Spatial Bayesian hierarchical modelling of extreme sea states

Colm Clancy[1,1], John O'Sullivan[1,1], Conor Sweeney[1,1], Frédéric Dias[1,1,1], Andrew C. Parnell[1,1]

[a] *School of Mathematics and Statistics, University College Dublin, Ireland*
[b] *UCD Earth Institute, University College Dublin, Ireland*
[c] *CMLA, ENS Cachan, CNRS, Université Paris-Saclay, 94235 Cachan, France*

Abstract

A Bayesian hierarchical framework is used to model extreme sea states, incorporating a latent spatial process to more effectively capture the spatial variation of the extremes. The model is applied to a 34-year hindcast of significant wave height off the west coast of Ireland. The generalised Pareto distribution is fitted to declustered peaks over a threshold given by the 99th percentile of the data. Return levels of significant wave height are computed and compared against those from a model based on the commonly-used maximum likelihood inference method. The Bayesian spatial model produces much smoother maps of return levels. Furthermore, this approach greatly reduces the uncertainty in the estimates, thus providing information on extremes which is more useful for practical applications.

Keywords: Bayesian hierarchical modelling, spatial modelling, extreme value analysis, ocean waves, significant wave height

1. Introduction

A detailed knowledge of the extreme sea states affecting a region is essential for any marine activity. For shipping, offshore and coastal installations, or the

*Corresponding author at: School of Mathematics and Statistics, University College Dublin, Belfield, Dublin 4, Ireland.

deployment of devices such as wave energy converters, it is crucial to have accurate information on the extremes likely to be encountered during operational lifetimes. These are typically expressed in terms of return levels and periods; for example, the level of significant wave height which is likely to occur on average once every 100 years. Extreme value theory provides statistical tools for such an analysis (Coles, 2001) and the methods have been widely applied in studies of ocean waves; reviews may be found in Vanem (2011) and Jonathan and Ewans (2013).

In Section **??** below we outline the background theory of extreme value analysis. Models of the extremes are often fitted to data-sets using a maximum likelihood approach. Although straightforward to implement, this can lead to large uncertainties in the parameter estimations and subsequent return levels (Vanem, 2015). Obviously, we wish to reduce the levels of uncertainty and obtain meaningful results which are of practical use. Bayesian inference allows for a more detailed analysis of this uncertainty, by providing complete probability distributions for the parameters (Gelman et al., 2013).

Our aim in this paper is to use Bayesian techniques to model the spatial variability of ocean wave extremes. We follow the approach of Cooley et al. (2007), who include a latent spatial process within a Bayesian hierarchical framework to capture the spatial dependence of precipitation extremes. This is described in detail in Section **??**. Such a model has been applied to the study of temperature extremes in the ocean by Oliver et al. (2014) but not to ocean wave data, to the best of the authors' knowledge.

We apply the statistical model to significant wave height data off the west coast of Ireland, obtained from a spectral wave model hindcast. Recently, O'Brien et al. (2013) provided a history of extreme wave events around Ireland, revealing an often severe environment. On the other hand, the seas off the west coast of Ireland have attracted interest due to their potential wave energy resources (Gallagher et al., 2016) and so an accurate description of the likely extremes is of both theoretical and practical relevance.

A description of the domain and data under study, along with model imple-

mentation details, is given in Section **??**. The results are presented in Section **??** with a discussion of conclusions in Section **??**.

2. Extreme Value Analysis

2.1. Background theory

There are a number of possible approaches to extreme value analysis. An introduction to the field may be found in Coles (2001). One fundamental method is the block maxima approach. We consider a sequence of independent and identically-distributed random variables, $Z_1, Z_2, \ldots$, and let $M_n = \max(Z_1, \ldots, Z_n)$ be the maximum over a block of n values; for example, we may take M_n to be the annual maxima in a multi-year set of significant wave height data. The extremal types theorem states that, under certain regularity conditions, the distribution function of the M_n will converge to a specific three-parameter form, known as the generalised extreme value (GEV) distribution.

A major disadvantage to this approach is the fact that, by using only the maxima from a given block size, we are discarding a lot of data. In this work we consider a data-set of hourly significant wave height, H_s. Modelling with, for example, annual maxima would be quite wasteful. An alternative is to model the excesses over a given threshold, u. We assume that our sequence of independent random variables, $Z_1, Z_2, \ldots$, satisfies the extremal types theorem described above. For large enough u, the distribution function of the exceedances $Y = Z - u$, conditional on $Z > u$, is approximately given by the generalised Pareto distribution (GPD)

$$F(y) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi} \tag{1}$$

defined on the set $\{y : y > 0 \text{ and } (1 + \xi \mathrm{y}/\sigma) > 0\}$. Here, ξ and σ are known as the shape and scale parameters, respectively, and have ranges $-\infty < \xi < \infty$ and $\sigma > 0$. For the limiting value when $\xi = 0$, we get the exponential distribution

$$F(y) = 1 - \exp\left(-\frac{y}{\sigma}\right)$$

These two methods of extreme value analysis have been applied extensively to ocean wave data from different sources. Examples of GEV models include Menéndez et al. (2009), who use monthly maxima of H_s from observational buoy data, and Izaguirre et al. (2011), in which monthly maxima are obtained from satellite altimeter missions. Threshold exceedance models of H_s with the GPD may be found in Caires and Sterl (2005), Vinoth and Young (2011) and Nicolae Lerma et al. (2015). In addition, a number of papers have compared the various approaches; see, for example, Caires (2011), Aarnes et al. (2012), Vanem (2015) and Clancy et al. (2015).

