Connecting climate model projections of global temperature change with the real world

Article

Published Version

www.reading.ac.uk/centaur

CONNECTING CLIMATE MODEL PROJECTIONS OF GLOBAL TEMPERATURE CHANGE WITH THE REAL WORLD

by Ed Hawkins and Rowan Sutton

Careful consideration of sensitivities to the choice of climate reference period is required to reliably compare climate models with observations and to produce robust projections of future climate.

Current state-of-the-art global climate models produce different values for Earth's mean temperature. For this reason, projections of changes in Earth's temperature over time are usually presented relative to a reference period, following Hansen et al. (1988). It is not widely appreciated, especially outside the climate science community, that the choice of reference period has important consequences for conclusions about such basic questions, such as: are

climate model simulations of the past consistent with observations and what do climate models predict for the future? The importance of these questions has been highlighted by the recent debate about the difference between observed and projected multimodel-mean warming of global surface temperatures (Stott et al. 2013; Huber and Knutti 2014; Schmidt et al. 2014).

In this article, we demonstrate how the choice of reference period affects the conclusions drawn in relation to both these questions and discuss consequences for connecting climate model projections of global temperature change to the real world. We also discuss further the reasons why anomalies have long been used for comparing observational data and model output, noting important limitations. Last, we discuss the implications for near-term and long-term projections of global-mean surface temperature provided in the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5) and provide recommendations for good practice.

WHY IS A REFERENCE PERIOD NEEDED FOR OBSERVATIONS? In 1935, the World Meteorological Organization (WMO) first discussed defining a recommended "normal" period to set a

AFFILIATIONS: HAWKINS AND SUTTON—NCAS-Climate, Department of Meteorology, University of Reading, Reading, United Kingdom

CORRESPONDING AUTHOR: Ed Hawkins, Department of Meteorology, University of Reading, Reading RG6 6BB, United Kingdom
E-mail: e.hawkins@reading.ac.uk

standard and allow comparisons between different observational data. The length of a normal period was chosen to be 30 years, and the period 1901–30 was selected initially (Trewin 2007). The WMO has recently adopted a two-tier approach, maintaining 1961–90 as a fixed standard reference period, along with a regularly updated period, which is currently 1981–2010.

Observation-based temperature datasets also use a reference period because mean temperatures can vary over very short spatial scales (~1 km), whereas the correlation scale for temperature anomalies is usually much larger (~1,000 km) (Hansen and Lebedeff 1987). Fewer stations are therefore required to estimate *changes* in global temperatures (Jones et al. 1997). For example, Callendar (1938) first demonstrated the Earth was warming using just 147 stations, and his calculations match modern estimates well (Hawkins and Jones 2013).

A number of factors enter the decision about an appropriate observational reference period, for example, to be representative of the most recent conditions but long enough not to be overly influenced by random fluctuations, to be a period the public can relate to, to not need updating too often, to maximize the number of observations available and be simple to calculate. In addition, the mean over a reference period is insufficient to represent the climate as higher-order statistics are also required (e.g., Landsberg 1944). Huang et al. (1996) demonstrated that a normal period updated every year was optimal for making predictions for the following year.

Reference periods are also used to aid communication of the unusual (or not) nature of an observation, such as an increase in global temperatures, or of a particular event such as an extreme flood or heatwave. For example, the warm global temperatures of 2014 were not particularly unusual compared to other years since 2000 but were very unusual compared to temperatures before 1900.

Observation quality is also important for the choice of reference period. The uncertainty on the observed estimate of global and regional temperatures is larger in the past, especially pre-1900. In the case of the Hadley Centre/Climatic Research Unit, version 4.3 (HadCRUT4.3), dataset (Morice et al. 2012), the reference period is 1961–90 because of the high availability of observations during this period. However, the surface temperature observations available for this period still do not cover the whole planet. Jones et al. (1999) estimated the observed 1961–90 global-mean temperature as 14.0° ± 0.5°C, and Fig. 1 illustrates that different atmospheric reanalyses have global-mean temperatures within that range.

WHY IS A REFERENCE PERIOD NEEDED FOR MODEL SIMULATIONS?

Simulating the absolute value of many climate variables, such as global-mean surface temperature, is challenging because they represent the balance between many different physical processes. It is not currently possible to tune global climate models (GCMs) to produce accurate values for all climate variables. However, some variables have a higher priority than others. For example, it is essential to produce a model with a near-zero net top-of-atmosphere (TOA) energy balance. Without such a balance the model climate drifts and does not provide a stable baseline against which to measure the response to changing radiative forcings. As discussed below, the simulated value for global-mean temperature matters, but it is less essential—for projections of global-mean temperature—to simulate the observed value precisely. Thus, global-mean temperature is generally given less weight than TOA energy balance when climate models are tuned (although there are some exceptions; e.g., Mauritsen et al. 2012). As a result, the range in simulated global-mean temperatures is far larger than the observational uncertainty.

It may be a surprise to some readers that an accurate simulation of global-mean temperature is not necessarily an essential prerequisite for accurate global temperature projections. Some supporting evidence comes from climate models and theoretical considerations. The mean global temperatures among simulations of the historical period with the latest phase 5 of the Coupled Model Intercomparison Project (CMIP5) GCMs (see appendix A) differ by up to 3 K (with a standard deviation of 0.7 K), but the changes over time are similar (Fig. 1). Importantly, there is no robust correlation between projected future warming and historical simulated mean global temperature in the CMIP5 simulations (Fig. 2), although there are no simulations with a high mean global temperature and large future warming in this particular set of GCMs. In addition, Mauritsen et al. (2012) created four different parallel versions of the MPI-ESM-LR, tuned differently, and found only modest variations in climate sensitivity across the ensemble. This evidence suggests that these differences in mean global temperature may not be crucial for projecting future global temperature changes, given current uncertainties in climate feedbacks [also see Fig. 9.42 of IPCC AR5 (Flato et al. 2013) and the blog discussion of Schmidt (2014)].

Theoretical insight into these climate model results is provided in appendix B, which uses a simple 1D energy balance model to show that differences in mean global temperature are relatively unimportant
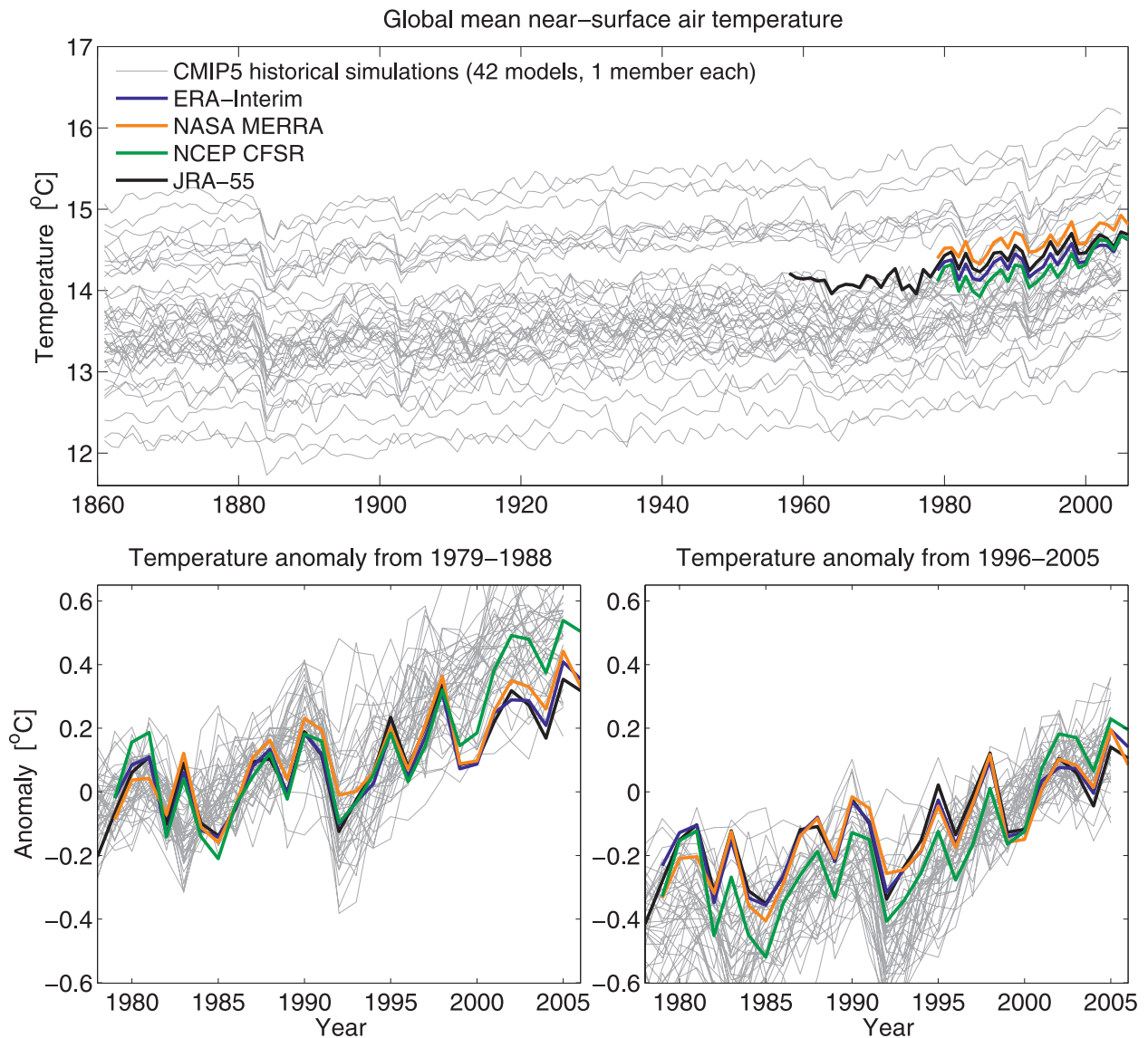
Fig. 1. (top) Global-mean 2-m air temperature from CMIP5 historical simulations (gray, 1861–2005) and various reanalysis estimates (colors; from Saha et al. 2010; Dee et al. 2011; Rienecker et al. 2011; Kobayashi et al. 2015). (bottom) Comparing the same data as temperature anomalies, using two reference periods (1979–88 and 1996–2005).

for projections of global-mean temperature, if the feedbacks are linear. However, there is much discussion in the literature about the extent to which feedbacks may be nonlinear and what the implications would be (e.g., Good et al. 2012; Gregory et al. 2015; Bloch-Johnson et al. 2015).

