Causes of Ocean Surface Temperature Changes in Atlantic and Pacific Tropical Cyclogenesis Regions

B.D. Santer,[1 *] T.M.L. Wigley,[2] P.J. Gleckler,[1] C. Bonfils,[3] M.F. Wehner,[4] K. AchutaRao,[1] T.P. Barnett,[5] J.S. Boyle,[1] W. Brüggemann,[6] M. Fiorino,[1] N. Gillett,[7] J.E. Hansen,[8] P.D. Jones,[7] S.A. Klein,[1] G.A. Meehl,[2] S.C.B. Raper,[9] R.W. Reynolds,[10] P.A. Stott,[11] K.E. Taylor,[1] and W.M. Washington[2]

[1] Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA; [2] National Center for Atmospheric Research, Boulder, CO 80303, USA; [3] University of California Merced, Merced, CA 95344, USA; [4] Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA; [5] Scripps Institution of Oceanography, La Jolla, CA 92037, USA; [6] Institut für Unternehmensforschung, Universität Hamburg, 22765 Hamburg, Germany; [7] Climatic Research Unit, University of East Anglia, Norwich, NR4 7TJ, UK; [8] NASA/Goddard Institute for Space Studies, New York, NY 10025, USA; [9] Dalton Research Institute, Manchester Metropolitan University, Manchester, M1 5GD, UK; [10] NOAA/National Climatic Data Center, Asheville, NC 28801, USA; [11] Hadley Centre for Climate Prediction and Research, U.K. Met. Office, Exeter, EX1 3PB, UK.

For submission to *Science*. Version: January 31, 2006

* To whom correspondence should be addressed. Email: santer1@llnl.gov

Previous research has identified links between changes in sea surface temperature (SST) and hurricane intensity. We use climate models to study the possible causes of SST changes in Atlantic and Pacific tropical cyclogenesis regions. The observed SST increases in these regions range from 0.32 to 0.67°C over the 20th century. The 22 climate models examined here suggest that century-timescale SST changes of this magnitude cannot be explained solely by unforced variability of the climate system, even under conservative assumptions regarding the magnitude of this variability. Model simulations that include external forcing by combined anthropogenic and natural factors are generally capable of replicating observed SST changes in both tropical cyclogenesis regions.

Hurricane activity is influenced by a variety of physical factors, such as sea surface temperatures (SSTs), wind shear, moisture availability, and atmospheric stability (*1*). Theory, observations, and modeling provide evidence of a direct link between changes in SSTs and hurricane intensity (*2*–*5*). One recent investigation found that secular SST changes in the Atlantic and Pacific tropical Cyclogenesis Regions (ACR, PCR) were highly correlated with a measure of hurricane intensity based on maximum wind speeds (*6*). This research raises an important question: what are the causes of past SST changes in areas where hurricanes develop?

The question of causality is timely in view of the unprecedented level of activity during the 2005 Atlantic hurricane season (*7*), preliminary evidence of a recent increase in the number of category 4 and 5 hurricanes (*8*), and conflicting estimates of the relative contributions of internal climate variability and external forcing to observed SST changes. While some studies suggest that 20th century SST changes in the ACR can be fully explained by internal variability of the climate system (*9*, *10*), other analyses find a substantial anthropogenic component in observed SST and ocean heat content changes (*7*, *11*–*13*).

Previous work has relied on observational data to assess the relative contributions of internal noise and external forcing to SST changes in tropical cyclogenesis regions (*7*, *9*, *10*). Such partitioning is difficult to achieve with observations alone. In the real world, we are performing an uncontrolled geophysical experiment, with humaninduced changes in external climate forcings (such as well-mixed greenhouse gases and various aerosol particles) superimposed on the ever-fluctuating "noise" of natural internal climate variability. We have no control experiment without anthropogenic forcings, which could be used to quantify climate noise. Systematic experimentation can only be performed with numerical models of the climate system.

We use 22 different climate models to estimate the magnitude of century-timescale SST changes arising from internally-generated variability and external forcing. Our focus is on SST changes in the ACR and PCR (*14*). We analyze both 20th century experiments with estimated historical changes in external forcings ("20CEN") and control simulations with no forcing changes (*15*). 20CEN forcings were not standardized across different modeling groups (*16*). The 20CEN results therefore reflect uncertainties in the applied forcings and in the physics and parameterizations of the models themselves. The most comprehensive experiments include changes in both natural external forcings (solar irradiance and volcanic dust loadings in the atmosphere) and in a wide variety of anthropogenic influences (such as well-mixed greenhouse gases, ozone, sulfate and black carbon aerosols, and land surface properties). All simulations were performed with coupled atmosphere-ocean General Circulation Models (A-OGCMs), in which SST changes are predicted.

Model SSTs are compared to the Extended Reconstructed SST (ERSST) dataset of the National Oceanic and Atmospheric Administration (NOAA) (*17*) and the

Hadley Centre sea ice and SST dataset (HadISST) (*18*). The aim of these comparisons is twofold: 1) to determine whether observed SST changes in the ACR and PCR can be explained by internally-generated variability estimated from control simulations; 2) to evaluate how successfully the 20CEN runs capture important features of the observed SST behavior in these two tropical cyclogenesis regions (*e.g.*, the climatological annual mean and seasonal cycle, high and low-frequency SST variability, and century-timescale SST trends). Our use of both ERSST and HadISST data provides information on structural uncertainties in the observations (*19*), which is often of key importance in model evaluation (*15*).

