# Panel and Multivariate Methods for Tests of Trend Equivalence in Climate Data Series

**Ross McKitrick**
(Corresponding Author)
Department of Economics
University of Guelph
rmckitri@uoguelph.ca

**Stephen McIntyre**
Toronto, Ontario
stephen.mcintyre@utoronto.ca

**Chad Herman**
Manalapan, NJ
chad.herman.us@gmail.com

Final Revised version
August 3, 2010

**In press, *Atmospheric Science Letters***

**Abstract**

We explain panel and multivariate regressions for comparing trends in climate data sets. They impose minimal restrictions on the covariance matrix and can embed multiple linear comparisons, which is a convenience in applied work. We present applications comparing post-1979 modeled and observed temperature trends in the tropical lower- and mid-troposphere. Results are sensitive to the sample length. In data spanning 1979 to 1999, observed trends are not significantly different from zero or from model projections. In data spanning 1979 to 2009 the observed trends are significant in some cases but tend to differ significantly from modeled trends.

Short title:
Panel and multivariate trend equivalence tests

Key words:
Trend comparisons, Panel regression, Tropical troposphere

# Panel and Multivariate Methods for Tests of Trend Equivalence in Climate Data Series

# 1 Introduction

Many issues of interest in climate analysis involve comparisons of trends across different data sets. This note explains regression-based methods that yield asymptotically valid parameter variances and covariances while providing a flexible testing framework. Obtaining linear trend coefficients is easy using ordinary least squares (OLS). Obtaining unbiased estimates of the parameter variances and covariances (collectively referred to as the covariance matrix) is more challenging, because the regression residuals may be autocorrelated within each panel, and both heteroskedastic (unequal variance) and correlated across panels. Regressions that use sequenced groups of time series observations are referred to as *panel* estimators (Davidson and MacKinnon 2002). They are convenient when panels are unbalanced, i.e. they do not all have the same numbers of observations, but they impose restrictions on the covariance matrix. A nonparametric method introduced by Vogelsang and Franses (2005) handles autocorrelation of unknown dimension, however it is only applicable to balanced panels.

We explain both methods and the tradeoffs between them. In Section 3 we apply them to a comparison of model temperature projections and observations in the tropical troposphere. We test trend significance as well as model-data equivalence. For discussions of the importance of modeling and climatological measurement issues related to the tropical atmosphere see Karl et al. (2006) Santer et al. (2005, 2008) and Douglass et al. (2007).

# 2 Methods

## 2.1 Introduction: two-equation case

We assume the data are stationary, though autocorrelated, upon detrending; in other words "trend stationary." Suppose there are two series of interest, $y_{1\tau}$ and $y_{2\tau}$, where $\tau$ = 1,…,*T*. Trends are fitted using:

$$y_{1\tau} = a_1 + b_1\tau + u_{1\tau} \qquad (1)$$

and

$$y_{2\tau} = a_2 + b_2\tau + u_{2\tau}\,. \tag{2}$$

A student's *t* test of slope equivalence is:

$$t = \frac{\hat{b}_1 - \hat{b}_2}{\sqrt{\tilde{s}_1^2 + \tilde{s}_2^2 - 2\,\mathrm{cov}(\hat{b}_1, \hat{b}_2)}} \tag{3}$$

where ^ denotes an ordinary least squares estimate, $\tilde{s}_i^2$ ($i$=1,2) denotes an autocorrelation-robust variance estimator for $\hat{b}_i$, and cov( $\hat{b}_1, \hat{b}_2$ ) is the estimated covariance between the trend terms.

Karl et al. (2006) drew attention to an apparent discrepancy between observed and model-generated temperature trends in the tropical atmosphere. Douglass et al. (2007) tested surface-matched differences (see SI) using

$$t = \frac{\hat{b}_1 - \hat{b}_2}{\tilde{s}_1} \tag{4}$$

where $\hat{b}_1$ denotes the trend through model ensemble means, $\hat{b}_2$ denotes the trend through observations and $\tilde{s}_1$ is the estimated standard error of $\hat{b}_1$. The test (4) incorrectly treats the observations as deterministic and assumes the model observations are independent across time. Santer et al. (2008) instead used

$$t = \frac{\hat{b}_1 - \hat{b}_2}{\sqrt{\frac{1+r_1}{1-r_1}\tilde{s}_1^2 + \frac{1+r_2}{1-r_2}\tilde{s}_2^2}} \tag{5}$$

where ~ denotes a least-squares estimate and $r_i$ denotes the first-order autoregressive (AR1) coefficient in series *i*. The ratio of AR1 terms is commonly referred to as an "effective degrees of freedom" adjustment (e.g. Santer et. al. 2000). Instead of a series providing *T* independent observations it is said to provide only $(1-r_i)T/(1+r_i)$ independent observations. The resulting variance corresponds to an estimate obtained using an AR1 model, but is not equivalent to that derived from higher-order autocorrelation models. Also, it does not yield a correct $2\,\mathrm{cov}(\hat{b}_1, \hat{b}_2)$ term (see SI), which was missing in both (4) and (5) anyway. While detrended climate model projections may be uncorrelated with observations, the assumption of no covariance among trend coefficients implies models have no low-frequency correspondence with observations in response to observed forcings, which seems overly pessimistic.

