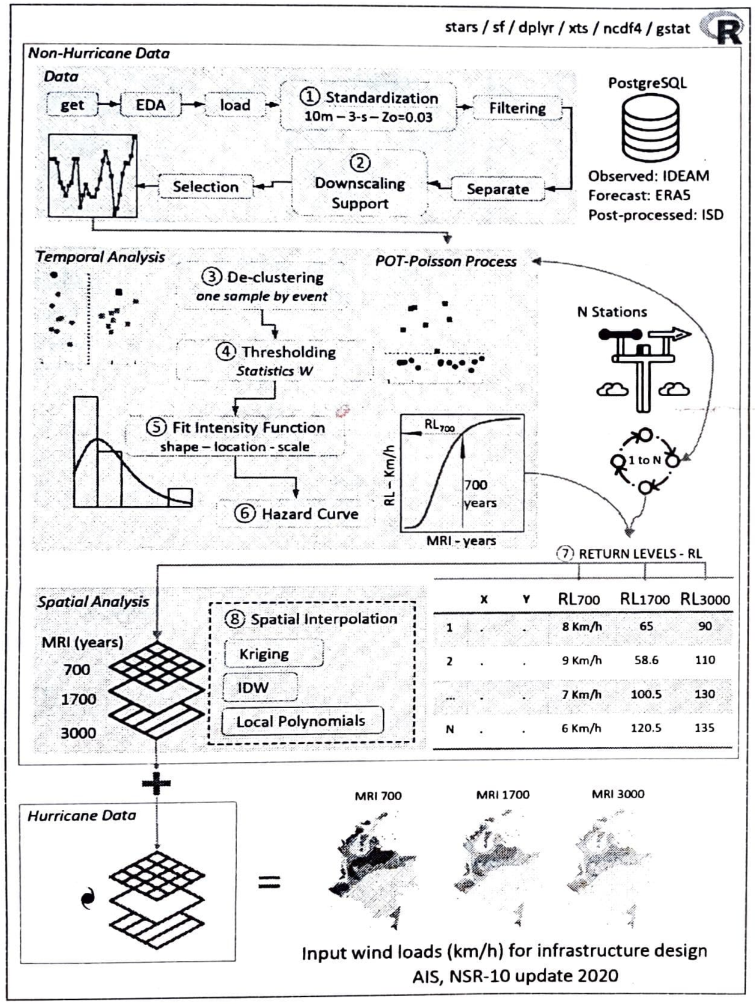

Figure 4.2: Methodology

Evitar el uso de guiones en todo el documento. Generan confusion porque se confunde con un signo menos

exposicion C se refiere a areas con algunas obstrucciones. Exposicion D si es espacio abierto perfecto.

4.1 Data Standardization

Analysis of extreme wind speeds requires data standardization as initial step. All input data must be standardized to represent three important conditions: a) anemometer height of 10 meters, b) open space terrain roughness, and c) averaging time of 3-seconds wind gust.

Parallel to the standardization activity described below, it is also important to consider for all stations involved in the analysis:

- *Separating*: As far as possible, identify each record of the time series, as thunderstorm (t) or non-thunderstorm (nt)
- *Filtering*: Remove wind speeds above 200 $\frac{Km}{h}$ and data pertaining to hurricane events, because the procedure with hurricane requires a different approach and need to be done independently

4.1.1 Anemometer Height (10 m)

According to the protocol for field data collection and location of methodological stations (IDEAM, 2005), the anemometer (wind sensor) in installed always to a fixed height of 10 meters from the surface, as is shown in Figure 4.3, ~~ergo~~, no height correction is needed.