We consider the observations first. In the smoothed ERSST and HadISST data (*20–22*), SSTs in the ACR were at record levels during the 2005 Atlantic hurricane season (Fig. 1A, S1A). The 2005 SST anomaly was smaller in the PCR, and not unprecedented (Fig. 1B, S1B). Observed SSTs in both tropical cyclogenesis regions have increased over the 20th century, with total linear changes in HadISST and ERSST of 0.41 and 0.67°C in the ACR (respectively) and 0.32 and 0.38°C in the PCR (Table 1). Differences between observational datasets primarily reflect the different procedures used by the NOAA and Hadley Centre groups to infill missing SST data (*17*, *18*).

Variability on sub- to multi-decadal timescales is superimposed on these overall increases in observed SSTs (Fig. 1A,B). Commonly-discussed sources of this variability are the El Niño/Southern Oscillation (ENSO) and the Atlantic Multidecadal

Oscillation (AMO) (*7*, *23*). In the ERSST and HadISST data, part of this variability is in phase with fluctuations in the optical depth of stratospheric aerosols produced by massive volcanic eruptions (*24*) (Figs. 1, S1). This result is consistent with the identification of volcanic effects (albeit at much larger spatial scales) in many different climate variables (*25–27*). The relationship between SST variability and stratospheric aerosol optical depth is clearer in the PCR than in the ACR, particularly for the eruption of Mt. Pinatubo in June 1991 (Figs. 1, S1). Regional differences in the observed SST changes after volcanic eruptions are expected, partly because of spatial differences in climate noise (*28*).

Eleven of the 22 historical forcing experiments included some representation of volcanic effects on climate (*16*). The 20CEN results in Fig. 1 are therefore partitioned into two sets, with and without volcanic forcing (V and No-V, respectively) (*29*). The pronounced differences between the V and No-V averages during major eruptions supports the observational evidence of volcanically-induced cooling of SSTs in both tropical cyclogenesis regions.

To assess whether the observed ACR and PCR trends could be due to climate noise alone, we used information from 22 model control runs to generate sampling distributions of the unforced SST trends in these regions (Fig. 2). For each control run, least-squares linear trends were estimated from successive 100-year segments of the ACR and PCR anomaly time series. By combining results from the 22 models, we obtained "multi-model" sampling distributions of unforced SST trends. These distributions were compared with observed and model-simulated SST trends over the 20th century. The null hypothesis that 20th century SST trends could be due to internal variability alone is rejected at the 5% level or better. This result holds for SST trends in all four datasets (ERSST, HadISST, V, and No-V) and in both tropical cyclogenesis regions, and is insensitive to analysis details (Table 1). Our significance testing strategy is conservative: residual control run drift (see Figs. S3, S4) was not subtracted prior to the estimation of trend sampling distributions, and inflates the standard error of the distribution. This makes it more difficult to reject the null hypothesis (*30*).

These results are only as reliable as the model-based estimates of century-timescale climate noise on which they are based. The p-values in Table 1 could be spuriously low if there were a systematic underestimate of internally-generated variability in the models used here – a possibility we tried to guard against by using a large number of models and a conservative significance testing procedure (*30*). Although we lack sufficiently long observational records to evaluate model estimates of century-timescale variability, the data are adequate for assessing simulated SST variability on sub-decadal to decadal timescales.

We use the 20CEN simulations to compare modeled and observed means, variability, and trends (Fig. 3). While most models systematically underestimate the climatological annual-mean SST in the ACR and PCR (Fig. 3A), there is no evidence of such a systematic underestimate in the temporal standard deviation of unfiltered SST anomalies, which is dominated by variability on interannual and ENSO timescales (Fig. 3B). In the ACR (PCR), roughly one-third (two-thirds) of the 60 20CEN realizations overestimate observed SST variability.

The model results in Figures 3A and B show apparent relationships between SST behavior in the ACR and PCR. SST biases in one tropical cyclogenesis region tend to be correlated with biases in the other region (Fig. 3A). There is an even stronger linear relationship between the amplitude of the high-frequency variability in the ACR and PCR (Fig. 3B). The apparent correlation of biases in geographically disparate regions may reflect common underlying causes, such as errors in the large-scale mean state and in the amplitude of tropically-coherent modes of variability. Note that different manifestations of climate noise have relatively little impact on the simulated means and high-frequency variability, as is evident from the small 'spread' between multiple realizations of any individual model's results (Figs. 3A,B).

Model performance in simulating variability on decadal and longer timescales is of most interest here (Fig. 3C). Variability on these timescales constitutes the background noise against which any slowly-evolving forced signal must be detected. In the ACR, the simulated standard deviation of the low-pass filtered (*22*) SST data is systematically lower than observed. Only 5 of the 22 models have 20CEN realizations with standard deviations close to or exceeding observed values. In the PCR, however, 21 of 22 models produce 20CEN realizations with greater than observed low-frequency SST variability. The implications of these results are discussed below.

Compared with Figures 3A and B, Fig. 3C displays much larger differences between the individual realizations of any given model's results. For example, the Parallel Climate Model (PCM) of the National Center for Atmospheric Research (*16*, *31*) has one 20CEN realization with low-frequency SST variability that is very similar to observed values (in both the ACR and PCR), while two other realizations have substantially lower PCR variability than either HadISST or ERSST. This illustrates that a large ensemble size (or long control run) is neccessary to obtain reliable model estimates of low-frequency SST variability. It also suggests that the observed low-frequency SST variability is difficult to determine reliably from the relatively short data records available.

