Modeling Non-stationary Processes of Extremes using Regression Trees

Morgan Huang [1] Paul Kushner [2] Karen Smith [3]

[1]Department of Computer and Mathematical Sciences, University of Toronto Scarborough [2]Department of Physics, University of Toronto [3]Department of Physical and Environmental Sciences, University of Toronto Scarborough

Abstract

We focus on modeling extreme-valued wind speeds which can help us provide statistical estimates on future wind speeds. Typically, extremal data can be easily modeled using extreme value theory (EVT), but usually assume stationary data. Modeling non-stationary processes, which are common in environmental applications, however, is a non-trivial task. In this study, we explore extreme wind events in Toronto between 1979-2020 using ERA5 reanalysis data, and aim to use regression trees to form quasi-stationary clusters, for which we use a point process approach with regression parameters using the clusters as binary variables to provide statistical conclusions on its return values.

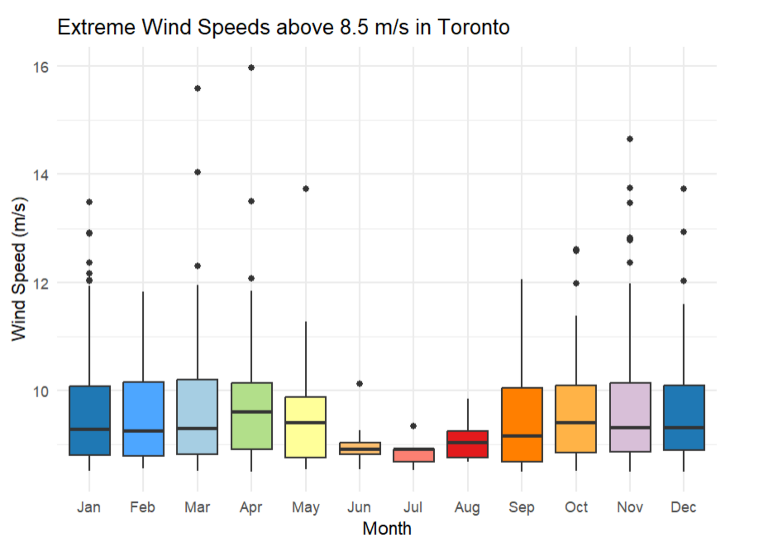

Figure 1: Boxplots of Extreme Wind Speeds showcasing seasonal cycle

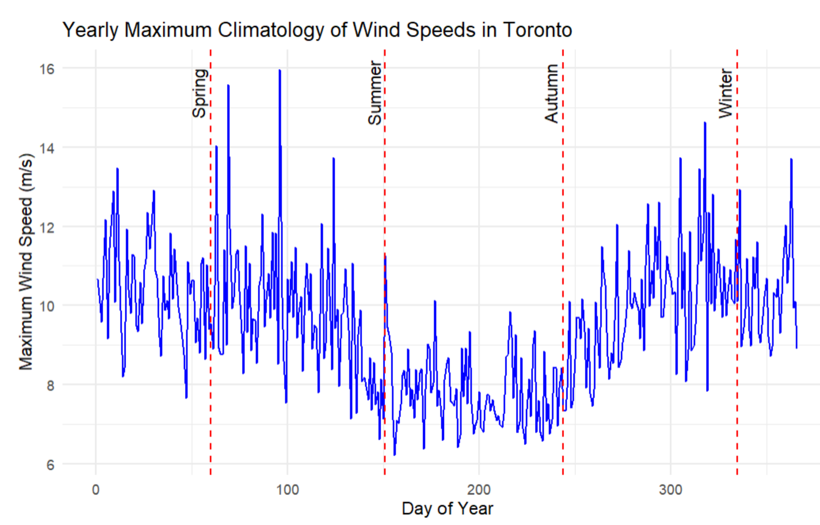

Figure 2: Daily Maximum Climatology of Wind Speeds between 1979-2020

Atmospheric Covariates

We would like to use covariates in our extreme value analysis of extreme winds not only to form stationary slices of our non-stationary data, but to also better understand what causes extreme wind events. To determine these covariates, we want to understand physical relationships between atmospheric variables and wind speed, and potential causes of extreme wind events.

