Extreme value analysis of output from complex environmental models

TIES, 2007, Mikulov

Adam Butler, Biomathematics & Statistics Scotland
(adam@bioss.ac.uk)

Complex environmental models

Atmosphere-Ocean General Circulation Models (climate)

Dynamic Global Vegetation Models (ecology)

Rainfall-Runoff models (hydrology)

Storm Surge Prediction Models (oceanography)

...and many others

- **Common features**

Process-based, mechanistic models of complex systems

Used to generate predictions of future environmental change

Evaluated numerically, often highly computer intensive to run

Spatio-temporal grid, outputs often high dimensional

- **Uncertainty and calibration**

Uncertainty analysis: propagation of uncertainty about inputs into uncertainty about outputs

Model calibration: quantifying model inadequacy by comparing outputs against observational data

- **Extreme events**

Increasing recognition that prediction of extreme events is a critical aspect of environmental risk assessment

Interest in model-based simulation of extreme events

Relatively little statistical work in this context

Example: a storm surge model

- Sea level = Mean sea level + Tide + **Surge** + Waves
- Surge is generated by wind and air pressure
- It is a key factor in coastal flooding:

  e.g. 1953 North Sea, 1993 Bangladesh
- **CSX**:

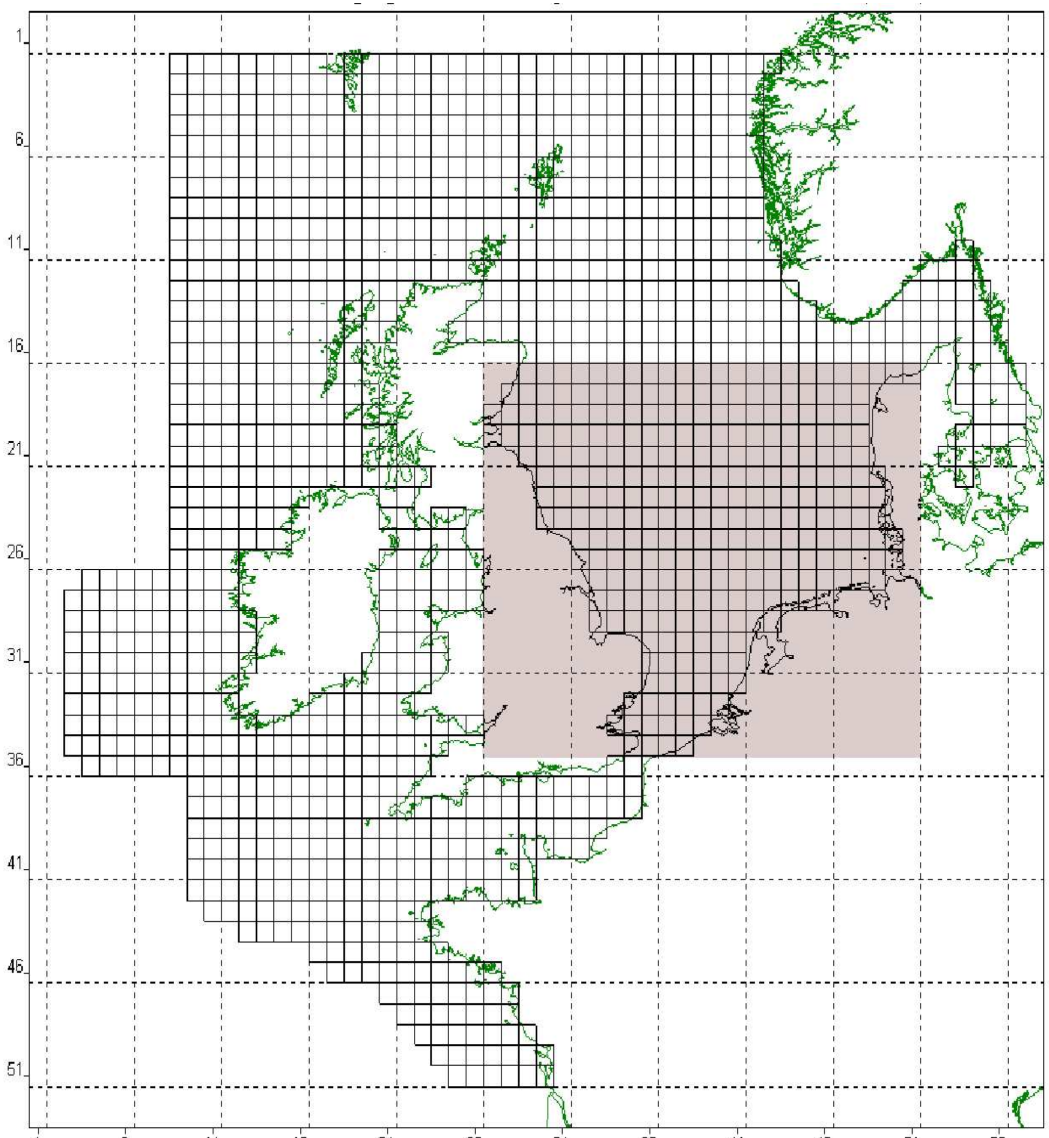
  a storm surge prediction model

  for the European Continental Shelf

  developed the Proudman Oceanographic Laboratory

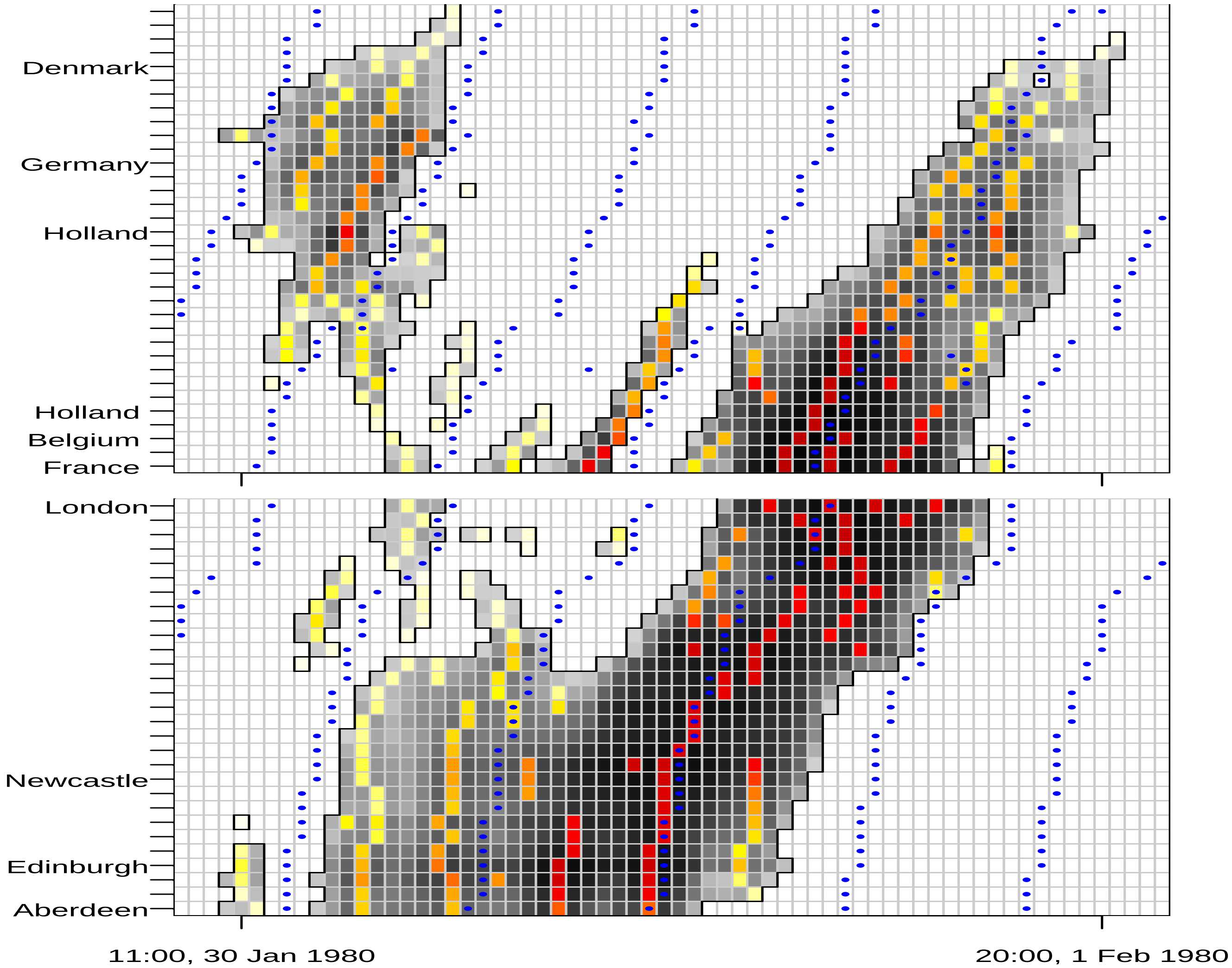
Spatial area covered by the CSX model

North Sea highlighted in purple

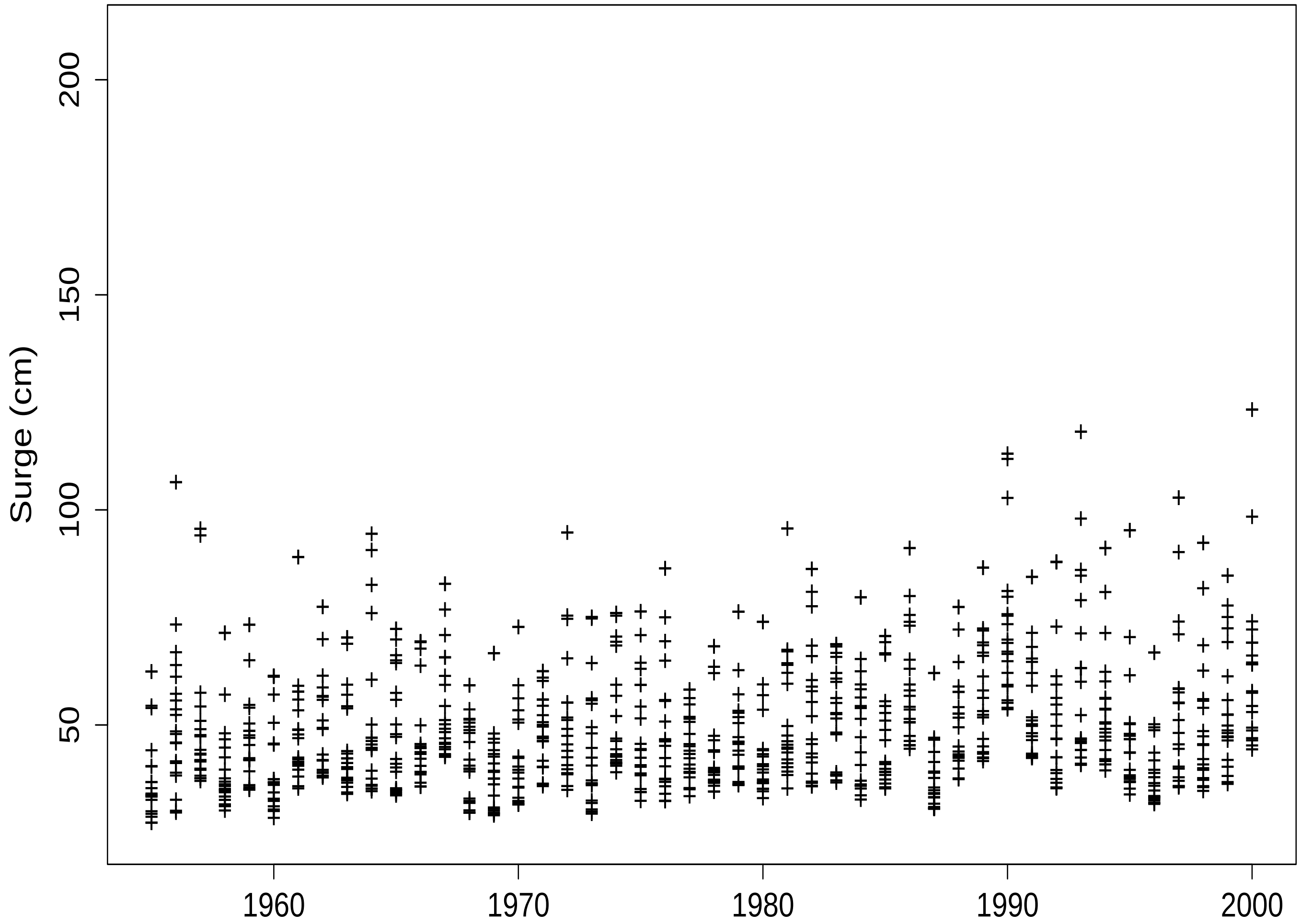


Simulated evolution of a storm surge

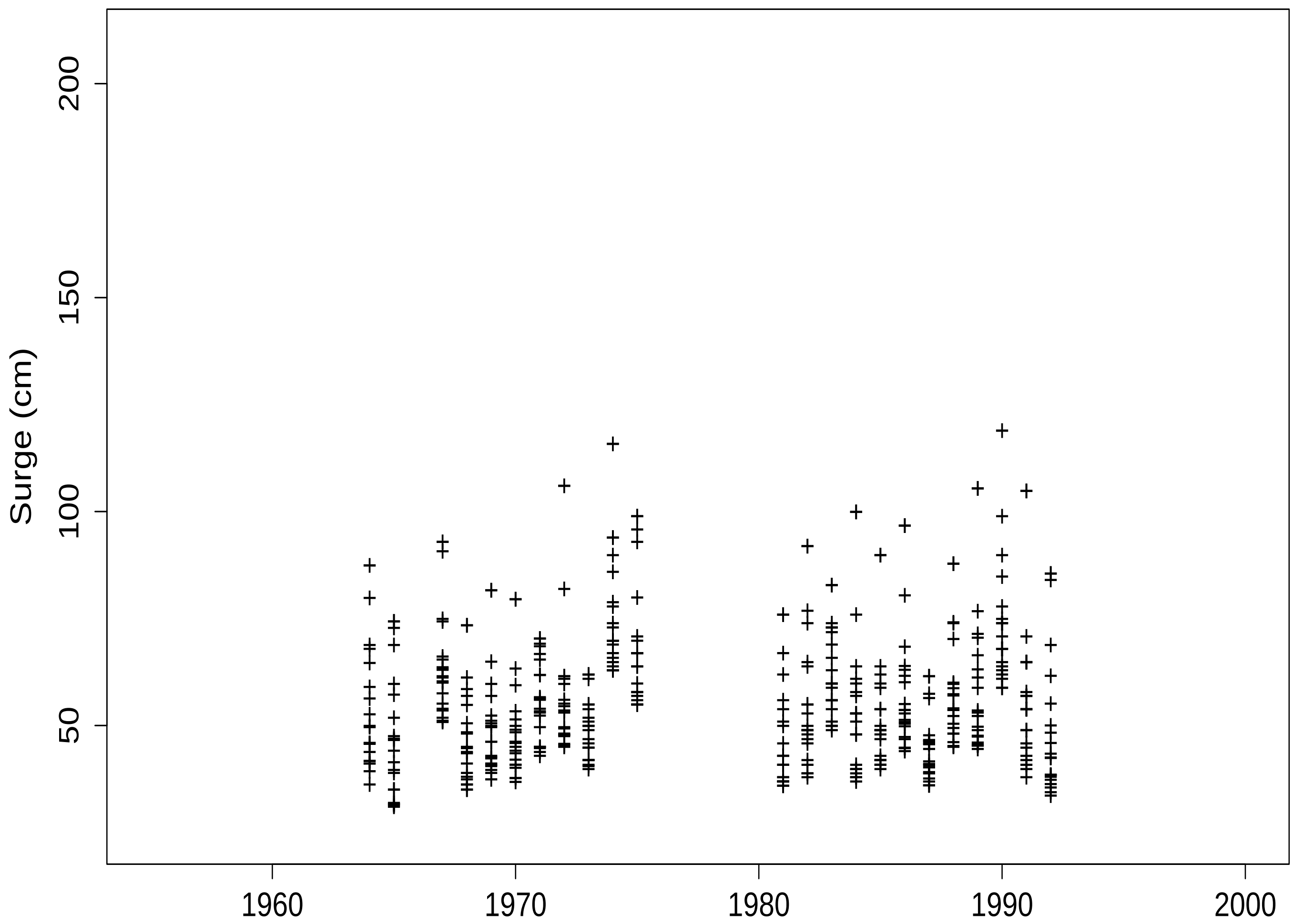
grey-yellow-red = extreme surge levels; blue = high tides



20 largest surges per year at Aberdeen: CSX model output



20 largest surges per year at Aberdeen: observational data



Extreme value theory

A methodology for drawing inferences about the statistical properties of rare events, using only data that refer to relatively extreme levels