Once we have the parameters of a distribution, we may compute the N-year return levels. For the GPD in (**??**), we have

$$P(Z > z|Z > u) = \left(1 + \frac{\xi(z-u)}{\sigma}\right)^{-1/\xi}. \tag{2}$$

We write $\zeta_u = P(Z > u)$ and can then find the return level z_m, the level which is exceeded on average once every m observations, by solving

$$P(Z > z_m) = \zeta_u \left(1 + \frac{\xi(z_m - u)}{\sigma}\right)^{-1/\xi} = \frac{1}{m}.$$

Letting $m = N\, n_y$, where n_y is the number of observations per year, we arrive at the following expression for the N-year return level:

$$z_N = u + \frac{\sigma}{\xi}\left[(N n_y \zeta_u)^{\xi} - 1\right] \tag{3}$$

For the case of the exponential distribution with $\xi = 0$, we have

$$z_N = u + \sigma \log\left(N n_y \zeta_u\right)$$

2.2. Model fitting

Given a set of data, we may fit one of the models described above. The maximum likelihood (ML) estimation method is commonly used. We can consider a set of n independent values, $z_1, \ldots, z_n$, to which we wish to fit a probability density function $f(z;\theta)$, where θ is a parameter of the distribution. The likelihood function is given by

$$L(\theta) = f(z|\theta) = \prod_{i=1}^{n} f(z_i;\theta)$$

The maximum likelihood estimator $\hat{\theta}$ is found by maximising the above likelihood function or, more usually, the logarithm of $L(\theta)$. Asymptotic properties of the ML estimate, which assume Gaussian behaviour, may then be used to compute confidence intervals. Furthermore, the so-called delta method provides confidence intervals for quantities derived from the parameter estimates; for example, the return levels in (**??**). Details of these, as well as other fitting methods used in extreme value modelling, are given in Coles (2001).

An alternative is to use Bayesian inference for parameter estimation (Gelman et al., 2013). Continuing the above example, we use Bayes' Theorem to write

$$f(\theta|z) \propto f(z|\theta)\, f(\theta) \tag{4}$$

Thus, we arrive at a *posterior distribution*, $f(\theta|z)$, from a combination of the likelihood of the data and a given prior distribution $f(\theta)$. Whereas the ML method gives a point estimate of a parameter, with the Bayesian approach the parameter is described by a complete distribution. This allows us to characterise the uncertainty in a natural way. Rather than appealing to asymptotic theory for confidence intervals, we may use, for example, the percentiles of the posterior distribution.

A detailed treatment of Bayesian methods may be found in Gelman et al. (2013). Coles (2001) provides a brief introduction to their use in extreme value analysis while Coles et al. (2003) further discuss their benefits over likelihood-based inference in modelling extremes. In the context of ocean wave modelling, Egozcue et al. (2005) and Scotto and Guedes Soares (2007) were among the first to apply a Bayesian approach; see Vanem (2011) for a review of various models of ocean extremes. The review of Jonathan and Ewans (2013) points to the growing use of Bayesian methods and their potential for ocean engineering applications.

Practical implementation of Bayesian inference can be computationally intensive, in particular the calculation of the proportionality constants in (**??**). The development of the Markov chain Monte Carlo (MCMC) technique has been hugely successful in making these methods viable. This algorithm may

be used to draw simulated samples from the desired posterior distributions (Geyer, 2011).

2.3. Spatial modelling of extremes

A number of authors have examined the spatial variation of extreme sea states, rather than focussing on one particular location. Fedele (2012), for example, considered space-time extremes of individual crest heights over a spatial region.

For studies involving extreme value modelling of significant wave height, the global or regional data-sets used have come from satellites (Vinoth and Young, 2011; Izaguirre et al., 2011) or model hindcasts and reanalyses (Caires and Sterl, 2005; Cañellas et al., 2007; Aarnes et al., 2012; Guo and Sheng, 2015). This extreme value analysis has been carried out at each individual point on some given spatial grid. However, this approach does not explicitly model the spatial variations since each point is treated independently. This current work improves upon this by incorporating a spatial model within a Bayesian hierarchical framework.

Previously, Vanem et al. (2012) used a Bayesian hierarchical structure in a spatio-temporal model of significant wave height data. Here, our specific interest is extremes. Cooley et al. (2007) presented a method for producing maps of extreme precipitation return levels in Colorado, using separate hierarchical models to model the intensity and frequency of events. Within both models, it was assumed that regional extreme precipitation is driven by a latent spatial process, defined by geographical and climatological covariates, and that effects not fully captured by the covariates are captured by the spatial structure in the hierarchies, using Gaussian processes. Inference was then conducted using an MCMC algorithm. This approach has since been used in oceanographic applications by Oliver et al. (2014), to analyse extremes of sea surface temperatures.

In this current work, we apply a similar model to extreme significant wave heights off the Irish west coast. This continues on from the work of Clancy et al. (2015), in which this region was examined using extreme value analysis

applied independently at each point. In the next section we describe the model in detail.

3. Model Details

The aim is to produce N-year return levels of significant wave height, H_s. We described earlier how these could be computed using (**??**). Given a dataset, we require a suitable threshold u, the parameters from the generalised Pareto distribution (GPD) for modelling the exceedances and their probability of occurrence ζ_u. The choice of threshold will be discussed later in Section **??**.

For the exceedances and the probability ζ_u, we follow the approach of Cooley et al. (2007) and employ a Bayesian hierarchical model with three layers. The first layer consists of modelling the data. The second describes the latent spatial process driving the extremes in the region while the third layer consists of the prior distributions on the parameters controlling the second.

Using Bayes theorem, under a three-layer hierarchical model the inference for the vector of parameters θ_1 (for the GPD of exceedances or the probabilities ζ_u) is given by

$$p\left(\theta_1|Z(x)\right) \propto p_1\left(Z(x)|\theta_1\right) p_2(\theta_1|\theta_2) p_3(\theta_2) \tag{5}$$

where the p_j are the probability densities with indices associated with the levels of the hierarchy and $Z(x)$ specifies the data at a given location x. We now describe the two hierarchical models. A directed acyclic graph (DAG) depicting the hierarchal structure of the models is given in Fig. **??**.

3.1. Modelling the threshold exceedances

3.1.1. Data layer

A GPD given by (**??**) is used to model the data at the first layer of the hierarchy. To ensure a positive scale parameter throughout the computations, we reparameterise with $\phi = \log \sigma$. At this level we thus have two spatially-varying parameters for the distribution, which we collectively write as $\theta_1 =$

$[\phi(x), \xi(x)]^T$. The first term in the hierarchy (**??**) is then derived from the density function for the GPD and given by the likelihood function

$$p_1\left(Z(x)|\theta_1\right) = \prod_{i=1}^{n_x}\prod_{k=1}^{n_i}\frac{1}{\exp\phi(x_i)}\left(1+\frac{\xi(x_i)z_k(x_i)}{\exp\phi(x_i)}\right)^{-1-1/\xi(x_i)}$$

where the indices i and k are such that $z_k(x_i)$ refers to the k-th exceedance at grid-point x_i. We have denoted the number of grid-points by n_x and the number of exceedances at each point x_i is then n_i.