Thermal Wind and Atmospheric Stability

In theory, wind is created by horizontal pressure gradients, where areas of high pressure exert some level of force towards areas of low pressure. From Figure 3, colder temperatures lead to a stronger decrease in pressure with respect to altitude than hotter temperatures, resulting in pressure gradients at different vertical altitudes. We calculate thermal wind, which uses temperature gradients to calculate the rate of change in horizontal wind speeds (U_g, V_g) with respect to altitude z:

$$\frac{\partial U_g}{\partial z} = -\frac{|g|}{T_v \cdot f_c}\frac{\partial T_v}{\partial y} \qquad \frac{\partial V_g}{\partial z} = \frac{|g|}{T_v \cdot f_c}\frac{\partial T_v}{\partial x}$$

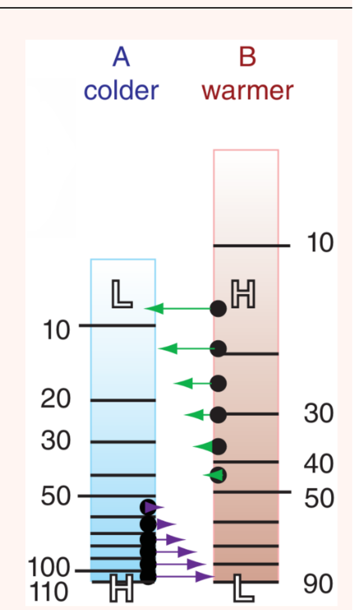

Figure 3: Difference in pressure is due to the hypsometric equation [7]

In addition to temperature gradients, an unstable atmosphere is heavily correlated with extreme wind events. Under unstable conditions, air parcels continue to rise to the top of the troposphere, which forces the jet stream winds to mix with surface winds. We can simply calculate the atmospheric stability by calculating the potential temperature gradient with respect to altitude.

Using covariates of atmospheric stability calculated in two layers (1000hPa-850hPA, 850hPa-500hPa), thermal wind gradients in the x and y direction, low level winds (winds at 850hPa), and the jet stream (winds at 500hPa), we fit a regression tree to find suitable discrete clusters. This makes our return level calculation easier and provides better information on the behavior of extreme winds without using interaction effects.

Methodology

We propose a point process approach in modeling extreme wind events. This is to overcome certain caveats that are present in other methods such as the block maxima approach, which is often wasteful of extremal data, and the peaks-over-threshold (POT) approach, where modeling non-stationarity often results in violating the threshold stability property (Eastoe and Tawn, 2009).

