

because it comes standardized from the source.

For ISD:

- Wind velocity from ISD comes from source as V_5 , that is, five seconds gust wind velocity. To standardize from V_5 to V_3 , using Durst curve, the correction factor is 1.03.
- Wind velocity V_5 from 74 ISD stations was standardized station by station, using procedure described in Surface Roughness at Open Space section, and Averaging Time 3-s Gust section.

For IDEAM:

- As the original variables obtained from IDEAM, do not represent gust speeds, it was necessary to start from *average hourly speed* V_{3600} , to obtain 3-s gust V_3 . To standardize from V_{3600} to V_3 , using Durst curve, the correction factor is 1.51.
- It was not possible to obtain the *average hourly speed* V_{3600} from IDEAM, see Table 2.2, but from *instantaneous wind velocity each 2 minutes - VV_AUT_2* it is possible to obtain a good estimator of V_{3600} , and from *instantaneous wind velocity each 10 minutes - VV_AUT_10* it is possible to obtain a poor estimator of V_{3600} .

5.1.2 Data Comparison

The available IDEAM data allowed two comparison processes, with quality data for few stations, and with low quality data, but available for all stations.

In both cases, to make the use of ISD and ERA5 viable, its time series are expected to be as similar as possible to IDEAM (field measurements). To verify this, two types of graphics were constructed:

- Time series overlay for the three sources. Not very effective method due to the large amount of data that makes the graphics unreadable.
- Scatter plot graphics comparing two different sources. Matching values by time, were sorted in ascending order, and put together on a scatter plot. The expected behavior in case of similarity in the data, is that all the points fall in a 45° line

IDEAM VV_AUT_2 - Quality Data Comparison

IDEAM VV_AUT_2 was available for twenty (20) stations, of which only twelve (12) were perfectly equivalent to ISD stations (see Table 5.1, and map in left panel of Figure 5.1). VV_AUT_2 dataset was transformed to V_{3600} (average hourly speed), averaging all 20 values available per hour. For twelve matching stations, wind velocity V_{3600} (transformation of VV_AUT_2), was standardized station by station, using procedure described in Surface Roughness at Open Space section and Averaging Time 3-s Gust, and finally, for the same twelve ISD and IDEAM standardized stations, a comparison was done against matching ERA5 stations (the corresponding cell in ERA5 that has within ISD and IDEAM locations).

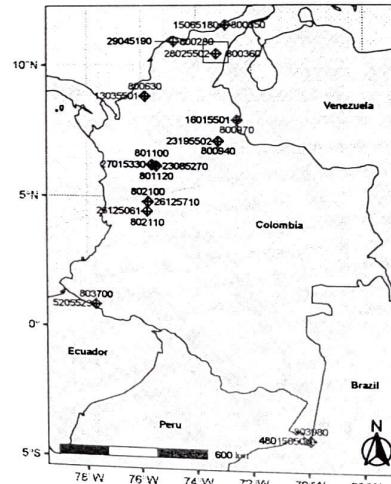
variable
instantaneous
each 2 minutes
(VV-AUT-2)

5.1. Data Standardization and Downscaling Support

Table 5.1: Quality Data Comparison

ISD ID	IDEAM ID	ERA5 ID, (col, row), [lon, lat]
803980	48015050	3320, (37, 68), [-70, -4.25]
803700	52055230	2309, (6, 48), [-77.75, 0.75]
802110	26125061	1582, (14, 33), [-73.75, 4.5]
802100	26125710	1533, (14, 32), [-73.75, 4.75]
801120	23085270	1240, (15, 26), [-73.5, 6.25]
801100	27015330	1240, (15, 26), [-73.5, 6.25]
800970	16015501	909, (27, 19), [-72.5, 8]
800940	23195502	1102, (24, 23), [-73.25, 7]
800330	13035501	749, (14, 16), [-70.75, 8.75]
800360	28025502	416, (24, 9), [-73.25, 10.5]
800350	15065180	221, (25, 5), [-73, 11.5]
800280	29045190	312, (18, 7), [-74.75, 11]

Quality Data Comparison
Twelve matching stations from IDEAM and ISD



Quality Data Comparison
Stations 28025502 from IDEAM, 800360 from ISD, and 416 from ERA5

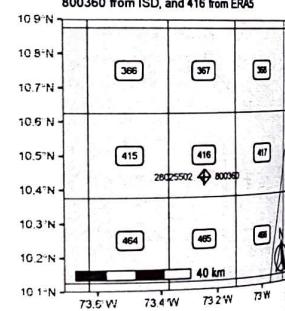


Figure 5.1: IDEAM VV_AUT_2 - Quality Data Comparison

The stations described in each row of the previous Table 5.1, were compared by generating scatter plots and common time series graphics. Stations 28025502 from IDEAM, 800360 from ISD, and 416 (cell with center point in -73.25° longitude, and 10.5° latitude) from ERA5

Que es? Velocidad
m/s?
Sera bueno unif.
undades de velocidad

5.1. Data Standardization and Downscaling Support

42

map in right panel of Figure 5.1, showed high correspondence, see Figure 5.2, because green regression line (empirical) is very similar to 45° line (theoretical). Unfortunately, in the other eleven stations, there was no high equivalence between sources.

