INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. (2007)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/joc.1651

A comparison of tropical temperature trends with model predictions

David H. Douglass,[a]* John R. Christy,[b] Benjamin D. Pearson[a†] and S. Fred Singer[c,d]

[a] *Department of Physics and Astronomy, University of Rochester, Rochester, NY 14627, USA*
[b] *Department of Atmospheric Science and Earth System Science Center, University of Alabama in Huntsville, Huntsville, AL 35899, USA*
[c] *Science and Environmental Policy Project, Arlington, VA 22202, USA*
[d] *University of Virginia, Charlottesville, VA 22903, USA*

ABSTRACT: We examine tropospheric temperature trends of 67 runs from 22 'Climate of the 20th Century' model simulations and try to reconcile them with the best available updated observations (in the tropics during the satellite era). Model results and observed temperature trends are in disagreement in most of the tropical troposphere, being separated by more than twice the uncertainty of the model mean. In layers near 5 km, the modelled trend is 100 to 300% higher than observed, and, above 8 km, modelled and observed trends have opposite signs. These conclusions contrast strongly with those of recent publications based on essentially the same data. Copyright © 2007 Royal Meteorological Society

KEY WORDS climate trend; troposphere; observations

Received 31 May 2007; Accepted 11 October 2007

1. Introduction

A panel convened by the National Research Council (2000) found for the satellite era (since 1979) '[a]pparently conflicting surface and tropospheric temperature trends' that could not be reconciled, with the Earth's surface warming faster than the lower troposphere. The panel concluded, after considering possible systematic errors that '[a] substantial disparity remains.' From a study of several independent observational datasets Douglass *et al.* (2004b) confirmed that the disparity was real and arose mostly in the tropical zone. Also, Douglass *et al.* (2004a) showed that three state-of-the-art General Circulation Models (GCMs) predicted a temperature trend that increased with altitude, reaching a maximum ratio to the surface trend ('amplification' factor R) as much as 1.5–2.0 at a pressure (altitude) about 200–400 hPa. This was in disagreement with observations, which showed flat or decreasing amplification factors with altitude.

In the Douglass *et al.* (2004a) study, only three observational datasets were considered, and the number of models was limited to the three most widely referenced. The present study includes all available datasets, and an Intergovernmental Panel on Climate Change (IPCC)-sponsored model inter-comparison project using the 'Climate of the 20th Century' (20CEN) forcing includes models from almost all the major modelling groups [Program for Climate Model Diagnosis and Intercomparison (PCMDI, 2005)]. The number of observational datasets and models constitutes a significant increase over the Douglass *et al.* (2004a) study, and thus, it is appropriate that a new analysis be made.

Santer *et al.* (2005) recently investigated the altitude dependence of temperature trends during the satellite era, emphasizing the tropical zone, where the characteristics are well-suited for model evaluation. They compared available observations with 19 of the models and suggest that any disparity between models and observations is due to residual errors in the observational datasets. In this article, we consider the observational results in 22 of the models that were available. As did Santer *et al.* (2005), we confine our study to the tropical zone – but we reach a different conclusion.

In Section 2 we describe the data and the models. In Section 3 we show that the models and the observations disagree to a statistically significant extent. In Section 4 we discuss efforts that have been made to resolve the disparity, and we summarize in Section 5.

2. Source material and definitions

Much of the Earth's global mean temperature variability originates in the tropics, which is also the place where the disparity between model results and observations is most apparent. For the models and most of the data, we define 'tropics' operationally as the region between 20°S and 20°N. In respect

* Correspondence to: David H. Douglass, Department of Physics and Astronomy, University of Rochester, Rochester, NY 14627, USA. E-mail: douglass@pas.rochester.edu
† Current address: VWR International, 5100 W. Henrietta Rd., Rochester, NY 14692, USA.

of the Integrated Global Radiosonde Archive (IGRA) and Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC) datasets (see below), the range is 30 °S–30 °N.

The influence of the major El Niño of 1997–1998 needs clarification. The models, free to produce El Niños at differing times and magnitudes, therefore, yield associated individual trends not directly comparable with observations. This results in a mismatch of model *versus* observed El Niño occurrences. Averaging over a number of simulations for a model is one way to minimize the influence of major El Niño events near the end of the record. In Douglass *et al.* (2004a) the data were truncated at 1996 to avoid the 1997–1998 event. Now the data extend to 2004, and thus, the impact of this significant El Niño is minimized. Hence, the influence of El Niños is effectively removed in both models and observations.

2.1. Observational datasets

We consider ten sets that measure temperature anomalies at various altitudes in the troposphere from the surface to the tropopause.

2.1.1. *Surface*

Three datasets were used: IPCC/HadCRUT2v (Jones and Moberg, 2003), Global Historical Climatology Network (GHCN, 2005), and the NASA Goddard Institute for Space Studies (GISS, 2005).

