Trend estimation in extremes of synthetic North Sea surges

Adam Butler†
Biomathematics & Statistics Scotland, Edinburgh, UK
Janet E. Heffernan and Jonathan A. Tawn
Lancaster University, UK
and Roger A. Flather
Proudman Oceanographic Laboratory, Liverpool, UK

Summary. Mechanistic models for complex atmospheric and hydrological processes are often used to simulate extreme natural events, usually in order to quantify the risks associated with these events. In this paper we use novel extreme value methods to analyse the statistical properties of output from a numerical storm surge model for the North Sea. The ‘model data’ constitute a reconstruction of the storm surge climate for the period 1955-2000 based upon a high quality meteorological dataset, and constitute the only available source of information on surge elevations at offshore and unmonitored coastal locations over this period. Previous studies have used extreme value methods to analyse storm surge characteristics, but we are able to extend and improve upon these analyses by using a local likelihood approach to provide a nonparametric description of temporal and spatial variations in the magnitude and frequency of storm surge events.

Keywords: Coastal flood risk; Extreme value theory; Extreme sea levels; Local likelihood; Statistical oceanography; Storminess; GEV model; Return level; Surge

1. Introduction

It is widely accepted that global mean sea levels have increased steadily over the past century or so as a result of an increase in the global mean atmospheric temperature. Continued increases in mean sea levels are predicted to have catastrophic impacts upon coastal communities around the world during the 21st century. Coastal flood events are almost invariably associated with extreme sea levels, however, and the true impact of climate change upon coastal flood risk will therefore depend critically upon the extent to which trends in mean sea level are indicative of trends in more extreme sea levels. There is some empirical evidence that levels of storminess may be increasing (e.g. WASA, 1998), and it has been hypothesised that this may lead to further increases in the level of coastal flood risk through an increase in the magnitude and frequency of “storm surge” events. Surges are transient distortions in the level of the sea surface which result from the action of wind and air pressure (Flather, 2001). The degree of distortion is known as the “surge elevation”, and may

†Address for correspondence: Adam Butler, Biomathematics & Statistics Scotland, James Clerk Maxwell Building, The King’s Building, Edinburgh EH9 3JZ, United Kingdom
E-mail: adam@bioss.ac.uk
be either positive or negative. Surge elevations are just one additive component of the overall sea level, together with surface waves, tides, and the mean sea level. Extreme positive surge elevations are a crucial contributory factor in coastal flooding along the shorelines of shallow and constrained basins such as the North Sea, the Bay of Bengal and the Gulf of Mexico. The evidence for historical changes in extreme surge elevations in the observational record is weak (Dixon and Tawn, 1992; Bijl et al., 1999). This is not perhaps surprising since the sparsity and inhomogeneity of the available data mean that such empirical studies typically have low statistical power (Pugh and Maul, 1999).

Many of the key aspects of storm surge dynamics can be encapsulated within deterministic geophysical models which are derived from approximations to the fundamental laws of fluid mechanics. Such “storm surge models” are routinely used for operational flood forecasting. Long runs from these models are now also, increasingly, being used to study statistical characteristics of both the past (Flather et al., 1998) and future (Woth et al., 2006) surge climate. If the inputs to storm surge models are provided by high quality historical meteorological data – as in this paper – then the outputs from these models provide the best available description of the past storm surge climate. From a statistical perspective the model output can effectively be regarded as spatially and temporally resolved storm surge elevation data (e.g. Langenberg et al., 1999).

We aim to detect and quantify changes over time in the magnitude and/or frequency of model-generated storm surge events in the North Sea, and to compare these changes with corresponding characteristics of surge elevation data from monitored coastal locations. Extreme value methods can be used to draw probabilistic inferences about unusually large – or small – values of a stochastic process, using only relatively extreme observations of that process. Such methods are typically used to draw inferences about events more extreme than those which have actually been observed (Coles, 2001). Extreme value models are widely used in disciplines such as hydrology, engineering, and finance, and have frequently been used to study the statistical characteristics of North Sea storm surges (Coles and Tawn, 1990; de Haan and de Ronde, 1998; Coles and Tawn, 2005). The statistical methods are motivated by a rich and well-established mathematical theory, describing the behaviour of stationary sequences at asymptotically high levels.

The statistical extreme value methodology can be easily extended to account for changes in the characteristics of extreme events over time if these changes can be well-approximated by a simple parametric form (Coles, 2001). The introduction of semiparametric and non-parametric estimation of changes in extreme values under weaker assumptions of temporal smoothness is relatively recent. Key approaches are based on smoothing splines (Pauli and Coles, 2001; Chavez-Demoulin and Davison, 2005), and on local likelihood (Hall and Tajvidi, 2000; Davison and Ramesh, 2000), but to date there have been few substantive applications of these methods. The model-generated dataset which we analyse here is large and highly structured, and raises various practical issues regarding the application of extreme value models to spatio-temporal data. The local regression model of Davison and Ramesh (2000) provides the basis for our analysis, but we also allow (Hall and Tajvidi, 2000) for the possibility that certain features of the extreme value parameters are homogeneous over time. This lets us account for the effects of the North Atlantic Oscillation, and also to identify changes in the magnitude and frequency of storm surges. We also extend the model to the spatio-temporal context. This extension is undertaken on the grounds that the pooling of spatial information in an extreme value analysis can improve the precision of estimators of extreme value characteristics at a particular site (Dixon et al., 1998).

