

# Antwort eines Mathematikers auf die BEST Studie

geschrieben von Douglas J. Keenan Und Andere | 31. Oktober 2011

Keenan (siehe Foto rechts) schreibt:

*The Economist* forderte mich auf, vier Forschungsstudien des BEST-Projektes zu kommentieren. Die vier Studien, die nicht veröffentlicht sind, haben folgende Titel:

- Decadal variations in the global atmospheric land temperatures
- Influence of urban heating on the global temperature land average using rural sites identified from MODIS classifications
- Berkeley Earth temperature averaging process
- Earth atmospheric land surface temperature and station quality

Es folgt Einiges der Korrespondenz, die wir hatten (man beachte: meine Kommentare wurden unter Zeitdruck geschrieben und sind nicht überarbeitet).

From: D.J. Keenan

To: Richard Muller [BEST Scientific Director]; Charlotte Wickham [BEST Statistical Scientist]

Cc: James Astill; Elizabeth Muller

Sent: 17 October 2011, 17:16

Subject: BEST papers

Anhang: Roe\_FeedbacksRev\_08.pdf; Cowpertwait & Metcalfe, 2009, sect 2-6-3.pdf; EmailtoDKeenan12Aug2011.pdf

Charlotte und Richard,

James Astill, Redakteur für Energie & Umwelt bei *The Economist*, fragte Liz Muller, ob es in Ordnung gehe, mir die BEST-Studien zu zeigen, und sie stimmte zu. Bislang habe ich zwei der Studien gelesen:

Decadal Variations in the Global Atmospheric Land Temperatures

- Influence of Urban Heating on the Global Temperature Land Average Using Rural Sites Identified from MODIS Classifications

Es folgen dazu einige Bemerkungen.

In der ersten Studie werden verschiedene Reihen verglichen und analysiert. Die Reihen wurden jedoch manchmal durch gleitende Mittelbildung geglättet. Geglättete Reihen können in den meisten statistischen Analysen nicht benutzt werden. Für einige Bemerkungen hierzu, die lediglich ein wenig statistisches Hintergrundwissen erfordern, sei hier auf die Blogbeiträge des Mathematikers (und

Statistikers) Matt Briggs hingewiesen:

Do not smooth times series, you hockey puck!

Do NOT smooth time series before computing forecast skill

Hier folgt ein Auszug daraus (Hervorhebung im Original):

Sofern die Daten nicht mit einem Fehler gemessen werden, **wird man die Reihen niemals, niemals unter keinen Umständen GLÄTTEN!** Falls man sie aus irgendwelchen bizarren Gründen doch glättet, **benutzt man diese geglätteten Reihen absolut NIEMALS als Input für andere Analysen!** Wird doch mit einem Fehler gemessen, kann man versuchen, sie zu modellieren (das heißt zu glätten), um vielleicht den Fehler abzuschätzen, aber selbst in diesen seltenen Fällen muss man eine Schätzung von außerhalb haben, d. h. die Abschätzung darf nicht aufgrund der aktuellen Daten erfolgen.

Falls man in einem Anfall von Schwachsinn die Daten von Zeitreihen *doch* glättet und sie dann *doch* als Eingangswerte für andere Analysen nimmt, **steigt die Wahrscheinlichkeit, sich selbst zum Narren zu halten, dramatisch!** Und zwar weil das Glätten Spurensignale erzeugt – Signale, die für andere analytische Methoden real aussehen.

Es scheint, dass dieses Problem vieles der statistischen Analyse in eurer Studie ungültig macht.

Es gibt noch ein weiteres, größeres Problem bei Ihren Studien. In statistischen Analysen zieht man Folgerungen nicht direkt aus den Daten. Vielmehr wird ein statistisches Modell in die Daten eingepasst, und die Folgerungen ergeben sich aus dem Modell. Wir sehen manchmal Aussagen wie „Die Daten nehmen signifikant zu“, aber das ist eine leere Phrase. Daten können nicht signifikant zunehmen, nur ein Trend in einem statistischen Modell kann das.

Ein statistisches Modell sollte sowohl hinsichtlich der Statistik als auch der zugrunde liegenden Wissenschaft plausibel sein. Die statistische Basis schließt in der Regel ein, das Modell mit anderen plausiblen Modellen oder die beobachteten Werte mit dem vom Modell vorhergesagten Werten zu vergleichen. Die Diskussion auf wissenschaftlicher Basis wird zum größten Teil aus den Texten in der Statistik kommen (weil die Texte die Statistiken erläutern), aber nichtsdestotrotz ist es unabdingbar, dass ein Modell wissenschaftlich plausibel ist. Falls die statistische und die wissenschaftliche Basis eines Modells in einer Analyse oder aus dem Zusammenhang nicht klar hervorgeht, dann sollte man Folgerungen aus dem Modell als unbegründet betrachten.

