Chapter 7
Group Comparison, Trends and Cluster Analysis to Understand Historical Precipitation

Raphael Abrahão

Introduction

In recent years, much research has been carried out regarding climate change and its consequences. The current concentrations of greenhouse gases are already a concern and several scientists predict that the average temperature of the planet can increase between 1.8 and 4.0 °C until the end of this century, which may cause dramatic environmental impacts (IPCC 2007; Malhi et al. 2009; Davidson et al. 2012). According to the IPCC (2014), warming of the climate system is unequivocal, and since the 1950s, many of the observed changes are unprecedented over decades to millennia.

This scenario of increasing temperatures is only one facet of climate change. Shifting in precipitation patterns (frequency and intensity) is another important change that needs to be deeper understood. Such changes in climate may impact ecosystems, trigger plagues and epidemics, threaten urban infrastructure, water and energy supply, as well as agriculture, especially in regions where shortage of water is already an issue (Knapp et al. 2002; Disch et al. 2012; Davidson et al. 2012; Durack et al. 2012; IPCC 2014).

Understanding the nature and extent of these impacts is crucial for the determination of adaptation policies towards avoiding or diminishing the negative impacts of climate change, as well as taking advantage of the positive impacts. However, to understand the potential impacts and plan adequately, we first need to understand how climate is changing.

R. Abrahão (✉)
Department of Renewable Energy Engineering (DEER), Center of Alternative and Renewable Energy (CEAR), Federal University of Paraíba (UFPB), Caixa Postal 5115, Cidade Universitária, João Pessoa 58051-970, PB, Brazil
e-mail: raphael@cear.ufpb.br

P. Grammelis (ed.), *Energy, Transportation and Global Warming*, Green Energy and Technology, DOI 10.1007/978-3-319-30127-3_7

Climate models are useful tools to project climate change scenarios. Although nowadays they are very comprehensive and can take into consideration several parameters and relationships, there is still a high level of uncertainty in the studies using these projections, especially for precipitation (Von Storch et al. 1993; Grillakis et al. 2011; Zhang et al. 2011).

In addition, in many parts of the world, relatively long historical climate datasets are available and little (and in many cases, none) use of this information is made towards a better understanding of climate change impacts. The high costs associated with installation and maintenance of these monitoring stations throughout decades could be better spent and deeper information about the dynamics of regional and local climate change could be obtained from the datasets.

A number of studies used historical data, confirming that climate change is not something that will happen in the future but is a current reality (Magnuson et al. 1997; Vincent and Gullett 1999; Alexander et al. 2006; Wang and Ding 2006). These studies are good examples of how deeper information about climate change can be obtained from real historical data. Furthermore, confidence in the projections from climate models is higher for temperature than for other climate elements such as precipitation; confidence is also higher at global and continental scales than for regional and local scales (AMS 2012).

Thus, can historical data provide sufficient information about climate change in order to understand the potential impacts and develop adaptation strategies? And, more importantly, are the methodologies effective to provide this information? The objective of this paper is to assess the use of cluster analysis, together with other more traditional methodologies such as group comparison and trend analysis, to expand knowledge about climate change on local and regional scales from historical data of precipitation.

Methodology

The methodology was based on three different approaches: (1) a general assessment of annual totals with comparison of past and recent periods, (2) trend analysis, and (3) the application of cluster analysis to analyse monthly and seasonal changes. The intention was not compare these approaches, but evidence the importance of using different approaches, instead of a single one, to assess climate changes that will affect the water cycle and others factors regionally.

Monthly rainfall and snowfall data were obtained from the Sudbury airport station (Canada), from January 1956 to December 2010 (55 complete years). Data was obtained from Environment Canada (Canadian National Climate Data and Information Archive, www.climate.weatheroffice.gc.ca) and it was subjected to a process of quality control before becoming available. The city of Sudbury is part of the province of Ontario and is located at 46°29′N and 81°00′W. The climate of the area is classified as humid continental (Dfb according to the Köppen-Geiger classification).

