University of British Columbia
Department of Statistics
Technical Report #246
May 2009

An Analysis of Alberta's climate. Part II: 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 is the second of two exploratory analyses of climate data for the Province of Alberta (precipitation and temperature) with the ultimate goal of developing new stochastic models for the processes involved. Its predecessor, Part I, presents a large number of findings that, like their confirmations in this report as well as its additional conclusions, provide foundations for statistical modeling these climate data and derivatives from them such as extreme values. However the data explored in Part I unlike those investigated here, had not been homogenized, that is adjusted for anomalies due to such things as changes in instrumentation over time. Thus its conclusions might better be viewed as hypothesis for further investigation and confirmation. However, on the positive side that data came from a very large number of monitoring sites covering the Province extensively.

Like Part I, Part II uses standard tools for exploratory data analysis. The lessons learned from that analysis are summarized as a basis for future modeling. Some of background material in Part I is repeated in this report for completeness.

1 Introduction

This report is the second of two exploratory analyses of climate data for the Province of Alberta (precipitation and temperature). Its predecessor, Part I, presents a large number of findings that, like their confirmations in this report, provide foundations for statistical modeling these climate data and derivatives from them such as extreme values. However the data explored in Part I unlike those investigated here, had not been homogenized, that is adjusted for anomalies due to such things as changes in instrumentation over time. Thus its conclusions might better be viewed as hypothesis for further investigation and confirmation. However those data come from a very large number of monitoring sites and cover the Province more extensively than the data addressed in Part II. Although the findings in this Part II differ in detail, they are broadly in agreement with those of Part I in spite of the differences in the two datasets involved. Moreover this second part gives the more detailed analysis the homogenized data support support and hence it includes additional findings.

The data includes measurements for daily maximum temperature (MT), daily minimum temperature (mt) and precipitation (PCPN). The temperature data have been provided to us by L.A. Vincent and the precipitation data, by Eva Mekis both at Environment Canada. This data set has been homogenized for changes of instrument, changes of station location and so on. More information about these data can be found in [3] and [4].

This report, like its predecessor, uses graphical and analytical tools to examine the behavior of selected climate variables. Looking at the data, we will see some interesting features that suggest future research and provide a foundation for statistical modeling.

Section 2 describes our data and in particular, provides plots of the geographical location of the stations from which the recorded measurements derive. In Section 3, we look at the daily and annual time series of temperatures and PCPNs. The normality of the distribution of annual values are investigated and the associations between different variables is investigated using simple regression analyses. We have also investigated how the seasonal patterns for different variables have changed over the course of the years. For example, the mean summer daily minimum temperature shows a significant increasing pattern over the course of the past century in Calgary and some other locations. Section 4 looks at the distribution of the daily values. For example, a normal distribution seems to describe well the temperature and a Gamma distribution, the daily precipitation values. Confidence intervals for the mean/standard deviation in the normal case and shape/scale parameters in the Gamma case are given. Section 5 looks at the spatial temporal correlation of different variables. In Section 6, we look at the extreme values, such as maximum annual temperature and annual minimum temperature.

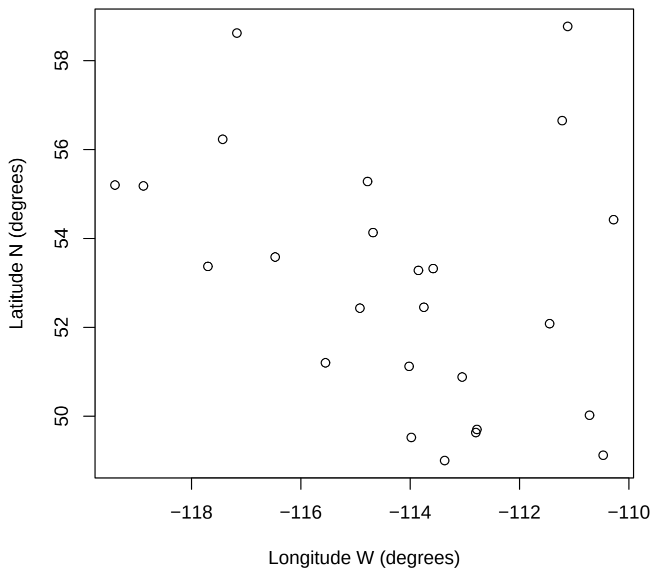

Figure 1: Alberta site locations for temperature (deg C) data. There are just 25 stations available with temperature data over Alberta, substantially fewer than in Part I.

2 Data description

The temperature data come from 25 stations operating between 1895 and 2006, that are distributed over Alberta. Their 47 precipitation counterparts recorded the PCPN data from 1895 to 2006 over variable recording intervals. For example, Caldwell provided PCPN data from 1911 to 1990. Figures 1 and 2 respectively depict the locations of the stations for temperature (both MT and mt) and PCPN. The number of years available for each station is plotted against the location in Figures 3 and 4. Another available variable for the location of the stations is the elevation and Figures 5 as well as 6 show their elevation in meters.

3 Temperature and precipitation

To get initial impressions of the data, we look at the time series of MT, mt, and PCPN at a single fixed location, the Calgary site since it has a long period of data available and includes both temperature and precipitation. Looking at the maximum and minimum temperature, we see a periodic trend over the year as shown in Figures 7 and 8, illustrating the MT and mt daily values from 2000 to 2003. A regular seasonal trend is seen in both processes.

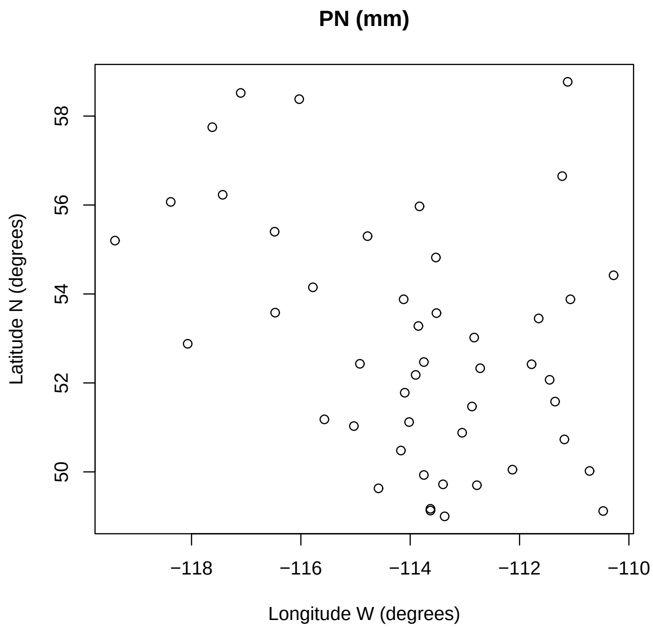

Figure 2: Alberta site locations for precipitation (mm) data. In all 47 monitoring stations provide data for the homogenized dataset.

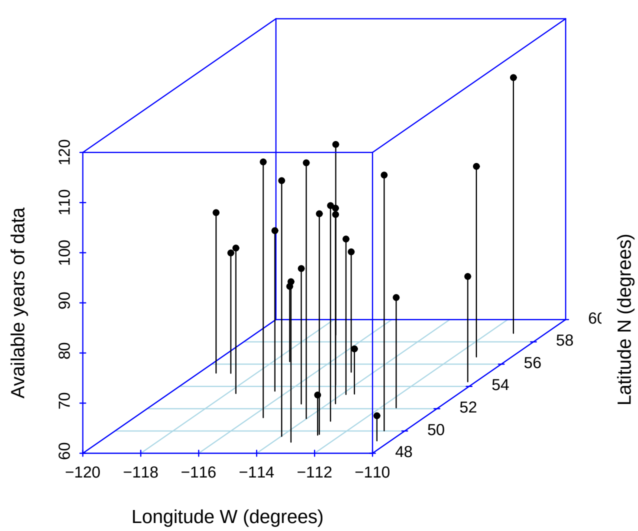

Figure 3: This "skyscraper plot" shows the number of years available for sites providing temperature (deg C) data.

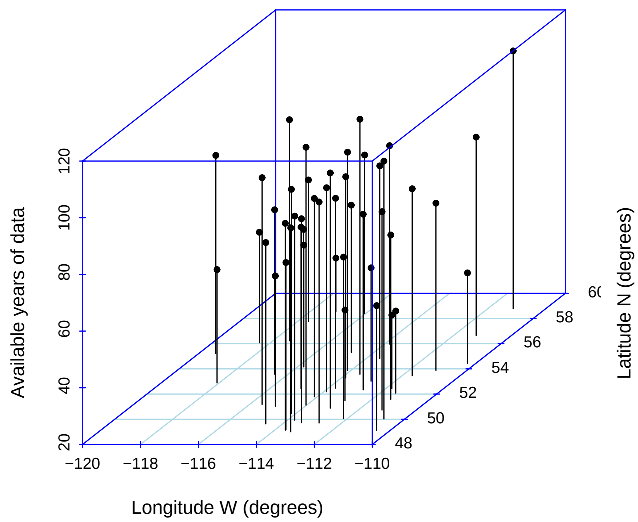

Figure 4: The number of years available for sites with PCPN (mm) data available.

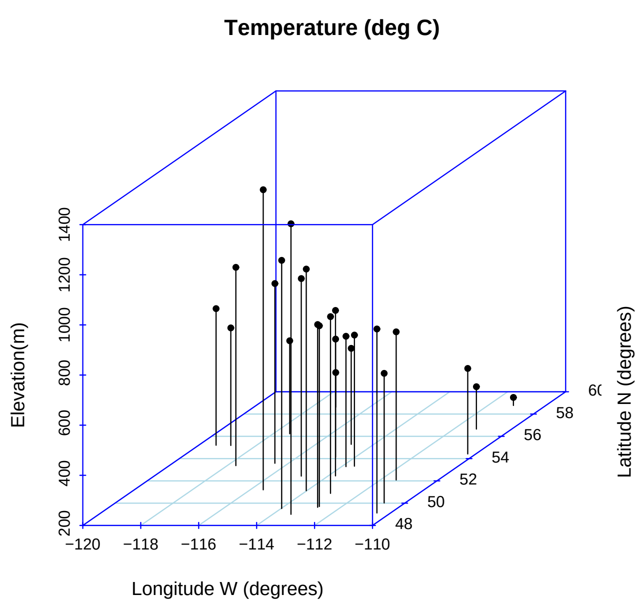

Figure 5: The elevation (meters) of sites with temperature data available.

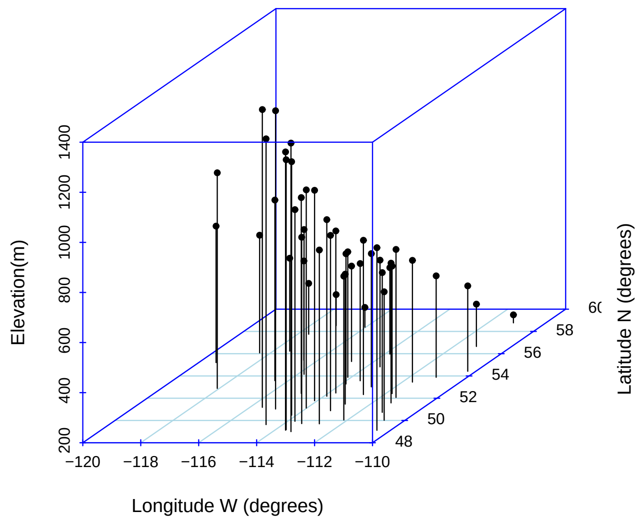

Figure 6: The elevation (meters) of the sites with precipitation data available.

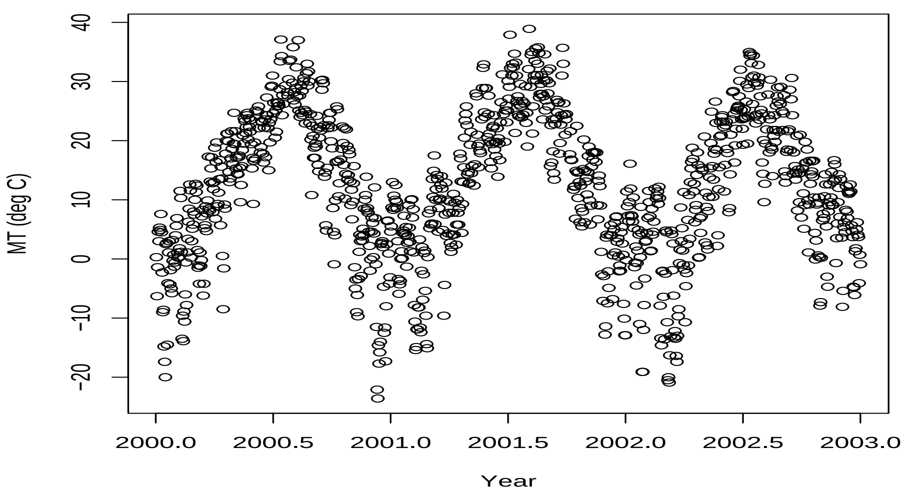

Figure 7: The time series of daily maximum temperature (deg C) for the Calgary site, 2000 to 2003. The dark band along the horizontal axis represents the zero precipitation days.

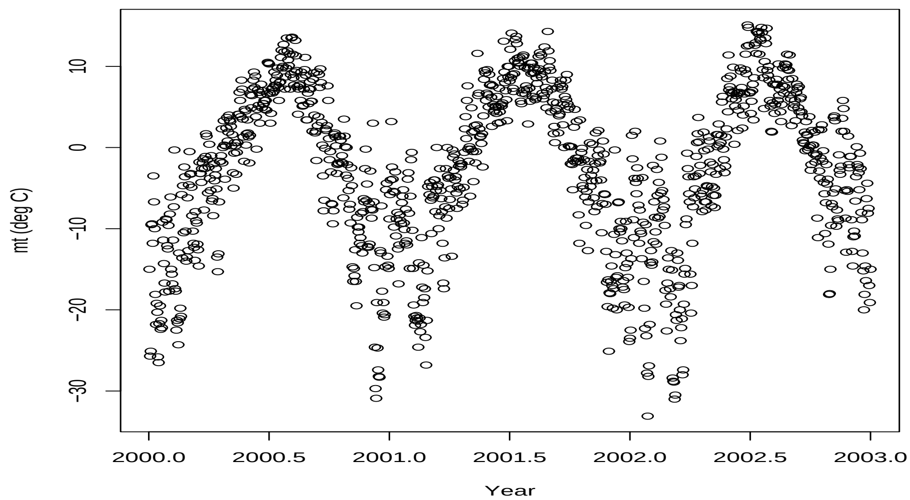

Figure 8: The time series of daily minimum temperature (deg C) for the Calgary site, 2000 to 2003.

