Analysis of Extreme Events

Sibusisiwe Khuluse

CSIR- Built Environment, Logistics and Quantitative Methods (CA)
Pretoria, South Africa.

Africa Women in Statistics Seminar, April 2009

Analysis of Risk

Risk analysis and decision analysis frameworks provide an array of tools to aid the decision making process. Probabilistic Risk Analysis (PRA) involves

- Estimation of probabilities.
- Determination of the distribution of the damage.
- Preparation of products that enable prediction of future risk events.

Extreme Value Theory

Extreme value theory (EVT) can be thought of as a tool to estimate the tail area of the distribution.

- Characterizes the probabilistic risk associated with an extreme event.
- In environmental studies, probabilistic risk often interpreted in terms of return intervals.

Classical Extreme Value Theory

Assume the X_i to be a sequence of independent random variables with common distribution F. The cornerstone of EVT is that, without any knowledge about F, a model exists that describes the behaviour of the largest (or smallest) member of the sample

$$M_n = \max(X_1, X_2, \ldots, X_n)$$

Classical Extreme Value Theory

Conditional on the existence of $\{a_n\}$ and $\{b_n\} > 0$, the *Fisher-Tippett theorem* states that the re-scaled sample maxima (or minima) converges in distribution to a variable whose distribution is only one of 3 types: I-Gumbel, II-Frechet and III-Weibull. These can be unified into

$$P\left(\frac{M_n - a_n}{b_n} \leq x\right) \longrightarrow \exp\left\{-\left(1+\xi\frac{x-\mu}{\sigma}\right)_+^{-\frac{1}{\xi}}\right\} \tag{1}$$

provided that ($y_+ = \max(y,0)$), $-\infty < \mu < \infty$ and $\sigma > 0$.

- This is the Generalized Extreme value (GEV) family of distributions.

From the Classical to the Threshold Exceedance Approach

An important consideration in classical EVT is the choice of block size *n*.

- Affects the trade-off between bias and variance, i.e. choice between accuracy or precision.

Criticism about the classical approach is that it is wasteful of data.

- Using only one observation per block, discarding the rest.

Alternative approach is the *threshold exceedance* approach.

- Essentially finding an approximate distribution for the series of excesses of a particular level (the threshold).

Generalized Pareto Distribution

Denote X_i by X. Suppose for large n, the Fisher-Tippett theorem holds. Then, for suitable threshold u,

$$P(X - u|X > u) \sim G(y; \sigma_u, \xi) = 1 - \left(1 + \xi \frac{y}{\sigma_u}\right)_+^{\frac{-1}{\xi}} \tag{2}$$

defined on $\{y : y > 0 \text{ and } \left(1 + \xi \frac{y}{\sigma_u}\right) > 0\}$, with

$$\sigma_u = \sigma + \xi(u - \mu) \tag{3}$$

- $G(\cdot)$ defines the Generalized Pareto distribution (GPD).

Generalized Pareto Distribution

The shape parameter, ξ, which is equivalent for the GEV and GPD, is determines the characteristics GPD.

- $\xi < 0$ - GPD has an upper bound $u - \sigma_u/\xi$
- $\xi > 0$ - GPD has no upper limit
- $\xi = 0$ interpreted as the limit $\xi \longrightarrow 0$, corresponds to the exponential distribution with parameter $1/\sigma_u$

Selecting the Threshold

Graphical methods of selecting an appropriate threshold.

1. The mean excess plot
   - Describes the conditional behaviour of the mean excess, as specified by

$$\mathrm{E}(X - u | X > u) = \frac{\sigma_u}{1+\xi} - \frac{\xi}{1+\xi} u \tag{4}$$

   - Suitable threshold selected from the plot as that value which is at the onset of linearity
2. Threshold stability plots
   - Threshold is selected by fitting the GPD across a range of thresholds, then assessing the stability of the parameter estimates.

