Extreme value analysis of decadal variations in storm surge elevations

Adam Butler\textsuperscript{a,*}, Janet E. Heffernan\textsuperscript{b}, Jonathan A. Tawn\textsuperscript{b}, Roger A. Flather\textsuperscript{c}, Kevin J. Horsburgh\textsuperscript{c}

\textsuperscript{a}Biomathematics \\& Statistics Scotland, Kings Buildings, Edinburgh EH9 3JZ, UK
\textsuperscript{b}Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK
\textsuperscript{c}Proudman Oceanographic Laboratory, 6 Brownlow Street, Liverpool L3 5DA, UK

Abstract

We use a novel statistical approach to analyse changes in the occurrence and severity of storm surge events in the southern and central North Sea over the period 1955-2000, using 1) output from a numerical storm surge model and 2) \textit{in situ} data on surge levels at sites for which sufficiently long observational records are available. The methodology provides a robust diagnostic tool for assessing the ability of models to reproduce the observed characteristics of storm surges, in a fashion that properly accounts both for variability within a single long run of the model and for recording error in observational data on sea levels.

The model re-analysis data show strong positive trends in the frequency and severity of storm surge events at locations in the north-eastern North Sea, whilst trends at locations in the southern and western North Sea appear to be dominated by decadal variability. Trends in \textit{in situ} data for sites along the North Sea coast are fairly synchronous with corresponding trends in the model re-analysis data, with the most serious discrepancies attributable to known problems in the observational...
1 Introduction

Climatologists have hypothesised that climate change may be leading to increased storminess within the NE Atlantic, and a string of empirical studies (Schinke, 1993; Schmidt & von Storch, 1993; von Storch et al., 1993; Stein & Hense, 1994; Lambert, 1996; Schmith et al., 1998; Alexandersson et al., 1998, 2000) have investigated whether the storm climate of the NE Atlantic has actually changed over the past century. The results of these studies are somewhat inconclusive, but there is some suggestion that the magnitude and frequency of storms may indeed have increased over the past century, and especially during the past thirty years or so.

There is virtually no empirical evidence, however, for any corresponding increase in storm surge elevations (Dixon & Tawn, 1992; Bijl et al., 1999). This absence of evidence should not necessarily be interpreted as an evidence of absence, since (Pugh & Maul, 1999) it would probably be difficult to detect even genuine trends within the relatively limited in situ data that are available. Trends in observational data may also be dominated by highly localised effects, especially since tide gauges are often based in estuarine ports which

---

* Corresponding author

Email address: adam@bioss.ac.uk (Adam Butler).
are susceptible to intensive site management (e.g. dredging).

Flather et al. (1998) show that, at least within the North Sea, numerical storm surge models are capable of generating outputs which have similar extreme characteristics to those within observational data, suggesting that surge models potentially can be used to study long-term changes in storm surge behaviour. This is the approach which we follow in this paper.

Langenberg et al. (1999) use output from a dynamical model of the North Sea to analyse trends in extreme high water levels, and find increasing trends along much of the European coast (but no significant trends along the British east coast). Using a hindcast of the surge climate for the period 1955-2000 – which is simply an updated version of that used by Flather et al. (1998) – we investigate whether the frequency and magnitude of storm surge events has changed over the past fifty years. Unlike Langenberg et al. (1999), who use winter high water levels, we concentrate solely upon the surge process (the observed sea level minus the tidal predictions), and use advanced statistical techniques to account for variability in the characteristics of simulated storm surge events. The methodology that we advocate has general applicability to the detection of changes in the magnitude and frequency of storm surge elevations, and of other quantities related to extreme sea levels.

We begin in Section 2 by briefly describing the numerical storm surge model and observational data which we use. In Section 3 we introduce the statistical methods which we use to analyse our data: we begin by giving an introduction to the area of extreme value theory, and then outline how extreme value methods can be used to estimate temporal trends in storm surge characteristics. We present the results of our analysis in Section 4, and then conclude with a
2 Observational and model re-analysis data

2.1 Model data: the CSX-DNMI hindcast

Numerical models for surge dynamics have improved rapidly over recent years (Bode & Hardy, 1997). The CSX model (Flather et al., 1991) was developed by the Proudman Oceanographic Laboratory (POL) as a two dimensional storm surge model for the north-west European continental shelf (see Figure 1). The CSX model is evaluated across a regular grid with a spatial resolution of 20’ in latitude and 30’ in longitude (approximately 35km by 35km). The model was used as an operational model for storm surge prediction from 1983 until 1991, but has since been superseded by models with higher spatial resolution.

