

An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes

Dian J. Seidel

Air Resources Laboratory, NOAA, Silver Spring, Maryland, USA

John R. Lanzante

Geophysical Fluid Dynamics Laboratory, NOAA, Princeton, New Jersey, USA

Received 3 December 2003; revised 3 March 2004; accepted 26 May 2004; published 29 July 2004.

[1] Historical changes in global atmospheric temperature are typically estimated using simple linear trends. This paper considers three alternative simple statistical models, each involving breakpoints (abrupt changes): a flat steps model, in which all changes occur abruptly; a piecewise linear model; and a sloped steps model, incorporating both abrupt changes and slopes during the periods between breakpoints. First- and second-order autoregressive models are used in combination with each of the above. Goodness of fit of the models is evaluated using the Schwarz Bayesian Information Criterion. These models are applied to the instrumental record of global monthly temperature anomalies at the surface and to the radiosonde and satellite records for the troposphere and stratosphere. The alternative models often provide a better fit to the observations than the simple linear model. Typically the two top-performing models have very close values of the Schwarz Bayesian Information Criterion. Usually the two models have the same basic form and the same net temperature change but with a different choice of autoregressive model. However, in some cases the best fits are from two different basic models, yielding different net temperature changes and suggesting different interpretations of the nature of those changes. For the surface data during 1900–2002 the sloped steps and piecewise linear models offer the best fits. Results for tropospheric data suggest that it is reasonable to consider most of the warming during 1958–2001 to have occurred at the time of the abrupt climate regime shift in 1977. Two fundamentally different, but equally valid, descriptions of stratospheric cooling were found: gradual linear change versus more abrupt ratcheting down of temperature concentrated in postvolcanic periods (~ 2 years after eruption). Because models incorporating abrupt changes can be as explanatory as simple linear trends, we suggest consideration of these alternatives in climate change detection and attribution studies. *INDEX TERMS*: 0350 Atmospheric Composition and Structure: Pressure, density, and temperature; 0370 Atmospheric Composition and Structure: Volcanic effects (8409); 1610 Global Change: Atmosphere (0315, 0325); 3309 Meteorology and Atmospheric Dynamics: Climatology (1620); *KEYWORDS*: atmospheric temperature change, abrupt changes, linear trends

Citation: Seidel, D. J., and J. R. Lanzante (2004), An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes, *J. Geophys. Res.*, 109, D14108, doi:10.1029/2003JD004414.

1. Introduction

[2] Identifying, quantifying, and understanding the nature of multidecadal temperature changes at the surface and in the troposphere and stratosphere has been the focus of numerous recent climate change studies and assessments summarized by Folland *et al.* [2001a] and the National Research Council (U.S.) Panel on Reconciling Temperature Observations [2000]. In most cases, temperature changes are statistically modeled using the simplest functional form possible, a straight line (often fit to anomaly data, with the mean annual

cycle removed), and quantified using the linear slope. However, it is widely recognized that global temperatures and other variables have not experienced monotonic linear changes and that the data exhibit both piecewise linear behavior and step-like changes [e.g., Karl *et al.*, 2000; Lindzen and Giannitsis, 2002; Christiansen, 2003; Lanzante *et al.*, 2003; Tomé and Miranda, 2004; Rodionov, 2004].

[3] Our exploration of nonmonotonic alternatives to characterize global temperature variations of the twentieth century is based on both empirical and theoretical considerations. In the troposphere an abrupt rise in temperature in ~ 1976 or 1977 and concomitant changes in global climate variables have been documented [Trenberth and Hurrell, 1994], and earlier regime changes have been suggested

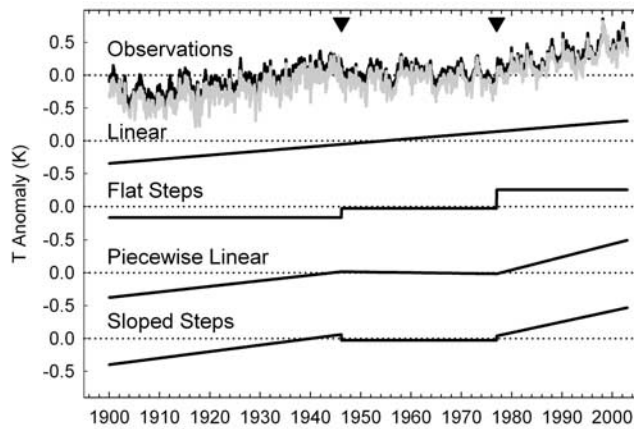


Figure 1. Global monthly surface temperature anomalies from the University of East Anglia (UEA (black)) and Global Historical Climatology Network (GHCN (grey)) data sets. Below the data are model fits to the average of the two data sets from four statistical models: linear, flat steps, piecewise linear, and sloped steps. The last three incorporate breakpoints, shown by the triangles at 1946 and 1977.

[Zhang *et al.*, 1997]. Although the definitive cause(s) are not known, current working theories have suggested that this phenomenon is a manifestation of internal climate variability. Global climate models including both natural and anthropogenic external forcings are able to simulate multidecadal periods of quasi-linear change in global surface temperature, with differing trends during different periods, including one period with little or no change [Delworth and Knutson, 2000; Zwiers and Weaver, 2000; Stott *et al.*, 2000; Hansen *et al.*, 2002; Meehl *et al.*, 2003; Broccoli *et al.*, 2003]. Step-like temperature changes have been observed in the stratosphere [Pawson *et al.*, 1998; Randel *et al.*, 2000; Chipperfield and Randel, 2003], although the largest forcings, associated with greenhouse gas increases and stratospheric ozone loss, are, if not linear, at least quasi-monotonic.

[4] Given the plausibility of temperature changes whose temporal behavior is more complex than a simple linear model, it is reasonable to explore other statistical models. In this study we consider several alternative models and demonstrate that in many cases they fit the data better than a simple linear trend. The alternative models can then be used to provide a somewhat different description of the low-frequency behavior of temperature and a somewhat different estimate of the overall change in temperature over the time period of consideration.

[5] Section 2 describes six global air temperature data sets. Section 3 outlines the four methods used here to model those data and describes the use of the Schwarz Bayesian Information Criterion to compare the model fits. Section 4 presents results from which conclusions are drawn in section 5.

2. Data

[6] The data used in this study are monthly global temperature anomaly time series from which attempts have been made to remove artificial temporal inhomogeneities. For the surface, and for each of five atmospheric layers, two

independently developed time series are used in the current analysis. The two data sets are averaged together to capture only their more robust features.

2.1. Surface Temperature

[7] Global surface temperature anomalies, for the period 1900–2002, are from two sources, both of which combine surface air temperatures over land and sea surface temperatures to form global data sets, but using different station networks, homogeneity adjustments, and data processing methods. Figures 1 and 2 show surface temperature anomaly data from the University of East Anglia/Climatic Research Unit [Folland *et al.*, 2001b; Jones *et al.*, 1999; Jones and Moberg, 2003] and from the National Oceanic and Atmospheric Administration (NOAA)/National Climatic Data Center's Global Historical Climatology Network (GHCN, version 2) [Peterson and Vose, 1997]. Because the two time series track each other closely, the average of the two faithfully represents the temperature changes in each of these two widely used data products.