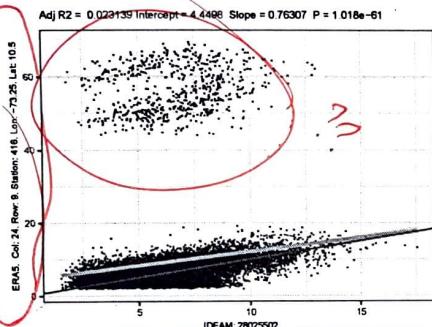


Figure 5.2: Quality Data Comparison. High similarity between sources

~~IDEAM VV_AUT_10 Non Quality Data Comparison (available in all IDEAM stations)~~

VV_AUT_10 was available for 204 stations, and despite that V_{3600} calculated from this source, is not an accurate or quality estimator, the standardization procedure was done to allow an additional comparison process, whose results are shown in the map displayed in Figure 5.3. Downscaling support was 'Good' comparing IDEAM and/or ISD stations with twenty-three (23) ERA5 stations (2261, 1971, 2066, 2020, 2260, 1875, 2213, 2637, 1442, 1583, 1501, 1582, 1381, 1493, 1485, 1397, 1338, 1055, 511, 1644, 515, 221, 1038), and 'Very Good' comparing IDEAM and/or ISD with five (5) ERA5 stations (265, 360, 78, 312, 416).

'Very Good' downscaling results for this non quality data comparison, are shown below:

- Table 5.2, shows in each row compared stations.
- Figure 5.4, shows an example of a very good time series plot for the ERA5 station 78 vs IDEAM stations 15075501 and 15079010.
- Figure 5.5, shows four different very good scatter plots, a) IDEAM 15015120 vs ERA5 265, b) IDEAM 29004520 vs ERA5 312, c) IDEAM 15079010 vs ERA5 78, and d) IDEAM 15075501 vs ERA5 78. Red line in each graphic represent the desired 45° line, where all points should fall, if the data sources would be exactly the same (theoretical behavior when there is equivalence of sources), and green line represents the linear regression line (empirical or real behavior when making the comparison).

No usar
bullets
para
la discusion

Las figuras deben tener claramente
identificados sus ejes.

5.1. Data Standardization and Downscaling Support

43

ERA5 grid, cells IDs from 1 to 3381 (49 cols, 69 rows)
ISD-IDEAM-ERA5 'poor data' comparison

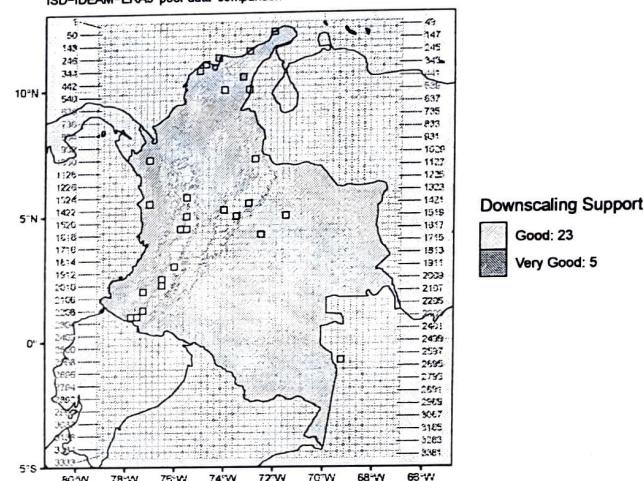


Figure 5.3: IDEAM VV_AUT_10 - Non Quality Data Comparison. Two different types of downscaling support: 'Good' and 'Very Good'

Table 5.2: Non quality data comparison. 'Very Good' downscaling support.

ISD ID	IDEAM ID	ERA5: ID, (col, row), [lon, lat]
NA	16015501	78, (29, 2), [-72, 12.25]
NA	15079010	78, (29, 2), [-72, 12.25]
NA	15075501	78, (29, 2), [-72, 12.25]
NA	15015120	265, (20, 6), [-74.25, 11.25]
NA	29004520	312, (18, 7), [-74.75, 11]
800280	29045190	312, (18, 7), [-74.75, 11]
NA	29045000	360, (17, 8), [-75, 10.75]
NA	28025502	416, (24, 9), [-73.25, 10.5]
800360	28035060	416, (24, 9), [-73.25, 10.5]

For some ISD stations in previous Table 5.2, the value 'NA' means that for the corresponding ERA5 and IDEAM station (same row), there is not a ISD station located inside the ERA5 cell space ($0.25^\circ \times 0.25^\circ$).

5.1. Data Standardization and Dowscale Support

44

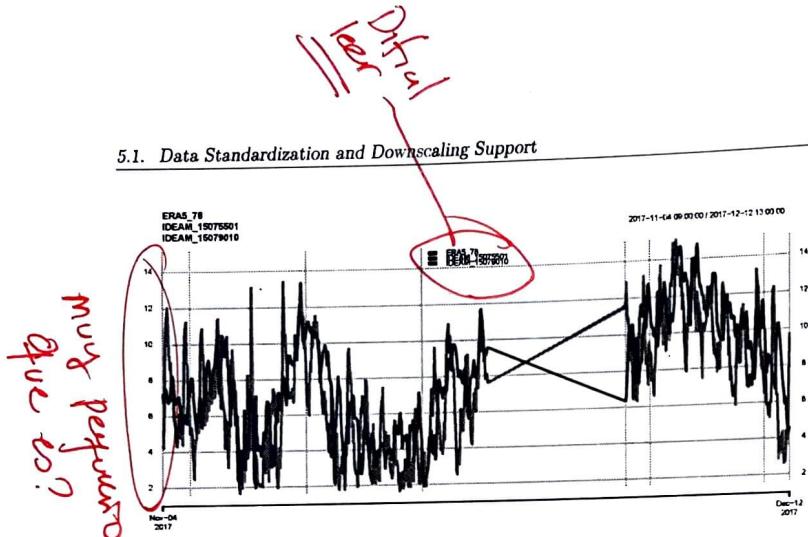


Figure 5.4: Non Quality Data Comparison. Time Series Graphic for 'Very Good' Dowscale Support