Flather et al. (1998) created a hindcast of the surge climate for the period 1955-1992 (the “CSX-DNMI hindcast”) by forcing the CSX model using a high-quality hindcast of the meteorology for this period - the DNMI hindcast, created by Det Norske Meteorologiske Institut (Reistad & Iden, 1995), and used in WASA (1998). There are problems with this hindcast in areas of data sparseness, but it provides a good representation of the surge climate within the relatively data-rich areas of the southern and central North Sea that are of interest to us (Langenberg et al., 1999).

Atmospheric pressure and wind values from the DNMI hindcast were converted into wind stresses by interpolating atmospheric pressure onto CSX grid,
and then applying the relationship given in Smith & Banke (1975). Tidal forcing for the CSX model included eight dominant periodic constituents (Q1, O1, P1, K1, N2, M2, S2 and K2), and allowed for 18.6 year variations in amplitude and phase. We make use of the same surge hindcast, now updated to cover the period 1955-2000.

We restrict attention to that part of the CSX grid which covers the southern and central North Sea. Our model re-analysis data consist of 403248 hourly surge residuals for each of the 259 grid cells within this region. Surge residuals are computed by deducting the levels obtained from running the CSX model with tidal only forcing from those obtained by running the CSX model with both tidal and meteorological forcing; surge residuals thereby automatically incorporate the effects of tide-surge interaction.

2.2 Observational data

We are interested, where possible, in comparing temporal trends in surge elevations from the model re-analysis data against corresponding trends in in situ records. Observational data are susceptible to systematic recording errors, however, and may also be dominated by highly localised trends (resulting, for example, from changing patterns of dredging). The process of removing tidal signals from observational sea level records in order to produce surge records can also be highly sensitive to any slight inaccuracies in the phase of the predicted tidal signal. It follows that asychroneities between observational and model re-analysis data are likely to be attributable to model inadequacy if they occur consistently across a number of sites, whilst discrepancies which occur only at a single site will typically be attributable to problems with the
observational record at that site.

The observational data which we use are derived from tide gauges at four locations along the British east coast (Aberdeen, Lowestoft, Southend and Dover), cover the period 1955-1992, and are available through the British Oceanographic Data Centre. Note that we have restricted attention only to the few sites for which sea level records are available for 20 or more of the years in the period 1955-1992. Dominant tidal signals have been removed from the observed sea level records using harmonic analysis, but underlying isostatic and eustatic trends in mean sea level have not been removed.

3 Statistical methodology

We analyse surge elevations from the CSX-DNMI hindcast, and, where available, from *in situ* tide gauge data, using statistical models from the field of extreme value analysis. Extreme value analysis is a widely-used statistical methodology (Coles, 2001) for drawing inferences about the extremes of a stochastic process using only data on relatively extreme values of that process. Using a statistical approach allows us to describe temporal trends in storm surge characteristics, whilst properly accounting for the effects of natural variability. Further details of the statistical methodology are given in Butler et al. (2006).

3.1 The generalised extreme value (GEV) model

Assume that we have \( n \) years of hourly surge data, \( t_1, \ldots, t_n \), and let \( x_{j}^{(1)} \) denote the annual maximum surge for year \( t_j \). A standard procedure would
be to assume that the annual maxima \( x_{1}^{(1)}, \ldots, x_{n}^{(1)} \) follow a GEV distribution with distribution function

\[
F(x^{(1)}; \theta) = \exp \left[ - \left( 1 + \xi \left( \frac{x^{(1)} - \mu}{\sigma} \right) \right)_{+}^{-1/\xi} \right],
\]

with parameter vector \( \theta = (\mu, \sigma, \xi) \). The three parameters correspond to a location parameter (\( \mu \)), a scale parameter (\( \sigma \)), and a shape parameter (\( \xi \)). The model is motivated by powerful mathematical theory (Leadbetter et al., 1983) describing the behaviour of sequences of random variables at asymptotically high levels, and provides a robust basis for statistical inferences about the magnitude and frequency of extreme events. We can estimate the parameters of the model using, for example, maximum likelihood, and can thereby obtain estimates for more easily interpretable quantities. In particular, the \( N \)-year return level

\[
q(N) = \mu - \frac{\sigma}{\xi} \left[ 1 - \left\{ - \log \left( \frac{N - 1}{N} \right) \right\}^{-\xi} \right],
\]

corresponds to the level which is exceeded with probability \( 1/N \) in any particular year, and is used as a design parameter in coastal engineering applications.