3.1.2. Process layer

Both $\phi(x)$ and $\xi(x)$ are modelled as Gaussian processes (Banerjee et al., 2014) and so the second term in (**??**) will take the form

$$p_2(\theta_1|\theta_2) = p_\phi(\phi(x)|\mu_\phi, \Sigma_\phi)p_\xi(\xi(x)|\mu_\xi, \Sigma_\xi)$$

where

$$p_\phi(\phi(x)|\mu_\phi, \Sigma_\phi) = \frac{1}{\sqrt{(2\pi)^{n_x}|\Sigma_\phi|}}\exp\left[-\frac{1}{2}(\phi-\mu_\phi)^T\Sigma_\phi^{-1}(\phi-\mu_\phi)\right].$$

A similar expression is used for $p_\xi(\xi(x)|\mu_\xi, \Sigma_\xi)$. Here $|\,.\,|$ denotes the determinant and θ_2 above represents all of the hyperparameters for μ_ϕ, μ_ξ, Σ_ϕ and Σ_ξ, to be discussed below.

A Gaussian process characterises an infinite-dimensional smooth surface such that any finite collection of n_x points on the surface follows a multivariate normal distribution (above) of dimension n_x. Such a smooth surface is an appropriate choice for the model parameters as we expect similar wave climates at nearby locations.

In addition to distance, the effect of any other covariates may be readily incorporated into the model. For m covariates $c^{(1)}, \ldots, c^{(m)}$, we write the mean vector in the general form

$$\mu_\phi = \mathcal{C}\alpha_\phi \tag{6}$$

where $\mathcal{C}$ is the $n_x \times (m+1)$ matrix

$$\mathcal{C} = \begin{pmatrix} 1 & c_1^{(1)} & c_1^{(2)} & \dots & c_1^{(m)} \\ 1 & c_2^{(1)} & c_2^{(2)} & \dots & c_2^{(m)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & c_{n_x}^{(1)} & c_{n_x}^{(2)} & \dots & c_{n_x}^{(m)} \end{pmatrix}$$

and the vector of coefficients is $\alpha_\phi = (\alpha_{\phi,0}, \alpha_{\phi,1}, \dots \alpha_{\phi,m})^T$. In this work we have used $m = 3$: the longitude, latitude and depth of a grid-point, and therefore we will have four coefficients $\alpha_{\phi,j}$.

The use of the Gaussian process also offers great flexibility through the choice of the covariance matrix Σ_ϕ. Here, we use the matrix given by

$$\Sigma_\phi = \varsigma_\phi^2 \, \mathcal{E} + \tau_\phi^2 I \tag{7}$$

where I is the identity matrix. The matrix $\mathcal{E}$ is given by an exponential correlation function and has components

$$\mathcal{E}_{i,j} = \exp\left(-d(i,j)^T \, \beta \, d(i,j)\right). \tag{8}$$

For two grid-points x_i and x_j, the vector $d(i,j)$ has two components given by the differences in longitude and latitude between x_i and x_j.

The 2×2 matrix β is symmetric positive definite. Its entries measure how quickly spatial dependence drops off in the two different directions. The other parameters of the covariance matrix appearing in (**??**) are the partial sill ς_ϕ^2 and the nugget parameter τ_ϕ^2. Further details on variogram analysis may be found in Zimmerman and Li (2012).

As mentioned, we assume the same Gaussian process model for the shape parameter ξ. Similar expressions as those above are used for μ_ξ and Σ_ξ.

3.1.3. Priors layer

Finally for the third layer in the hierarchical model, priors must be assigned to the hyperparameters in the second layer, which are assumed to be independent. For those in (**??**), a normal distribution with large variance was

selected: the covariates were re-scaled to be centred on zero and priors $\alpha_{\phi,i}$, $\alpha_{\xi,i} \sim \mathrm{N}(0, 50)$ were used. A lognormal prior was employed for the positive ς_ϕ^2 and τ_ϕ^2 parameters in (**??**); that is, their logarithm was assumed to have the normal distribution N(0,10).

For the entries of the matrix β in (**??**), a discrete uniform prior is assumed. We begin with a set of proposal values $v_\beta = \{0.001, 0.005, 0.01, 0.05, \ldots, 100, 500, 1000\}$, with all values being considered equally likely *a priori*. Within the MCMC algorithm, the entries of β are randomly proposed from v_β and accepted or rejected accordingly by the algorithm.

3.2. Modelling the frequency of exceedances

We now turn to ζ_u, which is defined as the probability that the threshold u is exceeded and is needed in (**??**) to compute return levels. For a given choice of threshold (discussed in Section **??**), we let $\zeta(x_i)$ be the exceedance probability at the location x_i. It is again assumed that there is a latent spatial process driving this and a hierarchical model is used, with data, process and prior layers.

At the data layer it is assumed that, at each grid-point i, the number of declustered threshold exceedances N_i is a binomial random variable with m_i trials (the total number of observations), each with a probability $\zeta(x_i)$ of being a cluster maximum. That is, $N_i \sim \mathrm{Bin}(m_i, \zeta(x_i))$, where

$$P(N = N_i) = \binom{m_i}{N_i} \zeta(x_i)^{N_i} (1 - \zeta(x_i))^{m_i - N_i}.$$

The process layer is similar to that of the GPD parameter $\phi(x)$. Following Diggle et al. (1998), $\zeta(x_i)$ is first transformed using a logit transformation, where

$$\zeta'(x_i) \equiv \mathrm{logit}(\zeta(\mathrm{x_i})) = \log\left(\frac{\zeta(\mathrm{x_i})}{1 - \zeta(\mathrm{x_i})}\right).$$

This is then modelled as a Gaussian process as before, with mean vector $\mu_{\zeta'}$ and covariance matrix $\Sigma_{\zeta'}$ taking the same form as in (**??**) and (**??**), respectively. The hyperparameters are then given the same prior distributions as described above in Section **??**.