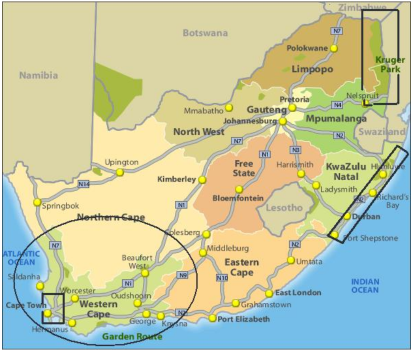

Figure: Map of South Africa with the study areas highlighted

- Western Cape
  - Climatologically diverse: Influence of the varied topography and it's location with respect to ocean currents.
  - Classified as a Mediterranean climate region.
  - Precipitation is of orographic and frontal nature.
- Coast of KwaZulu Natal
  - Popular tourist destination: Warm beaches.
  - Ports of Durban and Richards Bay handle the largest proportion of the country's exports.

Outline

Data

- Cape Town's daily rainfall series for the period 1958 – 2007.
- The 75th and 99th percentiles are 0.2 mm and 22.6 mm respectively.
- Highest recorded rainfall received on a single day is 93.7 mm (April 1993).
- Nearly the same amount fell over 3 days in July 2007, affecting nearly 40 000 people, mostly residents of Khayelitsha and Mitchell's Plain.

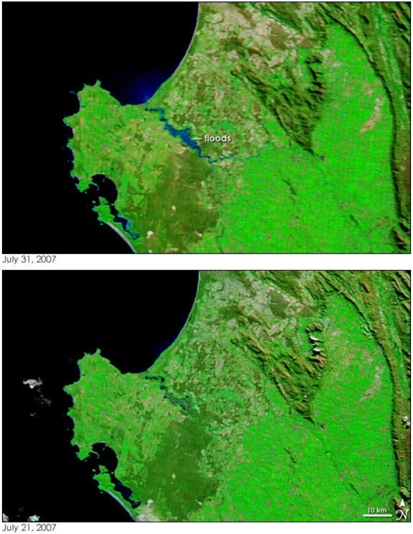

Figure: Swelling up of river near Saint Helena due to heavy rain in July 2007

Exploratory Analysis

From visual inspection of the scatter plot (Fig. 3), *no trend* in annual maximum and average rainfall could be detected.

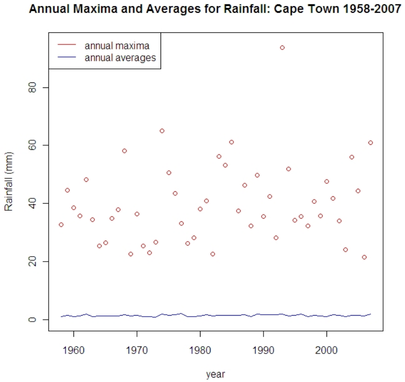

Figure: Scatter plot of the rainfall maximum and average series

Results from Fitting the GEV to Rainfall Maxima

Model	Parameter	Estimate	Std. error	95% CI
GEV Annual Max	μ	33.57	1.59	(30.45,36.69)
	σ	9.77	1.21	(7.40,12.14)
	ξ	0.07	0.12	(-0.11,0.37)
	50-yr r.l.	77.55	-	(64.62,108.54)
GEV Winter Max	μ	32.13	1.53	(29.13,35.13)
	σ	9.13	1.21	(6.76,11.50)
	ξ	0.15	0.14	(-0.07,0.55)
	50-yr r.l.	80.85	-	(64.29,122.60)

Table: GEV maximum likelihood estimates

Diagnostic Plots for the Annual Maxima Series

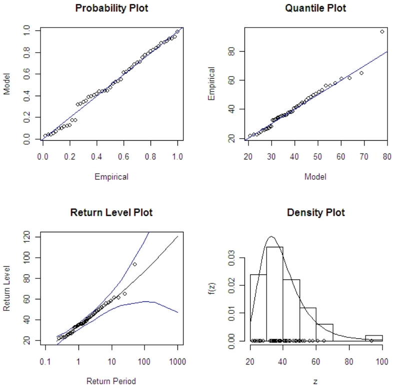

Figure: Diagnostic plots for checking the adequacy of the GEV model to annual rainfall maxima

Outline

Figure: Picture taken in Durban during the massive surf event that hit the coast of KZN in March 2007

NOAA Data for Durban

The March 2007 high surf was estimated to have resulted in R 400 million in damages. At the peak of the storm

- Significant wave height (HMO) reached 8.5 m.
- The water level was almost 40 cm above the predicted tide.