## 2.2 Panel Regressions

Equation (3) can be obtained using a panel regression. Suppose the dependent variable is the stacked vector $(y_1, y_2)'$, and we estimate the following equation:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + d_1 \begin{pmatrix} \tau \\ \tau \end{pmatrix} + d_2 \begin{pmatrix} 0 \\ \tau \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \qquad (6).$$

$(1\ 1)'$ denotes two stacked $T$-length vectors of ones. $(0\ 1)'$ denotes a vector of $T$ zeroes stacked on $T$ ones. This is called an indicator or "dummy variable," since it indicates (value=1) if the dependent variable is $y_2$. $(\tau\ \tau)'$ denotes a $2T$-length vector consisting of two $T$-length time trends and $(0\ \tau)'$ is $(\tau\ \tau)'$ times $(0\ 1)'$. A test of $\hat{d}_2 = 0$ in (6) can be shown to be equivalent to testing $\hat{b}_1 = \hat{b}_2$ (Kmenta 1986, see SI). Hence the $t$-statistic on $\hat{d}_2$ in equation (6) yields the test score (3).

To generalize the framework further, suppose we are comparing $m$ model-generated series and $o$ observational series, making the total number of series $N$=$m$+$o$. Each source $i$ yields $T_i \le T$ non-missing observations $y_{i\tau}$ over the interval $\tau$ = 1,…,$T$. Define an indicator variable $obs_{i\tau} = 0$ if the record is model-generated, and =1 if it is from an observational series. Denote the $i$-th vector as $y_i' = [y_{i1}, \ldots, y_{iT}]$. Stack these vectors into a single $NT \times 1$ vector **y** as follows:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \qquad (7).$$

Stack the trend vector $\tau' = [1, \ldots, T]$ $N$ times to get the $NT \times 1$ panel trend vector

$$\mathbf{t} = \begin{bmatrix} \tau \\ \vdots \\ \tau \end{bmatrix} \qquad (8).$$

The indicator, or dummy, variables are likewise stacked to form

$$\mathbf{d} = \begin{bmatrix} obs_1 \\ \vdots \\ obs_N \end{bmatrix} \qquad (9)$$

where $obs_i$ is $(obs_{i1}, \ldots, obs_{iT})'$. The regression equation is then written

$$\mathbf{y} = b_0 + b_1 \mathbf{t} + b_2 (\mathbf{d} \times \mathbf{t}) + b_3 \mathbf{d} + \mathbf{e} \qquad (10)$$

where **e** is an $NT \times 1$ residual vector with typical element $e_{i\tau}$. Note that all the "data" are on the left hand side, and the right hand side consists of dummy variables and trend terms.

When $obs_{ij} = 0$, $dy_{i\tau} / d\tau = \hat{b}_1$ and when $obs_{it} = 1$, $dy_{i\tau} / d\tau$ yields $(\hat{b}_1 + \hat{b}_2)$. Thus a *t* statistic on $\hat{b}_1$ will test whether the model trend is zero and a test of the linear restriction $\hat{b}_1 + \hat{b}_2 = 0$ indicates the significance of the observed slope. The *t* statistic on $\hat{b}_2$ tests whether the trend on observations differs significantly from the trend in models.

Equation (10) can be extended further. Suppose observations come from two different systems, such as satellites and weather balloons. Define two different indicator variables: $\mathbf{d}_1$, which equals 1 if an observation is from either system 1 or 2, and $\mathbf{d}_2$ which equals 1 only if the observation is from system 2. The regression equation becomes:

$$\mathbf{y} = b_0 + b_1\mathbf{t} + b_2(\mathbf{d}_1 \times \mathbf{t}) + b_3\mathbf{d}_1 + b_4(\mathbf{d}_2 \times \mathbf{t}) + b_5\mathbf{d}_2 + \mathbf{e} \qquad (11).$$

The estimated model trend is $\hat{b}_1$. The trend in observations from system 1 is $\hat{b}_1 + \hat{b}_2$ and from system 2 is $\hat{b}_1 + \hat{b}_2 + \hat{b}_4$. The *t* statistic on $\hat{b}_4$ tests whether the trend in the second observation system differs from that in the first, and so forth.

Hypothesis testing requires a valid estimator of $V(\mathbf{b})$, the covariance matrix of **b**. The general form is (Davidson and MacKinnon 2002)

$$V(\mathbf{b}) = E(\hat{\mathbf{b}} - \mathbf{b})^2 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \qquad (12).$$

where **X** denotes the right hand side variables in (11) and $\Omega = E(\mathbf{ee}')$. Obtaining a valid estimate of $\Omega$ requires modeling the cross- and within-panel covariances. For a panel *i* with *T* observations, define a matrix $\mathbf{A}_i$ of autoregressive weights using the panel-specific AR1 coefficient $\rho_i$:

$$\mathbf{A}_i = \begin{bmatrix} 1 & \rho_i & \rho_i^2 & \cdots & \rho_i^{T-1} \\ \rho_i & 1 & \rho_i & \cdots & \rho_i^{T-2} \\ \rho_i^2 & \rho_i & 1 & \cdots & \rho_i^{T-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_i^{T-1} & \rho_i^{T-2} & \rho_i^{T-3} & \cdots & 1 \end{bmatrix} \qquad (13).$$

