nature climate change

LETTERS

PUBLISHED ONLINE: 10 JUNE 2012 | DOI: 10.1038/NCLIMATE1553

Human-induced global ocean warming on multidecadal timescales

P. J. Gleckler[1]*, B. D. Santer[1], C. M. Domingues[2,3], D. W. Pierce[4], T. P. Barnett[4], J. A. Church[3], K. E. Taylor[1], K. M. AchutaRao[5], T. P. Boyer[6], M. Ishii[7] and P. M. Caldwell[1]

Large-scale increases in upper-ocean temperatures are evident in observational records[1]. Several studies have used well-established detection and attribution methods to demonstrate that the observed basin-scale temperature changes are consistent with model responses to anthropogenic forcing and inconsistent with model-based estimates of natural variability[2–5]. These studies relied on a single observational data set and employed results from only one or two models. Recent identification of systematic instrumental biases[6] in expendable bathythermograph data has led to improved estimates of ocean temperature variability and trends[7–9] and provide motivation to revisit earlier detection and attribution studies. We examine the causes of ocean warming using these improved observational estimates, together with results from a large multimodel archive of externally forced and unforced simulations. The time evolution of upper ocean temperature changes in the newer observational estimates is similar to that of the multimodel average of simulations that include the effects of volcanic eruptions. Our detection and attribution analysis systematically examines the sensitivity of results to a variety of model and data-processing choices. When global mean changes are included, we consistently obtain a positive identification (at the 1% significance level) of an anthropogenic fingerprint in observed upper-ocean temperature changes, thereby substantially strengthening existing detection and attribution evidence.

We examine volume average temperature anomalies (ΔT) for the upper 700 m of the global ocean (see Methods). Figure 1a compares uncorrected observational ΔT estimates ISH-UNCOR (ref. 10) and LEV-UNCOR (ref. 11) with improved versions, ISH (ref. 8) and LEV (ref. 9), which incorporate corrections for expendable bathythermograph (XBT) biases. The bias-corrected temperature analysis[7] from a third group (DOM) is also shown. Bias corrections have a substantial impact on the time evolution of ΔT, particularly during the 1970s–1980s, when they markedly reduce spurious decadal variability.

As shown below, these bias adjustments have important implications for detection and attribution (D&A) studies. Although there are no significant differences between the ΔT trends (which range from 0.022 to 0.028 °C per decade) in the three improved observational data sets, Fig. 1a illustrates that substantial structural uncertainties remain. The impact of different XBT bias corrections is a major source of this uncertainty[12].

Another important component of observational uncertainty relates to the sparseness of ocean temperature measurements and to the different methods used to objectively infill data where and when measurements are not available[13–15]. ISH and LEV use objective mapping techniques to carry out infilling, generating anomalies that are biased towards zero in data-sparse regions. The infilling method of DOM employs statistics of observed ocean variability estimated from altimeter data. We compare the spatially complete infilled estimates (ΔT_{IF}) with subsampled ΔT data (ΔT_{SS}) restricted to available *in situ* measurements (see Methods). Not surprisingly, the ΔT_{SS} variability in Fig. 1b is greater than that of ΔT_{IF}, particularly at the times/locations of the sparsest sampling (early in the record and in the southern oceans; Supplementary Fig. S1).

We use results from phase 3 of the Coupled Model Intercomparison Project (CMIP3; see Methods and Supplementary Information) to obtain information on the behaviour of ΔT in unforced (control) simulations and in externally forced twentieth-century runs (20CEN). External forcing is by a variety of anthropogenic factors (primarily greenhouse gases and sulphate aerosols). In some models, the applied forcing also includes natural changes in volcanic aerosols and solar irradiance. The seven CMIP3 models (with the data required for our analysis) incorporating the effects of volcanic eruptions (VOL) in the 20CEN simulations uptake less heat than the six that do not (NoV)[16].

Accounting for residual simulation drift (see Methods), the multimodel VOL global mean ΔT time series are within the spread of observational estimates over the entire observational record, whereas the warming in the NoV multimodel average is larger than observed in the most recent decades (Fig. 1c). Twenty-first-century ΔT changes in CMIP3 future projections are also shown[17], and are based on the SRES A1B scenario from the Intergovernmental Panel on Climate Change *Special Report on Emissions Scenarios* (SRES). The small discontinuity between the 20CEN and SRES A1B results arises because fewer simulations are available for the scenario runs and forcing discontinuities are known to exist in some simulations[18] (see Supplementary Information). Note that inclusion of volcanic forcing increases the simulated variability[7,15,19].

Figure 2 shows linear trends over 1960–1999 in observed and simulated ΔT_{SS} and ΔT_{IF} data. Results are for global averages and each of the six ocean basins. Observed ΔT_{IF} trends are generally smaller than their ΔT_{SS} counterparts, probably because the ΔT_{IF} results are biased low in data-sparse regions. Note

[1]Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Mail Code L-103, 7000 East Avenue, Livermore, California 94550, USA, [2]Antarctic and Climate Ecosystems Cooperative Research Centre, Hobart, Australia, [3]Centre for Australian Weather and Climate Research and Wealth from Oceans Flagship, CSIRO Marine and Atmospheric Research, GPO Box 1538, Hobart, Tasmania 7001, Australia, [4]Climate Research Division, Scripps Institution of Oceanography, Mail Stop 0224, La Jolla 92093, USA, [5]Indian Institute of Technology, Delhi 110 016, India, [6]National Oceanographic Data Center, National Oceanic and Atmospheric Administration, Silver Spring 20910, USA, [7]Climate Research Department, Meteorological Research Institute, 1-1, Nagamine, Tsukuba, Ibaraki 305-0052, Japan. *e-mail: gleckler1@llnl.gov.

