Analysis of
Extremes of Synthetic Surge Data

Climate extremes and risk reduction,
University College London,
30 November 2004

Jonathan Tawn and Adam Butler
(Lancaster University)

Joint work with:

Janet Heffernan (Lancaster University)

Roger Flather (Proudman Oceanographic Lab.)

Overview of storm surge work

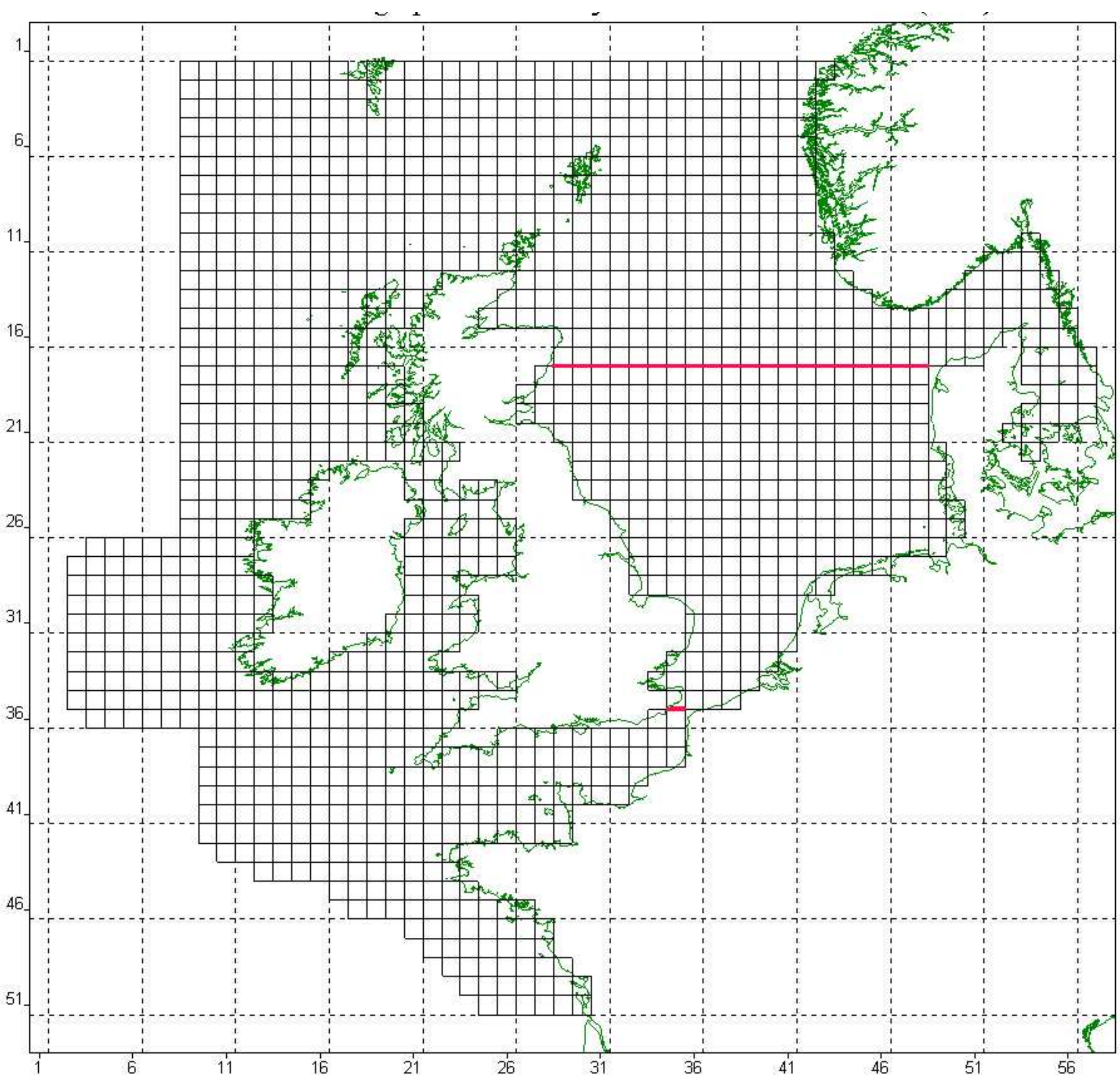
- Sea level = Mean sea level + Tide + Surge + Tide-surge interaction + Waves
- Storm surges are a source of coastal flood risk
- Hydrodynamical models are used to study surges
- We use spatial extreme value methods to analyse surges generated by a hydrodynamical model

Synthetic storm surge data

- We analyse hindcast output from the CSX model, a 2d numerical storm surge model for the European Continental Shelf
- Model forcing provided by DNMI pressure data for the period 1955-2000
- Previously analysed by Flather et al. (1998)

The **CSX** model grid

- Data for n years $t_1, \ldots, t_n$ at d sites $\mathbf{s}_1, \ldots, \mathbf{s}_d$
- $n = 46$, $d = 259$



The extreme value approach

- EV methods “let the tails speak for themselves”
- EV models are asymptotically-motivated, and so provide a robust basis for extrapolation
- EV models are parametric

The GEV model

- Annual maxima are often assumed to follow a Generalised Extreme Value (GEV) distribution
- The GEV has distribution function

$$G(x) = \exp\left[-\left\{1+\xi\left(\frac{x-\mu}{\sigma}\right)\right\}_+^{-1/\xi}\right],$$

  where μ, $\sigma > 0$ and ξ are location, scale and shape parameters respectively
- A design parameter is the level x such that

$$\mathbb{P}\left[\text{Annual max} > x\right] = 1/N;$$

  the "N-year return level"

The r-largest model

- Extension of the GEV model
- Uses r observations per year
- $x_{ij}^{(k)}$:= k-th largest peak in year t_j at site $\mathbf{s}_i$
- Assume $\mathbf{X}_{ij} := \left(X_{ij}^{(1)}, \ldots, X_{ij}^{(r)}\right)$ are jointly distributed according to the r-largest model (Smith, 1986; Tawn, 1988)
  with parameters $\theta_{ij} := (\mu_{ij}, \sigma_{ij}, \xi_{ij})$
- More robust to spatial and temporal variations than a threshold-based approach

