Multivariate Extreme Value Theory with an Application to Climate Data in the Western Cape Province

Lipika Bhagwandin (South Africa), Supervisor: Dr Şebnem Er (South Africa)
Lipika.Bhagwandin@alumni.uct.ac.za, sebnem.er@uct.ac.za
Department of Statistical Sciences, University of Cape Town

Introduction

- Extreme Value Theory (EVT) concerns the mathematical modelling of extreme events. A distinguishing factor of EVT is that the focus of the analysis is on the tails of a distribution. It allows one to make inference about extreme events occurring over a time period into the future. Accounting for multiple variables working together as predictors of extreme events provides a more accurate representation of how extreme events are ultimately caused from a statistical perspective.
- Multivariate EVT (MEVT) is used in this study to jointly model the extremes of weather variables. Specifically, to model pairwise combinations of rainfall, temperature and wind speed maxima from five weather stations across the Western Cape province in South Africa.

Methodology

- Multivariate analysis is split into two parts which look at the marginal distributions and dependence structure separately.
- Marginal analysis is completed using univariate techniques namely point process and threshold excess approaches through asymptotically dependent models.
  - Marginal distributions are transformed to a standardised distribution so that the data follows the multivariate extreme value distribution. Standard Fréchet margins are used in this study.
- Component Wise: Defining $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ as an independent sequence of random vectors (standard Fréchet margins) with a distribution function $F(x, y)$, the limiting joint distribution is

$$M_n = \left(\frac{max\{X_i\}}{n}, \frac{max\{Y_i\}}{n}\right) \quad \text{for } i=1,\ 2,..,n$$

$$Pr\{M_{x,n} \leq x, M_{x,n} \leq y\} \rightarrow G(x, y) \tag{1}$$

$$G(x, y) = \exp\left\{-\left(2\int_0^1 max\left(\frac{w}{x}, \frac{1-w}{y}\right)dH(w)\right)\right\} \tag{2}$$

$x, y > 0$, where G is a non-degenerate distribution function and H is a distribution function on [0,1] which satisfies the following mean constraint

$$\int_0^1 wdH(w) = \frac{1}{2} \tag{3}$$

- Threshold Excess: Defining $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ as an independent sequence of random observations (standard Fréchet margins) with a distribution function $F(x, y)$, the marginals have a distribution that is as follows

$$G(x) = 1 - \zeta_{u_x}\left[1 + \xi\left(\frac{x - u_x}{\sigma}\right)\right]^{-1/\xi} \quad x > u_x \tag{4}$$

$$G(y) = 1 - \zeta_{u_y}\left[1 + \xi\left(\frac{y - u_y}{\sigma}\right)\right]^{-1/\xi} \quad y > u_y \tag{5}$$

where u_x and u_y are suitable thresholds chosen using Mean Residual Life (MRL) plots. The bivariate case $G(x, y)$ follows on from (2).

- Point Process: Defining $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ as an independent sequence of random observations (standard Fréchet margins) with a distribution function $F(x, y)$, the marginals have a distribution that follows (1).

The point process sequence N_n is of the form

$$N_n = \left\{\left(\frac{x_1}{n}, \frac{y_1}{n}\right), \left(\frac{x_2}{n}, \frac{y_2}{n}\right), ..., \left(\frac{x_n}{n}, \frac{y_n}{n}\right)\right\}$$

Transforming to pseudo-polar coordinates where

$$r = x + y \quad \text{and} \quad w = \frac{x}{r} \tag{6}$$

the intensity function of the point process sequence is defined as

$$\lambda(r, w) = 2\frac{dH(w)}{r^2} \tag{7}$$

which relates back to equation (2).

- Parametric models are used to capture dependence structure between the variables.
  - The simplest model is the logistic family which is symmetric:

$$G(x, y) = \exp\{-(x^{-1/\alpha} + y^{-1/\alpha})^\alpha\} \quad x > 0, y > 0 \tag{8}$$

for $\alpha \in (0, 1)$. α is a measure of the dependence strength which represents independence $(\alpha = 1)$ and dependence $(\alpha \rightarrow 0)$.

- Parameter estimates are calculated using maximum likelihood estimation.
- Diagnostic measures include AIC, coefficients of tail dependence and extremal dependence.
- Component-wise maxima, threshold excess and point process models are explored and applied to weather data. The performance of the models are compared to each other in R using the *evd* package.