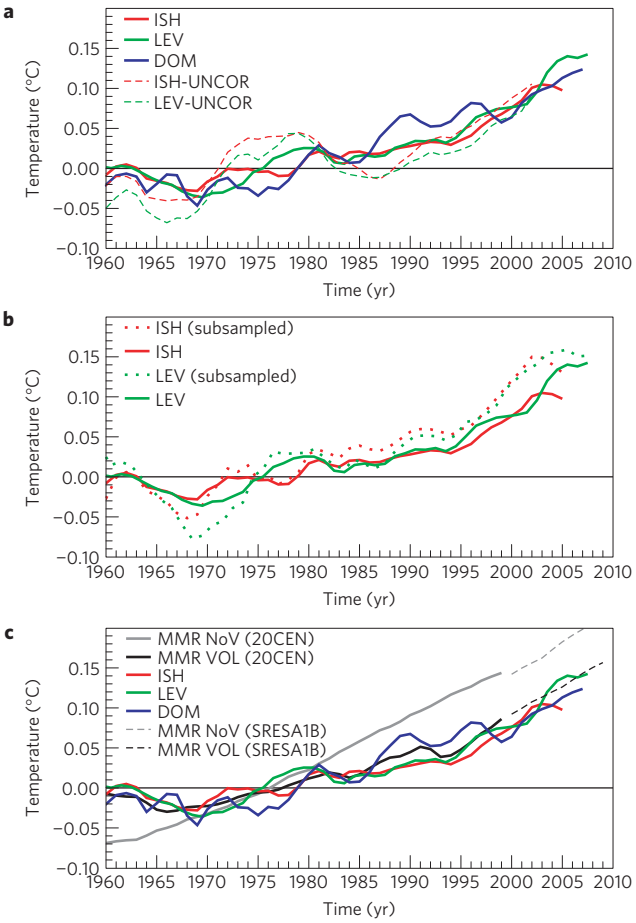

Figure 1 | Global mean ΔT (0–700 m) with respect to a 1957–1990 climatology. a, Estimates of Domingues *et al.*[7] (DOM), Ishii *et al.*[8] (ISH) and Levitus *et al.*[9] (LEV), all of which have been corrected for XBT biases. Earlier (uncorrected) estimates of Ishii *et al.*[10] (ISH-UNCOR) and Levitus *et al.*[11] (LEV-UNCOR) are also shown. **b**, ISH and LEV ΔT_{IF} (solid lines) and ΔT_{SS} (dotted lines) results. **c**, Recent observed ΔT estimates compared with the CMIP3 20CEN MMR for the subsets of models including VOL and NoV. MMR results are also shown for the CMIP3 SRES A1B scenarios, constructed from the same VOL and NoV subsets defined by the 20CEN models. The SRES A1B results include fewer model simulations than were available in the 20CEN MMRs. All time series are computed from spatially complete data, except the dotted lines in **b**. For visual display purposes only, all observational data are five-year running averages.

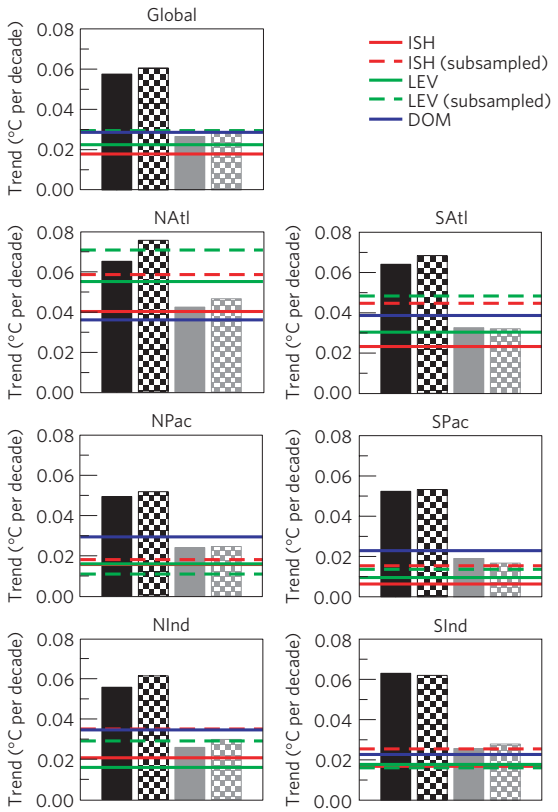

Figure 2 | Observed and simulated least-squares linear trends in ΔT over 1960–1999. Results are for global mean ΔT and for ΔT in six individual ocean basins. CMIP3 20CEN MMRs are based on different choices of applied external forcings (black bars for NoV, grey bars for VOL) and data coverage (spatially complete model data as solid bars; subsampled model data as checkered bars). Observations include infilled (solid lines) and subsampled (dashed lines) estimates for both Ishii *et al.*[8] and Levitus *et al.*[9]. Domingues *et al.*[7] estimates are available only for the infilled case. North Atlantic, NAtl; South Atlantic, SAtl; North Pacific, NPac; South Pacific, SPac; North Indian, NInd; South Indian; SInd.

that in both models and observations, the Atlantic warming is larger than in the Pacific.

