A fluctuation in surface temperature in historical context: reassessment and retrospective on the evidence

To cite this article: James S Risbey *et al* 2018 *Environ. Res. Lett.* **13** 123008

Environmental Research Letters

TOPICAL REVIEW

A fluctuation in surface temperature in historical context: reassessment and retrospective on the evidence

James S Risbey[1], Stephan Lewandowsky[1,2], Kevin Cowtan[3], Naomi Oreskes[4], Stefan Rahmstorf[5,6], Ari Jokimäki[7] and Grant Foster[8]

[1] CSIRO Oceans & Atmosphere, Hobart Tas, Australia
[2] University of Bristol, Bristol, UK and University of Western Australia, Crawley, WA, Australia
[3] Department of Chemistry, University of York, York, United Kingdom
[4] Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, United States of America
[5] Potsdam Institute for Climate Impacts Research, Potsdam, D14473, Germany
[6] University of Potsdam, Potsdam, Germany
[7] Independent researcher, Finland
[8] 1 Noyes Ct, Augusta, ME 04330, United States of America

E-mail: james.risbey@csiro.au

Keywords: climate variability, climate trends, temperature fluctuation, pause hiatus

OPEN ACCESS

RECEIVED
8 February 2017

REVISED
9 November 2018

ACCEPTED FOR PUBLICATION
23 November 2018

PUBLISHED
19 December 2018

Abstract

This work reviews the literature on an alleged global warming 'pause' in global mean surface temperature (GMST) to determine how it has been defined, what time intervals are used to characterise it, what data are used to measure it, and what methods used to assess it. We test for 'pauses', both in the normally understood meaning of the term to mean no warming trend, as well as for a 'pause' defined as a substantially slower trend in GMST. The tests are carried out with the historical versions of GMST that existed for each pause-interval tested, and with current versions of each of the GMST datasets. The tests are conducted following the common (but questionable) practice of breaking the linear fit at the start of the trend interval ('broken' trends), and also with trends that are continuous with the data bordering the trend interval. We also compare results when appropriate allowance is made for the selection bias problem. The results show that there is little or no statistical evidence for a lack of trend or slower trend in GMST using either the historical data or the current data. The perception that there was a 'pause' in GMST was bolstered by earlier biases in the data in combination with incomplete statistical testing.

> 'Nowadays any reference to polywater is always tinged with ridicule, but ten years ago many competent and experienced scientists were quite convinced of its reality. I can see no reason why the scientific and sociological issues raised by this unique episode should be shrouded in secrecy'—Felix Franks, Polywater.

1. Introduction

The Earth's climate varies on a vast range of temporal scales (National Research Council 1982, 1995). The persistent increase in greenhouse gases since the industrial revolution is imposing climate changes on timescales from decadal to centennial, and ultimately much longer too as the oceans and cryosphere respond to the changes in Earth's energy balance (Hansen *et al* 1985, Houghton *et al* 2001). The detection and attribution of greenhouse climate change (Mitchell *et al* 2001) deals with the identification of the 'signal' of the forced response to greenhouse gases from the 'noise' of variability of climate that occurs on the same decadal and multidecadal timescales. The greenhouse climate signal is always accompanied to some degree by 'noise' (variation) from other forcings of the climate system (such as due to changes in aerosol loading or solar variations) (Marotzke and Forster 2015) and by internal variations intrinsic to the coupled climate system (O'Kane *et al* 2013).

In recent years there have been more than two hundred articles in the climate literature discussing the notion of a 'pause' or 'hiatus' in greenhouse warming that is variously alleged to have taken place some time in the past couple of decades (Lewandowsky *et al* 2016). The form of alleged climate 'pause' varies across the literature, but essentially involves calculation of a short-term trend in global mean surface temperature (GMST) over a decade or two, which is then compared with either other periods in observed GMST (Stocker *et al* 2013), or with trends estimated from coupled climate model projections (Fyfe *et al* 2013, Risbey *et al* 2014). This review addresses the former issue (comparison of observed trends), while a companion review (Lewandowsky *et al* 2018) addresses the comparison with climate model expectations of trends.

When it first emerged, the concept of a global warming 'pause' was mostly cast in terms of the observational record as a period of slower than average warming (e.g. Stocker *et al* 2013). With time, usage broadened to include a comparison of observed warming rates with those inferred from model projections. The observations-based view of the 'pause' is perhaps more intuitively accessible, whereas the model-comparison view of the 'pause' allows for more complexity in matching variations in the forcing and the (model simulated) response to that forcing with observed trends. Neither definition turns out to be straight-forward in practice. We concern ourselves exclusively here with the first (observations-based) view of the 'pause'. As such, we do not consider the role of climate forcing, and we do not conduct any analysis directed at a causal understanding of fluctuations in GMST (which would require the use of climate models). We do not discount the worth of the model-comparison view of the 'pause', but the issues are complex enough that they require separate examination (Lewandowsky *et al* 2018). While the observations-based view of the 'pause' is intuitively appealing in that one can ostensibly 'see' a slowing of warming rate in (parts of) the GMST record, mere description is not the same as statistical evidence. The complexity here lies in the choices of data, periods, and tests employed to quantify whether any part of the record is indeed unusual. This review attempts to foreground some of those choices and their consequences.

The notion that global warming 'paused' is now entrenched in the journal literature (Stocker *et al* 2013). The 'pause' in warming is generally posited in this literature as an anomaly about climate that is inconsistent with rising greenhouse gases. Many pause-papers commence with the statement that despite rising levels of greenhouse gases, GMST has not increased since about 1998 (although the supposed start year varies) (Guemas *et al* 2013, Kosaka and Xie 2013, England *et al* 2014, Santer *et al* 2014). This alleged prima facie inconsistency is employed as one of the prime motivations in papers on the pause (Lewandowsky *et al* 2016). This review assesses the evidence for the 'pause' in the observed GMST record, as it is now, and as it was at the time the research was undertaken.

The review provides initial context by describing temperature fluctuations in the climatological literature and some issues in constructing observed series of GMST. The consequences of the uncertainties in GMST are described for assessment of short-term GMST trends. The review then proceeds by providing a series of retrospective constructions of short-term GMST trends on the basis of what was known about uncertainty in each of the major GMST series at different points in time compared to what is known now about uncertainty in each of these series. This retrospective analysis provides a framework in which to assess what was known (or could have been known at the time) when assessing the evidence for a 'pause' in global warming. The retrospective (historical) assessment of trends uses the versions of the GMST data that existed at the times when researchers carried out their assessments of trend-intervals.

Because the literature on the 'pause' is now so vast, the review treats the literature primarily as a database for statistical assessments. The sets of definitions implied for the 'pause' can be inferred from the pause-literature, which provides the range of intervals against which to assess potential pauses. We have attempted to summarise some of the key messages and the approach to statistical methods in this literature, but do not provide a chronological assessment of individual contributions in the sense more common in reviews. Our concern is with the definitions, data, and methods used and their implications for the conclusions drawn.

2. Climate fluctuations past and present

The field of climatology has long recognized that climate varies on decadal and longer time scales. The concept of a 'climate normal' was introduced in the early 20th century as a 30 year record or average of climate (Arguez and Vose 2011). The 30 year period was considered necessary to smooth out at least some of the known large decadal-scale variations in climate. The various GMST datasets have used a 30 year 'climate normal' period as a baseline against which to calculate anomalies for similar reasons. The literature on climate variability and change has recognized episodes or periods of multidecadal GMST variation throughout the 20th century (Handel and Risbey 1992, National Research Council 1995). Thus, the notion of fluctuations in GMST is not new and has been recognized as a confounding factor in attributing causes of decadal-scale GMST changes in all the IPCC reports since their inception in 1990. For example, the 1990 report (Houghton *et al* 1990) noted that

> 'Because of long period couplings between different components of the

> climate system, for example between ocean and atmosphere, the Earth's climate would still vary without being perturbed by any external influences. This *natural variability* could act to add to, or subtract from, any human-made warming.' (Here, the reference to unforced 'natural variability' is equivalent to 'internal variability' in modern usage.)

And the 1995 report (Houghton *et al* 1996) noted that for projections of climate change,

> '…decadal changes would include considerable natural variability.'

And that

> '…natural climate variability on long time-scales will continue to be problematic for CO_2 climate change analysis and detection.'

2.1. Present view of the present fluctuation

Given that climatologists were well aware that GMST fluctuates on decadal (and longer) time scales, the emergence of a claim in the climate literature from about 2009 that climate change as represented by GMST had entered a 'pause' or 'hiatus' was a strong claim. In effect, the claim was that the most recent decadal-scale fluctuation in GMST was somehow extraordinary or substantially different from past GMST fluctuations. This interpretation is consistent with the fact that the fluctuation was given a name ('pause' or 'hiatus') and with the claim frequently made in pause-papers that *this fluctuation* (but not others) was not consistent with the GMST response to increases in greenhouse gases (Lewandowsky *et al* 2016).

In order to assess the claims made about this particular fluctuation in the literature, we identified a set of 224 peer reviewed articles in the climate literature (through 2016) that referred to a 'pause' or 'hiatus' in GMST in the title or abstract. From this larger set, we constructed a subset of papers that defined a start and end date for any alleged pause, and which specified the GMST data used for analysis. This is the minimum amount of information needed to reproduce and test the claims of a 'pause' in these papers. The application of these criteria reduced the subset to 90 papers, which is the analysis subset used here and denoted 'pause-papers'. The number of papers published each year on the 'pause' is shown in figure 1(a) and rises substantially from 2013. The 'pause-research period' (as reflected by published papers) extends from about 2010 through the present.

Note that the 90 'pause-papers' are the subset that refer to a climate 'pause' and that provide sufficient information to reconstruct the nominal notion of the pause for that paper (the period used and the GMST dataset(s)). Many of these papers presuppose the existence of a 'pause' and address issues that are conditional upon its existence, without necessarily providing their own analysis or evidence for the identified 'pause'. The purpose of this literature set is to allow us to develop a picture of what the GMST pause is presumed to be in the literature, capturing areas of diversity and commonality. Further, the set of 'pause-papers' allow us to be inclusive in capturing all the different definitions used for the pause in GMST in our analysis here. The set of papers are listed in the appendix.

There is no single or dominant definition of the 'pause' in the literature (Lewandowsky *et al* 2015b). Many papers are not explicit about the period used to assess the pause or the criteria used to reach the conclusion that there is a pause. The distribution of start dates from the pause-papers (set of 90) for the 'pause' is shown in figure 1(b). These span a range from 1995 to 2004 illustrating the lack of consensus on this issue. Further, there is usually little or no statistical justification offered for choice of the start-year. This is a critical issue which we return to in section 3.4. Similarly, the durations presumed for pauses in the pause-papers span a range from about 10 to 20 years with a median of 15 years (figure 1(c)). The number of times a given year falls into the period defined as a pause across all the pause-papers is shown by the histogram in figure 1(d). The frequency profile of the histogram reveals a 'pause-period' in the literature spanning roughly 1998–2015.

The pause-period was selected by the authors of pause-studies to correspond to a period where the rate of warming is slower than the average longer-term warming rate. This period can be highlighted and placed in context by showing a sliding sequence of short-term trends in GMST through the modern period (figure 2). By colour-coding the trends as red/blue according to whether they are warming faster/slower than the longer-term rate of warming, it is apparent that there are persistent periods of faster and slower than average warming. The pause-period in the pause-literature shows up as the second slower than average warming period on the plot. The identification of a period of slower than average warming does not suffice to demonstrate that such a period is statistically unusual. For that more formal criteria would need to be applied.

Different criteria have been used to constitute a 'pause' in the pause-papers. Most early papers employ it in a manner consistent with the common sense usage to signify an absence of a warming trend (no trend). Later papers, however, often use it to signify a reduction in the warming trend, i.e. a slower than normal trend. This shift in definitions, by itself, might

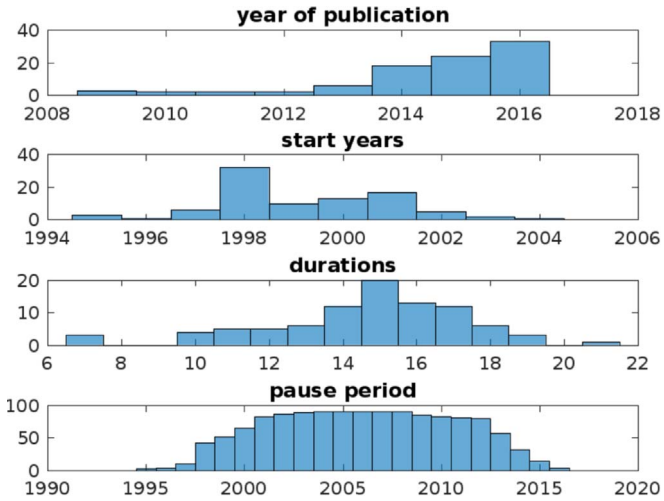

Figure 1. Histograms summarising characteristics of 'pause' definitions in the literature. Panel (a) shows the number of pause-papers published in the peer-reviewed literature each year between 2009 and 2016. (b) Is a histogram of the set of start dates for the pause-period inferred from the pause-papers. (c) Is a histogram of the durations of the pause-periods inferred from the pause-papers. (d) Is a histogram of the pause-period, which shows the number of times each year in the year axis is included in the pause-period across all the pause-papers.