Extreme value models are motivated by asymptotic theory, so EVT provides a relatively robust basis for extrapolation

Widely used in hydrology, finance and climatology

Application to model output is relatively recent, and throws up new methodological challenges

Typical extreme value problems

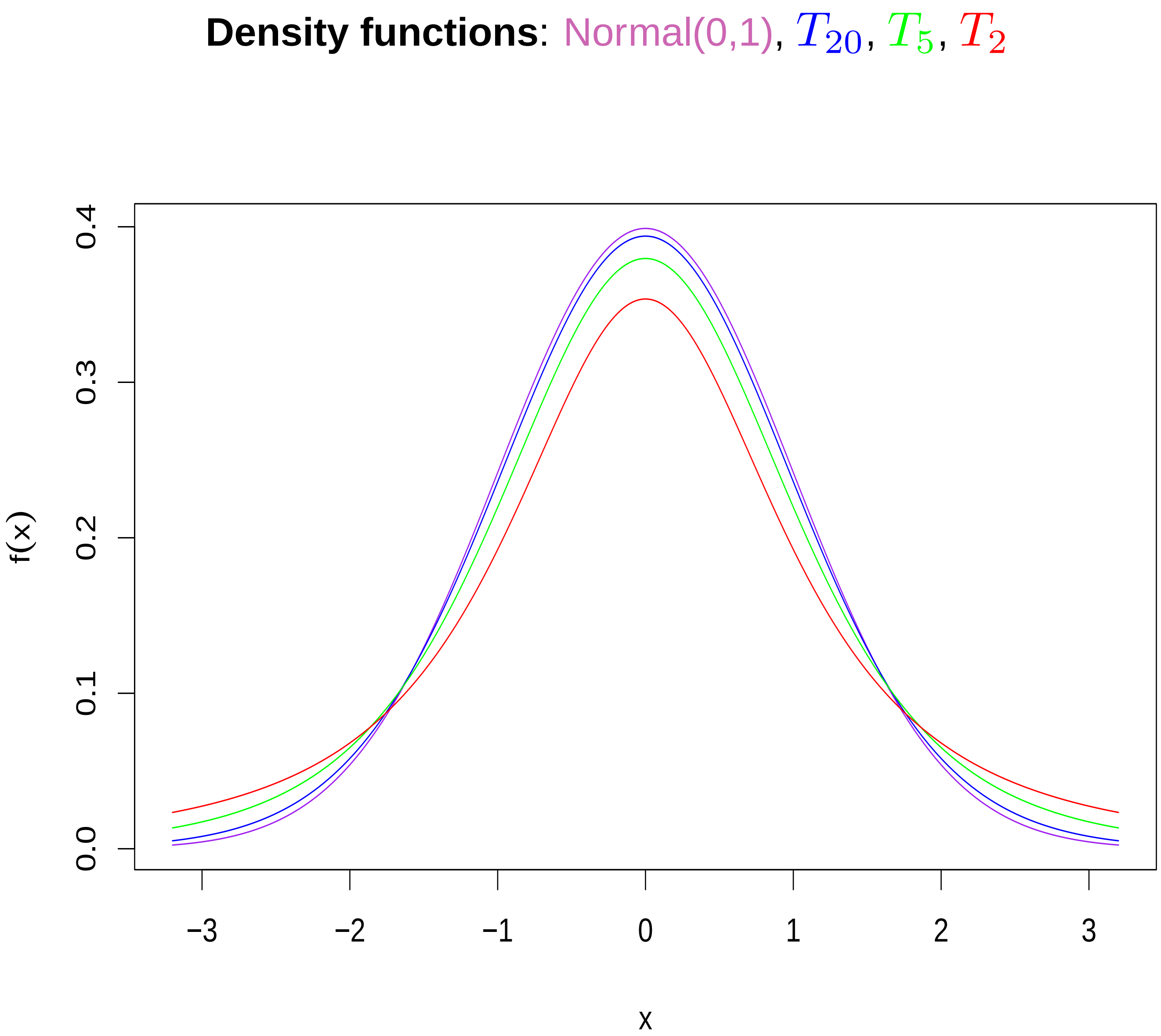
What is the probability, $\mathbf{P}(X > u)$, that a process X exceeds a threshold u during a particular period?

What is the value of the **return level**

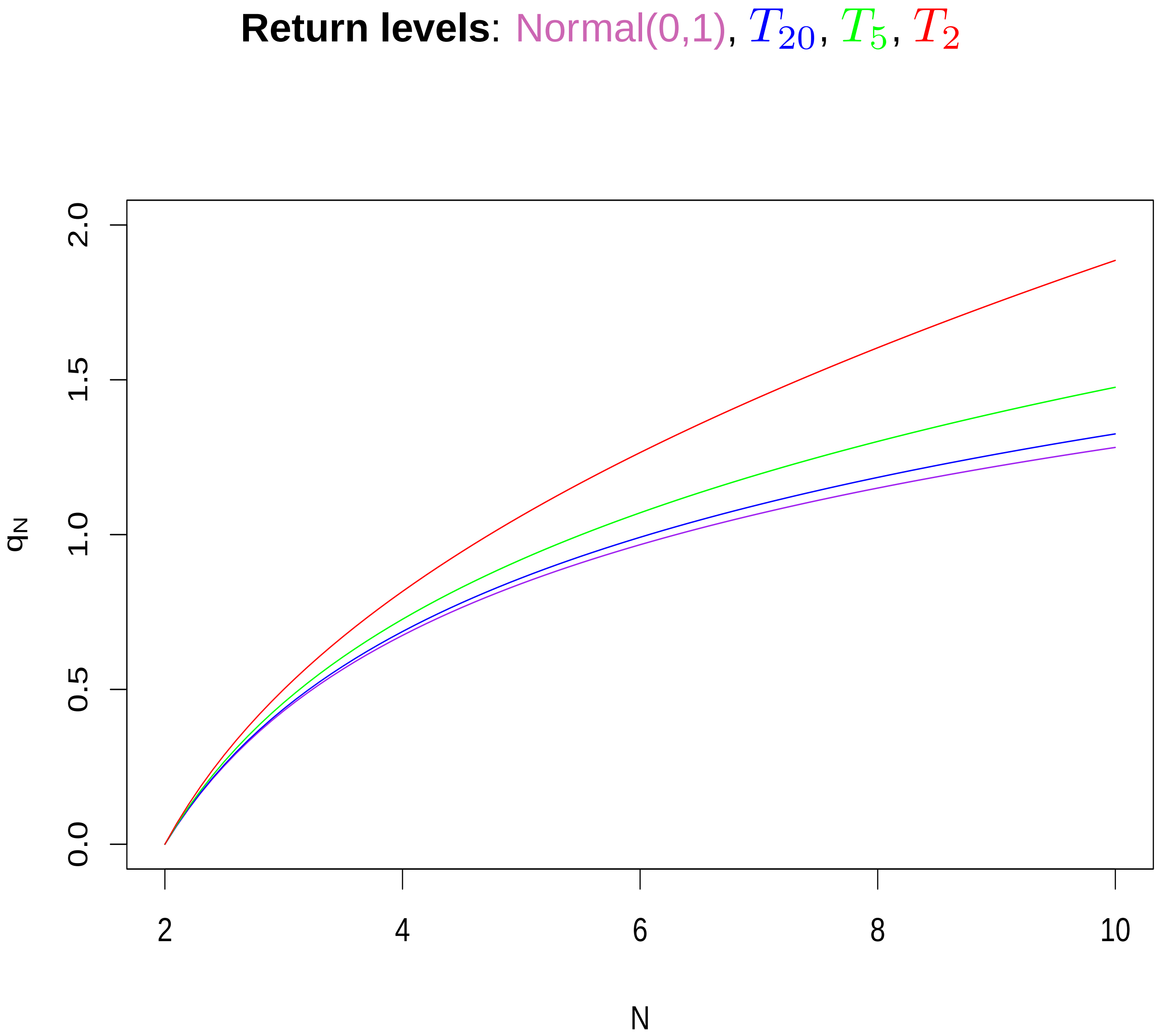
$$q_N = \{u : \mathbb{P}(X > u) = 1/N\}?$$

u and N will usually be large, so we are estimating the tail of the distribution of X.

