NOTES AND CORRESPONDENCE

Detection of Undocumented Changepoints: A Revision of the Two-Phase Regression Model

ROBERT LUND AND JAXK REEVES

Department of Statistics, University of Georgia, Athens, Georgia

5 November 2001 and 25 March 2002

ABSTRACT

Changepoints (inhomogeneities) are present in many climatic time series. Changepoints are physically plausible whenever a station location is moved, a recording instrument is changed, a new method of data collection is employed, an observer changes, etc. If the time of the changepoint is known, it is usually a straightforward task to adjust the series for the inhomogeneity. However, an undocumented changepoint time greatly complicates the analysis. This paper examines detection and adjustment of climatic series for undocumented changepoint times, primarily from single site data. The two-phase regression model techniques currently used are demonstrated to be biased toward the conclusion of an excessive number of unobserved changepoint times. A simple and easily applicable revision of this statistical method is introduced.

1. Introduction

Changepoints (discontinuities, inhomogeneities) are a ubiquitous feature of climatic series. Changepoints can arise from changes in recording instruments, station location moves, etc. Figure 1 of Jones et al. (1986) and Fig. 7 of Vincent (1998) show mean shifts in observed temperatures that are attributable to changepoints.

Our definition of a changepoint is a time where the mean of the series first undergoes a structural pattern change. This change may or may not induce a discontinuity in mean series values, but there is some pattern change (e.g., series trend slopes and/or the location parameters of the series could shift). We focus on changepoints in mean values only; variance changepoints, or more general structural changes in the marginal distributions of the series are not considered.

Changepoints can substantially alter conclusions made from climatic series. For example, Lund et al. (2001) show that changepoint information is the single most important factor for obtaining an accurate estimate of the linear temperature change rate for a fixed U.S. station. Specifically, it is not unusual for a U.S. temperature recording station to experience four or more changepoints over a century of operation. As a linear change rate is typically on the order of 1 or 2 °C century^{-1}, and each changepoint can induce a mean shift of a few degrees Celsius, even a small number of changepoints can seriously corrupt the estimation of a linear trend. Figure 1 illustrates the idea geometrically, depicting a series of length 1200 that experienced a single changepoint with a mean shift of 2.0°C at time 600. The true linear trend slope in this data is -1.0°C century^{-1}; the solid and broken lines represent least squares fitted lines to the data. Observe the differences in trend estimates: 1.894°C century^{-1} when the changepoint is ignored and -0.964°C century^{-1} when the changepoint is accounted for. Overall, one can appreciate the criticality of the changepoint issue.

The climatic literature on changepoints is by now extensive (cf. Thompson 1984; Solow 1987; Karl and Williams 1987; Gullet et al. 1991; Rhoades and Salinger 1993; Easterling and Peterson 1995; Vincent 1998). The statistical literature is also vast, with Page (1955), Kander and Zacks (1966), Hinkley (1969, 1971), Brown et al. (1975), Hawkins (1977), and Chen and Gupta (2000) composing a prominent historical sample.

In many cases, the time of the changepoint(s) is (are) known. The move of a temperature recording station, for example, is revealed from changes in the station longitude, latitude, or elevation records (station histories or metadata). Station histories, however, are notoriously incomplete, if they even exist. Hence, not all changepoint times are documented. An undocumented changepoint may be visually evident from a time series plot of the data in question. Still, many situations arise where

Corresponding author address: Dr. Robert Lund, Department of Statistics, University of Georgia, 204 Statistics Building, Athens, GA 30602-1952.
E-mail: lund@stat.uga.edu

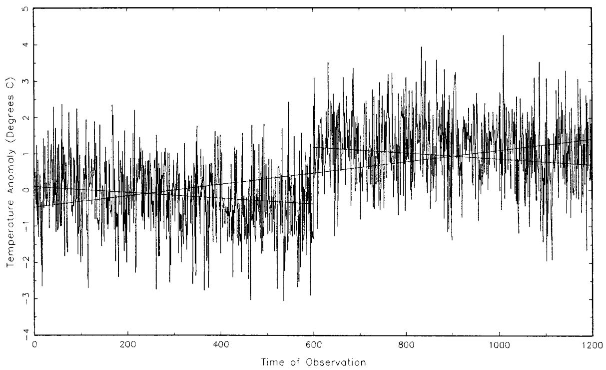

FIG. 1. Effect of changepoints on trend estimation.

the existence of a changepoint is debatable. This paper presents a revision of the two-phase linear regression test for changepoint detection at an undocumented time.

2. An $F_{\max}$ test statistic

We start with the simple two-phase linear regression scheme for a climatic series $\{X_t\}$ considered by Solow (1987), Easterling and Peterson (1995), and Vincent (1998; among others). This model can be written in the form:

$$X_t = \begin{cases} \mu_1 + \alpha_1 t + \epsilon_t, & 1 \le t \le c \\ \mu_2 + \alpha_2 t + \epsilon_t, & c < t \le n, \end{cases} \tag{2.1}$$

where $\{\epsilon_t\}$ is mean zero independent random error with a constant variance.

The model in (2.1) is viewed as a classic simple linear regression that allows for two phases. This allows for both step- ($\mu_1 \ne \mu_2$) and trend- ($\alpha_1 \ne \alpha_2$) type changepoints. Specifically, the time c is called a changepoint in (2.1) if $\mu_1 \ne \mu_2$ and/or $\alpha_1 \ne \alpha_2$. In most cases, there will be a discontinuity in the mean series values at the changepoint time c, but this need not always be so (Fig. 10 in section 5 gives a quadratic-based example where the changepoint represents more of a slowing of rate of increase than a discontinuity).