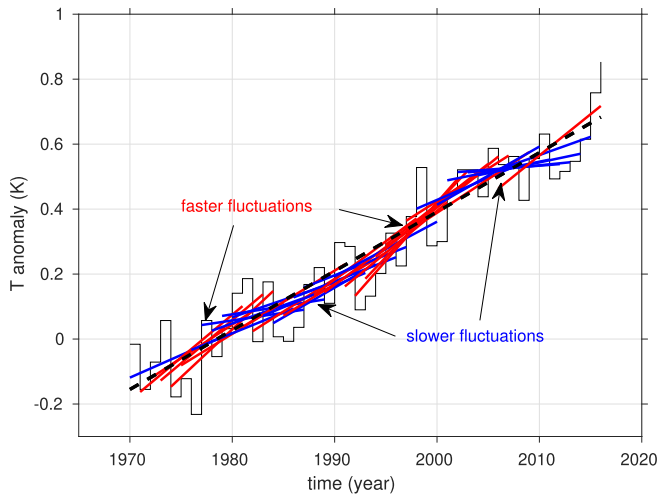

Figure 2. Time series of annual GMST anomaly from Cowtan and Way (stair plot). The black dashed line is a linear fit over the period. The thin red and blue lines are linear 11-year trend lines sliding over the period. These lines are red/blue when the slope of the 11-year fit is greater/less than the slope of the longer-term dash line. The choice of interval length here (11 years) is arbitrary, but all interval lengths used in the pause-papers will exhibit periods of faster and slower than average trend.

indicate a problem, as it shows that even at the time, the scientific community was unclear and inconsistent as to what the object of study was. In this paper we test both claims. To illustrate these definitions we have redrawn figure 2 in idealised form in figure 3. Here, we represent the GMST series (without interannual variation) in its idealised form as undergoing regular fluctuations about a long-term mean warming rate (the dashed black line). The fluctuating line is again coloured red when the trend is greater (warming faster) than the longer-term mean trend and blue when it is smaller (warming slower) than the mean trend. One expects short-term trends to fluctuate faster and slower through time than the longer-term trend as illustrated here. There has been little research attention on the faster fluctuation that preceded the slower fluctuation that is the target of the pause-papers (Rahmstorf *et al* 2007, Lewandowsky *et al* 2015a).

The 'slow' trend view of the pause (figure 3) is seldom defined formally in the pause-literature. It could refer (as in Stocker *et al* 2013) to a meaningful change in the trend (slope) of GMST in the pause-period relative to the longer-term trend that prevailed prior to the pause-period (*change in trend*). Alternatively, it could refer to a claim that the trend during the pause-period is unusual relative to trends of a similar length during the modern warming period (*unusual trend*). For example, the pause-period fluctuation in figures 2 and 3 could be assessed against slower than average fluctuation periods such as the prior one in the 1980s. We

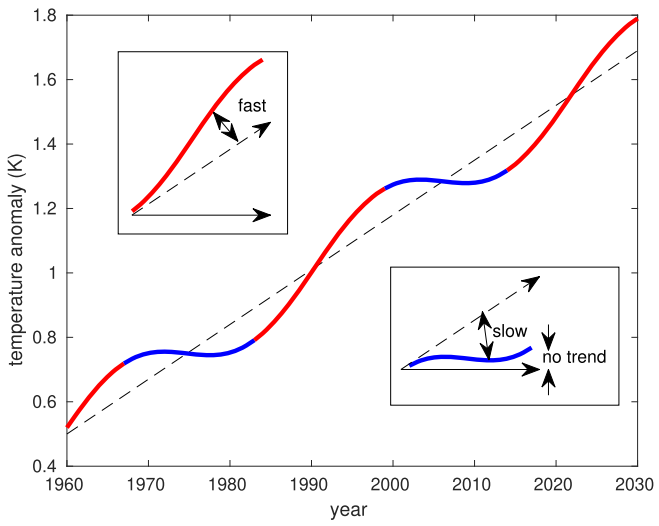

Figure 3. Idealised schematic of a smoothed global mean surface temperature series (blue/red line). The series is a linear trend plus sinusoidal variation to mimic multidecadal fluctuations. The dashed line is the linear fit component. The series is coloured red/blue when the local gradient is steeper/shallower than the linear fit. The inset boxes show a red segment where the slope is compared to the long-term linear slope, and a blue segment where the slope is compared to either the long term linear slope or to zero slope.

will restrict this comparison to fluctuations that occur through the period that GMST has been fairly steadily increasing (with fluctuations) to avoid including a large sample of the early record when the longer-term warming trend was much weaker. An objective way to determine how far back to include past fluctuations is to assess the GMST record for meaningful 'change-points' in trend (Cahill *et al* 2015). We have performed change-point analysis on each of the GMST records used here and find changes in each dataset near 1970. This is consistent with other analyses and with the choice often made in the literature to define the modern warming period. In all the analyses to follow we use the change-points (near 1970) particular to each dataset in assessing how unusual the recent slower fluctuation is.

3. Methods and data

The data used to construct records of GMST consist of a diverse set of observations of temperature collected over land (typically surface air temperature) and oceans (typically sea surface temperature (SST)) through time. The construction of GMST series requires the blending of these observations and removal of any known biases (debiasing) in the data (Karl *et al* 1989, Jones *et al* 1999). These efforts have been carried out principally by groups in the US and UK, and provide estimates of GMST back into the 19th century.

The time series of GMST from five of the principal groups constructing records of GMST is shown in figure 4. The five series exhibit clear variability at inter-annual and decadal to multidecadal scales, with a long-term warming trend. While there are some differences between the series as represented by the five different datasets here, they display very similar variability and long-term trends. As such, the differences between the datasets have historically been more of interest to specialists in the field, as they yield very similar views of the climate response to greenhouse gases.

More recently, the literature on the alleged 'pause' in GMST has brought about a shift in focus to consider short-term trends in this data (of typically 10–15 years duration). Short-term trends can be quite sensitive to small differences in end points in trend intervals, and thus the small differences between the GMST datasets can matter in determining trend magnitudes (Risbey and Lewandowsky 2017). All of the GMST data sets are evolving over time as they better account for measurement errors (Brohan *et al* 2006), extend coverage, add or change interpolation methods, and implement improved bias reduction on past data. We do not provide a review of these issues here, but we do single out a couple of issues that have played a role in assessing short-term trends over the pause-period. These are data coverage (Cowtan and Way 2014), and the bias reduction of SST data (Karl *et al* 2015, Hausfather *et al* 2017, Kent *et al* 2017). Improvements related to coverage and SST debiasing of the data over the past decade have resulted in changes to estimation of recent trends in some of the datasets. We provide an assessment of these issues here as it relates to claims that the recent temperature fluctuation represents a 'pause' in greenhouse warming.

Another critical issue in characterising short-term climate trends is their statistical treatment. Here again we single out two particular issues that have effectively confounded claims in the pause-literature about the prominence or otherwise of short-term GMST trends. The issues relate to the selection of a short interval to analyse the GMST trend that seems to depart from the

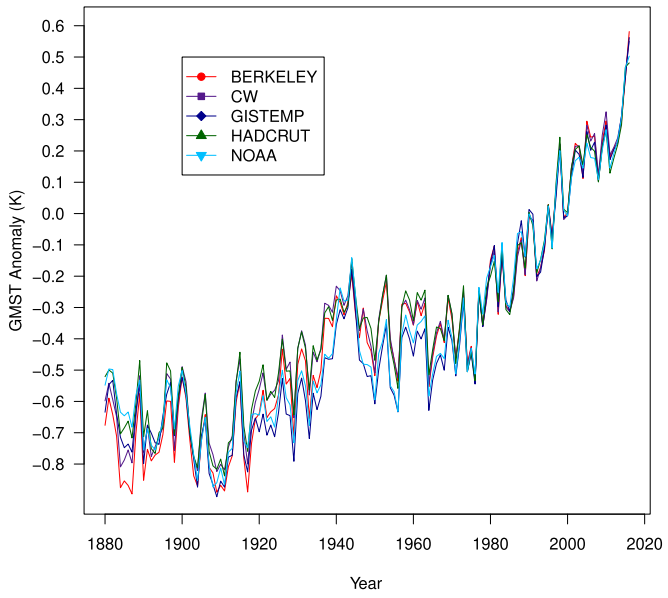

Figure 4. Annual mean global mean surface temperature series for each of the datasets shown based on versions of each set at the end of 2016. The baseline-period for calculating anomalies in each data set is 1981–2010.

Table 1. Data sets used to represent GMST and some characteristics related to data coverage. The global coverage is calculated for each dataset during 1981–2010 and is the average percentage of the global surface covered by grid cells with data. The release dates given are for when the data was available to the public. If no release date is given, the data set had been in use well before the period of research on the 'pause'. The number of versions of each dataset available for this research are Berkeley (7), Cowtan and Way (2), GISTEMP (113), HadCRUT (7), and NOAA (31).

Data set	Release	SST data	Coverage	Ship-buoy bias
Berkeley	Mar 2014	HadSST3	100%	Partially corrected[a]
Cowtan and Way	Nov 2013	HadSST3	100%	Partially corrected[a]
GISTEMP		HadISST1+OISST pre 2013 ERSSTv3 til mid 2015 ERSSTv4 after mid 2015	99.3%	Corrected mid-2015
HadCRUT3		HadSST2	82.0%	Uncorrected
HadCRUT4	Nov 2012	HadSST3	85.4%	Partially corrected[a]
NOAA		ERSSTv3 til mid 2015 ERSSTv4 after mid 2015	91.0%	Corrected mid-2015

[a] (Hausfather *et al* 2017).

long-term trend. To be fair in this comparison, one must properly account for the selection process (the 'selection bias' issue) and whether the trend in the interval is continuous or broken relative to neighbouring intervals (broken trends) (Rahmstorf *et al* 2017). These issues are described below and we assess their role in the interpretation of short-term GMST trends and the 'pause'.

3.1. Datasets

The main data for this review of short-term GMST trends are the five GMST series (as they existed at different points in time) that formed the basis of the trend assessments for papers on an alleged 'pause'. The datasets and some of their properties are described in table 1.

All of these datasets have undergone some forms of bias reduction effort over the past decade during which the climate community has focused on short-term GMST trends. This means that different versions of the data were in play at different times. The Berkeley (Rohde *et al* 2013), Cowtan and Way (Cowtan and Way 2014), and HadCRUT data use HadSST for their SST component, which was bias reduced in going from HadSST2 (Rayner *et al* 2006) to HadSST3 (Kennedy *et al* 2011). This change corresponded to the version change from HadCRUT3 (Brohan *et al* 2006) to HadCRUT4 (Morice *et al* 2012). Similarly GISTEMP (Hansen *et al* 2010) and NOAA (Smith *et al* 2008) use ERSST, which underwent bias reduction work from ERSSTv3 (Smith *et al* 2008) to ERSSTv4 (Huang *et al* 2015). The bias reduction in these SST datasets related to a range of issues including changes in ship-based SST measurement and coverage, and the increased role of buoy records (Karl *et al* 2015, Hausfather *et al* 2017). The earlier (less well bias-reduced) versions of

the SST records have a cool bias during the recent period, which does affect the magnitudes of short-term trends (section 4.1).

The differences in global coverage among the datasets (see table 1) also matter for determination of short-term trends. That is because the differences relate substantially to whether the Arctic region is well represented (Berkeley, Cowtan and Way, GISTEMP) or not (HadCRUT, NOAA), given that the Arctic has been warming fast enough relative to the global mean rate over the recent period to make a difference (Benestad 2008, Rahmstorf *et al* 2017). Note that Cowtan and Way is based on HadCRUT, except that Cowtan and Way include data coverage in the Arctic by applying kriging techniques to interpolate into the Arctic (Cowtan and Way 2014). The differences between Cowtan and Way and HadCRUT4 thus provide a direct measure of the role of data coverage, at least over recent decades when observational coverage is sufficient to properly support near-global temperature reconstruction.

The analyses conducted here have been repeated with all six of the datasets shown in table 1. For the sake of presentation we sometimes show results for just GISTEMP and HadCRUT. One reason for this is that we seek to provide a retrospective assessment of what the trends looked like at different points in time, and these two datasets were in use throughout the period of research on the 'pause', whereas some of the other datasets (Berkeley, Cowtan and Way) were only available after the start of this research. GISTEMP and HadCRUT are also good choices for contrasting lower data coverage (HadCRUT) and near complete coverage (GISTEMP). HadCRUT effectively provides the lowest estimate of short-term warming trends throughout the 'pause' research period and so provides a lower bound on what a pause-researcher with no insight into the differences between data sets might infer concerning the GMST trend.

All of the GMST datasets have been truncated here to the period 1880–2016. All data were converted to anomalies using a common reference period of 1981–2010. This reference period is suitable because the different SST records are most consistent over this period, and it avoids the recent changes in ship bias (Kent *et al* 2017, Hausfather *et al* 2017). All trends calculated here are linear trends using least squares regressions. The choice of linear trends matches usage in the literature, and can be justified over the period since about 1970 in which no new change points are detected in the GMST series (Cahill *et al* 2015, 2018). Versions of the GMST datasets were archived for analysis as they existed at different points in time over the 'pause' research period. This allows us to provide a set of 'historical' views of what the GMST trends looked like at different points in time as described in the next section. These data are available at: https://git.io/fAuos.

In some of the analysis here we break the GMST time series into a *baseline-period* and a *pause-period* to perform statistical tests. The baseline-period extends from the start of the modern warming period up to the beginning of each pause-period tested. The modern warming period is assumed to start at the last significant change point detected in each GMST series (Cahill *et al* 2015). These are 1970 for GISTEMP, NOAA, and Berkeley, and 1974 for HadCRUT and Cowtan and Way. A range of different intervals are used to test many different choices of pause-periods, with the range of intervals encompassing the set of periods inferred from the literature in figure 1(d).