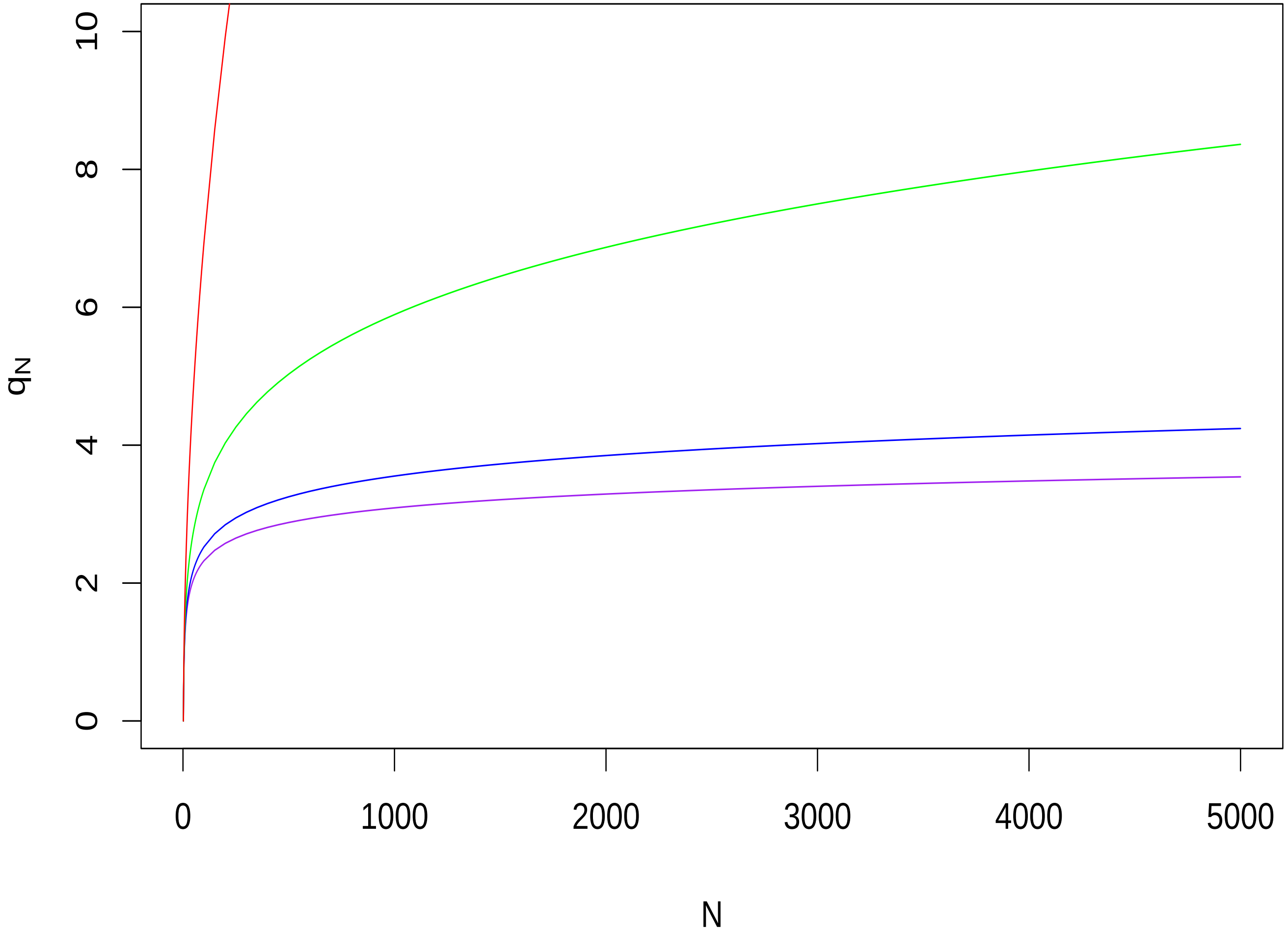
Little or no data are available on such rare events - we must extrapolate.



Density functions: Normal(0,1), T_{20}, T_5, T_2
f(x)
0.0
0.1
0.2
0.3
0.4
−3
−2
−1
0
1
2
3
x


Return levels: Normal(0,1), T_{20}, T_5, T_2
q_N
N
0.0
0.5
1.0
1.5
2.0
2
4
6
8
10

Return levels, again: Normal(0,1), T_{20}, T_5, T_2



A classic result

Consider an iid sequence $\{X_i : i = 1, 2, \ldots\}$

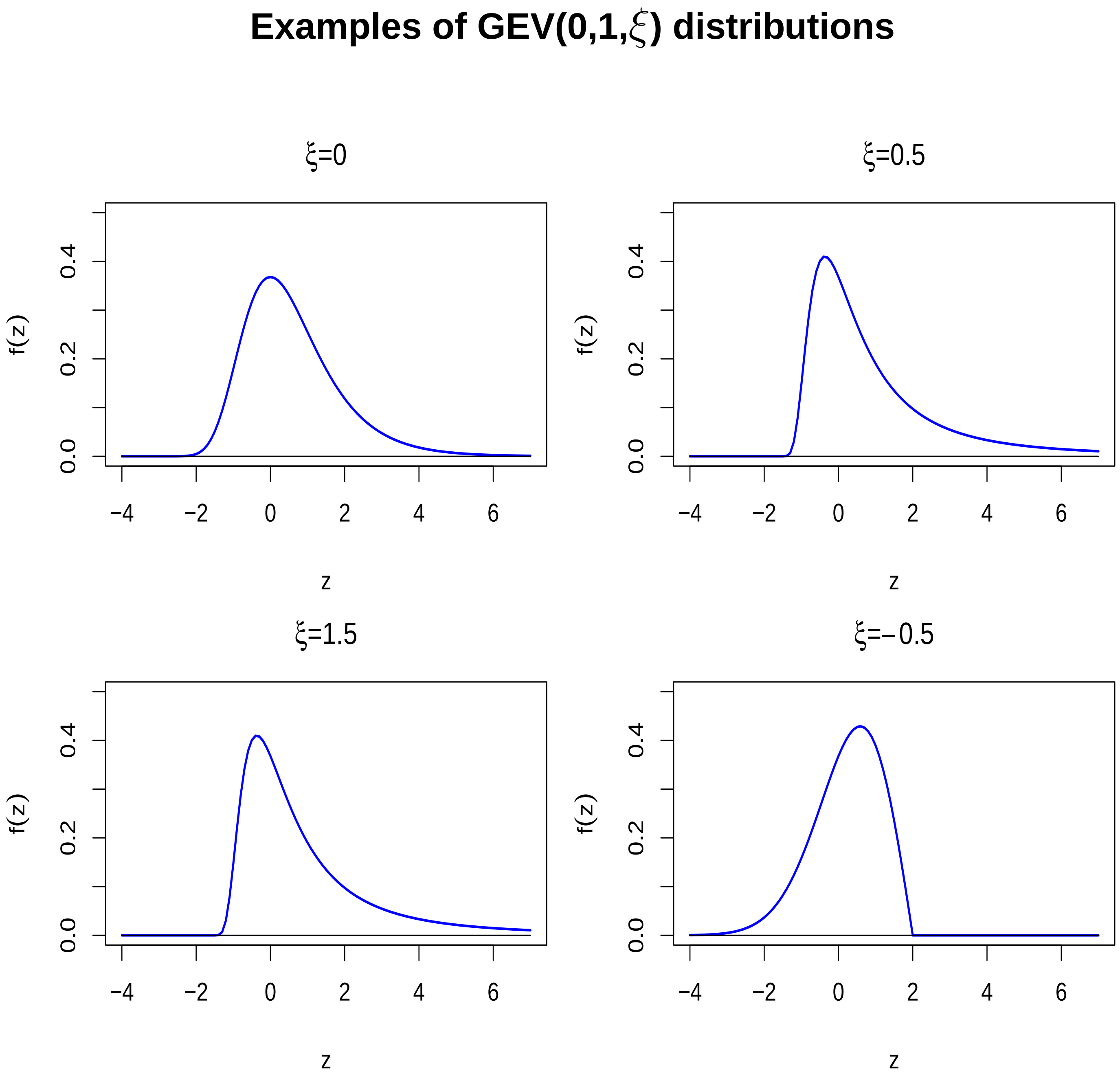
Under weak conditions, there exist $\{a_n\}$ and $\{b_n\}$ such that

$$F(z) = \lim_{n \to \infty} \mathbb{P}\left[\frac{\max(X_1, \ldots, X_n) - a_n}{b_n} \leq z\right]$$

is nondegenerate.

