Detecting and Removing Inhomogeneities from Long-Term Monthly Sea Level Pressure Time Series

KENNETH C. YOUNG

Institute of Atmospheric Physics, University of Arizona, Tucson, Arizona

(Manuscript received 6 February 1992, in final form 14 August 1992)

ABSTRACT

A method for the detection of and adjustment for multiple discontinuities in pressure time series is presented. The method relies on determining a reference pressure series using improved interpolation techniques and objectively comparing the observed and reference time series. The ability of the method to locate discontinuities in time and to estimate the magnitude of the discontinuity are examined using randomly generated time series with known properties.

The sea level pressure time series for Darwin (WMO 94120) is analyzed, and ten discontinuities are noted for the period 1869–1988. These are compared with previous examinations of the Darwin pressure time series.

I. Introduction

Many long-term sea level pressure time series are affected by station moves, instrument changes, or by changes in the time of observation. Similarly, the amalgamation of two or more nearby stations to form a single, long-term sea level pressure time series may introduce discontinuities. Sufficiently detailed station histories may not be available to locate or correct for inhomogeneities resulting from these factors.

Long-term sea level pressure time series are often used in constructing indices of the Southern Oscillation (e.g., Wright 1989). Such time series are also used in developing long-range climate forecast models (e.g., Christensen and Eilbert 1985). The presence of inhomogeneities in the sea level pressure time series reduces their usefulness by introducing extraneous noise that tends to obscure the signal being sought. Wright (1989) compares several indices of the Southern Oscillation based on sea surface temperature, rainfall, and sea level pressure, and identifies several inhomogeneities in the time series of the indices and suggests corrections to improve their homogeneity.

Several researchers have suggested and/or used a variety of corrections to the sea level pressures for Darwin (WMO 94120). Darwin is of particular interest since it constitutes part of a commonly used Southern Oscillation index with Tahiti (WMO 91938). Allan et al. (1991) provide an extension of the Darwin series back to March 1869 and document several instrument changes during the nineteenth century. Clayton (1944) provides evidence for a systematic change in the pressure at Darwin due to defects in the barometer that was replaced on 1 August 1931. Parker (1983) suggests a correction due to a change in the time of observation on 31 March 1939.

This paper presents an objective method for detecting and removing discontinuities in long-term sea level pressure time series. The method has been applied to 160 long-term sea level pressure stations worldwide, although the emphasis here will be on stations commonly used as indices of the Southern Oscillation due to the more general interest in pressures at these stations. The pressure record at Darwin is of particular interest due to the available documentation of the station history and the number of suggested corrections in the literature designed to improve the homogeneity of the pressure time series.

The method for detecting and removing discontinuities in the pressure time series in this paper is essentially that developed by Maronna and Yohai (1978) and used to analyze precipitation time series by Potter (1981). This method compares the observed time series with a reference time series that may be obtained through any of a number of interpolation methods. Easterling and Peterson (1992) review several methods for detecting discontinuities in annual temperature time series and conclude that the method developed by Maronna and Yohai is well suited for this purpose.

The sensitivity and accuracy of the method to small discontinuities lies in the skill of the interpolation

---

1 The Global Historical Climatological Network (GHCN) dataset provides Darwin pressures back to January 1882.

2 Inhomogeneities may be of two types, discontinuous or gradual. This method is designed to detect only discontinuous inhomogeneities, which will be termed discontinuities. This method may detect gradual inhomogeneities, although they would not be distinguished as such.
method. The three-way interpolation model described by Young (1992) is used to create the reference time series. The three-way model combines (1) multiple linear regression (MLR), (2) multiple discriminant analysis (MDA), and (3) normalized anomaly (NA) interpolation methods. Each method yields an estimate for an individual month; taking the median of the three estimated values will be shown to provide an interpolation skill equal to or better than any of the individual methods over the range of interstation correlations for the interpolation networks.

2. The sea level pressure database

The pressure data were obtained from the global historical climatological network (GHCN) dataset Vose et al. (1992) and the World Weather Disc. Station pressures were reduced to sea level for stations below 1500 m elevation using reported station elevations and monthly mean temperatures. Since the interpolation models require a minimum of 30 points, only stations with more than 30 years of record were included. A total of 1058 land stations were included in the interpolation network.

The Comprehensive Ocean–Atmosphere Data Set (COADS; Slutz et al. 1985) was analyzed in an effort to supplement the land-station network. A census of the $2^\circ \times 2^\circ$ gridded ship observations was undertaken for those months/grids reporting 10 or more, 20 or more, and 30 or more observations. A total of 157 "ship stations" was created, balancing the length of record, the average number of observations per month, and the location to provide a useful pressure record. The interpolation network comprised a total of 1215 stations.

In an effort to reduce the computational requirements while maintaining near-optimal interpolation models, a list of candidate stations was prepared for each set of interpolations. A set of interpolations refers to the interpolations required for a single station for a given month. For example, the 139 interpolations required for March at Darwin, from 1850 through 1988, constitutes a set. The list of candidate stations included 203 designated long-term stations, plus the 70 stations with the highest correlation with the base station for that month. Both positively and negatively correlated stations were included. Interpolation models were developed using only the candidate stations.

