



POTSDAM-INSTITUT FÜR
KLIMAFOLGENFORSCHUNG

Originally published as:

Rahmstorf, S., Foster, G., Cahill, N. (2017): Global temperature evolution: recent trends and some pitfalls. - Environmental Research Letters, 12, 054001

DOI: [10.1088/1748-9326/aa6825](https://doi.org/10.1088/1748-9326/aa6825)

Environmental Research Letters



LETTER

Global temperature evolution: recent trends and some pitfalls

OPEN ACCESS

RECEIVED
6 December 2016

REVISED
28 February 2017

ACCEPTED FOR PUBLICATION
21 March 2017

PUBLISHED
25 April 2017

Stefan Rahmstorf^{1,4}, Grant Foster² and Niamh Cahill³

¹ Potsdam Institute for Climate Impact Research, Potsdam, Germany

² 1 Noyes Ct, Augusta, ME 04330, United States of America

³ Department of Biostatistics and Epidemiology, UMASS Amherst, Amherst, MA, United States of America

⁴ Author to whom any correspondence should be addressed.

E-mail: rahmstorf@pik-potsdam.de

Keywords: global temperature, time series, slowdown, temperature records

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Abstract

Global surface temperatures continue to rise. In most surface temperature data sets, the years 2014, 2015 and again 2016 set new global heat records since the start of regular measurements. Never before have three record years occurred in a row. We show that this recent streak of record heat does not in itself provide statistical evidence for an acceleration of global warming, nor was it preceded by a ‘slowdown period’ with a significantly reduced rate of warming. Rather, the data are fully consistent with a steady global warming trend since the 1970s, superimposed with random, stationary, short-term variability. All recent variations in short-term trends are well within what was to be expected, based on the observed warming trend and the observed variability from the 1970s up to the year 2000. We discuss some pitfalls of statistical analysis of global temperatures which have led to incorrect claims of an unexpected or significant warming slowdown.

1. Introduction

Global-mean surface temperature (GMST) is the most important indicator of global climate change, because (i) it is directly related to the planetary energy balance (Fourier 1827) and increases quasi-linearly with cumulative greenhouse gas emissions (IPCC—Intergovernmental Panel on Climate Change 2013), and (ii) GMST is directly related to most climate impacts and risks (Arnell *et al* 2014). Hence there is a large interest in the time evolution of GMST, both in the scientific community and the general public (see e.g. Boykoff 2014, Lewandowsky *et al* 2015, Mooney 2013). Two facets of this high interest are frequent discussions about (i) whether the rise in GMST has accelerated or slowed down, and (ii) how well it agrees with various model projections. These are two separate issues; this paper deals with the former only, i.e. with analysis of possible trend changes in the observational data. Our goal is to provide a current analysis of GMST trends in the light of the recent series of three record-breaking years in a row in most data sets (never seen before in the instrumental record), and to point out two important pitfalls in analysing GMST trends.

While many scientific publications of the past years have discussed an alleged ‘hiatus’ or ‘slowdown’ and its possible causes, few have provided any statistical assessment of whether a significant trend change actually occurred. While it is clear and undisputed that the global temperature data show short periods of greater and smaller warming trends or even short periods of cooling, the key question is: is this just due to the ever-present noise, i.e. short-term variability in temperature? Or does it signify a change in behavior, e.g. in the underlying warming trend? In other words, are periods of particularly high or low warming trend significant, in that any of them is unexpected and requires further explanation than just the usual noise in the data? While it is a semantic question what the meaning of a ‘hiatus’ is, the question of significance is a well-defined scientific question.

Foster and Abraham (2015) applied ‘a barrage of statistical tests’ to the NASA GISTEMP data for 1970–2013 ‘to search for evidence of any significant departure from a linear increase at constant rate since 1970.’ In every case, the analysis not only failed to establish a trend change with statistical significance, it failed by a wide margin.