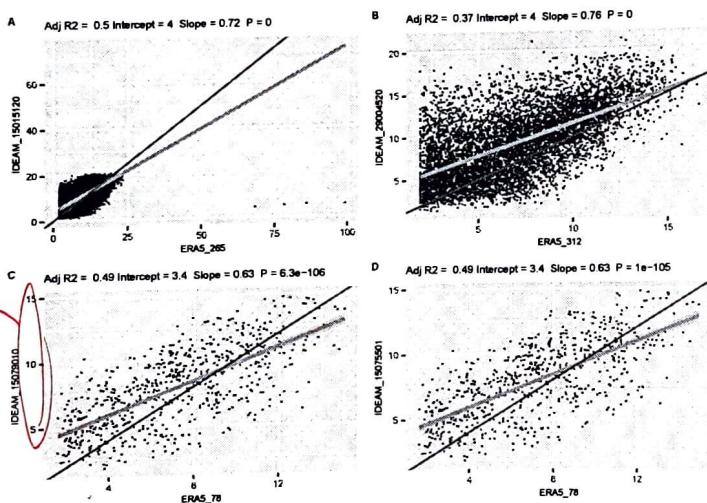


Figure 5.5: Non Quality Data Comparison: Scatter plots for 'Very Good' Dowscale Support

5.2. POT-PP for ISD Station 801120

5.2.1 POT-PP for ISD Station 801120

Figure 5.6 shows the satellite image (source Google Earth) of ISD station 801120, located in the international airport 'José María Córdova', municipality of Rio Negro (Antioquia, Colombia), with latitude 6.125° , and longitude -75.423° WGS84 coordinates. Red circles represent an influence radius of 800 meters. Table 5.3 shows different calculations related to correction factors applied to this station, using procedure described in sections Surface Roughness at Open Space, and Averaging Time 3-s Gust.

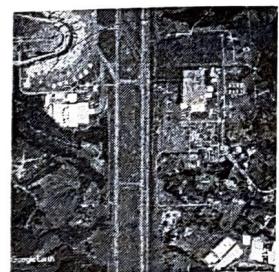


Figure 5.6: Location of ISD station 801120

Table 5.3: Corrections factors for ISD station 801120

Variable	Value
Roughness - Z_o	0.05
Empirical exponent - α	8.38
Gradient height - z_g	310.56
Exposure coefficient - K_z	0.88
$F_{exposition}$	1.07
Gust factor for V_3	1.03

5.2.1 Raw Data, De-clustering, and Thresholding

As storm information is not available for any of the data sources, all the data for the station was classified as *non-thunderstorm*. According to POT-PP method described in methodology, the first process applied to original time series -raw data-, is declustering, and then thresholding.

Non-thunderstorm raw data for ISD station 801120 has 2931 records, from 1986-12-12 00:00 to 2019-03-01 12:00:00, corresponding to a total amount time in days of 11730, and to an average number of events per year of 18.9, which means that the average duration of an event is 19.3 days (average size in days of a cluster). After declustering, and thresholding processes, the number of records decreases to 181. Time series graphics are shown in

¿Qué pasa si se usan los datos de ERA?

Si se reproduce la estación de Pintar se obtienen los mismos resultados?

Figure 5.7, showing the data before (left) and after (right) applying the mentioned processes. Detailed yearly statistics are reported in Table 5.4, also including summary for before (left), and after (right).

Table 5.4: Yearly statistics of raw data and declustered data for ISD station 801120

Year	Raw Data				Declustered Data			
	Count	Mean	Min	Max	Count	Mean	Min	Max
1986	63	45.2	27.9	163.3	7	106.4	43.8	163.3
1987	192	36.1	26.7	87.6	10	61.0	45.0	87.6
1988	234	43.8	26.7	90.4	23	64.2	45.0	90.4
1989	256	44.2	27.9	103.6	19	64.4	45.0	103.6
1990	250	44.9	26.7	103.6	21	67.2	45.0	103.6
1991	149	38.7	26.7	127.5	20	58.6	45.0	127.5
1992	126	35.2	26.3	81.7	9	52.6	43.8	81.7
1993	109	36.3	26.3	79.7	13	53.5	43.8	79.7
1994	124	36.8	26.7	79.7	12	56.1	45.0	79.7
1995	89	33.3	26.7	111.5	2	77.7	43.8	111.5
1996	70	35.6	26.7	87.6	6	65.7	43.8	87.6
1997	71	36.6	26.7	119.5	4	86.9	49.0	119.5
1998	65	33.8	27.9	61.4	2	54.6	47.8	61.4
1999	48	31.7	26.7	47.8	1	47.8	47.8	47.8
2000	69	33.4	26.7	87.6	3	68.3	55.8	87.6
2001	62	29.9	26.7	39.8	0	NA	NA	NA
2002	94	33.3	26.7	71.7	5	54.2	43.8	71.7
2003	78	31.5	26.7	71.7	1	71.7	71.7	71.7
2004	60	31.9	26.7	51.8	2	48.4	45.0	51.8
2005	59	33.3	26.7	94.4	2	69.1	43.8	94.4
2006	55	32.6	26.7	164.1	1	164.1	164.1	164.1
2007	25	29.8	26.7	39.0	0	NA	NA	NA
2008	13	36.1	26.7	96.4	1	96.4	96.4	96.4
2009	36	31.6	26.7	82.1	1	82.1	82.1	82.1
2010	31	43.0	27.9	119.5	8	83.0	61.4	119.5
2011	32	29.2	26.7	41.0	0	NA	NA	NA
2012	82	31.9	26.7	87.6	4	64.5	43.0	87.6
2013	91	29.7	26.7	37.0	0	NA	NA	NA
2014	95	30.1	26.7	47.8	1	47.8	47.8	47.8
2015	129	30.3	26.7	51.8	1	51.8	51.8	51.8
2016	33	30.7	26.7	87.6	1	87.6	87.6	87.6
2017	18	31.3	26.7	67.7	1	67.7	67.7	67.7
2018	22	31.0	26.7	39.8	0	NA	NA	NA
2019	1	28.7	28.7	28.7	0	NA	NA	NA

