4.3 ABRUPT GLOBAL TEMPERATURE CHANGE AND THE INSTRUMENTAL RECORD

Matthew J. Menne *
NOAA National Climatic Data Center, Asheville, North Carolina

1. INTRODUCTION

Changes in mean annual global surface temperatures during the period of reasonably widespread instrumental observations are widely recognized to be non-monotonic (Folland, et al., 2001a; Karl, et al., 2000; Seidel and Lanzante, 2004). Visual inspection of the sequence of annually averaged surface temperature series, like the NOAA/NCDC analysis (Quayle, et al., 1999) shown in Fig. 1, suggests that the rise in average surface temperatures during the instrumental period of record occurred during two distinct periods: from about 1910 to 1945 and since 1976. A slight negative or neutral trend from 1946 to about 1975 appears to separate the two periods of quasi-linear surface temperature increase. In spite of these apparent temperature regimes, net global surface temperature change is commonly estimated via a simple linear regression fit of the temperature curve (e.g., Folland et al. 2001; Levinson, 2005). For the period from 1880 to 2004, the net change thus estimated is around 0.6°C.

The perception of different trend regimes has, nevertheless, provided the motivation to formally identify breakpoints that mark transitions from one quasi-linear tendency period to another (Karl et al., 2000; Tome and Miranda, 2004). Identification of breakpoints is necessary, for example, to compare the rate of surface temperature change during two different regimes of the observational period or to compare rates of change in model simulations of climate during the instrumental period to the observational record. Moreover, estimates of net surface temperature change are themselves sensitive to the dimensions of the statistical model used to fit the observations (Seidel and Lanzante, 2004).

Most model descriptions of global mean surface temperature trends have not included abrupt (step-like) changes. In spite of this, a “sloped step” model such as discussed by Seidel and Lanzante (2004) may be appropriate for characterizing certain rapid changes in tropospheric temperatures, whether the cause is internal climate variability (e.g., Lorenz, 1968, 1990; Trenberth and Hurrell, 1994; Wigley, 1989; Zhang, et al., 1997) or sudden external forcings like volcanism (Robock and Moa, 1995). Moreover, in a comparison of three alternatives to simple linear trends, Seidel and Lanzante (2004) noted that a sloped-step model, that is, one in which periods of quasi-linear change are interrupted by abrupt changes, was a better model fit to the global mean surface temperature observations than a simple linear, piecewise linear (trend changes only) or “flat steps” (step changes only) model. In their analysis, the locations of the step changes were determined visually and with an effort to be

* *Corresponding authors address*: Matthew Menne, NOAA National Climatic Data Center, 151 Patton Avenue, Asheville, North Carolina 28801; e-mail: Matthew.Menne@noaa.gov.

consistent with previous research. Here, a model to the global temperature curve is similarly fit, but the *a priori* assumptions are minimized concerning the number and the instants of changepoints as well as the time interval between them. Moreover, given that trend changes and step changes may be easily confounded (both statistically and visually), a test for undocumented changepoints was used (e.g., Lund and Reeves, 2002) to identify more objectively where breaks in the global temperature curve most likely occur. The test model contains parameters for changes in mean and changes in trend. The model that is implied from the undocumented changepoint test then is compared to previously identified models.

2. METHODS

The test used here to identify probable breaks in the surface temperature series is a Two Phase Regression (TPR) test. Following the discussion by Lund and Reeves (2002), the formulation for a TPR model describing a series $\{Y_t\}$ is given by

$$Y_t = \begin{cases} \mu_1 + \alpha_1 t + \varepsilon_t, & 1 \le t \le c \\ \mu_2 + \alpha_2 t + \varepsilon_t, & c < t \le n \end{cases}. \quad (1)$$

