A Technique for the Identification of Inhomogeneities in Canadian Temperature Series

LUCIE A. VINCENT

Climate Research Branch, Atmospheric Environment Service, Downsview, Ontario, Canada

(Manuscript received 23 September 1996, in final form 13 August 1997)

ABSTRACT

A new technique has been developed for the identification of inhomogeneities in Canadian temperature series. The objective is to identify two types of inhomogeneities—nonclimatic steps and trends—in the series of a candidate station in the absence of prior knowledge of the time of site changes and to properly estimate their position in time and their magnitude. This new technique is based on the application of four linear regression models in order to determine whether the tested series is homogeneous, if there is a nonclimatic trend, a step, or trends before and/or after a step. The dependent variable is the series of the candidate station and the independent variables are the series of some neighboring stations. Additional independent variables are used to describe and measure steps and trends existing in the tested series but not in the neighboring series. After the application of each model, the residuals are analyzed in order to determine the fit of the model. If there is significant autocorrelation in the residuals, nonidentified inhomogeneities are suspected in the tested series and a different model is applied to the datasets. A model is finally accepted when the residuals are considered to be random variables. The description of the technique is presented along with some evaluation of its ability to identify inhomogeneities. Results are illustrated through the provision of an example of its application to archived temperature datasets.

1. Introduction

Reliable climatological time series are essential for the analysis of climate trends, climate variability, and for the detection of anthropogenic climate change. Research scientists have assembled climate datasets in order to attempt to identify regional, hemispheric, and global climate change (International Panel on Climate Change 1992). It is well known that most of the long-term climate datasets contain variations due to nonclimatic factors such as site relocations, urbanization, and many others. These nonclimatic factors often create inhomogeneities of various magnitudes and at different positions in time. Using datasets that are not adjusted for these nonclimatic variations can seriously affect the correct assessment of climate trends and variability, as well as interfere with the identification of any real climate change signal. Therefore, it becomes important to develop and improve techniques to properly identify and adjust for nonclimatic variations.

Preliminary homogeneity assessment of Canadian temperature series has indicated that most common types of inhomogeneities are nonclimatic steps and trends (Gullett et al. 1991). Steps are abrupt changes in the mean level and they are usually caused by relocation of the station, replacement of the instruments in use, or by some alteration in observing procedures. Trends are gradual increasing or decreasing changes over time, and they are usually associated with a cumulative natural or human-induced phenomenon such as the growth of the surrounding vegetation or the urbanization of the nearby area. The major causes of inhomogeneities in climatological series are well documented (Gullett et al. 1990; Karl and Williams 1987; Jones et al. 1986). Correction for small steps when in fact a trend exists in the series may lead to inappropriate climate representation.

Station history reports are maintained for each observing site. They are files detailing important information such as site location, instruments in use, height of instruments above the ground, and other pertinent characteristics of the observing practices. Some of these are available to some extent in digital form. Most are still available as paper copy only, especially for the earlier periods of time (Gullett et al. 1991). Such reports can be useful for the identification of inhomogeneities but they are frequently lacking essential information and they may not routinely document all causes of inhomogeneities. A technique that identifies the date of an inhomogeneity without knowing a priori the real time of a change at the station and that properly estimates the magnitude of the identified change is highly desirable.

A new technique for the identification of inhomogeneities in Canadian temperature series is presented. It has been developed with the following objectives.

Corresponding author address: Lucie A. Vincent, Climate Research Branch, Atmospheric Environment Service, 4905 Dufferin St., Downsview, Ontario M3H 5T4, Canada.
E-mail: Lucie.Vincent@ec.gc.ca

First, homogeneous and inhomogeneous intervals of time are identified in order to focus the search for inhomogeneities and to determine periods of time that can be used without alteration. Second, nonclimatic steps and trends are detected separately with a proper estimate of their position in time, magnitude, and statistical significance. Finally, the most probable time of occurrence of the inhomogeneity is identified in the absence of prior knowledge of the real time of change at the station in order to minimize the dependency on station history files. Considerable testing on simulated series shows that these objectives are met to a large extent; however, further testing is still required to fully determine the limitations of this new technique.

2. Background

In the literature, a number of techniques are described to detect inhomogeneities in climatological datasets (Easterling and Peterson 1995; Gullett et al. 1991; Sneyers 1989; Karl and Williams 1987; Alexandersson 1986; Jones et al. 1986; Mitchell 1961). Most of the techniques are based on the comparison of a candidate series with some reference series, and patterns are identified in the relationships between the series. However, objectives and identification processes vary from one technique to another. As an example, Jones et al. (1986) use a visual technique to identify the major inhomogeneities, and corrections are intended to be general adjustments only, suitable for continental or hemispheric-scale studies but not necessarily sufficient for local-scale studies. On the other hand, Karl and Williams (1987) use the station history reports to identify all potential discontinuities. This is a laborious procedure for the detection of inhomogeneities and frequently these reports do not provide sufficient information for the proper identification of all nonclimatic changes; ultimately, however, more local-scale climate change characteristics can be investigated using Karl's series.

An approach for homogeneity assessment of temperature series was previously investigated (Gullett et al. 1990) and an initial procedure based on multiple-phase regression models was developed (Vincent 1990). Regression models were used to identify multiple changepoints in the tested series, corresponding to potential inhomogeneities, and the F test was used to determine whether the changes were statistically significant. It was applied to test the homogeneity of the maximum and minimum temperature series of over 350 Canadian stations and the results are presented in Gullett et al. (1991). At the time, the objective was to assess only the homogeneity of the series; precision regarding location and magnitude of each inhomogeneity was not essential since data adjustments were not performed.

Another methodology for detecting undocumented discontinuities in climatological time series was recently presented by Easterling and Peterson (1995). Their main objective was to develop a methodology for application to global climate datasets that do not have adequate station history information for most stations. It is based on regression models with the difference between the base and reference series as the dependent variable and time as the independent variable. A two-phase regression approach is used to identify the position of the discontinuity and the significance of the change is established using the F test. The series is subdivided at the potential step and each subsection is then tested separately. The magnitude of the step is provided by the difference in the means of the difference series before and after the potential step, and the significance of each discontinuity is tested using multiresponse permutation procedures. Finally, the adjusted series is retested to verify that there are no more inhomogeneities. Easterling and Peterson have demonstrated that their methodology performed reasonably well (Easterling and Peterson 1995). They have as well indicated that homogenized datasets should be used with caution since different methodologies for the identification and correction of climate datasets could generate different adjustments and therefore could produce different analyses and results (Easterling et al. 1996).

