Extreme value methods for synthetic storm surge data

_at_ Southampton Oceanography Centre, 30 September 2004

Adam Butler (Lancaster University)
a.butler@lancaster.ac.uk

Extremes at Lancaster

- Project: *Modelling extreme events for multivariate and temporal processes*
- Dependence at extreme levels:
  Ledford & Tawn (1997),
  Coles, Heffernan & Tawn (1999),
  Heffernan & Tawn (2004)
- Clustering: simulation and estimation
- Extremal behaviour of time series models
- Models and diagnostics for spatial extremes

- Applications: regional flood estimation, pollution chemistry, maritime safety, finance

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

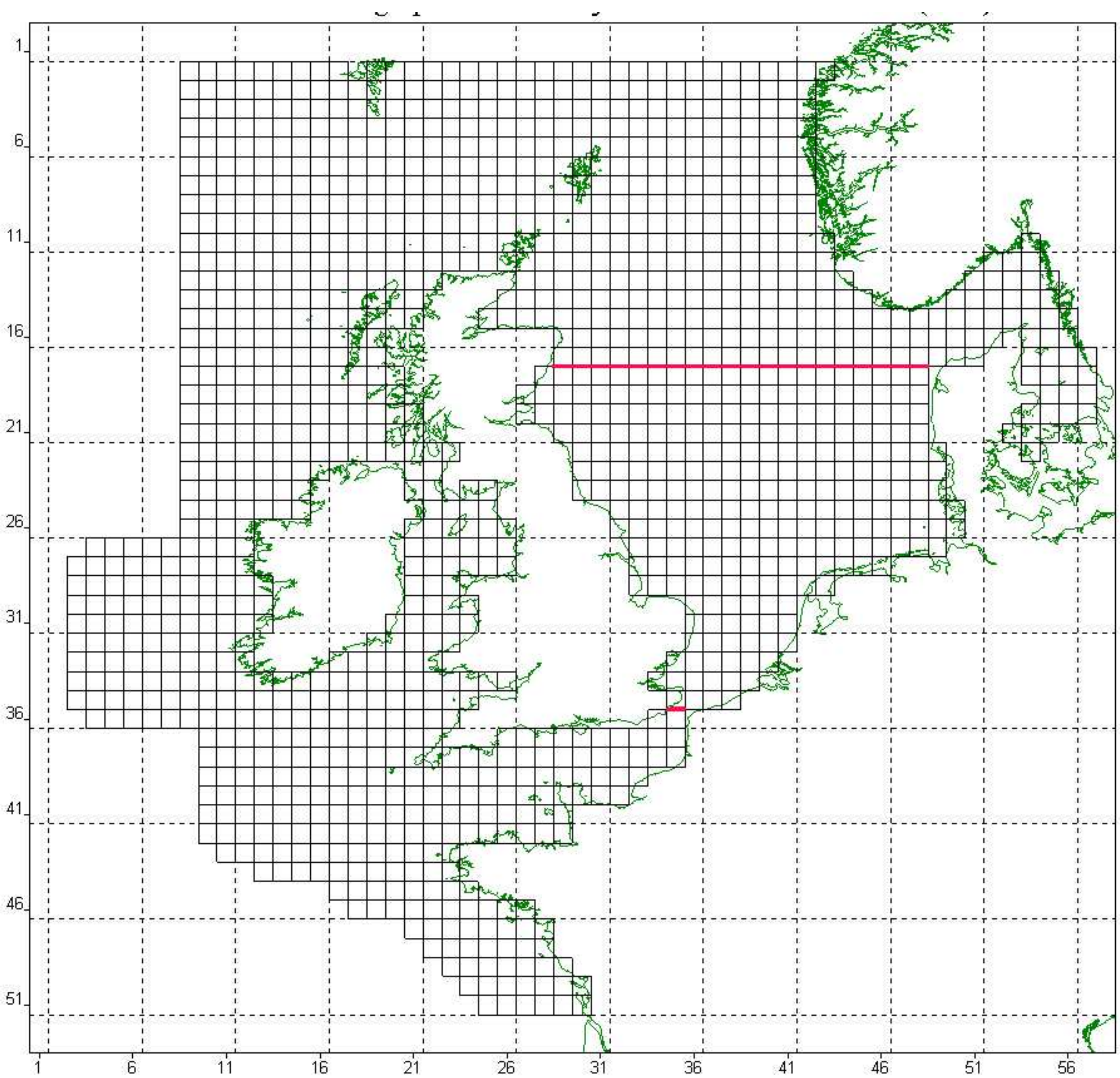
- Applied part of my PhD research
- Working with Janet Heffernan, Jonathan Tawn and Roger Flather (POL)

Synthetic storm surge data

- Numerical hydrodynamical models are used to study surge characteristics (Bode and Hardy, 1997)
- 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 1955-2000 (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"
- Univariate extremes are either...
  - extreme order statistics, or
  - exceedances of a high threshold
- EV models are asymptotically-motivated, and so provide a robust basis for extrapolation
- EV models are parametric

The r-largest model for extremes

- $x_{ij}^{(k)}$:= 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 (Weissman, 1978; Smith, 1986; Tawn, 1988)

