Extreme value analysis of the output from a deterministic hydrodynamical model

at the ***3rd International Symposium on Extreme Value Analysis: Theory and Practice, 19-23 July 2004, Aveiro, Portugal***

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

Jonathan A. Tawn (Lancaster University), j.tawn@lancaster.ac.uk

Janet E. Heffernan (Lancaster University), j.heffernan@lancaster.ac.uk

Roger A. Flather (Proudman Oceanographic Laboratory), raf@pol.ac.uk

Download from:
www.maths.lancs.ac.uk/-butler/presentations/eva2004.ps

Acknowledgements:

Barry Rowlingson (Lancaster University), David Blackman (POL)

Data sources:

Proudman Oceanographic Laboratory (POL), Climatic Research Unit (CRU), Danmarks Meteorologiske Institut (DMI), Rijksinstituut voor Kust en Zee (RIKZ), Det Norske Meteorologiske Institut (DNMI)

CONTENTS

Geophysical models

- Geophysical models for environmental processes
- Deterministic models, based upon discrete linear approximations to sets of fundamental physical equations
- Models are typically evaluated numerically
- Generate massive quantities of highly-structured output

Issues

- Model structure: parameterization, resolution
- Inputs: forcing data, initial/boundary conditions
- Validation against (often sparse) observational data
- Interpretation: statistical significance, attribution

Statistical analysis

- Full Monte Carlo analysis (`climateprediction.net`)
- Statistical emulation: the Bayesian SACCO methodology e.g. Oakley & O'Hagan (2002), Goldstein & Rougier (2004)
- 'Naive' statistical analysis of output e.g. Ferro et al. (2004)

2 STORM SURGES IN THE NORTH SEA

Sea levels (Pugh, 1987)

- Sea level = Mean sea level + Tide + Surge + Waves
- Surges (Flather, 2001) driven by wind and pressure.
- Surge levels linked to meteorology, depth & basin shape
- Tide and surge processes interact (Prandle and Wolf, 1978)

Changing storminess in the NE Atlantic

- Hypothesis: "Global warming" $\Longrightarrow$ increased storminess $\Longrightarrow$ increased storm surge levels (Pugh and Maul, 1999)
- Observational evidence:
  WASA (1998), Alexandersson et al. (2000) (storminess),
  Dixon and Tawn (1992), Bijl et al. (1999) (storm surges).

CSX-DNMI surge hindcast (Flather et al., 1998)

- CSX: a 2D storm surge model for the NE Atlantic, $\approx$35km spatial resolution (Flather et al., 1991)
- Meteorological forcing by interpolated DNMI pressure data for the period 1955-2000 (Reistad and Iden, 1995)
- An analysis of a similar hindcast (Langenberg et al., 1999) found increases in extreme sea levels on the European coast

Figure 1: CSX model grid.