Das statistische Modell, das in den meisten Analysen klimatologischer Zeitserien angewandt wird, ist eine gerade Linie (normalerweise steigend) mit Rauschen (d. h. Störungen), die AR(1) sind [AR steht hier also NICHT für Assessment Report! A. d. Übers.]. AR(1) steht kurz für

„Autoregression erster Ordnung“, was grob gesagt bedeutet, dass (allein) dieses Jahr eine direkte Auswirkung auf das nächste Jahr hat; ist zum Beispiel dieses Jahr sehr kalt, wird das nächste Jahr eine Tendenz in Richtung kälter als im Mittel aufweisen.

Dieses Modell – eine gerade Linie mit AR(1)-Rauschen – ist das beim IPCC angewandte Modell (siehe AR4: §I.3.A). Ebenso wurde dieses Modell durch das U.S. Climate Change Science Program benutzt (welches dem Kongress Rechenschaft ablegt) in seiner Analyse „Statistische Hinweise in Bezug auf Trends“ (hier). Und so weiter. Ein auf AR(1) basierendes Modell wurde außerdem auf zahlreiche klimatologische Zeitreihen von anderen Größen als der globalen Temperatur angewandt. Zum Beispiel auf die Pazifische Dekadische Oszillation, die in Ihrer Studie untersucht worden ist: Siehe das Begutachtungspapier von Roe (2008) im Anhang.

Obwohl ein auf AR(1) basierendes Modell verbreitet Anwendung fand, enthält es nichtsdestotrotz doch gravierende Probleme. Diese sind eigentlich so grundlegend, dass sie in einigen untergeordneten Texten über Zeitreihen diskutiert werden – zum Beispiel in *Time Series Analysis and Its Applications* (third edition, 2011) by R.H. Shumway & D.S. Stoffer (see Example 2.5; set exercises 3.33 and 5.3 elaborate).

In Australien beauftragte die Regierung die Garnaut Review, über die Klimaänderung zu berichten. Die Garnaut Review befragte Spezialisten für die Analyse von Zeitreihen, um die globalen Temperaturreihen zu analysieren. Unter Berücksichtigung des Berichtes dieser Spezialisten wie z. B. Shumway & Stoffer wurde das auf AR(1) basierende statistische Modell verworfen. Die statistische Analyse zeigt, dass dieses Modell zu einfach gestrickt ist, um mit der Komplexität globaler Temperaturreihen umzugehen.

Außerdem haben einige führende Klimatologen auf wissenschaftlicher Basis vehement darauf hingewiesen, dass das auf AR(1) basierende Modell unrealistisch und zu einfach ist (Foster et al., *GRL*, 2008).

Zusammenfassend muss man sagen, dass die meiste Forschung zur globalen Erwärmung auf einem statistischen Modell basiert, das man nicht verwenden sollte. Dies macht viele Analysen zur globalen Erwärmung ungültig. Ich habe ein op-ed piece im *Wall Street Journal* veröffentlicht, in dem ich diese Dinge erkläre, in einfachem Englisch, in diesem Jahr.

Das größte Zentrum zur Erforschung der globalen Erwärmung in UK ist das Hadley Centre. Dort arbeitet ein Statistiker, Doug McNeall. Nachdem mein op-ed piece erschienen war, gab es zwischen Doug McNeall und mir darüber einen Austausch von e-mails. Eine Kopie von einer seiner Botschaften findet sich im Anhang. Darin stellt er fest, dass das statistische Modell – eine gerade Linie mit AR(1)-Rauschen – „einfach ungeeignet“ ist. (Er glaubt immer noch, dass sich die Welt erwärmt, hauptsächlich aufgrund von Computersimulationen des globalen Klimasystems).

Obwohl man vom AR(1)-basierten Modell inzwischen weiß, dass es ungeeignet ist, weiß niemand, welches statistische Modell man benutzen sollte. Es gab verschiedene Studien hierzu in der wissenschaftlich begutachteten Literatur, in denen mögliche Lösungen vorgeschlagen worden sind, aber bisher hat kein Alternativmodell viel Akzeptanz erfahren.

Als ich vom BEST-Projekt erfuhr, hatte ich den Eindruck, dass es darin um statistische Dinge geht. Also war ich extrem neugierig zu sehen, welches statistische Modell verwendet wurde. Ich hatte angenommen, dass eine grundlegende statistische Expertise in das Projekt eingebracht worden ist, und ich hatte das Vertrauen dass zumindest eine große Verbesserung des AR(1)-basierten Modelles herauskommen würde. Tatsächlich habe ich das in einem Interview mit *The Register* im Juni gesagt.