The technique presented in this report is also based on the application of regression models. However, it differs from previous techniques in a number of ways. First, four regression models are used to identify the following situations in the tested series: whether it is homogeneous, if there is a nonclimatic trend, a step, or trends before and/or after a step. Second, the variables in the models are different: the dependent variable is the series of the candidate station and the independent variables are the series of a number of neighboring stations. Independent variables are also used to describe and measure steps and trends in the tested series; therefore, the magnitude of each inhomogeneity is provided by its corresponding estimate parameter. Finally, the most important difference from previous methods is that it uses the autocorrelation in the residuals to determine whether there is any inhomogeneity. Considerable testing of this technique has been done on simulated series, and recommendations for data adjustment are based on these results. Ongoing research and development of new techniques for the assessment and correction of climate datasets is important and should continue. The need for reliable climate datasets will increase as the impacts of present and future climate site automation will become evident.

3. Description of the technique

The technique consists of the application of four linear regression models. The first model determines whether the candidate series, also called the base series, is homogeneous for the tested interval of time. If it is homogeneous then the procedure stops and the three remaining models are not used. Otherwise, the second model is fitted to the data to establish if there is an overall trend in the base series. If the inhomogeneity is not an overall trend, then the third model is applied to identify a single step change. Finally, it might be necessary to apply the fourth model in order to determine if there are trends before and after the step, which provides an indication of multiple inhomogeneities in the base series. In this case, the series is subdivided at the position in time of the identified step and each segment is tested separately starting with the first model.

a. Model 1: Description of a homogeneous series

The following regression model is first considered to describe the base series:

$$y_i = a_1 + c_1x_{1i} + d_1x_{2i} + f_1x_{3i} + e_i \qquad i = 1, \ldots, n. \tag{1}$$

The dependent variable y_i is the temperature at the base station at time *i*. The independent variables x_{1i}, x_{2i}, x_{3i} are the temperatures of three reference series at time *i*. The reference series are usually the temperature at some neighboring stations. Since the number of neighbors varies from one base to another, the number of independent variables corresponding to the reference series varies from one situation to another. The reference series can also be a computed series derived from a number of stations such as the mean or the median of the neighboring series. In any case, the reference series should reflect as much as possible temperature variations only. In Canada, mostly in the northern part of the country, there is a large number of stations that can be considered isolated, with very few (perhaps two to four) distant neighbors to use for homogeneity testing (Gullett et al. 1991). For illustrative purposes only, three reference series are used in the equations presented in this report.

The parameters a_1, c_1, d_1, f_1 are estimated using the least squares method. The residuals e_i are the differences between the values of the base series and the fitted values given by the model. The common assumption is that the residuals are independent, normal, random variables with mean zero and constant variance σ^2 when the model describes correctly the data. However, since annual mean temperatures are often slightly correlated in time (see section 4), it is expected to have small serial correlation in the residuals. This problem is addressed in section 5a.

An informal way to evaluate the aptness of the model, or in other words to check if the model provides adequate description of the data, is to examine visually the graph of the residuals (Neter et al. 1985). When the residuals fluctuate in a random pattern around the zero line, the model is appropriate for the data, and in this case, the base is considered homogeneous (Fig. 1). On the other hand, if for a certain period of time they appear to be all on one side of the line and then all on the other side, the graph indicates the poor fit of the model. This change of pattern in the residuals can be associated with

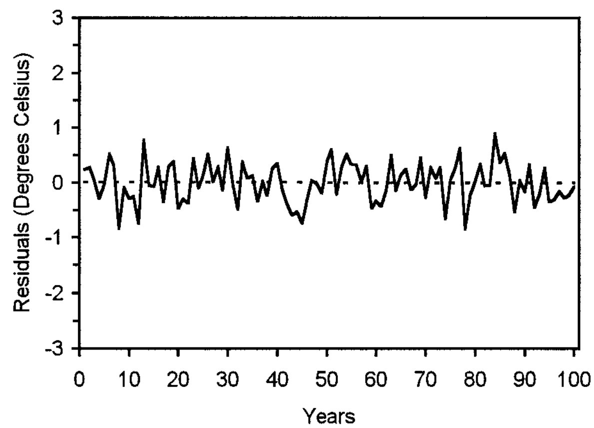

FIG. 1. Example of the residuals or differences between the values of a simulated homogenous base series and the fitted values from model 1. The dashed line represents the mean of the series.

an inhomogeneity at the base series. When such a situation occurs, it generates significant correlation between successive residual values, which is also called autocorrelation in the residuals.

The Durbin–Watson test has been developed to determine the statistical significance of the autocorrelation of lag one in the residuals, which is the correlation between residuals one distance apart (Neter et al. 1985; Draper and Smith 1981). The random error terms following a first-order autoregressive process are given as follows:

$$e_i = \rho e_{i-1} + u_i \qquad i = 1, \ldots, n,$$

where $|\rho| < 1$ and u_i are independent $N(0, \sigma^2)$. The usual test alternatives considered are H_0: $\rho = 0$ versus H_a: $\rho > 0$. The Durbin–Watson D statistic is calculated as follows:

$$D = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}.$$

Lower and upper bounds, d_1 and d_u, were obtained such as a value of D outside of these bounds leads to a definite decision: if $D > d_u$, conclude H_0; if $D < d_1$, conclude H_a; if $d_1 \leq D \leq d_u$, the test is inconclusive.

Inhomogeneities in the base series generate as well autocorrelation for residuals being several distances apart. An approach to determine the statistical significance of the autocorrelation at different lags has been suggested by Chatfield (1984). The autocorrelation coefficient at lag k is computed as follows:

$$r_k = \frac{\sum_{i=1}^{n-k} (e_i - \overline{e})(e_{i+k} - \overline{e})}{\sum_{i=1}^{n} (e_i - \overline{e})^2},$$

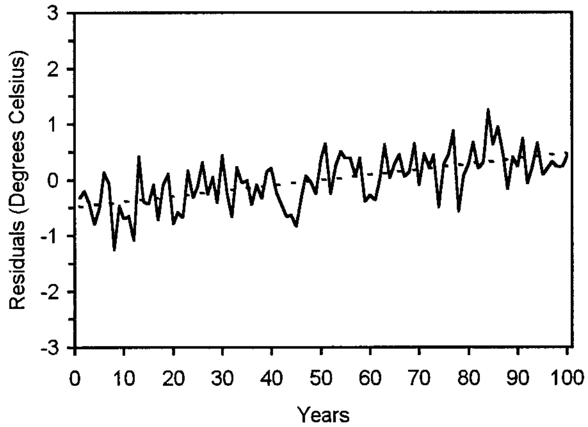

FIG. 2. Example of the residuals or differences between the values of a simulated base series with an overall trend and the fitted values from model 1. The dashed line represents the best-fit linear trend.