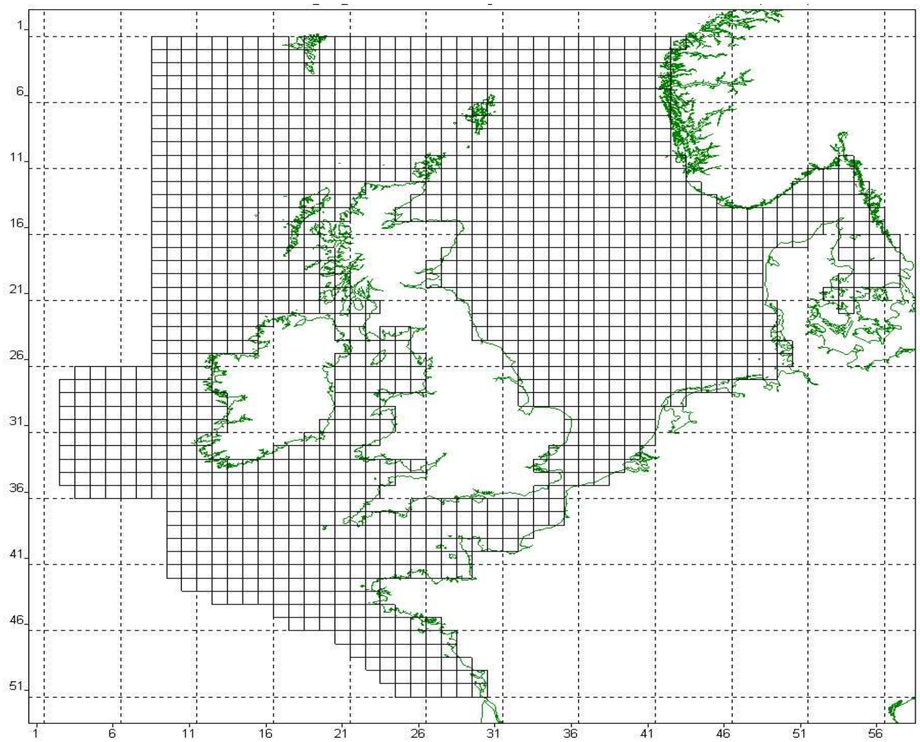


Figure 2: Water depth at CSX grid cells within the North Sea (m)

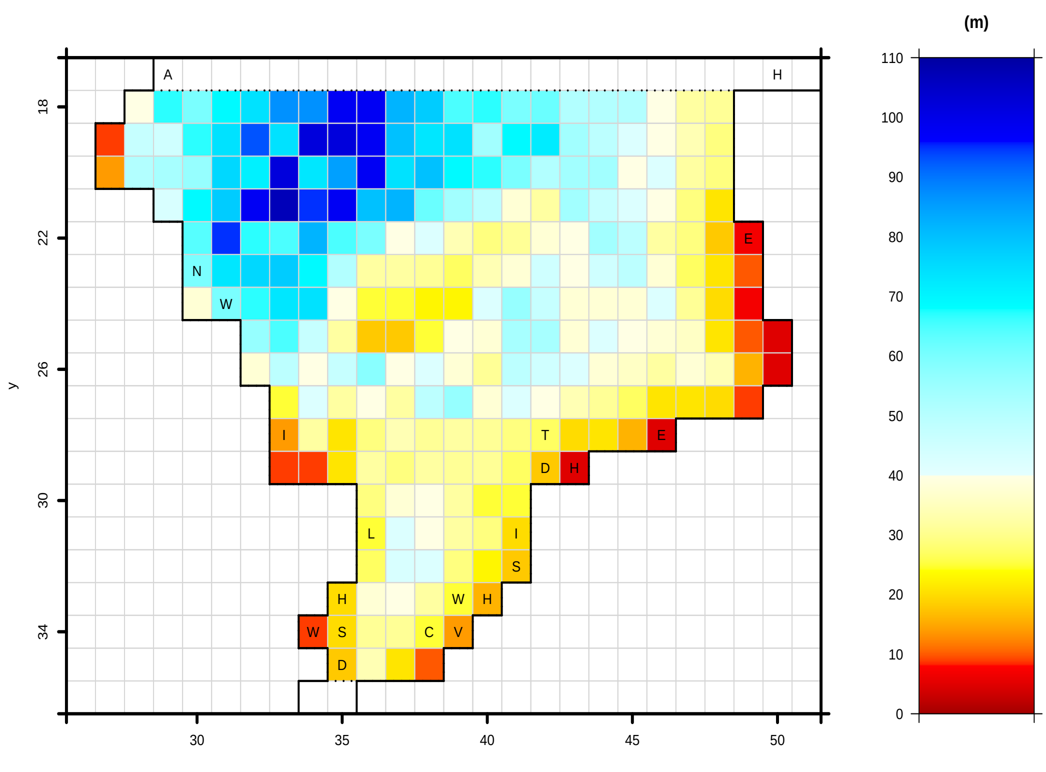


Figure 3: Maximum observed CSX-DNMI tide 1955-2000 (cm)