Under the null hypothesis of no changepoint, the two phases of the regression should be statistically equivalent and both the difference in means, $\mu_1 - \mu_2$, and the difference in slopes, $\alpha_1 - \alpha_2$, should be close to zero for each possible changepoint $c \in \{1,..,n\}$. In that case, a single or 'REDuced' phase of the regression is justified since $\mu_1 \approx \mu_2 \approx \mu_{RED}$ and $\alpha_1 \approx \alpha_2 \approx \alpha_{RED}$. To evaluate the null hypothesis of no changepoint versus the alternative hypothesis of an undocumented changepoint, an *F*-statistic is calculated at every time step in the series as

$$F_c = \frac{\left(\mathrm{SSE}_{\mathrm{RED}} - \mathrm{SSE}_{\mathrm{FULL}}\right)/2}{\left(\mathrm{SSE}_{\mathrm{FULL}}/(n-4)\right)}, \quad (2)$$

where SSE_{FULL} refers to the sum of the squared errors about each of the two phases (the two-phase or "FULL" model). The null hypothesis is rejected when the value of F_c is larger than would be expected from chance variation. If the instant of changepoint risk is unknown, for example, when there is an incomplete knowledge of the history of observation practice, every time step, *t*, must be evaluated as a potential changepoint, *c*. In this case, the magnitude of the *F*-statistic necessary to reject the null hypothesis of homogeneity must account for serial correlation in the calculated series of F_c's calculated using (2). Percentiles of the *F*-statistic appropriate for an undocumented changepoint can be obtained via simulations under the null hypothesis, in which the maximum of each series of F_c 's is recorded as

$$F_{\max} = \max_{1 \le c \le n} F_c . \quad (3)$$

The null hypothesis of a "one phase" time series is rejected when the value of $F_{\max}$ is greater than the percentile selected as the significance level. Here we use $F_{\max,0.95}$. As shown in Table 1 of Lund and Reeves (2002), percentiles of $F_{\max}$ are larger than their counterparts for the more usual *F*-statistic percentiles that are appropriate for testing a specific, known (observed) risk of changepoint. Consequently, changepoints that are significant with respect to $F_{\max}$ will also be significant with respect to the equivalent percentiles of the, e.g., $F_{3,n-4}$ statistic.

2.1 *Detecting multiple breaks in a series*

The presence of multiple breaks in a series can complicate the interpretation of $F_{\max}$. To resolve multiple breakpoints, hierarchic binary segmentation is often used in successive hypothesis testing. In this method, a series is split at the position of $F_{\max}$ when its value exceeds the critical (e.g., 0.05) value. Subsequences on either side of this split are then similarly evaluated for changepoints and additional splits are made, if necessary. This process is repeated recursively until the magnitude of $F_{\max}$ does not exceed the critical value in remaining segments or the sample size, *n*, in a segment is too small to test. This type of segmentation process is considered a "greedy" method because changepoints are selected to maximize the separation between segments at each split rather than evaluating all possible changepoint configurations iteratively to identify the optimal multi-way split. The solution is hierarchic because it will reliably converge to the optimal solution only when the true changepoints are hierarchic, which may not be the case (Hawkins, 2001). Optimal solutions, on the other hand, can be identified in a manner similar to the approach used by Karl et al. (2000), that is, by minimizing the observed deviation from the model response. A penalty function must be used to prevent "over-fitting" the observations and optimal approaches arguably work best when the analyst intervenes in establishing the ultimate solution (Lavielle, 1998; Cassinus and Mestre, 2004).

Because the interest here is in minimizing *a priori* assumptions about the number and instant of changepoints in the global temperature curve, a modest variation on successive hypothesis testing was used that compares more favorably to optimal solutions. In this "semi-hierarchic" splitting algorithm, each splitting step is followed by a merging step to test whether a split chosen at an earlier stage has lost its importance after subsequent breakpoints are identified (Hawkins, 1976; Menne and Williams, 2005).