It can be seen in the Table 5.4, that de-clustered data has zero records for some years. This situation is due to that all the data for each one of those years (2001, 2007, 2011, 2013, 2018, and 2019), belonged to a cluster that started the previous year, or finished the next year, and the unique chosen maximum value (the value representative for the cluster) was found in the previous or next year, but not in mentioned years of zero records.

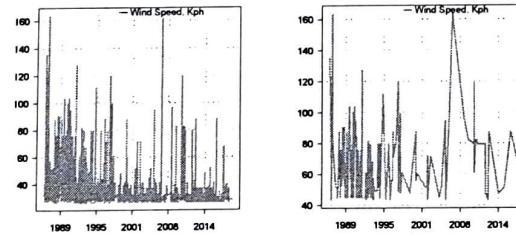


Figure 5.7: Non-Thunderstorm Time Series for ISD station 801120.
Left: Raw Data. Right: De-clustered Data

divide
various
or fractions
42
minutes
km/h

Using de-clustered data, and considering that it is only necessary to calculate optimal threshold for non-thunderstorm data, because there is no records classified as thunderstorm in any data source, many non-thunderstorm thresholds were tested, to choose the best one using the W statistic, as described in section thresholding of the methodology. Figure 5.8 shows a very good fit in resulting W-Statistic plot, for optimal non-thunderstorm threshold $b_{nt} = 42 \frac{km}{h}$, with a minimum W distance of 0.47, for ISD station 801120, where empirical values (black points) are very close or similar to theoretical values (red line).

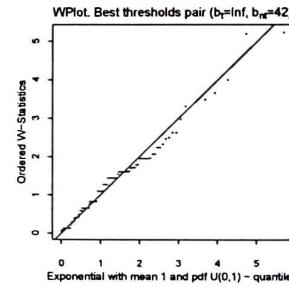


Figure 5.8: POT - Thresholding

5.2.2 Fitted *pdf* and *cdf*, and Goodness of Fit

Equation (3.6), defined in section POT-PP of the methodological framework, was used as intensity function $\lambda(t, y) = \lambda_{nt}(y)$ for POT-PP. When shape ζ_t is equal to zero (as it is in this study), an equivalent intensity function is described in Equation (4.3) defined in terms of the parameters location (ω_t), and scale (ψ_t). Related *pdf* and *cdf* functions are referenced

in Equations (4.1), where the domain D constraint the data above the threshold b , and the time to a non-thunderstorm period, and (3.9) respectively.

- Intensity function: $\frac{1}{\psi_{nt}} \exp\left(-\frac{(y-\omega_{nt})}{\psi_{nt}}\right)$
- pdf: $f(t, y) = \frac{\lambda(t, y)}{\int_b^Y \lambda(y, t) dy}$
- cdf: $F(t, y) = P(y \leq Y) = \frac{\int_b^y \lambda(y, t) dy}{\int_b^Y \lambda(y, t) dy}$

After fitting the intensity function to the domain D , the resulting parameters for ISD station 801120, are location ω_t equal to -55.62, and scale ψ_t equal to 23.4. Figure 5.9 shows the histogram and fitted pdf in panel A, Q-Q plot (theoretical quantiles against empirical ones) in panel B, empirical cumulative distribution against fitted cdf in panel C, and P-P plot (theoretical probabilities against empirical ones) in panel D. In all four panels, it can be seen that there is a very good visual correspondence between empirical data (points and histogram) and theoretical adjustment (lines).

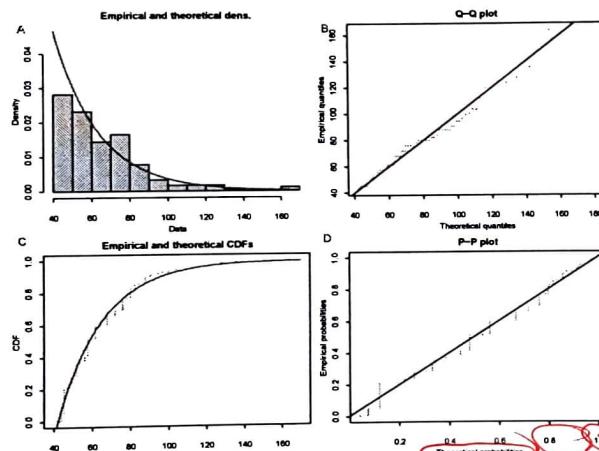


Figure 5.9: Graphic Diagnosis Of Goodness of Fit. Station 801120

Results of formal goodness of fit statistics for 'Kolmogorov-Smirnov D', 'Cramer-von Mises T' and 'Anderson-Darling A' are 0.089, 0.21, and 1.68 respectively. For a proposed null hypothesis, which indicates that the data conforms to a POT-PP, all resulting p-values using statistics D, T and A, confirm that there is no statistical evidence to ~~H0~~ reject stated hypothesis. Resulting p-value for statistic D is 0.11. Another available criteria to measure

5.2. POT-PP for ISD Station 801120

the quality of the fitted process are 'Akaike's Information Criterion', and 'Bayesian Information Criterion', with values 1505.2, and 1508.4 respectively. The Root Mean Square Error (RMSE), calculated using theoretical versus empirical cdf, is 0.023.

