Authors' Response to: Review: Dating Hiatuses: A statistical model of the recent slowdown in global warming – and the next one (esd-2018-81)

This paper uses econometric techniques to investigate the causes for a hiatus in warming and forecast whether another hiatus will occur in the future. The main results indicate that three factors drove the hiatus, the unusually warm year of 1998, the ENSO itself, and increases in tropospheric aerosols. With regard to the future, the authors conclude that the rate at which forcings rise will influence whether there is a hiatus in the future. These results (and to some degree the econometric results that are used to generate them) will interest the readers of *Earth System Dynamics.* As such, the manuscript should be considered for publication. The current form of the manuscript is close to being ready for publication. As described below, I have two substantive concerns and some minor concerns about the presentation/interpretation of results.

Thank you for your careful reading of our work and constructive comments, which have improved the manuscript! Please see our detailed responses below.

Substantive Comments

1. My first substantive concern focuses on the Oceanic Multidecadal Oscillation (OMO) and the empirical methods used to identify its effects.
   a. The OMO phenomenon needs to be described in more detail. Specifically, the authors need to define the OMO explicitly, briefly review the physical mechanisms thought to drive it, and how the OMO affects surface temperature.
   - **Added** a paragraph to the introduction: "Keenlyside et al. (2008) note an acceptance in the literature that the AMO and multidecadal periodic global temperature fluctuations more generally are related to the AMOC. Drijfhout et al. (2014) posit a physical link through changes in heat uptake across multiple oceans. Specifically, a weakening/strengthening of AMOC, meaning less/more convection and turbulent heat loss, leads to increased/decreased net heat uptake and therefore lower/higher surface temperatures. The change in net heat uptake occurs across multiple ocean basins, so we label the resulting multidecadal temperature cycle the Oceanic Multidecadal Oscillation (OMO). In other words, the OMO is the global sea surface temperature fluctuation resulting from changes in the thermohaline circulation. The OMO contrasts with the AMO in scope -- the latter is defined for the North Atlantic -- but the two may be highly correlated with similar periodicities."
   b. The authors also need to investigate the degree to which their results are sensitive to the methods used to estimate its effect on temperature.
   - **Added a paragraph and table** with additional investigations to the SOM: "The estimate of the aggregate effect of the OMO (net of forcings) reported in the paper for the optimal order, $(p_1,q_1)=(2,0)$, is 0.11°C (0.08,0.14)°C over 1998-2013, or a 41.7% (30.0,53.5)% exacerbation of the puzzle of the missing heat. As a robustness check on the effect of selection of p_1 and q_1, we varied each ±1 from the optimal order: (1,0), (3,0), (2,1), yielding estimates of the aggregate effects of 0.122°C (47.6% exacerbation), 0.092°C (35.9% exacerbation), and 0.029°C (11.3% exacerbation). The estimates with (1,0) or (2,0) seem the most plausible, because the additional terms of models with (3,0) and (2,1) generate estimates with very large magnitudes and opposing signs, a classic sign of near multicollinearity."

c. On page 8, line 1, the authors state that the OMO is estimated using a sine function. The authors should explain why a sine function is used.

- **Added several paragraphs** to the empirical results section: "The SOM (Supplementary Online Material) contains a detailed description of the methodology used to estimate the distribution of the OMO, the probability density function of which we denote by f(r) with support $[r^{-},r^{+}]$ (see the bottom panel of Figure S.3). Our methodology omits both long-run temporal temperature trends to avoid cointegration with h_{t} and idiosyncratic noise to avoid over-fitting very short-run fluctuations in GMT using sea-surface temperatures.

  Our method is intuitively similar to band-pass filters used to identify cycles in economic and other oscillating time series, but it is not executed in the frequency domain. Consistent with an energy balance model, filtering long-run stochastic trends (low frequency on the spectrum) is accomplished by regressing out the anthropogenic signal (see Figure S.1 of the SOM). In contrast, linear detrending (Enfield *et al.*, 2001) unnecessarily assumes a constant growth rate of the anthropogenic signal. More modern approaches, such as that of Trenberth and Shea (2006) and Lenton *et al.* (2017) also filter the anthropogenic signal, but they do so indirectly as a difference between two temperature series both subjected to the same stochastic trend.

  Detrending results in a noisy oscillation. High-frequency filtering could be accomplished using spectral methods if the time series were long enough or if the noise were "quiet" enough. However, it is difficult to identify a cycle with a 65-80 year period from a time series of 167 years in the presence of high-frequency cycles with substantial amplitudes, such as ENSO and the solar cycle. By focusing on a single periodic function, we narrow the desired frequency band enough to identify the cycle and to estimate it very transparently in the time domain. Specifically, we fit the stochastically detrended temperatures to a single periodic function (see Table S.1 and Figure S.4 of the SOM)."

d. Figure 1 suggests that the sine function is used to fit a time series for stochastically detrended GMT. But the authors do not explicitly define the data used to fit the sine function nor do the authors describe how they stochastically detrend global mean temperature. Because many readers are not econometricians, the authors need to show the equation(s) that they use to stochastically detrend global mean temperature.

- **Amended.** The description of our method for constructing the OMO is mostly contained in the SOM for the sake of brevity. We are concerned that too many technical details may be off-putting to some readers, but we wholeheartedly agree with the referee that the SOM should contain these details. We have now revised this part of the SOM, and the methodology should now be much clearer. If the explanation is still unclear, we can try again. (Or, if the editor and referee would prefer, we can bring these out of the SOM and into the paper at the expense of lengthening it.)

e. Finally, the authors need to investigate the degree to which this approach affects their results. My cursory review of the literature indicates that one set of authors calculate an index for the Atlantic Multidecadal Oscillation index from linearly detrended North Atlantic sea surface temperature anomalies while others identify the OMO signal using empirical orthogonal functions. How would the results reported in this manuscript change if they used one of these methods to estimate the effects of the OMO, instead of using the stochastically detrended global mean temperature?

- **Added R^2's for nonlinear regressions and relevant discussion** to the SOM: "The table also shows R^2's from the three regressions. These statistics should be interpreted with caution, both because they are R^2's from nonlinear regressions and because the regressions have the same functional form of time but different regressands. Nevertheless, we interpret them as evidence that the periodic function is a better approximation to the oscillation estimated using the proposed detrending method."
- **Added discussion** to the manuscript: "We note that homogeneous linear detrending results in a period of 72 years, while Trenberth and Shea's (2006) stochastically detrended AMO results in a period of 78 years. Our method, which has the advantages of relating the stochastic trend to forcings by way of a physical model and generates oscillations that are statistically more regular (see the SOM for details), generates an uncertainty interval that plausibly accounts for uncertainty in the stochastic trend."
- As we reported in the SOM, we tried alternative methods of detrending from the literature, including linear detrending and subtracting out alternative measurements of the stochastic temperature trend along the lines of Trenberth and Shea (2006). We hope that this is now clearer.

2. My second substantive concern focuses on the time series used to represent the cooling effects of tropospheric sulfates. In Figure 1, the authors identify the ability of various forcings to account for the missing heat, as represented by degrees Celsius anomaly from the base period. This is a very straightforward and understandable way to approach the problem. My issue here concerns the forcings used to simulate the model, which are the time series used to simulate the GISS model. In general, the forcings used to simulate the GISS model are highly stylized. They are largely linear with little variation in growth rates over time. This is especially true for reflective tropospheric aerosols. This linearity is different from the time series for anthropogenic sulfur emissions that are assembled by Steven J. Smith (and others) at the Pacific Northwest National Laboratory. These data are updated such that it is compatible with the sample period used by the authors (https://www.geosci-model-dev.net/11/369/2018/gmd-11-369-2018.pdf). I suggest that the authors investigate the degree to which their results are sensitive to the forcing used by redoing their analysis with the time series from the paper by Hoesly *et al*., (2017).

- **New analysis and section added to the SOM.** We acknowledge this important point and we really wrestled with how to address this comment. Another important difference between the Hansen data and the Hoesly data is that the former are expressed in W/m^2 and based on concentrations, while the latter are emissions data. Additional models – of the carbon cycle, e.g. – are required to convert emissions into forcings. Some authors, such as Kaufmann *et al.* (2006), have done this, but we are reluctant to add the additional models that would be required for most forcings, because (a) such an approach could just as easily be criticized, (b) estimation of the model contributes another layer of uncertainty to the data, and (c) most of the forcings should be measured without very much uncertainty, per Myhre *et al.* (2013), and we are already taking this uncertainty into account.

  Of course, tropospheric sulfates are exceptional, both because of the large uncertainty in the data and in their short residence times in the atmosphere, and we interpret the referee's concern to be mainly about measuring these. Because of the short residence time, it seems reasonable to equate concentrations with emissions within the time scale of a one-year increment. Please see the section we added to the SOM for additional details.

  The data sources *do* make a difference, and we now report this difference in the paper. How can we interpret this discrepancy at the end of the day? Clearly, Myhre *et al.* (2013) are right that there is much uncertainty associated with tropospheric aerosols! Our results included a very wide uncertainty interval around this effect, and the additional analysis supports perhaps an even wider uncertainty interval.
- We also **made minor revisions** to our input data and measures of uncertainty following a close comparison of the Hoesly and Hansen data. The numbers throughout the paper have changed a bit as a result.

Minor Comments

1. Page 3: "explain past multidecadal cooling or hiatus periods, such as the decades following the temperature spikes in about 1877 and 1943." On page 271 Kaufmann et al (2006) (cited by the authors) write "The radiative forcing of anthropogenic sulfur emissions increases at about the same rate as greenhouse gases between 1944 and 1976. As a result, there is relatively little net increase/decrease in total radiative forcing and therefore, global surface temperature. The timing of these temperature effects is consistent with results obtained from model simulations (Andronova and Schlesinger, 2000; Tett et al., 1999)." The authors should reconcile their statement about the hiatus with the explanation based on a slight decrease in total forcing.

- **Added** a paragraph: "Kaufmann *et al.* (2006) note about the period 1944-76 that the decrease in net radiative forcing resulting from an increase in anthropogenic sulfur emissions approximately offset the increase in net forcing from an increase in WMGHGs. Aside from the uncertainty in measuring forcing from sulfur emissions, this offset nicely shows the relationship between the OMO and temperatures: while the OMO declined by an average of 0.016°C per year, temperature also declined by an average of 0.012°C per year. The decline in temperature over this period may be explained both by a decline in the OMO and an increase in sulfur emissions."
- We appreciate that the referee directed our attention to this. We believe that Kaufmann *et al.* (2006) may have inadvertently overstated their case. Interannual variations in temperature are large – precisely the reason we filter out high-frequency noise when calculating the OMO. A small but persistent movement in temperature that correlates with the OMO is obscured. Our point does not undermine the point raised by those author about the offset caused by the different types of anthropogenic emissions.

2. Page 5 Lines 1-3 describe the method used to covert a change in W/m2 to temperature ($0.536 + 0.561 \times 0.431 \approx 0.777$,C). The authors should explain where 0.430 'comes from.'

- **Added:** "where 0.426 is the CCR estimate of α_1 in the model in (2) with $\gamma,\delta=0$ (see Table S.2 of the SOM)" [The 0.426 was 0.430, but these numbers have changed slightly.]

3. Page 11 "Our main findings for this period suggest that the three main factors driving the hiatus were (a) the unusually warm year of 1998, even conditional on the ENSO, (b) the ENSO itself, and (c) the increase in tropospheric aerosols during that period, though the latter is measured with a high degree of uncertainty" This is a very important component of the authors results, but these results are not really clear in the abstract. The authors should edit this abstract to make these results clearer. Also the abstract should highlight the result that the occurrence of a future hiatus depends on in part on the rate at which forcing grows.

- **Revised.** Thank you for this suggestion! We tried to execute this revision without lengthening the abstract too much.

Literature cited

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J.-I., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emission Data System (CEDS), Geosci. Model Dev. Discuss., doi:10.5194/gmd-2017-43, in review, 2017.

Dating Hiatuses: A Statistical Model of the Recent Slowdown in Global Warming – and the Next One

J. Isaac Miller[1] and Kyungsik Nam[1]

[1]University of Missouri

Correspondence: J. Isaac Miller (millerjisaac@missouri.edu)

Abstract. Much has been written about the so-called hiatus ~~or pause~~ or pause in global warming, also known as the stasis period, ~~the~~ the start of which is typically dated to 1998. HadCRUT4 global mean temperatures slightly decreased over 1998-2013, though a simple statistical model predicts that they should have grown by ~~0.016°~~0.016°C/yr, in proportion to the increases in concentrations of well-mixed greenhouse gases and ozone. We employ a statistical approach to assess the contributions of model forcings and natural variability to the hiatus. Our point estimates suggest that none of the model forcings explain more than a ~~fifth~~ third of the missing heat~~and that the~~, accounting for the upper bound of the confidence interval on the effect of tropospheric aerosols, which is the most prominent yet most uncertainly measured of the model forcings that could explain the missing heat. The El Niño Southern Oscillation (ENSO) explains ~~no more than a quarter~~ up to about a third of the missing heat~~. Two~~, and two thirds and possibly up to ~~84%~~ 81% is explained by the unusually high temperature of 1998. Looking forward, the simple model also fails to explain the large increases since then (~~0.087°~~0.087°C/year over 2013-2016). This period coincides with another El Niño, but the ENSO fails to satisfactorily account for the increase. We propose instead a semiparametric cointegrating statistical model that augments an energy balance model with a novel multibasin measure of the oceans' multidecadal temperatures cycles. The model partially explains the recent slowdown and explains ~~nearly~~ all of the subsequent warming. The natural cycle suggests the possibility – depending in part on the rate of increase of WMGHG concentrations – of a much longer hiatus over roughly 2023-2061, with potentially important implications for policy evaluation.

1 Introduction

There is a well-established physical and statistical link between temperatures and anthropogenic and natural climate forcings. A simple linear cointegrating regression of the HadCRUT4 global mean temperature anomaly (GMT) onto the radiative forcings given by Hansen *et al.* (2017) ~~correlates~~ explains 88% of the variation in mean temperature ~~with~~ using variations in these forcings. Constraining all but volcanic forcings to have a common coefficient in the regression explains 84%.

Over the period of 1998-2013, the second regression, estimated using a canonical cointegrating regression, predicts an increase of ~~0.241°~~0.239°C or 0.016°C/yr on average, in proportion to the increase in well-mixed greenhouse gases (WMGHGs) and ozone over this period. Instead, observed GMT slightly *decreased* by 0.024°C or 0.002°C/yr on average, earning this period the nicknames of the "stasis period," or "hiatus" or "pause" in global warming. The difference, measured in this way as ~~0.265°~~0.263°C or 0.018°C/yr, is the so-called "missing heat" of the hiatus, which is quite substantial in the context of the

aggregate temperature increase since the pre-industrial era of 0.85°C (IPCC, 2013). In contrast, ~~since 2013 (through 2016), temperatures have~~ over 2013-2016, temperatures increased by 0.087°C/yr, much *faster* than this simple statistical model predicts.

Our notion of hiatus is roughly consistent with that of Meehl *et al.* (2011), Kosaka and Xie (2013), and Drijfhout *et al.* (2014), who reference the apparent hiatus in global warming with respect to heat flux from greenhouse gases or model forcings more generally. Instead, some authors refer to the hiatus with respect to temperature changes or a trend over an earlier period (Schmidt *et al.*, 2014; Karl *et al.*, 2015; Yao *et al.*, 2016; and Medhaug *et al.*, 2017), while some authors refer to the hiatus without any explicit baseline. Linking missing heat to contemporaneous model forcings is physically appealing, and our empirical evidence suggests that our measure of missing heat comes from a cointegrating regression,[1] so the approach is statistically appealing, too. A slowdown or hiatus in global warming as we have defined it does not require a similar slowdown in forcings. On the contrary, such a hiatus is defined in spite of continuing increases in WMGHG concentrations.

What caused this hiatus? Various studies attribute it to one or more of (a) natural variability of the ocean cycles, particularly the Atlantic Multidecadal Oscillation (AMO), the Pacific Decadal Oscillation, and the El Niño Southern Oscillation (ENSO) (Kosaka and Xie, 2013; Steinman *et al.*, 2015; Yao *et al.*, 2016); (b) cooling from stratospheric aerosols released by volcanic activity (Vernier *et al.*, 2011; Neely *et al.*, 2013); (c) variability of the solar cycle (Huber and Knutti, 2014); (d) a change in the oceans' heat uptake and a weakening of the thermohaline circulation, particularly the Atlantic Meridional Overturning Circulation (AMOC) (Meehl *et al.*, 2011; Drijfhout *et al.*, 2014; Chen and Tung, 2014, 2016); (e) increased anthropogenic emissions of sulfur dioxide from bringing online a large number of coal-burning power plants in China (Kaufmann *et al.*, 2011); and (f) coverage bias or poor data more generally (Cowtan and Way, 2014; Karl *et al.*, 2015). Schmidt *et al.* (2014), Pretis *et al.* (2015), and Medhaug *et al.* (2017) emphasize the need to account for multiple explanations for the hiatus.

