PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Statistics of Coastal Flood Prevention

Stuart G. Coles and Jonathan A. Tawn

Phil. Trans. R. Soc. Lond. A 1990 **332**, 457-476
doi: 10.1098/rsta.1990.0126

Statistics of coastal flood prevention

By Stuart G. Coles and Jonathan A. Tawn
Department of Probability and Statistics, University of Sheffield, Sheffield S3 7RH, U.K.

For the design of sea defences the main statistical issue is to estimate quantiles of the distribution of annual maximum sea levels for all coastal sites. Traditional procedures independently analyse data from each individual site; thus known spatial properties of the meteorological and astronomical tidal components of sea level are not exploited. By spatial modelling of the marginal behaviour and inter-site dependence of sea level annual maxima around the British coast we are able to examine risk assessment for coastlines and the issue of sensitivity to climatic change.

1. Introduction

Public awareness of the issues surrounding coastal flooding has been heightened by the well-reported effects of a number of extreme storms, and speculation about the likely impact of climatic changes due to the greenhouse effect. Engineers and scientists share these same concerns, asking: 'To what height should a particular sea wall be built?'; 'What level of protection will it offer?'; 'If there is flooding in one locality what is the probability of flooding elsewhere?'; 'What effect would a rise in mean sea levels have on extreme events?' To answer such questions a detailed understanding of both the marginal and spatial behaviour of extreme sea levels is required.

The primary aim of this paper is to formulate a spatial model for the marginal distributions of extreme sea levels around mainland Britain, exploiting and extending recent statistical methodology for extreme values and taking account of the dynamics of the sea-level process. Temporal variability in the sea-level process is essentially due to two different physical mechanisms: astronomically induced tides, and meteorological storm surges. By accounting for the role of the tide, and by modelling the spatial coherence of the surge process, the spatial model improves standard marginal analyses and allows interpolation to sites with no data.

Typically, the marginal behaviour of sea level extremes is estimated by the use of the classical annual maxima approach. The distribution of the annual maximum sea level is taken to be generalized extreme value (GEV) $\mathscr{G}(\mu,\sigma,k)$, with distribution function

$$G(z) = \exp\left[-\{1-k(z-\mu)/\sigma\}^{\frac{1}{k}}\right], \quad 1-k(z-\mu)/\sigma > 0, \tag{1.1}$$

for parameters $\mu, \sigma > 0, k$. The case $k = 0$ is taken as the limit as $k \to 0$ giving the Gumbel distribution. Maximum likelihood is used to fit the model to the observed series of annual maxima at the site. The basis for the GEV model is that this family encompasses the complete class of limiting distributions of linearly normalized

Phil. Trans. R. Soc. Lond. A (1990) **332**, 457–476
Printed in Great Britain

maxima from weakly mixing stationary sequences (Leadbetter *et al.* 1983). Thus, in applications, it is implicitly assumed that the number of hourly observations in a year is sufficiently large to justify the asymptotic arguments. A statistic of interest for sea defence design is the return level, z_p, defined as the level exceeded by the annual maxima with probability p: i.e. $G(z_p) = 1-p$, so from (1.1)

$$z_p = \mu + \sigma k^{-1}[1-\{-\ln(1-p)\}^k]. \tag{1.2}$$

Substitution of the maximum likelihood estimates of μ, σ, k into (1.2) gives an estimate of z_p.

There are two apparent deficiencies of the annual maxima method as applied to sea levels. First, no allowance is made for the fact that the tidal component of the sea level is deterministically induced by astronomical forcing. An improved approach, which analyses the constituent parts of the process (tide and meteorological surge), is described by Tawn & Vassie (1989). This method fails, however, to reconcile the second deficiency of the annual maxima approach: that marginal analyses ignore the spatial coherence of the meteorological storm surge process.

A useful corollary of Tawn & Vassie's work is a formulation of the effect of the tide on the parameters of the annual maximum sea level distribution (Tawn 1990*a*). Exploiting this relation we construct, in §3, a spatial model for annual maximum sea levels where the parameters of the distribution at a site are functionally dependent on the tidal characteristics and geographical location of the site. Thus both tidal and spatial features of the sea-level process are modelled. In addition, it is known that in some shallow-water areas there is significant interaction between the tide and surge components (Prandle & Wolf 1978; Wolf 1978). Such effects are modelled empirically through functional forms for parameters, see §3*e*.

Estimates of return levels based on the above model assume the behaviour of the process will remain unchanged in future years. Currently, however, there is much debate about the impact on extreme sea levels of the greenhouse effect and associated climatic changes. Despite the pressing need to make provision for such effects the diverse, and often conflicting, nature of the various scientific hypotheses makes standard prediction techniques impossible. It is likely, though, that the overall structure of the statistical model for extremes will remain stable, and under this assumption in §4 we explore the sensitivity of return levels to changes in individual aspects of the model, thought to be plausible consequences of the greenhouse effect. It is then possible to identify which coastal regions are most vulnerable under the differing scientific conjectures for change.

Finally, to consider such issues as regional flooding, rather than localized flooding, it is necessary to account for spatial dependence in sea-level extremes. Multivariate extreme value theory provides a natural framework for modelling such dependence. The extent to which this theory can be exploited when interest is in many sites along a coastline is considered in §5.

2. The data and covariates

The data-set is an updated subset, edited for erroneous data, of that collected and analysed by Graff (1981) consisting of annual maximum sea levels, after surface waves have been averaged out, for 61 mainland British coastal sites, which are measured relative to a common datum, Ordnance Datum Newlyn (ODN). Details of the data and location of sites can be found in Graff's paper. A list of sites considered,

MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES
THE ROYAL SOCIETY A
PHILOSOPHICAL TRANSACTIONS OF

periods over which data were collected, and the available number of annual maxima for each site are given in table 1. These sites provide a wide coverage of open sea coastlines and a more detailed coverage of the Firth of Forth, Humber, Mersey, Thames and Wash estuarine regions.

Although extensive hourly sea level observations are available for some of these and other sites, we restrict attention to annual maxima. This is partly for simplicity, but mainly because the aim in this paper is to provide an exploratory spatial analysis of extreme sea levels rather than precise results for sea wall design.

The covariate information used consists of location, inter-site distances and an approximation to the tidal sequence at each site. Inter-site distances, in miles, were obtained by measuring the coastal convex hull between sites using Ordnance Survey maps, thus representing the distance surges travel between sites. The tidal sequence at time t hours, for any site i, can be written as

$$X_{i,t} = \sum_{j=1}^{m_i} H_{i,j} \cos(w_j t + \phi_{i,j}), \tag{2.1}$$

where the w_j are astronomically determined and m_i depends on the site; typically $60 \leqslant m_i \leqslant 102$, see Pugh (1987). We crudely represent the tide by $\tilde{X}_{i,t}$, the sum of the four dominant tidal constituents: M_2, the semi-diurnal lunar tide; S_2, the semi-diurnal solar tide; O_1, the diurnal lunar tide; and K_1, the diurnal luni-solar tide, with $w = 0.5059, 0.5234, 0.2625, 0.2428$ rad h^{-1} respectively. For these four constituents and for an extensive range of sites, $H_{i,j}$ and $\phi_{i,j}$ are given in the Admiralty tide tables, vol. 1 (1990). Where tidal information is not available a linear interpolate estimator for both parameters of each tidal constituent is used. For a site, i say, a useful measure of tidal range is given by mean high water springs, $\tilde{H}_i = H_{i,M_2} + H_{i,S_2}$, and an impression of the spatial variability of this statistic is given by figure 1.

3. Spatial modelling of annual maxima distributions

(*a*) *Separate site analyses*

Let $Z_{i,j}$ define the annual maximum sea level at site i in year j. In this section we consider estimation of the distribution of the annual maxima at site i, using the method described in §1. This requires the sequence $\{Z_{i,j}\}_j$ to be stationary, which is not the case for sea levels as long-term changes are caused by the following compounding factors: changes in global water levels, variations of storminess, localized or regional land rise or shrinkage, and dredging effects in estuaries (Lamb 1980; Pugh 1987). As the only detectable change in the distribution of the annual maxima, over time, is an approximately linear trend we extend the model to $Z_{i,j} \sim \mathscr{G}(\mu_{i,j}, \sigma_i, k_i)$, with $\mu_{i,j} = \alpha_i + j\beta_i$ (see Smith 1986; Tawn 1988*a*).

For the year 1990, table 1 gives the estimated return levels (1.2) with standard errors for $p = 0.1, 0.01, 0.001$ over all data sites. The standard errors depend on the shape of the tail of the annual maxima distribution as much as on the sample sizes, and so are larger on the east coast than the west coast. Lennon (1963*a*), Suthons (1963) and Graff (1981) have each done annual maxima analyses on subsets of these data using different methods for estimation and for handling trends. Unfortunately, none of these analyses gives a measure of precision to their estimates.

