Signal detection in global mean temperatures after "Paris": an uncertainty and sensitivity analysis

Hans Visser[1], Sönke Dangendorf[2], Detlef P. van Vuuren[1,3], Bram Bregman[4] and Arthur C. Petersen[5]

[1] PBL Netherlands Environmental Assessment Agency, Bilthoven, The Netherlands

[2] Research Institute for Water and Environment, University Siegen, Siegen, Germany

[3] Faculty of Geosciences, University Utrecht, Utrecht, The Netherlands

[4] Institute for Science, Innovation and Society, Radboud University, Nijmegen, The Netherlands

[5] STEaPP, University College London, London, Great Britain

Correspondence to: Hans Visser (hans.visser@pbl.nl)

Abstract. In December 2015, 195 countries agreed in Paris to 'hold the increase in global mean surface temperature (GMT) well below 2.0 °C above pre-industrial levels and to pursue efforts to limit the temperature increase to 1.5 °C'. Since large financial flows will be needed to keep GMTs below these targets, it is important to know how GMT has progressed since pre-industrial times. However, the Paris Agreement is not conclusive as for methods to calculate it. Should trend progression be deduced from GCM simulations or from instrumental records by (statistical) trend methods? Which simulations or GMT datasets should be chosen, and which trend models? What is 'pre-industrial', and finally, are the Paris targets formulated for total warming, originating from both natural and anthropogenic forcing, or do they refer to anthropogenic warming only? To find answers to these questions we performed an uncertainty and sensitivity analysis where datasets and model choices have been varied. For all cases we evaluated trend progression along with uncertainty information. To do so, we analysed four trend approaches and applied these to the five leading GMT products. We find GMT progression to be largely independent of various trend model approaches. However, GMT progression is significantly influenced by the choice of GMT datasets. Both sources of uncertainty are dominated by natural variability. As a parallel path, we calculated GMT progression from an ensemble of 106 GCM simulations. Mean progression derived from GCM-based GMTs appears to lie in the range of trend-dataset combinations. A difference between both approaches lies in the width of uncertainty bands: GCM simulations show a much wider spread. Finally, we discuss various choices for pre-industrial baselines and the role of warming definitions. Based on these findings we propose an estimate for signal progression in GMTs since pre-industrial.

1. Introduction

Global mean surface temperature (GMT) is undoubtedly one of the key indicators of climate change. Tollefson (2015) denotes the GMT indicator as 'the global thermostat'. Over the years many articles have been published in relation to GMT series and the patterns therein. These patterns combine an anthropogenic signal – induced by growing concentration of greenhouses and processes such as aerosol cooling – as well as natural variability. Natural variability can be regarded as a correlated noise process consisting of (i) internal random unforced (chaotic) variability and (ii) external radiatively forced changes. Here, internal variability is steered by short-term processes such as weather in the high latitudes or El Niño and La Niña, as well as by decadal processes such as the Interdecadal Pacific Oscillation (e.g., Trenberth, 2015; Fyfe et al. 2016; Xie, 2016; Meehl et al., 2016), and will result in correlated noise in GMTs (Mudelsee, 2014; Roberts et al., 2015). Externally forced variability is mainly due to volcanic eruptions and variations in solar irradiance. It influences global temperatures on annual to centennial scales (IPCC, 2013 - Ch. 10; Forster et al., 2013; Mann et al., 2016). A recent realization of internal variability led to a fierce debate in the popular media: GMTs were showing a claimed "slowdown", "pause" or "hiatus" from the year 1998 onwards (e.g., Lewandowski et al., 2015; Hedemann et al., 2017; Medhaug et al., 2017 - their figure 1).

GMTs has been a crucial indicator in climate negotiations for a long time and it has even become more so at the the 21st Conference of Parties (COP21) in Paris, December 2015. The final accord, approved by 195 countries, agreed on GMT targets which aim to avoid increases of 1.5 and 2.0 °C compared to pre-industrial temperatures (UN, 2015). IPCC (2014a) showed that meeting such GMT targets will require deep reductions of GHG emissions at the cost of high investments in mitigation measures worldwide. Given the fact that all goals are formulated on the basis of this single GMT indicator, the question arises: what is the current GMT level since pre-industrial?

So far, little attention has been paid to this topic. IPCC (2013), in its attempt to clarify the meaning of GMT measurements, applied linear trends to three different GMT datasets. They reported a trend progression $\Delta\mu$ of 0.85 [0.65, 1.06] °C for the period 1880-2012. The uncertainty range stands for 90% confidence limits, originating from differences in datasets, natural variability of the climate system (forced and unforced), and expert judgment (IPCC 2013 - Box 2.2). Hawkins et al. (2017) and Schurer et al. (2017) addressed the topic of trend progression *since pre-industrial* and quantified the role of various choices for pre-industrial baselines.

Hawkins et al. found that the period 1720-1800 would be the most suitable in physical terms, despite incomplete information about radiative forcings and very few direct observations during this time. Additionally, they concluded that the 1850-1900 period would be a reasonable surrogate for pre-industrial GMTs, being only 0.05 °C warmer than the 1720-1800 period. Subsequently, Hawkins et al. analyzed GMT progression since pre-industrial by calculating the GMT mean over the 20-year period 1986-2005 for various GMT products and other instrumental data (their figure 4). Trend progression itself was approximated in the study by multiple regression models with non-stationary explanatory variables such as historic GHG forcing curves or local temperature series (the Central England Temperature series or the De Bilt series). Schurer et al. found that GHGs had a significant warming effect on global temperatures if the period 1401-1800 is compared to 1850-1900: from 0.02 to 0.20 ºC (90% confidence limits). If all forcings are combined (GHG, solar, volcanic), they found a similar warming effect of 0.09 [0.03 - 0.19] °C.

In this article, we build on the work of Hawkins et al. but we do not base our GMT progression estimates on linear regression models with non-stationary regressors. The drawback of this approach is simply the linearity assumed, while the climate system is (highly) non-linear with a number of feedback processes. Therefore, we follow two other trend estimation approaches: (i) statistical trend models and (ii) global temperature trends derived from Global Climate Models (GCMs). Furthermore, we avoid methods or presentations based on subjectively selected time-windows (such as Moving Averages). The drawback of time windows is that averages over 21-year periods or similar do not give estimates for the beginning and ending of the sample period chosen (thus, we would have no trend estimates for the period 2007-2016).

A final topic we address is that of warming definitions. Should the Paris targets be interpreted as warming due to both anthropogenic and natural forcings, or as warming due to anthropogenic warming only? The terms 'global warming' or 'total warming' are interpreted in most literature as the sum of anthropogenic warming plus long-term (decadal to centennial) natural warming, consistent with the IPCC definition of climate change (IPCC Annex II, 2014). However, some researchers interpret 'global warming' as anthropogenic warming only, consistent with the definition proposed by UNFCCC in their article 1 (Otto et al., 2015; Millar et al. 2017). In both definitions, short-term natural variability – such as seen in "the hiatus period" – is smoothed from warming trends.

Our approach is that of an uncertainty and sensitivity analysis as promoted by Saltelli et al. (2004), Saisana et al. (2005) and Visser et al. (2015). We ask the following three major questions:

- How robust are estimates for GMT progression as for specific choices of trend modelling, use of GCMs and specific choices of GMT datasets?
- How do these choices influence uncertainties in GMT progression in relation to uncertainties due to forced and unforced natural variability?
- Does the choice for a specific pre-industrial baseline or period play a role? And are our estimates sensitive to forced natural variability on decadal to centennial scales? In other words, does it matter if we interpret the Paris targets as total warming, or as anthropogenic warming only?

