University of British Columbia
Department of Statistics
Technical Report #245
May 2009

An Analysis of Alberta's climate. Part I: Non-homogenized data.

by

Reza Hosseini[1], Nhu D Le[2] and James V Zidek[1]
[1] University of British Columbia
[2] British Columbia Cancer Agency & University of British Columbia

---

Deaprtment of Statistics, University of British Columbia

ABSTRACT

This report presents the first of two exploratory analyses of climate data for the Province of Alberta, specifically precipitation and temperature with the ultimate goal of developing new stochastic models for the processes involved. The data addressed in this report, Part I, has not been homogenized, that is enhanced to overcome the anomalies due to such things as changes in instrumentation and recording errors. However, on the positive side it comes from a very large number of monitoring sites that cover the Province extensively. The report uses standard tools for exploratory data analysis and presents findings about such things as extreme values, the temporal aggregation of climate data series, spatial as well as temporal trends.

1 Introduction

Recent years have seen an emerging worldwide concern about the state of the climate and undesirable trends that imply increasing environmental risk. Managing that risk is now seen as essential for health and welfare. Agriculture in particular contributes greatly to the latter and hence, ultimately to the former.

However, the great complexity of the processes confronted in environmental risk assessment and management along with the technological improvements in data capture, now imply the need for sophisticated new stochastic models for extracting and exploiting the information in those data for climate variables (such as precipitation and temperature). Yet the needed theory for space - time processes is only now being developed, despite the generally mature state of stochastic processes and spatial statistics. Thus we are led to develop the new theory for such processes reported in subsequent reports.

That work begins in this report, with a preliminary analysis of climate data obtained for the Province of Alberta, not because our ultimate focus is Alberta, but because that data provide a good base for out investigation, covering a long time period and a wide range of climatic regimes. That base will guide the choices made in developing that theory, in particular for temperature and precipitation (PCPN). Our specific motivation comes from agriculture largely because this work arose out of collaboration with scientists in that domain. There the environmental risks involve extreme events such as flooding and drought. That in turn leads us to look at extremes in the measurable responses representing the relevant environmental space–time processes.

This report using standard tools in exploratory data analysis such as graphical displays to investigate the climate variable. We leave inference to future reports a after suitable models have been developed. In fact the purpose of this report is to find reasonable assumptions on which to base such models. We believe that the analysis will also assist investigators working in other contexts. Some examples of investigations of climate variables in Canada and in particular Alberta are given in the bibliography. We have also included selectively previous work done on the modelling such processes where the work is deemed of relevance to our own.

However, as an important caveat, the dataset addressed in this report here has not been enhanced to remove anomalies such as those due to a change in instrumentation or more simply, a recording error. Such enhanced data are the subject of Part II in our series. However, of necessity, the latter involves a very small number of monitoring stations and hence, does not represent the spatial domain nearly as well as the data considered here in Part I.

To conclude, Section 2 describes the dataset. Section 3 is devoted to the exploratory analysis of daily, monthly and annual as well as extreme values. The long term trends, daily values distributions, spatial-temporal correlation and extreme value distributions are the main topics of this section.

2 The Data

The data used in this report were downloaded from the environment Canada website (http://www.weatheroffice.ec.gc.ca/) on two CDs giving daily temperature, precipitation, and snow on the ground measurements for various geographical locations in Canada. The National Climate Data and Information Archive, operated and maintained by Environment Canada, contains "official climate and weather observations for Canada" (quoting from the website). One of the CDs contains the data for eastern Canada (hereafter "2002East"), the other for western Canada (hereafter "2002West"). 2002East (108 MB ZIP) contains data for 6,774 locations in Ontario, Quebec, the Atlantic Provinces, Yukon Territory, Nunavut and the Northwest Territories while 2002West (106 MB ZIP) contains the data for 4,442 locations in British Columbia, Alberta, Saskatchewan, Manitoba, Yukon Territory, and the Northwest Territories.

The website includes amongst other things information about the CDs and a glossary of terms in climate literature including "precipitation" in particular:

> *Precipitation:* The sum of the total rainfall and the water equivalent of the total snowfall observed during the day.

We will use 2002West since it includes Alberta, the focus of our interest which arose out of collaboration with agricultural scientists (although this is purely incidental to our eventual purpose). The data is stored in a binary format in several files. Therefore the CD includes packages and their manuals to assist the user: "cdcd"; "cdex". cdcd helps the user visualize the data while "cdex" extracts it. However, "cdex" can only extract the data for one climate monitoring station at a time in a specific formats that are not always convenient for use in R (the well known statistical package used in our analysis) and other statistical softwares. These formats omit longitude, latitude and elevation. Hence, to get the data in a desired form, we need to read the binary files using another program written in Python by Bernhard Reiter and available online at:
http://www.intevation.de/∼bernhard/archiv/uwm/canadian_climate_cdformat/.
However, for unknown reasons this code fails to get the data for a lot of stations leading us to modify it significantly. The modified code is available at:
http://bayes.stat.ubc.ca/∼reza/Python%20code/.
After getting the data, we used our program to put the data into our desired formats. Moreover we included many new Python functions for other extraction purposes (Hosseini 2009).

A total of 1264 of the stations represented in 2002West lie in Alberta and they record: maximum temperature, minimum temperature, one-day rainfall, one-day snowfall, one-day precipitation, and snow depth at both a daily and monthly level. For each station, the data are aggregated and given for various time intervals. The earliest year for which data are reported in Alberta is 1877, the latest 2002. Of

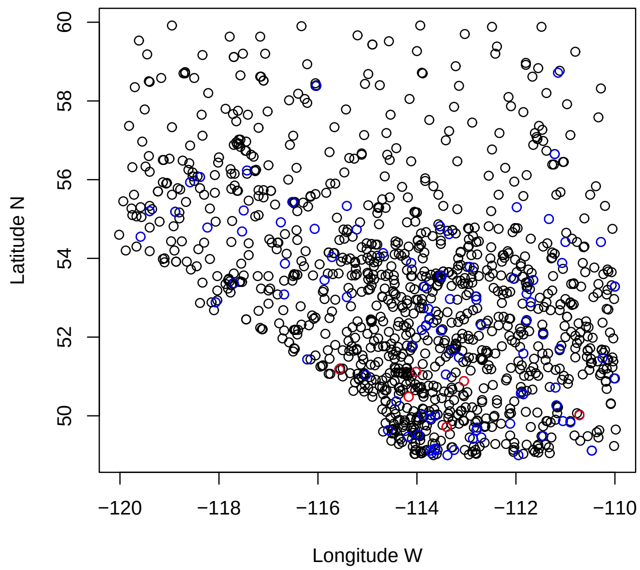

Figure 1: Location of the weather stations in Alberta. The red shows the stations with over 100 years of data and the blue the stations with more than 50 years.

the 1264 stations in Alberta, 1253 include precipitation (PCPN) data. The station with the longest PCPN data record includes 123 years of data. The average of number of years with PCPN data for Alberta is 20 years.

In all 96 stations have recorded more than 50 years of data, 34 more than 75 years and only 6 stations, more than 100.

Figure 1 shows the location of the available stations over Alberta. Figure 2 is the same plot, but with the number of the available years plotted as the height. In Figure 1, the locations with more than 100 years of PCPN data are shown in red and the stations with more than 50 years in blue. It can be seen in the plot that although stations over 100 years are centered at one corner, stations with over 50 years are rather spread out over the province.

The following stations have recorded more than 100 years of data in Alberta:

- Calgary Airport
- Fort MacLeod
- Gleichen
- High River
- Medicine Hat

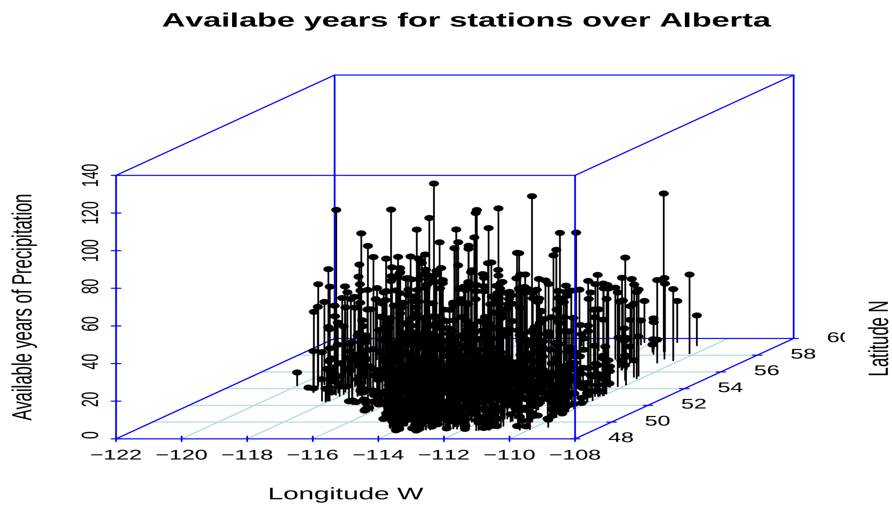

Figure 2: The number of available PCPN years for Alberta is represented by the height of the "skyscrapers" in this plot.

- Banff

3 Data analysis

For our analysis, we mostly use stations with a long time series of data, including the six noted above with more than 100 years of record. Two of these stations are the Calgary and Banff sites. We have singled out these stations in some parts of our analysis as representing two quite different physical environments within the Province. However for analyzing spatial patterns including spatial correlation more stations are needed. We picked the (102) stations which have complete data records from 1960 to 2000 for both daily PCPN and Maximum daily temperature and used all their data.

The daily maximum temperature (MT) and minimum temperature (mt) are given in degrees Celsius. Missing data are denoted by 'NA'.

The daily PCPN data has three possible values, Numeric (millimeters), NA and Trace, the latter meaning a day with $0 < \text{PCPN} < 0.2$. We investigated the percentage of days with 'Trace' amounts of precipitation notably for the Calgary site and found the latter had 11.15% of such days (excluding NA). To see if 0.2 is a significant amount compared to a wet day (i.e. a day with a recorded PCPN), we looked at the 5-number summary for wet days at two stations, Calgary's airport and Gleichen. The results are given in Table 1.

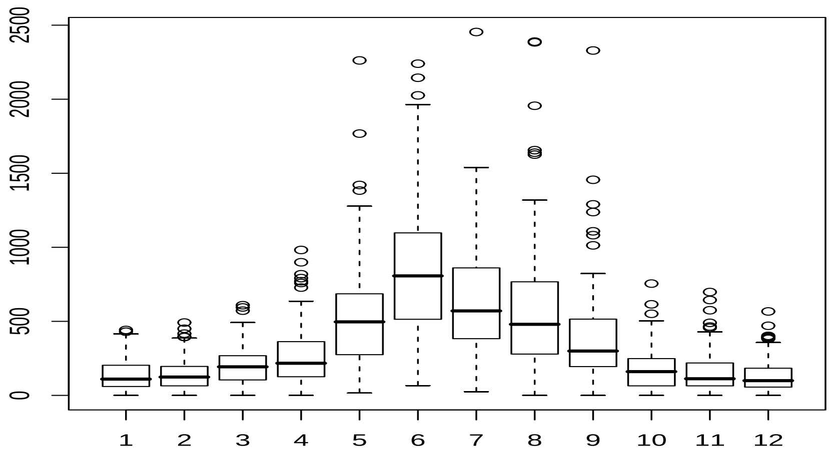

Figure 3: Calgary's airport monthly average precipitation over all the years.

Station	Min	1st Quartile	Median	Mean	Third Quartile	Max
Calgary airport	0.200	0.800	1.800	4.226	5.100	95.300
Gleichen	0.200	1.300	2.500	5.154	6.100	91.400

Table 1: The precipitation (mm) summary for wet days of two stations

That table shows the mean PCPN for a wet day to be more than 4 mm, a very large amount compared to 0.2 mm. Even the median is about 10 times larger than 0.2. On that basis 'Trace' amounts do indeed seem negligible. Thus in a lot of our modeling applications, we have for technical simplicity put the threshold at 0.2 instead of zero.

