ESTIMATES OF EXTREME SEA CONDITIONS

Final Report

EXTREME SEA-LEVELS AT THE UK A-CLASS SITES: OPTIMAL SITE-BY-SITE ANALYSES AND SPATIAL ANALYSES FOR THE EAST COAST.

Mark J. Dixon and Jonathan A. Tawn

Department of Mathematics and Statistics,

Lancaster University,

Lancaster LA1 4YF.

In collaboration with

The Proudman Oceanographic Laboratory,

Bidston Observatory,

Birkenhead,

Merseyside L43 7RA.

August 1995

CONTRACT

This report describes work funded by the Ministry of Agriculture, Fisheries and Food (Flood and Coastal Defence Division) under Commission FD 0303 with the NERC Proudman Oceanographic Laboratory (POL). POL's Nominated Officer was Mr G Alcock. The Ministry's Project Officer was Mr A C Polson. Publication does not imply endorsement by the Ministry of the report's conclusions or recommendations.

Summary

In this report the second stage of a three phase analysis of extreme sea-levels around the UK is described. This stage concerns the extension of the methods for the statistical analysis of extreme sea-levels to incorporate historical and spatial data, and their subsequent application to the UK sea-level data. Design level estimates are given for data sites on the east, west and south coasts, and at a regular grid along the east coast.

Topics covered

- A substantial extension to the Revised Joint Probability Method which includes
  - use of all types of available data from the site,
  - spatially interpretable parameters,
  - trends consistent with those observed in mean sea-levels.

  We term this method the Spatial Revised Joint Probability Method.
- Obtaining design estimates, in the form of return levels and trend estimates, for all 41 sites in the study based on data from the site of interest alone. For each site results for the best method of analysis are given.
- Case-studies for the spatial analysis of sea-level processes, such as:
  - extreme sea-level trends;
  - mean sea-level trends;
  - spatial modelling of the joint distribution of largest annual event data
  - the spatial variability and form of interaction between tides and large surges.
- The development of a spatial model for extreme sea-levels based on the Spatial Revised Joint Probabilities Method.
- The application of the spatial model to the UK east coast giving return level estimates for sites on a regular grid.

- Comparisons of the estimates from the spatial model with the best estimates from the analysis of data from the individual sites.

Findings and Conclusions

The conclusions concerning which method of analysis is most suitable to use for a particular site depend on whether the information available for the site is of the form of

1. data only for the site,
2. data for the site and from a spatial analysis for the associated coastline,
3. only a spatial analysis for the associated coastline,

The conclusions for each site are separated into these three cases:

1. When the only information available comes from data for the site, i.e. the case for sites on the south and west coasts in this report, then:

- The Spatial Revised Joint Probability Method is the best approach provided at least 5-10 years of high quality hourly data are available. If less data are available then the Joint Probabilities Method is the best approach.
- If there are less than 5-10 years of hourly data but there is also a long record of additional years of historical annual maximum data available at the site the above conclusion changes to use the Spatial Revised Joint Probability Method.
- If there are historical mean sea-level data available for the site that does not influence the selection of approach for estimating return levels.
- The trend is best estimated for the site using the Spatial Revised Joint Probability Method as all relevant data are pooled together. For sites with very short hourly records and no historical data information about the trend cannot be obtained.

2. When the information comes from both the site and a spatial model, i.e. the case for A-class and study sites on the east coast in this report, then:

- The spatial model return level estimates should be used provided the interpolated tidal series for the site given by the Proudman Oceanographic Laboratory's software agrees well with the tidal prediction obtained from data recorded at the site. For the east coast this is the case for all A-class sites and all sites on the open coast. When the interpolated tide is considered poor a special run of the spatial model is required with the tidal series based on predictions for the site replacing the spatially interpolated tidal series in the statistical analysis.

- The trend should be estimated using the spatial model. One situation in which the advice may differ is if the site has a long historical record of mean sea-levels or annual maximum (i.e. 40 years or more) and has a trend which is inconsistent with the spatial estimate due to neighbouring sites having different trends. In this case the trend may have a local effect which is detected by the site estimate but smoothed out by the spatial estimate, so then the site estimate should be used.

3. When the information comes from only a spatial analysis for the associated coastline, i.e. in this report these are sites on the east coast which are not study sites, then:

- The only source of return level estimates comes from the spatial model.
- The spatial estimate is relative to mean sea-level at the site but a spatial conversion factor to ODN is supplied.
- The spatial estimate is not calibrated at the site so it must be used with some caution. Ideally data or historical information for the site should be used to quantify the spatial estimate for the site.

Use of the report

The main results of the analysis are the best site-by-site estimates given in Part II, and the spatial estimate of return levels and trends along the east coast, given in Part IV. If interest is only in final estimates then only these components of the report need to be studied. A deeper understanding of the site-by-site methods is provided in Part I and the report of the first stage of the study (Dixon and Tawn, 1994). An appreciation of why we take the statistical approaches adopted in Part IV for the spatial model is given in Part III.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This report describes the second of three stages of the MAFF funded project *Estimates of Extreme Sea Conditions*. The broad aims of the overall study are to produce improved statistical methods for the analysis of extreme sea-levels and to systematically apply these to estimate design levels for all coastal sites using the method of analysis which best exploits all the available information. The project is broken down into the three stages as follows:

Stage 1:

The development, systematic application, and inter-comparison of three methods of extreme sea-level analysis: the r-largest, the Joint Probabilities Method, and the Revised Joint Probabilities Methods. Each of these methods uses only data from the site of interest so could only be applied for sites with long records. Application was restricted to periods of hourly data from the 22 sites:

Wick, Aberdeen, North Shields, Whitby, Immingham, Cromer, Lowestoft, Felixstowe, Southend, Sheerness, Dover, Newlyn, Ilfracombe, Avonmouth, Milford Haven, Fishguard, Holyhead, Heysham, Portpatrick, Ullapool, Stornoway and Lerwick.

Stage 2:

The extension of the Revised Joint Probabilities Method to incorporate all relevant types of data, such as historical annual maximum still water levels and spatial data from neighbouring sites. This method is applied to data from sites on the UK south and west coasts to give design level estimates on a site-by-site basis, and over a fine regular spatial grid along the UK east coast.

Data sites which are analysed here but were excluded from Stage 1 of the study are:

Leith, Harwich, Walton, Newhaven, Portsmouth, Weymouth, Devonport, Hinkley, Newport, Swansea, Mumbles, Barmouth, Liverpool, Port Erin, Workington, Millport, Islay, Tobermory and Kinlochbervie.

Stage 3:

The spatial methods developed in the second stage of the project will be extended to incorporate the additional information contained in a continuous 39 year run of a hydrodynamical model for tides and surges over the European continental shelf. This method will be used to provide design level estimates over a fine regular grid for the entire mainland UK coastline.

This current report will generally be self contained, although some reference to the report of the first stage of the project *Extreme Sea-Levels At The UK A-Class Sites: Site-By-Site Analyses*, referenced as Dixon and Tawn (1994), will inevitably be made. A brief summary of the objectives and findings of Stage 1 are given in Section 1.1, and an overview of this report is given in Section 1.2.

1.1 Summary of the first stage of the project

The objectives of the first stage of the work were:

- To describe in detail the existing statistical methodology for the site-by-site techniques and to compare these methods. These were given in Sections 4 and 5.

- Based on the wide experience gained through systematic application of the existing methods to a range of data sites, where the sea-level processes exhibit a variety of physical characteristics, to refine these methods so that they have better properties and more extensive applicability. These refinements were described in Section 6.

- To obtain the first systematically derived set of return level estimates, with associated measures of precision, at each A-class site for which site-by-site methods give reliable estimates. Such results, displayed in both graphical and tabular form, were given for each site in Sections 7 and 8.

- To develop substantive new methodology for the analysis of extreme sea-levels in the form of Joint Probability Methods for still water levels and waves. These were detailed in Section 9 of the report.

The principal findings obtained in that study were:

- At sites where the surge has a large variability at high tidal levels relative to the variability between high tides, then the r-largest, Joint Probability Method and Revised Joint Probability Methods each give broadly similar results. Generally, this is the case for sites in the South-East of the UK.

- At sites where the variability of high tides is large relative to the variability of the surge at high tidal levels, then both the annual maxima and r-largest methods under-estimate return levels, that is the methods over-estimate the return period of observed levels. Generally this is the case for sites on the UK west coast. The degree of under-estimation is reduced as the length of the annual maxima record is increased, but in some cases data from many nodal cycles of the tide are required before the bias is adequately reduced.

- Trends in extreme sea-level data were estimated, but the estimates are highly variable owing to the short period of hourly observation, and the use of only extreme value data in the estimation of trends.

- The Joint Probability and Revised Joint Probability Methods provide accurate estimates from much shorter series than the r-largest method.

- Of the Joint Probability and Revised Joint Probability Methods, the latter is the better provided there are at least 5-10 years of hourly observation.

In general the recommendation was that the best estimates of return levels, based on the site-by-site methods, are given by the Revised Joint Probability Method provided that sufficient high quality hourly data are available.

1.2 Outline of the report for Stage 2

Data from 41 sites are considered in this stage of the project. Dixon and Tawn (1994) found that the three site-by-site methods only produced reliable statistical results for 22 of these 41 sites. By better handling of trends in the statistical analysis, use of historical annual data, and the exploitation of spatial coherence of the extreme sea-level process through a spatial statistical analysis, results for the other 19 sites are obtained and improvements are made to estimates for the original 22 sites. Furthermore, by use of the spatial model developed for this report, estimates of extreme still water levels are given for sites at regular intervals along the UK east coast for which data are non-existent.

This work is achieved in four steps, related to the four parts of the report:

Part I:

- A substantial extension of the Revised Joint Probability Method as used by Dixon and Tawn (1994) to include all types of relevant available data from the site and such that the parameters of the model have a clear spatial interpretation. The new method is called the Spatial Revised Joint Probability Method.

Part II:

- Obtaining design estimates, in the form of return levels and trend estimates for all 41 sites in the study based on data from the site of interest only, including the 19 sites which were not studied by Dixon and Tawn (1994).
- For each site the results from best method of analysis are given.

Part III:

- The development of a spatial model for extreme sea-levels.
- Many case-studies are given to aid the development of the spatial extension of the model in Part I.

Part IV:

- The application of the spatial model along the UK east coast.
- Return level and trend estimates are given for a regular grid along the east coast.
- For A-class data sites on the east coast, comparisons of the spatial estimates are made with the best estimates from the analysis of data for the individual sites.

1.3 Use of estimates in this report

The estimates in this report have been produced in a systematic and consistent fashion for each site using the best currently available statistical methods. The methods are only able to extract information from the available data, so that even though a large data-base of sea-level elevations augmented with historical data is used, the resulting return level and trend estimates may be inaccurate or biased if the available data is an unrepresentative sample. In addition, the spatial method only provides improved estimates over marginal analyses if our assumption of spatial coherence is valid at the site under consideration. In most cases the data are representative, and the coherence assumption is valid. However, careful analysis carried out by an experienced coastal engineer who is able to use additional knowledge of the site which is not incorporated in the data-base used here will provide improved estimates at a particular site. Thus if such extra analyses are available, the recommendation is to use the estimates here only as a guide to obtain best estimates at a site.

Chapter 2

The UK sea-level data

2.1 Data availability and location of the sites

Data are available for all the A-class sites in the form of hourly measurements of still water level, which is defined as the observed sea-level after surface waves have been averaged out. These data are measured relative to Admiralty Chart Datum (ACD), which for the mainland sites can be converted to a common datum, Ordnance Datum Newlyn (ODN). Through the British Oceanographic Data Centre (BODC), POL hold these data for each of the sites now classified as A-class. Figure 2.1 shows a map of the UK with positions of 38 A-class sites indicated. Dixon and Tawn (1994) give a full description of the data from 22 of the 41 sites in this study. Those data sites exhibited the complete range of tidal and surge behaviour around the UK coastline. As the remaining 19 sites have similar tide and surge characteristics, the descriptive tide and surge density plots of Dixon and Tawn (1994) are not shown for the 19 new sites.

Of the 38 sites shown in Figure 2.1:

- No suitable data were available for Llandudno.
- Four sites not in the A-class network are included in the study; they are Harwich, Walton, Southend and Swansea. These sites are included to aid the assessment of the statistical procedures in regions of particular concern.
- There are 37 sites from the A-class network which have suitable data for at least some level of analysis.

In addition to hourly data, many of the sites in the study have data on historical annual maximum and/or annual mean sea-levels. Tables 2.1-2.2 show the availability of the hourly data and these additional data for all study sites, and Figure 2.2 gives a graphical display of which years have hourly data for these sites.

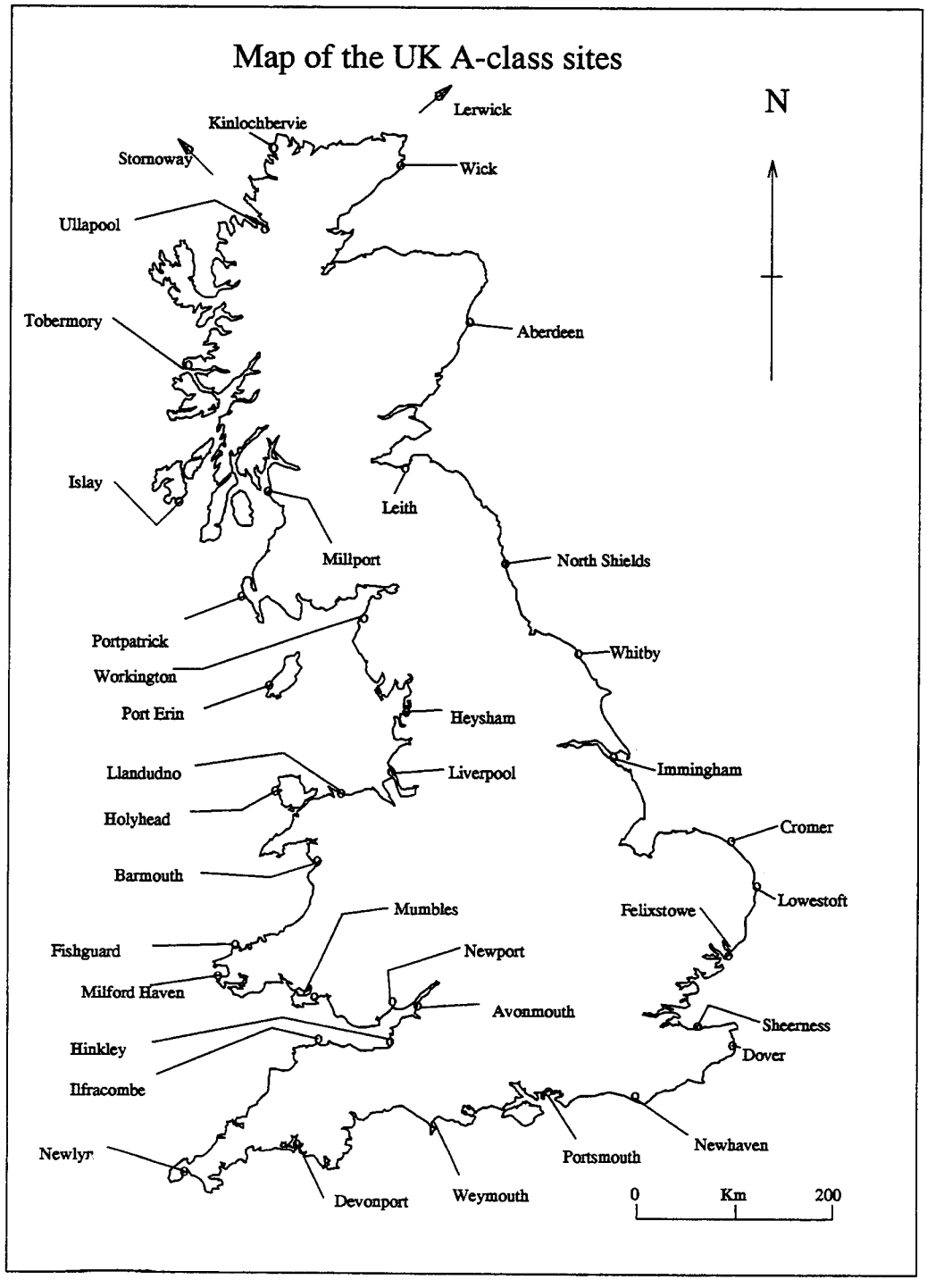

Figure 2.1: Map showing the locations of all of the UK A-class gauges.

Site	Hourly Data			Annual maxima			Mean sea-levels		
	Number	From	To	Number	From	To	Number	From	To
Wick	25	1965	1991	NA	NA	NA	24	1965	1991
Aberdeen	24	1946	1991	67	1908	1975	54	1932	1991
Leith	3	1989	1992	38	1939	1978	13	1956	1971
North Shields	22	1962	1991	35	1883	1978	88	1896	1991
Whitby	7	1981	1991	NA	NA	NA	8	1981	1991
Immingham	27	1964	1990	69	1920	1988	31	1960	1991
Cromer	4	1989	1992	NA	NA	NA	NA	NA	NA
Lowestoft	21	1970	1990	31	1953	1983	33	1956	1991
Felixstowe	4	1989	1992	NA	NA	NA	7	1981	1991
Harwich	16	1954	1975	51	1926	1976	13	1960	1972
Walton on the Naze	11	1967	1977	19	1968	1986	9	1969	1977
Southend	30	1929	1983	57	1929	1986	48	1933	1983
Sheerness	18	1965	1991	136	1819	1983	65	1834	1991
Dover	31	1958	1992	62	1912	1984	27	1961	1991

Table 2.1: Available data at the east coast study sites. NA denotes no data are available.

In certain case studies, in Chapters 7-10 of the report, other sites are considered, which although they are not in the 41 listed sites, have data on annual mean sea-levels and/or annual maximum sea-levels. These sites are identified in the relevant sections.

Throughout the analyses, the data pre-processing as described in Dixon and Tawn (1994, Section 2.4) is applied to all sites, except that the sites with short records are not removed from this study.

2.2 The sea-level process

For any site the observed hourly sea-level, at time t, Z_t, after averaging out surface waves, consists of three unobserved additive components: mean sea-level, Z_{0_t}, tidal level, X_t, and surge level, Y_t, so that

$$Z_t = Z_{0_t} + X_t + Y_t. \tag{2.2.1}$$

Although unobserved, these components can be estimated from long series of data.

The Mean Sea-level

For a particular year, the annual mean sea-level is usually obtained by averaging the hourly still water level observations from within the year (Pugh, 1987). However due to long period tides and non-zero mean levels for surges on an annual scale, this annual

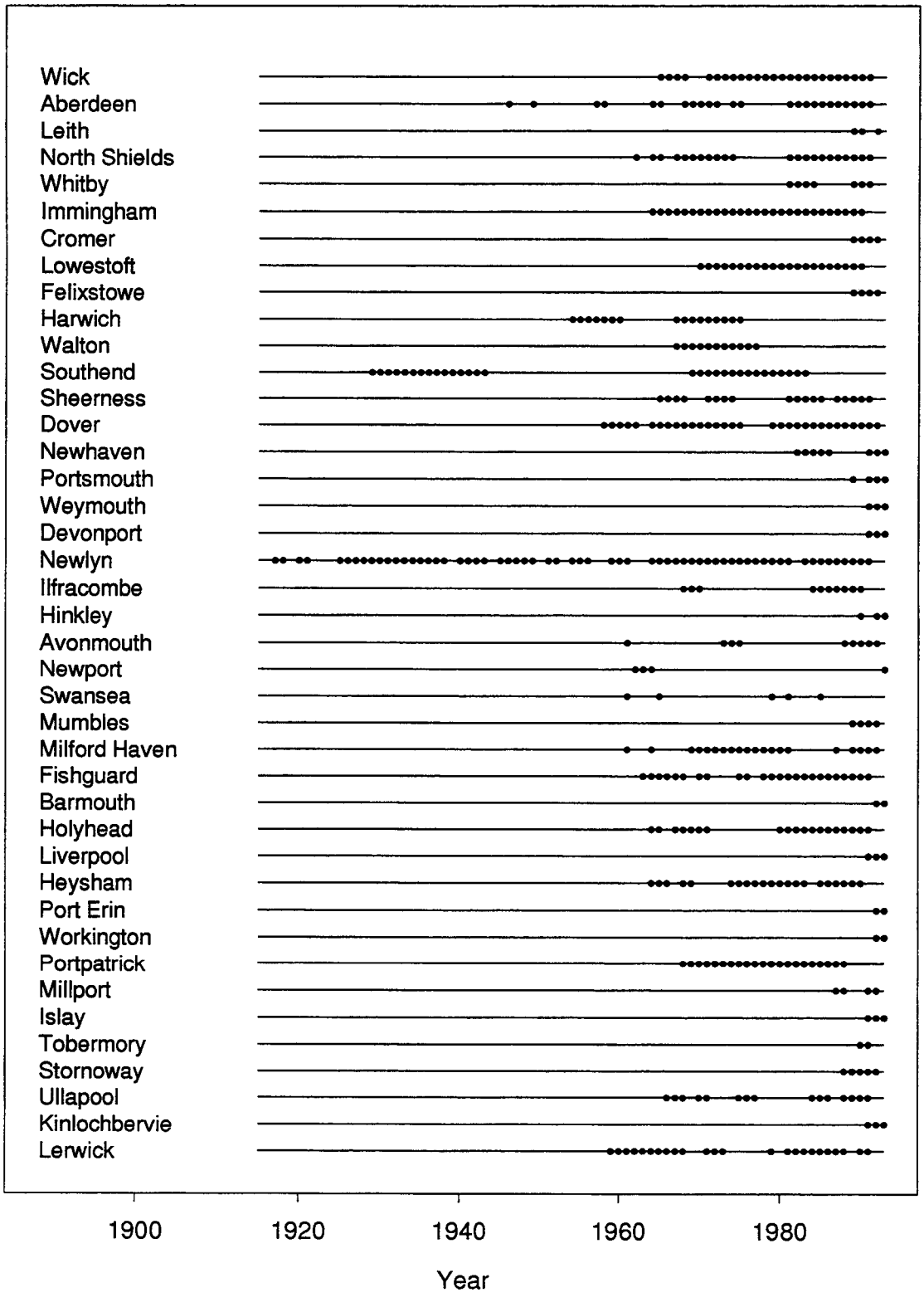

Figure 2.2: Information on the spans of available hourly data from each of the 41 sites in this study.

Site	Hourly Data			Annual maxima			Mean sea-levels		
	Number	From	To	Number	From	To	Number	From	To
Newhaven	8	1982	1993	60	1913	1976	NA	NA	NA
Portsmouth	4	1989	1993	104	1813	1975	NA	NA	NA
Weymouth	3	1991	1993	NA	NA	NA	NA	NA	NA
Devonport	3	1991	1993	38	1920	1977	28	1962	1990
Newlyn	62	1917	1991	61	1916	1976	77	1916	1991
Ilfracombe	10	1968	1990	NA	NA	NA	6	1984	1989
Hinkley	3	1990	1993	NA	NA	NA	NA	NA	NA
Avonmouth	9	1961	1992	64	1883	1986	NA	NA	NA
Newport	4	1962	1993	37	1899	1988	NA	NA	NA
Swansea	5	1961	1985	36	1936	1981	12	1961	1985
Mumbles	4	1989	1992	NA	NA	NA	NA	NA	NA
Milford Haven	20	1961	1992	26	1951	1977	15	1969	1991
Fishguard	24	1963	1991	NA	NA	NA	12	1975	1986
Barmouth	2	1992	1993	NA	NA	NA	NA	NA	NA
Holyhead	19	1964	1991	13	1959	1971	45	1938	1990
Liverpool	3	1991	1993	NA	NA	NA	25	1958	1983
Heysham	21	1964	1990	36	1940	1984	23	1962	1991
Port Erin	2	1992	1993	NA	NA	NA	NA	NA	NA
Workington	2	1992	1993	NA	NA	NA	NA	NA	NA
Portpatrick	21	1968	1988	NA	NA	NA	23	1968	1991
Millport	4	1987	1992	NA	NA	NA	NA	NA	NA
Islay	3	1991	1993	NA	NA	NA	NA	NA	NA
Tobermory	2	1990	1991	NA	NA	NA	4	1989	1991
Stornoway	5	1988	1992	NA	NA	NA	12	1977	1991
Ullapool	15	1966	1991	12	1963	1977	9	1983	1991
Kinlochbervie	3	1991	1993	NA	NA	NA	NA	NA	NA
Lerwick	24	1959	1991	11	1965	1976	34	1957	1991

Table 2.2: Available data at the south and west coast study sites. NA denotes no data are available.

mean is only a rough estimate of the mean sea-level, see Woodworth (1987, 1990) and Tawn *et al.* (1994). Estimation is further complicated by trends in mean sea-levels arising from a combination of an increase in the volume of water in the oceans resulting from 'greenhouse' expansion, and vertical movements of the land, upon which the tide gauge is fixed. The approach in this project, unless otherwise stated, is to ignore trends in mean

sea-level at the tidal analysis stage, and so allow the trend to be transferred into the surge series.

The Tidal Series

The tidal level is the component of the observed sea-level which is determined by astronomical forcing. It can be shown (Pugh, 1987) that the tidal level is

$$X_t = Z_0 + \sum_{i=1}^{m} f_{1_i} H_i \cos(\omega_i t + V_i + f_{2_i} - g_i). \tag{2.2.2}$$

Here m is the number of sinusoidal constituents, typically taken to be at least 60, Z_0 the mean level, H_i and g_i are the parameters corresponding to the amplitude and phase lag on the equilibrium tide of the ith constituent of the tide which vary with site; whereas for constituent i, the terms ω_i, V_i, $f_{1,i}$ and $f_{2,i}$ are the angular speed, the equilibrium phase, nodal amplitude and phase corrections respectively, which are known and common for all sites. For data from a specific site, the H_i and g_i parameters are estimated by standard tidal estimation techniques as used by Proudman Oceanographic Laboratory. Using the estimated tidal parameters, tidal series can be predicted at any past or future time and so the complete tidal series over the span of the data can be provided.

The Surge Series

The surge series is defined as the residual of the tidal series, i.e. the difference between the observed still water series and the predicted tide. The difference between observed levels and tidal predictions is due to meteorological forcing in the form of changes in air pressure and winds which cause surges to be generated (Pugh, 1987). As a consequence of the tidal analysis the surge sample has a zero average although may contain trends. On the east coast the surge variability generally increases with distance down the coast, reaching a maximum at the outer reaches of the Thames estuary, while on the west coast the surge is much less variable except at shallow water sites. The surge distribution is uni-modal, with the mode close to zero, and is positively skewed, as a consequence of large positive surge levels being more likely than large negative surges.

Part I

All data usage in tide-surge site-by-site method

Chapter 3

Background

Dixon and Tawn (1994) found that for sites with at least 5-10 years of good quality hourly data the best method of analysis of extreme sea-level data is the **Revised Joint Probability Method** (RJPM), and that the **Joint Probability Method** (JPM) is the most suitable for sites with less than 5-10 years of hourly data. The application of the RJPM and JPM in Dixon and Tawn (1994) was largely as described by Tawn and Vassie (1989) and Tawn (1992). However, there are a number of areas in which the RJPM can be improved to better handle trends and extended to incorporate historical data. To clarify later discussion we first briefly review the RJPM which is presented in full detail by Dixon and Tawn (1994, Sections 4.4 and 6.1).

3.1 A brief review of the RJPM

In the RJPM the distribution of the annual maximum still water level in year i is taken to be

$$G_i(z) = \left\{ \prod_{t=1}^{T} F_{Y,i|X}(z - X_t \mid X_t) \right\}^{\theta_l N/T} \tag{3.1.1}$$

where $F_{Y,i|X}$ is the distribution of the surge, Y, in year i conditional on the tide X; X_t is the hourly tidal level at time t; T and N are the number of hours in a nodal tidal cycle and a year respectively; and θ_l is the extremal index of the still water level series (so θ_l^{-1} is the mean duration of extreme still water level events). An important component of the method is the choice and estimation of the model for $F_{Y,i|X}$.

Dixon and Tawn (1994) develop functions of the tide, $a(X)$ and $b(X)$, which they term interaction functions, which are such that the location-scale normalisation of the surge series, Y_t,

$$S_t^* = [Y_t - a(X_t)]/b(X_t), \tag{3.1.2}$$

is approximately stationary with respect to the tide. Tide-surge interaction is removed by this normalisation, although the normalised surge series retains any long term trend that is present in the original surge series.

Standard extreme value techniques are then applied to the S_t^* series; for a high threshold, u, below the threshold, where the data are relatively dense, a non-parametric estimate of the distribution function is used whereas above the threshold a parametric model is used which is theoretically justified by extreme value theory. The model adopted for the distribution of S_t^*, in year i, is

$$F_{s,i}(y) = \begin{cases} \exp\{-(N\theta_s)^{-1}[1 - k_s(y - \mu_s(i))/\sigma_s]_+^{1/k_s}\} & \text{for} \quad y > u, \\ \tilde{F}_s(y) & \text{for} \quad y \leq u. \end{cases} \tag{3.1.3}$$

where θ_s is the extremal index of the stationary surge series,

$$\mu_s(i) = \alpha_s + \beta_s(i - i_0),$$

where i_0 is the base year, and $\mu_s(i), \sigma_s$ and k_s are the parameters of the annual maximum distribution for the stationary surge in year i. The parameter β_s is the trend in the extreme levels of the stationary surge series. The notation $[x]_+$ is used here, and throughout the report, to denote $\max(x, 0)$.

Taking the distribution of the transformed surge, S^*, to be (3.1.3) is equivalent to taking the distribution of the surge conditional on tidal level X to be

$$F_{Y,i|X}(y) = \begin{cases} \exp\{-(N\theta_s)^{-1}[1 - k_s(y - \mu_s(X,i))/\sigma_s(X)]_+^{1/k_s}\} & \text{for} \quad y > u(X), \\ \tilde{F}_s\{[y - a(X)]/b(X)\} & \text{for} \quad y \leq u(X), \end{cases} \tag{3.1.4}$$

where

$$\mu_s(X,i) = \mu_s(i)b(X) + a(X) = \alpha_s b(X) + a(X) + \beta_s(i - i_0),$$

and

$$\sigma_s(X) = \sigma_s b(X), \quad u(X) = a(X) + b(X)u.$$

Replacing the $F_{Y,i|X}$ term in equation (3.1.1) by the expression in equation (3.1.4) gives the distribution of the annual maximum still water level obtained by the RJPM to be

$$G_i(z) = \exp\left\{-\theta_l(T\theta_s)^{-1}\sum_{t=1}^{T}[1 - k_s(z - \mu_s(i)b(X_t) - a(X_t) - X_t)/\{\sigma_s b(X_t)\}]_+^{1/k_s}\right\} \tag{3.1.5}$$

for large z. Return level estimates are obtained by solving the equation

$$\hat{G}_i(z_i(p)) = 1 - p$$

for the return level $z_i(p)$. Here $\hat{G}_i$ is the distribution function given by equation (3.1.5) once the parameters are replaced by estimated values.

Key features of this method are the steps in the estimation of the parameters of the model. There are three types of parameters which require different methods of estimation. The first type of parameters are the interaction functions $a(X)$ and $b(X)$, the second type $(\alpha_s, \beta_s, \sigma_s, k_s)$ are associated with extreme surge levels, and the third type are the extremal indices θ_l and θ_s associated with temporal dependence in the still water levels and surge processes respectively.

Interaction functions: the function $a(X)$ was estimated as the 98% quantile of the surge distribution conditional on the tidal level being X, and $b(X)$ was estimated as the difference between the 99% and 98%, quantiles of the surge distribution conditional on the tidal level being X. These quantiles were derived from the empirical conditional distribution of surges given the tide obtained from pooling all the hourly surge data taking due account of the tidal level. In practice these quantile values were estimated for each of a number of tidal bands and the levels obtained were linearly interpolated to produce continuous curves.

Surge levels parameters: the extremal surge parameters $\alpha_s, \beta_s, \sigma_s$, and k_s were estimated using the r-largest approach (see Dixon and Tawn, 1994, Section 4.2) applied to the normalised surge levels. The surge data used were the $r = 8$ largest annual independent transformed surge values for each year of hourly data.

Extremal indices: estimates of the extremal index parameters, θ_l and θ_s are obtained by estimating the reciprocals θ_l^{-1} and θ_s^{-1}. Specifically the reciprocals are estimated by the average size of clusters of exceedances of a high threshold within an independent extreme event for the sea-level and surge processes respectively. Dixon and Tawn (1994, Section 6.2) proposed a procedure by which only the ratio of these parameters was estimated.

3.2 Limitations of the existing method

There are a number of limitations of the method as it stands, these are:

1. **Data usage**:
   In Dixon and Tawn (1994) only hourly data were used in the analysis of extreme sea-levels at the A-class sites. Typically, a site can have three different types of data: annual maximum, annual mean, and hourly sea-levels. The information about the process contained in historical data of annual mean and maximum sea-levels is

particularly valuable in terms of the trend in the process. Furthermore, since design levels have previously been based on the annual maximum analysis of the historical annual maximum data these data are thought to contain relevant information about the extreme sea-level process.

2. **Handling of trends**:
   The approach in Dixon and Tawn (1994) takes the trend to be present only in the extreme surges, therefore it ignores information about the trend from other features of the surge distribution, such as is given by the mean sea-level data. If the trend is common throughout the surge distribution the adopted approach is highly inefficient. In fact the trend in extreme surges is taken by Dixon and Tawn (1994) to be constant with respect to the tidal level for extremes of the stationary surge series, which implies that for the actual surge extremes the trend is $\beta_s b(X)$, i.e. it varies with the associated tidal level. This feature is artificial, and arises from the poor handling of the trend in the estimation of the interaction functions and empirical distribution of surges. In particular, in the estimation of the interaction functions the series is taken as stationary so a pooled estimate over years is used. Correspondingly, the values of the estimates of the interaction functions and the empirical surge distribution depend on the span of data used as the trend is incorporated in these estimates.

3. **Extension to spatial parameters**:
   In Parts I and II of this report the extreme sea-level methods that are applied are site-by-site methods, so spatial interpretation of parameters is unimportant. However, for the extension of these methods to spatial methods, the parameters must be spatially interpretable. This is not the case for the interaction functions and the empirical distribution of the surge, as estimated by Dixon and Tawn (1994), because these estimates depend on the span of data used due to the poor handling of the trends in their estimation, as identified above. As the spans of available data differ from site-to-site these parameters do not correspond to the same quantity at each site. Adjustments of these parameters are required to give spatially interpretable parameters.

3.3 Features to be incorporated in the new method

The three features, identified in Section 3.2, which require improvement in the RJPM are therefore the usage of available data, the manner in which trends in extremes are handled and spatial consideration of parameter interpretation.

To overcome these weaknesses the new RJPM should:

- make use of all the available data types at a site, and continue to exploit knowledge of the constituent tide and surge processes.
- make the key assumption that the trend in the surge process is common to all aspects of the process, i.e. the same trend occurs in mean surge levels, surge quantile levels, extreme surge levels, and surge quantiles conditional on any tidal level. This is a departure from the assumptions of Dixon and Tawn (1994), being a much stronger assumption, but is consistent with the findings in Dixon and Tawn (1994) where extreme sea-level trends were found to be poorly estimated, but not to be significantly different from the estimated mean sea-level trends.
- have model parameters for which a natural spatial interpretation can be given.

3.4 Outline to Part I

The methods used in the new RJPM, the **Spatial Revised Joint Probability Method, (SRJPM)** are relatively complicated so are given in detail, step-by-step, in Sections 4.1–4.7. In particular, in Section 4.1 the estimation of the interaction functions is developed. These functions are extended from the forms in Dixon and Tawn (1994) to incorporate trends in large surge quantiles. In Section 4.2 details are given of the estimation of the model parameters for the upper tail of the surge distribution conditional on the associated tidal level. The method differs from Dixon and Tawn (1994) as a point process method is used instead of the r-largest approach. In Section 4.3 details are given of how historical annual maximum data can be incorporated into the analysis. Similarly, in Section 4.4 we explain how historical data on the annual mean sea-level can be used to improve the estimation of the model.

Once the separate elements of the model have been described, in Section 4.5 these models are combined to give an overall likelihood for the model. This likelihood function is used as the basis for maximum likelihood estimation of the parameters of the model which in turn determine the return levels and other design quantities discussed in Section 4.6. Finally in Section 4.7, technicalities of the methods are discussed.

At each stage of the analysis we display results from a test site, Immingham, showing the impact of assumptions, and how the different information about the parameters of interest comes from the different sources of information.

Chapter 4

The Spatial RJPM (SRJPM)

4.1 Estimation of the interaction functions

Results obtained from the RJPM for sites with tide-surge interaction critically depend upon the interaction functions $a(X)$ and $b(X)$. As explained in Section 6.1 of Dixon and Tawn (1994), previously these functions were obtained from empirical estimates of quantiles of the surge distribution conditional on the tide level being X. Specifically, if $a_1(X)$ and $a_2(X)$ are the p_1 and p_2 quantiles ($p_1 < p_2$) of the surge distribution conditional on tide X then

$$a(X) = a_1(X) \tag{4.1.1}$$

and

$$b(X) = a_2(X) - a_1(X) \tag{4.1.2}$$

Dixon and Tawn (1994) took $p_1 = 98\%$ and $p_2 = 99\%$ over the entire span of the tide-surge data. In that study, the principle interest was in the interaction itself, and trends in the surge levels were ignored in the construction of the interaction functions. As stated above our model for extreme sea-levels is based on an assumption of a trend in the surge process which is common to all aspects of the series, so is present in both $a_1(X)$ and $a_2(X)$. In this section the model for interaction used by Dixon and Tawn (1994) is extended to allow both $a_1(X)$ and $a_2(X)$ to depend on time.

4.1.1 Model

Let $a_1(X,i)$ and $a_2(X,i)$ denote the true p_1 and p_2 quantiles of the surge distribution conditional on the tidal level being X in year i, i.e. $a_j(X,i)$ for $j = 1,2$ satisfy

$$\tilde{F}_{Y,i|X}(a_j(X,i)) = p_j$$

where $\tilde{F}_{Y,i|X}$ is an estimate of $F_{Y,i|X}$. It is assumed that there is a linear trend in the surge process with gradient β, so

$$a_j(X,i) = a_j(X,i_0) + \beta(i - i_0), \quad \text{for } j = 1,2, \tag{4.1.3}$$

where i_0 is the base year for the analysis. Note that the trend in both quantiles is identical, following from the underlying assumption of a common trend in the whole surge process.

The time dependent interaction functions $a(X,i)$ and $b(X,i)$ are naturally derived from $a_1(X,i)$ and $a_2(X,i)$, as in the time independent analysis of Dixon and Tawn (1994) given by (4.1.1) and (4.1.2), so

$$a(X,i) = a_1(X,i) = a_1(X,i_0) + \beta(i - i_0)$$

and

$$\begin{aligned} b(X,i) &= a_2(X,i) - a_1(X,i) \\ &= a_2(X,i_0) + \beta(i - i_0) - [a_1(X,i_0) + \beta(i - i_0)] \\ &= a_2(X,i_0) - a_1(X,i_0). \end{aligned}$$

So $a(X,i)$ is a linear function of time and $b(X,i)$ does not depend on time (or the trend). Subsequently we drop the time dependence and denote $b(X,i)$ by $b(X)$. Note that if there is no trend then both interaction functions are identical in form to those used by Dixon and Tawn (1994).

4.1.2 Estimation

In any year the realised value of a quantile of the surge distribution will differ from the linear form predicted by our model, so estimation of the interaction parameters requires statistical analysis of realised data on the quantiles. First these year by year quantile data must be obtained. As in Dixon and Tawn (1994) we partition the tidal range into 10 equi-probable tidal bands. Taking T_j to be the mid-point of the jth tidal band we let $A_{i,j}^{(l)}$ denote the realised quantile ($l = 1$ and 2 correspond to the p_1 and p_2 quantiles respectively) in year i at the tidal level T_j. This value is constructed from the hourly data using a kernel density estimate of the surge distribution conditional on the tidal level and integrating this to find the appropriate level. We assume that these empirical quantiles are distributed as

$$A_{i,j}^{(l)} \sim N(a_l(T_j,i), c_{j,l}^2) \text{ for } l = 1,2 \text{ and } j = 1,\ldots,10. \tag{4.1.4}$$

Here i is the year of interest. The realised quantile is taken to have the true value as its mean but varies around that value with a Gaussian error with variance $c_{j,l}^2$, i.e. a

tidal band and quantile specific variance. The choice of the Gaussian error model is based on the result that random variables which are empirical quantiles follow a Gaussian distribution for large sample sizes, David (1981).

From the underlying model for the true quantile, given by (4.1.3), with $\alpha_j = a_1(T_j, i_0)$, so (4.1.4) gives that

$$\begin{aligned} A_{i,j}^{(1)} &\sim N(a_1(T_j, i_0) + \beta(i - i_0), c_{j,1}^2) \\ &\sim N(\alpha_j + \beta(i - i_0), c_j^2) \end{aligned} \tag{4.1.5}$$

where $c_{j,1}^2 = c_j^2$. Also

$$\begin{aligned} A_{i,j}^{(2)} - A_{i,j}^{(1)} &\sim N(a_2(T_j, i) - a_1(T_j, i), d_j^2) \\ &\sim N(b(T_j), d_j^2), \end{aligned}$$

where d_j^2 is the variance of the difference in the random variables. Hence the statistical model for the empirical year-by-year estimates of the quantiles is that the p_1 quantile of the distribution is linear in time with a Gaussian error structure, whilst the difference in quantiles has constant mean, $b(X)$, through time.

The final assumption is that quantile data from year to year, and from tidal band to tidal band, are independent. The inference problem is now a standard statistical one in each case. The parameters (α_j, c_j^2) for $j = 1, \ldots, 10$ and the trend β are estimated by applying standard regression methods to the realised quantile data. The function $b(T_j)$ is estimated by the sample mean of the $A_{i,j}^{(2)} - A_{i,j}^{(1)}$ values over the years.

This provides estimates of the interaction functions $a(X, i)$ and $b(X)$ at the mid-points of the tidal bands only. The final stage of the inference is to estimate these functions for the entire range of possible values for the tide X. This is achieved using a two step procedure. First the tidal levels are transformed onto the interval $[0, 1]$. Specifically, the tidal level X_t is transformed to X_t^* at time t, using the relationship

$$X_t^* = \tilde{F}_{tide}(X_t),$$

where $\tilde{F}_{tide}$ is an empirical estimate of the tidal distribution function. Similarly define $T_j^* = \tilde{F}_{tide}(T_j)$. Application of a standard form of weighted kernel regression smoothing to the points $\{(T_j^*, \hat{\alpha}_j); j = 1, \ldots, 10\}$, where $\hat{\alpha}_j$ is the estimate of α_j gives a smooth estimated function $\tilde{a}(X^*, i_0)$. Transforming the tide back to the original scale gives an estimate $\tilde{a}(X, i_0)$. Estimation of $\tilde{b}(X)$ is similarly based on kernel regression smoothing on the transformed tidal scale. In conclusion, our estimates for the interaction functions $a(X, i)$ and $b(X)$, in the absence of other information, are

$$\tilde{a}(X, i_0) + \hat{\beta}(i - i_0) \text{ and } \tilde{b}(X)$$

respectively. Here $\hat{\beta}$ is the regression estimate of β obtained above.

4.1.3 Immingham example

For the Immingham hourly sea-level data the model described above is fitted to the derived surge quantile data conditional on the associated tidal level. This leads to a common trend estimate over tidal bands and estimates of interaction functions, $a(T_j, i_0) = \alpha_j$, and the standard deviations, c_j, associated with annual quantile data from each tidal band. These estimates are shown in Table 4.1 from which it is seen that for low tides, corresponding to low tidal bands, the surge quantiles are 20% larger than the corresponding surge quantiles for high tides (i.e. high tidal bands). This form of interaction was found previously for Immingham in Dixon and Tawn (1994, Figure 20) and is unusual for the east coast, as the more common form of interaction is for the largest conditional surge quantiles to occur at mid-tide (i.e. the middle tidal bands).

Tidal band	Interaction function in base year, $\hat{\alpha}_j$	$se(\hat{\alpha}_j)$	Standard deviation, $\hat{c}_j$	$se(\hat{c}_j)$
1	0.493	0.0171	0.0890	0.0123
2	0.502	0.0150	0.0779	0.0107
3	0.489	0.0136	0.0709	0.0097
4	0.498	0.0172	0.0892	0.0123
5	0.473	0.0162	0.0839	0.0115
6	0.475	0.0144	0.0750	0.0103
7	0.467	0.0133	0.0691	0.0095
8	0.458	0.0137	0.0713	0.0098
9	0.417	0.0119	0.0616	0.0085
10	0.408	0.0123	0.0644	0.0088
trend estimate, $\hat{\beta} = 2.917$mm/yr, with $se(\hat{\beta}) = 0.584$				

Table 4.1: Estimated interaction function parameters for Immingham: based on data of quantiles of conditional surge process. Unless stated the units for the results are metres.

Figure 4.1 shows the corresponding estimates $\hat{\alpha}_j$, for $j = 1, \ldots, 10$ plotted against the associated tidal level, shown on a uniform $(0, 1)$ scale (corresponding to the transformed tidal variable, X^*). Also shown on this figure is the smooth kernel regression function, $\tilde{a}(X^*, i_0)$, given on the transformed scale. This function fits the point estimates well and provides a natural smooth interpolation and extrapolation of these points.

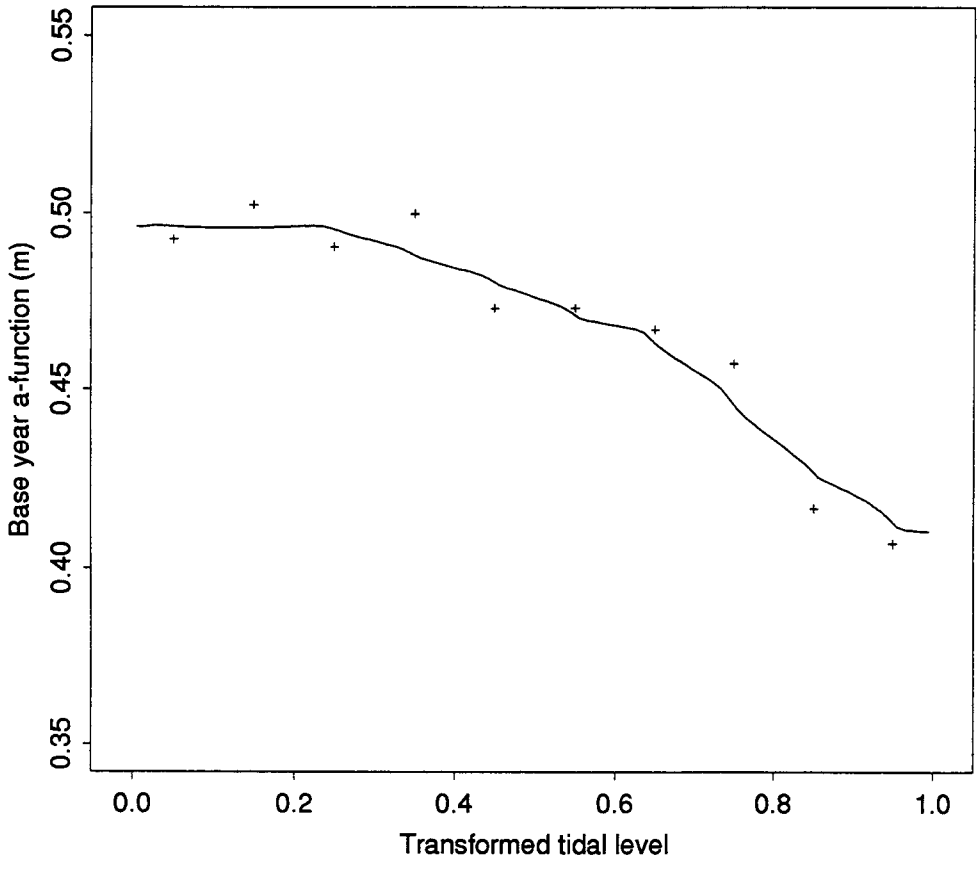

Figure 4.1: Base year a-function, i.e. $a(X^*, i_0)$, plotted against transformed tidal level, X^*.

4.2 The Point Process method

Most methods for the analysis of extreme sea-levels require some form of theoretically justified extreme value model in order to extrapolate the tail of the sea-level distribution with confidence. Dixon and Tawn (1994) based this part of the model on the asymptotic joint distribution of the r-largest values applied to annual events. Here we use the point process method of Smith (1989) which is generally a more appropriate and flexible approach. In cases of stationary sequences the r-largest and point process methods are basically identical. However when series exhibit trends and other forms of non-stationarity, such as is induced by tide-surge interactions, then the methods differ and the point process method is the more flexible and natural approach.

In this section we first describe the basic point process approach in a setting where the process is independent and identically distributed (i.i.d.). Then, in Section 4.2.1 we show how the methods can be extended to the estimation of extreme surges in the presence of trends and interaction.

For an i.i.d. process the starting point for the point process method is that the annual

maximum of the process is assumed to follow the generalised extreme value distribution, GEV(μ, σ, k), which has distribution function evaluated at x of

$$\exp\{-[1 - k(x - \mu)/\sigma]_+^{1/k}\} \tag{4.2.1}$$

where $\mu, \sigma(\sigma > 0)$, and k are location, scale, and shape parameters respectively, and $[x]_+ = \max(x, 0)$. The point process method specifies a model for the intensity of points above a high threshold, u say. Specifically, points which exceed u are taken to be points from a non-homogeneous Poisson process with intensity function $\lambda(x)$ for the level x. It follows that

$$\lambda(x) = \sigma^{-1}[1 - k(x - \mu)/\sigma]_+^{1/k-1}.$$

The likelihood for a realisation of such a process is

$$\Pr\{\text{ none of the points exceed } u\} \times \lambda(x_1) \times \ldots \times \lambda(x_{ne}) \tag{4.2.2}$$

where $x_1, \ldots, x_{ne}$ is an enumeration of the observed exceedances of the threshold u. Since the annual maximum of the process is GEV(μ, σ, k), it follows that

$$\Pr\{\text{ none of the points exceed } u\} = \exp\{-n_{yrs}[1 - k(u - \mu)/\sigma]_+^{1/k}\}.$$

where n_{yrs} is the number of years of observations. Maximising the point process likelihood, with respect to μ, σ, and k, provides estimates of the parameters of the distribution of the annual maximum of the process. In this case, the method is almost identical in form to the r-largest method. The difference between the methods is only in that here a pre-chosen threshold, u, is used to decide what constitutes an extreme value as opposed to using an extreme observation (the rth largest) of the process in the r largest method. So the point process method uses the same threshold through time, whereas in the r-largest annual events method the threshold varies with the r-largest annual value. Consequently, results from the two methods can differ since whether a value is classified as extreme or not, in the r-largest method, depends on the number of other independent extreme events in the year. The point process approach uses a consistent definition of an extreme value and so is preferable.

For later extension to the more complex case of surge extremes, it is helpful to write the likelihood for the point process in an alternative way. Breaking down the likelihood into the contribution from each observation we have that if the observation, y, is below the threshold the likelihood contribution is

$$L_1(y; \boldsymbol{\theta}) = \exp\left\{-\frac{1}{N}\left[1 - k\left(\frac{u - \mu}{\sigma}\right)\right]_+^{1/k}\right\},$$

where $\boldsymbol{\theta} = (\mu, \sigma, k)$ and N is the number of observations in a year. Alternatively, when the observation, y is above the threshold the likelihood contribution is

$$L_1(y;\boldsymbol{\theta}) = (\sigma N)^{-1}\left[1-k\left(\frac{y-\mu}{\sigma}\right)\right]_+^{1/k-1}\exp\left\{-\frac{1}{N}\left[1-k\left(\frac{u-\mu}{\sigma}\right)\right]_+^{1/k}\right\}.$$

The likelihood of the extremes of the process is then

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{N\times n_{yrs}} L_1(y_t, \boldsymbol{\theta}),$$

i.e. the product of the likelihood contributions from the individual observations. This overall likelihood is equivalent to equation (4.2.2) up to a scaling constant.

Now consider the modifications required when the process is stationary, for example when the hourly distribution is invariant to time but there is temporal-dependence. Extension of the point process approach to cover dependence in the observed extreme events is straightforward. In this case a data point is classified as extreme if it is an independent event maxima which exceeds the threshold, i.e. the analysis is based on de-clustered extreme values. Suppose that the distribution of the annual maximum is GEV(μ, σ, k), and the extremal index of the series is θ_s (i.e. θ_s^{-1} is the mean duration of extreme events). It follows that if the observation, y, is below the threshold, i.e. $y < u$, or y is not the maximum of an event, then the likelihood contribution is

$$L_2(y;\boldsymbol{\theta}) = \exp\left\{-\frac{1}{N\theta_s}\left[1-k\left(\frac{u-\mu}{\sigma}\right)\right]_+^{1/k}\right\},$$

where $\boldsymbol{\theta} = (\mu, \sigma, k)$ and N is the number of observations in a year. Alternatively, when the observation, y is above the threshold ($y \geq u$) and corresponds to an event maximum then the likelihood contribution is

$$L_2(y;\boldsymbol{\theta}) = (\sigma N\theta_s)^{-1}\left[1-k\left(\frac{y-\mu}{\sigma}\right)\right]_+^{1/k-1}\exp\left\{-\frac{1}{(N\theta_s)}\left[1-k\left(\frac{u-\mu}{\sigma}\right)\right]_+^{1/k}\right\},$$

The likelihood of the extremes of the process is then

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{N\times n_{yrs}} L_2(y_t, \boldsymbol{\theta}).$$

4.2.1 Application for extreme surges

The observed surge series deviates from the independent and identically distributed process discussed above in three ways:

1. temporal dependence,

2. non-stationarity in the form of long term trends

3. non-stationarity in the form of tide-surge interaction.

The first of these features is incorporated into the analysis by studying event maxima, as discussed above, which are temporally independent. The non-stationarity is incorporated by allowing the extreme value parameters to depend on the tidal state and year through the interaction functions of Section 4.1. Throughout this section the estimated interaction functions $\tilde{a}(X, i_0)$ and $\tilde{b}(X)$ are taken to be exact and known, but the trend β is taken as unknown, and i_0 is again taken to be the base year for the analysis. The extension for interaction is identical in form to the methods discussed in Section 6.1 of Dixon and Tawn (1994), differing only in the form of the point process method used and the estimation of the interaction functions to include trends.

We adopt the same model for the extremal dependence on tides as in Dixon and Tawn (1994, equation 6.3), so for tidal level X in year i the parameters μ, σ and k are replaced by $\mu(X, i), \sigma(X)$ and $k(X)$, where

$$\begin{aligned} \mu(X,i) &= \mu\tilde{b}(X) + \tilde{a}(X,i) \\ &= \mu\tilde{b}(X) + \tilde{a}(X,i_0) + \beta(i - i_0) \\ \sigma(X) &= \tilde{b}(X)\sigma \\ k(X) &= k. \end{aligned} \tag{4.2.3}$$

Here β is the trend, and the parameters μ, σ and k describe the extreme tail of the surge distribution once interaction features have been removed. Therefore this model corresponds to there being a linear trend in the extreme surge process together with changes in level and variation of the extreme surges which depend on the concomitant tidal level.

There are obvious connections with the models of Dixon and Tawn (1994) here: the relationships above imply that

$$S_t^* = \frac{[Y_t - \tilde{a}(X_t, i)]}{\tilde{b}(X_t)}$$

is a stationary surge sequence in the extremes. This is identical to Dixon and Tawn (1994, equation 6.2) except here trends are also included in the location of the normalisation.

The above framework assumes that the trend in the extreme surges is the same as in the whole of the surge process. If it is assumed that the trend in extreme surges is different from the rest of the surge series then an alternative trend term could be included. We shall not pursue that feature for the reasons given in the introduction to the section (Section 3.2).

The parameters of the extremal behaviour of the process are specified above. Before the corresponding point process likelihood can be derived in this case, the threshold which determines which surge observations are extreme must be chosen. The choice of a threshold in the non-stationary setting of extreme surges is more problematic than in the i.i.d. case considered in Section 4.1, since there a single value, u, defines an equally extreme point throughout time. Here we use a method of threshold choice based on a time-dependent threshold $u(X,i)$ which varies with tidal level, X, and year, i. This threshold is chosen to give an equally extreme level throughout time. Specifically, we take $u(X,i)$ to be a quantile of the conditional distribution of surges given the tidal level and year, so $u(X,i)$ satisfies

$$F_{Y,i|X}(u(X,i) \mid X) = 1 - p,$$

where $F_{Y,i|X}$ is the surge distribution function for year i conditional on the tidal level being X. This distribution function is estimated empirically; see Section 4.7 for the technical details of this. In applications we have taken $p = 0.005$.

Since applications of the methods require temporally independent extreme surge data, a declustering procedure is used. Extreme surge events are identified and declustered, relative to the time-dependent threshold, $u(X,i)$, to produce independent surge event maxima.

Now consider the likelihood function for the extremal surge process. The method of construction of the likelihood function is similar to the stationary case for the point process method. Suppose that the hourly surge observation, y, in year i occurs at a time corresponding to a tidal level of X. If y is below the threshold, i.e. $y < u(X,i)$, or y is not an event maximum, the likelihood contribution is

$$L_3(y;\boldsymbol{\theta}) = \exp\left\{-\frac{1}{N\theta_s}\left[1 - k\left(\frac{u(X,i)-\mu(X,i)}{\sigma(X)}\right)\right]_+^{1/k}\right\},$$

where N is the number of observations in a year, θ_s is the extremal index for the surge series, $\boldsymbol{\theta} = (\mu, \sigma, k, \beta)$ and $\mu(X)$ and $\sigma(X)$ are defined by equation (4.2.3). Alternatively, when the observation y is above the threshold (i.e. $y \geq u(X,i)$) and corresponds to an event maximum then the likelihood contribution is

$$\begin{aligned} L_3(y;\boldsymbol{\theta}) &= (\sigma(X)N\theta_s)^{-1}\left[1 - k\left(\frac{y-\mu(X,i)}{\sigma(X)}\right)\right]_+^{1/k-1} \\ &\times \exp\left\{-\frac{1}{(N\theta_s)}\left[1 - k\left(\frac{u(X,i)-\mu(X,i)}{\sigma(X)}\right)\right]_+^{1/k}\right\}, \end{aligned}$$

The likelihood of the extremes of the process is then

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{N\times n_{yrs}} L_3(y_t, \boldsymbol{\theta}).$$

4.2.2 Immingham example

In Section 4.1.3 the interaction functions were estimated to be

$$\tilde{a}(X,i) = \tilde{a}(X,i_0) + \hat{\beta}(i - i_0),$$

with estimated trend $\hat{\beta}$ and interaction function $\tilde{a}(X,i_0)$ in the base year i_0 obtained by using all the hourly data for Immingham. Here we take $\tilde{a}(X,i_0)$ as known, but estimate β and the point process parameters μ, σ and k using the point process likelihood (i.e. using extreme surge hourly data only). These parameter estimates are given in Table 4.2. The estimates show that if the trend is estimated using only extreme surge data from the

Parameter	Estimate	Standard Error
μ	6.505	0.343
σ	2.167	0.255
k	-0.1312	0.0683
β	1.523	1.561

Table 4.2: Maximum likelihood estimates and standard errors obtained using the point process method for Immingham. Estimates are derived taking the interaction function, $a(X,i_0)$ as known. The units of the trend estimate are mm/yr.

period of hourly observations then much larger uncertainty is obtained than from using the surge quantile data in Section 4.1.3. However, the trend estimates obtained from the two different sources are not significantly different from one another, suggesting that pooling the information should improve the trend estimate. This aspect is considered further in Section 4.5.

At this point there is little to be said about the estimates of the other parameters except to note that $\hat{k}$ is similar to that obtained by Dixon and Tawn (1994).

4.3 Annual maximum inclusion

Throughout Dixon and Tawn (1994), and this report so far, all the information about the extreme still water level process has been obtained from hourly observations of the tide and surge, and knowledge of the tidal properties. Prior to the development of joint probability methods by Pugh and Vassie (1979, 1980) all extreme sea-level analysis was based on annual maximum data, using the annual maximum method. Notable examples of this method are Graff (1981) and Coles and Tawn (1990). Dixon and Tawn (1994)

found that annual maximum methods (and the corresponding r-largest methods) are not ideal for still water level data as the inter-annual variation of the tides leads to invalid extrapolation. This error in the annual maximum method was found to be of particular importance for west coast sites.

Although the annual maximum method is not the most effective approach for estimating return levels, the information contained in the annual maximum data set are considered a valuable source of data additional to the hourly data for a site. Much more processing of charts and tide-gauge information is required to record and store hourly sea-levels than the extraction of just the maximum for each year. This means that although hourly records may be short at a site, there are often long historical records of annual maximum levels.

In this section we show how these historical annual maximum data can be incorporated naturally into the SRJPM. Specifically, by incorporating tidal knowledge, these data can give extra information about the parameters μ, σ and k of the extreme surge distribution, as well as the trend, β. Of most value is the information gained about the trend, since it is often more important to have sparse data over a long time span than high density data of limited extent. This applies here since the annual maximum data are often available for long time spans and for earlier years than the hourly data; see Tables 2.1 to 2.2.

The problem with making inferences about the extreme surge process from historical annual maximum still water levels is that the magnitude of the maximum sea-level in a year is recorded, but generally no information is available about either the time of the maximum or the associated tidal level. Therefore the associated annual maximum surge level is not known for that year, and worse, the annual maximum sea-level may not have corresponded to an extreme surge. This is the underlying reason why annual maxima methods are generally poor for extrapolation to long return periods. The information the annual maximum gives is a bound on the combinations of tide and surge in that year. This contains primarily information about levels in the year and hence over a large span of years provides information about the trend. In conjunction with knowledge of tidal levels that occurred over a year the annual maximum still water level also gives a bound on the surge levels that occurred in that year, and hence provide some information about the tail of the extreme surge distribution.

Dixon and Tawn (1994, Section 6.1) derived the distribution of the annual maximum still water level for a site with tide-surge interaction. There the distribution function for the tide was derived from a full nodal cycle of 19 years. As there can be a significant variation in tidal characteristics from year-to-year within the nodal cycle, the distribution of the annual maximum for a specific year is more correctly obtained by considering only the tides specific to each year.

4.3.1 Annual maximum model

Here we derive the distribution of the annual maximum still water in year i incorporating the features of the point process approach and the updated method for estimating the interaction functions.

Following Section 4.2, let $F_{Y,i|X}$ represent the distribution function of an hourly surge level, in year i, given that the associated tidal level is X, and let F_{tide} denote the distribution function of tidal levels. The function F_{tide} is now extended to $F_{tide,i}$ which represents the distribution function of tidal levels in year i. The distribution of the maximum hourly still water level in year i, $G_i(z)$, is

$$G_i(z) = \left[\prod_{j=1}^{N} F_{Y,i|X}(z - X_{t_{j,i}}|X = X_{t_{j,i}})\right]^{\theta_l}, \tag{4.3.1}$$

where θ_l is the still water level extremal index, and $t_{1,i}, \ldots, t_{N,i}$ is the sequence of N hourly times in year i. Rewriting equation (4.3.1) gives

$$G_i(z) = \exp\left\{\theta_l N \frac{\sum_{j=1}^{N} \log F_{Y,i|X}(z - X_{t_{j,i}}|X = X_{t_{j,i}})}{N}\right\}. \tag{4.3.2}$$

The term

$$\frac{\sum_{j=1}^{N} \log F_{Y,i|X}(z - X_{t_{j,i}}|X = X_{t_{j,i}})}{N}$$

in equation (4.3.2) is the average, over hourly tidal levels, of the log of the conditional surge distribution function. This average is more accurately given by the corresponding average over continuous time,

$$\int_0^1 \log F_{Y,i|X^*}(z - X|X^* = x)dx = \int_0^1 \log F_{Y,i|X^*}(z - F^{-1}_{tide,i}(x)|X^* = x)dx$$

where $F^{-1}_{tide,i}(x)$ is the inverse of the distribution of tidal levels in year i. As before X^* denotes the tide transformed onto a uniform $(0, 1)$ scale. It follows that

$$G_i(z) = \exp\{\theta_l N \int_0^1 \log F_{Y,i|X}(z - F^{-1}_{tide,i}(x)|X^* = x)dx\}. \tag{4.3.3}$$

Differentiation of the expression for the annual maximum sea-level distribution function gives the density, $g_i(z)$, which is given by

$$g_i(z) = G_i(z)\theta_l N \int_0^1 \frac{f_{Y,i|X^*}(z - F^{-1}_{tide,i}(x)|X^* = x)}{F_{Y,i|X^*}(z - F^{-1}_{tide,i}(x)|X^* = x)}dx, \tag{4.3.4}$$

where $f_{Y,i|X^*}$ is the density, for year i, of the surge level conditional on the transformed tidal level being X^*.

When the density $g_i(z)$ is evaluated at z_i, the observed annual maximum in year i is the likelihood contribution from year i when the only data available in that year is the annual maximum. The likelihood for the annual maximum data is given by

$$\prod_i g_i(z_i), \tag{4.3.5}$$

where the product is over those years which have annual maximum data.

Now consider the statistical models which determine $F_{Y,i|X^*}$ and $f_{Y,i|X^*}$. Based on the point process model, a natural model for $F_{Y,i|X}$ above the threshold is

$$F_{Y,i|X}(y|x) = \exp\left\{-\frac{1}{N\theta_s}\left[1-k\left(\frac{y-\mu(X,i)}{\sigma(X)}\right)\right]_+^{1/k}\right\} \quad \text{for } y \geq u(X,i), \tag{4.3.6}$$

where

$$\begin{aligned} \mu(X,i) &= \mu\tilde{b}(X) + \tilde{a}(X,i) \\ &= \mu\tilde{b}(X) + \tilde{a}(X,i_0) + \beta(i-i_0) \\ \sigma(X) &= \tilde{b}(X)\sigma. \end{aligned} \tag{4.3.7}$$

Below the threshold $u(X,i)$ the data are dense so a non-parametric estimator, $\tilde{F}_{Y,i|X}(y|x)$, of the conditional distribution function provides the best model. From these models for the distribution function the density function is derived by differentiation. There are a number of minor technical issues concerning the construction of the non-parametric estimator and the density estimator, a description of these is deferred until Section 4.7.

4.3.2 Immingham examples

Using all the annual maximum data series the likelihood function can be constructed, as in equation (4.3.5). This likelihood cannot be evaluated without knowledge of the the interaction functions $a(X,i_0)$ or $b(X)$. These functions are assumed known, and are taken to be be the estimated forms of Section 4.1.3. Similarly θ_s is fixed at its estimated value. Although hourly data are required to estimate these interaction functions, and θ_s, the data used differ from those used to construct the point process model. Holding the interaction functions fixed, the likelihood for the annual maxima is maximised with respect to (μ,σ,k,β). The resulting maximum likelihood estimates are given in Table 4.3.

Relative to the maximum likelihood estimates based on the point process model (see Table 4.2), the estimates of μ,σ and k are imprecise, showing the limited extent of information the historical annual maximum still water data contain in terms of the extreme surge parameters. The estimates of the trend parameter β has a smaller standard error

Parameter	Estimate	Standard Error
μ	7.355	1.016
σ	1.558	0.450
k	-0.0767	0.2243
β	4.336	0.635

Table 4.3: Maximum likelihood estimates and standard errors obtained using the annual maximum data alone for Immingham. Estimates are derived taking the interaction function, $a(X, i_0)$ as known. The units of the trend estimate are mm/yr.

than given by the point process data since the data cover a much longer span.

The reasons for the annual maximum distribution to vary from year to year are

- long term trends
- inter-annual variation in tidal levels.

As discussed above, Dixon and Tawn (1994) focussed on the first of these. The importance of the latter is seen in Figure 4.2. Figure 4.2 (a) shows the variation in the distribution of the tide, i.e. the variation of $F_{tide,i}$ with i, by considering the value of a fixed quantile of this distribution plotted over years. In particular, the 99% quantile of this distribution is plotted over a 19 year period for Immingham. Within this period the quantile value varies up to $15 - 20$cm. The nodal cycle of the tides induces an 18.61 year periodic behaviour in tidal quantiles, so the observed pattern of inter-annual variation is repeated over the years which are not shown on this figure.

The distribution function of the annual maximum still water level obtained by averaging the tidal distribution over a complete nodal cycle is given in Figure 4.2 (b), shown as the solid line. This is the distribution which Dixon and Tawn (1994) used to derive return levels, and is the obvious choice for use in design studies (see Section 4.6). Also shown on this figure are

1. the annual maximum curves given for specific years, corresponding to the years which give the lowest and highest 99% quantile in Figure 4.2 (a),
2. the observed annual maximum still water levels for Immingham, plotted as $\diamond$ or $\times$ depending on whether the 99% tidal quantile for that year is below or above the median 99% tidal quantile respectively.

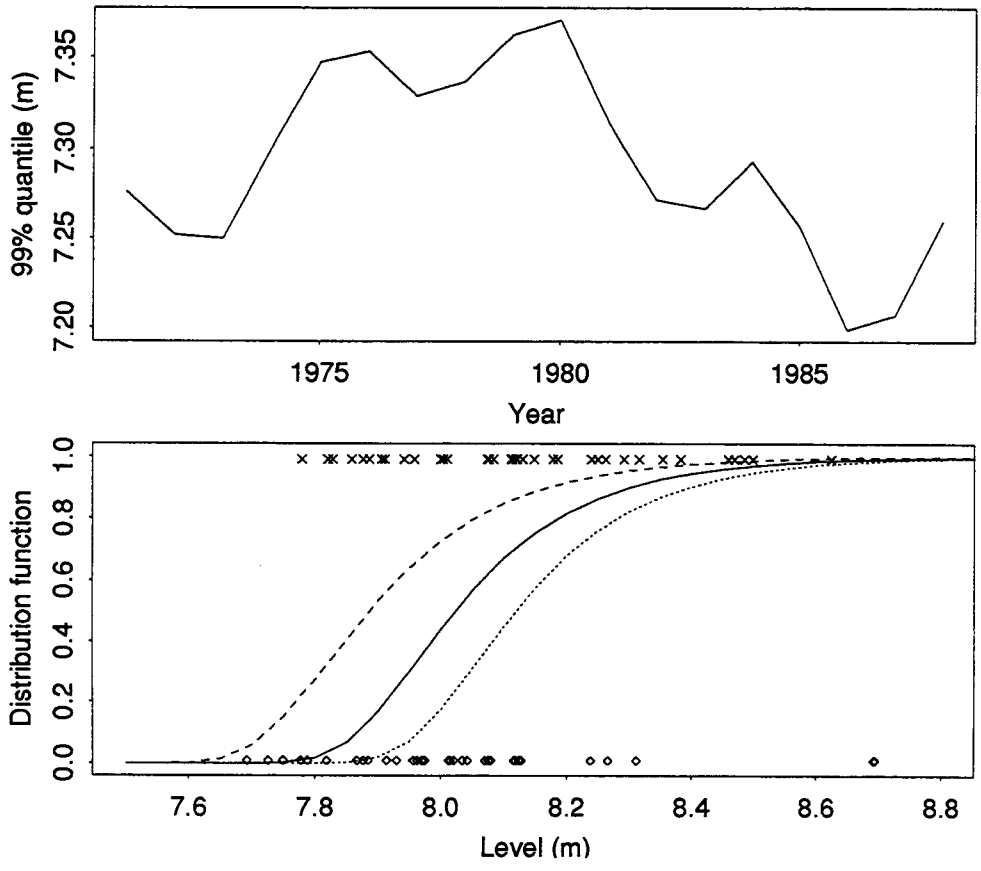

Figure 4.2: Annual variation in (a) 99% quantile of the tidal distribution, and (b) the annual maximum distribution function in 1990 using the lowest and highest annual tides (— — —) and (- - - -) respectively. The solid line is the annual maximum distribution based on 19 year average tidal distribution. Observed annual maxima, corrected to the year 1990, are shown as diamonds and crosses where they occur in low or high tidal years respectively.

To aid comparison of all the features on the plot each aspect has been corrected for the trend and so are all given to a common year, 1990. A number of important inferences can be drawn from this figure:

- The data points show that, to a large extent, years with the largest 99% tidal quantile produce the largest annual maximum still water levels.
- The shape of the distribution function of the annual maximum still water level for the different years is basically the same in spite of the different tidal characteristics in each year. The only real change corresponds to a shift of $20 - 30$cm, which is similar to the difference in the 99% tidal quantile for those years. Therefore, the variation in the distribution function of the annual maximum sea-level caused by the tidal component can largely be explained by a linear translation derived from the value of the 99% tidal quantile in any particular year.

- From the solid curve the annual maximum levels are seen to vary primarily in the range $7.8 - 8.4$m, i.e. an inter-annual variation of 0.6m. However 0.2m of this range is explained by year-to-year variations in the tide. Using the traditional annual maximum or r-largest methods this variation is incorporated into the procedure for extrapolation to long return levels. By comparison the JPM and RJPM exploit a knowledge of the separate tidal and surge distributions in the extrapolation process. As the extreme value theory methods used for extrapolation rely on the assumption of stationarity from year to year they are much more appropriate for the RJPM than the r-largest method. This is further evidence to support Dixon and Tawn (1994) preference for the use of the RJPM (and the JPM) over the r-largest method.

- If the 18.61 year average tidal distribution is used then the probability of occurrence of some of the observed data is zero. This feature is removed by using a tidal distribution which varies with year.

4.4 Mean sea-level inclusion

Just as historical annual maximum data reinforce the statistics derived from short term hourly data, annual mean sea-levels can be used to provide additional information of the trend in the surge process. Mean sea-level estimates are averages of many observations over the year and have considerably less inter-annual variability than either surge quantile data, $A_{i,j}^{(1)}$ of Section 4.1.3, or annual still water level maximum data of Section 4.3.2. This extra information is particularly important because we have made the critical assumption (Section 3.3) that any trend in the surge process is common throughout all aspects of the process. Hence mean sea-level data are the most informative data for the trend.

Let M_i denote the realisation of the mean-sea level variable in year i. Then we assume that

$$M_i \sim N(\alpha^* + \beta(i - i_0), e^2),$$

i.e. the mean sea-level follows a linear trend, of gradient β, with a Gaussian error distribution with variance e^2. Here, in the absence of other information about β, the best estimate of β comes from a standard regression analysis of the mean sea-level data. We consider further the use of mean sea-level data in one of the spatial analysis case studies given in Chapter 8.

4.4.1 Immingham Example

Using the mean sea-level data alone for Immingham an estimator of β can be obtained by simple regression. Table 4.4 gives the estimate and its standard error. Note that the

estimate is very similar to that from the point process model, which is based on only extreme surge data, so is consistent with the assumption of a common trend throughout all aspects the surge series.

Parameter	Estimate	Standard Error
β	1.513	0.632

Table 4.4: Maximum likelihood estimate and standard error for the trend (mm/yr) obtained using the annual mean sea-level data alone for Immingham.

4.5 Combined inference and model

As shown in the previous sections, information about the parameters of the extreme sea-level process is contained in data of different forms:

1. **Hourly data**: From a continuous period of hourly data, information about the conditional surge quantiles is obtained and provides the basis for estimation of the interaction functions $a(X, i)$ and $b(X)$. Furthermore, extreme surge data can be extracted in a declustered format for use in the point process method. The extreme surge data provide information about μ, σ, k and β and the interaction functions $a(X, i_0)$ and $b(X)$ which must be estimated, as they are not known in practice.

2. **Annual maximum sea-level data**: For many sites, in addition to periods of hourly data, historical data in the form of the observed annual maximum still water level are sometimes available. These data provide information about the extreme surge parameters μ, σ, k and β provided information is available to give a reasonable estimate of the interaction functions $a(X, i_0)$ and $b(X)$.

3. **Annual mean sea-level data**: In addition to periods of hourly data, for some sites annual mean sea-level data are available. These data provide information about the trend parameter, β, only.

4.5.1 Likelihood

The availability of these different different types of data varies from site-to-site. As this report focusses on the A-class sites all sites in this study have at least a year or more of hourly data. If hourly data are available for a particular year then it is possible to extract the annual mean sea-level and the annual maximum still water level. Annual maximum

still water level data are only considered for years in which hourly data are not available for the site, but mean sea-level data are considered for both years with hourly data and for the extra years, for which mean sea-level data are independently available.

It is assumed that declustered extreme surges, surge quantile estimates, and mean sea-level data are independent. The justification for this assumption is that the distribution of extreme values, intermediate quantiles and mean/median values are independent for large sample sizes (David, 1981). A consequence of this independence property is that the likelihood function for $\{a(X, i_0), \mu, \sigma, k, \beta\}$ is the product of the separate likelihood contributions for the three data forms for that year.

In a year which has no hourly data but only data for the mean-sea level and/or the annual maximum still water level, the likelihood function for such a year is the product of the appropriate likelihood contributions for the two data forms.

Let I_i, for $i = 1, 2, 3$ be the set of years with hourly data, mean sea-level data and the additional annual maximum still water level data respectively. From the above discussion $I_1 \subseteq I_2$ and $I_1 \cap I_3 = \emptyset$, i.e. there are at least as many years with mean sea-level data as hourly data and there is no overlap between years of hourly data and the additional annual maximum still water level data. If follows that the overall likelihood is

$$\left\{\prod_{I_1} L_{(\text{int})} L_{(\text{pp})}\right\} \left\{\prod_{I_2} L_{(\text{msl})}\right\} \left\{\prod_{I_3} L_{(\text{max})}\right\} \tag{4.5.1}$$

where $L_{(\text{int})}, L_{(\text{pp})}, L_{(\text{msl})}, L_{(\text{max})}$ are the likelihood functions for the interaction functions, the point process model, the mean sea-level and the annual maximum still water level respectively.

Estimation of $\{a(X, i_0), \mu, \sigma, k, \beta\}$ is then based on maximum likelihood estimation within the overall likelihood. This process of maximisation leads to different estimates than is given by the separate likelihood functions considered earlier; for example information about the trend is available from each of the different data formats. The resulting estimate of the trend is a mixture of the separate trend estimates and in turn leads to modified estimates of the other parameters. Since a trend estimate based on all the relevant data is much more accurate than that from any single data format improved estimates of the other parameters are obtained. This feature is particularly important for the point process model as the extreme surge parameters have most influence on the extrapolation to long return periods. This transfer of information from one data type to another is illustrated below based on the Immingham data.

4.5.2 Immingham example

Rather than immediately giving the estimates based on the information from the full data, we build up the likelihood from the different sources in stages, to best show how

the information is transferred between data formats. Table 4.5 shows estimates of the extreme surge parameters obtained by using the data for the point process model and the historical annual maximum still water levels. In this analysis the estimated base year interaction function, $a(X, i_0)$, is taken as known. The estimates differ from those given

Parameter	Estimate	Standard Error
μ	6.699	0.322
σ	2.213	0.218
k	-0.1221	0.0596
β	2.431	0.554

Table 4.5: Maximum likelihood estimates and standard errors obtained using the annual maximum data and point process model for Immingham. Estimates are derived taking the interaction function, $a(X, i_0)$ as known. The units of the trend estimate are mm/yr.

by the separate data forms, shown in Tables 4.2 and 4.3. In particular, the trend is a weighted average of the two separate trend estimates, with weights determined, in part, by the relative number of years of data of the two data forms. A number of features arise from having this combined trend estimate:

- Despite using data which are much more extreme than the conditional surge quantile data in the interaction function analysis, the resulting trend estimates are similar.
- The resulting trend estimate has a smaller standard error than either of the two separate estimates.
- The μ, σ and k parameter estimates, and their associated standard errors, are only slightly changed from the estimates given purely by the point process model. This shows that the annual maximum data provide only limited information about these extreme surge parameters.

Now consider a combined analysis of the interaction data, the point process data, and the annual maximum still water level data. Estimates of the parameters are given in Table 4.6. Again we first focus on the trend estimate. This falls between the estimate obtained in Table 4.5 and the estimate from the interaction analysis (see Table 4.1). The other features to note from this analysis are

- The standard error of the estimate of β is smaller than from either of the separate sources in this analysis.

Tidal band	Interaction function in base year, $\hat{\alpha}_j$	$se(\hat{\alpha}_j)$	Standard deviation, $\hat{c}_j$	$se(\hat{c}_j)$
1	0.489	0.0168	0.0895	0.0123
2	0.499	0.0148	0.0780	0.0107
3	0.489	0.0135	0.0710	0.0098
4	0.501	0.0170	0.0895	0.0123
5	0.473	0.0160	0.0842	0.0116
6	0.471	0.0143	0.0752	0.0103
7	0.466	0.0131	0.0693	0.0095
8	0.461	0.0135	0.0711	0.0098
9	0.420	0.0116	0.0613	0.0085
10	0.408	0.0120	0.0645	0.0089
location estimate, $\hat{\mu} = 6.479$, with $se(\hat{\mu}) = 0.313$				
scale estimate, $\hat{\sigma} = 2.098$, with $se(\hat{\sigma}) = 0.210$				
shape estimate, $\hat{k} = -0.1117$, with $se(\hat{k}) = 0.0602$				
trend estimate, $\hat{\beta} = 2.659$mm/yr, with $se(\hat{\beta}) = 0.384$				

Table 4.6: Estimated interaction function and extreme surge parameters for Immingham: based on data of quantiles of conditional surge process, extreme surges and historical annual maximum still water levels.

- Estimates of both σ and k are almost exactly the same as those from the joint point process and annual maximum analysis.
- There is a slight trade-off between the estimated values of μ and $a(X, i_0)$, this is seen to be reasonable from equation (4.3.7).

So far the analyses have excluded the mean sea-level data which provide information about the trend only. In Table 4.7 estimates of the parameters based on all the relevant data are given. As expected the main impact of including the mean-sea level data is a slightly lower estimate of the trend level, i.e. a trend which is more consistent with the mean sea-level data alone, and a reduction in the standard error of this estimate. Other parameters are barely influenced by these data being included. Since the historical mean

Tidal band	Interaction function in base year, $\hat{\alpha}_j$	$se(\hat{\alpha}_j)$	Standard deviation, $\hat{c}_j$	$se(\hat{c}_j)$
1	0.489	0.0170	0.0900	0.0124
2	0.500	0.0148	0.0780	0.0107
3	0.489	0.0136	0.0712	0.0098
4	0.500	0.0171	0.0898	0.0124
5	0.472	0.0161	0.0847	0.0116
6	0.472	0.0143	0.0753	0.0103
7	0.466	0.0132	0.0695	0.0095
8	0.460	0.0135	0.0708	0.0097
9	0.419	0.0115	0.0608	0.0089
10	0.408	0.0121	0.0646	0.0089
location estimate, $\hat{\mu} = 6.446$, with $se(\hat{\mu}) = 0.310$				
scale estimate, $\hat{\sigma} = 2.087$, with $se(\hat{\sigma}) = 0.207$				
shape estimate, $\hat{k} = -0.1120$, with $se(\hat{k}) = 0.0598$				
trend estimate, $\hat{\beta} = 2.337$mm/yr, with $se(\hat{\beta}) = 0.349$				

Table 4.7: Estimated interaction function and extreme surge parameters for Immingham: based on data of quantiles of conditional surge process, extreme surges, historical annual maximum still water levels, and mean sea-levels.

sea-level data primarily also provided information about the trend it is helpful to see if these data are important once the information from the mean sea-level data is included. To see this the model was refitted this time omitting the annual maximum data. The resulting estimates are shown in Table 4.8. There we see the trend estimate is largely

unchanged, although the standard error is larger, but some of the point process parameters are slightly modified.

Tidal band	Interaction function in base year, $\hat{\alpha}_j$	$se(\hat{\alpha}_j)$	Standard deviation, $\hat{c}_j$	$se(\hat{c}_j)$
1	0.489	0.0170	0.0902	0.0124
2	0.500	0.0148	0.0780	0.0107
3	0.489	0.0136	0.0713	0.0098
4	0.500	0.0171	0.0900	0.0124
5	0.473	0.0161	0.0848	0.0117
6	0.472	0.0143	0.0753	0.0104
7	0.467	0.0132	0.0696	0.0098
8	0.461	0.0135	0.0707	0.0097
9	0.419	0.0115	0.0607	0.0084
10	0.407	0.0124	0.0646	0.0089
location estimate, $\hat{\mu} = 6.495$, with $se(\hat{\mu}) = 0.346$				
scale estimate, $\hat{\sigma} = 2.169$, with $se(\hat{\sigma}) = 0.257$				
shape estimate, $\hat{k} = -0.1343$, with $se(\hat{k}) = 0.0692$				
trend estimate, $\hat{\beta} = 2.229$mm/yr, with $se(\hat{\beta}) = 0.425$				

Table 4.8: Estimated interaction function and extreme surge parameters for Immingham: based on data of quantiles of conditional surge process, extreme surges, and mean sea-levels.

4.5.3 Simplifications to the method of fit

The method of fit used to obtain the results in Table 4.7 is not suitable for repeated application to all study sites due to the computational complications, such as slow convergence, involved in simultaneously fitting such a high dimensional parametric model. The model has 26 parameters, these being the 24 parameters shown in Table 4.7 and the intercept and standard deviation for the mean sea-level trend component of the model (these are not shown in Table 4.7 as these have no relevance on the extreme sea-level process). Within the overall model fit the additional information carried by the point process, annual maximum and mean sea-level data about the interaction functions is minimal, so the estimates and associated standard errors for the interaction functions are little changed from the fit based on just the quantile data. In particular, α_j, c_j and their respective standard errors, for $j = 1, \ldots, 10$, are almost unchanged from Table 4.1 to Table 4.7. This suggests

that a substantial simplification is possible to the fitting procedure. Two features which determine how this simplification is achieved are

- the profile of the interaction function in the base year is unchanged by additional information provided by data other than the quantiles, this suggests that the estimate, obtained in Section 4.1, for the relationship between X and $a(X, i_0)$ can used in fitting the overall model,
- the values of c_j, for $j = 1, \ldots, 10$ are of no relevance to the extreme sea-level process.

This suggests estimating both $a(X, i_0)$ and c_j using the interaction function analysis of Section 4.1, and subsequently treating these parameters as fixed, i.e. $\tilde{a}(X, i_0)$ and $\tilde{c}_j$. The overall model is then fitted as described in Section 4.5 with

$$a(X, i_0) = \alpha + \tilde{a}(X, i_0). \tag{4.5.2}$$

Here the interaction function is taken to have the same profile with the tide, as given by the interaction function analysis, i.e. $\tilde{a}(X, i_0)$, but has an additional parameter, α, which allows for this profile to be shifted in level. The reason for including α is that in a regression analysis the trend and intercept parameters often exhibit dependence, so if the base year interaction level was fixed at a level based on a poor trend estimate given by the interaction function analysis this would bias the overall trend estimate obtained using potentially more informative data such as historical mean and annual maximum sea-level data. Generally though we expect $\tilde{a}(X, i_0)$ to be a good estimate of $a(X, i_0)$ so α should be small. Table 4.9 gives estimates obtained by applying this simplified model. Clearly the estimates and standard errors for these parameters are almost identical to

Parameter	Estimate	Standard Error
α	9.8×10^{-5}	2.6×10^{-3}
μ	6.451	0.308
σ	2.086	0.207
k	-0.1105	0.0595
β	2.339	0.347

Table 4.9: Maximum likelihood estimates and standard errors obtained using Immingham data with the profile of interaction function and the $\{c_j : j = 1, \ldots, 10\}$ fixed. The units of the trend estimate is mm/yr and for α is metres.

those in Table 4.7 from the full analysis. In this case the estimate of α is very small

as the trend estimate from the interaction function analysis of 2.917mm/yr is not too different from the overall trend estimate. These results confirm that the simplification of the fitting procedure is justified as it produces almost identical results to the full analysis yet it removes the substantial computational problems encountered previously.

In conclusion the fitting procedure we adopt is a two stage process:

Stage 1 involves fitting the interaction functions as described in Section 4.1. This is a 21 parameter maximisation, but since the parameters are nearly orthogonal convergence is very quick.

Stage 2 involves fixing the profile and standard deviation estimates from Stage 1 before fitting the overall model using (4.5.1), with $a(X, i_0)$ given by (4.5.2). This is a 7 parameter maximisation, so is also relatively straightforward.

4.6 Design levels

The design of coastal flood defences is based on criteria which require that the probability of flooding in a year is a specified level p. We have demonstrated that the distribution of the maximum still water level varies from year to year due to

1. long term trends
2. inter-annual variations in tide.

Dixon and Tawn (1994) discuss methods for incorporating trends into the design criteria. The simplest presentation of information is to specify the design level for a particular year and the estimated trend. From these the estimated design level in any year can be calculated.

The second feature of yearly variation is not discussed in Dixon and Tawn (1994), but the approach they use is the most natural to account for such variations. Suppose that there was no trend. From Section 4.3 it follows that the distribution of the annual maximum still water level in year i is

$$G_i(z) = \left[\prod_{j=1}^{N} F_{Y|X}(z - X_{t_{j,i}} | X = X_{t_{j,i}}) \right]^{\theta_l}. \tag{4.6.1}$$

Since still water level maxima from year to year are independent, it follows that the distribution of the maximum still water level over a nodal tidal cycle of 18.61 years (taken as 19 years) starting from year i is

$$G_i(z) G_{i+1}(z) \ldots G_{i+19}(z). \tag{4.6.2}$$

Yearly variations in the distribution of the annual maximum still water level which are due to the tide, are not important from design considerations so long as structures are designed to last at least one nodal cycle. Thus we can view this as the distribution of the maximum of 19 independent replications of a common yearly annual maximum design variable, with distribution $G(z)$ in year i. Then

$$[G(z)]^{19} = G_i(z)G_{i+1}(z)\dots G_{i+19}(z), \tag{4.6.3}$$

and so

$$\begin{aligned} G(z) &= [G_i(z)G_{i+1}(z)\dots G_{i+19}(z)]^{1/19} \\ &= \exp\{\theta_l N \int_0^1 \log F_{Y,i|X}(z - F_{tide}^{-1}(x)|X^* = x)dx\}. \end{aligned} \tag{4.6.4}$$

where the notation is as in equation (4.3.3).

From equation (4.3.1) we see that design levels will depend on the conditional surge distribution, for which a parametric model and a non-parametric model determines the form above and below a high threshold respectively. Models for the two components are discussed in Section 4.7.

4.7 Technicalities of the methods

This section contains a brief discussion of some of the technical steps of the methods discussed throughout Chapter 4. These steps are of limited interest, or relevance, for understanding the methods used but are important in the numerical evaluation and application of the methods.