Since there is no ‘true’ or ‘best’ trend approach (Visser et al., 2015), we explore four trend methods and apply these to five leading GMT products (similar to Hawkins et al.). This leads to a 4-by-5 matrix of GMT trend progressions since 1880. As a parallel path, we compare these trend progressions to those deduced from GCMs. We analyse an ensemble of 106 GCM experiments from the Coupled Model Intercomparison Project phase 5 (CMIP5), corrected for natural variability. GCMs are for a large part physics-based, in contrast to trend methods.

However, there are also drawbacks, the main one being that GCMs are only approximations to the real climate system and have considerable biases. Although GCMs are tuned to meet the main characteristics of instrumental data (Voosen, 2016), GMTs derived from GCMs still show a wide range of trend-progression estimates, as we show.

In the discussion section, we address the role of various assumptions as for pre-industrial baselines, and differences in trend progression if Paris targets are interpreted as 'total warming' versus 'anthropogenic warming'.

Our analysis is confined to historic data only (up to and including 2016). Examples for GMT projections have been given by IPCC (2013 - Ch. 12), Forster et al. (2013), Mann (2014) and Schurer et al. (2017). A short-term prediction model is given by Suckling et al. (2016).

2 Data and methods

2.1 Data

Various research groups have published global GMT datasets. IPCC (2013 - section 2.4.3) used three datasets, namely the HadCRUT4 series (Morice et al., 2012; Hope, 2016), the NOAA dataset (Vose et al., 2012) and the NASA/GISS dataset (Hansen et al 2010). In the analysis here, we instead use a recent update of the NOAA data (Karl et al., 2015). Karl et al. applied a number of corrections which mainly deal with sea surface temperatures, such as the change from buckets to engine intake thermometers. In addition, we added two series, i.e. the version of the HadCRUT4 data in which the missing data have been filled in as published by Cowtan and Way (2014) and the GMT series by Rohde et al. (2013). Note that these datasets are not independent. They start from roughly the same station data over land, and more importantly are based on only two SST analyses: HadSST3 and ERSST v4.

Cowtan and Way re-analysed the HadCRUT4 series by applying a statistical interpolation technique (Kriging) and satellite data for regions where data are sparse. Their series shows higher GMT values in recent decades than the non-interpolated HadCRUT4 series due to the more-than-average warming of the poles. The land part of the GMT data of Rohde et al. (2013; Berkeley Earth group of researchers) systematically addressed major concerns of global warming sceptics, mainly dealing with potential bias from data selection, data adjustment, poor station quality and the urban heat island effect. The ocean part (about 70%) is taken from HadSST3.

Next to the GMT data products we apply the stratospheric aerosol optical depth (AOD) index to explore the influence of volcanic dust. These data are from NASA and are available for the period 1850-2016 (Sato et al, 1993; Ridley et al., 2014).

Since two out of five GMT products start in the year 1880, we use the period 1880-2016 as our period of analysis. We return to this point in the discussion section. All data were downloaded from the institution websites with 2016 as the final year.

Next to these instrumental-data based GMTs we analyze three sets of GCM simulations all taken from CMIP5 (Taylor et al., 2012; IPCC, 2013 – Ch. 9-12). GMT is defined here as the global average of near-surface temperature (temperature at surface or 'tas' in short), in contrast to the observational datasets that use SST over sea for practical reasons (also denoted as 'blended temperature series'; Cowtan et al., 2015). The first set consists of GCM simulations where the input of greenhouse gases from 2005 onwards is taken from three Representative Concentration Pathways (RCPs): 4.5, 6.0 and 8.5 W/m^2 (Van Vuuren et al., 2011; IPCC, 2014 - section 12.4 and figure 12.5). These simulations cover the period 1861-2100. We have taken a set of 106 GCM simulations with one member per model (42 members for emission scenario RCP4.5, 25 members for RCP6.0 and 39 members for RCP8.5). GMTs from CMIP5 simulations are based on wide range of modeling differences such as climate sensitivities, cloud parametrization and aerosol forcing (e.g., IPCC 2013 - Ch. 9).

The second set that we have analyzed, consists of 37 GCM runs for natural variability, denoted as 'historicalNat'. These runs comprise forced and unforced natural variability but no GHG forcing (1860-2005). See Forster et al. (2013) for details. Finally, we analyzed 41 Pre-industrial Control (PiControl) runs with lengths varying between 200 and 1000 years. These runs simulate natural internal variability only. All CMIP5 runs were downloaded from the KNMI Climate Explorer website with one member per model (Trouet and Van Oldenborgh, 2013).

2.2 Trend modeling

The tracking of signals or trends in GMT series has a long history, and a wide range of methods have been applied to isolate long-term signals or 'trends'. We have summarized these in the Supplementary Material (table SM.1). As stated in the Introduction we choose statistical trend methods that allow for the quantification of trend progression where no window is needed and where uncertainty estimates are available for any incremental trend value. Furthermore, no specific period for pre-industrial has to be chosen (such as the mean of the 1851-1900 period or similar). 'Pre-industrial' is reflected in the choice of the start of the sample period only.

Based on these considerations we have selected four trend approaches for our sensitivity analysis: Ordinary Least Squares (OLS) linear trends, Integrated Random Walk (IRW) trends and two approaches with splines. The first trend - a linear fit by OLS - was chosen by IPCC (2013) as their main method. Uncertainties simply follow from the linear model:

$$\mathrm{var}(\Delta\mu_{2016}) = \mathrm{var}([a+b*2016] - [a+b*1880]) = 125^2 * \mathrm{var}(b),$$

where 'a' is the intercept and 'b' the slope. The variance of 'b' follows from the OLS equations. Next to that the variance estimate is corrected by calculating effective sample sizes, based on annual data (IPCC, 2013 - 2SM).

This correction is important since residuals are not white noise due to persistence in natural processes. The signal is therefore considered as noise with a large decorrelation scale in this approach.

The second trend approach that fulfils our uncertainty requirements, are sub-models from the class of Structural Time Series models (STMs), in combination with the Kalman filter (Harvey, 1989). From this group of models we choose the IRW trend model. The IRW trend model extends the linear regression trend line by a *flexible trend* while retaining all uncertainty information (Visser, 2004; Visser et al., 2012; Visser et al., 2015). Furthermore, the flexibility of the trend model is optimized by Maximum Likelihood (ML) optimization. The Kalman filter is the ideal filter here since it yields the so-called Minimum Mean Squared Estimator (MMSE) for the trend component in the model. The Kalman filter has been applied in many fields of research and is gaining popularity in climate research recently (e.g., Hay et al., 2015).

A third and fourth approach applies a combination of a trend model and the statistical structure of natural internal variability as derived from PiControl runs. It can be seen as a hybrid approach. To do so we have chosen the cubic spline trend model, a trend approach also applied in the AR5 (IPCC, 2013 - Box 2.2, figure 1). Smoothing splines are not statistical in nature and, thus, do not generate uncertainty estimates for GMT increments $\Delta\mu_{2016}$. However, uncertainty bands can be reconstructed by Monte Carlo (MC) simulations under the assumption of a given mean, variance and autocorrelation structure estimated directly from the underlying dataset (figure 1 and Mudelsee 2014 - section 3.3). To steer the flexibility of the cubic spline model we studied the correlation structure of internal variability. This correlation structure can be described by an AutoRegressive Moving Average (ARMA) model as proposed by Hunt (2011) and Roberts et al. (2015). They estimated ARMA models to a range of PiControl runs. Similarly, we analyzed 41 PiControl runs and found that variability can reasonably be characterized by AR(1) processes where the AR(1) parameter φ varies within the range [0.28 - 0.60], depending on the GCM run chosen (cf. Mudelsee, 2014 - section 2.1).

