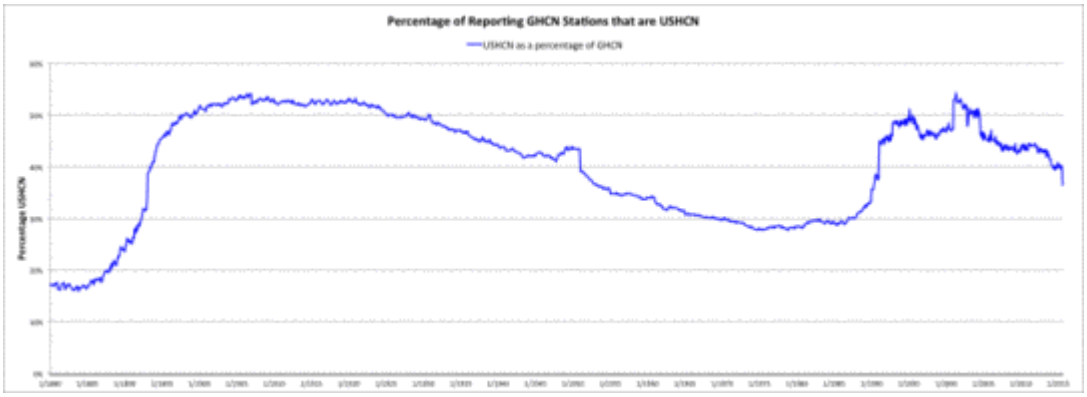
Etwa 92% (oder 99%) aller USHCN-Temperaturdaten bestehen aus geschätzten Werten

geschrieben von John Goetz | 2. Oktober 2015

In einem vorangegangenen Beitrag (hier, auf Deutsch beim EIKE hier) wurde gezeigt, dass die auf die GHCN-Daten angewendeten Adjustierungs-Modelle geschätzte Werte für etwa 66% der Informationen erzeugt hat, die den Verwendern der Daten zur Verfügung gestellt wurden wie etwa GISS. Weil die US-Daten einen relativ großen Beitrag zum Umfang der GHCN-Daten leisten, nimmt sich dieser Beitrag die Auswirkungen der Adjustierungs-Modelle auf die USHCN-Daten vor. Die Graphiken in diesem Beitrag verwenden den um 14 Uhr am 25.9.2015 heruntergeladenen Datensatz von der USHCN FTP Site.
Im USHCN V2.5-File ‚Readme' liest man: „Die Version USHCN 2.5 wird jetzt erzeugt mittels des gleichen Processing-Systems, das auch auf die monatliche Version der GHCN 3-Daten angewendet worden ist. Dieses Reprocessing besteht aus einem Konstruktions-Prozess, der die monatlichen Daten der Version 2.5 des USHCN in einer speziellen Prioritäts-Reihenfolge zusammenfasst (eine, die monatliche Daten bevorzugt, die direkt aus der jüngsten Version von GHCN-Daily berechnet wurden). Qualität kontrolliert die Daten, findet Inhomogenitäten und führt Adjustierungen durch, wo es möglich ist".
Es gibt drei wichtige Unterschiede im GHCN-Prozess. Erstens, der USHCN-Prozess erzeugt einheitliche Ergebnisse, die Schätzungen der Beobachtungszeit an jeder Station zeigen. Zweitens, USHCN wird Werte für fehlende Daten zu schätzen versuchen, ein Prozess, der bekannt ist unter der Bezeichnung *infilling* [im Folgenden mit „Auffüllung" übersetzt]. Derartige Daten werden jedoch vom GHCN nicht verwendet. Der dritte Unterschied ist, dass die homogenisierten Daten für die US-Stationen von den durch das GHCN adjustierten Daten der gleichen US-Stationen abweichen. Meiner Vermutung nach ist der Grund hierfür, dass die Homogenisierungs-Modelle des GHCN Daten über nationale Grenzen hinweg einbringen, während dies beim USHCN nicht der Fall ist. Dies muss noch näher untersucht werden.

