2.10 HOMOGENEITY ASSESSMENT OF CANADIAN PRECIPITATION DATA FOR JOINED STATIONS

Éva Mekis* and Lucie Vincent
Meteorological Service of Canada, Toronto, Ontario

1. INTRODUCTION

It is often essential to join climatological station observations in order to produce long time series that can be used for climate change analyses. The join of the observations can sometimes create artificial steps and adjustments are often needed to create a time series representing as much as possible climate variations in time. The first generation of Canadian adjusted daily precipitation time series was produced in 1996 and updated annually since then (Mekis and Hogg, 1999). Adjustments were applied on daily rain and snow separately at 495 stations across the country. The rehabilitation method includes adjustments for changes of rain gauge types, new snow density index was implemented and corrections for trace measurements were also included. When station observations were joined in time, adjustments were derived from Simple Ratio Method (Thom, 1966). Unfortunately, this method can be used only if there are common observations over a period of time. The main objective of this work is to investigate alternative methods (such as homogeneity methods) to determine the factors needed for the adjustment of joined precipitation observation segments. The already widely used Alexandersson (1986) and the newly developed Vincent (1998) methods for precipitation are applied to four case studies and the results are compared. Since automation of all climate observations is now a priority, it is hoped that a new procedure could eventually be used for combining human and automated precipitation observations.

2. DATA

The homogeneity techniques were applied on monthly and annual total rain and snow series computed from daily rain and snow extracted from the National Climate Data Archive of Environment Canada. The four cases are described in Table 1 with the following information: name of the two joined segments, year of the join, overlapping period (if available), distance and elevation difference between stations, installation date of the Type B rain gauge and beginning date of the synoptic measurement program. Since there was no overlapping period for Stettler North and Beatrice 2, it was impossible to apply the Simple Ratio Method for these two cases.

Table 1. Information regarding the case study locations

Station 1	Station 2	Join	Overlapping period From	Overlapping period To	Distance Difference [km/m]	Elevation Difference [km/m]	Type B Install.	Synoptic Start
Edmonton	Edmonton Municipal A	1938	1937/10	1943/06	2.2	13	1973/5	1937/10
Stettler	Stettler North	1977	no overlapping		3.9	2	1973/10	No
Beatrice	Beatrice 2	1979	no overlapping		1.3	7	1979/5	No
Bathurst	Bathurst	1965	1964/07	1972/02	7.7	54	1980/8	No

Reference series are used for homogeneity testing. Table 2 presents the surrounding stations included to produce the reference series with the accompanying information: distance from the base, elevation and elevation difference from the base, period of data and correlation on annual rain and snow. Two types of reference series were produced: the average of all surrounding stations (e.g. Ref_Avg) and the average of the highest correlated and longest series (e.g. Ref_R_1236: Reference series for Rain using stations 1, 2, 3, and 6). As indicated, the highest correlated stations could have been different for rain and snow.

*_Corresponding author address:_ Eva Mekis, Climate Research Branch, Meteorological Service of Canada, 4905 Dufferin Street, Downsview, ON, M3H 5T4; e-mail: Eva.Mekis@ec.gc.ca

Before the application of homogeneity techniques, it is important to establish the similarity between the variances of the segments to be joined. The F-test was applied to determine if the annual variances were significantly different from zero. For these four cases, for both rain and snow, the variances of the two segments were not significantly different.

3. METHODS

Three different methods were used to determine the magnitude of adjustment for the joined stations. The description of these methods is briefly presented. The methods were applied on the 12 monthly series and on the annual series separately.

Table 2. Summary of surrounding stations selection.