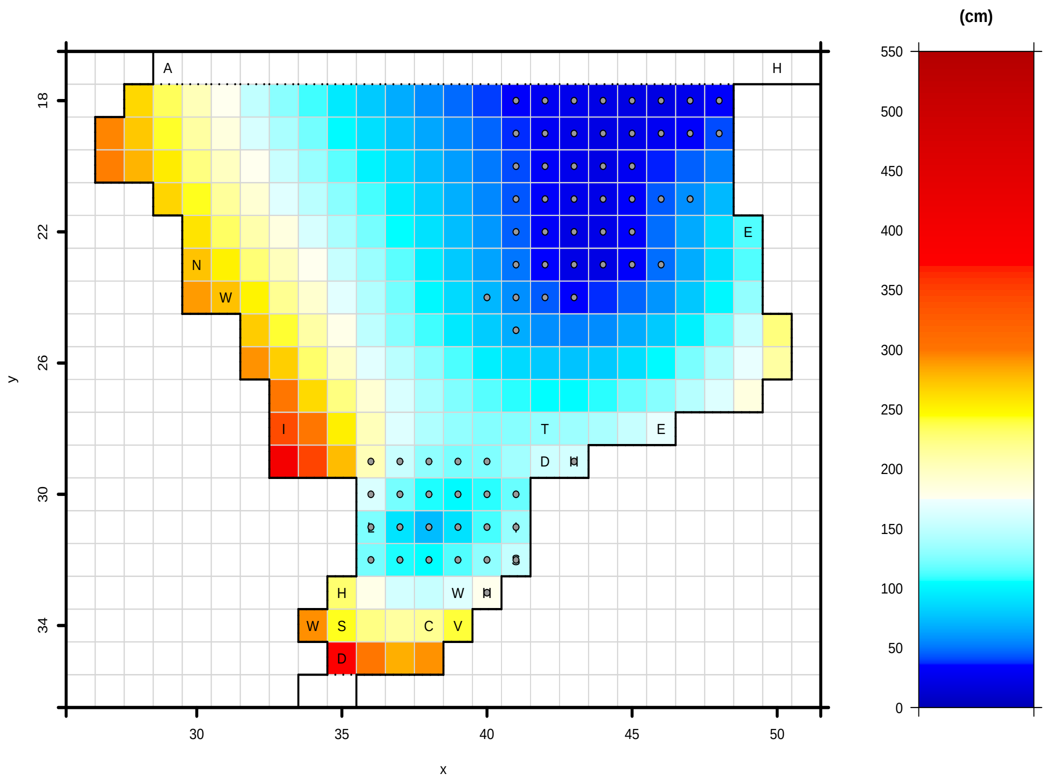
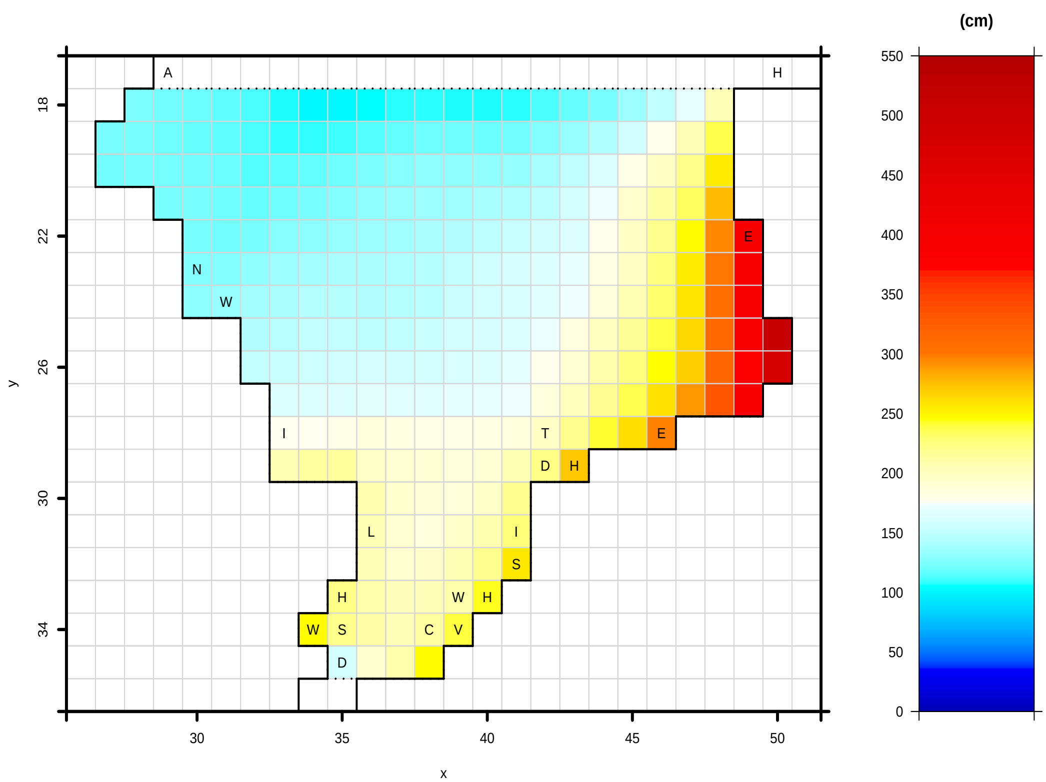


