

This is a repository copy of *Comment on "the impact of recent forcing and ocean heat uptake data on estimates of climate sensitivity"*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/163244/>

Version: Published Version

---

**Article:**

Cowtan, Kevin [orcid.org/0000-0002-0189-1437](https://orcid.org/0000-0002-0189-1437) and Jacobs, Peter (2020) Comment on "the impact of recent forcing and ocean heat uptake data on estimates of climate sensitivity". *Journal of Climate*. pp. 391-396. ISSN: 0894-8755

<https://doi.org/10.1175/JCLI-D-18-0316.1>

---

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

## CORRESPONDENCE

## Comment on “The Impact of Recent Forcing and Ocean Heat Uptake Data on Estimates of Climate Sensitivity”

KEVIN COWTAN

*Department of Chemistry, University of York, York, United Kingdom*

PETER JACOBS

*Department of Environmental Science and Policy, George Mason University, Fairfax, Virginia*

(Manuscript received 20 May 2018, in final form 11 October 2019)

## ABSTRACT

In 2018, Lewis and Curry presented a method for estimating the transient climate response (TCR) of the climate system from the temperature change between two time windows: an early baseline period in the nineteenth century and a modern period primarily in the twenty-first century. The results suggest a lower value of TCR than estimates from climate model simulations. Previous studies have identified uncertainty in the historical forcings, the impact of the time evolution of the forcing on temperature response, and observational issues as contributory factors to this disagreement. We investigate a further factor: uncertainty in the bias corrections applied to historical sea surface temperature data. This uncertainty can particularly affect the estimation of variables on decadal time scales and therefore affect the estimation of TCR using the window method as well as estimates of internal variability. We demonstrate that use of the whole historical record can mitigate the impacts of working with short time windows to some extent, particularly with respect to the early part of the record.

Several recent studies, including [Lewis and Curry \(2018\)](#) and [Otto et al. \(2013\)](#), use the ratio of the change in temperature to the change in forcing between two time windows as an estimator for transient climate response (TCR) and produce lower estimates of TCR than climate model simulations or other methods that are based on past change ([Knutti et al. 2017](#)). Previous studies have identified differences in the inferred forcings, differences in the temperature impact of historical versus transient forcing changes, and data type and coverage as potential explanatory factors for this difference ([Storelvmo et al. 2016](#); [Armour 2017](#); [Richardson et al. 2016](#)). In 2016 the authors of all of the major sea surface temperature (SST) datasets drew attention to major unresolved biases in historical sea surface temperature records ([Kent et al. 2017](#)), which may affect our understanding of both historical warming and internal variability. We demonstrate that these biases can also affect the results of the window

method when estimating TCR, and we explore to what extent this may be mitigated by using more of the data.

[Lewis and Curry \(2018\)](#) choose windows at the start and end of the historical temperature record as the basis for their TCR calculation. The early time window (1869–82) was nominally chosen to avoid major volcanic eruptions, in particular the Krakatoa eruption of 1883. However, coverage of the “water hemisphere” ([Boggs 1945](#)) is almost nonexistent in the 1860s [[Kennedy \(2014\)](#) and the Hadley Centre SST dataset, version 3 (HadSST3), gridded data]. Infilled records ([Hansen and Lebedeff 1987](#); [Rohde et al. 2013](#); [Cowtan and Way 2014](#)) can mitigate coverage issues for recent decades but can only meaningfully address data “holes” of up to ~1000 km in radius ([Hansen and Lebedeff 1987](#); [Cowtan et al. 2018](#)) and cannot reconstruct a missing hemisphere of data. Nineteenth-century temperatures are contingent on large “bucket corrections” to SST observations, the evolution of which are poorly constrained by metadata, and they show substantial differences between observational products ([Folland and Parker 1995](#); [Kent et al. 2017](#); [Cowtan et al. 2017](#)). An alternative early

---

*Corresponding author:* Kevin Cowtan, [kevin.cowtan@york.ac.uk](mailto:kevin.cowtan@york.ac.uk)

DOI: 10.1175/JCLI-D-18-0316.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

window (1930–50) used by [Lewis and Curry \(2018\)](#) spans the World War II period and is also the subject of large discrepancies among SST products ([Kent et al. 2017](#)).