Rajaratnam *et al* (2015) used four GMST data sets up to 2014 to perform statistical tests of four different hypotheses, namely ‘whether the recent period has demonstrated (i) a hiatus in the trend in global temperatures, (ii) a temperature trend that is statistically distinct from trends prior to the hiatus period, (iii) a ‘stalling’ of the global mean temperature, and (iv) a change in the distribution of the year-to-year temperature increases.’ They ‘conclude that the rate of warming over the past ≈ 15 yr is not appreciably different from the rate of warming prior to the recent period.’ They further find ‘overwhelming evidence’ against a ‘pause’ in warming (i.e. no trend) over the past ≈ 15 yr.

Cahill *et al* (2015) likewise analysed four GMST data sets up to 2014 using change point analysis, an established statistical technique to identify significant changes in trends in a data set. They found ‘no evidence of any detectable change in the global warming trend since ~ 1970 .’

Finally, Lewandowsky *et al* (2016) have investigated whether the most recent fluctuation in 15 yr trend value (as defined by a z-score) is unusual, by considering all possible 15 yr trends in GMST between 1970 and 2014. They find at least three instances of similar or greater fluctuations in 15 yr trends, the largest of all being the exceptionally rapid warming trend during 1992–2006. Incidentally, this large trend was noted by Rahmstorf *et al* (2007), who proposed ‘intrinsic variability within the climate system’ as the first candidate reason. Lewandowsky *et al* conclude that the pause period (comprising the 15 yr trends 1998–2012, 1999–2013 and 2000–2014) ‘is not unusual or extraordinary relative to other fluctuations and it does not stand out in any meaningful statistical sense.’

In contrast to these studies, Fyfe *et al* (2016) claimed that ‘the surface warming from 2001 to 2014 is significantly smaller than the baseline warming rate,’ where that baseline is 1972–2000. This claim was not backed up by statistical analysis, nor was any of the previous analysis cited that we discussed above.

In the following we will revisit the issue of trend changes in GMST with data up to 2016. We highlight two problems with some previous trend analyses: the multiple testing problem and the problem of using broken trends.

2. Data and terminology

One sense of the word ‘trend’ is the underlying value of a time series, the signal value as opposed to the noise. In the *additive* noise model the time series values are the sum of trend plus noise, i.e.

$$x_j = f(t_j) + \varepsilon_j, \quad (1)$$

where x_j are the data values, $f(t_j)$ the signal values, and ε_j the noise values. We adopt the common convention

that the signal value is the expected value of the data at a particular time

$$f(t_j) = \mathbf{E}(x|t_j), \quad (2)$$

which imposes the condition that the noise is zero-mean, i.e.

$$\mathbf{E}(\varepsilon_j) = 0 \quad \text{for all } j. \quad (3)$$

Another sense of the word trend, the one which we will adopt, is the *rate of change* of the underlying signal, i.e.

$$\beta(t_j) = \left(\frac{df}{dt} \right)_{t_j}. \quad (4)$$

Hence in the context of GMST, *trend* refers to the rate at which global temperature is changing.

Only if the rate of change is constant will the signal follow a straight line. Although this is rarely the case with complete precision, it very often happens that the noise level is high enough to make it impossible to establish trend change except over very long time spans. In such cases the trend is usually estimated by fitting the very same straight-line model. Particularly for global temperature time series, ‘trend’ usually refers to the estimated rate of change of the underlying temperature signal from linear regression. Given the signal-to-noise ratio of global temperature on decadal time scales, this is the most practical determination of the trend, and is the one most often cited in the literature.

Establishing acceleration or deceleration of global temperature means detecting and confirming a change in the trend. Since the trend is distinct from the noise, the influence of noise will always lead to apparent changes. Distinguishing those which are genuine from those induced by noise is the purpose of statistical significance testing.

