christys Einbettung nicht materiell geändert. Dies scheint die Vermutung von Judy Curry zu bestätigen, dass die Änderungen in Schmidts Graphik auf mehr zurückgehen als nur auf eine Änderung der Grundlinie. Eine Möglichkeit ist, dass die Änderung der visuellen Erscheinung Christys Verwendung von Ensemble-Mitteln für jedes Modell geschuldet ist anstatt individueller Läufe. Um dies zu testen, zeigt das vierte Paneel die Christy-Graphik unter Verwendung der Läufe. Und wieder scheint es mir nicht so, dass die ikonographische Entscheidung grundlegend ist für den visuellen Eindruck. Während der Spaghetti-Graph nicht eindeutig ist, kann der INM-CM4-Modelllauf als Blanko-„Kalt“-Modell in allen vier Paneelen unterschieden werden.

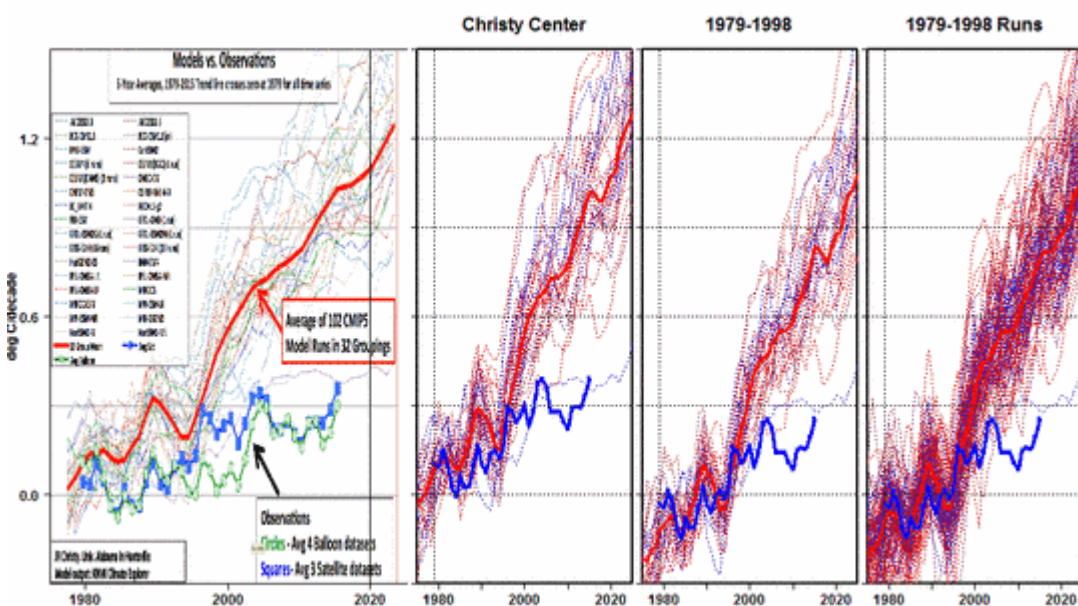


Abbildung 1: Christy-Graphik (links) und Variationen. Siehe Diskussion im Text. Die blaue Linie zeigt das Mittel von UAH 6.0 und RSS 3.3 TLT-tropische Daten.

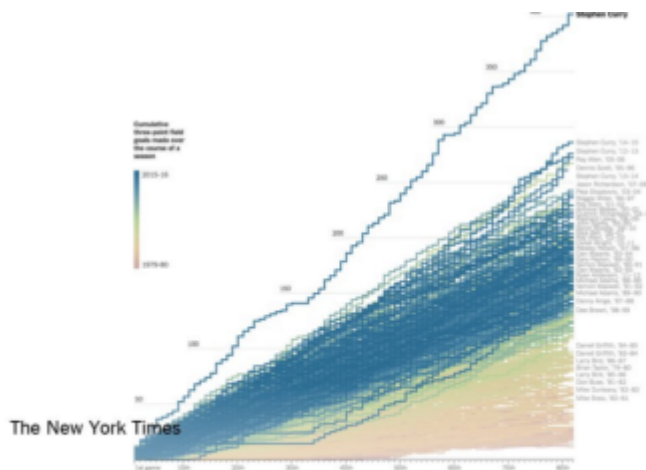
Schlussfolgerungen



Es ist nichts Mysteriöses an der Lücke zwischen Modellen und Beobachtungen zum Ende des Zeitraumes als eine Maßzahl für differierende Trends. Als Secretariat das Feld von 1973 Belmont um 25 Längen hinter sich ließ, haben sogar heutige Klimawissenschaftler nicht in Frage gestellt, dass Secretariat schneller gerannt ist als die anderen Pferde.

Nicht einmal Ben Santer hat in Frage gestellt, ob es eine „statistisch signifikante Differenz“ gegeben habe zwischen Steph Currys epischem *3-point shooting* in 2015-6 und Führern in anderen Jahreszeiten. Kürzlich hat Sports Illustrated der NYT die Lücke zwischen Steph Curry und früheren *3-point leaders* mittels einer Spaghetti-Graphik illustriert (siehe unten), die wie die Christy-Graphik den Vergleich begann mit einem gemeinsamen Ursprung. Die visuelle Kraft kommt von der Separierung zum Ende.

[In diesem Absatz konnte ich einige kursiv gesetzte Begriffe nicht vernünftig übersetzen. Anm. d. Übers.]



NYT Sports @NYTSports · Apr 16
Before Warriors play today, marvel at @StephenCurry30's shooting. He's the one at the top. nyti.ms/1qOL8X2

Falls NYT Sports die Reihe in die Mitte der Jahreszeit eingebettet hätte (nach Art von Bart Verheggen), dann wäre Currys Separierung zum Ende der Jahreszeit nur halb so groß. Falls NYT Sports die Reihe in der ersten Hälfte eingebettet hätte, wäre das Ende der Saison genauso reduziert. Offensichtlich sollte man derartige Versuche, die Separierung zu verringern, als lächerlich zurückweisen.

Es gibt eine wirkliche Diskrepanz zwischen Modellen und Beobachtungen in der tropischen Troposphäre. Falls der aktuelle Punkt die Differenz zu den Trends der Satelliten-Ära (seit 1979) ist, dann ist es vollkommen vernünftig, wie Carl Mears beobachtet hat, die Einbettung der Daten in eine frühe Referenzperiode zu verwenden, so wie der von Mears verwendete Zeitraum 1979 bis 1984 oder der von Christy und Spencer verwendete Zeitraum 1979 bis 1983, damit (Mears) „die Änderungen klarer hervortreten“.

Legt man die Worte von Schmidt aus, heißt das: Wenn man irgendetwas anderes macht, wird es in „verstecken“ hinauslaufen und in „Minimierung von Differenzen, um politisch zu punkten“, was nach Schmidts Worten „einen Partisanen kennzeichnet und nicht einen Wissenschaftler“.

Link:

<https://climateaudit.org/2016/04/19/gavin-schmidt-and-reference-period-trickery/>

Übersetzt von Chris Frey EIKE