Beitrag des USHCN zum GHCN

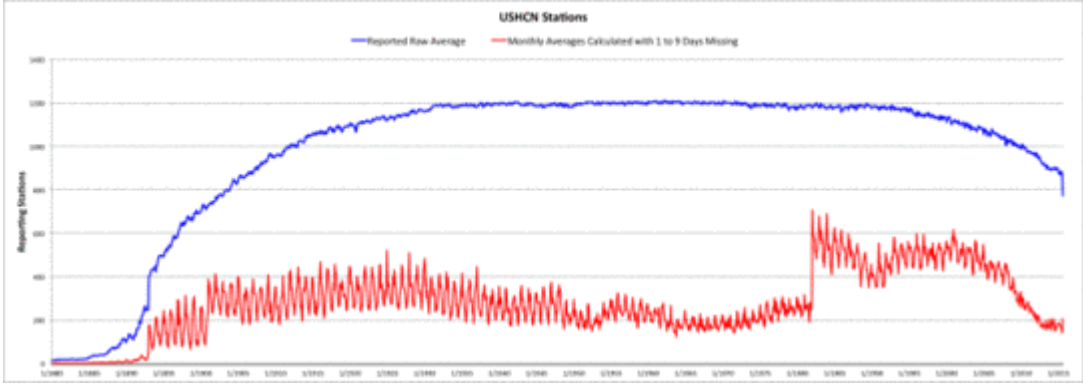
Im Kommentarbereich des o. g. Beitrages wies Tim Ball darauf hin, dass das USHCN eine unverhältnismäßig hohe Datenmenge zum GHCN-Datensatz beiträgt. Die erste Graphik unten zeigt diesen Beitrag im zeitlichen Verlauf. Man beachte, dass das US-Landgebiet (einschließlich Alaska und Hawaii) 6,62% der gesamten Festlandsfläche der Erde ausmacht.



Prozentzahl der Stationen im GHCN, die aus dem USHCN stammen

Wie viele der Daten sind modelliert?

Die folgende Graphik zeigt die Anzahl der Daten, in in der USHCN-Aufzeichnung für jeden Monat von Januar 1880 bis heute verfügbar sind. Auf der Y-Achse ist die Anzahl der Daten liefernden Stationen aufgetragen, so dass jeder Punkt auf der blauen Kurve die Anzahl der Monate repräsentiert, in denen das Monatsmittel aus unvollständigen täglichen Temperaturmessungen berechnet worden ist. USHCN berechnet einen monatlichen Mittelwert auch bei bis zu neun fehlenden Tagen und kennzeichnet den Monat mit einer Rangfolge von „a“ (1 fehlender Tag) bis zu „i“ (neun fehlende Tage). Wie man in der Graphik erkennt, wurden etwa 25% der monatlichen Werte berechnet, wenn einige Tage fehlen. Das offensichtlich jahreszeitliche Verhalten der roten Kurve verlangt nach weiteren Untersuchungen.