Statistical significance is a concept that is widely used but also critically discussed, mainly regarding the ambiguity of the threshold value (e.g. 90% or 95%) and the choice of the null hypothesis (see e.g. Nicholls 2001). The key to its usefulness is to clearly define what is meant. For our purpose, a significant slowdown or acceleration in global warming is a behavior of global-mean temperature which is highly unlikely to occur under the null hypothesis of a constant warming trend plus short-term random variations as observed in the past (where ‘past’ refers to a suitably defined baseline period). In other words, a significant change in warming trend refers to a temperature evolution which is unlikely to be a result of a simple continuation of the warming trend and random noise found in the baseline period. Our null hypothesis is thus that long-term trend and short-term noise continue unchanged. Any claim of a significant slowdown or acceleration

Table 1. Monte Carlo Results. Shown are the standard deviations and trends during the baseline period, the trends during the ‘slowdown’ periods, and the likelihood that a trend at least as small as the latter would be observed by chance if the baseline trend and standard deviation had continued unchanged.

Data set	1972–2000	1972–2000	2001–2014	2001–2014
	Std. dev.	Baseline trend	Slow trend	Likelihood
HadCRUT4	0.103 °C	0.0172 °C yr ⁻¹	0.0030 °C yr ⁻¹	31%
GISTEMP	0.103 °C	0.0173 °C yr ⁻¹	0.0074 °C yr ⁻¹	73%
Data set	1972–1999	1972–1999	2000–2012	2000–2012
	Std. dev.	Baseline trend	Slow trend	Likelihood
HadCRUT4	0.104 °C	0.0177 °C yr ⁻¹	0.0056 °C yr ⁻¹	65%
GISTEMP	0.104 °C	0.0179 °C yr ⁻¹	0.0103 °C yr ⁻¹	96%

would require data that are highly unlikely (e.g. 5% or 10% likelihood depending on the desired confidence level) to be consistent with this null hypothesis.

We consider five prominent global temperature data sets: (i) NASA GISTEMP (Hansen *et al* 2010, GISTEMP Team 2016), NOAA (Smith and Reynolds 2005, Smith *et al* 2008), HadCRUT4 (Morice *et al* 2012), the revision of HadCRUT by Cowtan and Way (2014), and the Berkeley Earth Surface Temperature (Rohde *et al* 2013). All are combined land+ocean series. Some of these (GISTEMP, Cowtan and Way, Berkeley Earth) aspire to provide a full global mean, by interpolating into some data-sparse regions of the globe, most notably the Arctic. The others simply ignore data gaps and average only over the data-covered part of world, which is systematically biased relative to the true global mean if the data-sparse regions deviate from global mean warming (which is well-documented for the Arctic, which recently has warmed two to three times faster than the globe).

3. Pitfalls in tests for trend change

Before we proceed to our change point analysis of global temperature trends up to the present, we discuss two important but underappreciated pitfalls waiting to trap those testing for a change in the trend of global temperature time series. These are the *multiple testing problem*, and failure to account for the additional degree of freedom when one uses a model with a jump discontinuity (a *broken trend*). We discuss each in turn.

3.1. Multiple testing problem

It is straight-forward to test by Monte Carlo simulations how likely it is that a linear warming trend as low as during some specified time interval (e.g. the interval 2001–2014 defined as ‘slowdown period’ by Fyfe *et al* (2016)) would occur under the null hypothesis, i.e. assuming a continuation of the same linear trend and variance found in a previous baseline period. We use a baseline period starting in 1972, as ~1972 marks the beginning of the most recent approximately linear phase of global warming as

identified objectively in the change point analysis of Cahill *et al* (2015). We first determine the linear trend and standard deviation of global temperature during the baseline period (see table 1). Subsequently we perform Monte Carlo simulations by generating 10000 realisations of time series consisting of this same linear trend plus white (Gaussian) noise with the standard deviation found in the baseline period. Note that the choice of uncorrelated white noise is (a) justified since the observed variability does appear to be close to white, and (b) conservative in the sense that any autocorrelation in the noise would make it more likely to obtain trends by chance that deviate strongly from the baseline trend, so that using auto-correlated noise would make it harder to reject the null hypothesis.

