

Tropical Temperature Trends in AMIP Simulations and the Impact of SST Uncertainties

T.J. Flannaghan¹, S. Fueglistaler^{1,2}, I.M. Held^{1,3}, S. Po-Chedley⁴, B.

Wyman³, M. Zhao⁵

Corresponding author: S. Fueglistaler, Department of Geosciences, Princeton University, Princeton, NJ 08544, USA. (stf@princeton.edu)

¹Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ.

²Department of Geosciences, Princeton University, Princeton, NJ.

³Geophysical Fluid Dynamics Laboratory, Princeton, NJ.

⁴Department of Atmospheric Sciences, University of Washington, Seattle, WA.

⁵University Cooperation for Atmospheric Research.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/2014JD022365

Abstract.

The comparison of trends in various climate indices in observations and models is of fundamental importance for judging the credibility of climate projections. Tropical tropospheric temperature trends have attracted particular attention as this comparison may suggest a model deficiency [Santer et al., 2005; Christy et al., 2007, 2010; Fu et al., 2011; Thorne et al., 2011].

One can think of this problem as composed of two parts: one focused on tropical surface temperature trends and the associated issues related to forcing, feedbacks, and ocean heat uptake; and a second part focusing on connections between surface and tropospheric temperatures and the vertical profile of trends in temperature. Here, we focus on the atmospheric component of the problem. We show that two ensembles of GFDL HiRAM model runs (similar results are shown for NCAR's CAM4 model) with different commonly used prescribed sea surface temperatures (SSTs), namely the HadISST1 and 'Hurrell' data sets, have a difference in upper tropical tropospheric temperature trends ($\sim 0.1\text{K}/\text{decade}$ at 300 hPa for the period 1984-2008) that is about a factor 3 larger than expected from moist adiabatic scaling of the tropical average SST trend difference. We show that this surprisingly large discrepancy in temperature trends is a consequence of SST trend differences being largest in regions of deep convection. Further, trends, and the degree of agreement with observations, not only depend on SST data set and the particular atmospheric temperature data set, but also on the period chosen for comparison. Due to the large impact on atmospheric temperatures, these sys-

tematic uncertainties in SSTs need to be resolved before the fidelity of climate models' tropical temperature trend profiles can be assessed.

Accepted Article

1. Introduction

Comparison of temperature trends from CMIP5 model runs with prescribed SSTs (referred to as AMIP simulations due to their historical role in the Atmospheric Model Intercomparison Project [Gates, 1992]) for the period 1981-2008 show that some models do better than others when compared to observations [Po-Chedley and Fu, 2012a]. Mitchell et al. [2013] show temperature trends of the CMIP5/ AMIP runs for the period 1979-2008, and argue that the model tropical tropospheric temperature trends are within the statistical uncertainty of observations. Conversely, Po-Chedley and Fu [2012a] note that nearly all AMIP models overestimate warming in the tropical upper troposphere, but those models that perform best when compared to the observations use the HadISST1 data set [Rayner, 2003], whereas the other models use a different data set [Hurrell et al., 2008], henceforth referred to as the Hurrell data set. Previous studies have discussed using AMIP simulations to analyse the implications of different SST data sets [Hurrell and Trenberth, 1999], and differences between Microwave Sounding Unit (MSU) products and SST reconstructions, using AMIP simulations to connect the two [Hurrell and Trenberth, 1997]. Here we run the Geophysical Fluid Dynamics Laboratory (GFDL) HiRAM model with the HadISST1 and Hurrell SST data sets, and show that subtle (but systematic) differences in the two SST data sets induce an unexpectedly large difference in upper tropospheric temperature trends, and that conclusions regarding trends are sensitive to period chosen. Similar results are obtained with the NCAR CAM4 model, pointing to a large uncertainty in atmospheric temperature trends induced by uncertainties in sea surface temperature data.

2. Methods

2.1. HiRAM model setup

The primary atmospheric model used in this study is the HiRAM model [Zhao et al., 2009] developed at the Geophysical Fluid Dynamics Laboratory, running at an approximate spatial resolution of 50 km in the horizontal and with 32 levels in the vertical, of which 10 are between 500 and 50 hPa. The moist convection parametrization is described by [Zhao et al., 2009, 2012]. The parametrized convection is rather strongly inhibited compared to many other models, resulting in a larger fraction of tropical rainfall occurring on resolved scales. This model has proven to be particularly useful in studying tropical cyclone statistics, including interannual variability, recent trends, and the response to climate change [Zhao et al., 2009; Held and Zhao, 2011; Zhao et al., 2012; Zhao and Held, 2012; Zhao et al., 2013]. The model uses a horizontal grid with the topology of a cube. Both 50 km (180 x 180 grid points on each face of the cube) and 25 km (360 x 360) versions of the model have been analysed. A three member ensemble of AMIP simulations from 1979-2008 using the 50 km model, and a two member ensemble of simulations with the 25 km model, were deposited in the CMIP5 archive [Hurrell et al., 2011]. These runs used the HadISST1 data set as boundary conditions for continuity with previous work. In this paper we focus on the 50 km model and make use of the 3-member ensemble in the CMIP5 archive plus an additional 3-member ensemble using the Hurrell data set, which is used by most of the other AMIP simulations in the archive.

All of these simulations include changing forcing agents: volcano-generated and anthropogenic aerosols, well-mixed greenhouse gases, prescribed time-varying ozone concentrations, and variations in the incoming solar irradiance. There are no prescribed changes

in land use. The sensitivity of model temperature trends to atmospheric forcings (as compared to the sensitivity to SSTs) is discussed in Section 3.3 below.

All quantities are computed using the ensemble average of the tropical average (20°S{20°N) data, unless otherwise stated. We calculate trends for the period 1981{2008 to avoid previously noted problems in the SST data before 1981 [Po-Chedley and Fu, 2012a], and for the period 1984{2008 to avoid suspiciously large differences in the two SST data sets over the period 1982/ 1983 identified below. The calculation of the statistical uncertainty in trends is described in Appendix A.

2.2. CAM4 model setup

In order to evaluate the sensitivity of results to choice of atmospheric general circulation model, we performed 2 additional model runs with the NCAR Community Earth System Model (CESM) version 1 with CAM4 atmospheric physics [Neale et al., 2010]. The model is run with 1.9 x 2.5 degree horizontal resolution and 26 vertical levels, with 1 run based on HadISST1 and 1 run based on the Hurrell SSTs. For both runs, the atmospheric forcings are perpetual year 2000 conditions (see Section 3.2 below).

2.3. Atmospheric temperature observations

The MSU temperature record is based on measurements from different satellites, and various groups have attempted to homogenize the measurements and remove biases. We use the Remote Sensing Systems (RSS) version 3.3 data [Mears and Wentz, 2009a, b], the University of Alabama in Huntsville (UAH) version 5.6 data [Christy et al., 2003], and the University of Washington (UW) data (Po-Chedley, S., T.J. Thorsen, Q. Fu; Removing diurnal cycle contamination in satellite-derived tropospheric temperatures: Understanding

tropical tropospheric trend discrepancies, *Journal of Climate*, in review, 2014). Problems in the UAH data homogenization have been pointed out [Po-Chedley and Fu, 2012a; Po-Chedley and Fu, 2012b, 2013], (Po-Chedley et al., in review), and the UAH data is shown here for comparison with previously published results.

