An Investigation of Non-stationary Nature of Environmental Data

by

Maulin Riteshbhai Amin

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Civil Engineering

Waterloo, Ontario, Canada, 2019

Author's Declartion

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Several investigations have shown that the Earth's climate is experiencing change, as evident by rising global temperatures, receding ice sheets and reducing glacier sizes. There is also empirical observation supporting changes in patterns of precipitation, high wind events and snow fall as the climate change effects become more pronounced. IPCC has outlined various emission scenarios under which the mean global temperature can increase from 2 to 8 °C, which is expected to have a significant impact on the social well-being of human population on the planet.

Structures are designed to withstand environment generated loads, such as high wind, snow, rain, ice and flood, over the design life of the structure. Currently, structural design codes assume that such environmental loads are stationary over the entire life of the structure ranging from 50 to 100 years. Since the assumption of stationary climate may not be tenable in future, researchers are taking in interest in modelling of non-stationary effects in environmental date for developing appropriate design load calculation methods.

The main objective of this thesis is to investigate probabilistic methods for modelling the non-stationary nature of environmental data and estimating design values, i.e., upper percentiles of extreme value distribution. The thesis focuses on non-stationary version of two commonly used methods, namely, annual maxima method and the peak over threshold (POT) method. The annual maxima method uses a non-stationary Gumbel distribution with time dependent parameters. The other more general model is based on stochastic process theory in which the arrival of events and the event magnitude are treated as probabilistic variables. The arrival process is modelled as the non-homogeneous Poisson process. The proposed approach is a generalization of currently used POT method which relies on asymptotic extreme value theory. Statistical test are applied to evaluate the presence of non-stationary effects.

The proposed approaches are illustrated by analysing the time series of daily maximum wind speed data from a selected set of stations. For the 10 weather stations analysed, the data analysis show that wind speed is decreasing at an average of 0.006625 KMPH annually over the 80 year period duration from 2020 to 2100.

Acknowledgements

I would like to thank my supervisor Dr. Mahesh Pandey for guiding me in the completion of this thesis. His invaluable technical insight through the course of this thesis assisted me a lot in the research process. It has been an honour working with him. I would also like to thank my thesis reviewers, Dr. Wei-Chau Xie and Dr. Shunde Yin, for their time and comments

My journey at the University of Waterloo would not have been possible without Anita Agrawal. I am grateful that she encouraged me to pursue this path. Thank you.

I would also like to extend my thanks to Noldainerick Manzana for helping me review and edit my thesis. I am also grateful to my friends Krishna Gelda, Harrison Oakes, Christopher Cunningham, John Wage, and Artyom Ogay for being there for me through thick and thin and providing support while I was working on this thesis.

Finally, I would also like to extend my gratitude to my family and my parents. They have contributed immensely to build me from the ground up.

Dedication

To Mum and Dad.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 General

Environmental loads of wind and precipitation are one of the dominant loads on structures in Canada as defined in by National Research Council Canada. As reported by Johnson (2017), 103 people died in the US because of hurricane Harvey and Irma. Moreover, these hurricanes also resulted in a combined structural damage of $200 billion. By analysing the historical data, these rare, extreme events can be predicted allowing us to make preparations to take preventative measures against damage to properties and life.

This means that any engineering work needs to account for extreme conditions; therefore, engineers, scientists, and statisticians have been drawn toward the field of extremes. The field of extremes, by definition, deals with the maxima or the minima of random variables.

The advance in technology has enabled NASA to collect enough facts to be able to prove that the global climate is changing rapidly (NASA, 2019). The evidence of climate change includes: rise in global temperature, warming oceans, shrinking ice sheets, retreating glaciers, decreasing snow cover, rise in sea level, declining Arctic sea ice, acidification of ocean, and extreme events. This thesis concentrates on extreme events as it is one of the factors that contribute toward the formulation of building codes of Canada. Apart from the improvement in technology, climate change is also one of the other factors that require the building codes to be updated regularly. That means that the future buildings we live and work in would need to be designed according to the updated building codes that account for the climate change.

The main objective of this thesis is to improve the extreme wind and snow estimation while also taking into account the non-stationary effects of climate change.

1.2 Motivations and Objectives

Management of the risk focusses on modelling the tail events; that is, the events that have a low probability, but have a high impact. Usually, the central part of our distribution becomes our focus while trying to estimate the whole distribution of losses, and the tails get neglected. Therefore, it is beneficial to have a model focussing on the tails of a distribution. Extreme Value Theory (EVT) is one such model.

In other words, the distribution of rare events can be described using EVT. It is especially useful in environmental, insurance, and financial applications. Each of these applications have the focal interest on the occurrence of rare, extreme events. Thus, EVT is used to study the behaviour of tail distributions when the scarcity of data makes it hard to infer the extremes. One of the popular EVT models is the Generalized Pareto Distribution. Over an adequately high threshold, it provides a reliable extrapolation of the exceedances. This is classified as the Stochastic Process Model. For the correct application of this method, it is very crucial to select the right threshold because it classifies which portion of the data is extreme and which is not.

A compromise between the actual operating conditions and the capacity of the structure provides the design values that ensure safety. At the end of the day, it is the aim of the designer to design a structure that withstands any catastrophic event during its lifetime. Therefore, it is always better to overestimate the capacity in comparison to the operating loads. In other words, the maximum value of the operating load should be less than the minimum value of the capacity of the structure or element. Thus, finding the values of the extreme distributions plays an important role in structural engineering. This is also why it is important to analyse the distribution of extremes as those are the values that affect the failure of any system (An, 2006).

The building codes list the calibrations necessary to ensure a consistent level of safety. However, the current code assumes that the environmental loads are generated by a stationary processes. We have already discussed the evidence of change in climate. This change in climate should also be accounted for while writing building design codes. Rather than assuming that the distribution parameters are time invariant, it becomes necessary to consider time variant distribution parameters to account for the climate change non-stationarity. This poses the research questions like: How to model non-stationary loads?

What should be the format of reliability-based design code under non-stationary loads? What should be the basis for specifying the target reliability? The scope of this thesis is to answer the first question by empirically investigating the non-stationary nature of wind speed data.

1.3 Environmental Data

Environment and Climate Change Canada (ECCC) hosts a File Transfer Server (FTP) with all historical weather data for numerous locations across Canada. Some of the data available to download include: temperature, precipitation, degree days, relative humidity, wind speed and direction, monthly summaries, averages, extremes, etc.

It can be accessed using the URL:

`climate.weater.gc.ca`

A `ReadMe.txt` file with instructions to download data automatically, after writing a few lines of code, is located at:

`ftp://ftp.tor.ec.gc.ca/Pub/Get_More_Data_Plus_de_donnees/Readme.txt`

There are several methods for extreme value analysis. As we have already discussed, the current analysis mainly focuses on the Gumbel model and the Stochastic Process model. The latter requires the selection of a distribution fitting the data. For this purpose, probability paper plot (PPP) method was employed (Saeed Far and Abd. Wahab, 2016). This model is also the simplest of the statistical model that can be used for extreme value analysis. Here, we first reduce the series of observations into a sequence of peaks-over-threshold values Z_1, ... , Z_M (Ross, 1987). The current wind speed data acquired from Environment Canada contains daily maximum wind speed data from which we delete the data below the chosen threshold z_0; thus, it results into a data set of peaks-over-threshold values of Z_1, ... , Z_M.

The data used for the analysis of wind speed measured only the wind speed, so the variables of wind direction and surface pressure were ignored. It was also evident from the change in station IDs that there was a change in measuring instrument. It can be assumed

that these changes mostly were the changes of instrumentation which are quite normal when it comes to technology. Even after the change in the Station ID, the station still was located at the same latitude and longitude; thus, confirming that it was just a change of the measuring instrument and the phasing out of the old one which called for a change in Station ID. The historical data also confirms this phasing out of the instruments by listing two instruments that share the same latitude and longitude for a year or two.

1.4 Thesis Organization

The five chapters of the thesis effectively present the research objectives along with providing relevant background information and supporting material.

Chapter 1 introduces the motivations and objectives of the thesis along with providing a brief general introduction of the requirement of the prediction/estimation methods. It also provides information about the type and source of data for the analysis shown in the thesis.

Chapter 2 details the background information necessary to develop a foundation knowledge for research. It then defines Extreme Value Theory along with the several Extreme Value distributions. Next, a brief summary of maximum likelihood method and probability paper plots is provided followed by a quick recap of the Poisson process. The chapter ends with an introduction to the two types of estimation methods used in the thesis which also acts an introduction to Chapter 3.

Chapter 3 discusses the application of Gumbel Model and Stochastic Process Model to the wind data. It also carries out hypothesis tests to test for the homogeneous Poisson process. These two models are used to calculate the stationary and non-stationary parameters that would be used in the following chapter.

Chapter 4 uses the stationary and non-stationary parameters calculated in the previous chapter to calculate the forecast high wind events until the year 2100. This helps us predict the nature of wind speed in the coming years.

Chapter 5 concludes the thesis by highlighting the major findings of the research along with supplying future recommendations and scope of this topic.

Chapter 2

Background

2.1 Introduction

Natural disasters force us to think whether knowing the future would have helped us prevent or prepare against the forces of natures. These natural calamities could include extraordinary dry spells, numerous forest fires, devastating earthquakes, massive snowfalls, and destructive hurricanes or floods. Therefore, a country like Japan that lies on the epicentre of earthquakes may want to design buildings that could withstand higher magnitudes of earthquakes so that the destruction to life and property could be avoided or could be kept as low as possible. Similarly, engineers in Canada would want to concentrate on wind and snow data as these two forces of nature result in the most amount of losses (Danard et al., 2003).

The force of a wind gust affects the wear and tear of buildings; therefore, the engineers are required to design structures that are able to withstand these winds. Hence, it is necessary to estimate the extreme values of these wind speeds during the design phase of the structure. This also requires the estimation of probabilities of such destructive winds during the life cycle of a structure. Extreme Value Analysis is employed in these scenarios to specify the probabilities of exceedances (Saeed Far and Abd. Wahab, 2016).

Historically, EVT has been applied mostly to flood related events. Haan (2006) even gives an example of the town of Delfzijl in Netherlands where 1877 severe storm surges had been observed over more than 100 years. As about half of Netherlands is below sea level it has to be protected against the sea by dikes. These sea dikes have to be able to withstand the flooding that might be caused by the severe storm surges. This flooding is mainly

a set of independent observations that are recorded under similar conditions of high-tide levels during the storm events. To avoid flooding, the problem involves designing the dikes high enough that the probability of flooding in a given year is 10^{-4}. For this problem, Haan (2006) suggests that using EVT provides a solid theoretical basis and framework for extrapolation. This is because EVT restricts the behaviour of the distribution function in the tail. EVT is quite general if we look at just one side of the coin. However, when we flip the coin we find out that EVT is sufficiently precise enough to help with extrapolation or estimation (Bruce, 2017) .

Currently the extreme events are fitted by stationary Gumbel distribution, but the change in climate is expected to introduce non-stationary loads; so, the data requires modelling using non-stationary loads to account for the change in climate. The two estimation methods that are used to analyse the wind data in this thesis are

1. Gumbel model or Annual maxima method
2. Stochastic Process model or Peaks-Over-Threshold method.

Both of these methods have been analysed for both stationary and non-stationary loads so that the effect of non-stationarity could be observed for both the instances.

Sufficiently large values of independent and identically distributed variates, as stated by a fundamental theorem in extreme value theory, are described by one of the three extreme value distributions: the Frechet distribution, the Gumbel distribution, and the reverse (negative) Weibull distribution. For a detailed exposition, refer Castillo (1988).

2.2 Extreme Value Theory

2.2.1 Introduction

Before applying EVT to a specific problem, it is important to ask as to when does EVT apply in a formal sense? The answer, fundamentally, is when the distribution to be modelled consists of extrema. From an overall distribution of data, the sampled minima or maxima are classified as extrema. In the words of Coles (2001) "The distinguishing feature of an extreme value analysis is the objective to quantify the stochastic behavior of a process at unusually large—or small—levels."

As listed by Castillo et al. (2005), some of the applications of EVT include: Ocean Engineering, Structural Engineering, Hydraulics Engineering, Meteorology, Material Strength,

Fatigue Strength, Electrical Strength of Materials, Highway Traffic, Corrosion Resistance, and Pollution Studies.

Some applications of EVT require us to estimate population central characteristics whereas some require us to estimate the maximum or minimum. Some examples of estimating the population central chracteristics include average rainfall, average temperature, median income, etc. Contrarily, estimating the maximum or minimum includes maximum flooding or minimum dam height, maximum earthquake intensity or minimum concrete strength, etc. It is important to consider extremes for engineering design because maximum values such as winds, floods, earthquakes, waves, etc. and minimum values such as strength, stress, etc. are the two key parameters that lead to the failure of any engineering system. Therefore, having a knowledge of the distribution of these maximum and minimum values results in getting good solutions to engineering design problems.

As discussed previously, engineering design always involves a compromise between safety and cost. There has to a perfect balance of both otherwise it may result in either unnecessary waste of money or a high probability of failure. It is also important to note that, in most cases, there is also a lack of availability of enough data (Davis et al., 1985).

Let us now assume $(s_1, s_2, ...)$ to be a sequence of i.i.d. samples. Thus, the maximum over an n-observation period is thus:

$$M_n = max(s_1, s_2, ...). \tag{2.1}$$

The approximate behaviour of M_n follows from the limit arguments associated with n approaching infinity for large values of n. An entire family of models can be calibrated from this observation via the observed extrema values of M_n. This chapter was summarized using the following references: Coles (2001), Haan (2006), Gumbel (1960), Reiss (2001), (Yue, 2000), Singh (2007) and Castillo et al. (2005).

2.2.2 The Extreme Value Theorem

The first Extreme Value Theorem is also known as the Fisher-Tippett Theorem. EVT indicates that the distribution of maximum or minimum can only assume limited forms given a well-behaved overall distribution just as the central limit theorem indicates that the random variable generated from a certain stochastic processes follows normal distribution. Let us first define this theorem to find the appropriate form.

Theorem 2.1.1

Assume $(s_1, s_2, ...)$ is a sequence of i.i.d samples and $M_n = max\{s_1, ...s_n\}$. If a sequence of pairs of real numbers (a_n, b_n) exists such that each $a_n > 0$ and

$$\lim_{x\to\infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F(x) \tag{2.2}$$

then if F is a non-degenerate distribution function, it belongs to one of three extreme value distributions: the Gumbel (I), Fréchet (II), or Reversed Weibull (III) distribution. Gumbel and Fréchet are for unbounded distributions and Reversed Weibull for bounded.

For modelling minima, a special consideration must be made. Minima can be transformed to maxima via $z_i = -s_i$, since the Fisher-Tippett Theorem applies to maxima.

Consider F to be an unknown distribution function of a random sample s. Thus, the conditional excess distribution function F_u of the variable s above a threshold u is defined as:

$$F_u(x) = P(s - u \leq x | s > u) = \frac{F(u + x) - F(u)}{1 - F(u)} \tag{2.3}$$

for $0 \leq x \leq s_F - u$ where s_F is either the finite or infinite right endpoint of the underlying distribution F . F_u describes the distribution of the excess values over u. Given this definition, the Pickands-Balkema-de Haan Theorem can be stated.

Theorem 2.1.2

Let $(s_1, s_2, ...)$ be a sequence of i.i.d. samples, and let F_u be their conditional excess distribution function. For a large class of underlying distribution functions F, and a large u, F_u is well approximated by the Generalized Pareto distribution:

$$F_u(x) \to GPD_F(x), as \text{ u} \to \infty \tag{2.4}$$

The Generalized Pareto distribution GPD_F is defined in the next section.

2.2.3 EVT Distributions

The Fisher-Tippett Theorem leads to three primary EVT distributions, and a number of other related distributions. Beyond the three primary extreme value distributions (Gumbel, Fréchet, and Reversed Weibull), the GEV, Rayleigh, and Pareto distributions are also useful for tail modelling. For the following distributions, assume κ, λ, and τ are the shape, scale, and location parameters, respectively.

Gumbel Distribution

The Gumbel is a distribution that applies when the data to be modeled are unbounded. It is used for modeling maxima (or minima, if the random variables are negated). Note that the shape of the Gumbel is not determined by the parameters—only the location and scale (which must be positive) are defined. The Probability Density Function (PDF) of the Gumbel distribution is given as:

$$f(x) = \frac{1}{\lambda} e^{-(z+e^{-z})} \tag{2.5}$$

where

$$z = \frac{x-\tau}{\lambda} \tag{2.6}$$

The CDF of the Gumbel distribution is given as:

$$F(x) = e^{-e^{-(x-\tau)/\lambda}} \tag{2.7}$$

Fréchet Distribution

The Fréchet is a distribution that applies when the data to be modeled are bounded from below and a heavy upper tail is desirable. Like the Gumbel distribution, it is used for modelling maxima (or minima, if the random variables are negated). The PDF of the Fréchet distribution is given as:

$$f(x) = \frac{\kappa}{\lambda} \left(\frac{x-\tau}{\kappa} \right)^{-1-\kappa} e^{-\left(\frac{x-\tau}{\lambda} \right)^{\kappa}} \tag{2.8}$$

The CDF of the Fréchet distribution is given as:

$$F(x) = e^{-\left(\frac{x-\tau}{\lambda} \right)^{\kappa}} \tag{2.9}$$

Weibull Distribution

The Weibull is a distribution that applies when the data to be modeled are bounded from below and the shape and scale parameters are positive. In contrast to the Gumbel and Fréchet distributions, the Weibull is used for modeling minima. The PDF of the two-parameter Weibull distribution is given as:

$$f(x) = \begin{cases} \frac{\kappa}{\lambda}\left(\frac{x}{\lambda}\right)^{\kappa-1} e^{(x/\lambda)^{\kappa}}, & x \geq 0 \\ 0, & x < 0. \end{cases} \tag{2.10}$$

The CDF of the Weibull distribution is given as:

$$F(x) = \begin{cases} 1 - e^{(x/\lambda)^{\kappa}}, & x \geq 0 \\ 0, & x < 0. \end{cases} \tag{2.11}$$

The Rayleigh distribution is a special case of the Weibull distribution, where the shape parameter $\kappa = 2$.

$$f(x) = \frac{x}{\lambda^2} e^{-x^2/2\lambda^2} \tag{2.12}$$

The CDF of the Rayleigh distribution is given as:

$$F(x) = 1 - e^{-x^2/2\lambda^2} \tag{2.13}$$

Fragoso and Turk [2013] have argued that the Rayleigh distribution can reduce sensitivity to the underlying distribution of data, and be computed efficiently, because the scale is the only parameter that must be computed.

Reverse Weibull Distribution

The reverse Weibull distribution is simply the opposite of the Weibull's non-degenerate distribution function, and is appropriate when the data are bounded from above. It follows directly from the Fisher-Tippett Theorem, and is thus used to model maxima.

$$F(x) = \begin{cases} e^{(x/\lambda)^{\kappa}}, & x < 0 \\ 1, & x \geq 0. \end{cases} \tag{2.14}$$

Generalized Extreme Value Distribution

The Generalized Extreme Value Distribution (GEV) is a generalization of the Gumbel, Fréchet and reverse Weibull distributions (all three are maximum EVT distributions). The PDF of the GEV distribution is

$$GEV(t) = \begin{cases} \frac{1}{\lambda} e^{-\nu^{-1/\kappa}} \nu^{-(1/\kappa+1)}, & \kappa \neq 0 \\ \frac{1}{\lambda} e^{-(x+e^{-x})}, & \kappa = 0. \end{cases} \tag{2.15}$$

where

$$x = \frac{t-\tau}{\lambda}, \nu = \left(1 + \kappa \frac{t-\tau}{\lambda}\right) \tag{2.16}$$

Different values of the shape parameter yield the extreme value type I, II, and III distributions. Specifically, the three cases $\kappa = 0$, $\kappa > 0$, and $\kappa < 0$ correspond to the Gumbel (I), Frechet (II), and Reversed Weibull (III) distributions. Thus the sign of κ can be used as an indicator of which distribution to choose. The CDF of the GEV is given as:

$$GEV(t) = \begin{cases} e^{-\left(1+\left(\frac{x-\tau}{\lambda}\right)\right)^{-1/\kappa}}, & \kappa \neq 0 \\ e^{-e^{-(x-\tau)/\lambda}}, & \kappa = 0. \end{cases} \tag{2.17}$$

Generalized Pareto Distribution

The Generalized Pareto Distribution (GPD) is a common distribution used to model the tails drawn from other distributions. It follows from the second extreme value theorem (i.e., the Pickands-Balkema-de Haan Theorem). Specifically, the modelling accounts for exceedances over a high threshold ($x - t$, where t is a threshold).

$$f(x) = \frac{1}{\lambda}(1 + \kappa z)^{-(1/\kappa+1)} \tag{2.18}$$

where

$$z = \frac{x-\tau}{\lambda} \tag{2.19}$$

2.3 Probability Paper Method

This section introduces probability plots and shows how they can be used to select a parent distribution for the given sample.

In daily practice, graphical methods are one of the important tools used by scientists and engineers as they are simple and easy to use. In words of Castillo (1988),"The basic idea of probability paper, of a given parametric family of distributions, is to modify the random variable, X, and probability, P, scales in such a way that the plot against X of any CDF, $F(x)$, belonging to that family, appears a straight line, such that no other CDF satisfies this property". However, this situation isn't possible in practice and is limited to just being theoretical. Usually a sample obtained from a random variable is known but not its CDF.

Thus, in probability papers, for a given family of distributions, the scales have been changed in such a manner that the cdfs represented graphically on paper appear as a straight line.

If $F(x;\theta)$ is a parametric family of CDFs with θ being the vector parameter, we have the following transformations:

$$\left.\begin{aligned}\xi &= g(x)\\ \eta &= h(y)\end{aligned}\right\} \tag{2.20}$$

The family of curves of equation

$$y = F(x;\theta) \tag{2.21}$$

become a straight line when transformed using Equation 2.20.

If $y = F(x;\theta)$ is written as

$$y = F(x;\theta) = h^{-1}(ag(x) + b) \leftrightarrow h(y) = ag(x) + b \tag{2.22}$$

Here, $g(x)$ and $h(xy)$ are functions and $h(y)$ is invertible.

It means, the transformation in Equation 2.20 changes Equation2.21 into a family of straight lines.

$$\eta = a\xi + b \tag{2.23}$$

where, η = reduced variate.