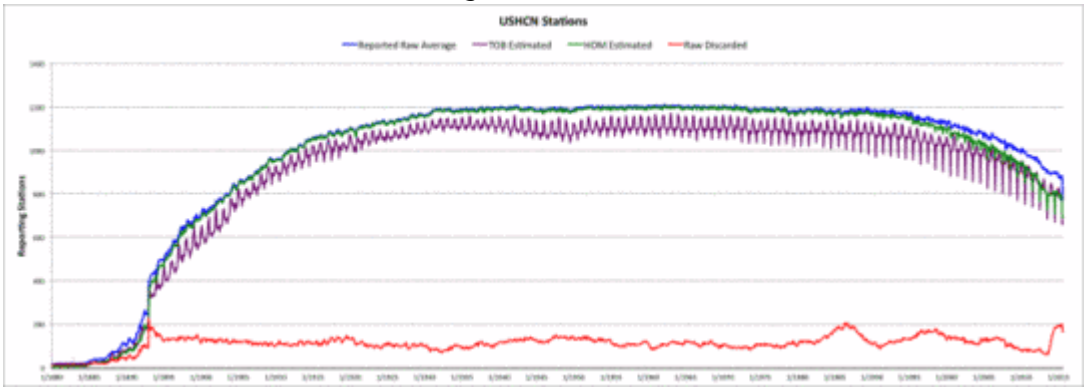


Aufzeichnende USHCN-Stationen

Die dritte Graphik zeigt das Ausmaß, mit dem die Adjustierungs-Modelle die USHCN-Daten beeinflussen. Die blaue Kurze zeigt wieder die für jeden Monat verfügbare Datenmenge im USHCN. Die violette Kurve zeigt die Anzahl von Messungen in jedem Monat, die infolge der Beobachtungszeiten geschätzt worden sind. Etwa 91% der USHCN-Daten weisen eine Beobachtungszeit-Schätzung auf. Die grüne Kurve zeigt die Anzahl von Messungen in jedem Monat, die mittels Homogenisierung geschätzt worden waren. Dies summiert sich zu etwa 99% in der Aufzeichnung. Wie oben schon erwähnt, unterscheiden sich die Schätzungen von GHCN und USHCN für US-Stationen. Im Falle des GHCN sind etwa 92% der US-Aufzeichnungen geschätzt.

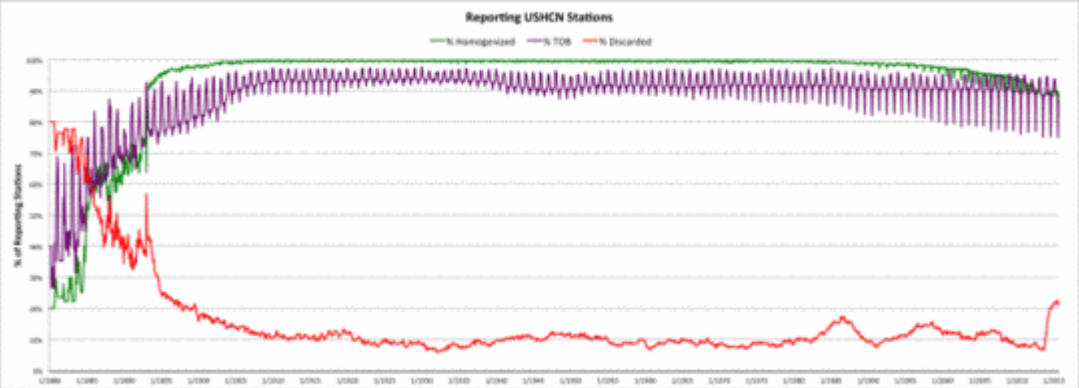
Die rote Kurve ist die Anzahl der Daten, die durch eine Kombination von Homogenisierung und GHCN ausgesondert worden sind. Gelegentlich sondert die Homogenisierung die Originaldaten ganz aus und ersetzt sie durch ungültige Werte (-9999). Öfter jedoch werden ausgesonderte Daten durch

gültige Werte berechnet aus benachbarten Stationen. Wenn dies geschieht, werden die homogenisierten Daten gekennzeichnet mit einem „E“. GHCN zeigt keine Daten, die auf diese Weise gekennzeichnet sind, weshalb sie in der roten Kurve als ausgesondert erscheinen.



