Spatio temporal analysis of extreme wind velocities for infrastructure desing. Case study Colombia

A Thesis
Presented to
The Division of Instituto for Geoinformatics - IFGI
University of Münster

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in Geospatial Technologies

Alexys Herleym Rodríguez Avellaneda

Jan 2020

Approved for the Division
(Faculty of Geosciences)

Dr. Edzer Pebesma

Dr. Juan C. Reyes\Dr. Sara Ribero

Acknowledgements

I want to thank a few people.

Preface

This is an example of a thesis setup to use the reed thesis document class (for LaTeX) and the R bookdown package, in general.

Table of Contents

List of Tables

List of Figures

Abstract

The preface pretty much says it all.

Second paragraph of abstract starts here.

Dedication

You can have a dedication here if you wish.

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, Simiu, Lombardo, & Levitan (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.

```
library(knitr)
hook_output = knit_hooks$get('output')
knit_hooks$set(output = function(x, options) {
  # this hook is used only when the linewidth option is not NULL
  if (!is.null(n <- options$linewidth)) {
    x = knitr:::split_lines(x)
    # any lines wider than n should be wrapped
    if (any(nchar(x) > n)) x = strwrap(x, width = n)
    x = paste(x, collapse = '\n')
  }
```

```
  hook_output(x, options)
})
```

```
# List of packages required for this analysis
pkg <- c("dplyr", "sf", "ggplot2","rnaturalearth", "rnaturalearthdata",
         "ggspatial", "kableExtra", "ncdf4", "stars", "magick", "RcmdrMisc",
         "knitr", "bookdown", "devtools")
# Check if packages are not installed and assign the
# names of the packages not installed to the variable new.pkg
new.pkg <- pkg[!(pkg %in% installed.packages())]
# If there are any packages in the list that aren't installed,
# install them
if (length(new.pkg))
  install.packages(new.pkg, repos = "http://cran.rstudio.com")
# Load packages (thesisdown will load all of the packages as well)
library("thesisdown")
library("dplyr")
library("sf")
library("ggplot2")
library("rnaturalearth")
library("rnaturalearthdata")
library("ggspatial")
#library("tibble")
library("knitr")
library("kableExtra")
library("ncdf4")
library("stars")
library("magick")
library("RcmdrMisc")
```

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 0.25º (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

