HOMOGENISATION OF A CANADIAN SURFACE PRESSURE DATABASE

Edward Graham* and Victoria Slonosky**

Climate Monitoring and Data Interpretation Division
Climate Research Branch, Meteorological Service of Canada
Environment Canada, 4905 Dufferin St
Downsview, Ontario, M3H 5T4, Canada

Current Affiliations:
*Département de Géographie
Université de Fribourg, Pérolles
CH-1700 Fribourg, Switzerland
Tel: 00-41-26-3009246
Fax: 00-41-26-3009746
email: edward.graham@unifr.ch

**Ouranos Consortium for Regional Climate Change and Adaptation
Department of Atmospheric and Oceanic Sciences
McGill University, 550 Sherbrooke St W
Montréal, Québec, H3A 1B9 Canada

Published in "Proceedings of Fourth Seminar for Homogenisation and Quality Control in Climatological databases. Budapest, Hungary. 6-10, October 2003"

ABSTRACT

This paper describes the collection, checking and homogenisation of a Canadian atmospheric surface pressure database. The object of the exercise was to create a database of monthly mean surface pressure for as many stations as possible across Canada as far back in time as possible. Data sources included the World Weather Records, Monthly Climatic Data for the World Bulletins, the Global Historical Climate Network and the electronic meteorological report archives of Environment Canada. Much of the earlier data was in paper form and had to be digitized by hand. Over 66,000 individual mean monthly pressure values were obtained, with a missing value rate of 5.9%. The homogenisation procedures used were the Standard Normal Homogeneity Test (SNHT; Alexandersson and Moberg 1997) and Multiple Comparison Analysis (MCA; as used by Slonosky et al 1999). In addition, simple subtraction of sea-level pressure from station-level pressure revealed a major inhomogeneity which took place in 1977, when computer generated pressure reduction tables were used for the first time by the Meteorological Service of Canada, and when the meteorological reporting procedure was brought into alignment with the World Meteorological Organisation's guidelines. As a result, the final homogenised database shows appreciable differences in trends compared to the unhomogenised series. The final database has been used by Slonosky & Graham (2003) in the statistical analysis of trends and variability of surface pressure across Canada during the 20th century.

1. INTRODUCTION

This project was conceived with the aim of analysing decadal to century scale variability of surface pressure over Canada. For the study of atmospheric circulation dynamics over these timescales, an extensive temporal, spatial and homogenous surface air pressure database is necessary. Many observations of surface pressure for Canada exist back to the 19th century, but the majority are in paper form. This project was the first attempt to gather all the different data sources together, and to digitise, quality control and homogenise them. The homogenisation methods are especially important as non-climatic changes, such as site relocations, instrument replacement, or changes in the observation practice, including changes in the time of observation or calculation procedure, can introduce biases of the same magnitude as the long term climatic variability of pressure into the series, leading to spurious trends and variability (Vincent et al. 2002, Slonosky et al. 1999, Peterson et al. 1998, Young 1993). This paper describes the collection, quality control and homogenisation part of the Canadian pressure database project. For a more detailed climatological analyses of the final database, readers are referred to Slonosky and Graham (2003), who have used the database in the analysis of decadal scale circulation variability over Canada and Greenland.

2. SOURCES OF DATA

The climatic variable of interest in this study was mean monthly surface air pressure. These data were all recorded at official observing stations of the Meteorological Service of Canada observational network, starting in 1841 for Toronto and in 1873 for a selection of other stations across Canada. In this project, we used data from 1873 onwards. There were four primary sources of data, namely:

a) The World Weather Records, published each decade by the Smithsonian Miscellaneous Collections. These contain monthly mean sea and station level air pressure data in paper form for assorted Canadian and Greenland stations from 1874 to 1990. Digitisation by hand was necessary.
b) The Monthly Climatic Data for the World Bulletins. These are individual monthly bulletins, containing sea and station level air pressure records in paper form from 1959 to 1999. Digitisation by hand was also necessary.
c) Global Historical Climate Network; a few Canadian stations are available in digital format from 1873-2000.
d) Environment Canada electronic database. This archive contains data for most Canadian reporting stations from 1953 onwards, but in hourly synoptic format. It is accessible in digital format on the internet by user interface (password-protected). Due to the hourly nature of the data, it was necessary to calculate monthly means of air pressure for each station from

this data. In the interests of data quality, months with fewer than 21 reporting days of data were discarded, as were days with fewer than three different synoptic reports.

Other miscellaneous sources of data included Climate Research Unit of the University of East Anglia. Data from Greenland was also included in this study, due to the existence of a few long term pressure records located in southern Greenland and the Davis Strait, close to north-eastern Canada. In total, over 66,000 individual mean monthly pressure values were obtained during the course of this project, with a missing value rate of 5.9%.