4. Implementation Details

4.1. Data

The data used in this study comes from the 34-year hindcast described in Gallagher et al. (2014). The third-generation spectral WAVEWATCH III version 4.11 model (Tolman, 2014) was used with an unstructured grid (Roland, 2008) to simulate the wave climate around Ireland for the period from 1979 to 2012. The model was forced with directional spectra and 10m wind data from the ERA-interim reanalysis of the European Centre for Medium-Range Weather Forecasts (Dee et al., 2011). The unstructured triangular grid consisted of approximately 15,000 nodes (Fig. **??**(a)) with horizontal resolution varying from 250m in the nearshore to 10km further offshore. The hourly fields produced were validated with observations from wave buoys and satellite altimeter data. The simulation was found to give excellent agreement for the significant wave height, defined here as $H_s = 4\sqrt{m_0}$, where m_0 is the zeroth moment of the directional wave spectrum.

Gallagher et al. (2014) reported a strong spatial variability in H_s. A preliminary extreme value analysis of this data-set using the annual maxima and GEV approach was carried out in Gallagher (2014). The highest extremes of H_s were found to occur off the west coast of Ireland along with a high level of uncertainty in the estimates. Similar results were found in Clancy et al. (2015), using a data-set from a coarser-resolution hindcast.

Following on from this, we will focus on the region off the west coast indicated in Fig. **??**. This domain contains $n_x = 334$ nodes and has a depth ranging from 39m to 1902m (see Fig. **??**(b)). In Fig. **??** we show the mean, 99th percentile and maximum of the hindcasted H_s fields. We can see that, even not far from the coast, we have maxima in excess of 14m, a level of sea state categorised as 'phenomenal' by the World Meteorological Organization (WMO, 2009).

4.2. Threshold choice and declustering

We wish to apply the GPD model of threshold exceedance to this data-set. The choice of an appropriate threshold is a non-trivial issue in extreme value

modelling and the subject of much ongoing research (Scarrott and MacDonald, 2012). With a low threshold, the asymptotic validity of the GPD may be violated, leading to bias. On the other hand, if the threshold is set too high we will be left with few data points for fitting the model, resulting in large variances. Coles (2001) describes commonly-used graphical methods for choosing the threshold for a single time series, based on the asymptotic properties of the GPD given in (**??**). However these are rather subjective and furthermore are not suited to modelling over a region with multiple locations.

Numerous other methods have been described in the literature. Thompson et al. (2009) automate the choice using a goodness-of-fit test for the distribution of successive parameter estimate differences as the threshold is increased. A quantile regression model was employed by Northrop and Jonathan (2011), while Dupuis (1998) used optimal bias robust estimators to fit the model and test for the validity of the threshold. In Tancredi et al. (2006), the authors incorporate the uncertainty from the threshold choice by including this within the model estimation procedure.

Here we adopt a more straightforward approach. At each grid-point, the 99th percentile of the H_s data series is selected as a threshold for modelling at that point. Taking a percentile-based threshold is convenient when dealing with a spatial array of data. In Clancy et al. (2015), the 97th percentile was used. Caires and Sterl (2005) examined both the 93rd and 97th percentile and found the higher to be more appropriate in general. Vanem (2015) tested thresholds based on even higher percentiles, above the 99th. The validity of the threshold may be assessed *a posteriori* by examining the fit of the model, and we discuss this choice further below.

Once we have chosen our threshold, we need to decluster the data to be used. This is necessary because the theoretical basis for the use of the GPD assumes that the exceedances are independent. Caires (2011) retains only the maxima of clusters of successive exceedances and additionally removes any peaks which occur less than 48 hours from another, these regarded as having been caused by the same storm system. Nicolae Lerma et al. (2015) varied this time between

48 and 72 hours and found no significant differences in their final results.

In this study we apply a similar, though slightly stricter, method of declustering to that of Caires (2011). Two successive sequences of exceedances are considered to be part of the same cluster and system if the time series drops below the threshold for 48 hours or less. We then use the peaks of each cluster for modelling with the GPD.

5. Results

The GPD was fitted to the data-set discussed above using the spatial Bayesian hierarchical model (BHM). Approximate draws from the posterior distribution of each parameter in the hierarchical model were obtained via the MCMC algorithm. Metropolis-Hastings steps were employed to update each parameter in turn, for each iteration of the MCMC algorithm. This involves drawing a potential value from an appropriate distribution and accepting or rejecting it according to the Hastings ratio (Geyer, 2011).

Three parallel chains were run for each model. Each simulation consisted of 20,000 iterations, of which the first 2,000 were considered as burn-in and consequently discarded. In order to reduce dependence amongst the remaining values, only every 10th was kept. Convergence of the resulting chains was established using the $\hat{R}$ criterion recommended by Gelman (1996), with values below the suggested criterion of 1.2 taken to imply convergence.

The model was implemented in R using a package called Rcpp (Eddelbuettel, 2013). This interface allows integration of R with C++ code, leading to appreciable reduction in the computational burden of the Metropolis-Hastings MCMC algorithm used.

We compare the results with those obtained by fitting the distribution independently at each grid-point with inference from a maximum likelihood (ML) method. For this we have used the Wave Analysis for Fatigue and Oceanography (WAFO) toolbox in Matlab (WAFO, 2011). In addition to spatial maps of the output, we have chosen four locations to focus on in more detail. These are

marked in Fig. **??**(b). Their locations and some hindcast details are listed in Table **??**.

As discussed in Section **??** above, the ML approach produces a single estimate of a given parameter and confidence intervals may be derived from its asymptotic properties. We will consider 95% confidence intervals for our estimates and present the lower and upper bounds of these intervals. The Bayesian model, on the other hand, yields a distribution for the parameter. From the values simulated by the MCMC algorithm, we will present the median value of this distribution and, again, confidence intervals bounded by the 2.5th and 97.5th percentiles.

5.1. Parameters of the GPD

In Fig. **??** we compare the estimates of the shape parameter of the GPD, ξ. The range of values are consistent with Clancy et al. (2015). Some differences are noticeable but, broadly speaking, the patterns are similar, with the highest values of ξ occurring in the band of latitude from 53°N to 54°N.