BEST hat sich aber weder des AR(1)-basierten Modells angenommen; noch hat es ein Modell verwendet, das mit einigen der Komplexitäten rechnet, an denen AR(1) gescheitert ist. Statt dessen hat BEST ein Modell verwendet, das sogar noch einfacher gestrickt ist als AR(1), ein Modell, das im Wesentlichen keine Strukturen in Zeitreihen zulässt. Im Einzelnen vermutet das bei BEST verwendete Modell, dass dieses Jahr keine Auswirkungen auf das nächste Jahr hat. Klimatologisch ist diese Vermutung ganz klar ungültig. Wie man einfach sehen kann, gilt es auch nicht in statistischer Hinsicht. Also sind die Schlussfolgerungen der beim BEST durchgeführten statistischen Analyse ohne Grundlage.

All dies geschah, obwohl zum Verständnis der entscheidenden Frage – welches statistische Modell sollte man verwenden? – lediglich ein einführendes Niveau zum Verstehen von Zeitreihen erforderlich ist. Die Frage ist so grundlegend, dass sie in der Einführung bei Shumway & Stoffer enthalten ist, siehe obigen Auszug. Ein weiterer Text, der das auf ähnliche Weise versucht, ist die *Introductory Time Series with R* von P.S.P. Cowpertwait & A.V. Metcalfe (2009); ein Abschnitt dieses Textes ist im Anhang. (Der Abschnitt sagt aus, dass aus statistischer Perspektive ein reines AR(4)-Modell für globale Temperaturen geeignet ist). Weder Shumway & Stoffer noch Cowpertwait & Metcalfe haben meines Wissens eine Agenda zur globalen Erwärmung. Vielmehr schreiben sie einführende Texte über Zeitreihen und geben Studenten praktische Beispiele; jeder Text schließt eine Reihe globaler Temperaturen als ein Beispiel ein.

Es gibt auch Bücher, die sich der statistischen Analyse von Klimadaten verschrieben haben und in denen das Modellieren von Zeitreihen im Detail beschrieben wird. Mein Bücherschrank enthält diese hier:

*Climate Time Series Analysis* (Mudelsee, 2010)

*Statistical Analysis in Climate Research* (von Storch & Zwiers, 2003)

*Statistical Methods in the Atmospheric Sciences* (Wilks, 2005)

*Univariate Time Series in Geosciences* (Gilgen, 2006)

Betrachtet man die zweite Studie über den Wärmeinseleffekt lautet die Schlussfolgerung, dass es eine gewisse städtische *Abkühlung* gegeben hat. Diese Schlussfolgerung widerspricht einem Jahrhundert Forschung ebenso wie der allgemeinen Erfahrung. Sie ist fast mit Sicherheit falsch. Und falls eine solche unerwartete Schlussfolgerung korrekt ist, dann sollten alle denkbaren Anstrengungen unternommen werden, um dem Leser zu zeigen, dass sie richtig sein muss.

Ich schlage eine alternative Erklärung vor. Machen Sie sich zunächst klar, dass in Ihrer Analyse als „sehr ländlich“ beschriebene Stationen tatsächlich einfach „Orte sind, die nicht durch eine bebaute Umgebung dominiert werden“. Mit anderen Worten, es könnte auch – und vermutlich ist das auch so – eine substantielle Urbanisation an jenen Stationen geben. Zweitens, nehmen Sie zur Kenntnis, dass Roy Spencer Beweise präsentiert hat, dass die Auswirkungen der Urbanisation hinsichtlich der Temperatur exponentiell mit steigender Bevölkerungszahl wachsen (hier)

Fasst man diese beiden Dinge zusammen, können wir erwarten, dass der UHI-Effekt an „sehr ländlich“ klassifizierten Stationen größer ist als an „städtisch“ klassifizierten. Und das ist tatsächlich das, was Ihre Analyse zeigt. Natürlich, sollte diese alternative Erklärung korrekt sein, können wir keinerlei Aussagen über die Größe des UHI-Effektes auf die Messung der mittleren Temperatur machen, wenn man die Annäherung in Ihrer Studie benutzt.

Es gibt noch andere, kleinere Probleme mit ihrer Studie. Im Einzelnen stellt der Abschnitt mit der Diskussion Folgendes fest:

Wir beobachten das Gegenteil eines städtischen Erwärmungseffektes während der Zeitspanne von 1950 bis 2010 mit einer Neigung von  $-0.19 \pm 0.19$  °C/100 Jahre. Dies ist statistisch nicht konsistent mit früheren Schätzungen, aber es zeigt, dass der Effekt sehr klein ist...