We examine the impact of the choice of dates for the early and late windows and evaluate the impact of using short data windows rather than all of the data. The potential impact of volcanic events is addressed by application of the window method not to the observed data but to the difference between the observations and the mean of climate model simulations from phase 5 of the Coupled Model Intercomparison Project (CMIP5) using data from the historical and representative concentration pathway 4.5 (RCP4.5) scenarios. Masking the model outputs to match the observational coverage also allows us to control for the impact of changing coverage. [Lehner et al. \(2016\)](#) suggest that climate model simulations overestimate the volcanic response, although this may be a result of internal variability and other factors masking the volcanic response ([Stevenson et al. 2017](#); [Liu et al. 2018](#)). Linear regression was therefore used to remove the residual volcanic contribution to the difference temperature series by using the stratospheric aerosol optical depth ([Sato et al. 1993](#)) convoluted with an exponential response function with an  $e$ -folding time of 1 yr (determined by fit to the data). No correction was made for internal variability; however, if an El Niño term is included in the regression the remaining short-term features in the variability of warming with window choice are slightly reduced.

For this analysis we will focus on the University of East Anglia Climatic Research Unit–Hadley Centre global land-plus-ocean temperature dataset, version 4 (HadCRUT4) as the temperature product ([Morice et al. 2012](#)); however, similar issues arise with the other temperature products, and in the case for the Extended Reconstructed Sea Surface Temperature (ERSST)-based products the problems in the early record are more serious ([Cowtan et al. 2017](#)). We also used temperature data from 36 CMIP5 models with, in total, 107 historical realizations, extended using RCP4.5 simulations for the period 2006–16 and regridded onto a common  $1^\circ \times 1^\circ$  grid. We calculated a multimodel mean gridded temperature series over the 107 simulations using monthly surface air temperature estimates, that is, the “tas” field ([Taylor et al. 2012](#)) in CMIP5 terminology (similar results are obtained if all of the simulations for each model are averaged and then an average is calculated across the models). We then converted the temperatures to temperature anomalies using a 1961–90 baseline. We averaged blocks of  $5 \times 5$  grid cells to match the  $5^\circ$  HadCRUT4 grid and calculated a gridded difference map series between the HadCRUT4 gridded observations and the multimodel mean. Last, we determined the mean temperature difference for

the common coverage region by using the cosine-weighted mean of the observed grid cells.

A comparison of early window dates for the HadCRUT4 temperature data ([Morice et al. 2012](#)) is shown in [Fig. 1a](#), fixing the late window to 1995–2016, which is the longer option suggested by [Lewis and Curry \(2018\)](#) and is less affected by an uncorrected bias in ship observations ([Hausfather et al. 2017](#)). Coordinates represent the start and end dates of the early window, with red regions indicating that observations warmed more than model results and blue regions indicating that modeled results warmed more than observations for a given choice of early window. Different window choices can lead to the conclusion that the model results show significantly faster warming than the observations do or that the observations warm slightly faster than the model results, and this discrepancy is much larger than changes arising from presence or absence of a historical volcanic eruption in the window.

The experiment was repeated using land data only in [Fig. 1b](#). In this case the University of East Anglia Climatic Research Unit–Hadley Centre land temperature dataset, version 4 (CRUTEM4), observations (modified by the Hadley Centre to account for urbanization and exposure biases) warm faster than the models for a window of any reasonable length. The HadSST3 observations are compared with the model marine air temperatures in [Fig. 1c](#): these two tests show that variability in the results for different window dates arises primarily from the ocean data.