The scale parameter, σ, from both models is shown in Fig. **??**. For both ξ and σ the spatial variation is noticeably smoother in the BHM compared with the ML. This is to be expected given the latent spatial processes in the BHM, with the covariances of parameters at different grid-points, given by (**??**) and (**??**), based on the distances between them. With the ML model we have simply fitted the distribution independently at each grid-point, with no relationship between neighbouring points.

The lower and upper bounds of the 95% confidence intervals for the estimates of ξ and σ are plotted in Figs. **??** and **??**, respectively. The benefits of the Bayesian approach are immediately clear: there is much less uncertainty in the parameter value estimates as evident from the much narrower confidence intervals. This can be seen in more detail in Fig. **??**, in which we show parameter estimates with confidence intervals for both models for the four points detailed in Table **??**. Looking at Figs. **??**(a) and **??**(b) for shape and scale, respectively, we see that at each of these four points the confidence intervals for the BHM

are contained within the larger intervals for the ML; in fact, this is true for ξ and σ at every grid-point.

The estimates in Fig. **??** show a negative shape parameter at all points for both models. The upper bounds of the ML fit in Fig. **??** allow for positive values. This has previously served as a motivation for fitting with a so-called Type I tail in which $\xi = 0$ is fixed (see Caires, 2011, for example). However, with the BHM model in Fig. **??**, we note that the tighter confidence intervals demand a strictly negative shape parameter.

5.2. Return levels

We now turn to the N-year return levels of significant wave height, which we denote by H_{sN}. In Figs. **??** and **??**, we present spatial plots of H_{s20} and H_{s100}, respectively. The estimates are given along with the lower and upper bounds of the 95% confidence interval. The overall patterns of the return levels are broadly similar for both models. The main differences can be seen in the size of the confidence intervals, which again are much narrower in the BHM. This is even more evident for H_{s100} in Fig. **??**.

As we extrapolate in time to longer return periods, we expect the uncertainty to grow. However, with the BHM we still have a much tighter confidence interval even at 100 years. Examining the values in Figs. **??**(c) and **??**(d) makes this more explicit. At Point 4, for example, the ML method produces a 100-year return level in the range of 13.53 to 17.55 metres. The BHM gives a sharper result of 15.24 to 15.85 metres.

We find that the highest extremes of significant wave height are to be expected roughly between 53°N and 54.5°N, with 100-year levels of close to 16m. This spatial pattern is consistent with the annual maxima GEV analysis of the data-set in Gallagher (2014). Although slightly higher values were reported (up to 18m), the annual maxima approach results in considerably higher levels of uncertainty in the estimates, due to the much smaller number of data-points used. In Clancy et al. (2015), a larger area was studied using a coarser-resolution hindcast but, nevertheless, this region (and further west) showed the most extreme

sea states.

5.3. Validation

Return level plots, as described in Coles (2001), are a useful diagnostic tool for assessing the fit of a model, in addition to illustrating the model estimates. The return level estimates, along with the bounds of the confidence intervals, are plotted against the return period, together with the empirical return level estimates from the data. These are shown in Fig. **??** for the four locations in Table **??** which we have been examining. The black curves are for the ML model, while the BHM is shown in red. The return levels from both models are quite similar for all return periods shown. However, once again we see clearly how the Bayesian model yields estimates with much less uncertainty. This is more and more evident as we extrapolate to longer return periods, such as 1000 years. Comparing with the empirical estimates shows a satisfactory fit, particularly for shorter return periods.

The Bayesian spatial model is further validated in Fig. **??** using a posterior predictive distribution (Gelman et al., 1996). This has been constructed as follows: at a given grid-point i, we have n_i exceedances over the threshold. We then randomly draw n_i values for the shape and scale parameters from the respective estimated posterior distributions; these describe n_i separate GPDs. For each of these, one value is generated. We now have n_i observed and n_i predicted exceedances. After ordering, the two sets should match since a good model will have predicted the observed values.

This is repeated for each of the $n_x = 334$ grid-points and a scatter-plot of the results is shown in Fig. **??**. We see a good match between the observed and predicted exceedances, with a correlation coefficient of 0.99, giving confidence in the validity of the fitted Bayesian model.

6. Discussion and Conclusions

We have applied a Bayesian hierarchical model (BHM) to a hindcast data-set in order to study the extremes of significant wave height off the west coast

of Ireland. Exceedances of H_s over a high threshold are modelled with the generalised Pareto distribution. The hierarchical model includes a latent spatial process which allows us to more effectively capture the spatial variation of the extremes. This approach was compared with a model which uses maximum likelihood (ML) inference and simply carries out an independent extreme value analysis on the time-series at each grid-point in a given domain.

The parameters of the fitted generalised Pareto distribution (GPD) were used to produce spatial maps of extrapolated return levels of H_s. Consistent with previous studies, we found that the highest extremes are to be expected in the latitude band roughly between 53°N and 54.5°N to the west of Ireland, with phenomenal sea states of around 16m estimated for the 100-year return level. A comparison of the two methods showed that the BHM produces much smoother estimates, as a result of the latent spatial process whereby the model parameters at a given grid-point are influenced by the neighbouring points.

A major advantage of the Bayesian approach is the fact that it formally handles parameter uncertainty, rather than relying on the approximate normality of the ML estimate. Parameter and return level estimates were analysed with confidence intervals bounded by the 2.5th and 97.5th percentiles. The BHM showed much a narrower intervals throughout, yielding much higher levels of certainty with the results. This is of crucial importance, as a single point estimate alone is of little practical value without a meaningful measure of uncertainty.

The estimates for both models showed a negative shape parameter ξ of the GPD, but the upper bounds of the wider ML confidence intervals allowed the possibility of a zero or positive shape. By fixing $\xi = 0$, the GPD simplifies to the exponential distribution and some parameter uncertainty is removed. However, the tighter confidence intervals of the BHM showed $\xi < 0$ at all points and thus provided no justification for reducing to the exponential distribution in this case. We note, in passing, that positive values of the shape parameter have been reported in other geographical regions; for example, in the tropics (Izaguirre et al., 2011).

