Spatio-temporal analysis of extreme wind velocities for infrastructure desing

Dissertation submitted in partial fulfillment of the requirements

for the Degree of Master of Science in Geospatial Technologies

Jan 2020

Alexys Herleym Rodríguez Avellaneda

alexyshr@gmail.com

https://github.com/alexyshr

Supervised by:

Prof. Dr. Edzer Pebesma

Institute for Geoinformatics

University of Münster - Germany

Co-supervised by:

Prof. Dr. Juan C. Reyes

Department of Civil and Environmental Engineering

Universidad de los Andes - Colombia

Co-supervised by:

Prof. Dr. Sara Ribero

Information Management School

Universidade Nova de Lisboa - Portugal

---

ifgi
Institut für Geoinformatik
Universität Münster

NOVA IMS Information Management School

UNIVERSITAT JAUME I

Declaration of Academic Integrity

I hereby confirm that this thesis on *Spatio-temporal analysis of extreme wind velocities for infrastructure desing* is solely my own work and that I have used no sources or aids other than the ones stated.

All passages in my thesis for which other sources, including electronic media, have been used, be it direct quotes or content references, have been acknowledged as such and the sources cited.

January 11, 2020

_____

I agree to have my thesis checked in order to rule out potential similarities with other works and to have my thesis stored in a database for this purpose.

January 11, 2020

_____

Acknowledgements

I would like to thank Prof. Dr. Edzer Pebesma, Prof. Dr. Juan C. Reyes and Prof. Dr. Sara Ribero for supervising my work and spending their valuable time for discussions and feedback. It was really a huge advantage for me to have this allways available support. It was a pleasure to work beside you. I want to thank Dr Adam Pintar and Engineer Juan David Sandoval for their support and contributions. My mother Ligia made possible all my achievements because she was allways there with love, support and valuable advice and contributions. I am grateful with all my heart. Thanks to my daughter Nicolle Chaely for its love, support, and always pleasant company. I would like to thank the European Union -'Erasmus Mundus Grant', because their funding allow me to fulfill this dream to go further with my accademic and professionals dreams.

I especially want to thank to Dr. Joaquín Huerta Guijarro because he allways was available to help and he was very friendly and receptive. Likewise, I want to thank some family members as Elsa Manrique, Barbara Avellaneda, and Kevin Martinez because they were allways interested in my activities and also they were an important source of motivation and support.

To all the beatuful people that shared with me different activities at the San Antonious church of Muenster, with special mention of father Alejandro Serrano Palacios and choir friends.

Preface

Models of extreme values are used for designing against the effects of extreme events like earthquakes, winds, rainfall, floods of different types of physical processes, avoiding widespread destruction and loss of lifes. This research presents a applied case of univariate extreme value analysis applied to wind velocities for infrastructure desing, consequently, the main interest are probabile future extreme wind events that structures need to be able to resist. This work in its teorethical and methodological component was directed by ASCE7-16 Engineers (2017) considering that output products will be used to update the chapter B.6, wind forces, of the Colombian earthquake resistant standard - NSR-10, maintaided by the Colombian Association of Seismic Engineering - AIS by its spahish acronym. ASCE7-16, defines four risk categories, which implies the use of different wind loads (represented in wind extreme values for different mean recurrence intervals) for structures that belong to each category, 3000 years of MRI for risk IV, 1700 years for risk III, and 700 years for risk II and I. This research has a particularly new situation regarding to the input data, and it is that not only time series of field measurements from meteorological stations are used (IDEAM data source), but also post-procesed information comming from the Integrated Surface Database - ISD (USA database based on IDEAM data source), and forecast reanalysis data from ERA5. This condition demanded a comparison of the different data sources in order to verify the feasibility of using ERA5 and ISD, with a previous activitie of sdandarization of wind velocities to reach the needed requirement of 3-s wind gust speed, 10 meters anemometers high and open space condition. At each station the used method Peaks Over Threshold - Poisson Process, required to identfy all the non-thunderstorm events in the non-hurricane dataset through a process of declustering, choose a suitable threshold level to leave for the analysis only the most extreme availables values, and then fit to the data a Gumbel extreme value distributon using maximun likelihood to find optimal parameters with the best goodness of fit. With the fitted model, it was possible to calculate return levels for required mean return intervals. Next, a process of spatial interpolation was done using Kriging, what allowed to have three continuous maps for the whole study area. Main interest writing this document, is help to readers to enter speedelly with the current details around wind extreme analysis.

Table of Contents

List of Tables

List of Figures

Abstract

For the input non-hurricane, non tornadic data in each available station of the study area (field measurement of forecast models), this research calculate extreme winds or return levels with three different mean recurrence intervals - MRI, 700, 1700, and 3000 years, with a change of being equaled or exceeded only one time in the corresponding MRI period. Then, continuous maps of wind extreme velocities are interpolated to cover the study area, which are mixed with existing wind extreme hurricane studies to be used as input loads for infrastructure desing.

Spatio-temporal analysis of historical wind data for infrastructrure design, namely, − from wind time series represented in forecast models over rectangular areas or pixels with a virtual station at its center, or field measurements at weather stations in specific coordinates around the study area −, calculate wind extreme magnitudes to be used as desing loads of structures of different risk categories (bridges, houses, buildings, hospitals, etc), requires the use of statistical extreme value analysis methodologies to create maps with different mean recurrence intervals (MRI), − short ones for less risky/important structures, and long ones for highly important structures.

Method used to calculate the return levels at each station the Peaks Overt Threshold - POT, using a non-homogeneous, bidimensional Poisson Process described, recomended by Engineers (2017), and developed and implemented in Pintar, Simiu, Lombardo, & Levitan (2015). To interpolate maps a geoestatistical procedure using Kriging was implemented, considering the model with the best goodness of fit from model parameters comparison.

List of Acronyms

I put all my efforts in Jesus' hands. I dedicate this work to my mother Ligia, my beautiful daughter Nicolle Chaely and my beautiful daughter Lucesita

ACF	Autocorrelation Function
AIC	Aikake's Information Criterion
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
DBAFS	Dockless Bike Availability Forecasting System
GPS	Global Positioning System
ID	Identification
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
MA	Moving Average
MAE	Mean Absolute Error
MLE	Maximum Likelihood Estimation
NFS	Naïve Forecasting System
PACF	Partial Autocorrelation Function
PDT	Pacific Daylight Saving Time
PST	Pacific Standard Time
PBSS	Public Bike Sharing Systems
RMSE	Root Mean Squared Error
RMSLE	Root Mean Squared Logarithmic Error
SFMTA	San Francisco Municipal Transportation Agency
SQL	Structured Query Language
STL	Seasonal Trend decomposition procedure based on Loess
WGS84	World Geodetic System 1984

Introduction

This research aims to create non-hurricane non-tornadic maps of extreme wind speeds for *three specific recurrence intervals* (700, 1700, and 3000 years) covering the Colombian territory. These maps will be combined with existing hurricane wind speed studies, to be used as input loads due to wind for infrastructure desing.

For each station with wind speeds time histories in the input data, extreme wind speed corresponding to each recurrence interval are calculated using a *Peaks Over Threshold* onwards *POT* extreme value model, then wind velocities with the same recurrence interval are *spatially interpolated* to generate continuous maps for the whole study area.

A wind speed linked to a *mean recurrence interval - MRI* of *N-years* (N-years return value or return period) is interpreted as the highest probable wind speed along the period of N-years. The annual probability of equal or exceed that wind speed is 1/N. The annual excedance probability for all velocity values in 700-years output map will be 1/700, for the 1700-years map will be 1/1700, and 1/3000 for the 3000-years final map.

There are different methods to model extreme value data, among them are a) sample maxima using a *Generalized Extreme Value Distribution* onwards *GEVD* (traditional method), b) POT using a *Generalized Pareto Distribution* onwards *GPD*, c) POT using a two-dimensional Poisson Process, that can be homomegenos, non-homogeneous, stationary, and non-stationary (originally know as *Point Process* approach), and d) POT Poisson-GPD. Following Pintar et al. (2015) in this research a *POT using a non-homogeneous non-stationary two-dimensional poisson proces* was selected, despide there is no R package available to apply this approach.

0.1 Background

To desing one structure, the horizontal forces wind and earthquake play an starring role. For the study area, Colombia, initially the wind force was considered with the decree 1984 as a fixed velocity $100\frac{Km}{h}$, later a continuos map with a return period of 50 years was included in the official design standard of the time (NSR-98), then, with the update to NSR-10, an additional map with return period of 700 yeas was included.

Extreme wind analysis is concerned with statistical methods applied to very high values of wind as random variable in a stochastic process, to allow statistical inference from historical data. Classical reference in this matter is Coles (2001), where a detailed study is done

about classical extreme value theory and models and threshold models. There are four main approaches to deal with extreme value analysis: - sample maxima associated to a Generalized Extreme Value Distribution - GEV, - excedances over threshold associated to a Generalized Paretto Distribution - GPD, - the Poisson-GPD, an homogeneous Poisson process fot the number of excedancess and a GPD fot the excess values, and the exceedances over threshold associated to a non- homomeneous bi-dimensional Poisson process, a Point process approach also known as Peaks Over Threshold - POT - Poisson process. Main details will be disscused here for each method, but as the last one is recommended in Asce2017, a more indeep explanation will be provided in for POT-Poisson Process.

0.1.1 Sample maxima

To work with random variables of sample maximun values, the used probability distribution function *pdf* is the GEV

$$H(y) = \exp\left\{-\left(1+\xi\frac{y-\mu}{\psi}\right)_+^{-\frac{1}{\xi}}\right\},$$

$(y+ = max(y,0))$ where μ is the location parameter, $\psi > 0$ is a scale parameter, and ξ is a shape parameter. GEV can be seen as the integration in the same *psf* of the Gumbel distribution (limit $\xi \to 0$), Fréchet distribution ($\xi > 0$), and Weibull distribution ($\xi < 0$).