All four trend methods are designed to smooth GMTs for annual to decadal natural variability (forced and unforced). However, if Paris targets should be interpreted as anthropogenic warming only, we should estimate the role of decadal to centennial forcings from volcanic and solar activity as well. To estimate the role of volcanic eruptions we have extended the OLS linear trend model and the IRW trend model by adding the AOD index as regressor (Visser and Molenaar, 1995; Visser et al., 2015 - figure 4).

3 Results

3.1 Sensitivity analysis trend methods and data products

Based on the 1880-2016 GMT sample period we have evaluated trend progression values $\Delta\mu_{2016}$ from 1880 up to 2016 along with uncertainties for all datasets and trend approaches. This yields the 4-by-5 matrix shown in table 1. As for linear trends we corrected uncertainty estimates by a factor $\sqrt{(1.60/0.40)} = 2.0$, analogous to the approach chosen in IPCC (2013 - Ch. 2, Sup. Mat.) since first-order autocorrelations lie around 0.60. Table 1 shows that the trend slopes for the dataset HadCRUT4, LOTI-NASA, NOAA-Karl and Cowtan and Way are close, where the lowest slope value is for the HadCRUT4 series. This dataset has poor coverage in the Arctic, where trends are much higher than the global mean. The steepest trend is found for the Berkeley Earth series, a remarkable result since the Berkeley Earth project was set-up to meet a range of critical comments from global warming sceptics. Identical patterns are found for the other trend models: lowest trend progression for the HadCRUT4 dataset and highest values for the Berkeley Earth dataset.

As for the IRW trend estimates we find reasonable flexible patterns which closely resemble the spline trend shown in IPCC (2013 - Ch.2: Box 2.2, figure 1b). An example for the HadCRUT4 dataset is shown in figure 2. Data, trend and uncertainties are shown in the upper panel. The trend increments [μ_t - μ_{t-1}] and [μ_t - μ_{1880}] are given in the middle left and right panel, respectively, along with uncertainties. The [μ_{2016} - μ_{1880}] value with uncertainty is taken as value in table 1. The lower left panel shows the innovations or one-step-ahead predictions errors which follow from the Kalman filter formulae. The lower right panel shows the autocorrelation function (ACF). We note that a prerequisite of Kalman filtering is that the one-step-ahead prediction errors follow a white noise process. The ACF shows an AR(1) value of 0.30 which is slightly significant. We applied a correction for compensating for this the violation by applying the approach of IPCC, as we did for linear trends: uncertainty bands are corrected by a factor $\sqrt{(1.30/0.70)}$ = 1.3.

As for smoothing splines we have estimated trends in GMT series such that the residual series exhibits an AR(1) process with a φ value of 0.28 and 0.60. Trend estimates based on the HadCRUT4 series are shown in figure 3. Both spline approaches show quite different trend patterns. The model shown in the upper panel of figure 3 is based on a slightly correlated noise process and - as for the IRW trend from figure 2 - closely resembles the spline trend shown in IPCC (2013 - Ch.2: Box 2.2, figure 1b). The model shown in the lower panel shows a parabolic shape. This parabolic pattern closely resembles the anthropogenic signal in GMT series as shown by IPCC (2013 - figure 10.1f), derived from 'historicalGHG' simulation runs (Forster et al., 2013).

It is interesting to note that none of the four trend methods show a sign of a 'hiatus', 'slowdown' or 'pause'. That is not surprising for the linear trend and the spline estimate with φ = 0.60 due to their stiff character. However, the IRW trend and spline with φ = 0.28 are more flexible and do not show any stabilisation pattern for recent years at all. We tested the residuals of the IRW trend model and these appear to be close to white noise (cf. lower panels of figure 1). This inference is consistent with recent findings on the hiatus (Marotzke et al., 2015; Hedemann et al., 2017; Medhaug et al., 2017; Rahmstorf et al., 2017).

Table 1 shows that differences between trend model and dataset combinations can be considerable. The lowest $\Delta\mu_{2016}$ value is found for the HadCRUT4 dataset in combination with the IRW trend model: 0.90 ± 0.18°C (± 2σ). The highest values are found for the Berkeley Earth dataset in combination with cubic spline interpolation and φ = 0.28: 1.12 ± 0.13 °C. These two extremes reveal that the range of $\Delta\mu_{2016}$ values due to datasets and trend models accounts for 0.22 °C. This range is somewhat lower than that due to natural variability alone. Based on 2σ limits, we find a low estimate of ± 0.12 °C, leading to a maximum range of 0.24 ºC (LOTI dataset in combination with cubic spline interpolation and $\varphi = 0.28$), and a high estimate of ± 0.19 °C, leading to a maximum range of 0.38 ºC (LOTI dataset and OLS linear trend).

To quantify the role of trend methods in more detail we have averaged trend estimates over the five GMT datasets and added it to table 1 (bottom row). It shows that the range of trend progressions is small: [0.97, 1.01] ºC. At the other hand, if we average *over trend methods*, the variability due to datasets is found (right column of table 1). The variability accounts for [0.92, 1.09] ºC. Clearly, variability due to GMT datasets is dominant over specific trend approaches.

3.2 Trend progression derived from GCM simulations

Trend progression derived from GCMs have been analyzed in a range of studies, e.g. IPCC(2013 - Ch. 10), Forster et al. (2013), Marotzke and Forster (2016), Mann et al. (2016) and Meehl et al. (2016). Here, we derive trend progression since pre-industrial by taking an ensemble of 106 GCM all-forcing simulations 1861-2016. We note that underlying models have quite different characteristics. However, we did not perform an extensive sensitivity analysis as for these factors, as for example in Visser et al. (2000).

Short-term forced and unforced natural variability in individual GCM simulations is smoothed by estimating splines to each individual simulation (both for $\varphi = 0.28$ and $\varphi = 0.60$, as in figure 3). In this way we find 106 values for $\Delta_{i,2016} \equiv y_{i,2016} - y_{i,1861}$. Results are shown in figure 4 (based on smoothing splines with $\varphi = 0.28$). The mean Δ_{2016} value is 1.15 ± 0.50 ºC (2σ) for smoothing all 106 curves with $\varphi = 0.28$ and 1.00 ± 0.50 ºC for smoothing with $\varphi = 0.60$. These values are consistent with those reported by Forster et al. (2013, table 3).

The GCM simulations analyzed here differ from data products as for their definition of temperatures ('tas only' versus blended temperatures). Cowtan et al. (2015) and Richards et al. (2016 - figure 1) showed that tas temperatures differ from blended temperatures by 0.10 ºC, for the period 1860-2009. Thus, mean GCM-derived warming estimates cover the ranges [1.00 to 1.15] ºC (tas) or [0.90 to 1.05] ºC (blended). We note that these ranges reasonably correspond to the range found in table 1.

4 Discussion

4.1 Uncertainty and sensitivity analysis

We make three comments concerning the robustness of the results given in section 3. First, as summarized in table SM.1 of the Supplementary Material section, a wide range of trend models exist in the literature, all with varying characteristics. The fact that many of these methods are not statistical in nature does not limit their application in the present context: the approach shown in figure 1 (creating surrogate GMT series by MC simulation) is also applicable to methods such as binomial filters or LOESS estimators. Therefore, we cannot rule out that the influence of trend modelling is underestimated in table 1. However, given the (i) small differences shown in the bottom row of table 1, and (ii) the wide uncertainty bands due to natural variability, we judge such an under-estimation to be relatively small.