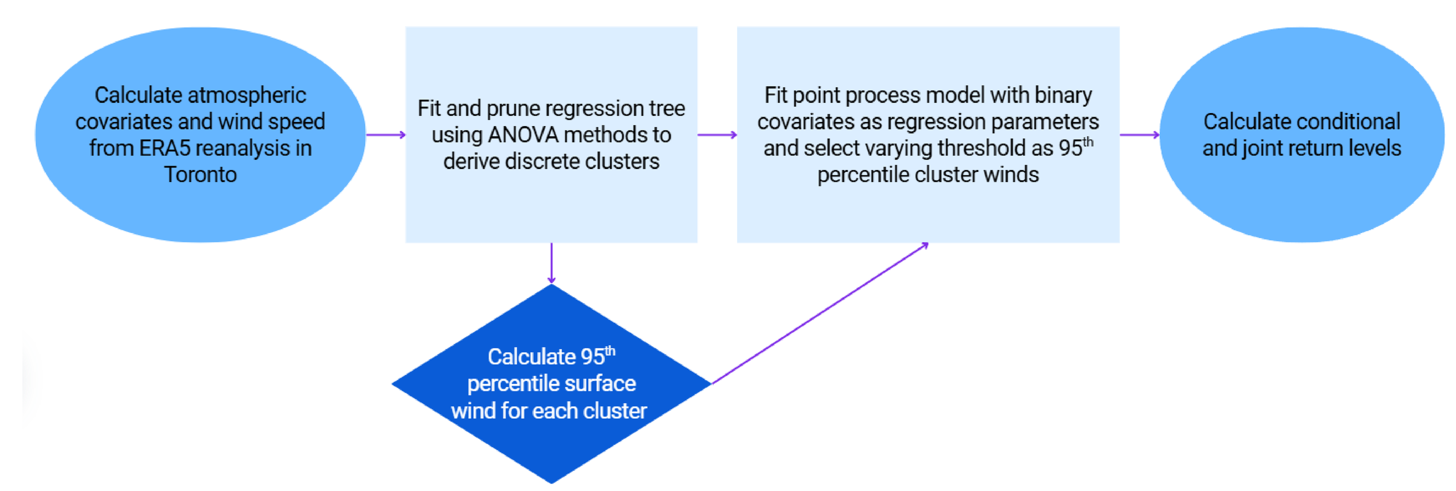

Figure 4: Flow Chart indicating the steps in fitting extreme value winds to a GEV distribution using the point process approach. We use ANOVA to fit our regression tree, but better theoretical basis can be found by setting the objective function as the negative log-likelihood (Farkas et. al 2024).

Results

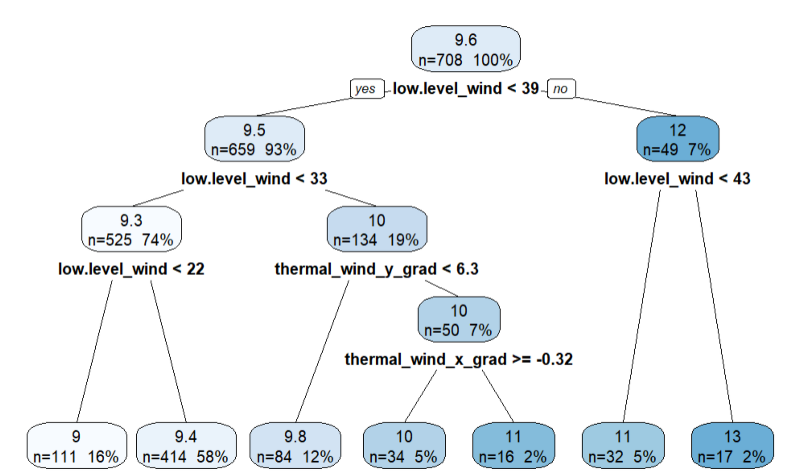

Figure 5: Regression Tree

After fitting and pruning the regression tree using ANOVA methods on extreme wind events (wind speeds of over 8.5 m/s), we obtain 7 different clusters, as shown in Figure 5. Thresholds for each cluster is heuristically chosen as the 95th percentile surface wind (Figure 6). Figure 7 shows the varying threshold with surface winds. Points above the threshold are considered extreme given the cluster, and are fitted into the point process model.