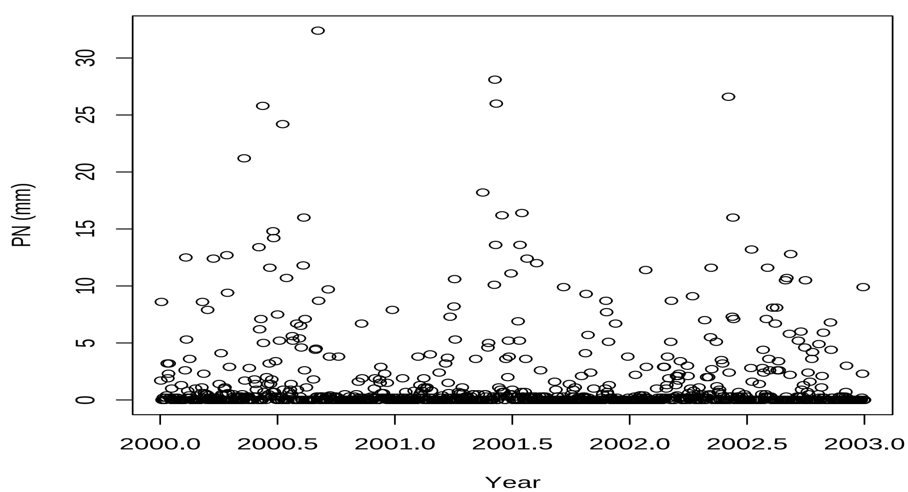

Figure 9: The time series of daily precipitation (mm) for Calgary 2000-2003.

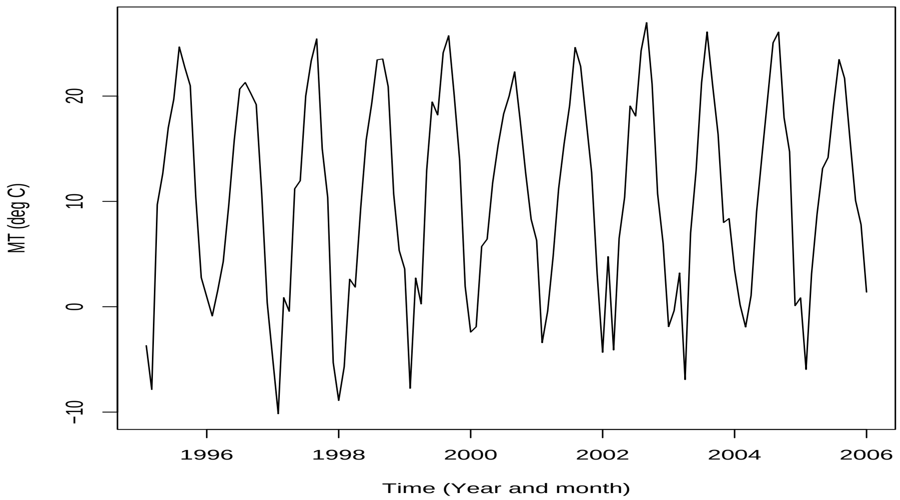

Figure 10: The time series of monthly maximum temperature (deg C) for Calgary 1995-2005.

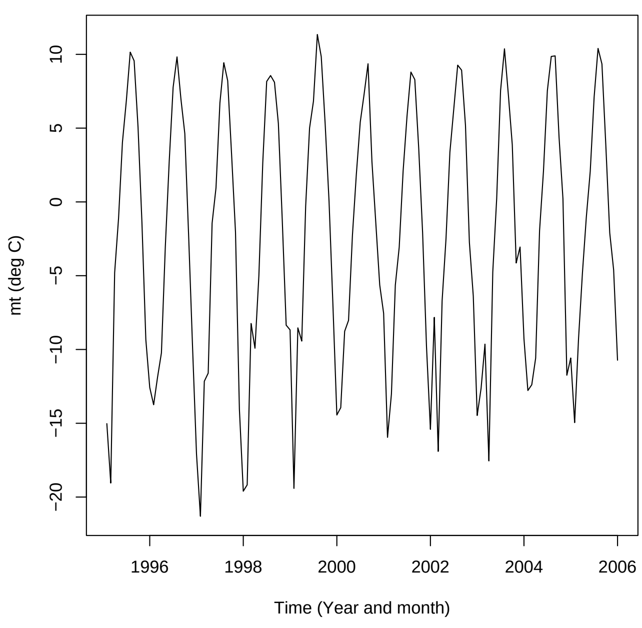

Figure 11: The time series of monthly minimum temperature means (deg C) for Calgary 1995-2005.

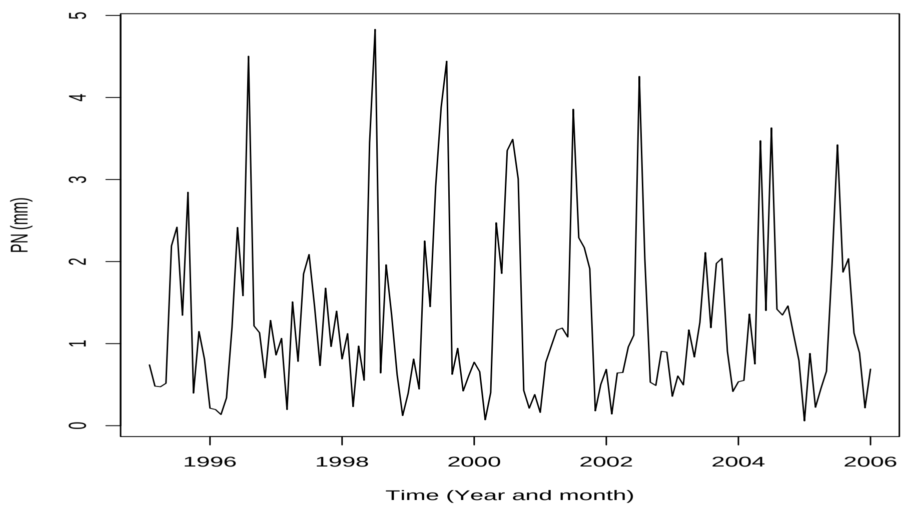

Figure 12: The time series of monthly precipitation means (mm) for Calgary 1995-2005.

Looking at the PCPN plot in Figure 9, we observe a large number of zeros. Moreover, seasonal patterns are hard to see by looking at daily values. To illustrate the seasonal patterns better, we look at the monthly averages for MT, mt and PCPN over the period 1995 to 2005 in Figures 10, 11 and 12. The seasonal patterns for precipitation can be seen clearly in Figure 12.

Next we look at the mean annual values of the three climate variables for all available years that have less than 10 missing days (Figures 13, 14 and 15). Table 1 gives a summary of these annual means.

Variable	min	1st Quartile	median	mean	3rd Quartile	max
MT (deg C)	7.59	9.64	10.37	10.36	11.19	13.46
mt (deg C)	-4.83	-3.40	-2.54	-2.66	-1.95	0.07
PCPN (mm)	0.68	1.12	1.28	1.29	1.39	2.51

Table 1: The summary statistics for the annual means of the three climate variables of central interest in our analysis.

Assuming stochastic normality and independence of the observations, we can obtain confidence intervals for the annual means of all three variables and these are given in Table 2. The confidence intervals are fairly narrow.

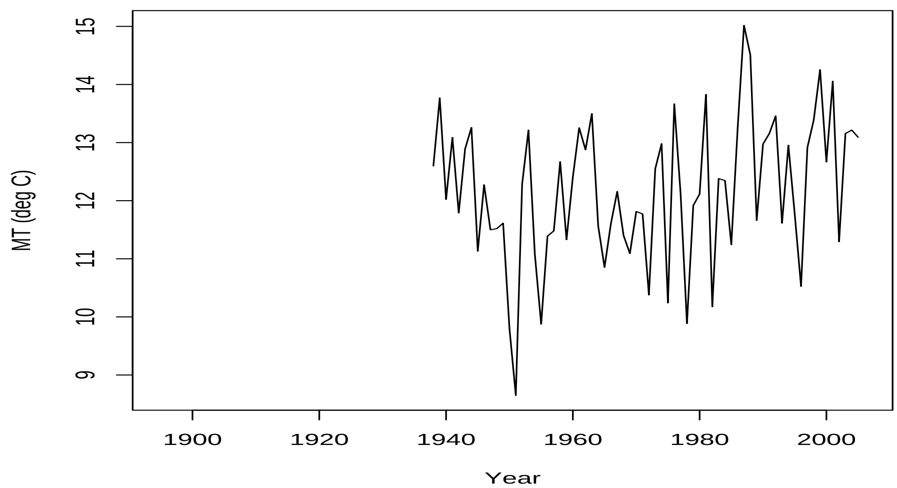

Figure 13: The annual mean maximum temperature (C) for Calgary site for all available years.

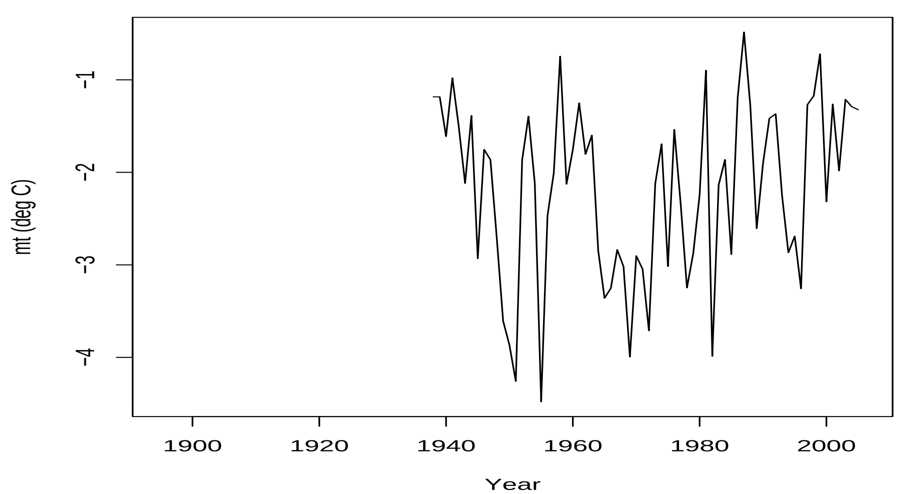

Figure 14: The annual mean minimum temperature (C) for Calgary site for all available years.

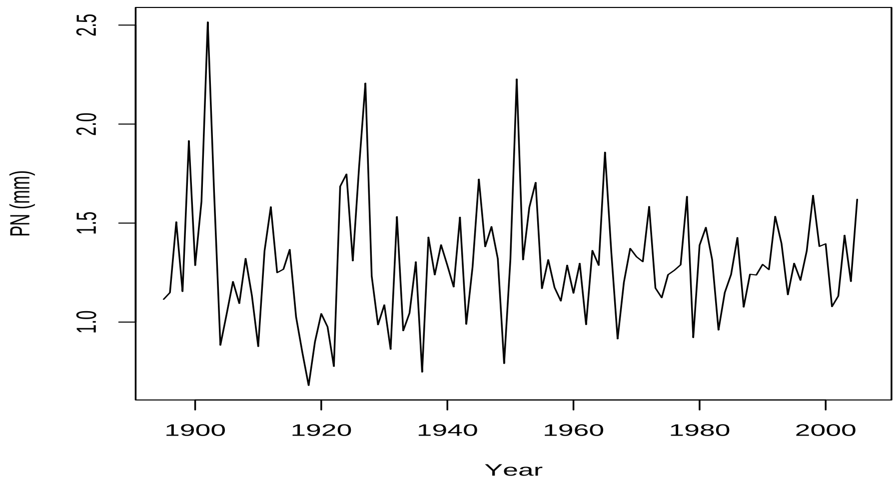

Figure 15: The annual mean precipitation (mm) for Calgary site for all available years.

Variable	95% confidence interval
MT (deg C)	(10.14, 10.57)
mt (deg C)	(-2.85, -2.47)
PCPN (mm)	(1.24, 1.35)

Table 2: Confidence intervals for the annual means of each of the climate variables in our study.

To investigate the shape of the distribution of annual means, we look at the histogram of each variable with a normal curve fitted in Figures 16, 18 and 20. The corresponding normal qq-plots are also given in Figures 17, 19 and 21 to asses the normality assumption. Both the histogram and the qq-plots for MT validate the normality assumptions. The histogram for *mt* is slightly left skewed. For PCPN, some deviation from the normality assumption is seen. This is expected since the daily PCPN process is very far from normal to start with. Hence, even averaging through the whole year has not quite given us a normal distribution.