This large between-realization variability is also relevant to comparisons of modeled and observed trends (Fig. 3D). In the ACR and PCR, 20 and 13 (respectively) of the 22 models have at least one realization of the 20th century SST trend that lies within the statistical confidence intervals of the observed results (Fig. 3D). There is no evidence of a systematic model deficiency in simulating the magnitude of 20th century SST trends in the Atlantic tropical cyclogenesis region. In contrast, nearly half of the simulated SST trends in the PCR are larger than the upper statistical confidence interval for the observed trends (*32*).

While not directly relevant to the issue of how well models replicate observed SST variability, it is instructive to consider model performance in simulating the climatological seasonal SST cycle in the ACR and PCR (Fig. 4). The phase and amplitude of the seasonal cycle are primarily driven by the seasonal migration of the thermal equator and the thermal inertia of the mixed layer. Models with V forcing successfully capture the phase and amplitude in the ACR, but slightly overestimate the observed amplitude in the PCR. A model cold bias throughout the seasonal cycle is apparent in both tropical cyclogenesis regions, consistent with Fig. 3A and Table 1.

A striking feature of Fig. 4 is the close correspondence between simulated and observed changes in the seasonal cycle from the first to the second half of the 20th century. Climatological mean SSTs increase in every month. In the absence of countervailing effects, such as increases in atmospheric stability (*33*) and vertical wind shear, these observed SST changes would tend to favor the extension of the Atlantic hurricane season (*34*). The situation is more ambiguous in the PCR, where SSTs are above 26°C throughout the year, and dynamic controls on hurricane activity may be relatively more important (*4*).

Although our work points towards a pronounced influence of external forcing on SST changes in Atlantic and Pacific tropical cyclogenesis regions, it does not separate and quantify the relative contributions of anthropogenic factors and natural external forcing (changes in solar irradiance and volcanic aerosols). Separation is difficult without "single forcing" experiments, in which key climate forcings are varied individually (rather than jointly, as in the 20CEN experiments). Single forcing experiments performed with PCM (*15*, *31*) indicate that increases in well-mixed greenhouse gases are the main driver of century-timescale increases in ACR and PCR SSTs (Fig. 5). PCM's greenhouse-gas induced warming is partly offset by the cooling effects of anthropogenic sulfate aerosol particles, while solar, volcanic, and ozone forcing make much smaller contributions to the simulated SST changes over the 20th century.

In summary, we find that current model estimates of internal climate variability cannot explain observed 20th century SST increases in either the Atlantic or Pacific tropical cyclogenesis regions. This conclusion is insensitive to existing uncertainties in model physics and parameterizations, and to the details of the procedure used to compare SST trends in observations and model control runs (*30*).

Our confidence in this conclusion would be undermined if models substantially underestimated the amplitude of natural internal climate variability. On decadal timescales, where observational records are of sufficient length to make useful model-data comparisons, most current models underestimate SST variability in the ACR and overestimate variability in the PCR (*35*). It is possible that biases of similar magnitude may also apply on the century timescales considered in Fig. 2. Even if they did, however, it is still highly unlikely that climate noise could fully explain the observed SST trends in the ACR (*36*). In the PCR, the evidence against an internal variability explanation is even stronger. The model overprediction of the PCR low-frequency SST variability implies that the observed PCR trends (which are already highly significant) are even less likely to be due to internal variability.

These results, together with other observational and modeling studies (*7*, *37*) do not support claims that internal climate noise accounts for all fluctuations in tropical Atlantic SSTs over the 20th century (*9*, *10*). Our work points towards a large externally-forced component of SST change in the Atlantic and Pacific tropical cyclogenesis regions. In both regions, model simulations with external forcing by combined natural and anthropogenic effects are broadly consistent with observed SST increases. The PCM experiments suggest that forcing by well-mixed greenhouse gases has been the dominant influence on century-timescale SST increases. We also find clear evidence of a volcanic influence on observed SST variability in the ACR and PCR.

Hurricanes are complex phenomena. Ocean surface temperatures are only one of a variety of factors that control their formation and evolution (*1*). Detailed analyses of changes in other large-scale conditions that affect tropical cyclogenesis (such as wind shear and vertical stability) are imperative in order to obtain a more complete understanding of how hurricane activity has changed and may continue to change in a warming world. Our research illustrates that models can be of considerable benefit in understanding the causes of such changes.

References and Notes

1. W. M. Gray, *Mon. Weath. Rev.* **96**, 669 (1968).

2. K. A. Emanuel, *Nature* **326**, 483 (1987).

3. G. J. Holland, *J. Atmos. Sci.* **54**, 2519 (1997).

4. S. C. B. Raper, in *Climate and Sea Level Change: Observations, Projections and Implications*, R. A. Warrick, E. M. Barrow, and T. M. L. Wigley, Eds. (Cambridge Univ. Press, Cambridge, 1993), pp. 192-212.