The 20 first years (1956–1975) and 20 last years of the dataset (1991–2010) were selected for evaluation of annual changes. Since the annual data presented normal distribution, the t-test was applied to compare both periods. The t-test performed a comparison of means to determine whether the means of the two periods where significantly different. A probability of less than a 5 % error ($p < 0.05$) was considered.

The Mann-Kendall test was used for trend analysis. A non-parametric test was selected because the monthly and seasonal data did not present normal distribution (Mann 1945; Kendall 1975). The trend estimation was performed applying the Sen's slope calculation (Sen 1968) and error probabilities between 0.1 % and 10 % were used.

Cluster analysis was carried out to analyze changes in the distribution of rainfall and snowfall throughout the year. The purpose was to place the objects (month-year combinations) into groups (or clusters) suggested by the data, not defined a priori. Thus, data was organized monthly, and each month was considered as a different variable, thereby obtaining data from 660 month-year combinations (12 months in 55 years). Years with similar precipitation distribution were placed into the same cluster. The cluster analysis was performed by standardizing the variables and using the Euclidean square distance as similarity measure (Hair et al. 1998). Due to its simplicity and computational efficiency, the Ward method was used to obtain hierarchical clustering (Hair et al. 1998).

In hierarchical cluster techniques, the objects are progressively aggregated until they form a single cluster. Each object begins in a cluster itself and then the closest clusters are merged to form a new cluster that replaces the two previous clusters. Merging of the two closest clusters is repeated until only one cluster is left (Ramos 2001; Munoz-Díaz and Rodrigo 2004). In some studies, similar results were obtained by applying cluster techniques and principal component analysis (Munoz-Díaz and Rodrigo 2004; Yin et al. 2011).

The "median year" of the objects (month-year combinations) was used as an indicator to compare recent and past periods. All statistical procedures were carried out with Statgraphics 15 and Makesens 1.0 softwares.

Results and Discussion

Changes in Annual Rainfall and Snowfall

The dataset suggests that annual rainfall is increasing in the Sudbury site. Even with the expected variability of rainfall between years, an increasing trend during recent years could be observed (Fig. 7.1). Significant statistical differences ($p < 0.05$) were observed when comparing the 1956–1975 period versus the 1991–2010 period, with average values of 608 and 680 mm/year respectively, which represents a 12 % increase in rainfall.

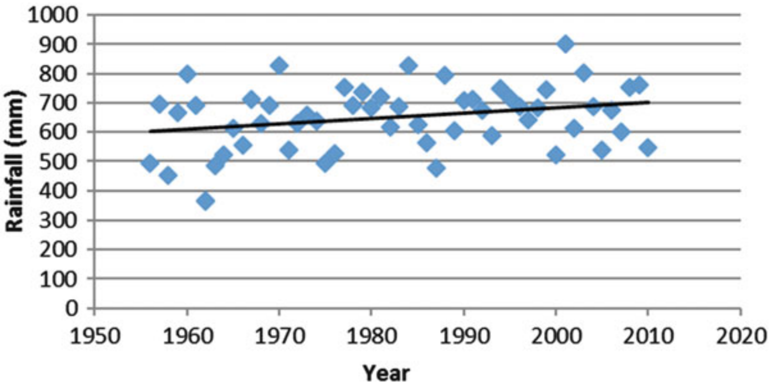

Fig. 7.1 Annual rainfall

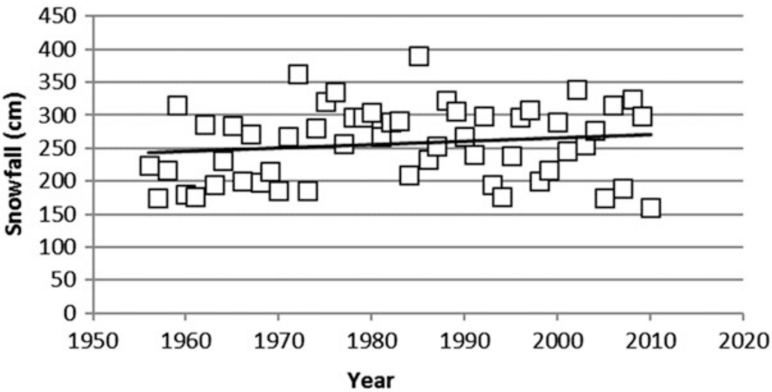

Fig. 7.2 Snowfall in the Sudbury site during the 1956–2010 period

Regarding snowfall, the trend was not that clear. Although a slight increase in snowfall could be observed (Fig. 7.2), there was no statistically significant difference between the means of the two periods ($p < 0.05$).

