A Quality Assurance System for Canadian Hourly Pressure Data

HUI WAN, XIAOLAN L. WANG, AND VAL R. SWAIL

Climate Research Division, Atmospheric Science and Technology Directorate, Environment Canada, Toronto, Ontario, Canada

(Manuscript received 25 April 2006, in final form 12 February 2007)

ABSTRACT

In this study a comprehensive quality assurance (QA) system, which includes the hydrostatic check combined with a statistical homogeneity test, is designed and applied to hourly pressure records (for 1953–2002) from 761 Canadian stations, to produce a high-quality database of hourly station and sea level pressures for various climate studies. The main principles of the QA system are described in detail, followed by a brief emphasis on the error correction algorithms. The general performance of the QA system and the main problems in the Canadian historical hourly pressure database are discussed and illustrated through various examples. The results show that there are serious systematic errors (i.e., sudden changes in the mean, or mean shifts) in the Canadian hourly pressure database, which are caused either by the use of incorrect station elevation values in the reduction of barometer readings to station or sea level pressure values (e.g., the "50-ft rule" or station relocation without updates to the station elevation), by transposing/swapping station and sea level pressure values, or by mistakes made in the archive data ingestion or data recording/digitization processes (e.g., use of a wrong base number). Random errors also exist and are mainly due to the transposition of two digits or miscoding of one or two digits. These errors must be corrected before the data are used in various climate studies, especially climate change–related studies.

1. Introduction

Climate change has become an important issue, because increasing evidence suggests consistent warming trends over the past century, with a faster warming rate over land compared to oceans (Houghton et al. 2001). More and more efforts have been devoted to the assessment of climate change and their impacts. However, one needs long-term homogeneous records of climate data to characterize climate variability and climate change in the past, and to validate numerical model simulations. It is imperative to conduct quality assurance and homogenization of climate data before these data are used for various climate studies, especially climate change–related studies.

Atmospheric circulation plays an essential role in the climate system because of its effects on the distribution of heat and moisture over the globe. Surface atmospheric pressure is an important variable that describes atmospheric circulation. Variations in surface pressure should also reflect variations in surface temperature, because the two variables are related to each other thermodynamically. Therefore, analysis of surface atmospheric pressure is critical to our understanding of climate variability and climate change.

Several studies on the collection and analysis of atmospheric pressure data have been carried out lately. As a result, several good quality pressure datasets of global or regional coverage have been developed, mainly to provide vital inputs for numerical model studies of global climatic variations and changes (e.g., Smith and Reynolds 2003; Kaplan et al. 2000; Allan et al. 1996; Trenberth and Paolino 1980). Many data quality–related problems were found and corrected in these studies. These problems include data errors and discontinuities or inhomogeneities, and high-latitude station data problems (which are reported to have arisen from a lack of data availability for the Arctic region).

In the mean time, there have been several studies using Canadian pressure data. Slonosky and Graham (2005) developed a Canadian monthly mean station pressure (SP) dataset with 71 stations that have data

Corresponding author address: Hui Wan, Climate Research Division, Atmospheric Science and Technology Directorate, Environment Canada, 4905 Dufferin Street, Toronto, ON M3H 5T4, Canada.
E-mail: hui.wan@ec.gc.ca

DOI: 10.1175/2007JAMC1484.1

records for 50–130 yr. They found strong correlations between the variability of atmosphere circulation and surface temperature anomalies. They also reported several major inhomogeneities in the dataset. Nkemdirim and Budikova (2001) examined trends in monthly mean sea level pressure (SLP) in western Canada using data from 51 stations for the period from 1956 to 1993, and reported a significant decline in annual mean and winter mean SLP over the Arctic.

However, the original records of surface atmospheric pressure are hourly measurements, from which the commonly used monthly or daily mean pressure values are derived. Unfortunately, the hourly pressure data archived in Environment Canada (EC) have not undergone a quality control (QC) or quality assurance (QA) procedure (except at times for which missing data are flagged). Slonosky and Graham (2005) corrected some problems in their analysis of monthly pressure data (although their corrections are not physically based and are applied to monthly data), while Nkemdirim and Budikova (2001) did not (and hence their results are most likely unreliable). A high-quality homogeneous pressure database is essential for various climate studies; hourly pressure data of high quality are particularly valuable for studying extremes such as atmospheric storminess. Therefore, the goal of the current study is to develop a comprehensive quality assurance system for hourly pressure data.

The necessity of applying a QA procedure to meteorological data has long been recognized. The earliest QA systems were developed for radiosonde data (Gandin 1988; Collins and Gandin 1990). However, more and more effort has been directed toward developing QA systems for high-temporal-resolution surface meteorological data, such as daily or hourly data (Kunkel et al. 1998; Graybeal et al. 2004; Shafer et al. 2000). A complex QA procedure consists of a series of checks on data, with the results obtained from these checks being used systematically to determine whether or not a value is suspicious and how to correct the suspicious value, if possible. Because not all flagged data are erroneous, a complex QA procedure should check all flagged data to screen out those most suspicious values (for correction or exclusion) and to remove flags from data that are deemed consistent with other reliable data. This procedure is usually called the decision-making method (DMM) (Gandin 1988; Graybeal et al. 2004). A modern complex QA system is used not only to identify but also to correct suspicious data whenever possible.

The EC digital archive contains pressure data from 1953 to date. For the early decades, data were digitized from original paper forms, without any quality control performed after digitization. Even for the real-time data (those from electronic reports), the QC procedure is quite limited according to the EC National Archive hourly data quality control documents published on the EC's Web site (Environment Canada 2004). Thus, a QA procedure for hourly pressure data is developed in this study with the goal of combining existing techniques with a statistical homogeneity test and fitting them to Canadian historical data.

In this study, we develop a QA procedure for Canadian hourly pressure data. The data, QA procedure, and homogeneity test used are described in sections 2, 3, and 4, respectively. Section 5 describes the error correction algorithms. The corrected data series are analyzed in section 6, with some concluding remarks in section 7.

2. Data

Surface atmospheric pressure is usually recorded for both the station elevation and mean sea level. Generally, atmospheric pressure values at the station elevation are called SP and are calculated from the station barometer readings. Mean SLP is derived from the SP, so that the barometric pressures for stations with different elevations can be compared at a common level (mean sea level) for synoptic purposes. Generally, SP data should be more reliable than SLP because fewer calculations are involved. However, SLP data have been used quite often for various purposes, such as constructing atmospheric circulation indicators (e.g., Wright 1984; Jones et al. 1999), developing long-range climate forecast models (e.g., Christensen and Eilbert 1985), and analyzing severe weather phenomena (e.g., Wang et al. 2006; Alexander et al. 2005). Therefore, high-quality data for both the station elevation and mean sea level are needed for various studies.

In this study we apply the QA system to as many stations as possible to support our interest in producing a gridded pressure dataset in the near future and in using the quality data for future generations of global reanalysis, such as the twentieth-century reanalysis project (Compo et al. 2006). There are 1085 stations available for both SP and SLP data in the EC data archive. Only stations with continuous records of at least 1 yr and at least eight reports per day were included in this study. (Although at most stations atmospheric pressure is reported hourly, with 24 measurements per day, some stations either have only one report every 3 or 6 h or have hourly reports for only part of day, e.g., from 0300 to 1600 UTC. The number of pressure reports per day could vary from station to station and/or from one period to another.) Because SLP data are derived from SP data, and we will use both elements for QA, the checking procedure will be ap-

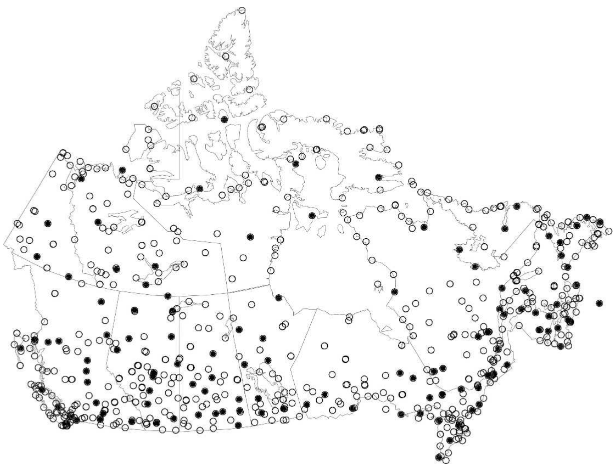

FIG. 1. Location of stations analyzed in the study. Solid dots indicate stations with more than 25 yr of data in the period of 1971–2000 (used in the selection of climatological thresholds).

plied to data only when both SP and SLP data are available. A total of 761 stations (see Fig. 1) are analyzed in the study.