References

[1] Beirlant, J. Goegebeur, Y. Teugels, J. Segers, J. De Waal, D. and Ferro, C. Statistics of Extremes.(2004).
[2] Coles, S. An Introduction to Statistical Modeling of Extreme Values.(2001), pp.143-166.
[3] Ledford, A. and Tawn, J. Statistics for near independence in multivariate extreme values.(1996).

Data

- Weather variables are daily maximum rainfall (mm), daily maximum temperature (^{0}C) and daily maximum wind speed (m/s) for five stations across the Western Cape province from 1965 to 2015.
- Data is broken up into the four seasons for stationarity: Summer (December - February), Autumn (March - May), Winter (June - August) and Spring (September - November).
- The five stations are Cape Town International Airport, George Airport, Langebaanweg, Plettenberg Bay and Vredendal. The results for Cape Town International Airport will be focused on for the purposes of this poster.

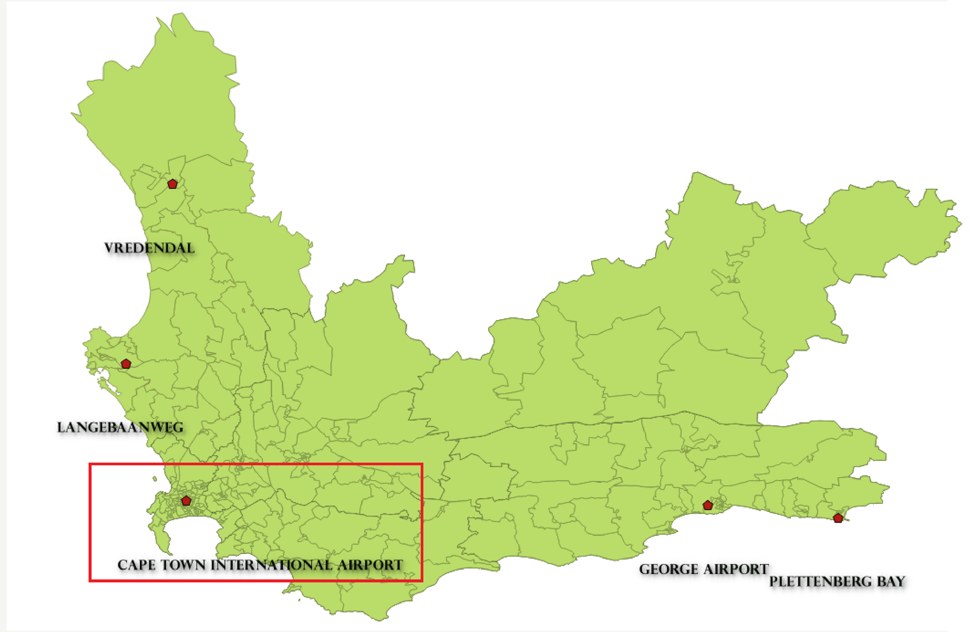

Figure 1: Map of the five stations across the Western Cape province.

Results

Figure 2 shows the transition of the raw data (wind speed and temperature maxima) to standard Fréchet margins and then plotted on log scale for Cape Town International Airport during spring.