0.1.2 Exceedances over threshold

If the researcher needs to work only with extreme values above an specific threshold, Pickands (1971) showed that the GEV has a GPD approximation where shape ξ parameter in previous equation is the same parameter for next equation for GPD,

$$G(y,\sigma,\xi) = 1-\left(1+\xi\frac{y}{\sigma}\right)_+^{-\frac{1}{\xi}},$$

Poisson-GPD If a rescale of the variable indices above the threshold is performed, then the exceedances over threshold approach can be seen as a point process, namely, an homogeneous Poisson Process where:

1. The number of exceedances above the threshold has a Poisson distribution with mean λ
2. The excees values follow a GPD with $N \leq 1$

Its cumulative distribution function *cdf* is

$$F(y) = \exp\left\{-\lambda\left(1+\xi\frac{y-\mu}{\sigma}\right)_+^{-\frac{1}{\xi}}\right\},$$

0.2 Research Aim and Objectives

0.3 Reseach Question

0.4 Thesis Document Structure

Chapter 1

Data

Input data is made up of three different sources a) IDEAM - Institute of Hydrology, Meteorology and Environmental Studies of Colombia `http://www.ideam.gov.co`, b) ISD - Integrated Surface Database `https://www.ncdc.noaa.gov/isd`, and c) ERA5 climate reanalysis `https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5`.

Table 1.1: Datasets description

Institution	Dataset	Details
IDEAM	Historical records at weather stations	IDEAM is responsible for the instalation, maintenance and management of all kind of weather stations located everywhere along the country
NOAA	ISD	ISD (Integrated Surface Database. NOAA's National Centers for Environmental Information - NCEI) Lite: A subset from the full ISD dataset containing eight common surface parameters in a fixed-width format free of duplicate values, sub-hourly data, and complicated flags.
ECMWF	ERA5	ERA5 is a reanalysis dataset with hourly estimates of atmospheric variables with horizontal resolution of 0.25^{o} (33 kilómeters), this is equally spaced cells every 0.25 degrees

Table 1.2: Datasets variables

Dataset	Variables	Description
IDEAM	vvmx_aut_60	Hourly wind maximun velocity
ISD	wind speed rate	Maximun hourly wind velocity. The rate of horizontal travel of air past a fixed point.
ERA5	fg10	10 metre wind gust since previous post-processing
	fsr	Forecast Surface Roughness

Table 1.3: Variables units and time

Variable	Units	Time	Stations
vvmx_aut_60	meters per second	Variable from 2001 until today. Irregular time series.	203
Wind speed	meters per second	Variable from 1941 until today. Note: There is too much variability in time (start, end, and time range) for each station. Irregutal time series.	101
fg10	meters per second	1979-Today	3381
fsr	meters per second	1979-Today	3381

Ideal data source to create extreme wind speeds maps should be field observed data from IDEAM, but there are not enough number of stations around the study area to represent all the local wind variability in a huge country with multiple variety of climates and and changing thermal floors, but there are other important motivatios to include different sources trying to improve output results:

- As just mentioned, low quantity of IDEAM stations
- There are uncertanties related to the way IDEAM anemometers are registering data, then comparison with other datasources are needed to be able to do appropriate data standardization, needed as a prerequisite to the analysis.
- There is no time continuity in the registration of IDEAM data. Historical time series are different and variable in each station.

Importance of ISD database for this study is based on the fact that post-procesed ISD database has wind extreme values, and it was used to create extreme wind maps for United States. ISD allows comparison with IDEAM records to take better decitions in order to do needed data standarization.

Despite that ERA5 data are not observed data, but forecast, its main advantage is data availability to assess the local climatic variance every 33 square kilometers.

1.1 IDEAM

Historical observed wind speeds from 203 in Colombia are managed by the official environmental authority IDEAM. Table 1.4 shows a sample of five IDEAM stations. Figure 1.1 shows a map of IDEAM stations.

Table 1.4: IDEAM Stations sample

Name[Code]	Latitud	Longitud
EMAS - AUT [26155230]	5.09	-75.51
SAN BENITO - AUT [25025380]	9.16	-75.04
AEROPUERTO ALFONSO LOPEZ - [28025502]	10.44	-73.25
TIBAITATA - AUT [21206990]	4.69	-74.21
ELDORADO CATAM - AUT [21205791]	4.71	-74.15

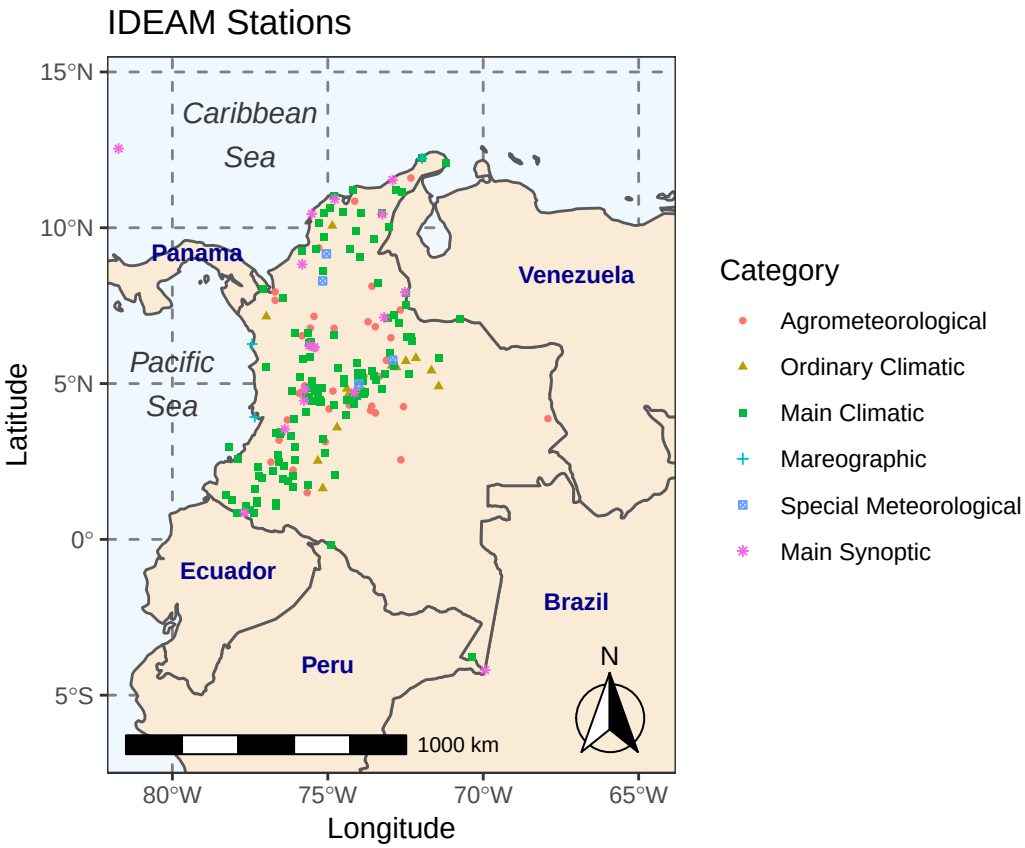

Figure 1.1: IDEAM Stations

Following, the time serie, autocorrelation function, and partial autocorrelation function, for IDEAM station “21205791” will be displayed.