To get an idea of the seasonal patterns for the Calgary station, we computed the average monthly PCPN over the years. Table 2 displays the results.

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
334	442	501	564	926	1120	844	754	590	377	359	336

Table 2: Monthly average precipitation (mm)

Observe that overall the month of June has highest precipitation level of all the months. The box plots in Figure 3 demonstrate this result even more conclusively and also show it to be the most variable of all the monthly averages over the years.

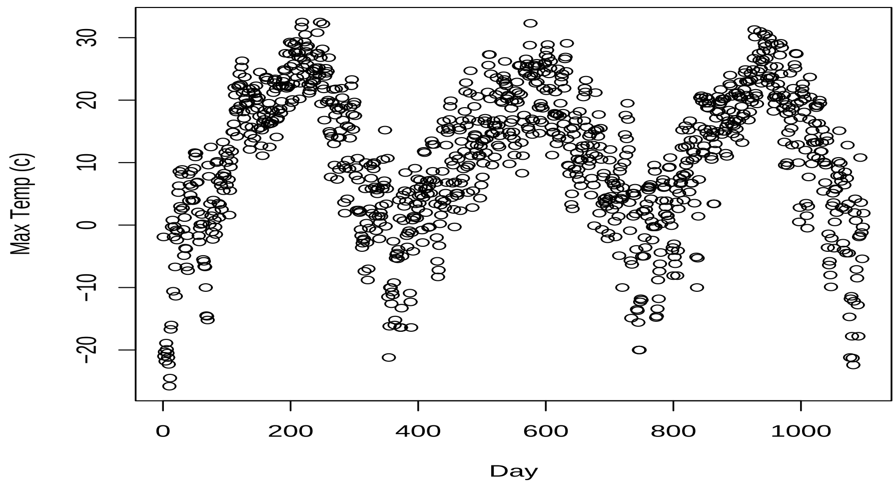

Figure 4: Calgary's maximum daily temperature (deg C), 1998-2000.

To gain further insight into the variation seen in the data we plot the daily time series for PCPN and MT for Calgary for the years 1998-2000 (Figures 4 and 5). The seasonal pattern are seen to be stronger for MT. At the same time the record for PCPN exhibits a lot of zeros, as indicated by the thick band along the x-axis. This points to the well–known need to model these two kinds of series very differently–precipitation values come from two distinctly different populations one consisting entirely of zeros.

Next, we plot the monthly means for 1990-2000 in Figures 6–8. The plots show the seasonal patterns for PCPN a lot better than the daily values. Finally, to see the annual variability in precipitation, we plot yearly averages (Figures 8 and 9). To find an appropriate time filter for seeing long term patterns, we have computed the temporal mean successively for intervals of length 1, 2, 5 and 10 years at the Calgary site. These figures show the decadal mean to be the best such filter–the time trend is now seen clearly.

3.1 Precipitation, maximum and minimum temperatures

In a spatial-temporal context like the one addressed by this report, it is not reasonable to assume that responses at different temporal and spatial points are identically distributed. Thus for example, we need to parameterize them accordingly so that for example we denote the maximum temperature for Calgary, January 1st, 2000 by

$$MT(s, d, y),$$

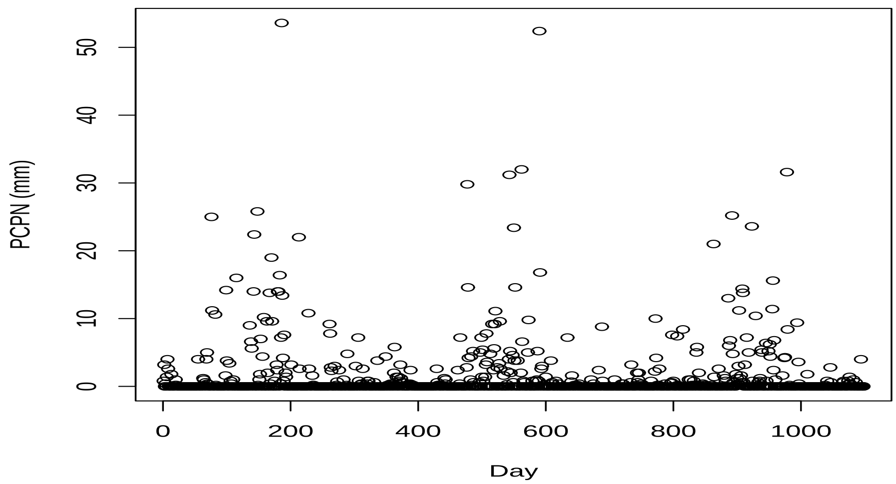

Figure 5: Calgary's daily precipitation totals, 1998-2000. The heavy black band along the x–axis represents zeros in the record.

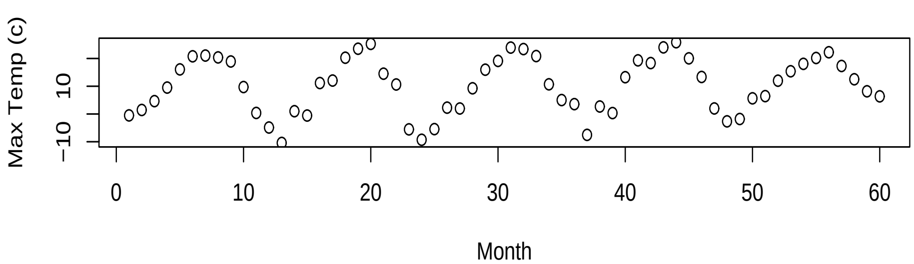

Figure 6: Monthly means of Calgary's maximum temperature (deg C), 1995-2000. Notice the strong seasonality in the series represented here.

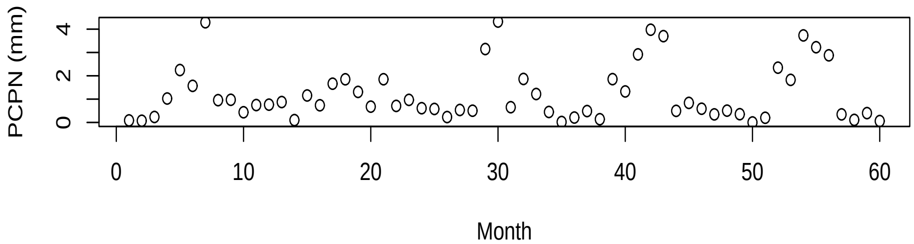

Figure 7: Monthly mean precipitation totals for Calgary, 1995-2000. These also exhibit a well–defined seasonal pattern but one that is less clearly defined than that for temperature even though these two variables are related.

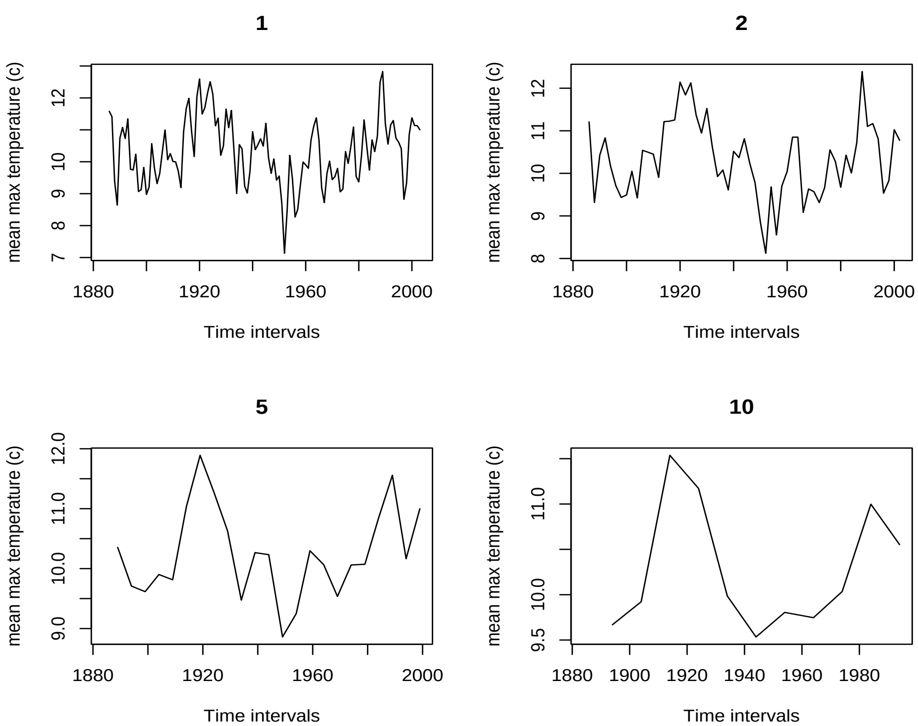

Figure 8: Calgary's mean maximum temperature (deg C) based on temporal aggregation intervals of 1,2,5 and 10 years. The 10 year filter seems to have removed the short term variation and displayed well the long term trend in the series. Notice the two peaks in this series revealing very warm conditions around 1915 and again in the 1990's.

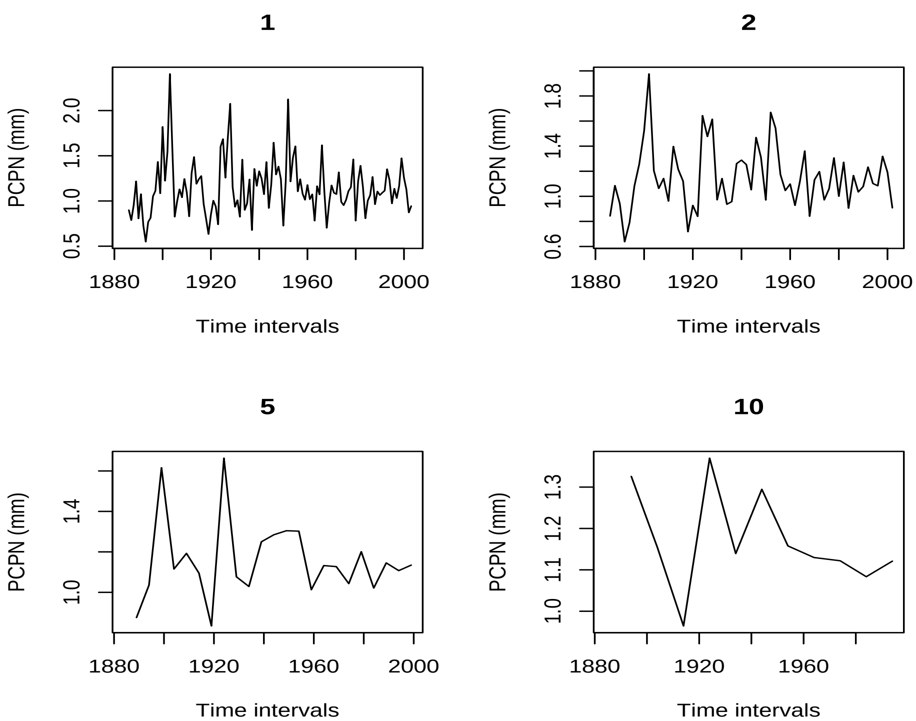

Figure 9: Like the analysis for the previous filter for Calgary's PCPN (mm), this figure shows the effects of filtering the precipitation series using means for time intervals of 1,2,5 and 10 years. Once again the decadel mean seems to be needed as a lo-pass filter to reveal long–term trends.

where, s=Calgary, d=01 and y=2000. Yet without further assumptions, would have at most a single observed response value at each space – time point, hardly enough to characterize its probability distribution.