Time-constant r-largest model

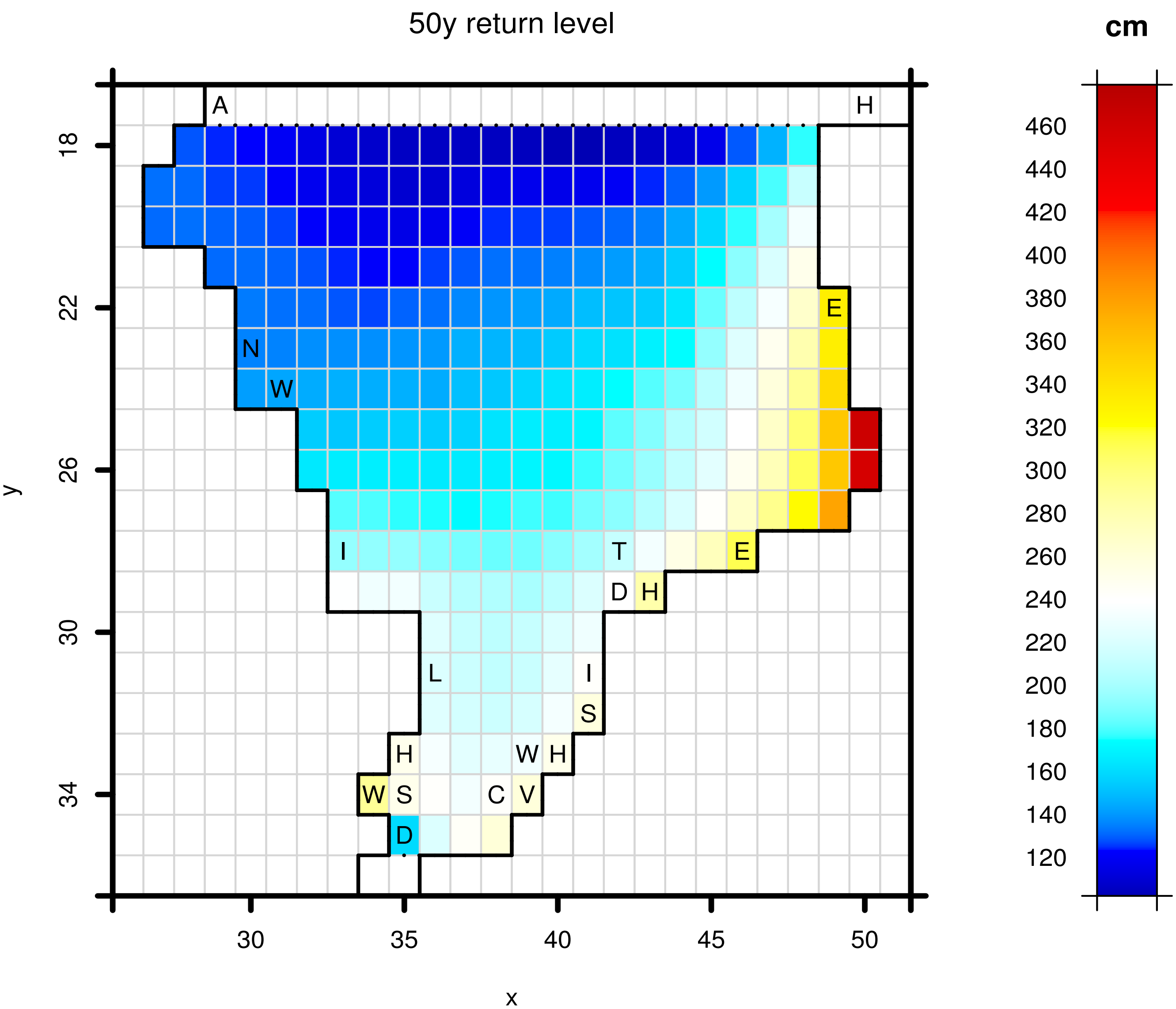
- Site-by-site analysis:
  Flather et al. (1998) assume that

  $$\theta_i := \theta_{i1} = \ldots = \theta_{in} \text{ for each site } \mathbf{s}_i$$

- Parameters constant in time, varying over space
- Flather et al. (1998) take $r = 7$
- We take $r = 20$, based on standard diagnostics

Spatial distribution of **50 year return level** estimates

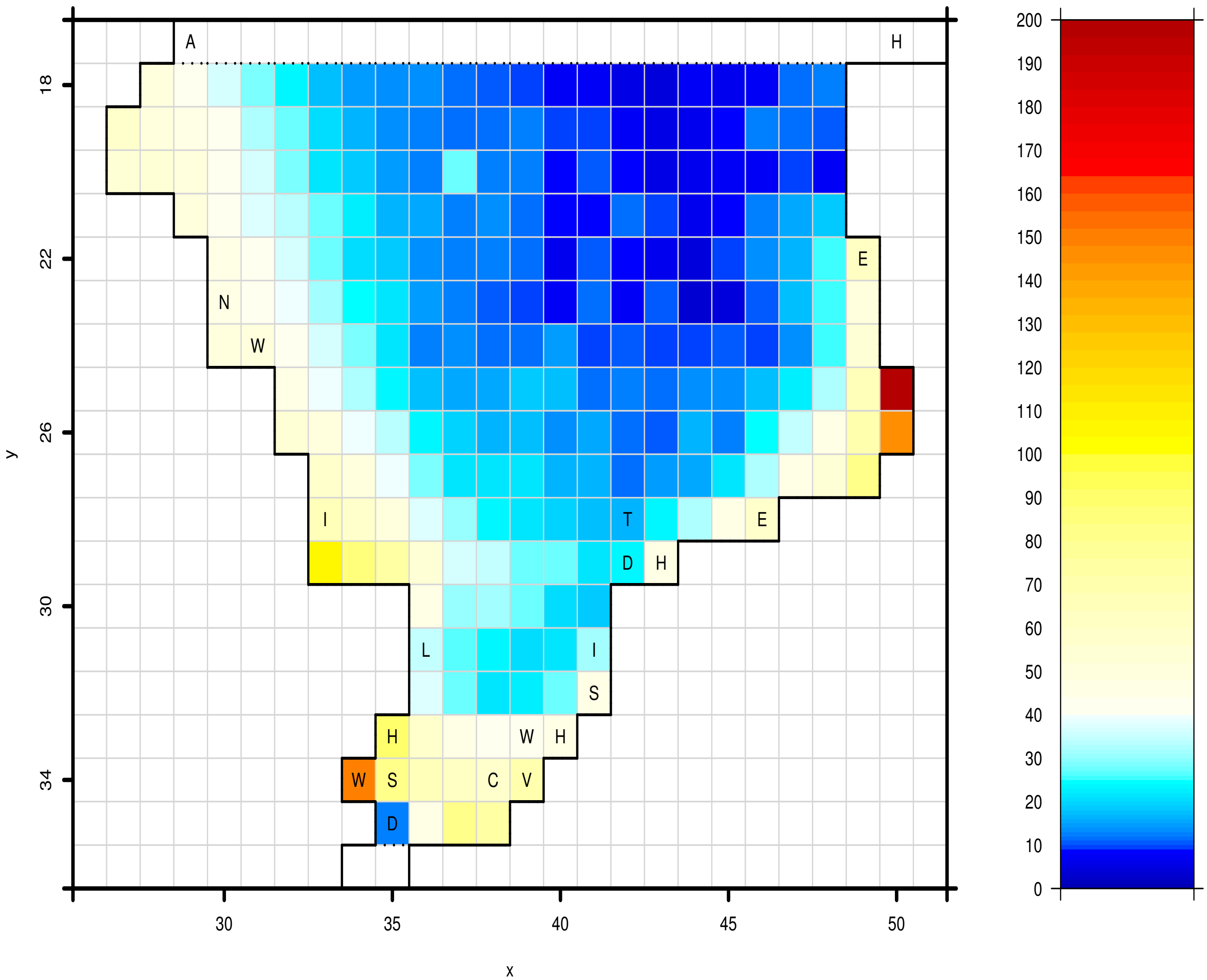
Estimates (in cm) from a time-constant r-largest model with $r = 20$