Then a model of $\Omega$ can be written

$$\Omega = \begin{bmatrix} \sigma_1^2\mathbf{A}_1 & \phi_{21}^2\mathbf{I}_2 & \cdots & \phi_{N1}^2\mathbf{I}_N \\ \phi_{21}^2\mathbf{I}_2 & \sigma_2^2\mathbf{A}_2 & \cdots & \phi_{N2}^2\mathbf{I}_N \\ \vdots & \vdots & & \vdots \\ \phi_{N1}^2\mathbf{I}_N & \phi_{N2}^2\mathbf{I}_N & \cdots & \sigma_N^2\mathbf{A}_N \end{bmatrix} \qquad (14)$$

where $\phi_{ij}^2$ denotes the covariance between series *i* and *j*, $\mathbf{I}_i$ denotes an identity matrix with dimension *T*, and $\sigma_i^2$ denotes the variance of series *i*. There are *N*(*N*-1)/2 covariances $\phi_{ij}^2$ in Equation (14) needing to be estimated, in addition to the variances and AR1 parameters. If some panels *j* are shorter than others ($T_j < T$) then the dimensions of the $\mathbf{A}_i$ matrices need to be adjusted accordingly. Some commercial statistical packages, like STATA, can accommodate unbalanced data sets.

### 2.3 Higher-order autocorrelations and multivariate trend models

Vogelsang and Franses (2005, herein VF05) derived two estimators for $\Omega$ that impose no parametric restrictions on the lag and correlation structure, as is done in (11). Suppose the *N* panels are used one-at-a-time in Equation (1), yielding OLS trend estimates $\hat{\mathbf{b}} = \hat{b}_1,...,\hat{b}_N$. Take the *N* residual series $u_{1\tau},...,u_{N\tau}$ and form the $T \times N$ matrix $\mathbf{U} = [u_{1\tau},...,u_{N\tau}]$. VF05 derive two transformations of $\mathbf{U}$ that converge in probability to a scalar multiple of $\Omega$. Of their two estimators we focus on the $F_2^*$ form which has higher power and is slightly easier to compute. It is obtained as follows. Denote $\mathbf{V} = \mathbf{U}'$ and take the columns $v_j$, for *j* = 1,…, *T*, each of length *N*. Define a vector $\hat{S}_\tau = \sum_{j=1}^{\tau} v_j$. Then VF05 show that

$$\hat{\Omega} = 2N^{-2}\sum_{\tau=1}^{T}\hat{S}_\tau \hat{S}_\tau{}' \qquad (15)$$

converges in probability to an unbiased estimate of $\Omega$, regardless of the form of autocorrelation and other departures from the independence assumption. For testing purposes linear restrictions on the slopes can be written in the matrix form $\mathbf{R}\hat{\mathbf{b}} = \mathbf{0}$ (see SI). The VF05 test statistic is

$$F_2^* = \mathbf{R}\hat{\mathbf{b}}\left[\mathbf{R}\eta^{-1}\hat{\Omega}\mathbf{R}'\right]^{-1}\mathbf{R}\hat{\mathbf{b}} / q \qquad (16)$$

where $\eta = \Sigma(t - \bar{t})^2$ and *q* is the number of restrictions, which in our examples always equals 1. Critical values for Equation (16) generated by Monte Carlo simulation are reported in VF05.

The VF05 approach improves on the panel method by providing robust trend variances and covariances regardless of the autocorrelation order and the structure of heteroskedasticity. However it requires balanced panels, which can be a limitation in some cases.

The VF05 statistic, as with all test statistics, has improved size as the sample size grows. Rejection probabilities also increase as $\rho \to 1$. Monte Carlo simulations in VF05 show that for *T*= 100, when *q*=1 and $\rho > 0.8$, just under 10% of $F_2^*$ scores exceed the 95$^{th}$ percentile, indicating a tendency to over-reject a true null, though this is an improvement compared to earlier alternatives. Each panel in our full sample has well over 100 observations, but a high $\rho$ value. Hence VF05 scores that are close to the critical values may overstate significance.

# 3 Empirical application

### 3.1 Data

We used the same archive of climate model simulations as Santer et al. (2008). The available group now includes 57 runs from 23 models. Each source provides data for both the lower troposphere (LT) and mid-troposphere (MT). Each model uses prescribed forcing inputs up to the end of the 20$^{th}$ century climate experiment (20C3M, see Santer et al. 2005). Projections forward use the A1B emission scenario. Table 1 lists the models, the number of runs in each ensemble mean and other details. We used four observational temperature series: two satellite-borne microwave sounding unit (MSU)-derived series and two balloon-borne radiosonde series. We use monthly data starting in 1979, covering the tropics from 20 degrees N to 20 degrees S. The MSU observations come from the University of Alabama-Huntsville (UAH, Spencer and Christy 1990) and Remote Sensing Systems Inc. (RSS, Mears et al 2003). The HadAT radiosonde series is an MSU-equivalent published on the Hadley Centre web site (http://hadobs.metoffice.com/hadat/msu_equivalents.html, Thorne et al. 2005). The Radiosonde Innovation Composite Homogenization (RICH) series is published by Haimberger et al. (2008) and is available at ftp://srvx6.img.univie.ac.at/pub/rich_gridded_2009.nc. We used the RICH gridded data and MSU-weights supplied by John Christy (pers. comm.) to construct MSU-equivalent series (see SI for details).

Our data start in January 1979 and end in December 2009. Thus we have $N$=27 panels, each with 372 monthly observations. Figure 1 displays the (smoothed) MSU series and the mean of the PCM model runs for comparison.

Douglass et al. (2007) and Santer et al. (2008) focused on trends from 1979 to about 1999, with some series extending a few years further. To compare with these results we first look at data ending in 1999, and then extend the sample to 2009. Since our panels are balanced we can generate results using both the VF05 and panel regression methods, but since the results are so similar we report only the VF05 results for the shorter 1979-1999 sample.

