TREND AND RETURN LEVEL OF EXTREME SNOW EVENTS IN NEW YORK CITY

by

Mintaek Lee

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in Mathematics

Boise State University

May 2017

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Mintaek Lee

Thesis Title: Trend and Return Level of Extreme Snow Events in New York City

Date of Final Oral Examination: 03 March 2017

The following individuals read and discussed the thesis submitted by student Mintaek Lee, and they evaluated his presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Jaechoul Lee, Ph.D.	Chair, Supervisory Committee
Partha S. Mukherjee, Ph.D.	Member, Supervisory Committee
Jodi L. Mead, Ph.D.	Member, Supervisory Committee

The final reading approval of the thesis was granted by Jaechoul Lee, Ph.D., Chair of the Supervisory Committee. The thesis was approved by the Graduate College.

나의 부모님께 이 논문을 헌정합니다

ACKNOWLEDGMENTS

I gratefully acknowledge the support and generosity of the following people. First and foremost, I would like to express my sincere gratitude to my advisor and mentor, Dr. Jaechoul Lee, for his continuing support, patience, and immense knowledge. Without him, this thesis would not have been possible. I also would like to thank my committee members, Dr. Partha Sarathi Mukherjee and Dr. Jodi Mead, for their services and valuable feedback which greatly improved my thesis.

My gratitude extends to my family and friends, particularly my parents, Yong Son Lee and Kwanseon Yoo, and my sister, Suhyun Lee, for their continuing support, encouragement, and motivation. A very special acknowledgment goes to my wife, Mary Lee, for not only proofreading my thesis for grammatical errors, but also continuously supporting me throughout the program in various ways.

Finally, I would like to thank the Department of Mathematics and Graduate College at Boise State University for the Graduate Residential Scholars Program, Summer Graduate Research Fellowship, and Graduate Teaching Assistantship awards which have supported my studies over the last two years.

ABSTRACT

A major winter storm brought up to 42 inches of snow in parts of the Mid-Atlantic and Northeast United States for January 22−24, 2016. The blizzard of January 2016 impacted about 102.8 million people, where at least 55 people died due to the snowstorm and it caused economic losses in a range of $500 million to $3 billion. This thesis studies two important aspects of extreme snow events: maximum snowfall and maximum snow depth. We apply extreme value methods to extreme snowfall and snow depth data from the New York City area to examine if there are any significant linear trends in extreme snow events and understand how likely the winter storm was in terms of return levels. We find that 87.5-th percentile snowfall and 75-th percentile snow depth have increased by 0.564 inches and 0.559 inches decade^{-1}, respectively, whereas the annual maximum snowfall and snow depth series show insignificant increases. Our analysis shows that the 2016 blizzard was indeed an extreme snow event equivalent to about a 40-year return level in the New York City area. Our methods are thoroughly illustrated with details and expressions for practitioners wishing to use extreme value methods in applications.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Extreme weather and climate events greatly impact human beings, societies, and ecosystems. In the United States, there have been 203 weather and climate related disasters from 1980 to 2016, each with at least $1 billion in damages (NCEI, 2017). Among these extreme weather and climate events, extreme snow events can disastrously affect urban life in particular. Recently, the blizzard of January 2016 brought an all-time record snowfall of 26.8 inches in Manhattan, resulting in significant damages/losses to many urban areas and life in the northeastern United States (National Weather Service, 2016). The 2016 blizzard raises an important question: Are we experiencing disastrous snowstorms more often than before? In contrast, many authors have reported decreasing trends in snowfall in various regions. Kunkel et al. (2009) found that snowfall is strongly decreasing in the mid-Atlantic coast. Burakowski et al. (2008) obtained a decreasing trend of 1.81 inches decade^{-1} in average winter snowfall in the northeastern United States. Huntington et al. (2004) verified a statistically significant decreasing trend in the ratio of snow to total precipitation (S/P) in the northeastern United States, concluding that the decreasing trend in S/P is mostly related to decreasing snowfall. According to these findings, the 2016 blizzard appears to be arguably contradictory to a common perception of global climate change: increasing temperature reduces snowfall.

We consider less-studied – but very critical to our life – questions, such as: are extreme snow events changing or not? If extreme snow events change, are they decreasing or increasing? How many snow events will we experience in the next 25 or 50 years? One should not necessarily conjecture that extreme snow events would show the same pattern as average snow events. In fact, extreme and mean statistics are statistically independent in large samples under some mild conditions (McCormick and Qi, 2000). This implies that the analysis of extreme statistics can exhibit different results from that of mean statistics. Furthermore, O'Gorman (2014) showed that for most land regions in the northern hemisphere, changes in extreme snowfall were very little even though mean snowfall decreased significantly based on climate simulations under high carbon dioxide emission scenarios.

Extreme value analysis requires the use of extreme value methods. Since extreme values usually do not follow the Gaussian distribution, the use of Gaussian-based methods does not produce accurate results for extreme values. Fawcett and Walshaw (2007) used the generalized Pareto (GP) distribution to model extreme sea surge heights in Newlyn, U.K., and compared methods dealing with temporally dependent data. Northrop and Jonathan (2011) used a quantile regression technique to model non-stationary thresholds of their extreme value model for hurricane-induced wave heights in the Gulf of Mexico. Lee et al. (2014) used the generalized extreme value (GEV) distribution with non-stationary location parameters and a changepoint technique to quantify linear trends for maximum and minimum temperatures in the contiguous United States. Panagoulia et al. (2014) used GEV and model selection criteria to analyze extreme precipitation in a mountainous Mesochora catchment in Greece. Rust et al. (2011) used GEV and bootstrap to estimate confidence intervals of return levels for floods in the southern Germany. Fawcett and Walshaw (2012) used

extremal index estimation and bootstrap to obtain confidence intervals of return levels for sea surge heights in Newlyn, U.K., and wind speeds in Bradfield, U.K.

More specifically, researchers have used extreme value methods to analyze extreme snowfall. For example, Blanchet et al. (2009) used a Poisson point process representation to analyze extreme snowfall data collected from 1999 to 2008 in the Swiss Alpine region. Makkonen et al. (2007) used climate simulations and GEV to compute 50-year return levels for snowfall in the Nordic area. López-Moreno et al. (2011) used climate simulations and GP distribution to analyze extreme snowfall in the Pyrenean mountain range located on the border of Spain and France.

We aim to evaluate whether the blizzard of January 2016 in the New York City area can be explained by using appropriate extreme value models, or if it was too much of an extreme event to be explained by such models. For a more comprehensive understanding of extreme snow events, we study two important aspects, namely snowfall and snow depth. We estimate return levels for extreme snowfall and snow depth in the New York City area and quantify trends in these characteristics, if any noticeable trends are identified.

The rest of this thesis proceeds as follows: Chapter 2 describes the snowfall and snow depth data used in our analysis, Chapter 3 explains the extreme value theory and methods applied to the data, Chapter 4 summarizes our findings on extreme snow events, and in Chapter 5 we conclude with some further comments.

CHAPTER 2

DATA

We study maximum snowfall and maximum snow depth in the New York City area to gain deeper insight into extreme snow events. The snowfall and snow depth data in this study were downloaded from the website of the Northeast Regional Climate Center (NRCC) in Ithaca, New York. We selected four weather stations from the study area: Central Park, Newark, La Guardia, and JFK. Table 2.1 shows the geographical descriptions of these four weather stations, and Figure 2.1 displays their spatial locations. We used the most recent 56 years of the NRCC's snowfall and snow depth data of the four stations, starting from July 1, 1959 and ending on June 30, 2015.

Table 2.1: Selected weather stations in the New York City area

Station	Full station name	Latitude	Longitude	Elevation
Central Park	NEW YORK CNTRL PK TWR	40.7789°	-73.9692°	132 ft
Newark	NEWARK LIBERTY INTL AP	40.6825°	-74.1694°	29 ft
La Guardia	LA GUARDIA AP	40.7794°	-73.8803°	39 ft
JFK	JFK INTL AP	40.6386°	-73.7622°	11 ft

NRCC's snowfall data contain daily maximum snowfall observations, where the maximum snowfall for a day is defined as the maximum amount of new snow and ice that have accumulated prior to melting or settling (National Weather Service, 2013).

Figure 2.1: Spatial location of selected weather stations

NRCC's snowfall measurements are recorded in inches and rounded to the nearest tenth of an inch. The snowfall measurements of less than 0.1 inches are typically recorded as "trace." We treat these nearly zero snowfall amounts as zero inches, which results in 3.9% of daily snowfall data as meaningful snowfall records ($\geq$0.1 inches). Although the NRCC data contain daily observations, many extreme snowfall events tend to last more than one day, producing multiple daily snowfall observations from a single snowstorm event. For example, a snowstorm in 2010 brought Central Park 9.4 inches of snow on February 25 and 11.5 inches of snow on February 26, totaling 20.9 inches of snow for the two consecutive days. To properly assess the snowfall amount from each snowstorm, we merged daily non-zero snowfall observations within each continuous snow event. If consecutive snowfall observations of greater than zero inches are not separated by a snowfall record of zero inches, we consider them to be from the same snow event.

Another important characteristic of snow events is daily snow depth, which is the

amount of snow, sleet, or ice present on the ground at 7 a.m. as measured with a measuring stick (National Weather Service, 2013). NRCC's snow depth data are measured in inches and rounded to the nearest inch. If less than half of the exposed ground is covered by snow or the measured depth is less than 0.5 inches, snow depth is recorded as "trace." We convert these observations to zero inches, yielding 6.0% of daily snow depth observations as meaningful records ($\geq$1 inch). We note that the snow depth observation for a given day tends to be affected by the temperature of the previous day. Therefore, we use the maximum temperature on the previous day as a predictor for snow depth. The maximum temperature data for the four locations are recorded in degrees Fahrenheit and rounded to the nearest degree.

NRCC's snow data contain missing values, but its missing rates are very low. For snowfall data, La Guardia has no missing days, Central Park and Newark both have two missing days, and JFK has seven missing days. For snow depth data, La Guardia and Newark both have no missing days, and JFK has four missing days. We find that the Central Park station has 1248 missing days, with most of the missing days occurring between 1996 and 2003. These account for about 6% of the daily data for the study period of 1959–2016. We do not know the reason why there are this many missing observations in Central Park during this seven-year period. With this stated, we exclude all missing data from our analysis as the overall missing rates are minimal.

We define a "snow year" as July 1 to June 30. For instance, the snow records from July 1, 1999 to June 30, 2000 are treated as the observations in the snow year of 1999. The observations up to June 30, 2015 are first used in our analysis to determine if our extreme value model could adequately predict the 2016 blizzard. Results from partial data up to June 30, 2015 are later compared to the results from the full data up to June 30, 2016, which includes observations from the January 2016 blizzard.

CHAPTER 3

METHODS

3.1 Block Maxima Methods

Suppose $D_1, \ldots, D_m$ are independent and identically distributed random variables taken from a population with a common distribution function $F(\cdot)$. Define $M_m = \max\{D_1, \ldots, D_m\}$ to be the maximum statistic of these m random variables. If there are sequences of constants $\{a_m\}$ and $\{b_m\}$, with $b_m > 0$, that rescale M_m such that

$$P\left(\frac{M_m - a_m}{b_m} \leq x\right) \longrightarrow G(x) \quad \text{as } m \to \infty,$$

where $G(\cdot)$ is a non-degenerate distribution function, then $G(\cdot)$ belongs to one of the following extreme value distribution families (Fisher-Tippett-Gnedenko Theorem):

$$\text{(Gumbel)} \quad G(x) = \exp\left\{-\exp\left[-\left(\frac{x-a}{b}\right)\right]\right\}, \quad -\infty < x < \infty,$$

$$\text{(Fréchet)} \quad G(x) = \begin{cases} 0, & \text{if } x \leq a; \\ \exp\left\{-\left(\frac{x-a}{b}\right)^{-\alpha}\right\}, & \text{if } x > a, \end{cases}$$

$$\text{(Weibull)} \quad G(x) = \begin{cases} \exp\left\{-\left[-\left(\frac{x-a}{b}\right)^{\alpha}\right]\right\}, & \text{if } x < a; \\ 0, & \text{if } x \geq a \end{cases}$$

for some constants a, $b > 0$, and $\alpha > 0$. These extreme value distribution families can be further unified into the following generalized extreme value (GEV) distribution (cf. Coles, 2001):

$$G(x) = \exp\left\{ -\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi} \right\}, \tag{3.1}$$

where $z_+ = \max\{z, 0\}$, $-\infty < \mu < \infty$, $\sigma > 0$, and $-\infty < \xi < \infty$. The unknown constants μ, σ, and ξ are called location, scale, and shape parameters, respectively.