Table 1. *Data sites, data spans and separate site based estimates and standard errors* (SE) *for return levels,* z_p *(metres with respect to ODN)*

	site, span of data and number of observations	$z_{0.1}$ (SE)	$z_{0.01}$ (SE)	$z_{0.001}$ (SE)
1	Ullapool 1963–1977 (12)	3.79 (0.29)	4.04 (0.46)	4.24 (0.74)
2	Gourock 1920–1978 (56)	3.01 (0.08)	3.26 (0.13)	3.43 (0.22)
3	Ardrossan 1944–1978 (35)	2.78 (0.16)	3.44 (0.45)	4.21 (1.11)
4	Silloth 1928–1978 (39)	6.23 (0.13)	6.69 (0.24)	7.08 (0.49)
5	Barrow 1920–1978 (19)	5.79 (0.13)	6.39 (0.60)	7.30 (1.94)
6	Heysham 1940–1984 (36)	6.13 (0.13)	6.93 (0.57)	8.13 (1.84)
7	Fleetwood 1930–1983 (48)	6.01 (0.08)	6.21 (0.08)	6.30 (0.10)
8	Gladstone Dock 1956–1977 (20)	6.08 (0.14)	6.21 (0.13)	6.25 (0.14)
9	Princes Pier 1941–1977 (37)	6.10 (0.14)	6.24 (0.10)	6.31 (0.08)
10	Georges Pier 1857–1903 (42)	5.82 (0.27)	6.12 (0.28)	6.33 (0.34)
11	Eastham Lock 1956–1977 (19)	6.43 (0.06)	6.48 (0.06)	6.49 (0.06)
12	Hilbre Island 1854–1981 (80)	5.51 (0.06)	5.78 (0.09)	5.96 (0.16)
13	Fishguard 1959–1977 (19)	3.23 (0.06)	3.28 (0.05)	3.29 (0.05)
14	Milford Haven 1954–1977 (23)	4.31 (0.13)	4.45 (0.15)	4.53 (0.23)
15	Swansea 1936–1981 (36)	5.76 (0.03)	5.80 (0.03)	5.81 (0.03)
16	Cardiff 1931–1982 (41)	7.44 (0.11)	7.71 (0.10)	7.86 (0.14)
17	Newport 1899–1988 (37)	7.93 (0.08)	8.27 (0.14)	8.50 (0.24)
18	Avonmouth 1924–1986 (61)	8.39 (0.09)	8.71 (0.15)	8.90 (0.22)
19	Newlyn 1916–1976 (61)	3.20 (0.04)	3.31 (0.05)	3.38 (0.07)
20	Devonport 1920–1977 (38)	3.15 (0.07)	3.31 (0.14)	3.41 (0.23)
21	Portland 1923–1977 (20)	1.90 (0.08)	2.05 (0.08)	2.16 (0.13)
22	Calshot 1930–1976 (42)	2.49 (0.06)	2.57 (0.08)	2.61 (0.09)
23	Southampton 1924–1975 (47)	2.68 (0.07)	2.89 (0.12)	3.04 (0.20)
24	Portsmouth 1813–1975 (104)	2.82 (0.04)	3.09 (0.09)	3.35 (0.20)
25	Newhaven 1913–1976 (60)	4.09 (0.05)	4.21 (0.06)	4.27 (0.08)
26	Pevensey 1953–1976 (24)	4.73 (0.14)	4.89 (0.16)	4.97 (0.19)
27	Rye 1949–1974 (16)	4.71 (0.13)	5.06 (0.31)	5.46 (0.74)
28	Dover 1912–1984 (62)	4.21 (0.08)	4.82 (0.30)	5.69 (0.83)
29	Margate 1968–1977 (10)	3.20 (0.52)	5.05 (3.68)	10.41 (19.61)
30	Sheerness 1819–1983 (136)	3.95 (0.06)	4.48 (0.15)	5.06 (0.35)
31	Tower Pier 1929–1977 (49)	5.21 (0.10)	5.75 (0.30)	6.40 (0.74)
32	Tilbury 1929–1977 (46)	4.55 (0.12)	5.25 (0.44)	6.23 (1.25)
33	Southend 1929–1986 (57)	4.01 (0.09)	4.53 (0.24)	5.09 (0.53)
34	Colchester 1944–1986 (43)	3.94 (0.07)	4.22 (0.18)	4.45 (0.37)
35	Holland-on-Sea 1934–1986 (53)	3.48 (0.12)	4.39 (0.54)	5.86 (1.74)
36	Walton 1968–1982 (15)	2.91 (0.13)	3.28 (0.39)	3.68 (0.90)
37	Harwich 1926–1976 (51)	3.31 (0.11)	3.92 (0.27)	4.63 (0.63)
38	Lowestoft 1953–1983 (31)	2.70 (0.17)	3.52 (0.49)	4.54 (1.24)
39	Gt Yarmouth 1899–1976 (77)	2.51 (0.11)	3.20 (0.29)	3.93 (0.67)
40	Kings Lynn 1860–1978 (119)	5.19 (0.06)	5.77 (0.18)	6.44 (0.44)
41	Wisbech Cut 1957–1978 (22)	5.23 (0.19)	5.95 (0.59)	7.01 (1.73)
42	Lawyers Sluice 1953–1978 (26)	5.27 (0.23)	6.15 (0.71)	7.37 (1.97)
43	Marsh Road Sluice 1953–1978 (17)	5.87 (0.22)	6.63 (0.70)	7.69 (1.90)
44	Boston 1920–1978 (59)	5.29 (0.07)	5.65 (0.15)	5.98 (0.32)
45	Grimsby 1920–1973 (54)	4.02 (0.11)	4.38 (0.16)	4.66 (0.27)
46	Immingham 1920–1988 (69)	4.56 (0.06)	4.92 (0.14)	5.23 (0.30)
47	Goole 1920–1978 (59)	5.75 (0.06)	5.92 (0.06)	5.99 (0.07)
48	Blacktoft 1921–1977 (56)	5.45 (0.07)	5.67 (0.09)	5.83 (0.14)
49	Brough 1922–1977 (56)	5.35 (0.07)	5.57 (0.10)	5.71 (0.18)
50	St Andrews Dock 1920–1973 (49)	4.95 (0.07)	5.17 (0.11)	5.34 (0.23)
51	Victoria Dock 1920–1969 (24)	4.88 (0.12)	5.13 (0.18)	5.31 (0.37)
52	Humber Dock 1920–1968 (42)	4.76 (0.11)	4.97 (0.15)	5.11 (0.21)
53	King Georges Dock 1922–1973 (36)	4.74 (0.09)	5.11 (0.21)	5.53 (0.49)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

site, span of data and number of observations	$z_{0.1}$ (SE)	$z_{0.01}$ (SE)	$z_{0.001}$ (SE)
54 Saltend Jetty 1965–1977 (13)	5.06 (0.25)	6.42 (2.45)	10.84 (14.97)
55 North Shields 1959–1978 (19)	3.54 (0.18)	4.53 (1.67)	7.38 (9.16)
56 Leith 1939–1978 (38)	3.55 (0.08)	3.70 (0.10)	3.79 (0.15)
57 Grangemouth 1934–1978 (34)	4.02 (0.11)	4.38 (0.16)	4.66 (0.27)
58 Rosyth 1914–1977 (31)	3.77 (0.08)	3.98 (0.14)	4.11 (0.23)
59 Kircaldy 1951–1978 (28)	3.42 (0.10)	3.65 (0.16)	3.83 (0.29)
60 Methil 1934–1977 (38)	3.82 (0.07)	4.05 (0.14)	4.26 (0.30)
61 Aberdeen 1908–1975 (67)	2.88 (0.05)	3.07 (0.08)	3.22 (0.13)