A good changepoint detection method will detect both step- and trend-type changepoints. Of course, changes in trend slopes could be rooted in true climate change, station relocations, urban heat effects, etc. In general, we will not attempt to assign cause to every found changepoint. Also, step- and trend-type changepoints cannot be unconfounded in general. One can, however, get a good feeling for whether the step- or trend-type changepoint dominates by plotting the estimated mean response against the data (cf. Figs. 3 and 10 below).

If a time $c \in \{2, \ldots, n-1\}$ is known to be the only changepoint time a priori, then least squares estimates of the trend parameters in (2.1) are

$$\hat{\alpha}_1 = \frac{\sum_{t=1}^{c} (t - \bar{t}_1)(X_t - \overline{X}_1)}{\sum_{t=1}^{c} (t - \bar{t}_1)^2} \quad \text{and}$$

$$\hat{\alpha}_2 = \frac{\sum_{t=c+1}^{n} (t - \bar{t}_2)(X_t - \overline{X}_2)}{\sum_{t=c+1}^{n} (t - \bar{t}_2)^2}, \tag{2.2}$$

where $\overline{X}_1 = c^{-1} \sum_{t=1}^{c} X_t$ and $\overline{X}_2 = (n-c)^{-1} \sum_{t=c+1}^{n} X_t$ are the average series values before and after time c, respectively; and $\bar{t}_1 = c^{-1} \sum_{t=1}^{c} t = (c+1)/2$ and $\bar{t}_2 = (n-c)^{-1} \sum_{t=c+1}^{n} t = (n+c+1)/2$ are the average time observations before and after time c, respectively. Least squares estimates of the location parameters μ_1 and μ_2 in (2.1) are

$$\hat{\mu}_1 = \overline{X}_1 - \hat{\alpha}_1 \bar{t}_1 \quad \text{and} \quad \hat{\mu}_2 = \overline{X}_2 - \hat{\alpha}_2 \bar{t}_2. \tag{2.3}$$

The denominators in (2.2) can be explicitly evaluated as

$$\sum_{t=1}^{c} (t - \bar{t}_1)^2 = \frac{c(c+1)(c-1)}{12} \quad \text{and}$$

$$\sum_{t=c+1}^{n} (t - \bar{t}_2)^2 = \frac{(n-c)(n-c+1)(n-c-1)}{12}. \tag{2.4}$$

To test a null hypothesis of no changepoints versus an alternative of an undocumented changepoint, a good test statistic must be developed. Under the null hypothesis, the regression parameters during the two phases must agree: $\alpha_1 = \alpha_2$ and $\mu_1 = \mu_2$. In this case, $\hat{\alpha}_1 - \hat{\alpha}_2$ and $\hat{\mu}_1 - \hat{\mu}_2$ should be statistically close to zero for each $c \in \{1, \ldots, n\}$. Rescaling this to a regression F statistic merely states that

$$F_c = \frac{(\mathrm{SSE}_{\mathrm{Red}} - \mathrm{SSE}_{\mathrm{Full}})/2}{\mathrm{SSE}_{\mathrm{Full}}/(n-4)} \tag{2.5}$$

should be small for each $c \in \{1, \ldots, n\}$ when there is no changepoint. In (2.5), $\mathrm{SSE}_{\mathrm{Full}}$ is the "full model" sum of squared errors computed from

$$\mathrm{SSE}_{\mathrm{Full}} = \sum_{t=1}^{c} (X_t - \hat{\mu}_1 - \hat{\alpha}_1 t)^2 + \sum_{t=c+1}^{n} (X_t - \hat{\mu}_2 - \hat{\alpha}_2 t)^2 \tag{2.6}$$

and $\mathrm{SSE}_{\mathrm{Red}}$ is the "reduced model" sum of squared errors

$$\mathrm{SSE}_{\mathrm{Red}} = \sum_{t=1}^{n} (X_t - \hat{\mu}_{\mathrm{Red}} - \hat{\alpha}_{\mathrm{Red}} t)^2, \tag{2.7}$$

where $\hat{\mu}_{\mathrm{Red}}$ and $\hat{\alpha}_{\mathrm{Red}}$ are estimated under the constraints $\mu_1 = \mu_2 = \mu_{\mathrm{Red}}$ and $\alpha_1 = \alpha_2 = \alpha_{\mathrm{Red}}$:

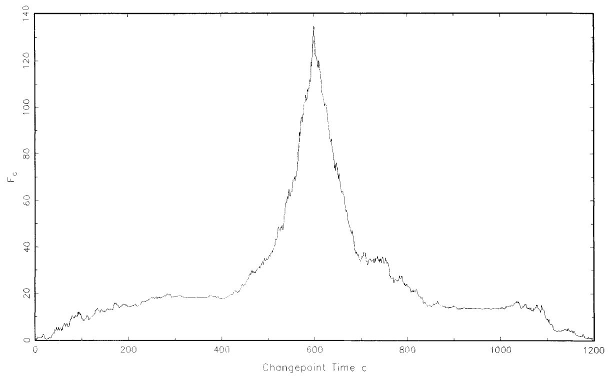

FIG. 2. The F statistics for Fig. 1 data.

$$\hat{\alpha}_{\mathrm{Red}} = 12\frac{\sum_{t=1}^{n}(X_t - \overline{X})t}{n(n+1)(n-1)} \quad \text{and}$$

$$\hat{\mu}_{\mathrm{Red}} = n^{-1}\sum_{t=1}^{n}(X_t - \hat{\alpha}_{\mathrm{Red}}t). \tag{2.8}$$

At the boundaries $c = 1$ and $c = n$, $\mathrm{SSE}_{\mathrm{Full}}$ is computed as follows. For $c = 1$, the first summation in (2.6) is interpreted as zero and the second summation represents the sum of squared errors after a line is fitted to $X_2, \ldots, X_n$. At $c = n$, the second sum in (2.6) is zero and the first is the sum of squared errors of a line fitted to $X_1, \ldots, X_{n-1}$.