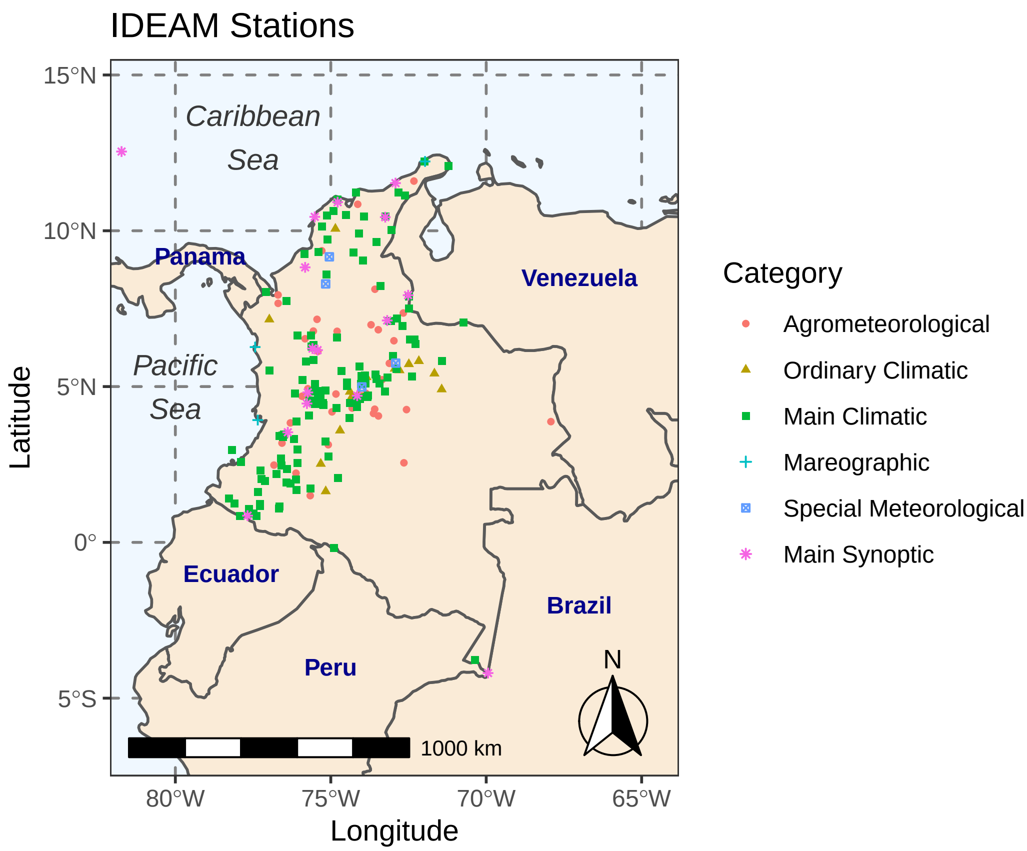

Figure 1.1: IDEAM Stations

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

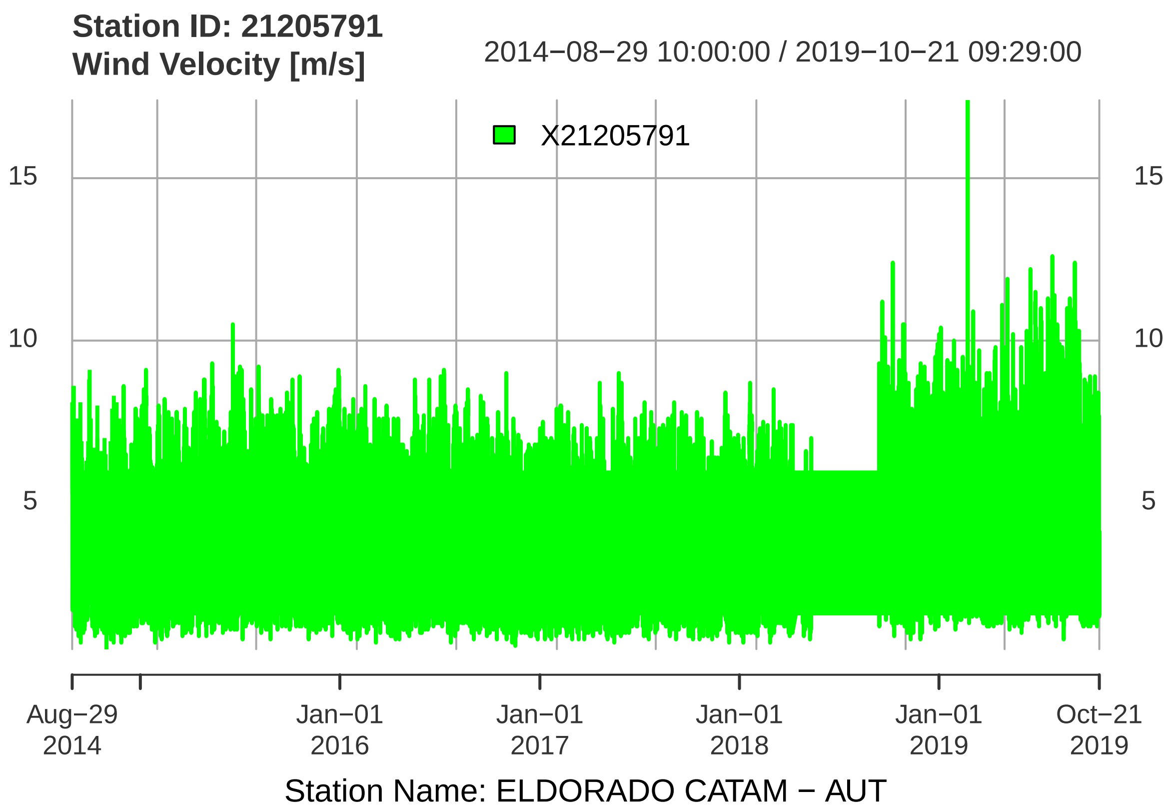

Figure 1.2: IDEAM Station - Time Serie

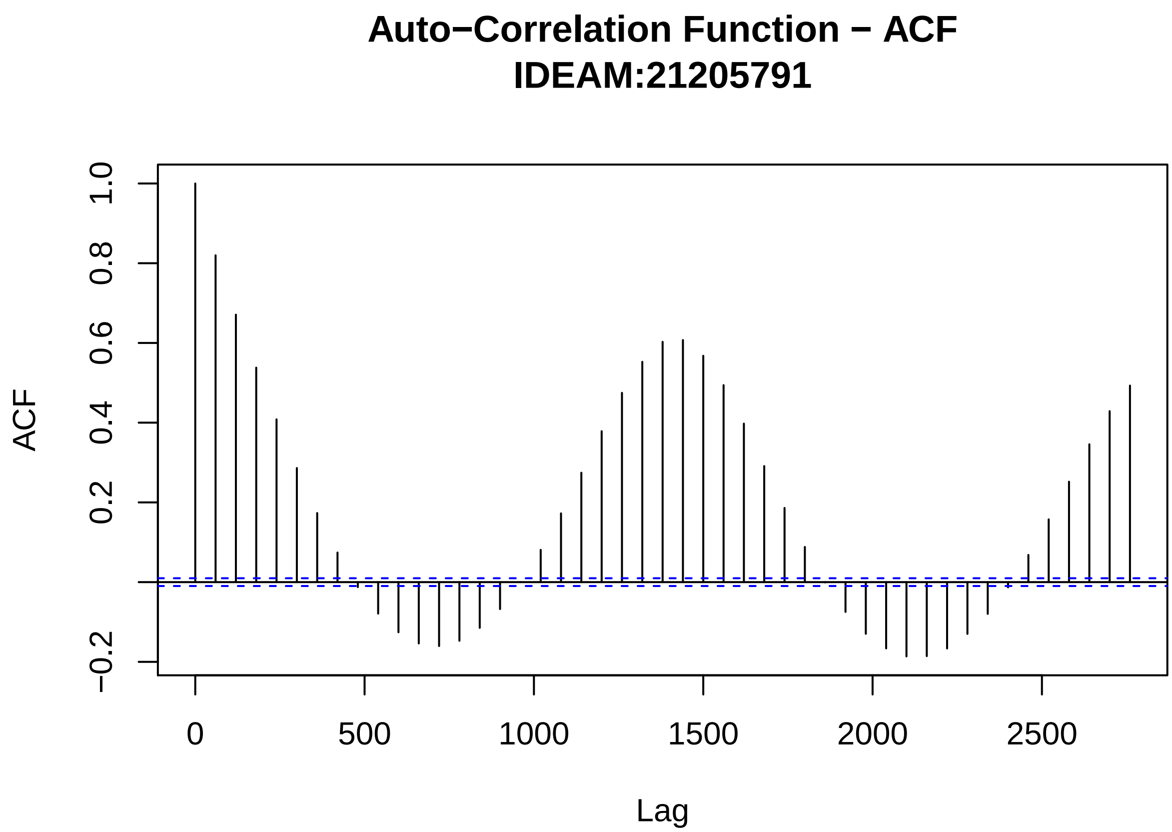