~~We propose a new method to measure the oceans' aggregate multidecadal cycle, which we call the~~ Keenlyside *et al.* (2008) note an acceptance in the literature that the AMO and multidecadal periodic global temperature fluctuations more generally are related to the AMOC. Drijfhout *et al.* (2014) posit a physical link through changes in heat uptake across multiple oceans. Specifically, a weakening/strengthening of AMOC, meaning less/more convection and turbulent heat loss, leads to increased/decreased net heat uptake and therefore lower/higher surface temperatures. The change in net heat uptake occurs across multiple ocean basins, so we label the resulting multidecadal temperature cycle the Oceanic Multidecadal Oscillation (OMO)~~,~~. In other words, the OMO is the global sea surface temperature fluctuation resulting from changes in the thermohaline circulation. The OMO contrasts with the AMO in scope – the latter is defined for the North Atlantic – but the two may be highly correlated with similar periodicities.

We propose a new method to measure the OMO, recognizing the possibility of heterogeneous long-run effects of anthropogenic forcings on ocean basins and allowing for a multibasin contribution to global mean temperatures, in the spirit of

[1]The regression of temperature anomalies on volcanic forcings, the sum of the remaining forcings, and an intercept yields a covariance stationary residual series. Augmented Dickey-Fuller tests with lag lengths up to four reject the null of no cointegration. ~~The GPH estimator suggests~~ Estimates of the memory parameter of about 0.49 using a lag truncation parameter up to 10 suggest the possibility that the residual series is stationary with long memory, ~~with memory parameter estimated between 0.44 and 0.49,~~ which also supports (fractional) cointegration.

Drijfhout *et al.* (2014) and Wyatt and Curry (2014). The method allows an improvement over the linear detrending method of Enfield *et al.* (2001) or a single ocean approach such as the AMO signal estimated by Trenberth and Shea (2006). Not only do we estimate the mean OMO~~for the globe~~, but we also estimate a global distribution representing the contribution of the OMO to spatially disaggregated sea surface anomalies.

We utilize a semiparametric cointegration statistical approach (Park *et al.*, 2010), with widely used and publicly available data sets to estimate an energy balance model (EBM) similar to the well-known model of North (1975) and North and Cahalan (1981). The estimated OMO enters the model nonparametrically, as does the quasi-periodic southern oscillation index (SOI), a common proxy for the ENSO. However, information criteria select a linear specification for the latter.

We find that the solar cycle and multidecadal ocean cycle have warmed rather than cooled over the period 1998-2013, so they cannot account for the missing heat. Volcanoes, tropospheric aerosols & surface albedo, and ENSO account for about 1%, 19%, and ~~23%~~ 24% respectively of the missing heat of the hiatus. The upper bounds on the uncertainty intervals for the latter two are both about one third. An even simpler explanation – that the hiatus is defined by starting in an unusually warm year, even taking into account El Niño – explains about two thirds of the missing heat, a result that echoes previous authors (Medhaug *et al.*, 2017, e.g.). A model that takes into account all but the residual just mentioned explains about ~~43%~~42%.

Roberts *et al.* (2015) speculate that the hiatus could last through the end of the decade, and Chen and Tung (2014) and Knutson *et al.* (2016) make stronger statements about its continuation. If so, then the unusually warm years of 2015-2016 are outliers and global temperatures can be expected to stabilize or cool in the next decade. On the contrary, our proposed model explains nearly all of the more recent record warm years, overshooting the record high anomaly of 0.773°C in 2016 by ~~only 0.021°~~less than 0.001°C. This result provides conclusive statistical evidence that the hiatus is over. In other words, we date the end of the recent hiatus prior to 2015.

Can we expect a future hiatus or slowdown? If so, when? We find that the two most influential non-seasonal drivers of global aggregate temperature change are the long-run contribution of WMGHGs and the fairly predictable OMO with a period of 76 years, consistent with the 65-80 year period estimated for the AMO in the literature (Knight *et al.*, 2005; Trenberth and Shea, 2006; Keenlyside *et al.*, 2008; Gulev *et al.*, 2013; Wyatt and Curry, 2014). Although the OMO cannot explain the recent hiatus, it can explain past multidecadal cooling or hiatus periods, such as the decades following the temperature spikes in about 1877 and 1943.

Kaufmann *et al.* (2006a) note about the period 1944-76 that the decrease in net radiative forcing resulting from an increase in anthropogenic sulfur emissions approximately offset the increase in net forcing from an increase in WMGHGs. Aside from the uncertainty in measuring forcing from sulfur emissions, this offset nicely shows the relationship between the OMO and temperatures: while the OMO declined by an average of 0.016°C per year, temperature also declined by an average of 0.012°C per year. The decline in temperature over this period may be explained both by a decline in the OMO and an increase in sulfur emissions.

If we condition the model on future forcings with growth rates similar to RCP8.5, we can expect temperatures to increase with a possible slowdown but without any future hiatus. However, if we condition on future forcings with the same average annual rate as that of the past 76 years, similarly to RCP6.0, we expect a multidecadal hiatus over approximately 2023-2061.

Note that our finding of a warm period separating the previous slowdown from the next one is exactly consistent with the recent projection of warming from 2018-2022 by Sévellec and Drijfhout (2018) using a different model and method.

2 Empirical Results

We utilize forcings data over 1850-2016 from Hansen *et al.* (2017).[2] We create two radiative forcing series: the sum of forcings from well-mixed greenhouse gases (primarily CO_2, CH_4, N_2O, and CFCs), ozone, tropospheric aerosols & surface albedo, and solar irradiance, denoted by h_1, and that from volcanoes, denoted by h_2. Shindell (2014) suggests the possibility that forcings due to aerosols and ozone may have effects that are different from those of WMGHGs. By aggregating all non-volcanic forcings into h_1, we are instead following Estrada *et al.* (2013), Pretis (2015), *inter alia*. A Wald test shows no statistically significant difference (p value of ~~0.31~~0.61) between models with and without the restriction imposed.[3]

Some authors, such as Estrada *et al.* (2013), ignore volcanoes in statistical estimation of EBMs. Leaving out volcanoes is statistically justified by the apparent uncorrelatedness of this series with the other forcings. Relegating that series to the error term may affect statistical uncertainty, but should not bias the estimates of the effect of h_1. Because one of our goals is to assess the impact of volcanic activity on the slowdown, we include volcanoes. However, we allow for a separate coefficient on h_2, in order to accommodate the suggestion of Lindzen and Giannitsis (1998) of a smaller sensitivity parameter for physical models that include volcanic forcing.

We use HadCRUT4 and HadSST3 temperature anomaly data, measured relative to 1961-1990, from Morice *et al.* (2012) and Kennedy *et al.* (2011a,b) respectively.[4] In order to estimate the global distribution of the OMO, monthly HadSST3 data observed over $5°$ latitude by ~~5°~~ 5° longitude are pooled into years over 1850-2016 (167 years of data). The HadCRUT4 data set combines HadSST3 for sea and CRUTEM4 for land, so the temperatures from HadCRUT4 and HadSST3 are comparable. However, using only HadSST3 for the distribution ensures that grid boxes containing both land and ocean stations will contain only ocean measurements in the distribution.

The SOM (Supplementary Online Material) contains a detailed description of the methodology used to estimate the distribution of the OMO~~. Specifically, $f(r)$ is~~ , the probability density function of ~~the OMO over the globe's oceans estimated with a~~ which we denote by $f_t(r)$ for year t with support $[r^-, r^+]$ ~~. Estimation of the OMO~~ (see the bottom panel of Figure S.3). Our methodology omits both long-run temporal temperature trends to avoid cointegration with h_t and idiosyncratic noise to avoid over-fitting very short-run fluctuations in GMT using sea-surface temperatures. It is intuitively similar to band-pass filters used to identify cycles in economic and other oscillating time series, but it is not executed in the frequency domain.

[2] Annual data for 1850-2015 downloaded from www.columbia.edu/~mhs119/Burden on May 15, 2017. See the Supplementary Online Matrial (SOM) for details on the extrapolation to 2016.

[3] The test is executed as qF, where $q = 3$ is the number of restrictions tested and F is the F-test of these restrictions based on Cochrane-Orcutt transformed regressions to accommodate an AR(1) error consistent with the bootstrapping strategy discussed below. The $q = 3$ restrictions are that the effects on temperature of a W/m^2 change in well-mixed greenhouse gases, ozone, tropospheric aerosols & surface albedo, and solar irradiance are equal.

[4] Ensemble median of HadCRUT.4.5.0.0 (annual unsmoothed globally averaged) and HadSST.3.1.1.0 (monthly globally disaggregated) downloaded from www.metoffice.gov.uk/hadobs on July 18, 2017 and April 4, 2017 respectively.

Consistent with an energy balance model, filtering long-run stochastic trends (low frequency on the spectrum) is accomplished by regressing out the anthropogenic signal (see Figure S.1 of the SOM). In contrast, linear detrending (Enfield *et al.*, 2001) unnecessarily assumes a constant growth rate of the anthropogenic signal. More modern approaches, such as that of Trenberth and Shea (2006) and Lenton *et al.* (2017) also filter the anthropogenic signal, but they do so indirectly as a difference between temperature series both subjected to the same stochastic trend.

Detrending results in a noisy oscillation. High-frequency filtering could be accomplished using spectral methods if the time series were long enough or if the noise were "quiet" enough. However, it is difficult to identify a cycle with a 65-80 year period from a time series of 167 years in the presence of high-frequency cycles with substantial amplitudes, such as ENSO and the solar cycle. By focusing on a single periodic function, we narrow the desired frequency band enough to identify the cycle and to estimate it very transparently in the time domain. Specifically, we fit the stochastically detrended temperatures to a single periodic function (see Table S.1 and Figure S.4 of the SOM).

We base our statistical model on an EBM given by

$$T^a = h^{*\prime}\alpha + \int_{r^-}^{r^+} b(r)f(r)dr + c(S) + \eta, \tag{1}$$

where T^a is the global mean temperature anomaly (GMT), $h^* = (1, h_1, h_2)'$ is global forcing, S is the Southern Oscillation Index (SOI, Ropelewski and Jones, 1987) used to proxy ENSO quasi-periodic cycles,[5] $\alpha = (\alpha_0, \alpha_1, \alpha_2)'$ is a coefficient vector, and η is an error term.

A detailed derivation of the EBM from a more familiar EBM similar to those of North (1975), North and Cahalan (1981), North *et al.* (1981), *inter alia* is provided in the SOM. A primary intuition for the derivation is that the nonlinear functions $b(r)$ and $c(S)$ allow the oceans' heat uptake to vary over multidecadal and interannual oscillations.

In order to estimate the EBM in (1) nonparametrically in $b(r)$ and $c(S)$, we attach time subscripts and write

$$T_t^a = h_t^{*\prime}\alpha + x_t'\gamma + w_t'\delta + \eta_t, \tag{2}$$

where $x_t = (x_{1t}, ..., x_{m_T t})' = \int_{r^-}^{r^+} b_{1:m_T}(r) f_t(r)dr$ with $b_{1:m_T}(r) = (b_1(r), ..., b_{m_T}(r))'$ and $w_t = (w_{1t}, ..., w_{m_S t})' = c_{1:m_S}(S_t)$ with $c_{1:m_S}(S_t) = (c_1(S_t), ..., c_{m_S}(S_t))'$, finite-order series approximation to $\int_{r^-}^{r^+} b(r)f_t(r)dr$ and $c(S_t)$, with m_S- and m_T-vectors of coefficients given by γ and δ respectively. The error term η_t contains both (serially correlated) stochastic forcings, along the lines of North *et al.* (1981), and any approximation errors from the series ~~expansions~~approximations.

2.1 The 1998-2013 Episode

The missing heat of the recent hiatus is defined above by the difference between the actual GMT in 2013 and the temperature predicted from increases in WMGHG and Ozone (G+Z, hereafter) alone using the restricted model with $\gamma, \delta = 0$ and starting in 1998. The GMT in 1998 was 0.536°C. Fixing 1998 as the starting year and based on an increase of climate forcings from G+Z of 0.561 W/m^2 over 1998-2013, the model predicts a GMT of ~~$0.536 + 0.561 \times 0.430 \simeq 0.777$°~~$0.536 + 0.561 \times 0.426 \simeq 0.775$°C

[5] Southern Oscillation Index downloaded from www.esrl.noaa.gov/psd/gcos_wgsp/ on April 13, 2018.

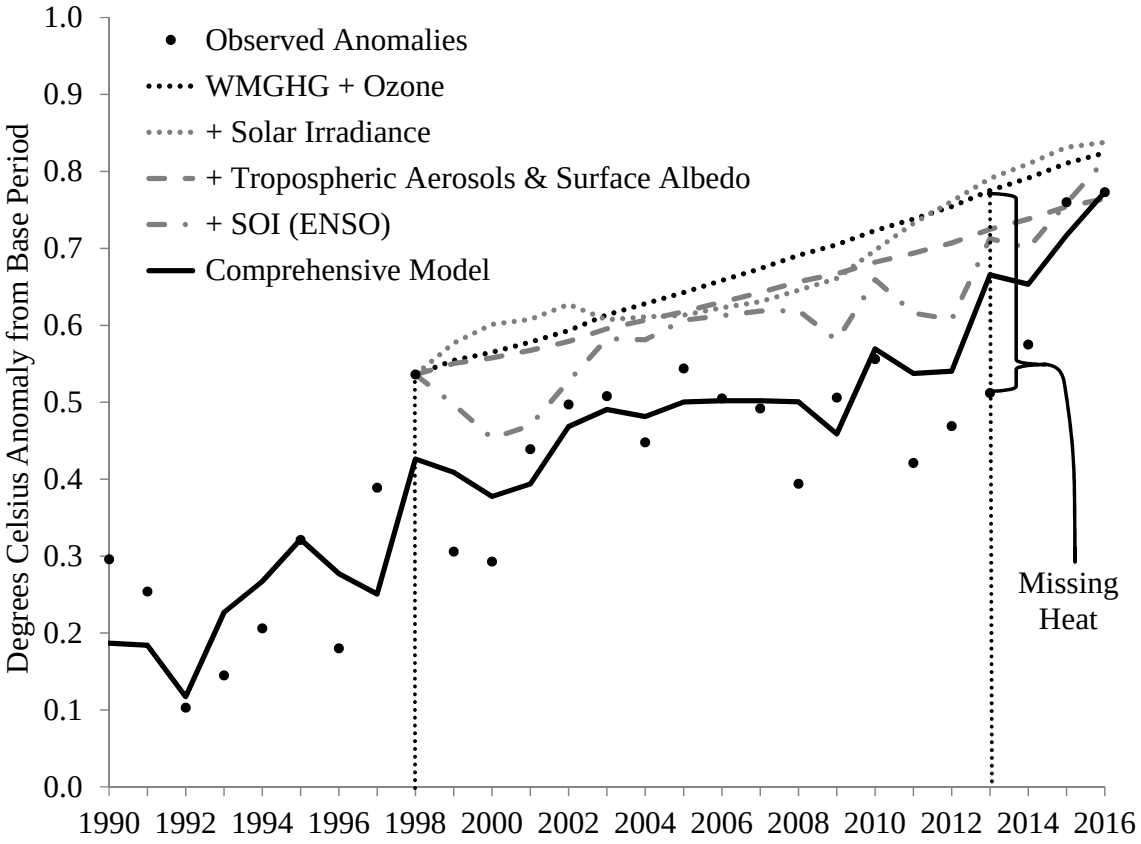

Figure 1. A Visual Anatomy of the 1998-2013 Episode. The hiatus is defined by the missing heat in 2013 relative to that predicted by increases in WMGHG and ozone forcings since 1998. Plots crossing the missing heat help to explain it (pictured: tropospheric aerosols & surface albedo, ENSO, and the comprehensive model), while those passing above the missing heat exacerbate it (pictured: solar irradiance).

in 2013, ~~with~~ where 0.426 is the CCR estimate of α_1 in the model in (2) with $\gamma, \delta = 0$ (see Table S.2 of the SOM), with a confidence interval of ~~$(0.750, 0.804)^\circ$~~ $(0.756, 0.795)^\circ$C.[6] In contrast, the observed GMT is 0.512°C in 2013, so that the difference, ~~$0.777 - 0.512 \simeq 0.265^\circ$C $(0.238, 0.292)^\circ$~~ $0.775 - 0.512 \simeq 0.263^\circ$C $(0.244, 0.283)^\circ$C, represents the missing heat. The 1998-2013 episode is illustrated by the missing heat in Figure 1.