Spatial distribution of **tide-surge interaction** effects

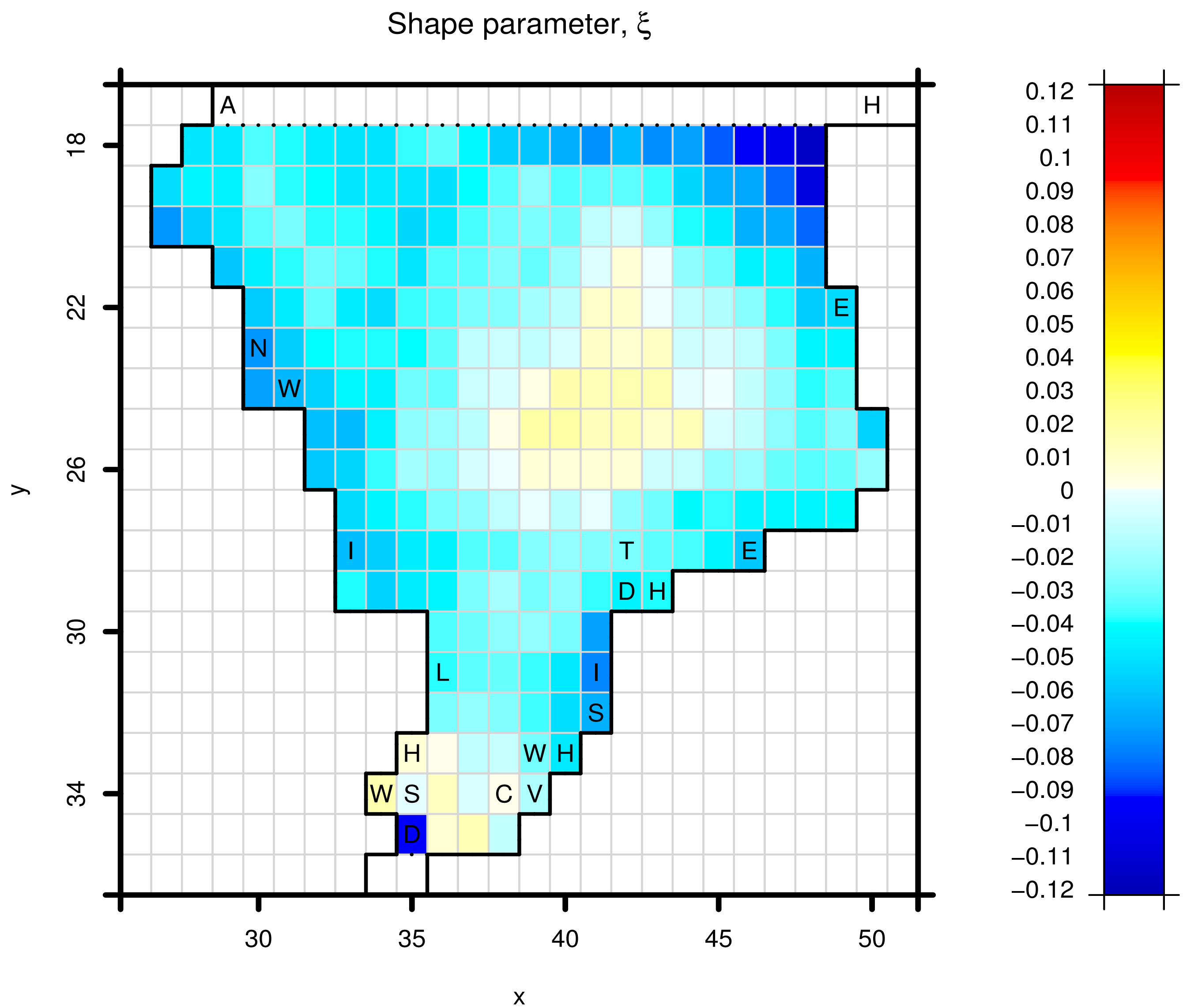
Estimates from a time-constant r-largest model with $r = 20$

50 year return level for surges - 50 year return level for high tide surges (cm)



Spatial distribution of **shape parameter** estimates

Estimates from a time-constant r-largest model with $r = 20$



Modelling temporal trends

Model temporal trends in extreme surges via:

- Simple parametric trend models (Coles, 2001)
- Changepoint models
- Nonparametric local likelihood methods (Hall and Tajvidi, 2000; Davison and Ramesh, 2000)

Nonparametric methods make weaker assumptions, but only parametric models allow extrapolation forward in time

Local likelihood: estimation

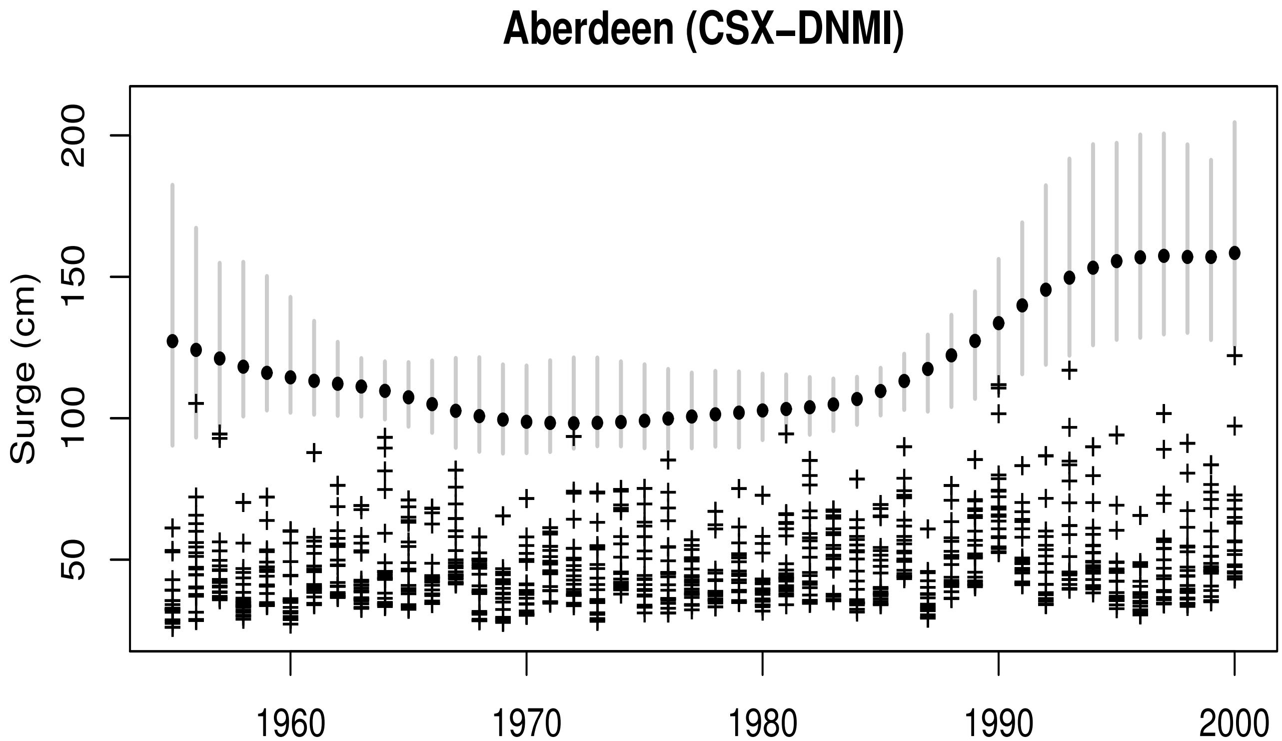
- Purely temporal smoothing
- Let $f(\mathbf{x}_{ij}; \theta_{ij})$ denote the density of $\mathbf{X}_{ij}$
- "Locally constant in time" estimator for θ_{ij}: maximise
$$\sum_{J=1}^{n} K(t_J - t_j; h_t) \log \left\{ f \left(\mathbf{x}_{iJ}; \theta_{ij} \right) \right\},$$
- K is a kernel function with bandwidth h_t
- "Local regression in time" estimator for θ_{ij}