2.2 *Comparison to previous solutions*

Like Seidel and Lanzante (2004), a modified form of the Schwartz Bayesian Information Criterion (BIC) statistic (Schwarz, 1978) was used to compare the model derived from successive hypothesis testing to previous solutions. The statistic, *S(q)*, quantifies the model's ability to replicate the time series while adhering to the principle of parsimony and is calculated as

$$S(q) = n_e \log[\frac{1}{n}\sum_{t=1}^{n}(T(t)-\hat{T}(t))^2] + q\log(n_e) \quad (4)$$

The BIC is proportional to the pooled sum of squares about each segments fit to the data and to the dimensions, *q*, of the model required to fit the segments. The lower the value of *S(q)*, the better the model fit. Model parsimony is rewarded via the penalty function, which is proportional to *q.* The value of *q* depends on the number of segments fit to the series as well as the nature of the breakpoints (i.e., change in mean, change in slope, or both). Considering the changepoint models evaluated by Seidel and Lanzante (2004), a flat steps model requires specification of the locations of *k* changepoints and the mean for each of the *k*+1 segments, or

$$q_{\text{flat steps}} = k + (k+1) = 2k + 1. \quad (5)$$

The piecewise linear model requires specification of the locations of *k* changepoints, the slope for each of the *k*+1 segments, plus the intercept of the first segment, so

$$q_{\text{piecewise linear}} = k + (k+1) + 1 = 2k + 2. \quad (6)$$

The sloped step model requires specification of locations of *k* breakpoints plus the slope and intercept for each of the *k*+1 segments, and

$$q_{\text{sloped steps}} = k + 2(k+1) = 3k + 2. \quad (7)$$

Given the presence of serial correlation in annual temperature values, model residuals are treated as a first order autoregressive process and the effective sample size, n_e, is calculated as

$$n_e = n\frac{(1-\rho_1)}{(1+\rho_1)}, \quad (8)$$

where ρ_1 is the lag 1 (1 year) autocorrelation coefficient.

3. RESULTS

In Fig. 2, the Karl et al. (2000) piecewise linear solution, with breakpoints in 1910, 1941 and 1975, is shown. A sloped-step analysis similar to the Seidel and Lanzante, 2004 monthly mean temperature analysis is shown in Fig. 3 for annual mean values. This sloped step model includes the breakpoints selected by Seidel and Lanzante (2004) at the years 1945 and 1977 in addition to one added in 1900 (the first year of their monthly analysis) to extend Seidel and Lanzante (2004) analysis back to 1880, consistent with Karl et al. (2000).

Like the Karl et al. (2000) solution presented in Fig. 2, three breakpoints and four segments are suggested by successive two-phase regression hypothesis testing using the hierarchical splitting algorithm, as shown in Fig. 4a. Instead of phase breakpoints occurring around the years 1910, 1941, and 1975, abrupt changes in the new analysis are identified at the years 1902, 1945 and 1963. The 1945 breakpoint is common to Seidel and Lanzante (2004) and this analysis. However, while the Seidel and Lanzante (2004) step changes in 1945 and 1977 are small and offsetting (-0.09 K in 1945 and 0.07 K in 1977), the steps in the new analysis are larger and all of the same sign (-0.22 K in 1902; -0.16 K in 1945; -0.21 K).

In Table 1, the values of the *S(q)* statistic are provided for four different model fits to the global surface temperature series. The four models include the simple linear, the Karl et al. (2000) piecewise linear, the sloped stepped model according to Seidel and Lanzante (2004) as well as for the new model identified using the successive hypothesis tests for breakpoints. The net temperature change for each of the four models is also provided. As shown in Table 1, the lowest *S(q)* value occurs for the new model fit, which suggests that the more objectively identified breakpoints result in a better fit of the annual mean temperature values. In addition, compared to the simple linear model, the estimates of net temperature change during the period of record is higher for piecewise linear and the sloped steps models.

