University of Glasgow

Statistical Modelling of Environmental Extremes

– Session 3 –

Dr Daniela Castro-Camilo

SCAN FOR COURSE MATERIAL

Institut Henri Poincaré, Sorbonne University – March 2022

Plan for today

Plan for today

- [9:0-10:30] **Session 3 and lab:** we learn how to tackle non-stationarity in threshold exceedances
  - R/RStudio required. Packages: ismev, evgam, quantreg, viridis.
  - Documents needed: SMEE_IHP2022_Session3HO.pdf, Practical3Lab.pdf.
  - Data needed: prcpFC.txt (same as Session 2)
- [11:00-12:30] **Session 3 cont'd. Session 4 and lab:** we learn how to model spatial exceedances using latent Gaussian models within the INLA framework.
  - R/RStudio required. Packages: INLA, viridis.
  - Documents needed: Practical4Lab.pdf.
  - Data needed: gpsim.txt and gplocs.txt.

What is this session about? – Challenges in threshold exceedance models

- We will learn how to apply the asymptotic results for threshold exceedances to environmental data facing temporal non-stationarity.
- Loosely speaking, if we can approximate the distribution of $M_n = \max\{X_1, \ldots, X_n\}$ by the GEV(μ, σ, ξ) distribution, then, for large enough u, the distribution of excesses over u can be approximated by the generalised Pareto distribution, i.e.,

$$\Pr(X - u \mid X > u) \sim H(y) = 1 - \left(1 + \xi \frac{y}{\sigma_u}\right)^{-1/\xi},$$

  defined on $\{y : y > 0 \text{ and } (1 + \xi y/\sigma_u) > 0\}$ where $\sigma_u = \sigma + \xi(u - \mu)$.
- In other words, if block maxima have approximating GEV distribution, then threshold excesses have a corresponding approximate distribution within the generalized Pareto family.
- In this session we will see how to still use this result when data are non-stationary.
- In the following we assume that exceedances are independent. We later reflect on this assumption and provide alternatives to relax it.

Non-stationary threshold exceedances

- Consider the prcpFC.txt dataset that contains daily precipitations (inches) from 1900 until 1999 in Fort Collins, Colorado, US.
- It is likely that exceedances are affected by seasons, years, etc. Therefore we need to do some exploration first to try to unveil temporal patterns.
- Figure 1 shows the data for the subperiod 1990-1999, with peaks over the fixed 1in threshold ($\sim$ 99.5% overall quantile).

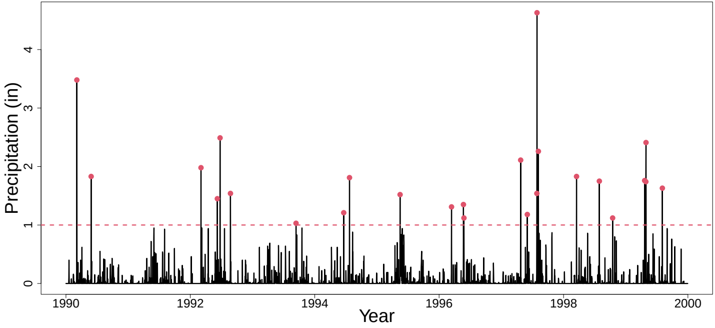

Figure 1: Daily precipitation in Colorado (in) for 1990-1999, with threshold excesses over 1in in red.

Non-stationary threshold exceedances

- Figure 2 shows the original data summarised using boxplots by month, with the 1in fixed threshold in red and the monthly-varying 99.5% threshold in green.

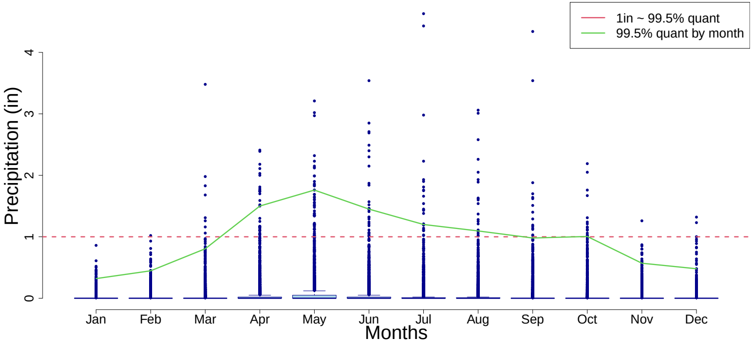

Figure 2: Boxplots of daily precipitation data in Colorado, by month.

- Figure 3 shows excesses over 1in summarised using boxplots by month with lines connecting medians and upper hinges.