The evidence discussed above summarizes the arguments that are typically presented in support of the common practice of using a reference period when comparing climate models with observations and when generating projections of global-mean temperature. However, it does not tell us what specific reference period we should choose and to what extent the choice matters. We turn to these issues next.

WHY DOES THE CHOICE OF REFERENCE PERIOD MATTER? It is standard practice when comparing simulations of climate change with observed changes, and with each other, to use a common reference period and define "anomalies," for example,

$$\Delta T(t) = T(t) - \overline{T_{\text{ref}}} \ , \qquad (1)$$

where $T(t)$ is a time series of a particular variable, $\overline{T_{\text{ref}}}$ is the time average over a reference period, and $\Delta T(t)$ is the anomaly. This procedure is usually performed on the observations and any model simulations for the same reference period.
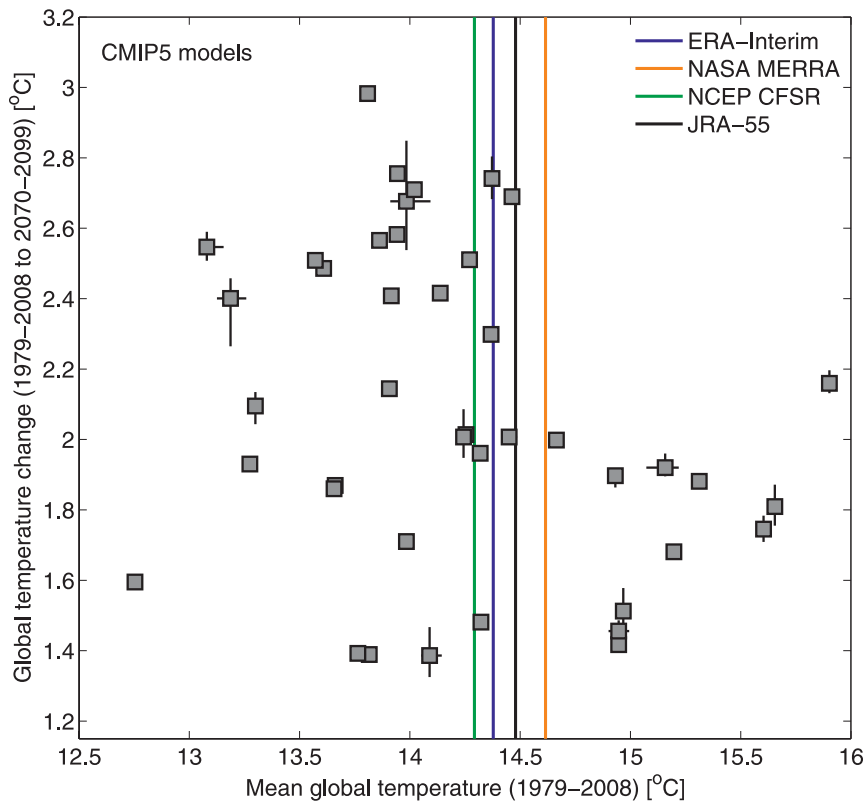
FIG. 2. Mean global surface temperature from CMIP5 historical simulations over 1979–2008 is not significantly correlated with the change from 1979–2008 to 2071–2100. The gray squares show the ensemble mean for each of the 42 CMIP5 models, with the error bars representing the minimum–maximum range from within each model's own ensemble, where available. The colored lines show the reanalysis estimates as in Fig. 1.

Figure 1 illustrates that the value of $\Delta T(t)$ changes when using different reference periods. In addition, the relative comparison of the different atmospheric reanalyses with each other, and with the simulations, also changes. For instance, in the example shown in Fig. 1, the simulations appear mostly warmer than the reanalyses with one choice of reference period but appear mostly cooler than the reanalyses with an alternative choice. Further, the NCEP CFSR reanalysis appears to be a slight outlier, but whether these differences are most apparent at the start or end of the simulation depends on the reference period. There is clearly sensitivity to the choice of reference period in any similar comparison (also see sidebar on "Illustrating the effect of reference period choice").

Evaluating historical simulations. One important test of the climate models used by the IPCC is their ability to simulate the

climate of the instrumental period since around 1850. This evaluation depends strongly on the choice of reference period used.

Figure 3 shows the CMIP5 simulations of global-mean temperature from 1861 to 2005, with different percentile ranges denoted by the blue bands. The HadCRUT4.3 observations (Morice et al. 2012) and associated uncertainties are shown in black and gray, respectively. However, HadCRUT4.3 is not spatially complete. Cowtan and Way (2014, hereafter CW14) recently used spatial interpolation to fill the gaps in HadCRUT4.3, and this CW14 dataset is shown in red (also see appendix C). The four panels perform the comparison with different reference periods, first using the whole period (1861–2005) and then using three different 30-yr periods.

The following observations may be made: First, the percentile ranges clearly change with the choice of reference period—they tend to be narrower during the chosen reference period than at other times. Second, the observations fall outside the 5%–95% ranges at different times when using the different reference periods.

TABLE 1. The number of years (in the 1861–2005 period) that the median of HadCRUT4.3 falls outside the specified CMIP5 confidence interval for different reference periods.

Reference period	>95%	>75%	<25%	<5%
Number expected	7	36	36	7
1861–2005	3	35	37	4
1861–90	1	10	50	8
1911–30	13	68	10	0
1961–90	1	27	31	2
1980–99	0	21	27	1
1976–2005	1	35	19	0
1986–2005	1	25	18	1

We illustrate the sensitivity to the choice of reference period with an analogy and schematic. Different time series can be thought of as stiff (and "wiggly") wires that are required to pass through a fixed length of tube. Different length reference periods correspond to tubes of different lengths, with longer tubes required to have wider diameters. There is little constraint on how the wires spread outside the tube, and for longer tubes, there is less constraint on how they vary within the tube, thanks to a larger

diameter. In the extreme, a tube that is one time point long would have zero diameter because all of the wires can be forced to pass through the same point. The constraint on where the wires are positioned vertically, relative to each other and relative to the tube, varies as the tube is slid horizontally along the loose bundle of wires.

Interpreting the wires as time series of annual-mean global-mean temperature illustrates the effect of choosing a reference period (Fig. SB1). Which is the warmest time series at later times

depends on the choice of reference period (or tube position). The black dashed lines show the range of possible futures for a larger set of time series demonstrating that the uncertainty shrinks for later reference periods (as discussed later in Fig. 7).

An animated version of Fig. SB1 is shown in Fig. ES1 (more information can be found online in the supplemental information available at http://dx.doi.org/10.1175/BAMS-D-14-00154.2), which also highlights the sensitivity to length of reference period.

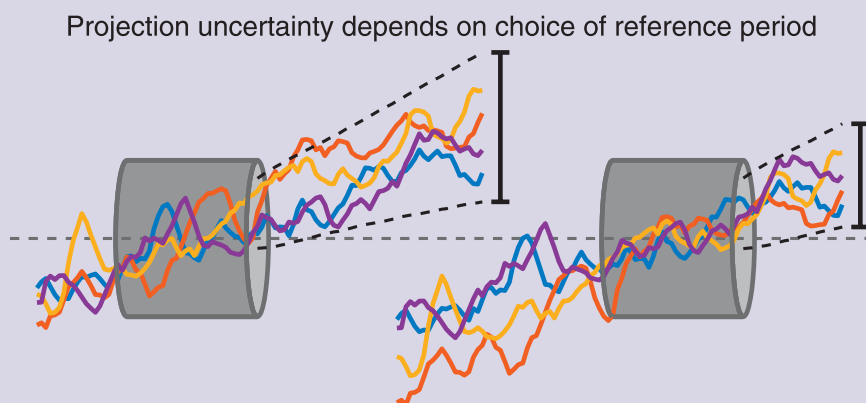## Projection uncertainty depends on choice of reference period

FIG. SB1. A schematic showing how a set of time series (or wires) behave for two different choices of reference period (or tube). Note that the ordering of the wires changes for the two different choices. The black dashed lines and error bars illustrate the range of a larger set of wires. An animated version of this figure is available in the online supplemental information.

Hence, any conclusions about the consistency of models and observations that may be inferred from analyses of this type are sensitive to the choice of reference period.

For example, there has been much attention on global temperatures over the past 15 years, which have risen more slowly than projected by the mean of the CMIP5 simulations (Fyfe et al. 2013). Figure 4 highlights that exactly where the most recent decade of observations falls within the CMIP5 simulated range is dependent on the choice of reference period but that they are always toward the lower end of the range. [However, appendix C highlights that part of the difference between the multimodel mean and observations is because the comparison is not quite like with like because of the incomplete coverage of the observations and the type of observation used (Hawkins 2013; Cowtan et al. 2015).]

The importance of this comparison is highlighted by an article in the media that stated that there is "irrefutable evidence that official predictions of global climate warming have been catastrophically flawed" (Rose 2013), based on a version of Fig. 4 for one particular choice of reference period that was published on a blog (Hawkins 2013). Other subsequent media articles more correctly discussed the implications of the most recent period using the same figure (e.g., *Economist* 2013).