Then $F(z)$ has a Generalised Extreme Value (GEV) distribution with parameters μ, σ and ξ:

$$F(z) = \exp\left[-\left\{1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right\}^{-1/\xi}\right].$$


Examples of GEV(0,1,ξ) distributions
ξ=0
ξ=0.5
ξ=1.5
ξ=−0.5
f(z)
0.0
0.2
0.4
−4
−2
0
2
4
6
z

Statistical inference & modelling

Apply GEV model to block maxima

Estimate model parameters by maximum likelihood

Thereby estimate return levels and exceedance probabilities

Other extreme value models

- Point process model describes the rate and distribution of exceedances of a high threshold u
- r-largest model describes the r largest values within a block, and generalises the GEV
- Both can be expressed in terms of the GEV parameters

$$\theta = (\mu, \sigma, \xi)$$

Key assumptions

An asymptotic model is valid at a finite but extreme level

We are looking at a sufficiently extreme level (u, n or r)

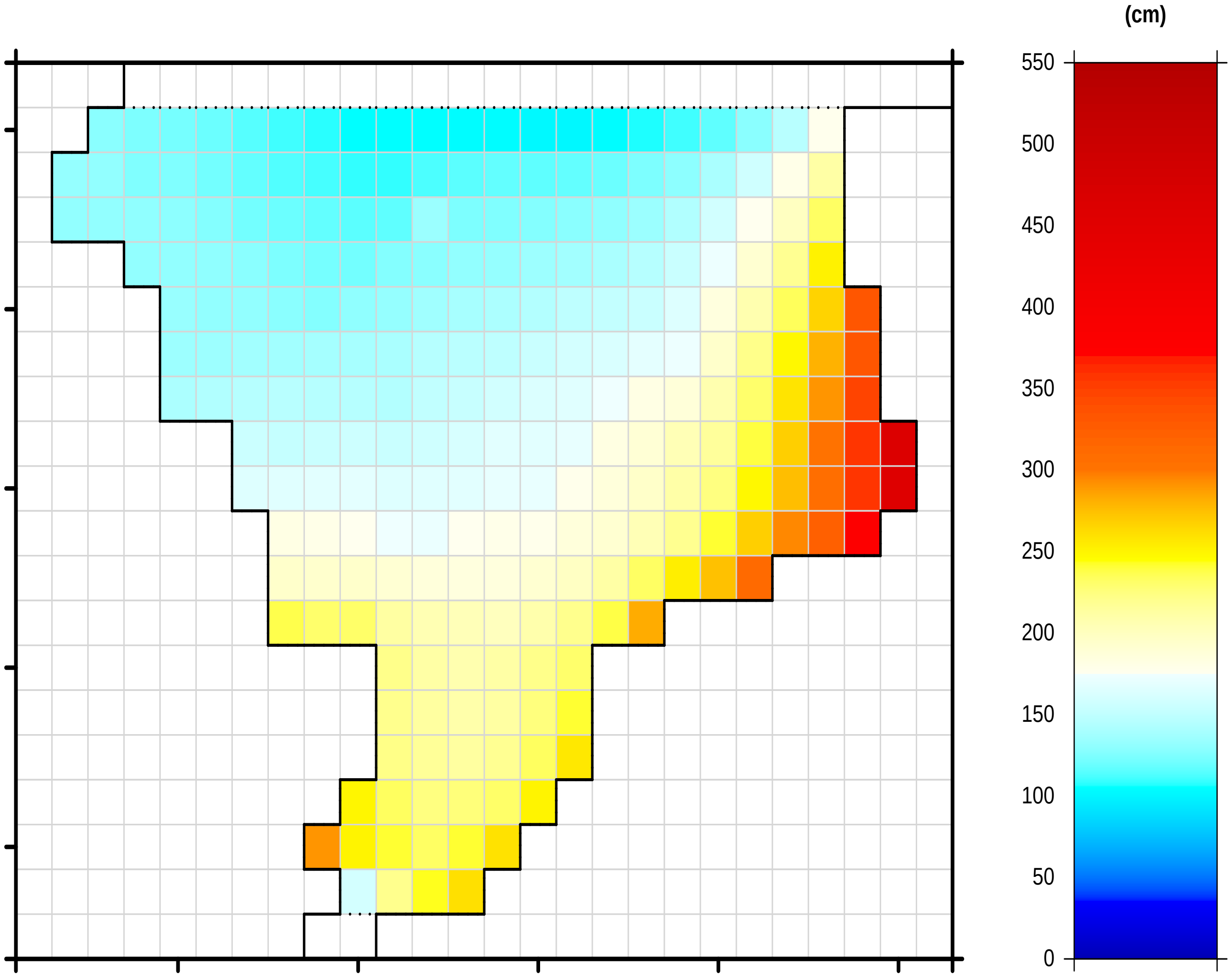
Extreme events are drawn from a common population

The data are free from outliers

Application to surges (Flather et al., 1998)

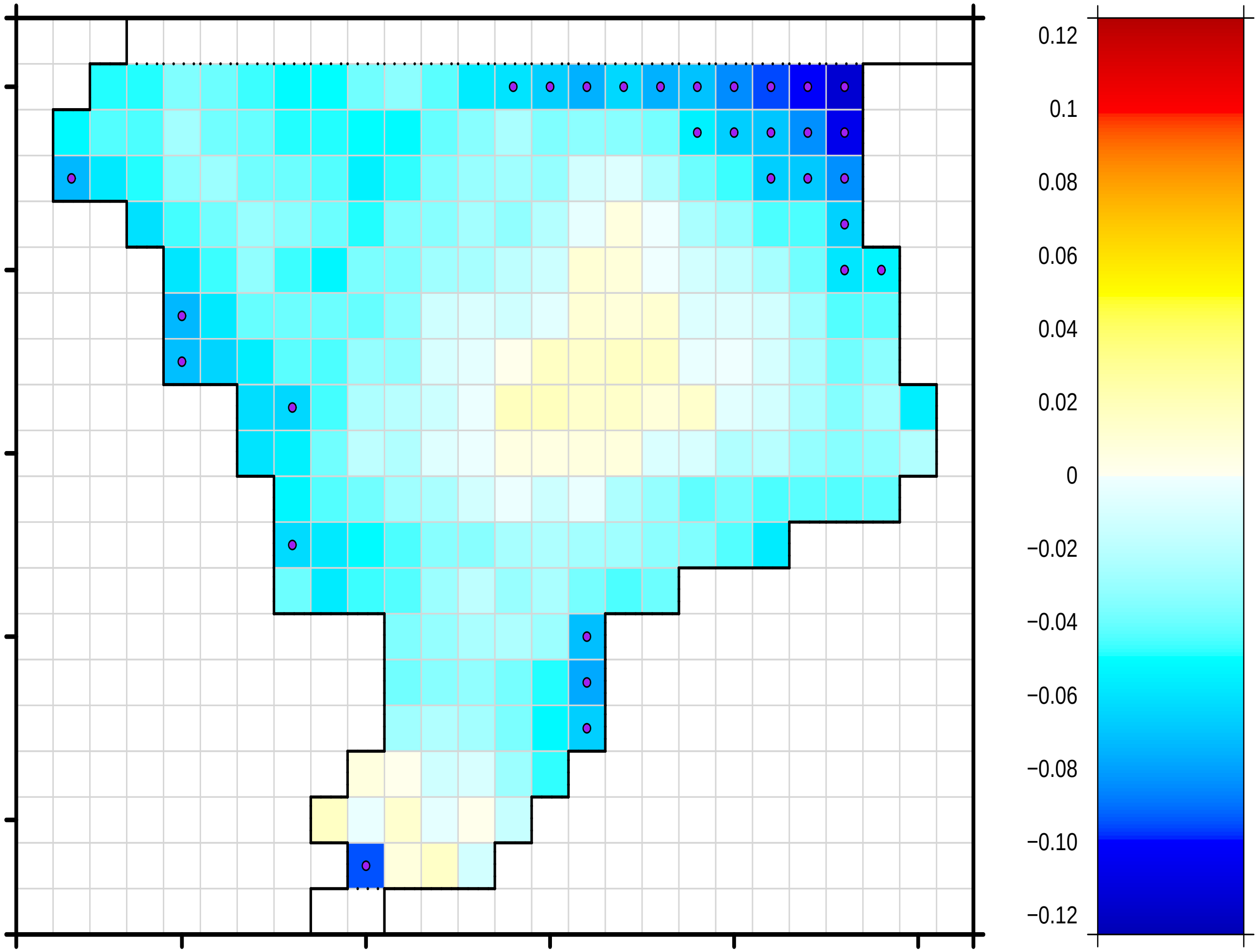
- Use the CSX model to generate a reconstruction of the storm climate for 1955-2000
- Apply EVT to output for each cell, modelling the r-largest storm surges per year
- Also apply to observational data, where available
- 50 year return levels estimated