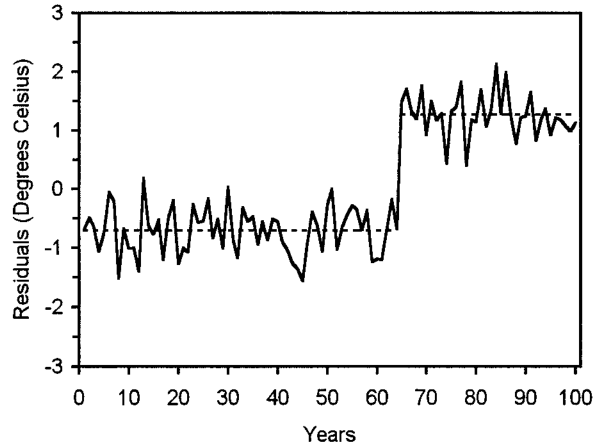

FIG. 3. Example of the residuals or differences between the values of a simulated base series with a step and the fitted values from model 1. The dashed line represents the mean of each segment.

where $\overline{e}$ is the residuals' average for n observations, $r_0 = 1$, and $|r_k| < 1$ for $k > 0$. The correlogram, which is the graph of the autocorrelation r_k against lag k, is obtained with its approximate 95% confidence interval given by $\pm 2n^{-1/2}$. Values of r_k outside the confidence interval are significantly different from zero at the 5% level. Consecutive significant values of r_k identified at low lags provide an indication of potential inhomogeneities in the base series.

The results provided by the Durbin–Watson test and by the correlogram are used in the following way. If there is no significant autocorrelation in the residuals, then the current model describes adequately the series. In this case, model 1 is accepted and the base series is homogeneous for the tested period of time. However, if there is significant autocorrelation in the residuals, then a different model is applied in order to identify the inhomogeneities in the base series.

b. Model 2: Description of a trend

When there is an overall trend in the base series, which does not occur in the reference series, the graph of the residuals after the application of model 1 shows a gradual change from one side of the zero line to the other (Fig. 2). To explain this gradual change, a variable representing a trend is introduced in model 1. The addition of this new independent variable forms a second model, which is given as follows:

$$y_i = a_2 + b_2 i + c_2 x_{1i} + d_2 x_{2i} + f_2 x_{3i} + e_i \qquad i = 1, \ldots, n. \tag{2}$$

The independent variable i takes successively the values 1 to n. It represents a position in time such as, for example, the first year, the second year, and up to the n year when annual values are used as variables. The parameter b_2 represents the slope of the regression line over time. The Durbin–Watson test and the correlogram are used to determine the significance of the autocorrelation in the residuals. If model 2 adequately describes the series, the significance of the trend is established using the common t statistic. The procedure stops and there is no need for further analysis. However, if there is significant autocorrelation in the residuals, consideration is given to the application of a different model.

c. Model 3: Description of a step

When a step occurs in the base series but not in the reference series, the graph of the residuals against time shows an abrupt change of level (Fig. 3). To describe the step in the base series, an independent variable, I, is introduced in model 1 instead of the variable, i, representing the trend. This new model is then written as follows:

$$y_i = a_3 + b_3 I + c_3 x_{1i} + d_3 x_{2i} + f_3 x_{3i} + e_i \qquad i = 1, \ldots, n. \tag{3}$$

The parameter b_3 provides the magnitude of the step. The independent variable I takes the following values:

$$I = 0 \quad \text{for} \quad i = 4, \ldots, p - 1,$$
$$I = 1 \quad \text{for} \quad i = p, \ldots, n - 3.$$

The value p is the position in time of a potential step change and it is called a changepoint. In theory, the changepoint p can take any value between and including 1 and n. However, in practice it is necessary to have some information before and after the step to correctly determine its position and magnitude. It is arbitrarily decided that the value of p will range from position 4 to $n - 3$, leaving three positions at the beginning and at the end of the series. The ability of the technique to properly identify a step at any position in time is discussed in section 5b.

Since the position of the step is unknown a priori, it is necessary to establish a procedure to identify the changepoint p. The approach consists of finding the value p that provides the minimum residual sum of squares for model 3. This idea came from a paper on climate change testing (Solow 1987) in which a test for detecting a change in the behavior of an annual temperature series based on a two-phase regression model is presented. This approach was as well used in our previous homogeneity assessment (Gullett et al. 1991; Vincent 1990). Model 3 is fitted to the datasets successively for p equal to 4 to $n - 3$. This procedure produces in this manner a sequence of residual sum of squares (RSS) values. The minimum RSS indicates the model with the best fit and the corresponding estimated changepoint p, as denoted by $\hat{p}_1$, represents the most probable position in time of the step in the base series.

The significance of the autocorrelation in the residuals is established by the Durbin–Watson test and the correlogram. When there is a step in the base series, the variable I describes the change in the mean level. The estimated parameter b_3 gives the magnitude of the step and the usual t statistic is used to determine its statistical significance. On the other hand, if the autocorrelation is still significant, it is most likely that there is more than one inhomogeneity in the base series. Model 4 is then used in an attempt to better explain the pattern in the series of residuals.

