

# Die Gefahr kurzer Datensätze

geschrieben von Chris Frey | 22. März 2023

[Willis Eschenbach](#)

Vor ein paar Monaten stieß ich auf eine weitere Behauptung, dass der Sonnenfleckenzyklus das Wetter hier unten an der Erdoberfläche beeinflusst, insbesondere die Meerestemperaturen in der El-Nino-Region des tropischen Pazifiks. Die [Arbeit](#) heißt Evidence of solar 11-year cycle from Sea Surface Temperature (SST), von Mazza und Canuto, im Folgenden MC2021 genannt. Ich habe darüber in meinem [Beitrag](#) „CEEMD Versus Joe Fourier“ geschrieben. In diesem Beitrag möchte ich meine Analyse ein wenig erweitern und einen der Gründe klären, warum die Behauptungen von MC2021 nicht stimmen.

Diejenigen, die meine Arbeiten lesen, wissen vielleicht, dass ich anfangs wirklich daran glaubte, dass Sonnenflecken das Wetter auf der Oberfläche beeinflussen. Als Kind hatte ich von William Herschels Behauptung aus dem Jahr 1801 gelesen, dass Sonnenflecken die Weizenpreise in England beeinflussen. Ich dachte also, dass es sehr einfach sein würde, Beweise dafür zu finden, dass die durch Sonnenflecken verursachten Schwankungen der Sonnenenergie tatsächlich das Wetter auf der Erdoberfläche beeinflussen.

Aber als ich mir die Daten zum ersten Mal ansah, fand ich ... nichts. Also habe ich weiter gesucht. Seitdem habe ich mir Dutzende von angeblichen Korrelationen angesehen und ... nichts gefunden. Nun, das ist nicht ganz richtig. Ich fand eine wissenschaftliche [Abhandlung](#) mit dem Titel „On The Insignificance Of Herschel's Sunspot Correlation“ (Über die Bedeutungslosigkeit von Herschels Sonnenflecken-Korrelation), in der der Autor nach stichhaltigen Beweisen für Herschels Behauptung suchte. Er fand ...

nichts.

Daher war ich an der MC2021-Studie interessiert. Darin heißt es:

*Nachdem wir Hunderte von Temperaturlaufzeichnungen der Erdoberfläche heruntergeladen und analysiert hatten, fanden wir in einigen wenigen Fällen eindeutige Beweise für die Signatur des 11-Jahres-Zyklus der Sonne, während diese in der überwiegenden Mehrheit der anderen Fälle nicht nachweisbar war, da sie unter anderen (saisonalen oder El-Nino-bedingten) Schwingungen oder Rauschen begraben war. Wir haben festgestellt, dass zwei Bedingungen am günstigsten sind, um die richtige Sonnensignatur in den Temperaturlaufzeichnungen zu finden: als Daumenregel:*

*Die tropischen Meerestemperaturen im Bereich von 5°N – 5°S. Das ist nicht verwunderlich, da die Sonnenstrahlen dort ihre Energie mit weniger*

Reflexion oder Streuung und mit einem optimalen Einfallswinkel auf die Wasseroberfläche übertragen.

Vergessen Sie Anomalien oder Indizes jeglicher Art; für unsere Ziele sind dies nur Datenverstopfungen. Schauen Sie, wann immer möglich, auf die realen Meerestemperaturen (SST).

Nach sorgfältiger Analyse der vielen Temperaturaufzeichnungen der gesamten Erdtemperatur sowie ausgewählter Regionen haben wir zwei Regionen ermittelt, die am stärksten vom 11-jährigen Sonnenzyklus betroffen sind. Bei beiden handelt es sich um die äquatorialen Ozeanregionen, die den Klimaforschern als El-Nino-3 und El-Nino-3-4 bekannt sind.

Hier ist ihr Diagramm, das die behauptete Beziehung zeigt:

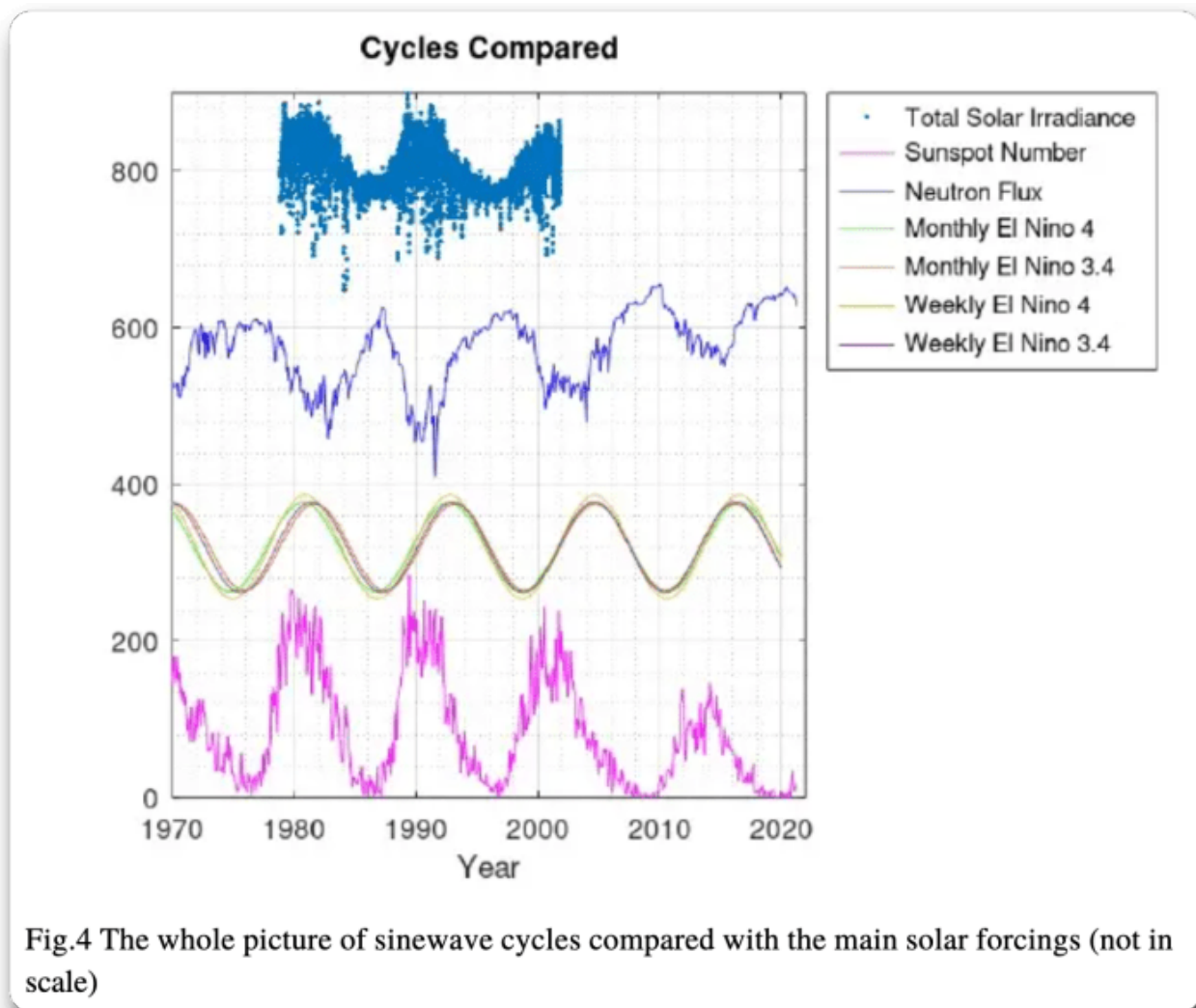


Abbildung 1. Abbildung 4 von MC2021, die die Beziehung zwischen Sonnenflecken und den Zyklen in den El-Nino-Regionen zeigt.

Also habe ich nachgesehen, ob ich ihre Ergebnisse reproduzieren kann.

Anstatt nur die Indizes Nino4 und Nino34 zu verwenden, habe ich auch den [Multivariaten ENSO-Index](#) (MEI) und den [Southern Oscillation Index](#) (SOI) herangezogen. Von all diesen Indizes ist bekannt, dass sie in gewisser Weise mit der El Nino/La Nina-Oszillation im tropischen Pazifik korrelieren. (Ich habe den NINO3-Index nicht verwendet, weil er nicht mit den anderen korreliert.) Hier ist das Ergebnis, beginnend im Jahr 1979, dem Beginn des MEI-Datensatzes:

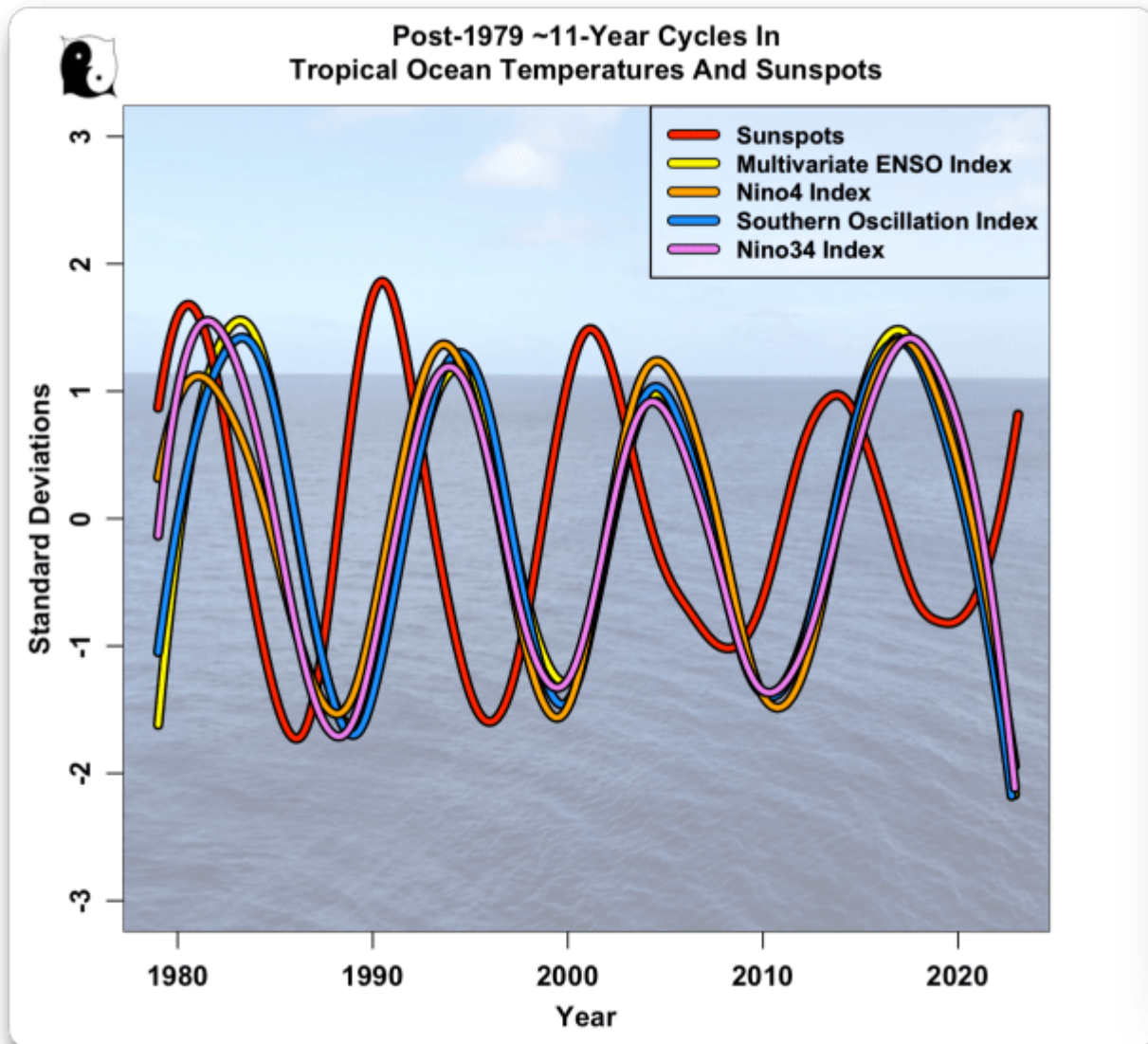


Abbildung 2. Ein Vergleich der zugrundeliegenden ~ 11-Jahres-Zyklen in den Sonnenflecken und dem tropischen Pazifik. Die Zyklen wurden mithilfe der vollständigen empirischen Ensemble-Mode-Zerlegung ermittelt.

Eine Sache ist sicher. Nino4, Nino34, MEI und SOI sind allesamt eindeutig unterschiedliche Messgrößen für das gleiche zugrunde liegende Phänomen. In jedem Datensatz können Sie die jüngsten, sehr langen La Nina-Bedingungen auf der rechten Seite der Grafik sehen, und die Datensätze stimmen durchweg gut miteinander überein.

Und sie stimmen alle recht gut mit den Sonnenflecken überein, mit einer Verzögerung von einigen Jahren zwischen den Sonnenflecken und den Indizes der tropischen Ozeane.

Was gibt es also nicht zu mögen?

Nun, was man nicht mögen kann, ist, dass die Datensätze sehr kurz sind. Meine Faustregel besagt, dass man mit nur drei Zyklen eines Phänomens nicht viel aussagen kann, und ich habe mich schon mehr als einmal von fünf Zyklen täuschen lassen. Und hier haben wir nur vier Zyklen.

Glücklicherweise reicht der MEI zwar nur bis 1979 zurück, die anderen drei Indizes reichen jedoch viel weiter zurück. Hier sind die vollständigen Datensätze, beginnend mit dem Jahr 1870:

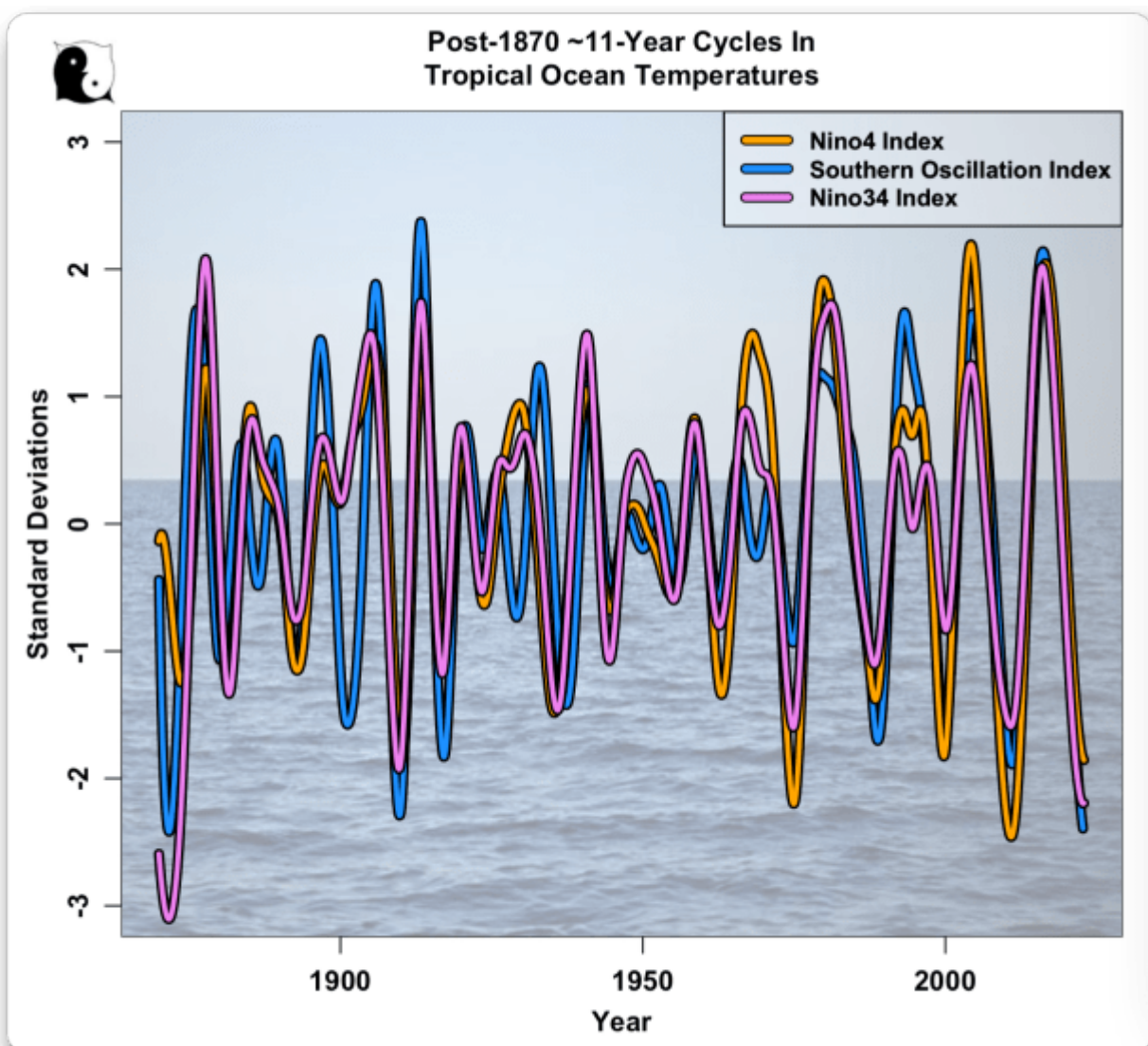


Abbildung 3. Vergleich der zugrunde liegenden ~ 11-jährigen Zyklen im tropischen Pazifik nach 1870. Die Zyklen wurden mithilfe der [Complete Empirical Ensemble Mode Decomposition](#) (CEEMD) bestimmt.



Da es eine recht gute Übereinstimmung zwischen den Zeitpunkten der Zyklen der drei Datensätze gibt, möchte ich ihren Durchschnitt verwenden, um die langfristigen Bedingungen im tropischen Pazifik darzustellen, und diesen mit den Sonnenfleckendaten vergleichen:

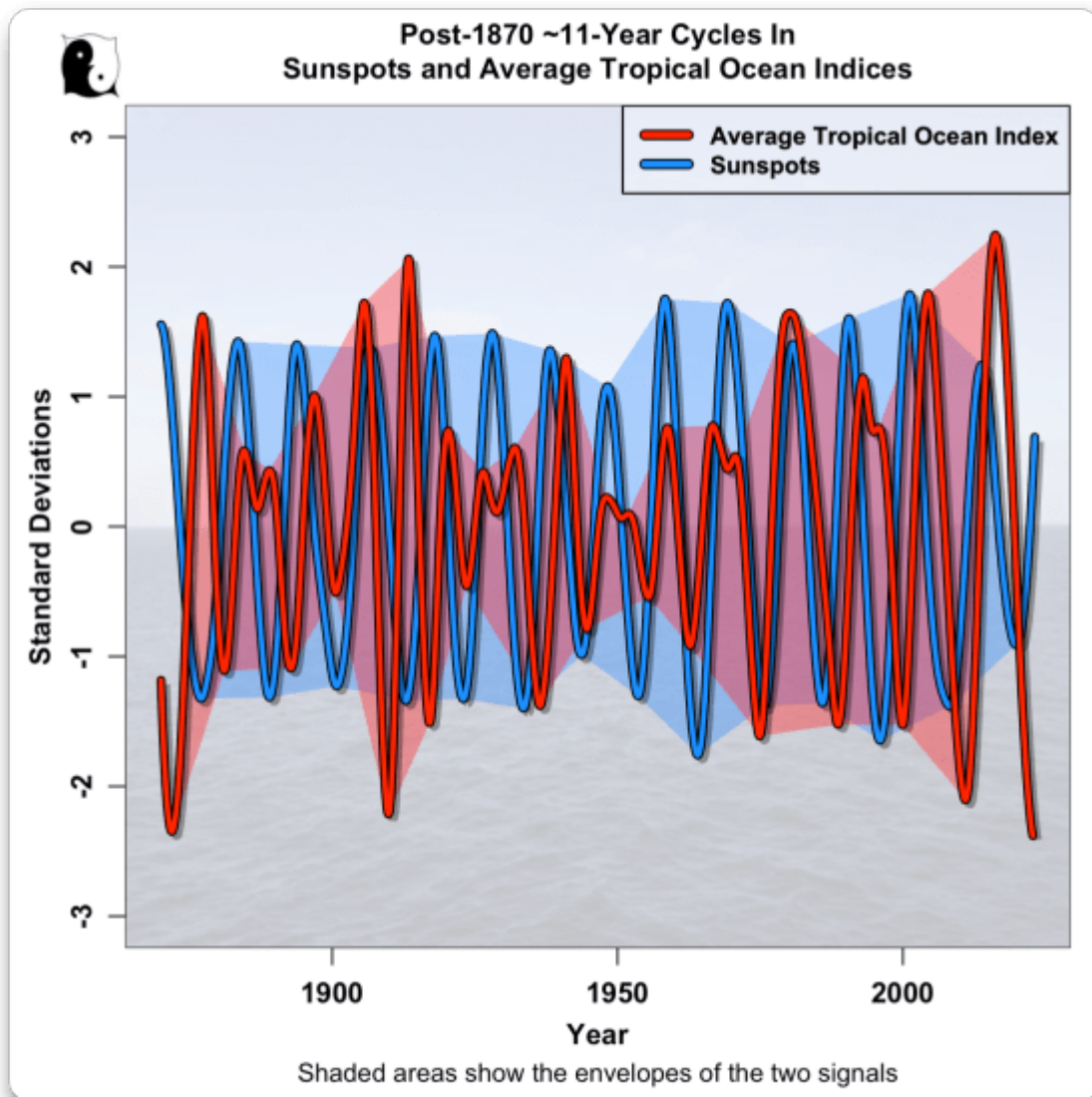


Abbildung 4. Ein Vergleich der Sonnenflecken nach 1870 und der zugrunde liegenden durchschnittlichen ~ 11-jährigen Zyklen im tropischen Pazifik. Die Zyklen wurden mithilfe der [Complete Empirical Ensemble Mode Decomposition](#) (CEEMD) bestimmt.

Ich bin sicher, Sie können die Schwierigkeit erkennen. Vor etwa 1945 ist der Ozean weit von der Phase der Sonnenflecken entfernt. Und je weiter man zurückgeht, desto größer wird die Diskrepanz zwischen den beiden Datensätzen. Hinzu kommt, dass die Hüllkurven der Signale sehr unterschiedlich sind. Man würde erwarten, dass, wenn das Sonnensignal stark ist, auch das Temperatursignal stark sein sollte ... aber das ist überhaupt nicht der Fall.

Dies lässt sich auch anhand der Fourier-Analyse erkennen. Der jüngere Teil des ENSO-Datensatzes nach 1960 weist einen eindeutigen 12-Jahres-Zyklus auf (nicht 11, sondern 12 Jahre, blaue Linie, rechtes Feld) ... aber wenn man sich den gesamten Datensatz bis 1870 ansieht (rote Linie, rechtes Feld), verschwindet dieser Zyklus im Rauschen und ändert sich in einen 13-Jahres-Zyklus. Beachten Sie, dass dies bei den Sonnenfleckendaten nicht der Fall ist (linkes Feld). Dort liegt der 11-Jahres-Zyklus immer deutlich über dem Rauschen und bleibt konstant bei 11 Jahren.

