Supplementary Materials

1 Data description

To check the potential impact on the identification of the recent slowdown from dataset choice and to guarantee comparable and general results, we attempted to collect as many datasets as possible. Finally, we obtained 32 datasets which cover almost all the available routinely updated global surface temperature datasets widely used in the hiatus research. The 32 datasets include 12 global merged land/marine surface temperature (MST) datasets and 20 sea surface temperature (SST) datasets. The 12 MST datasets include six observations and six reanalyses. There is no “surface temperature” in the reanalysis and thus the “2-meter temperature” was analyzed instead. The 20 SST datasets consist of eight observations, six satellites, and six reanalyses. These datasets are not completely independent. But different groups adopt different raw observations and employ different techniques to generate their respective products, which brings these datasets a certain degree of discrepancy. The relevant information for the 32 datasets could be found in Table S1. The global averaged annual anomalies against the 1981-2010 baseline from all the 32 temperature datasets were calculated to explore the evolution characteristics of the global surface temperature (Figs. S1-S2).

Most results of this study (sections 3a, 3c and 3d) was based on six observational MST anomaly estimates produced by incorporating land surface air temperature (LST) and SST. For one thing, both global warming and warming slowdown were commonly recognized based on the observed global mean surface temperature (GMST) derived from the observational combined land/marine surface temperature dataset. For another, the six datasets have enough temporal coverage for extracting the centennial trend and presenting the full picture of global temperature change. The Berkeley Earth Surface Temperature (BEST) combines the Berkeley Earth land-surface temperature and a reinterpolated version of the HadSST3 (Rohde et al. 2013a, 2013b). They provide two versions of global temperature anomalies, which differ in the treatment of the locations with sea ice. The temperature anomalies in the presence of sea ice are extrapolated from land-surface air temperature anomalies and sea-surface water temperature
anomalies, respectively. We adopted the air temperature version here. HadCRUT4 (version 4.6.0.0) near surface temperature is a blend of the CRUTEM4 LST dataset and the HadSST3 SST dataset (MORICE et al. 2012). Cowtan and Way (2014) produce a corrected version of HadCRUT4 (HadCRUT4krig, commonly referred as “Cowtan and Way” or “CW”) with complete coverage by filling the unobserved regions in HadCRUT4 using kriging reconstruction to reduce the influence of coverage bias of HadCRUT4. They give two versions of GMST anomaly time series. The later version (2.0) was used here. GISS Surface Temperature Analysis (GISTEMP v3, commonly referred as “GISS”) consists of land data from GHCN v3, SST data from ERSST5, and Antarctic station data from SCAR (Hansen et al. 2010, LenSSen et al. 2019). NOAA’s Merged Land-Ocean Surface Temperature analysis (MLOST version 3.5, commonly referred as “NOAA NCDC”, “NOAA”, and “NCDC”) unites GHCN v3 and ERSST4 (Smith et al. 2008, Vose et al. 2012). JMA’s land part is composed by GHCN before 2000 and CLIMAT messages archived at JMA after 2001, and its oceanic part is COBE-SST (Ishihara 2006).

2 The ensemble empirical mode decomposition

In order to eliminate the interference of natural variabilities and extract the time-varying anthropogenic trend underlying the multi-scale global temperature records, we employed the ensemble empirical mode decomposition (EEMD), which was developed based on the empirical mode decomposition (EMD). As the climate system is a complex and nonlinear coupling system, the variables involving are generally nonlinear and nonstationary, which do not meet the linearity and stationary assumptions required by most traditional time-frequency analysis tools. Therefore, it is difficult to get physically meaningful results by using traditional methods. To reveal the signal in nonlinear and nonstationary systems, Huang et al. (1998) proposed EMD method. Compared with traditional data analysis methods, EMD is based on the local characteristics of the data itself, and needs neither linear and stationary assumptions on data nor any external functions and parameters. Thus, the EMD is an adaptive and temporal local analysis tool which could be directly used for nonlinear and nonstationary data and help
obtain physically meaningful components.

The EMD decomposes any complicated data, \( x(t) \), into a few intrinsic mode functions (IMFs), \( c_i \), and a residual trend, \( r_n \).

\[
x(t) = \sum_{i=1}^{n} c_i + r_n
\]  

The IMFs, \( c_i \), are a set of narrow-spectrum fluctuations with different characteristic timescales. The remainder, \( r_n \), is a monotone function or a function with only one extreme value, from which no more periodic oscillatory can be extracted. It perfectly matches the definition of the intrinsic trend covering the whole data domain proposed by Wu et al. (2007).