Figure 4: Maximum observed CSX-DNMI surge 1955-2000 (cm)



Data

- At each site we have hourly data for n years $t_1, \ldots, t_n$.
- We use the r-largest approach - relative robustness
- Use the window declustering algorithm (Tawn, 1988) to extract r-largest storm peaks per year
- Let $x_j^{(k)}$ be the k-th largest storm peak in year t_j

The r-largest model

- Assume $x_j^{(k)}$ is a realisation from a random variable $X_j^{(k)}$.
- Assume $\mathbf{X}_j = \left(X_j^{(1)}, \ldots, X_j^{(r)}\right)$ are jointly distributed according to an r-largest model.
- Weissman (1978); Smith (1986); Tawn (1988).
- When $r = 1$ reduces to the GEV model for annual maxima.
- r must be small for the asymptotic motivation to be valid.
- Parameter vector $\theta_j = (\mu_j, \sigma_j, \xi_j)$: location, scale, shape.
- A typical design parameter is the N-year return level

$$z_j(N) = \mu_j - \frac{\sigma_j}{\xi_j}\left\{1 - \left[-\log\left(\frac{N-1}{N}\right)\right]^{\xi_j}\right\}.$$

Time constant

- Used by Flather et al. (1998) to analyse CSX-DNMI surges
- Inference via numerical maximum likelihood

Polynomial trend models

- $\mu(t)$ and $\sigma(t)$ each polynomial in time; $\xi(t)$ constant.
- $\mu(t)$ polynomial in time; $\sigma(t) = \nu\mu(t)$; $\xi(t)$ constant.
- For a high threshold u: rate of exceedance of u and expected exceedance of u each polynomial in time; $\xi(t)$ constant.

Non-smooth models

- Use North Atlantic Oscillation index as a covariate in $\mu(t)$
- Assume there is a changepoint in an unknown year t_J