2.1.2. *Radiosondes*

Coleman and Thorne (2005) provide a new analysis (HadAT2) of the Hadley radiosonde dataset. The trend values as a function of pressure (altitude) for the tropical zone (20 °S–20 °N) are listed in Table I. Free *et al.* (2005) have a new dataset based upon the 87-station set of Lanzante *et al.* (2003). They provide an updated analysis for the tropical zone (30 °S–30 °N) and different analyses of data from the IGRA, and the A and B versions of (RATPAC). We have chosen version B because it is consistent with IGRA. Haimberger (2006) uses a new technique for analysing the European Centre for Medium-Range Weather Forecasts dataset (RAOBCORE), which adjusts for changes in instrumentation for the 1184 radiosonde records. We use version 1.2 of his data for the tropics (20 °S–20 °N).

2.1.3. *Satellites*

The University of Alabama in Huntsville (UAH, 2007) (Christy and Norris, 2006) and Remote Sensing Systems (RSS) (Mears and Wentz, 2005) provide two independent analyses of the same Microwave Sounding Unit (MSU) data. These are the only two groups that produce reconstructed temperature anomalies for three different effective layers in the lower troposphere and lower stratosphere. T_{2LT} represents the lower troposphere and is a weighted mean, with the largest weights from the surface to 350 hPa (mean altitude 2.5 km), T_2 corresponds to the mid-troposphere with weights extending from the surface to 70 hPa (mean altitude 6.1 km) and T_4 corresponds to the lower stratosphere with weights extending from 120 to 20 hPa (mean altitude of 17 km). Here we will be considering only T_{2LT} and T_2. We include, also, the T_2 results of Vinnikov *et al.* (2006) (UMD), though this is their only product and is difficult to assess without gridded data.

2.1.4. *Simple statistical retrievals (SSRs)*

Johanson and Fu (2006) (JF) demonstrated that a statistical combination of MSU T_2 and T_4 produces a representation, T_{SSR}, of the deep troposphere (1000–100 hPa, mean pressure 550 hPa). The form of the equation is $(1 + a) \times (T_2) - a \times (T_4) = T_{SSR}$, where $a = 0.1$. In JF, T_{SSR} has shortcomings due to its reliance on the statistics of specific sets of radiosondes and specific time periods from which the value of '*a*' is determined (Christy and Norris, 2006). We do not use T_{SSR} in this study as the model stratospheric trends needed for T_{SSR} are difficult to accept as realistic. (Christy and Norris provided

Table I. Observed tropical temperature trends (milli °C/decade). 1979–2004. See note at bottom for explanation of the entries in the top row.

Altitude (pressure)*		T_{2LT}	T_2	'999'	850	700	500	400	300	250	200	150	100
Dataset	zone												
HadCRUT2v	20 °S–20 °N			124									
GHCN	24 °S–24 °N			129									
GISS	20 °S–20 °N			126									
UAH5.2	20 °S–20 °N	53	48										
RSS	20 °S–20 °N	151	133										
UMD	20 °S–20 °N		210										
HadAT2	20 °S–20 °N				61	28	14		75		−32	−147	−431
RAOBCORE	20 °S–20 °N				131	−15	−73	15	84	90	78	−71	−349
IGRA	30 °S–30 °N			176	98	60	33	92	93	43	−66	−166	−488
RATPAC	30 °S–30 °N			123	82	82	68	115	125	90	−17	−117	−392

* No pressure is assigned to the MSU values (T_{2LT} and T_2) because they result from weighted averages over a range of values (see text). The entry marked '999' represents the surface, given value 999 for plotting purposes. Other entries in the first row are pressure in hPa.

DOI: 10.1002/joc

evidence that SSR satellite data were not completely self-consistent as determined by radiosonde-calculated equivalent SSRs.)

2.2. Climate models

Recently, most of the modelling groups agreed to participate in a GCM Intercomparison Project, using the 20CEN forcing (PCMDI, 2005). Each of these models ran between one and nine simulations. We averaged these runs for each model to simulate removal of the El Niño/Southern Oscillation (ENSO) effect; these models cannot reproduce the observed time sequence of El Niño and La Niña events, except by chance (Santer *et al.*, 2003). Table II provides complete numerical results from our analysis of the 22 models at the surface and 12 different altitudes.

2.2.1. *Synthetic values of trends of T_{2LT} and T_2*

The various MSU temperature products are suitably weighted averages over a range of pressure levels. To compare model results with these values a corresponding averaging scheme over the model pressure levels has been developed. Appropriate weighting factors (Spencer and Christy, 1992) are applied to the model values at the different pressure levels to calculate synthetic values of trends of T_{2LT} and T_2. Such weights were also used by Santer *et al.* (2005) and Karl *et al.* (2006). Using these weighting factors we have calculated synthetic values of trends of T_{2LT} and T_2; these are in Table III.

2.3. Definitions

A *trend* is defined as the slope of a line that has been least-squares fit to the data. The ratio of a trend to the trend at the surface is called the 'amplification factor', R.