Changes in Monthly and Seasonal Distribution of Rainfall and Snowfall—Trend Analysis

Trend analysis also detected the annual changes presented in the previous section. A significant increasing trend in rainfall ($p < 0.10$) was quantified by the Sen's slope through a rate of 1.74 mm/year and no significant trend was detected for annual snowfall (Table 7.1).

Table 7.1 Rainfall and snowfall trends detected by the Mann-Kendall test and quantified by the Sen's slope for the Sudbury site during the 1956–2010 period

Period	Rainfall (mm/year)	Snowfall (cm/year)
January	0.05 ns	0.20 ns
February	–	0.17 ns
March	0.28[a]	−0.07 ns
April	0.22 ns	0.07 ns
May	0.58[a]	–
June	−0.13 ns	–
July	0.13 ns	–
August	0.36 ns	–
September	−0.18 ns	–
October	0.49[b]	–
November	0.15 ns	−0.10 ns
December	0.07 ns	0.17 ns
Winter	0.32[a]	0.67 ns
Spring	1.04[b]	0.06 ns
Summer	0.08 ns	–
Fall	0.35 ns	−0.15 ns
Annual	1.74[a]	0.60 ns

ns non significant
[a] $p < 0.10$
[b] $p < 0.05$
[c] $p < 0.01$
[d] $p < 0.001$

The trend analysis could be considered more comprehensive for evaluation of annual changes than the t-test group comparison, because all years of the dataset are included in the analysis and not only the 20 first and 20 last years. However, the expected variability of precipitation could hinder the detection of trends, which was not the case for the Sudbury site during the 1956–2010 period.

Changes in the distribution of precipitation throughout the year were also noted during the studied period. It means that the annual increasing trends observed for rainfall volumes were not uniform over the year. March and May presented significant ($p < 0.10$) positive trends with Sen's slopes of 0.28 and 0.58 mm/year respectively. October presented a more significant ($p < 0.05$) positive trend and a Sen's slope of 0.49 mm/year. The remaining months did not present significant trends.

Seasonally, significant increasing trends were observed in winter ($p < 0.10$) and spring ($p < 0.05$) through Sen's slopes of 0.32 and 1.04 mm/year respectively. The Mann-Kendall test did not detect significant trends in rainfall during summer and fall. Regarding snowfall, no significant trends were detected for any month or season.

Changes in Monthly and Seasonal Distribution of Rainfall and Snowfall—Cluster Analysis

For the monthly rainfall data from the Sudbury site, the cluster analysis applied separated the 55 years into three clusters (Table 7.2). Cluster 1 consisted of years 1956, 1959, 1960, 1961, 1962, 1964, 1970, 1972, 1976, 1978, 1987, 1992, 1997 and 2010. Cluster 2 consisted of years 1957, 1958, 1963, 1965, 1966, 1967, 1968, 1969, 1971, 1973, 1974, 1975, 1977, 1979, 1981, 1982, 1984, 1985, 1986, 1989, 1990, 1991, 1994, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007 and 2009. Finally, cluster 3 included years 1980, 1983, 1988, 1993, 1995, 1996 and 2008.

Each cluster included similar years regarding precipitation distribution over the year. It is important to note that the median year in Table 7.2 does not necessarily have to be included in the years within the cluster, since it is just an indicator for comparison between clusters representative of past and recent periods.

As the interannual variability of rainfall is often high, there were past and recent years within the three clusters. However, cluster 1 presented a greater concentration of earlier years (median year = 1971), while cluster 2 (median year = 1985) and cluster 3 (median year = 1993) gradually corresponded to more recent years.