2.2. Upper Air Temperature From Radiosondes

[8] Upper air temperatures for the period 1958–2001, for the tropospheric 850–300 hPa layer (Figure 3) and for the stratospheric 100–50 hPa layer (Figure 4), are from two radiosonde data sets. The UK Met Office's Hadley Center for Climate Prediction and Research HadRT2.1s data set is based on observations from radiosonde stations providing monthly temperature (CLIMAT TEMP) reports. Data at the 850, 700, 500, and 300 hPa levels were used to create 850–300 hPa layer mean anomalies, and data at the 100 and 50 hPa levels were used for the 100–50 hPa layer. Stratospheric data since 1979 have been adjusted globally using satellite data, but only at those stations where significant temperature changes were accompanied by known station

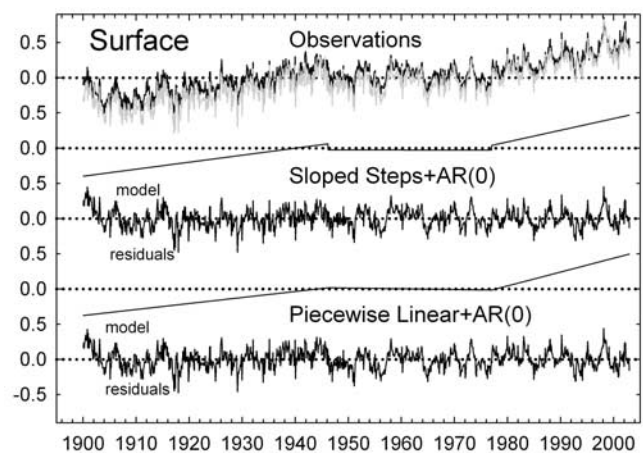


Figure 2. Global monthly surface temperature anomalies from the UEA (black) and GHCN (grey) data sets and the two best model fits to the average of the two data sets, selected on the basis of the Schwarz Bayesian Information Criterion and normally distributed residuals. The modeled time series and the residuals (observations minus model) are shown for the sloped steps model that does not include autoregressive behavior in the residuals (AR(0)) and for the piecewise linear model with AR(0).

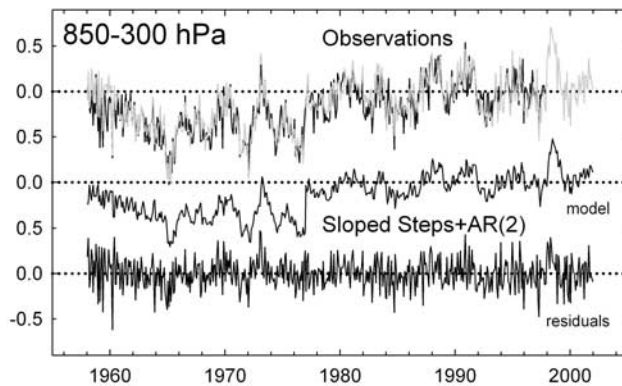


Figure 3. Global monthly 850–300 hPa temperature anomalies from HadRT2.1s (grey) and Lanzante-Klein-Seidel (LKS (black)) data. The best model fit to the average of the two data sets, sloped steps with second-order autoregressive models (AR(2)), and the residuals (observations minus model) are shown below the observations.

history events, using the method described by *Parker et al.* [1997]. The Lanzante-Klein-Seidel data set is based on 87 radiosonde stations with data adjustments for temporal homogeneity at all levels for the period 1948–1997 [Lanzante et al., 2003]. As with the HadRT2.1s data, layer means are based on pressure-level data, but, in this case, 400 and 70 hPa data were employed in addition to the levels mentioned above. The radiosonde network does not fully sample the globe, with notable gaps over ocean regions and in the Southern Hemisphere, so the global estimates, based on averaged zonal mean anomalies, are more representative of land areas in these two data products. As with the surface data the two independently derived data sets are highly correlated [Seidel et al., 2004], and we use the average of the two for the period 1958–1997, but we use only HadRT2.1s data for 1998–2001.

2.3. Upper Air Temperature From Satellite-Borne Microwave Sounding Unit

[9] For the period 1979–2001, microwave sounding unit (MSU) data from NOAA polar-orbiting satellites provide much better spatial coverage and horizontal resolution than radiosonde data. The data for channel 2 (shown in Figure 5) are mainly sensitive to tropospheric temperature, while the lower stratosphere is sampled by channel 4 (Figure 6). The two data sets employed here are from the University of Alabama in Huntsville (UAH) [Christy et al., 2003] and from Remote Sensing Systems, Inc. (RSS) [Mears et al., 2003]. Both groups prepare monthly, gridded, global temperature anomaly data, but they have different methods of creating homogeneous time series. Nevertheless, the global data sets are extremely highly correlated on both monthly and interannual timescales [Seidel et al., 2004], and, as with the in situ data, we examine the average of the two data sets.

3. Method

3.1. Statistical Models

[10] For the surface and for each of four layers (850–300, 100–50 hPa, microwave sounding unit channel 2 (MSU2),

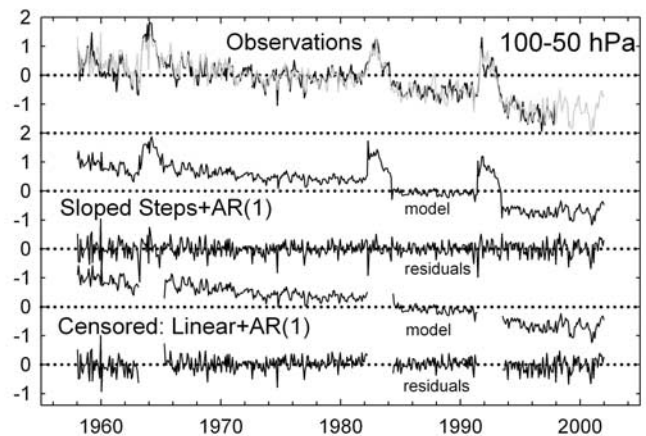


Figure 4. Global monthly 100–50 hPa temperature anomalies from HadRT2.1s (grey) and LKS (black) data. The best model fit to the average of the two data sets, sloped steps with a first-order autoregressive model (AR(1)), and the residuals (observations minus model) are shown below the observations. Also shown is the best model fit for the censored data, linear with AR(1), in which observations for the three 2 year periods following major volcanic eruptions are not included in the fitting procedures.

and microwave sounding unit channel 4 (MSU4)), the two-data set average monthly global temperature anomaly time series is modeled using four very simple mathematical forms, three of which employ breakpoints. The criteria for selecting the breakpoints are outlined below.