The 20CEN multimodel response (MMR) trends are more sensitive to the inclusion or neglect of volcanic forcing than to the use of spatially complete or subsampled data (Fig. 2). In most cases, the NoV ΔT_{SS} and ΔT_{IF} trends are larger than the corresponding observed estimates, whereas the VOL model results are bounded by the observational data. Recent evidence suggests that the century-scale ΔT changes in the VOL simulations may be biased low as a result of the neglect of pre-Krakatoa eruptions[20], but this will have little impact on the spatial structure of our normalized fingerprint (see below).

Before conducting our D&A analysis, it is important to verify that the models used here do not systematically underestimate natural variability, particularly on decadal timescales relevant to the detection of a slowly evolving ocean-warming signal. Such comparisons are hampered by the relatively short length of existing observational ΔT estimates. Similar comparisons with longer observational sea surface temperature records[21] show no evidence that CMIP3 models systematically underestimate observed sea surface temperature variability on 5–20-year timescales.

Here, we use the variability metrics S_5 and S_{10} to compare the combined spatiotemporal variability of simulated and observed five- and ten-year trends, respectively. This is done separately for ΔT_{SS} and ΔT_{IF}. For each ocean basin (and for the global mean), non-overlapping L-year trends in the ΔT time series, where L is length in time, are pooled together and the standard deviation is computed from the pooled samples (see Methods). For ten-year trends over 1960–1999, for example, there are 28 samples (four non-overlapping ten-year trends × seven regions). Because volcanic eruptions contribute to multiyear variability[7,15,19], the 20CEN VOL simulations (rather than the control runs) are the most appropriate integrations for comparing the amplitude of simulated and observed spatiotemporal variability.

With infilled observations and spatially complete model results, there is no evidence that the CMIP3 VOL models systematically underestimate the spatiotemporal variability of ten-year trends (Fig. 3a). The S_{10} result averaged for the three observations (AVGOBS) is virtually identical to the average of the individual S_{10} statistics from the VOL models.

As expected, the spatiotemporal variability of both models and observations is systematically larger in ΔT_{SS} (Fig. 3b). The observed