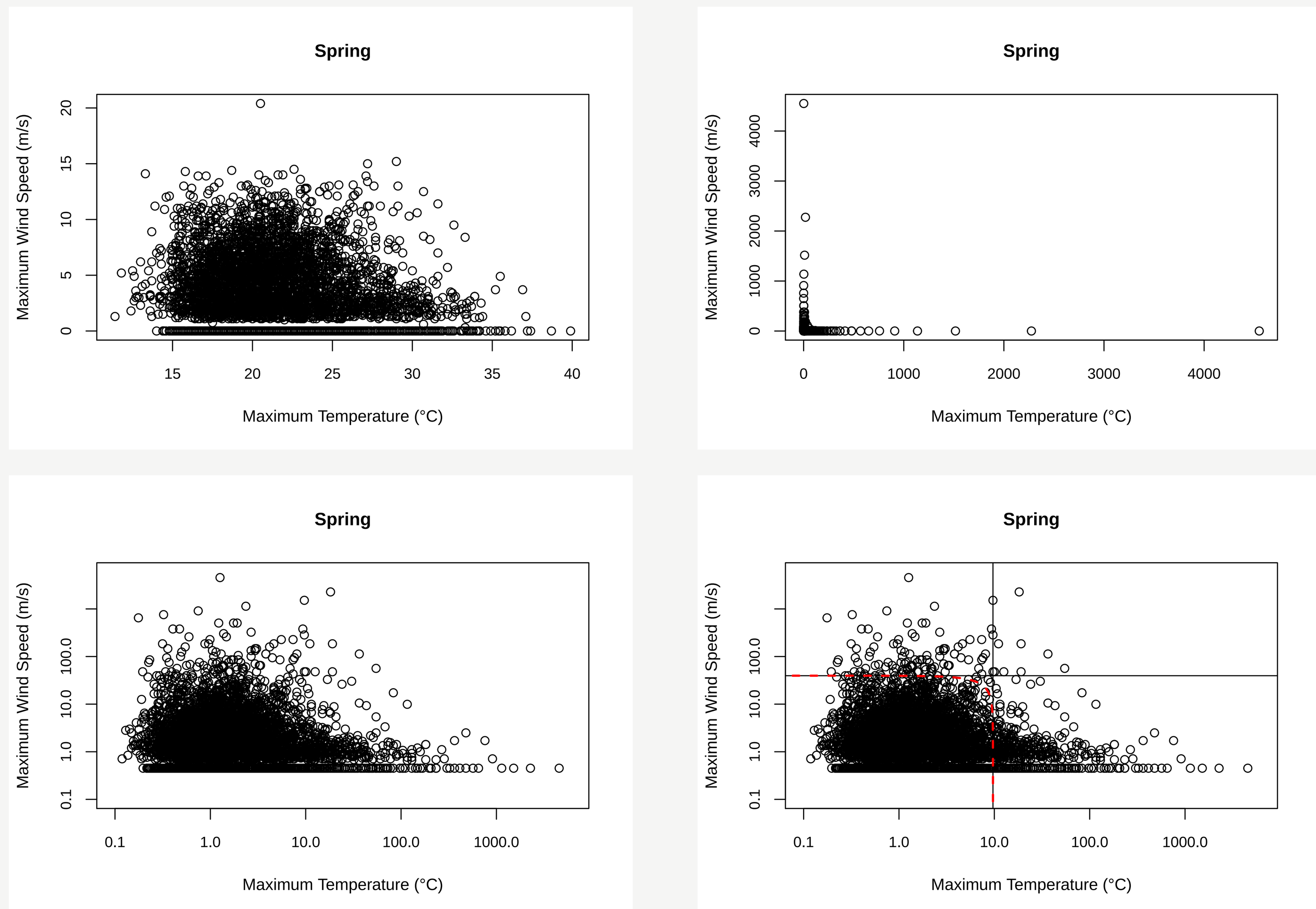

Figure 2: *Top left:* original data, *Top right:* transformed data , *Bottom left:* transformed data on log scale, *Bottom right:* transformed data on log scale with point process (- - -) and threshold excess thresholds (—).

Table 1: Parameter Estimates for Temperature and Wind Speed Maxima using the log model.

Season	Component-Wise		Threshold Excesses		Point Process	
	α	Std error	α	Std error	α	Std error
Summer	0.965	(0.067)	0.987	(0.009)	0.784	(0.008)
Autumn	0.803	(0.101)	0.915	(0.016)	0.752	(0.007)
Winter	0.969	(0.101)	0.877	(0.025)	0.801	(0.008)
Spring	0.999	(2 x 10^{-6})	0.962	(0.015)	0.754	(0.007)

- The results from the component-wise and threshold excess show α estimates close to 1 which indicates weak dependence between the variables at a single station throughout the seasons.
- The point process dependence estimates tend further away from 1 but still indicates weak dependence between the variables at a single station throughout the seasons. The difference in dependence estimates seen from the point process approach can be attributed to the curved threshold boundary which allows for more observations to analysed.

Conclusions & Future Work

- The component-wise model is wasteful of data and does not perform well in capturing the dependence between the weather extremes.
- The threshold excess and point process models are better performing models for jointly capturing the relationship between the maxima of different weather variables at a single location. However, asymptotically independent models seem to be more appropriate for this data.
- Extending to greater than two variables at a single location and including different dependence models.
- Exploring asymptotically independent models and non-parametric methods to model multivariate extremes.

Acknowledgments

The attendance of Lipika Bhagwandin at this Conference is sponsored by research funds of the late **Emeritus Professor Timothy Dunne.**
This presentation is in memory of a man who was an inspiration to young academics and students.

- South African Weather Services
- City of Cape Town