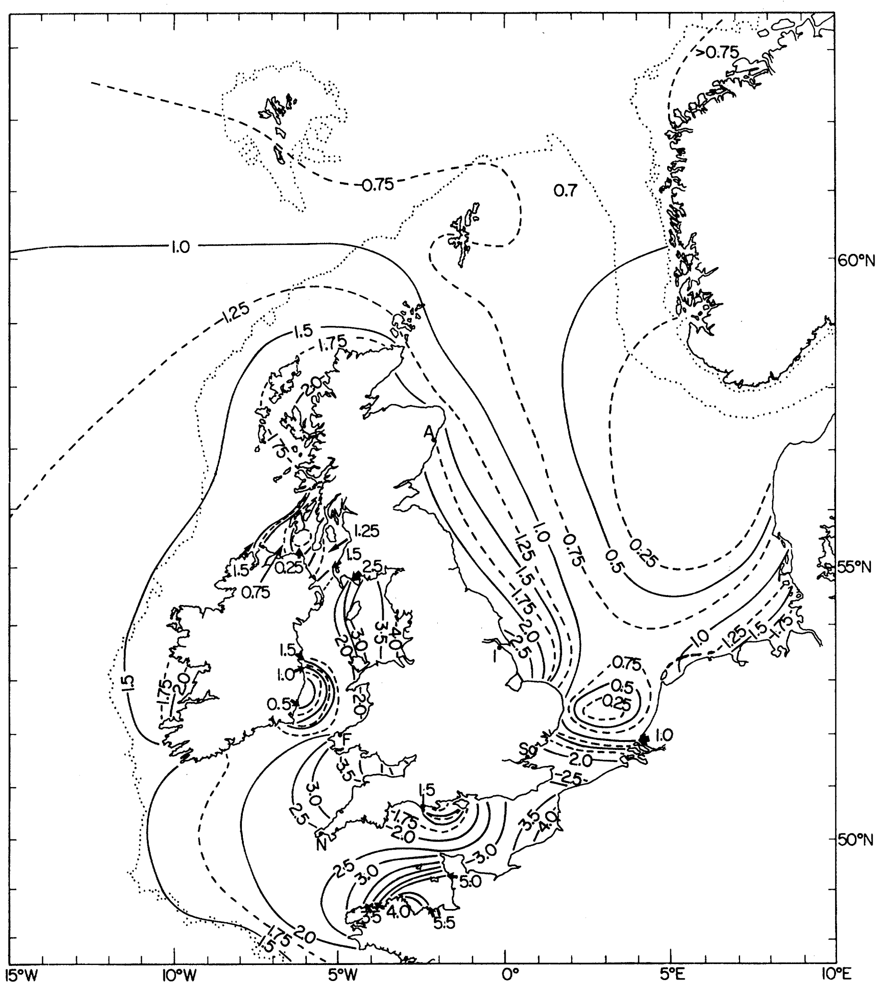
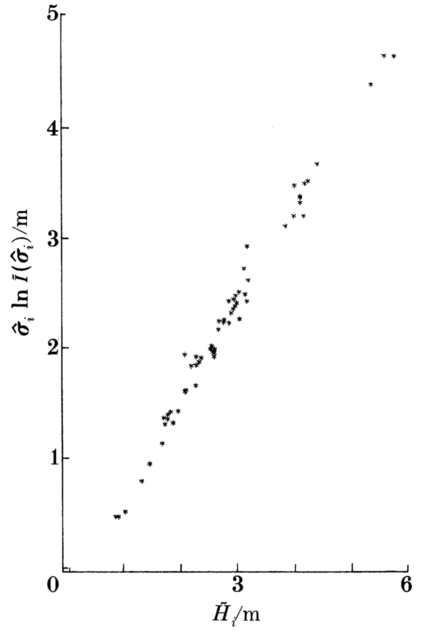


Figure 1. Map of $\tilde{H} = (H_{M_2} + H_{S_2})$, in metres, over the European continental shelf. A, I, So and F denote the sites Aberdeen, Immingham, Southend and Fishguard respectively.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES



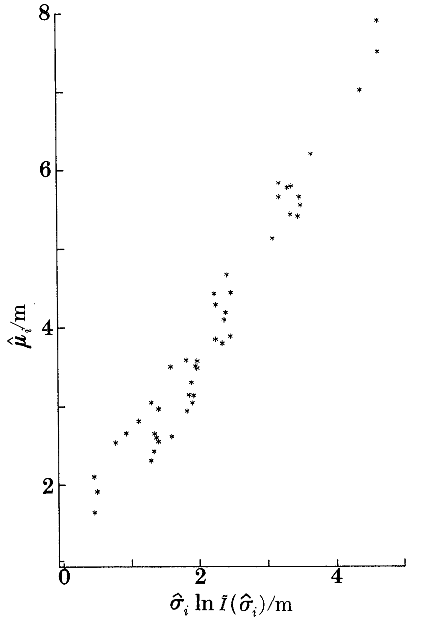


Figure 2 (left). $\hat{\sigma}_i \log_{10} \tilde{I}(\hat{\sigma}_i)$ against $\tilde{H}_i$ for all data sites – $\hat{\sigma}_i$ based on marginal estimates of σ_i.
Figure 3 (right). $\hat{\mu}_i$ against $\hat{\sigma}_i \ln \tilde{I}(\hat{\sigma}_i)$ for all sites excluding estuarine regions; $\hat{\mu}_i, \hat{\sigma}_i$ based on marginal estimates.

(*b*) *Tidal and regional heterogeneity*

By considering the sea level as an independent additive combination of tide and surge, Tawn (1990*a*) showed that, marginally for site i, a good approximation to the distribution of the annual maximum sea level when $k_i = 0$ is $Z_{i,j} \sim \mathscr{G}(\alpha_i + j\beta_i + \sigma_i \ln I(\sigma_i), \sigma_i, 0)$. Here

$$I(\sigma_i) = T^{-1} \sum_{t=1}^{T} \exp(X_{i,t}/\sigma_i),$$

where T is the full period of the tide and $X_{i,t}$ is the tide defined in (2.1). When $k_i \neq 0$, or if there is interaction between the tide and surge components, this structural dependence on the tide continues to hold due to the approximate invariance of the location parameter to k_i; thus we take

$$Z_{i,j} \sim \mathscr{G}(\alpha_i + j\beta_i + \sigma_i \ln \tilde{I}(\sigma_i), \sigma_i, k_i).$$

Here, $\tilde{I}(\sigma_i)$ is $I(\sigma_i)$ with $X_{i,t}$ replaced by the crude tidal sequence $\tilde{X}_{i,t}$, see §2. Figure 2 shows the strong dependence of $\sigma_i \ln \tilde{I}(\sigma_i)$, in which marginal estimates are used for σ_i, on $\tilde{H}_i$ over all sites.

The parameters α_i, σ_i, k_i are characteristics of the extreme surges that occur near high tides and so should reflect the spatial continuity of the surge process. Along the east and south coasts the surge process is particularly coherent (Davies & Flather 1978; George & Thomas 1976 respectively), while on the west coast the structure is more complex (Proctor & Flather 1989; Amin 1982). The trend parameter β_i, discussed in §3*a* varies widely around the coastline for our data, with most

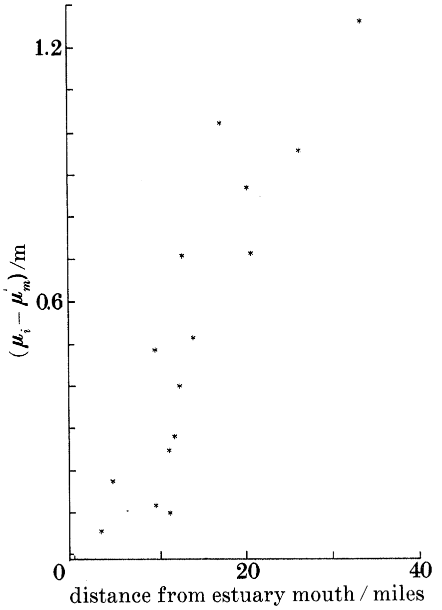


Figure 4. Increase in location parameter, $\hat{\mu}_i - \hat{\mu}_m$, where $\hat{\mu}_m$ is the location parameter at the estuary mouth, against distance from estuary mouth for all estuarine sites, $\hat{\mu}_i$ based on marginal estimates.

significant regional increases in the Humber, Mersey and Thames estuaries, possibly due to dredging. Elsewhere, estimates are generally not spatially stable suggesting that localized factors often mask regional changes. Linking the observed results to mean sea level (MSL) trends around Britain (Woodworth 1987) may clarify this difficult problem, but even MSL seems highly dependent on local characteristics. To quote Johnson (1929): 'along an irregular coast the MSL surface is an irregularly warped plane which is extremely sensitive to alterations in the form of the shore, the depths of the tidal channels and other shore features ... changes in these features ... cause local fluctuations in measured MSL ranging from fractions of an inch to ... one or two feet'. Because of the form of the trend the variability of both α_i and $\alpha_i + j\beta_i$ is partly due to site specific influences, so it is unrealistic to model these parameters spatially; instead we restrict attention to explaining the major sources of the variability of $\alpha_i + j\beta_i$ over sites for the year 1990.

In the above discussion the parameters $\alpha_i + j\beta_i$, σ_i and k_i are treated as if they were independent of the tide. It is conceivable, however, that the topographical features of the ocean and coast that determine the tidal parameters could also influence the marginal distribution of surges. For all open coastline data sites figure 3 shows the location parameter, μ_i for 1990 against $\sigma_i \ln \tilde{I}(\sigma_i)$, the direct tidal influence. The relation is remarkably linear with gradient 1.40, significantly larger than one, demonstrating that the marginal surge distribution is linked to tidal behaviour. This phenomenon, which does not appear to have been discussed in the oceanographic literature, is only identifiable through spatial analysis. The residuals are not spatially dependent although on some coastal stretches there is some pattern that could aid interpolation between data sites. For the estuarine sites figure 4 shows the influence of distance inland, and both the tide and surge at the estuary mouth, formalizing Graff's (1981, p. 410) empirical observation.

Thus the location parameter is time, tide and locality dependent. The parameter σ_i is also observed to be weakly dependent on $\sigma_i \ln \tilde{I}(\sigma_i)$. This, however, is due to the confounding effect of spatial variability in the tide and surge components; subsequently we find that once spatial features are modelled, σ_i and $\sigma_i \ln \tilde{I}(\sigma_i)$ are independent, as is the case for k_i and $\sigma_i \ln \tilde{I}(\sigma_i)$. Both σ_i and k_i exhibit spatial smoothness and it is this feature which we exploit in the remainder of §3.