3. The quality assurance system

The QA system proposed here consists of five components. These include checking for upper and lower climatological thresholds/limits, temporal pressure changes, and hydrostatic, temporal, and internal consistencies. For each station, all valid (nonmissing) values are subject to these five checks. Based on the results of these checks, a decision regarding either acceptance, correction, or rejection of the data is made.

a. Limits check (LC)

The climatological thresholds/limits check is a very commonly used checking procedure to identify outliers (e.g., Hubbard et al. 2005; Graybeal et al. 2004; Shafer et al. 2000). In this study, the climatological thresholds were determined as the lowest and highest values in the 1971–2000 period (for each station with at least 25 yr of data in this period), if these values are associated with acceptable values of 1-, 2-, and 3-h pressure changes as defined in the pressure changes check (see the next subsection). If the lowest or highest hourly value is associated with unacceptable values of pressure change (i.e., it does not pass the pressure changes check), we exclude it and check to see whether or not the second-lowest or -highest hourly value in this period can pass the pressure changes check. If not, we check the third-lowest/-highest value, and so on. This procedure goes on until the acceptable climatological thresholds are found. These thresholds are determined for each station for both station pressure and sea level pressure, separately. Note that the station-specific thresholds are also necessary for SLP, because the climatology of SLP also depends on the location relative to the climatological mean position of circulation modes such as the Aleutian or Icelandic low (the long-term mean SLP field is not even over the globe).

There are only 120 Canadian stations with at least 25 yr of hourly pressure data in the 1971–2000 period (see Fig. 1). Among these stations, the lower limits range from 942.0 to 981.1 hPa for SLP and from 846.7 to 968.2 hPa for SP, while the upper limits range from 1041.8 to 1078.8 hPa for SLP and from 917.1 to 1058.3 hPa for SP. Because the climatological limits were determined using data recorded in the 30-yr period from 1971 to 2000, extremes outside this period may exceed the thresholds. However, note that these thresholds are used only to screen out suspicious data for further analysis (to narrow the range of further checks); these suspicious data are not necessarily concluded as erroneous (and hence rejected) at the end of the procedure. Also, an arbitrary tolerance of 3.4 hPa (0.10 in. of Hg) was added to the thresholds for each station, which more or less alleviates the limits. For a station with a shorter data record, we use the lowest lower limit among its four "nearest" surrounding stations as its lower limit, and the highest upper limit as its upper limit. Station elevation is also considered; each of the four "nearest" stations must have an elevation difference from the short-term station that is less than 200 m (otherwise it is replaced by the next-nearest station; the 200-m limit is reasonable because it is used only in finding the most appropriate climatological limits). This limit for difference in elevation is important for setting the climatological limits of station pressure, especially for elevated stations.

b. Pressure changes check (PC)

The limits for 1-, 2-, and 3-h pressure changes (also called pressure tendency values) taken from the EC hourly data quality control document (Environment Canada 2004) are used in this study. They are 3.9 hPa h^{-1}, 6.9 hPa $(2\ h)^{-1}$, and 9.9 hPa $(3\ h)^{-1}$, respectively. These limits were developed in the early to mid-1990s by experienced meteorological technicians (D. Boudreau, National Archive and Data Management Division of Environment Canada, 2003, personal communication). Note that these limits are relatively low when compared with the limits used by other scientists in other countries [e.g., Shafer et al. (2000) and Meek and Hatfield (1994) use a limit of 10 hPa h^{-1}]. Thus, for very rare events, the true pressure tendency could exceed these limits [e.g., Le Blancq (2003) reported that 3-hourly station pressure tendency was 28.9 hPa on 11 February 2003 from 1000 to 1300 UTC at Sable Island, Nova Scotia, Canada]. We further check manually to determine whether or not the identified outlier is a true outlier when this check is used to find the climatological limits (see section 3a above). Generally, a flag is issued to a datum if at least one of the associated pressure tendency values exceeds its limit.

c. *Internal consistency check (IC)*

Basically, the SP and SLP values should not be equal for a long period of time for stations of nonzero station elevation. However, long periods of consecutive identical values of SP and SLP are seen in our pressure data archive. Slonosky and Graham (2005) reported discontinuities in the SP data series that are due to a change in the definition of "station elevation." The sixth edition of the "Manual of surface weather observation" ("MANOBS"; Environment Canada 1970) states that "the established elevation of Mean Sea Level (MSL) is arbitrarily assigned to stations at which the cistern height is less than 50 feet above MSL." The latest edition of the MANOBS (Environment Canada 1977) summarized that "prior to 1 January 1977 the term 'established elevation' was used" and that "an established elevation of zero metres (MSL) was assigned to all stations where the cistern elevation was less than 15 metres" (i.e., 50 ft). As a consequence, the station pressure and the sea level pressure were identical at these stations before January 1977. Therefore, an IC flag is activated when identical values of SP and SLP are found for at least 1 month. Actually, this "50-ft rule" problem could also lead us to flag a long run of consecutive hourly records during the hydrostatic check described below.

d. *Hydrostatic check (HC)*

The hydrostatic check has been used routinely in upper-air radiosonde data quality control (Gandin 1988; Collins and Gandin 1990). It plays a crucial role in identifying errors of either height, pressure, or temperature at mandatory isobaric surfaces. In this study, we use it alone to detect and correct random errors in both station and mean sea level pressure data, and we also combine it with a statistical homogeneity test to detect and correct systematic errors (see section 4 below).

For station pressure P_z and sea level pressure P_0, the hydrostatic check is based on the hydrostatic model {Saucier 1955, his Eq. [3.07(1)]}

$$Z = \ln\frac{P_0}{P_z} \times (T_0 + \overline{T}_{\mathrm{dry}}) \bigg/ \left(\frac{g}{R} - \frac{a}{2}\ln\frac{P_0}{P_z}\right), \qquad (1)$$

where Z is the station elevation (m), R is the gas constant for dry air, $T_0 = 273.15$ K, g is the acceleration of gravity, a is the standard lapse rate (0.0065°C m^{-1}), and $\overline{T}_{\mathrm{dry}}$ is the average of the current dry-bulb temperature and the dry-bulb temperature recorded 12 h earlier (°C).

However, since November 1976, the following formula is used in Canada for calculation of mean sea level pressure (Savdie 1982; WMO 1954):

$$P_0 = P_z \exp\left(\frac{gZ}{RT_{\mathrm{mv}}}\right), \qquad (2a)$$

where

$$T_{\mathrm{mv}} = (T_0 + \overline{T}_{\mathrm{dry}}) + \frac{aZ}{2} + e_s C_h(Z) + F(\overline{T}_{\mathrm{dry}}), \qquad (2b)$$

$$e_s = (\overline{T}_{\mathrm{dry}} + T_0)^{-0.000\,14\overline{T}^2_{\mathrm{dry}} + 0.0116\overline{T}_{\mathrm{dry}} + 0.279},$$

$$C_h(Z) = 2.8322 \times 10^{-9}Z^2 + 2.225 \times 10^{-5}Z + 0.107\,43,$$

and

$$F(\overline{T}_{\mathrm{dry}}) = b_1\overline{T}^2_{\mathrm{dry}} + b_2\overline{T}_{\mathrm{dry}} + b_3.$$

The T_{mv} is the mean virtual temperature of the fictitious air column between the height of the station and the mean sea level. The third term in (2b) represents a humidity correction, where e_s is the surface vapor pressure and $C_h(Z)$ is the humidity correction factor (a function of Z). The last term $F(\overline{T}_{\mathrm{dry}})$ accounts for correction of plateau effects (see Savdie 1982), and b_1, b_2, and b_3 are plateau correction parameters specific to each station. Note that neglecting humidity and plateau correction, the combination of Eqs. (2a) and (2b) is equivalent to (1).