SST observations generally come from the top 10 m of the ocean and should strictly be compared with temperatures at a corresponding depth in the models. [Cowtan et al. \(2015\)](#) used the CMIP5 ocean surface temperature field (“tos” in CMIP5 nomenclature) for this purpose. Lewis and Curry argue that this field is not the top layer of the bulk ocean surface temperature. [Richardson et al. \(2016\)](#) examined 28 model configurations: in 22 of these configurations the tos field is identical (18 cases) or almost identical (4 cases) to the top layer of the bulk ocean temperature (“thetao” in CMIP5 nomenclature). We extend this analysis to 33 model configurations: for 20 of 33 model configurations the sea surface temperature field is essentially identical to the top layer of the bulk ocean temperature field. For 12 further model configurations (ACCESS1.0, ACCESS1.3, BCC\_CSM1.1, CSIRO-Mk3.6-0, EC-EARTH, GISS-E2-H, GISS-E2-H-CC, GISS-E2-R, GISS-E2-R-CC, MPI-ESM-LR, MPI-ESM-MR, MRI-CGCM3, MRI-ESM1, and NorESM1-M; expansions of acronyms are available online at <http://www.ametsoc.org/PubsAcronymList>) the differences between tos and the upper thetao are noiselike and do not impact the trend. The remaining model (GFDL-ESM2G)

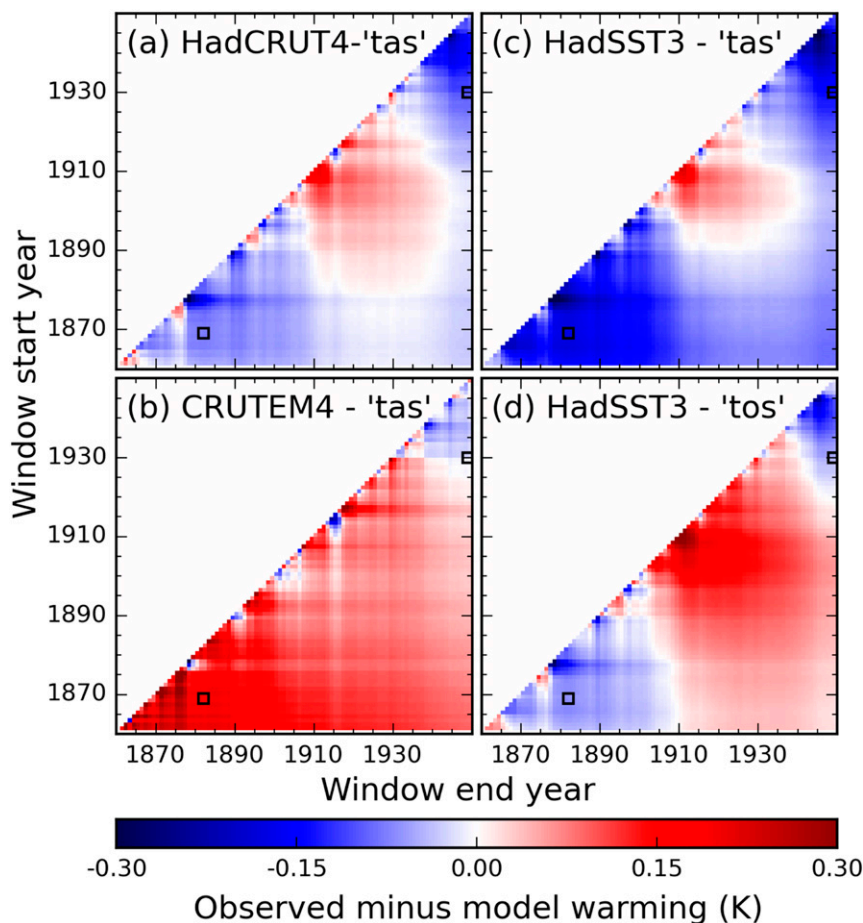


FIG. 1. Comparison of temperature change in observations relative to models for a range of early window start and end dates (the late window is fixed at 1995–2016): (a) blended temperature observations are compared with surface air temperatures from models, (b) land air temperature observations are compared with surface air temperatures from models, (c) sea surface temperature observations are compared with surface air temperatures from models, and (d) sea surface temperature observations are compared with sea surface temperatures from models. Open squares mark the [Lewis and Curry \(2018\)](#) windows. The difference between (c) and (d) is from SSTs warming slower than air temperatures. Red regions indicate that observations warm more than the model results, and blue regions indicate that modeled results warm more than observations, using a given window as a baseline. The plots in (a) and (c) incorporate the misleading comparison of marine air temperatures from models to SST observations, while (b) and (d) are like-with-like comparisons for land and ocean, respectively.

shows large differences between *tos* and upper *thetao* that are suggestive of a data deposition or processing error.