Table 1 lists the percentage of Monte Carlo simulations that show at least one interval of the same length with a trend at least as low as that found during two alleged slowdown periods in the observational data, i.e. 2000–2012 and 2001–2014. We show these results for two data sets: GISTEMP as it is typical for data sets including (partly interpolated) coverage of the whole globe, and HadCRUT4 as the extreme case of a data set with a large gap of missing data in the Arctic (the region of most rapid recent warming), which leads to particularly low recent warming trends (Cowtan and Way 2014).

For ‘slowdown period’ 2001–2014 we tested how many of 10000 Monte Carlo realizations of 43 yr of data (1972–2014) show at least one 14 yr interval with a trend as low or lower than 2001–2014. For ‘slowdown period’ 2000–2012 we tested how many of 10000 Monte Carlo realizations of 41 yr of data (1972–2012) show at least one 13 yr interval with a trend as low or lower than 2000–2012.

The results show that even for the HadCRUT data, the chances of getting such a low trend as observed during 2001–2014 are 31%, so this ‘slowdown’ is far from significant by any standard. For the GISTEMP data, the chances of finding a period with a trend as low as observed during 2001–2014 are 73%, so there is nothing remotely remarkable about this. If one uses the slightly different time interval 2000–2012 (suggested by some for a possibly significant slowdown)

and GISTEMP, then it would have been a statistically significant event at the 95% level if one had *not* found a trend as low as observed! *That* would have falsified the null hypothesis of ongoing linear warming trend plus noise, not the fact that such a low trend was in fact observed.

Note that conservative assumptions have been made, i.e. white noise was used and the number ‘14’ (looking at 14 yr trends) was taken as a given, although 14 was also chosen after the fact, because a particularly low 14 yr trend appeared. So if investigated more elaborately and rigorously, the likelihood of just by chance getting a slow trend period as observed would be greater still.

It is important to understand that a simple comparison of the trend values of the ‘baseline’ and the ‘slowdown’ periods, finding that their uncertainty ranges do not overlap (Hawkins 2016, Santer *et al* 2000) does not provide evidence for a significant slowdown. That would be the case only if the ‘slowdown period’ were one randomly drawn sample—for one random period it would indeed be unlikely to encounter such a low trend just by chance. For the HadCRUT data the chance would be $< 2\%$, for GISTEMP $< 10\%$. However, the period 2001–2014 was *not* randomly drawn: it was specially selected *because of* its low trend from many possible time intervals. This is a well-known and not uncommon statistical mistake: the failure to account in a significance analysis for the fact that a particular number is not one randomly drawn sample but has been specifically selected because of its value. Consider making a test to check whether two dice are loaded towards producing low numbers, rolling those dice once, and they both roll a one. This indeed is a randomly drawn sample and it would provide some support for the suspicion that the dice are loaded, given that it is an event that has less than 5% probability of occurring just by chance with unloaded dice. But if you roll those two dice many times until you finally find one occurrence of two ones, then this event has no significance whatsoever. There is nothing rare about finding one such event in many trials with unloaded dice. This pitfall is known as the multiple comparison or multiple testing problem (Wikipedia 2016). A common approach to correct for multiple testing is the Bonferroni correction (Dunn 1961).

In summary, there is nothing significant or unusual about the interval of lesser warming trend that started around the turn of the century.

3.2. Broken trends problem

The discussion has so far used broken (i.e. discontinuous) trend lines. This is a further problem of many past analyses, also tending to enhance the (in this case false) impression of a significant slowdown.