5.2.3 Hazard Curve and Return Levels - RL

Hazard curve is the solution to Equation (3.8), but eliminating from it elements related to thunderstorms, the equation is simplified to $A_{nt} \int_{Y_N}^{\infty} \lambda_{nt}(y) dy = \frac{1}{N}$, where A_{nt} is the average time of non-thunderstorm events by year, and Y_N is the return level or extreme wind velocity corresponding to the N-years return period or MRI. Replacing in this equation the intensity function λ_{nt} , and solving Y_N for all possible values of MRI, will provide hazard curve displayed in Figure 5.10.

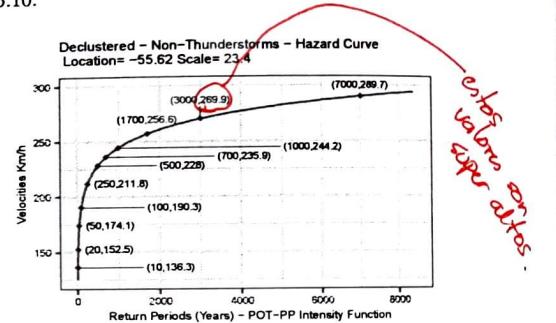


Figure 5.10: Hazard Curve. Station 801120

Table 5.5: Return Levels -RL for typical Mean Return Intervals - MRI. ISD station 801120

MRI	Return Level	Velocidad en km/h
10	136.30	
20	152.48	
50	174.10	
100	190.32	
250	211.76	
500	227.98	
700	235.85	
1000	244.20	
1700	256.61	
3000	269.90	
7000	289.73	

Ahí, sería bueno mirar en esta estación que pasa si al azar definimos ~~a la mitad~~ de los datos como tormenta. Solo para ver el efecto sobre el resultado.

Return levels of interest for this research, correspond to 700, 1700 and 3000 years of MRI, however, due to the mechanism of integration with existing hurricane study information, described in methodology, it is necessary to extract for all stations values related to typical return periods, as shown in the Table 5.5.

5.2.4 Comparison with POT-GPD and Common Extreme Value Distributions

To enable a comparison between, a) POT-PP (previous section), b) POT-GPD, and c) the fitting process of common extreme value distributions (GPA, GEV, GUM) without using POT method, ~~that is~~, using the generic concept of hazard function hf (see theoretical framework), a whole automation process was done to calculate return levels and errors using mentioned alternatives, bearing in mind that in all cases *maximum likelihood* was used to calculate the parameters.

Resulting return levels and errors for POT-GPD, using R packages extRemes - Gilliland (2019), ismev - Janet E. Heffernan with R port & Alec G. Stephenson. (2018), evd - Stephenson (2002), Renext - Deville & IRSN (2016), evir - Pfaff & McNeil (2018), and fExtremes - Wuertz. Setz, & Chalabi (2017), are reported in Table 5.6. Similarly are shown in Table 5.7, return levels calculated from the adjustment of the probability distributions GPA, GEV, and Gumbel.

Table 5.6: POT-GPD. Return Levels in Kph km/h

PACKAGE	RETURN LEVELS FOR TYPICAL MRIs										RMSE	
	10	20	50	100	250	500	700	1000	1700	3000		
extRemes	155.6	169.3	187.2	200.4	217.6	230.3	236.4	242.8	252.2	262.1	276.6	0.057
ismeV	155.5	169.3	187.1	200.4	217.5	230.1	236.2	242.6	252.0	261.9	276.4	0.057
evd	155.6	169.3	187.2	200.4	217.6	230.3	236.4	242.7	252.2	262.1	276.6	0.057
Renext Renouv	155.6	169.3	187.2	200.4	217.6	230.3	236.4	242.7	252.2	262.1	276.6	0.057
evir	155.0	168.5	185.4	198.4	215.1	227.3	233.1	239.2	248.2	257.6	271.3	0.058
fExtremes	155.5	169.3	187.2	200.4	217.5	230.2	236.3	242.6	252.0	261.9	276.5	0.057
Renext 2 parameters	200.8	203.9	206.5	207.8	208.9	209.4	209.6	209.7	209.9	210.1	210.3	0.337

Table 5.7: Common Extreme Value Distributions. Return Levels in Kph km/h

EVD	RETURN LEVELS FOR TYPICAL MRIs										RMSE	
	10	20	50	100	250	500	700	1000	1700	3000		
NAME	10	20	50	100	250	500	700	1000	1700	3000	7000	
gpa Generalized Pareto	149.6	160.6	174.2	183.9	195.8	204.2	208.2	212.2	218.0	223.9	232.2	0.018
gev Generalized Extreme Value	172.5	198.8	239.2	274.8	329.5	377.8	403.5	432.7	479.9	536.0	631.7	0.058
gum Gumbel	140.9	152.1	167.0	178.2	193.0	204.3	209.7	215.5	221.1	233.3	247.0	0.067

es bueno usar
siempre km/h para
velocidades en todo
el documento,

(¿Hay resultados de
ERA5?)

Sería bueno
contextualizar con
las que se han usado
a nivel mundial para
estos tipos de análisis.