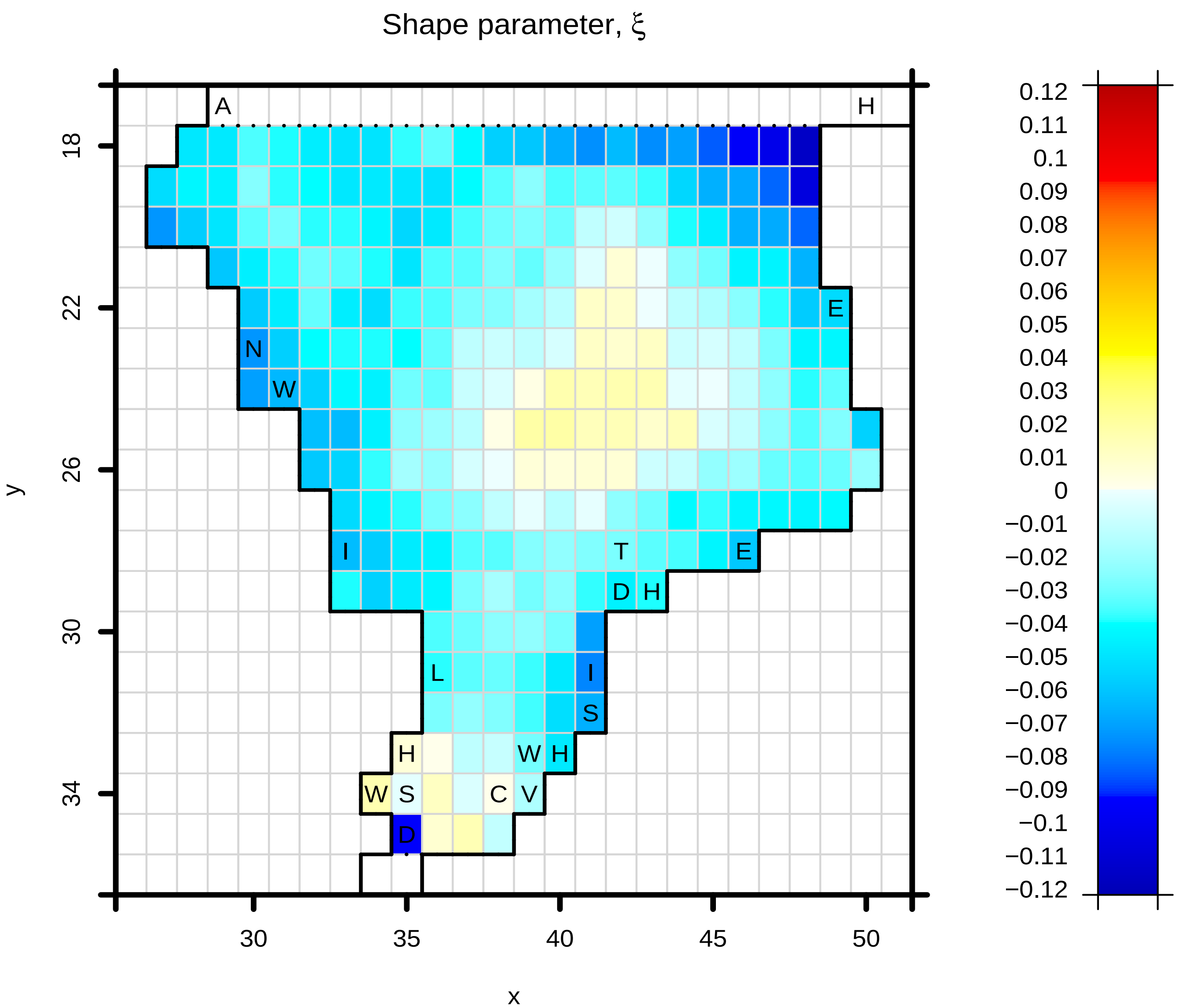
- r-largest model has parameters $\theta_{ij} := (\mu_{ij}, \sigma_{ij}, \xi_{ij})$; (location, scale and shape parameters)
- A design parameter is the level x_{ij} such that

$$\mathbb{P}\left[X_{ij}^{(1)} > x_{ij}\right] = 1/N;$$

  the "N-year return level"

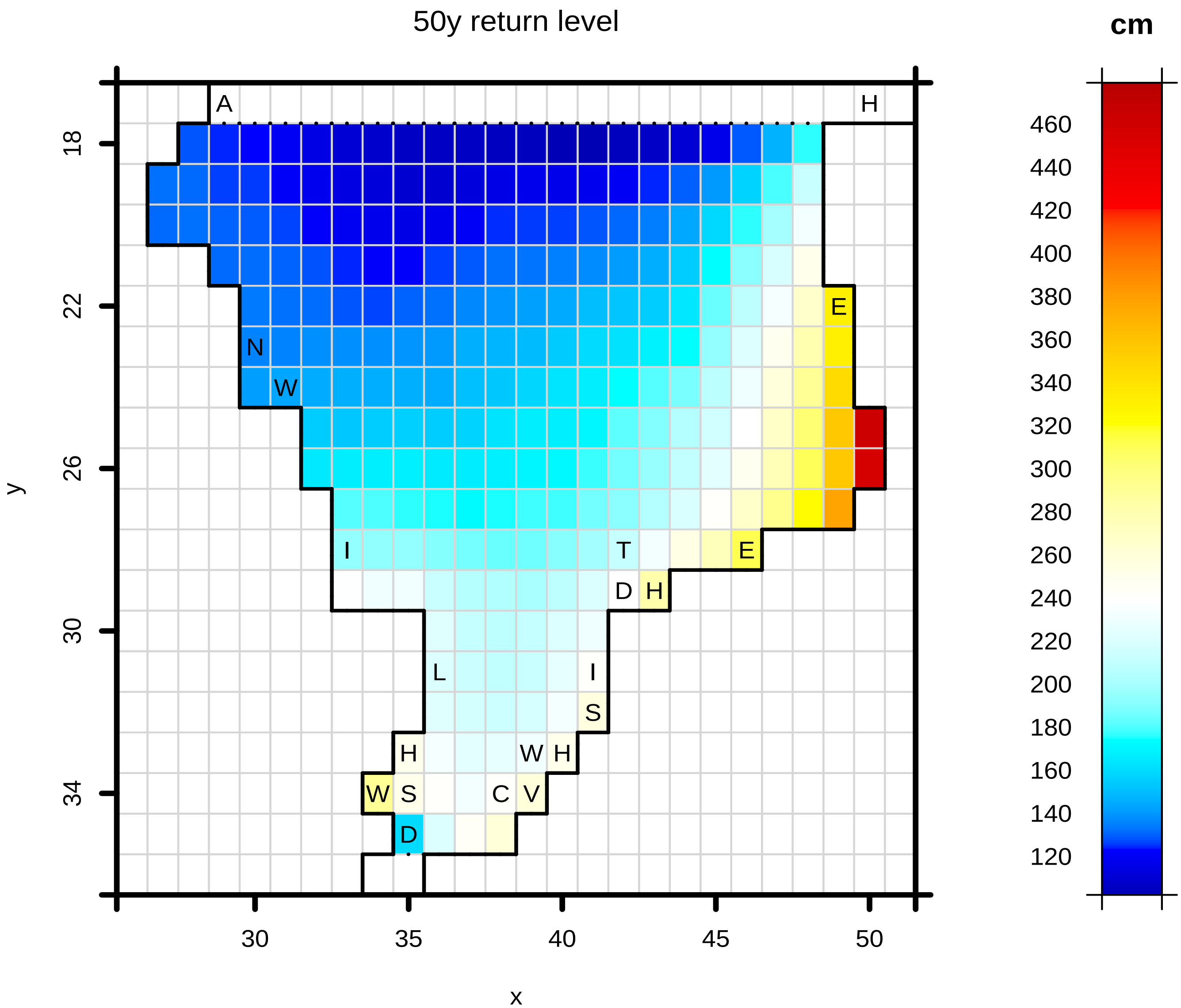
Spatial distribution of extreme value parameter estimates