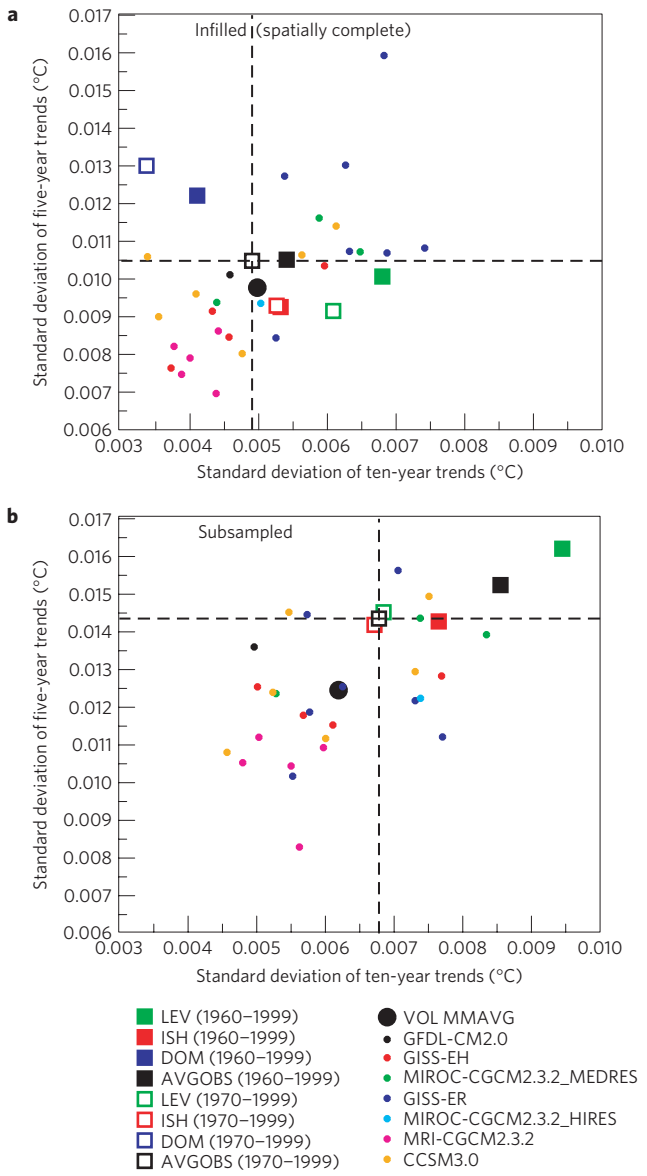

Figure 3 | Observed and simulated variability metrics S_5 and S_{10}, which provide information on the combined spatiotemporal variability of five- and ten-year ΔT trends, respectively. For details of the calculation of S_5 and S_{10}, refer to the Methods section. **a**, Results for infilled observations and model data. **b**, Results for subsampled observations and models (using a space-time mask of observed data coverage). Individual observed S_5 and S_{10} estimates are identified as coloured squares (filled for 1960–1999, open for 1970–1999). The average of the variability statistics from the three individual observational data sets (AVGOBS) is also shown. All model results are for a 1960–1999 analysis period. Model S_5 and S_{10} values from individual VOL model simulations are shown as small coloured circles. The VOL model average of the S_5 and S_{10} results is denoted by a solid black circle. Where more than one realization of the 20CEN experiment is available for an individual model, these realizations are averaged together before averaging across models. Dashed lines intersect at the average of the observed variability estimates (for the 1970–1999 analysis period).

S_5 and S_{10} variability estimates are sensitive to the choice of analysis period. Over 1960–1999, LEV and ISH show substantial variability differences, both for S_5 and S_{10}. For the period 1970–1999, however, values of the variability metrics are very similar in LEV and ISH, which reflects the fact that the two groups rely on very similar raw data. The larger differences (LEV versus ISH) over the longer analysis period are probably related to an observing system transformation in the late 1960s from predominantly mechanical bathythermographs (0.5 °C accuracy and 250 m maximum depth) to XBTs (0.1 °C accuracy, with maximum depths of 460–760 m), with XBTs in widespread use by 1970 (Supplementary Fig. S2). The lower geospatial data distribution in the 1960s, both horizontally and with depth, allows for more deviation in ΔT due to differences in the analyses (for example, quality control, bias corrections, gridding) employed by LEV and ISH. The S_{10} results suggest that on decadal timescales, the CMIP3 VOL models may underestimate the (poorly constrained) observed spatiotemporal variability of ΔT_{SS} by 25–28% (depending on the choice of analysis period). We consider below whether noise errors of this magnitude affect our ability to identify an anthropogenic fingerprint in observations.

Previous D&A studies involving ΔT relied on space–time detection methods[2–5]. Such approaches combine spatial and temporal information into a single vector. We apply a standard pattern-based method[22], which treats spatial and temporal aspects of the ΔT detection problem separately, and facilitates comparison of simulated and observed basin-scale patterns of temperature change. This method has been successfully employed in the identification of anthropogenic effects on atmospheric temperature, tropopause height and water vapour[23].

We use the first empirical orthogonal function (EOF) of the 20CEN MMR to estimate the simulated spatial pattern of response (the fingerprint) to external forcings (Fig. 4a). The fingerprint was estimated over the period 1960–1999. The spatial structure of the (normalized) fingerprint is relatively insensitive to the choice of analysis period (see Methods). In the VOL and NoV MMRs (and in each of the three observed ΔT data sets), the leading EOF primarily captures the pronounced change in mean state and exhibits warming in all ocean basins, with consistently larger warming in the Atlantic than in the Pacific (Fig. 2). The model fingerprints are more similar to the leading EOF of DOM ΔT than to the leading EOF of the other two observational estimates.

Multimodel noise estimates are obtained by pooling the preindustrial control runs of individual VOL and NoV models (see Methods and Supplementary Fig. S3). In contrast to the model fingerprints, the leading EOF derived from the pooled control runs does not have the same sign in all basins (Fig. 4b). The pooled VOL and NoV control runs show differences in the structure of the leading noise mode. Because the control runs do not include time-varying external forcings, these differences must be the result of structural differences in the models (for example, in model physics, resolution and parameterizations).

Our aim is to search for time-increasing correspondence between the model-predicted ocean-warming fingerprint and the observational data sets, and then to determine whether such correspondence could be due to natural variability alone. This involves projecting the model (20CEN and preindustrial control runs) and observational ΔT datasets onto the VOL model fingerprint pattern, yielding pseudo-principal component projection time series. By fitting overlapping trends of various values of L to these pseudo-principal component time series, we can examine the behaviour of signal-to-noise (S/N) ratios as a function of timescale and determine the detection time—the time at which S/N rises above (and remains above) a stipulated (1% or 5%) significance level[23] (see Methods). In the following, we selected 1970 (rather than 1960) as the start date for calculating trends from pseudo-principal component time series, because of the relative paucity of *in situ* ΔT profiles in the 1960s (Supplementary Fig. S2).