km minuscula

therefore

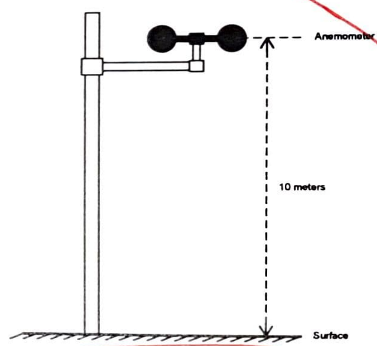

Figure 4.3: Anemometer height - 10 m

Explicar la que presenta la Figura

4.1.2 Surface Roughness at Open Terrain (0.03 m)

Due to the effects that the terrain has on wind speed, a correction should be applied if the station is located in a geographical space considered "not open terrain". When terrain is open, the roughness corresponds to 0.03 meters. There are some alternative methodologies to calculate the roughness, Masters, ~~Vickery, Bacon, & Rappaport~~ (2010) uses the station

et al

for example,

data, but the separation of the measurements should not exceed one minute (something difficult to obtain) ~~and~~ Lettau (1969) uses an empirical equation that is recommended in ~~Engineers (2017)~~ (page 743, equation C26.7-1), which was used here.

ASCE7-16

Sin cursiva

$$\mathit{Roughness} = z_0 = 0.5 * H_{ob} * \frac{S_{ob}}{A_{ob}}$$

signo x no * de hecho en este caso no se requiere ningún signo

w

Where H_{ob} is the average height of the obstacles, S_{ob} is the average vertical area perpendicular to the wind of the obstacles, and A_{ob} is the average area of the terrain occupied by each obstruction. ~~Then,~~ the empirical exponent α, gradient height z_g, and exposure coefficient K_z, ~~corresponding to equations C26.10-3, C26.10-4, and C26.10-1 si of Engineers (2017)~~, are used to calculate the correction factor $F_{exposition}$, ~~verifying that~~ z_0 units are in meters.

for

$$\alpha = 5.65 \times z_0^{-0.133}$$

no se requiere

$$z_g = 450 \times z_0^{0.125}$$

$$K_z = 2.01 \times \left(\frac{z}{z_g}\right)$$

$$F_{exposition} = \frac{0.951434}{K_z}$$

According to

~~Following~~ NIST (2012), calculation of roughness needs to be weighted according to the predominance of wind magnitude in eight directions (north, south, east, west, north-east, north-west, south-east, and south-west), see Figure 4.4, using a detailed aerial photo or satellite ~~image inside~~ a radius of 800 meters around the station location, as shown in Figure ??, with south direction highlighted.

No se entiende

Figure 4.4 shows....

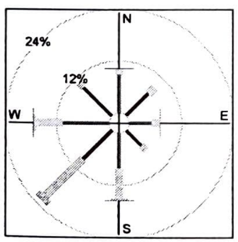

Figure 4.4: Wind rose with wind percentages in eight directions for a generic station.

Figure 4.5: Digital imagery for 'Vanguardia' ISD station (USAF:802340), located in 'Villavicencio' airport, with four (south, north, east, and west) 45 degree sectors highlighted. Radius of the circular zone is 800 meters

No se había mencionado antes. Difícil seguir el flujo de la lectura

Figure ?? shows extreme conditions for roughness, open space in left image (ISD Station 804070), closed space in center image (ISD Station 803000), and a typical example where Lettau procedure is needed. Lettau equation need to be applied to each direction and then the final z_o value is the weighted average, using historical wind percentage. See Figure 4.7 showing the strokes made to calculate the different areas for two Colombian stations. Information about wind percentage per direction at each station were obtained from IDEAM (1999).

Figure 4.6: Roughness values: 0.03 for open space (left), 0.1 for closed space (center), and areas where Lettau equation is needed because roughness is different in each direction (right).

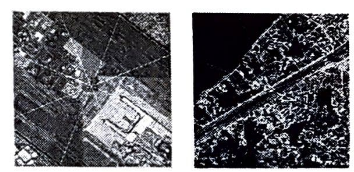

Figure 4.7: Lettau calculation. In red the area occupied by the obstacles, and in blue the perpendicular area. Source Triana (2019)

4.1.3 Averaging Time: 3-s Gust

ASCE7-16

of

To transform hourly mean wind velocity V_{3600}, to 3-s gust velocity V_3, ~~Engineers (2017)~~ recommends to use ~~C. S.~~ Durst (1960). ~~See Wind Loads Requirements.~~ As the axis x represents duration t of the gust, what is done is to look there for the value 3 seconds, and read the corresponding gust factor $\frac{V_t}{V_{3600}}$, ~~this is, the value in the axis y, then~~ (for this case

$$V_t = V_{3\,seconds} = (gust\,factor) \times V_{3600\,seconds}$$

It is valid only for open terrain 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 in seconds.