Estimated 50y return levels, based on an r-largest fit at each site



Shape parameter estimates, based on an r-largest fit at each site

Pink circles denote values significantly different from zero



Trends in extremes

Interest in detecting and estimating changes in the characteristics of extreme environmental events

e.g. let Z_t denote the largest event in year t, and assume that $Z_t \sim$GEV(θ_t)

Recent papers have used nonparametric regression to describe trends over time in the GEV parameters, θ_t:

- Local likelihood: Davison & Ramesh (2000); Hall & Tajvidi (2000)
- Penalised likelihood: Pauli & Coles (2001); Chavez-Demoulin & Davison (2005)
- Bayesian state space modelling: Gaetan & Grigoletto (2004)

Trends in storm surges (Butler *et al.*, 2007a)

- Use local likelihood version of r-largest model
- Estimate θ_t by maximising

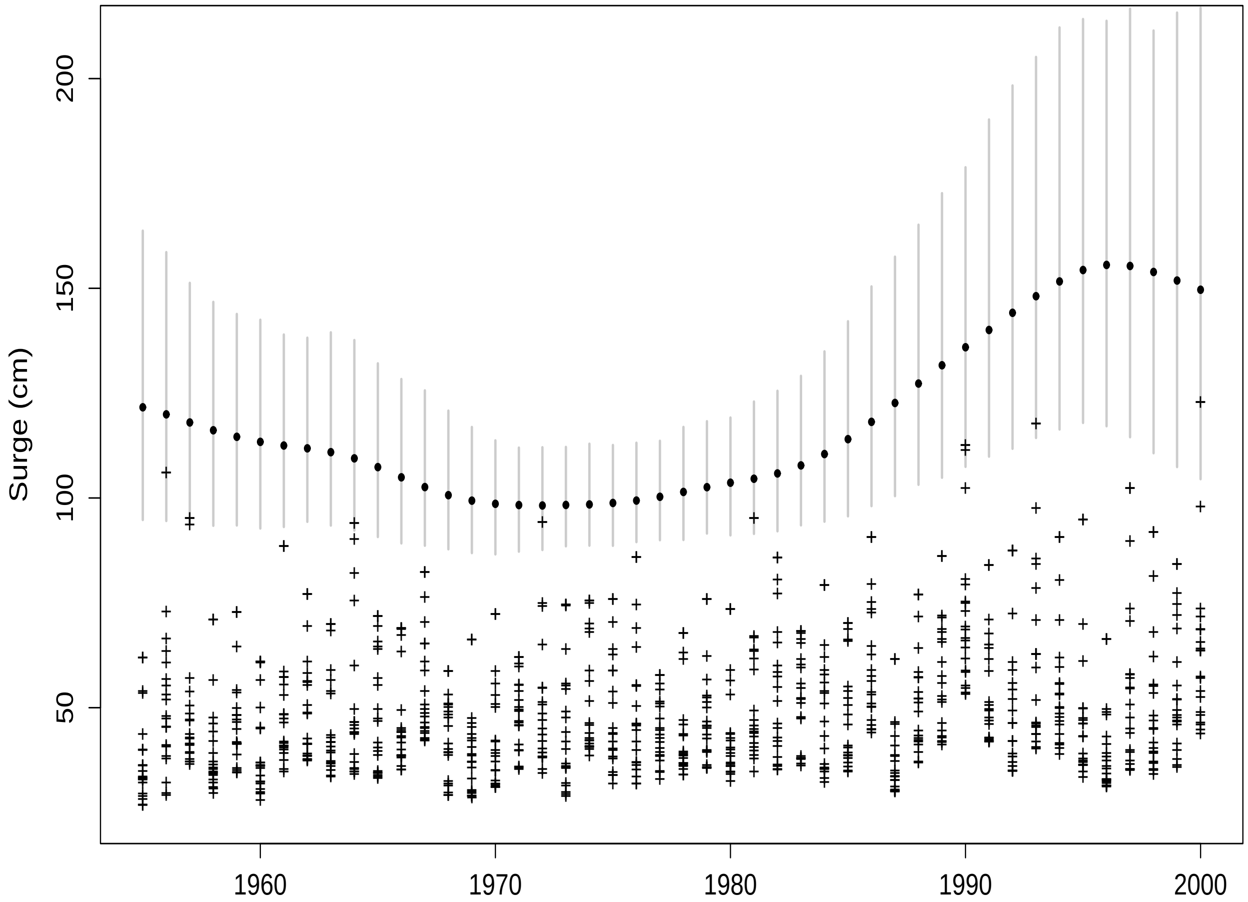
$$\sum_T w_{tT} \log f(z_T; \theta_t)$$

- Use a Gaussian kernel, $w_{tT} = \phi([t - T]/h)$
- Assess sensitivity to values of r and the bandwidth h
- Also consider models in which only some parameters are time varying

Estimated trends in 50 year return levels

Model output for Aberdeen

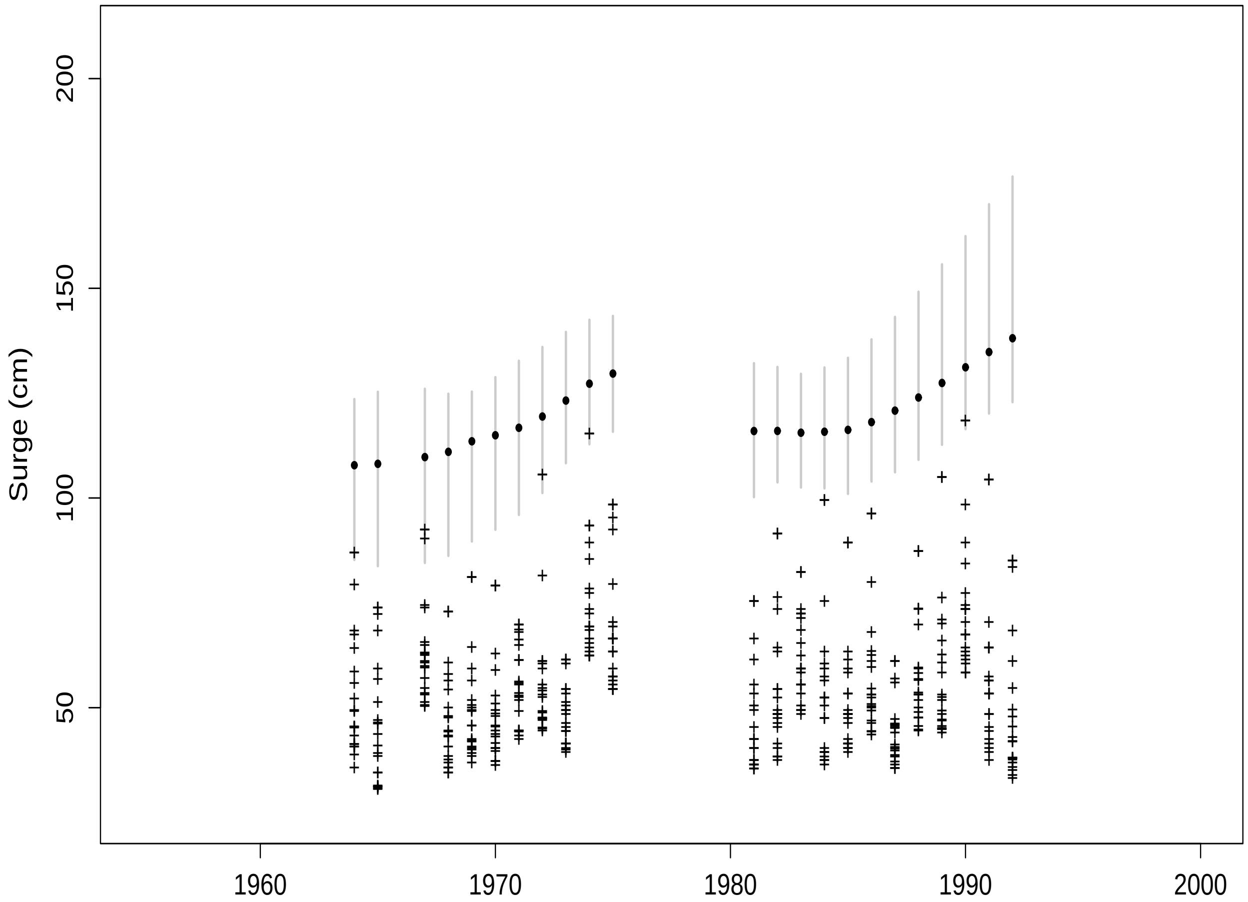
95% confidence intervals also shown (grey)