One way to try to explain the missing heat is to "turn on" some of the other forcings in the model. To that end, we estimate the model in (2) with both $\gamma, \delta \neq 0$ (unrestricted) and $\gamma, \delta = 0$ (restricted). Least squares is expected to be consistent, but we use the canonical cointegrating regression approach of Park *et al.* (2010) ~~,~~ in order to estimate the coefficients asymptotically normally and the standard errors consistently for cointegrated temperatures and forcings. A number of previous studies have

[6] The intervals throughout the paper are given with 90% confidence, in keeping with those for the forcings given by the IPCC (Myhre *et al.*, 2013). The intervals reflect not only statistical uncertainty from the regression error but also uncertainty in the underlying data. Details of the construction of these intervals are given in the SOM.

provided physical and statistical evidence in favor of a cointegrating relationship: Kaufmann *et al.* (2006a, 2006b, 2010, 2013), Pretis (2015), *inter alia*. As explained in the SOM, this procedure also corrects for uncertainty in the forcings data.

Adding only volcanoes to G+Z decreases forcings by 0.036 W/m^2. Predicted GMT decreases by only 0.003°C ~~(0.001, 0.005)°~~ (0.001, 0.0 or about 1.1% ~~(0.3, 1.9)%~~ (0.5, 1.7)% of (the point estimate of) the missing heat. In the data, forcing from stratospheric aerosols is attributed to volcanic activity, while forcing from tropospheric aerosols is attributed to anthropogenic sulfur dioxide emissions. Vernier *et al.* (2011) refute previous studies that attribute an increase in stratospheric aerosols to emissions. While those authors do not discuss the effect of volcanic activity on the hiatus directly, their Figure 5 suggests that the stratospheric aerosol levels from Mt. Pinatubo subsided until about 1997, and the increase since then has been relatively small.

Similarly, adding only tropospheric aerosols & surface albedo to G+Z decreases forcings by 0.118 W/m^2. Predicted GMT decreases by ~~0.051°C (0.005, 0.097)°~~0.050°C (0.030, 0.070)°C, or 19.1% ~~(1.7, 36.5)%~~ (11.5, 26.8)% of the missing heat. An alternative measure of forcings from tropospheric aerosols based on the sulfur emissions data of Hoesly *et al.* (2018) predicts GMT *increasing* by 0.003°C (−0.014, 0.020)°C, or exacerbating the missing heat by 1.3% (−5.7, 8.3)%. The SOM contains more details of how the alternative data were employed and additional empirical results based on them.

That anthropogenic aerosol emissions appear to explain some of the missing heat using the Hansen *et al.* (2017) data is consistent with the findings of Storelvmo *et al.* (2016) ~~. In spite of the uncertainty in measuring forcings from anthropogenic emissions, our results suggest that these may play a larger role than natural volcanic emissions, which is more consistent with the findings of Kaufmann~~ and Kaufmann *et al.* (~~2011)than those of Vernier~~ 2011). Unfortunately, although the interval estimate using these tropospheric aerosol data explicitly account for measurement error, it does not cover the point estimate using the data of Hoesly *et al.* (~~2011). In any case, our results suggest that neither of these two explanations explain very much of the missing heat~~2018), and vice versa. In short, the effect is quite uncertain using either data set.

Solar irradiance *increases* forcings by 0.037 W/m^2, so that the predicted GMT *increases* by 0.016°C ~~(0.012, 0.019)°~~ (0.013, 0.018)°C, exacerbating the missing heat by 6.0% ~~(4.7, 7.3)%~~(5.1, 6.9)%. There is a decline from 2002-2006, but the net effect of solar over 1998-2013 is to increase temperature – not to decrease it. To the extent that solar contributed to the slowdown by decreasing temperatures, the results suggest that solar alone ~~cannot be~~ is not sufficient. This finding is not inconsistent with that of Schmidt *et al.* (2014), who examine solar in conjunction with other forcings as an explanation.

The preceding explanations are model forcings, and none of them satisfactorily account for the slowdown either alone or in concert. As previous authors have pointed out, natural variability may play a role, and we now turn to measures of two such types: the OMO and the ENSO.

In order to examine multidecadal oscillations from the OMO as a possible explanation, we let $\gamma \neq 0$, but keep $\delta = 0$. The regressor vector x_t is correlated with the other forcings, and we want to capture the partial effect of the OMO while retaining the total effect of G+Z. In order to do so, we employ the fitted residuals from regressing x_t onto the other forcings as a regressor in the model, rather than using x_t itself. The two approaches – using x_t or its fitted residuals – yield equivalent model fits, but using the fitted residuals fixes the coefficient vector α.

The oscillation exacerbates the missing heat by 0.110°C ~~(0.078, 0.143)°C or 41.6% (29.3, 53.9)%~~(0.079, 0.141)°C or 41.7% (30.0, 53.5)%. By itself, the fitted OMO worsens the puzzle in the sense that the predicted temperature in 2013 in-

creases to ~~$0.777 + 0.110 = 0.887$°~~$0.775 + 0.110 = 0.885$°C. The reason for the increase is that the OMO appears to be increasing rather than decreasing during this period. This result contrasts sharply with that of Yao *et al.* (2016), who attribute the hiatus to a much shorter oceanic cycle of 60 years.

The ENSO is quasiperiodic with a period of about 5-6 years. However, roughly every three El Niño episodes are so-called "super El Niños" with much higher amplitudes than the intervening episodes. In other words, the ENSO also oscillates at a decadal scale, roughly 15-18 years. The last two peaks of the longer oscillation were in 1997-98 and 2015-16, coinciding with the beginning and end of the 1998-2013 episode. Letting $\delta \neq 0$ but keeping $\gamma = 0$ yields an increase in the normalized and orthogonalized SOI of 1.162 and thus a decrease of 0.062°C ~~(0.011, 0.113)°~~$(0.029, 0.095)$°C, so that the ENSO explains ~~23.4% (4.3, 42.5) %~~ 23.7% $(11.1, 36.3)\%$ of the missing heat.

All of the explanations so far ignore to some extent that the starting year matters, as has been pointed out by previous authors (Medhaug *et al.*, 2017, e.g.). Not only was 1998 an El Niño year, it was an anomalously warm one. Suppose that the temperature anomaly in 1998 had been equal to that in 1997, 0.389°C. The same exercise of defining the hiatus using growth rates predicted by G+Z results in a 2013 temperature anomaly of ~~0.630°~~$0.389 + 0.561 \times 0.426 = 0.628$°C, a decrease of ~~$0.777 - 0.630 = 0.147$°~~$0.775 - 0.628 = 0.147$°C, which explains ~~55.5%~~ 55.9% of the missing heat. In other words, half of the puzzle is explained simply by the construction of the puzzle.

The counterfactual of setting the 1998 temperature to that of 1997 seems effective in explaining the slowdown, but it is extremely *ad hoc*. A similar result is obtained more formally by fitting the model forcings but no natural variation – i.e., by estimating the model in (2) with $\gamma, \delta = 0$. This model predicts the temperature in 1998 to be ~~0.400°~~0.395°C – close to that in 1997 – and increasing to ~~0.604°~~0.597°C by 2013. In other words, a simple model, including all forcings but without the OMO or the ENSO, explains ~~0.173°C (0.124, 0.223)°~~0.178°C $(0.155, 0.202)$°C, or ~~65.4% (46.8, 84.0) %~~ 67.7% $(58.7, 76.6)\%$ – more than half – of the missing heat. Looking at the most comprehensive model with $\delta, \gamma \neq 0$ gives a qualitatively similar result, explaining ~~42.6% (25.1, 60.2) %~~41.6% $(27.5, 55.8)\%$.

This finding suggests that the unusually warm year of 1998 – a residual in the model – accounts for most of the apparent slowdown between 1998-2013. It is consistent with the finding of Kosaka and Xie (2013), in the sense that the El Niño year is necessarily followed by La Niña cooling. However, with the results of Pretis *et al.* (2015) and those using the SOI above in mind, we certainly cannot attribute the slowdown to the ENSO uniquely.

Yet there is a new problem given by the ~~recent~~ high GMTs of 0.760°C in 2015 and 0.773°C in 2016. The restricted model undershoots these temperatures by more than 0.1°C. Are these ~~again~~ simply outliers, as 1998 was? A natural explanation is the ENSO, because 2015 and 2016 were El Niño years. The model with $\delta \neq 0$ and $\gamma = 0$ – i.e., with SOI but no ~~AMO~~ OMO – undershoots 2015 by ~~0.113°~~more than 0.1°C and 2016 by ~~0.057°~~just less than 0.1°C. In other words, accounting for ENSO does about as poorly as not accounting for ENSO in predicting the temperature in 2015, but improves the prediction for 2016.

Finally, consider the proposed comprehensive model with ~~$\delta, \gamma \neq 0$~~$\delta, \gamma \neq 0$. The model undershoots 2015 by only ~~0.044°~~0.042°C, while overshooting 2016 by ~~only 0.021°~~less than 0.001°C (see Figure 1). We interpret these numbers to mean that the recent high temperatures of 2015 and 2016 are attributable more to the smooth, multidecadal, and somewhat predictable OMO than to the higher-frequency quasi-periodic ENSO. As a result, we can say that 2015 and 2016 were not outliers, and that increases

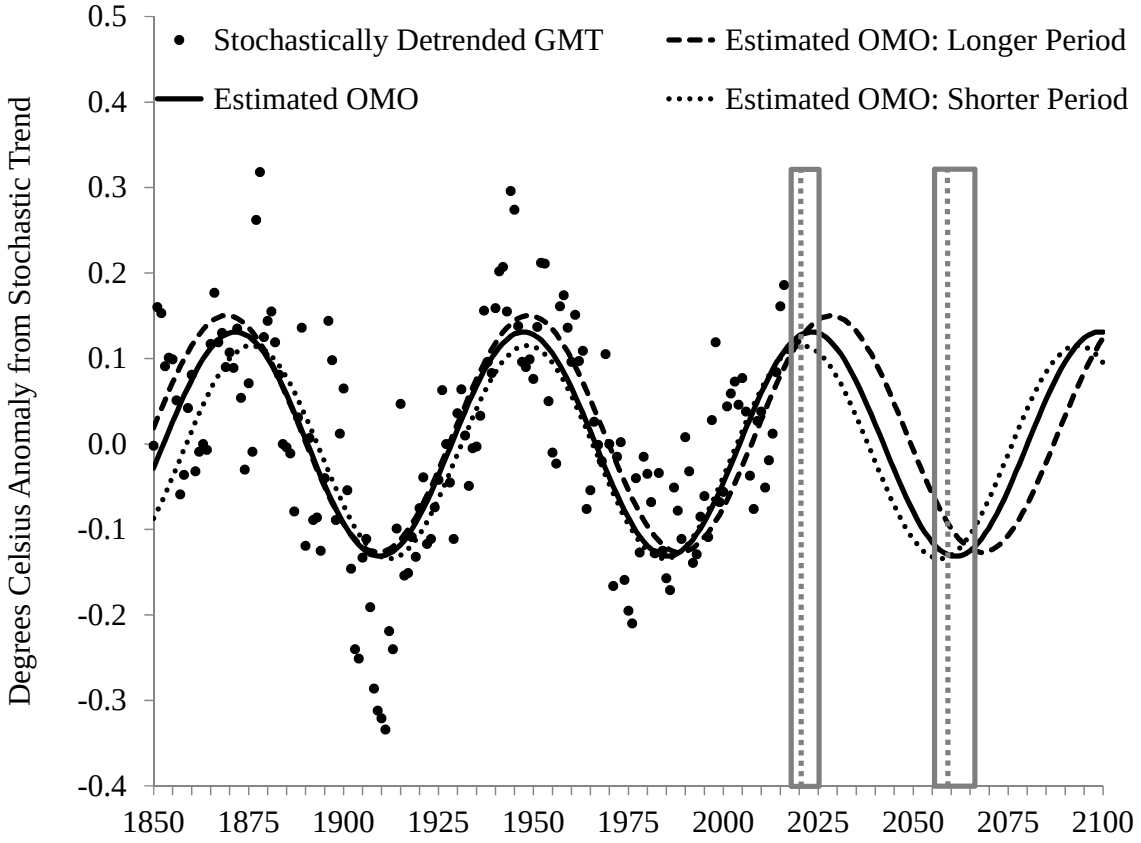

Figure 2. OMO, 1850-2100. Estimated and predicted Oceanic Multidecadal Oscillation, with 90% confidence interval on the next downturn and subsequent upturn.

in global mean temperatures may be expected to continue as the OMO continues to put upward pressure on temperatures. Put more simply, the hiatus that appeared to begin in 1998 ended in 2013.

2.2 The 2023-2061 Episode

Wyatt and Curry (2014) emphasize that, although evidence supports a secularly varying oscillation like the one that we estimate, future external forcings may alter the amplitude and period of the cycle. Linear detrending may overemphasize this possibility by giving a stochastically trending series with secular oscillations the appearance of a secularly trending series with quasi-periodic or stochastic oscillations. If the long-run trend is indeed anthropogenic, the former is more appropriate than the latter.

We fit a sine function and predict it to 2100, as shown in Figure 2. After crossing zero before 2005, the sine function continues to increase for roughly $76/4 = 19$ years until about 2023, and it then *decreases* for about 38 years until roughly 2061. Figure 2 shows sine functions reflecting a lower and upper 90% confidence interval for the estimated period. This interval is not a

prediction interval for a future year, so the plots do not straddle that of the point forecasts. Nor is it constructed from standard errors, which do not reflect correlation of the estimates of the period and phase shift.

Rather, we rely on an AR(1) bootstrap strategy in the spirit of Poppick *et al.* (2017), which is described in the SOM. The 90% bootstrap confidence interval of the estimated period of 76 years is 73 to 80 years. We date the next peak in the sign function as 2023 – likely falling in the interval 2021-2028 – and the next minimum as 2061 – likely falling in the interval 2057-2068. We note that homogeneous linear detrending results in a period of 72 years, while Trenberth and Shea's (2006) stochastically detrended AMO results in a period of 78 years. Our method, which has the advantages of relating the stochastic trend to forcings by way of a physical model and generates oscillations that are statistically more regular (see the SOM for details), generates an uncertainty interval that plausibly accounts for uncertainty in the stochastic trend.

A downturn in the temperatures due to the ocean cycle implies a slowdown but not necessarily a hiatus in global warming, because the upward trend in forcings may more than offset the downturn. The model in (2) may be used to forecast temperature anomalies conditional on changes in one or more forcings. In our forecasts, we condition on volcanic activity remaining at its 2016 level and the SOI remaining at its temporal mean over 1850-2016. We do not forecast the ENSO, because it is not periodic enough that long-run forecasts of the ENSO would be very accurate and its estimated effect on temperature is not as large as that from the OMO.

We consider two possible scenarios for non-volcanic forcings, which are closely related to RCP8.5 and RCP6.0. The data in 2016 have already deviated from the RCPs, so we simply match the cumulative growth rates of all forcings CO_2 equivalents of all anthropogenic forcings under the two scenarios starting in 2016. RCP8.5 is considered by many to be "business-as-usual," and the average annual growth implied by RCP8.5 is ~~0.054~~ 0.053 W/m^2/yr. Forcings would have to grow at a sustained rate much faster than the recent growth of ~~0.047~~ 0.029 W/m^2/yr ~~since 2013~~ over 2013-16 – i.e., since the end of the hiatus and beginning of the recent El Niño period. On the other hand, RCP6.0 has an average annual growth of ~~0.022~~ 0.021 W/m^2/yr, similar to ~~0.025~~ 0.024 W/m^2/yr over the last 76 years – one complete period of the OMO – in order to filter out any multidecadal cyclicality in the forcings themselves.

Two points bear discussion. First, we are ignoring the recalcitrant component of warming (Held *et al.*, 2010), nor are we using a dynamic model to try to capture short-run dynamics. As a result, our model is set up to make conditional forecasts of roughly 5-90 years from the end of the sample. Second, forecasts are conditional on the scenarios mentioned above, but we make no attempt to forecast individual forcings, such as solar or WMGHGs.

Figure 3 shows the sample paths of the conditional forecasts under the two scenarios. Under RCP8.5, anthropogenic forcings increase so much that downturns in the OMO cycle are never again powerful enough to force a hiatus in global warming. The global temperature anomaly increases by ~~about~~ 0.022°C/yr ~~(0.020, 0.025)°~~ (0.019, 0.024)°C/yr on average to ~~2.652°~~ nearly 3°C over the base period by 2100.