4.7.1 Interaction functions

Throughout the development of the interaction functions $a(X, i)$ and $b(X)$, the X variable denotes the tide, and the domain of these functions is from the lowest astronomical tide to the highest astronomical tide. Consequently this range changes from site-to-site as the tidal characteristics differ. It is therefore difficult to compare interaction functions from one site to another. To remove this difficulty, and to enable spatial modelling of the interaction functions, the interaction functions are considered in the form $a(X^*, i)$ and $b(X^*)$ where X^* is a transformation of X given by

$$X^* = \tilde{F}_{tide}(X),$$

where $\tilde{F}_{tide}$ is an empirical estimate of the tidal distribution function. Thus X^* is the tide transformed monotonically onto a $[0,1]$ range. This range is identical for all sites so the interaction functions can be spatially interpreted. This scale was previously found helpful when deriving a smooth estimate of the interaction functions over the entire tidal range.

From now on we will use X and not distinguish between X and X^*, as these variables simply map from one to the other, and the range of the variable shows which is being used. Most frequently the transformed tide will be used for the reasons discussed above.

4.7.2 Non-parametric estimation of the conditional surge density

The distribution of the annual maximum still water level from which design levels are estimated depends on the distribution function of the surge conditional on the associated tidal level, see equation (4.6.4). Therefore indirectly it depends on the density of the conditional surge both above and below the threshold. Furthermore, the density of the annual maximum still water level, required for the inclusion of annual maximum data (see Section 4.3) explicitly depends on this density.

In Section 4.2, the model for the density above the threshold $u(X,i)$ is parametric but below the threshold no explicit non-parametric model was given. Here the non-parametric estimate of the density $f_{Y,i|X}$, i.e. the density of the surge in year i conditional on the tidal level X is described. In the estimation of this density the whole range of surges are considered, however for applications this estimate will only be applied for surges below the threshold $u(X,i)$ in year i for an associated tidal level of X, as above this threshold the parametric model is used.

We need an estimate of this density for years with hourly data, years for which design levels are required, and years with annual maximum still water level data only. Therefore, generally there will be insufficient data in a year upon which to base a non-parametric estimate of this density. The approach used is to pool information over years. This is not a direct step as the density in year i, $f_{Y,i|X}$, varies linearly with year i and is given by

$$f_{Y,i|X}(y|X) = f_{Y,i_0|X}(y - \beta(i - i_0)|X), \tag{4.7.1}$$

where i_0 is a base year and β is the trend. This suggests that a suitable approach is to de-trend all the data to the base year, and pool the resulting data to non-parametrically estimate the surge density conditional on the tide for that base year, i.e. obtain $\tilde{f}_{Y,i_0|X}$, then use

$$\tilde{f}_{Y,i|X}(y|X) = \tilde{f}_{Y,i_0|X}(y - \beta(i - i_0)|X), \tag{4.7.2}$$

to provide the density estimate for year i. There is one obvious complication with this method, namely the use of the trend, β. The trend parameter is unknown and is itself

one of the key parameters to be estimated. An initial estimate of the trend β_{init} must be used in the steps above. Given the relatively narrow span of hourly data any reasonable estimate for β_{init} is sufficient here.

4.7.3 Non-parametric estimation of the conditional surge distribution

The surge distribution function for year i conditional on the tidal level being X is $F_{Y,i|X}$. As shown in Section 4.2 a parametric model exists for this function when the surge is above the threshold, but below the threshold a non-parametric estimator must be used. Using the definition that

$$F_{Y,i_0|X}(y|X) = \int_{-\infty}^{y} f_{Y,i_0|X}(z|X)dz,$$

the distribution function in the base year can be estimated by integrating the non-parametric estimate of the density obtained in Section 4.7.2, so that

$$\tilde{F}_{Y,i_0|X}(y|X) = \int_{-\infty}^{y} \tilde{f}_{Y,i_0|X}(z|X)dz.$$

It follows that

$$\tilde{F}_{Y,i|X}(y|X) = \tilde{F}_{Y,i_0|X}(y - \beta_{init}(i - i_0)|X).$$

Two final complications are that once the non-parametric density estimate $\tilde{f}_{Y,i|X}$ below the threshold is combined with the parametric model above the threshold the resulting estimate may not integrate to one exactly, and that $\tilde{F}_{Y,i|X}$ may not tie up perfectly with the parametric estimate of the conditional surge distribution at the threshold. A slight refinement is needed in each case. Letting

$$\hat{F}_{Y,i|X}(y|X) = \frac{F^{(p)}_{Y,i|X}(u(X,i)|X)}{\tilde{F}_{Y,i|X}(u(X,i)|X)} \tilde{F}_{Y,i|X}(y|X) \text{ for } y \leq u(X,i), \tag{4.7.3}$$

where $F^{(p)}_{Y,i|X}$ is the parametric estimate of the distribution. It follows that the refined non-parametric estimate $\hat{F}_{Y,i|X}$ and the parametric estimate $F^{(p)}_{Y,i|X}$ are continuous at the threshold. Furthermore differentiating equation (4.7.3) gives

$$\hat{f}_{Y,i|X}(y|X) = \frac{F^{(p)}_{Y,i|X}(u(X,i)|X)}{\tilde{F}_{Y,i|X}(u(X,i)|X)} \tilde{f}_{Y,i|X}(y|X) \text{ for } y \leq u(X,i), \tag{4.7.4}$$

which, when combined with the parametric model above the threshold, integrates to one.

Part II

Study site results using site-by-site based models

Chapter 5

Site-by-site results

In this section we apply four site-by-site methods for estimating return levels, these being

1. the new Spatial Revised Joint Probability Method (SRJPM),
2. the Revised Joint Probability Method (RJPM),
3. the Joint Probability Method (JPM),
4. the r-largest method.

The first method is developed in Part I of this report and the last three methods are as given by Dixon and Tawn (1994). The three existing methods are applied to 26 of the current study sites in Dixon and Tawn (1994). As the hourly data used for the previous analysis is identical to that used here, the estimates are as given in the previous report. Exceptions are Immingham and Newlyn where some errors were found in the previous data handling technique so the methods have been repeated thus providing more correct estimates.

For the additional study sites the three existing methods are applied as is the SRJPM to all study sites. Figures 5.1-5.7 show the four estimates, to ACD, for each site.

5.1 New sites

First we examine estimates for the sites not considered in Dixon and Tawn (1994) which were obtained by analysis using the existing methods. Of these sites Harwich and Walton have over 10 years of hourly data; Newhaven and Swansea have between 5-10 years of hourly data; and the 15 remaining sites have less than 5 years of hourly data. The recommendations made in Dixon and Tawn (1994) suggest using the RJPM for Harwich and

Walton, depending on the estimates using either of the RJPM or the JPM for Newhaven and Swansea, and using the JPM for the 15 remaining sites. The RJPM appears reasonable for both Newhaven and Swansea as well as Harwich and Walton. In some cases the RJPM gives reasonable results for sites with less than 5 years of data, such as Leith and Portsmouth, and adequate results considering the data length for Hinkley, Newport, Mumbles, Millport and Islay. For the other short record sites the RJPM gives a poor estimate. Of the existing methods we recommend that the JPM should always be used when less than 5 years of hourly data are available.

5.2 Results for the SRJPM

Unlike the existing RJPM this method exploits historical data in the form of annual maximum still water levels and annual mean sea-levels, as well as better utilising the hourly data when estimating the trend and extremal parameters which influence the tail of the annual maximum distribution. The method is applied to all sites as shown for the example site of Immingham in Part I. The statistical analysis provides parameter estimates and standard errors which are easily converted to return levels with associated standard errors using the methods described in Section 4.6. The resulting return level estimates are shown on Figures 5.1-5.7. This method did not work for all sites as the fitting algorithm failed to converge when there were insufficient data to identify all aspects of the model as it is highly structured. Sites where no estimates are obtained are Port Erin, Workington and Tobermory.

The SRJPM is computationally slow when there are a large number of additional historical annual maximum values for the site. To speed up the fitting procedures the annual maximum data from the 1800's were omitted from the analysis. This reduced the information available for only three of the sites, Sheerness, Portsmouth and Avonmouth. The loss of information by this exclusion of data is minimal as they are of limited additional value once the other annual maxima are included into the analysis.

5.2.1 Long record sites

For sites with over 10 years of hourly data the SRJPM estimate is generally similar to the existing RJPM estimate despite the introduction of the additional information provided by the annual maximum and mean sea-level data. Immingham is a case where little change occurs in the return level estimates despite the addition of 69 years of annual maximum values and 31 years of mean sea-levels. In such cases the gain in using the historical data is mainly in terms of improved trend estimation, as is seen in Section 5.3.

For some sites a noticeable change does occur. The change in the return level estimates for Wick, Fishguard, Ullapool and Lerwick is primarily due to the inclusion of mean sea-level data which has improved the trend handling. In each case there is a shift in the return level curve but its shape of the return level curve is unchanged. In the other cases the difference is due to a combination of the information from mean and maximum data. For Aberdeen, Lowestoft, Milford Haven and Heysham the resulting curve for the SRJPM is steeper than that from the RJPM, whereas for Walton, Southend and Sheerness it is shallower. However, even for these sites there is no statistical difference once standard errors are accounted for. At this stage the only basis we have for assessment is to see how they compare with estimates for neighbouring sites which have similar tidal characteristics. For Walton, using Harwich as a reference, the SRJPM estimate is better than the previous estimate in terms of the gradient of the curve. For Sheerness and Southend the SRJPM estimates are more consistent with each other than previously, so generally the SRJPM estimates appear to be an improvement on the RJPM estimates.

5.2.2 Sites with 5-10 years of hourly data

The return level estimates are changed little for Ilfracombe and Swansea, but for Whitby, Newhaven and Avonmouth the primary reason for the changes is the additional information about the trend which induces a shift in the return level curve. In each case the SRJPM estimate appears to be an improvement on both the RJPM and JPM.

5.2.3 Sites with 1-4 years of hourly data

If no additional historical data are available then generally the SRJPM does not give improved estimates relative to the RJPM, so the JPM is recommended for Cromer, Felixstowe, Weymouth, Hinkley, Mumbles, Barmouth, Port Erin, Workington, Millport, Islay, Tobermory and Kinlochbervie. Of the other short hourly record sites Leith, Portsmouth and Devonport are much improved by incorporating historical annual maximum and mean sea-level data. In these cases the SRJPM is recommended. Even though Newport has holdings of annual maximum data the SRJPM is poor so the JPM is recommended there. This leaves Liverpool which has historical data only in the form of 25 years of mean sea-levels. The SRJPM is a significant improvement on the RJPM and has a return level curve which is similar in shape to that for the JPM except much lower. It is not clear which method to use in this case as the mean sea-level data could be adjusting a biased trend in the JPM. Without further information, which in theory could be obtained from annual maximum analyses for neighbouring Merseyside sites (see Graff, 1981, and Coles and Tawn, 1990), we opt for the JPM estimates which are generally more reliable for

short data sets.

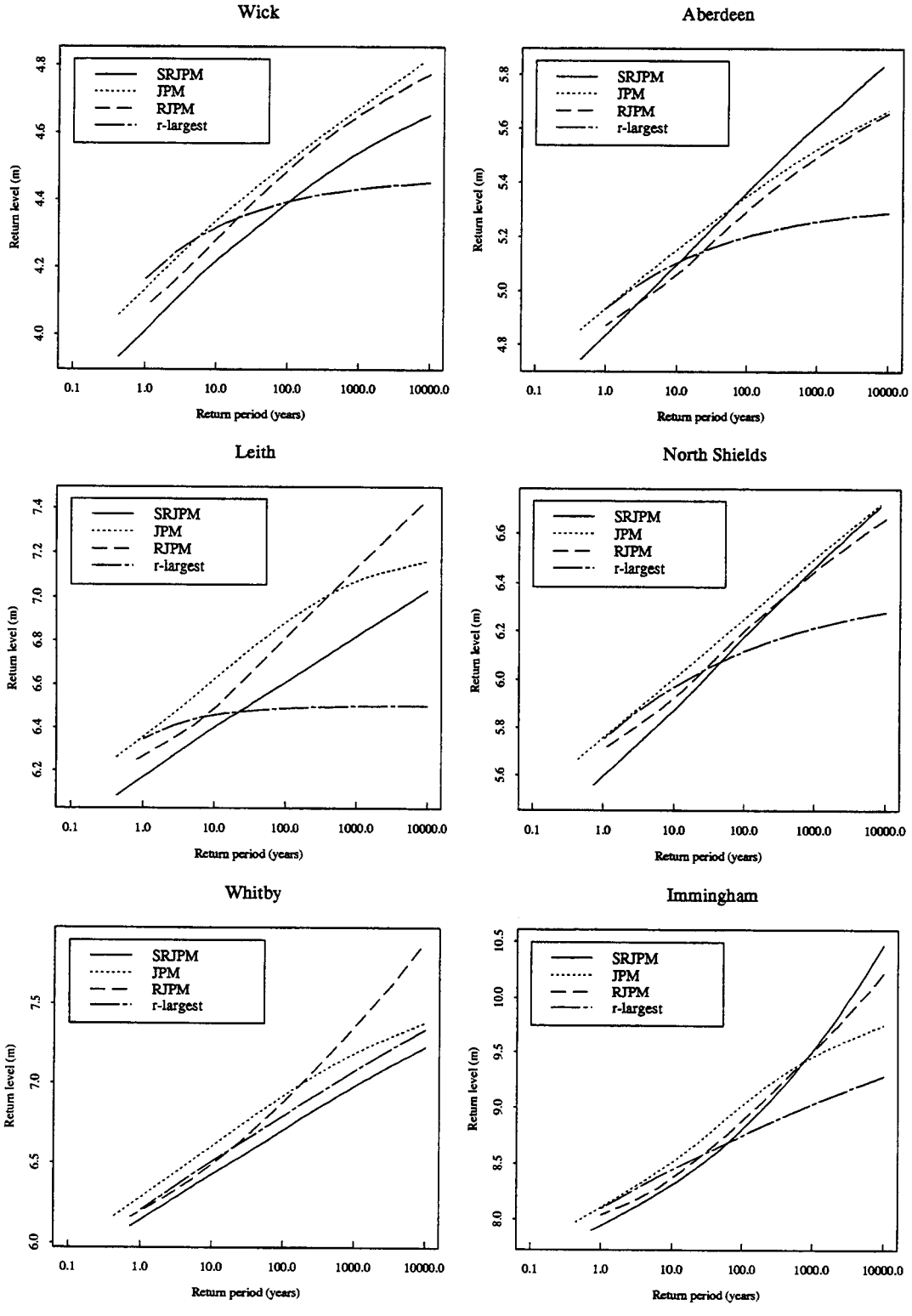

Figure 5.1: Return level plots (in metres relative to ACD) for 1990 for the 4 site-by-site methods: Wick to Immingham.

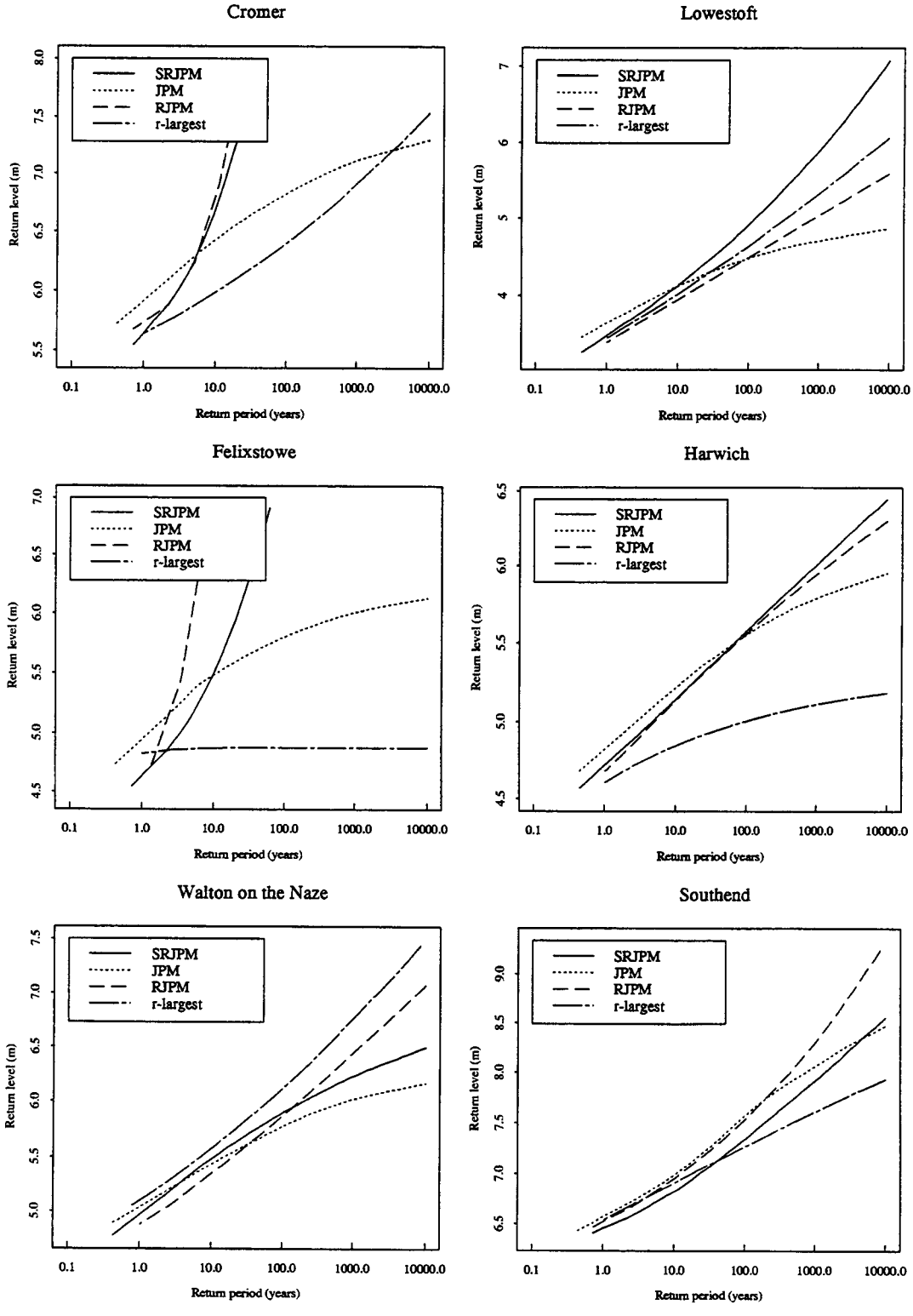

Figure 5.2: Return level plots (in metres relative to ACD) for 1990 for the 4 site-by-site methods: Cromer to Southend.

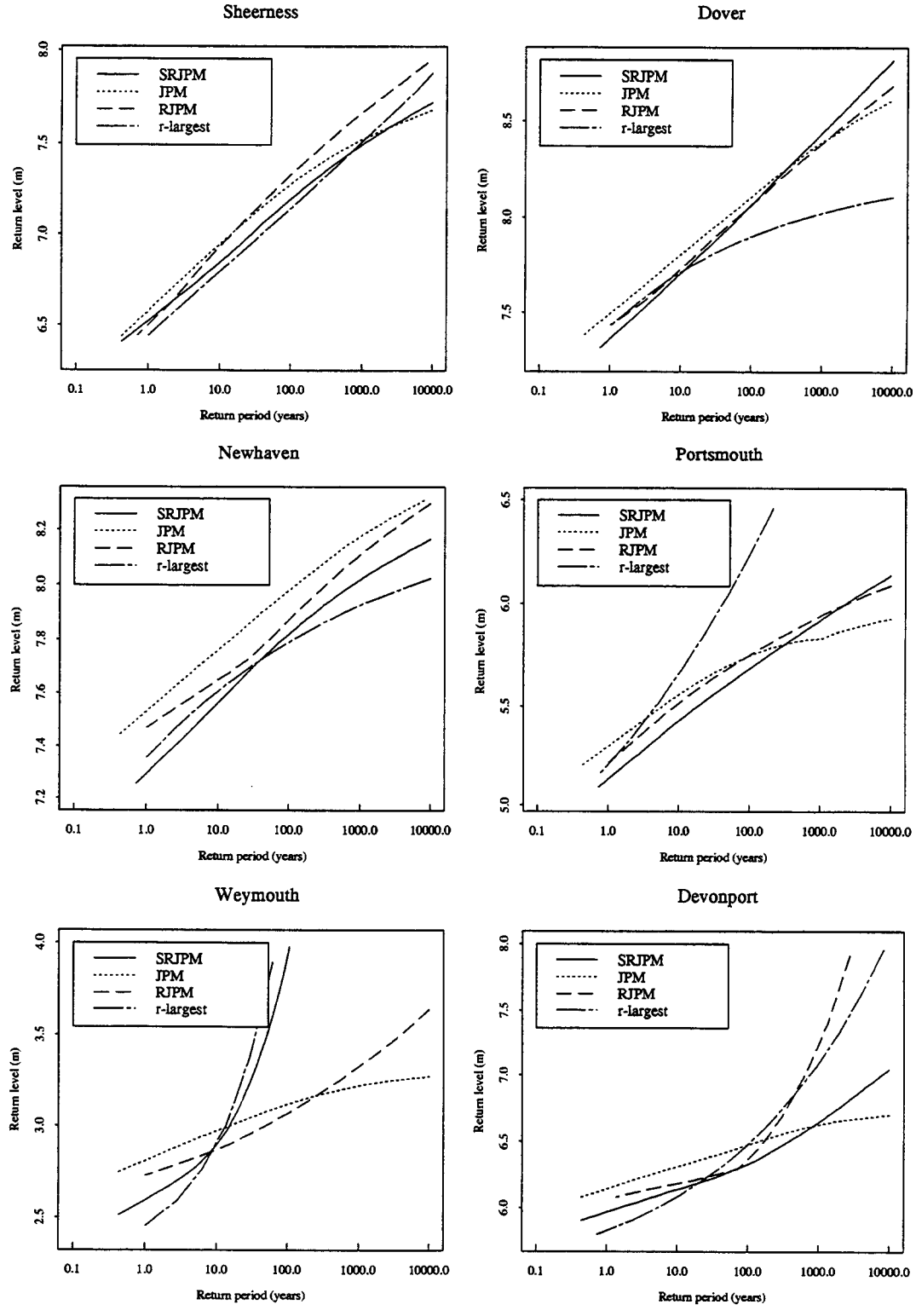

Figure 5.3: Return level plots (in metres relative to ACD) for 1990 for the 4 site-by-site methods: Sheerness to Devonport.

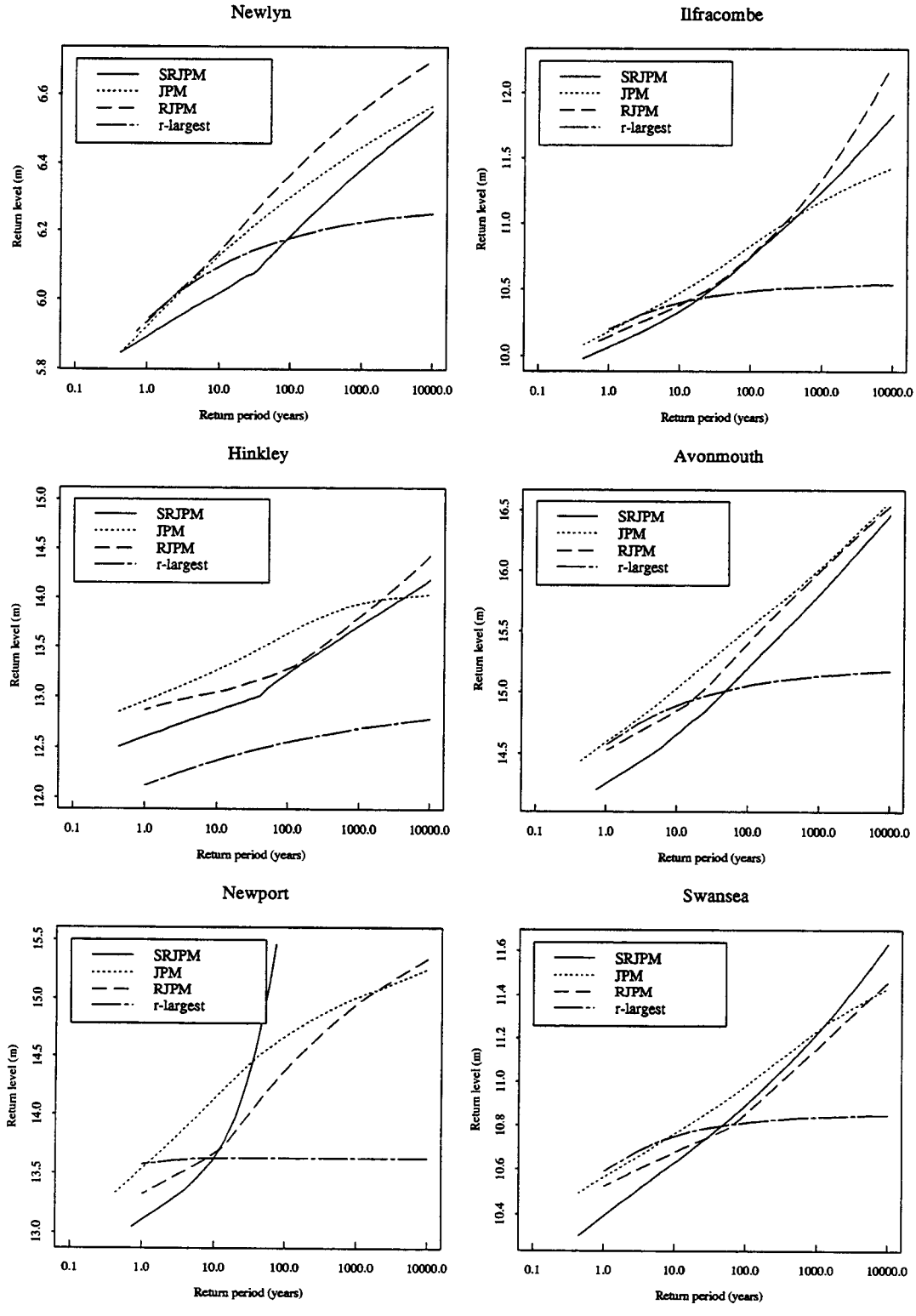

Figure 5.4: Return level plots (in metres relative to ACD) for 1990 for the 4 site-by-site methods: Newlyn to Swansea.

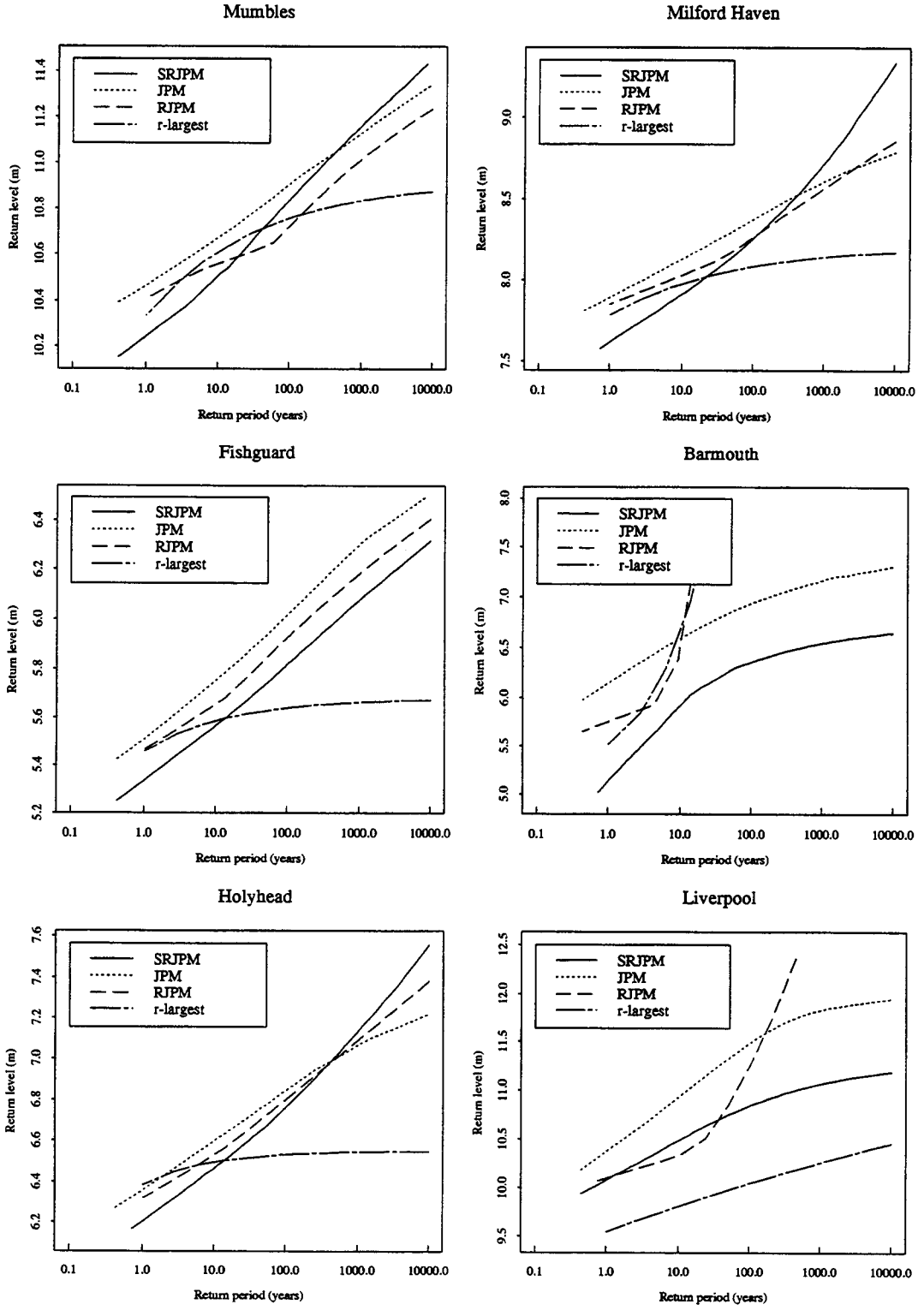

Figure 5.5: Return level plots (in metres relative to ACD) for 1990 for the 4 site-by-site methods: Mumbles to Liverpool.

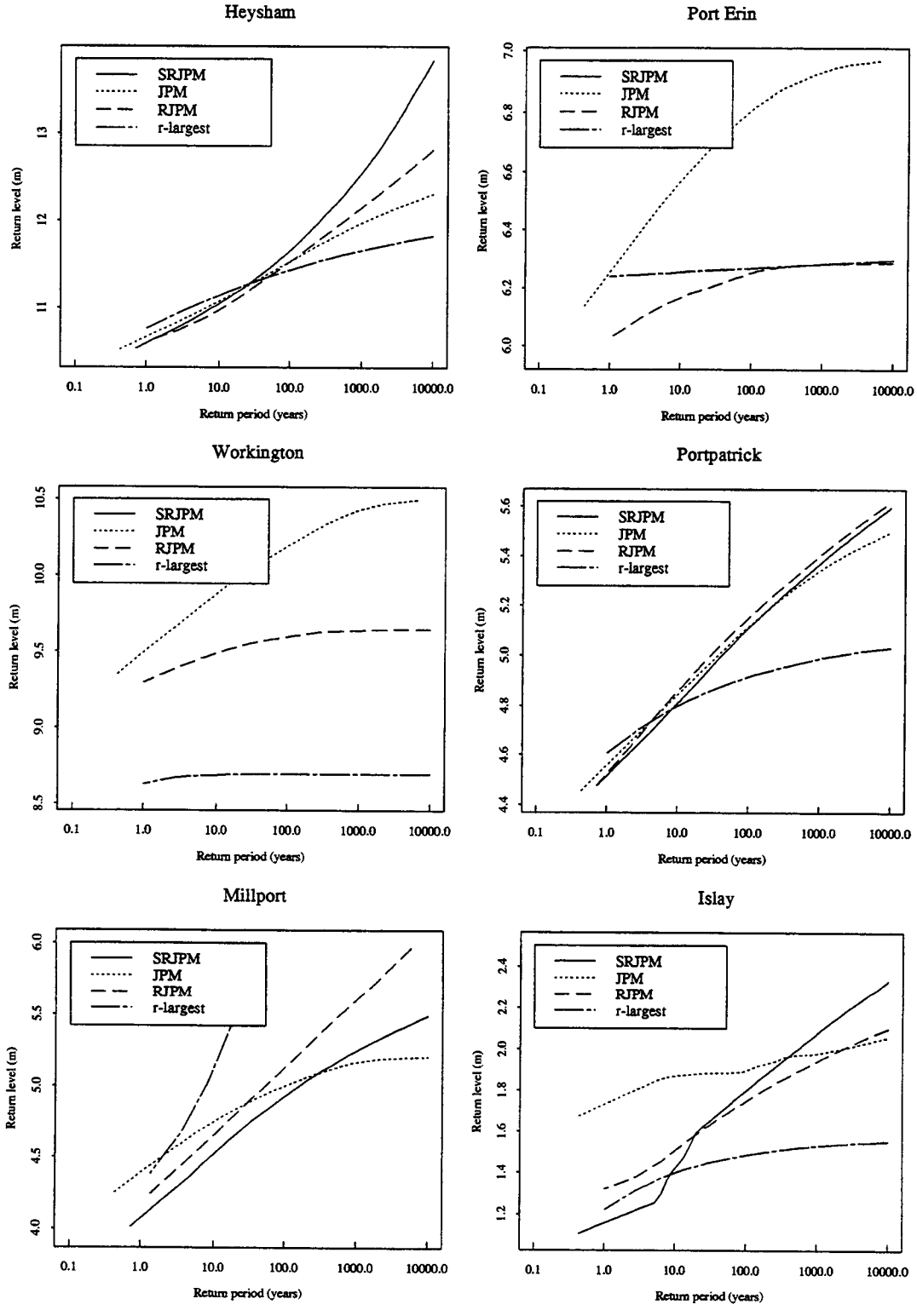

Figure 5.6: Return level plots (in metres relative to ACD) for 1990 for the 4 site-by-site methods: Heysham to Islay.

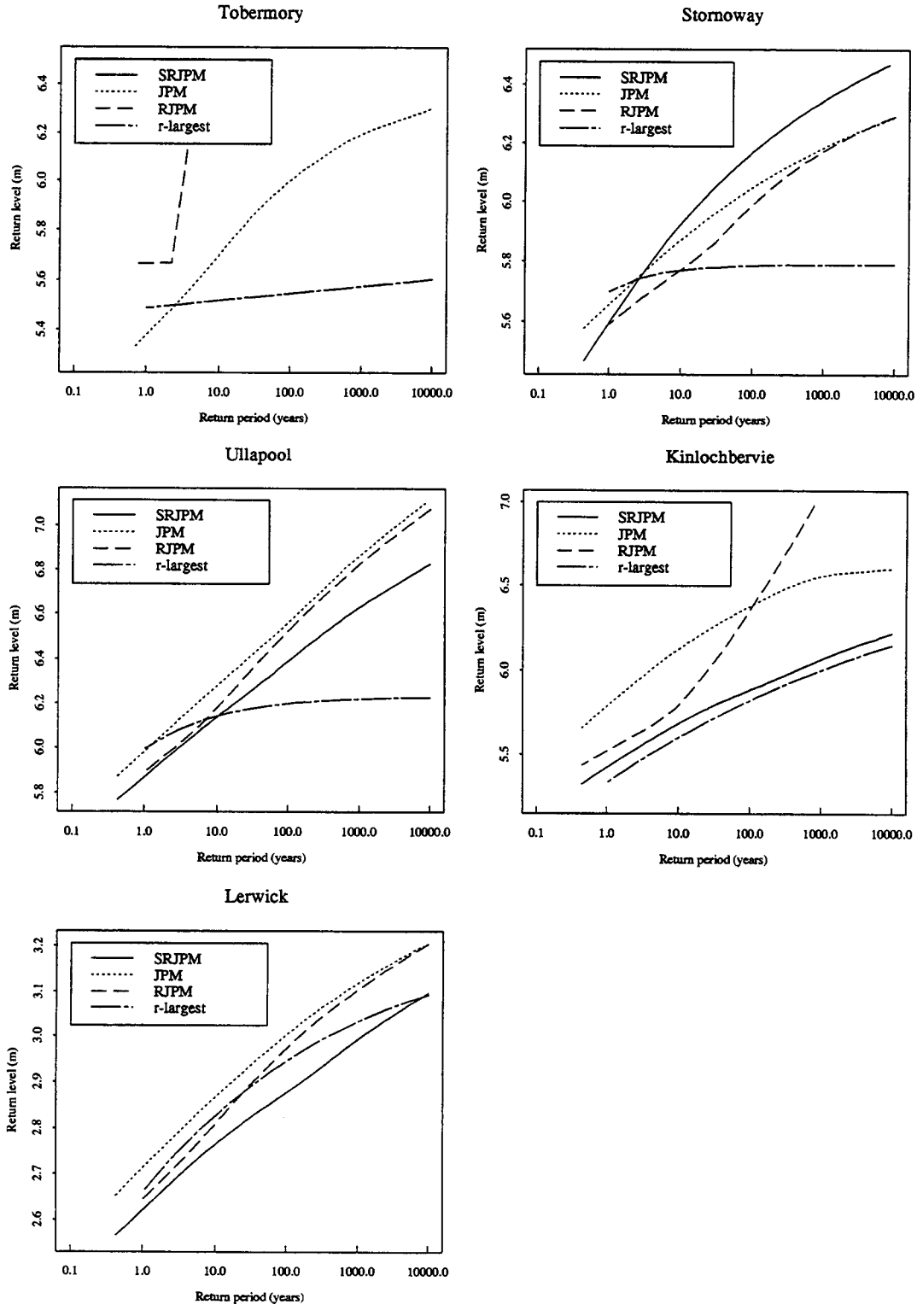

Figure 5.7: Return level plots (in metres relative to ACD) for 1990 for the 4 site-by-site methods: Tobermory to Lerwick.

5.3 Best estimates for each site

The best estimates for each site are given in two formats. Here we have tables for trends and return levels each with associated standard errors, and in the Appendix the return level curves for the method considered best at each site are given in greater detail to aid the evaluation of return levels for non-standard return periods.

First consider the best site-by-site estimates of the trend. As the JPM does not account for the trend, and the RJPM and r-largest method only use extreme values to estimate the trend the best estimate is always provided by the SRJPM. Table 5.1 shows the estimates from the SRJPM applied to sites which either have more than 5 years of hourly data, or have long historical records of annual mean sea-levels or maximum levels. The sites for which no reliable trend estimate can yet be given are Cromer, Felixstowe, Weymouth, Hinkley, Mumbles, Barmouth, Port Erin, Workington, Millport, Islay, Tobermory and Kinlochbervie. For such sites trend information must be inferred from neighbouring sites. In this report (Section 12.2) a spatial trend is developed for the whole east coast, providing trend estimates for both Cromer and Felixstowe. However no such estimate is available for the south and west coast sites listed above, so information about the trend in this region must be obtained from other sources, such as the spatial annual maximum extreme sea-level trend obtained in Chapter 7.

Of the trend estimates themselves there are a few notable features relative to the previous recommendations in Dixon and Tawn (1994):

- The trend estimates have standard errors which are generally one third the size of those given in Dixon and Tawn (1994), hence 95% confidence intervals for the trends are much tighter, giving greater confidence in using the trend to predict return levels in future years.

- North Shields and Sheerness have the largest reduction in standard error for the trend, in each case being about one tenth the standard error in Dixon and Tawn (1994).

- The trend estimate values are less spatially variable, with Sheerness, Avonmouth, Ilfracombe, Heysham, and Portpatrick over 2mm/year lower, and Ullapool 2mm/year higher, than previously.

- The trend estimate for sites with short hourly records but long historical data sets are generally very reliable. These sites are Leith, Newhaven, Portsmouth, Newport, Swansea and Liverpool. In contrast the methods of Dixon and Tawn (1994) give a very poor trend estimate for these sites as these historical data are excluded from the analysis.

Site	Trend, β	s.e.
Wick	3.851	0.4741
Aberdeen	0.585	0.1741
Leith	2.653	1.5743
North Shields	2.126	0.1454
Whitby	7.519	1.6910
Immingham	2.339	0.3471
Lowestoft	1.990	0.5507
Harwich	1.192	0.8346
Walton on the Naze	-0.093	1.6770
Southend	2.330	0.1911
Sheerness	1.712	0.0960
Dover	3.203	0.2771
Newhaven	2.395	0.3314
Portsmouth	-2.013	0.6687
Devonport	3.221	1.3217
Newlyn	2.830	0.2364
Ilfracombe	2.814	0.6511
Avonmouth	-1.166	0.6020
Newport	-0.964	0.3055
Swansea	3.020	0.5722
Milford Haven	-0.729	0.5253
Fishguard	1.549	0.3799
Holyhead	3.831	0.3439
Liverpool	2.683	0.8114
Heysham	4.205	0.9574
Portpatrick	7.909	0.7859
Stornoway	3.592	1.1567
Ullapool	6.143	0.7236
Lerwick	0.116	0.3003

Table 5.1: Site-by-site trend estimates (in mm/year) and standard errors (s.e.) for the study sites with either more than 5 years of hourly data or long historical data series. The estimates are from the SRJPM.

- The largest trend estimate is that for Portpatrick, however in Chapter 8, there is evidence to suggest that this trend is biased due to a datum shift.

Now consider return level estimates. In Section 5.2 the best method of extreme sea-level analysis was chosen for each site. For return periods of 10, 25, 50, 100, 250, 500, 1000, 10000 years return level estimates from the best method are given in Tables 5.2 - 5.5. The estimates are given relative to ACD for the year 1990. Estimates of return levels for years other than 1990 need to be adjusted for the trend at the site of interest. This adjustment is made as in Dixon and Tawn (1994, Section 8.5). These tables show the standard errors for the estimates when the SRJPM is the recommended approach. In cases where the JPM is the proposed method no standard errors are available, as signified by the NA in the table. There is no clear pattern of changes in the standard errors obtained under SRJPM from those under the SRJPM. However, if only those cases for which large changes in standard errors for the 10000 year return level are considered, a clearer picture emerges.

- For both Aberdeen and North Shields the SRJPM has a standard error of about twice the standard error for the RJPM, so is worse in each case. However, the previous estimates were very precise so the SRJPM standard errors may be more realistic.

- For Whitby, Lowestoft, Southend, Sheerness, Ilfracombe and Holyhead the standard errors for the SRJPM are more precise than for the RJPM, i.e. often less than half the standard errors from the RJPM.

- For Leith, Newhaven, Portsmouth, Devonport, Swansea and Stornoway standard errors are obtained for the SRJPM whereas using the methods of Dixon and Tawn (1994) only the JPM would have been recommended for these sites, and hence no standard errors would previously have been available.

In conclusion the standard errors for the SRJPM have been reduced in value from the values obtained using the RJPM for those sites which were imprecisely estimated using the RJPM. For those sites which the RJPM gave very small standard errors the SRJPM appears to give larger estimates, possibly better reflecting the true uncertainty. In general the bias as well as the variance of estimates is important. This is more difficult to calibrate but the better the model captures the known physics of the processes, the more confident we are that the bias is reduced. This again supports the use of the SRJPM.

For the sites on the south and west coasts of the UK the return level estimates given in this section are the best available. For the east coast there are alternative estimates obtained by a spatial analysis of the extreme sea-level data, that being the focus of the

Site	Method	10	s.e.	25	s.e.	50	s.e.	100	s.e.
Wick	SRJPM	4.21	0.009	4.29	0.013	4.33	0.015	4.40	0.019
Aberdeen	SRJPM	5.09	0.019	5.20	0.030	5.27	0.038	5.37	0.057
Leith	SRJPM	6.40	0.022	6.48	0.029	6.53	0.035	6.62	0.049
North Shields	SRJPM	5.87	0.024	5.99	0.039	6.06	0.051	6.18	0.076
Whitby	SRJPM	6.42	0.029	6.53	0.045	6.60	0.057	6.71	0.080
Immingham	SRJPM	8.30	0.024	8.48	0.041	8.60	0.056	8.82	0.085
Cromer	JPM	6.41	NA	6.59	NA	6.69	NA	6.83	NA
Lowestoft	SRJPM	4.11	0.090	4.40	0.120	4.59	0.140	4.93	0.174
Felixstowe	JPM	5.47	NA	5.62	NA	5.70	NA	5.81	NA
Harwich	SRJPM	5.13	0.075	5.31	0.120	5.41	0.155	5.59	0.226
Walton	SRJPM	5.46	0.074	5.65	0.103	5.75	0.125	5.90	0.169
Southend	SRJPM	6.81	0.034	7.01	0.052	7.14	0.064	7.35	0.084
Sheerness	SRJPM	6.83	0.026	6.98	0.038	7.07	0.044	7.20	0.056
Dover	SRJPM	7.69	0.018	7.84	0.029	7.93	0.038	8.08	0.059

Table 5.2: Return levels (in metres relative to ACD) for 1990 obtained using the best site-by-site methods: return periods are 10, 25, 50 and 100 years for the east coast sites.

remainder of the report. For east coast sites both site-by-site estimates and spatial model estimates are available, so a choice of approach has to be made. We will see in Part IV of this report that the spatial estimates for these sites are a considerable improvement over the site-by-site estimates, so that method should generally be used.

Site	Method	250	s.e.	500	s.e.	1000	s.e.	10000	s.e.
Wick	SRJPM	4.46	0.025	4.49	0.029	4.54	0.035	4.66	0.063
Aberdeen	SRJPM	5.47	0.082	5.53	0.100	5.61	0.126	5.85	0.233
Leith	SRJPM	6.70	0.073	6.75	0.088	6.81	0.111	7.03	0.190
North Shields	SRJPM	6.30	0.108	6.37	0.131	6.46	0.165	6.73	0.302
Whitby	SRJPM	6.82	0.110	6.89	0.132	6.97	0.165	7.23	0.303
Immingham	SRJPM	9.08	0.118	9.25	0.141	9.49	0.174	10.47	0.309
Cromer	JPM	6.96	NA	7.03	NA	7.10	NA	7.30	NA
Lowestoft	SRJPM	5.29	0.212	5.52	0.236	5.85	0.270	7.07	0.396
Felixstowe	JPM	5.90	NA	5.95	NA	6.00	NA	6.13	NA
Harwich	SRJPM	5.76	0.315	5.86	0.376	6.00	0.469	6.45	0.840
Walton	SRJPM	6.04	0.219	6.12	0.252	6.22	0.298	6.50	0.456
Southend	SRJPM	7.58	0.106	7.72	0.120	7.90	0.140	8.55	0.212
Sheerness	SRJPM	7.33	0.071	7.40	0.082	7.49	0.097	7.73	0.156
Dover	SRJPM	8.22	0.087	8.31	0.108	8.43	0.140	8.81	0.278

Table 5.3: Return levels (in metres relative to ACD) for 1990 obtained using the best site-by-site methods: return periods of 250, 500, 1000 and 10000 years for the east coast sites.

Site	Method	10	s.e.	25	s.e.	50	s.e.	100	s.e.
Newhaven	SRJPM	7.55	0.015	7.67	0.021	7.73	0.024	7.83	0.029
Portsmouth	SRJPM	5.42	0.034	5.53	0.046	5.59	0.057	5.69	0.079
Weymouth	JPM	2.97	NA	3.03	NA	3.06	NA	3.12	NA
Devonport	SRJPM	6.13	0.017	6.20	0.036	6.25	0.065	6.34	0.112
Newlyn	SRJPM	6.01	0.004	6.07	0.011	6.10	0.014	6.19	0.020
Ilfracombe	SRJPM	10.33	0.020	10.47	0.038	10.57	0.051	10.75	0.080
Hinkley	JPM	13.26	NA	13.40	NA	13.49	NA	13.65	NA
Avonmouth	SRJPM	14.65	0.036	14.84	0.053	14.98	0.064	15.21	0.083
Newport	JPM	14.11	NA	14.37	NA	14.50	NA	14.68	NA
Swansea	SRJPM	10.62	0.032	10.73	0.033	10.79	0.041	10.90	0.061
Mumbles	JPM	10.66	NA	10.75	NA	10.81	NA	10.91	NA
Milford Haven	SRJPM	7.91	0.015	8.03	0.026	8.12	0.037	8.26	0.066
Fishguard	SRJPM	5.56	0.010	5.66	0.017	5.72	0.023	5.82	0.036
Barmouth	JPM	6.58	NA	6.74	NA	6.82	NA	6.94	NA
Holyhead	SRJPM	6.46	0.017	6.57	0.027	6.64	0.036	6.77	0.056
Liverpool	JPM	10.92	NA	11.15	NA	11.28	NA	11.49	NA
Heysham	SRJPM	11.03	0.033	11.26	0.057	11.40	0.077	11.68	0.128
Port Erin	JPM	6.56	NA	6.67	NA	6.73	NA	6.81	NA
Workington	JPM	9.87	NA	10.01	NA	10.09	NA	10.20	NA
Portpatrick	SRJPM	4.80	0.034	4.93	0.046	5.00	0.054	5.12	0.071
Millport	JPM	4.74	NA	4.86	NA	4.92	NA	5.01	NA
Islay	JPM	1.87	NA	1.88	NA	1.89	NA	1.90	NA
Tobermory	JPM	5.69	NA	5.83	NA	5.90	NA	6.01	NA
Stornoway	SRJPM	5.92	0.024	6.03	0.028	6.09	0.030	6.17	0.034
Ullapool	SRJPM	6.13	0.016	6.23	0.024	6.29	0.030	6.40	0.040
Kinlochbervie	JPM	6.12	NA	6.23	NA	6.29	NA	6.38	NA
Lerwick	SRJPM	2.76	0.006	2.81	0.008	2.84	0.009	2.88	0.016

Table 5.4: Return levels (in metres relative to ACD) for 1990 obtained using the best site-by-site methods: return periods of 10, 25, 50 and 100 years for the south and west coast sites.

Site	Method	250	s.e.	500	s.e.	1000	s.e.	10000	s.e.
Newhaven	SRJPM	7.91	0.035	7.96	0.040	8.01	0.048	8.17	0.084
Portsmouth	SRJPM	5.79	0.109	5.84	0.130	5.91	0.162	6.14	0.291
Weymouth	JPM	3.16	NA	3.19	NA	3.22	NA	3.27	NA
Devonport	SRJPM	6.45	0.163	6.53	0.196	6.64	0.243	7.05	0.418
Newlyn	SRJPM	6.27	0.028	6.32	0.034	6.38	0.044	6.55	0.084
Ilfracombe	SRJPM	10.95	0.116	11.07	0.140	11.24	0.174	11.84	0.295
Hinkley	JPM	13.79	NA	13.86	NA	13.94	NA	14.04	NA
Avonmouth	SRJPM	15.45	0.133	15.60	0.165	15.79	0.210	16.46	0.365
Newport	JPM	14.83	NA	14.91	NA	15.00	NA	15.26	NA
Swansea	SRJPM	11.02	0.118	11.10	0.157	11.21	0.216	11.63	0.449
Mumbles	JPM	11.00	NA	11.05	NA	11.12	NA	11.35	NA
Milford Haven	SRJPM	8.43	0.104	8.54	0.130	8.70	0.168	9.34	0.324
Fishguard	SRJPM	5.93	0.052	5.99	0.064	6.07	0.083	6.32	0.167
Barmouth	JPM	7.04	NA	7.09	NA	7.16	NA	7.30	NA
Holyhead	SRJPM	6.91	0.079	7.00	0.093	7.12	0.114	7.56	0.189
Liverpool	JPM	11.67	NA	11.75	NA	11.82	NA	11.95	NA
Heysham	SRJPM	11.99	0.188	12.20	0.229	12.52	0.289	13.79	0.526
Port Erin	JPM	6.87	NA	6.90	NA	6.93	NA	6.97	NA
Workington	JPM	10.31	NA	10.36	NA	10.42	NA	10.50	NA
Portpatrick	SRJPM	5.23	0.090	5.29	0.102	5.37	0.120	5.60	0.188
Millport	JPM	5.08	NA	5.12	NA	5.17	NA	5.21	NA
Islay	JPM	1.95	NA	1.97	NA	1.98	NA	2.06	NA
Tobermory	JPM	6.09	NA	6.14	NA	6.19	NA	6.30	NA
Stornoway	SRJPM	6.25	0.039	6.29	0.043	6.34	0.048	6.48	0.069
Ullapool	SRJPM	6.50	0.053	6.55	0.062	6.62	0.076	6.83	0.135
Kinlochbervie	JPM	6.46	NA	6.51	NA	6.55	NA	6.60	NA
Lerwick	SRJPM	2.92	0.023	2.95	0.027	2.99	0.034	3.10	0.063

Table 5.5: Return levels (in metres relative to ACD) for 1990 obtained using the best site-by-site methods: return periods of 250, 500, 1000 and 10000 years for the south and west coast sites.

Part III

Spatial extension of the full model

Chapter 6

Introduction to spatial methods

Parts III and IV of this report concern the development and application of methods of extreme sea-level analysis that use information from neighbouring sites along a coastline, as well as data from the site of interest, for obtaining trend and return level estimates. Such an approach is termed a spatial method. Success in extending and applying the site-by-site methods to spatial procedures requires the extreme sea-level process to have spatially coherent characteristics along a coastline. Results in Dixon and Tawn (1994) and in Part II of this report, show that this is the case for the data from the study sites on the east coast.

In developing, and using, a spatial model in effect all that is being used is the additional physical knowledge about the sea-level process, i.e. separately the tide and surge themselves vary smoothly along a coastline, which implies that the the distribution of extreme sea-levels must vary smoothly along the coastline. Similarly, as isostatic and eustatic trends vary smoothly spatially then so should the trend in extreme sea-levels. Therefore, extension of the site-by-site methods to spatial procedures corresponds to the use of improved oceanographic and geological modelling within the statistical analysis.

The relationships derived between the extreme sea-level process for different sites using spatial methods necessarily provide a continuous model for extreme sea-levels along the coastline. Furthermore, as spatial methods incorporate all the relevant data into the estimation of the spatial coherent extremal characteristics, the precision of estimates of trends and return levels will be improved. Specifically, the benefits of adopting a spatial approach vary with site:

- **For sites with long records:** good estimates of return levels are usually adequately provided by a site-by-site analysis, such as the RJPM or the SRJPM. Specifically, estimates obtained by spatial methods are likely to provide similar return level estimates to those obtained using site-by-site methods for return periods up to 100-250 years, but for longer return periods and trends spatial methods should

provide better estimates;

- **For sites with short spans of data:** return level estimates obtained using site-by-site methods are generally quite poor because the data provide no information about trends or for return periods longer than 50 years. Furthermore, even return level estimates for short return periods are the subject of potential bias due to unusual conditions over the sampling period. Spatial methods produce better estimates of trends and return levels and overcome sampling bias which may be present, thus improving estimates for short return periods;

- **For sites along the coastline with no data records:** site-by-site methods cannot be used. However, for such sites estimates of return levels and trends are obtained automatically by use of spatial methods. This therefore is the most important consequence of adopting a spatial method.

In Dixon and Tawn (1994) spatial coherence around the UK, as observed from the data sites, was found to be much greater along the east coast than the west coast. On the south coast little could be inferred owing to the absence of sites with sufficient data. There are two reasons for this difference between the east and west coasts:

- the spatial coverage of A-class sites, with sufficient data for extreme value considerations, is extensive on the east coast and sparse on the west coast,

- the surge process is more coherent on the east coast, as typically surges are externally generated and propagate down that coast gradually amplifying in level as they travel south; whereas on the west coast surges are more localised in extent, approaching from offshore and not propagating along the coastline.

A consequence of these features is that a more dense spatial network of data-sites, than that given by the A-class sites, is required to apply spatial methods to south and west coast data. A possible source of additional data for these coasts, which increases the temporal and spatial extent of the available data from that given by the data sites, is from hydrodynamical models. These synthesised data form the basis of the third phase of the work by Lancaster University for MAFF, in which a 39 year run of the hydrodynamical model will be used, and these data interfaced with historical annual maxima and hourly observations at the analysis stage. Based on data from the study sites alone, a spatial model for extreme sea-levels can only be developed for the UK east coast.

6.1 Illustrative example

The crudest approach to spatial modelling is to take the return level estimates, obtained using the site-by-site methods applied to the study sites, and spatially smooth these. The method of smoothing needs to account for the different uncertainties from site to site, so at the data sites

- for sites with long records, and typically small standard errors, the smoothed return level estimate is almost identical to the estimates obtained using the site-by-site methods;
- for sites with short records, and typically large standard errors, the smoothed return level estimates can differ considerably from the estimates obtained using the site-by-site methods depending on information from neighbouring sites.

For intermediate sites with no data the spatially smoothed estimate can take any form, in that it primarily interpolates estimates between sites. However at this crudest level of spatial modelling the spatial interpolation takes no account of known physical changes in processes between data sites, and so is unlikely to provide good estimates in coastal regions where the tidal characteristics are known to change significantly, and non-linearly, along the coastline. Figures 6.1-6.3 show spatial maps for the east coast, indicating distance from Wick, of such a spatial model providing return level estimates. Specifically such spatial estimates of 10, 100, 1000, and 10000 year return levels shown in Figures 6.1–6.3: based on the r-largest method, the JPM, and the RJPM respectively. All three figures show the spatial estimate obtained for 1990, and in addition, Figures 6.1–6.3 show this estimate for the year 2100. Since there are no standard errors for return levels obtained by the JPM the weighting using in the spatial interpolation is based on the relative lengths of data records for the data sites. Each of the methods gives similar results for the 1990 estimate of return levels in terms of their basic shape. The characteristics of each curve are a gradual increase in levels down the east coast to the region around Immingham, followed by a significant decrease to Lowestoft and a subsequent rise through to Dover. This pattern is similar for the four return periods shown in the plots. The spatial estimate is seen to be close to the site-by-site estimates unless the confidence intervals for the estimates are large, or when there is are neighbouring sites with considerably more information; see for example Harwich in Figures 6.1 and 6.3.

There are two critical weaknesses of this approach to spatial estimation of return levels, which can be seen from these figures:

1. **Incorporating knowledge of spatial tides**: the continuous estimate obtained from smoothing return level estimates for the data sites reflects the positions of the

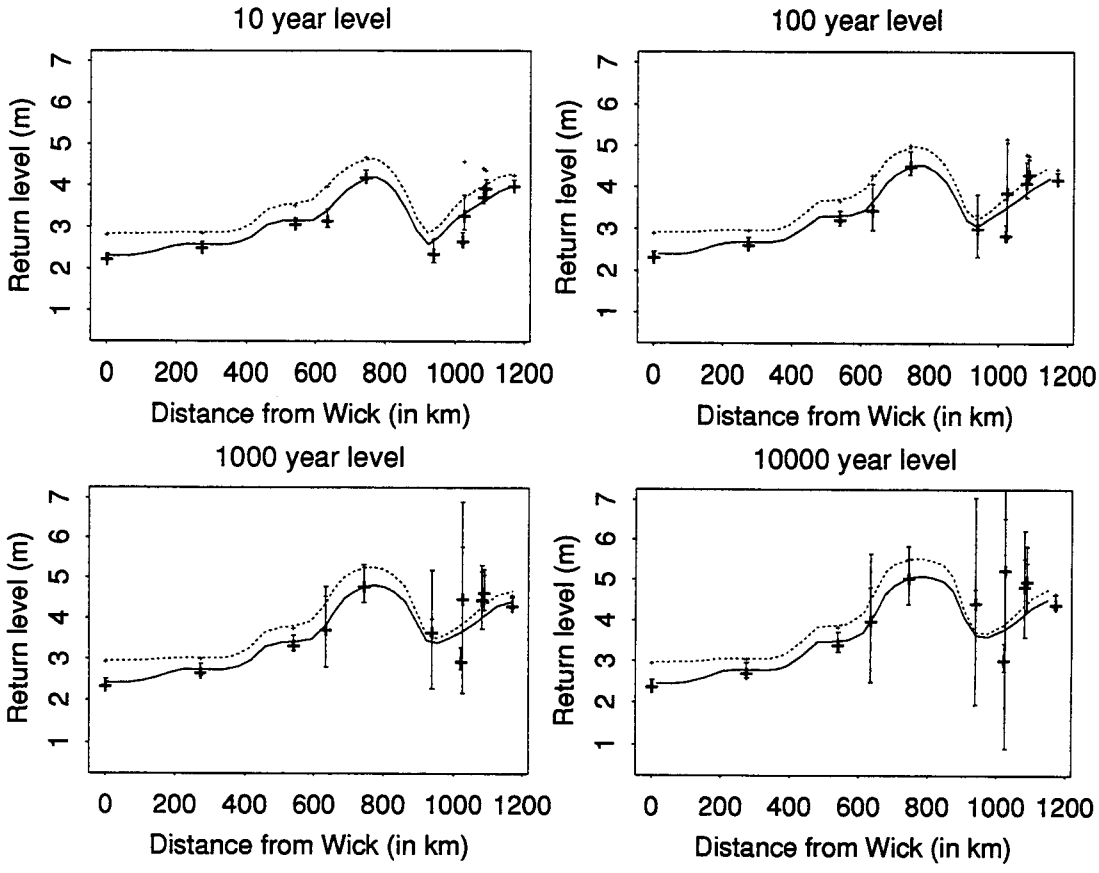

Figure 6.1: Crude smoothed return levels (relative to mean sea-level in 1990) plotted against distance from Wick (in km) for 1990 (solid line) and 2100 (broken line) using the r-largest method. The site-by-site estimates for 1990 and 2100 are also displayed with 95% confidence intervals given for 1990.

data sites. For example, Immingham is in an estuary and so has a larger tidal range than neighbouring sites on the open coast. By spatially smoothing return levels the whole region around Immingham has high levels, yet this is artificial as the tides are only higher in the Humber estuary, and the surges should change smoothly along the neighbouring open coast. The Immingham example shows that high tidal levels at a site relative to the local coastal stretch can lead to over-estimation of return levels along that coastline. Opposite situations also occur. For example, the spatial estimates show no rise in levels around regions such as the Moray Firth, the Firth of Forth, the Wash and the Thames, yet it is known that the tidal levels are non-linear in these regions and can be larger than on the surrounding open coast. When no data sites are in these non-linear regions, the crude approach to spatial smoothing adopted here cannot recover such knowledge.

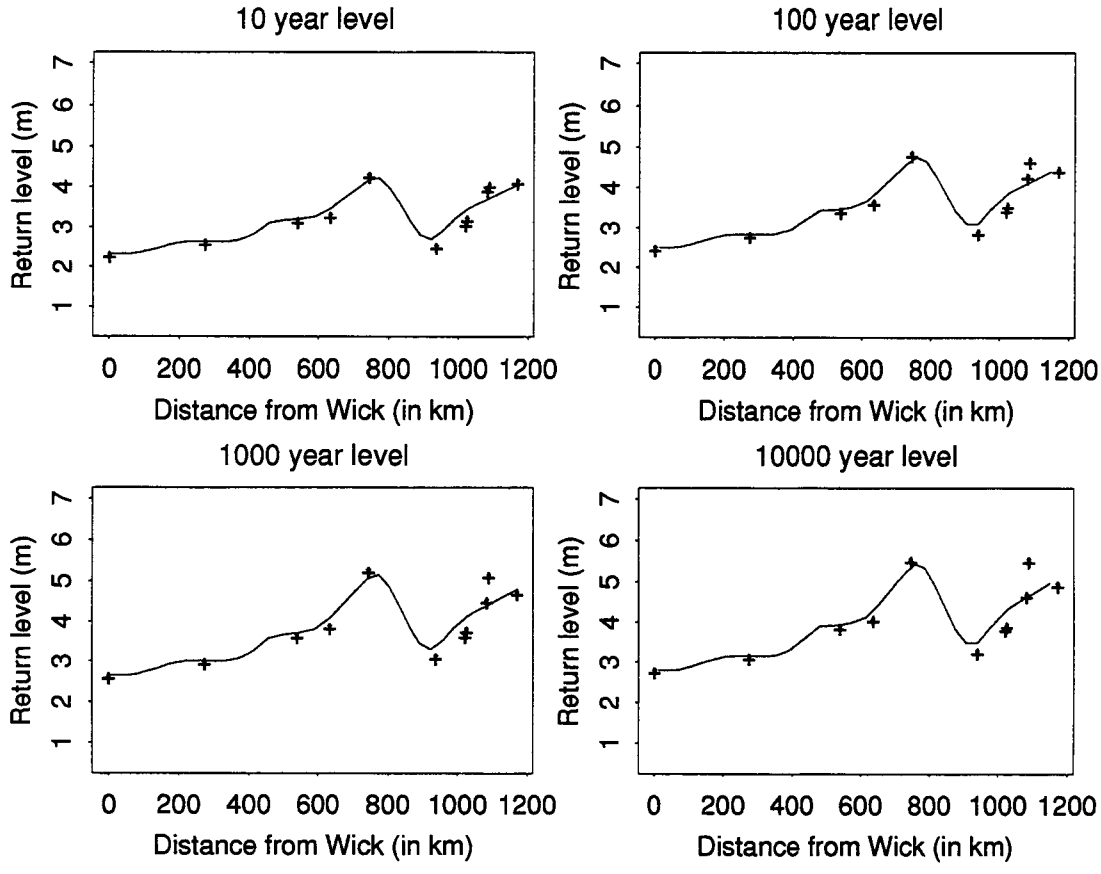

Figure 6.2: Crude smoothed return levels (relative to mean sea-level in 1990) plotted against distance from Wick (in km) for 1990 using the joint probabilities method. The site-by-site estimates for 1990 and 2100 are also displayed.

2. **Handling of trends**: now consider estimation of return levels for the year 2100. Adopting the procedure of earlier, we estimate the return level for 1990 for each site and adjust these to the year 2100 by using the site-by-site estimates of the trend. The spatial estimate for 2100 is then obtained by spatially smoothing these estimates. For some sites the trend estimates are quite poor, so the resulting site-by-site return level estimates for 2100 are also poor, exhibiting less spatial smoothness than for 1990. Correspondingly, the spatially smoothed estimate captures some of this roughness. For example, in Figure 6.3 the 10 year return level plot shows that the trend estimate for Whitby is probably much too large and that this has distorted the spatial estimate for all sites on the coast between North Shields and Immingham. This demonstrates how the spatial smoothing depends on the year of interest, a feature which is undesirable when trends are estimated so badly using site-by-site estimates.

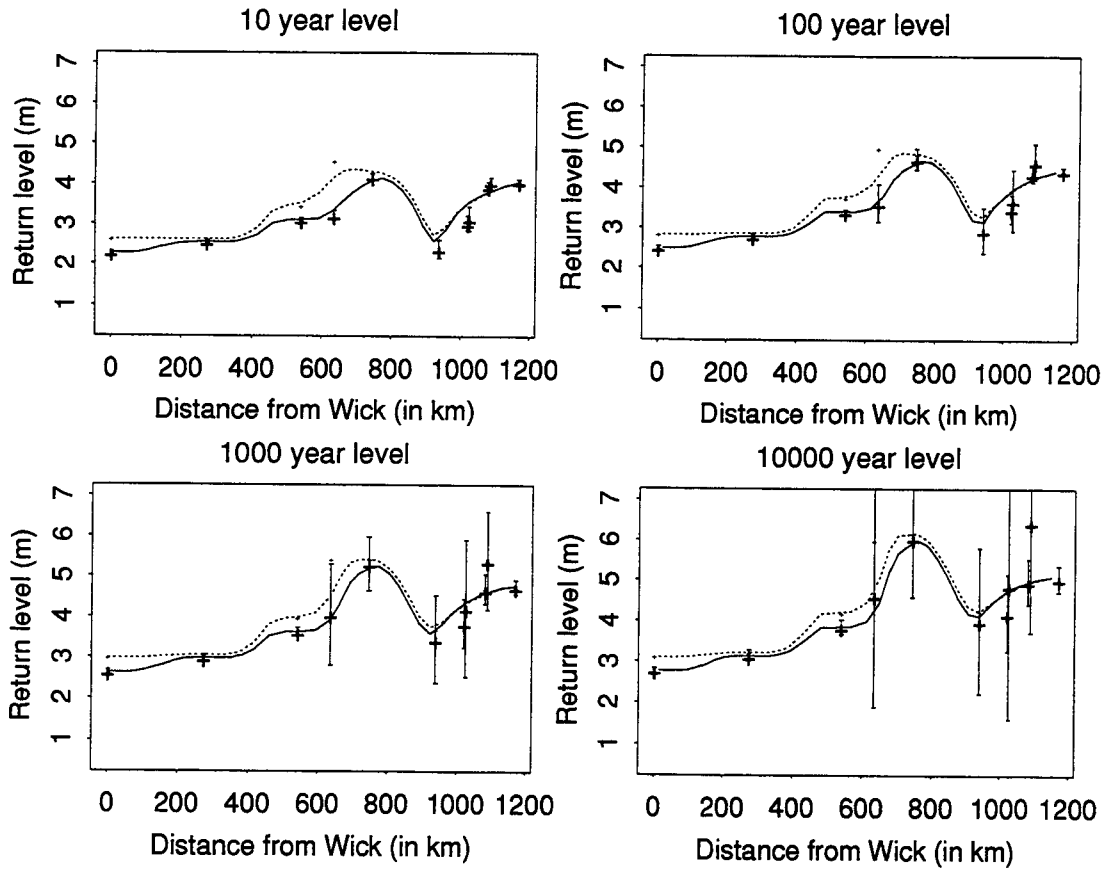

Figure 6.3: Crude smoothed return levels (relative to mean sea-level in 1990) plotted against distance from Wick (in km) for 1990 (solid line) and 2100 (broken line) using the RJPM. The site-by-site estimates for 1990 and 2100 are also displayed with 95% confidence intervals given for 1990.

6.2 Outline for the spatial model

Instead of taking the simplistic approach outlined above we adopt a more sophisticated method of spatial analysis which exploits the known separate spatial variations in the tidal and surge characteristics. Since the role of the tide, in determining the distribution of the extreme still water level, is only exploited in the JPM and RJPM/SRJPM we do not consider the r-largest approach further. Furthermore, of the joint probability approaches we focus on extending spatially the SRJPM only. Our justification for this is that Dixon and Tawn (1994) show that the JPM does not provide statistical errors, it is restrictive in the form of extrapolation to extreme return levels due to the use of the empirical distribution of surges, and it cannot be extended to include historical annual maximum data, as shown in Chapter 6; and the RJPM because it also does not include historical data or have spatially interpretable parameters.

The spatial model for the SRJPM involves spatial estimation of the parameters present in the site-by-site model described in Part I. Here we briefly discuss these parameters and outline the features which would be considerably improved if a spatial approach to estimation was adopted. Separate spatial models are developed for each of the parameters of the SRJPM, with each model exploiting the coherence of the underlying physical processes of tide, surge, and tide-surge interaction. Together these determine the still water level extremes. Combining the smooth spatial models for each of these underlying parameters of the SRJPM leads to the resulting estimation and interpolation of still water return levels and trends being significantly improved.

Parameters of the SRJPM are:

Tidal sequence

In order to give estimates of return levels at any point along the east coast, the tidal sequence is required for each site of the application of the spatial model. Some form of spatial interpolation of the tidal sequences at each data-site is required. Since the tidal sequence is important in determining return levels for short return periods at a site, a more sophisticated interpolation scheme is used. This scheme combines tidal information from the European continental shelf hydrodynamical model at a fine grid with tidal information from the A-class sites; details of the adopted approach are given in Section 12.1.

Tide-Surge Interaction

In Dixon and Tawn (1994) and Parts I and II of this report tide-surge interaction was found to be important for the entire length of the east coast, with the degree of interaction varying from site to site in a coherent manner. Interaction was modelled through the functions $a(X, i)$ and $b(X)$. Here these functions are modelled separately providing spatial analogues of the present interaction functions, i.e. they are defined at any point along the coast. These change smoothly with distance along the coast and the tidal state. This component of the SRJPM is discussed in Section 12.3.

The Surge Distribution

Once spatial versions of the interaction functions have been developed then spatial modelling of the surge distribution is reduced to spatial modelling of the upper tail of the distribution of the transformed surge variable, S^*. For each site

$$S_t^* = (Y_t - a(X_t, i))/b(X_t).$$

This transformed surge series is stationary in time.

The aim of separating the modelling of extreme surges into the two stages of the interaction functions and the distribution of extremes of the S^* variable is to explain the principle sources of spatial variation through the interaction functions, which are estimable from the bulk of the data, and leave only a simple spatial structure to be explained in the tail of the distribution of S^*, which must be estimated from extreme values. Consequently, even sites which provide poor site-by-site return level estimates can be used to estimate the interaction functions, thereby improving the spatial interpolation of these by increasing the spatial density of sites used.

Although the approach used in Chapter 12 of this report is slightly different from that described above, the general philosophy of the approach is the same. A point process method is used for the extremes of the surge series with $a(X_t, i_0)$ and $b(X_t)$ taken as covariates; details of the spatial estimation of the point process parameters is discussed in Section 12.4 for the east coast data.

Extreme trend parameter

Dixon and Tawn (1994) found that accurate estimation of the trend in extreme sea-levels is difficult using only hourly data separately from site to site as these data are of limited span. The approach developed in Part I and adopted in Part II was to augment these hourly data by information from historical annual maximum still water levels and/or mean sea-level data which are of longer spans. Trends in extreme still water levels are due to eustatic and isostatic trends. Eustatic trends are constant along the east coast and trends in extreme sea-levels should therefore exhibit spatial variations consistent with the pattern of isostatic trends. Trend estimation is improved by incorporating knowledge of this spatial pattern. There are two possibilities:

- use estimates of isostatic trends obtained by Shennan (1989), or some other source,
- assume only that there is a smooth spatial pattern to the trend, without taking a specific form.

These methods are assessed for spatial estimation of extreme sea-level trends in Chapters 8, 9, and in Section 12.2.

Extremal Index

In Dixon and Tawn (1994) the ratio of surge and still water level extremal indices is estimated and found to exhibit smooth spatial variation. Here, in Section 12.5, it is found necessary to separately estimate each of the extremal index parameters as

well; hence separate spatial models for the individual indices as well as for the ratio are considered.

6.3 Choice of spatial approach

Suppose an estimate of a parameter, $\theta(x)$ say, is required at location x, but estimates are available only at the d study sites with locations $\{x_1, \ldots, x_d\}$. This is the underlying situation in all spatial methods, where $\theta(x)$ may correspond to a return level at position x along a coastline, or could be a parameter of the SRJPM along the coastline. There are primarily two methods of spatial analysis to give estimates $\theta(x)$ at any location x along the coast:

- Parametric
- Non-parametric.

Both methods involve two steps: specifying a family of suitable functions, and estimating these functions using the data/information from the data-sites. The key distinction between the two methods is in terms of the first step. In an application of the parametric approach a simple class of analytical functions, e.g. low order polynomials, is used, with the function depending on a few unknown parameters. Non-parametric methods use the class of smoothly varying functions such as splines, smoothed splines or kernel regression functions (Silverman, 1985; Hardle, 1989; and Chu and Marron, 1991).

Whether adopting a parametric or non-parametric approach there are two choices for the method of estimation:

- Likelihood methods
- Weighted least squares.

Estimation based on the methods of likelihood and weighted least squares both involve obtaining 'best' parameter values subject to some criteria. In the case of likelihood methods the parameter estimates are taken to be those values of the parameters which maximise the probability of obtaining the spatial data, whereas in the weighted least squares method the distance between the spatial parameter estimate and the site-by-site estimates is minimised. In particular, the likelihood methods require the specification of the joint distribution of observations over sites along the coast to form the likelihood function. This function is then maximised with respect to the parameters of $\theta(x)$ in a parametric approach, or with respect to the class of smooth functions (subject to some roughness criteria) for non-parametric methods. Alternatively, the weighted least squares methods

focus on the site-by-site estimates $\{\hat{\theta}(x_1),\ldots,\hat{\theta}(x_d)\}$, of the parameters $\{\theta(x_1),\ldots,\theta(x_d)\}$, and their respective standard errors $\{se(\hat{\theta}(x_1)),\ldots,se(\hat{\theta}(x_d))\}$. Specifically, for parametric methods the function

$$\sum_{i,j=1}^{d}(\hat{\theta}(x_i)-\theta(x_i))\Sigma_{i,j}(\hat{\theta}(x_j)-\theta(x_j)), \tag{6.3.1}$$

where $\Sigma_{i,j}$ is the (i,j)th entry of the inverse of the variance-covariance matrix of the site-by-site estimates $\{\hat{\theta}(x_1),\ldots,\hat{\theta}(x_d)\}$, requires minimisation over the parameters of $\theta(x)$. Alternatively for non-parametric methods the function

$$\sum_{i,j=1}^{d}(\hat{\theta}(x_i)-\theta(x_i))\Sigma_{i,j}(\hat{\theta}(x_j)-\theta(x_j))+ \text{ a roughness penalty}, \tag{6.3.2}$$

where $\Sigma_{i,j}$ is as in equation (6.3.1), and requires minimisation over the class of suitably smooth functions. The roughness penalty in equation (6.3.2) usually takes the form of

$$\lambda \times \int_A [\theta'(x)]^2 dx$$

where A is the set corresponding to locations of the coastline of interest and $\theta'(x)$ is the derivative of $\theta(x)$ at x. Here λ is a roughness parameter which controls the roughness allowed for $\theta(x)$, so large values of λ cause only very smooth functions to be admissible, whereas a small choice for λ enables much rougher functions for $\theta(x)$ to be considered.

In applications of the weighted least squares method, independence is often assumed so that

$$\Sigma_{i,j}=0 \text{ if } i\neq j \text{ and } \Sigma_{i,i}=1/[se(\hat{\theta}(x_i))]^2 \text{ for } i=1,\ldots,d.$$

In this case, equation (6.3.1) reduces to

$$\sum_{i=1}^{d}\left(\frac{\hat{\theta}(x_i)-\theta(x_i)}{se(\hat{\theta}(x_i))}\right)^2,$$

and equation (6.3.2) becomes

$$\sum_{i=1}^{d}\left(\frac{\hat{\theta}(x_i)-\theta(x_i)}{se(\hat{\theta}(x_i))}\right)^2+\lambda\times\int_A[\theta'(x)]^2dx.$$

This is a useful assumption as the site-by-site estimates are obtained without providing information about correlation of estimates with estimates at different sites. In practice parameter estimates will exhibit dependence between sites as the data they are derived from are spatially dependent. Thus independence is a working assumption which makes the analysis easier. The consequences of the assumption are that estimates are unbiased but standard errors of the fitted spatial model may be too small, so non-standard methods are required to estimate these.

6.3.1 Properties

Each form of smoothing and estimation has features which may make it the more suitable in certain spatial applications. These features are:

- **Ease of application:**
  Likelihood methods are complex to apply in high dimensional problems with varying degrees of missing data as they require detailed modelling of the spatial dependence structure of the extreme sea-level process. By comparison weighted least squares methods are easy to implement as they require only the output parameter estimates from the site-by-site methods.

- **Multiparameter cases:**
  Likelihood methods handle interdependence between parameters at different sites or of different aspects at a site naturally. As weighted least squares takes each parameter separately there is a risk of inappropriate spatial smoothing if some parameters are highly dependent.

- **Precision estimates:**
  Likelihood methods can provide estimates of precision of the parameters or a combination of parameters in a more efficient way, and can exploit any dependence in the parameters in a straightforward way. Although estimated errors are available for the non-parametric methods, they are generally less reliable.

- **Complexity of functions/flexibility:**
  Parametric models are ideal models if the spatial variation is simple in form as then low dimensional polynomial models, which have only a few parameters to estimate, can be used. When the parameter of interest exhibits complex spatial structure parametric methods are likely to over-simplify the spatial representation. Non-parametric methods impose no predetermined form and allow the function to be determined by the structure at the observed sites subject to a smoothness criteria, and hence is more flexible.

- **Extensions to applications in future studies:**
  The methods adopted in this study will be used in the third stage of the study so must be extendible to higher dimensions and complex spatial structures for the underlying process. Methods which are at their limits in the present study will not therefore be applicable in the next stage, so will be difficult to implement in this or future studies.

It is not immediately apparent which of these methods is likely to be the most appropriate for the present, or any future, spatial analysis of the UK east coast extreme sea-level data using the SRJPM. Therefore rather than make an arbitrary choice we first examine the viability of the various approaches by using a range of case studies which provide considerable evidence in favour of using non-parametric spatial methods based on a weighted least squares approach. These case studies are given in Chapters 7 to 10.

6.4 Outline for the rest of the report

Chapters 7 – 10 describe case studies associated with the development of a spatial analysis for aspects of the model for extreme sea-levels developed in Part I of this report. These studies are of:

- Extreme sea-level trends
- Mean sea-level trends
- Spatial modelling of r-largest data
- Tide-surge interaction around the UK.

In Chapter 11 the findings from the case studies which have importance in the development of a spatial method for extreme sea-level analysis are reviewed, and the spatial approach that is to be adopted is described.

Based on the findings of Chapter 11, the chosen spatial method is applied in Part IV, with Chapter 12 discussing the application of the method to the separate components of the model, and Chapter 13 showing the final estimates in graphical and tabular forms. If interest is purely in terms of *results* for the east coast, the reader should proceed to Part IV.

Chapter 7

Case study I: extreme sea-level trends

7.1 Introduction

Return level estimates for future years, obtained using the methods described throughout the report, are sensitive to trends in extreme sea-levels and so detailed knowledge about them is vital for the design of coastal flood prevention schemes. In this case study we describe a spatial model for extreme sea-level trends using annual maximum data. This provides an estimate of the extreme sea-level trend around the entire UK coastline.

The sea-level is recorded using tide gauges, which measure the height of the ocean surface relative to the adjacent land. The *eustatic sea-level trend* is defined to be the trend in the actual sea-level, i.e. the trend in the recorded sea-level minus the land trend. Trends in eustatic extreme sea-level can arise due to trends in (*i*) eustatic mean sea level, (*ii*) the number of extreme storms, (*iii*) the variability of extreme storms, and (*iv*) the tidal range, each of which may contribute significantly. The majority of previous studies of sea-level trends have concentrated on the analysis of mean sea-level trends, so the effects of (*ii*), (*iii*) and (*iv*) are ignored, and consequently these studies fail to provide the principal information required by coastal engineers and planners (Pugh, 1987, Chapter 8; Rossiter, 1969).

In this case study trends in extreme sea-level are studied using sea-level annual maximum data from 62 UK sites: Figure 7.1 shows the positions of these data sites. The number of annual maximum data available for each site ranges from 10, at Margate, to 136, at Sheerness, with a mean of 43 per site. These data are described in detail by Graff (1981) and Coles and Tawn (1990). Authorities involved in coastal design and management are primarily concerned with observed trends, whereas climatologists and oceanographers are mainly interested in eustatic trends, so both observed and eustatic

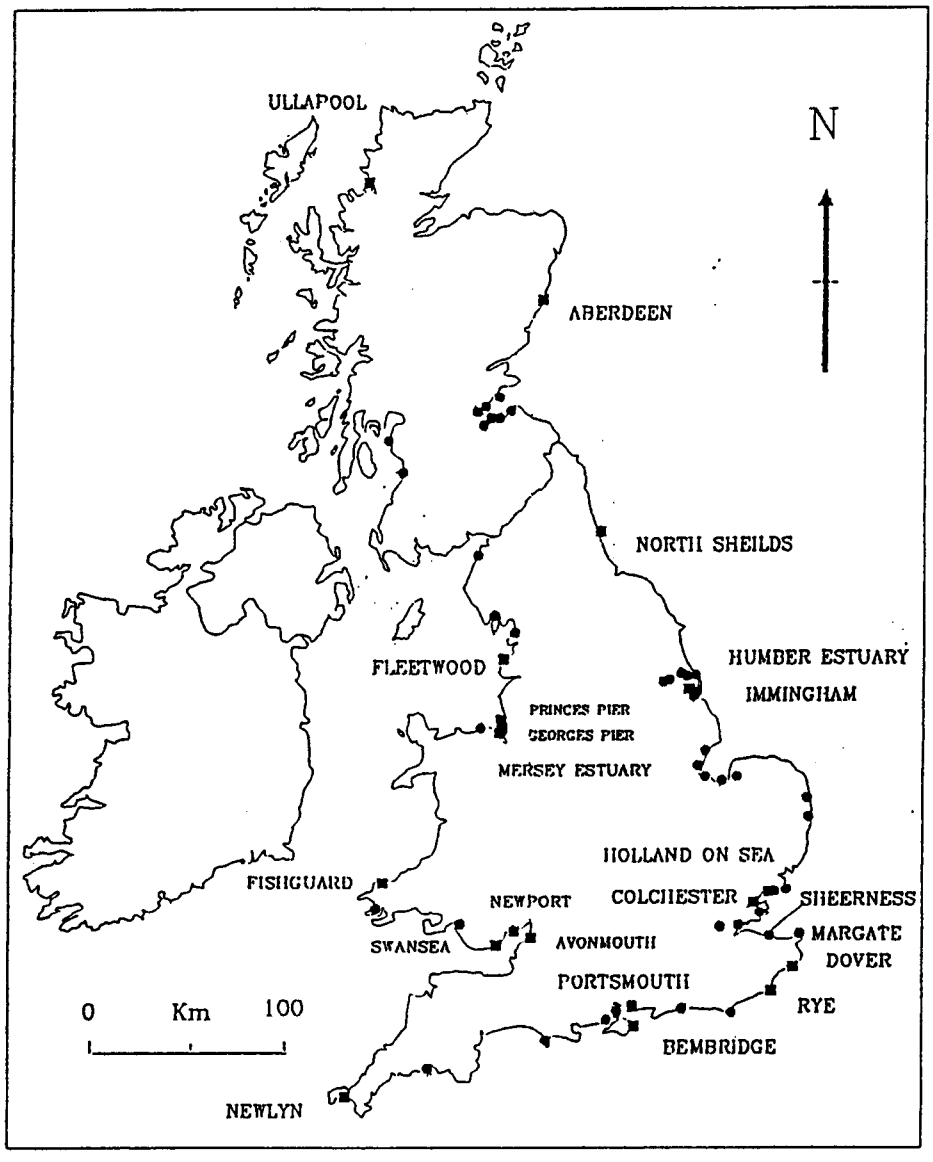

Figure 7.1: Map showing the locations of the 62 UK annual maximum sites.

trends are important and are examined in this section.

In order to estimate trends accurately from historical tide gauge data, it is essential to have high quality instrumental maintenance and control of the tidal datum. For instance, a slipping datum induces a spurious trend in the annual maximum series. Since trends in annual maximum data are of the order of -3 to 6mm per year, errors of this kind can be overlooked yet may have an influential effect. We attempt to reduce the effects of these problems by using a spatial model for trends which essentially spatially smoothes trends over local sites.

For the estimation of the extreme sea-level trend at a site, a simple approach is to use standard linear regression techniques to fit a linear model to the annual maxima data (Blackman and Graff, 1978). However, this fails to exploit knowledge about either the form of the error distribution or the spatial dependence of the data. Therefore typically only the longest series are analysed, as 'poor' trend estimates can be obtained from short series. This approach wastes much data and provides only a limited knowledge of the spatial variation of long term extreme sea-level trends around the coast. Figure 7.2 shows data and trend estimates for two south coast sites, illustrating a situation where exploiting spatial dependence can improve trend estimation. The trend estimate for Bembridge, ignoring information from the neighbouring site – Portsmouth, is large and negative. Over the period of the Bembridge data, the Portsmouth series shows a negative trend, however viewed over a longer time scale these data are consistent with the positive trend obtained for the site. By accounting for spatial dependence and exploiting the regional coherence of extreme sea-levels along the south coast, without any assumption of spatial smoothness for trends, the estimated trend for Bembridge is found to be more consistent with that for Portsmouth.

In Section 7.2, a smooth spatial estimate of trends in UK extreme sea-level data is obtained. To aid discussion within this case study, in Sections 7.2.1 and 7.2.2 respectively, the annual maximum analysis, applied site-by-site, and the spatial version of this approach are briefly summarised. In Section 7.2.3, a non-parametric model for the spatial trend estimates is developed, giving a smooth estimate around the entire UK coastline. This estimate has significant spatial variation with large positive trends in the South East and small or negative trends around the Scottish coast.

Shennan and Woodworth (1992) suggest that eustatic mean sea level trends should be approximately constant over the relatively small region of the British Isles, the main variation between sites being due to relative vertical land movements. The extent to which the variation found in Section 7.2 is due to this factor is examined in Section 7.3 and results suggest that UK eustatic extreme sea-level trends are approximately constant (1.0 to 1.2mm per year) around the coastline. This estimate is consistent with findings

of Shennan and Woodworth (1992) for eustatic mean sea-level trends. At present, there seems to be no evidence for trends in existing extreme sea-level data, other than in mean sea-level. This is discussed in Section 7.4.

7.2 Trends in extreme sea-levels

7.2.1 Marginal analysis

In this section the classical annual maxima approach is applied to the 62 UK sites. Let Z_t denote the annual maximum in year t for the site of interest. We restrict attention to linear trends with identically distributed errors so that the annual maximum sea-level in year t is taken to be GEV$(\alpha + \beta t, \sigma, k)$, with distribution function

$$\Pr\{Z_t \leq z\} = \exp\left[-\{1 - k(z - \alpha - \beta t)/\sigma\}^{\frac{1}{k}}\right], \text{ subject to } 1 - k(z - \alpha - \beta t)/\sigma > 0, \tag{7.2.1}$$

where $\alpha, \beta, \sigma(>0), k$, are the intercept, trend, scale and shape parameters respectively.

Under this model the first two moments are easily shown to be

$$\begin{aligned} E(Z_t) &= \alpha + \beta t + \sigma k^{-1}[1 - \Gamma(k+1)] \\ Var(Z_t) &= \sigma^2 k^{-2}[\Gamma(2k+1) - \Gamma(k+1)^2] \end{aligned} \tag{7.2.2}$$

where Γ is the gamma function. Typically for UK sea-levels k lies in the range -0.1 to 0.35 varying slowly around the coast, whereas σ takes values in the range 0.1 to 0.35m, see Coles and Tawn (1990) for values at specific sites. From (7.2.2) it can be seen that $E(Z_t)$ varies linearly with time but Var(Z_t) is independent of time. Thus a suitable model for Z_t is of the form of a linear model over time with independent and identically distributed GEV error structure. Estimation of this model is via maximum likelihood with the resulting trend estimates, obtained by separately applying this method to each site in the study, shown in Figure 7.3.