Now, if a changepoint is present at time c, F_c, and hence $\max_{1\le c\le n} F_c$, should be large (statistically). Furthermore, $\max_{1\le c\le n} F_c$ will be small when F_c is small for each $c \in \{1, \ldots, n\}$, as is the case (statistically) under the null hypothesis. Hence, the existence of an undocumented changepoint is concluded when

$$F_{\max} = \max_{1\le c\le n} F_c \tag{2.9}$$

is too large to be attributed to chance variation. The time of the most prominent changepoint is estimated as the argument(s) c that maximize(s) F_c. Figure 2 plots values of F_c for the data in Fig. 1. The maximum F statistic is prominent at time 600, the true time of the changepoint, with an $F_{\max}$ statistic of 134.56.

To quantify how large $F_{\max}$ should be to conclude that an undocumented changepoint exists, the distribution of $F_{\max}$ under the null hypothesis must be derived. This is where mistakes have been propagated in the literature. Under the null hypothesis and Gaussian errors $\{\epsilon_t\}$, F_c has an F distribution with 2 numerator degrees of freedom and $n - 4$ denominator degrees of freedom for each c. Hence, under the null hypothesis, $F_{\max}$ behaves as the maximum of F statistics. However, the F_cs are not independent; rather, F_c and F_{c-1} are positively correlated (similar) for a fixed c. This dependence greatly complicates the analysis.

Hinkley (1969, 1971) examined the above issue and reported that the null hypothesis distribution of $F_{\max}$ was, under the constraint that the two regression lines meet at the changepoint time (unrealistic in this setup), approximately an $F_{3,n-4}$ distribution; that is, an F distribution with 3 numerator degrees of freedom and $n - 4$ denominator degrees of freedom. Hinkley likely reasoned that c is unknown and could be regarded as a third free parameter under the null, thereby adding 1 numerator degree of freedom. Solow (1987) applied Hinkley's distributional claim in the climate literature without reexamination of the issue. Easterling and Peterson (1995) quoted Solow and dropped the ''meeting constraint'' altogether. Vincent (1998) quoted Easterling and Peterson (1995).

In the next section, we will compute the true percentiles of the $F_{\max}$ distribution under the null hypothesis of no change. The comparisons show that the true distribution of $F_{\max}$ is much larger (stochastically) than that of the $F_{3,n-4}$ distribution. From a practical standpoint, using an $F_{3,n-4}$ null hypothesis distribution leads to overestimation of the number of unobserved changepoints. This has climatic repercussions. For example, adjusting a temperature record for too many changepoints eliminates natural variability in the record and may make insignificant anomalies appear significant.

TABLE 1. The $F_{\max}$ and $F_{3,n-4}$ percentiles.

n	$F_{\max,0.90}$	$F_{3,n-4,0.90}$	$F_{\max,0.95}$	$F_{3,n-4,0.95}$	$F_{\max,0.99}$	$F_{3,n-4,0.99}$
10	8.39	3.29	11.56	4.76	22.38	9.78
25	6.10	2.36	7.37	3.07	10.55	4.87
50	5.91	2.20	6.92	2.81	9.31	4.24
75	5.94	2.16	6.88	2.73	9.07	4.07
100	5.99	2.14	6.91	2.70	8.98	3.99
200	6.14	2.12	7.01	2.65	8.96	3.88
300	6.26	2.10	7.11	2.64	9.03	3.85
400	6.33	2.10	7.18	2.63	9.08	3.83
500	6.39	2.09	7.24	2.62	9.10	3.82
750	6.53	2.09	7.37	2.62	9.22	3.81
1000	6.57	2.09	7.42	2.61	9.26	3.80
2500	6.79	2.09	7.65	2.61	9.51	3.79
5000	6.98	2.08	7.85	2.61	9.68	3.79

3. The $F_{\max}$ percentiles

It is difficult to derive the null hypothesis distribution of $F_{\max}$ mathematically. However, null hypothesis $F_{\max}$ percentiles (critical values) can easily be obtained via simulation. One draw from the $F_{\max}$ distribution under the null hypothesis is obtained by simulating a time series satisfying (2.1) with $\mu_1 = \mu_2$ and $\alpha_1 = \alpha_2$, and then computing the $F_{\max}$ statistic. It is worth noting that the null hypothesis $F_{\max}$ distribution does not depend on the particular values of μ_1 and α_1 (regression F statistics are scale independent). This procedure is repeated many times in a Monte Carlo sense; the result is a sample from which the $F_{\max}$ critical values can be very accurately estimated.

Table 1 lists the 90th, 95th, and 99th percentiles of the $F_{\max}$ distribution as a function of the series length

n and compares these values with the $F_{3,n-4}$ percentiles. The $F_{\max}$ percentiles were obtained by simulating 100 000 $F_{\max}$ values for each series length n; 1 000 000 $F_{\max}$ values were in fact generated for all table entries with $n \leq 750$. Such simulated percentiles are quite accurate; in fact one standard error of estimation for each of the Table 1 entries is bounded by 0.010. Generation of Table 1 took over one month on a Pentium personal computer. The $F_{3,n-4}$ percentiles were obtained by inverting the true $F_{3,n-4}$ cumulative distribution function.