Falls die beiden Schätzungen nicht konsistent sind, widersprechen sie einander. Mit anderen Worten, zumindest eine von ihnen muss falsch sein. Also kann die eine Schätzung nicht benutzt werden, um eine Folgerung aus der anderen zu „verifizieren“. Das hat nichts mit Statistik zu tun. Es ist logisch!

Sincerely, Doug

\* \* \* \* \*

Douglas J. Keenan

<http://www.informath.org>

---

From: Richard Muller  
To: James Astill  
Cc: Elizabeth Muller  
Sent: 17 October 2011, 23:33

Subject: Re: BEST papers

Hallo James,

Sie haben die Kopie einer e-mail empfangen, die DJ Keenan an mich und Charlotte geschrieben hat. Er spricht darin eine Menge Dinge an, die eine Antwort erfordern. Einige spiegeln Missverständnisse, andere spiegeln einfach Meinungsverschiedenheiten unter Experten im Bereich der Statistik. Da diese Dinge wieder und immer wieder aufkommen dürften, bereiten wir einen Menüpunkt FAQ vor, den wir auf unserer Website platzieren.

Keenan sagt, dass er unsere lange Studie zu statistischen Methoden noch nicht gelesen hat. Ich denke, wenn er sie liest, wird er mehr bereit sein, die Erfahrung und die Sorgfalt anzuerkennen, die wir für diese Analyse aufgewendet haben. David Brillinger, unser leitender Berater zu Statistiken, warnte uns davor, dass wir bei Vermeidung des Jargons der Statistik andere Statistiker in die Irre führen können, indem sie denken könnten, wir hätten eine naive Annäherung durchgeführt. Aber wir haben uns entschlossen, in einem beiläufigen Stil zu schreiben, speziell um es uns zu ermöglichen die größere Welt der Geophysiker und Klimawissenschaftler zu erreichen, die den Jargon nicht verstehen. Falls Keenan die Methoden der Studie liest, wird er mehr anerkennen, was wir getan haben.

Auch ist es wichtig zu erkennen, dass wir kein neues Feld der Wissenschaft erzeugen, sondern etwas zu einem solchen Feld hinzufügen, das eine lange Geschichte hat. In der Vergangenheit habe ich Folgendes entdeckt: wenn man es vermeidet, Methoden aus der Vergangenheit anzuwenden, verstehen die Schlüsselwissenschaftler im Feld nicht, was man gemacht hat. Als mein liebstes Beispiel zitiere ich eine Studie von mir, in der ich geschrieben habe, dass ich Daten genutzt habe, die ungleich in der Zeit verteilt waren; also führte ich ein Lomb periodogramm [?] durch. Die Studie wurde durch Schiedsrichter zurückgewiesen, die argumentierten, dass ich eine „obskure“ Annäherung benutzt habe und einfach die traditionelle Interpolation hätte durchführen sollen, gefolgt von einer Blackman-Tukey-Analyse. In Zukunft werde ich es so machen, immer mit der Vorsicht, auch eine Lomb-Analyse durchzuführen, um sicherzustellen, dass sich keine Unterschiede ergeben.

In seinem Eingangskommentar geht es um das Glätten von Daten. Es gibt sicherlich Statistiker, die diesem Vorgehen vehement widersprechen, aber es gab auch Spitzenstatistiker, die es unterstützten. Hierzu gehören David Brillinger und sein Mentor, der große John Tukey. Tukey revolutioniert das Feld der Datenanalyse für die Wissenschaft, und seine Methoden dominieren viele Bereiche der physikalischen Wissenschaft.

Tukey argumentierte, dass dieses Glätten eine Version von "Vorab-Weißwäsche" war, ein nützliches Verfahren, um reale, aber uninteressante Eigenheiten aus den Daten zu entfernen. Er zeigte, dass diese Vorab-

Weißwäsche zu einem robusteren Ergebnis führt. Dies ist im Endeffekt das, was er in seiner Studie zu dekadischen Variationen gemacht hatte. Die langen Änderungen der Zeiträume standen nicht im Brennpunkt unserer Studie, also haben wir eine Anpassung der größten Wahrscheinlichkeit durchgeführt, sie entfernt und dann die übrigen Daten untersucht.

Zitat von Keenan: „Falls man in einem Anfall von Schwachsinn die Daten von Zeitreihen *doch* glättet und sie dann *doch* als Eingangswerte für andere Analysen nimmt, **steigt die Wahrscheinlichkeit, sich selbst zum Narren zu halten, dramatisch!** Und zwar weil das Glätten Spurensignale erzeugt – Signale, die für andere analytische Methoden real aussehen.“ Dann zieht er ein Fazit, das nicht aus diesem Zitat abgeleitet ist; er sagt: „Dieses Problem scheint Vieles der statistischen Analyse in Ihrer Studie ungültig zu machen“.