We plot all three variables (annual mean MT, mt and PCPN) in the same graph, Figure 22. As shown in that figure, MT and mt show the same trends over time. To get an idea of how the two variables are related, we fit a regression line, taking mt as response and MT as the explanatory variable. As seen in Figure 23, the regression fit looks very good. We repeat this analysis, this time taking MT as explanatory variable and PCPN as response. As shown in Figure 24, the fit remains reasonably good, although the association is not as strong. As shown

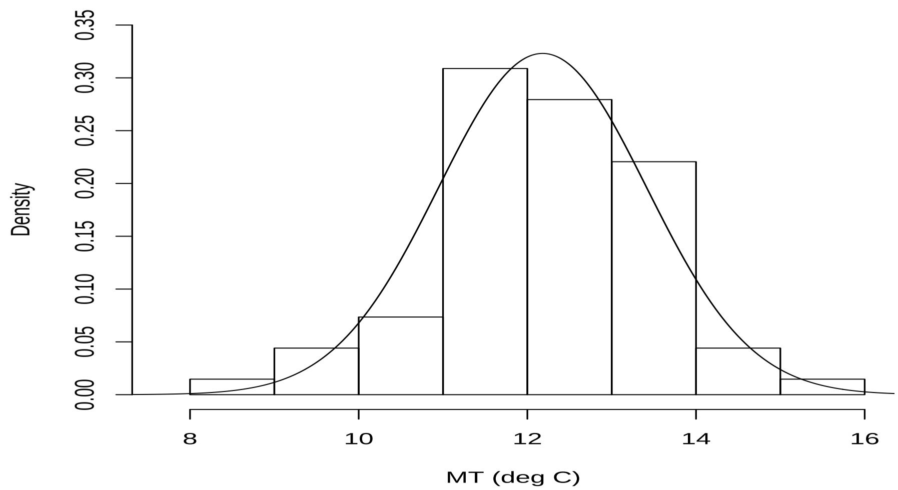

Figure 16: The histogram of annual maximum temperature means (deg C) for Calgary with a normal curve fitted to the data.

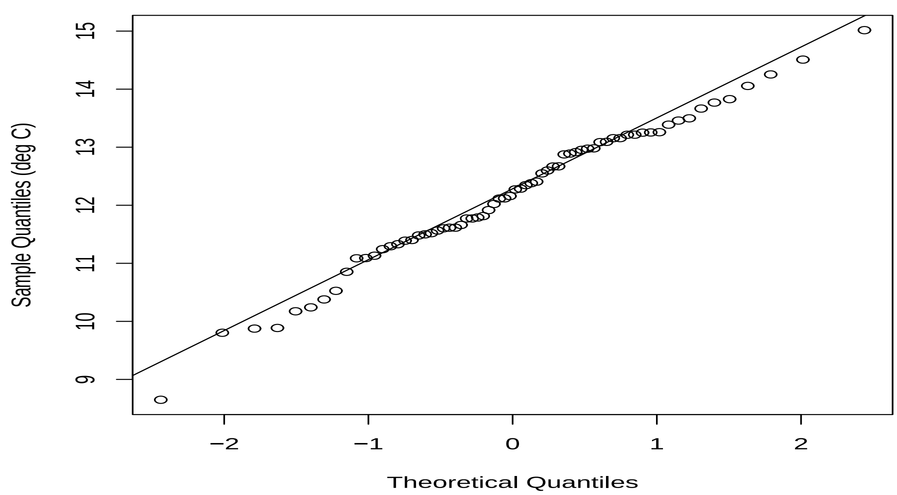

Figure 17: The normal qq-plot for annual maximum temperature means (deg C) for Calgary.

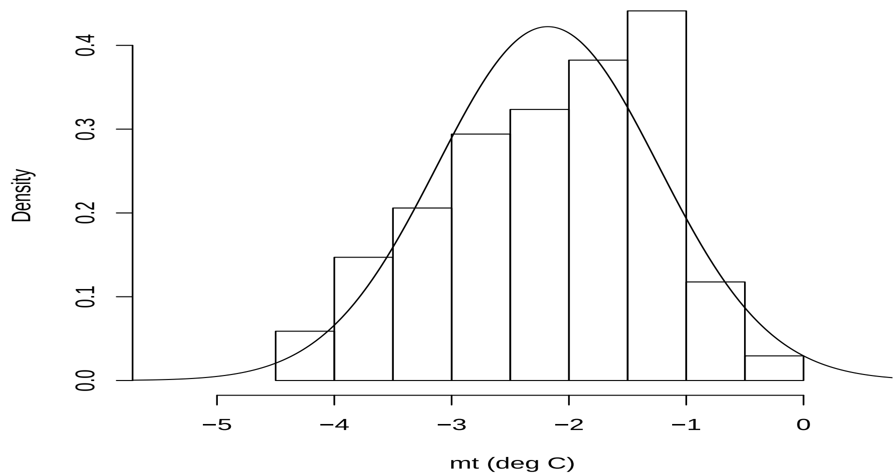

Figure 18: The histogram of annual minimum temperature means (deg C) for Calgary with normal curve fitted to the data.

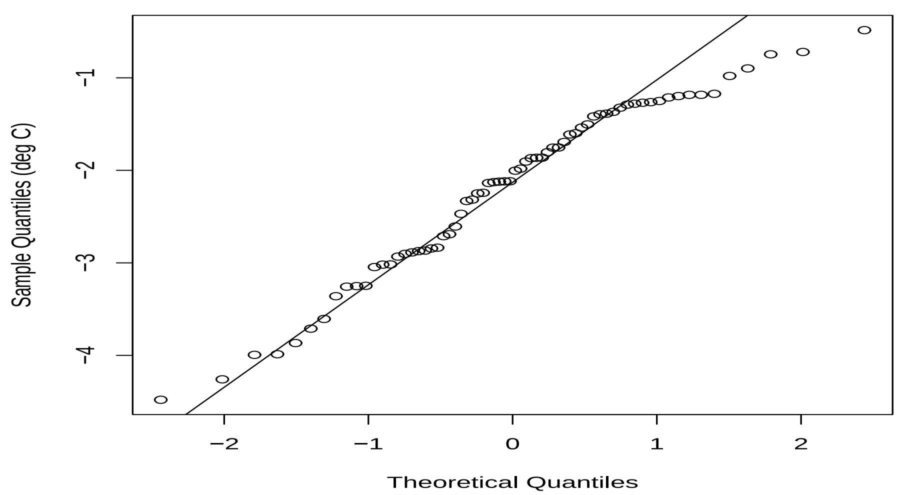

Figure 19: The normal qq-plot for annual minimum temperature means (deg C) for Calgary.

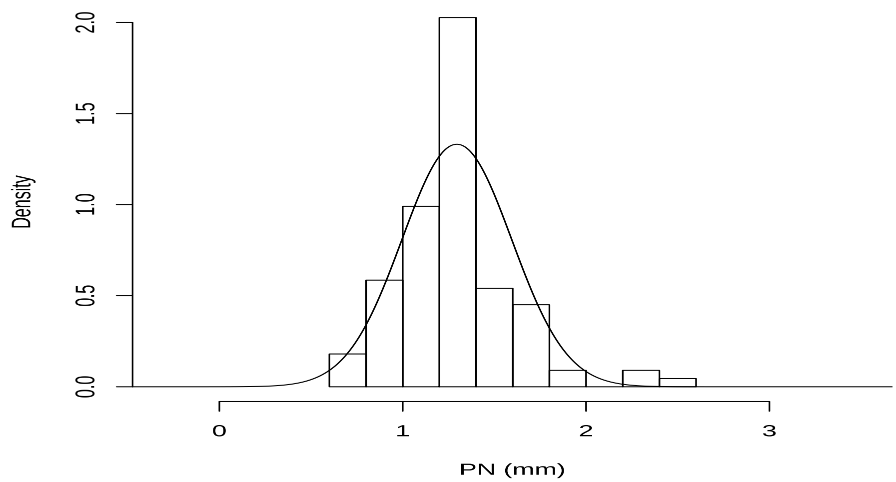

Figure 20: The histogram of annual precipitation means (mm) for Calgary with normal curve fitted to the data.

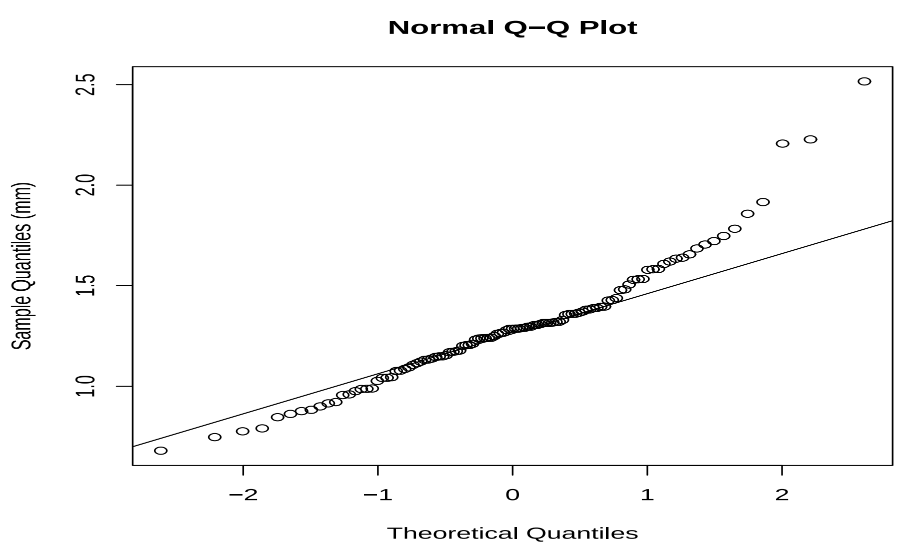

Figure 21: The normal qq-plot for annual precipitation means for Calgary.

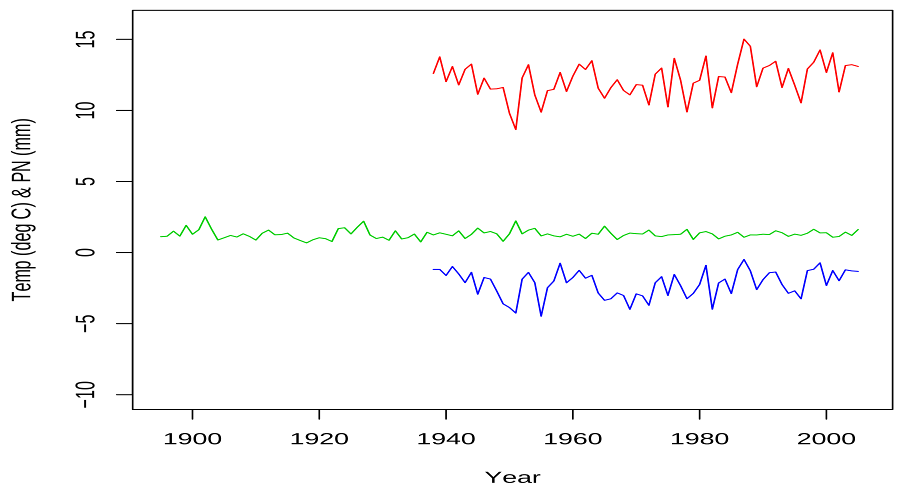

Figure 22: The time series plots of maximum temperature (deg C), minimum temperature (deg C) and precipitation (mm) annual means for Calgary. The time series plot in the bottom is minimum temperature, the one in the middle is precipitation and the top is maximum temperature.

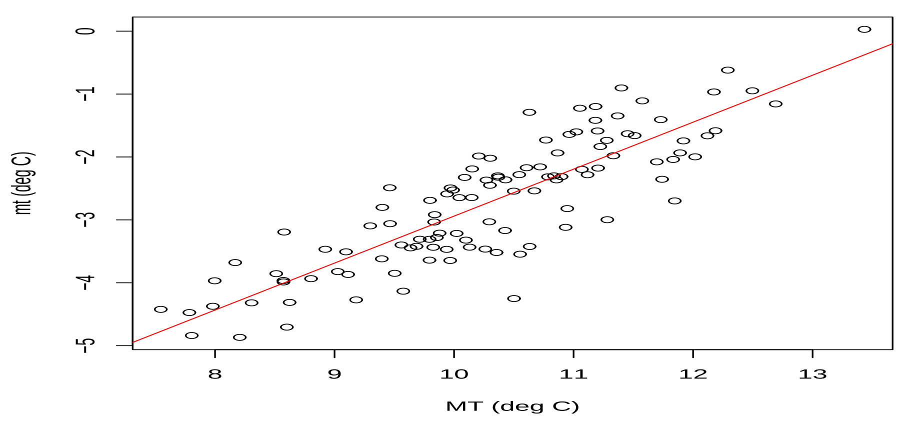

Figure 23: The regression line fitted to maximum temperature and minimum temperature annual means for Calgary.

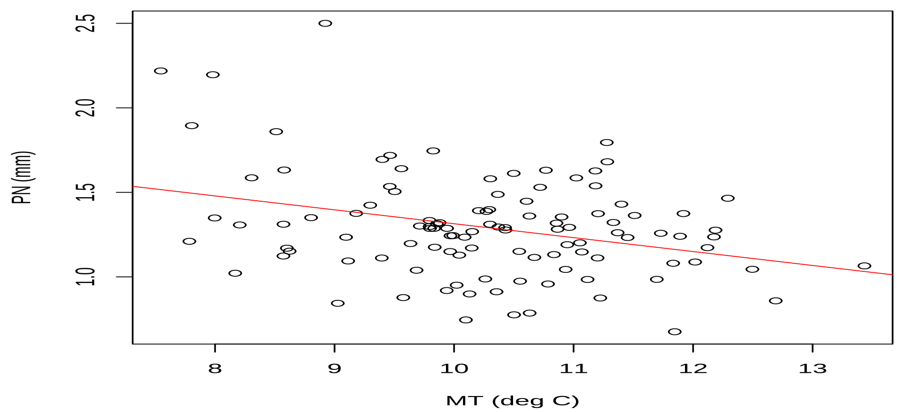

Figure 24: The regression line fitted to maximum temperature and precipitation annual means for Calgary.

in Table 3, both fits are significant. One can criticize use of a simple regression since the independence assumption might not be satisfied. Finding a more reliable and sensible relationship between the variables needs a multivariate model taking account of correlation and other aspects of the processes. Also note that these are annual averages which are not as correlated as daily values over time as seen in the annual time series plots.

Variables	Intercept	Slope	p-value for Intercept	p-value for slope
mt (deg C)	-10.40	0.746	2×10^{-16}	2×10^{-16}
PCPN (mm)	2.13	-0.082	1.49×10^{-14}	0.0005

Table 3: Lines fitted to annual mean minimum temperature and annual mean precipitation against annual mean maximum temperature.