The MSU instrument is processed into channels that have weight in the lower troposphere (TLT), the mid troposphere (TMT) and the lower stratosphere (TLS). The TMT channel has a significant stratospheric component (which contaminates the tropospheric trends with stratospheric trends that are controlled by very different processes), much of which can be removed by taking a linear combination of TMT and TLS [Fu et al., 2004], resulting in the TTT channel (also referred to as T24). Equivalent quantities are calculated for each ensemble using the weighting functions provided by RSS.

2.4. Sea surface temperature data

The two SST data sets (HadISST1 and Hurrell) are based on a very similar set of observations. Both are based on ship tracks and buoy data from the Comprehensive Ocean-Atmosphere Data Set (COADS) project [Woodruff et al., 1987]. After November 1981, satellite data from the Advanced Very High Resolution Radiometer (AVHRR) instrument is assimilated into both products. However, the methods for assimilation and bias correction differ between the two. HadISST1 uses the Reduced Space Optimum Interpolation method [Kaplan et al., 1997], whereas Hurrell uses the NOAA Optimum Interpolation v2 (OI v2) product [Reynolds et al., 2002]. The OI v2 procedure has a higher spatial resolution, and therefore the Hurrell data set represents regions with high gradients in SST such as the Gulf Stream much better than the HadISST1 data set. However, it is

not a priori clear if more detail in the SSTs is important for the tropical tropospheric warming problem or which data set is more suitable for such studies.

3. Results

3.1. Temperature trends

Figure 1a shows the deseasonalized TTT upper tropospheric temperatures averaged over $20^{\circ}\text{S}\{20^{\circ}\text{N}$ from the AMIP configurations used in this study, and the RSS TTT data. AMIP simulations at first glance closely match observations, especially regarding interannual variability as noted previously [Hurrell and Trenberth, 1997], not surprisingly given the tight vertical coupling in the tropical troposphere. The standard picture of the tropics involves convection placing the atmosphere on a moist adiabat, with wave dynamics maintaining temperatures that are horizontally uniform, by adjusting the temperature of the part of the tropical atmosphere not actively convecting to more or less closely match this moist adiabat. But one needs to ask "which moist adiabat?" given the spatially inhomogeneous temperature trends at the surface [Sobel et al., 2002].

Figure 1b shows the tropical mean ($20^{\circ}\text{S}\{20^{\circ}\text{N}$) monthly mean of both SST data sets (deseasonalized), with the difference shown in Figure 1c. Before November 1981, the data sets are offset by approximately a constant, as during this period Hurrell uses HadISST1 combined with a fixed climatology. After November 1981, AVHRR satellite data are used, and the two data sets diverge rapidly, with a large discrepancy between the two during 1982 and 1983. Whatever the explanation for the relatively large differences in the 1981-1983 period, this divergence motivates us to consider both the 1981 and 1984 starting points when comparing model trends with observed trends. From 1984 onwards, the difference between the two is more stable until the early 1990's, when first Hurrell

cools relative to HadISST1, then warms in the late 1990's, and finally cools again after 2005.

Figure 2(a,c) shows the profiles of the two ensemble mean temperature trends, the trends in the TTT channel for the RSS (black), UAH (green) and UW (purple) MSU data sets, and the TTT trends for the two model ensembles. The TTT trend for each ensemble is plotted at the pressure level where the ensemble temperature trend is equal to the ensemble TTT trend. The UAH, RSS and UW trends are plotted at the average of the levels used for each ensemble. The levels have no further significance other than allowing us to plot the MSU data on the same figure as the ensemble temperature trends.

All numerical values are listed in Table 1.

For the 1981{2008 period, the HadISST1 ensemble is very consistent with the RSS MSU data for the TTT channel, while the Hurrell ensemble agrees very well with the UW MSU data. Conversely, the trend in the UAH MSU data is much smaller than either model result. For the 1984{2008 period, the trend in the RSS MSU data is similar, and the trend in the UW MSU data is nearly identical to the Hurrell model ensemble trend. Conversely, the trend in the UAH data is similar to, but still smaller than the HadISST1 model ensemble trend, which in turn is nearly a factor 2 smaller than the Hurrell model ensemble trend. We have argued above that the SSTs from 1981 to 1983 are less reliable. Consequently, the excellent agreement between the HadISST1 ensemble and RSS for the period 1981{2008 noted by Po-Chedley and Fu [2012a] may be fortuitous. Without preferring one SST data for use as boundary conditions in these AMIP simulations and one particular MSU data set, no conclusion is possible concerning the possibility that the model tropospheric trends amplify too much with increasing height.

The differences in simulated tropospheric temperature trends between model runs using HadISST1 and Hurrell are significant compared to the trends themselves, focusing attention on the uncertainty in the SST data sets. In both periods, the tropospheric trend profile (Figure 2, blue and red solid lines) shows the trend increase with height qualitatively similar to that of a moist adiabat. Figure 2(b,d) shows the ratio of the difference in atmospheric trend to the difference in the SST trend between the two ensembles (black). Relative to the trend difference in the two SST data sets, the atmospheric temperature trends show an amplification that peaks around 200 hPa by a factor of ~ 6 . The green dotted lines in Figure 2(b,d) show the amplification of surface temperature trends with height as expected from simple moist adiabatic scaling. Compared to the expectation from simple moist adiabatic scaling of the trend difference in tropical average SSTs, the atmospheric trend difference is about a factor 3 too large. While moist adiabatic scaling is not expected to perfectly capture tropical temperature profile changes (see e.g. discussion in Singh and O’Gorman [2012]), this very large discrepancy is disturbing. Indeed, one might have expected that moist adiabatic scaling performs better for the trend difference (where only SSTs differ) than for the trends in each calculation where also atmospheric forcings vary with time.

We show in Section 4 below that this conundrum can be resolved when considering the relation between the distributions of deep convection and sea surface temperatures, but first discuss some sensitivities of the model results.

3.2. AMIP versus coupled model runs

The limitations of AMIP simulations have been discussed e.g., by Douville [2005]; Copsey et al. [2006]; Emanuel and Sobel [2013]. The importance of these limitations

depends on the problem being addressed. We argue here that trends in tropical tropospheric temperatures can be studied with AMIP simulations, based on the analysis of an atmosphere/land model with the lower boundary conditions provided by a coupled model using the identical atmosphere/land components. This perfect model test is performed using the coupled CM2.1 model [Delworth et al., 2006].

We find that the AMIP framework results in upper tropospheric temperatures that are systematically warmer than the coupled model, by roughly 0.15 K at 300hPa (not shown). Results for the vertical profile of trends in the tropical troposphere are shown in Figure 3(a,b). The AMIP model simulates the fully coupled model's trends well, underestimating them by a few percent. For example, at 300 hPa the trend is $1.0:4\text{K}/\text{decade}$ and the trend difference is $1.0:035\text{K}/\text{decade}$. (This coupled model generates larger temperature trends than AMIP models because it does not simulate the recent hiatus in warming of ocean surface temperatures.) This difference in trends is statistically significant in the upper troposphere. While this difference is interesting and not well understood, it is an order of magnitude smaller than the trend. Further tests of the limitations of the AMIP framework are desirable, but we show below that both the HiRAM and CAM4 atmospheric temperature trend differences are well understood without having to consider this problem.