A second comment concerns a source of uncertainty dealing with the choice for year or period that can be regarded as 'pre-industrial'. As for the analyses in section 3.1, we have chosen for the year 1880 as low end of the sample period, simply because two out of five GMT products start in 1880 (NASA and NOAA). This choice is consistent with that made by IPCC (2013) as for historic trend progression (without claiming this to be 'since pre-industrial'). In section 3.2 we have chosen the year 1861 as low end of the sample period, again since simulations are available from that year onwards.

Would our results and conclusions from table 1 or figure 4 be different if the sample period would be enlarged, starting in 1400, 1720 or 1850? Strictly spoken, we cannot answer this question since we cannot extend our analyses to these starting years due to data availability. As for instrumental dataset, we could perform some analyses from 1850 onwards but GMT estimates become inaccurate for these early decades. However, estimates based on GCM simulations are given by Hawkins et al. (2017) and Schurer et al. (2017).

Hawkins et al. show that the GMT difference between the two periods 1720-1800 and 1850-1900 is small, around 0.05 ºC, lying on the edge of statistical significance. Additionally to their analysis we compared GMT mean values over three periods: 1850-1900, 1860-1880 and 1880-1900, based on the HadCRUT4 dataset. The mean values appear to be similar: -0.31 ± 0.03 °C, -0.31 ± 0.06 °C and -0.32 ± 0.05 °C, respectively (2σ limits). These differences are small if compared to the uncertainties due to natural variability, shown in table 1. These results suggest that the choice for 1720-1800, 1850-1900, 1860-1880 or 1880-1900 as 'pre-industrial' will have a small influence to the findings presented here. At the other hand, Schurer et al. show from GCM simulations that global warming is underestimated by 0.09 [0.03, 0.19] ºC if the period 1401-1800 is chosen as pre-industrial baseline (compared to the period 1850-1900). Their estimate for the influence of GHG only lies close to these estimates, in the range from 0.02 to 0.20 ºC. We conclude that recent simulations point to an underestimation of global warming if calculated relative to late nineteenth century estimates. The underestimation lies around 0.10 ºC.

A third comment deals with differences in warming definitions as mentioned in the Introduction. If the Paris targets should be interpreted as anthropogenic warming only, we should estimate these contributions as well. Clearly, the incremental estimates $\Delta\mu_{2016}$ shown in Table 1 do not contain corrections for decadal to centennial natural forcings from solar and volcanic activity. To estimate the role of volcanic activity on the estimates given in table 1 we have extended the OLS linear trend and the IRW trend model with a regression component where GMT series are regressed on the OAD index shown in figure 5. Results are summarized in table SM.2. The table shows that incremental estimates $\Delta\mu_{2016}$ are overestimated by 0.02 ºC for linear trends and by 0.04 ºC for IRW trends.

To estimate the role of long-term solar activity we did not choose for the time-series approach above since any explanatory variable in a regression model with some long-term trend will correlate and 'explain' the long-term trend in the dependent variable. Therefore, we refer to GCM estimates for the role of solar activity.

IPCC (2013) estimates the role of solar variability to be small and on the edge of significance. Incremental solar forcing for the period 1750-2011 accounts for 2 [0, 4] % of GHG forcing (Figure SPM.5 and Box 10.2). Schurer at al. (2017 - figure S3) estimate the incremental contribution of solar forcing on GMTs to be 0.07 [0.02, 0.12] ºC. This estimate compares the period 1850-1900 to 1990-2000. Furthermore, the long-term influence of volcanic activity is non-significant in their simulations (their figure S2).

Next to these estimates we analyzed an ensemble of 37 GCM simulations with natural forcing only ('historicalNat'; IPCC, 2013 - figures 10.1 and 10.7; Forster et al., 2013 - fig 2). The mean curve with 2 standard errors (SEs) is shown in figure SM.2, along with major volcanic eruptions (eruptions with a Volcanic Explosivity Index of 5 and 6). Mean trend progression for these 37 runs accounts for 0.078 ± 0.030 ºC (2 SE), 1861-2005.

From these inferences we conclude that the difference between total warming and 'anthropogenic warming lies around 0.10 ºC with an uncertainty range of [0.0, 0.14].

4.2 Policy recommendation

Schurer et al. (2017) end their article with the recommendation that a consensus be reached as to what is meant by pre-industrial temperatures. In this way, the chance would be reduced of conclusions that appear contradictory being reached by different studies. Furthermore, it would allow for a more clearly defined framework for policymakers and stakeholders. We fully agree with this recommendation. However, our uncertainty and sensitivity analysis has shown that the choice of a proper pre-industrial baseline is not the only parameter that could lead to contradictory results. Decisions around data products and GCM simulations, various time series techniques, or warming assumptions should be taken into account as well.

Here, we make the following policy proposal which aims to be a reasonable compromise. First, we propose to base GMT warming estimates on data products rather than GCM simulations. Our argumentation is that Δ_{2016}

values based on GCM simulations show a wide range of warming estimates (figure 4). We note that even wider ranges are found for *absolute* GMT estimates (CMIP5 estimates for the mean GMT value over the period 1961-1990 show a range of 2.5 ºC according to IPCC 2013 - figure 9-8). Another argument is that simulation estimates from CMIP5 are accurate up to the year 2005 (estimates for 2006-2016 apply approximations for GHG concentrations, and no volcanic and solar activity).

Second, since warming estimates vary as a function of the GMT data product chosen (tabel 1), we propose to estimate trends on the annual averages of all five data products.

Third, we found that the choice for specific trend methods plays a minor role, with largest differences between stiff and more flexible trend models. Therefore, we propose to apply a flexible and a stiff trend method and average the warming estimates found.

Fourth, two studies on the role of pre-industrial baselines have been published recently. Schurer et al. (2017) find a GHG-induced warming in the range [0.02, 0.20] ºC if the period 1401-1800 is compared to the period 1850-1900. Hawkins et al. (2017) define the period 1720-1800 as a reasonable baseline for pre-industrial and find small non-significant differences between the period 1720-1800 and 1850-1900. We choose to follow the baseline proposed by Hawkins et al. Since all five GMT data products have data from 1880 onwards and GMT mean values for 1850-1900 and 1880-1900 are of equal size (based on the HadCRUT4 data product), we propose to analyse trend progression from 1880 onwards.

Finally, we propose to interpret global warming in the context of "Paris" as the sum of natural and anthropogenic warming, consistent with the IPCC definition of climate change. One argument for this choice is that ecological systems and human society will respond to total warming and induced shifts in climate extremes *regardless of its origin*.

From these choices it follows that trend progression Δ_{2016} accounts for 1.00 ± 0.13 ºC (bottom row of table 1).

5 Conclusions

We have addressed the issue of signal progression of GMT in relation to the GMT targets agreed upon in Paris in December 2015. Although these targets are clearly defined – avoiding increments of 1.5 and 2.0 °C – there remain a number of (scientific) questions unaswered in the agreement. We have identified five aspects of the accord which hamper an exact quantification of GMT progression: (i) the use of instrumental data and trend methods versus GCM-derived progression, (ii) the role of varying datasets, (iii) the role of varying trend methods, (iv) the role of varying choices for pre-industrial and (v) the role of warming definitions. Since there is no 'true' or 'best' approach

(Visser et al., 2015), we have chosen to perform an uncertainty and sensitivity on GMT progression as propagated by Saltelli et al. (2004) and related articles. This allows us to test the robustness of various trend progression claims.