The separation into three clusters highlighted the temporal increase in annual rainfall, with cluster 1 adding up to 611 mm/year; cluster 2 to 656 mm/year and cluster 3, representing the most recent years, adding up to 701 mm/year. These results of increased rainfall over time had been indicated by the methods applied previously, however, through cluster analysis, changes in the seasonal distribution of rainfall could also be observed, with little interference from the interannual variability of the data. Thereby, the most interesting result from the cluster analysis was the observation of changes in the distribution of rainfall throughout the year. Increases were observed in most months, but at different scales (Table 7.2). Decreases were observed during the months of June, July and September despite the annual increases. Since these months belong to the rainy period of the site (May–October), the shift may have consequences on local water and agricultural management.

Observing these changes through the seasons, it is easier to verify the gradual influence of climate changes on the annual variability of rainfall (Fig. 7.3). Changes were very important during winter and spring, with increases of 309 % and 65 % respectively, when comparing clusters 1 and 3. During the fall, increases were not very important (6 %), while in summer, which is the most important rainy season in Sudbury, the 15 % reduction represented a decrease of 41 mm in rainfall.

For snowfall, cluster analysis separated the 55 years into two clusters. Cluster 1 comprised years 1956, 1957, 1958, 1959, 1960, 1961, 1963, 1964, 1966, 1968, 1969, 1970, 1973, 1977, 1980, 1984, 1986, 1987, 1991, 1993, 1994, 1995, 1998, 2003, 2005, 2007 and 2010, and cluster 2 consisted of years 1962, 1965, 1967, 1971, 1972, 1974, 1975, 1976, 1978, 1979, 1981, 1982, 1983, 1985, 1988, 1989, 1990, 1992, 1996, 1997, 1999, 2000, 2001, 2002, 2004, 2006, 2008 and 2009. The median years of the two clusters were close (1977 and 1987), implying in that the

Table 7.2 Three clusters discriminated by the cluster analysis based on monthly rainfall data in Sudbury site from 1956 to 2010

Cluster	Median Year	January (mm)	February (mm)	March (mm)	April (mm)	May (mm)	June (mm)	July (mm)	August (mm)	September (mm)	October (mm)	November (mm)	December (mm)	Sum (mm)
1	1971	6	0	11	37	67	86	86	95	129	51	35	7	611
2	1985	7	9	30	47	72	86	79	75	94	81	57	19	656
3	1993	36	5	25	69	97	50	70	106	104	82	41	16	701

Median year and annual sum correspondent of each cluster are also provided

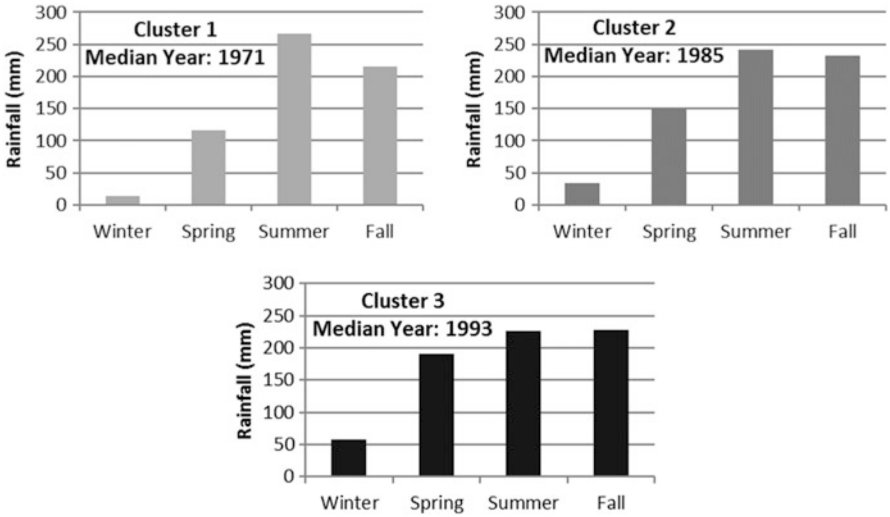

Fig. 7.3 Trends of changes in rainfall distribution throughout the seasons in the Sudbury site from 1956 to 2010, represented by the three clusters discriminated by cluster analysis

trends of change in snowfall distribution over the year were not as clear as for rainfall.