Block maxima methods use a sequence of these maximum statistics. To elaborate, we express data $\{D_1, \ldots, D_N\}$ as a set of n blocks of block size m:

$$\left\{(D_1, \ldots, D_m), (D_{m+1}, \ldots, D_{2m}), \ldots, (D_{(n-1)m+1}, \ldots, D_{nm})\right\},$$

where $N = nm$. We then take a maximum statistic within each block and denote the maximum statistic as $M_{m,i}$ for i-th block. The previous extreme value theorem implies that if block size m is large enough, the GEV distribution can be an approximate probability distribution for the maximum statistics $\{M_{m,1}, \ldots, M_{m,n}\}$ *regardless of* the distribution function $F(\cdot)$ from which the original sample $\{D_1, \ldots, D_{nm}\}$ is taken. In practice, the maximum statistic $M_{m,i}$ is taken as the maximum of data values recorded for a meaningful period, most often years. The maximum likelihood estimates of GEV parameters are found by maximizing the log likelihood function:

$$\ell(\mu, \sigma, \xi) = \begin{cases} -n\log(\sigma) - \sum_{i=1}^n \left(1 + \xi\left(\frac{x_i - \mu}{\sigma}\right)\right)^{-1/\xi} \\ \quad - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \log\left(1 + \xi\left(\frac{x_i - \mu}{\sigma}\right)\right) \end{cases}, \quad \text{if } \xi \neq 0; \\ -n\log(\sigma) - \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right) \\ \quad - \sum_{i=1}^n \exp\left(-\left(\frac{x_i - \mu}{\sigma}\right)\right), \qquad \text{if } \xi = 0, \end{cases} \tag{3.2}$$

using numerical methods.

Return levels are crucial quantities in many extreme value studies. The return level associated with the return period of K years, denoted by x_K, is the expected level that is to be exceeded on average once over following K years. For annual maximum data, the K-year return level is expressed as

$$x_K = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - \{-\ln(1-K^{-1})\}^{-\xi}\right], & \text{if } \xi \neq 0; \\ \mu - \sigma \ln\{-\ln(1-K^{-1})\}, & \text{if } \xi = 0. \end{cases} \tag{3.3}$$

Substituting the GEV parameters (μ, σ, ξ) in (3.3) with their maximum likelihood estimates finds the maximum likelihood estimate for the return level x_K.

3.2 Threshold Exceedances Methods

Suppose $X_1, \ldots, X_N$ are independent and identically distributed random variables with distribution function $F(\cdot)$. Let X be any of these X_i's, and assume $F(\cdot)$ satisfies the Fisher-Tippett-Gnedenko Theorem. For a predetermined high threshold u, one considers the exceedance over u, denoted by $Y = X - u$. It is known from the extreme value theory that, for large enough u, if the exceedance Y is positive ($Y > 0$, or equivalently $X > u$), the conditional distribution function of Y can be approximated by the generalized Pareto (GP) distribution:

$$H(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma^*}\right)^{-1/\xi}, & \text{if } \xi \neq 0; \\ 1 - \exp\left(-\frac{y}{\sigma^*}\right), & \text{if } \xi = 0, \end{cases}$$

where $y > 0$, $1 + \xi y/\sigma^* > 0$, and $\sigma^* = \sigma + \xi(u - \mu)$.

Threshold exceedances methods use all observations above the threshold u. To

elaborate, suppose n data values are greater than u and denote them by $X_1^*, \ldots, X_n^*$ (that is, $X_i^* > u$). Then, consider the exceedances by $Y_i = X_i^* - u$. Extreme value theory implies that, if u is large enough, the GP distribution can be an approximate probability distribution for the exceedances $\{Y_1, \ldots, Y_n\}$ *regardless of* $F(\cdot)$ from which the original sample $\{X_1, \ldots, X_N\}$ is taken.

Choice of an adequate threshold is very important in the threshold exceedances methods. A threshold that is too low may lead to bias, as it is likely to violate the asymptotic basis of the model (Coles, 2001). Conversely a threshold that is too high would result in high variance, as there would be fewer observations that can be used in the model estimation process. A common practice is to find the lowest threshold for which the GP model provides an appropriate fitting. Mean residual life (MRL) plots or parameter stability plots are commonly used in practice to visually find such a threshold.

The MRL plot is constructed by plotting all possible thresholds against the mean of exceedances from the corresponding threshold. To elaborate, let $u_1 < \cdots < u_j < \max\{X_1, \ldots, X_N\}$ be all possible thresholds for X_i. The MRL plot will consist of points

$$\left(u_i, \frac{1}{n_{(i)}} \sum_{k=1}^{n_{(i)}} Y_k^{(i)} \right),$$

where u_i is a threshold for $i = 1, \ldots, j$, $X_1^{*(i)}, \ldots, X_{n_{(i)}}^{*(i)}$ are $n_{(i)}$ data values greater than u_i, and $Y_k^{(i)} = X_k^{*(i)} - u_i$ for $k = 1, \ldots, n_{(i)}$ are exceedances from u_i. The parameter stability plot is constructed by plotting all possible thresholds against the maximum likelihood estimate of σ^* and ξ obtained from $Y_1^{(i)}, \ldots, Y_{n_{(i)}}^{(i)}$ for $i = 1, \ldots, j$. For the thresholds $u^* < \ldots < u_j$ on which the GP model is appropriate, the MRL plot should be approximately linear and the parameter stability plot should be stable

(Coles, 2001). Such threshold u^* is usually considered as an adequate choice for the threshold of the given data.

Threshold exceedances methods are different from block maxima methods in that threshold methods use all observations above a certain threshold, instead of using only the maximum statistics within preset blocks. Because a larger number of extreme data is often used, threshold exceedances methods can perform better than block maxima methods. This is very important for extreme value studies due to the limited frequency of extreme events in many practices. The two GP parameters, σ^* and ξ, are often estimated by maximizing the log likelihood function:

$$\ell(\sigma^*, \xi) = \begin{cases} -n\log(\sigma^*) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{n} \log\left(1 + \xi\left(\frac{x_i - \mu}{\sigma^*}\right)\right), & \text{if } \xi \neq 0; \\ -n\log(\sigma^*) - \frac{1}{\sigma^*} \sum_{i=1}^{n} x_i, & \text{if } \xi = 0, \end{cases} \tag{3.4}$$

using numerical methods.

Denote m as the number of observations per year and $p_u^* = P(X > u)$ as the over-threshold probability for a threshold u. Based on the GP distribution, the return level x_K is expressed as

$$x_K = \begin{cases} u + \frac{\sigma^*}{\xi}[(Kmp_u^*)^{\xi} - 1], & \text{if } \xi \neq 0; \\ u + \sigma^* \ln(Kmp_u^*), & \text{if } \xi = 0, \end{cases}$$

if K is large enough such that $x_K > u$. The probability p_u^* is often estimated by an empirical probability $\hat{p}_u^* = n/N$.

3.3 Bias Correction for Temporal Correlation

While most daily climate data are correlated over time, their annual maxima observations are weakly correlated in many cases, not inducing biased estimation caused by temporal dependence. However, the threshold exceedance series is more noticeably correlated. For example, snow depth threshold exceedances, as will be explained in Section 4.2, exhibit strong temporal dependence. Since the asymptotic result in Section 3.2 is supported under the assumption that each X_i is independent and identically distributed, return level estimates will be biased if the data are dependent.

To correct this bias caused by temporal dependence, past authors have incorporated an additional parameter. Suppose $D_1, \ldots, D_m$ are independent and identically distributed random variables and $D_1^*, \ldots, D_m^*$ are stationary random variables from the same distribution function $F(\cdot)$. Define $M_m = \max\{D_1, \ldots, D_m\}$ and $M_m^* = \max\{D_1^*, \ldots, D_m^*\}$. Under some mild conditions,

$$P\left(\frac{M_m - a_m}{b_m} \leq x\right) \longrightarrow G(x) \quad \text{as } m \to \infty,$$

if and only if

$$P\left(\frac{M_m^* - a_m}{b_m} \leq x\right) \longrightarrow G^\theta(x) \quad \text{as } m \to \infty,$$

for sequences of constants $\{a_m\}$ and $\{b_m\} > 0$, and a constant θ such that $0 < \theta \leq 1$ (cf. Coles, 2001). The unknown constant θ is called extremal index. Leadbetter (1983) interprets this parameter as the reciprocal of the limiting mean cluster size for stationary series, equating to 1 for perfectly independent series and converging to 0 for completely dependent series.

To estimate the extremal index, we use the intervals estimator (Ferro and Segers, 2003), although we do not particularly favor one single estimation method. Fawcett and Walshaw (2012) showed that this estimator performs very well compared to other methods in their simulation study. To elaborate, define $1 \leq \tau_1 < \cdots < \tau_n \leq N$ to be the times when threshold exceedances $\{X_1^*, \ldots, X_n^*\}$ were observed. Denote $\Delta_i = \tau_{i+1} - \tau_i$ to be the interexceedance time for $i = 1, \ldots, n-1$. Ferro and Segers (2003) derived the limiting distribution of the interexceedance times from a strictly stationary sequence of random variables and used the moment-based approach to obtain the estimator of θ as follows:

$$\hat{\theta} = \min\{\hat{\theta}^*, 1\}$$

where

$$\hat{\theta}^* = \begin{cases} \dfrac{2(\sum_{i=1}^{n-1} \Delta_i)^2}{(n-1)\sum_{i=1}^{n-1} \Delta_i^2}, & \text{if } \max\{\Delta_i\} \leq 2; \\ \dfrac{2(\sum_{i=1}^{n-1} (\Delta_i - 1))^2}{(n-1)\sum_{i=1}^{n-1} (\Delta_i - 1)(\Delta_i - 2)}, & \text{if } \max\{\Delta_i\} > 2. \end{cases}$$

Ferro and Segers (2003) also introduced a bootstrapping scheme to construct an approximate distribution of $\hat{\theta}$. Since $\hat{\theta}^{-1}$ represents the limiting mean cluster size, one can assume that there are approximately independent $L = \lceil \theta n \rceil$ clusters. The bootstrap scheme then divides interexceedance times into two categories: intracluster and intercluster times. Intracluster times are defined as a set of the top $L-1$ largest entries from Δ_i, expressed as $\{\Delta_{(1)}, \ldots, \Delta_{(L-1)}\}$. In case of $\Delta_{(L-1)} = \Delta_{(L)}$, reduce L until $\Delta_{(L-1)} > \Delta_{(L)}$. Then, intracluster times would be a collection of L sets of approximately independent interexceedance times, each divided by intercluster times. Each set of intracluster times is expressed as $\mathcal{A}_j = \{\Delta_{i_{j-1}+1}, \ldots, \Delta_{i_j+1}\}$, where $i_0 = 0$

and $i_C = n$, and $\mathcal{A}_j = \{\ \}$, if $i_{j-1} + 1 = i_j + 1$. To form a bootstrap replication of interexceedance times, the scheme randomly samples $L-1$ elements from intercluster times and L sets from intracluster times. Then, the estimated extremal index $\hat{\theta}^{(b)}$ is found from the b-th bootstrap sample for $b = 1, \ldots, B$.

If exceedances are clustered ($\theta < 1$), the frequency of each cluster of exceedances, instead of each individual exceedance, should be taken into account in computing return levels (Coles, 2001). For large enough K such that $x_K^* > u$, the return level x_K^* for dependent series is expressed as

$$x_K^* = \begin{cases} u + \frac{\sigma^*}{\xi}[(Kmp_u^*\theta)^\xi - 1], & \text{if } \xi \neq 0; \\ u + \sigma^* \ln(Kmp_u^*\theta), & \text{if } \xi = 0. \end{cases} \tag{3.5}$$

Note that $x_K^* = x_K$ when $\theta = 1$. The estimate for x_K^* is obtained by using the associated GP parameter estimates in (3.5).

3.4 Uncertainty Correction for Spatial and Temporal Correlation

Snow data recorded from one station are spatially correlated with its neighboring stations. In practice, the maximum likelihood estimation often uses a log-likelihood function constructed under the assumption that the data are independent and identically distributed, and finds the values of the unknown parameters by equating the first partial derivatives of log-likelihood function to zero. Although this assumption could not be perfectly met in real-world data, a common practice is to use these score function equations as a set of the estimating equations for the unknown parameters. Using this approach, Smith (1990) developed a method that finds the maximum

likelihood estimates under the independent and identical distribution assumption and then adjusts the standard error of the maximum likelihood estimates for spatial dependence. We propose extending this approach to our data with spatial and temporal dependence.