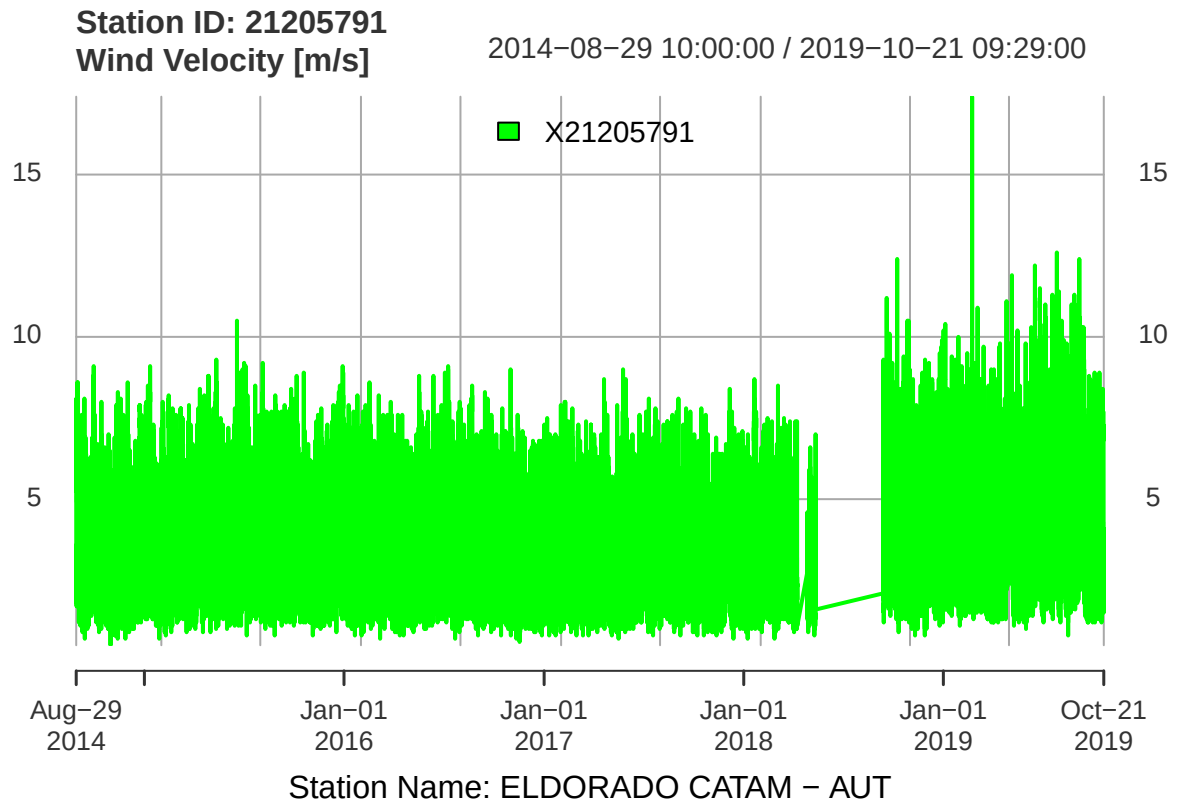

Figure 1.2: IDEAM Station - Time Serie

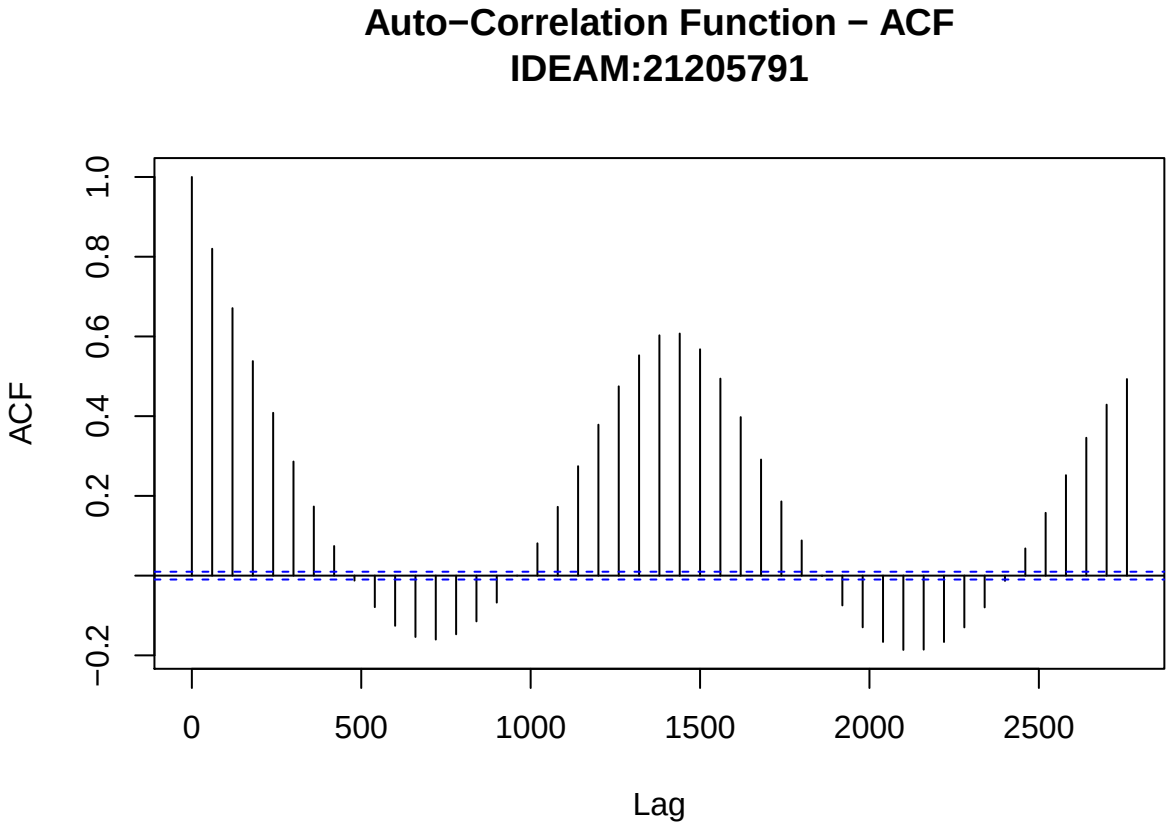

Figure 1.3: IDEAM Station ACF

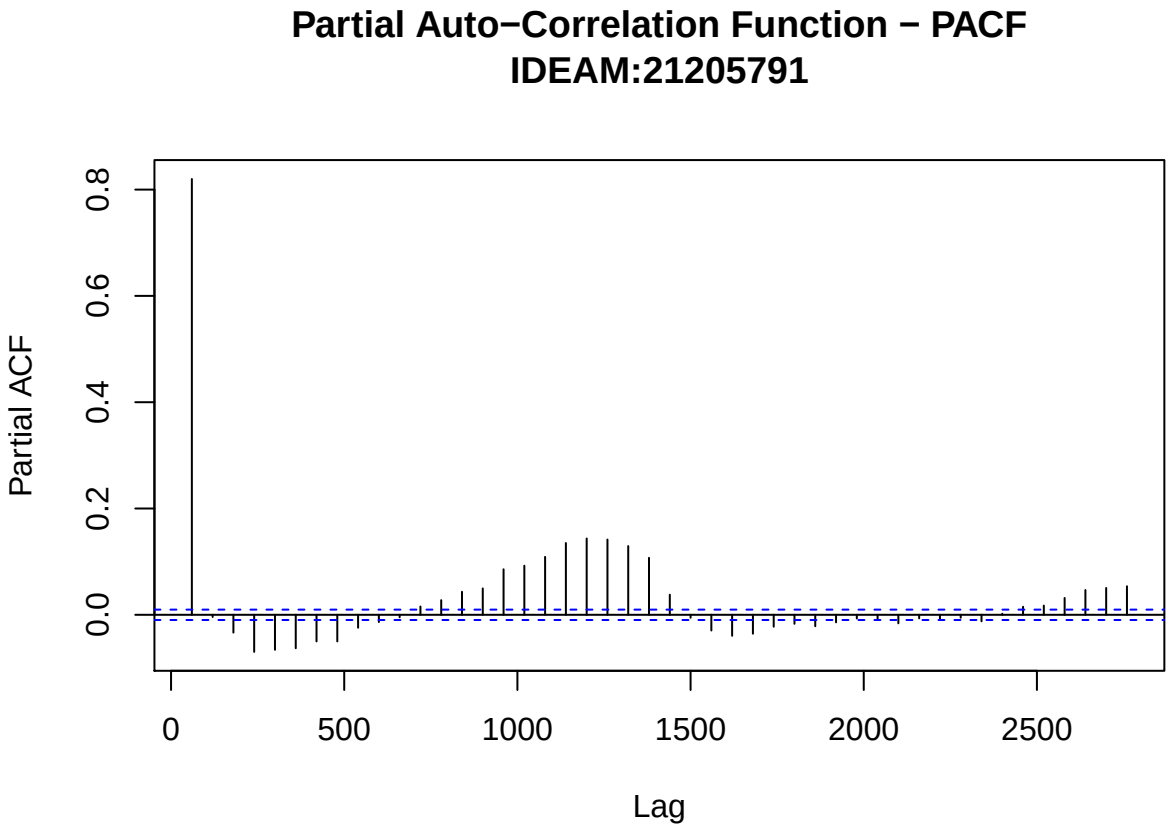

Figure 1.4: IDEAM Station PACF

1.2 ISD

ISD is a database with environmental variables among then extreme wind speeds. ISD has data for the whole planet, and is based on observed data at metereological stations in each country, which means that for Colombia is based on IDEAM data. Main advantage is data availability at neighbor countries and specialized postprocesing made by NOAA's National Centers for Environmental Information - NCEI in United States, which facilitates its use.Table 1.5 shows a sample of five ISD stations. Figure 1.5 shows a map of ISD stations.

Table 1.5: ISD Stations sample

Code	Name	Latitud	Longitud
804400	BARINAS	8.62	-70.22
800810	ALTO CURICHE	7.05	-76.35
801000	BAHIA SOLANO / JOSE MUTIS	6.18	-77.40
802590	ALFONSO BONILLA ARAGON INTL	3.54	-76.38
803150	BENITO SALAS	2.95	-75.29

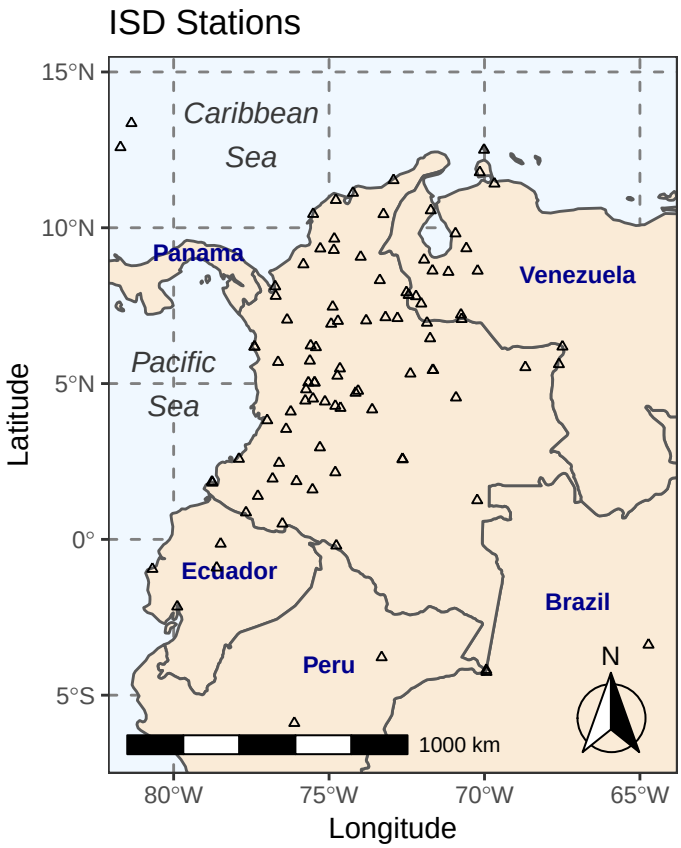

Figure 1.5: ISD Stations

Following, the time serie, autocorrelation function, and partial autocorrelation function, for ISD station “802590” will be displayed.

```
select "mydatetime", "802590" as "X802590" from isd_lite_unstack where "802590" IS NOT NUL
```