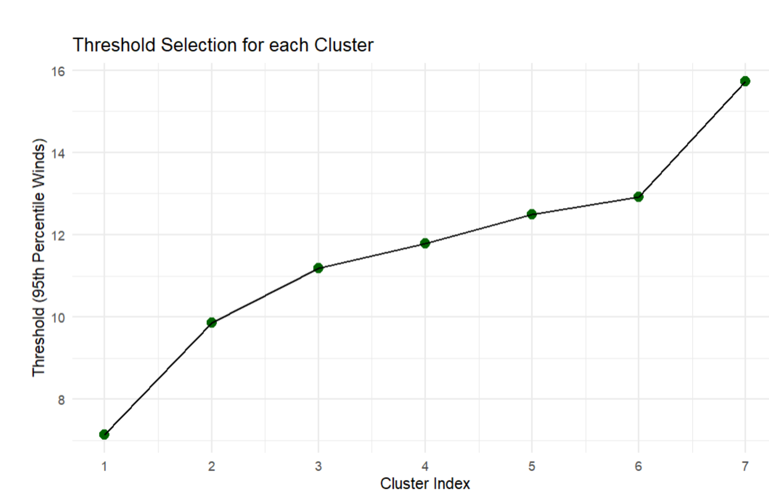

Figure 6: Threshold Selection (95th percentile) for each cluster

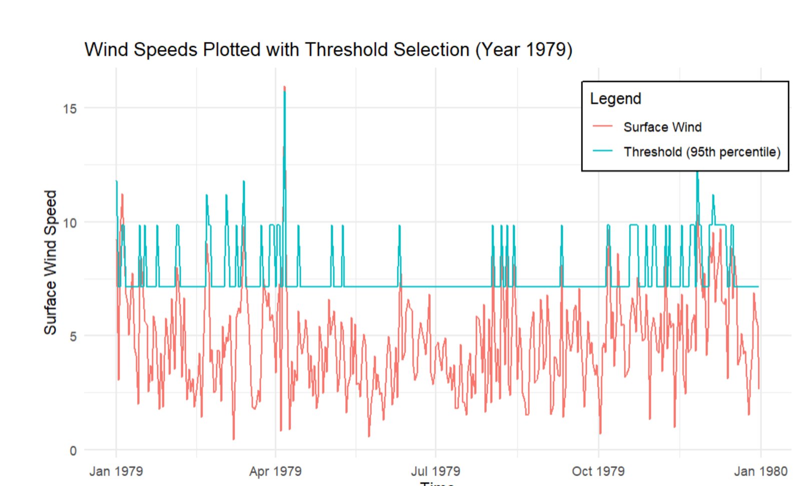

Figure 7: Plot of Varying Threshold and Wind Speed in the year 1979

We can perform model diagnostics on our point process model using a QQ-plot as shown on the right. Since our parameters vary with time, standardization is required. In this case, we convert the GEV parameters to parameters of a Generalized Pareto distribution, then standardize to a standard exponential distribution and plot its theoretical quantiles.

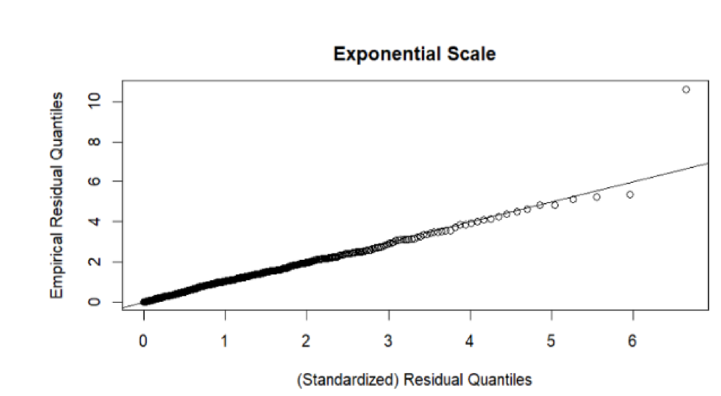

Figure 8: QQ-plot matches almost perfectly with the exception of one outlier

Return Levels

Working with a non-stationary process, we can immediately calculate conditional return levels $z_{p,k}$, where the probability that $z_{p,k}$ is exceeded for a given year is given by p, conditioned on the fact that all data points in a given year belong to cluster C_k:

$$\mathbb{P}(Z > z_{p,x} | Z \in C_k) = p$$

If we are interested in general return levels regardless of cluster partitions C_k, then we can calculate z_p by solving the following equation:

$$\mathbb{P}(Z > z_p) = \sum_{k=1}^{m} \mathbb{P}(Z > z_p | Z \in C_k)\mathbb{P}(Z \in C_k) = p$$

We thus calculate conditional return levels for each cluster k along with its bootstrap 95% confidence intervals on the left, and then we solve the general return level equation with 95% bootstrap confidence plotted on the right:

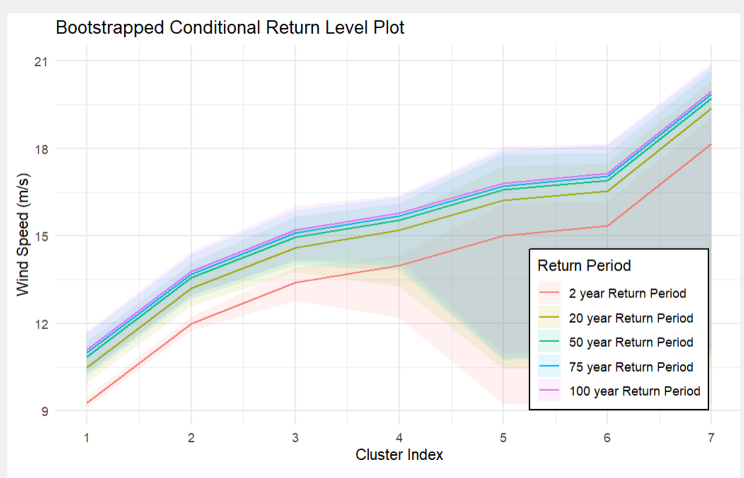

Figure 9: Conditional Return Levels for each cluster for return period of 2, 20, 50, 75, and 100 years

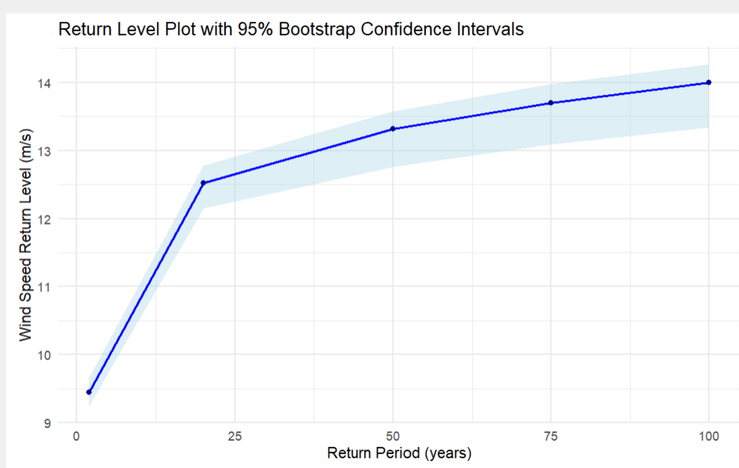

Figure 10: General Return Level plots

Assuming that the 40-year historical distribution of wind speeds do not change in the future, we expect that a wind speed of 14 m/s will be exceeded once every 100 years.

Discussion

Assuming that the historical distribution will be equal to future distributions of wind speeds is a strong assumption, especially with regards to climate change (Milly et al. 2008). Since extrapolation to future processes is not favorable (Cooley, 2013), environmental studies must focus on providing EVT analysis on not only historical, but future climate model projections. In this study, we were only able to provide information about historical extremes, since regular climate models (resolution of around 1 degree latitude and longitude) show no obvious trend in extreme wind speeds. However, (Morris et. al 2024) recently showed some increases in extreme wind speeds in Southern Ontario, by using a "variable-resolution" grid which increases resolution towards the area of study. This paper showed that extreme wind events are not caused by global effects of extratropical cyclones, but rather from locally reduced atmospheric static stability which cannot be detected by regular climate models.

Acknowledgments

This project is supported by the Data Sciences Institute, University of Toronto. The author thanks Paul Kushner and Karen Smith for their guidance on this project.

References

[1] Stuart Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer London, 2001. [2] Daniel Cooley. Return periods and return levels under climate change. *Extremes in a Changing Climate*, pages 97–114, 2013. [3] Jonathan A. Tawn Emma F. Eastoe. Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 58(1):25–45, 2009. [4] Michael Morris et al. Resolution dependence of extreme wind speed projections in the great lakes region. *Journal of Climate*, 37:3153–3171, 2024. [5] Paul CD Milly et al. Stationarity is dead: Whither water management? *Science*, 319:573–574, 2008. [6] Sebastien Farkas et al. Generalized pareto regression trees for extreme event analysis. *Journal of Extremes*, 27:437–477, 2024. [7] Roland Stull. *Practical Meteorology: An Algebra-based Survey of Atmospheric Science*. Dept. of Earth, Ocean Atmospheric Sciences, University of British Columbia, 2017.

https://www.linkedin.com/in/morganhuang237/

SUDS Showcase 2025, 10th floor, 700 University Avenue, Toronto

morgan.huang@mail.utoronto.ca