Modelling extremes of environmental data

Malvina Bozhidarova

STOR-i, Lancaster University

September 4,2020

Why is extreme value theory important?

- Case study: Fukushima Daiichi Nuclear Power Plant.

Figure: https://www.industryglobalnews24.com/images/fukushima-joins-hand-with-chernobyl-in-the-renewable-revitalization.jpeg

Background Theory

- What is an "extreme event"?

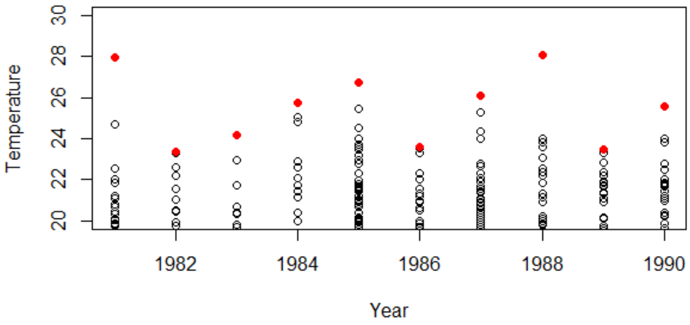

Figure: Year maxima

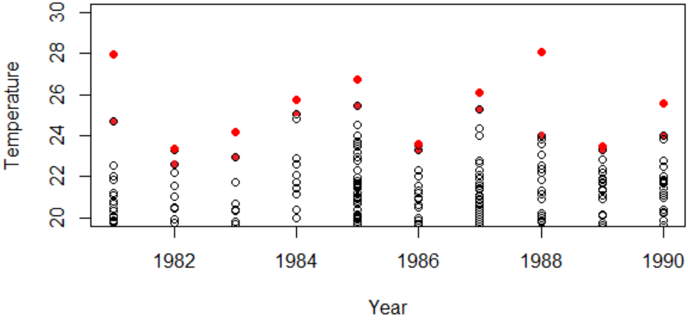

Figure: R-largest values

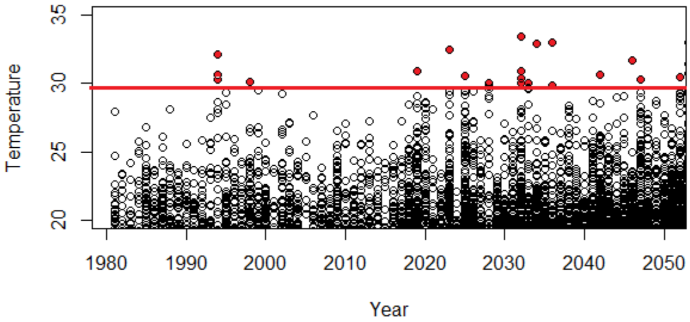

Figure: Peaks over threshold

Non-stationarity

- By definition, there is little data for extreme events, any analysis relies heavily upon theoretical results.
- When non-stationarity is present, the standard theory cannot be applied.
- Non-stationarity is a common feature in many environmental datasets (what we are working with).

Example of non-stationarity

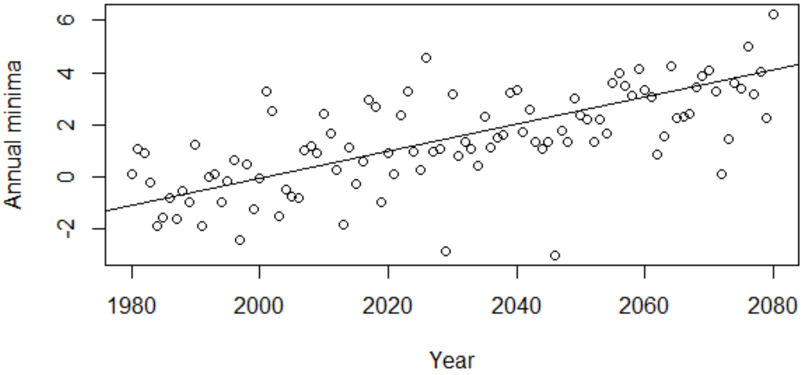

Figure: Long term trend

Our goal

- To model and simulate non-stationary extreme data
- To estimate non-stationary return levels
- To define and estimate uncertainty of non-stationary return levels

