

# Unsicherheiten bei Messungen der Wassertemperatur und in Datensätzen

geschrieben von Judith Curry | 21. November 2013

Die erste Studie von John Kennedy vom UK Met Office enthält eine kompakte und dringend nötige Ungenauigkeits-Analyse der Messungen und Analysen von Wassertemperaturen SST:

## **Eine Übersicht über in-situ-Messungen und Datensätze der Wassertemperatur**

John Kennedy

**Abstract:** *Archive mit In-Situ-Messungen der Wassertemperatur (SST) reichen über 160 Jahre zurück. Die Qualität der Messungen ist variabel und die Fläche der vermessenen Ozeane begrenzt, vor allem zu Beginn der Reihe und während der beiden Weltkriege.*

*SST-Messungen und darauf basierende gerasterte Datensätze werden in vielen Untersuchungen verwendet, so dass das Verständnis für und die Abschätzung der Unsicherheiten vital ist. Ziel dieser Übersicht ist es, die verschiedenen Komponenten aufzuzeigen, die zu der Gesamt-Unsicherheit der SST-Messungen beitragen sowie der daraus abgeleiteten Datensätze. Dabei bezweckt sie auch, gegenwärtige Lücken des Verständnisses zu identifizieren. Unsicherheiten entstehen auf der Ebene individueller Messungen sowohl durch systematische als auch durch Zufallseffekte. Obwohl diese eingehend untersucht worden sind, schreitet die Verbesserung der Fehlermodelle voran. Jüngste Verbesserungen gab es hinsichtlich des Verständnisses allgegenwärtiger systematischer Fehler, die die Abschätzung langzeitlicher Trends und der Variabilität betreffen. Allerdings sind die Adjustierungen zur Minimierung dieser systematischen Fehler unsicher, und diese Unsicherheiten sind vor den siebziger Jahren größer und vor allem in den Zeiträumen um den Zweiten Weltkrieg, und zwar infolge des Fehlens zuverlässiger Metadaten. Die Unsicherheiten in Zusammenhang mit der Auswahl der statistischen Methoden, mit denen man global vollständige SST-Datensätze erzeugen möchte, wurden mit Hilfe unterschiedlicher Analysemethoden erkundet, aber sie enthalten nicht die jüngsten Einsichten von Messfehlern, und sie verlangen nach einem ordentlichen Eckpfeiler zum Vergleich, mit dem ihre Eignung objektiv bestimmt werden kann. Diesen Problemen kann man begegnen durch die Erzeugung neuer SST-Analysen [new end-to-end SST analyses] und durch die Gewinnung und Digitalisierung von Daten und Metadaten aus Schiffslogbüchern und anderer zeitnaher Unterlagen.*

Veröffentlicht in *Reviews of Geophysics*, [Link zum Abstract](#) und zum [ganzen Manuskript](#).

Auszüge:

Bei der Verwendung der SST-Beobachtungen und der darauf basierenden Analysen ist es wichtig, die darin enthaltenen Unsicherheiten sowie die Hypothesen und statistischen Methoden zu verstehen, die zur Erzeugung beigetragen haben. Mit dieser Übersicht beabsichtige ich, die verschiedenen Komponenten aufzuzeigen, die zu der Gesamtunsicherheit der SST-Messungen und der daraus abgeleiteten Datensätze beitragen. Auch Wissenslücken möchte ich damit schließen.

Abschnitt 2 beschreibt eine Klassifizierung der Unsicherheiten. Die Einteilungen nach Klassen sind nicht definitiv und auch nicht ganz eindeutig. Allerdings spiegeln sie die Art und Weise, mit der man sich dieser Unsicherheiten in der Literatur angenommen hat. Und sie bieten einen nützlichen Rahmen, an Unsicherheiten in den SST-Datensätzen zu denken. ...

