UNIVERSITY *of* York

This is a repository copy of *Pacific variability reconciles observed and modelled global mean temperature increase since 1950*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/168042/

Version: Published Version

Article:

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Pacific variability reconciles observed and modelled global mean temperature increase since 1950

Martin B. Stolpe[1] · **Kevin Cowtan**[2] · **Iselin Medhaug**[1] · **Reto Knutti**[1]

Abstract

Global mean temperature change simulated by climate models deviates from the observed temperature increase during decadal-scale periods in the past. In particular, warming during the 'global warming hiatus' in the early twenty-first century appears overestimated in CMIP5 and CMIP6 multi-model means. We examine the role of equatorial Pacific variability in these divergences since 1950 by comparing 18 studies that quantify the Pacific contribution to the 'hiatus' and earlier periods and by investigating the reasons for differing results. During the 'global warming hiatus' from 1992 to 2012, the estimated contributions differ by a factor of five, with multiple linear regression approaches generally indicating a smaller contribution of Pacific variability to global temperature than climate model experiments where the simulated tropical Pacific sea surface temperature (SST) or wind stress anomalies are nudged towards observations. These so-called pacemaker experiments suggest that the 'hiatus' is fully explained and possibly over-explained by Pacific variability. Most of the spread across the studies can be attributed to two factors: neglecting the forced signal in tropical Pacific SST, which is often the case in multiple regression studies but not in pacemaker experiments, underestimates the Pacific contribution to global temperature change by a factor of two during the 'hiatus'; the sensitivity with which the global temperature responds to Pacific variability varies by a factor of two between models on a decadal time scale, questioning the robustness of single model pacemaker experiments. Once we have accounted for these factors, the CMIP5 mean warming adjusted for Pacific variability reproduces the observed annual global mean temperature closely, with a correlation coefficient of 0.985 from 1950 to 2018. The CMIP6 ensemble performs less favourably but improves if the models with the highest transient climate response are omitted from the ensemble mean.

Keywords CMIP5 · CMIP6 · Internal variability · Decadal variability · Global mean temperature · ENSO

1 Introduction

The overall modelled global temperature increase of the past seven decades agrees well with what has been observed (Fig. 1a), but differences occurred on decadal to multidecadal time scales (Dai et al. 2015; Meehl et al. 2016). While the climate models participating in the Coupled Model Intercomparison Project phase 5 (CMIP5; Taylor et al. 2012) and

phase 6 (CMIP6; Eyring et al. 2016) capture the observed global mean surface temperature (GMST) trend from the 1940s/early 1950s to the mid-1970s reasonably well (Folland et al. 2018; Papalexiou et al. 2020), they tend to overestimate the warming from the mid-1970s to present-day (Tokarska et al. 2020). On shorter time scales, further differences become apparent (Fig. 1b). Whereas Earth appeared to warm faster than simulated by models from the 1970s to the 1990s, models overestimate the GMST increase during the so-called 'global warming hiatus' from the 1990s to the early twenty-first century (Medhaug et al. 2017). On decadal to multidecadal time scales, episodes of differences in the rate of modelled and observed global temperature change are related mostly to internal climate variability (Dai et al. 2015; Kosaka and Xie 2016), uncertainty in radiative forcing (Haustein et al. 2019; Marotzke and Forster 2015), and observational uncertainty (Karl et al. 2015).
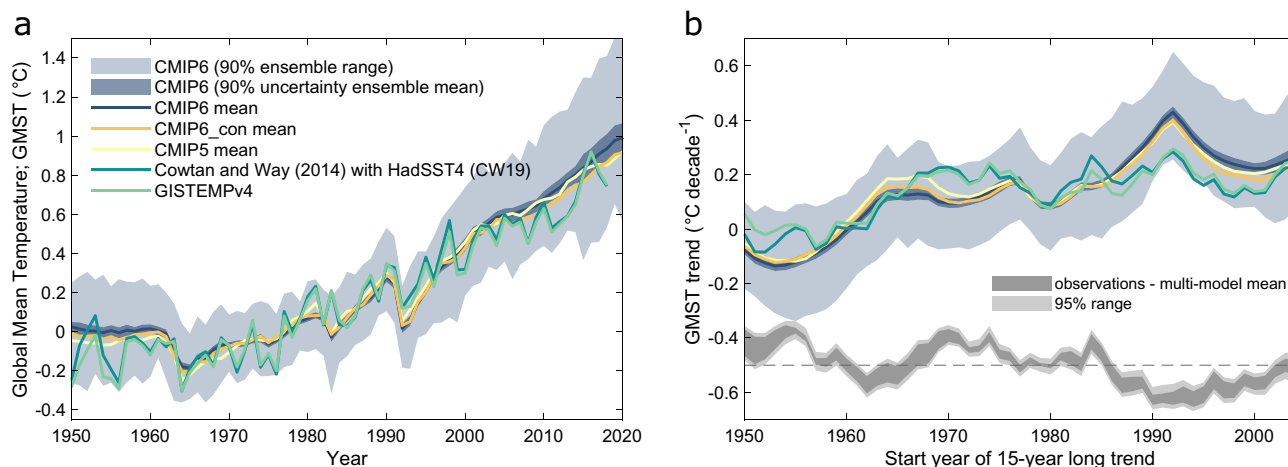
✉ Martin B. Stolpe
  martin.stolpe@env.ethz.ch

1 Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland

2 Department of Chemistry, University of York, York, UK

Fig. 1 **a** Observed and modelled annual global mean surface temperature (GMST) anomalies since 1950 (reference period is 1961–1990). For the 90% ensemble range (i.e., from 5 to 95%) we use one member per model, otherwise we first average the realizations of each model. To estimate the uncertainty of the ensemble mean we resample the CMIP6 models with replacement. The uncertainty for the other multi-model ensembles is similar, but not shown for clarity. CMIP6_con is a constrained multi-model ensemble that only includes CMIP6 models with a transient climate response of < 2.4 °C. **b** Running 15-year long trends in observed and modelled GMST and the difference between the two. The dark grey range indicates the minimum to maximum difference between the different GMST observational datasets and multi-model means, and in light grey the 95% range (i.e., from 2.5 to 97.5%) across the different combinations with resampled ensemble means is shown. The difference is shifted by − 0.5 °C per decade as indicated by the horizontal dashed line

The 'global warming hiatus', the most recent period of divergence between observed and modelled global warming, has been examined extensively and is explained by a combination of factors: under-representation of the fastest warming regions in the observational record (Cowtan and Way 2014); a biased comparison between observations and models, where sea surface temperature (SST) was assumed to warm at the same rate as marine air temperature (Cowtan et al. 2015); uncorrected biases in SST datasets related to changes in measurement instruments over time (Hausfather et al. 2017; Karl et al. 2015); mismatches in the radiative forcing, caused mainly by solar variability and omitted small volcanic eruptions (e.g., Folland et al. 2018; Huber and Knutti 2014; Ridley et al. 2014; Schmidt et al. 2018; Schmidt et al. 2014); and internal variability in the climate system (e.g., England et al. 2014; Kosaka and Xie 2013) causing the excess heat to be transported more efficiently away from the atmosphere to the deeper ocean (e.g., Meehl et al. 2011; Watanabe et al. 2013). Modelled warming during recent decades might also be too large if models overestimate the transient climate response (TCR) (Jiménez-de-la-Cuesta and Mauritsen 2019), but Huber and Knutti (2014) and Santer et al. (2017) found little evidence for a biased response of the CMIP5 ensemble to external forcing. Climate sensitivity has, however, increased from CMIP5 to CMIP6 (Meehl et al. 2020; Zelinka et al. 2020) and the CMIP6 ensemble shows greater warming than CMIP5 since the 1970s, which increases the discrepancy with observed warming during the 'global warming hiatus' (Fig. 1).

Internal variability in the Pacific is believed to have played a key role in decadal-scale differences between observed and modelled warming, and numerous studies quantified its contribution to the 'global warming hiatus' and earlier episodes. During the 'hiatus' period, Pacific trade winds strengthened, thereby intensified the Pacific shallow ocean overturning cells and increased upwelling of cooler waters in the central and eastern Pacific (England et al. 2014; Maher et al. 2018). This lowered the SSTs in the eastern Pacific and led to increased heat uptake into the subsurface western Pacific Ocean (England et al. 2014). Part of the heat has been transported by an enhanced Indonesian throughflow from the Pacific into the subsurface Indian Ocean, leading to a warming of the Indian Ocean (Lee et al. 2015). These processes reduced the pace of global warming during the 'hiatus' period relative to the ensemble mean of climate model simulations where internal variability is averaged out (Kosaka and Xie 2013).

Here, we aim to reconcile differing conclusions on the importance of Pacific variability during periods of accelerated and reduced rates of warming. While a consensus emerged that Pacific variability has contributed to the 'global warming hiatus' (Medhaug et al. 2017), estimates of how much differ greatly, and range from fully explaining it (e.g., Kosaka and Xie 2016; Peyser et al. 2016), to most of it (Swingedouw et al. 2017), around half (Huber and Knutti 2014), around a fourth (Meehl et al. 2016), to even less than that (Chylek et al. 2016; Tung and Chen 2018).

After having introduced the data and methods in Sect. 2, we review different approaches of quantifying the influence of Pacific variability on the global mean temperature in Sect. 3. In Sect. 4 we quantitatively compare the estimated Pacific contribution to global mean temperature during the past seven decades, but with a focus on the 'global warming hiatus', from 18 studies and examine which factors led to differing conclusions on the importance of Pacific variability. We demonstrate that the difference between observed and modelled warming is dominated by Pacific variability. Adjusting the multi-model ensemble means for the effect of Pacific variability therefore allows us to uncover biases in the simulated global mean forced response unrelated to Pacific variability. Whereas the variability-adjusted CMIP5 multi-model mean closely resembles observed warming, the CMIP6 ensemble mean appears to overestimate warming during recent decades.

2 Data and methods

We focus on global mean surface temperature (GMST hereafter) as one of the key metrics of global climate change, but note that it is an incomplete measure of the warming of the globe since most of the heat is stored in the oceans. The underlying assumptions throughout the paper are that the forced signal and internal variability are independent and approximately add up linearly. Under strong forcing, these assumptions may break down (Brown et al. 2017; Olonscheck and Notz 2017).

We restrict our analysis to the period from 1950 to 2018 as the tropical Pacific SSTs were only poorly sampled prior to that (Deser et al. 2010) and to avoid large SST biases during the World War II period and prior (Chan et al. 2019; Cowtan et al. 2018; Thompson et al. 2008). Nonetheless, the examined period covers most of the human-made climate change, as around three-quarters of the overall anthropogenic global warming took place after 1950 (Haustein et al. 2017).

Further, we limit the analysis to annual mean (January–December) and global mean values. Explaining differences between observed and simulated annual mean GMST does not guarantee that the individual seasons are explained as well (Deser et al. 2017; Douville et al. 2015) nor that the simulated temperature pattern agrees fully with the observed pattern (Deser et al. 2017; Xu et al. 2020).

2.1 Observational data

To quantify GMST changes, we use two spatially interpolated datasets, CW19 and GISTEMPv14. CW19 is an updated version of the Cowtan and Way (2014) dataset, which combines CRUTEM4.6 (Jones et al. 2012) over

land with HadSST4 (Kennedy et al. 2019) over the ocean and infills regions of missing observations by kriging. The initially published Cowtan and Way (2014) dataset used HadSST3, and to differentiate the two versions we refer to it as CW19 in the following. GISTEMPv4 (Hansen et al. 2010; Lenssen et al. 2019) represents SSTs by ERSSTv5 (Huang et al. 2017a) and land air temperatures are based on NOAA GHCNv4 (Menne et al. 2018). GISTEMPv4 extrapolates temperature anomalies between stations which are up to 1200 km apart and thereby achieves nearly complete spatial coverage after 1950. Because measurement coverage since 1950 is relatively high, statistical infilling is able to alleviate biases in global warming arising from poor coverage (Benestad et al. 2019), such as underestimated warming during the 'global warming hiatus' (Cowtan and Way 2014; Huang et al. 2017b). We compare GISTEMPv4 and CW19 with other spatially interpolated GMST datasets in the supporting information (Fig. S1). While there are some differences between the datasets, our results do not depend strongly on the choice of GMST dataset.

To characterize Pacific variability, we use indices based on tropical SST and wind stress. We calculate indices of oceanic variability from two SST datasets, COBE-SST2 (Hirahara et al. 2014) and ERSSTv5. Both datasets are spatially interpolated. Among other SST datasets, ERSSTv5 and COBE-SST2 are most consistent with independent ocean profile data used for evaluating the datasets after 1950 (Huang et al. 2018), and are not affected by a cooling bias in recent years seen in other SST datasets (Hausfather et al. 2017). We represent the El Niño-Southern Oscillation (ENSO) by the observed monthly-mean equatorial Pacific SSTs in the Nino3.4 region (5° S–5 °N, 170°–120° W; Trenberth 1997). The simulated ENSO variability might be spatially displaced compared to the observed variability, and therefore we also use a larger region in the tropical, central to eastern Pacific, typical of what has been used in tropical Pacific pacemaker experiments (15° N–15° S, 180°–90° W; hereafter the pacemaker region).

We quantify tropical Pacific zonal mean eastward wind stress variability using four modern reanalyses, ERA5 (Hersbach and Dee 2016; covering the period 1979–2018), JRA-55 (Kobayashi et al. 2015; 1958–2018), MERRA2 (Gelaro et al. 2017; 1980–2018), and the NOAA-CIRES 20th Century Reanalysis (20CR) v2c (Compo et al. 2011; 1851–2014). ERA5, JRA-55 and MERRA2 assimilate different observation types and variables whereas 20CR only assimilates surface pressure observations and may, therefore, be less susceptible to changes in the observational system, but is also less tightly constrained by observations. For calculating the zonal mean wind stress variability, we use 180°–150° W and 6° S–6° N, a region where England et al. (2014) found maximum regression between the observed Interdecadal Pacific Oscillation (IPO) and

wind stress variability. Again, we also use a larger region, 150° E–150° W and 10° S–10° N, as in Saenko et al. (2016), to assess the robustness of our results. In the following, we refer to these as the England and Saenko regions. All regions are displayed in Fig. 2.

For the GMST and SST observations we use 1961–1990 as the reference period, whereas we compute anomalies with respect to 1982–2010 for the wind stress reanalyses fields.