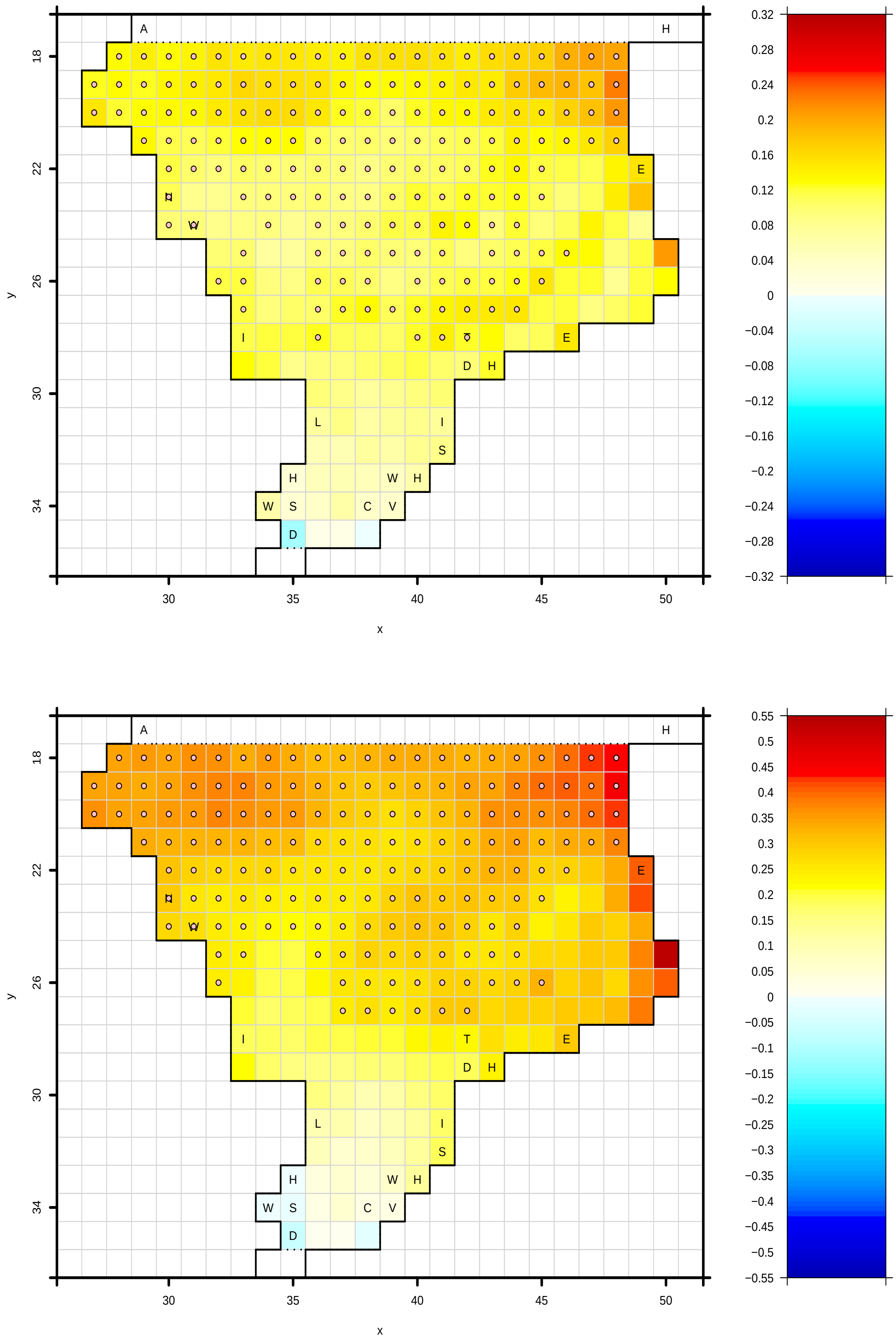
Local trend modelling

- Local likelihood, a kernel approach to nonparametric regression (Hall and Tajvidi, 2000; Davison and Ramesh, 2000).
- Alternative penalized likelihood approach: Pauli and Coles (2001), Chavez-Demoulin and Davison (2004)

Figure 5: Spatial distribution of estimates for the trend parameter μ_β (cm/y) in the r-largest models for which

Upper: $\mu(t) = \mu_\alpha + \mu_\beta t$, $\sigma(t) = \sigma, \xi(t) = \xi$

Lower: $\mu(t) = \mu_\alpha + \mu_\beta t$, $\sigma(t) = \nu\mu(t), \xi(t) = \xi$



Local constant estimation

- To estimate θ_j maximise the local likelihood function
$$\sum_{J=1}^{n} K(t_J - t_j; h) \log \{f(\mathbf{x}_J; \theta_j)\},$$
where $f(\mathbf{x}_j; \theta_j)$ denotes the density of $\mathbf{X}_j$.
- K is a kernel function with bandwidth h.
- Can fix some parameters over time e.g. fix $\xi(t)$ or $\sigma(t)/\mu(t)$.

Local regression

- Assume a regression model for the behaviour of θ_j in the local region around t_j.
- Can reduce bias close to the ends of the time range.