3.2. Historical and hindsight trends

We could, and have, examined GMST trends over the pause-period using the latest available GMST data (through 2016 here). We term this view of the trends the '*hindsight*' view since the current (hindsight) GMST data has the benefit of any and all bias reductions that have taken place in the preceding periods. While 'hindsight' provides the best current view of the pause-period trends, the calculation of trends during the pause-period necessarily relied on the versions of the data that were available at the points in time when the research was conducted. To be fair to researchers at any given point in time, we have also calculated a set of '*historically-conditioned*' trends for each of the GMST datasets. The historically-conditioned trends use the versions of each of the datasets that were current at the time the trend was calculated.

The concept of 'historically-conditioned' trends is illustrated for the HadCRUT dataset in figure 5. This figure shows the trend value for trends starting in 1998 and ending at the points in time marked on the *x*-axis (vantage year). The solid line is the series of historically-conditioned trend values and uses only data that was available up to the time of the vantage year. Different versions of the HadCRUT data are indicated by the dots on the trend line. The trend line shows one particularly large jump up (of about 0.05 K/decade) just after 2012, corresponding to the switch from HadCRUT3 to HadCRUT4 (where the ship-buoy bias went from uncorrected to partially corrected (table 1)). The thin lines in figure 5 show the trends calculated back to the earlier vantage years *as if* the newer versions of the data existed in the earlier periods. The difference between the thick historically-conditioned trend line and the thin (hindsight) lines is thus an indication of how the trends change between different versions of the datasets. In this case, the differences between HadCRUT3 and HadCRUT4 are quite marked as indicated by the large differences between the solid and thin lines.

3.3. Continuous and broken trends

In the calculation in figure 5 we performed the trend calculations as traditional in the pause-literature

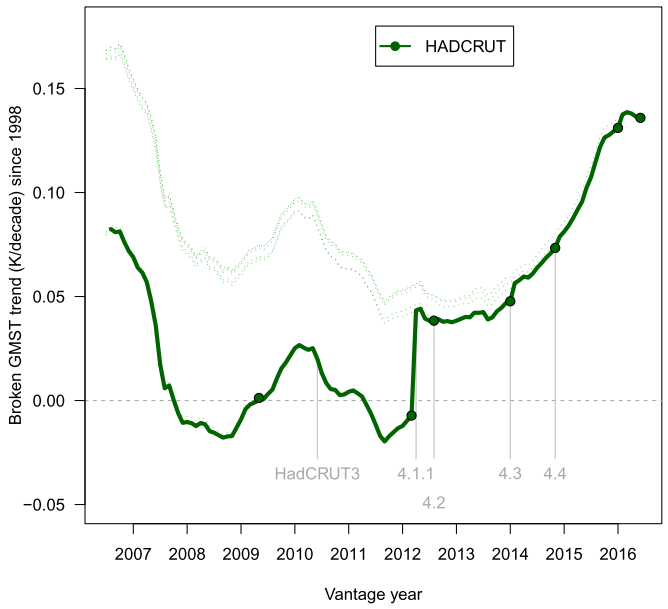

Figure 5. Historically conditioned trends. The solid line is the best fit least squares trend in HadCRUT data from a start year of 1998 to end year (vantage year) as shown on the *x*-axis. The thick line corresponds to the version of HadCRUT current at the current vantage year, with the vertical grey lines indicating which versions applied. The thin dotted lines provide retrospective trends back to trend end year of 2007 to show what the trends would have been if the current version of the data had been available earlier to compute the trend. The green dots indicate when an update to the dataset was available. Only some of the updates constitute new versions. The trends are incremented from monthly data which results in some fine scale variation in the trends.

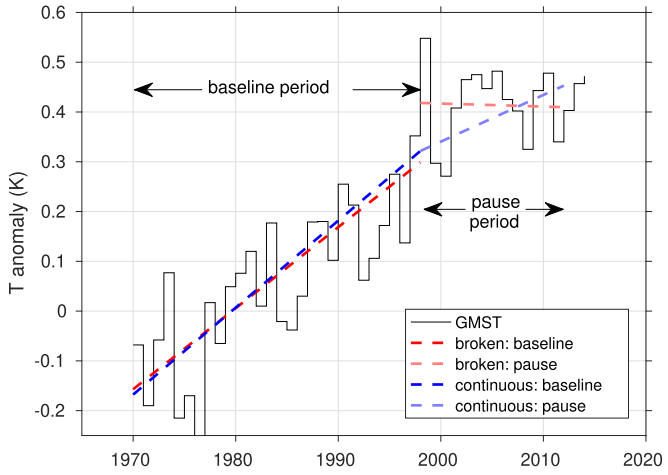

Figure 6. GMST annual mean series for HadCRUT3 (black line). The best fit broken trends pre and post 1998 are shown as red lines: darker red for the baseline-period 1970–1998 and lighter red for the pause-period 1998–2012. The best fit continuous trends about 1998 are shown as blue lines: darker blue for 1970–1998 and lighter blue for 1998–2012.

using a simple trend between a start and end date. However, when the trend is fitted to the data in this way (without regard for the years preceding or following the chosen start and end years respectively), then the isolated trend is *'broken'*, meaning not continuous with, trends in the remainder of the data. This has implications for the testing of the data (Rahmstorf *et al* 2017).

An example of 'broken' and 'continuous' trends in the HadCRUT3 GMST data is shown in figure 6. Here the trend in the pause-period from 1998 to 2012 is shown as a broken-trend (not continuous with prior trends) by the dashed red best-fit trend line over the period. The red broken-trend line for the baseline-period 1970–1998 preceding the broken-pause trend exhibits a jump discontinuity at the common year in 1998. This introduces an extra degree of freedom into the trend analysis which affects the assessment of statistical significance. Such a jump should be explicitly mentioned (e.g. 'temperature jumped upward and then remained flat', rather than just stating it remained flat), and it would normally require some physical justification as to why such a jump in the series should be modelled here (Rahmstorf *et al* 2017). Such allowance and justification is largely absent from the pause-literature that purports to find a pause. A

more parsimonious and physical assumption that does not introduce the extra degree of freedom is to model the trends as continuous trends as shown by the dashed blue trend lines in figure 6. The change in slope of the continuous trend line is much less severe during the pause-period. In testing short-term GMST trends against the hypothesis of 'no trend' we will show results for both broken and continuous trends.

3.4. Selection bias

In the set of pause-literature supporting the notion of a pause in GMST, there is often little or no explanation for why the pause-period used was chosen. While there are differences in the periods chosen to examine GMST for a pause (figure 1), the periods have in common the property that they roughly cover the interval from the late 1990s to early 2010s when GMST was fluctuating with a slower short-term trend than the long-term trend (Risbey *et al* 2015) (as represented by the 'blue' period in figure 3). It is clear from this commonality of pause-periods that the period was not randomly chosen or drawn. Rather, the pause-period was selected (from many possible time intervals) because of its lower trend (Rahmstorf *et al* 2017), as evident in figure 2. Any analysis of the significance of such a period must take into account the fact that it was selected on the basis of its value rather than randomly drawn. This is the 'selection bias' problem. This problem is not accounted for in any of the pause-papers that claim to have found a significant slowdown of warming. Since frequentist hypothesis testing requires a sampling plan that is 'blind' to the nature of the data, selecting a subset of data on the basis of its value and then testing it will have the effect of artificially raising the presumed significance of the pause-periods chosen.

Appropriate corrections for selection bias are described in Rahmstorf *et al* (2017) and performed as one variant of the analysis here. The selection bias problem is referred to as 'multiple testing' (Ventura *et al* 2004, Wilks 2006) in Rahmstorf *et al*, since overcoming the bias in the selection period requires one to perform multiple tests for different start and end times of the tested period.

The procedure used to account for selection bias must address the issue that we have only a half century or so of relatively enhanced greenhouse warming (the modern warming period), and thus few samples from which to test the unusual nature of the pause-period. The remedy applied here is to generate Monte Carlo samples from the modern period as follows: for each of the five datasets the longer-term (baseline) trend is fitted to the period post the change-point determined circa 1970 for each dataset up to the start of each pause-period selected (usually close to 1998). The standard deviation of the residuals about the baseline fit is calculated. We then generate synthetic realisations of GMST over the period encompassing both the baseline and the pause-period using the same linear trend as the baseline-period plus white noise with the same standard deviation as the residuals[9]. This procedure is repeated to give 1000 synthetic realisations. We then compare the magnitude of the pause-period trend with all trends of the same length that occur (at any time) through the 1000 realisations. We report the result as the percentage of realisations that contain a trend with a magnitude smaller than the pause-period trend. We take the view here that a minimal requirement for a trend to be unusually weak (paused) with this procedure is that fewer than 5% of all realisations sampled contain a less-positive trend interval than the selected pause-period trend. If this is not satisfied, then the trend is not very unusual in relation to what one would expect to find in the case of a constant warming trend superimposed by random interannual variability.

4. Results

4.1. Historically-conditioned and broken trends

The various GMST datasets were updated during the pause-research period, resulting in different views of short-term trends for different versions of these datasets (section 3.2). An illustration of the effect of these changes on short-term trends is shown in figure 7, which plots the magnitude of the trend since 1998 in each of the five datasets. Noticeable jumps in the trend value (solid lines) occur for HadCRUT, NOAA, and GISTEMP as they undergo the bias reductions to SST described in section 3.1.

The period from 2012 to 2014 is of particular interest, since it spans the completion of the 5th IPCC assessment report. The HadCRUT, NOAA, and GISTEMP trends appear to be in good agreement over this period, however this agreement is illusory, because the NOAA and GISTEMP records did not include corrections for the SST bias until late 2014. The apparent agreement in trends arises from HadCRUT4 underestimating the rate of warming due to incomplete area coverage and the remaining datasets underestimating warming due to the uncorrected SST data.

Some reduction in the spread of trend values across the datasets would be expected for later vantage years as the actual trend interval considered is longer. However, this is not the main factor in this case. If one traces the thin lines back to 2007 (showing what the trends would have been then if the later versions of the datasets had been available), the spread in trend values across datasets in 2007 is reduced by about half from

[9] Note that the Monte Carlo simulation was repeated with ARMA lag 1 models as candidates for modelling of the noise. The best-fitting model from the set {white noise, ARMA(1, 0), ARMA(0, 1), ARMA(1, 1)} was chosen on the basis of AIC for each design cell. The results are nearly identical to those performed with the white noise model, and so only those for the white noise model are shown here.

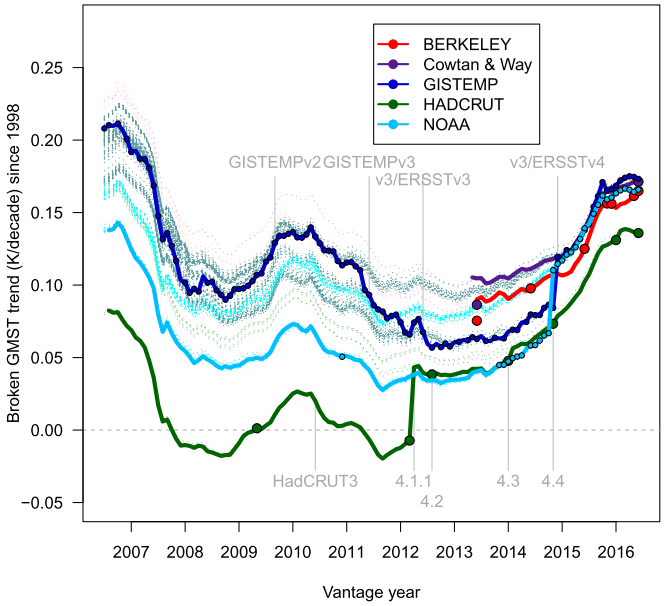

Figure 7. Historically conditioned trends as in figure 5, but for all 5 datasets. The major jumps in the trends occur when HadCRUT shift from HadCRUT3 to HadCRUT4 (green curve) and when GISTEMP and NOAA shift from ERSSTv3 to ERSSTv4 (light blue and navy blue curves). The trends in this figure are all 'broken trends' (section 3.3) in that they start in 1998 without a requirement that the trend start value is continuous with trends before that time.

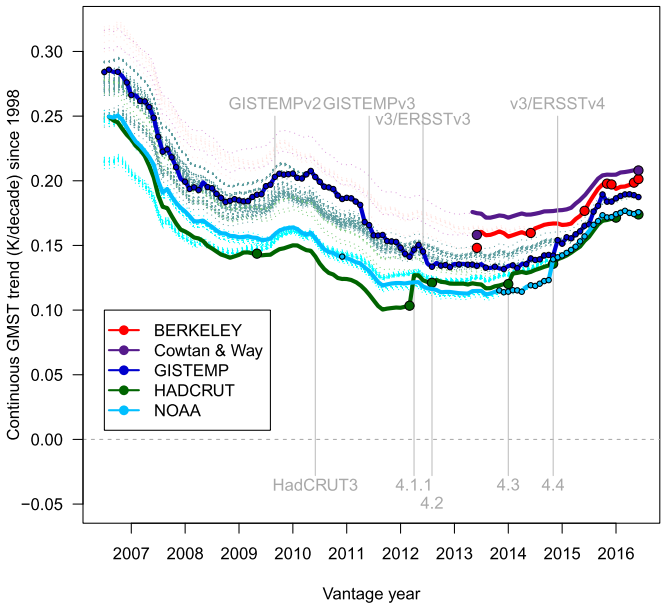

Figure 8. As in figure 7 except that the trends are 'continuous trends'. That is, the trend commencing in 1998 must be continous with trends up to that point.

the historical spread (thick lines) of about 0.1 K/decade to the hindsight spread (thin lines) of about 0.05 K/decade.