Return level plot

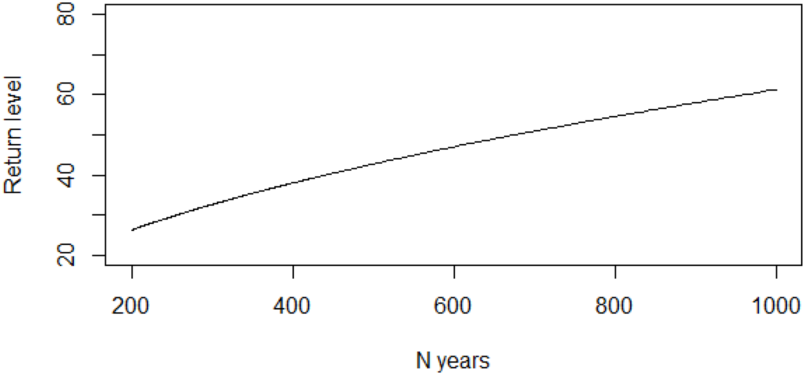

Figure: Return level plot for stationary data

Non-stationary return levels

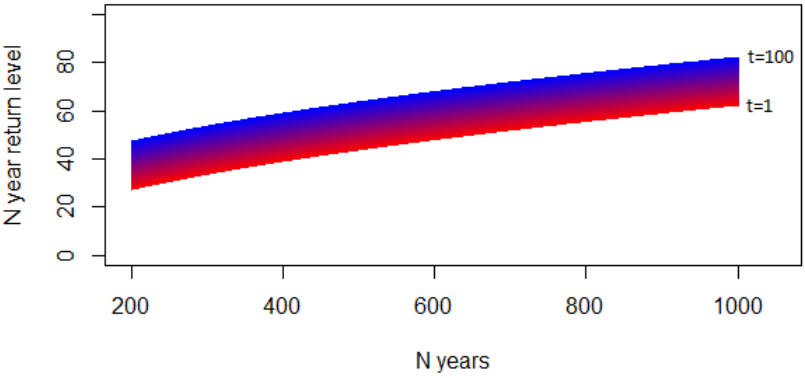

Figure: Return level plots over 100 years

Heysham Temperature Data

	Date	Maximum air temperature at 1.5m
106	1981-03-16	9.38296
107	1981-03-17	6.51260
108	1981-03-18	4.26260
109	1981-03-19	6.00015
110	1981-03-20	10.19670
111	1981-03-21	9.14102
112	1981-03-22	5.90493
113	1981-03-23	6.25869

Figure: Heysham temperature data

- Fit different non-stationary GEV models to the data.
- GEV model is used to model block maxima data.
- Compare models using likelihood ratio test.
- Best model appears to be GEV$(\mu_0 + \mu_1 t, \sigma, \xi)$.
- MLE estimates are:

$\hat{\mu_0}$	23.20662050
$\hat{\mu_1}$	0.07911825
$\hat{\sigma}$	3.30779629
$\hat{\xi}$	-0.36780430

Return level plots for each year

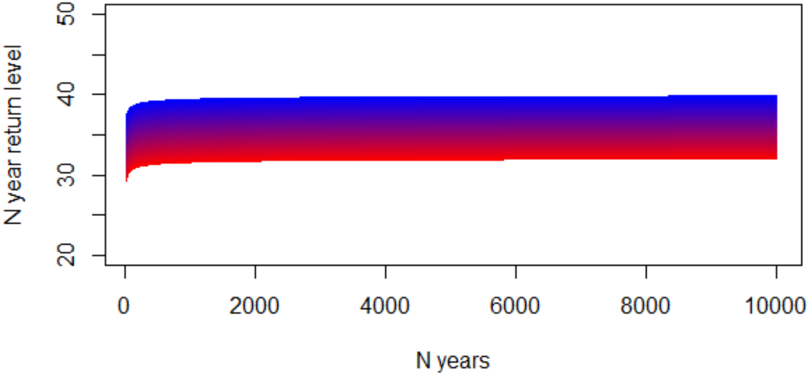

Figure: Return level plots for 1980-2080

Uncertainty of the return levels

- Parametric bootstrap
- Using projected data
- Scaling data and then bootstrap
- Delta method

Parametric bootstrap

- Fit model to annual maxima
- Using simulations from this model we refit GEV and obtain return level estimates.
- We get a sample of return level estimates for each year
- Find 95% CI for each year

Delta Method

$$\mathrm{Var}(\hat{z}_p) \approx \nabla z_p^T V \nabla z_p, \tag{3.11}$$

where V is the variance-covariance matrix of $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ and

$$\begin{aligned} \nabla z_p^T &= \left[\frac{\partial z_p}{\partial \mu}, \frac{\partial z_p}{\partial \sigma}, \frac{\partial z_p}{\partial \xi} \right] \\ &= \left[1, \ -\xi^{-1}(1 - y_p^{-\xi}), \ \sigma\xi^{-2}(1 - y_p^{-\xi}) - \sigma\xi^{-1} y_p^{-\xi} \log y_p \right] \end{aligned}$$

evaluated at $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$.