2.2 Model data

To estimate unforced climate variability, we use pre-industrial control (piControl) simulations from 33 CMIP5 models and 35 CMIP6 models consisting of 18,797 and 21,740 simulated years, respectively. The models used are listed in Table S1 in the supporting information. We linearly detrend the control simulations to remove residual model drift.

We compare the observed warming against the means of CMIP5 and CMIP6 models. For the CMIP5 ensemble, 86 historical members from 38 models are available, and we extend these with the RCP8.5 scenario from 2006 onward (the choice of the scenario, however, does not make a large difference for 2006–2018; England et al. 2015). The CMIP6 ensemble consists of 47 models with 347 realizations. Besides their improved model physics, and a generally higher resolution, the CMIP6 models are forced with updated external forcings. From 2015 onward we use simulations under SSP5-8.5 forcing for which a subset of 34

models provides simulations. We compute the historical multi-model means by first averaging the members of each model to then average across the models (including only the first member of each model results in a very similar multi-model average). The uncertainty around the multi-model ensemble means is quantified by resampling the ensembles with replacement.

When comparing simulated to observed warming, we blend absolute SST over oceans with absolute near-surface air temperature over land and sea ice as in observations (procedure as introduced by Cowtan et al. 2015) for the CMIP5 models. Clarke and Richardson (2020) show that the blending bias expressed as the ratio of ΔGSAT, the global mean air temperature change, to ΔGMST, the blended temperature change, is similar for the CMIP5 and CMIP6 ensembles. Therefore, we approximate the CMIP6 ΔGMST as:

$$\Delta\text{GMST}_{\text{CMIP6}} = \Delta\text{GSAT}_{\text{CMIP6}} + (\Delta\text{GMST}_{\text{CMIP5}} - \Delta\text{GSAT}_{\text{CMIP5}})$$
$$+ \ \beta * (\Delta\text{GSAT}_{\text{CMIP6}} - \Delta\text{GSAT}_{\text{CMIP5}}).$$

The last term on the right-hand side accounts for the difference in warming between the two ensembles. β is the ratio of (ΔGMST – ΔGSAT)/ΔGSAT, and we estimate it to be – 0.048 during 1900–2018 for the CMIP5 mean, consistent with other assessments (Richardson et al. 2018). For ERSSTv5, night-time marine air temperature observations are used to correct biases in SST measurements, but it is not obvious how this influences the differential warming

Fig. 2 The four regions in the tropical Pacific used in this study to obtain indices of internal variability: The Nino3.4 and pacemaker regions for SST variability, and the England and Saenko regions for wind stress variability
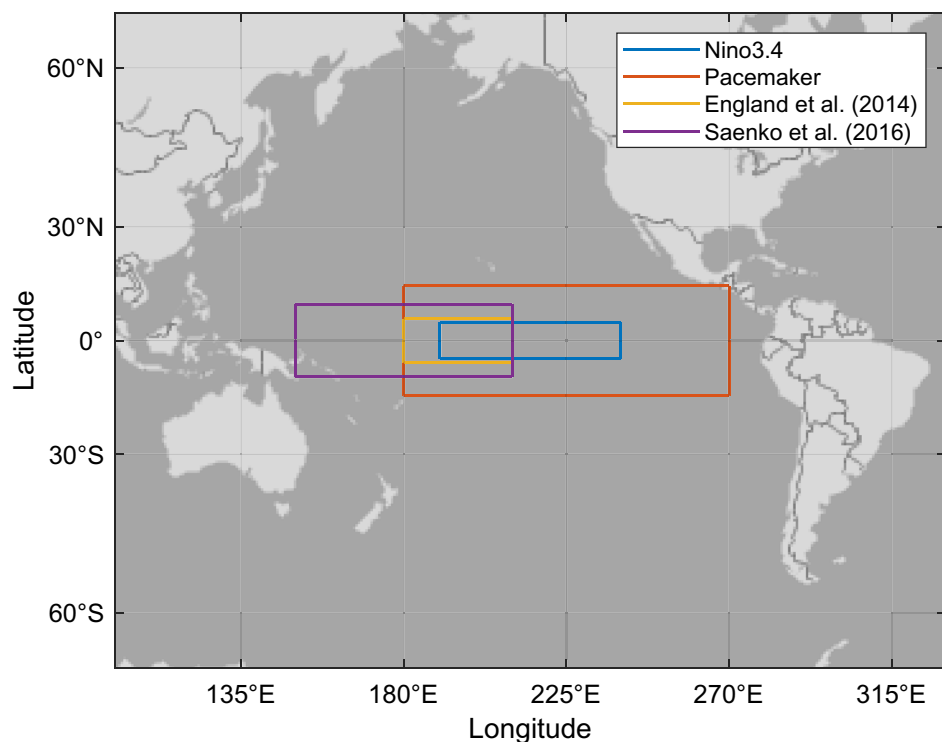
between SSTs and marine air temperatures (Cowtan et al. 2015), and we therefore also compare GISTEMPv4 against blended model output.

For the historical simulations, we use the same reference periods as for the observations and assume that biases in the simulated climatological GMST project only weakly onto the simulated climate change (Hawkins and Sutton 2016; Stolpe et al. 2019).

2.3 Method

We compare published studies that estimate the influence of tropical Pacific variability on global temperature using various methods (see Sect. 3). We restrict this comparison to studies that use a measure of Pacific variability, such as SST or surface wind variability, and relate this to GMST, and exclude studies that decompose GMST statistically into different components and then relate these to spatial patterns and modes of variability (e.g., Chen and Tung 2018; Dai and Wang 2018; Wei et al. 2019). Further we only consider studies that use a measure of tropical Pacific variability and omit studies that solely use other measures of Pacific internal variability [e.g., the IPO as in Meehl et al. (2016) or Power et al. (2017) or the Pacific Decadal Oscillation, PDO, as in Steinman et al. (2015)] to remove one source of potential spread between estimates. This leaves us with 18 studies that we use for the comparison—which we argue is a representative sample of the literature. All studies are summarized in Table S2.

Most of the data is provided by the corresponding authors (see acknowledgments) and for a few studies we extracted the data from figures in the individual papers. The data of Peyser et al. (2016) was not available to us, and we therefore computed it following their methodology (see supporting information). Within CMIP6, two projects include experiments dedicated to quantifying the role of time-evolving equatorial Pacific variability to regional and global temperature variability: The Decadal Climate Prediction Project (Boer et al. 2016; DCPP; "dcppC-pac-pacemaker" experiment) and the Global Monsoons Model Intercomparison Project (Zhou et al. 2016; GMMIP, "hist-resIPO" experiment). Data from one model, IPSL-CM6A-LR, following the dcppC-pac-pacemaker experiment was available and is also included in our analysis. At the time of writing, three models performed the hist-resIPO experiment, but each model only provided very small ensemble sizes (Table S2) and therefore the data is not included in our analysis.

To understand differences between the studies, and to test assumptions made when relating Pacific variability and GMST, we use variability analogues (Huber and Knutti 2014; Stolpe et al. 2017). Variability analogues are periods from piControl simulations for which the simulated, unforced variability matches the observed Pacific variability.

The mean over the selected analogues in another variable, e.g., in GMST, then provides an estimate of the influence of Pacific variability on the examined quantity.

We measure tropical Pacific variability based both on SST and wind stress from the regions shown in Fig. 2 and standardize the observed and modelled time series of Pacific variability. This allows a better comparison between time series with different amounts of variability but has only a small impact on our results. Starting from January 1950 we select the 20 variability analogues from the more than 40,000 piControl years that have the smallest root-mean-square error (RMSE) with the observed time series of Pacific variability over a period of 40 months. We shift by 1 month and again select the best matching analogues starting from February 1950 and repeat this until we sampled the whole observational period. Then we average all the overlapping global near-surface air temperature fields of these variability analogues for every month and compute global mean time series. We assume that the relationship between Pacific variability and GMST is the same under present-day and pre-industrial conditions.

3 Comparison of methods to quantify the Pacific imprint on GMST

Methods to quantify the tropical Pacific influence on GMST can be broadly categorized into three groups. Regression-based approaches, pacemaker experiments, and freely evolving climate model simulations. In Table S2 we list key features of studies based on these methods.

In pacemaker experiments (Boer et al. 2016; Deser et al. 2017; Douville et al. 2015; Kosaka and Xie 2013, 2016; Swingedouw et al. 2017; Zhou et al. 2016), modelled SSTs in the central to eastern equatorial Pacific are nudged towards observed SST anomalies. The influence of Pacific variability is then estimated as the difference in GMST evolution between two ensembles: A freely evolving initial condition ensemble with historical changes in external radiative forcing, and an ensemble with nudged tropical SSTs, but the same historical forcing. Kosaka and Xie (2016) argue that the obtained estimate of GMST variability induced by the Pacific is independent of the climate model's radiative forcing and climate sensitivity, and Douville et al. (2015) propose that it depends instead on the model-specific relationship between equatorial Pacific variability and GMST and in particular how strongly the equatorial variability projects onto high latitudes. Wang et al. (2017) similarly find that the tropical impact on high-latitude air temperature varies strongly between models on a decadal time scale and Deser et al. (2017) argue that atmospheric teleconnections forced by the tropical Pacific, which are not captured by all models, are of relevance for Eurasian and North American

boreal winter temperature and thereby also GMST. The link between equatorial Pacific SST and global temperature has been shown to be overestimated in some models (e.g., in CanESM2, see Saenko et al. 2016). The SST-pacemaker approach creates an artificial heat sink in the Pacific which could bias, if large, the estimated Pacific contribution to global temperature variability (Tung and Chen 2018), but Kosaka and Xie (2016) argue this is not a major concern.

The key differences between the pacemaker studies are the climate models used, the number of ensemble members, which determines how much variability originating from outside the Pacific is still present, the SST dataset, and the restoring time scale. A larger restoring time scale allows the model to evolve more freely, and results in a more physical higher-frequency ocean-atmosphere interaction (Swingedouw et al. 2017). The larger restoring time scale has also been recommended for the CMIP6 DCPP pacemaker experiments (Boer et al. 2016).

To remedy the issue of artificial heat uptake, England et al. (2014), Delworth et al. (2015), Douville et al. (2015), Gastineau et al. (2019), Oka and Watanabe (2017), Svendsen et al. (2018), and Watanabe et al. (2014) prescribe wind stress trends instead of SST. This is done either over tropical oceans (Watanabe et al. 2014) or only over the tropical Pacific. The wind nudging experimental design comes with the disadvantage of larger uncertainties in wind stress from reanalyses compared to SST reconstructions. The agreement between reanalyses in decadal tropical Pacific wind stress trends is relatively poor in the early 20th century, but improves over time (Kajtar et al. 2018). Further, the SST evolution in the tropical Pacific is not constrained. While year-to-year variability in tropical Pacific SST is usually well-matched when wind fields are prescribed (Douville et al. 2015; Gastineau et al. 2019), decadal trends have been shown to be biased in some models (e.g., Gastineau et al. 2019), which then might lead to a biased GMST influence. Most wind-stress studies use ERA-interim, for which the robustness of tropical Pacific wind trends has been evaluated against several observational datasets (de Boisséson et al. 2014), but it is limited to the period after 1979. Watanabe et al. (2014), Oka and Watanabe (2017), and Svendsen et al. (2018) use JRA-55 and 20CR which cover a longer period.

In regression-based approaches (Folland et al. 2018; Foster and Rahmstorf 2011; Hu and Fedorov 2017; Johansson et al. 2015; Lean 2018; Rypdal 2018; Saenko et al. 2016; Schmidt et al. 2014) the observed GMST, or its rate of change, is modelled by a number of predictors that are linearly combined. These studies differ in which predictors they include, whether the predictors are filtered, for example with a long-memory response (Rypdal 2018), an e-folding response profile (Folland et al. 2018), or by using a mixed-layer model (e.g., Thompson et al. 2008), and the temporal resolution of the predictors, i.e., monthly or annual. The

regression models typically consist of anthropogenic forcing, solar variability, stratospheric aerosol optical thickness as a measure of volcanic activity, and a term to describe tropical Pacific variability. Some studies include further modes of internal variability, such as the Atlantic Multidecadal Variability (AMV; e.g., Chylek et al. 2016; Folland et al. 2018; Rypdal 2018), or the Arctic Oscillation (AO; e.g., Folland et al. 2018), which can make a difference if predictors are correlated. Contrary to other regression analyses, Chylek et al. (2014) propose a minimal regression model with only greenhouse gas forcing and the AMV as predictors of annual GMST, but Rypdal (2018) argues that including the AMV at annual resolution misattributes GMST variability caused by ENSO variability and volcanic forcing to the AMV. This is because ENSO variability and volcanic forcing also influence North Atlantic SST, and accordingly the predictors are correlated. Regarding the Pacific variability term, regression-based studies often use the Nino3.4 index or the Multivariate ENSO index (MEI; Wolter and Timlin 2011). The latter is a more comprehensive way of characterizing ENSO variability than a purely oceanic, SST based index as it combines environmental variables from both the atmosphere and the ocean. Exceptions are Saenko et al. (2016) who instead use western tropical Pacific zonal wind stress as a predictor and Peyser et al. (2016) who relate the east-minus-west difference in tropical Pacific sea surface height to GMST variability.

A general concern of regression models is that the effect of predictors on GMST is not physically constrained and hence much of the variance is explained by construction, although a connection might not exist in reality. The more predictors are included, the larger the risks of overfitting. Kosaka and Xie (2016), Peyser et al. (2016), and Wang et al. (2017) argue that models simulate a GMST response to tropical Pacific variability that is around 20–45% larger on a decadal time scale than on an interannual time scale. Regression models are constructed on observations that are dominated by the latter and therefore might underestimate the amplitude of Pacific variability on GMST on a decadal time scale (Wang et al. 2017). Whereas Kajtar et al. (2019) demonstrate that the sensitivity of decadal GMST variability on the IPO is similar in the CMIP5 multi-model mean and observations, Haustein et al. (2019) reconcile observed and modelled warming during the twentieth century without a time scale dependent GMST response on tropical Pacific variability.