The hydrostatic residuals R_z are defined as

$$R_z = Z_m - Z, \qquad (3)$$

where Z is the recorded current station elevation, which is taken from Environment Canada's station information system (SIS) [i.e., a metadata database that also includes historical station information, such as previous station location(s) (latitude–longitude), elevation(s), etc.] and can be assumed correct with good confidence (because the current station elevations in our SIS are very accurate), and Z_m is the estimation of the station elevation obtained by substituting the related hourly P_0, P_z, or $\overline{T}_{\mathrm{dry}}$ values in model (1). In the absence of data error(s), R_z values shall be very close to zero. A tolerance of R_z is used here to allow for small errors in the value of either P_0 or P_z, or even in the dry-bulb temperature $\overline{T}_{\mathrm{dry}}$ or the recorded station elevation Z (but undocumented large elevation changes can still be identified by this check).

To be more confident about the results of a hydrostatic check that uses dry-bulb temperatures, the limits check is also performed on dry-bulb temperature [which should range between −55° and 40°C, according

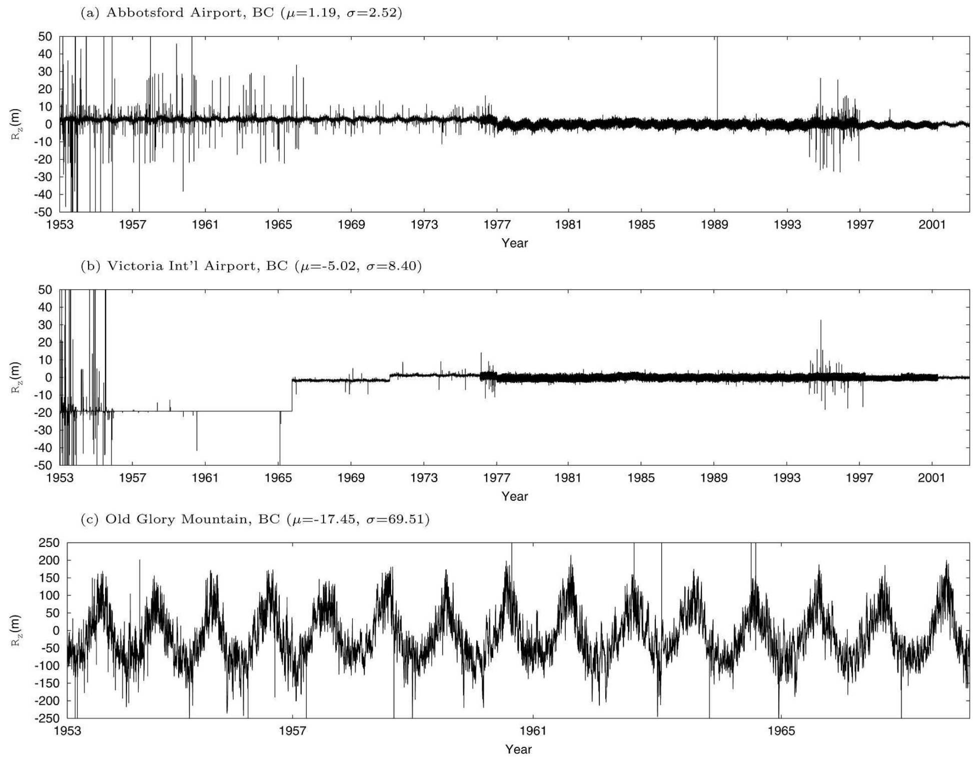

FIG. 2. Time series of differences between the recorded and the estimated station elevation for hourly observations at the indicated stations in British Columbia: (a) Abbotsford Airport, (b) Victoria International Airport, and (c) Old Glory Mountain.

to the EC hourly data QC document; see Environment Canada (2004)]. The results show that outliers are found only for 34 out of 761 stations, and the outlier rate is fairly low in general. We flag the R_z values that are associated with outlier(s) of dry-bulb temperature for further analysis. As an additional measure, we also tested the sensitivity of the estimated station elevation in (1) to errors in pressure and temperature records separately, and found that the estimated station elevation is more sensitive to pressure errors than to temperature errors. For example, for a station at 400-m elevation, an error of 1 hPa in the pressure data (station- or sea level pressure) will result in a difference of 8.5 m in the estimated elevation Z_m, while an error of 1°C in the temperature data will only result in a difference of 1.5 m in Z_m. Therefore, in this study, we assume that the recorded hourly dry-bulb temperature values are correct in general, but carefully analyze those values associated with outlier temperature(s). All hourly pressure data (both P_0 and P_z) associated with an R_z value that is greater than its tolerance are flagged for further analysis as a result of this hydrostatic check.

For each station, the tolerance of R_z is determined by the so-called sigma test (Hubbard et al. 2005; Shewhart 1980),

$$\mu - \gamma\sigma \leq R_z \leq \mu + \gamma\sigma, \tag{4}$$

where μ and σ are the mean and standard deviation of the hourly R_z time series, respectively, and γ is a parameter that defines the tolerance in terms of σ. A value passes the hydrostatic check if the above relationship holds. Clearly, the tolerance range of R_z depends on the estimates of μ and σ. Generally, in the absence of data errors, both μ and σ should be near zero (cf. Fig. 2a). However, it was found that many data problems and errors could significantly bias the estimates of μ and σ. For example, as shown in Fig. 2b, a clear step was found in the time series of R_z on 3 October 1965, with most of the residuals (in absolute value) before 1965 equal to the station elevation (19.2 m), which is apparently an error caused by the 50-ft-rule problem (cf. section 3c). Because step changes in the R_z time series could significantly affect the estimates of μ and σ, and hence the R_z tolerance used in the hydrostatic check, we need to identify and correct step changes in the R_z series first, so that more realistic R_z tolerance can be determined for screening random errors. That is, we need to identify and correct systematic errors in the related pressure series first. We do so by combining a statistical homogeneity test with the hydrostatic check,

TABLE 1. Station pressure (P_z) and sea level pressure (P_0) recorded at Nanaimo, British Columbia, from 2100 UTC 3 Apr 1954 to 0300 UTC 4 April 1954, and the results of applying the five checks on these data.

	P_z/P_0 (hPa)	LC flag (P_z/P_0)	PC flag (P_z/P_0)	HC flag (P_z/P_0)	TC flag (P_z/P_0)	IC flag (P_z/P_0)	Total flags (P_z/P_0)
2100 UTC	1006.8/1010.6	0/0	0/0	0/0	0/0	0/0	0/0
2200 UTC	1006.8/1010.6	0/0	1/0	0/0	0/0	0/0	1/0
2300 UTC	1006.7/1010.5	0/0	1/0	0/0	0/0	0/0	1/0
0000 UTC	1016.4/1010.2	0/0	1/0	1/1	0/0	0/0	2/1
0100 UTC	1005.7/1009.5	0/0	1/0	0/0	0/0	0/0	1/0
0200 UTC	1004.8/1008.6	0/0	1/0	0/0	0/0	0/0	1/0
0300 UTC	1003.9/1007.7	0/0	1/0	0/0	0/0	0/0	1/0

which is the most innovative part of the QA system proposed here and is described later in section 4.