Now, since Smith's method is a basis of our treatment for spatial and temporal dependence, we illustrate this method. Suppose that there are multiple stations in a region and each station has n years' of data recorded. Assume the log-likelihood function for a p-dimensional vector of unknown model parameters, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^{\mathrm{T}}$, from all stations' data is expressed as

$$\ell_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} h_i(\boldsymbol{\theta}), \tag{3.6}$$

where h_i is the contribution of all stations' i-th year data to the log-likelihood function ℓ_n. Note that h_i's are treated as independent in (3.6). Let $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_p)^{\mathrm{T}}$ denote the maximum likelihood estimate of $\boldsymbol{\theta}$, and $\boldsymbol{\theta}_0 = (\theta_{1,0}, \ldots, \theta_{p,0})^{\mathrm{T}}$ be the true value of $\boldsymbol{\theta}$. Applying Taylor expansion to the estimating equation $\nabla \ell_n(\hat{\boldsymbol{\theta}}) = \mathbf{0}$ produces

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \approx -[\nabla^2 \ell_n(\boldsymbol{\theta}_0)]^{-1} \nabla \ell_n(\boldsymbol{\theta}_0),$$

where ∇ and ∇^2 represent gradient and Hessian

$$\nabla \ell_n(\boldsymbol{\theta}_0) = \begin{pmatrix} \dfrac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \theta_{1,0}} \\ \vdots \\ \dfrac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \theta_{p,0}} \end{pmatrix}, \qquad \nabla^2 \ell_n(\boldsymbol{\theta}_0) = \begin{pmatrix} \dfrac{\partial^2 \ell_n(\boldsymbol{\theta}_0)}{\partial \theta_{1,0}^2} & \cdots & \dfrac{\partial^2 \ell_n(\boldsymbol{\theta}_0)}{\partial \theta_{1,0} \partial \theta_{p,0}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 \ell_n(\boldsymbol{\theta}_0)}{\partial \theta_{p,0} \partial \theta_{1,0}} & \cdots & \dfrac{\partial^2 \ell_n(\boldsymbol{\theta}_0)}{\partial \theta_{p,0}^2} \end{pmatrix},$$

respectively. Approximating each entry in the Hessian by its expected value gives

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \approx H^{-1} \nabla \ell_n(\boldsymbol{\theta}_0), \tag{3.7}$$

where $H = -E[\nabla^2 \ell_n(\boldsymbol{\theta}_0)]$ is the Fisher information. Taking variance to both sides of (3.7) produces the variance of maximum likelihood estimators:

$$Cov(\hat{\boldsymbol{\theta}}) \approx H^{-1} V H^{-1}, \tag{3.8}$$

where $V = Cov(\nabla \ell_n(\boldsymbol{\theta}_0))$. If the assumed model is correct (that is, if the data are spatially independent), then one obtains $V = H$, resulting in the conventional approximation $Cov(\hat{\boldsymbol{\theta}}) \approx H^{-1}$. The Fisher information H then can be approximated by the observed information matrix $H \approx -\nabla^2 \ell_n(\hat{\boldsymbol{\theta}})$, which is typically calculated from software by using a quasi-Newton or similar algorithm.

Now, suppose the data, which construct ℓ_n in (3.6), are spatially dependent among stations, but the contribution h_i's are independent from each year i. This assumption can make sense in practice. For example, consider annual maximum snowfall series observed at several stations in a study region. Maximum snowfall values are spatially correlated among the stations each year but are nearly uncorrelated over years at each station. Under the assumption that h_i's have an identical distribution, V is reexpressed as

$$\begin{aligned} V &= Cov\left(\nabla \ell_n(\boldsymbol{\theta}_0)\right) \\ &= Cov\left(\sum_{i=1}^{n} \nabla h_i(\boldsymbol{\theta})\right) \\ &= \sum_{i=1}^{n} Cov\left(\nabla h_i(\boldsymbol{\theta})\right) \\ &= n Cov\left(\nabla h_1(\boldsymbol{\theta})\right). \end{aligned}$$

Here, $Cov(\nabla h_1(\boldsymbol{\theta}))$ can be estimated by the empirical covariance matrix of the observed gradient values $\{\nabla h_1(\hat{\boldsymbol{\theta}}), \ldots, \nabla h_n(\hat{\boldsymbol{\theta}})\}$ as follows:

$$\widehat{Cov}\big(\nabla h_1(\boldsymbol{\theta})\big) = \frac{1}{n}\begin{pmatrix} \sum_{i=1}^n (u_{i,1} - \bar{u}_{\cdot,1})^2 & \cdots & \sum_{i=1}^n (u_{i,1} - \bar{u}_{\cdot,1})(u_{i,p} - \bar{u}_{\cdot,p}) \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n (u_{i,p} - \bar{u}_{\cdot,p})(u_{i,1} - \bar{u}_{\cdot,1}) & \cdots & \sum_{i=1}^n (u_{i,p} - \bar{u}_{\cdot,p})^2 \end{pmatrix},$$

where $u_{i,j} = \frac{\partial}{\partial \theta_j} h_i(\boldsymbol{\theta})$ and $\bar{u}_{\cdot,j} = \frac{1}{n}\sum_{i=1}^n u_{i,j}$. Therefore, using $\hat{V}$ as an estimate of V in (3.8) produces the estimated variance of maximum likelihood estimators:

$$\widehat{Cov}(\hat{\boldsymbol{\theta}}) = \hat{H}^{-1}\hat{V}\hat{H}^{-1}, \tag{3.9}$$

where $\hat{H}$ is the approximation of H using a quasi-Newton algorithm and is usually calculated by software under independent and identically distributed assumption. Smith (1990) shows that if data are spatially correlated, the variance in (3.9) is more accurate than the variance under independent and identically distributed assumption.

Temporal dependence in extreme values can be a negligible issue for block maxima methods when block size is large enough. Our block maxima method uses annual maxima, which represents a large block size of 365 days, and the autocorrelation in annual maxima at lag 1 (one year) are in fact very low (-0.041 in Central Park, 0.014 in Newark, -0.045 in La Guardia, and -0.188 in JFK for annual snowfall maxima). However, threshold exceedances methods must take into account temporal dependence to obtain accurate results. When the exceedances over a given threshold are considered, these exceedances can be significantly autocorrelated for a short period of time. Fitting the independence-assumed GP distribution to these autocorrelated threshold exceedances can be problematic.

To address temporal dependence in threshold exceedances, decorrelation techniques have been developed. These techniques aim to extract a set of independent values from all threshold exceedances. Past studies often applied the "*runs-declustering*" algorithm that considers all exceedances within a user-specific parameter to be a cluster and used only the maximum value of all exceedances within each cluster (Davison and Smith, 1990). However, Fawcett and Walshaw (2007) showed that the declustering method can systemically incur bias in model parameter estimates, producing underestimated return levels when threshold exceedances are strongly autocorrelated. They instead used all threshold exceedances and proposed modifying the standard error of the maximum likelihood estimates for autocorrelation in the threshold exceedances by incorporating Smith's method. Fawcett and Walshaw (2012) showed that using all threshold exceedances with Smith's adjustment method performs better than the declustering method for return level estimation, provided that the extremal index is appropriately estimated.

Extremes are rare by their nature. With the temporal correlation in threshold exceedances taken into account, we can further improve estimation accuracy by incorporating all neighboring, spatially correlated stations into a threshold exceedance model. Thus, this approach can use the maximum possible number of extreme observations available to us. We propose extending Smith's method to this spatially and temporally dependent extreme data, as Smith's method has been shown to perform well under spatial dependence and under temporal dependence. In our study, the log-likelihood function for $\boldsymbol{\theta}$ in (3.6) is reexpressed as

$$\ell_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} h_i(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{t=1}^{n_{ij}} g_{ijt}(\boldsymbol{\theta}), \tag{3.10}$$

where g_{ijt} is a contribution from the t-th observation from station j in i-th year to ℓ_n, m is the number of weather stations, and n_{ij} is the number of observations in station j for i-th year. We assume in (3.10) that each contribution h_i, which is constructed from observations with some dependence, is independent of each other. If the data are spatially and temporally dependent but independent in each year, we can construct each h_i and proceed as usual.

3.5 Confidence Intervals for Return Levels

The delta method is often used to construct asymptotic confidence intervals for a function of parameters. Denote $\hat{\boldsymbol{\theta}}$ by a vector of maximum likelihood estimators and $\boldsymbol{\theta}_0$ a vector of true values of the unknown parameters. Suppose $\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \to N(\mathbf{0}, \boldsymbol{\Sigma})$ in distribution. For a given function $g(\cdot)$, if the p-dimensional $\nabla g(\boldsymbol{\theta_0})$ exists and is non-zero, then

$$\sqrt{n}(g(\hat{\boldsymbol{\theta}}) - g(\boldsymbol{\theta})) \longrightarrow N(0, \nabla g(\boldsymbol{\theta}_0)^{\mathrm{T}} \boldsymbol{\Sigma} \nabla g(\boldsymbol{\theta}_0))$$

in distribution (cf. Casella and Berger, 2001). Rust et al. (2011) obtained the asymptotic variance of the return level x_K in (3.3) for the Gumbel distribution (equivalently a GEV distribution with $\xi = 0$) as follows:

$$Var(\hat{x}_K) \approx \beta_{11} - (\beta_{12} + \beta_{21}) \log\left(-\log\left(1 - \frac{1}{K}\right)\right) + \beta_{22}\left[\log\left(-\log\left(1 - \frac{1}{K}\right)\right)\right]^2,$$

where β_{ij} is the (i, j)-th element of the Fisher information matrix inverse H^{-1}. However, the asymptotic confidence intervals based on the delta method often fail to reach targeted coverage probabilities, inadequately capturing the sampling variability of the maximum likelihood estimators (Rust et al., 2011). In addition, finding the

asymptotic distribution for the return level is very difficult for the GEV distribution with $\xi \neq 0$ or GP distribution settings.

We consider bootstrap confidence intervals. To obtain a bootstrap confidence interval, B number of replications, each of size n, need to be obtained from the data. The simple resampling method would be to select a random sample of size n with replacement from the data for each iteration. However, this simple resampling method often fails to approximate the true distribution of data when the data is not independent and identically distributed (cf. Givens and Hoeting, 2013). There are many techniques available to address this problem. We choose the moving block bootstrap, first introduced by Künsch (1989), due to its effectiveness and simplicity. To implement the moving block bootstrap, $n - l - 1$ overlapping boxes, each sized at l, needs to be defined from the data. n/l boxes of length l are then randomly sampled from these boxes to obtain a bootstrap sample. For example, if we have data $\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$ with sample size $n = 9$ and choose $l = 3$, a possible bootstrap sample would be $\{x_3, x_4, x_5, x_4, x_5, x_6, x_1, x_2, x_3\}$. From the b-th bootstrap sample for $b = 1, \ldots, B$, maximum likelihood estimates of the model parameters are found to compute the return level estimate, $\hat{x}_K^{(b)}$.

The classical percentile bootstrap method uses the $\alpha/2$-th lower and upper quantiles of the return level bootstrap estimates $\{\hat{x}_K^{(1)}, \ldots, \hat{x}_K^{(B)}\}$ as confidence intervals' endpoints. However, the percentile bootstrap method often produces biased confidence intervals if the bootstrap distribution is skewed (DiCiccio and Efron, 1996). Efron (1987) developed the bias-corrected and accelerated (BCa) bootstrap method to correct bias due to skewness. The BCa method first computes the bias-correction constant z_{BC} by

$$z_{\mathrm{BC}} = \Phi^{-1}\left(\frac{1}{B}\sum_{b=1}^{B} I(\hat{x}_K^{(b)} < \hat{x}_K)\right),$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, $I(E)$ is an indicator function which returns 1 if E is true and 0 otherwise, and $\hat{x}_K$ is the estimated return level from the original data. Next, the acceleration constant c_{A} is computed by

$$c_{\mathrm{A}} = \frac{\sum_{i=1}^{n}(\tilde{x}_K^{(-i)} - \tilde{x}_K^{(\cdot)})^3}{6\big[\sum_{i=1}^{n}(\tilde{x}_K^{(-i)} - \tilde{x}_K^{(\cdot)})^2\big]^{3/2}},$$

where $\tilde{x}_K^{(-i)}$ is the delete-1 jackknife estimate of x_K with i-th observation deleted from the data and $\tilde{x}_K^{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n}\tilde{x}_K^{(-i)}$ (cf. Givens and Hoeting, 2013). Then, the $(1-\alpha)\times 100\%$ BCa confidence interval uses the following quartiles of the return level bootstrap estimates as the interval endpoints:

$$\Phi\left(z_{\mathrm{BC}} + \frac{z_{\mathrm{BC}} - z_{\alpha/2}}{1 - c_{\mathrm{A}}(z_{\mathrm{BC}} - z_{\alpha/2})}\right), \qquad \Phi\left(z_{\mathrm{BC}} + \frac{z_{\mathrm{BC}} + z_{\alpha/2}}{1 - c_{\mathrm{A}}(z_{\mathrm{BC}} + z_{\alpha/2})}\right),$$

where $z_{\alpha/2}$ is the $\alpha/2$-th upper quantile of the standard normal distribution. We used this method to obtain confidence intervals for return levels.

CHAPTER 4

RESULTS

4.1 Annual Maximum Snowfall Results

We apply the block maxima method to the annual maximum snowfall series for each of the four stations selected from the New York City area. To elaborate, let M_{it} be the maximum snowfall observation at station i during snow year t. In order to account for possible geographical location effects on snowfall, we assume that $\{M_{it}\}$ follows a GEV(μ_i, σ_i, ξ_i) distribution, with μ_i as the GEV location parameter for each of the Central Park, Newark, La Guardia, and JFK stations. For model selection, we use the corrected Akaike information criterion (AIC_C), defined as:

$$AIC_C = 2k - 2\ell + \frac{2k(k+1)}{n-k-1},$$

where ℓ is the maximum log likelihood value for model, k is the number of model parameters, and n is the sample size. The model with the lowest AIC_C value is preferred. We find that a GEV model with different σ_i's and ξ_i's does not significantly improve a GEV model fit. Specifically, a model with different μ_i's, σ_i's, and ξ_i's returned the AIC_C value of 1376.398, whereas a model with different μ_i's but σ and ξ fixed returned the AIC_C value of 1362.943. Hereafter, we consider only those models with different μ_i's but σ and ξ fixed for all four stations.