5. T. R. Knutson, R. E. Tuleya, *J. Clim.* **17**, 3477 (2004).

6. K. A. Emanuel, *Nature* **436**, 686 (2005).

7. K. E. Trenberth, D. J. Shea, *Science* (submitted).

8. P. J. Webster, G. J. Holland, J. A. Curry, H.-R. Chang, *Science* **309**, 1844 (2005).

9. S. B. Goldenberg *et al.*, *Science* **293**, 474 (2001).

10. M. Chelliah, G. D. Bell, *J. Climate* **17**, 1777 (2004).

11. T. P. Barnett, D. W. Pierce, R. Schnur, *Science* **292**, 270 (2001).

12. S. Levitus *et al.*, *Science* **292**, 267 (2001).

13. T. P. Barnett *et al.*, *Science* **309**, 284 (2005).

14. The Atlantic and Pacific tropical cyclogenesis regions used here are identical to those defined in (*6*). Gridded, monthly-mean model and observational SST data were spatially-averaged over 6°N-18°N, 60°W-20°W (ACR) and over 5°N-15°N, 180°E-130°E (PCR). All spatial averages were area weighted, properly accounting for both complete and fractional grid-cells within the region. Since a small portion of the ACR extends into South America, all data points over land were masked out prior to calculation of spatial averages.

15. B. D. Santer *et al.*, *Science* **309**, 1551 (2005).

16. Materials and methods are available as supporting information at *Science* Online.

17. T. M. Smith, R. W. Reynolds, *J. Clim.* **17**, 2466 (2004).

18. N. A. Rayner *et al.*, *J. Climate* (in press).

19. P. W. Thorne, D. E. Parker, J. R. Christy, C. A. Mears, *Bull. Amer. Met. Soc.* (in press).

20. P. Lynch, X.-Y. Huang, *Mon. Weath. Rev.* **120**, 1019 (1992).

21. M. E. Mann, *Geophys. Res. Lett.* **31**, L07214, doi:10.1029/2004GL019569 (2004).

22. For visual display purposes, both the modeled and observed SST data in Figs. 1 and S1 were smoothed using a Lynch/Huang (LH) digital filter (*20*) with a win-

dow width of $T_w = 21$ months, corresponding to a half-power point of 25 months. Variability is damped on interannual and ENSO timescales, while information on the SST response to volcanic forcing is largely preserved. The overall linear trend was subtracted prior to filtering, and reinserted after filtering. To avoid loss of data, $(T_w - 1)/2$ points at the beginning and end of the time series were "reflected", as described in (*21*). To estimate modeled and observed variability on decadal and longer timescales (Fig. 3C), we applied the same LH filter to the detrended SST anomaly data and set T_w to 145 months. This yields a half-power point at 119 months. The response functions for both T_w choices are shown in Fig. S2.

23. J. R. Knight *et al.*, *Geophys. Res. Lett.* **32**, doi:10.1029/2005GL024233 (2005).

24. M. Sato *et al.*, *J. Geophys. Res.* **98**, 22987 (1993).

25. N. P. Gillett *et al.*, *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL020044 (2004).

26. A. Robock, *Rev. Geophys.* **38**, 191 (2000).

27. P. J. Gleckler *et al.*, *Nature* (in press).

28. T. M. L. Wigley *et al.*, *Geophys. Res. Lett.* **27**, 4101 (2000).

29. Ensembles of the 20CEN simulations were performed with 13 of the 22 models analyzed here (*16*). Each ensemble contains multiple realizations of the same experiment, differing only in their initial conditions, but with identical changes in

external forcings. This yields many different realizations of the climate "signal" (the response to the imposed forcing changes) plus climate noise. Averaging over multiple realizations reduces noise and facilitates signal estimation. Here, we calculate averages over V and No-V 20CEN runs. In each case, $\overline{X}$ is the arithmetic mean of the ensemble means (for the models for which these are available) and of individual realizations, *i.e.*, $\overline{X} = \frac{1}{N}\sum_{j=1}^{N}\overline{X}_j$, where N is the total number of V or No-V models (11 here), and $\overline{X}_j$ is the ensemble mean signal (or individual realization) of the j^{th} model. This weighting avoids undue emphasis on results from a single model with a large number of realizations.

30. Note that the sampling distributions in Fig. 2 are slightly asymmetrical, with longer negative tails. This is because the initial SST drift (at least in the ACR and PCR) is often negative in the model control runs examined here (see, *e.g.*, results for GISS-EH in Fig. S3 and CNRM-CM3 in Fig. S4). Another conservative aspect of the significance testing analysis is our use of two-tailed tests, which explicitly consider these longer negative tails. Our conclusions regarding the significance of 20th century SST trends are insensitive to whether we use one- or two-tailed tests, and to our inclusion of residual cooling in estimates of *bona fide* natural internal variability.

31. W. M. Washington *et al.*, *Clim. Dyn.* **16**, 321 (2000).

32. In some models, SST trends in individual 20CEN realizations are negative in the

ACR (see, *e.g.*, MIROC3.2(medres) and UKMO-HadGEM1 results in Fig. 3D). This situation never arises in the PCR, where 20CEN trends are invariably positive and significantly different from zero (at the 5% level or better) in 59 out of 60 cases (*c.f.* 50 out of 60 cases in the Atlantic) (*16*).

33. L. Bengtsson, M. Botzet, M. Esch, *Tellus* **48A**, 57 (1996).

34. Such an extension was observed this year, with the development of tropical storm Zeta in late December 2005. The Atlantic hurricane season ends officially on November 30th.