5.3 Wind Maps

Using calculated return levels in all available ISD stations, continuous maps covering the study area, were created using spatial interpolation techniques as described in methodology. As calculated return levels for ERA5, represent predefined square cells of 0.25° decimal degrees, no interpolation process was necessary in this reanalysis dataset.

5.3.1 Existing Hurricane Maps

The Colombian consulting firm Ingeniar Ltda, following the methodology described in CIMNE (2015), and CIMNE (2017), has provided raster wind maps for return periods 10, 20, 50, 100, 250, 500, 700, 1000, 1700, 3000, and 7000 years, resulting from the probabilistic study of winds due to hurricanes in the Colombian Caribbean Sea and the surrounding continental zone. Figure 5.11 shows three of mentioned maps.

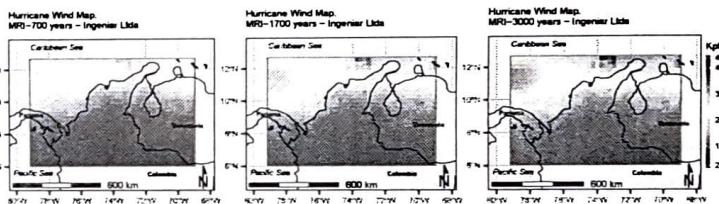


Figure 5.11: Ingenier Hurricane Wind Maps.

5.3.2 Non-Hurricane Maps

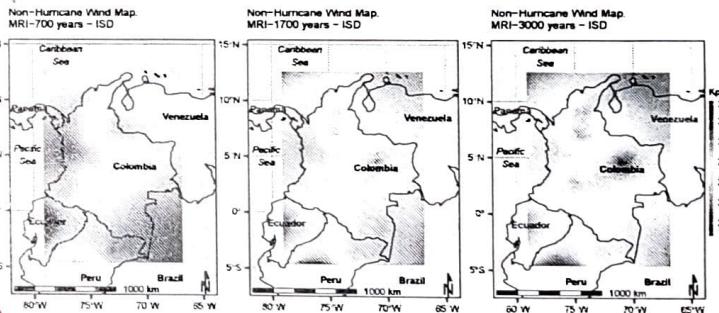


Figure 5.12: ISD Non-Hurricane Wind Maps.

5.3. Wind Maps

52

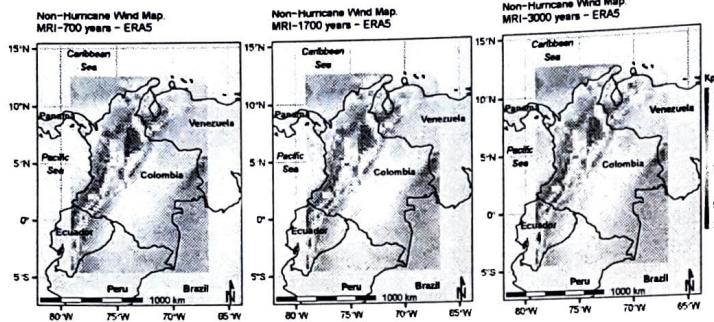


Figure 5.13: ERA5 Non-Hurricane Wind Maps.

5.3.3 Combined Maps

Following the procedure described in integration with Hurricane data, final wind maps are created, combining existing data for hurricane studies, and non-hurricane maps produced in this study.

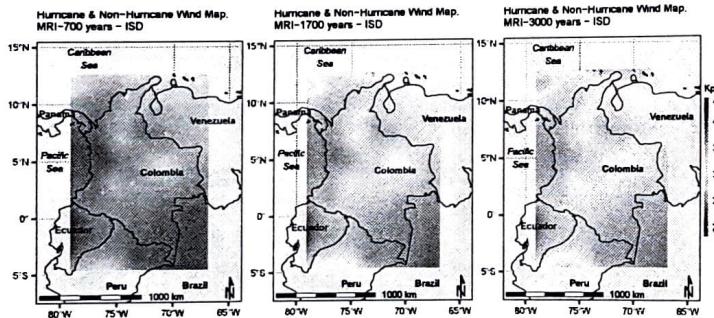


Figure 5.14: ISD Hurricane & Non-Hurricane Wind Maps.

Coto
se puede
simplificar
cálculos
después.

5.4. Final Discussion and Future Work

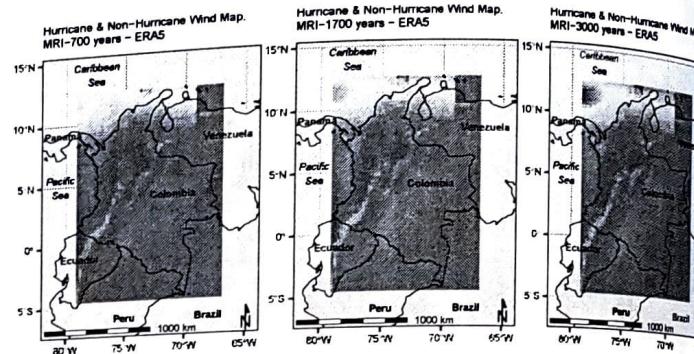


Figure 5.15: ERA5 Hurricane & Non-Hurricane Wind Maps.

5.4 Final Discussion and Future Work

Regarding the comparison of the data, it must be remembered that the basis of comparison is, that is, the one that represents the truth in the field - IDEAM field measurements, were fully available, what disturbed the process since before starting it. On the other hand, there are many uncertainties with respect to the model that represents the ISD database, because first, the available documentation does not specify whether it is an average or a gust variable, second, the comparative graphs showed that ISD database did not represent a continuous variable (vertical or horizontal stripes in scatter plots), and finally, the comparisons against IDEAM never showed good results.