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



Spatial distribution of extreme value parameter estimates

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



Temporal trends in extremes

- Site-by-site analysis: Flather et al. (1998) assume that $\theta_i := \theta_{i1} = \ldots = \theta_{in}$

- Model temporal trends in extreme surge elevations (Butler et al., 2004)
- Parametric trend models (Coles, 2001)
- Local regression via local likelihood: Hall and Tajvidi (2000), Davison and Ramesh (2000)
- Alternative penalized likelihood (smoothing spline) approach - developed in the EV context by: Chavez-Demoulin (1999), Pauli and Coles (2001)

Local likelihood: estimation

- Let $f(\mathbf{x}_{ij};\theta_{ij})$ denote the density of $\mathbf{X}_{ij}$
- Local constant estimator for θ_{ij}: maximise

$$\sum_{J=1}^{n} K(t_J - t_j; h) \log\left\{f\left(\mathbf{x}_{iJ};\theta_{ij}\right)\right\},$$

- K is a kernel function with bandwidth h
- Local regression estimator for θ_j: model θ_j as a function of time using a local regression model.

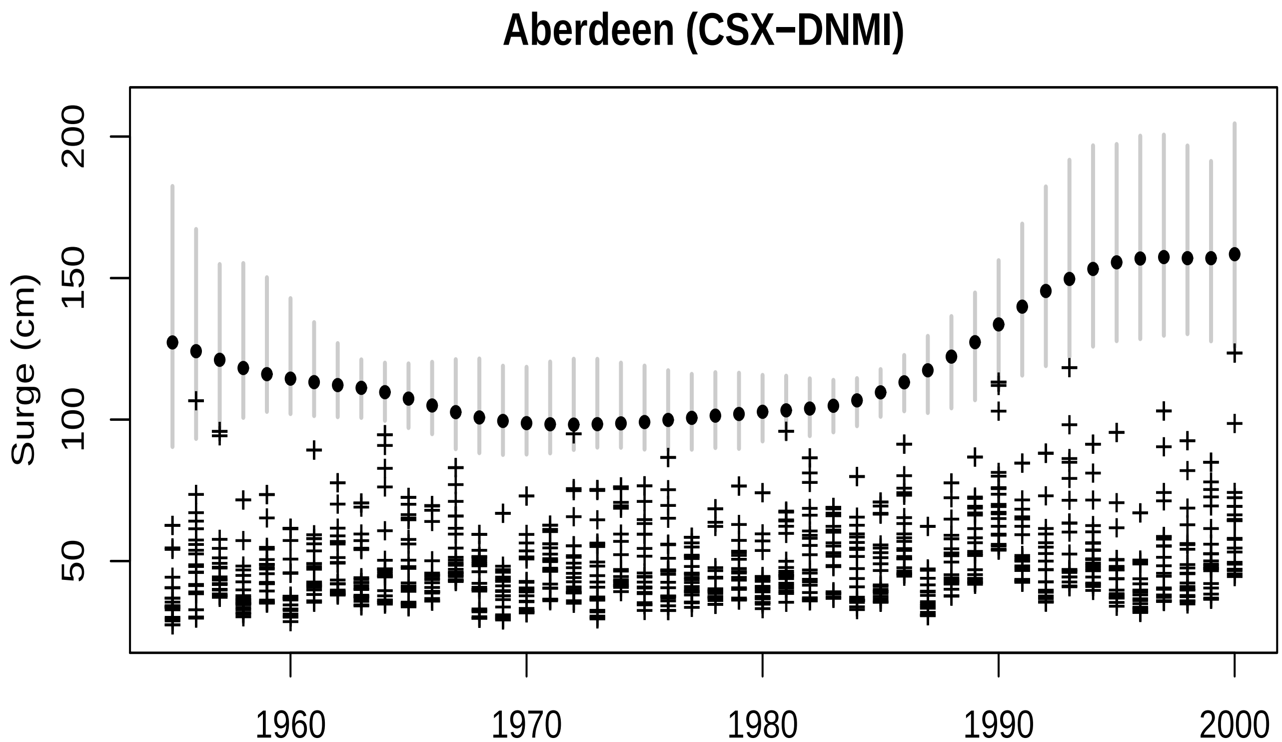
Local likelihood: statistical issues

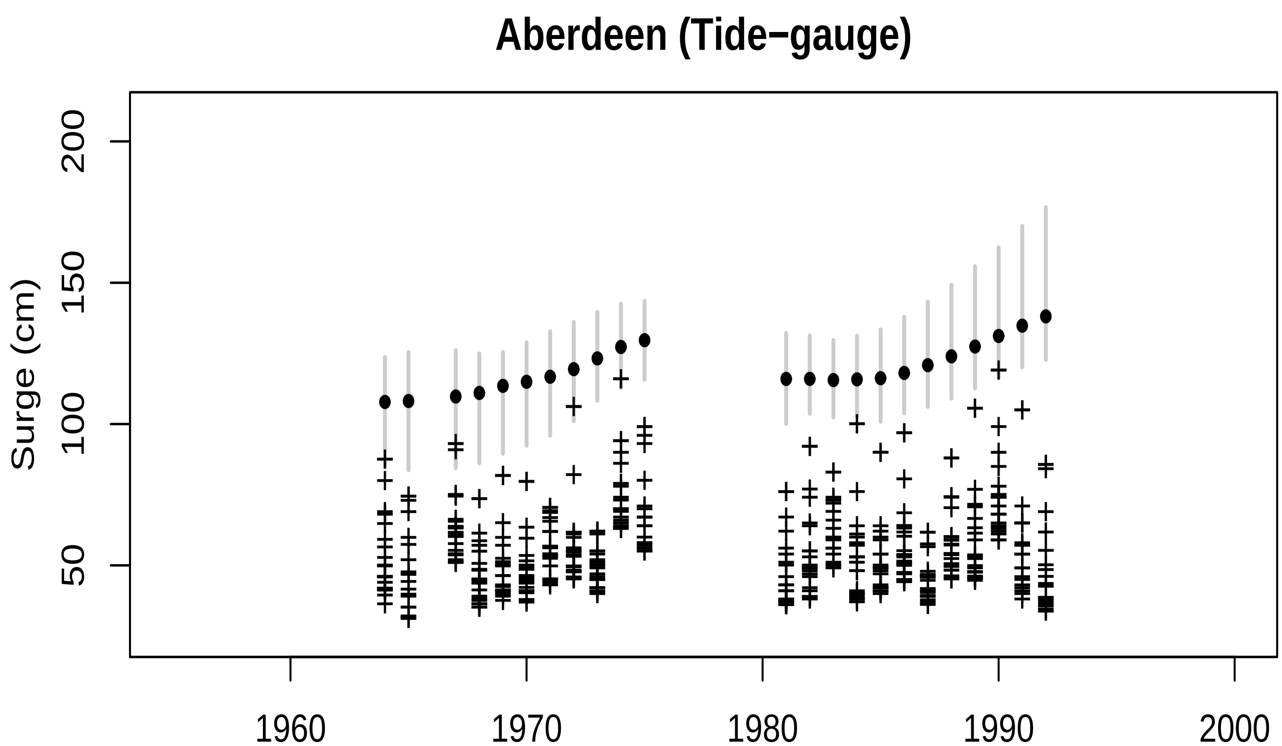
- Selection of the kernel bandwidth h
- Variability bands: use the semiparametric bootstrap (Davison and Ramesh, 2000)
- Model diagnostics: quantile-quantile plots, for each $k = 1, \ldots, r$