There are different frameworks to interpret multimodel ensembles of climate simulations (e.g., Annan and Hargreaves 2010; Sanderson and Knutti 2012). Here, we consider a simple way of evaluating the reliability of the CMIP5 ensemble by examining whether the observations fall within each percentile the appropriate number of years. For example, in a reliable ensemble, the observations should be above

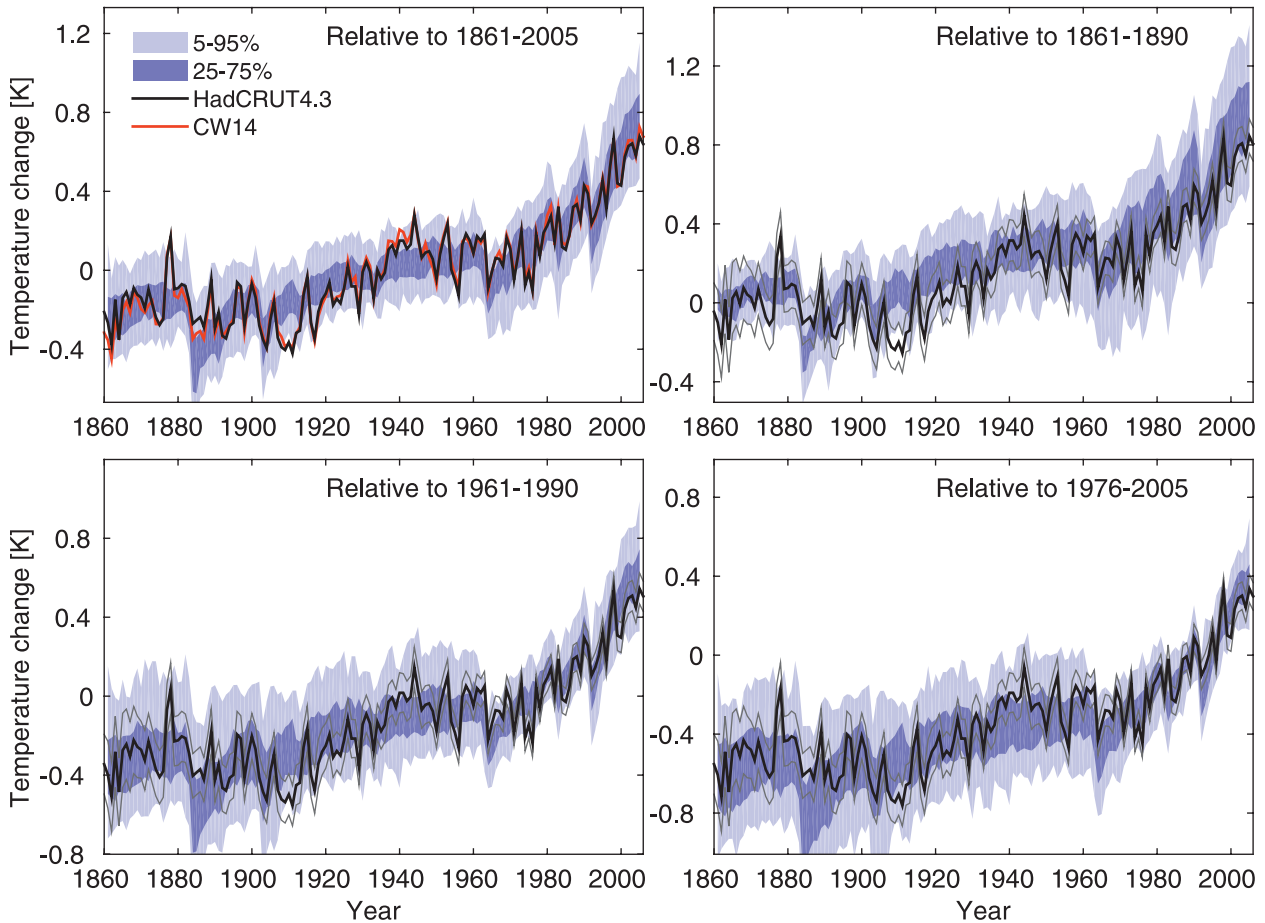Global mean surface air temperature: CMIP5 historical & observations

FIG. 3. Comparing CMIP5 historical simulations (42 models) with the HadCRUT4.3 (Morice et al. 2012) and CW14 observations using four different reference periods. (top left) Using 1861–2005 (the whole period). (top right),(bottom) Using different 30-yr periods as labeled, with the HadCRUT4.3 uncertainties in gray.

the 95th (or below the 5th) percentile about 1 year in 20 (or 5% of the time). Table 1 shows the number of years that fall outside various percentile ranges for a range of reference periods, including the whole period (1861–2005) and subperiods. When using the whole period, the CMIP5 ensemble is close to reliable, and perhaps slightly too wide in the tails, as indicated by the smaller number of years outside the 5th and 95th percentiles than expected. However, when evaluated using other reference periods, the ensemble appears far from reliable. This behavior is likely due to the phasing of internal variability in the particular realization of climate that we have observed and will also be influenced by errors in the specified historical forcings and in the simulated response to those forcings. One might conclude that the reference period should be as long as possible to reduce the influence of variability, but a counter argument is that both the forcing uncertainties (e.g., Carslaw et al. 2013; Stevens 2013)

and observation uncertainties (Morice et al. 2012) may be larger further back in time (also see appendix C). An alternative approach to assess reliability is to use trends in temperature, which are independent of the reference period (e.g., van Oldenborgh et al. 2013; Marotzke and Forster 2015). However, the analysis of trends is also influenced by the forcing uncertainties.

Projections of global-mean temperature. Future projections derived from climate models are similarly sensitive to the choice of reference period. The IPCC AR5 used a 1986–2005 reference period for generating climate projections, but previous assessment reports used earlier periods. To understand the impact of the choice of reference period on projections it is helpful to express results relative to a common baseline.

In the United Nations Framework Convention on Climate Change (UNFCCC) process, the change in

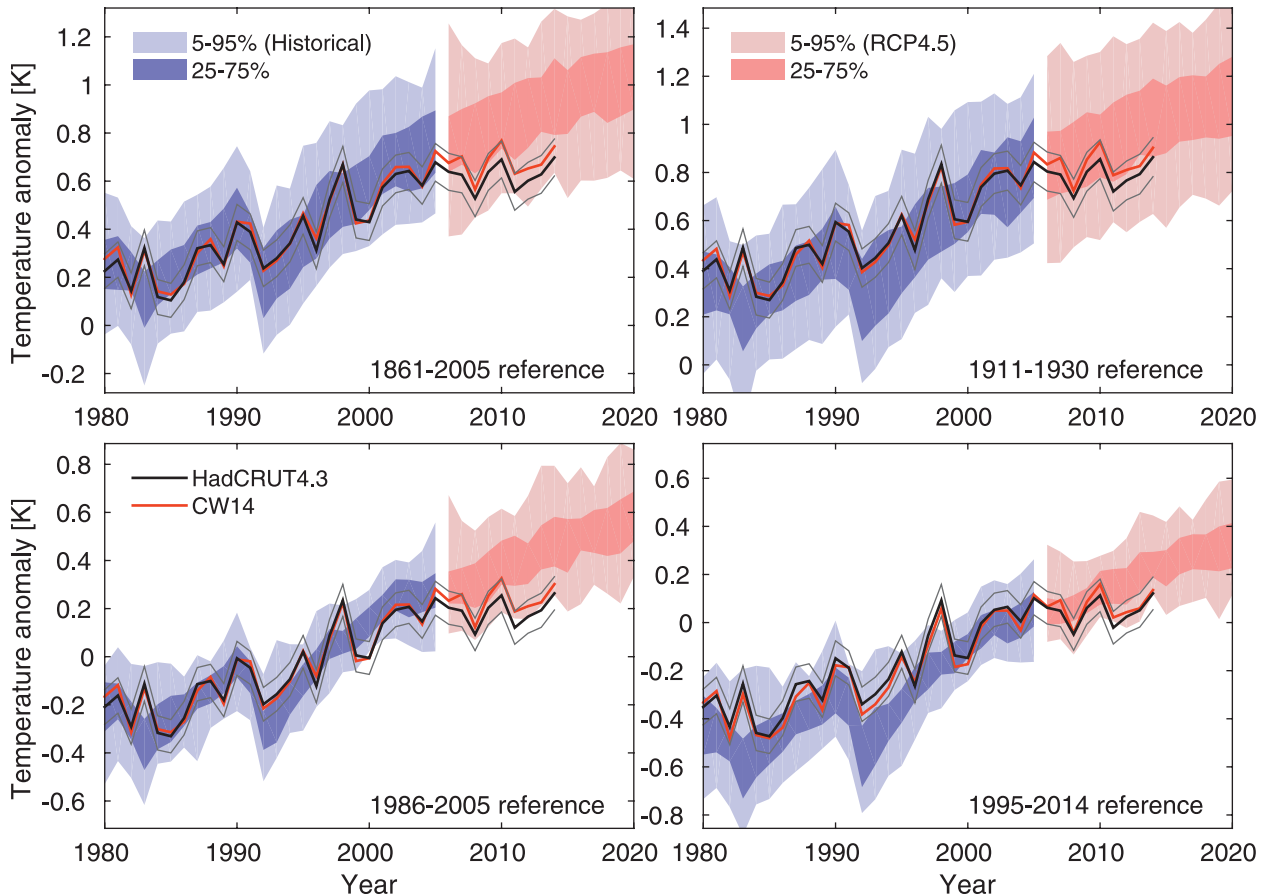Global mean surface air temperature: CMIP5 historical & observations

Fɪɢ. 4. Comparing CMIP5 historical and RCP4.5 simulations (42 models) with the HadCRUT4.3 (Morice et al. 2012) and CW14 observations, using four different reference periods as labeled. The position of the observations within the CMIP5 ensemble depends strongly on the reference period choice.

global-mean temperature since preindustrial times has become an important metric for discussions of mitigation policy. A difficulty with this metric is that preindustrial climate is not well defined because of a lack of observations before 1850 and a nonstationary climate due to natural external forcings such as solar variability and volcanic eruptions. However, a pragmatic approach is to use an early period in the instrumental record, such as 1850–1900, to define a baseline. Such a baseline should not strictly be described as preindustrial but does provide a useful reference point and was used in IPCC AR5 (Kirtman et al. 2013).

Projections relative to such a "preindustrial" baseline ΔW_{future} can be constructed in two ways. First, the raw model output can be referenced to the preindustrial period:

$$\Delta W_{\text{future}}(t) = T_{\text{mod}}(t) - \overline{T_{\text{mod,pre-ind}}} . \qquad (2)$$

This is perhaps the simplest method but may not be optimal because both the observations and radiative forcings in the past are uncertain, as discussed above. Instead, Joshi et al. (2011) constructed projections by combining the observed warming from a preindustrial period to a recent reference period and used the model projections to project future warming relative to the same recent reference period. This reduces the impact of the uncertainty in past radiative forcings and ties the projections to more recent observations. This approach was also used by Vautard et al. (2014) when considering changes in European temperatures and was adopted by the IPCC AR5 (Kirtman et al. 2013).