The third class of methods uses readily available climate model simulations to establish the effect of Pacific variability on GMST. This can be done either by searching for variability analogues from piControl simulations (Huber and Knutti 2014), regressing GMST on indices of Pacific variability in control (Peyser et al. 2016; Wang et al. 2017) or historical simulations (Kajtar et al. 2019), or by only including historical model simulations that are in the right ENSO phase to the ensemble mean (Risbey et al. 2014). The latter approach gives an ensemble mean that is closer to the observed warming

(Risbey et al. 2014). The adjusted ensemble mean is, however, based on fewer models and may be less robust. A strength of the variability analogues method is that it retains the physical connection across variables, and does not induce an artificial heat sink. A limitation, however, is the sparsity of piControl simulations. With increasing length of the analogues, i.e., analogues that track the observed variability over a longer period, it becomes harder to find suitable analogues, and accordingly the Pacific SST trend is underestimated (Fig. S2 in the supporting information). Therefore, we here search for relatively short analogues of 40 months length to ensure that the selected analogues follow the observed variability closely. Also, the number of analogues selected at each time step is somewhat arbitrary, but the results are only weakly dependent on this threshold (Fig. S3). Further, the simulated relationship between Pacific and GMST variability could be biased and is not constrained by observations. This bias is, however, as we show later, likely substantially lower in multi-model analyses than in pacemaker experiments, which usually rely on a single climate model.

4 Results

4.1 Simulated and observed global mean temperature change

We start by examining the long-term GMST increase simulated by the CMIP5 and CMIP6 ensemble means

and whether there is evidence for a mismatch with the observed warming. Such a discrepancy might indicate that internal variability alone is insufficient in reconciling observed and modelled warming, but that there is a bias in radiative forcing or in how sensitive the climate responds to the imposed forcing.

The 1950–2018 long-term warming of both the CMIP5 and CMIP6 ensemble agrees well with what has been observed, but the CMIP6 ensemble mean shows greater warming than CMIP5 from around 1970 onward for trends ending in 2018. CMIP6 ensemble mean trends starting from 1975 and later even fall outside the observational 2σ envelope, although some bootstrapped samples of the ensemble mean remain within the observational uncertainty (Fig. 3a). This enhanced warming of the CMIP6 ensemble during the last decades is likely caused by its 11% higher ensemble mean TCR (2.03 °C compared to 1.83 °C for CMIP5; Table S1. Note that TCR estimates are not available for all models) (Flynn and Mauritsen 2020; Nijsse et al. 2020; Tokarska et al. 2020).

Several recent studies observationally constrained TCR to central values from 1.3 to 2.05 °C, with upper 95th percentiles of the constrained distributions in the range 1.9–2.4 °C (Fig. 3b). 20% of the CMIP6 models (9 out of 45 models with available TCR estimate; Table S1), but only one CMIP5 model, simulate a TCR that exceeds 2.4 °C, which is more than we would expect if the models were random samples of the observationally constrained TCR distributions,
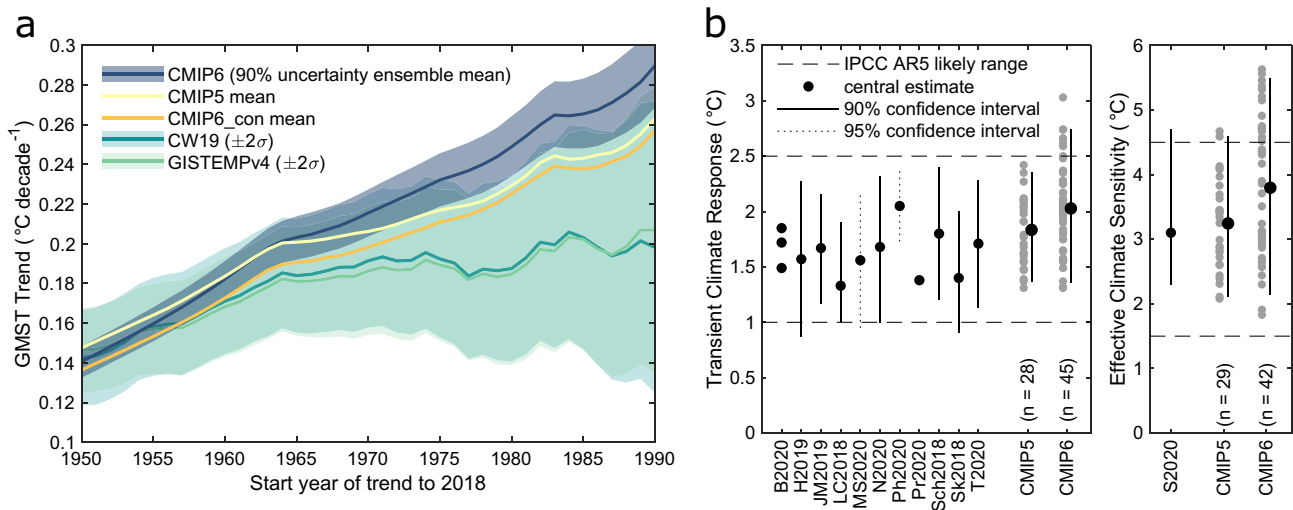
Fig. 3 **a** Trends in observed and modelled GMST with different start years that all end in year 2018. For the observational uncertainty we treat the monthly residual from 1950 to 2018 as an ARMA(1,1) process following the approach of Foster and Rahmstorf (2011). CMIP6 uncertainty is estimated from resampling the ensemble mean. **b** Transient climate response (TCR) simulated by the CMIP5 and CMIP6 models (Table S1) compared to observationally constrained estimates from B2020 (Bruns et al. 2020), H2019 (Haustein et al. 2019),

JM2019 (Jiménez-de-la-Cuesta and Mauritsen 2019), LC2018 (Lewis and Curry 2018), MS2020 (Montamat and Stock 2020), N2020 (Nijsse et al. 2020), Ph2020 (Phillips et al. 2020), Pr2020 (Pretis 2020), Sch2018 (Schurer et al. 2018), Sk2018 (Skeie et al. 2018), and T2020 (Tokarska et al. 2020). The estimate of T2020 is based on the observed 1981–2017 temperature increase. Simulated effective climate sensitivity (ECS) is compared to the baseline assessment of S2020 (Sherwood et al. 2020)

suggesting that some CMIP6 models overestimate the response to external forcing. We therefore also consider a constrained CMIP6 ensemble, CMIP6_con, for which we conservatively only include models with a TCR of less than 2.4 °C. CMIP6_con consists of 27 models (Table S1) with a mean TCR of 1.83 °C, virtually identical to that of the CMIP5 ensemble, and expectedly it simulates temperature trends from 1965 and onward which are closer to the observed trends (Fig. 3a).

Despite the overall agreement in long-term warming, and the potentially overestimated warming from CMIP6 to which we come back in Sect. 6, differences remain on decadal time scales (Fig. 1b) and we therefore examine the contribution of the tropical Pacific to these. We start with a general overview of the Pacific contribution to GMST since 1950 and then examine the 'global warming hiatus' in detail. We first present results of the published literature to then examine why studies come to differing conclusions on the GMST imprint of Pacific variability.

4.2 Study intercomparison: Pacific influence on global mean temperature

4.2.1 Overview: Pacific contribution 1950 to present-day

Across the assessed studies, there is broad agreement that tropical Pacific variability has contributed to the difference between observed and modelled warming since 1950. For the period from 1960 to 2012, the estimated tropical Pacific influence on GMST (in °C) from the examined studies is positively correlated with the difference between observed and modelled annual mean GMST (mean $r = 0.69$, with a range of 0.52–0.85; see Fig. S4 for details). After smoothing the time series with a running 5-year mean to emphasize lower frequency variability, the average correlation with the residual remains similar, $r = 0.72$, but the spread among the studies increases ($r = 0.35$–0.93) (Fig. S4).

Pacific variability may therefore explain a significant part of the difference between observed and modelled GMST

Fig. 4 Difference between running 15-year long trends in observed and modelled GMST. The dark grey range indicates the minimum to maximum difference between the different observational GMST datasets and multi-model means, and in light grey the 95% range across the different combinations with resampled ensemble means is shown. Running GMST trends for **a** studies using a multiple linear regression approach and **b** for SST and wind stress pacemaker experiments. The estimates of England et al. (2014), Huber and Knutti (2014), Peyser et al. (2016), Schmidt et al. (2014), and Swingedouw et al. (2017) only cover a short period and are therefore not included here. Figure S6 in the supporting information shows running 10- and 20-year long trends
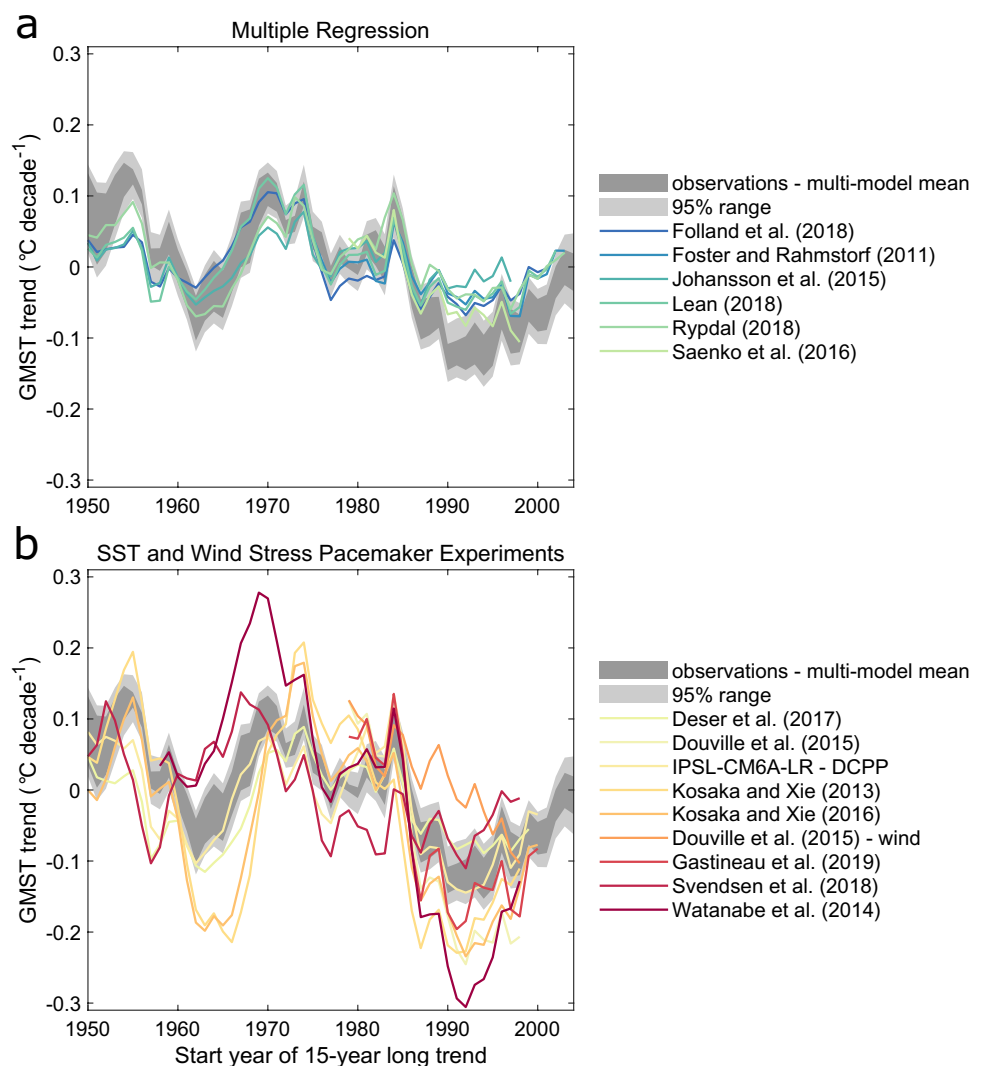
increase on different time scales, but there is a large spread between the studies. Whereas the multiple linear regression studies indicate a consistent Pacific imprint on running 15-year long GMST trends (Fig. 4a), the variability across the SST and wind stress nudging experiments is considerably larger (Fig. 4b). The regression-based Pacific variability estimates reconcile observed and modelled warming during most of the examined period, except for temperature trends starting in the 1950s and during the 'global warming hiatus' although there is similarity in the temporal evolution during these periods. The pacemaker experiments display a larger spread and greater amplitude of GMST trends related to Pacific variability than the multiple regression estimates. Compared to the difference between observed and modelled GMST changes, several pacemaker experiments appear to overestimate the Pacific influence on GMST, but also suggest that potentially the complete discrepancy between observed and modelled running 15-year trends can be attributed to Pacific variability. These findings are similar when we instead examine running 10-, and 20-year long trends (Fig. S6).

### 4.2.2 Pacific contribution to the 'global warming hiatus'

Next, we exemplarily examine the Pacific contribution to the 'global warming hiatus'. This period is not only covered by

all the studies, shows particularly large trends in the Pacific, but also the spread between the different studies is large. We use two different periods to define the 'hiatus', 1997–2012 and 1992–2012. The latter period is motivated by the onset of the acceleration in Pacific trade winds (England et al. 2014). The 'hiatus' ended with the years 2014–2016, when a strong El Niño released large amounts of heat from the north-western tropical Pacific (Yin et al. 2018), but we use 2012 as the end-year because most studies in our intercomparison only include data until then (cf. Table S2).

During 1997–2012, the observed warming was between −0.21 and −0.11 °C smaller than the modelled temperature increase of the multi-model ensemble means when calculated as the linear trend multiplied by its length (Fig. 5a). For the years from 1992 to 2012, this difference ranges from −0.28 to −0.19 °C (Fig. 5b). For both periods GISTEMPv4 shows slightly greater warming than CW19, but the uncertainty in modelled warming is larger than observational uncertainty. CMIP6 shows greater warming than CMIP5 and CMIP6_con and accordingly the largest divergence from observed GMST.

All inspected studies agree that the tropical Pacific acted to lower GMST trends during the early twenty-first century. The spread among the studies is, however, substantial and reaches a factor of more than five for 1992–2012 (Fig. 5b).

