This research follows the methodological approach defined in chapter — of the ASCE7-16 standard (ASCE, 2017). ~~ASCE7~~ ~~This document use~~ ~~Design wind velocities in this document depends on~~ ~~The document defines for risk categories~~ The ASCE7-16 considers design wind velocities for various mean recurrence intervals (MRI), depending on the risk category (RC) of the structure, as ~~follows~~: follows: MRI = 700 years for risk category I and II, 1700 years for RC III, and 3000 for RC IV.

Acknowledgements

Special thanks to Prof. Dr. **Edzer Pebesma**, first, for all the contributions to the open source community, considering that main work in this thesis was done using his R packages, especially *sf*, *stars* and *gstat*, and second, for all high level knowledge transmitted through the subjects *Spatial Data Science with R* and *Analysis of Spatio-Temporal Data*, which were the motivation and basis to carry out the investigation.

Special thanks to Prof. Dr. **Juan C Reyes** for his contribution in selecting the research topic, and great contributions in information, methodology and support.

I would like to thank to:

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 to have that support always available, and a pleasure to work beside you. Dr. **Adam Pintar**, for sharing its related POT-PP R Code, and for devoting much of his time to reviewing and commenting on my progress. Dr. **Joaquín Huerta Guijarro**, ~~because he always was~~ available to help ~~and he was very friendly and receptive~~ me. Dr. **Christoph Brox**, because he was beside me in the difficult moments of the incident and the surgery. **European Union** -'Erasmus Mundus Grant', because their funding allow me to fulfill this dream to go further with my academic and professionals dreams. Engineer **Juan David Sandoval** for its valuable contributions. My mother **Ligia** made possible all my achievements, because she was always there with love, support, and valuable advice, I am grateful with all my heart. My daughter **Nicolle Chaely** for its love, support, and always pleasant company. Family members as **Elsa Manrique**, **Barbara Avellaneda**, and **Kevin Martinez**, because they were an important source of motivation and support. To all the beautiful people that shared with me different activities at **San Antonius Church of Münster**, with special mention of father **Alejandro Serrano Palacios** for support and friendship, and **choir friends**.

for ~~being~~ being receptive ~~and~~ friendly and

revisar formato de referencias usualmente se usa solo un apellido Beirlant et al (2004), cuando son mas de dos autores

Las referencias usualmente van al comienzo y final de las oraciones

Hay dos opciones usar entre [ ] numeros o (Beirlant et al, 2004, Haigh & Wahl, 2019)

Apellido ( ) solo se usa al comienzo de una oración

Alex, por fa revisar formato de agradecimientos en journals

Por ejemplo:
The author is thankful to.... for.....
Gratitute is extended to.... for.....
The outstanding help of..... is highly appreciated.
The author thanks......

~~Preface~~ Summary

~~*Models of*~~ models E*xtreme values* are used for ~~designing against the effects of~~ estimating design forces of *extreme events* like earthquakes, winds, rainfall, floods, engineering etc. ~~of different types of physical processes, see~~ Beirlant, Goegebeur, Teugels, & Segers (2004), a~~voiding widespread destruction and loss of lives, see~~ Haigh & Wahl (2019). ~~This research presents a applied case of univariate extreme value analysis, explained in detail in Smith (2004), applied to wind velocities for infrastructure design.~~ Consequently, the main interest are probable future ~~more~~ extreme ~~wind~~ events that structures need to be able to resist.

This research presents an application of univariate extreme value analysis to estimate wind velocities for infrastructure design. (Smith, 2004)

~~This work in its theoretical 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 structure design norm, see Ministerio de Vivienda (2010), maintained by the Colombian Association of Seismic Engineering - AIS by its Spanish 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.~~ of the four risk categories

design

~~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-processed information coming from the Integrated Surface Database ~~- ISD (USA database based on IDEAM data source), see~~ (~~Smith, Lott, & Vose (2011)~~), and ~~forecast reanalysis data from ERA5, see European Centre For Medium-Range Weather Forecasts (2017)~~. This condition demanded a comparison of the different data sources, in order to verify the feasibility in the use of ERA5 and ISD, ~~with a previous process of standardization of wind velocities (only for IDEAM and ISD)~~, to ~~reach the needed~~ meet the requirement of 3-s wind gust speed, 10 meters anemometer height, and terrain open space condition.