In the EMD, the decomposition is implemented by a sifting process which is the central to produce the IMFs. The detailed description of the sifting procedures can be found in Huang et al. (1998). In sifting process, a stoppage criterion is needed for stopping further sifting. Here we adopted a fixed-number sifting stoppage criterion (Wu and Huang 2009, 2010), which is different with the original Cauchy type criterion used by Huang et al. (1998).

The EMD has been a reliable and effective scale-decomposition tool widely applied in many scientific and engineering studies. However, mode mixing, occurring because of signal intermittency, sometimes results in unstable decompositions and difficulty in physical interpretation of the IMFs. To overcome this problem, Wu and Huang (2009) developed the EEMD. They added different white noise time series to the target data to form an ensemble of noise-added signals, then decomposed each noise-added time series by EEMD and treated the ensemble mean of obtained IMFs as the true IMFs. The addition of white noises helps the ensemble exhaust all possible solutions and makes the signals with similar scales settle in one IMF. The true IMFs rise before our eyes when the added white noises cancel each other out after taking an average over enough trials. It is a truly noise-assisted data analysis method and greatly improves the original EMD. Benefiting from its unique properties, the EEMD is outstandingly suitable for analyzing data from the complex and coupled climate system, and
particularly powerful in extracting low-frequency oscillations and the intrinsic secular trend.

Fig. S3 shows the EEMD decompositions of the GMSTs derived from six observational MST datasets and the mean GMST time series. Each GMST time series was decomposed into five IMFs and an intrinsic trend. Owing to robust adaptivity and locality of the EEMD method, the components from six datasets are in good accordance, especially in the recent decades, though the datasets with different length were produced by different procedures. And the components from the mean GMST can fully represent those from six datasets. For simplicity, the mean of the six GMSTs was used as a representative dataset in sections 3a and 3d.

3 Statistical significance and robustness of intrinsic trend

Here we focused on the statistical significance test of the EEMD trend, as the variabilities reflected by IMFs were treated as noises to the secular warming trend. The significance of IMFs can be tested by two distinct approaches proposed by Wu and Huang (2004) and Wu et al. (2011), respectively. We assessed the statistical significance of the EEMD trend according to the approach developed by Ji et al. (2014) based on a Monte Carlo method. We generated 5000 samples of white noise with the same length as the GMST (168 years), and further generated 5000 red noise series having the same lag-1 auto-correlation as the GMST. Then we applied EEMD to each noise series to extract its overall trend. The metric for testing significance was defined as the trends increment/decrement at any temporal from its starting value (at 1850). Thus, we obtained numerically the distribution of the metric. Then we can tell whether the secular trend of the data is statistically significant by comparing the metric of the observed GMST with that of the noises. The EEMD trends of GMST and noises and their distributions at 2017 were plotted in Fig. S4. The trend increment of the GMST at 2017 is above two-SD spread of white and red noises. As the PDFs of trend increments of white and red noises at any given temporal location are approximately normally distributed, we could calculate the 95% confidence intervals based on the normal distribution. At 2017, the EEMD trend increment of the GMST is beyond the 95% confidence intervals of the distributions of the white and red
noises, the EEMD trend of observed GMST is thus considered statistically significant with 95% confidence level.

The EEMD-determined intrinsic trend is more robust than the least-squares linear trend. Fig. S5 displays the sensitivity of the EEMD trends and the least-squares fits of the GMST time series to the dataset choice and end points of data. At 2017, the spread of the increments of linear trends derived from six datasets is 0.26 °C, almost twice that of the EEMD trends of 0.14 °C. With respect to the sensitivity to the end date of the data, the contrast between the two spreads is even starker. The spread of the increments of the linear trends (0.28 °C) is more than four times that of the EEMD trends (0.06 °C). This means the linear trend can easily change when alternate dataset is used or new data added. In contrast to the changeable linear trend, the EEMD-determined intrinsic trend is quite robust, as it is insensitive to the dataset choices and hardly changes with the addition of new data.

4 Statistical significances of linear trend

Here, the linear warming rate was obtained by calculating the least-squares linear trend of the GMST time series. The statistical significance of the individual trends and the trend differences were tested according to Santer et al. (2000).