The trends shown in figure 7 are all broken trends. The same data and trends are replotted in figure 8 ensuring that all trends are continuous with the prior period (section 3.3). The results change markedly. The trend values are all higher and do not fall below 0.1 K/decade for any dataset (even HadCRUT3) for any vantage year. This is in contrast to the results for broken trends (figure 7) where HadCRUT3 trends are near zero for vantage years from 2008 to 2012. The spread between trend values for earlier versions of the GISTEMP, NOAA, and HadCRUT datasets is substantially reduced in figure 8 using continuous trends. Taken together, the shift to high positive trend values and the reduction in spread across datasets make it clear that use of continuous trends would not have supported the view of GMST 'pausing' at any point in time here, for any dataset.

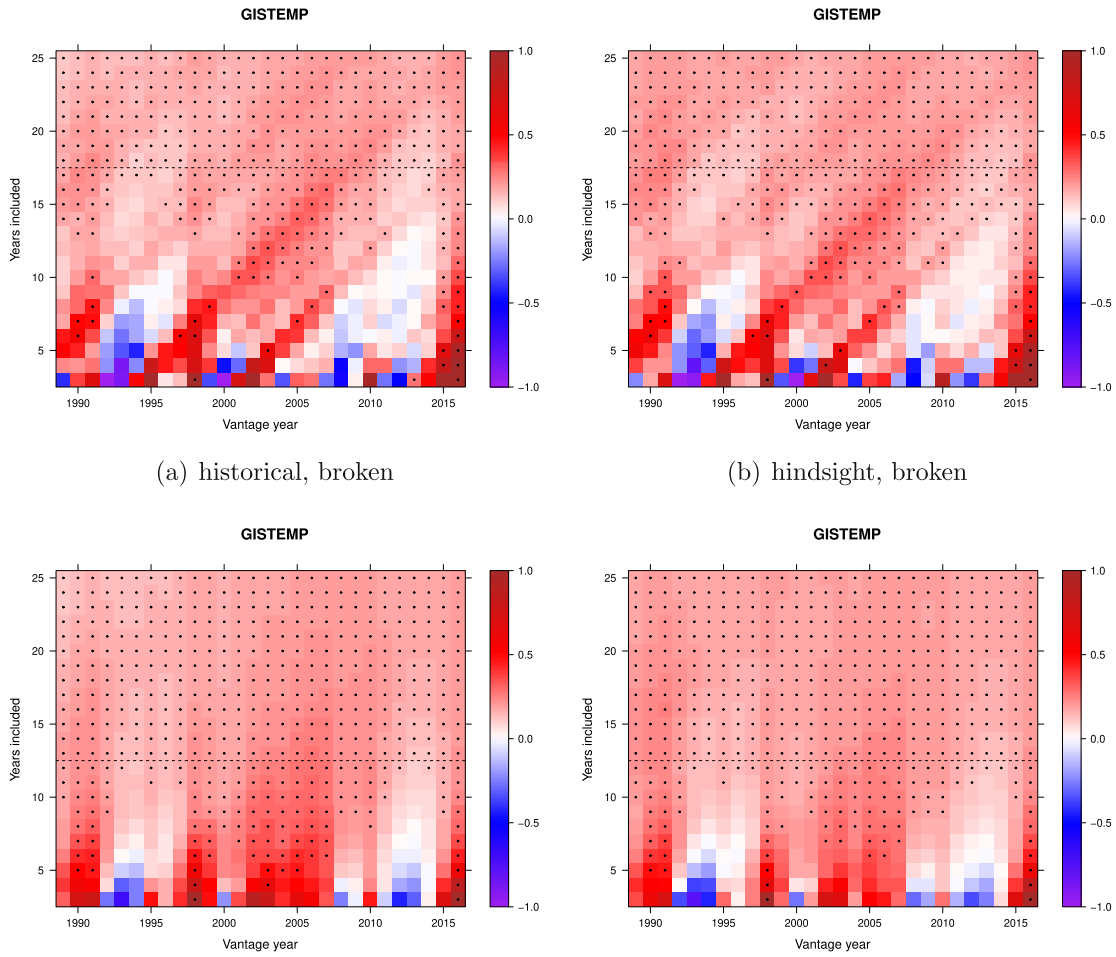

Figure 9. The plot shows the magnitude of trends (K/decade) in GISTEMP GMST (shaded) for a matrix of intervals with duration from 3 to 25 years. Trend magnitudes are capped at ± 1 K for plotting. The last year of the trend interval is given by the 'vantage year' on the x-axis. The number of years in the interval is indicated on the y-axis. For example, the top right corner of the matrix corresponds to the trend of interval length 25 years, ending in 2016. The dots indicate which trends are significant ($p < 0.05$) in an ordinary least squares analysis of annual means. The horizontal dashed line is placed at the interval duration above which trends from all vantage years in the plot are significant. The top left panel (a) is for historical data (the versions of the GISTEMP data that were current for the given vantage year) and broken trends. The top right panel (b) computes all trends from the version of GISTEMP available in 2017 (hindsight) and uses broken trends. The bottom left panel (c) uses historical GISTEMP data and continuous trends. The bottom right panel uses hindsight (2017) GISTEMP data and ensures that all trends are continuous.

4.2. Assessment of no-trend

In this section we provide a systematic assessment from historical and hindsight perspectives on whether one could make a statistical determination that there was no-trend in GMST during the pause-period. For this assessment we show results for both the HadCRUT and GISTEMP data, since both datasets have been heavily used through the pause-research period and HadCRUT provides a lower bound on the magnitudes of pause-period trends.

We assessed a matrix of trends from 3 to 25 years duration from vantage points (i.e. the last year of the trend interval) between 1989 and 2016 as shown in figure 9 for GISTEMP. Note that this set of intervals includes all those used in the pause-literature for the 'pause' along with earlier intervals to provide further context for the pause-period. The earliest vantage year considered here is 1989 so as not to include intervals that are substantially outside the period of modern warming. The colour scale shows positive trends in red and negative trends in blue. It is clear right away that trends less than about 10 years duration are 'noisy' in the sense that they could be of either sign. It is generally regarded that about 17 year intervals are needed to obtain sufficient power to detect a signal in GMST (Lewandowsky *et al* 2015b) or tropospheric temperature (Santer *et al* 2011). The trends significant at the level $p < 0.05$ in figure 9 are represented by black dots in the matrix. If there is no black dot in a square here, one has failed to reject the hypothesis that there is no trend in the data. In the landscape of trends provided by these diagrams one is interested in whether it takes longer to reject the hypothesis of zero trend during the pause-period than at other times.

For GISTEMP using broken trends (figures 9(a), (b)) two major periods of non-significant trend show

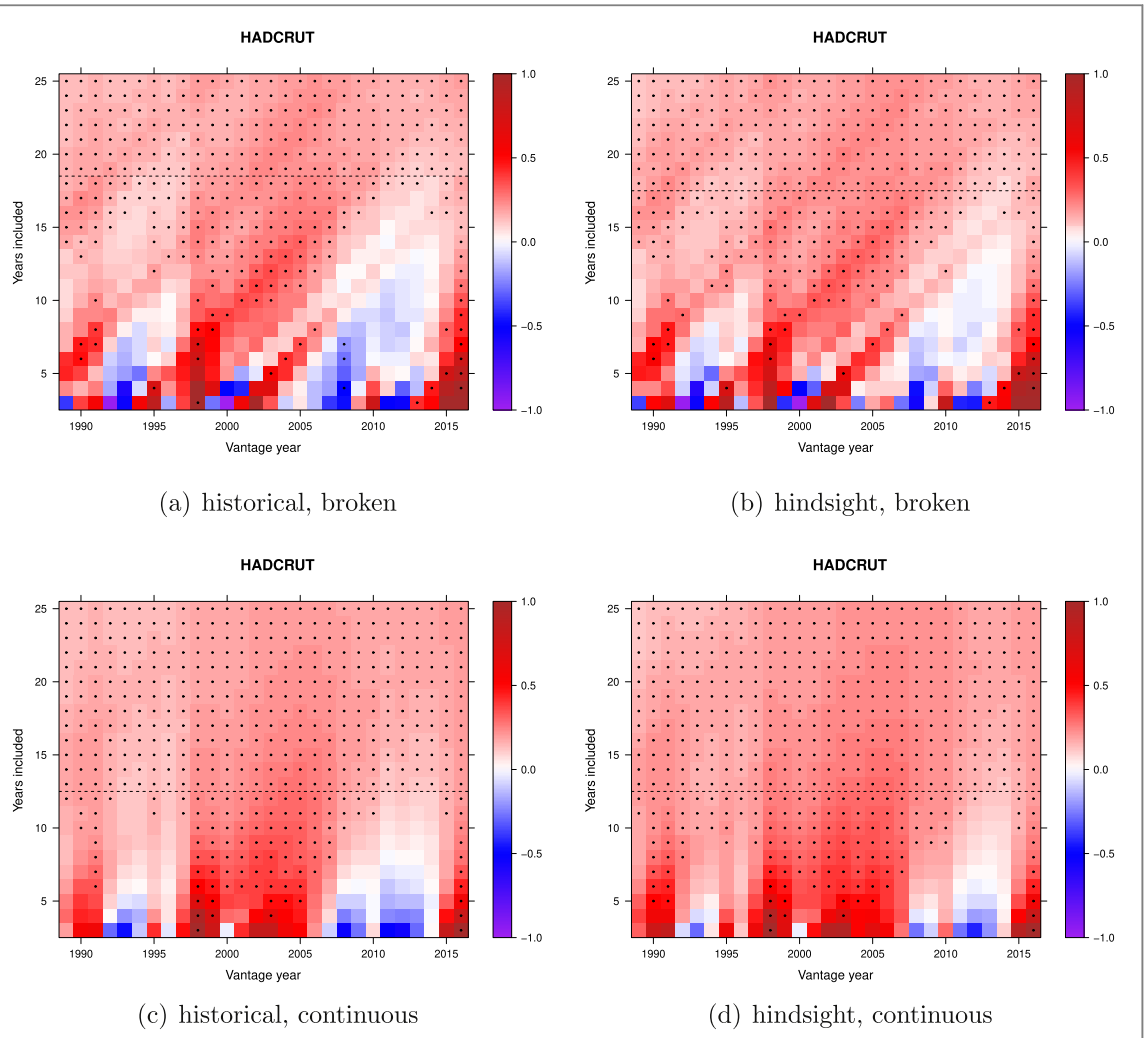

Figure 10. As in figure 9, but for HadCRUT.

up (represented by intervals extending to 17 years to reject the no-trend hypothesis). The second of these periods corresponds to the pause-period. Neither of these periods seem statistically unusual. For GISTEMP there is also little difference in this picture whether one uses historical (Figure 9(a)) or hindsight (figure 9(b)) versions of the data to calculate the trends. When switching from broken to continuous trends (figures 9(c), (d)) any weakly significant trend in GMST is even less pronounced, and it takes only 12 years to reject the no-trend hypothesis during even the slower fluctuation periods.

For HadCRUT (figure 10) the matrix of trend magnitudes and intervals to reject the no-trend hypothesis is broadly similar to that for GISTEMP. That is, with broken trends (figures 10(a), (b)) there are two slower than average fluctuations evident in the matrix of trends, and the time needed to reject zero-trend is similar to GISTEMP and similar using historical or hindsight versions of the data. Further, the use of continuous trends for the analysis (figures 10(c), (d)) reduces the interval needed to reject zero-trends to only about 12 years. In short, the trend matrices for GISTEMP and HadCRUT both show that there are no unusual or remarkable periods where it takes longer than expected to eliminate the no-trend hypothesis. This conclusion is not sensitive to whether one used historical data or not, or whether one used broken or continuous trends.

4.3. Assessment of an unusual trend

The view of the 'pause' as a significant slowing of trend can be assessed either as a change in trend or as an unusual trend. In this section we address the 'unusual trend' definition, and in the following section we address the 'change in trend' definition.