Several comments about the Table 1 values are worth making. First, the $F_{\max}$ percentiles are at least twice as large as their $F_{3,n-4}$ counterparts. Hence, use of the $F_{3,n-4}$ percentiles can cause a gross overestimation of the number of undocumented changepoints. Second, whereas the $(1 - \alpha) \times 100$ percentiles of the $F_{\max}$ distribution increase for $n \geq 100$, the $(1 - \alpha) \times 100$ percentiles of the $F_{3,n-4}$ distribution actually decrease in n. In fact, 3 times the $F_{3,n-4}$ distribution converges to a chi-squared distribution with 3 degrees of freedom as $n \to \infty$. In particular, $\lim_{n\to\infty} F_{3,n-4,0.95} = 2.60$, whereas $\lim_{n\to\infty} F_{\max,0.95} = \infty$. The above monotonicity and limiting properties mathematically demonstrate the inappropriateness of the $F_{3,n-4}$ distribution for undocumented changepoint detection.

For series lengths n that are not listed in Table 1, a linear interpolation of the percentiles appears to work reasonably well. Alternatively, we have fitted the curve

$$\hat{F}_{\max,0.95} = 3.5642 + 2\ln[\ln(n-4)]\left\{1 + \frac{4\ln[\ln(n-4)]}{n-4}\right\}, \quad n \geq 100, \tag{3.1}$$

to the 95th percentiles in Table 1. This curve, whose form is extracted from extreme value theory, fits the percentiles reasonably well for large n; we do not recommend its application unless $n \geq 100$.

Tracing the history of the $F_{\max}$ versus $F_{3,n-4}$ issue, it is clear that the asymptotic normality in Hinkley (1969, 1971) does not hold as claimed. Further error in the methods is attributed to Easterling's and Peterson's (1995) removal of Hinkley's requirement that the regression lines meet at the phase boundaries. Whereas Easterling and Peterson (1995) also apply numerous statistical tests (t tests, Durbin–Watson, multiresponse permutation, etc.) to candidate changepoint times, none of these methods fundamentally accounts for the fact that the changepoint time is unknown; hence, they should not be expected to work well in general.

We conjecture that a reasonable approximating limiting distribution of $F_{\max}$ (as $n \to \infty$) would involve the Gumbel extreme value law:

$$\Pr[F_{\max} \leq x] = \exp\{-[e^{-(x-a_n)/b_n}]\}, \tag{3.2}$$

for appropriately selected scaling constants a_n and b_n. Our conjecture is based on the fact that scaled maxima of independent F statistics must converge to the Gumbel distribution. The F statistics involved in (2.9) are very dependent, so naive application of classic extreme value results will be invalid. Nonetheless, from the extreme value laws for dependent random sequences in Leadbetter et al. (1983), the dependence in F_c (as c varies) does not appear (at first glance) strong enough to alter the overall form of the limiting distribution. It would, however, change the scaling constants a_n and b_n. Another plausible form for the limit distribution of $F_{\max}$ would involve the supremum of a Brownian Bridge process. Of course, with the Table 1 values and (3.1), there is no practical reason to consider this issue further.

4. Extensions and limitations of the methods

The above methods were developed for single site data, but they can easily accommodate reference series. If a reference series $\{R_t\}$ has a constant mean in time—$E[R_t] \equiv c$ for some constant c and all t—then one can apply the above methods to the differenced series $\{X_t - R_t\}$. The changepoint times in $\{X_t\}$ and $\{X_t - R_t\}$ will be identical. The assumption that $E[R_t]$ has a constant mean is important. For if $\{X_t\}$ and $\{R_t\}$ are experiencing similar climate change, the trend slopes of $\{X_t\}$ and $\{R_t\}$ will be approximately the same, and $\{X_t - R_t\}$ will have no appreciable trend.

After a changepoint time is detected, it is a simple matter to adjust the series for the changepoint. To put both phases of the series on a common mean zero basis, merely examine the "residuals" $\{X_t^*\}$ defined by

$$X_t^* = \begin{cases} X_t - \hat{\mu}_1 - \hat{\alpha}_1 t, & 1 \leq t \leq c \\ X_t - \hat{\mu}_2 - \hat{\alpha}_2 t, & c < t \leq n. \end{cases} \tag{4.1}$$

To convert all measurements to a current day basis, examine $\{X_t^{**}\}$ defined by

$$X_t^{**} = \begin{cases} X_t + (\hat{\mu}_2 - \hat{\mu}_1) + (\hat{\alpha}_2 - \hat{\alpha}_1)t, & 1 \leq t \leq c \\ X_t, & c < t \leq n. \end{cases} \tag{4.2}$$

The above analysis extends beyond the piecewise linear response form in (2.1). For a simpler example, the piecewise constant regression model:

$$X_t = \begin{cases} \mu_1 + \epsilon_t, & 1 \leq t \leq c \\ \mu_2 + \epsilon_t, & c < t \leq n, \end{cases} \tag{4.3}$$

could be considered if one is sure there are no trends in the series. In this case, the $F_{\max}$ statistic in (2.9) would, under the null hypothesis of no undocumented changepoint, behave as the maximum of n dependent F statistics with 1 numerator degree of freedom and $n - 2$ denominator degrees of freedom. The 1 numerator degree of freedom is due to the difference between parameter counts in full (two parameters) and reduced (one parameter) models; 2 degrees of freedom are lost from n total in the denominator from estimation of μ_1 and

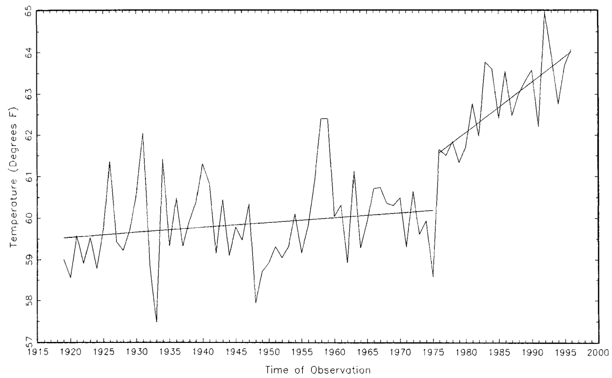

FIG. 3. Chula Vista annual temperatures.