Approaches based on instrumental data. We find that trend values for GMT progression 1880-2016 vary considerately, from 0.90 ºC (HadCRUT4 dataset in combination with the IRW trend model) to 1.12 ºC (Berkeley Earth dataset in combination with cubic spline interpolation and $\varphi = 0.28$). The two extremes reveal that the range of $\Delta\mu_{2016}$ values due to datasets and trend models accounts for 0.22 °C. This range is smaller than that due to natural variability alone. Based on 2σ limits, we find a low estimate of 0.24 °C (LOTI dataset in combination with cubic spline interpolation and $\varphi = 0.28$) and a high estimate of 0.38 °C (LOTI dataset and OLS linear trend). Furthermore, variability due to various GMT products dominates the variability due to specific trend approaches.

Approaches based on GCMs. We find that mean trend progressions lie within the range of estimates from instrumental data. However, the uncertainty bands for 106 simulations are much wider than those derived from instrumental trend estimates. Here, GCM variability stems from a wide range of modeling assumptions such as climate sensitivities, cloud parameterization and aerosol forcing (e.g., IPCC, 2013 - Ch. 9), rather than from natural variability.

The choice of a pre-industrial period. Recent studies have shown that GHG warming prior to 1880 or 1850 cannot be neglected. Schurer et al. (2017) estimate that early warming (1401-1800 compared to 1850-1900) accounts for 0.09 [0.03, 0.19] ºC. The role of solar and volcanic activity is minimal in this comparison.

Interpretation of Paris targets as being 'total warming' or 'anthropogenic warming only'. We find that the role of solar and volcanic activity is small on centennial scale. This contribution lies around 0.10 ºC (+0.03 ºC from volcanic activity and +0.07 ºC from solar activity).

Hiatus. As a side result of our trend analyses we note that no signs of an 'hiatus', 'slowdown' or 'pause' can be discerned in GMT trend progression. This inference is consistent with recent findings (Marotzke et al. 2015, Hedemann et al. 2017, Medhaug et al. 2017, Rahmstorf et al. 2017).

Policy recommendation. Schurer et al. (2017) recommend that a concensus be reached as to what is meant by pre-industrial temperatures. Our analysis shows that other sources of uncertainties should be taken into account as well. If not, contradictory results will appear in different studies with direct consequences for CO_2 reductions to hold GMTs below the Paris targets. Our proposal shows a GMT progression Δ_{2016} of 1.00 ºC.

Table 1. Trend increments $\Delta\mu_{2016}$ along with 2σ confidence limits. Increments are given for five GMT series and four trend approaches.

GMT dataset	GMT progression $\Delta\mu_{2016}$ with 2σ confidence limits (°C)				
	OLS linear trend	**IRW trend**	**Spline with φ=0.28**	**Spline with φ=0.60**	**Mean progression**
HadCRUT4, CRU	0.90 (± 0.18)	0.93 (± 0.17)	0.94 (± 0.12)	0.92 (± 0.14)	0.92
HadCRUT4, Cowtan and Way	0.96 (± 0.17)	1.06 (± 0.17)	1.06 (± 0.12)	0.98 (± 0.15)	1.02
LOTI series, NASA	0.98 (± 0.19)	1.02 (± 0.18)	1.01 (± 0.12)	0.99 (± 0.14)	1.00
Karl *et al* (2015), NOAA	0.95 (± 0.19)	0.96 (± 0.19)	0.94 (± 0.14)	0.95 (± 0.14)	0.95
Berkeley Earth Project	1.04 (± 0.17)	1.12 (± 0.17)	1.12 (± 0.13)	1.06 (± 0.14)	1.09
Mean progression	0.97	1.02	1.01	0.98	1.00

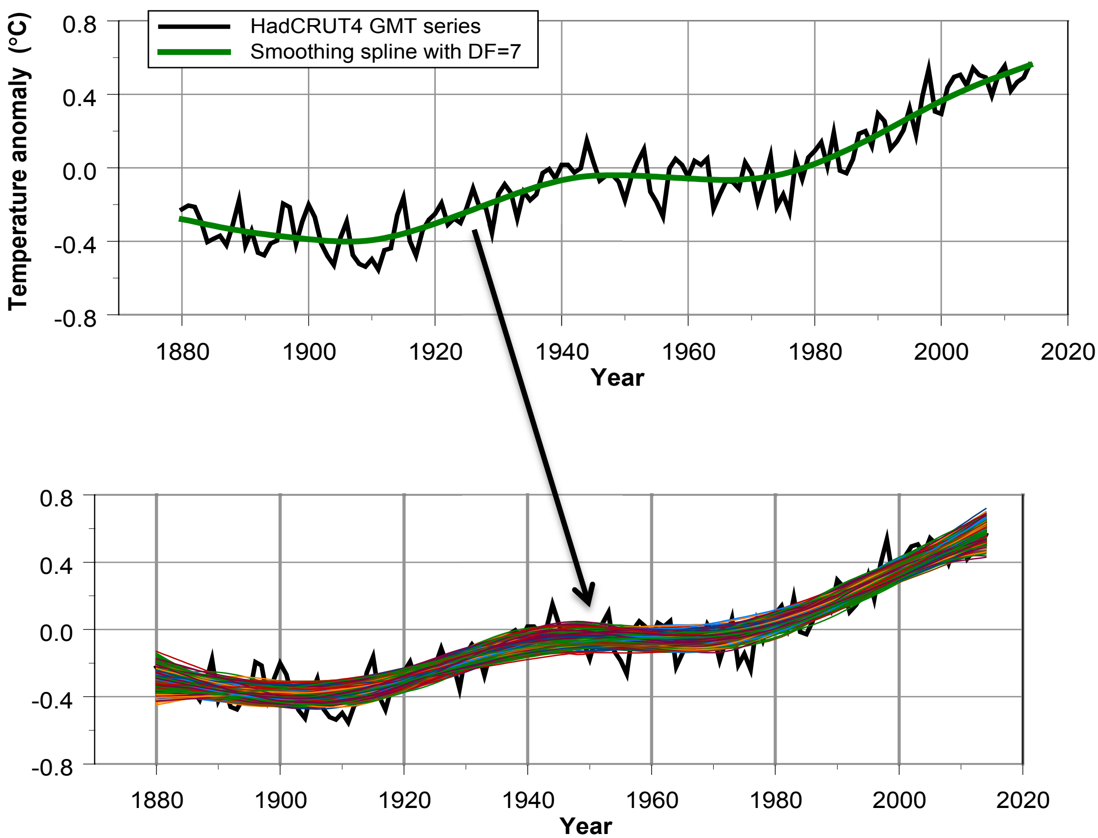

Figure 1. Construction of 1000 surrogate trend series by MC simulation, based on cubic splines. The AR(1) parameter estimated on the residuals of the spline model in the upper panel, accounts for 0.28. A surrogate GMT series $\hat{y}_{i,t}$ is formed by simulating a new residual series $r_{i,t}$ based on the AR(1) process with φ= 0.28, and adding it to the estimated spline (green line upper panel). Then, a spline trend $\mu_{i,t}$ is estimated for each surrogate $\hat{y}_{i,t}$. As an illustration we have plotted 1000 of such trends $\mu_{1,t}$, ... , $\mu_{1000,t}$ in the lower panel. Now, confidence limits can be estimated for any μ_t based on the values $\mu_{1,t}$, ... , $\mu_{1000,t}$. These confidence limits can be based on standard deviations or percentiles. Similarly, confidence limits can be calculated for the increment [μ_{2016} - μ_{1880}], based on the values [$\mu_{1,2014}$ - $\mu_{1,1880}$] ,...... , [$\mu_{1000,2014}$ - $\mu_{1000,1880}$] (Mudelsee, 2014 - Sections 3.3.3 and 3.4).