	STATION NAME	Region	Distance [km]	Elev [m]	Elev. Dif. [m]	Period From	Period To	Annual Corr. Rain	Annual Corr. Snow
	EDMONTON MUNICIPAL A	ALTA		668		1895	2000		
1	CALMAR	ALTA	38	720	52	1914	2000	0.75	0.83
2	THORSBY	ALTA	53	744	76	1932	1968	0.72	0.86
3	SION	ALTA	53	701	33	1906	2000	0.66	0.55
4	WETASKIWIN	ALTA	68	756	88	1902	1975	0.61	0.69
5	RADWAY	ALTA	70	633	35	1922	1951	0.65	0.39
6	CAMROSE	ALTA	76	675	7	1928	1941	0.83	0.44
	Ref_Avg					1902	2000	**0.75**	**0.73**
	Ref_R_1236, Ref_S_124					1906	2000	**0.74**	**0.63**
	STETTLER NORTH	ALTA		816		1918	2000		
1	FORESTBURG PLANT SITE	ALTA	42	663	153	1967	2000	0.74	0.65
2	HUXLEY EAST	ALTA	45	869	53	1970	1980	0.58	0.82
3	SCOLLARD	ALTA	46	853	37	1969	1998	0.56	0.67
4	PINE LAKE	ALTA	56	945	129	1960	1981	0.53	0.49
5	CASTOR	ALTA	60	808	8	1969	2000	0.65	0.53
6	SULLIVAN LAKE	ALTA	61	811	5	1971	2000	0.68	0.56
7	HUXLEY	ALTA	69	914	98	1966	2000	0.70	0.54
8	CAMROSE 2	ALTA	69	738	78	1947	2000	0.67	0.57
9	TROCHU EQUITY	ALTA	69	854	38	1954	1990	0.56	0.32
10	RED DEER	ALTA	75	847	31	1974	2000	0.41	0.56
11	GWYNNE	ALTA	75	768	48	1975	2000	0.31	0.55
12	LACOMBE CDA	ALTA	72	847	31	1908	2000	0.52	0.69
13	CAMROSE	ALTA	78	739	77	1946	2000	0.61	0.31
	Ref_Avg					1907	2000	**0.63**	**0.69**
	Ref_R_157, Ref_S_123-12					1966	2000	**0.87**	**0.72**
	BEATRICE 2	ONT		297		1895	2000		
1	MILFORD BAY	ONT	8	252	45	1965	1984	0.91	0.88
2	MUSKOKA A	ONT	20	282	15	1934	2000	0.82	0.83
3	HUNTSVILLE WPCP	ONT	30	321	24	1960	2000	0.80	0.76
4	DORSET MOE	ONT	38	323	26	1976	2000	0.72	0.82
5	DWIGHT	ONT	48	404	107	1973	2000	0.67	0.71
6	WEST GUILFORD	ONT	56	328	31	1968	1987	0.60	0.88
7	COLDWATER WARMINSTER	ONT	57	285	12	1971	2000	0.43	0.67
8	ORILLIA TS	ONT	57	220	77	1965	2000	0.68	0.71
	Ref_Avg					1934	2000	**0.81**	**0.87**
	Ref_R_12, Ref_S_12346					1934	2000	**0.84**	**0.88**
	BATHURST A	NB		58		1922	1999		
1	NEPISIGUIT FALLS	NB	28	106	48	1922	2000	0.89	0.65
2	LITTLE RIVER MINE	NB	47	341	283	1956	1992	0.84	0.71
3	NEW RICHMOND	QUE	59	47	11	1940	1994	0.55	0.70
4	ST ELZEAR DE BONAVENTURE	QUE	66	229	171	1948	2000	0.63	0.45
5	CHATHAM NBEPC	NB	70	4	54	1960	1980	0.73	0.67
6	PORT DANIEL	QUE	81	69	11	1927	2000	0.62	0.71
7	RENOUS	NB	95	46	12	1953	1978	0.85	0.59
8	ST ALEXIS DE MATAPEDIA	QUE	106	274	216	1927	2000	0.80	0.37
	Ref_Avg					1922	2000	**0.89**	**0.68**
	Ref_R_1257, Ref_S_26					1927	2000	**0.90**	**0.74**

3.1 Simple Ratio Method

This method is used to obtain monthly adjustments from two stations with overlapping period (Thom, 1966). Adjustments are determined independently for rain and snow:

$$x_{1,i,m} = x_{1,i,m} \cdot \left(\sum_{j=1}^{n} x_{2,j,m} / \sum_{j=1}^{n} x_{1,j,m} \right)$$

where
- i — first date of overlapping period
- j — days in overlapping period
- m — month identifier
- x_1, x_2 — meas. at the first and second stations.