Normal probability paper

The cdf of normal population is given by,

$$F(x;\theta) = \Phi\left(\frac{x-\mu}{\sigma}\right) \tag{2.24}$$

Here, $\theta = (\mu, \sigma)$, μ = mean, σ = standard deviation, and $\Phi(x)$ = CDF of the standard normal $N(0,1)$ population.

Using Equations 2.22 and 2.24,

$$\left.\begin{aligned} \xi &= g(x) = x \\ \eta &= h(y) = \Phi^{-1}(y) \end{aligned}\right\} \tag{2.25}$$

and

$$\left.\begin{aligned} a &= \frac{1}{\sigma} \\ b &= \frac{-\mu}{\sigma} \end{aligned}\right\} \tag{2.26}$$

This results in a family of straight lines

$$\eta = a\xi + b = \frac{\xi + \mu}{\sigma} \tag{2.27}$$

Log-normal probability paper

When X is transformed to $log(X)$ we get log-normal probability paper. The only change here is that the abscissas axis becomes logarithmic.

Gumbel probability paper

The CDF curve for Gumbel is given by

$$y = F(x;\alpha,\beta) = exp\left[-exp\left(-\frac{x-\alpha}{\beta}\right)\right]; \quad -\infty < x < \infty \tag{2.28}$$

Taking logarithms twice of $1/y$ we get

$$-log\left[log\left(\frac{1}{y}\right)\right] = \frac{x-\alpha}{\beta} \tag{2.29}$$

Comparing Equation 2.29 with Equations 2.20 and 2.22 gives

$$\begin{aligned} \xi &= g(x) = x \\ \eta &= h(y) = -log\left[log\left(\frac{1}{y}\right)\right] = -log[-log(y)] \end{aligned} \tag{2.30}$$

and

$$\left.\begin{aligned} a &= 1/\beta \\ b &= -\alpha/\beta \end{aligned}\right\} \tag{2.31}$$

This shows that the Equation 2.30 changes Equation 2.31 into a family of straight lines

$$\eta = a\xi + b = \frac{\eta-\alpha}{\beta} \tag{2.32}$$

Similarly, we can derive probability papers for other distributions (Wilk and Gnanadesikan, 1968).

2.4 Annual Maxima Method

2.4.1 Stationary Gumbel Model

Here an Extreme Value distribution is fitted to the annual maximum wind speeds. For simplicity of this and the following section, we will only use the term wind speed. It is also assumed that the maxima from different years are serially independent. In comparison to daily or hourly measurements, the maxima taken over different years are more likely to be independent (Cohen, 1986). However, this also reduces the number of observations to be used for the statistical analysis. Therefore, the estimates obtained using Annual Maxima are much more variable in comparison to any other method (Perrin et al., 2006).

If m_j is the yearly maxima of the rounded wind speed recordings then F is the distribution of the yearly maximum wind speed. Therefore, the T-year wind u_T is given by

$$T(1 - F(u_T) = 1 \tag{2.33}$$

Here, m_j is assumed to have the Extreme Value distribution function

$$G(X) = exp\left\{-\left(1 + \gamma\frac{x-\mu}{\sigma}\right)_+^{-1/\gamma}\right\}$$

Here, μ is the location parameter, σ is the scale parameter, γ is the shape parameter, and $+$ represents the "positive" part. This means that those x-values that make the expression within the inner parenthesis negative is replaced by zero (Wolinski and Pytlowany, 2012). The density is given by

$$g(x) = \frac{1}{\sigma}\left(1 + \gamma\frac{x-\mu}{\sigma}\right)_+^{-(1/\gamma)-1} exp\left\{-\left(1 + \gamma\frac{x-\mu}{\sigma}\right)|_+^{-(1/\gamma)}\right\} \tag{2.34}$$

This becomes a Gumbel distribution function when $\gamma = 0$. That is

$$G_0(x) = exp(-exp(-\frac{x-\mu}{\sigma})) \tag{2.35}$$

The Gumbel distribution has been found to fit many types of data sets; therefore, it reserves a special place in EVT (Hong et al., 2014). This thesis makes use of Gumbel distribution while applying Annual Maxima method.

Gumbel probability paper is the standard starting point when the limit distribution of data is unknown because Gumbel type occupies a central position between Weibull and Frechet papers. Moreover, the plots of Weibull or Frechet probability paper requires the knowledge of the threshold parameter (Castillo, 1988). Gumbel model is one of the most commonly used probabilistic models when dealing with extreme wind speeds (Hong et al., 2013). It is also known as extreme value type I distribution. The Gumbel distribution has a CDF curve given by

$$y = F(x; \alpha, \beta) = e^{-e^{-\frac{x-\alpha}{\beta}}} \quad ; \quad \infty < x < \infty \tag{2.36}$$

where,
$F(x)$ = the cumulative distribution function,
x = value of the random variable X,
α = location parameter, and
β = scale parameter.

Equation 2.36 is used to plot the Gumbel probability paper for the annual maximum wind speed in km/h (KMPH). The process of plotting the PPPs is similar to that shown in Section 2.3. Castillo (1988) suggests that the Gumbel probablity paper of wind speed data will show a slight concave shape and does not show a linear trend in its range. The greater value of the wind speeds are accurately fitted by the Gumbel model (Kang et al., 2015). The right tail, however, will show a linear trend, which suggests a Gumbel type limit distribution.

Return periods help us estimate the risk based on historical data.

$$x_T = \alpha + \beta y_T \tag{2.37}$$

where, $y_T = -ln\left(-ln\left(1 - \frac{1}{T}\right)\right)$ and $\frac{1}{T}$ = the probability of getting a wind speed with a return period of y-years in one year.

2.4.2 Non-stationary Gumbel Model

The characteristics of non-stationary processes changes through time. A suitable model Z_t, the annual maximum wind speed in year t might be

$$Z_t = G(\alpha(t), \beta) \tag{2.38}$$

where for parameters a_0 and a_1,

$$\alpha(t) = a_0 + a_1 t \tag{2.39}$$

Thus, the location parameter for an appropriate extreme value model is modelled as a linear trend for variations through time in the observed process. Here a_1 is the parameter for the annual rate of change in annual maximum wind-speed.

Estimation by Maximum Likelihood Method

This method involves maximizing the likelihood of the observed sample which can be used to derive point or interval estimates. The scope of this section is limited to point estimation because this thesis does not use interval estimation.

Let us assume that $X = \{X_1, X_2...\}$ are independent and *iid* random variables having a common parametric family of pdfs $f(x;\theta)$ and cdfs $F(x;\theta)$. Here,

$$\theta = \{\theta_1, \theta_2, ..., \theta_k\} \in \Theta$$

This is a vector-valued parameter of dimension k in the parameter space Θ.

As the variables in X are independent, their joint probability distribution function is given by

$$f(x|\theta) = \prod_{i=1}^{n} f(x_i;\theta).$$

The values of $x = \{x_1, x_2, ..., x_n\}$ become known when the sample has been collected. Thus the above function can be viewed as a function of θ given x. This function, called the *likelihood function*, is written as

$$L(\theta|x) = \prod_{i=1}^{n} f(x_i;\theta).$$

Mathematically, it is often easier to deal with *loglikelihood function* rather than the likelihood function. The loglikelihood function is given by

$$\ell(\theta|x) = \log L(\theta|x) = \sum_{i=1}^{n} f(x_i;\theta).$$

The maximum likelihood of θ is obtained by maximizing either of the two functions above with respect to θ based on whether loglikelihood is required or likelihood is required.

Maximum likelihood approach has the adaptability to changes in model structure. That puts it at an advantage when it comes to parameter estimation.

Using a non-stationary Gumbel model to describe the distribution of Z_t for $t = 1, 2, ..., m$:

$$Z_t = GEV(\alpha(t), \beta(t)) \tag{2.40}$$

The likelihood of which is simply,

$$L(\eta) = \prod_{t=1}^{m} g(z_t; \alpha(t), \beta(t)) \tag{2.41}$$

where, $g(z_t; \alpha(t), \beta(t)) =$ the Gumbel density function with parameters $\alpha(t), \beta(t)$ evaluated at z_t.

Therefore, for the Gumbel case of $\xi = 0$, the MLE is given by

$$\ell(\alpha, \beta) = -mlog\beta - \sum_{t=1}^{m} \left(\frac{z_t - \alpha(t)}{\beta} \right) - \sum_{t=1}^{m} exp\left\{ - \left(\frac{z_t - \alpha(t)}{\beta} \right) \right\} \tag{2.42}$$

where $\alpha(t)$ is given by Equation 2.40.

The `imsev` pacakge of R developed using the algorithms written by Coles (2001) could be used easily in MATLAB using the `system()` command of MATLAB.

Percentile of the distribution

CDF for Gumbel distribution is given by,

$$F(x) = e^{-e^{-\left(\frac{x - \alpha}{\beta} \right)}} \tag{2.43}$$

So, multiplying both the sides by ln,

$$\begin{aligned} -\ln[-\ln(p)] &= \frac{x_p - \alpha}{\beta} \\ -\beta[\ln[-\ln(p)]] &= x_p - \alpha \end{aligned} \tag{2.44}$$

$$\therefore x_p = \alpha - \beta[\ln[-\ln(p)]] \tag{2.45}$$

Thus for non-stationary case from Section 3.2.2,

$$x_p = \alpha(t) - \beta[\ln[-\ln(p)]] \tag{2.46}$$

Considering the parameters a_0 and a_1 from equation 2.40,

$$x_p = a_0 - a_1(t) - \beta[\ln[-\ln(p)]] \tag{2.47}$$

2.5 Stochastic Process Model

This is one of the simplest statistical models for the analysis of extreme values. In this model, the number of observations are reduced to only peaks-over-threshold values Y_1, Y_2, ... , Y_k. Here, the peaks over threshold values exceed a particular threshold value y_0 (Ross, 1987).

When the peak magnitudes Y_i are independent and identically distributed, provided each variable Y_i - y_0 has an exponential distribution with the parameter λ, it results in the simplest analysis of this model. A very high threshold makes it hard to get enough observations, and a very low threshold will take into account non-extreme values. Therefore, for this model, the threshold value should be neither too high nor too low. The threshold value, here onwards, will be denoted by u.

The Stochastic Process model is the probabilistic model of the

1. the arrival process and
2. the peak wind speed.

Assume that $x_i,\ i = 1, 2, ..., N$ is the wind speed data, and u be the value of the threshold. That means,

$$y = x - u \tag{2.48}$$

We know that GPD is defined by

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi}, & \xi \neq 0. \\ 1 - exp\left(-\frac{x}{\beta}\right), & \xi = 0. \end{cases} \tag{2.49}$$

and are valid

$$\begin{aligned} \beta > 0 \; and \; x \geq 0 \; &\rightarrow \; \xi \geq 0 \\ 0 \leq x \leq -\frac{\beta}{\xi} \; &\rightarrow \; \xi < 0 \end{aligned} \tag{2.50}$$

where ξ is the shape parameter, and β is the scale parameter.

The distribution function of exceedances above threshold u for all $u < x_F$ is,

$$F_u(x) = P\{X - u \leq x \mid X > u\}, x \geq 0 \tag{2.51}$$

F_u, by the conditional probabilities, can be defined as

$$F_u(x) = \begin{cases} \dfrac{F(u+x) - F(u)}{1 - F(u)}, & x \geq 0. \\ 0, & \text{else.} \end{cases} \tag{2.52}$$

Putting the exceedances in relation to GPD,

$$F_u(y) \approx G_{\xi,\beta}(y) \tag{2.53}$$

That means,

$$F(x) = (1 - F(u)) \cdot \; G_{\xi,\beta}(x - u) + F(u) \quad for \quad x > u \tag{2.54}$$

where,

$$F(u) = \frac{N - m}{N} \tag{2.55}$$

Here, N is the number of observations and m is the number of exceedadnces.

Therefore, the estimation is given by -

$$\widehat{F(x)} = 1 - \frac{m}{N}\left(1 + \frac{\hat{\xi}(x-u)}{\hat{\beta}}\right)^{-1/\hat{\xi}} \tag{2.56}$$

where, $\hat{\xi}, \hat{\beta}$ = estimated parameters of GPD.

In general, this model assumes that once we have selected a threshold value, all the values or samples over the threshold are peaks (Rosso, 2015).

2.5.1 Poisson Process

This section briefly summarizes both the nonhomogeneous Poisson process (NHPP) and the homogeneous Poisson process which will be used while analysing the data using POT method.

A discrete distribution on the non-negative integers is called the Poisson distribution. A random variable X is said to have a Poisson distribution if it is a discrete random variable that has the probability mass function (pmf)

$$p(x) = P(X = x) = \frac{\phi^x \exp(-\phi)}{x!}; \quad x = 0, 1, 2, ...$$

If X is a discrete random variable having the above equation as pmf then we can write $X \sim POI(\phi)$.

The mean and the variance of the Poisson distribution are given by $E(X) = \phi$ and $V(X) = \phi$ respectively.

Moreover, for a Poisson process, the random variable $N(a, b]$ has a Poisson distribution with mean

$$\int_a^b \lambda(x)dx$$

This is a Poisson process with a constant intensity function. The HPP cannot be used to model systems that deteriorate or improve because the intensity function is constant. These situations call for the application of a Poisson process with non-constant intensity function.

HPP is very closely related to the exponential distribution. As stated by Rigdon (2000),"A process is an HPP with intensity λ, if and only if the times between failure are *iid* exponential random variables with mean $1/\lambda$".

The intensity density function of HPP is given by $\lambda(x) = \lambda$.

The intensity function, on the other hand, is a function $\lambda(x)$ that when integrated over a given time t gives the expected value of the number of events occurring in that time. This can be written, mathematically, as

$$\Lambda(t) = \int_t \lambda(x)dx.$$

Here, the expected number of events occurring in time t is given by $\Lambda(t)$.

HPP model fitting requires statistical estimation of only a single parameter called the failure rate (λ).

$$\lambda = n/t \tag{2.57}$$

If n is the number of shock events observed in a total observation period of t, then the failure rate of the the HPP model can be calculated using Equation 2.57.

The arrival of shock events is not constant if the system is deteriorating. For such systems, HPP is no longer applicable.

As stated by Rigdon (2000),"The nonhomogeneouos Poisson process (NHPP) is a Poisson process whose intensity function is non-constant". A point process is said to be NHPP with intensity function $\lambda(x)$ if $N(A)$, the number of events, occuring in A is a Poisson random variable. This means,

$$N(A) \sim Po(\Lambda(t))$$

For NHPP, the failure rate is a non-linear function of time. NHPP model is fitted to the failure data using **Power Law** model. The mean or the expected number of events for a Power Law model is given by

$$E[N(t)] = \lambda t^{\beta}$$

The process is called a **power law process**, when the intensity function has the form $\lambda(t) = (\beta/\theta)(t/\theta)^{\beta-1}$. Here, $\beta > 0$ and $\theta > 0$.

CROW-AMSAA method is used to fit the NHPP Power Law model graphically. In general, it can be used to estimate the parameters λ and β.

Here, parameter β is interpreted as follows:

- $\beta >1$, the number of events is increasing over time (NHPP)
- $\beta = 1$, a stable process (the mean number of events is constant over time; HPP)
- $\beta <1$, the number of events is decreasing over time (NHPP)

This is the simplest test to verify for NHPP or HPP. However, sometimes it is required to test for the HPP process in much detail when the value of β is very close to the value of one but not exactly one. That is when the tests outlined in the following section are used.

CROW-AMSAA Plots

Dr. Larry H, Crow developed this method at the U.S. Army Material Systems Analysis Activity (AMSAA), thereby it came to be known as CROW-AMSAA. CROW-AMSAA plots, according to the definition are simple power curves with cumulative failures on the $y-axis$ and the cumulative time on the $x-axis$. When data is plotted on a log paper, it results in a straight line (Dawson, 2011).

The equation $N(t) = \lambda * t^{\beta}$ is used to plot the simple regression line. The slope provides the β statistic, whereas the $y-axis$ intercept at time $t = 1$ provides λ. The statistic λ, here, is the occurrence of high wind events at time equal to 1. This is simply a hypothetical value to allow to be able to forecast future high wind events.

2.6 Distribution of Maximum Load

Load arrival process is an NHPP with the rate function $\lambda(t)$. The mean rate function is given by

$$\Lambda(t) = \int_0^t \lambda(x)dx \tag{2.58}$$

In the interval $(0, t)$, load arrives at random times $s_1, s_2, ..., s_{N(t)}$. Here, $N(t)$ is the number of events in $(0, t)$. The distribution of maximum magnitude of shock in $(0, t)$ is given by $X_{max}(t)$. If $[X_{max}(t) \leq u]$ then all the shocks in $(0, t)$ could be denoted by

$$[X_1 \leq u, X_2 \leq u, ..., X_{N(t)} \leq u] \tag{2.59}$$

This is illustrated in the Figure 2.1 where the random times are the years on the *y-axis*. The shocks $X_1, ..., X_{N(t)}$ are shown as stem plots, and the threshold u is shown as a solid line.

Therefore,

$$\begin{aligned} P[X_{max}(t) \leq u] &= P[X_1 \leq u, X_2 \leq u, ..., X_{N(t)} \leq u] \\ &= \sum_{n=0}^{\infty} P[X_1 \leq u, ..., X_n \leq u; N(t) = n] \end{aligned} \tag{2.60}$$

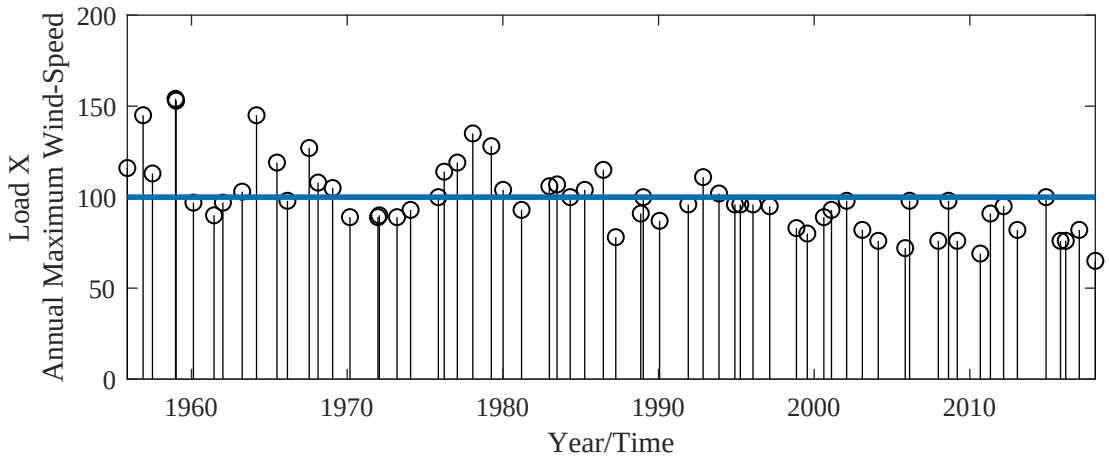


Figure 2.1: Annual Max Wind Speed vs. Years

Using the independence of X and $N(t)$,

$$\begin{aligned} P[X_{max}(t) \leq u] &= \sum_{n=0}^{\infty} P[X_1 \leq u, ..., X_n \leq u; N(t) = n] \\ &= \sum_{n=0}^{\infty} [F_x(u)]^n P[N(t) = n] \end{aligned} \tag{2.61}$$

For the NHPP model

$$P[N(t) = n] = \frac{[\Lambda(t)]^n e^{-\Lambda(t)}}{n!} \tag{2.62}$$

Substituting Equation 2.62 into 2.61,

$$\begin{aligned} P[X_{max}(t) \leq u] &= F_{max}(u, t) \\ &= exp\left(-\Lambda(t)(1 - F_X(u))\right) \end{aligned} \tag{2.63}$$

This means that $F_{max}(X, t)$ is the cdf of the maximum value distribution in the interval $(0, t)$.

2.6.1 Percentiles of Maximum Value

Suppose,

$$\begin{aligned} P[X_{max}(t) \leq x_p] &= p \\ or \quad F_{max}(x_p, t) &= p \end{aligned} \tag{2.64}$$

where p is the p^{th} percentile of $X_{max}(t)$.

Substituting Equation 2.63 into 2.64 gives

$$\begin{aligned} exp\left(-\Lambda(t)(1 - F_X(u))\right) &= p \\ 1 - F_x(x_p) &= \frac{-ln\ p}{\Lambda(t)} \end{aligned} \tag{2.65}$$

Therefore,

$$F_x(x_p) = 1 + \frac{ln\ p}{\Lambda(t)} \tag{2.66}$$

Depending on the distribution of shock height X or $F_x(x)$, substituting the cdf of $F(x)$ into the Equation 2.64 gives us the percentile for the given sample of data.

In the POT method, the threshold was selected to be u, so to model shock above the threshold we have to consider $x = y - u$, where y is the original data.

Our wind-speed data follows exponential distribution. Therefore,

$$F_x(x) = 1 - e^{-\lambda x} \tag{2.67}$$

Substituting Equation 2.67 into 2.66,

$$\begin{aligned} 1 - e^{-\lambda x_p} &= 1 + \frac{ln\ p}{\Lambda(t)} \\ -e^{-\lambda x_p} &= \frac{ln\ p}{\Lambda(t)} \end{aligned} \tag{2.68}$$

$$\therefore x_p = \left(\frac{-1}{\lambda}\right) ln\left[\frac{-ln\ p}{\Lambda(t)}\right] \tag{2.69}$$

Where for exponential distribution,

$$\Lambda(t) = \lambda(t)^{\beta}. \tag{2.70}$$

Now, we know that x_p is the shock above the threshold u. This means that the p^{th} percentile of real data is given by

$$y_p = x_p + u \tag{2.71}$$

Here, y_p gives the predicted wind-speed for the future years in km/h (KMPH).

We know from Equation 2.69 that $x_p = \left(-\frac{1}{\lambda}\right)\left(ln\left[-\frac{ln(p)}{\Lambda(t)}\right]\right)$.

Generalizing this equation to find the distribution of maximum wind speed in any general interval (t_1, t_2),

$$\begin{aligned} \Lambda(t_1, t_2) &= \Lambda(0, t_2) - \Lambda(0, t_1) \\ &= \lambda(t_2)^{\beta} - \lambda(t_1)^{\beta} \end{aligned} \tag{2.72}$$

where, t_1 and t_2 are in days, parameters λ and β are the values from Table 3.11.

Thus Equation 2.70 can be re-written using Equation 2.72

$$x_p = \left(-\frac{1}{\lambda}\right)\left(ln\left[-\frac{ln(p)}{\Lambda(t_1, t_2)}\right]\right) \tag{2.73}$$

Accounting for the threshold u again, the solution obtained from using Equation 2.73 can be used to calculate y_p using Equation 2.71.