For the models, we calculate the mean, standard deviation (σ), and estimate of the uncertainty of the mean (σ_{SE}) of the predictions of the trends at various altitude levels. We assume that σ_{SE} and standard deviation are related by $\sigma_{SE} = \sigma/\sqrt{N-1}$, where $N = 22$ is the number of independent models. A case could be made that N should be greater than 22 since there are 67 realizations. In Figure 1 we show the mean of the model predictions and its $2\sigma_{SE}$ uncertainty limits. Thus, in a repeat of the 22-model computational runs one would expect that a new mean would lie between these limits with 95% probability.

Table II. (a). Temperature trends for 22 CGCM Models with 20CEN forcing. The numbered models are fully identified in Table II(b).

		Pressure (hPa)–>												
		Surface	1000	925	850	700	600	500	400	300	250	200	150	100
Model	Sims.*	Trends (milli °C/decade)												
1	9	128	303	121	177	161	172	190	216	247	263	268	243	40
2	5	125	1507	113	112	123	126	138	148	140	105	2	−114	−161
3	5	311	318	336	346	376	422	484	596	672	673	642	594	253
4	5	95	92	99	99	131	179	158	184	212	224	182	169	−3
5	5	210	302	224	215	249	264	293	343	391	408	400	319	75
6	4	119	118	148	175	189	214	238	283	365	406	425	393	−33
7	4	112	460	107	123	122	130	155	183	213	228	225	211	0
8	3	86	62	57	58	82	95	108	134	160	163	155	137	100
9	3	142	143	148	150	149	162	200	234	273	284	282	258	163
10	3	189	114	200	210	225	238	269	316	345	348	347	308	53
11	3	244	403	270	278	309	331	377	449	503	481	461	405	75
12	3	80	173	114	115	102	98	124	150	161	164	166	142	4
13	2	162	155	170**	182	225	218	221	282	352	360	340	277	−39
14	2	171	293	190	197	252	245	268	328	376	367	326	278	69
15	2	163	213	174	181	199	204	226	271	307	299	255	166	53
16	2	119	128	124	140	151	176	197	228	271	289	306	260	120
17	2	219	−1268	199	223	259	283	321	373	427	454	479	465	280
18	1	117	117	126	148	163	159	180	207	227	225	203	200	163
19	1	230	220	267	283	313	346	410	506	561	554	526	521	244
20	1	191	151	176	194	212	237	254	304	387	410	400	367	314
21	1	191	328	241	222	193	187	215	255	300	316	327	304	90
22	1	28	24	46	73	27	−26	−26	−1	20	24	32	−1	−136
Total simulations:	67													
Average		156	198	166	177	191	203	227	272	314	320	307	268	78
Std. Dev. (σ)		64	443	72	70	82	96	109	131	148	149	154	160	124

* 'Sims.' refers to the number of simulations over which averages within a model were taken.
** The value for model 13 at $p = 925$ hPa was interpolated.

Table II. (b). Symbols and sources of the models numbered in Table II(a).

Model	Name	Institution
1	GISS_er	Goddard Institute for Space Studies, USA (GISS)
2	NCAR-CCSM3	National Center for Atmospheric Research, USA (NCAR)
3	CCCma-CGCM3.1T47	Canadian Centre for Climate Modeling and Analysis (CCCMA)
4	GISS_eh	GISS
5	MRI-CGCM2.3.2	Meteorological Research Institute, Japan
6	bcc_cm1	China Institute of Atomic Energy
7	PCM	NCAR
8	ECHAM5	Max-Planck Institute, Germany
9	FGOALS-g1.0	Institute for Atmospheric Phyics, China
10	GFDL-2.0	Geophysical Fluid Dynamic Laboratory, Princeton (GFDL)
11	GFDL-2.1	GFDL
12	MIROC3.2_Merdes	Center for Climate System Research, Japan (CCSR)
13	HADCM3	Hadley Center for Climate Prediction and Research, UK
14	HadGEM1	Hadley
15	CSIRO_MK3.0	Commonwealth Science and Industrial Research. Australia
16	GISS_aom	GISS
17	ISPL_CM4	Institute Simon-Pierre LaPlace, France
18	bccr_bcm2.0	Bjerknes Center for Climate Research, Norway
19	CCCma-CGCM3.1(T63)	CCCMA
20	CNRM-CM3	Meteo-France/centre National de Research Meteirologique
21	MIROC3.2_Hires	CCSR
22	INCM_3_0	Institute for Numerical Mathematics, Russia

Table III. Comparison of trends of T_{2LT} and T_2 (in °C/decade).

	T_{2LT}	T_2
MSU	0.053	0.048
RSS	0.151	0.133
Models*	0.214 ± 0.040	0.228 ± 0.050

* Uncertainties are $\pm 2\sigma_{SE}$.

2.3.1. *Statistical Significance*

A*greement* means that an observed value's stated uncertainty overlaps the $2\sigma_{SE}$ uncertainty of the models.