Nevertheless a slowdown is predicted until about 2061 under RCP8.5, after which point temperatures growth is predicted to accelerate to a much faster rate over multiple decades than that of the historical record. Of course, our forecasts are conditional on ENSO being unrealistically flat. A hiatus could again result from a well-timed super El Niño, such as that in 1997-98. ~~We~~

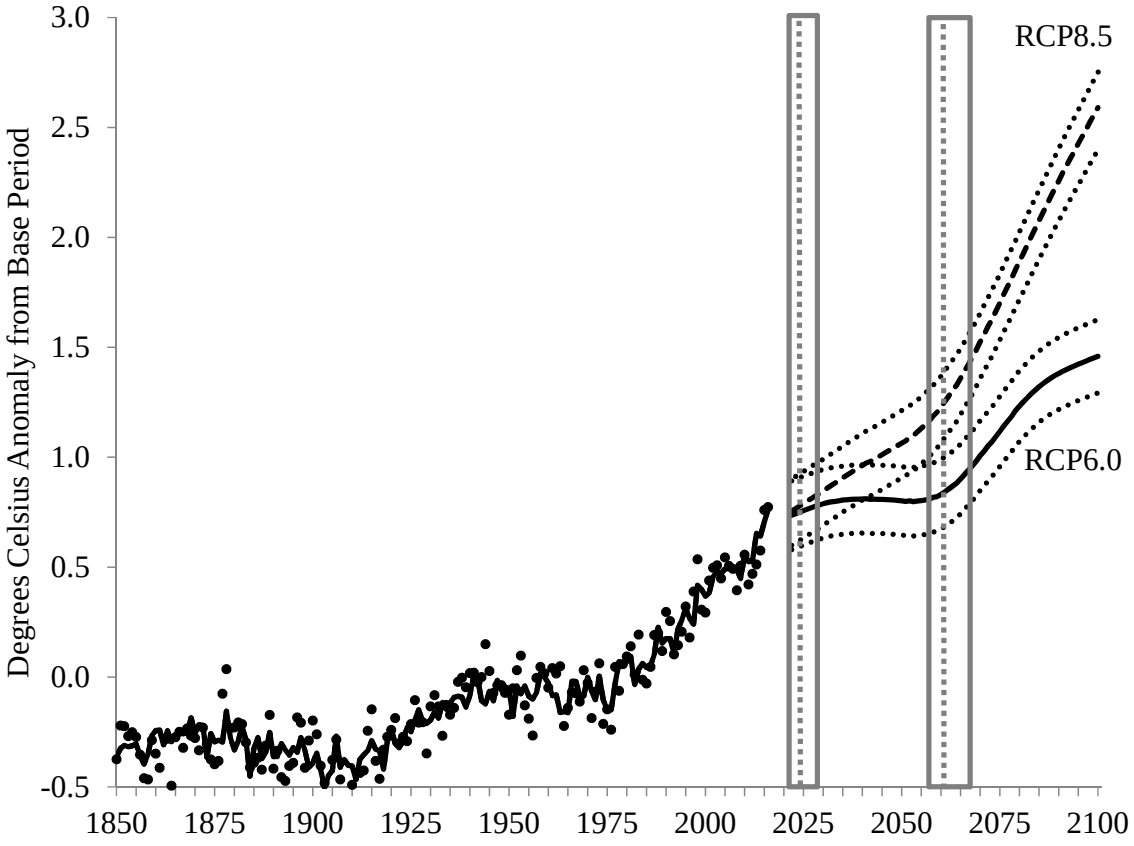

Figure 3. Conditional Forecasts of Temperature Anomalies, 2022-2100. Until 2016, dots represent actual temperature anomaly data and the solid line represents the temperature anomaly predicted by the model. Scenario labeled RCP8.5 uses RCP8.5 growth rates for anthropogenic forcings from a starting point of 2016, and similar for RCP6.0. 90% confidence bands~~generated using methodology described in the SOM~~ are shown.

~~believe that the loss in forecast accuracy from conditioning our temperature forecasts on those of the SOI would exceed the benefit from doing so.~~

Under RCP6.0, the temperature increases by about ~~0.09°~~0.008°C/yr ~~(0.007, 0.011)°~~(0.006, 0.010)°C/yr on average. By 2100, temperature anomalies increase to ~~1.496°C (1.329, 1.664)°~~1.459°C (1.292, 1.626)°C, which is ~~1.801°C (1.634, 1.969)°~~1.764°C (1.598, 1.931)°C above pre-industrial temperatures, because the base period is ~~approximately~~ 0.305°C above pre-industrial temperatures, approximated by the 1850-1879 average. While this ~~may still be~~ interval is still below 2°C, it exceeds the Paris Agreement goal of 1.5°C.

We see a substantial ebb and flow of the effect of the OMO cycle on temperatures under RCP6.0. Between 2023 and 2061, the dates identified of the next maximum and minimum of the OMO, temperature is predicted to grow by only 0.0001°C/yr – i.e., virtually no growth. In contrast to the average annual growth of anthropogenic forcings of 0.022 W/m^2/yr under RCP6.0,

this projection clearly suggests a future hiatus period that is much longer than the 1998-2013 episode. However, a very crude rule-of-thumb forecast suggests the possibility of a super El Niño in approximately 2034, which could break up the hiatus predicted by the OMO.

The variation in temperatures from the OMO is estimated to be ~~0.26°C (0.25, 0.28~~0.262°C (0.249, 0.277)°C. At its predicted nadir in 2061, temperatures are predicted have increased since 2023 by ~~0.11°C (0.11, 0.11)°~~0.100°C (0.098, 0.102)°C under RCP6.0 or ~~0.50°C (0.49, 0.51~~0.486°C (0.479, 0.494)°C under RCP8.5, meaning that they would have increased by ~~0.37°~~$0.262 + 0.100 = 0.362$°C under RCP6.0 or ~~0.76°~~$0.262 + 0.486 = 0.748$°C under RCP8.5 without the OMO. Based on the point estimates, we expect the variation of the OMO to mask the underlying warming trend by ~~34%~~ $1 - 0.486/0.748 = 35\%$ under RCP8.5 and ~~70%~~ $1 - 0.100/0.362 = 72\%$ under RCP6.0.

3 Summary and Implications for Policy Evaluation

It is no exaggeration to say that the 1998-2013 apparent hiatus in the otherwise evident trend of warming global mean temperatures has generated controversy. From a scientific point of view, a number of researchers have put forth differing explanations backed up by plausible physical models joined with sound statistical methods. Because of the critical importance of climate change to human systems – economic, political, etc. – the controversy has spilled over into the arena of public and political debate, where the lack of warming is viewed as empirical validation by those skeptical of global warming. Lack of consensus about the cause only adds to such doubt.

In this paper, we disentangle some of the causes of the 1998-2013 hiatus ~~*and*~~ and subsequent run-up in temperatures using a modern statistical technique, a semiparametric cointegrating regression, based on an energy balance model. Our main findings for this period suggest that the three main factors driving the hiatus were (a) the unusually warm year of 1998, even conditional on the ENSO, (b) the ENSO itself, and (c) the increase in tropospheric aerosols during that period, though the latter is measured with a high degree of uncertainty. Other potential causes that we investigate had considerably less impact or else an accelerating rather than confounding impact on rising temperatures. Our statistical model not only explains much of the hiatus but also explains the rapid warming since 2013. We find that this warming marks the end of the hiatus, in contrast to some findings in the literature (Chen and Tung, 2014, Knutson *et al.*, 2016, e.g.) but consistently with that of Sévellec and Drijfhout (2018)

Further, fitting the mean of the distribution of detrended ocean temperature anomalies (an oceanic multidecadal oscillation) to a periodic function enables us to make forecasts of the global mean temperature conditional on forcing scenarios. If forcings grow at the same rate as they have for the past 76 years, the estimated period of the OMO, we can expect a longer hiatus in global warming from about 2023 to about 2061, roughly 3-4 decades. The controversy of the recent 15-year hiatus is a precursor to that which may result from a much longer one. Kaufmann *et al.* (2017) recently showed a correlation between climate skepticism and locally cooler (or less warm) temperatures in the US. If the lack of warming indeed drives doubt, three decades of no warming is indeed a long period to fuel skepticism. Nevertheless, on the current trajectory, we can expect the decades following the next hiatus to push well past the 1.5°C goal of the Paris Agreement and even past 2°C.

Even though our model makes use of spatially disaggregated sea surface temperatures, our results have nothing to say directly about regional differences in temperature oscillations. As emphasized by Kaufmann *et al.* (2017) and many other authors, the effects are spatially heterogeneous. Based on estimated warming trends displayed in Figure S.1 of the SOM, we speculate that the effect of a future hiatus will be more noticeable in the vicinity of the Pacific and Indian Oceans than in the vicinity of the Atlantic Ocean, because the latter has more strongly increasing trends.

It may be useful to assign a probability to the possibility of a future multidecadal hiatus. Such a forecast would require more information and entail more uncertainty than the conditional forecasts above, because a probability distribution would be needed for future forcings. Rather than try to forecast forcings, one could base such a forecast on, say, expert opinion of the likelihood of forcing scenarios. Suppose, for example, that one believes that forcings will increase at an average rate of w per year, where w is a random variable symmetrically distributed around RCP6.0. Figure 3 suggests that, roughly speaking, scenarios with weaker growth will result in a future hiatus, while those with stronger growth will not. Ignoring the uncertainty associated with the conditional forecasts, one with such a prior could make a prediction that a multidecadal hiatus will occur with a probability of roughly 50%. We emphasize, however, the inherent uncertainty in such an exercise, even taking into account our allowance for uncertainty in the data and estimates.

Our forecasts are *conditional* on hypothetical concentration pathways. We cannot and do not suggest that policy should be based on our results. Rather, we seek to inform scientists and policymakers of the possibility of a warming hiatus due to a natural cycle. Such a cycle may be expected to have a confounding effect on policy evaluation, because a natural downturn may be mistaken for the effectiveness of mitigation. Quasi-experimental statistical evaluation of such policies must take into account this effect to avoid mistaking a failed policy for a successful one.

Code and data availability. Data, code, and appendices are available as Supplementary Online Material.

Competing interests. The authors declare no competing interests.

Acknowledgements. The authors appreciate useful feedback from Buz Brock, Neil Ericsson, David Hendry, Luke Jackson, and participants of the 2017 Conference on Econometric Models of Climate Change (Nuffield College, University of Oxford), a seminar at the Korea Energy Economics Institute, and a colloquium at the University of Missouri. All errors are our own.

References

Chen, X. and Tung, K.K.: Varying planetary heat sink led to global-warming slowdown and acceleration, Science, 345, 897-903, https://doi.org/10.1126/science.1254937, 2014.

Chen, X. and Tung, K.K.: Variations in ocean heat uptake during surface warming hiatus, Nat Commun, 7:12541, https://doi.org/10.1038/ncomms12541, 2016.

Cowtan, K. and Way, R.G.: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends, Q J Roy Meteor Soc, 140, 1935-1944, https://doi.org/10.1002/qj.2297, 2014.

Drijfhout, S.S., Blaker, A.T., Josey, S.A., Nurser, A.J.G., Sinha, B., and Balmaseda, M.A.: Surface warming hiatus caused by increased heat uptake across multiple ocean basins, Geophys Res Lett, 41, 7868-7874, https://doi.org/10.1002/2014GL061456, 2014.

Enfield, D.B., Mestas-Nunez, A.M., and Trimble, P.J.: The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S., Geophys Res Lett, 28, 2077-2080, https://doi.org/10.1029/2000GL012745, 2001.

Estrada, F., Perron, P., and Martínez-López, B.: Statistically derived contributions of diverse human influences to twentieth-century temperature changes, Nat Geosci, 6, 1050-1055, https://doi.org/10.1038/ngeo1999, 2013.

Gulev, S.K., Latif, M., Keenlyside, N., Park, W., and Koltermann, K.P.: North Atlantic Ocean control on surface heat flux on multidecadal timescales, Nature, 499, 464-468, https://doi.org/10.1038/nature12268, 2013.

Hansen, J., Sato, M., Kharecha, P., von Schuckmann, K., Beerling, D.J., Cao, J., Marcott, S., Masson-Delmotte, V., Prather, M.J., Rohling, E.J., Shakun, J., Smith, P., Lacis, A., Russell, G., and Ruedy, R.: Young people's burden: requirement of negative CO2 emissions, Earth Syst Dynam, 8, 577-616, https://doi.org/10.5194/esd-8-577-2017, 2017.

Held, I.M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., and Vallis, G.K.: Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing, J Climate, 23, 2418-2427, https://doi.org/10.1175/2009JCLI3466.1, 2010.

Hoesly, R.M., Smith, S.J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J.J., Vu, L., Andres, R J., Bolt, R.M., Bond, T.C., Dawidowski, L., Kholod, N., Kurokawa, J.-I., Li, M., Liu, L., Lu, Z., Moura, M.C.P., O'Rourke, P.R., and Zhang, Q.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emission Data System (CEDS), Geosci. Model Dev., 11, 369-408, https://doi.org/10.5194/gmd-11-369-2018, 2018.

Huber, M. and Knutti, R.: Natural variability, radiative forcing and climate response in the recent hiatus reconciled, Nat Geosci, 7, 651-656, https://doi.org/doi.org/10.1038/ngeo2228, 2014.

IPCC: Summary for Policymakers, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P.M., Cambridge University Press, Cambridge, 1-27, 2013.

Karl, T.R., Arguez, A., Huang, B., Lawrimore, J.H., McMahon, J.R., Menne, M.J., Peterson, T.C., Vose, R.S., and Zhang, H.-M.: Possible artifacts of data biases in the recent global surface warming hiatus, Science, 348, 1469-1472, https://doi.org/10.1126/science.aaa5632, 2015.

Kaufmann, R.K., Kauppi, H., Mann, M.L., and Stock, J.H.: Reconciling anthropogenic climate change with observed temperature 1998-2008, P Natl Acad Sci USA, 108, 11790-11793, https://doi.org/10.1073/pnas.1102467108 , 2011.

Kaufmann, R.K., Kauppi, H., Mann, M.L., and Stock, J.H.: Does temperature contain a stochastic trend: linking statistical results to physical mechanisms, Climatic Change, 118, 729-743, https://doi.org/10.1007/s10584-012-0683-2, 2013.

Kaufmann, R.K., Kauppi, H., and Stock, J.H.: Emissions, concentrations and temperature: a time series analysis, Climatic Change, 77, 249-278, https://doi.org/10.1007/s10584-006-9062-1, 2006a.

Kaufmann, R.K., Kauppi, H., and Stock, J.H.: The relationship between radiative forcing and temperature: what do statistical analyses of the instrumental temperature record measure? Climatic Change, 77, 279-289, https://doi.org/10.1007/s10584-006-9063-0, 2006b.

Kaufmann, R.K., Kauppi, H., and Stock, J.H.: Does temperature contain a stochastic trend? Evaluating conflicting statistical results, Climatic Change, 101, 395-405, https://doi.org/10.1007/s10584-009-9711-2, 2010.

Kaufmann, R.K., Mann, M.L., Gopal, S., Liederman, J.A., Howe, P.D., Pretis, F., Tang, X., and Gilmore, M.: Spatial heterogeneity of climate change as an experiential basis for skepticism, P Natl Acad Sci USA, 114, 61-71, https://doi.org/10.1073/pnas.1607032113 , 2017.

Keenlyside, N.S., Latif, M., Jungclaus, J., Kornblueh, L., and Roeckner, E.: Advancing decadal-scale climate prediction in the North Atlantic sector, Nature, 453, 84-88, https://doi.org/10.1038/nature06921, 2008.

Kennedy, J.J., Rayner, N.A., Smith, R.O., Saunby, M., and Parker, D.E.: Reassessing biases and other uncertainties in sea-surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties, J Geophys Res, 116, D14103, https://doi.org/10.1029/2010JD015218, 2011a.

Kennedy, J.J., Rayner, N.A., Smith, R.O., Saunby, M., and Parker, D.E.: Reassessing biases and other uncertainties in sea-surface temperature observations measured in situ since 1850: 2. Biases and homogenisation, J Geophys Res, 116, D14104, https://doi.org/10.1029/2010JD015220, 2011b.

Knight, J.R., Allan, R.J., Folland, C.K., Vellinga, M., and Mann, M.E.: A signature of persistent natural thermohaline circulation cycles in observed climate, Geophys Res Lett, 32, L20708, https://doi.org/10.1029/2005GL024233, 2005.

Knutson, T.R., Zhang, R., and Horowitz, L.W.: Prospects for a prolonged slowdown in global warming in the early 21st century, Nature Commun 7, 13676, https://doi.org/10.1038/ncomms13676, 2016.

Kosaka, Y. and Xie, S.-P.: Recent global-warming hiatus tied to equatorial Pacific surface cooling, Nature, 501, 403-416, https://doi.org/10.1038/nature12534, 2013.