3. The interpolation models

The MDA and MLR interpolation models are discussed in Young (1992); a brief summary of these is given below. The normalized anomaly model is similar to that described by Karl and Williams (1987), with modifications discussed below. Each of the models is adaptive in the sense that each provides an optimal interpolation based on the stations available for each individual interpolation. This is accomplished by determining a sublist of stations from the candidate list that has values available for the date the interpolation is desired and then by creating interpolation models based on those stations.

a. Description of the three interpolation models

The normalized anomaly model selects the three available stations from the candidate list with the highest correlation with the base station. A station is considered available for a given interpolation if it has an observed value of pressure for that date. The historical means and standard deviations are determined for the base station and the three surrounding stations. Each surrounding station provides an estimate of the pressure at the base station as

$$\tilde{P}_o = \tilde{P}_i + S_i \frac{\sigma_o}{\sigma_i} (P_i - \tilde{P}_i),$$

where the overlines represent averages, and the subscripts o and i refer to the base and surrounding stations, respectively. The variable $S_i$ has a value of +1, if the correlation coefficient is positive, and has a value of −1, if the correlation coefficient is negative. This extends the normalized anomaly method to permit utilization of negatively correlated stations.

The three estimates of the pressure at the base station are weighted by the square of the $t$ statistic following the procedure described by Young (1992) for the normal ratio method. The weight is given by

$$w_i = \frac{r_i^2 (n_i - 2)}{1 - r_i^2},$$

where $r_i$ is the correlation coefficient between the base station and the $i$th surrounding station, and $n_i$ is the number of points the correlation coefficient is based on. A minimum of 30 points was required.

The multiple linear regression (MLR) and multiple discriminant-analysis (MDA) models were both limited to a maximum of four surrounding stations and a minimum of 30 points was required. As described in Young (1992), the MLR selection is based on the improvement in the multiple correlation coefficient rather than in the reduction in the residual sum of squares, since the number of available points varies from test to test.

---

3 The GHCN pressure dataset was obtained from Tom Peterson at the National Climatic Data Center.

4 WeatherDisc Associates, 4584 NE 89th, Seattle, WA 98115.

5 COADS was obtained from the National Center for Atmospheric Research.

6 The term "base station" refers to the station for which the interpolation is desired. The term "surrounding station" is used to identify those stations that are being used to make the interpolation; such stations may or may not be in close geographical proximity to the base station.
The MDA model is based on Miller (1962), with modifications described in Young (1992). For the purpose of interpolation, the pressures at the base station represent the predictands, and those at the surrounding stations represent the predictors. The sorted predictands are partitioned into roughly equal-sized groups. The optimal number of groups must be determined, although the interpolation model is not overly sensitive to the number of groups. For this study, ten groups were determined to be near optimal.

The MDA interpolation involves selection of a number of historical analogs. The optimal number of analogs varies with the number of points the model is based on and the degree of correlation between the base and surrounding stations. Models based on poorly correlated stations generally use more analogs than models based on highly correlated stations. Models based on a greater number of points generally use more analogs. The optimal number of analogs is determined separately for each interpolation and is based on maximizing the mean absolute error interpolation skill over the database from which the model was developed.

b. Evaluation of the interpolation models

Interpolations were carried out for 160 long-term sea level pressure stations worldwide. Interpolated values were determined for each month from January 1850 through December 1988 for each station. Interpolated values were compared to the available observed values in order to evaluate the skill of the three interpolation methods. For each interpolation, the median of the three values provided by the three models was determined. The median of the interpolated values is referred to as the "three-way model."

The skill of the interpolation must clearly depend on the degree of correlation between the base and surrounding stations. The square of the multiple correlation coefficient \( R^2 \) from the MLR model was chosen to represent this degree of correlation. Brownlee (1965) defines \( R^2 \) as

\[
R^2 = \frac{\text{sum of squares due to regression}}{\text{total sum of squares}}.
\]

Fifty stations were used to evaluate the interpolation models. These stations were selected to provide a wide range of values of \( R^2 \) and tend to include a relatively larger fraction of the stations with lower values of \( R^2 \). A total of 67 857 interpolations are represented in this analysis; the breakdown by intervals of \( R^2 \) is given in Table 1.

The interpolation skill was determined for both the mean absolute error (MAE) and the root-mean-square error (RMSE). The interpolation skill is compared to substitution of the long-term average (climatological) value as the interpolated value. This provides a climatological mean absolute error (CMAE) and a climatological root-mean-square error (CRMSE). The interpolation skill is the fractional reduction in the MAE or RMSE compared to climatology and is calculated as

\[
\text{skill} = \frac{\text{CMAE} - \text{MAE}}{\text{CMAE}}
\]

or

\[
\text{skill} = \frac{\text{CRMSE} - \text{RMSE}}{\text{CRMSE}}.
\]

The interpolation skills as functions of \( R^2 \) for each of the three models and for the three-way model are shown in Figs. 1 and 2. In general, the MDA model is superior to the MLR and NA models for \( R^2 < 0.95 \), while the MLR model tends to be superior for \( R^2 > 0.98 \). The MDA model exhibits useful skill (above 20%) for \( R^2 \) as low as 0.4, whereas the MLR and NA models are of little use for \( R^2 \) below 0.7. The three-way model tends to equal or exceed the best of the individual models over the entire range of \( R^2 \). The three-way interpolation model is used for the remainder of this study.