Next we look at the change in the seasonal means for all three variables. As we noted above there are missing data, particularly near the beginning of the time series. This has caused the gap at the beginning of most plots. To get a longer time series of means, we first compute the monthly means allowing 3 missing days and then compute the yearly mean using the monthly means. This procedure is reasonable since days close to each other have similar values. We do the regression analysis for three locations: Calgary, Banff and Medicine Hat. We fit the regression line to annual means, spring means, summer means, fall means and winter means for each of MT, mt and PCPN with respect to time. The results are given in Tables 4, 5 and 6. (We have only included fits that turned out to be significant.) Note that PCPN does not appear in any of the tables. Annual minimum temperature and summer mean temperature show an increase in all three locations. Figure 25

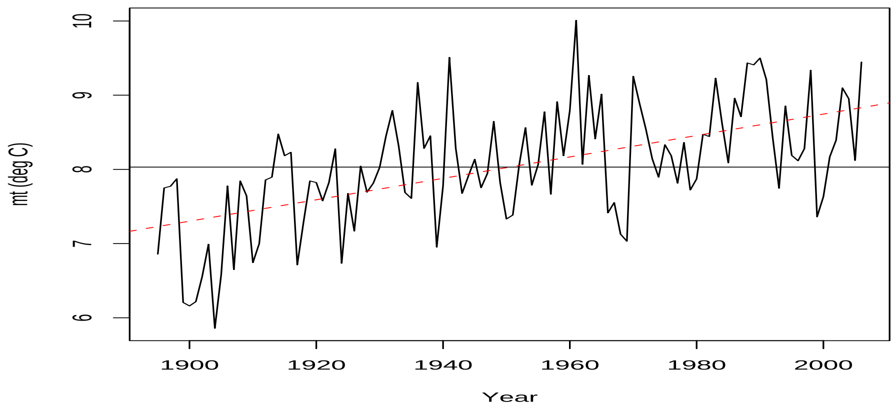

Figure 25: The regression line fitted to summer minimum temperature means against time for Calgary.

depicts one of the time series (mt summer mean for Calgary) with the regression line fitted.

Variable	Season	Intercept	Slope	p-value for the intercept	p-value for the slope
mt (deg C)	Year	-24.72	0.112	$2 \times 10-05$	0.0001
mt (deg C)	Spring	-30.05	0.138	0.0008	0.0024
mt (deg C)	Summer	-20.11	0.0144	6×10^{-7}	3×10^{-11}

Table 4: The regression line parameters for the fitted lines for each variable with respect to time for the Calgary site.

Variable	Season	Intercept	Slope	p-value for the intercept	p-value for the slope
MT (deg C)	Year	-12.99	0.0105	0.019	0.0002
MT (deg C)	Spring	-17.0	0.0048	0.075	0.009
MT (deg C)	Summer	255.5	-1.1	0.0006	0.646
MT (deg C)	Fall	-12.64	0.0106	0.19	0.0326
mt (deg C)	Year	-37.0	0.01666	$2 \times 10-10$	2×10^{-8}
mt (deg C)	Spring	-49.8	0.0229	5×10^{-9}	10^{-7}
mt (deg C)	Summer	-36.8	0.0212	2×10^{-15}	2×10^{-16}

Table 5: The regression line parameters for the fitted lines for each variable with respect to time for the Banff site.

Variable	Season	Intercept	Slope	p-value for the intercept	p-value for the slope
MT (deg C)	Year	-24.6	0.0185	0.00102	3×10^{-6}
MT (deg C)	Spring	-34.24	0.0235	0.009	0.0005
mt (deg C)	Year	-39.98	0.0197	$5 \times 10-10$	2×10^{-9}
mt (deg C)	Spring	-39.81	0.0196	5×10^{-5}	9×10^{-5}
mt (deg C)	Summer	-10.93	0.0112	0.0199	7×10^{-6}
mt (deg C)	Fall	-24.66	0.0122	0.0110	0.0137

Table 6: The regression line parameters for the fitted lines for each variable with respect to time at the Medicine Hat site.

4 Daily values and their distributions

This section focuses on daily values for all three variables. To that end, we pick four days, Jan 1st, April 1st, July 1st and October 1st, to span the year's climate. Let us look at the time series, histograms and normal qq-plots for each variable over the years. Figures 26 to 31 give the results. In fact the plots validate the assumption of normality for daily MT and mt for the selected days of the year. We also tried the first day of each month and observed similar results.

We plot the histogram for PCPN as well (Figure 33). The distribution is far from normal because of high frequency of no PCPN (dry) days.

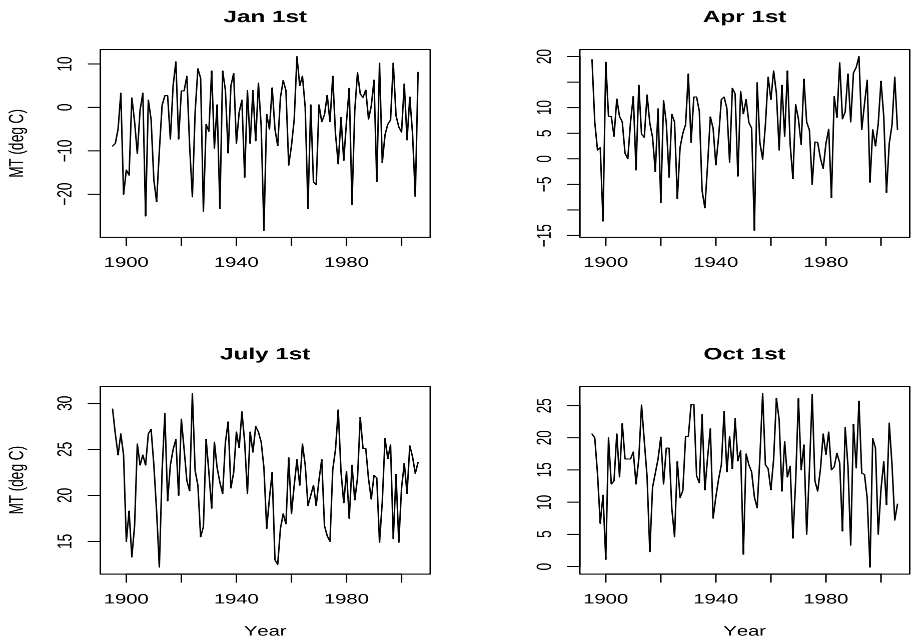

Figure 26: The time series over the years of daily maximum temperatures for Calgary for four given dates: January 1st, April 1st, July 1st and October 1st. Observe the appreciable amount of noise and lack of any distinct overall trends,

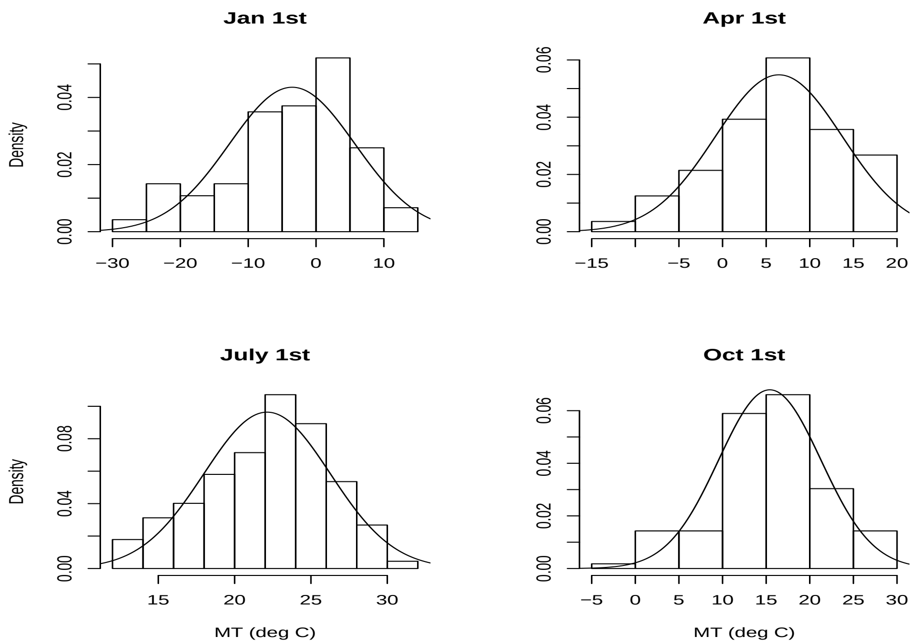

Figure 27: The histograms of daily maximum temperature for Calgary for four given dates over the years of our study: January 1st, April 1st, July 1st and October 1st. Notice the symmetry of the histograms as shown by comparison with best fitting normal distribution.

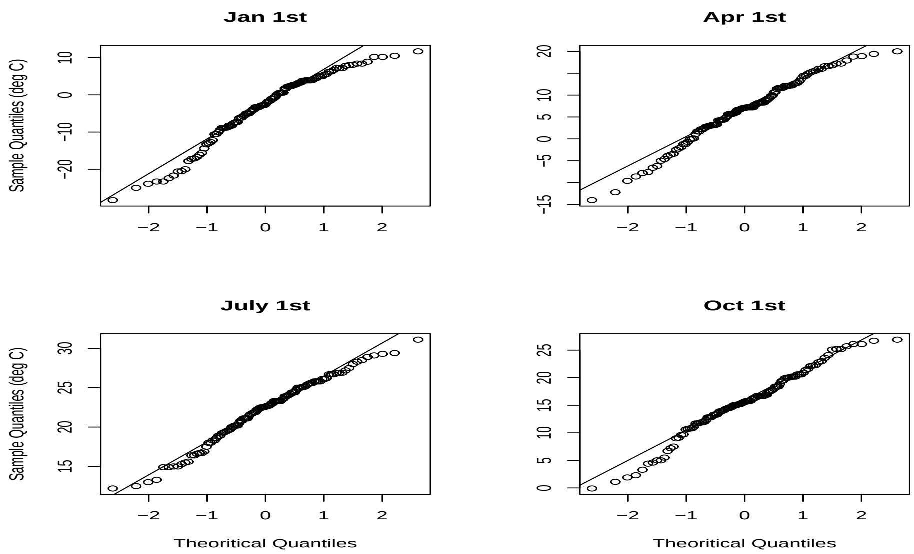

Figure 28: The normal qq-plots of for the data distribution shown in the previous figure, shows in some cases a modest departure from normality in the tails of the data distributions. Surprisingly, the all important right hand tails seem lighter than the best fitting normal tails: the 97.5$^{\text{th}}$ % ile of the empirical distribution is reached earlier than that of the corresponding normal, as one moves into the righthand tail.

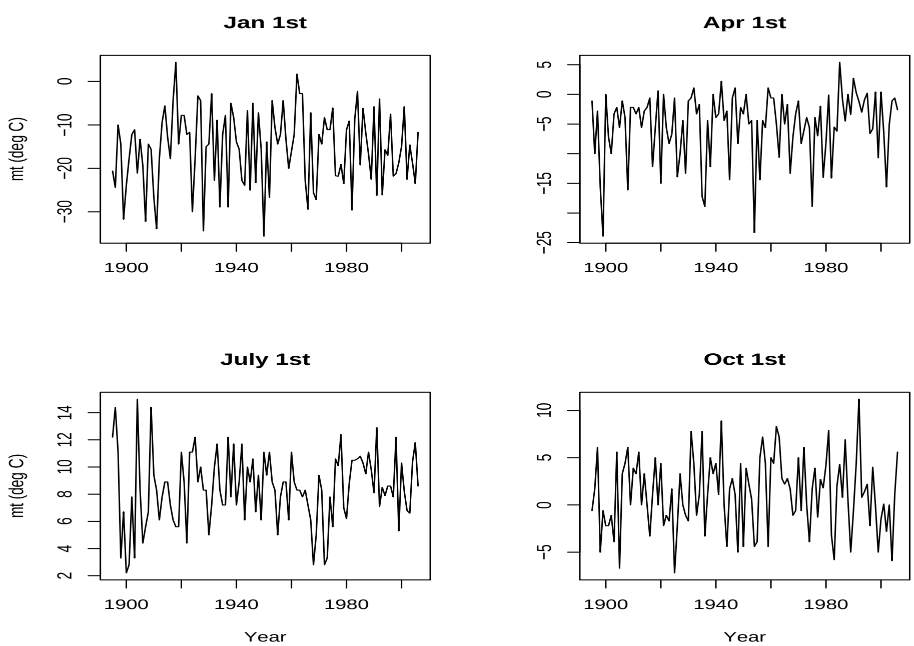

Figure 29: The time series of daily minimum temperature for Calgary for four given dates: January 1st, April 1st, July 1st and October 1st.

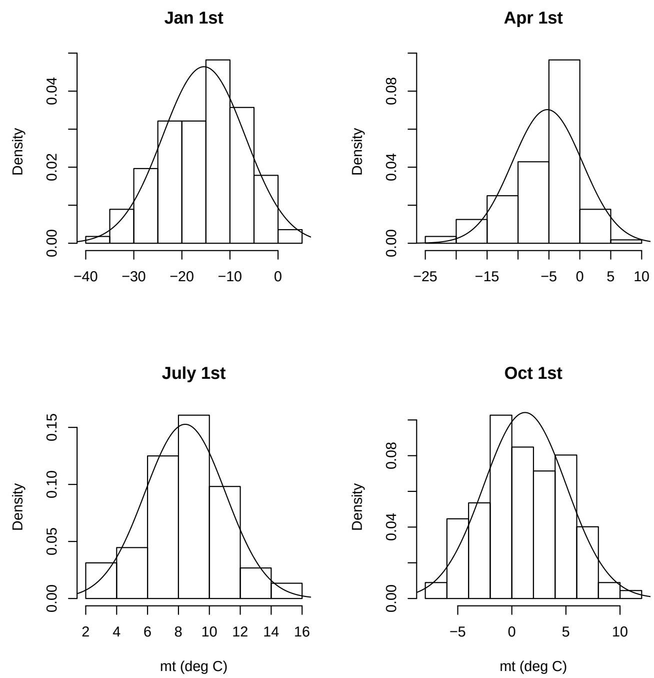

Figure 30: The histograms of daily minimum temperature for Calgary for four given dates: January 1st, April 1st, July 1st and October 1st.