Using this approach, the observed warming up to the chosen reference period is

$$\Delta W_{\text{obs}}(t) = \overline{T_{\text{obs,ref}}} - \overline{T_{\text{obs,pre-ind}}} , \qquad (3)$$

and the simulated temperature anomaly above the preindustrial baseline is then

$$\Delta W_{\mathrm{future}}(t) = \Delta T_{\mathrm{mod}}(t) + \Delta W_{\mathrm{obs}}, \qquad (4)$$

$$= T_{\mathrm{mod}}(t) - \overline{T_{\mathrm{mod,ref}}} + \overline{T_{\mathrm{obs,ref}}} - \overline{T_{\mathrm{obs,pre\text{-}ind}}}. \qquad (5)$$

As both T_{mod} and $\overline{T_{\mathrm{obs,pre\text{-}ind}}}$ are reference period independent, an important quantity is

$$\Delta W_{\mathrm{diff}} = \overline{T_{\mathrm{obs,ref}}} - \overline{T_{\mathrm{mod,ref}}}. \qquad (6)$$

If this quantity were constant for any choice of reference period, then the reference period would not matter. However, it is not constant because that would require a perfect correlation between T_{obs} and T_{mod}.

Figure 5 shows the effect of choice of reference period for four different GCMs. The projected global temperatures have a strong dependence on the reference period for some GCMs: the impact on projected temperature changes relative to the baseline can be as much as 0.5 K. Other models show much less sensitivity. The black line, which is often the warmest, uses an early reference period and is close to the simple approach of Eq. (2). In HadGEM2-ES, it is the chosen IPCC AR5 reference period that is warmest. Note that a strong dependence on reference period is likely due to an incorrect simulation of the forcings or feedbacks, but a weak dependence could simply be due to cancelling errors. Large-amplitude internal decadal variability may also be important.

Figure 6 shows how ΔW_{diff} changes during the historical simulations for various CMIP5 models using rolling 30-yr reference periods. It is clear that different models behave in very different ways, warming more or less than the observations at different times.

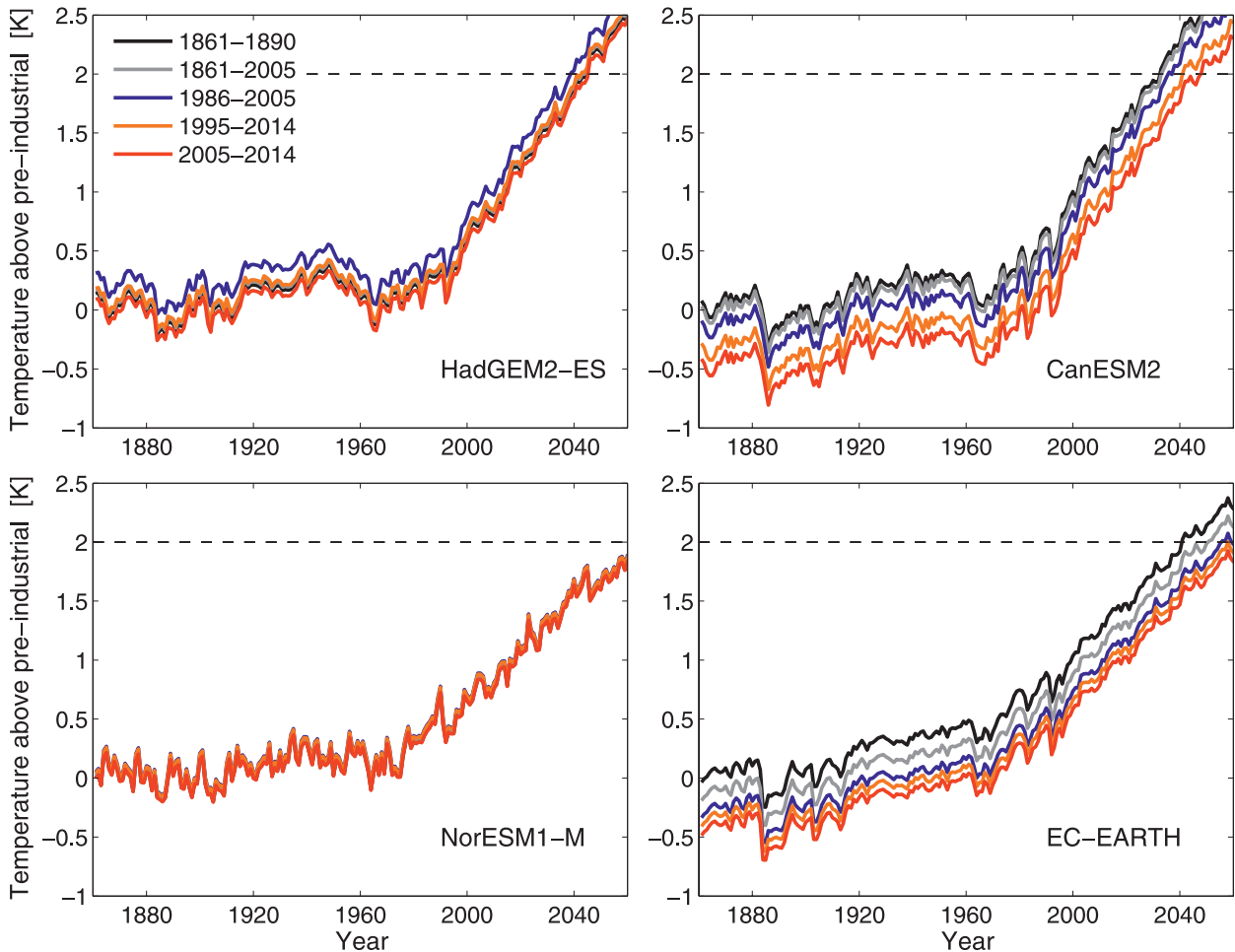## Global mean surface air temperature above pre–industrial using various reference periods

FIG. 5. **Comparing projections (RCP4.5) using five different reference periods for four example GCMs, showing the ensemble means. Note that the observed change is used as the anomaly from 1850 to 1900 to the chosen reference period. Most of the CMIP5 GCMs behave more like CanESM2 and EC-EARTH than NorESM1-M.**

Note, for example, that ΔW_{diff} generally increases in the early to mid-twentieth century, when the observations warm faster than the simulations and vice versa during the recent "slowdown."

This metric may help understand how the CMIP5 models are responding differently to different types of radiative forcing. For example, consider two models with the same overall warming, but with different amplitudes of response to aerosols or volcanic eruptions. The evolution of temperature change over the twentieth century will look different and ΔW_{diff} will change substantially with time. This metric merits further investigation.

Assessing multimodel projections. From the previous results, it is clear that the choice of reference period will influence both the mean and the range of the projected global-mean temperature. Figure 7 shows projections of global-mean temperature using representative concentration pathway (RCP) 4.5 (Thomson et al. 2011) for four different reference periods. Note that using an early 1861–90 reference period produces a larger magnitude and wider range for the future than more recent reference periods. The reduction in ensemble spread when using a more recent reference period is because there is less time for the ensemble to diverge (see appendix D). The 1986–2005 period was used by IPCC AR5, and $\Delta W_{\text{obs}} = 0.61$ K in this case (Kirtman et al. 2013). Updating to a more recent 1995–2014 reference period reduces the projections by about 0.1 K. The choice of reference period affects the bounds on the projected range of global-mean temperature for specific time periods (e.g., 2016–35, 2046–65, and 2080–99, as indicated in the figure and as used in AR5) by up to 0.2 K. As a proportion of the total projected change (relative to the reference period or to the baseline), this sensitivity is considerably larger for the near term (2016–35) than for the long term (2080–99).

The most commonly used magnitude of global temperature change discussed in the context of climate change policy is 2°C above preindustrial. Using the projections and an early 1861–90 reference period, the projected median year of crossing this threshold is 2049. However, as discussed above, this choice is not likely to be optimal for making future projections. Using alternative choices, the projected median year of crossing this threshold changes from 2052 using the 1861–2005 reference period to 2063 for the most recent reference period—an apparent delay of a decade.

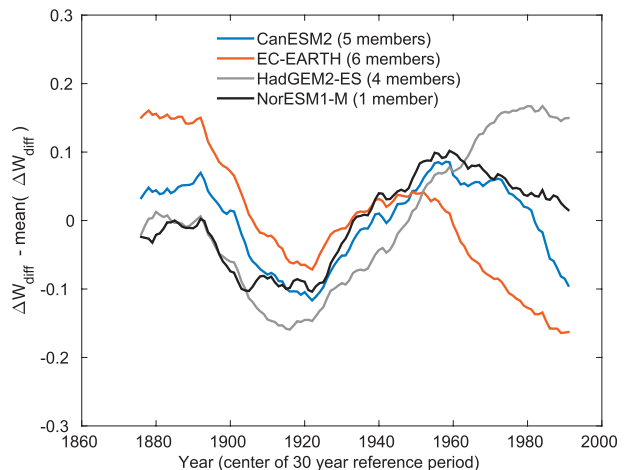Which is the most appropriate reference period to use to make such projections? There is no

FIG. 6. The change in ΔW_{diff} from its time mean for rolling 30-yr reference periods. Note the different behavior for the different models, which are shown in the colors for the ensemble mean.

straightforward answer as it will depend on the role of natural variability in recent observed changes as well as the simulated response to both greenhouse gases and volcanic eruptions. It will also depend on the quality of the observations and radiative forcings as discussed above. An additional issue is sensitivity to the length of the reference period. This is explored in appendix D, which suggests that the 20-yr period length used by the IPCC AR5 is a reasonable choice, but the optimal choice depends on the climate variable and region of interest. In summary, the sensitivity to choice of the reference period in any similar analysis needs to be examined.