A number of constraints were applied for the proper application of this method on precipitation:

- Number of common days (≥ 0 rain or snow values) is computed in each month. If the number of common days is less than 30, the annual ratio is used instead.
- For annual ratio minimum 1 year is required, otherwise the value of the annual ratio is set to 1.0.
- Limit of acceptable ratios are: $0.5 < R < 2.0$. If it is not fulfilled, then the annual ratio value is used instead.
- Trace values are not included in the monthly ratio computation.

3.2 Vincent Method

This method is based on regression models and can be used to identify steps in climatological time series (Vincent, 1998). The models were applied for the date corresponding to the join of the stations (also called "forced step"). A series of ratios q_i between the tested series and the reference series was produced for each month and the year. Outliers - defined by the values outside of the limits mean $\pm$ 3 standard deviations - were flagged and not used in the regression models. The series were further standardized using the 1961-1990 mean and standard deviation:

$$z_i = \frac{q_i - \overline{q}}{s_q}$$

The dependent variable in the regression models is the standardized ratio z_i and the independent variables are the time and an indicator variable describing the step. The indicator variable provides the magnitude of the step at the date of the join.

3.3 Alexandersson Method

This method is based on likelihood ratios and can be used to identify steps in precipitation time series (Alexandersson, 1986). The method was applied on the same standardized ratio series, as described in section 3.2. The year could not be forced to the joining year. The method finds instead the most probable date of the change. The outliers were *not* removed, however the outliers can be identified using the following interval:

$$[\, q_{0.25} - 1.5 \times (q_{0.75} - q_{0.25}),\ q_{0.75} + 1.5 \times (q_{0.75} - q_{0.25}) \,]$$

where $q_{0.25}$ and $q_{0.75}$ are 25% and 75% quantiles.

The magnitude of the step is given by the difference between the average of the ratios before and after the step.

4. RESULTS

The results are presented in Tables 3.1 and 3.2 for rain and snow respectively. Months with less than 10 greater-than-zero events at the tested station are excluded from the results (mainly summer snow and winter rain months) since there are not enough events. The values represent the adjustment factors, which can be used to adjust the first segment. The significant changes at the 0.05 level found around or at the joining year are highlighted in the Tables. No significance test was performed on the Simple Ratio Method. Since the date of change is forced in the Vincent method, the year is not specified. For Alexandersson method, the first year identified is also given. The value is highlighted, if the change identified is close to the joining year. Occasionally with splitting the series at the first inhomogeneity found, further information was obtained about a possible change around the critical joining year. The results of the Alexandersson method were obtained by using the ANCLIM program (Štepánek, 2000).

4.1 Edmonton

For the annual rain series, none of the methods founds any significant change at or around 1938. The closest year found by Alexandersson was in 1944. In July, significant change was found by Vincent-aver with a ratio of 0.83. The Alexandersson test specified 1939 with the ratio 0.7, but the statistic values have not reached the threshold. Studying in details results of Alexandersson, it was found that the T_o are really close to the limit with two peaks around 1937. If the outliers would have been removed from the ratios (Figure 1a), the results could have been similar to the Vincent's results, where the outliers were removed.

For the snow series, there was no significant step identified (Figure 1b). The only significant change was detected in November by Vincent-aver and Vincent-124. Alexandersson method did not identify an even no-significant change close to the date. The difference can be also due to the removal of 7 outliers from the ratio series in the Vincent method.