3.3. Atmospheric forcings

The HiRAM model runs analysed in this paper are simulations forced with changing greenhouse gases and aerosols along with the prescribed sea surface temperature. To assess the relative importance of the sea surface temperature forcing to the other forcings, we run the model (3 ensemble members) with HadISST1 sea surface temperatures, but with

unchanging greenhouse gases and aerosols (referred to as 'unforced runs'). Figure 3(c) shows the trend profile for both ensembles as well as for the forced Hurrell ensemble. The largest differences between the forced/unforced ensembles are in the lower troposphere, where the unforced ensemble has a lower trend, and above 200 hPa where the difference is primarily due to ozone trends. The difference in lower tropospheric trends is mostly due to the trends over land surfaces. Figure 3(d) shows that the effect of the greenhouse gas and aerosol forcings on the mid and upper troposphere is much less than the effect of changing the SST forcing data set. The difference between the ensembles is not statistically significant at the 95% level above 500 hPa, and is smaller than results from similar calculations previously reported [Santer et al., 2005], possibly due to differences in treatment of ozone.

3.4. CAM4 model results

Figure 3(e,f) shows that the CAM4 model runs give very similar trends as the GFDL/HiRAM ensemble model runs. Most importantly, the tropical tropospheric trend differences between the Hurrell and HadISST1 based runs peak around 300 hPa at 0:10K/decade for the period 1984-2008 as do the GFDL/HiRAM runs (see Figure 2(c), difference between red and blue solid lines). In Section 4, we show that for both models these differences are a consequence of differences in the two SST data sets, with model differences playing only a secondary role.

3.5. Sensitivity of trends to period

Figure 2 shows trends in terms of the period 1981-2008 to allow comparison with previously published results, and for the period 1984-2008 which omits the period 1982/83

that we identified as particularly problematic. Figure 4 shows the sensitivity of the trend to the start date for all start dates up to 1996. The trends in all quantities vary with start year, which may be expected for data with substantial interannual variability. More importantly, the differences between the HadISST1 and Hurrell based ensembles are also dependent on the start date, which indicates that the year to year variability in the difference between the two SST datasets (Figure 1c) has a substantial impact on atmospheric temperature trends.

Figure 4(a) shows that the temperature trends determined with ordinary least squares regression from the 3 MSU data sets are typically between those of the HadISST1 (blue) and Hurrell (red) ensembles. For a start date between 1981 and 1990, the Hurrell ensemble trends are remarkably consistent with those of the UW MSU data, but after the early 1990's all MSU data sets agree better with the HadISST1 ensemble mean. The same calculation with the more robust Theil-Sen estimator for trends [Sen, 1968; Lanzante, 1996], which is less sensitive to start/end date, shows better agreement with the Hurrell ensemble mean for all MSU data from the mid-1980's onwards (Figure 4b).

4. Discussion

Tropical deep convection occurs preferentially over regions of anomalously high SSTs, and tropical average SSTs may evolve differently than SSTs in regions of deep convection. In order to characterise the tropical average surface conditions at the locations of deep convection, we average the SSTs with a weighting given by the precipitation distribution (as in Sobel et al. [2002]), which is equivalent to the column integrated latent heating

distribution. This precipitation-weighted temperature T_p is defined as

$$T_p = \frac{\sum T_s P_i}{\sum P_i}; \quad (1)$$

where P is precipitation, \bar{h}_t is the tropical (20°S – 20°N) oceanic average, and T_s is the sea surface temperature. T_p is defined for each month using the model generated precipitation for that month. We also define an SST weighted by the climatological mean (seasonally varying) precipitation, T_{pc} .

Figure 5 shows the timeseries of average SST difference, precipitation-weighted SST difference and the tropospheric temperature difference scaled to allow more direct comparison with the SST differences. The latter quantity is defined by getting the tropical average tropospheric temperature difference from 400 hPa to 200 hPa with the moist adiabatic model (see Appendix B) to give an equivalent surface temperature. By doing so, the amplitude of the upper tropospheric temperature variability is approximately scaled to that of the SST and precipitation weighted SST, such that the amplitudes of the three timeseries can be compared directly. The figure shows that the upper tropospheric temperature difference between the two ensembles evolves similarly as the difference in tropical average SSTs (with a correlation of 0.74). Both quantities are smaller from 1984 until the late 1990's, and are larger from the late 1990's onwards. However, the magnitude of this change in the SST difference is smaller than in the scaled atmospheric temperature difference. Conversely, the difference in precipitation weighted SSTs not only captures the magnitude of this transition better, but also captures more interannual variability such as the evolution of the difference during the 1997/1998 El-Niño. Correspondingly, the correlation between the upper tropospheric temperature difference between the two ensembles and the precipitation-weighted SSTs is substantially larger (0.93).

Returning to Figure 2, the green solid lines in Figures 2(b,d) show the ratio of the difference in atmospheric temperature trend and the difference in the precipitation-weighted SST trend (T_p). The much better agreement with the moist adiabatic scaling (green dotted) indicates that the temperature trend profile differences can be explained by moist adiabatic scaling from the precipitation-weighted SST trend difference.

Figure 6 shows a map of the trend differences of annual mean SSTs (Hurrell minus HadISST1) for the period 1984-2008. For the period 1981-2008 the pattern is very similar, but the amplitudes are smaller as expected from the smaller difference in tropical average SST trends (see Table 1). Also shown is the climatological mean 300 K temperature isoline, and the region of the warmest quartile (thick black contour) based on HadISST1 data (contours are very similar for Hurrell).

Figure 7 shows the trend in each percentile of the two SST data sets. The two figures show that the two SSTs have trends that differ most in the warmest regions. This explains the very large atmospheric trend differences between the HadISST1 and Hurrell ensembles, since the warmest regions are also the regions of deep convection.

The differences between the SSTs weighted by the climatological mean seasonal cycle of precipitation (T_{pc} , triangles in Figure 7) are larger than the differences in area average SSTs (circles), but are still smaller than when fully considering the temporal co-variation in SSTs and rainfall (T_p , diamonds). Hence, Figure 6 provides a qualitative indication where trend differences between the two SSTs matter most for atmospheric temperature trends, but for the quantitatively correct estimate the exact relation between SSTs and precipitation distribution is important.

The same mechanism also explains the trend differences in the two CAM4 model calculations. For the period 1984-2008, the two CAM4 model calculations have a trend difference in atmospheric temperatures of 0.10K/decade at 300 hPa (Section 3.4), and a difference in precipitation weighted SSTs of 0.041K/decade (the corresponding value for the HiRAM ensembles is 0.051K/decade; i.e. the difference between 0.116K/decade and 0.065K/decade, see Table 1). Hence, just as for the HiRAM model, the amplification ratio around 300 hPa (0.10K/decade / 0.041K/decade) of 1.2:4 agrees well with the expectations based on the moist adiabatic scaling.

5. Conclusions

Rather subtle differences in SST data have large implications for atmospheric temperature trends in models using those SSTs as boundary conditions. Compared to expectations based on moist adiabatic temperature scaling of tropical average trend differences between HadISST1 and Hurrell SSTs, we find that the tropical tropospheric trend difference in AMIP model (HiRAM and CAM4) calculations using these SSTs is about a factor 3 too large. However, we show that the model atmospheric temperature trend difference can be explained by SSTs weighted with the model rainfall, which reveals that trends in the two SSTs differ substantially more (namely by that factor 3) in the important regions of deep convection than in the tropical average. With current SST uncertainties, one cannot conclude that atmospheric general circulation models have systematic biases in the tropical temperature trend profile. Due to the non-random nature of the SST differences (and likely also atmospheric temperature data differences), the level of agreement between model and observation depends on SST data set, temperature data set, and period used to

calculate trends. We conclude that resolving the discrepancies between SSTs is imperative to understand trends in tropical climate in recent decades.