Anzahl der Messstationen und Ausmaß der Schätzungen

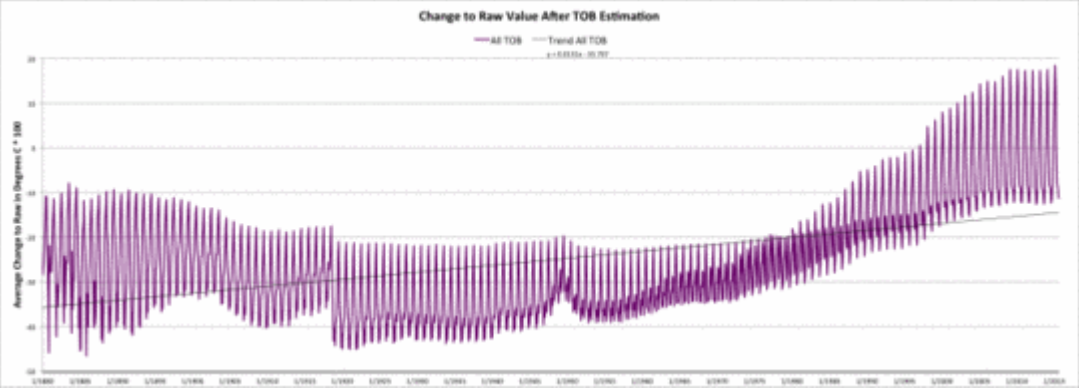
Die nächste Graphik zeigt die drei Datensätze (Beobachtungszeit, homogenisiert, ausgesondert) als Prozentzahl der Gesamtzahl der Stationen.



Ausmaß von USHCN-Schätzungen als Prozentzahl aller eingehenden Messstationen.

Die Auswirkung der Modelle

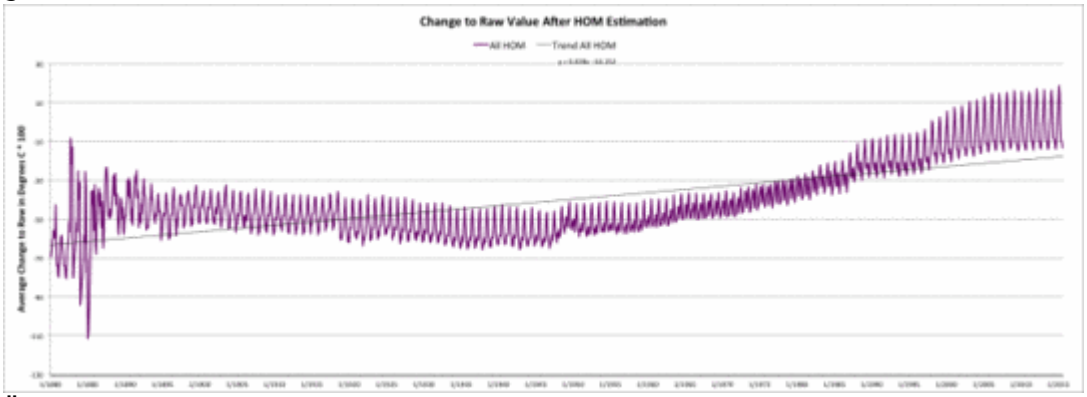
Die fünfte Graphik zeigt die mittlere Änderung der Rohdaten infolge des Beobachtungszeit-Adjustierungsmodells, wenn man diese durch einen geschätzten Wert ersetzt. Die Kurve schließt alle Schätzungen ein, einschließlich der 9% der Fälle, in denen der Beobachtungszeit-Wert gleich dem Wert der Rohdaten ist.



Änderung der USHCN-Rohdatenwerte nach der Beobachtungszeit-Schätzung

Die sechste Graphik zeigt die mittlere Änderung der Rohwerte infolge des Homogenisierungs-Modells. Die Kurve schließt alle Schätzungen ein, einschließlich des 1% aller Werte, in denen der homogenisierte Wert

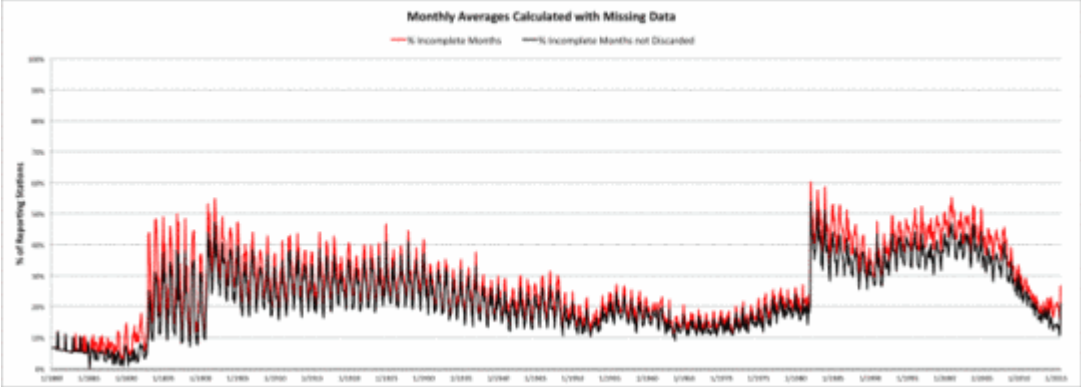
gleich ist dem Rohdatenwert.