Most of the SST-based pacemaker studies indicate that the model—observation difference during both periods can
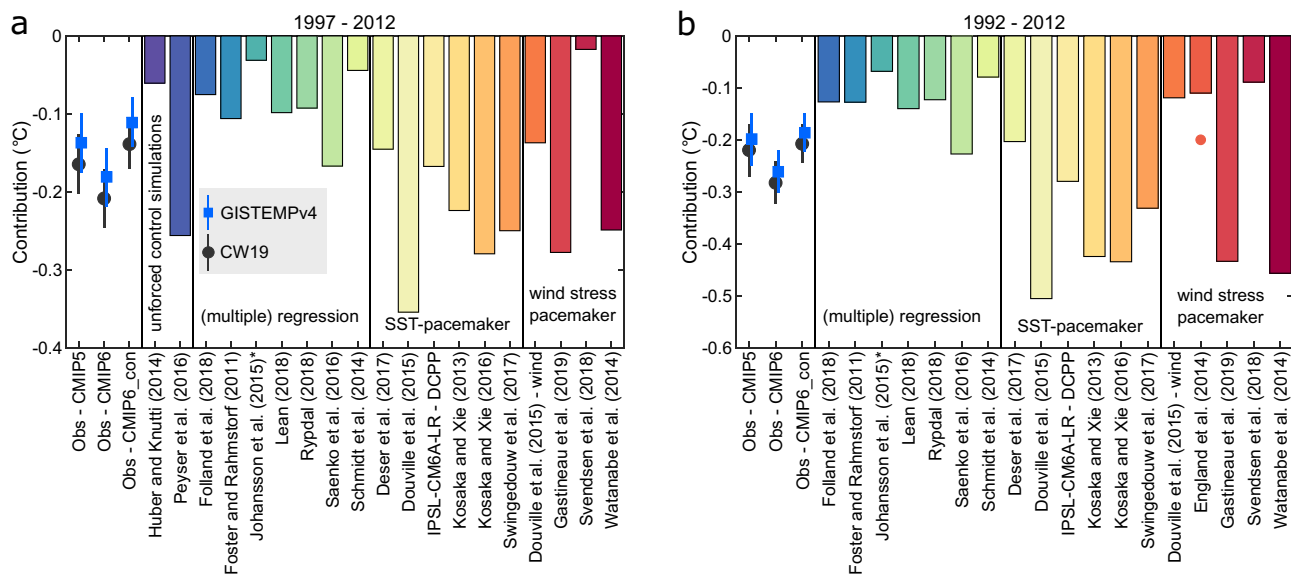
Fig. 5 The contribution of Pacific variability to GMST as estimated by different studies during the period from **a** 1997 to 2012 and **b** 1992 to 2012 (only best estimates are shown), calculated as the least squares linear trend multiplied by 16 years and 21 years, respectively (bars). The time series of Johansson et al. (2015) ends in 2011 and is therefore marked by an asterisk. The lower and upper

(orange dot) estimates of England et al. (2014) come from intermediate and full complexity climate models, respectively. Note that we implemented the method of Peyser et al. (2016) as they examine a different period (details in supporting information). For the difference between observed and modelled warming, we show the 90% range across the resampled ensemble means (whiskers)

be fully attributed to Pacific variability even under the high CMIP6 warming. The spread among the pacemaker studies is large, and several experiments find a contribution that is significantly larger than the actual model—observation difference. The experiments of Douville et al. (2015) show a cooling that is more than twice as large as that of the pacemaker experiment of Deser et al. (2017), although they mention that their model simulates a "weak but realistic ENSO-GMST relationship" on an annual time scale. On average the SST-pacemaker experiments simulate a cooling of 0.36 °C during 1992–2012, which is larger than the actual difference of 0.19–0.28 °C between observed and simulated warming.

Experiments where wind stress is overridden by reanalysis values span an even wider range. While Douville et al. (2015) find a contribution of −0.12 °C for 1992–2012, the experiments of Watanabe et al. (2014) and Gastineau et al. (2019) show about four times larger influences on GMST. Douville et al. (2015) argue that MIROC5, the model used by Watanabe et al. (2014), has an unrealistically strong correlation between Pacific SST and GMST. In the wind stress experiment of Gastineau et al. (2019) the simulated cooling in the Nino3.4 region is more than twice as large as observed. This may be one reason why they find a larger Pacific contribution to the 'hiatus' than Swingedouw et al. (2017) who nudge eastern tropical SSTs in the same model. For the examined wind stress experiments, the global temperature decrease related to Pacific variability is on average 0.25 °C, which is sufficient to fully explain the observation—model discrepancy from 1992 to 2012. Among the wind stress pacemaker studies, the experiment of Svendsen et al. (2018) shows the smallest Pacific contribution during both periods. The use of 20CR, which shows a somewhat smaller intensification in Pacific trade winds than the other reanalyses during the 'global warming hiatus', or the relatively small ensemble size (Table S2) resulting in some internal variability from outside the Pacific still being present, might contribute to this.

The regression-based studies and variability analogues of Huber and Knutti (2014) show a smaller Pacific influence during the 'global warming hiatus', and these studies typically find additional contributions to the 'hiatus' from solar variability (e.g., Folland et al. 2018; Huber and Knutti 2014; Lean 2018) and stratospheric aerosols (Huber and Knutti 2014). Apart from the study of Saenko et al. (2016), which uses wind stress as a predictor and finds a considerably larger Pacific contribution than the other multiple regression studies, the Pacific induced cooling is smaller than in any SST-based pacemaker experiment. The GMST cooling estimated by multiple linear regression is on average 0.12 °C, or only about half of the difference between observed and modelled warming during 1992–2012 (estimates of Saenko et al. (2016) and Johansson et al. (2015) not included here). Adjusting the regression-based estimate upwards by about

40%, which is how much stronger GMST responds to tropical Pacific variability on average on a decadal than an interannual time scale according to CMIP5 models (Wang et al. 2017), still results in an estimate less than half of that of the SST-pacemaker experiments. This suggests that there are further fundamental differences between the methods. The regression of modelled GMST onto the east to west tropical Pacific sea surface height gradient (Peyser et al. 2016), results in a significantly larger suppression of the global mean temperature increase from 1997 to 2012, roughly in line with SST-based pacemaker experiment. Due to the short observational record it is difficult to assess how realistically models simulate this relationship. The regression coefficient is, however, similar across CMIP5 and CMIP6 models (supporting information).

In the following, we examine the differences between multiple linear regression and SST-based pacemaker experiments by inspecting the roles of anthropogenic forcings on tropical Pacific SST (Sect. 4.3) and of model uncertainty (Sect. 4.4). Then, we assess whether tropical Pacific wind variability leads to consistent results, and whether or not the differences between wind pacemaker experiments can be explained by observational uncertainty (Sect. 4.5).

4.3 Influence of forced trend in Pacific SST

Several studies assume there is no or a negligible forced signal in the tropical Pacific SSTs and directly use the area-mean SST as a predictor for GMST (cf. Table S2). Sometimes a long-term linear trend is removed (e.g., Hu and Fedorov 2017), and while this removes some external forcing, it probably does not remove all (e.g., Mann et al. 2014). If anthropogenic forcing is not fully accounted for, the tropical Pacific cooling due to internal variability during the 'global warming hiatus' period is underestimated by attributing it to external forcing. The MEI also shows a long-term positive trend, and might therefore contain a forced signal (Lewis and Curry 2018). Since the MEI is a combination of several variables, some of which might be less influenced by external forcing, it is not obvious where this signal originates. In pacemaker experiments the forced signal is implicitly removed as it is present in both the nudged and the freely evolving experiment. These experiments, however, rely on the faithful representation of the forced signal by a single climate model.

We estimate the forced signal in the tropical Pacific with the method of Turkington et al. (2019) and by using multimodel means. Turkington et al. (2019) quantify the forced signal as the linear trend in tropical SSTs from 30° S to 30° N over 1962–2011, a period chosen such that there is little trend in the IPO. The rate of global warming, however, increased significantly in the early 1970s (Rahmstorf et al. 2017) related to changes in anthropogenic forcing.
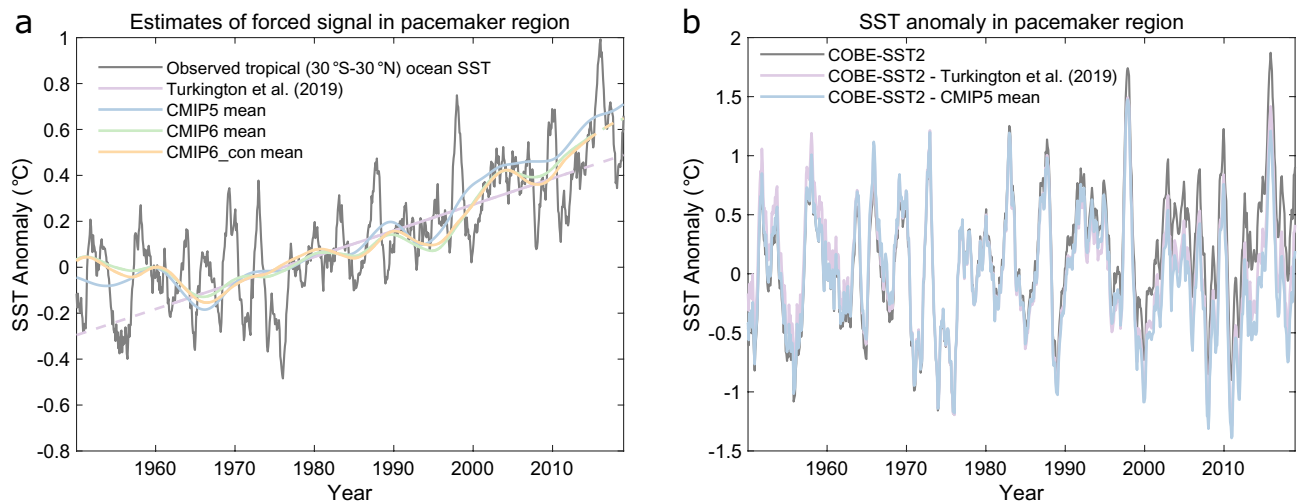
Fig. 6 a Different estimates of the forced signal in the pacemaker region (with respect to 1961–1990) from 1950 to 2018. The observed monthly COBE-SST2 tropical ocean SST evolution (30° S–30° N) used to estimate the forced response with the Turkington et al. (2019) approach is shown in grey. We indicate extrapolation with the dashed part of the line. The CMIP ensemble mean time series are smoothed with a loess smoother. **b** Observed area-mean SST time series in the pacemaker region when the estimated forced signals are removed (reference period is 1961–1990). Results for the Nino3.4 region, with CMIP6 or CMIP6_con instead of CMIP5 to represent the forced signal, and with ERSSTv5 instead of COBE-SST2 are all similar and therefore not shown

This might bias the linear trend low. The approach further assumes that the tropical-wide forced trend is similar to the forced signal in the eastern tropical Pacific. Different to Turkington et al. (2019), we only estimate one trend over the whole period instead of separately for every month.

Alternatively, we directly take the multi-model means over the respective regions as estimates of the forced signal. Some higher frequency variability is still present, and we therefore smooth the multi-model means with a loess smoother, although that also partly removes the volcanic signal (Fig. 6a). The multi-model means show similar forced responses, and relative to that the Turkington method underestimates the forced signal in the past two decades, yet the multi-model means could be biased. It has therefore been suggested to scale either the regional multi-model mean SST (e.g., Frankcombe et al. 2015) or the simulated GMST (e.g., Kajtar et al. 2019) against the observed SST evolution in the region of interest by means of linear regression. In the case of the eastern equatorial Pacific SST the scaling factor is, however, not stable, but decreases when the most recent decades are included. This is not surprising given the occurrence of the 'global warming hiatus'. Nonetheless, the scaling approach indicates a forced signal that is in between the Turkington et al. approach and the unscaled regional ensemble means (Fig. S8). Since we sample a reasonable range of the forced response uncertainty, we do not include the scaling approach in the following analysis.

We subtract the estimates of the forced signal from the observed SSTs to obtain the actual internal variability. After the correction, SSTs become colder during the late period

of the observational record (Fig. 6b). The approach of Turkington et al. (2019) spuriously warms the mid-twentieth century, which is a limitation of the linear extrapolation. Using these forcing corrected SST time series, we estimate the Pacific contribution to GMST with the variability analogues method.

Overall, the running 15-year long trends associated with Pacific variability exhibit agreement with the difference between observed and modelled warming, irrespective of the method used to estimate the forced response, but differences become evident during recent decades (Fig. 7a). During the 'global warming hiatus' the uncorrected, raw SST index underestimates the Pacific contribution as is expected from the presence of a forced signal (Fig. 6b).

From 1992 to 2012 (Fig. 7b), Pacific variability reduces the GMST increase by 0.09–0.10 °C when the forced warming of the Pacific is not accounted for. This is similar to the results of Rypdal (2018) and Huber and Knutti (2014), who directly use the uncorrected Nino3.4 index. With the Turkington et al. (2019) method, the Pacific contribution increases to 0.14–0.17 °C, and with the multi-model means removed, to 0.18–0.22 °C (with CMIP5 mean subtracted), 0.17–0.23 °C (with CMIP6), and 0.16–0.21 °C (with CMIP6_con). The results from the different climate model ensembles are similar, because their forced responses are similar (Fig. 6a). This demonstrates the importance of removing the forced signal: It approximately doubles the influence of the Pacific during the 'hiatus'.
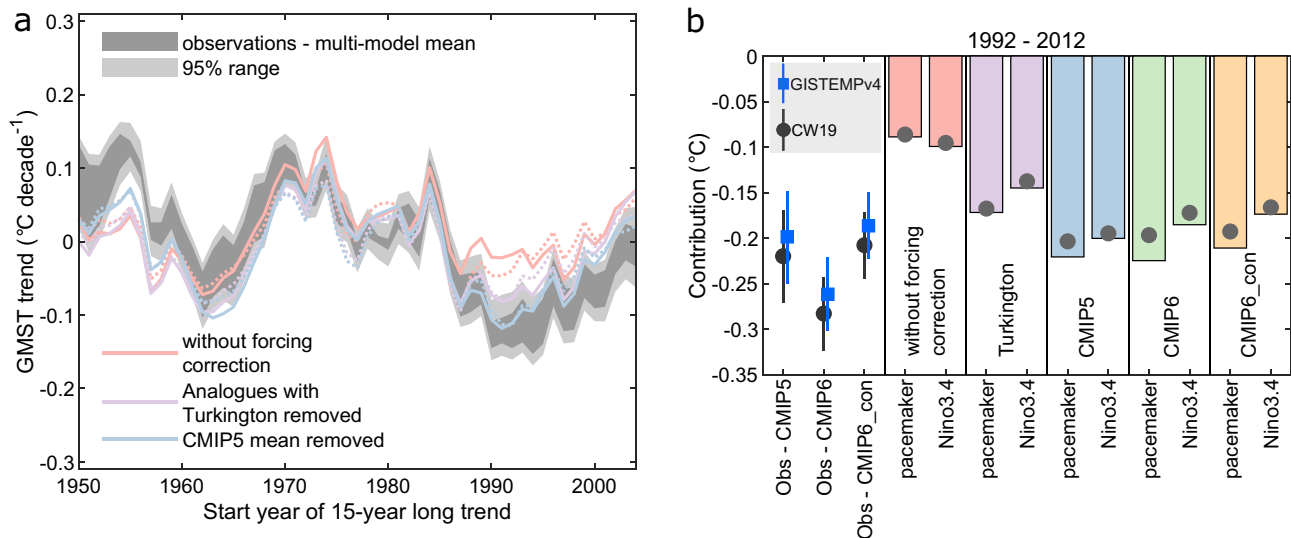
Fig. 7 a Difference between running 15-year long trends in observed and modelled GMST compared to the contribution of equatorial Pacific SST variability (average of COBE-SST2 and ERSSTv5) to GMST using the method of variability analogues. Solid lines indicate analogues that were selected based on variability in the pacemaker region, and dotted lines analogues based on the Nino3.4 region. The different colours represent methods of removing the forced response from observed SST. The CMIP6 and CMIP6_con estimates are omitted here for clarity as they are similar to the CMIP5 ensemble mean (Fig. 6a). The dark grey range indicates the minimum to maximum difference between the different observational GMST datasets and

multi-model means, and in light grey the 95% range across the different combinations with resampled ensemble means is shown. Similar figures, but for running 10-year and 20-year long trends are shown in the supporting information Figure S9. **b** Contribution of Pacific variability to GMST during the period from 1992 to 2012 calculated as the least squares linear trend multiplied by 21 years (bars). The grey dots in (**b**) indicate estimates for which the modelled and observed time series were not standardized prior to selecting the analogues. For the difference between observed and modelled warming, we show the 90% range across the resampled ensemble means (whiskers)