Table 1 summarizes the data. 1979-2009 trends in degrees C decade$^{-1}$ are shown for the LT and MT levels, with accompanying standard errors, for all ensemble means and observational series. Each series was centered and the trend regression allowed for a six-lag autoregressive process, denoted AR6. Table 1 (final column) shows that in 17 of the 23 models, and in all 4 observational series, autocorrelation at lags greater than one were observed in at least one atmospheric layer. Hence an AR1 error specification is likely inadequate. Extended autocorrelation lags were also observed in the individual model runs.

All climate models were forced with 20$^{th}$ century greenhouse gas and sulfate levels: other assumed forcings are listed in Table 1.

### 3.2 Multivariate trend test results

We weighted each model by the number of runs in its ensemble to adjust for the effect of combining runs into an average, though our conclusions would be unchanged if we weighted each model equally.

Table 2 presents tests of trend significance for the observational series. On data ending in 1999 the VF05 test shows the four observational series are insignificant at both the LT and MT layers individually and averaged together (Column 'Obs'). By extending the data to 2009 the $F_2^*$ score of combined significance at the LT layer rises from 12.50 to 76.66, thus attaining significance at 5%. All observed LT series are individually significant, except UAH which is significant at 10%. At the MT layer, extending the sample raises the combined $F_2^*$ score from 5.06 to 23.77, which is significant at 10%. UAH and Hadley series are insignificant, RICH is marginal and RSS is individually significant at 5%.

Trend comparison results are in Table 3. The second column ("Obs") shows that at both the LT and MT layers, on data ending in 1999 the difference between models and observations is only marginally significant, echoing the findings of Santer et al. (2008). But with the addition of another decade of data the results change, such that the differences between models and observations now exceed the 99% critical value. As shown in Table 1 and Section 3.3, the model trends are about twice as large as observations in the LT layer, and about four times as large in the MT layer.

At both the LT and MT layers, on data ending in either 1999 or 2009, the VF05 tests show that the balloon data are not significantly different from the MSU data, but within the satellite category, the RSS and UAH data are significantly different. Possible reasons for RSS/UAH differences include treatment of intersatellite calibration, orbital decay and other processing issues (see, e.g., Santer et al. 2005, Christy and Norris 2009, Karl et al. 2006).

### 3.3 Panel regressions tests

In cases where one or more series is not of full length the VF05 test will not work. The panel-corrected standard error estimator in the STATA program (command xtpcse) allows an unbalanced panel in the estimate of Equation (11), however it imposes an AR1 assumption. For comparison purposes we report these results on data ending in 2009. We again weighted each observation by the number of runs in the ensemble mean. None of the conclusions depend on this step.

In Table 2 the panel estimator at the LT layer shows that the observations as a group (column 2) exhibit a significant trend of 0.110 C decade$^{-1}$, compared to a model trend (column 9) of 0.272 C decade$^{-1}$. The balloon and MSU series are each jointly significant ($p$ = 0.026 and 0.042 respectively). In the MT layer the model trend (0.253 C decade$^{-1}$) remains significant. The mean observed trend is only 0.057 C decade$^{-1}$. The panel-estimated standard error implies it is insignificant ($p$=0.272) while the VF05 score implies significance at 10%. Among observational series only RSS is individually significant, echoing the VF05 results. The MSU and balloon series are each jointly insignificant. Figures 2 and 3 show the trend magnitudes.

In Table 3, the $p$ values of the test scores on a hypothesis of equality between the indicated trends are shown in the bottom row. On data ending in 2009, the trend differences between models and observations (column 2) are significant in both the LT ($p$=0.002) and MT ($p$ = 0.000) layers, as was the case with the VF05 tests. The model-observation difference is significant for all data products at both layers except for the RSS series in the LT layer ($p$=0.059).

In the final columns of Table 3 we test the differences among the observational series. As was the case with the VF05 tests, the balloons and MSU series are not significantly different from each other ($p$=0.880), but within the MSU category, the RSS and UAH series are significantly different ($p$=0.000).

# 4 Discussion and conclusions

Econometric tools are increasingly being used for climate data sets (see, e.g., Fomby and Vogelsang 2002, Mills 2010). We present two econometric methods for trend comparisons between data sets. Both add flexibility for multivariate comparisons and provide improved treatment of complex error structures. The multivariate testing method of Vogelsang and Franses (2005) yields the more robust estimator of the covariance matrix, but requires balanced data panels. Panel regression methods can accommodate comparisons of series of unequal lengths, but software limitations typically limit treatment of within-panel autocorrelation to the AR1 case. In our example the two methods yielded similar conclusions, indicating that the AR1 approximation in the panel model was likely not overly restrictive. In general, however, for the purpose of multivariate trend comparisons in climatology, we particularly recommend that the VF05 method enter the empirical toolkit.

In our example on temperatures in the tropical troposphere, on data ending in 1999 we find the trend differences between models and observations are only marginally significant, partially confirming the view of Santer et al. (2008) against Douglass et al. (2007). The observed temperature trends themselves are statistically insignificant. Over the 1979 to 2009 interval, in the LT layer, observed trends are jointly significant and three of four data sets have individually significant trends. In the MT layer two of four data sets have individually significant trends and the trends are jointly insignificant or marginal depending on the test used. Over the interval 1979 to 2009, model-projected temperature trends are two to four times larger than observed trends in both the lower and mid-troposphere and the differences are statistically significant at the 99% level.