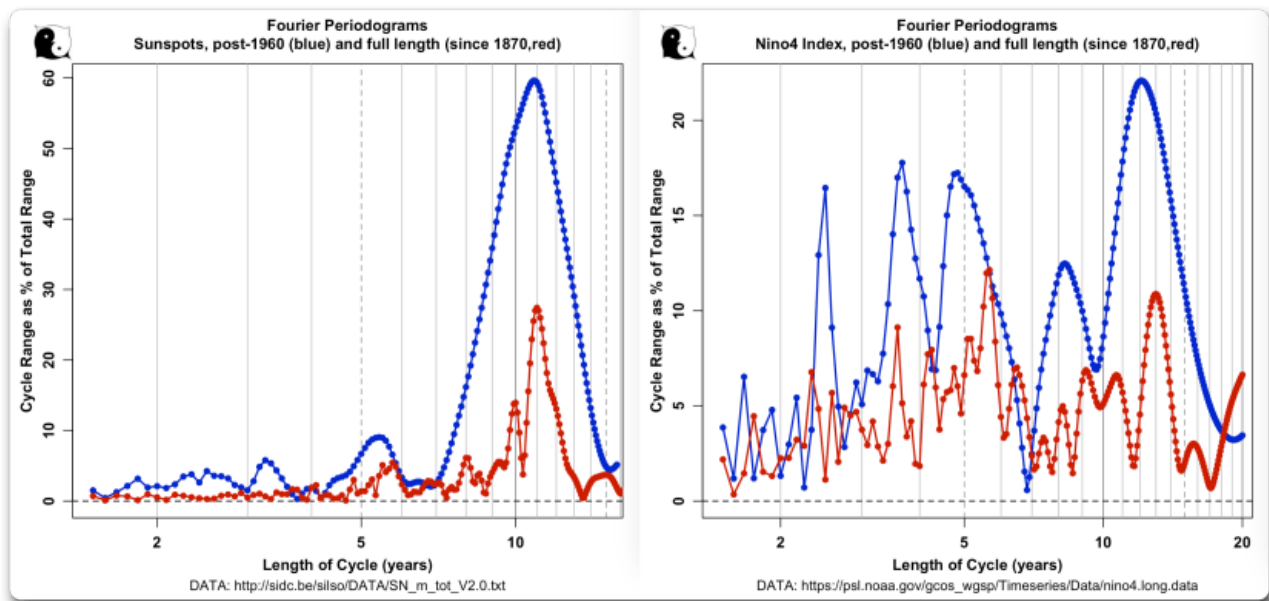


Abbildung 5. Fourier-Periodogramme der Sonnenflecken in voller Länge (rot) und nach 1960 (linkes Feld) sowie der Niño4-Index (rechtes Feld).

Leider haben wir, die wir uns mit Klimawissenschaften beschäftigen, es mit einem äußerst komplexen System zu tun. Das Klima setzt sich aus sechs großen Teilsystemen zusammen: Atmosphäre, Biosphäre, Hydrosphäre, Kryosphäre, Lithosphäre und Elektrosphäre. Jedes dieser Teilsysteme weist interne Resonanzen und Zyklen auf, die auf verschiedenen Zeitskalen von Millisekunden bis zu Millionen von Jahren auftreten. Und nicht nur das: Alle Teilsysteme tauschen auf denselben Zeitskalen in regelmäßigen, unregelmäßigen und zufälligen Abständen Energie aus. Schließlich wird das System von einer ständig wechselnden Energiequelle gespeist.

Selbst das IPCC räumt ein, dass dieses System völlig chaotisch ist, ein Chaos, das seit Millionen von Jahren besteht. Infolgedessen sehen wir oft das, was ich „Pseudozyklen“ nenne. Das sind zyklische Schwankungen in bestimmten Datensätzen. Es handelt sich jedoch nicht um echte Zyklen – sie treten ohne Vorwarnung auf, dauern einige Zeit an und verschwinden dann wieder, um durch andere Pseudozyklen ersetzt zu werden.

Dieses Problem wird durch die Tatsache verschärft, dass so viele unserer

wetterbezogenen Datensätze so kurz sind, oft kürzer als ein Menschenleben. Der MSU-Datensatz für die Temperatur der unteren Troposphäre umfasst nur 44 Jahre an Daten, kürzer als die meisten Menschenleben. Dasselbe gilt für den Multivariaten Enso-Index und auch für die Daten nach 1970, die in dem hier diskutierten Papier MC2021 verwendet werden.

Und diese Überschneidung von kurzen Datensätzen und Pseudozyklen führt zu vielen Behauptungen über zyklisches Verhalten in Temperatur-Datensätzen, insbesondere in Bezug auf Sonnenflecken, wo es in Wirklichkeit keine echten, unveränderlichen Zyklen gibt.

Es gibt ein letztes Problem. Dies hat mit der Tatsache zu tun, dass natürliche Wetterdaten oft „autokorreliert“ sind. Das bedeutet, dass die Temperatur von heute oft mit der von gestern zusammenhängt, und die Temperatur von diesem Monat oft mit der des letzten Monats.