Natürlich ist das unlogisch. Nur weil die Glättung die Wahrscheinlichkeit zunehmen lässt, dass wir uns selbst betrügen, heißt das noch lange nicht, dass wir das auch tun. Es gibt einen echten Wert geglätteter Daten, und ja, man muss sich vor den Fallen in Acht nehmen. Wenn Sie das tun, gibt es einen echten Vorteil der Glättung. In meinem Buch zum Thema “Ice Ages and Astronomical Causes” habe ich das detailliert beschrieben. Der größte Teil dieses Buches handelt von den Fallen und Fallgruben, in die andere bereits gefallen sind.

Keenan fährt fort: „In statistischen Analysen werden keine Folgerungen aus den Daten direkt abgeleitet. Vielmehr wird ein statistisches Modell an die Daten angepasst, und Folgerungen werden aus dem Modell gezogen“. Dem stimme ich von ganzem Herzen zu! Er könnte durcheinander gekommen sein, weil wir die Sprache der Physik und Geophysik benutzt haben und nicht die der Statistik. Er sagt weiter: „Dies macht vieles von den Analysen zur globalen Erwärmung ungültig.“ Wenn wir weiterkommen wollen, ist es nicht gut, einfach das meiste der früheren Arbeiten schlecht zu machen. Also haben wir sorgfältiger gearbeitet und gültige statistische Methoden angewandt, aber unsere Studie schreiben wir so, dass frühere Bearbeiter auf diesem Gebiet verstehen, was wir sagen. Unsere Hoffnung ist es teilweise, die Methoden allgemein voranzubringen.

Unglücklicherweise lautet Keenans Schlussfolgerung, dass es praktisch keine gültigen Arbeiten im Bereich Klima gegeben hatte, dass man ein besseres Modell braucht, ohne dass er aber weiß, wie dieses Modell aussehen könnte. Er sagt: „Zusammenfassend muss man sagen, dass die meiste Forschung zur globalen Erwärmung auf einem statistischen Modell basiert, das man nicht verwenden sollte. Dies macht viele Analysen zur globalen Erwärmung ungültig. Ich habe ein op-ed piece im *Wall Street Journal* veröffentlicht, in dem ich diese Dinge erkläre, in einfachem Englisch, in diesem Jahr.“

Hier folgt dieses Zitat, in dem er im Grunde ausdrückt, dass keine Analyse bzgl. der globalen Erwärmung unter seinen statistischen Standards gültig ist: „Obwohl man vom AR(1)-basierten Modell inzwischen

weiß, dass es ungeeignet ist, weiß niemand, welches statistische Modell man benutzen sollte. Es gab verschiedene Studien hierzu in der wissenschaftlich begutachteten Literatur, in denen mögliche Lösungen vorgeschlagen worden sind, aber bisher hat kein Alternativmodell viel Akzeptanz erfahren.“

Was er sagt ist, dass statistische Modelle nicht geeignet sind, um zu zeigen, ob es eine globale Erwärmung oder Abkühlung oder was auch immer gibt. Das ist eine sehr starke Schlussfolgerung, und sie spiegelt meiner Meinung nach seine übertriebene Pedanterie statistischer Methoden. Er kann und wird jede Studie in der Vergangenheit und in der Zukunft aus den gleichen Gründen kritisieren. Wir könnten genauso gut unsere Versuche aufgeben, die globale Erwärmung zu evaluieren, bis wir ein „Modell“ finden, dem Keenan zustimmt – aber er hilft uns nicht dabei, ein solches Modell zu finden.

Tatsächlich zeigt ein schneller Blick auf seine Website, dass es in der Liste seiner Veröffentlichungen fast ausschließlich um Analysen geht, die zeigen, dass die anderen Studien falsch sind. Ich habe den starken Verdacht, dass Keenan jedes Modell ablehnt, das wir benutzt haben.

Er gibt einigen Klagen klaren Ausdruck. Er zitiert aus unserer Studie, wo wir sagen: „Wir beobachten das Gegenteil eines städtischen Erwärmungseffektes während der Zeitspanne von 1950 bis 2010 mit einer Neigung von  $-0.19 \pm 0.19$  °C/100 Jahre. Dies ist statistisch nicht konsistent mit früheren Schätzungen, aber es zeigt, dass der Effekt sehr klein ist...“

Dann beklagt er sich: Wenn die beiden Schätzungen nicht konsistent sind, widersprechen sie einander. Mit anderen Worten, mindestens eine von ihnen muss falsch sein. Also kann es nicht benutzt werden, um die Ergebnisse des anderen Modells zu „verifizieren“. Das hat nichts mit Statistik zu tun. Das ist logisch!