Lindzen, R.S. and Giannitsis, C.: On the climatic implications of volcanic cooling, J of Geophys Res, 103, 5929-5941, https://doi.org/10.1029/98JD00125 , 1998.

Medhaug, I., Stolpe, M.B., Fischer, E.M., and Knutti, R.: Reconciling controversies about the 'global warming hiatus,' Nature, 545, 41-56, https://doi.org/10.1038/nature22315, 2017.

Meehl, G.A., Arblaster, J.M., Fasullo, J.T., Hu, A., and Trenberth, K.E.: Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods, Nat Clim Change, 1, 360-364, https://doi.org/10.1038/nclimate1229, 2011.

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 dataset, J Geophys Res, 117, D08101, https://doi.org/10.1029/2011JD017187, 2012.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., and Zhang, H.: Anthropogenic and natural radiative forcing., in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P.M., Cambridge University Press, Cambridge, 659-740, 2013.

Neely III, R.R., Toon, O.B., Solomon, S., Vernier, J.-P., Alvarez, C., English, J.M., Rosenlof, K.H., Mills, M.J., Bardeen, C.G., Daniel, J.S., and Thayer, J.P.: Recent anthropogenic increases in SO2 from Asia have minimal impact on stratospheric aerosol, Geophys Res Lett, 40, 999-1004, https://doi.org/10.1002/grl.50263, 2013.

North, G.R.: Theory of energy-balance climate models, J Atmos Sci, 32, 2033-2043, https://doi.org/10.1175/1520-0469(1975)032<2033:TOEBCM>2.0.CO;2, 1975.

North, G.R. and Cahalan, R.F.: Predictability in a solvable stochastic climate model, J Atmos Sci, 38, 504-513, https://doi.org/10.1175/1520-0469(1981)038<0504:PIASSC>2.0.CO;2, 1981.

North, G.R., Cahalan, R.F., and Coakley, Jr., J.A.: Energy balance climate models, Rev Geophys and Space Ge, 19, 91-121, https://doi.org/10.1029/RG019i001p00091, 1981.

Park, J.Y., Shin, K., and Whang, Y.J.: A semiparametric cointegrating regression: Investigating the effects of age distributions on consumption and saving, J Econometrics, 157, 165-178, https://doi.org/10.1016/j.jeconom.2009.10.032, 2010.

Poppick, A., Moyer, E.J., and Stein, M.L.: Estimating trends in the global mean temperature record, Advances in Statistical Climatology, Meteorology, and Oceanography, 3, 33-53, https://doi.org/10.5194/ascmo-3-33-2017, 2017.

Pretis, F.: Econometric Models of Climate Systems: The Equivalence of Two-Component Energy Balance Models and Cointegrated VARs, University of Oxford, Department of Economics Discussion Paper, Number 750, 2015.

Pretis, F., Mann, M.L., and Kaufmann, R.K.: Testing competing models of the temperature hiatus: assessing the effects of conditioning variables and temporal uncertainties through sample-wide break detection, Climatic Change, 131, 705-718, https://doi.org/10.1007/s10584-015-1391-5, 2015.

Roberts, C.D., Palmer, M.D., McNeall, D., and Collins, M.: Quantifying the likelihood of a continued hiatus in global warming, Nat Clim Change, 5, 337-342, https://doi.org/10.1038/nclimate2531, 2015.

Ropelewski, C.F. and Jones, P.D.: An extension of the Tahiti-Darwin Southern Oscillation Index, Mon Weather Rev, 115, 2161-2165, https://doi.org/10.1175/1520-0493(1987)115<2161:AEOTTS>2.0.CO;2, 1987.

Schmidt, G.A., Shindell, D.T., and Tsigaridis, K.: Reconciling warming trends, Nat Geosci, 7, 158-160, https://doi.org/10.1038/ngeo2105, 2014.

Sévellec F. and Drijfhout, S.S.: A novel probabilistic forecast system predicting anomalously warm 2018-2022 reinforcing the long-term global warming trend, Nat Commun, 9:3024, https://doi.org/10.1038/s41467-018-05442-8, 2018.

Shindell, D.T.: Inhomogeneous forcing and transient climate sensitivity, Nat Clim Change, 4, 274-277, https://doi.org/10.1038/nclimate2136, 2014.

Steinman, B.A., Mann, M.E., and Miller, S.K.: Atlantic and Pacific multidecadal oscillations and Northern Hemisphere temperatures, Science, 347, 988-991, https://doi.org/10.1126/science.1257856, 2015.

Storelvmo, T., Leirvik, T., Lohmann, U., Phillips, P.C.B., and Wild, M.: Disentangling greenhouse warming and aerosol cooling to reveal Earth's climate sensitivity, Nat Geosci, 9, 286-289, https://doi.org/10.1038/ngeo2670, 2016.

Timmermann, A., Oberhuber, J., Bacher, A., Esch, M., Latif, M., and Roeckner, E.: Increased El Niño frequency in a climate model forced by future greenhouse warming, Nature, 398, 694-697, https://doi.org/10.1038/19505, 1999.

Trenberth, K.E. and Shea, D.J.: Atlantic hurricanes and natural variability in 2005, Geophys Res Lett, 33, L12704, https://doi.org/10.1029/2006GL026894, 2006.

Vernier, J.-P., Thomason, L.W., Pommereau, J.-P., Bourassa, A., Pelon, J., Garnier, A., Hauchecorne, A., Blanot, L., Trepte, C., Degenstein, D., and Vargas, F.: Major influence of tropical volcanic eruptions on the stratospheric aerosol layer during the last decade, Geophys Res Lett, 38, L12807, https://doi.org/10.1029/2011GL047563, 2011.

Wyatt, M.A. and Curry, J.A.: Role for Eurasian Arctic shelf sea ice in a secularly varying hemispheric climate signal during the 20th century, Clim Dynam, 42, 2763-2782, https://doi.org/10.1007/s00382-013-1950-2, 2014.

Yao, S.-L., Huang, G., Wu, R.-G., and Qu, X.: The global warming hiatus – a natural product of interactions of a secular warming trend and a multi-decadal oscillation, Theor Appl Climatol, 123, 349-360, https://doi.org/10.1007/s00704-014-1358-x, 2016.

Supplementary Online Materials for "Dating Hiatuses: A Statistical Model of the Recent Slowdown in Global Warming – and the Next One"

J. Isaac Miller[1] and Kyungsik Nam[1]

[1]University of Missouri

Correspondence: J. Isaac Miller (millerjisaac@missouri.edu)

S.1 Updating the Forcing Series

We update the forcing series of Hansen *et al.* (2017) to 2016 as follows. We regress the first three series, CO_2, CH_4, N_2O, (in W/m^2) onto the natural log of the series given by NOAA[1] in ppm for carbon dioxide and in ppb for methane and nitrous oxide. We then predict the 2016 ~~forcings~~ forcing using the natural log of the 2016 NOAA data. The 2016 solar forcing is updated by imposing the 2015-16 percent change from NASA.[2]

In contrast ~~,~~ to CO_2, CH_4, and N_2O, forcings from CFCs are changing very slowly and, so we set 2016 to be the average of 2004-2015. We take the same moving average approach to estimate the 2016 forcings for ozone ~~, tropospheric aerosols & surface albedo,~~ and volcanoes as for CFCs. Volcanic activity over this period was not trivial, as noted by Vernier *et al.* (2011) and Neely *et al.* (2013), but neither are there any major eruptions on the order of Mount Pinatubo in 1991. ~~Finally, solar data in~~

Forcings from tropospheric aerosols & surface albedo near the end of the sample are changing smoothly, but not as slowly as those from CFCs, ozone, and volcanoes, so we set the 2016 ~~is updated by imposing the 2015-16 percent change from NASA.~~[3] forcing to be that of 2015 plus the average annual growth rate over 2004-2015.

S.2 Estimating the Oceanic Multidecadal Oscillation

Ocean cycles in mean temperature data – and the AMO in particular – have been estimated a number of ways in the literature. A key problem in estimating the cycle is removing the long-run trend due to global climate change. A common method for this purpose is linear detrending of GMT (Enfield *et al.*, 2001; Wyatt and Curry, 2014), although linear detrending has been criticized for this purpose by the IPCC (Bindoff *et al.*, 2013). Approaches using stochastic trends include those of Trenberth and Shea (2006), who use temperatures in other oceans to detrend the Atlantic, and Lenton *et al.* (2017), who use global mean temperature to detrend regions in the Atlantic and Pacific.

[1]Downloaded from www.columbia.edu/~mhs119/GHGs on May 15, 2017.

[2]Downloaded from solarscience.msfc.nasa.gov on May 15, 2017.

[3]~~Downloaded from on May 15, 2017.~~

Another problem that we must avoid is over-fitting the statistical model in equation (2) in the paper. As an example of over-fitting, consider 2016, which was an unusually warm El Niño year. A variable constructed by simply detrending sea surface temperatures would have a particularly high value for 2016. Regressing GMT onto detrended GMT or sea surface temperatures would show a superficially good fit, in the sense that the model could not distinguish between secular cyclical variability and idiosyncratic noise.

Our approach handles these two problems by filtering out both long-run and short-run information from the time series of temperature anomaly distributions. We decompose the temperature anomaly T^a into a long-run trend component T^τ, a stationary multidecadally oscillating component T^s, and a noise component T^n, so that $T^a = T^\tau + T^s + T^n$. Similar to the AMO, we refer to T^s as the Oceanic Multidecadal Oscillation (OMO).

~~In order to estimate T^s~~To estimate the OMO, we first divide up the HadSST3 data into ~~oceans~~five ocean regions: North Atlantic ($i = 1$), South Atlantic ($i = 2$), North Pacific ($i = 3$), South Pacific ($i = 4$), and Indian ($i = 5$), ~~Indian,~~ defined according to NOAA.[3] We then calculate the mean ~~for each ocean~~ temperature anomaly $\bar{T}^a_{it}$ for each region $i = 1, \ldots, 5$ at each year ~~and then detrend those means. Rather than using a linear time trend or using the trends from other oceans, we regress these means into~~ $t = 1850, \ldots, 2016$ and then estimate a stochastic trend for each region. The trends are estimated by linear regressions of $\bar{T}^a_{it}$ onto a constant and WMGHGs ~~, reflecting models in the literature,~~ in W/m^2 relative to 1850 (h^{WMGHG}_t), which may be written as

$$\bar{T}^a_{it} = \alpha_{0i} + \alpha_{1i} h^{WMGHG}_t + v_{it}$$

for $i = 1, \ldots, 5$ and $t = 1850, \ldots, 2016$. They are identified by the predicted mean temperature anomaly $T^\tau_{it} = \hat{\alpha}_{0i} + \hat{\alpha}_{1i} h^{WMGHG}_t$ from each of these five regressions.

Our stochastic trend approach using WMGHGs is justified by a widely accepted physical link between temperatures and forcings from energy balance models like that in ~~our~~ equation (2)~~, which assume a long-run relationship between temperature and WMGHGs. Doing so for each ocean separately allows for heterogeneous local~~ . In contrast, the linear deterministic trend approach replaces h^{WMGHG}_t with t in the regression above. In order for the linear trend approach to be justified by a physical model, forcings would have to increase by a constant amount over time, which is justified neither by the data nor by the process by which WMGHG emissions cumulate to concentrations. Moreover, varying coefficients over i allow for heterogeneous multibasin climate sensitivity – i.e., it allows for the sea surface temperatures in different ~~oceans~~ ocean regions to be influenced by WMGHG differently over the long term.

Figure S.1 shows the predicted temperature trends T^τ_{it} from these regressions. Note that the Pacific and Indian Oceans share a common warming trend, while those of the Atlantic are quite a bit different. The Atlantic is warming faster than the other oceans, with the South Atlantic catching up to the Indian and Pacific from a colder starting point and the North Atlantic becoming increasingly warmer.

~~**Number of Observations per Year from HadSST Data.** Number of observation available to estimate annual probability density functions.~~

[3] See www.nodc.noaa.gov/woce/woce_v3/wocedata_1/woce-uot/summary/bound.htm.

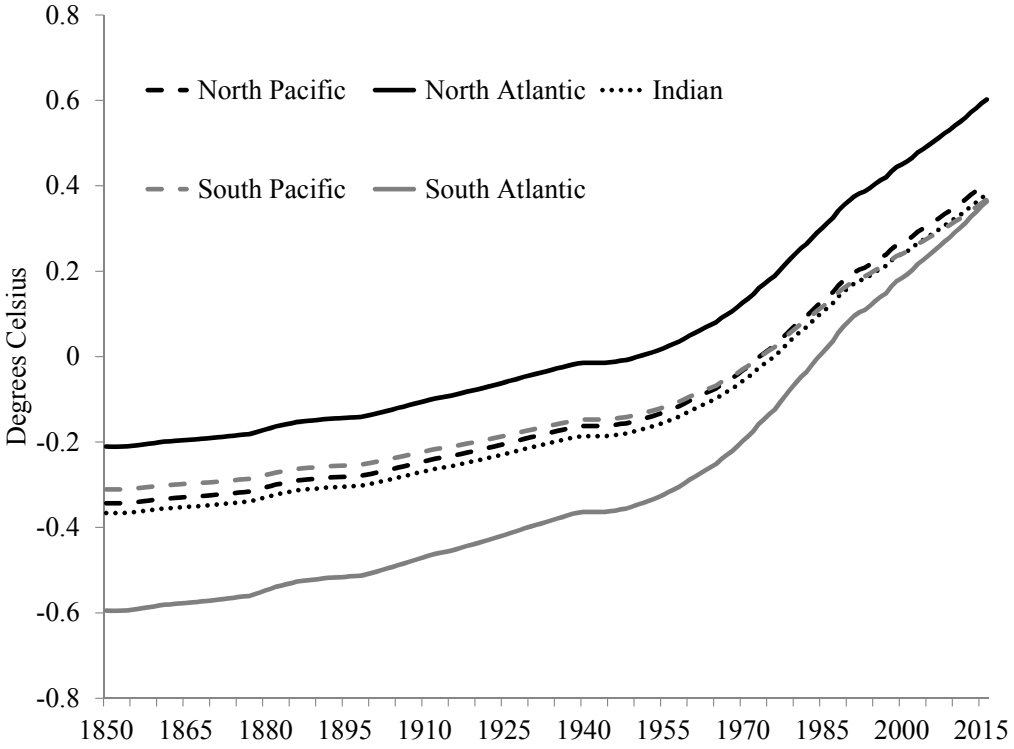

Figure S.1. Heterogeneous Oceanic Temperature Trends. Fitted anomalies from regressing average temperature anomalies of five ocean regions onto a constant and WMGHGs.

In order to detrend the distribution of sea surface temperatures, we ~~simply~~ subtract the estimated trend ~~for each ocean~~ T^{τ}_{it} for each region from the series of ~~temperatures~~ temperature anomalies in each of the $5°$ by $5°$ boxes in that ocean. ~~The maximum number of observations in the HadSST3 data is $36 \times 72 = 2{,}592$ per month and $2{,}592 \times 12 = 31{,}104$ per year~~Mathematically, this is $T^{s+n}_{k(i)t} = T^{a}_{k(i)t} - T^{\tau}_{it}$, where $k(i)$ indexes boxes in each ocean region i. Figure S.2 shows the actual number available for each year of our sample. A maximum of $17{,}391$ is attained in 1979. Prior to the 1970's the number of observations generally increases over time, but with noticeable dips during major international disruptions, such as World War II, World War I, and the American Civil War.