Once all the mean shifts (systematic errors) are identified and corrected, the mean and standard deviation (μ and σ) of the new R_z time series (calculated from the corrected pressure data) can be used in (4) to set the R_z tolerance for the screening of random errors. With the more accurate estimates of μ and σ, the value of γ in (4) can now be selected by predetermining the upper limit for the random error rate, as practiced in Hubbard et al. (2005). In this study, the upper limit of random error rate is set to 0.2‰ for all stations analyzed. That is, we cap the random error rate uniformly across the country, rather than using a fixed γ value; the values of γ are determined in such a way that an upper limit of 0.2‰ random error rate is kept for each and every station (thus, for a station with 50-yr hourly observations, there will be 87 data flagged for further investigation). Note that the uniform rate of 0.2‰ is used just as an upper limit of random error rate. The actual rate of random errors that are corrected as a result of this procedure does vary from station to station, and systematic errors are distinguished from random errors (they are identified and corrected first).

e. Temporal consistency check (TC)

A temporal consistency check is to detect errors in the form of an unusually long run of constant value. Usually a timing window is used to detect inordinately long periods of constant pressure data (Graybeal et al. 2004; Meek and Hatfield 1994). In this study, if a constant pressure value runs consecutively for 12 h or longer in duration, all of these hours are flagged as a result of the temporal consistency check. This check is found to be useful for detecting errors arising from a damaged barometer or careless observing.

f. Decision-making method

We apply the five checks described above to hourly station and sea level pressure data (P_z and P_0) recorded at each of the 761 stations. As a result, many values could be flagged in one, several, or all of the five checks. However, not all flagged values are erroneous data. For example, a value can be flagged because of an error in the value recorded either 1–3 h earlier or later that cause the related pressure change to exceed its limit. One needs to analyze both adjacent flagged values and the number of flags on each value to determine the most suspicious one(s) for correction or exclusion. Such an analysis also leads to the removal of flags on values that are deemed correct. Thus, this decision-making procedure is an important step in climate data quality assurance (Graybeal et al. 2004; Collins and Gandin 1990). Because the QA system is only applied to two elements, the decision-making system is not very complicated. For example, a station pressure of 1006.4 hPa at 0000 UTC 4 April 1954 was miscoded as 1016.4 hPa, which caused eight flags on P_z and/or P_0 as shown in Table 1. Usually a datum with the highest count of flags is most suspicious, and all flags on values adjacent to that datum can often be removed (e.g., the value 1016.4 is flagged in the final database and all other data in Table 1 are cleared of flags). This is the base of our automatic DMM.

Occasionally, the total counts of flags for the two elements (P_z and P_0) are the same and we do not have enough information to judge which element is more suspicious. For example, a valid P_0 of 1021.5 hPa is miscoded as 1025.1 hPa, which is a mild error and is not severe enough to raise the LC/PC/TC/IC flags, only enough to raise the HC flag. In this case, we cannot determine which element (P_z or P_0) is erroneous; thus, both the P_z and P_0 values are flagged and further inspected manually.

4. Identification of systematic errors

As shown in Fig. 2b and discussed earlier in section 3d, there are mean shifts in the hydrostatic residual R_z series, which reflect mean shifts (systematic errors) in the related pressure series.

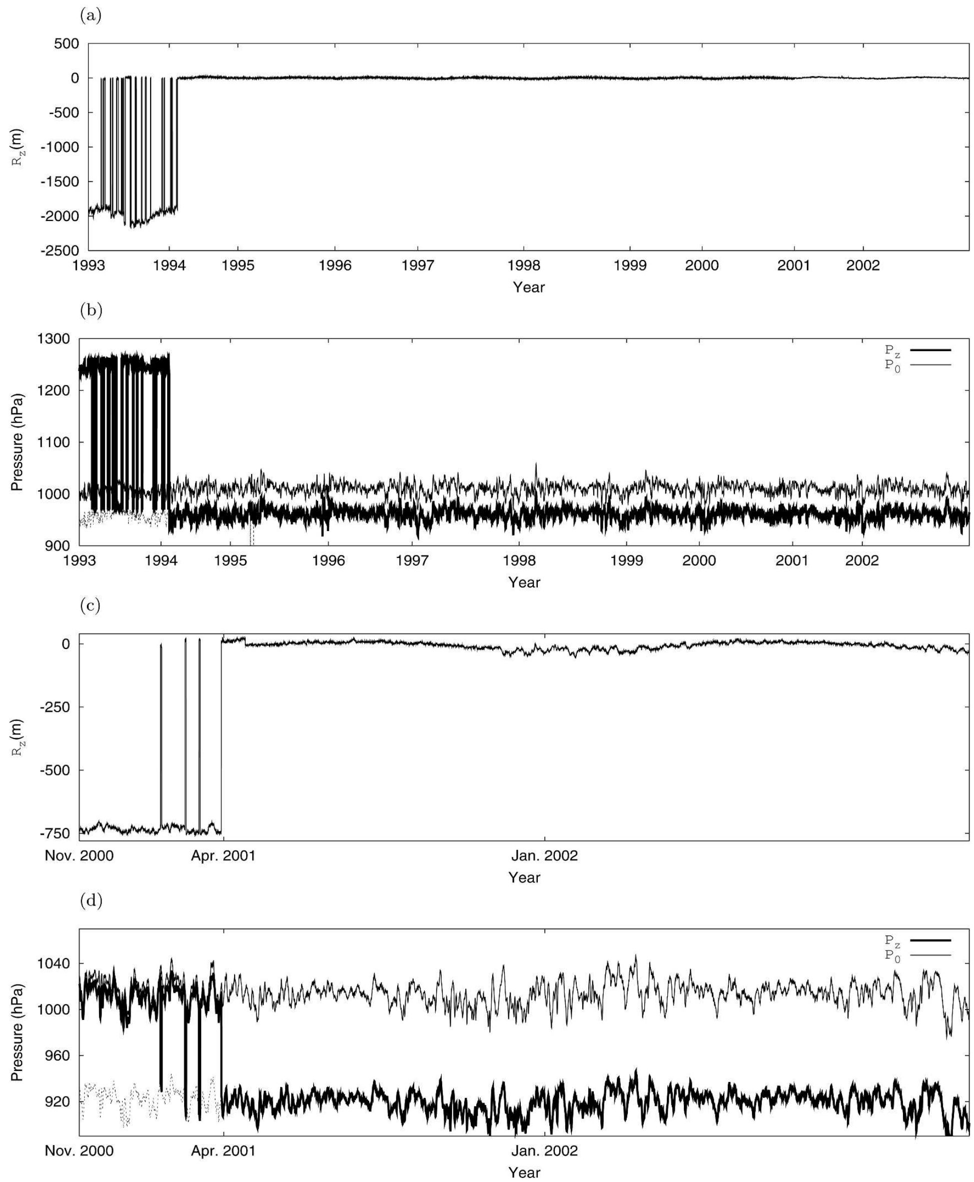

FIG. 3. Time series of R_z, P_0, and P_z for the selected period of hourly observations at (a), (b) Cape Hooper, Nunavut, Canada and (c), (d) Dease Lake Limited Weather Information Service, British Columbia. The dashed curves in (b) and (d) show the corrected P_z values.

Figure 3 shows two more examples of systematic errors in the hydrostatic residual and station pressure series. This type of error was found for many stations, especially the Arctic stations, and is not due to the 50-ft-rule problem. For some unknown reason (maybe an error in the archive data ingestion), the station pressure values for the period from 1992 up to 2002 were wrongly loaded for about 40 stations, including 18 Arctic stations. The associated R_z values (Figs. 3a,c) are incredibly high, showing a clear step change that would be easy to detect statistically, and most of the associated P_z values (see Figs. 3b,d) are unrealistically high and obviously wrong.