Appendix A: Statistical uncertainty

For each model configuration, we have three ensemble members. This is a small sample size and as such it is difficult to estimate the ensemble spread which is needed to construct confidence intervals. In total, we have three ensembles of three runs (Hurrell, HadISST1, HadISST1 without greenhouse gases and aerosols) that all use the same atmospheric model. We therefore make the assumption that the spread about the true mean in each ensemble follows the same normal distribution. This is a reasonable assumption because ensemble spread is mainly a function of the "weather", which is expected to be similar in each ensemble. We can then estimate the standard deviation of the difference from the ensemble mean using all three ensembles. There are only 6 degrees of freedom because each ensemble mean must be removed. In addition, we subtract 0.5 from the number of degrees of freedom used in the standard error formula to remove bias [Brugger, 1969].

Therefore, our estimate for the ensemble spread is

$$\hat{\sigma}_x = \sqrt{\frac{1}{6 - 0.5} \sum_{i=1}^3 \sum_{j=1}^3 (x_{i;j} - \bar{x}_i)^2}; \quad (\text{A1})$$

where x is the quantity for which the ensemble spread is being estimated, i is the ensemble number and j is the run number inside each ensemble. The \bar{x}_i is the ensemble mean for each ensemble i . This estimator performs better than if we had computed the ensemble spread from a single ensemble. Each ensemble member is independent, so the standard error of the ensemble mean $\hat{\sigma}_x$, is given by $\hat{\sigma}_x = \hat{\sigma}_x / \sqrt{3}$. When computing the difference in quantities between the ensembles (e.g. in Figure 2ii), we estimate the standard error

by assuming both ensemble means are independent and drawn from a normal distribution with standard deviation σ_x . This gives a standard error for the difference of $\sigma_{\text{diff};x} = \frac{P}{2\sigma_x}$. Once the standard error of a quantity is estimated, we assume that the variation is Gaussian and use the standard two-sided 95% confidence interval at $\pm 1.96\sigma$.

The trends presented in this paper are computed using ordinary least squares linear regression. Similar results are obtained when we use the Theil-Sen slope estimator [Sen, 1968] which is less sensitive to outliers than ordinary least squares [Lanzante, 1996]. The Theil-Sen estimator is useful because some of the largest differences between the ensembles and the SST data sets are at the start and end of the period.

Appendix B: Moist Adiabatic Model

The moist adiabatic can be derived from near surface temperature and relative humidity [Stone and Carlson, 1979], along with assumptions on how the phase of water varies (ice, liquid water or a mixture). We use the approximations given in the ECMWF IFS documentation (Cycle 40) for the latent heats and vapor pressure, defined as

$$e_{\text{sat}} = a_1 \exp \left[a_3 \frac{T - T_0}{T - a_4} \right]; \quad (\text{B1})$$

where $T_0 = 273.16 \text{ K}$, and the parameters a_i set according to Buck [1981] in the case of liquid water ($a_1 = 611.21 \text{ Pa}$, $a_3 = 17.502$ and $a_4 = 32.19 \text{ K}$), and according to Alduchov and Eskridge [1996] in the case of ice ($a_1 = 611.21 \text{ Pa}$, $a_3 = 22.587$ and $a_4 = -40.7 \text{ K}$).

The IFS documentation assumes that the fraction of liquid water α changes according to

$$\begin{aligned} \alpha &= 0 & T &\geq T_{\text{ice}}; \\ \alpha &= \frac{T - T_{\text{ice}}}{T_0 - T_{\text{ice}}} & T_{\text{ice}} < T < T_0; \\ \alpha &= 1 & T > T_0; \end{aligned} \quad (\text{B2})$$

where $T_{\text{ice}} = 250.16 \text{ K}$.

Near the surface, the relative humidity is typically 80% in the tropics. Hence, the moist adiabatic profiles shown are initialized at the surface with a relative humidity of 80%, and constrained to match the model temperature at the 925 hPa level, as an approximation to boundary layer conditions. The initial relative humidity of 80% is held fixed, and does not have a strong effect on the results as it is approximately constant throughout the period.

Acknowledgments. We thank 3 anonymous reviewers for helpful suggestions. This work was supported by DOE grant SC0006841. We thank John Lanzante and Tom Knutson for helpful comments on an earlier version of this manuscript. We thank Qiang Fu for fruitful discussions and sharing the UW MSU analysis. We thank RSS and UAH for providing MSU data, retrieved from <http://www.remss.com> and <http://www.ncdc.noaa.gov/temp-and-precip/msu/>, respectively. We thank the Hadley Centre for the HadISST1 data, retrieved from <http://www.metoffice.gov.uk/hadobs/hadisst/data/download.html>. We thank NCAR for the Hurrell SSTs, retrieved from <https://climatedataguide.ucar.edu/climate-data/merged-hadley-noaaoi-sea-surface-temperature-sea-ice-concentration-hurrell-et-al-2008>. All data will be made available upon request (contact: Stephan Fueglistaler, stf@princeton.edu).

References

- Alduchov, O., and R. Eskridge (1996), Improved Magnus form approximation of saturation vapor pressure, *Journal of Applied Meteorology*, 35(4), 601{609.
- Brugger, R. M. (1969), A Note on Unbiased Estimation of the Standard Deviation, *The American Statistician*, 23(4), 32{32, doi:10.1080/00031305.1969.10481865.

Buck, A. (1981), New equations for computing vapor pressure and enhancement factor, *Journal of Applied Meteorology*, 20(12), 1527{1532.

Christy, J. R., R. W. Spencer, W. B. Norris, W. D. Braswell, and D. E. Parker (2003), Error Estimates of Version 5.0 of MSUAMSU Bulk Atmospheric Temperatures, *Journal of Atmospheric and Oceanic Technology*, 20(5), 613{629, doi:10.1175/1520-0426(2003)20<613:EEOVOM>2.0.CO;2.

Christy, J. R., W. B. Norris, R. W. Spencer, and J. J. Hnilo (2007), Tropospheric temperature change since 1979 from tropical radiosonde and satellite measurements, *Journal of Geophysical Research*, 112(D6), D06,102, doi:10.1029/2005JD006881.

Christy, J. R., B. Herman, R. Pielke, P. Klotzbach, R. T. McNider, J. J. Hnilo, R. W. Spencer, T. Chase, and D. Douglass (2010), What Do Observational Datasets Say about Modeled Tropospheric Temperature Trends since 1979?, *Remote Sensing*, 2(9), 2148{2169, doi:10.3390/rs2092148.

Copsey, D., R. Sutton, and J. R. Knight (2006), Recent trends in sea level pressure in the Indian Ocean region, *Geophysical Research Letters*, 33(19), L19,712, doi:10.1029/2006GL027175.