A standard nonparametric density estimation technique (Gaussian kernel with Silverman bandwidth) is used to estimate the density $f^{s+n}_t(r)$ of heterogeneously detrended sea surface temperature anomalies ~~. Detrending removes T^{τ}~~ T^{s+n}_{kt}, where k now indexes boxes across all ocean regions: in set notation, $\{T^{s+n}_{kt}\} = \cup_{i=1}^{5}\{T^{s+n}_{k(i)t}\}$. The detrending procedure above has removed T^{τ}_{it}, so that $f^{s+n}_t(r)$ reflects the density of $T^s + T^n$ across all of the ocean boxes. We omit 0.5% of the outliers in each tail, which we believe is an adequate threshold to ameliorate well-known boundary problems from kernel density estimation without substantively altering the moments of the distribution. The support of the density of outliers becomes $r \in [-2.995, 3.085]$. We

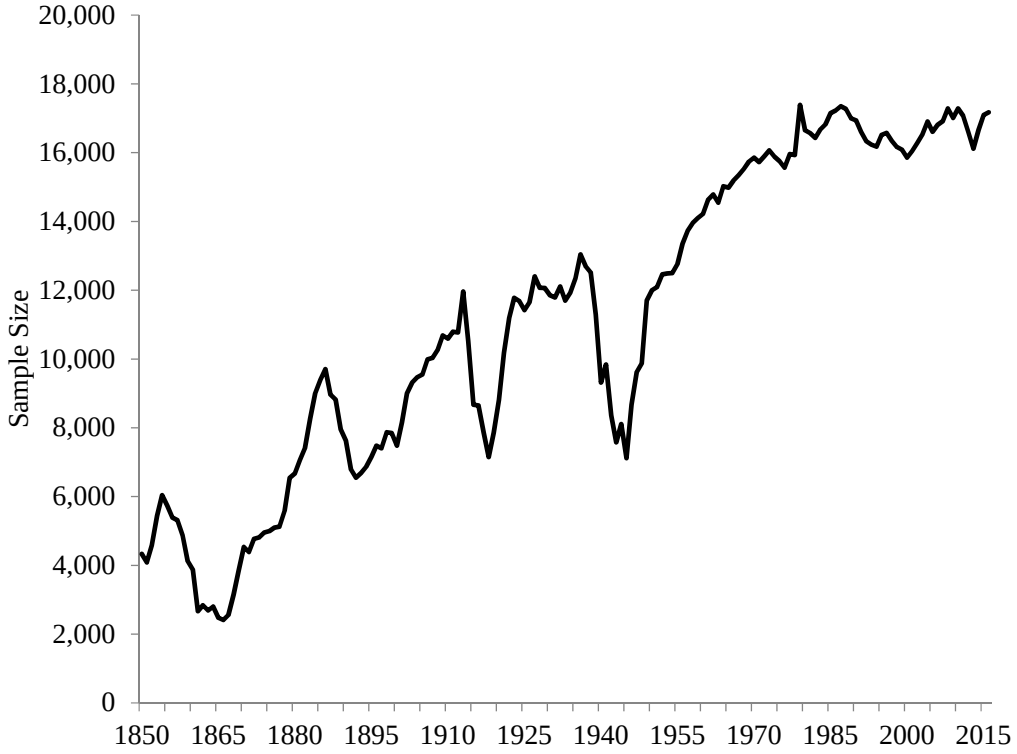

Figure S.2. Number of Observations per Year from HadSST Data. Number of observation available to estimate annual probability density functions.

abbreviate the endpoints of the support simply by $-$ for r^- and $+$ for r^+ throughout the SOM. Figure S.3 (top panel) shows the density ~~$f_t^{s+n}(r)$~~ $f_t^{s+n}(r)$ of the stationary temperature distribution for each year.

Next, we smooth $f_t^{s+n}(r)$ to obtain $f_t^s(r)$ by removing short-run noise T^n. To ~~do so~~this end, we first calculate the spatial mean $\int_-^+ r f_t^{s+n}(r)dr$, which could be referred to as the heterogeneously detrended oceanic mean temperature: it approximates the mean of T_{kt}^{s+n} over k. We fit the result to a single sine function, estimating

$$\int_-^+ r f_t^{s+n}(r)dr = \theta_1 \sin(\theta_2(t/T) + \theta_3) + \theta_4 + e_t, \tag{S.1}$$

using nonlinear least squares. ~~We abbreviate the endpoints of the integral simply by $-$ for r^- and $+$ for r^+ throughout the SOM.~~

Nonlinear least squares estimates a periodic function with an amplitude of $0.130°$C, a vertical shift of $-0.000°$C (nearly zero), a period given by $2\pi/13.75 \times T \simeq 76$ years, and a phase shift given by $0.21/13.75 \times T \simeq 3$ years (Table S.1). This period is roughly consistent with the Wyatt-Curry "stadium wave" with a half-period cooling regime of 31-38 years from about

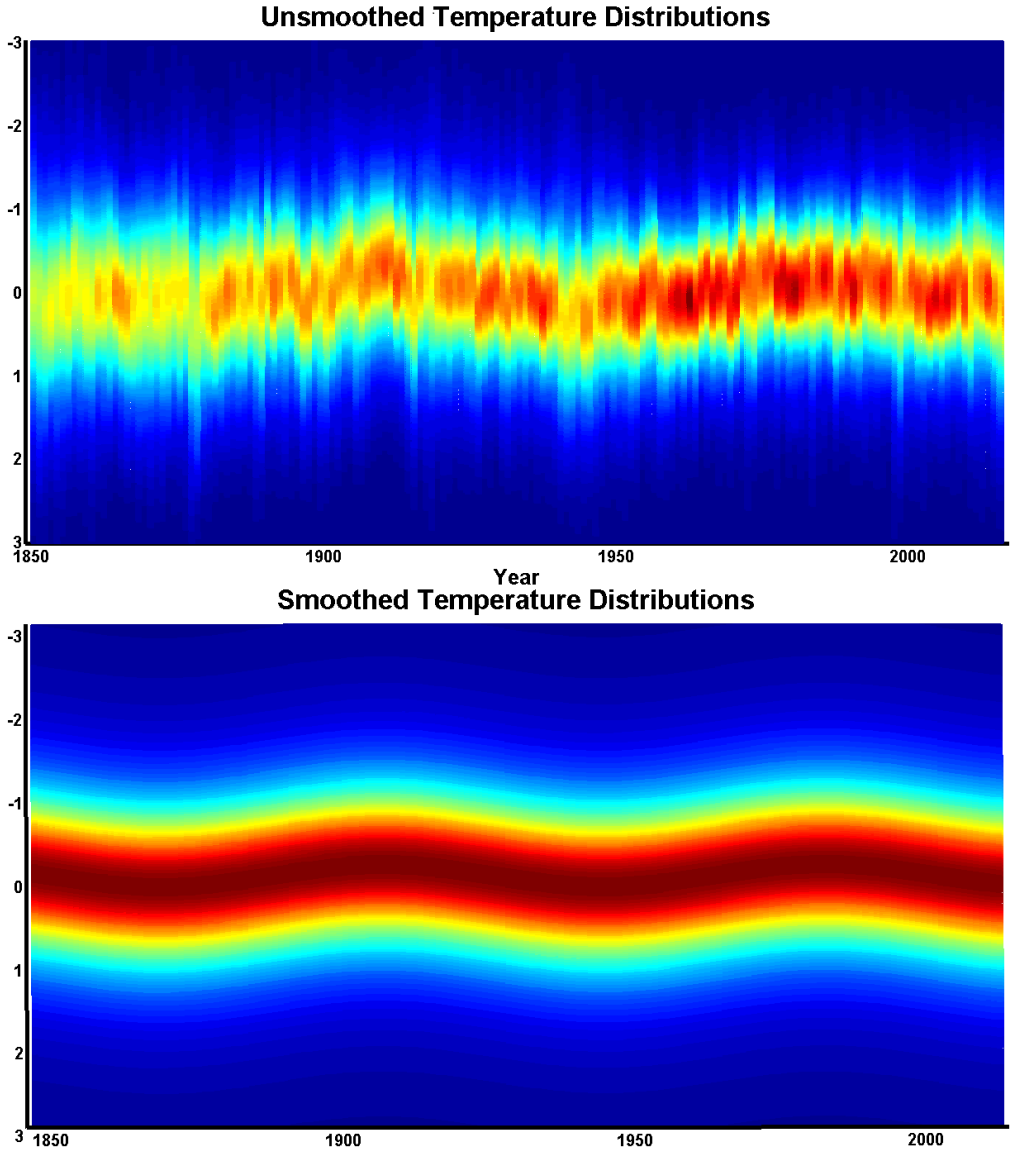

Figure S.3. ~~Unsmoothed and Smoothed Anomaly Distributions Over Time.~~ Unsmoothed and Smoothed Detrended Anomaly Distributions Over Time. Top: Unsmoothed anomaly distribution from heterogeneously detrended sea surface temperature anomalies. Bottom: Smoothed anomaly distribution from fitting the unsmoothed distribution to a sine function, used to represent the OMO.

	Linear/Homog		Trenberth-Shea		WMGHG/Heterog	
	est.	s.e.	est.	s.e.	est.	s.e.
θ_1	0.13	0.11	0.12	0.13	0.13	0.11
θ_2	14.66	2.71	13.53	4.99	13.75	2.95
θ_3	−0.11	1.61	0.28	3.47	−0.21	1.67
θ_4	−0.03	0.08	−0.03	0.10	−0.00	0.08
R^2	0.42		0.34		0.53	

Table S.1. Periodic Function Estimation Results. Results from fitting the ~~WMGHG/heterogeneously detrended OMO,~~ linearly/homogeneously detrended OMO, ~~and~~ the Trenberth-Shea (2006) AMO, and the WMGHG/heterogeneously detrended OMO to the ~~nonlinear regression~~ periodic function in (S.1).

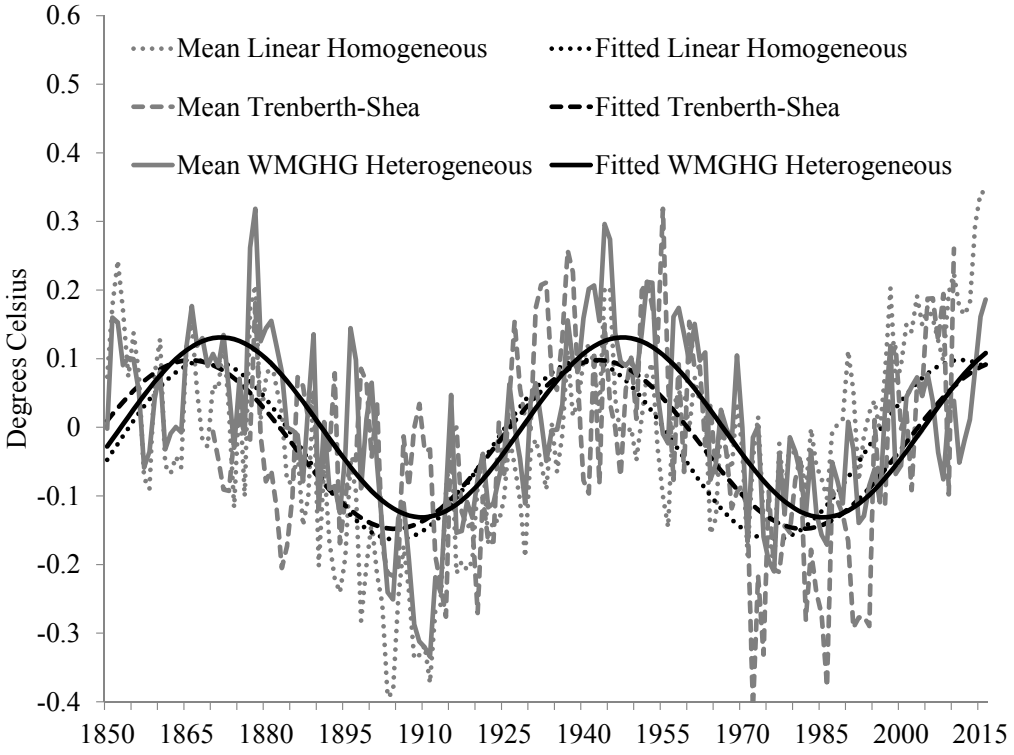

Figure S.4. Mean and Fitted Oscillations. Mean and fitted oscillations using data detrended by a single linear trend, using ~~heterogeneous WMGHG detrending, and using~~ data from Trenberth and Shea (2006), and using data detrended by allowing heterogeneous WMGHG trends.

1940 to about 1975. The years over the sample in which the OMO has a neutral effect – neither cooling nor warming – are (approximately) 1852, 1928, and 2004.

As a comparison, Table S.1 and Figure S.4 compare the OMO and periodic function estimated in this manner with an OMO and periodic function estimated using linear detrending and with the AMO signal of Trenberth and Shea (2006)[4] and similarly estimated periodic function. The linear detrending method ~~estimates~~ yields a shorter period of 72 years, while the Trenberth-Shea AMO signal has a longer period of 78 years. Although the AMO signal has a longer period, the phase shift is negative, so that the next peak occurs just after the end of the sample. In contrast, the linear detrending method shows a peak in about 2011 – in stark contrast to the recent high temperatures in 2015 and 2016. The table also shows R^2's from the three regressions. These statistics should be interpreted with caution, both because they are R^2's from nonlinear regressions and because the regressions have the same functional form of time but different regressands. Nevertheless, we interpret them as evidence that the periodic function is a better approximation to the oscillation estimated using the proposed detrending method.

Now, in order to ~~estimate~~ identify the distribution of T^s from that of T^s+T^n, we construct a distribution that is changing only in its mean over time. To ~~that~~ this end, we first create a measure of the average distribution $f^n(r)$ of "de-cycled" anomalies T^n with trend and periodic function removed. The density $f_t^{s+n}(r)$ is already detrended, but in order to remove the multidecadal cycle, we change the support by subtracting the heterogeneously detrended oceanic mean temperature $T^s = \hat{\theta}_1 \sin(\hat{\theta}_2(t/T) + \hat{\theta}_3) + \hat{\theta}_4$ estimated from equation (S.1) from each detrended temperature anomaly in each year. For example, the density function estimated for a detrended temperature anomaly of $1°$C in year t becomes the "de-cycled" density function estimate at $(1 - T^s)°$C and, most importantly, the density at $(T^s)°$C becomes $0°$C. We then average the densities at each temperature anomaly r across the sample, 1850-2016, obtaining a "de-cycled" density $f^n(r)$.

Finally, we create a series of estimated densities ~~$f_t(r)$~~ $f_t^s(r)$, smoothed versions of $f_t^{s+n}(r)$. ~~In order to do so~~To that end, we reverse the procedure described above by adding the heterogeneously detrended oceanic mean temperature T^s to $f^n(r)$ in order change the support back to ~~r~~ $r \in [-2.995, 3.085]$. We then remove outlying anomalies outside the original support. In this way, we estimate the smoothed density ~~$f_t(r)$~~ $f_t(r) = f_t^s(r)$ of the OMO, displayed in the bottom panel of Figure S.3. The methodology clearly extracts a density that (a) appears to be stationary, devoid of a long-run stochastic or deterministic trend from warming, (b) appears to be smooth, devoid of idiosyncratic noise, and (c) appears to capture the multidecadal cycle.

Disaggregated historical temperature data are measured with bias and uncertainty. By design, our method for smoothing the cyclical component eliminates any idiosyncratic/short-run uncertainty in these data. If there is a static bias throughout the time span, it is picked up by a non-zero estimate of θ_4 in equation (S.1), which we estimate to be nearly zero (Table S.1). Since we use θ_4 to build the cyclical component, any bias in our estimate of θ_4 is picked by the intercept α_0 in equations (1) and (2), but should not affect the other coefficients estimates used to make our inferences. However, if the bias changes over time in a non-idiosyncratic way, modeling it would require a more involved strategy, which we leave for future research.

S.3 Energy Balance Model

We define location $\ell \in G = L \cup O$ where the set G is all locations on the globe and L and O are sets of land and ocean locations. Location ℓ may be given as a latitude-longitude pair, in which case the integrals over ℓ below become double integrals over

[4]Downloaded from www.cgd.ucar.edu/cas/catalog/climind/AMO.html on July 17, 2017.

latitude and longitude. Letting the index $j = G, L, O$, $n_j = \int_j d\mu_\ell$ with counting measure μ denotes the number of locations in each set.

Adapting the energy balance model (EBM) of North (1975) and North and Cahalan (1981) to accommodate external forcing, we may write

$$C(\ell)dT_t(\ell) = QS(\ell)a(\ell) - (A + BT_t(\ell)) + D(\ell, T_t(\ell)) + h_t(\ell) + \varepsilon_t(\ell), \tag{S.2}$$

where C is heat capacity, T_t is temperature at time t, Q is the solar constant, S is solar irradiance, a is co-albedo, $A + BT_t$ is emitted energy, D is a linear-in-temperature approximation to the heat diffusion term in their model, h_t is radiative forcing, and ε_t is stochastic forcing. The stochastic forcing term is assumed by North *et al.* (1981) to be idiosyncratic, but we do not require this assumption.[5,6]

Historical temperature data sets typically express temperature in terms of anomalies from a base period in order to ameliorate well-known measurement errors. Accordingly, we decompose temperature as temperature during a base period b plus the temperature anomaly, $T_t = T_b + T_t^a$. The two forcing components may also be decomposed into base plus anomaly, expressed as $h_t = h_b + h_t^a$ and $\varepsilon_t = \varepsilon_b + \varepsilon_t^a$. Adapting the EBM in (S.2) to accommodate temperature anomalies allows

$$(C + B)T_t^a - CT_{t-1}^a = D(\ell, T_t^a) + h_t^a + \varepsilon_t^a, \tag{S.3}$$

by subtracting $C(\ell)dT_b(\ell) = 0$ from both sides, discretizing the derivative to a unit increment, and suppressing the location argument for now.