The abscissa in the plot is site distance measured anti-clockwise from Ullapool, see Figure 7.1. As the graph wraps around, site distances of 0 and 3935 km both correspond to Ullapool. Some spatial smoothness in these trend estimates is observed, but this feature is obscured by the estimates obtained at sites with short data series, poor data quality, or biased sampling. In particular, trend estimates obtained at Georges Pier, Princes Pier, Swansea, Bembridge, and North Shields are not consistent with those at neighbouring sites. The last three sites each have shape parameter estimates which are atypical for their respective coasts.

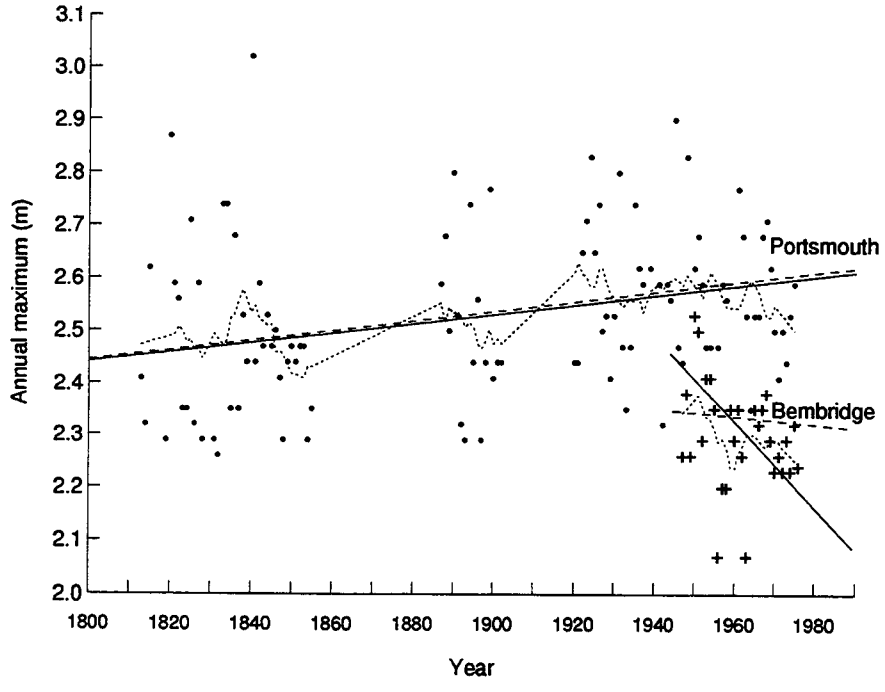

Figure 7.2: Plot of annual maximum against year for two south coast sites–Portsmouth (•) and Bembridge (+). The various lines shown for each site are ..... a least squares type smoothing of the annual maxima; —— the regression line as obtained by marginal estimation (Section 7.2.1); –·–·– the regression line as obtained by spatial estimation (Section 7.2.2).

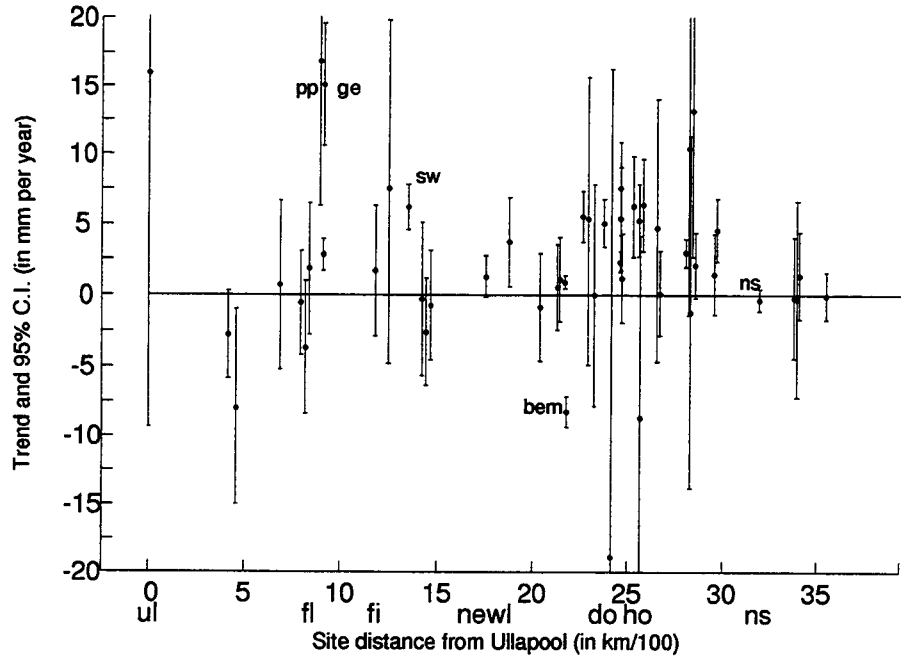

Figure 7.3: Trend estimates with 95% confidence intervals obtained by the marginal annual maxima method, described in Section 7.2.1. These estimates are plotted against anticlockwise distance from Ullapool. The poor fitting sites are seen to be: Georges Pier (ge), Princes Pier (pp), Swansea (sw), Bembridge (bem), and North Shields (ns), each having trend estimates which are inconsistent with those at neighbouring sites. The end sites of the separate coastal stretches, defined in Table 7.1, are indicated by site name abbreviations on the abscissa (fl–Fleetwood, fi–Fishguard, do–Dover, ho–Holland-on-Sea). The Humber estuary sites are not displayed here as they all have 'effectively' the same coastal distance so all the trend estimates merge together.

7.2.2 Spatial analysis

In order to clarify notation and later discussion within this case study, a brief summary of the spatial annual maxima method (Coles and Tawn, 1990) is given. The approach is to model the joint distribution of the annual maxima over data sites accounting for dependence between sites, and to model the changes in each of the parameters of the marginal distribution over sites in a way which is consistent with the properties of the underlying generating process.

Let $Z_{i,t}$ be the annual maximum at site i for year t, so that $Z_{i,t} \sim \text{GEV}(\mu_{i,t}, \sigma_i, k_i)$,

Section of coast	First and Last sites in stretch	l	$L_F - L_S$	p
Scottish West	Ullapool–Fleetwood	5	4.65	> 0.075
West	Gladstone Dock–Eastham Lock	3	4.10	> 0.100
South West	Fishguard–Avonmouth	5	10.14	< 0.005
South	Newlyn–Rye	9	17.10	< 0.005
South East	Dover–Colchester	6	17.62	< 0.001
East	Holland (on–Sea)–Immingham	11	19.28	< 0.001
Humber Estuary	Goole–Saltend Jetty	7	30.68	< 0.001
Scottish East	North Shields–Aberdeen	6	4.72	> 0.100

Table 7.1: Information on the likelihood ratio tests for homogeneous trends. Here L_F and L_S are the maximum log likelihoods when the trend parameters are unconstrained within a stretch and under the homogeneous trend model respectively. Also l is the degrees of freedom, i.e. the number of sites in the region minus one, and p is the significance level of the likelihood ratio test.

with $\mu_{i,t} = \alpha_i + t\beta_i$. The intercept parameters, α_i, are not spatially modelled since they depend mainly on the deterministic tidal sequence at site i, see Coles and Tawn (1990, Figure 7.3). A spatial model for this parameter is developed in Chapter 9 for the Humber estuary case study. The scale and shape parameters, σ_i and k_i respectively, are extremal characteristics of the family of surges that could occur around the time of high tide (Tawn, 1992). This family contains all surges for sites where the tide and surge are independent, whereas it is restricted to surges that occur at certain stages of the tide at sites which exhibit tide–surge interaction. Since the surge and tide–surge interaction are spatially coherent processes, this should be reflected in both σ_i and k_i. This expected spatial smoothness of σ_i and k_i is exploited by using a model with site position as a covariate, and applying it to the 8 stretches of coast defined in Table 7.1. The best fitting form of the model for each stretch is then determined by various diagnostic tests.

The advantages of this spatial approach over the marginal approach of Section 7.2.1 are two-fold. First by identifying the underlying relationship between parameters over sites, increased precision of the GEV parameters is obtained and estimates are robust to erroneous data typical in historical records for annual maximum data. Secondly, accounting for spatial dependence in the analysis allows a transfer of information between any two sites that have non-zero dependence. This information transfer can give a considerable increase in the precision of parameter estimates, and is most influential when some sites have missing data or when one site has a longer series of data than another. In Chapter 8 this is illustrated in the case of mean sea-levels using a simple regression example. Information transfer also helps to overcome the problems associated with the

tendency of tide gauges to fail in extreme conditions since information transfer effectively replaces these missing values.

On the basis of Figure 7.3, Coles and Tawn (1990) suggest that the trend parameters, β_i, do not display enough spatial stability, and are too sensitive to local effects, to be sensibly modelled spatially. Therefore they estimate the trend using the marginal method described in Section 7.2.1. There are two drawbacks in doing this: the first is that poor marginal estimates are obtained at some sites with short data series; and the second is that, as Coles and Tawn (1990) treat the trends as known in the spatial model, the standard errors of subsequent return level estimates are under–estimated. It is found that when the trends are estimated within the spatial model framework, considerably improved spatial stability of the estimates is obtained.

7.2.3 Spatial modelling of the trend parameters β_i

The models in Section 7.2.2 are extended to allow simultaneous estimation of the trend parameters within a joint likelihood analysis. As with the marginal trend estimation, this gives a separate, independent, estimate of the trend for each site; however in this case the error distribution, modelled spatially, accounts for both the spatial dependence and the coherence of the scale and shape parameters. In other words, smooth changes in the variability of the error distribution are modelled. The resulting trend estimates, plotted in Figure 7.4, have greater spatial stability than the marginal trend estimates (Figure 7.3), particularly estimates at Bembridge, Swansea, and North Shields which are now consistent with neighbouring sites. This improved stability arises even though there are no such restrictions imposed by the spatial model.

The simplest model for spatial trends is a single homogeneous trend over all sites, but Figure 7.4 shows this is clearly unsuitable. For exploratory purposes, a piecewise homogeneous model for the trend is fitted by constraining the β_i to be constant over sites within a stretch. These homogeneous trend estimates are shown in Figure 7.4. To test the statistical significance of these models, a likelihood ratio test (Cox and Hinkley 1974, Chapter 4.3) of β_i homogeneous against β_i free for each site i in the stretch is used. Results of these tests, summarised in Table 7.1, show that a homogeneous model for β_i can be taken in the Scottish West, West, and Scottish East coast, but that there is significant evidence of heterogeneity along all other stretches.

To accommodate both the observed variation in the trend estimates over data sites and the expected coherence in the underlying trend, an estimator which is both smooth and flexible is required. Here this is obtained using a local linear weighted least squares estimator which is then smoothed. Let d_i be the anti-clockwise distance of site i from Ullapool, with $d_1 < d_2 < ... < d_{n_s}$, where n_s is the number of sites in the study, and let β_i

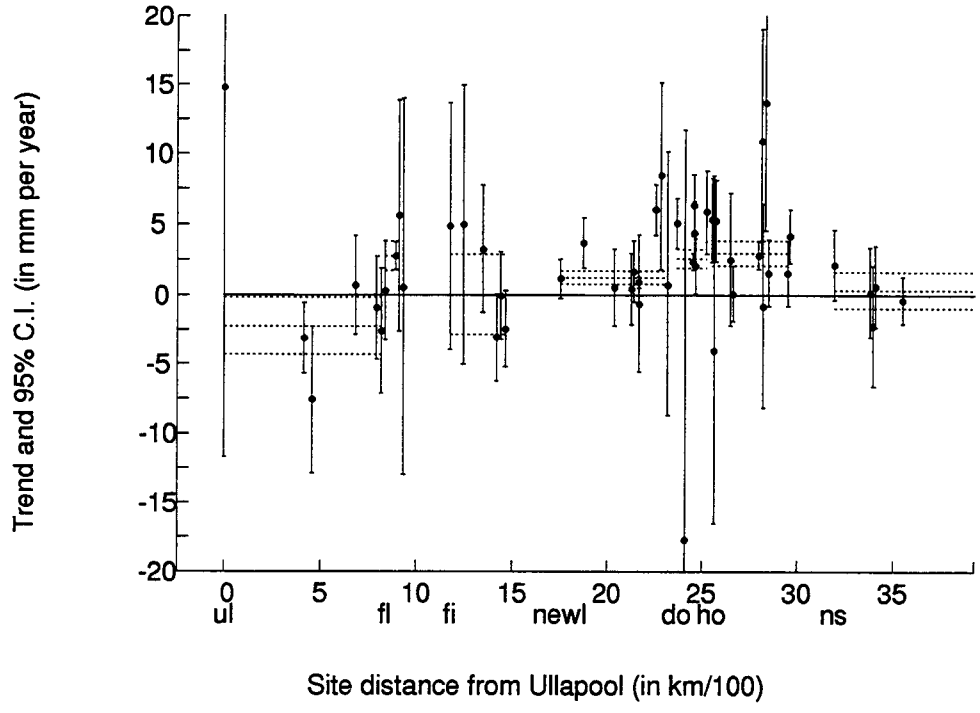

Figure 7.4: Trend estimates with 95% confidence intervals obtained by the spatial annual maxima method described in Section 7.2.2. The abscissa is the same as in Figure 7.3. The improvement in spatial smoothness is seen, particularly at Bembridge, Swansea, and North Shields as these now have trend estimates which are consistent with those at neighbouring sites. Georges Pier, Princes Pier and the Humber estuary sites are not included here. The estimate and 95% confidence bands for the piecewise homogeneous model, obtained in Section 7.2.3, are shown by the broken lines.

be the trend for site i. Owing to the cyclic nature of the problem, β_i and d_i are defined for all i by

$$\beta_i = \beta_{i \bmod n_s}, \quad d_i = c(i - i \bmod n_s)n_s^{-1} + d_{i \bmod n_s},$$

where c is the coastal length ($c = 3935$ km). Define the *neighbourhood* of site i, $N_m(i)$ as

$$N_m(i) = \{j : i - m \leq j \leq i + m\},$$

where m, an integer, is to be chosen. Then the local linear least squares model applied to site i is

$$\beta_j = \gamma_{i,0} + \gamma_{i,1} d_j + \epsilon_j \qquad \text{for all } j \in N_m(i),$$

where the ϵ_j are taken as independent random variables with zero mean and variance

v_j^2 $(v_j = v_{j \bmod n_s})$. Here v_j is the standard error of the spatial model trend estimator for site j. To account for the different precision of the separate trend estimates, weighted least squares (Rawlings, 1988, Section 11.5.1) is used to obtain estimates $\hat{\gamma}_{i,0}$, $\hat{\gamma}_{i,1}$ and $\hat{\beta}_i$ and the associated standard error of $\hat{\beta}_i$.

We apply this procedure separately to each site, taking $m = 3$. The choice of m is made subjectively in that a large choice for m gives an estimate that looks over-smoothed, and a smaller m leads to a rough spatial estimate. This gives the spatial estimate, displayed in Figure 7.5, which shows significant spatial variation in trends around the coast. Large trends are observed along the south east coast, and small or negative trends can be seen around the Scottish coast. Note that the confidence bounds shown are not confidence *bands* in the sense that the spatial estimate will lie within them 95% of the time, but pointwise confidence bounds which give, for any position, the associated 95% confidence interval for the trend.

A clearer general impression of trends around the coastline is given by smoothing the local least squares estimate using a kernel smoother with a Gaussian kernel function (Silverman, 1986, Section 2.4). The estimate together with confidence bounds, obtained by the same technique, are shown in Figure 7.6.

7.2.4 Estuarine sites

Estuarine sites are examined separately since, as suggested by Coles and Tawn (1990), their trends are typically larger than those from neighbouring sites. In particular, Princes Pier, Georges Pier and the Humber estuary sites are not included in the spatial analysis presented in Section 7.2.3. A separate fit is obtained for sites along the Humber estuary, see Figure 7.7. The fitted homogeneous trend for the Humber estuary is 5.6mm per year (with standard error 1.1mm per year) in comparison with 2.8mm per year predicted for the neighbouring coastline. The Humber estuary is used as an example in application of another spatial model in Chapter 9 and estimation of the trend is considered further there.

7.3 Comparison with crustal movement

In this section, the influence of vertical land movement on the spatial variation in extreme sea-level trends is examined. Based on geological data, Shennan (1989, Figure 9) gives land trend estimates throughout Britain for the late Holocene period, in the form of a contour map. By digitising the contour plot, and interpolating the estimates using a cubic spline, a smooth spatial estimate for the entire coastline is obtained. Shennan and Woodworth (1992) take the land trend estimates to be stable through time and valid over

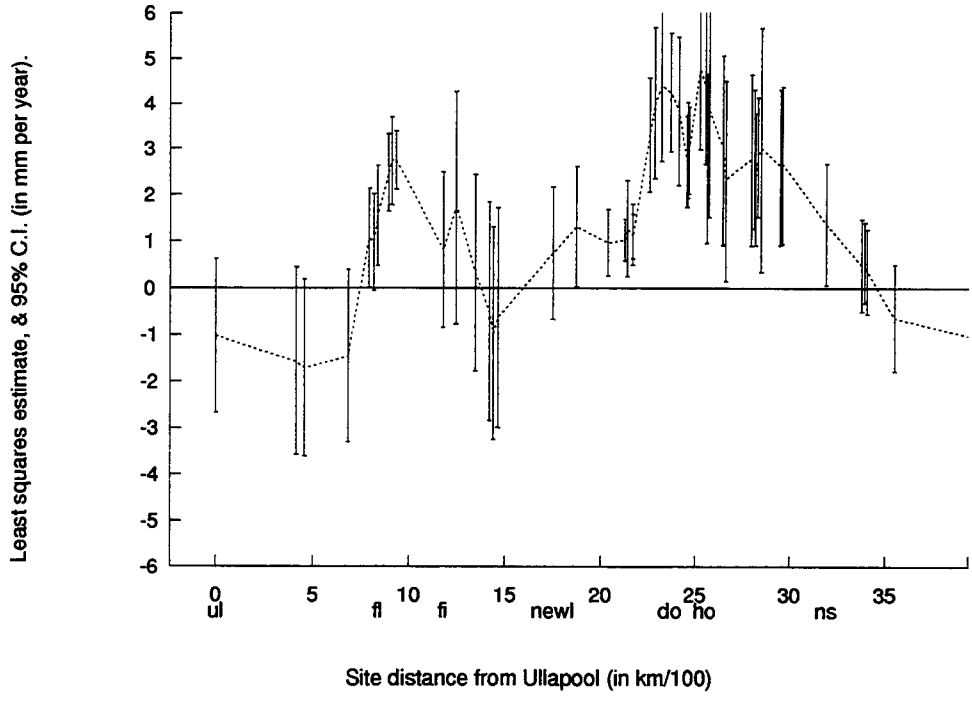

Figure 7.5: Spatial trend estimates and 95% confidence bounds obtained by the local least squares procedure described in Section 7.2.3. The abscissa is the same as in Figure 7.3. The estimates have been linearly interpolated to give a clearer impression.

the period of the tide gauge data. The behaviour of eustatic extreme sea-level trends therefore can be examined by adding the smooth vertical land movement trend estimate to the kernel estimate obtained in Section 7.2.3. The choice of interpolation used for the vertical land trend estimates has little effect on the eustatic extreme sea-level trend estimate. The result shown in Figure 7.8, is approximately constant, 1.0 to 1.2mm per year, suggesting that the spatial variation in UK observed extreme sea-level trends is almost entirely due to land movement trends. These findings are similar to those for UK eustatic mean sea-level trends (Shennan and Woodworth, 1992 and Woodworth and Jarvis, 1991) and are consistent with estimates for global mean sea-level rise obtained by Gornitz and Lebedeff (1987) and other researchers, see Robin (1986, Table 7.1); although more recent estimates are slightly larger than this, for example Trupin and Wahr (1990) who obtain an estimate of 1.1 to 1.9mm per year. Throughout the study, it is important to note that both the estimated eustatic mean sea-level trends of Shennan and Woodworth (1992), and the eustatic extreme sea level trends in Figure 7.8 may contain a systematic

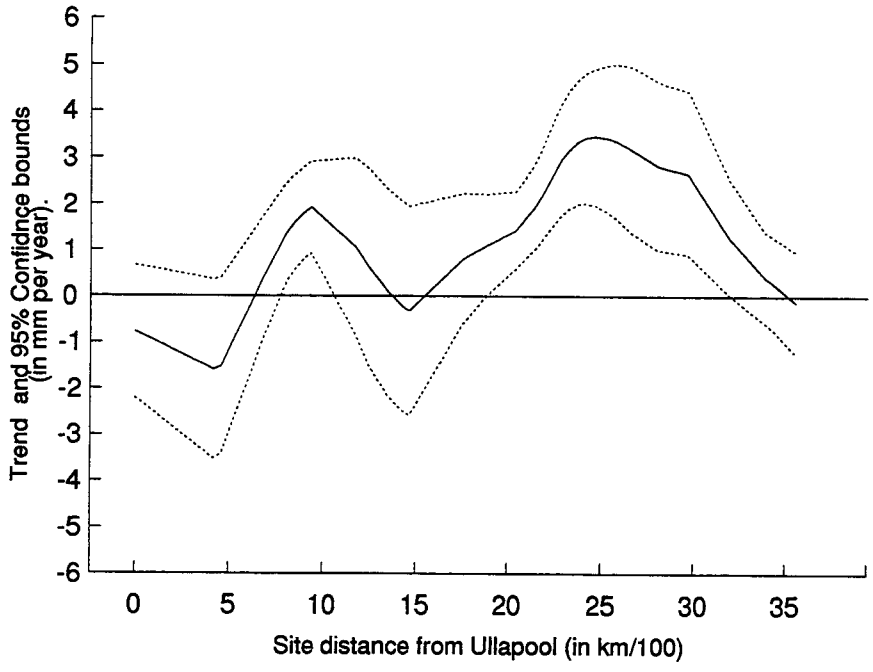

Figure 7.6: Trend estimates, obtained using a kernel smoothing of the local least squares estimates, as described in Section 7.2.3. The abscissa is the same as in Figure 7.3. The clear spatial pattern of trends around the UK is evident from this figure; observed trends are large and positive in the South and South East, and small or negative around the Scottish coast.

error due to the Holocene sea-level change model used to infer land movements from geological data (Shennan, 1989). Although this method of estimating land level trends is not very accurate, it is the best available method at present.

7.4 Discussion

7.4.1 Comparison with mean sea-level trends

The results obtained in Section 7.3 suggest that UK extreme sea-level and UK mean sea-level trends are similar in both spatial behaviour and value; that is there is no significant evidence for effects due to changes in characteristics of extreme storms or tidal range, the latter being in agreement with Woodworth *et al.* (1991). Similar results are obtained for Australia (Tawn and Mitchell, 1994). This suggests that current information about trends

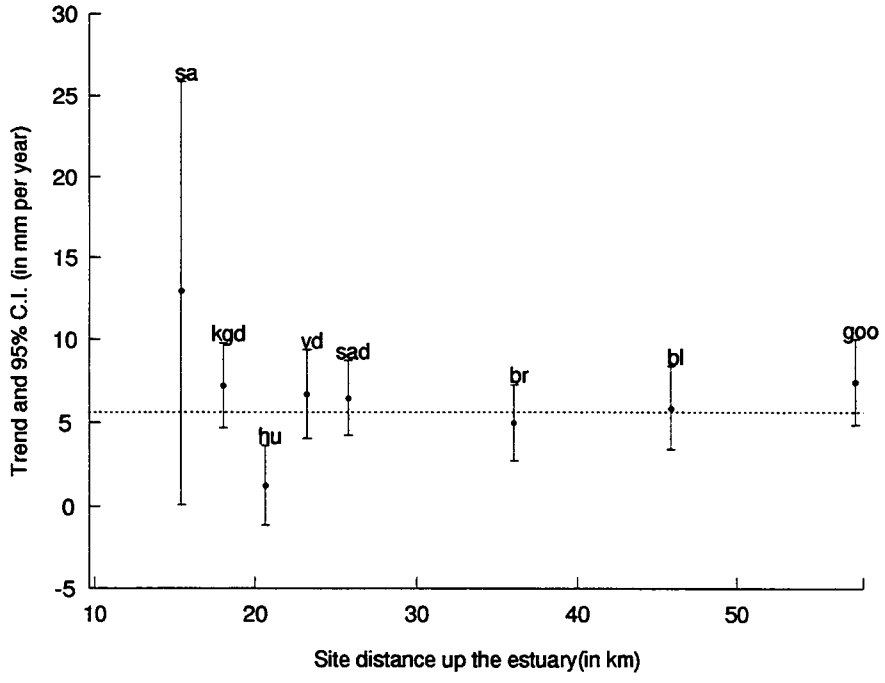

Figure 7.7: Trend estimates for sites in the Humber estuary obtained from the spatial model of Section 7.2.2 plotted against distance up the estuary. The estimate of the homogeneous trend for the sites is indicated by the broken line. Site name abbreviations are: sa (Saltend Jetty); kgd (King Georges Dock); hu (Humber Dock); vd (Victoria Dock); sad (St Andrews Dock); br (Brough); bl (Blacktoft); and goo (Goole).

in mean sea-level is applicable to trends in extreme sea-level and vice-versa, although in the event of climate change, this need not remain the case. This finding motivates the study in Chapter 8 where mean sea-level trends are examined on a spatial scale.

Further evidence for the equivalence of mean sea-level and extreme sea-level trends is provided by the separate comparison of trend estimates obtained from mean and annual maximum data series at each site. Accounting for the precision of the estimates of mean and extreme sea-level trends, there are no systematic differences between the two. For example, at Sheerness (with 42 observations of annual mean sea-level over the period 1901–1963, and 136 observations of annual maximum sea-level over the period 1819–1983) the estimated mean sea-level trend is 2.36 (0.17)mm per year whereas the estimated extreme sea-level trend using the spatial model is 2.36 (0.30) and 2.73 (0.51)mm per year, obtained with and without local trend smoothing respectively. The figures in parentheses

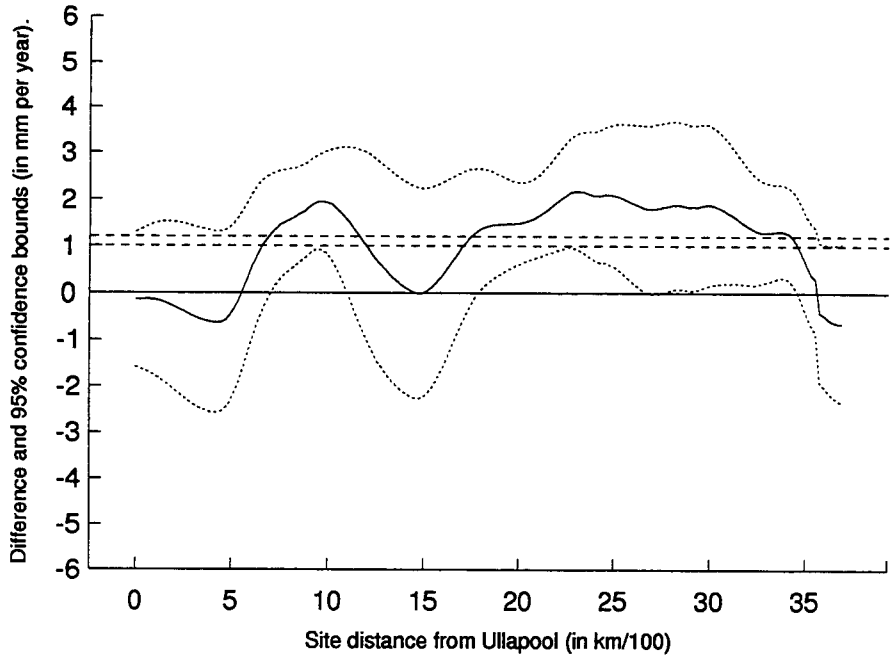

Figure 7.8: Eustatic extreme sea level trends obtained by adding vertical crustal movement to the kernel estimate of observed extreme sea level trends. The abscissa is the same as in Figure 7.3. Homogeneous eustatic extreme sea level trend estimates between 1.0 and 1.2mm per year lie inside the pointwise 95% confidence intervals for the entire coastline. These values are shown on the plot.

are the associated standard errors. A possible reason for the slight, but non–significant, increase under local smoothing is due to large trends at neighbouring sites (Tower Pier and Tilbury) in the Thames Estuary. With hindsight these sites should possibly have been classified as estuary sites.

Assuming trends in mean sea-level and extreme sea-level are the same, the trend estimator can be based on either data set. The main advantage of using mean sea-level data is that the inter–annual variation is smaller than in annual maximum data so, for a site with the same length record in both the mean and the maximum series, a more accurate trend estimator is obtained. However the extra data processing involved to obtain mean sea-level data and the requirement of high quality observations throughout the year means that historical archives often contain annual maximum levels but not annual mean levels and consequently there are more annual maximum than annual mean

data. Thus the extreme sea-level approach provides both increased temporal information and greater spatial coverage of the coastline.

Possible reservations about the inclusion of these extra sites for trend estimation are that either the tidal datum or the observations may be of poor quality. These potential errors should have only a small effect on the spatial trend estimate obtained in Section 7.2.3. This is because each additional site appears to have a trend which is consistent with a neighbouring site for which the tidal datum is known to be of high quality, and because the spatial smoothing reduces the effect of bad quality data from any individual site.

Currently, there is concern that the rate of rise of global sea-level may be increasing as a result of climate change and the greenhouse effect, so there is a need to study accelerations in sea-levels as well as linear trends. The result of Woodworth (1990), that there is no evidence of acceleration in mean sea-level, need not apply to extreme sea-levels since accelerations in the eustatic component of extreme sea-levels may differ from those in mean sea levels even if the linear trend component for each variable is the same. As a preliminary test for accelerations in the eustatic component of extreme sea-levels, the spatial model is used to test for a homogeneous quadratic term in the trend within each of the coastal stretches defined in Table 7.1. Based on likelihood ratio tests, no significant evidence is found for accelerations in extreme sea-level trends for any of the stretches. Because of the large inter–annual variability of annual maximum levels, it could be many years before evidence of extreme sea-level accelerations becomes significant from observational data. This is considered further, using mean sea-level data, in Chapter 8.

7.4.2 Spatial averaging

In mean sea-level trend studies, a summary statistic of interest is the spatial average eustatic mean sea-level trend. Two possible causes of bias in estimates of this statistic are a non–uniform distribution of the data sites, and a non–zero average vertical land movement trend over the region under consideration. Estimates obtained by averaging site values of mean sea-level trend are dominated by regions of highest site density, so biased estimates would arise if that region was atypical (Barnett 1984). Various approaches aimed at overcoming these problems have been proposed: Gornitz and Lebedeff (1987) and Shennan and Woodworth (1992) use geological data from nearby the tide gauges to infer land movements, while Peltier and Tushingham (1989) and Trupin and Wahr (1990) employ geodynamic models of the Earth. Woodworth (1987) and Woodworth and Jarvis (1991) side–step the problem by studying simply relative sea-level trends between sites, in effect relative vertical land movements.

The summary statistic for the UK extreme sea-level trends is estimated here by viewing

the trend as a function of coastal distance from a site. The required average is then given by the integral of the associated smooth spatial trend estimate (Figure 7.6), normalised by coastal length, which gives a value of 1.0mm per year. A better estimate is obtained by first adding vertical land movements, and then integrating the resulting smooth spatial trend estimate (Figure 7.8), normalised by coastal length, producing an average value of 1.1mm per year. Thus in the UK there seems to be only a small bias due to the average vertical land movement trend being non–zero, although generally the bias could be much larger. The effect of the non–uniform site distribution is substantial for the UK sites in this study where, for example, averaging site values gives an average trend of 1.9mm per year. This over–estimation is due to most sites being in the South East where trends are largest, as is clear from Figure 7.5. This further emphasises the importance of this spatial approach.

7.5 Conclusions

In this case study the spatial coherence of the sea-level process has been exploited to aid estimation of trends in sea-level annual maximum data. A clear picture of the general pattern in extreme sea-level trends around the UK coastline is obtained on the basis of data from 62 sites.

Extreme sea-level trends are found to have two principal components: changes in eustatic mean sea-level and in land level. The former appears to be homogeneous (approximately 1.1mm per year) over the entire open coastline; whereas the latter has significant spatial variation in the form of North-South tilting of the UK consistent with glacial isostatic uplift in the North and land subsidence in the South and South East. Finally, there is some evidence to suggest that estuary sites have different trends than sites along the neighbouring open coastline. The next case study investigates these effects using mean sea-level data within a spatial model.

Throughout the study only annual maximum observations have been analysed even though extensive hourly sea-level data are available at some of the sites. The analysis could potentially be extended to use all the largest independent levels from each year to give a more robust estimator with increased precision. Smith (1986, 1989) and Tawn (1988) explain how this is achieved marginally and in Chapter 9 a spatial method, which uses more data than only the maximum, is described and applied to data from the Humber estuary.

Chapter 8

Case study II: mean sea-level trends

8.1 Introduction

The extreme sea-level trend at a site consists of the mean sea-level trend plus the trend in the behaviour of extreme storms. Results in Chapter 7 suggest that, for most UK sites, the trend in extreme storms is negligible so that the extreme sea-level trend is just the mean sea-level trend. In this section a spatial model for mean sea-levels is developed for UK east coast data.

Since the mean sea-levels arise from averaging the sea-level process, the sampling variability for mean sea-level data is typically much smaller than for extreme data. Consequently using mean sea-level data should lead to an increased precision of trend estimates over estimates based on extreme sea-level data provided that a suitable model can be developed. Since mean sea-level data arise from an averaging process they have a marginal distribution and joint characteristics over a set of sites which are naturally described by a relatively simple parametric statistical distribution. Using such a model for mean sea-levels over sites leads to three important results.

1. **Increased information on extreme sea-level trends.**
   The purposes of examining mean sea-levels in this study is to obtain extra information about extreme sea-level trends so that adjustments to future design levels can be more accurately predicted. Trends in mean sea-levels are similar to trends in extreme sea-levels yet can be estimated more precisely. It follows that spatial modelling of mean sea-level trends gives a clearer picture of both the spatial variation and magnitude of extreme sea-level trends along a coastline; here the UK east coast.

2. **Testing for non-linear trends**
   The flexibility of the mean sea-level model allows information about any form of

acceleration in trends to be pooled from all the sites along a coastline. This combination of information increases the ability to assess evidence for accelerations in trends.

3. **Estimation of regional/global trends**
   By incorporating the spatially varying estimated land level trend, obtained from Shennan (1989), the mean sea-level model can be used to provide an estimate of the homogeneous eustatic sea-level trend along the UK east coast. The eustatic trend is of interest to oceanographers studying sea-level changes on regional and global scales.

The breakdown of the work in this case study on spatial modelling of mean sea-level trends is as follows: the data are east coast UK mean sea-level data and are described in Section 8.2. Application of a simple univariate model to these data is described in Section 8.3. In Section 8.3 the gains of using a spatial model are highlighted by reference to the separate site applications. In Section 8.4 a simple analytical example is used to show how information is transferred using the spatial model, and a multivariate model is applied to the east coast data. Evidence for accelerations is assessed and eustatic trends are examined.

As measurements of still water levels are made using tide gauges, or some equivalent land based measuring equipment, the reference level is the local land level. As discussed in Chapter 7 this is not constant as land levels also have trends due to geological effects. We investigate this for mean sea-level trends using the same land level trend estimate given in Section 7.3. Finally, since mean sea-level data can sometimes have datum shifts due to poor tide gauge management, in Section 8.5, we examine models which could be used to reduce the impact of such changes on trends.

8.2 Mean sea-level data

As described in Chapter 2, for any site, the observed sea-level process, Z_t, at time t, is the composition of four physically distinct processes: mean sea-level, M_t; astronomical tide, X_t; surge, Y_t; and surface waves, W_t, such that

$$Z_t = M_t + X_t + Y_t + W_t. \tag{8.2.1}$$

The mean sea-level, M_t, is a low frequency component, with variations on an annual scale; W_t is high frequency, with variations of the order of a few seconds; Y_t and X_t are intermediate in terms of frequency. The surge, tide, and wave components are defined to have zero mean, so that M_t incorporates the mean for these components.

The mean sea-level, M_t, at time t, is derived by analysis of the observed still water levels to produce a series of annual estimates. This is achieved typically by either taking annual averages of observations made regularly in time, or by first using a low-pass numerical filter to largely remove the effects of semi-diurnal and diurnal tides before annual averaging (Pugh, 1987).

For each site, annual estimates of mean sea-levels are obtained using some form of averaging over an annual series of still water levels. Such data are available from the Permanent Service for Mean Sea-Level, based at the Proudman Oceanographic Laboratory, which holds a large data-base of annual mean sea-levels from over 1300 sites world-wide (Pugh *et al.*, 1987; Woodworth *et al.*, 1990).

In this case study the east coast of the UK subset of this data-base is used. Figure 8.1 shows the location of these sites. The length and span of the data series varies over sites; Figures 8.2-8.5 give a graphical summary of the data with sites plotted in order of increasing distance from Wick. Although the quality of the data is generally high, possible systematic errors can arise from either the handling of the tide gauge; timing errors due to consistent incorrect positioning of the chart on the drum of the gauge; or from datum control errors. The first two sources of systematic error should be identifiable in preliminary data analysis before the mean sea-level data are entered to the data-base, although datum errors are likely to be small and may not be identified. As small datum shifts may be highly influential in trend estimation, methods which protect against this potential source of error should be employed; this feature is addressed in Section 8.5.

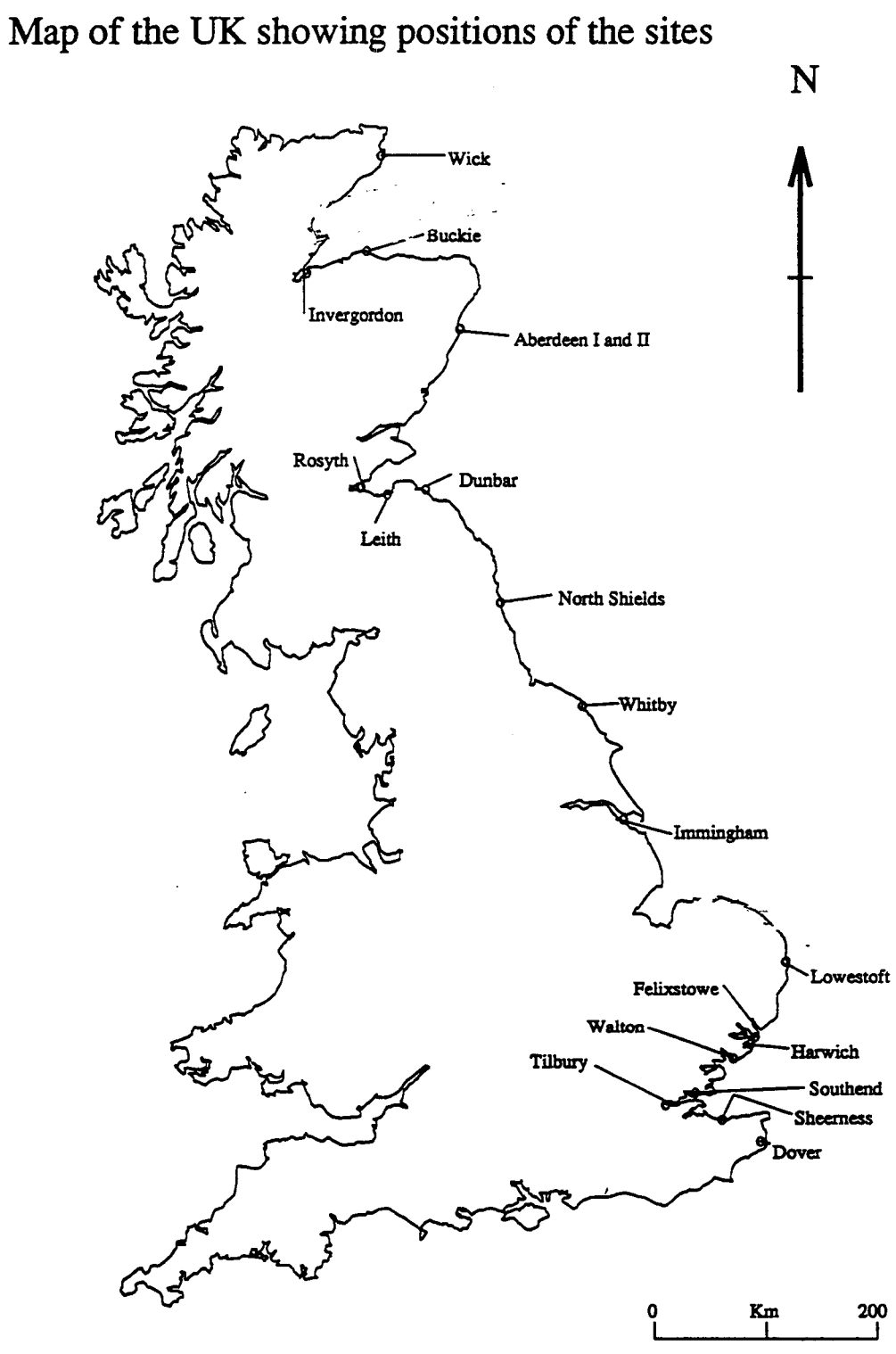

Figure 8.1: Map showing the locations of the UK mean sea-level sites on the east coast.

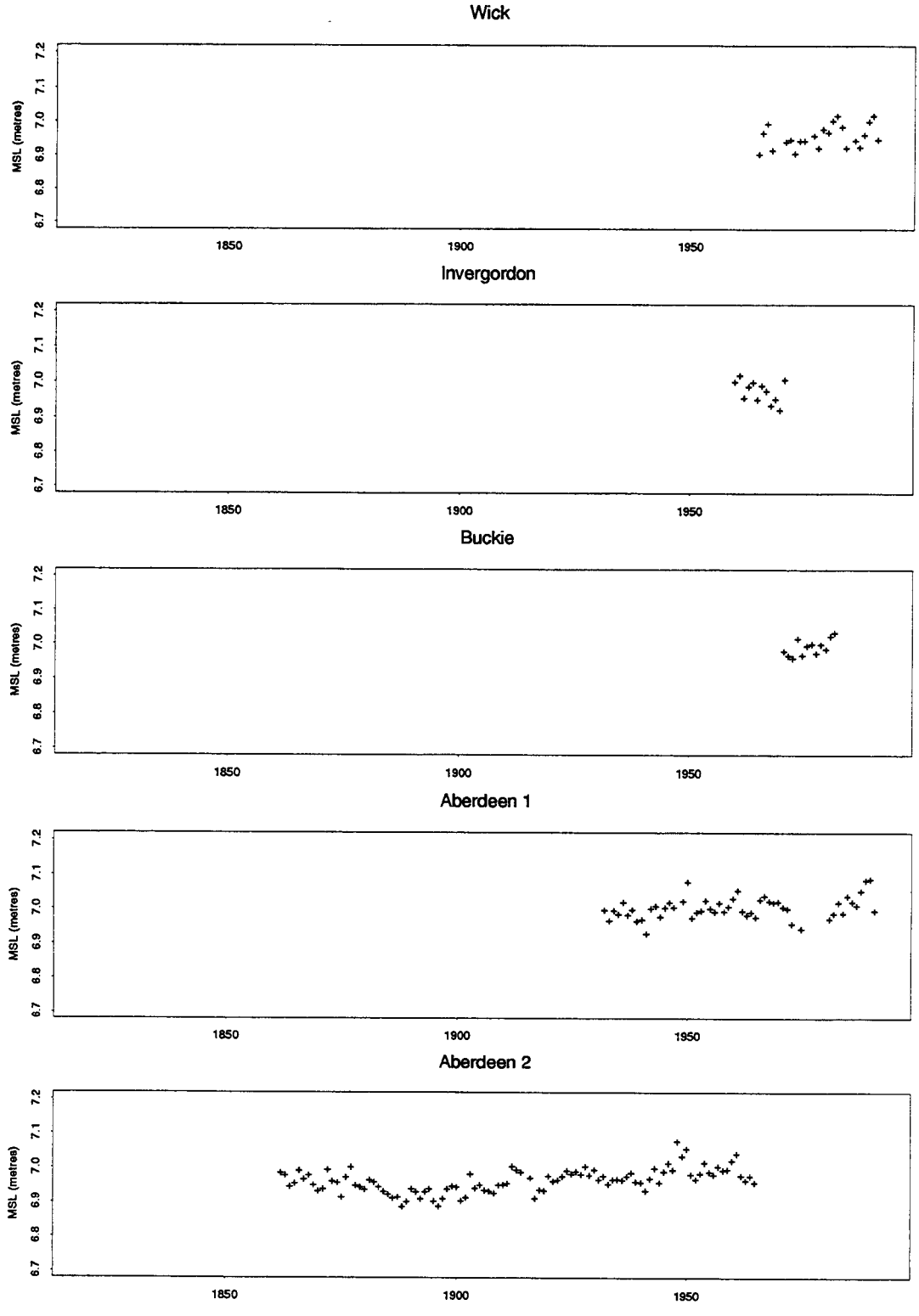

Figure 8.2: The mean sea-level data, in metres, plotted using common axes, for the UK east coast data sites from Wick to Aberdeen. The data are plotted relative to the Revised Local Reference level.

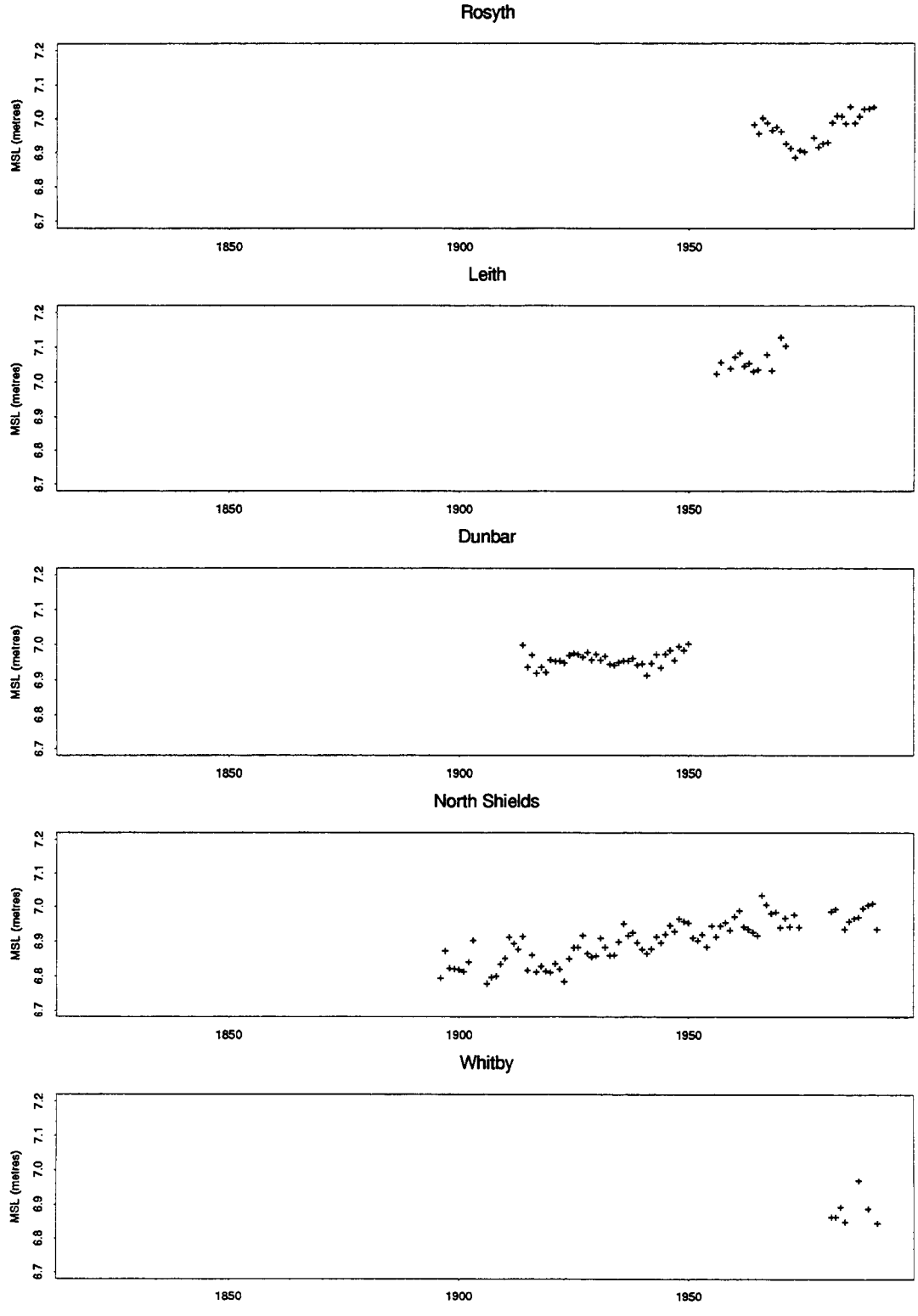

Figure 8.3: The mean sea-level data, in metres, plotted using common axes, for the UK east coast data sites from Rosyth to Whitby. The data are plotted relative to the Revised Local Reference level.

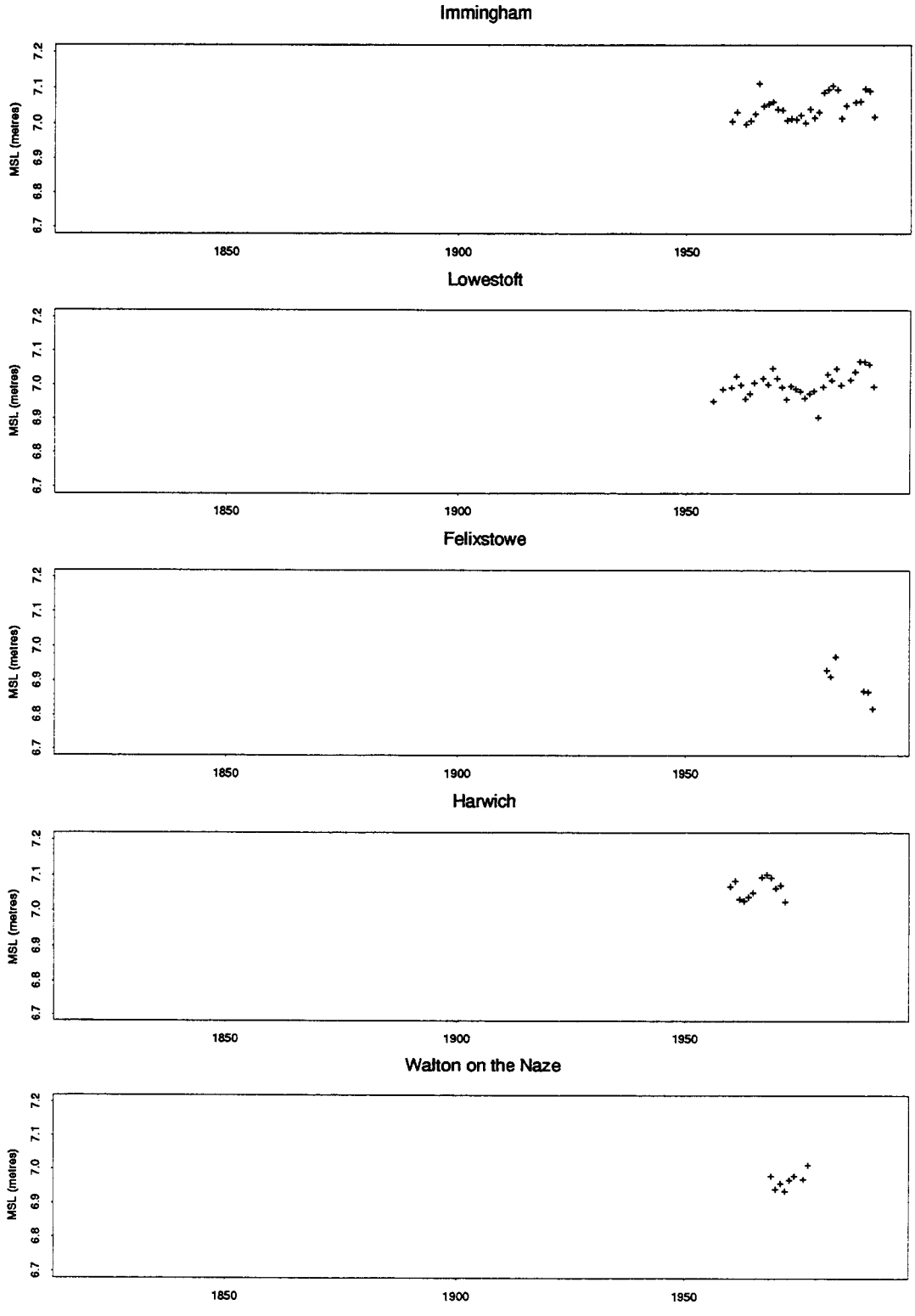

Figure 8.4: The mean sea-level data, in metres, plotted using common axes, for the UK east coast data sites from Immingham to Walton. The data are plotted relative to the Revised Local Reference level.

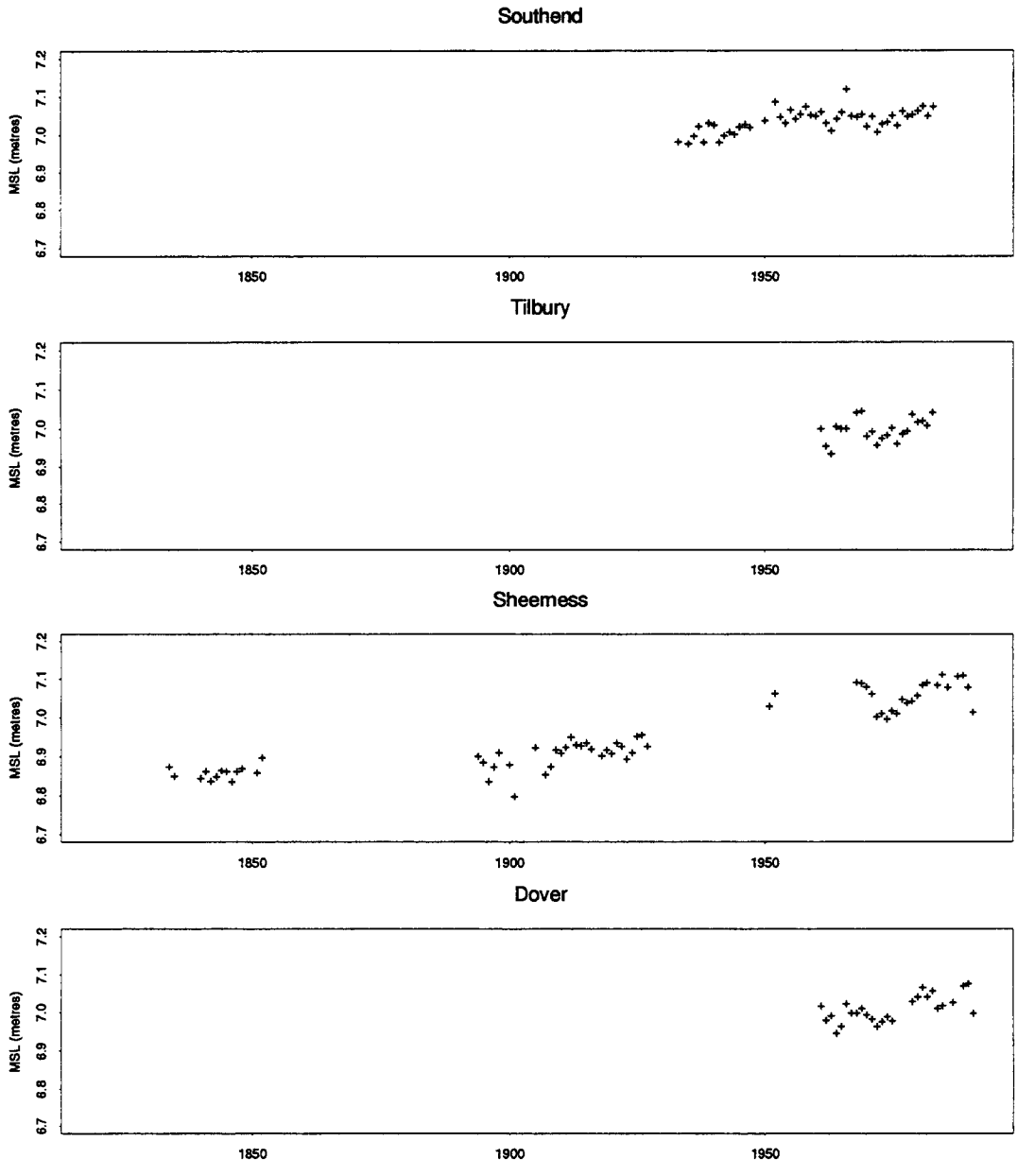

Figure 8.5: The mean sea-level data, in metres, plotted using common axes, for the UK east coast data sites from Southend to Dover. The data are plotted relative to the Revised Local Reference level.

8.3 Existing marginal analyses

Existing procedures for estimating mean sea-level trends are marginal analyses in that only data from the site of interest are used. In this section we outline a simple procedure for marginal mean sea-level trend estimation, and illustrate it by application to the data shown in Figures 8.2-8.5. This motivates the spatial model of the next section. Other methods which have been applied to mean sea-level data are briefly reviewed at the end of the section.

From (8.2.1) the annual average still water level is of the approximate form

$$\frac{1}{k}\sum_{t=1}^{k}(M_t + X_t + Y_t) = \overline{M} + \overline{X} + \overline{Y}, \tag{8.3.1}$$

where k is the annual number of observations and M_t, X_t, and Y_t are as in (8.2.1). Thus from (8.3.1) an annual mean sea-level observation consists of a signal, in terms of the long term mean sea-level $\overline{M}$, with the error being the annual average tide and surge, $\overline{X} + \overline{Y}$. By definition of the processes involved, the long term mean of $\overline{X} + \overline{Y}$ is zero, so the error in $\overline{M}$ is zero mean. First we setup some notation which is used throughout the case study. One model for the observed annual mean sea-level in year t at site position x is

x	the convex coastal distance, in km, from Wick,
t_0	base year of 1800,
$Z_t(x)$	observed annual mean sea-level,
$M_t(x)$	the mean sea-level signal,
$\epsilon_t(x)$	a temporally independent and identically distributed normal error, with zero mean and variance $\sigma^2(x)$.

$$Z_t(x) = M_t(x) + \epsilon_t(x). \tag{8.3.2}$$

with signal $M_t(x)$ corresponding to $\overline{M}$, and white noise error structure. The error term arises from an averaging operation so arguments based on the central limit theorem are used to justify taking $\epsilon_t(x)$ to be normally distributed.

Two forms of model for the mean sea-level component that have been explored are a linear model

$$M_t(x) = \alpha(x) + \beta(x)t, \tag{8.3.3}$$

and a model which incorporates an acceleration term either via a piece-wise linear form, considered by Woodworth (1990), or a quadratic model

$$M_t(x) = \alpha(x) + \beta(x)t + \gamma(x)t^2. \tag{8.3.4}$$

These models for $M_t(x)$, which are based on the environmental and geological processes which govern trends, are simple and yet capture all the relevant features of the mean sea-level process.

Thompson (1980) and Woodworth (1987) fitted this model with the linear trend, (8.3.3), by maximum likelihood, and their approach is illustrated using our UK east coast data. The data are shown in Figures 8.2-8.5 with sites displayed in order of increasing distance x. Estimates of intercept, $\alpha(x)$, trend, $\beta(x)$, and standard deviation, $\sigma(x)$, obtained from fitting the model with linear trend, are shown in Figure 8.6, where estimated trends are typically between $0 - 3$mm/yr. There is some regularity in the estimated $\beta(x)$, with generally larger estimates in southern England and smaller values along the Scottish coast, i.e. for large and small x respectively. For many sites, estimates of the trend parameters, and the corresponding intercepts, are quite poor, but exhibit negative association which is typical between these parameter estimates.

Woodworth (1987) applied this to only a small subset of these sites since there were worries over quality and datum control, and the length of the record. In particular, the pairs of sites Aberdeen I and II, and Sheerness and Southend were each combined to provide an increased sample. Although this seems a potential source of bias, Figure 8.6 shows there is sufficient spatial smoothness in estimates to pool local data. A potentially more efficient approach is to preserve the spatial nature of the data by fitting explicit spatial models, a theme which is developed in Section 8.4.

As discussed in Chapter 7, the trend in the sea-level, as measured by tide gauges, is the sum of the two components, eustatic sea-level trend and land level trend. Within our region the eustatic sea-level trend should be homogeneous over all sites, and the land level trend varies with site position. Therefore we take the sea-level trend, $\beta(x)$, to be

$$\beta(x) = \beta_l(x) + \beta_e \tag{8.3.5}$$

where $\beta_l(x)$ is the land level trend at position x and β_e is the eustatic sea-level trend. Shennan and Woodworth (1992) use (8.3.5) together with estimates of $\beta_l(x)$, obtained by Shennan (1989), to obtain a series of β_e estimates, which are then averaged to give a regional estimate of eustatic sea-level trend. Some authors estimate eustatic mean sea-level trends from spatial averages of estimates of $\beta(x)$ which may lead to bias as discussed in Section 7.4. Whichever approach is used, owing to the spatial dependence of the data, estimates which exploit the spatial nature of the problem should provide an improved estimate, together with an unbiased estimate of the standard error.

Finally consider the quadratic model (8.3.4). Both Woodworth (1990) and Douglas (1992) fit this form in each case finding no evidence of accelerations, i.e. their estimates of $\gamma(x)$ do not differ significantly from zero and show no spatial pattern. Since land level trends are linear over the observational period, and acceleration in eustatic sea-levels, if

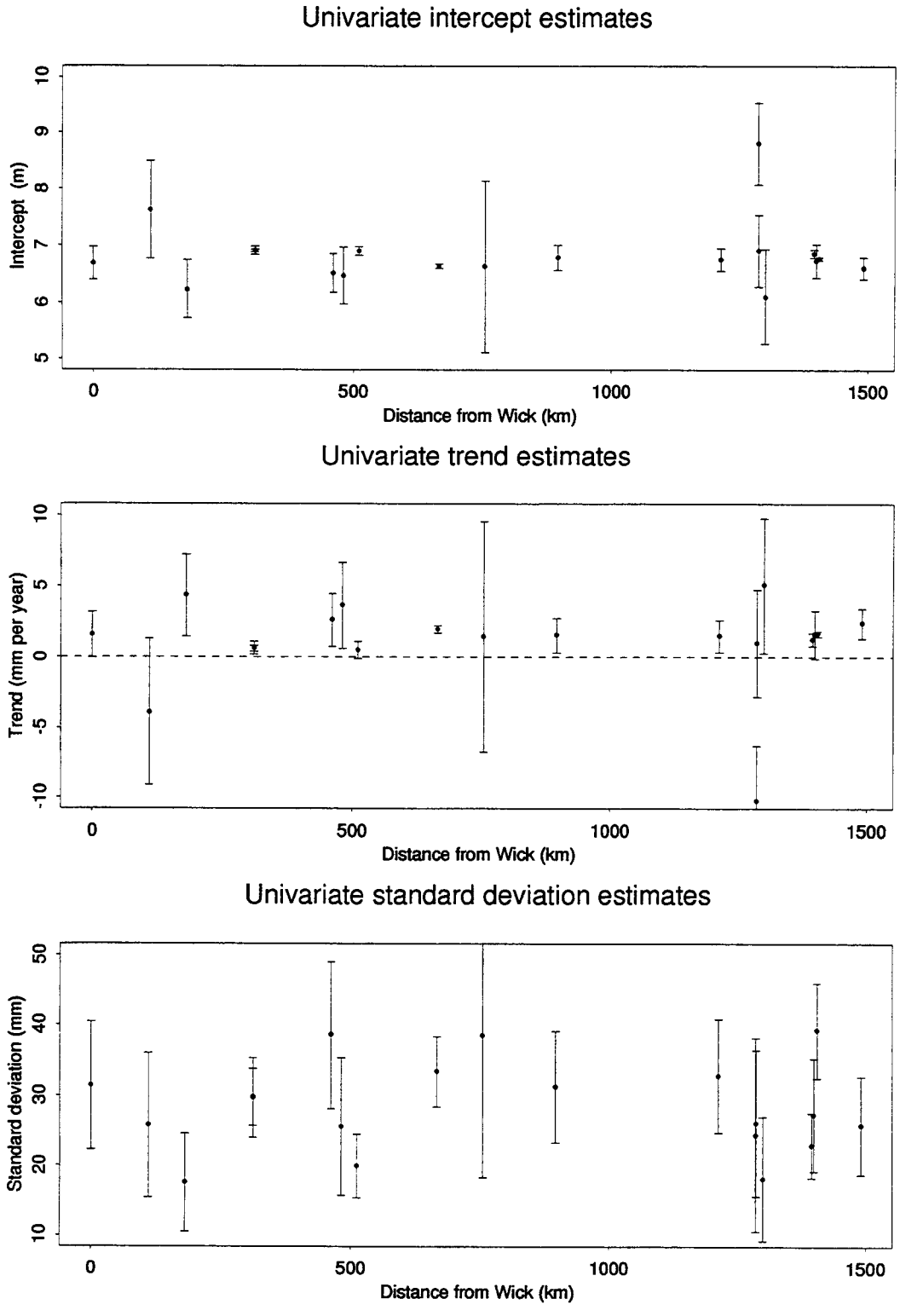

Figure 8.6: Marginal model based estimates, together with 95% confidence intervals, of $\alpha(x), \beta(x), \sigma(x)$ plotted against coastal distance from Wick, x.

present, will be homogeneous over the region, then $\gamma(x)$ should be regionally homogeneous, i.e.

$$\gamma(x) = \gamma. \tag{8.3.6}$$

This feature is not exploited by the existing analyses, but can be by spatial analyses. This is important since evidence of accelerations from any one site is likely to be extremely weak, yet viewed spatially may be significant.

This approach to modelling mean sea-level trends is the simplest possible as it ignores tidal variation within tides and additional information provided by numerical models of the surge process over recent years. Models to incorporate such information are discussed in Tawn *et al.* (1994).

8.4 Spatial analyses

In this section we examine the joint distribution of mean sea-levels over data sites and assess the benefits of treating the problem spatially through a simple analytical example together with an application of a multivariate model to the UK east coast data.

The model in Section 8.3 is extended by introducing the joint distribution, over the d sites, of the errors $\{\epsilon_{t,i} = \epsilon_t(x_i) : i = 1, \ldots, d\}$, where x_i denotes the position of site i. Again the central limit theorem justifies the use of a multivariate normal distribution, with variance-covariance matrix $\Sigma = (\sigma_{ij})_{i,j=1,\ldots,d}$. Then marginally, for site i the annual mean sea-level, $Z_{t,i} = Z_t(x_i)$, follows a normal distribution with mean

$$M_{t,i} = M_t(x_i) = \alpha(x_i) + \beta(x_i)t = \alpha_i + \beta_i t$$

and variance σ_{ii}.

The benefits of a spatial analysis arise from combining information over sites. It is most apparent when there are missing data, as in Figures 8.2-8.5, since there is a transfer of information between sites which partially compensates for the loss of data. Here we illustrate this analytically by using a simple bivariate case ($d = 2$) for model (8.3.2) with linear form (8.3.3). Let the annual mean sea-levels at the sites be $Z_{t_1,1}, \ldots, Z_{t_n,1}$ and $Z_{t_1,2}, \ldots, Z_{t_m,2}$, where $m < n$ and the t_i are year indices, so site 2 has missing data in relation to site 1. Take ρ to be the correlation between $Z_{t,1}$ and $Z_{t,2}$, i.e. $\rho = \sigma_{12}/(\sigma_{11}\sigma_{22})^{1/2}$, and $(\hat{\alpha}_k, \hat{\beta}_k)$ to be the maximum likelihood estimates of (α_k, β_k) for $k = 1, 2$, then it can be shown that

$$\hat{\beta}_1 = \frac{S_{t_n 1}}{S_{t_n t_n}} \quad \text{and} \quad \text{Var}(\hat{\beta}_1) = \frac{\sigma_{11}}{S_{t_n t_n}}$$

and that $\hat{\beta}_2$ satisfies

$$\hat{\beta}_2 = \frac{S_{t_m 2}}{S_{t_m t_m}} + r_m \frac{S^*_{t1}}{S_{t_m t_m}} \quad \text{and} \quad \text{Var}(\hat{\beta}_2) = \frac{\sigma_{22}}{S_{t_m t_m}} \left[1 - \rho^2 \left(1 - \frac{S_{t_m t_m}}{S_{t_n t_n}} \right) \right]. \tag{8.4.1}$$

Here, for $k = 1, 2$,

$$S_{t_j t_j} = \sum_{i=1}^{j} (t_i - \bar{t}_j)^2, \quad S_{t_j k} = \sum_{i=1}^{j} (t_i - \bar{t}_j)(z_{t_i,k} - \bar{z}_{j,k}), \quad S^*_{tk} = \sum_{i=m+1}^{n} (t_i - \bar{t}_m)(z_{t_i,k} - \hat{M}_{t_i,k})$$

and

$$r_m = [\sum_{i=1}^{m} (z_{t_i,1} - \hat{M}_{t_i,1})^2]^{-1} \sum_{i=1}^{m} (z_{t_i,1} - \hat{M}_{t_i,1})(z_{t_i,2} - \hat{M}_{t_i,2}),$$

with

$$\bar{t}_j = j^{-1} \sum_{i=1}^{j} t_i, \quad \bar{z}_{j,k} = j^{-1} \sum_{i=1}^{j} z_{t_i,k} \quad \text{and} \quad \hat{M}_{t,k} = \hat{\alpha}_k + \hat{\beta}_k t.$$

The trend estimate and standard error for the longer Z_1 series are unaffected by knowledge about the shorter Z_2 series, whereas the trend estimate for the Z_2 series is the marginal estimate adjusted, by the second term in equation (8.4.1), for transferred information from the Z_1 series. The elements of the second term are of interest: S^*_{t1} measures how unrepresentative the Z_1 series is over the additional observation period relative to during the overlapping observation period; r_m is a measure of dependence between the two series during the overlapping period; and $S_{t_m t_m}$ a measure of the information from the overlapping period. If either the additional observations are unrepresentative, there is no correlation in the overlapping series, the period of overlapping data is large, or the residual variance of the Z_1 series is large relative to the Z_2 series then there is limited information transfer in terms of the trend estimate value. However if there is non–zero correlation then $\text{Var}(\hat{\beta}_2)$ is reduced from the variance of the marginal trend estimator.

In the general multivariate case the analytical expressions are more complex, and consequently more difficult to interpret, but information continues to be transferred, see Johnston (1984) and Zellner (1962, 1963, 1972). This is now illustrated by application to the UK east coast, which requires the likelihood for the data, accounting for missing values. Define δ_{ij} to be an indicator function of whether or not data are observed at site j, in year i, $A_i = \{j : \delta_{ij} = 1, j = 1, \ldots, d\}$, d_i to be the number of sites with data in year i, i.e. $d_i = |A_i|$, and n to be the number of years. Then the overall likelihood, for parameter $\boldsymbol{\theta}$, is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} L_i(\boldsymbol{\theta}) \tag{8.4.2}$$

where

$$L_i(\boldsymbol{\theta}) = (2\pi)^{-d_i/2} |\Sigma_{(i)}|^{-1/2} \exp\{-\frac{1}{2} (\mathbf{z}_{(i)} - \boldsymbol{\mu}_{(i)})^T \Sigma^{-1}_{(i)} (\mathbf{z}_{(i)} - \boldsymbol{\mu}_{(i)})\},$$

i.e. L_i is the likelihood for an observation from a d_i-dimensional multivariate normal distribution. Here

$$\begin{aligned}\boldsymbol{\mu}_{(i)} &= \{(M_{i,k_1},\ldots,M_{i,d_i}) : k_j \in A_i, j = 1,\ldots,d_i\},\\ \mathbf{z}_{(i)} &= \{(z_{i,k_1},\ldots,z_{i,d_i}) : k_j \in A_i, j = 1,\ldots,d_i\}\end{aligned}$$

and

$$\Sigma_{(i)} = (\sigma_{jk})_{j,k\in A_i}.$$

Owing to the spatial coherence of the mean sea-level process, the correlation between sites i and j, $\rho_{ij} = \sigma_{ij}/(\sigma_{ii}\sigma_{jj})^{1/2}$, should vary as a smooth function of the inter-site distance, $|x_i - x_j|$. Figure 8.7 shows estimates of ρ_{ij}, obtained by maximising (8.4.2) with respect to $\boldsymbol{\theta}$ for $d = 2$, against the inter-site distance for pairs of sites. The form of the smooth local average curve on the figure motivates our selection of

$$\rho_{ij} = \rho_0 \exp\{-\rho_1|x_i - x_j|\}, \tag{8.4.3}$$

with $0 \leq \rho_0 \leq 1$ and $\rho_1 > 0$, i.e. an exponential correlation model with a nugget effect. With ρ_{ij} taken as (8.4.3) and all the other parameters unconstrained over sites, maximising likelihood (8.4.2) gives the estimates shown in Figure 8.8 when $M_{t,i}$ is taken to be linear. Here, $\hat{\rho}_0$ and $\hat{\rho}_1$ are 0.7(0.04) and $2.3 \times 10^{-4}(9.0 \times 10^{-5})$ (figures in brackets are standard errors), but despite this high correlation, the trend estimates are only slightly changed from those obtained by the univariate analyses, see Figure 8.9 for a comparison. In particular, the estimates for the longer series remain largely unchanged while the poor estimates from sites with short records become either more or less consistent with neighbouring estimates depending on how typical their observed data are with the fitted correlation model (8.4.3) and data from neighbouring sites, as is seen from (8.4.1).

It is clear from Figure 8.8 that there is still limited spatial smoothness in the trend estimates, and that the intercept, $\alpha(x)$, continues to be negatively correlated with $\beta(x)$. As $\alpha(x)$ corresponds to the mean sea-level at the base year, 1800, it should take a smooth form along the coast provided that data are to similar datum. Due to the complex form of spatial variation of $\alpha(x)$, parametric models are not suitable, so some form of non-parametric model could be used, in which case the regression model would be of a generalised additive form (Hastie and Tibshirani, 1990) and estimation could proceed using associated techniques. As we are not specifically interested in estimating $\alpha(x)$, instead we simply specify that it must vary smoothly along the coast by using a penalised likelihood

$$L_{\text{pen}}(\boldsymbol{\theta}) = L(\boldsymbol{\theta}) - \lambda \sum_{i=1}^{d} [\alpha'(x_i)]^2 \tag{8.4.4}$$

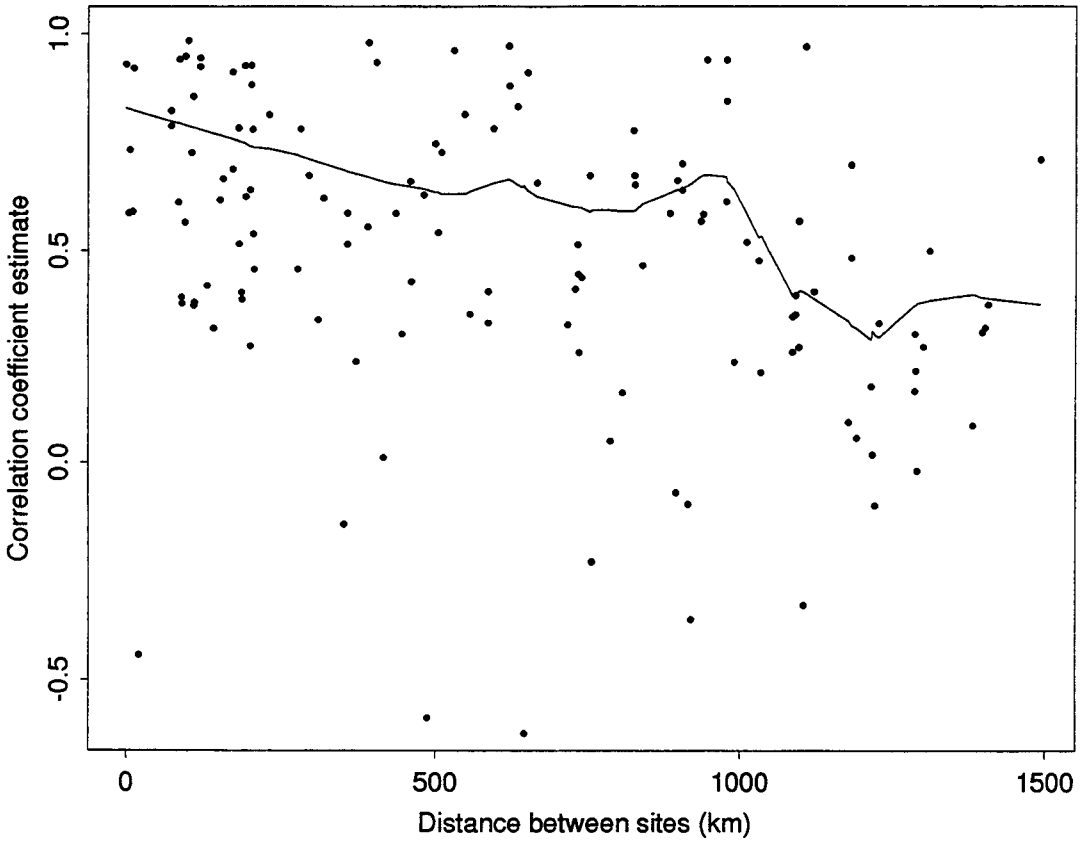

Figure 8.7: Inter-site correlation plotted against distance. The points correspond to the maximum likelihood estimates from fitting the bivariate normal model (8.4.2) to each pair of sites for which there are observations which overlap in time. The plotted curve is a simple smoother of these points which accounts for weighting given by the respective standard errors associated with each estimate. For clarity, estimates with large standard errors have been omitted from the plot, although they have been accounted for in the smoothing.

where $L(\boldsymbol{\theta})$ is given by (8.4.2), the second term penalises spatial roughness in the $\alpha(x)$ estimates, and λ is a tuning constant. Figure 8.9 shows estimates of $\beta(x)$ obtained from maximising the penalised likelihood (8.4.4), for correlation model (8.4.3), with $\lambda = 10^{-4}$. This leads to a significant smoothing of the trend estimates without imposing an explicit smoothness on them.

Now consider the estimation of the eustatic mean sea-level trend. Following Shennan and Woodworth (1992) we use (8.3.5) with $\beta_l(x)$ estimated by Shennan (1989). Then β_e is obtained by maximising the penalised likelihood (8.4.4), giving $\hat{\beta}_e = 1.17$mm/yr with standard error 0.07mm/yr. This estimate and the associated 95% confidence interval are shown on Figure 8.10, together with the penalised likelihood estimates, and confidence intervals, for $\beta(x)$ adjusted by estimates of $\beta_l(x)$.

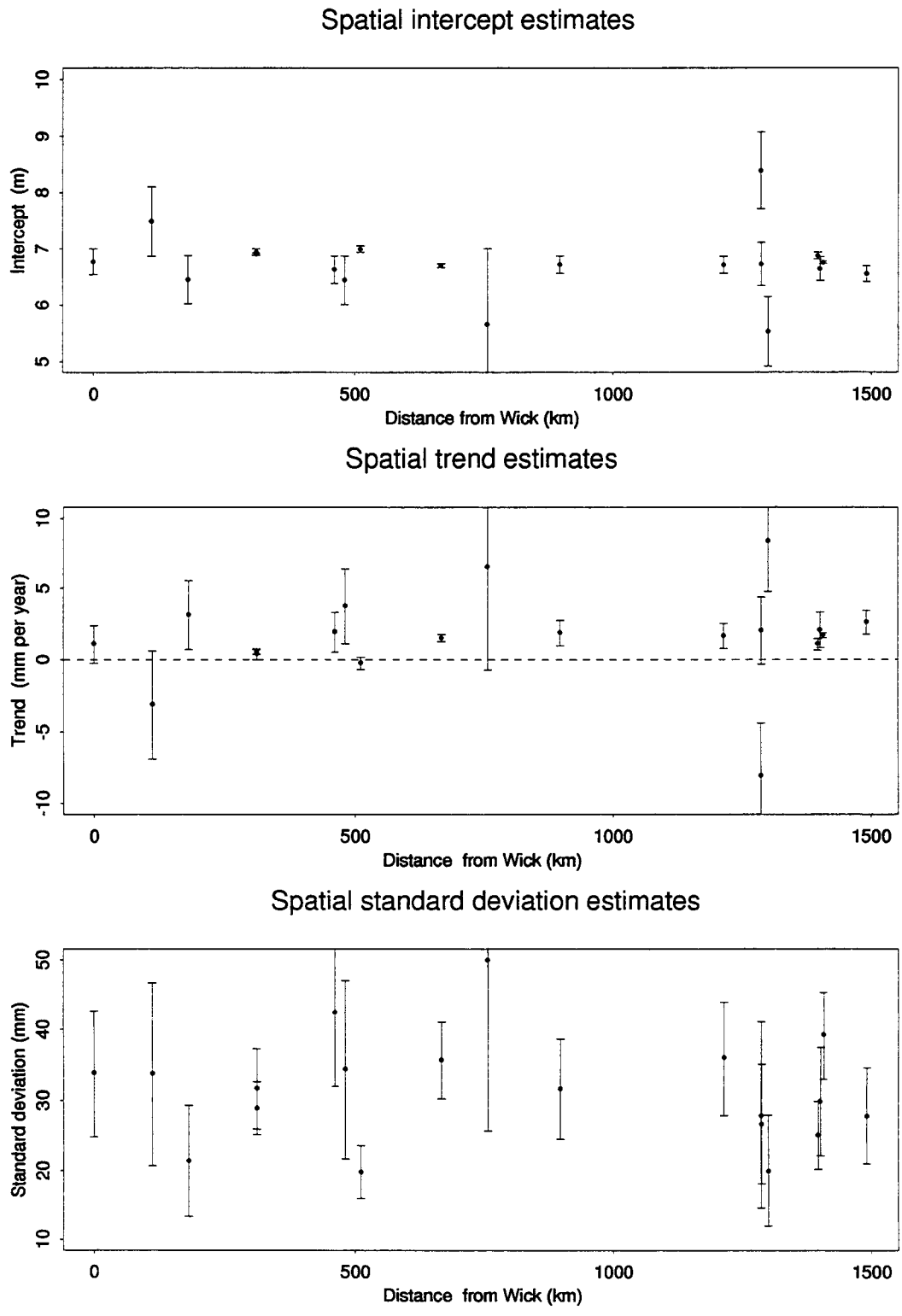

Figure 8.8: Spatial model based estimates, together with 95% confidence intervals, of $\alpha(x), \beta(x)$, and $\sigma(x)$ plotted against coastal distance from Wick, x. These estimates are maximum likelihood estimates of the parameters of the multivariate normal model spatial correlation model.

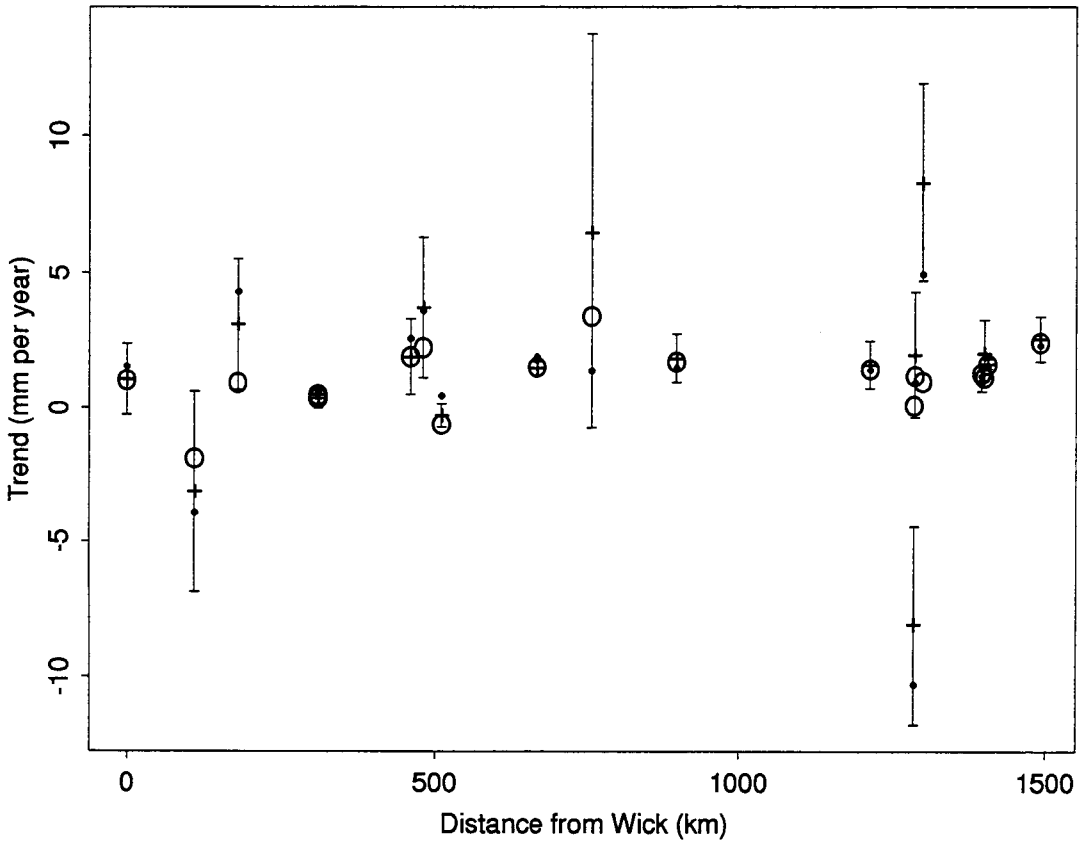

Figure 8.9: Marginal, spatial and penalised likelihood based estimates of the trend, $\beta(x)$, plotted against coastal distance from Wick, x. Spatial estimates are denoted by +, and are given with associated 95% confidence intervals. Marginal estimates are shown by •, and the penalised likelihood estimates by ○.

This figure suggests that the land level trends in the South East, i.e. large x, may be inaccurate, and consequently may have led to an under-estimation of the eustatic mean sea-level trend. Our estimate of β_e is consistent with previous findings (Barnett, 1984; Woodworth, 1987; Peltier and Tushingham, 1989) but is more precise as a result of our use of spatial analyses. Accelerations can be studied similarly: following (8.3.6), the quadratic trend parameter, γ, can be estimated using the penalised likelihood (8.4.4) giving $\hat{\gamma} = 9.9 \times 10^{-4}$mm/yr^2, with standard error 14.5×10^{-4}mm/yr^2. Thus even though the spatial properties of the data are exploited in this analysis, no significant evidence of accelerations is obtained. Assuming that this current estimated linear trend continues throughout the next century, the model here gives an estimate of the increase in eustatic mean sea-level by 2100 of 0.12m which is much smaller than the original IPCC (1990) estimate of 0.66m and the revised estimate by Wigley and Raper (1992) of 0.50m which were obtained under climate change scenarios. Based on Figure 8.10 our estimate appears

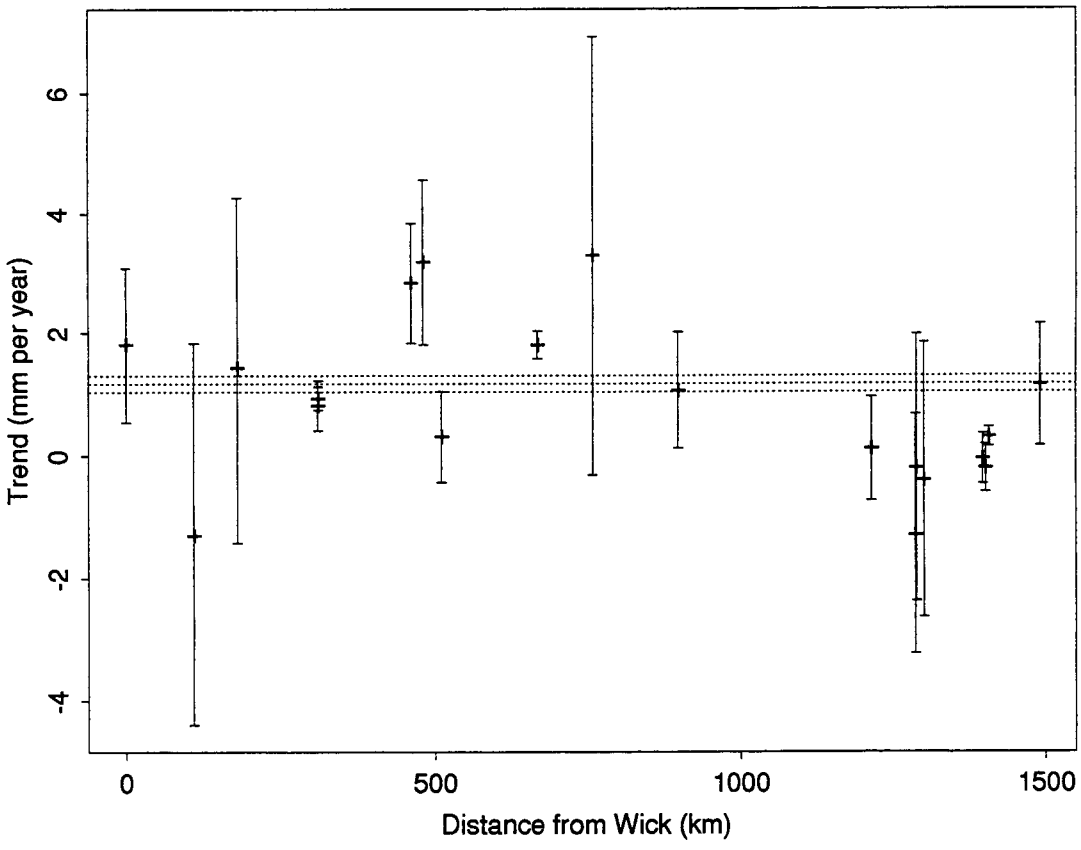

Figure 8.10: Estimated eustatic sea-level trend. The plot shows estimates, together with 95% confidence intervals, obtained using the penalised likelihood, where the estimates have been adjusted by the estimated land level trend (Shennan, 1989). Also shown on the plot, the dotted lines, are the estimate, and 95% confidence interval, of a common regional eustatic sea-level trend, obtained using the penalised likelihood.

to be a slight under-estimate, whereas estimates obtained using climate models appear high, but are much reduced from the original claims of $1 - 2$m.

Maximum likelihood in all these models was computationally intensive due to the high dimensionality of the parameter space and, as a consequence of missing data, the need to invert as many as n different variance-covariance matrices at each iteration in the maximisation routine. Evaluation of standard errors also proved troublesome numerically with problems involving rounding errors that caused the hessian matrix to be negative definite, i.e. not all the eigenvalues of the matrix were positive so negative standard errors resulted.