Our methods assume trends are linear. We found no evidence for nonlinearity on the observed data, but some on modeled data in the MT. Also, the fact that the results are sensitive to the end date suggests that they might also be sensitive to the start date. Since the satellite data are unavailable prior to 1979 we cannot extend these series earlier. Interpretation of trend comparisons should therefore make reference to the time period analysed, which, ideally, should have some intrinsic interest. In this case the 1979-2009 interval is a 31-year span during which the upward trend in surface data strongly suggests a climate-scale warming process. As noted in the studies cited in the introduction, comparing models to observations in the tropical troposphere is an important aspect of testing explanations of the origins of surface warming.

**References**

Christy, J. R., W. B. Norris. 2009. Discontinuity Issues with Radiosonde and Satellite Temperatures in the Australian Region 1979–2006. *J. Ocean and Atmos. Tech*. 26 508-522 DOI: 10.1175/2008JTECHA1126.1

Davidson, R. and J.G. MacKinnon 2004, *Econometric Theory and Methods*. Toronto: Oxford.

Douglass, D. H., J. R. Christy, B. D. Pearson, and S. F. Singer. 2007. A comparison of tropical temperature trends with model predictions. *Intl J Climatology* Vol 28(13) 1693—1701 DOI 10.1002/joc.1651

Fomby, T. and Vogelsang, T.J., 2002. The application of size robust trend analysis to global warming temperature series. *Journal of Climate* **15**, pp. 117–123.

Haimberger, L., C. Tavolato, S. Sperka, 2008: Towards the elimination of the warm bias in historic radiosonde temperature records - some new results from a comprehensive intercomparison of upper air data. *J. Climate*. 21, 4586—4606 DOI 10.1175/2008JCLI1929.1.

Karl, T. R., Susan J. Hassol, Christopher D. Miller, and William L. Murray. 2006. *Temperature Trends in the Lower Atmosphere: Steps for Understanding and Reconciling Differences.* Synthesis and Assessment Product. Climate Change Science Program and the Subcommittee on Global Change Research. http://www.climatescience.gov/Library/sap/sap1-1/finalreport/sap1-1-final-all.pdf. Accessed August 3 2010.

Kmenta, Jan (1986). *Elements of Econometrics* 2nd ed. New York: MacMillan.

Mears, C. A., M. C. Schabel, and F. J. Wentz. 2003. A reanalysis of the MSU channel 2 tropospheric temperature record. *J. Climate* 16, no. 22: 3650-3664.

Mears, C. A., and F. J. Wentz. 2005. The effect of diurnal correction on satellite-derived lower tropospheric temperature. *Science* 309: 1548-1551.

Mills T. C. (2010) Skinning a cat: alternative models of representing temperature trends. *Climatic Change* 101: 415-426, DOI 10.1007/s10584-010-9801-1.

Santer, B. D., P. W. Thorne, L. Haimberger, K. E. Taylor, T. M. L. Wigley, J. R. Lanzante, S. Solomon, M. Free, P. J. Gleckler, and P. D. Jones. 2008. Consistency of Modelled and Observed Temperature Trends in the Tropical Troposphere. *Int. J. Climatol.* Vol 28(13) 1703—1722 DOI: 10.1002/joc.1756

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, G. A. Schmidt 2005. Amplification of Surface Temperature Trends and Variability in the Tropical Atmosphere. *Science* Vol. 309. no. 5740, pp. 1551 – 1556 DOI: 10.1126/science.1114867

Santer, B. D., T. M. L. Wigley, J. S. Boyle, D. J. Gaffen, J. J. Hnilo, D. Nychka, D. E. Parker, and K. E. Taylor. 2000 Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. *J Geophys Res* Vol. 105, No. D6, 7337-7356, March 27, 2000

Spencer, R. W., and J. R. Christy. 1990. Precise monitoring of global temperature trends from satellites. *Science* 247, no. 4950: 1558-1562.

Thorne, P. W., D. E. Parker, S. F. B. Tett, P. D. Jones, M. McCarthy, H. Coleman, P. Brohan, and J. R. Knight, 2005: Revisiting radiosonde upper-air temperatures from 1958 to 2002. *J. Geophys. Res.*, 110, D18105, doi:10.1029/2004JD005753.

Vogelsang, Timothy and Philip Hans Franses (2005) Testing for Common Deterministic Trend Slopes. *Journal of Econometrics* 126 (2005) 1—24.