3. Analysis

3.1. Trends

From the observational data we determine trends for the period Jan 1979–Dec 2004. The datasets are those considered by Santer *et al*. (2005), with the addition of the new set RAOBCORE. We calculate trends at 13 altitude levels between the surface and the tropopause for each of the models, for the period 1979–1999, the last year considered in many of the models.

Tables I and II show the results of all trend calculations for the tropics, including averages and standard deviation (SD) values, as a function of pressure (altitude). Figure 1 shows these temperature trends as a function of pressure (altitude) from the surface (∼1000 hPa) to the tropopause (∼100 hPa/17 km).

3.1.1. *Observations*

Figure 1 shows three independent surface and four radiosonde results. One sees that all radiosonde trends are initially positive at the surface, generally decrease with pressure/altitude up to about 300 hPa/8 km, and then decrease rapidly. Thus, for the radiosonde observations, R is generally less than or equal to 1.0. The MSU values are shown in the right panel of Figure 1. There it is seen that all of the observational values are less than the synthetic values from the models. Only the value from UMD is within the uncertainty of the model calculation.

3.1.2. *Models*

Each of the 22 models had been run with between one and nine simulations of the 20CEN forcings. We averaged over these for each model for the surface and for 12 pressure levels from 1000 and 100 hPa (Table II), and then computed the mean and SD for the 22-model ensemble. In addition, the values of the synthetic T_{2LT} and T_2 were computed (see Table III).

3.1.3. *Altitude dependence*

The 22-ensemble mean shown in Figure 1 increases with altitude, reaches a maximum R-value of about 2.1 at about 250 hPa, and then decreases. Modelled trends are more than twice those seen in the radiosonde data. Beyond ∼200 hPa the observed trends are much lower and become negative. These results are in good agreement with those of Douglass *et al*. (2004a).

The surface is obviously of interest because that is where we live. However, an essential place to compare

DOI: 10.1002/joc

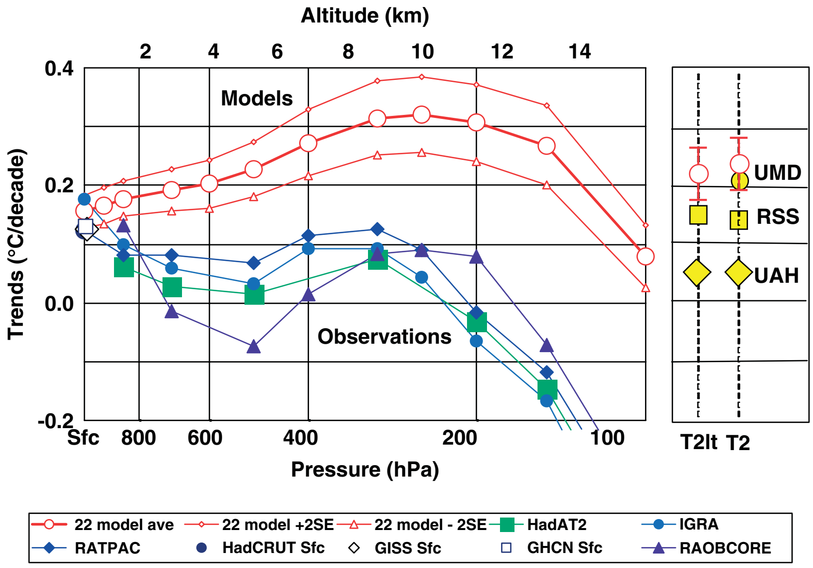

Figure 1. Temperature trends for the satellite era. Plot of temperature trend (°C/decade) against pressure (altitude) from the data in Tables I and II. The HadCRUT2v surface trend value is a large blue circle. The GHCN and the GISS surface values are the open rectangle and diamond. The four radiosonde results (IGRA, RATPAC, HadAT2, and RAOBCORE) are shown in blue, light blue, green, and purple respectively. The two UAH MSU data points are shown as gold-filled diamonds and the RSS MSU data points as gold-filled squares. The MSU UMD data point is gold circle. The 22-model ensemble average is a solid red line. The 22-model average $\pm 2\sigma_{SE}$ are shown as lighter red lines. Some of the values for the models for 1000 hPa are not consistent with the surface value or the value at 925 hPa. This is probably because some model values for p = 1000 hPa are unrealistic; they may be below the surface. So instead of using the values for p = 1000 hPa we used the surface values. MSU values of T_{2LT} and T_2 are shown in the panel to the right. UAH values are yellow-filled diamonds, RSS are yellow-filled squares, and UMD is a yellow-filled circle. Synthetic model values from Table III are shown as white-filled circles, with $2\sigma_{SE}$ uncertainty limits as error bars. From the text, the uncertainties of the observational datasets are: surface, ±0.04 °C/decade; radiosondes, ±0.10 °C/decade; MSU satellites, ±0.10 °C/decade. At the surface, the mean of the models and the observations are seen to agree within the uncertainties; hence the overlap of symbols.

observations with greenhouse models is the layer between 450 and 750 hPa (Schneider *et al.*, 1999), where the presence of water vapour is most important. Lindzen (2000) has called this the 'characteristic emission layer (CEL)'. In this region, the disparity between the model average and the observations is in every case outside the $2\sigma_{SE}$ confidence level. From 500 to 150 hPa, model R values range from +1.5 to +2.1, while in radiosonde observations the range is −1.3 to +0.8. Given the trend uncertainty of 0.04 °C/decade quoted by Free *et al.* (2005) and the estimate of ±0.07 for HadAT2, the uncertainties do not overlap according to the $2\sigma_{SE}$ test and are thus in disagreement.