Änderung der USHCN-Rohdaten nach der Homogenisierungs-Schätzung

Unvollständige Monate

Wie zuvor beschrieben berechnet das USHCN ein monatliches Mittel, falls bis zu neun Tage mit Daten fehlen. Die folgende Graphik zeigt die Prozentzahl der Monate, in der die Aufzeichnung unvollständig ist (rote Kurve) sowie die Prozentzahl der Monate, die erhalten bleiben, nachdem man die Adjustierungs-Modelle angewendet hatte (schwarze Kurve). Es ist offensichtlich, dass unvollständige Monate nicht oft ausgesondert wurden.



Anzahl der monatlichen USHCN-Mittelwerte, die aus unvollständigen täglichen Beobachtungen berechnet wurden.

Die nächste Graphik zeigt die mittlere Anzahl von Tagen, die fehlten, als die täglichen Aufzeichnungen der Monate unvollständig waren. Nach einer gewissen Volatilität vor dem Jahr 1900 beträgt die mittlere Anzahl fehlender Tage in unvollständigen Monaten zwei (6,5%).

Anzahl der fehlenden Tage bei unvollständigen monatlichen Mittelwerten im USHCN

Ein Wort zum Auffüllen

Die USHCN-Modelle erzeugen Schätzungen für einige fehlende Monate und

ersetzen gelegentlich einen kompletten Monat durch eine Schätzung, falls es zu viele Inhomogenitäten gibt. Die letzte Graphik zeigt die Häufigkeit, mit der dies in der USHCN-Aufzeichnung vorkommt. Die blaue Kurve zeigt die Anzahl der nicht existierenden Messungen, die durch den Auffüllungsprozess geschätzt worden sind. Die violette Linie zeigt die Anzahl bestehender Messungen, die ausgesondert und durch den Auffüllungsprozess ersetzt worden sind. Vor dem Jahr 1920 kam es sehr häufig zur Schätzung fehlender Daten. Seitdem ist es öfter zur Ersetzung fehlender Daten gekommen, als dass fehlende Daten geschätzt worden sind. Aufgefüllte Daten sind in den Adjustierungs-Schätzungen des GHCN nicht enthalten.

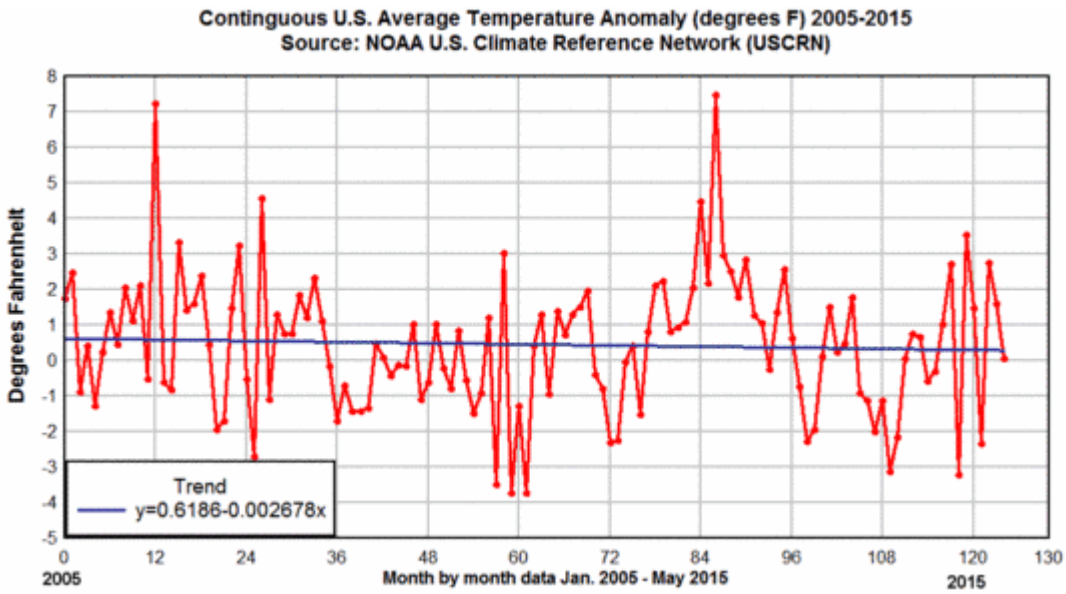
Anzahl der Auffüllung fehlender Daten im USHCN.