(c) *Existing regional methods*

To estimate return levels at sites with no observations, or for the estimation of extremely high return levels, methods have been developed that exploit the spatially homogeneous features of the physical process of interest. The idea behind these approaches is that within a homogeneous region extreme events are generated by similar physical mechanisms, so it is valid to use regional data for analyses at individual sites. Such techniques are widely used by hydrologists concerned with extreme rainfall, but regional methods for the analysis of extreme sea levels have been far more elementary. Here we review briefly these procedures before extending them to our application in §3*d*, *e*.

For extreme sea levels Lennon (1963*a*) found empirically that an index equivalent to $(z_{i,0.01} - \tilde{H}_i)/2\tilde{H}_i$ changes smoothly along the west coast of Britain and proposed that for site j, with tidal information only, $z_{j,0.01}$ could be obtained by interpolating estimated indices between data sites. Graff (1981, fig. 25) found that this index was not spatially stable for other coasts, and since from §3*b* the location parameter varies with tide, while σ_i and k_i vary smoothly and independently of the tide, then

$$z_{i,p} - 1.4\sigma_i \ln \tilde{I}(\sigma_i) \approx z_{i,p} - 1.23\tilde{H}_i$$

should be reasonably stable for open coastlines.

For the analysis of extreme rainfalls NERC (1975) and Jenkinson (1977) have proposed standardizing annual maxima data, for each site, by dividing by the respective sample mean annual maxima. Possibly after initial checks for dependence (Buishand 1984), these standardized data are then pooled and a common distribution fitted, treating the data as independent. Moore (1987) and Buishand (1989) have extended this method to enable the GEV parameters to be described by data site covariates, and for covariate effects to be estimated simultaneously in the likelihood, see also Davison & Smith (1990). Fitzgerald (1989) has suggested estimating regional GEV parameters by averaging the corresponding marginal parameters over sites within the homogeneous region. The latter method is liable to lead to significant errors as no account is taken of the precision of each estimate, nor of the dependencies between the GEV parameters at each site.

Each of the regional methods described above fails to account for spatial dependence; thus the standard errors of estimates are significantly underestimated. Ignoring spatial dependence also leads to biased estimates if at some sites there is bias in the selection of data, or if data were not recorded in the year of a highly influential extreme event. There is some evidence of both these features in our data-set.

(d) *The model*

A complete description of the marginal and spatial dependence structures of sea level annual maxima is given by the joint distribution function over sites

$$G(z_1, \ldots, z_{61}) = \Pr(Z_1 \leqslant z_1, \ldots, Z_{61} \leqslant z_{61}).$$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Appropriate models for G are the class of multivariate extreme value distributions, see §5*a*, but the high dimensionality of the current problem makes a general approach intractable without further assumptions. In particular we will assume that:

(i) spatially, the process of sea-level annual maxima is a first-order Markov process: for ordered sites i, j, k, Z_i and Z_k are conditionally independent given Z_j, and

(ii) the bivariate dependence structure of annual maxima between neighbouring sites is of logistic form, see §5*a*.

Assumption (i) is plausible because of the nature of surge progression (Davies & Flather 1978, although care is needed around estuarine regions where the linearity of the process is more suspect. For $\mathscr{G}(\mu_i, \sigma_i, k_i), i = 1, 2$, margins the density function of the logistic form chosen in (ii) is given by

$$f(x_1, x_2) = (\sigma_1 \sigma_2)^{-1} \tilde{x}_1^{(r-k_1)} \tilde{x}_2^{(r-k_2)} (\tilde{x}_1^r + \tilde{x}_2^r)^{(-2+1/r)} \{(\tilde{x}_1^r + \tilde{x}_2^r)^{1/r} + r - 1\} \exp\{-(\tilde{x}_1^r + \tilde{x}_2^r)^{1/r}\}, \tag{3.1}$$

where $r \geqslant 1$ and $\tilde{x}_i = [1 - k_i(x_i - \mu_i)/\sigma_i]^{1/k_i}$, for $i = 1, 2$. The choice of the logistic structure is based on its flexibility, allowing dependence to range from independence ($r = 1$), to complete dependence ($r \rightarrow \infty$). In addition, Coles & Tawn (1991) find a model incorporating assumptions (i) and (ii) to provide a good representation of extreme surge data from the east coast.

One complication in the estimation procedure is that in each year a collection of sites have missing data, the composition of the collection varying from year to year. As an approximation to the true joint distribution we take the 'neighbours' of a site, as required in assumption (ii), to vary yearly, being chosen as the closest geographical sites (one either side) with data in that particular year.

Violations of assumptions (i) and (ii) are unlikely to be critical here as primary interest is in the model for marginal distributions, not in the dependence structure itself. Thus the role of the model for spatial dependence is simply to account for dependence in the empirical realizations of the series to which the model is fitted.

In formulating the likelihood function for this model we use the generic notation $f(\cdot)$ to denote a (joint) probability density function, the appropriate random variables being apparent from the argument of f. Marginal and dependence parameters are denoted by $\boldsymbol{\theta}$ and $\boldsymbol{r}$ respectively. Now for a coastal stretch with ordered sites $\{1, \ldots, N\}$, let $I_j = \{j_1, \ldots, j_{N_j}\}, j = 1, \ldots, n$, be the ordered subset of N_j sites with data in year j. Then the full likelihood may be written as

$$\begin{aligned} L(\boldsymbol{\theta}, \boldsymbol{r}; z_{i,j} : j = 1, \ldots, n, i \in I_j) &= \prod_{j=1}^{n} f(z_{j_1}, \ldots, z_{jN_j}; \boldsymbol{\theta}, \boldsymbol{r}) \\ &= \prod_{j=1}^{n} f(z_{j_1}; \boldsymbol{\theta}) \prod_{l=2}^{N_j} f(z_{j_l} \mid z_{j_{l-1}}; \boldsymbol{\theta}, \boldsymbol{r}) \\ &= \prod_{j=1}^{n} f(z_{j_1}, z_{j_2}; \boldsymbol{\theta}, \boldsymbol{r}) \prod_{l=2}^{N_j - 1} \frac{f(z_{j_l}, z_{j_{l+1}}; \boldsymbol{\theta}, \boldsymbol{r})}{f(z_{j_l}; \boldsymbol{\theta})}. \end{aligned} \tag{3.2}$$

Although it would be preferable to estimate each of the logistic parameters $r_{j,k}$, between sites j and k, as part of the full likelihood (3.2), this is computationally unfeasible. Thus estimates of $r_{j,k}$, based on separate bivariate estimation, are substituted into (3.2). An alternative technique is to use estimates of the logistic parameters, $r_{j,k}$, based on the spatial dependence model proposed in §5*c*.

Table 2. *Information about various spatial models for the east-coast sites, 35–46*
(L_S is the maximized log likelihood for the spatial model: a and b denote $L_i(M)-L_i(S)$ and $z_{i,0.01}(M)-z_{i,0.01}(S)$ respectively.)

model...	C		D		E	
L_S	190.90		166.09		170.47	
(unknown parameters)	(36)		(14)		(16)	
i	a	b	a	b	a	b
35	1.28	0.52	1.06	0.45	1.26	0.49
36	0.25	−0.09	1.59	−0.32	1.62	−0.28
37	0.60	0.16	0.50	0.13	0.55	0.14
38	1.00	0.39	2.13	0.50	1.46	0.46
39	0.61	0.22	5.45	0.35	2.78	0.30
40	0.22	0.10	0.16	0.08	0.40	0.05
41	1.24	0.48	1.30	0.25	1.40	0.24
42	1.10	0.56	1.05	0.50	0.89	0.50
43	0.82	0.46	0.56	0.31	0.64	0.31
44	0.38	0.08	1.90	−0.27	2.11	−0.24
45	0.09	−0.01	2.38	−0.34	0.58	−0.16
46	0.46	0.10	0.93	−0.25	0.64	−0.04

(*e*) *Model and results*

In this section we describe a general covariate model for σ_i and k_i, the parameter $\mu_{i,j}$ having been discussed in §3*b*. Following Davison (1984), we take

$$\sigma_i = \exp\{\beta_0+\beta_1 d_i+\beta_2 d_i^2+\beta_3 d_i^*\},\quad k_i = \gamma_0+\gamma_1 d_i+\gamma_2 d_i^*,\quad i=1,\dots,N,$$

with higher-order polynomial terms being found unnecessary. Here d_i is the coastal distance of site i from a reference position and, for estuarine sites, d_i^* is the distance inland from the estuary mouth. In cases where no simple parametric model could be found a change point model is used.

The likelihood (3.2), evaluated with these marginal models, enables assessment of the spatial homogeneity–coherence of features in annual maxima data. This likelihood provides information about the global fit of the model, but not about the fit at particular data sites. A useful diagnostic tool for this purpose is the difference $L_i(M)-L_i(S)$ in the marginal log-likelihoods for each data site, i, evaluated respectively at the separate marginal maximum likelihood estimates, discussed in §3*a*, and at the maximum likelihood estimates for the site from the spatial model fit. The corresponding change $z_{i,p}(M)-z_{i,p}(S)$ in the estimates of return levels provides further diagnostic information.