3. COMMON PROBLEMS AND EXAMPLES

As was to be expected, there were many problems encountered during the preliminary checking and quality control procedure. Table 1 lists the main problems in order of magnitude. The biggest problems encountered were related to unreported changes in station or barometer location and elevation. In the absence of metadata (information on the barometer or station history), we had to use standard homogenisation techniques to correct for such changes (these techniques are outlined in section 5).

Another major problem encountered was the so-called “50-foot” rule. This was a bizarre rule which had been introduced into Canada during the 1930s (at the start of aviation). It stated that stations with an altitude of between 0 and 50 feet above sea level should report station pressure as being identical to sea-level pressure i.e. station pressure was also reduced to sea level (McMaster 1975, Upton 1972). However, stations at an altitude greater than 50 feet did not have this correction applied. Furthermore, a site re-location of less than 50 feet in height resulted in no new established elevation for the station in question – a correction factor was assumed to have been added. This rule was abolished in 1976 under World Meteorological Organisation guidelines (Environment Canada 1976). It also coincided with metrification and computerisation of the pressure reduction tables at same time, adding to additional complications. McMaster (1975) states that mean sea level pressures would not be affected, but station pressures (for those stations previously at an altitude between sea level and 50 feet) would experience a slight drop (McMaster 1975). Infact, analysis of this database shows that the correction to station level pressure records was often the order of 1mb or more. As a result of these reporting changes between 1976 and 1977, the vast majority of station records (77%) showed some sort of spurious jump at the 1976/1977 boundary.

Other errors uncovered during the course of this project included differences between the World Weather Record values and the Environment Canada electronic archives, due to differences in the rounding procedure. There were also frequent errors resulting from confusion between sea and station level pressures, and between metres and feet in altitude (see Table 1).

Problem	Example
Unreported changes in station location and/or station /barometer elevation	Standard homogenisation techniques were used to correct for these errors.
“50-foot” rule	All Canadian stations with an altitude between 0 and 50 feet were affected from ~1930 to 1976 (see text for more details).
Computerisation	Computer generated pressure reduction tables were used for the first time in 1976 (replacing primitive desk calculators) – the majority of Canadian stations were affected.
Conversion/rounding errors	Between 1951 and 1970, the World Weather Records (WWR) were rounded to the nearest millibar; the averages are 0.05mb higher than the Environment Canada synoptic archives (which were rounded to the nearest 0.1mb).
Confusion between feet and metres	This is a common problem (but especially in 1976 when metrification was adopted in Canada). For example, Clyde station was given an altitude of 26 feet in the WWRs of the 1940s but the WWRs of the 1990s were still listing it as 26 metres.
Confusion between “sea” and “station”	Sea level pressure was sometimes given as station pressure, and vice versa (human typing error).
“sea” and “station” data transposed	The WWR for Greenland have identical values for sea and station level pressure, regardless of station location or altitude.
Random differences between GHCN and WWR (e.g. typos, outliers)	These were usually easily identified by visual analysis.

Table 1: common problems and errors uncovered during the collection and preliminary quality control of the Canadian atmospheric pressure database. Examples of each problem are given in the right-hand column.

4. PROVENANCE

After a rigorous procedure of preliminary checking and quality control, 73 individual station series were retained from Canada and Greenland. The stations selected are shown in Figure 1. As can be seen, the longest station series (with start dates before 1900; red colour) are to be found primarily in southern Canada and southern Greenland. Those stations with records from between 1900 and 1930 (blue colour) are generally clustered in north-western Canada. The stations with the shortest series (only 50 or 60 years) are found almost exclusively in the Canadian Arctic i.e. we do not have reliable data from the Arctic from before the second world war.

Station level observations were selected in preference to sea level values, as it was considered that station level pressures were more reliable. Fewer calculations are involved in obtaining station level pressures, and therefore there were fewer opportunities for calculation related inhomogeneities to occur. In several cases only sea or station level observations were available for certain portions of a station record; in these cases, either the station information or, if the station information was unavailable, monthly transfer functions were used to relate the segments and produce a uniform station or sea level record.