As a tentative hypothesis we could assume equality in the daily response distributions from one year to the next. For example, January 1 on successive years, 2000, 2001, $\cdots$ yields independent and identically distributed observations of the response (random variable) $X(s, d)$ for $d = 1$. In fact we might be inclined to adopt an even stronger assumption since the autocorrelation in the precipitation series remains significant for lags of only a few days while that for maximum temperature remains significant for lags of less than 30 days (see Figures 20 and 22). Undoubtedly that working hypothesis is too restrictive due to climate change, and we need to allow for change. Moreover, introducing a year effect would allow us separately model the within–year and between–year responses. For instance, we could model the year effect by the annual mean and the within year distribution by removing that effect through subtraction, division or both, the random response variable at the fixed location s and day d thus becoming:

$$Y(s, d, y) = X(s, d, y) - \sum_{i=1}^{365} X(s, i, y)/365$$

or

$$Y(s, d, y) = \frac{X(s, d, y)}{\sum_{i=1}^{365} X(s, i, y)/365}.$$

However to keep things simple for our exploratory analysis we adopt the tentative hypothesis above as a working assumption. With that starting point we plotted histograms of observed yearly responses for successive days $d = 1, \ldots, 366$ and locations s. We then fitted Gaussian distributions to maximum daily temperature (MT; Figure 10) and Gamma distributions to daily precipitation (PCPN) above threshold (Figure 12) in accord with common practice. The qq-plots for MT and PCPN in Figures 11 and 14, respectively, show the fits of the assumed distributions to be reasonably good.

3.2 Spatial correlation

This subsection explores spatial correlation patterns for both daily precipitation (PCPN) and the daily maximum temperature (MT) against the geodesic intersite distances between the monitoring sites (i.e. stations) for a fixed day of the year. Figures 15 and 16 show that correlation. However, plotting the dark cloud of all the available points would obscure that pattern. Thus each figure depicts just 500 points randomly picked from among the large number available. The plots show a decreasing trend in spatial correlation with respect to the distance as one might intuitively expect. However they also show that distance alone fails to capture all of the spatial structure.

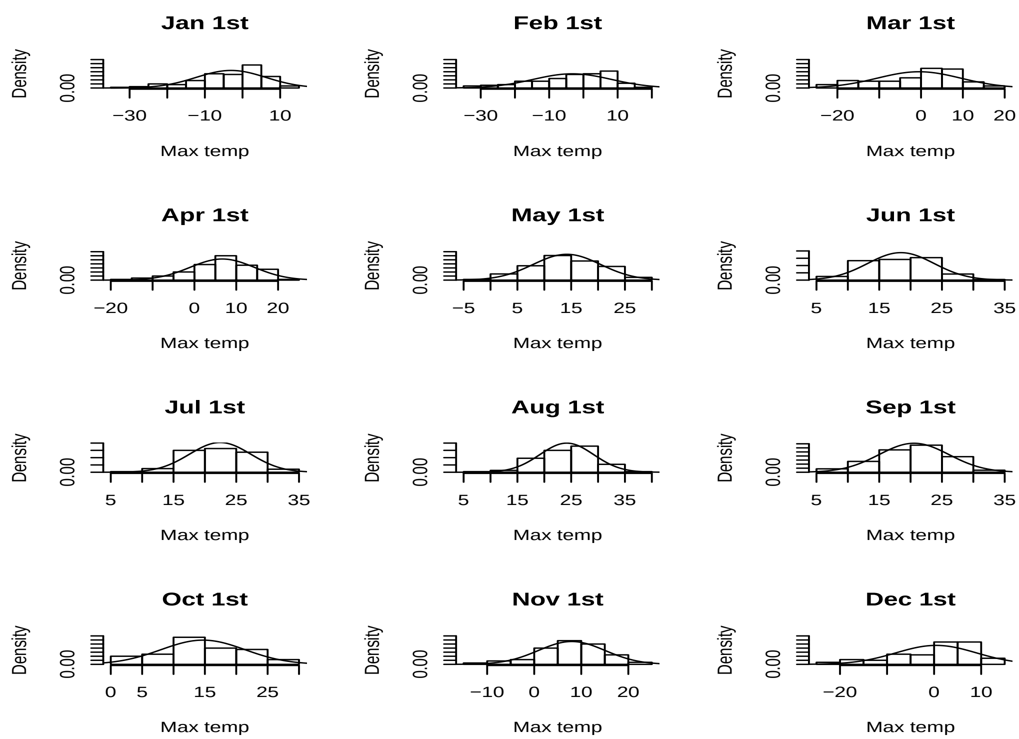

Figure 10: Histograms for Calgary's maximum daily temperature (C) for the first day of each month, based on data collected over the years. Also seen are the best Gaussian fits to these histograms.

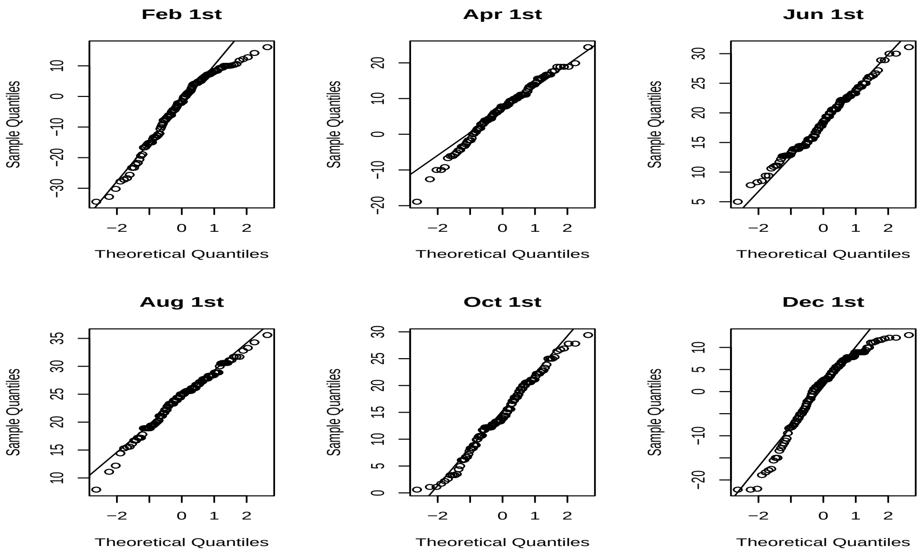

Figure 11: Normal qq-plots for Calgary's maximum daily temperature (C) data for the first day of 6 selected months. Generally the Gaussian approximation seems satisfactory for these data.

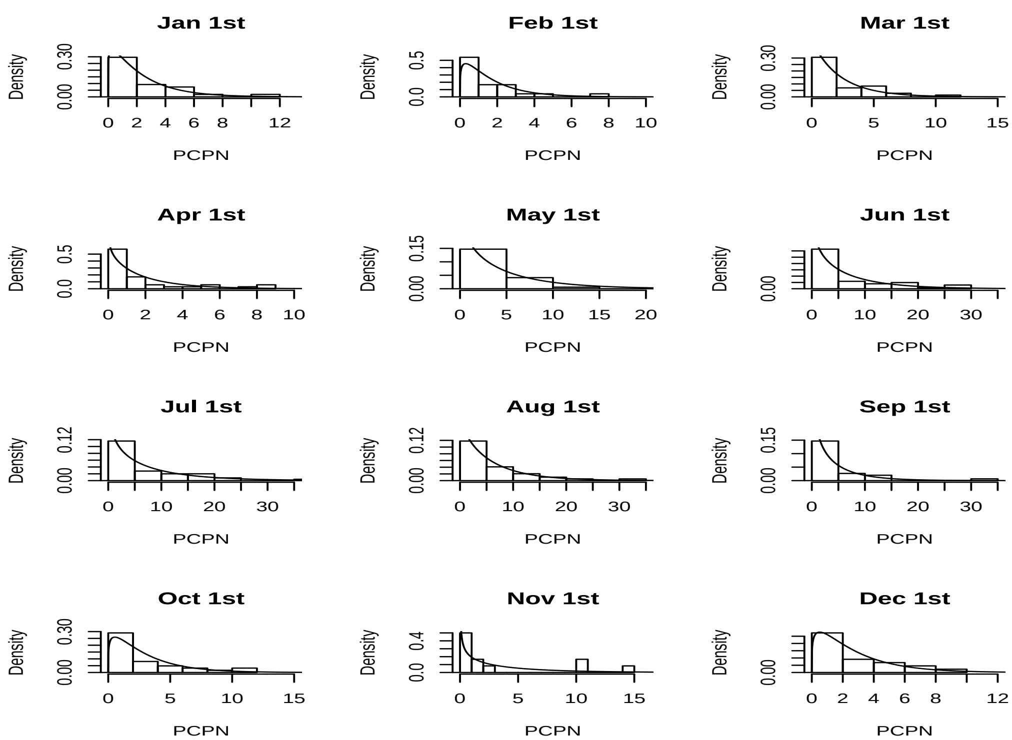

Figure 12: Gamma distribution fits to the histograms of Calgary's PCPN (mm) data for the first day of each month after deleting the zeros.

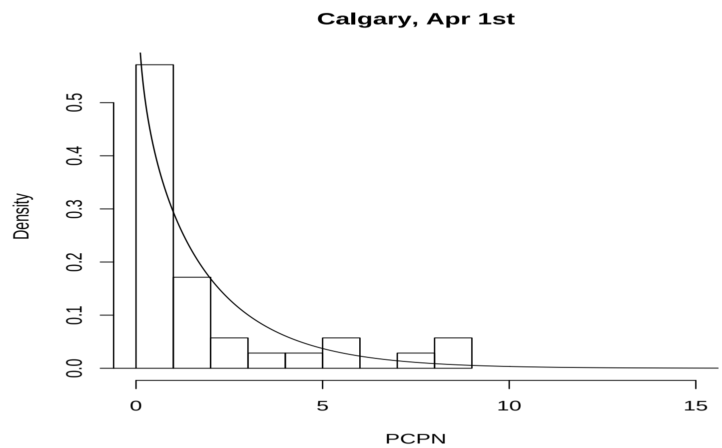

Figure 13: The Gamma distribution fitted to the Calgary's total precipitation (PCPN in mm) for the first day of April after removing the zeros. A good approximation is seen.

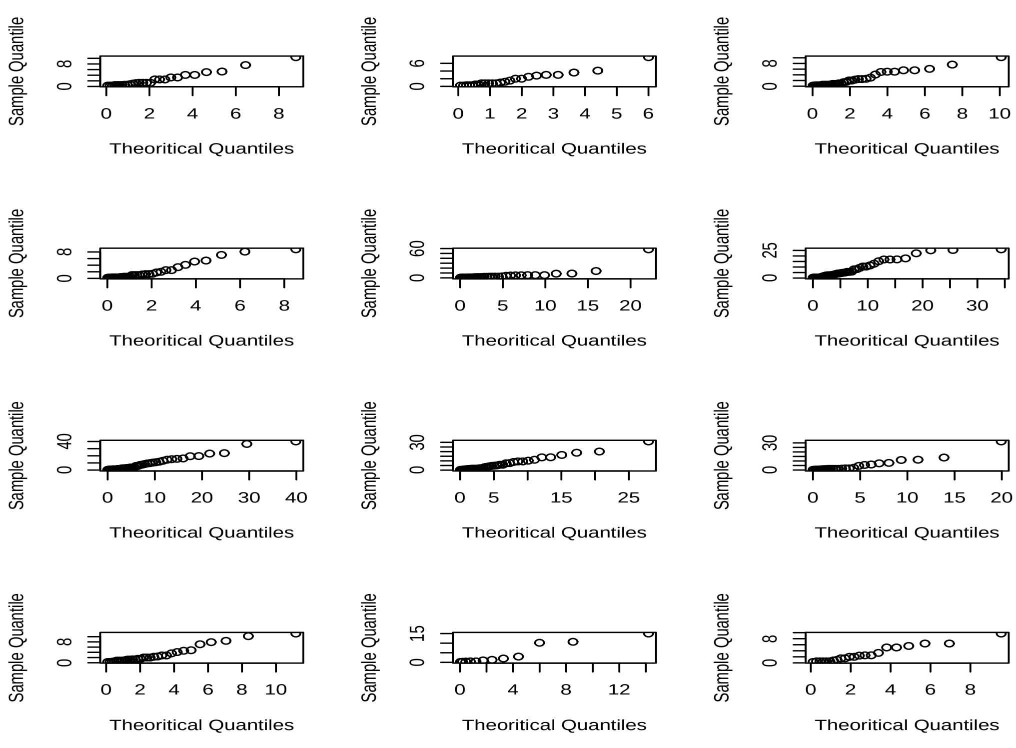

Figure 14: The Gamma distribution's qq-plots for Calgary's daily precipitation (PCPN) data over the years broken down by month. Overall the approximation seems satisfactory.