Statistical issues

- Bandwidth selection
- Uncertainty estimation: based on the semiparametric bootstrap procedure proposed by Davison and Ramesh (2000)
- Diagnostics: use a “de-standardized” quantile-quantile plot

Figure 6: Spatial distribution of 'relative' 50y return level estimates from a local constant model ($r = 20$, $h = 3.5$y) in specific years.
'Relative': divided by the time-constant 50y r.l. estimate for that site

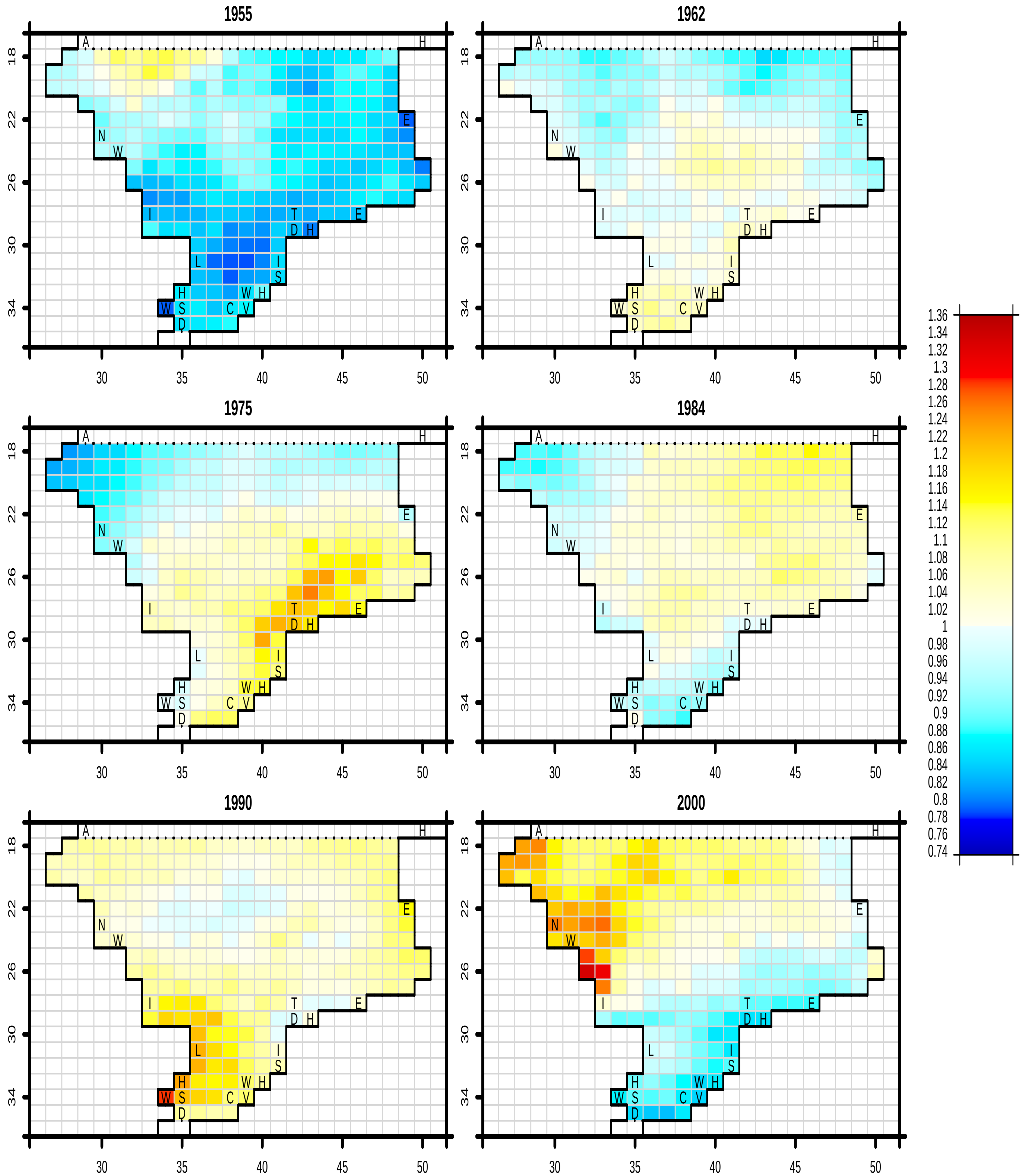


Figure 7: Spatial distribution of raw 50y return level estimates, from a local constant model (r=20, h=3.5y) in specific years.