[11] As an example, Figure 1 shows the four model fits to the surface data, with breakpoints selected in 1946 and 1977. The first is a linear fit to the data for the full period of record. The other three models involve fitting linear seg-

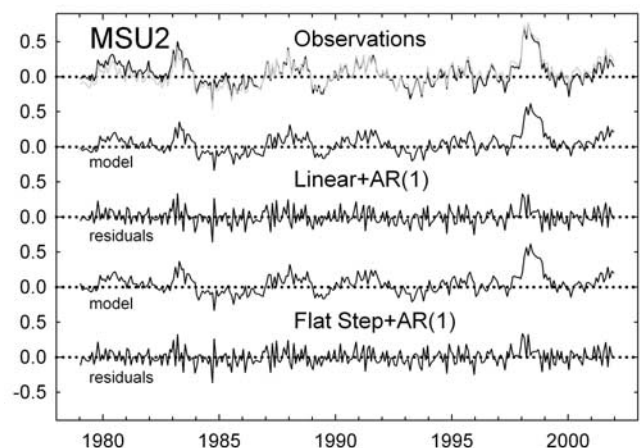


Figure 5. Global monthly temperature anomalies for 1979–2001 from microwave sounding unit (MSU) channel 2 from the University of Alabama in Huntsville (UAH (black)) and Remote Sensing Systems, Inc. (RSS (grey)) data sets. The two best model fits to the average of the two data sets (linear with AR(1) and flat step with AR(1)) and the residuals (observations minus model) are shown below the observations.

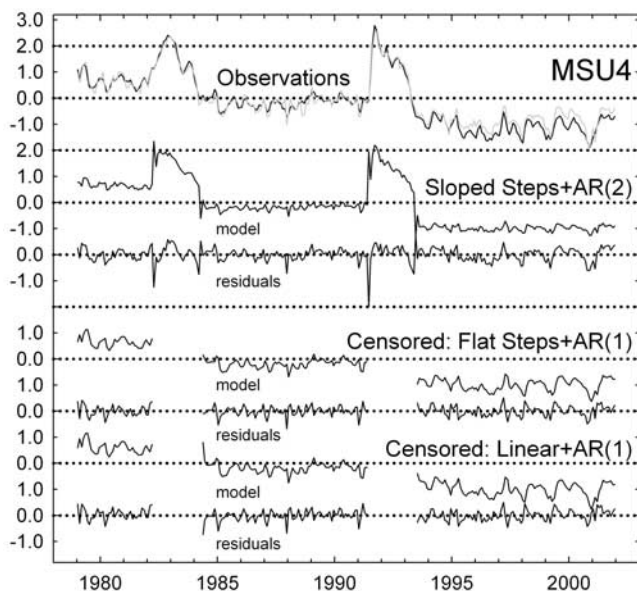


Figure 6. Global monthly MSU channel 4 temperature anomalies from RSS (grey) and UAH (black) data. The best model fit to the average of the two data sets, sloped steps with AR(2), and the residuals (observations minus model) are shown below the observations. Also shown are the two best model fits for the censored data (flat steps with AR(1) and linear with AR(1)), in which the observations for the two 2 year periods following major volcanic eruptions are not included in the fitting procedures.

ments to the data between the breakpoints. Thus the second model is a series of nonsloping segments interrupted by instantaneous upward or downward level shifts at each breakpoint. The third model is a piecewise linear function, with the number of linear segments being one greater than the number of breakpoints, N_b . The fourth model resembles the second, but the segments between the level shifts may have nonzero linear slopes. For brevity, the four models are termed linear, flat steps, piecewise linear, and sloped steps in the remainder of the paper. For all but the flat steps model, the linear slopes are determined using least squares regression, with constraints on the intercepts in the piecewise linear model to ensure that the data segments connect at the breakpoints. In the flat steps model the level shifts at the breakpoints are based on differences in mean temperature anomalies between the segments preceding and following the breakpoint. We augment each of the four basic models with autoregressive components, extensively used in previous analyses of atmospheric data, which tend to be highly temporally autocorrelated. We use both first- and second-order autoregressive models, AR(1) and AR(2), as discussed in section 3.2.

[12] The breakpoints were initially identified by visual inspection of the time series and then refined. The guiding principles in selecting breakpoints are as follows: (1) minimization of the number of breakpoints so as to maximize the simplicity of the fitting model, (2) consistency between the breakpoints used in this study and those identified in previous research, and (3) support from an objective, nonparametric statistical method developed

by Lanzante [1996]. At the surface, breakpoints were identified at January 1946, in accord with Folland *et al.* [2001a], and January 1977, 1 year after the breakpoint in Folland *et al.* [2001a] and Karl *et al.* [2000]. In the 850–300 hPa layer a single breakpoint in January 1977 was selected, and none were selected for the MSU2 layer (where the data begin in 1979). The 1977 breakpoint nominally identifies a climate shift apparent in the Pacific Ocean [Miller *et al.*, 1994] and in many other features of the climate system [Trenberth, 1990; Trenberth and Hurrell, 1994]. For the stratospheric layers (100–50 hPa and MSU4), breakpoints were chosen at the months of three major volcanic eruptions (when temperatures rise abruptly) and 2 years following the eruption (when they fall, albeit somewhat less abruptly). The eruptions were Mt. Agung in March 1963, El Chichón in April 1982, and Mt. Pinatubo in June 1991.

3.2. Comparison of Models Using the Schwarz Bayesian Information Criterion

[13] To evaluate the four models, we employ a modified version of the Schwarz Bayesian Information Criterion [Schwarz, 1978; Priestley, 1981]. The criterion, $S(q)$, provides an objective and quantitative method of comparing the statistical models and is evaluated as follows:

$$S(q) = n_{\text{eff}} \log \left\{ \frac{1}{n} \sum_{t=1}^n [T_{\text{obs}}(t) - T_{\text{mod}}(t)]^2 \right\} + q \log(n_{\text{eff}}), \quad (1)$$

where t is time, n is the number of data points (in this case months) in the time series, $T_{\text{obs}}(t)$ is the observed temperature anomaly time series, and $T_{\text{mod}}(t)$ is the modeled time series. The first term on the right hand side of equation (1) is proportional to the mean square error of the residuals and thus measures how well the model replicates the data. The second term is a penalty factor based on the number of fitting parameters, q , in the model and so allows $S(q)$ to reward a model's parsimony. Thus the lower the value of $S(q)$, the better the model fit.

[14] Our implementation of $S(q)$ substitutes the effective sample size, n_{eff} , for n , in an attempt to account for the effects of temporal coherence. We estimate n_{eff} as

$$n_{\text{eff}} = n(1 - \rho_1)/(1 + \rho_1) \quad (2)$$

to account for the fact that each monthly T_{obs} value does not represent an independent observation, owing to the strong lag-one autocorrelation ρ_1 [Laurmann and Gates, 1977]. We compute ρ_1 for the residuals (observations minus model) from each of the simple model fits so that each model is evaluated based on n_{eff} appropriate to that model.