### 3.2 The \( r \)-largest value model

The GEV model makes highly inefficient use of the available data, since it only allows us to use one observation per year. An extension of this model allows us to use the \( r \) largest values per year, \( x = (x^{(1)}, \ldots, x^{(r)}) \), where \( r \geq 1 \) is chosen \emph{a priori} (see Section 3.4). The joint probability density function for
the $r$-largest model is of the form

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

where the parameters $\mu$, $\sigma$ and $\xi$ correspond exactly to the parameters of the GEV model.

If we assume that the parameters of the model are constant over time, then maximum likelihood estimates for these parameters can be obtained by maximising the loglikelihood $P_{n \in [1]} \log f(x_j; \theta)$ over $\theta$, where the vector $x_j$ contains the $r$-largest values for year $t_j$. Note that the loglikelihood is a sum of contributions from each year due to an assumption of independence of extreme surges from year to year.

We need to ensure that the $r$-largest values which we select for a particular year are drawn from distinct (and statistically independent) storm events. We adopt the declustering algorithm of Tawn (1988), which sequentially selects the $k = 1, \ldots, r$ largest value within a year under the restriction that the $k$-th largest value must lie at least $s$ hours away from each of the 1st, $\ldots, (k-1)$-th largest values. The value of $s$ must also be chosen a priori (see Section 3.4).

The $r$-largest model provides an alternative to the more usual ‘POT’ approach of fitting the generalised pareto distribution (GPD) to peak exceedances of a high threshold $u$. The POT approach cannot easily be used in the presence of temporal and spatial variability, because a separate threshold must be selected for each year and site. The $r$-largest approach is, in contrast, robust to temporal and spatial variations, because it relies upon a purely relative definition of what constitutes an extreme value. For example, we would need to take account of temporal trends in mean sea level when choosing a value of
u, whereas the definition of an extreme in the \( r \)-largest model is automatically invariant to the effect of any such trend.

3.3 Local regression models

We can model the parameters of an extreme value model as a function of time (Coles, 2001; Katz et al., 2002). Let \( x_j^{(k)} \) denote the \( k \)-th largest surge elevation in year \( t_j \), where \( k \leq r \), and assume that \( x_j = (x_j^{(1)}, \ldots, x_j^{(r)}) \) follows an \( r \)-largest distribution with parameter vector \( \theta_j = (\mu_j, \sigma_j, \xi_j) \).

Parametric regression models would impose a specific structure upon the form of the temporal trend in the extreme value parameters (i.e. how \( \theta_j \) changes with \( t_j \)), whereas nonparametric regression models impose only the - relatively much weaker - assumption that the parameter values \( \theta_j \) vary smoothly as functions of time \( t_j \). The past few years have seen the rapid development of a nonparametric regression methodology within the context of extreme value modelling (Hall & Tajvidi, 2000; Davison & Ramesh, 2000; Pauli & Coles, 2001; Gaetan & Grigoletto, 2004; Chavez-Demoulin & Davison, 2005), but there have so far been few substantive applications of these methods in the scientific literature. In this paper we adopt the widely-used local likelihood approach to nonparametric regression (Tibshirani & Hastie, 1987; Fan et al., 1998), using and adapting the methodology developed by Davison & Ramesh (2000) and Hall & Tajvidi (2000).

The key idea of the local likelihood approach is that we can estimate the parameters \( \theta_j = (\mu_j, \sigma_j, \xi_j) \) for year \( t_j \) by maximising a weighted sum of log-likelihood contributions from all timepoints \( t_1, \ldots, t_n \). Weights are determined
by a kernel function, $K(t-t_j; h)$, which attributes greater weight to time points that are close to $t_j$ than to time points which are distant from $t_j$.