A principal weakness of the proposed methodology presented in this section is that we take a crude estimate of land level trend to be exact. These land trend estimates were obtained from geological data and are subject to large uncertainties. Current scientific work

in this area focusses on the use of global positioning satellites and absolute gravimetry to obtain accurate geodetic measurements of land level movements. We now describe how these measurements can be incorporated into the analysis to obtain improved estimates of $\beta_l(x)$. Since the measurements are subject to small random errors, observations are required over a number of years for each site. Letting $S_t(x)$ denote the measurement in year t at site x then

$$S_t(x) = s_0(x) + \beta_l(x)t + \epsilon_t(x)$$

where $s_0(x)$ is the base year level and $\epsilon_t(x)$ is the noise term. The likelihood for the measurements over sites and years, $L_{\text{sat}}(\beta(x_i); i = 1, \ldots, d)$, can be developed using arguments similar to those used earlier in this section. Then the joint likelihood of mean sea-level observations and satellite based geodetic measurements is

$$L_{\text{pen}}(\boldsymbol{\theta}) L_{\text{sat}}(\beta(x_i); i = 1, \ldots, d), \tag{8.4.5}$$

since the two measurement processes are independent. In (8.4.5) the trend parameter of $\boldsymbol{\theta}$ is $\beta_l(x_i) + \beta_e$ for $i = 1, \ldots, d$, so information about $\beta_l(x)$ is obtained from both components of the likelihood function, and this leads to improved estimates of both $\beta_l(x)$ and β_e by comparison to using the separate analysis of the data on mean sea-levels and land levels from satellites.

8.5 Errors in the datum

Throughout the analysis in Sections 8.3 and 8.4 the data for each site are assumed to be measured relative to a fixed datum level. However, datum errors, in the form of shifts, do occasionally occur during the upgrading of tidal gauges or through poor management of the gauge. If such shifts in level are not detected they can lead to bias in trend estimates. Figure 8.11 illustrates this for the mean sea-level data at Portpatrick, a site on the Scottish west coast, where a datum shift appears to occur in 1976. These data were analysed in Part II, and this potential datum error may explain the large trend estimate obtained there.

If undetected the trend would be approximately 5mm/yr, as opposed to $1 - 2$mm/yr given by the data before and after 1976. Similarly, if for Sheerness (see Figures 8.2-8.5) datum errors occurred in the long periods with no data then this would be highly influential. To avoid such errors, sites where there are worries about the datum control are often excluded from the analysis (Woodworth, 1987). The inefficiency of this approach motivates the development of models which are robust to datum changes, yet remain able to detect small trends. Ideas for such models are outlined in this section. In essence

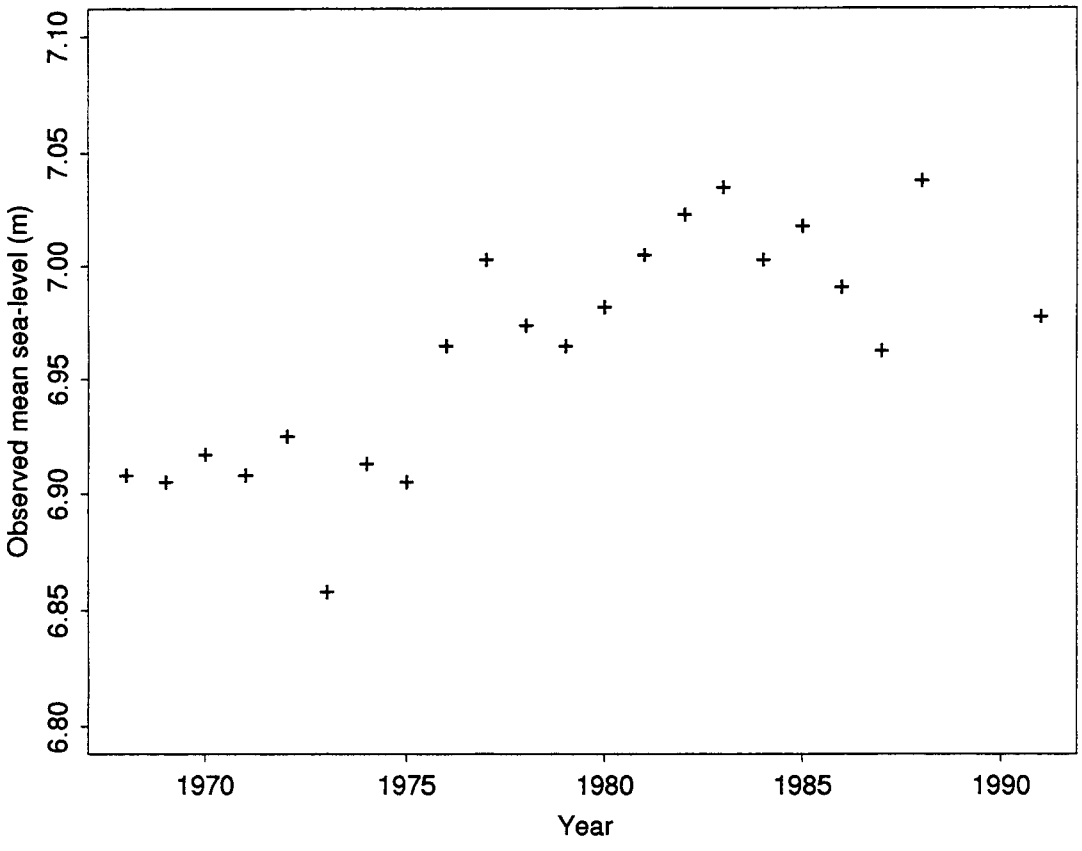

Figure 8.11: The Portpatrick mean sea-level series which exhibits a possible datum shift in 1976.

such methods must be change-point resistant as datum shifts lead to change-points in the incremental process $Z_{t+1}(x) - Z_t(x)$.

A simple approach based on analysing the univariate series $\{Z_{t+1}(x) - Z_t(x)\}$, separately for each site, will be unlikely to detect change points owing to the large inter-annual variability of mean sea-levels, i.e. the variance of $\epsilon_{t+1}(x) - \epsilon_t(x)$. Figure 8.11 illustrates why this is the case: the largest incremental rise occurs in 1988 and this does not appear to be a change-point. This shows how difficult it is to detect small datum shifts using only the data from the site. Again this suggests a spatial approach will be beneficial.

The buddy checking procedure of Woodworth (1987) could be used here since if there were a shift, c, in year t^* at site x, then (8.3.2) gives

$$Z_t(x) - Z_t(y) = \begin{cases} \alpha(x) - \alpha(y) + [\beta_l(x) - \beta_l(y)]t + \epsilon_t(x) - \epsilon_t(y) & \text{if } t < t^* \\ c + \alpha(x) - \alpha(y) + [\beta_l(x) - \beta_l(y)]t + \epsilon_t(x) - \epsilon_t(y) & \text{if } t \geq t^* \end{cases}$$

which amounts to analysing spatial increments, between pairs of sites, for change-points. Buddy checking is likely to be an improvement over analysing temporal increments as

$\epsilon_t(x) - \epsilon_t(y)$ will have a smaller variance than $\epsilon_{t+1}(x) - \epsilon_t(x)$ owing to spatial dependence for x and y sufficiently close together.

These approaches to detect datum shifts have two drawbacks; firstly they cannot exploit the complete spatial structure of the problem, since only multiple pairwise comparisons can be assessed via buddy checking, and secondly knowledge of the maintenance history of the gauges cannot be incorporated. The first deficiency can be overcome using the following model: assume there exists a single change-point of size c^* at time t^* and site x^*, and referring to (8.3.2), define

$$Z_t(x^*) = \begin{cases} M_t(x^*) + \epsilon_t(x^*) & \text{if } t < t^* \\ c^* + M_t(x^*) + \epsilon_t(x^*) & \text{if } t \geq t^* \end{cases} \tag{8.5.1}$$

and $Z_t(x) = M_t(x) + \epsilon_t(x)$ for all other x. Here c^*, t^* and x^* are parameters of the model, so for example the model can be extended to have n_c change-points at the expense of additional parameters $\{(c_i^*, t_i^*, x_i^*); i = 1, \ldots, n_c\}$. Estimation of the model would be similar to before but with the change-point parameters also needing to be estimated. In essence this extends the buddy checking procedure to all the sites and incorporates a model for the shifts at the detected sites and times.

The issue of exploiting the additional knowledge of the times of potential datum errors is best handled through a Bayesian prior. Bayesian change-point problems have been studied previously (Smith, 1975), however here the prior for the time of the change-point is informative as it relates to the history of the gauge maintenance. The novel aspect in this case arises from the feature that when gauges are replaced, or have a fault, there will be intervals without a gauge, and hence no data. Therefore if a site has a period of missing data, we assign it a higher prior probability for a change point than during a period of continuous data. At first sight this may appear contrary to the Bayesian paradigm, but of course it is not since missing data are non-informative in the likelihood formulation.

Suitable priors for change-point times are yet to be formulated but models should capture the above feature, together with independence over sites, and incorporate the feature that change points are thought to be more likely further in the past. With the flexibility of Bayesian modelling using Gibbs sampling (Gelfand and Smith, 1990) many possible forms for the prior can be examined, and thus the positions of possible change-points identified and fitted simultaneously within the Bayesian analysis, through model (8.5.1). These trend estimates will be more robust than existing estimates and no efficiency should be lost. Application at a computational level is still required.

8.6 Conclusions

In this case study a procedure is proposed for obtaining improved mean sea-level trend estimates by using spatial statistical models. A precise estimate $(1.0, 1.3)$mm/yr for the eustatic mean sea-level trend was obtained, and this is of interest in its own right, particularly to oceanographers studying global sea-level trends.

Based on the data used here, no significant evidence is found for an acceleration in the mean sea-level process; however it is clear that acceleration effects, such as those predicted by climatological models, may not be detected using the mean sea-level data available at the present time as the acceleration is still predicted to be very small.

Since trends in extreme sea-levels and mean sea-levels were found to to very similar in Chapter 7, the findings of this case study are of use in studying extreme sea-levels, as the estimates trends are more accurate than extreme sea-level trend estimates at sites. This trend estimation significantly improves the estimation of future return levels for sites along the UK east coast.

Chapter 9

Case study III: spatial modelling of r-largest data

9.1 Introduction

The purpose of this case study is to investigate the properties and feasibility of a spatial version of the r-largest method. The method is applied to a set of data from the Humber estuary.

The model is simultaneously fitted to the data from all of the sites along the estuary, using a joint likelihood. As the data from each site are themselves extreme values the natural model for these data is a multivariate extreme values model (Coles and Tawn, 1991). However, application of multivariate extreme value models is not straightforward, particularly for r-largest data from each site. One particular approach, which has been successfully used for other applications, is to fit a joint model over sites assuming

- that the extremal data are independent over sites
- that the parameters of the distribution for each site vary smoothly from site-to-site, and can be described by a spatial parametric form.

Examination of the data clearly shows that this assumption of independent spatial data is incorrect for extreme still water levels in the Humber, a feature typical in other applications of this approach. Parameter estimates are unbiased under this incorrect dependence model, but standard errors of the parameter estimates will be under-estimated by the standard method of calculation. This is corrected by using a method proposed by Smith (1994), whereby the spatial independence likelihood model is used for estimation but a complex procedure is then used to adjust the standard errors for spatial dependence. This procedure is widely used in other statistical areas, and has been successfully applied in

the extreme value context by Coles and Walshaw (1994) and Coles and Tawn (1995). Therefore it is adopted here as it has the benefit of avoiding the need to explicitly model spatial dependence in the likelihood. Details of the adjustment procedure are not given here.

The breakdown of the work in this case study is as follows:
The Humber data and covariate information are described in Section 9.2. As we are interested in providing estimates of return levels at any location along the Humber we need the tidal series at each site. The methods used to obtain this are described in Section 9.3. In Section 9.4 a standard univariate r-largest method is applied to all sites. This motivates the choices of form of regression models and covariates in the spatial application in Section 9.5. Comparisons of the results with other methods of analysis applied to the Humber estuary are given in Section 9.6.

Throughout this case study, we take Z_i to be the annual maximum in year i. From the standard r-largest method (see Dixon and Tawn, 1994, Section 4.2) we have that

$$\begin{aligned} \Pr\{Z_i \leq z\} &= \exp\{-[1-k(z-\mu_i)/\sigma]_+^{1/k}\} \\ &= G_i(z) \end{aligned} \tag{9.1.1}$$

where $\mu_i, \sigma(\sigma > 0)$ and k are the location, scale and shape parameters respectively. A linear trend is used here, so

$$\mu_i = \alpha + \beta(i - i_0), \tag{9.1.2}$$

with the base year taken as $i_0 = 1919$ throughout the chapter. Let

$$G_i'(z) = \frac{d}{dz} G_i(z),$$

and $Z_i^{(j)}$ denote the jth largest still water level in year i. Finally note the return level, $z_p(t)$, is defined by

$$\Pr\{Z_i \leq z_p(i)\} = G_i(z_p(i)) = 1 - p$$

for given p, and then

$$z_p(i) = \alpha + \beta(i - i_0) + \sigma\{1 - [-\log(1-p)]^k\}/k. \tag{9.1.3}$$

Throughout, all return levels will be given for the year 1990, i.e. $i = 1990 - 1919 = 71$, so the return level for any other year, say year t, is

$$z_p(t) = z_p(1990) + \beta(t - 1990).$$

9.2 Humber estuary: available data

The data used in this study are still water levels for 14 sites along the Humber estuary. There are two sources for these data:

- the NRA data–base: this consists of series of high water levels.
- the historical annual maximum still water level series: these data are held by the Proudman Oceanographic Laboratory but were originally collected and analysed by Graff (1981).

For the purposes of extreme level analysis, using the joint distribution of the r – largest order statistics method (Tawn, 1988), the data from the NRA data–base are pre–processed before use. The pre–processing involves extracting from each site, and each year, the top r (here $r = 7$) extreme independent levels. The method of extracting the top independent levels is similar to that used in the application of the r-largest method in Dixon and Tawn (1994, Section 4.2). The slight difference is that here the data are not hourly sea-levels, but high water levels instead, so are approximately every 12 hours. In line with the declustering storm length choice used in Section 9.7, two high water levels are defined to be independent only if they are not consecutive.

Some years where high water data are available are excluded owing to missing data. Our criteria for omission was that if less than 75% of high water data were recorded in the winter periods of January – March combined with October – December inclusive, then r-largest values are not extracted from the NRA data–base; again this is in line with application of the r-largest method in Dixon and Tawn (1994).

Table 9.1 shows the sites with data which are analysed in this study, together with the associated data types and years over which these data are available. Although the largest annual order statistic and the annual maximum in years where both were available should be identical there are many cases where they differ. As no systematic reason could be determined for them differing we retain the r–largest annual event data in years for which both exist to exploit as much data as possible. To summarise, for each site, the extremal data set consists of years containing one of the following:

- the 7–largest independent annual events,
- the annual maximum,
- no data.

The data sites for the case study (omitting some of the docks at Hull) are shown on a map of the Humber estuary in Figure 9.1.

Site	Distance	7-largest	Annual Max.
Spurn Point	0.0	1973–89 16 years	no data
Grimsby	11.8	no data	1920-73 54 years
Immingham	20.8	1956–88 28 years	1920–88 69 years
Saltend Jetty	31.7	no data	1965–77 13 years
King Georges Dock	33.7	no data	1922–73 36 years
Victoria Dock	36.5	no data	1920–69 24 years
Humber Dock/Hull	37.9	1977–89 11 years	1920–68 42 years
St. Andrews Dock	39.2	no data	1920–73 49 years
South Ferriby	51.5	1968–79 12 years	no data
Brough	57.4	1951–89 35 years	1922–77 56 years
Blackfoft	67.5	1923–89 66 years	1921–77 56 years
Goole	78.9	1920–89 78 years	1920–78 59 years
Barmby	92.6	1972–87 8 years	no data
Selby	96.2	1972–86 11 years	no data

Table 9.1: Data sites, distance in kilometres along the estuary reference axis and exploited data.

9.3 Humber estuary: tides

The methods in both Chapter 9 and 12 require information about the tidal series at each data site. So, for interpolation of return level estimates to new sites, for example on a regular grid along the estuary, information about the tidal series at every site of interest is required. This section gives a summary of the methods developed by the Proudman Oceanographic Laboratory to give the required tidal series estimates in the case of the Humber data.

Due to the introduction of complexities in the tide, arising from the shallow water effects in the estuary beyond Goole, the tidal analysis was restricted to the Spurn Point to Goole stretch of the estuary. To interpolate the tides along the Humber estuary from Spurn Point to Goole the following procedure is adopted:

1. Reference Ports

Harmonic constituents were derived from observed data for 6 reference ports:

Spurn Point
Immingham
Hull
Humber Bridge
Blacktoft
Goole

2. Reference axis for the Humber

From these 6 reference stations it was required to produce the harmonic constituents at regular spatial intervals along the estuary which we chose to be at 0.8 kilometre increments. This incremental distance can be changed but it was felt unnecessary as the tidal amplitude only changes by about 5 cm over this incremental distance. It was essential to define a reference axis for the estuary from a small scale map. This axis was used to determine the position of the reference stations, together with the other data sites, and any intermediate interpolated positions. No attempt has been made to consider the estuary as two dimensional - the position of a station on either bank is simply found by using the normal to the reference axis. As far as possible, the axis was chosen to follow the deep water channel along the Humber. The positions of the data sites are given in Table 9.1, with the additional reference site Humber Bridge having a distance of 40.0 kilometres.

3. Interpolation of harmonic constituents

Two approaches for the required interpolation were considered:

(a) Interpolation of the tidal constituents directly

Using the 6 base reference ports distributed along the estuary, a low order Chebychev

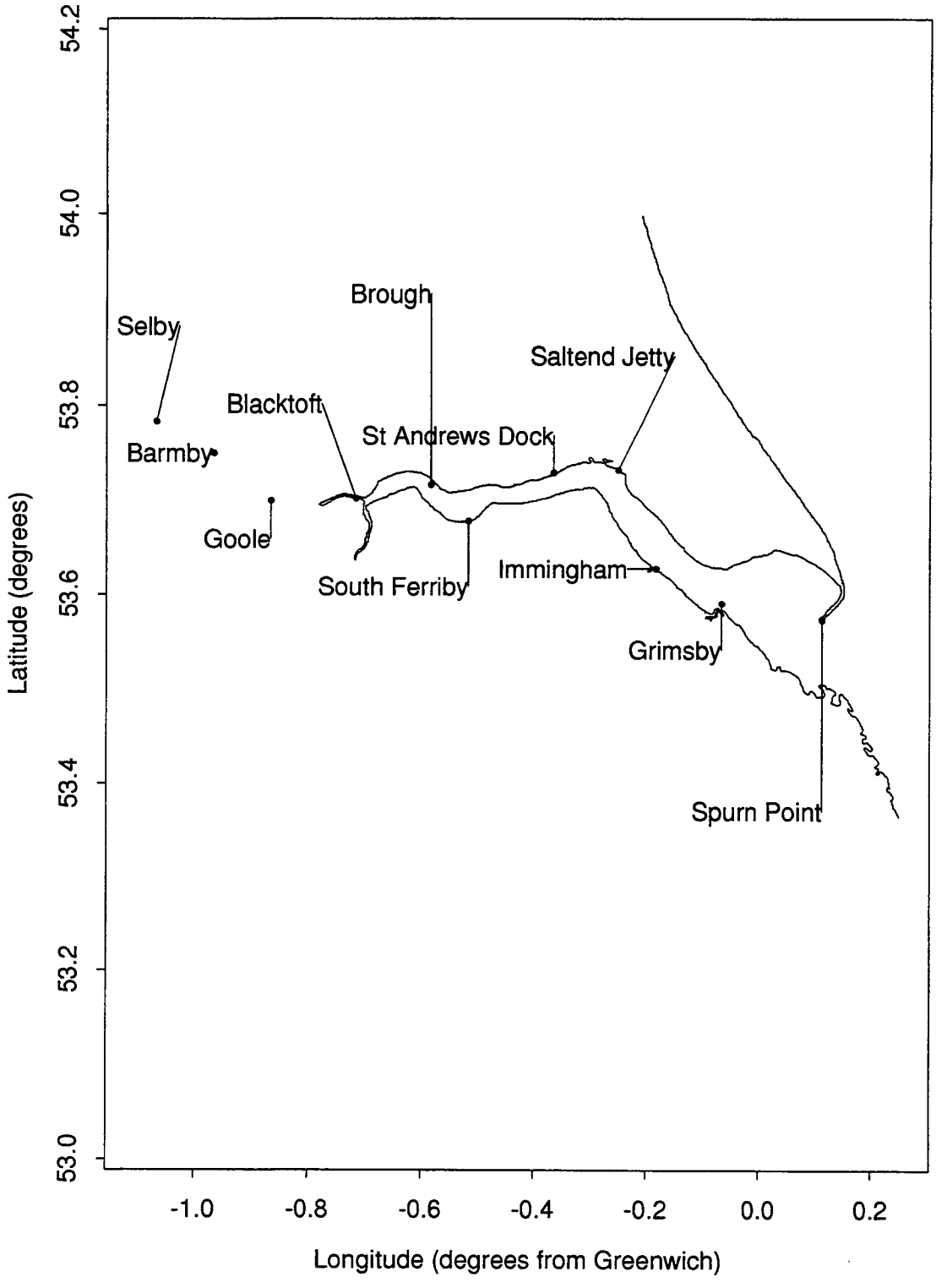

Figure 9.1: Map of the Humber estuary data sites.

polynomial was fitted separately to each of 60 tidal constituent. These curves were then used to interpolate to points 0.8 kilometres apart. In this manner a set of harmonic constituents were determined for the whole estuary. This technique was found to perform better for the major constituents than for the minor terms as one might expect. However even for the major constituents this was found to be less accurate than required particularly in the upper reaches of the estuary. The reason is that the tidal regime becomes very complex upstream and it would require many more reference points to represent the changes in linear and shallow water terms correctly.

(b) Interpolation of the tidal time series

A better approach, and the one that was adopted finally, was to interpolate the tidal time series rather than the harmonic constituents. This is rather a 'hammer and nut' approach but gave better results. A tidal prediction covering 1 year was prepared for the 6 reference ports using the appropriate harmonic constituents (60). A spatial interpolation across the 6 ports was then performed for the first hourly sea level height resulting in a set of 85 interpolated values up the estuary. This was repeated successively through the 8760 hourly heights creating a data set of 1 year of predictions for each of the interpolated positions. A low order (4) Chebychev polynomial and a cubic spline technique were both tried as interpolation schemes and both proved to be satisfactory but the cubic spline gave slightly better results. A tidal analysis was then performed on each of these positions using the 1 year of interpolated tidal data to produce 60 harmonic constituents for each position.

Whichever method of tidal interpolation is used it is impossible to determine the accuracy of the resulting interpolated harmonic constituents. However some idea can be gained from a comparison of the interpolated constituents at the reference points with the actual values, as the interpolation curves were not constrained to pass through the reference points but only to fit them in a least squares sense. Values for M_2 are shown in Table 9.2, for method (b), with amplitudes in metres and phases in degrees. These

Site	Actual		Interpolated	
	Ampl	Phase	Ampl	Phase
Spurn Point	2.134	151.1	2.141	151.1
Immingham	2.272	161.7	2.277	161.3
Hull	2.412	173.5	2.415	173.1
Humber Bridge	2.475	181.5	2.483	181.4
Blacktoft	2.012	202.4	2.023	202.3
Goole	1.923	187.4	1.929	187.3

Table 9.2: Comparison of observed and interpolated features of the M_2 tide.

errors appear relatively small, but could of course be larger between the reference points. However from an examination of the curve fitting visually the maximum error in the range of the tide at the interpolated positions appears to be of the order 10 cm in the tidal range.

9.4 Tidal covariate information

Based on a theoretical model and findings with data, Coles and Tawn (1990) showed generally that the tide only affects the location parameter of the GEV, though the intercept α, see equation (9.1.3), whereas the other parameters changed smoothly with coastal distance. Here we examine which characteristic of the tide is a suitable covariate for estuary sites.

Consultation with Proudman Oceanographic Laboratory revealed that the most likely dependence on tidal information is through one of three functions of the harmonic constituents of the tide at the site. Figure 9.2 shows the three possible tidal functions:

- (a) the sum of the amplitudes of M_2, S_2, M_4, M_6 and MS_f.
- (b) the sum of the amplitudes of M_2, S_2.
- (c) the sum of the amplitudes of M_2, S_2 added to the product of the amplitude of M_4 and the cosine of the phase difference between M_2 and M_4.

Coles and Tawn (1990) effectively used form (b) but their analysis was primarily restricted to the open coast. Based on knowledge that as tides progress up an estuary the energy in M_2 is converted to energy in M_4, M_6 and MS_f it was considered worth considering form (a). Form (c) is also given, as M_2 and M_4 are phase locked.

9.5 Separate–site extreme value application

In this section the standard r-largest method in the case of linear trends (see Dixon and Tawn, 1994, Section 4.2) is applied to these Humber data. Combining r-largest and annual maximum still water level data the likelihood for a site is

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{n} \left\{ \left[\prod_{i=1}^{r_{i_j}} \frac{G'_{i_j}(z_{i_j}^{(i)})}{G_{i_j}(z_{i_j}^{(i)})} \right] G_{i_j}(z_{i_j}^{(r_{i_j})}) \right\}, \tag{9.5.1}$$

where $\boldsymbol{\theta} = (\alpha, \beta, \sigma, k)$, $i_1, \ldots, i_n$ are the n years in which data are observed, and

$$r_t = \begin{cases} 1 & \text{if only annual maximum available in year t} \\ 7 & \text{if r–largest events are available in year t.} \end{cases}$$

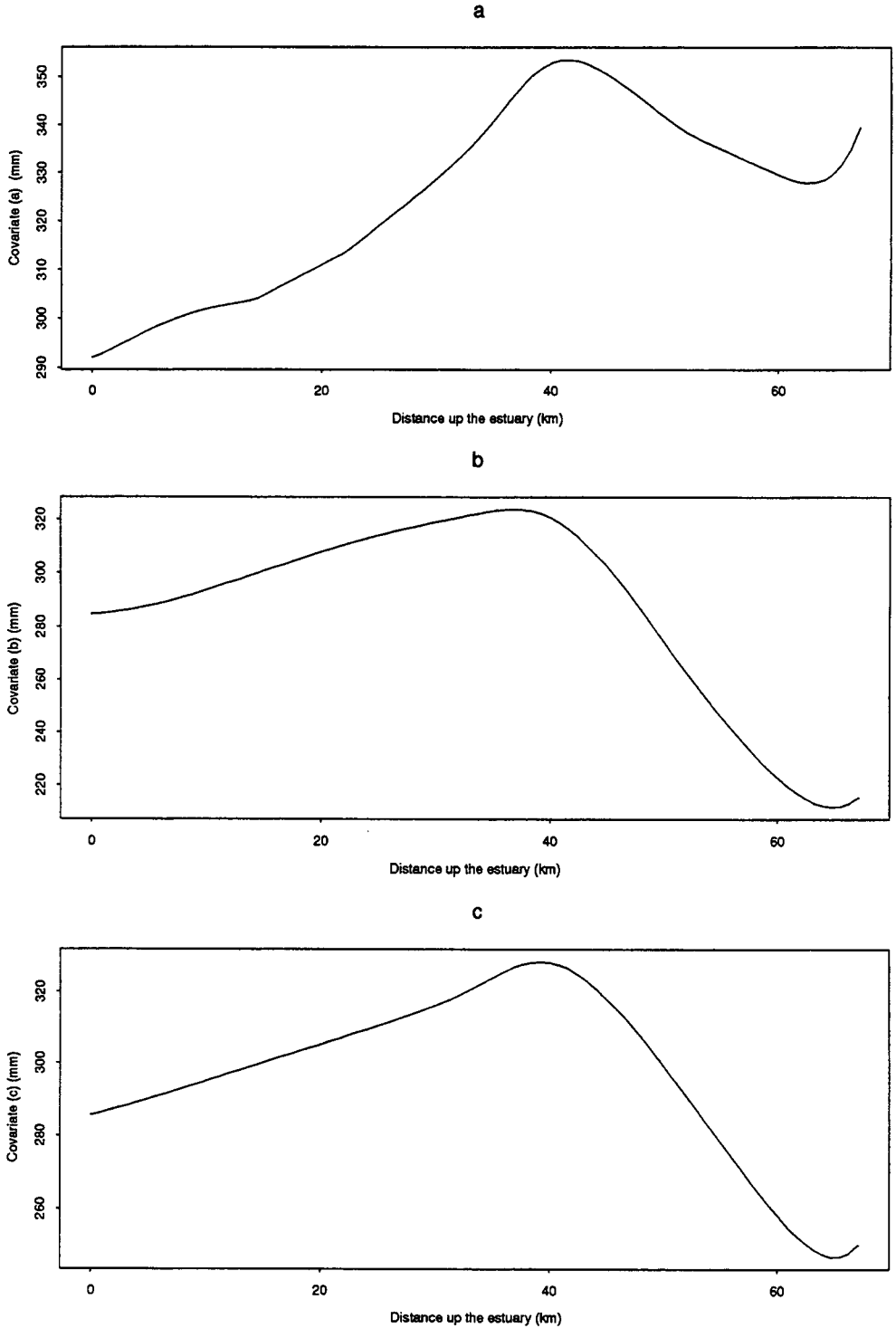

Figure 9.2: Tidal covariates plotted against distance up the estuary.

Figure 9.3 shows the maximum likelihood estimates together with associated 95% confidence intervals for the individual GEV parameters α, β, σ, and k, the intercept, trend, scale and shape parameters respectively. In this figure each parameter is plotted against the reference axis – estuarine distance. Similarly Figure 9.4 shows return levels $p = 0.1, 0.01, 0.001$ and 0.0001 plotted against the reference distance.

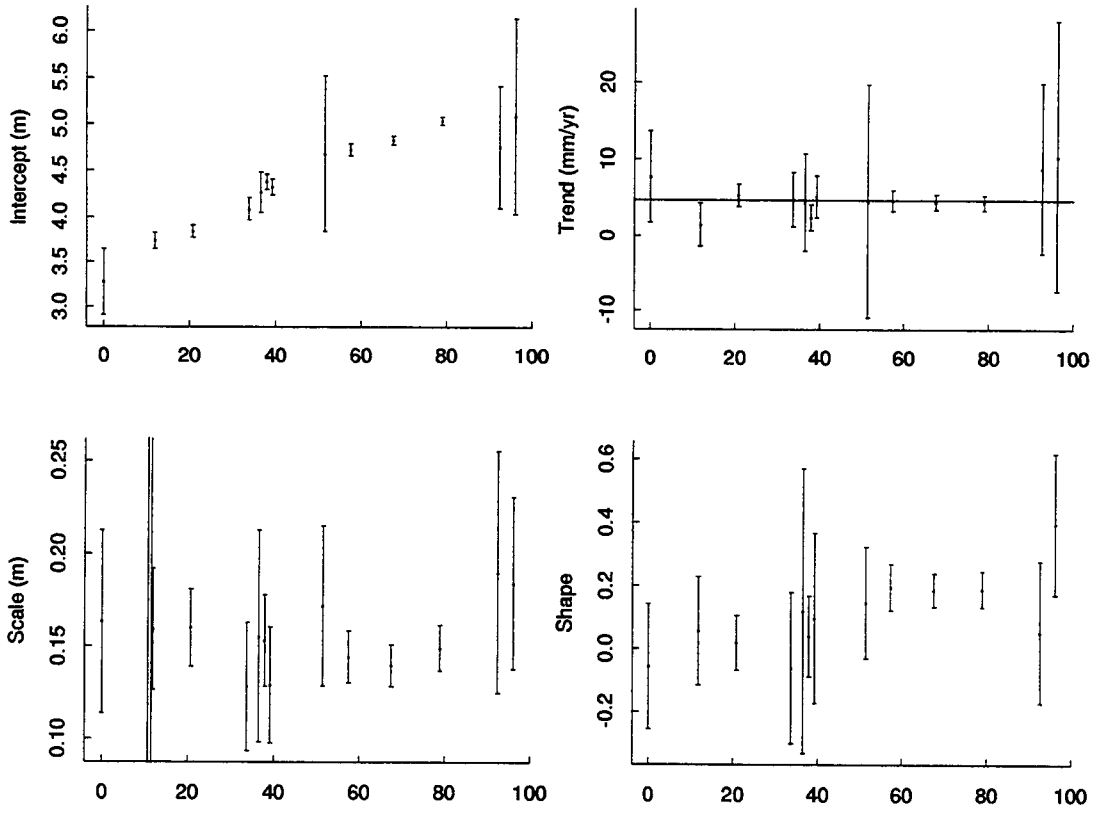

Figure 9.3: Maximum likelihood estimates, with 95% confidence intervals, for the GEV parameters obtained from univariate analysis at each site. The abscissa on the plots denotes estuarine distance in km.

The estimates for Saltend Jetty are excluded from the figures due to the large uncertainties associated with the estimates – $z_p(1990) = 5.06$ (with standard error 0.25), 6.42 (2.45) and 10.84 (14.97) for $p = 0.1, 0.01$ and 0.001 respectively. Despite this feature this site is retained in the data–base as it is considered to be consistent with the other data, with the large uncertainty arising due to the limited duration of the series.

In Figure 9.3, poorly estimated values of the linear trend parameters (α, β) occur in pairs at a site due to the correlation between estimates of α and β. Similarly problems occur with σ and k. Despite these complications with individual parameter interpretation, from Figure 9.3:

- the trend parameter, β, shows no significant variation along the estuary,
- the parameter α, increases up the estuary,
- the scale parameter, σ, appears relatively stable over the estuary, with some evidence of a quadratic relationship with estuarine distance,
- the shape parameter, k, increases, becoming significantly positive, as estuarine distance is increased.
- Comparing the tidal forms of Figure 9.2 with the α plot in Figure 9.3 only form (a) captures the variability of α, and so this is the only tidal covariate considered later.

These empirical findings motivate the development of the spatial model in the next section in which distance is clearly an important covariate but tidal function plays less of a role.

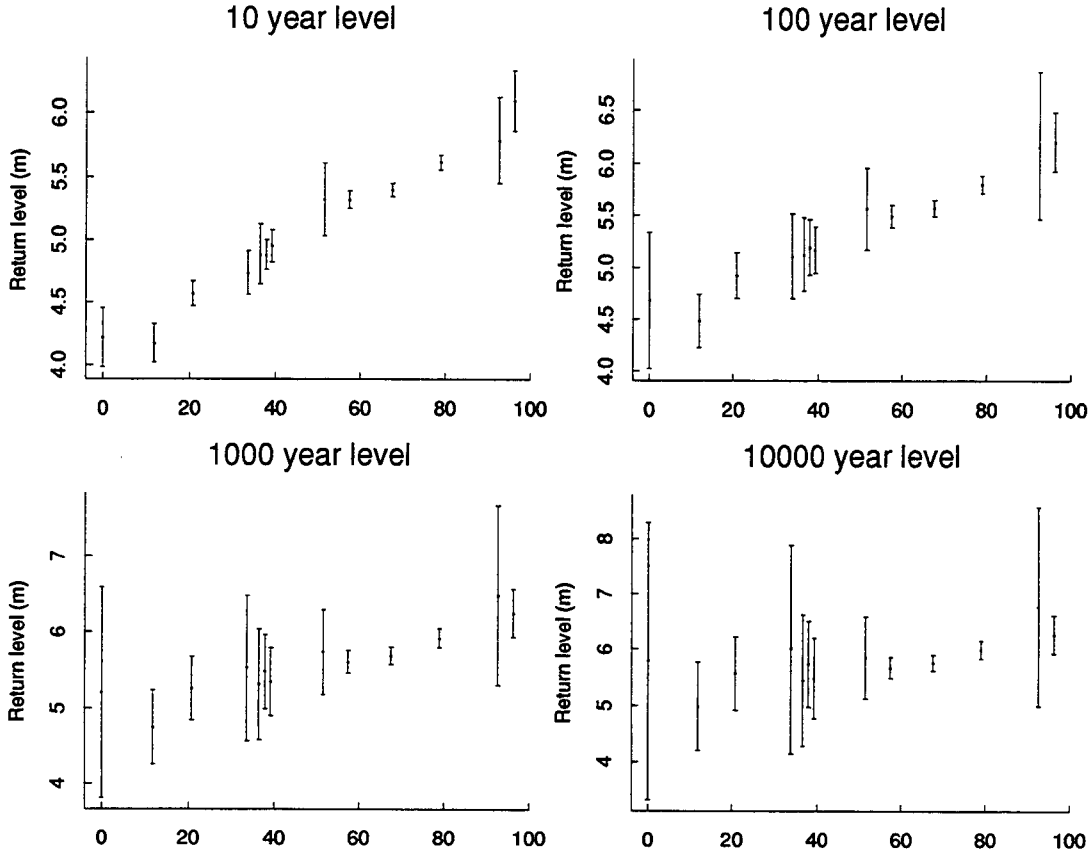

Figure 9.4: Maximum likelihood estimates, with 95% confidence intervals, for return levels obtained from univariate analysis at each site. The abscissa on the plots denotes estuarine distance in km.

9.6 Spatial extreme value methods

The extreme value data for each site are assumed to be independent and a multivariate model for all sites is fitted using regression covariate techniques to link marginal parameters. Thus this spatial model is not fully multivariate as spatial independence is assumed even though dependence is evident in the data. As discussed in the introduction of this case study, this method works well and has been applied in other areas. Let $L_{s_m}(\theta_{s_m})$ denote the likelihood for site $s_m (m = 1, \ldots, 13)$, with parameters $\theta_{s_m} = (\alpha_{s_m}, \beta_{s_m}, \sigma_{s_m}, k_{s_m})$. Then under the assumption of independence over sites, the joint likelihood for all sites is

$$\prod_{m=1}^{13} L_{s_m}(\theta_{s_m}). \tag{9.6.1}$$

The approach is to apply regression techniques using this joint likelihood with covariates of estuarine distance and tidal information described in Section 9.3. Generally, with regression parameter vector ϕ, and covariates d_{s_m} and T_{s_m} (corresponding to estuarine distance and tidal function at site s_m respectively), we have

$$\theta_{s_m} = h(d_{s_m}, T_{s_m}; \phi) \tag{9.6.2}$$

i.e.

$$(\alpha_{s_m}, \beta_{s_m}, \sigma_{s_m}, k_{s_m}) = (h_1(d_{s_m}, T_{s_m}; \phi_1), h_2(d_{s_m}, T_{s_m}; \phi_2), h_3(d_{s_m}, T_{s_m}; \phi_3), h_4(d_{s_m}, T_{s_m}; \phi_4)),$$

where $h_i, i = 1, \ldots, 4$ are some functions of d_{s_m} and T_{s_m}. This model is a good representation of the process if the process is spatially coherent such that the resulting parameters of the GEV change systematically along the estuary as a function of estuarine distance and tidal function at each site. The approach is to search for some functions of the covariates, using the univariate plots against distance as a basis, so that a good model fit is obtained. Once they have been chosen, the parameters ϕ are estimated by maximising the likelihood

$$L^*(\phi) = \prod_{m=1}^{13} L_{s_m}\left(h(d_{s_m}, T_{s_m}; \phi)\right), \tag{9.6.3}$$

with L_{s_m} and h defined as in equations (9.6.1) and (9.6.2).

First covariate models are fitted to β, σ and k, leaving α unconstrained from site to site. Our reason for this is that α is highly influenced by β, yet to fit β requires the other parameters. Details of all the intermediate regression forms fitted and tested are not given here; the models used for β, σ and k are decided as:

- $\beta_{s_m} = h_2(d_{s_m}; \phi_2) = \phi_2$, constant with respect to d_{s_m},
- $\log \sigma_{s_m} = h_3(d_{s_m}; \phi_3)$, quadratic in d_{s_m},

- $k_{s_m} = h_4(d_{s_m} : \phi_4)$, linear in d_{s_m}.

The points on Figure 9.5 show the corresponding fourteen α_{s_m} parameter estimates under this model.

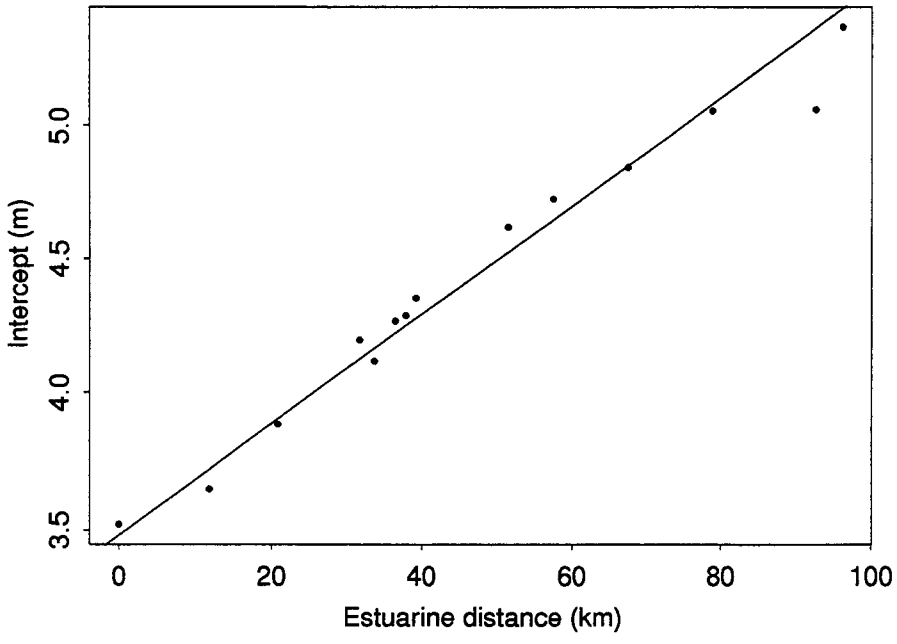

Figure 9.5: The intercept parameter plotted against distance for the spatial fit (solid line), and the unconstrained spatial fit (points).

The linearity of the plot suggests that once estuarine distance is accounted for, the tidal characteristic is not informative. This is confirmed by the use of likelihood ratio tests in nested models of regression parameters for tidal information and so the model for the intercept is

- $\alpha_{s_m} = h_1(d_{s_m}; \phi_1)$, linear in d_{s_m}.

The fitted form for h_1 is shown on Figure 9.5 whilst the fitted forms for h_2, h_3 and h_4 are shown on Figure 9.6 (which also shows the separate site estimates obtained in Section 9.5). The fitted model, for site s, is

$$\begin{aligned} \hat{\alpha}_s &= 3.482 + 0.0203 d_s \\ \hat{\beta}_s &= 0.00465 \end{aligned}$$

$$\hat{\sigma}_s = \exp\{-1.592 - 0.01518 d_s + 0.00016 d_s^2\}$$
$$\hat{k}_s = -0.0648 + 0.0035 d_s$$

with respective standard errors $(3.5 \times 10^{-2}, 2.0 \times 10^{-4}, 6.44 \times 10^{-4}, 4.9 \times 10^{-2}, 1.55 \times 10^{-3}, 1.7 \times 10^{-5}, 4.1 \times 10^{-2}, 6.2 \times 10^{-4})$.

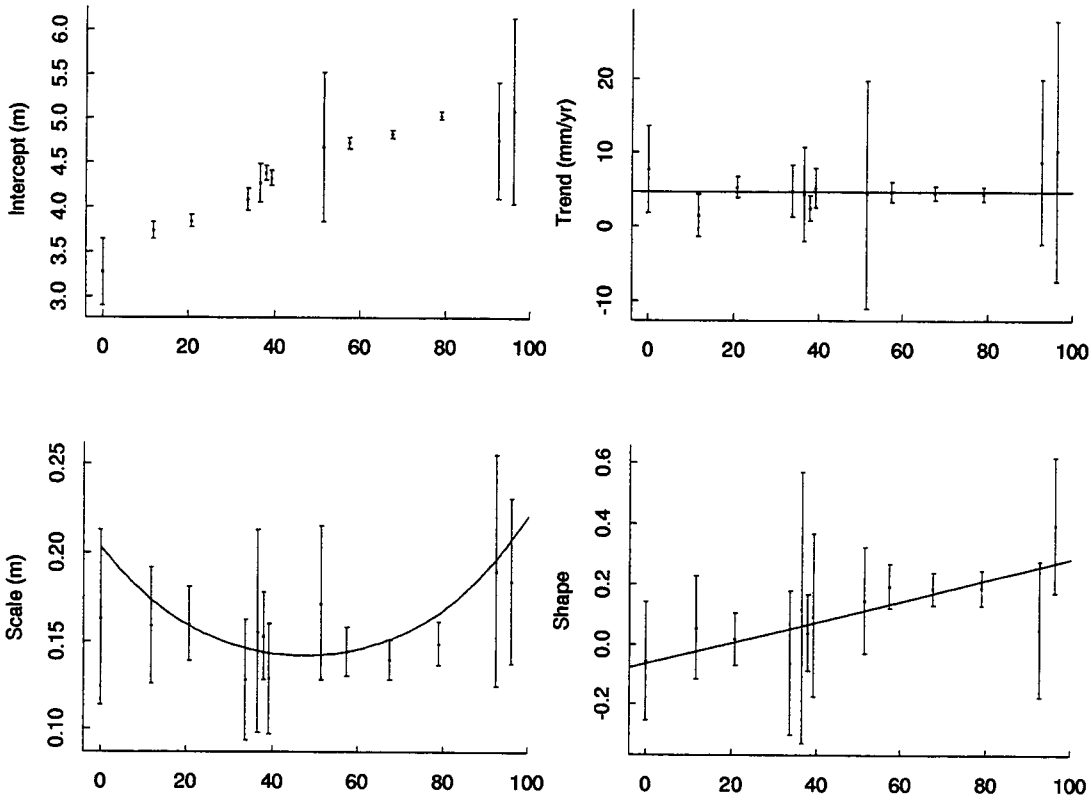

Figure 9.6: Maximum likelihood estimates for trend, scale and shape parameters, obtained from the spatial analysis. The univariate estimates are also shown on the plots. The abscissa on the plots denotes estuarine distance in km.

Both Figures 9.5 and 9.6 show that the spatial model fits the information from the separate sites very well. Figure 9.7 shows the corresponding spatially estimated return levels, with associated 95% confidence intervals. Clearly these fit the return level estimates from the separate site analyses well but have the benefit of being much more precise and exhibiting smooth spatial variability.

A key advantage of using likelihood (9.6.2) for estimation is that data from all sites are used in the estimation leading to improved parameter precision and the incorporation of physical knowledge about the spatial coherence of the process. As a result of the spatial estimation we now have a model for the GEV parameters and hence return levels at any

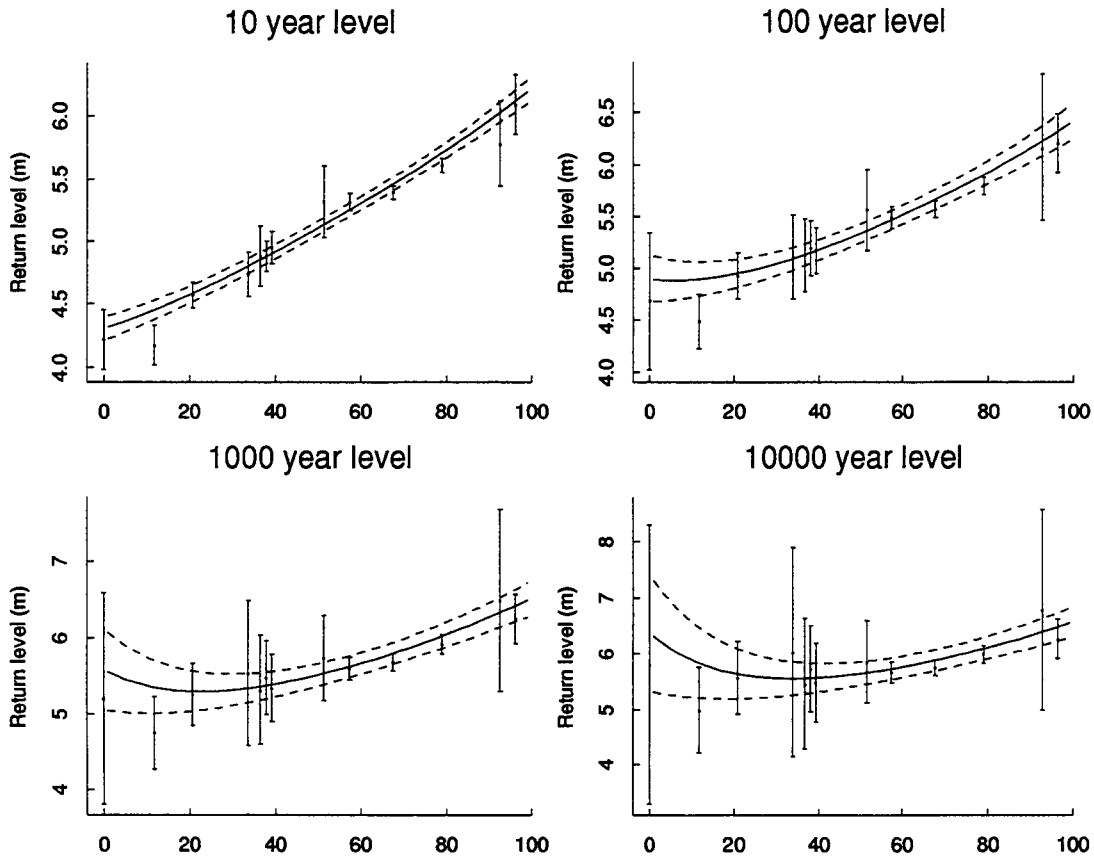

Figure 9.7: Maximum likelihood estimates, with 95% confidence intervals, for return levels obtained from the spatial analysis. The univariate estimates are also shown on the plots. The abscissa on the plots denotes estuarine distance in km.

estuarine position, thus the return level at a new position s, for year t, is given by

$$h_1(d_s,T_s;\hat{\phi}_1)+h_2(d_s,T_s;\hat{\phi}_2)(t-i_0)+h_3(d_s,T_s;\hat{\phi}_3)\{1-[-\log(1-p)]^{h_4(d_s,T_s;\hat{\phi}_4)}\}/h_4(d_s,T_s;\hat{\phi}_4),$$

where $\hat{\phi}_i$ is the estimate of ϕ_i.

In Table 9.3 the values of the spatially estimated return levels, as shown on Figure 9.7, are given at 5 km intervals along the Humber estuary reference axis.

9.7 Comparison with estimates from other analyses

It is of interest to compare both the trend and return level estimates obtained here with other estimates for the Humber region.

First consider trends. Our estimate is 4.65 mm/year with 95% confidence interval of (3.37, 5.93). In Section 7.2.4, where only the annual maximum data from Table 9.1 were used, a trend estimate of 5.6 mm/year with 95% confidence interval of (3.4, 7.8) was

Distance	$z_{0.1}(1990)$	$z_{0.01}(1990)$	$z_{0.001}(1990)$	$z_{0.0001}(1990)$
1.0	4.32 (0.05)	4.90 (0.11)	5.56 (0.26)	6.32 (0.51)
5.0	4.36 (0.04)	4.89 (0.10)	5.46 (0.23)	6.10 (0.42)
10.0	4.43 (0.04)	4.89 (0.09)	5.38 (0.19)	5.90 (0.34)
15.0	4.50 (0.04)	4.91 (0.08)	5.33 (0.16)	5.76 (0.28)
20.0	4.58 (0.04)	4.95 (0.07)	5.31 (0.14)	5.66 (0.24)
25.0	4.66 (0.03)	4.99 (0.06)	5.30 (0.12)	5.60 (0.20)
30.0	4.74 (0.03)	5.05 (0.06)	5.32 (0.11)	5.57 (0.17)
35.0	4.83 (0.03)	5.11 (0.05)	5.35 (0.10)	5.56 (0.15)
40.0	4.92 (0.03)	5.18 (0.05)	5.40 (0.09)	5.58 (0.13)
45.0	5.01 (0.03)	5.26 (0.05)	5.45 (0.08)	5.61 (0.12)
50.0	5.11 (0.03)	5.34 (0.05)	5.52 (0.08)	5.65 (0.11)
55.0	5.21 (0.03)	5.43 (0.05)	5.59 (0.07)	5.71 (0.10)
60.0	5.31 (0.03)	5.52 (0.05)	5.67 (0.07)	5.78 (0.10)
65.0	5.41 (0.03)	5.62 (0.05)	5.76 (0.07)	5.85 (0.10)
70.0	5.52 (0.03)	5.72 (0.05)	5.85 (0.07)	5.93 (0.10)
75.0	5.63 (0.03)	5.82 (0.05)	5.95 (0.08)	6.02 (0.10)
80.0	5.74 (0.03)	5.93 (0.06)	6.05 (0.08)	6.12 (0.10)
85.0	5.85 (0.03)	6.05 (0.06)	6.16 (0.09)	6.22 (0.11)
90.0	5.97 (0.04)	6.17 (0.07)	6.27 (0.10)	6.33 (0.12)
95.0	6.09 (0.04)	6.29 (0.08)	6.40 (0.11)	6.45 (0.13)
100.0	6.22 (0.05)	6.42 (0.09)	6.52 (0.12)	6.58 (0.14)

Table 9.3: Spatial estimates of return levels (and standard errors), in metres relative to ODN, for the Humber estuary at 5 kilometre increments up the estuary.

obtained. This is consistent with this analysis which uses extended data sets and hence achieves a more accurate estimate.

Coles and Tawn (1990) use the annual maximum data and give return level estimates for nine of the sites in the present study. These results are based on separate estimation from site-to-site. Coles and Tawn (1990) also describe a spatial method of analysis which is applied to the Humber data: these are given here in Table 9.4. The reason that

Site	$z_{0.1}(1990)$	$z_{0.01}(1990)$	$z_{0.001}(1990)$
Immingham	4.55 (0.04)	4.97 (0.08)	5.39 (0.15)
Saltend Jetty	4.95 (0.03)	5.22 (0.05)	5.44 (0.10)
King Georges Dock	4.75 (0.03)	5.01 (0.05)	5.23 (0.09)
Humber Dock	4.75 (0.03)	5.01 (0.04)	5.24 (0.08)
Victoria Dock	4.85 (0.03)	5.12 (0.04)	5.34 (0.08)
St. Andrews Dock	4.99 (0.03)	5.26 (0.04)	5.47 (0.07)
Brough	5.36 (0.03)	5.63 (0.04)	5.84 (0.07)
Blacktoft	5.45 (0.03)	5.72 (0.05)	5.92 (0.09)
Goole	5.75 (0.03)	6.01 (0.06)	6.21 (0.12)

Table 9.4: Return level estimates (and standard errors) obtained by Coles and Tawn (1990).

some of the estimates in Table 9.4 are more accurate than given by the current analysis is that Coles and Tawn (1990) treated the trend as known in the statistical analysis – thus removing some of the true variability. The best fitting spatial model found by Coles and Tawn is identical in structure to the one found here except there $\log \sigma$ was taken to be linear with estuarine distance. Despite this the results are in a good agreement with those found here. The key advantage of the current approach is that, unlike in Coles and Tawn (1990) and Chapter 7, the location intercept parameter, α, is spatially modelled as well as β, σ and k, enabling estimates of return levels to be obtained for any site along the estuary.

Chapter 10

Case study IV: tide-surge interaction

10.1 Introduction

Throughout the report we have seen that estimation of extreme sea-levels in practice is complicated by deviations from the the ideal assumptions of independent and identically distributed surges. The most important deviation is the dependence between tides and surges. In shallow water areas dynamic processes, such as bottom friction, cause the tidal and surge components of the sea-level process to interact; in particular, surge values that occur at the time of a high tide tend to be damped, whereas surge values at rising tide typically are amplified. Accounting for interaction in the modelling of extreme surges is important since ignoring this feature and proceeding as if the processes were independent is liable to result in significant overestimation of extreme sea levels.

In the RJPM, summarised in Section 3.1, interaction is incorporated by normalising the surge series, and effectively applying standard extreme value techniques to this series, which does not depend on the tidal level. The normalisation is achieved by scaling the surge series by the tidal interaction functions a and b. It is known that the form of interaction tends to be similar at neighbouring sites, especially along the east coast of the UK (Dixon and Tawn, 1994, Section 6). In addition, estimation of these interaction functions can be imprecise, especially at sites with short data series. This suggests that a spatial model that allows a transfer of information on interaction from site to site would be beneficial. In this section we extract information on only interaction from each site, and examine its spatial behaviour along the east coast of the UK.

The standard oceanographic approach used to assess the level of tide–surge interaction is the examination of the standard deviation of the conditional surge process (Prandle and Wolf, 1978; Pugh, 1987; Walden *et al.*, 1982). Since here interest is in the behaviour of

the extreme surges, interaction is characterised using non-parametrically estimated high quantiles of the distribution of surges conditional on tidal level. Trends in the surge process are ignored throughout the section since interest here is purely in the tide-surge interaction, so the estimation of these functions is as in Dixon and Tawn (1994) rather than using the methods in Chapter 4 for the SRJPM.

Let Y_t and X_t be the hourly surge and corresponding tide series. Suppose that there exist functions of tidal level, $a(X)$ and $b(X)$ such that the location-scale normalisation of the surge series,

$$S_t^* = [Y_t - a(X_t)]/b(X_t) \tag{10.1.1}$$

is a stationary series, independent of tidal level. Then the methods for estimating the surge distribution in the tide-surge independence case can be applied on this transformed series to give estimates of the parameters of the annual maximum distribution of the transformed surge variable. These estimates can then be recombined with $a(X)$ and $b(X)$ to transform back and give an estimate of the distribution of the surge for all tidal levels, $F_{Y|X}(y|X)$.

The focus of this section is the estimation of the interaction functions $a(X)$ and $b(X)$. Following Dixon and Tawn (1994), if $a_1(X)$ and $a_2(X)$ are the 98% and 99% quantiles of the surge distribution conditional on tidal level X, with $a_1(X) < a_2(X)$,

$$a(X) = a_1(X) \quad \text{and} \quad b(X) = a_2(X) - a_1(X). \tag{10.1.2}$$

Estimation of these conditional quantiles, $a_1(X)$ and $a_2(X)$ can be imprecise, especially at sites with short data series. In Section 10.2 we investigate how a spatial model for the interaction functions can be used to provide better estimates of a and b at sites, and along a coastline. In Section 10.3 the method is applied to the east coast using data from eight of the A-class sites. The results provide increased understanding of the interaction process over tide and space and confirm hydrodynamical model forms for the spatial form of interaction along this coastline.

10.2 Spatially modelling extreme surges in the presence of tides

Based on the estimated interaction functions $a(X)$ and $b(X)$, the aim is to investigate how the form of tide-surge interaction varies along a coastline. The interaction functions correspond to physical features of the process, and since the sea-level processes are generally spatially coherent, a and b should vary smoothly from site to site, and exhibit dependence across nearby sites.

For each site, $a(X)$ and $b(X)$ both depend upon a combination of two physically distinct aspects of the process; the interaction between tides and extreme surges and the marginal distribution of extreme surges. Thus although these interaction functions should be spatially reasonably smooth, they provide little physical insight into the separate mechanisms of **interaction** and **marginal heterogeneity** along the coastline.

To help examine the form of spatial variation in the interaction, we separate these two processes of interaction and surge amplification. As, by definition, the surge has zero mean the principal change in marginal distribution from site to site is a scale change, so one way of distinguishing the two physical processes is to remove the amplification effect on the extremes by standardising the marginal extreme surge distribution at each site. This standardisation could be achieved by dividing the surge series by the sample standard deviation, but since our interest is in the upper tail, an alternative measure of variation, which is of more relevance, is a high quantile, q_s say. Hence the surge series is standardised at each site by scaling by q_s, i.e.

$$Y_t^* = Y_t/q_s. \tag{10.2.1}$$

Now define new interaction functions, a^* and b^*, by

$$a^*(X) = a_1^*(X) \quad \text{and} \quad b^*(X) = a_2^*(X) - a_1^*(X) \tag{10.2.2}$$

where $a_1^*(X)$ and $a_2^*(X)$ are the 98% and 99% quantiles of the normalised surge distribution conditional on tidal level, F_{Y^*}. The functions $a^*(X)$ and $b^*(X)$ measure only the interaction of the extreme surges and tide. These functions can then be mapped against coastal distance, d, and non-parametric spatial models $a^*(X,d)$ and $b^*(X,d)$ fitted to provide a model of the interaction process at any tidal level and coastal position. Similarly q_s can be mapped against coastal distance, and a non-parametric spatial model of marginal amplification $q_s(d)$ obtained for any coastal site. Consequently if we are interested in the interaction functions $a(X,d)$ and $b(X,d)$, at any coastal distance, these may be obtained from our separate spatial estimates of amplification and interaction by the property

$$a(X,d) = q_s(d)a^*(X,d) \quad \text{and} \quad b(X,d) = q_s(d)b^*(X,d). \tag{10.2.3}$$

10.3 Application to the east coast

The example data used here are hourly surge levels with corresponding tidal levels for eight of the sites in Table 2.1 along the east coast of the UK, from Wick to Sheerness. In order of coastal distance from Wick the sites used are Wick, Aberdeen, North Shields, Whitby, Immingham, Lowestoft, Southend, and Sheerness, shown on Figure 10.1. These

have been selected for this case study as they have long series of data and interaction can be well estimated.

First we construct our interaction functions $a^*(X)$, $b^*(X)$ for each site before considering the spatial versions $a^*(X,d), b^*(X,d)$. To aid comparison between sites, for each site we use the transformed tidal level X^*, as defined in Section 4.7, i.e.

$$X^* = F_{tide}(X) \tag{10.3.1}$$

where $F_{tide}(X)$ is the tidal distribution function. For the remainder of this section, we will estimate the interaction functions based on X^*, i.e. $a^*(X^*)$ and $b^*(X^*)$.

Define $f_{Y^*|X^*}$ to be the conditional density of the normalised surge conditional on the transformed tide. Then $a_1^*(X^*)$ and $a_2^*(X^*)$, the 98% and 99% quantiles of this conditional distribution, satisfy

$$q_i = \int_{-\infty}^{a_i(X^*)} f_{Y^*|X^*}(s)ds, \tag{10.3.2}$$

where $q_1 = 0.98$ and $q_2 = 0.99$. For a given q_i, $a_i^*(X^*)$ can be estimated from (10.3.2) if $f_{Y^*|X^*}$ is replaced by a bivariate kernel density estimate. The sampling variability of this estimator of $a_i^*(X^*)$ can be obtained by applying this estimation procedure repeatedly to blocks (years) of data, which produces a sample of i.i.d. estimates. Similarly, for these east coast data, the 98% surge quantile was found to be a suitable choice for q_s. For each of the eight sites, Figures 10.2 and 10.3 show the estimated functions $a^*(X^*)$ and $b^*(X^*)$ which are obtained from estimates of $a_i(X^*)$, $X^* \in [0,1]$.

Note that if the tide and surge were independent then for all X, $a^*(X^*) = 1$ and $b^*(X^*)$ is a constant. A clear interpretation can be given to $a^*(X^*)$ in particular, since the cases $a^*(X_0^*) = 1, < 1$ and > 1 correspond to the surge tail being unchanged, reduced and amplified respectively, at tidal state X_0^*, relative to the marginal surge distribution.

Due to the transformation of the tide and the surge normalisation in Section 10.2, both $a^*(X^*)$ and $b^*(X^*)$ can be compared over sites. The eight plots in Figures 10.2 and 10.3 show a gradual change in the profile of the functions over the sites, suggesting that interaction is a coherently changing process along this coast.

In Figure 10.4 estimates of $a^*(X_0^*)$ for each site, together with a kernel regression estimate, weighted by the standard deviation of the corresponding site based estimates, are plotted against coastal distance for values of $X_0^* = 0.1, 0.3, 0.5, 0.7$ and 0.9. These transformed tidal levels, X_0^* correspond to low tide through to high tide in steps of equal tidal state.

Also shown on Figure 10.4 are the weighted spatially smoothed estimates of $a^*(X_0^*, d)$. It is clear from Figure 10.4(a) that in the lowest tidal state there is a gradual change along the coastline from dampening of the surge at Wick to amplification at Immingham

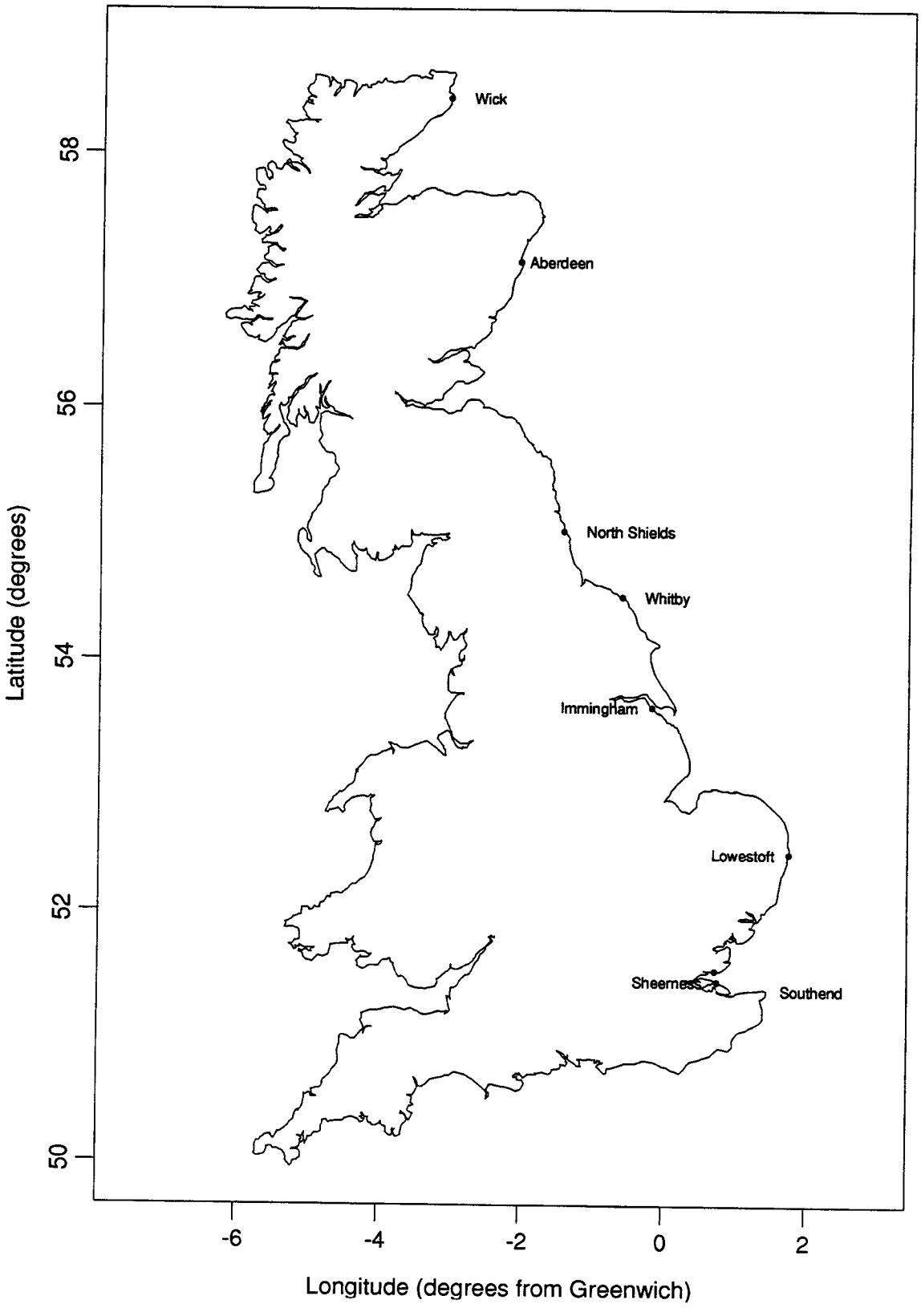

Figure 10.1: Map of the UK showing sites used in the tide-surge interaction east coast case study.

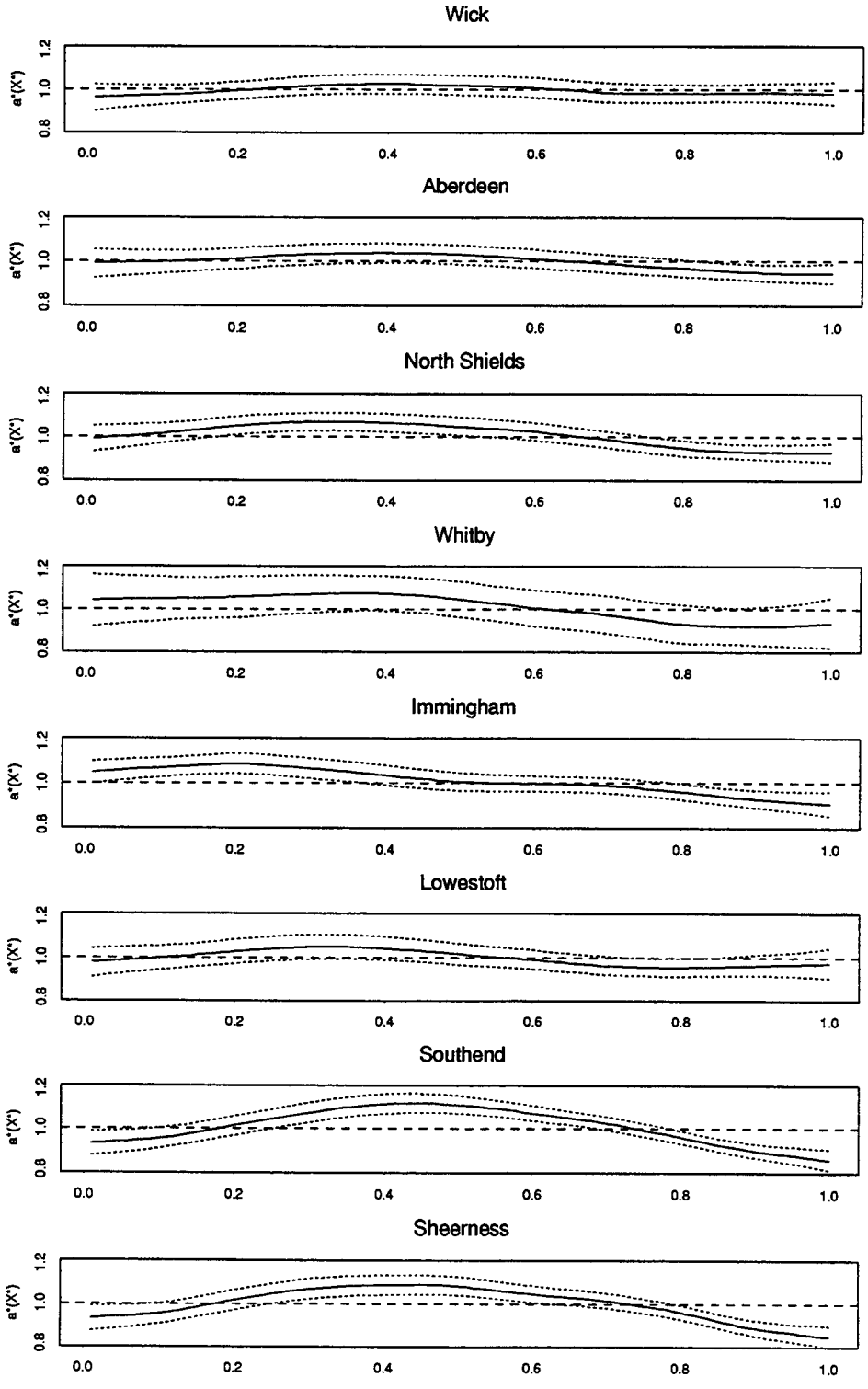

Figure 10.2: Estimates of the function $a^*(X^*)$, for each of the eight sites in increasing order of distance from Wick. The abscissa is standardised tidal level and the dotted lines are 95% pointwise confidence intervals. The dashed line $a^*(X^*) = 1$ is shown as a guide.

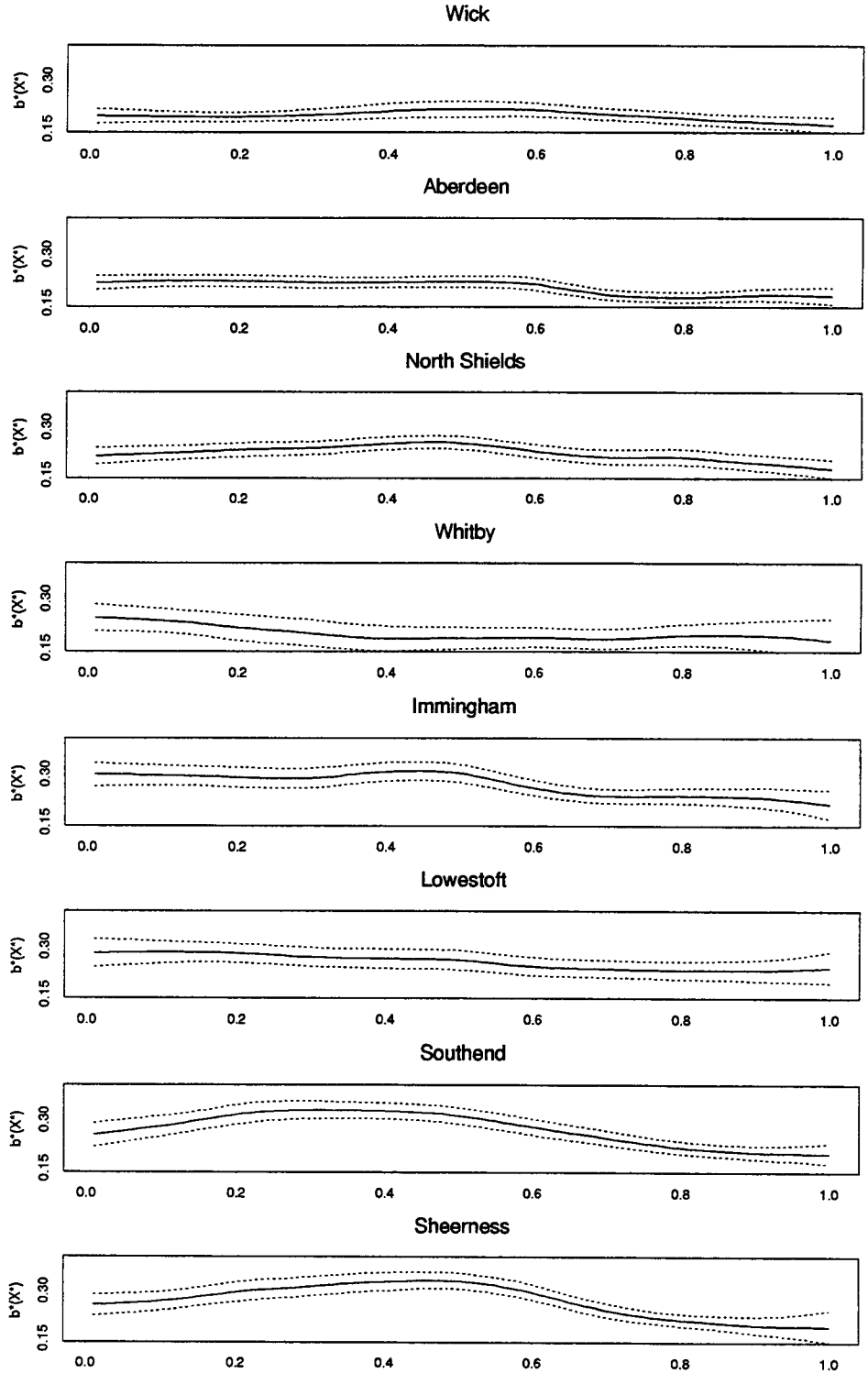

Figure 10.3: Estimates of the function $b^*(X^*)$, for each of the eight sites in increasing order of distance from Wick. The abscissa is standardised tidal level and the dotted lines are 95% pointwise confidence intervals.

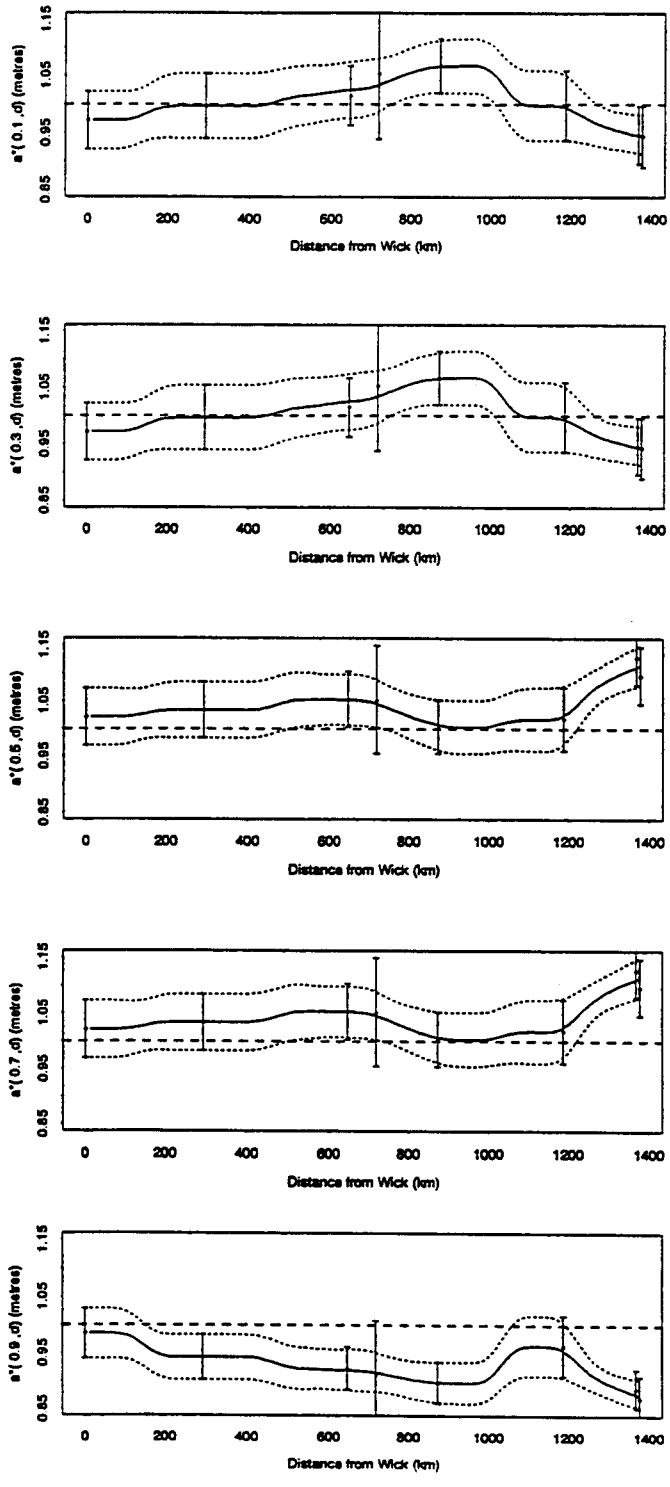

Figure 10.4: Estimates of the function $a^*(X^*)$ at tides $X^* = 0.1, 0.3, 0.5, 0.7$, and 0.9 with 95% confidence intervals, against distance (in km) from Wick, obtained from the estimates in Figure 10.2. Figures (a) to (e) correspond to $X^* = 0.1, 0.3, 0.5, 0.7$, and 0.9 respectively. The dotted lines are 95% pointwise confidence intervals. The dashed line $a^*(X^*, d) = 1$ is shown for each figure.

returning to a dampened form towards Sheerness. At mid-tidal levels, Figures 10.4(b)-(d), there is slight amplification at Wick which decreases to the Immingham–Lowestoft region before significantly increasing to Southend and Sheerness. The interaction process at high tidal states is the most important for extreme sea-level studies since the largest sea-levels typically result from large surges occuring at high tides. Figure 10.4(e) shows that a dampening effect of various magnitudes occurs for all the coastline at high tides. Furthermore the spatially estimated profile is remarkably similar to findings in Wolf (1978) where purely interaction terms were extracted from a hydrodynamical model of the sea-level process in the North Sea.

As described so far, these models provide information only about the interaction process. To obtain estimates of quantities more relevant to extreme sea-level analyses, these estimates need to be recombined with the marginal specific component of the model $q_s(d)$. Figure 10.5 shows estimates of q_s together with a smooth spatial estimate obtained by applying weighted kernel regression. What this figure shows is a gradual amplification of surges down the east coast with a significant increase around the Wash region between Immingham and Lowestoft. Viewed without either the smoothing or the confidence intervals, the estimate for Whitby looks inconsistent. One reason for this is that the sampling period for Whitby is the 1980s whereas most other sites also include 1970s data, which is generally recognised as being a milder decade in terms of storm surges, hence the smoothing process corrects for this sampling bias. In Section 12.2 we shall see that trends are another reason for this feature.

We have now obtained separately estimated smoothed spatial estimates for the two fundamental physical components of the conditional surge process: amplification, $q_s(d)$, and interaction, $a^*(X^*, d), b^*(X^*, d)$. Consequently, spatial estimates of the interaction functions, $a(X^*, d)$ and $b(X^*, d)$, for the tail of the conditional surge distribution can be obtained from equation (10.3.2). Alternatively $a(X^*, d)$ and $b(X^*, d)$ can be estimated directly and smoothed spatially. This latter approach has the disadvantage that although the two physical processes of amplification and interaction are separately spatially coherent there is no reason that when combined the resulting process should be as simple to interpolate. Figure 10.6 shows that the estimates obtained from both approaches when $X_0^* = 0.9$ give very similar results. The advantage of decomposition of the interaction functions is minimal in this case although generally the effect may be greater.

To summarise the findings in this study of interaction, provided that the functions a and b are such that (10.1.1) is independent of tidal level, and such that a and b can be estimated independently from the extreme tail estimation, then interaction can be examined separately by removing the amplification. This study also reveals that there is a simple spatially coherent pattern of interaction at extreme levels along the east coast.

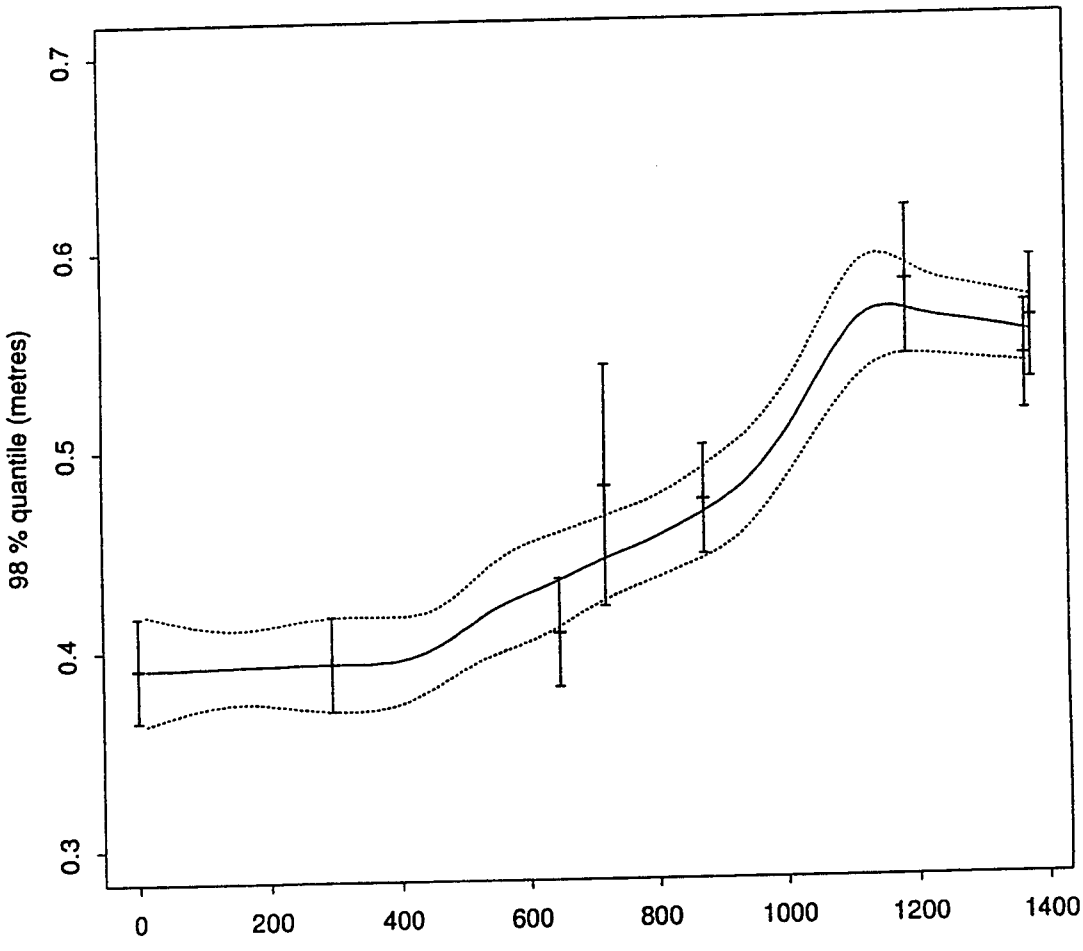

Figure 10.5: Estimates and associated 95% confidence intervals of the 98% quantile of the marginal surge distribution for each of the eight sites along the coast, against coastal distance. The solid (and dotted) lines represent the spatial smoothing (and associated 95% confidence intervals).

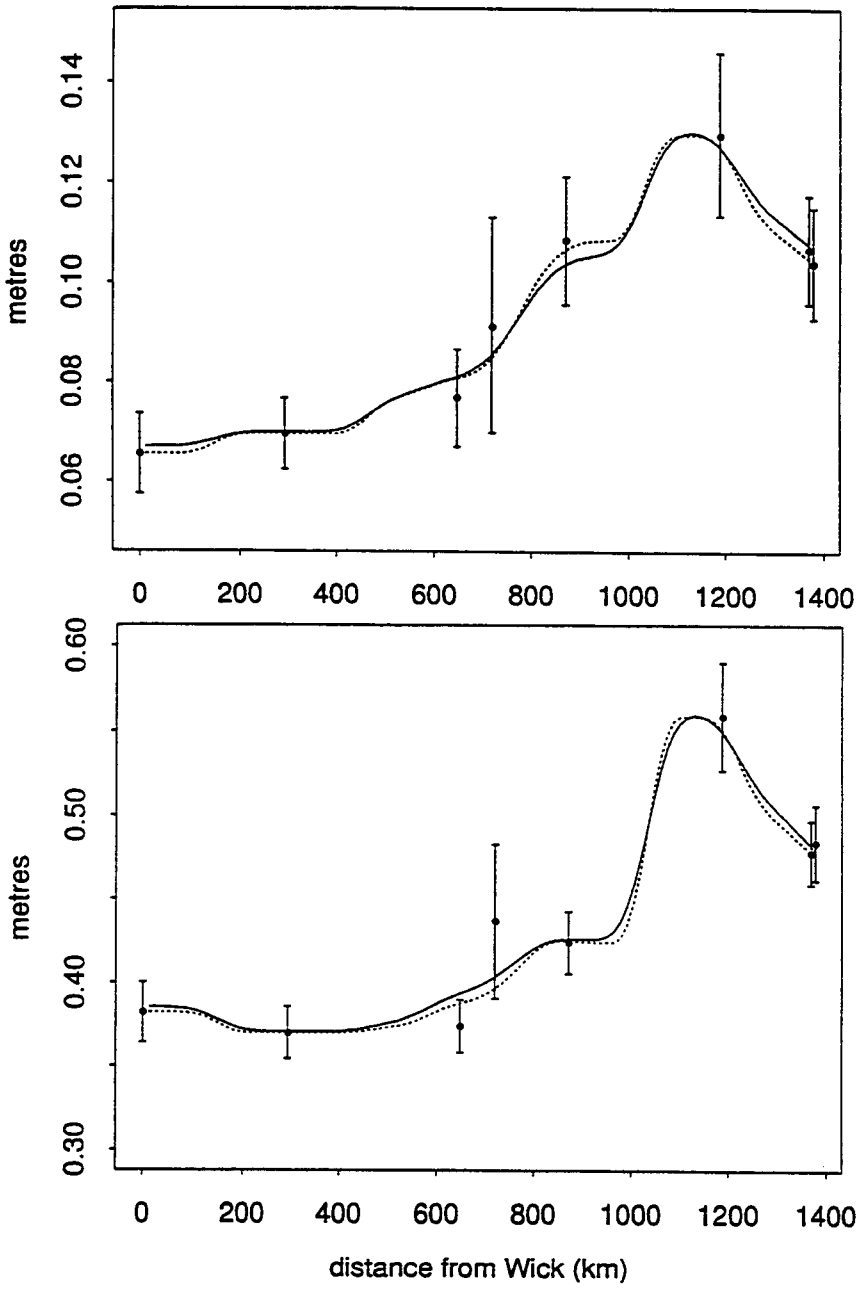

Figure 10.6: Estimates for $a^*(0.9)$ and $b^*(0.9)$, in figures (a) and (b) respectively, at each site with 95% confidence intervals. The lines are spatial estimates for $a^*(0.9, d)$ and $b^*(0.9, d)$ obtained by two methods: the broken line is the spatial estimate obtained by the direct smoothing of the $a^*(0.9)$ estimates, and the solid line is the estimate obtained by use of equation (10.1.1).

Chapter 11

Conclusions from the case studies

The findings from the four case studies, in conjunction with the site-by-site methods and results in Parts I and II, provide the basic framework for the spatial method of analysis used in Part IV. The key choices are whether to use

1. parametric or non-parametric functions for the parameters of the SRJPM, or
2. a likelihood analysis or a weighted least squares method for fitting the spatial model for the parameters,

and these aspects are considered separately in Sections 11.1 and 11.2. Also important is the handling of the trend estimation in the spatial analysis. In Section 11.3 our strategy for modelling the trend spatial is discussed.

11.1 Parametric or non-parametric parameter functions

The analysis in the case studies for mean sea-level trends and spatial modelling of the r-largest data in the Humber estuary are completely parametrically based; the analysis in the case studies for tide-surge interaction along the east coast is completely non-parametric; whereas the analysis of extreme sea-level trends involve a mixture of parametric and non-parametric models.

Parametric methods were applied in cases where the spatial variation of the parameter of interest was viewed to be simple enough to be well represented by a low-order parametric model. All the resulting models are such that they

- have simple spatial forms which can be represented in closed form,
- provide as much accuracy for the intermediate sites as at the data sites,

- provide good confidence intervals.