The effect of window choice for observed and modeled SSTs (as opposed to modeled air temperatures) is shown in [Fig. 1d](#). Use of model SSTs increases the warming of the observations relative to the models by approximately  $0.1^{\circ}\text{C}$  for any choice of window.

The period from 1850 to 1930 represents a change from the use of wooden buckets to poorly insulated canvas buckets in the measurement of SSTs, the latter requiring a large bias correction. The early features of

[Figs. 1c and 1d](#) could be explained if this change occurred primarily between 1890 and 1910, as suggested by comparison of SSTs with coastal weather station observations ([Jones et al. 1991](#); [Folland and Parker 1995](#); [Cowtan et al. 2017](#)). After World War II, HadSST3 may be affected by incorrect inference of some observation types and other biases ([Carella et al. 2018](#); [Davis et al. 2018](#)). Internal multidecadal variability may also contribute to the features of [Fig. 1d](#), although the Pacific contribution is likely to be small in the nineteenth century because of poor coverage, and

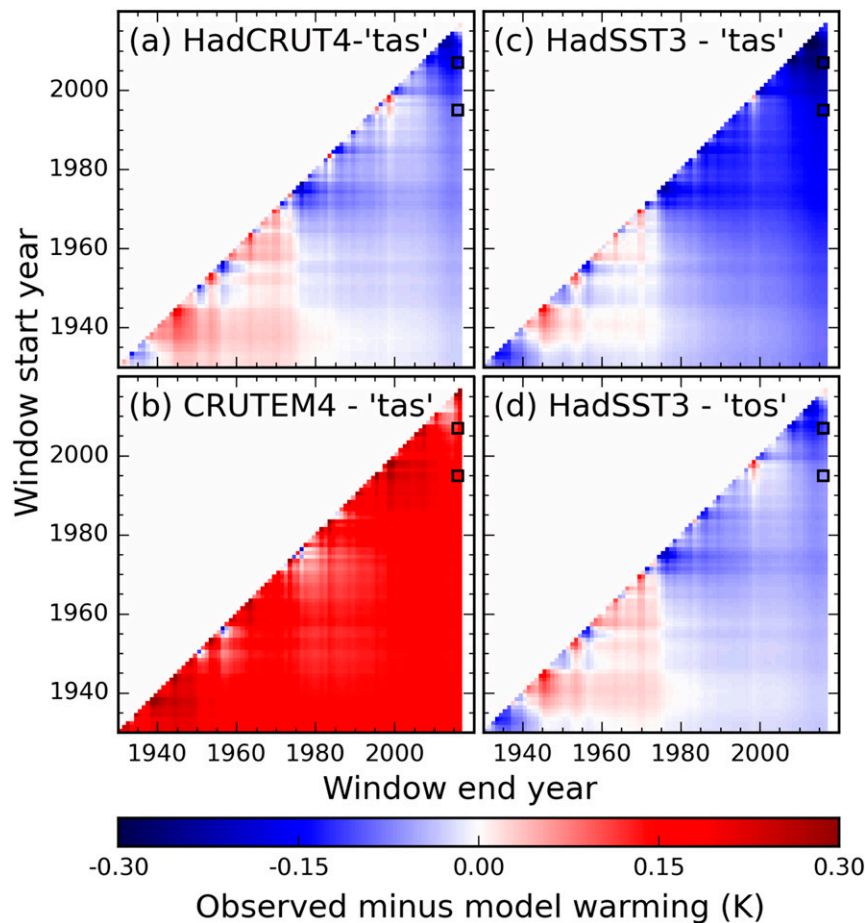


FIG. 2. As in Fig. 1, but exploring different choices for the late window while holding the early window fixed at 1869–82.

the coastal temperature difference is not localized to either the Pacific or Atlantic Oceans.