Er interpretiert unsere Aussage falsch. Unsere Schlussfolgerung basiert auf unserer Analyse. Wir glauben, dass diese korrekt ist. Die Tatsache, dass sie mit früheren Schätzungen inkonsistent ist, schließt ein, dass eine davon falsch ist. Natürlich glauben wir dies von den früheren Schätzungen. Wir glauben nicht, dass die früheren Schätzungen mehr waren als „Rateschätzungen“ [unübersetzbares Wortspiel: statt „estimates“ steht im Original „guestimates“. A. d. Übers.], und darum gibt es keinen „statistischen“ Widerspruch.

Er beklagt sich: Betrachtet man die zweite Studie über den Wärmeinseleffekt lautet die Schlussfolgerung, dass es eine gewisse städtische *Abkühlung* gegeben hat. Diese Schlussfolgerung widerspricht einem Jahrhundert Forschung ebenso wie der allgemeinen Erfahrung. Sie ist fast mit Sicherheit falsch. Und falls eine solche unerwartete Schlussfolgerung korrekt ist, dann sollten alle denkbaren Anstrengungen unternommen werden, um dem Leser zu zeigen, dass sie richtig sein muss.



Er zieht eine starke Schlussfolgerung aus einem Effekt, der nur bei einer Standardabweichung signifikant ist. Er hätte uns niemals behaupten lassen, dass  $-0,19 \pm 0,19^\circ\text{C}$  pro Jahrhundert auf eine städtische Abkühlung hinweist. Ich bin überrascht, dass ein Statistiker argumentiert, dass ein statistisch so insignifikanter Effekt auf Abkühlung hinweist.

Bitte seien Sie vorsichtig, wem Sie diese e-mail zeigen. Wir sind wirklich daran interessiert, besser zu sein als alle anderen Analysten in diesem Bereich, und ich befürchte, wenn Teile dieser Mail aus dem Zusammenhang gerissen gelesen werden, dass man diese falsch interpretieren könnte.

Rich

---

From: D.J. Keenan  
To: James Astill  
Sent: 18 October, 2011 17:53  
Subject: Re: BEST papers

James,

Um auf den Kern der Sache zu kommen, es sieht so aus, als ob Rich [Richard Mullen] und ich in Übereinstimmung sind. Hier folgt ein Auszug aus seiner Antwort.

Keenan: „In statistischen Analysen werden keine Folgerungen aus den Daten direkt abgeleitet. Vielmehr wird ein statistisches Modell an die Daten angepasst, und Folgerungen werden aus dem Modell gezogen“. Dem stimme ich von ganzem Herzen zu!

Und daher lautet die Frage: War das statistische Modell, das für die Analyse benutzt worden ist, eine gute Wahl? Falls nicht – da ihre Schlussfolgerungen auf diesem Modell basieren – müssen ihre Schlussfolgerungen unbegründet sein.

Tatsächlich wurde das von ihnen verwendete statistische Modell von nahezu jedem abgelehnt. Im Besonderen wurde es sowohl vom IPCC als auch vom CCSP abgelehnt, wie ich schon vorher sagte. Ich kenne keine Arbeit, die ihre Wahl des Modells unterstützt: sie haben das Modell einfach übernommen, ohne zu versuchen zu ergründen, ob es gerechtfertigt ist, was mit Sicherheit falsch ist.

(Es ist seit Jahrzehnten bekannt, dass das statistische Modell, das sie verwendet haben, nicht benutzt werden sollte. Obwohl die statistischen Probleme mit diesem Modell seit langem eindeutig waren, kannte niemand den physikalischen Grund dafür. Dann hat Klaus Hasselmann im Jahre 1976 eine Studie veröffentlicht, in der er den Grund erklärt. Die Studie ist berühmt und wurde seitdem über 1000 mal zitiert.)

Man kann darüber streiten, welches statistische Modell man benutzen kann. Gewiss ist jedoch, dass das Modell, das BEST benutzt hat, abgelehnt werden sollte. Daraus folgt, dass die Schlussfolgerungen unbegründet sind.

Hinsichtlich der Glättung erfordert die Situation hier nur wenig Statistik, um sie zu verstehen. Man betrachte das Beispiel, das Matt Briggs hier vorgestellt hat. Wir nehmen zwei Reihen, jede vollständig abgeschlossen. Wir berechnen die Korrelation der beiden Reihen: Das wird zu einer Korrelation um 0 tendieren. Dann glätten wir beide Reihen und berechnen die Korrelation der beiden geglätteten Reihen: Sie wird größer sein als zuvor. Je mehr wir die beiden Reihen glätten, umso größer wird die Korrelation. Und doch haben wir mit zwei vollständig abgeschlossenen Reihen begonnen. Das ist keine Meinungssache, es ist faktisch. Und doch berechnet BEST die Korrelation geglätteter Reihen.