colocar el valor del factor

4.2 Downscaling Support

necessary

~~As it happens in this study.~~ Where it is intended to complement the local/regional wind analysis, with data from ISD (output data of a model for extreme winds), and ERA5 re-analysis dataset (large scale forecast data), it is required to probe by means of *comparisons* (exploratory data analysis and/or statistical measures) that those sources (modeled and forecast) are similar to IDEAM field measurements.

The proposed mechanism in the search ~~for~~ from downscaling support is, a) the creation of *common time series graphs*, where time series ~~overlays for~~ all data sources, are expected to be similar, and b) the elaboration of *scatter plots graphics*, which are generated matching two sources ~~by~~ in time (sorted in ascending order by wind velocity), ~~and that~~, ~~visually will allow to~~ evaluate ~~about~~ data similarity between ~~two~~ sources, when all the points fall very close to a 45 degree line. In both cases, the strategy for station matching, could be one of the following:

By visual inspection is possible to

1. *Manual matching*, doing a detailed analysis station by station (only for ISD and IDEAM). While it is true that ISD is based on IDEAM, their names and locations are somewhat different, for this reason, it is necessary to read information available from each source, and decide station by station, about its correspondence.
2. *Intersection matching*, between ISD and IDEAM point stations and ERA5 cells. All ISD and IDEAM stations falling inside a ERA5 cell, will be compared between them.

4.3 ~~Peaks Over Threshold - Poisson Process (POT-PP)~~ Temporal analysis

Similar to ~~how~~ how the adjustment of statistical data to a normal distribution ~~works in order~~ is done to make inferences ~~considering deviations from the mean~~, ~~here~~ only some part of the data (those that are extreme - over a high threshold - POT), need to be fitted to a PP considering extreme deviations from the mean. While in the first case (normal distribution) the inferences are for events similar to the samples, in this case, when working with extreme value theory, the

in extreme value analysis

inferences will be for more extreme events than any previously observed or measured. ~~In the theoretical framework section are described the main elements of POT-PP.~~

In summary, POT means only to work with extreme values, and PP means to adjust data to a ~~pdf~~ PDF, which depends on an intensity function $\lambda(t,y)$, where t is time, y is wind extreme velocity. As ~~it~~ shown in Figure 3.8, in a POT-PP approach with domain D, all the observations follow a Poisson distribution with mean $\int_D \lambda(t,y)\,dt\,dy$. Main advantage of POT-PP is that it is designed to consider storm and not-storm events independently (for each disjoint sub-domain D_1 or D_2 inside D, the observations in D_1 or D_2 are independent random variables), but in the end use them both for the inferences.

$$\text{pdf} = f(t, y|\eta) = \frac{\lambda(t,y)}{\int_D \lambda(t,y)\,dt\,dy} \tag{4.1}$$

4.3.1 Declustering

To make the assumptions of PP more justifiable, it is important to have only one sample per event, the highest one. For instance, if a hypothetical storm started at 11:30 in the morning and ended at 12:30 in the afternoon, and the time series for that event has thirty wind measurements (one each two minutes), it is necessary to leave only the stronger or maximum value; ~~and~~ this process is called de-clustering (~~see~~ Figure 4.8). POT-PP defines that all the adjacent observations separated by six hours (6) or less in the case of thunderstorm events, and four (4) days or less, in the case of non-thunderstorm events belong to the same cluster.