Although some systematic errors can be easily identified through visualization of the R_z and pressure series together (as shown in Fig. 3), it is common practice to detect them statistically. Considering the nature/physics of the R_z time series, for the vast majority of stations it is reasonable to assume that R_z has an independent identical Gaussian (IID Gaussian) distribution with mean μ and variance σ^2 under the null hypothesis of no systematic errors (note that this assumption is violated in the cases of elevated stations; details are given later in this section). Thus, testing whether or not there is a step change in the R_z time series for the period from N_1 to N_2 ($1 \le N_1 < N_2 \le N$; $n = N_2 - N_1 + 1$) is to test

$$H_0\colon R_z(t) = \mu + \varepsilon_t, \tag{5}$$

against

$$H_a\colon R_z(t) = \begin{cases} \mu_1 + \varepsilon_t, & N_1 \le t \le c \\ \mu_2 + \varepsilon_t, & c < t \le N_2 \end{cases}, \tag{6}$$

where step size $\Delta = \mu_2 - \mu_1 \neq 0$ and ε_t denotes an IID Gaussian variable of mean zero and variance σ^2. In the first homogeneity test (i.e., at the beginning of the process), $N_1 = 1$ and $N_2 = N$. In the successive tests, either N_1 or N_2 or both of them are set to the changepoints identified in the previous test(s); that is, the time series is segmented at the newly identified changepoint and a successive test is applied to each new segment of the time series [N_1 and N_2 are the first and last data points of the segment being tested; see Wang and Feng (2007) for the details].

Here, detection of an undocumented step change can be done with the $T_{\max}$ statistic as in the standard normal homogeneity test (SNHT; Alexandersson 1986), or equivalently using the following $F_{\max}$ statistic:

$$F_{\max} = \max_{N_1 \le c \le N_2} F_c, \tag{7}$$

where

$$F_c = \frac{(\mathrm{SSE}_0 - \mathrm{SSE}_a)/1}{\mathrm{SSE}_a/(n-2)}, \tag{8}$$

and

$$\mathrm{SSE}_0 = \sum_{t=N_1}^{N_2} [R_z(t) - \hat{\mu}]^2$$

$$\mathrm{SSE}_a = \sum_{t=N_1}^{c} [R_z(t) - \hat{\mu}_1]^2 + \sum_{t=c+1}^{N_2} [R_z(t) - \hat{\mu}_2]^2. \tag{9}$$

Similar to those in Wang (2003) and Lund and Reeves (2002), the critical values of the $F_{\max}$ statistic here are obtained from 10 million simulations under H_0 for each series length n. Also, the F_c statistic above, which has an F distribution with $(1, n-2)$ degrees of freedom under H_0, can be used to assess significance of a documented step change (i.e., one that is supported by metadata) at time c. Both the $F_{\max}$ and F_c statistics, along with metadata (if available), are used in this study to identify R_z time series that have a significant step change. These R_z time series are further investigated, along with the related P_z and P_0 time series, to identify the cause and correct for the step change (via correcting the erroneous pressure values). Note that, in the absence of data error, the R_z time series should be random, with zero trend and no climate signal in general. These features of the R_z series render the use of reference series in the homogeneity test unnecessary. Actually, this is the case where the no-trend assumption of SNHT (and its equivalent tests, such as the one outlined in this paragraph) truly holds.

In general, both station relocation, without an update to the large change in station elevation for pressure reduction, and a change in observing instrument (e.g., sensor used in automatic stations) are often the causes for sudden changes in the mean of the R_z time series (and, hence, large μ and σ values). As shown in Fig. 4a, the R_z time series for the Lytton (British Columbia, Canada) station shows a clear step change on 1 July 1989, which was found to have arisen from a relocation of station with a decrease of 27.4 m in station elevation that was not accounted for in the calculation of station pressure from barometer readings (i.e., the elevation of the old site, which is 27.4 m higher than the elevation of the new site, was used in the calculation). This step change can also be identified from the original time series of P_z (Fig. 4b).

Figures 2c and 4a also show examples in which the assumption of IID Gaussian distribution for the R_z time series is violated. Specifically, the R_z time series exhibit a clear annual cycle (periodic variation). Our further investigation reveals that all such cases are associated with highly elevated stations (e.g., Old Glory Mountain, which has an elevation of 2347 m), which indicates that this very likely reflects a problem with the sea level pressure reduction (cf. Mohr 2004; Pauley 1998). The reduction of station pressure to mean sea level assumes a fictitious air column between the height of the station and the mean sea level. Usually, the air temperature decreases with increasing height from the surface; the rate of such a temperature decrease with increasing elevation is called the temperature lapse rate. However, the mean temperature of the fictitious air column is unknown, and is usually approximated in Canada by using a standard temperature lapse rate and $\overline{T}_{\mathrm{dry}}$ (cf. Savdie 1982). Also, a plateau correction (i.e., a correction for the plateau effects) has been added since November 1976 for all stations in Canada (Savdie 1982) in an attempt to get approximately the same amplitude of the annual variation of sea level pressure at all stations, regardless of their elevation (WMO 1964; Mohr 2004). However, as a result of the standard pressure reduction method, including the plateau correction, misleading sea level pressure values can be obtained for high-

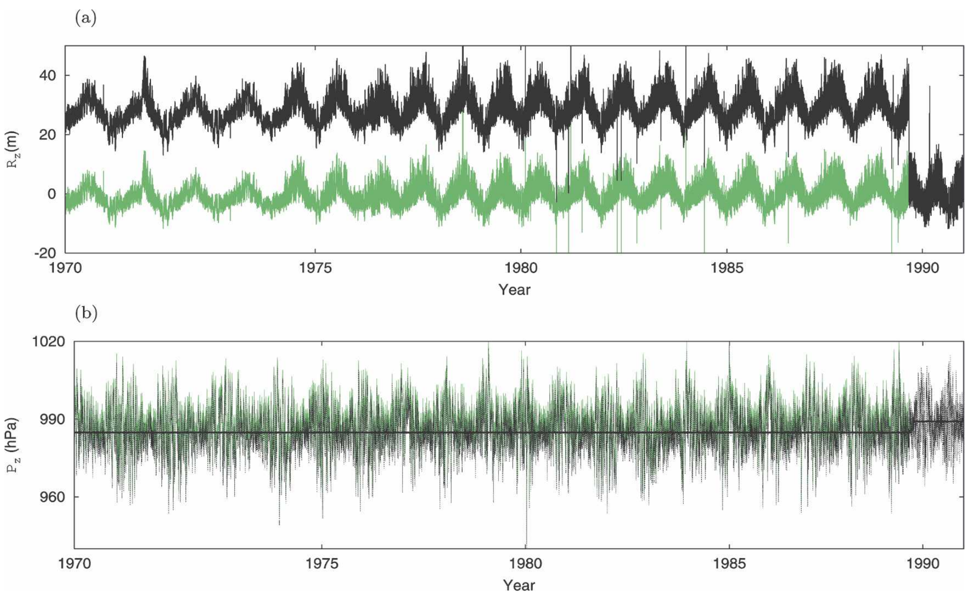

FIG. 4. Time series of (a) R_z and (b) P_z for the selected period of hourly observations at Lytton. The green curves indicate the adjusted values. The thick line in (b) shows the mean value of raw P_z before and after the changepoint.

altitude stations (Mohr 2004). We also notice that, in general, the plateau correction results in a slight increase in the variance of R_z for low-elevation stations (see period 1976–2001 in Figs. 2a,b), and a slight decrease for high-elevation stations (not shown). However, the discontinuity in variance is of a very small magnitude ($\pm$1 m of standard deviation of R_z, in general) when compared with the other systematic errors or with random errors we are trying to correct using this QA system. This study aims at correcting discontinuities in the mean, rather than in the variance (the later needs completely different statistical tests). Evaluation of the existing pressure reduction method and correction of the pressure reduction problem are beyond the scope of this paper. We need to be aware of this problem and keep in mind that a large R_z variance does not always correspond to discontinuities in the mean. We plotted and visually examined all of the R_z time series of large variance to determine the cause. To improve the validity of the IID Gaussian assumption for these time series for the homogeneity test, we estimate and remove the annual cycle from such a R_z time series before applying the homogeneity test to it (the series of departures from the annual cycle should be much closer to an IID Gaussian series).