The same successive hypothesis testing algorithm was applied to the global mean surface temperature series based on archives maintained by institutions in the United Kingdom (Folland et al., 2001b; Jones, 1994; Jones and Moberg, 2003; Jones et al., 2001; Parker et al., 1995). As shown in Fig. 5, an identical set of breakpoint years was obtained. In some respects this is not surprising since there is dependency between the UK analysis and (Quayle et al., 1999).

4. DISCUSSION AND CONCLUSIONS

We are reluctant to suggest that this new model provides a more accurate description of the low frequency behavior of global surface temperatures than previous descriptions until further analysis is done to investigate possible causes behind the alternative set of breakpoints. Further analysis should include evaluation of step-type signals at higher geographic resolution to determine whether there is any physical support for this set of abrupt breakpoint years. Nevertheless, the new model does raise some interesting points. For example, as a consequence of resolving abrupt changes at the years 1945 and 1963 in the new analysis, no changepoint occurs at the 1977 climate shift (Trenberth and Hurrell, 1994]; Zhang et al., 1997), at least in the global mean series. In fact, following the abrupt break in 1963, the year 1977 is not a statistically significant breakpoint (0.05 significance level) even with respect to a TPR test for a known changepoint risk in which a smaller critical value for the *F*-statistic is appropriate. As a result, the new model suggests that the most recent warming phase dates to 1964, rather than to the late 1970s. In addition, in this new description, none of the temperature trend phases is negative. The trend between 1946 and 1963, however, is not statistically significant.

5. AKNOWLEDGEMENTS

Special thanks to Anne Waple of NCDC and STG, Inc. for helpful discussions concerning the low frequency variability of global surface temperatures.

6. REFERENCES

Cassinus, H., and O. Mestre (2004), Dectection and correction of artificial shifts in climate series, Part III, *Journal of the Royal Statistical Society, Series C*, *53*, 405-425.

Folland, C. K., et al. (2001a), Observed climate variability and change, in *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernment Panel on Climate Change*, edited by J. T. Houghton, et al., p. 881, Cambridge University Press, Cambridge, UK.

Folland, C. K., et al. (2001b), Global temperature change and its uncertainties since 1861, *Geophysical Research Letters*, *28*, 2621-2624.

Hawkins, D. M. (1976), Point estimation of the parameters of piecewise regression models, *Applied Statistics*, *25*, 51-57.

Jones, P. D. (1994), Hemispheric surface air temperature variations: a reanalyis and an update to 1993, *Journal of Climate*, 1794-1802.

Jones, P. D., and A. Moberg (2003), Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001, *Journal of Climate*, *16*, 206-223.

Jones, P. D., et al. (2001), Adjusting for sampling density in grid box land and ocean surface temperature time series, *Journal of Geophysical Research-Atmospheres*, *106*, 3371-3380.

Karl, T. R., et al. (2000), The record breaking global temperatures of 1997 and 1998: Evidence for an increase in the rate of global warming? *Geophysical Research Letters*, *27*, 719-722.

Lavielle, M. (1998), Optimal segmentation of random processes, *IEEE Transactions on Signal Processing*, *46*, 1365-1373.

Levinson, D. H. (2005), State of the Climate in 2004, *Bulletin of the American Meteorological Society*, *86*, S1-S86.

Lorenz, E. N. (1968), Climatic determinism, *Meteorological Monographs*, *8*, 1-3.

Lorenz, E. N. (1990), Can chaos and intransitivity lead to interannual variability? *Tellus*, *42A*, 378-389.

Lund, R., and J. Reeves (2002), Detection of undocumented changepoints: A revision of the two-phase regression model, *Journal of Climate*, *15*, 2547-2554.

Menne, M.J., and C. N. Williams (2005), Detection of undocumented changepoints using multiple test statistics and reference series, *Journal of Climate*, *18*, 4271-4286.

Parker, D. E., et al. (1995), Marine surface temperature: observed variations and data requirements, *Climatic Change*, *31*, 559-600.

Quayle, R. G., et al. (1999), An operational near-real-time global temperature index, *Geophysical Research Letters*, *26*, 333-335.