Nevertheless, changes in snowfall distribution could be observed for December, January and February, the 3 months of higher snowfall values in Sudbury. Although the annual sum of snowfall did not change significantly, more recent years presented lower values in December (13 %), and higher values in January (38 %) and especially in February (116 %), when values more than doubled (Fig. 7.4).

The application of cluster and trend analysis evidenced that the increasing annual trends for rainfall volumes in the Sudbury area were not uniform over the year. This annual increase was mainly in winter and spring. Decreases in summer rainfall were detected only through cluster analysis. According to cluster analysis, although years are currently more humid, summers are becoming drier in the location. Previous studies have detected and/or projected changes in rainfall in Sudbury and nearby sites (Magnuson et al. 1997; OCCIAR 2010; IEESC 2012; Charron 2014), however, the changes in seasonality were not considered in depth.

Regarding snow precipitation, while group comparison and trend analysis did not indicate any significant trends, cluster analysis showed clear changes for the months of greatest snowfall (December, January and February). Reductions in December and increases in January and February were observed. This means that the climate of the site is changing towards later winters regarding snowfall, which may also be related to changes in other variables (e.g., temperature).

Although it was not the focus of this study, changes in extreme events should also be investigated. According to Folland et al. (2002), in regions where annual rainfall has increased it is very likely that there have been even more pronounced increases in heavy and extreme rainfall events. And the converse is also true.

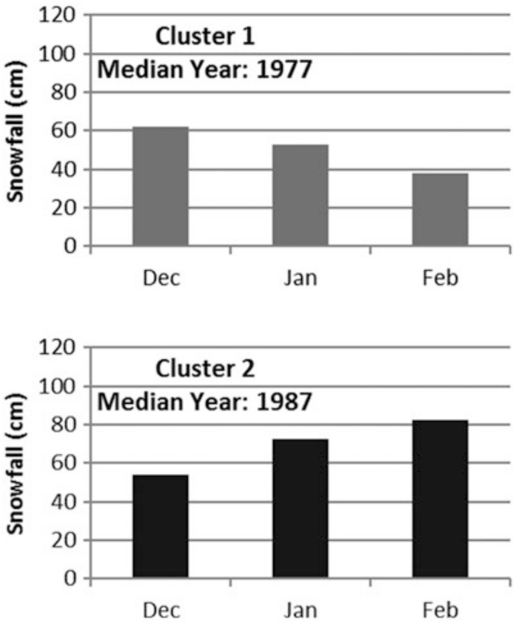

Fig. 7.4 Trends of changes in snowfall distribution throughout December, January and February in the Sudbury site from 1956 to 2010, represented by the two clusters discriminated by cluster analysis

In some regions, heavy and extreme precipitation events have increased, despite the fact that total precipitation has decreased or remained constant.

Conclusions

The methodologies used in this study demonstrated that the detection of annual changes is only the initial step in a more comprehensive understanding of climate change, which also includes complex seasonal and monthly changes. In many areas of the world, the absence of detected changes or significant trends from annual data may give a false idea of the absence of climate change in the location (e.g., snowfall in the site of this study). However, the results presented herein indicate that the inclusion of simple methods, such as cluster analysis, can contribute to a better understanding of seasonal and monthly climate changes.

Acknowledgement This work was supported by the Ontario Centre for Climate Impacts and Adaptation Resources (OCCIAR) and Environment Canada.

References

Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein, A. M. G., et al. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research: Atmospheres, 111*, 1–22.

AMS, American Meteorological Society. (2012). *Climate change: An information statement of the American Meteorological Society*. Adopted by AMS Council in 20 August 2012, Boston.

Charron, I. (2014). *A guidebook on climate scenarios: Using climate information to guide adaptation research and decisions*. Montreal: Ouranos.

Davidson, E. A., de Araújo, A. C., Artaxo, P., Balch, J. K., Brown, I. F., Bustamante, M. C., et al. (2012). The Amazon basin in transition. *Nature, 481*(7381), 321–328.