3.2. Uncertainties

3.2.1. Surface

The three observed trends are quite close to each other. There are possibly systematic errors introduced by urban heat-island and land-use effects (Pielke *et al.*, 2002; Kalnay and Cai, 2003) that may contribute a positive bias, though these are estimated as being within ±0.04 °C/decade (Jones and Moberg, 2003).

3.2.2. Radiosondes

Free *et al.* (2005) estimate the uncertainties in the trend values as 0.03–0.04 °C/decade for pressures in the range 700–150 hPa. For HadAT2, using their figure 10, we estimate the uncertainties as 0.07 °C/decade over the range 850–200 hPa. For RAOBCORE, Haimberger (2006) gives the uncertainty as 0.05 °C/decade for the range 300–850 hPa.

Several investigators revised the radiosonde datasets to reduce possible impacts of changing instrumentation and processing algorithms on long-term trends. Sherwood *et al.* (2005) have suggested biases arising from daytime solar heating. These effects have been addressed by *Christy et al.* (2007) and by Haimberger (2006). Sherwood *et al.* (2005) suggested that, over time, general improvements in the radiosonde instrumentation, particularly the response to solar heating, has led to negative biases in the daytime trends *vs* nighttime trends in unadjusted tropical stations. Christy *et al.* (2007) specifically examined this aspect for the tropical tropospheric layer and indeed confirmed a spuriously negative trend component in composited, unadjusted daytime data, but also discovered a likely spuriously positive trend in unadjusted nighttime measurements. Christy *et al.* (2007) adjusted day and night readings using both UAH and RSS satellite data on individual stations. Both RATPAC and HadAT2 compared very well with the adjusted datasets, being within ±0.05 °C/decade, indicating that main cooling effect of the radiosonde changes were evidently detected and eliminated in both. Haimberger (2006) has also studied the daytime/nighttime bias and finds that 'The spatiotemporal consistency of the global radiosonde dataset is improved by these adjustments and spurious large day-night differences are removed.' Thus, the error estimates

DOI: 10.1002/joc

stated by Free *et al.* (2005); Haimberger (2006), and Coleman and Thorne (2005) are quite reasonable, so that the trend values are very likely to be accurate within ±0.10 °C/decade.

3.2.3. MSU satellite measurements

Thorne *et al.* (2005a) consider the uncertainties in climate-trend measurements; dataset construction methodologies can add bias, which they call structural uncertainty (SU). We take the difference between MSU UAH and MSU RSS trend values, ∼0.1 °C/decade, as an estimate of SU.

Much has been made of the disparity between the trends from RSS and UAH (Santer *et al.*, 2005) – caused by differences in adjustments to account for time-varying biases. Christy and Norris (2006) find that UAH trends are consistent with a high-quality set of radiosondes (VIZ radiosondes) for T_{2LT} at the level of ±0.06 and for T_2 at the level ±0.04. For RSS the corresponding values are ±0.12(T_{2LT}) and ±0.10(T_2). For T_{2LT}, Christy *et al.* (2007) give a tropical precision of ±0.07 °C/decade, based on internal data-processing choices and external comparison with six additional datasets, which all agreed with UAH to within ±0.04. Mears and Wentz (2005) estimate the tropical RSS T_{2LT} error range as ±0.09. Thus, there is evidence to assign slightly more confidence to the UAH analysis.

UMD does not provide statistics of inter-satellite error reduction, and, since the data are not in a form to perform direct radiosonde comparison tests, we are unable to estimate its error characteristics. In the later discussion we indicate the likelihood of spurious warming. UMD data were not in a form to allow detailed analysis such as provided in Christy and Norris (2006) to generate an error estimate.

3.2.4. Model results

The uncertainties of the trends at the various pressure levels and for the surface are determined by $\pm 2\sigma_{SE}$, defined above, and are plotted in Figure 1. The values and the uncertainties of the synthetic T_{2LT} and the T_2 are listed in Table III. It is important to understand the design of this experiment. As will be shown, the mean surface trend of the model simulations is very close to the observed surface trend. This fortuitous result grants the opportunity to answer the question, 'How well do modelled upper air temperature tendencies compare with observations?' With such a large number of simulations in the sample size, and a consistency between models and observations at the surface, we are able to determine with high confidence what models suggest in answer to this question.