Delworth, T.L., A.J. Broccoli, A. Rosati, R.J. Stouffer, V. Balaji, J.A. Beesley, W.F. Cooke, K.W. Dixon, J.P. Dunne, K.A. Dunne, J.W. Durachta, K.L. Findell, P. Ginoux, A. Gnanadesikan, C.T. Gordon, S.M. Griffes, R. Gudgel, M.J. Harrison, I.M. Held, R.S. Hemler, L.W. Horowitz, S.A. Klein, T.R. Knutson, P.J. Kushner, A.R. Langenhorst, H.-C. Lee, S.-J. Lin, J. Lu, S. Malyshev, P.C.D. Milly, V. Ramaswamy, J.L. Russell, M.D. Schwarzkopf, E. Shevliakova, J.J. Sirutis, M.J. Spelman, W.F. Stern, M. Winton, A.T. Wittenberg, B. Wyman, F. Zeng, and R. Zhang (2006), GFDL's CM2 Global

Coupled Climate Models. Part I: Formulation and Simulation Characteristics, *Journal of Climate*, 19(5), DOI:10.1175/JCLI3629.1.

Douville, H. (2005), Limitations of time-slice experiments for predicting regional climate change over South Asia, *Climate Dynamics*, 24(4), 373{391, doi:10.1007/s00382-004-0509-7.

Emanuel, K., and A. Sobel (2013), Response of tropical sea surface temperature, precipitation, and tropical cyclone-related variables to changes in global and local forcing, *Journal of Advances in Modeling Earth Systems*, 5 (2), 447{458, doi:10.1002/jame.20032.

Fu, Q., C. M. Johanson, S. G. Warren, and D. J. Seidel (2004), Contribution of stratospheric cooling to satellite-inferred tropospheric temperature trends., *Nature*, 429(6987), 55{8, doi:10.1038/nature02524.

Fu, Q., S. Manabe, and C. M. Johanson (2011), On the warming in the tropical upper troposphere: Models versus observations, *Geophysical Research Letters*, 38(15), n/a{n/a, doi:10.1029/2011GL048101.

Gates, W. L. (1992), AMIP: The Atmospheric Model Intercomparison Project, *Bulletin of the American Meteorological Society*, 73(12), 1962{1970, doi:10.1175/1520-0477(1992)073<1962:ATAMIP>2.0.CO;2.

Held, I. M., and M. Zhao (2011), The Response of Tropical Cyclone Statistics to an Increase in CO₂ with Fixed Sea Surface Temperatures, *Journal of Climate*, 24(20), 5353{5364, doi:10.1175/JCLI-D-11-00050.1.

Hurrell, J., M. Visbeck, and P. Pirani (2011), WCRP Coupled Model Intercomparison Project-Phase 5-CMIP5, *CLIVAR Exchanges*, 15(56).

Hurrell, J. W., and K. E. Trenberth (1997), Spurious trends in satellite MSU temperatures from merging different satellite records, *Nature*, 386(6621), 164{167, doi: 10.1038/386164a0.

Hurrell, J. W., and K. E. Trenberth (1999), Global Sea Surface Temperature Analyses: Multiple Problems and Their Implications for Climate Analysis, Modeling, and Reanalysis, *Bulletin of the American Meteorological Society*, 80(12), 2661{2678, doi: 10.1175/1520-0477(1999)080<2661:GSSTAM>2.0.CO;2.

Hurrell, J. W., J. J. Hack, D. Shea, J. M. Caron, and J. Rosinski (2008), A New Sea Surface Temperature and Sea Ice Boundary Dataset for the Community Atmosphere Model, *Journal of Climate*, 21(19), 5145{5153, doi:10.1175/2008JCLI2292.1.

Kaplan, A., Y. Kushnir, M. A. Cane, and M. B. Blumenthal (1997), Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures, *Journal of Geophysical Research*, 102(C13), 27,835, doi:10.1029/97JC01734.

Lanzante, J. (1996), Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data, *International Journal of Climatology*, 16(11), 1197{1226, doi: 10.1002/(SICI)1097-0088(199611)16:11<1197::AID-JOC89>3.0.CO;2-L.

Mears, C. A., and F. J. Wentz (2009a), Construction of the Remote Sensing Systems V3.2 Atmospheric Temperature Records from the MSU and AMSU Microwave Sounders, *Journal of Atmospheric and Oceanic Technology*, 26(6), 1040{1056, doi: 10.1175/2008JTECHA1176.1.

Mears, C. A., and F. J. Wentz (2009b), Construction of the RSS V3.2 Lower-Tropospheric Temperature Dataset from the MSU and AMSU Microwave Sounders, *Journal of Atmo-*

spheric and Oceanic Technology, 26(8), 1493-1509, doi:10.1175/2009JTECHA1237.1.

Mitchell, D.M., P.W. Thorne, P.A. Stott, and L.J. Gray (2013), Revisiting the controversial issue of tropical tropospheric temperature trends, *Geophys. Res. Letts.*, 40, 1-6, doi:10.1002/grl.50465.

Neale, R. B., J. H. Richter, A. J. Conley, S. Park, P. H. Lauritzen, A. Gettelman, D. L. Williamson, P. J. Rasch, S. J. Vavrus, M. A. Taylor, W. D. Collins, M. Zhang, and S. Lin (2010), Description of the near community atmosphere model (cam 4.0), NCAR technical note, NCAR/TN-485+STR.

Po-Chedley, S., and Q. Fu (2012a), Discrepancies in tropical upper tropospheric warming between atmospheric circulation models and satellites, *Environmental Research Letters*, 7(4), 044018, doi:10.1088/1748-9326/7/4/044018.

Po-Chedley, S and Q. Fu (2012b), A Bias in the Midtropospheric Channel Warm Target Factor on the NOAA-9 Microwave Sounding Unit, *Journal of Atmospheric and Oceanic Technology*, 29, 646-656, doi: 10.1175/JTECH-D-11-00147.1.

Po-Chedley, S and Q. Fu (2013), Reply to Comments on 'A Bias in the Midtropospheric Channel Warm Target Factor on the NOAA-9 Microwave Sounding Unit, *Journal of Atmospheric and Oceanic Technology*, 30, 1014-1020, doi: 10.1175/JTECH-D-12-00131.1.

Rayner, N. A. (2003), Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *Journal of Geophysical Research*, 108(D14), 4407, doi:10.1029/2002JD002670.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang (2002), An Improved In Situ and Satellite SST Analysis for Climate, *Journal of Climate*, 15(13), 1609-1625, doi:10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2.