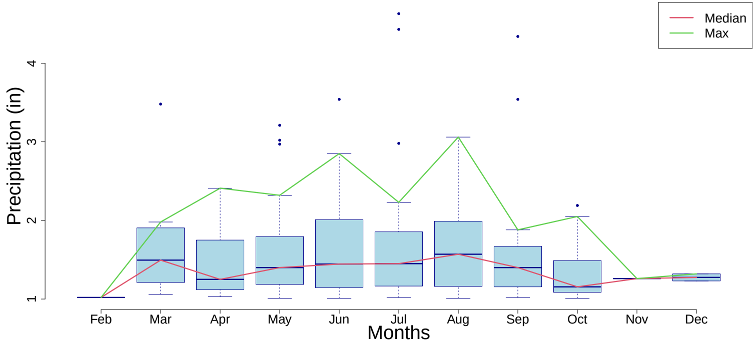

Figure 3: Boxplots of excesses over 1in, by year.

- Figure 4 shows excesses over 1in summarised using boxplots by year.

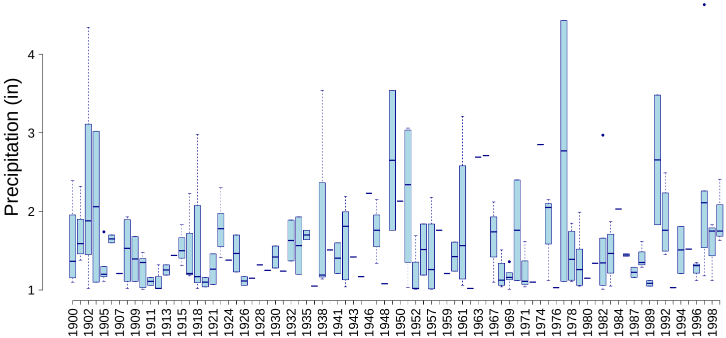

Figure 4: Boxplots of excesses over 1in, by month.

Non-stationary threshold exceedances

- The plots highlight the fact that exceedances vary with time/months/years.
- There are different ways to address these sources of temporal non-stationarity: we can assume time-varying parameters, or we can assume time-varying thresholds.
- Here we will consider different approaches and discuss their pros and cons:

1. **Naive approach:** assuming iid and stationary threshold. We will see ways to choose the threshold.
2. **Slightly less naive approach:** linear trend in $\log(\sigma_u)$.
3. **Single-season approach:** we focus on the season that gives rise to the “most extreme” extremes. For example, July-Aug-Sept and fit a stationary model with fixed threshold.
4. **Seasonal piecewise approach (motivated by a comment from last week!):** we fit 12 different stationary GP models to exceedances from each month, using month-specific thresholds.
5. **Smoothly varying seasonal parameters using GAMs:** we let the GP paramaters to flexibly vary with months/years.
6. **Threshold exceedances using quantile regression**: all the above approaches (except, perhaps, 4) can also assume a continuously time-varying threshold $u(t)$. Here we will show how to obtain such threshold using Non-parametric quantile regression.

Dependence in threshold exceedances – comments

- In the previous analysis we assumed that threshold exceedances were independent.
- This is usually an unrealistic assumption for environmental applications, as threshold exceedances tend to cluster (heatwaves, storms, etc).
- You learned in the Risk Analytics course that one way deal with clusters of exceedances is to decluster, i.e., filtering the dependent observations to obtain a set of threshold excesses that are approximately independent. This works by:
    1. Using an empirical rule to define clusters of exceedances;
    2. identifying the maximum excess within each cluster;
    3. assuming cluster maxima to be independent, with conditional excess distribution given by the generalized Pareto distribution;
    4. fitting the generalized Pareto distribution to the cluster maxima.

Dependence in threshold exceedances – comments

Is declustering a good idea?

- Declustering is still widely used, but it has a considerable drawback: is wasteful of data.
- Moreover, Fawcett ad Walshaw (2007) showed that declustering is liable to incur serious bias in the estimation of parameters, as well as the return levels.

How dependence affects estimation?

- Point estimates obtained by ML are usually unchanged.
- The major effect of dependence is seen in the standard errors associated with parameter estimates: they are usually *too optimistic*, i.e., too small. This means we underestimate the uncertainty of the estimates.
- So, if we want to use all exceedances (and avoid declustering), we need a way to *inflate* the standard errors to account for dependence in the data.
- Two ways in which this can be achieved:
  - Making appropriate adjustments to the standard errors (Smith, 1991): a modified covariance matrix for the parameters is produced using empirical information on dependence.
  - Explicitly modelling temporal dependence (see Fawcett and Walshaw (2006) for an application assuming a stationary first-order Markov chain within each season.

References

1. Fawcett, L., & Walshaw, D. (2006). Markov chain models for extreme wind speeds. *Environmetrics: The official journal of the International Environmetrics Society*, 17(8), 795-809.

2. Fawcett, L., & Walshaw, D. (2007). Improved estimation for temporally clustered extremes. *Environmetrics: The official journal of the International Environmetrics Society*, 18(2), 173-188.

3. Smith RL. 1991. Regional estimation from spatially dependent data. Preprint. (https://www.rls.sites.oasis.unc.edu/postscript/rs/re