The choice of an appropriate bandwidth is critical when applying nonparametric re-
Trend estimation in extremes of synthetic North Sea surges

Regression models. In the context of storm surge behaviour different levels of smoothing will highlight changes resulting from distinct physical processes operating at different temporal scales, so the choice of bandwidth depends primarily on the scientific objectives of the analysis. We concentrate here upon detecting decadal and longer-term variations in extreme surge levels by using a Gaussian kernel with a fixed bandwidth (standard deviation) of $h = 3.5$ years. Levels of storminess in the NE Atlantic and storm surge characteristics in the North Sea are both known to exhibit substantial variability at decadal levels (see WASA, 1998 and Bijl et al., 1999 respectively). The aim of the current analysis is to describe the nature of these smooth variations without making a priori assumptions about their form. We briefly discuss the extent to which our results are sensitive to bandwidth.

We describe the model-generated and observational surge elevation data in Section 2. In Section 3 we briefly review the standard extreme value models, and in Section 4 we outline the local likelihood approach for allowing the parameters of these models to vary smoothly over time. We present methods for introducing structural constraints into the local regression model in Section 5, and in Section 6 extend the approach to the spatio-temporal context. We conclude with a discussion of some more general scientific and statistical issues. Note that the detailed results and scientific interpretation of the analyses in Sections 4 and 5.2 are discussed in more detail within a companion article in an oceanographic journal (Butler et al., 2006), but that the more sophisticated analyses of Sections 5.1 and 6 are not dealt with in that paper. A fuller account of the methods and results is also given in an unpublished PhD thesis (Butler, 2005).

2. Surge elevation data

We analyse surge elevations taken from a single run of the CSX model, a two-dimensional hydrodynamical model which describes surge and tide dynamics across the European Continental Shelf. This deterministic model simulates surge elevations conditionally upon the values of meteorological inputs. Flather et al. (1998) used the model to generate a reconstruction of the surge climate for the period 1955-1992 by interpolating meteorological data from Reistad and Iden (1995) onto the CSX grid. We analyse an updated version of this reconstruction, which has been extended to include the period 1993-2000. The CSX model has a spatial resolution of approximately 35km by 35km. It is therefore capable of describing large-scale spatial variations in storm surge elevations, but not of resolving more localised features within coastal or estuarine areas. We restrict attention to model outputs for grid cells in the southern and central North Sea (Figure 1), the region of the shelf considered to be at greatest risk from coastal flooding.

Our “model data” therefore consist of 403248 simulated hourly surge elevations for each of 259 grid cells. We follow Flather et al. (1998) in analysing contemporaneous observational data on sea levels, derived from tide gauges at coastal locations. We restrict our attention to three sites on the UK coast for which there are relatively long series of data. Dominant tidal signals have been removed from these hourly data at the pre-processing stage, but long term trends in mean sea level have not been removed. The twenty largest storm surge elevations per year are plotted in Figure 2 for the model data and observational data at these three locations. Note that the model data apply to grid cells, so some discrepancy is to be expected as we are comparing model values which apply to areas with observational data taken at point locations.
Fig. 1. Area covered by the CSX model grid, which has a resolution of 35 by 35km. In this paper we focus only on output for 259 grid cells in the southern and central North Sea.
Fig. 2. The 20 largest surge elevations per year based on the CSX model and observational data for three ports on the British east coast (crosses). Estimated 50 year return levels from fitting the local $r$-largest model of Section 4 to these data using a bandwidth of $h = 3.5$ years (black dots), together with 95% variability bands (grey).
3. Extreme value models

Extreme value methods typically adopt one of two alternative approaches by modelling either the exceedances of a high threshold, or, after division of the data into blocks, block maxima. We adopt a generalisation of the latter strategy, and, for each spatial location, model the 20 largest independent surge elevations occurring within a year. Note that this “r-largest” approach defines an extreme value in purely relative terms, whereas the threshold exceedance approach defines an extreme event in absolute terms. It is consequently more natural and straightforward to use the r-largest approach when extending the analysis temporally and spatially, as we shall do in Sections 4 and 6 respectively, since the relative definition of an extreme event is robust to nonstationarities in the underlying process. Our use of the r-largest approach relies, however, on the assumption that stationarity of extreme events is a reasonable approximation within each year at all sites. This assumption is likely to be reasonable in this application because trends are of small magnitude relative to intra-annual variability and because extreme surge events occur within a relatively well-defined winter period.

Let $X_1, \ldots, X_n$ be a stationary sequence of random variables which satisfy a weak long-range dependence condition and have common distribution function $F$. Assume that there exist sequences of normalising constants $a_n > 0$ and $b_n$ such that

$$\mathbb{P} \left( \frac{\max(X_1, \ldots, X_n) - b_n}{a_n} \leq z \right) \to G(z; \theta) \text{ as } n \to \infty,$$

where $G$ is a non-degenerate distribution. Then $G$ will be a Generalised Extreme Value (GEV) distribution of the form

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

where $\theta = (\mu, \sigma, \xi)$ are location, scale and shape parameters respectively, and where $x_+ = \max(x, 0)$ (see Leadbetter et al., 1983; Coles, 2001). Positive values of the shape parameter ($\xi > 0$) arise when $F$ is heavy tailed (e.g. student $t$ or Cauchy), negative values ($\xi < 0$) correspond to light tailed distributions such as the uniform ($\xi = 1$), and $\xi = 0$ arises when the upper tail of $F$ decays exponentially.