Die Antwort nutzt rhetorische Techniken, um zu vermeiden (Zitat) „Nur weil die Glättung die Wahrscheinlichkeit zuzunehmen lässt, dass wir uns selbst betrügen, heißt noch lange nicht, dass wir das auch tun.“ Diese Feststellung stimmt, stellt aber nicht den obigen Punkt in Frage.

Hinsichtlich der Studie zum UHI enthielt meine Botschaft auch das Folgende.

Es gibt andere, kleinere Probleme mit Ihrer Studie. Im Einzelnen wird im Abschnitt über die Diskussion das Folgende gesagt.

Wir beobachten das Gegenteil eines städtischen Erwärmungseffektes während der Zeitspanne von 1950 bis 2010 mit einer Neigung von  $-0.19 \pm 0.19$  °C/100 Jahre. Dies ist statistisch nicht konsistent mit früheren Schätzungen, aber es zeigt, dass der Effekt sehr klein ist...

Wenn die zwei Schätzungen nicht konsistent sind, dann widersprechen sie einander. Mit anderen Worten, mindestens eines von ihnen muss falsch sein. Also kann man nicht eine Schätzung benutzen, um eine Folgerung aus der anderen zu „verifizieren“. Das hat nichts mit Statistik zu tun. Es ist logisch!

In der Antwort wird behauptet: „Die Tatsache, dass (die Schlussfolgerung der Studie) inkonsistent mit früheren Schätzungen ist schließt ein, dass eine falsch ist“. Die Behauptung ist offensichtlich absurd.

In der Antwort werde ich auch dafür kritisiert, dass ich „eine starke Schlussfolgerung aus einem Effekt gezogen habe, der nur für eine Standardabweichung signifikant ist“. Diese Schlussfolgerung habe ich nicht gezogen, sondern in ihrer Studie wurde mir das in den Mund gelegt, und zwar indem sie sagt, dass der Effekt „im Gegensatz zum Erwarteten steht, wenn der UHI anomale Erwärmung hinzufügen würde“, und dass „natürliche Erklärungen irgendeine aktuelle Form der ‚urbanen Abkühlung‘ erforderlich machen“, um dann mögliche Gründe zu beschreiben wie z. B. „Wenn eine Asphaltoberfläche durch Beton ersetzt wird, können wir

erwarten, dass die Absorption von Sonnenwärme abnimmt, was insgesamt zu einem Abkühlungseffekt führt“.

Man beachte, dass die Antwort nicht die alternative Erklärung erwähnt, die meine Botschaft hinsichtlich der UHI-Ergebnisse vorgeschlagen hatte. Diese Erklärung, die auf einer Analyse von Roy Spencer basiert (den ich in meiner Botschaft erwähnt hatte), impliziert, dass wir keine Folgerung über die Größe des UHI-Effektes auf die Messung der mittleren Temperatur ableiten können, wenn man die Annäherung in ihrer Studie zur Hand nimmt.

Ich habe rasch die Methoden ihrer Studie überflogen. Sie betrifft keinen meiner Kritikpunkte.

Rich nennt auch sein Buch über die Gründe der Eiszeiten. Lesen Sie bitte mein op-ed piece im *Wall Street Journal*, und beachten Sie vor allem die Diskussion zu den Abbildungen 6 und 7. Sein Buch behauptet, die Daten in Abbildung 6 zu analysieren: das Ziel des Buches ist es zu erklären, warum die Ähnlichkeit der zwei Linien so schwach ist. Um den Mechanismus zu verstehen, ist es tatsächlich nur notwendig, eine einfache Subtraktion durchzuführen – wie ich in meinem op-ed piece erkläre. Kurz gesagt, die Analyse in seinem Buch ist außerordentlich inkompetent – und man muss nur Subtraktion verstehen, um das zu sehen.

Die Person, die die Datenanalyse in jenem Buch durchgeführt hat, ist die gleiche Person, die für die Datenanalyse beim BEST verantwortlich ist. Die Datenanalyse in den BEST-Studien würde einen Grundkurs über statistische Zeitreihen im dritten Studienjahr nicht überstehen.