To specifically test the null hypothesis of no change, model 1 is compared with a model describing a change, such as model 3, using the F test (Vincent 1990; Neter et al. 1985). When models 1 and 3 are fitted to the datasets, the residual sum of squares RSS1 and RSS3 are obtained, respectively. The hypothesis H_0: there is no change versus H_1: there is a change at $\hat{p}_1$ considered. The F statistic is calculated as follows:

$$F^* = \frac{(\mathrm{RSS1} - \mathrm{RSS3})/(\mathrm{DF1} - \mathrm{DF3})}{\mathrm{RSS3}/\mathrm{DF3}},$$

which, when H_0 holds, follows approximately the F distribution with DF3 − DF1 and DF3 degrees of freedom. DF1 is equal to $n - 4$ since four parameters are estimated in model 1 while DF3 is $n - 5$ because five parameters are estimated in model 3. Using a risk of type I error α, the decision rule is to reject H_0 if $F^* > F(1 - \alpha; 1, n - 5)$.

d. Model 4: Description of trends before and after a step

To identify additional inhomogeneities in the base series, model 4 is fitted to the datasets. The model describes the following pattern: a trend from the beginning of the series to a changepoint, a potential change of level, and a second trend from the changepoint to the end of the series (Fig. 4). This model is described as follows:

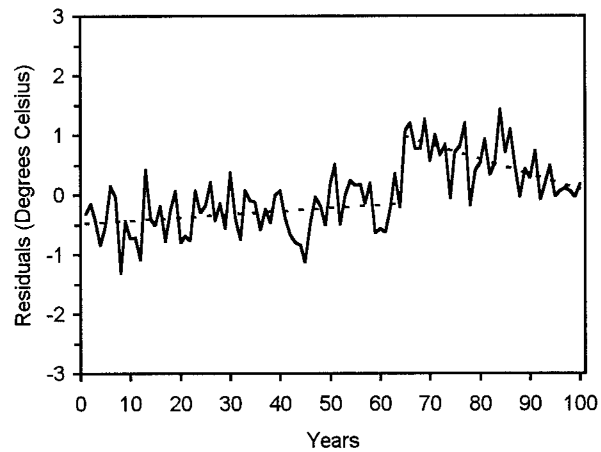

FIG. 4. Example of the residuals or differences between the values of a simulated base series with trends before and after a step and the fitted values from model 1. The dashed line represents the best-fit linear trend of each segment.

$$\begin{aligned} y_i &= a_4 + b_4 i I_1 + a_5 I_2 + b_5 i I_2 + c_4 x_{1i} + d_4 x_{2i} \\ &\quad + f_4 x_{3i} + e_i \qquad i = 1, \ldots, n \\ I_1 &= 1 \quad \text{and} \quad I_2 = 0 \quad \text{for} \quad i = 4, \ldots, p - 1 \\ I_1 &= 0 \quad \text{and} \quad I_2 = 1 \quad \text{for} \quad i = p, \ldots, n - 3. \end{aligned} \tag{4}$$

Model 4 is fitted to the datasets successively for p equal to 4 to $n - 3$. The best fit is determined by the minimum RSS and the corresponding estimated changepoint p, as denoted by $\hat{p}_2$, indicates the most probable position in time of the change.

The Durbin–Watson test and the correlogram are used once again to determine the significance of the autocorrelation of the residuals. If there is a change in the behavior of the trend, including or not a step, then model 4 adequately describes the inhomogeneities in the base series. The estimated parameters b_4 and b_5 correspond to the slopes before and after $\hat{p}_2$, respectively, and the associated t statistic is used to establish the significance of each trend. The estimated magnitude of the step at $\hat{p}_2$, denoted by $\hat{m}$, is calculated with the following equation:

$$\hat{m} = (a_4 + a_5 + b_5\hat{p}_2) - (a_4 + b_4(\hat{p}_2 - 1)). \tag{5}$$

If there is still significant autocorrelation after the application of model 4, then all inhomogeneities may not have been correctly identified in the base series. In this case, the base and reference series are subdivided at $\hat{p}_2$ and each segment is tested separately starting with model 1.

4. Simulation of annual temperature series

Simulated temperature datasets are used to test the ability of the technique to correctly identify inhomogeneities in a base series and to properly estimate their position in time as well as their magnitude. To ensure that simulated datasets are representative of some climate characteristics, it was first necessary to investigate real temperature series at a number of Canadian sites. The selection of stations was based on spatial coverage, length of record, and data completeness with missing values being replaced using a known technique (Thom 1966). Preliminary assessment of annual mean maximum and minimum temperature datasets indicated that the selected series were relatively homogeneous over the chosen period of time (Gullett et al. 1991). Statistical parameters such as means, standard deviations, autocorrelations, and linear trends were computed. The autocorrelation of lag one in the annual temperature series was small ranging from -0.2 to 0.2 with a very few exceptions at 0.3. This investigation of the real temperature series provided examples of the wide range of temperatures that are typical across the country, as well as the variability existing in the annual datasets. A detailed description of the different Canadian climates including temperature averages and extremes is presented in Phillips (1990). The general climatological pattern provided by the linear trends is in agreement with the findings of Skinner and Gullett (1993) showing cooling along the east coast and warming in the western regions of the country over the last four decades.

The methodology used for the creation of simulated annual temperatures series has been described by Easterling and Peterson (1992). It produces rough simulations of real temperature series sufficient for identification of inhomogeneities but not representative of real climate variations. It has been chosen for comparison purposes since other techniques have already been tested using this type of simulated datasets (Easterling and Peterson 1992). Random numbers normally distributed with mean 0.0 and standard deviation 1.0 are generated from an AR(1) model with an autoregressive parameter equal to 0.1, producing a series of 100 time elements (or 100 yr). This procedure is repeated four times to create four simulated series, one for the base and three for reference. Each reference series is then cross correlated with the base series at a coefficient level of 0.70–0.90 by multiplying the values of the base series by 1.5 and adding them to the values of the reference series. After this process, each reference series is restandardized. A constant value can be added to every element of the series to form simulated series of different overall means, and an increasing small value can be used to simulate an overall trend. Statistical parameters derived from a large number of simulated series were in general agreement with the parameters obtained from the investigation of the real temperature series.

5. Evaluation of the technique

To evaluate the performance of the technique and to determine its limitations, it is necessary to test and analyze a large number of simulated series representing a variety of situations. This section presents the results of some tests to define the situations in which it performs the best and to associate a level of confidence to its performance. Further testing is required for a more complete evaluation of this new technique.

a. Ability to identify a homogeneous series

Since the annual mean maximum and minimum temperatures are slightly correlated in time, it is expected to have small serial correlation in the residuals. The following test provides the percentage of times that significant autocorrelation is identified in the residuals when series slightly correlated in time are used as independent and dependent variables in the model.