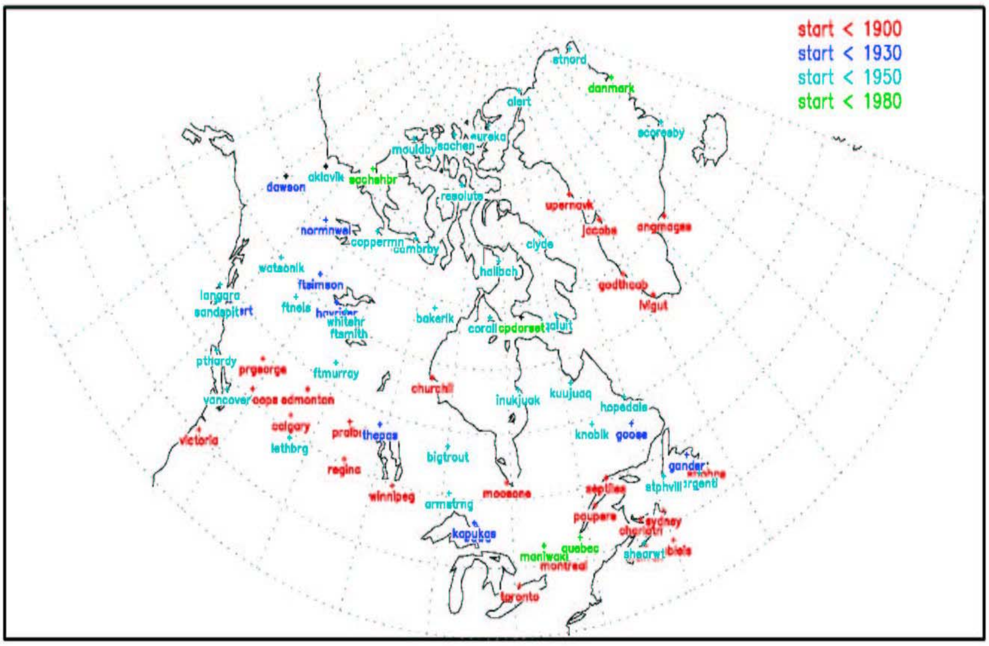
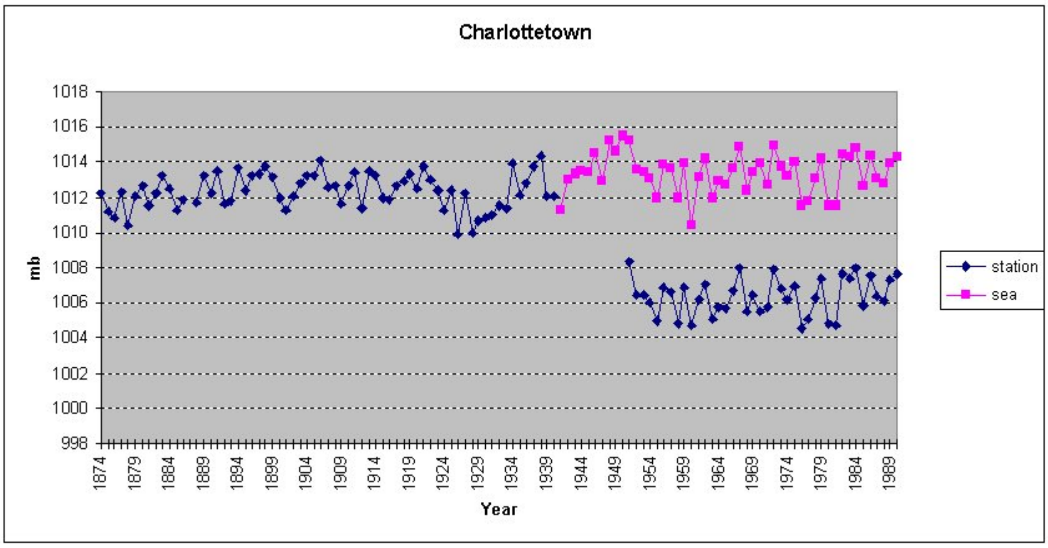


Figure 1. Location of station series.