In environmental applications the GEV model is typically used to model annual maxima, with $a_n$ and $b_n$ being absorbed into the unknown parameters $\mu$ and $\sigma$. The GEV parameters can be estimated directly using, for example, maximum likelihood. A widely used design parameter in the context of coastal engineering is the $T$-year “return level”,

$$S_T = \mu - \frac{\sigma}{\xi} \left[ 1 - \left\{ -\log \left( \frac{T-1}{T} \right) \right\}^{-\xi} \right],$$

which corresponds to the level that is exceeded with probability $1/T$ in any particular year. Here $T$ is typically taken to be larger than the number of years for which data are available, in which case inference about the return level $S_T$ involves extrapolation. Our primary interest in this paper will be to estimate the 50 year return level for surge elevations, $S_{50}$. Extreme value methods provide a relatively robust basis for such extrapolation as they make weak assumptions about the form of the distribution of the surge.
The GEV model uses only one value per year, and so makes inefficient use of the available data. Using a similar asymptotic theory to that which motivated the GEV model, we can extend this model to make use of the $r$-largest independent observations per year for fixed $r$ with $r \geq 1$. Let $M_n^{(k)}$ denote the $k$-th largest independent value of the sequence $X_1, \ldots, X_n$ where $1 \leq k \leq r$. Assume that (1) holds with $G$ being a GEV distribution with parameters $\theta = (\mu, \sigma, \xi)$. It follows that

$$P\left(\frac{M_n^{(k)} - b_n}{a_n} \leq z\right) \rightarrow H_k (z^{(k)}; \theta) = G \left( z^{(k)}; \theta \right) \sum_{j=0}^{k-1} \frac{[- \log G (z^{(k)}; \theta)]^j}{j!}$$

and that the limiting joint density of \{(M_n^{(1)} - b_n)/a_n, \ldots, (M_n^{(r)} - b_n)/a_n\} is

$$h(z^{(1)}, \ldots, z^{(r)}; \theta) = \exp \left[ - \left\{ 1 + \xi \left( \frac{z^{(r)} - \mu}{\sigma} \right) \right\}^{-\frac{1}{\xi}} \right] \prod_{k=1}^{r} \frac{1}{\sigma} \left\{ 1 + \xi \left( \frac{z^{(k)} - \mu}{\sigma} \right) \right\}^{-\frac{1}{\xi} - 1}$$

where $z^{(1)} \geq z^{(2)} \geq \ldots \geq z^{(r)}$. The parameters of this $r$-largest model are the same parameters as for the GEV model and hence inference above the GEV parameters is made more efficient by use of the $r$-largest model.

The validity of the $r$-largest approach depends on selecting a value of $r$ that is in accord with asymptotic theory, and on being able to identify the $r$-largest independent values within a stationary sequence. Choosing a value for $r$ constitutes a trade-off between bias and variance: using too few values per block leads to imprecise estimates; using too many values leads to bias due to the asymptotic model not being appropriate for the data because $r/n$ is not sufficiently close to 0. Parameter stability plots (Coles, 2001) provide a subjective but useful tool for selecting $r$: they are constructed by graphing estimates of the extreme value parameters $\mu$, $\sigma$ and $\xi$ as a function of $r$, and are then used to search for the highest value of $r$ below which estimates for all three parameters appear to be stable. We use a value of $r = 20$ throughout this paper, based on assessing parameter stability plots for model-generated surges at an arbitrarily selected subset of around 25 grid cells (not shown). As hourly surge data exhibit strong autocorrelation, we use a declustering algorithm (Tawn, 1988) to ensure that the $r$-largest values that we select from a year are separated from each other by an interval of at least 25 hours.

We follow Flather et al. (1998) in applying an $r$ largest analysis to surge elevations generated by the CSX model for particular grid cells, and, where they are available, to the corresponding observational data. In Figure 3 we plot maximum likelihood estimates of the 50 year return level $S_{50}$ based on model-generated surge data, for each of the 259 grid cells within the study region. Note that we fit the extreme value model separately at each of the 259 cells, ignoring the effects of spatial dependence; we return to this issue in Section 6. Return level estimates vary substantially across space, from less than 100cm around the Scottish east coast to more than 500cm within the Wadden Sea (near the German coast). Extreme surge elevations are greatest in areas of shallow water, although the relationship with water depth is complicated. The patterns in Figure 3 are well known to oceanographers (e.g. Flather et al., 1998).

The estimates for the location and shape parameters $\mu$ and $\sigma$ are not shown, but exhibit very similar spatial patterns to those seen for $S_{50}$. The shape parameter $\xi$ quantifies a more subtle aspect of extremal behaviour, and is the critical parameter in determining the estimates of return levels $S_T$ when $T$ is very large (e.g. $T \gg 50$). Shape parameter
Fig. 3. Maximum likelihood estimates (in cm) of 50 year return levels $S_{50}$ obtained by fitting the $r$-largest model of Section 3 with $r = 20$ to surge elevations generated by the CSX model. The analysis is performed separately for each grid cell. Letters denote the locations of selected ports: A = Aberdeen, S = North Shields, I = Immingham, L = Lowestoft, T = Tilbury, D = Dover, V = Vlissingen, N = Den Helder, M = Eemshaven, E = Esbjerg, H = Hirtshals.
estimates are negative across most of the study region (Figure 4), suggesting that the distribution of storm surges is light tailed with a finite upper endpoint. Estimates are significantly less than zero (at the 0.05 level) for thirty grid cells, mainly in the north-east corner in an area around the Danish coast. A few cells in the centre of the North Sea and in the Thames Estuary and English Channel have shape parameter estimates which are slightly positive, but none are significantly greater than zero. The shape parameter estimates change fairly smoothly in space. This property is exploited in Section 6.