Figure 5a shows how the VOL, NoV and observational ΔT signals evolve as L increases from 10 to 39 years. The AVGOBS of the three XBT corrected data sets is also shown, together with a simple estimate of signal uncertainty obtained from the variability in

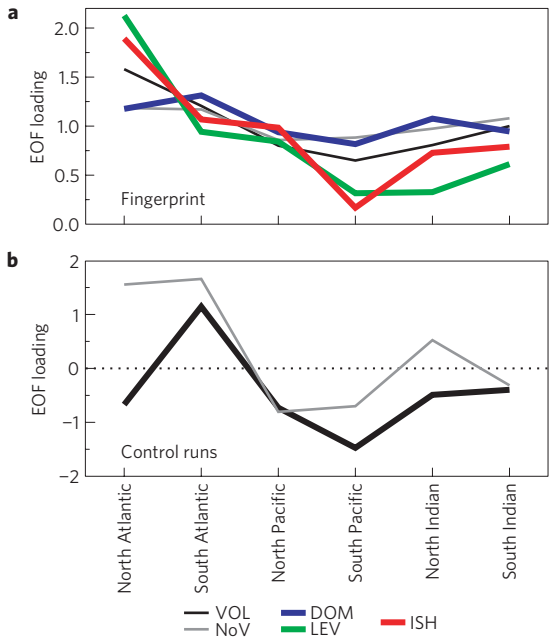

Figure 4 | EOF analysis of basin-average upper-ocean temperature changes. Basin-scale structure of externally forced fingerprints from 20CEN runs (**a**) and leading noise mode from pooled control runs (**b**). For calculation details, refer to the Methods section. The leading EOF for each of the three different observational ΔT estimates is also shown in **a**. Observational results are for infilled data sets and all model results are for spatially complete data, with removal of model drift based on a cubic fit to control-run data, and the global mean included in all data.

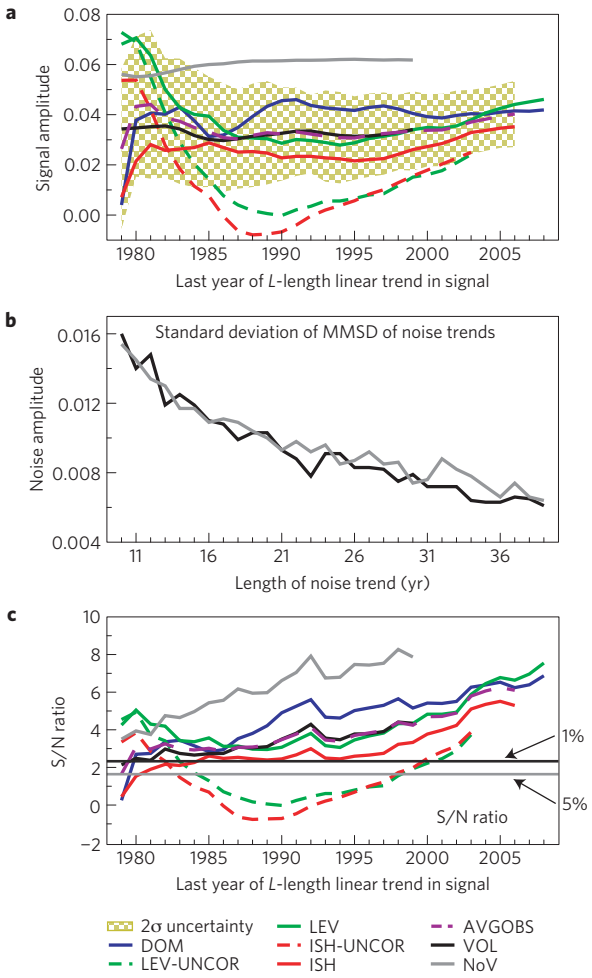

Figure 5 | S/N analysis of basin-scale changes in ΔT. The analysis relies on pseudo-principal component time series, obtained by projecting modelled and observed basin-scale ΔT data onto a common basis set (the VOL model fingerprint). **a**, L-length linear trends calculated from the signal pseudo-principal component time series—that is, from projections (onto the VOL model fingerprint) of newer, XBT-corrected ΔT data (solid red, green and blue lines); older, uncorrected ΔT data (dashed red and green lines); and VOL and NoV MMRs (black and grey solid lines). The average of the three newer observational signal estimates is also shown (AVGOBS; dashed purple line), together with a two-σ estimate of signal uncertainty obtained from the MMSD of the pooled VOL model control-run pseudo-principal components (see Methods). The start date for the calculation of signal trends is 1970 and the initial trend length is ten years. **b**, Noise amplitude as a function of trend length L. For each value of L (which ranges from 10 to 39 years), the noise estimate is the standard deviation of the sampling distribution of L-year, non-overlapping trends obtained from the VOL and NoV control-run pseudo-principal components. **c**, S/N ratio as a function of increasing trend length L. A common VOL model noise estimate (the black line in **b**) was used to calculate S/N. The 1% and 5% significance thresholds are shown (as horizontal black and grey lines respectively) and assume a Gaussian distribution of noise trends in the VOL control-run pseudo-principal components. All observational estimates are infilled, all model data are spatially complete and the global mean is included in all data.

the pooled VOL model control-run pseudo-principal components (see Methods). The NoV signal strength is unrealistically large on timescales greater than 15 years. The uncorrected observational ΔT estimates are also outside of the uncertainty range on the AVGOBS results (except on the longest timescales). The time evolution of the trends in the uncorrected observational data sets is inconsistent with the more linear change evident in the newer data sets and in the model results. In contrast, the evolution of signal trends as a function of L is remarkably similar in the VOL models and in AVGOBS.