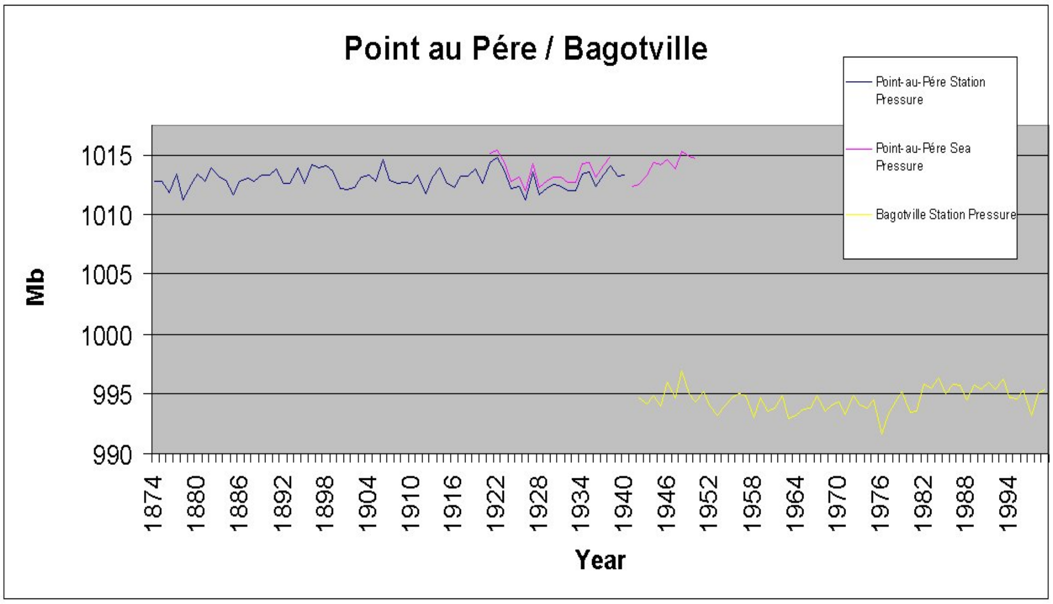


Figure 2: a) Mean annual atmospheric pressure (mb) records for Charlottetown, Prince Edward Island, 1874-1998. Station pressure is available from 1874-1940 and 1951-1998, but there is a gap from 1941 to 1950. Fortunately, a sea-level pressure record overlaps with the station pressure record, and using monthly transfer functions a complete station level pressure record is obtainable. b) Point-au-Père is available as a combination of station level and sea level series from 1874 to 1950, but records then cease. However, the nearby station of Bagotville commences in 1942, and using monthly transfer functions for the overlapping period, a complete station level pressure record is obtainable.

The example of Charlottetown (Prince Edward Island) is shown in figure 2(a). Station level pressure was available from 1874 onwards but stopped in 1940. Station data resumed again in 1951, but with a mean long term value of about 6mb lower than before, most likely due to a station re-location. Fortunately, a sea level pressure record overlaps the broken period, and using monthly transfer functions, a complete station level pressure record was obtained.

Several observation series which started relatively early ended abruptly; when possible, nearby stations were used to complete the series and produce a composite series of longer duration. Monthly transfer function were again calculated using the overlapping portions of both series to produce the composite series; the earlier segments were reduced to conform with the later, most modern segment. The example of Point-au-Père, Québec, is shown in figure 2(b). Station level data are available from 1874 to 1940, with a sea level pressure record from 1921 to 1950, after which records cease. However, the nearby station of Bagotville started recording data in 1942, and using monthly transfer functions, all three segments of atmospheric pressure can be reduced to one continuous record through to the present.

With regard to sea-level pressure reports, it is worth noting that J.G. Potter (1955) made a decision to exclude all data prior to 1940 when constructing monthly mean sea level pressure maps for Canada (Potter 1955). Potter notes that prior to 1940s, barometers were not inspected regularly, and “periods of 20 years or more passed without reports on the index”.

5. HOMOGENISATION METHODS

A database is said to be “homogenous” if the internal variations are caused by weather and climate. In reality, however, the internal variability caused by weather and climate may be large enough to mask or obscure step changes or trends caused by inhomogenieties. In order to track down such inhomogenieties, the following methods were employed in this project.

5.1. Subtraction of Station Pressure from Sea Level Pressure

Simple and straightforward, this method gave considerable insight into an individual station’s history. The subtraction of the station level series from the sea level series was plotted visually. Several inhomogeneities were discovered using this technique, including a nation-wide inhomogeneity which otherwise may have gone undetected.

In November 1976, computer generated pressure reduction tables were used for the first time by the Meteorological Service of Canada (formerly the Atmospheric Environment Service), replacing the previous calculations which were made either manually or by using primitive desk calculators. The algorithms used to calculate the station pressure correction and the station to sea pressure corrections were also adapted to World Meteorological Organisation (WMO) guidelines at this time, including the addition of plateau correction (Savdie 1982). Also at the same time, the “50-foot” rule (as discussed in section 3) was abolished (Environment Canada 1976, McMaster 1975, Upton 1972). The result was a noticeable jump in both sea level and station level pressure records, but especially in the sea level record across almost all parts of Canada. The sea level correction was sometimes of the order of several millibars. Coastal and low-lying stations which had been subject to the “50-foot” rule saw a strong drop in station pressures at this time. These discontinuities were usually noticeable upon subtraction of station pressure from sea level pressure (see figures 3a and 3b). If undetected, this would have led to spurious trends of atmospheric pressure across large parts of Canada. A total of 50 stations (more than three-quarters) showed some sort of jump across this discontinuity, with a mean correction factor of 0.92mb. The highest correction factor was 5.5mb at Dawson in north-western Canada.