35. Missing or incorrectly-specified forcings also influence the model-versus-observed variability differences shown in Fig. 3C. For example, the observed decadal variability in ACR and PCR SSTs receives a contribution from volcanic forcing (see Figs. 1, S1), which is neglected in the No-V group of models. This missing forcing must contribute to the No-V models' underestimate of observed SST variability in the ACR.

36. In the ACR, the standard deviation of the sampling distribution of unforced century-timescale SST trends is 0.14°C/century. HadISST and ERSST trends over the 20th century are (respectively) *ca.* 3 and 5 standard deviations away from the mean of the sampling distribution. The models used here would have to underestimate century-timescale SST variability in the ACR by a factor of two or more in order to achieve a non-significant result for the ERSST trend.

37. J. E. Hansen *et al.*, *J. Geophys. Res.* (submitted).

38. Work at Lawrence Livermore National Laboratory (LLNL) was performed under the auspices of the U.S. Dept. of Energy, Environmental Sciences Division, contract W-7405-ENG-48. A portion of this study was supported by the U.S. Dept. of Energy, Office of Biological and Environmental Research, as part of its Climate Change Prediction Program. TMLW was supported by NOAA Office of Climate Programs ("Climate Change Data and Detection") grant NA87GP0105. The authors acknowledge the international modeling groups for providing their data for analysis, the JSC/CLIVAR Working Group on Coupled Modelling (WGCM) and their Coupled Model Intercomparison Project (CMIP) and Climate Simulation Panel for organizing the model data analysis activity, and the IPCC WG1 TSU for technical support. The IPCC Data Archive at Lawrence Livermore National Laboratory is supported by the Office of Science, U.S. Department of Energy. HadISST data were provided by John Kennedy at the Hadley Centre. The authors thank Charles Doutriaux (PCMDI) for computational support.

Figure 1: Time series of monthly-mean, spatially-averaged SST anomalies for the Atlantic (**A**) and Pacific (**B**) tropical cyclogenesis regions (*14*). Observational results are from the NOAA ERSST dataset (*17*). Results for a second observational dataset (HadISST) (*18*) are very similar, and shown in Fig. S1. Model data are from a total of 60 realizations of 20th century climate change (performed with 22 different models), and have been partitioned into two groups, with and without volcanic forcing (V and No-V). All model data were low-pass filtered prior to averaging (*22*). ERSST data were smoothed with the same filter. The yellow and grey envelopes are the 1σ and 2σ confidence intervals for the V averages, calculated (with the smoothed data) at each time t using a sample size $N = 11$ (*29*). Since most of the 20CEN experiments end in December 1999, the V and No-V averages are only calculated until that month. ERSST data are shown through December 2005. All SST anomalies were defined relative to climatological monthly means over 1900 through 1909. This reference period was chosen for visual display purposes only, and has no impact on subsequent trend analyses or variability estimates. An estimate of the stratospheric aerosol optical depth (SAOD) (*24*) is given in (**C**). Dashed vertical lines denote the times of maximum SAOD during major volcanic eruptions. Note that the amplitude of the observed and simulated SST variability is not directly comparable, since the latter was damped by averaging over different realizations and models (*29*).

Figure 2: Comparison between unforced and externally-forced SST changes in the Atlantic (**A**) and Pacific (**B**) tropical cyclogenesis regions. Time series of ACR and

PCR SST anomalies (*14*) were calculated from 22 different model control runs, with anomalies defined relative to the smoothed initial state of the control run (Figs. S3, S4). For each control run, least-squares linear trends were fitted to overlapping 100-year segments of the ACR and PCR anomaly time series. Successive trends overlapped by 90 years. This procedure yields a total of 698 unforced SST trends for each tropical cyclogenesis region (Table S2). The unforced trends are plotted in the form of histograms. Very similar histograms are obtained if trends are fitted to non-overlapping 100-year segments of control run SST data. Also plotted are the observed ERSST and HadISST trends over 1900 to 1999 and the forced trends from the model 20CEN experiments (partitioned into V and No-V averages; Table 1).

Figure 3: Comparison of basic statistical properties of simulated and observed SSTs in the Atlantic and Pacific tropical cyclogenesis regions. Results are for climatological annual means (**A**), temporal standard deviations of unfiltered (**B**) and filtered (*20*) anomaly data (**C**), and least squares linear trends over 1900 to 1999 (**D**). For each statistic, ACR and PCR results are displayed in the form of scatter plots. Model results are individual 20CEN realizations, and are partitioned into V and No-V models (colored circles and triangles, respectively). Observations are from the ERSST and HadISST datasets. All calculations involve monthly-mean, spatially-averaged anomaly data for the period January 1900 through December 1999. For anomaly definition and sources of data, refer to Fig. 1 and (*16*). The brown horizonal and vertical lines in panels **A**-**C** are at the locations of the ERSST and HadISST values, and

facilitate visual comparison of the modeled and observed results. The black crosses centered on the observed trends in panel **D** are the 2σ trend confidence intervals, adjusted for temporal autocorrelation effects (*16*). The brown dashed lines in **D** denote the upper and lower limits of these confidence intervals.

Figure 4: Simulated and observed changes in the climatological seasonal cycle of SSTs in the Atlantic (**A**) and Pacific (**B**) tropical cyclogenesis regions. The dashed and solid lines are the climatological seasonal cycles over the first and second halves of the 20th century. Simulated results are for V models only. The 1σ and 2σ confidence intervals are for the 1950-1999 V averages. The dotted line at 26°C indicates a frequently-quoted "threshold" SST value required for tropical cyclogenesis (*1*, *8*).