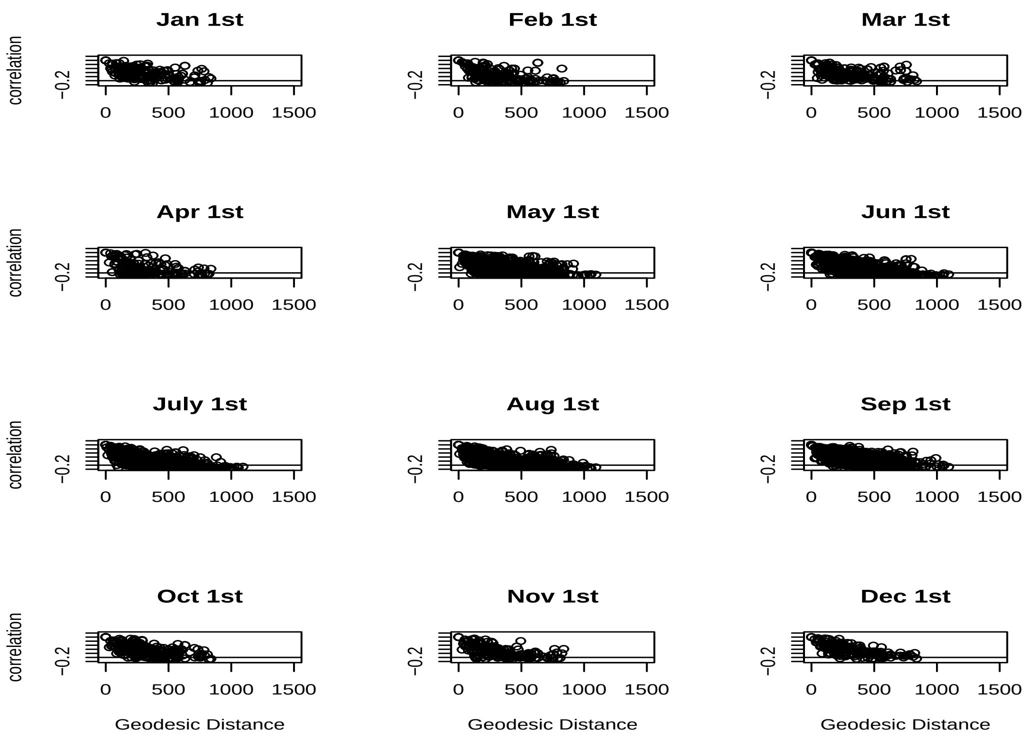

Figure 15: This figure displays for selected days and stations, the estimated spatial correlation for daily precipitation (mm) plotted as a function of intersite Geodesic distance (km). The data come from the years 1960 to 2000. Note the monotone decreasing trend with some evidence of anisotropy, perhaps due to the effect of elevation.

In contrast. MT's plot shows a persistently high correlation over the entire spatial domain, again in agreement with intuition. The line $y = 0$ helps to show that MT's spatial correlation is positive over the entire spatial domain despite the effects of sampling error. In contrast spatial correlation for PCPN turns out to be negative although of small magnitude and hence likely the result of sampling error. (Note: We for each fixed day d, we have computed the intersite spatial correlations only between stations that share at least 40 common observations over the years.)

3.3 Correlation patterns over time

This subsection looks at correlation trends over days of the year for selected pairs of monitoring sites. First we compute those correlations for fixed dates Jan 1st, Jan 2nd, ... for both PCPN and MT (Figures 17 and 18). These figures show the estimated correlation for Calgary paired with each of 5 other locations. PCPN shows a rather irregular (non-constant) trend, possibly due to the small number of

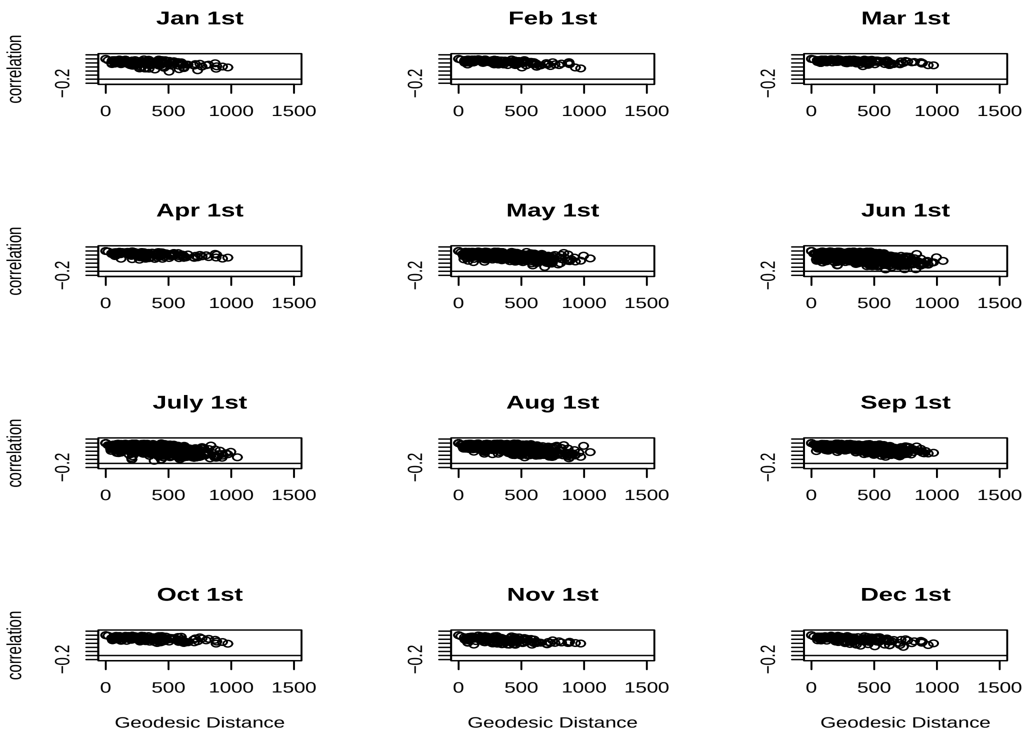

Figure 16: The estimated spatial correlation of maximum daily temperature (C) for selected days as well as well as stations, plotted against Geodesic intersite distance (km) and computed from data from the years 1960 to 2000. Note the persistence of the correlation over the spatial domain.

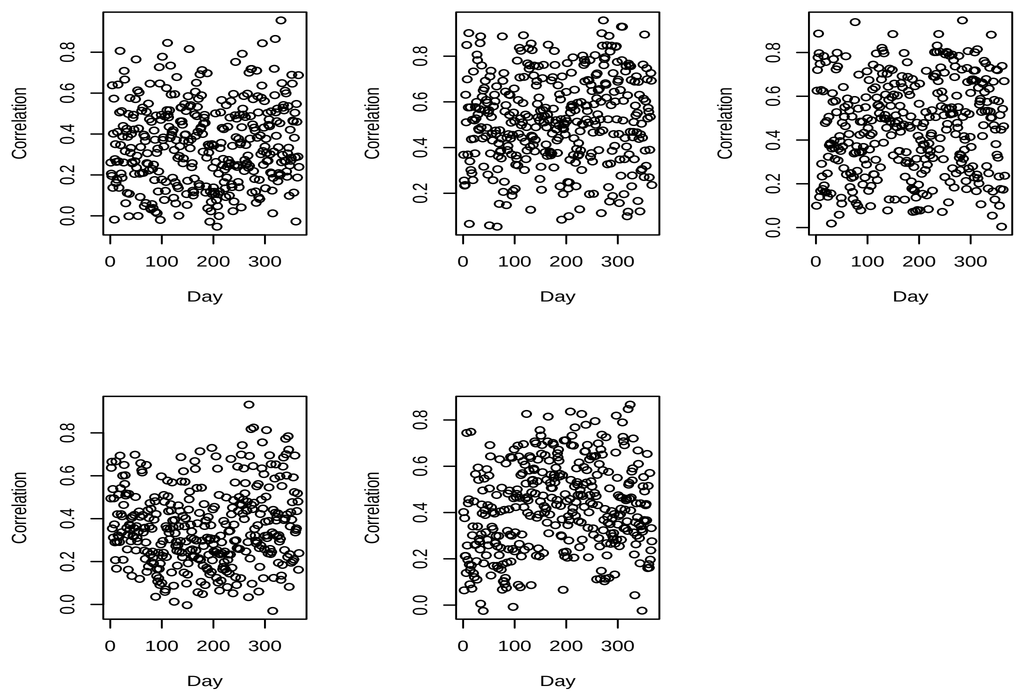

Figure 17: Estimated spatial correlations between Calgary's daily precipitation (mm) above the threshold and that at 5 other selected sites based on more than 100 years of data. Note the apparent lack of trend in these correlations over days of the year.

available data points (100 roughly including many zeros). In contrast, for MT we see a fairly constant intersite correlation over time. Note however, the apparently weaker spatial correlation during the summer.

Having examined spatial correlation patterns over time, we now turn to temporal correlation over space. More precisely, we investigate the persistence of temporal correlation over time (days) in the responses at different specified locations. For definiteness we pick a specific day, Jan 1 and see how strongly responses on that day autocorrelate with those on Jan 2nd, Jan 3rd,... in the same year. We can even go beyond that year and see if Jan 1st of the next year is autocorrelated with the same response a year earlier. Thus for both PCPN and MT we estimate the autocorrelation of Jan 1st's responses with each of the following 732 (2 years) days up to Jan 1 two years hence. In other words we investigate days as far apart as two years. Figures 19 and 21 show the results. For the PCPN process we use all the values including the zero's. They indicate that autocorrelation for PCPN remains significant for only a few days and that of MT for less than 20 days. This supports our assumption of response independence from year–to–year.

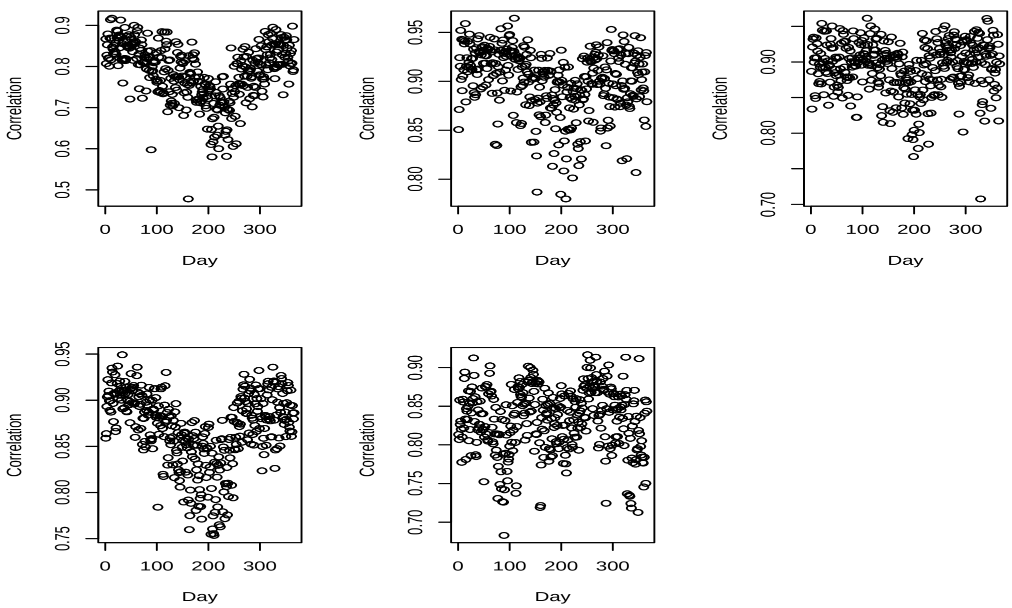

Figure 18: Estimated spatial correlations between Calgary's daily maximum temperature (C) and that at 5 other selected sites based on more than 100 years of data. Unlike precipitation, the maximum temperature field does display some trend over the year, with some evidence of a weakening in the spatial correlation in the summer.