As shown in Fig. 5, significant step change(s) in the R_z time series are found to have mainly arisen from either the 50-ft-rule problem (see those marked with a square), a long run of obviously wrong P_z values (see those marked with a circle), or a station relocation without updates to the changed elevation (see those marked with a triangle). The absolute values of the mean and standard deviation of R_z time series calculated from raw (uncorrected) pressure data, shown in Figs. 5 and 6a, respectively, are particularly large at the stations of obviously wrong P_z values, much larger than those at other problematic stations. In other words, the effects of these errors are much larger than those that are due to the 50-ft-rule problem or station relocation.

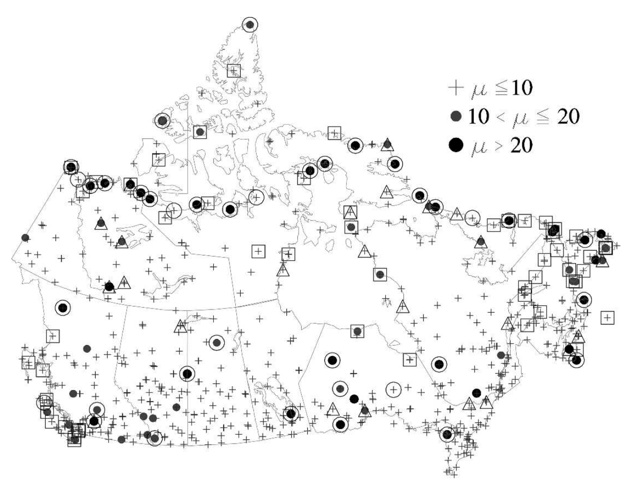

FIG. 5. Absolute values of the mean (m) of R_z time series calculated from raw pressure data. Stations of large step change(s) in the R_z time series are marked to indicate the cause being the 50-ft-rule problem (square), a long run of obviously wrong P_z values (circle), and station relocation without updates to the station elevation (triangle) in the calculation of station pressure.

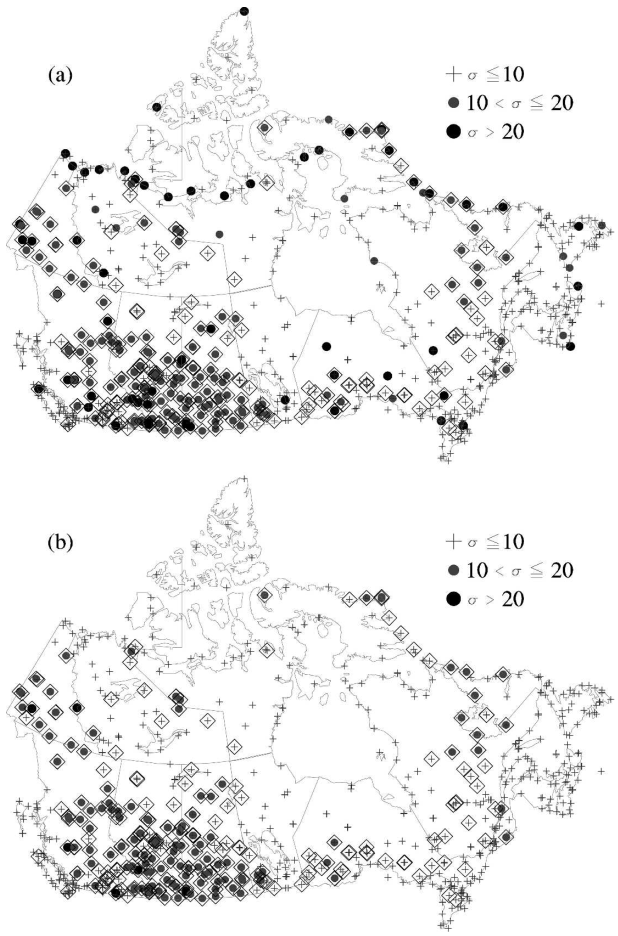

FIG. 6. The standard deviation (m) of R_z time series calculated from the (a) raw and (b) corrected pressure data. The diamond signs indicate stations of elevation greater than 305 m.

5. Correction of errors

Errors in meteorological data are very complicated and not easy to correct. Nevertheless, we should try our best not to reject, but to be able to correct erroneous data, especially for data-sparse regions (e.g., the Arctic region). An automatic error correction system is designed in the study.

It is highly desirable to know what caused the errors before we start to correct them. Table 2 lists the four types of errors that are most often found in our digital hourly pressure database, in addition to those that lead to a significant step change in the R_z time series. The vast majority of errors are of E1 and E2 (see Table 2). The E3 error is a profound problem in the Canadian hourly pressure data that were digitized from paper archives. In Canada, hourly pressure values used to be recorded (manually on paper) in tenths of hectopascal, and only the last three digits were recorded (e.g., "132" for a pressure of 10132, or "587" for 9587; unit: 0.1 hPa). The omitted base number (10 000 or 9000, or even 8000) needs to be added back during the digitization of our paper archives. Unfortunately, it is not always easy to determine which base number should be added, and the algorithm used to do so makes mistakes. This is why this type of error occurs and can be very hard (even impossible) to correct. This type of error sometimes persists for several hours or days, or even months (cf. Fig. 7), and can be mistaken as systematic biases caused by station relocation or instrument change, etc. Unfortunately, the same base number problem affects the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis dataset for the period from 1948 to 1967 (NCEP–NCAR 2006). Usually this type of error will not cause any exceedance of the climatological limits; these errors will only be detected by the hydrostatic check (combined with statistical homogeneity test) or by the PC check (i.e., the first and last erroneous data usually cannot pass the PC check). Therefore, it is sometimes impossible for us to determine which of P_z or P_0 is in error. A visual inspection of the time series segment often helps identify this type of error, which we do in this study.

TABLE 2. Errors most often found in the Canadian digital pressure database.

Type	Description
E1	One digit is miscoded (e.g., 1 is mistaken for 0, 2 for 3, 1 for 7)
E2	Digits are transposed (e.g., 1032.5 entered as 1035.2)
E3	Wrong base number added (e.g., "73" is taken as 907.3 hPa when it should be 1007.3 hPa)
E4	Station pressure and sea level pressure are transposed

a. Correction of systematic errors

The hydrostatic check combined with a statistical homogeneity test described in section 4 above is very useful in identifying and correcting systematic errors that lead to a significant step change in R_z time series, such as those caused by the 50-ft-rule problem, by a long run of obviously wrong P_z values (e.g., those shown in Fig. 3), and by station relocation without updates to the changed elevation. We found that all of the systematic step changes in R_z time series are associated with erroneous P_z (but correct P_0) values. Correction of this kind of systematic errors is relatively straightforward. These systematic errors have one common feature, that is, they are due to a change/error in elevation Z. Theoretically, we can simply use the correct station elevation and the hydrostatic model to calculate the correct val-

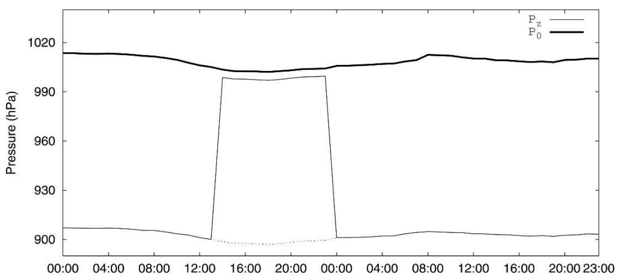

FIG. 7. An example of using a wrong base number when digitizing station pressure data recorded at Red Deer Airport (Alberta) from 0000 UTC 20 Apr to 2300 UTC 21 April 1953. The dashed line shows the correct values.

