

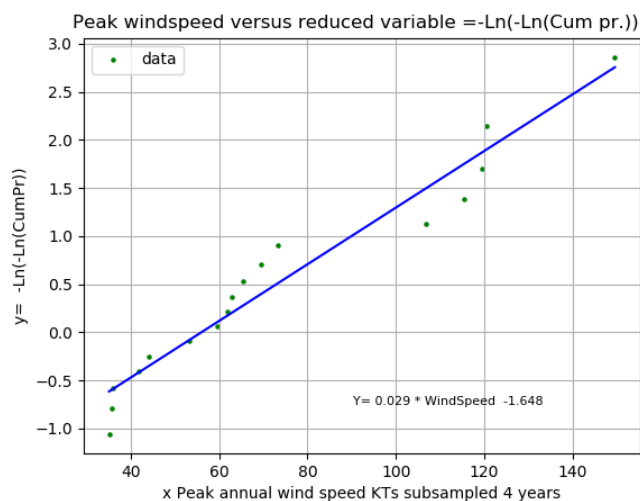
Description for the Tropical Cyclone recurrence interval programs written in python by John Rupp.

These programs and data are a current revisiting of a past study that modelled the effects of tropical cyclones from the western Pacific Ocean for a given latitude, longitude point. The methodology is described in: "A Technique for estimating recurrence intervals of tropical cyclone-related high winds in the tropics: results for Guam. Journal of Applied Meteorology, 35, 628-637. 1996"

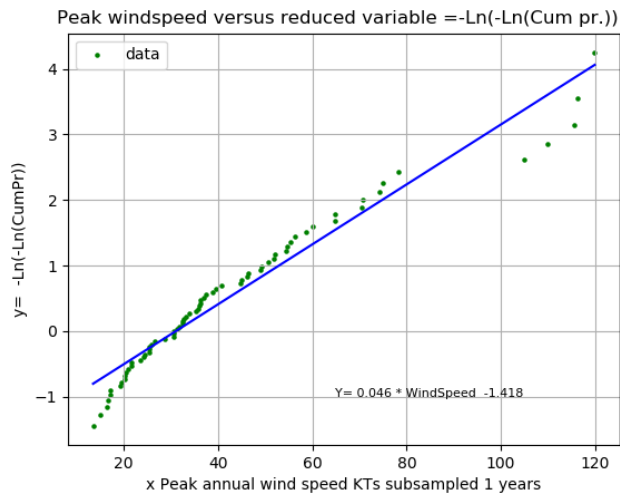
The technique uses a model of the wind radius profile around a tropical cyclone. This when combined with a dataset of historical tropical cyclone best track data, can give a reasonable accurate reconstruction of historical winds that would have been experienced in locations that didn't have wind recorders to capture passing cyclones. It gives a modeled reconstruction of what would have been felt. The Best track data is from the Joint Typhoon Warning Center which covers a period from 1945 to 2014.

To start the statistical analysis the first step was to extract the highest estimated wind speed for each year in the dataset using a reference point of latitude and longitude. These values were then ranked from highest to lowest and a function calculated on the cumulative probability of $n/(N+1)$ to arrive at a reduced variate. The reduced variate was the double natural logarithm of the cumulative probability. In this form when the data was plotted on logarithmic paper the data points were linear. This is one of the assumptions of the Gumbel Extreme value distribution type 2. As long as the data came from a distribution that was of an exponential form and that the sample was sufficiently large this distribution could be used to model extreme values.

When these values are plotted as shown below a linear model could be built to model the recurrence interval as a function of wind speed.



Sometimes there is not enough activity during a single year to produce data that would be considered in the upper tail of the distribution. When this occurs it shows the curved line or a second order feature. In this case the data can be sampled over time scales longer than a single year until the fit is linear. See the feature below that shows a curved fit of the data points when sampled using a peak wind annually.



This is an important consideration because if this technique is applied to areas that don't experience a lot of tropical cyclones then you may not be getting accurate estimates especially on the higher range of wind speeds. When there is a good fit, even if sampled over multiple years as the interval, the results can be algebraically transformed for an annual 1 year value. Below are such results for Guam.

Recurrence interval of peak annual Tropical cyclone winds for Guam

Wind Speed Kts	Return Period in years
30.0	2.41
40.0	3.03
50.0	3.88
60.0	5.02
70.0	6.56
80.0	8.62
90.0	11.39
100.0	15.12
110.0	20.12
120.0	26.83
130.0	35.84
140.0	47.93
150.0	64.17

Here are the steps I used to generate the data and charts.

- 1) Run the LoadDB_1_hour_data.py. This will create a SQLite database using python. It loads over 350000+ records that represent tropical cyclone best track data interpolated to every hour.

- 2) Run the `readfrommdb_1hr_data.py`. It will prompt you for a station latitude and longitude. Make sure East longitudes are entered as a negative number. The output is a file of the highest estimated wind for each year.
- 3) For the last part of the analysis run the `Plot_TC_RP.py` script. It's hard-coded for a file name but you can modify that to your file. This script will generate 2 charts as listed earlier in this document.
- 4) If your results show a substantial second order curve to the fit you may need to sample your data over a longer timeframe. If that's the case use the utility "`highest_in_n.py`". it's output is a vector of numbers that is the highest of each pair going through the list of numbers by 2. If you go this route you need to set a parameter in the "`Plot_TC_RP.py`" script to change the value of the "grouping" variable. It adjusts for sampling other than a single year and adjust the results to get to the correct return period/recurrence interval for a single year probability.