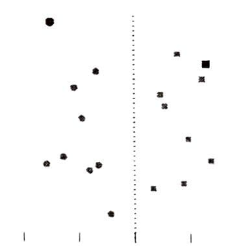

Figure 4.8: De-clustering in PP. Two thunderstorm clusters are shown. Separation between adjacent observations inside the clusters are always equal or less than six hours. Distance between the last event in the first cluster and the first event in the second cluster is larger than six hours. Only red samples are used to fit the PP, but in addition a POT (thresholding) process is needed

4.3.2 Thresholding

As the POT model requires to work only with the most extreme values in the time series, it is necessary to select a threshold to filter out small values. ~~Selection of threshold value imply two effects in the model~~. Bias is high when a low threshold is selected (many exceedances) because the asymptotic support is weak. Opposite situation happens for high thresholds where variance is potentially high, ~~so~~ according to Davison & Smith (1990), it is ~~needed~~ to select a threshold value, consistent with model structure.

necessary

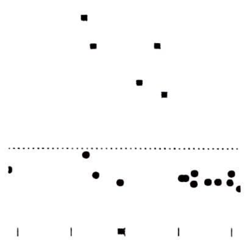

Figure 4.9: POT - Thresholding

Selection of the thresholds pairs, one for thunderstorm, and one for non-thunderstorm, is based in W transformation described in threshold selection section. W-statistic is done comparing the ordered empirical result of applying $W = -log(1-U)$ to the data axis y in Figure 4.10, with the theoretical quantiles of an exponential variable with uniform distribution between 0 and 1, axis x in same figure. ~~W-statistic is the highest vertical~~ distance between the 45° line and the points in the graphic. The best thresholds pairs returns the minimum value for W-statistics after testing, in an iterative process, with many threshold pairs combinations.

esta oración no tiene verbo
No se entiende

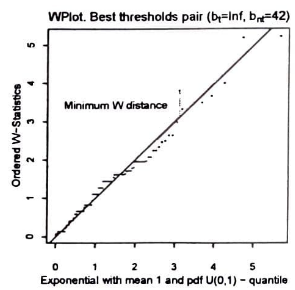

Figure 4.10: POT - Thresholding

4.3.3 Exclude No-Data Periods

PP requires to remove long periods of time when stations were not recording or failing. Proposed time in Pintar et al. (2015) is 180 days, namely, to remove all the gaps from the time series larger than six months.

4.3.4 Fit Intensity Function

Probability density function *pdf*, and cumulative distribution function *cdf*, of the PP, depend of the intensity function, and are shown in Equation (4.1), and Equation (3.9), respectively.

To facilitate the estimation of the parameters for the PP intensity function, parameter *shape* $= \zeta_t$ is taken to be zero in Equation (3.6), then doing the limit, the resulting intensity function is the same as the the GEV type I or Gumbel distribution,

$$\frac{1}{\psi_t} \exp\left\{\frac{-(y-\omega_t)}{\psi_t}\right\} \tag{4.2}$$

In this study, used intensity functions are ~~shown in next Equation (4.3)~~:

Sin cursiva

$$\lambda(y,t)\begin{cases}\frac{1}{\psi_s}\exp\left(\frac{-(y-\omega_s)}{\psi_s}\right), & for\ t\ in\ thunderstorm\ period\\ \frac{1}{\psi_{nt}}\exp\left(\frac{-(y-\omega_{nt})}{\psi_{nt}}\right), & for\ t\ in\ non-thunderstorm\ period\end{cases} \tag{4.3}$$

Las ecuaciones hacen parte del texto por lo tanto no es necesario referirse con su número. Esto solo se hace cuando están en otra sección.

As is shown in 4.11, the fitting process involve finding the best group of parameters of the intensity function, in such a way that the red curve (*pdf* of the PP, based in intensity function) be as tight as possible to the shape of the data histogram. As is described in POT-PP, optimal parameters to do the fitting process of the intensity function are calculated using *maximum likelihood*.

aquí se puede incluir la forma presentada antes

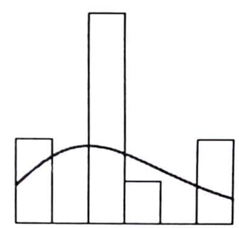

Figure 4.11: POT - PP intensity function fitting process

4.3.5 Hazard Curve ~~- Return Levels - RL~~

If Equation (3.8), Y_N is solved using estimated parameters of the intensity function, and a hazard curve is constructed as shown in Figure 4.12, where axis x represents annual exceedance probability $P_e = \frac{1}{N}$, and axis y represents the RL Y_N for the corresponding N-years return period, then it will be possible to have the extreme return wind velocity level for any given return period going from axis x to axis y through the curve.