Figure: Cole's book, page 57, Delta Method

Using projected data

- Produce 12 different data sets
- Find return levels for each data set
- Build 95% CIs

```
1980-12-01,6.05337,7.84292,9.75039,12.1378,4.97939,11.7072,7.83413,2.57729,2.20303,7.27603,7.48257,9.21816
1980-12-02,4.65322,6.38247,6.19155,10.8007,4.9728,10.0375,6.31631,2.90029,5.17446,4.88101,8.10049,9.49087
1980-12-03,2.48013,5.46108,7.20278,11.9799,3.58291,7.76553,10.2621,1.08584,2.78726,1.41519,8.30606,9.40078
1980-12-04,2.85439,9.77554,5.4125,13.5553,2.2875,6.96377,10.3236,-5.22007,3.16982,1.26455,7.32168,11.3544
```

Figure: Heysham temperature projected data

Results

- All methods give very close results for the CIs.
- Using projected data gives wider CIs (due to higher variance).

Year	Lower	Upper
1980	28.20638	37.67593
1981	28.29646	37.77433
1982	28.38632	37.87295
1983	28.47597	37.97179
1984	28.56540	38.07084
1985	28.65462	38.17010
1986	28.74363	38.26958
1987	28.83242	38.36927
1988	28.92100	38.46917
1989	29.00936	38.56929
1990	29.09752	38.66961
1991	29.18546	38.77015
1992	29.27320	38.87090
1993	29.36073	38.97185
1994	29.44805	39.07301

Figure: 95% CIs for 10 000-year return level, using projected data

Year	Lower	Upper
1980	30.05475	33.08563
1981	30.14567	33.15295
1982	30.23644	33.22042
1983	30.32705	33.28805
1984	30.41750	33.35583
1985	30.50778	33.42378
1986	30.59791	33.49190
1987	30.68786	33.56018
1988	30.77764	33.62864
1989	30.86724	33.69728
1990	30.95666	33.76609
1991	31.04589	33.83510
1992	31.13493	33.90429
1993	31.22378	33.97368
1994	31.31243	34.04327

Figure: 95% CIs for 10 000-year return level, using delta method

Results

- For all 4 methods we see that for higher values of N, we get that CIs for 1980 and 2080 become closer.

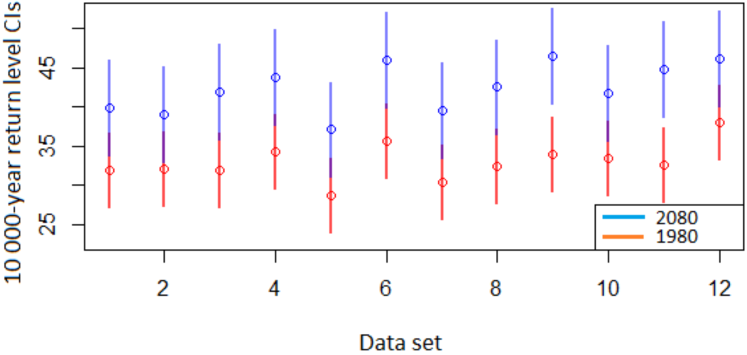

Figure: 95% CIs for 10 000-year return level intersection, using projected data

Results using projected data

- For higher-year return level we get larger overlap of the CIs

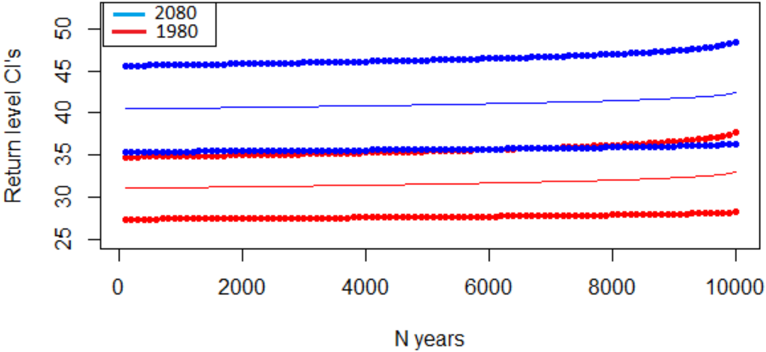

Figure: CI's for 1000 to 10 000-years return levels

Future ideas

- Do the same analysis for all different years
- Try to use r-largest values or threshold exceedance methods
- Do analysis for different data (e.g.humidity data)

- Thank you for listening!
- Any questions?