4. A nonparametric model for temporal trend

Simple models for temporal trend can easily be incorporated into the parameters of an extreme value model (Coles, 2001). Preliminary analyses using parametric polynomial and changepoint regression models suggest that standard parametric forms are not capable of adequately describing temporal trends in the extremes of surges generated by the CSX model. We adopt a nonparametric regression approach which lets us estimate trends under the much weaker assumption that the GEV model parameters vary smoothly in time, without making any more specific assumptions about the structural form of this variation. We adopt the local likelihood approach to nonparametric regression which has been adapted for the extreme value context by Hall and Tajvidi (2000) and Davison and Ramesh (2000).

We begin by presenting the general inferential procedure, and then outline how this is applied to the $r$ largest model. Let $l(\theta_{t_j}; z_j)$ denote the log-likelihood associated with an observation $z_j$ taken at time $t_j$ from a statistical model with parameter vector $\theta_{t_j}$, and
assume that \( z_1, \ldots, z_N \) are independent data that follow this model. Let \( K \) denote a kernel function which determines the weights given to each observation \( z_j \) in the estimation of \( \theta_t \). The bandwidth \( h \) of this function controls the temporal smoothness of \( \theta_t \). Finally, let \( R \) denote a vector-valued local regression function, which may depend upon parameters \( \tau_t \) and which must fulfill the constraint that \( R(0; \tau_t) = 0 \) for each component of the vector and for all \( \tau_t \). For any \( t \in [t_1, t_N] \) a local likelihood estimator for \( \theta_t \) can then be found by maximising the weighted sum of log-likelihood contributions,

\[
\sum_{j=1}^{N} K(t_j - t; h) l[\theta_t + R(t_j - t; \tau_t); z_j],
\]

with respect to \( \theta_t \) and \( \tau_t \).

For this section the data \( z_j = (z_j^{(1)}, \ldots, z_j^{(r)}) \) denote the \( r \) largest surge elevations in year \( t_j \) from a single location, where \( j = 1, \ldots, N \); we generalise to the spatio-temporal context in Section 6. We assume that \( z_j \) follows an \( r \)-largest distribution with parameter vector \( \theta_j = (\mu_j, \sigma_j, \xi_j) \). We follow Hall and Tajvidi (2000) and assume that the location and scale parameters \( \mu_t \) and \( \sigma_t \) are local linear in time whilst the shape parameter \( \xi_t \) is local constant in time. We adopt a Gaussian kernel function \( K \) centred on zero and with standard deviation \( h \). The resulting local log-likelihood function is

\[
\sum_{j=1}^{N} \phi \left( \frac{t_j - t}{h} \right) l[(\mu_t + \alpha_t(t_j - t), \sigma_t + \beta_t(t_j - t), \xi_t); z_j].
\]

The local likelihood estimator \( \hat{\theta}_t \) for \( \theta_t \) is obtained by maximising (3) with respect to the parameters of interest \( \theta_t = (\mu_t, \sigma_t, \xi_t) \) and the nuisance parameters \( \tau_t = (\alpha_t, \beta_t) \).

We quantify estimation uncertainty using the semiparametric bootstrap scheme of Davison and Ramesh (2000). The construction of bootstrap confidence intervals is complicated by the presence of bias in the local likelihood estimators (even asymptotically). We use standard methods to construct bands which capture variability in the parameter estimates without correcting for this bias (“variability bands”: Bowman and Azzalini, 1997), but it is important to note that these bands should be interpreted as confidence intervals about \( E(\hat{\theta}_t) \) rather than \( \hat{\theta}_t \).

The bandwidth \( h \) determines the smoothness of the fitted local regression model, with larger values of \( h \) corresponding to stronger levels of smoothing: using \( h = 0 \) means that the parameters \( \theta_{t_j} \) are estimated using only data from year \( t_j \), while taking \( h \to \infty \) makes the location and scale parameters linear in time and the shape parameter constant over time. For bandwidths of less than two years we find that the estimates of change from the CSX model data are noisy and are dominated by the influence of a few exceptionally large events. For relatively large bandwidths of more than around five years, the estimates change slowly and are predominantly indicative of long-term trends in surge elevations, with almost all decadal variations being removed. Intermediate bandwidths of around two to five years tend to reflect decadal variations, and indicate the factors driving longer term trends in behaviour.

Automatic methods of bandwidth selection are designed to determine the level of smoothing that is best supported by the available data. We attempted to use the likelihood cross validation criterion for this purpose, but found that it had poor performance in the context of our model. This may be related to the known difficulties associated with using the
LCV criterion for kernel density estimation in the presence of either outliers or heavy tails (Silverman, 1986).

