The Effect Of Temporal Dependence And Seasonality On Return Level Estimates Of Excessive Rainfall

Sibusisiwe Khuluse
CSIR - Built Environment
Wits University - Department of Statistics and Actuarial Science

August 20, 2009

Outline

Outline

Study Area

Figure: Satellite image of Cape Town

Problem Statement

Figure: Informal settlements are vulnerable to flooding

The need for Extreme Value Theory

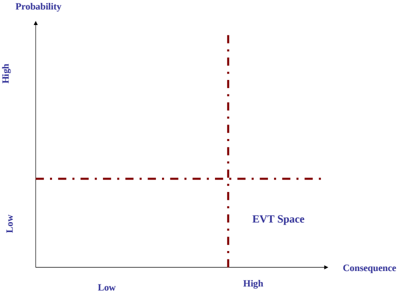


Figure: The domain of application Extreme Value Theory

Outline

The Classical Approach

Assume X_i is a sequence of independent random variables with common distribution F. Without any knowledge of F, a model exists that describes the behaviour of the largest (or smallest) member of the sample

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

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 the Generalized Extreme Value (GEV) family of distributions

$$P\left(\frac{M_n - a_n}{b_n} \leq x\right) \longrightarrow \begin{cases} \exp\left(-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right) & 1+\xi\frac{x-\mu}{\sigma} > 0,\ \xi \neq 0 \\ \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) & x \in \mathbb{R},\ \xi \to 0 \end{cases} \tag{1}$$

where $-\infty < \mu < \infty$ and $\sigma > 0$.

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).

Approximate Distribution for Threshold Exceedances

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

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

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

$$\sigma_u = \sigma + \xi(u - \mu) \quad (3)$$

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

iSibalo Capacity Building Programme CSIR our future through science

The Return Level Parameter

The return level is that level of the process, which we expect to be exceeded on average once every N years.

$$q_N = \begin{cases} u + \frac{\sigma_u}{\xi}\left[(\lambda N)^{\xi} - 1\right] & \xi \neq 0 \\ u + \sigma_u \log(\lambda N) & \xi = 0 \end{cases} \quad (4)$$

Exceedance process is assumed to be Poisson with rate λ (per year), estimated by $\hat{\lambda} = m/n$.

Issues to Consider

- The data are often incomplete due to measuring instrument failure, relocation of measuring sites, etc.
- Careful consideration has to be taken in selecting the threshold.
- The length of the data that is available is often shorter than the prediction horizon.

Outline

Data

Rainfall data from Cape Town International Airport ($-33°97'$S, $18°60'$E, 44 m altitude).

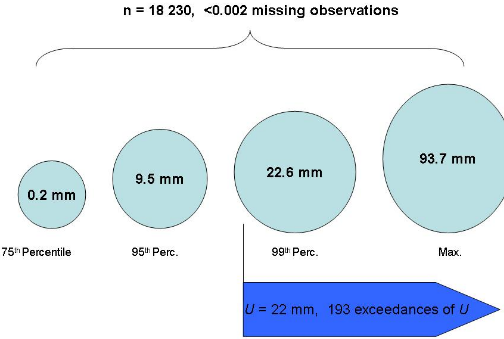


Figure: Description of the rainfall data

Data

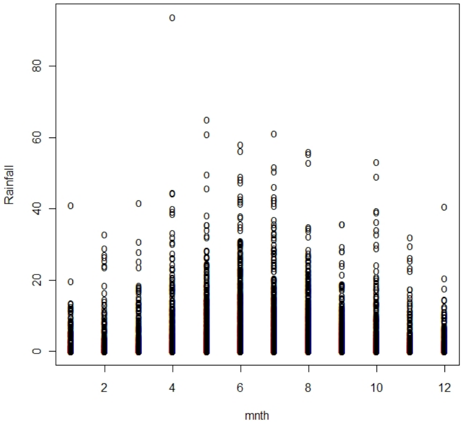


Figure: Rainfall series by month (1958-2007 Cape Town Int)

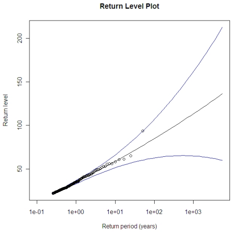
Results

The estimated extremal index $\hat{\theta} = 0.78(0.69, 0.89)$

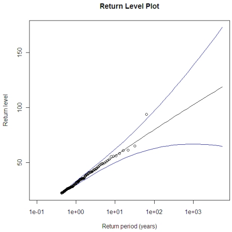
Model (m)	$\hat{\sigma}$(s.e.)	$\hat{\xi}$(c.i.)	$\hat{q}_{25}$	$\hat{q}_{50}$
GP0 (193)	9.37 (0.98)	0.04 (-0.08,0.22)	68.91 (60.36,83.41)	76.79 (65.50,93.72)
GP1 (149)	11.17 (1.25)	-0.02 (-0.14,0.17)	68.69 (58.65,80.09)	72.81 (63.83,89.55)
GP2 (144)	9.74 (1.19)	0.05 (-0.09,0.26)	67.06 (58.38,83.23)	75.47 (63.86,94.65)
GP3 (31)	11.72 (2.95)	-0.26 (-0.61,0.22)	45.07 (40.07,56.81)	48.74 (42.95,67.99)

Table: Parameter estimates and the accompanying measures of uncertainty

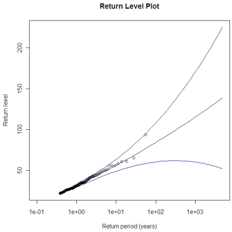
Return Level Plots



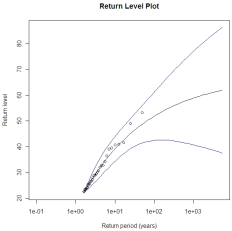
(a) Annual Rainfall (assuming independence)



(b) Annual Rainfall (assuming dependence)



(c) Annual Winter Rainfall



(d) Annual Summer Rainfall

Conclusion

- Clustering extremes should be appropriately treated to ensure better accuracy - especially for long-range return level estimation.
- Where the series shows strong seasonal signal, more insight can be gained by analyzing the behaviour of the extremes of the process for each season.
- To understand the extreme rainfall patterns in the region, extension of the analysis to nearby sites is necessary.

Outline

Acknowledgements

- South African Weather Service for the data.
- Statistics South Africa for the funding through the ISIBALO Capacity Building initiative.
- Mark Dowdeswell for his guidance.

References

- Coles S.G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.
- Beirlant J., Goegebeur Y., Segers J. and Teugels J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley.
- 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.
- International Council for Science Regional Office for Africa (2007). *Science Plan on Natural and Human-Induced Hazards and Disasters in sub-Saharan Africa*.
- 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.