A similar experiment was conducted for the late window while holding the early window fixed at 1869–82 (Fig. 2). When using land data alone, all window choices of reasonable length lead to faster warming of the observations than the models. The sea surface temperature data show slower warming in the observations except for windows ending before 1975, because of the unusual warmth of HadSST3 relative to both models and ERSST between 1950 and 1980 (Kent et al. 2017; Cowtan et al. 2017; Carella et al. 2018; Davis et al. 2018). Windows starting after 2005 show a greater difference between observations and models: a residual bias in the sea surface temperatures for recent years (Hausfather et al. 2017) and the overestimation of forcings (Huber and Knutti 2014; Tatebe et al. 2019; Volodin and Gritsun 2018) are expected to contribute to a difference between modeled and observed warming for windows running to the present.

Multidecadal biases are present in all current SST products, including the ERSST temperature data (Huang et al. 2017) that are used in the other main temperature products not used by Lewis and Curry. ERSST shows little or no evidence of a lower early bias due to the use of wooden buckets (Kent et al. 2017), in contradiction of the observational metadata, suggesting the need for caution with respect to nineteenth-century temperatures in this product. ERSST is cooler than HadSST3 for the period 1930–50, except during World War II when it is too warm as a result of an uncorrected bias in the marine air temperatures and temporal smoothing in the ERSST algorithm suppressing the World War II bias correction (Cowtan et al. 2017).

The results of the window method are influenced by decisions concerning the criteria for window selection. We analyze the effect of window selection by evaluating the regression coefficient that fits the multimodel mean temperature change for the RCP4.5 simulations to the observational data, comparing model land air temperatures



with land-based observations, model marine air temperatures with SST observations, and model SSTs with SST observations. Regression coefficients fitting the corrected model data to the observations were determined for different data selections and are given in Table 1, with values of greater than 1 indicating observations warming faster than the models, and vice versa.

Land temperature observations warm faster than the models for any of the chosen data selections, with some variation resulting from window choice (i.e., the values in the CRUTEM/tas column of Table 1 are always greater than unity). SST observations warm more slowly than modeled marine air temperatures for long windows running to the present (i.e., the values in the HadSST/tas column are less than unity). SST observations warm slightly faster than modeled SSTs for long windows (i.e., the values in the HadSST/tos column are greater than unity). Regression coefficients using model SSTs are typically ~15% higher than those using marine air temperatures (based on the ratios of the HadSST/tos to the HadSST/tas columns when using long windows). Observed SSTs warm more quickly than modeled SSTs prior to the twenty-first century, but the difference is reduced on inclusion of the last 20 years of data, consistent with the underestimation of recent SST observations and the overestimation of forcings. The inclusion of the intervening decades of data mitigates most of the variability resulting from choice of the early window, but has limited benefit with respect to the late window because the rapid temperature change at the end of the record gives the final decades greater leverage in determining the regression coefficient.

In summary, the use of short time windows and the difference between air and sea surface warming, as indicated by temperature, can influence conclusions concerning whether observations are warming faster than indicated by models, with the differences primarily arising in the sea surface temperatures. Since warming in model results is strongly correlated with forcing, this also impacts TCR estimates determined using window methods. In comparisons between observations and climate model simulations, use of longer spans of data can reduce the impact of early window choice, but varying the end point of the data still affects the results (with the implication that conclusions from historical data can change in future). It is *vital* that use of historical temperature data for the estimation of climate sensitivity or internal variability be informed by the literature on the limitations and biases in those products, which generally incorporates *more recent results* than the datasets themselves. On the basis of current data it is not possible to conclude that models show faster warming than observations do, and as a result discrepancies

TABLE 1. Regression coefficients that scale the multimodel mean temperature change to fit the observations for the comparison of land air temperatures (“tas”) with land-based observations, model marine air temperatures (“tas”) with SST observations, and model SSTs (“tos”) with SST observations. The rows provide values for different subsets of the data, with the first six rows using two 20-yr windows and the last six rows using a single longer window. Values that are greater than 1 indicate that the observations are warming faster than the models.