We follow Hall & Tajvidi (2000) in adopting a model which is locally linear in the location and scale parameters, but locally constant in the shape parameter. For year $t_j$ we numerically maximise the resulting local loglikelihood function

$$\sum_{j=1}^{n} K(t_j - t_j; h) \log f(x_j; [\mu_j + \alpha_j(t_j - t_j), \sigma_j + \beta_j(t_j - t_j), \xi_j])$$

over the parameter vector $(\mu_j, \sigma_j, \xi_j, \alpha_j, \beta_j)$. The values of $\mu_j$, $\sigma_j$ and $\xi_j$ associated with this maximum provide a local likelihood estimator $\hat{\theta}_j(h)$ for $\theta_j$. The regression parameters $\alpha_j$ and $\beta_j$ are nuisance parameters which quantify the rate at which the parameters $\mu_j$ and $\sigma_j$ change over time. We take the kernel function $K(t-t_j; h)$ to be a Gaussian density function with mean 0 and standard deviation $h$, so that the bandwidth $h$ determines the overall degree of smoothing.

### 3.4 Selection of $r$, $s$, and $h$

Our statistical model relies on us adopting suitable choices for the number of storms per year, $r$, the minimum storm separation, $s$, and the bandwidth, $h$.

The asymptotic justification for the $r$-largest model relies upon $r$ being small relative to the number of observations within a year. Larger values of $r$, however, enable us to use more data, and so lead to more efficient (i.e. less uncertain) estimates for the parameters of the model. By plotting estimates of the model parameters as a function of $r$ - a 'parameter stability plot' (Coles, 2001) - we have a graphical tool for selecting the largest value of $r$ below which the parameter estimates appear to be stable, indicating that below this value the
asymptotic model is approximately valid. We adopt a value of \( r = 20 \) throughout this paper, based on parameter stability plots for model re-analysis data at around 25 arbitrarily selected sites in the North Sea. Note that analyses based on smaller values of \( r \) - such as the value \( r = 7 \) used by Flather et al. (1998) - yield parameter estimates which are fairly consistent with, but substantially more variable than, those obtained using \( r = 20 \).

Trial analyses (based on the same sites as above) suggest that estimates for the extreme value parameters are largely insensitive to the value of \( s \), so long as this is taken somewhere in the range 20-40 hours. We adopt \( s = 25 \) hours, slightly smaller than the value \( s = 34 \) used by Flather et al. (1998).

The bandwidth \( h \) determines the smoothness of the local regression model for temporal trend - with larger values of \( h \) corresponding to stronger levels of smoothing - and the choice of \( h \) is an important aspect of model choice. We adopt a bandwidth of \( h = 3.5 \) years throughout this paper; a kernel with this bandwidth assigns approximately 95% of weight to a 14 year window around the year of interest. This choice of bandwidth reflects both the objectives of our analysis, and our prior knowledge of the physical system: we wish to smooth out year-to-year variations in storm surge behaviour, and to concentrate attention upon the decadal and longer term trends which might result from gradual changes in levels of storminess. A range of automatic criteria for selecting \( h \) are available, but preliminary results using the likelihood cross validation criterion suggest that these will tend to select values of the bandwidth (\( h = 10 - 20 \) years) which are so large that they will inevitably smooth out the decadal-level variations which we are predominantly interested in detecting. We have fitted our model using a range of alternative bandwidths (\( h = 2, 5 \) and 10 years). We find that the detailed results of our analysis depend reasonably
strongly upon the particular bandwidth that we have adopted, for the reasons that we have just outlined, but that the qualitative forms of the long-term trends in storm surge characteristics that we detect are largely insensitive to the choice of bandwidth.

3.5 Quantifying uncertainty & assessing model performance

Bootstrapping is a general methodology for computing the distribution of an estimator via simulation (Davison & Hinkley, 1997). We use the semi-parametric bootstrap scheme of Davison & Ramesh (2000) to quantify uncertainties in the local likelihood estimators $\hat{\theta}_1(h), \ldots, \hat{\theta}_n(h)$. The scheme involves:

(1) for each year $t_j$ transforming the data onto a uniform scale via

$$u_j^{(1)} = F_1(x_j^{(1)}; \hat{\theta}_j(h)), \ldots, u_j^{(r)} = F_r(x_j^{(r)}; \hat{\theta}_j(h)),$$

where $F_k(x^{(k)}; \theta)$ denotes the marginal distribution function for $x^{(k)}$;