The results for each coast are presented and discussed below, but, for presentation, we describe in detail only the model for the southern east coast (12 sites, from Immingham to Holland-on-Sea inclusive). Table 2 contains information on maximum likelihood estimation of various models for this coast. Since we restrict the likelihood (3.2) by fixing the dependence parameters, $\boldsymbol{r}$, the global maximum of the log-likelihood, L_S, is 190.90, occurring when maximizing with respect to 36 unknown parameters μ_i, σ_i and k_i unconstrained over sites, see model C in table 2.

The conventional annual maxima method, ignoring spatial dependence, is also equivalent to maximizing (3.2) with 36 marginal parameters but with $\boldsymbol{r}=\mathbf{1}$: this gives $L_S = 21.69$, model A. This change in likelihood is due to a combination of omitting the spatial dependence, and poor marginal parameter estimates from the

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Table 3. *Best fitting spatial models for the west- and south-coast sites*

(The parameters are as in (3.3) and the reference sites for distances are Fishguard in the 13–18 case, and Newlyn in the 19–27 case. a and b denote $L_i(M)-L_i(S)$ and $z_{i,0.01}(M)-z_{i,0.01}(S)$ respectively. The symbol $*$ denotes γ_0 has a change point model, $\gamma_0 = 0.331$ if $i = 13, 14, 15$, and $\gamma_0 = 0.148$ if $i = 16, 17, 18$.)

sites… best fitting spatial model and estimates	1–6 $\hat{\beta}_0 = -1.619$ $\hat{\gamma}_0 = -0.021$			7–12 $\hat{\beta}_0 = -1.586$ $\hat{\gamma}_0 = 0.178$ $\hat{\gamma}_2 = 0.040$			13–18 $\hat{\beta}_0 = -2.051$ $\hat{\beta}_1 = 0.040$ $\hat{\gamma}_0 = *$			19–27 $\hat{\beta}_0 = -2.241$ $\hat{\beta}_1 = 0.009$ $\hat{\gamma}_0 = 0.122$		
	i	a	b	i	a	b	i	a	b	i	a	b
	1	1.11	−0.41	7	0.99	−0.08	13	1.52	−0.06	19	1.30	−0.09
	2	1.54	−0.34	8	1.18	−0.20	14	0.10	0.03	20	0.22	−0.02
	3	0.41	0.23	9	1.72	−0.19	15	4.46	−0.21	21	1.99	−0.03
	4	2.07	0.02	10	0.31	0.09	16	0.45	−0.08	22	2.67	−0.19
	5	1.56	0.04	11	1.21	−0.08	17	0.20	−0.05	23	0.00	0.00
	6	0.85	0.33	12	0.02	0.02	18	0.82	−0.05	24	1.22	0.07
	—	—	—	—	—	—	—	—	—	25	2.76	−0.17
	—	—	—	—	—	—	—	—	—	26	1.83	−0.05
	—	—	—	—	—	—	—	—	—	27	0.81	0.11

separate annual maxima analyses. A fairer way to evaluate the marginal fit of the annual maxima approach is to calculate L_S using the marginal estimates from §3a, but with $\boldsymbol{r}$ fixed: this gives $L_S = 173.42$, model B. Therefore, even if all marginal parameters are taken to be unrelated from site to site, accounting for spatial dependence leads to significant improvements in marginal estimation due to spatial transfer of information. Further illustration of this is given by Coles (1990).

Homogeneity of σ_i and k_i along this coast, $\hat{\beta}_0 = -1.641$, $\hat{\gamma}_0 = -0.040$, model D, gives a significantly worse fit by comparison with model C when tested by standard likelihood methods, but no worse than by comparison with model B. A significantly better fitting model has $\ln \sigma_i$ taken as a quadratic in coastal distance from Immingham, $\hat{\beta}_0 = -1.800$, $\hat{\beta}_1 = 3.64 \times 10^{-3}$, $\hat{\beta}_2 = -1.19 \times 10^{-5}$, and $\hat{\gamma}_0 = -0.023$, model E. This model agrees remarkably well with known dynamical properties of this coastal stretch: gradual surge amplification during southward propagation, and a conflicting effect due to changes in tide–surge interaction (Wolf 1978).

Models C, D and E, which account for spatial dependence, yield return level estimates which, for almost all data sites, are less than those obtained in §3a: thus the effect of the influential 1953 extreme storm on east coast return level estimates, assessed spatially, is diminished. Consequently model E is as good a fit of the marginal data as the annual maxima method provides, yet reflects known physical properties. Although in terms of likelihood model E is significantly worse than model C, it is substantially more robust to erroneous data which are often evident in historical data. In addition, model E gives return level estimates which are similar to those from model C, as can be seen from table 2, but are also more precise with standard errors for $z_{i,0.01}$ ranging from 0.07 to 0.10 m over these sites, for comparison see table 1.

Tables 3 and 4 contain information about the best fitting spatial model for the west and south coasts, and the remaining east-coast sites respectively. Of most interest in these results is that a relatively simple spatial model provides such a good fit for each

Table 4. *Best fitting spatial models for the remaining east-coast sites*

(The parameters are as in (3.3) and the reference site for distances in 55–61 case is Aberdeen. a and b denote $L_i(M)-L_i(S)$ and $z_{i,0.01}(M)-z_{i,0.01}(S)$ respectively.)

sites ... best fitting spatial model and estimates	30–36 $\hat{\beta}_0=-1.736$ $\hat{\gamma}_0=-0.010$			49–56 $\hat{\beta}_0=-1.996$ $\hat{\beta}_3=0.010$ $\hat{\gamma}_0=-0.057$ $\hat{\gamma}_2=0.006$			55–61 $\hat{\beta}_0=-2.087$ $\hat{\beta}_1=0.058$ $\hat{\beta}_3=0.020$ $\hat{\gamma}_0=0.115$		
	i	a	b	i	a	b	i	a	b
	30	0.83	0.08	49	2.10	−0.09	55	0.01	−0.01
	31	0.47	1.48	50	0.14	−0.05	56	0.67	0.00
	32	0.61	0.02	51	0.32	−0.08	57	0.22	0.05
	33	0.19	−0.01	52	0.87	−0.14	58	0.26	−0.09
	34	0.27	0.20	53	0.23	−0.07	59	0.24	0.08
	35	0.37	0.01	54	3.07	−0.11	60	0.65	−0.06
	36	1.06	−0.35	55	0.34	0.05	61	2.12	0.82
				56	1.51	1.12			

data site, with the exception of Humber Dock and Swansea and to a lesser extent Calshot and Newhaven, whose distributions may be influenced by the Isle of Wight. For estuarine sites, as with the location parameter, the distance from the estuary mouth to the data site is important: with increasing distance inland k_i increases in the Mersey and Humber whereas σ_i increases in the Humber and the Firth of Forth. The separate models for the coastal stretches are almost continuous over model boundaries despite this not being a constraint when fitting the models. However, we recommend interpolation along the north and west coasts of Wales and the northern coast of Scotland be done with great care as there is limited data in these regions to validate the model.

4. The impact of climatic change

Much publicity has been given to the potentially catastrophic influence of the greenhouse effect on sea levels. Typically attention has been restricted to changes in mean sea level though changes that occur in the behaviour of extreme sea levels are of more importance in terms of coastal flooding. Lamb (1980) has indicated historical connections between coastal flooding and changes in both global temperature and the distribution of wind directions. In this section we examine the sensitivity of the results of §3 to various suggested consequences of the greenhouse effect and assess the resulting increased risk of flooding.

For simplicity we consider separately the influence on annual maxima return levels of changes in mean sea level, the number of extreme storm events and the severity of storm events. In practice a combination of such factors may be important, as may changes in tidal behaviour and tide–surge interaction induced by mean level change.

The following model relates the annual maxima distribution to the distributions of the number and magnitude of extreme events in a year. For a high threshold, u, greater than the largest tidal amplitude, suppose that in a year the sea level exceeds u on M independent occasions with the maxima from each independent exceedance being $Y_1, \ldots, Y_M$, where $Y_i > u$ for all i. The times of exceedances are taken to follow a Poisson process, so M follows a Poisson distribution, with mean λ say. Conditionally

Table 5. *Impact of change in mean sea level on existing sea defences of height $z_{0.01}$ (metres with respect to ODN)*

k	$z^*_{0.01} - z_{0.01} = u^* - u$	p^{*-1} ($\sigma = 0.1$)	($\sigma = 0.2$)	($\sigma = 0.3$)
−0.1	0.1	52.34	72.70	80.94
−0.1	0.3	12.70	37.30	52.34
−0.1	0.5	2.78	18.36	33.24
0.1	0.1	23.37	46.93	59.98
0.1	0.3	2.58	12.31	23.37
0.1	0.5	1.03	4.06	10.07
0.3	0.1	7.78	21.44	33.11
0.3	0.3	1.25	3.78	7.78
0.3	0.5	1.00	1.58	3.12

on M, the exceedances $Y_1 - u, Y_2 - u, \ldots, Y_M - u$ are taken to be independent and identically distributed with distribution function

$$H(y) = 1-(1-\beta y/\alpha)^{1/\beta}, \quad 1-\beta y/\alpha > 0,$$

where $\alpha > 0$, and β is arbitrary. For $\beta = 0$, taken as the limit as $\beta \to 0$, H is the exponential distribution with mean α, but more generally is the generalized Pareto distribution which has been shown, both theoretically and in practice, to provide a good model for exceedances of a high threshold (see Pickands 1975; Davison & Smith 1990 respectively). Thus, for $y > u$, the annual maxima distribution function, G, satisfies

$$\begin{aligned} G(y) &= \sum_{m=0}^{\infty} \Pr\{\max(Y_1, \ldots, Y_m) \leqslant y \,|\, M = m\} \Pr\{M = m\} \\ &= \sum_{m=0}^{\infty} [H(y-u)]^m \,\mathrm{e}^{-\lambda} \lambda^m / m! \\ &= \exp\{-\lambda[1-\beta(y-u)/\alpha]^{1/\beta}\}, \end{aligned} \tag{4.1}$$

which is a $\mathscr{G}(u+\alpha\beta^{-1}(1-\lambda^{-\beta}), \alpha\lambda^{-\beta}, \beta)$ distribution.