Interpolation procedures, in general, tend toward the long-term mean values.\(^7\) For example, if there is little or no correlation between the base station and the surrounding stations, the minimum interpolation error is obtained by using interpolated values close to the long-term mean. Thus, an interpolated time series will exhibit a reduction in variance compared to the observed time series. Figure 3 illustrates the reduction in variance as a function of \( R^2 \) for each of the methods. The MLR method is noted to be superior (least reduction in variance) for \( R^2 > 0.7 \), while the NA method is superior for lower \( R^2 \) values. For the purposes of detecting and correcting discontinuities in the pressure time series, a small MAE is desired and the reduction in the variance is less important. For the purposes of determining the frequency of extreme events in such a time series for

\(^7\) For the MDA method, the interpolated value is the median of the analogs, which implies the MDA method tends toward the median rather than the mean as the correlation between the base and the surrounding stations decreases.
Fig. 1. Interpolation skill based on the mean absolute error for the three individual models and the three-way model as a function of $R^2$. Interpolation skill represents the fractional improvement over the substitution of the long-term average value. Plots are based on 67 857 interpolations distributed as indicated in Table 1.

Fig. 2. Interpolation skill based on the root-mean-square error for the three individual models and the three-way model as a function of $R^2$, as for Fig. 1.
example, the MLR method would be more appropriate since the reduction in variance is smaller.

4. Analyzing the pressure time series for discontinuities

Discontinuities are detected by comparing the observed time series with a reference time series generated using the interpolation techniques described in the previous section. Tests of the methods used to detect discontinuities were conducted using simulated sea level pressure data in an effort to ascertain the accuracy in locating the time and determining the magnitude of discontinuities and the likelihood of identifying spurious discontinuities. Finally, applications of the method to several actual sea level pressure time series are shown.

a. Methods for detecting discontinuities

The basic method used here for detecting discontinuities was developed by Maronna and Yohai (1978) and described by Potter (1981). In essence, a series that is the difference between the observed and reference time series is developed. At any point, this time series can be split into two portions; the means of the two portions are compared and a student t-statistic is calculated for the difference in means. If this is done for all points (excluding the end points), a plot of the t statistic as a function of time may be obtained (see Fig. 4).

The point at which the maximum in the t statistic occurs is identified as a discontinuity. Only those candidate discontinuities exhibiting t statistics that are significant at a predetermined level of significance are accepted. The average difference between the observed and reference time series is determined for the portion of the time series up to and including the discontinuity and for the portion following the discontinuity. The magnitude of the adjustment is estimated as the difference between these two averages.

Multiple discontinuities are identified and isolated using a stepwise procedure in which the time series is adjusted for the discontinuity associated with the maximum t statistic at each step. The adjusted series is then reanalyzed for additional discontinuities, and the procedure is repeated until no significant discontinuities remain. Once all significant discontinuities have been identified, the adjustments required to bring each segment into agreement with the most recent segment are calculated. This procedure is illustrated in section d below.
b. Simulated time series of sea level pressure

Simulated sea level pressure time series were generated assuming normal distributions with monthly means and standard deviations based on Bergen, Norway (WMO 01317), Charleston, South Carolina (WMO 72208), and Apia, Samoa (WMO 91762), as given in Table 2. These stations were chosen since their records did not appear to contain major discontinuities, and they provide a wide range in standard deviations.

A one-hundred-year series of monthly sea level pressure was simulated for both the observed and reference series. A series of 1200 normally distributed independent values, with a mean of zero and a standard deviation of one, was generated using a pseudorandom number generator. The desired autocorrelation was achieved by averaging in a fraction of the previous value to each value, that is, to achieve an autocorrelation (lag of one month) of roughly 0.07, the current value was calculated as

$$x_i^* = 0.93x_i + 0.07x_{i-1}^*,$$

where the unstarred values are the independent values, and the starred values are the autocorrelated values.
The cross correlation between the observed and referenced time series reflects the degree of correlation between the base and the surrounding stations. The desired degree of cross correlation was obtained by averaging in a fraction of each observed series value to the corresponding reference-series value. For example, to achieve a correlation coefficient of 0.95 between the observed and reference series, 

\[ r'_{ij} = 0.75x_{ij}^* + 0.25x_{ij}^* \]

where \( r \) and \( o \) are the reference and observed series values, and the primed values incorporate both autocorrelation and cross correlation. The proper fraction to achieve a given level of cross correlation was determined by trial and error.

These standardized time series were converted to monthly sea level pressure time series by multiplying each value by the observed standard deviation for that month and adding in the average pressure for that month; that is, the first value was converted to pressure using the mean and standard deviation for January, and the second value was converted using the February mean and standard deviation, etc. The reduction in variance associated with the interpolation was introduced into the reference time series during this step by reducing the standard deviation used in converting the standardized reference time series by the square root of the desired reduction in variance.