- Santer, B. D., T. M. L. Wigley, C. Mears, F. J. Wentz, S. a. Klein, D. J. Seidel, K. E. Taylor, P. W. Thorne, M. F. Wehner, P. J. Gleckler, J. S. Boyle, W. D. Collins, K. W. Dixon, C. Doutriaux, M. Free, Q. Fu, J. E. Hansen, G. S. Jones, R. Ruedy, T. R. Karl, J. R. Lanzante, G. a. Meehl, V. Ramaswamy, G. Russell, and G. a. Schmidt (2005), Amplification of surface temperature trends and variability in the tropical atmosphere., *Science* (New York, N.Y.), 309(5740), 1551-1556, doi:10.1126/science.1114867.
- Sen, P. (1968), Estimates of the regression coefficient based on Kendall's tau, *Journal of the American Statistical Association*, 63(324), 1379-1389.
- Singh, M.S., P.A. O'Gorman (2012), Upward Shift of the Atmospheric General Circulation under Global Warming: Theory and Simulations, *J. Clim.*, 25, 8259-8276, doi:10.1175/JCLI-D-11-00699.1
- Sobel, A. H., I. M. Held, and C. S. Bretherton (2002), The ENSO Signal in Tropical Tropospheric Temperature, *Journal of Climate*, 15(18), 2702-2706, doi:10.1175/1520-0442(2002)015<2702:TESITT>2.0.CO;2.
- Stone, P., and J. Carlson (1979), Atmospheric lapse rate regimes and their parameterization, *Journal of the Atmospheric Sciences*, 36, 415-423.
- Thorne, P. W., J. R. Lanzante, T. C. Peterson, D. J. Seidel, and K. P. Shine (2011), Tropospheric temperature trends: history of an ongoing controversy, *Wiley Interdisciplinary Reviews: Climate Change*, 2(1), 66-88, doi:10.1002/wcc.80.
- Woodruff, S. D., R. J. Slutz, R. L. Jenne, and P. M. Steurer (1987), A Comprehensive Ocean-Atmosphere Data Set, *Bulletin of the American Meteorological Society*, 68(10), 1239-1250, doi:10.1175/1520-0477(1987)068<1239:ACOADS>2.0.CO;2.

Accepted Article
Zhao, M., and I. M. Held (2012), TC-Permitting GCM Simulations of Hurricane Frequency Response to Sea Surface Temperature Anomalies Projected for the Late-Twenty-First Century, *Journal of Climate*, 25 (8), 2995-3009, doi:10.1175/JCLI-D-11-00313.1.

Zhao, M., I. M. Held, S.-J. Lin, and G. a. Vecchi (2009), Simulations of Global Hurricane Climatology, Interannual Variability, and Response to Global Warming Using a 50-km Resolution GCM, *Journal of Climate*, 22 (24), 6653-6678, doi:10.1175/2009JCLI3049.1.

Zhao, M., I. M. Held, and S.-J. Lin (2012), Some Counterintuitive Dependencies of Tropical Cyclone Frequency on Parameters in a GCM, *Journal of the Atmospheric Sciences*, 69 (7), 2272-2283, doi:10.1175/JAS-D-11-0238.1.

Zhao, M., I. M. Held, and G. A. Vecchi (2013), Robust direct effect of increasing atmospheric CO₂ concentration on global tropical cyclone frequency: a multi-model inter-comparison, *U.S. CLIVAR Variations*, 11 (3), 17-23.

Table 1. Trends (in K/decade) in tropical average (20°S – 20°N) MSU channels over both time periods considered in this study for the HadISST1 ensemble, the Hurrell ensemble, and the RSS, UAH and UW MSU data sets. The UW TTT channel data was constructed using the RSS TLS data. The SST trend and precipitation-weighted SST trend over each period is also shown for the two ensembles.

	Channel	HadISST1	Hurrell	RSS	UAH	UW
1981–2008	TLT	0.139	0.170	0.147	0.091	{
	TMT	0.121	0.159	0.114	0.042	0.141
	TLS	-0.246	-0.239	-0.346	-0.418	{
	TTT	0.158	0.199	0.160	0.088	0.187
	SST	0.062	0.073	{	{	{
	P-weighted SST	0.065	0.099	{	{	{
1984–2008	TLT	0.135	0.191	0.179	0.130	{
	TMT	0.119	0.187	0.159	0.090	0.183
	TLS	-0.201	-0.187	-0.257	-0.332	{
	TTT	0.151	0.224	0.200	0.132	0.223
	SST	0.073	0.092	{	{	{
	P-weighted SST	0.065	0.116	{	{	{

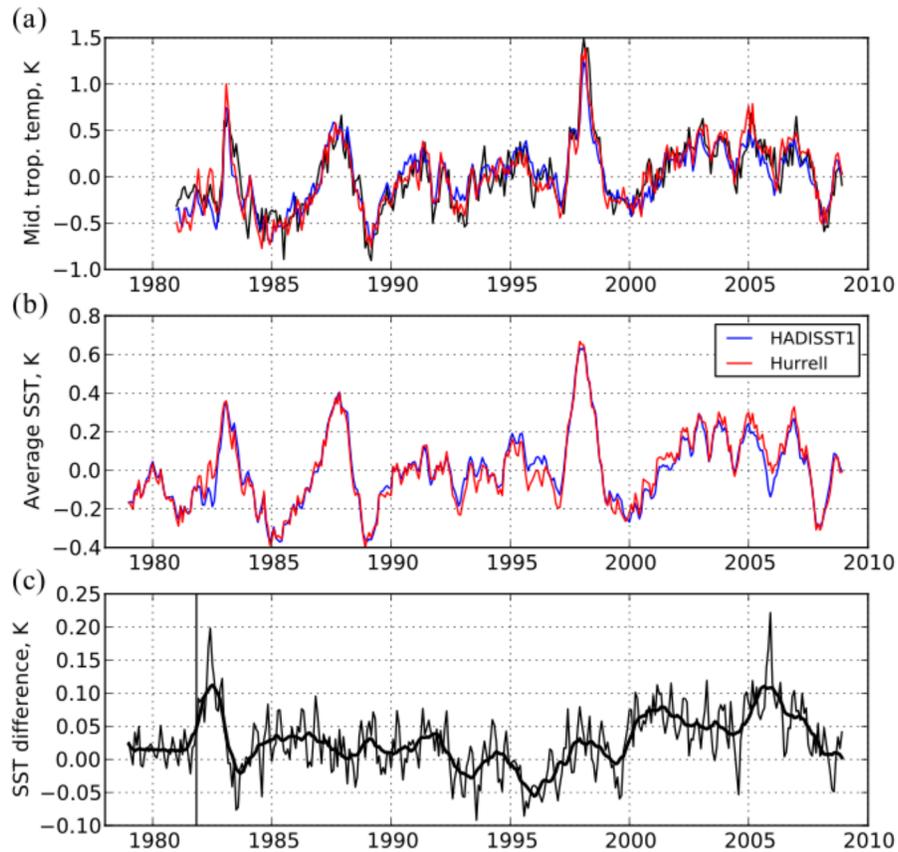


Figure 1. (a) Deseasonalized monthly mean mid-tropospheric temperatures (TTT channel; see text) from the HadISST1 ensemble (blue), the Hurrell ensemble (red) and the RSS MSU data set (black) $20^{\circ}\text{S}\{20^{\circ}\text{N}$ average. (b) Deseasonalized monthly mean HadISST1 and Hurrell SST $20^{\circ}\text{S}\{20^{\circ}\text{N}$ average from 1979 to 2008. (c) The monthly mean difference Hurrell minus HadISST1 (thin), and the difference with a 12 month running mean applied (thick). The vertical marker indicates November 1981, when the AVHRR satellite observations begin.