Kosaka and Xie (2013) additionally performed a SST-based pacemaker experiment with fixed radiative forcings, which conceptually corresponds to using observed, uncorrected SST within a regression model or within variability analogues. The difference between the two types of experiments reveals a similar effect as discussed, and instead of cooling by more than 0.4 °C during 1992–2012 (Fig. 5b), global temperatures are reduced by only 0.13 °C when forcings are kept constant.

Removing the forced response from observed SST naturally makes a smaller difference for earlier periods, because the radiative forcing was weaker. Nonetheless, the effect cannot be neglected. Depending on whether variability cools or warms the planet during a certain period, accounting for positive radiative forcing will either increase or decrease the Pacific influence. For the period 1974–1995, as an example, the observed warming was greater than modelled, and the Pacific variability shows a positive trend. Part of the Pacific warming is externally driven and removing it lowers the internal variability contribution to the warming during this period. We suspect this is one reason why the SST-based pacemaker and regression-based estimates overlap during this period (Fig. S7), but not during the 'global warming hiatus' (Fig. 5b).

4.4 Contribution of model uncertainty

While the method of estimating and removing the forced response from Pacific SST is important, in particular during the recent 'hiatus' period, it does not explain why some SST-pacemaker studies find a Pacific contribution that is even larger. To examine this, we rank the climate models according to their sensitivities with which the simulated GMST responds to variability in the tropical Pacific on a decadal time scale. We estimate the sensitivities by regressing GMST trends on pacemaker SST trends in each model's control simulation for 16-year long trends (corresponding to the period 1997–2012), 21-year long (1992–2012) and 22-year long trends (1974–1995). Then, we multiply these sensitivities with the observed SST change during each period. AMV may also affect GMST on these time scales, and including the modelled AMV as an additional predictor decreases the spread in regression coefficients relating the Pacific variability influence to GMST somewhat, but does not fundamentally change the conclusions.

This analysis demonstrates, in agreement with Wang et al. (2017), that there are considerable differences in how strongly tropical Pacific SST variability projects onto global mean temperatures between models. On the examined time
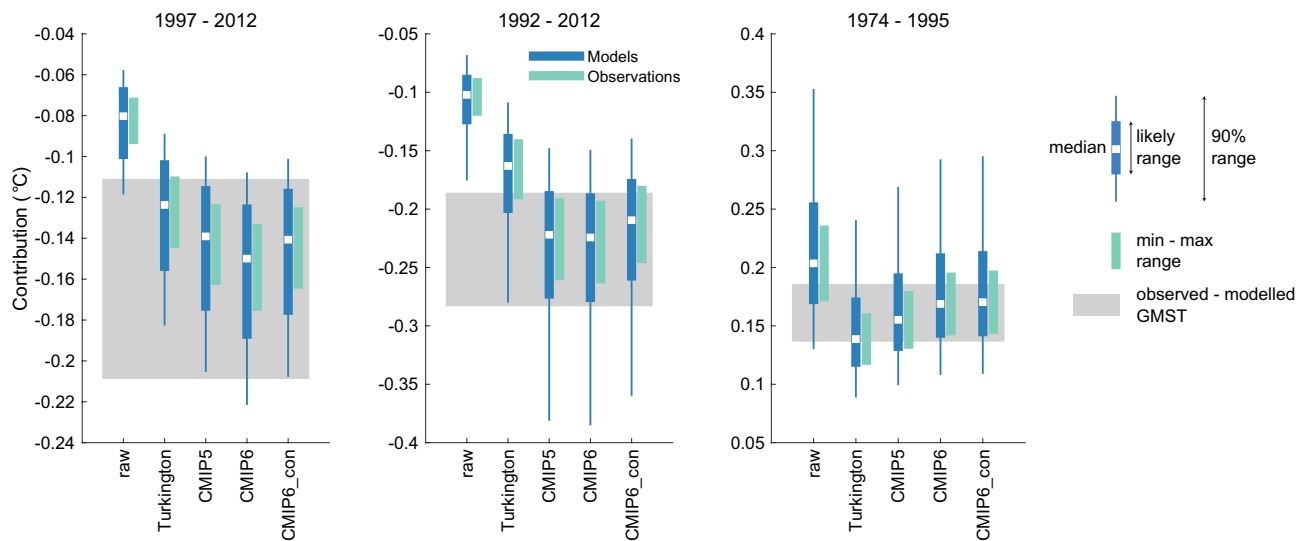
Fig. 8 Contribution of Pacific variability to GMST depending on model sensitivity and method of estimating the forced signal (i.e., raw, that is without forcing correction, Turkington, CMIP5, CMIP6, and CMIP6_con). The sensitivity with which GMST responds to Pacific internal variability is estimated from the SST in the pace-maker region of CMIP5 and CMIP6 control simulations with a length of at least 400 years (Table S1) and is multiplied with the observed SST trend (average of COBE-SST2 and ERSSTv5 with the different estimates of the forced signal removed as indicated). Shown are the likely (17–83%) and 90% (5–95%) ranges. For the observational estimate we display the minimum to maximum range across 36 estimates of the sensitivity with which GMST responds to Pacific variability based on combinations of GMST dataset, SST dataset, and method of removing the forced signal from the GMST and SST. The grey shaded areas indicate the difference between observed and multi-model mean warming during each period

scales, there is at least a factor of two difference between low- and high-sensitivity models (Fig. 8). The combination of model uncertainty and forced signal explains most of the spread across the studies examined: if the forced signal is not accounted for, even high-sensitivity models would not be able to fully explain the 'global warming hiatus' (which is also the case for the control experiment of Kosaka and Xie (2013) as discussed in the previous section). In absolute terms, the spread between low- and high-sensitivity models is relatively small if the forced signal is not removed. However, if it is accounted for, the sensitivity of a single model becomes more important, and the 'hiatus' contribution ranges from less than 0.15 °C to around 0.40 °C cooling (Fig. 8; middle), as we indeed observe (Fig. 5b). From 1974 to 1995, when Pacific variability acted to increase GMST, removing the forced response reduces Pacific warming and accordingly the model spread shrinks.

To verify whether modelled sensitivities are supported by observations, we subtract estimates of the forced signal, based on the multi-model ensemble means, from the SST and GMST time series from 1900 to 2018 and then follow the procedure as described. As shown in Fig. 8, observationally based estimates are comparable to that of the multi-model median, consistent with a similar analysis of Kajtar et al. (2019) for the IPO. Model-based estimates of the Pacific contribution to GMST variability should therefore be based on a sufficiently large number of climate models,

either by repeating pacemaker experiments with different models, or by using variability analogues from a large set of control simulations.

4.5 Variability analogues on wind trends

Motivated by the large spread across wind-stress based pacemaker experiments (Figs. 5, S7), we next address the question, whether these experiments are expected to result in a similar Pacific contribution as SST-based setups, and what role observational uncertainty plays. Therefore, we first examine the wind response in analogues selected from tropical Pacific SST.

In these analogues, the standardized variability in zonal wind stress in the western tropical Pacific agrees well with observed variability (Fig. 9a), but 15-year long trends in absolute wind stress are generally underestimated (Fig. 9b). Swingedouw et al. (2017) observe a similar underestimation in wind stress trends when they nudge towards observed SSTs in their pacemaker experiment. This might indicate too weak SST-wind coupling in the climate model ensemble, other sources of internal variability (Swingedouw et al. 2017), or that the wind stress trends are partly driven by processes outside the Pacific, such as tropical Atlantic Ocean SST variations (e.g., Chikamoto et al. 2016; McGregor et al. 2014).
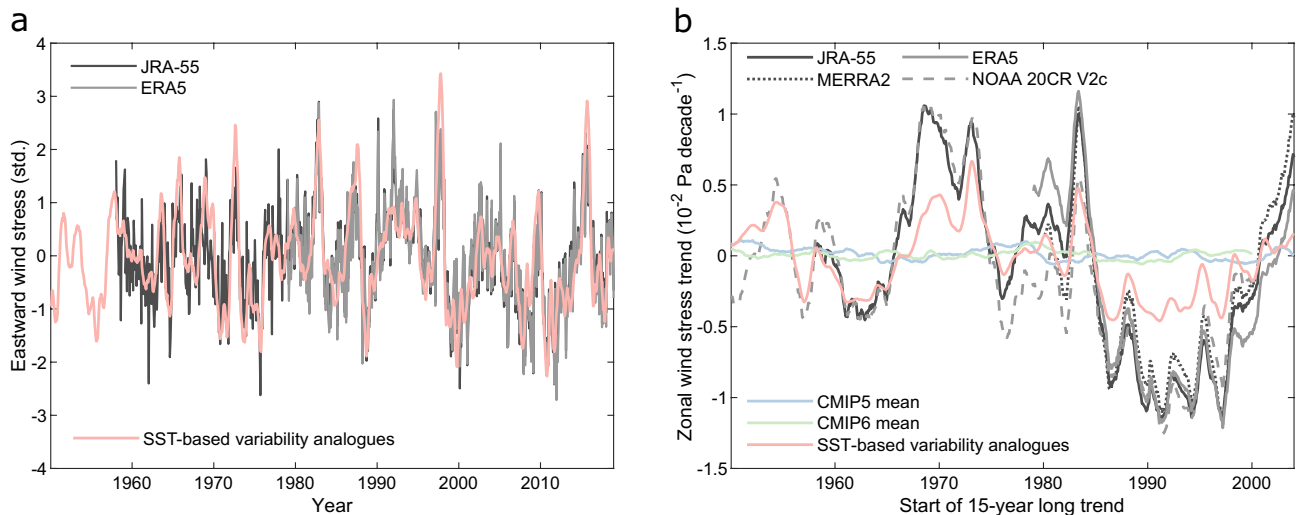
Fig. 9 a Monthly mean zonal wind stress anomalies in the western equatorial Pacific (Saenko region; cf. Figure 2) from SST-based variability analogues (with multi-model means removed from pacemaker region) and two reanalyses, JRA-55 and ERA5. The time series are standardized over the period 1982–2010. The correlation coefficients between the wind stress variability from SST-analogues and from reanalyses are 0.71 and 0.75 with JRA-55 and ERA5, respectively.

b Running 15-year long trends in wind stress from SST-based variability analogues, from historical CMIP5 and CMIP6 ensemble means, and from four reanalysis datasets (Saenko region; trends start every month). While the standardized month-to-month wind stress variability is well-matched, its long-term trends are underestimated by the SST analogues
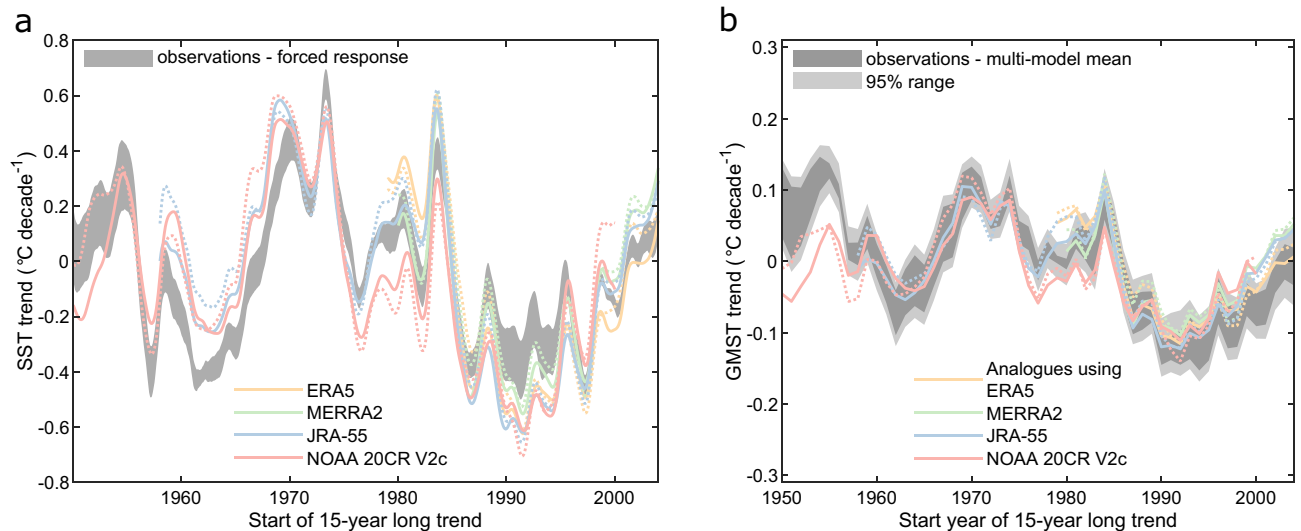
Fig. 10 a Observed (with forced response removed) running 15-year long SST trends in the pacemaker region starting every month. The grey shading indicates the minimum to maximum difference between different SST datasets and methods of removing the forced trend from the observations. The coloured lines show the SST trends for analogues selected based on the observed wind stress in the Sae-nko region (solid) and in the England region (dashed) from different reanalyses. b Difference between running 15-year long trends in observed and modelled GMST compared to the contribution of

wind variability estimated using the method of variability analogues. Solid lines indicate analogues that were selected based on variability in the Saenko region, and dotted lines analogues based on the England region. The dark grey range indicates the minimum to maximum difference between the different observational GMST datasets and multi-model means, and in light grey the 95% range across the different combinations with resampled ensemble means is shown. Similar figures, but for running 10-year and 20-year long trends are shown in the supporting information Figure S9

There is only a negligible forced Pacific wind stress trend in the historical simulations (Fig. 9b), consistent with the findings of Watanabe et al. (2014), which circumvents the

need for estimating and removing the forced signal when searching for wind-based variability analogues. Unlike the SST time series, we do not standardize the wind stress prior

to selecting analogues, as the standardized wind stress variability is already well captured when searching for SST-based analogues, but the absolute trend is underestimated.