In both VOL and NoV models, the amplitude of the internally generated variability decays markedly with increasing L (Fig. 5b). This decrease in the noise amplitude is the primary driver of the pronounced increase in S/N ratios as L increases (Fig. 5c). As expected from Fig. 5a, the multimodel VOL results are nearly superimposed on the AVGOBS estimate, whereas S/N ratios for the NoV and uncorrected observational data are inconsistent with the newer estimates. For all three corrected estimates, the S/N is consistently above a 1% significance level, with ratios greater than four by 2003.

Results in Fig. 5 are for ΔT_{IF}. Use of subsampled data has little impact on our estimated S/N ratios, which still consistently exceed the 1% significance level (Supplementary Fig. S4). The relative insensitivity of S/N ratios to use of subsampled or infilled data results from both observed trends and variability estimates generally being larger for ΔT_{SS} than ΔT_{IF}. In other words, the larger trends compensate for the larger multimodel noise estimates obtained with subsampled data.

Our S/N results are also relatively insensitive to the application of different methods for removing residual drift in simulated ΔT (arising from incomplete simulation spin-up[24–26]; Supplementary Fig. S4). Repeating each of these D&A sensitivity tests with a 1960 start date alters the time evolution of the signal trends, particularly for the earlier uncorrected observational estimates, but again yields S/N ratios well above four for the longest trend lengths (Supplementary Fig. S5).

How would model variability errors affect these results? We note that even our lowest S/N ratios exceed four. To convert these highly

significant results to results that failed to achieve significance at the 5% level would require that the simulated multidecadal ΔT variability used here (for estimating the noise component of S/N ratios) was at least a factor of two smaller than in observations. There is no evidence (from Fig. 3) of an error of this magnitude. On the ten-year timescale for which meaningful comparisons with observed variability are feasible, simulated values of S_{10} for subsampled data (infilled data) are only 25–28% (7–10%) smaller than in observations.

The choice of multimodel fingerprint (VOL or NoV) does not change the overall picture of highly significant S/N ratios (Supplementary Fig. S6). However, the observations project more strongly onto the VOL fingerprint than the NoV fingerprint. The inclusion of volcanic forcings may contribute to this improved agreement, but it may also be related to differences in the physics and parameterizations of the models comprising the VOL and NoV subsets (as well as to differences in other, non-volcanic external forcings).

We also repeated our D&A analysis after first removing the time-evolving global mean temperature change from all data sets. This is a more stringent test of the similarity between modelled and observed temperature changes. In the mean-removed case, we still obtain positive detection of the VOL and NoV model fingerprints in observations in roughly half of the D&A tests. This indicates that there is useful signal information in the subglobal pattern of ΔT, such as the larger warming in the Atlantic than in the Pacific—a feature common to models and observations (Fig. 2).

We have identified a human-induced fingerprint in observed estimates of upper-ocean warming on multidecadal timescales, confirming the results of previous D&A work[2–5]. Our results are robust to the use of multiple bias-corrected observational data sets, to use of infilled or subsampled data, to model signal and noise uncertainties and to different technical choices in simulation drift removal and in the application of our D&A method. There is evidence from our variability comparisons that the models used here may underestimate observed decadal scale variability of basin-average upper-ocean temperatures. However, this variability underestimate would have to be smaller than observed by a factor of more than two to negate our positive identification of an anthropogenic fingerprint in the observed multidecadal warming of the upper 700 m of the oceans. Our analysis provides no evidence of a noise error of this magnitude.

Methods

Ocean temperature data analysis. We examine anomalies of volume average temperature ΔT rather than ocean heat content. Use of ΔT enables us to assess the impact of spatiotemporal changes in the availability of historical temperature data, that is, we do not have to rely on objective infilling where temperature measurements are unavailable. We compare data-only ΔT_{SS} results with results based on the analysis of infilled data sets ΔT_{IF}. As in previous studies[3,4,14,15], observed ΔT_{SS} estimates are based on globally gridded (1° × 1° latitude/longitude) products, not raw measurements. The observed and simulated historical anomalies are with respect to a 1957–1990 climatology and all control-run anomalies are with respect to the overall time mean of each model's control run. Annual means of all model ocean temperature data have been interpolated to the spatial grid and standard vertical depth levels of the observational data to facilitate the subsampling of simulation output at observational grid cells with valid data. We focus on the top 700 m of the ocean column (where the observations are concentrated) and compute basin-scale ΔT changes in the North Atlantic, South Atlantic, North Pacific, South Pacific, North Indian and South Indian oceans.

Model subsampling. Each 20CEN simulation is subsampled in the same manner as previous studies[15] using the 1960–1999 ISH mask of data availability (in space and time). A similar procedure is followed for each control simulation, but in this case the observational mask is repeatedly applied to all possible non-overlapping segments of the entire control. Further information is provided in the Supplementary Information.