Years	CRUTEM/tas	HadSST/tas	HadSST/tos
1861–80; 1998–2017	1.22	0.81	0.92
1881–1900; 1998–2017	1.12	0.83	0.95
1901–20; 1998–2017	1.05	1.05	1.21
1861–80; 1978–97	1.41	0.85	0.96
1881–1900; 1978–97	1.19	0.88	1.02
1901–20; 1978–97	1.06	1.30	1.52
1861–2017	1.11	0.90	1.03
1881–2017	1.07	0.91	1.05
1901–2017	1.04	0.94	1.09
1861–1997	1.18	1.03	1.17
1881–1997	1.13	1.10	1.27
1901–97	1.04	1.24	1.46

between model-based TCR estimates and those deduced from the observations must arise primarily from inconsistencies in TCR evaluation method, incompatibility of modeled and observed temperature estimates, and/or differences between the modeled and historical forcings.

The data computer code used in this paper, along with additional figures, is available online (<https://doi.org/10.15124/92466e73-6012-4ab4-ad10-cd7f4c075cb3>).

## REFERENCES

- Armour, K., 2017: Energy budget constraints on climate sensitivity in light of inconstant climate feedbacks. *Nat. Climate Change*, **7**, 331–335, <https://doi.org/10.1038/nclimate3278>.
- Boggs, S., 1945: This hemisphere. *J. Geogr.*, **44**, 345–355, <https://doi.org/10.1080/00221344508986498>.
- Carella, G., J. Kennedy, D. Berry, S. Hirahara, C. Merchant, S. Morak-Bozzo, and E. Kent, 2018: Estimating sea surface temperature measurement methods using characteristic differences in the diurnal cycle. *Geophys. Res. Lett.*, **45**, 363–371, <https://doi.org/10.1002/2017GL076475>.
- Cowan, K., and R. Way, 2014: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quart. J. Roy. Meteor. Soc.*, **140**, 1935–1944, <https://doi.org/10.1002/qj.2297>.
- , and Coauthors, 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>.
- , R. Rohde, and Z. Hausfather, 2017: Evaluating biases in sea surface temperature records using coastal weather stations. *Quart. J. Roy. Meteor. Soc.*, **144**, 670–681, <https://doi.org/10.1002/qj.3235>.
- , P. Jacobs, P. Thorne, and R. Wilkinson, 2018: Statistical analysis of coverage error in simple global temperature