The eastern equatorial Pacific SST response is overestimated in some wind-stress experiments (England et al. 2014; Gastineau et al. 2019), and we examine whether this is also the case with the variability analogues. As depicted in Fig. 10a, 15-year long trends in tropical Pacific SST are generally well-matched in the analogues based on wind variability. While SST trends starting in the 1960s tend to be underestimated in the analogues, trends starting from the early 1990s are overestimated. It is striking that the overestimation in the latter period is less severe than the underestimation of wind stress trends in SST-based analogues during the same period (cf. Figure 9b). We do not investigate this further, but it could be related to the set of models selected to provide analogues, and to a general tendency of the analogues to underestimate large changes in the target quantity.

Given that the wind stress analogues realistically reproduce eastern Pacific SST, we still consider them a useful setup for quantifying the Pacific influence on GMST. Except for trends starting in the 1950s, the effect of Pacific variability estimated from wind analogues reconciles observed and modelled GMST changes (Fig. 10b), and is similar to estimates based on SST-analogues after accounting for the forced response (Fig. 7a). Observational uncertainty is probably not the main reason for the differences among the wind-stress pacemaker experiments, because the different reanalyses show similar trends (Fig. 9b) and accordingly similar imprints on GMST during the studied period (Fig. 10b).

However, we only examine the wind variability over a relatively small region, whereas the pacemaker experiments prescribe winds over a much larger region (Table S2), and significant differences might exist elsewhere.

To examine the spread among wind-based pacemaker experiments further, we regress modelled GMST trends against wind stress trends (i.e., as in Sect. 4.4 for tropical Pacific SST). This reveals that the spread in regression slopes between models is even larger than with tropical Pacific SST (Fig. S10), which might not be surprising given that the wind variability is less directly connected to GMST variability (Saenko et al. 2016). This reaffirms the need for repeating pacemaker experiments with multiple different climate models.

5 Discussion

While we demonstrate that models that track the observed tropical Pacific variability (either by nudging or by searching for sufficiently short analogues) closely resemble the observed GMST changes, we have not yet examined whether models are actually capable of doing so on their own. Figure 11 compares both the observed SST trends in the tropical Pacific (Fig. 11b), from which we remove different estimates of the forced signal, and wind stress trends from reanalyses (Fig. 11a) with modelled variability from piControl simulations during the 'global warming hiatus'. Consistent with the results of England et al. (2014), CMIP5 models are unable to simulate the intensification of trade winds during the 'global
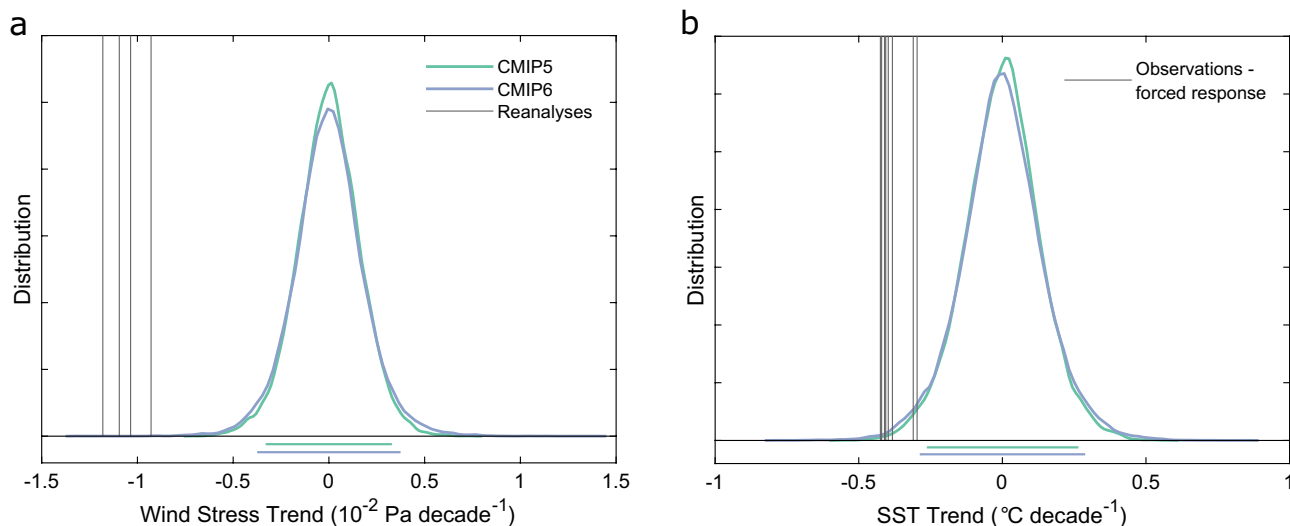
Fig. 11 Kernel probability density estimates of modelled 21-year long piControl trends in **a** zonal wind stress in the Saenko region and **b** SST in the pacemaker region with estimates of the observed changes during 1992–2012. For **a** the observed changes are the differ-

ent reanalyses and for **b** combinations of observational SST datasets and methods to remove the forced response. The horizontal bars indicate the ±2σ range of modelled trends

warming hiatus' (Fig. 11a). Observed wind stress trends starting in the early 1990s are not seen in any CMIP5 model irrespective of the reanalysis. Although the CMIP6 ensemble displays a broader range of 21-year long Pacific trade wind trends, only one model, MIROC6, simulates trends within the observed range.

Climate models generally appear to underestimate Pacific trade wind variability on a decadal time scale (Kajtar et al. 2018; Kociuba and Power 2015), possibly related to model biases in Atlantic mean SST, which mute the contribution of Atlantic SST variations to Pacific trade wind trends (Kajtar et al. 2018; McGregor et al. 2018), or there is an external, forced contribution to wind stress that models miss (cf. Figure 9b). In particular, the origin of the tropical Atlantic warming during the 'global warming hiatus', whether internal variability or anthropogenically forced, matters for the interpretation of the Pacific trade wind acceleration (McGregor et al. 2018). It has further been suggested that anthropogenic aerosols play a role during the 'global warming hiatus' by altering the trade winds and the state of the PDO (Smith et al. 2016; Takahashi and Watanabe 2016), but the robustness of these results has been questioned (Kuntz and Schrag 2016; Oudar et al. 2018). Further, observational uncertainties exist, and these are particularly large in the pre-satellite era (Bordbar et al. 2017; Kajtar et al. 2018). Reanalyses have been argued to overestimate the intensification of the Pacific Walker circulation during the past decades when compared to satellite-derived estimates (Chung et al. 2019), but de Boisséson et al. (2014) found the tropical Pacific trade wind trends, which are the surface expression

of the Walker circulation cell, to be robust in different observations including reanalyses.

Regarding the eastern tropical Pacific SST trends, Fyfe and Gillett (2014) found that none of 117 historical CMIP5 simulations reproduces the observed regional cooling from 1993 to 2012, and Kajtar et al. (2019) argue that the CMIP5 ensemble underestimates IPO variability on time scales greater than 10 years. When searching for 21-year long trends in the piControl simulations, both the CMIP5 and CMIP6 ensembles show trends as observed during the 'global warming hiatus' from 1992 to 2012 (Fig. 11b), but with a higher frequency in the CMIP6 ensemble. Whereas 9–23 out of 33 CMIP5 models (depending on the dataset, and the method of removing the forced response from the observations) are capable of simulating cooling equal to or exceeding the observed eastern tropical Pacific cooling, 13–28 out of 35 CMIP6 models do so.

There is some evidence that the response of the tropical Pacific SST to anthropogenic forcing is biased in CMIP5 models (Coats and Karnauskas 2017), with too much warming in the eastern tropical Pacific, possibly related to a climatological cold bias of models' equatorial cold tongues (Seager et al. 2019). If that is the case, the Pacific cooling during the 'global warming hiatus' from internal variability is smaller, as discussed in Sect. 4.3. This, in turn, would imply that less of the discrepancy between observed and modelled warming during the 'global warming hiatus' can be attributed to Pacific internal variability (cf. Figure 7b). A more detailed analysis of
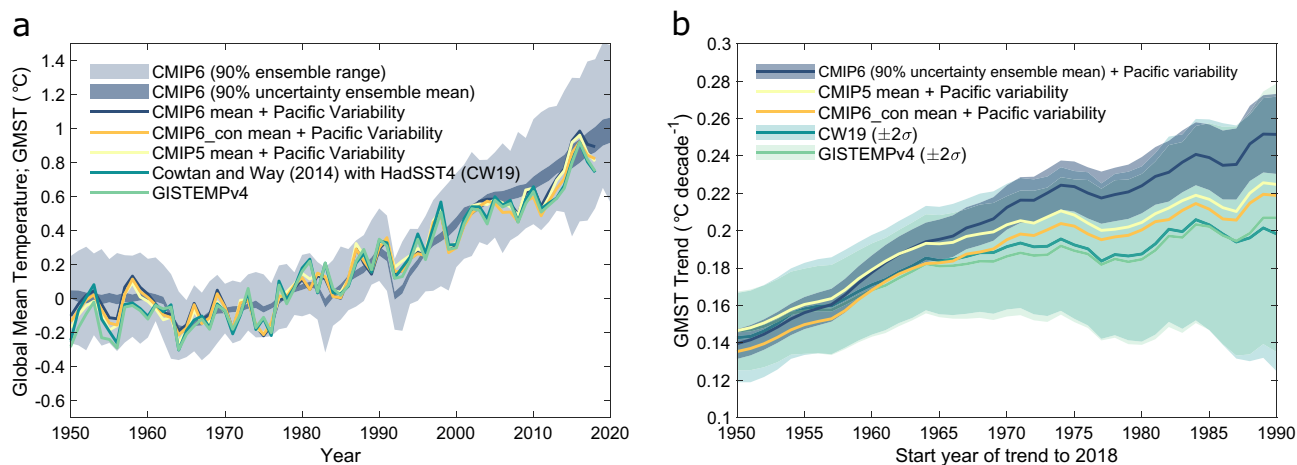
Fig. 12 a Observed and modelled annual GMST anomalies since 1950 (reference period is 1961–1990). For the 90% ensemble spread we use one member per model, otherwise we first average the realizations of each model. An estimate of the Pacific variability influence on GMST based on the average of analogues from both COBE-SST2 and ERSSTv5 in the pacemaker region is added to the climate model ensemble means. We remove the regional CMIP6_con mean from the observed eastern tropical Pacific SST prior to selecting the variability analogues. **b** Trends in observed and modelled GMST with different start years that all end in year 2018. Like Fig. 3a, but with added Pacific variability to the multi-model means. The results are similar when an estimate of Pacific variability based on wind stress is added to the multi-model means instead (Fig. S11) or an estimate based on the multi-model median regression coefficient (Fig. S12)

the CMIP6 ensemble will be required to assess whether the cold tongue bias has improved from CMIP5 and how to best quantify the forced signal in the tropical Pacific. This issue is common to all methods of quantifying the Pacific influence on GMST based on observed SST. Better constraining the forced signal in the tropical Pacific will help to reduce the uncertainty of the influence of internal Pacific variability on GMST during the 'hiatus' and in the future.

6 Conclusion: reconciling observed and modelled global mean temperature

While the three multi-model means track the overall observed GMST change from 1950 to 2018 well, with Pearson correlation coefficients of 0.96, 0.94, and 0.95, for CMIP5, CMIP6 and CMIP6_con, respectively, they do not match the year-to-year variability (Fig. 1a), and differences exist on decadal time scales (Fig. 1b). After adding in the influence of the tropical Pacific, keeping the potential limitations discussed in the previous section in mind, most of the discrepancies between observed and modelled GMST are resolved, and the correlation coefficients increase to 0.99 (with CMIP5), 0.97 (CMIP6), and 0.98 (CMIP6_con) (Fig. 12a). In agreement with Lehner et al. (2016), the apparent mismatch between simulated and observed volcanic cooling is also resolved after adjusting the multi-model means for the effect of Pacific variability.

Whereas the GMST trends of the CMIP5 and CMIP6_con ensemble means are similar to the observed trends after accounting for Pacific variability, the CMIP6 warming remains larger than observed during the past four decades (Fig. 12b). This is consistent with growing evidence that some CMIP6 models with a high TCR, present in CMIP6 but not in CMIP6_con, overestimate past and future warming (Brunner et al. 2020; Dittus et al. 2020; Liang et al. 2020; Nijsse et al. 2020; Tokarska et al. 2020; Winton et al. 2020), and supports a TCR of around or slightly below 1.8 °C. However, we stress that uncertainties in the tropical forced signal (Fig. 6) and the sensitivity with which GMST responds to Pacific variability (Fig. 8) remain.

The good agreement between observed and modelled warming since 1950 does not necessarily imply that they agree for the right reasons. First, there is significant uncertainty in the forced signal, as is evident from the ensemble spread, but this is not represented when analysing multi-model ensemble means, and similar forced signals can be achieved with various combinations of aerosol forcing and climate sensitivity (Kiehl 2007; Knutti 2008). Second, observational uncertainty remains (e.g., Davis et al. 2019), which might influence both the difference between modelled and observed warming, and the indices of internal variability used in this study. Third, further modes of internal variability might affect the observed global mean temperature evolution. In particular, the Atlantic Multi-decadal Variability (AMV) has been argued to contribute significantly to GMST variations (e.g., Chen and Tung 2018; Chylek et al. 2016; Wu et al. 2019), although this view has been challenged repeatedly (e.g., Booth et al. 2012; Haustein et al. 2019). In the supporting information we estimate the AMV contribution to GMST and show that it does not reduce the difference between models and observations, but on the contrary, increases it. Further research is required to better understand this, but we note that the AMV contribution to GMST is relatively small after 1980 irrespectively.