Special Case of HPP

NHPP also has a special case HPP with parameter λ.

We know that rate,

$$\lambda(t) = \frac{d}{dt}\Lambda(t) = \frac{d}{dt}(\alpha t^{\beta}) = \alpha\beta t^{\beta-1} \tag{2.74}$$

Therefore, in HPP

$$\Lambda(t) = \alpha t, \quad \text{because} \quad \beta = 1 \tag{2.75}$$

In the previous pages of this thesis the parameter α has been denoted by λ, so using the same notation for consistency, the above equation could be written as

$$\Lambda(t) = \lambda t \tag{2.76}$$

where,

$$\begin{aligned} \lambda &= \frac{\text{number of failures}}{\text{observation period}} \\ &= \frac{\text{total number of wind events above threshold } u}{\text{total observation period in days}} \end{aligned} \tag{2.77}$$

Based on Equation 2.76, for the case of any general interval of (t_1, t_2),

$$\begin{aligned} \Lambda(t_1, t_2) &= \Lambda(0, t_2) - \Lambda(0, t_1) \\ &= \lambda(t_2) - \lambda(t_1) \end{aligned} \tag{2.78}$$

Therefore, after substituting Equation 2.78 into 2.73, we can calculate the percentile for the HPP case using Equation 2.71

2.7 Statistical Test for HPP

The simplest statistical model to describe a failure in a system is using a Homogeneous Poisson Process (HPP). HPP implies the following two things:

- The system does not deteriorate or exhibit reliability improvement in global time
- The system does not wear out in local time.

HPP would not be an appropriate model for the system if even one of the above two is not satisfied. The Graphical method of Section 3.4.1 can be used to identify a HPP. One of the earliest tests called the Laplace test could also be used for this purpose.

For our case of collecting Wind Speed data, the collection of data was terminated at December 2018. Thus, the data is said to be **time truncated** because the testing stopped at a predetermined time t.

Laplace Test

For the time truncated case, the test statistic for the Laplace test is

$$L = \frac{\sum_{i=1}^{n} T_i/n - t/2}{t/\sqrt{12n}} \tag{2.79}$$

Here, t is the predetermined time, and the random variables $T_1 < T_2 < \cdots < T_n$ are distributed as n order statistics from a distribution that is uniform on the interval $(0,t)$.

The null hypothesis

$$H_0\text{: The process is an HPP}$$

is rejected if

$$L < -z_{\alpha/2} \text{ or } L > z_{\alpha/2}$$

A large value of L indicates that the system is deteriorating, and a small value indicates that the system is improving.

2.7.1 Test of Hypothesis for β

For a time truncated case, the number of failures N is random whereas the time the testing stops t is fixed. The limitations of the Graphical tests can be overcome using these procedures. Here the MLE is equal to

$$\hat{\beta} = \frac{N}{\sum_{i=1}^{N} log(t/t_i)} \tag{2.80}$$

and the conditionally unbiased estimator for β is

$$\bar{\beta} = \frac{N-1}{N}\hat{\beta} \tag{2.81}$$

We begin the procedure of statistical inference for our time truncated case with the testing of the hypothesis for β.

The quantity $2n\beta/\hat{\beta}$ has a chi-square distribution with $2n$ degrees of freedom for a time truncated case at time t and N $= n$.

Here the null hypothesis, H_0: HPP process, is rejected if

$$2n\beta/\hat{\beta} < \chi^2_{1-\alpha/2} \quad or \quad 2n\beta/\hat{\beta} > \chi^2_{\alpha/2} \tag{2.82}$$

or

$$\hat{\beta} < \frac{2n\beta_0}{\chi^2_{\alpha/2}} \quad or \quad \hat{\beta} > \frac{2n\beta_0}{\chi^2_{1-\alpha/2}} \tag{2.83}$$

2.7.2 Cramer-von Mises Test

Cramer-von Mises (CVM) test is used here to formally test the adequacy of the power law process. Here, the hypothesis is as follows

H_0: Power Law Process is an adequate model
H_1: Power Law Process is not an adequate model

Small samples do not provide enough information, so they might not lead to the rejection of the null hypothesis, or they might. Sometimes, even large samples are prone to lead to the rejection of null hypothesis even if the data follows the power law process closely. Because of this reason, we use charts to determine whether our time truncated data follows a power law model. The reason behind using the CVM plots for the hypothesis is because of the limitation of the CVM test critical value table relative to the size of our samples. The tables that exist have the maximum sample size of 100 which is not enough for the current weather data (Stephens, 1970).

The Cramer-Von Mises test statistic is given by

$$C^2_R = \frac{1}{12n} + \sum_{i=1}^{n} \left(\hat{R}_i - \frac{2i-1}{2n} \right)^2 \tag{2.84}$$

where

$$\hat{R}_i = \left(\frac{t_i}{t} \right)^{\bar{\beta}} \tag{2.85}$$

Here, we plot $\hat{R}_i$ on the horizontal axis, and its expectation $E(\hat{R}_i)$ on the vertical axis.

$$E\left(\hat{R}_i\right) = \frac{2i-1}{2n} \tag{2.86}$$

This plot is approximately linear along 45° line going through the origin when the data follows a power law process. Such a plot with points lying close to the diagonal line going through the origin indicates a good fit.

Efficiency of the estimator

For a HPP, the estimate of intensity, for time truncated case, is given by

$$\bar{u} = \frac{n}{t_n} \tag{2.87}$$

whereas, for NHPP the estimate of intensity is given by

$$\hat{u}\,(t) = \frac{N\hat{\beta}}{t}. \tag{2.88}$$

Thus, the **efficiency** is given by

$$\text{Efficiency} = \frac{\text{MSE of } \hat{u}}{\text{MSE of } \bar{u}}. \tag{2.89}$$

Here $\bar{u}$ is a better estimator when efficiency is greater than 1. Contrarily, the efficiency of less than 1 indicates that $\hat{u}$ is a better estimator. Based on this, it is advisable to assume a power law process for a larger sample size and a HPP for a small sample size.

2.8 Conclusion

The global consensus among scientists about the change in global climate has motivated the non-stationary modelling of extreme wind speed. Anything that is built will need to survive through this non-stationary climate. That is the strength of the buildings would need to be formulated in accordance with the forecasted wind speed. The two models of Annual Maxima and Stochastic Process model would need to be analysed from the non-stationary point of view as these are the most commonly used models for the extreme wind speed data. This Chapter provided a background knowledge for these two methods. For the non-stationary Gumbel model, the parameter a_1 provides the annual rate of change in the annual maximum wind speed, whereas for the non-stationary Stochastic Process model, the parameter λ provides the daily rate of change in the daily maximum wind speed.

Chapter 3

Application - Wind Data

3.1 Introduction

This chapter discusses the application of two estimation methods of the Gumbel model and the Stochastic process model(Holmes and Moriarty, 1999; Lechner et al., 1992; Simiu and Heckert, 1996) of extreme value analysis and applies it to the wind speed data that Environment Canada collects. This wind speed data is collected in the intervals of hours, days, months, and years. This data is available to download using Cygwin64 Terminal. Table 3.1 lists the stations used for the analysis of wind data. Most stations have multiple station IDs to accommodate the introduction of new equipment and retiring of the old one. This is evident from the example of the station of the Region of Waterloo. Waterloo Wellington A (Station ID 4832) operated from 1970 to 2003, when it comes to collecting the daily weather data, and was located at the latitude and longitude of 43.45 and -80.38 degrees respectively. Region of Waterloo Int'l Airport (Station ID 32008) located at the latitude and longitude of 43.46 and -80.38 degrees respectively, operated from 2002 to 2011. Based on this, the location can be deemed to be the same with just the change in the Station ID and name.

The latitude and longitude data, in degrees, was retrieved from Environment Canada's FTP url:

ftp://client_climate@ftp.tor.ec.gc.ca/Pub/Get_More_Data_Plus_de_donnees/Station%20Inventory%20EN.csv

Table 3.1: List of Weather Stations

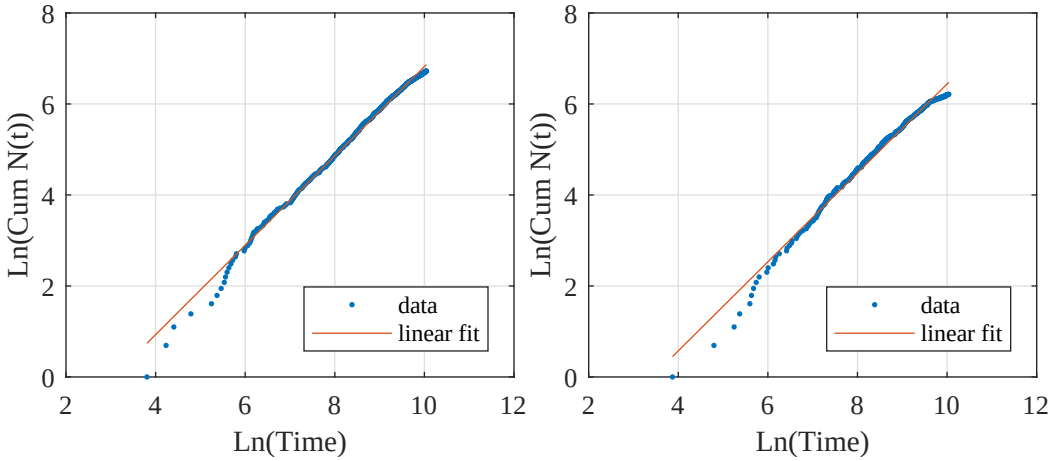
Station	Equivalent Name	Station ID	First Year	Last Year
London	London Int'l Airport	4789	1940	2012
	London A	50093	2012	2018
Toronto Pearson	Toronto Lester B. Pearson Int'l A	5097	1937	2013
	Toronto Intl A	51459	2013	2018
Region of Waterloo	Waterloo Wellington A	4832	1970	2003
	Region of Waterloo Int'l Airport	32008	2002	2011
	Kitchener/Waterloo	48569	2010	2018
Trenton	Trenton A	5126	1955	2018
Wiarton	Wiarton A	4538	1947	2014
	Wiarton A	53139	2014	2018
Sarnia	Sarnia Airport	4589	1947	2014
	Sarnia Climate	44323	2006	2018
Toronto Islands	Toronto Island A	5085	1957	2006
	Toronto City Centre	30547	2006	2014
	Toronto City Centre	48549	2010	2018
Hamilton	Hamilton A	4932	1959	2011
	Hamilton A	49908	2011	2018
Kapuskasing	Kapuskasing A	4157	1961	2014
	Kapuskasing A	52604	2014	2018
Sudbury	Sudbury A	4132	1957	2013
	Sudbury A	50840	2013	2018

3.2 Annual Maxima Method

3.2.1 Stationary Model

As we have already discussed, Gumbel probability paper is the standard starting point when the limit distribution of data is unknown

Equation 2.36 is used to plot the Gumbel probability paper for the annual maximum wind speed in km/h (KMPH). The process of plotting the PPPs is similar to that shown in the previous sections. Figure 3.1 shows an example PPP for the Trenton wind speed data. The plots for the wind speed data from the remaining nine weather stations is available for reference in Appendix A.4. The main goal behind using these PPPs is to calculate the stationary Gumbel parameters.



(a) Trenton: 60 KMPH (b) Trenton: 70 KMPH

Figure 3.1: Stationary Gumbel PPP

3.2.2 Non-stationary Model

Based on the details discussed in Section 3.2.2, the `system()` command of MATLAB is used to run the `imsev` pacakge of R to calculate the non-stationary parameters listed in Table 3.3 (Martinez, 2008). These parameters would be later on used to make the projections.

Table 3.2: Stationary Gumbel Parameters using PPP

Station Name	Parameters
Trenton	$\alpha = 90.288$, $\beta = 16.142$
Region of Waterloo	$\alpha = 88.312$, $\beta = 9.779$
London	$\alpha = 96.256$, $\beta = 13.025$
Hamilton	$\alpha = 96.302$, $\beta = 9.904$
Sarnia	$\alpha = 87.885$, $\beta = 13.289$
Wiarton	$\alpha = 90.821$, $\beta = 9.647$
Toronto Island	$\alpha = 91.765$, $\beta = 8.539$
Toronto Pearson	$\alpha = 93.234$, $\beta = 7.428$
Kapuskasing	$\alpha = 79.824$, $\beta = 9.384$
Sudbury	$\alpha = 85.530$, $\beta = 14.361$

Table 3.3: Non-Stationary Gumbel Parameters using MLE

Station Name	Parameters
Trenton	$a_0 = 108.307$, $a_1 = -0.472$, $\beta = 11.480$
Region of Waterloo	$a_0 = 99.985$, $a_1 = -0.707$, $\beta = 18.839$
London	$a_0 = 117.956$, $a_1 = -2.186$, $\beta = 19.947$
Hamilton	$a_0 = 106.323$, $a_1 = -0.380$, $\beta = 9.543$
Sarnia	$a_0 = 94.534$, $a_1 = -0.240$, $\beta = 12.105$
Wiarton	$a_0 = 96.937$, $a_1 = -0.225$, $\beta = 9.944$
Toronto Island	$a_0 = 95.357$, $a_1 = -0.055$, $\beta = 8.432$
Toronto Pearson	$a_0 = 98.105$, $a_1 = -0.207$, $\beta = 7.775$
Kapuskasing	$a_0 = 87.105$, $a_1 = -0.260$, $\beta = 10.991$
Sudbury	$a_0 = 104.934$, $a_1 = -0.570$, $\beta = 11.554$

3.3 Stochastic Process Model

POT allows for the dataset to be enlarged because it considers the largest order statistics exceeding sufficiently high threshold from the collected data. This helps with reducing the sampling uncertainty. Simiu and Heckert (1996) has already applied this technique to estimate the design of wind speed in the US.

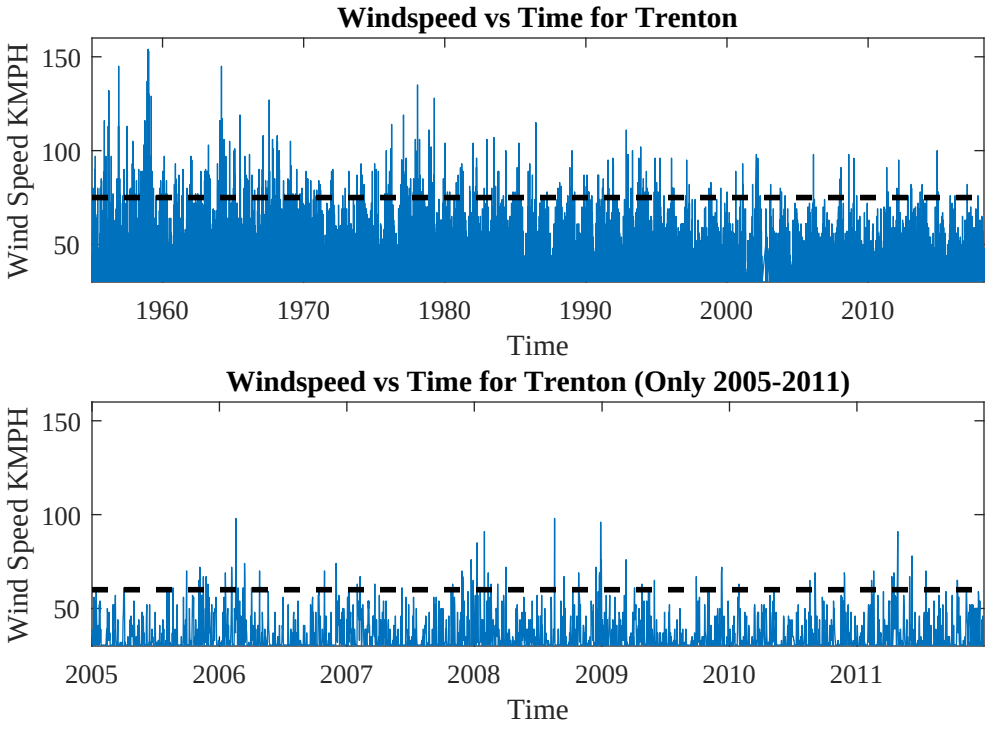


Figure 3.2: Wind Speed vs. Time Plots

3.3.1 Distribution of Event Magnitude

Figure 3.2 shows two plots of Wind Speed against Time. Here, the average rough peaks are considered as storm events. These averagely rough peaks could be viewed as the threshold of our data. The selection of threshold should be such that only true peaks are selected.The distribution of the selected extremes will not converge to the generalized pareto distribution asymptote if true peaks are not selected. However, the threshold should also not be so

Table 3.4: Maximum Wind Speed in descending order for Trenton Weather Station

Date (yyyy-mm-dd)	**Total Wind Gust** (KMPH)
1958-12-22	154
1959-01-05	153
1956-11-21	145
1964-03-05	145
1978-01-26	135
.	.
.	.
.	.
2015-10-29	76
2016-02-29	76
2005-11-06	72
2010-09-03	69
2018-01-23	65

high that there is not enough data for the satisfactory analysis and the determination of distribution of parameters (Lee et al., 2011).

In the figure on the top, the time extends from 1955 to 2018. Here, because the data is so condensed the threshold comes out to be around 75 KMPH. However, in the plot below, with the time-line of years extending from 2005 to 2011, the threshold is visible to be about 60 KMPH. These two thresholds have been marked using a horizontal line in the Figure 3.2. Visually finding a threshold, thus, is not effective because we are getting two different values. For this reason, we find threshold using the method outlined in this section.

The first step in finding the threshold involves finding the annual maximum data points. To help with this, first, the data is arranged in a descending order as shown in Table 3.4. It lists the Total Wind Speed in kilometres per hour (KMPH) for the Trenton Station only, so 9 other similar tables are put together for the remaining 9 stations on our list. The data from these 9 stations has not been listed for convenience as only the minimax and the maximax values from these tables help with narrowing down a threshold value. For the station of Trenton, from Table 3.4 the minimax is 65 KMPH, whereas the maximax is 154 KMPH.

The minimax and maximax data for the 10 stations is listed in Table 3.5. The thresh-

Table 3.5: Minimax and Maximax Wind Speed in KMPH for each of the 10 stations

Minimax Date (yyyy-mm-dd)	**Minimax Wind Gust** (KMPH)	**Station Name**	**Maximax Date** (yyyy-mm-dd)	**Maximax Wind Gust** (Km/h)
2018-01-23	69	Toronto Airport	1978-01-26	115
2018-01-23	65	Trenton	1958-12-22	154
2018-01-17	67	Toronto Island	1978-01-26	126
2001-04-12	80	London	1992-06-17	148
2018-01-26	72	Wiarton	1984-04-30	126
2002-10-19	63	KW	1981-03-30	120
2011-12-15	78	Hamilton	1978-01-26	133
2006-07-30	65	Sarnia	1991-03-27	159
1996-01-27	59	Kapuskasing	1975-10-25	106
2004-03-20	63	Sudbury	1964-04-14	137

old is selected using this table. We are going to use two threshold values: 60KMPH and 70KMPH. Here, only the minimax values of the wind speed were used to decide the threshold for storm events. Anything below these two values was considered to be a normal meteorological occurrence and not a storm event.

A MATLAB function is written to get the data ready to plot the probability paper plots. The first step involves removing all the data below the threshold of 60KMPH. A universal MATLAB function is written for this so that it can be reused for all the 10 stations. This function also forms the basis of the MATLAB code to plot the probability paper plots (Sa, 2003).

Apart from removing all the data below 60 KMPH, this MATLAB function also calculates the time interval or inter-arrival time between the events and removes any inter-arrival time that is less than 7 days. This is because a storm event has been assumed to last for about 7 days. Deleting below 7 days ensures that we are looking at different storm events. This function is then used for all the remaining 9 stations to form 10 tables, in total, similar to Table 3.6.

60KMPH Threshold

The MATLAB code for this section is based on the one developed for the Section 3.3.1. Here, we delete the data below 61KMPH. The reason behind deleting the data below

Table 3.6: Example of extracted data above 60 KMPH for Trenton

Date (yyyy-mm-dd)	**Total Wind Gust** (KMPH)	**Inter-arrival time** (days)
1955-03-13	66	45
1955-04-06	66	9
1955-04-19	64	13
1955-05-27	64	38
1955-08-05	80	69
1955-08-30	85	25
.	.	.
.	.	.
.	.	.
2017-10-15	76	54
2017-10-30	63	15
2017-11-09	63	10
2017-12-05	63	26
2018-01-23	65	48

61KMPH is because upon deleting the data shifts; that means, we have to offset this shift in data to obtain the correct probability paper plots. This is achieved by subtracting 60KMPH from the extracted data. The Probability Paper Plots, comparing the four distributions of Exponential, Log Normal, Weibull, and Gumbel, for the Wind Gust data above 60KMPH are shown in Figure 3.3. From this figure, it can be seen that the data set follows Exponential Distribution.

The distribution parameters and R-Squared values of the four distributions are summarized in the Table 3.7. The R-Squared values in this table have been calculated using the MATLAB function `fitlm`. This table also helps us compare how close is the mean and the variance for the Trenton Weather Station relative to the four distributions: Lognormal, Exponential, Weibull, and Gumbel.

Upon repeating the process of distribution selection for the remaining nine stations, we find out that the distribution of data stays the same as the one selected for Trenton Station; that is, the Wind Gust data follows an Exponential distribution and the Wind Inter-Arrival data follows a Log Normal distribution. Table 3.8 lists the distribution parameters for the 9 weather stations not covered in the previous sections.