Estimating variability of simulated and observed five- and ten-year trends. We focus on the combined spatiotemporal variability of ΔT for two reasons: the number of non-overlapping ten-year trends in any individual basin is small, thus hampering reliable estimation of decadal scale temporal variability in that basin; and for D&A purposes (particularly in the projection of model control-run noise information onto the searched-for fingerprint), it is important for models to reliably portray not only the global mean amplitude of observed decadal timescale ΔT variability, but also the observed basin-scale covariability of ΔT changes. As volcanic forcing can influence variability on multiyear timescales, the VOL model 20CEN simulations are most suitable for direct comparison with observations. After first removing the overall linear trend for each basin, we compute non-overlapping five- and ten-year trends in each basin's regression residuals. We pool these results and calculate the standard deviations of five-year and ten-year trends, yielding the metrics S_5 and S_{10} (respectively). This procedure is carried out for two different analysis periods: 1960–1999 and 1970–1999. For the period 1960–1999, there are four non-overlapping trends (1960–1969, 1970–1979, 1980–1989 and 1990–1999) and seven basins (see Fig. 2), yielding a sample size of 28.

Accounting for simulation drift in individual model signal and noise estimates. Residual drift associated with the incomplete spin-up of model control runs is removed from all ΔT basin-average time series. Quadratic and cubic fits are computed for the entire control, yielding a drift estimate. This drift is then removed from the original control, yielding an estimate of the true model noise. For each 20CEN simulation, there is a contemporaneous section of the corresponding control and a contemporaneous section of the control-drift estimate. This section of the control drift is removed from the 20CEN simulation.

Multimodel fingerprints, noise estimates and detection time. CMIP3 20CEN runs (nominally 1870 to 1999) are averaged together to produce a MMR. If more than one realization of the 20CEN experiment is available for an individual model, these realizations are averaged together before averaging across models. We partition our MMRs by considering separately those models that included the impacts of VOL (CCSM3.0, GFDL-CM2.0, GISS-EH, GISS-ER, MIROC-CGCM2.3.2, MRI-CGCM2.3.2) and those that did not (CCCma-CGCM3.1, CNRM-CM3, CSIRO-Mk3.0, GISS-AOM, FGOALS-g1.0 and UKMO-HadCM3). As defined here, the fingerprint is the first EOF of the MMR of ΔT in the six ocean basins, calculated over 1960–1999. Fingerprints are computed separately for the VOL and NoV MMRs. Use of alternative periods for estimating the (normalized) fingerprint has little impact on its spatial structure. Before the calculation of EOFs, the ΔT anomalies were weighted by the volume of each basin. We also carried out our analysis with both normalized and non-normalized ΔT data, where normalization in each basin is by the temporal standard deviation of the basin-average ΔT time series from the pooled VOL or NoV control runs. Normalization has little impact on our D&A results. For simplicity, we show only non-normalized results.

Our multimodel noise estimates are based on concatenating all available control data for a given subset (for example, the VOL models with subsampling and quadratic-drift removal). An alternative approach would be to select a time slice of an equal length that is determined by the shortest control available. Previous work with column-integrated atmospheric water vapour (using the same CMIP3 multimodel archive) has demonstrated that this choice does not have a significant impact on the overall D&A results[23]. The uncertainty estimate for the AVGOBS results in Fig. 5a was generated as follows. We projected the basin-average upper-ocean temperature changes from the VOL concatenated control runs onto the searched-for fingerprint. This is equivalent to calculating an uncentered spatial covariance between the time-invariant fingerprint (which was estimated from the VOL twentieth-century simulations) and the time-varying control-run temperature changes. The resulting projection time series, $N(t)$, provides information about unforced changes in pattern similarity. We fit L-year, non-overlapping trends to $N(t)$, with L varying from 10, 11, 12,…39 years. For each trend length L, we calculated σ, the standard deviation of the multimodel sampling distribution (MMSD) of L-year trends. The shaded envelope represents the AVGOBS signal trends (which are also a function of the trend length L) $+/-2\times\sigma$. The σ uncertainty estimate could also have been applied to the VOL signal trends, which would have meant truncating the uncertainty envelope in 1999, at the end of the VOL runs. For visual display purposes, we chose to apply this uncertainty estimate to the AVGOBS results, which end at a later date (2006).

We estimate detection times by first projecting (a six-basin scalar product) the ΔT observations and pooled multimodel noise estimates onto the time-invariant model fingerprints. These pseudo-principal component times series are the signal and noise projection time series Z (t) and N (t), respectively. Fitting trends of increasing length L to Z (t) and N (t) allows us to calculate S/N ratios as a function of trend length. For a given value of L, the noise is the standard deviation of the sampling distribution of L-length trends in N (t). The detection time is defined as the year at which S/N exceeds and remains above a stipulated 5% significance threshold. The assumed start time for calculating trends in Z (t) is 1970. We also explore the impact on S/N ratios of use of an earlier 1960 start date (Supplementary Fig. S5).