Our GEV estimation results are summarized in Table 4.1. We consider four different GEV models with varying number of μ_i's by iteratively merging closest μ_i's. We choose Model 3 for further analysis as it has the lowest AIC_C value. We then compute the standard errors of GEV maximum likelihood estimates. The naïve standard errors (ignoring spatial dependence among the four stations) and the corrected standard errors by Smith's method (including spatial correlation for the stations) are displayed in the table next to the parameter estimates. The spatial correlation corrected standard errors are greater than the corresponding naïve standard errors, implicating that there are more uncertainty in the GEV parameter estimates once the spatial correlation is considered.

Table 4.1: Stationary GEV estimates for annual maximum snowfall data with their associated standard errors in parentheses (*left*: naïve, *right*: corrected)

		Model 1	Model 2	Model 3	Model 4
μ	μ_1	6.952 (0.484, 0.545)	6.863 (0.375, 0.589)	6.748 (0.322, 0.577)	6.556 (0.293, 0.571)
	μ_2	6.760 (0.512, 0.658)			
	μ_3	6.526 (0.491, 0.578)	6.535 (0.491, 0.573)		
	μ_4	5.981 (0.488, 0.578)	5.988 (0.489, 0.574)	5.990 (0.489, 0.573)	
σ		3.788 (0.229, 0.482)	3.793 (0.228, 0.481)	3.797 (0.228, 0.480)	3.823 (0.230, 0.473)
ξ		0.168 (0.059, 0.085)	0.165 (0.058, 0.084)	0.165 (0.058, 0.084)	0.162 (0.059, 0.081)
ℓ		-674.212	-674.254	-674.426	-675.487
AIC_C		1362.943	1360.895	1359.127	1359.157

Snowfall data could show non-stationary characteristics, such as long-term trends. We consider this issue by including adequate parameters into our models. Specifically, to quantify a possible long-term linear trend in the annual maximum snowfall series at station i, we reparameterize the GEV location parameter μ in (3.1) to have a linear trend:

$$\mu_{it} = \mu_i + \beta \left(\frac{t - 1958}{10} \right)$$

for snow year $t = 1959, \ldots, 2015$. Here, the trend parameter β is interpreted as the expected change in maximum snowfall over a decade:

$$E(M_{i,t+10}) - E(M_{i,t}) = \left[\mu_{i,t+10} + \frac{\sigma}{\xi}(\Gamma(1-\xi) - 1) \right] - \left[\mu_{i,t} + \frac{\sigma}{\xi}(\Gamma(1-\xi) - 1) \right] = \beta,$$

when $\xi < 1$. We assume that the geospatial trends in maximum snowfall series are the same for the four stations. Results from the GEV maximum likelihood estimation are summarized in Table 4.2. We will only consider Model 3 for further analysis as this model is parsimonious and still a good fit for the data based on the AIC_C values. The estimated linear trend in Model 3 is 0.232 inches decade^{-1}. Based on 56 annual maximum snowfall observations from each of the four stations, the trend estimate β is not significantly different from zero with the corrected standard error 0.292 used for significance test. However, we note that this estimate could be more statistically meaningful when we have more snow year data.

Table 4.2: Non-stationary GEV estimates for annual maximum snowfall data with their associated standard errors in parentheses (*left*: naïve, *right*: corrected)

		Model 1	Model 2	Model 3	Model 4
μ	μ_1	6.281 (0.656, 0.911)	6.214 (0.562, 0.950)	6.114 (0.523, 0.929)	5.962 (0.514, 0.927)
	μ_2	6.181 (0.656, 1.010)			
	μ_3	5.953 (0.639, 0.911)	5.918 (0.632, 0.905)		
	μ_4	5.287 (0.677, 0.930)	5.288 (0.671, 0.926)	5.293 (0.671, 0.923)	
	β	0.246 (0.154, 0.298)	0.231 (0.152, 0.293)	0.232 (0.152, 0.292)	0.213 (0.153, 0.297)
σ		3.796 (0.233, 0.473)	3.768 (0.229, 0.459)	3.772 (0.230, 0.458)	3.807 (0.230, 0.454)
ξ		0.156 (0.061, 0.083)	0.169 (0.061, 0.081)	0.169 (0.061, 0.080)	0.162 (0.060, 0.077)
ℓ		-673.136	-673.121	-673.263	-674.523
AIC_C		1360.791	1358.629	1356.801	1357.229

Estimated return levels for 25, 50, 75, and 100 years based on the selected stationary and non-stationary GEV models are summarized in Table 4.3. For interval estimation of return levels, the typical confidence interval based on asymptotic theory, expressed as an estimate ± 1.96 standard error, can be considered. However, as the shape parameter estimate is positive, the GEV distribution is right skewed, resulting in that the return level distribution to be also right skewed. We instead consider bootstrap confidence intervals to approximate this right skewness. To further correct bias in conventional percentile bootstrap methods, we compute the BCa bootstrap confidence intervals with 10000 replications, producing 95% BCa bootstrap intervals as shown in Table 4.3. Although maximum likelihood estimates of return levels calculated with and without trends are different, it is worth mentioning that their 95% confidence intervals are almost identical. We believe this further verifies that there are no significant long term linear trends in the annual maximum snowfall series at this point. We find that the total amount of snowfall brought by the January 2016 snowstorm is about equivalent to a 50-year return level for Central Park, Newark, and

Table 4.3: GEV annual maximum snowfall return level estimates and their associated 95% BCa bootstrap confidence intervals

Without trend	Actual	K_{25}	K_{50}	K_{75}	K_{100}
Central Park Newark La Guardia	27.5 24.5 28.2	22.753 (20.296, 26.057)	27.561 (24.184, 33.624)	30.622 (26.547, 39.103)	32.918 (28.259, 43.536)
JFK	30.4	22.996 (19.440, 25.147)	26.802 (23.364, 32.635)	29.864 (25.782, 38.017)	32.160 (27.502, 42.378)
With trend	Actual	K_{25}	K_{50}	K_{75}	K_{100}
Central Park Newark La Guardia	27.5 24.5 28.2	24.000 (20.849, 26.885)	29.421 (25.093, 34.080)	33.090 (27.721, 39.199)	35.990 (29.727, 43.316)
JFK	30.4	23.178 (19.993, 26.171)	28.599 (24.240, 33.256)	32.269 (26.966, 38.400)	35.169 (28.922, 42.522)

La Guardia; and 75-year return level for JFK when linear trends are not considered. However, if linear trends are included, the actual snow observations are between 25 and 50-year return levels for Central Park, Newark, and La Guardia; and between 50 and 75-year return levels for JFK, respectively.

4.2 87.5-th Percentile Snowfall Results

Now, we apply the threshold exceedances method for the four selected stations. As discussed in Sections 3.1 and 3.2, the block maxima methods use only block maximum observations whereas the threshold exceedances methods use all extreme observations above a certain threshold. Due to a lack of enough extreme events, threshold exceedances methods can provide additional meaningful information in understanding extreme events. To apply the threshold exceedances method, we first need to determine an adequate threshold u for the daily non-zero snowfall series. The MRL plot in Figure 4.1 and parameter stability plots in Figure 4.2 suggest that a threshold of 87.5-th percentile of all non-zero snowfall observations (around 6 inches) could be an adequate threshold when all observations are considered. As the National Weather Service Forecast Office in New York City issues a winter storm warning when 6 inches or more of snow is expected in a 12 hour period, or at least 8 inches for a 24 hour period, this threshold choice can also be justified on practical viewpoints. We choose thresholds for non-zero snowfall series to be 87.5-th percentile so that 12.5% of observations would exceed it.

We then fit a GP distribution to all exceedances over a threshold. For this, let X_{it} be the snowfall observation for day t at station i. Assume that u_i is the predetermined threshold associated with station i. If $X_{it} > u_i$, we denote such X_{it} by X^*_{it}. The exceedance over threshold u_i is then expressed as $Y_{it} = X^*_{it} - u_i$. By extreme value

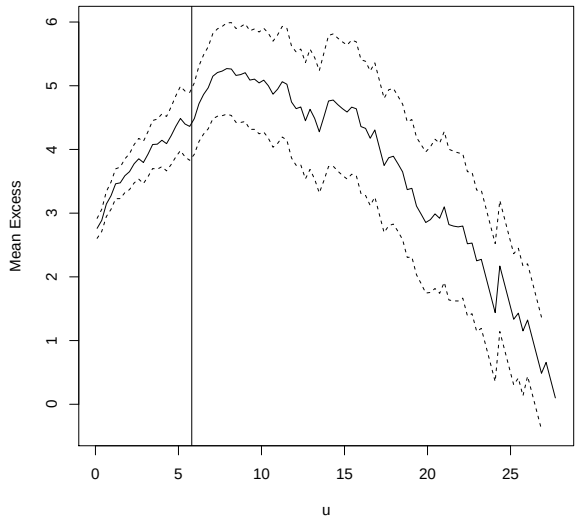


Figure 4.1: Mean residual life plot for all non-zero snowfall observations (vertical line: 87.5-th percentile, u: threshold)

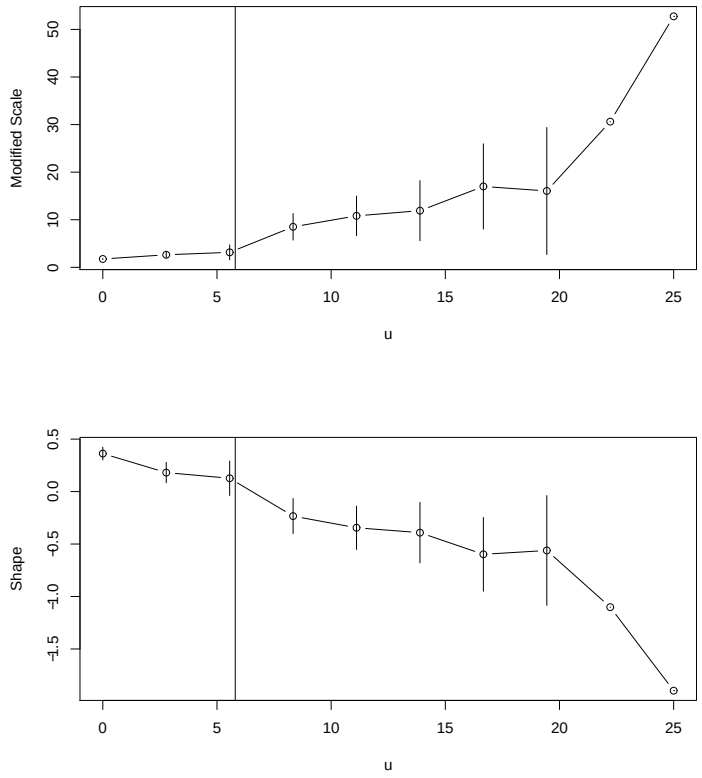


Figure 4.2: Parameter stability plots for GP fit to all non-zero snowfall observations (vertical line: 87.5-th percentile, u: threshold)

theory, we assume that these positive Y_{it}'s have a GP(σ_i^*, ξ_i) distribution for each station i. As in the previous GEV-based block maxima method, we further assume that the GP scale and shape parameters, σ_i^*'s and ξ_i's, are identical within each i, since these four stations are located close to each other. In fact, our GP estimation for different scale and shape parameters does not produce a significantly improved model. Also, our estimates of ξ_i's are not significantly different from zero for all model cases (Models 1, 2, 3, and 4), leading to set ξ_i's to be zero in our GP analysis. We consider only those GP models with identical $\sigma_i^* = \sigma^*$ and $\xi_i = 0$. We also comment that since non-zero snowfall observations are clustered during various periods from one day to five days and since we combined consecutive snowfall events in a single duration, an adjustment is needed to calculate the over-threshold probability $p_u^* = P(X > u)$. We estimate p_u^* by n^*/N^*, where n^* is the total number of days that threshold-exceeding snowfall events occurred and N^* is the total number of days that any nonzero snowfall events occurred. Furthermore, since we combined consecutive snowfall events, serial correlation is negligible and the independence assumption therefore appears to be valid.

Our threshold exceedances method proceeds by following two steps. First, a quantile regression model is applied to obtain an optimal threshold. Second, GP-based maximum likelihood estimation is performed for over-threshold exceedances. Quantile regression methods estimate the conditional quantiles for the probability distribution of a response variable, producing a fitted curve that corresponds to various percentage points for the response variable (Koenker, 2005). We obtain stationary thresholds by fitting a 87.5-th quantile regression with different intercepts, u_i, from all non-zero snowfall observations. Different quantile regression models are fitted by merging similar u_i's, and their corresponding GP parameters, σ_i^*, are estimated under the

assumption that $\xi_i = 0$. Table 4.4 summarizes our two-step estimation results. We choose Model 3 for further analysis among the four models we considered, as this model is parsimonious and still produces a good fit for the data. We comment that since all four models have different sample observations, the log-likelihood values for each model should not be interpreted in the same manner as the GEV cases. Thus, it would be unreasonable to compare AIC_C values. The standard errors for quantile regression parameters are computed using the "*se=boot*" option within the "*summary.rq*" function in the R package "*quantreg*," which implements one of the bootstrap approaches (Koenker, 2016). The default "*xy-pair*" method is used in our studies. The standard errors for the maximum likelihood estimate of GP scale parameter σ is computed using two different approaches. The first approach is the naïve method that ignores spatial and temporal dependences in snowfall observations of the four locations. The second approach is the corrected method, which extends Smith's method to account for both spatial and temporal dependences. The corrected method inflated the GP standard errors by a factor of two, which is similar to our finding in the GEV cases. This implies that spatial and temporal dependence contributes to larger uncertainty in GP parameter estimates.