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

Estimates from a local linear r-largest model with $h = 3.5$y and $r = 20$

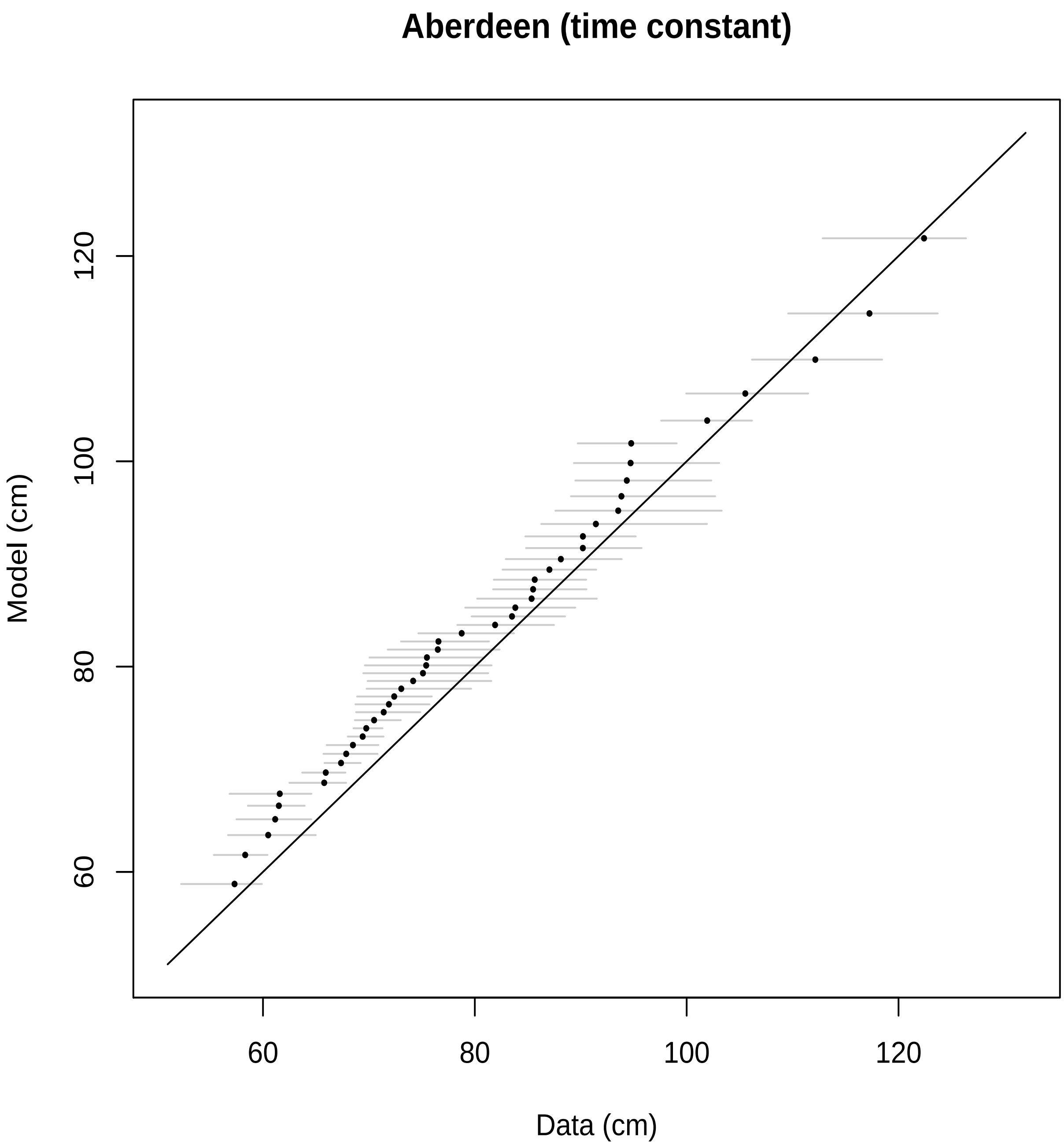
95% pointwise variability bands shown (grey)