For the remainder of this section we use z^*_p to denote the updated return level, exceeded by the annual maxima with probability p, under specified changes to the model (4.1). The corresponding level of protection provided by defences designed to the present return level, z_p, we take to be p^*, where $z^*_{p^*} = z_p$: that is, p^* is the probability of the annual maximum exceeding z_p given the new conditions.

(*a*) *Changes in mean sea level*

If climatic change affects only mean sea level, model (4.1) will apply provided the height of the threshold u is raised to a level u^*, where $u^* - u$ is the change in mean sea level. Thus $z^*_p = z_p + u^* - u$ and $p^* = 1 - \exp(-\{[-\ln(1-p)]^k + k(u^*-u)/\sigma\}^{\frac{1}{k}})$.

(*b*) *Increase in the number of extreme storms*

A change in the occurrence rate of extreme storms, assuming the distribution of extreme storms is unchanged can be modelled by letting the mean number of exceedances of the threshold change from λ to $\lambda(1+\gamma)$. From (4.1) it follows that

$$z^*_p = z_p + \sigma k^{-1}[1-(1+\gamma)^{-k}][-\ln(1-p)]^k, \quad p^* = 1-(1-p)^{(1+\gamma)}.$$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Table 6. *Impact of an increase in the number of extreme storms on existing sea defences of height* $z_{0.01}$ *(metres with respect to ODN)*

		$z^*_{0.01}-z_{0.01}$			
k	γ	$(\sigma=0.1)$	$(\sigma=0.2)$	$(\sigma=0.3)$	p^{*-1}
−0.1	0.2	0.03	0.06	0.09	83.42
−0.1	0.4	0.05	0.11	0.16	71.57
−0.1	0.6	0.08	0.15	0.23	62.69
0.1	0.2	0.01	0.02	0.03	83.42
0.1	0.4	0.02	0.04	0.06	71.57
0.1	0.6	0.03	0.06	0.09	62.69
0.3	0.2	0.01	0.01	0.01	83.42
0.3	0.4	0.01	0.02	0.02	71.57
0.3	0.6	0.01	0.02	0.03	62.69

(c) *Increase in the severity of storms*

The key effects of an increase in the variability of storm surges are likely to be an increase in the number of extreme sea levels above a high threshold, covered in §4*b*, and an increase in the severity of each storm. The latter can be characterized by examining changes in the moments or quantiles of exceedances of the threshold. If $Y-u$ is a generalized Pareto random variable

$$E(Y-u)=\alpha(\beta+1)^{-1},\quad \operatorname{var}(Y-u)=\alpha^2(\beta+1)^{-2}(2\beta+1)^{-1}.$$

Changes in β are complex and not easily observed, but may be of considerable importance for extensive extrapolation. We examine only the case where α changes to $\alpha(1+\delta)$, i.e. all quantiles of the exceedance distribution are changed by the same multiplicative factor. From (4.1) it follows that

$$z^*_p=z_p+\delta(z_p-u),\quad p^*=1-\exp\left(-(1+\delta)^{-\frac{1}{k}}\{\delta k\sigma^{-1}(\mu-u)+\delta+[-\ln(1-p)]^k\}^{\frac{1}{k}}\right).$$

To overcome the difficulty that u is not an observed parameter with annual maxima data we take u as $z_{0.9}$.

(d) *Summary of climatic effects*

Tables 5, 6 and 7 illustrate, respectively, the effects of these three factors for a range of marginal parameters which are consistent with our estimates. The sensitivity of the results to each assumption is measured by the change in return level, $z^*_p-z_p$, and the resulting level of protection, p^{*-1}. For changes in mean sea level $z^*_p-z_p$ is invariant to σ and k, whereas p^{*-1} is a decreasing function of σ^{-1}, u^*-u, and k. For an increase in the number of extreme storms p^{*-1} decreases as γ increases but is invariant to σ and k, whereas $z^*_p-z_p$ is an increasing function of γ, σ, and $-k$. For an increase in the severity of storms p^{*-1} decreases with increasing δ, σ, or $-k$.

Accordingly, the south coast and eastern Scotland are most vulnerable to an increase in mean level, while the southern east coast, the Bristol Channel and western Scotland are most vulnerable to an increase in the severity of storms. All sites would be equally affected by an increased number of storms though the associated reduction in the level of protection is substantially less than for either of the other factors. Note that the change in characteristics of storms may differ from coast to coast.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Table 7. *Impact of an increase in the severity of extreme storms on existing sea defences of height* $z_{0.01}$ (*metres with respect to ODN*)

		$(z^*_{0.01}-z_{0.01}, p^{*-1})$		
k	δ	$(\sigma = 0.1)$	$(\sigma = 0.2)$	$(\sigma = 0.3)$
−0.1	0.2	(0.13, 5.19)	(0.27, 1.71)	(0.40, 1.01)
−0.1	0.4	(0.27, 1.02)	(0.53, 1.00)	(0.80, 1.00)
−0.1	0.6	(0.40, 1.00)	(0.80, 1.00)	(1.20, 1.00)
0.1	0.2	(0.09, 32.45)	(0.18, 3.79)	(0.27, 1.06)
0.1	0.4	(0.18, 15.76)	(0.36, 1.16)	(0.55, 1.00)
0.1	0.6	(0.27, 9.57)	(0.55, 1.00)	(0.82, 1.00)
0.3	0.2	(0.07, 69.32)	(0.14, 32.87)	(0.21, 18.01)
0.3	0.4	(0.14, 54.61)	(0.28, 17.92)	(0.41, 8.00)
0.3	0.6	(0.21, 46.17)	(0.41, 12.19)	(0.62, 4.97)

5. Spatial/multivariate dependence

When assessing the risk of flooding for a coastline the probability of no flooding at any point along the coast is the main concern; therefore knowledge of the marginal distributions of sea level extremes must be supplemented by information about spatial dependence. Detailed modelling of the spatial dependence requires the theory of max-stable processes (de Haan 1984) for which statistical methods are yet to be developed. Since our data sites are widely distributed, and the process is spatially coherent, we consider the almost equivalent problem of estimating the probability that there will be no flooding at a collection of data sites along the coastline. In §3 a simplistic model for inter-site dependence was proposed; here using results from multivariate extreme value theory we examine further the spatial distribution of extreme sea levels.

(*a*) *Multivariate distribution of annual maxima*

Multivariate extreme value theory concerns the limiting joint distribution of the vectors of componentwise linearly normalized maxima from weakly mixing stationary vector sequences. Both theoretical and statistical aspects of such distributions have received recent attention (Resnick 1987; Joe 1989; Smith *et al.* 1990; Tawn 1988*b*, 1990*b*). Based on asymptotic arguments similar to those justifying the use of the GEV distribution for univariate annual maxima we model the joint distribution of annual maxima at N sites using a multivariate extreme value distribution. In particular, if site i has marginal annual maxima distribution G_i, with $Z_i \sim \mathscr{G}(\mu_i, \sigma_i, k_i)$ for $i = 1, \dots, N$, then

$$\Pr\{Z_i \leqslant z_i : i = 1, \dots, N\} = \left[\prod_{i=1}^{N} G_i(z_i)\right]^{B(\boldsymbol{w})}, \tag{5.1}$$

where $\boldsymbol{w} = (w_1, \dots, w_N)$ with $w_i = \ln G_i(z_i)/\ln[\prod_{j=1}^{N} G_j(z_j)]$ for $i = 1, \dots, N$. The dependence function, $B(\boldsymbol{w})$, belongs to a subclass of convex functions satisfying $\max(w_1, \dots, w_N) \leqslant B(\boldsymbol{w}) \leqslant 1$ (see Pickands 1981; Tawn 1990*b*).