With respect to ERA5 database, although the comparative results showed greater similarity, it should be remembered that each record in the time series does not represent a point value, on the contrary, it represents a square cell of 0.25 decimal degrees. The IDEAM station with which the comparison was made, can fall into any location of the cell, and consider only a very local condition, that is not represented by an averaged forecast for the whole cell size. This considering that Colombia is a tropical region with a widely diverse terrain (mountains, plains, rivers, forests, etc) and climate. So the possible similarity between IDEAM and ERA5 is limited by this condition.

The main difference between POT-PP and POT-GPD is that in the latter, wind quantiles are adjusted to a GPD, and the time is adjusted to a Poisson Process (1D), while in the former, time and magnitude are adjusted to a Poisson Process (2D). If there is no weather classification available (storm and no storm) in the wind time series, POT-PP loses advantages and resembles in potential and scope to POT-GPD, because the intensity function, varying only in magnitude, becomes similar to a GPD. For this reason, POT-PP method is really useful, if historical classification of time series is available, record by record, in stochastic and non-stochastic.

This classification by time of historical series, is useful because it allows to define more precisely the average rate of events per year (Poisson process rate), which in POT-PP is represented by the average amount of events time per year, this is, components A_t and A_{nt} of Equation (3.8) - $A_t \int_{Y_N}^{\infty} \lambda_t(y) dy + A_{nt} \int_{Y_N}^{\infty} \lambda_{nt}(y) dy = \frac{1}{N}$ -, used to calculate return levels Y_N .

By the lack of thunderstorm and non-thunderstorm information, is impossible to calculate which part of the annual time belongs to storm and which to ~~not~~ storm. As the available data were all assumed as non-thunderstorm data, this average time of events per year will always result in a fixed wrong value of 365 days, the maximum possible value. For ISD, this condition is reflected in high and unlikely final results.

However, this same condition did not affect the results in the ERA5 database in the same way. Although ~~also~~ in ERA5, all the data were classified as "non-thunderstorm", and the average time of events was always 365 days, an additional condition made the final result more realistic. Contrary to what happens in the ISD database, where time series have many gaps and there is lack of information, used ERA5 database has the full time series, hour by hour, from 1979 to 2019. Following the theory behind POT-PP, this implies that there is only one cluster for the whole time series, which would leave a single data after the de-clustering, canceling the entire subsequent process. Our proposed solution was to work only with one data per week, the maximum, which implies that the de-clustering process will have no effect, since it is based on 4-day gaps, see Declustering in Methodology section, resulting in more events above the threshold, exactly 48 events for each ~~one~~ of the 40 years of history, which translates into greater averaging of the final wind data.

One of the objectives of the investigation was to compare different methods in the calculation of return levels, and that was achieved using in all cases, both POT-GPD, and POT-PP. The Poisson for POT-GPD, does not accept data classification by time (storm or non-storm), because it is one-dimensional and data must represent a single general type of event: wind. Despite that, it is important to emphasize that the shortcomings in the calculation of the Poisson process rate, similarly affected the application of both methods, so in all cases, the results have the same limitation, and the use of POT-GPD does not represent best quality in the results.

To improve the quality of extreme wind analysis in future research, the inclusion of seasonal effects is recommended, this can be done in two ways, first, using a separate POT-PP for each season, and second, model the Poisson process parameters (location, scale and shape) as sinusoidal functions of time. Finally, it is possible to include more formal statistics (not only graphics), to face the downscaling challenge.

Chapter 6

Conclusions

Final maps using ERA5 forecast database, see Combined Maps in Results and Discussion section, representing return periods of 700, 1700, and 3000 years, are the extreme velocities needed as input load for the design of structures of different use category in the study area. Nevertheless, by one hand, full data from the IDEAM source is needed to enable the validation of downscaling support, on the other hand, it is essential to include in the study the classification of thunderstorm and non-thunderstorm data to achieve more realistic results, and finally, an additional conservative calibration process is needed, where to each municipality is assigned only a wind velocity, in order to define final values that will be part of the structure design norm.

Other general conclusions of the investigation are:

- In the absence of wind field measurements, alternative data sources as ISD and ERA5 can be a viable source of data to calculate extreme wind events, but always must be searched for statistic or graphic support for the downscaling issue, and at the end a process of calibration is needed for each particular case.
- A powerful R tool is available to do extreme value analysis using POT-PP and POT-GPD approaches
- Results of this research could be the starting point of a process to update wind maps in countries with information issues.
- Output results of this research will contribute to reduce the risk of infrastructure collapse, representing a favorable impact in human lives, material losses, and disaster prevention.

For a detailed analysis of the results, refer to Results and Discussion section, and for a discussion about the project and its relevant topics, refer to Final Discussion

Paper para una mejor
discusión con
soluciones con
"ballets"

oración demusicalizada

Unificar formato de referencias.

Appendix E

User Manual

One complete example will be shown here to run the entire process from the beginning (data standardization), to the end (wind map creation and final map creation), using ISD station X.