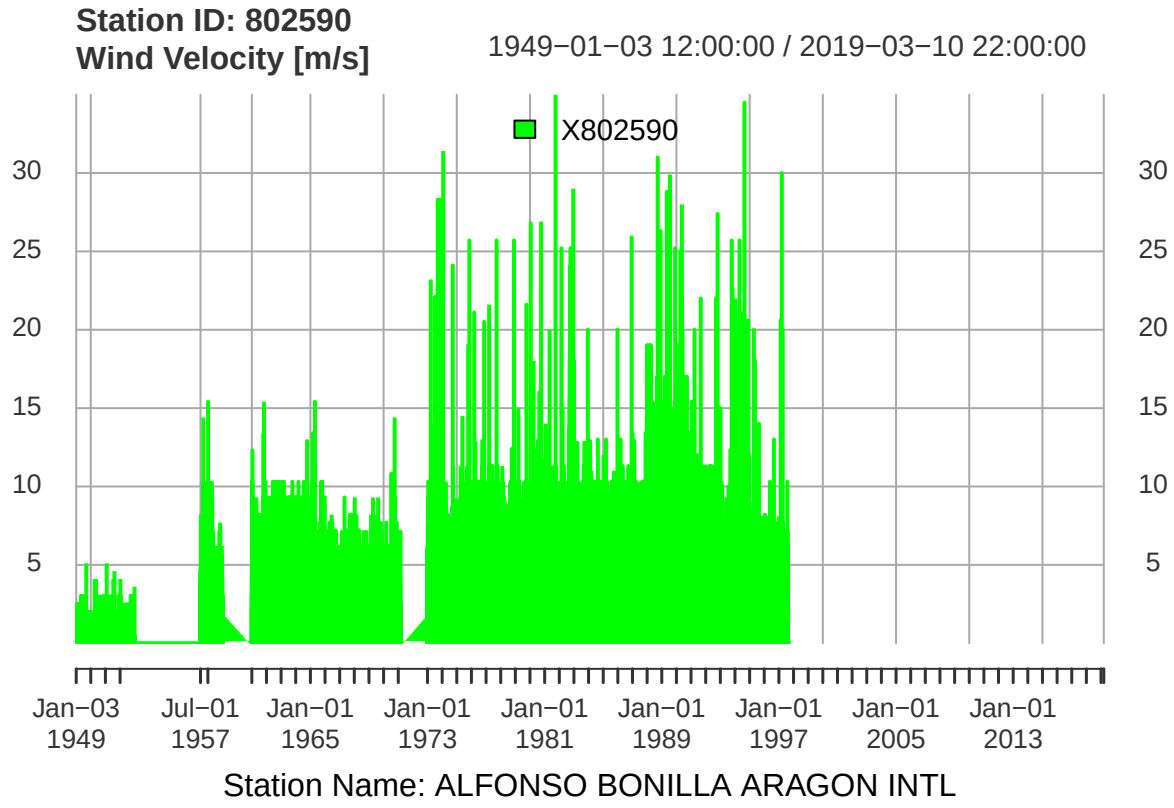

Figure 1.6: ISD Station - Time Serie

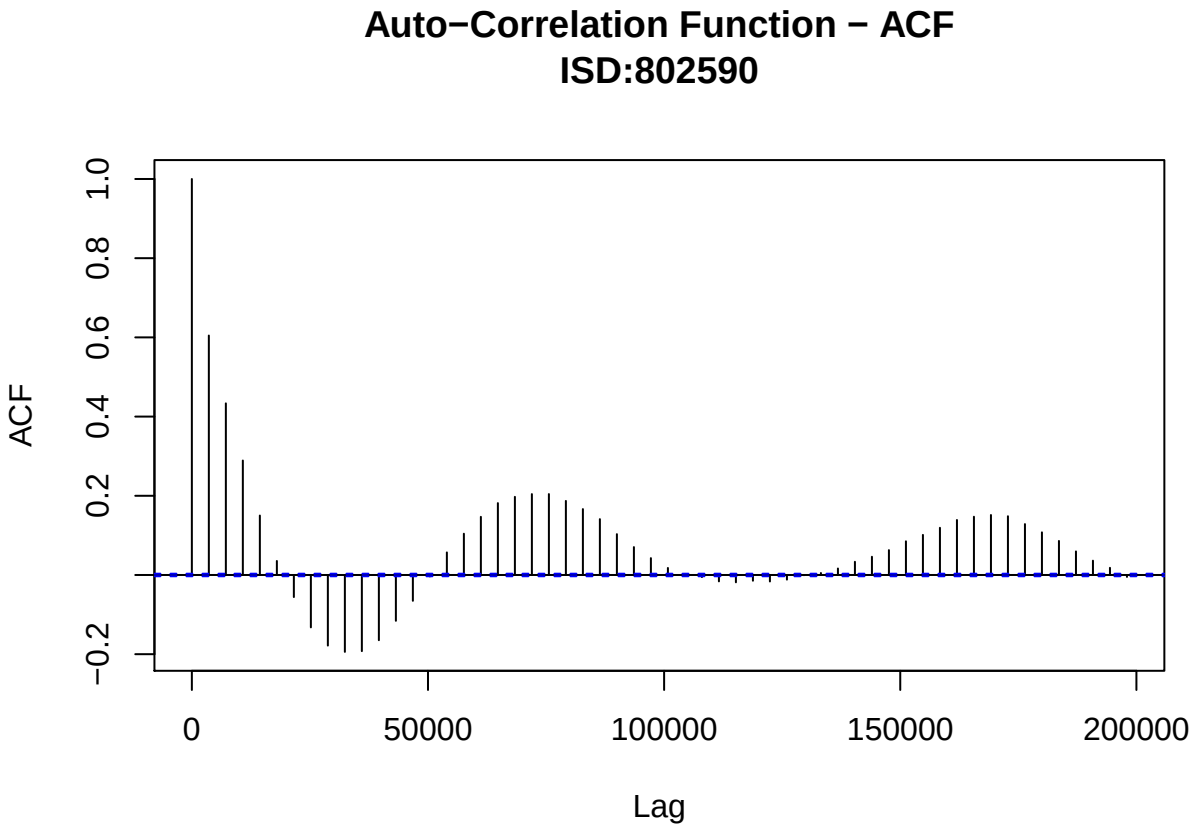

Figure 1.7: ISD Station ACF

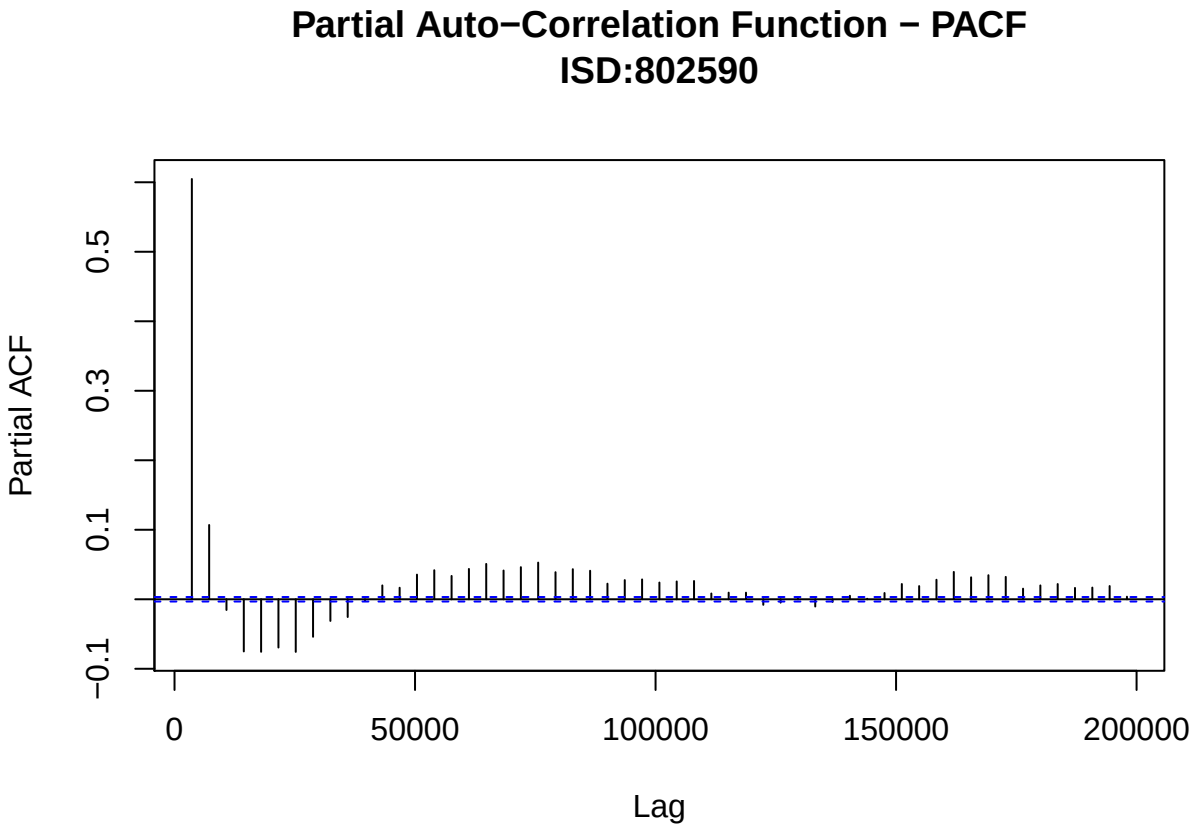

Figure 1.8: IDEAM Station PACF

1.3 ERA5

ERA5 is forecast reanalysis data procesed by the *European Centre for Medium-Range Weather Forecasts* - ECMWF with wind speeds time series in square cells *matrix of pixels* of 0.25 degrees (33 km) covering the whole plannet. For the study area was extracted a raster of 69 rows by 49 XXX columns in format NetCDF. Figure 1.9 shows a map of ERA5 stations (cells centers).

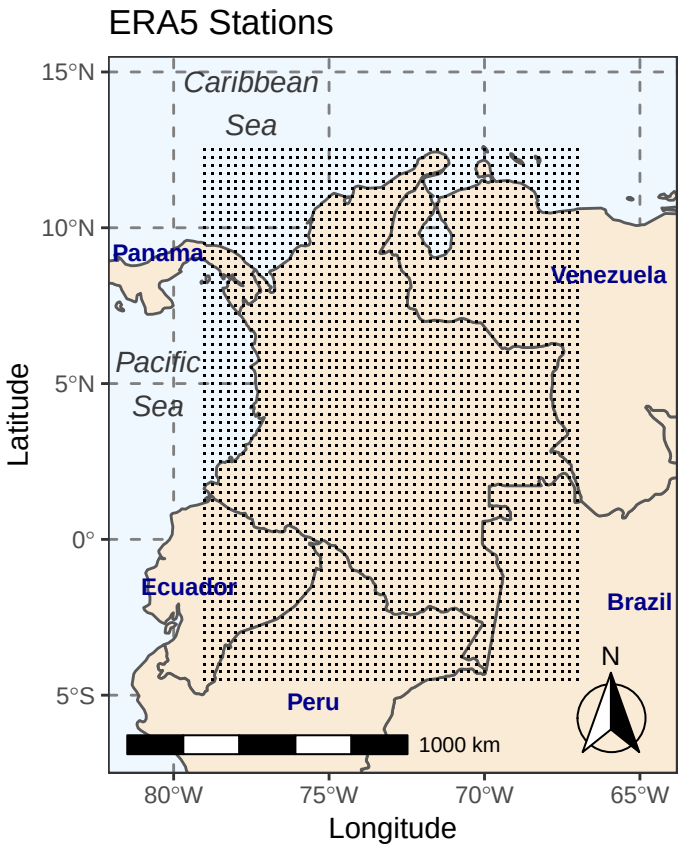

Figure 1.9: ERA5 Stations (cells centers)

1.4 Data Download and Organization

1.5 Data Standarzation

Analysis of extreme wind speeds requieres data standarizaton as initial step. All input data must be standarized to represent three important conditions: a) anemometer height of 10 meters, b) open space roughness, and c) averaging time of 3-seconds wind gust. Data for analysis must represent 3-s peak wind speeds 10 meters heigh above the surface, in open terrain. * 10 mts anemometer height * Open space terrain roughness * 3-s gust averaging time

Chapter 2

Theoretical Framework

2.1 Probability Concepts

Poisson process is an stochastic method that relies in the concepts of probability distributions. The main functions related to probability for extreme value analysis will be described below.