[15] The number of fitting parameters, q , depends on the model. For the linear model, $q_{\text{linear}} = 2$, the slope and intercept of the line. For the other models, q depends on the number of breakpoints, N_b , each of whose location must be specified. For the flat steps model, there are N_b breaks (each of which must be specified), defining $N_b + 1$ segments, for which mean levels must be determined, so

$$q_{\text{flat steps}} = N_b + (N_b + 1) = 2N_b + 1. \quad (3)$$

For the piecewise linear model we specify the locations of the breakpoints, the slope of each segment, and the intercept of the first segment, so that

$$q_{\text{piecewise linear}} = N_b + 1(N_b + 1) + 1 = 2N_b + 2. \quad (4)$$

The least parsimonious of our four models is the sloped steps model, in which the locations of the breaks, and the slopes and intercepts of each segment must be specified:

$$q_{\text{sloped steps}} = N_b + 2(N_b + 1) = 3N_b + 2. \quad (5)$$

If we go beyond these simple linear fits to model the autoregressive behavior of the residuals, we must augment q by one for an AR(1) process in which

$$T' \equiv T_{\text{obs}}(t) - T_{\text{mod}}(t) = \rho_1 T'(t-1), \quad (6)$$

or we must augment q by two for an AR(2) process where

$$T'(t) = a_1 T'(t-1) + a_2 T'(t-2), \quad (7)$$

where

$$a_1 = \rho_1(1 - \rho_2)/(1 - \rho_1^2) \quad (8)$$

$$a_2 = (\rho_2 - \rho_1^2)/(1 - \rho_1^2), \quad (9)$$

and ρ_2 is the lag-two autocorrelation.

[16] To test whether a given model's residuals are normally distributed (and that the breakpoints do not introduce spikes in the time series of residuals), we assess the goodness of fit of the residuals to a Gaussian distribution, both with removal of the AR(1) and AR(2) behavior and with a model that does not include autoregressive behavior in the residuals (AR(0)), using the Anderson-Darling statistic [Anderson and Darling, 1954] and test the null hypothesis of normally distributed residuals. We eliminate from further consideration any model for which the null hypothesis is rejected at the 1% significance level.

[17] In summary, we select the best model as the one with the lowest value of $S(q)$, which ensures an optimal balance of small mean square error and model parsimony, provided that the residuals are normally distributed. As shown in section 4, in many cases the difference between the lowest and second-lowest values of $S(q)$ was very small, so we present both models as reasonable choices.

4. Results

4.1. Surface Temperature During 1900–2002

[18] Figures 7a and 8a show the RMS error of the residuals, and the Schwarz Bayesian Information Criterion, respectively, for the model fits to the surface temperature time series for 1900–2002 as represented by the average of the UEA and GHCN data sets (Figures 1 and 2). Without considering the autoregressive behavior of the residuals (AR(0) (light grey)), the piecewise linear and sloped steps models result in lower RMS errors than the linear or flat steps models. Incorporating either AR(1) or AR(2)

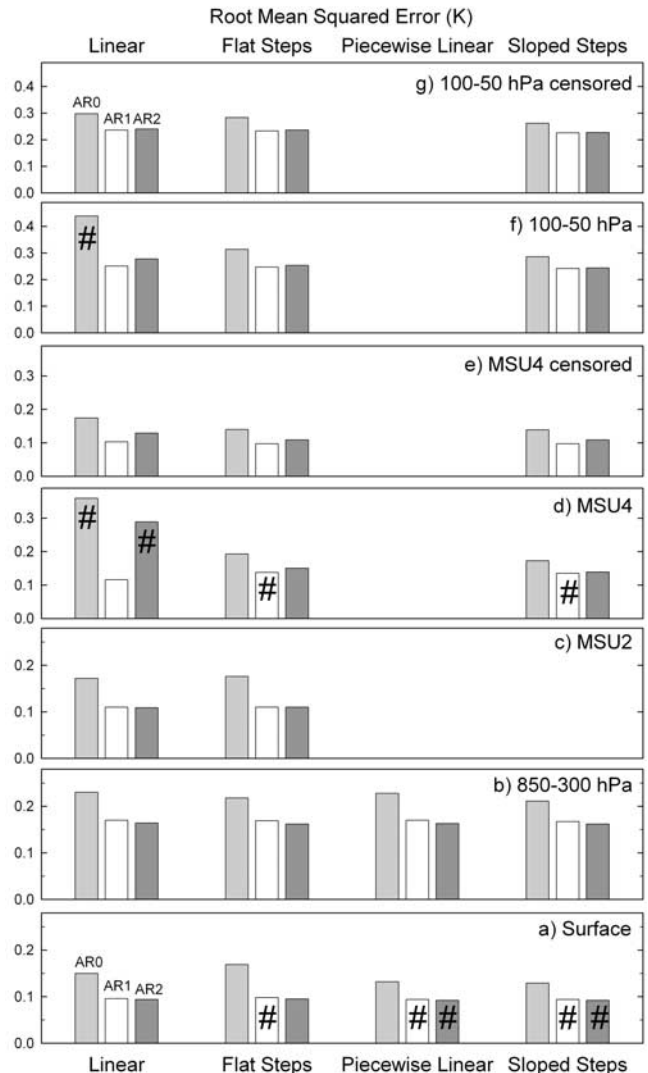


Figure 7. Root-mean-square error (K) of the model fits to the observations for (a) surface, (b) 850–300 hPa, (c) MSU channel 2, (d) MSU channel 4, (e) MSU channel 4 data censored to disregard data for 2 years following major volcanic eruptions, (f) 100–50 hPa, and (g) 100–50 hPa censored as in Figure 7e. For each of the four simple models (linear, flat steps, piecewise linear, and sloped steps) the residuals are either not modeled at all (AR(0) (light grey)) or are modeled as autoregressive processes (AR(1) (white), AR(2) (dark grey)). The number symbols indicate non-Gaussian distributions of the residuals.

processes results in reducing the RMS error by about one third to 0.09 K for all four models, although in some cases this results in residuals that are not normally distributed at the 1% level (indicated by the number symbol in Figures 7 and 8). The lowest $S(q)$ value for the surface data is for the sloped steps model with AR(0), the second-lowest value is for the piecewise linear model with AR(0), and the difference in $S(q)$ is 4.3% (Figure 8a and Table 1).

[19] Figure 2 depicts both of these model fits and the time series of the residuals. The “steps” in the sloped steps model are actually very small (−0.09 K in 1945 and 0.07 K