Because $C, B > 0$, $\pi = C/(C + B) < 1$, so that the autoregressive component may be inverted. Doing so yields

$$T_t^a = (C + B)^{-1} \sum_{i=0}^{\infty} \pi^i [D(\ell, T_{t-i}^a) + h_{t-i}^a + \varepsilon_{t-i}^a] \simeq \frac{1}{B}[D(\ell, T_t^a) + h_t^a + \varepsilon_t^a],$$

where the approximation results from a Beveridge-Nelson-type decomposition (see Phillips and Solo, 1992). The approximation is more valid when the data are cointegrated or cotrending, in which case the neglected terms have a lower asymptotic order.

What does the stochastic forcing term ε_t^a represent? Aside from noisy measurement of the data, it also accounts for otherwise missing components of temperature changes. Most notably missing are natural variability, such as changes in the ocean heat uptake, and other natural cycles, such as ENSO. As proxies, we employ the OMO, given by $T_t^s(\ell)$, and the SOI often used as a proxy for ENSO, given by S_t. Timmermann *et al.* (1999) note the possibility that external forcings may correlate with more frequent and/or severe ENSO cycles. By including the SOI in the model, we are implicitly assuming that the correlation is reflected in the SOI.

An alternative way to capture natural variability might be to allow for separate meridional ocean transport, along the lines of Rose and Marshall (2009), which could likely be accomplished along the lines of Pretis (2015) using a model that cointegrates

[5] North and Cahalan (1981) assume co-albedo to be a function of temperature as well as latitude, but more recent studies show that co-albedo is effectively constant in temperature at a given latitude (Stephens *et al.*, 2015; Stevens and Schwartz, 2012).

[6] Alexeev *et al.* (2005) note the effectiveness of modeling the diffusion coefficient D as a function of temperature in order to capture polar amplification, which is not our aim.

surface temperature with deep ocean heat content. However, keeping in mind that our aim is to model hiatus periods that may be sparsely distributed over the historical record, the short time span over which ocean heat content is measured precludes this approach.

In such a model, surface temperatures, deep ocean heat content, and forcings share a single stochastic trend, so that the marginal value of the deep ocean heat content relative to forcings is natural variability. Hence, omitting deep ocean heat content does not cause a spurious regression, but rather relegates this stationary variability to the error term. Our proxies, the OMO and SOI, allow us to explicitly model the primary multidecadal and interannual sources of this variability.

Because the EBM does not explicitly include these indicators, it is natural to model them nonparametrically. To this end, we specify the model as

$$T_a = \alpha_0 + \alpha_1[h_t(\ell) + D(\ell, T^a(\ell))] + b(T_t^s(\ell)) + c(S_t) + \eta_t(\ell), \tag{S.4}$$

where

$$b(T_t^s(\ell)) = \begin{cases} \sum_{i=1}^{m_T} \gamma_i^T b_i(T_t^s(\ell)) \text{ for } \ell \in O \\ 0 \text{ for } \ell \in L \end{cases}$$

$$c(S_t) = \sum_{i=1}^{m_S} \gamma_i^S c_i(S_t)$$

are two generic series expansions intended to capture possibly nonlinear effects of these indicators. The OMO captures variability over the ocean, so this component is set to zero over land. The SOI is a single indicator. The last term $\eta_t(\ell)$ contains the original stochastic forcings and their lags and an allowance for finite-order approximation error of the two expansions. We may think of this term representing residual forcings, and it almost certainly exhibits temporal correlation.

Now we aggregate across locations to obtain a global model. To this end, let $O(T^s) = \{\ell \in O : T^s(\ell) = T^s\}$ be a subset of O over which the $T^s(\ell)$ has the same numerical value, let $n_{O(T^s)} = \int_{O(T^s)} d\mu_\ell$ denote the number of locations in $O(T^s)$ that have the value T^s, and let f be the probability density function of $T^s(\ell)$ in O with support $[r^-, r^+]$. Note that

$$\int_{-}^{+} b(r)f(r)ds = \int_{-}^{+} \left[n_{O(T^s)}^{-1} \int_{O(T^s)} b(T^s(\ell))d\mu_\ell \right] f(T^s)dT^s = n_O^{-1} \int_O b(T^s(\ell))d\mu_\ell,$$

which means we can aggregate all of the functions b across ocean locations or we can aggregate all of the functions b with the same observed argument and then aggregate them again with weights given by the frequency of each argument. Defining $T^a = n_G^{-1} \int_G T^a(\ell)d\mu_\ell$, $h = n_G^{-1} \int_G h(\ell)d\mu_\ell$, $\eta = n_G^{-1} \int_G \eta(\ell)d\mu_\ell$, integrating across locations, and noting that the diffusion term $D(\ell, T^a(\ell))$ is constrained to integrate to zero by the first law yields the EBM in equation (1) in the paper.

S.4 Estimating the Energy Balance Model

We approximate the functions b and c nonparametrically using a series of polynomial and trigonometric functions known as the flexible Fourier functional form, which Park *et al.* (2010) analyze using a semiparametric cointegrating regression much

	OLS	CCR		OLS	CCR	
	est.	est.	s.e.	est.	est.	s.e.
α_0	-0.315	-0.313	0.018	-14.254	-13.881	13.130
α_1	0.413	0.426	0.020	0.431	0.440	0.015
α_2	0.079	0.081	0.026	0.054	0.055	0.021
γ_1				9.278	9.025	9.557
γ_2				-8.687	-8.446	9.705
δ_1						
α_0	-0.212	-0.213	0.037	-20.037	-19.339	12.474
α_1	0.408	0.419	0.019	0.426	0.431	0.015
α_2	0.093	0.094	0.026	0.069	0.069	0.020
γ_1				13.540	13.043	8.370
γ_2				-12.989	-12.494	8.500
δ_1	-0.055	-0.054	0.017	-0.062	-0.061	0.012

Table S.2. EBM Estimation Results. Results from estimating the model in (2) in the paper using least squares (OLS) and asymptotically normal canonical cointegrating regression estimates (CCR) (Park *et al.*, 2010) with $p_1 = 0, 2$, $p_2 = 0, 1$, and $q_1, q_2 = 0$ in (S.4).

like ours. This form may be written as

$$
\begin{aligned}
b_j^v(v) &= v^j \text{ for } j = 1, ..., p_1 \\
&= \cos 2\pi kv \text{ for } j = p_1 + 2k - 1 \text{ and } k = 1, ..., q_1 \\
&= \sin 2\pi kv \text{ for } j = p_1 + 2k \text{ and } k = 1, ..., q_1
\end{aligned}
$$

and analogously for $c_j^v(v)$, for $v \in [0,1]$. Using this notation, $m_T = p_1 + 2q_1$ and $m_S = p_2 + 2q_2$. It is important that these functions are defined over the unit interval, so let $b_j(r) = (r^+ - r^-)b_j^v((r - r^-)/(r^+ - r^-))$ and $c_j(S) = (S^+ - S^-)c_j^v((S - S^-)/(S^+ - S^-))$, where S^+ and S^- are the maximum and minimum observed SOI. Thus,

$$\int b_j(r) f_t(r) dr = \int (r^+ - r^-) b_j^v((r - r^-)/(r^+ - r^-)) f_t(r) dr$$

holds, making estimation convenient by simply multiplying b_j^v by the range of values in the domain of the OMO.

The optimal orders $(p_1, q_1, p_2, q_2) = (2, 0, 1, 0)$ – i.e., $m_T = 2$ and $m_S = 1$ – are jointly determined by Schwarz-Bayesian and Hannan-Quinn information criteria evaluated using least squares with p_1, p_2 up to 3 and q_1, q_2 up to 2. With $m_S = 1$, SOI enters linearly and the regressor is thus simply $(S_t - S^-)$.[7] The estimation results are given in Table S.2.

[7] All models are expected to be cointegrated, because we simply add stationary series to the model with $(p_1, q_1, p_2, q_2) = 0$, which we found to be ~~stationary~~ cointegrating (see footnote 1 in the paper). More formally, we run residual-based augmented Dickey-Fuller tests which ~~strongly rejected~~ reject the null of no cointegration up to four lags.

Coefficient Estimates Aggregate Effects		
	v^1	
v^2	v^3	
$\cos 2\pi kv$	$\sin 2\pi kv$	
°C	% missing heat	
$(p,q) = (1,0)$	0.75	
	0.122	−47.6
$(p,q) = (2,0)$	13.54	
−12.99		
	0.113	−44.1
$(p,q) = (3,0)$	−176.56	
447.34	−335.76	
	0.092	−35.9
$(p,q) = (2,1)$	1236.48	
−1502.58		
103.87	−60.31	
	0.029	−11.3

Table S.3. Robustness Checks on Order of the OMO Expansion. Results from estimating the model in (2) in the paper using least squares with $p_1 = 1,2,3$, with $q = 0$fixed and $q_1 = 0,1$ with $p_1 = 2$ fixed in (S.4).

The estimate of the aggregate effect of the OMO (net of forcings) reported in the paper for the optimal order, $(p_1,q_1) = (2,0)$, is 0.11°C $(0.08,0.14)$°C over 1998-2013, or a 41.7% $(30.0,53.5)\%$ exacerbation of the puzzle of the missing heat. As a robustness check on the effect of selection of p_1 and q_1, we varied each ± 1 from the optimal order: $(1,0)$, $(3,0)$, $(2,1)$, yielding estimates of the aggregate effects of 0.122°C (47.6% exacerbation), 0.092°C (35.9% exacerbation), and 0.029°C (11.3% exacerbation), by way of least squares. The estimates with $(1,0)$ or $(2,0)$ seem the most plausible, because the additional terms of models with $(3,0)$ and $(2,1)$ generate estimates with very large magnitudes and opposing signs, a classic sign of near multicollinearity.

S.5 Estimation of Uncertainties

S.5.1 Uncertainties from Estimating the OMO

Estimating the OMO relies on a statistical approximation, and we employ a parametric bootstrap strategy similar to that of Poppick *et al.* (2017) to account for uncertainty in estimation. Specifically, after fitting the periodic function in (S.1), we fit the residuals to an AR(1) and redraw from the residuals of the fitted AR(1). We re-create the regressand using a re-created AR(1) error with fitted autoregressive parameter (0.49). We then re-estimate all the parameters of the nonlinear regression. We conduct 999 bootstrap replications in this manner, and the sample paths plotted in Figure 2 ~~in~~ of the paper reflect the periodic functions with 0.05 and 0.95 quantiles of θ_2 in (S.1), which determines the period.

Our bootstrap differs from a typical bootstrap in that we do not re-estimate only the period and restrict the other parameters to their fitted values. Re-estimating all of the parameters allows uncertainty about the period to be correlated with uncertainty about the phase shift in particular, because uncertainty in both parameters affects our dating of the function's optima.

S.5.2 Approximations to Uncertainties in Forcings

Myhre *et al.* (2013) estimate forcings (in W/m^2 ~~)~~ with 90% confidence~~for 2010 as follows: from WMGHGs: 2.83 (2.54 to 3.12) ; from ozone: 0.07 (0.02 to 0.12); from tropospheric aerosols & surface albedo: −0.35 (−0.85 to 0.15); and from solar irradiance: 0.05 (0.00 to 0.10)~~) from WMGHGs as 2.83 ± 0.29, or $\pm 10\%$. Volcanic forcings are estimated to be -0.06 (-0.08 to -0.04) over 1999-2002 and -0.11 (-0.15 to -0.08) over 2008-2011~~.~~, or $\pm 33\%$. Similarly, forcings from solar irradiance are estimated to be -0.05 ± 0.05, or $\pm 100\%$. The data from Hansen *et al.* (2017) for ozone are calculated as the simple average of forcings from tropospheric and stratospheric ozone (see footnote c of their Table A1). The total is estimated by Myhre *et al.* (2013) ~~state that uncertainty about forcing from WMGHGs is about $\pm 10\%$, and indeed this roughly corresponds to the numerical values given for 2010. The numerical values given for volcanic forcings over the two periods correspond to $\pm 33\%$ within rounding error. If we make a similar assumption regarding the other forcings , then uncertainties about forcings from ozone, tropospheric aerosols & surface albedo , and solar are $\pm 70\%$, $\pm 143\%$, and $\pm 100\%$ respectively in 2010.~~ to be 0.35 ± 0.20, so that the average is 0.175 ± 0.10, or $\pm 57\%$. Myhre *et al.* (2013) estimate forcings from changes in surface albedo as -0.15 ± 0.10 and from tropospheric aerosols as -0.9 (–1.9 to –0.1). Recent data from Hansen *et al.* (2017) put the total close to -1.15, so we consider the uncertainty to be -1.15 ± 1.00, where $1.00 = 0.10 + (-0.1 - (-1.9))/2$, or $\pm 87\%$.

We assume that these percentages are roughly the same each year, as they are for WMGHGs and appear to be for volcanoes. In this way, we generate intervals for each forcing in each year to be consistent with the given data~~, in the spirit of Poppick *et al.* (2017) but without a bootstrap~~. This procedure inherently but realistically allows heteroskedasticity, because the uncertainty grows as the forcing's value grows. However, it does not allow for heteroskedasticity due to more precise measurements over time, ~~e.g.,~~ as noted by Myhre *et al.* (2013) for solar~~.~~, e.g. We further assume that the uncertainties in the forcings are purely idiosyncratic in the sense that they are neither time-dependent nor mutually dependent.

Approximating the distributions of these uncertainties by a Gaussian distribution centered at the reported value, we can estimate the variance by dividing the difference in the quantiles by 2×1.645 and squaring the result, which reverses the

formula $\pm 1.645\sqrt{\mathbf{var}(v_t)}$ to calculate 90% intervals from the variance of a mean-zero Gaussian random variable v_t. Doing so generates a 5×5 variance/covariance matrix for each time period, with the estimated variance of each of the five forcings along the main diagonal and zeros elsewhere.

The average of the diagonals, reflecting the variances of the uncertainty for each forcing (WMGHGs, ozone, aerosols, solar, and volcanic respectively), is estimated to be $(0.007,$ ~~$0.001, 0.284, 0.002, 0.012)'$~~ $0.001, 0.105, 0.002, 0.012)'$ over 1850-2016 and ~~$(0.029, 0.003, 0.969, 0.004, 0.000)'$~~ $(0.029, 0.002, 0.359, 0.004, 0.000)'$ over 1999-2013, the recent fifteen-year hiatus period. As expected, forcings from aerosols are estimated to be the most uncertain. The uncertainties over the hiatus period are generally larger than those over the whole sample, reflecting the larger magnitudes of the forcings near the end of the sample. In contrast, volcanic forcings are nearly zero during the hiatus period, reflecting the absence of a volcanic eruption with a major impact on global climate.

S.5.3 Contribution of Regressor Uncertainty to Coefficient Estimators

Uncertainty in the regressors may be treated as "classical measurement error" in the parlance of the econometrics literature, which is known to cause bias in the coefficient estimates. Because we observe forcings with error, we may denote our observation of forcings by $h_t = h_t^0 + v_t$, where $h_t^0 = (h_{1t}^0, h_{2t}^0)'$, with h_{1t}^0 and h_{2t}^0 denoting respectively non-volcanic and volcanic forcings if the forcings could be observed without uncertainty. Similarly, $v_t = (v_{1t}, v_{2t})'$ such that v_{1t} and v_{2t} are respectively the sum of uncertainties about non-volcanic forcings, estimated as described above, and uncertainty about volcanic forcing. v_t has a mean of zero and its components have variances given by $\sigma_{v1,t}^2$ and $\sigma_{v2,t}^2$. The former is the sum of the variances of the non-volcanic forcings, as the covariances are assumed to be zero.