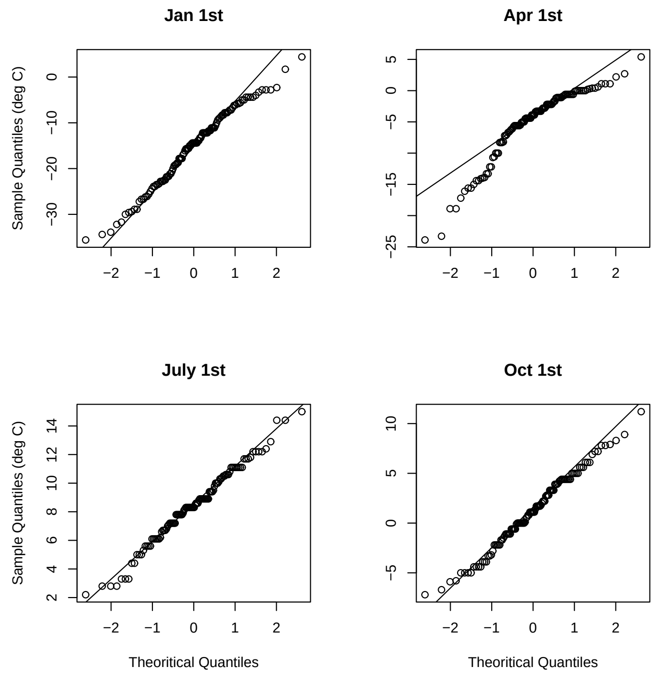

Figure 31: The normal qq-plots of daily minimum temperature for Calgary for four given dates: January 1st, April 1st, July 1st and October 1st.

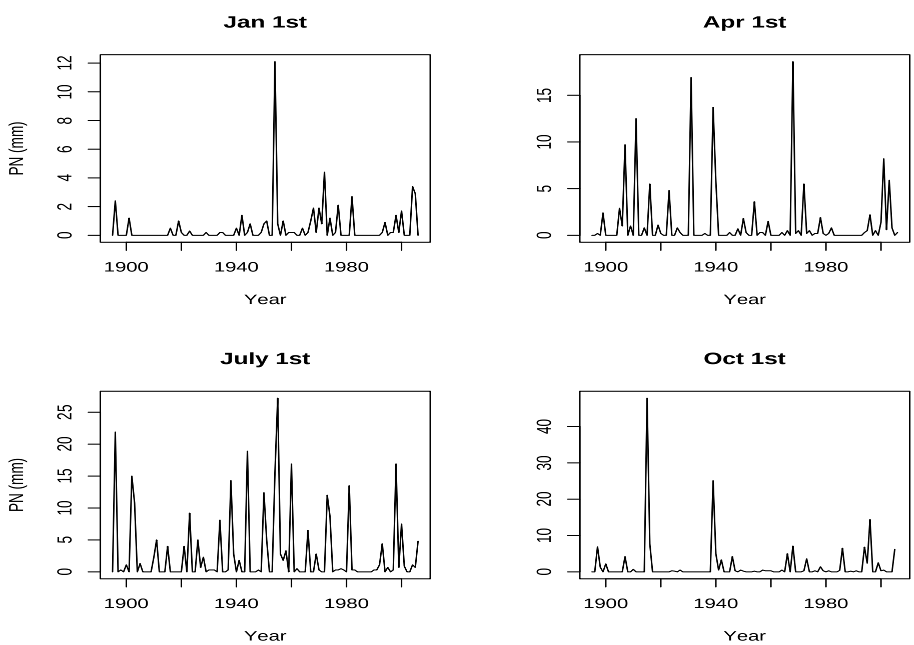

Figure 32: The time series of daily precipitation for Calgary for four given dates: January 1st, April 1st, July 1st and October 1st.

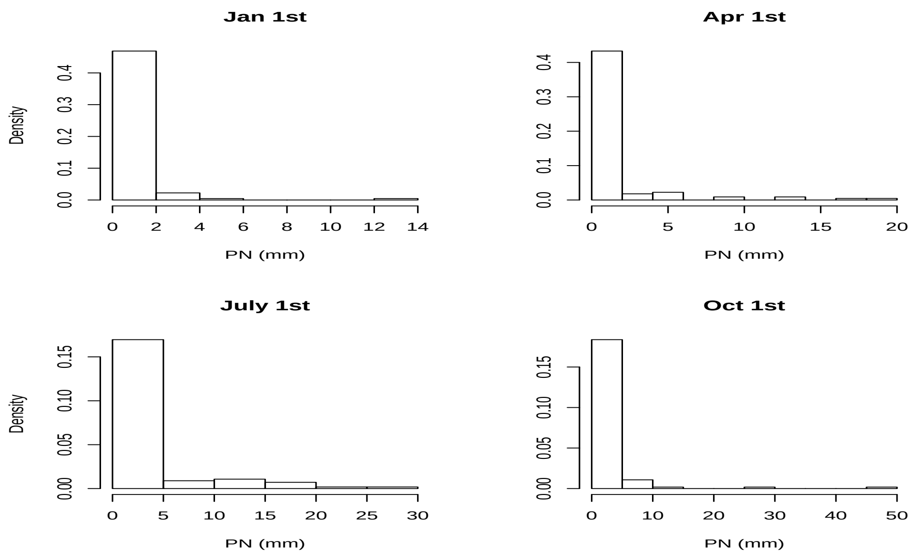

Figure 33: The histograms of daily precipitation for Calgary for four given dates: January 1st, April 1st, July 1st and October 1st.

Next we use the available years to compute the confidence intervals for the mean of every given day of the year for MT and mt. For PCPN, we construct the confidence intervals for the probability of a PCPN day, one with PCPN> 0.2 (mm). This threshold is justified in Part I (Hosseini et al. 2009) and amounts below are negligible. Figures 34 to 36 give the confidence intervals for the means. The confidence interval for the standard deviations (obtained by bootstrap techniques) are given in Figures 37 to 39. A regular seasonal pattern is seen. For example the maxima for MT and mt happen around the 200th day of the year (in July) and the minima, at the beginning of the year. Comparing the plots of means and standard deviations, we observe that warmer days have smaller standard deviation than colder days. For example the minimum standard deviations for the maximum and minimum temperatures happen around the 200th day of the year which correspond to the warmest period of the year. The plots also show that a simple periodic function seems to suffice for modeling the seasonal patterns. Unlike MT and mt, for the 0-1 PCPN process, the standard deviation is the highest in June, when the probability of precipitation is close to $\frac{1}{2}$.

As shown above, the estimated distribution of daily PCPN values is far from normal. Thus this time, after removing the zeros, we fit a Gamma distribution to PCPN (Figure 42). The Gamma qq-plots are given in Figure 43 and reveal a good fit.

Figure 34: The confidence intervals for the daily mean maximum temperature (deg C). Dashed line shows the upper bound and the solid line the lower bound of the confidence intervals.

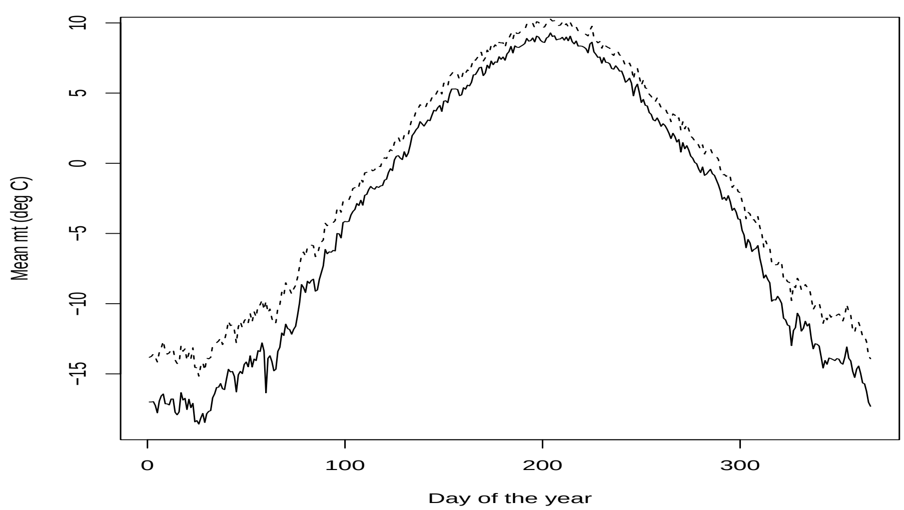

Figure 35: The confidence intervals for the daily mean mt (deg C). The dashed line shows the upper bound and the solid line, the lower bound of the confidence intervals.

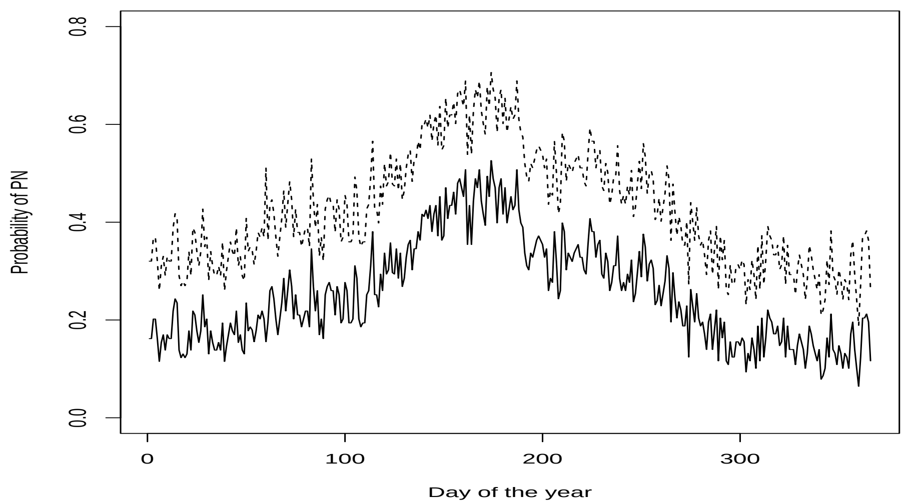

Figure 36: The confidence intervals for the estimated probability of PCPN (mm) for the days of the year. The dashed line shows the upper bound and the solid line, the lower bound of the confidence intervals.

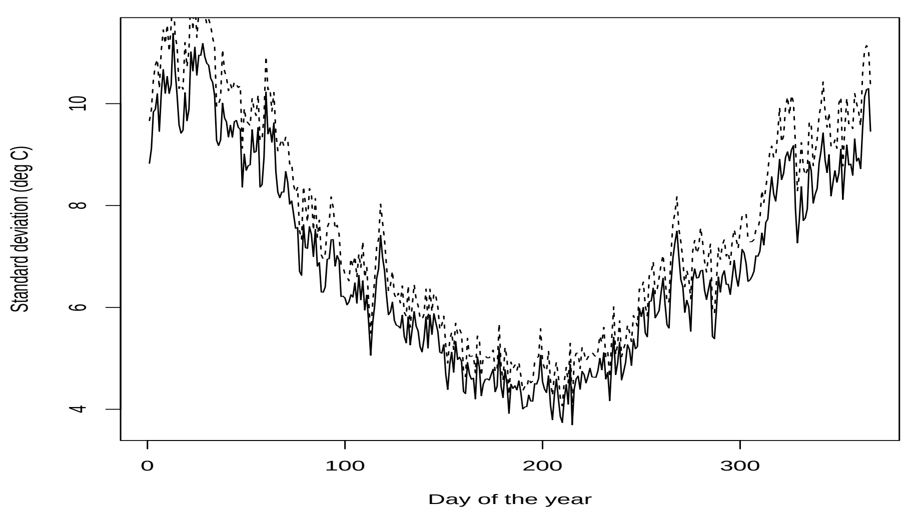

Figure 37: The confidence intervals for the standard deviation of each day of the year for the maximum daily temperature (deg C). The dashed line shows the upper bound and the solid line, the lower bound of the confidence intervals.

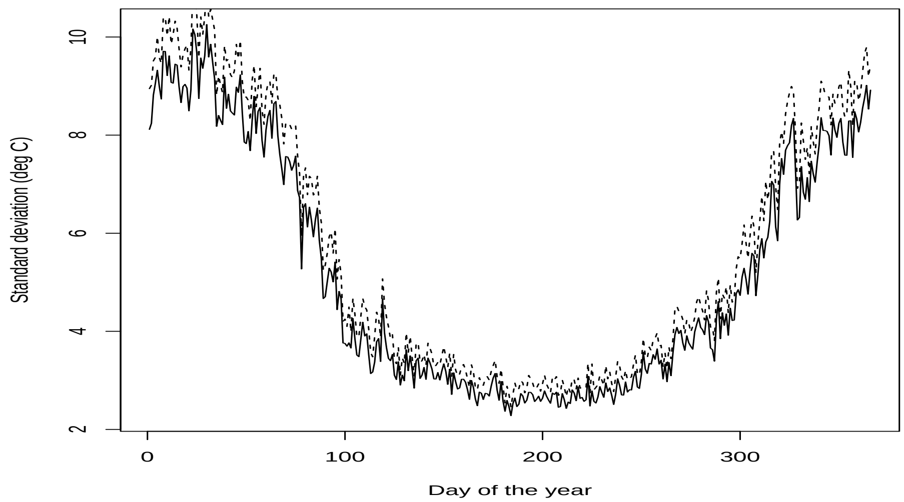

Figure 38: The confidence intervals for the standard deviation of each day of the year for minimum daily temperature (deg C). The dashed line shows the upper bound and the solid line, the lower bound of the confidence intervals.