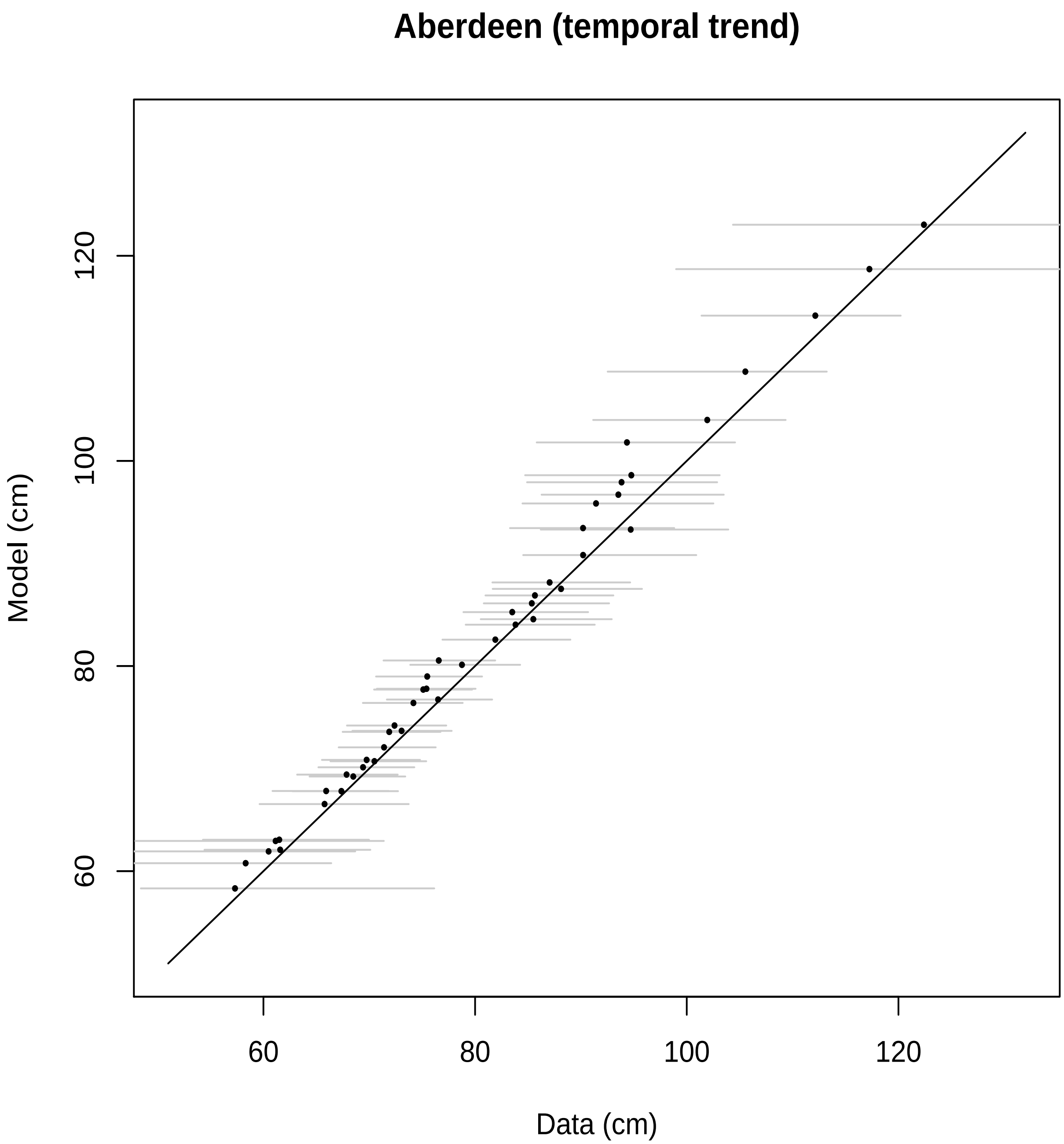
Q-Q plot for r-largest model fitted to CSX-DNMI surges

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



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

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



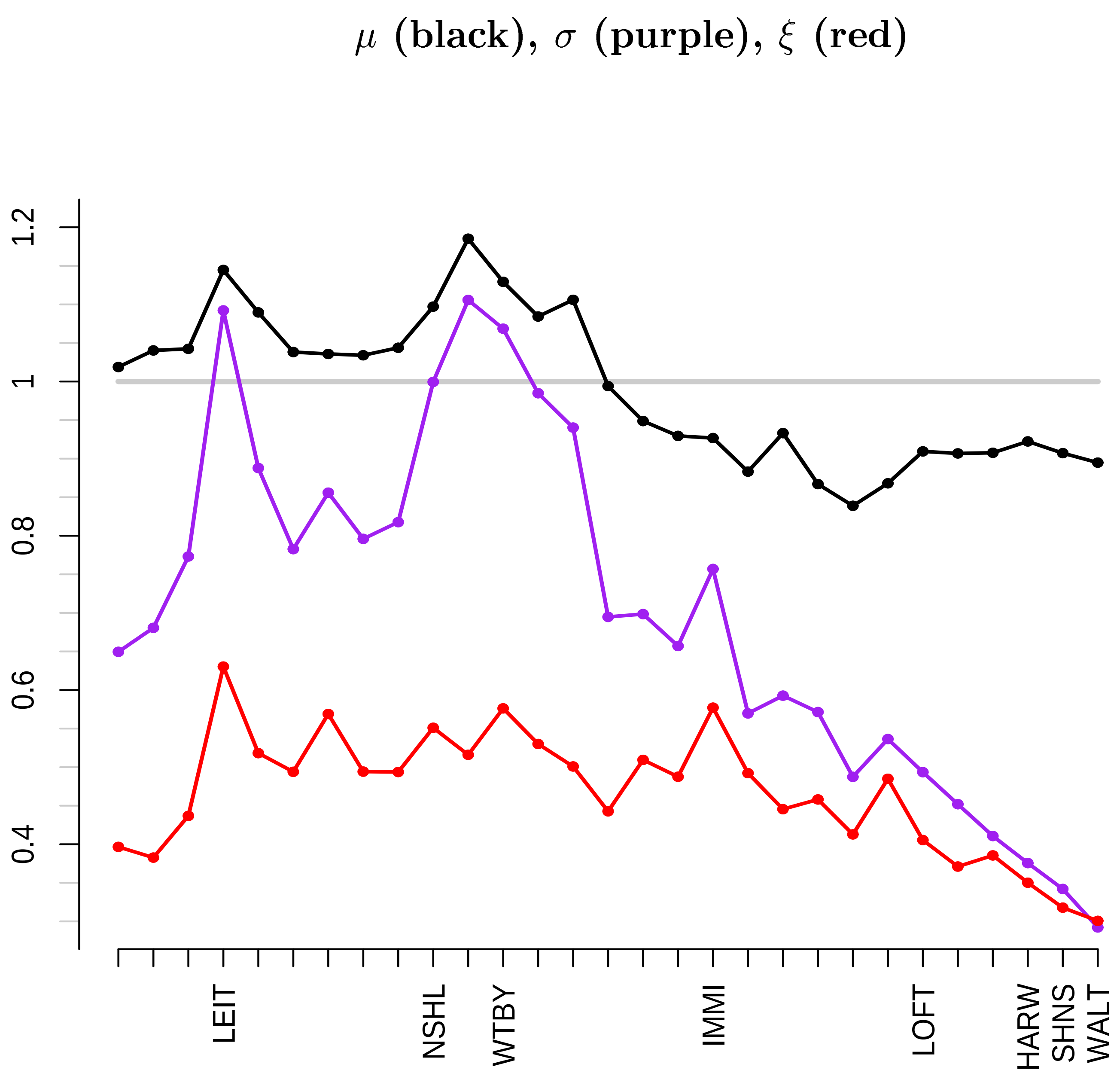
Incorporating spatial information (1)