Table 4.4: Stationary GP estimates for non-zero snowfall data and their associated standard errors in parentheses (*left*: naïve, *right*: corrected) for σ^* with the assumption that $\xi = 0$

		Model 1	Model 2	Model 3	Model 4
u	u_1	6.000 (0.384)	6.000 (0.384)	5.900 (0.205)	5.800 (0.181)
	u_2	5.700 (0.455)	5.800 (0.314)		
	u_3	5.900 (0.408)			
	u_4	5.500 (0.483)	5.500 (0.483)	5.500 (0.483)	
σ^*		4.652 (0.280, 0.493)	4.601 (0.275, 0.491)	4.555 (0.272, 0.487)	4.542 (0.270, 0.486)
ℓ		-702.831	-707.339	-707.047	-708.760

Under climate change scenarios, use of constant thresholds can cause inaccurate results. For example, a constant threshold that is adequate for observations from 1960 could be too low or too high for observations from 2010. Recent studies (Kyselỳ et al., 2010; Northrop and Jonathan, 2011; Jonathan et al., 2014) used quantile regression methods to obtain time-dependent thresholds for threshold modeling. Therefore, we consider non-stationary thresholds to quantify a possible long-term linear trend in the over-threshold snowfall exceedances in the New York City area. We assume that the trends in non-zero snowfall series are identical for the four stations. We estimate the time-dependent threshold u_{it} by fitting a 87.5-th quantile regression model with different intercepts u_i's and a linear trend β as follows:

$$u_{it} = u_i + \beta \left(\frac{t}{365.25 \times 10} \right).$$

Here, t denotes the day t^* in which $X^*_{it^*} > 0$. Note $t^* = 1, \ldots, 20454$ with $t^* = 1$ as July 1, 1959 (the first day of snow year 1959), and $t^* = 20454$ as June 30, 2016 (the last day of snow year 2015 and last day of the study period). If a snowfall event had lasted for more than one day, the average of corresponding t^* is used as the value for t. This time-dependent threshold u_{it} can more likely keep the over-threshold probability p^*_u constant over the entire study period. Table 4.5 summarizes our quantile regression and GP-based maximum likelihood estimation results. Similar to the stationary case, we only consider Model 3 for further analysis. The estimated linear trend in Model 3 is 0.564 inches decade^{-1}, which is about 2.4 times greater than the GEV estimated trend for annual maximum snowfall series. In addition, the standard error for Model 3 with all exceedances is much smaller than the standard error for the GEV trend estimate with annual maximum series. This is possibly due to the fact that the

quantile regression and GP model uses all non-zero observations, whereas the GEV model uses only annual maxima, therefore contributing to the reduction of variance. The estimated trend in all exceedances is determined to be statistically significant, implying that 87.5-th percentile snowfall have increased by 0.564 inches decade^{-1}. This is in contrast to the annual maximum snowfall, which shows an insignificant increase by 0.232 inches decade^{-1}. For further analysis, only non-stationary GP models are considered as they have shown to offer significantly improved fitting over stationary GP models.

Table 4.5: Non-stationary GP estimates for non-zero snowfall data and their associated standard errors in parentheses (*left*: naïve, *right*: corrected) for σ^* with the assumption that $\xi = 0$

		Model 1	Model 2	Model 3	Model 4
u	u_1	4.652 (0.559)	4.644 (0.600)	4.633 (0.607)	4.109 (0.363)
	u_2	4.044 (0.487)	4.112 (0.395)	3.986 (0.381)	
	u_3	4.232 (0.490)			
	u_4	3.919 (0.512)	3.850 (0.455)		
	β	0.525 (0.122)	0.541 (0.110)	0.564 (0.116)	0.548 (0.111)
σ^*		4.582 (0.272, 0.500)	4.529 (0.268, 0.499)	4.527 (0.269, 0.502)	4.618 (0.275, 0.513)
ℓ		-713.739	-717.976	-712.870	-716.002

Estimated return levels for 25, 50, 75, and 100 years from the selected non-stationary GP model along with their 95% confidence intervals computed from the BCa bootstrap method with 10000 replications are shown in Table 4.6. Since serial correlations on our 87.5-th percentile snowfall series are negligible, a standard paired bootstrap is employed in lieu of the moving block bootstrap. Return level estimates computed from non-stationary GP models align fairly well with those from GEV models. We find that the actual snowfall observations from the blizzard are about equivalent to a 25-year return level for Central Park, less than a 25-year return level for Newark, and between 25 and 50-year return levels for La Guardia and JFK.

Table 4.6: Non-stationary GP snowfall return level estimates and their associated 95% BCa bootstrap confidence intervals

	Actual	K_{25}	K_{50}	K_{75}	K_{100}
Central Park	27.5	27.216 (24.199, 30.121)	31.809 (28.224, 35.319)	35.070 (31.008, 39.097)	37.789 (33.256, 42.308)
Newark La Guardia JFK	24.5 28.2 30.4	27.365 (24.701, 30.196)	31.958 (28.717, 35.416)	35.219 (31.543, 39.242)	37.938 (33.799, 42.426)

4.3 Annual Maximum Snow Depth Results

To study a more comprehensive characteristic of snow events, we now consider snow depth. The block maxima method is first applied to the annual maximum snow depth series recorded at the four selected stations. Let M_{it} be the maximum snow depth observation at station i during snow year t. Since the snow depth measurement in a day is affected by the temperature of the previous day, we include temperature data in our analysis. Let T_{it} be the maximum temperature observation at station i in the previous day when M_{it} was observed. Specifically, the Central Park station had its maximum snow depth observation for the snow year 2014 recorded on March 6, 2015. We use the maximum temperature observation on March 5, 2015 as the corresponding temperature observation for the March 6 snow depth. We assume that $\{M_{ij}\}$ follows a GEV(μ_{it}, σ, ξ) distribution with μ_{it} modeled as

$$\mu_{it} = \mu_i + \nu(T_{it} - 32).$$

Here, the parameter ν represents the expected change in maximum snow depth for every one degree increase of the previous day's maximum temperature starting from 32°F. Geographical proximities influence snow depth by this model for μ_{it}. We assume

that all four stations have identical σ, ξ, and ν, as verified in our preliminary analysis.

Four different models are fitted by combining μ_i's, producing the results in Table 4.7. Of these models, we choose Model 3 for our analysis as this model balances well between goodness-of-fit and parsimoniousness based on the AIC_C values. Both naïve standard errors and corrected standard errors by Smith's method are reported along with maximum likelihood estimates of model parameters. Overall, these results are similar to those from GEV snowfall analysis, but the shape parameter estimates in annual maximum snow depth series are greater than the corresponding estimates in annual maximum snowfall series.

Table 4.7: Stationary GEV estimates for annual maximum snow depth data with their associated standard errors in parentheses (*left*: naïve, *right*: corrected)

		Model 1	Model 2	Model 3	Model 4
μ	μ_1	6.514 (0.490, 0.705)	7.112 (0.521, 0.887)	6.440 (0.366, 0.674)	6.119 (0.290, 0.706)
	μ_2	6.370 (0.459, 0.627)	6.848 (0.487, 0.781)		
	μ_3	5.883 (0.455, 0.584)	6.090 (0.366, 0.644)	5.816 (0.347, 0.595)	
	μ_4	5.748 (0.453, 0.594)			
	ν	-0.215 (0.039, 0.069)	-0.235 (0.043, 0.077)	-0.217 (0.039, 0.070)	-0.217 (0.039, 0.237)
σ		3.586 (0.241, 0.444)	3.779 (0.252, 0.517)	3.585 (0.240, 0.461)	3.614 (0.241, 0.945)
ξ		0.240 (0.075, 0.102)	0.182 (0.072, 0.107)	0.238 (0.074, 0.105)	0.232 (0.073, 0.351)
ℓ		-655.275	-656.408	-655.330	-656.429
AIC_C		1325.081	1325.212	1320.942	1321.045

We estimate a long-term linear trend in snow depth data by modeling the GEV location parameter μ in (3.1) as

$$\mu_{it} = \mu_i + \beta\left(\frac{t-1958}{10}\right) + \nu(T_{it} - 32)$$

for snow year $t = 1959, \ldots, 2015$. The parameter β is interpreted as the expected change in maximum snow depth over a decade under the assumption that temperature

remains unchanged. We further assume that the trends in the maximum snow depth series are the same for the four stations. Table 4.8 summarizes our GEV estimation results. Overall, the results are similar to ones from models without trends. Model 3 is selected for further analysis since it has the lowest AIC_C value. The estimated linear trend from Model 3 is 0.157± 1.96 (0.165 *naïve*, 0.292 *corrected*) inches decade^{-1}. Although this estimate is not significantly different from zero, we note that this trend will be also used to estimate return levels.

Table 4.8: Non-stationary GEV estimates for annual maximum snow depth data with their standard errors in parentheses (*left*: naïve, *right*: corrected)

		Model 1	Model 2	Model 3	Model 4
μ	μ_1	6.471 (0.618, 0.769)	6.317 (0.608, 0.757)	6.378 (0.579, 0.824)	5.733 (0.458, 0.709)
	μ_2	6.206 (0.618, 0.838)	6.321 (0.595, 0.881)		
	μ_3	5.273 (0.615, 0.650)	5.674 (0.493, 0.718)	5.612 (0.559, 0.737)	
	μ_4	5.421 (0.616, 0.691)			
	β	0.098 (0.151, 0.264)	0.143 (0.144, 0.277)	0.157 (0.165, 0.292)	0.155 (0.146, 0.280)
	ν	-0.221 (0.045, 0.078)	-0.225 (0.042, 0.077)	-0.241 (0.048, 0.085)	-0.225 (0.043, 0.079)
σ		3.721 (0.262, 0.519)	3.634 (0.237, 0.485)	3.891 (0.274, 0.565)	3.644 (0.243, 0.478)
ξ		0.234 (0.083, 0.129)	0.179 (0.070, 0.112)	0.142 (0.074, 0.116)	0.212 (0.075, 0.116)
ℓ		-655.587	-655.343	-655.764	-665.825
AIC_C		1327.860	1325.217	1323.924	1341.932

As snow depth for a day is dependent on the temperature reading on that day, return levels also depend on temperatures. Here, we consider three different temperature scenarios: (1) the lowest temperature ever recorded during the study period in each station (9°F for Central Park, 5°F for Newark, and 8°F for both La Guardia and JFK), (2) 32°F, and (3) 40°F. Each scenario is chosen for its own reason. The lowest temperature (minimum) scenario considers return levels under the most extreme temperature case. The 32°F scenario is a reasonable choice as it is a freezing point.

The 40°F scenario is chosen as this is approximately the average daily maximum temperature for the days with non-zero snow depth observed.

Under the three temperature scenarios, return levels for 25, 50, 75, and 100 years are calculated from the selected stationary and non-stationary GEV models. Table 4.9 summarizes these return level results along with their associated 95% BCa bootstrap confidence intervals from 10000 replications. Similar to snowfall, the fitted GEV distribution is right-skewed as the shape parameter is positive. To take into account the right skewness in the GEV return level distribution, bootstrap confidence intervals are used instead of the asymptotic confidence interval. We find that maximum likelihood estimates of return levels and their respective 95% confidence intervals calculated with linear trends are consistently lower than ones calculated without linear trends. This results seem to be contradictory because the estimated linear trend β, although statistically not significant, is positive. We believe this is due to the fact that the estimated ξ from non-stationary Model 3 is much smaller than the estimated ξ from stationary Model 3, which contributed to fitted return levels from non-stationary models to be smaller, even with a positive slope. Regardless, we do find results from stationary and non-stationary models are similar and we believe that this further supports our finding that a significant long term linear trend in the annual maximum snow depth series does not exist at this point. Lastly, we find that the maximum snow depth observation recorded from the 2016 snowstorm is less than a 25-year return level for Central Park and Newark; and about equivalent to a 75-year return level for La Guardia and JFK under the 32°F temperature scenario. However, if we consider the 40°F temperature scenario, the actual snow observations are about equivalent to a 25-year return level for Central Park and Newark; and less than a 50-year return level for La Guardia and JFK. Furthermore, the actual snow

Table 4.9: GEV annual maximum snow depth return level estimates and their associated 95% BCa bootstrap confidence intervals under three different temperature scenarios