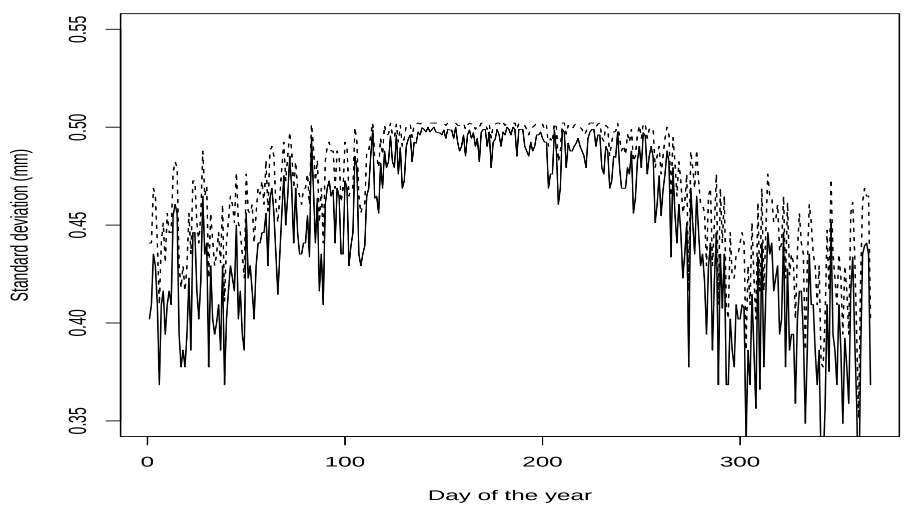

Figure 39: The confidence intervals for standard deviation of each day of the year for the probability of precipitation (mm) (for the 0-1 PCPN process). The dashed line shows the upper bound and the solid line, the lower bound of the confidence intervals. The plot shows $sd \leq 1/2$, the maximum value of $sd = \sqrt{p(1-p)}$.

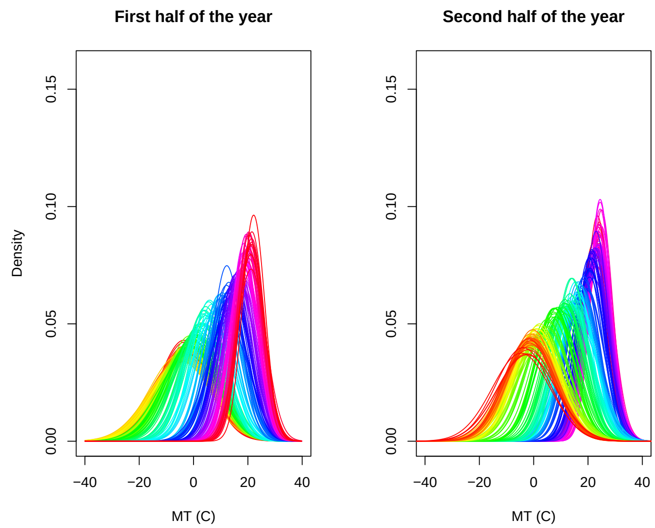

Figure 40: The estimated distribution of MT (C) for each day of the year from Jan 1st to Dec 1st. The year has been divided to two halves. In each half rainbow colors are used to show the change of the distribution.

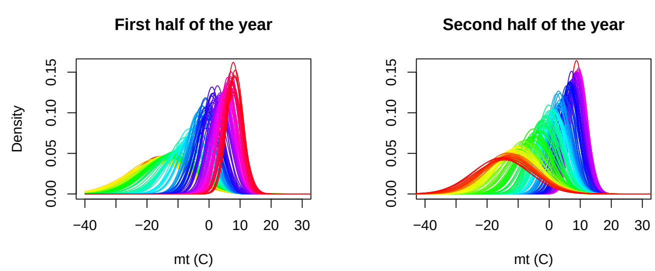

Figure 41: The estimated distribution of mt (C) for each day of the year from Jan 1st to Dec 1st. The year has been divided to two halves. In each half rainbow colors are used to show the change of the distribution.

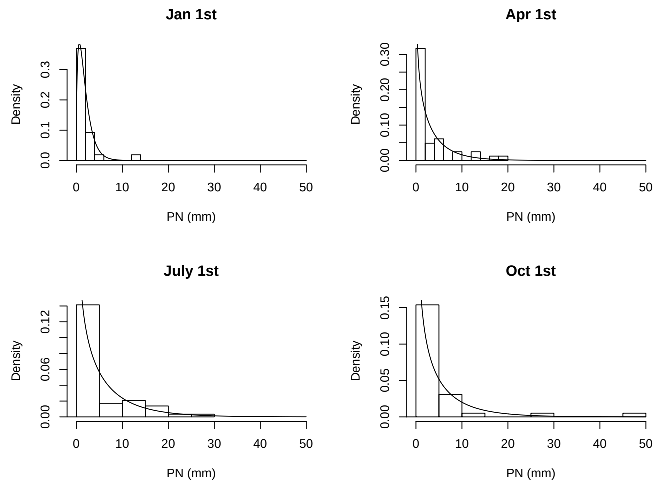

Figure 42: The histograms of daily precipitation greater than 0.2 mm for Calgary on which is superimposed, the best fitting gamma density curve obtained using the maximum likelihood method.

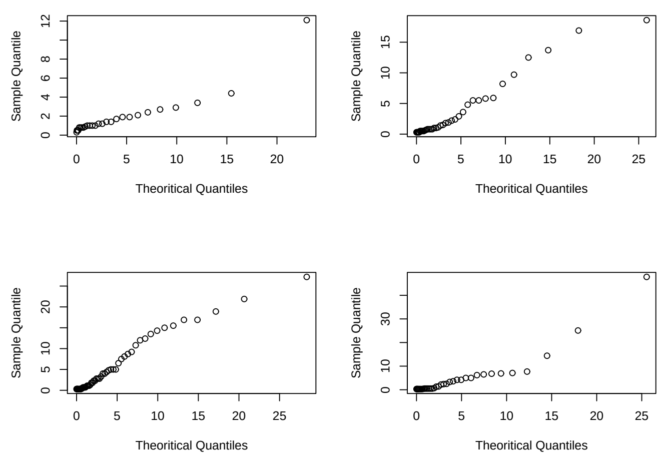

Figure 43: The qq-plots of daily precipitation greater than 0.2 mm for Calgary with the gamma distribution fitted using the maximum likelihood method.

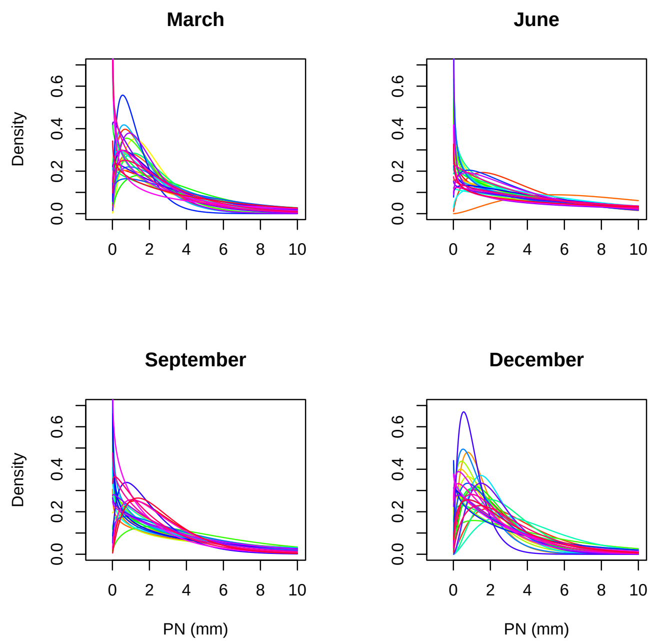

Figure 44: The gamma fit to the empirical distribution of precipitation (mm) for each day over a 4 month period. In each month rainbow colors are used to show the change of the distribution.

Figures 40, 41 and 44 reveal the result of our investigation of the change in the distribution over a period of time. For MT and mt, we have done that for the course of the year. The figures show how the distribution deforms continuously over the year. We can also notice changes in the mean and standard deviation over the year. For PCPN, we have done the same only for 4 different months because of high irregularity in the process.

Next we look at the parameters of the Gamma distribution fitted to the PCPN distribution over the course of a year. If we use maximum likelihood (ML) parameter estimates, as we have done to select the gamma density curve, the confidence intervals obtained by bootstrap method will be very wide (and tend rapidly to infinity). Hence, we use the simpler method of moments (MOM) estimates to obtain confidence intervals. The MOM confidence intervals are given in Figure 46. Since there is no closed form for the ML estimates, we need to use Newton's method to compute the required maxima. However MOM gives us closed form solutions. This advantage might explain the better behavior of MOM estimates in forming the confidence intervals. However even the MOM confidence intervals do not look satisfactory and are rather wide and irregular specially at the beginning and end of the year.

We can also consider the 0-1 process of PCPN (1 for wet and 0 for dry)

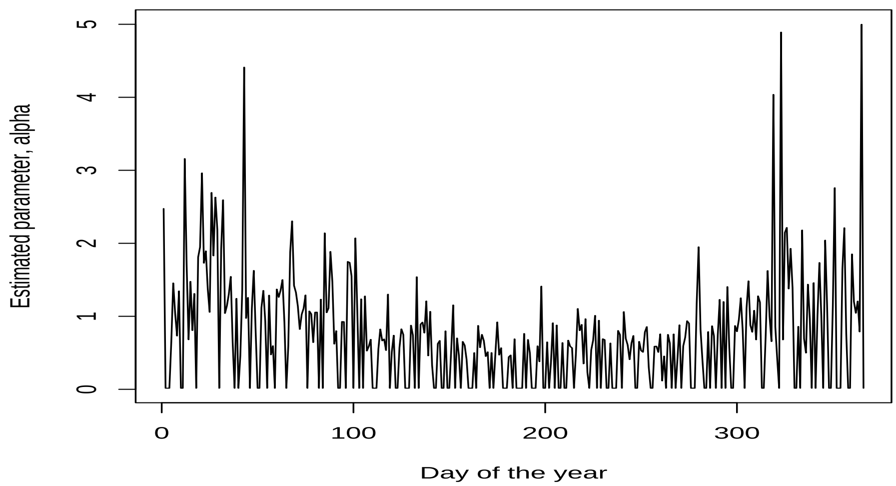

Figure 45: The maximum likelihood estimates for the gamma distribution's shape parameter.

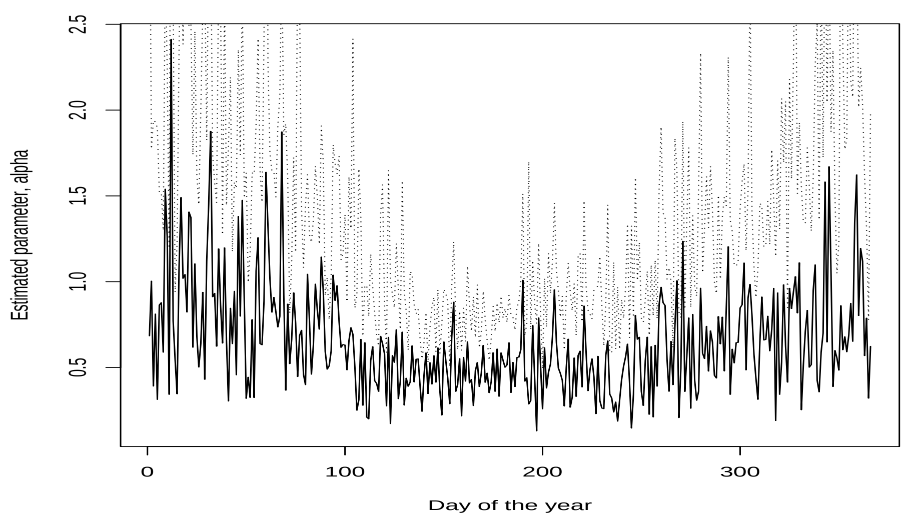

Figure 46: The confidence interval for the method of moments estimate for the gamma distribution's shape parameter. The dotted line is the upper bound and the solid line the lower bound.

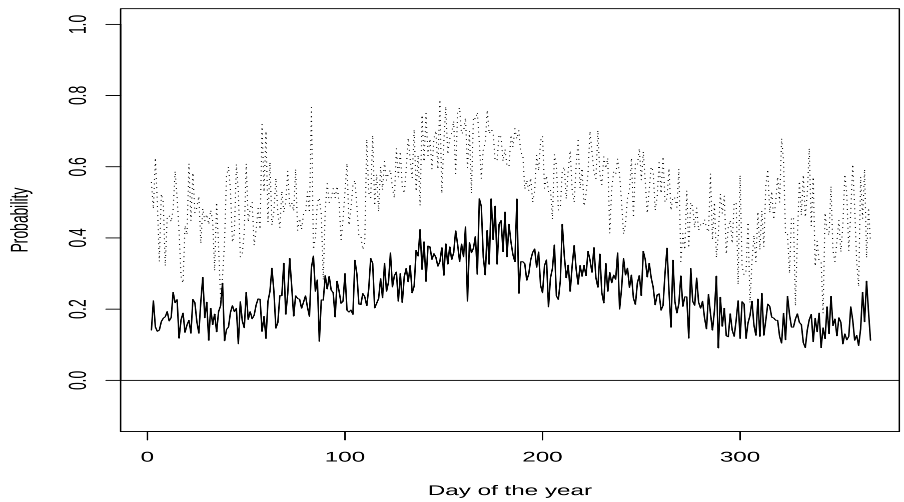

Figure 47: The Markov chain transition probabilities when the PCPN process is viewed as a binary sequence of precip–noprecip days (1 and 0 respectively). The dotted line is the "PCPN if PCPN=1" ($\hat{p_{11}}$) and the dashed is "PCPN if PCPN=0" ($\hat{p_{01}}$) on the previous day.

and compute the transition probabilities for PCPN (Figure cal-PN-transition). The figure shows the probability of PCPN is changing continuously over the year and can be parameterized by a simple periodic function.

Considering the 0-1 process induced by PCPN as a Markov chain leads to the interesting question as its order. In search of an answer, denote by 1 and 0 respectively, a PCPN occurrence and nonoccurrence. Also let $p_{x_{t-r}\cdots x_t}(t)$, where $x_i = 1, 0$, denote the probability of observing x_t on day t of the year conditional on the chain $(x_{t-r} \cdots x_{t-1})$. In Figure 47, we have plotted the estimated $\hat{p}_{11}(t)$ and $\hat{p}_{01}(t)$ for different days of the year. The clear gap between these two estimated probabilities indicate that a first order Markov chain should be preferred to a 0^{th}–order chain. Figures 48 and 49 show $\hat{p}_{111}$ plotted against $\hat{p}_{011}$ and $\hat{p}_{001}$ against $\hat{p}_{101}$. The estimated probabilities seem to be close and overlap heavily over the course of the year. Hence a first order Markov seems to suffice for describing the binary process induced by PCPN.