Figure 1(a) shows a model with broken trends applied to HadCRUT4 data. A naive statistical analysis

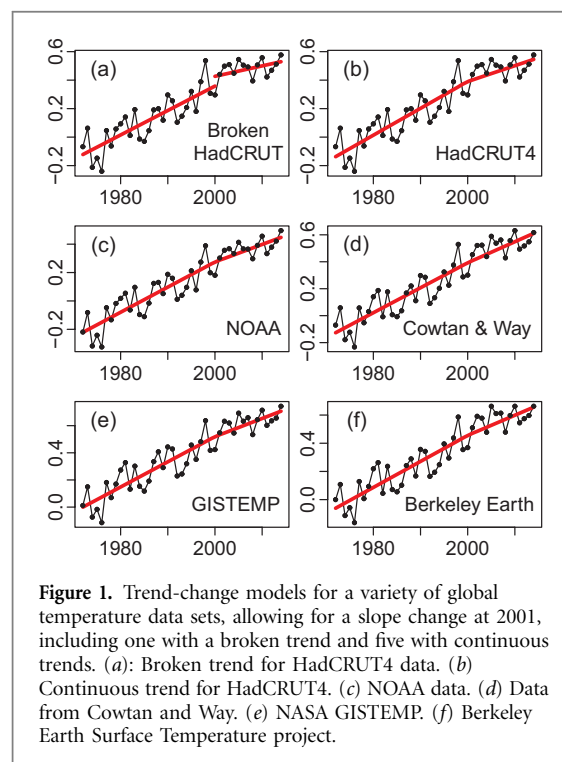


Figure 1. Trend-change models for a variety of global temperature data sets, allowing for a slope change at 2001, including one with a broken trend and five with continuous trends. (a): Broken trend for HadCRUT4 data. (b) Continuous trend for HadCRUT4. (c) NOAA data. (d) Data from Cowtan and Way. (e) NASA GISTEMP. (f) Berkeley Earth Surface Temperature project.

suggests that the change is real because the two linear segments have significantly different slopes.

However, the underlying model includes more degrees of freedom than just a change of slope, it includes a change of intercept as well, which is not accounted for in the naive statistical comparison. The proper approach is to account for both added degrees of freedom, as is done by the *Chow test* (Chow 1960). For HadCRUT4 data it returns a p -value of 0.0635, which fails statistical significance at 95% confidence, while for NOAA data the p -value is 0.295, nowhere near significance. Note that these p -values have not yet included allowance for the multiple testing problem, so that either of the two pitfalls alone invalidates claims of a significant slowdown.

There are also grounds to suspect that the ‘broken trend’ model is unphysical. If instead one allows for a slope change at 2001 but requires the model to be continuous, it yields the models shown in figures 1(b) through (f). None of the continuous trends even gives the visual impression of meaningful trend change, and more important, when tested for significance none achieves even 80% confidence for a trend change in 2001.

4. Change point analysis of recent warming

Finally, we present a change point analysis of global temperature including the latest values for 2015 and 2016. This analysis models a time series as piecewise linear sections and objectively estimates where/when changes in data trends occur. The model forces each line segment to connect, avoiding unphysical

Table 2. Trends and change-point times from analysis of five global temperature data sets.