Das Problem ist, dass autokorrelierte Daten oft Zyklen enthalten ... Zyklen, die etwas bedeuten können oder auch nicht. Hier ist zum Beispiel ein zufälliger autokorrelierter Datensatz eines Typs, der „Fractional Gaussian Noise“ oder „FGN“ genannt wird. Wie der Name schon sagt, handelt es sich nur um Rauschen, nicht um sinnvolle Signale. Beachten Sie, wie sehr er beispielsweise einem natürlichen Temperaturdatensatz ähnelt:

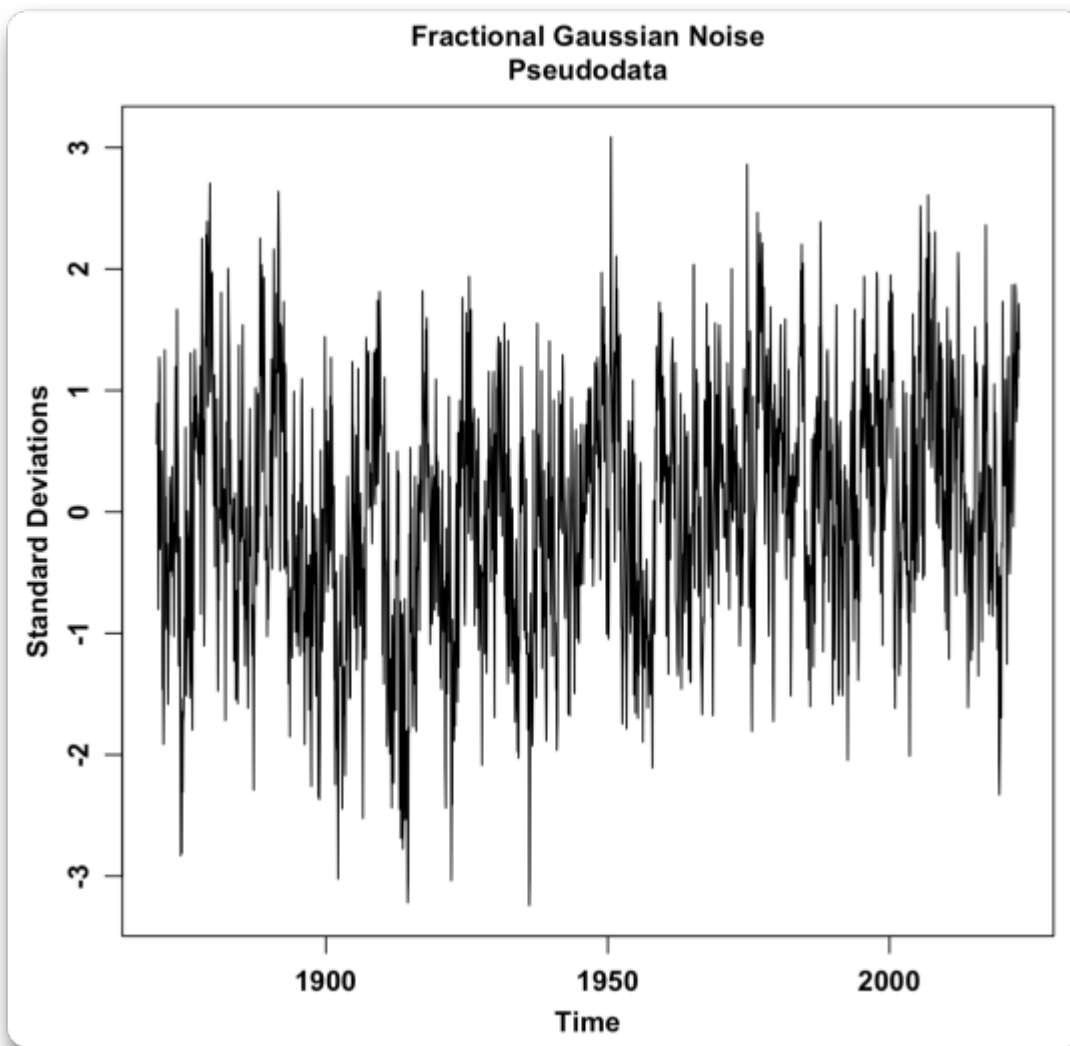


Abbildung 6. Ein Beispiel für zufälliges fraktionales Gauß'sches Rauschen (fractional gaussian noise, FGN).

Und hier ist die CEEMD-Zerlegung desselben FGN-Datensatzes. Denken Sie daran, dass es sich hier nur um Rauschen handelt und keine tatsächlichen Signale vorhanden sind:



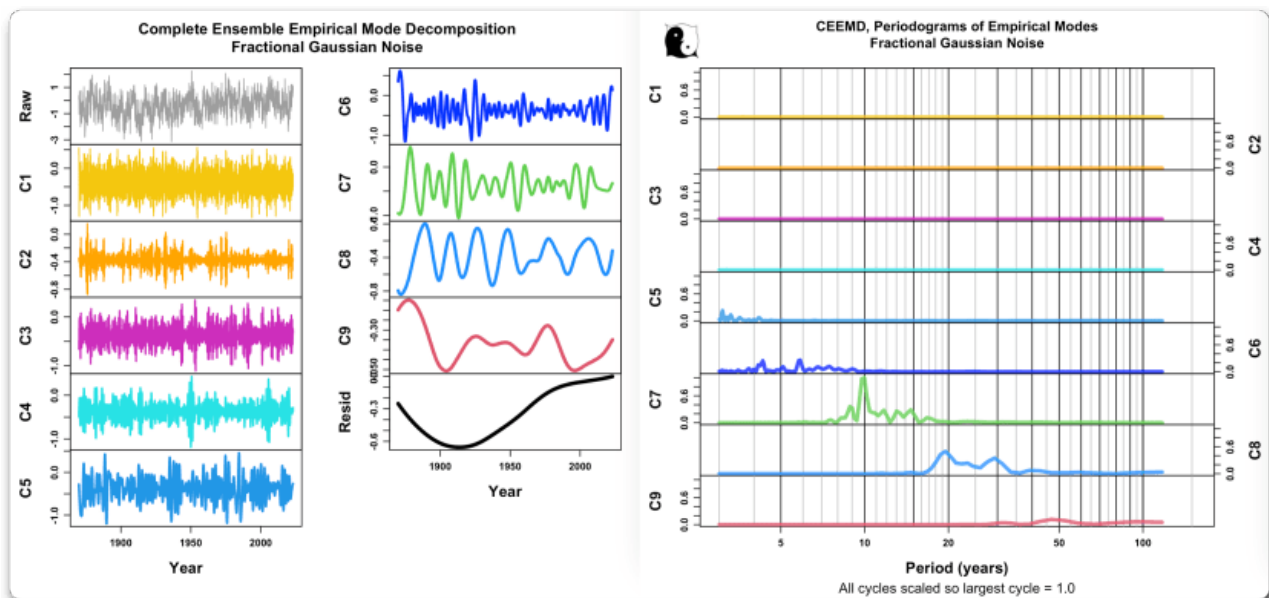


Abbildung 7. CEEMD-Analyse eines Beispiels von zufälligem fraktionalem Gauß'schen Rauschen (FGN).

Also ... was sehen wir hier? Nun, das linke Feld zeigt die einzelnen Signale, die sich aus der Zerlegung des FGN-Rauschens ergeben. Wie Sie sehen können, gibt es Signale bei einer Vielzahl von Frequenzen, mit einem Rest, der zeigt, was übrig bleibt, nachdem alle regulären Signale entfernt wurden.

Das rechte Feld zeigt dagegen „Periodogramme“ für jedes der Signale im linken Feld mit den entsprechenden Farben. Die Periodogramme zeigen, welches die stärksten Zyklen in den zerlegten Signalen sind. Man erkennt: der stärkste Zyklus liegt bei 10 Jahren und ist in Grün dargestellt. Sie können dieses Signal auf dem linken Feld sehen. Das muss von den Sonnenflecken kommen! ... oh, Moment, das ist nur zufälliges Gauß'sches Rauschen ...

Die Zehnjahreszyklen sind jedoch weit von einem regelmäßigen Signal entfernt. Sowohl die Amplitude als auch die Zykluslänge ändern sich ständig, wie man an der grünen Linie im linken Feld erkennen kann. Dies ist auch an den kleineren Spitzen in der grünen Linie auf der rechten Seite zu erkennen.

Das ist es, was ich einen „Pseudozyklus“ genannt habe. „Pseudo“ deshalb, weil es sich nicht um einen unveränderlichen, dauerhaften Zyklus handelt. Stattdessen ist es nur einer der vielen Zyklen, die wir in allen autokorrelierten Daten finden, bedeutungslose Zyklen, die erscheinen, sich ändern und wieder verschwinden.

Zu beachten ist auch, dass das FGN-Rauschen einen anhaltenden Trend enthält, die schwarze Linie unten rechts im linken Feld. Dies ist das Ergebnis der gleichen Sache, der Autokorrelation. Wie in dem [AGU-Artikel](#) „Nature's Style: Naturally Trendy“ beschrieben können Trends in

natürlichen Datensätzen genauso wenig bedeuten wie der Trend, den diese FGN-Daten zeigen.

Aus diesem Grund müssen in der Klimawissenschaft die statistischen Behauptungen immer und jedes Mal um die Autokorrelation bereinigt werden ... zum Beispiel sagt die Statistik des oben gezeigten FGN-Signals ohne die Bereinigung um die Autokorrelation, dass der Trend sehr aussagekräftig ist, mit einem p-Wert von  $<2e-16$ .

Aber wenn wir die Autokorrelation bereinigen und die Bonferroni-Berechnung verwenden, um die Tatsache zu berücksichtigen, dass ich 5 FGN-Datensätze untersucht habe, um diesen einen zu finden, stellt sich heraus, dass er überhaupt nicht statistisch signifikant ist, mit einem Autokorrelations- und Bonferroni-bereinigten p-Wert von 0,075 ... also doch nur zufälliges Rauschen.

Klimawissenschaft. Versteckte Schlaglöcher überall.

---

*Es folgt noch eine Liste von Beiträgen verschiedener Autoren dazu aus früheren Jahren.*

Link:

<https://wattsupwiththat.com/2023/03/18/the-danger-of-short-datasets/>

Übersetzt von Christian Freuer für das EIKE