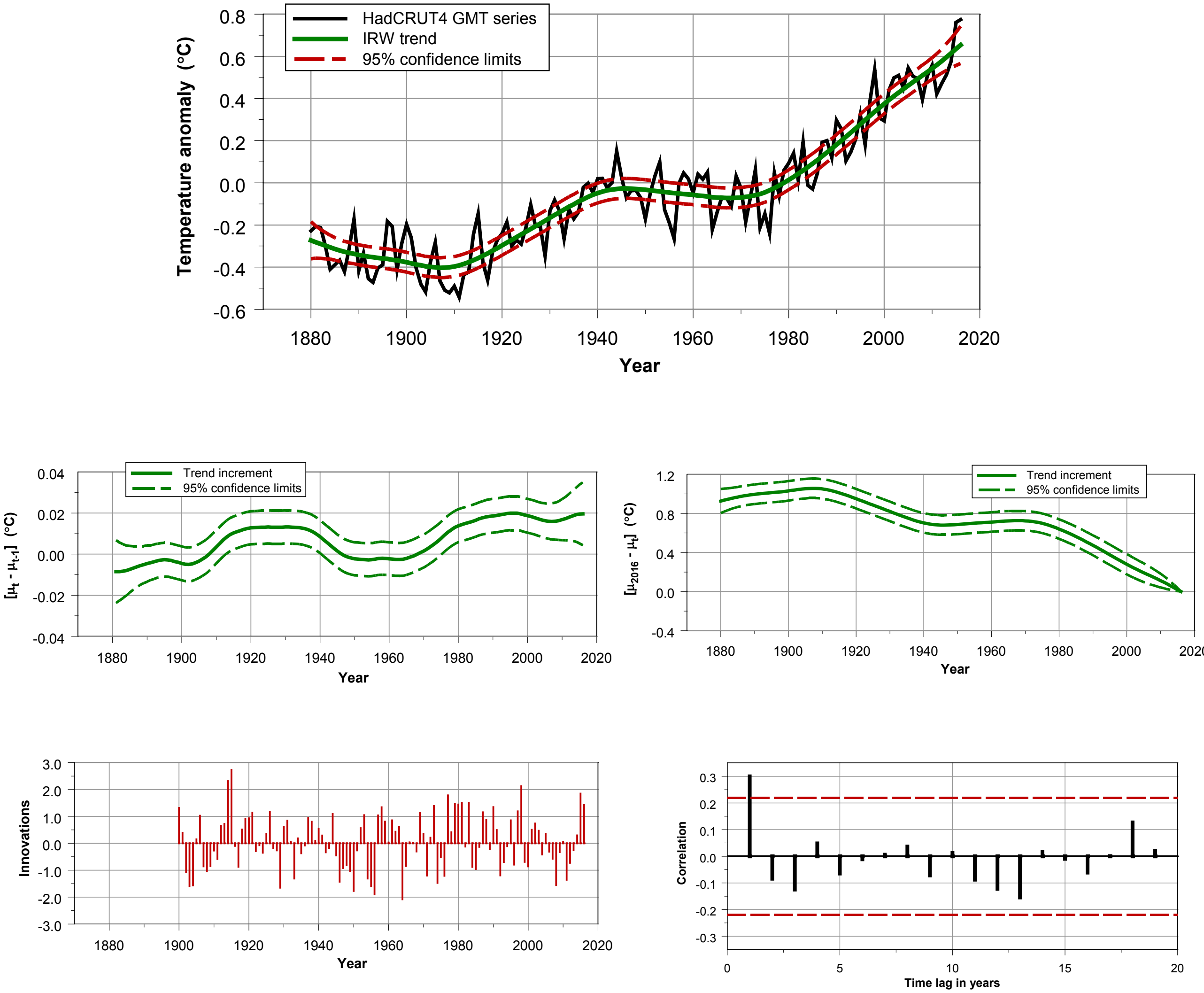

Figure 2. Results for the IRW trend model as applied to the HadCRUT4 series. Period: 1880-2016. The upper panel shows the trend (green line) along with 95% confidence limits (red dashed lines). The trend increments $[\mu_t - \mu_{t-1}]$ are given in the middle left panel along with uncertainties. Idem the $[\mu_t - \mu_{1880}]$ values in the middle right panel. The lower left panel shows the innovations or one-step-ahead predictions errors which follow from the Kalman filter formulae. The lower right panel shows the autocorrelation function (ACF).

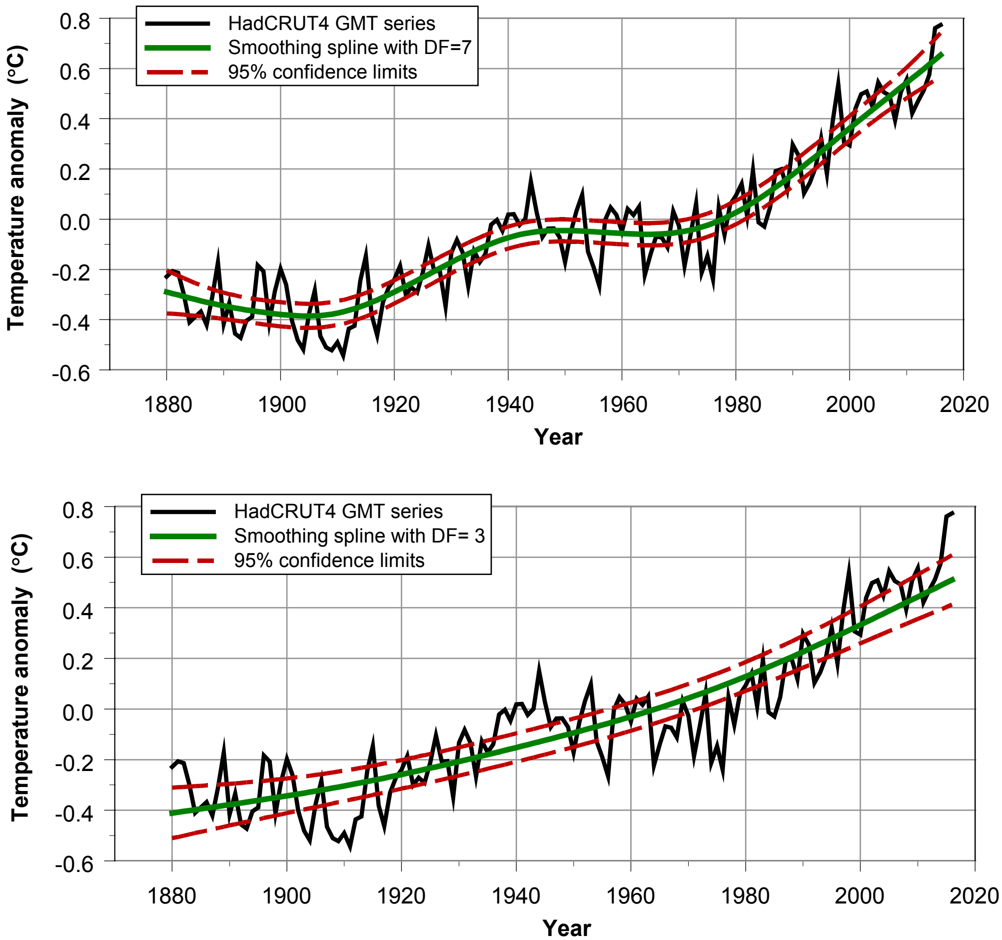

Figure 3. Two smoothing spline estimates for the HadCRUT4 GMT series, with uncertainties generated by MC simulation. All confidence limits are based on 1000 surrogate GMT series following the approach set out in Mudelsee (2014 - Section 3.3.3). Upper panel: AR(1) parameter chosen as φ = 0.28 (equivalent to 7 degrees of freedom), the low end of φ values within CMIP5 PiControl runs. Lower panel: AR(1) parameter chosen as φ = 0.60, the high end of φ values (DF=3).