At each station, the ~~used method~~ Peaks Over Threshold - Poisson Process, required to identify all the non-thunderstorm events in the non-hurricane dataset, through a process of declustering, choosing a suitable threshold level to leave for the analysis only the most extreme values available, and then, fit to the data a intensity function, using maximum likelihood to find optimal parameters with the best goodness of fit. With the fitted model, it was possible to calculate return ~~levels~~ wind velocities for the required mean return intervals. ~~Next~~ Finally, a process of spatial interpolation was done using Kriging, what allowed to have three continuous maps for the whole study area.

no parece adecuado tanta referencia en el resumen

Alex creo que solo debería haber uno de estos dos: Preface ó Abstract Integrarlos
Son redundantes

Table of Contents

No numerar un solo item

List of Tables

mejorar nombres iniciar con mayuscula

Colocar un nombre mas descriptivo.

List of Figures

Pensé que el Preface era el resumen solo dejar uno de los dos

Abstract

Integrarlos

For the input non-hurricane, non tornadic data in each available station of the study area, this research calculate extreme winds or return levels for 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 combined with existing wind extreme hurricane studies, to be used as input loads for infrastructure design.

MRI

The development of this research focused in non-hurricane data covers three main areas, *downscaling support*, *temporal analysis*, and *spatial analysis*, and includes ~~in the end~~ an integration process with *existing results of hurricane studies*, ~~which all together~~, allow to generate extreme winds maps with different ~~mean recurrence intervals~~ (MRIs) for the design of structures of different risk categories, ~~namely~~, less risky/important structures for short MRIs (700 and 1700 years), and highly important structures for the longest MRI of 3000 years.

i.e.

The integration of all these results

Due to the specific characteristics of the study area, where there is lack of historical wind measurements ~~(IDEAM data source)~~, it became necessary to look for *alternative data sources*, ISD ~~(model, based on IDEAM)~~, and ERA5 (forecast data), ~~which~~ resulted in ~~the~~ downscaling issue, ~~and~~ that was confronted from a graphic comparison of all sources by matching stations, in the search of adequate support for the use of complementary data. The result of the comparison showed little similarity between the different sources. Prior to the comparison process, ISD and IDEAM data sources were standardized to represent 3-second wind gust, 10 meters anemometer height, and terrain open space roughness.

a

This

Tienen oraciones demasiado largas

The method of temporal analysis used to calculate the return levels at each station, from the historical wind time series, is the Peaks Over Threshold POT, using a non-homogeneous, bi-dimensional Poisson Process PP, recommended by ~~Engineers~~ (2017), and developed ~~and implemented in~~ Pintar, ~~Simiu, Lombardo, & Levitan~~ (2015), considering ~~from a~~ maximum likelihood adjustment, the model ~~with the best goodness of fit~~. Main components of ~~this matters~~ are de-clustering, thresholding, intensity function fitting, hazard curve, and return levels calculation.

temporal analysis model?

et. al

ASCE

Non-hurricane maps where created for data sources ISD and ERA5, using Kriging as spatial interpolation method, ~~and~~ after the integration with hurricane studies, the results for ERA5 showed the most reliable final maps, despite limitations in input data. Due to the limitation in the classification of storm and non-storm data, ISD final map showed ~~very~~ very high wind values, which are ~~quite~~ unlikely.

This process

to obtain the parameters of

This shortcoming may be corrected when storm data is available.

previas

the

of

Las introducciones que escribo usualmente estan organizadas de la siguiente forma:

Primeros parrafos (secciones): El objetivo es convencer al lector de la importancia del problema general de investigación mediante estadísticas, experiencias previas, impactos documentados, etc.