For the analysis 11 years (1997-2008) of simulated wave data from the National Oceanic and Atmospheric Administration (NOAA) was used.

Threshold Selection

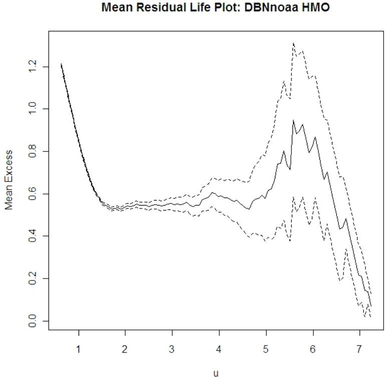

Figure: Mean Residual Life Plot

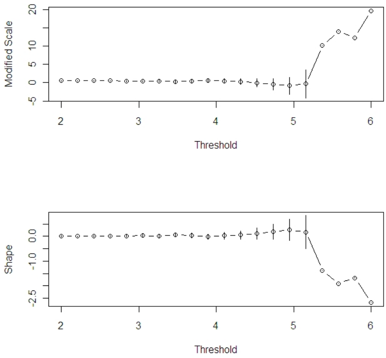

Figure: Threshold Stability Plot

Results of the Threshold Exceedance Analysis

HMO exceedances of 3.8 m were approximated by the GPD. An estimate of the exceedance rate (per year) was 29.

Parameter	Estimate	Std. error	95% CI
σ_u	0.60	0.05	(7.40,12.14)
ξ	-0.01	0.05	(-0.10,0.11)
50-yr r.l.	8.05	-	(7.18,9.01)
100-yr r.l.	8.44	-	(7.40, 9.49)

Table: MLE of the GPD model of HMO values in excess of 3.8 m

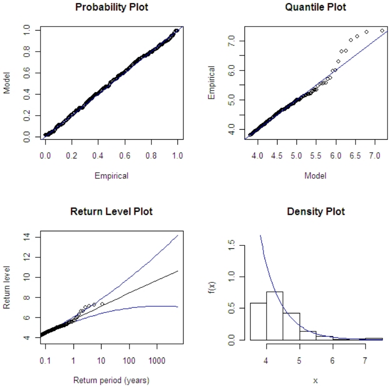

Figure: Diagnostic plots for the GPD model of the exceedances

The methodology provided by Extreme Value Theory can be a powerful tool in risk analysis, however this can be demeaned as the result of:

- The paucity of extreme events, as well as in some instances, the lack of historical records of such events.
- Failure to exercise caution during the model-building stage.
- Extrapolation too far beyond the range of the data.

To derive the most out of this approach, development and understanding of advanced EVT methods becomes necessary. Further, there is a strong need for close collaboration between the domain expert and the modeller/ analyst.

For Further Reading

1. Beirlant J., Goegebeur Y., Segers J. and Teugels J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley.
2. Coles S.G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.
3. International Council for Science Regional Office for Africa (2007). *Science Plan on Natural and Human-Induced Hazards and Disasters in sub-Saharan Africa*.
4. IPCC- Working Group II (2007). Editors: Parry M.L., Canziani O.F., Palutikof J.P., Van der Linden P.J. and Hanson C.E. *Climate Change 2007: Impacts, Adaptation and Vulnerability*. Cambridge University Press.
5. Paté-Cornell M.E. and Dillon R.L. (2006). *The Respective Roles of Risk and Decision Analyses in Decision Support*. Decision Analysis, 3(4):220-232.

CSIR
our future through science

"The theory of probabilities is at bottom nothing but common sense reduced to calculus." -Laplace, Théorie analytique des probabilités, 1820

Thank you!