(2) for each $j = 1, \ldots, n$ taking $\phi_j$ to be a random integer between 1 and $n$;

(3) for each year $t_j$ transforming back onto the original scale via

$$\tilde{x}_j^{(1)} = F_1^{-1}(u_{\phi_j}^{(1)}; \hat{\theta}_j(h)), \ldots, \tilde{x}_j^{(r)} = F_r^{-1}(u_{\phi_j}^{(r)}; \hat{\theta}_j(h)),$$

where $F_k^{-1}$ denotes the inverse of $F_k$;

(4) fitting the local regression of Section 3.3 to the resampled data $\tilde{x}_1, \ldots, \tilde{x}_n$.

This procedure is repeated $B$ times, allowing us to obtain $B$ independent estimates of $\theta_1(h), \ldots, \theta_n(h)$. If the number of bootstrap samples $B$ is sufficiently large then, under weak assumptions, the empirical distribution of these resampled estimates will provide a good approximation to the true distribution.
of $\hat{\theta}_1(h), \ldots, \hat{\theta}_n(h)$. The estimates, may, for example, be used to construct ‘variability bands’ (Bowman & Azzalini, 1997). Such bands are designed to represent uncertainty within the parameter estimates - they are similar to confidence intervals, except that they make no attempt to adjust for the bias involved in estimation.

Quantile-Quantile (Q-Q) plots are a widely used diagnostic 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 statistical model at extreme levels, so Q-Q plots are particularly appropriate for assessing the fit of extreme value models. The parameters of our $r$-largest model vary over time, so we must standardise the fitted distributions for different years onto a common scale before assessing model fit. For a specific value of $k$ we can construct a Q-Q plot by plotting

$$r_j^{(k)} \text{ against } F_k^{-1} \left( \frac{j}{n+1} ; \hat{\theta}_j(h) \right) \text{ for } j = 1, \ldots, n,$$

where $k \leq r$, and where $r_j^{(k)}$ denotes the $j$-th largest value in the sequence $x_1^{(k)}, \ldots, x_n^{(k)}$.

4 Results and discussion

In this section we report the results from fitting our local $r$-largest model to model re-analysis and observational surge data for the period 1955-2000. The statistical model which we use is local linear in the location and scale parameters and local constant in the shape parameter, uses $r = 20$ observations per year, and has a bandwidth of $h = 3.5$ years. Further details of the analyses and results can be found in Butler (2005).
Estimates for 50 year return levels from fitting our local regression model to model re-analysis and observational data at Aberdeen, Lowestoft and Dover are shown in Figure 2, together with 95% pointwise variability bands. Note that the model re-analysis data actually relate to the nearest grid cell to the named port, and not to any specific point location. The modelled elevation is the value over a grid box with an area of approximately $1200 \text{ km}^2$ and this will generally differ from a corresponding tide gauge value at a specific location. Consequently the model does not reproduce local variability in fine detail but comparisons (e.g. Flather et al 1998) show that the DNMI-CSX model reproduces surge peak and variability well at open coastal locations. It is these large scale variabilities that are of interest here.

Return levels for model re-analysis data at Aberdeen show a decrease from 1955 to 1970 followed by a rapid increase from 1980 to 1995, resulting in an overall increase over the period 1955-2000. The increase from 1980-1990 is reflected to a lesser degree in the (short) observational record at this location. The observational data also record an increase in storm surge elevations of similar magnitude over the period 1965-1975, but this is likely to be a spurious trend resulting from the re-siting of the gauge in 1973.

Return levels from the model re-analysis data for Lowestoft increase during 1955-1960 and 1980-1990 and decrease during 1990-2000, again leading to an overall increase over the period 1955-2000, and these variations are highly synchronous with corresponding variations in the observational record. Re-
turn levels for model re-analysis data at Southend (not shown) show similar patterns to those at Lowestoft. At Southend the corresponding trends in the observational data are radically different to those in the model re-analysis data; historical damage to the tide gauge site resulted in several relocations of the gauge at this site which, combined with timing errors identified in the early part of the dataset, render trends in the *in situ* data unreliable.

Finally, return levels for model re-analysis data at Dover increase from 1955-1960 and decrease from 1990-2000, showing no overall net change over the period 1955-2000; these variations are moderately synchronous with the corresponding variations in the observational record.