The threshold chosen for this work was the 99th percentile of the H_s data at

each grid-point. As noted, this approach has been used by a number of authors, with various ranges of percentiles tested. Initially, we fitted the model using the 97th percentile, following previous work in Clancy et al. (2015). However, when analysing the validity of the fit, as discussed in Section **??**, we found that both models were greatly underestimating the higher return levels. By switching to the 99th percentile, higher values of the shape parameter were found, resulting in a much more satisfactory fit. The drawback of increasing the threshold is the reduction in the number of exceedance data-points. Nevertheless, even after declustering the data with this higher threshold, we were still left with roughly 160 exceedances per grid-point with which to fit the models.

In addition to a correlation between points based on the distance, we have allowed the distributions of the parameters to be affected by the depth at a given location. However, this was found to have little influence and omitting this covariate gave almost identical results. Other covariates may be added to the model. For example, time-dependent covariates can be used to study the trends in extremes with a changing climate (Caires et al., 2006). Such an investigation could be a possible extension to this current work.

Acknowledgements

The authors are grateful to Sarah Gallagher for her helpful advice on the hindcast data. This work was supported by the European Research Council (ERC) under the research project ERC-2011-AdG 290562-MULTIWAVE and Science Foundation Ireland under grant number SFI/12/ERC/E2227, as well as by the Environmental Protection Agency grant number 2012-CCRP-PhD.3.

Point	Lon.	Lat.	Depth	Mean	99th	Max.
1	-9.66	52.88	68.84	2.42	7.18	13.51
2	-11.22	53.10	134.74	3.12	8.36	15.21
3	-11.60	53.97	329.90	3.21	8.54	15.46
4	-10.70	53.91	130.17	3.12	8.41	15.24

Table 1: The longitude, latitude (both in degrees) and depth (metres) of each of the four points indicated in Fig. **??**(b), along with the mean, 99th percentile and maximum significant wave height (metres) from the 34-year hindcast.

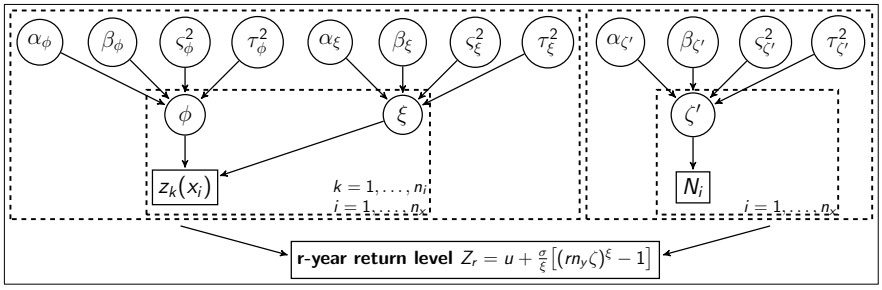
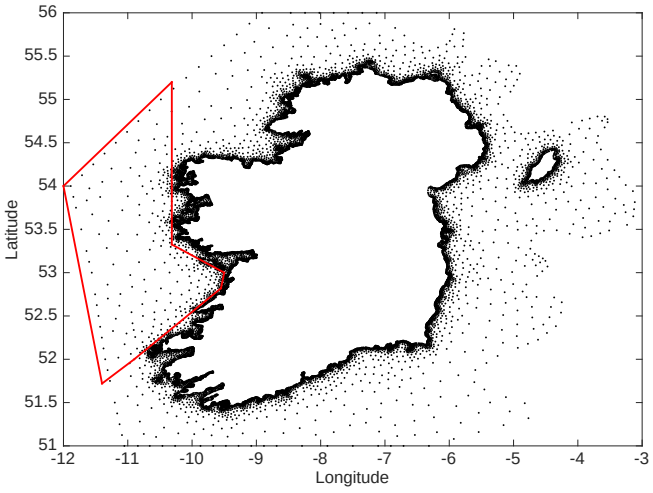
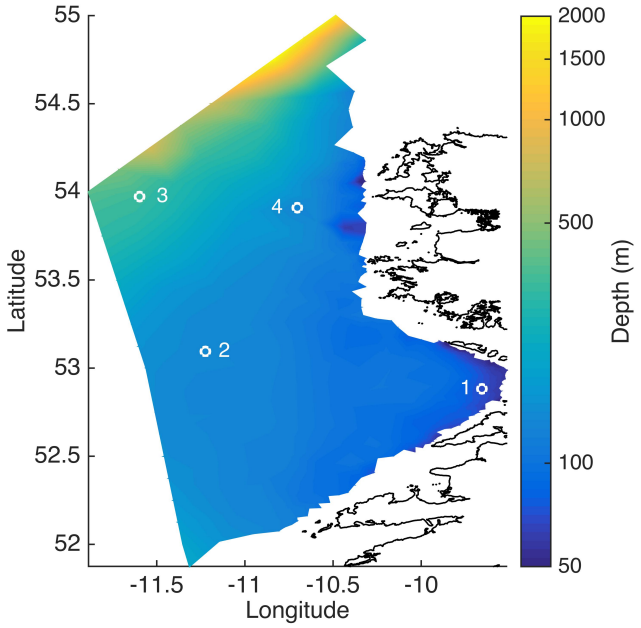


Figure 1: A directed acyclic graph (DAG) of the Bayesian hierarchical model fitted to the wave data. Details of each layer and the parameters involved may be found in the text.



(a)



(b)

Figure 2: (a) the computational grid used for the hindcast, with our region of study outlined. (b) the depth of this region in metres, plotted on a logarithmic scale, with the four locations from Table **??** marked.