4. Discussion and conclusions

4.1. Evaluating the extent of agreement between models and observations

Our results indicate the following, using the $2\sigma_{SE}$ criterion of consistency: (1) In all cases, radiosonde trends are inconsistent with model trends, except at the surface. (2) In all cases UAH and RSS satellite trends are inconsistent with model trends. (3) The UMD T_2 product trend is consistent with model trends.

Case 1: Evidence for disagreement: There is only one dataset, UMD T_2, that does not show inconsistency between observations and models. But this case may be discounted, thus implying complete disagreement. We note, first, that T_2 represents a layer that includes temperatures from the lower stratosphere. In order for UMD T_2 to be a consistent representation of the entire atmosphere, the trends of the lower stratosphere must be significantly more positive than any observations to date have indicated. But all observed stratospheric trends, for example by MSU T_4 from UAH and RSS, are significantly negative. Also, radiosonde trends are even more profoundly negative – and all of these observations are consistent with physical theory of ozone depletion and a rising tropopause. Thus, there is good evidence that UMD T_2 is spuriously warm. Summing up, then, there is a plausible case to be made that the observational trends are completely inconsistent with model trends, except at the surface.

Case 2. Is some agreement possible? To establish this we would essentially require evidence that the radiosonde, UAH, and RSS-trends are significantly more positive than observed. Even if the extreme of the confidence intervals of Christy and Norris (2006) and Christy *et al.* (2007) were applied to UAH data, the results would still show inconsistency between UAH and the model average. Further, we would require evidence of time-dependent biases that are negative across a wide range of radiosonde types throughout the tropics and within the US VIZ radiosondes network. (Recall that Sherwood *et al.* (2005) examined unadjusted radiosonde observations, whereas RATPAC, RAOBCORE, and HadAT2 used here have been adjusted.) Additionally, the six datasets examined in Christy *et al.* (2007) (not reported on here) support the less positive trends represented by RATPAC and HadAT2. While these adjusted datasets likely retain some errors, we do not see any indication that the errors would be so pervasive as to imply a trend error greater than 0.1–0.4 °C/decade above 500 hPa.

Case 3: Is any conclusion possible? The answer may be 'no' if our experimental design is flawed. To assess the extent of agreement, we have attempted to create the most reasonable and defensible comparison between models and observations. To come to no conclusion is a defensible action if the observational datasets present a range from which no confident assessment of observational accuracy may be made. The assumption here is that every dataset is equally likely to have minimal error. If this is the case, then pronouncements that model and observational data are consistent cannot be made and the many recommendations for improvements in upper-air observations contained in the CCSP report (Karl *et al.*, 2006) should be addressed forthwith. *Our view, however, is that the weight of the current evidence, as outlined*

DOI: 10.1002/joc

above, supports the conclusion that no model-observation agreement exists.

4.2. Efforts to remove the disparity between observations and models

Santer *et al*. (2005) have argued that the model results are consistent with observations and that the disparity between the models and observations is 'removed' because their ranges of uncertainties overlap. They define 'range' as the region between the minimum and maximum of the simulations among the various models. However, 'range' is not statistically meaningful. Further – and more importantly – it is not robust; a single bad outlier can increase the range value of model results to include the observations. A more robust estimate of model variability is the uncertainity of the mean of a sufficiently large sample.

We shall now discuss the range issue in more detail. Recently, Karl *et al*. (2006) compared trends – using the same models, simulations, and datasets as those of Santer *et al*. (2005). Their conclusions are that '... [W]hile these data are consistent with the results from climate models at the global scale, discrepancies in the tropics remain to be resolved.' The discrepancies are: that on decadal time scales '...[a]lmost all model simulations show greater warming aloft [while]... most observation show greater warming at the surface.' They offer two possible explanations: (1) The models are wrong, and (2) The datasets are biased. They conclude that '[T]he new evidence in this Report favours the second explanation.' Our analysis of the 'new evidence' indicates that their choice of explanation 2 is not justified. In fact, their own analysis supports the conclusions of our present article. In their chapter 5, Karl *et al*. calculate trends of T_S (surface) and T_{2LT} (synthetic) for each of 49 model simulations, and show a histogram of the values of $T_S - T_{2LT}$ along with the values from the observations (Figure 5.4G of *CCSP-SAP-1.1*). It is clear from their figure that all the observations deviate from the mean – as we have also found. To support their claim of agreement, the uncertainties of the models and observations would have to overlap. If one defines the uncertainty of the model results as from $-2\sigma_{SE}$ to $+2\sigma_{SE}$, then they do not overlap. However, Karl *et al*. define uncertainty as the *range* of values, which is considerably larger. How much larger? In Table IV we show that the ratio of range to $4\sigma_{SE}$ is 5.8 for the Karl *et al*. 49-value suite of simulations, and 9.0 for our suite of 67 values. 'Range' thus overestimates the uncertainty by large factors. See Table IV for more discussion.

Thus the use of the 'range' definition of uncertainty allows misleading statements to be made, such as 'Discrepancies between the datasets and the models have been reduced.'