5.2. Standard Normal Homogeneity Test and HadSLP

The Standard Normal Homogeneity Test (SNHT; Alexandersson and Moberg 1997) is a test which uses differences in standard deviations between a candidate series and a “pure” reference series to find breaks. As we did not have a reference series in this project, we compared our Canadian pressure dataset with the UK Met Office Hadley Centre’s northern hemisphere HadSLP gridded dataset, which is available for the period 1873-1998. It is acknowledged, however, that the HadSLP dataset is by no means homogenous – especially for the data sparse regions of northern Canada and the Arctic. It transpires that the HadSLP grid was digitised from old weather charts, in which meteorologists may have over-extrapolated the extension of lower latitude weather features, and may also have assumed the predominance of a virtual, quasi-permanent “Arctic High” (see Jones 1987). In practice, this method of homogenisation was suitable only in the detecting of significant jumps and outliers. Often, it demonstrated the shortcomings of the HadSLP gridded dataset. This procedure also assumed that no change in climate occurred between periods of jumps

To compare the datasets, a fortran program called gridcheck was run on the database, using the HadSLP grid as a reference series (the HadSLP grid has a size of 5° by 10°). The nearest value of the grid was interpolated for each station location. Outliers and jumps were detected by using the following equation (differences of greater than 4mb were flagged):

$$outlier = (stnvalue - stnmean) - (gridvalue - gridmean)$$

e.g. (980mb – 990mb) – (1014mb – 1020mb) = -10 – (-6) = -4

Sometimes, we had to infill data from the HadSLP grid, in order to make a continuous station record (but it was only undertaken in special circumstances for a few stations). Smoothing of the infilled HadSLP values was necessary, however, as the internal variability of HadSLP grid was greater than that of the station series that needed infilling. The following two equations were used:

$$gridvalue = gridraw - (\Delta diff)$$

$$newstnvalue = \frac{(gridvalue - \overline{grid})}{grid\sigma}.stn\sigma + \overline{grid}$$

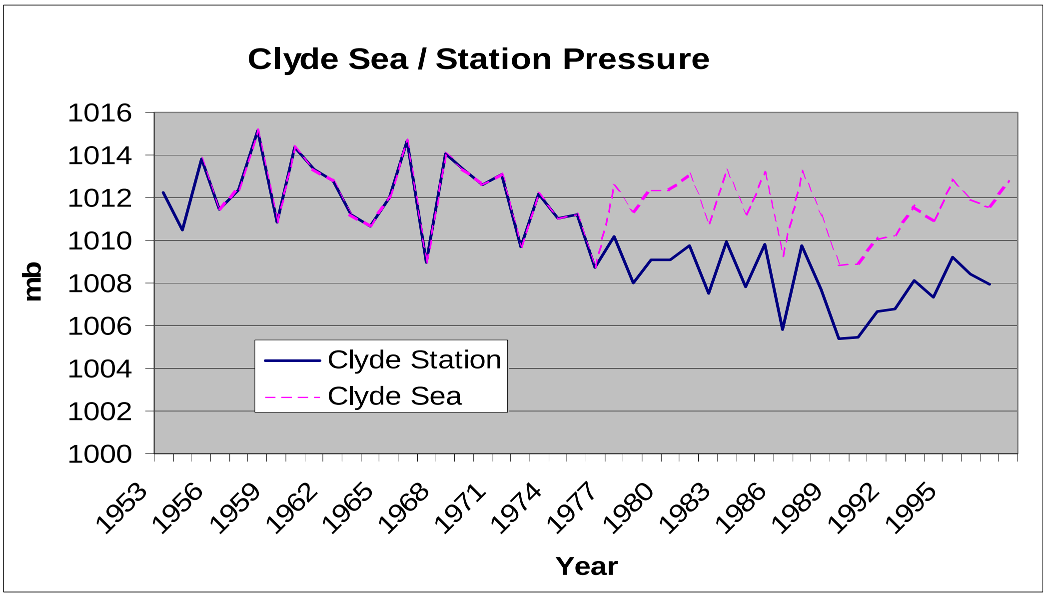
gridvalue is simply the raw interpolated grid value (*gridraw*) minus the mean difference between grid and station data for each month. Then we subtract the mean value of the grid for that point, and divide by the grid's standard deviation. Then we multiply by the station series standard deviation and re-add the mean value for the grid at that point again. This gives is a new value for the station record. This method simply reduces the variability of the infilled period.