Data		mean	2.5%	25%	Pct. 50%	75%	97.5%
GISTEMP	Rate 1	−0.005	−0.008	−0.006	−0.005	−0.003	−0.000
	Rate 2	0.014	0.009	0.011	0.012	0.015	0.027
	Rate 3	−0.003	−0.009	−0.004	−0.002	−0.001	0.002
	Rate 4	0.018	0.016	0.017	0.018	0.019	0.021
	CP 1	1915	1910	1911	1914	1918	1930
	CP 2	1943	1938	1942	1943	1944	1947
	CP 3	1969	1964	1967	1969	1971	1975
NOAA	Rate 1	−0.007	−0.011	−0.009	−0.007	−0.006	−0.003
	Rate 2	0.012	0.009	0.011	0.012	0.013	0.015
	Rate 3	−0.003	−0.008	−0.004	−0.003	−0.001	0.002
	Rate 4	0.017	0.015	0.017	0.017	0.018	0.020
	CP 1	1911	1907	1909	1910	1912	1917
	CP 2	1943	1939	1942	1943	1944	1947
	CP 3	1970	1964	1968	1969	1971	1975
HadCRUT	Rate 1	−0.002	−0.003	−0.002	−0.002	−0.001	−0.000
	Rate 2	0.013	0.009	0.011	0.013	0.014	0.018
	Rate 3	−0.002	−0.007	−0.004	−0.002	−0.001	0.001
	Rate 4	0.018	0.016	0.017	0.018	0.019	0.021
	CP 1	1913	1908	1911	1913	1915	1919
	CP 2	1943	1936	1941	1943	1944	1949
	CP 3	1975	1971	1974	1975	1976	1979
Cowtan & Way	Rate 1	−0.001	−0.002	−0.001	−0.001	−0.000	0.001
	Rate 2	0.014	0.010	0.012	0.014	0.016	0.021
	Rate 3	−0.003	−0.006	−0.004	−0.003	−0.002	0.000
	Rate 4	0.019	0.017	0.018	0.019	0.020	0.022
	CP 1	1914	1908	1912	1914	1917	1921
	CP 2	1942	1936	1940	1942	1943	1947
	CP 3	1975	1971	1974	1976	1977	1980
Berkeley Earth	Rate 1	0.000	−0.002	−0.001	0.000	0.000	0.001
	Rate 2	0.016	0.011	0.013	0.015	0.017	0.023
	Rate 3	−0.002	−0.006	−0.003	−0.002	−0.000	0.002
	Rate 4	0.019	0.016	0.018	0.019	0.020	0.022
	CP 1	1916	1909	1913	1916	1918	1923
	CP 2	1942	1937	1941	1942	1944	1949
	CP 3	1975	1970	1973	1975	1977	1982

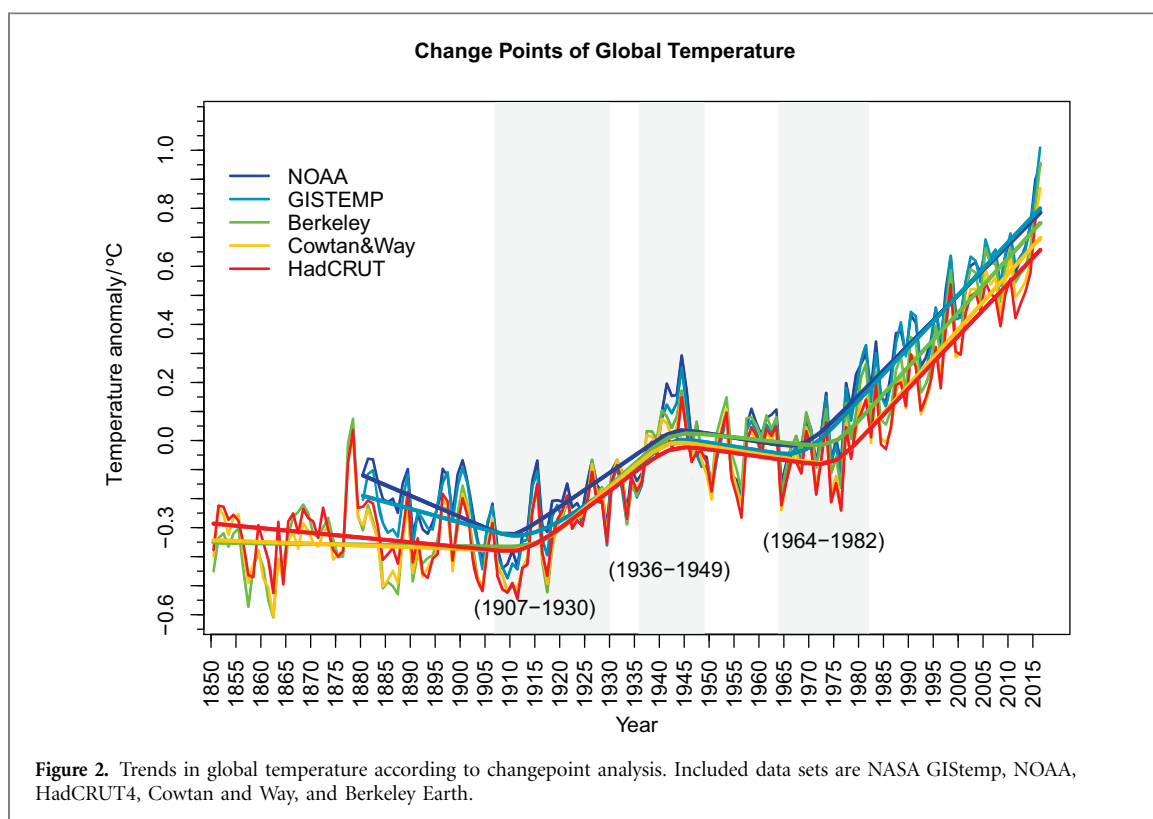
discontinuities. If the data do not support a trend change then such will not be detected. Analysis as in Cahill *et al* (2015) has been extended to include data from 2015 and 2016 in order to investigate whether or not an acceleration has recently taken place; the estimated trend values are shown in figure 2. This question is increasingly asked by journalists, given the third record-hot year in a row.