Figure 5: Contribution of different external forcings to SST changes in the Atlantic (**A**) and Pacific tropical cyclogenesis regions (**B**). Results are from a 20CEN run and from single-forcing experiments performed with the Parallel Climate Model (PCM) (*31*). Each result is the low-pass filtered average of a four-member ensemble, with window width T_w set to 145 months (*22*). For anomaly definition, refer to Fig. 1. Stratospheric aerosol optical depth (*24*) is also shown (**C**).

Table 1: Statistics for modeled and observed SSTs in the Atlantic and Pacific tropical cyclogenesis regions (*14*). Results are for climatological annual means (°C) and least-squares linear trends (°C/century). Means and trends were calculated over the period January 1900 through December 1999 using monthly-mean, spatially-averaged

anomaly data (with anomalies defined as in Fig. 1). Model results are from 20CEN integrations performed with 22 different climate models, partitioned into V and No-V groups (*16*). For the V and No-V means and trends, the 1σ standard deviations are shown, in each case based on a sample size $N = 11$. Observational results are from the ERSST (*17*) and HadISST (*18*) data sets. The 2σ confidence intervals on the observed trends are adjusted for temporal autocorrelation effects (*16*). The p-values are estimates of the probability that the SST trend over the 20th century could be due to (model-simulated) natural internal variability alone. Probabilities are based on one- and two-tailed tests (p-value(1) and p-value(2), respectively; see Fig. 2).

Table 1: Statistics for modeled and observed SSTs in the Atlantic and Pacific tropical cyclogenesis regions.

Region	Statistic	V models	No-V models	ERSST	HadISST
ACR	Mean	25.36 ± 0.46	25.55 ± 1.02	26.35	26.49
ACR	Trend	0.37 ± 0.25	0.55 ± 0.26	0.67 ± 0.29	0.41 ± 0.26
ACR	p-value(1)	0.012	0.000	0.000	0.012
ACR	p-value(2)	0.048	0.000	0.000	0.048
PCR	Mean	27.74 ± 0.33	27.81 ± 0.96	28.43	28.48
PCR	Trend	0.59 ± 0.22	0.60 ± 0.29	0.38 ± 0.15	0.32 ± 0.13
PCR	p-value(1)	0.000	0.000	0.000	0.000
PCR	p-value(2)	0.012	0.012	0.012	0.024

Supporting Online Material

Observational Data

We used version 2 of the NOAA Extended Reconstructed Sea Surface Temperature dataset (ERSST) (*S1*) and version 2 of the Hadley Centre Sea Ice and SST data (*S2*). ERSST data were available from January 1854 to present in the form of monthly means on a regular $2° \times 2°$ latitude/longitude grid. Reconstruction of high-frequency SST anomalies involved fitting to a set of spatial modes. HadISST data were available as monthly means on a $1° \times 1°$ latitude/longitude grid, and spanned the period from January 1870 to present. Reconstruction of SSTs in data-sparse locations relied on a two-stage reduced-space optimal interpolation procedure. Further details of the ERSST and HadISST datasets are available online (*S3*, *S4*).

Modeling groups contributing to IPCC database

At the time this research was conducted, 15 modeling groups had performed a wide range of simulations in support of the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC AR4). Climate data from these simulations were made available to the scientific community through the U.S. Dept. of Energy's Program for Climate Model Diagnosis and Intercomparion (PCMDI). Six modeling groups provided results for at least two different model configurations. Results from a total of 22 different climate models were analyzed.

We considered two sets of simulations here: pre-industrial control runs, and 20CEN experiments with historical changes in a number of different anthropogenic and natural forcings. In IPCC terminology, these integrations are referred to as “picntrl” and “20c3m” (respectively).

Official designations of the 15 modeling groups listed below (with official model acronyms in brackets):

1. Bjerknes Center for Climate Research, Norway [BCCR-BCM2.0].

2. Canadian Centre for Climate Modelling and Analysis, Canada [CCCma-CGCM3.1(T47) and CCCma-CGCM3.1(T63)].

3. National Center for Atmospheric Research, U.S.A. [CCSM3 and PCM].

4. Météo-France/Centre National de Recherches Météorologiques, France [CNRM-CM3].

5. Commonwealth Scientific and Industrial Research Organization (CSIRO) Atmospheric Research, Australia [CSIRO-Mk3.0].

6. Max-Planck Institute for Meteorology, Germany [ECHAM5/MPI-OM].

7. Meteorological Institute of the University of Bonn, Meteorological Research Institute of the Korean Meteorological Agency, and Model and Data group, Germany/Korea [MIUB/ECHO-G].

8. Institute for Atmospheric Physics, China [FGOALS-g1.0].

9. Geophysical Fluid Dynamics Laboratory, U.S.A. [GFDL-CM2.0 and GFDL-CM2.1].

10. Goddard Institute for Space Studies, U.S.A. [GISS-AOM, GISS-EH, and GISS-ER].

11. Institute for Numerical Mathematics, Russia [INM-CM3.0].

12. Institute Pierre Simon Laplace, France [IPSL-CM4].

13. Center for Climate System Research, National Institute for Environmental Studies, and Frontier Research Center for Global Change, Japan [MIROC-CGCM2.3.2(medres) and MIROC-CGCM2.3.2(hires)].