Projections of global-mean temperature presented in the IPCC AR5. The IPCC AR5 presented assessed likely ranges for global-mean temperature in the near-term (Kirtman et al. 2013) and long term (Collins et al. 2013), where "likely" refers to >66% probability of occurrence. For the long term, ranges were presented for each of the RCP scenarios. Importantly, the likely ranges were based on an assessment of the all the evidence available at the time. This evidence included, but was not limited to, CMIP5 climate model projections expressed relative to the 1986–2005 reference period. For the near-term assessment, the sensitivity to the choice of reference period was discussed explicitly (Kirtman et al. 2013, section 11.3.6.3). Taken together with other lines of evidence, this resulted in the assessed likely range for global-mean temperature in 2016–35 being significantly cooler than was suggested by the "raw" CMIP5 projections expressed relative to 1986–2005. For the long-term assessment, the sensitivity to the reference period was not discussed
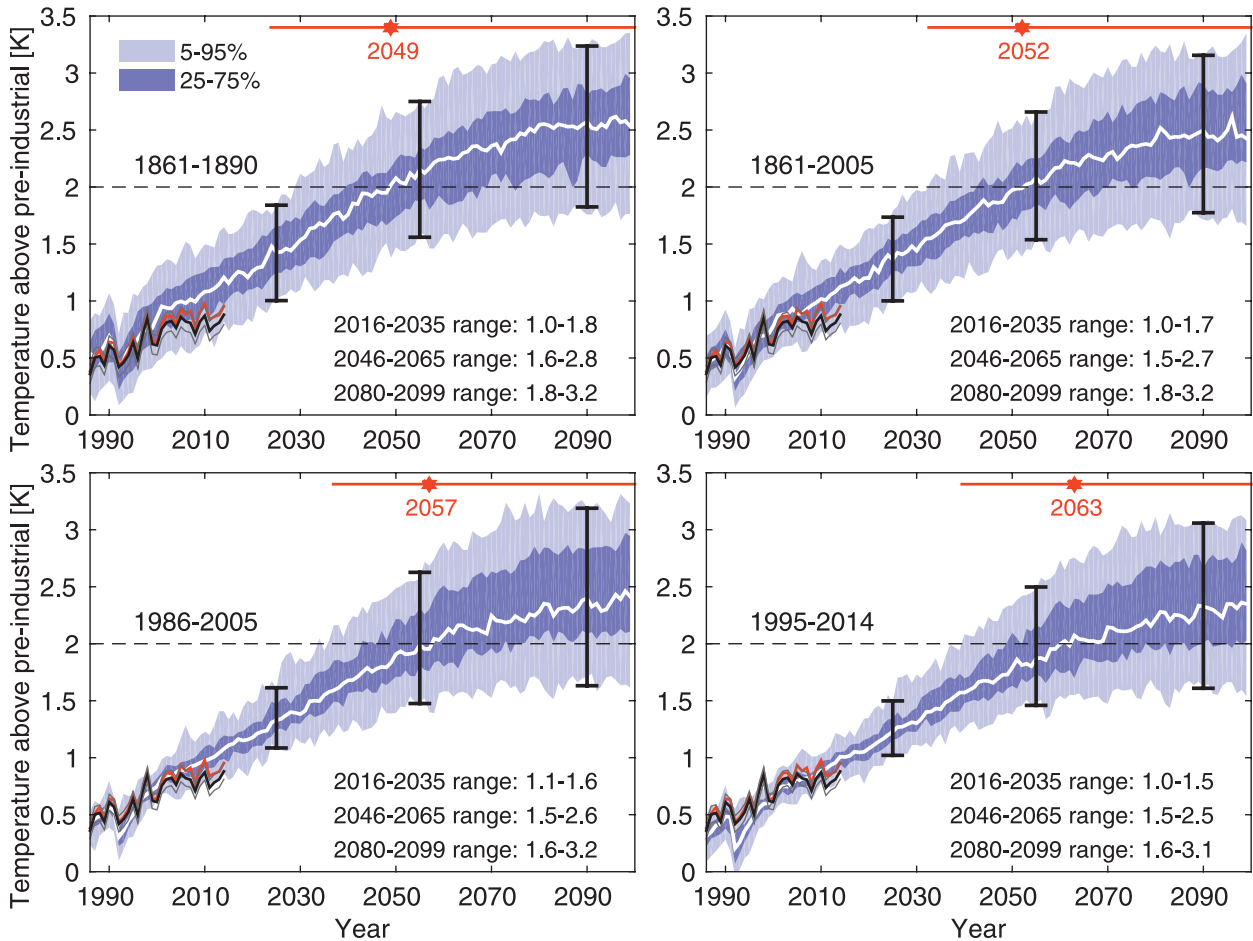
FIG. 7. Comparing CMIP5 projections (RCP4.5, 42 models) for the future using four different reference periods. Red stars indicate projected time of crossing 2 K above preindustrial (defined as 1850–1900) with red bars representing the 5%–95% range. Black error bars show 5%–95% temperature ranges for defined periods as indicated in legend. Observations (black) and observational uncertainties (gray) shown for HadCRUT4.3 (Morice et al. 2012). The CW14 observations are shown in red.

explicitly, but as noted above, this sensitivity is a much smaller proportion of the change signal than is the case for the near term, except for RCP2.6. In addition, the sensitivities described in this article are unlikely to affect any of the IPCC AR5 assessment statements on the likelihood of crossing particular temperature levels by certain times because these assessments were based on conservative assumptions.

REGIONAL TEMPERATURES AND OTHER CLIMATE VARIABLES. The previous sections have focused entirely on projections of global-mean temperature, which is an important variable for summarizing future climatic changes. However, the impacts of climate change depend strongly on changes in regional temperatures, precipitation, and other climate variables. The use of a reference period for

such projections raises more fundamental questions, since the (previously discussed) arguments advanced to justify this approach for global-mean temperature cannot be readily transferred to regional scales.

For example, the simulated mean temperature may be a critical issue in regions where phase transitions between water and ice are common, for example, in the presence of sea ice (e.g., Wang and Overland 2009; Mahlstein and Knutti 2012) or permafrost. Mean temperature, rainfall, and evapotranspiration will all likely be important in regions where soil moisture may become limited. Further work is needed to examine the sensitivity of regional climate projections to errors in the simulation of the mean state.

The implications for climate impact studies are profound. If, for example, daily output from GCM simulations is used as an input to a climate impact

model or if a temperature threshold is used to calculate integrated measures of temperature exceedance, then an adequate simulation of the mean and variance (at least) of the variables used is necessary. Often this criteria is not met, and various bias correction techniques are adopted that add additional uncertainties (e.g., Christensen et al. 2008; Piani et al. 2010; Ho et al. 2012; Hawkins et al. 2013; Koehler et al. 2013). A case-by-case approach is required to assess the implications of mean state errors for climate impact studies.

Issues concerning the availability and quality of observational records are also challenging for regional projections. This is partly because optimal reference periods are typically much longer, as the variability is larger relative to forced changes (see appendix D).

SUMMARY AND RECOMMENDATIONS.

Because climate models produce different values for Earth's global-mean surface temperature, it is standard practice to define a reference period when comparing simulations and projections of temperature change with observations and with each other. While there are some justifications for this approach, it necessarily involves approximations that have limited validity. Further investigating the limitations of this approach is an important area for further research. In addition, this article has highlighted the following points:

1) There is no perfect choice of reference period, but relevant considerations include
   (i)   the need for a sufficiently long time period to reduce the effects of multidecadal natural climate fluctuations,
   (ii) the quality and global coverage of the available observations, and
   (iii) the quality of information about past radiative forcings that drive climate change.
   The first point argues for using as long a period as possible, whereas the second and third points argue in favor of using a recent period for which better quality and more complete observations are available.
2) Conclusions concerning (i) the consistency of simulations with observations (e.g., over the recent slowdown period, 1998–2013) and (ii) the magnitude of projected future changes in climate both exhibit sensitivity to the choice of reference period.
3) A strong recommendation is that any studies that seek to draw quantitative conclusions from

analyses that involve the use of a reference period should explicitly examine the robustness of those conclusions to alternative choices of reference period. This approach was taken in the assessment of near-term (2016–35) changes in global-mean temperature in the IPCC AR5 (Kirtman et al. 2013) but has not been used systematically in climate research. An alternative approach is to focus on trends in climate that do not require the definition of a reference period [see box 11.2 of Kirtman et al. (2013), van Oldenborgh et al. (2013), or Marotzke and Forster (2015)].
4) When presenting temperature projections relative to a fixed baseline, the impact of the choice of reference period can be several tenths of kelvins for some models. This is a significant issue for near-term projections of climate change but less significant for longer-term projections of climate change. Similarly, the reference period choice affects the projected ensemble spread of the CMIP5 models by up to 0.2 K. The same sensitivity can affect estimates of the time at which policy relevant temperature targets (e.g., 2 K above preindustrial climate) may be exceeded by as much as 15 years.
5) The optimal length and timing of the reference period for producing projections depends on the climate variable under consideration. The most recent 20 years [as used by IPCC Fourth Assessment Report (AR4) and AR5] is a reasonable choice for global-mean temperature (see appendix D). For other variables, such as precipitation, the optimal reference periods are likely to be much longer, as the variability is large relative to the changes.
6) The issues associated with the use of anomalies relative to a reference period are particularly serious for regional climate projections. Errors in simulating the mean (and higher-order moments) of regional climate variables may have consequences for regional climate and impact projections that need to be assessed on a case-by-case basis.