Estimated trends in 50 year return levels

Observational data for Aberdeen

95% confidence intervals also shown (grey)

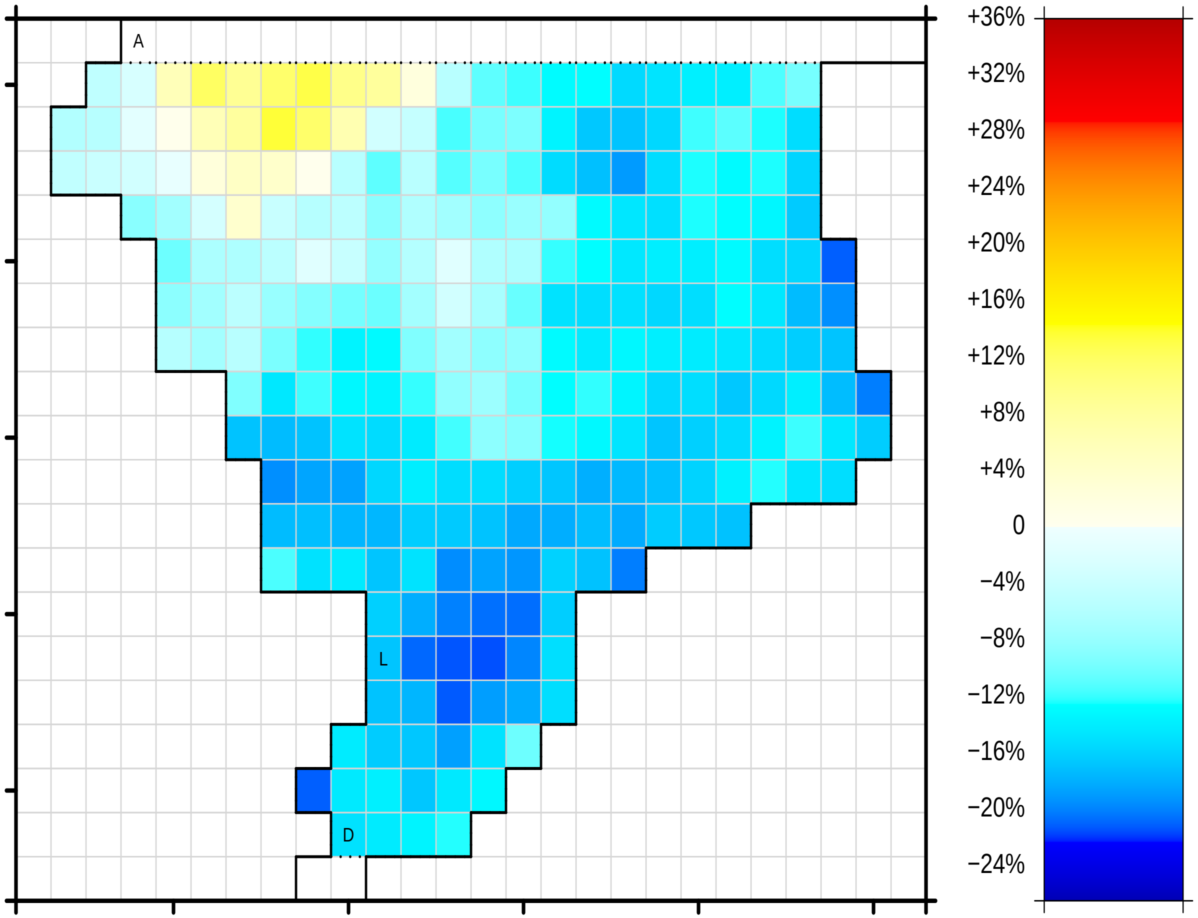


Spatio-temporal smoothing (Butler *et al*., 2007b)

- More efficient estimation of trends by smoothing parameters θ_{st} over space as well as time
- Need to ensure baseline spatial pattern is preserved

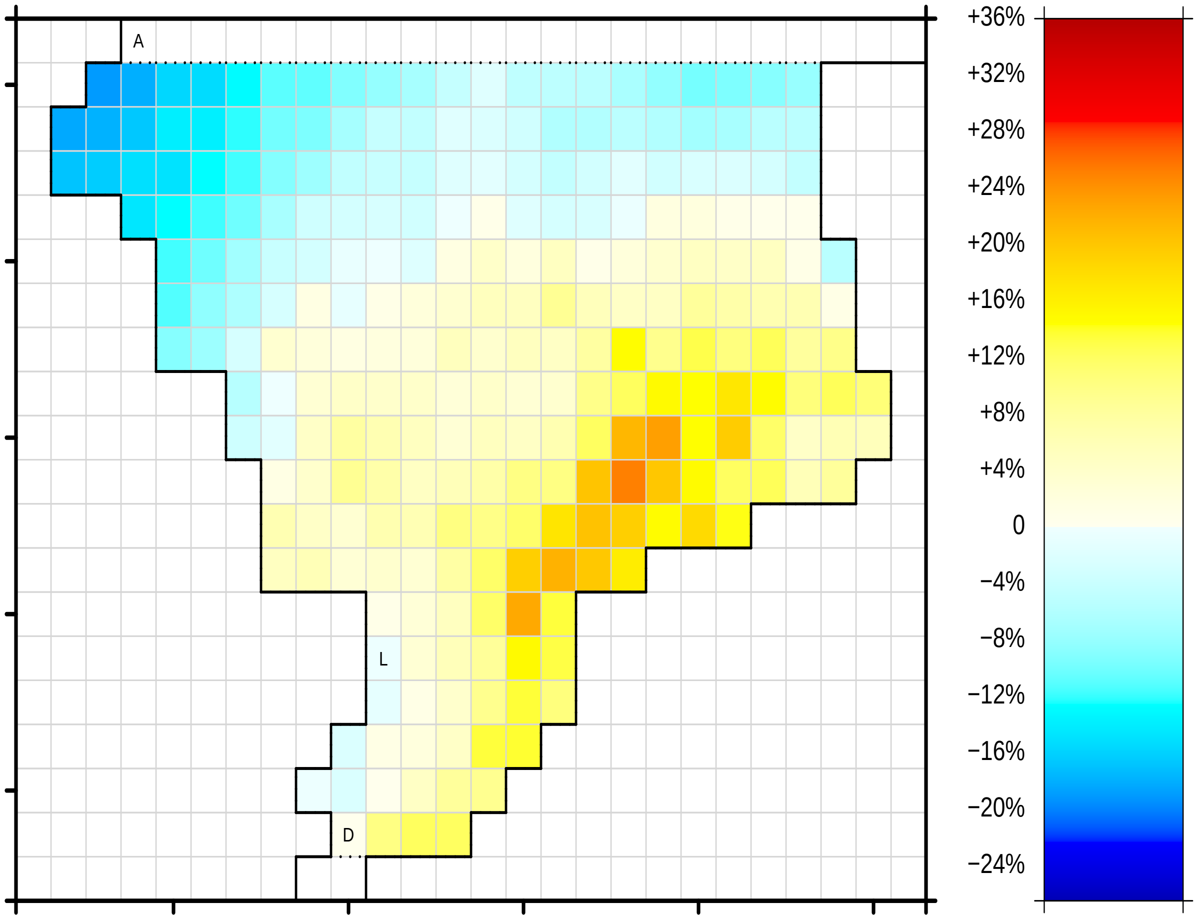
Spatio-temporal trends in estimated 50y return levels