A bandwidth of $h = 3.5$ years was used to estimate 50 year return levels shown in Figure 5 for all 259 grid cells in six years (selected to give a representative indication of trends over the whole period). A kernel with this bandwidth assigns 95% of weight to timepoints within a 14 year window. We adopt this bandwidth throughout Sections 4 to 6. To highlight the temporal variations we plot percentage deviations relative to the estimated return level for each cell based on the time homogeneous model of Section 3, stripping out information about spatial variations in the overall magnitude of storm surge elevations.

The graphs highlight three periods that have fairly distinct trends. Between 1955 and 1975 return levels show modest increases across the southern and eastern parts of the study region, but decrease slightly across the northern and western parts, the most rapid changes occurring between 1955 and 1962. Between 1975 and 1990 levels increase substantially in the west (along the whole length of the east coast of Britain), decrease in an area around the Dutch coast, and are relatively static throughout most of the north and west of the study region. The most dramatic changes take place between 1990 to 2000, with levels increasing substantially in the north and west whilst decreasing substantially in the south and east. These changes are most rapid along the east coast of Britain. Certain cells show trends which are distinct from these regional trends, most notably in the Wadden Sea, the Thames Estuary and the English Channel.

4.1. **Comparison of model output with observational data**

Our local regression model for extreme values can be used as a diagnostic check of whether changes in the characteristics of model-generated storm surges are synchronous with corresponding changes in observational data. Figure 2 shows estimates of 50 year return levels, together with 95% pointwise variability bands, under the fitted local regression models for observational data at Aberdeen, Lowestoft and Dover and for model-generated data at the nearest grid cells to these locations.

Estimates derived from model-generated data for Aberdeen show a decrease from 1955 to 1970 followed by a rapid increase from 1980 to 1995, resulting in a substantial overall increase over the period 1955-2000. The increase from 1980-1990 is reflected to a lesser degree in the (short) observational record at Aberdeen, but the observational data also contain an increase of similar magnitude over the period 1965-1975 that is not synchronous with any changes in the model-generated data. Return level estimates from model-generated data at Lowestoft increase during 1955-1960 and 1980-1990 and decrease during 1990-2000, leading to a slight overall increase for the period 1955-2000. These variations are highly synchronous with corresponding variations in the observational record at Lowestoft. Estimates from model-generated data at Dover increase from 1955-1960 and decrease from 1990-2000, showing no overall change for the period 1955-2000. These variations are less synchronous with corresponding variations in the observational record than at the other two sites.

Some of the discrepancies between the observational data and the CSX model are attributable to known issues with the observational record. For example, the Aberdeen tide gauge was re-sited in 1975 as it was producing faulty readings, so the increase in the observational data from 1965 to 1975 may well be spurious. Other discrepancies are attributable to known issues with the CSX model: the lack of synchronicity between model-generated and observational data at Dover probably results from the fact that the Straits of Dover are poorly resolved by a model that has a relatively coarse spatial resolution.
**Fig. 5.** Variations in estimates of 50 year return levels obtained by fitting the local $r$-largest model of Section 4 to surge elevations generated by the CSX model, using $r = 20$ observations per year and a bandwidth of $h = 3.5$ years. For each cell percentage deviations are shown relative to the estimate obtained using the time homogeneous model of Section 3, for each of six selected years.
4.2. Graphical diagnostics for the statistical model

A different aspect of model diagnosis involves checking that the local regression model itself provides an adequate fit to the surge data, whether model-generated or observational. Quantile-Quantile (Q-Q) plots are a widely used graphical tool for comparing the distributional properties of a fitted statistical model against the empirical distribution of the data. The quantile scale acts to focus attention on the performance of the model at extreme levels, so Q-Q plots are particularly appropriate for assessing the fit of extreme value models.

The construction of the Q-Q plot is complicated by the presence of temporal nonstationarity, and we can either overcome this by transforming the model and data onto a standard scale, such as a uniform scale, or by transforming both the model and data onto the original scale of the data. More specifically, we can either plot

\[ H_k \left( z_j^{(k)}; \hat{\theta}_t \right) \text{ against } \hat{\zeta}_j^{(k)}/(N + 1). \]  

or

\[ z_j^{(k)} \text{ against } H_k^{-1} \left( \hat{\zeta}_j^{(k)}/(N + 1); \hat{\theta}_t \right), \]  

where \( \hat{\zeta}_j^{(k)} \) denotes the rank of \( H_k(z_j^{(k)}; \hat{\theta}_t) \) in the sequence

\[ H_k(z_1^{(k)}; \hat{\theta}_{t_1}), \ldots, H_k(z_N^{(k)}; \hat{\theta}_{t_N}). \]

Note that (5) can be obtained from (4) by applying \( H_k(\bullet; \hat{\theta}(t_j)) \) to both axes of the graph. The two plots highlight distinct aspects of model fit, but only the latter can be used to compare models with different levels of complexity because the model information is lost in the standardisation procedure used to create the former graph. In Figure 6 we use (5) to compare the performance of the local regression model against the \( r \) largest model from Section 3 in which the parameters are constant over time. The local regression model produces consistently more accurate estimates than the time homogeneous model, since the points lie closer to the diagonal. It also has greater estimation uncertainty, represented by the wider variability bands. The local regression model consequently exhibits adequate performance for the years and grid cells shown, whereas the local constant model shows evidence of inadequate fit for the synthetic (CSX model generated) surge data at Aberdeen and Lowestoft.