The presence of an annual pressure cycle introduces autocorrelation into the pressure time series. Autocorrelation coefficients were calculated from the observed-pressure time series for each of the three stations by subtracting the monthly mean and dividing by the monthly standard deviation before determining the autocorrelation, that is, the autocorrelations were calculated from standardized time series. The autocorrelation coefficients for a one-month lag for the three stations were -0.01, 0.07, and 0.45, respectively.

c. Tests using simulated time series

The simulated time series used in these tests assumed that the interpolation model achieved an \( R^2 \) value of 0.9 with a reduction in variance of 0.7, with the exception of one test in which a less skillful interpolation was simulated. More than 60% of the interpolations listed in Table 1 achieved \( R^2 \) values in excess of 0.9, making this choice a fairly representative one. Unless otherwise indicated, simulations were based on Charleston monthly means and standard deviations.

The likelihood of the method to produce spurious discontinuities was found to increase as the autocorrelation of the time series increased. Tests based on ten simulated time series, without introduced discontinuities for both Bergen and Charleston, showed no spurious discontinuities (false alarms). However, 12 of the 20 simulations for Apia (autocorrelation = 0.45) produced spurious discontinuities when a 95% level of significance was used. This was reduced to a 25% false-alarm rate when a 99% level of significance was used. Four of the simulations for Apia produced two spurious discontinuities each. The average absolute value of the adjustments was 0.14 mb, ranging from 0.04 mb to 0.22 mb. Introduction of a 0.45 autocorrelation into the Charleston-and-Bergen simulated time series produced similar false-alarm rates.

The precision in locating a discontinuity was found to decrease as the variance of the monthly values increased (see Table 3). An introduced discontinuity of 0.5 mb at year 48, month 7, was detected for all 20 simulations for Apia, although a 25% false-alarm rate was found. The root-mean-square error (RMSE) in locating the discontinuity was 5.4 months. Only one of the 20 simulations failed to locate the discontinuity within 10 months, and the discontinuity for six simulations was located from between 48/6 and 48/8.

All twenty of the simulations for Charleston detected the introduced discontinuity, and no spurious discontinuities were detected. The RMSE in locating the discontinuity was 5.6 months. For Bergen, two of the 20 simulations failed to detect any discontinuities at the 95% level and three others were in error by more than four years. If a 99% level of significance had been used,

---

8 It should be noted that introduction of discontinuities into unautocorrelated time series produced autocorrelation in proportion to the magnitude of the discontinuity; that is, large discontinuities resulted in a substantial increase in the autocorrelation.
Table 3. Probability of detection (POD) and false alarm rates (FAR) for simulated sea level pressure time series. The root-mean-square error in locating a 0.5-mb discontinuity and the estimated adjustment are given. Results are based on 20 simulations of 100 years each for each station. The POD and FAR are based on a threshold level of 99% significance.

<table>
<thead>
<tr>
<th>Station</th>
<th>POD</th>
<th>FAR</th>
<th>Time (mos)</th>
<th>Adjustment (mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>API191762</td>
<td>100%</td>
<td>25%</td>
<td>5.4</td>
<td>0.50 ± 0.08</td>
</tr>
<tr>
<td>CHA722038</td>
<td>100%</td>
<td>0%</td>
<td>5.6</td>
<td>0.51 ± 0.05</td>
</tr>
<tr>
<td>BER01317</td>
<td>80%</td>
<td>0%</td>
<td>32.3</td>
<td>0.55 ± 0.11</td>
</tr>
</tbody>
</table>

the probability of detection would have been 80%. The RMSE in locating the discontinuity for the 18 simulations was 32.3 months. Thus, increasing the variance of the monthly values decreases the probability of identifying a discontinuity and increases the error in locating the discontinuity. The estimated adjustments for Bergen, Charleston, and Apia were 0.55 ± 0.11 mb, 0.51 ± 0.05 mb, and 0.50 ± 0.08 mb, respectively.

The accuracy in locating the discontinuity improves markedly as the magnitude of the discontinuity increases. The RMSE in locating a 0.3-mb discontinuity located at year 48/7 was 39.1 months, compared to 5.6 months for locating a 0.5-mb discontinuity, and 2.3 months for locating a 1.0-mb discontinuity. The estimated adjustments were 0.31 ± 0.05 mb, 0.51 ± 0.05 mb, and 1.00 ± 0.05 mb, respectively. Easterling and Peterson (1992) also found that the accuracy in locating the discontinuity improved as the magnitude of the discontinuity was increased.

It was slightly more difficult to detect small discontinuities using shorter time series, although the accuracy in locating the discontinuities was slightly better. Analyses based on simulations of 50-year time series located a 0.5-mb discontinuity with an RMSE of 4.9 months and a 0.3-mb discontinuity with an RMSE of 14.2 months, compared to RMSE of 5.6 months and 39.1 months for the 100-year series, respectively. However, the value of the t statistic was reduced from 54.6 to 37.8 for the 0.3-mb discontinuity with a POD of 95% (for a 95% significance-level threshold). The estimated adjustments based on the 50-year series were virtually identical to those based on the 100-year series.

Discontinuities located near the ends of the time series were more difficult to locate than discontinuities near the middle. Discontinuities of 0.5 mb introduced at year 5/1 and at year 10/1 exhibited RMSE in locating the discontinuity of 10 months compared to 5.6 months for 0.5-mb discontinuities introduced at year 48/7. The uncertainty in the estimated adjustment also was greater; the estimated adjustment for the discontinuity at year 5 was 0.51 ± 0.11 mb, compared to an estimated adjustment of 0.51 ± 0.05 mb when the discontinuity was at 48/7.