A pause-period trend would be unusual if it were very unlikely to find similar length trends with such a weak magnitude during the modern warming period. The Monte Carlo testing procedure described in section 3.4 has been applied to each of the GMST data sets over a combination of pause-segments of varying lengths and start times sufficient to span the range of pause-definitions found in section 2.1 (figure 1(d)). For each pause-segment there is a baseline-segment spanning the interval from the change point detected in the dataset up to the pause-segment. The Monte Carlo series is generated over the combination of these

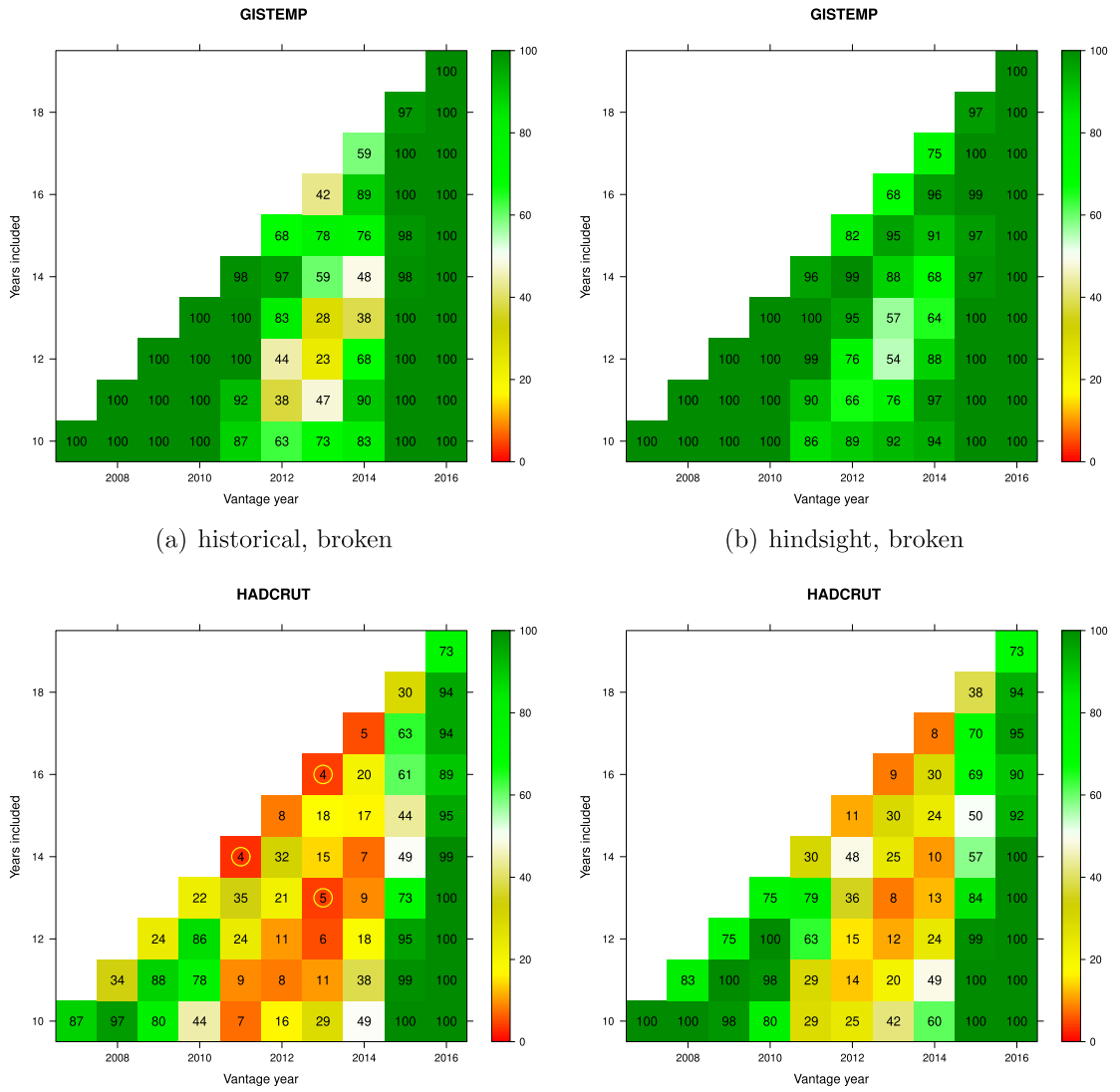

Figure 11. Multiple tests of a matrix of pause-interval trends. Each element of the matrix corresponds to a trend-interval spanning the last year of the interval (vantage year) back through the number of years listed on the *y*-axis. For example, the top right element of the matrix is for the 19 year trend-interval ending in 2016. Each of these trends is compared with a population of trends. The population is generated by taking the baseline-interval from the change point in the dataset signifying the modern warming period to the start of each trend-interval. The residuals to a best fit from this baseline-interval are then used to generate 1000 Monte Carlo realisations of a series from the beginning of the baseline-interval to the end of the trend-interval. The magnitude of the trend-interval is then assessed against a population of all intervals of the same length that occur any time in the Monte Carlo realisations. The shading denotes the percentage of realisations that contain an interval of a lower trend magnitude than the pause-interval trend tested. The numbers in the squares are the actual percentages. Where a yellow circle is present it denotes that fewer than 5% of the realisations in the Monte Carlo sample contain a lower magnitude trend interval than the tested pause-interval trend. For broken trends for GISTEMP using (a) historical data and (b) hindsight data, and for HadCRUT using (c) historical data, and (d) hindsight data.

two periods and provides the basis to assess whether the pause-segment is unusual.

The results for GISTEMP and HadCRUT are shown in figure 11. A matrix of pause-segments are represented. The vantage year (*x*-axis) is the last year of each pause-segment tested, and the number of years included (*y*-axis) defines how far back the pause-interval extends. For every pause-segment represented by an element in the matrix we have tested intervals of the same length throughout the Monte Carlo realisations of the series. The left column of figure 11 uses the historical data as they existed for each interval of the matrix, and the right column is for the hindsight version of each dataset.

For GISTEMP the pause-period trends are not at all unusual as shown by the high proportions of realisations in the Monte Carlo sample that contain a trend with magnitude lower than that in the pause-period. This conclusion holds whether considering historical or hindsight versions of GISTEMP. For HadCRUT the results show some differences from GISTEMP, but the conclusions are substantially the same. For the hindsight version of HadCRUT4 the pause-periods are not unusual and never drop below

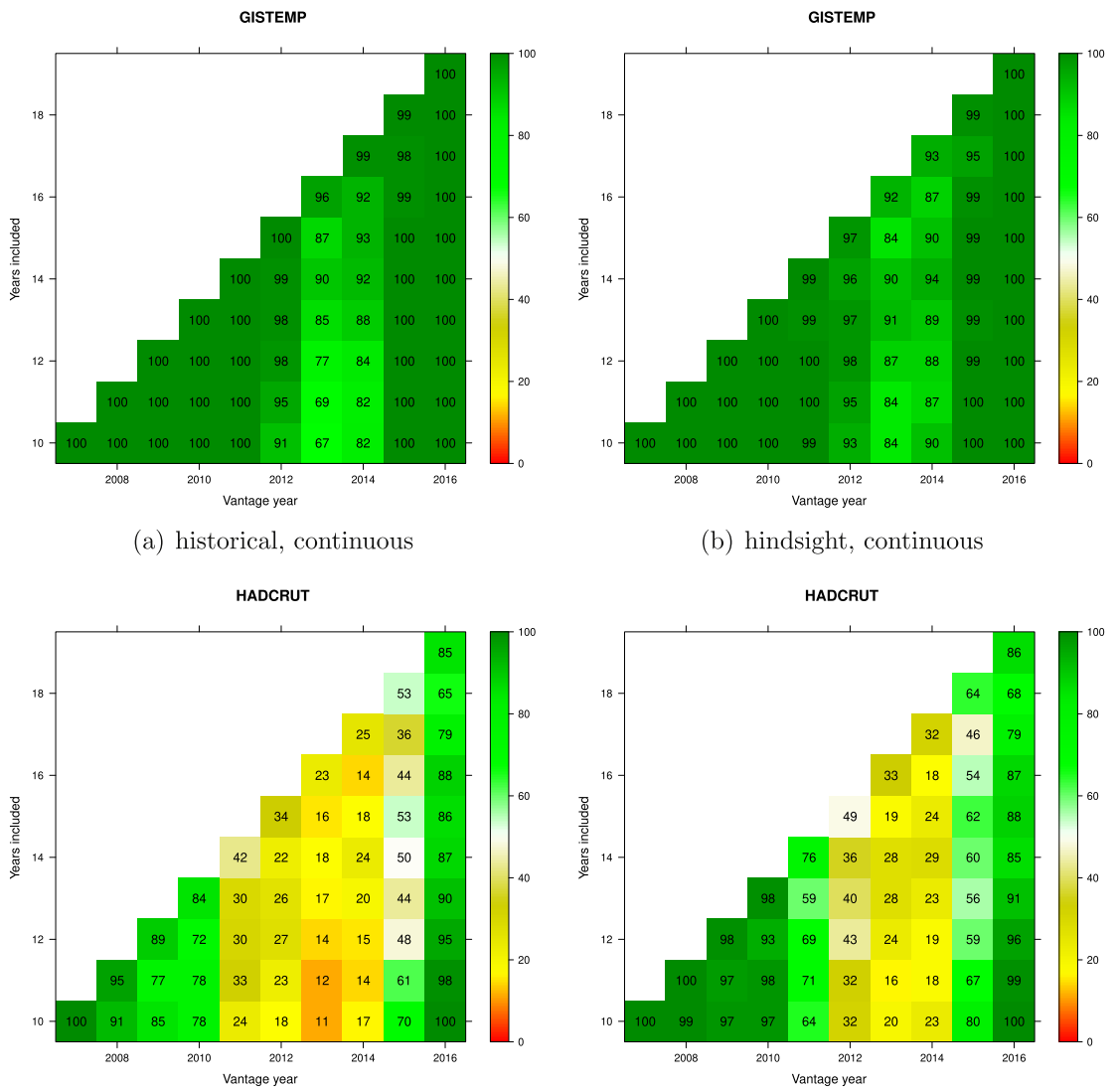

Figure 12. As in figure 11 but for continuous trends.

8% of the Monte Carlo realisations. The same result also generally holds for the historical HadCRUT, but there are a few isolated choices of pause-interval (given by the yellow circles in figure 11) where there is a smaller than 5% chance of finding a lower magnitude trend in the Monte Carlo sample. For two choices of pause-intervals in the HadCRUT historical Monte Carlo data there is a 1 in 25 chance of obtaining a lower trend magnitude than the pause-interval. Such odds are not that unusual given that the analysis involves multiple trials by testing different possible durations of slowdown intervals, which increases the likelihood of finding one by chance. To be unusual, one would expect to see a more sustained set of intervals about these intervals that are also indicative of low odds weak trends. That is not the case for even the HadCRUT historical data, where those few occasions where the odds drop below 5% are among intervals where the trends are more typical. As such, the evidence that the pause-intervals are unusual is weak, even in the most favourable configuration (HadCRUT historical) for such evidence.

4.4. Assessment of a change in trend

The assessment of unusual trends above allowed each pause-segment tested to be 'broken'. A more reasonable test is to ensure that each segment tested is continuous with the data that precedes it (section 3.3). When the Monte Carlo assessment is carried out with continuous trends it then becomes a test of a change in trend between the baseline-segment and the pause-segment. When the baseline-segment and pause-segment share the overlapping year in common without a jump (continuous trend), then the proportion of Monte Carlo series containing a lower magnitude trend than the pause-segment provides a statistical measure of the change in trend between the baseline and pause segments.

The results for Monte Carlo tests with continuous trends are shown in figure 12. As in figure 11 the tests

are shown for GISTEMP and HadCRUT with both historical (left column) and hindsight (right column) data. In practice it makes no real difference to the results whether hindsight or historical data are used. For both cases and both datasets the change in trend is not unusual. Even for HadCRUT historical data, the pause-segment from the change in trend is always larger in trend than the trends in at least 10% of other Monte Carlo realisations. Thus, the evidence to support a change in trend in GMST during the pause-period is similarly lacking.

5. Review of evidence

5.1. Types of evidence

In this section we review the evidence for a 'pause', as it was asserted in the GMST record. We review the evidence through time as it depended on different versions of the GMST record, on the number of years available in the record, and on the methods applied to assess the record. Use of the term 'evidence' implies that the information has some substantive meaning to the nature of the phenomenon asserted—in this case the claim of an unusual and noteworthy period of global temperature trend that has such different characteristics from prior temperature fluctuations that it warrants its own name and can be posited as a form of counter-evidence to global warming (e.g. Guemas *et al* 2013, Kosaka and Xie 2013, England *et al* 2014, Santer *et al* 2014). The evidence for this could be strong, or partial, or not at all. Evidence can also be current in the sense that it continues to be sustained by data and reason. Evidence can also be 'apparent' in the sense that it appears (or appeared) to support the existence of the phenomenon, but upon closer inspection turns out not to be substantive.

5.2. Analysis choices and evidence

In section 2 we noted that the 'pause' is typically neither clearly defined nor consistently defined in the literature. It is possible to characterise the range of pause-period definitions by surveying what is used to assess 'pauses' in the pause-literature (figure 1). All our assessments of pause-periods sampled the entire range of pause-periods used. The views of the 'pause' for observations in the literature divide into assessments of 'no-trend' or a 'slow-trend' as illustrated in figure 13. Much of the pause-literature models the trends as 'broken' trends, but does not take into account the additional degree of freedom introduced by that choice (section 3.3), nor the need to account for selection bias (section 3.4). The branches in figure 13 represent allowance for those choices in examination of pause-trends. For the 'slow-trend' definition of the 'pause', use of broken trends amounts to search for unusual trends (section 4.3), whereas use of continuous trends tests for a change in trend (section 4.4).

In any assessment of pause-trends one must select sources of GMST data. Many studies use a single data source, though it is prudent, given the sensitivity of trends to uncertainties in the data, to sample multiple sources. The HadCRUT, GISTEMP, and NOAA datasets were available to researchers throughout the entire pause-research period. Versions of the Cowtan and Way and Berkeley data came online during the pause-research period, and were thus only partly available (see table 1). We represent the data-choice available to researchers by the penultimate branches, HadCRUT, GISTEMP, and Other (NOAA, Cowtan and Way, Berkeley) in figure 13. Thus, descending the tree in the figure, a typical researcher makes choices (explicitly or implicitly) about how to define the 'pause' (no-trend or slow-trend), how to model the pause-interval (as broken or continuous trends), which (and how many) datasets to use (HadCRUT, GISTEMP, Other), and what versions to use for the data with what foresight about corrections to the data (historical, hindsight). For example, a researcher who chose to define the 'pause' as no-trend and selected isolated intervals to test trends (broken trends) using HadCRUT3 data would be following the left-most branches of the tree. These assessments could be made at various points in time during the pause-research period. The bottom rows in figure 13 represent assessments made for each year from 2010 through 2016.