2.1.1 Probability Density Function - *pdf*

Pdf defines the probability that a continuos variable falls between two points, this is, in *pdf* the proability is related to the area below the curve (integral) between two points, as for continuos probability distributions the probability at a single point is zero. The term density is directly related to the probability of a portion of the curve, if the density function has high values the probability will be greater in comparison with the same portion of curve for low values.

$$\int_a^b f(x)dx = Pr[a \leq X \leq b]$$

Equation (2.1) is the Gumbel *pdf*.

$$\mathrm{f(x)} = \frac{1}{\beta} \exp\left\{-\frac{\mathrm{x}-\mu}{\beta}\right\} \exp\left\{-\exp\left\{-\left(\frac{\mathrm{x}-\mu}{\beta}\right)\right\}\right\}, \quad -\infty < \mathrm{x} < \infty \tag{2.1}$$

where $\exp\{.\} \mapsto \mathrm{e}^{\{.\}}$, β is the scale parameter, and μ is the location parameter. Location (μ) has the effect to shift the *pdf* to left or right along 'x' axis, thus, if location value is changed the effect is a movement of *pdf* to the left (small value for location), or to the right (big value for location). Scale has the effect to stretch ($\beta > 1$) of compress ($0 < \beta < 1$) the *pdf*, if scale parameter is close to zero the pdf approaches a spike.

Figure 2.1 shows *pdf* with location (μ) = 100 and scale (β) = 40, using equation (2.1).

```
location = 100
scale = 40
.x <- seq(0, 300, length.out=1000)
pdfG <- function(x) {
  1/location *exp(-(x-location)/scale)*exp(-exp(-(x-location)/scale))
  }
.y = pdfG(.x)
plot(.x, .y, col="green", lty=4,
     xlab="Velocities Km/h", ylab="Density Function - Gumbel Distribution",
     main=paste("Gumbel - Density Function Gumbel Distribution\n", "Location=",
     round(location,2), " Scale=", round(scale,2)), type="l",
     cex.axis = 0.5, cex.lab= 0.6, cex.main=0.7, cex.sub=0.6)
```

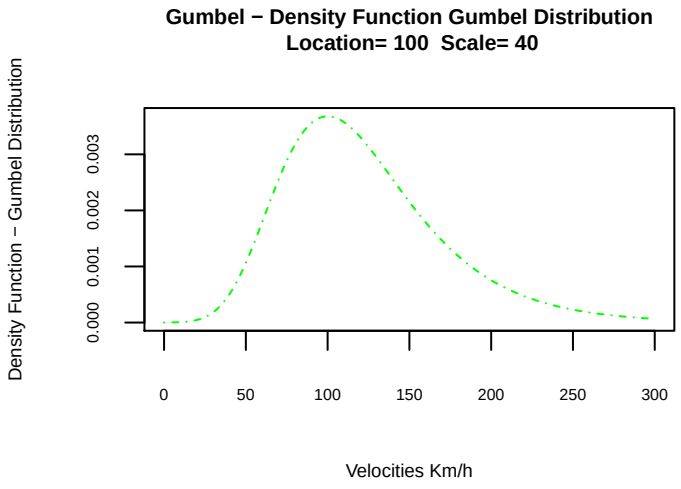

Figure 2.1: Gumbel pdf

Figure 2.2 shows *pdf* with location (μ) = 100 and scale (β) = 40, using function `dgumbel` of the package `RcmdrMisc`

```
location = 100
scale = 40
.x <- seq(0, 300, length.out=1000)
dfG = dgumbel(.x, location=location, scale=scale)
plot(.x, dfG, col="red", lty=4,
     xlab="Velocities Km/h", ylab="Density Function - Gumbel Distribution",
     main=paste("Gumbel - Density Function Gumbel Distribution\n", "Location=",
     round(location,2), " Scale=", round(scale,2)), type="l",
     cex.axis = 0.5, cex.lab= 0.6, cex.main=0.7, cex.sub=0.6)
```

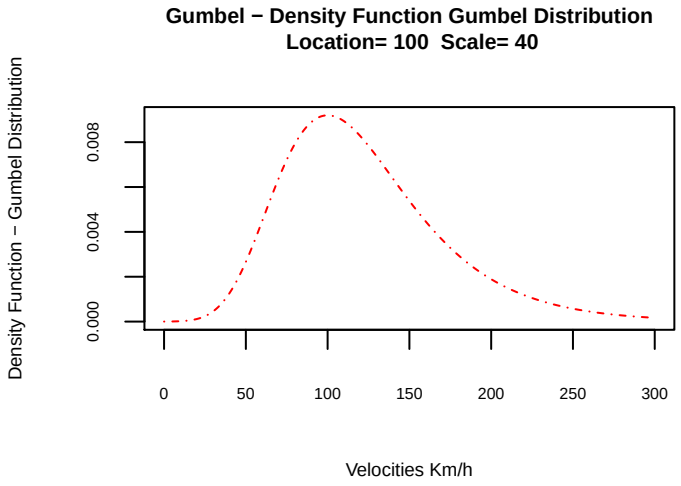

Figure 2.2: Gumbel pdf - dgumbel function

2.1.2 Cumulative Distribution Funtcion - *cdf*

Cdf is the probability of taking a value less than or equal to x. That is

$$F(x) = Pr[X < x] = \alpha$$

For a continuous variable, *cdf* can be expressed as the integral of its *pdf*.

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

Equation (2.2) is the Gumbel *cdf*.

$$\mathrm{F(x)} = \exp\left\{-\exp\left[-\left(\frac{\mathrm{x}-\mu}{\beta}\right)\right]\right\}, \quad -\infty < \mathrm{x} < \infty \tag{2.2}$$

Figure 2.3 shows Gumbel *cdf* with location (μ) = 100 and scale (β) = 40, using equation (2.2). As previously done with *pdf*, similar result can be achieved using function `pgumbel` of package `RcmdrMisc`.

```
location = 100
scale = 40
.x <- seq(0, 300, length.out=1000)
cdfG <- function(x) {
  exp(-exp(-(x-location)/scale))
  }
.y = cdfG(.x)
plot(.x, .y, col="green", lty=4,
     xlab="Velocities Km/h", ylab="Probability",
     main=paste("Gumbel - Cumulative Distribution Function\n", "Location=",
     round(location,2), " Scale=", round(scale,2)), type="l",
     cex.axis = 0.5, cex.lab= 0.6, cex.main=0.7, cex.sub=0.6)
```

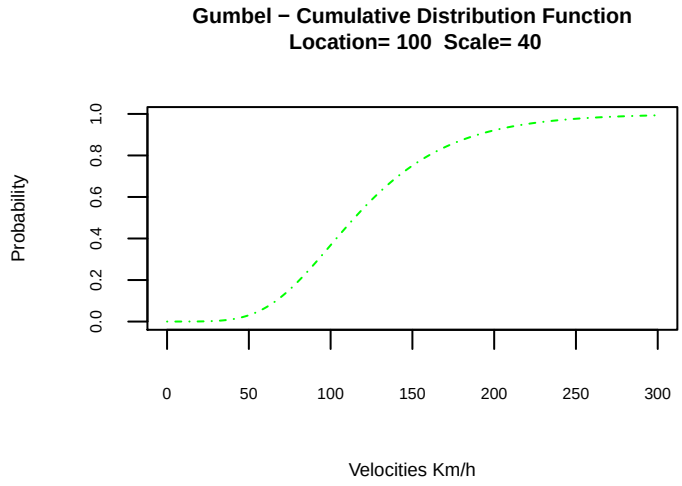

Figure 2.3: Gumbel cdf

2.1.3 Percent Point Function - *ppf*

Ppf is the inverse of *cdf*, also called the *quantile* function. This is, from a specific probability get the corresponding value x of the variable.

$$x = G(\alpha) = G(F(x))$$

Equation (2.3) is the Gumbel *ppf*.

$$\mathrm{G}(\alpha) = \mu - \beta \ln(-\ln(\alpha)) \quad 0 < \alpha < 1 \tag{2.3}$$

Figure 2.4 shows Gumbel *ppf*, using equation (2.3). Similar result can be achieved using function `qgumbel` of package `RcmdrMisc`.

```
location = 100
scale = 40
.x <- seq(0, 1, length.out=1000)
ppfG <- function(x) {
  location - (scale*log(-log(x)))
  }
.y = ppfG(.x)
plot(.x, .y, col="green", lty=4,
     ylab="Velocities Km/h", xlab="Probability",
     main=paste("Gumbel - Percent Point Function\n", "Location=",
     round(location,2), " Scale=", round(scale,2)), type="l",
     cex.axis = 0.5, cex.lab= 0.6, cex.main=0.7, cex.sub=0.6)
```

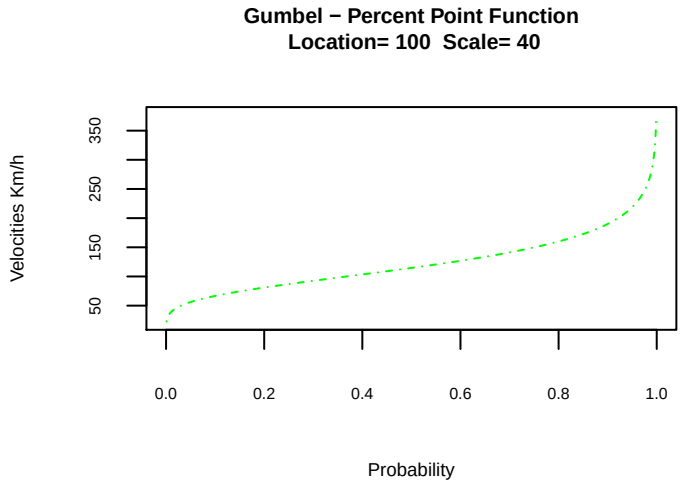

Figure 2.4: Gumbel cdf

2.1.4 Hazard Function - *hf*

Using $S(x) = 1 - F(x)$ as survival function -*sf*, the probability that a variable takes a value greather than x $S(x) = Pr[X > x] = 1 - F(x)$, the *hf* is the ratio between *pdf* and *sf*.

$$h(x) = \frac{f(x)}{S(x)} = \frac{f(x)}{1 - F(x)}$$

Equation (2.4) is the Gumbel *ppf*.

$$\mathrm{h(x)} = \frac{1}{\beta} \frac{\exp(-(\mathrm{x} - \mu)/\beta)}{\exp(\exp(-(\mathrm{x} - \mu)/\beta)) - 1} \tag{2.4}$$