The RMSE in locating the discontinuity was markedly greater for less skillful interpolations. A simulation for an interpolation $R^2 = 0.4$, with a reduction in variance of 0.43, resulted in an error in locating a 0.5-mb discontinuity of 65 months compared to 5.6 months for the interpolation with $R^2 = 0.9$. The estimated adjustment was 0.53 ± 0.09 mb compared to 0.51 ± 0.05 mb for the $R^2 = 0.9$ simulation.

Limited tests with multiple, introduced discontinuities suggested that the presence of additional discontinuities had relatively little effect on the detection and adjustment of discontinuities. Twenty simulations were tested with 0.5 mb added from year 1/1 through year 20/3, and 0.3 mb subtracted from year 20/4 through year 48/7. The discontinuity at 20/3 was located with an error of 6.1 months compared to an error in locating a 0.5-mb single discontinuity at year 48/7 of 5.6 months. The estimated adjustment for the period from year 1/1 through year 20/3 was 0.49 ± 0.07 mb, compared to 0.51 ± 0.05 mb for the single discontinuity simulations.

The RMSE in locating the smaller discontinuity of 0.3 mb at year 48/7 was 39.0 months, compared to the RMSE of 39.1 months in locating a single discontinuity at the same date. The estimated adjustment was 0.31 ± 0.06 mb for the period from year 20/4 through 48/7. The probability of detection was 100%; the false-alarm rate, at a 95% level of significance, was 30% and was reduced to a 5% false-alarm rate at a 99% level of significance.

d. Analysis of the pressure time series for Darwin

Figure 4 is a plot of the $t$ statistic for Darwin from 1868 through 1920 (based on the time series from 1869 through 1898). The horizontal lines superimposed on the plot of the $t$ statistic represent the 95% and 99% significance levels estimated from the table provided by Potter (1981). The bar graph at the top represents the month-by-month difference between the observed and referenced time series ($\Delta P$), with values above the center line indicating that the observed value is greater than the reference value. It is evident from this figure that there is more than one discontinuity.

The Darwin time series shown in Fig. 4 has a maximum $t$ statistic (39.6) at April 1875. The adjustment required to make the average $\Delta P$ for the period 1869 through April 1875 equal to that for the period subsequent to April 1875 is −0.52 mb. This adjustment was applied to all the observed pressures through April 1875. This in effect removes the discontinuity at April 1875.

The analysis was then repeated on the adjusted time series with the results shown in Fig. 5. Now, the peak in the $t$ statistic (68.4) is at November 1898. The calculated adjustment of +0.39 mb was applied to all of the observed pressures through November 1898, and

---

10 Data for 1869–1881 was obtained from Allan et al. (1991).
the analysis was repeated as shown in Fig. 6. Now the peak in the $t$ statistic (56.1) is at December 1881, requiring a correction of $-0.44$ mb. Figure 7 shows the results of the fourth analysis. Note how the bar graph illustrates the improvement in the homogeneity of the time series with each adjustment.

For time series with several discontinuities, it is recommended that smaller portions of the time series be examined separately to confirm discontinuities identified in the previous analysis and to locate other possible discontinuities that may have been masked by the identified discontinuities. For example, the partition of the Darwin pressure series from January 1882 through April 1917 was used to verify and pinpoint the 1891 and 1898 discontinuities. This analysis, which is shown in Fig. 8, pinpoints the 1898 discontinuity at September rather than November. The 1891 discontinuity was shifted by one month, from February to January. Other partitions failed to find additional discontinuities.

Once all of the discontinuities significant at the 1% level were identified, the average $\Delta P$ for each segment was calculated. The adjustments required to bring each segment into alignment with the most recent segment

---

**Fig. 5.** Analysis of the Darwin pressure time series adjusted for the discontinuity detected in the previous analysis shown in Fig. 4. The peak is located at November 1898 with $t = 68.4$. 
Fig. 6. Analysis of the Darwin pressure time series adjusted for the discontinuities detected in the previous analyses shown in Figs. 4 and 5. The peak is located at November 1881, with $t = 56.1$.

(October 1931 to date) were determined using the difference in the average $\Delta P$ values between that segment and the most recent segment. These adjustments are shown in Table 4.

The average $R^2$ values for each year for the Darwin interpolations are plotted in Fig. 9. For the period 1869–1881, the $R^2$ values are close to 0.6, suggesting that discontinuities smaller than 0.5 mb can be detected. The results in Table 4 show that all of the discontinuities, except that at April 1875, are considerably larger than 0.5 mb, suggesting that these discontinuities are located with an error of a few months and the adjustments are reliable to within 0.2 mb. The discontinuities at January 1891 and September 1898 are 0.5 mb, with interpolation $R^2$ values slightly greater than 0.8, suggesting these are located with an error of 6 to 8 months and the adjustments are within 0.1 mb.