Table 3.1 Homogeneity test result – Rain

Rain - Methods		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Ann
Edmonton - 1938	Simple Ratio				0.92	1.01	0.91	0.99	0.97	0.88	0.88			0.95
	Vincent-aver				1.08	**0.70**	0.96	**0.83**	0.87	0.97	0.77			0.89
	Vincent-1236				1.14	**0.67**	1.00	0.86	0.88	0.94	0.90			0.92
	Alex-aver					1952 0.6								1944 0.9
	Alex-1236					1945 0.7		1939 0.7						
Stettler - 1977	Simple Ratio	no overlaping period available												
	Vincent-aver			1.24	1.07	1.22	1.15	1.12	1.09	1.24	1.15	1.06		1.10
	Vincent-157			0.97	1.65	1.24	1.15	1.10	1.05	1.27	1.51	**0.79**		1.08
	Alex-aver													1966 1.2
	Alex-157													
Beatrice 2 - 1979	Simple Ratio	no overlaping period available												
	Vincent-aver	2.08	1.79	1.24	**1.26**	1.13	1.04	1.26	0.98	1.04	**1.23**	1.12	**1.49**	**1.15**
	Vincent-12	1.90	1.30	1.20	1.14	1.01	0.98	1.09	1.06	0.98	1.10	1.05	1.47	1.06
	Alex-aver				1985 1.1							1961 0.9		**1984** **1.1**
	Alex-12													
Bathurst - 1965	Simple Ratio	0.99	0.96	0.99	1.06	1.00	1.03	1.00	0.98	1.00	1.01	1.06	1.07	1.01
	Vincent-aver	1.58	1.29	1.05	1.08	0.93	0.89	0.89	1.07	0.77	0.89	**1.14**	**1.37**	0.95
	Vincent-1257	1.18	0.97	1.59	1.61	1.06	1.08	1.01	**1.31**	0.99	1.05	1.24	1.82	**1.04**
	Alex-aver										1991 1.3	**1968** **1.2**		1933 1.2
	Alex-1257					1991 1.3					1986 1.2	1986 1.3		1937 1.1

Table 3.2 Homogeneity test result – Snow

Snow - Methods		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Ann
Edmonton - 1938	Simple Ratio	0.93	0.97	0.90	0.88	0.65				0.75	0.94	0.98	0.77	0.91
	Vincent-aver	1.00	1.19	1.04	1.34	1.54				1.05	1.05	**0.74**	1.17	1.01
	Vincent-124	1.19	1.18	0.99	1.37	2.21				1.22	1.22	**0.71**	1.26	1.00
	Alex-aver													
	Alex-124	1921 0.9												1923 0.7
Stettler - 1977	Simple Ratio	no overlaping period available												
	Vincent-aver	0.89	1.22	0.90	1.21	1.71				**1.27**	**1.59**	1.16	1.13	0.97
	Vincent-123&12	**1.21**	**1.98**	1.34	1.45	1.25				1.41	**1.61**	1.31	1.87	**1.26**
	Alex-aver			1948 0.7							1983 2.4		1923 0.7	
	Alex-123&12	1991 2.5	**1976** **2.00**			**1977** **0.8 N**				1984 2.6	1989 1.4			1984 1.4
Beatrice 2 - 1979	Simple Ratio	no overlaping period available												
	Vincent-aver	1.00	0.95	**0.61**	0.72	0.97					0.97	0.81	**0.93**	**0.91**
	Vincent-12346	1.04	0.96	**0.61**	0.71	0.92					0.96	0.77	0.98	0.94
	Alex-aver			**1957** **0.8**	1964 1.00								1968 0.8	**1978** **0.9**
	Alex-12346			1957 0.8	1964 0.9									
Bathurst - 1965	Simple Ratio	0.99	1.00	0.99	1.01	0.93					0.97	1.00	1.02	1.00
	Vincent-aver	1.00	**1.13**	**1.24**	1.42	1.29					1.28	0.90	1.04	**1.15**
	Vincent-26	**0.98**	**1.06**	**1.13**	1.08	0.94					1.09	0.92	0.96	**1.03**
	Alex-aver		1975 1.3	1943 1.5										**1969** **1.1**
	Alex-26	1944 1.5		**1968** **1.3**										1971 1.2

The results provided by the homogeneity techniques are not in agreement with the SRM results on the annual series, as the latter founds adjustments of 0.95 and 0.91 in annual rain and snow respectively.

In conclusion, no significant change has been found in Edmonton series.