5.3. Multiple Comparison Analysis

Due to the shortcomings of the previous method, Multiple Comparison Analysis (MCA; Slonosky et al,1999) was used as the next stage in the homogenisation procedure. The method involves the selection and comparison of data from the four nearest stations in each direction from a candidate station. It is a semi-objective iterative technique based upon graphical inspection of difference series between neighbouring stations. For each candidate station, four neighbouring stations were chosen for comparison, one in each cardinal direction (if a station was at a boundary or edge, simply the four nearest stations were chosen). Four difference series were calculated, and the graphical results plotted together (see figure 4 for an example). If a jump or discontinuity occurred in more than two difference plots, the jump was attributed to the candidate station. All stations were inspected, and all stations were then adjusted, and the process repeated to ensure that the discontinuities were correctly attributed. Adjustment factors were calculated to adjust identified inhomogeneous periods to the modern portion of the series. This method also assumes that no change in climate occurs between periods of jumps. An evaluation of this method, compared to the SNHT method (Alexandersson and Moberg 1997) and the Bayesian method developed by Caussinus and Mestre (CMT; 1996) is given in Slonosky et al. (1999). For pressure data, generally speaking, MCA is better than CMT, which in turn is better than the SNHT and the raw data series respectively.

As a final method of homogenisation, we did a visual check for “bulls eyes” – these are outliers that escaped the above homogenisation methods. This is a rather subjective method and only a few corrections were made.

Figure 3; a) Clyde mean annual sea level (dotted pink line) and station level (dark blue solid line) atmospheric pressure 1953-1999. b) subtraction of the two series. Note the discontinuity in late 1976 when computer generated pressure reduction tables were used for the first time (these introduced new algorithms in the calculation of sea level pressures, causing a rise). The "50-foot" rule was also abolished in 1976, leading to a drop in station-level pressure.



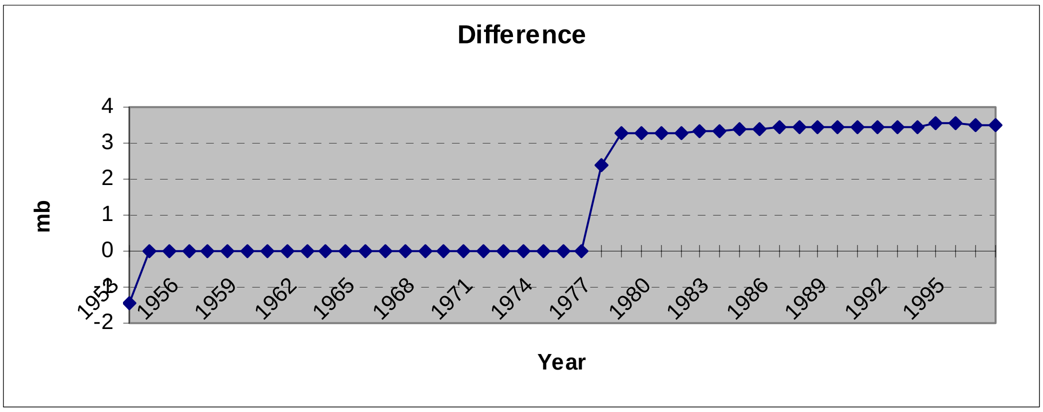
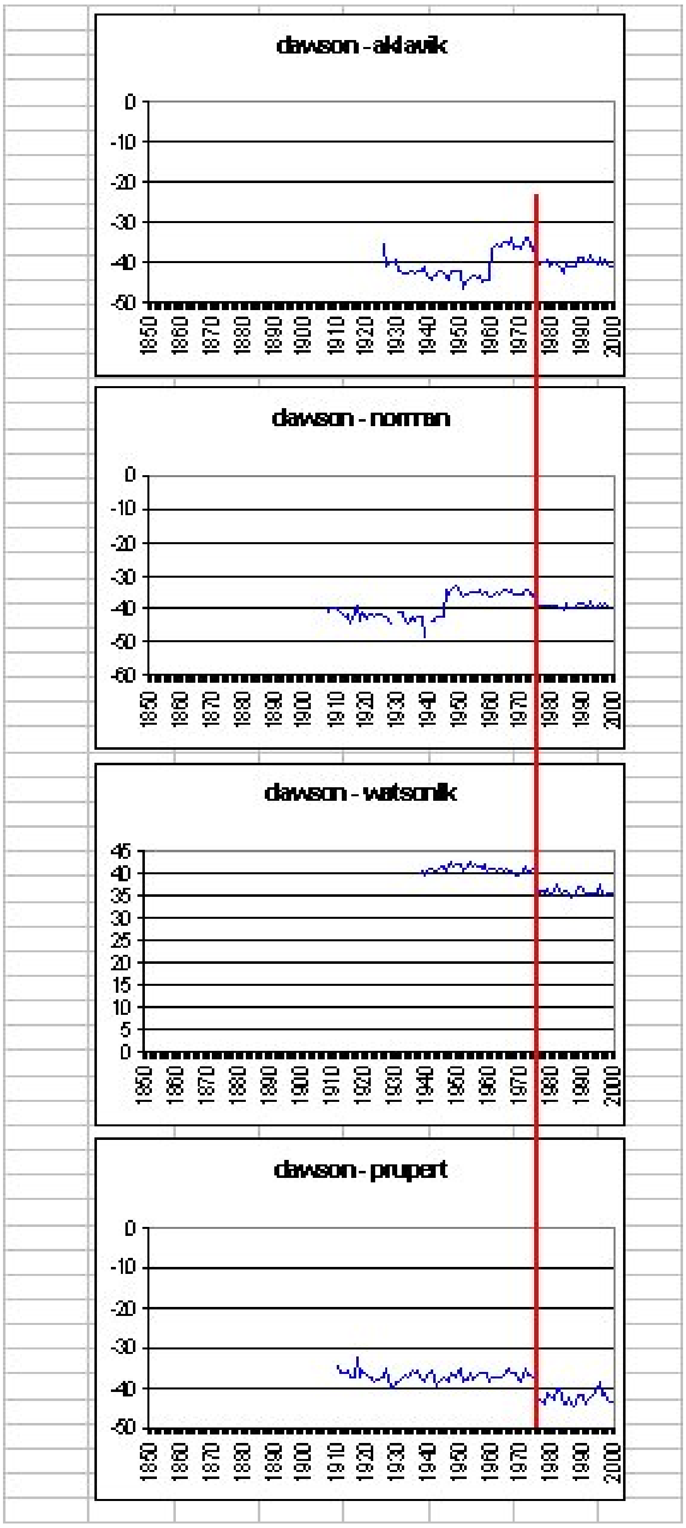