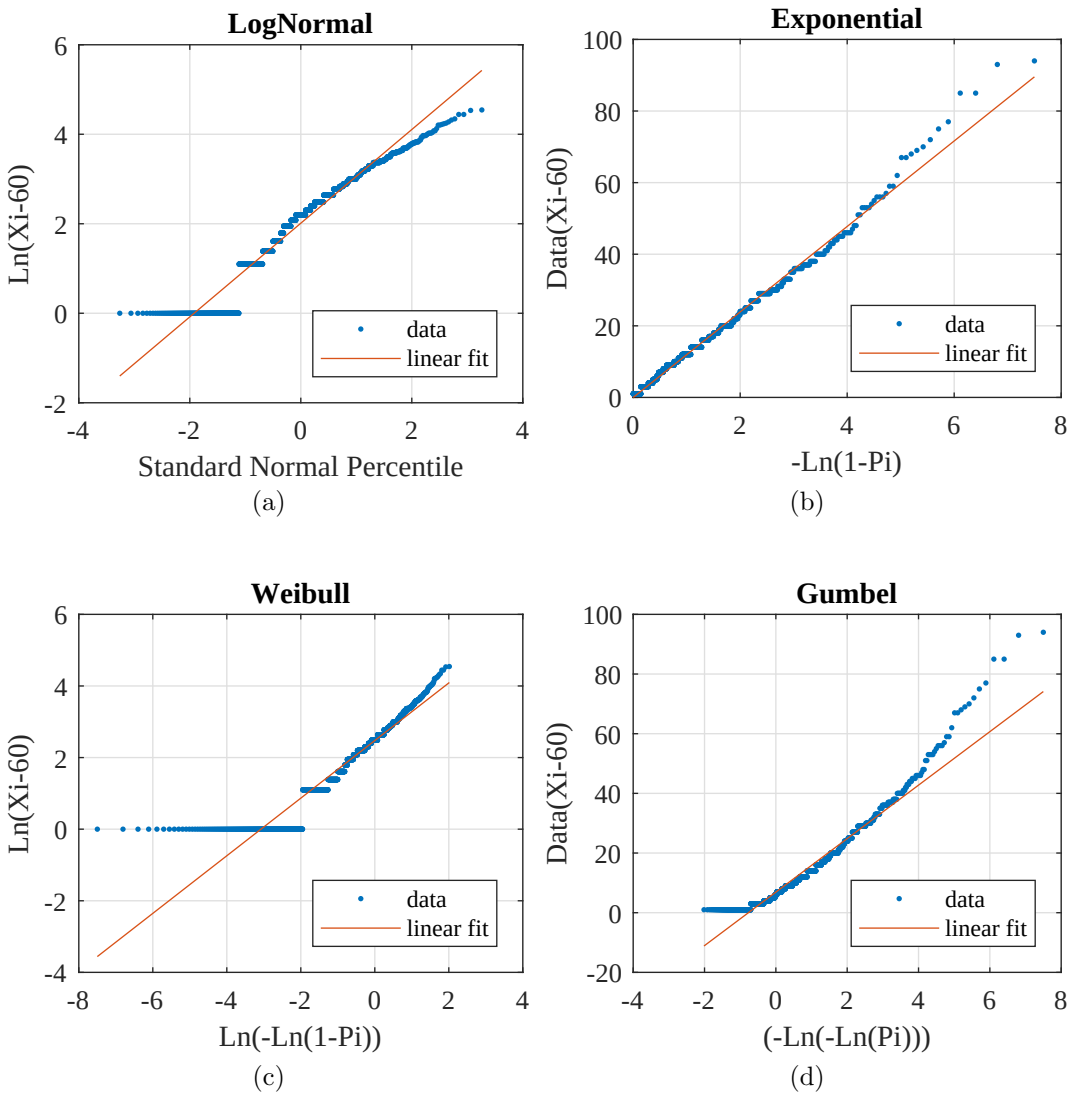


Figure 3.3: Probability paper plots with Wind Speed threshold 60KMPH: Trenton

Table 3.7: Distribution Parameters and Statistics with Wind Speed threshold 60KMPH: Trenton

Distribution	**Parameters**	**R-Squared**	**Mean**	**Variance**	**COV**
Log Normal	$\zeta = 1.04$, $\lambda = 2.01$	0.948	12.92	331.78	1.40
Exponential	$\lambda = 0.083$	0.993	11.93	142.34	1.00
Weibull	$\alpha = 1.24$ $\beta = 11.87$	0.915	11.07	80.26	0.80
Gumbel	$\alpha = 6.89$ $\beta = 8.95$	0.949	12.06	131.97	0.95

Table 3.8: PPP Parameters for Ontario Weather Stations with a threshold of 60KMPH

Station	λ
Waterloo	0.102
London	0.083
Hamilton	0.092
Sarnia	0.093
Wiarton	0.095
Toronto Island	0.097
Toronto Pearson	0.096
Kapuskasing	0.110
Sudbury	0.091

70KMPH Threshold

Here, we delete the data below 71KMPH. Following this, the steps to execute this part are the same as in Section 3.3.1.

The same procedure as the previous section is followed to plot the the probability paper plots of the Figure 3.4. As can be observed from the figure, the data set follows Exponential Distribution. Table 3.9 summarizes the distribution parameters and R-Squared values of the distributions in the Figure 3.4.

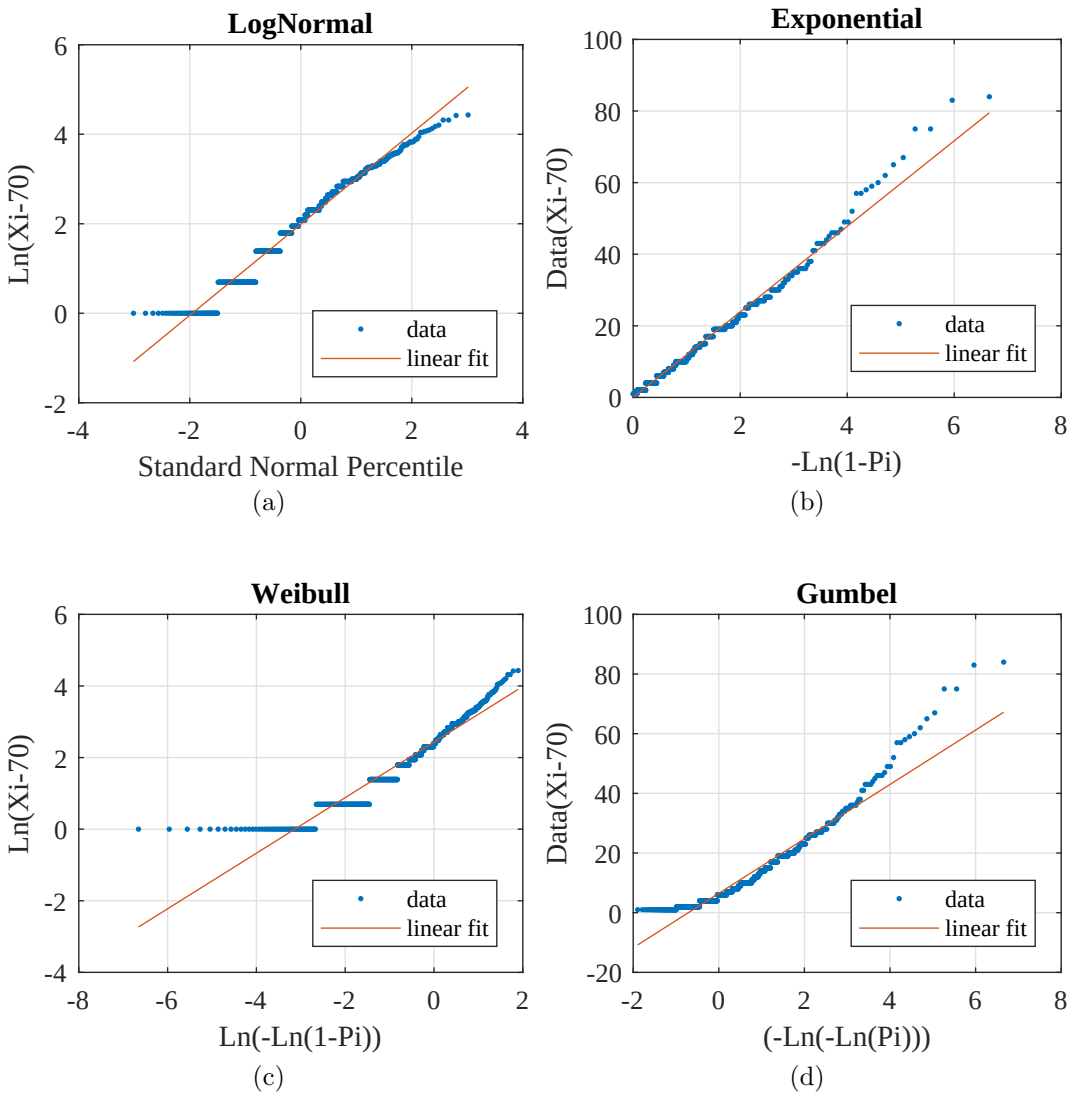


Figure 3.4: Probability paper plots with Wind Speed threshold 70KMPH: Trenton

Table 3.9: Distribution Parameters and Statistics with Wind Speed threshold 70KMPH: Trenton

Distribution	**Parameters**	**R-Squared**	**Mean**	**Variance**	**COV**
Log Normal	$\zeta = 1.01$, $\lambda = 1.98$	0.973	12.23	271.89	1.34
Exponential	$\lambda = 0.083$	0.989	11.93	142.46	1.00
Weibull	$\alpha = 1.28$ $\beta = 11.38$	0.920	10.53	67.89	0.78
Gumbel	$\alpha = 6.47$ $\beta = 9.12$	0.930	11.73	136.89	0.99

Similar to the previous section of 60KMPH, here, the wind speed data follows an Exponential distribution and the Wind Inter-Arrival data follows a Log Normal distribution. The only exception to this scenario is the Toronto Station Wind Inter-Arrival data above 70KMPH which follows an Exponential Distribution if we use the R^2 value to select the distribution; otherwise, it follows the Log Normal Distribution. Table 3.10 lists the distribution parameters for the 9 weather stations not covered in the previous sections.

Table 3.10: PPP Parameters for Ontario Weather Stations with a threshold of 70KMPH

Station	λ
Region of Waterloo	0.105
London	0.081
Hamilton	0.095
Sarnia	0.092
Wiarton	0.098
Toronto Island	0.102
Toronto Pearson	0.101
Kapuskasing	0.109
Sudbury	0.091

When it comes to the Inter-Arrival PPP for the Toronto Station with the wind data above 70KMPH, unlike the previous cases, the PPP turns out to be Exponential if we consider the R-Squared value to be of significance. Exponential Distribution Parameter for this case is $\lambda = 0.024677$ with $R^2 = 0.991373$.

3.4 NHPP Arrival Process of High Wind Events

Graphical methods are often used to select a reasonable model. These graphical methods find out whether the time between failures is getting longer or shorter. In terms of wind speed, we find out whether the time between storm like events is getting longer or shorter. Based on this, one from the two commonly applied models is selected: Homogeneous Poisson Process (HPP) or the Power Law Process.

3.4.1 Graphical Methods

The simplest graphical method involves plotting the cumulative wind inter-arrival time t_i against the cumulative number abnormal of wind speed events N(t_i). If the plot is linear then the system is said to be stable over the time data was collected. This means that a HPP or a renewal process is an appropriate model for this scenario. A renewal process is not a suitable model when the plot has a concave up or concave down curvature. Systems with such a curvature are modeled by a Nonhomogenous Poisson Process (NHPP) or a nonstationary process. NHPP is also referred to as a **Power Law Process**. It is appropriate for systems that are improving or deteriorating. Figure A.1 shows the plots using this simple graphical method. All the figures show either a slight concave up or concave down curvature.

3.4.2 CROW-AMSAA Plots

As discussed in the previous sections and by Dawson (2011) The NHPP Power Law model states that

$$E[N(t)] = \lambda t^{\beta} \tag{3.1}$$

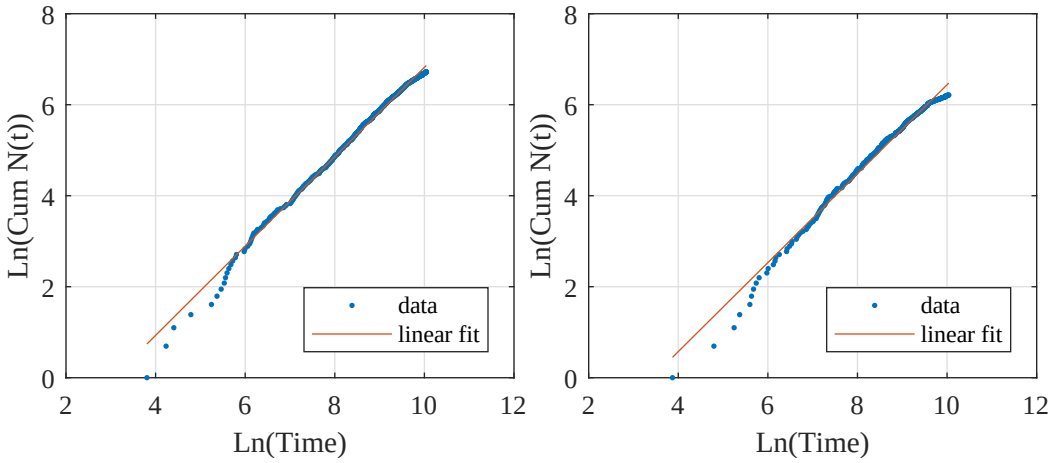
Taking the natural log in Equation 3.1, we get

$$\ln(E[N(t)]) = \ln(\lambda t^{\beta}) \tag{3.2}$$

Equation 3.2 can be simplified to

$$\ln(E[N(t)]) = \ln(\lambda) + \beta[ln(t)] \tag{3.3}$$

Thus, we can now use the Equation 3.3 to find the NHPP parameters by plotting *ln(E[N(t)])* vs. *ln(t)*. In other words, we are using CROW-AMSAA method for fitting NHPP Power Low model graphically.



(a) Trenton: 60 KMPH (b) Trenton: 70 KMPH

Figure 3.5: 60KMPH and 70KMPH NHPP Plots for Trenton

For example, when we plot the NHPP for the Trenton Station wind speed data above 60KMPH we get the following linear equation -

$$\ln(E[N(t)]) = -2.9967 + 0.9817[ln(t)] \tag{3.4}$$

Comparing equation 3.4 with 3.3, we see that

$$\lambda = e^{-2.9967} = 0.0499; \beta = 0.9817 \tag{3.5}$$

Similarly, we can find the parameters for the remaining sets of data. Table 3.11 summarizes the NHPP parameters for the 10 Ontario Stations.

We know that, parameter β is interpreted as follows:

- β >1, the number of events is increasing over time (NHPP)
- $\beta = 1$, a stable process (the mean number of events is constant over time; HPP)
- β <1, the number of events is decreasing over time (NHPP)

Figure 3.5 shows the Wind Inter-Arrival fitted with NHPP Power Law model for the wind speed data from Trenton. Figures in the Appendix A.2 show the Wind Inter-Arrival fitted with NHPP Power Law model for the remaining nine weather stations. As can be observed from the table, the values of β is either greater or less than 1. Thus, we can categorize all the weather stations under a NHPP.

Table 3.11: NHPP Parameters for the ten Ontario weather stations for the Wind Inter-Arrival data above, both, 60KMPH and 70KMPH.

Station Name	λ	β	R^2	**Interpretation**
Trenton (60KMPH)	0.0499	0.9817	0.990	NHPP
Trenton (70KMPH)	0.0355	0.9775	0.984	NHPP
Region of Waterloo (60KMPH)	0.0708	0.9181	0.993	NHPP
Region of Waterloo (70KMPH)	0.0260	0.9708	0.988	NHPP
London (60KMPH)	0.0564	0.9485	0.996	NHPP
London (70KMPH)	0.0756	0.8509	0.982	NHPP
Hamilton (60KMPH)	0.0854	0.9251	0.999	NHPP
Hamilton (70KMPH)	0.1285	0.8453	0.996	NHPP
Sarnia (60KMPH)	0.0414	0.9868	0.989	NHPP
Sarnia (70KMPH)	0.0048	1.1751	0.970	NHPP
Wiarton (60KMPH)	0.1148	0.8786	0.997	NHPP
Wiarton (70KMPH)	0.0866	0.8580	0.996	NHPP
Toronto Island (60KMPH)	0.0500	0.9752	0.996	NHPP
Toronto Island (70KMPH)	0.0191	1.0399	0.993	NHPP
Toronto Pearson (60KMPH)	0.0525	0.9730	0.995	NHPP
Toronto Pearson (70KMPH)	0.0348	0.9712	0.997	NHPP
Kapuskasing (60KMPH)	0.0053	1.1677	0.979	NHPP
Kapuskasing (70KMPH)	0.0042	1.1082	0.973	NHPP
Sudbury (60KMPH)	0.1053	0.8992	0.990	NHPP
Sudbury (70KMPH)	0.1500	0.8283	0.969	NHPP

3.4.3 Statistical Test for HPP

From the previous section, it can be seen that the values of β are very close to one, but not exactly one. Therefore, we perform the following tests to verify whether the data could be modelled as HPP.

Laplace Test

We know the null hypothesis

$$H_0\text{: The process is an HPP}$$

is rejected if

$$L < -z_{\alpha/2} \text{ or } L > z_{\alpha/2}$$

A large value of L indicates that the system is deteriorating, and a small value indicates that the system is improving.

Table 3.12 shows the results of the Laplace test for the 10 weather stations. It can be seen that for all the stations the P-value is less than 0.025. This indicates a rejection of null hypothesis.

Test of Hypothesis for β

As discussed in Chapter 2, we begin the procedure of statistical inference for our time truncated case with the testing of the hypothesis for β.

Here the null hypothesis, H_0: HPP process, is rejected if

$$2n\beta/\hat{\beta} < \chi^2_{1-\alpha/2} \quad or \quad 2n\beta/\hat{\beta} > \chi^2_{\alpha/2} \tag{3.6}$$

or

$$\hat{\beta} < \frac{2n\beta_0}{\chi^2_{\alpha/2}} \quad or \quad \hat{\beta} > \frac{2n\beta_0}{\chi^2_{1-\alpha/2}} \tag{3.7}$$

The results of size α test of the Hypothesis $H_0 : \beta = 1$ versus $H_0 : \beta \neq 1$ are shown in the Table 3.13. According to the results, the process is governed by a power law process except for the stations of the Region of Waterloo and London. Samples for both the stations are smaller in size in comparison to the other stations. Sometimes small samples do not provide enough information. This could be the reason behind there not being enough evidence to reject the null hypothesis for the two cases of the Region of Waterloo and London.

Table 3.12: Laplace Test Results

Station Name	L	P-value	Hypothesis
Trenton	-12.55	0.0001	Reject H_0
Region of Waterloo	-2.25	0.0241	Reject H_0
London	-2.54	0.0109	Reject H_0
Hamilton	-7.57	0.0001	Reject H_0
Sarnia	-3.64	0.0003	Reject H_0
Wiarton	-5.65	0.0001	Reject H_0
Toronto Island	-5.99	0.0001	Reject H_0
Toronto Pearson	-4.71	0.0001	Reject H_0
Kapuskasing	-2.70	0.0069	Reject H_0
Sudbury	-13.95	0.0001	Reject H_0

Table 3.13: Results of Hypothesis test for β

Station Name	Reject H_0 if	$\hat{\beta}$	Hypothesis
Trenton	$\hat{\beta}$<1.525 or $\hat{\beta}$ >1.765	0.7581	Reject H_0
Region of Waterloo	$\hat{\beta}$<0.965 or $\hat{\beta}$ >1.150	1.0032	Not enough evidence to reject H_0
London	$\hat{\beta}$<0.8915 or $\hat{\beta}$ >1.091	0.9123	Not enough evidence to reject H_0
Hamilton	$\hat{\beta}$<1.283 or $\hat{\beta}$ >1.529	0.8157	Reject H_0
Sarnia	$\hat{\beta}$<1.097 or $\hat{\beta}$ >1.308	0.9259	Reject H_0
Wiarton	$\hat{\beta}$<1.116 or $\hat{\beta}$ >1.330	0.8170	Reject H_0
Toronto Island	$\hat{\beta}$<1.207 or $\hat{\beta}$ >1.439	0.8568	Reject H_0
Toronto Pearson	$\hat{\beta}$<1.347 or $\hat{\beta}$ >1.605	0.8978	Reject H_0
Kapuskasing	$\hat{\beta}$<0.928 or $\hat{\beta}$ >1.106	1.0017	Not enough evidence to reject H_0
Sudbury	$\hat{\beta}$<1.391 or $\hat{\beta}$ >1.658	0.6700	Reject H_0

Cramer-von Mises test

As discussed in the previous Chapters, the Cramer-Von Mises test statistic is given by

$$C_R^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(\hat{R}_i - \frac{2i-1}{2n} \right)^2 \tag{3.8}$$

where

$$\hat{R}_i = \left(\frac{t_i}{t} \right)^{\bar{\beta}} \tag{3.9}$$

Here, we plot $\hat{R}_i$ on the horizontal axis, and its expectation $E(\hat{R}_i)$ on the vertical axis.

$$E\left(\hat{R}_i\right) = \frac{2i-1}{2n} \tag{3.10}$$

This plot is approximately linear along 45° line going through the origin when the data follows a power law process. Such a plot with points lying close to the diagonal line going through the origin indicates a good fit. As can be seen from the plots in the figures of Appendix A.3, all except for the Region of Waterloo and London have scatter plot points following a line going through the origin. This again gives us the same results as that in the Section **??**. As a conclusion, we can easily say that power law model governs the Wind Speed data obtained from the eight weather stations within the province of Ontario.

Efficiency of the estimator

The **efficiency** is given by

$$\text{Efficiency} = \frac{\text{MSE of } \hat{u}}{\text{MSE of } \bar{u}}. \tag{3.11}$$

The results for the efficiency calculations are shown in the Table 3.14. From 3.14, we can see that efficiency is less than 1 for all the cases except for the Region of Waterloo. Therefore, we can say that power law process is an adequate model for the process as it is followed by the data from a maximum number of weather stations.

Table 3.14: Efficiency of estimator $\bar{u}$ relative to $\hat{u}$

Station Name	N	t	$\hat{\beta}$	$\bar{u}$	$\hat{u}$	Efficiency
Trenton	831	22721	0.7581	0.0365	0.0277	0.759
Region of Waterloo	526	13882	1.0032	0.0379	0.0380	1.003
London	369	10002	0.9123	0.0369	0.0336	0.910
Hamilton	699	15906	0.8157	0.0439	0.0358	0.816
Sarnia	598	16059	0.9259	0.0372	0.0345	0.927
Wiarton	608	16517	0.8170	0.0368	0.0301	0.817
Toronto Island	658	15741	0.8568	0.0418	0.0358	0.856
Toronto Pearson	734	17250	0.8978	0.0425	0.0382	0.898
Kapuskasing	506	20539	1.0017	0.02463	0.02467	1.001
Sudbury	758	22170	0.6700	0.0341	0.0229	0.669

3.5 Conclusion

The basis of this analysis was to find out whether the power law process governs the Wind Speed data gathered from the ten Ontario meteorological stations. Based on the Graphical methods of Section 3.4.1 and tests from Sections 2.7 and 2.7.1 we can say that the power law process is an adequate model for the process as it is followed by the data from a maximum number of weather stations. These tests were conducted because the value of the parameter β for NHPP was coming out to be very close to 1 but not exactly equal to 1. Moreover, from the example of Trenton the non-stationary daily rate of change of wind speed for the Stochastic Process model was 0.0355 KMPH (λ) whereas the annual rate of change according to the annual maxima model was 0.472 KMPH (a_1). Using the same parameters, we can interpret this rate of change for the remaining weather stations.