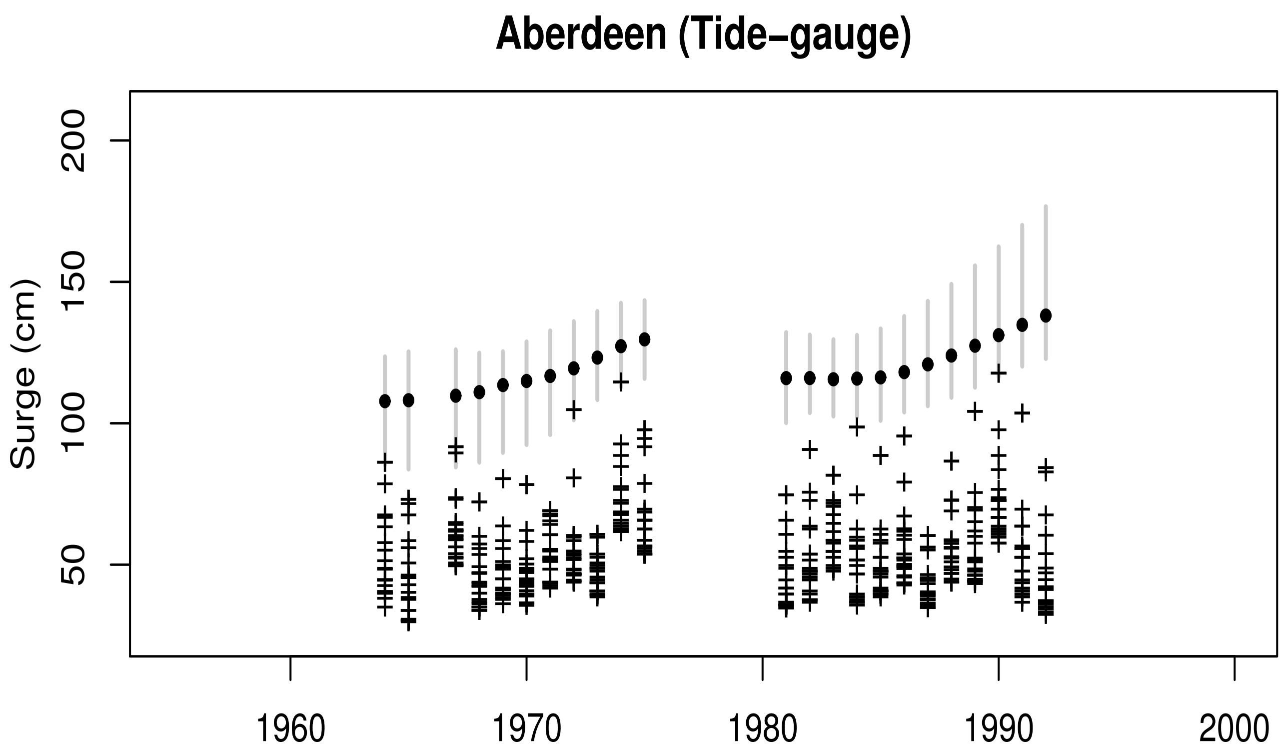
Local likelihood: statistical issues

- Selection of the kernel bandwidth h_t
- Variability bands: use the semiparametric bootstrap (Davison and Ramesh, 2000)
- Model diagnostics: quantile-quantile plots, for each $k = 1, \ldots, r$

20-largest surges per year, & estimated **50y return levels**

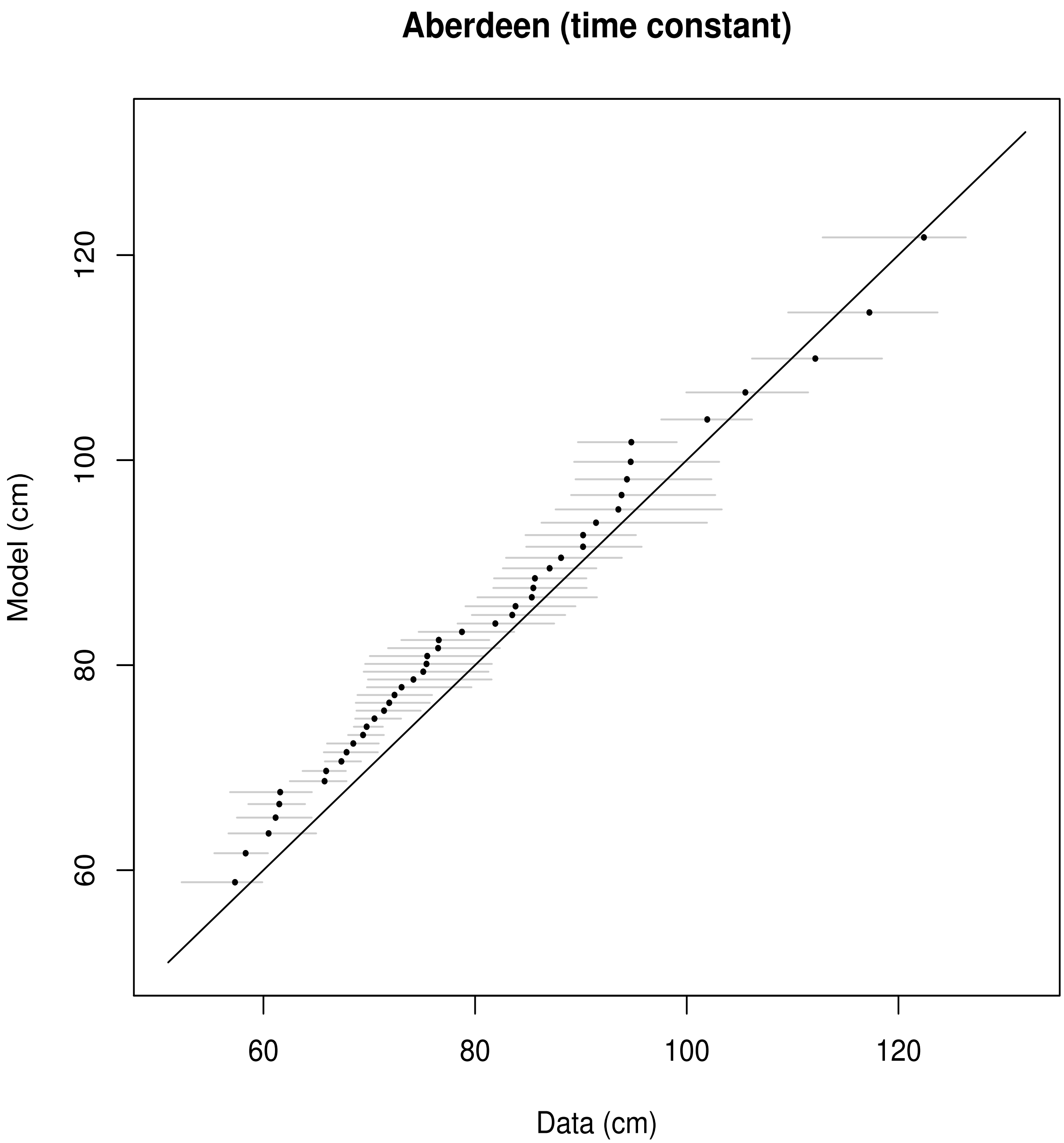
Estimates from a local linear r-largest model with $h_t = 3.5$y and $r = 20$
95% pointwise variability bands shown (grey)