ues and use them to replace the corresponding erroneous P_z values. However, stations with these systematic errors could be in the elevated areas (except for those of the 50-ft-rule problem), and hence their R_z time series could have large periodic variations, such as those shown in Fig. 4a (which are due to the elevated area pressure reduction problem; see discussions in section 4). Replacement of erroneous P_z values with the corresponding P_z values calculated using the correct elevation would dampen the periodic feature of the R_z time series, which is not desired here. In this case, the desirable correction is the difference $\Delta = \overline{P}_z^c - \overline{P}_z^e$, where $\overline{P}_z^c$ denotes the mean (over the period of wrong elevation) of the calculated P_z values, and $\overline{P}_z^e$ denotes the mean of the erroneous P_z values. That is to say, we just need to add $\Delta = \overline{P}_z^c - \overline{P}_z^e$ to the erroneous P_z values to obtain the corrected P_z values. For example, for the Lytton case shown in Fig. 4, we add $\Delta = 3.4$ hPa to all the P_z values before 1 July 1989. Such an adjustment corrects for the systematic error, while retaining the periodic feature of R_z time series for highly elevated stations (see Fig. 4a). Of course, random errors are still to be identified and corrected.

b. Correction of isolated simple errors

Errors of E1 or E2 (see Table 2) are usually isolated cases (i.e., the values before and after it are correct for both elements) that are easy to correct, and hence are called simple errors. The algorithm we use to correct an isolated case of simple error is outlined in Fig. 8. First, we determine if the erroneous datum is an isolated error. If the answer is yes, we use the hydrostatic model (1) or (2a) and (2b) to estimate the correct value, depending on which element is in error. We use the recorded station elevation here, and a plateau correction was added in T_{mv} if the error occurs after November 1976 (the time the plateau correction was introduced in Canada), using the plateau correction parameters taken from the EC archive. Then, we compare this estimated pressure value with the original (erroneous) one, and compare its associated pressure change pattern with the corresponding pattern of the other element (the two elements should have the same pattern of pressure change) to see if we can determine the cause of error, and hence the correct value. As shown in Tables 3 and 4, if replacement of a digit or a transposition of two digits in the original data would make it approximately equal to the estimated value and ensure a consistency of pressure change between the two elements, this is a simple error; we apply the correction and flag it as "cor-

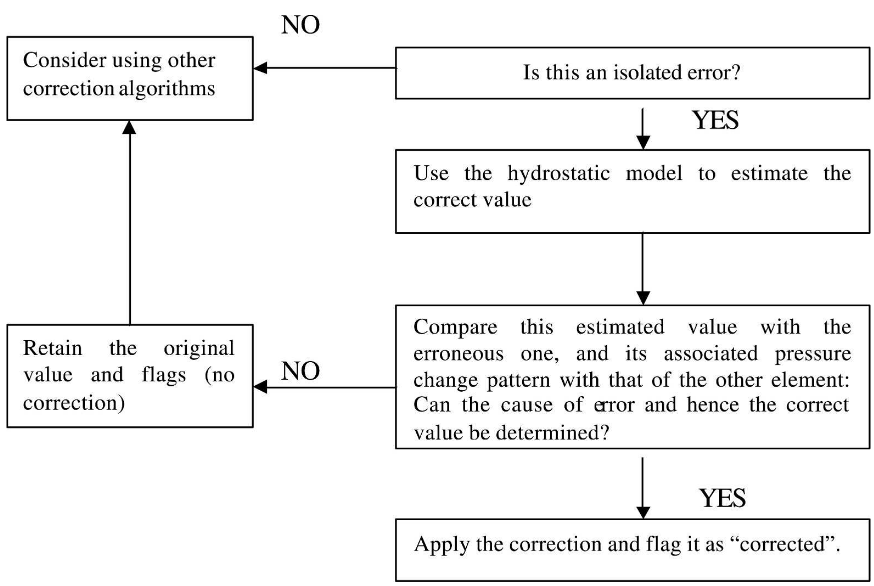

FIG. 8. An algorithm for correcting isolated simple errors.

TABLE 3. An example of the E1 error: "1029.1" was miskeyed as "1024.1" (16 Jan 1953 at station 7016294).

	2000 UTC	2100 UTC	2200 UTC	Model value	Correct value
P_0	1027.3	1024.1	1029.9	1029.1	1029.1
P_z	1017.6	1019.4	1020.2		

TABLE 5. An example of more than two digits in error "846.6" (4 Feb 1976 at station 1018642).

	0800 UTC	0900 UTC	1000 UTC	Model value	Correction
P_0	1023.5	846.6	1024.0	1023.8	1023.8
P_z	1027.4	1027.8	1027.9		

rected." If this is not a simple error and we are not able to determine the cause or the correct value, or if this is not an isolated case of error, we consider using other error correction algorithms (see the following subsections).

c. *Correction of isolated but complex errors*

Sometimes an isolated error is not a simple error (such as E1 or E2). For example, the value 846.6 in Table 5 is obviously wrong, inconsistent with either the corresponding or neighboring hourly P_z or P_0 values. The hydrostatic model estimate of the correct value is 1023.8, which would ensure a consistent pressure change pattern for both elements here and would pass the pressure limit check if it were used to replace the erroneous value 846.6. In other words, it is reasonable to replace 846.6 with 1023.8 in this case. Thus, we apply the correction and flag it as corrected.

d. *Human–machine interactive corrections*

The existing QA methods are often not able to correct erroneous data completely automatically. Human–machine interactive correction is usually applied when the automatic decision-making method cannot determine which element is in error. In this case, one needs to analyze manually the flag types and the original data for both elements to determine which element is in error and to estimate the correct value(s). In most cases, the correction is set to the value estimated using the hydrostatic model. For example, our analysis of the data shown in Table 6 reveals that the P_z value of 1001.6 was mistaken as the P_0 value, whose reasonable estimate is 996.7.

Correction of those systematic errors described in section 5a also requires human–machine interaction. The corresponding R_z time series, along with both P_0 and P_z time series, are plotted (as shown earlier in Fig. 3) and visually examined to determine the error type and its cause, because the automatic decision-making system is not able to determine which element (P_0 or P_z) is in error in this case, although the hydrostatic check is powerful in identifying and correcting this type of error.

Finally, there is a very small number of suspicious reports that even a specialist was not able to correct. This situation usually occurs when the hydrostatic check cannot be performed because of a missing element (e.g., dry-bulb temperature) that is needed as input to the hydrostatic model. In this case, we set the data as missing if they do not pass the climatological limits check. Otherwise, we accept them without any correction.

6. Analysis of the corrected data series

The QA approach described above is applied to each station for both pressure levels. Corrected data are stored with their corresponding flags. However, a second iteration of the QA was run with corrected data in order to detect any wrong corrections or erroneous data that went undetected at the first run.

The rate of random errors identified for most stations (systematic errors that were corrected as described in section 5a were not counted here) is less than 1‰. Of more than 1.8×10^8 hourly pressure data values (both levels) processed, approximately 4.1×10^6 (or 2.3%) data values (including systematic errors) have been corrected. About 30% of the detected errors can be automatically corrected, while human–machine interactive correction is needed to correct the other 70%.

TABLE 4. An example of the E2 error: "59.2" was miskeyed as "52.9" (25 Nov 1965 at station 4019080).

	0600 UTC	0700 UTC	0800 UTC	Model value	Correct value
P_0	959.5	952.9	958.6	957.4	959.2
P_z	1024.7	1024.1	1023.5		

TABLE 6. An example of mistakenly reporting the same value for both sea level pressure P_0 and station pressure P_z "1001.6" (28 Oct 1954 at station 7113534).

	1900 UTC	2200 UTC	0100 UTC	Model value	Correction
P_0	996.3	1001.6	998.2	996.7	996.7
P_z	1000.9	1001.6	1002.8		

As shown in Fig. 6b, the standard deviation of the R_z time series calculated using corrected station and mean sea level pressure data is much smaller, showing a more organized pattern in comparison with Fig. 6a. Large values are now seen only at the elevated stations.

The hydrostatic check plays an important role in the whole QA system. About 50% of all of the errors detected using the QA system were identified through this check; more specifically, all of the detected systematic errors were identified by the hydrostatic check in combination with the statistical homogeneity test, plus 20%–30% of the detected random errors were identified using the hydrostatic relationship alone. Also, our results show that it is reasonable to assume that the hourly dry-bulb temperature data used in the hydrostatic check are correct. The hydrostatic method can also be helpful in detecting inhomogeneities in atmospheric pressure data caused by station relocation, observer change, and so on, as shown in Fig. 4.