Während dieser gesamten Studie wird unterschieden zwischen einem Fehler und einer Unsicherheit. Der Fehler in einer Messung ist der Unterschied zwischen einem idealisierten „wahren Wert“ und dem gemessenen Wert. Er ist nicht bestimmbar. Die Unsicherheit einer Messung ist definiert als der „Parameter zusammen mit dem Ergebnis der Messung, der die Verteilung der Werte charakterisiert, die man vernünftigerweise der Messgröße zuordnen kann“. In diesem Sinne ist Unsicherheit in der folgenden Diskussion gemeint. Das ist nicht notwendigerweise so wie in den zitierten Studien. Allgemein wird das Wort Fehler als Synonym für Unsicherheit gesehen, wie es in den üblichen Phrasen Standardfehler und Analysefehler zum Ausdruck kommt.

Grob gesagt wurden Fehler in individuellen SST-Beobachtungen aufgeteilt in zwei Gruppen: beobachtete zufällige Fehler [random observational errors] und systematische Fehler. Obwohl dies ein bequemer Weg ist, mit den Unsicherheiten umzugehen, werden sich allgemein Fehler in den SST-Messungen ein wenig durch gleiche Charakteristiken auszeichnen.

Zufällige Fehler treten aus vielen Gründen auf: falsche Ablesung des Thermometers, Rundungsfehler, die Schwierigkeit, ein Thermometer mit größerer Präzision als der vorgegebenen Skaleneinteilung zu lesen, falsch aufgezeichnete Werte, Fehler bei der Digitalisierung und Rauschen des Messfühlers. Obwohl all diese Fehler eine Einzelmessung beeinträchtigen, tendieren sie dazu, sich gegenseitig aufzuheben, wenn große Datenmengen miteinander verglichen werden. Daher ist der Beitrag von beobachteten Randwert-Fehlern zu den Unsicherheiten der globalen mittleren SST viel kleiner als der Beitrag eines Randwertfehlers zur Unsicherheit einer Einzelbeobachtung, selbst in Jahren mit geringster Anzahl von Messungen. Dennoch können, wenn nur wenige Beobachtungen vorliegen, beobachtete Randwert-Fehler einen wichtigen Beitrag zur Gesamtunsicherheit leisten.

Systematische Fehler sind viel problematischer, weil deren Auswirkungen

relativ ausgeprägter sind, wenn größere Datenmengen zusammengeführt werden. Systematische Fehler können auftreten, weil ein bestimmtes Thermometer falsch kalibriert ist oder an einer ungeeigneten Stelle steht. Egal wie viele Daten man mittelt von einem Thermometer, das wegen falscher Kalibrierung stets 1 K zu viel anzeigt, wird den Fehler nicht unter dieses Niveau reduzieren, es sei denn durch Zufall. Allerdings wird der systematische Fehler in vielen Fällen von der Umgebung des Thermometers im Einzelnen abhängen und daher von Schiff zu Schiff verschieden sein. In diesem Falle wird die Mittelbildung von Messungen vieler verschiedener Schiffe oder Bojen dazu tendieren, den Beitrag systematischer Fehler zur Unsicherheit des Mittels zu reduzieren.

Kennedy et al. (2011b) haben zwei Formen der Unsicherheit betrachtet: die grid-box sampling-Unsicherheit und large scale sampling-Unsicherheit.\* Grid-box sampling-Unsicherheit bezieht sich auf die Unsicherheit der Schätzung der mittleren Wassertemperatur in einem Gebiet innerhalb einer Gitterbox mit einer endlichen und oftmals kleinen Anzahl von Messungen. Large scale Sampling-Unsicherheit bezieht sich auf die Unsicherheit, die aus der Abschätzung der Wassertemperatur eines größeren Gebietes resultieren mit vielen Gitterboxen ohne eine Messung. Obwohl diese beiden Arten von Unsicherheit eng zusammenhängen, ist es oft einfacher, die Grid-box sampling-Unsicherheit abzuschätzen, wo man es nur mit der Variabilität innerhalb einer Gitterbox zu tun hat. Bei der großräumigen Sampling-Unsicherheit muss man das große Spektrum der Variabilität im globalen Maßstab berücksichtigen.