14. Meteorological Research Institute, Japan [MRI-CGCM2.3.2].

15. Hadley Centre for Climate Prediction and Research, U.K. [UKMO-HadCM3 and UKMO-HadGEM1].

Forcings used in 20CEN runs

Details of the natural and anthropogenic forcings used by differing modeling groups in their IPCC 20CEN simulations are given in Table S1. This Table was compiled using information that participating modeling centers provided to PCMDI (see http://www-pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation.php).
All model acronyms used in the Table are defined in the previous Section.

A total of 11 different forcings are listed in Table S1. A letter 'Y' denotes inclusion of a specific forcing. As used here, 'inclusion' signifies the specification of time-varying forcings, with changes on interannual and longer timescales. Forcings that were varied over the seasonal cycle only, or not at all, are identified with a dash. A question mark indicates a case where there is uncertainty regarding inclusion of the forcing.

Results in Table 1 are stratified by inclusion or omission of volcanic forcing (V and No-V, respectively). Nine of the 11 V models explicitly incorporated volcanic aerosols. Two models – MRI-CGCM2.3.2 and MIUB/ECHO-G – represented volcanic effects in a more indirect manner, using estimated volcanic forcing data from (*S5*) and (*S6*) (respectively) to adjust the solar irradiance at the top of the model atmosphere.

While all 15 modeling groups used very similar changes in well-mixed greenhouse gases, the changes in other forcings were not prescribed as part of the experimental design. In practice, each group employed different combinations of 20th century forcings, and often used different datasets for specifying individual forcings. End dates for the experiment varied between groups, and ranged from 1999 to 2003.

Calculation of temporal standard deviations

All temporal standard deviations in Figs. 3B and C were estimated from linearly detrended data. This was done because some of the model simulations examined here (and the ERSST data in the ACR) have large century-timescale SST trends, which inflate the temporal variance.

Calculation of Confidence Intervals for Linear Trends

The black crosses in Fig. 3D are the 'adjusted' 2σ confidence intervals for b, the slope parameter of the estimated least-squares linear trend in the observed data (*S7*). The adjustment for temporal autocorrelation assumes a lag-1 autocorrelation structure of the trend residuals, $e(t)$. The lag-1 autocorrelation coefficient of $e(t)$ is used to compute an effective sample size, n_e , and to adjust s_b , the standard error of b. Strong temporal autocorrelation of $e(t)$ results in $n_e \ll n$ (the actual number of time samples) and inflates s_b.

Testing Whether Linear Trends are Significantly Different from Zero

Tests of the null hypothesis that b is not significantly different from zero are mentioned in the discussion of Fig. 3D. These tests involve the ratio b/s_b, which is assumed to be distributed as Student's t. Some studies have assumed (incorrectly) that values of $e(t)$ are statistically independent, thus biasing estimates of s_b and trend significance. All significance estimates for the trends in Fig. 3D involve a one-tailed Student's t-test of the null hypothesis that b is not significantly different from zero, with n_e used for calculating s_b and determining the critical t value (*S7*).

Captions for Figures in Supporting Online Material

Figure S1: Time series of monthly-mean, spatially-averaged SST anomalies for the Atlantic (**A**) and Pacific tropical cyclogenesis regions (**B**). Results are for ERSST (*S1*) and HadISST data (*S2*) and were low-pass filtered, with filtering options as for Fig 1. To facilitate dataset intercomparison, anomalies are defined relative to climatological monthly means over 1971-2000, a period of relatively complete and stable observational coverage. Stratospheric aerosol optical depth (**C**) is from (*S5*).

Figure S2: Response function for the Lynch/Huang digital filter for two different choices of the window width T_w (21 and 145 months). These yield half-power at 25 and 119 months, respectively (*S8*).

Figure S3: Time series of unforced variations in monthly-mean, spatially-averaged SST anomalies for the Atlantic tropical cyclogenesis region. Results are from the IPCC AR4 pre-industrial control runs (V models only). Since modeling groups assumed different start dates for their pre-industrial control runs (see Table S2), the first month of each control integration was arbitrarily set to January 1800. This facilitates variability comparisons across models. All anomalies were defined relative to climatological monthly means over the initial decade of the control, which visually highlights control run drift. Low-pass filtering options are as for Fig. 4, with window width $T_w = 145$ months.

Figure S4: As for Fig. S3, but for No-V models.

Captions for Tables in Supporting Online Material

Table S1: Forcings used in IPCC 20CEN simulations. Results are partitioned into V and No-V models (first and last 11 rows, respectively).

Table S2: Technical details of IPCC 20CEN runs and pre-industrial control integrations. The AGCM resolution is given for both spectral models (in terms of the triangular truncation; *e.g.*, T30, T42, *etc.*) and grid-point models (in terms of the latitude-longitude spacing of grid-points). N_r is the number of realizations that were used for calculating 20CEN ensemble means. CTL_1, CTL_N, and L are (respectively) the first year, last year, and length (in years) of the pre-industrial control runs employed for estimating the sampling distributions of unforced SST trends shown in Fig. 2. $N_c(1)$ and $N_c(2)$ are (respectively) the number of non-overlapping and overlapping 100-year linear trends estimated from each control run. For $N_c(2)$, successive trends overlap by 90 years. Note that the HadGEM1 control run, which commenced in December 1859 and ended in December 2099, has missing data in June 1926 and April 2099. The HadGEM1 control run data used here span the period January 1928 to December 2098.