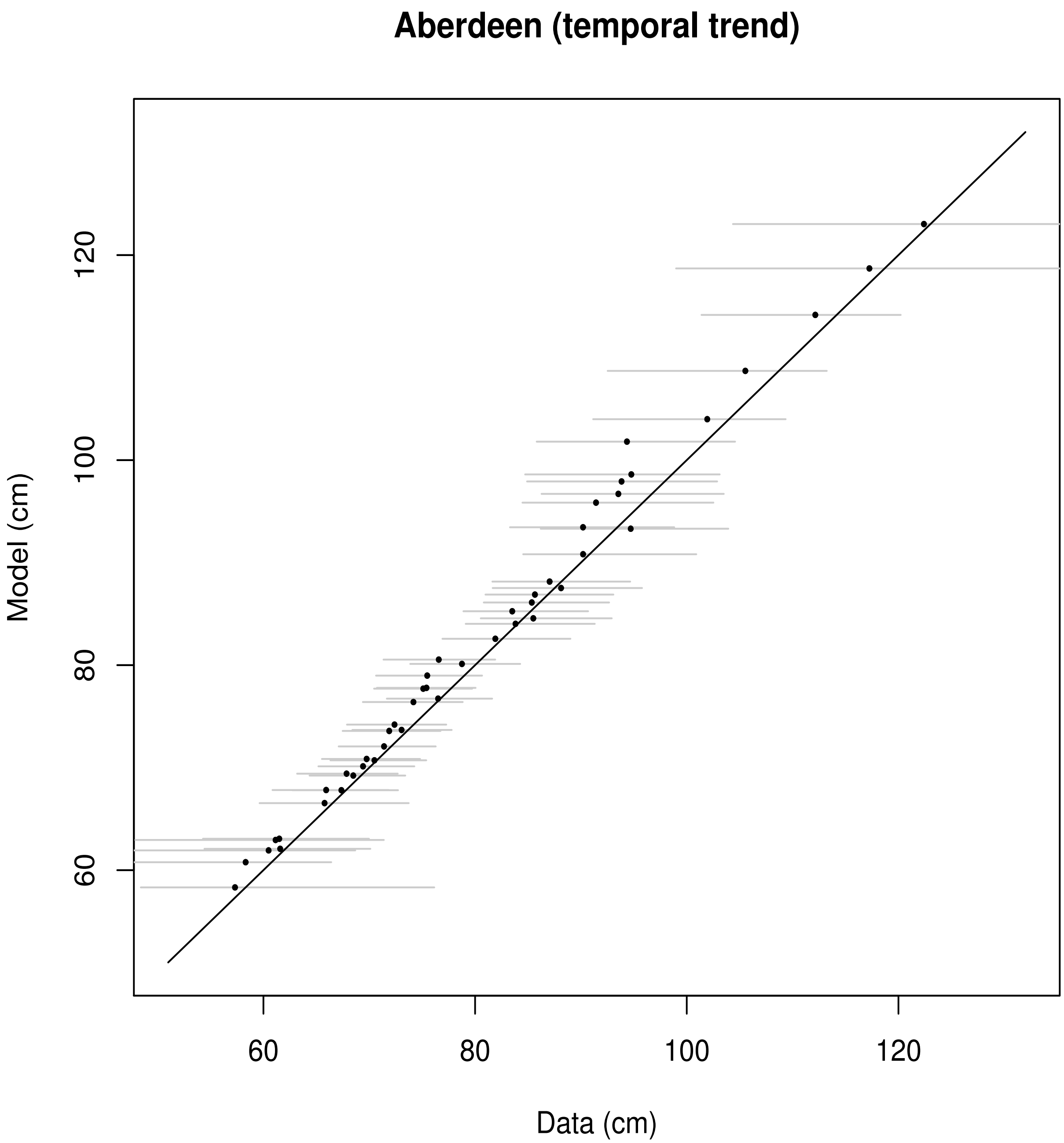
Q-Q plot for r-largest model fitted to CSX-DNMI surges

$k = 1$; $r = 20$; 95% pointwise variability bands shown (grey)