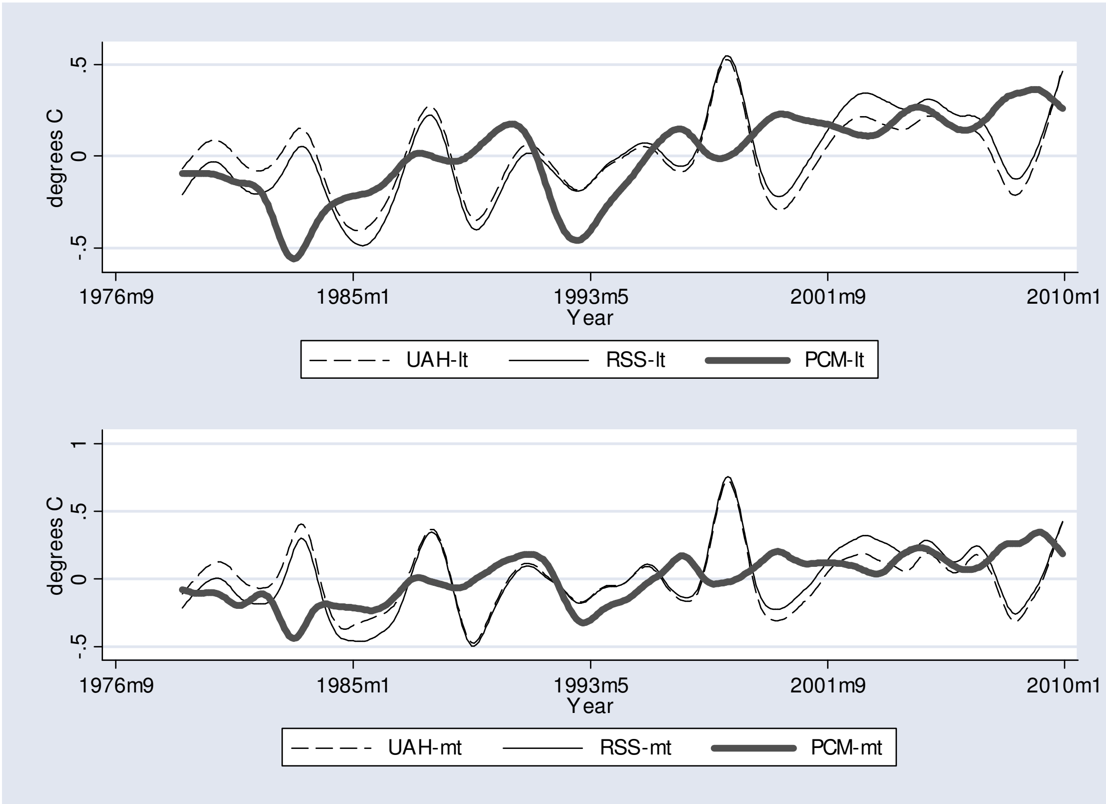


**Figure 1:** UAH (thin dashed) and RSS (thin solid) satellite series 1979:1 to 2008:9. Thick line: Model 21 ensemble mean. Series smoothed using Hodrick-Prescott filter with smoothing parameter $\lambda = 200$ . Top: lower troposphere LT, Bottom: mid-troposphere MT.