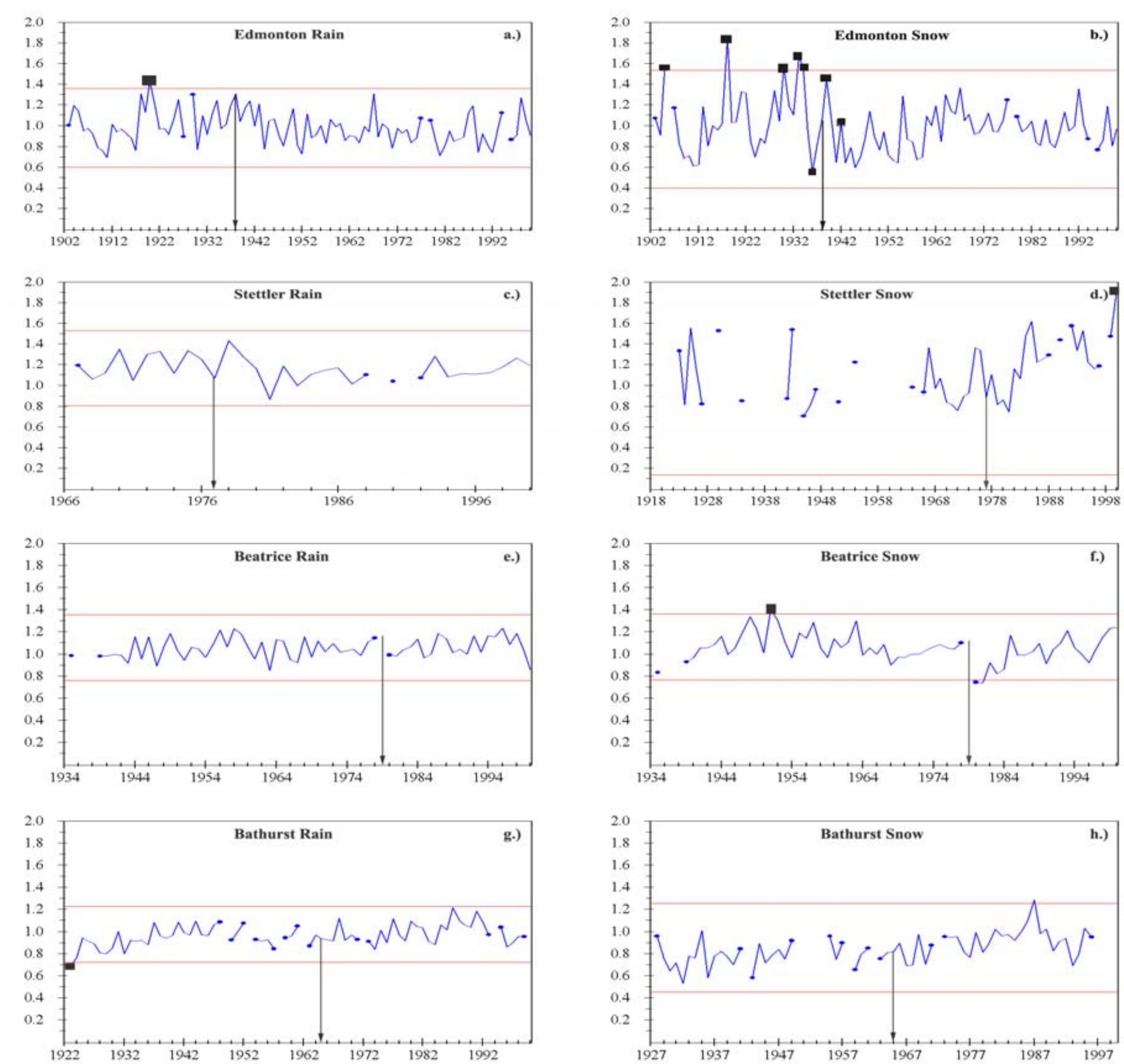
4.2 Stettler North

The results show that the join of the stations in 1977 did not create any significant step in the annual total rain series. The ratios graph shows a very consistent series (Figure 1c). For the monthly rain, no significant steps were detected overall (with the exception of November Vincent-157).

The analysis of annual snow shows no significant step (with the exception of Vincent-123&12). However, the ratios graph shows several missing values (Figure 1d). Some steps were sometimes identified by one or two methods on the monthly series, but never by all of three methods and this provides an indication that the join has not created a step.