Figure 1.3: IDEAM Station ACF

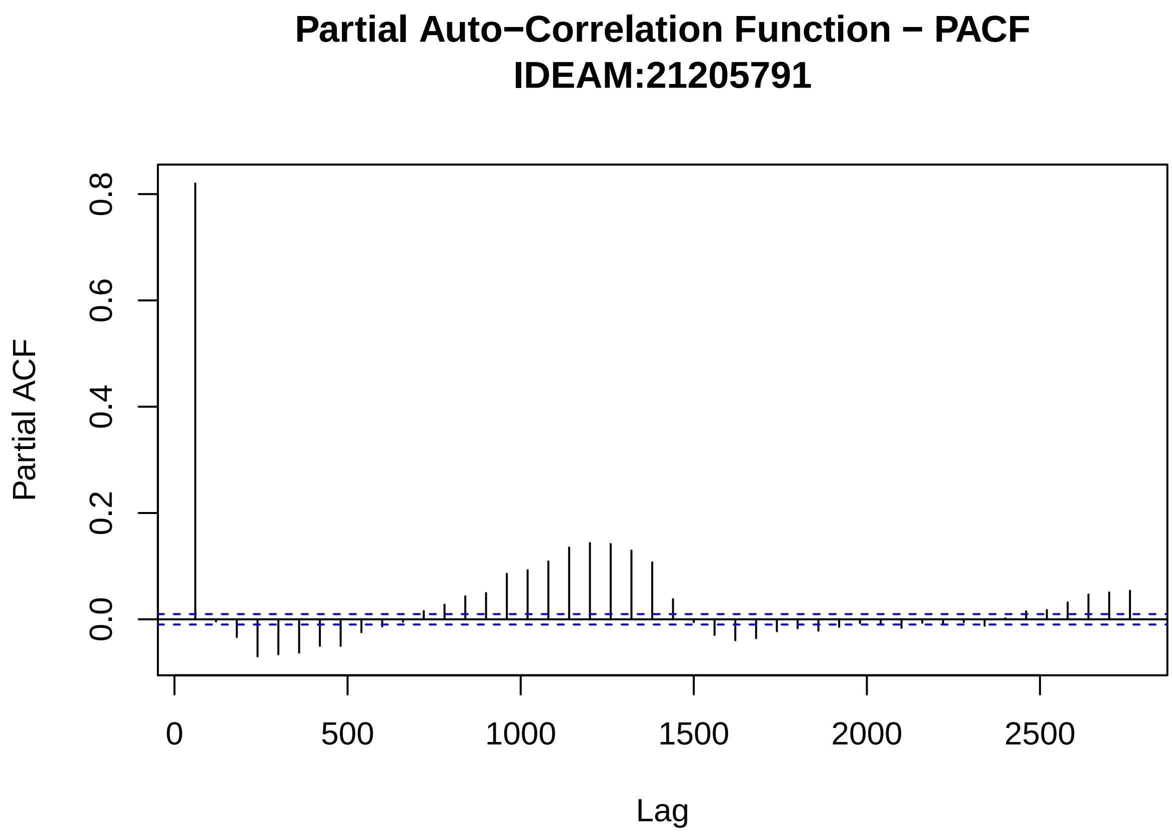

Figure 1.4: IDEAM Station PACF

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

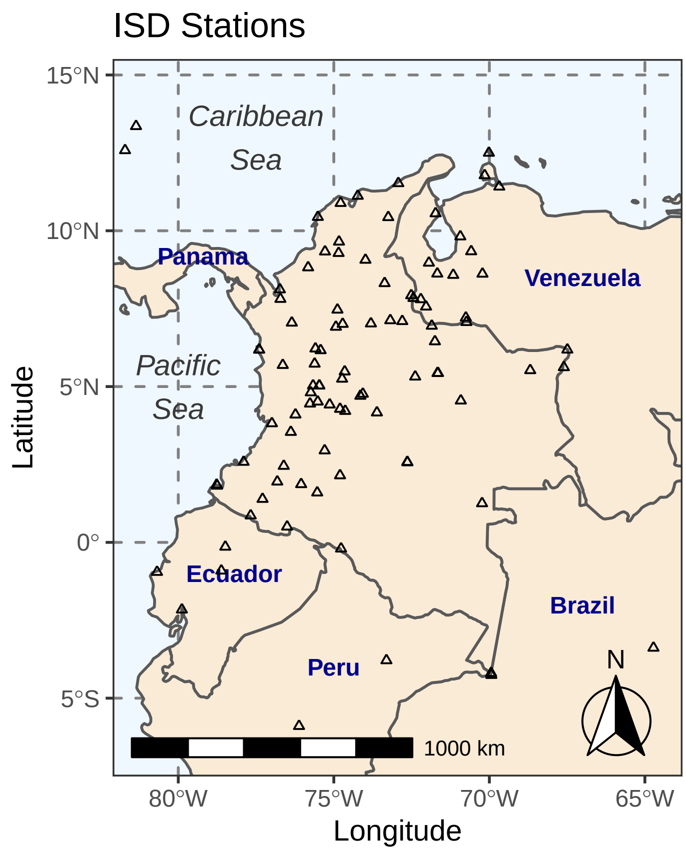

Figure 1.5: ISD Stations