However in the case of the tide-surge interaction study the spatial variation in the interaction functions cannot be described by such a simple representation, so instead the data are allowed to determine the precise spatial function through a non-parametric kernel regression smoothing. Similarly, in the extreme sea-level trends study although simple smooth parametric relationships are a good representation for the scale and shape parameters of the generalised extreme value distribution, the trend and the intercept parameters exhibit a complex spatial pattern which cannot be expressed as any low order polynomial. There the trend parameter has spatial variation due to land level trends, which once removed leads to a much simpler form. This is not the case for the intercept parameter as tidal covariates cannot explain all the variation over the spatial scale under study. When the spatial scale is restricted to a local region, such as the Humber estuary case study, the variation is simple for spatial modelling of r-largest data.

In conclusion, as the SRJPM critically depends on the interaction functions, particularly on the east coast where interaction is an important component of the extreme sea-level process, we must have spatial models for the parameters which capture the complex variation exhibited by these functions. A non-parametric approach is adopted on the grounds of

- flexibility: when the estimates at the data sites suggest a complex spatial model is needed the non-parametric method takes such a form.

- robustness to model choice: although a level of smoothness needs specifying, that is a more natural choice than selecting between two parametric models and adopting a particular form.

11.2 Likelihood or weighted least squares fitting methods

Only in the mean sea-level trends and extreme sea-level trends case studies are we able initially to adopt a likelihood based approach to the spatial analysis, which fully accounts for spatial dependence in the estimates. Although likelihood methods are used in the Humber study, spatial dependence cannot be explicitly modelled owing to the lack of time history information in the data-base. The interaction study is not suitable for a likelihood based analysis as no statistical model is available upon which to base the likelihood function.

Of the two studies where likelihood methods are used, alternative fitting methods are ultimately used to extract the important information. For example, in the extreme

sea-level trends study a weighted least squares type approach is used whereas for the mean sea-level trends study a smoothness penalty function has to be imposed to obtain reasonable results. In each case the likelihood based methods are at their limit in terms of speed of fit and other computational problems, and so to extend these analyses to larger scale problems seems infeasible.

Owing to the high dimensionality of the data structure and the complexity in specifying a full likelihood based model for a spatial revised joint probability method we use the simpler weighted least squares type of method for fitting our spatial models.

11.3 Handling of trends

A key finding from the case studies concerns the handling of trends in a spatial analysis. In particular, due to short record spans if there is a poor estimate of the trend at one of more sites this can significantly distort the true smooth spatial pattern of other parameters of the model. Examples where this is most evident are

1. In the extreme sea-level trends study, Chapter 7, the poor trend estimate for Bembridge in the annual maximum analysis caused the corresponding scale and shape parameter estimates to differ significantly from those at neighbouring sites.

2. In the mean sea-level trends study, Chapter 8, the poor trend estimates for Walton, Harwich and Felixstowe caused the corresponding intercept parameter estimates to differ significantly from those at the neighbouring sites for which a reliable trend estimate could be obtained.

3. In the study of the Humber estuary, Chapter 9, when a common trend was fitted for all sites, the intercept parameters had a very simple linear dependence on distance up the estuary, yet when no spatial estimate of the trend was used the intercept parameter had a more complex spatial variability, and so there was no obvious spatial model for that parameter, see Figures 9.3–9.7.

4. In the tide-surge interaction study, Chapter 10, trends were not considered and the spatial inconsistency of the Whitby estimate was partially due trends.

These findings suggests that a good estimate of the trend is required before any other aspect of the extreme sea-level process is considered spatially. Therefore the approach we take in Part IV of the report for the spatial analysis of the east coast data is to first spatially estimate the trend using all possible sources of data. This spatial estimate of the trend will then be used when all the other parameters of the model are considered.

Part IV

The spatial model: application and results

Chapter 12

Application to east coast – SRJPM parameters

12.1 Tides

12.1.1 General description

Firstly we describe a procedure to determine the tides at any point on the coastline. In general tides are only known accurately where a set of measurements exist: in this case the objective is to produce tidal values for locations where there have never been measurements. To achieve the required degree of accuracy, the problem was approached by using the sites of the National Tide Gauge Network, and any suitable intermediate ports, as reference points. Between these reference, or primary, points the tide is generated by a spatial interpolation scheme which is discussed in detail below. The tide at the primary stations is of course known to a high degree of accuracy. At intermediate points an attempt is made to maintain this accuracy by employing output from a numerical model as a base function for the interpolation scheme.

Previous work on the Humber Estuary (discussed in Chapter 9) has shown that it was possible to generate tides at intermediate points in an estuary situation using a set of primary stations and a conventional but non-linear interpolation scheme. Cubic splines were fitted to tidal predictions for each of six reference points and used to derive an assumed prediction for the remainder of the estuary.

A pure interpolation scheme of this type was considered insufficiently accurate for the east coast as the distance between the ports is much greater than in the estuary situation. Instead the primary stations were used in conjunction with the output from the European shelf 12km numerical tidal model (Flather and Smith, 1993) which acted as a base function for the interpolation. In this way the model represented the properties of

the tide and controlled its propagation characteristics between the various primary points.

Differences were calculated between the model and each primary point as a representation of the relationship between the two. These differences were then spatially interpolated using 'tilt' and 'bias' coefficients and used to adjust the model tide at each of the intermediate points. In practice the technique was applied in the time domain although it could in principal be used in frequency space. Interpolation of the harmonic constants, amplitude, h, and phase, g, or to be exact in combined form $h\cos g$ and $h\sin g$, proved to be insufficiently accurate, as it was found to perform better for the major constituents than for the minor terms.

By this technique it was possible to generate the tidal distribution for sites at a regular spacing along the coastline. For sites where observed data was available this was found to be a near representation of the actual distribution, and at intermediate sites the change in distribution was found to capture physical changes which were known to occur in those coastal regions. Although it was not possible to calculate confidence limits directly, an attempt was made to estimate the accuracy of the method. This is discussed below.

12.1.2 Shoreline distance metric

The first requirement of the method is that a reference coastline is adopted and adhered to. All primary stations, interpolated points, and many calculations are related to this coastline. It must be in a digitised form to be useful for calculating along track distances which raises the question of resolution. Obviously it must be sufficiently accurate such that all stations will actually lie on the digitised form of the coastline. However an over accurate representation leads to problems of tides in estuaries and in rivers. The method is not designed with estuaries or rivers in mind as the model grid does not extend this far. When it is necessary to generate tides in shallow water/non-linear areas, alternative procedures such as the methods adopted in the Humber estuary case study (Chapter 9) can be used.

We have adopted a relatively high quality version of the east coast, accurate to approximately 100 metres, and named it the Shoreline Distance Metric. The problem of estuaries has been resolved by limiting any chosen position to a fixed distance from the nearest model grid point.

The coastline is a subset of the World Vector Shoreline compiled by the U.S. Defence Mapping Agency and supplied by the British Oceanographic Data Centre. It has been reorganised only in the sense that each grid point follows the natural coastline in sequential order. Originally the World Vector Shoreline is not ordered, parts of Cornwall are adjacent to Yorkshire in the data file, for example. Our ordering begins at a point near Wick and progresses clockwise around the mainland.

The coastline can be used to calculate the distance between points following the coastline as opposed to the geographic distance between the points. This coastline has therefore been designated the Shoreline Distance Metric (SDM). In a similar fashion the model grid points adjacent to the coast follow a Model Distance Metric (MDM). This has a coarser resolution of approximately 12km. Model grid points are identified with the nearest point on the SDM thus making it easier to extract the appropriate model harmonic constituents for the interpolation scheme.

12.1.3 Primary stations

Table 12.1 contains a list of tide gauge stations and their geographic locations which were used as primary reference points. They include most of the National Tide Gauge stations on the east coast and several other ports for which a good quality sea level record is available. The suitability of a site as a reference station was judged on the quality of its data. Generally at least a year of good quality tidal measurements had to be available before it was considered applicable.

All the stations have been analysed at one time or another for their harmonic constants. These constants act as the source of tidal predictions for the reference stations. At each station the number of constants differs according to the complexity of the tidal regime. In the very non-linear areas of the southern North Sea additional shallow water constituents are included to represent the severe distortion of the tide in this area. At a minimum the number of constituents used for any primary station is 60 but is more usually 100.

The stations have been chosen to give a reasonable spatial coverage down the east coast. Stations which are either too close together or are in the inner reaches of an estuary relative to the hydrodynamic model have been eliminated.

12.1.4 Hydrodynamic model grid

The harmonic constituents were extracted from the shelf surge-tide operational model - CS3 - which has a resolution of $1/60^\circ$ in longitude by $1/90^\circ$ in latitude, corresponding to 12km (Flather and Smith, 1993). The model is forced by fifteen constituents on the boundary and analysed for 50 model generated constituents at each grid point. For the present purpose only the model grid points adjacent to the coast are of interest and have been extracted (Figure 12.1).

Using 50 model constituents is not at odds with the number of constituents from the primary stations as they give a reasonable representation of the tide. Any departures from this simplified tidal representation is absorbed by the adjustment process, of bias and tilting, in the interpolation scheme.

Place	Latitude	Longitude	No. of harmonic constants
Wick	58°26N	3°05W	102
Cromarty Firth	57°41N	4°01W	60
Buckie	57°41N	2°57W	61
Aberdeen	57°09N	2°05W	101
Leith	55°59N	3°11W	108
St Abbs	55°54N	2°07W	62
Seahouses	55°35N	1°39W	62
River Tyne (North Shields)	55°00N	1°26W	102
River Tees Entrance	54°38N	1°09W	66
Whitby	54°29N	0°37W	98
Scarborough	54°17N	0°23W	60
Bull Sand Fort	53°33N	0°04E	101
Hull	53°44N	0°21W	102
Immingham	53°38N	0°11W	110
Inner Dowsing	53°19N	0°33E	95
Cromer	52°56N	1°18E	104
Hunstanton	52°56N	0°29E	62
Lowestoft	52°28N	1°45E	109
Felixstowe	51°57N	1°21E	107
Harwich	51°57N	1°17E	110
Walton-on-the-Naze	51°51N	1°16E	70
Coryton	51°32N	0°31E	100
Southend	51°31N	0°43E	106
Shivering Sands	51°30N	1°05E	97
North Woolwich	51°30N	0°05E	100
London Bridge	51°30N	0°05W	100
Tilbury	51°27N	0°22E	100
Sheerness	51°27N	0°45E	111
Margate	51°24N	1°23E	99
Ramsgate	51°20N	1°25E	60
Dover	51°07N	1°19E	113
Newhaven	50°47N	0°04E	54

Table 12.1: Tide gauge stations used as reference points for the east coast spatial interpolation scheme for tides.

Figure 12.1: Map showing the British coastline taken from the World Vector Shoreline. The solid line represents the Shoreline Distance Metric used to locate coastal stations and their shoreline distances from an origin near Wick. The grid boxes are those from the CS3 shelf surge-tide operational model. Only the grid boxes adjacent to the coastline are shown.

12.1.5 The difference interpolation scheme

Software has been produced by the Proudman Oceanographic Laboratory to perform the interpolation scheme automatically. It works in the following manner.

A point (X, Y) is chosen by the operator on the SDM at which the tides are required. The software then calculates the equivalent point on the MDM say (x, y). In practice (x, y) will probably lie between two adjacent model points (x_1, y_1) and (x_2, y_2). Interpolation of $h\cos g$ and $h\sin g$ between (x_1, y_1) and (x_2, y_2) is therefore required as follows

$$\begin{aligned}(r_1/r_2)[h_2\cos(g_2) - h_1\cos(g_1)] + h_1\cos(g_1) &= h\cos(g)\\ \text{and } (r_1/r_2)[h_2\sin(g_2) - h_1\sin(g_1)] + h_1\cos(g_1) &= h\sin(g)\end{aligned}$$

where

$$r_1 = [(x - x_1)^2 + (y - y_1)^2]^{1/2} \text{ and } r_2 = [(x_2 - x_1)^2 + (y_2 - y_1)^2]^{1/2}.$$

One year of predictions is generated from these interpolated values of $h\sin(g)$ and $h\cos(g)$. This is the model time series that has to be adjusted by the estimated tilt and bias correction. We denote this series, at time t, by $c(t)$.

12.1.6 Estimating the tilt and bias

Two reference stations are chosen on the SDM that lie on either side of (X, Y), positions A and B. One year of predictions are generated for each of these two sites and for the two equivalent points on the MDM.

The differences between the SDM and MDM series are formed. We denote these difference series by $a(t)$ and $b(t)$ at time t, and let R_{ba} be the distance between reference stations A and B on MDM, and R_{xa} be the distance between reference station A and (x, y) on MDM. For each hour, t, of the year we have:

$$\begin{aligned}\text{Bias}(t) &= a(t)\\ \text{Tilt}(t) &= (b(t) - a(t))/R_{ba}.\end{aligned}$$

The resulting estimated tidal series is then

$$C(t) = c(t) + R_{xa} \times \text{Tilt}(t) + \text{Bias}(t).$$

The process is repeated for each hour of the year to produce an estimated tidal time series $C(t)$ for the desired position (X, Y). This series can then be analysed to provide its harmonic constituents.

12.1.7 Tests

Some tests were performed to assess the accuracy of the method. In Table 12.2 the derived values from the interpolation scheme are compared with actual values for the harmonic constituents. In particular, these results show that the interpolated amplitude and phases of the constituents $M2$, $S2$, $K1$ and $O1$ agree well with actual values.

site	grid reference	source	M2		S2		K1		O1	
			h	g	h	g	h	g	h	g
Invergordon	57°41N 4°10W	A	136	335	48	012	11	180	12	034
		B	133	336	47	013	12	179	12	034
Blyth	55°07N 1°29W	A	160	087	55	126	13	234	14	081
		B	158	087	53	129	11	242	14	080
Dunbar	56°00N 2°31W	A	161	055	56	096	11	219	13	067
		B	161	056	55	097	11	221	14	065
Walton	51°51N 1°16E	A	143	330	40	024	11	358	13	177
		B	141	331	40	024	10	357	13	177

Table 12.2: Constituents $M2$, $S2$, $K1$ and $O1$ in centimetres and degrees. Rows denoted A and B refer to values from the Admiralty Tide Tables and from the interpolation scheme respectively. The terms h and g denote amplitude and phase respectively.

12.1.8 Spatial mapping of tides

Figure 12.2 shows the 90%, 95%, 99% and 99.9% tidal quantiles for each data site (shown as a + symbol) and the continuous spatial tidal quantile curves for the whole coastline, obtained from the Proudman Oceanographic Laboratory's tidal interpolation routine discussed earlier in this section. Note that in Figure 12.2 the quantiles are based on a zero mean sea-level at each site, and are independent of the datum used to obtain the measured data for each site. Out of all the data sites only the quantiles for Harwich and Walton are not near perfectly described by the spatial estimate, and even in these cases the spatial estimate provides only a slight underestimation.

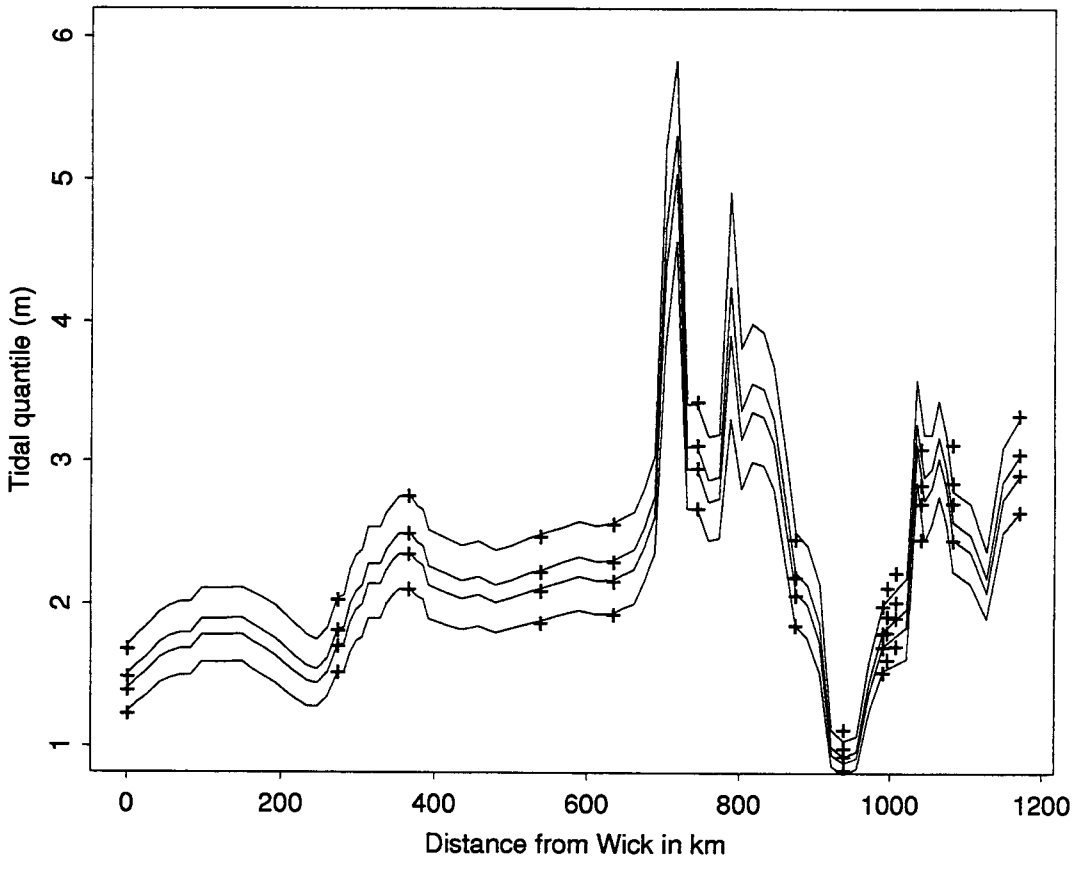

Figure 12.2: Quantiles of the tidal distribution at the sites, and at all distances obtained using the tide interpolation model. The quantiles shown are the 90%, 95%, 99% and the 99.9% levels. The + symbols indicate the corresponding quantiles, based on tidal predictions from actual observations, for the data sites.

The interpolation provided by the tidal scheme is highly non-linear between data sites in the areas around the Firth of Forth, the Humber estuary, the Wash, and the southern North Sea. This plot is based on estimates of the tidal quantile at 20km intervals along the east coast. In the Humber and Wash regions the tide changes at such a large rate

that 20km intervals are too coarse for a smooth plot, and so a higher resolution grid is required in these regions. In such regions the local bathymetry which is responsible for the non-linear spatial changes in the tidal characteristics will also influence aspects of the surge distribution. The spatial estimate of return levels that we obtain in Chapter 13 is based on data from insufficient sites in these complex coastal regions to calibrate any change in surge distribution, so we are not confident in resolving the present spatial estimate to a finer scale. In the third stage of the project the spatial estimate exploits surge data which possess much greater spatial resolution over these regions as synthetic data from the tide-surge numerical models are available. These complex coastal regions will be further studied in Stage 3 of the project.

12.2 Trends

Motivated by the findings of the spatial case studies, see Chapter 11, the first random component of the extreme sea-level process we spatially model is the trend. The two processes which influence the spatial behaviour of observed extreme sea-level trends are

- eustatic trends
- isostatic trends (vertical land level trends).

As interest here is in the observable trend at a site, i.e. the composition of the isostatic and eustatic trends, the case studies have shown that it is better to spatially estimate this overall trend rather than to separately spatially estimate these two components.

It is vital to estimate the trends as accurately as possible as they influence the spatial smoothness of other parameters and the precision of design levels for years in the future. There are two sources of information about trends:

- **Trends at sites in the study:** a key theme of Part I was the development of a model for extreme sea-levels at a site which exploited all relevant information from that site. The information took the form of data on hourly sea-levels and historical data on annual maximum and mean sea-levels. Of the extra information provided by the historical data, most was found to be in terms of the trend at the site.
- **Trends at intermediate sites:** in the case studies of Part III, sites which only have historical data available are used. These data are found to be particularly helpful in spatially mapping the trend.

In this study we base the spatial estimate of the trend on the best available estimate of the trend at sites along the east coast. These site-by-site trend estimates, and respective

95% confidence intervals are shown on Figure 12.3. The site label abbreviations given on this, and subsequent figures in Part IV are defined in Tables 12.3 and 12.4. For each site the trend estimate is obtained using one of the following methods:

- the SRJPM of Part I (estimates are given in Part II),
- the annual maximum method (estimates are given in Chapter 7),
- the mean sea-level data (estimates are given in Chapter 8).
- a weighted average of estimates of the annual maximum and mean sea level trends when both are available but hourly data are not.

Information concerning the sites and form of the additional data used to provide estimates for sites which are not part of the study is given in Table 12.4.

Site	Abbreviation
Wick	wic
Aberdeen	abe
Leith	lei
North Shields	nor
Whitby	whi
Immingham	imm
Cromer	cro
Lowestoft	low
Felixstowe	fel
Harwich	har
Walton on the Naze	wal
Southend	sou
Sheerness	she
Dover	dov

Table 12.3: Site abbreviations for sites with hourly data which are used in the figures in Part IV.

These site-by-site estimates generally reflect the true value of the trend, typically lying in a range $0 - 6$ mm/year, with some notable exceptions such as Cromer, Felixstowe and Margate which have large negative trends, and Saltend Jetty, Marsh Road Sluice and Wisbech which have large positive trends. The length of record, the form of data, and the self consistency of the data at the site affect the uncertainty of the estimates which is much

Site	Abbrev.	Annual maxima			Mean sea-levels		
		Number	From	To	Number	From	To
Invergordon	inv	NA	NA	NA	12	1960	1971
Buckie	buc	NA	NA	NA	13	1971	1982
Methil	met	38	1934	1977	NA	NA	NA
Kirkaldy	kir	28	1951	1978	NA	NA	NA
Rosyth	ros	31	1914	1977	27	1964	1990
Grangemouth	gra	34	1934	1978	NA	NA	NA
Dunbar	dun	NA	NA	NA	38	1914	1950
Goole	goo	59	1920	1978	NA	NA	NA
Blacktoft	bla	56	1921	1977	NA	NA	NA
Brough	bro	56	1922	1977	NA	NA	NA
St Andrews Dock	and	13	1965	1977	NA	NA	NA
Humber Dock	hum	42	1920	1968	NA	NA	NA
King Georges Dock	kin	36	1922	1973	NA	NA	NA
Saltend Jetty	sal	13	1965	1977	NA	NA	NA
Grimsby	gri	54	1920	1973	NA	NA	NA
Boston	bos	59	1920	1978	NA	NA	NA
Lawyers Sluice	law	26	1953	1978	NA	NA	NA
Wisbech	wis	22	1957	1978	NA	NA	NA
Kings Lynn	ki2	119	1860	1978	NA	NA	NA
Grt Yarmouth	yar	78	1899	1976	NA	NA	NA
Holland on Sea	ho2	53	1934	1986	NA	NA	NA
Colchester	col	43	1944	1986	NA	NA	NA
Tilbury	til	46	1929	1977	23	1961	1983
Tower Pier	tow	49	1929	1977	NA	NA	NA
Margate	mar	10	1968	1977	NA	NA	NA

Table 12.4: Available annual maximum and mean sea-level data at the additional east coast sites.

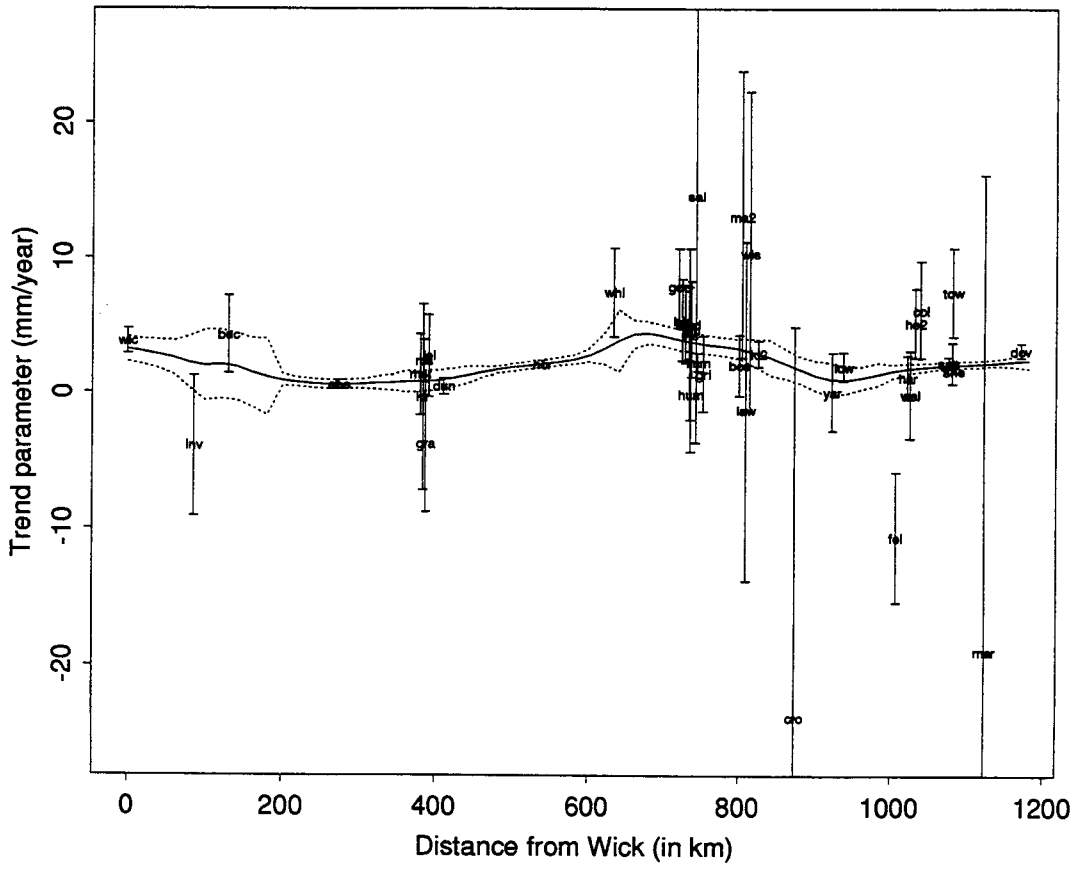

Figure 12.3: Trend parameter against distance; the site-by-site estimates for the best method at each site and the spatial estimate based on all site-by-site estimates with approximate 95% confidence intervals.

larger at some sites than others. So some sites with long series of annual maximum values, such as Kings Lynn, can have smaller confidence intervals than sites with a reasonable holding of hourly data, such as Whitby. Generally though the well-established A-class sites give the most precise trend estimates.

Also shown on Figure 12.3 is the fitted smooth spatial trend with associated 95% confidence interval. As discussed earlier, the form of spatial modelling that is adopted weights the spatial fit to the site-by-site estimates in accord with the standard error at each site and the proximity and value of trend estimates from neighbouring sites. It is seen that for sites which are quite a distance from neighbouring sites, such as Wick, Aberdeen and North Shields, the spatial trend is basically the site-by-site trend with the confidence interval also being that of the site-by-site estimate. In other words, for isolated sites which have a precise site-by-site trend estimate, the spatial trend is approximately the same as the site-by-site estimate. When the site-by-site estimate for an isolated site is poor, then the spatial information plays more of a role in producing the spatial estimate.

For example, for Whitby the spatial trend is lower than the estimate for the site, and for Invergordon and Buckie the spatial trend simply averages the separate site-by-site estimates of the trend.

The contrasting situation is when sites are grouped closely together, such as occurs in the Firth of Forth, the Humber, the Wash and the region around the Thames estuary. Then the spatial trend acts as a local weighted average of the site-by-site estimates so largely ignores erroneous trend estimates, such as Saltend Jetty, Cromer, Felixstowe and Margate.

As shown on the scale of Figure 12.3 the spatial trend is difficult to interpret as it is relatively constant. Figure 12.4 shows this spatial trend estimate plotted on a more appropriate scale without the site-by-site estimates. This plot gives a clearer picture of

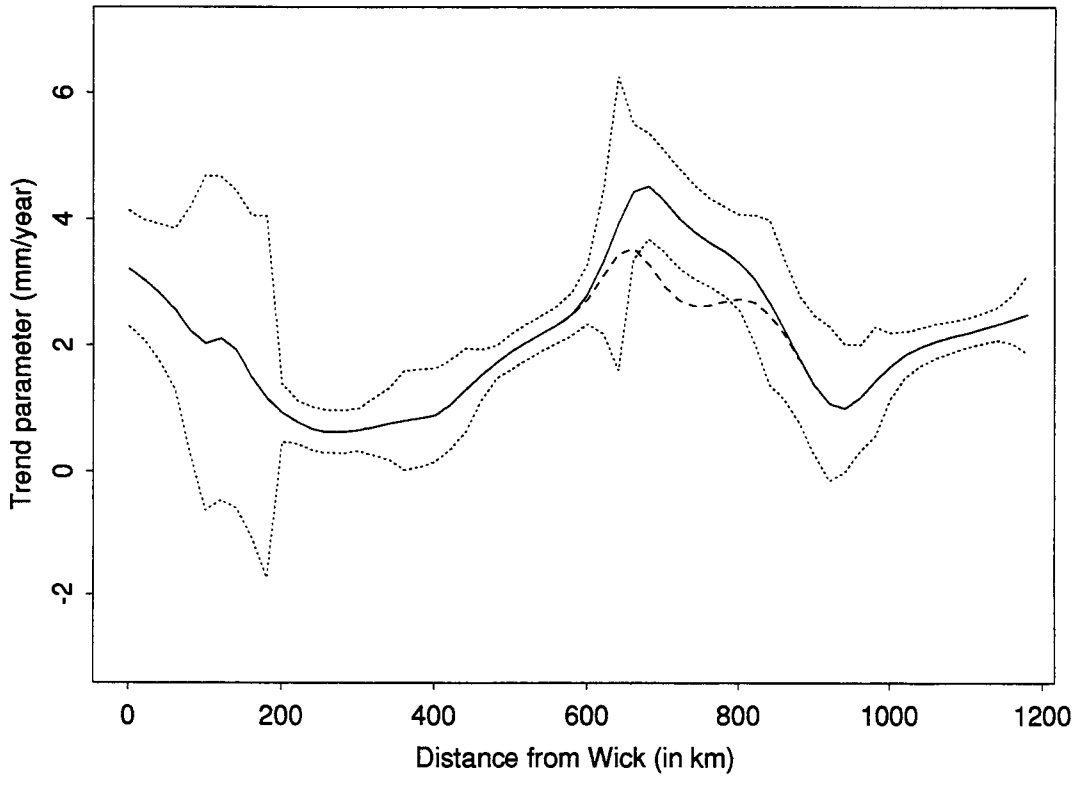

Figure 12.4: The spatial trend estimate plotted against distance. The continuous and broken lines are the spatial estimates including and excluding the Humber estuary data respectively. The dotted lines show the associated 95% confidence intervals for the continuous line spatial estimate.

the spatial trend estimate. The trend is at its highest in and around the Humber and at Wick, with low points in southern Scotland and around Lowestoft. As there is so little

data in the proximity of Wick the trend estimate is largely influenced by the estimate at the site. As Wick is the boundary of our spatial analysis there is no scope to use spatial information from further north even if suitable data exist. The spatial trend value is contrary to information about the isostatic trend in northern Scotland but is consistent with the findings in the case study for mean sea-level data, which showed that the trend at Wick is much greater than predicted by the combined trend given by isostatic and eustatic changes.

The large estimated trends around the Humber estuary are more problematic. In the case study on annual maximum sea-level trends the trend in the Humber was found to be significantly larger than on the neighbouring open coast. Using increased data, the case study of the Humber estuary confirmed the presence of a homogeneous trend along the length of the Humber. From the site-by-site estimates in Figure 12.3 we see that the estimates for all sites in the Humber region, other than Humber Dock, Immingham and Grimsby, are above the spatial trend. This suggests that these estuary site trend estimates are biasing the spatial trend estimate. To examine this, estimation of the spatial trend is repeated omitting all Humber estuary sites other than Immingham and Grimsby. The resulting trend estimate is shown as the dashed line on Figure 12.4. This estimate is coincident with the previous estimate apart from in the interval from North Shields to Cromer, where it is up to 1mm/year less than the previous estimate. This updated estimate lies outside the confidence interval for much of that coastal stretch, showing that the Humber estuary sites are highly influencing the spatial trend estimation.

The spatial trend estimate for the east coast is taken to be the second spatial estimate, for all open coastline sites; this estimate with associated 95% confidence interval is shown in Figure 12.5. The spatial trend estimate for the Humber is a homogeneous estimate of 4.65mm/year from Immingham to Goole, this being the estimate obtained in the case study of Chapter 9. The 95% confidence interval for the trend in the Humber estuary is $(3.37, 5.93)$mm/year, which does not overlap the corresponding confidence interval for the spatial trend estimate along the neighbouring open coast (see Figure 12.5).

This spatial estimate is more precise than the trend estimates obtained using data for a single site at a time. In the remainder of the analysis the trend is fixed at the spatial estimate before estimating the other features of the model. As discussed in Section 11.3 this use of the spatial model for trend instead of the site estimates should give site-by-site estimates of the other parameters which exhibit greater spatial smoothness and thus simplify the physical interpretation of the parameters.

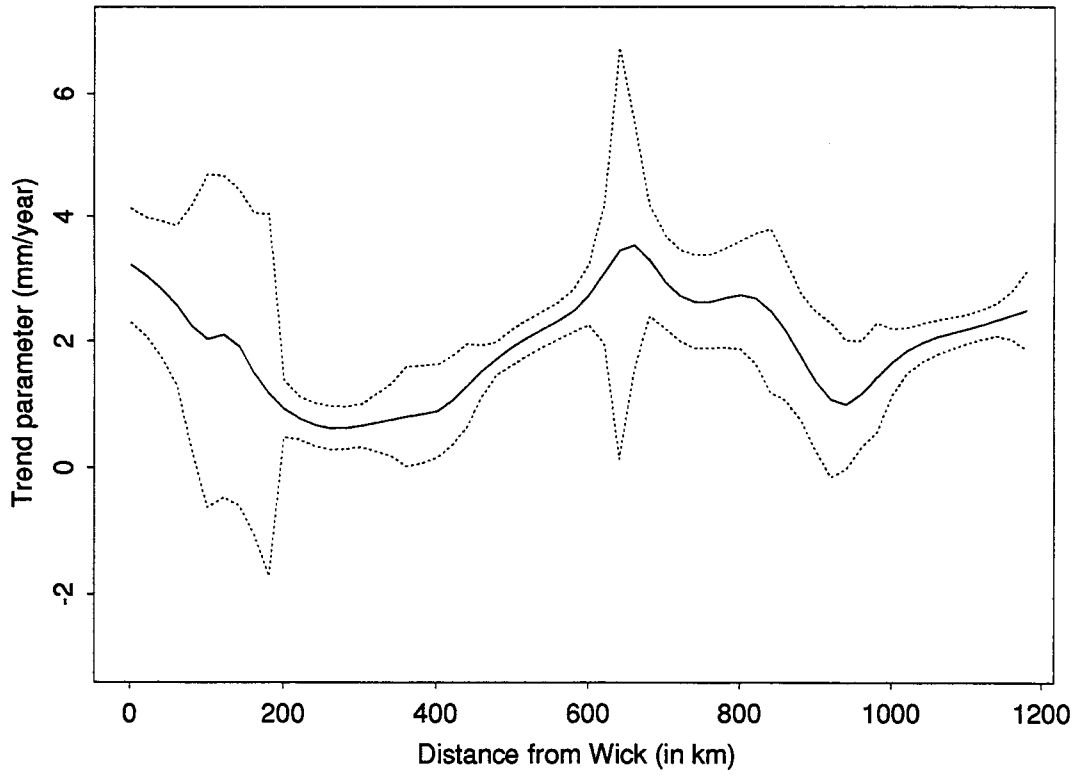

Figure 12.5: The spatial trend estimate and associated 95% confidence intervals obtained from all site-by-site estimates except for the Humber estuary.

12.3 Interaction functions

In this section we estimate the two interaction functions $a(X, i)$ and $b(X, i)$ for any site along the east coast. As we will only show results for these functions for a single year, 1990, we drop the dependence on i from the notation. So $a(X) = a(X, 1990) = a(X, i_0) + \beta(1990 - i_0)$.

Following the notation of the case study on tide-surge interaction (Chapter 10) we denote these functions, at distance d from Wick, by $a(X, d)$ and $b(X, d)$. The second index denotes distance and not the year as used in Part I. Also denote the 98% quantile of the surge distribution for a site at distance d from Wick by $q_s(d)$. In this section we develop spatial models for $a(X, d)$ and $b(X, d)$ in 1990. In the case of $a(X, d)$ two different approaches will be considered, whereas based on the findings in Chapter 10 only one approach is used for $b(X, d)$. In addition denote the interaction function $a(X, d)$ normalised by $q_s(d)$ by

$$a^*(X, d) = \frac{a(X, d)}{q_s(d)},$$

and let T_j, $j = 1, \ldots, 10$ be the mid-point of the jth equi-probable tidal band, i.e. T_j satisfies $F_{tide}(T_j) = (j - 0.05)/10$.

In Figure 12.6 the site-by-site estimates of $a(T_j)$, for $j = 1, \ldots, 10$, are plotted for 1990. These estimates depend strongly on the trend at each site and here the spatial trend is used. The corresponding plot (not shown) obtained using the site-by-site trend estimates shows markedly more variability, so Figure 12.6 goes a long way to justify our choice to working with the trend fixed at the spatial estimate. Also shown on Figure 12.6 is the spatial estimate of these interaction functions, i.e. the spatial estimate of $a(X, d)$, shown here for $a(T_j, d)$, with $j = 1, \ldots, 10$. As seen in Chapter 10 this function varies smoothly in space and with respect to tidal level. The spatial profile of the function is similar for all tidal levels, except for possibly the top tidal band. The general pattern is that there are near constant levels from Wick-Leith and Immingham-Dover with a steep increase between these two regions. Interaction is largest in the Immingham-Dover region, so for this coastal stretch at these tidal levels the surge is typically larger than for the rest of the coast. In the top tidal band, the profile shows a single peak between Cromer and Lowestoft, suggesting that for high tidal levels at a site, the surge is notably larger there than elsewhere on the coast.

As measured by $a(X, d)$ the interaction function combines two separate physical mechanisms: pure tide-surge interaction and marginal heterogeneity. As seen in Chapter 10, the interaction function $a(X, d)$ can be decomposed into statistics which separately measure these two mechanisms:

- $a^*(X, d)$ which measures purely tide-surge interaction effects
- $q_s(d)$ which measures purely the amplification of the surge.

Figure 12.7 shows site-by-site and spatial estimates of $a^*(T_j, d)$ for $j = 1, \ldots, 10$. When this function is equal to one this corresponds to there being no interaction effect, i.e. the surge level at that tidal state behaves as if the surge and tide were independent. When this function is greater than, or less than, one this corresponds to interaction amplifying, and dampening, the surge respectively. Most important for extreme sea-levels is that at the top tidal band, i.e. for high tides, interaction dampens the surge everywhere, with the least effect in the North and around Lowestoft. At mid-tide there is an amplification of the surge in the southern North Sea region. The spatial estimate of $a^*(T, d)$ is similar to that obtained in the tide-surge interaction case study in Figure 10.4.

Figure 12.8 shows the 98% surge quantile, $q_s(d)$, plotted in the form of site-by-site and spatial estimates. This shows a gradual increase in surge variability down the east coast with a near constant maximum level in the Cromer to Sheerness region. Four features of this figure are notably different from Figure 10.5:

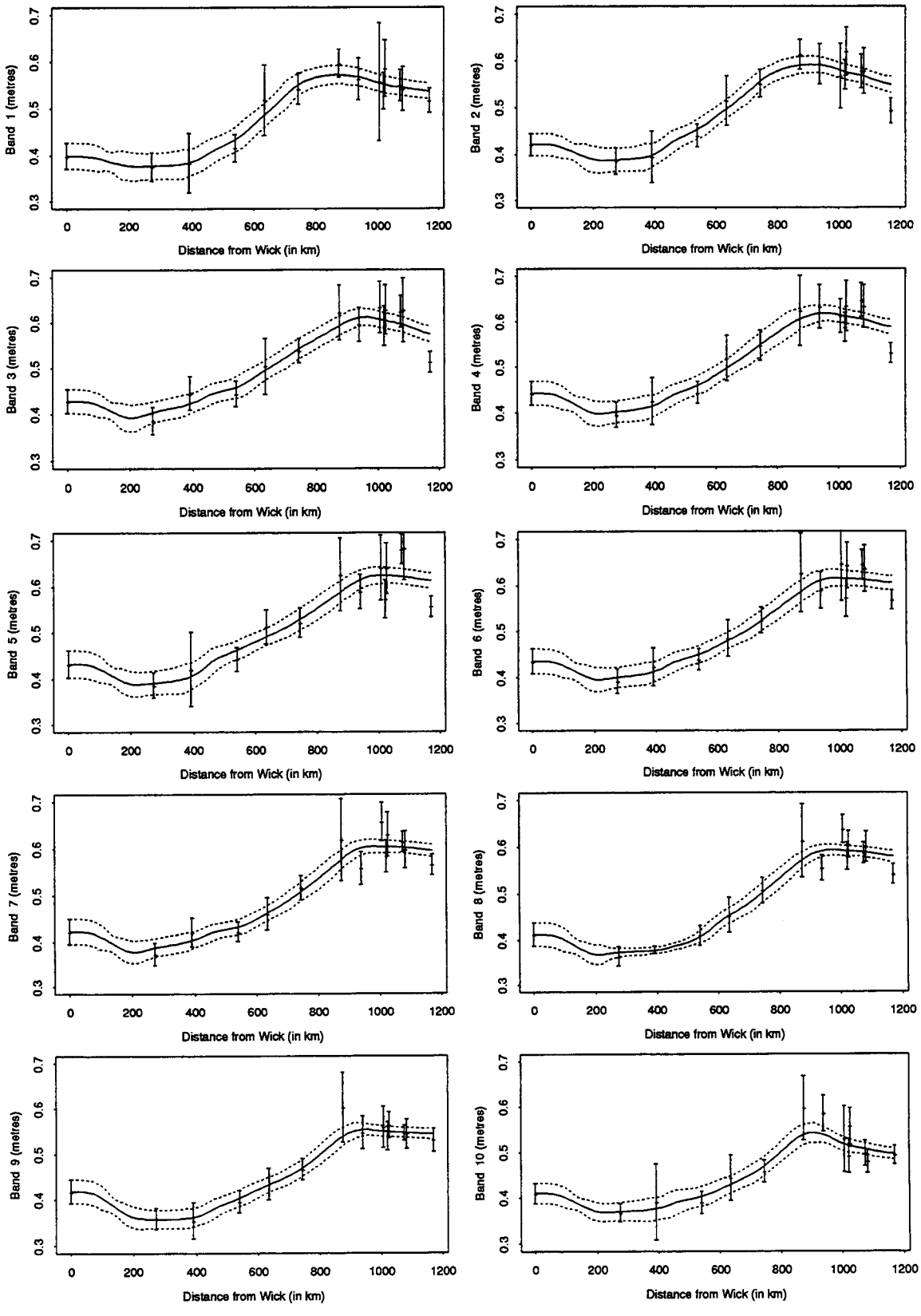

Figure 12.6: Maximum likelihood estimates of the 10 tidal dependence a-functions in the base year of 1990 and the spatial estimate of these parameters.

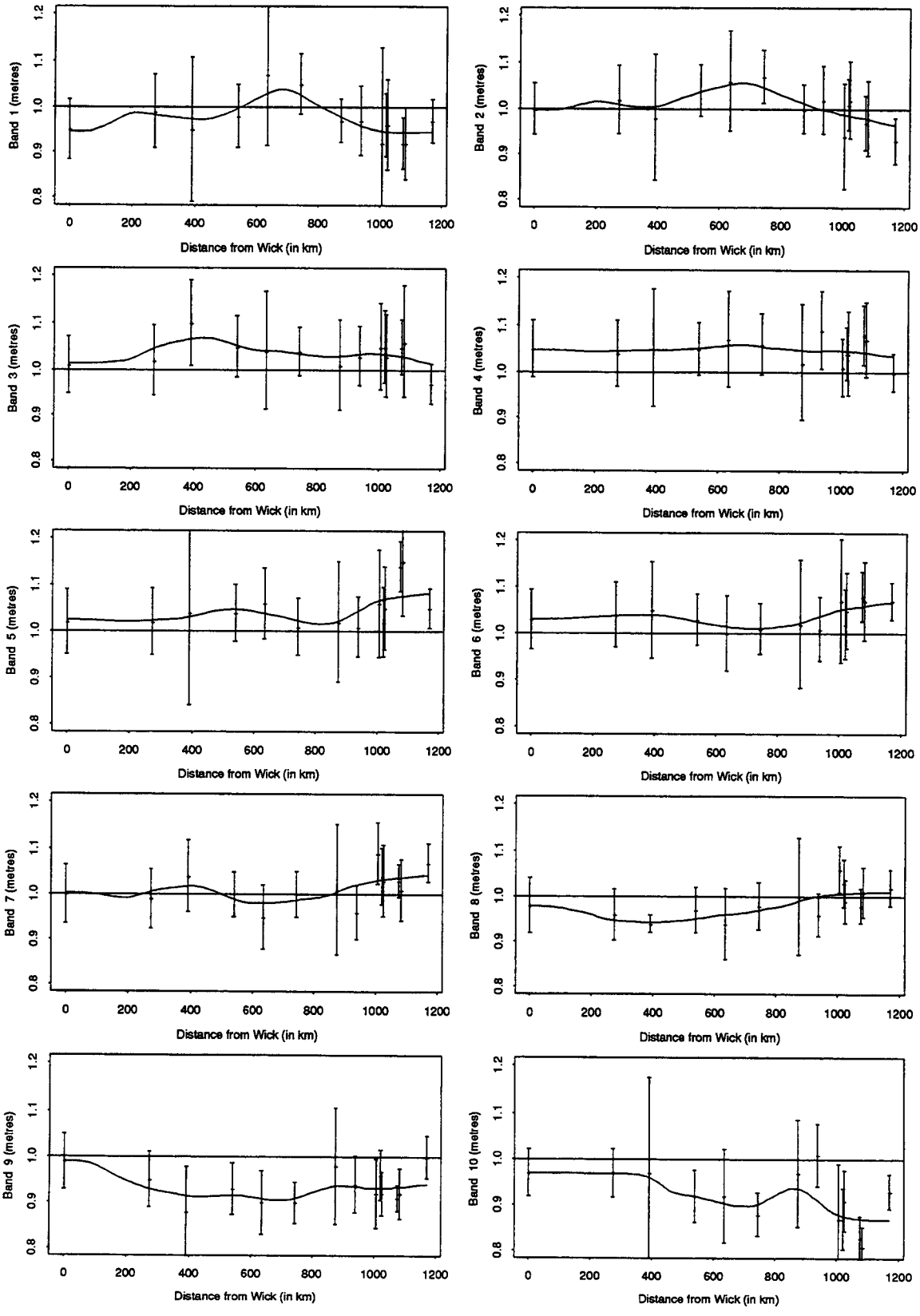

Figure 12.7: Maximum likelihood estimates of the 10 tidal dependence a-functions in the base year of 1990 normalised by the 98% quantile and the spatial estimate of this function.

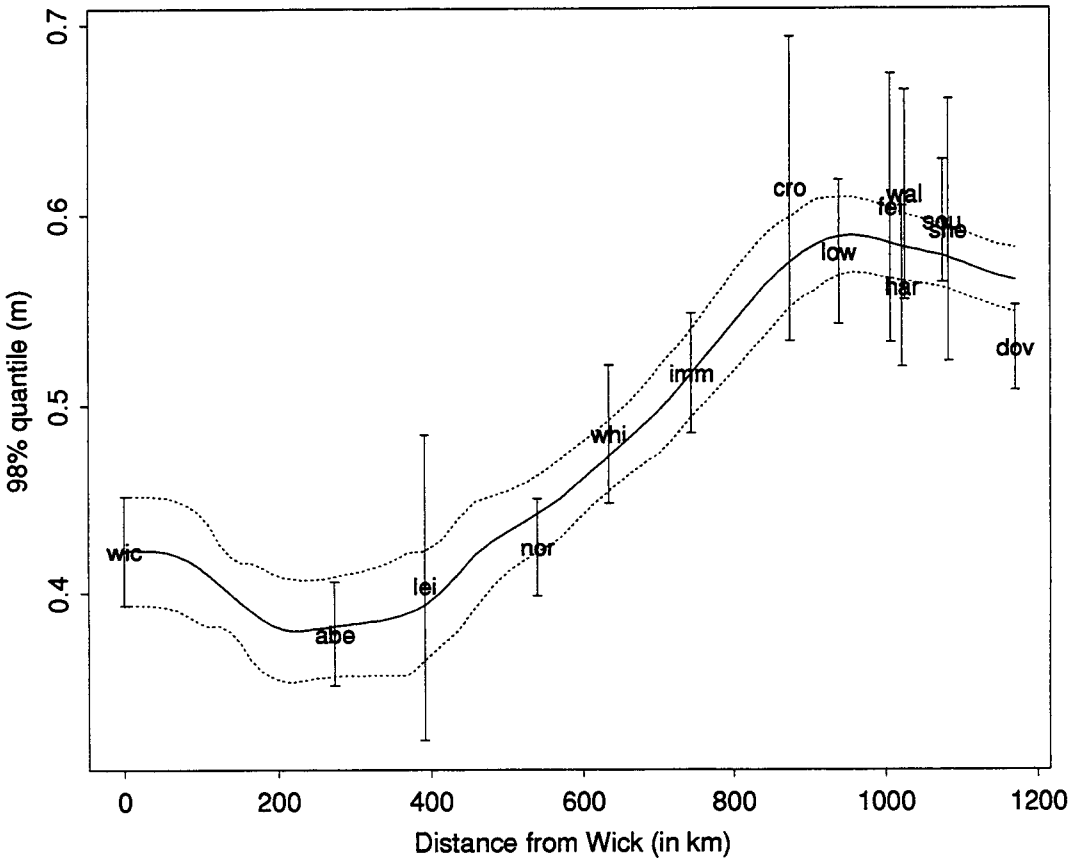

Figure 12.8: Site-by-site and spatial estimates, with associated 95% confidence intervals, of the 98% quantile of the marginal surge distribution in the base year of 1990.

- the estimate for the Wick-Aberdeen stretch is larger in the present analysis than previously. The reason for this is the high spatial trend in this region relative to that for southern Scotland. In the case study analysis trends were ignored which resulted in Wick being similar to Aberdeen. We have some reservations about the spatial trend estimate in this region, as discussed in Section 12.2, and this feature may be purely an artifact of the spatial trend.

- The site-by-site quantile estimate for Whitby in Figure 12.8 is spatially consistent with the corresponding estimates from neighbouring sites. That was not the case in Figure 10.5 where the Whitby estimate was considerably higher than neighbouring sites. The reason there was that the Whitby data comes from a different sampling period than the other sites (i.e. a more recent start date). Due to the positive trend being ignored in the case study that analysis provided too large an estimate relative to the other sites. The methods of Section 4.1 incorporate trends in the estimation of the interaction functions so naturally adjust for sampling bias in the data collection.

- The extra sites of Leith, Cromer, Felixstowe, Harwich and Walton are consistent with the spatial interpolation of the quantile given in Figure 10.5.

- The range of the coastline in the two analyses is different. The current analysis extends the case study range to include the Sheerness to Dover stretch. The quantiles in this region are lower than in the Cromer to Sheerness stretch.

The amplification, $q_s(d)$, and interaction, $a^*(X,d)$, components of the function $a(X,d)$ are of interest in their own right, providing a clear picture of the spatial variation in the characteristics of the interaction process. However, for the extreme sea-level analysis they are only of interest in combination, i.e. $q_s(d)a^*(X,d)$.

Two forms of spatial modelling of $a(X,d)$ are

1. spatially smooth the raw site-by-site estimates of $a(X,d)$, as is done in Figure 12.6;

2. separately spatially smooth the site-by-site estimates of $a^*(X,d)$ and $q_s(d)$, and then combine these estimates to give $a(X,d) = q_s(d)a^*(X,d)$, i.e. combine the estimates shown in Figures 12.7 and 12.8.

These two approaches were considered in the case study for interaction but were applied to a subset of sites and the analysis ignored trends. The resulting estimates of $a(X,d)$ from these approaches are shown in Figure 12.9 for the current data and analysis. Despite the improved handling of trends the findings are largely similar to those of Chapter 10, i.e. that there is only a little difference between the methods. The one difference between the methods concerns the estimate around Lowestoft. The second method produces a lower estimate which is below the lower endpoint of the 95% confidence interval of the site-by-site estimate of $q_s(d)$. We adopt the simpler first method for the remainder of the study.

Now consider the $b(X,d)$ function. By construction this interaction function is independent of the trend so is little changed from the results in Chapter 10. Site-by-site and spatial estimates are shown in Figure 12.10 for a range of tidal quantiles. The spatial profile of $b(X,d)$ is similar to that of $a(X,d)$ in that it is smooth with a steep increase from Leith to Cromer and approximately constant at each end of the coast. This pattern is similar for all tidal bands.

12.4 Point Process parameters

There are two aspects to the point process model: the threshold $u(X,i)$ which determines the transition from non-parametric to parametric model for the surge distribution in year i for tidal level X; and the point process parameters (μ, σ, k). The base year for the

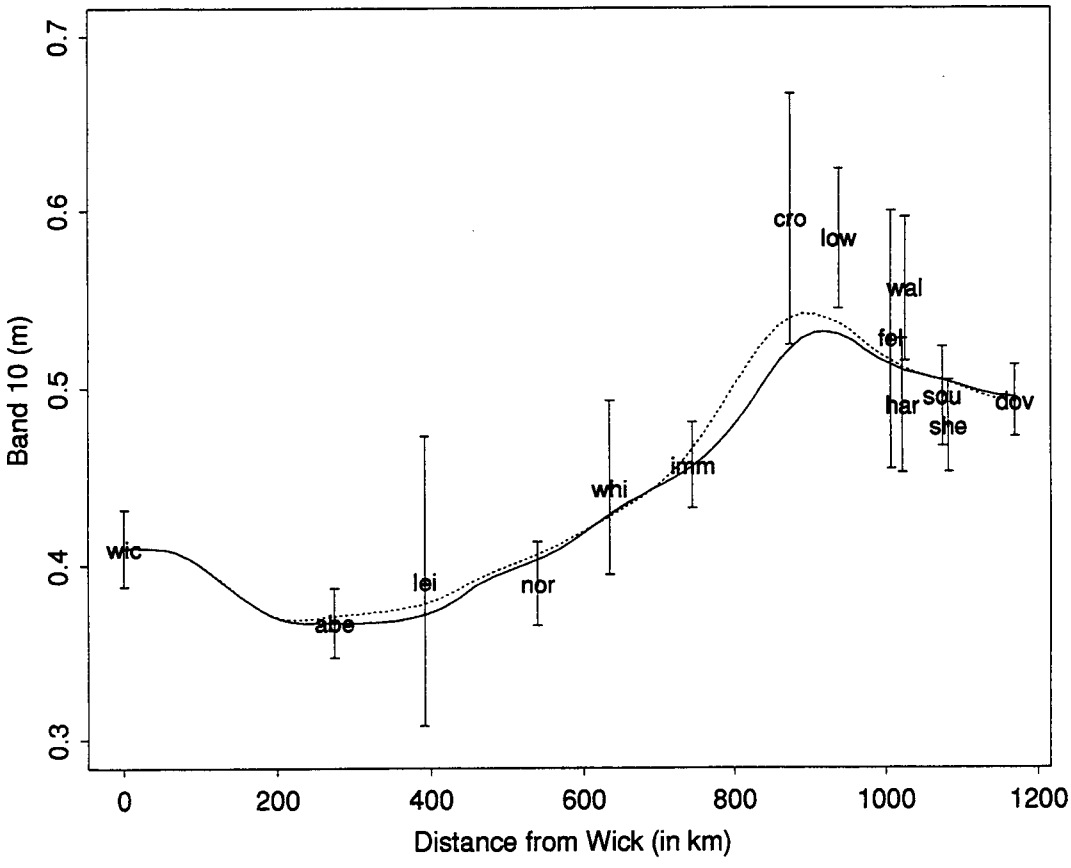

Figure 12.9: Site-by-site and two spatial estimates of $a(T_{10}, d)$. The solid curve is the estimate based on spatial smoothing of the site-by-site estimates, and the dotted curve is the estimate based on the separate smotthing of the normalised a-functions, and the 98% quantile. The estimate is for the base year of 1990.

analysis throughout Chapter 12 is 1990 so we drop the dependence of the threshold on i. In this section, we are interested in developing spatial models for $u(X, d), \mu(d), \sigma(d)$ and $k(d)$, where d is the distance from Wick.

First consider the threshold $u(X, d)$. Site-by-site estimates of the threshold are shown in Figure 12.11 for different tidal quantiles. As the threshold is taken to be the 99.5% quantile of the distribution of the surge given the tidal level is X, it is likely to be similar in form to $a(X, d)$ which is the corresponding 98% quantile of this distribution. The spatial estimate of the threshold is shown on this figure and takes the expected form.

Now consider the site-by-site and spatial estimates of $\mu(d), \sigma(d)$ and $k(d)$ shown in Figures 12.12–12.14. Recall the two reasons for development of the interaction functions:

- to produce a transformed surge process that is independent of the tidal level,
- to produce a transformed surge process which spatially was near homogeneous, or

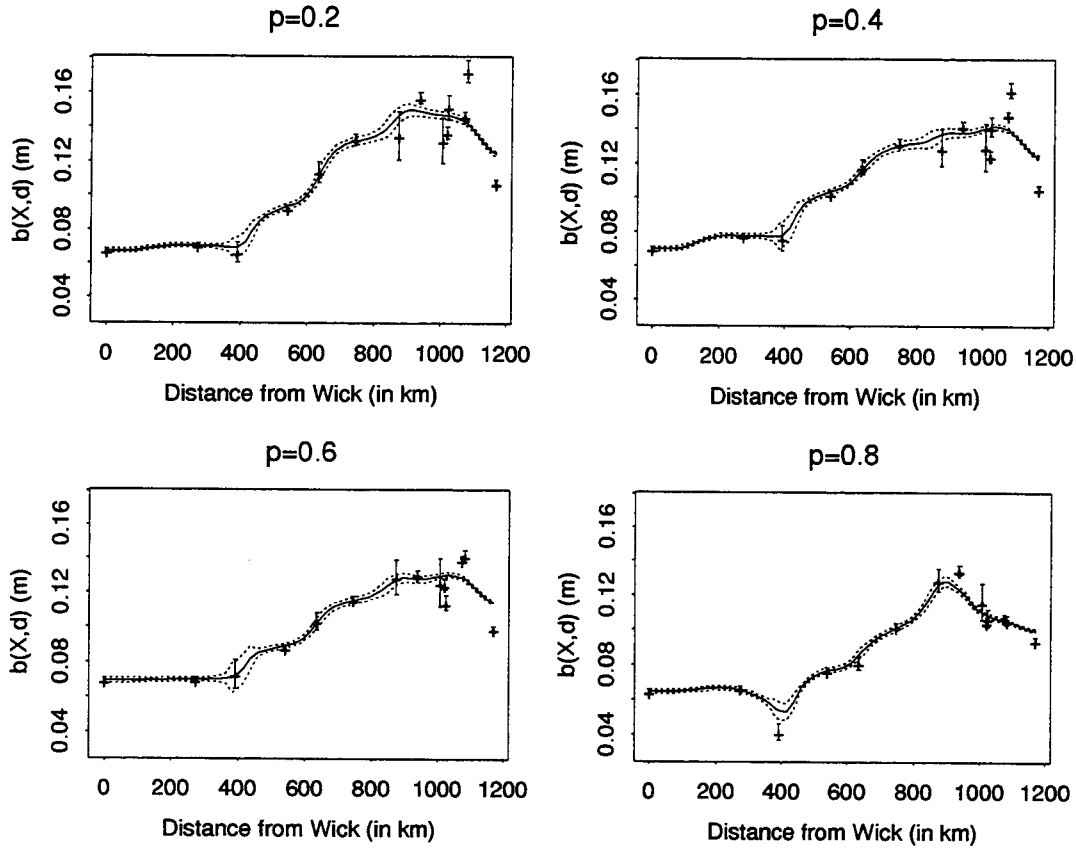

Figure 12.10: Site-by-site and spatial estimates of the tidal dependence $b(X, d)$ functions. The plots are for tidal levels, X, such that $F_{tide}(X) = p$, with $p = 0.2, 0.4, 0.6$ and 0.8.

which had parameters that changed very slowly and smoothly along a coastline.

Results in Dixon and Tawn (1994, Chapter 6) showed the the first objective was met. Here Figures 12.12–12.14 show that the second objective has also been achieved. In particular, in the site-by-site estimates of these parameters there is a gradual increase in both $\mu(d)$ and $\sigma(d)$ with distance from Wick, and $k(d)$ decreases down the coast. In addition to this general pattern, there is some finer structure in the spatial variation of these parameters between Immingham and Lowestoft. This takes the form of a local minimum in the location and shape parameters, and a local maximum in the scale parameter. The spatial estimate captures these features well, with the associated confidence interval reflecting site-by-site estimates at the data-sites when they occur in isolation, but giving considerably tighter intervals when the site has limited data (such as Whitby) and when the site is near other sites (such as on the Lowestoft-Dover stretch).

The smooth transition of these parameters along this coast gives the intended confidence in their spatial interpolation. This is particularly important as these three parameters are critical in the extrapolation to long period return levels. The spatial estimate of

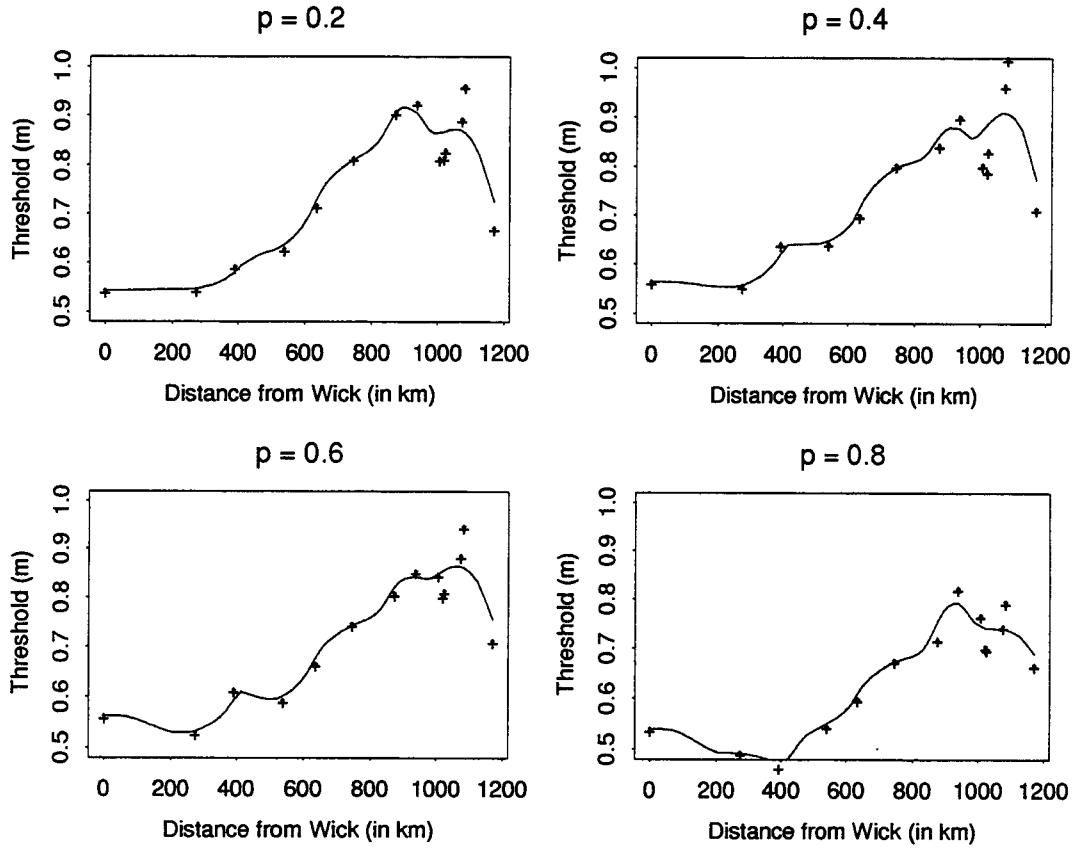

Figure 12.11: Site-by-site and spatial values of the threshold, $u(X, d)$, which determines the transition from non-parametric to parametric modelling for the surge distribution. The estimates shown correspond to the threshold for X such that $F_{tide}(X) = p$, where $p = 0.2, 0.4, 0.6$ and 0.8. The spatial estimate is a weighted regression with weights inversely proportional to the number of years of data at the sites.

the shape parameter is particularly interesting starting as a positive value, 0.15, at Wick, showing evidence for an upper endpoint to the distribution; decreasing to zero just north of Immingham from where it remains basically at that level, which corresponds to the transformed surge variable having an exponential tail with no upper endpoint.

One possible explanation for the local structure in the $\mu(d)$ and $\sigma(d)$ plots between Immingham and Lowestoft is the use of the interaction function $b(X)$. Here the site-by-site estimate of this function are used and that differs from the spatial estimate of Figure 12.10 in this region for high tidal levels. Since for year i at tidal level X the surge location and scale parameters are

$$\begin{aligned} \mu(X, i) &= \mu b(X) + a(X, i_0) + \beta(i - i_0) \\ \sigma(X) &= \sigma b(X), \end{aligned}$$

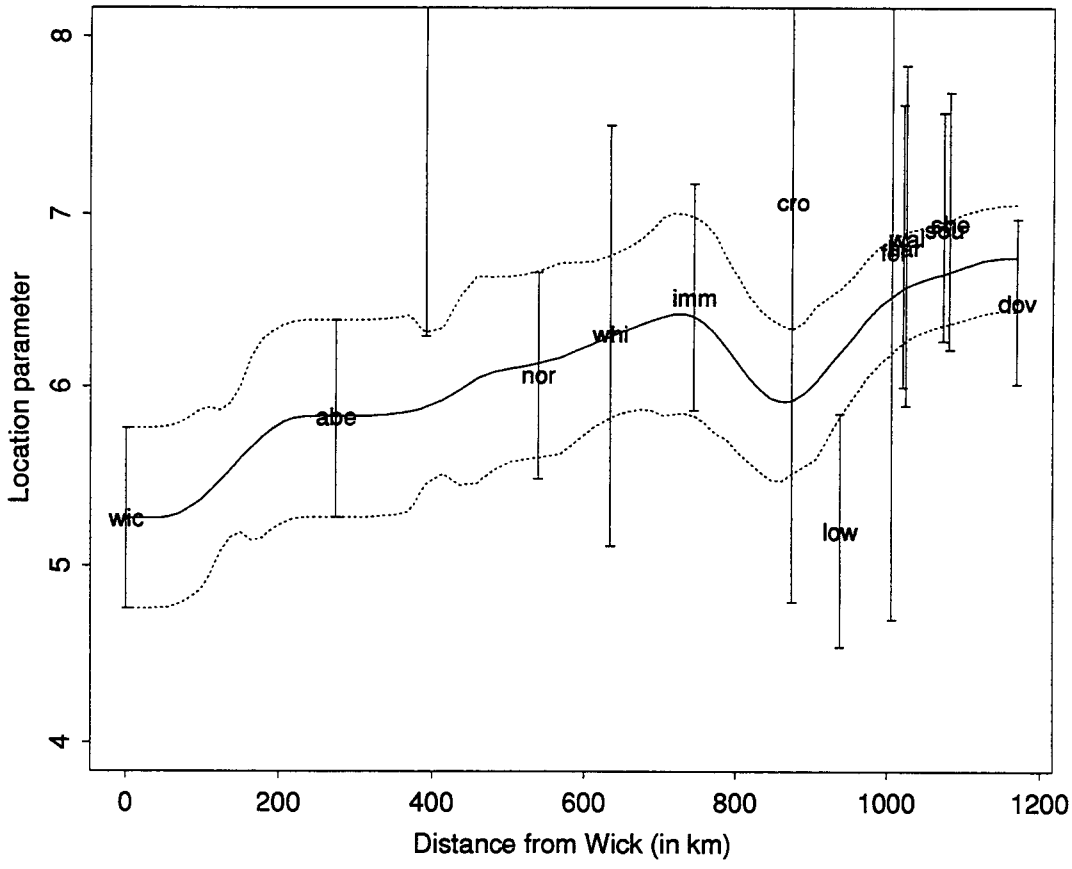

Figure 12.12: The point process location parameter for the surge, μ, against distance; the site-by-site estimates and the spatial estimate with approximate 95% confidence intervals.

errors in the estimation of $b(X)$ influences both μ and σ when viewed spatially. This suggests that by using the spatially smoothed estimate of $b(X)$ the corresponding estimate of $\mu(d)$ would be smoothed. We have not used the spatial estimate of $b(X)$ as the resulting spatial model for $\sigma(d)$ is less smooth than the estimate shown in Figure 12.13.

The spatial estimate is based on the same weighted kernel regression estimator of the site-by-site estimates as used for the interaction functions, except here the site-by-site estimates for Leith, Cromer and Felixstowe play no role in the fitting procedure as the estimates for these sites are based on the 4 years or less of hourly data (Leith also has many years annual maximum). They carry so little information about these extremal parameters that their inclusion should hardly change the spatial estimates, however as they may introduce bias due to their spurious values they were omitted from the weighted kernel regression. The site-by-site estimates are shown on the figures for these 3 sites. With the exception of the location parameter for Leith, the confidence intervals for site-by-site and spatial estimates overlap well and the spatial estimate appears to give a much improved estimate at these sites. The site-by-site location parameter estimate for Leith

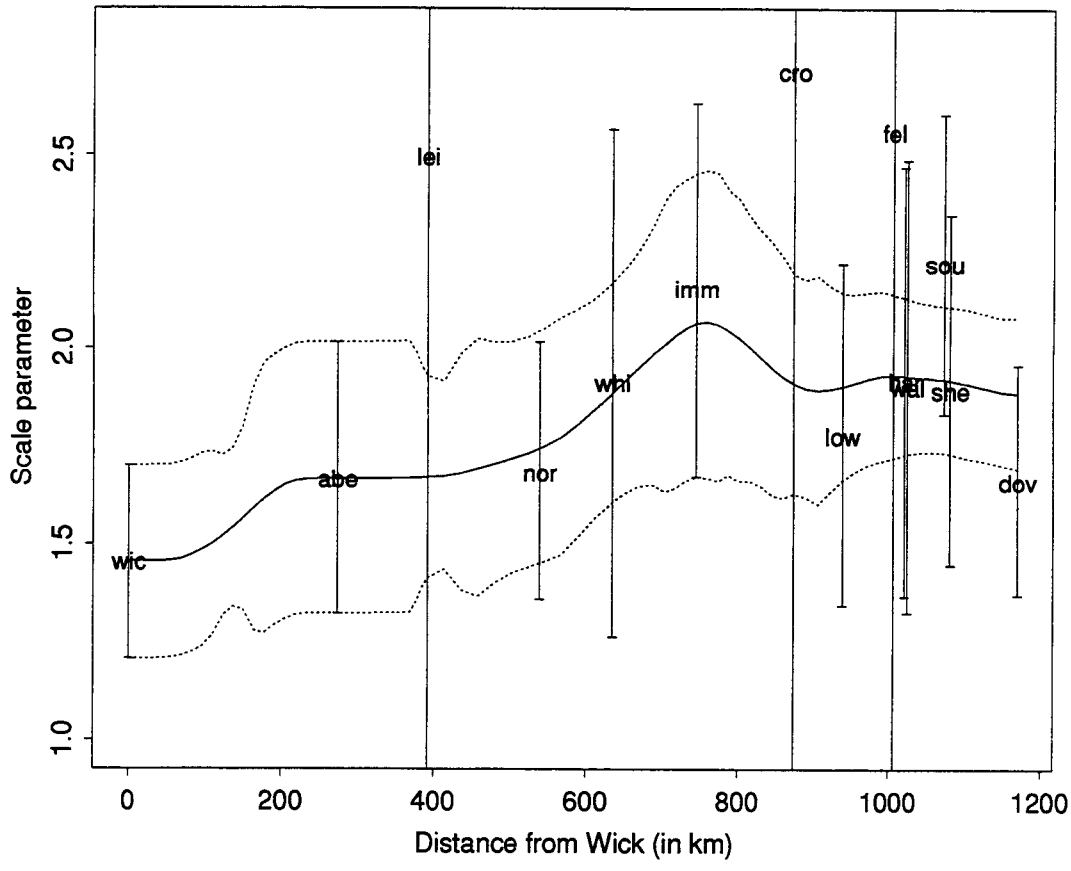

Figure 12.13: The point process scale parameter for the surge, σ against distance; the site-by-site estimates and the spatial estimate with approximate 95% confidence intervals.

is not well explained by the spatial estimate but its value appears spurious since if this were a true feature we would have seen a similar property in the surge quantile estimate of Figure 12.8. We adopt the spatial estimate here.

So far we have noted the spatial variation in the extremal parameters without discussing the impact on the tail of the surge distribution. The near linear relationship with distance down the coast for each of the three parameters corresponds to the transformed surge variable

$$\frac{Y(d) - a(X,d)}{b(X,d)},$$

having quantiles which increase down the east coast, i.e. with d. The additional local feature previously observed for the Immingham-Lowestoft region is slightly more complex. There the location parameter deceases, suggesting a shorter tail for the transformed surge variable, whereas the scale and shape parameter changes both correspond to a longer tail in that region. In combination these parameters determine the overall length of the tail of the transformed surge. However, most influential are the scale and shape parameters

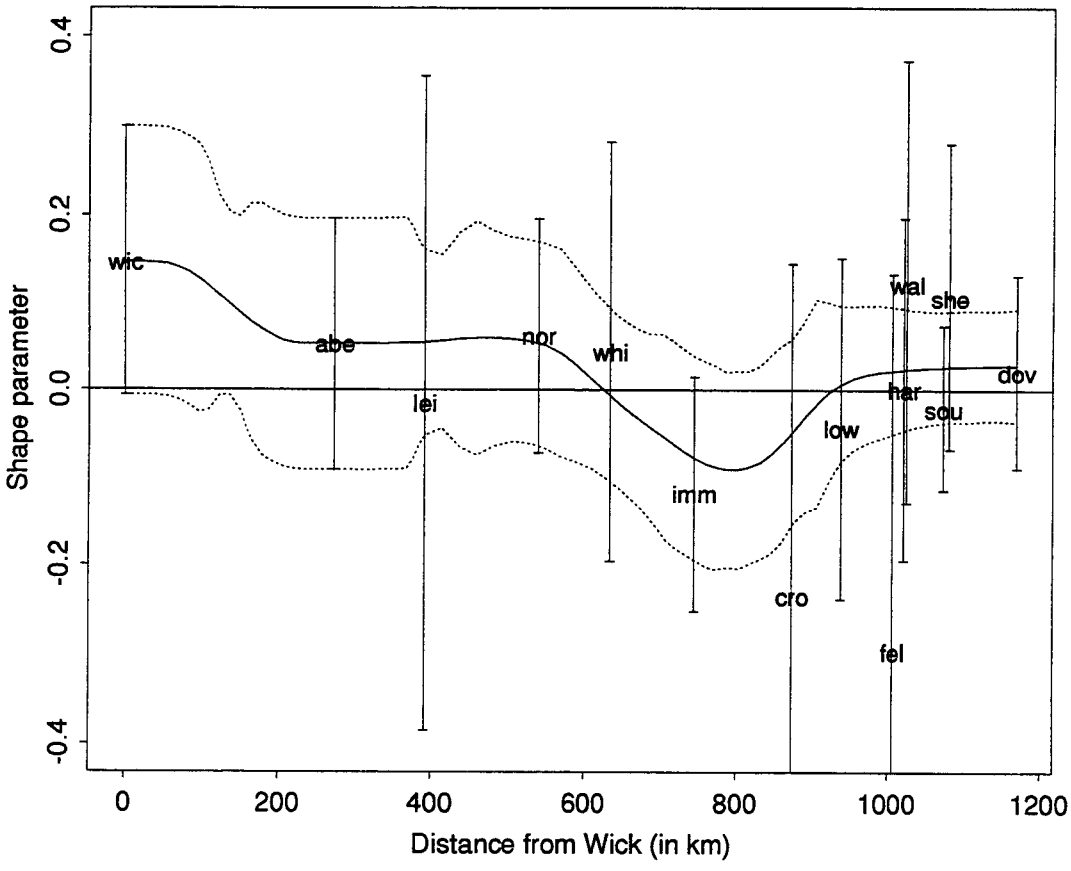

Figure 12.14: The point process shape parameter for the surge, k, against distance; the site-by-site estimates and the spatial estimate with approximate 95% confidence intervals. The line $k = 0$ is shown as a guide.

which lead to the transformed surge having its longest tail in the Immingham to Lowestoft region.

12.5 Extremal indices

Dixon and Tawn (1994, Section 6.2) developed an estimation procedure for the ratio of extremal indices, θ_l/θ_s using information from the site of interest. Site-by-site estimates of this ratio are shown on Figure 12.15 together with a smooth spatial estimate which suggests that this ratio is nearly constant, approximately 0.5, for this entire coastline. Figures 12.16 and 12.17 show the corresponding plots for the individual extremal index parameters, θ_s and θ_l. Since these parameters in ratio are nearly constant it is not surprising that these estimates are similar, showing a gradual increase down the coast, in a pattern similar to the tidal characteristics at these sites. Dixon and Tawn (1994) showed that the extremal indices have a minimal role, relative to the extremal parameters

considered in Section 12.4, in determining return levels so we have not considered it necessary to attempt to explicitly model this observed variation with tides to go beyond the crude spatial estimates shown in Figures 12.16 and 12.17.

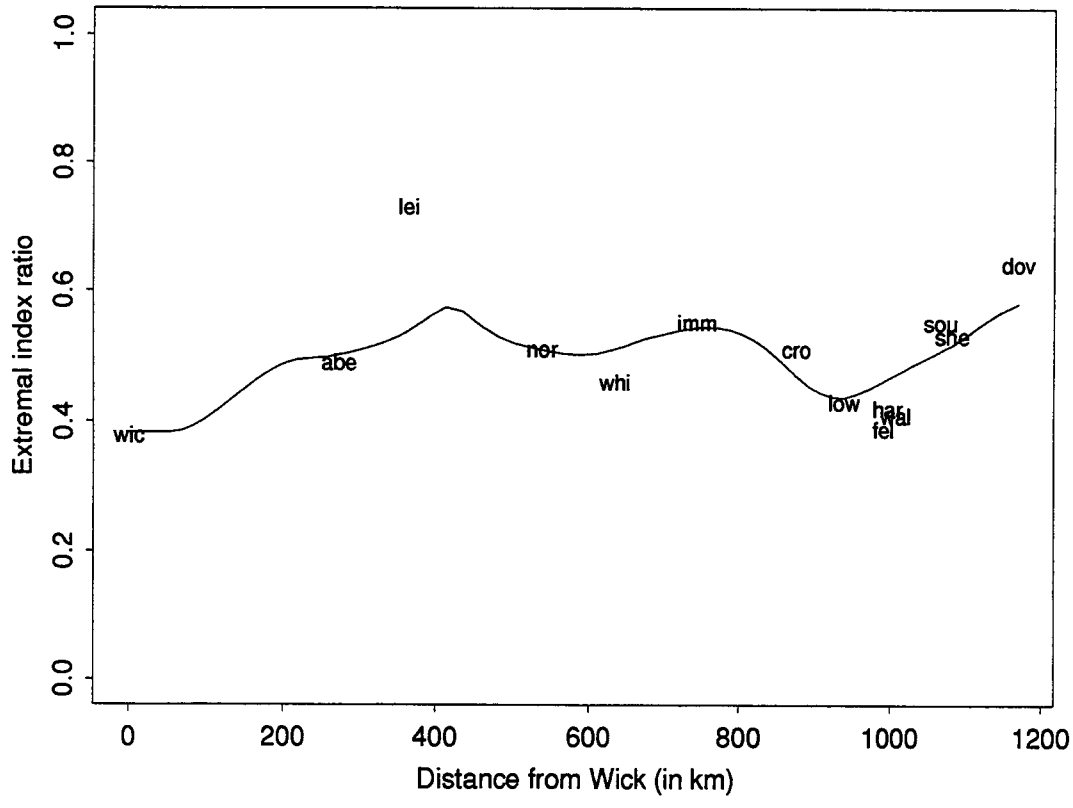

Figure 12.15: Site-by-site and spatial estimates of the extremal index ratio, θ_l/θ_s. The spatial estimate is a weighted regression estimate with weights inversely proportional to the number of years of data at the sites.

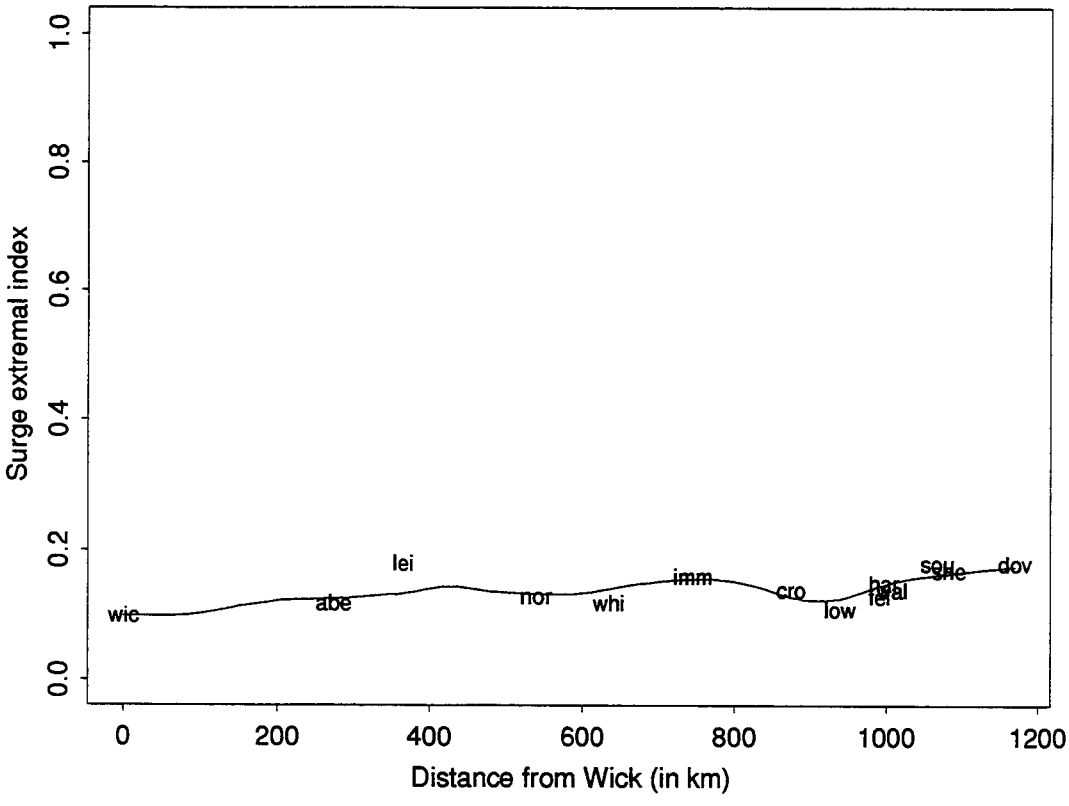

Figure 12.16: Site-by-site and spatial estimates of the surge extremal index, θ_s. The spatial estimate is a weighted regression estimate with weights inversely proportional to the number of years of data at the sites.

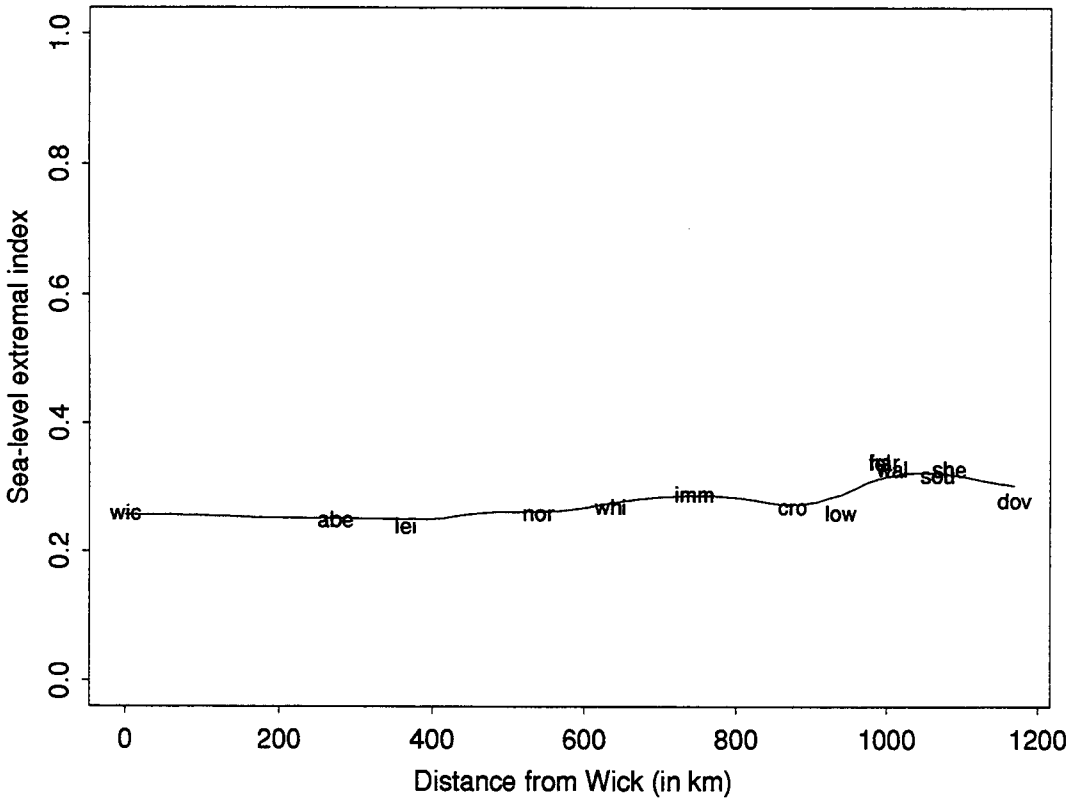

Figure 12.17: Site-by-site and spatial estimates of the sea-level extremal index, θ_l. The spatial estimate is a weighted regression estimate with weights inversely proportional to the number of years of data at the sites.

Chapter 13

Application to east coast – return levels

In Chapter 12 spatial models have been derived for all the parameters of the SRJPM. Each of these parameters has a clear spatial interpretation as they are independent of the datum used to record the sea-level data. In this section we combine these spatial parameters with the tidal series to give design levels using the methods discussed in Section 4.6 that were applied in Part II for the site-by-site return level estimates.

The one complication we have in mapping the resulting return levels spatially concerns the datum the tidal levels are taken relative to. The sea-level data are typically measured relative to Admiralty Chart Datum (ACD) or Ordnance Datum Newlyn (ODN), so the tidal analysis produces predictions relative to the datum of the observations. The datum dependent parameter in the tidal analysis is the Z_0 term in equation (2.2.2). The Proudman Oceanographic Laboratory's spatial tidal interpolation method, described in Section 12.1, is relative to mean sea-level, i.e. is the tidal prediction minus the Z_0 term. Therefore when the tides from the spatial interpolation are combined with the spatial parameters from the SRJPM the return level estimates are relative to mean sea-level at each site. Relative to this datum level return levels should vary smoothly spatially. The spatial estimate of the return level curves for return periods of $10, 100, 1000$ and 10000 years are shown in the four figures, Figure 13.1-13.4 for the entire east coast. In Section 13.2 we discuss the conversion of these estimates to the datum of interest.

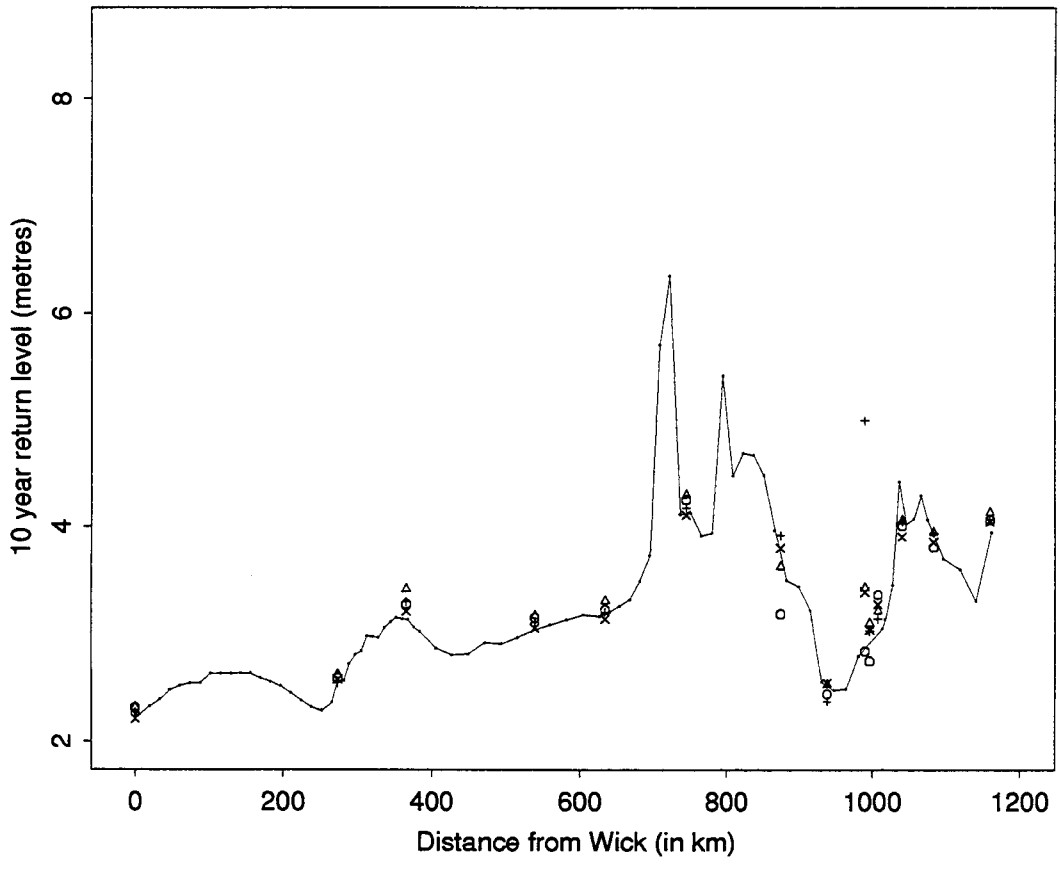

Figure 13.1: Spatial estimate of the 10 year return levels. The symbols $\triangle$, $+$, $\bigcirc$, and $\times$ represent estimates obtained from the JPM, the RJPM, the r-largest and the SRJPM given in Part II. The estimates are to the datum of mean sea-level at each site.