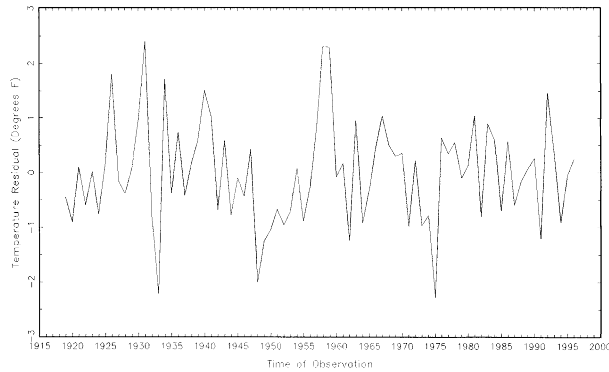

FIG. 5. Chula Vista annual residuals (with $c = 1975$).

μ_2. In this case, the $F_{\max}$ statistic is the square of a $t_{\max}$ statistic (cf. Hawkins 1977; Potter 1981; and Alexandersson 1986 for the statistical and climate origins).

In the next section, we will have cause to examine a series with the quadratic regression model:

$$X_t = \begin{cases} \mu_1 + \alpha_1 t + \beta_1 t^2 + \epsilon_t, & 1 \le t \le c \\ \mu_2 + \alpha_2 t + \beta_2 t^2 + \epsilon_t, & c < t \le n, \end{cases} \tag{4.4}$$

for the presence of undocumented changepoints. In this case, the $F_{\max}$ statistic would be distributed as the maximum of n-dependent F statistics each having 3 numerator degrees of freedom and $n - 6$ denominator degrees of freedom. The 3 numerator degrees of freedom represent the difference between parameter counts in full and reduced models; 6 degrees of freedom are lost from n total in the denominator due to estimation of the six regression parameters in the full model in (4.4).

Vincent (1998) presents examples where indicator structured response forms are appropriate for $E[X_t]$. Whereas quadratic and higher-order polynomial response functions for $E[X_t]$ do not typically admit explicit parameter estimates such as those in (2.2) and (2.3), sum of squared errors are easily computed by writing the regression in a general linear model form. This is all that is needed to compute the F statistics in (2.5).

Eventual application of the above methods to monthly and more general periodic series is envisioned (see also Gullet et al. 1991). For such pursuits, forms for $E[X_t]$ with a periodic component would be more appropriate. General problems with periodic time series in climatology are discussed in Lund et al. (1995). For temperature series, we expect changepoints to be easier to detect during summer, when variability is lower, than during winter, when variability is higher (see Gullet et al. 1991).

A series could have multiple undocumented changepoints. Clearly, one could have three or more phases in (2.1). In practice, one could eliminate the most prominent changepoint first, adjust the series for the mean shift due to that changepoint, and then successively test for further changepoints. This process would be repeated iteratively until all changepoints are identified. Alternatively, percentiles for the joint distribution of the largest and second largest values among $F_1, \ldots, F_n$ could be investigated.

The $F_{\max}$ percentiles in Table 1 were computed for errors $\{\epsilon_t\}$ that are independent and normally distributed (Gaussian). Due to the Central Limit theorem, normality is plausible for monthly or yearly series that are aggregated from daily values via summation or averaging. If

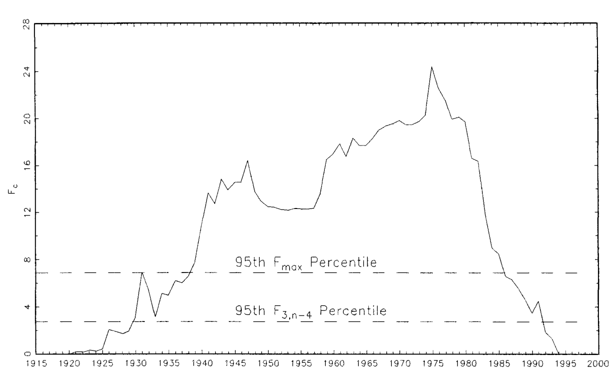

FIG. 4. Chula Vista annual F statistics.

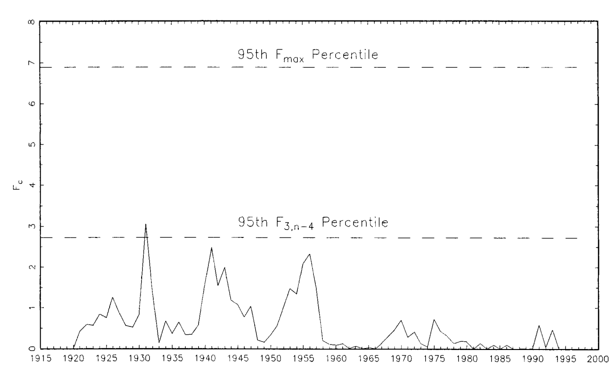

FIG. 6. Chula Vista annual residual F statistics.

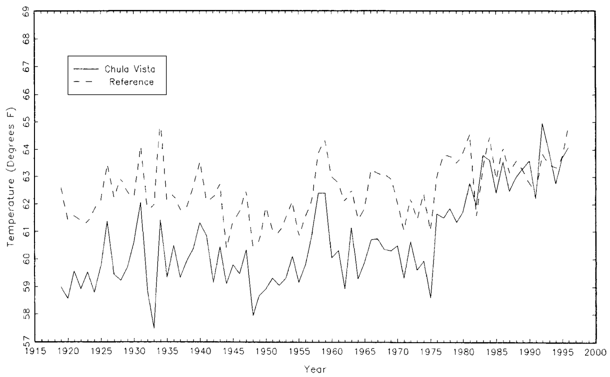

FIG. 7. Chula Vista and reference temperatures.