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 N
```

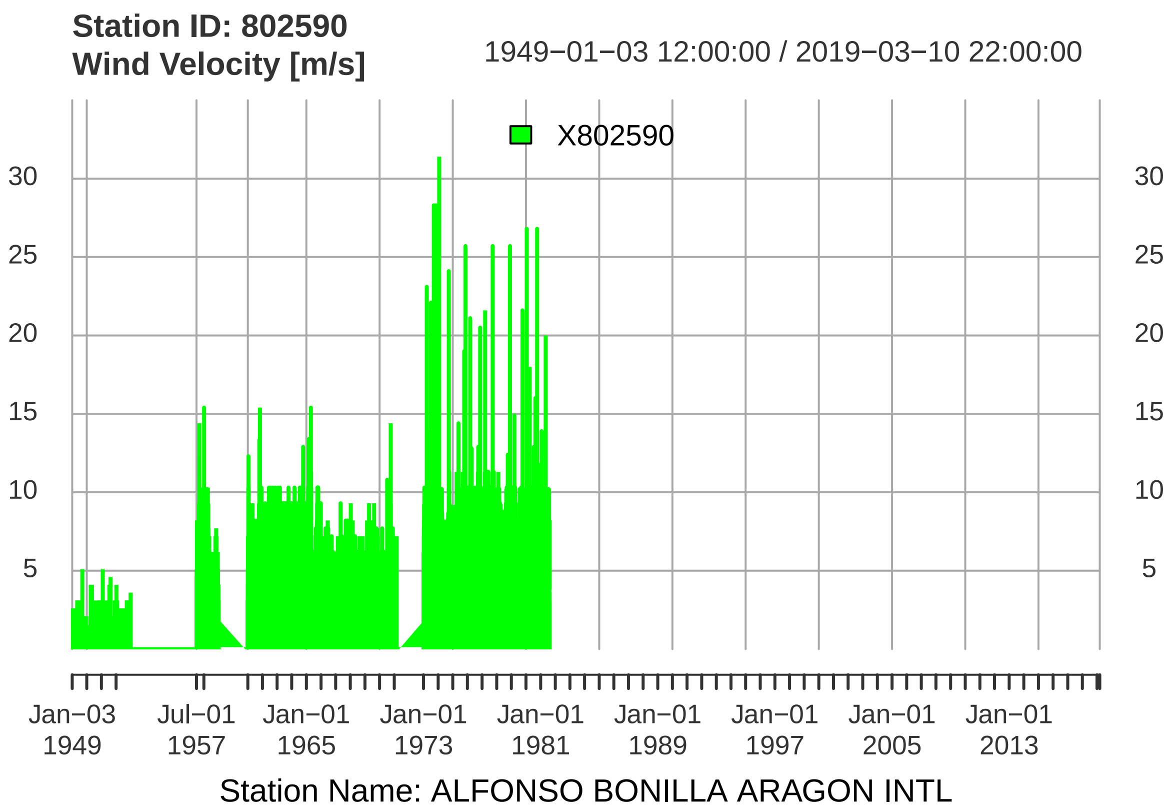

Figure 1.6: ISD Station - Time Serie

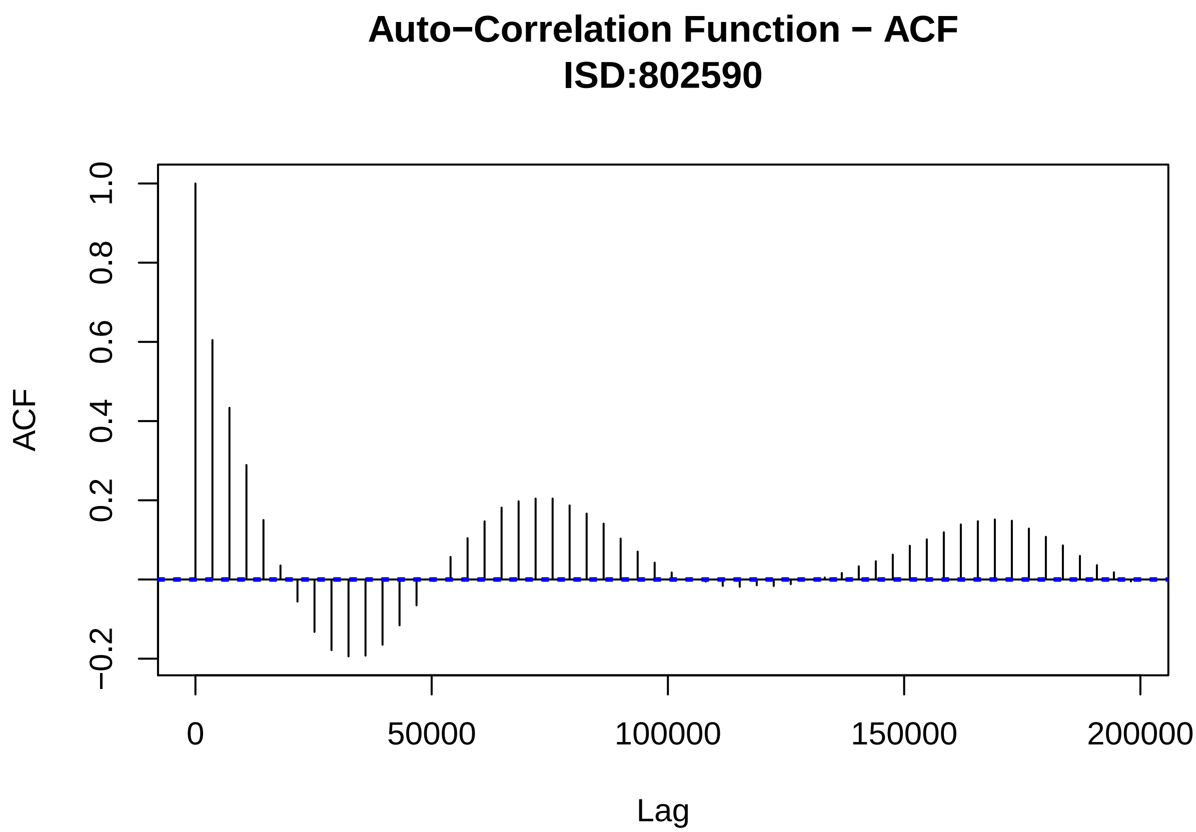

Figure 1.7: ISD Station ACF

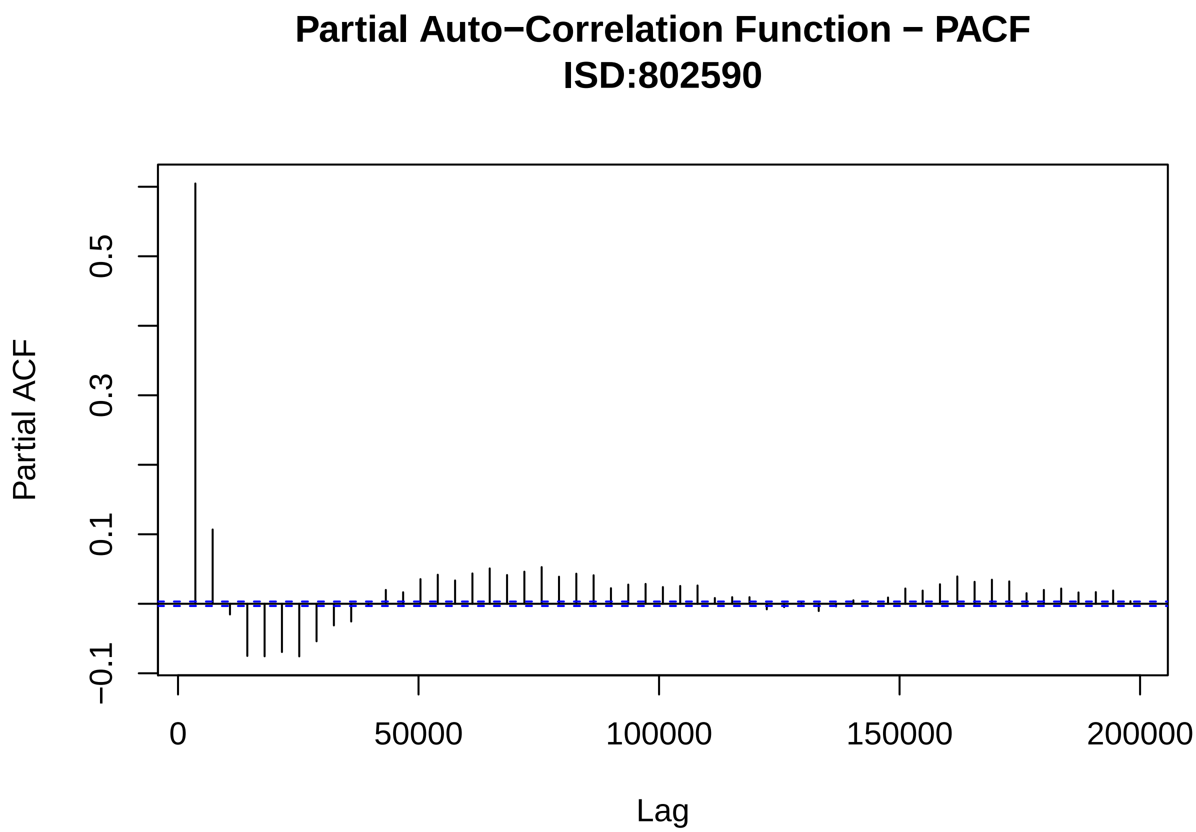