Note that the variability bands at all three sites appear to be consistent with a model in which there are no changes at all over time, but that simpler parametric extreme value models imply that the increase in storm surge elevations at Aberdeen is highly statistically significant. These findings are not inconsistent - the variability bands about the fitted local regression model are pointwise, and so only provide an indication of statistical significance for an individual year. At Aberdeen the local regression model provides strong cumulative evidence for an increase in storm surge elevations over the course of the entire period 1955-2000.

**Figure 3 about here.**

### 4.2 Spatial variations in temporal trend

In Figure 3 we plot estimated 50 year return levels obtained by fitting our local regression model to model re-analysis data at each of 259 grid cells throughout
the North Sea; estimates are shown for six relatively representative years.

The return level estimates for a particular year at a particular site are standardised is such a way that a level of one corresponds to the 50 year return level which we would estimate for that site if we assumed that storm surge characteristics were constant over time. This standardisation highlights temporal variations in the levels, with values less than one corresponding to years in which surge levels were relatively low (compared to the whole period 1955-2000) at that location, and values greater than one corresponding to years in which surge elevations were relatively high.

The maps indicate rapidly increasing levels in storm surge elevations from 1955 to 1962 across the entire region, except for an area in the NW corner of the study area (off and around the Scottish east coast). Trends during the period from 1962 to 1984 are rather complicated, and relatively subtle: levels along the Dutch coast increase from 1962 to 1975 but decrease from 1975 to 1984, whilst levels along the Scottish coast decrease from 1962 to 1975 but then increase from 1975 to 1984. The period from 1984 to 1990 witnesses a rapid increase in storm surge elevations along the British east coast south of the Humber, but relatively few changes elsewhere. From 1990 to 2000 we see a rapid increase in storm surge elevations across the north and west of the study region, but a rapid decrease in storm surge elevations across the south and east of the region.

The maps in Figure 3 present new insights into the impact of climate variations on extreme surge levels in the North Sea from 1955 to 2000. Only point estimates are shown in the figure, but the associated uncertainties must be accounted for when drawing conclusions about the strength of the temporal
trends. Our bootstrap methods enable us to quantify levels of uncertainty within our estimates at each site for each year, but cannot easily be used to draw formal statistical inferences simultaneously over space and time. However, analyses based upon fitting extreme value models with polynomial trends over time suggest that the strongest of the temporal trends shown in Figure 3 are highly statistically significant (Butler, 2005).

We have focused so far on the estimation of trends in 50 year return levels, although our modelling approach could equally well be used to produce estimates for any other return level. Estimated trends in return levels are not straightforward to interpret, however, and the estimated trends in the values of the location, scale and shape parameters tend to give deeper scientific insights into the mechanisms which are generating the trends. It is known for example (Coles & Tawn, 1990) that trends in the mean level of surges would lead \( \sigma_j \) and \( \xi_j \) to be constant over time \( t_j \), whilst trends in the variability of surges would lead \( \sigma_j/\mu_j \) and \( \xi_j \) to be constant over time. Trends in the frequency of storm surges would, if associated with no change in the magnitude of storm surges, lead \( \xi_j \) and \( \mu_j - \sigma_j/\xi_j \) to be constant over time. The parameter estimates from our analyses (not shown) suggest that the temporal trends in return levels which we have reported are due to complex changes in the distribution of storm surge characteristics, and cannot be attributed to any simple trend in the mean, variance or frequency of storm surges.

4.3 Assessment of model fit

We need to check that the extreme value models of Section 3 do indeed provide a reasonable description of the (model re-analysis or observational) data.
In Figure 4 we use Q-Q plots to assess the fit of the basic $r$-largest model of Section 3.2 (in which the parameters are constant over time) and the local regression model of Section 3.3 (in which the parameters vary over time) to model re-analysis data for Aberdeen, Lowestoft and Dover. Points on the Q-Q plots which lie close to the diagonal line indicate a good level of agreement between the data and the statistical model, whilst points that lie far from the diagonal line indicate that the statistical model is failing to capture some important aspect of the data. We can also use boostrapping to generate variability bands about the model-based quantiles, so that the Q-Q plot is additionally able to reflect the uncertainty involved in fitting the model to the data.