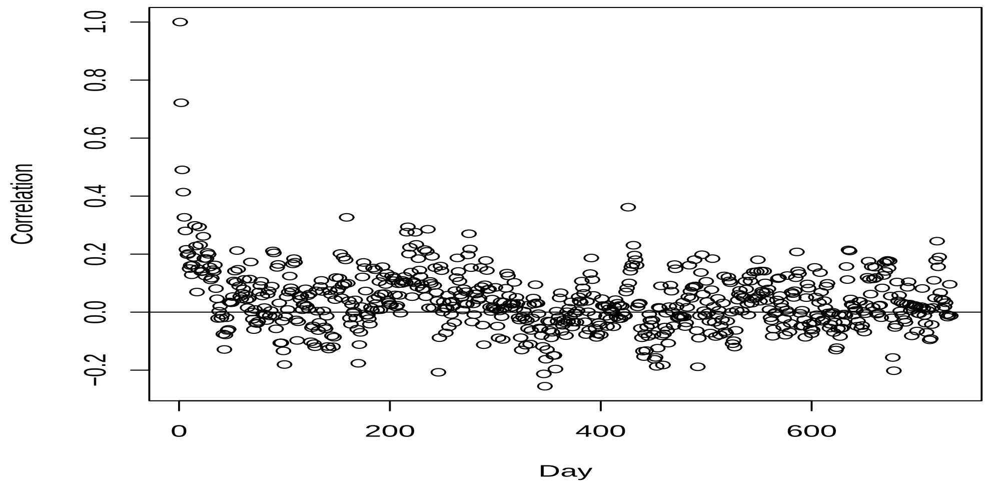

Figure 19: The estimated autocorrelation of Calgary's daily maximum temperature (C) on Jan 1st with each of 732 subsequent days. Notice the sharp decline to near zero as the time lag approaches 20 days.

The previous analysis was done for only Jan 1st. To see if the same properties hold for other days of the year, we examine four more days during the year. Figures 20 and 22 display the results, confirming our findings for Jan 1st.

3.4 Extreme events

This subsection addresses the final topic in our exploratory analysis, extreme events for climate data, specifically daily precipitation (PCPN in mm) and maximum daily temperature (MT in degrees C). In particular, our analysis seeks a foundation for a peak–over–threshold (POT) on which to build a models for those extremes. Such models commonly model the height of those peaks above a specified threshold, as a generalized Pareto distribution (GPD). But the threshold much be found by ad hoc methods. After experimenting with various options, we settled on the quantiles of the large datasets for PCPN and MT coming from all 102 stations described earlier and all years. Tables 3 and 4 give the results for various quantile levels.

Percentile	95	90	80	70	60
Threshold (mm)	9.7	5.1	1.6	0.3	0

Table 3: Thresholds for PCPN (mm) estimated as quantiles of the PCPN dataset for all 102 stations used in our analysis and all years, 1960 to 2000.

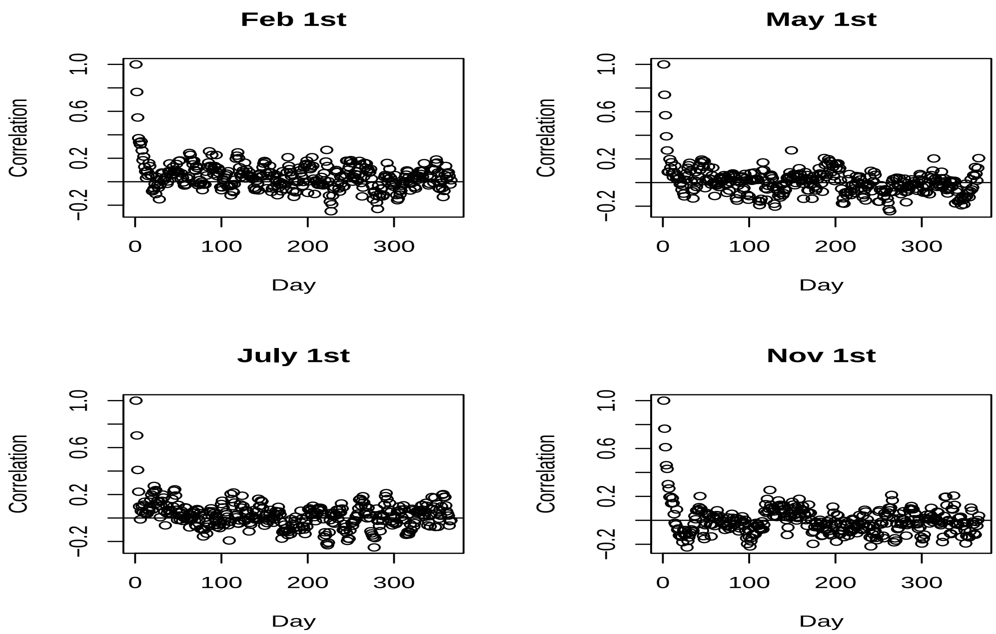

Figure 20: The estimated autocorrelation function for Calgary's maximum daily temperature (C) using 4 different selected start days. These results confirm that the process has a relatively short memory of less than 20 days.

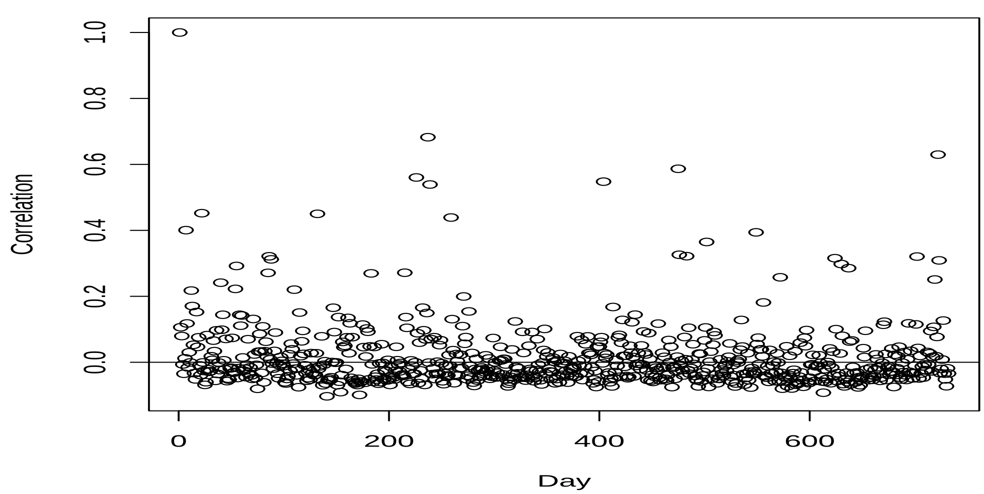

Figure 21: The estimated autocorrelation of Calgary's daily precipitation (mm) on Jan 1st with each of 732 subsequent days. Notice the sharp decline to near zero after just a few days.

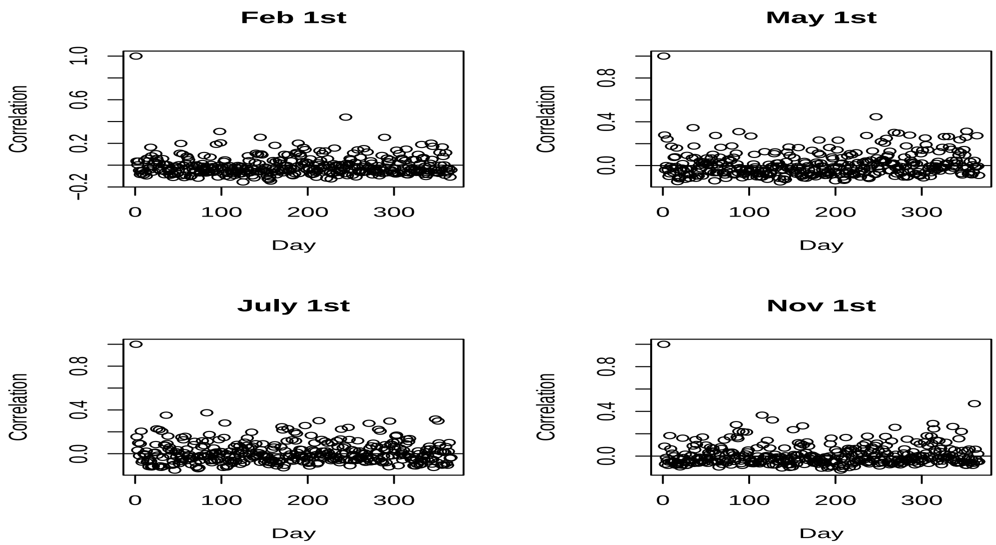

Figure 22: The estimated autocorrelation function for Calgary's daily precipitation (mm) using 4 different selected start days and only when precipitation exceeds the threshold. These results confirm that the process has a relatively short memory of just a few days.

Percentile	99	98	95	90	85	80
Threshold (C)	31	29.4	27	24.4	23	21.5

Table 4: Thresholds for MT (C) estimated as quantiles of the MT dataset for all 102 stations used in our analysis and all years, 1960 to 2000.

We must now determine which of these quantiles to use in order to ensure that the GPD approximates the POT sample distribution well. In fact, we found not only that the fit was fairly robust under varying thresholds, but as well, that the exponential distribution, the simplest GPD, provides a good approximation for various thresholds. Figures 23 and 25 give Calgary's results. Figures 24 and 26 confirm the accuracy of that approximation.

We we come to an issue of fundamental importance in modeling the environmental field of extreme values for our climate variables, more specifically their POT values. How do the (estimated) rate parameters for their exponential distributions vary over space? Recall that an exponential density function is given by

$$f(x; \lambda, u) = \lambda \exp -\lambda(x - u);\ x > u, \lambda > 0$$

where, u, $x - u$ and λ are respectively the threshold, the POT and the rate parameter. Intuition suggests stations in close proximity have similar $\hat{\lambda}$ estimates. To check our intuition, we use stations with PCPN and MT data records for 1960 to 2000.

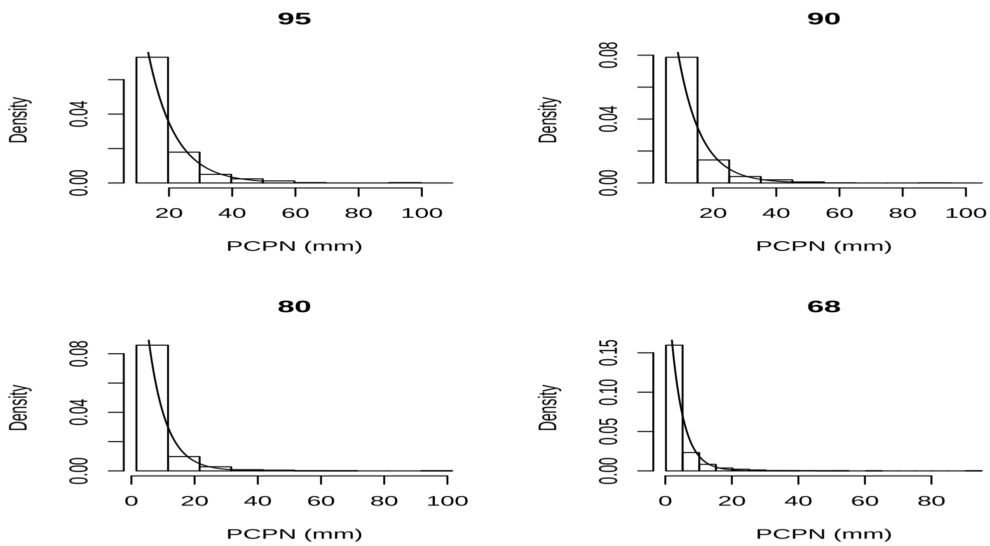

Figure 23: Exponential distributions fitted to POTs for Calgary's PCPN and different choices for the quantiles. The distribution fits well thereby bypassing the need to resort to the more class of generalized Pareto distributions.