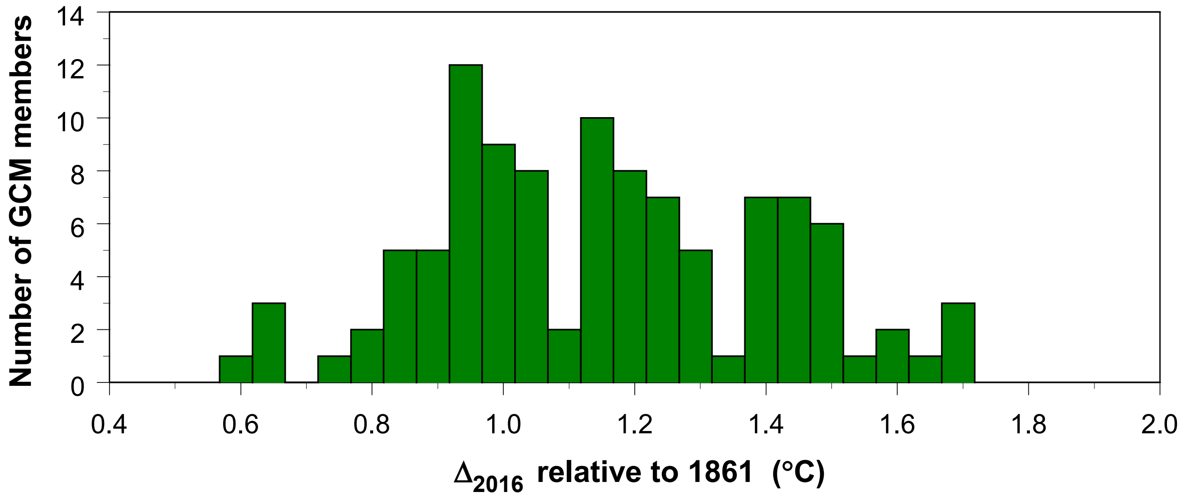

Figure 4. Histogram based on 106 GCM $\Delta_{i,2016}$ values, relative to 1861. Mean value is 1.15 ± 0.50 ºC (2σ). Individual GCM curves were smoothed by splines where the AR(1) parameter is chosen as $\varphi = 0.28$ (equivalent to 7 degrees of freedom), the low end of φ values within CMIP5 PiControl runs.

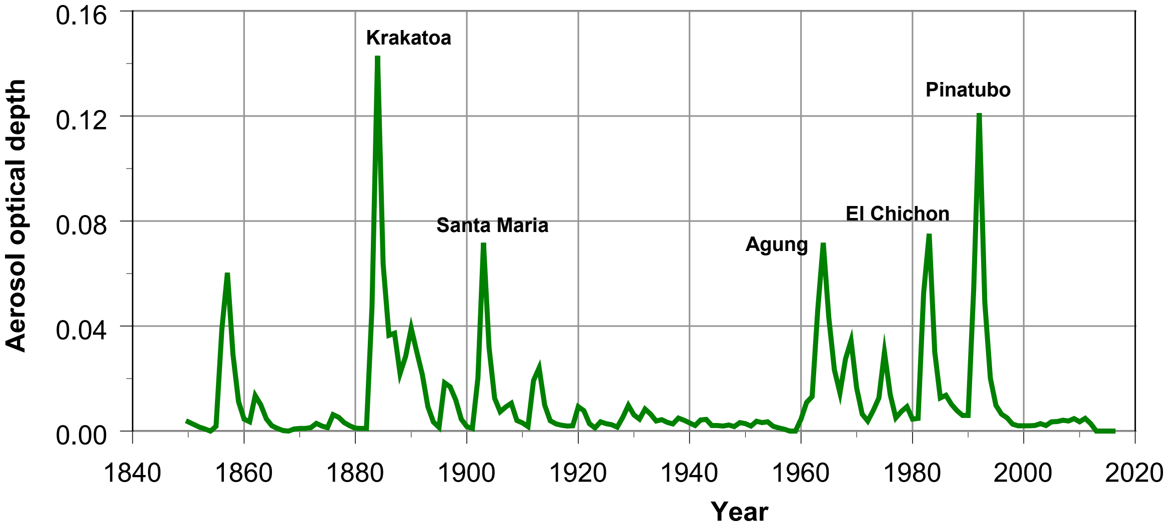

Figure 5. The AOD index series as introduced by Sato et al. (1993). Period is 1850-2016.

Code availability. IRW trends have been estimated by the TrendSpotter software. This software package is freely available from the first author. Splines have been estimated by the statistical package S-Plus, version 8.2. The scripts which are highly similar to R, are available from the first author.

Data availability. All five GMT datasets are open access and have been downloaded from the authors websites. All CMIP5 runs named in Section 2.1 were downloaded from the KNMI Climate Explorer website with one member per model (Trouet and Van Oldenborgh, 2013). The names of individual GCMs can be found there as well. Please see https://climexp.knmi.nl/cmip5_indices.cgi?id=someone@somewhere . Data used for the graphical presentations in this article can be gained from the first author.

The Supplement related to this article is available online

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgments. We thank Geert Jan van Oldenborgh (KNMI, Climate Explorer) for thorough comments on an early version of this text. Furthermore, we thank Peter Thorne (Maynooth University), an anonymous reviewer and Lenny Smith (The London School of Economics and Political Science) for important comments on the manuscript.

References

Climate Analytics: The 1.5 ºC temperature limit in the Paris Agreement and 2016 temperature records, Briefing paper, 2016.

Cowtan, K. and Way, R. G.: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends, Q.J.R. Meteor. Soc., 140, 1935-1944, 2014.

Cowtan, K., Hausfather, Z., Hawkins, E., Jacobs, P., Mann, M.E., Miller, S. K., Steinman, B. A., Stolpe, M. B. and Way, R. G.: Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures, GRL, doi:10.1002/2015GL064888 6527-6534, 2015.

Forster, P.M., Andrews, T., Good, P., Gregory, P.M., Jackson, L. S., Zelinka, M.: Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models, JGR: Atmospheres 118, 1139-1150, 2013.

Fyfe, J. C., Meehl, G. A., England, M. H., et al: Making sense of the early-2000s warming slowdown, Nature Climate Change, 6, 224-228, 2016.

Hansen, J., Ruedy, R., Sato, M. and Lo, K.: Global surface temperature change, Reviews of Geophysics, 48, RG4004, 2010.

Harvey A.C.: Forecasting, structural time series models and the Kalman filter, Cambridge University Press, Cambridge United Kingdom, 1989.

Hawkins, E., Ortega, P., Suckling, E., Schurer, A., Hegerl, G., Jones, P., Joshi, M., Osborn, T., Masson-Delmotte, V., Mignon, J., Thorne, P. and Van Oldenborgh, G.: Estimating changes in global temperature since the pre-industrial period, Bull. Amer. Meteor. Soc., doi:10.1175/BAMS-D-16-0007, preliminary version on Internet, 2017.

Hay, C.C., Marrow, E., Kopp, R. E. and Mitrivica, J.X.: Probabilistic reanalysis of twentieth-century sea-level rise, Nature, 517, 481–484, 2015.