Annual temperature series of 100 yr are simulated following the procedure described in section 4 (with an autocorrelation of about 0.1). No additional value is introduced in the base series at this time in order to simulate a homogeneous series. Correlation coefficients between the base and references vary from 0.70 to 0.90. Model 1 is fitted to the datasets and the Durbin–Watson test is applied to determine the statistical significance of the autocorrelation of lag one in the residuals. This process is repeated 1000 times. Results show that 13.6% of the time, the technique detects significant autocorrelation when no inhomogeneity is introduced in the base series. This suggests that 13.6% of the time, the null hypothesis of no change in the base series is rejected when in fact it is true (type I error). In these cases, models 2 and 3 are applied to identify the type and magnitude of the inhomogeneity. Significant steps greater than 0.4°C, ranging from 0.2°C to 0.4°C and from 0.0°C to 0.2°C are found in 1.2%, 3.3%, and 6.1% of the series, respectively; no significant step or trend is identified in 3.0% of the base series even if the autocorrelation of lag one is significant.

b. Ability to identify the position and magnitude of a step

A single step is introduced in the base series by adding a constant value to each element of the series for a specific interval of time. Ten positions have been selected as starting dates for the step, and in all cases, the ending date is the end of the series. They were chosen to further cover the beginning and the end of the series since it is expected that the technique would have more difficulty to correctly identify an inhomogeneity located near one of the extremities of the series. The magnitude varies from about one-quarter to twice the standard deviation imposed in each series. It is expected that smaller steps would be more difficult to identify than larger ones. Over all, a total of 80 situations are analyzed (10 positions by eight magnitudes). Each situation is tested separately 1000 times and each time a new set of simulated series is used. Results are summarized for the first five positions, since positions and results were found symmetrical.

TABLE 1. Percentage of simulated base series with the date of the step identified correctly, and identified correctly within two positions, for various positions and magnitudes of a step.

Magnitude	Position in time 5	10	15	20	35
a. Date identified correctly					
2.00	99.3	98.8	99.9	99.3	99.2
1.75	97.9	98.4	98.7	97.3	97.5
1.50	95.5	94.4	96.3	96.2	95.7
1.25	89.6	91.0	88.7	90.5	90.0
1.00	78.8	79.0	79.9	78.1	80.9
0.75	59.4	64.5	63.8	60.7	63.4
0.50	35.4	37.1	36.8	40.0	38.5
0.25	12.6	11.5	12.4	11.5	12.6
b. Date identified correctly to within two positions					
2.00	99.9	99.9	99.9	99.9	99.9
1.75	99.9	99.9	99.9	99.9	99.9
1.50	99.4	99.9	99.9	99.9	99.9
1.25	99.0	99.5	99.4	99.5	99.6
1.00	95.8	97.0	98.1	97.1	97.7
0.75	85.5	92.4	92.0	91.5	91.9
0.50	64.8	69.9	72.6	74.6	75.2
0.25	27.8	31.6	33.7	34.8	34.7

TABLE 2. Percentage of simulated base series with the magnitude of the step identified correctly to within 0.1°C, and within 0.2°C, for various positions and magnitudes of a step.

Magnitude	Position in time 5	10	15	20	35
a. Magnitude identified correctly to within 0.1°C					
2.00	38.5	52.4	61.8	68.5	75.5
1.75	37.2	49.8	63.3	67.7	75.2
1.50	36.1	53.2	63.0	65.7	74.0
1.25	37.9	51.8	61.8	67.9	75.1
1.00	37.6	51.4	59.6	66.4	78.0
0.75	33.5	49.9	61.7	68.3	73.8
0.50	29.7	49.8	56.4	66.8	76.7
0.25	35.6	44.4	53.4	61.4	66.5
b. Magnitude identified correctly to within 0.2°C					
2.00	67.7	84.7	91.5	94.6	97.8
1.75	68.2	83.1	92.5	95.2	98.3
1.50	65.4	84.4	91.1	95.3	97.6
1.25	69.8	85.6	90.2	93.9	98.0
1.00	64.7	83.6	92.0	95.1	98.5
0.75	61.0	82.5	91.2	95.5	97.9
0.50	55.8	83.0	88.2	92.3	98.5
0.25	53.9	66.1	78.0	85.4	92.5

To measure the similarity between the series before the application of the technique, correlation coefficients are calculated from the first difference series (Peterson and Easterling 1994). The first difference series is the series of the differences between each element of the series and its previous value. It is a special type of filter for removing steps and linear trends in a time series (Box and Jenkins 1976). In these simulations, the correlation coefficients on the first difference series between base and reference stations range from 0.70 to 0.90.

Table 1 presents the percentage of simulated base series with the date of the step identified correctly, and identified correctly to within two positions (or years), respectively. The results of Table 1a show that the technique correctly identifies the date of a step of magnitudes ranging from 1.25° to 2.00°C almost every time (88.7%–99.9%) at all positions. The correct date of a step of 0.75° and 1.00°C is identified most of the time (59.4%–80.9%), and the correct date of a step of 0.25° and 0.50°C are not frequently identified. Table 1b shows that the technique identifies correctly to within 2 yr the date of a step of 0.75°–2.00°C almost every time (85.5%–99.9%). The correct date of a step of 0.50°C is frequently identified within 2 yr (64.8%–75.2%), and a step of 0.25°C is not identified very often. Contrary to expectation, the technique does not have much more difficulty to retrieve the date near the extremities of the series.

In Table 2, the percentage of simulated base series with the magnitude of the step identified correctly to within 0.1°C and to within 0.2°C is presented. The results show that the technique has more difficulty to identify the correct magnitude of the step as its position gets closer to one of the extremities. In Table 2a, when the step is located near the middle of the series, such as for positions 35 (or 65), its magnitude within 0.1°C is frequently identified (66.5%–78.0%). In Table 2b, the magnitude of the step to within 0.2°C is correctly identified almost all the time (83.0%–98.5%) for magnitudes ranging from 0.50° to 2.00°C and for positions ranging from 10 to 90.