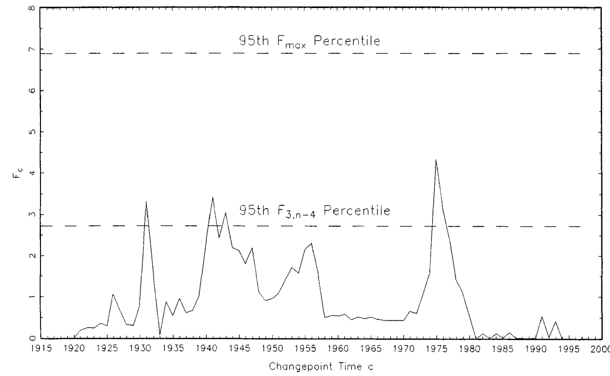

FIG. 9. Chula Vista minus reference residual F statistics.

$\{\epsilon_t\}$ is independent but non-Gaussian, then the percentiles in Table 1 are approximations, becoming exact as $n \to \infty$. In practice, the Table 1 percentiles are very reasonable for large n unless the distribution of ϵ_t is heavy tailed (having infinite higher-order moments). A more serious practical concern involves correlations in $\{\epsilon_t\}$. Heavily correlated errors could change the Table 1 percentiles. It would be a serious endeavor to investigate dependence of the Table 1 percentiles on autocorrelations in $\{\epsilon_t\}$; we do not attempt such a study here.

5. Application of the methods

This section presents two applications of the above methods to climatic series. Our first application, intended as a base check of the methods, considers a series where metadata is available. The second application moves to a more nebulous situation.

The first series we investigate contains temperatures from Chula Vista, California, for the 78-yr period 1919–96 inclusive. The annual averages of this data are plotted in Fig. 3. The accompanying metadata reveal three possible changepoints during this period. Specifically, the station began operation in 1919 at longitude 117°6′, latitude 32°36′, at an elevation of 9 ft with a cotton region shelter with maximum and minimum thermometers. In 1966, the instrumentation was replaced with a thermograph. In 1982, the station was physically moved to longitude 117°5′, latitude 32°37′, and the elevation changed to 56 ft. In 1985, the station reverted to using maximum and minimum thermometers. Hence, changepoints in 1966, 1982, and 1985 are plausible, although the station physically moved only once (in 1982).

It is instructive to proceed as if the changepoint times are unknown. Figure 4 shows the F statistics of the annual data for the entire 78-yr record. The dashed lines depict the 95th percentiles of the $F_{3,n-4}$ and $F_{\max}$ distributions (2.827 and 6.942, respectively). Both the $F_{3,n-4}$ and $F_{\max}$ 95th percentiles strongly suggest that one or more changepoints exist. The maximum F statistic occurs at 1975—between the times of the first two candidate changepoint times—with $F_{\max} = 24.36$. The estimated mean response (an estimate of $E[X_t]$ with $c = 1975$) is plotted against the yearly data in Fig. 3; this piecewise line appears to fit the data well.

The residuals in (4.1) were computed and are plotted in Fig. 5. The F statistics for this series are plotted in Fig. 6. The 95th percentile of the $F_{3,n-4}$ distribution is exceeded by the F statistic at time 1931 whereas the 95th percentile of the $F_{\max}$ distribution is not exceeded by any F statistic. Hence, the $F_{3,n-4}$ criteria indicates another changepoint but the $F_{\max}$ criteria suggests that our analysis is complete. There is nothing in the me-

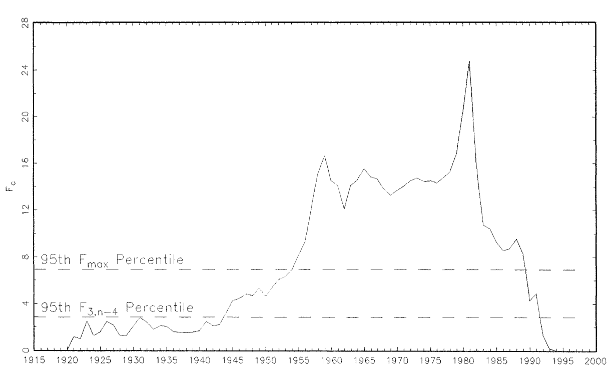

FIG. 8. Chula Vista minus reference F statistics.

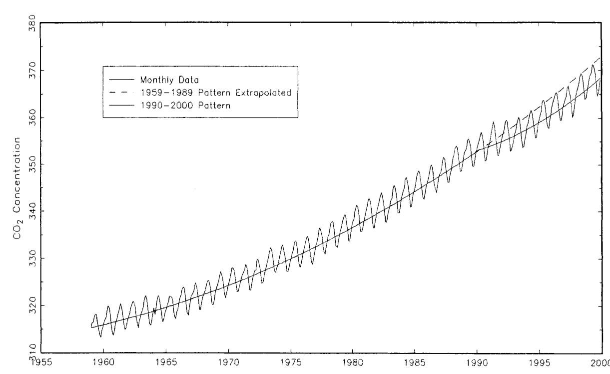

FIG. 10. Mauna Loa monthly carbon dioxide concentrations.