Additionally, we point out a related and misleading feature of CCSP-SAP-1.1. By selecting the *range* of model outputs, comparisons against observations were shown which included some model simulations with very small upper-air trends because their surface trends were likewise unrealistically small. But these few results were not consistent with surface observations at all and should not have been utilized in the comparison. Our experimental design is more rigorous. *We are comparing the best possible estimate of model-produced upper-air trends that are consistent with the magnitude of the observed surface trend.* With this pre-condition in place (granted to us by the fact the mean of the modelled surface trends was very close to observations) the upper air comparisons become informative and not confused by one or two model runs which are *de facto* inconsistent with observed surface trends.

There is an enormous ongoing effort to find errors in the observations that would reduce the disagreement with the models. Here, the task is daunting since the various datasets are independently constructed and one needs to find something wrong with each one of them. In regard to the observations, Thorne *et al*. (2005b) say '...*As a community we must assume that the latest dataset versions are the best estimates based upon investigators' knowledge and experience using the data*'. We agree: the values given are the values we should use.

Recently, Thorne *et al*. (2007) performed a limited comparison between observations of tropical layer–mean temperatures and the same quantity in a series of Hadley Centre climate model simulations. The model uncertainties are even less than reported here. They found that for the period in question (1979–2004), a comparison of the surface–tropospheric temperature relationship between the model estimate and all but one

Table IV. Difference between modelled surface trends and (weighted) troposphere trends, $T_S - T_{2LT}$(synthetic) (°C/decade).

	Average[b]	SD	SE	max	min	Range	Range/ ($4\sigma_{SE}$)
Karl *et al*. 19 models (49 simulations)	−0.060	0.028	0.0066	0.005	−0.145	0.150	5.9
This article (22 models, 67 simulations)	+0.0053	0.054	0.0117	0.197	−0.211	0.408	9.0
This article (20 models, outliers removed[a]) (61 simulations)	+0.0006	0.024	0.0055	0.108	−0.062	0.170	7.8

SD, standard deviation; SE, uncertainty of the mean.

[a] The effect of outliers (models 2 and 17) is to increase the 'range'. This illustrates that range is not a robust statistical measure and is unsuitable for comparing model results with observations.

[b] Although the difference of the averages of Karl *et al*. and this article are of opposite sign, they are close to one another. One would expect the SD and SE for the 22-model suite to be smaller. However, if the outliers are very large they can cause the opposite effect, as seen in the table.

DOI: 10.1002/joc

of the observational datasets showed that a discrepancy existed of the same type as demonstrated in this article.

5. Summary

We have tested the proposition that greenhouse model simulations and trend observations can be reconciled. Our conclusion is that the present evidence, with the application of a robust statistical test, supports rejection of this proposition. (The use of tropical tropospheric temperature trends as a metric for this test is important, as this region represents the CEL and provides a clear signature of the trajectory of the climate system under enhanced greenhouse forcing.) On the whole, the evidence indicates that model trends in the troposphere are very likely inconsistent with observations that indicate that, since 1979, there is no significant long-term amplification factor relative to the surface. If these results continue to be supported, then future projections of temperature change, as depicted in the present suite of climate models, are likely too high.

In summary, the debate in this field revolves around the idea of discrepancy in surface and tropospheric trends in the tropics where vertical convection dominates heat transfer. Models are very consistent, as this article demonstrates, in showing a significant difference between surface and tropospheric trends, with tropospheric temperature trends warming faster than the surface. What is new in this article is the determination of a very robust estimate of the magnitude of the model trends at each atmospheric layer. These are compared with several equally robust updated estimates of trends from observations which disagree with trends from the models.

The last 25 years constitute a period of more complete and accurate observations and more realistic modelling efforts. Yet the models are seen to disagree with the observations. We suggest, therefore, that projections of future climate based on these models be viewed with much caution.

Acknowledgements

We thank R. S. Knox for many valuable conversations. We thank P. Thorne, C. Mears, and R. Courtney for valuable communications. We particularly thank Leopold Haimberger and Melissa Free for providing radiosonde data. We also thank J. Lawrimore and C. Tankersley of NOAA for providing the GHCN tropical anomalies. We acknowledge the international modelling groups for providing their data for analysis, the Program for Climate Model Diagnosis and Inter-comparison (PCMDI) for collecting and archiving the model data, 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.

References

Christy JR, Norris WB. 2006. Satellite and VIZ-radiosonde intercomparison for diagnosis of non-climate influences. *Journal of Atmospheric and Oceanic Technology* **23**: 1181–1194.

Christy JR, Norris WB, Spencer RW, Hnilo JJ. 2007. Tropospheric temperature change since 1979 from tropical radiosonde and satellite measurements. *Journal of Geophysical Research-Atmospheres* **112**: D06102, DOI:10.1029/2005JD006881.

Coleman H, Thorne PW. 2005. HadAT: An update to 2005 and development of the dataset website. Data at <http://hadobs.metoffice.com/hadat/update_report.pdf>.