Figure 1.8: IDEAM Station PACF

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).

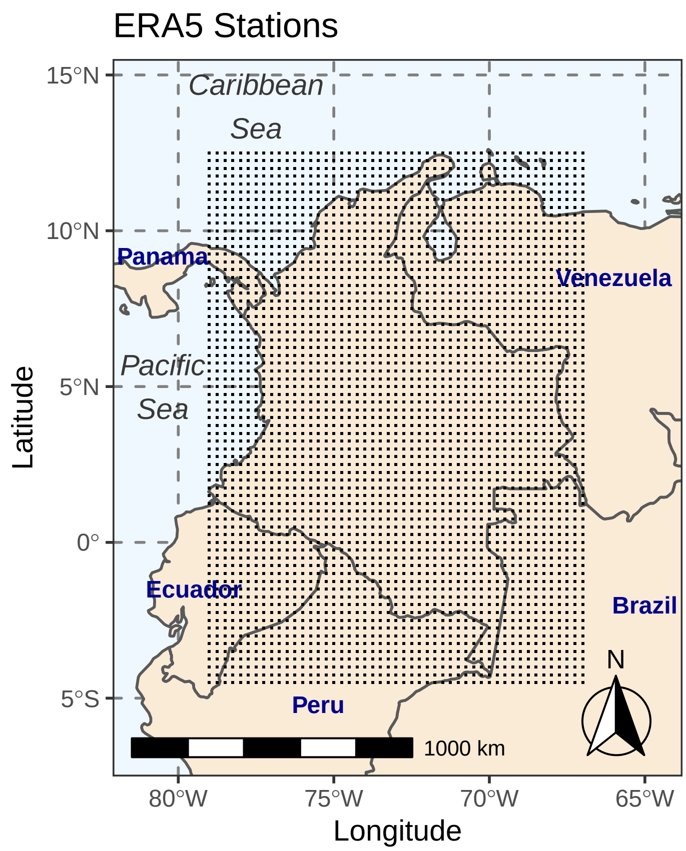

Figure 1.9: ERA5 Stations (cells centers)

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)
```