APPENDIX A: CMIP5 MODELS

ACCESS1.0	Australian Community Climate and Earth-System Simulator, version 1.0
ACCESS1.3	Australian Community Climate and Earth-System Simulator, version 1.3
BCC_CSM1.1	Beijing Climate Center, Climate System Model, version 1.1
BCC_CMS1.1(m)	Beijing Climate Center, Climate System Model, version 1.1 (moderate resolution)
BNU-ESM	Beijing Normal University–Earth System Model
CCSM4	Community Climate System Model, version 4
CESM1(BGC)	Community Earth System Model, version 1 (Biogeochemistry)
CESM1(CAM5)	Community Earth System Model, version 1 (Community Atmosphere Model, version 5)
CMCC-CM	Centro Euro-Mediterraneo per I Cambiamenti Climatici Climate Model
CMCC-CMS	Centro Euro-Mediterraneo per I Cambiamenti Climatici Stratosphere-resolving Climate Model
CNRM-CM5	Centre National de Recherches Météorologiques Coupled Global Climate Model, version 5
CSIRO Mk3.6.0	Commonwealth Scientific and Industrial Research Organisation Mark 3.6.0
CanESM2	Second Generation Canadian Earth System Model
EC-EARTH	European Consortium Earth System Model
FGOALS-g2.0	Flexible Global Ocean–Atmosphere–Land System Model, gridpoint version 2.0
FIO-ESM	First Institute of Oceanography Earth System Model
GFDL CM3	Geophysical Fluid Dynamics Laboratory Climate Model, version 3
GFDL-ESM2G	Geophysical Fluid Dynamics Laboratory Earth System Model with Generalized Ocean Layer Dynamics (GOLD) component
GFDL-ESM2M	Geophysical Fluid Dynamics Laboratory Earth System Model with Modular Ocean Model (MOM), version 4 component
GISS-E2-H	Goddard Institute for Space Studies (GISS) Model E2, coupled with Hybrid Coordinate Ocean Model (HYCOM)
GISS-E2-H-CC	GISS Model E2, coupled with either HYCOM and interactive terrestrial carbon cycle (and oceanic biogeochemistry)
GISS-E2-R	GISS Model E2, coupled with the Russell ocean model
GISS-E2-R-CC	GISS Model E2, coupled with Russell and interactive terrestrial carbon cycle (and oceanic biogeochemistry)
HadGEM2-AO	Hadley Centre Global Environment Model, version 2—Atmosphere and Ocean
HadGEM2-CC	Hadley Centre Global Environment Model, version 2—Carbon Cycle
HadGEM2-ES	Hadley Centre Global Environment Model, version 2—Earth System
INM-CM4.0	Institute of Numerical Mathematics Coupled Model, version 4.0
IPSL-CM5A-LR	L'Institut Pierre-Simon Laplace Coupled Model, version 5A, low resolution
IPSL-CM5A-MR	L'Institut Pierre-Simon Laplace Coupled Model, version 5A, midresolution
IPSL-CM5B-LR	L'Institut Pierre-Simon Laplace Coupled Model, version 5B, low resolution
MIROC5	Model for Interdisciplinary Research on Climate, version 5
MIROC-ESM	Model for Interdisciplinary Research on Climate, Earth System Model
MIROC-ESM-CHEM	Model for Interdisciplinary Research on Climate, Earth System Model, Chemistry Coupled
MPI-ESM-LR	Max Planck Institute Earth System Model, low resolution
MPI-ESM-MR	Max Planck Institute Earth System Model, medium resolution
MRI-CGCM3	Meteorological Research Institute Coupled Atmosphere–Ocean General Circulation Model, version 3
NorESM1-ME	NorESM1-M with carbon cycling (and biogeochemistry)
NorESM1-M	Norwegian Earth System Model, version 1 (intermediate resolution)

APPENDIX B: DOES GLOBAL MEAN TEMPERATURE MATTER FOR CLIMATE SENSITIVITY?

Inspired by Schmidt (2007), a simple 1D energy balance model of Earth can be written as

$$\text{surface: } G = S + \lambda A , \qquad (B1)$$

$$\text{atmosphere: } \lambda G = 2\lambda A , \text{ and} \qquad (B2)$$

$$\text{top of atmosphere (TOA): } S = (1 - \lambda)G + \lambda A , \qquad (B3)$$

where λ is the emissivity of the atmosphere (i.e., the strength of the greenhouse effect), and $S = S_*(1 - a)/4$, where a is Earth's albedo and S_* is the solar constant. In terms of temperature, $A = \sigma T_a^4$ and $G = \sigma T_g^4$, where σ is the Stefan–Boltzmann constant with T_a and T_g representing the temperature of the atmosphere and surface, respectively.

Eliminating A,

$$G = \frac{S}{1 - \lambda/2} , \qquad (B4)$$

and the surface temperature,

$$T_g = \sqrt[4]{\frac{S}{\sigma(1 - \lambda/2)}} . \qquad (B5)$$

Figure B1a shows how the mean surface temperature changes with a and λ. As a "standard" model, $T_g = 14.0°C$ for $a = 0.3$ and $\lambda = 0.7643$. For similar reference models with a fixed albedo ($a = 0.3$), a range of global temperatures ($T_g = 13°$ and $15°C$) can be produced for small changes in λ (0.7470 and 0.7814, respectively). Alternatively, for a fixed emissivity ($\lambda = 0.7643$), $T_g = 13°$ and $15°C$ for $a = 0.31$ and 0.29, respectively. Note that the observed albedo is 0.29–0.30 (Stephens et al. 2015) and that the same global temperature can be produced with widely different parameter settings.

A change in forcing can be introduced by varying S_* to change S. In the case where there are no feedbacks, then

$$\Delta G = \frac{\Delta S}{1 - \lambda/2} . \qquad (B6)$$

Figure B1b shows that the warming for a $\Delta S = 1 \text{ W m}^{-2}$ forcing change is rather insensitive to the initial mean global temperature. For example, in the range of reference models given above with initial temperatures from 13° to 15°C, this no feedback (or Planck) sensitivity only varies from 0.298 to 0.305 K (W m⁻²)⁻¹.

The Planck sensitivity is amplified by various feedbacks (albedo, water vapor, lapse rate, clouds, etc.). As a simple example, we parameterize the albedo and emissivity to be linearly dependent on temperature change:
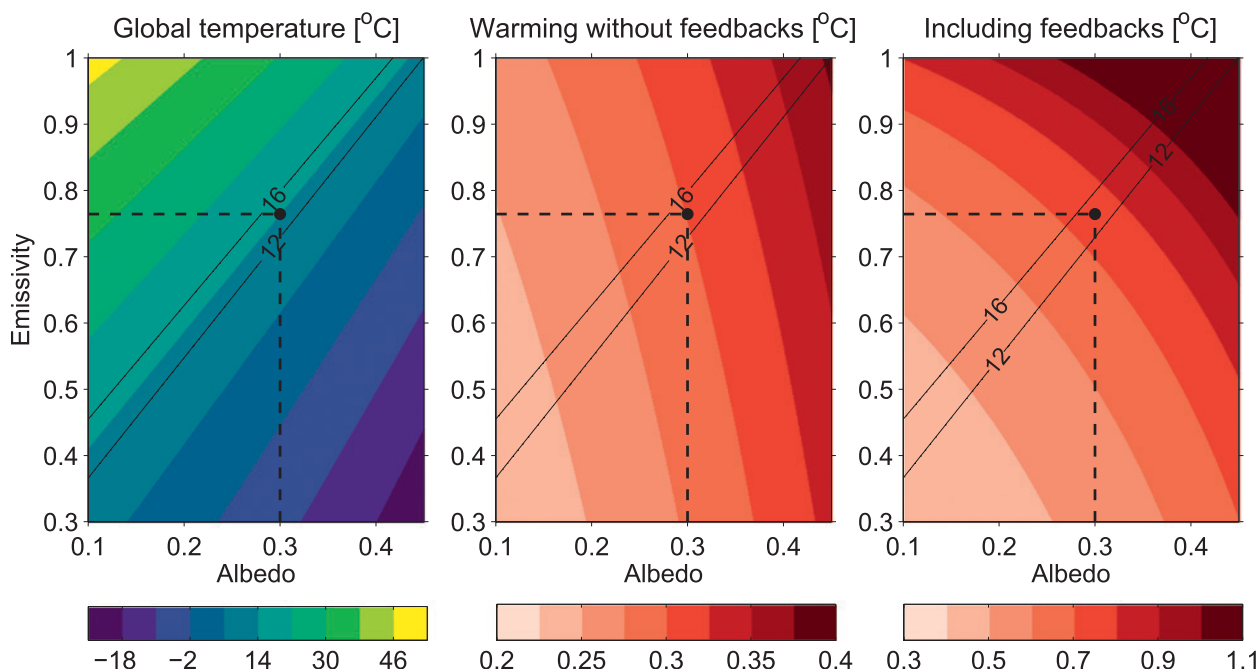

FIG. B1. Simple 1D energy balance model for Earth showing (left) mean global temperature as a function of albedo and emissivity parameters and the corresponding warming for a 1 W m⁻² forcing (middle) without feedbacks and (right) including feedbacks. Black contours represent constant 12° and 16°C global temperatures in all panels, with a standard model (14°C) indicated with the filled circle and dashed lines.

Comparing HadCRUT4.3 and Cowtan & Way global temperatures
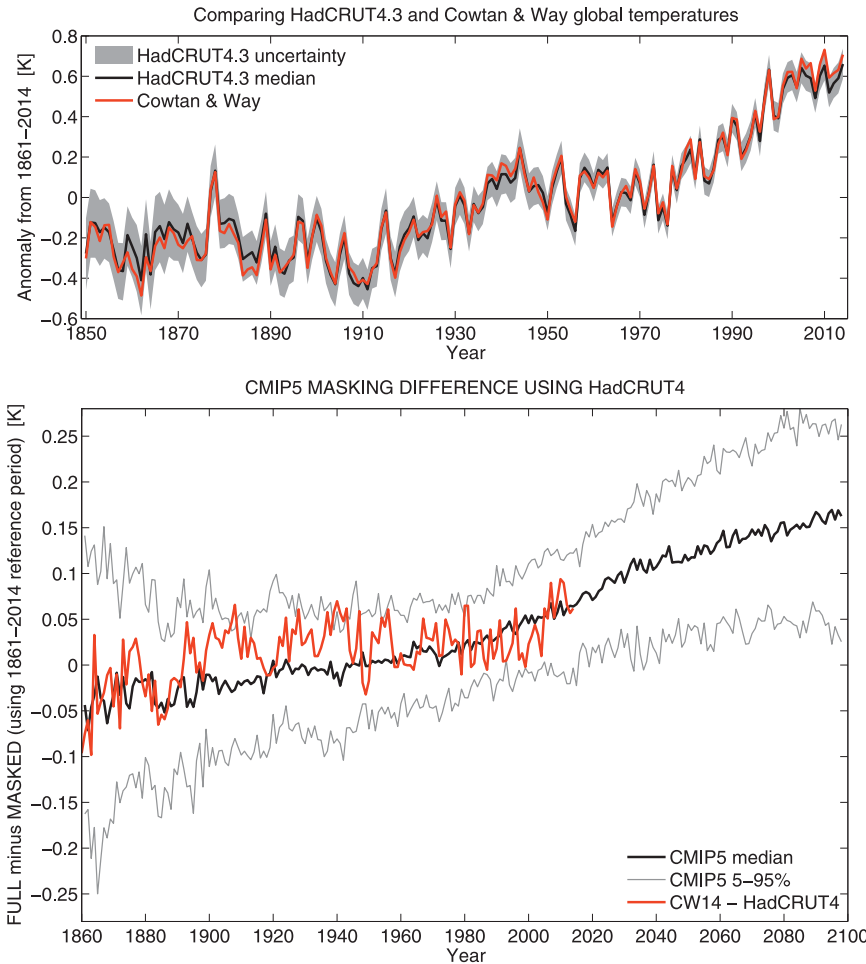
CMIP5 MASKING DIFFERENCE USING HadCRUT4