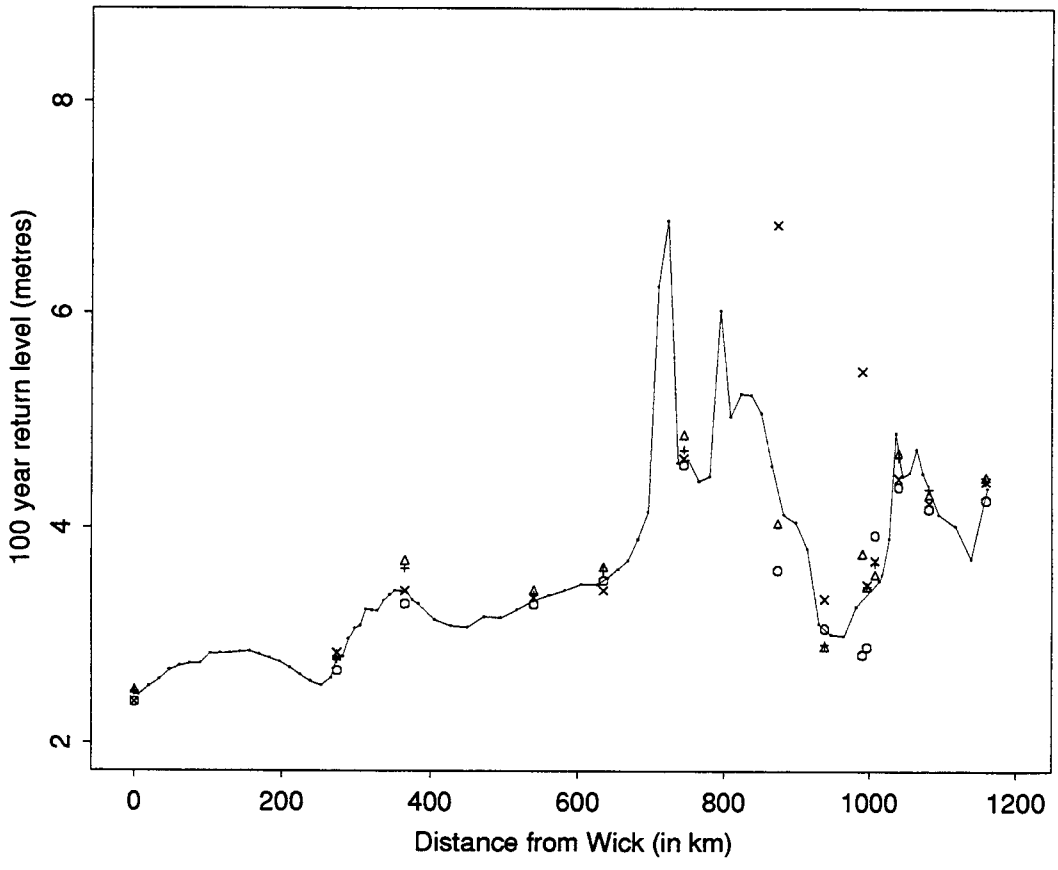

Figure 13.2: Spatial estimate of the 100 year return levels. The symbols $\triangle$, $+$, $\bigcirc$, and $\times$ represent estimates obtained from the JPM, the RJPM, the r-largest and the SRJPM given in Part II. The estimates are to the datum of mean sea-level at each site.

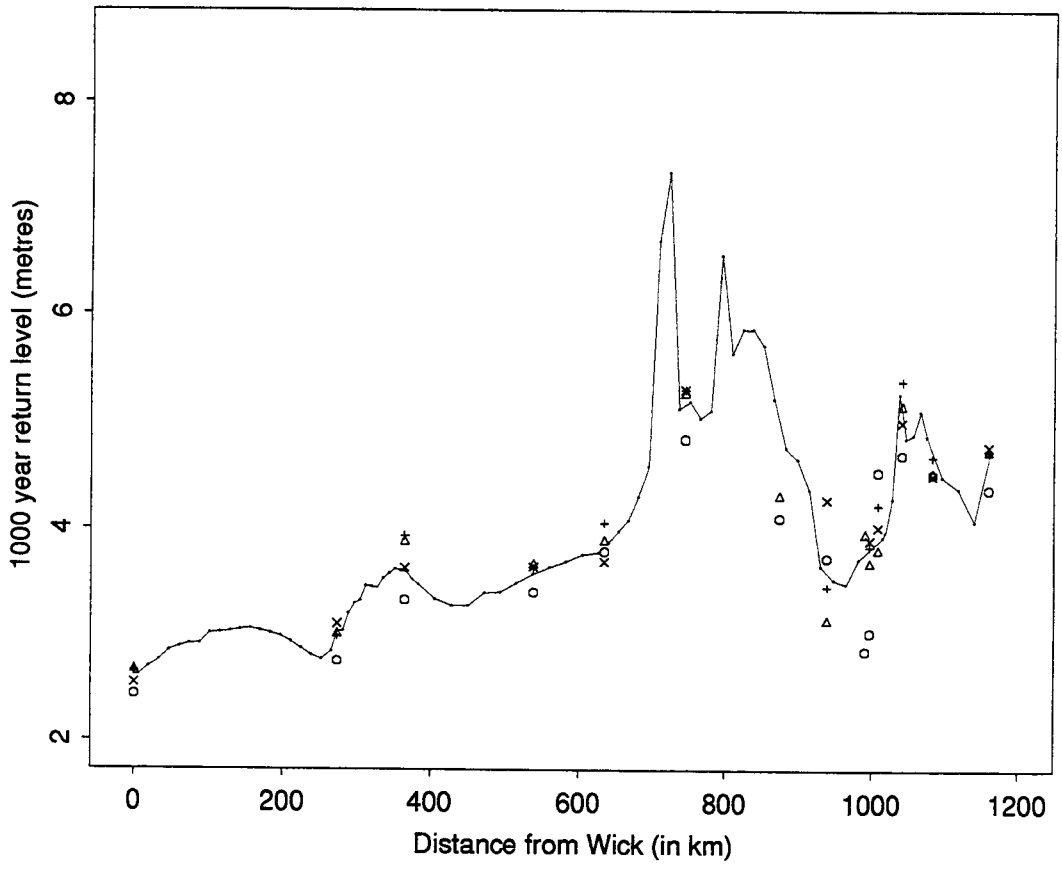

Figure 13.3: Spatial estimate of the 1000 year return levels. The symbols $\triangle$, $+$, $\bigcirc$, and $\times$ represent estimates obtained from the JPM, the RJPM, the r-largest and the SRJPM given in Part II. The estimates are to the datum of mean sea-level at each site.

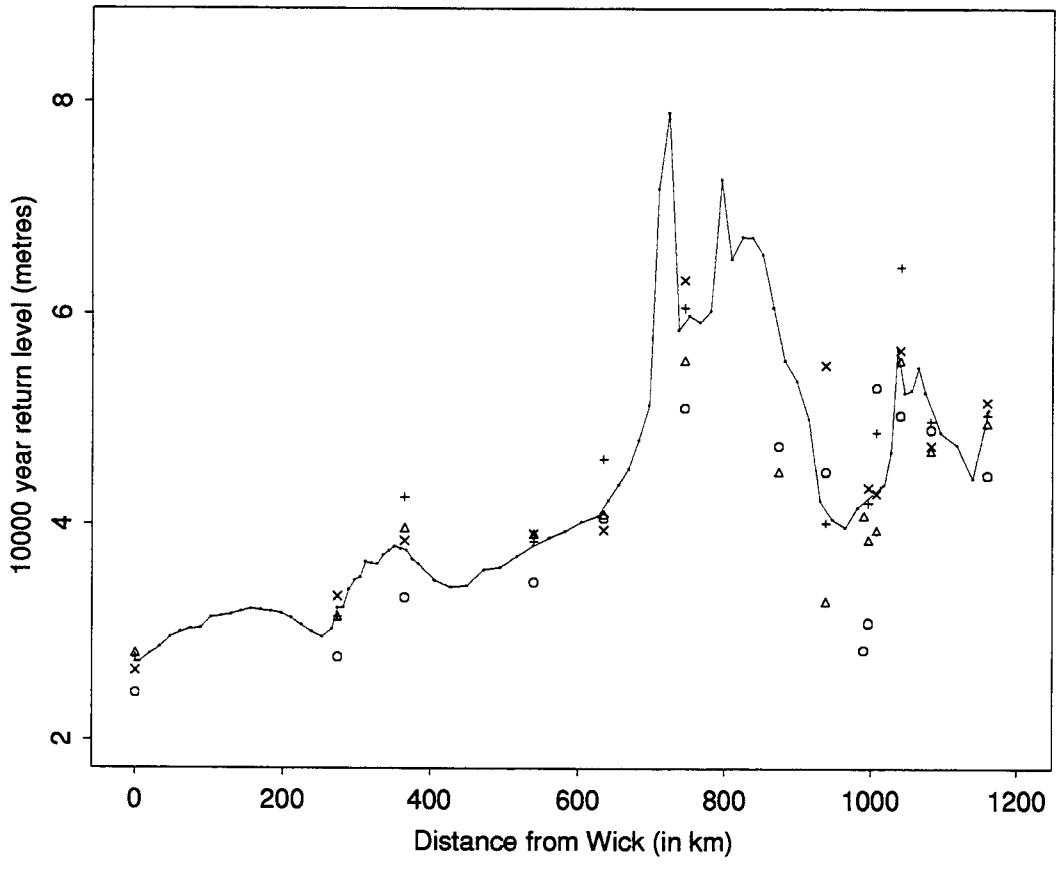

Figure 13.4: Spatial estimate of the 10000 year return levels. The symbols △, + , ○, and × represent estimates obtained from the JPM, the RJPM, the r-largest and the SRJPM given in Part II. The estimates are to the datum of mean sea-level at each site.

These figures also show the corresponding estimates based on the four site-by-site methods (JPM, RJPM, r-largest, and the SRJPM). These are the site-by-site estimates given in Part II but presented here relative to the mean sea-level datum rather than the ACD used in Part II.

The spatial estimates reflect the site-by-site estimates well at the data sites and vary between sites in a similar way to the spatial tidal quantile profiles of Figure 12.2. A number of points of interest are apparent from these four figures:

1. The spatial estimate always lies between two of the site-by-site estimates for each study site with the exception of Leith and Dover at the 10 year return period and Walton at the 10 and 100 year return periods.

2. The spatial interpolation of the spatial return level estimate is markedly different from that given by the spatial smoothing the site-by-site return level estimates shown in Figures 6.1-6.3.

3. Although the spatial estimate is based on the SRJPM it does not pass though the site-by-site estimates based on that method as the spatial estimates of the parameters such as the trend, interaction functions and point process parameters differ from the site-by-site estimates. However, the 10 year return level the spatial estimate is very close to the site-by-site estimates for sites with long records and differs by up to 40cm for sites with less data such as Felixstowe. This is increasingly noticeable for the longer return periods, for which the Cromer and Felixstowe SRJPM estimates are off the top of the plot for both the 1000 and 10000 year return periods.

4. Dixon and Tawn (1994) suggested that the r-largest method under-estimates long period return levels for sites when the inter-annual variation in the tides is large relative to the variation in the surge. This important finding is immediately obvious from the figures, particularly Figure 13.4. If the r-largest method had been applied serious under-design would occur for all sites north of Cromer, and for some other sites such as Harwich and Dover. The notable exception is Lowestoft, for which the r-largest estimate is well above the spatial estimate.

5. Of the site-by-site methods Dixon and Tawn (1994) studied, the RJPM was considered the best provided at least 5-10 years of hourly data were available. Relative to the better spatial method developed here we see that the advice is reliable. When the data record is quite long for return periods of 1000 and 10000 years the RJPM estimate is improved by use of the spatial estimate at some sites.

6. Of the site-by-site methods Dixon and Tawn (1994) studied, the JPM was considered the best if less than 5-10 years of hourly data were available. Relative to the better

spatial method developed here we see that the advice is again reliable. For example at Cromer and Felixstowe the best site-by-site method for the 10, 100 and 1000 year return periods is the JPM.

7. The JPM shows the slight over-estimation predicted in Dixon and Tawn (1994) for most sites. The JPM also has the weakness that when the tidal range is small relative to the surge variation return levels for long return periods are under-estimated. The main example of this given in Dixon and Tawn (1994) study was Lowestoft, but examining the 10000 year return level estimates relative to the spatial estimate shows that the JPM under-estimates for sites from Immingham to Walton.

8. Figure 13.5 shows the changes that occur in the spatial return level estimate from the 10 year level to the more extreme return periods. The spatial smoothness of

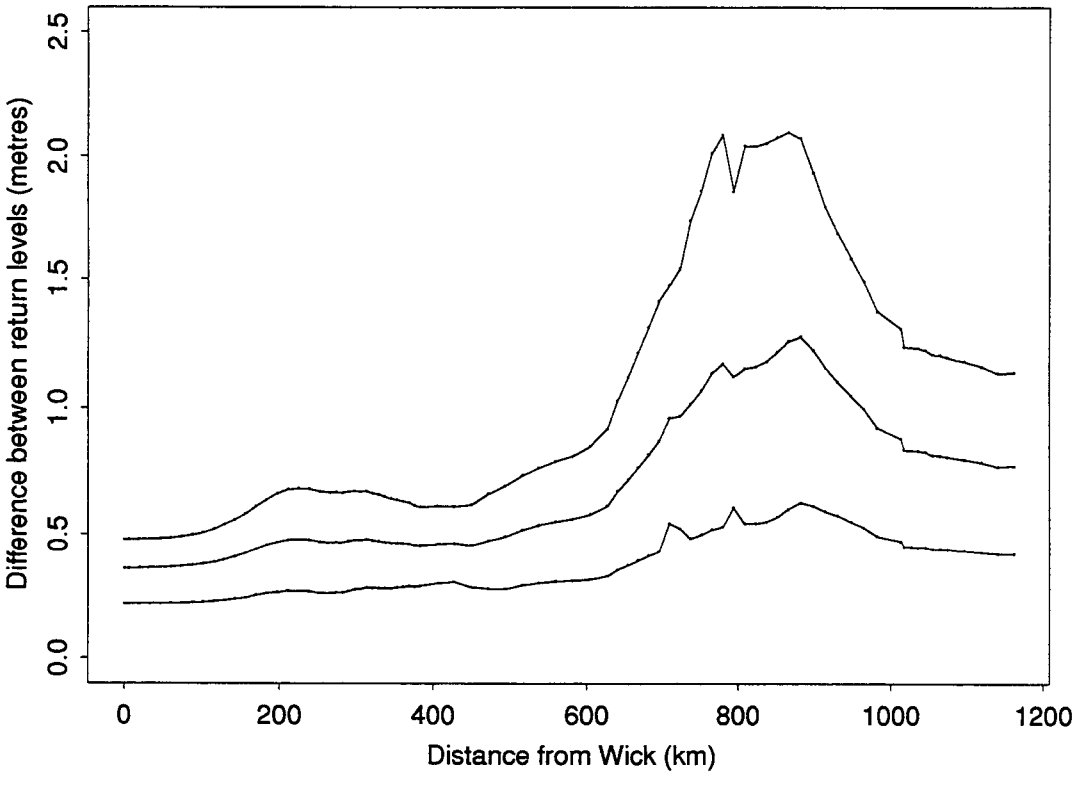

Figure 13.5: Differences in the spatial return level estimate from one return period to another. The three curves, top to bottom, correspond to the 10000 year-10 year estimate, the 1000-10 year estimate and the 100-10 year estimate.

this figure shows that this quantity is largely independent of the tide and is most influenced by the point process parameters of the extreme surge tail. We see the

combined impact of the extremal parameters $\mu(d), \sigma(d)$ and $k(d)$ which lead to the extreme sea-level distribution having its longest tail in the Immingham to Sheerness region, with the maximum difference occuring in the Immingham to Cromer region of the coast.

Figures 13.1-13.4 are ideal for assessing the basic properties of the spatial estimate relative to the site-by-site estimates. However those figures are of little use for examining the spatial estimate for any site or coastal stretch due to the large spatial variation in levels over the whole east coast, and the local variation in the shallow water estuary regions. Therefore in Figures 13.7-13.14 the spatial estimate is shown for sites on a 20km grid for short coastal stretches of five grid sites at a time. The boundaries of these stretches are shown in Figure 13.6, as are the five regions of complex, non-linear, tidal changes along the coast where the spatial estimate varies most rapidly. Each of the 83 grid locations is identified throughout with a reference number $(1 - 83)$ and latitude and longitude point, see Tables 13.7 - 13.9.

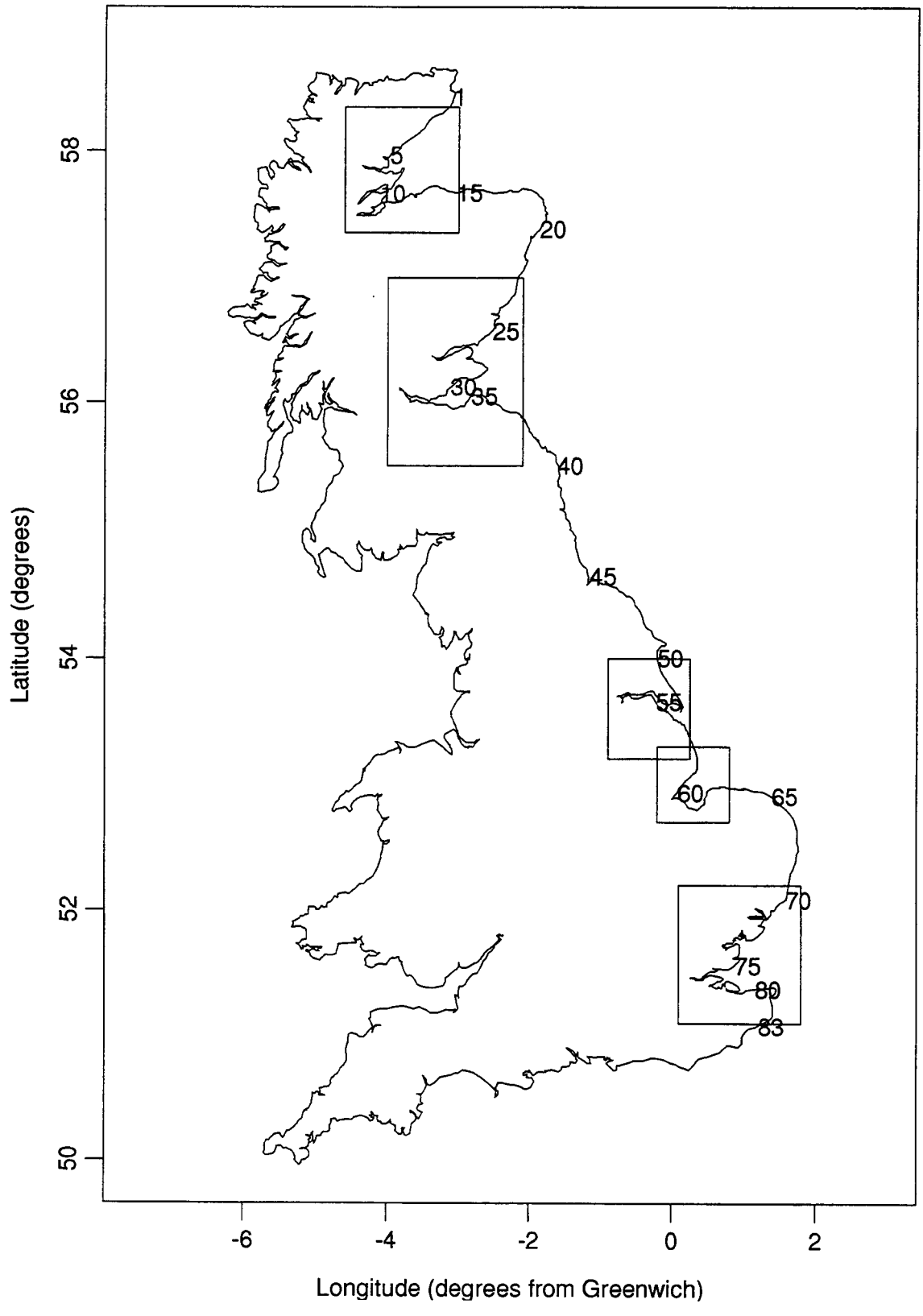

Figure 13.6: UK map showing the five regions to be enlarged.

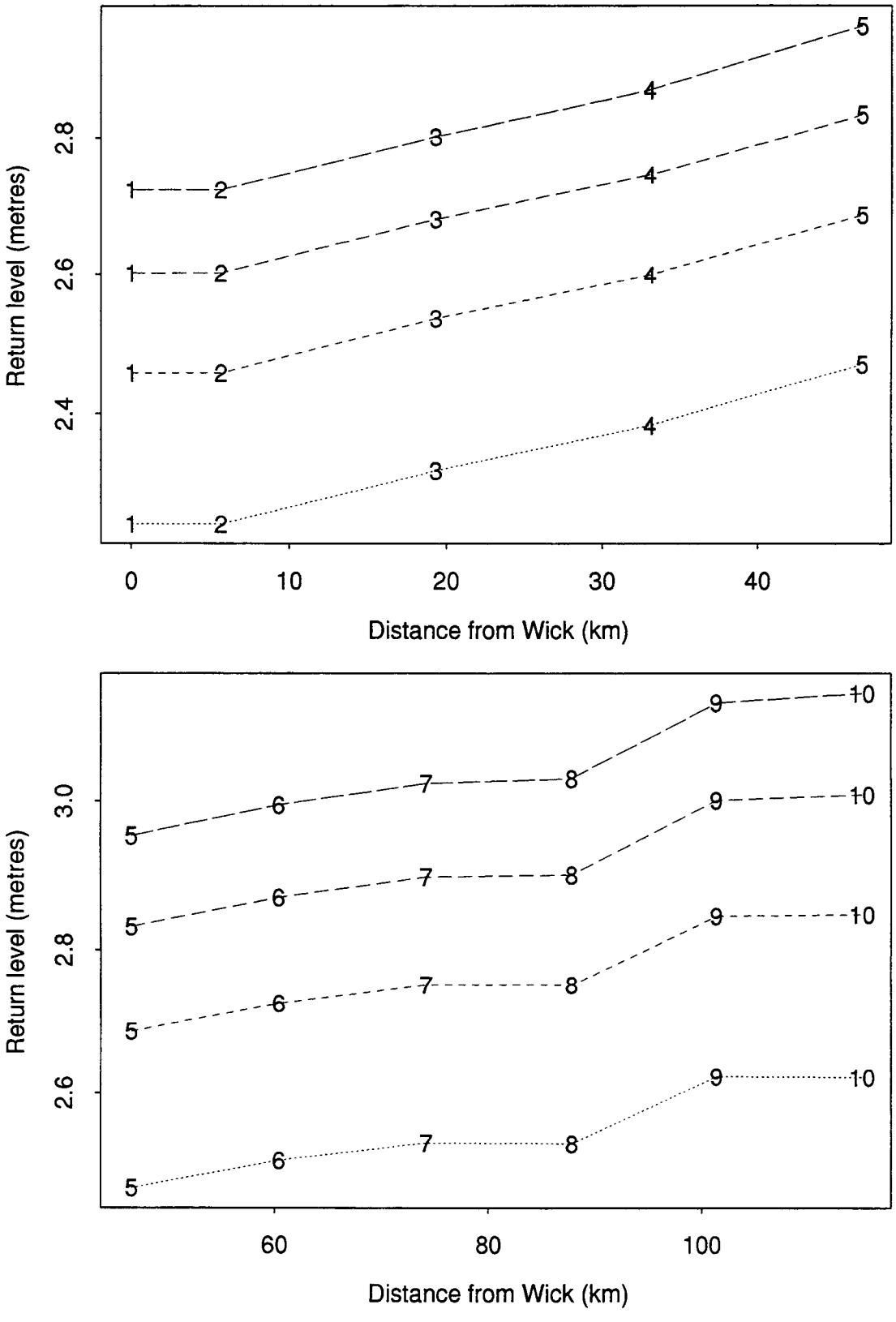

Figure 13.7: Spatial return level estimates in metres, relative to mean sea-level in 1990, for the east coast grid: site numbers 0 to 5 and 5 to 10. The four curves (bottom to top) are the 10, 100, 1000, and the 10000 year return levels.

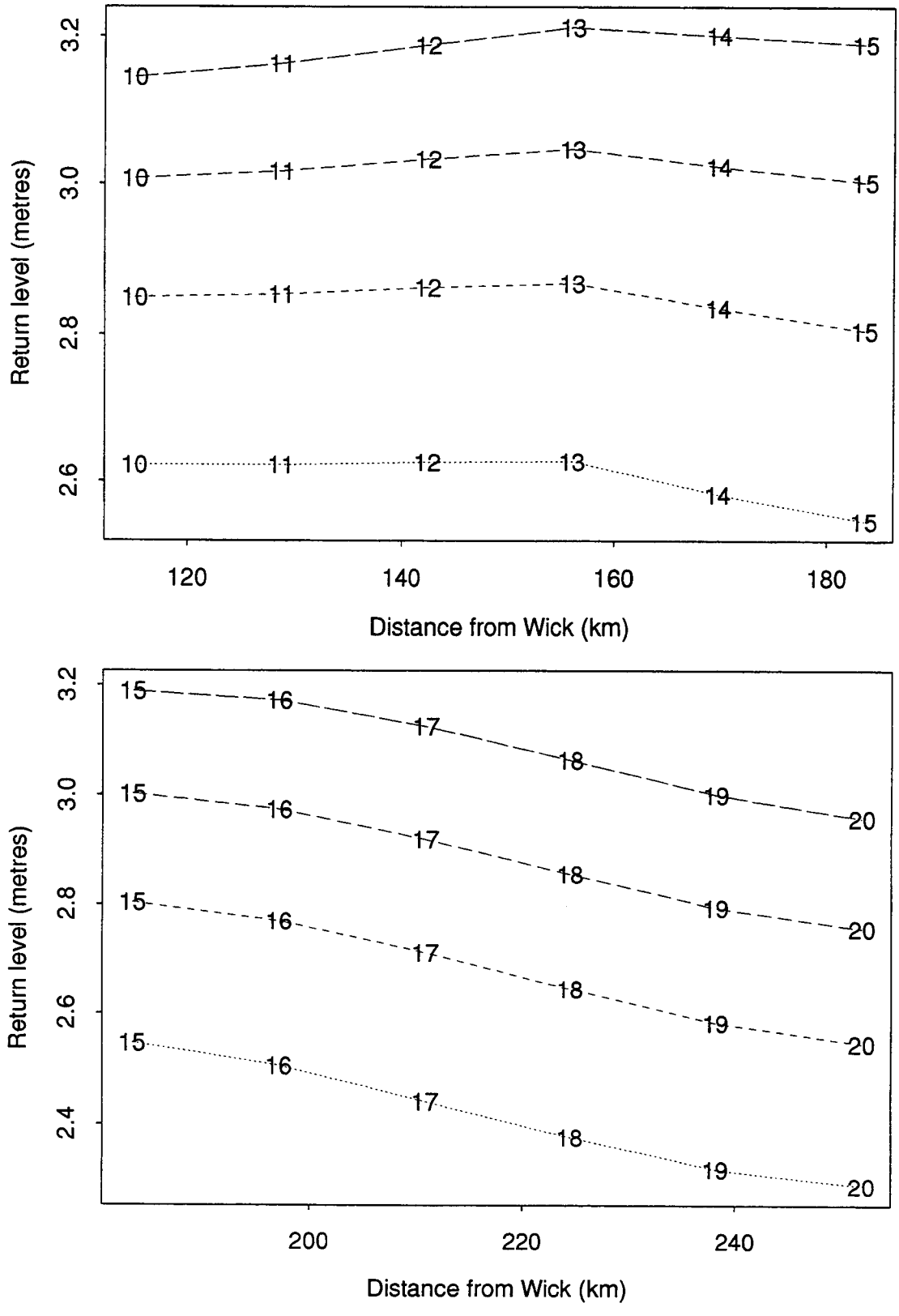

Figure 13.8: Spatial return level estimates in metres, relative to mean sea-level in 1990, for the east coast grid: site numbers 10 to 15 and 15 to 20. The four curves (bottom to top) are the 10, 100, 1000, and the 10000 year return levels.

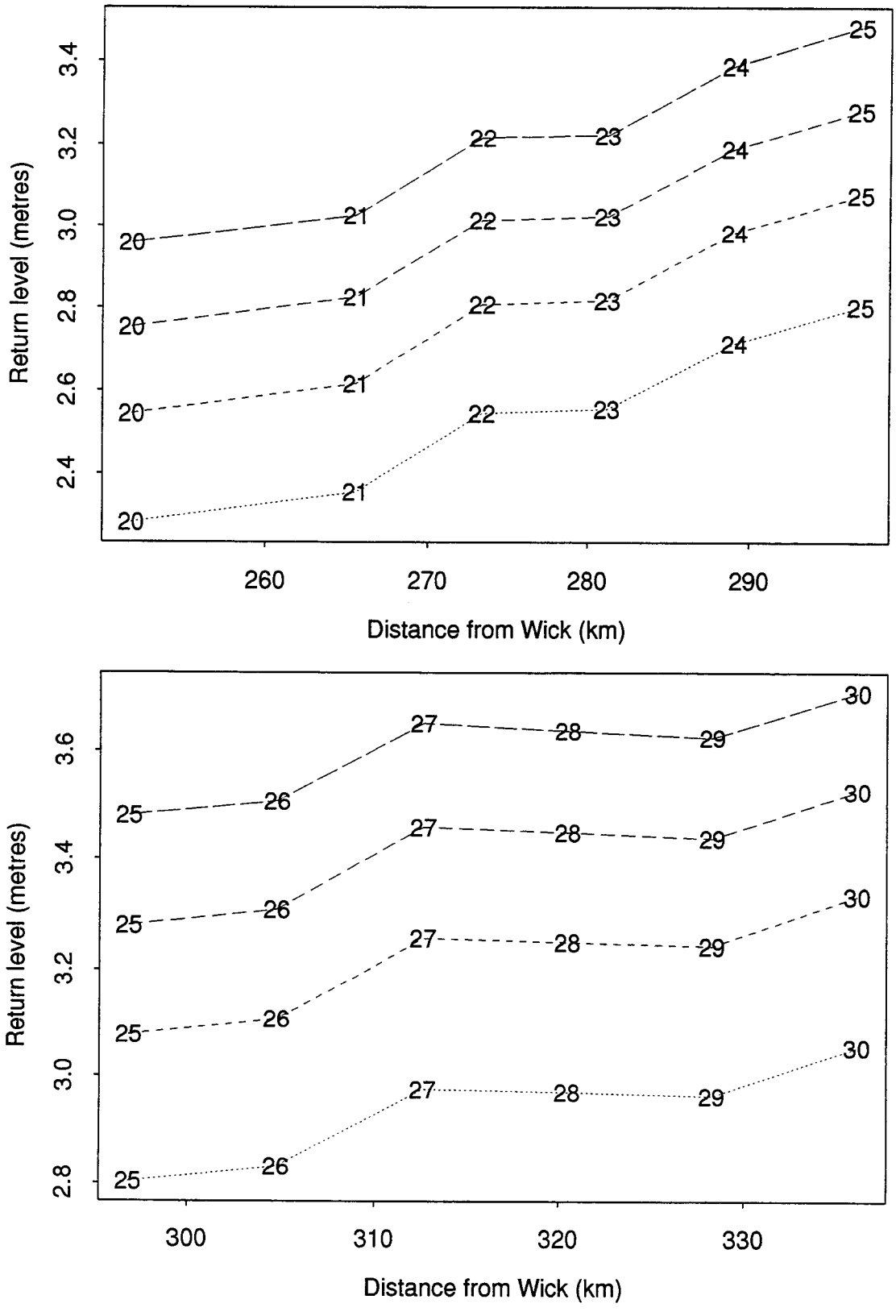

Figure 13.9: Spatial return level estimates in metres, relative to mean sea-level in 1990, for the east coast grid: site numbers 20 to 25 and 25 to 30. The four curves (bottom to top) are the 10, 100, 1000, and the 10000 year return levels.

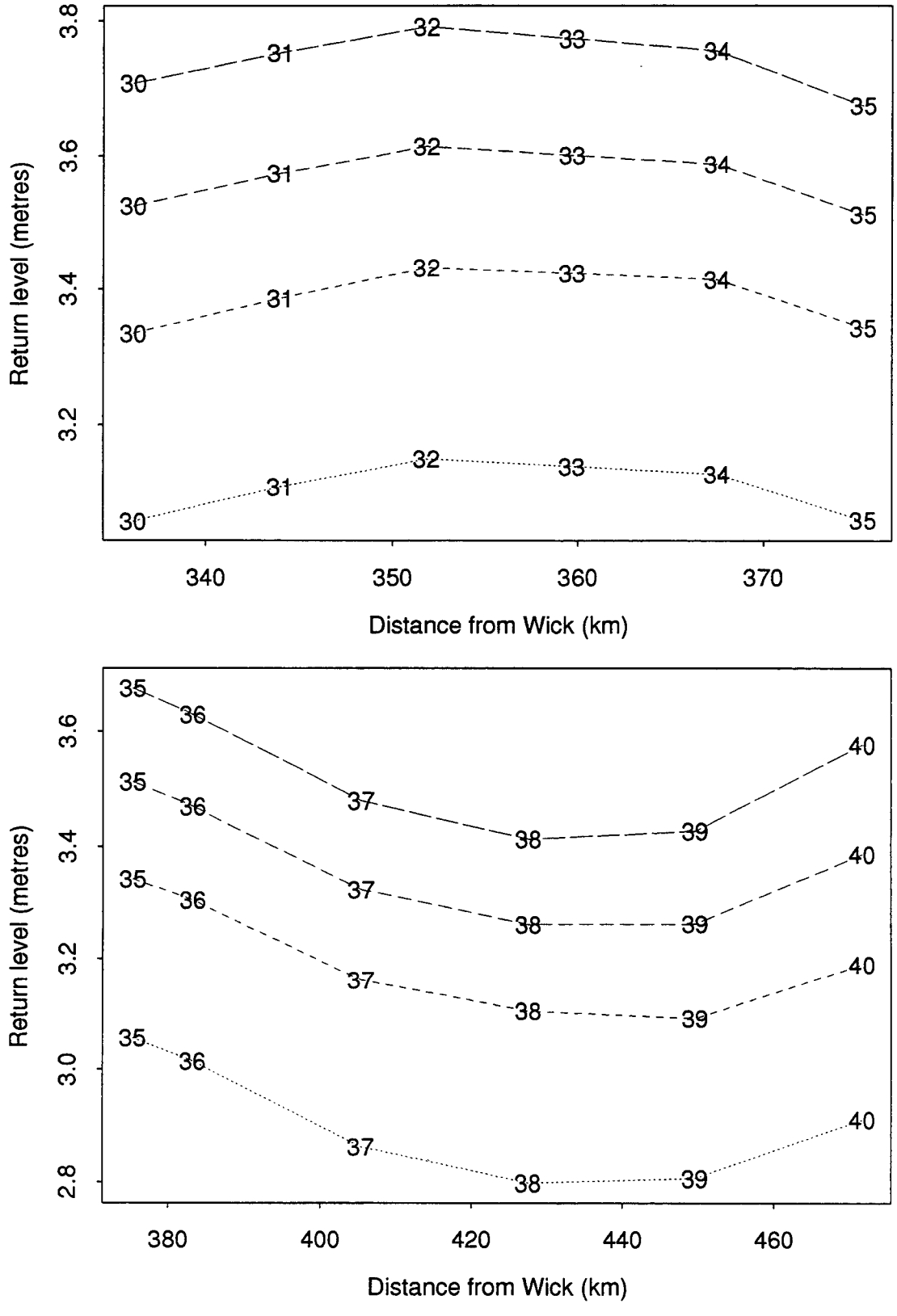

Figure 13.10: Spatial return level estimates in metres, relative to mean sea-level in 1990, for the east coast grid: site numbers 30 to 35 and 35 to 40. The four curves (bottom to top) are the 10, 100, 1000, and the 10000 year return levels.

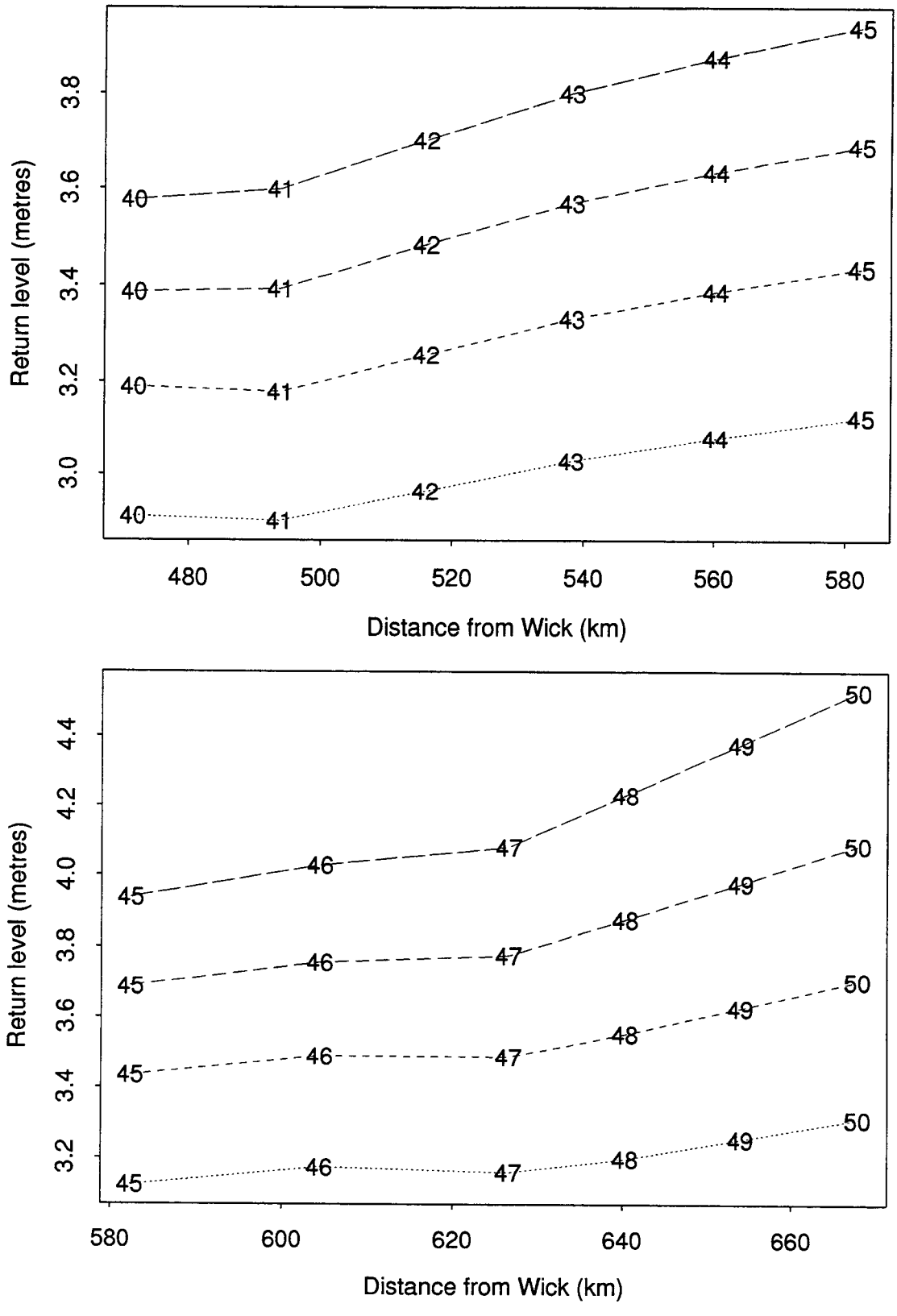

Figure 13.11: Spatial return level estimates in metres, relative to mean sea-level in 1990, for the east coast grid: site numbers 40 to 45 and 45 to 50. The four curves (bottom to top) are the 10, 100, 1000, and the 10000 year return levels.

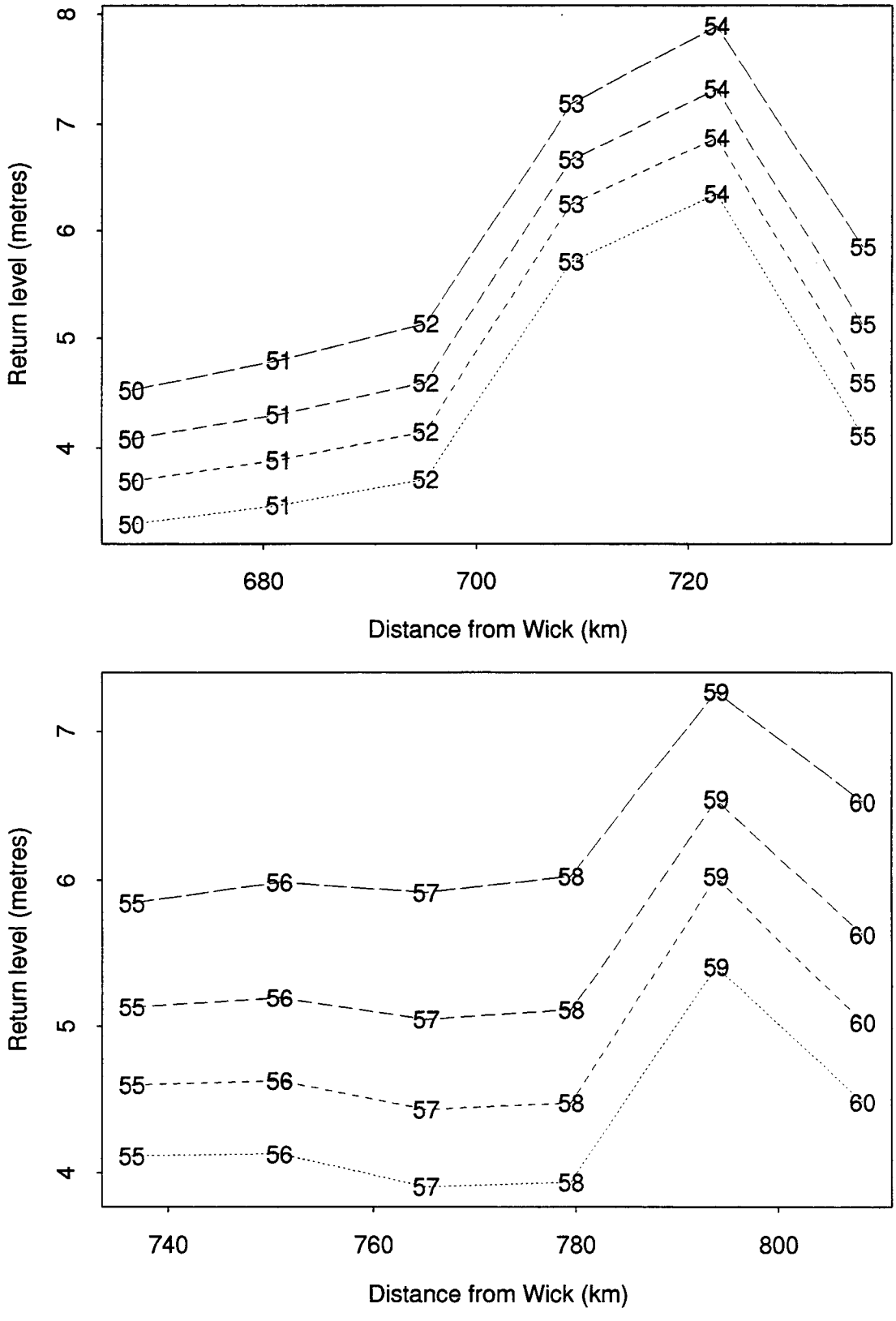

Figure 13.12: Spatial return level estimates in metres, relative to mean sea-level in 1990, for the east coast grid: site numbers 50 to 55 and 55 to 60. The four curves (bottom to top) are the 10, 100, 1000, and the 10000 year return levels.

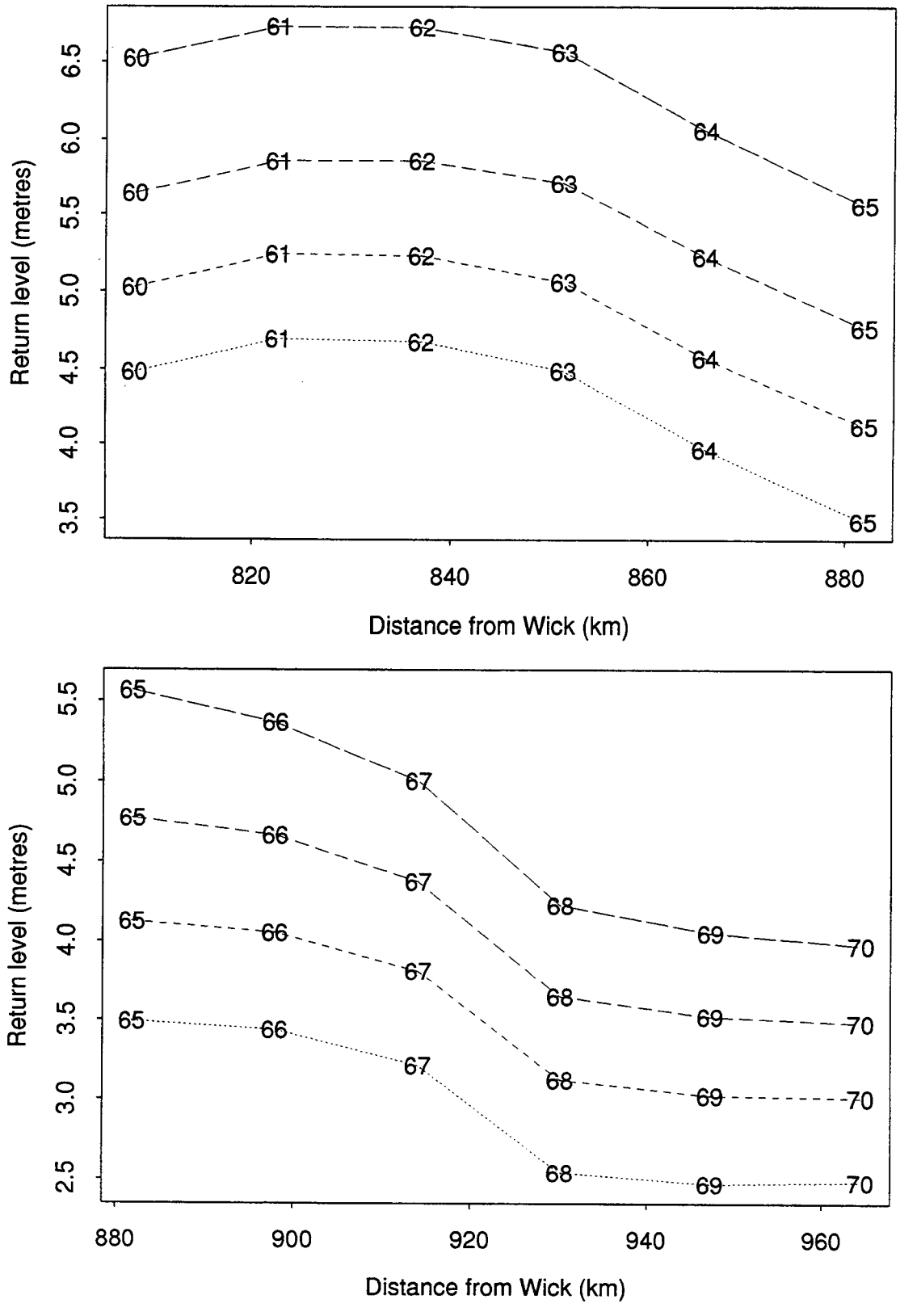

Figure 13.13: Spatial return level estimates in metres, relative to mean sea-level in 1990, for the east coast grid: site numbers 60 to 65 and 65 to 70. The four curves (bottom to top) are the 10, 100, 1000, and the 10000 year return levels.

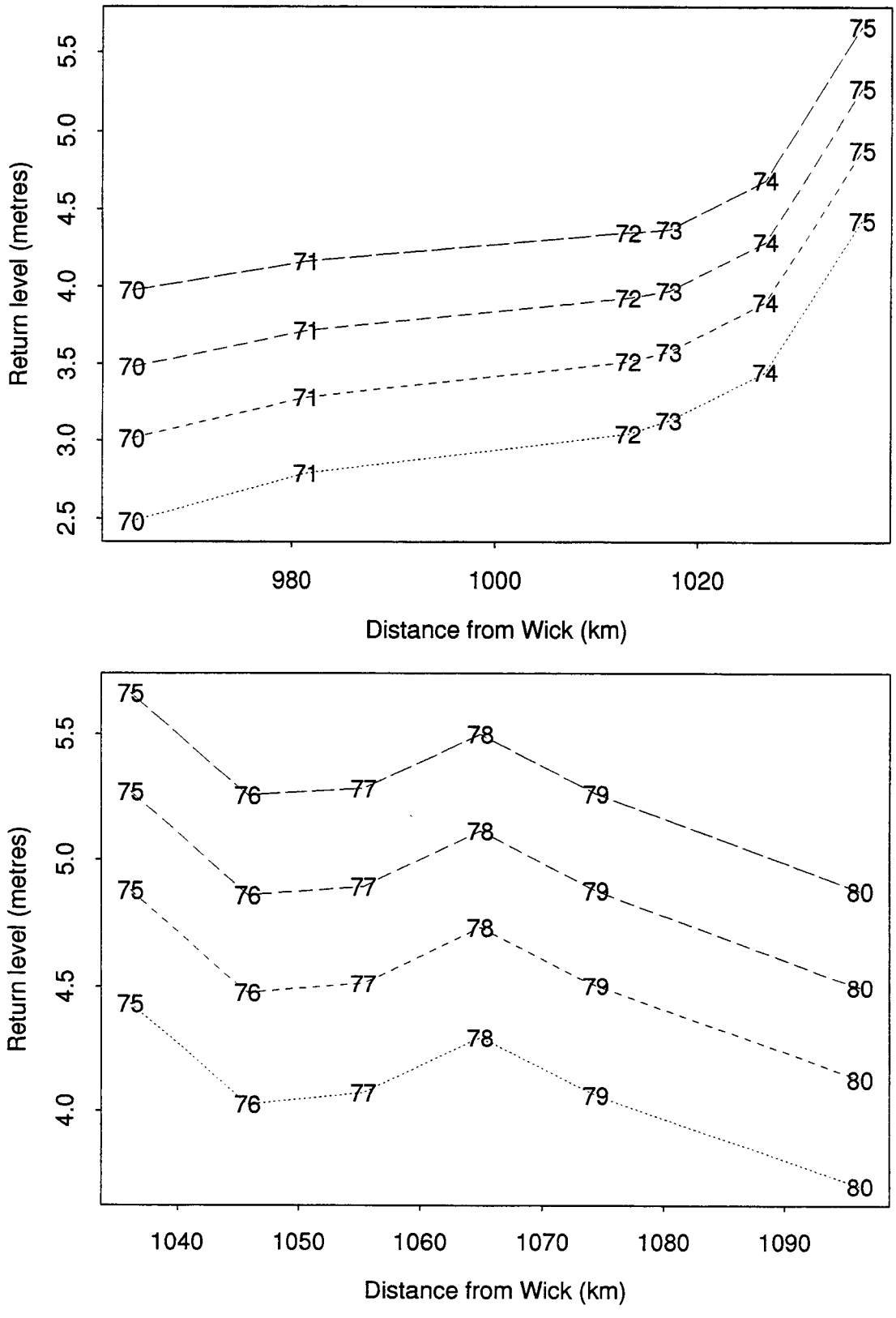

Figure 13.14: Spatial return level estimates in metres, relative to mean sea-level in 1990, for the east coast grid: site numbers 70 to 75 and 75 to 80. The four curves (bottom to top) are the 10, 100, 1000, and the 10000 year return levels.

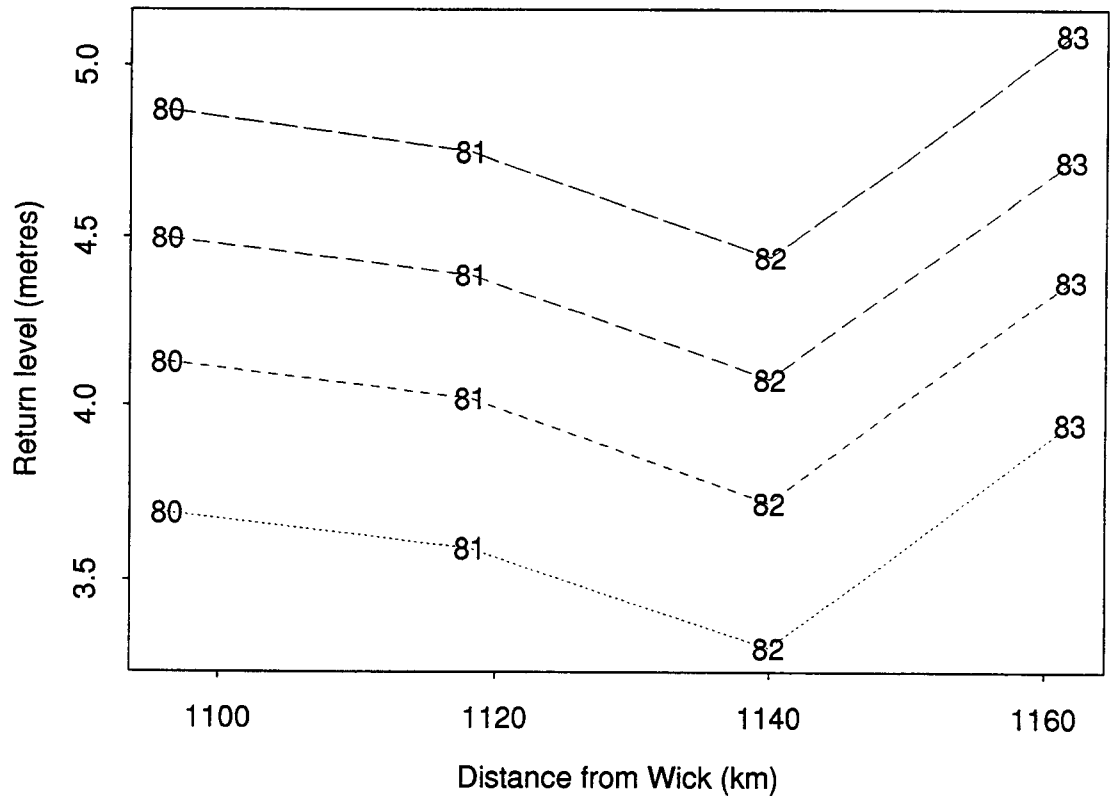

Figure 13.15: Spatial return level estimates in metres, relative to mean sea-level in 1990, for the east coast grid: site numbers 80 to 83. The four curves (bottom to top) are the 10, 100, 1000, and the 10000 year return levels.

The location of the grid sites is vital in determining the return level estimate in the five complex coastal regions, so the exact position of these sites is shown on separate maps of these regions in Figures 13.16-13.19. Combining the information from the maps and the spatial estimate for the stretches of five consecutive grid sites (coastal stretches of approximately 100km in length) we see how the spatial model performs within these regions. First recall that in fact the spatial model does not model the tides into small estuaries and inlets as it is based on interpolating tidal predictions from a 12km hydrodynamical model, so coastal structure of a localised scale is ignored. These within the estuary regions we are looking for large scale changes rather than local features that may be implied by the precise locations of the 20km grid. Also of importance is the non-linearity of the one dimensional metric we have used for the coastline. This has led to the grid producing many sites in estuaries and areas of high local coastal structure, i.e. at a denser frequency than on the smooth regions of open coast where the 20km spacing is easily obtained. Finally note that over the scale of five coastal regions the extremal surge parameters and interaction functions vary smoothly and slowly so that any observed spatial variation in return levels must be due to the predicted tidal series in these regions.

Now consider the five regions in turn. In the Moray Firth region (Figure 13.16, sites 2-15) the spatial estimate changes very smoothly apart from between grid sites 8 and 10. The precise reason for this jump in level is unclear but it probably reflects the resolution of the hydrodynamical model in this region, i.e. this is the narrowest part of the estuary. In the Firth of Forth region (Figure 13.17, sites 23-37) the spatial estimate also varies quite smoothly with the maximum value obtained at the furthest grid point up the estuary. In the Humber estuary region (Figure 13.18, sites 51-57) sites 53 and 54 which are well up the estuary give much larger estimates than for the neighbouring open coastline. Recall that the spatial interpolation of the tide is not too reliable up estuaries and the high non-linearity in this region also casts doubt over the spatial interpolation of the SRJPM parameters between these sites. In fact the agreement with the return level estimates for the Humber given in Chapter 9 is poor beyond the Humber Bridge. This suggests that the Humber case study return level estimates should be used for sites up the estuary beyond the Humber Bridge. However, the good agreement between the Immingham site-by-site estimate and the spatial model suggests that the spatial model is fine for use in the outer reaches of the estuary. In the Wash region (Figure 13.19, sites 58-63) grid site 59 has a notably different spatial estimate than the neighbouring estimates on either side. This is probably due to the non-linearities of the tide in the entrance to the Wash, as indicated by the results in Pugh and Vassie (1976). Finally, in the Thames estuary region (Figure 13.20, sites 70-83) sites 75-78 are notably different from the neighbouring sites. Again this is due to these being at the mouth of the estuary, and so in the most shallow water.

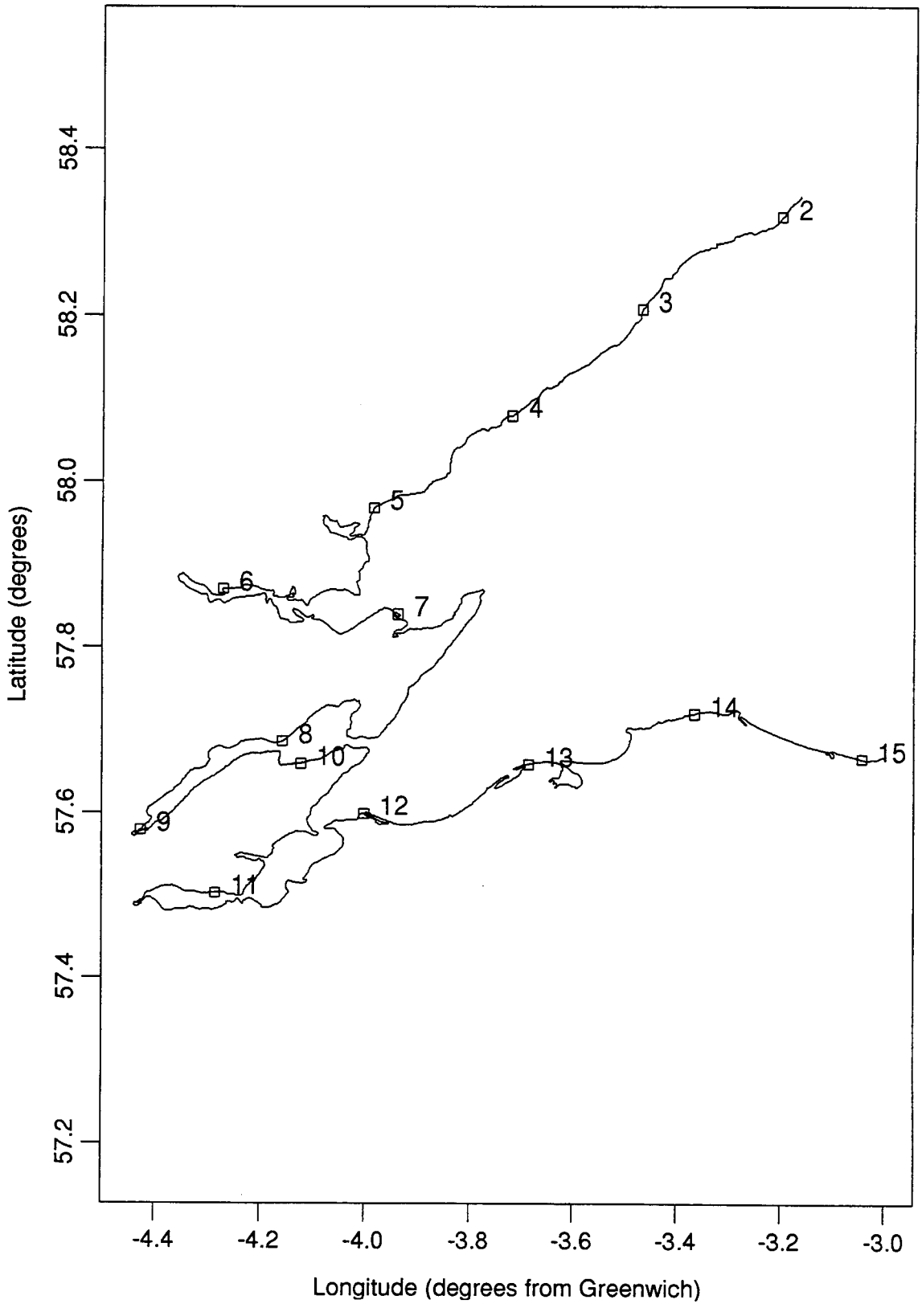

Figure 13.16: Moray Firth. □ and the numeric labels indicate sites on the 20km grid.

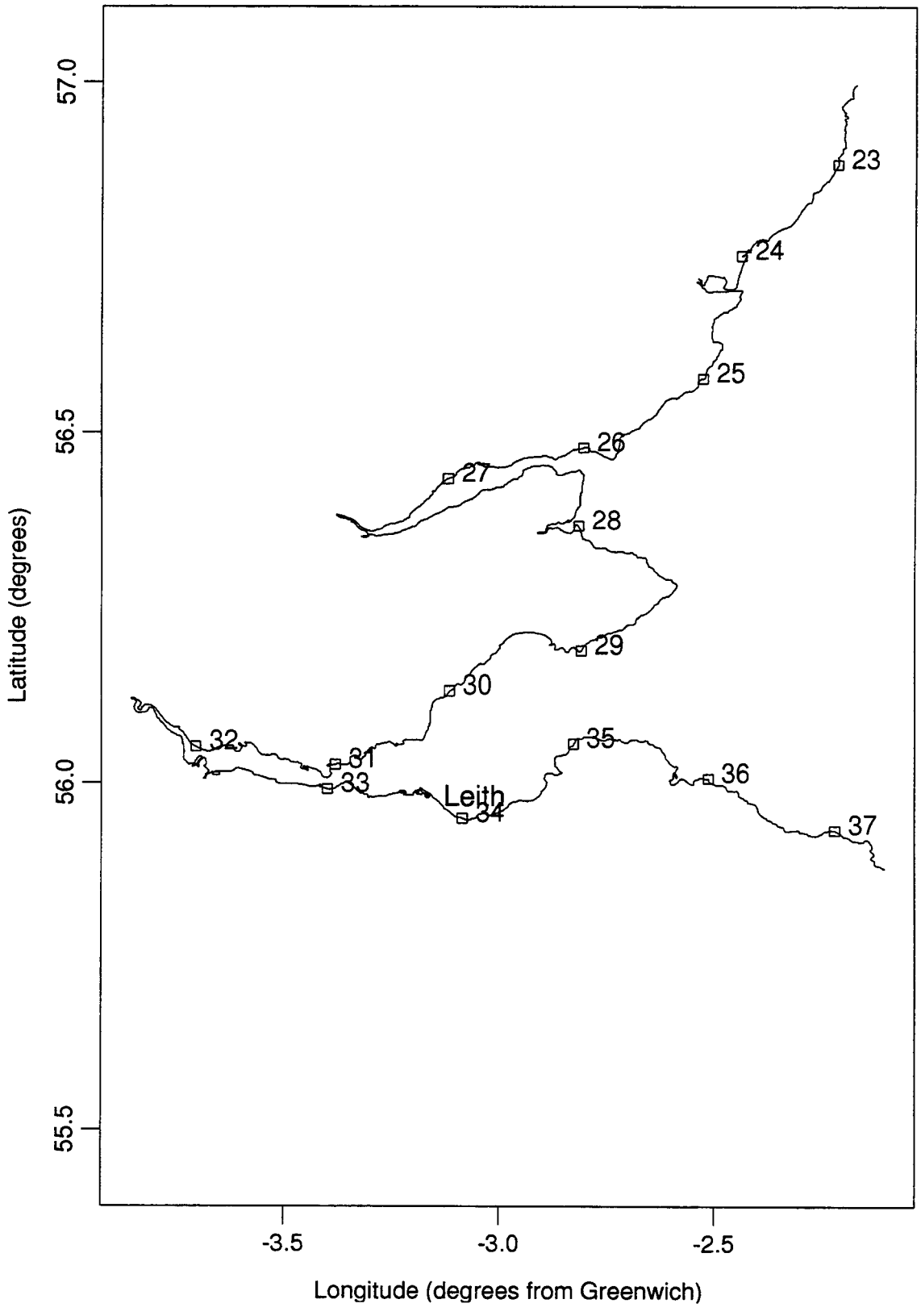

Figure 13.17: Firth of Forth. □ and the numeric labels indicate sites on the 20km grid and (•) indicates the study sites in this region.

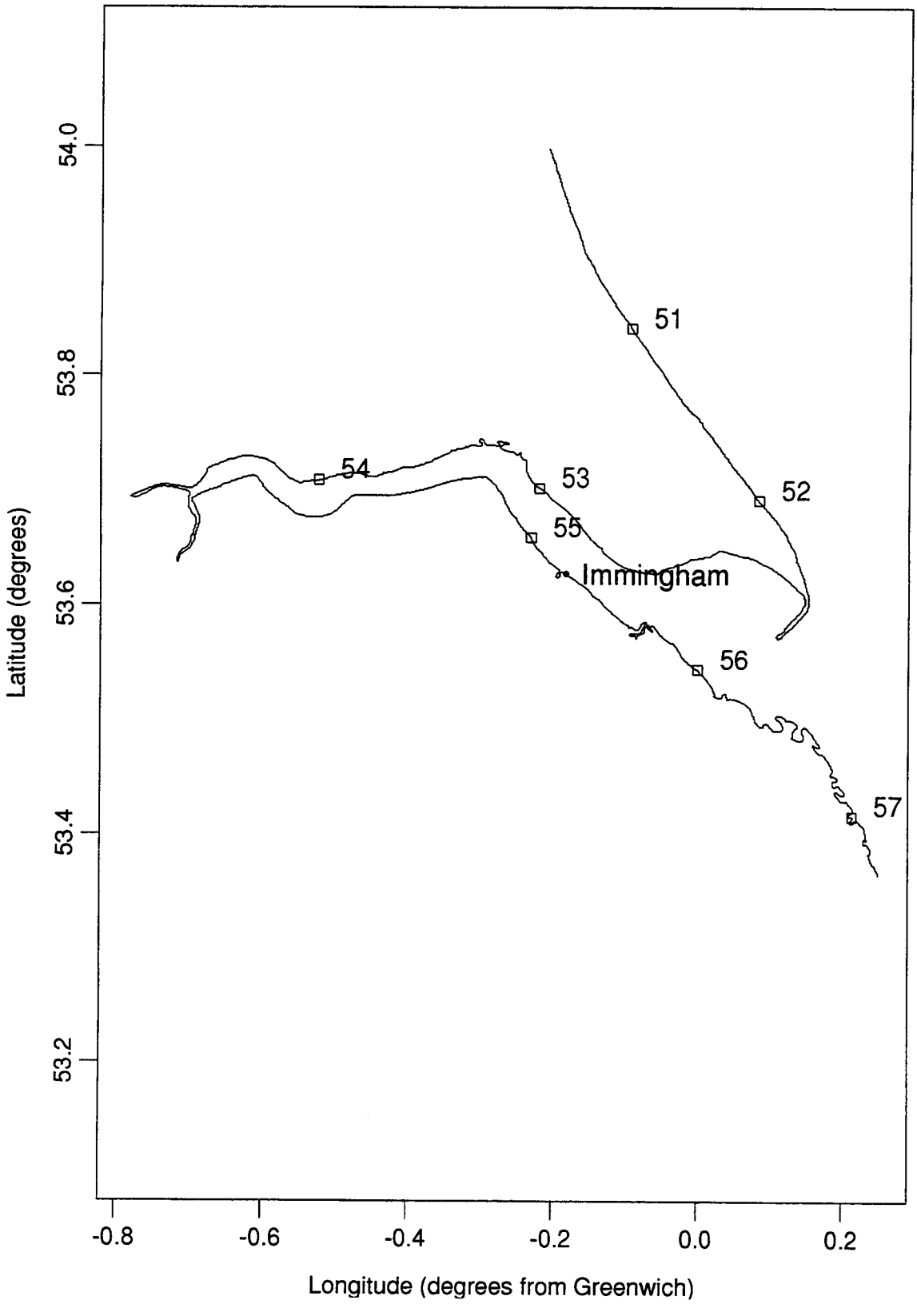

Figure 13.18: Humber Estuary. □ and the numeric labels indicate sites on the 20km grid and (•) indicates the study sites in this region.

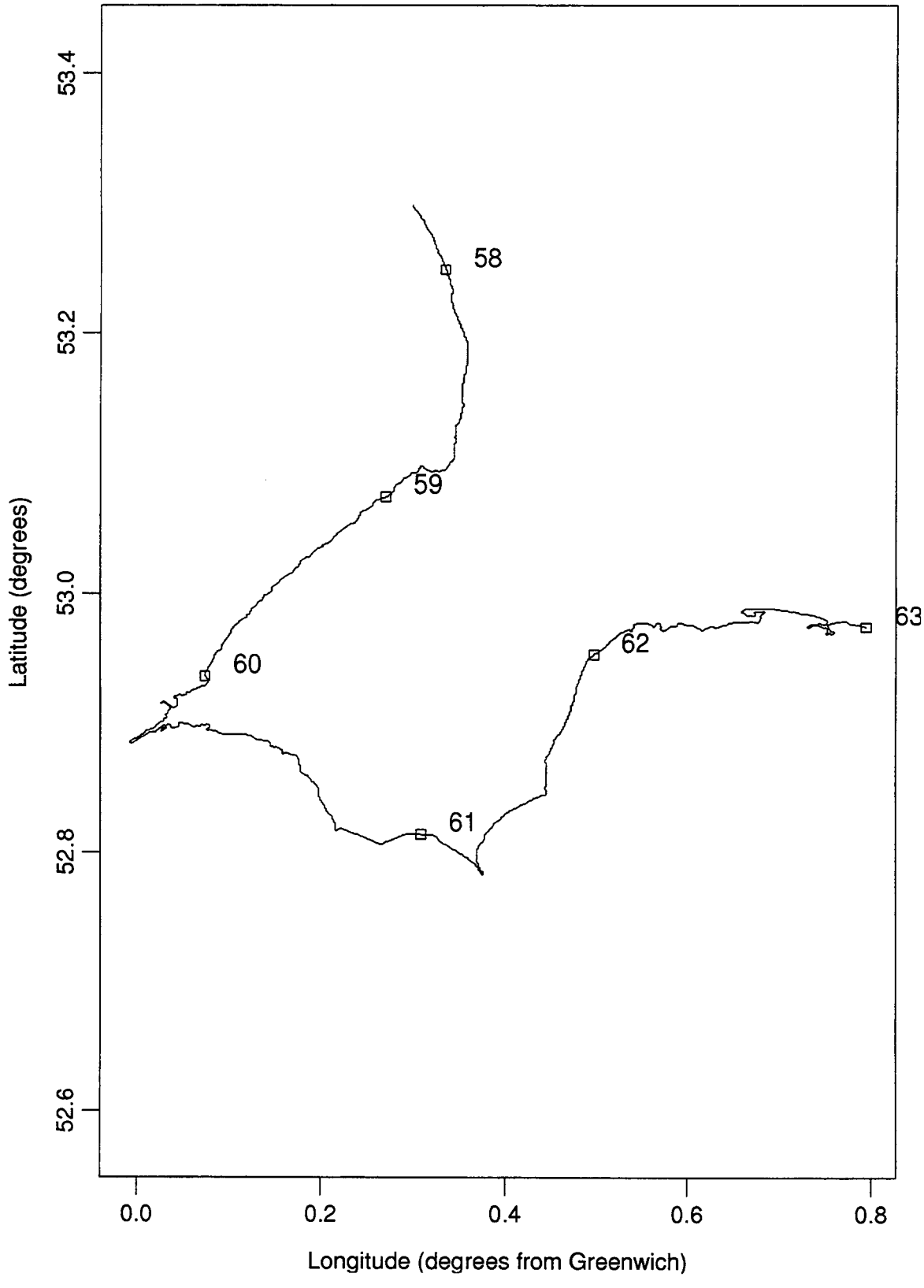

Figure 13.19: The Wash. □ and the numeric labels indicate sites on the 20km grid.

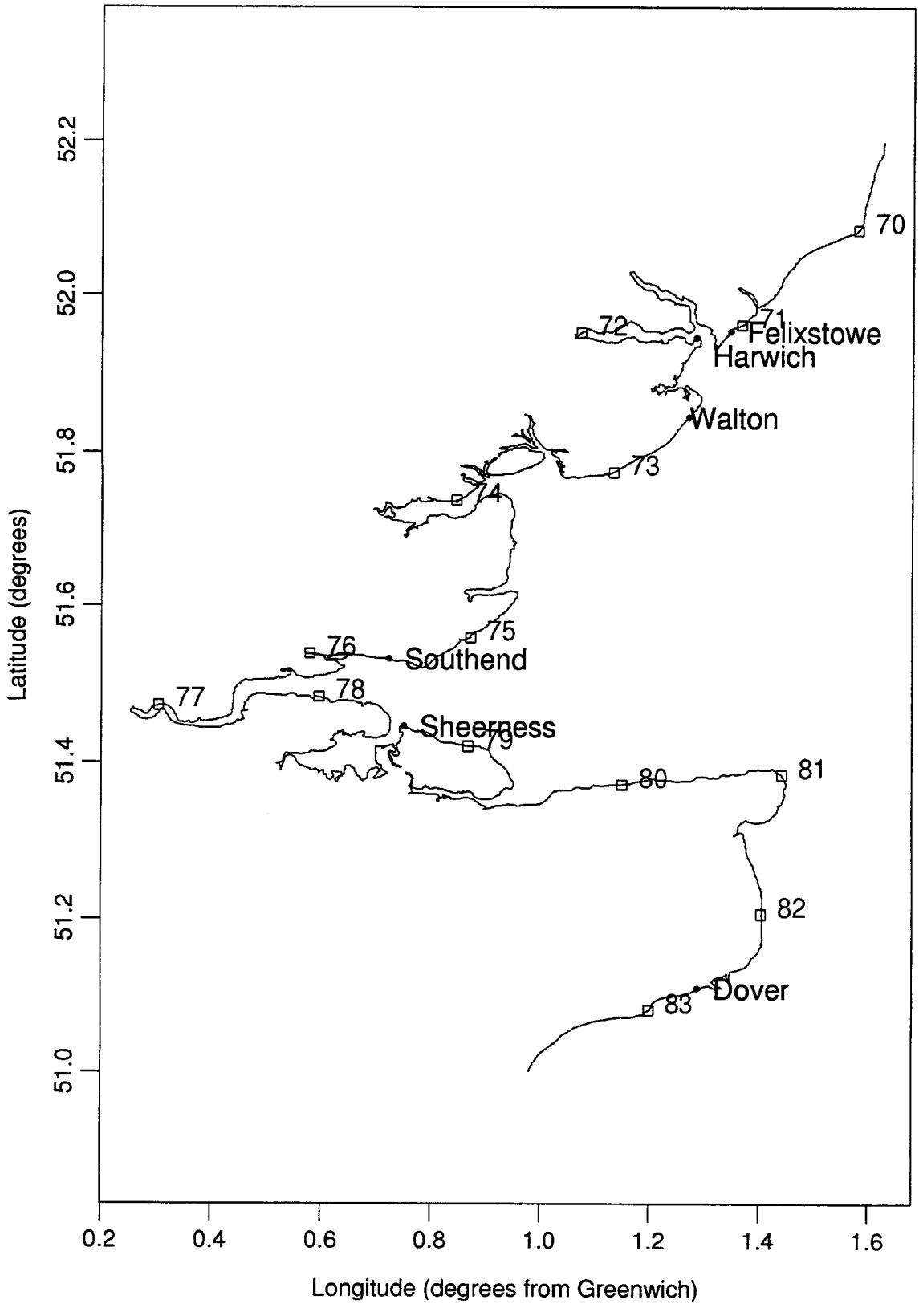

Figure 13.20: The Thames estuary region. □ and the numeric labels indicate sites on the 20km grid and (•) indicates the study sites in this region.

Detailed output from the spatial analysis is given in Sections 13.1 and 13.3 where tables of return level estimates and trends are given for the 20km grid sites and the study sites respectively.

Throughout Chapter 13 we have not obtained standard errors and confidence intervals for the spatial estimate of the return level curves as they provide a measure of the uncertainty due to parameter estimation. From Figures 13.1-13.4 the key component for spatially interpolating return levels is seen to be the spatial interpolation of the tides. This is a deterministic interpolation for which the quality of the interpolation is determined by the accuracy of the numerical model of the tidal process. This accuracy cannot be easily calibrated. If the tidal interpolation can be completely trusted then the standard errors of the spatial return level estimates can be reasonably based on the most reliable neighbouring site-by-site estimate.

13.1 Results for the east coast grid

In this section we give tables of the spatial return level estimates and spatial trend estimates for each grid site on the 20km grid. These return level estimates are given relative to mean sea-level at each site. Information to allow these estimates to be converted to ODN, or some other datum of interest, are given in Section 13.2.

Site			l year return level			
Number	long	lat	l=10	l=25	l=50	l=100
1	-3.08	58.43	2.24	2.34	2.39	2.46
2	-3.20	58.32	2.24	2.34	2.39	2.46
3	-3.47	58.21	2.32	2.42	2.46	2.54
4	-3.72	58.08	2.38	2.48	2.53	2.60
5	-3.98	57.97	2.47	2.57	2.62	2.69
6	-4.27	57.87	2.51	2.61	2.65	2.73
7	-3.94	57.84	2.53	2.63	2.68	2.75
8	-4.16	57.69	2.53	2.63	2.68	2.75
9	-4.43	57.58	2.62	2.72	2.77	2.85
10	-4.12	57.66	2.62	2.72	2.77	2.85
11	-4.29	57.50	2.62	2.72	2.77	2.85
12	-4.00	57.60	2.63	2.73	2.78	2.86
13	-3.69	57.66	2.63	2.73	2.78	2.87
14	-3.37	57.72	2.58	2.69	2.75	2.83
15	-3.04	57.67	2.55	2.66	2.71	2.80
16	-2.71	57.69	2.51	2.62	2.68	2.77
17	-2.37	57.68	2.44	2.56	2.62	2.71
18	-2.03	57.69	2.38	2.49	2.55	2.64
19	-1.81	57.56	2.32	2.43	2.49	2.58
20	-1.87	57.38	2.28	2.39	2.45	2.54
21	-2.05	57.23	2.36	2.47	2.52	2.62
22	-2.11	57.05	2.55	2.66	2.72	2.81
23	-2.21	56.88	2.56	2.67	2.73	2.82
24	-2.44	56.75	2.72	2.83	2.89	2.98
25	-2.53	56.58	2.80	2.93	2.99	3.08
26	-2.80	56.48	2.83	2.96	3.02	3.11
27	-3.12	56.43	2.98	3.11	3.17	3.26
28	-2.81	56.36	2.97	3.11	3.16	3.25
29	-2.81	56.18	2.96	3.10	3.16	3.24
30	-3.12	56.13	3.05	3.19	3.25	3.33

Table 13.1: Return level estimates, in metres relative to mean sea-level in 1990, obtained from the spatial model for the east coast grid: sites 1-30. Return levels are given for return periods of 10, 25, 50 and 100 years.

Site			l year return level			
Number	long	lat	l=10	l=25	l=50	l=100
31	-3.38	56.03	3.11	3.25	3.30	3.39
32	-3.70	56.05	3.15	3.29	3.35	3.43
33	-3.40	55.99	3.14	3.28	3.34	3.42
34	-3.08	55.95	3.13	3.27	3.33	3.41
35	-2.83	56.05	3.06	3.20	3.27	3.34
36	-2.51	56.01	3.02	3.16	3.23	3.30
37	-2.22	55.93	2.86	3.01	3.08	3.16
38	-2.01	55.79	2.80	2.95	3.02	3.10
39	-1.83	55.64	2.81	2.94	3.01	3.09
40	-1.61	55.51	2.91	3.04	3.10	3.19
41	-1.55	55.33	2.90	3.02	3.08	3.18
42	-1.52	55.15	2.96	3.09	3.15	3.26
43	-1.38	54.98	3.03	3.16	3.22	3.33
44	-1.31	54.80	3.08	3.21	3.28	3.39
45	-1.16	54.65	3.12	3.25	3.32	3.44
46	-0.88	54.57	3.17	3.30	3.38	3.49
47	-0.60	54.49	3.16	3.29	3.37	3.49
48	-0.42	54.34	3.20	3.34	3.42	3.55
49	-0.27	54.18	3.25	3.40	3.49	3.63
50	-0.21	54.01	3.31	3.46	3.55	3.70
51	-0.09	53.84	3.49	3.64	3.74	3.90
52	0.08	53.69	3.72	3.88	3.98	4.15
53	-0.22	53.70	5.71	5.99	6.10	6.26
54	-0.52	53.71	6.36	6.62	6.71	6.88
55	-0.23	53.66	4.12	4.30	4.41	4.60
56	0.00	53.54	4.13	4.31	4.42	4.63
57	0.21	53.42	3.91	4.10	4.22	4.43
58	0.33	53.25	3.94	4.13	4.25	4.48
59	0.27	53.07	5.42	5.73	5.85	6.03
60	0.07	52.94	4.48	4.71	4.82	5.03

Table 13.2: Return level estimates, in metres relative to mean sea-level in 1990, obtained from the spatial model for the east coast grid: sites 31-60. Return levels are given for return periods of 10, 25, 50 and 100 years.

Site			l year return level			
Number	long	lat	l=10	l=25	l=50	l=100
61	0.31	52.81	4.70	4.91	5.03	5.24
62	0.50	52.95	4.68	4.88	5.01	5.23
63	0.79	52.97	4.49	4.70	4.83	5.06
64	1.09	52.96	3.97	4.19	4.33	4.57
65	1.38	52.91	3.49	3.73	3.88	4.13
66	1.62	52.78	3.44	3.67	3.81	4.06
67	1.74	52.62	3.21	3.44	3.58	3.81
68	1.73	52.44	2.55	2.77	2.90	3.12
69	1.63	52.26	2.47	2.69	2.81	3.02
70	1.58	52.08	2.48	2.69	2.81	3.01
71	1.37	51.96	2.79	2.98	3.10	3.28
72	1.07	51.95	3.05	3.23	3.34	3.52
73	1.13	51.77	3.13	3.31	3.42	3.58
74	0.84	51.74	3.45	3.63	3.74	3.90
75	0.87	51.56	4.43	4.61	4.71	4.88
76	0.58	51.54	4.03	4.20	4.31	4.48
77	0.30	51.47	4.07	4.25	4.35	4.51
78	0.59	51.48	4.29	4.47	4.57	4.73
79	0.87	51.42	4.06	4.24	4.34	4.50
80	1.15	51.37	3.69	3.87	3.96	4.13
81	1.44	51.38	3.59	3.76	3.86	4.02
82	1.40	51.21	3.30	3.47	3.56	3.72
83	1.20	51.08	3.95	4.12	4.21	4.37

Table 13.3: Return level estimates, in metres relative to mean sea-level in 1990, obtained from the spatial model for the east coast grid: sites 61-83. Return levels are given for return periods of 10, 25, 50 and 100 years.

Site			l year return level			
Number	long	lat	l=250	l=500	l=1000	l=10000
1	-3.08	58.43	2.52	2.56	2.60	2.72
2	-3.20	58.32	2.52	2.56	2.60	2.72
3	-3.47	58.21	2.60	2.64	2.68	2.80
4	-3.72	58.08	2.67	2.70	2.75	2.87
5	-3.98	57.97	2.75	2.79	2.83	2.95
6	-4.27	57.87	2.79	2.83	2.87	2.99
7	-3.94	57.84	2.82	2.85	2.90	3.02
8	-4.16	57.69	2.82	2.86	2.90	3.03
9	-4.43	57.58	2.91	2.95	3.00	3.13
10	-4.12	57.66	2.92	2.96	3.01	3.15
11	-4.29	57.50	2.93	2.97	3.02	3.16
12	-4.00	57.60	2.94	2.98	3.03	3.19
13	-3.69	57.66	2.95	2.99	3.05	3.21
14	-3.37	57.72	2.92	2.96	3.02	3.20
15	-3.04	57.67	2.89	2.94	3.00	3.19
16	-2.71	57.69	2.86	2.91	2.98	3.17
17	-2.37	57.68	2.80	2.85	2.92	3.13
18	-2.03	57.69	2.74	2.79	2.86	3.06
19	-1.81	57.56	2.67	2.73	2.79	3.00
20	-1.87	57.38	2.63	2.69	2.76	2.96
21	-2.05	57.23	2.71	2.76	2.83	3.03
22	-2.11	57.05	2.90	2.95	3.02	3.22
23	-2.21	56.88	2.91	2.96	3.02	3.22
24	-2.44	56.75	3.07	3.12	3.19	3.39
25	-2.53	56.58	3.17	3.21	3.28	3.48
26	-2.80	56.48	3.19	3.24	3.31	3.51
27	-3.12	56.43	3.34	3.39	3.46	3.65
28	-2.81	56.36	3.33	3.38	3.45	3.64
29	-2.81	56.18	3.33	3.38	3.44	3.63
30	-3.12	56.13	3.41	3.46	3.52	3.71

Table 13.4: Return level estimates, in metres relative to mean sea-level in 1990, obtained from the spatial model for the east coast grid: sites 1-30. Return levels are given for return periods of 250, 500, 1000, 10000 years.

Site			l year return level			
Number	long	lat	l=250	l=500	l=1000	l=10000
31	-3.38	56.03	3.47	3.51	3.57	3.75
32	-3.70	56.05	3.51	3.56	3.61	3.79
33	-3.40	55.99	3.50	3.54	3.60	3.78
34	-3.08	55.95	3.49	3.53	3.59	3.76
35	-2.83	56.05	3.42	3.46	3.51	3.68
36	-2.51	56.01	3.38	3.42	3.47	3.63
37	-2.22	55.93	3.23	3.27	3.32	3.48
38	-2.01	55.79	3.17	3.21	3.26	3.41
39	-1.83	55.64	3.17	3.21	3.26	3.43
40	-1.61	55.51	3.27	3.32	3.39	3.58
41	-1.55	55.33	3.27	3.32	3.39	3.60
42	-1.52	55.15	3.35	3.41	3.48	3.70
43	-1.38	54.98	3.43	3.49	3.57	3.80
44	-1.31	54.80	3.49	3.55	3.63	3.87
45	-1.16	54.65	3.54	3.61	3.69	3.94
46	-0.88	54.57	3.60	3.67	3.76	4.03
47	-0.60	54.49	3.61	3.68	3.78	4.08
48	-0.42	54.34	3.69	3.77	3.88	4.23
49	-0.27	54.18	3.77	3.86	3.98	4.38
50	-0.21	54.01	3.86	3.95	4.08	4.53
51	-0.09	53.84	4.06	4.16	4.31	4.80
52	0.08	53.69	4.33	4.44	4.60	5.14
53	-0.22	53.70	6.43	6.53	6.67	7.19
54	-0.52	53.71	7.06	7.17	7.33	7.90
55	-0.23	53.66	4.81	4.94	5.13	5.85
56	0.00	53.54	4.85	5.00	5.20	5.99
57	0.21	53.42	4.67	4.83	5.05	5.92
58	0.33	53.25	4.72	4.88	5.12	6.03
59	0.27	53.07	6.23	6.36	6.54	7.28
60	0.07	52.94	5.26	5.42	5.64	6.53

Table 13.5: Return level estimates, in metres relative to mean sea-level in 1990, obtained from the spatial model for the east coast grid: sites 31-60. Return levels are given for return periods of 250, 500, 1000, 10000 years.

Site			l year return level			
Number	long	lat	l=250	l=500	l=1000	l=10000
61	0.31	52.81	5.48	5.64	5.86	6.74
62	0.50	52.95	5.48	5.64	5.86	6.73
63	0.79	52.97	5.32	5.48	5.71	6.57
64	1.09	52.96	4.83	5.00	5.23	6.06
65	1.38	52.91	4.39	4.55	4.77	5.57
66	1.62	52.78	4.30	4.46	4.66	5.37
67	1.74	52.62	4.04	4.19	4.37	5.01
68	1.73	52.44	3.34	3.48	3.65	4.23
69	1.63	52.26	3.23	3.36	3.52	4.05
70	1.58	52.08	3.21	3.33	3.48	3.97
71	1.37	51.96	3.47	3.57	3.72	4.16
72	1.07	51.95	3.69	3.80	3.93	4.35
73	1.13	51.77	3.75	3.85	3.98	4.37
74	0.84	51.74	4.07	4.17	4.29	4.69
75	0.87	51.56	5.05	5.14	5.27	5.67
76	0.58	51.54	4.64	4.74	4.86	5.25
77	0.30	51.47	4.67	4.77	4.89	5.28
78	0.59	51.48	4.90	4.99	5.11	5.50
79	0.87	51.42	4.66	4.75	4.88	5.26
80	1.15	51.37	4.29	4.38	4.50	4.88
81	1.44	51.38	4.18	4.27	4.39	4.76
82	1.40	51.21	3.87	3.96	4.08	4.44
83	1.20	51.08	4.52	4.61	4.73	5.09

Table 13.6: Return level estimates, in metres relative to mean sea-level in 1990, obtained from the spatial model for the east coast grid: sites 61-83. Return levels are given for return periods of 250, 500, 1000, 10000 years.