Chapter 4

Forecast of Maximum Wind Speed

4.1 Introduction

This chapter evaluates the 95^{th} percentile of annual maximum wind speed using Annual Maxima and Stochastic Process model. From Figure 4.1 the extreme wind events show a slight overall decrease in annual maximum wind speeds over the last few years. To further verify this trend, we compare the two models in the following sections. Table 4.1 provides a summary of all the important parameters.

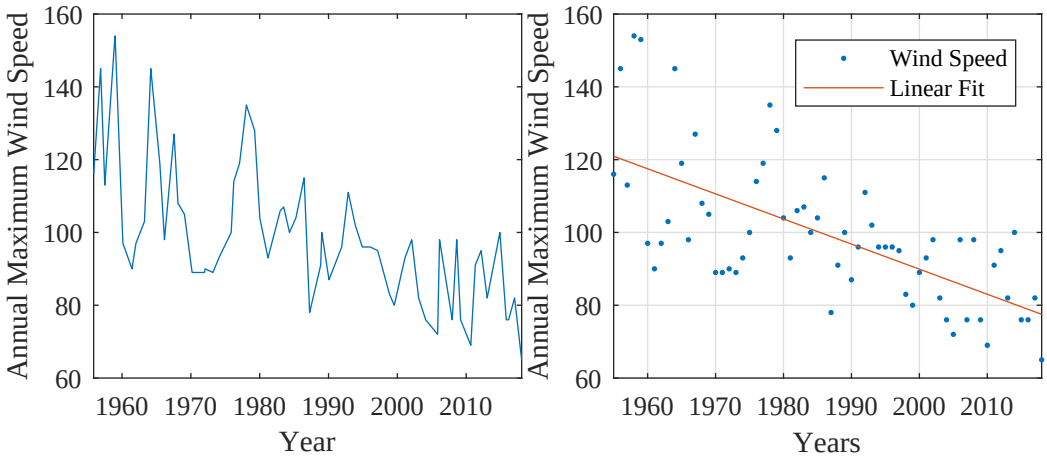
Table 4.1: Model parameters - Summary

Station	**Location function**	a_0	a_1	β	λ
Trenton	Constant	90.288		16.142	
	G-Linear	108.307	-0.472	11.480	
	HPP				0.0214
	NHPP			0.9775	0.0355
Region of Waterloo	Constant	88.312		9.779	
	G-Linear	99.985	-0.707	18.839	
	HPP				0.0183
	NHPP			0.9708	0.0260
London	Constant	96.256		13.025	
	G-Linear	117.956	-2.186	19.947	
	HPP				0.0143
	NHPP			0.8509	0.0756

Station	Location function	a_0	a_1	β	λ
Hamilton	Constant	96.302		9.904	
	G-Linear	106.323	-0.380	9.543	
	HPP				0.0291
	NHPP			0.8453	0.1285
Sarnia	Constant	97.885		13.289	
	G-Linear	94.534	-0.240	12.105	
	HPP				0.0200
	NHPP			1.1751	0.0048
Wiarton	Constant	90.821		9.647	
	G-Linear	96.937	-0.225	9.944	
	HPP				0.0210
	NHPP			0.8580	0.0866
Toronto Island	Constant	91.765		8.539	
	G-Linear	95.357	-0.055	8.432	
	HPP				0.0258
	NHPP			1.0399	0.0191
Toronto Pearson	Constant	93.234		7.428	
	G-Linear	98.105	-0.207	7.775	
	HPP				0.0265
	NHPP			0.9712	0.0348
Kapuskasing	Constant	79.824		9.384	
	G-Linear	87.105	-0.260	10.991	
	HPP				0.0102
	NHPP			1.1082	0.0042
Sudbury	Constant	85.530		14.361	
	G-Linear	104.934	-0.570	11.554	
	HPP				0.0198
	NHPP			0.8283	0.1500

4.1.1 Annual Maxima Method

From the calculations above, we know that our wind speed data for the weather station of Trenton is for the duration of 64 years. Thus, for year 2020 the duration would change to 66 and would then increment by 1 to make predictions for the years following 2020. This is shown in the Table B.1.



(a) (b) with Trend-line

Figure 4.1: Annual Max Wind Speed vs. Years

Therefore, substituting these values of t and the parameters from Table 3.3 in Equation 2.47

$$x_{0.95} = 173.592 \; km/h$$

Performing the same calculation for annually using Equation **??**, we get

$$x_p = 142.880 \; km/h$$

Similarly, we can make predictions for the other years the results of which are shown in Table B.1.

pda

4.1.2 Stochastic Process Model

95^{th} percentile: NHPP

Our wind-speed data starts in year 1955 and ends in year 2018.

$$\therefore \text{Total number of years} = \text{Stop Year} - \text{Start Year} + 1 = 64$$

Thus to calculate t in days for Equation 2.69,

$$(64 \times 365) + 16 = 23376$$

Here the number 16 accounts for the additional days for the leap years between the interval of 1955 to 2018. Based on this, we can say that the year 2020 will have 24106 days.

Considering the threshold of $u = 70$KMPH, from Table 3.11 we have $\lambda = 0.0355$ and $\beta = 0.9775$.

Therefore using Equation 2.70,

$$\Lambda(t) = \lambda(t)^{\beta} = 681.294$$

Now, we know from Table 3.9 that the scale parameter $\lambda = 0.083$ for exponential distribution of wind-speed data above the threshold of 70KMPH. Moreover, as our current calculation is for the 95^{th} percentile, we take $p = 0.95$.

Substituting the scale parameter for exponential distribution and the other above calculted values in Equation 2.69,

$$x_{0.95} = 114.388$$

From Equation 2.71,

$$y_{0.95} = 184.388\ km/h$$

Similarly we solve for the remaining projection years. The solutions to the above calculations for years 2020 to 2100 are shown in the Table B.2.

Moving on to finding the distribution (t_1, t_2), where $t_1 = 24106$ days and $t_2 = 24471$ days, we can now calculate the value of $\Lambda(t_1, t_2)$ using Equation 2.72 and the values of $\Lambda(t)$ from Table B.2.

$$\begin{aligned}
\Lambda(t_1, t_2) &= \Lambda(0, t_2) - \Lambda(0, t_1) \\
&= \lambda(t_2)^{\beta} - \lambda(t_1)^{\beta} \\
&= 691.362 - 681.294 \\
&= 10.067
\end{aligned}$$

Using the above solution, we can now calculate $x_{0.95}$ and $y_{0.95}$ for the interval (t_1, t_2) using Equations 2.73 and 2.71.

$$\therefore \textbf{Annual } y_{0.95} = 133.608 \; km/h$$

Similarly, we can now calculate the remaining values shown in Table B.2.

95^{th} percentile: HPP

In total, the number of wind events above the threshold of $u = 70$ KMPH are 501. We also know from the previous section that for the duration starting from year 1955 and ending in 2018, our total observation period is 23376 days. Substituting these values in Equation 2.77,

$$\lambda = 0.0214$$

Using Equations 2.70 and 2.71 we get

$$y_{0.95} = 181.055 \; km/h.$$

Now using the values in Table B.3 in Equation 2.78, we get

$$\Lambda(t_1, t_2) = 7.823$$

Thus for the interval (t_1, t_2), using Equations 2.73 and 2.71, we get

$$\textbf{Annual } y_{0.95} = 130.570 \; km/h.$$

Projections for the years starting from 2020 to 2100 are shown in Table B.3.

4.2 Model Comparison

Figure 4.2 shows three plots in total. Plot (a) was plot using the values from Table B.1. Plot (b) was plot using the values from Table B.2. Finally, plot (c) was plot using the values from Table B.3. A summary of these three tables is also available in Table 4.1.

Table 4.2: Forecast Comparison: Trenton

Model	% Change	Forecast Year 2020	Forecast Year 2100
Gumbel-Linear	0.472 KMPH	111.222 KMPH	73.422 KMPH
NHPP	0.035 KMPH	133.608 KMPH	133.381 KMPH

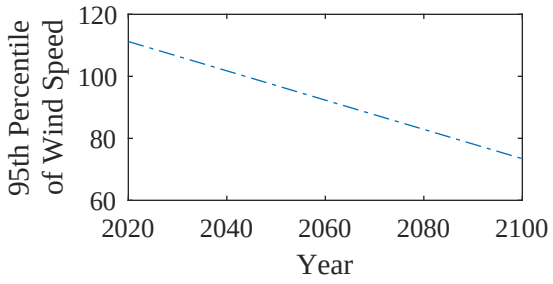
Plot (b) for NHPP clearly shows a downward trend; that is, it shows a decrease in wind-speed. However, this downward trend is very minor. Looking at the values of Annual $x_{0.95}$ from Table B.2 we can see that only the numbers after the decimal point vary keeping the wind-speed to stay around 133.XXX km/h. The plots (a) for the non-stationary Gumbel shows a very steep decline. This verifies the fact that the annual maximum wind speed is reducing. Stationary model of HPP in Figure (c) on the other hand, does not show any change. Essentially, it says that the maximum wind-speed events should stay constant.

Because the predicted wind-speed via NHPP varies around 133 km/h, if we plot all the three plots on a single plot for comparison, the plot (b) for NHPP appears to be almost a straight line. That is also the reason why it was chosen to plot the three cases separately just so that trend is visible clearly just for this one instance. Figure 4.3 shows the 95th percentile plots for the 10 weather stations. Table 4.2 shows the amount by which the 95th percentile of annual maximum wind speed will reduce.

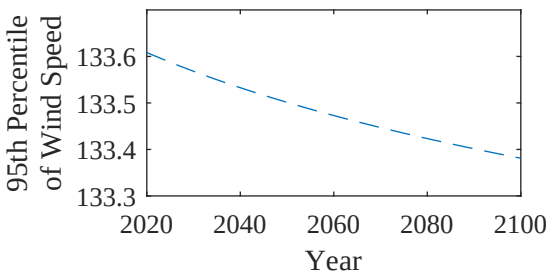
The National Building Code of Canada (NBC) used annual maximum data fitted over Gumbel distribution to calculate the 1 in 50 year chance of being exceeded in any year. The comparision of these 1 in 50 (1/50) year wind speed with the non-stationary Gumbel model is shown in the Table 4.3. While the NBC only predicts the wind speed for a 1/50 year using a stationary model, the current analysis of this thesis allows us to look at the decreasing trend of the wind speeds in the coming years.

Table 4.3: Comparison of National Building Code of Canada 1/50 year probability of wind speed with the stationary and non-stationary Gumbel model

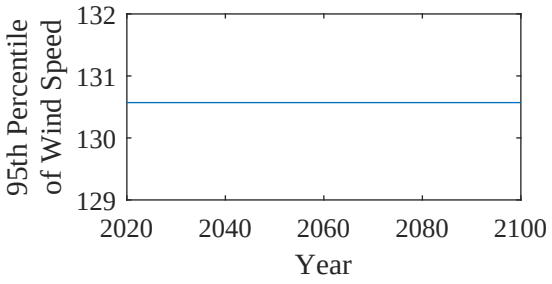
Station	NBC 1/50 (KMPH)	Stationary Gumbel (KMPH)	Non-Stationary Gumbel	
			Year 2020 (KMPH)	Year 2100 (KMPH)
Trenton	97.20	90.28	111.22	72.94
Region of Waterloo	86.04	88.31	121.29	64.02
London	97.20	96.25	125.47	113.71
Hamilton	96.12	96.30	116.01	85.53
Sarnia	97.20	97.88	118.69	99.19
Wiarton	97.20	90.82	114.98	96.74
Toronto Island	93.96	91.76	110.63	93.86
Toronto Pearson	93.96	93.23	117.55	113.04
Kapuskasing	78.84	79.82	104.12	83.01
Sudbury	96.12	85.53	102.71	56.47



(a) Annual Gumbel Prediction for Gumbel

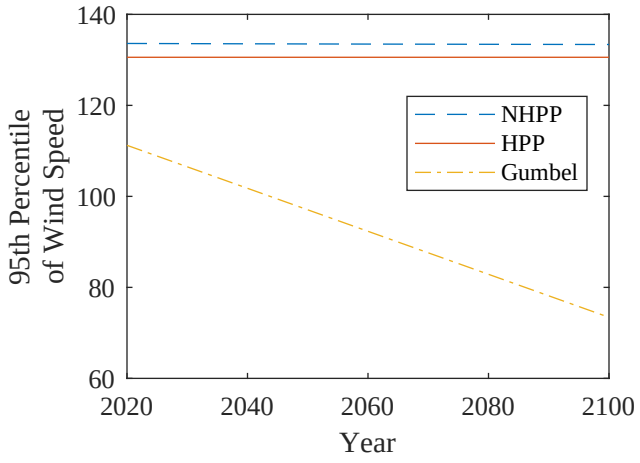


(b) Annual NHPP Prediction

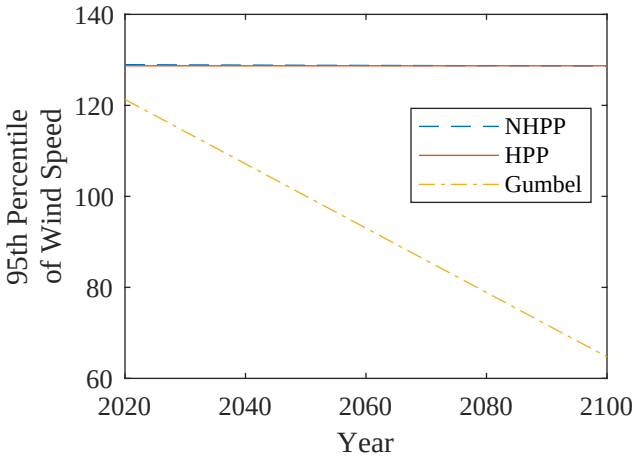


(c) Annual HPP Prediction for HPP

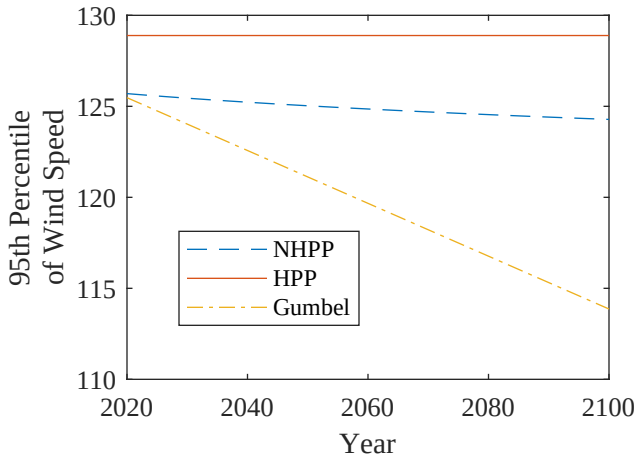
Figure 4.2: Plots of 95th Percentile vs. Year: Trenton



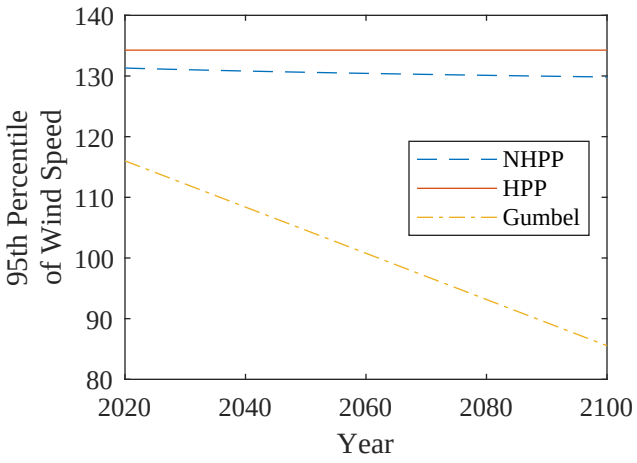
(a) Trenton



(b) Region of Waterloo

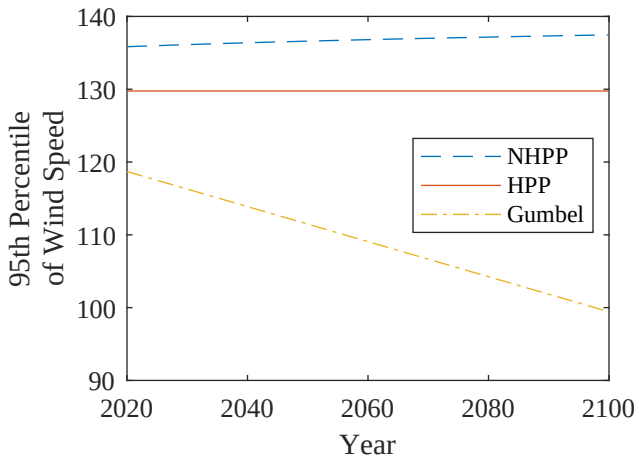


(c) London

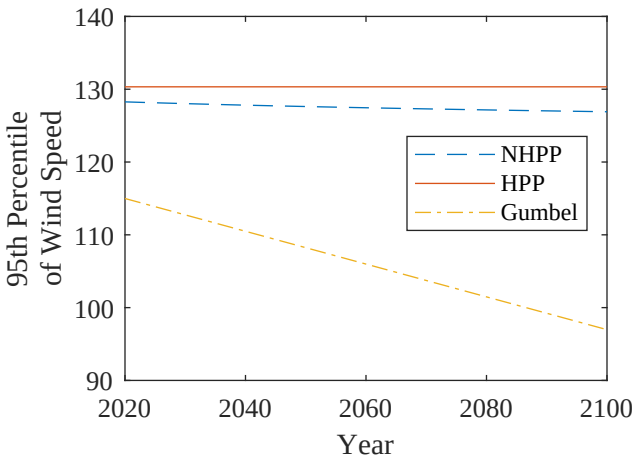


(d) Hamilton

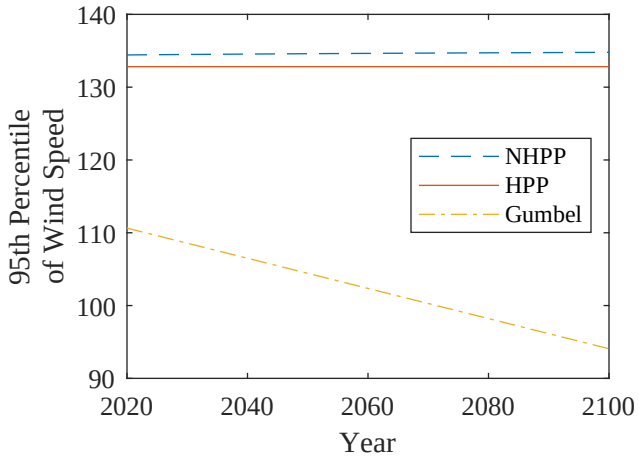
Figure 4.3: 95th Percentile Plots



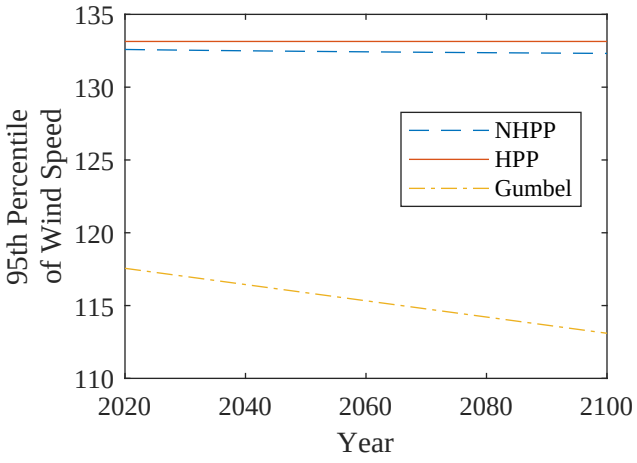
(e) Sarnia



(f) Wiarton

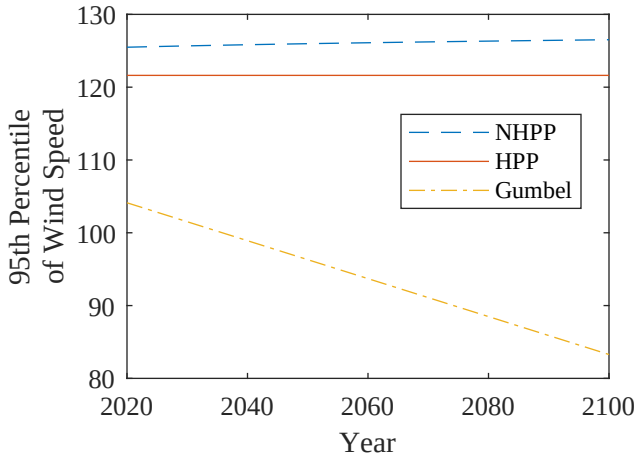


(g) Toronto Island

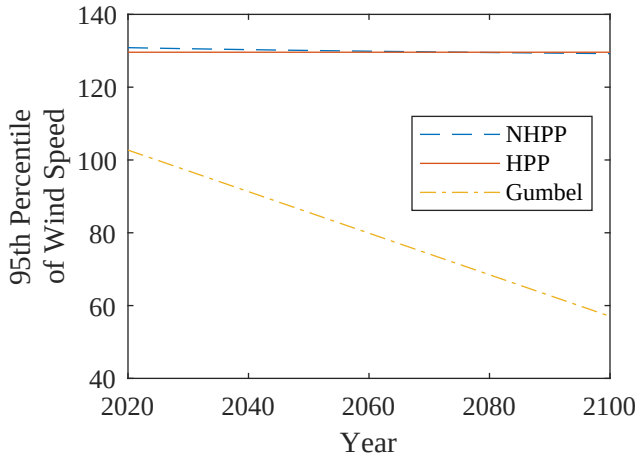


(h) Toronto Pearson

Figure 4.3 (cont.): 95th Percentile Plots



(i) Kapuskasing



(j) Sudbury

Figure 4.3 (cont.): 95th Percentile Plots

4.3 Conclusion

Overall, the non-stationary plots show a downward trend. This indicates that when we assume extreme wind events to be non-stationary, these events are forecasted to reduce gradually over the coming years. This is also evident from the non-stationary parameters calculated in the previous Chapter. For the example of Trenton, the 95th percentile of annual maximum wind speed is expected to reduce at the rate of 0.0059 KMPH annually over the period of 80 years starting from 2020 until 2100.