Q-Q plot for r-largest model fitted to CSX-DNMI surges

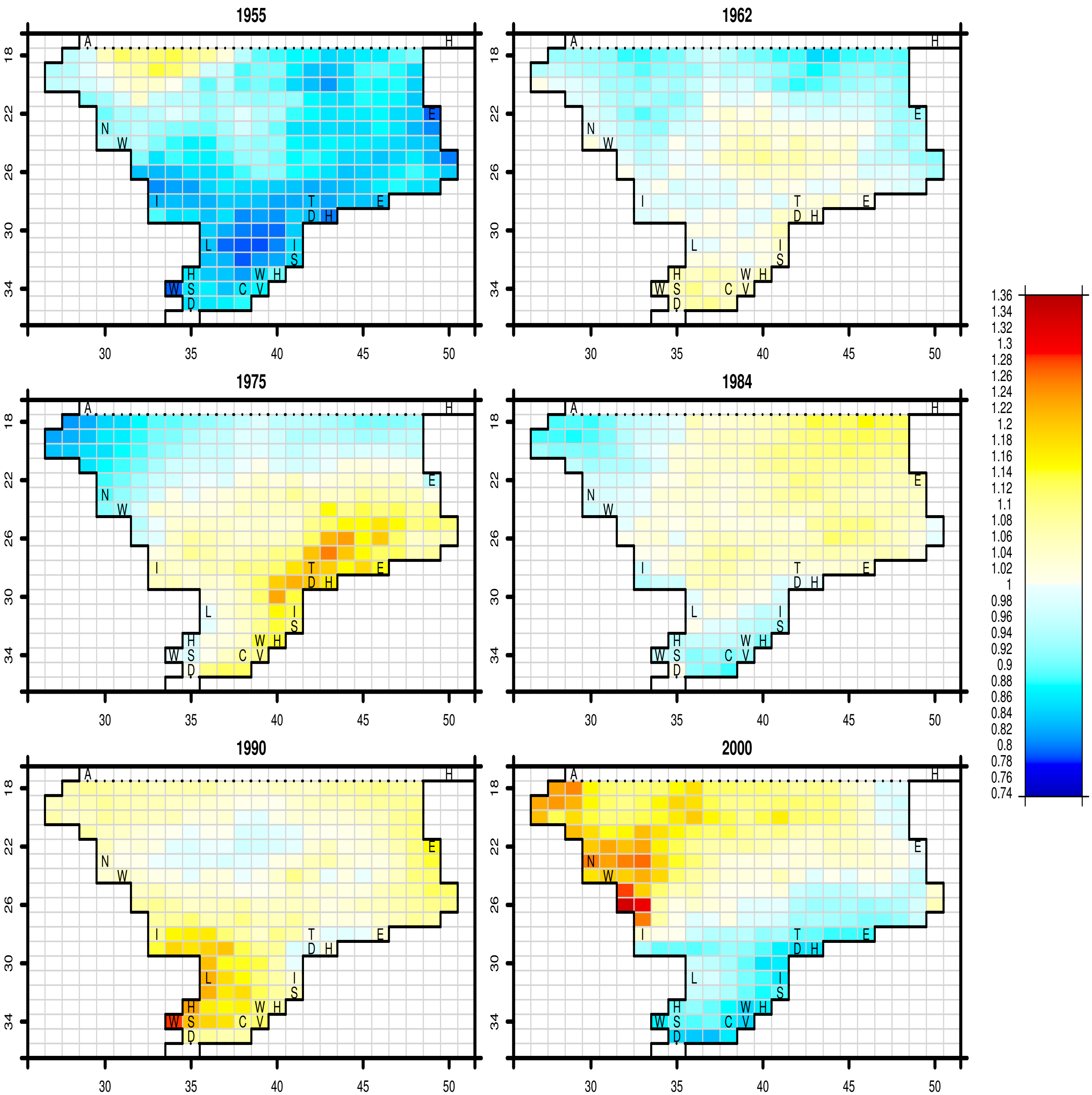
$h_t = 3.5$; $k = 1$; $r = 20$; 95% pointwise variability bands shown (grey)



Scaled spatio-temporal variations in **50y return levels** in the North Sea, after temporal smoothing

Estimates from a local r-largest model at each site, with $h_t = 3.5$ & $r = 20$

Estimates are scaled to be one for the time-constant r-largest model



North Atlantic Oscillation

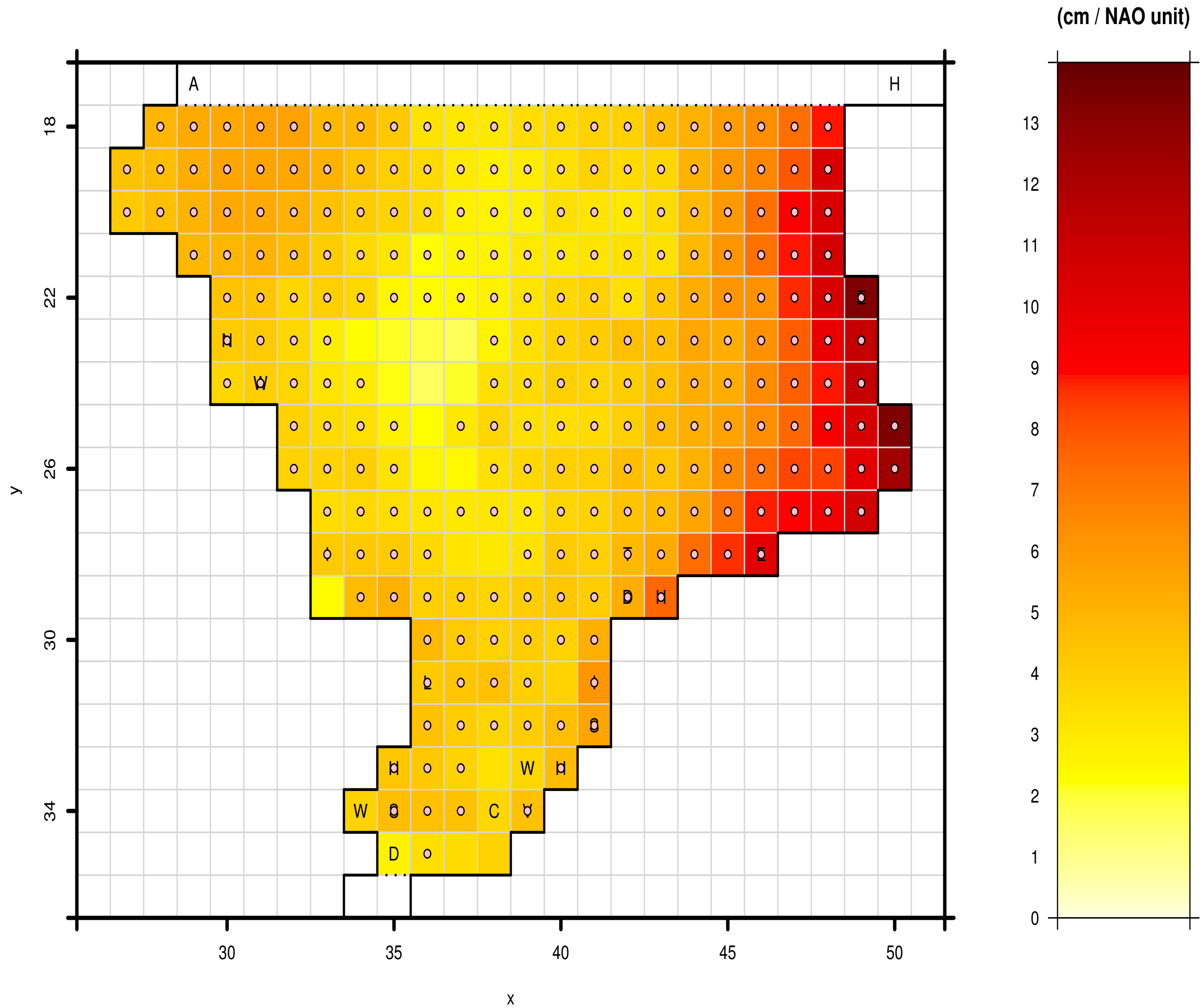
- The NAO: a mode of large-scale climate variability within the North Atlantic
- Defined as the normalised pressure difference between Reykjavik and Gibraltar
- Can use as a covariate to explain temporal trends in storm surge elevations

Incorporating the **annual NAO index** as a covariate

Estimates of β_i from an r-largest model at each site,

with $\mu_{ij} = \mu_i + \beta_i \text{NAO}_j$, $\sigma_{ij} = \sigma_i$, $\xi_{ij} = \xi_i$