5 Correlation

The correlation in a spatial-temporal process can depend on time and space. In this section, we study the temporal and spatial patterns of the correlation function separately.

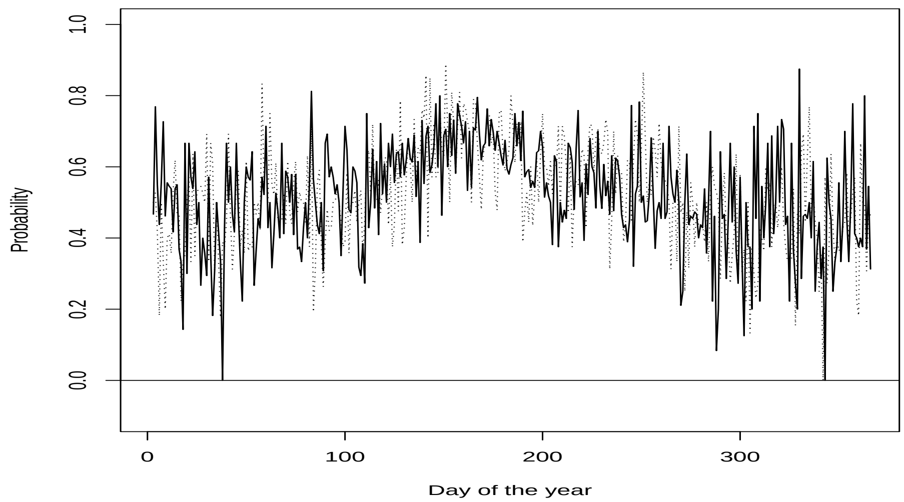

Figure 48: This figure plots the 3rd order transition probabilities, $\hat{p}_{111}$ (solid) against $\hat{p}_{011}$ (dotted), for the binary process induced by the PCPN process.

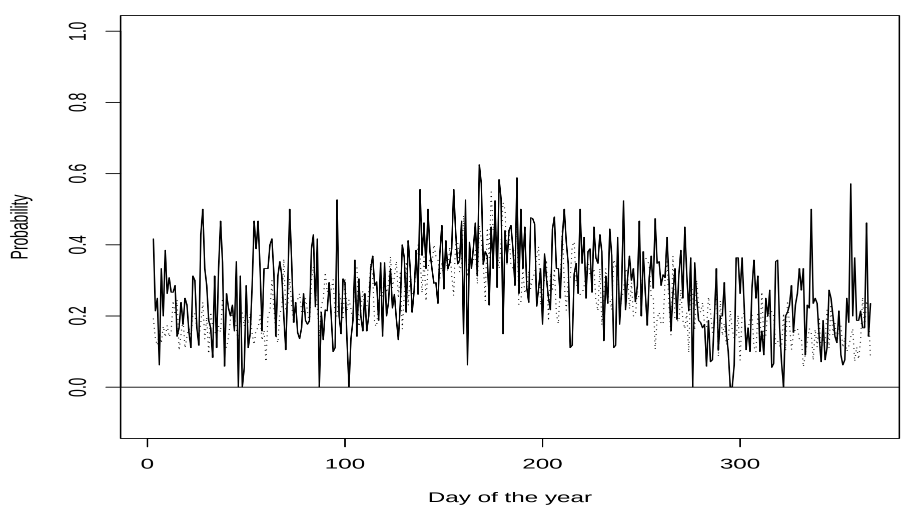

Figure 49: This figure plots the 3rd order transition probabilities, $\hat{p}_{001}$ (solid) against $\hat{p}_{101}$ (dotted), for the binary process induced by the PCPN process.

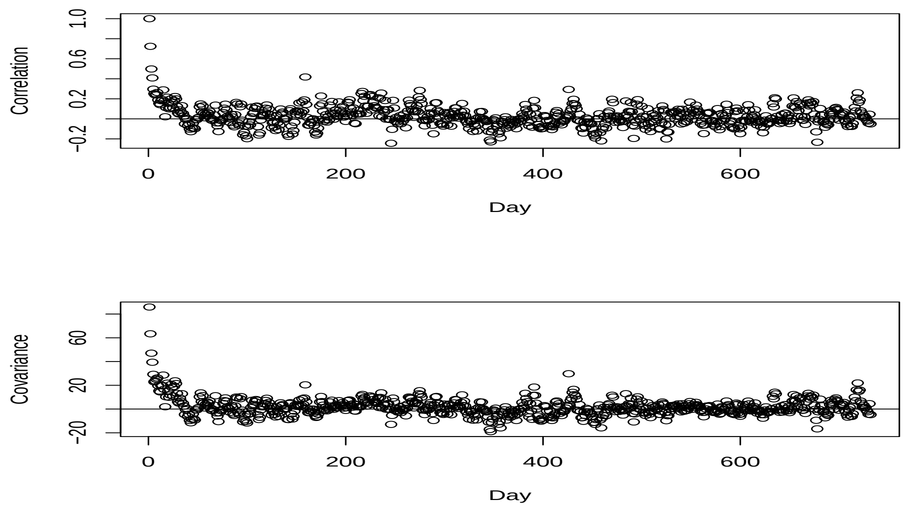

Figure 50: The correlation and covariance plots for the maximum daily temperature (deg C) at the Calgary site on Jan 1st and that on each of the 732 subsequent days.

5.1 Temporal correlation

Here we look at the correlation/covariance of the variables as a function of time. The location is taken to be the Calgary site for reasons given above in an earlier analysis in this report. First we look at the correlation/covariance of a given day and its subsequent days, starting with Jan 1st and computing its correlation/covariance with the following days: Jan 2nd, Jan 3rd and so on. Figures 51 to 53 show the results, in particular a decreasing trend over time. However the increase is far from linear and in fact it looks to be exponentially decreasing. The plots also show that only a few subsequent days are possibly correlated and in particular two days that are one year apart can be considered independent, a refined version of the conclusion reached in Part I (Hosseini et al. 2009). This conclusion will be used in building a spatial-temporal model.

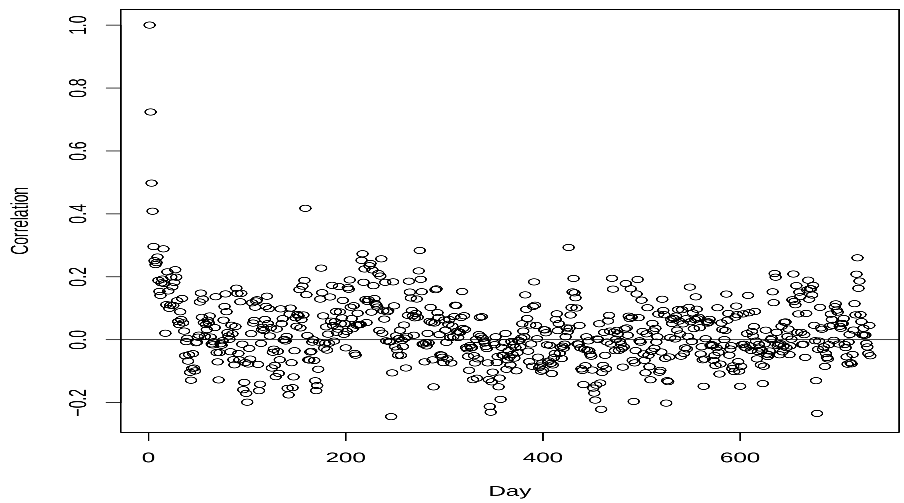

Figure 51: The correlation plot for the maximum daily temperature (deg C) at the Calgary site on Jan 1st and that for each of the 732 subsequent days.

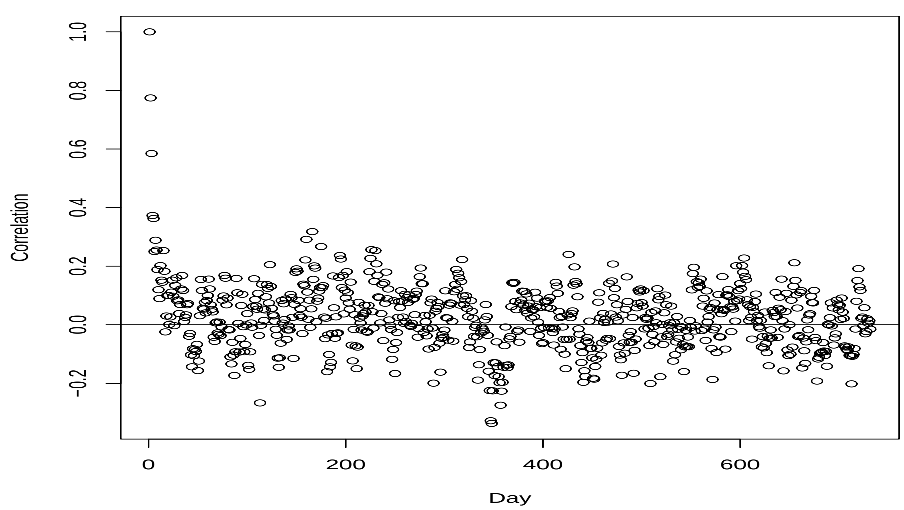

Figure 52: The correlation plot for minimum temperature (deg C) at the Calgary site on Jan 1st and that for each of the 732 subsequent days.

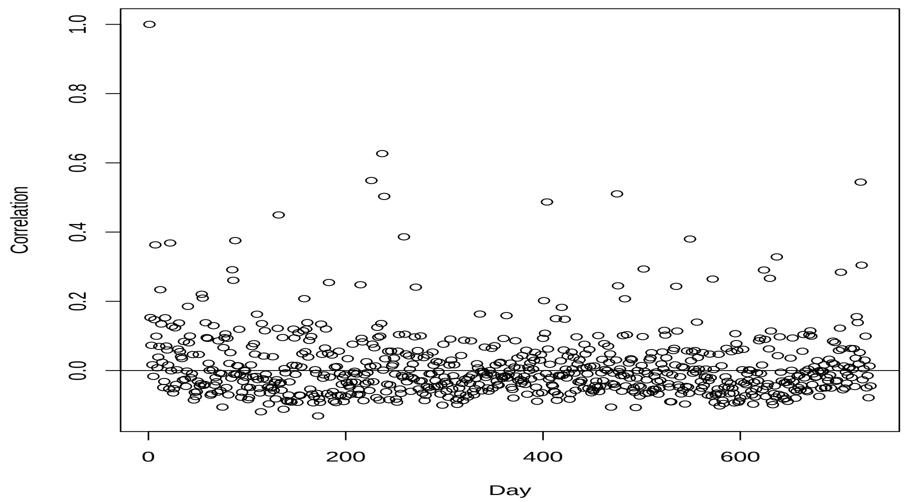

Figure 53: The correlation plot for precipitation (mm) at the Calgary site on Jan 1st and that for each of the 732 subsequent days.

Next we look at the correlation of responses on other days of the year with that on their 30 consecutive days. Our goal: to see if the correlation function has the same behavior over the course of a year. We pick, Feb 1st, April 1st, July 1st, Oct 1st as our start days for this analysis. Figures 54 and 56 show similar patterns.

Finally, we look at the correlation of two fixed locations over the course of the year (by changing the day). The results are given in Figures 57 and 58. Strong correlation and clear seasonal patterns are seen for MT and mt. This seems to indicate in particular that the temperature process is not stationary. The correlation in the middle of the year around day 200 which correspond to the summer season seems to be smaller than the correlation at the beginning and end of the year which corresponds to the cold season.

5.2 Spatial correlation

This subsection looks at spatial correlation by fixing a few dates: January 1st, April 1st, July 1st and Oct 1st distributed over the year's climate regimes. We plot the correlation with respect to the geodesic distance (km) on the surface of the earth. Figures 59 to 62 show the results for MT, mt, PCPN and the 0-1 PCPN binary process respectively. For MT and mt, we observe a clear decreasing trend with respect to distance. The trend for $PCPN$ does not seem to be regular.

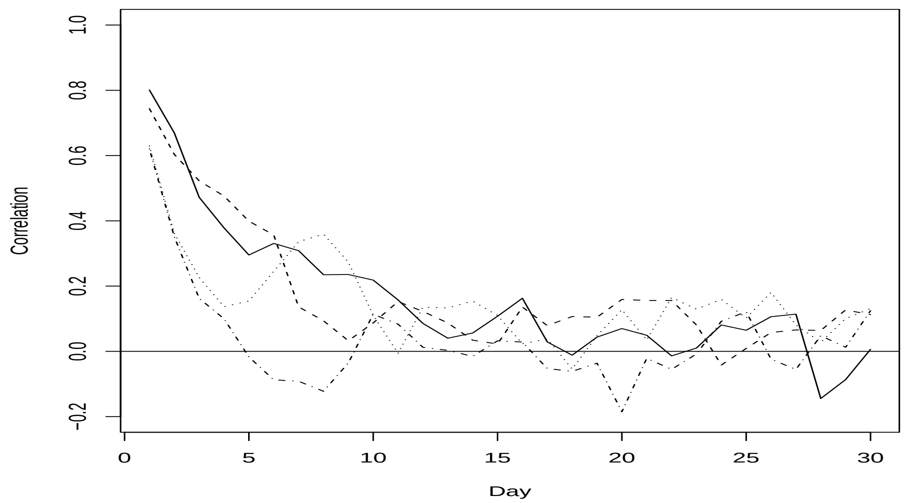

Figure 54: The correlation plot for maximum temperature (deg C) over Calgary for Feb 1st (solid), April 1st (dashed), July 1st (dotted) and Oct 1st (dot dash) and 30 consequent days.