5. Incorporating structural constraints

Local likelihood methods are intuitively simple, and relatively straightforward to apply and to extend to non-standard situations. However, they tend to be less flexible than the alternative approach to nonparametric regression based on fitting smoothing splines via penalised likelihood. In particular, local regression models cannot explicitly control the degree of smoothing that is applied to each parameter, because smoothing takes place within the likelihood space rather than the parameter space. The local modelling approach, can, however, be made more flexible by adapting it to allow for structural constraints within the parameter space, for example allowing some parameters to remain fixed over time. The possibility of smoothing only a subset of the parameters was noted by Hall and Tajvidi (2000).

Assume that the parameters \( \theta_t = (\lambda_t, \delta) \) of a model can be partitioned in terms of a set of parameters \( \lambda_t \) which we wish to smooth over time and a set of parameters \( \delta \) whose
Fig. 6. Quantile-Quantile plots to assess the fit of the time homogeneous $r$-largest model of Section 3 (left) and the local $r$-largest model of Section 4 (right) to surge elevations generated by the CSX model for three ports on the British east coast. Plots shown are for the annual maxima ($k = 1$); 95% variability bands are also shown (grey lines).
values are time-homogeneous. We present an iterative procedure for optimising the local likelihood, similar to that used by Hall and Tajvidi (2000), Pauli and Coles (2001) and Chavez-Demoulin and Davison (2005), which alternates between (a) estimating $\lambda_t$ by local likelihood whilst holding the value of $\delta$ fixed and (b) estimating $\delta$ by maximum likelihood whilst holding the value of $\lambda_t$ fixed. The procedure is initialised by estimating $\lambda_t$ and $\delta_t$ simultaneously using maximum likelihood, assuming that $\lambda_t = \lambda$ (i.e. without the use of any smoothing). For example, if we require the shape parameter to be time-homogeneous then the objective function is the local likelihood
\[
N \sum_{j=1}^{N} \phi \left( \frac{t_j - t}{h} \right) l[(\mu_t + \alpha_t (t_j - t), \sigma_t + \beta_t (t_j - t), \xi); z_j].
\] (6)
If the location parameter contains a linear covariate $A_t$ with a time homogeneous coefficient $\rho$, so that $\mu_t = \mu'_t + \rho A_t$, then the local likelihood is
\[
N \sum_{j=1}^{N} \phi \left( \frac{t_j - t}{h} \right) l[(\mu'_t + \alpha_t (t_j - t) + \rho A_t, \sigma_t + \beta_t (t_j - t), \xi); z_j].
\] (7)

We now examine a variety of different structural constraints that let us interrogate aspects of the process that are of particular interest.

5.1. Disentangling changes in magnitude of extreme events from changes in frequency
The analysis of Section 4 identified changes over time in the characteristics of storm surge elevations. We now assess the extent to which these changes can be attributed to changes within the individual parameters of the extreme value model, $\mu_t$, $\sigma_t$ and $\xi_t$, by considering submodels in which certain of these parameters are held constant over time. This approach allows us to determine whether the changes are due to simple changes in the underlying surge process.

The first submodel considered allows the location and scale parameters $\mu_t$ and $\sigma_t$ to vary freely, but constrains the shape parameter to be constant over time ($\xi_t = \xi$), and amounts to an assumption that the tail properties of the hourly surge process are constant in time. The model is fitted by maximising local likelihood (6). It is common in extreme value analyses to make the assumption of constant shape parameter (Coles and Tawn, 1990), corresponding to the assumption that changes over time (or space) in the statistical properties of the underlying process result only from shifts in the location and scale parameters of that process.

Our second submodel also constrains $\xi$ to be constant over time, but now imposes the additional constraint that the ratio $\eta = \sigma_t / \mu_t$ is constant over time. This model imposes an assumption that trends over time in any two different return levels $T_1$ and $T_2$ are parallel on the log scale, so that the ratio $S_{T_1} / S_{T_2}$ remains constant over time. Such a model for the extremes would arise if variations in the statistical properties of the underlying process result solely from a scaling up or down in the overall magnitude of that process. It has been widely hypothesised that the key impact of climate change in many environmental systems will be to increase the level of variability. This submodel provides a basis for testing the hypothesis that changes in the characteristics of extreme events over time can be attributed solely to changes in the variability of the underlying process.
Fig. 7. Estimated 50 year return levels from fitting the constrained local regression models of Section 5.1 (left hand side) and Section 5.2 (right hand side) to surge elevations generated by the CSX model for three ports on the British east coast. For the model of Section 5.1 we show estimates obtained by constraining the shape parameter $\xi_t$ to be constant over time (red), along with corresponding 95% variability bands (pink). For the model of Section 5.2 we show estimates obtained by incorporating the MLE of the effect of NAO into the location parameter as an offset term (red), together with estimates obtained by incorporating the upper and lower endpoints of a 95% confidence interval for the effect of the NAO as offset terms (pink). We also, for the purposes of comparison, include all of the information from Figure 2.
In Figure 7 (left hand side) we apply the shape parameter constrained submodel to CSX-generated surge data for Aberdeen, Lowestoft and Dover, and compare the results against those obtained using the unconstrained model of Section 4. The results that we present are based upon a single iteration of the inferential procedure, but we have checked that similar results will be obtained by running the procedure for a larger number of iterations. We see that there are fairly substantial variations between the constrained model and the unconstrained model, with the estimates from the unconstrained model frequently lying close to the edge of the 95% variability band associated with the constrained model and sometimes lying outside this band. Estimates for the second model in which $\xi_t$ and $\sigma_t/\mu_t$ are both constrained (not shown) are similar to those obtained using the model in which only the shape parameter $\xi_t$ is constrained, with the differences between the two constrained models being substantially smaller than the differences between the constrained and unconstrained models. We conclude from the analysis that trends over time are complex, probably resulting from distinct trends in all three of the extreme value parameters, and that simpler models which hold some of these parameters fixed are not capable of reproducing the qualitative features of the variations that are described by the model from Section 4.