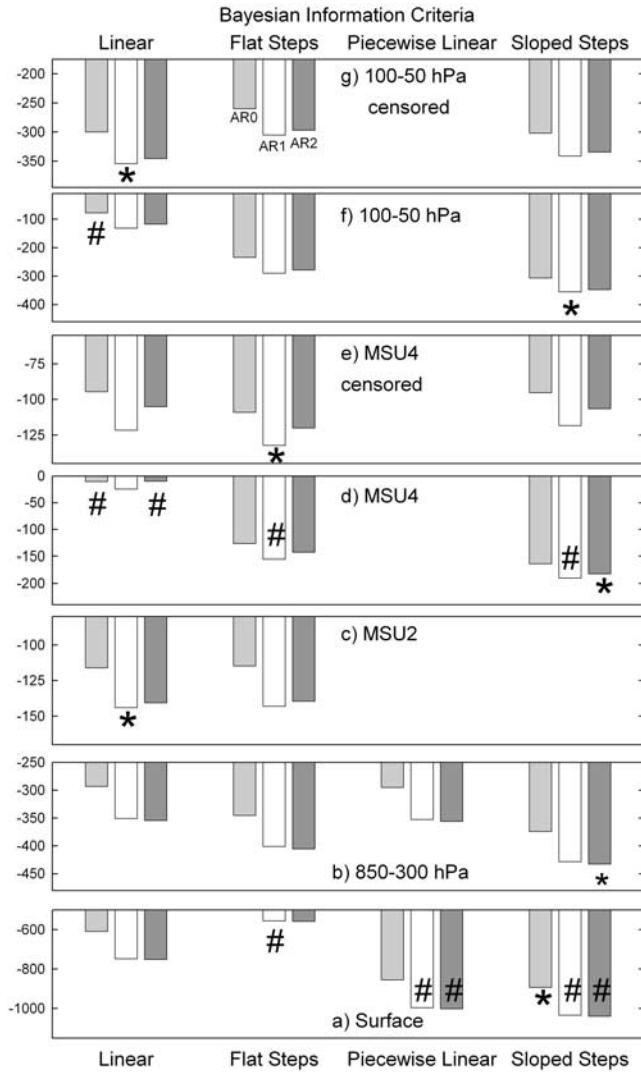


Figure 8. As in Figure 7, but for the Schwarz Bayesian Information Criterion, $S(q)$. Lower values of $S(q)$ indicate a better model. The best model for each level is identified with an asterisk.

in 1977), and most of the temperature change in that model is due to the increases during 1900–1945 and during 1977–2002. Thus, in this case, the sloped steps and piecewise linear models are very similar, and both yield a net warming of 0.87 K (Table 1). In comparison, our linear model, with a slope of 0.063 ± 0.007 K/decade, yields a net warming of 0.66 K (Table 1), 31% smaller than the warming in the two best fit models. (The uncertainty in linear slopes is reported as ± 1 standard deviation of the slope estimate, based on the residuals (observations minus linear model fit) without taking into account autoregressive behavior in the residuals.) The slopes of the three segments of the sloped steps model fit are, in chronological order, 0.100, -0.001 , and 0.166 K/decade. These results affirm the piecewise linear approach taken by *Karl et al.* [2000] and *Folland et al.* [2001a], with an increased rate of warming since 1977 than in previous decades.

4.2. Tropospheric Temperature From Radiosondes During 1958–2001

[20] As shown in Figure 7b the incorporation of AR(1) and AR(2) modeling reduces the RMS error of the residuals in the 850–300 hPa radiosonde temperature anomalies for 1958–2001, compared with the AR(0) versions of our four basic models, and these reductions are larger than the differences in RMS among the four basic models. The lowest $S(q)$ value is for the sloped steps model with AR(2), and the second-lowest value is for the same model with AR(1), so Figure 3 shows only the best fitting model and its residuals. The model yields trends of -0.107 and 0.072 K/decade before and after the 1977 breakpoint, respectively, and an upward jump of 0.35 K at the breakpoint, yielding a net warming of 0.32 K. These results suggest that it is reasonable to propose that much of the tropospheric warming during 1958–2001 occurs at the 1977 breakpoint, as *Lindzen and Giannitsis* [2002] argue, and as suggested by *Gaffen et al.* [2000] for the tropical troposphere. In comparison with the sloped steps model the net warming from the linear model (with a trend of 0.117 ± 0.022 K/decade) is 0.52 K, 62% larger (Table 1).

4.3. Stratospheric Temperature From Radiosondes During 1958–2001

[21] The record of 100–50 hPa temperatures from radiosondes (Figure 4) shows cooling during 1958–2001, punc-

Table 1. Best Statistical Models to Describe Global Temperature Changes for Each Atmospheric Level^a

Level	Censored	Period	ΔT_{lin}	N_b	Best Model	ΔT_1	Second-Best Choice	ΔT_2	$\Delta S, \%$
Surface	no	1900–2002	0.66	2	Sloped Steps plus AR(0)	0.87	Piecewise Linear plus AR(0)	0.87	4.3
850–300 hPa	no	1958–2001	0.52	1	Sloped Steps plus AR(2)	0.32	Sloped Steps plus AR(1)	0.32	0.9
100–50 hPa	no	1958–2001	-1.81	6	Sloped Steps plus AR(1)	-1.82	Sloped Steps plus AR(2)	-1.82	2.1
100–50 hPa	yes	1958–2001	-1.9	6	Linear plus AR(1)	-1.90	Linear plus AR(2)	-1.90	2.6
MSU2	no	1979–2001	0.13	0	Linear plus AR(1)	0.13	Flat Step plus AR(1)	0.00	0.7
MSU4	no	1979–2001	-1.13	4	Sloped Steps plus AR(2)	-0.88	Sloped Steps plus AR(0)	-0.88	10.5
MSU4	yes	1979–2001	-0.99	4	Flat Steps plus AR(1)	-0.83	Linear plus AR(1)	-0.99	8.1
850–300 hPa	no	1979–2001	0.14	0	Linear plus AR(1)	0.14	Linear plus AR(2)	0.14	0.3
100–50 hPa	no	1979–2001	-1.67	4	Sloped Steps plus AR(1)	-1.18	Sloped Steps plus AR(2)	-1.18	3.9
100–50 hPa	yes	1979–2001	-1.48	4	Linear plus AR(1)	-1.48	Linear plus AR(2)	-1.48	5.7

^aAlso shown are second-best statistical models and net temperature change (K) from each (ΔT_1 and ΔT_2) over the data period. Results are based on the Schwarz Bayesian Information Criterion, $S(q)$. Also shown are the net temperature change (K) from the linear model (ΔT_{lin}), the number of breakpoints (N_b) identified in the time series, and the percentage difference in $S(q)$ between the two best models ΔS . If “volcanically perturbed” periods have been

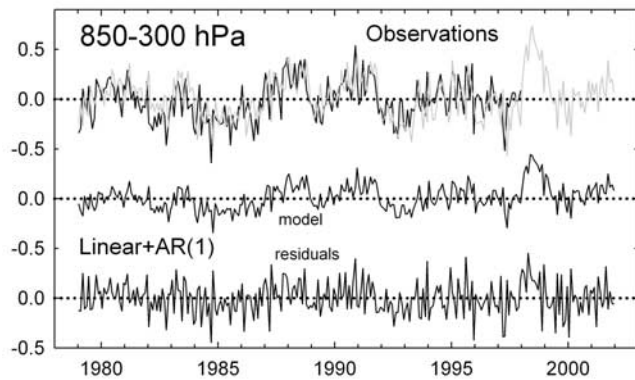


Figure 9. The same as in Figure 3, but only considering the radiosonde data for 1979–2001, in which case the best fitting model is linear with AR(1).

tuated by abrupt warming (lasting ~ 2 years) associated with major volcanic eruptions. This behavior is best modeled by the sloped steps model with either AR(1) or AR(2). The $S(q)$ values for these models differ by only 2.1% (Table 1 and Figure 8f), and Figure 4 shows sloped steps with AR(1).