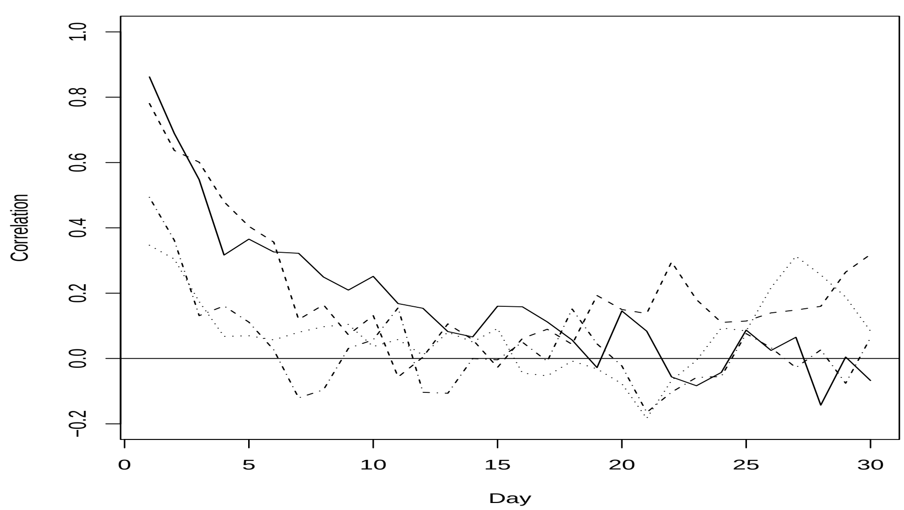

Figure 55: The correlation plot for minimum temperature (deg C) over Calgary for Feb 1st (solid), April 1st (dashed), July 1st (dotted) and Oct 1st (dot dash) and 30 consequent days.

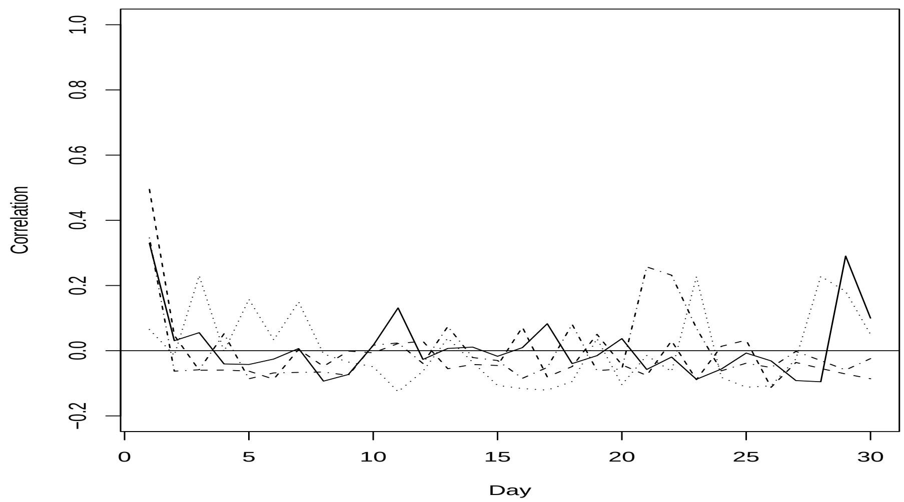

Figure 56: The correlation plot for precipitation (mm) over Calgary for Feb 1st (solid), April 1st (dashed), July 1st (dotted) and Oct 1st (dot dashed) and 30 consequent days.

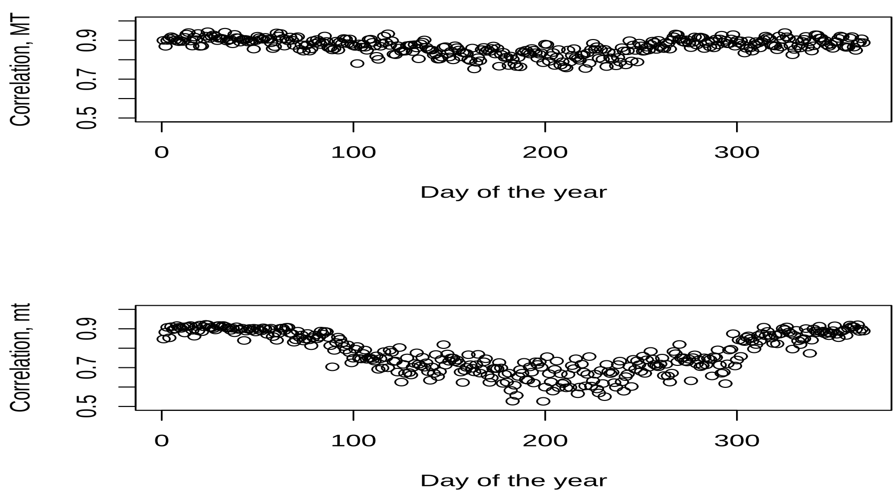

Figure 57: The correlation plot for maximum temperature and minimum temperature (deg C) between Calgary and Medicine Hat.

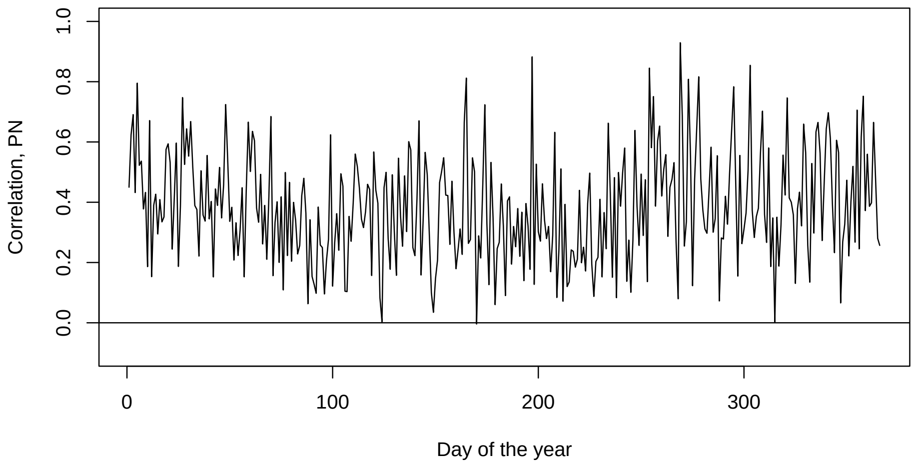

Figure 58: The correlation plot for precipitation (mm) between Calgary and Medicine Hat.

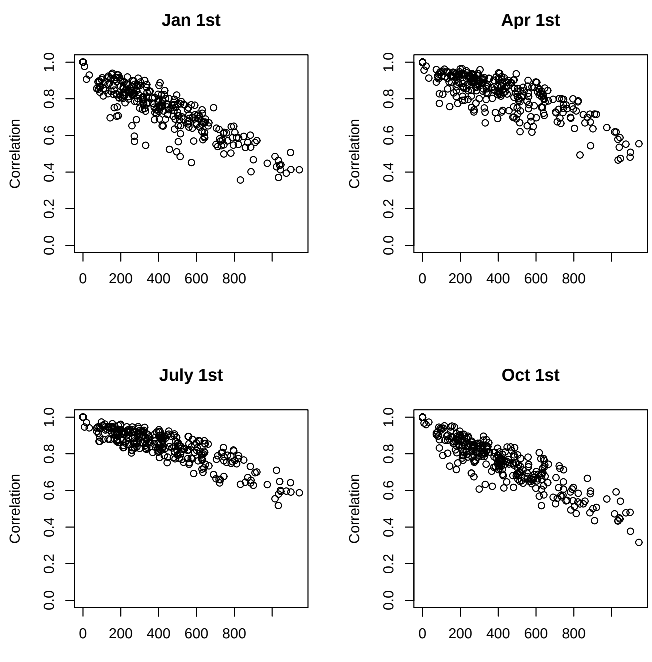

Figure 59: The correlation plot for maximum temperature (deg C) with respect to distance (km).

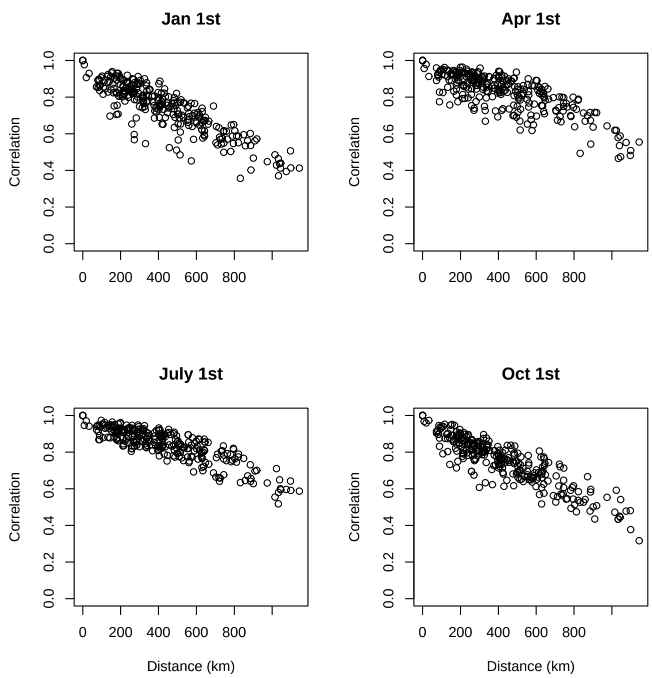

Figure 60: The correlation plot for minimum temperature (deg C) with respect to distance (km).

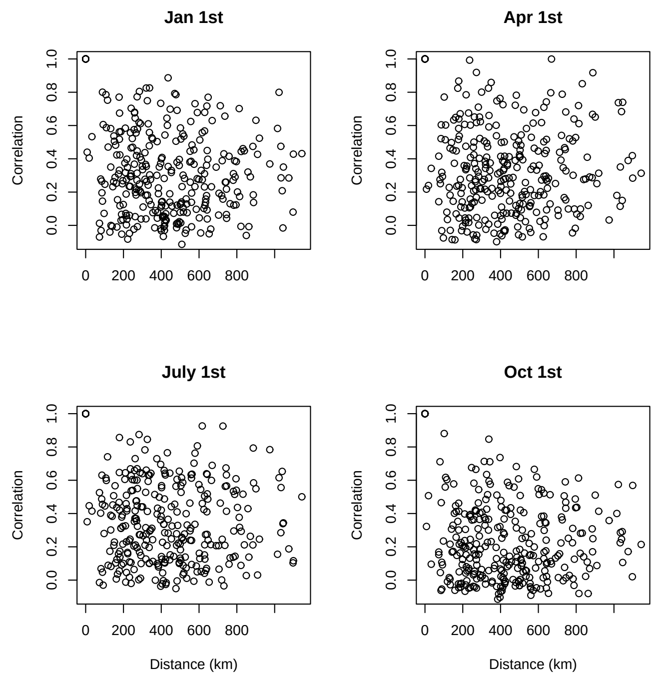

Figure 61: The correlation plot for precipitation (mm) with respect to distance (km).

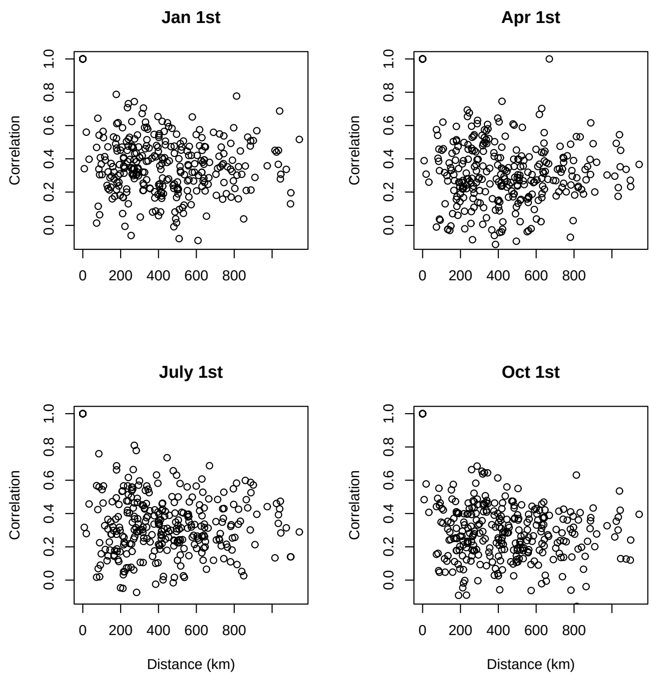

Figure 62: The correlation plot for precipitation (mm) 0-1 process with respect to distance (km).

6 Lessons learned and concluding remarks

In this report we performed some analysis on the homogenized data. Some of that was a repetition of the analysis done on the non-homogenized data [1]. The following conclusions confirm the results of the analysis on the non-homogenized data. It also includes some new conclusions based on the analysis on the homogenized data:

- There is a strong seasonal trend in temperature and precipitation processes. See Figures 7, 8, 11 and 36.
- The summer average min temperature has increased over several locations over the past century. See Figure 25.
- mt and MT highly correlated. See Figure 23.
- The distribution of daily maximum temperature and minimum temperature is close to Gaussian with some deviations seen on the tails. See Figures 27 and 29.
- The temperature process in Alberta is less variable in the warm seasons and the converse holds for the precipitation process. See Figures 37, 38 and 39.
- The distribution of the daily temperature varies continuously over the course of the year. This could not be shown for precipitation. (It might be because we need more data.)
- The correlation between two sites depends on the time of the year. They are more correlated in cold seasons. This might be because there are more global weather regimes in the cold seasons influencing the whole region.
- The correlation over time for MT, mt and PCPN seems stationary and is decreasing with a nonlinear trend (exponentially) with the time difference.
- The spatial correlation for MT and mt is strong and is decreasing almost linearly with respect to the geodesic distance.
- The spatial correlation for PCPN is not strong. It might be because the sites are too faraway to capture the spatial correlation for PCPN.

These results should help researchers build spatial-temporal models to make inference about the climate.

References

[1] Hosseini, R., Le, N. D. and Zidek, J.V. (2009). An Analysis of Alberta's climate. Part I: Nonhomogenized data. TR 245, Department of Statistics, University of British Columbia.

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

[3] Mekis, E. and Hogg, W. D. (1999) Rehabilitation And Analysis Of Canadian Daily Precipitation Time Series. Atmosphere-Ocean 37(1), pp. 53-85.

[4] Vincent, L.A., Zhang, x., Bonsal, B. R. and Hogg W.D. (2002) Homogenization of daily temperatures over Canada. Journal of Climate, 15, 1322-1334