Im Zusammenhang mit der Unsicherheit der Wassertemperatur-[Daten] sind unbekannte Unbekannte die Dinge, die man leicht übersieht. Ihrem Wesen nach sind unbekannte Unbekannte nicht quantifizierbar; sie repräsentieren die tieferen Unsicherheiten, die allen wissenschaftlichen Unternehmungen anhaften. Mit dieser Übersicht hoffe ich zeigen zu können, dass die Bandbreite für Revolutionen in unserem Verständnis begrenzt ist. Nichtsdestotrotz kann eine Verbesserung durch eine kontinuierliche Entwicklung unseres Verständnisses nur erfolgen, wenn wir akzeptieren, dass unser Verständnis unvollständig ist. Unbekannte Unbekannte werden nur mit kontinuierlichen, sorgfältigen und manchmal erfinderischen Untersuchungen der Daten und Metadaten ans Tageslicht kommen.

**Kommentar von Judith Curry:** Das Uncertain T. Monster [das ist hier als „Name“ zu verstehen] ist SEHR erfreut über diese umfassende Diskussion der Unsicherheiten. Die größten Herausforderungen (in der Studie ausführlich beschrieben) sind die Annahme struktureller Unsicherheiten der Analysemethoden und die Kombination aller Unsicherheiten. Jedwede Applikation dieser Daten (einschließlich der Trendanalyse) erfordert die Berücksichtigung dieser Dinge.

In der zweiten Studie wird versucht, das Unsicherheits-Monster zu schlachten.

**Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends**

[etwa: Betrachtung des Bias' in den Temperaturreihen von HadCRUT4 und

dessen Auswirkungen auf jüngste Temperaturtrends]

Kevin Cowtan und Robert Way

**Abstract:** Eine unvollständige globale Abdeckung ist eine potentielle Bias-Quelle bei globalen Temperatur-Rekonstruktionen, falls die nicht mit Daten erfassten Gebiete nicht gleichmäßig auf der Planetenoberfläche verteilt sind. Der meistverwendete HadCRUT4-Datensatz erfasst im Mittel 84% des Globus' während der letzten Jahrzehnte, wobei sich die datenfreien Gebiete um die Pole und in Afrika konzentrieren. Drei bestehende Rekonstruktionen mit einer fast globalen Überdeckung werden untersucht, wobei jede zeigt, dass HadCRUT4 anfällig für Bias ist infolge der Behandlung datenfreier Gebiete. Zwei alternative Vorgehensweisen, die globale Temperatur zu rekonstruieren, werden untersucht. Eine davon basiert auf einem optimalen Interpolations-Algorithmus und die andere auf einer Hybrid-Methode [?] mit zusätzlichen Informationen aus Satellitendaten. Die Methoden werden validiert auf Basis ihrer Eignung, fehlende Daten zu rekonstruieren. Beide Methoden führen zu besseren Ergebnissen als wenn man die datenfreien Regionen ausschließt, wobei vor allem die Hybrid-Methode gut geeignet ist für diese datenfreien Gebiete. Temperaturtrends der Hybrid-Methode und der Rohdaten von HadCRUT4 werden verglichen. Der weithin genannte Trend seit 1997 ist nach der Hybrid-Methode zweieinhalb mal größer als der korrespondierende Trend in den HadCRUT4-Daten. Der Bias verursacht einen kühlenden Bias der Temperaturen der letzten Zeit verglichen mit Ende der neunziger Jahre. Trends, die in den Jahren 1997 oder 1998 beginnen, sind hinsichtlich des globalen Trends besonders verzerrt. Das wird verschärft durch das starke El Nino-Ereignis von 1997/1998, der auch dazu tendiert, Trends zu unterdrücken, die in jenen Jahren begannen.

Veröffentlicht von der Royal Meteorological Society, [link zum abstract](#).

Es gibt eine Website mit Daten und Metadaten und auch ein erklärendes youtube video.