From these results, it is concluded that when the first difference series are cross correlated at a coefficient level between 0.70 and 0.90, the technique correctly identifies, at least 82.5% of the time, the date of the step within 2 yr, and the magnitude of the step within 0.2°C, for positions ranging from 10 to 90, and magnitudes ranging from 0.75° to 2.00°C. The position and magnitude of steps of 0.50°C are frequently identified, and the position of steps of 0.25°C is not correctly identified very often.

c. Remarks

Various techniques for detecting and adjusting for artificial discontinuities in climatological time series were presented by Easterling and Peterson (1995, 1992). One of their objectives was to evaluate and compare the ability of different techniques to properly identify the year of a discontinuity. Rigorous comparison between other techniques and ours is not presented here. Results from testing simulated datasets show that steps equal to or greater than 1.0°C are well identified by all techniques. This finding is not surprising since the standard deviation of all simulated series is about 1.0, references are well correlated with the base, and steps of 1.0°C are relatively large under these conditions. However, all techniques have much more difficulty to correctly identify steps of 0.5°C; therefore, caution should be used when adjusting steps of that magnitude.

Inhomogeneities in the reference series can obscure the proper identification of inhomogeneities in the base series. Preliminary testing on simulated series revealed that this technique identifies correctly the position of the step in the base even if steps are introduced in one or more reference series; however, as the steps in the reference series get larger, the magnitude of the identified step in the base becomes incorrect. In application to real climate datasets, this problem is minimized by examining the graphs of the differences between the base and the reference series. By comparing station pairs, inhomogeneities belonging to the base and those associated with the neighboring series become apparent (Jones et al. 1986). Therefore, it is possible either to correct large inhomogeneities in neighboring series before testing the base or to discard neighboring series that appear to have too many problems. To eliminate the impact of inhomogeneities in reference series, Peterson and Easterling (1994) have developed a methodology to create a representative climate reference series; however, this method was not considered in the present study.

Trends occurring in the base series and not in the reference series can also be detected by this technique. Preliminary testing on simulated series indicates that the beginning date and the magnitude of a trend are correctly identified when a step occurs at the beginning of the trend. Changepoints associated only with a change in magnitude or direction of a trend are not always detected. In application, it is often difficult to retrieve from the station history files the cause of a human-induced trend. Further testing is required to fully determine the ability of the technique to identify nonclimatic trends. As well, for a more complete evaluation of this technique, further tests are needed, for example, to determine the joint statistical significance when multiple inhomogeneities are identified in the base series. However, even if the precise size of the tests are unknown, the objectives of this technique are served nonetheless.

6. Example of application to real datasets

A procedure for homogeneity assessment and data correction of real climatological series is established. The major tasks involve the selection of high quality and climatically representative neighboring stations, estimation of missing values, detection and measurement of inhomogeneities, identification of the causes of the inhomogeneities through investigation of station history files, correction of datasets, and evaluation of the adjusted series with respect to its neighbors. The technique has addressed a number of these tasks and an approach for homogeneity assessment and data adjustment of real climate datasets is now presented.

The annual mean maximum temperatures of Collegeville, Nova Scotia, are tested for homogeneity for the period 1916–95. This station was selected for its long-term program of observations, its high quality data, and its few gaps of missing values; as well, a sufficient number of neighboring stations were available for the full period of time for homogeneity testing. Its station history file also contains detailed information, which could be helpful for the verification of the cause of the identified inhomogeneities. The annual mean maximum temperature series is presented in Fig. 5: it shows a decreasing linear trend over the tested interval of time.

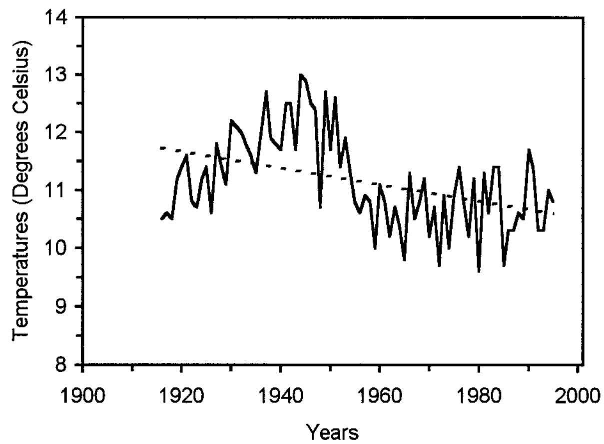

FIG. 5. Annual mean maximum temperatures for Collegeville, 1916–95. The dashed line represents the best-fit linear trend.

Neighboring stations are carefully selected to represent as much as possible the climate of the area. Distance from the tested site, elevation, landscape features, vegetation, and amount of missing data are considered. Since the technique requires a similar interval of time for the base and neighboring series, length of record is often a prime factor of selection. Missing monthly values are estimated using a known technique (Thom 1966), which has been used in previous work (Gullett et al. 1991; Gullett et al. 1990). Four stations are initially chosen as potential candidates. Correlations computed from the first difference series indicate that the climate patterns presented by the series without inhomogeneities are similar, since all coefficients are greater than 0.70. However, correlations on the annual values, ranging from 0.45 to 0.65, reveal disagreements between some of the series and suggest potential inhomogeneities either at the base or neighboring sites. To assess the suitability of each neighboring series, graphs of the differences between the annual values of the base and each neighbor are produced. In this example, there is one neighbor with a large step near the beginning of its record and it is rejected from the analysis. Based on correlations and graphs, three neighbors are finally chosen for testing the homogeneity of Collegeville.

Model 1 is first fitted to the series for the period 1916–95. The autocorrelation in the residuals are significantly different from zero at several consecutive low lags (Fig.

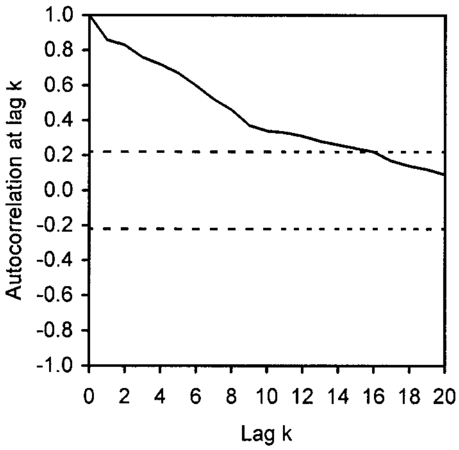

FIG. 6. Autocorrelations of the residuals at different lags after fitting model 1 to the annual mean maximum temperatures of Collegeville, 1916–95. The dashed lines represent the approximate 95% confidence interval.