13.2 Conversion of the spatial estimate to datum of interest

Now consider the conversion of the spatial estimate from an estimate relative to the mean sea-level datum to one relative to the datum of interest. Here we illustrate how this is achieved when the datum of interest is ODN. To convert the estimate two steps are required:

1. Convert the return level estimate for 1990 to ODN by an additive adjustment for the mean sea-level in 1990 relative to ODN. The adjustment factor is shown for the entire east coast in Figure 13.21. That figure shows estimates of the mean sea-level

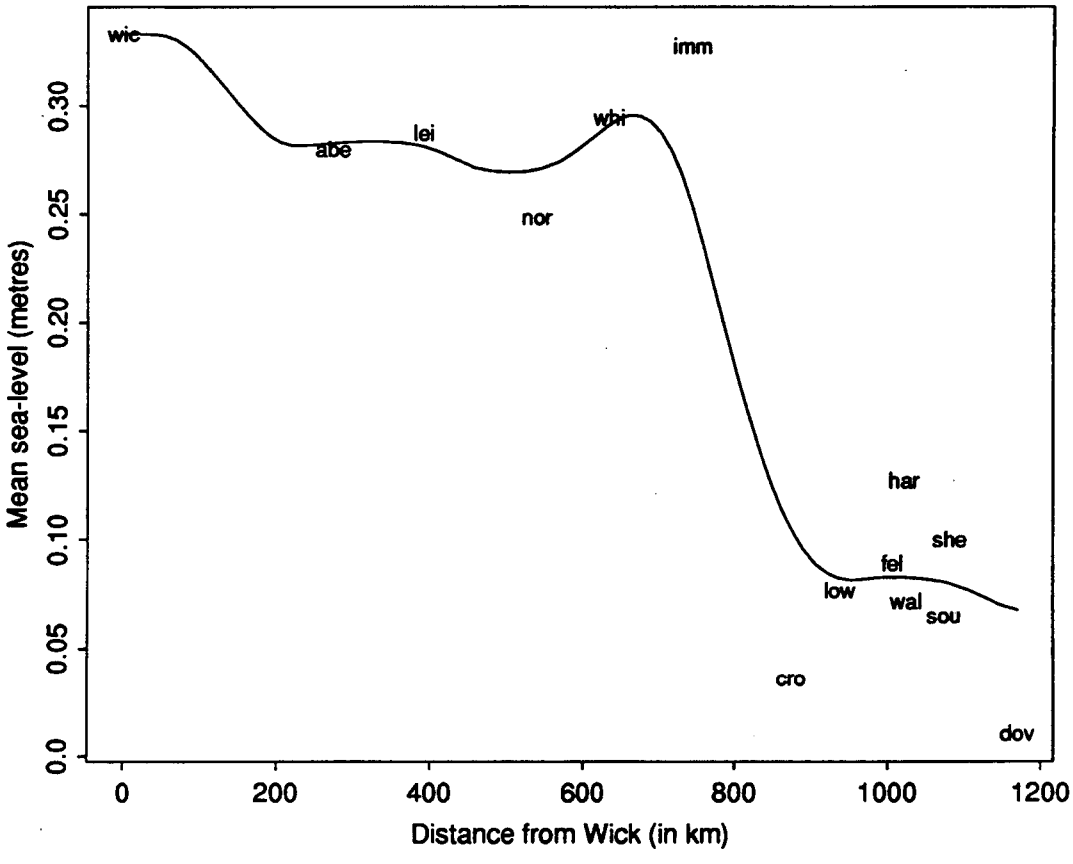

Figure 13.21: Mean sea-level in 1990 against distance from Wick; the site-by-site estimates at each site and the spatial estimate.

in 1990 for each study site relative to ODN. The site-by-site estimates were obtained by predicting the mean sea-level in 1990 at the sites using the spatial estimate of the trend (see Section 12.2). In Figure 13.21 we see that there is a small, but systematic, difference between the mean sea-level datum and ODN, with a drift in

value from near 35cm at Wick to approximately 7cm at Dover. This drift feature is a well known inconsistency in the mean sea-level to ODN relationship, and this figure nicely shows that it is probably due to an error in the levelling of ODN between Immingham and Cromer. To aid the conversion the adjustment factor for each of the 83 grid sites are also given in Tables 13.7-13.9. These values have been obtained by spatially smoothing the site-by-site estimates of the conversion factor, i.e. the values corresponding to the smooth curve on Figure 13.21.

2. Convert the return level estimate for 1990, which is relative to ODN, to the year of interest, using the spatial estimate of the trend given in Tables 13.7-13.9.

Number	long	lat	MSL in 1990	trend (mm per yr)
1	-3.083	58.433	0.334	3.220
2	-3.199	58.318	0.334	3.097
3	-3.469	58.208	0.334	2.967
4	-3.718	58.080	0.333	2.827
5	-3.984	57.968	0.333	2.651
6	-4.272	57.869	0.331	2.445
7	-3.937	57.839	0.328	2.219
8	-4.158	57.688	0.323	2.082
9	-4.428	57.580	0.318	2.061
10	-4.123	57.660	0.313	2.087
11	-4.287	57.503	0.307	1.968
12	-4.002	57.599	0.301	1.728
13	-3.685	57.660	0.295	1.448
14	-3.367	57.721	0.290	1.217
15	-3.043	57.667	0.286	1.044
16	-2.709	57.690	0.283	0.902
17	-2.370	57.675	0.282	0.797
18	-2.033	57.694	0.282	0.721
19	-1.808	57.561	0.282	0.665
20	-1.869	57.384	0.282	0.630
21	-2.048	57.231	0.283	0.626
22	-2.109	57.053	0.283	0.626
23	-2.214	56.883	0.283	0.639
24	-2.438	56.751	0.283	0.652
25	-2.527	56.578	0.283	0.668
26	-2.803	56.479	0.283	0.685
27	-3.118	56.433	0.283	0.703
28	-2.814	56.364	0.283	0.724
29	-2.808	56.184	0.283	0.745
30	-3.115	56.129	0.283	0.766

Table 13.7: Conversion factor (in metres) for the mean sea-level in 1990 to ODN and the spatial estimate of the sea-level trend (in mm/year): for the east coast grid sites 1-30.

Number	long	lat	MSL in 1990	trend (mm per yr)
31	-3.380	56.026	0.283	0.787
32	-3.702	56.051	0.283	0.808
33	-3.398	55.991	0.283	0.822
34	-3.085	55.949	0.282	0.836
35	-2.826	56.055	0.282	0.850
36	-2.513	56.006	0.281	0.869
37	-2.218	55.931	0.279	0.994
38	-2.013	55.792	0.275	1.230
39	-1.833	55.639	0.271	1.491
40	-1.614	55.506	0.270	1.718
41	-1.554	55.326	0.269	1.908
42	-1.518	55.147	0.270	2.075
43	-1.381	54.985	0.271	2.230
44	-1.313	54.804	0.274	2.394
45	-1.159	54.648	0.279	2.652
46	-0.877	54.571	0.285	3.101
47	-0.600	54.489	0.291	3.730
48	-0.423	54.340	0.294	4.111
49	-0.271	54.183	0.296	4.432
50	-0.211	54.006	0.295	4.494
51	-0.093	53.840	0.293	4.419
52	0.084	53.693	0.287	4.253
53	-0.220	53.704	0.279	4.067
54	-0.523	53.710	0.268	3.910
55	-0.231	53.660	0.253	3.773
56	0.000	53.544	0.235	3.655
57	0.213	53.416	0.216	3.546
58	0.334	53.249	0.197	3.429
59	0.270	53.074	0.178	3.296
60	0.074	52.936	0.160	3.107

Table 13.8: Conversion factor (in metres) for the mean sea-level in 1990 to ODN and the spatial estimate of the sea-level trend (in mm/year): for the east coast grid sites 31-60.

Number	lat	long	MSL in 1990	trend (mm per yr)
61	0.308	52.814	0.143	2.871
62	0.497	52.953	0.128	2.598
63	0.795	52.974	0.115	2.286
64	1.093	52.958	0.105	1.948
65	1.380	52.906	0.096	1.593
66	1.619	52.783	0.089	1.293
67	1.744	52.620	0.084	1.072
68	1.734	52.438	0.082	1.012
69	1.632	52.262	0.081	1.119
70	1.580	52.085	0.082	1.317
71	1.367	51.961	0.083	1.534
72	1.073	51.951	0.083	1.855
73	1.133	51.775	0.083	1.880
74	0.845	51.739	0.082	1.941
75	0.871	51.558	0.082	1.996
76	0.578	51.538	0.082	2.037
77	0.305	51.472	0.081	2.076
78	0.595	51.483	0.081	2.107
79	0.867	51.420	0.080	2.138
80	1.150	51.372	0.077	2.207
81	1.442	51.385	0.074	2.288
82	1.404	51.207	0.070	2.366
83	1.198	51.081	0.068	2.447

Table 13.9: Conversion factor (in metres) for the mean sea-level in 1990 to ODN and the spatial estimate of the sea-level trend (in mm/year): for the east coast grid sites 61-83.

In any application for a site which is not a study site there are two approaches possible:

1. Obtain a conversion factor by suitable interpolation of the values in Tables 13.7-13.9 and use that to adjust the estimate.

2. Obtain a short record of data from the site of interest, measured relative to ODN, and use this to estimate the mean sea-level over the time period of observation. Using the spatial trend estimate of Figure 12.5 adjust the mean sea-level estimate for the site to 1990. This provides the adjustment factor to convert from mean sea-level datum to ODN.

Converting results from ODN to ACD or another datum of interest is straightforward.

13.3 Results for study sites

In this section we give tables of the spatial return level estimates and spatial trend estimates for each study site on the east coast. These return level estimates are given relative to ACD. As study sites do not fall exactly on the 20km grid over the east coast special results have to be derived. Three approaches to obtain return level estimates for the study sites are

1. Interpolate the spatial return level estimates that are provided for the 20km grid in Tables 13.1-13.6. Similarly the trend estimate for the site should be obtained by interpolating the spatial trend estimate of Tables 13.7-13.9.

2. First obtain the tidal series for the study site by using the Proudman Oceanographic Laboratory software for the prediction of the tidal series at any east coast site. Combining this series with the spatial models for the interaction functions and extremal parameters developed in Sections 12.2-12.4 gives the information necessary to derive the return levels of interest.

3. Use the observed hourly data to perform a standard tidal analysis and prediction for the site. The method in point 2 can be used with the spatially predicted tide replaced by the predictions from the site data.

In applications where the Proudman Oceanographic Laboratory's spatial prediction of tides is inadequate, for example in complex estuary regions, the use of tides predicted from data at the site of interest is vital, so the third approach is best in these cases. When the tidal processes or return levels change rapidly and non-linearly along a coastal stretch then either the second or the third approach are required. However, the first approach is quite reasonable for most applications, and is the only possibility for those without access to additional information than contained within this report. Therefore we shall adopt the first approach to compare our spatial return level and trend estimates for the data sites. The spatial model estimates for return levels are given in Tables 13.11 and 13.12, and for trends in Table 13.10. Furthermore Figures 16.1-16.41 of the Appendix show the spatial return level curves together with the best site-by-site estimates.

Site	Trend	s.e.
Wick	3.220	0.4675
Aberdeen	0.626	0.1741
Leith	0.818	0.4000
North Shields	2.184	0.1454
Whitby	3.730	0.9882
Immingham	3.773	0.3774
Cromer	1.948	0.5328
Lowestoft	1.012	0.5328
Felixstowe	1.541	0.3636
Harwich	1.613	0.3118
Walton on the Naze	1.727	0.2429
Southend	1.975	0.1547
Sheerness	2.138	0.1285
Dover	2.408	0.1921

Table 13.10: Spatial trend estimates (in mm/year) and standard errors at the east coast study sites

In Section 12.2 and Chapter 13 we have already made a general comparison between the site-by-site and spatial model estimates for the trends and return levels respectively, so little remains to be said generally. Specific comments based on Tables 5.1 and 13.10 concerning trend estimation are:

- the spatial estimate has smaller standard errors than the site-by-site estimates,
- the spatial estimate provides trend estimates for Cromer and Felixstowe, neither of which were estimable using site-by-site methods,
- poor site-by-site estimates, such as for Whitby and Walton, have been improved by use of the spatial estimate.

From Tables 5.2, 5.4, 13.11 and 13.12 (but most easily seen through Figures 16.1-16.41) we see that

- For the majority of the long record sites (Wick, Aberdeen, North Shields, Immingham and Sheerness) there is a good agreement between the spatial and site-by-site estimates,
- For the other long record sites the agreement is reasonable. For Southend and Dover there is a discrepancy at short return periods, presumably due to the use

Site	10	25	50	100
Wick	4.25	4.35	4.40	4.47
Aberdeen	5.07	5.18	5.24	5.33
Leith	6.32	6.46	6.52	6.60
North Shields	5.88	6.01	6.08	6.19
Whitby	6.48	6.62	6.70	6.83
Immingham	8.31	8.49	8.60	8.81
Cromer	6.28	6.52	6.66	6.91
Lowestoft	4.04	4.25	4.38	4.59
Felixstowe	5.09	5.28	5.38	5.56
Harwich	5.16	5.35	5.46	5.63
Walton on the Naze	5.29	5.48	5.58	5.75
Southend	7.10	7.28	7.38	7.55
Sheerness	6.67	6.84	6.94	7.10
Dover	7.91	8.07	8.17	8.33

Table 13.11: Return levels for the study sites, in metres relative to ACD, for 1990. Estimates are obtained from the spatial model by interpolating the grid site estimates. Given for return periods of 10, 25, 50, and 100 years.

of the spatial tidal estimate, whereas for Lowestoft the agreement for short return periods is excellent but for long return periods is poor. The reason for this is unclear, although it could be that either the site-by-site estimate is over-estimating the return levels due to a sampling bias in the data period (1953 onwards), or that the spatial model under-estimates in the region around Lowestoft due to over-smoothing of the spatial estimate of the surge variability (see Figure 12.8),

- For sites with moderate records (Leith, Whitby, Harwich and Walton) there is a good agreement in the estimates,

- For the short record sites (Cromer and Felixstowe) the spatial estimate is much better than the site-by-site estimate from the JPM. However, within the critical region of 100-1000 year return periods the estimates differ by less than 50cm.

Briefly, in conclusion, our spatial model performs well for the estimation of return levels at study sites as well as providing the only possible estimation method for intermediate sites with no data. The good agreement we have found between the spatial and site-by-site estimates at the study sites is expected for return level estimates. The real benefit of

Site	250	500	1000	10000
Wick	4.53	4.57	4.61	4.73
Aberdeen	5.42	5.47	5.54	5.74
Leith	6.68	6.72	6.78	6.95
North Shields	6.30	6.36	6.43	6.67
Whitby	6.97	7.05	7.16	7.51
Immingham	9.03	9.18	9.38	10.17
Cromer	7.17	7.34	7.56	8.35
Lowestoft	4.80	4.92	5.09	5.62
Felixstowe	5.74	5.84	5.97	6.40
Harwich	5.81	5.91	6.04	6.46
Walton on the Naze	5.92	6.02	6.15	6.56
Southend	7.71	7.81	7.93	8.33
Sheerness	7.26	7.35	7.47	7.85
Dover	8.48	8.57	8.68	9.05

Table 13.12: Return levels for the study sites, in metres relative to ACD, for 1990. Estimates are obtained from the spatial model by interpolating the grid site estimates. Given for return periods of 250, 500, 1000, and 10000 years.

the spatial approach for the study sites is in terms of trends, as the spatial estimate is a substantial improvement over the previously available estimates, and for the estimation of long period return levels for sites with short records.

Chapter 14

Acknowledgements

We thank the Ministry of Agriculture Fisheries and Food for funding this work. We have found considerable support from colleagues and POL staff. In particular at POL, we thank Ian Vassie who made a large contribution in the development of the POL tidal interpolation software, and provided much helpful and highly appreciated advice for other aspects of the work. Sheila Shaw and David Blackman have been most helpful in supplying data and answering questions concerning datum, data availability; and Graham Alcock has always been very helpful in giving advice and support.

Finally at Lancaster University we gratefully acknowledge the help that Barry Rowlingson has provided through advice in computing throughout the project and aspects concerning the preparation of the report.

Chapter 15

References

Barnett, T. P. (1984). The estimation of 'global' sea level change: A problem of uniqueness. *J. Geophys. Res.*, **89**, 7980–7988.

Blackman, D. L. and Graff, J. (1978). Analysis of maximum sea levels in Southern England. *Proc. 16th Coastal Engrg. Conf., Hamburg, 1978, Vol.1*, American Soc. Civ. Engrs., 931–947.

Chu, C.-K. and Marron, J. S. (1991). Choosing a kernel regression estimator. *Statist. Sci.*, **6**, 404–436.

Coles, S. G. and Tawn, J. A. (1990). Statistics of coastal flood prevention. *Phil. Trans. R. Soc. Lond., A*, **332**, 457–476.

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *J. R. Statist. Soc., B,* **53**, 377–392.

Coles, S. G. and Tawn, J. A. (1995). Modelling extremes of the areal rainfall process. To appear in *J. R. Statist. Soc., B,*

Coles, S. G. and Walshaw, D. (1994). Directional modelling of extreme wind speeds. *Appl. Statist.*, **43**, 139–158.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics.* Chapman and Hall, London.

Cressie, N. (1991). *Statistics for Spatial Data.* John Wiley and Sons.

David, H. A. (1981). *Order Statistics*, 2nd ed. New York: Wiley.

Dixon, M. J. and Tawn, J. A. (1994). *Estimates Of Extreme Sea Conditions: Extreme Sea-Levels At The UK A-Class Sites: Site-By-Site Analyses.* Proudman Oceanographic Laboratory Internal document No. **65**.

Douglas, B. C. (1992). Global sea-level acceleration. *J. Geophys. Res.*, **97**, 12699–12706.

Flather R. A. and Smith, J. (1993). Recent progress with storm surge models - results for January and February 1993. In *Proc. of MAFF Conf. of River and Coastal Engineers.* University of Loughborough 5-7 July 1993.

Gelfand, A. and Smith, A. F. M. (1990). Sampling-based approaches to calculating

marginal densities. *J. Amer. Statist. Assoc.*, **85**, 398–409.

Gornitz, V. and Lebedeff, S. (1987). Global sea-level changes during the last century. In *Sea-Level Change and Coastal Evolution*, No. 41 pp3–16, eds Nummedal, D., Pilkey, O. H. and Howard, J. D., Society for Economic Paleontologists and Mineralogists, Tulsa, Oklahoma, SEPM Special Publication.

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.

Hardle, W. (1989). *Applied Nonparametric Regression.* Cambridge University Press.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalised Additive Models.* London, Chapman and Hall.

Intergovernmental Panel on Climate Change (IPCC) (1990). *The IPCC Scientific Assessment*, eds. Houghton, J. T., Jenkins, G. J. and Ephraums, J. J., Cambridge University Press, Cambridge.

Johnston, J. (1984). *Econometric Methods*, 3rd ed., McGraw-Hill, New York.

Peltier, W. R. and Tushingham, A. M. (1989). Global sea level rise and the greenhouse effect: might they be connected?. *Science*, **244**, 806–810.

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

Pugh, D. T. (1987). *Tides, Surges and Mean Sea Level: a Handbook for Engineers and Scientists.* Wiley, Chichester.

Pugh, D. T., Spencer, N. E. and Woodworth, P. L. (1987). *Data Holdings of the Permanent Service for Mean Sea-Level*, Permanent Service for Mean Sea-Level, Bidston, pp156.

Pugh, D. T. and Vassie, J. M. (1976). Tide and surge propagation off-shore in the Dowsing region of the North Sea. *Dt. hydrog. Z.*, **29**, 163–213.

Pugh, D. T. and Vassie, J. M. (1979). Extreme sea-levels from tide and surge probability. Proceedings 16th Coastal Engineering Conference, 1978, Hamburg. *American Society of Civil Engineers, New York*, **1**, 911–930.

Pugh, D. T. and Vassie, J. M. (1980). Applications of the joint probability method for extreme sea level computations. *Proc. Instn. Civ. Engrs., Part 2*, **69**, 959–975.

Rawlings, J. O. (1988). *Applied Regression Analysis: A Research Tool.* Wadsworth and Brooks/Cole.

Robin, G. de Q. (1986). Changing the sea level. In *The Greenhouse Effect, Climatic Change and Ecosystems*, eds Bolin, B., Doos, J., Jager J., and Warrick, R. A., SCOPE 29.

Rossiter, J. R. (1969). Tidal regime in the Thames. *The Dock and Harbour Authority*, **49**, 461-462.

Shennan, I. (1989). Holocene crustal movements and sea-level changes in Great Britain.

J. Quart. Sci., **4**, 77–89.

Shennan, I. and Woodworth, P. L. (1992). A comparison of late Holocene and twentieth century sea-level trends from the U.K. and North sea region. To be published in *Geophys. J. Int.*.

Silverman, B. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J. R. Statist. Soc, B*, **47**, 1–52.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

Smith, A. F. M. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika*, **62**, 407–416.

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

Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend estimation in ground level ozone. *Statist. Sci.*,**4**, 367–393.

Smith, R. L. (1994). Regional estimation from spatially dependent data. Submitted to *Biometrika.*

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

Tawn J. A. (1992). Estimating probabilities of extreme sea levels. *Appl. Statist.*, **41**, 77–93.

Tawn, J. A., Dixon, M. J., and Woodworth, P. L. (1994). Trends in sea-levels, in *Statistics for the Environment 2: Water Related Issues*, 147–181, eds. V. Barnett and F. K. Turkman, Wiley: Chichester.

Tawn, J. A. and Mitchell, W. M. (1994). A spatial analysis of Australian extreme sea levels. Submitted.

Tawn, J. A. and Vassie, J. M. (1989). Extreme sea levels: the joint probabilities method revistied and revised. *Proc. Istn. Civ. Engrs., Part 2*, **87**, 429–442.

Thompson, K. R. (1980). An analysis of British monthly mean sea-levels. *Geophys. J. R. astr. Soc.*, **63**, 57–73.

Trupin, A. and Wahr, J. (1990). Spectroscopic analysis of global tide gauge sea level data. *Geophys. J. Int.*, **100**, 441–453.

Walden, A. T. Prescott, P. and Webber, N. B. (1982). The examination of surge–tide interaction at two ports on the central south coast of England. *Coastal Eng.*, **6**, 59–70.

Wigley, T. M. L. and Raper, S. C. B. (1992). Implications for climate and sea-level of revised IPCC scenarios. *Nature (London)*, **357**, 293–300.

Wolf, J. (1978). Interaction of tide and surge in a semi–infinite uniform channel, with an 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. Geo.*, **11**, 57–87.

Woodworth, P. L. (1990). A search for accelerations in records of European mean sea level. *Int. J. Clim.*, **10**, 129–143.

Woodworth, P. L. and Jarvis, J. (1991). A feasibility study of the use of short historical and short modern tide gauge records to investigate long term sea level changes in the British Isles. *Proudman Oceanographic Laboratory*, Internal document No. 23.

Woodworth, P. L., Shaw, S. M. and Blackman, D. L. (1991). Secular trends in mean tidal range around the British Isles and along the adjacent European coastline. *Geophys. J. Int.*, **104**, 593–609.

Woodworth, P. L., Spencer, N. E. and Alcock, G. A. (1990). On the availability of the European mean sea-level data. *Int. hyd. Rev.*, **67**, 131–146.

Zellner, A. (1962). Seemingly unrelated regression. *J. Amer. Statist. Ass.*, **57**, 348–368.

Zellner, A. (1963). Seemingly unrelated regression. *J. Amer. Statist. Ass.*, **58**, 977–992.

Zellner, A. (1972). Seemingly unrelated regression. *J. Amer. Statist. Ass.*, **67**, 225.

Chapter 16

Appendices

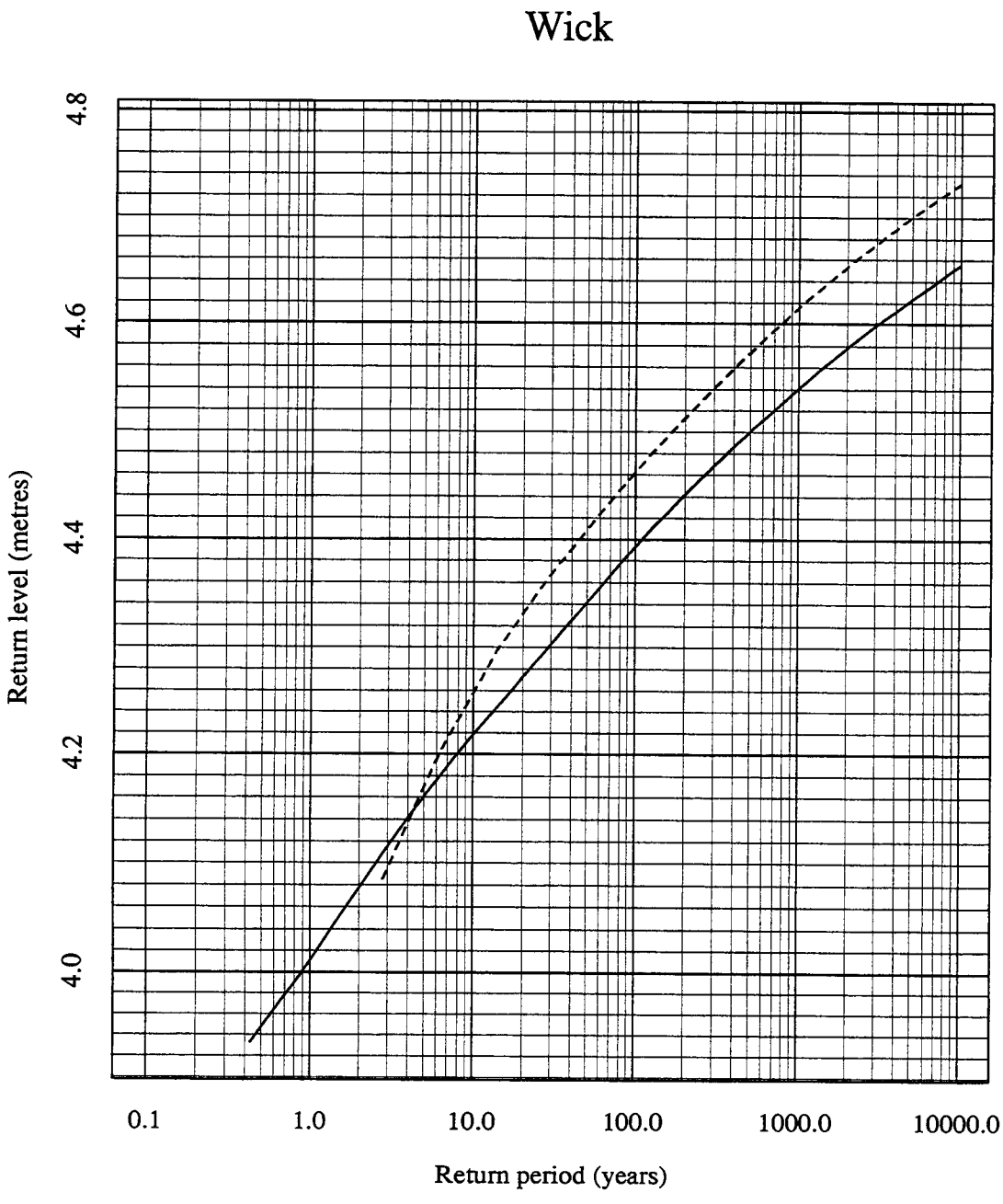

Figure 16.1: Port Diagram for the best site-by-site method (solid line) and the spatial model (broken line) for Wick relative to ACD in 1990.

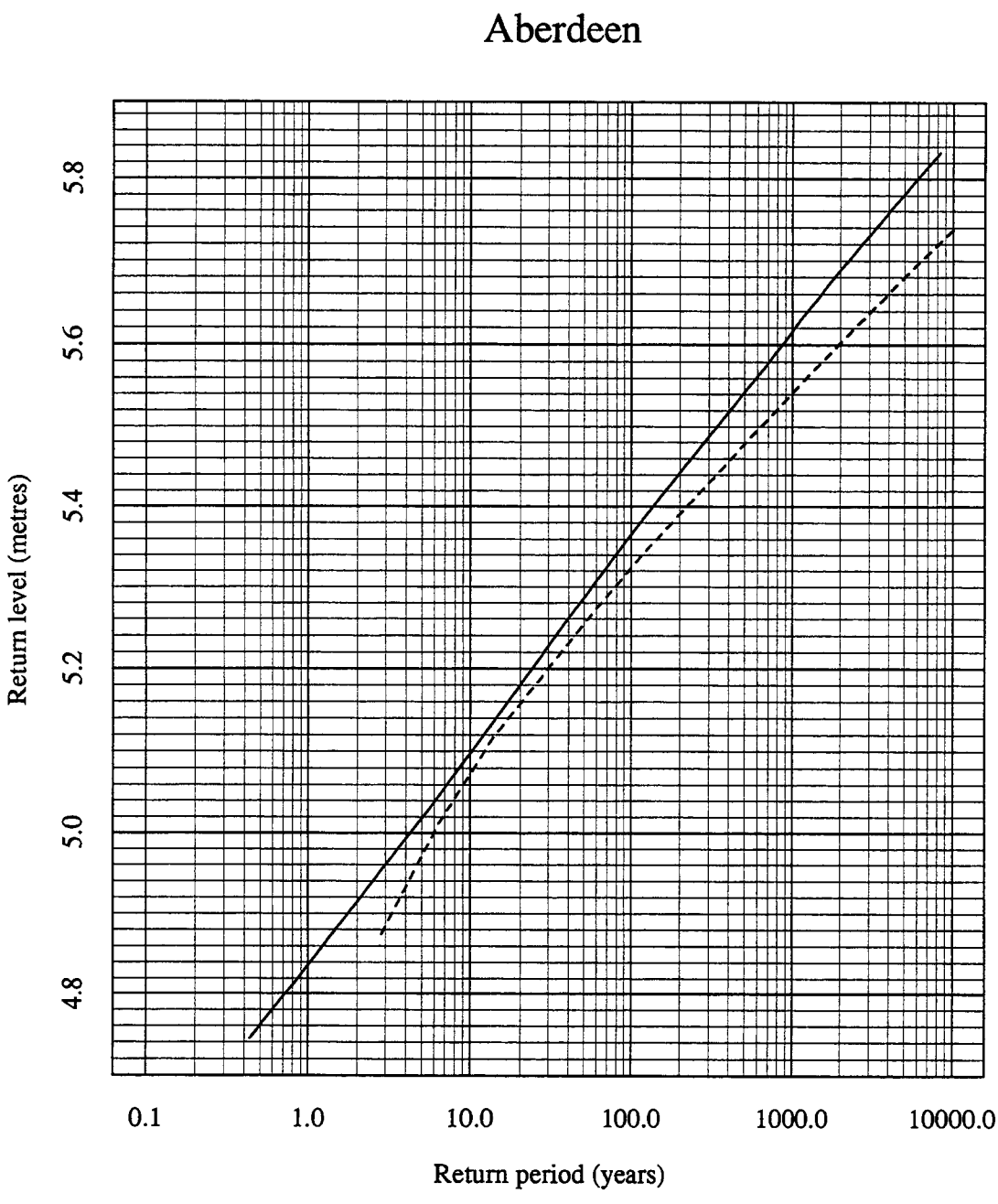

Figure 16.2: Port Diagram for the best site-by-site method (solid line) and the spatial model (broken line) for Aberdeen relative to ACD in 1990.

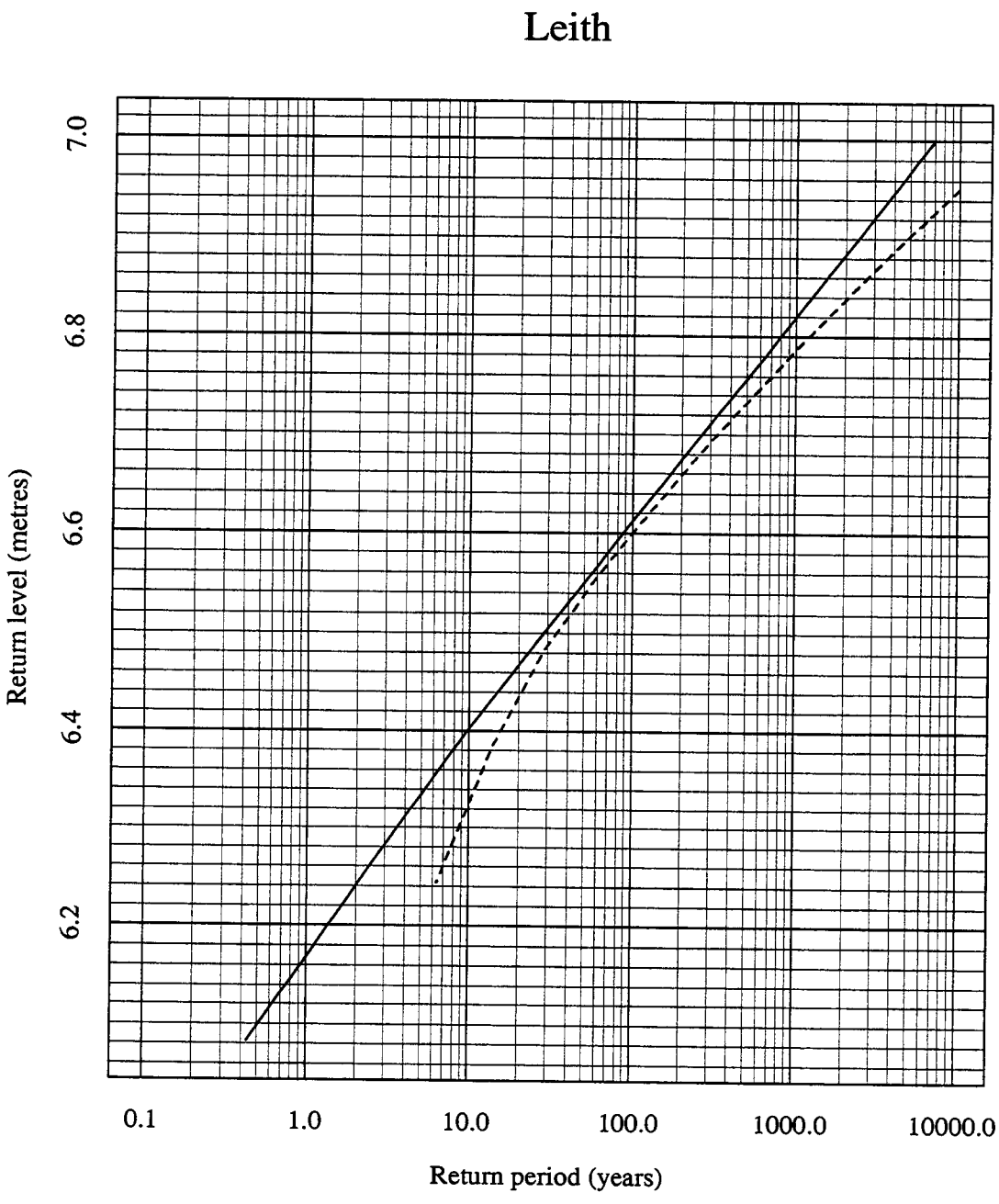

Figure 16.3: Port Diagram for the best site-by-site method (solid line) and the spatial model (broken line) for Leith relative to ACD in 1990.

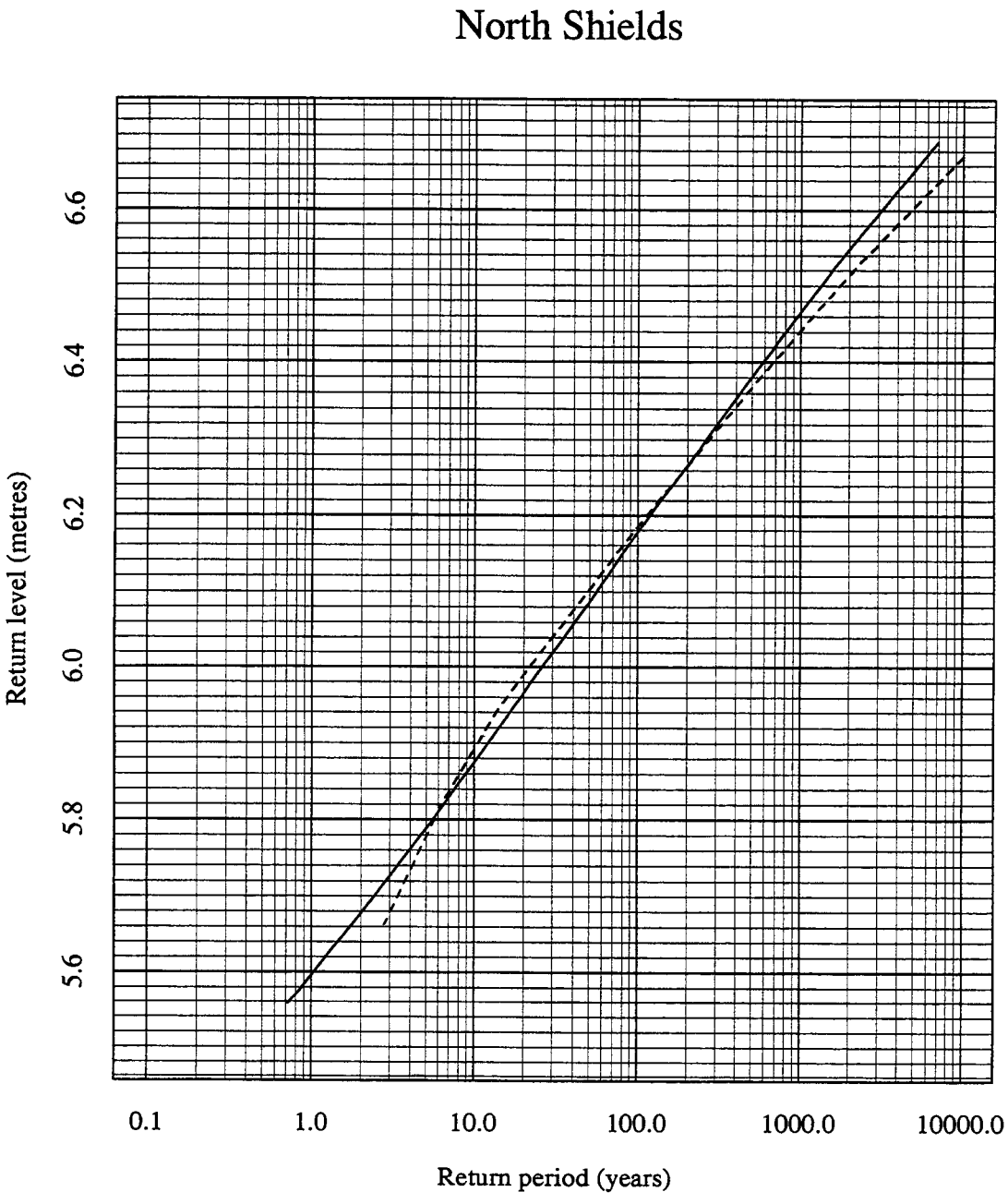

Figure 16.4: Port Diagram for the best site-by-site method (solid line) and the spatial model (broken line) for North Shields relative to ACD in 1990.

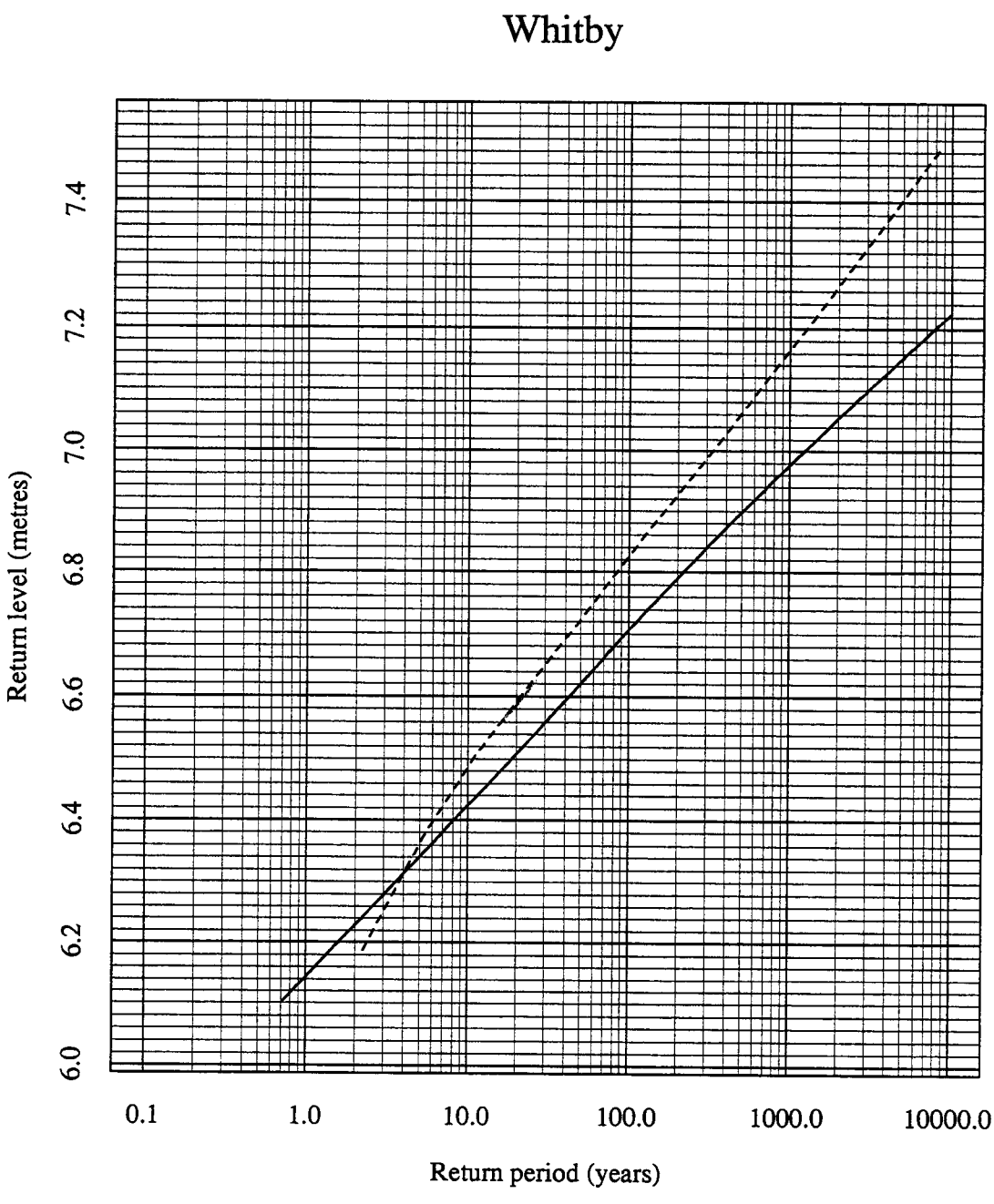

Figure 16.5: Port Diagram for the best site-by-site method (solid line) and the spatial model (broken line) for Whitby relative to ACD in 1990.

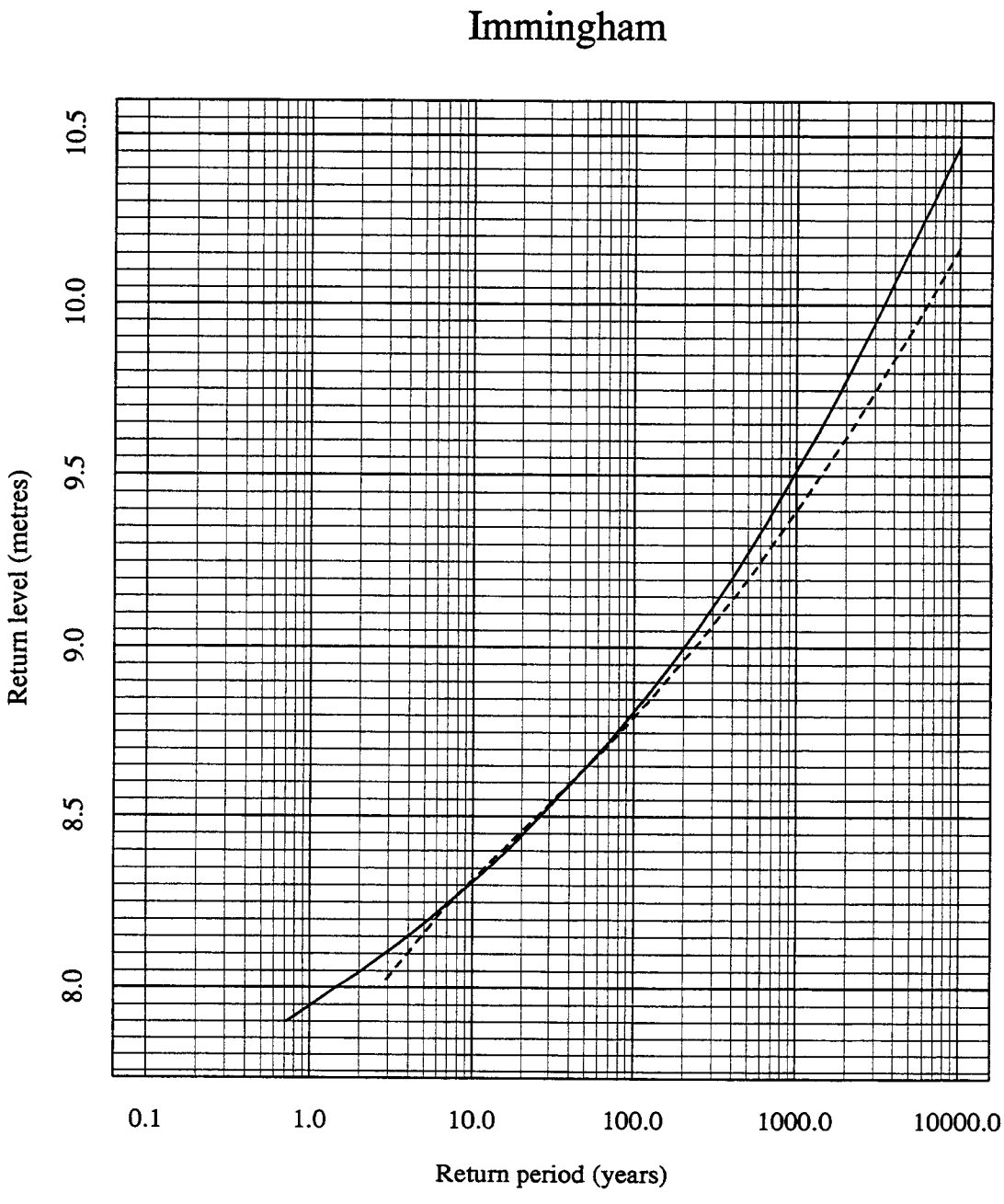

Figure 16.6: Port Diagram for the best site-by-site method (solid line) and the spatial model (broken line) for Immingham relative to ACD in 1990.

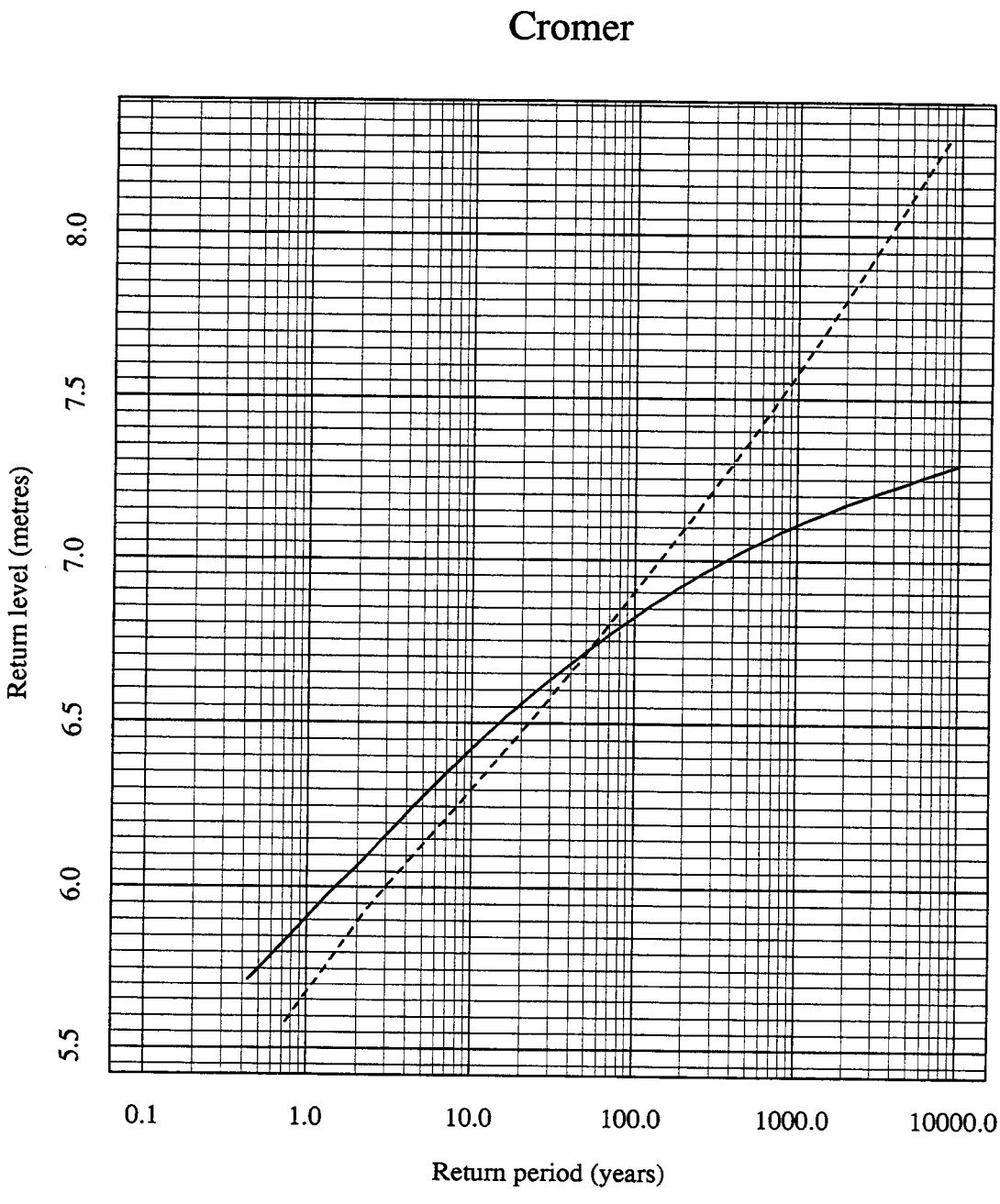

Figure 16.7: Port Diagram for the best site-by-site method (solid line) and the spatial model (broken line) for Cromer relative to ACD in 1990.

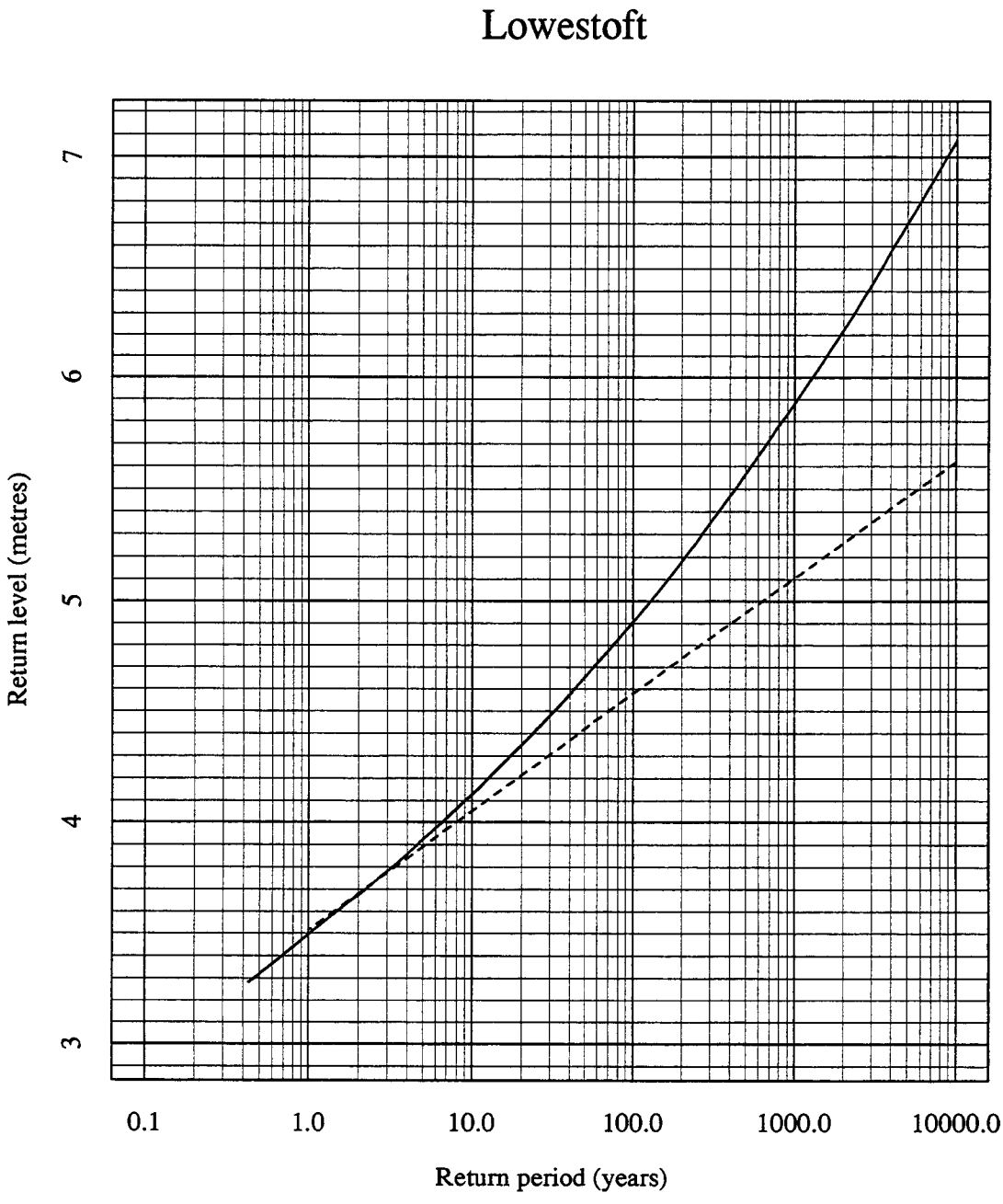

Figure 16.8: Port Diagram for the best site-by-site method (solid line) and the spatial model (broken line) for Lowestoft relative to ACD in 1990.

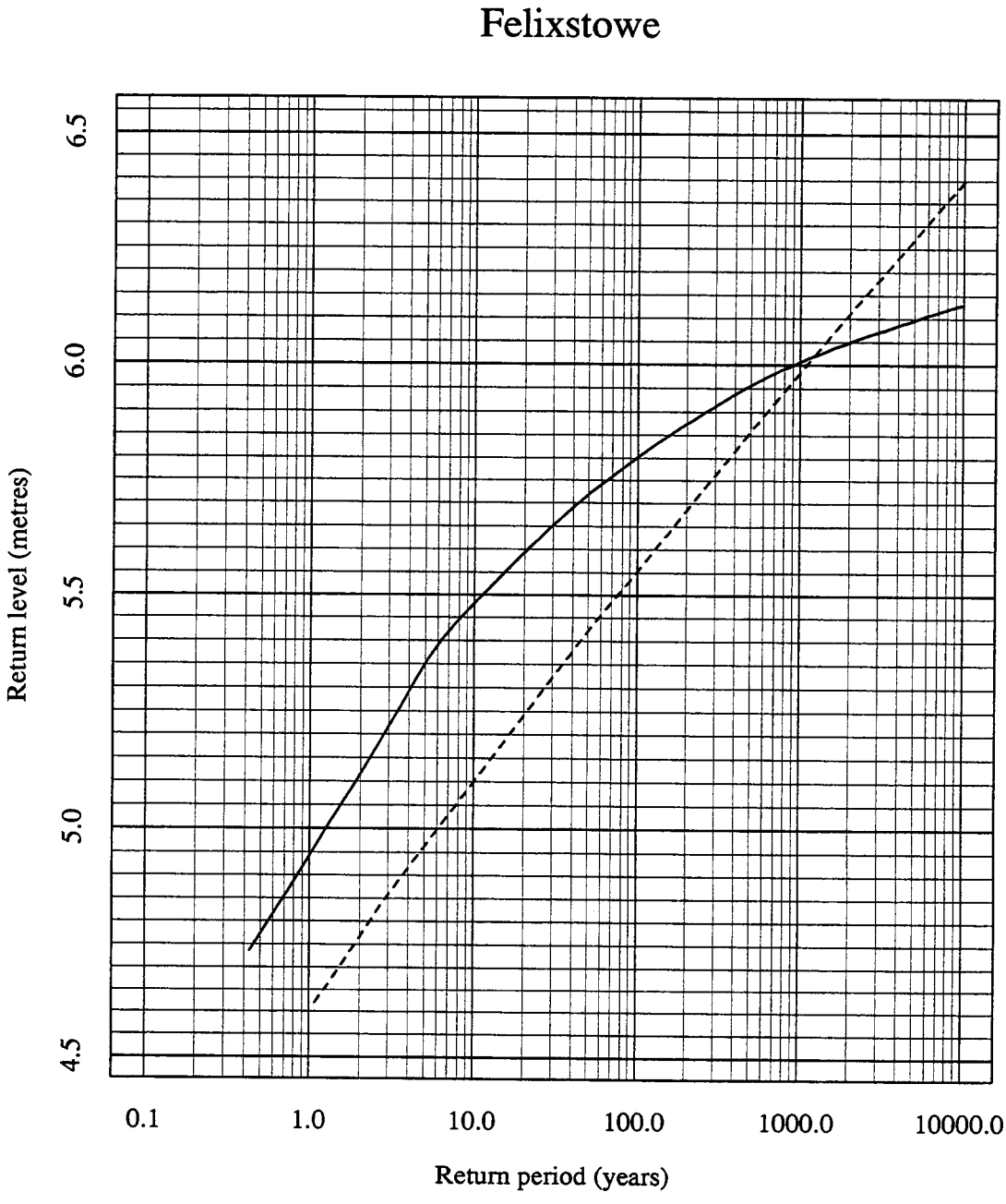

Figure 16.9: Port Diagram for the best site-by-site method (solid line) and the spatial model (broken line) for Felixstowe relative to ACD in 1990.

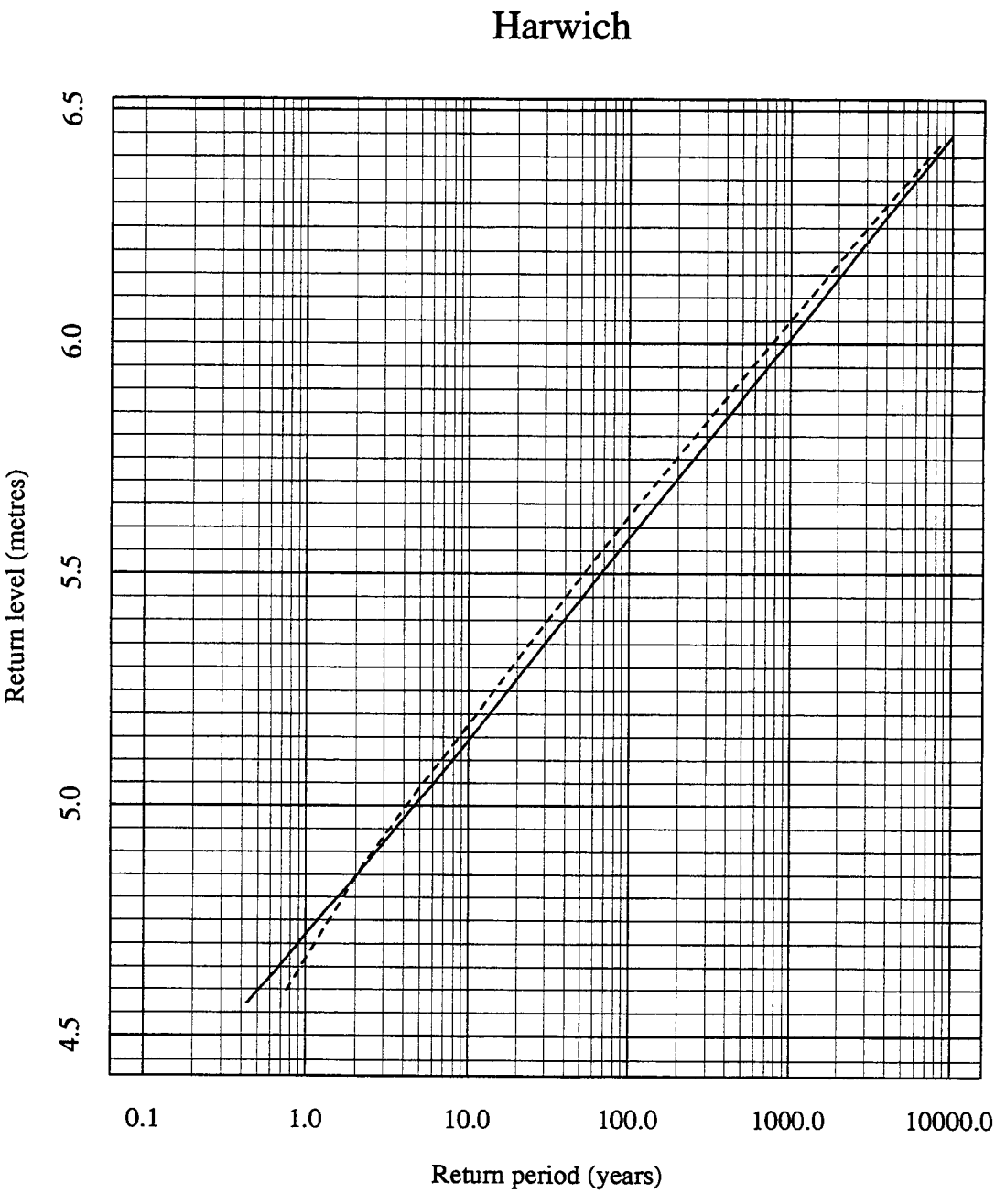

Figure 16.10: Port Diagram for the best site-by-site method (solid line) and the spatial model (broken line) for Harwich relative to ACD in 1990.

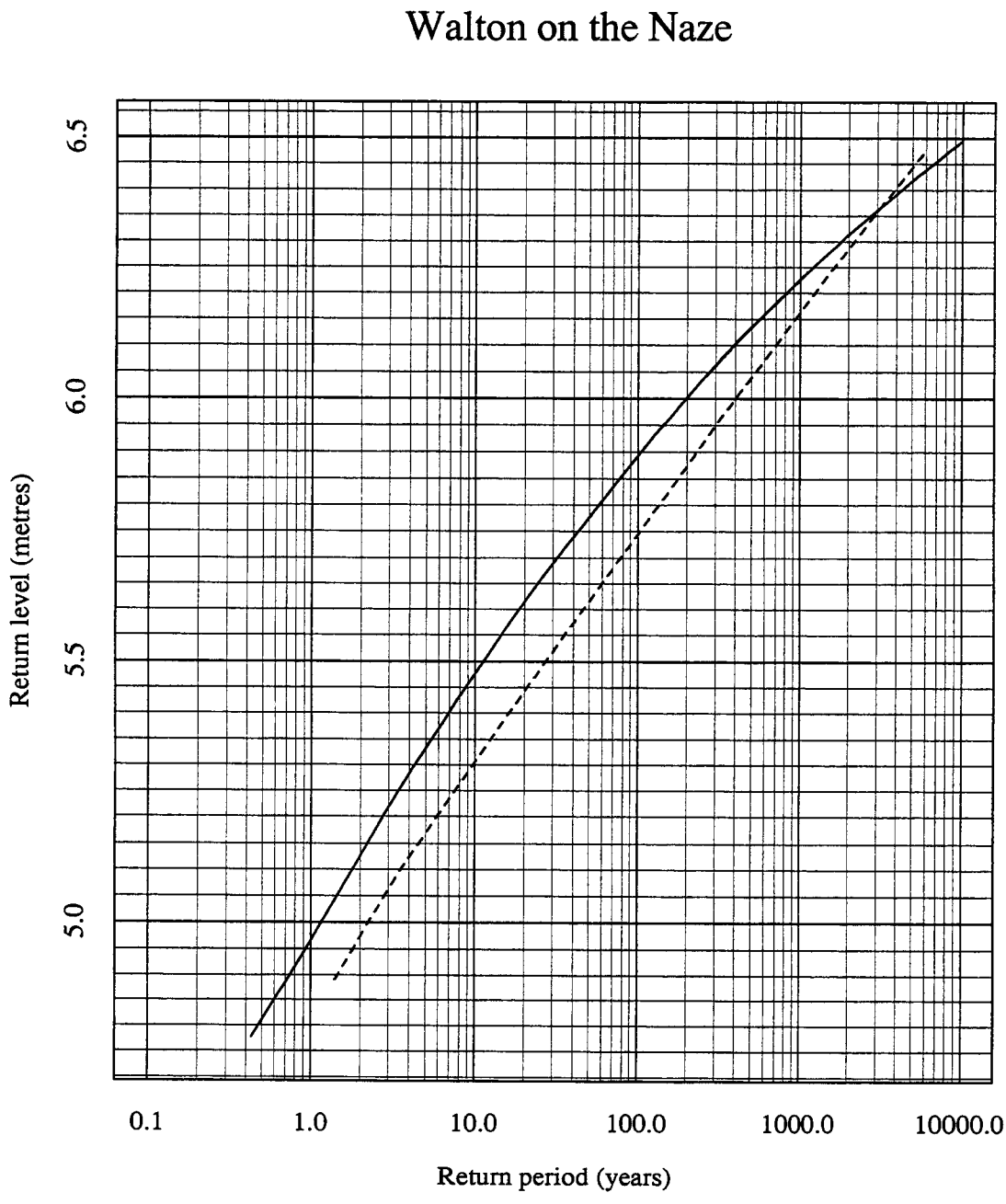

Figure 16.11: Port Diagram for the best site-by-site method (solid line) and the spatial model (broken line) for Walton on the Naze relative to ACD in 1990.

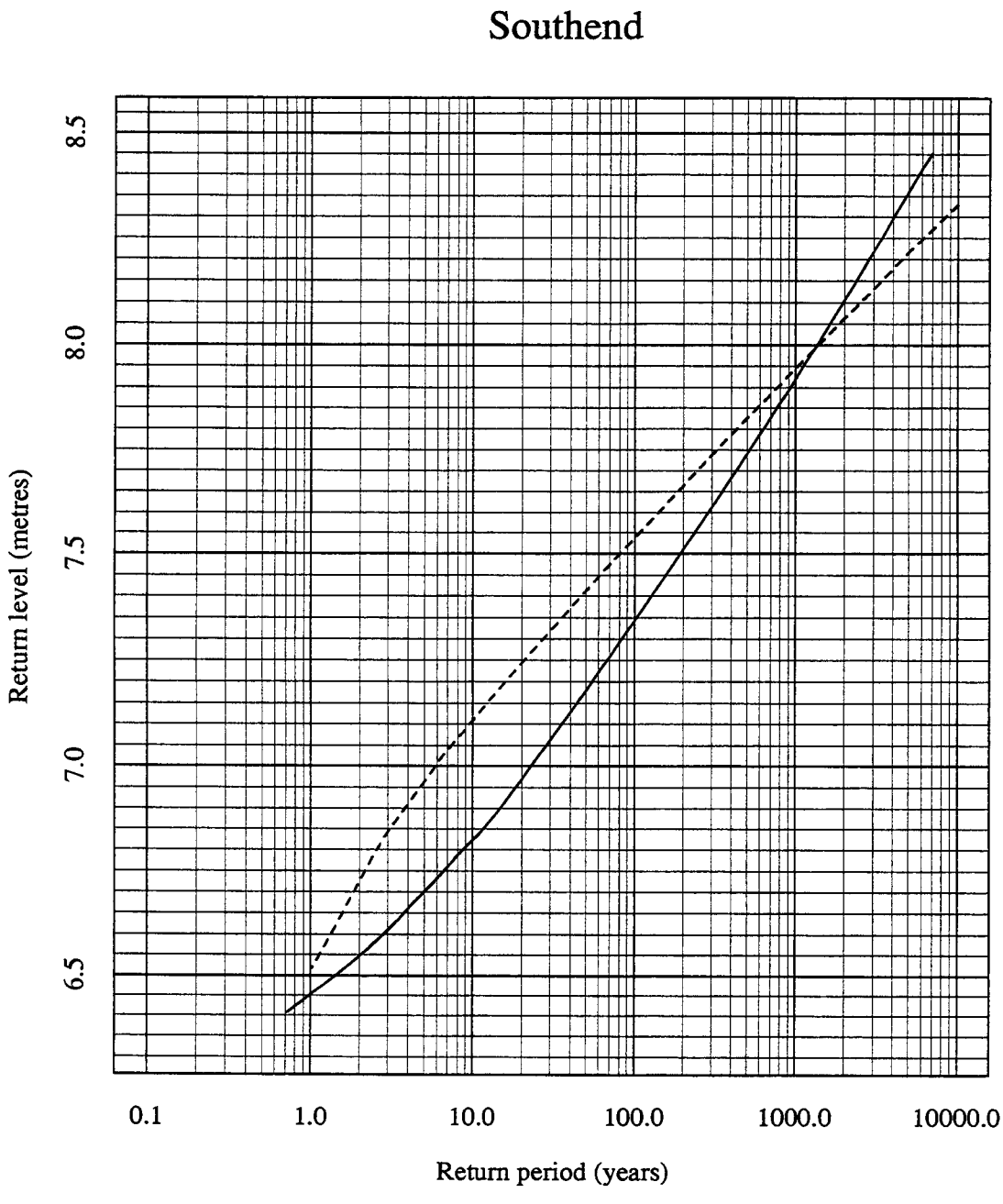

Figure 16.12: Port Diagram for the best site-by-site method (solid line) and the spatial model (broken line) for Southend relative to ACD in 1990.

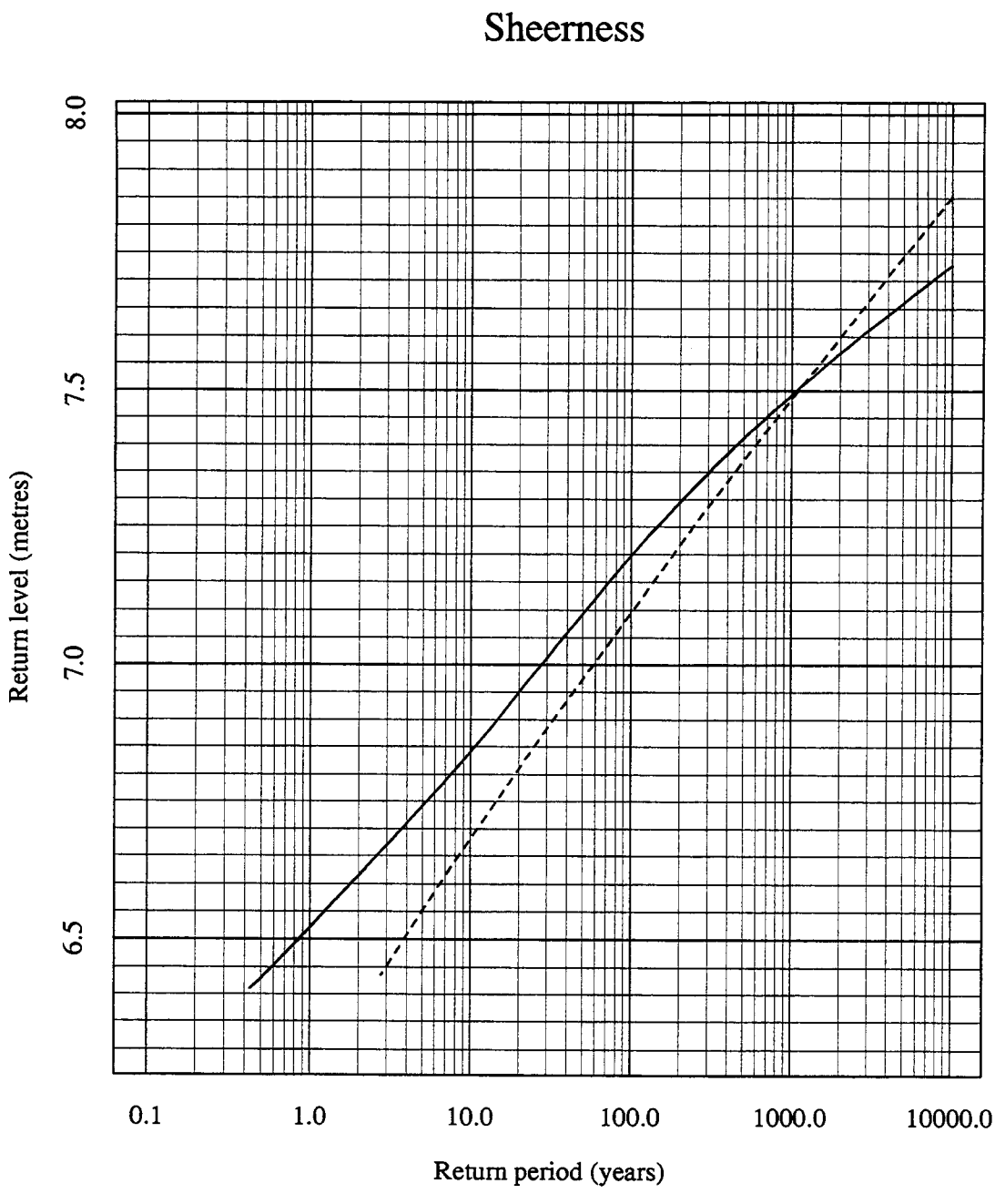

Figure 16.13: Port Diagram for the best site-by-site method (solid line) and the spatial model (broken line) for Sheerness relative to ACD in 1990.

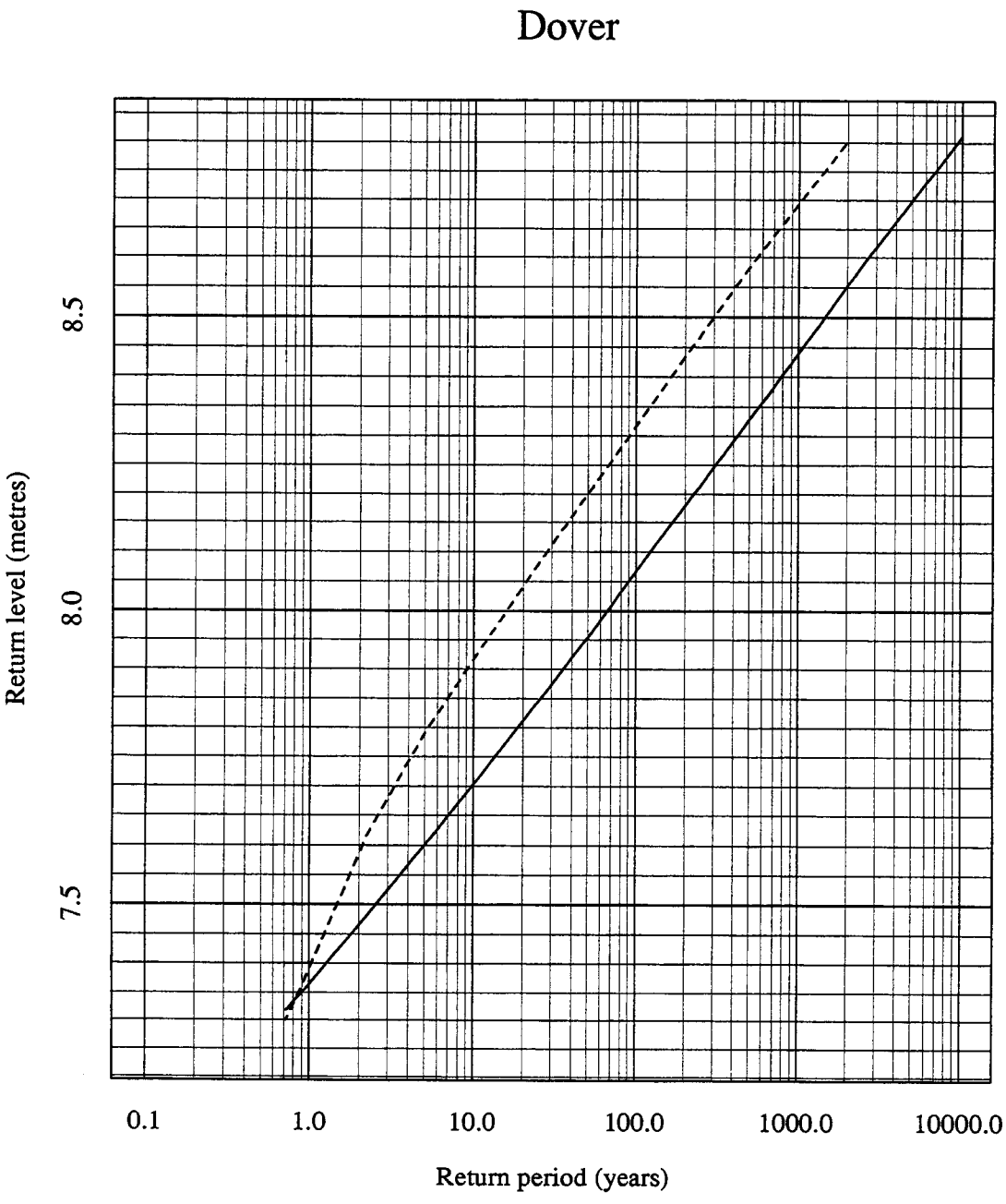

Figure 16.14: Port Diagram for the best site-by-site method (solid line) and the spatial model (broken line) for Dover relative to ACD in 1990.

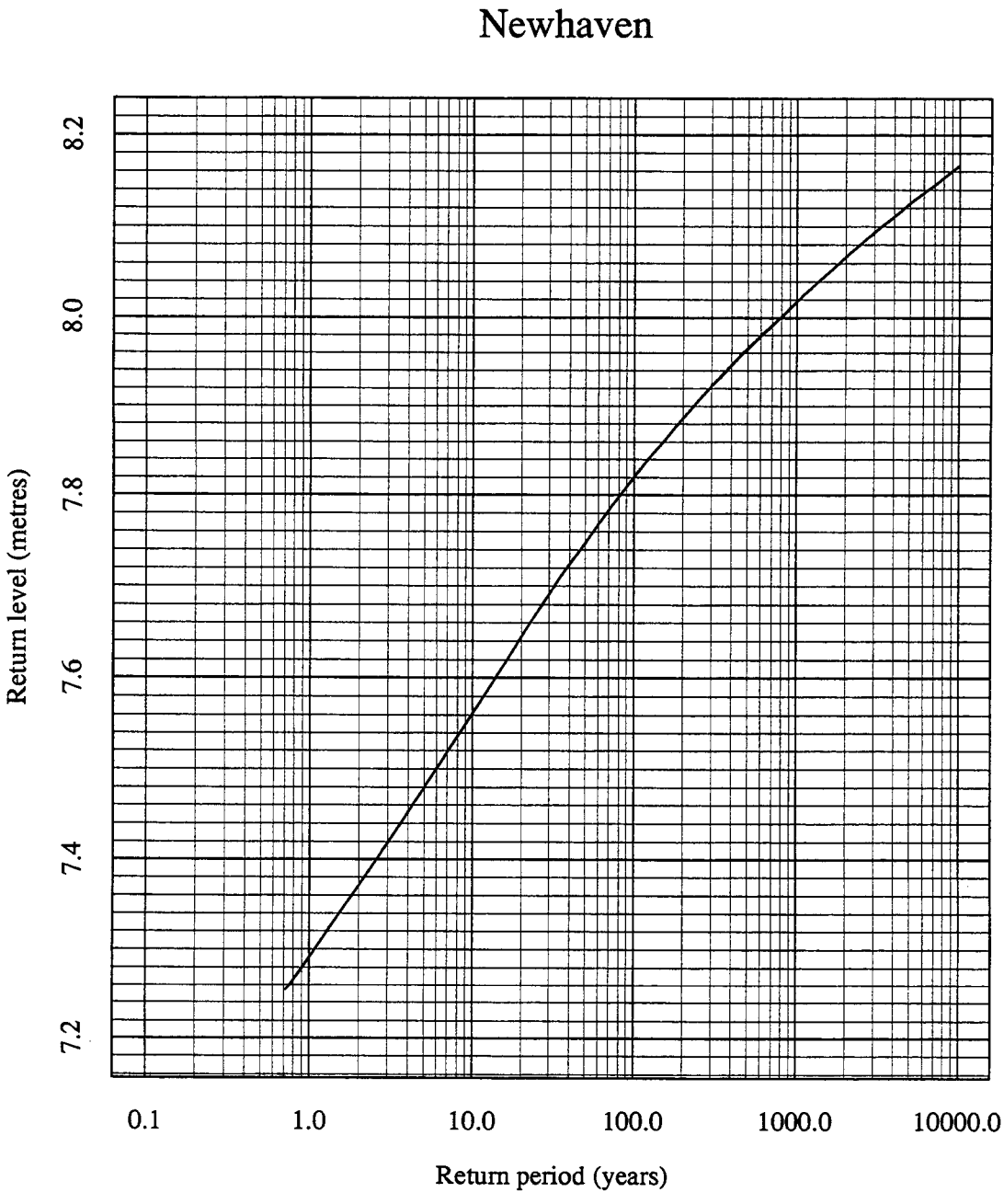

Figure 16.15: Port Diagram for the best site-by-site method for Newhaven relative to ACD in 1990.

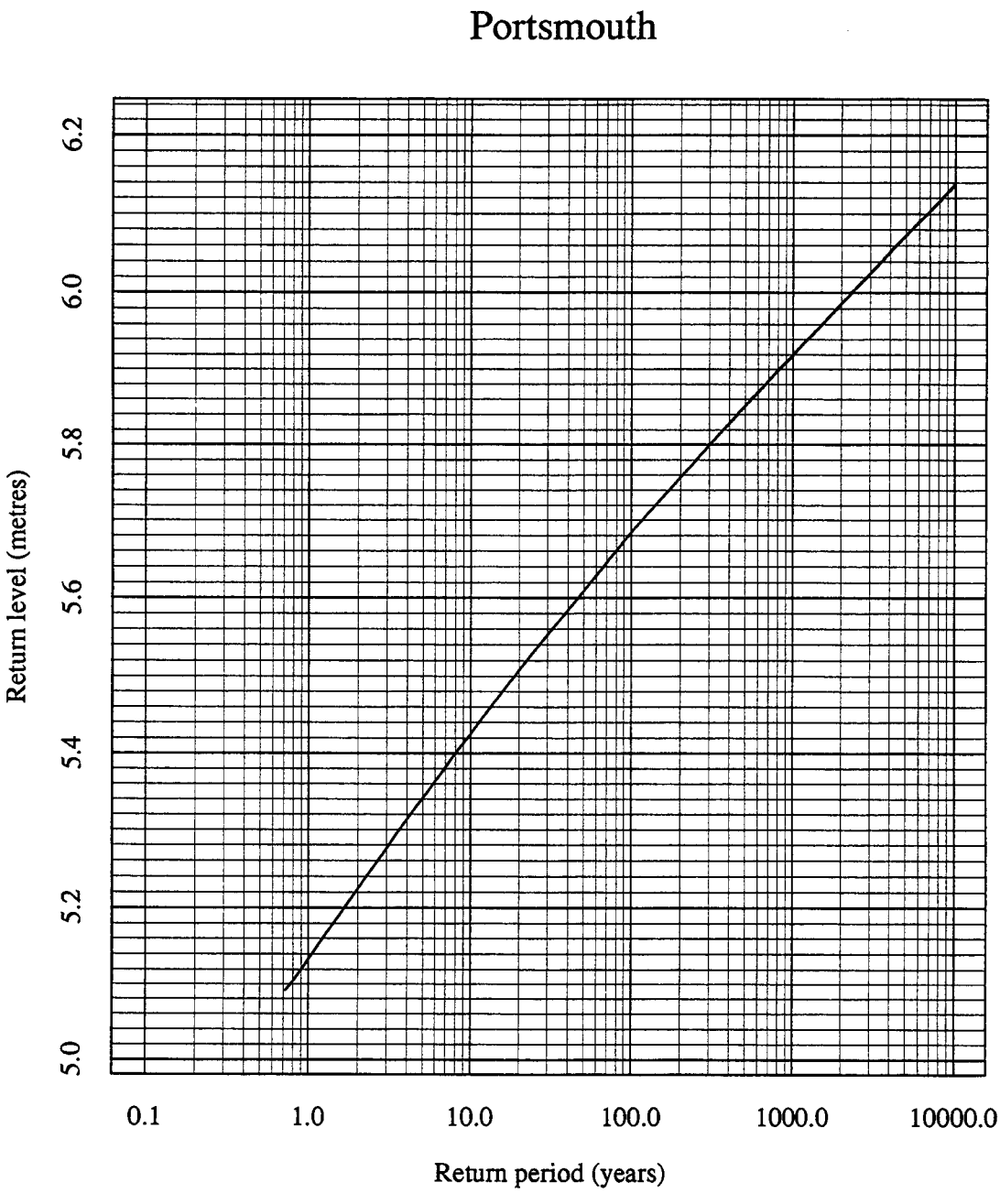

Figure 16.16: Port Diagram for the best site-by-site method for Portsmouth relative to ACD in 1990.

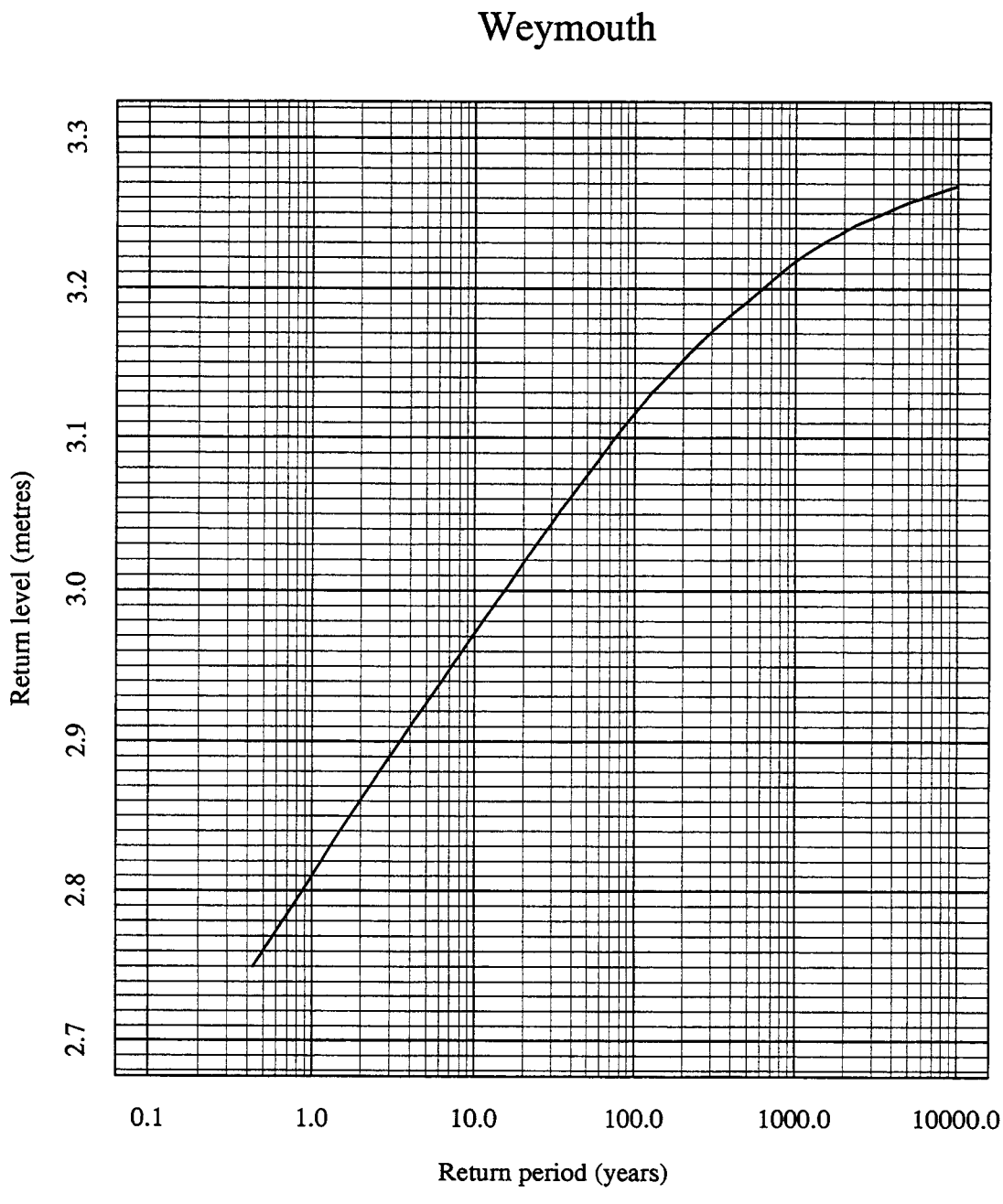

Figure 16.17: Port Diagram for the best site-by-site method for Weymouth relative to ACD in 1990.