Received 11 November 2011; accepted 30 April 2012; published online 10 June 2012

References

1. Levitus, S., Antonov, J., Boyer, T. & Stephens, C. Warming of the world ocean. *Science* **287,** 2225–2229 (2000).
2. Barnett, T., Pierce, D. & Schnur, R. Detection of anthropogenic climate change in the world's oceans. *Science* **292,** 270–274 (2001).
3. Barnett, T. *et al.* Penetration of human-induced warming into the world's oceans. *Science* **309,** 284–287 (2005).
4. Pierce, D. W. *et al.* Anthropogenic warming of the oceans: Observations and model results. *J. Clim.* **19,** 1873–1900 (2006).
5. Palmer, M., Good, S., Haines, K., Rayner, N. & Stott, P. A new perspective on warming of the global oceans. *Geophys. Res. Lett.* **36,** L20709 (2009).
6. Gouretski, V. & Koltermann, K. How much is the ocean really warming? *Geophys. Res. Lett.* **34,** L01610 (2007).
7. Domingues, C. *et al.* Improved estimates of upper-ocean warming and multi-decadal sea-level rise. *Nature* **453,** 1091–1094 (2008).
8. Ishii, M. & Kimoto, M. Reevaluation of historical ocean heat content variations with time-varying XBT and MBT depth bias corrections. *J. Oceanogr.* **65,** 287–299 (2009).
9. Levitus, S. *et al.* Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems. *Geophys. Res. Lett.* **36,** L07608 (2009).
10. Ishii, M., Kimoto, M., Sakamoto, K. & Iwasaki, S. Steric sea level changes estimated from historical ocean subsurface temperature and salinity analyses. *J. Oceanogr.* **62,** 155–170 (2006).
11. Levitus, S., Antonov, J. & Boyer, T. Warming of the world ocean, 1955–2003. *Geophys. Res. Lett.* **32,** L02604 (2005).
12. Lyman, J. *et al.* Robust warming of the global upper ocean. *Nature* **465,** 334–337 (2010).
13. Gregory, J., Banks, H., Stott, P., Lowe, J. & Palmer, M. Simulated and observed decadal variability in ocean heat content. *Geophys. Res. Lett.* **31,** L15312 (2004).
14. AchutaRao, K. *et al.* Variability of ocean heat uptake: Reconciling observations and models. *J. Geophys. Res.* **111,** C05019 (2006).
15. AchutaRao, K. M. *et al.* Simulated and observed variability in ocean temperature and heat content. *Proc. Natl Acad. Sci. USA* **104,** 10768–10773 (2007).
16. Gleckler, P. *et al.* Krakatoa's signature persists in the ocean. *Nature* **439,** 675 (2006).
17. IPCC *Special Report on Emissions Scenarios* (eds Nakicenovic, N. & Swart, R.) (Cambridge Univ. Press, 2000).
18. Arblaster, J., Meehl, G. & Karoly, D. Future climate change in the Southern Hemisphere: Competing effects of ozone and greenhouse gases. *Geophys. Res. Lett.* **38,** L02701 (2011).
19. Church, J., White, N. & Arblaster, J. Significant decadal-scale impact of volcanic eruptions on sea level and ocean heat content. *Nature* **438,** 74–77 (2005).
20. Gregory, J. Long-term effect of volcanic forcing on ocean heat content. *Geophys. Res. Lett.* **37,** L22701 (2010).
21. Santer, B. D. *et al.* Separating signal and noise in atmospheric temperature changes: The importance of timescale. *J. Geophys. Res.* **116,** D22105 (2011).
22. Hasselmann, K. in *Meteorology of Tropical Oceans* (ed. Shawn, D. B.) 251–259 (Roy. Met. Soc., 1979).
23. Santer, B. *et al.* Identification of human-induced changes in atmospheric moisture content. *Proc. Natl Acad. Sci. USA* **104,** 15248–15253 (2007).
24. Gleckler, P. *et al.* Krakatoa lives: The effect of volcanic eruptions on ocean heat content and thermal expansion. *Geophys. Res. Lett.* **33,** L17702 (2006).
25. Gregory, J. *et al.* Comparison of results from several AOGCMs for global and regional sea-level change 1900-2100. *Clim. Dynam.* **18,** 225–240 (2001).
26. Gregory, J. M., Lowe, J. A. & Tett, S. F. B. Simulated global-mean sea level changes over the last half-millennium. *J. Clim.* **19,** 4576–4591 (2006).

Acknowledgements

We acknowledge the climate model development groups for providing their simulation output for analysis, PCMDI for collecting and archiving this data and the World Climate Research Programme's Working Group on Coupled Modelling for organizing the model data analysis activity. The CMIP3 multimodel data set is supported by the Office of Science, US Department of Energy. Work at Lawrence Livermore National Laboratory (by P.J.G., P.M.C., B.D.S. and K.E.T.) was carried out under the auspices of the US Department of Energy under contract DE-AC52-07NA27344. C.M.D. was partly financially supported by a CSIRO Office of the Chief Executive Postdoctoral Fellowship, the Australian Climate Change Science Program and the Australian Academy of Science (Scientific Visit to North America Program). D.W.P. was partially funded by the US Department of Energy's International ad hoc Detection and Attribution Group (IDAG).

Author contributions

P.J.G., B.D.S., D.W.P., T.P.Barnett and K.E.T. designed the research. C.M.D., T.P.Boyer and M.I. contributed new analytic tools. P.J.G., B.D.S., C.M.D., J.A.C., T.P.Boyer, P.M.C. and K.M.A. analysed the data. P.J.G., B.D.S., C.M.D., J.A.C., D.W.P. and K.E.T. wrote the paper.

Additional information

The authors declare no competing financial interests. Supplementary information accompanies this paper on www.nature.com/natureclimatechange. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.J.G.