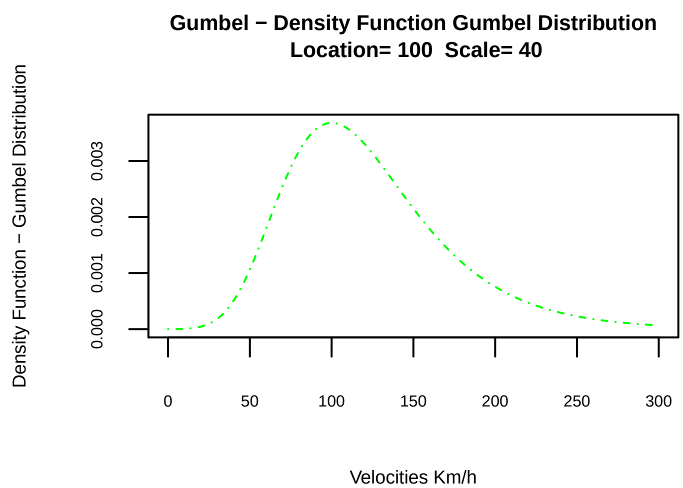

Figure 2.1: Gumbel pdf

Figure 2.2 shows *pdf* with location $(\mu) = 100$ and scale $(\beta) = 40$, 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)
```

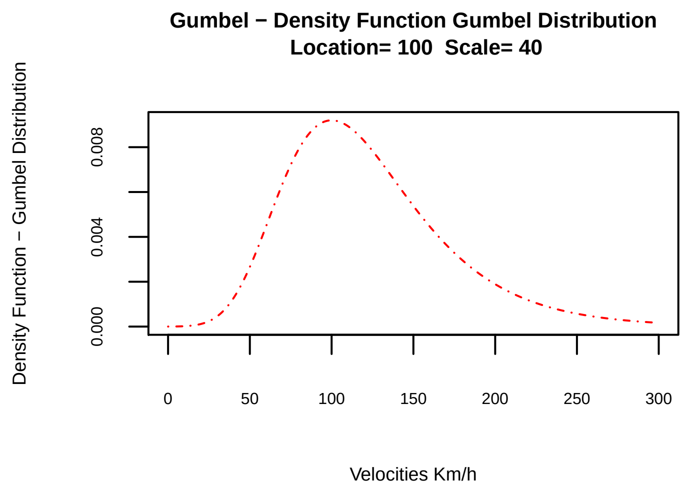

Figure 2.2: Gumbel pdf - dgumbel function

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 $(\mu) = 100$ and scale $(\beta) = 40$, 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)
```