Robock, A., and J. Moa (1995), The volcanic signal in surface temperature records, *Journal of Climate*, *8*, 1086-1103.

Schwarz, G. (1978), Estimating the dimension of a model, *The Annals of Statistics*, *6*, 461-464.

Seidel, D. J., and J. R. Lanzante (2004), An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes, *Journal of Geophysical Research-Atmospheres*, *109*.

Tomé, A. R., and P. M. A. Miranda (2004), Piecewise linear fitting and trend changing points of climate parameters, *Geophysical Research Letters*, *31*.

Trenberth, K., and J. Hurrell (1994), Decadal atmospheric-ocean variations in the Pacific, *Climate Dynamics*, *9*.

Wigley, T. M. L. (1989), Climate variation on the 10-100 year time scale: observations and possible causes, in *Global Changes of the Past*, edited by R. S. Bradley, pp. 83-101, UCAR/Office for Interdisciplinary Earth Studies, Boulder.

Zhang, Y., et al. (1997), ENSO-like interdecadal variability: 1900-93, *Journal of Climate*, *10*, 1004-1020.

Table 1. Value of the Swartz Bayesian Information Criterion, *S(q)*, and the net temperature change, ΔT, for four models describing the low frequency changes in global mean surface temperature.

Model	***S(q)***	**ΔT**
Linear	-115.2	0.61 K
Piecewise Linear (Karl et al., 2000)	-141.6	0.63 K
Sloped Steps (Seidel and Lanzante, 2004)	-145.0	0.71K
Sloped Steps (this analysis)	-179.7	0.69 K

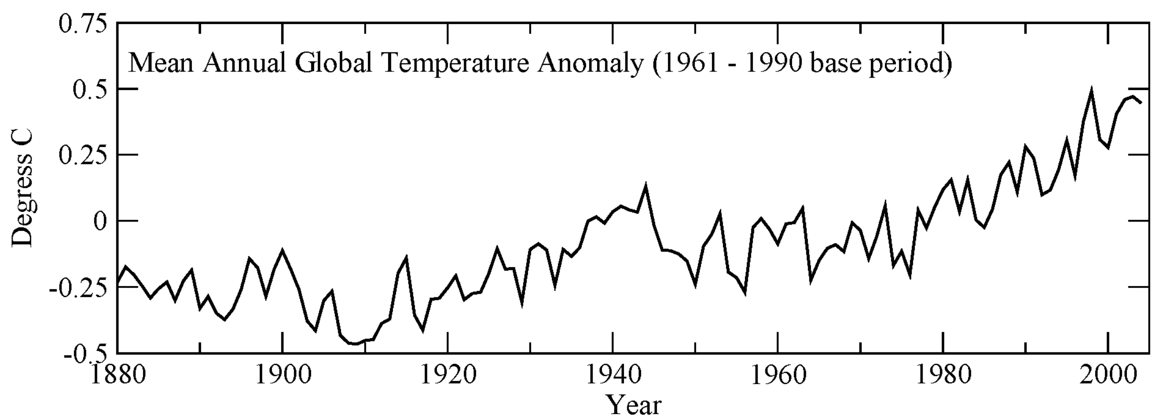

Figure 1. Mean annual global surface temperatures from NOAA/NCDC archives [*Quayle, et al.*, 1999].