[22] The sloped steps model yields a net cooling of 1.82 K (almost identical to the linear model), but it is mostly accomplished during the three 2 year periods following volcanic eruptions. Specifically, the slopes (trends) during the non-volcanically perturbed periods are generally small and quite varied: -1.089 K/decade before the 1963 Mt. Agung eruption, -0.227 K/decade between 1965 and the 1982 El Chichón eruption, -0.099 K/decade between 1984 and the 1991 Mt. Pinatubo eruption, and -0.031 K/decade from 1993 to 2001. The net temperature change associated with the three eruptions is more significant. For each eruption an abrupt upward temperature shift was followed by a 2 year period of strong cooling (~ 0.5 to 0.9 K/decade) and an abrupt downward shift, yielding a net 2 year cooling of 0.21 K from Mt. Agung, 0.33 K from El Chichón, and 0.60 K from Mt. Pinatubo. The combined effect of the three volcanoes accounts for 63% of the net 1958–2001 cooling. For each eruption the initial warming is smaller than the subsequent cooling, which leads to a “ratcheting down” of stratospheric temperature, as suggested by *Pawson et al.* [1998].

[23] Because of the relatively short periods of volcanically perturbed conditions in the stratosphere and the large and abrupt temperature increases associated with the eruptions, the piecewise linear model is not suited to the 100–50 hPa and MSU4 data and was not applied. Therefore no results for that model appear in Figures 7 or 8 for the stratospheric levels. However, when we “censor” the observational data and do not include the observations during the 2 year volcanically perturbed periods, we are able to assess whether a single linear trend, a series of flat steps, or a series of segments with potentially different trends (similar to the piecewise linear model) provides the best fit. When the data are censored, the best models are linear, with AR(1) and AR(2), with a very similar net cooling (1.90 K) as when the full time series is considered (Figures 4 and 8 and Table 1).

[24] Although it is beyond the scope of this paper to speculate on physical mechanisms that might cause temper-

atures to behave in the manner shown by the statistical models, we note that a gradual cooling, as given by the linear model, suggests a gradual forcing, such as long-term changes in the gaseous composition of the atmosphere. Stratospheric ozone loss and increases in well-mixed greenhouse gases [*Ramaswamy et al.*, 2001; *Hansen et al.*, 2002; *Hare et al.*, 2004; *Shine et al.*, 2003], or gradually increasing stratospheric water vapor [*Forster and Shine*, 2002], have been suggested as causes of the observed cooling. A sloped steps model, however, suggests punctuated forcing associated with the volcanic eruptions, such as short-lived elevated concentrations of volcanic aerosols or perhaps of water vapor [*Joshi and Shine*, 2003], and related perturbed chemistry or radiative effects.

4.4. Tropospheric Temperature From Satellites and Radiosondes During 1979–2001

[25] As can be seen in Figure 5, the MSU channel 2 data for 1979–2001 have no obvious breakpoints and so cannot easily be divided into two or more segments. Therefore the only models used to fit these data are the linear model (which yields a slope of 0.053 ± 0.002 K/decade and a net temperature increase of 0.13 K, Table 1) and the flat steps model, which is essentially the linear model with zero slope. The linear model with AR(1) yields a lower $S(q)$ value than the flat step model, although the difference is a paltry 0.7% (Table 1 and Figure 8c).

[26] Examining the radiosonde data for the same time period (Figure 9 and Table 1), we find the linear models (with AR(1) and AR(2)) provide the best fits and yield a net warming of 0.14 K, in good agreement with the 0.13 K warming in the MSU2 data. Unlike the other time series discussed in sections 4.1–4.3 and in section 4.5, in which long-term (multidecadal) signals are dominant, the chief temperature variations in the MSU2 and radiosonde records for 1979–2001 are on interannual timescales and are best explained by an AR(1) process (Figures 7c and 8c) with or without an explicit trend.

4.5. Stratospheric Temperature From Satellites and Radiosondes During 1979–2001

[27] The RSS and UAH stratospheric data (Figure 6) are in better accord and have much better global sampling than the radiosonde-based stratospheric data (Figure 4). Therefore they may offer a more reliable perspective on the manner in which the global stratosphere has cooled since 1979. The best models are the sloped steps with AR(2) or AR(0) (Figures 6 and 8, Table 1), with a net cooling of 0.88 K. This is 22% less than the 1.13 K cooling in the linear model (with a slope of -0.470 ± 0.383 K/decade).

[28] Even more prominently than with the radiosonde stratospheric data (Figure 4) the cooling in the sloped steps model for MSU4 data is a result of ratcheting down of temperature following volcanic eruptions. In this case, El Chichón is associated with a net cooling of 0.40 K, and Mt. Pinatubo is associated with a net cooling of 0.43 K, which combine to explain 94% of the total cooling. The modeled trend between the two eruptions is $+0.085$ K/decade (a slight warming), and post-Pinatubo it is only -0.006 K/decade.

[29] When we censor the MSU4 data and remove from consideration the periods following the El Chichón and Mt.

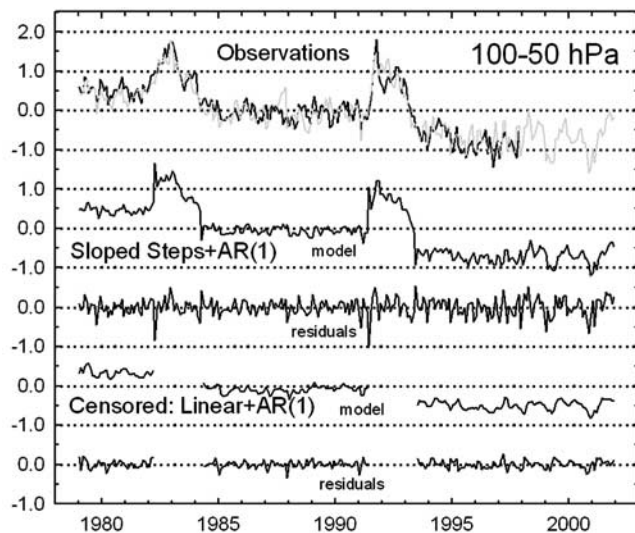


Figure 10. The same as Figure 4, but only considering the radiosonde data for 1979–2001.

Pinatubo eruptions, we find the two best models to be flat steps with AR(1) and linear with AR(1) (Table 1 and Figure 6). These two models yield net coolings of 0.83 and 0.99 K, respectively (Table 1). In the flat steps case the lack of trend between eruptions is inherent in the model, whereas the linear model achieves all the cooling gradually during those same periods. Thus these two choices, with $S(q)$ differing by 8.1%, offer fundamentally different views of the nature of recent stratospheric cooling.

[30] The 100–50 hPa radiosonde data over the same 1979–2001 time period (Figure 10 and Table 1) are best modeled with the sloped steps model with AR(1) or AR(2) for the uncensored case and linear with AR(1) or AR(2) for the censored case. The net cooling in all these cases is larger than seen in the MSU4 data (Table 1), consistent with the findings of Seidel *et al.* [2004]. This might be related to the vertically broader layer sampled by MSU channel 4, which extends both above and below the 100–50 hPa layer, or to the poorer spatial sampling of the sonde data, or to uncorrected biases in one or more data sets. For the sloped steps model applied to the uncensored data, the net cooling is again mainly associated with the eruptions, with El Chichón and Mt. Pinatubo contributing 0.50 and 0.60 K, respectively, to the overall 1.18 K temperature decrease. This is 0.49 K less than the net cooling of 1.67 K obtained from the linear model, which again underscores the sensitivity of net temperature change to model choice.