Without Trend / Minimum	Actual	K_{25}	K_{50}	K_{75}	K_{100}
Central Park	22	28.622 (25.184, 31.802)	34.495 (29.749, 39.618)	38.388 (32.561, 45.540)	41.381 (34.599, 50.400)
Newark	20	29.492 (25.924, 32.785)	35.365 (30.501, 40.425)	39.258 (33.391, 46.314)	42.251 (35.479, 51.145)
La Guardia	27	28.216	34.088	37.981	40.975
JFK	28	(24.779, 31.502)	(29.331, 39.273)	(32.122, 45.158)	(34.161, 50.083)
With Trend / Minimum	**Actual**	K_{25}	K_{50}	K_{75}	K_{100}
Central Park	22	28.955 (25.379, 32.594)	33.883 (28.329, 37.605)	37.127 (29.168, 41.251)	39.640 (31.429, 44.261)
Newark	20	29.920 (26.199, 33.841)	34.848 (29.121, 38.689)	38.092 (30.895, 42.233)	40.605 (32.265, 45.229)
La Guardia	27	28.431	33.359	36.603	39.116
JFK	28	(24.688, 31.979)	(27.675, 37.017)	(29.629, 40.677)	(31.052, 43.626)
Without Trend / 32°F	**Actual**	K_{25}	K_{50}	K_{75}	K_{100}
Central Park	22	23.621	29.493	33.386	36.379
Newark	20	(20.558, 27.102)	(24.788, 35.615)	(27.394, 41.739)	(29.311, 46.734)
La Guardia	27	22.997	28.869	32.762	35.755
JFK	28	(19.906, 26.473)	(24.121, 35.001)	(26.733, 41.133)	(28.663, 46.040)
With Trend / 32°F	**Actual**	K_{25}	K_{50}	K_{75}	K_{100}
Central Park	22	23.404	28.332	31.576	34.089
Newark	20	(20.060, 26.394)	(23.049, 32.032)	(24.894, 35.863)	(26.251, 38.970)
La Guardia	27	22.639	27.566	30.811	33.324
JFK	28	(19.191, 25.444)	(22.217, 31.099)	(24.053, 34.998)	(25.437, 38.100)
Without Trend / 40°F	**Actual**	K_{25}	K_{50}	K_{75}	K_{100}
Central Park	22	21.881	27.753	31.646	34.640
Newark	20	(18.669, 25.840)	(22.871, 34.445)	(25.477, 40.616)	(27.403, 45.577)
La Guardia	27	21.257	27.129	31.022	34.015
JFK	28	(17.991, 25.187)	(22.161, 33.753)	(24.816, 39.956)	(26.694, 44.769)
With Trend / 40°F	**Actual**	K_{25}	K_{50}	K_{75}	K_{100}
Central Park	22	21.473	26.401	29.645	32.158
Newark	20	(17.930, 24.578)	(20.983, 30.278)	(22.786, 34.233)	(23.079, 37.248)
La Guardia	27	20.708	25.636	28.880	31.393
JFK	28	(17.130, 23.630)	(20.068, 29.384)	(21.817, 33.237)	(22.891, 36.323)

observations are significantly less than a 25-year return level for Central Park and Newark; and about equivalent to a 25-year return level for La Guardia and JFK under the minimum temperature scenario.

4.4 75-th Percentile Snow Depth Results

The threshold exceedances method is now applied to snow depth series recorded at the four stations. To determine an adequate threshold, an MRL plot is again examined as shown in Figure 4.3. The MRL plot suggests that 75-th percentile of all non-zero snow depth observations (around 6 inches) could be an adequate threshold. We choose the 75-th percentile as thresholds for non-zero snow depth series so that 25% of observations would exceed it. Similar to the snowfall case, this is again consistent with the practical viewpoint. A GP distribution is then fitted to all exceedances over the threshold. Scale and shape parameters, σ^* and ξ, respectively, are assumed to be

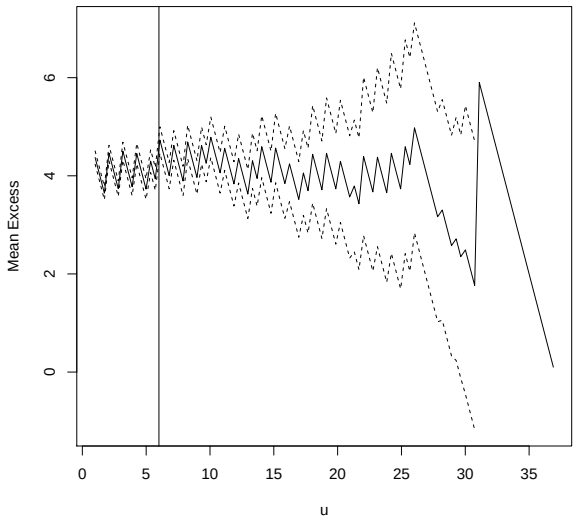


Figure 4.3: Mean residual life plot for all non-zero snow depth observations (vertical line: 75-th percentile, u: threshold)

identical across all stations. Also, ξ is set to be zero since our maximum likelihood estimate of ξ is very close to zero. We assume that the exceedances Y_{it} over u_i for day t at station i have a GP$(\sigma^*, \xi = 0, \theta_i)$ distribution. Since each observation here is equally weighted at one day each, no adjustment is needed for the over-threshold probability p^*. Since there are strong serial correlations up to around lag 50 (50 days), extremal indexes are also estimated using the intervals estimator.

To obtain different thresholds for the four stations, we fit a 75-th quantile regression model with different intercepts u_i's and a regression coefficient ν as follows:

$$u_{it} = u_i + \nu(T_{it} - 32),$$

where t denotes the day t^* in which non-zero snow depth observation was recorded ($X^*_{it^*} > 0$ for $t^* = 1, \ldots, 20454$ with $t^* = 1$ as July 1, 1959 and $t^* = 20454$ as June 30, 2016). Models 1–4 are fitted by iteratively merging smallest differences in u_i's and θ_i's. Results are summarized in Table 4.10. We choose Model 3 for further analysis since it appears to be the most parsimonious and providing an adequate fit for the data. The estimated coefficient associated with temperature in Model 3 is -0.1, implicating that maximum snow depth decreases by 0.1 inches on average as temperature increases by 1°F from 32°F. The standard errors for quantile regression parameters are computed using the "*se=boot*" option within the "*summary.rq*" function in the R package "*quantreg*." Again, the default "*xy-pair*" method is used in our studies. The standard errors for extremal indexes are computed using the bootstrapping strategy from Section 3.3. Here, the threshold exceedance series $\{Y_{it}\}$ exhibit both spatial and temporal dependences. For maximum likelihood estimate of scale parameter, naïve standard error (ignoring both spatial and temporal

dependences) and corrected standard error using Smith's method to account for both spatial and temporal dependences are reported. We note that the corrected GP standard errors are about three times larger than their naïve standard errors, implicating that both temporal dependence and spatial dependence contribute to additional uncertainty to GP parameter estimates.

Table 4.10: Stationary GP estimates for non-zero snow depth data with their associated standard errors in parentheses (*left*: naïve, *right*: corrected) for σ^* with the assumption that $\xi = 0$

		Model 1	Model 2	Model 3	Model 4
u	u_1	6.400 (0.209)	6.323 (0.211)	6.300 (0.209)	5.700 (0.102)
	u_2	5.600 (0.170)	5.613 (0.169)	5.500 (0.112)	
	u_3	5.300 (0.219)	5.355 (0.155)		
	u_4	5.400 (0.232)			
	ν	-0.100 (0.012)	-0.097 (0.012)	-0.100 (0.012)	-0.100 (0.012)
σ^*		4.149 (0.119, 0.381)	4.147 (0.119, 0.377)	4.113 (0.118, 0.373)	4.215 (0.122, 0.370)
θ	θ_1	0.149 (0.038)	0.149 (0.038)	0.152 (0.027)	0.175 (0.024)
	θ_2	0.155 (0.031)	0.155 (0.031)		
	θ_3	0.203 (0.038)	0.204 (0.029)	0.204 (0.029)	
	θ_4	0.206 (0.035)			
ℓ		-2950.912	-2957.669	-2954.969	-2928.933

To quantify a possible long-term linear trend in the threshold exceedances of snow depth series in the New York City area, we include a linear slope term in our quantile regression model. Different thresholds are obtained for the four stations by fitting a 75-th quantile regression model with different intercepts u_i's, a linear trend β, and a regression coefficient ν as follows:

$$u_{it} = u_i + \beta \left(\frac{t}{365.25 \times 10} \right) + \nu(T_{it} - 32).$$

It is assumed that β and ν are the identical across all stations. Results from

quantile regression, GP-based maximum likelihood estimation, and intervals estimator for extremal indexes are summarized in Table 4.11. For a similar reason to the stationary case, only Model 3 is considered for further analysis. The estimated temperature term in Model 3 is -0.102 inches per 1°F increase from 32°F, which is about two-fifths of the GEV temperature parameter estimate for the annual maximum snow depth series, paired with the previous day's maximum temperature series. This result implies that annual maximum snow depth observations are more susceptible to temperature changes than 75-th percentile snow depth observations. The trend estimate is 0.559 inches decade^{-1}, which is about 3.5 times greater than the GEV trend estimate. Unlike the GEV trend estimate, the quantile regression trend estimate is statistically significant, implying that the 75-th percentile snow depth has increased by 0.559 inches decade^{-1}. This is in contrast to the annual maximum snow depth, which shows an insignificant increase of about 0.157 inches decade^{-1}. Similar to

Table 4.11: Non-stationary GP estimates for non-zero snow depth data and their associated standard errors in parentheses (*left*: naïve, *right*: corrected) for σ^* with the assumption that $\xi = 0$

		Model 1	Model 2	Model 3	Model 4
u	u_1	4.725 (0.248)	4.755 (0.245)	4.767 (0.233)	3.924 (0.142)
	u_2	4.031 (0.175)	4.035 (0.175)	3.769 (0.143)	
	u_3	3.464 (0.214)	3.612 (0.182)		
	u_4	3.727 (0.209)			
	β	0.572 (0.046)	0.565 (0.045)	0.559 (0.042)	0.577 (0.044)
	ν	-0.103 (0.009)	-0.103 (0.009)	-0.102 (0.009)	-0.101 (0.011)
σ^*		4.081 (0.117, 0.408)	4.126 (0.120, 0.405)	4.109 (0.118, 0.407)	4.178 (0.120, 0.402)
θ	θ_1	0.188 (0.040)	0.188 (0.040)	0.173 (0.024)	0.193 (0.022)
	θ_2	0.161 (0.030)	0.161 (0.030)		
	θ_3	0.220 (0.043)	0.219 (0.031)	0.219 (0.031)	
	θ_4	0.218 (0.034)			
ℓ		-2928.777	-2844.294	-2932.307	-2954.788

snowfall, non-stationary GP models offer significantly improved fitting over stationary GP models. For this reason, only non-stationary GP models are considered for further analysis.

Estimated return levels for 25, 50, 75, and 100 years from the selected non-stationary GP model are obtained under the three different temperature scenarios as explained in the Section 4.3. These return level estimates are accompanied by their 95% confidence intervals computed from the BCa bootstrap method with 10000 replications. Estimation results are summarized in Table 4.12. All zero and nonzero observations are resampled using moving block bootstrap with the block size 50 in each bootstrap iteration due to high serial correlation. Furthermore, Ferro and Segers (2003) showed that $\hat{\theta}$ obtained by the intervals estimator converges in probability to true θ under some mild conditions. For this reason, the mean of $B = 10000$ samples of $\hat{\theta}^{(b)}$ where $b = (1, \ldots, B)$ are used in place of extremal index in each BCa bootstrap replication (i.e., standard bootstrap and delete-1 jackknife) of return levels. Return levels along with the 95% confidence intervals estimated from GEV and non-stationary GP models aligned fairly well. We do note that non-stationary GP models returned consistently lower return levels than GEV models which we believe is because we explicitly corrected serial correlation using extremal index in non-stationary GP settings. Finally, we find that the maximum snow depth observation recorded from the blizzard is less than a 25-year return level for Central Park and Newark; and between 50 and 75-year return levels for La Guardia and JFK under the 32°F temperature scenario. However, if we consider the 40°F temperature scenario, the actual snow observations are about equivalent to a 25-year return level for Central Park and Newark; and a 75-year return level for La Guardia and JFK. Furthermore, the actual snow observations are noticeably less than a 25-year return level for Central Park and

Newark; and between 25 and 50-year return levels for La Guardia and JFK under the minimum temperature scenario.