Im Guardian findet sich ein ausführlicher Artikel. Auszüge daraus:

*Es gibt große Gebiete mit Lücken in der Datenerfassung, hauptsächlich in der Arktis, der Antarktis und Afrika, wo es nur ganz vereinzelt Stationen gibt, an denen die Temperatur gemessen wird.*

*Die Temperaturaufzeichnungen von NASA's GISTEMP versuchen, diese Meldelücke zu schließen, indem Temperaturmessungen in der Umgebung auf diese Gebiete extrapoliert werden. Allerdings fehlen den NASA-Daten Korrekturen von der Art, wie Wassertemperaturen gemessen werden – ein Problem, das bislang nur vom Met. Office angegangen worden ist.*

*In ihrer Studie wenden Cowtan & Way das Kriging-Verfahren an, um die Lücken in den Messungen zu füllen, allerdings tun sie das sowohl für das Festland als auch für die Ozeane. In einem zweiten Anlauf nutzen sie auch die nahezu globale Erfassung von Satellitendaten, indem sie die*

Satelliten-Temperaturmessungen der University of Huntsville in Alabama (UAH) mit den verfügbaren Oberflächendaten kombinieren, um die Lücken mit einem ‚Hybrid‘-Datensatz zu füllen. Sie fanden heraus, dass das Kriging-Verfahren die besten Ergebnisse liefert, um Temperaturen über den Ozeanen zu bestimmen, während die Hybrid-Methode am Besten auf dem Festland und vor allem auch über Meereis funktioniert, das große Teile der datenfreien Gebiete ausmacht.

Cowtan & Way untersuchen die Behauptung eines ‚Stillstands‘ der globalen Erwärmung der letzten 16 Jahre, indem sie die Trends von 1997 bis 2012 betrachten. Während aus den HadCRUT4-Daten während dieser Zeit lediglich ein Trend von  $0,046^{\circ}\text{C}$  pro Jahrzehnt hervorgeht – die NASA nennt  $0,08^{\circ}\text{C}$  pro Dekade – schätzen das neue Kriging-Verfahren und die Hybrid-Datensätze einen Trend von jeweils  $0,11^{\circ}\text{C}$  bzw.  $0,12^{\circ}\text{C}$  pro Jahrzehnt.

Diese Ergebnisse zeigen, dass die Verlangsamung der Erwärmung der globalen mittleren Temperatur nicht so signifikant ist wie ursprünglich gedacht. Die Erwärmung hat sich verlangsamt, zum größten Teil deswegen, weil mehr der Erwärmung während des letzten Jahrzehnts in die Ozeane transferiert worden ist (hier). Allerdings erfolgt diese Art der vorübergehenden Verlangsamung der Erwärmung (und Beschleunigungen) auf regelmäßiger Basis infolge kurzfristiger natürlicher Einflüsse.

Die Ergebnisse dieser Studie haben auch Auswirkungen auf die jüngste Forschung. Zum Beispiel ergibt die Korrektur des jüngsten Kalt-Bias‘, dass die globale Temperatur nicht so weit vom Mittel der Modellprojektionen entfernt liegt als ursprünglich gedacht. Sie liegen mit Sicherheit noch innerhalb der Bandbreite der individuellen Temperatursimulationen der Klimamodelle (hier). Jüngste Studien, die ursprünglich zu dem Ergebnis gelangt waren, dass das globale Klima etwas weniger sensitiv auf zunehmende Treibhausgase reagiert als vorher gedacht, könnten ebenfalls die tatsächliche Klimasensitivität unterschätzt haben.

Das ist natürlich nur eine einzige Studie, beeilt sich Dr. Cowtan anzumerken.

„Kein schwieriges wissenschaftliches Problem wird jemals in einer einzigen Studie gelöst. Ich erwarte nicht, dass unsere Studie das letzte Wort hierzu ist, aber ich hoffe, dass wir die Diskussion vorangebracht haben“.

Beispiele für Tweets hierzu:

Dana Nuccitelli: *Diese neue Studie killt den Mythos des Stillstands der globalen Erwärmung.*

John Kennedy: *Die Ironie ist, dass die zum Zerschlagen der HadCRUT4-Daten verwendete Studie annimmt, dass HadCRUT4 dort korrekt ist, wo es Daten gibt.*

Die Studie erregt viel Aufmerksamkeit in den Medien, und auch ich selbst bekomme Fragen von Reportern.

### **Beurteilung von Judith Curry**

Schauen wir zuerst auf die drei Methoden, die sie anwenden, um Datenlücken zu füllen, vor allem in Afrika sowie in der Arktis und der Antarktis.