5.2. Accounting for the North Atlantic Oscillation

Can variations over time in the magnitude and frequency of model-generated storm surges be attributed to established modes of climate variability, such as the North Atlantic Oscillation (NAO)? The NAO index quantifies the average normalised pressure difference between Reykjavik (Iceland) and either Ponta Delgada (the Azores) or Gibraltar, and represents an established mode of large-scale climate variability within the North Atlantic. Monthly and annual indices for the NAO are produced each year by the Climatic Research Unit (Norwich, UK) and indices for winter months are known to be strongly correlated to levels of storm activity in the North Atlantic (e.g. Rogers, 1997).

For each grid cell of the CSX model we account for the effects of the NAO using a simple parametric model in which the location parameter of the $r$ largest model is a linear function of the NAO index $A_t$. The NAO index is a common covariate for all sites, but we already have seen that trends at different sites are asynchronous. We therefore allow for site-specific trends by assuming that $\mu_t = \mu'_t + \rho A_t$, where $\mu'_t$ is expected to vary smoothly over time. The model is fitted by maximising local likelihood (7), so that we smooth over the parameters $(\mu'_t, \sigma_t, \xi_t)$ rather than over $(\mu_t, \sigma_t, \xi_t)$.

We investigate the impact of the NAO upon model-generated surge elevations at each grid cell in the North Sea by regressing the location parameter of the $r$-largest model (with $r = 20$) onto both annual and monthly indices for the NAO. We find a strong positive relationship between extreme model-generated surges and the annual NAO index, with the parameter estimates $\hat{\rho}$ being statistically significant at the 95% level for around 90% of grid cells within the study area. There are also strong positive relationships with NAO indices for each of the individual winter months from December to March. The local regression estimates of 50 year return levels derived using a bandwidth of $h = 3.5$ years are, however, similar to those obtained in Section 4 without accounting for the NAO (Figure 7, right hand side). One plausible interpretation is that extreme surge levels exhibit variations over a range of temporal scales, and that the relatively short term variations which are attributable to the NAO are distinct from the decadal and long-term variations explored in Sections 4 and 5.1.
6. Spatio-temporal modelling

Spatial information can be used to improve the precision of extreme value parameter estimates (Coles and Tawn, 1990), and, in particular, to improve the precision of estimated trends in the characteristics of extreme events (Dixon et al., 1998). Storm surge elevations at nearby locations are known to be strongly dependent, even at extreme levels. The dependence arises because nearby sites tend to (a) have similar physical characteristics and therefore be affected by storms with similar characteristics, and (b) be affected by the same storm surges events. We deal with the former source of dependence by modelling the parameters of the extreme value model as a smooth function of spatial location. We use a resampling scheme to account for the residual spatial dependence that is induced by the latter source.

We allow for smooth spatial variations in the parameters by extending the local regression model of Section 4 to the spatio-temporal context. Recall from Figure 8, however, that surge elevations show a high degree of spatial variation across the North Sea, resulting predominantly from variations in water depth. We do not wish to smooth these well-established, and relatively well-estimated, “average” patterns of spatial variation, but rather to smooth spatially the more subtle temporal trends in storm surge behaviour which may result from changes in storm activity. We therefore impose an assumption of local constancy in time and space over appropriately transformed, scale-free, versions of the extreme value parameters, with inference based on an analogous procedure to that which we introduced for the temporal context in Section 5. The idea of using reparameterisation to achieve different levels of smoothing in different parameters of a local regression model appears to be novel, and potentially has wider applicability.

Let \( z_{ij} = (z_{ij}^{(1)}, \ldots, z_{ij}^{(r)}) \) denote the \( r \) largest surge elevations for time \( t_j \) at location \( s_i \), where we have data for \( j = 1, \ldots, N \) years and \( i = 1, \ldots, D \) sites. For simplicity we assume here that space is one-dimensional, so that \( s_i \) is scalar, but the approach is easily extended to the situation in which spatial location is specified using two coordinates. We assume that \( z_{ij} \) follows an \( r \)-largest model with parameters \( \mu_{t_j,s_i}, \sigma_{t_j,s_i} \) and \( \xi_{t_j,s_i} \), and adopt a three stage procedure for the estimation of these parameters:

(a) For each site \( I = 1, \ldots, D \) maximise

\[
\sum_{j=1}^{N} l((\mu_{s_I}, \sigma_{s_I}, \xi_{s_I}); z_{Ij})
\]

in order to obtain maximum likelihood estimates \( \hat{\mu}_{s_I}, \hat{\sigma}_{s_I} \) and \( \hat{\xi}_{s_I} \) based on the time homogeneous model of Section 3.