Table 4.12: Non-stationary GP snow depth return level estimates and their associated 95% BCa bootstrap confidence intervals under three different temperature scenarios

Minimum	Actual	K_{25}	K_{50}	K_{75}	K_{100}
Central Park	22	24.595 (21.826, 26.918)	28.881 (25.689, 31.590)	31.958 (28.357, 35.025)	34.544 (30.506, 37.978)
Newark	20	23.870 (21.613, 25.806)	28.157 (25.443, 30.499)	31.234 (28.108, 33.930)	33.819 (30.320, 36.916)
La Guardia	27	24.524	28.811	31.888	34.473
JFK	28	(22.194, 26.544)	(26.016, 31.234)	(28.662, 34.689)	(30.894, 37.696)
32°F	Actual	K_{25}	K_{50}	K_{75}	K_{100}
Central Park	22	22.249 (19.438, 24.586)	26.535 (23.235, 29.294)	29.612 (25.852, 32.806)	32.198 (28.032, 35.804)
Newark	20	21.116 (18.859, 23.036)	25.403 (22.588, 27.808)	28.479 (25.218, 31.335)	31.065 (27.391, 34.417)
La Guardia	27	22.076	26.363	29.439	32.025
JFK	28	(19.749, 24.101)	(23.465, 28.880)	(26.127, 32.387)	(28.300, 35.422)
40°F	Actual	K_{25}	K_{50}	K_{75}	K_{100}
Central Park	22	21.432 (18.564, 23.773)	25.719 (22.311, 28.517)	28.796 (24.949, 32.021)	31.382 (27.154, 35.061)
Newark	20	20.300 (18.000, 22.271)	24.587 (21.706, 27.072)	27.663 (24.341, 30.598)	30.249 (26.515, 33.685)
La Guardia	27	21.260	25.547	28.623	31.209
JFK	28	(18.883, 23.328)	(22.612, 28.117)	(25.250, 31.642)	(27.424, 34.718)

CHAPTER 5

DISCUSSION

5.1 Comments

To illustrate the necessity of using appropriate extreme value analysis techniques for extreme data, we fit a gamma distribution model to the NRCC's snow data. Only non-zero daily snowfall and snow depth observations are used to fit gamma distribution models. Then, the return period K_x, which is an inverse of return levels based on the actual observations from the blizzard, are computed. Table 5.1 summarizes the estimates of shape parameter α and rate parameter β for Gamma models and the corresponding return periods K_x in years. We note that Gamma models produce unrealistic return periods for the actual observations from the blizzard, therefore Gamma models do not provide adequate fits for snow data. This result illustrates why appropriate extreme value distributions need to be used for the snow data.

Table 5.1: Gamma estimation results and return years

	Snowfall			Snow depth		
	α	β	K_x	α	β	K_x
Central Park	0.847	0.286	444 years	1.472	0.321	93 years
Newark	0.701	0.264	135 years	1.543	0.359	70 years
La Guardia	0.788	0.302	926 years	1.536	0.394	738 years
JFK	0.780	0.320	3547 years	1.513	0.374	1195 years

We now consider return periods of the 2016 blizzard. To obtain useful return period scenarios, we use the three extreme value models as illustrated in Chapters 3 and 4: (1) stationary GEV block maxima method, (2) non-stationary GEV block maxima method with linear trend, and (3) non-stationary GP threshold exceedances method with extremal index and linear trend. We estimate return periods based on the actual snowfall observations from January 22, 2016 to January 24, 2016. Additionally, we estimate return periods based on the actual snow depth observations on January 24, 2016 paired with the maximum temperatures recorded on January 23, 2016. Table 5.2 summarizes these estimated return periods. All three methods returned very similar return years, suggesting that they all provide adequate fits to the data. Results indicate that snow depth observations from the blizzard were generally not as extreme compared to snowfall observations, which can be explained by increasing trends in temperatures paired with extreme snow depth observations.

Table 5.2: Return periods for the actual snowfall and snow depth observations from the January 2016 blizzard

Snowfall	Actual Observation	GEV		GP
		Stationary	Non-stationary	Non-stationary
Central Park	27.5 inches	50 years	40 years	26 years
Newark	24.5 inches	32 years	27 years	18 years
La Guardia	28.2 inches	55 years	43 years	32 years
JFK	30.4 inches	80 years	56 years	45 years
Snow depth	**Actual Observation**	**GEV**		**GP**
		Stationary	Non-stationary	Non-stationary
Central Park	22 inches & 35°F	22 years	23 years	25 years
Newark	20 inches & 36°F	17 years	17 years	21 years
La Guardia	27 inches & 36°F	45 years	48 years	57 years
JFK	28 inches & 35°F	49 years	53 years	64 years

Central Park and Newark had much lower snow depth observations recorded for the January 2016 blizzard (22 and 20 inches, respectively) compared to La Guardia and JFK (27 and 28 inches, respectively). This resulted in Central Park and Newark having much lower return periods computed from all three methods than La Guardia and JFK. Similarly, JFK had higher snowfall records (30.4 inches) than any of the other three stations (27.5, 24.5, and 28.2 inches), which resulted in JFK having the highest return periods for snowfall.

For comparison purposes, all data up to June 30, 2016 are now applied to the three selected extreme value models. For stationary and non-stationary GEV models, the addition of maximum snowfall and snow depth observations from snow year 2015 resulted in about a 20% increase of the maximum likelihood estimate of ξ. For non-stationary setting, although the estimated linear trends did increase by 16% for snowfall and 64% for snow depth, they were still not significantly different from zero with the corrected standard errors. Overall, these changes resulted in an increase of up to four inches in return levels all across stations. For non-stationary GP models, maximum likelihood estimates of σ and linear trends saw little to no increase after the addition of non-zero snowfall and snow depth observations from snow year 2015. However, it was still enough to cause an increase of up to three inches in return levels all across stations.

5.2 Conclusion

We found insignificant upward trends for both annual maximum snowfall and annual maximum snow depth series in the New York City area. It would be possible to obtain more significant trends in annual maxima when more data become available.

Even with this, we found that 87.5-th percentile for snowfall and 75-th percentile for snow depth series have significantly increased using the quantile regression and GP models.

It is evident that the blizzard of January 2016 in the New York City area was indeed an extreme event; however, this was only about a once-in-forty-years event in terms of return periods, and conventional extreme value analysis techniques were adequate enough in explicating the event. Our findings indicate that although magnitudes of snow events that are strong enough to be annual maxima may not increase over time, we may observe even higher 87.5-th percentile snowfall and 75-th percentile snow depth in the next 50 years or so.

5.3 Future Work

In this thesis, we focused on the analysis of extreme snow data using established statistical methodologies. There are several research topics originating from our work in this thesis that we can consider in the future.

We fitted extreme value models under the assumption that data are independent and identically distributed, then merged similar model parameters for simplicity. We extended Smith's method to adjust underestimated standard errors for model parameters due to spatial and temporal dependence. This approach is fairly straightforward and much easier to implement than directly accounting for spatial and temporal dependence in the modeling processes. We may consider comparing how this approach performs compared to approaches directly modeling spatial and temporal dependence.

When more weather stations from wider geographical regions are considered, our setting may lead to overfitting. The spatio-temporal modeling techniques, with some

region segmentation scheme, can be insightful in avoiding this overfitting problem for a possible extension of this study covering a wider geographical region (e.g., Le and Zidek, 2006; Cressie and Wikle, 2015). Some approaches that can be used to directly model for spatial and temporal dependence are max-stable process (e.g., Buishand et al., 2008; Blanchet and Davison, 2011; Reich and Shaby, 2012) and Bayesian hierarchical modeling (e.g., Cooley, Nychka, and Naveau, 2007; Sang and Gelfand, 2009; Eastoe, 2009; Banerjee, Carlin, and Gelfand, 2015).

As discussed in Section 3.2, choosing appropriate thresholds in GP models is important, and the process of making those choices can be subjective. Northrop et al. (2017) used a Bayesian cross-validation method to quantify the bias-variance tradeoff when selecting a stationary threshold under the assumption that data are independent and identically distributed. We may consider developing a method applicable for the selection of non-stationary thresholds for identically distributed data with spatial and temporal dependence.

Frequency and severity of extreme snowfall and snow depth can be affected by various factors, such as elevation, spatial region, temperature, humidity, or sunshine duration. Although temperature was included in snow depth analysis, it was only considered in the location parameter estimation process. Multivariate extensions of this thesis can provide additional meaningful information in gaining deeper understanding in extreme snow events (e.g., Coles and Tawn, 1991; Schlather and Tawn, 2003; Blanchet and Davison, 2011).

Rust et al. (2011) showed that weakly correlated series also satisfy the Fisher-Tippett-Gnedenko theorem. This means that block maxima methods can be applied to not only uncorrelated series but also weakly correlated series. Furthermore, extremal index can be used to allow GEV and GP distributions to be applied to the

short memory time series (Leadbetter, 1983; Coles, 2001). However, whether or not similar results can be obtained when dealing with the long memory time series has not been well studied at this point. We leave these explorations for future work.

REFERENCES

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2015). *Hierarchical modeling and analysis for spatial data* (2nd ed.). Boca Raton, FL: CRC Press.

Blanchet, J. and Davison, A. C. (2011). Spatial modeling of extreme snow depth. *Ann. Appl. Stat.*, 5(3), 1699–1725.

Blanchet, J., Marty, C., and Lehning, M. (2009). Extreme value statistics of snowfall in the Swiss Alpine region. *Water Resour. Res.*, 45. W05424.

Buishand, T. A., de Haan, L., and Zhou, C. (2008). On spatial extremes: with application to a rainfall problem. *Ann. Appl. Stat.*, 2(2), 624–642.

Burakowski, E. A., Wake, C. P., Braswell, B., and Brown, D. P. (2008). Trends in wintertime climate in the northeastern United States: 1965-2005. *J. Geophys. Res.*, 113. D20114.

Casella, G. and Berger, R. L. (2001). *Statistical inference* (2nd ed.). Pacific Grove, CA: Duxbury.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. London, UK: Springer-Verlag.

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *J. R. Stat. Soc. B*, 53(2), 377–392.

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *J. Amer. Statist. Assoc.*, 102(479), 824–840.

Cressie, N. and Wikle, C. K. (2015). *Statistics for spatio-temporal data*. Hoboken, NJ: John Wiley & Sons.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *J. R. Stat. Soc. B*, 52(3), 393–442.

DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals (with discussion). *Statist. Sci.*, 11(3), 189–228.

Eastoe, E. F. (2009). A hierarchical model for non-stationary multivariate extremes: a case study of surface-level ozone and NO_x data in the UK. *Environmetrics*, 20(4), 428–444.

Efron, B. (1987). Better bootstrap confidence intervals (with discussion). *J. Amer. Statist. Assoc.*, 82(397), 171–200.

Fawcett, L. and Walshaw, D. (2007). Improved estimation for temporally clustered extremes. *Environmetrics*, 18(2), 173–188.

Fawcett, L. and Walshaw, D. (2012). Estimating return levels from serially dependent extremes. *Environmetrics*, 23(3), 272–283.

Ferro, C. A. T. and Segers, J. (2003). Inference for clusters of extreme values. *J. R. Stat. Soc. B*, 65(2), 545–556.

Givens, G. H. and Hoeting, J. A. (2013). *Computational statistics* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Huntington, T. G., Hodgkins, G. A., Keim, B. D., and Dudley, R. W. (2004). Changes in the proportion of precipitation occurring as snow in New England (1949–2000). *J. Clim.*, 17(13), 2626–2636.

Jonathan, P., Ewans, K., and Randell, D. (2014). Non-stationary conditional extremes of northern North Sea storm characteristics. *Environmetrics*, 25(3), 172–188.

Koenker, R. (2005). *Quantile regression*. Cambridge, UK: Cambridge University Press.

Koenker, R. (2016). *quantreg: Quantile Regression*. R package version 5.29.

Kunkel, K. E., Palecki, M., Ensor, L., Hubbard, K. G., Robinson, D., Redmond, K., and Easterling, D. (2009). Trends in twentieth-century U.S. snowfall using a quality-controlled dataset. *J. Atmos. Oceanic Technol.*, 26(1), 33–44.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.*, 17(3), 1217–1241.

Kyselỳ, J., Picek, J., and Beranová, R. (2010). Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. *Glob. Planet. Change*, 72(1), 55–68.

Le, N. D. and Zidek, J. V. (2006). *Statistical analysis of environmental space-time processes*. New York, NY: Springer, New York.

Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete*, 65(2), 291–306.

Lee, J., Li, S., and Lund, R. (2014). Trends in extreme U.S. temperatures. *J. Clim.*, 27(11), 4209–4225.

López-Moreno, J. I., Goyette, S., Vicente-Serrano, S. M., and Beniston, M. (2011). Effects of climate change on the intensity and frequency of heavy snowfall events in the Pyrenees. *Clim. Change*, 105(3), 1–20.

Makkonen, L., Ruokolainen, L., Räisänen, J., and Tikanmäki, M. (2007). Regional climate model estimates for changes in Nordic extreme events. *Geophysica*, 43(1–2), 25–48.

McCormick, W. P. and Qi, Y. (2000). Asymptotic distribution for the sum and maximum of Gaussian processes. *J. Appl. Probab.*, 37(4), 958–971.

National Weather Service (2013). *Snow measurement guidelines for National Weather Service surface observing programs*. Silver Spring, MD: Office of Climate, Water and Weather Services.

National Weather Service (2016). *Evaluation of reported snowfall at local climatological data stations during the East Coast blizzard of January 22-23, 2016*. Silver Spring, MD: Office of Climate, Water and Weather Services.

NOAA National Centers for Environmental Information (NCEI) (2017). *U.S. billion-dollar weather and climate disasters*. `http://www.ncdc.noaa.gov/billions/`.

Northrop, P. J., Attalides, N., and Jonathan, P. (2017). Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity. *J. R. Stat. Soc. C*, 66(1), 93–120.

Northrop, P. J. and Jonathan, P. (2011). Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*, 22(7), 799–809.

O'Gorman, P. A. (2014). Contrasting responses of mean and extreme snowfall to climate change. *Nature*, 512, 416–418.

Panagoulia, D., Economou, P., and Caroni, C. (2014). Stationary and nonstationary generalized extreme value modelling of extreme precipitation over a mountainous area under climate change. *Environmetrics*, 25(1), 29–43.