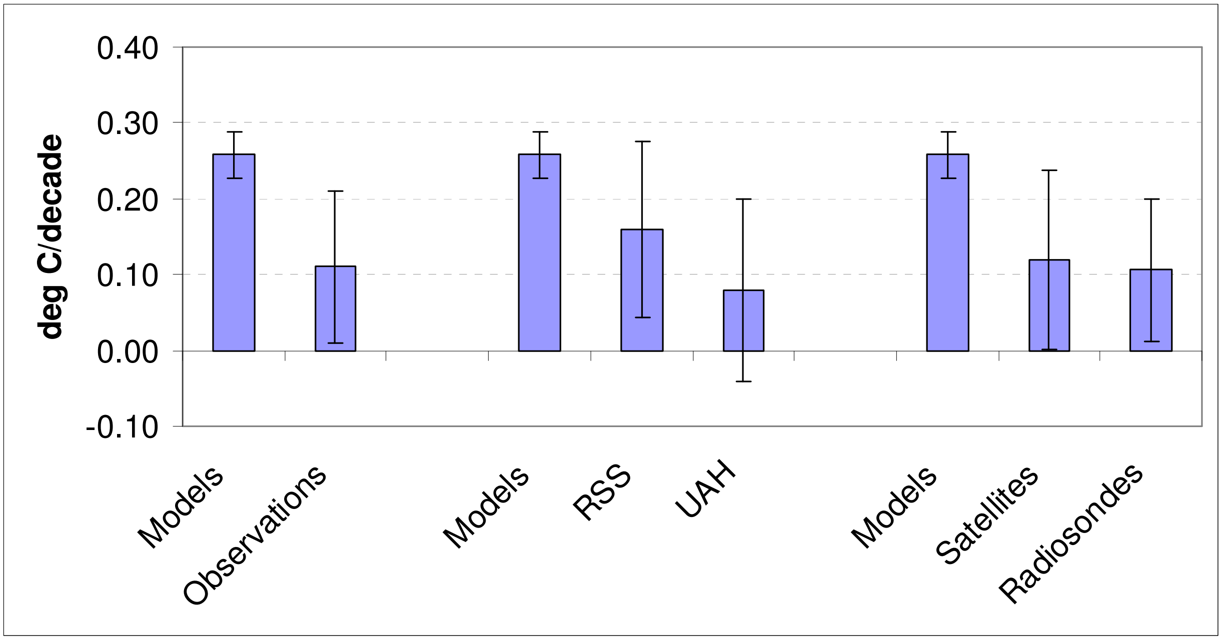


**Figure 2:** Modeled and estimated trends (1979-2009, C decade$^{-1}$) in the tropics, lower troposphere (LT) layer

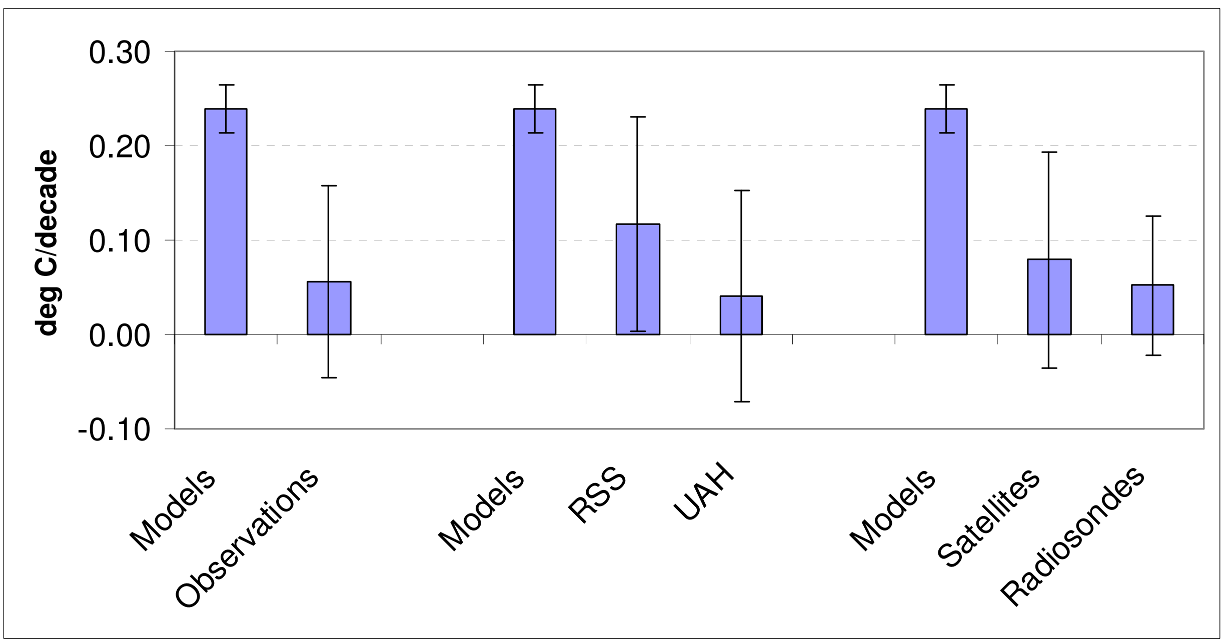


**Figure 3:** Modeled and estimated trends (1979-2009, C decade$^{-1}$) in the tropics, mid-troposphere (MT) layer

| Panel | Model/ Obs Name | Extra Forcings | No. Runs | LT Trend (Std Dev) | MT Trend (Std Dev) | AR Coeffs LT / MT |
|---|---|---|---|---|---|---|
| 1 | BCCR BCM2.0 | O | 1 | 0.210** (0.058) | 0.211** (0.053) | 1,2/ 1 |
| 2 | CCCMA3.1-T47 | NA | 5 | 0.363** (0.021) | 0.380** (0.020) | 1,2,3,5 / 1,3 |
| 3 | CCCMA3.1-T63 | NA | 1 | 0.419** (0.041) | 0.444** (0.039) | 1,6 / 1,6 |
| 4 | CNRM3.0 | O | 1 | 0.258** (0.085) | 0.326** (0.111) | 1,6 / 1,3,6 |
| 5 | CSIRO3.0 | | 1 | 0.162* (0.083) | 0.30 (0.083) | 1,3 / 1,3 |
| 6 | CSIRO3.5 | | 1 | 0.305** (0.103) | 0.288** (0.109) | 1,2,6 / 1,2,6 |
| 7 | GFDL2.0 | O, LU, SO, V | 1 | 0.229** (0.099) | 0.225** (0.104) | 1,6 / 1,6 |
| 8 | GFDL2.1 | O, LU, SO, V | 1 | 0.188 (0.115) | 0.193 (0.126) | 1 / 1,4,5 |
| 9 | GISS_AOM | | 2 | 0.127 (0.091) | 0.123 (0.095) | 1 / 1 |
| 10 | GISS_EH | O, LU, SO, V | 6 | 0.277** (0.047) | 0.261** (0.043) | 1 / 1 |
| 11 | GISS_ER | O, LU, SO, V | 5 | 0.258** (0.065) | 0.230** (0.043) | 1,3,4,6 / 1,4 |
| 12 | IAP_FGOALS1.0 | | 3 | 0.273* (0.037) | 0.259** (0.028) | 1 / 1 |
| 13 | ECHAM4 | | 1 | 0.290** (0.033) | 0.270** (0.028) | 1,4 / 1 |
| 14 | INMCM3.0 | SO, V | 1 | 0.185** (0.076) | 0.186** (0.081) | 1,4,6 / 1,6 |
| 15 | IPSL_CM4 | | 1 | 0.203** (0.077) | 0.202** (0.082) | 1,3,6 / 1,3,6 |
| 16 | MIROC3.2_T106 | O, LU, SO, V | 1 | 0.100 (0.078) | 0.102 (0.084) | 1,6 /1,6 |
| 17 | MIROC3.2_T42 | O, LU, SO, V | 3 | 0.280** (0..037) | 0.284** (0.039) | 1/1 |
| 18 | MPI2.3.2a | SO,V | 5 | 0.277** (0.060) | 0.232** (0.057) | 1,2 / 1,2,6 |
| 19 | ECHAM5 | O | 4 | 0.227** (0.044) | 0.224** (0.045) | 1 / 1 |
| 20 | CCSM3.0 | O,SO,V | 7 | 0.320** (0.050) | 0.285** (0.044) | 1 /1,6 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 21 | PCM_B06.57 | O, SO, V | 4 | 0.178* (0.043) | 0.142** (0.023) | 1,2,3 / 1,2 |
| 22 | HADCM3 | O | 1 | 0.204** (0.060) | 0.186** (0.063) | 1,2,4,6 /1,6 |
| 23 | HADGEM1 | O, LU, SO, V | 1 | 0.258** (0.058) | 0.270** (0.056) | 1 /1 |
| 24 | UAH | | | 0.070 (0.058) | 0.040 (0.062) | 1,2 /1,2 |
| 25 | RSS | | | 0.157** (0.058) | 0.117* (0.065) | 1,2 / 1,2 |
| 26 | HadAT | | | 0.097* (0.053) | 0.020 (0.066) | 1,2 / 1,2 |
| 27 | RICH | | | 0.114** (0.050) | 0.072 (0.059) | 1,2 / 1,2 |