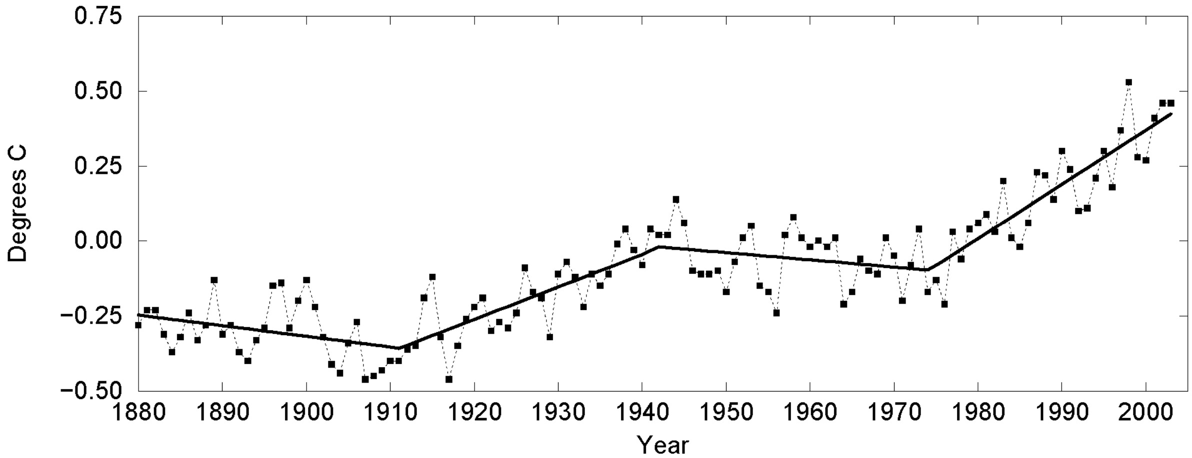

Figure 2. Piecewise linear model of the NOAA/NCDC temperature record (solid line) identified using an optimal technique (trend breakpoints only) described by [*Karl, et al.*, 2000; *Tomé and Miranda*, 2004].

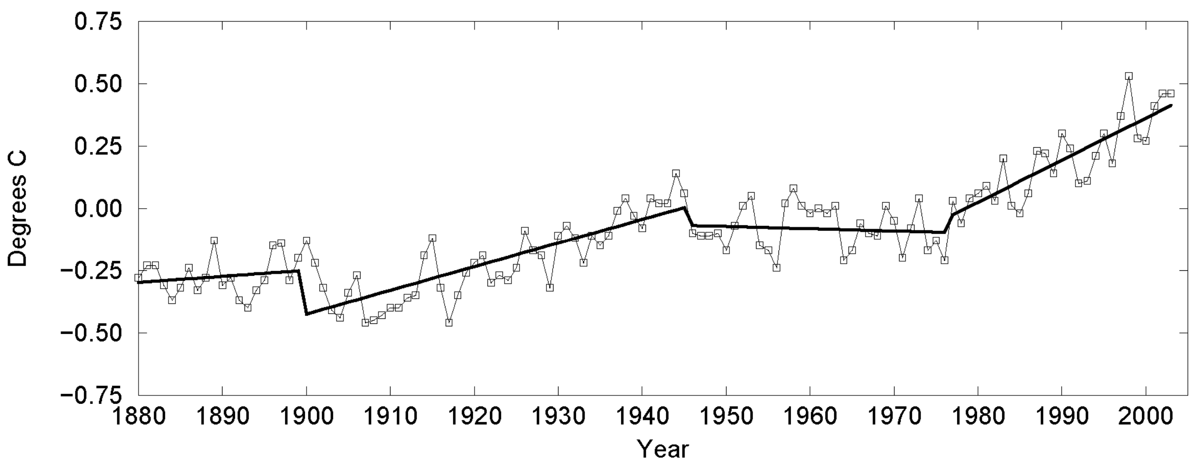

Figure 3. Sloped step model describing low frequency variability in global mean surface temperatures in the [*Quayle, et al.*, 1999] analysis using 1900 and the years identified by [*Seidel and Lanzante*, 2004] as breakpoints.