- 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 constrained equal at all sites on the British east coast, relative to a model where ξ is unconstrained



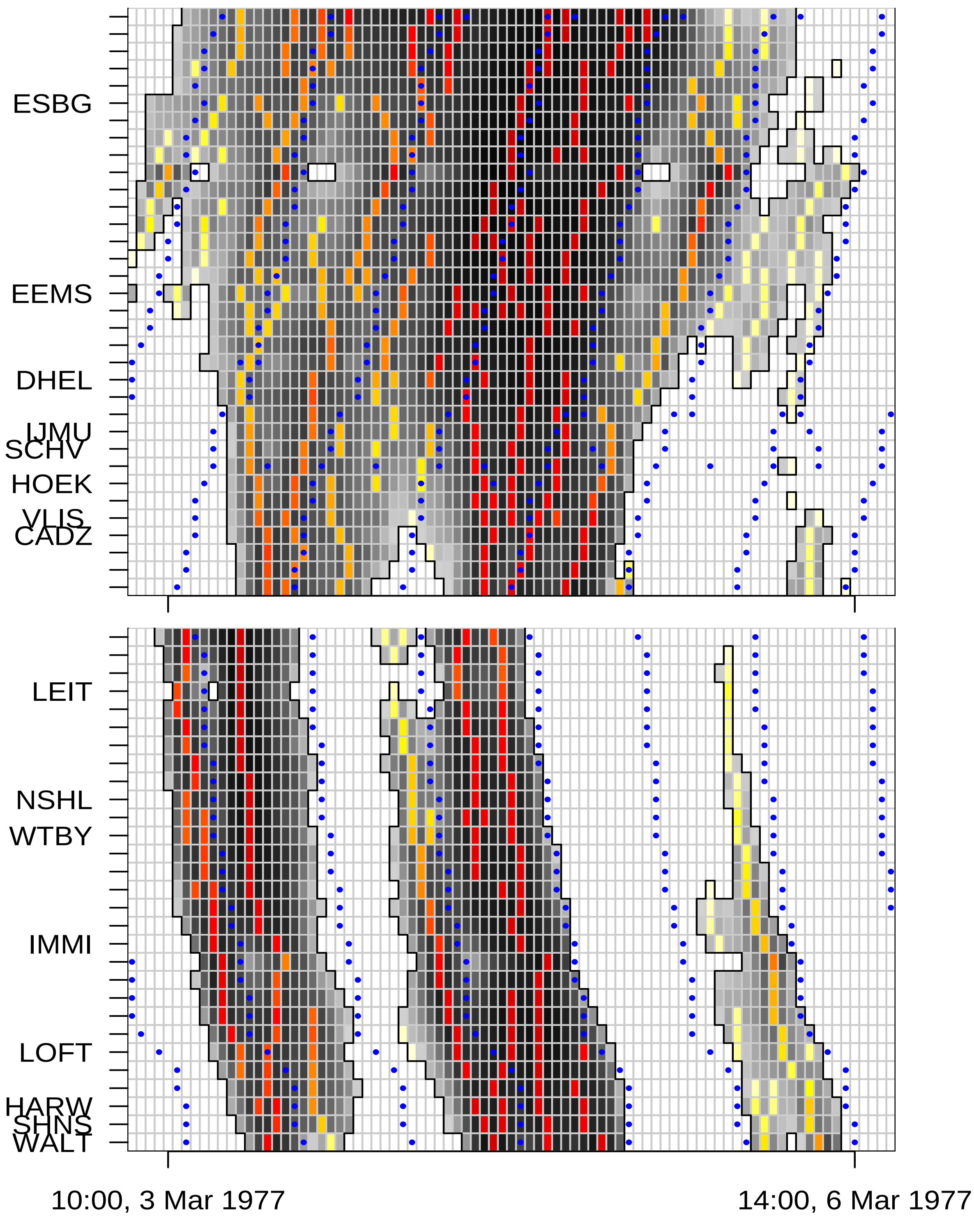
Incorporating spatial information (2)

- Use multivariate extreme value methods to model dependence at extreme levels between $X_1, \ldots, X_d$