7. Concluding remarks

To build a high-quality database for atmospheric pressure (at both station and sea levels) in Canada, we have developed a comprehensive QA system that includes the hydrostatic check combined with a statistical homogeneity test, which was applied to hourly pressure data recorded in the last 50 yr at 761 Canadian stations. The combination of a physically based model with a statistical test is shown to be very powerful in detecting both random and systematic errors in pressure data and provides physically based, more accurate estimates of the adjustment/correction needed.

The results show that there are serious systematic errors in the Canadian historical atmospheric pressure data and that random error(s) are present for almost every station. Systematic errors are found to be caused either by the use of wrong station elevation values in the reduction of barometer readings to station or sea level pressure values (e.g., the 50-ft rule or station relocation without updating the station elevation), by transposing/swapping station and sea level pressure values, or by mistakes made in the archive data ingestion or data recording/digitization processes (e.g., use of a wrong base number). Fortunately, a vast majority of these errors can be detected and corrected by the QA system with either an automatic or interactive correcting method. The corrected P_0 and P_z data should be much more reliable and better suited for various climate studies, including their use in producing a 100-yr reanalysis (Compo et al. 2006).

It is also noticed that the introduction of the plateau correction in 1977 and the digital barometer (Vaisala barometer) around 2001 appear to cause small discontinuities in the pressure variance, which are not corrected in this study. The current QA system is designed only to detect and correct random errors and discontinuities in the mean (mean shifts).

Acknowledgments. The authors thank Dr. Gilbert Compo and Mr. Amir Shabbar for their helpful comments/suggestions on an earlier version of this manuscript. The three anonymous reviewers and editor Julie A. Winkler are also acknowledged for their helpful comments.

REFERENCES

Alexander, L. V., S. F. B. Tett, and T. Jonsson, 2005: Recent observed changes in severe storms over the United Kingdom and Iceland. *Geophys. Res. Lett.,* **32,** L13704, doi:10.1029/2005GL022371.

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

Allan, R. J., J. A. Lindesay, and D. E. Parker, 1996: *El Niño, Southern Oscillation and Climatic Variability*. CSIRO Publishing, 405 pp.

Christensen, R. A., and R. F. Eilbert, 1985: Seasonal precipitation forecasting with a 6–7 month lead time in the Pacific Northwest using an information theoretic model. *Mon. Wea. Rev.,* **113,** 502–519.

Collins, W. G., and L. S. Gandin, 1990: Comprehensive hydrostatic quality control at the National Meteorological Center. *Mon. Wea. Rev.,* **118,** 2752–2767.

Compo, G. P., J. S. Whitaker, and P. D. Sardeshmukh, 2006: Feasibility of a 100-year reanalysis using only surface pressure data. *Bull. Amer. Meteor. Soc.,* **87,** 175–190.

Environment Canada, 1970: Manual of surface weather observations. 6th ed. 307 pp.

——, 1977: Manual of surface weather observations. 7th ed. 419 pp.

——, cited 2004: The National Archives System QC program checks for hourly SA's. [Available internally at http://nadm.ontario.int.ec.gc.ca/Intranet/nadm/nas/quality_control/hourlies_e.html.]

Gandin, L. S., 1988: Complex quality control of meteorological observations. *Mon. Wea. Rev.,* **116,** 1137–1156.

Graybeal, D. Y., A. T. DeGaetano, and K. L. Eggleston, 2004: Complex quality assurance of historical hourly surface airways meteorological data. *J. Atmos. Oceanic Technol.,* **21,** 1156–1169.

Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson, Eds., 2001: *Climate Change 2001: The Scientific Basis.* Cambridge University Press, 881 pp.

Hubbard, K. G., S. Goddard, W. D. Sorensen, N. Wells, and T. T. Osugi, 2005: Performance of quality assurance procedures for an applied climate information system. *J. Atmos. Oceanic Technol.,* **22,** 105–112.

Jones, P. D., M. J. Salinger, and A. B. Mullan, 1999: Extratropical circulation indices in the Southern Hemisphere based on station data. *Int. J. Climatol.,* **19,** 1301–1317.

Kaplan, A., Y. Kushnir, and M. A. Cane, 2000: Reduced space optimal interpolation of historical marine sea level pressure: 1854–1992. *J. Climate,* **13,** 2987–3002.

Kunkel, K. E., and Coauthors, 1998: An expanded digital daily database for climatic resources applications in the midwestern United States. *Bull. Amer. Meteor. Soc.,* **79,** 1357–1366.

Le Blancq, F. W., 2003: Newfoundland low and pressure tendencies, 11 February 2003. *Weather,* **58,** 239–242.

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

Meek, D. W., and J. L. Hatfield, 1994: Data quality checking for single station meteorological databases. *Agric. For. Meteor.,* **69,** 85–109.

Mohr, M., 2004: Problems with the mean sea level pressure field over the western United States. *Mon. Wea. Rev.,* **132,** 1952–1965.

NCEP–NCAR, cited 2006: Reanalysis PSFC problem 1948–1967. [Available online at http://wwwt.emc.ncep.noaa.gov/gmb/bkistler/psfc/psfc.html.]

Nkemdirim, L. C., and D. Budikova, 2001: Trends in sea level pressure across western Canada. *J. Geophys. Res.,* **106,** 11 801–11 812.

Pauley, P. M., 1998: An example of uncertainty in sea level pressure reduction. *Wea. Forecasting,* **13,** 833–850.

Saucier, W. J., 1955*: Principles of Meteorological Analysis*. University of Chicago Press, 438 pp.

Savdie, I., 1982: AES barometry program. Network Planning and Standards Division, Data Acquisitions Services Branch, Environment Canada Tech. Record 9, 12 pp.

Shafer, M. A., C. A. Fiebrich, D. S. Arndt, S. E. Frederickson, and T. W. Hughes, 2000: Quality assurance procedures in the Oklahoma Mesonetwork. *J. Atmos. Oceanic Technol.,* **17,** 474–494.

Shewhart, W. A., 1980: *Economic Control of Quality of Manufactured Product.* American Society for Quality Control, 501 pp.

Slonosky, V. C., and E. Graham, 2005: Canadian pressure observations and circulation variability: Links to air temperature. *Int. J. Climatol.,* **25,** 1473–1492.

Smith, T. M., and R. W. Reynolds, 2004: Reconstruction of monthly mean oceanic sea level pressure based on COADS and station data (1854–1997). *J. Atmos. Oceanic Technol.,* **21,** 1272–1282.

Trenberth, K. E., and D. A. Paolino Jr., 1980: The Northern Hemisphere sea-level pressure data set: Trends, errors and discontinuities. *Mon. Wea. Rev.,* **108,** 855–872.

Wang, X. L., 2003: Comments on "Detection of undocumented changepoints: A revision of the two-phase regression model." *J. Climate,* **16,** 3383–3385.

——, and Y. Feng, cited 2007: RHtestV2 user manual. Climate Research Division, Atmospheric Science and Technology Directorate, Science and Technology Branch, Environment Canada, Toronto, ON, Canada, 19 pp. [Available online at http://cccma.seos.uvic.ca/ETCCDMI/RHtest/RHtestV2_UserManual.doc.]

——, H. Wan, and V. R. Swail, 2006: Observed changes in cyclone activity in Canada and their relationships to major circulation regimes. *J. Climate,* **19,** 896–915.

WMO, 1954: Reduction of atmospheric pressure. Tech. Note 7, 35 pp.

WMO, 1964: Note on the standardization of pressure reduction methods in the international network of synoptic stations. Tech. Note 61, 42 pp.

Wright, P. B., 1984: Relationships between indices of the Southern Oscillation. *Mon. Wea. Rev.,* **112,** 1913–1919.