Supporting References and Notes

S1. T. M. Smith, R. W. Reynolds, *J. Clim.* **17**, 2466 (2004).

S2. N. A. Rayner *et al.*, *J. Climate* (in press).

S3. http://www.ncdc.noaa.gov/oa/climate/research/sst/sst.html#grid.

S4. http://www.hadobs.org.

S5. M. Sato *et al.*, *J. Geophys. Res.* **98**, 22987 (1993).

S6. T. J. Crowley, *Science* **289**, 270 (2000).

S7. B. D. Santer *et al.*, *J. Geophys. Res.* **105**, 7337 (2000).

S8. P. Lynch, X.-Y. Huang, *Mon. Weath. Rev.* **120**, 1019 (1992).

Table S1: Forcings used in IPCC simulations of 20th century climate change.

	Model	G	O	SD	SI	BC	OC	MD	SS	LU	SO	VL
1	CCSM3	Y	Y	Y	-	Y	Y	-	-	-	Y	Y
2	GFDL-CM2.0	Y	Y	Y	-	Y	Y	-	-	Y	Y	Y
3	GFDL-CM2.1	Y	Y	Y	-	Y	Y	-	-	Y	Y	Y
4	GISS-EH	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
5	GISS-ER	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
6	MIROC3.2(medres)	Y	Y	Y	?	Y	Y	Y	Y	Y	Y	Y
7	MIROC3.2(hires)	Y	Y	Y	?	Y	Y	Y	Y	Y	Y	Y
8	MIUB/ECHO-G	Y	-	Y	Y	-	-	-	-	-	Y	Y
9	MRI-CGCM2.3.2	Y	-	Y	-	-	-	-	-	-	Y	Y
10	PCM	Y	Y	Y	-	-	-	-	-	-	Y	Y
11	UKMO-HadGEM1	Y	Y	Y	Y	Y	Y	-	-	Y	Y	Y
1	BCCR-BCM2.0	Y	-	Y	-	-	-	-	-	-	-	-
2	CCCma-CGCM3.1(T47)	Y	-	Y	-	-	-	-	-	-	-	-
3	CCCma-CGCM3.1(T63)	Y	-	Y	-	-	-	-	-	-	-	-
4	CNRM-CM3	Y	Y	Y	-	Y	-	-	-	-	-	-
5	CSIRO-Mk3.0	Y	-	Y	-	?	?	?	?	?	?	-
6	ECHAM5/MPI-OM	Y	Y	Y	Y	-	-	-	-	-	-	-
7	FGOALS-g1.0	Y	-	Y	?	-	-	-	-	-	-	-
8	GISS-AOM	Y	-	Y	-	-	-	-	Y	-	-	-
9	INM-CM3.0	Y	-	Y	-	-	-	-	-	-	Y	-
10	IPSL-CM4	Y	-	Y	Y	-	-	-	-	-	-	-
11	UKMO-HadCM3	Y	Y	Y	Y	-	-	-	-	-	-	-

G = Well-mixed greenhouse gases

O = Tropospheric and stratospheric ozone

SD = Sulfate aerosol direct effects

SI = Sulfate aerosol indirect effects

BC = Black carbon

OC = Organic carbon

MD = Mineral dust

SS = Sea salt

LU = Land use change

SO = Solar irradiance

VL = Volcanic aerosols.

Table S2: Technical details of IPCC 20CEN runs and pre-industrial control integrations.

	Model	AGCM resolution	N_r	CTL_1	CTL_N	L	$N_c(1)$	$N_c(2)$
1	CCSM3	T85	5	280	509	230	2	14
2	GFDL-CM2.0	$2.0° \times 2.5°$	3	1	500	500	5	41
3	GFDL-CM2.1	$2.0° \times 2.5°$	3	1	500	500	5	41
4	GISS-EH	$4.0° \times 5.0°$	5	1880	2279	400	4	31
5	GISS-ER	$4.0° \times 5.0°$	5	1901	2400	500	5	41
6	MIROC3.2(medres)	T42	3	2300	2799	500	5	41
7	MIROC3.2(hires)	T106	1	1	100	100	1	1
8	MIUB/ECHO-G	T30	5	1860	2200	341	3	25
9	MRI-CGCM2.3.2	T42	5	1851	2200	350	3	26
10	PCM	T42	4	451	1079	629	6	53
11	UKMO-HadGEM1	$1.25° \times 1.875°$	1	1927	2098	172	1	8
1	BCCR-BCM2.0	T63	1	1850	2099	250	2	16
2	CCCma-CGCM3.1(T47)	T47	5	1850	2850	1001	10	91
3	CCCma-CGCM3.1(T63)	T63	1	1850	2199	350	3	26
4	CNRM-CM3	T63	1	1930	2429	500	5	41
5	CSIRO-Mk3.0	T63	1	1871	2250	380	3	29
6	ECHAM5/MPI-OM	T63	3	2150	2655	506	5	41
7	FGOALS-g1.0	T42	3	1850	2199	350	3	26
8	GISS-AOM	$3.0° \times 4.0°$	2	1850	2100	251	2	16
9	INM-CM3.0	$4.0° \times 5.0°$	1	1871	2200	330	3	24
10	IPSL-CM4	$2.5° \times 3.75°$	1	1860	2359	500	5	41
11	UKMO-HadCM3	$2.5° \times 3.75°$	1	1859	2200	342	3	25
	TOTAL	-	60	-	-	-	84	698