Chapter 5

Summary

Global warming and climate change are being discussed all over the globe these days because it factually evident that these changes are imperiling the planet. Prevention is better than cure. Therefore, it is of our best interest to utilize the historical data collected by Environment Canada to predict what the future holds. The focus of this thesis is to use the wind speed data available on Environment Canada website to investigate probabilistic methods for modelling the non-stationary nature of wind data.

The predictions made by this thesis can help the construction industry design buildings that are able to withstand the changes in the climate that the Earth's atmosphere will experience in the coming years. Unlike electronics, buildings are made to last. The findings of this thesis, hence, can also help while calibrating the future versions of building codes of Canada.

The current code calibrations are made under the assumption that the environmental loads are generated by a stationary process. Under stationarity, each year has a uniform annual probability of failure. This is not the case when we consider climate to be non-stationary because of global warming. The 95th and 99th percentile plots of the power law model confirm that the wind speed is decreasing at a very slower pace, on average. This average is much higher for few of stations where this decreasing trend is clearly visible in the comparison plots of Annual Maxima and Stochastic Process models. It is also important to point out, here, that the stations of Kapuskasing and Sarnia also exhibited an upward trend for the wind speed analysis. Furthermore, it is also necessary to note that for both the wind speed data analysis, the NHPP parameter a_1 includes the values of 1, so there is no evidence to reject the HPP model.

Based on the percentile plots for the Stochastic Process model, the probability of failure

will change over the service life of buildings. That means that reliability would depend on a specific calendar period. Climate change is a gradual process. If measures are are eventually put in place to curb global warming, this change in climate will reduce gradually as well. Therefore, the reliability in 2020-2050 could be different from the reliability in 2050-2100.

This thesis uses weather stations only from the province of Ontario. This research could be extended to identify the changes in the wind speed for several other provinces of Canada. Again, the results from that could be compared to the current analysis to provide an overall perspective on weather change.

According to NBC the average 1/15 year chance of wind speed exceeding in a year is by 93.45KMPH. This on average falls between the values of the model parameters for wind speed data that provide us with the forecasted rate at which the wind speed will change over the coming years. For the non-stationary case of Annual Maxima model, the parameter a_1 provides the annual rate of change of wind speed. Therefore, for the example of Trenton the 95th percentile of annual maximum wind speed will reduce at 0.0059 KMPH annually over the period of 80 years starting from 2020 until 2100. The rate of change can be identified for the remaining weather stations using the same model parameters.

References

An, Y. (2006). *Extreme value models for the estimation of design wind speed*. University of Waterloo, Waterloo, Ont.

Bruce, P. C. (2017). *Practical statistics for data scientists : 50 essential concepts*. O'Reilly Media, Inc., Sebastopol, CA, first edition. edition.

Castillo, E. (1988). *Extreme value theory in engineering*. Statistical modeling and decision science. Academic Press, Boston ; Toronto.

Castillo, E., Hadi, A. S., Balakrishnan, N., and Sarabia, J. M. (2005). *Extreme value and related models with applications in engineering and science*. Wiley series in probability and statistics. Wiley, Hoboken, N.J.

Cohen, J. P. (1986). Large sample theory for fitting an approximating gumbel model to maxima. *Sankhya: The Indian Journal of Statistics, Series A (1961-2002)*, 48(3):372–392.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer series in statistics. Springer, London ; New York.

Danard, M., Munro, A., and Murty, T. (2003). Storm surge hazard in canada. *Natural Hazards*, 28(2):407–434.

Davis, R. A., Leadbetter, M. R., Lindgren, G., and Rootzen, H. (1985). Extremes and related properties of random sequences and processes. *Journal of the American Statistical Association*, 80(389).

Dawson, J. W. (2011). The use of crow-amsaa plots to assess mishap trends.

de Brito, E., Silva, G. O., Cordeiro, G. M., and Demétrio, C. G. B. (2016). The mcdonald gumbel model. *Communications in Statistics: Theory and Methods*, 45(11):3367–3382.

DeBock, D. J., Liel, A. B., Harris, J. R., Ellingwood, B. R., and Torrents, J. M. (2017). Reliability-based design snow loads. i: site-specific probability models for ground snow loads.(report)(author abstract). *Journal of Structural Engineering*, 143(7).

Finkenstadt, B. and Rootzen, H. (2004). *Extreme values in finance, telecommunications, and the environment*. Monographs on statistics and applied probability ; 99. Chapman and Hall/CRC, Boca Raton, Fla.

Gorgoso-Varela, J. and Rojo-Alboreca, A. (2014). Use of gumbel and weibull functions to model extreme values of diameter distributions in forest stands. *Annals of Forest Science*, 71(7):741–750.

Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55(292):698–707.

Haan, L. d. L. (2006). *Extreme value theory an introduction*. Springer series in operations research. Springer, New York.

Holmes, J. and Moriarty, W. (1999). Application of the generalized pareto distribution to extreme value analysis in wind engineering. *Journal of Wind Engineering and Industrial Aerodynamics*, 83(1-3):1–10.

Hong, H., Li, S., and Mara, T. (2013). Performance of the generalized least-squares method for the gumbel distribution and its application to annual maximum wind speeds. *Journal of Wind Engineering & Industrial Aerodynamics*, 119:121–132.

Hong, H., Mara, T., Morris, R., Li, S., and Ye, W. (2014). Basis for recommending an update of wind velocity pressures in canadian design codes. *Canadian Journal of Civil Engineering*, 41(3):206–221.

Johnson, D. (2017). Is 2017 the worst hurricane season ever?

Kang, D., Ko, K., and Huh, J. (2015). Determination of extreme wind values using the gumbel distribution. *Energy*, 86:51–58.

Lechner, J. A., Leigh, S. D., and Simiu, E. (1992). Recent approaches to extreme value estimation with application to wind speeds. part i: the pickands method. *Journal of Wind Engineering and Industrial Aerodynamics*, 41(1-3):509–519.

Lee, B.-H., Ahn, D.-J., Kim, H.-G., and Ha, Y.-C. (2011). An estimation of the extreme wind speed using the korea wind map. *Renewable Energy*, 42(C).

Lombardo, F. T. and Ayyub, B. M. (2015). Analysis of washington, dc, wind and temperature extremes with examination of climate change for engineering applications. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 1(1):04014005.

Martinez, W. L. (2008). *Computational statistics handbook with MATLAB*. Chapman and Hall, Boca Raton, FL, second edition edition.

NASA (2019). Climate change evidence: How do we know?

Perrin, O., Rootzén, H., and Taesler, R. (2006). A discussion of statistical methods used to estimate extreme wind speeds. *Theoretical and Applied Climatology*, 85(3):203–215.

Pes, M. P., Pereira, E. B., Marengo, J. A., Martins, F. R., Heinemann, D., and Schmidt, M. (2017). Climate trends on the extreme winds in brazil. *Renewable Energy*, 109:110–120.

Reiss, R.-D. R.-D. (2001). *Statistical analysis of extreme values : from insurance, finance, hydrology, and other fields*. Birkhauser Verlag, Basel ; Boston, 2nd ed. edition.

Rigdon, S. E. (2000). *Statistical methods for the reliability of repairable systems*. Wiley series in probability and statistics. Applied probability and statistics. Wiley, New York ; Chichester [England].

Ross, W. H. (1987). A peaks-over-threshold analysis of extreme wind speeds. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 15(4):328–335.

Rosso, G. (2015). Extreme value theory for time series using peak-over-threshold method.

Sa, J. P. M. d. (2003). *Applied statistics using SPSS, STATISTICA, and MATLAB*. Springer, Berlin ; New York.

Saeed Far, S. and Abd. Wahab, A. K. (2016). Evaluation of peaks-over-threshold method. *Ocean Science Discussions*, 2016:1–25.

Simiu, E. and Heckert, N. (1996). Extreme wind distribution tails: a "peaks over threshold" approach. *Journal of Structural Engineering*, 122(5).

Singh, V. P. V. P. (2007). *Risk and reliability analysis : a handbook for civil and environmental engineers*. ASCE Press/American Society of Civil Engineers, Reston, Va.

Stephens, M. A. (1970). Use of the kolmogorov-smirnov, cramer-von mises and related statistics without extensive tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 32(1):115–122.

Tukey, J. W. J. W. (1977). *Exploratory data analysis*. Addison-Wesley series in behavioral science. Addison-Wesley Pub. Co. [1977], Reading, Mass.

Wan, H., Wang, X. L., and Swail, V. R. (2010). Homogenization and trend analysis of canadian near-surface wind speeds. *Journal of Climate*, 23(5):1209–1225.

Wilk, M. B. and Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika*, 55(1).

Wolinski, S. and Pytlowany, T. (2012). Evaluation of load values using the gumbel model. *Archives of Civil Engineering*, 58(2):199–208.

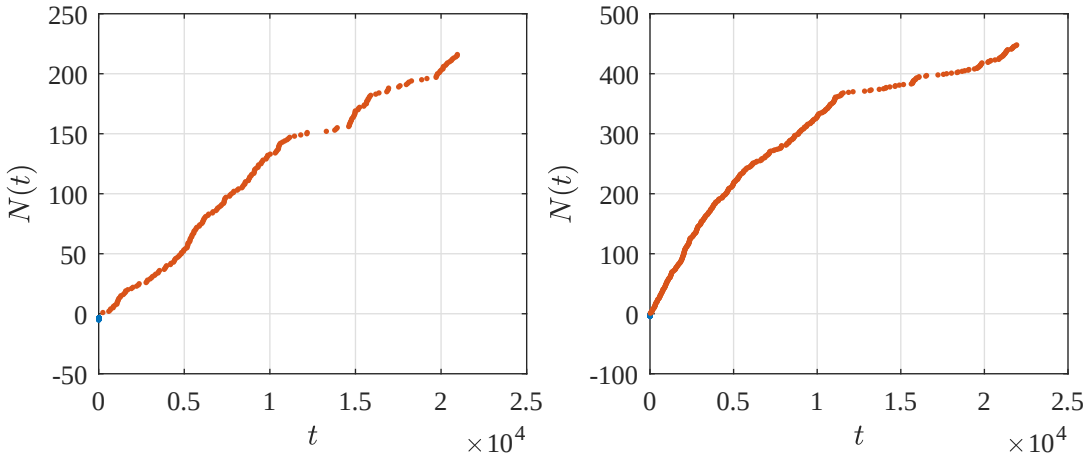
Yue, S. (2000). The gumbel mixed model applied to storm frequency analysis. *Water Resources Management*, 14(5):377–389.

APPENDICES

Appendix A

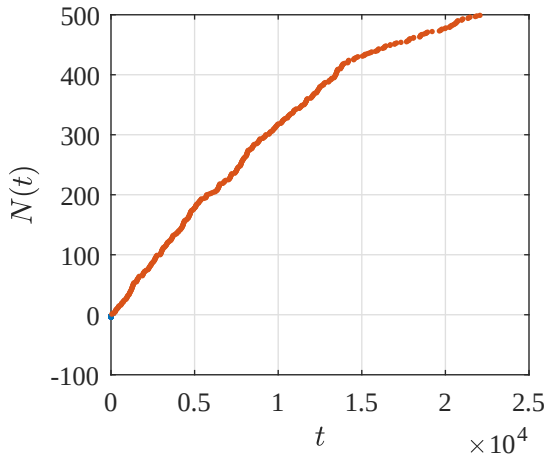
Plots

A.1 N(t) against t

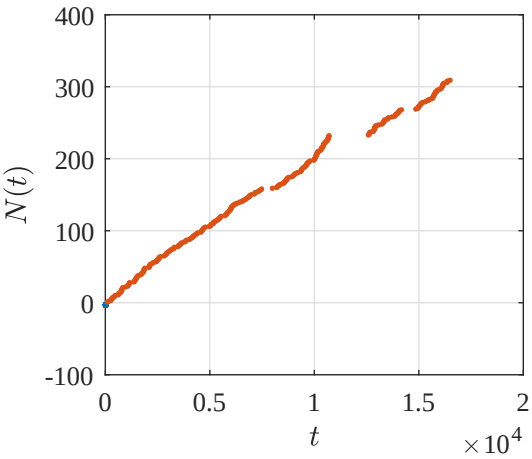


(a) Kapuskasing (b) Sudbury

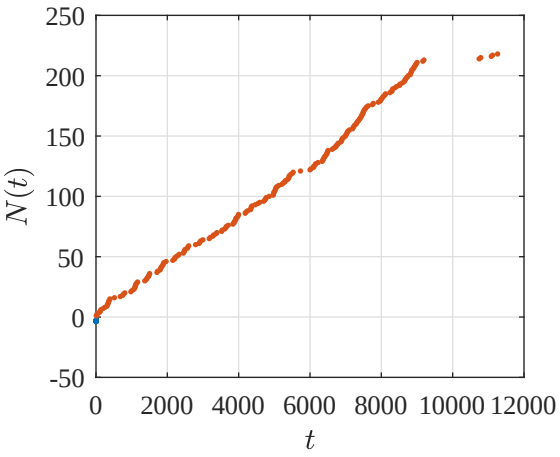
Figure A.1: Plots of N(t) against t



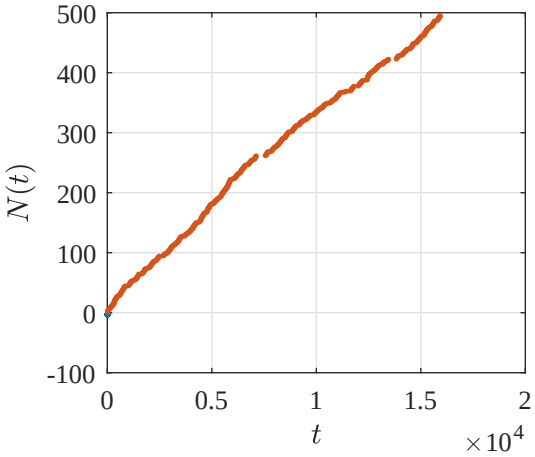
(c) Trenton



(d) Region of Waterloo

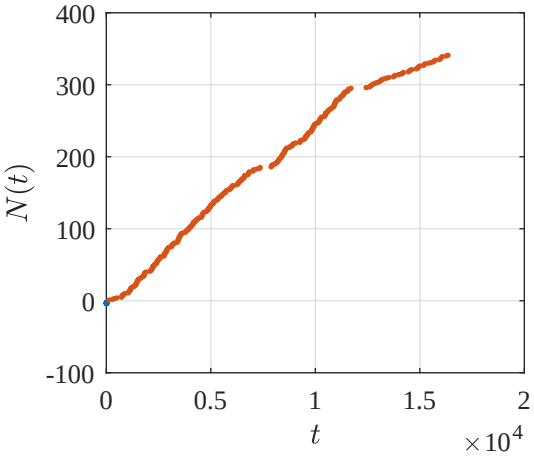


(e) London

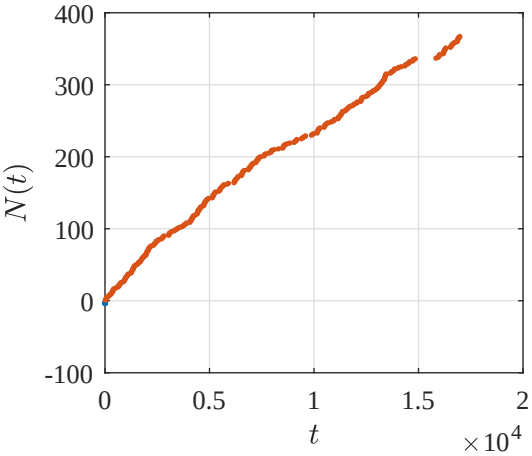


(f) Hamilton

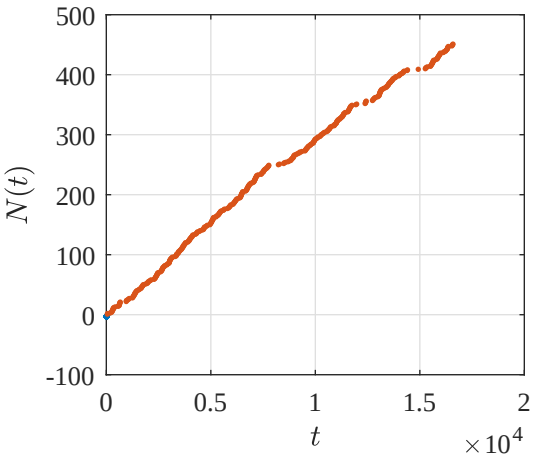
Figure A.1 (cont.): Plots of N(t) against t



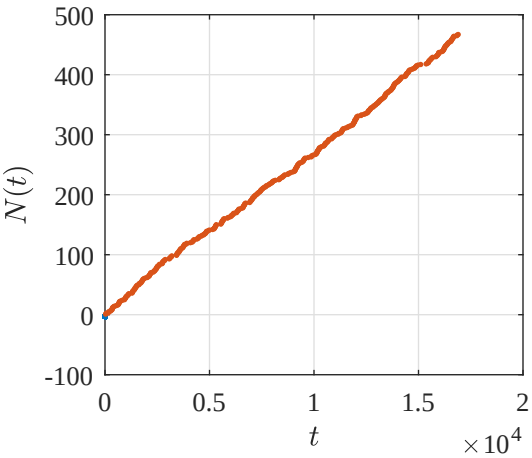
(g) Sarnia



(h) Wiarton



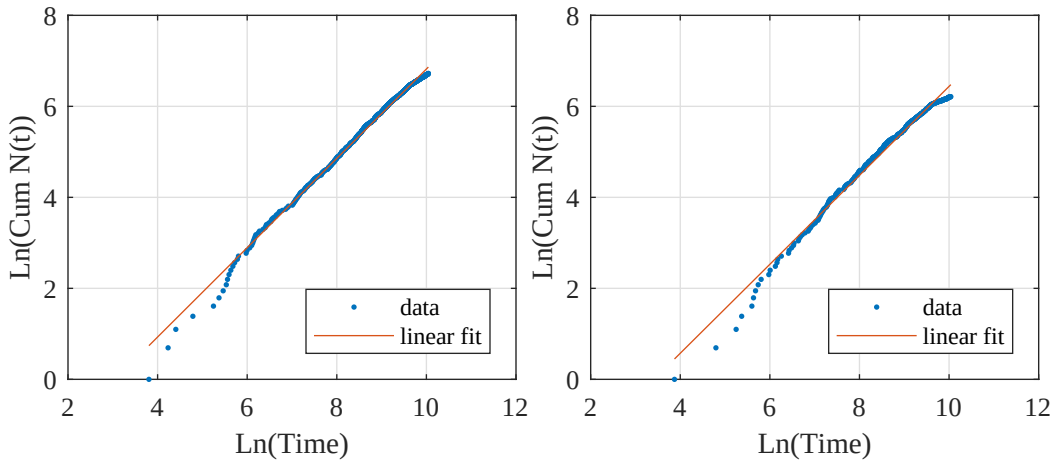
(i) Toronto Island



(j) Toronto Airport

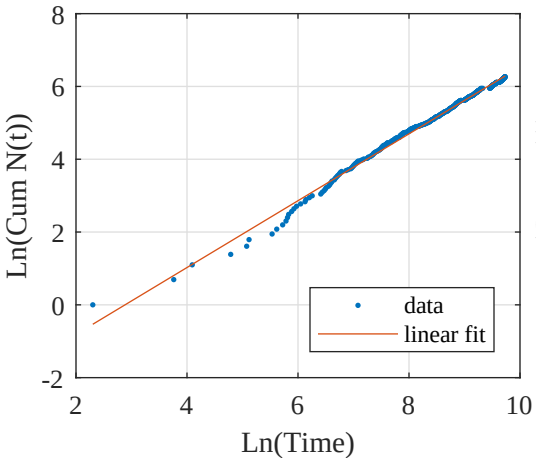
Figure A.1 (cont.): Plots of N(t) against t