No recent (post-1980) change-point was found in any of the five data sets, with three change points suitably capturing the climate signal, suggesting that the recent hot years are a continuation of the existing trend, augmented by noise. The 2016 value seems visually extreme, but does not yet provide statistical evidence for a trend change. Of course, future temperature development might provide evidence that an acceleration indeed happened around 2014, but the data up until now do not. Moreover, the extreme heat of 2016 has

almost certainly been enhanced by an El Niño event in the tropical Pacific which is now over, so that we expect a lower temperature again in 2017. Table 2 lists the mean trend for each time span, together with the likely values at a variety of percentiles, as well as the change-point times with their percentiles, for all five data sets.

5. Discussion and conclusions

Short-term fluctuations are unavoidable in global temperature. That episodes will occur which visually seem—sometimes strongly—to represent a change in the underlying trend, is therefore not merely possible, but inevitable. Because fluctuation is ubiquitous, differentiating between genuine trend change and appearances which are merely the manifestation of ‘noise’ is important.



Our purpose has been to determine what can and cannot be said about trends and their changes, based on the temperature data records *only*. We find that the public discussion of time intervals within the range 1998–2014 as somehow unusual or unexpected, as indicated by terms like ‘hiatus’, ‘pause’ and ‘slowdown’, has no support in rigorous study of the temperature data. Nor does recent talk of sudden acceleration based on three record-hot years in a row and the exceptional value in 2016. Both the alleged slowdown and the suspected acceleration are in fact well within the expected range of behavior for a constant trend plus the usual ‘noise’.

The fact that global temperature data do not reveal any significant trend changes since the acceleration in the 1970s does not rule out that subtle trend changes may nevertheless have occurred; it merely shows that these were not large enough to emerge from the ‘noise’ of short-term fluctuations.

By physical arguments, by model simulations, or by correlation analyses with additional data (e.g. El Niño/Southern Oscillation indices or solar forcing data) it is possible to identify specific physical causes of temperature fluctuations, and this is a fruitful topic of ongoing climate research (Foster and Rahmstorf 2011, Kosaka and Xie 2013, England *et al* 2014, Suckling *et al* 2016) which helps us to understand natural climate variability. However, this is distinct from the question of whether a significant trend change has occurred in the temperature data as such. That is not the case. It is unfortunate that a major public and media discussion has revolved around an alleged significant and unexpected slowdown in the

rate of global warming, for which there never was a statistical basis in the measured global surface temperature data.