- estimators. *Dyn. Stat. Climate Syst.*, **3**, dzy003, <https://doi.org/10.1093/climsys/dzy003>.
- Davis, L. L. B., D. W. J. Thompson, J. J. Kennedy, and E. C. Kent, 2018: The importance of unresolved biases in 20th century sea-surface temperature observations. *Bull. Amer. Meteor. Soc.*, **100**, 621–629, <https://doi.org/10.1175/BAMS-D-18-0104.1>.
- Folland, C., and D. Parker, 1995: Correction of instrumental biases in historical sea surface temperature data. *Quart. J. Roy. Meteor. Soc.*, **121**, 319–367, <https://doi.org/10.1002/qj.49712152206>.
- Hansen, J., and S. Lebedeff, 1987: Global trends of measured surface air temperature. *J. Geophys. Res.*, **92**, 13 345–13 372, <https://doi.org/10.1029/JD092iD11p13345>.
- Hausfather, Z., K. Cowtan, D. C. Clarke, P. Jacobs, M. Richardson, and R. Rohde, 2017: Assessing recent warming using instrumentally homogeneous sea surface temperature records. *Sci. Adv.*, **3**, e1601207, <https://doi.org/10.1126/sciadv.1601207>.
- Huang, B., and Coauthors, 2017: Extended Reconstructed Sea Surface Temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *J. Climate*, **30**, 8179–8205, <https://doi.org/10.1175/JCLI-D-16-0836.1>.
- Huber, M., and R. Knutti, 2014: Natural variability, radiative forcing and climate response in the recent hiatus reconciled. *Nat. Geosci.*, **7**, 651–656, <https://doi.org/10.1038/ngeo2228>.
- Jones, P., T. Wigley, and G. Farmer, 1991: Marine and land temperature data sets: A comparison and a look at recent trends. *Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations*, M. Schlesinger, Ed., Elsevier, 153–172.
- Kennedy, J., 2014: A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Rev. Geophys.*, **52**, 1–32, <https://doi.org/10.1002/2013RG000434>.
- Kent, E. C., and Coauthors, 2017: A call for new approaches to quantifying biases in observations of sea-surface temperature. *Bull. Amer. Meteor. Soc.*, **98**, 1601–1616, <https://doi.org/10.1175/BAMS-D-15-00251.1>.
- Knutti, R., M. Rugenstein, and G. Hegerl, 2017: Beyond equilibrium climate sensitivity. *Nat. Geosci.*, **10**, 727–736, <https://doi.org/10.1038/ngeo3017>.
- Lehner, F., A. Schurer, G. Hegerl, C. Deser, and T. Frölicher, 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>.
- Lewis, N., and J. Curry, 2018: The impact of recent forcing and ocean heat uptake data on estimates of climate sensitivity. *J. Climate*, **31**, 6051–6071, <https://doi.org/10.1175/JCLI-D-17-0667.1>.
- Liu, F., J. Li, B. Wang, J. Liu, T. Li, G. Huang, and Z. Wang, 2018: Divergent El Niño responses to volcanic eruptions at different latitudes over the past millennium. *Climate Dyn.*, **50**, 3799–3812, <https://doi.org/10.1007/s00382-017-3846-z>.
- Morice, C., J. Kennedy, N. Rayner, and P. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geophys. Res.*, **117**, D08101, <https://doi.org/10.1029/2011JD017187>.
- Otto, A., and Coauthors, 2013: Energy budget constraints on climate response. *Nat. Geosci.*, **6**, 415–416, <https://doi.org/10.1038/ngeo1836>.
- Richardson, M., K. Cowtan, E. Hawkins, and M. Stolpe, 2016: Reconciled climate response estimates from climate models and the energy budget of Earth. *Nat. Climate Change*, **6**, 931–935, <https://doi.org/10.1038/nclimate3066>.
- Rohde, R., and Coauthors, 2013: Berkeley Earth temperature averaging process. *Geoinf. Geostat. Overview*, **13**, 20–100, <https://doi.org/10.4172/2327-4581.1000103>.
- Sato, M., J. Hansen, M. McCormick, and J. Pollack, 1993: Stratospheric aerosol optical depths, 1850–1990. *J. Geophys. Res.*, **98**, 22 987–22 994, <https://doi.org/10.1029/93JD02553>.
- Stevenson, S., J. Fasullo, B. Otto-Bliesner, R. Tomas, and C. Gao, 2017: Role of eruption season in reconciling model and proxy responses to tropical volcanism. *Proc. Natl. Acad. Sci. USA*, **114**, 1822–1826, <https://doi.org/10.1073/pnas.1612505114>.
- Storelvmo, T., T. Leirvik, U. Lohmann, P. Phillips, and M. Wild, 2016: Disentangling greenhouse warming and aerosol cooling to reveal Earth's climate sensitivity. *Nat. Geosci.*, **9**, 286–289, <https://doi.org/10.1038/ngeo2670>.
- Tatebe, H., and Coauthors, 2019: Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geosci. Model Dev.*, **12**, 2727–2765, <https://doi.org/10.5194/gmd-12-2727-2019>.
- Taylor, K., R. Stouffer, and G. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, <https://doi.org/10.1175/BAMS-D-11-00094.1>.
- Volodin, E., and A. Gritsun, 2018: Simulation of observed climate changes in 1850–2014 with climate model INM-CM5. *Earth Syst. Dyn.*, **9**, 1235–1242, <https://doi.org/10.5194/esd-9-1235-2018>.