Oraciones largas demasiado y difíciles de leer y entender.

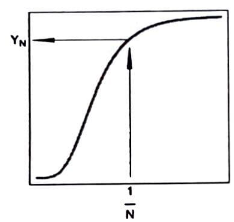

Figure 4.12: POT - PP fitting process

No está bien conectado con la oración anterior

~~Two alternatives approaches for RL~~

There is an equation that allows direct calculations of return levels, and also it is possible to use the quantile function of Gumbel when shape parameter equals to zero, but it is important to emphasize that Equation (4.4), and the use of Gumbel quantile function for RL calculations, is only valid when the analysis of POT-PP includes only one type of event (thunderstorm or non-thunderstorm), and the average estimated duration time of the event in a year is considered to be one (independent of the units in which time is processed), namely, the values for parameters A_t or A_{nt} of Equation (3.8) are equal to one.

Instead of solving Equation (3.8), next Equation (4.4) can be used replacing directly the PP parameters and the N return periods to create the hazard curve and get RL.

$$Y_N = \frac{\psi}{\varsigma}\left[-\log\left(\frac{N-1}{N}\right)\right]^{-\varsigma} - \frac{\psi}{\varsigma} + \omega \qquad (4.4)$$

As for this research $\varsigma = 0$, return levels Y_N can be calculated using the Gumbel quantile function, but using $(1 - \frac{1}{N})$ as probability.

Muy confuso. No se entiende. La ecuación se muestra tarde no cuando se menciona por primera vez.

4.4 Spatial Interpolation

Probabilistic (Kriging) and deterministic (IDW, local polynomials) techniques are used to create maps for return levels with same return period. Interpolation with Kriging requires verification of minimum procedures to ensure proper use of the method, ~~for instance~~.

No se entiende su relación con el problema específico.

- Structural analysis, which includes data normality check, for example, with Kolmogorov Smirnov or Shapiro Wilk goodness of fit tests, and if needed, data transformation to ensure data normality, e.g. using Box-Cox, and in addition, trend analysis to verify the need for trend modeling, in subsequent steps
- Semivariance Analysis: Use of available tools like cloud semivariogram, experimental semivariogram, directional semivariograms to verify isotropy or anisotropy, and different theoretical semivariograms, to ensure the best model of spatial autocorrelation, as a preliminary step to interpolation.
- Kriging Predictions: Use of different types of Kriging predictors, like simple, ordinary, universal, based on the results of the structural analysis.
- Cross Validation: Use of statistics like root mean square, average standard error, mean standardized, and root mean square standardized, that allow to measure the quality of the predictions and the magnitude of the errors.

Possible advantage of deterministic methods, is a better assessment of the local variability of spatial autocorrelation. It can also be considered with IDW or local polynomials a detailed assessment of structural analysis and cross validation. At the end of the spatial interpolation analysis all the predictions can be compared to select the most suitable result.

Main references in this research related to this matter using **R software** are E. Pebesma & Graeler (2019), Pebesma (2004), and Gräler, Pebesma, & Heuvelink (2016). For the implementation of spatial statistics using vector or raster format, see E. Pebesma (2019a), E. Pebesma (2019b), and Pebesma (2018).

4.5 Integration with Hurricane Data

ASCE 7-16 ~~Engineers (2017)~~ proposes the equation C26.5-2 for combination of statistically independent events, of non-hurricane and hurricane wind speed data.

$$P_e(y > Y_N) = 1 - P_{NH}(y < Y_N) \times P_H(y < Y_N) \qquad (4.5)$$

Where $P_e(y > Y_N)$ is the annual exceedance probability for the combined wind hazards, $P_{NH}(y < Y_N)$ is the annual non-exceedance probability for non-hurricane winds, and $P_H(y < Y_N)$ is the annual non-exceedance probability for hurricane winds.