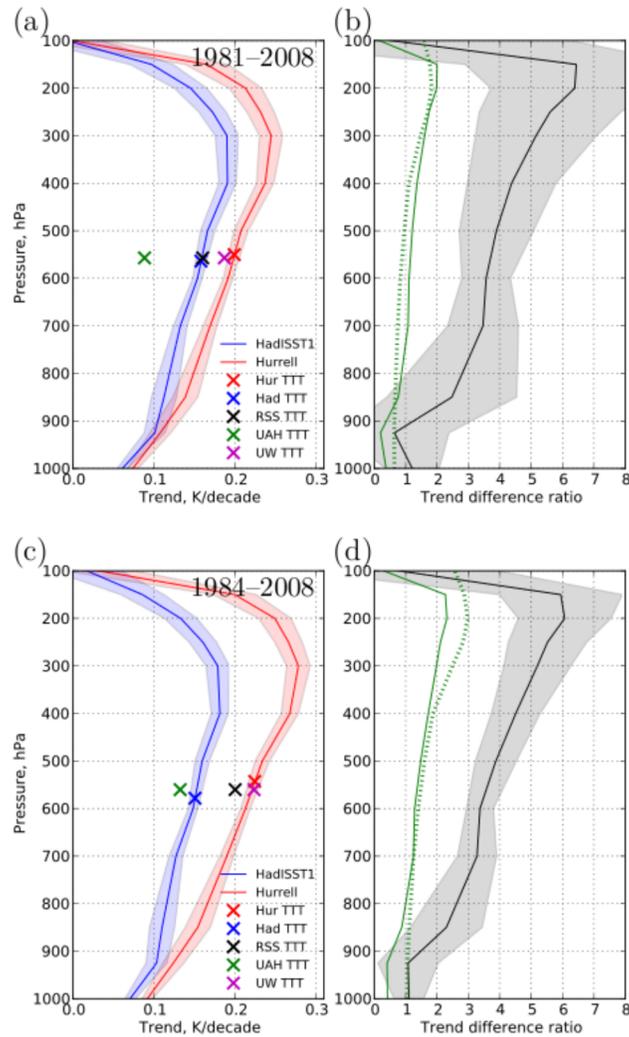


Figure 2. Atmospheric trends for the two time periods 1981{2009 (a,b) and 1984{2009 (c,d). Panels a and c show profiles of temperature trend for each ensemble (solid lines). The symbols show the trend for different MSU channels (black is RSS data, green is UAH data, magenta is UW data, blue and red are the equivalent model quantities). Panels b and d show profiles of the ratio of the difference in atmospheric trend to the difference in the SST trend between the two ensembles (black), and the ratio using precipitation-weighted SST in place of average SST (green solid). The trend difference ratio consistent with a moist adiabat is also shown (green dotted). The shading shows the 95% confidence interval derived from the ensemble spread (see Appendix A).

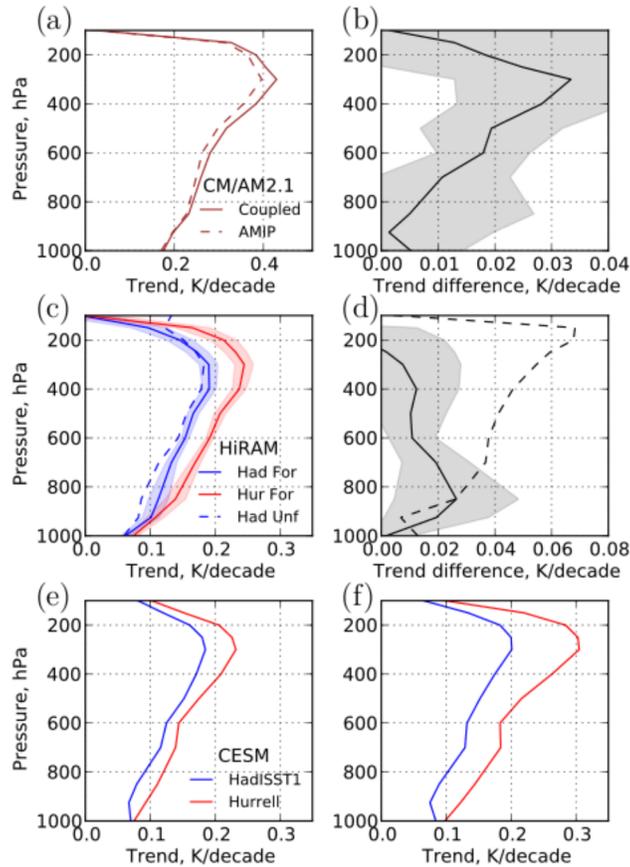


Figure 3. Model sensitivities. (a) Profiles of the linear trends in atmospheric temperatures in a coupled run of CM2.1 (solid), and a corresponding AMIP run (dashed) using the same atmospheric and land models but with prescribed SSTs identical to those in the CM2.1 run. (b) Profile of the trend difference between the coupled and AMIP run shown in panel (a). The shading shows the 95% confidence interval. (c) Temperature trend profiles (for the period 1981–2008) in the HadISST1 ('Had For'; blue, solid) and Hurrell ('Hur For'; red, solid) HiRAM model runs (data as in Figure 2), and an ensemble using HadISST1 but no atmospheric forcings ('Had Unf'). The shadings show the 95% confidence intervals. (d) Differences between trends shown in panel (c). Solid line: HadISST1 minus HadISST1 unforced; dashed line: Hurrell minus HadISST1 model run. The shading shows the 95% confidence interval. (e) Temperature trend profiles (for the period 1981–2008, single model runs) using CESM/ CAM4 with HadISST1 and Hurrell SSTs, both without atmospheric forcings. (f) As panel (e), but for the period 1984–2008.

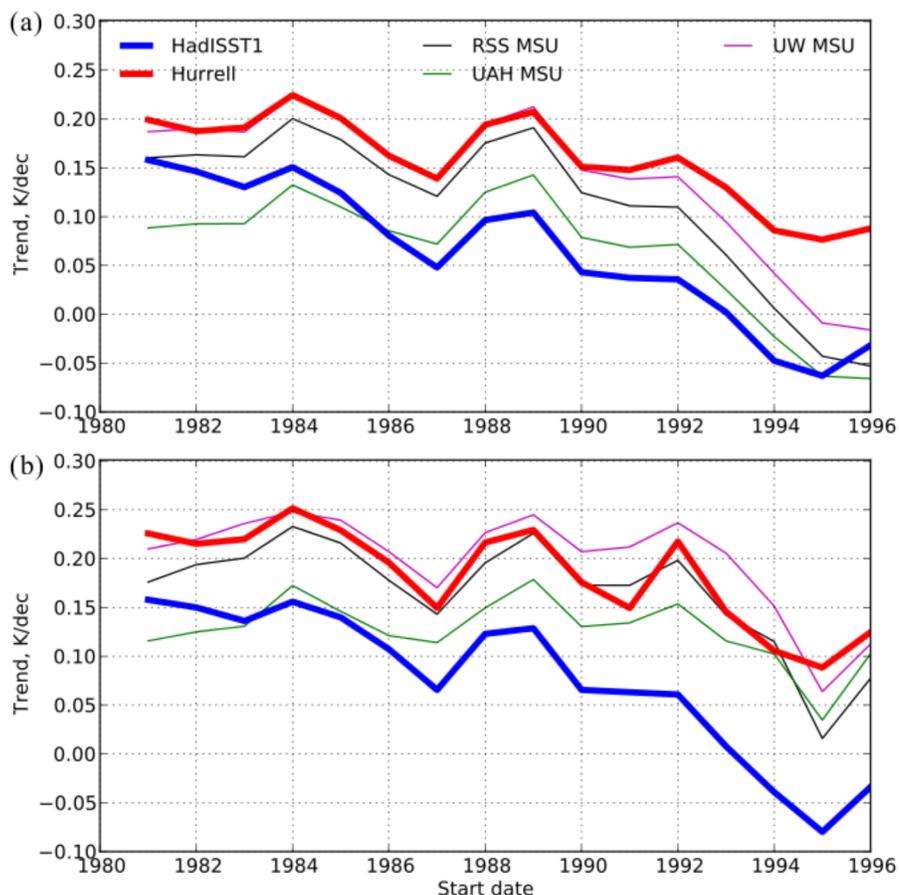


Figure 4. Mid-tropospheric temperature trends (TTT channel) in the HadISST1 ensemble (blue), Hurrell ensemble (red), and the MSU TTT data from RSS (black), UAH (green) and UW (magenta), computed for different start dates and shown as function of start date (abscissa). The end date is 2008 in all cases. (a) The linear trend computed using ordinary least squares regression. (b) Theil-Sen estimator of the slope.