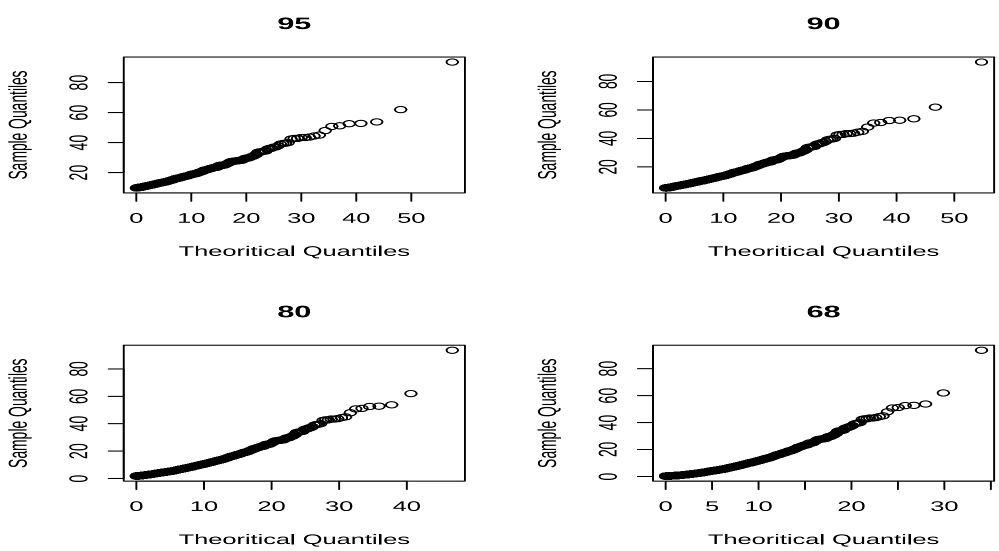

Figure 24: qq-plots for exponential fits to the peak–over–threshold values observed for different thresholds and Calgary's PCPN. This plot confirms the quality of fit observed in the previous figure.

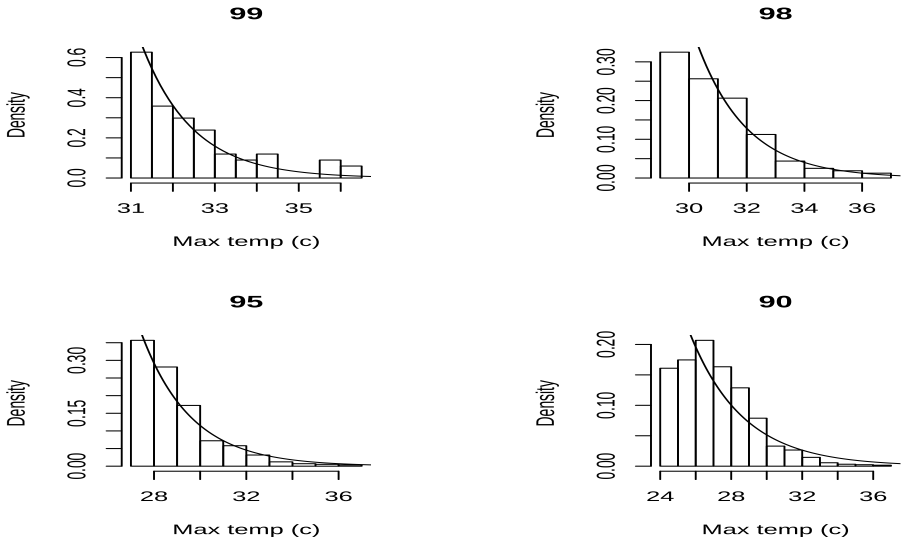

Figure 25: Exponential distributions fitted to POTs for Calgary's MT and different choices for the quantiles. The distribution fits well thereby bypassing the need to resort to generalized Pareto distributions with even more parameters than the exponential.

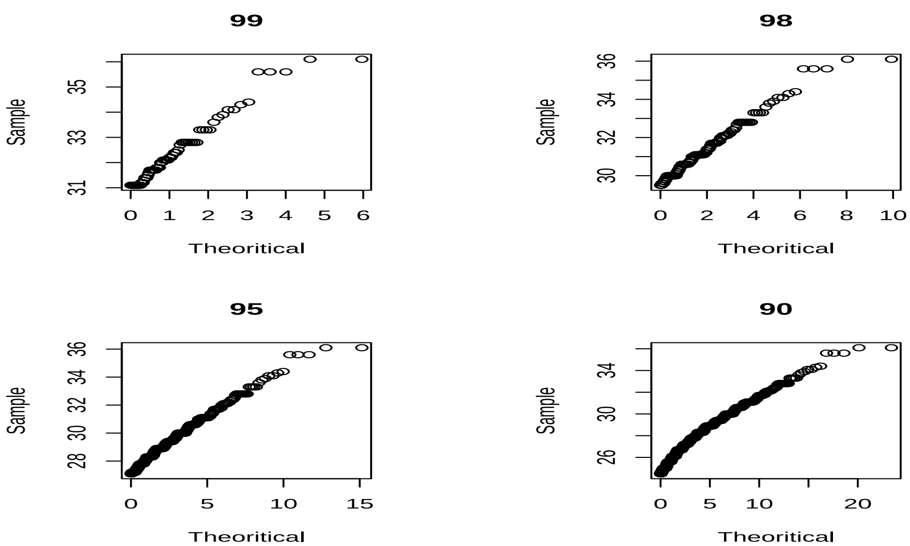

Figure 26: qq-plots for exponential fits to the peak–over–threshold values observed for different thresholds and Calgary's MT. This plot confirms the quality of fit observed in the previous figure.

For estimates generated from these data we use the semivariance, a standard geostatistical tool, to assess their spatial structure. In general the empirical semivariance is given by

$$\gamma(h) = \frac{\sum_{d(x,y)\approx h}(z(x) - z(y))^2}{2n(h)},$$

where $z(s)$ denotes a datum at location s, h, an intersite distance between any two points on the Earth's surface, and $n(h)$, the number of data point pairs at a distance of approximately h units apart.

Figures 27 and 28 display the variogram plots for PCPN and MT. Note that the mean numbers of data points available for estimating the parameters for the POT distributions for PCPN and MT are respectively, 2786 and 143, well above the minimum of 10 usually specified for estimating λ at any given single location. The means of the estimated parameter $\hat{\lambda}$ for PCPN and MT are respectively, 0.19 and 0.77 while the variances of $\hat{\lambda}$ for PCPN and MT are respectively, 0.04 and 30.23. Note the in the plots the distance is standardized (between 0 and 1).

The variogram plots confirm our intuition and show that indeed the estimates for the exponential approximations are close to one another over very wide spatial ranges. Notice the generally strong continuity implied by the plot, especially for small intersite distances. At the same time the big blips in the plot at certain distance point to possible irregularities in the field for certain distances.

3.5 Maximum yearly temperatures and precipitation.

This section looks at the maximum daily temperature (MT) and precipitation (PCPN) over the year. We also look at the dates during the year at which extreme values of these variables are realized. To be precise, define

$$YMT(s, d, y) = \max_{d=1,2,\cdots,366} MT(s, d, y)$$

and YPCPN, the annual maximum of daily precipitation, in the same fashion. Although other time intervals may be of interest, we restrict ourselves to the year in this preliminary analysis as that will be the period of principal interest.

We start by looking at the YMT time series for Calgary and Banff given in Figure 29. Figure 30 shows the same kind of results for six other selected sites for comparison and they show similar patterns to the previous two. Thus for YPCPN we present just the analogous plots the first two stations are these are seen in Figure 32. Overall the plots show YMT to be less variable through the years than YPCPN.

Next, we look at the date that these extreme happens over the years (Figures 31 and 33). The date seems to be more variable than the actual values.

Finally Figure 34 displays histograms of YMT for six selected stations while Figure 36 does the same for its logarithmic transformation. Figures 35 and 37 show normal qq-plots for these two variables. They show the Gaussian to be a better

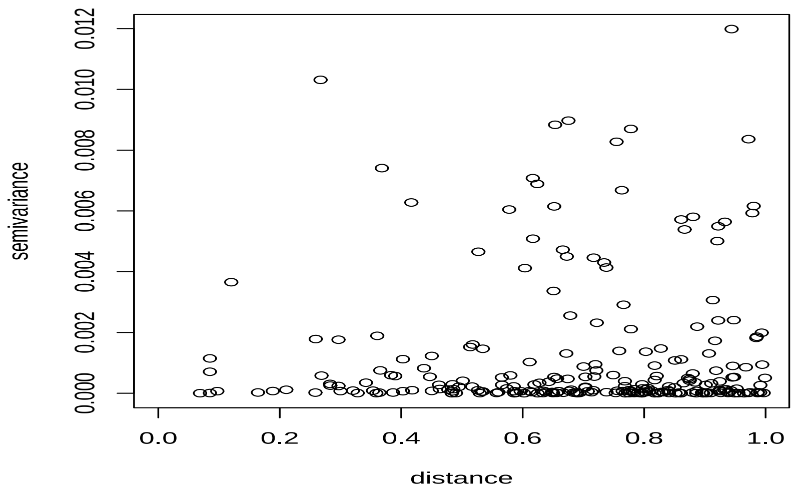

Figure 27: Variogram plot for the estimates of the exponential distribution parameter $\hat{\lambda}$ that best fits the height of the peaks of daily precipitation (PCPN in mm) over a threshold of 0.2 (mm). Notice the generally strong continuity implied by the plot, especially for small intersite distances. At the same time the big blips in the plot at certain distance point to possible irregularities in the field for certain distances.

approximation of the untransformed data than the untransformed. In both cases the distributions seem to be symmetric.

4 Lessons learned and concluding remarks

Any conclusions reached on the basis of the analyses described in this report much be viewed as tentative since the data include anomalies such as possible changes due to changes in instrumentation. Moreover, it was not possible to examine all of the stations in detail, forcing us to focus on a few representative of very different physical environments.

Nevertheless, the data do point strongly to certain conclusions and these we now summarize by category.

Precipitation

- For all practical purposes, we may take the lower threshold for precipitation as 2mm. Generally precipitation totals greatly exceed that amount. Moreover, the gamma distribution approximates these measurable amounts rather well. Finally the autocorrelation in these amounts remains significant for only lags of a few days.

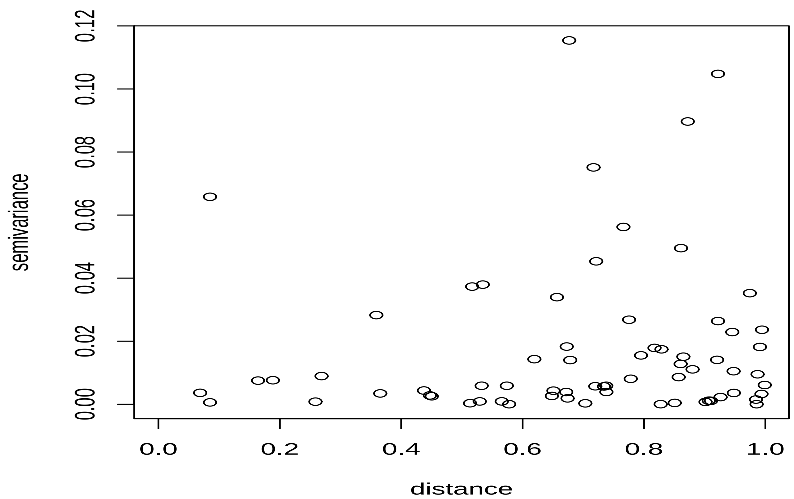

Figure 28: Variogram plot for the estimates of the exponential distribution parameter $\hat{\lambda}$ that best fits the height of the peaks of maximum daily temperature (MT in degrees C) over a threshold of 31 (^{o} C). Notice the generally strong continuity implied by the plot, especially for small intersite distances. At the same time the big blips in the plot at certain distance point to possible irregularities in the field for certain distances.