Figure 2.5 shows Gumbel *hf*, using equation (2.4).

```
location = 100
scale = 40
.x <- seq(0, 3000, length.out=1000)
hfG <- function(x) {
  (1/scale)*(exp(-(x-location)/scale))/(exp(exp(-(x-location)/scale))-1)
  }
.y = hfG(.x)
plot(.x, .y, col="green", lty=4,
     xlab="Velocities Km/h", ylab="Hazard",
     main=paste("Gumbel - Hazard Function\n", "Location=",
     round(location,2), " Scale=", round(scale,2)), type="l",
     cex.axis = 0.5, cex.lab= 0.6, cex.main=0.7, cex.sub=0.6)
```

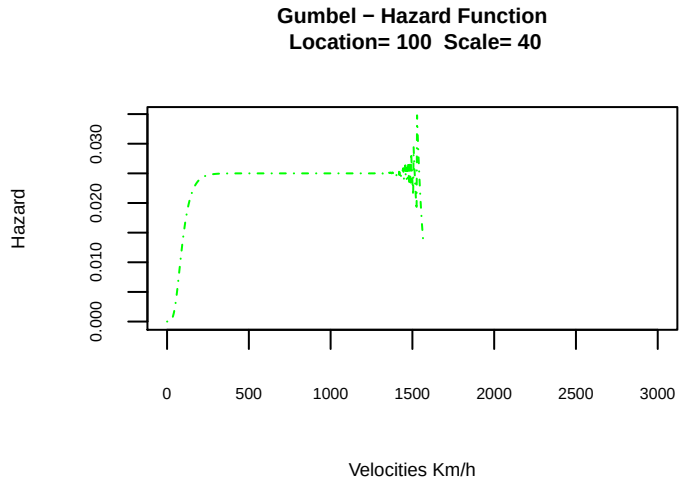

Figure 2.5: Gumbel cdf

```
#library(reliaR)
#plot(.x, hgumbel(.x, mu=location, sigma=scale))
#plot(.x, hra.gumbel(.x, mu=location, sigma=scale))
```

2.2 Statistical Concepts For Extreme Analysis

In order to approach the extreme value analysis, some statistical concepts are needed to understand the theoretical framework behind this knowledge area. In this section will be introduced the concepts annual excedance probability, mean recurrence interval - MRI, exposure time, and compound probability for any given exposure time and MRI.

As an hypotetical example, a simulated database of extreme wind speed will be used. This database is supposed to have 10.000 years of simulated wind speeds.

2.2.1 Annual Excedance Probability - Pe

Using the previously described database, a question arises to calculate the probability to exceed the highest probable loss due to the simulated winds. It is possible to conclude that there is only one event grather or equal (in this case equal) to the higest probable causing loss in 10.000 years, and it is the *highest wind*. If we sort the database by wind magnitude in descending order (small winds last), the question is solved calculating the annual excedance probability Pe with next formula

Figure 2.6: Sorted Winds by Magnitud - wind simulation database

$$P_e = \frac{Event\,index\,after\,descending\,sorting}{Years\,of\,simulations} = \frac{1}{10.000} = 0.001 = 0.01\%$$

because the highest wind will be the first in the sorted list. Same exercise can be done with all winds to construct the annual exedance probability curve, that in this case will represent the probability to equal or exceed different probable losses due to wind.

2.2.2 Return Period - Mean Recurrence Interval

Continuing with the previous section, if the inverse of the excedance probability is taken, the return period (in years) is obtained. The return period or Mean Recurrence Interval - MRI is associated with an specific return level (wind extreme velocity). MRI is the numbers of years (N) needed to obtain 63% of change that the corresponding return level will occur at least one time in that period. The return level is expected to be exceeded on average once every N-years. The annual excedance probability of the return level corresponding to N-years of MRI, is $P_e = \frac{1}{MRI} = \frac{1}{N}$.

For an specific wind extreme event A, the probability that the event will occur in a period equal to MRI years is 63%. If we analyse for the same period a strongest wind extreme event B, its orrurence probability will be lest than 67%. If the purpose of this researh is to design infrastructure considering wind loads, the structure will be more resistant to wind if we desing with stronger winds, this is high MRIs, and low annual excedance probability. Common approach for infrastructure design, considering any type of load (earthquake, wind, etc) is to choose high MRI according to the importance/use/risk/type of the structure. For highly important structures lik hospitals or coliseums, where the risk of collapse must be diminished, the MRI used to design is higher in comparison to common structures (for instance a normal house), which implies less risks for its use and importance.

$$\mathrm{P_e} = \begin{cases} 1 - \exp\left(-\frac{1}{MRI}\right), \; for\, MRI < 10\, years \\ \frac{1}{MRI}, \; for\, MRI \geq 10\, years \end{cases}$$

2.2.3 Compound Excedance Probability - Pn

If time of exposure is consider, understood as time the structure will be in use, it is possible to have a compound probability P_n, where n is the exposure period. P_n is the probability that the extreme wind speed will be equaled or exceeded at least one time in n years, and is related with the occurrence probability, but also is posible to calculate the non-ocurrence compound probability (probability that the event will not occur).

$$\mathrm{P_n} = \begin{cases} 1 - \left(1 - \frac{1}{MRI}\right)^n, \; occurrence\, probability \\ \left(1 - \frac{1}{MRI}\right)^n, \; non - occurrence\, probability \end{cases}$$

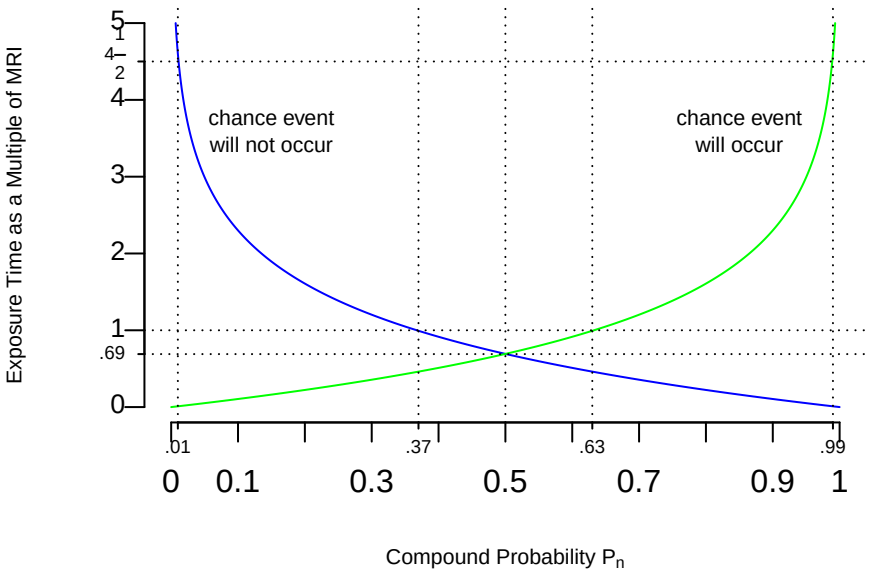

Figure 2.7: Compound Probability

If it is consider exposure time as a multiple of return period, the resulting figure 2.7, shows that:

- When exposure time is .69% of the return period, then probability (ocurrence and non-occurrence) will be 50%
- As was stated previously, whe exposure time is equal to return period, then the probability that the extreme wind speed (return level) occur is 63%, and 37% for the non occurrence probability.
- If exposure time is 4.5 times the return period, there is a 99% of change that the return level will occur.

The example discused here was presented as an instrument to introduce important concepts, nonetheless,there are specialized approaches to deal with extreme value analysis which will be discussed in Extreme Valua Analysis Overview and more in detail in Peaks Over Threshold - Poisson Process. In summary, is neccesary to fit the data over a specific threshold to an extreme value distribution, and P_e wil be $1 - F(y)$, with F(y) as the *cdf*, and MRI as $\frac{1}{1-F(y)}$.

2.3 Extreme Value Analysis Overview

Analysis of extreme values is related with statistical inference to calculate probabilities of extreme events. Main methods to analize extreme data are ephochal, Peaks Over Threshold - POT, and extreme index. The epochal method, also known as block maxima, uses the most extreme value for a specific frame of time, tipically, one year. POT is based in the selection of a single threshold value to do the analysis only with values above the threshold. But there are different POT aproaches, the most commond one is Generalized Paretto Distribution - POT-GPD, but also it is possible to use the Poisson process approach.

In both methods (Epochal and POT), the first step is to fit the data to an appropiate probability distribution model, among them the most used are, - Extreme Value Type I (Gumbel), Extreme Value Type II (Frechet), Weibull, Generalized Pareto - GPD, and Generalized Extreme Value - GEV.

Distribution models are fitted based in the estimation of its parameters, mommonly called location, scale and shape, nonetheless each model has its own parameters names. There are different methods to estimate parameters, among them, - method of moments (modified moments - see Kubler (1994), and L moments - see Hosking & Wallis (1997)), - method of maximum likelihood MLE, see Harris & Stocker (1998), which is problematic for GPD and GEV, - probability plot correlation coeficient, and - elemental percentiles (for GPD and GEV)

Once cadidate parameters are available, it is neccesary to assess the goodness of fit of the selected model, using one of the next methods, - Kolmogorov-Smirnov (KS) goodnes of fit test, and - Anderson-Darling goodness of fit test. Here a visual assesment is also useful using a probability plot or a kernel density plot with the fitted *pdf* overlaid.

The main use of the fitted model is the estimation of mean return intervals - MRI, and extreme wind speeds (return levels),

$$MRI = \frac{1}{1 - F(y)}$$

with $F(y)$ as the *cdf*. If $1 - F(y)$ is the annual excedance probability, MRI is its inverse, see Simiu & Scanlan (1996) for more details about MRI. If y is solved from previos equation using a given MRI of N-years, its value represents the Y_N wind speed return level,

$$Y_N = G\left(1 - \frac{1}{\lambda N}\right)$$

where G is the *ppf* (quantile function) and λ is the number of wind spees over the threshold per year.

The CRAN Task View "Extreme Value Analysis" `https://cran.r-project.org/web/views/ExtremeValue.html` shows available **R** for block maxima, POT by GPD, and external indexes estimation aproaches. Most important to consider are `evd`, `extremes`, `evir`, `POT`, `extremeStat`, `ismev`, and `Renext`.