References

- ADB. (2014). *Guidelines for wind resource assessment: Best practices for countries in wind development*. Asian Development Bank. Retrieved from https://www.ebook.de/product/30686652/guidelines_for_wind_resource_assessment.html
- Beirlant, J., Goegebeur, Y., Teugels, J., & Segers, J. (2004). *Statistics of extremes: Theory and applications*. John Wiley & Sons, Ltd. <http://doi.org/10.1002/047002382>
- CIMNE, I. (2015). *Update on the probabilistic modelling of natural risks at global level: Global risk model* (technical report). The United Nations Office for Disaster Risk Reduction - UNISDR. Retrieved from <https://www.preventionweb.net/english/hazard/2015/en/bdocs/CIMNE-INGENIAR,%202014a.pdf>
- CIMNE, I., ITEC. (2017). *Metodología de modelación probabilista de riesgos naturales* (technical report No. ERN-CAPRA-T1-3). CAPRA- Probabilistic Risk Assessment Initiative. Retrieved from <https://ecapra.org/sites/default/files/documents/ERN-CAPRA-T1-3%20Modelos%20de%20Evaluaci%C3%B3n%20de%20Amenazas.pdf>
- Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer London. <http://doi.org/10.1007/978-1-4471-3675-0>
- Comarazamy, D. (2005). *Disaster mitigation in health facilities. Wind effects. Structural issues*. Pan American Health Organization. Retrieved from <http://www.paho.org/english/guiones/structural.pdf>
- Council, N. R. (1994). *Hurricane Hugo, Puerto Rico, the Virgin Islands, and Charleston, South Carolina, September 17–22, 1989*. In (pp. 247–257). Washington, DC: National Academy Press. <http://doi.org/10.17226/1993>
- 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>
- Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(3), 393–442. Retrieved from <http://www.jstor.org/stable/2345667>
- Deville, Y., & IRSN. (2016). *Renext: Renewal method for extreme values extrapolation*. Retrieved from <https://CRAN.R-project.org/package=Renext>

revisar

ASCE/SEI

Engineers, A. S. 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

European Centre For Medium-Range Weather Forecasts. (2017). ERA5 reanalysis. UCAR/NCAR - Research Data Archive. <http://doi.org/10.5065/D6X34W69>

Gilleland, E. (2019). *ExtRemes: Extreme value analysis*. Retrieved from <https://CRAN.R-project.org/package=extRemes>

Gräler, B., Pebesma, E., & Heuvelink, G. (2016). Spatio-temporal interpolation using gstat. *The R Journal*, 8(1), 204–218. Retrieved from <https://journal.r-project.org/archive/2016/RJ-2016-014/index.html>

Haigh, I. D., & Wahl, T. (2019). Advances in extreme value analysis and application to natural hazards. *Natural Hazards*, 98(3), 819–822. <http://doi.org/10.1007/s11069-019-03628-x>

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>

IDEAM. (1999, June). Aeronautical information. Annual wind regime. Web Page. Retrieved from <http://bart.ideam.gov.co/cliciu/rosas/viento.htm>

IDEAM. (2005). *Protocolo toma de datos de campo y emplazamiento de estaciones meteorológicas*.

Janet E. Heffernan with R port, O. S. functions written by, & Alec G. Stephenson., R documentation provided by. (2018). *Ismev: An introduction to statistical modeling of extreme values*. Retrieved from <https://CRAN.R-project.org/package=ismev>

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

Lettau, H. (1969). Note on aerodynamic roughness-parameter estimation on the basis of roughness-element description. *Journal of Applied Meteorology*, 8(5), 828–832. [http://doi.org/10.1175/1520-0450\(1969\)008%3C0828:NOARPE%3E2.0.CO;2](http://doi.org/10.1175/1520-0450(1969)008%3C0828:NOARPE%3E2.0.CO;2)

Masters, F. J., Vickery, P. J., Bacon, P., & Rappaport, E. N. (2010). Toward objective, standardized intensity estimates from surface wind speed observations. *Bulletin of the American Meteorological Society*, 91(12), 1665–1682. <http://doi.org/10.1175/2010bams2942.1>

Ministerio de Vivienda, C. y T. (2010). *Reglamento Colombiano de Construcción Sísmica*

33

NSR

Resistente - nsr 10. Carrera 20 # 84-14, oficina 502, Bogotá.: Asociación Colombiana de Ingeniería Sísmica. Comisión Asesora Permanente.

NIST. (2012, February). Standardized extreme wind speed database for the United States. Web Page. Retrieved from https://www.itl.nist.gov/div898/winds/NIST_TN/nist_tn.htm

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1), 439–446. <http://doi.org/10.32614/RJ-2018-009>

Pebesma, E. (2019a). *Sf: Simple features for r*. Retrieved from <https://CRAN.R-project.org/package=sf>

Pebesma, E. (2019b). *Stars: Spatiotemporal arrays, raster and vector data cubes*.

Pebesma, E., & Graeler, B. (2019). *Gstat: Spatial and spatio-temporal geostatistical modelling, prediction and simulation*. Retrieved from <https://CRAN.R-project.org/package=gstat>

Pebesma, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Computers & Geosciences*, 30, 683–691.

Pfaff, B., & McNeil, A. (2018). *Evir: Extreme values in r*. Retrieved from <https://CRAN.R-project.org/package=evir>

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 MeasurementMaps of non-hurricane non-tornadic wind speeds with specific mean recurrence intervals for the contiguous united states using a two-dimensional poisson process extreme value model and local regression results*. National Institute of Standards; Technology.

Rezapour, M., & Baldock, T. E. (2014). Classification of hurricane hazards: The importance of rainfall. *Weather and Forecasting*, 29(6), 1319–1331. <http://doi.org/10.1175/waf-d-14-00014.1>

Roberts, S. (2012). *Wind wizard: Alan G. Davenport and the art of wind engineering*. Princeton University Press. Retrieved from <https://books.google.de/books?id=e2eYDwAAQBAJ>

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, A., Lott, N., & Vose, R. (2011). The integrated surface database: Recent developments and partnerships. *Bulletin of the American Meteorological Society*, 92(6), 704–708. <http://doi.org/10.1175/2011bams3015.1>

Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to