Consider a time series, $x(t)$, where $t$ varies from 1 to $n_t$. The linear trend, $b$, can be estimated by least-squares regression:

$$\hat{x}(t) = a + bt$$

where $\hat{x}(t)$ is the least-squares fit to $x(t)$.

$$e(t) = x(t) - \hat{x}(t)$$

$e(t)$ is the regression residuals which are assumed statistically independent. The variance of the regression residuals is defined as

$$s_e^2 = \frac{1}{n_e - 2} \sum_{t=1}^{n_t} e(t)^2$$

$n_e$ is the effective sample size calculated using lag-1 autocorrelation coefficient of $e(t)$, $r_1$.
The standard error of linear trend is

\[ s_b = \frac{S_e}{\left[ \sum_{t=1}^{n_t} (t - \bar{t})^2 \right]^{1/2}} \]  \hspace{1cm} (5)

Where \( \bar{t} \) is the mean of time index \( t \).

The ratio between the trend estimate and its standard error, \( t_b \), is distributed as Student’s \( t \).

\[ t_b = \frac{b}{s_b} \]  \hspace{1cm} (6)

Whether the trend is significant is determined by comparing \( t_b \) with a critical value, \( t_{crit} \), for a stipulated significance level, \( \alpha \), and \( n_e - 2 \) degrees of freedom. The null hypothesis of no trend is rejected when \( t_b > t_{crit} \). By using this method, the statistical significance of the individual trends in Figs. 2, 3 and 6 were tested at \( \alpha = 0.05 \) (i.e., 95% confidence level). Note the EEMD-determined intrinsic trend is monotonic with \( r_1 \approx 1 \), which does not meet the assumption of statistically independent regression residuals. So, this method shouldn’t apply to the intrinsic trend. The significances of linear trends of the intrinsic trend in Fig. 6b were assessed by Mann-Kendall test (Mann 1945, Kendall 1955) instead.

Based on the significance test of individual trends, Santer et al. (2000) provided two approaches to assess the statistical significance of the trend difference. In the first approach, the significance of trend difference is tested by examining whether there is overlap between the confidence intervals of two trends. This method tests whether the two trends are drawn from the same population. The trend confidence interval is defined as

\[ CI = b \pm t_{inv} \cdot s_b \]  \hspace{1cm} (7)

\( t_{inv} \) is calculated by inverting Student’s \( t \) distribution for given degrees of freedom (\( n_e \)) and significance level (\( \alpha = 0.05 \)). Using this method, we calculated the 95% confidence intervals of the trends during 1998-2012 and 1975-1997 to examine whether the trend during hiatus period differs significantly from that during reference period (Fig. S6). The second method examines whether there is a significant trend in the difference time series, \( d(t) = x_1(t) - \)
This method tests whether the differences in two time series have a significant effect on the trends. Using this method, we assessed whether datasets choice has a significant impact on the linear trend during 1998-2012 by examining whether there is a significant trend in the difference time series between various datasets (Fig. 3). Also, we used this method to test the trend differences between observation and simulation (Fig. 5).

5 Selection of hiatus period and reference period

After discussing the influence of the period choice in section 3a, we attempted to check the impact of the dataset choice in section 3b. Thus, to remove the potential influence from period choice and to exclusively concentrate on the trend differences between various datasets, a fixed hiatus period and a fixed reference period are needed. Also, a similar method was adopted in section 3c to check the influence from hiatus definition. In previous studies, 1998 and 2012 is the most popular start and end years, and the length of 1998-2012 represents the typical scale of hiatus period (Fig.1 and Table 1 in the manuscript). Especially, IPCC AR5 used 1998-2012 to refer to the global warming hiatus period. To keep consistent with the previous studies, we selected 1998-2012 as a fixed hiatus period in section 3b. Since the period 1998-2012 is particularly between two super El Niño events of 1997-1998 and 2015-2016, the linear trend over this decadal-scale period is greatly changeable. To reduce the potential impact of the reference period on the magnitude of the warming slowdown, the linear trend of the reference period should be relatively stable. So, the reference period shouldn’t be decadal scale. But the reference period shouldn't be too long either. Since there is a “big hiatus” from the mid-1940s to the mid-1970s, if the reference period includes the big hiatus, the warming rate would be unusually low and impede the identification of the recent slowdown. Fyfe et al. (2016) pointed out that the long-term reference period such as 1950-1999 is problematic with the big hiatus artificially included, and the previous warming surge is a more physically interpretable baseline. All things considered, in section b the reference period was temporarily fixed as 1975-1997, which represents the observed average state of the previous warming surge.
Table S1 Summary of global surface temperature datasets.