2.4 Peaks Over Threshold - Poisson Process

According to Pintar et al. (2015) the stochastic poisson process is mainly defined by its intensity function. As the intensity function is nos uniform over the domain, the poisson process considered here is non-homogeneous, and due to the intensity function dependance of magnitud and time, it is also bi-dimmensional. Poisson Process was described for the first time in Pickands (1971), then extended in Smith (1989).

$$\lambda\,(\mathrm{y},\mathrm{t}) \begin{cases} \lambda_t(y), & for\, t\, in\, thunderstorm\, period \\ \lambda_n t(y), & for\, t\, in\, non - thunderstorm\, period \end{cases} \tag{2.5}$$

Generic equation (2.5) shows the intensity function, which is defined in the domain $D = D_t \cup D_{nt}$, and allow to fit the poisson process at each station to the observed data $\{t_i, y_i\}_{i=1}^{I}$ for al the times (t_i) of threshold crossing observations and its corresponding wind speeds magnitudes (y_i). Thus, only data above the threshold is used.

Intensity function of the Poisson Process is defined in Smith (2004),

$$\frac{1}{\psi_t}\left(1 + \zeta_t \frac{y - \omega_t}{\psi_t}\right)_+^{-\frac{1}{\zeta_t}-1}$$

Where ζ_t controls the tail lengh of the intensity function at a given time t, but to facilitate the estimation of the parameters then ζ_t is taken to be zero, then doing the limit, the resulting intensity function is the same as the the GEV type I or Gumbel distribution,

$$\frac{1}{\psi_t} \exp\left\{\frac{-(y - \omega_t)}{\psi_t}\right\}$$

In this study, the used intensity functions are shown in ecuation (2.6).

$$\lambda\,(\mathrm{y},\mathrm{t}) \begin{cases} \dfrac{1}{\psi_s} \exp\left\{\dfrac{-(y-\omega_s)}{\psi_s}\right\}, & for\, t\, in\, thunderstorm\, period \\ \dfrac{1}{\psi_{nt}} \exp\left\{\dfrac{-(y-\omega_{nt})}{\psi_{nt}}\right\}, & for\, t\, in\, non-thunderstorm\, period \end{cases} \tag{2.6}$$

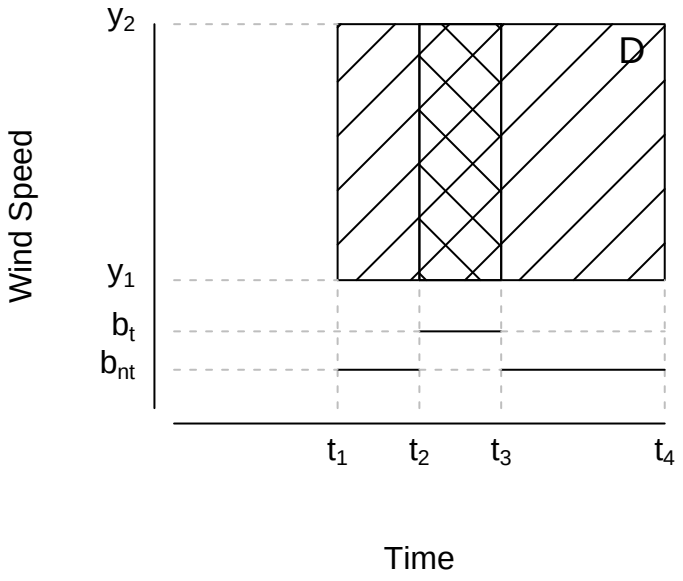

Figure 2.8: Domanin off the Poisson Process

Figure 2.8 represent the domain D of the Poisson process. In time, the domain represents the station service period from first sample t_1 to last sample t_4. D is the union of all thunderstorm periods D_t (from t_2 to t_3), and all non-thunderstorm periods D_{nt} (periods t_1 to t_2 and t_3 to t_4). In magnitud, only thunderstorm data above its threshold b_t, and only non-tunderstorm data above its threshold b_{nt} are used.

Thunderstoms and non-thunderstorms are modeled independently:

1. Observations in domain D follow a Poisson distribution with mean $\int_D \lambda(t,y)\,dt\,dy$
2. For each disjoint subdomain D_1 or D_2 inside D, the observations in D_1 or D_2 are independent random variables.

Visual representation of the intensity function for the Poisson Process can be seen in figure 2.9. In vertical axis, two surfaces were drawn representing independent intensity functions for thunderstorm $\lambda_t(y)$ and for non-thunderstorm $\lambda_{nt}(y)$. The volume under each surface

for its corresponding time periods and peak (over threshold) velocities, is the mean of the Poisson Process.

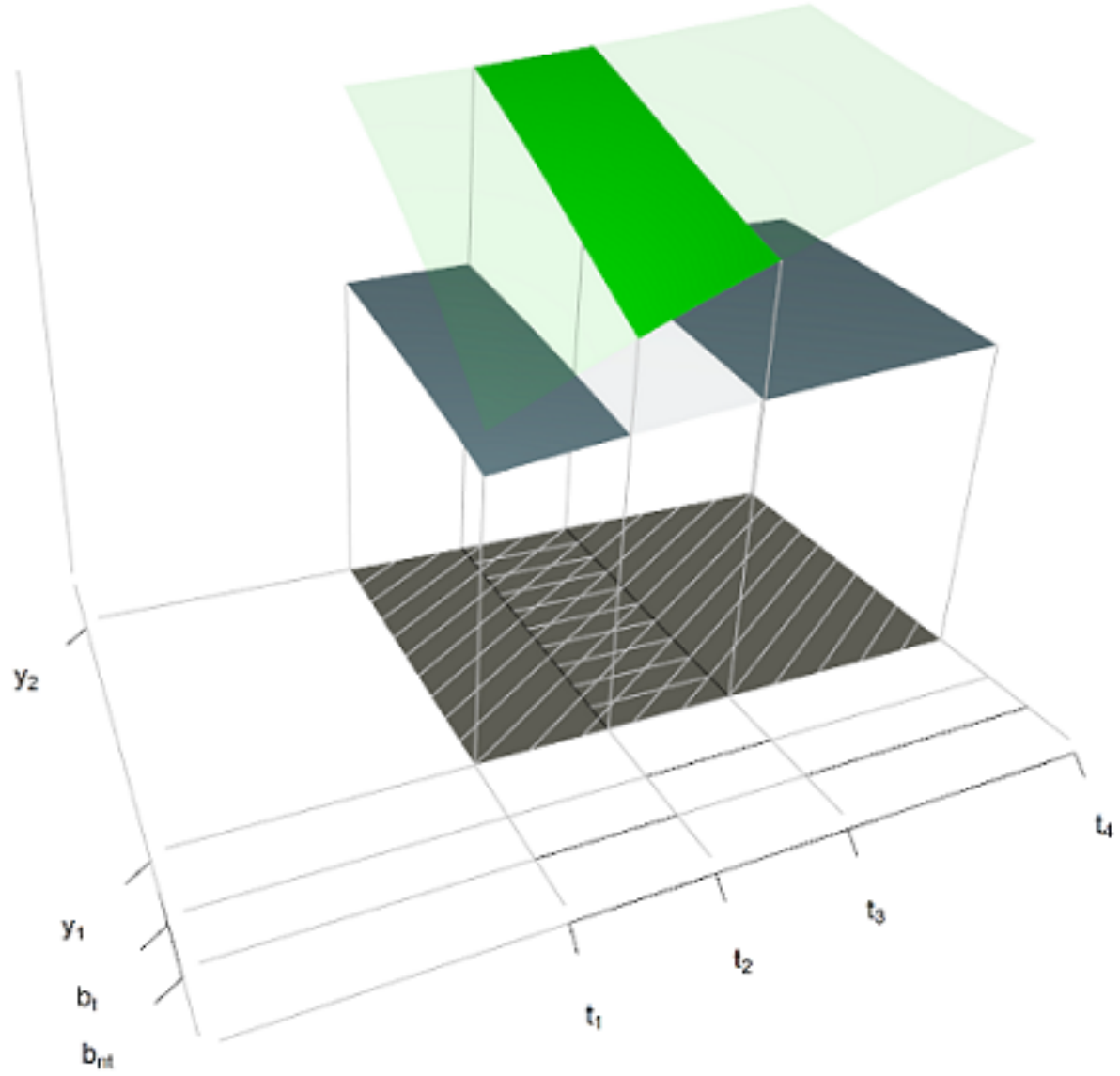

Figure 2.9: Volume under surfaces represents the mean of the Poisson process

The method of maximun likelihood es used to estimate the parameters of the Poisson process, the selected vector of parameters η are the $\hat{\eta}$ values that maximizes the function

$$\mathrm{L}(\eta) = \left(\prod_{\mathrm{i}=1}^{\mathrm{I}} \lambda \left(\mathrm{y_i}, \mathrm{t_i} \right) \right) \exp \left\{ - \int_{\mathrm{D}} \lambda \left(\mathrm{y}, \mathrm{t} \right) \mathrm{dy}\, \mathrm{dt} \right\} \tag{2.7}$$

$\hat{\eta}$ values need to be calculated using a numericall approach because there is not analytical solution available.

Once the Poisson process is fittet to the data, the model will provide extreme wind velocities (return levels), for different return periods (mean recurrence intervals).

A Y_N extreme wind velocity, called the return level (RL) belonging to the N-years return period, has a expected frequency to occur or to be exceeded (annual excedance probability)$P_e = \frac{1}{N}$, and also has a probability that the event does not occur (annual non-excedance probability) $P_{ne} = 1 - \frac{1}{N}$. Y_N will be the resulting value of the G (ppf or quantile) function using a probability equal to P_{ne}. $Y_N = quantile(y, p = P_{ne}) = G(x, p = P_{ne}) = ppf(x, p = P_{ne})$. As for this study $\zeta = 0$, the G function to use is the Gumbel quantile function. Y_N can be undestood as the wind extreme value expected to be exceeded on average once every N years.

For different POT approaches, as POT-GPD described –, the value of the probability passed to the G function, has to be modified with the λ parameter, as is described in next equation. λ is the number of wind speed over the threshold per year.

$$Y_N = G\left(y, 1 - \frac{1}{\lambda\, N}\right)$$

For the Poisson process Y_N is also the solution to the next equation, which is defined in terms of the intensity function,

$$\int_{\mathrm{Y_N}}^{\infty}\int_{0}^{1} \lambda\,(\mathrm{y},\mathrm{t})\,\mathrm{dydt} = \mathrm{A_t}\int_{\mathrm{Y_N}}^{\infty} \lambda_{\mathrm{t}}\,(\mathrm{y})\;\mathrm{dy} + \mathrm{A_{nt}}\int_{\mathrm{Y_N}}^{\infty} \lambda_{\mathrm{nt}}\,(\mathrm{y})\;\mathrm{dy} = \frac{1}{\mathrm{N}} \tag{2.8}$$

where A_t, is the multiplication of the average number of thunderstorm per year and the average lengh of a thunderstorm (taken to be 1 hour as defined in Pintar et al. (2015)), and $A_{nt} = 1 - A_t$. The average length of a non-thunderstorm event is variable, and it is adjusted in each station to guarantee that $A_{nt} + A_t = 1$

The same thunderstorm event in considered to occur if the time lag distance between sucesive thunderstorm samples is small than six hours, and for non-thunderstorm this time is 4 days. For the Poisson process, all the measurements belonging to the same event (thunderstorm or non tunderstorm), need to be declustered to leave only one maximun value. In other words, the number of thunderstorm in the time serie is the number of time lag distances grather than 6 hours, and for non-thunderstorm grather than 4 days.