Figure 4: Multiple Comparison Analysis (MCA) for Dawson (Yukon). When the station series is compared with that of the four neighbouring stations (Aklavik, Norman Wells, Watson Lake and Prince Rupert) a distinct jump occurs in 1976 in all four comparisons (marked by the red line) – this jump can therefore be attributed to Dawson.



6. A FEW RESULTS

The final homogenised database shows appreciable differences in trends of atmospheric pressure when compared to the unhomogenised data. Table 2 lists the regional trends of standardized anomaly pressure data before and after the homogenization process. In the unhomogenised series, statistically significant trends (at the 95% confidence level, p-value < 0.05) are evident in south-eastern and south-western parts of Canada (South of 55N, east of 100W; South of 55N, west of 100W, respectively). The final database (after homogenisation), however, shows that these trends are spurious, and that no significant trends in atmospheric pressure are evident in any part of Canada during the 20th century.

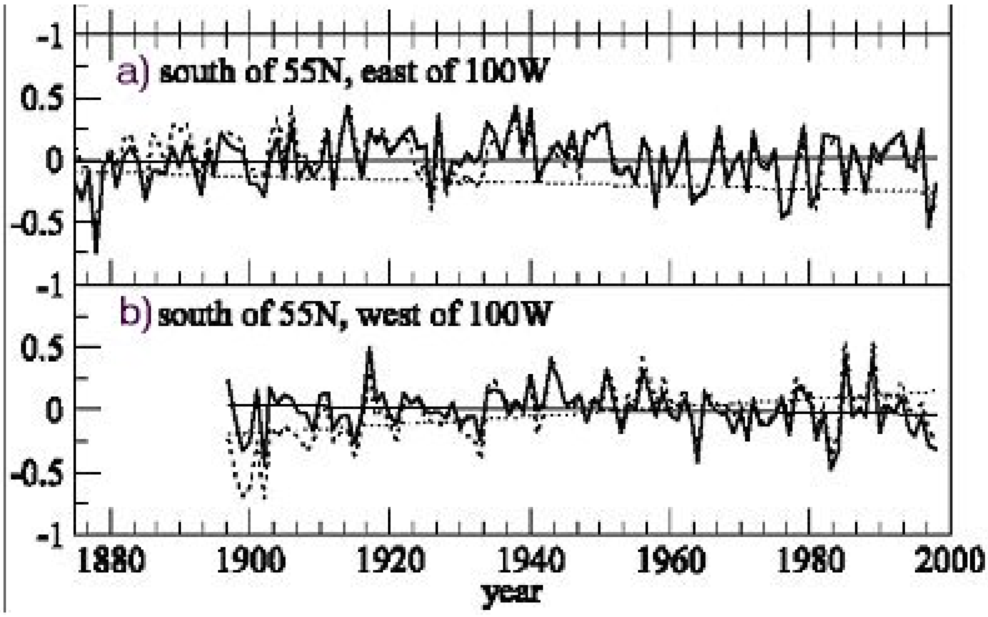
Table 2. Trends in original and homogenized data