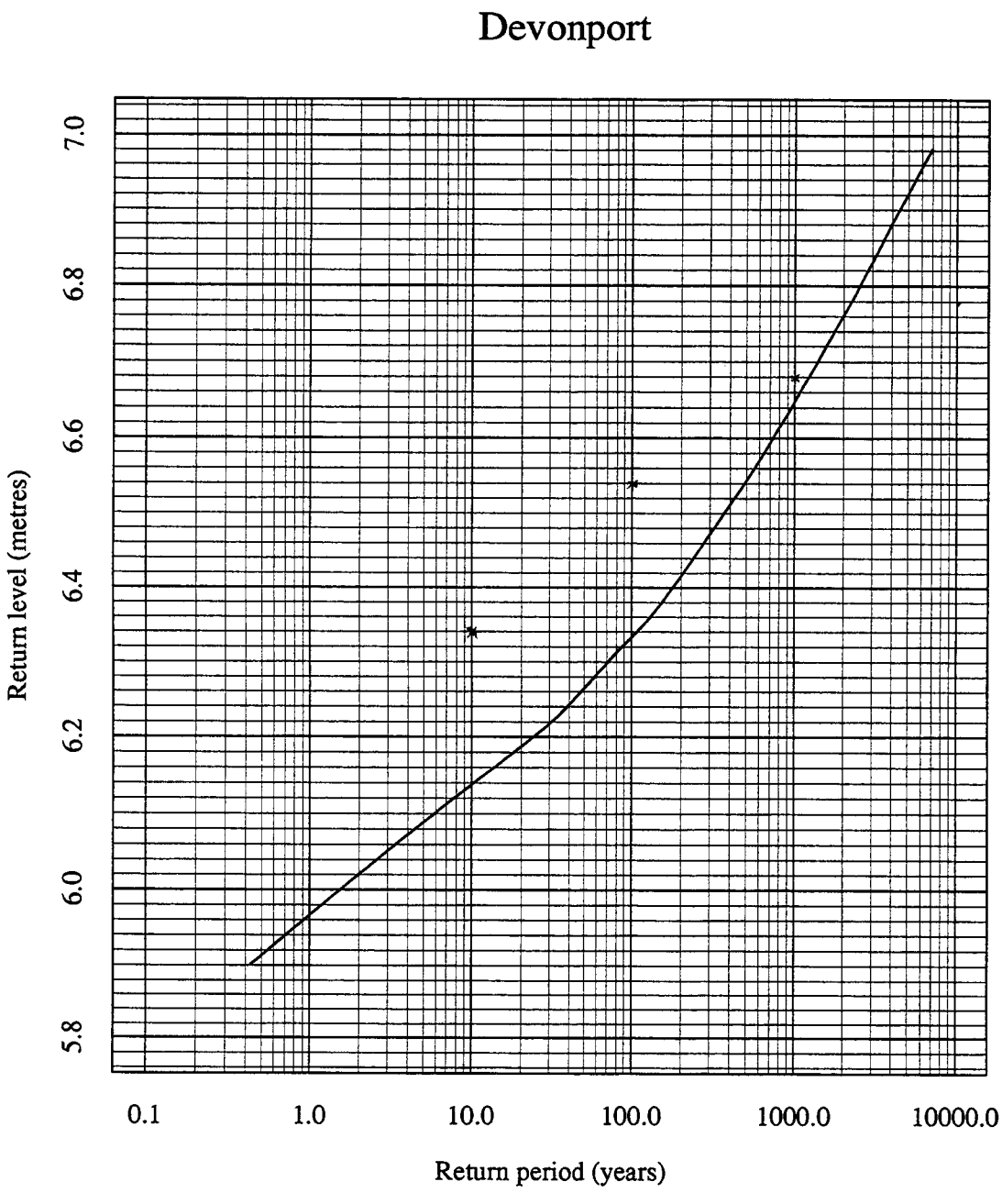

Figure 16.18: Port Diagram for the best site-by-site method for Devonport relative to ACD in 1990.

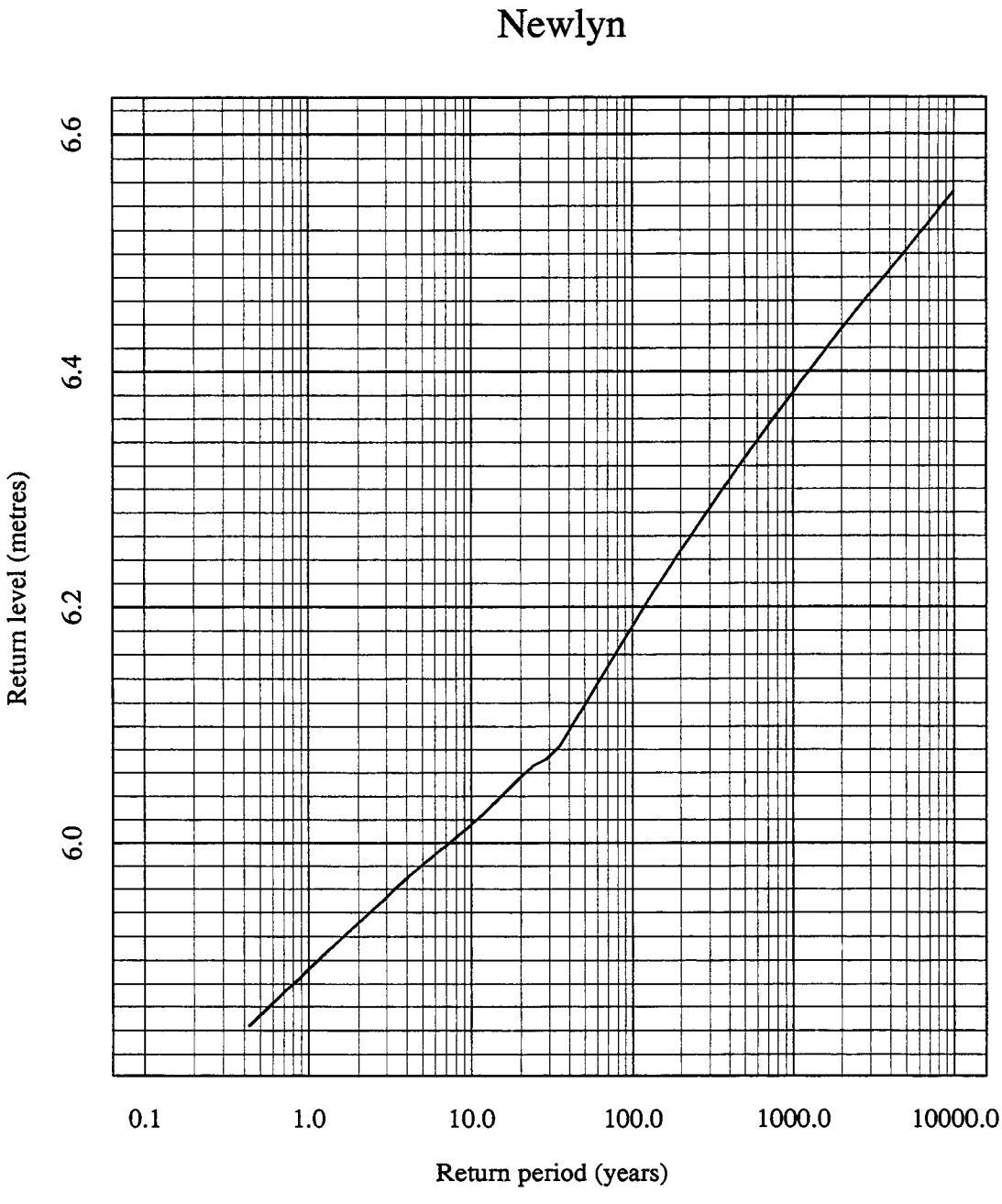

Figure 16.19: Port Diagram for the best site-by-site method for Newlyn relative to ACD in 1990.

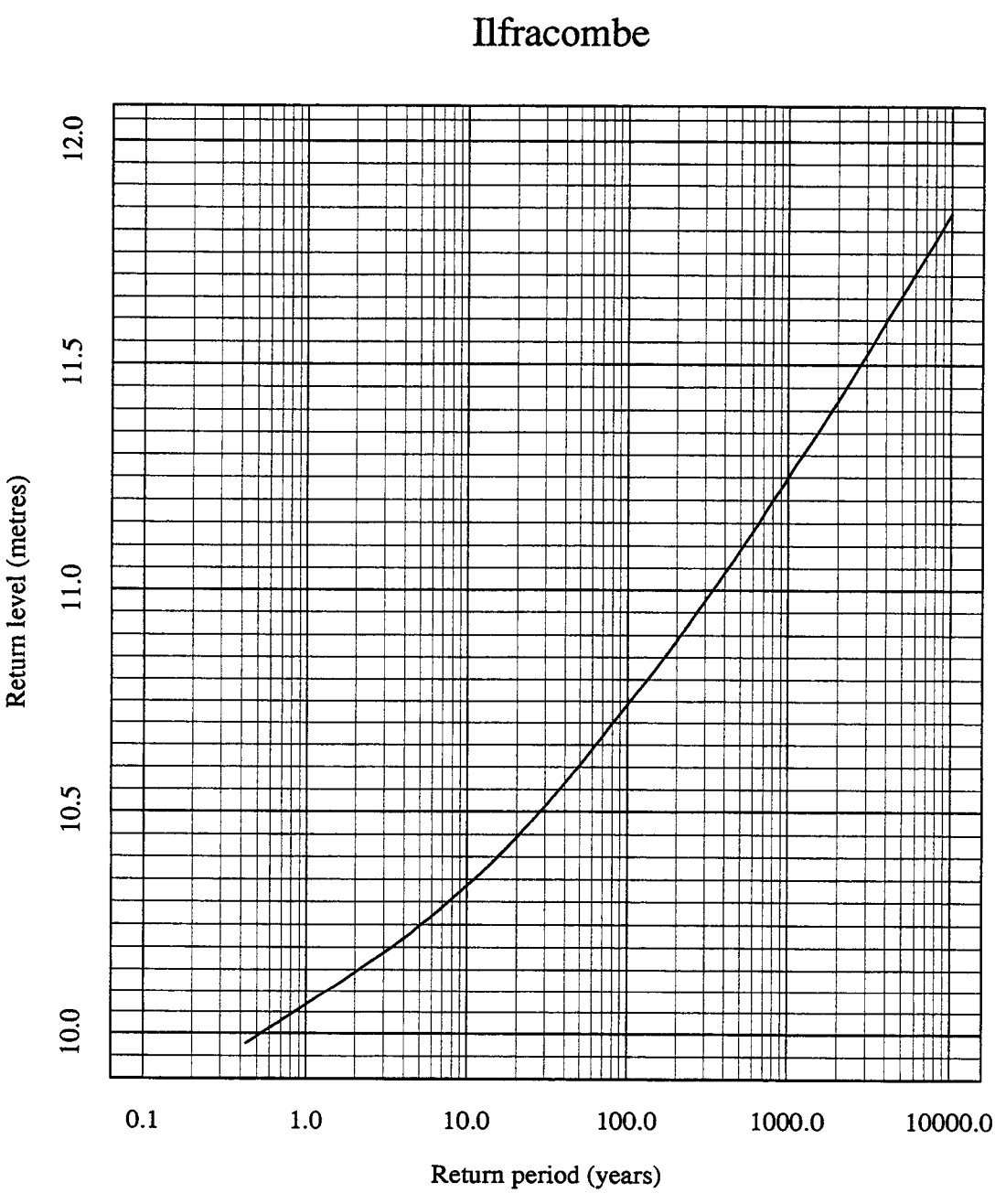

Figure 16.20: Port Diagram for the best site-by-site method for Ilfracombe relative to ACD in 1990.

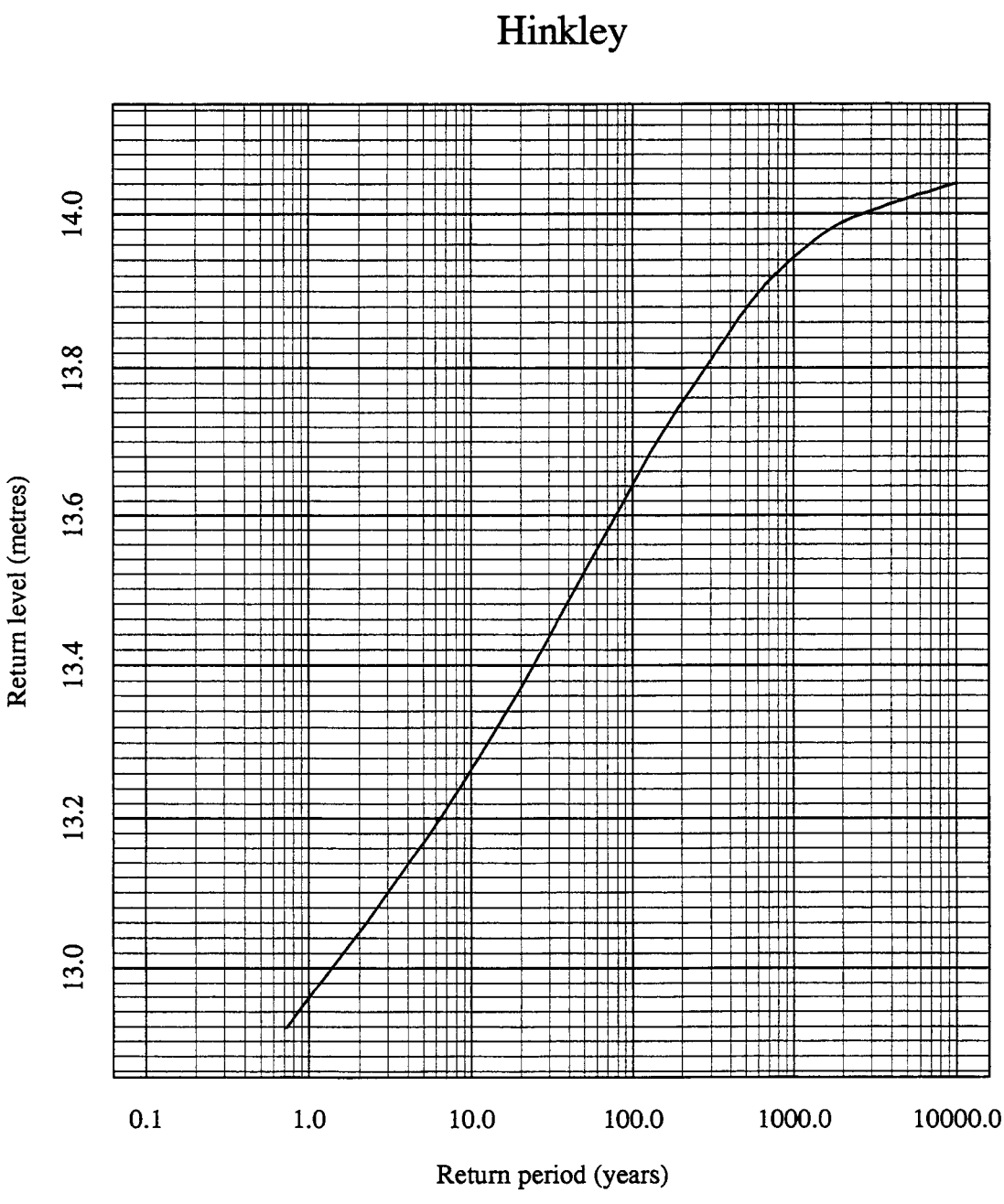

Figure 16.21: Port Diagram for the best site-by-site method for Hinkley relative to ACD in 1990.

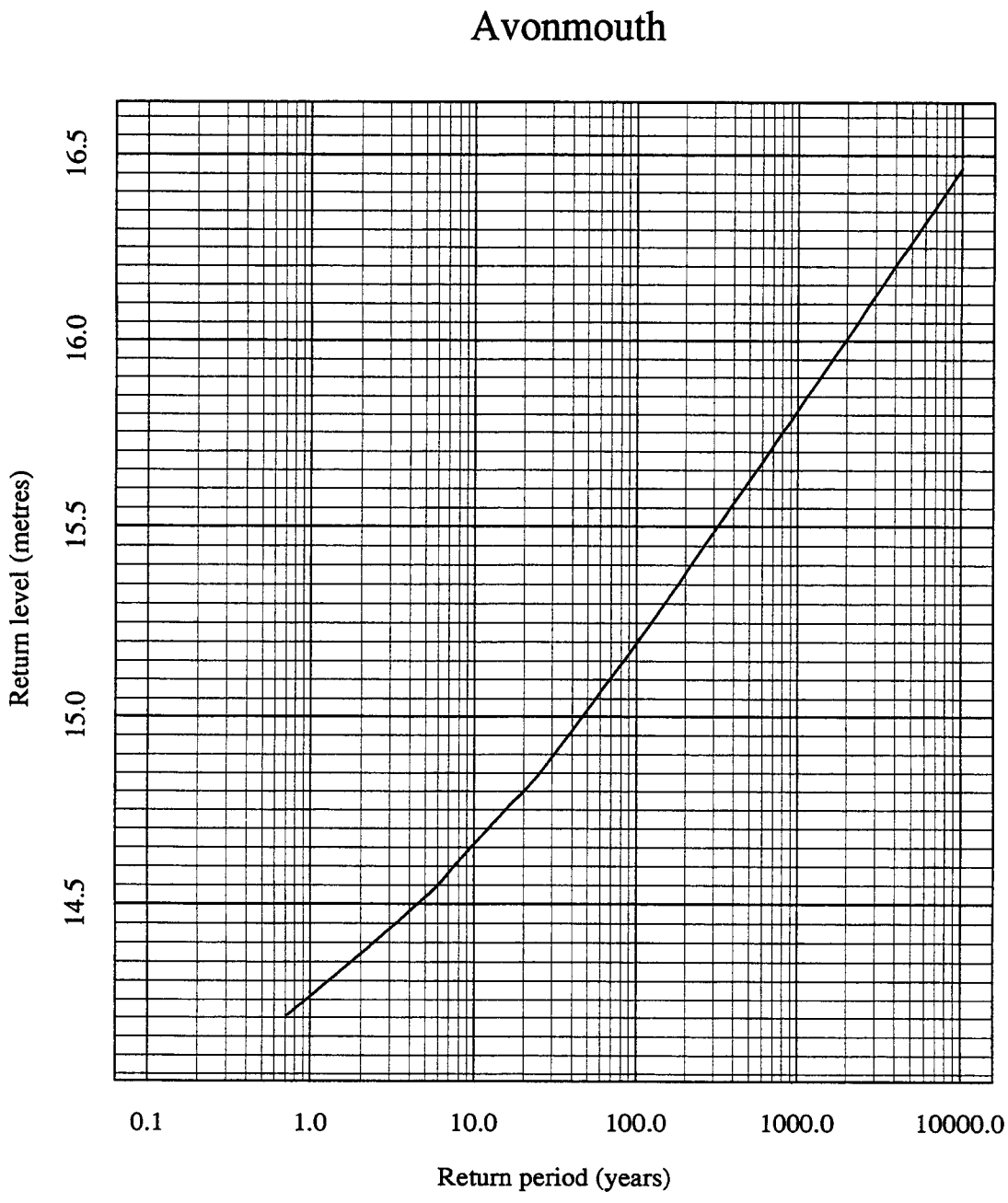

Figure 16.22: Port Diagram for the best site-by-site method for Avonmouth relative to ACD in 1990.

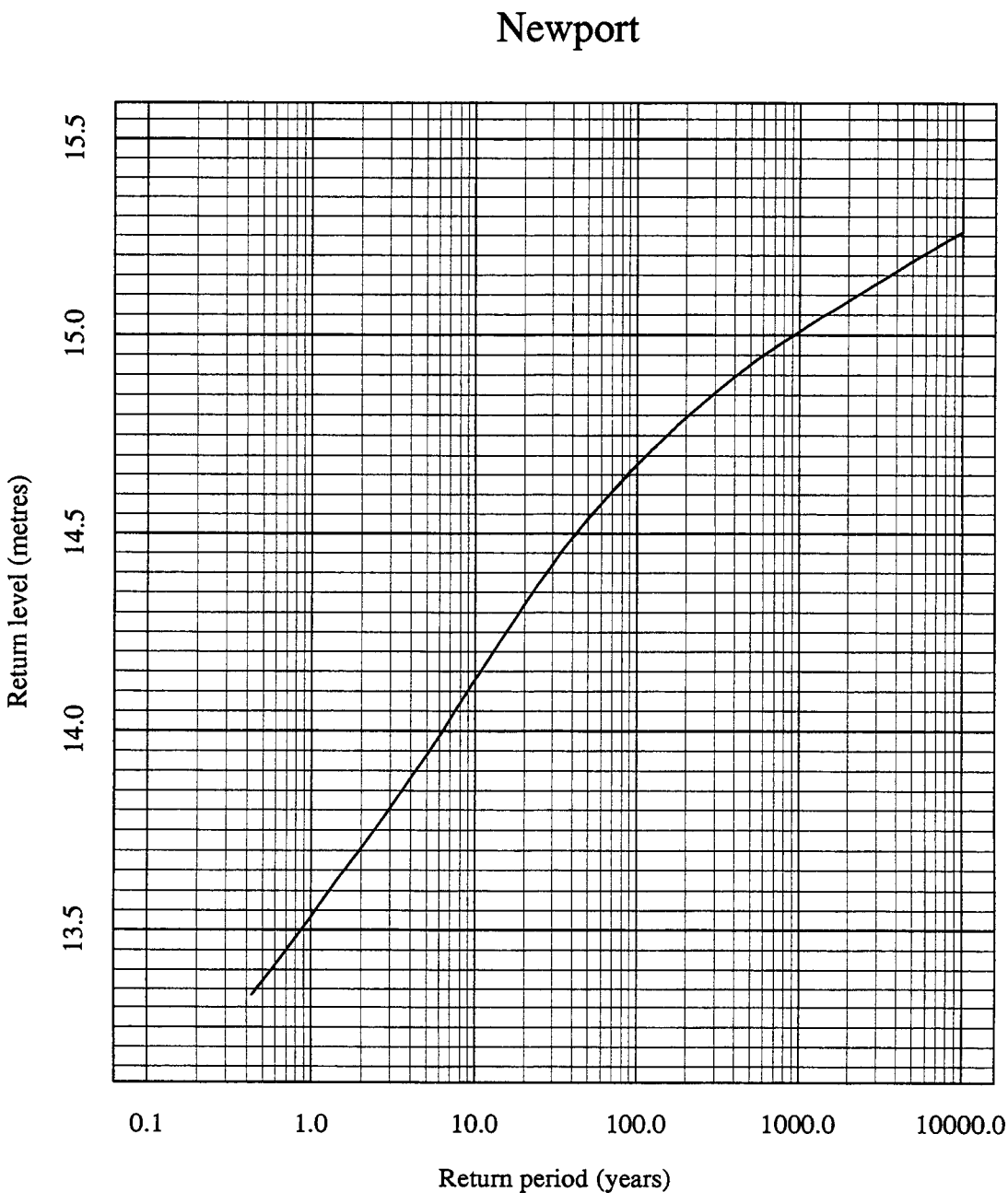

Figure 16.23: Port Diagram for the best site-by-site method for Newport relative to ACD in 1990.

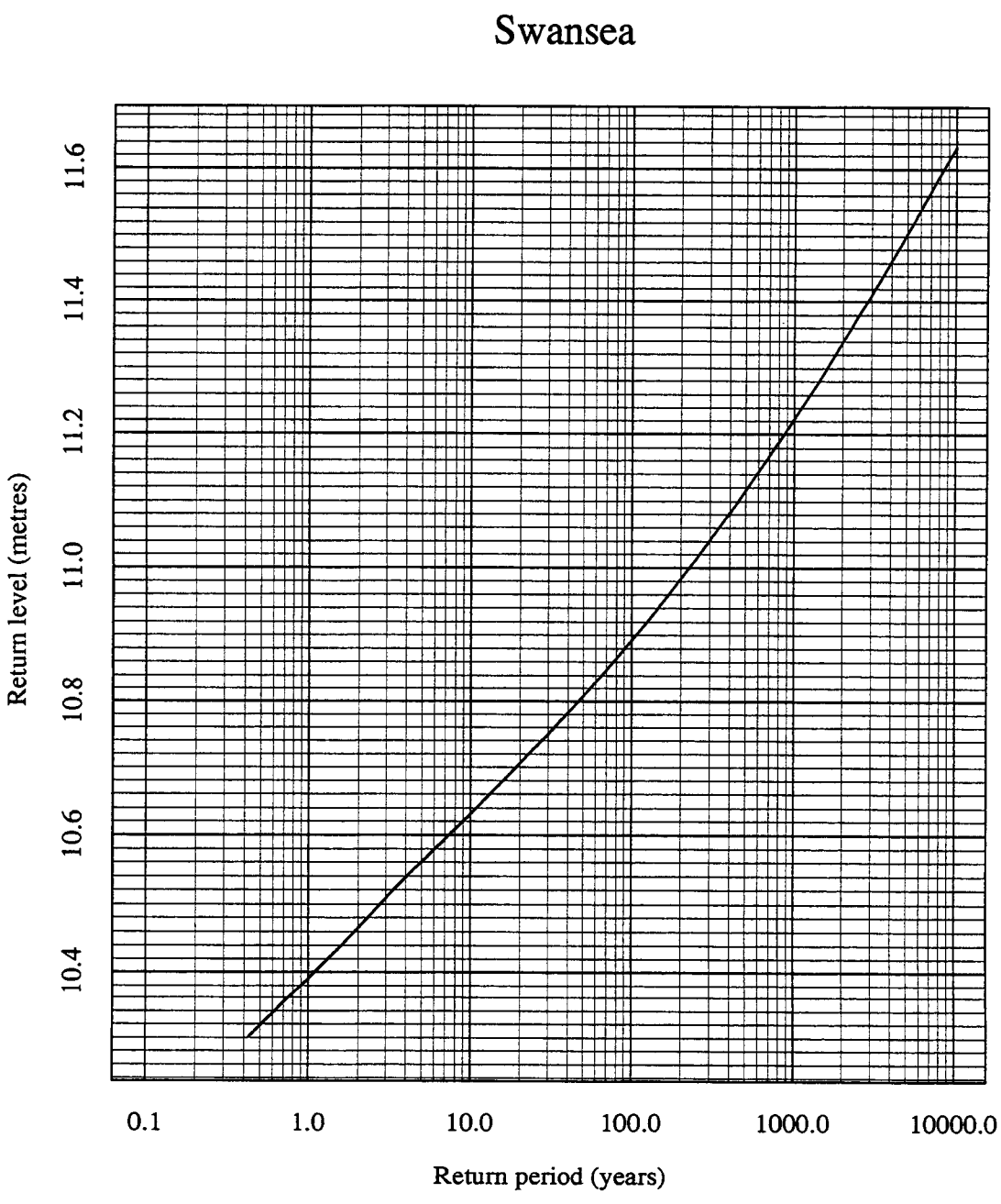

Figure 16.24: Port Diagram for the best site-by-site method for Swansea relative to ACD in 1990.

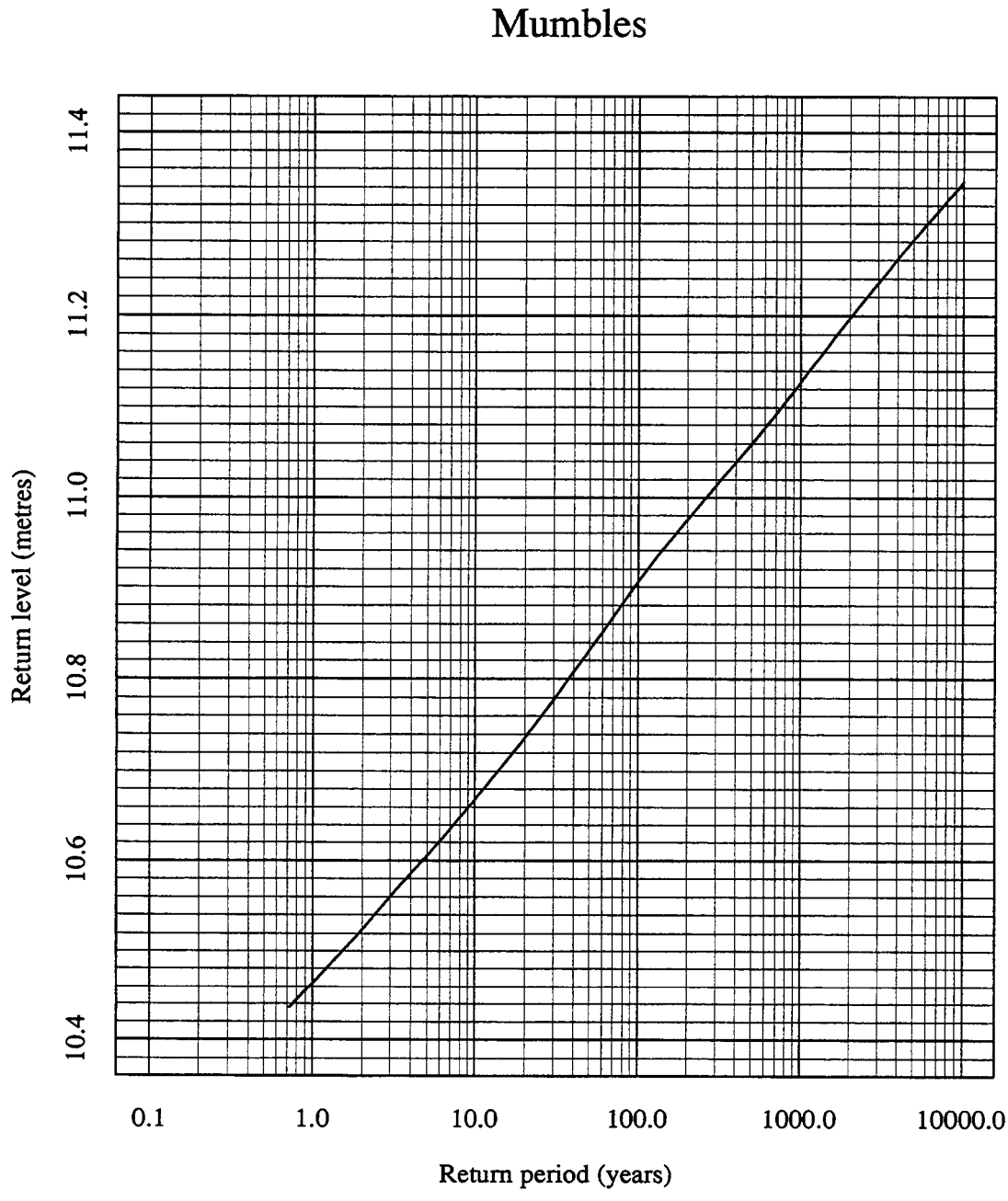

Figure 16.25: Port Diagram for the best site-by-site method for Mumbles relative to ACD in 1990.

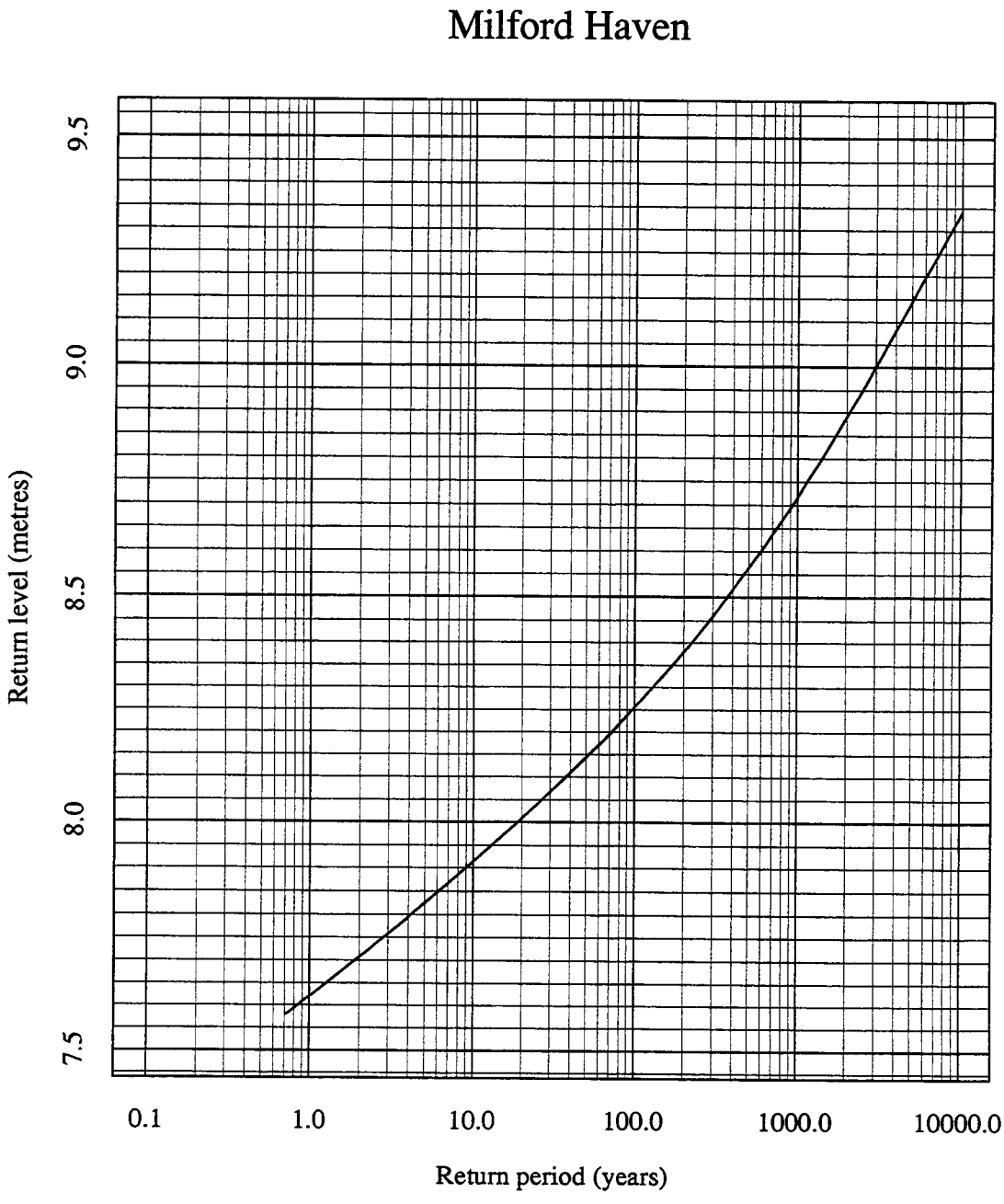

Figure 16.26: Port Diagram for the best site-by-site method for Milford Haven relative to ACD in 1990.

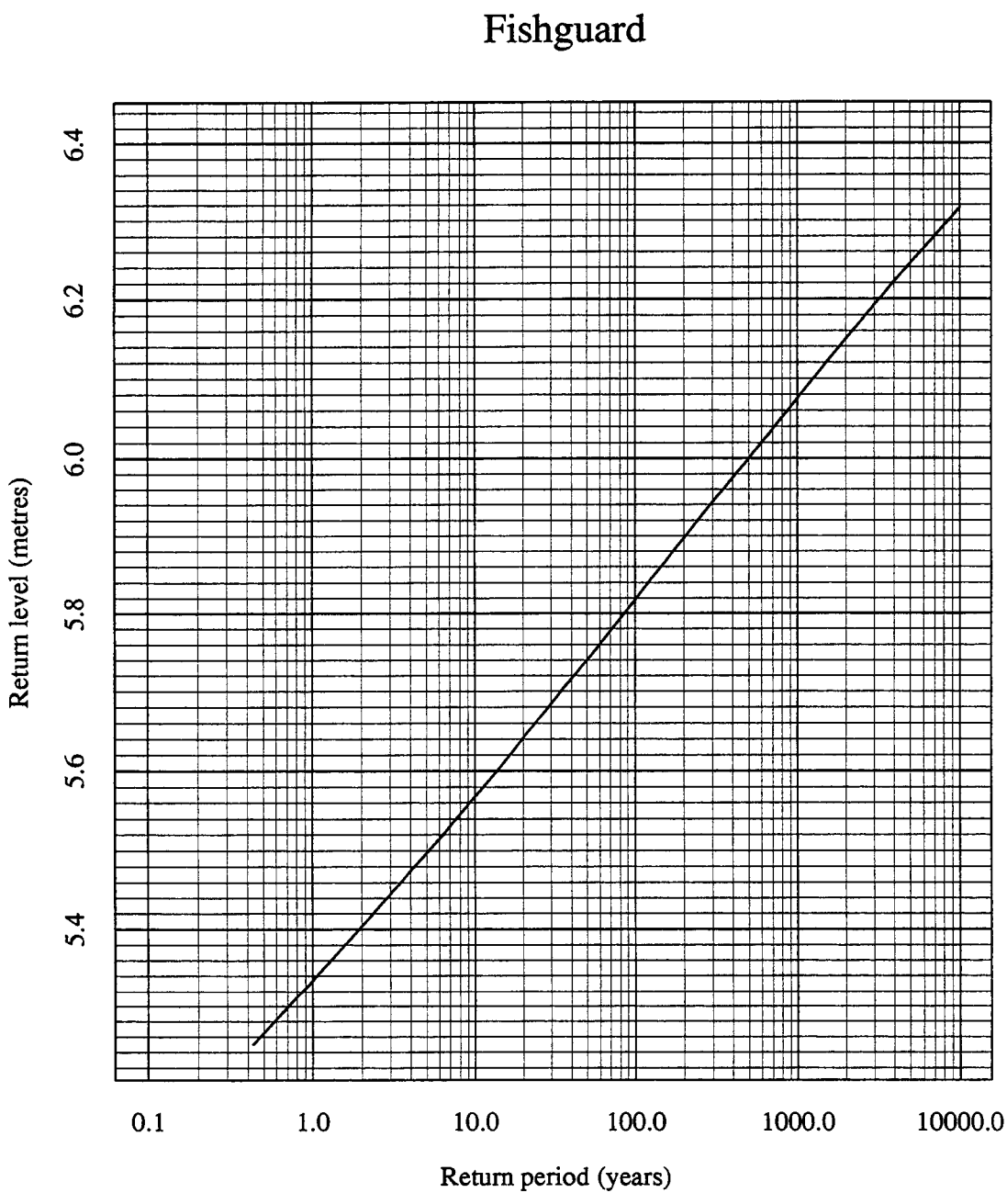

Figure 16.27: Port Diagram for the best site-by-site method for Fishguard relative to ACD in 1990.

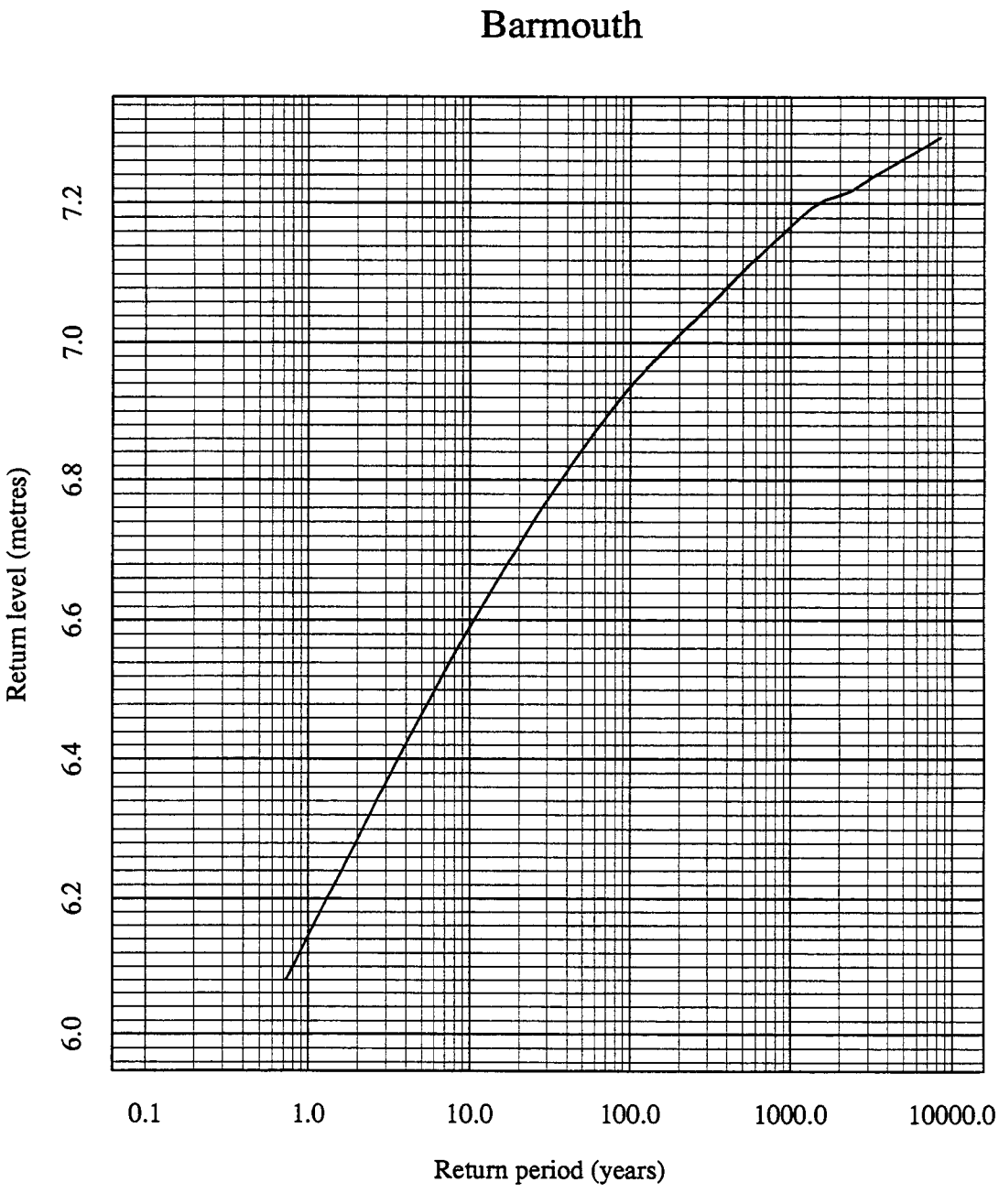

Figure 16.28: Port Diagram for the best site-by-site method for Barmouth relative to ACD in 1990.

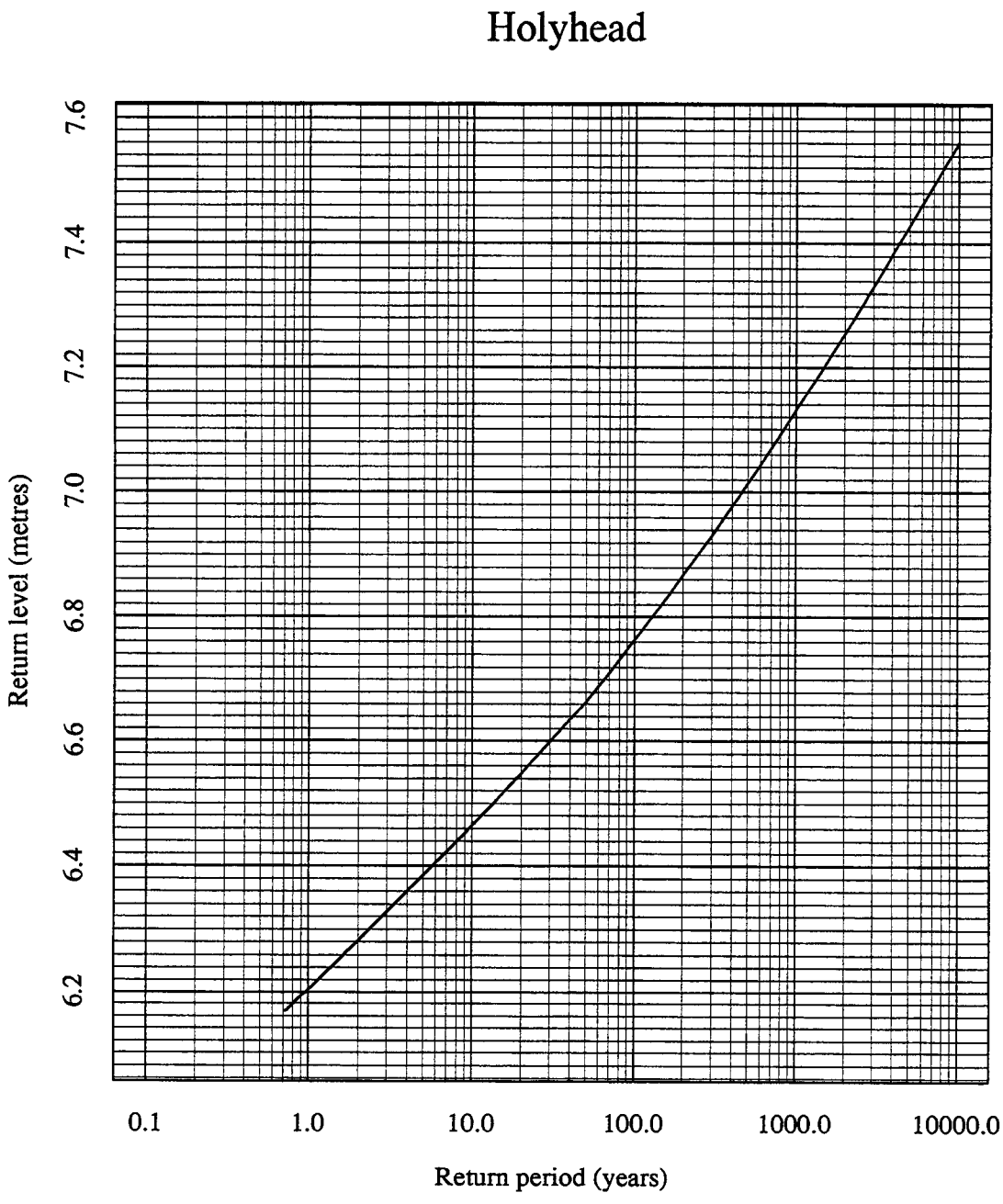

Figure 16.29: Port Diagram for the best site-by-site method for Holyhead relative to ACD in 1990.

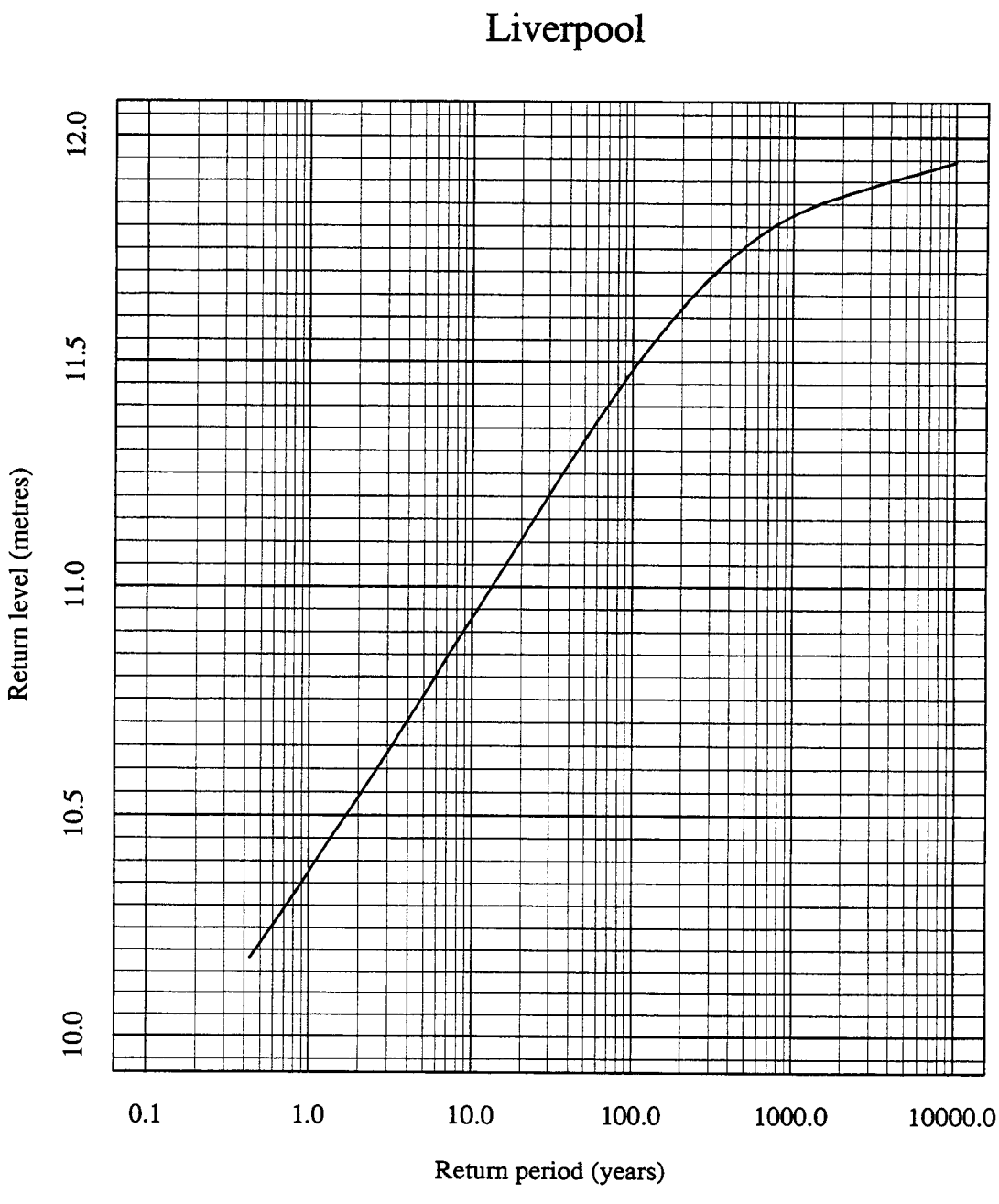

Figure 16.30: Port Diagram for the best site-by-site method for Liverpool relative to ACD in 1990.

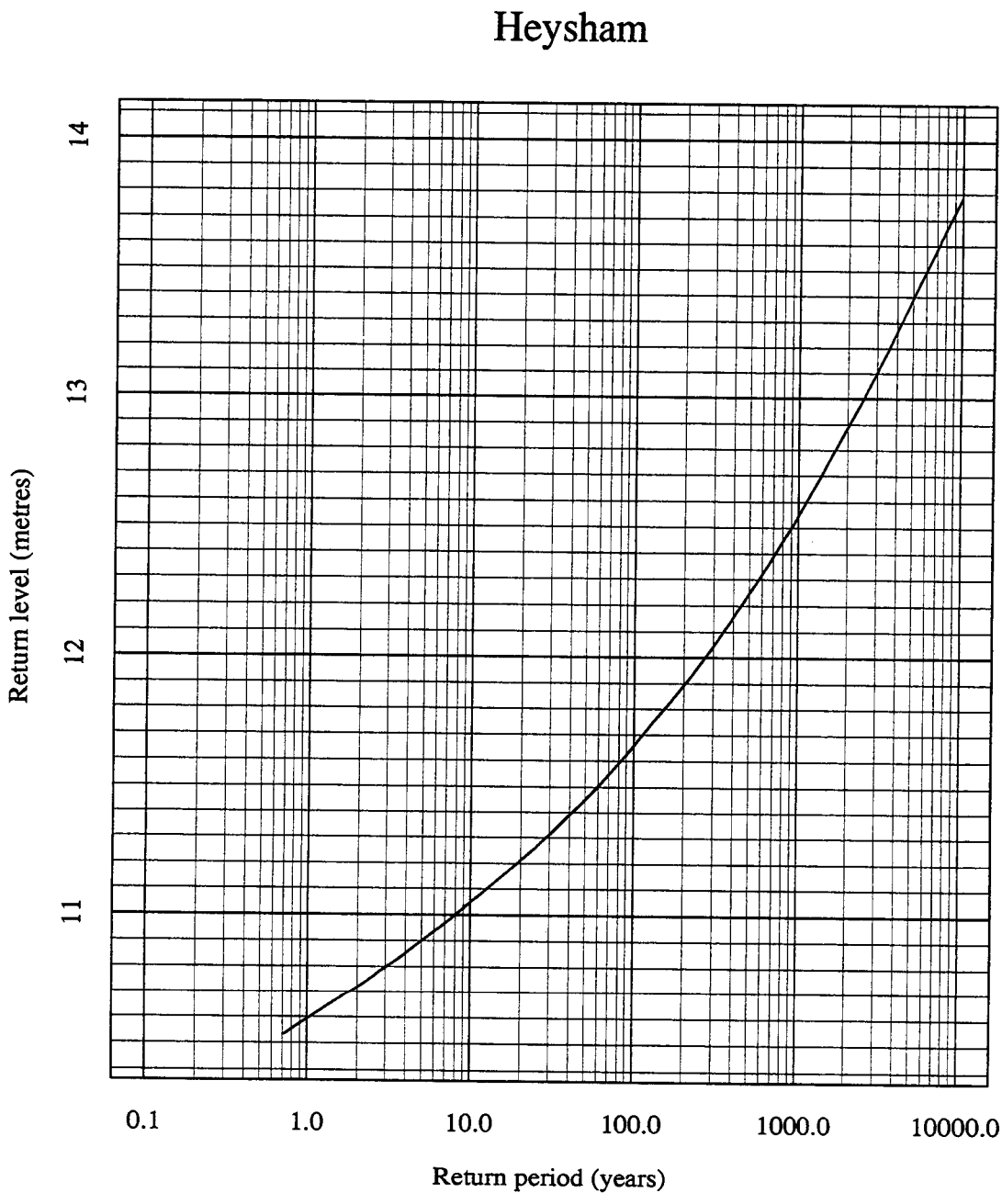

Figure 16.31: Port Diagram for the best site-by-site method for Heysham relative to ACD in 1990.

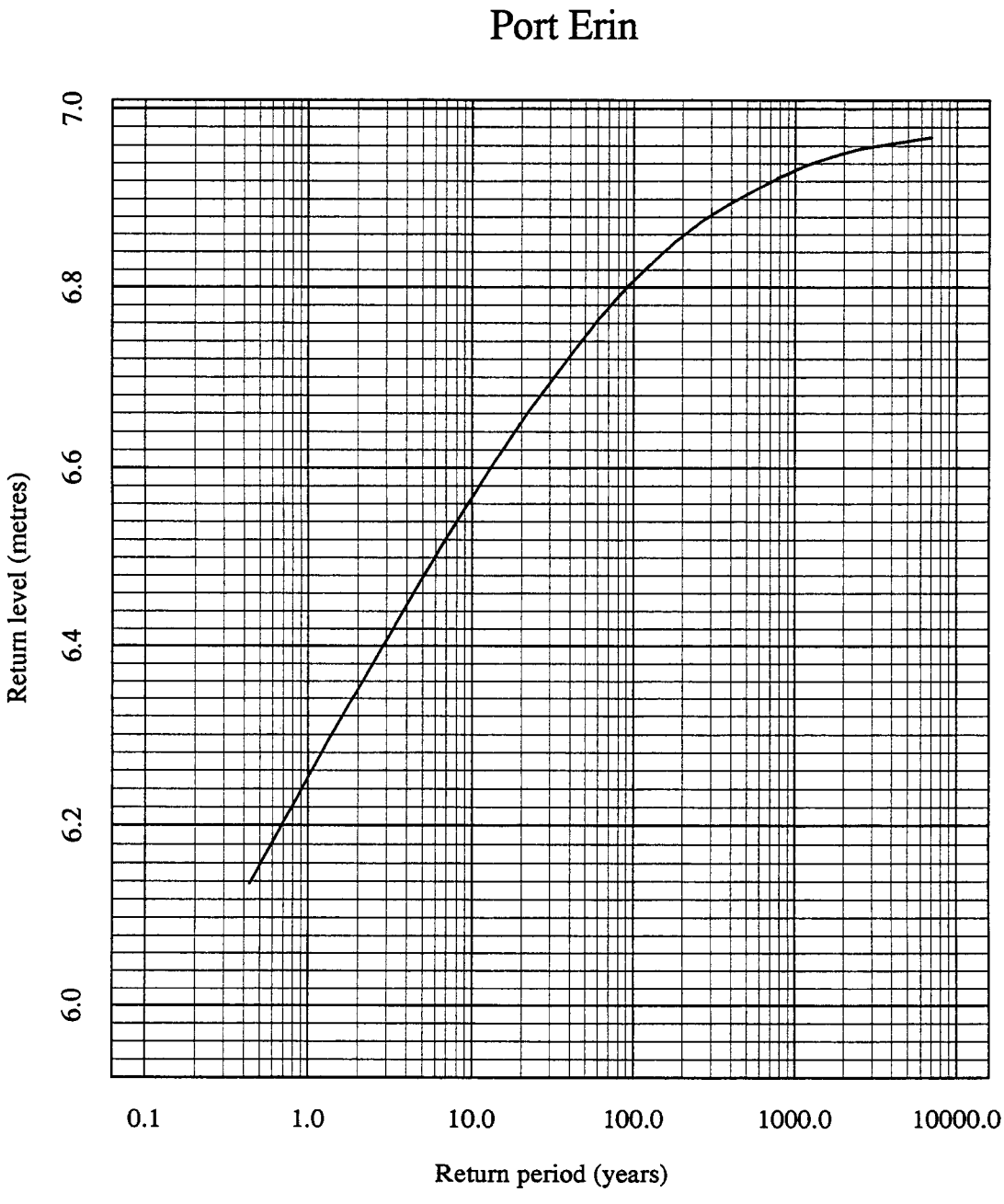

Figure 16.32: Port Diagram for the best site-by-site method for Port Erin relative to ACD in 1990.

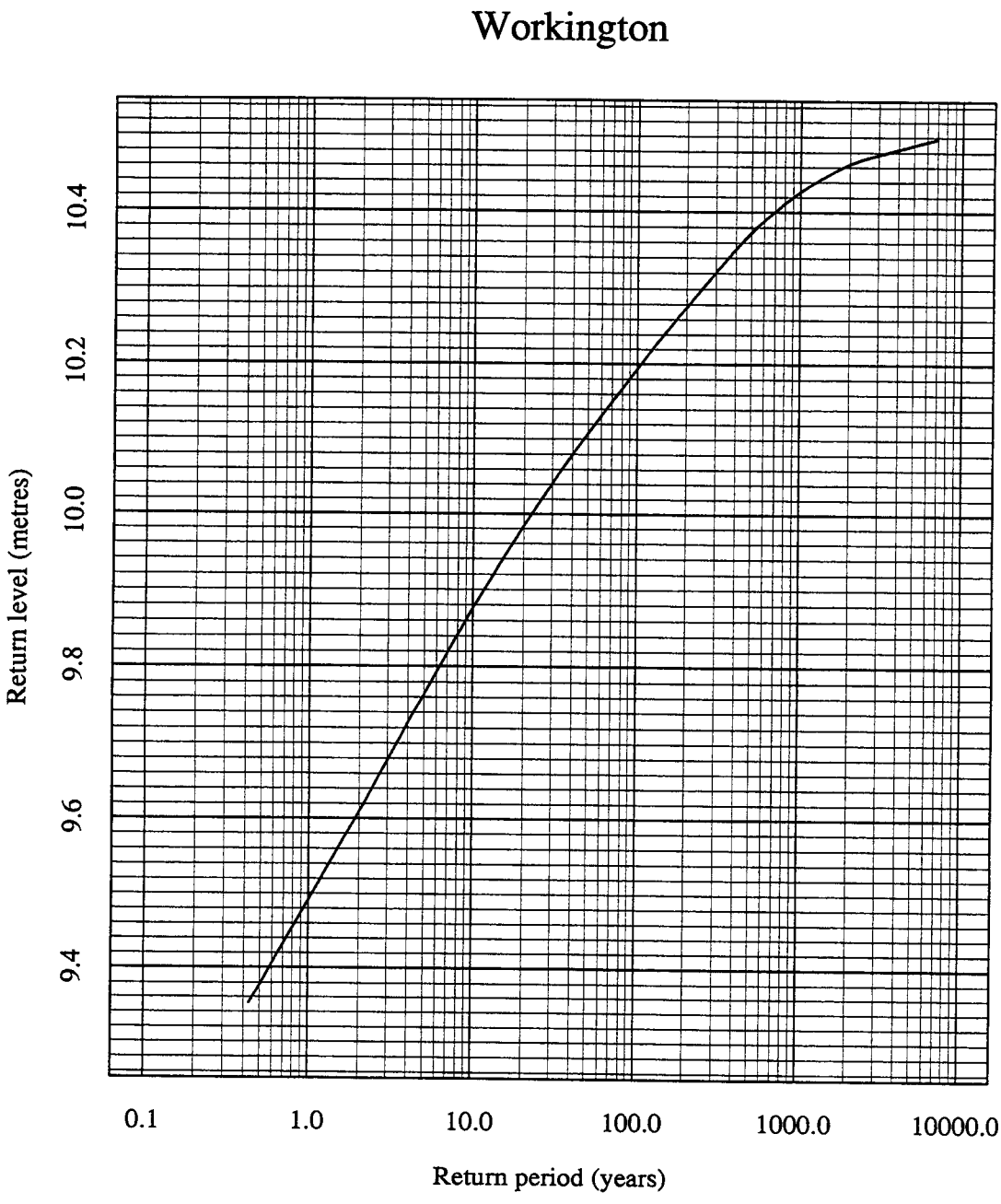

Figure 16.33: Port Diagram for the best site-by-site method for Workington relative to ACD in 1990.

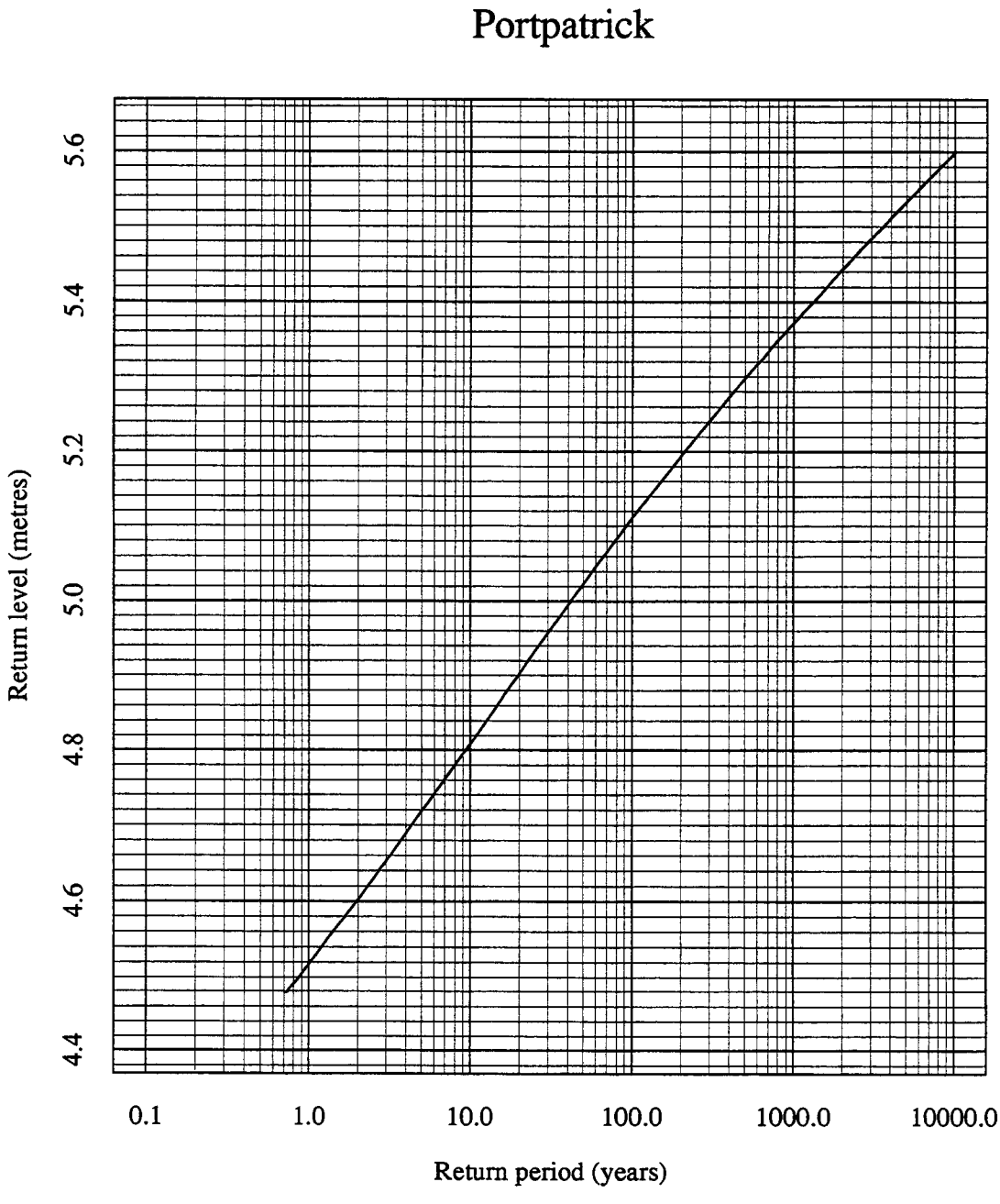

Figure 16.34: Port Diagram for the best site-by-site method for Portpatrick relative to ACD in 1990.

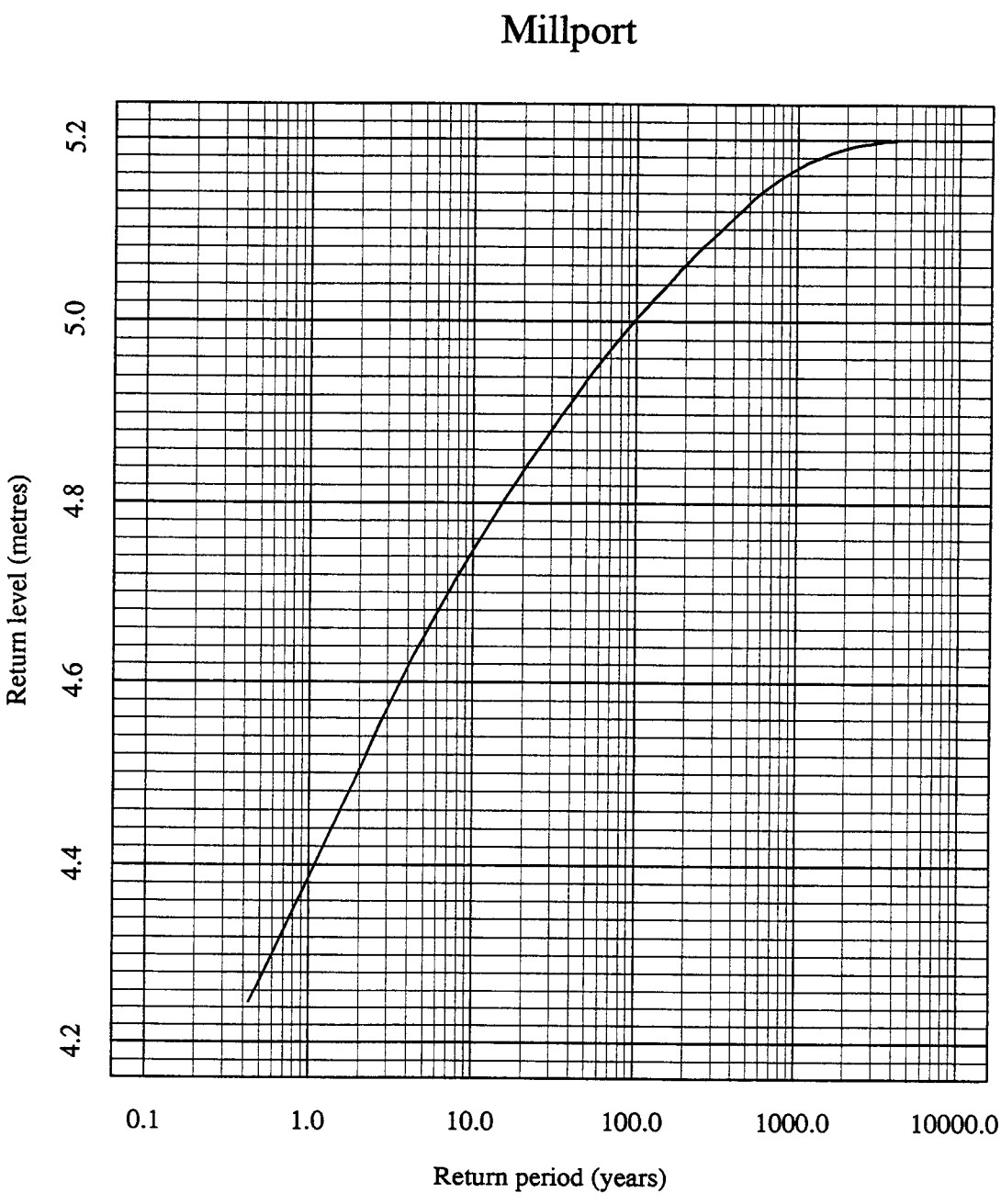

Figure 16.35: Port Diagram for the best site-by-site method for Millport relative to ACD in 1990.

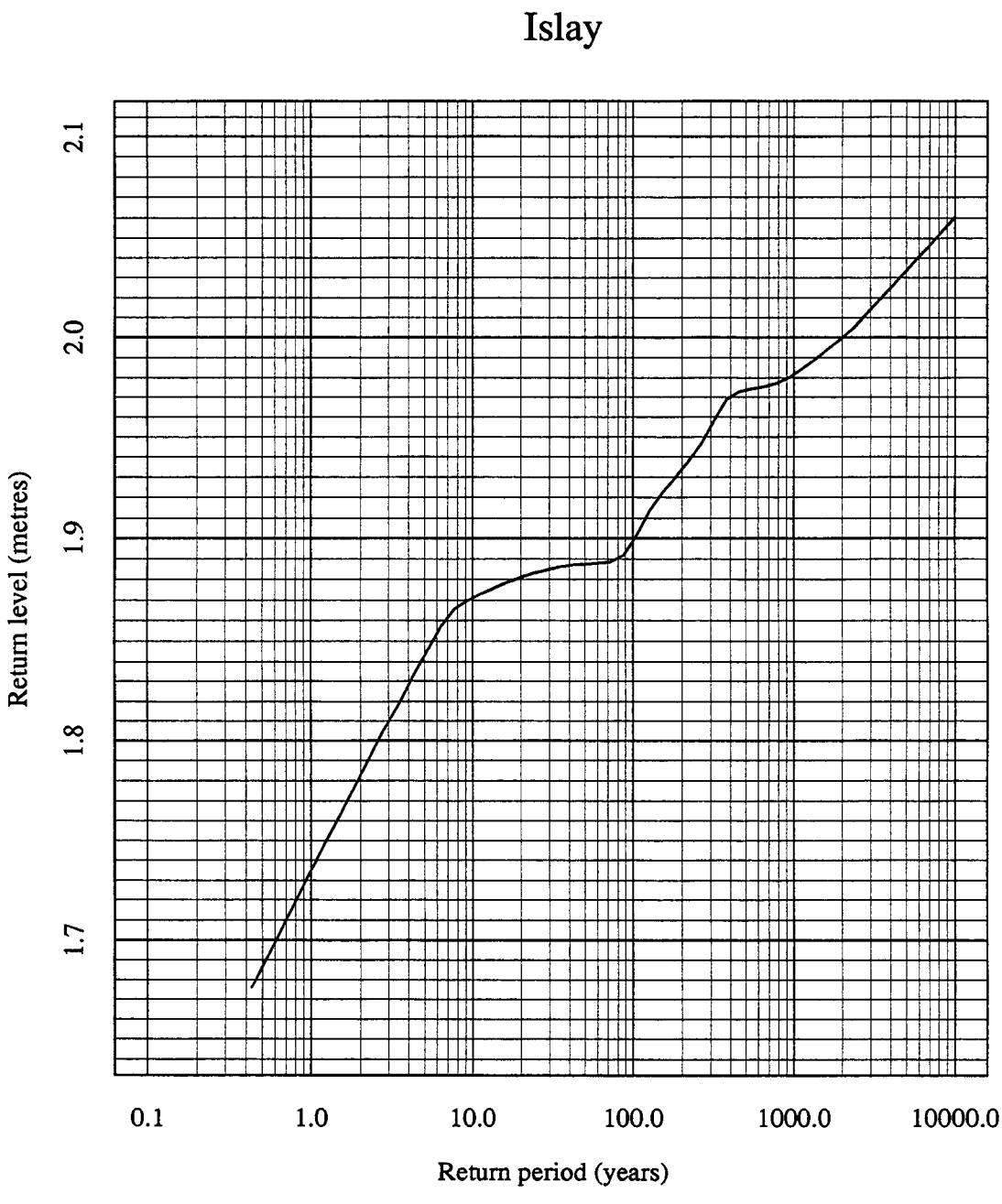

Figure 16.36: Port Diagram for the best site-by-site method for Islay relative to ACD in 1990.

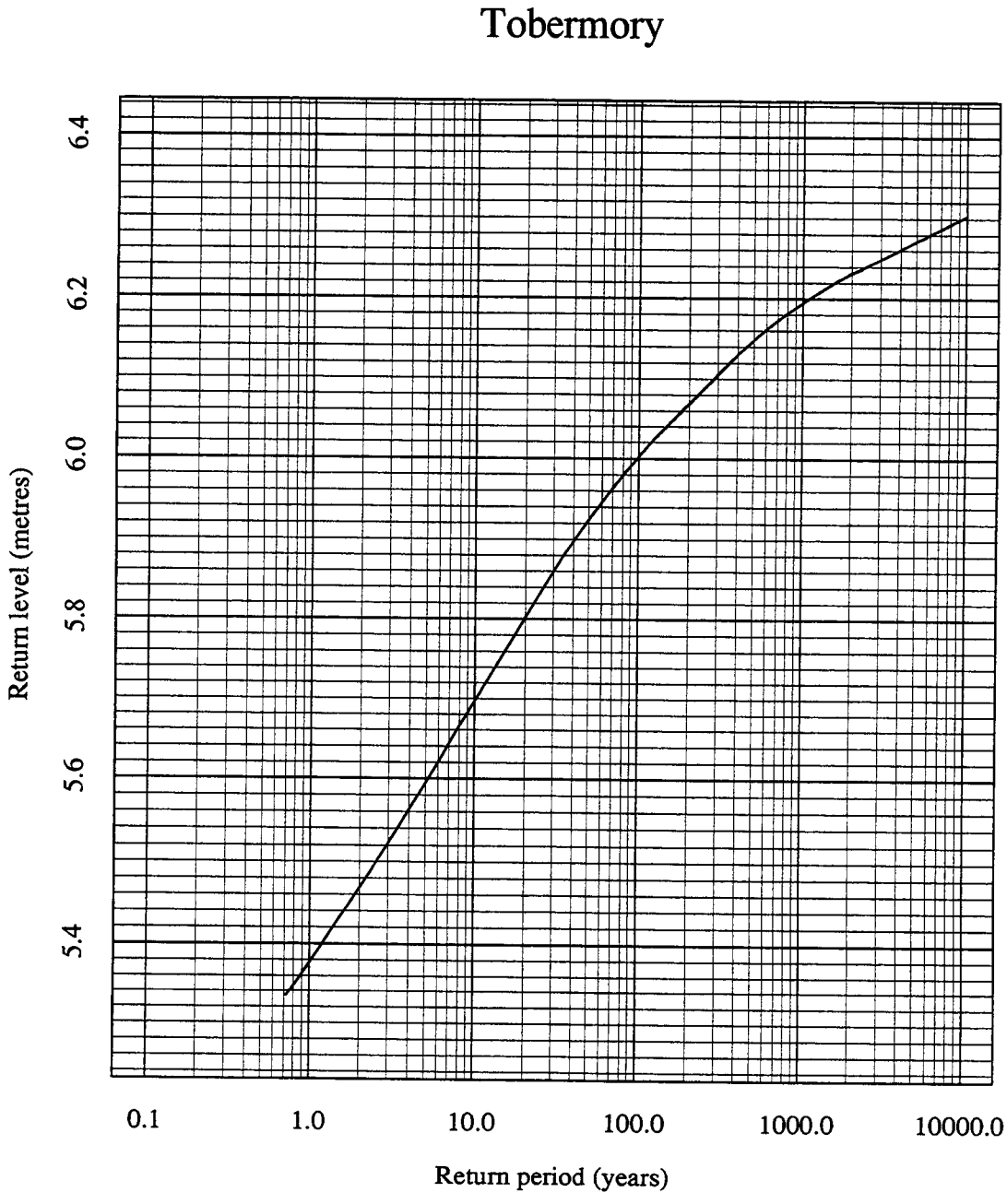

Figure 16.37: Port Diagram for the best site-by-site method for Tobermory relative to ACD in 1990.

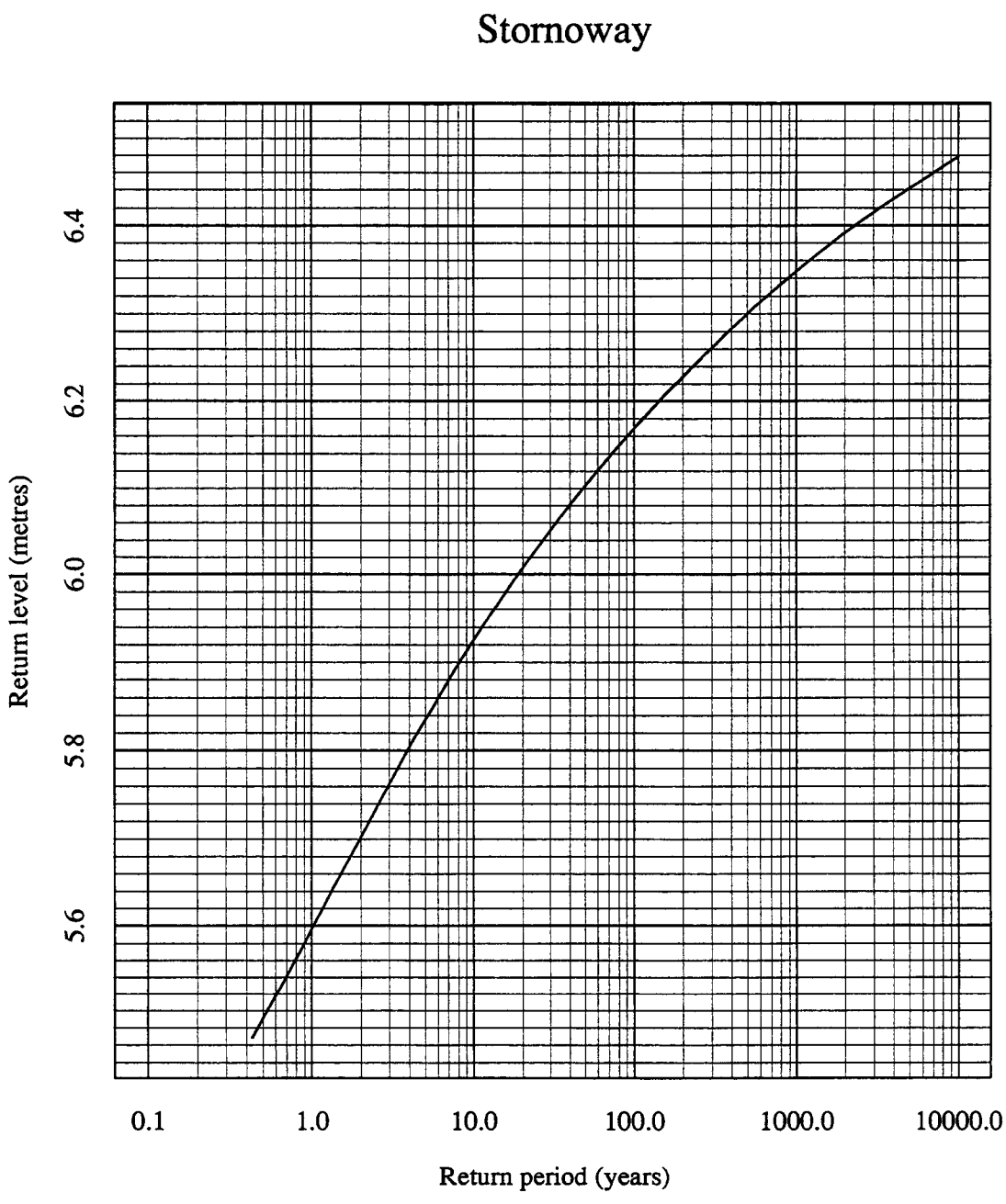

Figure 16.38: Port Diagram for the best site-by-site method for Stornoway relative to ACD in 1990.

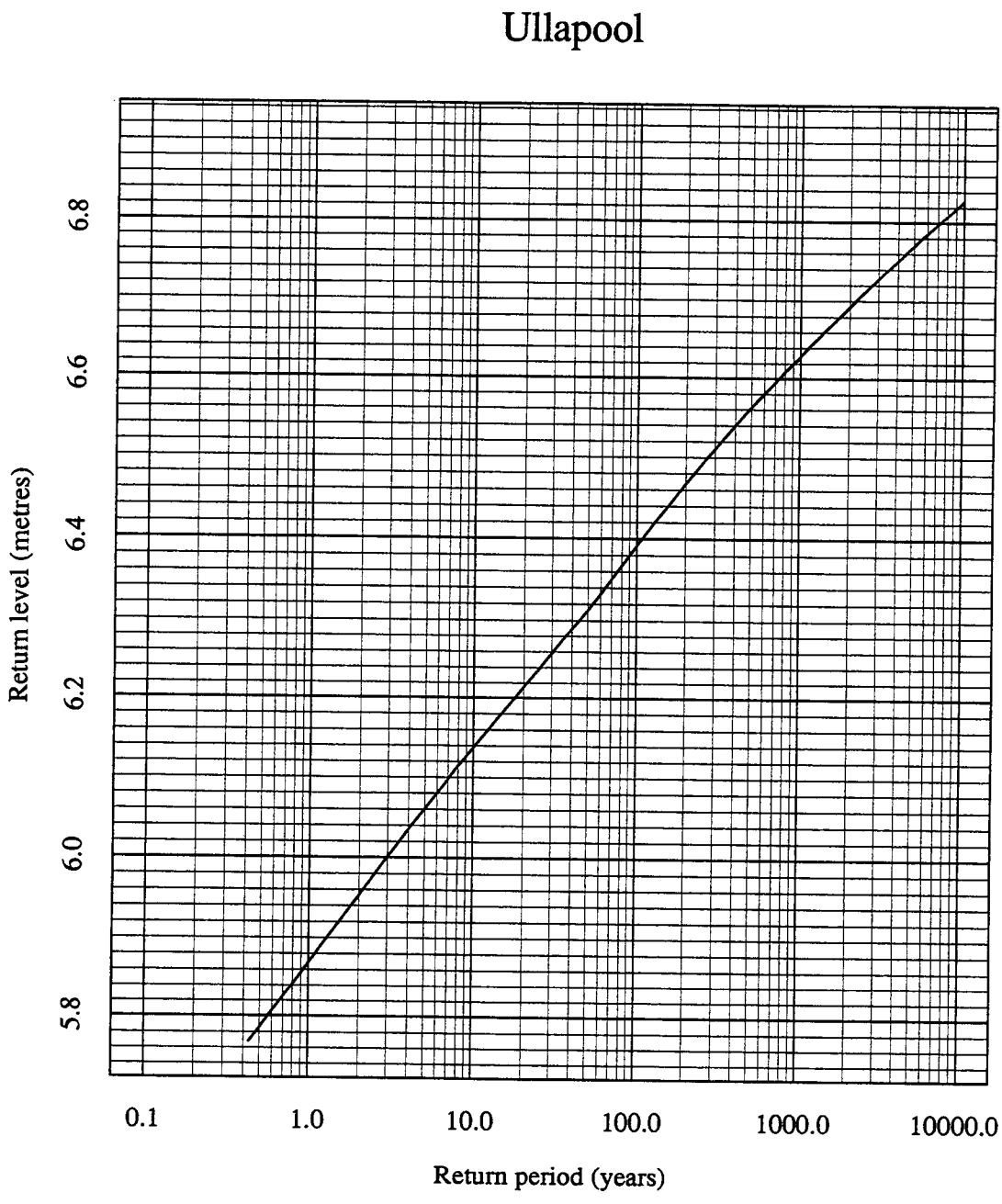

Figure 16.39: Port Diagram for the best site-by-site method for Ullapool relative to ACD in 1990.

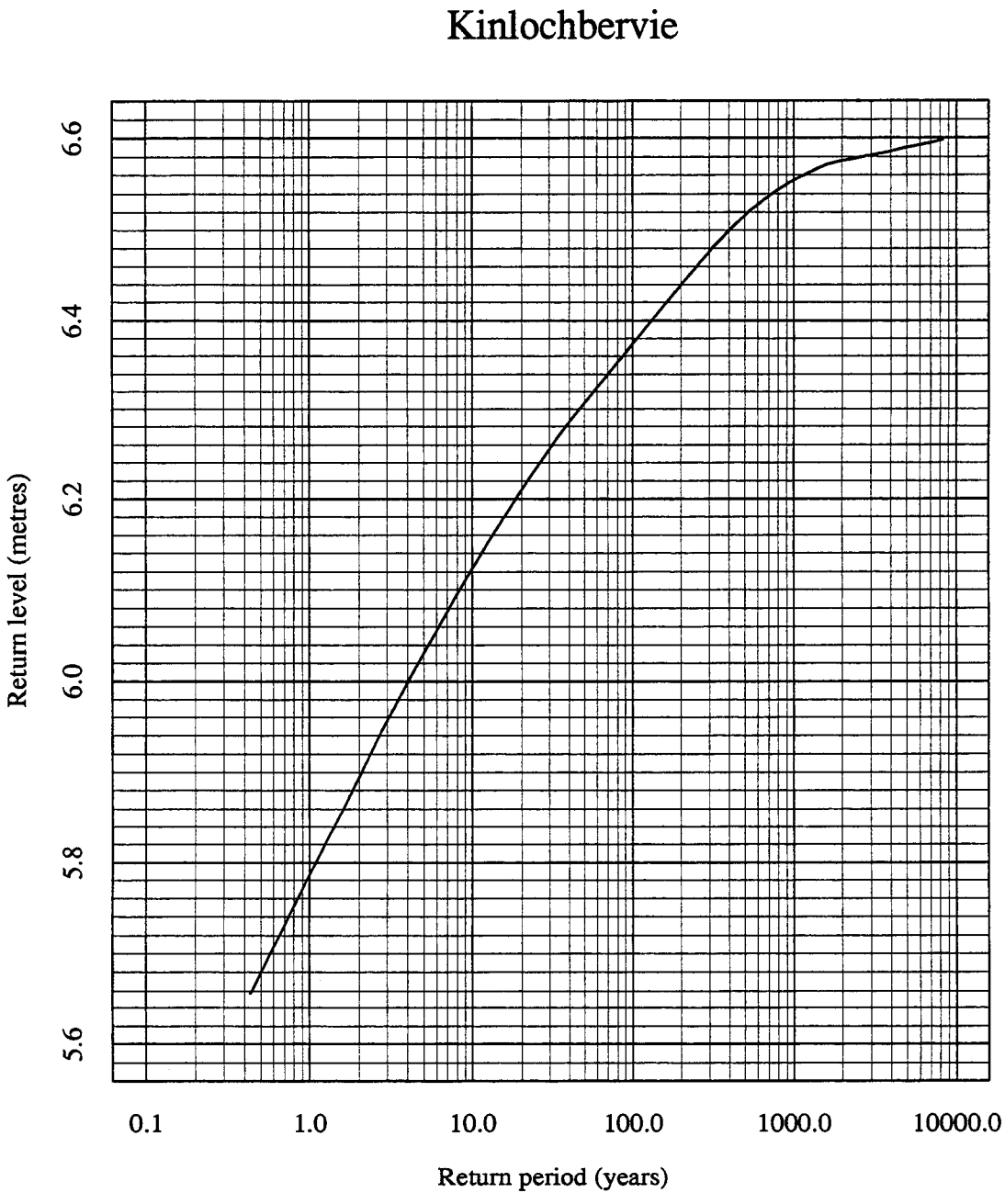

Figure 16.40: Port Diagram for the best site-by-site method for Kinlochbervie relative to ACD in 1990.

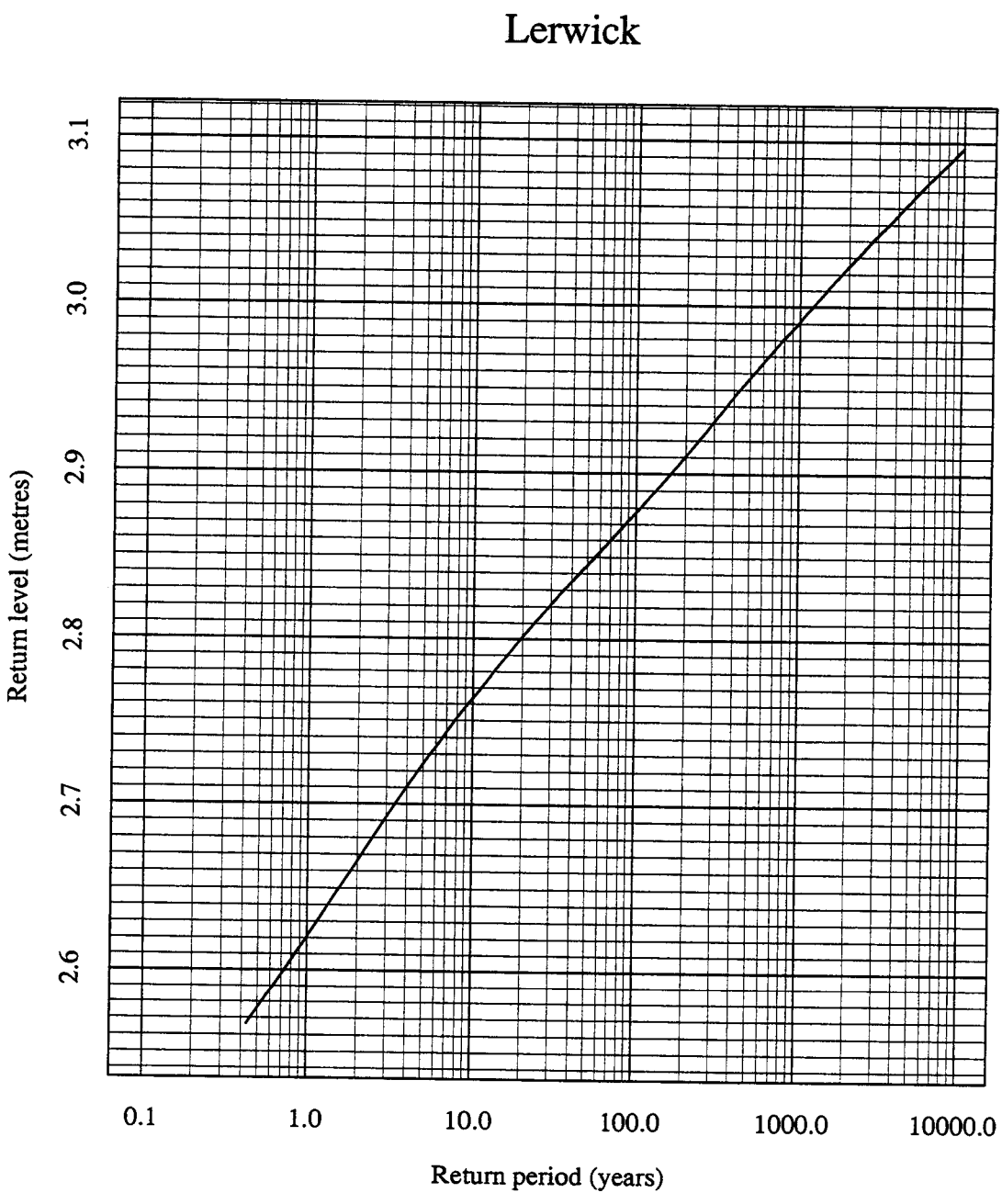

Figure 16.41: Port Diagram for the best site-by-site method for Lerwick relative to ACD in 1990.