Suppose site i has coastal defences of height z_{i,p_i}, where $G_i(z_{i,p_i}) = 1-p_i$. Then by (5.1) the probability of no flooding, in a given year, at any of the N sites is $[\prod_{i=1}^{N}(1-p_i)]^{B(w)}$ with $w_j = \ln(1-p_j)/\sum_{i=1}^{N}\ln(1-p_i)$. Consequently where each site has equal protection, i.e. $p_j = p$ for all j, as is the criterion for Dutch sea defences

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

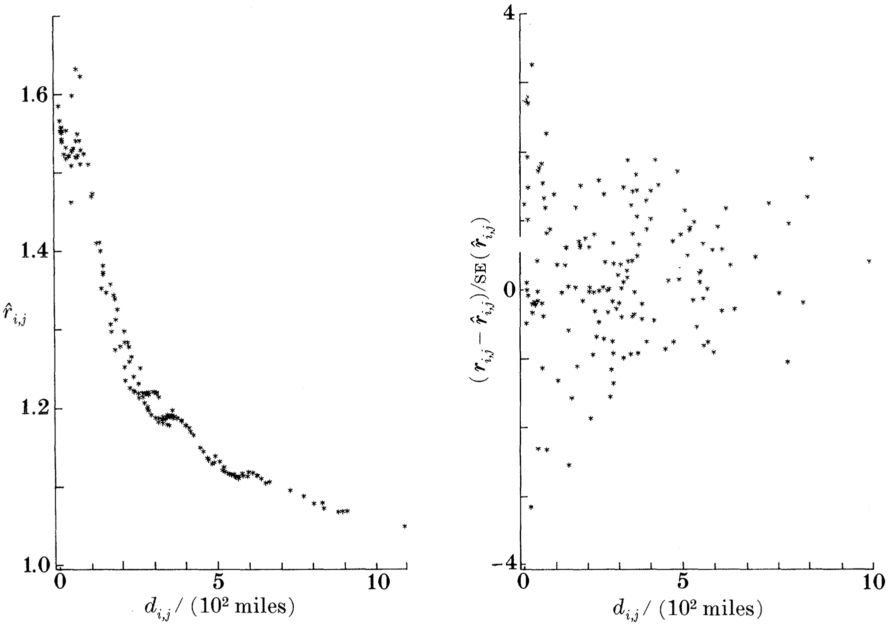


Figure 5 (left). Fitted values of logistic parameter $\hat{r}_{i,j}$, model (5.3), against distance between site pairs, for all west coast sites.

Figure 6 (right). Standardized residuals, $(r_{i,j}-\hat{r}_{i,j})/\text{SE}(\hat{r}_{i,j})$, of logistic parameter $r_{i,j}$, against distance between site pairs for west coast model (5.3).

Table 8. *Parameter estimates of the spatial dependence model* (*5.3*)

	α	$\beta_1 \times 10^3$	$\beta_2 \times 10^2$	ρ	γ
all sites	1.000	0.814	−0.107	−0.107	0.499
west coast	1.000	1.15	4.65	−0.0297	0.609
south coast	0.946	1.84	−8.91	0.0882	0.390
east coast	0.988	1.68	−1.04	0.341	0.851

(Dekkers & de Haan 1989), then (5.1) reduces to $(1-p)^{N_e}$. Here $N_e = NB(N^{-1}, \ldots, N^{-1})$, $1 \leqslant N_e \leqslant N$, which Reed & Dales (1988) call the effective number of independent sites. They also show that N/N_e is approximately the expected number of the N sites flooded, given flooding occurs for at least site. For British coastal defences p_j is chosen on the basis of a cost benefit analysis of flood prevention at the site; thus p_j is smallest in the locality of coastal nuclear power stations and towns, and largest where the area of potential flooding is arable land. Estimation of the dependence function can be achieved by using parametric or non-parametric methods (see Smith *et al.* 1990). For tractability, we use parametric methods in the bivariate case (Tawn 1988*b*) and non-parametric methods for higher dimensions (Pickands 1981). The set of all bivariate extreme value distributions cannot be finitely parametrized, thus subclasses of the complete family are used. As in §3 we use the logistic model with

$$B(w, 1-w) = [w^r + (1-w)^r]^{1/r}, \quad r \geqslant 1. \tag{5.2}$$

An associated dependence measure is given by $2[1-B(\frac{1}{2},\frac{1}{2})] = 2-2^{1/r}$ corresponding to the limit as $p \to 0$ of $\Pr\{Z_i > z_{i,p} \mid Z_j > z_{j,p}\}$ with r depending on i and j, see Tawn (1988*b*).

(*b*) *Factors influencing spatial dependence*

Spatial dependence of extreme sea levels is due to the propagation of storm surges along coastlines, and the progression of the meteorological conditions which generate storm surges (Heaps 1967, 1983). Therefore if there were no tidal component of sea level the key factor influencing inter-site dependence would be coastal distance between sites, with statistics of the depression tracks which generate extreme surges as a less important covariate (Lennon 1963*b*). However, since extreme sea levels arise from extreme surges occurring near high tide the timing of the extreme surge at a site is vital. Consequently, the time for surge progression, the duration of the extreme surges at each site, and the spatial pattern of tidal phases are also important factors. The first two of these are identifiable only by detailed analysis of hourly observations at each site, which is beyond the scope of this preliminary analysis. On the other hand, the tidal phase can be well approximated by the phase, ϕ_{M_2}, of the semi-diurnal lunar constituent, which is given in the Admiralty tide tables (1990) and plotted spatially by Flather (1987).

(*c*) *Model, estimation and results*

Since the set of all bivariate dependencies provides useful information about the full dependence structure, we use this summary information to determine how the factors discussed in §5*b* affect inter-site dependence for extreme sea levels. To examine this we model the bivariate logistic dependence parameter between sites i and j, $r_{i,j}$, in terms of these factors. Motivated partly by the form of the correlation coefficient $(1-r^{-2})$ for the logistic model when the margins are Gumbel, also by the models for spatial correlation proposed in the discussion of Haslett & Raftery (1989), and by the properties of the surge and tide processes, we take

$$r_{i,j}(\alpha,\beta_1,\beta_2,\rho,\gamma) = \{1-\alpha\exp[-\beta_1 d_{i,j}-\beta_2\cos(\phi_{i,M_2}-\phi_{j,M_2}+d_{i,j}\rho)]\}^{-\gamma}, \quad (5.3)$$

with $0<\alpha\leqslant 1$. Here $d_{i,j}$ is the coastal distance between sites i and j and ϕ_{i,M_2} is the phase of M_2 at site i. The purpose of the cosine term is to account for surge progression between sites, $d_{i,j}\rho$ representing the time for progression, and $\phi_{i,M_2}-\phi_{j,M_2}$ the time lag between high tide at the two sites. With hourly surge data the estimation of the $d_{i,j}\rho$ term would be much improved. The form of (5.3) corresponds to correlation between Gumbel margins when $\gamma=\frac{1}{2}$, however we found the extra flexibility introduced by γ to be necessary.

Our method of estimation is motivated by unpublished work by R. L. Smith. For all possible pairs of data sites, with data from overlapping periods, we fitted the joint distribution (5.1) with logistic model for the dependence structure by maximum likelihood, in each case obtaining $\hat{r}_{i,j}$ and the variance of $\hat{r}_{i,j}$, in the form of the observed Fisher information (Tawn 1988*b*). Model (5.3) was fitted using weighted least squares, minimizing

$$\sum_{i,j}(\hat{r}_{i,j}-r_{i,j})^2/\operatorname{var}(\hat{r}_{i,j}).$$

This procedure was done for east-, west- and south-coast pairs of sites separately, and for all pairs of sites, to examine regional variations due to different types of storm progression. For each case parameter estimates are given in table 8. Figures 5 and 6 show, respectively, the fitted values and standardized residuals from model (5.3) for sites on the west coast. Fluctuations around the general trend in figure 5, especially at short distances, are indicative of the magnitude of the tidal phase component in

MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES
THE ROYAL SOCIETY A
PHILOSOPHICAL TRANSACTIONS OF

the model. In figure 6 the increased variability at short distances suggests that local topography and variations in surge progression may be a dominant influence on dependence for short distances.

In addition to the local dependence, which decays with distance along the coastline, some distantly separated coastal stretches exhibit significantly stronger dependence than is given by the fitted model (5.3). This is because we use a distance metric that does not account for meteorological progression, and so in some cases is inadequate at long distances. For example, the west coast, from Silloth to Avonmouth, is dependent with both the south coast, from Calshot to Pevensey Bay, and the Firth of Forth regions, whereas the south coast, from Devonport to Portsmouth, is dependent with the Humber and Wash regions.

To estimate joint probabilities, such as (5.1) with $N > 2$, we use an unbiased form of Pickands' (1981) estimator:

$$\hat{B}(\boldsymbol{w}) = (n_N - 1)\left\{\sum_{j=1}^{n_N} \min_{i=1,\dots,N} (\tilde{x}_{i,j}/w_i)\right\}^{-1}, \quad w_1 + \dots + w_N = 1 \tag{5.4}$$

with variance $[B(\boldsymbol{w})]^2/(n_N-2)$. Here n_N is the number of years with data for each of the N sites and $\tilde{x}_{i,j} = [1-k_i(z_{i,j}-\mu_{i,j})/\sigma_i]^{1/k_i}$. When estimating (5.1) with $z_i = z_{i,p}$ for all i, i.e. a common protection level is given to all data sites, the problem reduces to evaluating the effective number of independent sites, N_e. For the coastal stretches examined in §3*e* $\hat{N}_e$ and (standard error (SE) of $\hat{N}_e$) are 2.11 (0.94) for sites 1–6, 3.07 (0.82) for 7–12, 1.45 (0.55) for 13–18, 4.48 (1.58) for 19–27, 1.93 (0.79) for 28–34, 2.96 (1.21) for 35–46, 1.93 (1.36) for 47–54 and 2.58 (0.86) for 55–61, using (5.4). These estimates illustrate the strong spatial dependence along coastlines. The relatively large standard errors arise because only simultaneous data from each site can be used in evaluating this estimator, however, omitting sites with limited data is found only to reduce the standard errors without substantially changing estimates.