Deviations between the model and data are fairly substantial for estimates obtained using the no-trend model, relative to the (rather narrow) variability bands associated with this model, indicating that such a model fails to adequately describe the statistical features of storm surges in the model re-analysis data. Deviations between the model and data for the local regression model, are, in contrast, typically quite small, at least relative to the (substantially wider) variability bands associated with this model. The greatest improvement in fit is at Aberdeen, which we have already seen to be the site with the most pronounced temporal trends, with more modest improvements at Lowestoft and Dover.
The North Atlantic Oscillation (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 by the Climatic Research Unit, with indices for winter months known to be strongly correlated to levels of storm activity in the North Atlantic (Kaas et al., 1996, Rogers, 1997, and Schmith et al., 1998). Wakelin et al. (2003) reported that NAO indices are also strong correlated with non-extreme sea levels on the European continental shelf.

We have investigated the impact of the NAO upon model re-analysis data 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 and then estimating the four parameters of the resulting regression models using maximum likelihood. We have found a strong positive relationship between extreme surges and the annual NAO, with this relationship being statistically significant (at the 95% level) at around 90% of grid cells within the North Sea - mirroring the strong relationship at non-extreme levels reported by Wakelin et al. (2003). There are also strong positive relationships with NAO indices for each of the individual winter months from December to March. We have also investigated whether the long-term trends in extreme surge elevations which we have estimated using our local regression model remained once we have accounted for this effect of the NAO. We treated the estimated regression coefficients from regressing the location parameter of the \( r \)-largest model onto the NAO signal as a fixed “offset” term within a local re-
gression model for model re-analysis data at each location, and then estimated
the remaining parameters of the model using local likelihood. We found that
the resulting estimates for temporal trend did not differ qualitatively from
those which we obtained without accounting for the NAO.

4.5 Accounting for tide-surge interactions

The most severe coastal flood events occur when extreme surge elevations coin-
cide with high water of spring tides. Tide and surge processes are dynamically
coupled, with interactions resulting from non-linearity including the effects of
friction and advection. In certain environments tide-surge interactions act to
decrease surge levels at high water whilst increasing surge levels on the ris-
ing tide (Prandle & Wolf, 1978). Such interactions are strongest in areas of
shallow water and with a large tidal range.

We have used extreme value methods to analyse the 20 largest (declustered)
model re-analysis surge elevations to occur on a high tide - i.e. within 30
minutes of high water - for each year and site. Estimates for 50 year surge
return levels will inevitably be lower when we restrict attention only to those
surges which occur in conjunction with a particular tidal state, but we find
that the magnitude of the reduction varies markedly between locations. At off-
shore locations in the central North Sea estimates are reduced by only 0-20cm,
but within the shallow coastal waters of the Wash, the Thames Estuary and
Heligoland Bay estimates are reduced by 100-200cm. Despite the sometimes
substantial effect of tide-surge interaction upon extreme surge elevations, we
find that estimates for temporal trends in surge elevations at high tides are
qualitatively similar to those which we estimate when ignoring tidal informa-

20
5 Conclusions

We have used a novel statistical approach to study temporal trends in the magnitude and frequency of storm surge elevations in the North Sea over the past fifty years, using both observational data and model re-analysis data.

We have seen that the model re-analysis data show strong positive trends in the north western part of our study region (around the Scottish coast), and provide weak evidence for positive trends across the central and north-eastern part of our study region (including the German Bight). Surge signals in the Southern Bight, in contrast, appear to be almost entirely dominated by decadal signals. The most pronounced temporal changes appear to occur during the periods 1955-1960 (decreases in the NW, but rapid increases everywhere else), 1980-1990 (rapid increases in the south and west, few changes elsewhere), and 1990-2000 (rapid increases in the north and west, rapid decreases in the east and south). Similar trends apply when we restrict attention only to surge levels associated with high tides. Extreme surge elevations are strongly related to NAO indices across most of the North Sea, but we have seen that the decadal and longer-term trends remain once we have accounted for the effect of NAO.

Our results differ from those of Langenberg et al. (1999), whose dynamical hindcast analysis found no significant trends along the British coastline but trends in winter mean high waters along the Danish and German coasts; the differences may be due to our selection of residual as the variable of interest. The second approach of Langenberg et al. (1999) was to use regression tech-
niques applied over a century of data where they found very small positive trends everywhere. There is no comparable analysis in this paper, and in any case their regression was applied to a very limited atmospheric variable (mean sea level pressure at 5° resolution).