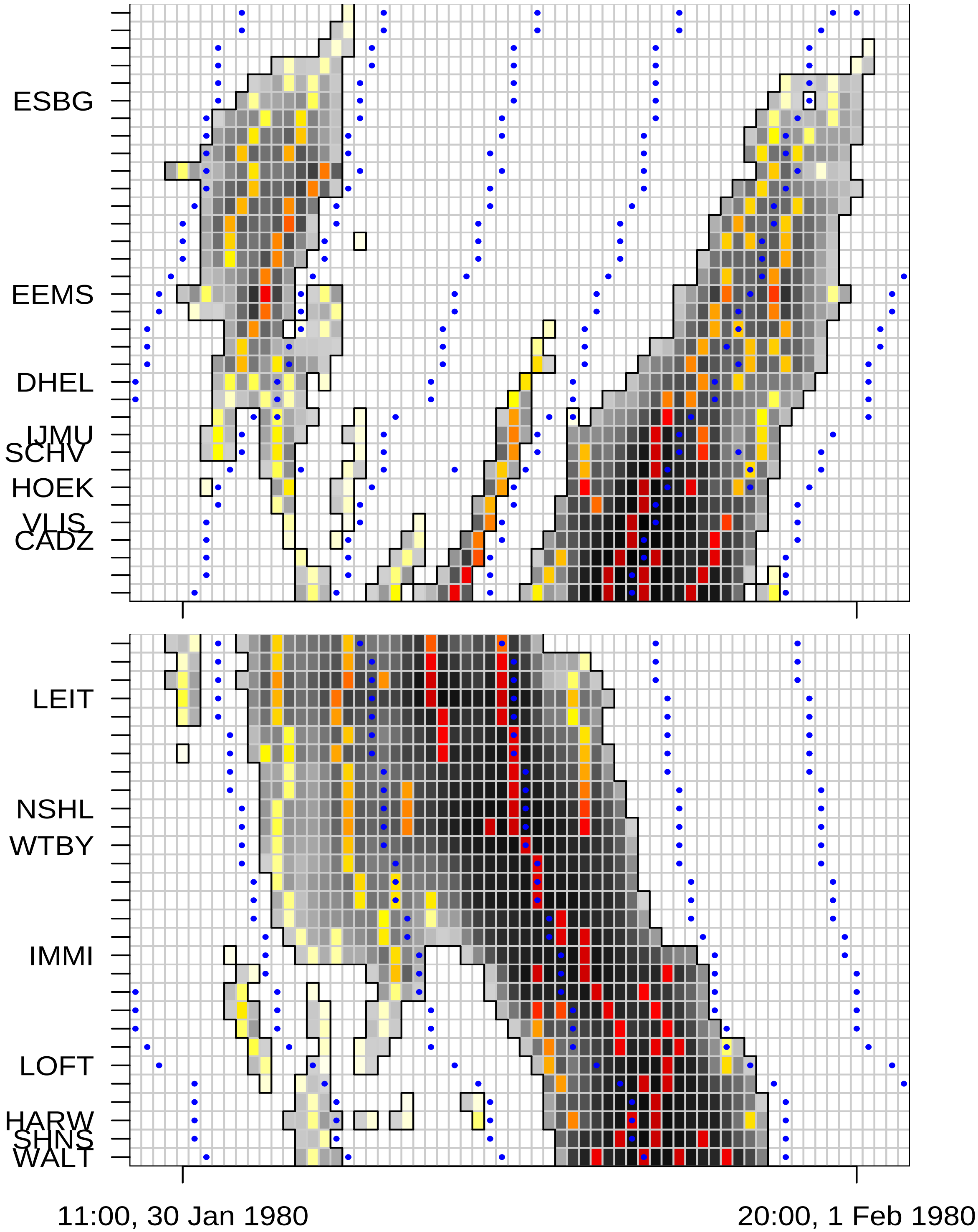
- Classical methods (Resnick, 1987) are only suitable if data are simultaneously extreme in $X_1, \ldots, X_d$ (asymptotic dependence)
- Recent methods (Ledford and Tawn, 1997) allow for the possibility of asymptotic independence
- Heffernan and Tawn (2004) consider the behaviour of $\mathbf{X}_{-i}|(X_i = x_i)$ as $x_i \to \infty$

- Definition of a multivariate extreme is complicated by temporal dependence...

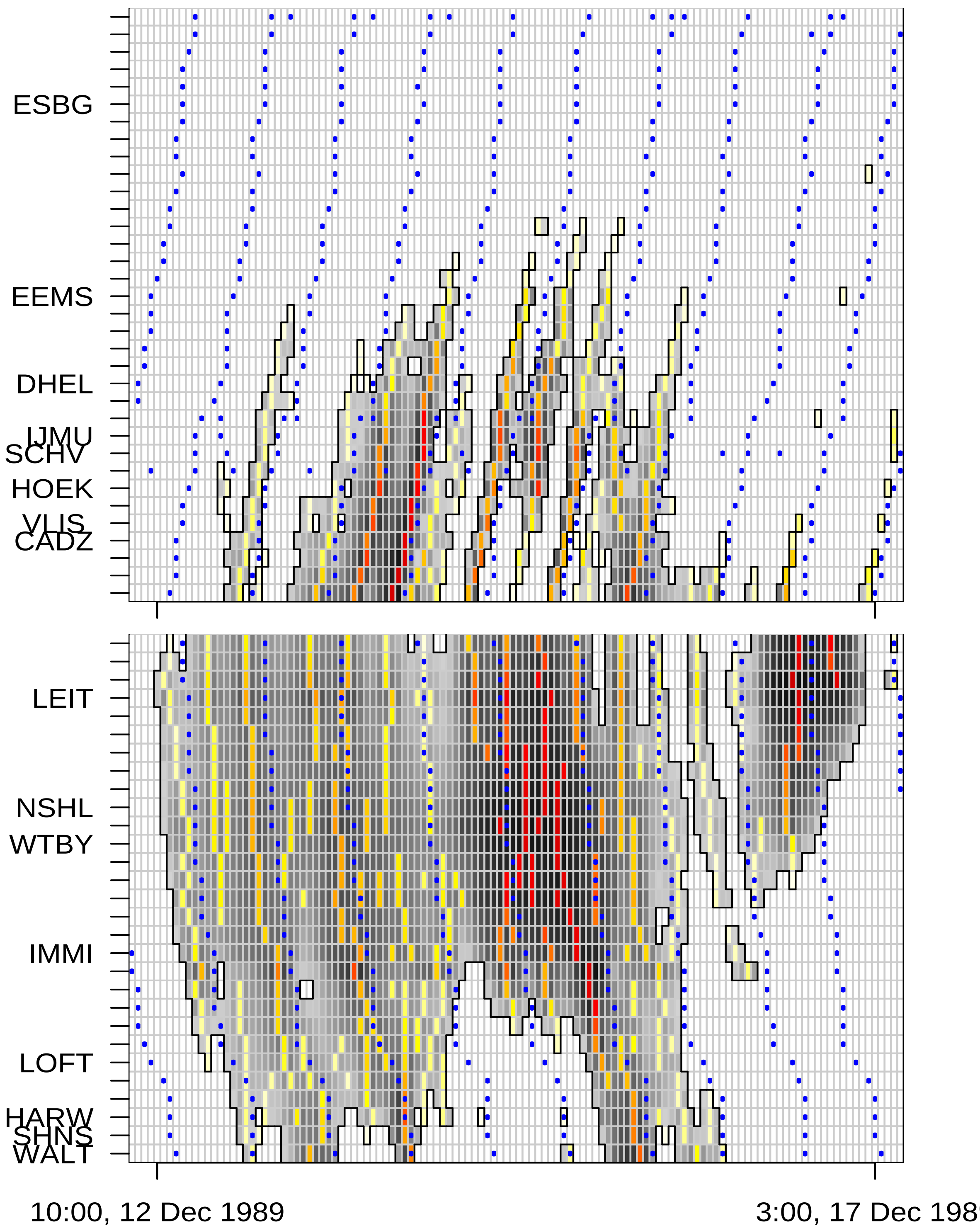
Spatio-temporal profile for a North Sea storm surge