Parrafos intermedios (secciones): El objetivo es presentar toda la literatura disponible sobre el problema y establecer claramente sus vacios de investigación. ~~Usual~~ Usualmente se incluyen entre 15 y 30 referencias utilizando un estilo unificado y aprobado por algun estandar

Ultimos parrafos (secciones): En respuesta a los vacios de investigación identificados previamente, este ultimo parrafo (seccion) convence al lector de la relevancia del estudio y su impacto. Aqui se resaltan los aportes (a la ciencia y estado de conocimiento) y lo novedoso de la investigación.

List of Acronyms

AIS	~~Seismic~~ Colombian Earthquake Engineering Association
ASCE	American Society of Civil Engineers
ASCE7-16	ASCE/SEI Design Loads Standard
cdf	Cumulative Distribution Function
EDA	Exploratory Data Analysis
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA5	ECMWF climate reanalysis dataset
EVD	Extreme Value Distribution (GEVD, GEV)
GEVD	Generalized Extreme Value Distribution (EVD, GEV)
GEV	Generalized Extreme Value Distribution (GEVD, EVD)
GPD	Generalized Pareto Distribution
hf	Hazard Function
IDEAM	Institute of Hydrology, Meteorology and Environmental Studies
IDW	Inverse Distance Weighted
ISD	Integrated Surface Database
MRI	Mean Return Interval or Return Period
NSR	Seismic Resistant Norm
NOAA	National Oceanic and Atmospheric Administration
NetCDF	Network Common Data Form
NCEI	NOAA's National Centers for Environmental Information
P_e	Annual Exceedance Probability
pdf	Probability Distribution Function
P_n	Compound Exceedance Probability
POT	Peaks Over Threshold
ppf	Percent Point Function (Quantile)
PP	Poisson Process
POT-GPD	Peaks Over Threshold - Generalized Pareto Distribution
POT-PP	Peaks Over Threshold - Poisson Process
RL	Return Level
RMSE	Root Mean Squared Error
SEI	Structural Engineering Institute
SQL	Structured Query Language
WGS84	World Geodetic System 1984

Chapter 1

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 study area (Colombian territory). These maps will be combined with existing hurricane wind speed studies, to be used as input loads due to wind for infrastructure design.

For each station with wind speeds time series in the input data, following Pintar et al. (2015), extreme wind speeds corresponding to each recurrence interval are calculated using a *Peaks Over Threshold Poisson Process* extreme value model, onwards *POT-PP*, 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, see Engineers (2017). The annual probability of equal or exceed that wind speed is 1/N. The annual exceedance 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.

esto parece nuevamente un resumen debería estar al final de la introducción

1.1 Context and Background

To design ~~a~~ an specific structure, ~~the~~ horizontal forces (wind and earthquake) play an starring role. For ~~the study area,~~ Colombia, initially, ~~the~~ wind force ~~was considered as~~ were calculated considering a fixed velocity - value 100 ~~Km/h~~, later, a continuous map with a return period of 50 years was included in the official design standard. ~~Then~~, an additional map with a return period of 700 year was ~~included~~ added. ~~see~~ (Ministerio de Vivienda, 2010).

Afterwards,

km/h minuscula kilo; mayuscula es Kelvin

In the context of this study, 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. namely, assess from the ordered sample of wind velocities, the probability of wind events that are more extreme than the ones previously observed and included in the mentioned input sample. ~~Classical reference in this matter is~~ Coles (2001) ~~where~~ presents a detailed study ~~is done~~ about classical extreme value theory and models

Organizar en una oracion nueva

velocity

redundante

especialmente preposiciones

Alex la redaccion en general tiene problemas de oraciones demasiado largas, redundancia y no buen uso de palabras. Voy a limitar mi revision a cosas mas generales.