FIG. C1. Comparing global-mean surface temperatures. (top) HadCRUT4.3 and CW14. (bottom) Difference between simulation of global temperatures using the full data and only where HadCRUT4.3 observations are available (masked). For 2015 onward, the HadCRUT4.3 mask of Dec 2014 is assumed. A reference period of 1861–2014 is used in both panels.

$$a = k(T - T_o) + a_0 , \qquad \text{(B7)}$$

$$\lambda = m(T - T_o) + \lambda_0 . \qquad \text{(B8)}$$

If we set $k = -0.003$ and $m = 0.005$, then Fig. B1c shows the corresponding temperature change for a 1 W m^{-2} forcing. The ratio of k and m is set by the changes in a and λ for the range of reference models discussed above. The resulting warming is larger than without the feedbacks but still rather independent of global-mean temperature; there is a near orthogonality between the contours of mean temperature and climate sensitivity. For our reference model, the climate sensitivity with feedbacks is 2.78 K for a forcing of 3.7 W m^{-2}, equivalent to a doubling of CO_2.

Although this is only a toy model of global climate, it provides some simple physical explanations for why

climate sensitivity may not depend strongly on global-mean temperature as long as it does not vary too much from the observed value, as seen in the CMIP5 models (Fig. 2). However, Bloch-Johnson et al. (2015) discuss the possible consequences of nonlinear dependence of feedbacks on temperature and find slightly larger sensitivity to the mean state.

APPENDIX C: CONSIDERING OBSERVATIONAL ISSUES.
HadCRUT4.3 is not a spatially complete dataset, as observations are not available everywhere. Therefore, comparing HadCRUT4.3 with the full global temperature from GCMs is not necessarily a fair comparison. To test the sensitivity to the lack of complete observational coverage, Fig. C1 first compares the global temperatures using HadCRUT4.3 (Morice et al. 2012) and the interpolated version of CW14. The estimates of CW14 fall inside the HadCRUT4.3 uncertainties for the vast majority of years. The most recent decade, however, is at the upper edge of the uncertainties, suggesting that the missing regions are warming more rapidly than the global average in the last few years.

Figure C1 also shows the effect on simulated global temperatures when computed only where there is observational coverage in HadCRUT4.3, on a month-by-month basis. This "masking" introduces a slight cool bias in the simulations; that is, the simulations warm less when masked with the observational coverage, again implying that the real world may have actually warmed slightly more than observed with HadCRUT4.3. In addition, the lack of complete coverage may reduce the measured long-term future observed change by around 0.07 K and the near term by around 0.02 K (Fig. C1), assuming the observational coverage does not improve.

Figure C1 also shows the difference between CW14 and HadCRUT4.3, which broadly matches the estimates from the simulations. The differences between the two datasets apparent in the last few years are not particularly large compared to the range of corrections expected from the CMIP5 simulations.

In addition, Cowtan et al. (2015) highlight a further complication with such a comparison. The observations are constructed from sea surface temperatures (SSTs) over the ocean and near-surface air temperature over the land, whereas the models are normally presented using averaged air temperatures everywhere. As the SSTs warm slightly slower than the corresponding air temperatures over the oceans, this results in the simulated changes using air temperatures being only slightly larger than when using SSTs over the ocean.

Overall, the masking and surface-type effects account for around a third of the difference between the multimodel mean and HadCRUT4.3 when using a 1961–90 reference period (Cowtan et al. 2015).

APPENDIX D: HOW LONG SHOULD A REFERENCE PERIOD BE? To consider the question of an appropriate length of reference period (or tube; see the sidebar on "Illustrating the effect of reference period choice") to make future projections, we use a toy simulator of temperatures. We assume that temperature θ changes linearly with time t as

$$\theta = \alpha t + \varepsilon(t) \qquad (D1)$$

from 1970 to 2100. We consider realizations of temperature using different models (or sensitivity α) sampled from $N(\hat{\alpha}, \sigma^2)$ and red noise ε with variance γ^2 and the AR(1) parameter fixed at 0.5.

We generate 1,000 realizations (or climate models) of temperature (with and without the noise component). These simulations are then referenced to different periods of length $L = 1$–30 years but all ending in 2005. The shortest reference period is then only using 2005, and the longest is 1976–2005.

The total uncertainty (using the standard deviation across the 1,000 realizations) in future temperatures can
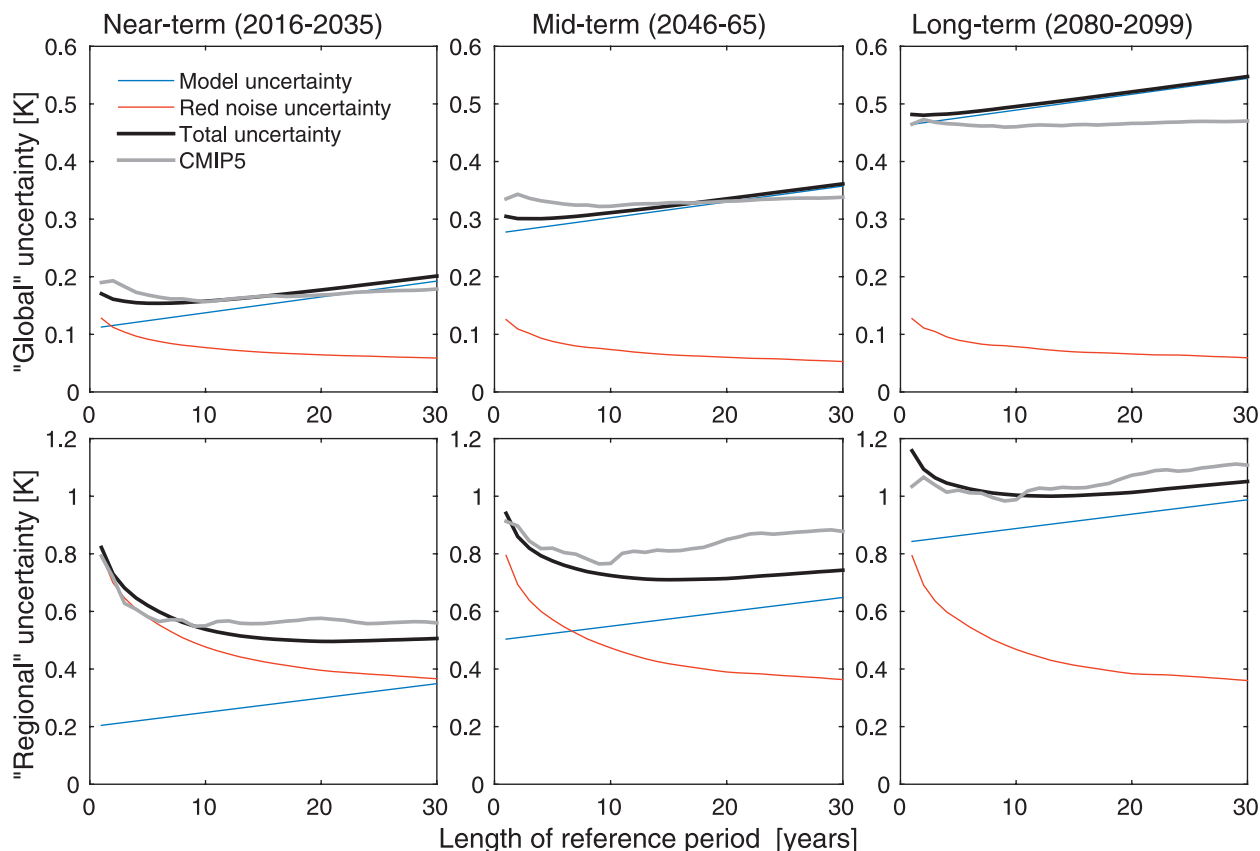
CMIP5 and toy simulator uncertainty

FIG. D1. Defining an optimal reference period. Uncertainty in CMIP5 projections as function of length of reference period, always ending in 2005 (gray). (top) Global mean surface air temperature (SAT). (bottom) European land SAT. The total uncertainty (black) and two components of uncertainty (blue, red) using a toy simulator are also shown.

be separated into components due to different model sensitivity and that due to the noise. Figure D1 shows the total uncertainty for different future time periods (black), the uncertainty due to the model sensitivity (blue) and noise (red) for three different future time periods (columns), and two different sets of toy simulator parameters (rows). The two sets of parameters are chosen to approximately represent global-mean temperature (top row; $\hat{\alpha} = 0.21$ K decade^{-1}, $\sigma = 0.055$ K decade^{-1}, and $\gamma = 0.12$ K) and European land temperature (bottom row; $\hat{\alpha} = 0.23$ K decade^{-1}, $\sigma = 0.10$ K decade^{-1}, and $\gamma = 0.75$ K) in the CMIP5 GCMs.

As L increases, the noise uncertainty component decreases because of averaging over more years in the reference period. The model sensitivity uncertainty component increases linearly with L because there is more time between the middle of the reference period and the verification time, allowing for a longer period of model uncertainty growth. The total uncertainty therefore has a minimum. For global temperature, this minimum occurs for $L = 1$–5 yr, whereas for regional temperature, the optimal L is longer, around 15–20 yr, depending on the verification time. For climate variables with larger variability, such as precipitation, the optimal L may increase further.