The Simple Ratio Method could not be used, since no overlapping period was available.

Figure 1. Ratio plots of annual rain and snow (base divided by highest correlated reference series). The confidence limits are determined by Alexandersson method. The values covered with black squares are removed as an outlier in Vincent's method. The joining year is marked with arrow.



4.3 Beatrice

The joining year 1979 is identified on both rain and snow by both Vincent and Alexandersson methods using the average of the surrounding stations. However, the ratios plots indicate more change for snow than for rain (Figures 1f and e respectively). For rain, the significant change happens in April. Vincent method suggests relatively higher adjustment (1.26) compare to Alexandersson (1.10). In October, Vincent identifies a significant change while Alexandersson did not. Since none of the tests identifies outliers in this month, the results are more comparable. Studying the Alexandersson results further, there is a T_o value in 1982 which is located very close to the significance level. The detection of 1982, instead of 1979, is not a contradiction with the Vincent method since the year was forced in the Vincent test and it is uncertain what would have been the first choice for this method.

For snow, the annual adjustments found with both methods are amazingly close to each other. The distribution of the adjustment is not so close on the monthly basis, the difference could have been again in the fact that on annual basis no outlier was found and removed (e.g. in December, the ratio in 1985 was removed by the Vincent technique).

The results cannot be compared with the Simple Ratio Method, since no overlapping period was available.

4.4 Bathurst

The results show that the join in 1965 has not created a significant step in the annual total rain (Figure 1g), since most methods did not identify significant steps in 1965 (with the exception of Vincent-1257 with a ratio of 1.04). No significant steps were identified in the monthly series by most methods either (Vincent-1257 in August and Vincent-aver in November and December).

However, the join of the station observations has created a significant step in 1965 in the annual total snow identified by Vincent-aver, Vincent-26, Alex-aver, and to a certain extend Alex-26 with a step ratio of 1.15, 1.03, 1.10, 1.20 respectively (Figure 1h). The suggested adjustment factors are higher compared to the value 1.0 by the Simple Ratio Method. On monthly basis, it is clear that the join has the most impact on the first three months of the year.

5. CONCLUSION

The homogeneity tests provide an important tool to determine if the join of two precipitation observation stations creates an artificial step in the time series. They use a completely different and independent source of data, namely the available surrounding station information. The selection of surrounding stations for the reference series is very important and becomes a time consuming part of the study. The Simple Ratio Method also provides another very valuable information from the overlapping measurements. When both surrounding stations and overlapping period are available, homogeneity methods and the Simple Ratio Method should be used together leading to an acceptable answer – namely the necessity for the application of some adjustments.

In Table 1 Type-B rain gauge installation date and the starting date of the synoptic measurement program are also included. Since the homogeneity methods are applied on the raw data (not like SRM method), the homogeneity methods could have identified cases of concurrent events within the same year. For example, in Edmonton the synoptic measurement programs started in 1937/10 and the stations are joined in 1938. For Beatrice, the Type B gauge was installed in 1979/05 and the stations are joined during the same year.

The study of precipitation homogeneity testing has just started and there is still lot to do. The Simple Ratio Method can be improved by computing gradually changing monthly adjustment factors or using further constraints in the computation of adjustment factors. For the homogeneity methods, the same definition for outlier identification should be used and they should be removed in the Alexandersson method as well. It is also important to introduce a minimum number of greater than zero events (if there is not enough greater than zero events, the sample is not representative and the results should not be used).

6. REFERENCES

Alexandersson, H., 1986: A homogeneity test applied to precipitation data. J. Climatology, 6, 661-675.

Mekis, E. and W.D. Hogg, 1999: Rehabilitation and analysis of Canadian daily precipitation time series. Atmosphere-Ocean, 37, 53-85.

Štepánek P., 2000: ANCLIM - Software for homogenization and time series analysis (for Windows 95/NT). Department of Geography, Masaryk University, Brno, http://www.sci.muni.cz/~pest/

Thom, H.C.S., 1966: Some methods of climatological analysis. WMO Technical Note No. 81. 7-9.

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