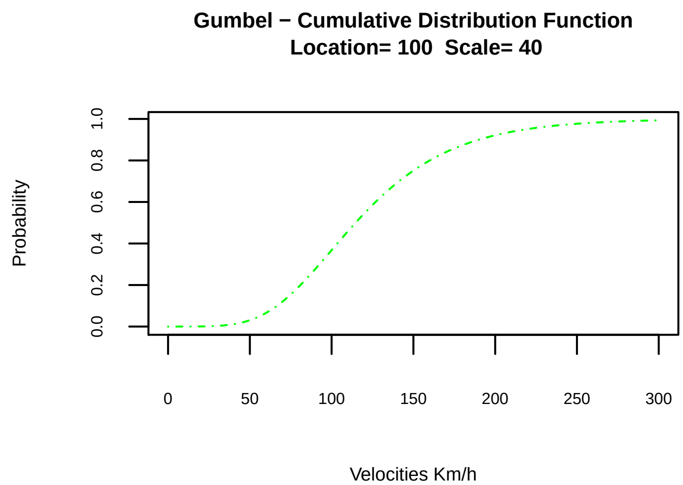

Figure 2.3: Gumbel cdf

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)
```

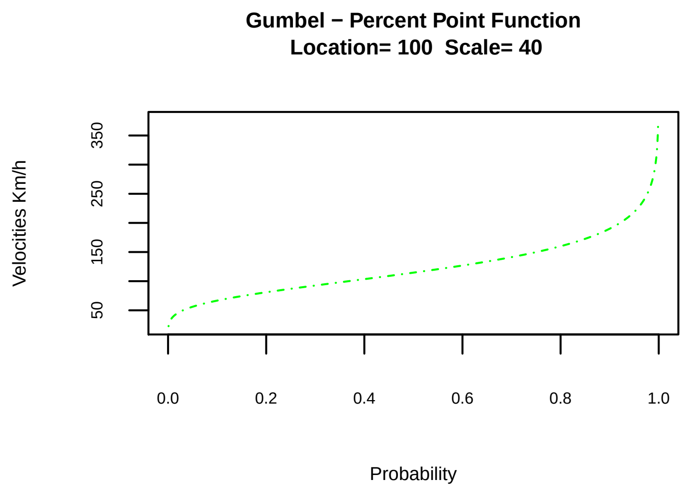

Figure 2.4: Gumbel cdf

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)
```

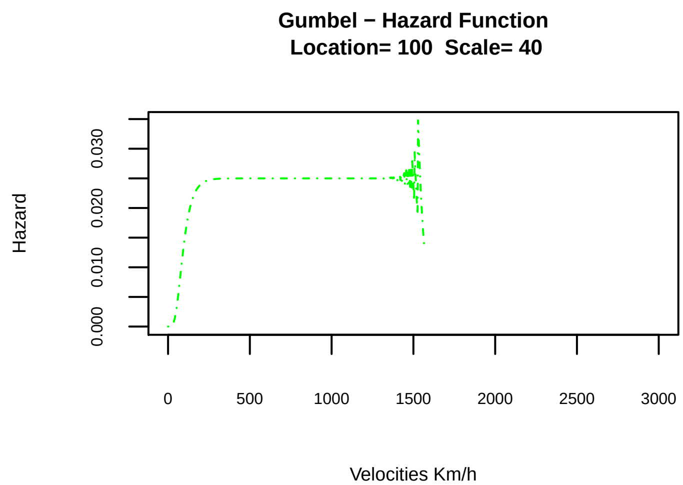

Figure 2.5: Gumbel cdf

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

Introductory concepts for statistical analysis of extreme values

2.1.5 Annual Excedance Probability - Pa

2.1.6 Return Period

2.1.7 Compound Excedance Probability - Pn

2.2 Extreme Value Analysis Overview

2.3 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).

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

Generic equation (**??**) 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).

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

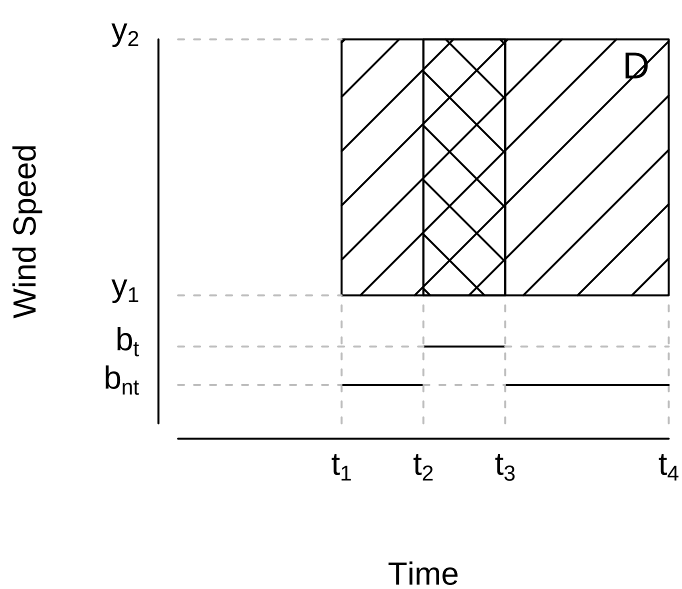

Figure 2.6: Domanin off the Poisson Process

Figure 2.6 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 2.7. 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.