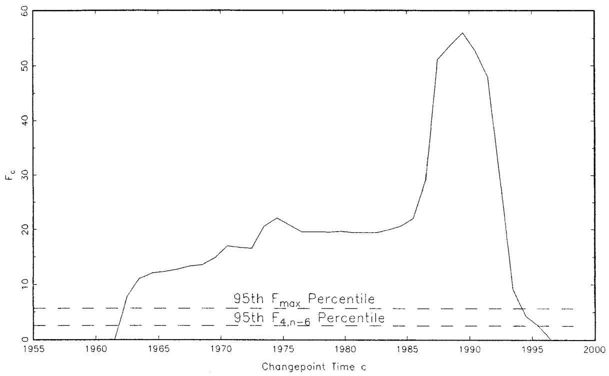

FIG. 11. Mauna Loa annual F statistics.

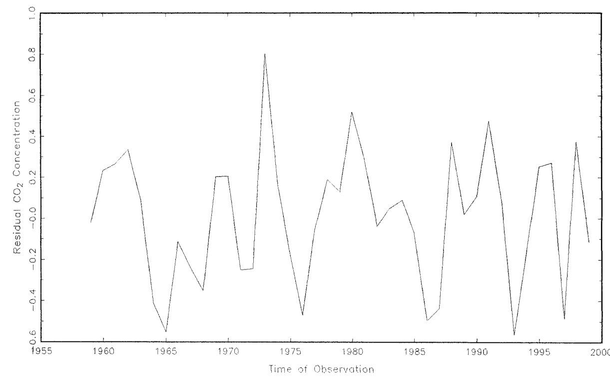

FIG. 12. Mauna Loa annual residuals (with $c = 1989$).

tadata to suggest a changepoint around 1931. In fact, the reference series comparisons below suggest that the 1931 changepoint is more rooted in chance variation than an actual changepoint. Overall, one sees the importance of the methods used; specifically, a $F_{3,n-4}$ criteria would lead to more changepoints, which in this case appear fictitious.

It is instructive to compare the above analysis to that obtained from reference series comparisons. A reference series $\{R_t\}$ was constructed by averaging five nearby series from stations at Avalon, Cuyamaca, Indio, Irvine, and Redlands, California. These stations report reasonably high quality data during the entire Chula Vista period of record. This reference series is plotted in Fig. 7 against the Chula Vista data for comparison's sake. Notice how the two series track well through the 1970s, with a change in their relative behavior shortly thereafter. The variability of the reference series is less than that of the Chula Vista series due to averaging. Of course, averaging reduces variability in general.

The F statistics for $\{X_t - R_t\}$ are plotted in Fig. 8. Both $F_{3,n-4}$ and F_{max} methods suggest the presence of a changepoint at time $c = 1981$ (very close to the 1982 station location move). The residuals in (4.1) were computed for $\{X_t - R_t\}$ with $c = 1981$. Figure 9 plots the F statistics of these residuals. Here, the $F_{3,n-4}$ 95% threshold is exceeded by *four F* statistics, whereas the corresponding F_{max} threshold is not exceeded anywhere. Again F_{max} and $F_{3,n-4}$ methods arrive at conflicting conclusions, with $F_{3,n-4}$ methods suggesting additional changepoints. As there is no physically compelling reason to declare a changepoint in the 1930s–40s, or in 1975 (especially in view of the closeby 1981 changepoint), we attribute the exceedence of the $F_{3,n-4}$ threshold to an unrealistic method.

Before continuing, it is worth making some points. First, if one knows the changepoint times beforehand, then t- and F-based sample mean differences comprise good statistical methods (the usual disclaimers on correlations apply). But if the changepoint times are truly unknown, one should revert to a F_{max} paradigm. For given a long enough series, it is not surprising to find an arbitrary preset short-length pattern somewhere in the series, if one is allowed to look anywhere for the pattern. From Table 1, it is clear that any changepoint declared by an F_{max} method will also be a changepoint in an $F_{3,n-4}$ analysis. The higher threshold of the F_{max} method precisely accounts for the extra variability due to the unknown changepoint time. In truth, one has candidate changepoint times for the Chula Vista series; a fair analysis would make use of this information.

Second, the residuals of $\{X_t\}$ were partitioned into pre- and post-1975 segments and the residuals of $\{X_t - R_t\}$ were partitioned into pre- and post-1982 segments. Each of these four segments were analyzed for additional changepoints. No further changepoints were found; hence, we omit the details of this analysis. Finally, it is worth noting that the instrumentation changes of 1966 and 1985 did not give rise to formal changepoints.

Now consider carbon dioxide concentrations reported at Mauna Loa, Hawaii, during the 41-yr period 1959–99 inclusive. Figure 10 displays the monthly values of this series. Annual averages were computed to eliminate the periodicities in the series; the F statistics with the quadratic model in (4.4) were computed for these annual averages and are displayed in Fig. 11. Justification for a nonzero quadratic component in this data through the year 1990 is contained in Lund et al. (1995). There is

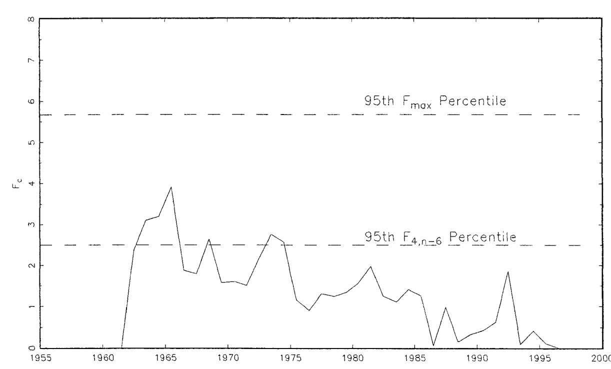

FIG. 13. Mauna Loa annual residual F statistics.

no nearby reference station available for this data over the entire period of record.