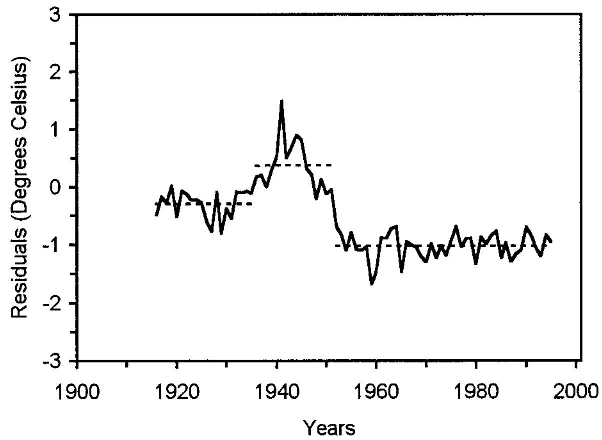

FIG. 7. Residuals obtained from fitting model 1 to the annual mean maximum temperatures of Collegeville, 1916–95. The dashed line represents the mean of each segment.

6), and it is concluded that the base is not homogeneous for the tested period of time. The graph of the residuals against time confirms this result (Fig. 7). Model 3 identifies a decreasing step of 1.3°C in 1952 while model 4 identifies a decreasing step of 1.6°C in 1952 with an increasing trend from 1916 to 1951. When this first interval is tested, an increasing step of 0.6°C is identified in 1936 by both models 3 and 4. The interval 1952–95 is homogeneous since the autocorrelation in the residuals is not significant after fitting model 1 to the corresponding period of time.

Before adjusting the base series, it is preferable to establish if the identified inhomogeneities are real and correctable. In theory, all nonclimatic steps should be corrected regardless of their magnitudes. However, in practice, techniques have more difficulty to correctly identify small steps. This new technique is reliable to identify steps of 0.75°C and above while steps of 0.50°C are frequently identified and steps of 0.25°C are not identified very often. Considering these results, the following rules are applied for adjusting real climate datasets: steps of about 0.75°C and above are always corrected, steps of about 0.50°C are corrected only if it is possible to determine their cause using the station history files, and steps of about 0.25°C and under are not corrected. Investigation of the station history reports for Collegeville revealed a change in observer and a small site relocation in 1936. Written comments from the inspector indicated that the maximum temperatures were too high after 1936. A second change in observer took place at the beginning of the 1950s with a site relocation of about 10 km north of the previous site. The inspector subsequently reported that the new data had little in common with the data from the former sites.

Adjustments are applied to bring each segment into agreement with the most recent homogeneous part of the series. First, segment 1936–51 is adjusted to segment 1952–95 by adding a mean value of about -1.6 to the annual values of segment 1936–51. A different correction factor is computed for each of the 12 months. To accomplish this, model 3 is applied to the series of the January values for the period 1936–95, to estimate the magnitude of adjustment in 1952. This is repeated for the 11 other series of monthly values in order to obtain 12 monthly correction factors, their average producing a correction of -1.6 for the annual values in 1952. Then, segment 1916–35 is adjusted to the corrected segment 1936–95 by adding a mean value of -1.0 to the annual values of segment 1916–35. The adjusted series is retested for homogeneity: the autocorrelation in the residuals are not significantly different from zero after fitting model 1 (Fig. 8), and it is concluded that the adjusted dataset is homogeneous. The graph of the residuals against time confirms this result (Fig. 9). To

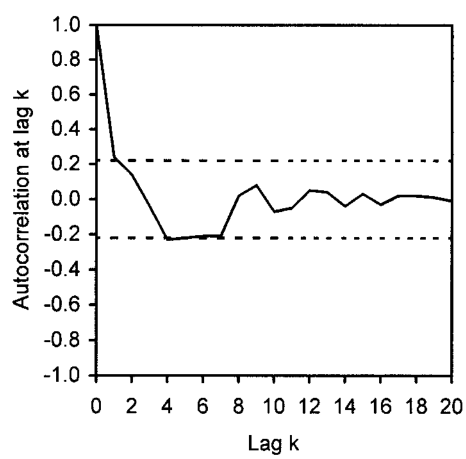

FIG. 8. Autocorrelations of the residuals at different lags after fitting model 1 to the adjusted annual mean maximum temperatures of Collegeville, 1916–95. The dashed lines represent the approximate 95% confidence interval.

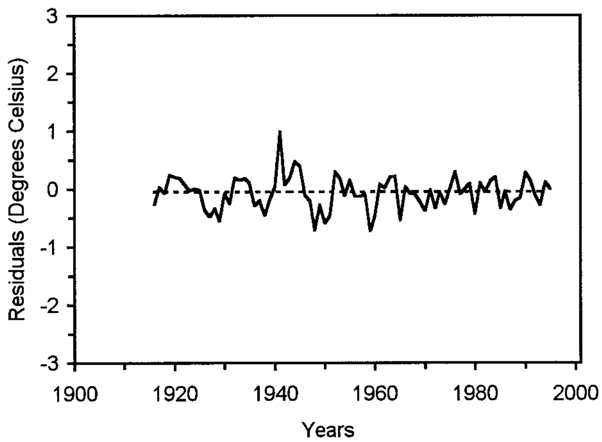

FIG. 9. Residuals obtained from fitting model 1 to the adjusted annual mean maximum temperatures of Collegeville, 1916–95. The dashed line represents the mean of the series.

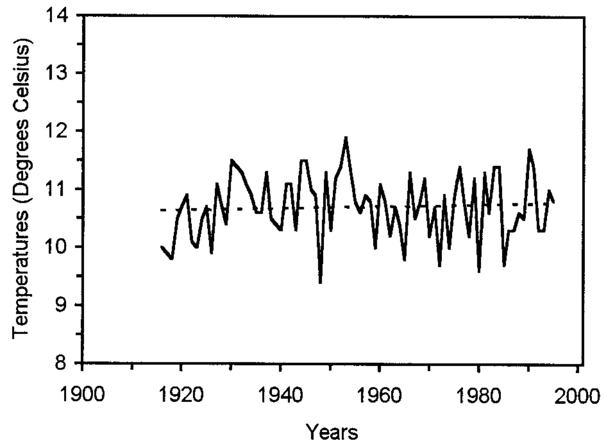

FIG. 10. Adjusted annual mean maximum temperatures for Collegeville, 1916–95. The dashed line represents the best-fit linear trend.

validate the adjusted series, the mean, standard deviation and linear trend are obtained at the base and the three neighboring sites for comparison purposes. The adjusted series of Collegeville shows a slight positive linear trend over the tested interval of time (Fig. 10). Increasing linear trends of about this magnitude are as well observed in the neighboring series. It is concluded that the adjusted series of Collegeville reflects more closely the climate variation observed in the area.