Douglass DH, Pearson BD, Singer SF. 2004a. Altitude dependence of atmospheric temperature trends: climate models versus observation. *Geophysical Research Letters* **31**: L13208, DOI:10.1029/2004GL020103.

Douglass DH, Pearson BD, Singer SF, Knappenberger PC, Michaels PJ. 2004b. Disparity of tropospheric and surface temperature trends: new evidence. *Geophysical Research Letters* **31**: L13207, DOI:10.1029/2004GL020212.

Free M, Seidel DJ, Angell JK. 2005. Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC): a new dataset of large-area anomaly time series. *Journal of Geophysical Research* **110**: D22101, DOI:10.1029/2005/D006169.

GHCN. 2005. Global temperature anomalies found at: http://lwf.ncdc.noaa.gov/oa/climate/research/anomalies/anomalies.html#anomalies.

GISS. 2005. Temperature anomalies at: http://data.giss.nasa.gov/gistemp/graphs/Fig.B.txt.

Haimberger L. 2006. Homogenization of radiosonde temperature time series using innovation statistics. *Journal of Climate* **20**: 1377–1402.

Johanson CM, Fu Q. 2006. Robustness of tropospheric temperature trends from MSU channel 2 and 4. *Journal of Climate* **19**: 4234–4242.

Jones PD, Moberg A. 2003. Hemispheric and large-scale surface air temperature variations: an extensive revision and an update to 2001. *Journal of Climate* **16**: 206–223, HadCRUT2v is the designation for version 2.

Kalnay E, Cai M. 2003. Impact of urbanization and land-use change on climate. *Nature* **423**: 528–531.

Karl TR, Hassol SJ, Miller CD, Murray WL (eds). 2006. *Temperature Trends in the Lower Atmosphere: Steps for Understanding and Reconciling Differences. A Report by the Climate Change Science Program (CCSP) and the Subcommittee on Global Change Research*: Washington, DC, US Government. (Report at http://www.climatescience.gov/Library/sap/sap1-1/finalreport/default.htm).

Lanzante JR, *et al*. 2003. Temporal homogenization of monthly radiosonde data. Part II: trends, sensitivities, and MSU comparison. *Journal of Climate* **16**: 224–240.

Lindzen R. 2000. *Climate Forecasting: When Models are Qualitatively Wrong*. Lecture at the George C. Marshall Institute: Washington, DC; 2000.

Mears CA, Wentz FJ. 2005. The effect of Diurnal Correction on the satellite-derived Lower tropospheric temperature. *Science* **309**: 1548–1551, Data at http://www.ssmi.com/msu/msu_data_description.html.

National Research Council. 2000. *Reconciling Observations of Global Temperature Change*. National Academy Press: Washington, DC.

Pielke RA Sr, *et al*. 2002. The influence of land-use change and landscape dynamics on the climate system: relevance to climate-change policy beyond the radiative effects of greenhouse gases. *Philosophical Transactions of the Royal Society of London, Series A* **360**: 1–15.

PCMDI. 2005. The 20CEN forcing consists of a variety of natural and "human-caused" climate forcings that have changed over the last century. The results of the 20CEN forcings of the various models are archived by the US Department of Energy, http://www-pcmdi.llnl.gov/ipcc.

Santer BD, *et al*. 2003. Influence of satellite data uncertainties on the detection of externally forced climate change. *Science* **300**: 1280–1284.

Santer BD, *et al*. 2005. Amplification of surface temperature trends and variability in the tropical atmosphere. *Science* **309**: 1551–1555.

Schneider EE, Kirtman BP, Lindzen RE. 1999. Tropospheric Water Vapor and Climate Sensitivity. *Bulletin of the American Meteorological Society* **54**: 1649–1658.

Sherwood SC, Lanzante JR, Meyer CL. 2005. Radiosonde daytime biases and late – 20th century warming. *Science* **309**: 1556–1559.

DOI: 10.1002/joc

Spencer RW, Christy JR. 1992. Precision and radiosonde validation of satellite gridpoint temperature anomalies. Part I: MSU Channel 2. *Journal of Climate* **5**: 847–857.

Thorne PW, *et al*. 2005a. Revisiting radiosonde upper-air temperatures from 1958 to 2002. *Journal of Geophysical Research* **110**: D18105, DOI:10.1029/2004JD005753.

Thorne PW, *et al*. 2005b. Uncertainties in climate trends: lessons from upper-air temperature records. *Bulletin of the American Meteorological Society* **86**: 1437–1442, DOI:10.1175/BAM-86-19-1437.

Thorne PW, *et al*. 2007. Tropical vertical temperature trends: a real discrepancy? *Geophysical Research Letters* **34**: L16702, DOI:10.1029/2004GL029875.

UAH. 2007. MSU anomalies TLT(2LT) and T2 are found at (version 5.2), http://vortex.nsstc.uah.edu/public/msu/.

Vinnikov KY, *et al*. 2006. Temperature trends at the surface and in the troposphere. *Journal of Geophysical Research* **111**: D03106, DOI:10.1029/2005JD006392.