A.2 CROW-AMSAA Plots

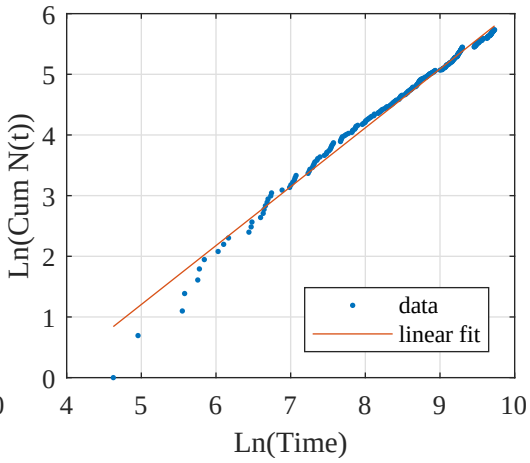


(a) Trenton: 60 KMPH

(b) Trenton: 70 KMPH

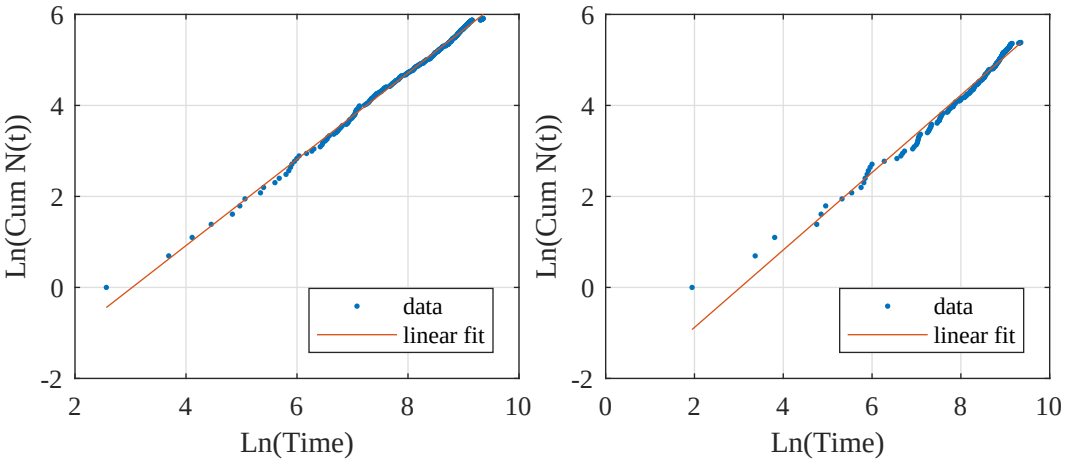


(c) Region of Waterloo: 60 KMPH



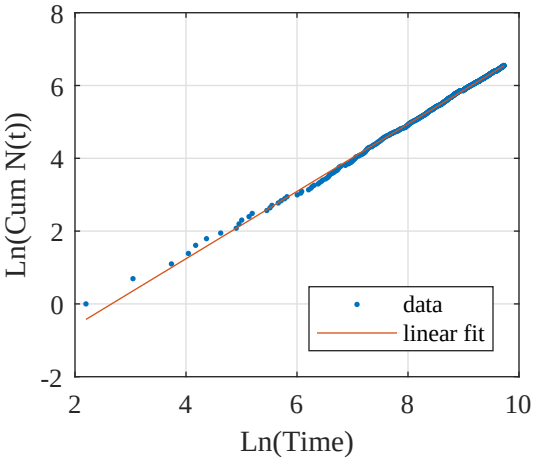
(d) Region of Waterloo: 70 KMPH

Figure A.2: CROW-AMSAA Plots to estimate NHPP paramters

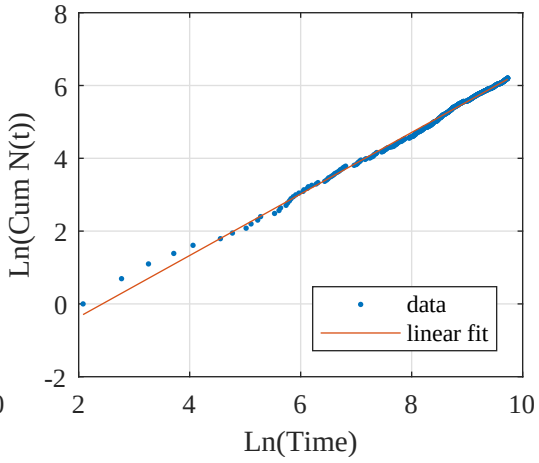


(e) London: 60 KMPH

(f) London: 70 KMPH

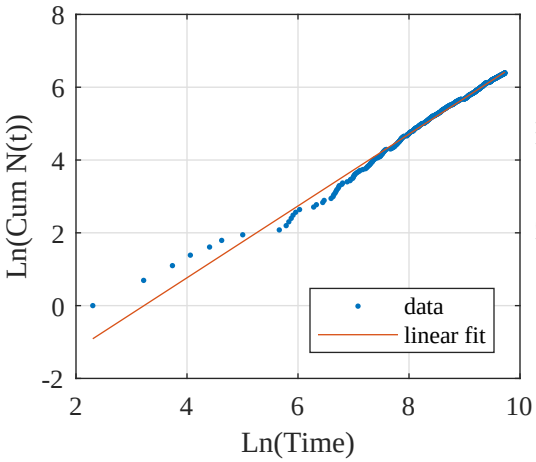


(g) Hamilton: 60 KMPH

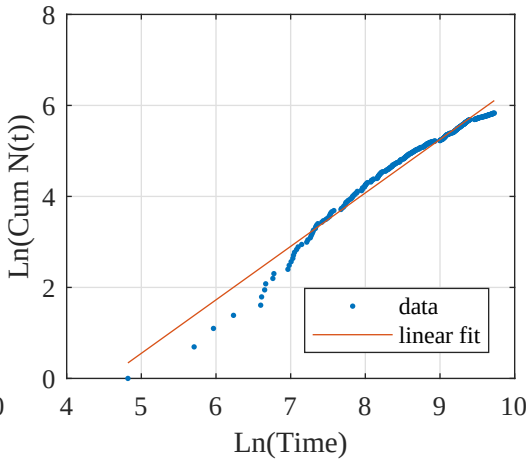


(h) Hamilton: 70 KMPH

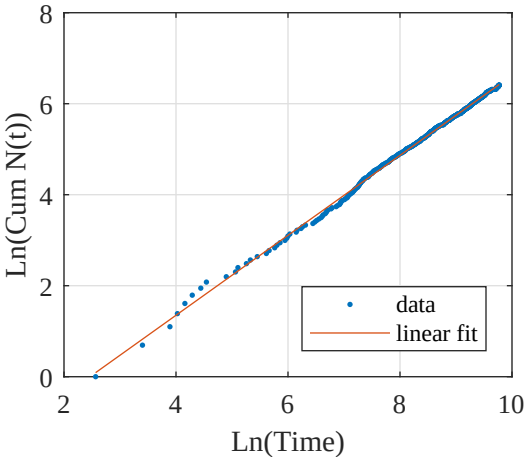
Figure A.2 (cont.): CROW-AMSAA Plots to estimate NHPP paramters



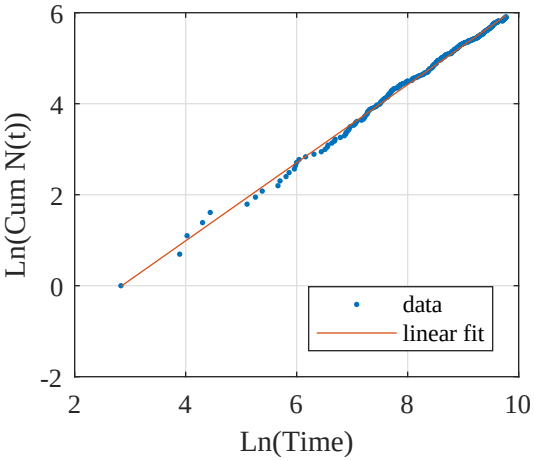
(i) Sarnia: 60 KMPH



(j) Sarnia: 70 KMPH

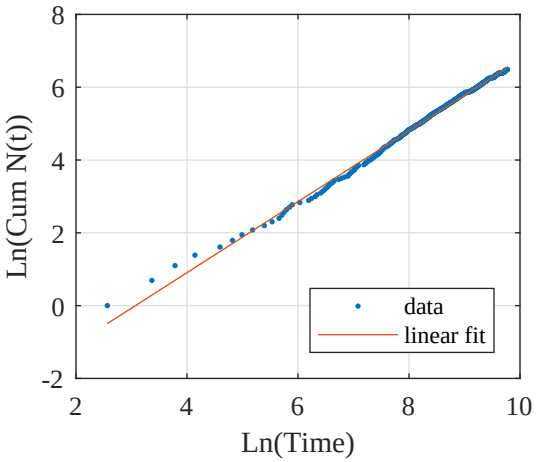


(k) Wiarton: 60 KMPH

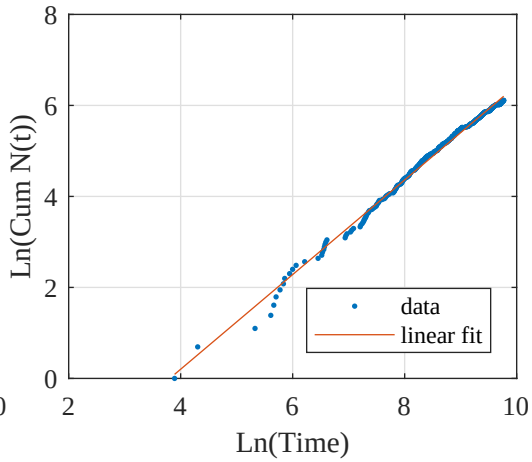


(l) Wiarton: 70 KMPH

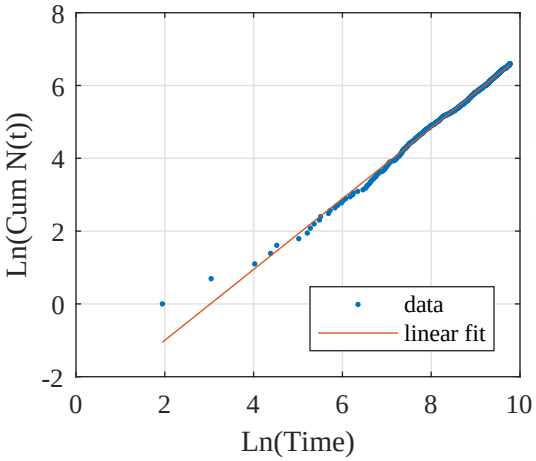
Figure A.2 (cont.): CROW-AMSAA Plots to estimate NHPP paramters



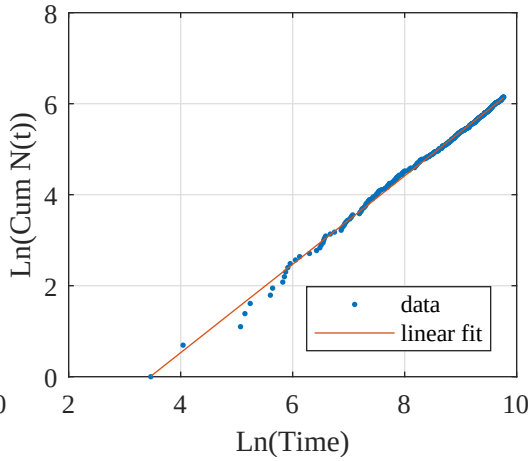
(m) Toronto Island: 60 KMPH



(n) Toronto Island: 70 KMPH

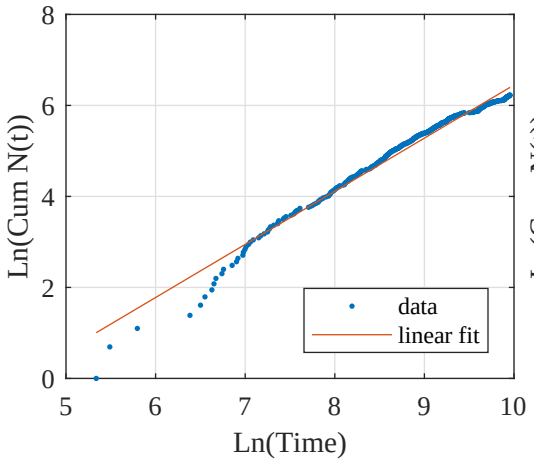


(o) Toronto Pearson: 60 KMPH

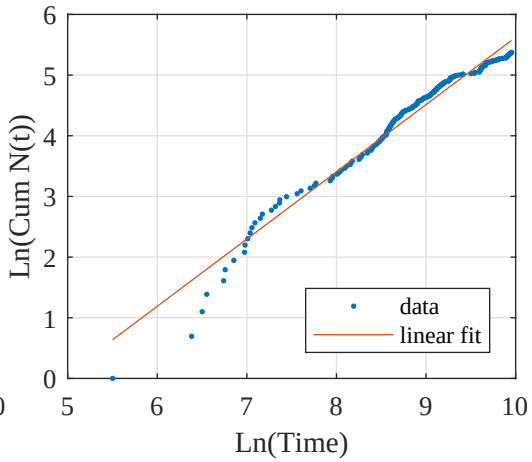


(p) Toronto Pearson: 70 KMPH

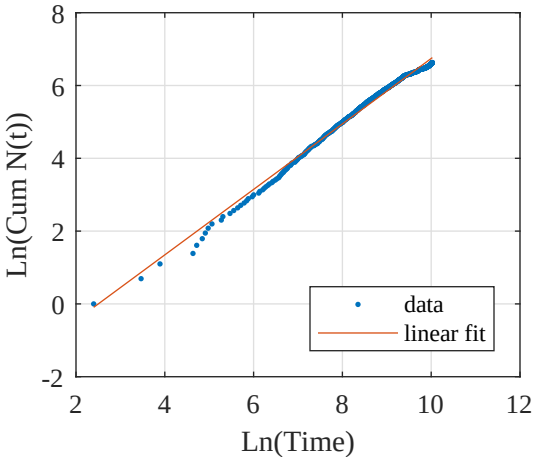
Figure A.2 (cont.): CROW-AMSAA Plots to estimate NHPP paramters



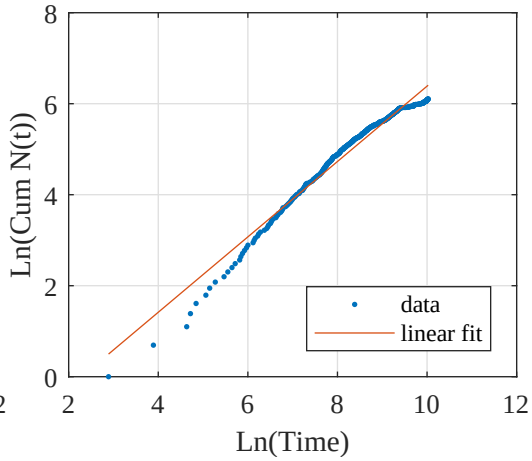
(q) Kapuskasing: 60 KMPH



(r) Kapuskasing: 70 KMPH



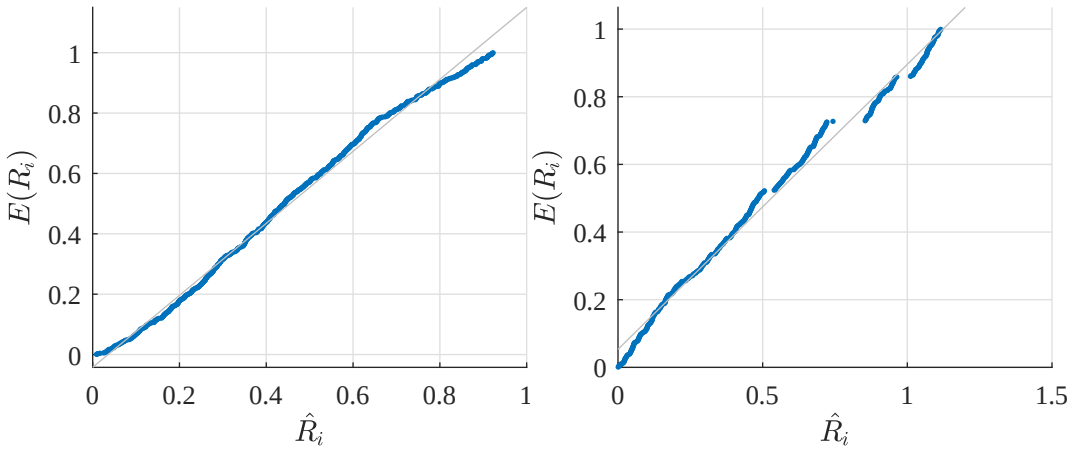
(s) Sudbury: 60 KMPH



(t) Sudbury: 70 KMPH

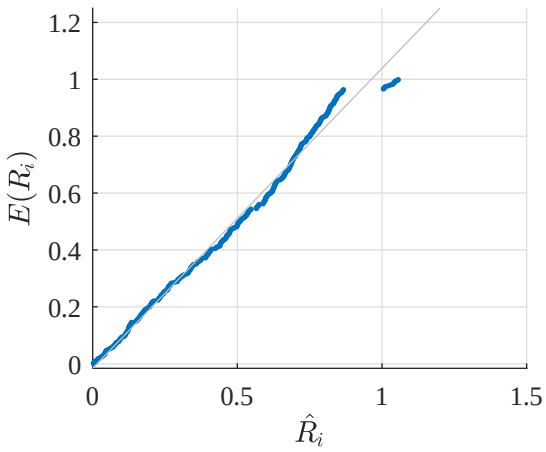
Figure A.2 (cont.): CROW-AMSAA Plots to estimate NHPP paramters

A.3 CVM test Plots

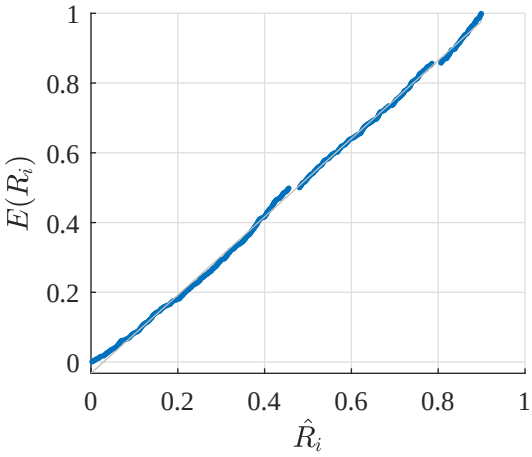


(a) Trenton

(b) Region of Waterloo

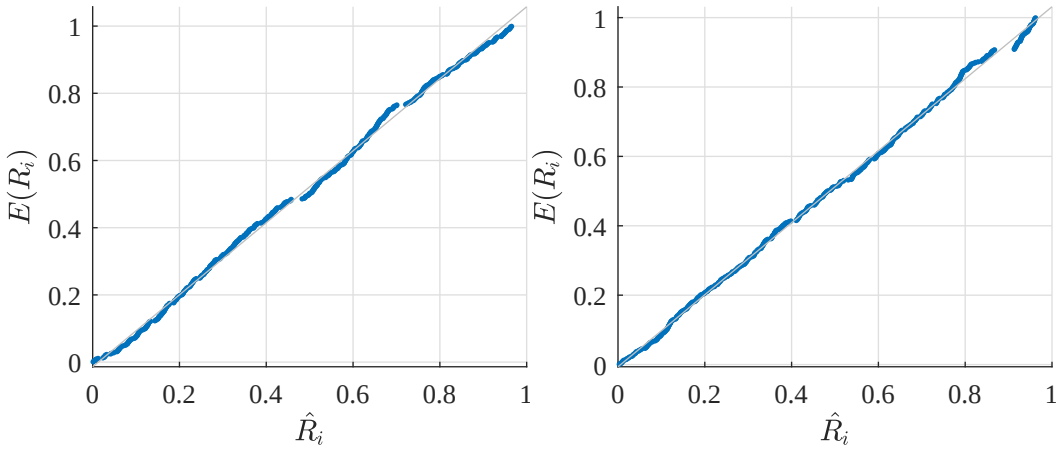


(c) London



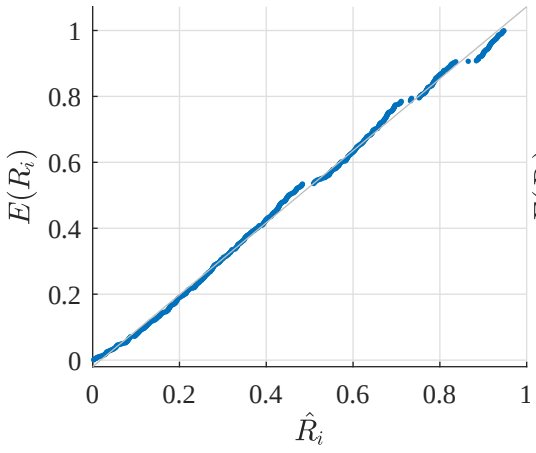
(d) Hamilton

Figure A.3: Plots for Cramer-von Mises test

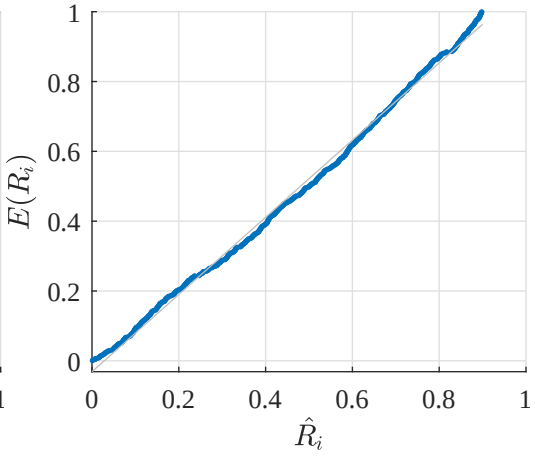


(e) Sarnia

(f) Wiarton

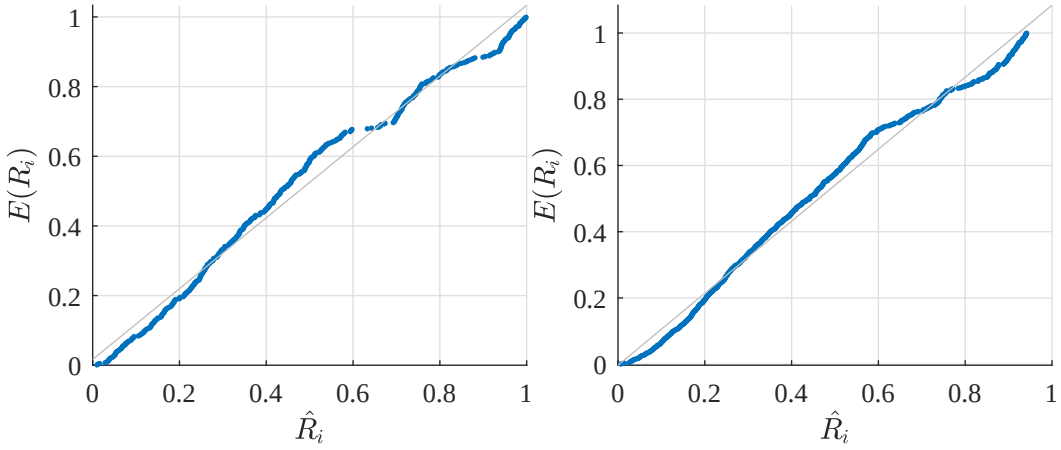


(g) Toronto Island



(h) Toronto Pearson

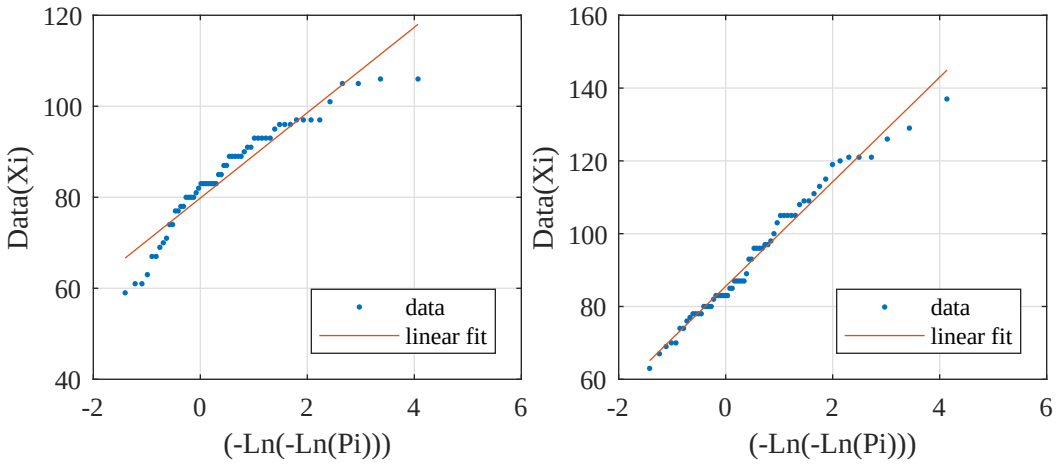
Figure A.3 (cont.): Plots for Cramer-von Mises test