We have used the toy model to demonstrate that there are competing effects when choosing a reference period for making projections. A similar procedure can be performed for simulated global temperature and European land temperatures to test whether these effects are seen in the CMIP5 GCMs. In this case, the total CMIP5 uncertainty is shown in Fig. D1 as the gray lines. This should not be expected to match the toy simulator perfectly because the CMIP5 trends are nonlinear. However, a similar structure in the change in total uncertainty for different L is seen but with less sensitivity to L. For global and regional temperatures a reference period length of 10 and 20 years, respectively, is close to optimal. The IPCC AR5 decision to use $L = 20$ appears to have been a good choice for presenting future changes in temperature.

REFERENCES

Annan, J. D., and J. C. Hargreaves, 2010: Reliability of the CMIP3 ensemble. *Geophys. Res. Lett.*, **37**, L02703, doi:10.1029/2009GL041994.

Bloch-Johnson, J., R. T. Pierrehumbert, and D. S. Abbot, 2015: Feedback temperature dependence determines the risk of high warming. *Geophys. Res. Lett.*, **42**, 4973–4980, doi:10.1002/2015GL064240.

Callendar, G. S., 1938: The artificial production of carbon dioxide and its influence on temperature. *Quart. J. Roy. Meteor. Soc.*, **64**, 223–240, doi:10.1002/qj.49706427503.

Carslaw, K. S., and Coauthors, 2013: Large contribution of natural aerosols to uncertainty in indirect forcing. *Nature*, **503**, 67–71, doi:10.1038/nature12674.

Christensen, J. H., F. Boberg, O. B. Christensen, and P. Lucas-Picher, 2008: On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophys. Res. Lett.*, **35**, L20709, doi:10.1029/2008GL035694.

Collins, M., and Coauthors, 2013: Long-term climate change: Projections, commitments and irreversibility. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 1029–1136.

Cowtan, K., and R. G. Way, 2014: Coverage bias in the hadcrut4 temperature series and its impact on recent temperature trends. *Quart J. Roy. Meteor. Soc.*, **140**, 1935–1944, doi: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, doi:10.1002/2015GL064888.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.

Economist, 2013: A sensitive matter. *Economist,* 30 March. [Available online at www.economist.com/news/science-and-technology/21574461-climate-may-be-heating-up-less-response-greenhouse-gas-emissions.]

Flato, G., and Coauthors, 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 741–866.

Fyfe, J. C., N. P. Gillett, and F. W. Zwiers, 2013: Overestimated global warming over the past 20 years. *Nat. Climate Change*, **3**, 767–769, doi:10.1038/nclimate1972.

Good, P., W. Ingram, F. H. Lambert, J. A. Lowe, J. M. Gregory, M. J. Webb, M. A. Ringer, and P. Wu, 2012: A step-response approach for predicting and understanding non-linear precipitation changes. *Climate Dyn.*, **39**, 2789–2803, doi:10.1007/s00382-012-1571-1.

Gregory, J. M., T. Andrews, and P. Good, 2015: The inconstancy of the transient climate response parameter under increasing CO_2. *Philos. Trans. Roy. Soc. London*, **373**, 20140417, doi:10.1098/rsta.2014.0417.

Hansen, J., and S. Lebedeff, 1987: Global trends of measured surface air temperature. *J. Geophys. Res.*, **92**, 13 345–13 372, doi:10.1029/JD092iD11p13345.

——, I. Fung, A. Lacis, D. Rind, S. Lebedeff, R. Ruedy, G. Russell, and P. Stone, 1988: Global climate changes

as forecast by Goddard Institute for Space Studies three-dimensional model. *J. Geophys. Res.*, **93**, 9341–9364, doi:10.1029/JD093iD08p09341.

Hawkins, E., 2013: Updated comparison of simulations and observations. Accessed 1 August 2015. [Available online at www.climate-lab-book.ac.uk/2013/updated -comparison-of-simulations-and-observations/.]

——, and P. D. Jones, 2013: On increasing global temperatures: 75 years after Callendar. *Quart. J. Roy. Meteor. Soc.*, **139**, 1961–1963, doi:10.1002/qj.2178.

——, T. E. Fricker, A. J. Challinor, C. A. T. Ferro, C. K. Ho, and T. M. Osborne, 2013: Increasing influence of heat stress on French maize yields from the 1960s to the 2030s. *Global Change Biol.*, **19**, 937–947, doi:10.1111/gcb.12069.

Ho, C. K., D. B. Stephenson, M. Collins, C. A. T. Ferro, and S. J. Brown, 2012: Calibration strategies: A source of additional uncertainty in climate change projections. *Bull. Amer. Meteor. Soc.*, **93**, 21–26, doi:10.1175/2011BAMS3110.1.

Huang, J., H. M. van den Dool, and A. G. Barnston, 1996: Long-lead seasonal temperature prediction using optimal climate normals. *J. Climate*, **9**, 809–817, doi:10.1175/1520-0442(1996)009<0809:LLSTPU>2 .0.CO;2.

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

Jones, P. D., T. J. Osborn, and K. R. Briffa, 1997: Estimating sampling errors in large-scale temperature averages. *J. Climate*, **10**, 2548–2568, doi:10.1175/1520 -0442(1997)010<2548:ESEILS>2.0.CO;2.

——, M. New, D. E. Parker, S. Martin, and I. G. Rigor, 1999: Surface air temperature and its changes over the past 150 years. *Rev. Geophys.*, **37**, 173–199, doi:10.1029/1999RG900002.

Joshi, M., E. Hawkins, R. Sutton, J. Lowe, and D. Frame, 2011: Projections of when temperature change will exceed 2°C above pre-industrial levels. *Nat. Climate Change*, **1**, 407–412, doi:10.1038/nclimate1261.

Kirtman, B., and Coauthors, 2013: Near-term climate change: Projections and predictability. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 953–1028.

Kobayashi, S., and Coauthors, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, doi:10.2151/jmsj.2015-001.

Koehler, A.-K., A. J. Challinor, E. Hawkins, and S. Asseng, 2013: Influences of increasing temperature on Indian wheat: Quantifying limits to predictability. *Environ. Res. Lett.*, **8**, 034016, doi:10.1088/1748 -9326/8/3/034016.

Landsberg, H., 1944: On the relation between mean cloudiness and the number of clear and cloudy days in the United States. *Eos, Trans. Amer. Geophys. Union*, **25**, 456–457, doi:10.1029/TR025i003p00456.

Mahlstein, I., and R. Knutti, 2012: September Arctic sea ice predicted to disappear near 2°C global warming above present. *J. Geophys. Res.*, **117**, D06104, doi:10.1029/2011JD016709.

Marotzke, J., and P. Forster, 2015: Forcing, feedback and internal variability in global temperature trends. *Nature*, **517**, 565–570, doi:10.1038/nature14117.

Mauritsen, T., and Coauthors, 2012: Tuning the climate of a global model. *J. Adv. Model. Earth Syst.*, **4**, M00A01, doi:10.1029/2012MS000154.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geophys. Res. Atmos.*, **117**, D08101, doi:10.1029/2011JD017187.

Piani, C., G. Weedon, M. Best, S. Gomes, P. Viterbo, S. Hagemann, and J. Haerter, 2010: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *J. Hydrol.*, **395**, 199–215, doi:10.1016/j .jhydrol.2010.10.024.

Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, doi:10.1175/JCLI-D-11-00015.1.

Rose, D., 2013: The Great Green Con no. 1: The hard proof that finally shows global warming forecasts that are costing you billions were WRONG all along. *Daily Mail*, 16 March. [Available online at www .dailymail.co.uk/news/article-2294560/The-great -green-1-The-hard-proof-finally-shows-global-warm ing-forecasts-costing-billions-WRONG-along.html.]

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, doi:10.1175/2010BAMS3001.1.

Sanderson, B. M., and R. Knutti, 2012: On the interpretation of constrained climate model ensembles. *Geophys. Res. Lett.*, **39**, L16708, doi:10.1029/2012GL052665.

Schmidt, G., 2007: Learning from a simple model. Accessed 1 August 2015. [Available online at www .realclimate.org/index.php/archives/2007/04 /learning-from-a-simple-model/.]

——, 2014: Absolute temperatures and relative anomalies. Accessed 1 August 2015. [Available online at www.realclimate.org/index.php/archives/2014/12 /absolute-temperatures-and-relative-anomalies/.]

——, D. T. Shindell, and K. Tsigaridis, 2014: Reconciling warming trends. *Nat. Geosci.*, **7**, 158–160, doi:10.1038 /ngeo2105.

Stephens, G., D. O'Brien, P. J. Webster, P. Pilewski, S. Kato, and J. Li, 2015: The albedo of Earth. *Rev. Geophys.*, **53**, 141–163, doi:10.1002/2014RG000449.

Stevens, B., 2013: Aerosols: Uncertain then, irrelevant now. *Nature*, **503**, 47–48, doi:10.1038/503047a.

Stott, P., P. Good, G. Jones, N. Gillett, and E. Hawkins, 2013: The upper end of climate model temperature projections is inconsistent with past warming. *Environ. Res. Lett.*, **8**, 014024, doi:10.1088/1748-9326/8/1/014024.

Thomson, A. M., and Coauthors, 2011: RCP4.5: A pathway for stabilization of radiative forcing by 2100. *Climatic Change*, **109**, 77–94, doi:10.1007/s10584-011-0151-4.

Trewin, B., 2007: The role of climatological normals in a changing climate. World Meteorological Organization Tech. Rep. WCDMP 61, 46 pp. [Available online at www.wmo.int/datastat/documents/WCDMPNo61_1.pdf.]

van Oldenborgh, G. J., F. J. Doblas Reyes, S. S. Drijfhout, and E. Hawkins, 2013: Reliability of regional climate model trends. *Environ. Res. Lett.*, **8**, 014055, doi:10.1088/1748-9326/8/1/014055.

Vautard, R., and Coauthors, 2014: The European climate under a 2°C global warming. *Environ. Res. Lett.*, **9**, 034006, doi:10.1088/1748-9326/9/3/034006.

Wang, M., and J. E. Overland, 2009: A sea ice free summer arctic within 30 years? *Geophys. Res. Lett.*, **36**, L07502, doi:10.1029/2009GL037820.