(b) Maximise the weighted sum of log-likelihoods

\[
\sum_{I=1}^{D} \sum_{j=1}^{N} \phi \left( \frac{s_I - s_i}{\nu} \right) \phi \left( \frac{t_j - t_j}{h} \right) l((\mu_{s_I}, \gamma_{t_j,s_i}, \sigma_{s_I}, \omega_{t_j,s_i}, \xi_{t_j,s_i}); z_{Ij})
\]

to obtain estimates \( \hat{\gamma}_{t_j,s_i}, \hat{\omega}_{t_j,s_i} \) and \( \hat{\xi}_{t_j,s_i} \) for the parameters \( \gamma_{t_j,s_i}, \omega_{t_j,s_i}, \) and \( \xi_{t_j,s_i} \).

(c) Estimate \( \mu_{s_i,t_j} \) by \( \hat{\mu}_{s_i}, \hat{\gamma}_{t_j,s_i}, \) estimate \( \sigma_{s_i,t_j} \) by \( \hat{\sigma}_{s_i} \), \( \hat{\omega}_{t_j,s_i} \), and estimate \( \xi_{s_i,t_j} \) by \( \hat{\xi}_{t_j,s_i} \).

The role of the \( \hat{\mu}_I \) and \( \hat{\sigma}_I \) terms is to ensure that the assumption of local constancy over space is approximately valid for the parameters \( \gamma_{t_j,s_i} \) and \( \omega_{t_j,s_i} \). Note that this approach
is a spatio-temporal analogue of the approach that we introduced in Section 5 for dealing with structural constraints on the parameter space. It is straightforward to quantify uncertainty within the model by bootstrapping the data over time whilst keeping the spatial structure fixed, thus ensuring that the resampled datasets preserve the spatial structure of the observed data (Davison and Ramesh, 2000).

We apply our spatio-temporal model to a subset of the CSX model data, restricting attention to cells along the East coast of Britain. The level of spatial smoothing cannot be selected using standard statistical techniques for bandwidth selection because of the presence of residual spatial dependence (Opsomer et al., 2001). We therefore compute the estimates for various, relatively modest, degrees of spatial smoothing, and plot the resulting trends in Figure 8 for selected coastal locations (with a fixed temporal bandwidth of \( h = 3.5 \) years). Estimates of the temporal variations at each specific location become more precise as the degree of spatial smoothing is increased (not shown), and, as we would expect, the estimated trends at the different sites also tend to become more similar. The qualitative nature of the temporal variations at each site remains unchanged by the process of spatial smoothing, however.

7. Discussion

We have outlined an approach for using extreme value methods to analyse variations in the magnitude and frequency of storm surges generated by a deterministic numerical model, and have used the methodology in order to provide a description of changes in North Sea storm surge characteristics over the second half of the twentieth century. Our approach is based on using a local regression model to describe temporal trends in the parameters of an extreme value model under a weak assumption of smoothness. The analysis reveals subtle long-term and decadal variations that could not easily be described using standard parametric models, and demonstrates the impact of the NAO on these variations. The development of nonparametric regression models for the extreme value context remains an active area of research, and more sophisticated approaches such as Bayesian state space modelling (Gaetan and Grigoletto, 2004) could offer alternative insights into the changing characteristics of extreme values.

Environmental models typically generate data on a regular spatial grid. We incorporated assumptions of spatial smoothness into our analysis by adapting the local regression model for the spatio-temporal context, whilst using a bootstrap scheme to correct for the effects of residual spatial dependence upon uncertainty estimation. Latham (2006) performed an alternative spatial analysis of the same data, based on constructing an explicit model to describe the dependence of extreme surges at nearby locations, and it would be interesting to see if this approach could also be extended to allow for the estimation of temporal trend.

The results of our analysis are predicated on the assumption that the behaviour of a single run from the CSX model accurately reflects the behaviour of real storm surges. This is clearly naïve: the meteorological inputs to the model are based on observational data and are consequently uncertain; the model requires aggregation of surge characteristics onto a relatively coarse grid that ignores the effects of local bathymetry and estuarine effects; and the model provides an approximate and imperfect description of the true physical processes which underpin surge dynamics. We have partially validated our results through a comparison with independent observational data on sea levels, but further progress depends upon systematically identifying and quantifying the sources of bias and uncertainty within
Fig. 8. Variations in estimates of 50 year return levels from fitting the spatio-temporal model of Section 6 to cells along the British east coast (left) and along the coast of the European mainland (right), for a temporal bandwidth of $h = 3.5$ years and a range of spatial bandwidths, $v = 1, 2$ and $3$ cells. Percentage deviations are relative to the estimates obtained using the time homogeneous model of Section 3, as in Figure 5. The letters refer to the locations of selected ports, as in Figure 3.
the storm surge model itself.

Acknowledgements

Butler was funded by the Engineering and Physical Sciences Research Council and the Beauclerk Memorial Fund, and undertook this work whilst studying for a doctorate at Lancaster University. The work of Heffernan and Tawn was funded through the EPSRC-NERC Environmental Mathematics and Statistics initiative. Barry Rowlingson, David Blackwell, Kevin Horsburgh and Mark Latham provided valuable assistance with various aspects of the analysis. Tide gauge data were provided by the British Oceanographic Data Centre, Rijksinstituut voor Kust en Zee and Danmarks Meteorologiske Institut, meteorological data by the Norwegian Meteorological Institute, and NAO indices by the Climatic Research Unit.

References