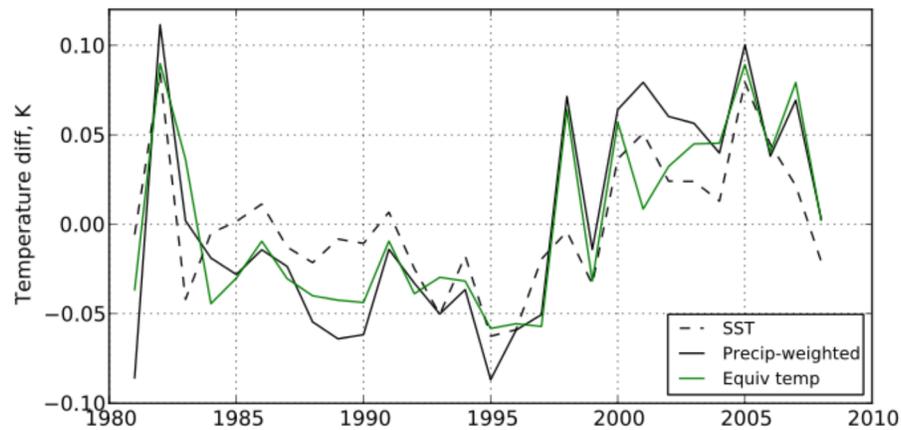


Figure 5. Timeseries of the difference between the Hurrell and HadISST 1 (Hurrell & HadISST 1) ensembles in tropical average SST difference (black dashed), precipitation-weighted SST difference (black solid), and a scaled version (see text) of the atmospheric temperature difference between the two model ensembles. All quantities have their time mean removed, and annual means have been taken.

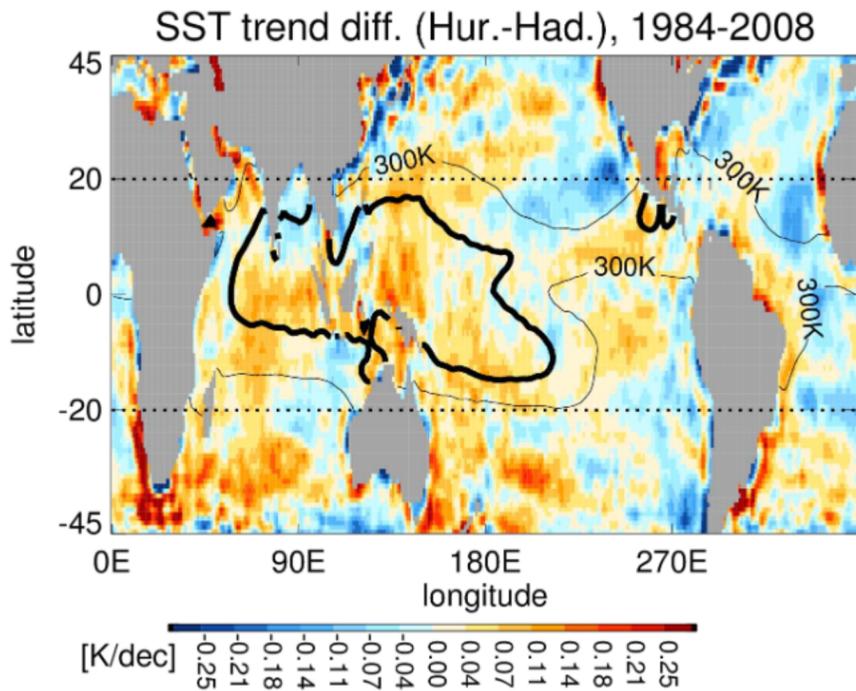


Figure 6. SST trend difference (Hurrell minus HadISST1) for the period 1984-2008 (pattern very similar for the period 1981-2008) based on annual mean data. Also shown are the climatological mean 300 K temperature isoline (thin black contour), and the warmest quartile (thick black contour), based on HadISST1 data (contours nearly identical for Hurrell data). The black dotted lines show the $20^{\circ}\text{S}\{20^{\circ}\text{N}$ latitude belt.

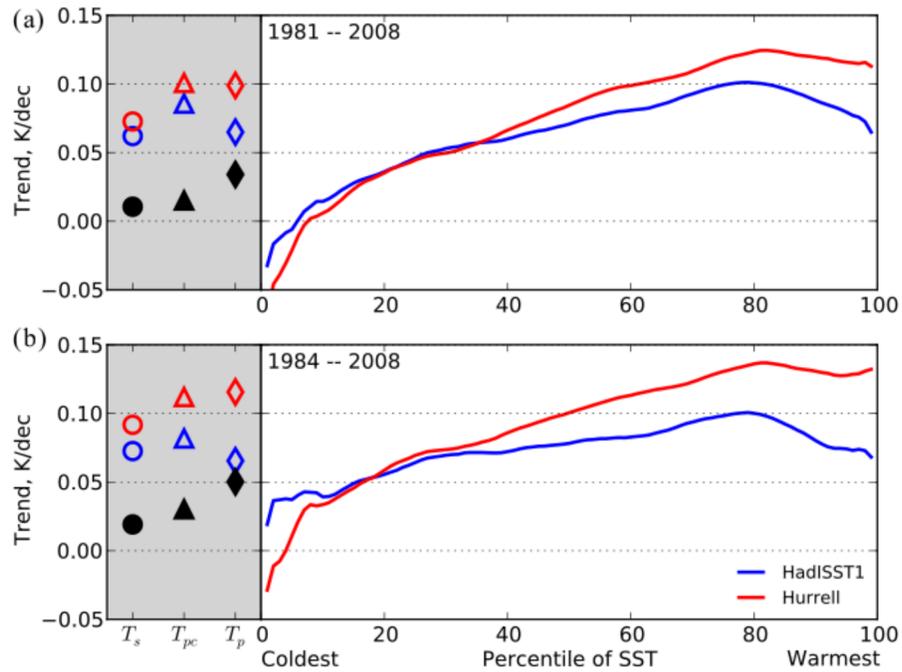


Figure 7. Trends of the percentiles of tropical SST in the HadISST1 (blue) and Hurrell (red) data sets over (a) 1981-2008 and (b) 1984-2008. The symbols show the trend in average SST T_s (circles), SST weighted with the climatological precipitation T_{pc} (triangles), and the precipitation weighted SST T_p (diamonds). The black filled symbols show the difference between the Hurrell based and HadISST1 based calculations.