In addition to detecting and providing adjustments for discontinuities in the sea level pressure time series, the time series of $\Delta P$ values was analyzed for single anomalous values. The program scanned for locally extreme values of $\Delta P$ and then calculated the mean
and standard deviation of the fifteen $\Delta P$ values prior to and following the locally extreme value. The $\chi^2$ statistic was calculated as

$$\chi^2 = \frac{(\Delta P)_x - \bar{\Delta P})^2}{\sigma_{\Delta P}},$$

where $(\Delta P)_x$ is the locally extreme value.

The reported sea level pressure of 1011.1 mb for July 1979 was found to be 3.7 mb less than the interpolated value, giving a $\chi^2$ value of 253. The reported pressure of July 1979 would have been a record low pressure for that month, given the adjustments to the pressure time series noted in Table 4; the reported pressure for July 1979 was discarded. The observed pressures for December 1886 of 1003.4 mb (interpolated 1007.4 mb) and for February 1953 of 1007.3 mb (interpolated 1008.7 mb), with $\chi^2$ values of 36 and 44, respectively, were noted but not discarded.

e. Comparison with previously reported discontinuities for Darwin

Several of the discontinuities located by this analysis coincide with changes in the barometer at Darwin, some of which have been previously associated with
discontinuities in the Darwin pressure record. Previously suggested discontinuities resulting from changes in the times of observation were not found in this analysis.

The discontinuity located by this method at September 1898 has been previously noted by several researchers. Trenberth (1984) used an adjustment of +1.0 mb to the Darwin sea level pressures from January 1882 to August 1898 based on Troup (1965), although Troup did not provide precise information as to the correction or the month. Troup referenced Kidson (1925) as the source of the correction. The barometer change in August 1898 is also documented by Allan et al. (1991).

This analysis suggests there are two adjustments to be applied for the period from January 1882 through 1898. The adjustment for the period from January 1882 through January 1891 is +0.9 mb, which is in good agreement with Trenberth’s adjustment of +1.0 mb. However, the period from February 1891 through September 1898 should be adjusted only by +0.4 mb rather than 1 mb. Allan et al. (1991) note a barometer change
TABLE 4. Adjustments required to remove significant discontinuities from the reported sea level pressure time series for Darwin (WMO 94120).

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>Adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>March 1869</td>
<td>December 1869</td>
<td>+1.0 mb</td>
</tr>
<tr>
<td>January 1870</td>
<td>March 1871</td>
<td>−1.6 mb</td>
</tr>
<tr>
<td>April 1871</td>
<td>January 1873</td>
<td>+0.7 mb</td>
</tr>
<tr>
<td>February 1873</td>
<td>October 1873</td>
<td>−2.2 mb</td>
</tr>
<tr>
<td>November 1873</td>
<td>April 1875</td>
<td>−0.3 mb</td>
</tr>
<tr>
<td>May 1875</td>
<td>November 1881</td>
<td>−0.1 mb</td>
</tr>
<tr>
<td>December 1881</td>
<td>January 1891</td>
<td>−0.9 mb</td>
</tr>
<tr>
<td>February 1891</td>
<td>September 1898</td>
<td>+0.4 mb</td>
</tr>
<tr>
<td>October 1898</td>
<td>April 1917</td>
<td>−0.1 mb</td>
</tr>
<tr>
<td>May 1917</td>
<td>September 1931</td>
<td>+0.2 mb</td>
</tr>
<tr>
<td>October 1931</td>
<td>date</td>
<td>0.0 mb</td>
</tr>
</tbody>
</table>

in August 1891 that may be the cause of the discontinuity at February 1891.

The discontinuity located by this analysis at September 1931 may have resulted from a barometer change in August 1931 reported by Clayton (1944). Trenberth and Shea (1987) note a correction for a problem due to etched lines in the barometer cistern that restricted the movement of the mercury, resulting in a reduction in the observed magnitude of the diurnal (and semidiurnal) pressure range, as explained by Clayton (1944). Clayton suggested a gradually increasing correction to be applied from 1914 through 1 August 1931 when the barometer was replaced. Trenberth and Shea (1987) note that although this correction was supposedly applied, the corrected values appear to be identical with the original values. A small adjustment for the period May 1917 through October 1931 (see Table 4) may reflect this correction.

No evidence was found in this analysis for a discontinuity associated with a change in the time of observation after 31 March 1939 noted by Parker (1983), who estimated an adjustment of +0.2 mb for the period prior to the change. Wright (1984) notes this and other changes in the times of observations but did not apply any corrections.

Allan et al. (1991) report a barometer change in March 1877 as well as changes in the cistern height during 1878–1879 and 1894–1895. None of these can be associated with discontinuities found by this analysis. The discontinuities that mark the adjustment of −2.2 mb for the period from February 1873 through October 1873 are marginally significant at the 1% level, which suggests this adjustment may be spurious. However, the discontinuity at December 1869 has a t statistic of 55.4, which is highly significant, suggesting the adjustment of −1.6 mb from January 1870 through March 1871 represents a real discontinuity.

Fig. 9. Interpolation network correlation as measured by $R^2$ for Djakarta (96745), Darwin (94120), Tahiti (91938), and Poona (43063). Values represent the average for each year. The associated interpolation skill and reduction in variance can be determined from Figs. 1 through 3.
TABLE 5. Adjustments required to remove significant discontinuities from the reported sea level pressure time series for Djakarta (WMO 96745).