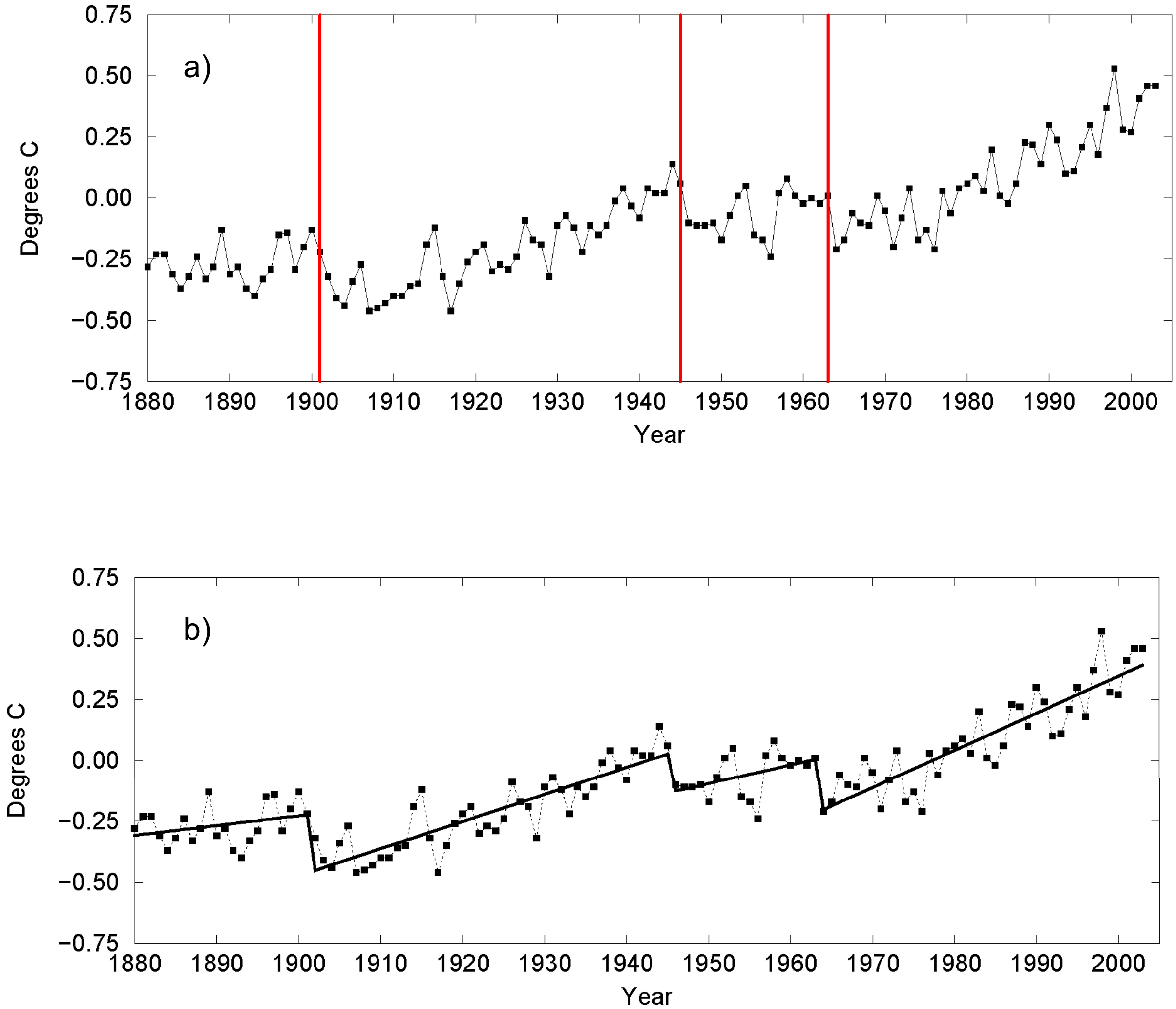

Figure 4. a) breakpoint years (red bars) identified using successive hypothesis testing and the two phase regression (TPR) test for undocumented changepoints; b) model describing low frequency variability in global mean surface temperatures in the [*Quayle, et al.*, 1999] analysis using the breakpoints shown in Fig. 4a.