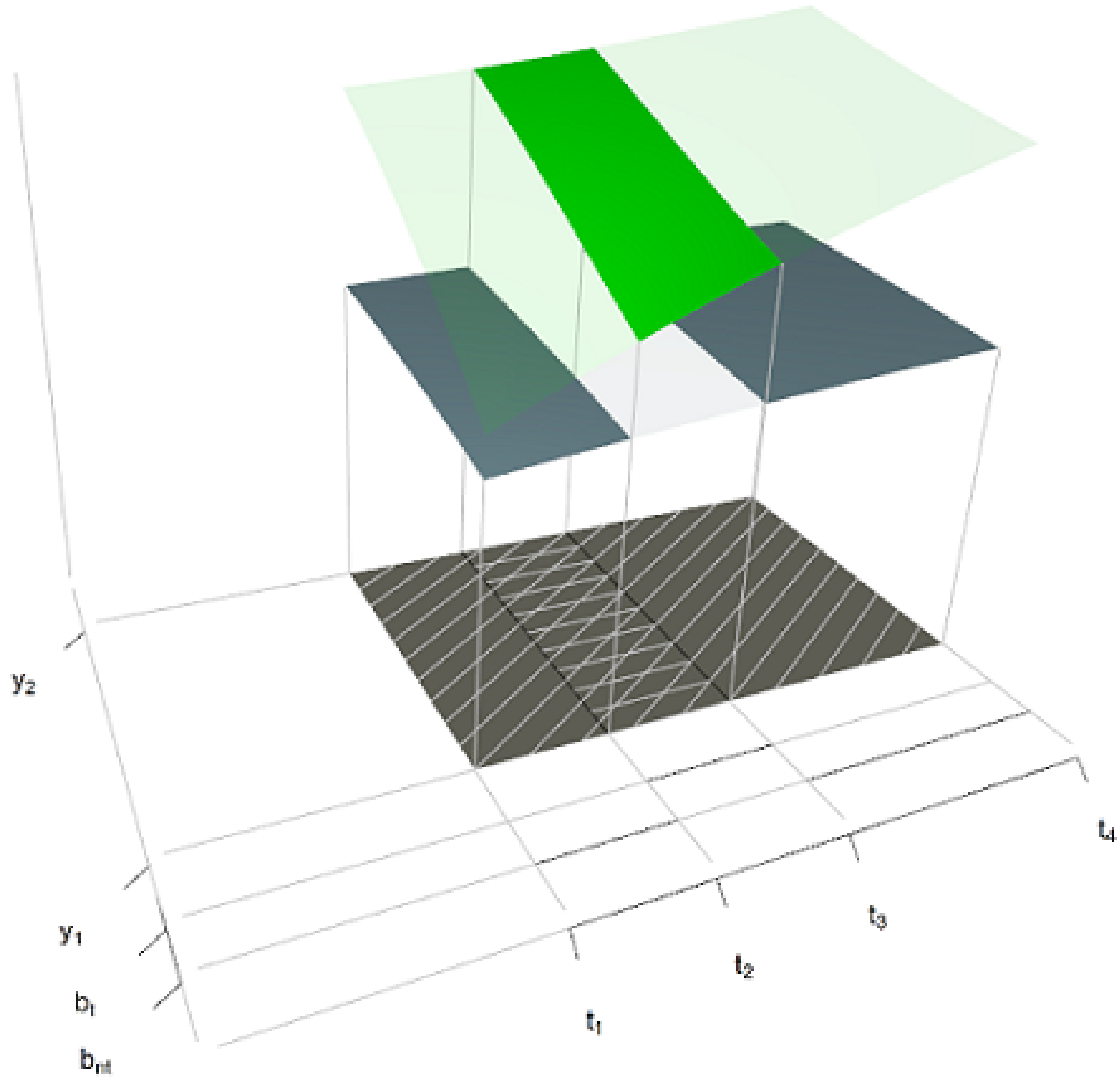

Figure 2.7: Volume under surfaces represents the mean of the Poisson process

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,

$$\int_{\mathrm{Y_N}}^{\infty}\int_0^1 \lambda\left(\mathrm{y},\mathrm{t}\right)\mathrm{dydt} = \mathrm{A_t}\int_{\mathrm{Y_N}}^{\infty}\lambda_{\mathrm{t}}\left(\mathrm{y}\right)\,\mathrm{dy} + \mathrm{A_{nt}}\int_{\mathrm{Y_N}}^{\infty}\lambda_{\mathrm{nt}}\left(\mathrm{y}\right)\,\mathrm{dy} = \frac{1}{\mathrm{N}} \tag{2.8}$$

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

$$U = F(Y)$$

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

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

The First Appendix

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")
  }
```

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

Appendix B

The Second Appendix, for Fun

References

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.

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