We have shown that differences in the estimated Pacific contribution to GMST arise primarily from the method of how (and if) the forced signal in the tropical Pacific is accounted for and from the sensitivity of GMST on tropical Pacific SST and wind variability on a decadal time scale which varies substantially across models, in agreement with Wang et al. (2017) and Bordbar et al. (2019). When we account for this, we can reproduce most of the spread across the 18 studies examined. Further factors, such as a time-scale dependence of the Pacific influence on GMST (Peyser et al. 2016; Wang et al. 2017) likely further contribute to the differences between the studies.

Our results demonstrate that pacemaker experiments using a single model should be interpreted with caution. Based on observations, we have demonstrated that high sensitivities between tropical Pacific SST and GMST as simulated by some models are unlikely, but observations support a sensitivity similar to the multi-model median, consistent with Kajtar et al. (2019). To obtain robust estimates, pacemaker experiments should therefore be repeated with several models, as is currently done within CMIP6. For IPSL-CM5A-LR, the wind-stress simulation indicates a greater Pacific contribution to the 'hiatus' than the SST-nudging experiment (Gastineau et al. 2019; Swingedouw et al. 2017), whereas it is the opposite for CNRM-CM5 (Douville et al. 2015). Artificial heat uptake in SST-based pacemaker experiments is therefore presumably not a major concern compared to other uncertainties when quantifying the Pacific effect on GMST.

For multiple-regression approaches we recommend careful evaluation of whether the predictors contain a forced signal, how it influences the results, and how to best remove it prior to the analysis. There might be cases where it is justifiable not removing the forced signal, e.g., if a minimum Pacific contribution to the 'global warming hiatus' is estimated, but the assumptions should be clearly stated.

References

Benestad RE, Erlandsen HB, Mezghani A, Parding KM (2019) Geographical distribution of thermometers gives the appearance of lower historical global warming. Geophys Res Lett 46:7654–7662. https://doi.org/10.1029/2019gl083474

Boer GJ et al (2016) The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. Geosci Model Dev 9:3751–3777. https://doi.org/10.5194/gmd-9-3751-2016

Booth BBB, Dunstone NJ, Halloran PR, Andrews T, Bellouin N (2012) Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. Nature 484:228–232. https://doi.org/10.1038/nature10946

Bordbar MH et al (2019) Uncertainty in near-term global surface warming linked to tropical Pacific climate variability. Nat Commun 10:1990. https://doi.org/10.1038/s41467-019-09761-2

Bordbar MH, Martin T, Latif M, Park W (2017) Role of internal variability in recent decadal to multidecadal tropical Pacific climate changes. Geophys Res Lett 44:4246–4255. https://doi.org/10.1002/2016gl072355

Brown PT, Ming Y, Li WH, Hill SA (2017) Change in the magnitude and mechanisms of global temperature variability with warming. Nat Climate Change 7:743–748. https://doi.org/10.1038/Nclimate3381

Brunner L, Pendergrass AG, Lehner F, Merrifield AL, Lorenz R, Knutti R (2020) Reduced global warming from CMIP6 projections when weighting models by performance and independence. Earth Syst Dynam Discuss 2020:1–23. https://doi.org/10.5194/esd-2020-23

Bruns SB, Csereklyei Z, Stern DI (2020) A multicointegration model of global climate change. J Econ 214:175–197. https://doi.org/10.1016/j.jeconom.2019.05.010

Chan D, Kent EC, Berry DI, Huybers P (2019) Correcting datasets leads to more homogeneous early-twentieth-century sea surface warming. Nature 571:393–397. https://doi.org/10.1038/s41586-019-1349-2

Chen XY, Tung KK (2018) Global-mean surface temperature variability: space-time perspective from rotated EOFs. Clim Dyn 51:1719–1732. https://doi.org/10.1007/s00382-017-3979-0

Chikamoto Y, Mochizuki T, Timmermann A, Kimoto M, Watanabe M (2016) Potential tropical Atlantic impacts on Pacific decadal climate trends. Geophys Res Lett 43:7143–7151. https://doi.org/10.1002/2016gl069544

Chung ES, Timmermann A, Soden BJ, Ha KJ, Shi L, John VO (2019) Reconciling opposing Walker circulation trends in observations and model projections. Nat Climate Change 9:405–412. https://doi.org/10.1038/s41558-019-0446-4

Chylek P, Klett JD, Lesins G, Dubey MK, Hengartner N (2014) The Atlantic Multidecadal Oscillation as a dominant factor of oceanic influence on climate. Geophys Res Lett 41:1689–1697. https://doi.org/10.1002/2014gl059274

Chylek P, Klett JD, Dubey MK, Hengartner N (2016) The role of Atlantic Multi-decadal Oscillation in the global mean temperature variability. Clim Dyn 47:3271–3279. https://doi.org/10.1007/s00382-016-3025-7

Clarke DC, Richardson M (2020) The benefits of continuous local regression for quantifying global warming. submitted. https://doi.org/10.1002/essoar.10502294.1

Coats S, Karnauskas KB (2017) Are simulated and observed twentieth century tropical Pacific sea surface temperature trends

significant relative to internal variability? Geophys Res Lett 44:9928–9937. https://doi.org/10.1002/2017gl074622

Compo GP et al (2011) The twentieth century reanalysis project. Q J Roy Meteorol Soc 137:1–28. https://doi.org/10.1002/qj.776

Cowtan K et al (2015) Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures. Geophys Res Lett 42:6526–6534. https://doi.org/10.1002/2015gl064888

Cowtan K, Way RG (2014) Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. Q J Roy Meteorol Soc 140:1935–1944. https://doi.org/10.1002/qj.2297

Cowtan K, Rohde R, Hausfather Z (2018) Evaluating biases in sea surface temperature records using coastal weather stations. Q J Roy Meteorol Soc 144:670–681. https://doi.org/10.1002/qj.3235

Dai X-G, Wang P (2018) Identifying the early 2000s hiatus associated with internal climate variability. Sci Rep. https://doi.org/10.1038/s41598-018-31862-z

Dai A, Fyfe JC, Xie S-P, Dai X (2015) Decadal modulation of global surface temperature by internal climate variability. Nat Climate Change 5:555–559. https://doi.org/10.1038/nclimate2605

Davis LLB, Thompson DWJ, Kennedy JJ, Kent EC (2019) The importance of unresolved biases in twentieth-century sea surface temperature observations. Bull Am Meteorol Soc 100:621–629. https://doi.org/10.1175/Bams-D-18-0104.1

de Boisséson E, Balmaseda MA, Abdalla S, Källén E, Janssen PAEM (2014) How robust is the recent strengthening of the Tropical Pacific trade winds? Geophys Res Lett 41:4398–4405. https://doi.org/10.1002/2014gl060257

Delworth TL, Zeng FR, Rosati A, Vecchi GA, Wittenberg AT (2015) A link between the hiatus in global warming and North American drought. J Clim 28:3834–3845. https://doi.org/10.1175/Jcli-D-14-00616.1

Deser C, Alexander MA, Xie SP, Phillips AS (2010) Sea surface temperature variability: patterns and mechanisms. Ann Rev Marine Sci 2:115–143. https://doi.org/10.1146/annurev-marine-120408-151453

Deser C, Guo R, Lehner F (2017) The relative contributions of tropical Pacific sea surface temperatures and atmospheric internal variability to the recent global warming hiatus. Geophys Res Lett 44:7945–7954. https://doi.org/10.1002/2017GL074273

Dittus AJ, Hawkins E, Wilcox LJ, Sutton R, Smith CJ, Andrews MB, Forster PM (2020) Sensitivity of historical climate simulations to uncertain aerosol forcing. Geophys Res Lett 47:e2019GL085806. https://doi.org/10.1029/2019gl085806

Douville H, Voldoire A, Geoffroy O (2015) The recent global warming hiatus: What is the role of Pacific variability? Geophys Res Lett 42:880–888. https://doi.org/10.1002/2014gl062775

England MH et al (2014) Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. Nat Climate Change 4:222–227. https://doi.org/10.1038/nclimate2106

England MH, Kajtar JB, Maher N (2015) Robust warming projections despite the recent hiatus. Nat Climate Change 5:394–396. https://doi.org/10.1038/nclimate2575

Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, Taylor KE (2016) Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geosci Model Dev 9:1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Flynn CM, Mauritsen T (2020) On the climate sensitivity and historical warming evolution in recent coupled model ensembles. Atmos Chem Phys 20:7829–7842. https://doi.org/10.5194/acp-20-7829-2020

Folland CK, Boucher O, Colman A, Parker DE (2018) Causes of irregularities in trends of global mean surface temperature since the late 19th century. Sci Adv. https://doi.org/10.1126/sciadv.aao5297

Foster G, Rahmstorf S (2011) Global temperature evolution 1979–2010. Environ Res Lett 6:044022. https://doi.org/10.1088/1748-9326/6/4/044022

Frankcombe LM, England MH, Mann ME, Steinman BA (2015) Separating internal variability from the externally forced climate response. J Clim 28:8184–8202. https://doi.org/10.1175/Jcli-D-15-0069.1

Fyfe JC, Gillett NP (2014) Recent observed and simulated warming. Nat Climate Change 4:150–151. https://doi.org/10.1038/nclimate2111

Gastineau G, Friedman AR, Khodri M, Vialard J (2019) Global ocean heat content redistribution during the 1998–2012 Interdecadal Pacific Oscillation negative phase. Clim Dyn 53:1187–1208. https://doi.org/10.1007/s00382-018-4387-9

Gelaro R et al (2017) The modern-era retrospective analysis for research and applications, Version 2 (MERRA-2). J Clim 30:5419–5454. https://doi.org/10.1175/Jcli-D-16-0758.1

Hansen J, Ruedy R, Sato M, Lo K (2010) Global surface temperature change. Rev Geophys. https://doi.org/10.1029/2010rg000345

Hausfather Z, Cowtan K, Clarke DC, Jacobs P, Richardson M, Rohde R (2017) Assessing recent warming using instrumentally homogeneous sea surface temperature records. Sci Adv. https://doi.org/10.1126/sciadv.1601207

Haustein K et al (2019) A limited role for unforced internal variability in 20th century warming. J Clim 32:4893–4917. https://doi.org/10.1175/jcli-d-18-0555.1

Haustein K, Allen MR, Forster PM, Otto FEL, Mitchell DM, Matthews HD, Frame DJ (2017) A real-time Global Warming Index. Sci Rep 7:15417. https://doi.org/10.1038/s41598-017-14828-5

Hawkins E, Sutton R (2016) Connecting climate model projections of global temperature change with the real world. Bull Am Meteorol Soc 97:963–980. https://doi.org/10.1175/Bams-D-14-00154.1

Hersbach H, Dee DP (2016) ERA5 reanalysis is in production. ECMWF Newsletter 147

Hirahara S, Ishii M, Fukuda Y (2014) Centennial-scale sea surface temperature analysis and its uncertainty. J Clim 27:57–75. https://doi.org/10.1175/Jcli-D-12-00837.1

Huang BY et al (2017) Extended reconstructed sea surface temperature, Version 5 (ERSSTv5): upgrades, validations, and intercomparisons. J Clim 30:8179–8205. https://doi.org/10.1175/Jcli-D-16-0836.1

Huang JB et al (2017b) Recently amplified arctic warming has contributed to a continual global warming trend. Nat Climate Change 7:875–879. https://doi.org/10.1038/s41558-017-0009-5

Huang BY et al (2018) Evaluating SST analyses with independent ocean profile observations. J Clim 31:5015–5030. https://doi.org/10.1175/Jcli-D-17-0824.1

Huber M, Knutti R (2014) Natural variability, radiative forcing and climate response in the recent hiatus reconciled. Nat Geosci 7:651–656. https://doi.org/10.1038/Ngeo2228

Hu SN, Fedorov AV (2017) The extreme El Nino of 2015–2016 and the end of global warming hiatus. Geophys Res Lett 44:3816–3824. https://doi.org/10.1002/2017gl072908

Jiménez-de-la-Cuesta D, Mauritsen T (2019) Emergent constraints on Earth's transient and equilibrium response to doubled CO2 from post-1970s global warming. Nat Geosci 12:902–905. https://doi.org/10.1038/s41561-019-0463-y

Johansson DJA, O'Neill BC, Tebaldi C, Häggström O (2015) Equilibrium climate sensitivity in light of observations over the warming hiatus. Nat Clim Change 5:449–453. https://doi.org/10.1038/nclimate2573

Jones PD, Lister DH, Osborn TJ, Harpham C, Salmon M, Morice CP (2012) Hemispheric and large-scale land-surface air temperature

variations: an extensive revision and an update to 2010. J Geophys Res-Atmos. https://doi.org/10.1029/2011jd017139

Kajtar JB, Santoso A, McGregor S, England MH, Baillie Z (2018) Model under-representation of decadal Pacific trade wind trends and its link to tropical Atlantic bias. Clim Dyn 50:1471–1484. https://doi.org/10.1007/s00382-017-3699-5

Kajtar JB, Collins M, Frankcombe LM, England MH, Osborn TJ, Juniper M (2019) Global mean surface temperature response to large-scale patterns of variability in observations and CMIP5. Geophys Res Lett 46:2232–2241. https://doi.org/10.1029/2018gl081462

Karl TR et al (2015) Possible artifacts of data biases in the recent global surface warming hiatus. Science 348:1469–1472. https://doi.org/10.1126/science.aaa5632

Kennedy JJ, Rayner NA, Atkinson CP, Killick RE (2019) An ensemble data set of sea surface temperature change from 1850: the met office hadley centre HadSST.4.0.0.0. data set. J Geophys Res-Atmos 124:7719–7763. https://doi.org/10.1029/2018jd029867

Kiehl JT (2007) Twentieth century climate model response and climate sensitivity. Geophys Res Lett. https://doi.org/10.1029/2007gl031383