**Table 1: Summary of data series.** Each row refers to model ensemble mean (rows 1—23) or observational series (rows 24—27). All models forced with 20th century greenhouse gases and direct sulfate effects. Rows 10, 11, 19, 22 and 23 also include indirect sulfate effects. 'Extra forcing' column indicates which models included other forcings: ozone depletion (O), solar changes (SO), land use (LU), volcanic eruptions (V). NA: information not supplied to PCMDI. No. runs: indicates number of individual realizations in the ensemble mean. LT and MT trends based on linear regression allowing 6 autoregressive terms. Std errors in brackets. * significant at 10%. ** significant at 5%. AR Coefs: the autoregressive lags that were significant ($p$<0.05) for LT / MT layers respectively.

| | Tests of Trend Significance | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Obs | MSU | UAH | RSS | BAL | HAD | RICH | Models |
| LT | | | | | | | | |
| *VF Method* | | | | | | | | |
| 1979-1999 $F_2^*$ | 12.50 | | 3.98 | 25.47* | | 7.85 | 15.79 | |
| 1979-2009 $F_2^*$ | 76.66** | | 27.92* | 118.79** | | 55.16** | 93.12** | |
| *Panel Method* | | | | | | | | |
| 1979-2009 | | | | | | | | |
| Trend ($^o$C/decade) | 0.110** | 0.120** | 0.079 | 0.159** | 0.105** | | | 0.272** |
| Std Error | 0.050 | 0.059 | 0.060 | 0.058 | 0.047 | | | 0.013 |
| *p* | 0.027 | 0.042 | 0.186 | 0.006 | 0.026 | | | 0.000 |
| MT | | | | | | | | |
| *VF Method* | | | | | | | | |
| 1979-1999 $F_2^*$ | 5.06 | | 1.55 | 19.36 | | 0.27 | 10.08 | |
| 1979-2009 $F_2^*$ | 23.77* | | 6.21 | 62.96** | | 0.26 | 41.43* | |
| *Panel Method* | | | | | | | | |
| 1979-2009 | | | | | | | | |
| Trend ($^o$C/decade) | 0.057 | 0.079 | 0.041 | 0.117** | 0.043 | | | 0.253** |
| Std Error | 0.051 | 0.057 | 0.056 | 0.057 | 0.049 | | | 0.012 |
| *p* | 0.272 | 0.166 | 0.466 | 0.039 | 0.389 | | | 0.000 |

**Table 2:** Trend significance tests using nonparametric covariance estimator on balanced panels and panel regression on unbalanced panels. *VF Method*: Shown are Vogelsang and Franses (2005) $F_2^*$ test scores. 90% critical value is 20.14, 95% critical value is 41.53, 99% critical value is 83.96. *Panel Method* refers to panel regression results. Shown are: the trend in C decade$^{-1}$, the standard error of the trend and the *p* value of a test of H0: trend=0. Top block: LT, or lower troposphere. Bottom block: MT or mid-troposphere. See text for discussion of column groupings. Headings: Obs=average of all observational series; MSU=combined satellite record; UAH=University of Alabama-Huntsville; RSS=Remote Sensing Systems; BAL=combined balloon (radiosonde) series; HAD=HadAT balloon series; RICH=Haimberger balloon series; Models=average of 23 ensemble means. * denotes significant at 10% level, ** denotes significant at 5% level.

| | Tests of Difference from Models | | | | | RSS *vs* | BAL *vs* |
|---|---|---|---|---|---|---|---|
| | Obs | MSU | UAH | RSS | BAL | UAH | MSU |
| LT | | | | | | | |
| *VF Method* | | | | | | | |
| 1979-1999 | 24.96* | | | | | 1990.10** | 4.51 |
| 1979-2009 | 188.55** | | | | | 399.85** | 2.06 |
| *Panel (p)* | | | | | | | |
| *1979-2009* | 0.002** | 0.012** | 0.002** | 0.059* | 0.001** | 0.000** | 0.880 |
| MT | | | | | | | |
| *VF Method* | | | | | | | |
| 1979-1999 | 35.48* | | | | | 1203.37** | 10.18 |
| 1979-2009 | 257.67** | | | | | 229.35** | 13.91 |
| *Panel (p)* | | | | | | | |
| *1979-2009* | 0.000** | 0.003** | 0.000** | 0.019** | 0.000** | 0.000** | 0.243 |

**Table 3:** Trend difference tests using nonparametric covariance estimator on balanced panels and panel regression on unbalanced panels. *VF* group results: Vogelsang and Franses (2005) *F2* test scores, 90% critical value is 20.14, 95% critical value is 41.53, 99% critical value is 83.96. *Panel (p)* refers to panel regression results. Shown are the *p* values of a test of whether indicated trend difference = 0. Top block: LT, or lower troposphere. Bottom block: MT or mid-troposphere. See text for discussion of column groupings. Headings: See Table 2 legend. * denotes significant at 10% level, ** denotes significant at 5% level.