Schlussfolgerung

Das US-Festland macht 6,62% der Landfläche der Erde aus, doch stammen von hier 39% der Daten im GHCN-Netzwerk. Von 1880 bis zur Gegenwart wurden etwa 99% aller Temperaturdaten im homogenisierten USHCN-Output geschätzt (mit Unterschieden zu den Original-Rohdaten). Etwa 92% der Temperaturdaten im Beobachtungszeitpunkt-Output des USHCN sind geschätzt worden. Die Adjustierungsmodelle des GHCN schätzen etwa 92% der US-Temperaturen, aber diese Schätzungen passen weder zu den USHCN-Beobachtungszeit-Schätzungen noch zu den homogenisierten Schätzungen. Die Homogenisierungs-Schätzung führt einen positiven Temperaturtrend von etwa 0,34°C pro Jahrhundert relativ zu den USHCN-Rohdaten ein. Die Beobachtungszeit-Schätzungen führen einen positiven Temperaturtrend von etwa 0,16°C pro Jahrhundert ein. Diese sind nicht additiv. Der Homogenisierungs-Trend steht bereits für den Beobachtungszeit-Trend.

Anmerkung: Das US-Klima-Referenznetzwerk, von Anfang an vorgesehen zur freien Verfügung für JEDWEDE Adjustierung von Daten, zeigt keinerlei Trend, wie ich bereits im Juni 2015 in einem Beitrag mit dem Titel [übersetzt] „Trotz der Versuche, ihn global zu beseitigen, existiert der „Stillstand“ immer noch in ursprünglichen US-Temperaturdaten“ (hier) erläutert habe.

Plottet man die Daten aus jenem Netzwerk, sieht das so aus:



Bemerkungen von Anthony Watts hierzu: Natürlich haben Tom Karl und Tom Peterson von NOAA/NCDC (jetzt NCEI) niemals diese USCRN-Daten an das Tageslicht gebracht, weder in einer Presseerklärung noch in einem Klimazustands-Bericht für die Medien. Sie wurden auf eine Nebenseite der Website verbannt und niemals erwähnt. Wenn es um Behauptungen bzgl. wärmstes Jahr/Monat/Tag jemals geht, werden stattdessen der Öffentlichkeit ausschließlich diese stark adjustierten, höchst unsicheren USHCN/GHCN-Daten in diesen regelmäßigen Berichten gezeigt. Man fragt sich, warum NOAA NCDC/NCEI Millionen Dollar ausgegeben haben, um ein Klimazustands-Netzwerk für die USA zu erschaffen und dieses dann niemals zur Information der Öffentlichkeit herangezogen wird. Vielleicht liegt es daran, dass es nicht die gewünschten Ergebnisse zeigt?
Link:
http://wattsupwiththat.com/2015/09/27/approximately-92-or-99-of-ushcn-surface-temperature-data-consists-of-estimated-values/
Übersetzt von Chris Frey EIKE