For the Mauna Loa series, an observed F_{max} statistic of 56.02 was obtained. As the trend in this data shows quadratic behavior, the true null hypothesis distribution of F_{max} is that of 41 dependent F statistics, each of which has 3 numerator degrees of freedom and $n - 6 = 35$ denominator degrees of freedom. The 95th percentile of the F_{max} distribution in this case is 5.67, whereas the 95th percentile of the $F_{4,n-6}$ distribution is 2.64. Both methods suggest that an unobserved changepoint is present in the late 1980s to early 1990s, with the maximum F statistic occurring at 1989. The estimated mean response function for this fit with $c = 1989$ is plotted against the monthly data in Fig. 10 and appears to fit the data well. Notice that the changepoint is not a discontinuity, but rather represents a slowing in the rate of increase. A continuation of the pre-1990 pattern is plotted in Fig. 10 for comparison's sake. Whereas the metadata do not suggest a changepoint in the late 1980s or early 1990s, it is plausible that the effects of the Mount Pinatubo volcanic eruption are being seen here. Of course, causes of this changepoint are debatable [see the Intergovernmental Panel on Climate Change's report (IPCC 2001) for additional discussion].

Figures 12 and 13 display the annual residuals [versions of $\{X_t^*\}$ computed with the quadratic response in (4.4) and $c = 1989$] and their corresponding F statistics. As no F statistic for the residual series exceeds the 95th F_{max} percentile, one concludes that the piecewise quadratic trend response in (4.4) with $c = 1989$ is reasonable and that there are no further changepoints in the data. One should note that use of the F criteria, rather than an F_{max} criteria, would lead to the inclusion of additional unwarranted changepoints.

6. Conclusions

Use of the $F_{3,n-4}$ criteria can lead to inclusion of many undocumented changepoints that are, in truth, fictitious. We believe that this has indeed happened with many temperature and climate series. We suggest that series for which undocumented changepoints have been detected by the $F_{3,n-4}$ criteria be reevaluated using the F_{max} criteria of this article.

Acknowledgments. The authors acknowledge support from the National Science Foundation Grants DMS 0071383 and DMS9895344. We thank Lynne Seymour and Hicham El-Hassani for useful conversations on the topic. We are indebted to the two referees of this work; their comments greatly improved this manuscript.

REFERENCES

Alexandersson, H., 1986: A homogeneity test applied to precipitation data. *J. Climatol.,* **6,** 661–675.

Brown, R. L., J. Durbin, and J. M. Evans, 1975: Techniques for testing the constancy of regression relationships over time (with discussion). *J. Roy. Stat. Soc.,* **37B,** 149–192.

Chen, J., and A. K. Gupta, 2000: *Parametric Statistical Change Point Analysis.* Birkhauser, 240 pp.

Easterling, D. R., and T. C. Peterson, 1995: A new method for detecting undocumented discontinuities in climatological time series. *Int. J. Climatol.,* **15,** 369–377.

Gullet, D. W., L. Vincent, and L. Malone, 1991: Homogeneity testing of monthly temperature series: Application of multi-phase regression models with mathematical changepoints. Canadian Climate Centre Rep. 91-10 Atmospheric Environment Service, Downsview, ON, Canada, 47 pp.

Hawkins, D. M., 1977: Testing a sequence of observations for a shift in location. *J. Amer. Stat. Assoc.,* **72,** 180–186.

Hinkley, D. V., 1969: Inference about the intersection in two-phase regression. *Biometrika,* **56,** 495–504.

——, 1971: Inference in two-phase regression. *J. Amer. Stat. Assoc.,* **66,** 736–743.

IPCC, cited 2001: Annual Report. [Available online at http://www.ipcc.ch/pub/reports.html.]

Jones, P. D., S. C. B. Raper, R. S. Bradley, H. F. Diaz, P. M. Kelly, and T. M. L. Wigley, 1986: Northern Hemisphere surface air temperature variations: 1851–1984. *J. Climate Appl. Meteor.,* **25,** 161–179.

Kander, Z., and S. Zacks, 1966: Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points. *Ann. Math. Stat.,* **37,** 1196–1210.

Karl, T. R., and C. N. Williams Jr., 1987: An approach to adjusting climatological time series for discontinuous inhomogeneities. *J. Climate Appl. Meteor.,* **26,** 1744–1763.

Leadbetter, M. R., G. Lindgren, and H. Rootzen, 1983: *Extremes and Related Properties of Random Sequences and Processes.* Springer-Verlag, 336 pp.

Lund, R. B., H. L. Hurd, P. Bloomfield, and R. L. Smith, 1995: Climatological time series with periodic correlation. *J. Climate,* **8,** 2787–2809.

——, L. Seymour, and K. Kafadar, 2001: Temperature trends in the United States. *Environmetrics,* **12,** 1–18.

Page, E. S., 1955: A test for a change in a parameter occurring at an unknown point. *Biometrika,* **42,** 523–527.

Potter, K. W., 1981: Illustration of a new test for detecting a shift in mean in precipitation series. *Mon. Wea. Rev.,* **109,** 2040–2045.

Rhoades, D. A., and M. J. Salinger, 1993: Adjustment of temperature and rainfall records for site changes. *Int. J. Climatol.,* **13,** 899–913.

Solow, A. R., 1987: Testing for climate change: An application of the two-phase regression model. *J. Climate Appl. Meteor.,* **26,** 1401–1405.

Thompson, C. S., 1984: Homogeneity analysis of rainfall series: An application of the use of a realistic rainfall model. *J. Climatol.,* **4,** 609–619.

Vincent, L. A., 1998: A technique for the identification of inhomogeneities in Canadian temperature series. *J. Climate,* **11,** 1094–1104.