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>Adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 1866</td>
<td>June 1899</td>
<td>+0.1 mb</td>
</tr>
<tr>
<td>July 1899</td>
<td>June 1948</td>
<td>+0.2 mb</td>
</tr>
<tr>
<td>July 1948</td>
<td>April 1963</td>
<td>+0.4 mb</td>
</tr>
<tr>
<td>May 1963</td>
<td>December 1985</td>
<td>0.0 mb</td>
</tr>
<tr>
<td>January 1986</td>
<td>date</td>
<td>−0.5 mb</td>
</tr>
</tbody>
</table>

f. Analyses for other stations

Five other long-term pressure stations that have been used by various researchers to describe the Southern Oscillation were analyzed for discontinuities. Of these, Tahiti (WMO 91938) and Easter Island (WMO 85469) appear to be homogeneous; that is, the records did not contain discontinuities significant at the 1% level, although Tahiti did have an apparent discontinuity at April 1881, significant at −2%. The August 1934 pressure reported for Tahiti appears to be too low, with \( \chi^2 = 25 \). Trenberth and Shea (1987) did not find discontinuities in these records either.

The record for Djakarta (WMO 96745) was found to contain four relatively small discontinuities detailed in Table 5. The record for Apia (WMO 91762) was found to contain two discontinuities detailed in Table 6. No discontinuities were found by Trenberth and Shea (1987) for Djakarta for the period 1900 through 1984 or for Apia for the period 1890 through 1984. The record for Santiago (WMO 85574) was found to contain five discontinuities detailed in Table 7. Trenberth and Shea (1987) noted one discontinuity for Santiago from 1906 through 1984.

The pressure record at Poona (WMO 43063) was analyzed for comparison with the discontinuities and the adjustments given by Trenberth and Shea (1987). Their method is a subjective method of examining a single time series of pressure anomalies for discontinuities, and then calculating the adjustment required to bring each segment into alignment. They note two discontinuities for the period 1893 through 1984, one at December 1909 and the other at December 1960. This analysis failed to find the discontinuities suggested by Trenberth and Shea (1987) but did identify three other discontinuities detailed in Table 8.

Figure 9 suggests a high degree of reliability for the interpolation models for Poona, with \( R^2 \) values close to 0.96 for the period since 1900. Examination of Trenberth and Shea’s (1987) Fig. A1 shows a decrease of approximately 3 mb near 1960, which is identified as a discontinuity. Examination of the pressure record for nearby Bombay (WMO 43057) shows a similar drop in the pressure. If a systematic change in the manner of reporting sea level pressures was instituted for all the Indian stations at this time, then the interpolation methods described here would fail to detect such a discontinuity.

TABLE 7. Adjustments required to remove significant discontinuities from the reported sea level pressure time series for Santiago (WMO 85574).

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>Adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 1861</td>
<td>July 1871</td>
<td>+0.3 mb</td>
</tr>
<tr>
<td>August 1871</td>
<td>March 1876</td>
<td>−0.5 mb</td>
</tr>
<tr>
<td>April 1876</td>
<td>February 1898</td>
<td>+0.6 mb</td>
</tr>
<tr>
<td>March 1898</td>
<td>May 1903</td>
<td>+1.4 mb</td>
</tr>
<tr>
<td>June 1903</td>
<td>March 1977</td>
<td>+0.5 mb</td>
</tr>
<tr>
<td>April 1977</td>
<td>date</td>
<td>0.0</td>
</tr>
</tbody>
</table>

TABLE 6. Adjustments required to remove significant discontinuities from the reported sea level pressure time series for Apia (WMO #91762).

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>Adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 1890</td>
<td>October 1910</td>
<td>0.0 mb</td>
</tr>
<tr>
<td>November 1910</td>
<td>October 1917</td>
<td>−1.1 mb</td>
</tr>
<tr>
<td>November 1917</td>
<td>date</td>
<td>0.0 mb</td>
</tr>
</tbody>
</table>

5. Summary and conclusions

The method of detecting discontinuities in time series developed by Maronna and Yohai (1978), and used by Potter (1981) for locating single discontinuities in precipitation time series, has been applied successfully to locating multiple discontinuities in monthly average sea level pressure time series. Tests of the method were conducted using simulated sea level pressure time series having a range of variance and autocorrelation with known discontinuities.

The accuracy in locating discontinuities depends on the magnitude of the discontinuity, the variance of the underlying time series, and to a lesser extent, on how close the discontinuity is to the ends of the time series. Whereas the RMSE in locating a 0.5-mb discontinuity in simulated time series with relatively low variance (standard deviations less than 2 mb) was generally less than 6 months, the RMSE was more than 30 months for a simulated time series with a relatively high variance (standard deviation of more than 5 mb).

The probability of detecting a 0.5-mb discontinuity

TABLE 8. Adjustments required to remove significant discontinuities from the reported sea level pressure time series for Poona (WMO 43063).