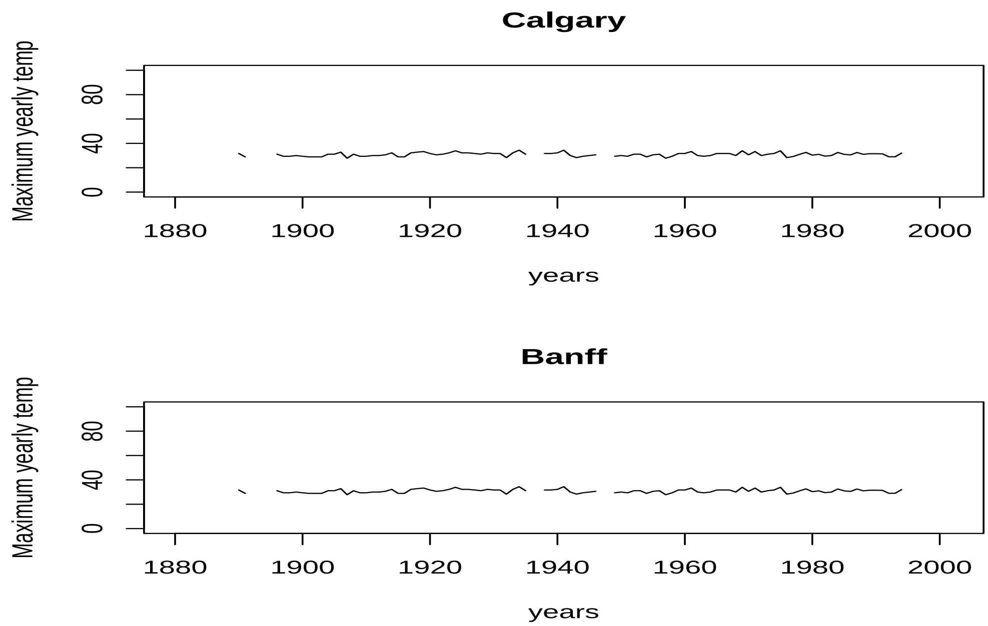

Figure 29: Time series plots of the annual maxima of the daily maximum temperature series at the Banff and Calgary stations.

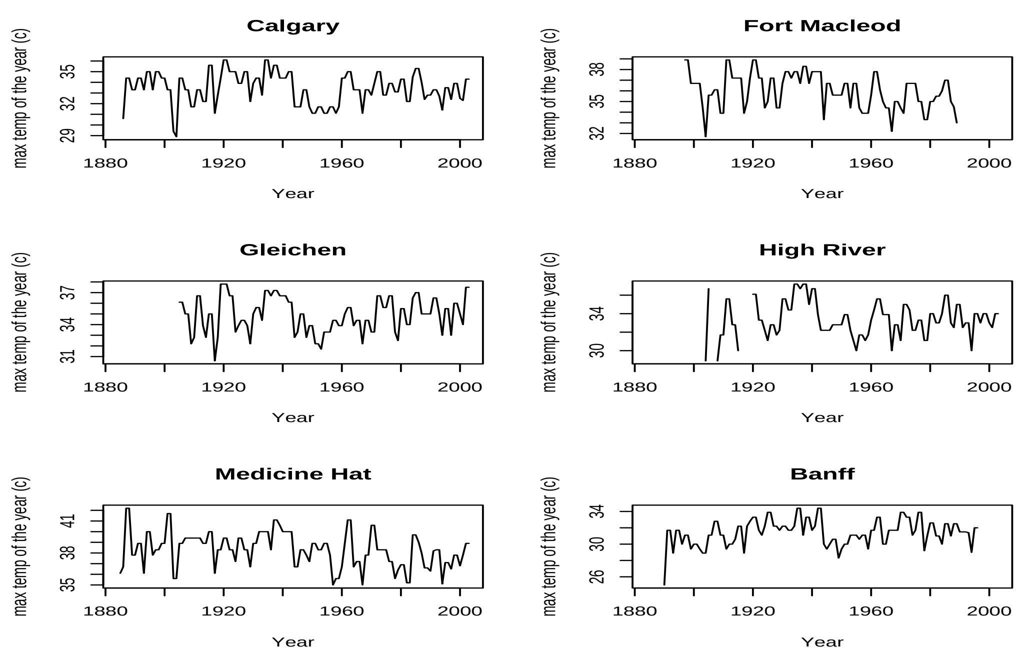

Figure 30: Time series plots of the yearly maxima of the maximum daily temperature at 6 selected Alberta monitoring stations.

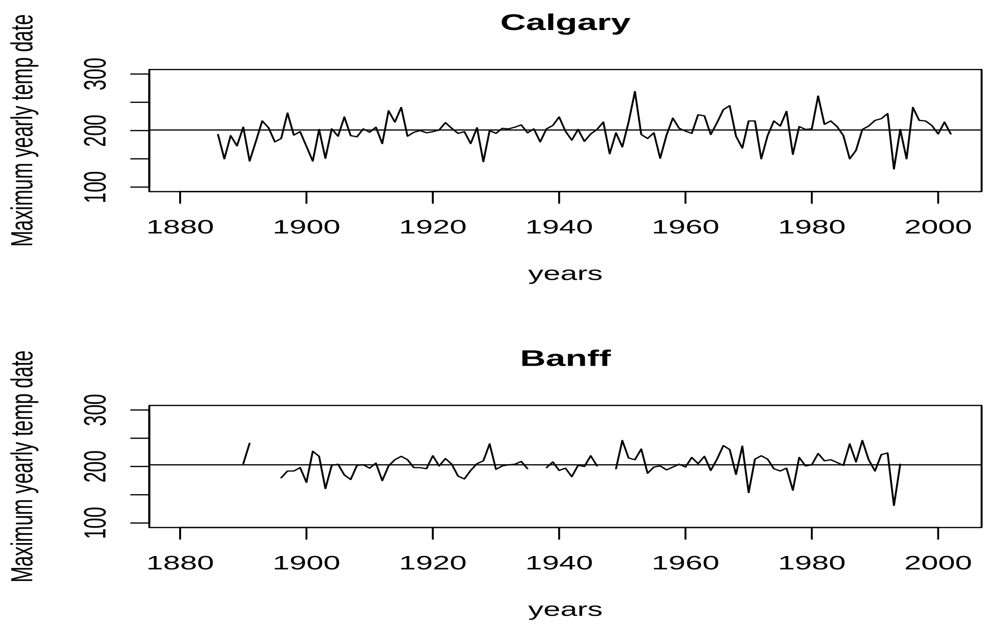

Figure 31: The day–of–the–year when the Banff and Calgary sites achieve their overall maximum of maximum daily temperatures.

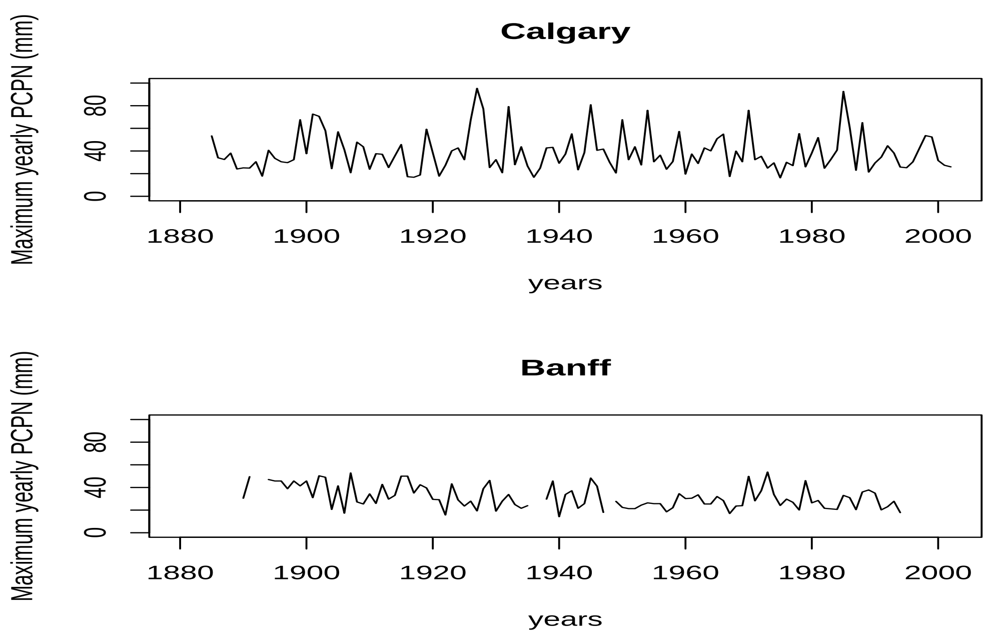

Figure 32: A time series plot of the annual maxima of the daily total precipitation (mm) measured at the Banff and Calgary stations.

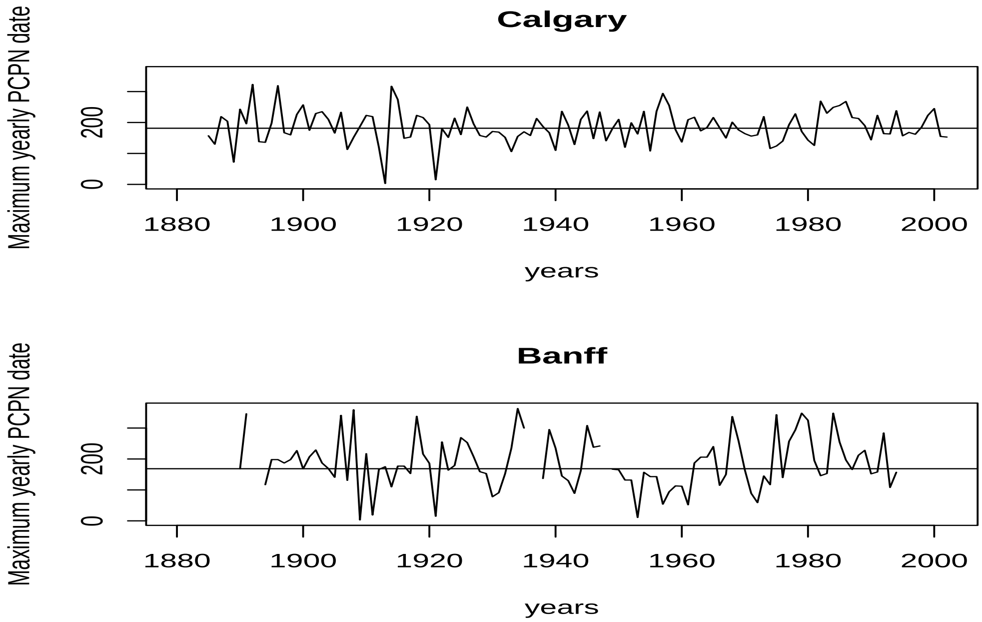

Figure 33: The day–of–the–year when the Banff and Calgary sites achieve their overall maximum of daily precipitation totals.

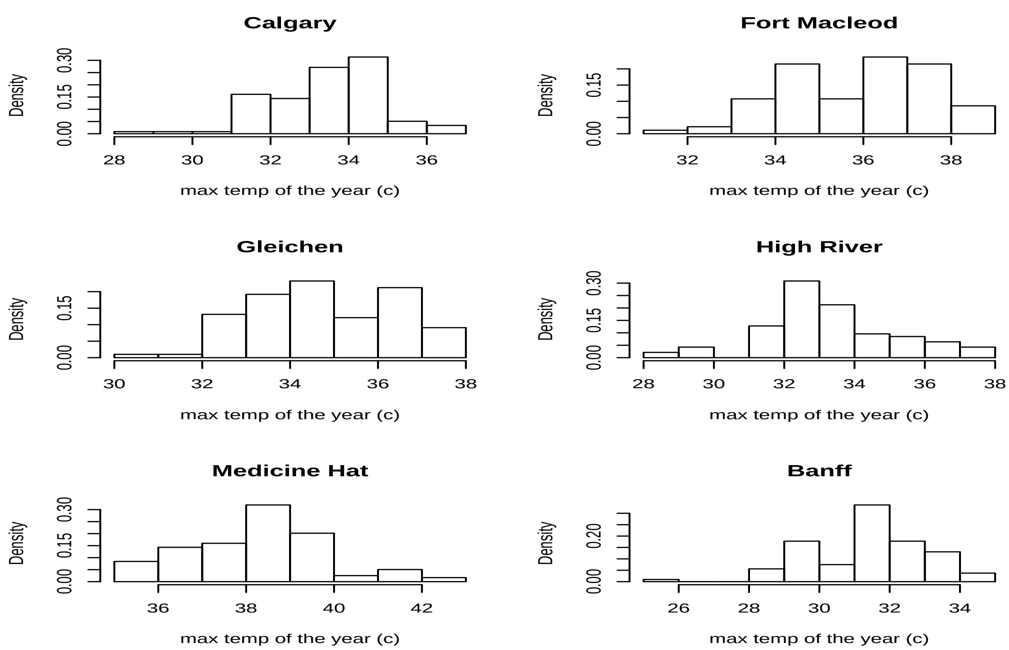

Figure 34: Histograms of the yearly maxima of maximum daily temperatures at six selected sites.