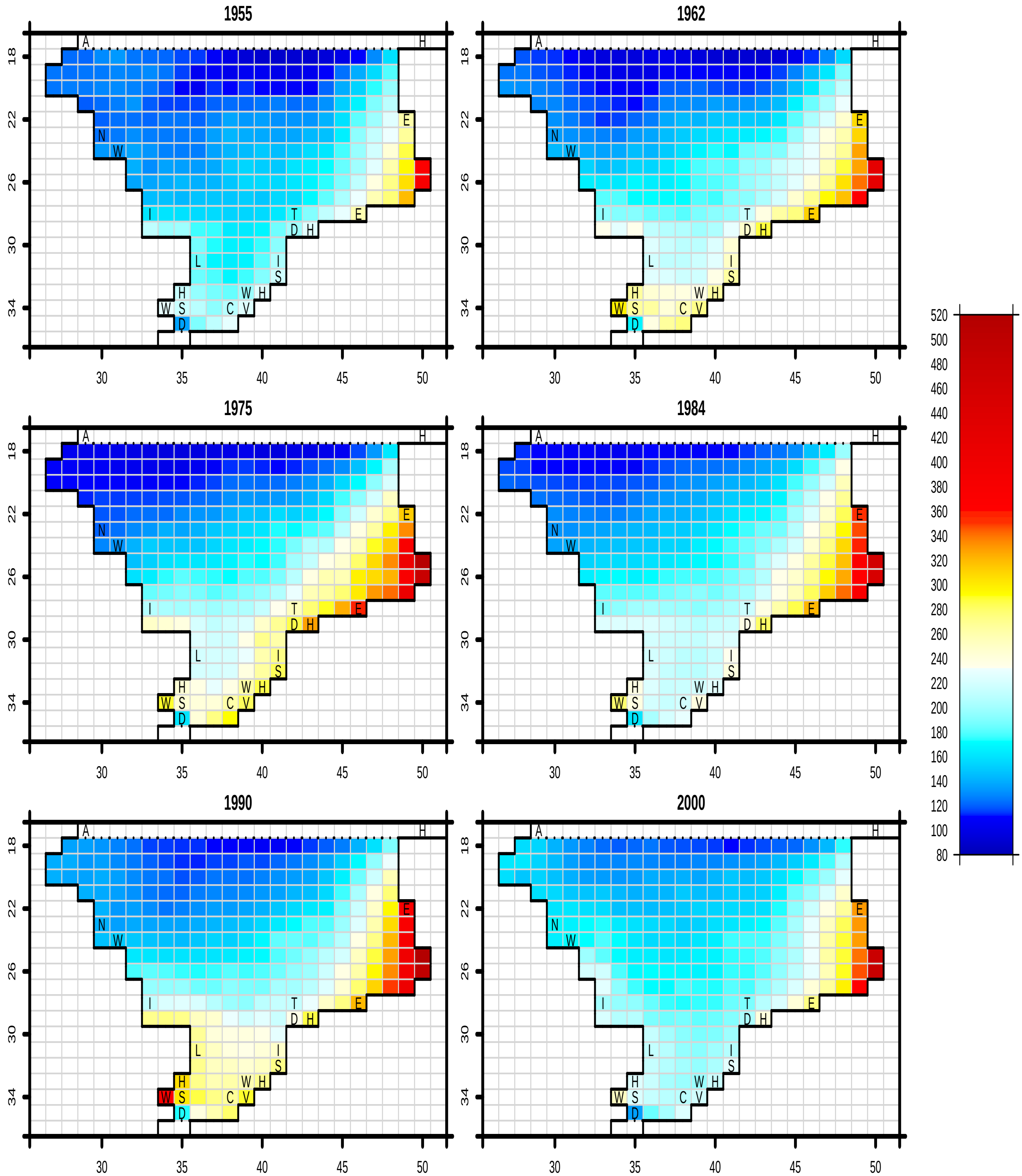
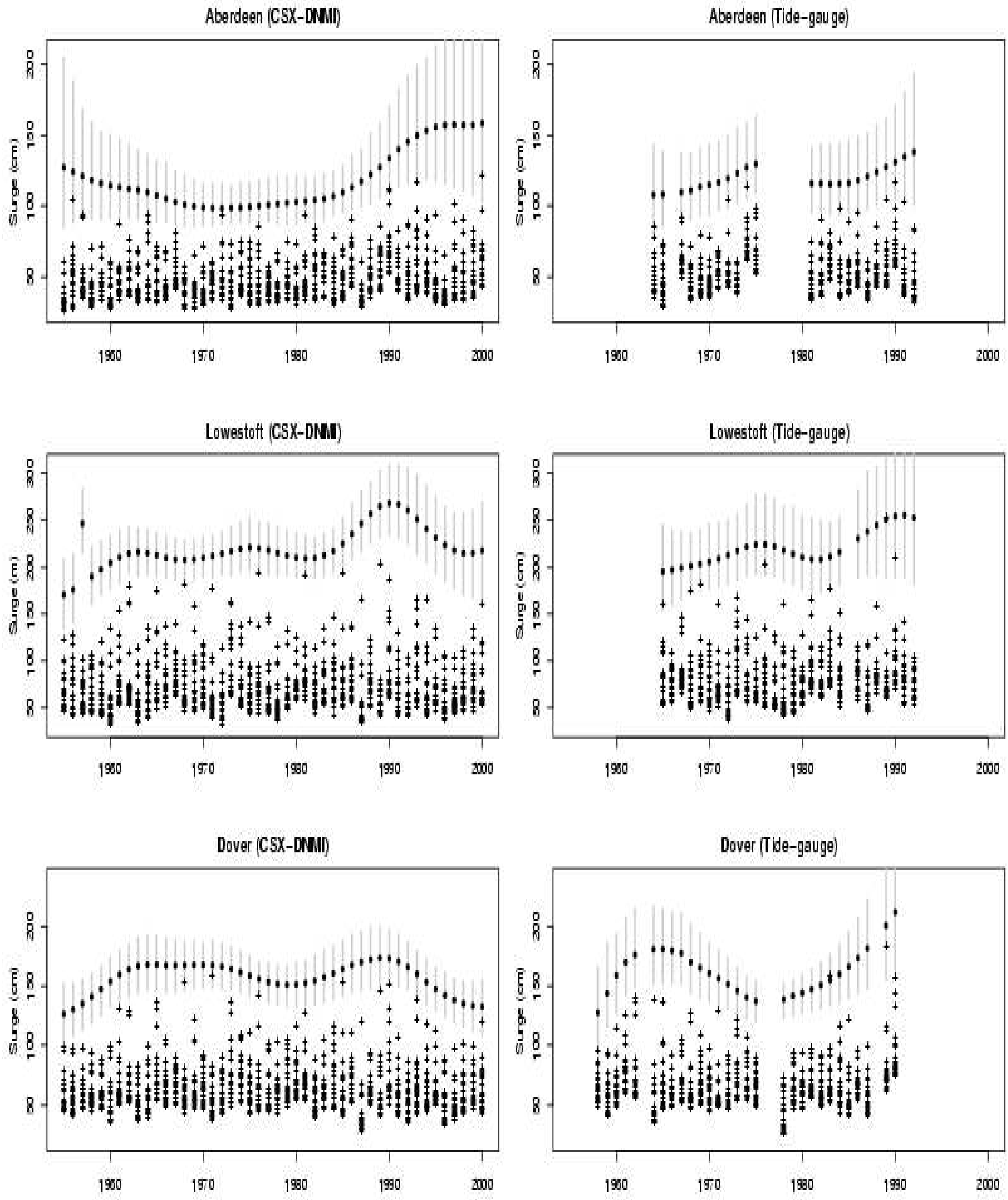


Figure 8: 20-largest surges per year from CSX-DNMI and tide gauge data (+), & estimated 50y return levels from a local linear r-largest model with $h = 3.5$y (•). 95% variability bands also shown.



6 CONCLUSIONS

Trends in storm surges

- Forth, Tyne, Dogger: strong overall increase 1955-2000, dominated by rapid increase 1980-2000
- German Bight: moderate overall increase 1955-2000, but marked decrease 1990-2000
- Thames, Humber: no clear long-term trend

Extension: incorporating spatial information

- Exploit spatial information to reduce uncertainty in trend estimates: Dixon and Tawn (1992), Barao and Tawn (1999)
- Smoothing / linking parameters over space (Smith, 1999)
- Explicit modelling of dependence

Wider context

- Case study in the application of extreme value methods to geophysical model output