Zu guter letzt, ein Kommentar zu Temperaturlaufzeichnungen am Boden wäre angebracht. Wir haben Satellitenaufzeichnungen für die letzten Jahrzehnte, und die stimmen gut mit den Bodenaufzeichnungen überein. Wir haben auch gute Beweise, dass die Welt vor 100 bis 150 Jahren kälter war als heute. Hauptsächlich aus diesen Gründen glaube ich, dass die Temperaturlaufzeichnungen – von NASA, NOAA, Hadley/CRU und jetzt BEST – möglicherweise im Großen und Ganzen richtig sind.

Cheers, Doug

---

From: James Astill  
To: D.J. Keenan  
Sent: 18 October 2011, 17:57  
Subject: Re: BEST papers

Hallo Doug

Vielen Dank! Heißt das, Sie misstrauen der BEST-Methodik in hohem Maße, stimmen aber ihrer wichtigsten Schlussfolgerung zu, d. h. der Temperaturlaufzeichnung am Boden?

best  
James

James Astill  
Energy & Environment Editor

---

From: D.J. Keenan  
To: James Astill  
Sent: 18 October 2011, 18:41  
Subject: Re: BEST papers

James,

Ja, ich stimme zu, dass die BEST-Temperaturaufzeichnungen sehr wahrscheinlich grob richtig sind, zumindest während der letzten 120 Jahre oder so. Dies gilt für das allgemeine Aussehen ihrer Kurve, nicht aber für ihre Abschätzung der Unsicherheiten.

Cheers, Doug

---

From: D.J. Keenan  
To: James Astill  
Sent: 20 October, 2011 13:11  
Subject: Re: BEST papers

James,

Irgendjemand hat mir gerade die BEST-Presseerklärung zugesandt und mich nach meinem Kommentar dazu gefragt. Die Presseerklärung beginnt mit dem folgenden Statement:

Die globale Erwärmung ist einer heute veröffentlichten grundlegenden Studie zufolge real. Trotz Einwänden von Skeptikern der Klimaänderung findet die BEST-Studie verlässliche Beweise für einen Anstieg der globalen Mitteltemperatur an der Erdoberfläche von etwa 1°C seit Mitte der fünfziger Jahre.

Der zweite Satz könnte zutreffen. Der erste Satz jedoch ist weder implizit im zweiten Satz enthalten, noch folgt er aus den Analysen in den Forschungsunterlagen.

Zu zeigen, dass die „globale Erwärmung real“ ist, erfordert viel mehr als zu zeigen, dass die mittlere globale Temperatur auf dem Festland seit Mitte der fünfziger Jahre um 1°C gestiegen ist. Zur Illustration, die Temperatur 2010 lag höher als 2009, aber das für sich allein ist kein Beweis für die globale Erwärmung: Die Temperaturzunahme könnte offensichtlich auf Änderungen der Randbedingungen zurückzuführen sein.

Ähnlich kann auch die Temperaturzunahme seit den fünfziger Jahren auf Änderungen der Randbedingungen zurückzuführen sein.

Um zu zeigen, dass die Zunahme der Temperaturen seit den fünfziger Jahren nicht auf derartige Fluktuationen zurückzuführen ist, ist eine gültige statistische Analyse der Temperaturen erforderlich. Das BEST-Team hat dies nicht getan.

Ich möchte etwas betonen. Nehmen wir an, jemand sagt „ $2+2=5$ “. Dann ist es nicht bloß meine Meinung, dass das, was dieser Jemand sagt, falsch ist; sondern vielmehr ist es falsch, was er gesagt hat. Ganz ähnlich ist es nicht nur meine Meinung, dass die statistische Analyse vom BEST ernsthaft ungültig ist; sondern vielmehr, die statistische Analyse vom BEST ist ungültig.

Cheers, Doug

---

From: James Astill  
To: D.J. Keenan  
Sent: 20 October 2011, 13:19  
Subject: Re: BEST papers

Hallo Doug

Vielen Dank für Ihre Gedanken dazu. Es wird interessant sein zu beobachten, wie die BEST-Studien durch den Prozess der Begutachtung kommen. Bitte halten Sie mich auf dem Laufenden!

best

james  
James Astill  
Energy & Environment Editor

---

Ein Artikel über das BEST wurde in der Ausgabe von *The Economist* am 22. Oktober veröffentlicht. Der Artikel, autorisiert von James Astill, erwähnt die oben genannten Punkte nicht. Die Unterschlagzeile lautet: „eine neue Studie zu Temperaturaufzeichnungen lässt wenig Raum für Zweifler. Die Welt erwärmt sich.“ Der Einführungssatz lautet: „Für jene, die noch in Frage stellen, ob die Welt sich wirklich erwärmt, ist es nötig zu glauben, dass die instrumentellen Aufzeichnungen falsch sind“.

*Douglas J. Keenan*

Übersetzt von Chris Frey für EIKE