5. Conclusions

[31] Monthly global temperature anomaly time series for the surface, for two tropospheric layers, and for two stratospheric layers were modeled using four simple statistical models incorporating linear slopes and instantaneous step changes. The breakpoints were selected to be as few as possible while including step-like changes already documented in the literature, and their exact timing was determined using an objective statistical method. The Schwarz Bayesian Information Criterion ($S(q)$, which takes into

account both the reduction in mean square error and the number of fitting parameters) was used to determine which models provided the best fit to the observations, with the provision that the residuals (observations minus model) be normally distributed, as determined by the Anderson-Darling test. The main conclusions of this analysis are as follows:

[32] 1. Accounting for the autoregressive behavior of the residuals, with an AR(1) or AR(2) model, has the greatest effect in reducing the mean square error of the residuals and $S(q)$.

[33] 2. In six of the ten cases examined (including data for five different regions in the vertical, considering two time periods for the radiosonde data, and considering both complete and censored stratospheric data), the best fitting models were not a simple linear fit but involved breakpoints. Two of the four cases in which linear models were chosen were for the troposphere during 1979–2001, a period in which we identified no potential breakpoints.

[34] 3. The frequent choice of the sloped steps and flat steps models highlights the importance of abrupt changes. The choice of the sloped steps model for the tropospheric data suggests it is reasonable to consider most of the warming during 1958–2001 to have occurred at the time of the climate “regime shift,” modeled here at the start of 1977.

[35] 4. For the satellite data period, 1979–2001, linear models are the best choices for both the MSU2 and radiosonde 850–300 hPa data, and the flat step model is a reasonable alternative in the MSU2 case. Since we identified no breakpoints in this period, these models essentially suggest either gradual warming or (for the flat step model for MSU2) no discernible temperature trend. The linear model yields a 0.13 K warming from MSU2 and 0.14 K from radiosondes.

[36] 5. The sloped steps models provide the best fits to stratospheric radiosonde and MSU4 data when the full time series are considered. When the data are censored to eliminate the 2 year periods following major volcanic eruptions, linear models are a better fit. These two models suggest fundamentally different pathways of stratospheric cooling: gradual change versus ratcheting down of temperature owing to volcanic warming and postvolcanic cooling of greater magnitude over the 2 years following the eruption, with little change during the periods between eruptions.

[37] 6. The net temperature change is sensitive to the choice of model fit. Our best fit models yield more surface warming, less tropospheric warming, and generally less stratospheric cooling than simple linear fits.

[38] Although we have combined several statistical techniques to select models in an objective fashion, the reader should be aware that our approach involves some underlying assumptions and is based on some subjective decisions. First, the theory underpinning autoregressive modeling presumes statistical stationarity in the data. (A random variable, or a random process, is said to be stationary if all its statistical parameters are independent of time [Von Storch and Zwiers, 1999].) To meet this requirement, we first fit nonstationary models (linear, flat steps, piecewise linear, or sloped steps) and then apply autoregression to the residuals from the models. Second, there are alternatives to

the Schwarz Bayesian Information Criterion $S(q)$ for selecting models. For example, the widely used Akaike Information Criterion tends to select more complex models than the Bayesian [Priestley, 1981]. Nevertheless, $S(q)$ identified the more complex models in our study fairly often, rather than the simpler linear model, which suggests these results may be robust to the choice of information criterion. Third, we have substituted an effective sample size in our implementation of $S(q)$. Although such a substitution is commonly used, some cautions have been raised by Thiebaut and Zwiers [1984]. Although there is no unique approach, the above caveats notwithstanding, we believe that our implementation is reasonable and yields plausible alternatives to the simple linear trend perspective. This does not discount the possible validity of even more complex models [e.g., Woodward and Gray, 1995]. However, the relative simplicity of the models used here is an advantage in that they may be more amenable to physical interpretation than more complex models.

[39] Because the models incorporating breakpoints provide reasonable alternatives to linear trends, climate change detection and attribution studies could consider the possibility of using these alternative statistical models, or similar constructs, to compare observations with climate model simulations. In the stratospheric case it seems worthwhile to try to separate the effects of step-like changes associated with volcanic eruptions from gradual changes, which may have different causes. On the other hand, the ratcheting down of stratospheric temperatures may suggest a complex interaction among different processes affecting temperature. For the troposphere the nature of the modeled 1977 upward shift in temperature is unclear, and we suggest that attempts to attribute the warming over the past half century to natural or anthropogenic effects should consider the sloped or flat steps model as well as the traditional simple linear model to describe the change.

[40] **Acknowledgments.** The authors thank Mike Wallace (University of Washington) and an anonymous reviewer for their helpful suggestions.