In a cointegrating model like ours, bias in the long-run relationship is not hard to fix. In fact, although it was not designed to do so, the feasible Canonical Cointegrating Regression (CCR) methodology of Park *et al.* (2010) already takes into account this bias. Using a closely related model, Miller (2010, Theorem 2) shows the CCR estimator to be consistent, asymptotically normal, and asymptotically unbiased, with a variance that takes into account the measurement uncertainty.[8]

S.5.4 Contribution of Uncertainty to Forecasts

Explaining the contributions of the uncertainty to the missing heat of the 1998-2013 episode requires a measure of in-sample fit of T_t^a for some arbitrary time period $t = 0$, given by ~~$\hat{T}_0^a = h_0^{*\prime}\hat{\alpha} + x_0'\hat{\gamma} + w_0\delta$~~ $\hat{T}_0^a = h_0^{*\prime}\hat{\alpha} + x_0'\hat{\gamma} + w_0\hat{\delta}$. For simplicity, denote the right-hand side by $z_0'\hat{\pi}$ with ~~$z_t = (h_t^{*\prime}, x_t', w_t)' = (1, h_{1t}, h_{2t}, x_t', w_t)'$ and $\pi = (\alpha', \gamma', \delta)' = (\alpha_0, \alpha_1, \alpha_2, \gamma', \delta)'$ and let $z_t^0 = (1, h_{1t}^0, h_{2t}^0, x_t', w_t)'$~~ $z_t = (h_t^{*\prime}, x_t', w_t)' = (1, h_{1t}, h_{2t}, x_t', w_t)'$ and $\pi = (\alpha', \gamma', \delta)' = (\alpha_0, \alpha_1, \alpha_2, \gamma', \delta)'$ and let $z_t^0 = (1, h_{1t}^0, h_{2t}^0, x_t', w_t)$ and $\varpi_t = (0, v_{1t}, v_{2t}, 0, 0)$, such that $z_t = z_t^0 + \varpi_t$. The variance of the uncertainty in ~~$\hat{T}_0^a$~~ $\hat{T}_0^a$ is given by

$$\mathbf{var}(\hat{T}_0^a | z_0^0) = z_0^{0\prime}\mathbf{var}(\hat{\pi}|z_0^0)z_0^0 + \mathbf{var}(\varpi_0'\hat{\pi}|z_0^0) + 2z_0^{0\prime}\mathbf{cov}(\hat{\pi}, \varpi_0'\hat{\pi}|z_0^0)$$

[8] We do not model uncertainty in measuring temperatures~~, which is expected to be smaller than uncertainties in the forcings~~. ~~Also, we~~ Nor do ~~not~~ we model the effect of uncertainty in the volcanic forcings on the coefficient estimate, because the uncertainty is much smaller than the uncertainty for the other forcings – nearly zero – over the 1998-2013 period.

using this notation.

If we could observe z_0^0, a (Gaussian) 90% uncertainty interval for ~~$\hat{T}_0^a$ would be~~ $\hat{T}_0^a$ would be

$$z_0^{0\prime}\hat{\pi} \pm 1.645\sqrt{z_0^{0\prime}\mathbf{var}(\hat{\pi}|z_0^0)z_0^0}$$

from the first term. Instead, rewrite the variance as

$$\begin{aligned}\mathbf{var}(\hat{T}_0^a|z_0^0) &= z_0^{0\prime}\mathbf{var}(\hat{\pi}|z_0^0)z_0^0 + \mathbf{var}(\hat{\alpha}_1|z_0^0)\bar{\sigma}_{1v}^2 + \mathbf{var}(\hat{\alpha}_2|z_0^0)\bar{\sigma}_{2v}^2 \\ &+ \left[2z_0^{0\prime}\mathbf{cov}(\hat{\pi}, \varpi_0'\hat{\pi}|z_0^0)\right] + \left[\mathbf{var}(\varpi_0'\hat{\pi}|z_0^0) - (\mathbf{var}(\hat{\alpha}_1|z_0^0)\bar{\sigma}_{1v}^2 + \mathbf{var}(\hat{\alpha}_2|z_0^0)\bar{\sigma}_{2v}^2)\right]\end{aligned}$$

where $\bar{\sigma}_{1v}^2$ and $\bar{\sigma}_{2v}^2$ are temporal averages that estimate the variances $\sigma_{v1,t}^2$ and $\sigma_{v2,t}^2$ at $t = 0$. Specifically, we use the averages over 1999-2013 given above, so that $\bar{\sigma}_{2v}^2$ is effectively zero.

Ignoring the two terms in brackets and setting $\bar{\sigma}_{2v}^2 = 0$, a (Gaussian) 90% uncertainty interval for $\hat{T}_0^a$ given by

$$z_0'\hat{\pi} \pm 1.645\sqrt{z_0^{0\prime}\mathbf{var}(\hat{\pi}|z_0^0)z_0^0 + \mathbf{var}(\hat{\alpha}_1|z_0^0)\bar{\sigma}_{1v}^2}$$

takes into account uncertainty in the non-volcanic forcings. Because the uncertainty in the regressors is correlated with the uncertainty in the estimator, the bracketed terms are not zero, but we expect that they will be small.

As with predictions from any linear model, the variance of the out-of-sample conditional forecasts is augmented by the estimated variance of the error term η_t. In that case, we drop $\mathbf{var}(\hat{\alpha}_1|z_0^0)\bar{\sigma}_{1v}^2$, because we are conditioning on specific data and there are no measurement errors, and ~~because~~ we use least squares rather than CCR, in order to minimize mean squared forecast error.

S.6 Results with an Alternative Measure of Aerosol Forcing

As noted in the paper, the data on forcings from tropospheric aerosols are particularly uncertain and the data from Hansen *et al.* (2017) are intentionally smoothed. As an alternative data source, we consider the data on anthropogenic emissions of sulfur dioxide from Hoesly *et al.* (2018). Because sulfur dioxide has such a short residence time in the atmosphere, the natural log of emissions should be a reasonably good proxy for forcings within a year, in contrast to well-mixed greenhouse gases, the residence times of which require the natural log of concentrations cumulated from emissions in order to approximate forcings.

To convert the log of emissions into a measure in W/m^2, we regress the series of forcings from tropospheric aerosols and surface albedo of Hansen *et al.* (2017) onto an intercept and the log emissions series. The fitted prediction from this regression is the linear transformation of the log emissions data that best approximates the scale of the forcings from both tropospheric aerosols and surface albedo, and it may be expressed in W/m^2 as desired. Figure (S.5) shows a comparison of the two series, as well as the uncertainty interval calculated for the former series using $\pm 87\%$ as discussed above. The magnitude of the uncertainty clearly allows for either measure.

Table (S.4) shows our key results for the hiatus calculated with the two measures. All of the results in the paper are calculated using the data from Hansen *et al.* (2017), but we also report in the paper the contribution of tropospheric aerosols and surface

Contributions (°C) from:	
Hansen et al. (2017)	
Hoesly et al. (2018)	
Volcanoes	−0.003
(−0.001, −0.004)	−0.002
(−0.001, −0.004)	
TASA	−0.050
(−0.030, −0.070)	0.003
(−0.014, 0.020)	
Solar	0.016
(0.013, 0.018)	0.014
(0.012, 0.017)	
OMO	0.110
(0.079, 0.141)	0.083
(0.050, 0.117)	
SOI	−0.062
(−0.029, −0.095)	−0.063
(−0.030, −0.097)	
Model in (2) with $\gamma, \delta = 0$	−0.178
(−0.155, −0.202)	−0.108
(−0.085, −0.131)	
Model in (2) with $\gamma, \delta \neq 0$	−0.110
(−0.072, −0.147)	−0.068
(−0.023, −0.113)	

Table S.4. Estimated Contributions of Key Potential Explanations of the 1998-2013 Episode. Negative temperatures help to explain the missing heat, while positive temperatures exacerbate the puzzle. Tropospheric aerosols & surface albedo are abbreviated by TASA. 90% confidence intevals are shown in parentheses.

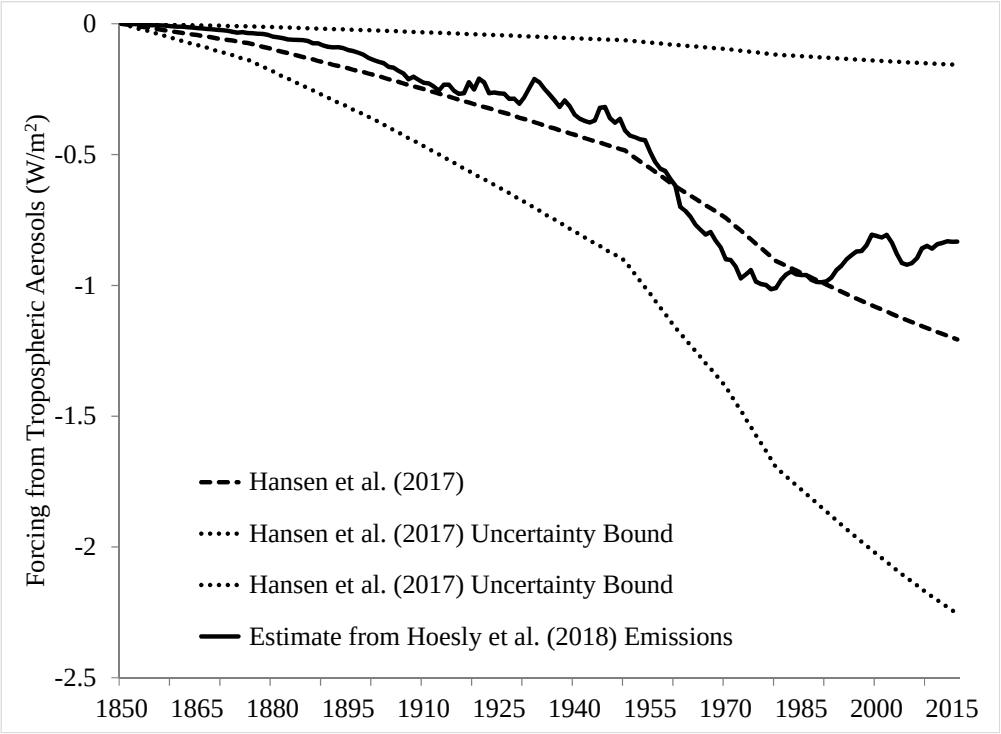

Figure S.5. Alternative Estimates of Forcings from Tropospheric Aeresols and Surface Albedo. Original forcings from Hansen *et al.* (2017) plotted with 90% uncertainty intervals of ±87% against estimated forcings from Hoesly *et al.* (2018).

albedo (TASA) using the data calculated from that of Hoesly *et al.* (2018). Because the alternative measure of TASA is aggregated with the other non-volcanic forcings in the model in (2), all of the estimates are influenced by the choice of TASA forcing measure. That of the contribution of TASA obviously is influenced the most. The difference in effect is so large that the intervals do not overlap and the direction of the effect of TASA in explaining the hiatus is effectively unknown – it may exacerbate rather than help to explain the hiatus.

A statistical mystery remains. The two measures are apparently within measurement error of each other in Figure (S.5), and stationary measurement error does not affect asymptotic inference about cointegration relationships. We have constructed our intervals to allow for stationary measurement error over 1998-2013, and we report cointegration diagnostics in footnote 1 of the paper. The diagnostics are not substantively different using the alternative measure, suggesting that the measurement error is in fact stationary and that both estimates are consistent. So, we should expect very similar estimates of α_1 in (2) and thus similar estimates of the contribution of TASA in Table (S.4). Yet they are not similar. Moreover, Figure (S.5) suggests that, in spite of the diagnostics, the difference between measures may not be stationary. We leave further investigation of this discrepancy for future research.

References

Alexeev, V.A., Langen, P.L., and Bates, J.R.: Polar amplification of surface warming on an aquaplanet in "ghost forcing" experiments without sea ice feedbacks, Clim Dynam, 24, 655-666, https://doi.org/10.1007/s00382-005-0018-3, 2005.

Bindoff, N.L., Stott, P.A., AchutaRao, K.M., Allen, M.R., Gillett, N., Gutzler, D., Hansingo, K., Hegerl, G., Hu, Y., Jain, S., Mokhov, I.I., Overland, J., Perlwitz, J., Sebbari, R., and Zhang, X.: Detection and attribution of climate change: From global to regional, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P.M., Cambridge University Press, Cambridge, 867-952, 2013.

Enfield, D.B., Mestas-Nunez, A.M., and Trimble, P.J.: The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S., Geophys Res Lett, 28, 2077-2080, https://doi.org/10.1029/2000GL012745, 2001.

Hansen, J., Sato, M., Kharecha, P., von Schuckmann, K., Beerling, D.J., Cao, J., Marcott, S., Masson-Delmotte, V., Prather, M.J., Rohling, E.J., Shakun, J., Smith, P., Lacis, A., Russell, G., and Ruedy, R.: Young people's burden: requirement of negative CO2 emissions, Earth Syst Dynam, 8, 577-616, https://doi.org/10.5194/esd-8-577-2017, 2017.

Hoesly, R.M., Smith, S.J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J.J., Vu, L., Andres, R J., Bolt, R.M., Bond, T.C., Dawidowski, L., Kholod, N., Kurokawa, J.-I., Li, M., Liu, L., Lu, Z., Moura, M.C.P., O'Rourke, P.R., and Zhang, Q.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emission Data System (CEDS), Geosci. Model Dev., 11, 369-408, https://doi.org/10.5194/gmd-11-369-2018, 2018.

Lenton, T.M., Dakos, V., Bathiany, S., and Scheffer, M.: Observed trends in the magnitude and persistence of monthly temperature variability, Sci Rep-UK, 7, 5940, https://doi.org/10.1038/s41598-017-06382-x, 2017.

Miller, J.I.: Cointegrating regressions with messy regressors and an application to mixed-frequency series, J Time Ser Anal, 31, 255-277, https://doi.org/10.1111/j.1467-9892.2010.00660.x, 2010.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., and Zhang, H.: Anthropogenic and natural radiative forcing., in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P.M., Cambridge University Press, Cambridge, 659-740, 2013.

Neely III, R.R., Toon, O.B., Solomon, S., Vernier, J.-P., Alvarez, C., English, J.M., Rosenlof, K.H., Mills, M.J., Bardeen, C.G., Daniel, J.S., and Thayer, J.P.: Recent anthropogenic increases in SO2 from Asia have minimal impact on stratospheric aerosol, Geophys Res Lett, 40, 999-1004, https://doi.org/10.1002/grl.50263, 2013.

North, G.R.: Theory of energy-balance climate models, J Atmos Sci, 32, 2033-2043, https://doi.org/10.1175/1520-0469(1975)032<2033:TOEBCM>2.0.CO;2, 1975.

North, G.R. and Cahalan, R.F.: Predictability in a solvable stochastic climate model, J Atmos Sci, 38, 504-513, https://doi.org/10.1175/1520-0469(1981)038<0504:PIASSC>2.0.CO;2, 1981.

North, G.R., Cahalan, R.F., and Coakley, Jr., J.A.: Energy balance climate models, Rev Geophys and Space Ge, 19, 91-121, https://doi.org/10.1029/RG019i001p00091, 1981.

Park, J.Y., Shin, K., and Whang, Y.J.: A semiparametric cointegrating regression: Investigating the effects of age distributions on consumption and saving, J Econometrics, 157, 165-178, https://doi.org/10.1016/j.jeconom.2009.10.032, 2010.

Phillips, P.C.B., and Solo, V.: Asymptotics for linear processes, Ann Stat, 20, 971-1001, https://doi.org/10.1214/aos/1176348666, 1992.

Poppick, A., Moyer, E.J., and Stein, M.L.: Estimating trends in the global mean temperature record, Advances in Statistical Climatology, Meteorology, and Oceanography, 3, 33-53, https://doi.org/10.5194/ascmo-3-33-2017, 2017.

Pretis, F.: Econometric Models of Climate Systems: The Equivalence of Two-Component Energy Balance Models and Cointegrated VARs, University of Oxford, Department of Economics Discussion Paper, Number 750, 2015.

Rose, B.E.J. and Marshall, J.: Ocean heat transport, sea ice, and multiple climate states: Insights from energy balance models, J Atmos Sci, 66, 2828-2843, https://doi.org/10.1175/2009JAS3039.1, 2009.

Stephens, G.L., O'Brien, D., Webster, P.J., Pilewski, P., Kato, S. and Li, J.: The albedo of Earth, Rev Geophys, 53, https://doi.org/10.1002/2014RG000449, 2015.

Stevens, B. and Schwartz, S.E.: Observing and modeling Earth's energy flows, Surv Geophys, 33, 779-816, https://doi.org/10.1007/s10712-012-9184-0, 2012.

Timmermann, A., Oberhuber, J., Bacher, A., Esch, M., Latif, M., and Roeckner, E.: Increased El Niño frequency in a climate model forced by future greenhouse warming, Nature, 398, 694-697, https://doi.org/10.1038/19505, 1999.

Trenberth, K.E. and Shea, D.J.: Atlantic hurricanes and natural variability in 2005, Geophys Res Lett, 33, L12704, https://doi.org/10.1029/2006GL026894, 2006.

Vernier, J.-P., Thomason, L.W., Pommereau, J.-P., Bourassa, A., Pelon, J., Garnier, A., Hauchecorne, A., Blanot, L., Trepte, C., Degenstein, D., and Vargas, F.: Major influence of tropical volcanic eruptions on the stratospheric aerosol layer during the last decade, Geophys Res Lett, 38, L12807, https://doi.org/10.1029/2011GL047563, 2011.

Wyatt, M.A. and Curry, J.A.: Role for Eurasian Arctic shelf sea ice in a secularly varying hemispheric climate signal during the 20th century, Clim Dynam, 42, 2763-2782, https://doi.org/10.1007/s00382-013-1950-2, 2014.