7. Conclusions

A new technique for the identification of inhomogeneities in Canadian temperature series has been presented. Its general approach is quite different from other techniques commonly used for homogeneity assessment since it uses the autocorrelation in the residuals to determine whether there are inhomogeneities in the tested series. At first, it considers the entire period of time and then it systematically divides the series into homogeneous segments. Each segment is defined by some changepoints, and each changepoint corresponds to either an abrupt change in mean level or a change in the behavior of the trend. The station history files can be used, when available, to determine the cause of the inhomogeneities. Application of this technique to other climate elements, such as for total precipitation, may be possible; however, a proper investigation of the time series of the element is first required to facilitate the interpretation of the results.

A number of objectives were established to enhance the application of this technique to climatological datasets. It has been shown that it identifies homogeneous and inhomogeneous intervals of time. This feature is useful for focusing the search for inhomogeneities and for determining reliable segments that can be used without alteration. The technique also identifies and measures two types of inhomogeneities: steps and trends. Results have shown that when base and reference series are reasonably well correlated, the technique identifies the date within 2 yr and the magnitude within 0.2°C of steps of magnitude 0.75°C and larger. Steps of 0.5°C are frequently identified and steps of 0.25°C are not identified very often. Preliminary results also show that the beginning date and the magnitude of a trend are identified when a step occurs at the beginning of the trend; however, further testing is required to fully assess the ability of the technique to identify nonclimatic trends. The most probable date of a change is detected by fitting regression models for a sequence of potential changepoints, and the one providing the minimum RSS value becomes the estimated date of the inhomogeneity. It has been shown that this approach is quite reliable and it also minimizes the use of the station history files since they are used only for the verification of the cause of the inhomogeneities and not for the actual identification process.

Future work will involve the application of this new technique to assess and correct Canadian temperature datasets in order to create a new version of a Canadian historical climate database. Homogeneity assessment and correction of inhomogeneities are essential to identify or to produce datasets that are reliable for climate change studies; however, inappropriate adjustment can lead to erroneous conclusions. Verification of the climatic representation of each "new" adjusted series thus becomes crucial to the success of the operation.

Acknowledgments. The author would like to thank Mr. Don W. Gullett, Mrs. Leslie H. Malone, and Dr. Francis W. Zwiers of the Atmospheric Environment Service, and Prof. Georges Monette of York University for providing encouragement, guidance, and reviews throughout the duration of the project. The author would like also to thank Dr. Edward Epstein, Dr. David R. Easterling, and an anonymous reviewer for their valuable comments and suggestions.

REFERENCES

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

Box, G. E. P., and G. M. Jenkins, 1976: *Time Series Analysis, Forecasting and Control.* Holden-Day, 575 pp.

Chatfield, C., 1984: *The Analysis of Time Series, An Introduction.* 3d ed. Chapman and Hall, 286 pp.

Draper, N., and H. Smith, 1981: *Applied Regression Analysis.* 2d ed. John Wiley and Sons, 709 pp.

Easterling, D. R., and T. C. Peterson, 1992: Techniques for detecting and adjusting for artificial discontinuities in climatological time series: A review. *Proc. Fifth International Meeting on Statistical Climatology*, Toronto, ON, Canada, Amer. Meteor. Soc., J28–J32.

——, and ——, 1995: A new method for detecting undocumented discontinuities in climatological time series. *Int. J. Climatol.,* **15,** 369–377.

——, ——, and T. R. Karl, 1996: On the development and use of homogenized climate datasets. *J. Climate,* **9,** 1429–1434.

Gullett, D. W., L. Vincent, and P. J. F. Sajecki, 1990: Testing for homogeneity in temperature time series at Canadian climate stations. CCC Rep. 90-4, Atmospheric Environment Service, 43 pp. [Available from the Climate Research Branch, Atmospheric Environment Service, 4905 Dufferin St., Downsview, ON M3H 5T4, Canada.]

——, ——, and L. H. Malone, 1991: Homogeneity testing of monthly temperature series. Application of multiple-phase regression models with mathematical changepoints. CCC Rep. 91-10, Atmospheric Environment Service, 47 pp. [Available from the Climate Research Branch, Atmospheric Environment Service, 4905 Dufferin St., Downsview, ON M3H 5T4, Canada.]

International Panel on Climate Change, 1992: *Climate Change 1992, The Supplementary Report to The IPCC Scientific Assessment.* Cambridge University Press, 200 pp.

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: 1851–84. *J. Climate Appl. Meteor.,* **25,** 161–179.

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.

Mitchell, J. M., Jr., 1961: The measurement of secular temperature change in the eastern United States. Research Paper 43, Office of Climatology, U.S. Weather Bureau, Washington, DC, 80 pp.

Neter, J., W. Wasserman, and M. H. Kutner, 1985: *Applied Linear Statistical Models.* 2d ed. Irwin, 1127 pp.

Peterson, T. C., and D. R. Easterling, 1994: Creation of homogeneous composite climatological reference series. *Int. J. Climatol.,* **14,** 671–680.

Phillips, D., 1990: *The Climates of Canada.* Environment Canada, 176 pp.

Skinner, W. R., and D. W. Gullett, 1993: Trends of daily maximum and minimum temperature in Canada during the past century. *Climatol. Bull.,* **27,** 63–77.

Sneyers, R., 1989: On the homogeneity of long series of observation. Preprints, *Fourth Int. Meeting on Statistical Climatology,* Rotoria, New Zealand, Amer. Meteor. Soc., 10–14.

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

Thom, H. C. S., 1966: *Some methods of climatological analysis.* Tech. Note 81, WMO, 53 pp. [Available from World Meteorological Organization, 41 Avenue Giuseppe-Motta, Case Postale 2300, CH-1211 Geneva 2, Switzerland.]

Vincent, L., 1990: Time series analysis: Testing the homogeneity of monthly temperature series. Survey Paper 90-05, York University, 50 pp. [Available from York University, 4700 Keele St., North York, ON M3J 1P3, Canada.]