(i) Kapuskasing (j) Sudbury

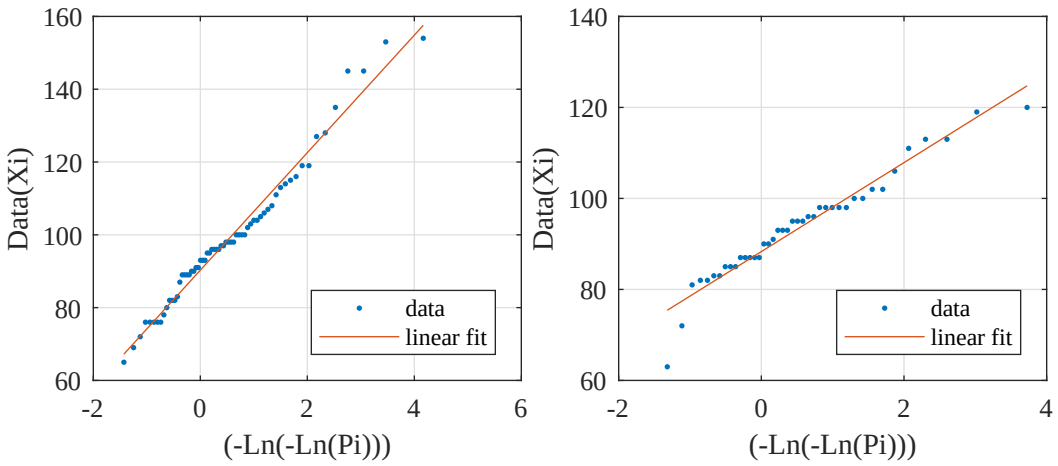
Figure A.3 (cont.): Plots for Cramer-von Mises test

A.4 Stationary Gumbel Model Plots

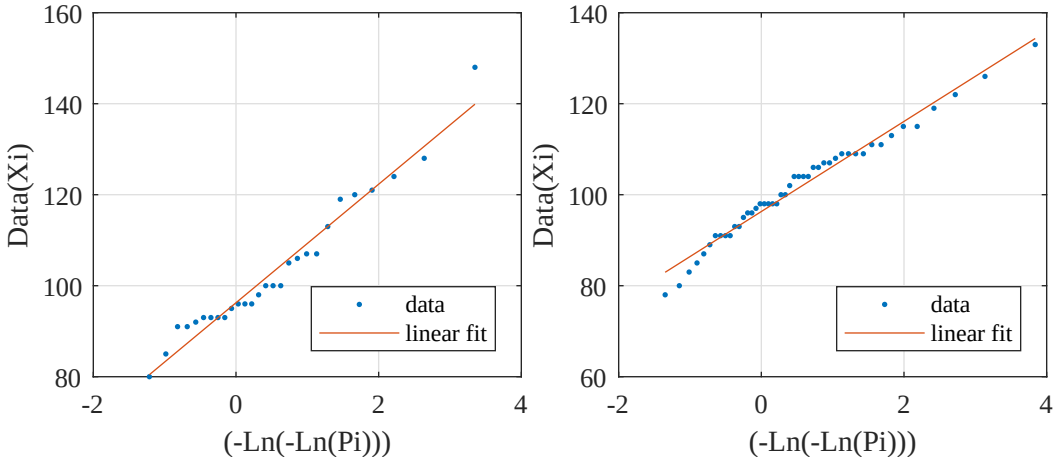


(a) Kapuskasing (b) Sudbury

Figure A.4: Plots for Gumbel Model

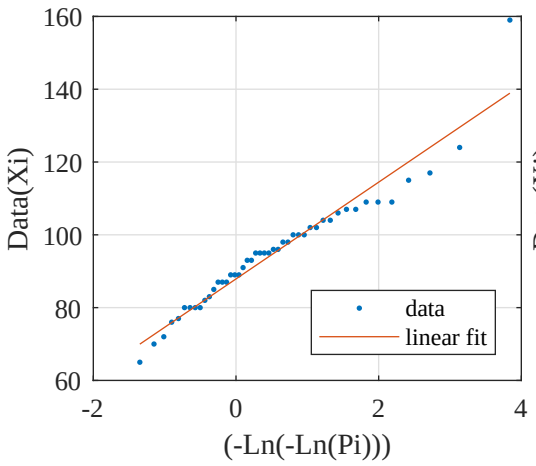


(c) Trenton (d) Region of Waterloo

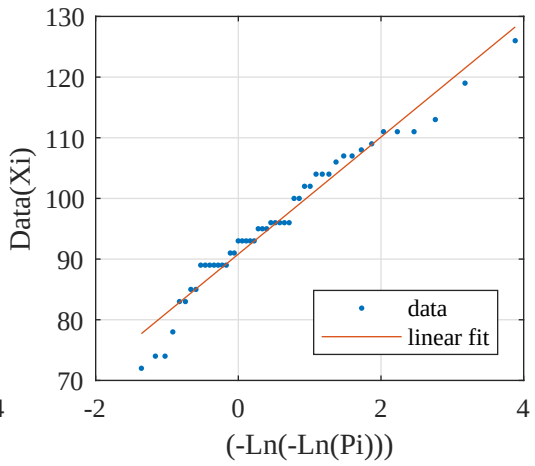


(e) London (f) Hamilton

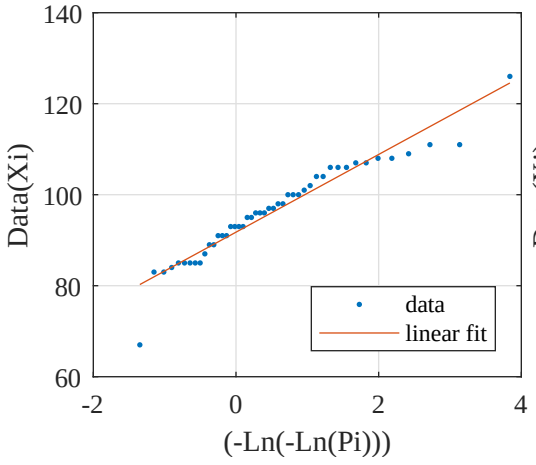
Figure A.4 (cont.): Plots for Gumbel Model



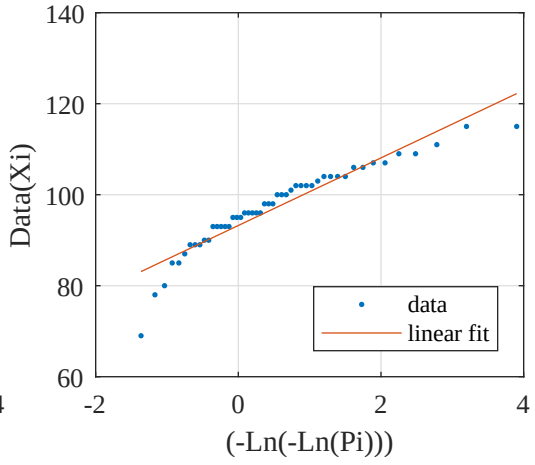
(g) Sarnia



(h) Wiarton



(i) Toronto Island



(j) Toronto Pearson

Figure A.4 (cont.): Plots for Gumbel Model

Appendix B

Tabular Solutions for 95th percentile using Gumbel, NHPP, and HPP

B.1 95th Percentile using Gumbel: Trenton

Table B.1: 95th Percentile using Non-Stationary Gumbel: Trenton

Year	**Number of Year t**	**Annual x_p**
2020	66	111.222
2021	67	110.750
2022	68	110.277
2023	69	109.805
2024	70	109.332
2025	71	108.860
2026	72	108.387
2027	73	107.915
2028	74	107.442
2029	75	106.970
2030	76	106.497
2031	77	106.025
2032	78	105.552
2033	79	105.080
2034	80	104.607

Table B.1 continued from previous page

Year	Number of Year t	Annual x_p
2035	81	104.135
2036	82	103.662
2037	83	103.190
2038	84	102.717
2039	85	102.245
2040	86	101.772
2041	87	101.300
2042	88	100.827
2043	89	100.355
2044	90	99.882
2045	91	99.410
2046	92	98.937
2047	93	98.465
2048	94	97.992
2049	95	97.520
2050	96	97.047
2051	97	96.575
2052	98	96.102
2053	99	95.630
2054	100	95.157
2055	101	94.685
2056	102	94.212
2057	103	93.740
2058	104	93.267
2059	105	92.795
2060	106	92.322
2061	107	91.850
2062	108	91.377
2063	109	90.905
2064	110	90.432
2065	111	89.960
2066	112	89.487
2067	113	89.015
2068	114	88.542

Table B.1 continued from previous page

Year	Number of Year t	Annual x_p
2069	115	88.070
2070	116	87.597
2071	117	87.125
2072	118	86.652
2073	119	86.180
2074	120	85.707
2075	121	85.235
2076	122	84.762
2077	123	84.290
2078	124	83.817
2079	125	83.345
2080	126	82.872
2081	127	82.400
2082	128	81.927
2083	129	81.455
2084	130	80.982
2085	131	80.510
2086	132	80.037
2087	133	79.565
2088	134	79.092
2089	135	78.620
2090	136	78.147
2091	137	77.675
2092	138	77.202
2093	139	76.730
2094	140	76.257
2095	141	75.785
2096	142	75.312
2097	143	74.840
2098	144	74.367
2099	145	73.895
2100	146	73.422

B.2 95th Percentile using NHPP-POT: Trenton

Table B.2: 95th Percentile using NHPP-POT: Trenton

Year	**Days** t	$\Lambda(t)$	$x_{0.95}$	$y_{0.95}$	**Annual** $\Lambda(t)$	**Annual** $x_{0.95}$	**Annual** $y_{0.95}$
2020	24106	681.294	114.388	184.388	10.067	63.608	133.608
2021	24471	691.362	114.565	184.565	10.064	63.604	133.604
2022	24836	701.426	114.739	184.739	10.060	63.600	133.600
2023	25201	711.486	114.910	184.910	10.057	63.596	133.596
2024	25566	721.543	115.079	185.079	10.053	63.592	133.592
2025	25931	731.596	115.246	185.246	10.050	63.588	133.588
2026	26296	741.646	115.410	185.410	10.047	63.584	133.584
2027	26661	751.693	115.573	185.573	10.043	63.580	133.580
2028	27026	761.736	115.732	185.732	10.040	63.576	133.576
2029	27391	771.776	115.890	185.890	10.037	63.572	133.572
2030	27756	781.813	116.046	186.046	10.034	63.568	133.568
2031	28121	791.847	116.200	186.200	10.031	63.564	133.564
2032	28486	801.878	116.351	186.351	10.028	63.561	133.561
2033	28851	811.905	116.501	186.501	10.025	63.557	133.557
2034	29216	821.930	116.649	186.649	10.022	63.553	133.553
2035	29581	831.951	116.795	186.795	10.019	63.550	133.550
2036	29946	841.970	116.939	186.939	10.016	63.546	133.546
2037	30311	851.986	117.082	187.082	10.013	63.543	133.543
2038	30676	861.999	117.222	187.222	10.010	63.539	133.539
2039	31041	872.009	117.361	187.361	10.007	63.536	133.536
2040	31406	882.016	117.499	187.499	10.004	63.533	133.533
2041	31771	892.020	117.635	187.635	10.002	63.529	133.529
2042	32136	902.022	117.769	187.769	9.999	63.526	133.526
2043	32501	912.021	117.902	187.902	9.996	63.523	133.523
2044	32866	922.017	118.033	188.033	9.994	63.520	133.520
2045	33231	932.011	118.163	188.163	9.991	63.517	133.517
2046	33596	942.001	118.292	188.292	9.988	63.513	133.513
2047	33961	951.990	118.419	188.419	9.986	63.510	133.510
2048	34326	961.976	118.544	188.544	9.983	63.507	133.507
2049	34691	971.959	118.669	188.669	9.981	63.504	133.504
2050	35056	981.940	118.792	188.792	9.978	63.501	133.501

Table B.2 continued from previous page

Year	**Days** t	$\Lambda(t)$	$x_{0.95}$	$y_{0.95}$	**Annual** $\Lambda(t)$	**Annual** $x_{0.95}$	**Annual** $y_{0.95}$
2051	35421	991.918	118.914	188.914	9.976	63.498	133.498
2052	35786	1001.894	119.034	189.034	9.973	63.495	133.495
2053	36151	1011.867	119.154	189.154	9.971	63.492	133.492
2054	36516	1021.838	119.272	189.272	9.969	63.490	133.490
2055	36881	1031.807	119.389	189.389	9.966	63.487	133.487
2056	37246	1041.773	119.505	189.505	9.964	63.484	133.484
2057	37611	1051.737	119.619	189.619	9.962	63.481	133.481
2058	37976	1061.698	119.733	189.733	9.959	63.478	133.478
2059	38341	1071.658	119.845	189.845	9.957	63.476	133.476
2060	38706	1081.615	119.957	189.957	9.955	63.473	133.473
2061	39071	1091.570	120.067	190.067	9.953	63.470	133.470
2062	39436	1101.522	120.176	190.176	9.950	63.468	133.468
2063	39801	1111.473	120.285	190.285	9.948	63.465	133.465
2064	40166	1121.421	120.392	190.392	9.946	63.462	133.462
2065	40531	1131.367	120.499	190.499	9.944	63.460	133.460
2066	40896	1141.311	120.604	190.604	9.942	63.457	133.457
2067	41261	1151.252	120.708	190.708	9.940	63.455	133.455
2068	41626	1161.192	120.812	190.812	9.938	63.452	133.452
2069	41991	1171.130	120.915	190.915	9.936	63.450	133.450
2070	42356	1181.065	121.016	191.016	9.933	63.447	133.447
2071	42721	1190.999	121.117	191.117	9.931	63.445	133.445
2072	43086	1200.930	121.217	191.217	9.929	63.442	133.442
2073	43451	1210.860	121.317	191.317	9.927	63.440	133.440
2074	43816	1220.787	121.415	191.415	9.925	63.437	133.437
2075	44181	1230.712	121.513	191.513	9.924	63.435	133.435
2076	44546	1240.636	121.609	191.609	9.922	63.433	133.433
2077	44911	1250.558	121.705	191.705	9.920	63.430	133.430
2078	45276	1260.477	121.800	191.800	9.918	63.428	133.428
2079	45641	1270.395	121.895	191.895	9.916	63.426	133.426
2080	46006	1280.311	121.989	191.989	9.914	63.423	133.423
2081	46371	1290.225	122.082	192.082	9.912	63.421	133.421
2082	46736	1300.137	122.174	192.174	9.910	63.419	133.419
2083	47101	1310.047	122.265	192.265	9.908	63.417	133.417
2084	47466	1319.955	122.356	192.356	9.907	63.414	133.414

Table B.2 continued from previous page

Year	Days t	$\Lambda(t)$	$x_{0.95}$	$y_{0.95}$	Annual $\Lambda(t)$	Annual $x_{0.95}$	Annual $y_{0.95}$
2085	47831	1329.862	122.446	192.446	9.905	63.412	133.412
2086	48196	1339.767	122.535	192.535	9.903	63.410	133.410
2087	48561	1349.670	122.624	192.624	9.901	63.408	133.408
2088	48926	1359.571	122.712	192.712	9.899	63.406	133.406
2089	49291	1369.470	122.800	192.800	9.898	63.404	133.404
2090	49656	1379.368	122.886	192.886	9.896	63.401	133.401
2091	50021	1389.264	122.973	192.973	9.894	63.399	133.399
2092	50386	1399.158	123.058	193.058	9.893	63.397	133.397
2093	50751	1409.051	123.143	193.143	9.891	63.395	133.395
2094	51116	1418.941	123.227	193.227	9.889	63.393	133.393
2095	51481	1428.831	123.311	193.311	9.887	63.391	133.391
2096	51846	1438.718	123.394	193.394	9.886	63.389	133.389
2097	52211	1448.604	123.477	193.477	9.884	63.387	133.387
2098	52576	1458.488	123.558	193.558	9.882	63.385	133.385
2099	52941	1468.370	123.640	193.640	9.881	63.383	133.383
2100	53306	1478.251	123.721	193.721	9.879	63.381	133.381

B.3 95th Percentile using HPP-POT: Trenton

Table B.3: 95th Percentile using HPP-POT: Trenton

Year	Days t	$\Lambda(t)$	x_p	y_p	Annual $\Lambda(t)$	Annual x_p	Annual y_p
2020	24106	516.646	111.055	181.055	7.823	60.569	130.570
2021	24471	524.468	111.236	181.236	7.823	60.569	130.570
2022	24836	532.291	111.414	181.414	7.823	60.569	130.570
2023	25201	540.114	111.590	181.590	7.823	60.569	130.570
2024	25566	547.937	111.763	181.763	7.823	60.569	130.570
2025	25931	555.759	111.934	181.934	7.823	60.569	130.570
2026	26296	563.582	112.103	182.103	7.823	60.569	130.570
2027	26661	571.405	112.269	182.269	7.823	60.569	130.570
2028	27026	579.228	112.432	182.432	7.823	60.569	130.570
2029	27391	587.050	112.594	182.594	7.823	60.569	130.570
2030	27756	594.873	112.754	182.754	7.823	60.569	130.570
2031	28121	602.696	112.911	182.911	7.823	60.569	130.570

Table B.3 continued from previous page

Year	**Days** t	$\Lambda(t)$	x_p	y_p	**Annual** $\Lambda(t)$	**Annual** x_p	**Annual** y_p
2032	28486	610.519	113.066	183.066	7.823	60.569	130.570
2033	28851	618.342	113.220	183.220	7.823	60.569	130.570
2034	29216	626.164	113.371	183.371	7.823	60.569	130.570
2035	29581	633.987	113.521	183.521	7.823	60.569	130.570
2036	29946	641.810	113.669	183.669	7.823	60.569	130.570
2037	30311	649.633	113.814	183.814	7.823	60.569	130.570
2038	30676	657.455	113.959	183.959	7.823	60.569	130.570
2039	31041	665.278	114.101	184.101	7.823	60.569	130.570
2040	31406	673.101	114.242	184.242	7.823	60.569	130.570
2041	31771	680.924	114.381	184.381	7.823	60.569	130.570
2042	32136	688.746	114.519	184.519	7.823	60.569	130.570
2043	32501	696.569	114.655	184.655	7.823	60.569	130.570
2044	32866	704.392	114.790	184.790	7.823	60.569	130.570
2045	33231	712.215	114.923	184.923	7.823	60.569	130.570
2046	33596	720.037	115.054	185.054	7.823	60.569	130.570
2047	33961	727.860	115.184	185.184	7.823	60.569	130.570
2048	34326	735.683	115.313	185.313	7.823	60.569	130.570
2049	34691	743.506	115.441	185.441	7.823	60.569	130.570
2050	35056	751.329	115.567	185.567	7.823	60.569	130.570
2051	35421	759.151	115.692	185.692	7.823	60.569	130.570
2052	35786	766.974	115.815	185.815	7.823	60.569	130.570
2053	36151	774.797	115.937	185.937	7.823	60.569	130.570
2054	36516	782.620	116.058	186.058	7.823	60.569	130.570
2055	36881	790.442	116.178	186.178	7.823	60.569	130.570
2056	37246	798.265	116.297	186.297	7.823	60.569	130.570
2057	37611	806.088	116.414	186.414	7.823	60.569	130.570
2058	37976	813.911	116.531	186.531	7.823	60.569	130.570
2059	38341	821.733	116.646	186.646	7.823	60.569	130.570
2060	38706	829.556	116.760	186.760	7.823	60.569	130.570
2061	39071	837.379	116.873	186.873	7.823	60.569	130.570
2062	39436	845.202	116.985	186.985	7.823	60.569	130.570
2063	39801	853.025	117.096	187.096	7.823	60.569	130.570
2064	40166	860.847	117.206	187.206	7.823	60.569	130.570
2065	40531	868.670	117.315	187.315	7.823	60.569	130.570

Table B.3 continued from previous page

Year	**Days** t	$\Lambda(t)$	x_p	y_p	**Annual** $\Lambda(t)$	**Annual** x_p	**Annual** y_p
2066	40896	876.493	117.423	187.423	7.823	60.569	130.570
2067	41261	884.316	117.530	187.530	7.823	60.569	130.570
2068	41626	892.138	117.636	187.636	7.823	60.569	130.570
2069	41991	899.961	117.742	187.742	7.823	60.569	130.570
2070	42356	907.784	117.846	187.846	7.823	60.569	130.570
2071	42721	915.607	117.949	187.949	7.823	60.569	130.570
2072	43086	923.429	118.052	188.052	7.823	60.569	130.570
2073	43451	931.252	118.153	188.153	7.823	60.569	130.570
2074	43816	939.075	118.254	188.254	7.823	60.569	130.570
2075	44181	946.898	118.354	188.354	7.823	60.569	130.570
2076	44546	954.720	118.453	188.453	7.823	60.569	130.570
2077	44911	962.543	118.551	188.551	7.823	60.569	130.570
2078	45276	970.366	118.649	188.649	7.823	60.569	130.570
2079	45641	978.189	118.746	188.746	7.823	60.569	130.570
2080	46006	986.012	118.842	188.842	7.823	60.569	130.570
2081	46371	993.834	118.937	188.937	7.823	60.569	130.570
2082	46736	1001.657	119.031	189.031	7.823	60.569	130.570
2083	47101	1009.480	119.125	189.125	7.823	60.569	130.570
2084	47466	1017.303	119.218	189.218	7.823	60.569	130.570
2085	47831	1025.125	119.310	189.310	7.823	60.569	130.570
2086	48196	1032.948	119.402	189.402	7.823	60.569	130.570
2087	48561	1040.771	119.493	189.493	7.823	60.569	130.570
2088	48926	1048.594	119.583	189.583	7.823	60.569	130.570
2089	49291	1056.416	119.673	189.673	7.823	60.569	130.570
2090	49656	1064.239	119.762	189.762	7.823	60.569	130.570
2091	50021	1072.062	119.850	189.850	7.823	60.569	130.570
2092	50386	1079.885	119.937	189.937	7.823	60.569	130.570
2093	50751	1087.708	120.024	190.024	7.823	60.569	130.570
2094	51116	1095.530	120.111	190.111	7.823	60.569	130.570
2095	51481	1103.353	120.196	190.196	7.823	60.569	130.570
2096	51846	1111.176	120.282	190.282	7.823	60.569	130.570
2097	52211	1118.999	120.366	190.366	7.823	60.569	130.570
2098	52576	1126.821	120.450	190.450	7.823	60.569	130.570
2099	52941	1134.644	120.533	190.533	7.823	60.569	130.570

Table B.3 continued from previous page

Year	**Days** t	$\Lambda(t)$	x_p	y_p	**Annual** $\Lambda(t)$	**Annual** x_p	**Annual** y_p
2100	53306	1142.467	120.616	190.616	7.823	60.569	130.570