Since the GMST datasets changed through time during the pause-research period, we kept track of the 'historical' data that was available at the time any pause-research was conducted, and made sure that one line of our analysis used only historical data. We also redid all assessments using the most recent versions of each dataset, termed 'hindsight' here. In practice, some datasets incorporated improvements before others. Further, some of the deficiencies in the historical datasets were known at the time. For example, the effect of a lack of Arctic coverage on assessment of GMST trends was known before the pause-research period (Benestad 2008, Simmons *et al* 2010), and was addressed in some datasets, but not others (table 1). The presence of a bias in the SSTs arising from the increase in buoy observations was also known prior to most of the pause literature (Smith *et al* 2008). As such, even when using purely historical data, researchers often have some knowledge of limitations in the data used, of improvements available, and of the effects of those changes. That is, the 'historical' perspective is not entirely blind to the 'hindsight' data, and thus in practice the historical researcher sits some way between these perspectives.

From the 'hindsight' (current data) perspective, the results of this study are unanimous in showing no evidence for a statistical 'pause' in GMST. This unanimity is represented by the bottom rows (years 2010–2016) in figure 13 for all the 'hindsight' branches. The open green symbol on these branches is used to indicate little or no statistical evidence. Using

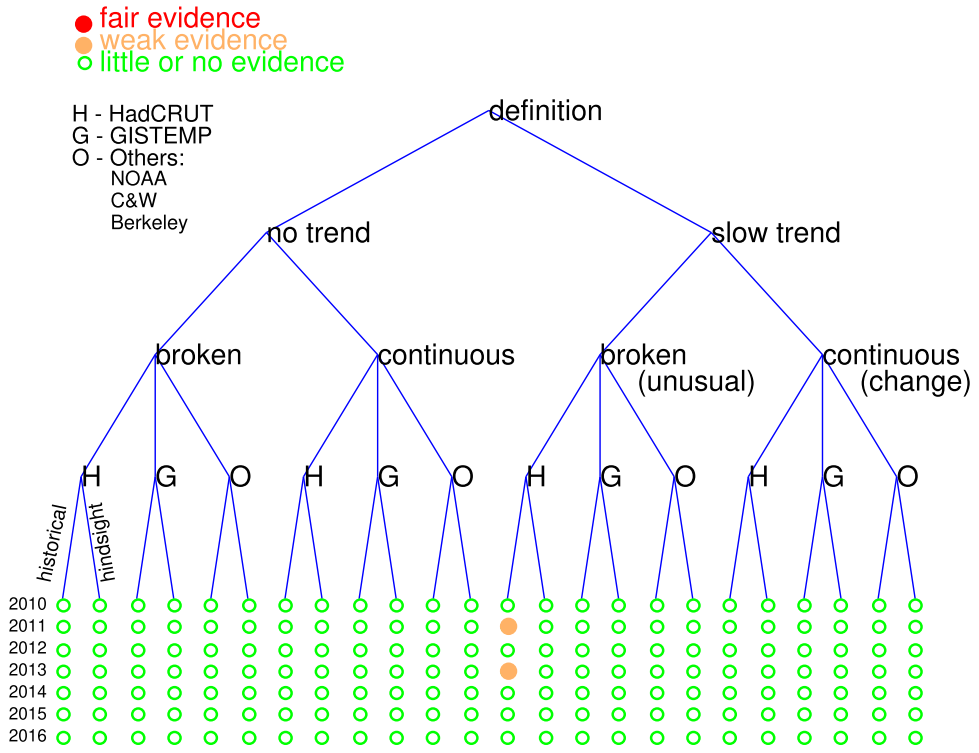

Figure 13. Tree representation of choices to represent and test pause-periods. The 'pause' is defined as either no-trend or a slow-trend. The trends can be measured as 'broken' or 'continuous' trends. The data used to assess the trends can come from HadCRUT, GISTEMP, or other datasets. The bottom branch represents the use of 'historical' versions of the datasets as they existed, or contemporary versions providing full dataset 'hindsight'. The colour coded circles at the bottom of the tree indicate our assessment of the level of evidence (fair, weak, little or no) for the tests undertaken for each set of choices in the tree. The 'year' rows are for assessments undertaken at each year in time.

hindsight GMST it does not matter how one defines the 'pause' (as a lack of trend or as a slow trend), whether one models the trends as broken or continuous, or even which version of GMST one uses (HadCRUT, GISTEMP, NOAA, Cowtan and Way, or Berkeley). For any given set of choices on the above, the result is the same in showing a lack of statistical evidence for a 'pause'. That is, looking back using current data we can't find any evidence for a 'pause', even using the most generous (and statistically dubious—broken trends, no selection bias correction) assumptions of how to model and define the 'pause' (no trend, slow trend).

Moving to the purely 'historical' data perspective, the result of the combination of tests from section 4 is substantially the same. For the no-trend definition of the 'pause' the interval length in years required to obtain significant trends is longer if using broken trends than continuous trends. However, even for broken trends, the interval is about 17 years, consistent with the result in the literature that it takes about this long to establish a signal (Santer *et al* 2011, Lewandowsky *et al* 2015b). This conclusion does not depend on the dataset used. Since there is nothing statistically unusual in this result, we have classified it as 'little or no evidence' in figure 13.

Redefining the 'pause' from no-trend to a slower than average trend introduces ambiguity into the definition of the 'pause' (Lewandowsky *et al* 2015b). It also creates confusion by using a common-language term in an uncommon manner. However, even if we accept the 'slow trend' definition of the 'pause' in historical data there is still little indication of a statistically unusual pause. The closest any of the tests comes to showing evidence is for the use of broken trends to test for an unusual trend. In the sole case of a choice of historical HadCRUT data there are a few isolated trend intervals that occur in the Monte Carlo realisations at the 1 in 20 to 1 in 25 level (4%–5%) for low trend values. We have judged this to be weak evidence in figure 13 as such levels of occurrence are not very extreme and are not sustained outside a few intervals. Further, the length of the intervals reaching this level (13, 14, and 16 years) is less than that typically required to demonstrate signal in GMST trends. And, in any of the other datasets available at the time, the pause-segment trend values for these particular intervals are even more common.

The vast preponderance of outcomes summarised in figure 13 shows that there is little or no evidence (now or then) for a lack of trend or slowing of trend in GMST during the alleged pause-period. In order to infer even minimal statistical evidence for a 'pause' a researcher must have accepted *all* of the following: that the term 'pause' refers to a change in the rate of, rather than a cessation of warming; that a broken trend

implying an upward jump in temperature at the start of the pause-period is the best way to detect a change in rate of warming; that HadCRUT is a better representation of GMST (than other data sources) despite known limitations in coverage; and that isolated intervals suffice to make the case. One may ask whether this isolated case of 'weak evidence' in figure 13 among all possible choices and outcomes is consonant with declaration of a 'pause' in GMST? The case against this is strong and includes the following points:

- The requirement to coin a new climate phenomenon, 'a pause' in observed GMST, which allegedly ran counter to greenhouse warming expectations is that the period in question should be quite exceptional or statistically unusual. The period alleged does not meet that requirement.
- Researchers knew that the climate fluctuates naturally on the time scales considered and knew to expect faster and slower than average warming periods spanning a decade or two.
- Even in the most favourable case for a pause, using HadCRUT3, the pause-period was not very exceptional.
- Researchers knew that short-term trends were sensitive to uncertainties in the GMST data, and that other GMST datasets were even less remarkable in their pause-intervals.
- There were reasons to view the other GMST datasets as good/better alternatives to HadCRUT for trend examination. This included updated corrections and better spatial coverage. The wisdom of this has been confirmed as HadCRUT trends move closer to the other datasets when HadCRUT is updated (figure 7).
- The pause-research literature did not reach a consensus on what the 'pause' actually was (figure 1), and the pause-definitions shifted through time.
- The pause-research literature did not generate robust statistical evidence for a 'pause'.

5.3. Alternative reviews of the evidence

With GMST now returning to a period where decadal trends are fluctuating steeper (faster) than the longer term warming rate (red periods in figures 2 and 3), various researchers have reviewed the evidence for a 'pause' in the prior slower than average warming rate fluctuation. The comprehensive review of Medhaug *et al* (2017) is agnostic about a pause in GMST observations and concludes that it depends 'on the time period considered, the dataset, and the hypothesis tested'. We agree with that only to the extent that there is a sensitivity to these factors. The choice of dataset and the statistical tests used contributed to a perception of a 'pause' for one dataset and for some questionable statistical tests using that dataset only. That apparent evidence was weak as discussed above and does not withstand more rigorous scrutiny with more complete or updated datasets, nor with appropriate statistical tests. Medhaug *et al* argue that 'the diverging conclusions' (about the reality of the pause) 'do not need to be inconsistent'. We argue that they are inconsistent because we do not accept that equally valid conclusions about the 'pause' in GMST can be reached using incomplete statistical methods and subsets of the data with known additional biases.

The review of Fyfe *et al* (2016) is mostly addressed at the view of a pause as a discrepancy between observations and models (see Lewandowsky *et al* 2018 for analysis of this view of the 'pause'), but concludes on the issue of a 'slowdown' in the GMST record that the 'pause' has a sound scientific basis and is supported by observations. They argue that any baseline period used to assess a pause-period in GMST must commence from 1972, not earlier. That is consistent with our choice here to use the last significant break point in the GMST record, circa 1970 (depending on dataset), to mark the beginning of the baseline-period. Fyfe *et al* argue that using this baseline period, the trend over 2001–2014 is significantly smaller than the baseline warming rate. It is not clear whether the testing underlying this conclusion took into account testing for continuous (versus broken) trends, or for correction for the selection bias problem. Our analysis, which does take these issues into account, does not support their claim to find a soundly-based slowdown in the observational record (see also Rahmstorf *et al* 2017). Some of the pause-literature has not made clear the statistical tests applied, and in some cases the evidence offered has been simple visual inspection of curves without any statistical support. The 'visual' evidence for a 'pause' may seem compelling if the series is truncated in particular ways, but it does not withstand substantive statistical scrutiny.

6. Discussion and conclusions

In learning lessons from the pause-episode in the GMST record we can describe some elements of the pause-timeline and its consequences. The origin of the 'pause' lay in contrarian narratives about the climate (Mooney 2013, Lewandowsky *et al* 2015a). With the 'pause' (or 'hiatus'), a false narrative about an alleged inconsistency between natural fluctuations of global temperature and ongoing global warming was inserted into climate discussion. Once the notion of a 'pause' was established, some of the major journals gave prominent feature to articles about it (Nature 2017). The IPCC formalised the 'pause/hiatus' for the climate community in its 5th assessment report by defining and accepting it as an observed fact about the climate system (Stocker *et al* 2013) [Box TS.3]. Many climatologists also adopted the 'pause' or 'hiatus' into

their own language about climate change. The adoption of these terms by the mainstream research community gave the 'pause' further legitimacy, even though they often explained that it was not unusual in the context of natural variability. Whether intended or not, this fed the public narrative that there was a 'pause' in global warming (Mooney 2013). To complete the cycle, researchers and climate institutions have now declared the pause to be 'over', thereby reinforcing the notion that it once existed (Xie and Kosaka 2017, Met Office 2017).

In hindsight, with current GMST datasets, there is no statistical evidence for a 'pause'. That is the case regardless of which dataset is used and even using statistical tests that inflate the significance of the results. Global warming did not pause in observations (according to any common usage of the term or in statistical terms), but clearly we need to understand how and why scientists came to the conclusion that it had in order to avoid future episodes of this kind. To this end, we pose a series of counterfactual questions about the evidence on the 'pause' in GMST.

Looking back, did the evidence depend on earlier versions of the GMST data? This question hinges upon the use of HadCRUT3 rather than any of the other GMST datasets, for only in HadCRUT3 was there even weak, isolated evidence. If HadCRUT4 had existed when HadCRUT3 did, it is unlikely that the initial claims of a 'pause' would have been made. As such, one can conclude that the use of one of the earlier GMST datasets (HadCRUT3) contributed to the perception of a 'pause'. Given the existence of known shortcomings in this data at the time (related to global coverage and SST biases), that raises the issue of communicating data uncertainties (Brohan *et al* 2006) and their implications more broadly between GMST data providers and users.

Alternatively, one can ask whether the evidence depended on the statistical methods and assumptions used to test for a 'pause'. Suppose for example that the use of continuous trends and selection bias testing had been standard at the time the pause-research was first carried out. In that case there would have been no statistical evidence for a pause, even using HadCRUT3 GMST data, and the issue would presumably not have gained any currency in the research community. Thus, the use of inappropriate statistical tests also contributed to the perception of a 'pause'. That also raises issues for the research community about the need to formulate definitions of new phenomena in terms of clear, quantifiable metrics, and in avoiding the common pitfalls in trend analysis (Miller 2013). Some recommendations along these lines for addressing future (inevitable) fluctuations in GMST trend might include:

- Any description of a new form of climate fluctuation should include a clear and generalisable definition of the phenomenon. This would include criteria for identifying onset and decay of the phenomenon.
- The definition should make clear the features that make the fluctuation unusual and whether it has a statistical or physical basis or both.
- The statistical assessment of the phenomenon should include some assessment of the sensitivity to the statistical methods employed and to the sources and major biases in the underlying data.