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>Adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 1875</td>
<td>June 1876</td>
<td>−3.3 mb</td>
</tr>
<tr>
<td>July 1876</td>
<td>March 1893</td>
<td>+0.8 mb</td>
</tr>
<tr>
<td>April 1893</td>
<td>December 1949</td>
<td>−0.1 mb</td>
</tr>
<tr>
<td>January 1950</td>
<td>January 1951</td>
<td>+1.4 mb</td>
</tr>
<tr>
<td>February 1951</td>
<td>June 1967</td>
<td>+0.2 mb</td>
</tr>
<tr>
<td>July 1967</td>
<td>date</td>
<td>0.0</td>
</tr>
</tbody>
</table>
was 100% for time series with relatively small variances, whereas the probability of detection was 80% for series with larger variances using a 99% level of significance for acceptance. The probability of detecting a discontinuity of a given magnitude is somewhat lower for shorter time series. Given a discontinuity, the uncertainty in the estimated adjustment was relatively insensitive to the variance of the series or the magnitude of the discontinuity. The RMSE of the estimated adjustment was generally less than 0.1 mb.

The detection of spurious discontinuities was found to be dependent on the autocorrelation of the time series. Time series with low autocorrelations (less than 0.1 for a one-month lag) did not produce any spurious detections. Tests using the autocorrelation of 0.45 calculated from the Apia (WMO 91762) sea level pressure time series exhibited false-alarm rates of roughly 25%, using a threshold significance level of 99%. Higher autocorrelations appear to be associated with tropical stations, suggesting that this method may introduce spurious adjustments when applied to tropical sea level pressure time series, although the magnitude of the spurious adjustments is generally between 0.1 and 0.2 mb.

The presence of multiple discontinuities was found to have relatively little effect on the probability of their detection or the accuracy in locating the discontinuity. The procedure used to identify multiple discontinuities is basically objective but not completely algorithmic, that is, when several discontinuities are present, it is recommended that segments of the time series be analyzed; the choice of segments is subjective. In most cases with several discontinuities, it is preferable to verify the discontinuities that are found by analyzing the complete time series, by analyzing appropriate segments of the time series.

The skill of the interpolation method was found to have a strong effect on locating the discontinuity. The method was able to locate a 0.5-mb discontinuity with a RMSE of 5.6 months for a “typical” level of interpolation skill \( R^2 = 0.9 \). When a much less skillful interpolation was simulated \( R^2 = 0.4 \), the RMSE in locating the discontinuity increased to 65 months. The uncertainty in the estimated adjustment was relatively unaffected.

The choice of interpolation models is dependent on the purpose to which the interpolated time series is to be applied. The three-way or median interpolation model provides the maximum interpolation skill, defined as a reduction in the mean absolute error or the root-mean-square error. However, this is achieved at the expense of a larger reduction in the variance, that is, the three-way model achieves its skill by not “going out on a limb.” If the purpose of the interpolation is to analyze extreme events, then multiple linear regression or the normalized anomaly methods retain more of the variance, although the overall interpolation skill is reduced.

Most of the long-term sea level pressure time series analyzed contain significant discontinuities. Adjustments of 5 mb or more were found on a number of occasions, particularly during the nineteenth century. The sea level pressure time series for Sverdlovsk (WMO 28440) and Dunedin (WMO 93894) were of such poor quality during the nineteenth century that the entire record was discarded. The record at Dunedin seriously affected the reference time series for Wellington. The question of which station should be discarded was resolved by removing each from the list of available stations for the other and examining each time series for discontinuities.

Abrupt changes in climate over extensive regions have been noted by van Loon and Madden (1983) and by Trenberth (1990). Trenberth found an abrupt 2-mb decrease in sea level pressure over a large portion of the north Pacific Ocean occurring around 1976. Since this method relies on surrounding stations to provide the reference time series, it is highly likely that both the base station and the surrounding stations would experience similar changes. Thus, such an abrupt change in the observed time series would be reflected in the reference time series such that no discontinuity would be noted.

On the opposite side, if such a regional change were the result of changing the time of observation for many stations simultaneously, it is possible this method would fail to detect such a discontinuity. If most or all of the surrounding stations were affected, as well as the base station, then the discontinuity would be reflected in the reference time series and no discontinuity would be detected. This is a possible explanation for the failure of this method to detect the discontinuity of roughly 3 mb near 1960 for Poona (WMO 43063), noted by Trenberth and Shea (1987).

The presence of numerous and significant discontinuities in sea level pressure time series increases the difficulty of identifying teleconnections, such as those associated with the Southern Oscillation. Such teleconnections not only illustrate the interconnectedness of the atmosphere but are important in developing useful long-range forecast models.

The methods presented here may be directly applied to the analysis of time series of other parameters, such as temperature or precipitation. Relatively little modification of the existing programs is required to analyze long-term monthly temperature time series for discontinuities. The analysis of precipitation time series would require replacing the normalized anomaly method with the normal ratio method as described by Young (1992). These methods also can be applied to daily temperature (or precipitation) data.

Acknowledgments. This research was supported in part by grants from Edwin X. Berry and Associates and the Salt River Project. The assistance of Tom Peterson of the National Climatic Data Center (NCDC)
in supplying the GHCN pressure data is gratefully acknowledged. Discussions with David Easterling, also of the NCDC, regarding the construction of simulated time series with known autocorrelation and cross correlation were quite valuable. The able assistance of Gene Petrescu in handling both the land and ocean pressure data is acknowledged. The figures were drafted by James Abbott.

REFERENCES