All stations:	Original (mb per decade)	Homogenized (mb per decade)
North of 67N	-0.0285	-0.0123
North of 55N	0.0000	-0.0121
North of 55N, east of 100W	0.0053	-0.0078
South of 55N, east of 100W	-0.0117*	0.0030
North of 55N, west of 100W	-0.0130	-0.0128
South of 55N, west of 100W	0.0324*	-0.0130

*statistically significant at the 95% confidence level

These results emphasise the extreme importance of rigorous homogenisation in any climate data analysis, as there are statistically significant trends in the unhomogenised series for southern Canada. Figure 5 shows these trends more clearly in a graphical format for south-eastern and south-western Canada. The dotted lines show the spurious significant trends in the unhomogenised data (upward trend in southwestern Canada, downward trend in southeastern Canada); the solid lines show the final non-significant trends in the all areas of the homogenised series. EOF analyses has also be performed on the Canadian Stations pressure dataset, and a full statistical and climatological analyses of the database is given in Slonosky & Graham (2003), where they present and analyse trends and variability of surface pressure across Canada during the 20th century.

Figure 5; area-averaged standardized pressure anomalies for before and after homogenisation, for a) all stations south of 55N, east of 100W, and b) all stations south of 55N, west of 100W. The dotted lines show the statistical significant trends before homogenisation; the solid lines show the no-significant trends after homogenisation.



7. ACKNOWLEDGEMENTS

The authors would like to thank the organisers of the 4th seminar for homogenisation and quality control in climatological databases (Budapest, Hungary, 6-10 October 2003), especially Sándor Szalai and Tamás Szentimrey. We would also like to thank Martin Beniston (University of Fribourg) for allowing time to research and write this article. During the data retrieval phase of the project, Roberta McCarthy and Malcolm Geast of Environment Canada where paramount in providing data, information and sources.

8. REFERENCES

Alexandersson, H., and A. Moberg, 1997: Homogenization of Swedish temperature data. Part I: A homogeneity test for linear trends. *Int. J. Climatol.,* **17,** 25-34.

Caussinus, H. and O. Mestre, 1996: Towards new tools and methodologies for relative homogeneity testing. *First Seminar for Homogenization of surface climatological data*, Budapest, Hungary, Hungarian Meteorological Service, 62-82.

Environment Canada, 1976: *Internal Memorandum No. OBS 2-76*, Toronto, Canada, Atmospheric Environment Service, November 1976, 4pp.

Jones, P.D., 1987: The early twentieth century Arctic high – fact or fiction?. Climate Dynamics, 1, 63-75.

McMaster, R.S.,1975: *Memorandum No. 8437-1 from ACNS to ACNC*, Toronto, Canada, Data Standards Division, Government of Canada, 22 September 1975, 2pp.

Peterson, T.C. D.R. Easterling, T.R. Karl, P. Groisman, N. Nicholls, N. Plummer, S. Torok, I. Auer, R.Boehm, D. Gullett, L Vincent, R. Heino, H. Tuomenvirta, O. Mestre, T. Szentimrey, J. Salinger, E. Førland, I. Hanssen-Bauer, H. Alexandersson, P. Jones and D. Parker, 1998: Homogeneity adjustments of *in situ* atmospheric climate data: a review. *Int. J. Climatol.,* **18,** 1493-1517.

Potter, J.G., 1955: *Monthly mean sea level pressure maps for Canada.* Toronto, Canada, CIR-2710, Tec-224, Meteorological Division, Department of Transport, 18 October 1955, 10pp.

Savdie, I.,1982: *AES Barometry Program*, Technical Record No. 9. Toronto, Canada. Network Planning and Standards Division, Data Acquisition Services Branch, Environment Canada, August 1982, pp 1-11.

Slonosky, V.C., P.D. Jones, and T.D. Davies, 1999: Homogenization techniques for European monthly surface pressure series. *J. Clim.,* **12,** 2658-2672.

Thompson, D.W.J., and J.M. Wallace, 1997: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.,* **25,** 1297-1300.

Slonosky, V.C. and E. Graham, 2003: Canadian Surface Pressure and Atmospheric Circulation Variability: A Homogeneity Exercise, submitted to Journal of Climate.

Upton, F.T., 1972: *Headquarters Standards and Procedures for Barometry in Canada*. Toronto, Canada. Atmospheric Environment Service, 17pp.

Vincent, L.A., X. Zhang, B.R. Bonsal, W.D. Hogg, 2002: Homogenization of daily temperatures over Canada. *J. Clim.,* **15,** 1322-1334.

Young, K.C., 1993. Detecting and removing inhomogeneities from long-term sea level pressure time series. J. Clim, 6,1205-1220.