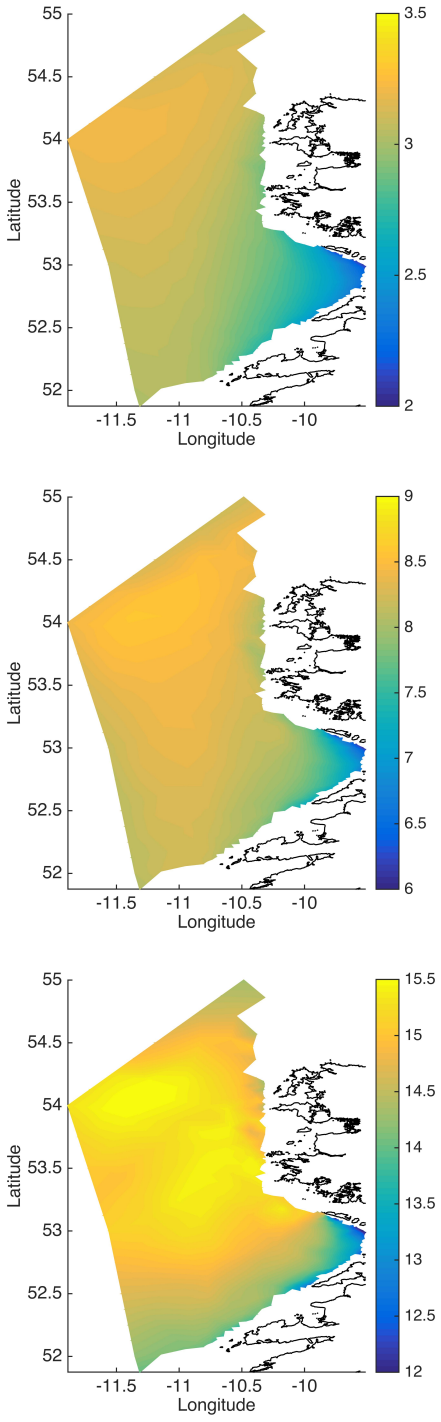


Figure 3: Top to bottom: mean, 99th percentile and maximum significant wave heights in metres, for the 1979–2012 hindcast. Note the differing scales.

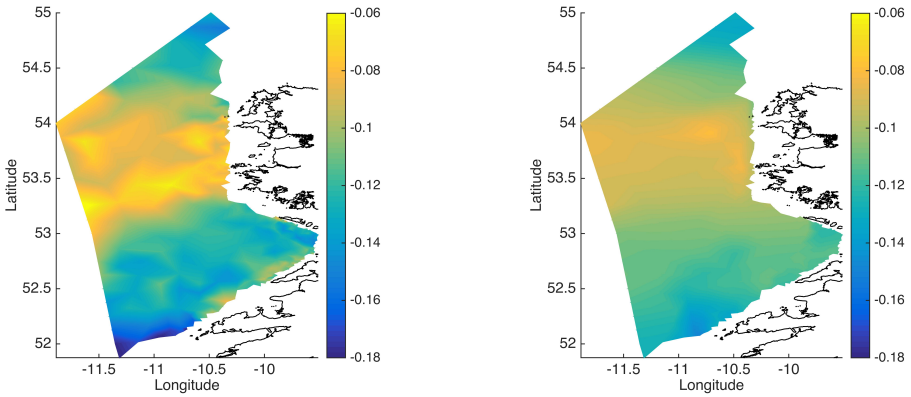


Figure 4: Estimates of the shape parameter ξ, using (left) the maximum likelihood (ML) method and (right) the Bayesian (BHM) model.

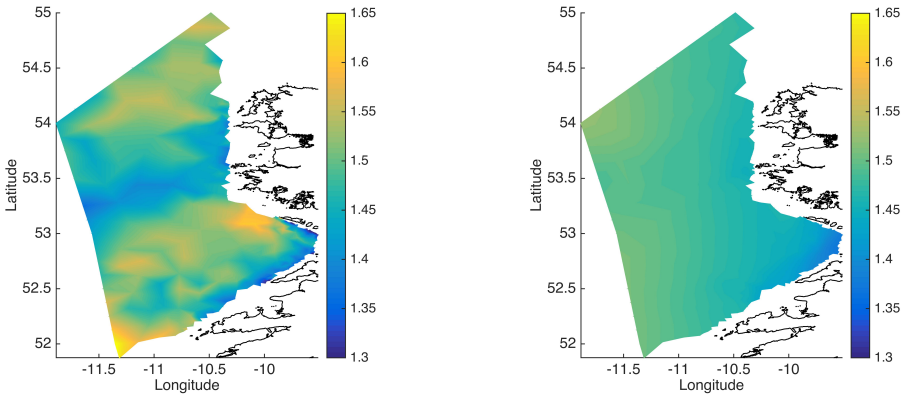


Figure 5: Estimates of the scale parameter σ, using (left) the maximum likelihood (ML) method and (right) the Bayesian (BHM) model.

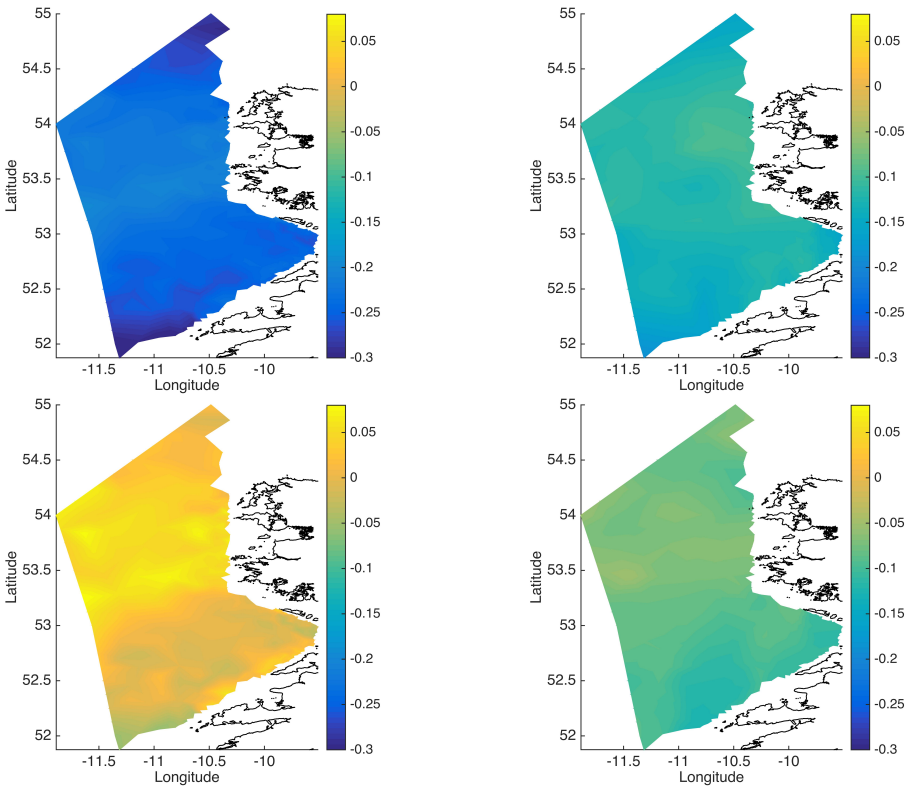


Figure 6: Confidence interval bounds for the shape parameter ξ using the ML method (left) and the BHM (right), given by the 2.5th (above) and 97.5th (below) percentiles of the estimates.