Dots denote sites $\mathbf{s}_i$ for which β_i is significant at the 95% level



Incorporating spatial information

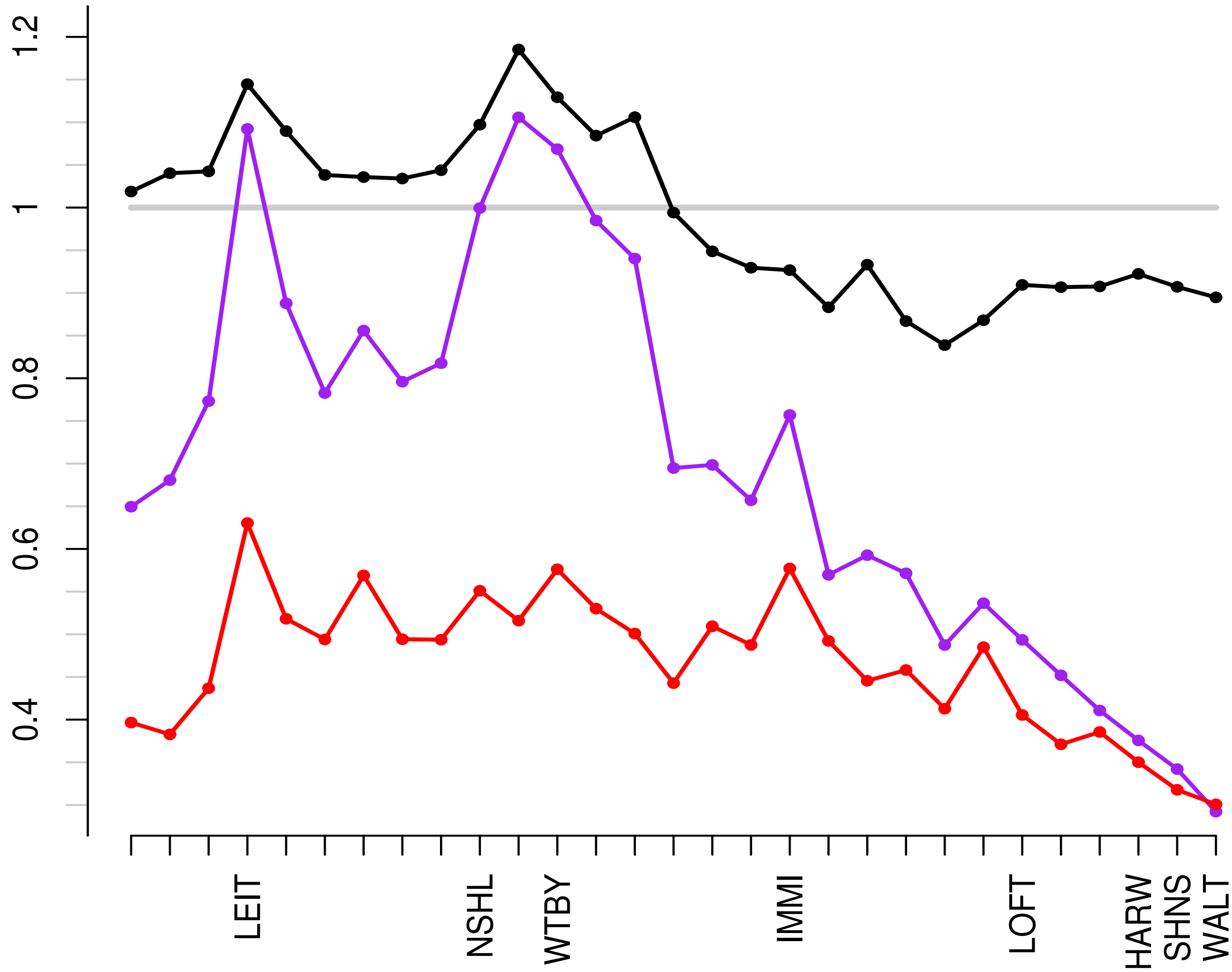
- Purely spatial smoothing: $\theta_{ij} = \theta_i$.
- Model the marginal parameter surface $\theta_i := (\mu_i, \sigma_i, \xi_i)$ as a smooth function of $\mathbf{s}_i$
  - Spatial linkage: e.g. $\xi_1 = \ldots = \xi_d$
  - Spatial smoothing: e.g. ξ_i is locally polynomial

- Aim: to improve efficiency of the marginal parameter estimates
- Uncertainty assessment and model diagnostics need to account for residual spatial dependence

Efficiency gains from linking ξ over **space**

Variance from an r-largest model with ξ_i constrained equal at all sites on the British east coast, relative to a model where ξ_i is unconstrained

μ_i (black), σ_i (purple), ξ_i (red)