Disch, J., Kay, P., & Mortsch, L. (2012). A resiliency assessment of Ontario's low-water response mechanism: Implications for addressing management of low-water under potential future climate change. *Canadian Water Resources Journal, 37*(2), 105–123.

Durack, P. J., Wijffels, S. E., & Matear, R. J. (2012). Ocean salinities reveal strong global water cycle intensification during 1950 to 2000. *Science, 336*(6080), 455–458.

Folland, C. K., Karl, T. R., & Salinger, M. J. (2002). Observed climate variability and change. *Weather, 57*(8), 269–278.

Grillakis, M. G., Koutroulis, A. G., & Tsanis, I. K. (2011). Climate change impact on the hydrology of Spencer Creek watershed in Southern Ontario, Canada. *Journal of Hydrology, 409*(1–2), 1–19.

Hair, J., Anderson, R., Tatham, R., & Black, W. (1998). *Multivariate data analysis*. Englewood Cliffs, NJ: Prentice-Hall.

IEESC, Institute for Energy, Environment and Sustainable Communities. (2012). *Producing high-resolution (25 km × 25 km) probabilistic climate change projections over Ontario using UK PRECIS*. Regina: University of Regina.

IPCC, Intergovernmental Panel on Climate Change. (2007). *Climate change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge/New York: Cambridge University Press.

IPCC, Intergovernmental Panel on Climate Change. (2014). Summary for policymakers. In *Climate change 2014: Impacts, adaptation, and vulnerability*. Cambridge/New York: Cambridge University Press.

Kendall, M. G. (1975). *Rank correlation methods*. London: Griffin.

Knapp, A. K., Fay, P. A., Blair, J. M., Collins, S. L., Smith, M. D., Carlisle, J. D., et al. (2002). Rainfall variability, carbon cycling, and plant species diversity in a mesic grassland. *Science, 298*(5601), 2202–2205.

Magnuson, J. J., Webster, K. E., Assel, R. A., Bowser, C. J., Dillon, P. J., Eaton, J. G., et al. (1997). Potential effects of climate changes on aquatic systems: Laurentian Great lakes and Precambrian shield region. *Hydrological Processes, 11*(8), 825–871.

Malhi, Y., Aragao, L. E., Galbraith, D., Huntingford, C., Fisher, R., Zelazowski, P., et al. (2009). Exploring the likelihood and mechanism of a climate-change-induced dieback of the Amazon rainforest. *Proceedings of the National Academy of Sciences of the United States of America, 106*(49), 20610–20615.

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica, 13*, 245–259.

Munoz-Díaz, D., & Rodrigo, F. S. (2004). Spatio-temporal patterns of seasonal rainfall in Spain (1912–2000) using cluster and principal component analysis: Comparison. *Annales Geophysicae, 22*(5), 1435–1448.

OCCIAR, Ontario Centre for Climate Impacts and Adaptation Resources. (2010). *Climate change and conservation authorities in northern Ontario*. Workshop report. OCCIAR, Sudbury.

Ramos, M. C. (2001). Divisive and hierarchical clustering techniques to analyse variability of rainfall distribution patterns in a Mediterranean region. *Atmospheric Research, 57*(2), 123–138.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association, 63*, 1379–1389.

Vincent, L. A., & Gullett, D. W. (1999). Canadian historical and homogeneous temperature datasets for climate change analyses. *International Journal of Climatology, 19*(1), 1375–1388.

Von Storch, H., Zorita, E., & Cubasch, U. (1993). Downscaling of global climate change estimates to regional scales: An application to Iberian rainfall in wintertime. *Journal of Climate, 6*, 1161–1171.

Wang, B., & Ding, Q. (2006). Changes in global monsoon precipitation over the past 56 years. *Geophysical Research Letters, 33*, 1–4.

Yin, H., Gao, Y., & Fan, C. (2011). Distribution, sources and ecological risk assessment of heavy metals in surface sediments from Lake Taihu, China. *Environmental Research Letters, 6*(4), 1–11.

Zhang, H., Huang, G. H., Wang, D., & Zhang, X. (2011). Uncertainty assessment of climate change impacts on the hydrology of small prairie wetlands. *Journal of Hydrology, 396*(1–2), 94–103.