Hedemann, C., Mauritsen, T., Jungclaus, J., Marotzke, J.: The subtle origins of surface-warming hiatuses, Nature Climate Change, 7, 336-339, 2017.

Hope, M.: Temperature spiral goes viral, Nature Climate Change, 6, 657, 2016.

Hunt, B.G.: The role of natural climatic variation in perturbing the observed global mean temperature trend, Clim. Dyn., 36, 509-521, 2011.

IPCC: Climate Change 2013: the Physical Science Basis. Contribution of Working Group I to the Fifths Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T. F., Qin, D., Plattner G. K., et al. (eds)] Cambridge: Cambridge University Press, 2013.

IPCC: Climate Change 2014:Mitigation of climate change. Contribution of Working Group III to the Fifths Assessment Report of the Intergovernmental Panel on Climate Change [Edenhofer, O., Pichs-Madruga, R. et al. (eds)] Cambridge: Cambridge University Press, 2014a.

IPCC: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the IPCC [Pechauri, R. K. and Meyer, L. A. (eds)] IPCC Geneva, Switzerland, 2014b.

IPCC Annex II: Glossary, Contribution of Working Groups I, II and III to the Fifth Assessment Report of the IPCC [Pechauri, R. K. and Meyer, L. A. (eds)] IPCC Geneva, Switzerland, 2014.

Karl, T. R., Arguez, A., Huang, B., et al: Possible artifacts of data biases in the recent global surface warming hiatus, Science, 348, 1469-1472, 2015.

Kosaka, Y. and Xie, S.P.: Recent global-warming hiatus tied to equatorial Pacific surface cooling, Nature, 501, 403-408, 2013.

Lewandowsky, S., Oreskes, N., Risbey, J. S., Newell, B. R.: Seepage: climate change denial and its effect on the scientific community. Global Environm. Change, 33: 1-13, 2015.

Mann, M. E.: Earth will cross the climate danger threshold by 2036, Scientific American, April issue, 2014.

Mann, M. E., Rahmstorf, S., Steinman, B. A., Tingley M. and Miller, S.K.: The likelihood of recent record warmth, Nature Scientific Reports, DOI:10.1038/srep19831, 2016.

Marotzke, J., Forster, P. M.: Forcing, feedback and internal variability in global temperature trends, Nature, 517, 565-570, 2015.

Medhaug, I., Stolpe, M. B., Fischer, E. M. and Knutti, R.: Reconciling controversies about the 'global warming hiatus', Nature, 545, 41-47, 2017.

Meehl, G. A., Hu, A., Santer, B. D. and Xie, S-P: Contribution of the Interdecadal Pacific Oscillation to twentieth-century global surface temperature trends, Nature Climate Change, 6, 1005-1008, 2016.

Millar, R.J., Fuglestvedt, J. S., Friedlingstein P., Rogelj, J., Grubb, M. J., Matthews, H. D., Skeie R. B., Forster, P. M., Frame, D. J., Allen, M.R.: Emission budgets and pathways consistent with limiting warming to 1.5 ºC, Nature Geoscience, 10, 741-747, 2017.

Morice, C. P., Kennedy, J. J., Rayner, N. A. and Jones, P.D.: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 data set, J. of Geophys. Res., 117**,** doi:10.1029/2011JD017187, 2012.

Mudelsee, M.: *Climate time series analysis: classical statistical and bootstrap methods,* Springer, New York, USA, 2014.

Otto, R. E. L., Frame, D. J., Otto, A., Allen, M. R.: Embracing uncertainty in climate change policy, Nature Clim. Change, doi:10.1038/NCLIMATE2716, 2015.

Rahmstorf, S., Forster, G. and Cahill, N.: Global temperature evolution: recent trends and some pitfalls, Environ. Res. Lett., 12, 054001, 2017.

Ridley, D. A., Solomon, S., Barnes, J. E., et al.: Total volcanic stratospheric aerosol optical depths and implications for global climate change, Geophys. Res. Letters, 41, doi:10.1002/2014GL061541, 2014.

Roberts, C. D., Palmer, M. D., McNeall, D., Collins, M.: Quantifying the likelihood of a continued hiatus in global warming, Nature Clim. Change, doi:10.1038/NCLIMATE2531, 2015.

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., Rosenfeld, A., Wurtele, J., Curry, J., Wickham, C., Mosher, S.: Berkeley Earth temperature averaging process, Geoinformatics & Geostatistics: An Overview 1/2, 1-13, 2013.

Saisana, M., Saltelli, A., Tarantola, S.: Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators, J. R. Statist. Soc. A, 168, 307-323, 2005.

Saltelli, A., Tarantola, S., Campolongo, F. and Ratto, M.: *Sensitivity analysis in practice* Wiley & Sons, Chichester UK, 2004.

Sato, M., Hansen, J. E., McCormick, M. P., Pollack, J. B.: Stratospheric aerosol optical depths 1850-1990, J. of Geophys. Res., 98, 22987-22994, 1993.

Schurer, A.P., Mann, M.E., Hawkins, E., Tett, S.F.B., Hegerl, G.C.: Importance of the pre-industrial baseline for likelihood of exceeding Paris goals, Nature Clim. Change, doi:10.1038/NCLIME3345, 2017.

Suckling, E. B., Van Oldenborgh, G. J., Eden, J. M. and Hawkins, E.: An empirical model for probabilistic decadal prediction: global attribution and regional hindcasts, Clim Dyn, doi:10.1007/s00382-016-3255-8, 2016.

Taylor, K. E., Stouffer, R.J., Meehl, G. A.: An overview of CMIP5 and the experiment design, BAMS april issue, 485-498, 2012.

Thomson, A. M. et al: RCP4.5: a pathway for stabilization of radiative forcing by 2100, Clim. Change, 109,77-94, 2011.

Tollefson, J.: The 2 °C dream, Nature 527, 436-438, 2015.

Trenberth, K. E.: Has there been a hiatus? Science, 349, 691-692, 2015.

Trouet, V. and Van Oldenborgh, G. J.: KNMI Climate Explorer: a web-based research tool for high-resolution paleoclimatology, Tree-Ring Research, 69(1), 3-13, 2013.

UN: Adoption of the Paris Agreement. FCCC/CP/2015/L.g/Rev.1, 2015. http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf

Van Vuuren, D., Edmonds, J., Kainuma M., et al: The representative concentration pathways: an overview, Clim. Change, 109, 5-31, 2011.

Visser, H.: Estimation and detection of flexible trends, Atmosph. Environ. 38, 4135-4145, 2004.

Visser, H. and Molenaar, J.: Trend estimation and regression analysis in climatological time series: an application of structural time series models and the Kalman filter, J. of Climate, 8(5), 969-979, 1995.

Visser H, Folkert R. J. M., Hoekstra J. and De Wolf J. J.: Identifying key sources of uncertainty in climate change projections, Clim. Change, 45, 421-457, 2000.

Visser, H. and Petersen, A.C.: Inferences on weather extremes and weather-related disasters: a review of statistical methods, Clim. of the Past, 8, 265-286, 2012.

Visser, H., Dangendorf, S., Petersen, A.C.: A review of trend models applied to sea level data with reference to the "acceleration-deceleration debate", J. of Geophys. Res.: Oceans, 120, 3873-3895, doi:10.1002/2015JC010716, 2015.

Voosen, P.: Climate scientists open up their black boxes to scrutiny, Science, 354, 401-402, 2016.

Vose RS et al.: NOAA's merged land-ocean surface temperature analysis, Bull. Am. Meteor. Soc., 93, 1677- 1685, 2012.

Xie, S. P.: Leading the hiatus research surge, Nature Clim. Change, 6, 345-346, 2016.