The trends which we have detected are clearly reliant upon the accuracy of the CSX model, and of the hindcast meteorological data which were used to force it. The CSX-DNMI data were the best available at the time of the analysis. Although there are more consistent reanalysis fields available now (e.g. ERA-40) they do not improve significantly on the spatial resolution here, or on the temporal resolution (6 hourly). Improvements in computer power since the inception of this project mean that it is now possible to repeat the analysis using a surge model with a finer spatial resolution, but the resolution of the meteorology will always be a limiting factor in reconciling the model output with the variation of sea level seen in an observational record.

The key focus of the current paper, however, has been to demonstrate a novel statistical approach for the analysis of long-term variability in extreme sea levels, and to show how this technique can be used as a diagnostic tool for the magnitude and frequency of storm surges generated by sophisticated numerical models. The approach has general applicability for detecting and exploring temporal and spatial trends in the magnitude and frequency of extreme events, and could easily be extended to deal with more complicated problems and/or to incorporate more physical information into the statistical analysis. In particular, it is known that we can reduce the uncertainties in parameter estimates from extreme value models by explicitly incorporating knowledge of spatial dependence (Dixon & Tawn, 1992; Dixon et al., 1998), and preliminary investigations using the datasets from this paper suggest that these benefits
can be substantial within an oceanographic context (Butler et al., 2006). The key practical issues with the methodology revolve around the choice of $r$ and the choice of $h$, and it is important to recognise that these quantities need to be carefully selected on the basis of both (a) the empirical properties of the dataset at hand, and (b) the scientific objectives of the analysis. Overall, the methodology which we have presented provides a flexible and statistically rigorous approach for quantifying changes in the extremal properties of oceanographic processes.

**Acknowledgements**

The first author was jointly 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. Barry Rowlingson and David Blackman provided technical assistance in processing the output from the CSX model. Tide gauge data for the United Kingdom were provided by the British Oceanographic Data Centre, and NAO indices by the Climatic Research Unit.

**References**


Long-term trend variations of the storm climate over NW Europe. The Global Atmosphere and Ocean System, 6: 97-120.

Bijl, W. 1997., Impact of a wind climate change on the surge in the southern North Sea. Climate Research, 8: 45-49.


the North Atlantic Oscillation and climate variability of Northern Europe.
Journal of Climate, 10: 1635-1647.

Schinke, H., 1993. On the occurrence of deep cyclones over Europe and the


1875-1995 re-analysed. Climate Dynamics, 14: 529-536.

Smith, R.L., 1986. Extreme value theory based on the $r$ largest annual events.

Smith, S.D. and Banke, E., 1975. Variation of the sea surface drag coefficient

Stein, O. and Hense, A., 1994. A reconstructed time series of the number of
extreme low pressure events since 1880. Meteorologische Zeitschrift, NF3: 43-46.


of the American Statistical Association, 82: 559-567.

von Storch, H., Guddal, J., Iden, K., Jonsson, T., Perlwitz, J., Reistad, M., de
in the North Atlantic. Technical report 116, Max-Planck Institute for Me-
teorology, Hamburg, Germany, unpublished.


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. Temporal trends in observational and model re-analysis surge elevation data at three locations on the British East Coast. The 20 largest values per year are shown (crosses), together with estimated 50 year return levels from the local $r$-largest model (dots). 95% variability bands are also shown (grey lines).

Fig. 3. Estimated 50 year return levels from fitting the local $r$-largest model to model re-analysis data at locations throughout the North Sea. Results are shown for six representative years. Return level estimates for each year at each location are scaled by dividing by the corresponding estimate obtained from fitting a time-constant $r$-largest model to data for that location (also using $r = 20$). Locations of selected ports are indicated: Aberdeen (A), Whitby (W), Immingham (I), Lowestoft (L), Dover (D), Vlissingen (V), West Terschelling (T), Eemshaven (M), Esbjerg (E), Hirtshals (H).

Fig. 4. Q-Q plots to assess the fit of time constant and local $r$-largest models to model re-analysis data at three locations on the British east coast. Plots shown are for the annual maxima ($k = 1$), but have been produced for all values $k = 1, \ldots, 20$. 95% variability bands are also shown (grey lines).
Figure 1
Figure 3