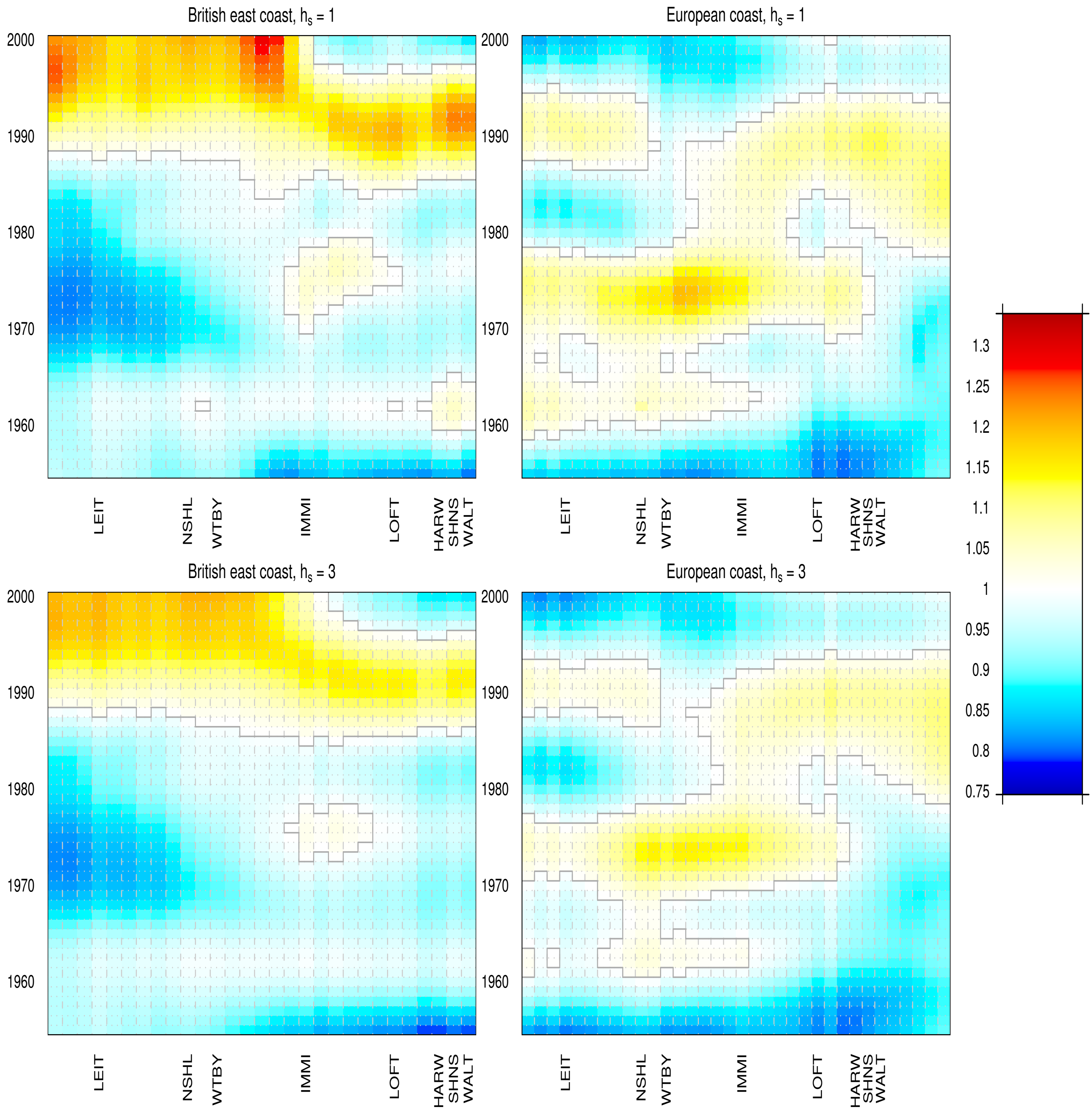


Spatio-temporal modelling

- Model θ_{ij} as a smooth function of time t_j *and* location $\mathbf{s}_i$
- Use a novel extension of the local likelihood methodology
- Multivariate normal kernel with a temporal bandwidth h_t and a spatial bandwidth $\mathbf{h}_s$
- Spatial smoothing reduces the uncertainty associated with estimates of temporal trend

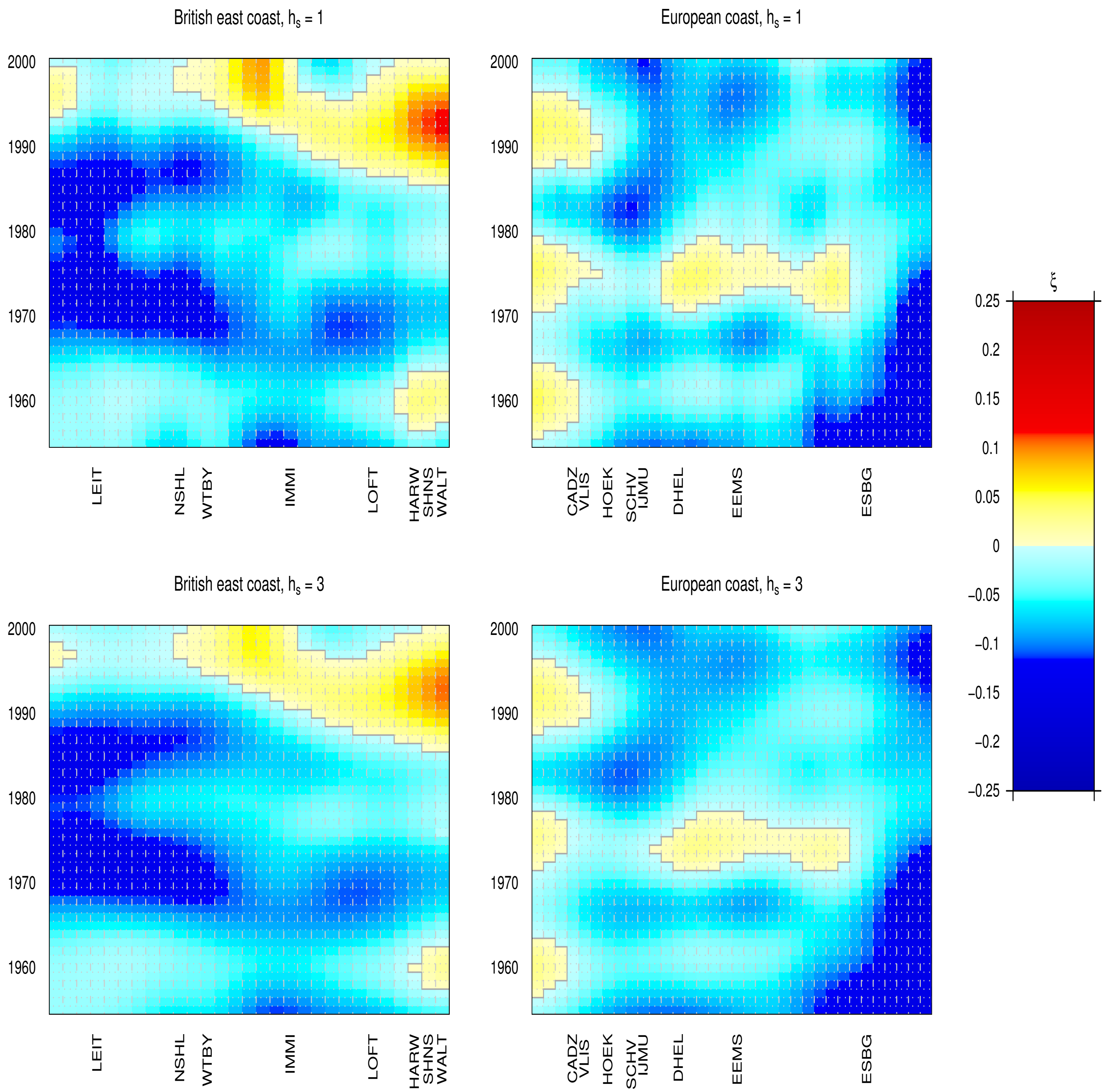
Spatio-temporal variations in **50y return levels** along North Sea coasts, after spatio-temporal smoothing

Estimates from a local spatio-temporal r-largest model,
with $h_t = 3.5$, $r = 20$ and various spatial bandwidths h_s
Estimates are scaled to be one for the time-constant r-largest model



Spatio-temporal variations in **shape parameter estimates** along North Sea coasts, after spatio-temporal smoothing

Estimates from a local spatio-temporal r-largest model, with $h_t = 3.5$, $r = 20$ and various spatial bandwidths h_s



Further work

- Use multivariate extreme value methods to model spatial dependence at extreme levels
- Compare extremal characteristics of surge data generated under different climate scenarios
- Explore the impact of model resolution

Bibliography

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

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

Flather, R. A., Smith, J. A., Richards, J. C., Bell, C. and Blackman, D. L. (1998) Direct estimates of extreme storm surge elevations from a 40-year numerical model simulation and from observations. *Global Atmos. Ocean Syst.*, 165–176.

Hall, P. and Tajvidi, N. (2000) Distribution and dependence function estimation for bivariate extreme value distributions. *Bernoulli*, **6**, 835–844.

Smith, R. L. (1986) Extreme value theory based on the r largest annual events. *J. Hydrol.*, **86**, 27–43.

Tawn, J. A. (1988) An extreme value theory model for dependent observations. *J. Hydrol.*, **101**, 227–250.