###Threshold Selection

$$U = F(Y)$$
$$W = -log(1 - U)$$

2.5 Wind Loads Requirements

As the output maps of this research will be used as input loads for infrastructure design, the methodology used for its creation, need to be consistent with Colombian official wind loads

requirements. Today (2020), the Colombian norm that defines wind loads is the Seismic Resistant Standard 2010 - NSR-10 by its acronym of Spanish, see XXX. Chapter related to wind loads is B.6. NSR-10 was created and is maintained by the Colombian Association of Seismic Engineering - AIS.

NSR-10 is mainly based in the USA norm American Society of Civil Engineers 7-16, minimum design loads and associated criteria for buildings and other structures - ASCE7-16, see Engineers (2017). Under these circumstances, ASCE7-16 defines the minimum requirements of the research products. Especially the chapter C26 - "wind loads - general requirements", C26.5 "wind hazard map", and C26.7 "Exposure" - pages 733 to 747. Wind speeds requirements of ASCE7-16 are based in the combination of independent non-hurricane analysis, and hurricane wind speeds simulations models. The focus of this research will be the analysis of non-hurricane wind data, however, existing results of hurricane studies will be used to present final maps with both components. In ASCE7-16, for non-hurricane wind speed, the procedure is mainly based on Pintar et al. (2015).

ASCE7-16 (page 734), requires the calculation of wind extreme return levels for specific return periods according to the risk category of the structure to be designed: risk category I - 300 years, risk category II - 700 years, risk category III - 1700 years, risk category IV - 3000 years. NSR-10 only requires 700, 1700 and 3000 years. In addition, extreme wind speeds for those MRI need to correspond to: - 3 second gust speeds, - at 33 ft (10 meters) above the ground, and - exposure category C (open space).

- Risk IV - This are 'indispensable buildings' that involve sustancial risk. This structures that can handle toxic or explosive substances.
- Risk III - There is sustancial risk because this structures that can handle toxic or explosive substances, can cause a serious economical impact, or masive interruption of activities if they fail
- Risk II - Category 'by default', and correspond to structures not classified in others categories.
- Risk I - This structures represent low risk for people lifes

To standarize wind speeds to gust speeds ASCE7-16 proposes the curve Durst (see C. S. Durst (1960), and figure 2.10). It is valid only for open terran conditions. Durst curve shows in axis y the gust factor $\frac{V_t}{V_{3600}}$, a ration between any wind gust averaged at t seconds, V_t, and the hourly averaged wind speed V_{3600}, and in the axis x the duration t of the gust.

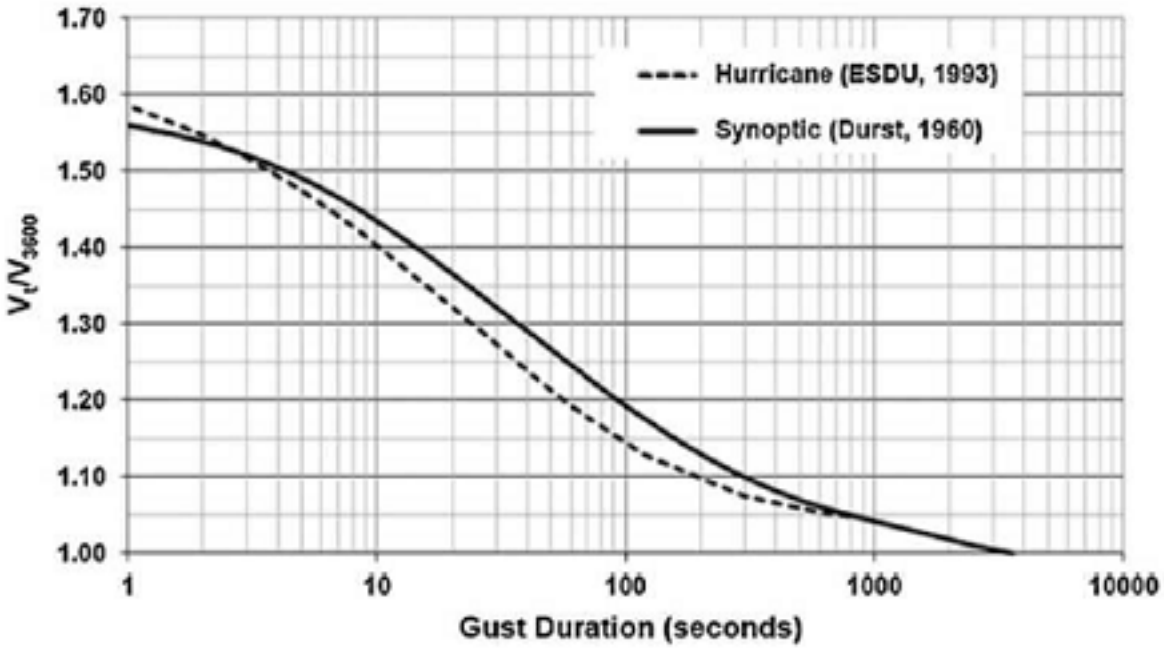

Figure 2.10: Maximun speeds averaged over t (sec), to hourly mean speed

Chapter 3

Methodology

3.1 Input Data Selection and Standarization

3.1.1 Data Selection

3.1.2 Data Standarization

Anemometer height - 10 m

Surface Roughness - 0.03 m

Averaging Time - 3-s gust

3.1.3 Data Filterng

3.2 Fit data to a POT - Poisson Process

3.2.1 Data Requirements

3.2.2 Exploratory Data Analysis and Data Preparation

Declustering of observations

Exclude no-data periods

Threshold selection

3.2.3 Parameters Estimation

Intensity function

Density function

Distribution function

Maximun likelihood estimation

3.2.4 Velocities at Return Periods

3.3 spatial Interpolation

Conclusion

If we don't want Conclusion to have a chapter number next to it, we can add the `{-}` attribute.

More info

And here's some other random info: the first paragraph after a chapter title or section head *shouldn't be* indented, because indents are to tell the reader that you're starting a new paragraph. Since that's obvious after a chapter or section title, proper typesetting doesn't add an indent there.

Appendix A

R Code

This first appendix includes all of the R chunks of code that were hidden throughout the document (using the `include = FALSE` chunk tag) to help with readibility and/or setup.

In the main Rmd file

```
# This chunk ensures that the thesisdown package is
# installed and loaded. This thesisdown package includes
# the template files for the thesis.
if(!require(devtools))
  install.packages("devtools", repos = "http://cran.rstudio.com")
if(!require(thesisdown))
  devtools::install_github("ismayc/thesisdown")
library(thesisdown)
```

In Chapter 3:

```
# This chunk ensures that the thesisdown package is
# installed and loaded. This thesisdown package includes
# the template files for the thesis and also two functions
# used for labeling and referencing
if(!require(devtools))
  install.packages("devtools", repos = "http://cran.rstudio.com")
if(!require(dplyr))
    install.packages("dplyr", repos = "http://cran.rstudio.com")
if(!require(ggplot2))
    install.packages("ggplot2", repos = "http://cran.rstudio.com")
if(!require(ggplot2))
    install.packages("bookdown", repos = "http://cran.rstudio.com")
if(!require(thesisdown)){
  library(devtools)
  devtools::install_github("ismayc/thesisdown")
  }
```

```
library(thesisdown)
flights <- read.csv("data/flights.csv")
```

Appendix B

The Second Appendix

References

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer London. http://doi.org/10.1007/978-1-4471-3675-0

C. S. Durst, B. A., O. B.E. (1960). Wind speeds over short periods of time. *The Meteorological Magazine*, *89*(1056), 181–187. Retrieved from https://www.depts.ttu.edu/nwi/Pubs/ReportsJournals/ReportsJournals/Windspeeds.pdf

Engineers, A. S. O. C. (2017). *Minimum design loads and associated criteria for buildings and other structures (asce7-16)*. American Society of Civil Engineers. Retrieved from https://www.ebook.de/de/product/35017614/american_society_of_civil_engineers_minimum_design_loads_and_associated_criteria_for_buildings_and_other_structures_7_16.html

Harris, J. W., & Stocker, H. (1998). Maximum likelihood method. In *Handbook of mathematics and computational science* (p. 824). Springer-Verlag.

Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis*. Cambridge University Press. http://doi.org/10.1017/cbo9780511529443

Kubler, J. (1994). *Computational Statistics &Amp; Data Analysis*, *18*(4), 473–474. Retrieved from https://EconPapers.repec.org/RePEc:eee:csdana:v:18:y:1994:i:4:p:473-474

Pickands, J. (1971). The two-dimensional poisson process and extremal processes. *Journal of Applied Probability*, *8*(4), 745–756. http://doi.org/10.2307/3212238

Pintar, A. L., Simiu, E., Lombardo, F. T., & Levitan, M. L. (2015). *Simple guide for evaluating and expressing the uncertainty of NIST MeasuremenMaps of non-hurricane non-tornadic wind speeds with specified mean recurrence intervals for the contiguous united states using a two-dimensional poisson process extreme value model and local regressiont results*. National Institute of Standards; Technology.

Simiu, E., & Scanlan, R. H. (1996). *Wind effects on structures : Fundamentals and applications to design* (3rd ed.). New York : John Wiley. Retrieved from http://lib.ugent.be/catalog/rug01:001267836

Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science*, *4*(4), 367–377. http://doi.

org/10.1214/ss/1177012400

Smith, R. L. (2004). Extreme values in finance, telecommunications, and the environment (chapman & hall/crc monographs on statistics and applied probability). In B. F. inkenstädt & H. Rootzén (Eds.), (pp. 1–78). Chapman; Hall/CRC. Retrieved from https://www.amazon.com/Telecommunications-Environment-Monographs-Statistics-Probability/dp/1584884118?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=1584884118