1955: % deviation from time-constant estimate of q_{50}



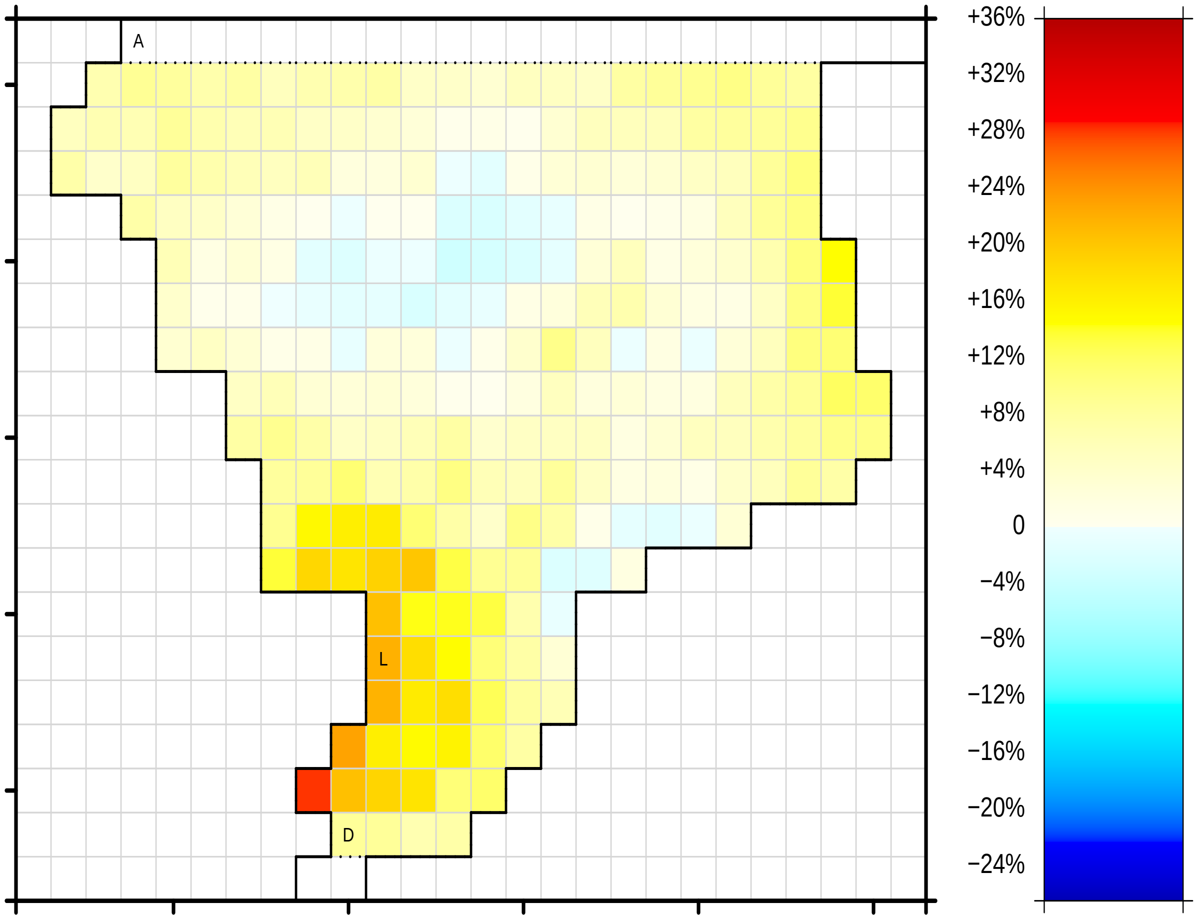
Spatio-temporal trends in estimated 50y return levels

1975: % deviation from time-constant estimate of q_{50}



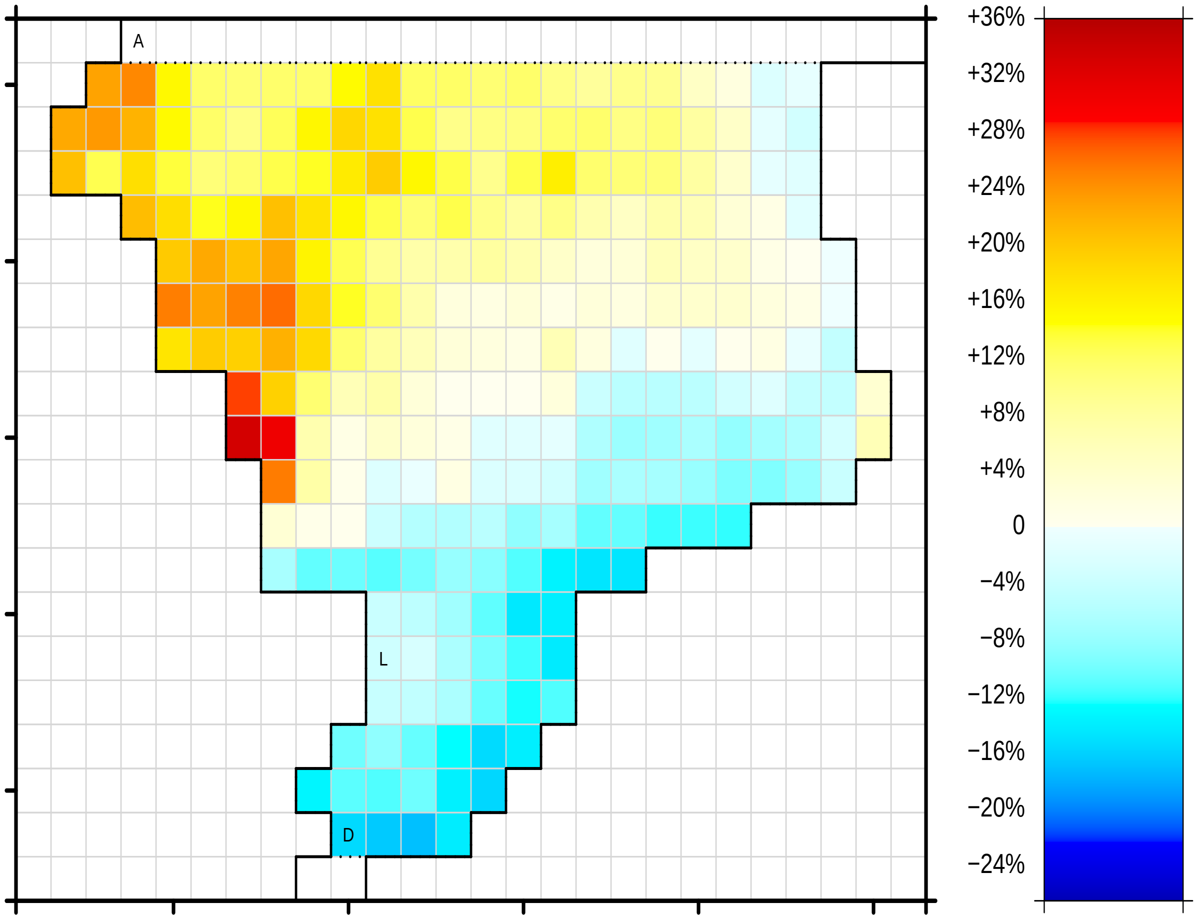
Spatio-temporal trends in estimated 50y return levels

1990: % deviation from time-constant estimate of q_{50}



Spatio-temporal trends in estimated 50y return levels

2000: % deviation from time-constant estimate of q_{50}



Spatio-temporal smoothing (Butler *et al*., 2007b)

- A modified version of local likelihood, based on reparameterisation

- Estimate θ_{st} by maximising

$$\sum_{S} \sum_{T} w_{sStT} \log f(\mathbf{z}_{ST}; \theta^*_{st} b_S),$$

over θ^*_{st} for fixed b_S.

Statistical challenges

Analysing a single model run using EVT

- Dealing with complex trends over space and time
- Spatial and temporal dependence, multivariate extremes, regional risk assessment (Latham, 2006)
- Automatic identification of extreme events

Analysing multiple runs

Combine EVT with existing statistical methods for analysis of complex models - emulation and calibration

Emulation of extremes

- Simulate input λ from a prior distribution $\pi(\lambda)$
- Generate output $G(\lambda)$ by running the model, G
- Assume an extreme value model, $\max(G(\lambda)) \sim \text{GEV}(\theta_\lambda)$
- Assume a Gaussian process for the parameters, $\theta_\lambda \sim N(\mu, \Sigma)$.

Some references

Coles, S.G. (2001) *An Introduction to Statistical Modelling of Extreme Values.* Springer.

Davison, A. C. & Ramesh, N. I. (2000) Smoothing sample extremes. *J. Roy. Statist. Soc. Ser. B*, **92**, 191-208.

Butler, A., Heffernan, J.E., Tawn, J.A., Flather, R.A. and Horsburgh, K.J. (2007) Extreme value analysis of decadal variations in storm surge elevations. *Journal of Marine Systems*, **67**, 189-200.

Butler, A., Heffernan, J.E., Tawn, J.A. & Flather, R.A. (2007) Trend estimation in extremes of North Sea surges. *J. Roy. Statist. Soc. Series C*, **56**(4).

Latham (2006) *Statistical methodology for the extreme values of dependent processes.* PhD thesis, Lancaster University.