<table>
<thead>
<tr>
<th>Collection</th>
<th>Dataset</th>
<th>Temporal Coverage</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>MST</td>
<td>BEST</td>
<td>1850-2017</td>
<td><a href="http://berkeleyearth.org/data/">http://berkeleyearth.org/data/</a></td>
</tr>
<tr>
<td>Main</td>
<td>GISTEMP</td>
<td>1880-2017</td>
<td><a href="https://data.giss.nasa.gov/gistemp/">https://data.giss.nasa.gov/gistemp/</a></td>
</tr>
<tr>
<td></td>
<td>MST</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reanalysis</td>
<td>ERA-Interim</td>
<td>1979-2017</td>
<td><a href="https://www.ecmwf.int/en/forecasts/datasets/archive-datasets/reanalysis-datasets/era-interim">https://www.ecmwf.int/en/forecasts/datasets/archive-datasets/reanalysis-datasets/era-interim</a></td>
</tr>
<tr>
<td></td>
<td>NOAA 20CR</td>
<td>1851-2014</td>
<td><a href="https://www.esrl.noaa.gov/psd/gridded/data/20thC_ReanV2c.monolevel.m.html">https://www.esrl.noaa.gov/psd/gridded/data/20thC_ReanV2c.monolevel.m.html</a></td>
</tr>
<tr>
<td>Reanalysis</td>
<td>EAR-Interim</td>
<td>1979-2017</td>
<td><a href="https://www.ecmwf.int/en/forecasts/datasets/archive-datasets/era-interim">https://www.ecmwf.int/en/forecasts/datasets/archive-datasets/era-interim</a></td>
</tr>
</tbody>
</table>
Figure S1 GMST annual anomalies.

The GMST anomalies (°C, relative to 1981-2010) time series during 1850-2017 derived from six observations (a) and six reanalyses (b). The black line and gray spread represent the mean time series of six datasets and twice standard deviations (SDs, gray dotted line of right coordinate) apart from the mean time series. The red and blue time intervals highlight the warming surge period (1975-1997) and hiatus period (1998-2012), respectively.
Figure S2 Global mean SST annual anomalies.

Same as Figure S1, but for SST anomalies (°C, relative to 1981-2010) derived from eight observations (a), six satellites (b), and six reanalyses (c).
Figure S3 The EEMD components of six GMST time series derived from observational MST datasets and their mean time series.
Figure S4 EEMD trends and the distributions of the trend increments at 2017.

The upper panel plots the EEMD trends increments from their starting values (i.e., values at 1850). The EEMD trends are derived from mean GMST of six datasets (bold blue line), 5000 white noise series (thin gray line), and 5000 red noise series (thin pink line). Two bold gray (pink) lines are two SDs depart from the mean value values (i.e., zero). The two lower panels show the distributions of the white and red noise trends increments at 2017, respectively. The dotted lines represent the 95% confidence intervals. The blue lines indicate the observed GMST trend increment at 2017.
Figure S5 Sensitivity of the EEMD-determined intrinsic trends and the least-squares linear trends to datasets (a and b) and the end dates (c and d) of GMST data.

The colorful lines are the EEMD trends and linear trends calculated based on the GMSTs derived from different datasets (a and b), or with different end dates ranging from 1980 to 2017 (c and d) which simulates situations when new data are added. The vertical distances of the gray shaded boxes represent the spread of the trend increments at the last common year.
Figure S6 95% confidence intervals for linear trends of observed GMSTs and SSTs during 1975-1997 and 1998-2012.

At the right edge, the “mean” dataset is the average of the left six observations, and is treated as an individual dataset here. The trend of the mean dataset shown here is equal to the mean of six trends derived from different datasets shown in Figs. 3a and 3b. However, the error bars indicate the 95% confidence intervals for linear trend, which are distinct from the error bars in Figs. 3a and 3b presenting the dataset’s uncertainty.
Figure S7 The warming rates of real hiatus periods.

Same as Figure 4a, near-zero or negative warming rate according to definition 1 based on six MST datasets.
Figure S8 The warming rates of warming slowdown periods.

Same as Figure 4b, warming rates slower than that warming rate of previous warming surge during 1975-1997 according to definition 2a based on six MST datasets.
Figure S9 The warming rates of slower-than-average warming periods.

Same as Figure 4c, warming rates slower than long-term warming rate during 1950-1999 according to definition 2b based on six MST datasets.
**Figure S10 The warming rates of slower-than-average warming periods.**

Same as Figure 4d, warming rates slower than long-term warming rate during 1900-1999 according to definition 2b based on six MST datasets.
References