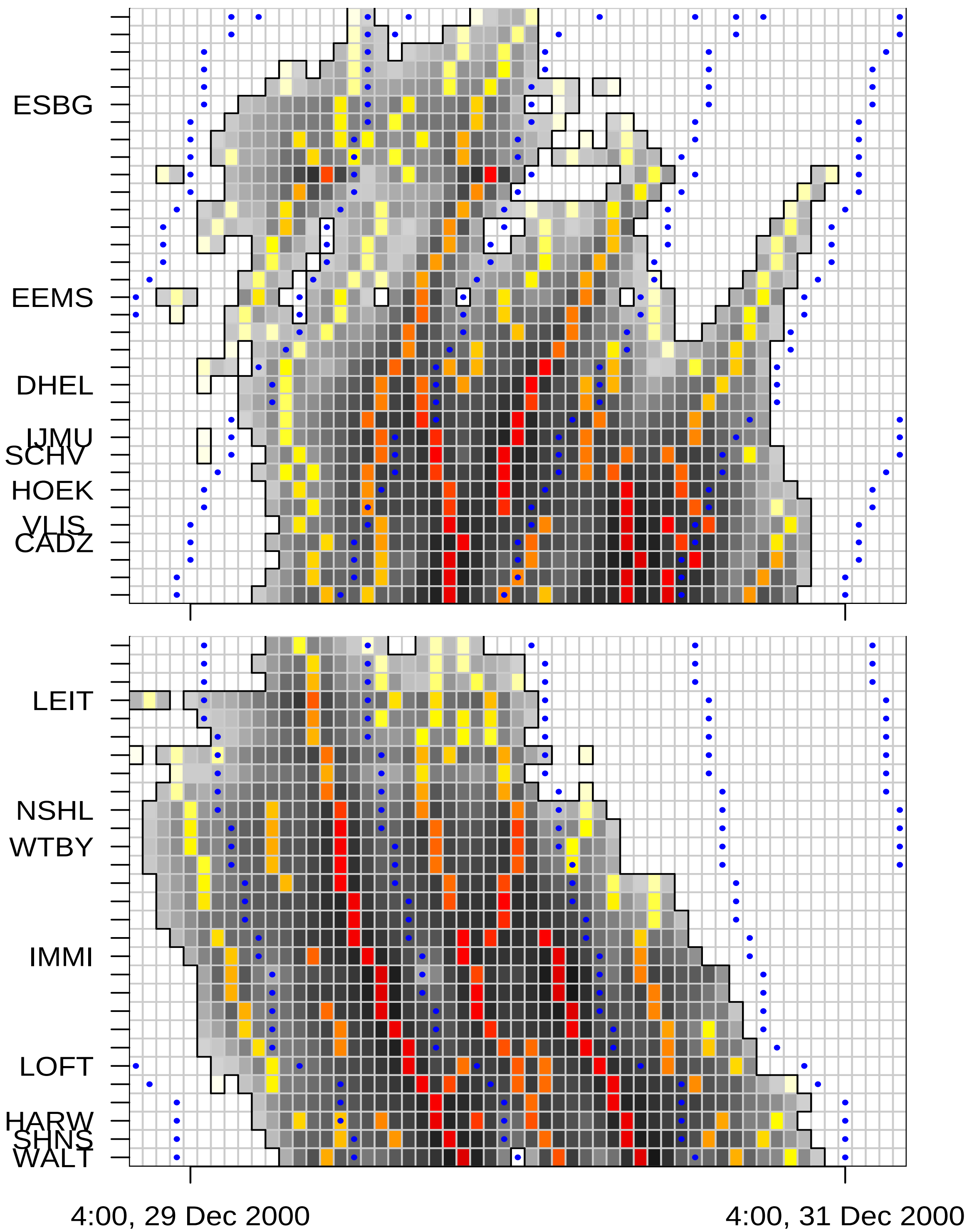
Spatio-temporal profile for a North Sea storm surge



Spatio-temporal profile for a North Sea storm surge



Spatio-temporal profile for a North Sea storm surge



Conditional extremes of a Markov chain

- Assume that $\mathbf{X}$ has standard Gumbel margins
- Heffernan and Tawn (2004) assume that there exist vector-valued functions $\alpha_{|i}$ and $\beta_{|i}$ such that

$$F(\alpha_{|i}(x_i) + \beta_{|i}(x_i)\mathbf{z}_{|i} | x_i) \to G(\mathbf{z}_{|i})$$

  as $x_i \to \infty$, where G has nondegenerate margins

- We have theoretical results which characterise the structure of $\alpha_{|i}$, $\beta_{|i}$ and $\mathbf{z}_{|i}$ when $\mathbf{X}$ is Markov
- We aim to use our results to develop a Markov model for dependence in spatial extremes

Bibliography

Bode, L. and Hardy, T. A. (1997) Progress and recent developments in storm surge modelling. *J. Hydraul. Eng.*, **123**, 315–331.

Butler, A., Heffernan, J. E. amd Flather, R. A. and Tawn, J. A. (2004) Spatial estimation of extremal trends in North Sea surge elevations. In *Proceedings of the IMA meeting on Flood Risk Assessment, University of Bath, 7-8 September 2004.*

Chavez-Demoulin, V. (1999) *Two Problems in Environmental Statistics: Capture-Recapture Analysis and Smooth Extremal models.* Ph.D. thesis, Swiss Federal Institute of Technology (Laussanne).

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

Heffernan, J. E. and Tawn, J. A. (2004) A conditional approach for multivariate extreme values (with discussion). *J. Roy. Statist. Soc. Ser. B*, **66**, 497–546.

Ledford, A. W. and Tawn, J. A. (1997) Modelling dependence within joint tail regions. *J. Roy. Statist. Soc. Ser. B*, **59**, 475–499.

Pauli, F. and Coles, S. G. (2001) Penalized likelihood inference in extreme value analyses. *J. Appl. Statist.*, **28**, 547–560.

Resnick, S. I. (1987) *Extreme Values, Regular Variation, and Point Processes.* Berlin: Springer-Verlag.

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.

Weissman, I. (1978) Estimation of parameters and large quantities based on the k largest observations. *J. Amer. Statist. Assoc.*, **73**, 812–815.