Researchers have noted that whether the 'pause' was real or not, it helped generate research on the mechanisms of climate variability on decadal time scales, and thus increased understanding about the climate system (Lewandowsky *et al* 2015a, 2015b, 2016, Fyfe *et al* 2016, Medhaug *et al* 2017, Nature 2017). While this is true, it is also important to ask what has been lost by the invention of a 'pause' in global warming? We will never fully know the answer to this question, but it is clear that the climate-research community's self-declaration of a 'pause' in global warming has created additional confusion for the public and policy-system about the pace and urgency of climate change. This in turn may have contributed to reduced momentum for action to prevent greenhouse climate change, even if only a bit and if only by some years. That lost momentum is likely to be counted in higher total emissions of greenhouse gases before climate stabilisation (Allen *et al* 2009, Meinshausen *et al* 2009). The full costs of that are unknownable, but the risks are substantial (World Bank 2012, Hansen *et al* 2016). That is, there are costs, and there are perspectives upon which it matters whether the 'pause' was real or not. The effort to deconstruct the basis for the 'pause' is not strictly academic and provides some salient lessons for the science.

Appendix

List of papers used to define 'pause' periods in the literature through 2016.

Adam O, Schneider T and Harnik N 2014 Role of changes in mean temperatures versus temperature gradients in the recent widening of the hadley circulation *Journal of Climate* **27** 7450–61

Allan P R, Liu C, Loeb G N, Palmer D M, Roberts M, Smith D and Vidale, P-L 2014 Changes in global net radiative imbalance 1985–2012 *Geophysical Research Letters* **41** 5588–97

Amaya J D, Xie S-P, Miller J A and McPhaden J M 2015 Seasonality of tropical pacific decadal trends associated with the XXI century global warming hiatus *Journal of Geophysical Research* **120** 6782–98

An W, Hou S, Zhang W, Wu S, Xu H, Pang H, Wang Y and Liu Y 2016 Possible recent warming hiatus on the northwestern tibetan plateau derived from ice core records *Scientific Reports* **6** 32813

Brown T P, Li W, Cordero C E and Mauget A S 2015 Comparing the model-simulated global warming signal to observations using empirical estimates of unforced noise *Scientific Reports* **5** 9957

Chakrabarty K D and Peshin K S 2013 Global warming and solar anomaly *Indian Journal of Radio and Space Physics* **42**

Chikamoto Y, Mochizuki T, Timmermann A, Kimoto M and Watanabe, M 2016 Potential tropical atlantic impacts on pacific decadal climate trends *Geophysical Research Letters* **43** 7143–51

Crowley J T, Obrochta P S and Liu J 2014 Recent global temperature 'plateau' in the context of a new proxy reconstruction *Earth's Future* **2** 281–94

Dai A, Fyfe C J, Xie S-P and Dai X 2015 Decadal modulation of global surface temperature by internal climate variability *Nature Climate Change* **5** 555–9

Delworth L T, Zeng F, Rosati A, Vecchi A G and Wittenberg T A 2015 A link between the hiatus in global warming and north american drought *Journal of Climate* **28** 3834–45

Duan A and Xiao Z 2015 Does the climate warming hiatus exist over the tibetan plateau? *Scientific Reports* **5** 13711

Dunstone J N 2014 A perspective on sustained marine observations for climate modelling and prediction *Philosophical Transactions A* **372**

Easterling R D and Wehner F M 2009 Is the climate warming or cooling? *Geophysical Research Letters* **36** L08706

England H M, Kajtar B J and Maher N 2015 Robust warming projections despite the recent hiatus *Nature Climate Change* **5** 394–6

England H M, McGregor S, Spence P, Meehl A G, Timmermann A, Cai W, Gupta S A, McPhaden J M, Purich A and Santoso A 2014 Recent intensification of wind-driven circulation in the pacific and the ongoing warming hiatus *Nature Climate Change* **4** 222–7

Estrada F, Perron P and Martnez-Lpez B 2013 Statistically derived contributions of diverse human influences to twentieth-century temperature changes *Nature Geoscience* **6** 1050–55

Furuoka F 2016 An econometric analysis of global warming hiatus *Applied Economics Letters* **NA**

Fyfe C J, Gillett P N and Zwiers W F 2013 Overestimated global warming over the past 20 years *Nature Climate Change* **3** 767–9

Fyfe C J, Meehl A G, England H M, Mann E M, Santer D B, Flato M G, Hawkins E, Gillett P N, Xie S-P, Kosaka Y and Swart C N 2016 Making sense of the early-2000s warming slowdown *Nature Climate Change* **6** 224–8

Gervais F 2016 Anthropogenic co2 warming challenged by 60-year cycle *Earth-Science Reviews* **155** 129–35

Gettelman A, Shindell T D and Lamarque F J 2015 Impact of aerosol radiative effects on 2000–2010 surface temperatures *Climate Dynamics* **45** 2165–79

Gleisner H, Thejll P, Christiansen B and Nielsen K J 2015 Recent global warming hiatus dominated by low-latitude temperature trends in surface and troposphere data *Geophysical Research Letters* **42** 510–7

Gu G, Adler F R and Huffman J G 2016 Long-term changes/trends in surface temperature and precipitation during the satellite era (1979–2012) *Climate Dynamics* **46** 1091–1105

Guan X, Huang J, Guo R and Lin P 2015 The role of dynamically induced variability in the recent warming trend slowdown over the northern hemisphere *Scientific Reports* **5** 12669

Hawkins E, Edwards T and McNeall D 2014 Pause for thought *Nature Climate Change* **4** 154–6

Haywood M J, Jones A and Jones S G 2014 The impact of volcanic eruptions in the period 2000–2013 on global mean temperature trends evaluated in the hadgem2-es climate model *Atmospheric Science Letters* **15** 92–6

Huang J, Xie Y, Guan X, Li D and Ji F 2017 The dynamics of the warming hiatus over the northern hemisphere *Climate Dynamics* **48** 429–46

Huber M and Knutti R 2014 Natural variability radiative forcing and climate response in the recent hiatus reconciled *Nature Geoscience* **7** 651–6

Hunt G B 2011 The role of natural climatic variation in perturbing the observed global mean temperature trend *Climate Dynamics* **36** 509–21

Johansson A J D, O'Neill C B, Tebaldi C and Hggstrm O 2015 Equilibrium climate sensitivity in light of observations over the warming hiatus *Nature Climate Change* **5** 449–53

Kamae Y, Shiogama H, Watanabe M and Kimoto M 2014 Attributing the increase in northern hemisphere hot summers since the late XXth century *Geophysical Research Letters* **41** 5192–99

Kaufmann K R, Kauppi H, Mann L M and Stock H J 2011 Reconciling anthropogenic climate change with observed temperature 1998–2008 *PNAS* **108** 11790–3

Kay E J, Deser C, Phillips A, Mai A, Hannay C, Strand G, Arblaster M J, Bates C S, Danabasoglu G, Edwards J, Holland M, Kushner P, Lamarque J-F, Lawrence D, Lindsay K, Middleton A, Munoz E, Neale R, Oleson K, Polvani L and Vertenstein M 2015 The community earth system model (cesm) large ensemble project: A community resource for studying climate change in the presence of internal climate variability *Bulletin of the American Meteorological Society* **96** 1333–49

Knutson R T, Zhang R and Horowitz W L 2016 Prospects for a prolonged slowdown in global warming in the early XXI century *Nature Communications* **7** 13676

Kosaka Y and Xie S-P 2013 Recent global-warming hiatus tied to equatorial pacific surface cooling *Nature* **501** 403–7

Kosaka Y and Xie S-P 2016 The tropical pacific as a key pacemaker of the variable rates of global warming *Nature Geoscience* **9** 669–73

Kumar S, III K L J, Pan Z and Sheffield J 2016 Twentieth century temperature trends in cmip3, cmip5, and cesm-le climate simulations: Spatial-temporal uncertainties differences, and their potential sources *Journal of Geophysical Research* **121** 9561–75

Lean L J and Rind H D 2009 How will earth's surface temperature change in future decades? *Geophysical Research Letters* **36** L15708

Leggett L M W and Ball D A 2015 Granger causality from changes in level of atmospheric co2 to global surface temperature and the el nio?southern oscillation and a candidate mechanism in global photosynthesis *Atmospheric Chemistry and Physics* **15** 11571–92

Lewandowsky S, Risbey S J and Oreskes N 2016 The 'pause' in global warming: Turning a routine fluctuation into a problem for science *Bulletin of the American Meteorological Society* **97** 723–33

Li C, Stevens B and Marotzke J 2015 Eurasian winter cooling in the warming hiatus of 1998–2012 *Geophysical Research Letters* **42** 8131–39

Li W T and Baker C N 2016 Detecting warming hiatus periods in CMIP5 climate model projections *International Journal of Atmospheric Sciences* **2016** 9657659

Lin M and Huybers P 2016 Revisiting whether recent surface temperature trends agree with the cmip5 ensemble *Journal of Climate* **29** 8673–87

Lin Y and Franzke E L C 2015 Scale-dependency of the global mean surface temperature trend and its implication for the recent hiatus of global warming *Scientific Reports* **5** 12971

Lovejoy S 2014 Return periods of global climate fluctuations and the pause *Geophysical Research Letters* **41** 4704–10

Lovejoy S 2015 Using scaling for macroweather forecasting including the pause *Geophysical Research Letters* **42** 7148–55

Macias D, Stips A and Garcia-Gorriz E 2014 Application of the singular spectrum analysis technique to study the recent hiatus on the global surface temperature record *PLoS ONE* **9** e107222

Mann E M, Steinman A B, Miller K S, Frankcombe M L, England H M and Cheung H A 2016 Predictability of the recent slowdown and subsequent recovery of large-scale surface warming using statistical methods *Geophysical Research Letters* **43** 3459–67

Marotzke J and Forster M P 2015 Forcing feedback and internal variability in global temperature trends *Nature* **517** 565–70

Meehl A G, Hu A, Santer D B and Xie S-P 2016 Contribution of the interdecadal pacific oscillation to twentieth-century global surface temperature trends *Nature Climate Change* **6** 1005–8

Meehl A G, Teng H and Arblaster M J 2014 Climate model simulations of the observed early-2000s hiatus of global warming *Nature Climate Change* **4** 898–902

Meehl G A, Hu A and Teng H 2016 Initialized decadal prediction for transition to positive phase of the interdecadal pacific oscillation *Nature Communications* **7** 11718

Meehl G A and Teng H 2012 Case studies for initialized decadal hindcasts and predictions for the pacific region *Geophysical Research Letters* **39** L22705

Parker A 2014 The 'present global warming hiatus' is part of a quasi-60 years oscillation in the worldwide average temperatures in the downwards phase *Environmental Science: An Indian Journal* **9** 14–22

Parker A 2015 The 'artefacts' of data biases in surface temperatures are hiding the hiatus *American Journal of Geophysics, Geochemistry and Geosystems* **1** 66–70

Pasini A, Triacca U and Attanasio A 2016 Evidence for the role of the atlantic multidecadal oscillation and the ocean heat uptake in hiatus prediction *Theoretical and Applied Climatology* **NA**

Peyser E C, Yin J, Landerer W F and Cole E J 2016 Pacific sea level rise patterns and global surface temperature variability *Geophysical Research Letters* **43** 8662–9

Power S, Delage F, Wang G, Smith I and Kociuba G 2016 Apparent limitations in the ability of cmip5 climate models to simulate recent multi-decadal change in surface temperature: implications for global temperature projections *Climate Dynamics* **NA**

Pretis F, Mann L M and Kaufmann K R 2015 Testing competing models of the temperature hiatus: assessing the effects of conditioning variables and temporal

uncertainties through sample-wide break detection *Climatic Change* **131** 705–18

Quirk T 2012 Did the global temperature trend change at the end of the 1990s? *Asia-Pacific Journal of Atmospheric Sciences* **48** 339–44

Rackow T, Goessling F H, Jung T, Sidorenko D, Semmler T, Barbi D and Handorf D 2016 Towards multi-resolution global climate modelling with echam6-fesom. part ii: climate variability *Climate Dynamics* **NA**

Risbey S J, Lewandowsky S, Langlais C, Monselesan P D, O′Kane J T and Oreskes N 2014 Well-estimated global surface warming in climate projections selected for ENSO phase *Nature Climate Change* **4** 835–40

Roberts D C, Palmer D M, McNeall D and Collins M 2015 Quantifying the likelihood of a continued hiatus in global warming *Nature Climate Change* **5** 337–42

Saenko A O, Fyfe C J, Swart C N, Lee G W and England H M 2016 Influence of tropical wind on global temperature from months to decades *Climate Dynamics* **47** 2193–2203

Saffioti C, Fischer M E and Knutti R 2015 Contributions of atmospheric circulation variability and data coverage bias to the warming hiatus *Geophysical Research Letters* **42** 2385–91

Schmidt A G, Shindell T D and Tsigaridis K 2014 Reconciling warming trends *Nature Geoscience* **7** 158–60

Schurer P A, Hegerl C G and Obrochta P S 2015 Determining the likelihood of pauses and surges in global warming *Geophysical Research Letters* **42** 5974–82

Seneviratne I S, Donat G M, Mueller B and Alexander V L 2014 No pause in the increase of hot temperature extremes *Nature Climate Change* **4** 161–3

Shi Y, Zhai P and Jiang Z 2016 Multi-sliding time windows based changing trend of mean temperature and its association with the global-warming hiatus *Journal of Meteorological Research* **30** 232–41

Smith M D, Booth B B B, Dunstone J N, Eade R, Hermanson L, Jones S G, Scaife A A, Sheen L K and Thompson V 2016 Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown *Nature Climate Change* **6** 936–40

Solomon S, Rosenlof H K, Portmann W R, Daniel S J, Davis M S, Sanford J T and Plattner G-K 2010 Contributions of stratospheric water vapor to decadal changes in the rate of global warming *Science* **327** 1219–23