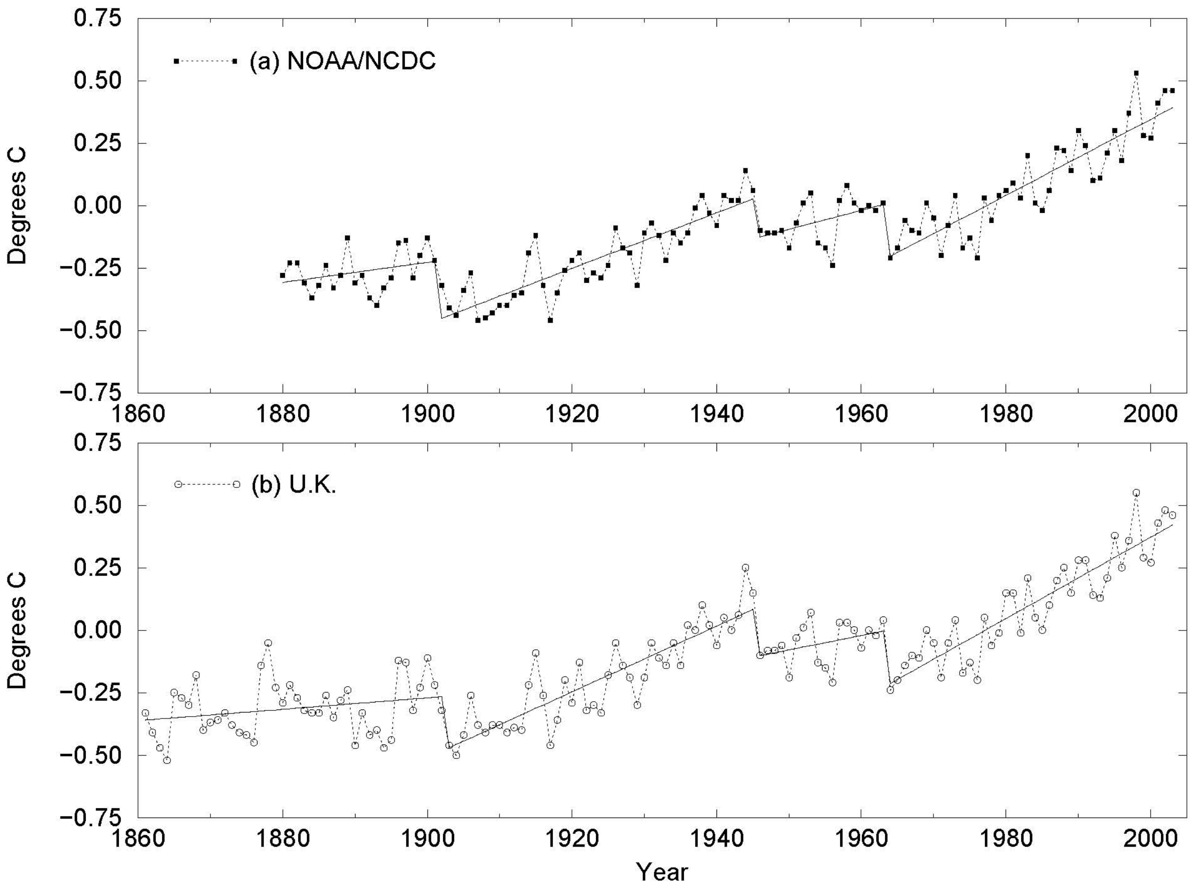

Figure 5. (a) Model describing low frequency variability in global mean surface temperatures in the [*Quayle, et al.*, 1999] analysis with breakpoints identified using successive hypothesis testing and the TPR test for undocumented changepoints (reproduced from Fig. 4b); (b) as in Fig. 5a, except for the surface temperature analysis described by [*Folland, et al.*, 2001b; *Jones*, 1994; *Jones and Moberg*, 2003; *Jones, et al.*, 2001; *Parker, et al.*, 1995].