References

- Arnell N *et al* 2014 Global-scale climate impact functions: the relationship between climate forcing and impact *Clim. Change* **134** 475–87
- Boykoff M T 2014 Media discourse on the climate slowdown *Nat. Clim. Change* **4** 156–8
- Cahill N *et al* 2015 Change points of global temperature *Environ. Res. Lett.* **10** 084002
- Chow G C 1960 Tests of equality between sets of coefficients in two linear regressions *Econometrica* **28** 591–605
- Cowtan K and Way R G 2014 *Q. J. R. Meteorol. Soc.* **140** 1935–44 (Dataset accessed: 10 February 2017) (www-users.york.ac.uk/kdc3/papers/coverage2013/)
- England M H *et al* 2014 Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus *Nat. Clim. Change* **4** 222227
- Foster G and Abraham J 2015 Lack of evidence for a slowdown in global temperature *US CLIVAR* **13** 6–9
- Foster G and Rahmstorf S 2011 Global temperature evolution 1979–2010 *Environ. Res. Lett.* **6** 044022
- Fourier J J 1827 MEMOIRE sur les temperatures du globe terrestre et des espaces planetaires *Memoires d l'Academie Royale des Sciences de l'Institut de France* **VII** 570–604
- Eyre J C *et al* 2016 Making sense of the early-2000s warming slowdown *Nat. Clim. Change* **6** 224–8
- GISTEMP Team 2016 *GISS Surface Temperature Analysis (GISTEMP)* (NASA Goddard Institute for Space Studies) (Dataset accessed: 10 February 2017) (<http://data.giss.nasa.gov/gistemp/>)
- Hansen J *et al* 2010 Global surface temperature change *Rev. Geophys.* **48** RG4004
- Hawkins E 2016 Slowdown discussion. Retrieved from www.climate-lab-book.ac.uk/2016/slowdown-discussion/, permanently archived at www.webcitation.org/6kLGdv8Ax

- IPCC 2013 Climate change: the physical science basis *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, ed Stocker TF *et al* (Cambridge: Cambridge University Press) (<http://doi.org/10.1017/CBO9781107415324>)
- Kosaka Y and Xie S-P 2013 *Nature* **501** 403407
- Lewandowsky S, Oreskes N, Risbey J S, Newell B R and Smithson M 2015 Seepage: climate change denial and its effect on the scientific community *Glob. Environ. Change* **33** 1–13
- Lewandowsky S *et al* 2016 The ‘pause’ in global warming: turning a routine fluctuation into a problem for science *B. Am. Meteorol. Soc.* **97** 723–33
- Mooney C 2013 Who created the global warming ‘pause’? Grist (<http://grist.org/climate-energy/who-created-the-global-warming-pause/>)
- Morice C P *et al* 2012 Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 dataset *J. Geophys. Res.* **117** D08101 (Dataset accessed: 10 February 2017) (www.metoffice.gov.uk/hadobs/hadcrut4/data/current/time-series/)
- Nicholls N 2001 The insignificance of significance testing *Bull. Amer. Meteorol. Soc.* **82** 981–6
- Rahmstorf S *et al* 2007 Recent climate observations compared to projections *Science* **316** 709
- Rajaratnam B *et al* 2015 Debunking the climate hiatus *Clim. Change* **133** 129–40
- Rohde R *et al* 2013 A new estimate of the average earth surface land temperature spanning 1753 to 2011 *Geoinfor. Geostat: An Overview* 1.1 (Dataset accessed: 10 February 2017) (www.scitechnol.com/2327-4581/2327-4581-1-101.pdf)
- Santer B D *et al* 2000 Statistical significance of trends and trend differences in layer-average atmospheric temperature time series *J. Geophys. Res.* **105** 733756
- Smith T M and Reynolds R W 2005 A global merged land air and sea surface temperature reconstruction based on historical observations (1880–1997) *J. Clim.* **18** 202136 (Dataset accessed 10 February 2017) (www1.ncdc.noaa.gov/pub/data/noaa/globaltemp/operational/timeseries/)
- Smith T M *et al* 2008 Improvements to NOAA’s historical merged land-ocean surface temperature analysis (1880–2006) *J. Clim.* **21** 228393
- Suckling E B *et al* 2016 An empirical model for probabilistic decadal prediction: global attribution and regional hindcasts *Clim. Dyn.* **48** 3115–38
- Wikipedia 2016 *Multiple comparisons problem* (https://en.wikipedia.org/wiki/Multiple_comparisons_problem)