En mi opinion 1.1.1 a 1.1.3 deberia estar en el capitulo 3. Creo que es demasiado especifico para estar en una introducción. Ver mi opinion de la informacion que deberia contener la intro en la pagina anterior.

and threshold models. In general,

There are four ~~main~~ approaches to deal with extreme value analysis (Smith (2004)): (a) sample maxima associated to a Generalized Extreme Value Distribution - GEV (traditional method); (b) exceedances over threshold associated to a Generalized Pareto Distribution, onwards *POT-GPD*; (c) the Poisson-GPD, an homogeneous Poisson process for the number of exceedances, and a GPD for the excess values, and (d) the exceedances over threshold associated to a non-homogeneous non-stationary bi-dimensional Poisson process (a Point process approach also known as POT-PP). Main details will be discussed here for each method, but as the last one is recommended in ~~Asce2017~~ ASCE 7-16, therefore a more indeed explanation will be provided ~~in POT-Poisson Process~~ for this process. ~~There is a whole section with the details about the background of this research, see Theoretical Framework~~

1.1.1 Sample Maxima

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

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

esta solo aplica si $\xi \neq 0$

significado?

que significa? Pertenece a H?

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

GEV was developed within extreme value theory to combine Gumbel, Fréchet and Weibull families (also known as type I, II and III)

1.1.2 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~~ generalized Pareto distribution approximation:

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

$\frac{y-\mu}{\sigma}$

Si se incluyen aqui tambien colocarlos en H

1.1.3 Poisson-GPD

If a rescale of the variable indexes above the threshold is performed, then the exceedances over threshold approach can be seen as a point process, ~~namely~~ i.e. an homogeneous Poisson Process where:

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

que es?? no se ha definido

Its cumulative distribution function *cdf* is:

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

significado

1.2 Problem Statement and Motvation

Wind forces are important for infrastructure design, ~~see~~ Comarazamy (2005). For a civil engineer, designer of different types of structures, main forces to consider when designing a structure, for instance a bridge or a building, are a) dead load due to the weight of the structure, and b) live load due to earthquake and wind. For ~~the study area, a developing country~~ Colombia, the structure design standard has defined in great detail, all aspects related to seismic forces, and dead forces, but lack of detail in ~~wind forces~~ design wind velocities. Actually, current map is 20 years outdated, and ~~it~~ is not appropriate for all types of structures, because it only includes two return periods. Additionally, it is well known that in recent years there have been accelerated changes in the climate of the planet, including issues related to winds, aspect that is reflected in frequent failures of structures due to wind forces, see Council (1994), as is stated in Rezapour & Baldock (2014), wind forces are able to completely destroy different types of infrastructures, reason why last five decades the way to assess wind loads in structural design has had remarkable changes, see Roberts (2012).

Demasiado "wordy" y redundante

Oraciones demasiado largas y con problemas de puntuación

A complete study of extreme wind forces, need to address separately ~~and using different scientific approaches, hurricane and non-hurricane data,~~ to ~~allow a~~ include in the final ~~research~~ product ~~as~~ the integration of ~~the~~ results ~~in~~ from both fronts ~~, see Engineers (2017)~~ This hurricane and non-hurricane data (ASCE, 2017). In the study area, hurricane winds are only present inland in the Caribbean Sea, therefore, only affects directly ~~to~~ 'San Andres y Providencia' island - one (1) of thirty-three (33) states. In 1102 of 1103 municipalities (more than 99%), ~~the issue of~~ only non-hurricane winds ~~is the only one~~ are relevant. ~~and in addition, this lacks recent studies and research,~~ However, all municipalities located near to the northern onshore border may be impacted by side effects of hurricanes.

As a note of clarification on the motivation to carry out the research, the author of this thesis is a civil engineer, from Colombia (the study area), and has developed previous research work with 'Universidad de los Andes', related to geoinformatics, and analysis and evaluation of natural risks. Due to the proximity of the University with the Colombian Association of Seismic Engineering - AIS, the opportunity to contribute to the update of the standard has arisen.

No se si se necesite este aclaración

1.3 Knowledge Gap

Nowadays, methodologies to deal with the inference of extreme wind maps are quite mature and advanced, and many of them are already implemented and ready for use, reason ~~why~~ for this the main contribution ~~in~~ this research is not related to the theoretical foundations of the methods themselves, but to application of the method in a particular case ~~in developing countries, where the lack of data plays a decisive role in achieving the results, see~~ where good quality data is not available (ADB (2014)). Thereby,

If ~~shape~~ the shape ξ and location μ are both zero, the GPD is equivalent to the exponential distribution