Reich, B. J. and Shaby, B. A. (2012). A hierarchical max-stable spatial model for extreme precipitation. *Ann. Appl. Stat.*, 6(4), 1430–1451.

Rust, H. W., Kallache, M., Schellnhuber, H. J., and Kropp, J. P. (2011). Confidence Intervals for Flood Return Level Estimates Assuming Long-Range Dependence. In J. Kropp & H. J. Schellnhube (Eds.), *In Extremis: Disruptive Events and Trends in Climate and Hydrology* (pp. 60–88). Heidelberg, Germany: Springer-Verlag.

Sang, H. and Gelfand, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environ. Ecol. Stat.*, 16(3), 407–426.

Schlather, M. and Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values: properties and inference. *Biometrika*, 90(1), 139–156.

Smith, R. L. (1990). *Regional estimation from spatially dependent data*. Unpublished manuscript.

APPENDIX A

GRADIENTS OF GEV AND GP DISTRIBUTIONS

In Section 3.4, the empirical covariance matrix of the observed gradient values are used to estimate V in (3.8) to obtain the estimated corrected variance of maximum likelihood estimates in (3.9). We show the expressions for gradients of GEV and GP distributions in this Appendix.

Gradient of the log likelihood function for GEV distribution is obtained by taking partial derivatives of (3.2) in terms of each unknown parameter as:

$$\begin{aligned}
\frac{\partial l(\mu,\sigma,\xi)}{\partial \mu} &= -\frac{1}{\sigma}\sum_{i=1}^{n}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-1-1/\xi} + (1+\xi)\sum_{i=1}^{n}\frac{1}{\xi(x_i-\mu)+\sigma},\\
\frac{\partial l(\mu,\sigma,\xi)}{\partial \sigma} &= -\frac{n}{\sigma} - \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma^2}\right)\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-1-1/\xi}\\
&\quad + \left(\frac{1+\xi}{\sigma}\right)\sum_{i=1}^{n}\frac{x_i-\mu}{\xi(x_i-\mu)+\sigma},\\
\frac{\partial l(\mu,\sigma,\xi)}{\partial \xi} &= \frac{1}{\xi^2}\sum_{i=1}^{n}\log\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]\\
&\quad - \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-1}\\
&\quad - \frac{1}{\xi^2}\sum_{i=1}^{n}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-1/\xi}\log\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]\\
&\quad + \frac{1}{\xi}\sum_{i=1}^{n}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-1/\xi}\left(\frac{x_i-\mu}{\sigma}\right)\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-1},
\end{aligned}$$

if $\xi \neq 0$, and

$$\frac{\partial l(\mu,\sigma)}{\partial \mu} = \frac{1}{\sigma}\sum_{i=1}^{n}\left\{1-\exp\left[-\left(\frac{x_i-\mu}{\sigma}\right)\right]\right\},$$
$$\frac{\partial l(\mu,\sigma)}{\partial \sigma} = -\frac{n}{\sigma}+\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma^2}\right)\left\{1-\exp\left[-\left(\frac{x_i-\mu}{\sigma}\right)\right]\right\},$$

if $\xi = 0$.

Gradient of the log likelihood function for GP distribution is obtained by taking partial derivatives of (3.4) in terms of each unknown parameter as:

$$\frac{\partial l(\sigma^*,\xi)}{\partial \sigma^*} = -\frac{n}{\sigma^*}+\left(\frac{1+\xi}{\sigma^*}\right)\sum_{i=1}^{n}\frac{\left(\frac{x_i-\mu}{\sigma^*}\right)}{\xi\left(\frac{x_i-\mu}{\sigma^*}\right)+1},$$
$$\frac{\partial l(\sigma^*,\xi)}{\partial \xi} = -\left(1+\frac{1}{\xi}\right)\sum_{i=1}^{n}\frac{\left(\frac{x_i-\mu}{\sigma^*}\right)}{\xi\left(\frac{x_i-\mu}{\sigma^*}\right)+1}+\frac{1}{\xi^2}\sum_{i=1}^{n}\log\left[\xi\left(\frac{x_i-\mu}{\sigma^*}\right)+1\right],$$

if $\xi \neq 0$, and

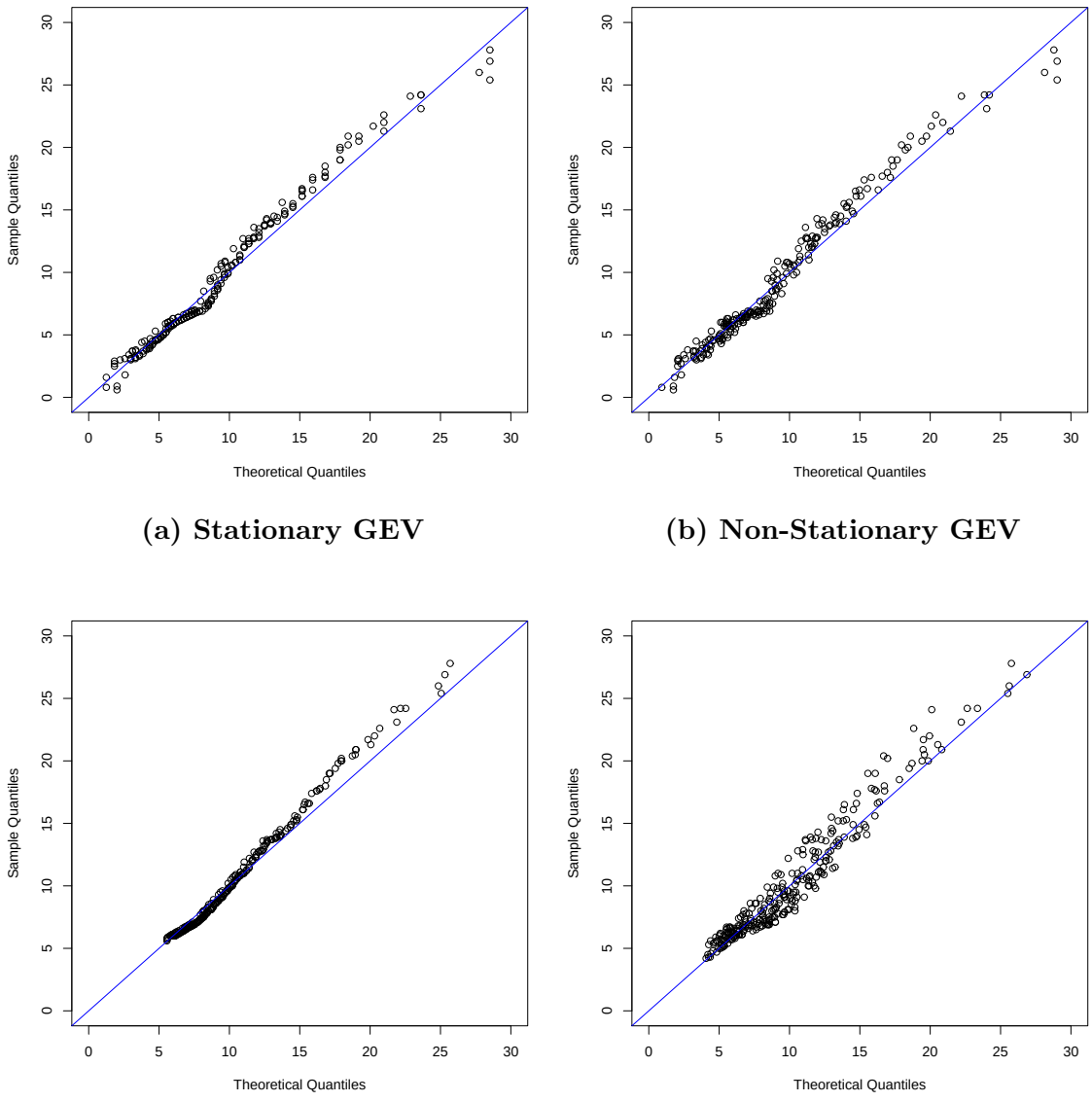
$$\frac{\partial l(\sigma^*)}{\partial \sigma^*} = -\frac{n}{\sigma^*}+\frac{1}{\sigma^{*2}}\sum_{i=1}^{n}x_i,$$

if $\xi = 0$.

APPENDIX B

QUANTILE-QUANTILE PLOTS FOR GEV AND GP MODELS

To visually examine if the snow data follow appropriate extreme value distributions, Quantile-Quantile plots of fits of (1) stationary GEV, (2) non-stationary GEV, (3) stationary GP, and (4) non-stationary GP distributions from Chapter 4 to snowfall and snow depth data are produced. Figures B.1 and B.2 suggest that annual maximum snowfall and annual maximum snow depth series follow the GEV distribution, and exceedances over the 87.5-th percentile snowfall and exceedances over the 75-th percentile snow depth series follow the GP distribution.



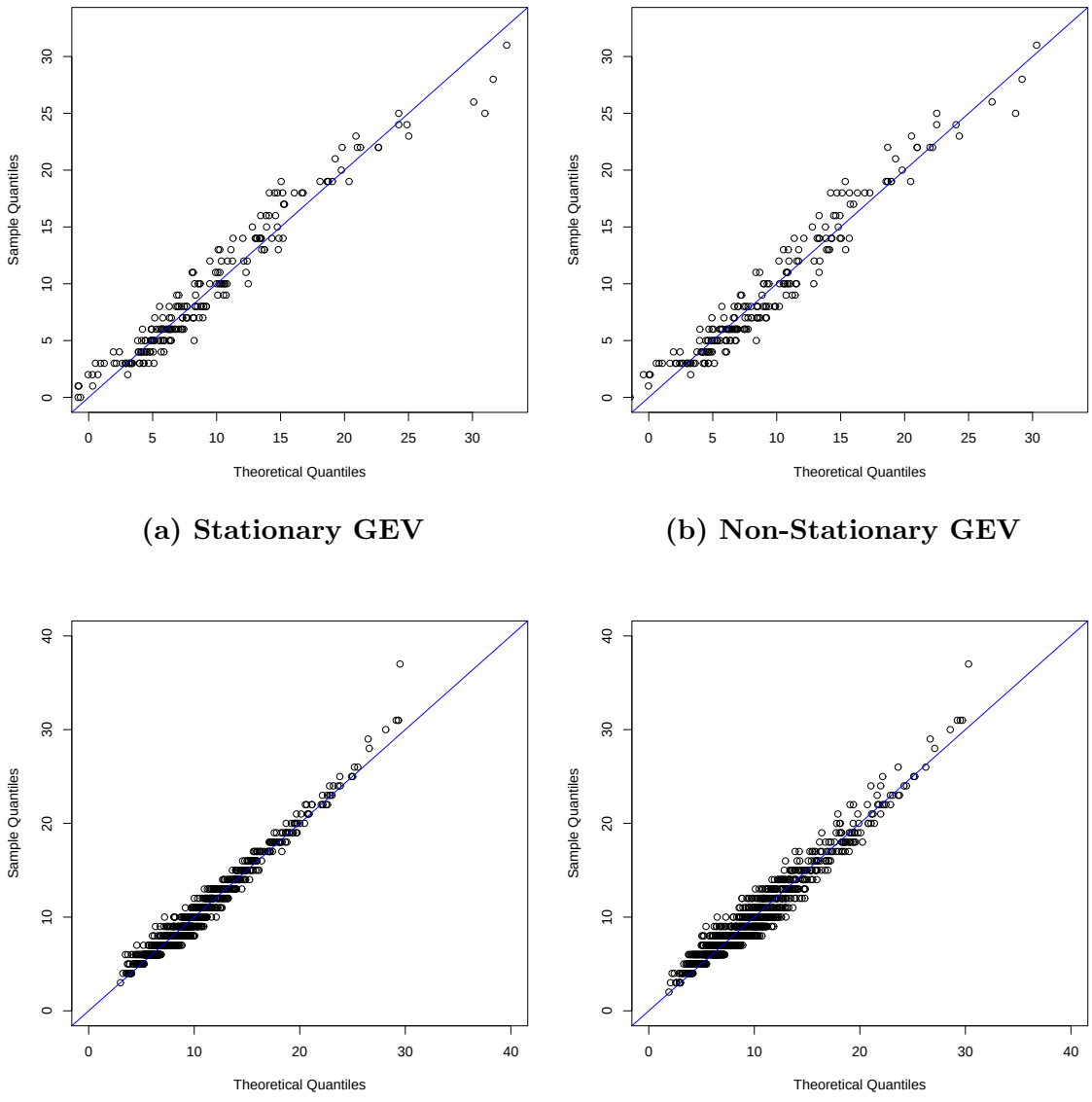
(a) Stationary GEV

(b) Non-Stationary GEV

(c) Stationary GP

(d) Non-Stationary GP

Figure B.1: Quantile-Quantile plots of fit of GEV to annual maximum snowfall and GP to exceedances over the 87.5-th percentile snowfall



(a) Stationary GEV

(b) Non-Stationary GEV

(c) Stationary GP

(d) Non-Stationary GP

Figure B.2: Quantile-Quantile plots of fit of GEV to annual maximum snow depth and GP to exceedances over the 75-th percentile snow depth