1. Kriging
2. UAH-Satellitenanalysen der Lufttemperatur
3. Reanalysen von NCAR NCEP

Der größte Unterschied in ihrem rekonstruierten globalen Mittel kommt aus der Arktis, so dass ich mich darauf konzentriere (wobei ich gerade diesbezüglich einige Erfahrung habe).

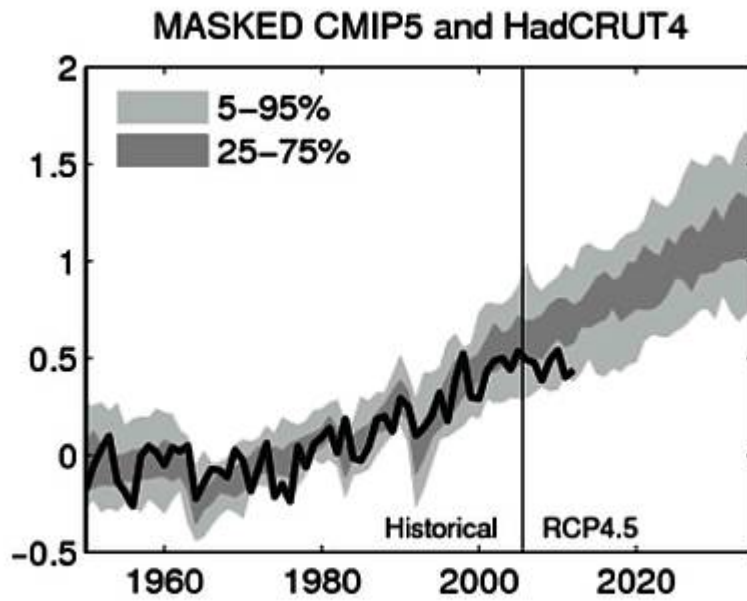
Erstens, Kriging. Das Kriging-Verfahren über Eis auf Land und Wasser macht physikalisch keinen Sinn. Während in der Studie Rigor et al. (2000) genannt werden, die ‚eine gewisse‘ Korrelation im Winter zwischen Festland und Meereis bis zu einer Höhe von 1000 km gefunden haben, würde ich in anderen Jahreszeiten keine Korrelation erwarten.

Zweitens, UAH-Satellitenanalysen. Sie sind unbrauchbar in höheren Breiten bei Temperaturinversionen, und sie sind unbrauchbar über Meereis, das räumlich eine sehr komplexe Verteilung der Abstrahlung im Mikrowellenbereich aufweist). Hoffentlich wird sich John Christy dieses Problems annehmen.

Drittens, Reanalysen in der Arktis. Man betrachte Abbildung 1 in dieser Studie, die einem das Gefühl für die Größenordnung der Gitterpunktfehler an einem Punkt während eines jährlichen Zyklus‘ gibt. Zwar sind sie in gewisser Weise brauchbar, aber Reanalysen sind unbrauchbar, um Trends aufzuzeigen, die temporären Inhomogenitäten in den Datensätzen geschuldet sind.

Also denke ich, dass die Analyse von Cowtan und Way nichts zusätzlich zu unserem Verständnis des globalen Temperaturfeldes und des ‚Stillstands‘ hinzufügt.

Unter dem Strich bleibt die Graphik von Ed Hawkins, in der Klimamodell-Simulationen in Gebieten verglichen werden, in denen Beobachtungen existieren. Das ist der geeignete Weg, um Klimamodelle mit Beobachtungen und Messungen zu vergleichen, und es bleibt die Tatsache, dass Klimamodelle und Beobachtungen stark voneinander abweichen.



Gibt es irgendetwas Nützliches von Cowtan und Way? Nun, sie machen deutlich, dass wir versuchen sollten, auf irgendeine Weise Temperatur-Variationen über dem Arktischen Ozean zu erfassen. Daran wird gegenwärtig geforscht.

Link:

<http://judithcurry.com/2013/11/13/uncertainty-in-sst-measurements-and-data-sets/>

Übersetzt von Chris Frey EIKE