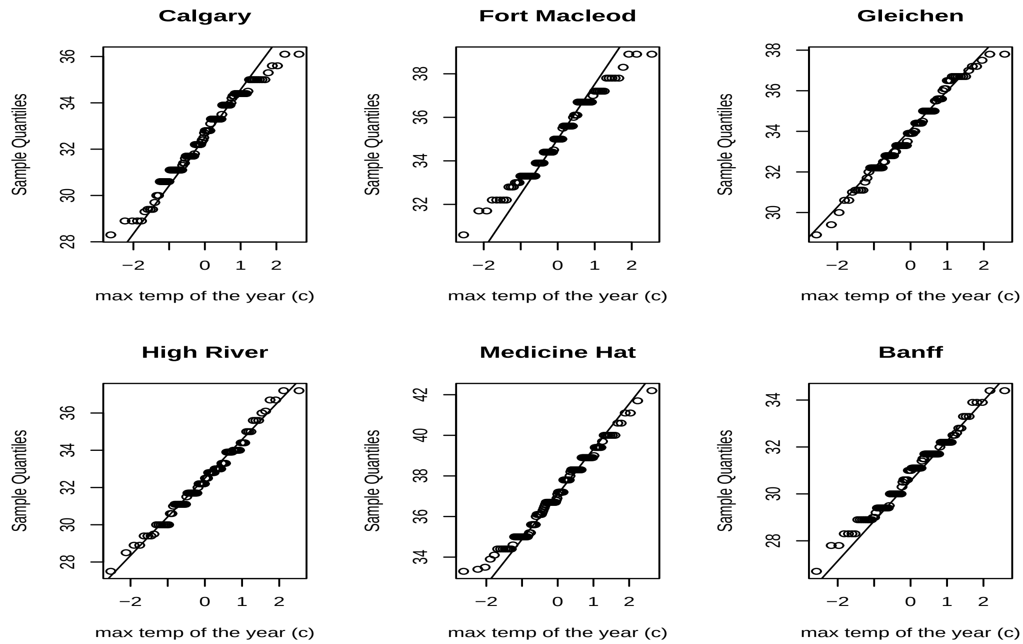

Figure 35: Normal qq-plots of the yearly maxima of daily maximum temperatures at 6 stations. Notice that the normal distribution fits the empirical distribution fairly well.

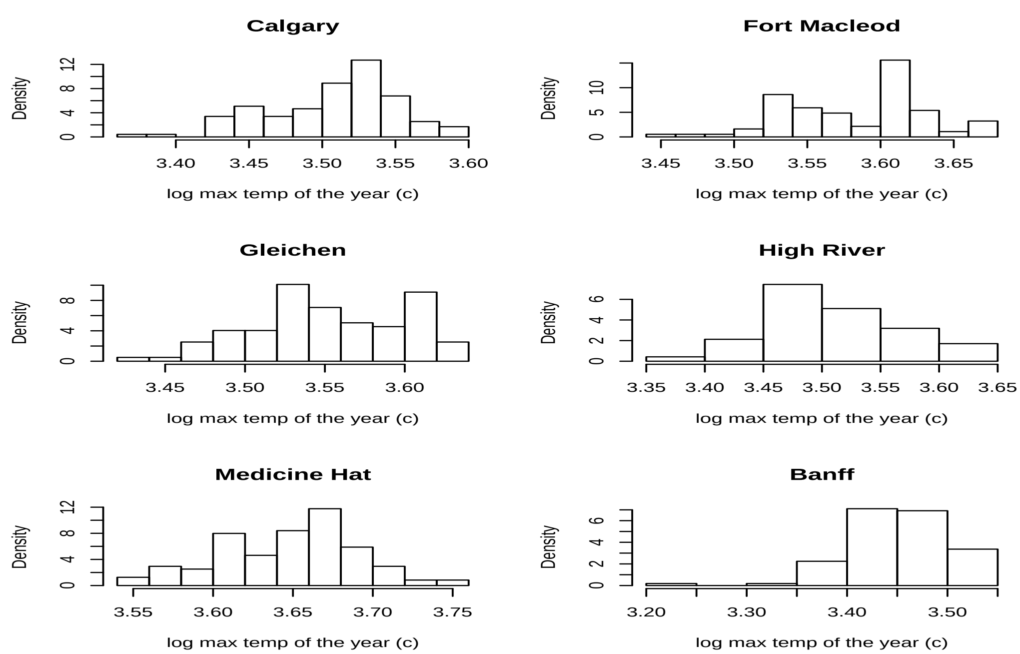

Figure 36: Histograms of the yearly maxima of maximum daily temperatures after logarithmic transformations for six selected sites.

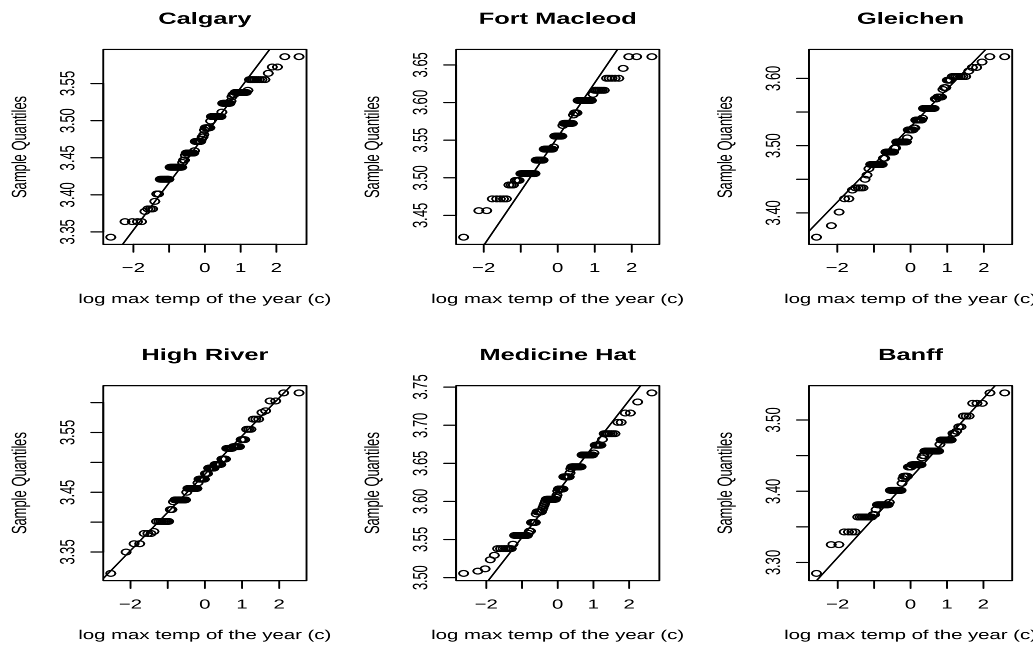

Figure 37: Normal qq-plots of the yearly maxima of daily precipitation at six selected Alberta monitoring stations. Notice that the normal distribution fits the empirical distribution fairly well.

- Space – time precipitation models must take into account that the measurable responses for this variable come from two very different populations, one consisting entirely of zeros.
- The Calgary site points to the June's monthly average as having not only the highest precipitation levels but also the most variable one over the years.
- The 10 year filter seems to have removed the short term variation and displayed well the long term trends in the series.
- The estimated intersite spatial correlation function for the space–time field of total daily precipitation above the threshold appears to be monotonically decreasing in agreement with intuition. On the other hand evidence of anisotropy is also seen.

Temperature

- The Gaussian distribution provides a reasonably good approximation to the observed sequence of maximum daily temperature data. Significant autocorrelation in the observed sequence last no more than 30 days.
- The estimated spatial correlation function for maximum temperature shows a strong persistence over the spatial domain as intuition might suggest.
- As in the case of precipitation, the decadel average seems the best low pass filter. We see two temperature peaks with very warm weather around 1915 and again in the 1990's.

Extreme events

- For both temperature and precipitation, the empirical distributions of the heights of peaks over specified threshold values are well approximated by the exponential distribution, in accordance with the statistical theory of extreme value.
- Geostatistical analysis shows the field of estimated parameters over the spatial domain of the exponential distributions in the previous bullet tends to be pretty flat. At the same intriguing blips appear for certain pairs of locations underlying that field and this issue warrants further investigation in future work.
- Although the size of the extremes do not vary a lot from year–to–year, the date at which they are realized does. Moreover the annual maximum temperature seems more variable from year–to–year than the corresponding value for precipitation.

In Part II of this analysis now in preparation, we turn to an enhanced version of the dataset treated in this report and carry out a more followup analysis albeit with a much smaller dataset.

References

[1] Cox, D. R., F.R.S and Isham, V. (1988) A simple spatial-temporal model of rainfall, Proc. R. Soc. Lond. A 415, pp 317-328.

[2] Embrechts, P., Kluppelberg, C., Mikosch, T., (1997) Modelling Extremal Events for Insurance and Finance, Springer.

[3] Hosseini, R. (2009). Extracting Canadian climate data. TR # 244, Department of Statistics, University of British Columbia.

[4] Rodriguez-Iturbe, I., Cox, D. R., F.R.S and Isham, V., (1988) A simple spatial-temporal model of rainfall, Proc. R. Soc. Lond. A 417, pp 283-298.

[5] Sampson, P. D. and Guttorp, P. (1992) Nonparametric Estimation of Non-stationary Spatial Covariance Structure, Journal of the American Statistical Association, Vol. 87, No 417, Theory and Methods, pp 108-119.

[6] Sanso, Bruno and Guenni, Leylys (1999) Venezuelan rainfall data analysed by using a Bayesian space-time model, Appl. Statist. vol. 48, part 3, pp 345-362.

[7] Sanso, Bruno and Guenni, Leylys, A Non-stationary Multisite Model for Rainfall.

[8] Shen, S. P., Dzikowski, P., Li, G. and Griffith, D. (2001) Interpolation of 1961-1997 Daily Temperature and Precipitation Data onto Alberta Ploygons of Ecodistrict and Soil Landscapes of Canada, Journal of applied meteorology, Vol. 40, pp 2162-2177.

[9] Shen, S. S. P and Yin, H. (2005) Temporal and Spatial Changes Of the Agroclimate in Alberta, Canada, from 1901 to 2002, Journal of Applied Meteorology, Vol. 44, pp 1090-1105.

[10] Wikle, C. K. Spatio-temporal Methods in Climatology, To appear in Encylopedia of Life Support Systems.

[11] Yaglom, A. M. (1981) Correlation theory of Stationary and Related Random Functions, Springer-Verlag.

[12] Bras, R. L., Iturbe, I. (1984) Random functions and Hydrology, Addison-Wesly.

[13] Yin, H. (2005) Statistical data analysis of Alberta Agroclimate and Global Surface Air Temperature. PhD Thesis, University of Alberta, Edmonton, AB T6G 2G1, Canada.

[14] Zhang, Z. and Switzer, P. (2007) Stochastic space-time regional rainfall modeling adapted to historical rain gauge data, Water Resources Research, Vol. 43, W03441.