References

- Anderson, T., and D. Darling (1954), A test of goodness of fit, *J. Am. Stat. Assoc.*, *49*, 765–769.
- Broccoli, A. J., K. W. Dixon, T. L. Delworth, T. R. Knutson, R. J. Stouffer, and F. Zeng (2003), Twentieth-century temperature and precipitation trends in ensemble climate simulations including natural and anthropogenic forcing, *J. Geophys. Res.*, *108*(D24), 4798, doi:10.1029/2003JD003812.
- Chipperfield, M. P., and W. J. Randel (2003), Global ozone: Past and future, in *Scientific Assessment of Ozone Depletion: 2002, Global Ozone Research and Monitoring Project, Rep. 47*, World Meteorol. Organ., Geneva, Switzerland.
- Christiansen, B. (2003), Evidence for nonlinear climate change: Two stratospheric regimes and a regime shift, *J. Clim.*, *16*, 3681–3690.
- Christy, J. R., R. W. Spencer, W. B. Norris, W. D. Braswell, and D. E. Parker (2003), Error estimates of version 5.0 of MSU/AMSU bulk atmospheric temperatures, *J. Atmos. Oceanic Tech.*, *20*, 613–629.
- Delworth, T. L., and T. R. Knutson (2000), Simulation of early 20th century global warming, *Science*, *287*, 2246–2250.
- Folland, C. K., T. R. Karl, J. R. Christy, R. A. Clarke, G. V. Gruza, J. Jouzel, M. E. Mann, J. Oerlemans, M. J. Salinger, and S.-W. Wang (2001a), Observed climate variability and change, in *Climate Change 2001: The Scientific Basis*, edited by J. T. Houghton et al., 881 pp., Cambridge Univ. Press, New York.
- Folland, C. K., et al. (2001b), Global temperature change and its uncertainties since 1861, *Geophys. Res. Lett.*, *28*(13), 2621–2624.
- Forster, P. M. de F., and K. P. Shine (2002), Assessing the climate impact of trends in stratospheric water vapor, *Geophys. Res. Lett.*, *29*(6), 1086, doi:10.1029/2001GL013909.
- Gaffen, D. J., B. D. Santer, J. T. Boyle, J. R. Christy, D. E. Graham, and M. J. Ross (2000), Multi-decadal changes in the vertical temperature structure of the tropical troposphere, *Science*, *287*, 1239–1241.
- Hansen, J., et al. (2002), Climate forcings in Goddard Institute for Space Studies SI2000 simulations, *J. Geophys. Res.*, *107*(D18), 4347, doi:10.1029/2001JD001143.
- Hare, S. H. E., L. J. Gray, W. A. Lahoz, A. O'Neill, and L. Steenman-Clark (2004), Can stratospheric temperature trends be attributed to ozone depletion?, *J. Geophys. Res.*, *109*, D05111, doi:10.1029/2003JD003897.
- Jones, P. D., and A. Moberg (2003), Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001, *J. Clim.*, *16*, 206–223.
- Jones, P. D., M. New, D. E. Parker, S. Martin, and I. G. Rigor (1999), Surface air temperature and its variations over the last 150 years, *Rev. Geophys.*, *37*, 173–199.
- Joshi, M. M., and K. P. Shine (2003), A GCM study of volcanic eruptions as a cause of increased stratospheric water vapor, *J. Clim.*, *16*, 3525–3534.
- Karl, T. R., R. W. Knight, and B. Baker (2000), The record breaking global temperatures of 1997 and 1998: Evidence for an increase in the rate of global warming?, *Geophys. Res. Lett.*, *27*(5), 719–722.
- Lanzante, J. R. (1996), Resistant, robust and nonparametric techniques for analysis of climate data: Theory and examples, including applications to historical radiosonde station data, *Int. J. Climatol.*, *16*, 1197–1226.
- Lanzante, J. R., S. A. Klein, and D. J. Seidel (2003), Temporal homogenization of monthly radiosonde temperature data. Part I: Methodology, *J. Clim.*, *16*, 224–240.
- Laurmann, J., and L. Gates (1977), Statistical considerations in the evaluation of climatic experiments with atmospheric general circulation models, *J. Atmos. Sci.*, *34*, 1187–1199.
- Lindzen, R. S., and C. Giannitsis (2002), Reconciling observations of global temperature change, *Geophys. Res. Lett.*, *29*(12), 1583, doi:10.1029/2001GL014074.
- Mears, C. A., M. C. Schabel, and F. J. Wentz (2003), A reanalysis of the MSU channel 2 tropospheric temperature record, *J. Clim.*, *16*, 3650–3664.
- Meehl, G. A., W. M. Washington, T. M. L. Wigley, J. M. Arblaster, and A. Dai (2003), Solar and greenhouse gas forcing and climate response in the twentieth century, *J. Clim.*, *16*, 426–444.
- Miller, A. J., D. R. Cayan, T. P. Barnett, N. E. Graham, and J. M. Oberhuber (1994), The 1976–77 climate shift of the Pacific Ocean, *Oceanography*, *7*, 21–26.
- National Research Council (U.S.) Panel on Reconciling Temperature Observations (2000), *Reconciling Observations of Global Temperature Change*, 85 pp., Natl. Acad. Press, Washington, D. C.
- Parker, D. E., M. Gordon, D. P. N. Cullum, D. M. H. Sexton, C. K. Folland, and N. Rayner (1997), A new global gridded radiosonde temperature data base and recent temperature trends, *Geophys. Res. Lett.*, *24*(12), 1499–1502.
- Pawson, S., K. Labitzke, and S. Leder (1998), Stepwise changes in stratospheric temperature, *Geophys. Res. Lett.*, *25*(12), 2157–2160.
- Peterson, T. C., and R. S. Vose (1997), An overview of the Global Historical Climatology Network temperature database, *Bull. Am. Meteorol. Soc.*, *78*, 2837–2849.
- Priestley, M. (1981), *Spectral Analysis and Times Series: Univariate Series*, vol. 1, Academic, San Diego, Calif.
- Ramaswamy, V., et al. (2001), Stratospheric temperature trends: Observations and model simulations, *Rev. Geophys.*, *39*, 71–122.
- Randel, W. J., F. Wu, J. M. Russell III, J. M. Zawodny, and J. Nash (2000), Interannual changes in stratospheric constituents and global circulation derived from satellite data, in *Atmospheric Science Across the Stratosphere*, *Geophys. Monogr. Ser.*, vol. 123, pp. 271–285, AGU, Washington, D. C.
- Rodionov, S. N. (2004), A sequential algorithm for testing climate regime shifts, *Geophys. Res. Lett.*, *31*, L09204, doi:10.1029/2004GL019448.
- Schwarz, G. (1978), Estimating the dimension of a model, *Ann. Stat.*, *6*, 461–464.
- Seidel, D. J., et al. (2004), Uncertainty in signals of large-scale climate variations in radiosonde and satellite upper-air temperature datasets, *J. Clim.*, *17*, 2225–2240.
- Shine, K. P., et al. (2003), A comparison of model-predicted trends in stratospheric temperatures, *Q. J. R. Meteorol. Soc.*, *129*, 1565–1588, doi:10.1256/qj.02.186.
- Stott, P. A., S. F. B. Tett, G. S. Jones, M. R. Allen, J. F. B. Mitchell, and G. J. Jenkins (2000), External control of 20th century temperature by natural and anthropogenic forcings, *Science*, *290*, 2133–2137.
- Thiebaut, H., and F. Zwiers (1984), The interpretation and estimation of effective sample size, *J. Clim. Appl. Meteorol.*, *23*, 800–811.
- Tomé, A., and P. M. A. Miranda (2004), Piecewise linear fitting and trend changing points of climate parameters, *Geophys. Res. Lett.*, *31*, L02207, doi:10.1029/2003GL019100.

- Trenberth, K. E. (1990), Recent observed interdecadal climate changes in the Northern Hemisphere, *Bull. Am. Meteorol. Soc.*, *71*, 988–993.
- Trenberth, K. E., and J. W. Hurrell (1994), Decadal atmosphere-ocean variations in the Pacific, *Clim. Dyn.*, *9*, 303–319.
- Von Storch, H., and F. W. Zwiers (1999), *Statistical Analysis in Climate Research*, 484 pp., Cambridge Univ. Press, New York.
- Woodward, W., and H. Gray (1995), Selecting a model for detecting the presence of a trend, *J. Clim.*, *8*, 1929–1937.
- Zhang, Y., J. M. Wallace, and D. S. Battisti (1997), ENSO-like interdecadal variability: 1900–93, *J. Clim.*, *10*, 1004–1020.
- Zwiers, F. W., and A. J. Weaver (2000), The causes of 20th century warming, *Science*, *290*, 2081–2083.

J. R. Lanzante, NOAA/Geophysical Fluid Dynamics Lab (GFDL), Forrestal Campus, Route 1 North, Princeton, NJ 08542, USA. (john.lanzante@noaa.gov)

D. J. Seidel, NOAA Air Resources Laboratory (R/ARL), 1315 East West Highway, Silver Spring, MD 20910, USA. (dian.seidel@noaa.gov)