Song J, Wang Y and Tang J 2016 A hiatus of the greenhouse effect *Scientific Reports* **6** 33315

Steinman A B, Mann E M and Miller K S 2015 Atlantic and pacific multidecadal oscillations and northern hemisphere temperatures *Science* **347** 988–991

Swanson L K and Tsonis A A 2009 Has the climate recently shifted? *Geophysical Research Letters* **36** L06711

Thoma M, Greatbatch J R, Kadow C and Gerdes R 2015 Decadal hindcasts initialized using observed surface wind stress: Evaluation and prediction out to 2024 *Geophysical Research Letters* **42** 6454–61

Thorne P, Outten S, Bethke I and Seland y 2015 Investigating the recent apparent hiatus in surface temperature increases: Part 2. comparison of model ensembles to observational estimates *Journal of Geophysical Research* **120** 8597–620

Trenberth E K and Fasullo T J 2013 An apparent hiatus in global warming? *Earth's Future* **1** 19–32

Trenberth E K, Fasullo T J, Branstator G and Phillips S A 2014 Seasonal aspects of the recent pause in surface warming *Nature Climate Change* **4** 911–16

Wang S, Wen X, Luo Y, Tang G, Zhao Z and Huang J 2010 Does the global warming pause in the last decade: 1999–2008? *Advances in Climate Change Research* **1** 49–54

Wang Y, Su H, Jiang H J, Livesey J N, Santee L M, Froidevaux L, Read G W and Anderson J 2016 The linkage between stratospheric water vapor and surface temperature in an observation-constrained coupled general circulation model *Climate Dynamics* **NA**

Watanabe M, Kamae Y, Yoshimori M, Oka A, Sato M, Ishii M, Mochizuki T and Kimoto M 2013 Strengthening of ocean heat uptake efficiency associated with the recent climate hiatus *Geophysical Research Letters* **40** 3175–79

Watanabe M, Shiogama H, Tatebe H, Hayashi M, Ishii M and Kimoto M 2014 Contribution of natural decadal variability to global warming acceleration and hiatus *Nature Climate Change* **4** 893–7

Wei M and Qiao F 2016 Attribution analysis for the failure of cmip5 climate models to simulate the recent global warming hiatus *Science China Earth Sciences* **60** 397–408

Yao S-L, Huang G, Wu R-G and Qu X 2016 The global warming hiatus—a natural product of interactions of a secular warming trend and a multi-decadal oscillation *Theoretical and Applied Climatology* **123** 349–60

Zeng X and Geil K 2016 Global warming projection in the XXI century based on an observational data-driven model *Geophysical Research Letters* **43** 10947–54

Zhao L, Xu J, Powell A, Guo D, Shi C, Shao M and Wang D 2016 Investigation on the tendencies of the

land–ocean warming contrast in the recent decades *IEEE Geoscience and Remote Sensing Letters* **13** 1522–26

Zhou C and Wang K 2016a Coldest temperature extreme monotonically increased and hottest extreme oscillated over northern hemisphere land during last 114 years *Scientific Reports* **6** 25721

Zhou C and Wang K 2016b Spatiotemporal divergence of the warming hiatus over land based on different definitions of mean temperature *Scientific Reports* **6** 31789

Zhou Y, Luo M and Leung Y 2016 On the detection of precipitation dependence on temperature *Geophysical Research Letters* **43** 4555–65

Zhu Y, Wang T and Wang H 2016 Relative contribution of the anthropogenic forcing and natural variability to the interdecadal shift of climate during the late 1970s and 1990s *Science Bulletin* **61** 416–24

ORCID iDs

James S Risbey https://orcid.org/0000-0003-3202-9142
Stephan Lewandowsky https://orcid.org/0000-0003-1655-2013
Kevin Cowtan https://orcid.org/0000-0002-0189-1437
Naomi Oreskes https://orcid.org/0000-0001-6122-9843
Stefan Rahmstorf https://orcid.org/0000-0001-6786-7723

References

Allen M, Frame D, Huntingford C, Jones C, Lowe J, Meinshausen M and Meinshausen N 2009 Warming caused by cumulative carbon emissions towards the trillionth tonne *Nature* **458** 1163–6

Arguez A and Vose R 2011 The definition of the standard WMO climate normal: the key to deriving alternative climate normals *Bull. Am. Meteorol. Soc.* **92** 699–704

Benestad R 2008 Mind the gap! (http://realclimate.org/index.php/archives/2008/11/mind-the-gap/)

Brohan P, Kennedy J, Harris I, Tett S and Jones P 2006 Uncertainty estimates in regional and global observed temperature changes: a new data set from 1850 *J. Geophys. Res.* **111** 1–21

Cahill N 2018 No 'slowdown' in the story of global warming *Significance* **15** 24–7

Cahill N, Rahmstorf S and Parnell A 2015 Change points of global temperature *Environ. Res. Lett.* **10** 1–6

Cowtan K and Way R 2014 Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends *Q. J. R. Meteorol. Soc.* **140** 1935–44

England M, McGregor S, Spence P, Meehl G, Timmermann A, Cai W, Gupta A S, McPhaden M, Purich A and Santoso A 2014 Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus *Nat. Clim. Change* **4** 222–7

Fyfe J, Gillett N and Zwiers F 2013 Overestimated global warming over the past 20 years *Nat. Clim. Change* **3** 767–9

Fyfe J *et al* 2016 Making sense of the early-2000s warming slowdown *Nat. Clim. Change* **6** 224–8

Guemas V, Doblas-Reyes F, Andreu-Burillo I and Asif M 2013 Retrospective prediction of the global warming slowdown in the past decade *Nat. Clim. Change* **3** 649–53

Handel M and Risbey J 1992 An annotated bibliography on the greenhouse effect and climate change *Clim. Change* **21** 91–255

Hansen J, Ruedy R, Sato M and Lo K 2010 Global surface temperature change *Rev. Geophys.* **48** 1–29

Hansen J, Russell G, Lacis A, Fung I, Rind D and Stone P 1985 Climate response times: dependence on climate sensitivity and ocean mixing *Science* **229** 857–9

Hansen J *et al* 2016 Ice melt, sea level rise and superstorms: evidence from paleoclimate data, climate modeling, and modern observations that 2 °C global warming could be dangerous *Atmos. Chem. Phys.* **16** 3761–812

Hausfather Z, Cowtan K, Clarke D, Jacobs P, Richardson M and Rohde R 2017 Assessing recent warming using instrumentally homogeneous sea surface temperature records *Sci. Adv.* **3** 1–13

Houghton J T, Jenkins G and Ephraums J (ed) 1990 *Climate Change: The IPCC Scientific Assessment* (Cambridge: Cambridge University Press)

Houghton J T *et al* (ed) 1996 *Climate Change 1995: The Science of Climate Change* (Cambridge: Cambridge University Press)

Houghton J T *et al* (ed) 2001 *Climate Change 2001: The Scientific Basis* (Cambridge: Cambridge University Press)

Huang B, Thorne P, Smith T, Liu W, Lawrimore J, Banzon V, Zhang H, Peterson T and Menne M 2015 Further exploring and quantifying uncertainties for extended reconstructed sea surface temperature (ERSST) version 4 *J. Clim.* **29** 3119–42

Jones P D, New M, Parker D, Martin S and Rigor I 1999 Surface air temperature and its changes over the past 150 years *Rev. Geophys.* **37** 173–99

Karl T, Arguez A, Huang B, Lawrimore J, McMahon J, Menne M, Peterson T, Vose R and Zhang H 2015 Possible artifacts of data biases in the recent global surface warming hiatus *Science* **348** 1469–72

Karl T, Tarpley D, Quayle R, Diaz H, Robinson D and Bradley R 1989 The recent climate record: what it can and cannot tell us *Rev. Geophys.* **27** 405–30

Kennedy J, Rayner N, Smith R, Saunby M and Parker D 2011 Reassessing biases and other uncertainties in sea-surface temperature observations since 1850: 1. Measurement and sampling errors *J. Geophys. Res.* **116** 1–11

Kent E *et al* 2017 A call for new approaches to quantifying biases in observations of sea-surface temperature *Bull. Am. Meteorol. Soc.* **198** 1601–16

Kosaka Y and Xie S-P 2013 Recent global-warming hiatus tied to equatorial Pacific surface cooling *Nature* **501** 403–7

Lewandowsky S, Oreskes N, Risbey J, Newell B and Smithson M 2015a Seepage: climate change denial and its effect on the scientific community *Global Environ. Change* **33** 1–13

Lewandowsky S, Cowtan K, Risbey J, Mann M, Steinman B, Oreskes N and Rahmstorf S 2018 The 'pause' in global warming in historical context: comparing models to observations *Environ. Res. Lett.* **13** 123007

Lewandowsky S, Risbey J and Oreskes N 2015b On the definition and identifiability of the alleged 'hiatus' in global warming *Sci. Rep.* **5** 1–12

Lewandowsky S, Risbey J and Oreskes N 2016 The 'pause' in global warming: turning a routine fluctuation into a problem for science *Bull. Am. Meteorol. Soc.* **97** 723–33

Marotzke J and Forster P 2015 Forcing, feedback, and internal variability in global temperature trends *Nature* **517** 565–70

Medhaug I, Stolpe M, Fischer E and Knutti R 2017 Reconciling controversies about the 'global warming hiatus' *Nature* **545** 41–7

Meinshausen M, Meinshausen N, Hare W, Raper S, Frieler K, Knutti R, Frame D and Allen M 2009 Greenhouse-gas emission targets for limiting global warming to 2 °C *Nature* **458** 1158–63

Met Office 2017 A Pacific flip triggers the end of the recent slowdown (http://metoffice.gov.uk/news/releases/2017/a-pacific-flip-triggers-the-end-of-the-recent-slowdown)

Miller B 2013 When is consensus knowledge based? Distinguishing shared knowledge from mere agreement *Synthese* **190** 1293–316

Mitchell J *et al* 2001 Detection of climate change and attribution of causes *Climate Change 2001: The Scientific Basis* (Cambridge: Cambridge University Press) ch 12 pp 695–738

Mooney C 2013 Who created the global warming pause? (http://grist.org/climate-energy/who-created-the-global-warming-pause/)

Morice C, Kennedy J, Rayner N and Jones P 2012 Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 data set *J. Geophys. Res.* **117** 1–22

National Research Council 1982 *Climate in Earth History* (Washington, DC: National Academy Press)

National Research Council 1995 *Natural Climate Variability on Decade-to-Century Time Scales* (Washington, DC: National Academies Press)

Nature 2017 What pause? *Nature* **545** 7652

O'Kane T, Matear R, Chamberlain M, Risbey J, Sloyan B and Horenko I 2013 Decadal variability in an OGCM Southern Ocean: Intrinsic modes, forced modes and metastable states *Ocean Model.* **69** 1–21

Rahmstorf S, Cazenave A, Church J, Hansen J, Keeling R, Parker D and Somerville R 2007 Recent climate observations compared to projections *Science* **316** 709

Rahmstorf S, Foster G and Cahill N 2017 Global temperature evolution: recent trends and some pitfalls *Environ. Res. Lett.* **12** 1–7

Rayner N, Brohan P, Parker D, Folland C, Kennedy J, Vanicek M, Ansell T and Tett S 2006 Improved analyses of changes and uncertainties in sea surface temperature measured *in situ* since the mid-nineteenth century: the HadSST2 dataset *J. Clim.* **19** 446–69

Risbey J and Lewandowsky S 2017 The 'pause' unpacked *Nature* **545** 37–9

Risbey J, Lewandowsky S, Hunter J and Monselesan D 2015 Betting strategies on fluctuations in the transient response of greenhouse warming *Phil. Trans. R. Soc.* A **373** 14–27

Risbey J, Lewandowsky S, Langlais C, Monselesan D, O'Kane T and Oreskes N 2014 Well-estimated global surface warming in climate projections selected for ENSO phase *Nat. Clim. Change* **4** 835–40

Rohde R, Muller R, Jacobsen R, Muller E, Perlmutter S, Rosenfeld A, Wurtele J, Groom D and Wickham C 2013 A new estimate of the average Earth surface land temperature spanning 1753 to 2011 *Geoinform. Geostat.: Overview* **1** 1–7

Santer B *et al* 2011 Separating signal and noise in atmospheric temperature changes: the importance of timescale *J. Geophys. Res.* **116** 1984–2012

Santer B *et al* 2014 Volcanic contribution to decadal changes in tropospheric temperature *Nat. Geosci.* **7** 185–9

Simmons A, Willett K, Jones P, Thorne P and Dee D 2010 Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: inferences from reanalyses and monthly gridded observational data sets *J. Geophys. Res.* **115** 1–10

Smith T, Reynolds R, Peterson T and Lawrimore J 2008 Improvements to NOAA's historical merged land-ocean surface temperatures analysis (1880–2006) *J. Clim.* **21** 2283–96

Stocker T *et al* 2013 Technical summary *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge: Cambridge University Press) pp 33–115

Ventura V, Paciorek C and Risbey J 2004 Controlling the proportion of falsely rejected hypotheses when conducting multiple tests with climatological data *J. Clim.* **17** 4343–56

Wilks D 2006 On field significance and the false discovery rate *J. Appl. Meteorol. Clim.* **45** 1181–9

World Bank 2012 *Turn Down the Heat: Why a 4 °C Warmer World Must be Avoided* (Technical Report) (Washington, DC: The World Bank)

Xie S and Kosaka Y 2017 What caused the global surface warming hiatus of 1998–2013? *Curr. Clim. Change Rep.* **3** 128–40