To illustrate the case where more protection is given to some sites than others consider the sites 30–33 in the Thames estuary. More protection is required nearer London so take $p_{30} = p_{33} = 0.01$ and $p_{31} = p_{32} = 0.001$ giving (5.1) as 0.9851 with SE 0.0025; whereas for $p_j = \bar{p} = 0.0055$, $j = 30, \dots, 33$ gives (5.1) as 0.9902 with SE 0.0016. As the two estimates evaluate (5.4) at different points it is clear that in general the complete multivariate dependence structure is important for the optimal design of coastline flood prevention schemes.

6. Discussion and conclusions

Throughout the paper discussion has been restricted to annual maxima of sea levels after waves have been averaged out. We have simultaneously modelled the dependence of the parameters of the distribution of such data on the positions and tidal characteristics of the sites, accounting for the influence of local coastal topography, spatial dependence and varying data spans. Our key findings are that the topographical features which determine the tidal parameters also influence the marginal distribution of surges on open coastlines, with more complex relationships in estuarine regions; that some aspects of the model are locally influenced whereas others are spatially coherent; and that some aspects of the model are more sensitive to climatic change than others.

However, there is considerable scope to extend this preliminary analysis using extensive hourly sea level, tide and surge data, since these contain much information about the physical process of extremes which is lost by analysing only annual maxima data, see §5*b*. The additional data on extreme events can also be used to improve estimation of return levels (Tawn 1990*a*). A further aspect of the statistical analysis for sea defence design is the incorporation of extreme wave data. Only very limited studies of wave data, mainly obtained from oceanographic numerical models, and their dependence with extreme surge data have been done. We see this to be a valuable and fruitful area for future use of statistical arguments in oceanography.

We thank Dr C. W. Anderson, Dr R. A. Flather and Dr J. M. Vassie for many helpful discussions. We are also indebted to Mr J. Graff without whose considerable effort in collecting the extensive historical data these analyses would not have been possible. S.G.C. was supported by a Science and Engineering Research Council studentship and J.A.T. was partly supported by a Nuffield Foundation grant.

References

Admiralty 1990 *Tide tables*, vol. 1. Taunton: Hydrographic Department.

Amin, M. 1982 On analysis and forecasting of surges on the west coast of Great Britain. *Geophys. Jl R. astr. Soc.* **68**, 79–94.

Buishand, T. A. 1984 Bivariate extreme-value data and the station year method. *J. Hydrol.* **69**, 77–95.

Buishand, T. A. 1989 Statistics of extremes in climatology. *Statistica Neerlandica* **43**, 1–30.

Coles, S. G. 1990 Discussion of paper by A. C. Davison and R. L. Smith. *Jl R. statist. Soc.* B **52**. (In the press.)

Coles, S. G. & Tawn, J. A. 1991 Modelling extreme multivariate events. *Jl R. statist. Soc.* B **53**. (In the press.)

Davies, A. M. & Flather, R. A. 1978 Application of numerical models of the north-west European continental shelf and the North Sea to the computation of the storm surges of November to December 1973. *Dt. hydrogr. Z. Erganzungshelf* A **14**. (72 pages.)

Davison, A. C. 1984 Modelling excesses over high thresholds, with an application. In *Statistical extremes and applications* (ed. J. Tiago de Oliveira), pp. 131–153. Dordrecht: Reidel.

Davison, A. C. & Smith, R. L. 1990 Models for exceedances over high thresholds (with discussion). *Jl R. statist. Soc.* B **52**. (In the press.)

Dekkers, A. L. M. & Haan, L. De 1989 On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.* **17**, 1795–1832.

Fitzgerald, D. L. 1989 Single station and regional analysis of daily rainfall extremes. *Stochastic Hydrol. Hydraul.* **3**, 281–291.

Flather, R. A. 1987 Estimates of extreme conditions of tide and surge using a numerical model of the north-west European continental shelf. *Estuarine Coastal Shelf Sci.* **24**, 69–93.

George, K. J. & Thomas, D. K. 1976 Two notable storm surges of the south coast of England. *Hydrographic J.* **2**, 13–16.

Graff, J. 1981 An investigation of the frequency distributions of annual sea level maxima at ports around Great Britain. *Estuarine Coastal Shelf Sci.* **12**, 389–449.

de Haan, L. 1984 A spectral representation for max-stable processes. *Ann. Probab.* **12**, 1194–1204.

Haslett, J. & Raftery, A. E. 1989 Space-time modelling with long memory dependence: assessing Ireland's wind power resource. *Appl. Statist.* **38**, 1–50.

Heaps, N. S. 1967 Storm surges. In *Oceanogr. Mar. Biol. A. Rev.* (ed. H. Barnes), vol. 5, pp. 11–47. London: Allen and Unwin.

Heaps, N. S. 1983 Storm surges, 1967–1982. *Geophys. Jl R. astr. Soc.* **74**, 331–376.

Jenkinson, A. F. 1977 The analysis of meteorological and other geophysical extremes. *Met 013 Tech. Note* 58. Bracknell: Meteorological Office.

Joe, H. 1989 Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statist. Probab. Lett.* **9**, 75–82.

Johnson, D. 1929 Studies of mean sea level. *Bull. National Res. Council*, no. 70. Washington, D.C.: National Academy of Sciences.

Lamb, H. H. 1980 Climate fluctuations in historical times and their connexion with transgressions of the sea, storm floods and other coastal changes. In *Transgressies en Occupatiegeschiedenis in de kustgebieden van Nederland en Belgie* (ed. A. Verhulst & M. K. E. Gottschalk), pp. 251–290. Belgissches Centrum voor Landelijke Geschiedenis.

Leadbetter, M. R., Lindgren, G. & Rootzen, H. 1983 *Extremes and related properties of random sequences and series*. New York: Springer-Verlag.

Lennon, G. W. 1963*a* A frequency investigation of abnormally high tidal levels at certain west coast ports. *Proc. Inst. Civ. Engrs.* **25**, 451–484.

Lennon, G. W. 1963*b* The identification of weather conditions associated with the generation of major storm surges along the west coast of the British Isles. *Q. Jl R. met. Soc.* **89**, 381–394.

Moore, R. J. 1987 Combined regional flood frequency analysis and regression on catchment characteristics by maximum likelihood estimation. In *Regional flood frequency analysis* (ed. V. P. Singh), pp. 119–131. Dordrecht: Reidel.

NERC 1975 *The flood studies report*. London: The National Environment Research Council.

Pickands, J. 1975 Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119–131.

Pickands, J. 1981 Multivariate extreme value distributions. *Bull. Int. statist. Inst.*, 859–878.

Prandle, D. & Wolf, J. 1978 The interaction of surge and tide in the North Sea and River Thames. *Geophys. Jl R. astr. Soc.* **55**, 203–216.

Proctor, R. & Flather, R. A. 1989 Storm surge prediction in the Bristol Channel – the floods of 13 December 1981. *Continental Shelf Res.* **9**, 889–918.

Pugh, D. T. 1987 *Tides, surges and mean sea level*. Chichester: Wiley.

Reed, D. W. & Dales, M. Y. 1988 Regional rainfall risk: a study of spatial dependence. In *Proc. IAHR Symp.* Birmingham.

Resnick, S. I. 1987 *Extreme values, regular variation, and point processes*. New York: Springer-Verlag.

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

Smith, R. L., Tawn, J. A. & Yuen, H. K. 1990 Statistics of multivariate extremes. *Int. statist. Rev.* **58**, 47–58.

Suthons, C. T. 1963 Frequency of occurrence of abnormally high sea levels on the east and south coasts of England. *Proc. Inst. Civ. Engrs* **25**, 433–450.

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

Tawn, J. A. 1988*b* Bivariate extreme value theory: Models and estimation. *Biometrika* **75**, 397–415.

Tawn, J. A. 1990*a* Estimating probabilities of extreme sea levels. *Appl. Statist.* (Submitted.)

Tawn, J. A. 1990*b* Modelling multivariate extreme value distributions. *Biometrika* **77**, 245–253.

Tawn, J. A. & Vassie, J. M. 1989 Extreme sea levels: the joint probabilities method revisited and revised. *Proc. Inst. Civ. Engrs* **87**, 429–442.

Wolf, J. 1978 Interaction of tide and surge in a semi-infinite uniform channel, with application to surge propagation down the east coast of Britain. *Appl. math. Model* **2**, 245–253.

Woodworth, P. L. 1987 Trends in U.K. mean sea level. *Mar. Geodesy* **11**, 57–87.