- To understand Equation (4.5), it is important to remember that to calculate return level Y_N, for a given N-year return period, the exceedance probability $\frac{1}{N}$ of Y_N is calculated. Then, the non-exceedance probability for Y_N is $\left(1 - \frac{1}{N}\right)$. The procedure consist in the creation of a new hazard curve, calculating all $P_e(y > Y_N)$ values for different Y_N return levels, combining hazard curves from non-thunderstorm and thunderstorm data.

Equation (4.5) can be expressed in terms of only exceedance probabilities, $P_e = 1 - (1 - P_{nh}) * (1 - P_h)$, where P_{nh} is the the annual exceedance probability for non-hurricane winds, and P_h is the annual exceedance probability for hurricane winds. A graphical explanation of the procedure to calculate the combined P_e for the return level $30\frac{Km}{h}$, is shown in next

Figure 4.13. For each cell in the study area, it is necessary to calculate a new combined hazard curve, this is, all the P_e values corresponding all different return levels (~~see right table in~~ Figure 4.13).

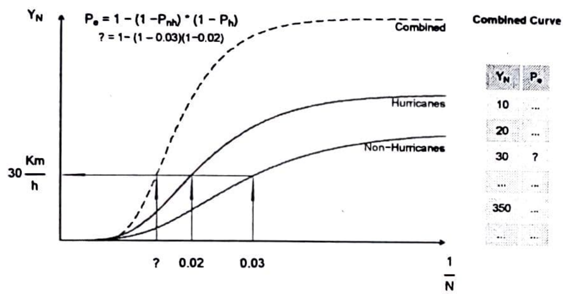

Figure 4.13: Integration Hurricane and Non-Hurricane Data

After the combined hazard curve is created, a new process of spatial interpolation need to be accomplished. In case of absence of hazard curves for stations, but availability of return levels maps, it becomes necessary to recreate hazard curves cell by cell, to apply Equation (4.5). In this case, ~~are required~~ as many maps as possible for different return periods~~, in order~~ to estimate detailed enough hazard curves from return level values (cell values).

are required

~~are required~~

Alex, esto es una oración de 7 renglones. Se debe partir por lo menos en tres.

Alex, cual es ~~su~~ la opinion del Prof. Pebesma con respecto a hacer un mapa sin huracanes y despues otro nuevo con huracanes?

No es mejor generar los datos brutos ~~y~~ al final hacer un solo mapa?

el sujeto va antes del verbo

No hay sujeto

Chapter 5

Results and Discussion

will be shown first,

In this section, ~~will be shown first~~, the data source comparison (post standardization process) to face the downscaling issue by using ERA5 and ISD database; then, the resulting process of fitting a POT-PP in the ISD station 801120, which includes revision of intensity function parameters, goodness of fit, hazard curve, return levels, and comparison with POT-GPD results, next, non-hurricane maps outputs, which includes results for ISD and ERA5 stations, after that, output maps combining hurricane and non-hurricane results will be displayed, and finally, a detailed discussion of the retults and future work is highlighted.

mejorar

5.1 Data Standardization and Downscaling Support

Looking for a statistical justification in the use of ISD (model) and ERA5 databases (forecast), as input data for this study, and considering the *downscaling approach* ~~described in the downscaling supportsection of methodology~~, data sources ISD and IDEAM were standardized to enable comparison. Standardization consisted of transforming the data to be equivalent to V_3 (3-s gust, 10 meters anemometer height, and open space roughness. In the comparison process, ~~for coincident stations by spatial location,~~ it was checked if the velocity values (standardized) in the three sources, for equal dates, were similar in magnitude.

presented previously

stations and

5.1.1 Data Standardization

None of the sources required anemometer height standardization. Lettau (1969) was used for roughness standardization of ISD and IDEAM, applying the method station by station. Gust velocities standardization was done using Durst curve~~, and in order~~ to obtain V_3 from Durst curve, it was required to start from V_{3600} (average hourly speed), or from a different wind gust speed, for instance V_5 (5-s gust)

For ERA5:

- Variable *10m wind gust - 10fg* of ERA5 data source does not need any standardization,

Falta redactar algo aquí,

No se entiende