Knutti R (2008) Why are climate models reproducing the observed global surface warming so well? Geophys Res Lett 35:L18704. https://doi.org/10.1029/2008gl034932

Kobayashi S et al (2015) The JRA-55 reanalysis: general specifications and basic characteristics. J Meteorol Soc Jpn 93:5–48. https://doi.org/10.2151/jmsj.2015-001

Kociuba G, Power SB (2015) Inability of CMIP5 models to simulate recent strengthening of the walker circulation: implications for projections. J Clim 28:20–35. https://doi.org/10.1175/Jcli-D-13-00752.1

Kosaka Y, Xie SP (2013) Recent global-warming hiatus tied to equatorial Pacific surface cooling. Nature 501:403–407. https://doi.org/10.1038/nature12534

Kosaka Y, Xie S-P (2016) The tropical Pacific as a key pacemaker of the variable rates of global warming. Nat Geosci 9:669–673. https://doi.org/10.1038/ngeo2770

Kuntz LB, Schrag DP (2016) Impact of Asian aerosol forcing on tropical Pacific circulation and the relationship to global temperature trends. J Geophys Res-Atmos 121:14403–14413. https://doi.org/10.1002/2016jd025430

Lean JL (2018) Observation-based detection and attribution of 21st century climate change. Wires Clim Change. https://doi.org/10.1002/wcc.511

Lee S-K, Park W, Baringer MO, Gordon AL, Huber B, Liu Y (2015) Pacific origin of the abrupt increase in Indian Ocean heat content during the warming hiatus. Nat Geosci 8:445–449. https://doi.org/10.1038/ngeo2438

Lehner F, Schurer AP, Hegerl GC, Deser C, Frölicher TL (2016) The importance of ENSO phase during volcanic eruptions for detection and attribution. Geophys Res Lett 43:2851–2858. https://doi.org/10.1002/2016gl067935

Lenssen NJL, Schmidt GA, Hansen JE, Menne MJ, Persin A, Ruedy R, Zyss D (2019) Improvements in the GISTEMP Uncertainty Model. J Geophys Res-Atmos 124:6307–6326. https://doi.org/10.1029/2018jd029522

Lewis N, Curry J (2018) The impact of recent forcing and ocean heat uptake data on estimates of climate sensitivity. J Clim 31:6051–6071. https://doi.org/10.1175/Jcli-D-17-0667.1

Liang Y, Gillett NP, Monahan AH (2020) Climate model projections of 21st century global warming constrained using the observed warming trend. Geophys Res Lett. https://doi.org/10.1029/2019gl086757

Maher N, England MH, Gupta AS, Spence P (2018) Role of Pacific trade winds in driving ocean temperatures during the recent slowdown and projections under a wind trend reversal. Clim Dyn 51:321–336. https://doi.org/10.1007/s00382-017-3923-3

Mann ME, Steinman BA, Miller SK (2014) On forced temperature changes, internal variability, and the AMO. Geophys Res Lett 41:3211–3219. https://doi.org/10.1002/2014gl059233

Marotzke J, Forster PM (2015) Forcing, feedback and internal variability in global temperature trends. Nature 517:565–570. https://doi.org/10.1038/nature14117

McGregor S, Timmermann A, Stuecker MF, England MH, Merrifield M, Jin F-F, Chikamoto Y (2014) Recent Walker circulation strengthening and Pacific cooling amplified by Atlantic warming. Nat Clim Change 4:888–892. https://doi.org/10.1038/nclimate2330

McGregor S, Stuecker MF, Kajtar JB, England MH, Collins M (2018) Model tropical Atlantic biases underpin diminished Pacific decadal variability. Nat Clim Change 8:493–498. https://doi.org/10.1038/s41558-018-0163-4

Medhaug I, Stolpe MB, Fischer EM, Knutti R (2017) Reconciling controversies about the 'global warming hiatus'. Nature 545:41–47. https://doi.org/10.1038/nature22315

Meehl GA et al (2020) Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. Sci Adv 6:eaba1981. https://doi.org/10.1126/sciadv.aba1981

Meehl GA, Arblaster JM, Fasullo JT, Hu A, Trenberth KE (2011) 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

Meehl GA, Hu A, Santer BD, Xie S-P (2016) Contribution of the Interdecadal Pacific Oscillation to twentieth-century global surface temperature trends. Nat Clim Change 6:1005–1008. https://doi.org/10.1038/nclimate3107

Menne MJ, Williams CN, Gleason BE, Rennie JJ, Lawrimore JH (2018) The global historical climatology network monthly temperature dataset, Version 4. J Clim 31:9835–9854. https://doi.org/10.1175/Jcli-D-18-0094.1

Montamat G, Stock JH (2020) Quasi-experimental estimates of the transient climate response using observational data. Clim Change. https://doi.org/10.1007/s10584-019-02589-1

Nijsse FJMM, Cox PM, Williamson MS (2020) Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models. Earth Syst Dynam 11:737–750. https://doi.org/10.5194/esd-11-737-2020

Oka A, Watanabe M (2017) The post-2002 global surface warming slowdown caused by the subtropical Southern Ocean heating acceleration. Geophys Res Lett 44:3319–3327. https://doi.org/10.1002/2016gl072184

Olonscheck D, Notz D (2017) Consistently estimating internal climate variability from climate model simulations. J Clim 30:9555–9573. https://doi.org/10.1175/Jcli-D-16-0428.1

Oudar T, Kushner PJ, Fyfe JC, Sigmond M (2018) No impact of anthropogenic aerosols on early 21st century global temperature trends in a large initial-condition ensemble. Geophys Res Lett 45:9245–9252. https://doi.org/10.1029/2018gl078841

Papalexiou SM, Rajulapati CR, Clark MP, Lehner F (2020) Robustness of CMIP6 historical global mean temperature simulations: trends, long-term persistence, autocorrelation, and distributional shape. Earth's Future. https://doi.org/10.1029/2020ef001667

Peyser CE, Yin J, Landerer FW, Cole JE (2016) Pacific sea level rise patterns and global surface temperature variability. Geophys Res Lett 43:8662–8669. https://doi.org/10.1002/2016gl069401

Phillips PCB, Leirvik T, Storelvmo T (2020) Econometric estimates of Earth's transient climate sensitivity. J Econ 214:6–32. https://doi.org/10.1016/j.jeconom.2019.05.002

Power S, Delage F, Wang GM, Smith I, Kociuba G (2017) Apparent limitations in the ability of CMIP5 climate models to simulate recent multi-decadal change in surface temperature: implications

for global temperature projections. Clim Dyn 49:53–69. https://doi.org/10.1007/s00382-016-3326-x

Pretis F (2020) Econometric modelling of climate systems: The equivalence of energy balance models and cointegrated vector autoregressions. J Econ 214:256–273. https://doi.org/10.1016/j.jeconom.2019.05.013

Rahmstorf S, Foster G, Cahill N (2017) Global temperature evolution: recent trends and some pitfalls. Environ Res Lett. https://doi.org/10.1088/1748-9326/aa6825

Richardson M, Cowtan K, Millar RJ (2018) Global temperature definition affects achievement of long-term climate goals. Environ Res Lett 13:054004. https://doi.org/10.1088/1748-9326/aab305

Ridley DA et al (2014) Total volcanic stratospheric aerosol optical depths and implications for global climate change. Geophys Res Lett 41:7763–7769. https://doi.org/10.1002/2014gl061541

Risbey JS, Lewandowsky S, Langlais C, Monselesan DP, O'Kane TJ, Oreskes N (2014) Well-estimated global surface warming in climate projections selected for ENSO phase. Nat Climate Change 4:835–840. https://doi.org/10.1038/nclimate2310

Rypdal K (2018) The life and death of the recent global surface warming hiatus parsimoniously explained. Climate. https://doi.org/10.3390/cli6030064

Saenko OA, Fyfe JC, Swart NC, Lee WG, England MH (2016) Influence of tropical wind on global temperature from months to decades. Clim Dyn 47:2193–2203. https://doi.org/10.1007/s00382-015-2958-6

Santer BD et al (2017) Causes of differences in model and satellite tropospheric warming rates. Nat Geosci 10:478–485. https://doi.org/10.1038/ngeo2973

Schmidt A et al (2018) Volcanic radiative forcing from 1979 to 2015. J Geophys Res-Atmos 123:12491–12508. https://doi.org/10.1029/2018jd028776

Schmidt GA, Shindell DT, Tsigaridis K (2014) Reconciling warming trends. Nat Geosci 7:158–160. https://doi.org/10.1038/ngeo2105

Schurer A, Hegerl G, Ribes A, Polson D, Morice C, Tett S (2018) Estimating the transient climate response from observed warming. J Clim 31:8645–8663. https://doi.org/10.1175/jcli-d-17-0717.1

Seager R, Cane M, Henderson N, Lee DE, Abernathey R, Zhang HH (2019) Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. Nat Clim Change 9:517–522. https://doi.org/10.1038/s41558-019-0505-x

Sherwood S et al (2020) An assessment of Earth's climate sensitivity using multiple lines of evidence. Rev Geophys. https://doi.org/10.1029/2019rg000678

Skeie RB, Berntsen T, Aldrin M, Holden M, Myhre G (2018) Climate sensitivity estimates – sensitivity to radiative forcing time series and observational data. Earth Syst Dynam 9:879–894. https://doi.org/10.5194/esd-9-879-2018

Smith DM et al (2016) Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. Nat Clim Change 6:936–940. https://doi.org/10.1038/nclimate3058

Steinman BA, Mann ME, Miller SK (2015) Atlantic and Pacific multidecadal oscillations and Northern Hemisphere temperatures. Science 347:988–991. https://doi.org/10.1126/science.1257856

Stolpe MB, Medhaug I, Knutti R (2017) Contribution of Atlantic and Pacific multidecadal variability to twentieth-century temperature changes. J Clim 30:6279–6295. https://doi.org/10.1175/Jcli-D-16-0803.1

Stolpe MB, Medhaug I, Beyerle U, Knutti R (2019) Weak dependence of future global mean warming on the background climate state. Clim Dyn 53:5079–5099. https://doi.org/10.1007/s00382-019-04849-3

Svendsen L, Keenlyside N, Bethke I, Gao YQ, Omrani NE (2018) Pacific contribution to the early twentieth-century warming in the Arctic. Nat Clim Change 8:793–797. https://doi.org/10.1038/s41558-018-0247-1

Swingedouw D, Mignot J, Guilyardi E, Nguyen S, Ormieres L (2017) Tentative reconstruction of the 1998–2012 hiatus in global temperature warming using the IPSL-CM5A-LR climate model. CR Geosci 349:369–379. https://doi.org/10.1016/j.crte.2017.09.014

Takahashi C, Watanabe M (2016) Pacific trade winds accelerated by aerosol forcing over the past two decades. Nat Clim Change 6:768–772. https://doi.org/10.1038/nclimate2996

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bull Am Meteorol Soc 93:485–498. https://doi.org/10.1175/Bams-D-11-00094.1

Thompson DWJ, Kennedy JJ, Wallace JM, Jones PD (2008) A large discontinuity in the mid-twentieth century in observed global-mean surface temperature. Nature 453:646–645. https://doi.org/10.1038/nature06982

Tokarska KB, Stolpe MB, Sippel S, Fischer EM, Smith CJ, Lehner F, Knutti R (2020) Past warming trend constrains future warming in CMIP6 models. Sci Adv 6:eaaz9549. https://doi.org/10.1126/sciadv.aaz9549

Trenberth KE (1997) The definition of El Nino. Bull Am Meteorol Soc 78:2771–2777. https://doi.org/10.1175/1520-0477(1997)078<2771:Tdoeno>2.0.Co;2

Tung KK, Chen XY (2018) Understanding the recent global surface warming slowdown. A review. Climate 6:19. https://doi.org/10.3390/cli6040082

Turkington T, Timbal B, Rahmat R (2019) The impact of global warming on sea surface temperature based El Niño–Southern Oscillation monitoring indices. Int J Climatol 39:1092–1103. https://doi.org/10.1002/joc.5864

Wang C-Y, Xie S-P, Kosaka Y, Liu Q, Zheng X-T (2017) Global influence of tropical Pacific variability with implications for global warming slowdown. J Clim 30:2679–2695. https://doi.org/10.1175/jcli-d-15-0496.1

Watanabe M et al (2013) Strengthening of ocean heat uptake efficiency associated with the recent climate hiatus. Geophys Res Lett 40:3175–3179. https://doi.org/10.1002/grl.50541

Watanabe M, Shiogama H, Tatebe H, Hayashi M, Ishii M, Kimoto M (2014) Contribution of natural decadal variability to global warming acceleration and hiatus. Nat Clim Change 4:893–897. https://doi.org/10.1038/nclimate2355

Wei M, Qiao F, Guo Y, Deng J, Song Z, Shu Q, Yang X (2019) Quantifying the importance of interannual, interdecadal and multidecadal climate natural variabilities in the modulation of global warming rates. Clim Dyn 53:6715–6727. https://doi.org/10.1007/s00382-019-04955-2

Winton M et al (2020) Climate sensitivity of GFDL's CM4.0. J Adv Model Earth Syst 12:e2019MS001838. https://doi.org/10.1029/2019ms001838

Wolter K, Timlin MS (2011) El Nino/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext). Int J Climatol 31:1074–1087. https://doi.org/10.1002/joc.2336

Wu T, Hu A, Gao F, Zhang J, Meehl GA (2019) New insights into natural variability and anthropogenic forcing of global/regional climate evolution. npj Clim Atmos Sci 2:18. https://doi.org/10.1038/s41612-019-0075-7

Xu X, He S, Furevik T, Gao Y, Wang H, Li F, Ogawa F (2020) Oceanic forcing of the global warming slowdown in multi-model simulations. Int J Climatol. https://doi.org/10.1002/joc.6548

Yin JJ, Overpeck J, Peyser C, Stouffer R (2018) Big jump of record warm global mean surface temperature in 2014–2016 related to unusually large oceanic heat releases. Geophys Res Lett 45:1069–1078. https://doi.org/10.1002/2017gl076500

Zelinka MD et al (2020) Causes of higher climate sensitivity in CMIP6 models. Geophys Res Lett 47:e2019GL085782. https://doi.org/10.1029/2019gl085782

Zhou TJ et al (2016) GMMIP (v1.0) contribution to CMIP6: global monsoons model inter-comparison project. Geosci Model Dev 9:3589–3604. https://doi.org/10.5194/gmd-9-3589-2016