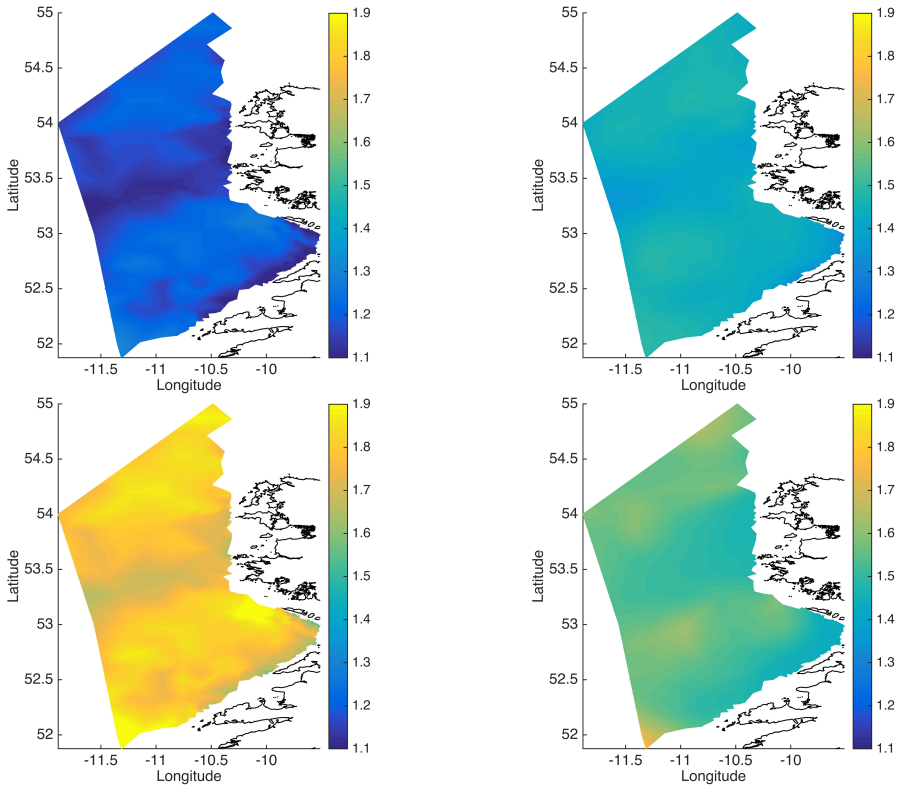


Figure 7: Confidence interval bounds for the scale parameter σ using the ML method (left) and the BHM (right), given by the 2.5th (above) and 97.5th (below) percentiles of the estimates.

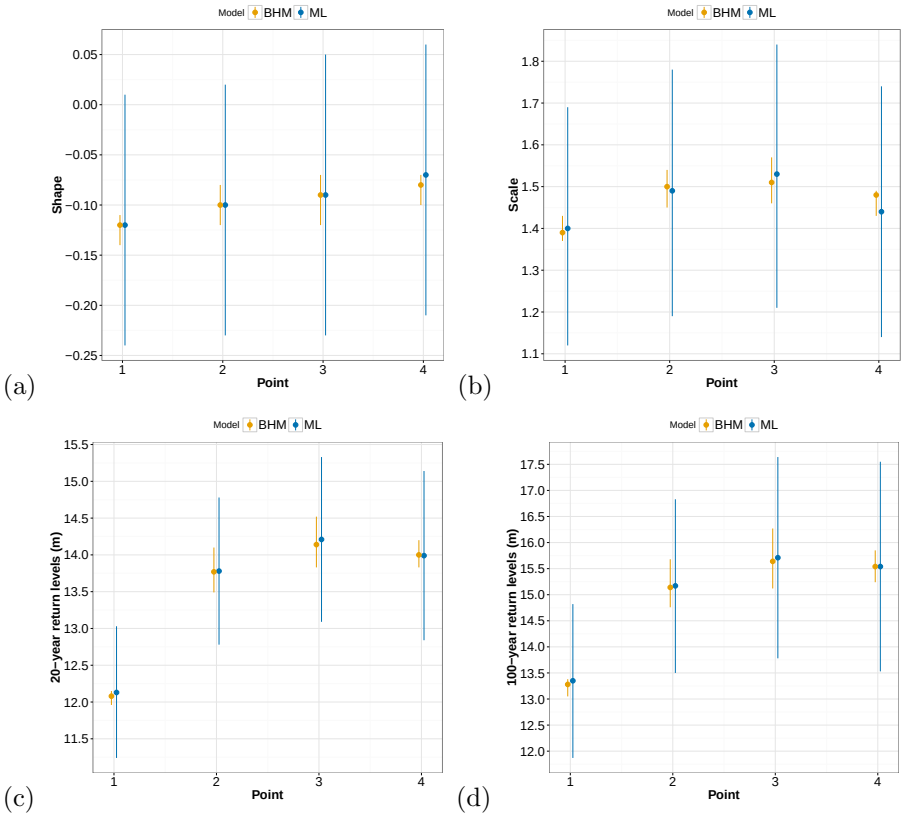


Figure 8: Comparison of estimates from both models for the points marked in Fig. **??**(b) and detailed in Table **??**: (a) shape and (b) scale parameters, along with the (c) 20-year and (d) 100-year return levels. The dot marks the estimate, with the vertical lines indicating the confidence interval bounded by the 2.5th and 97.5th percentile.

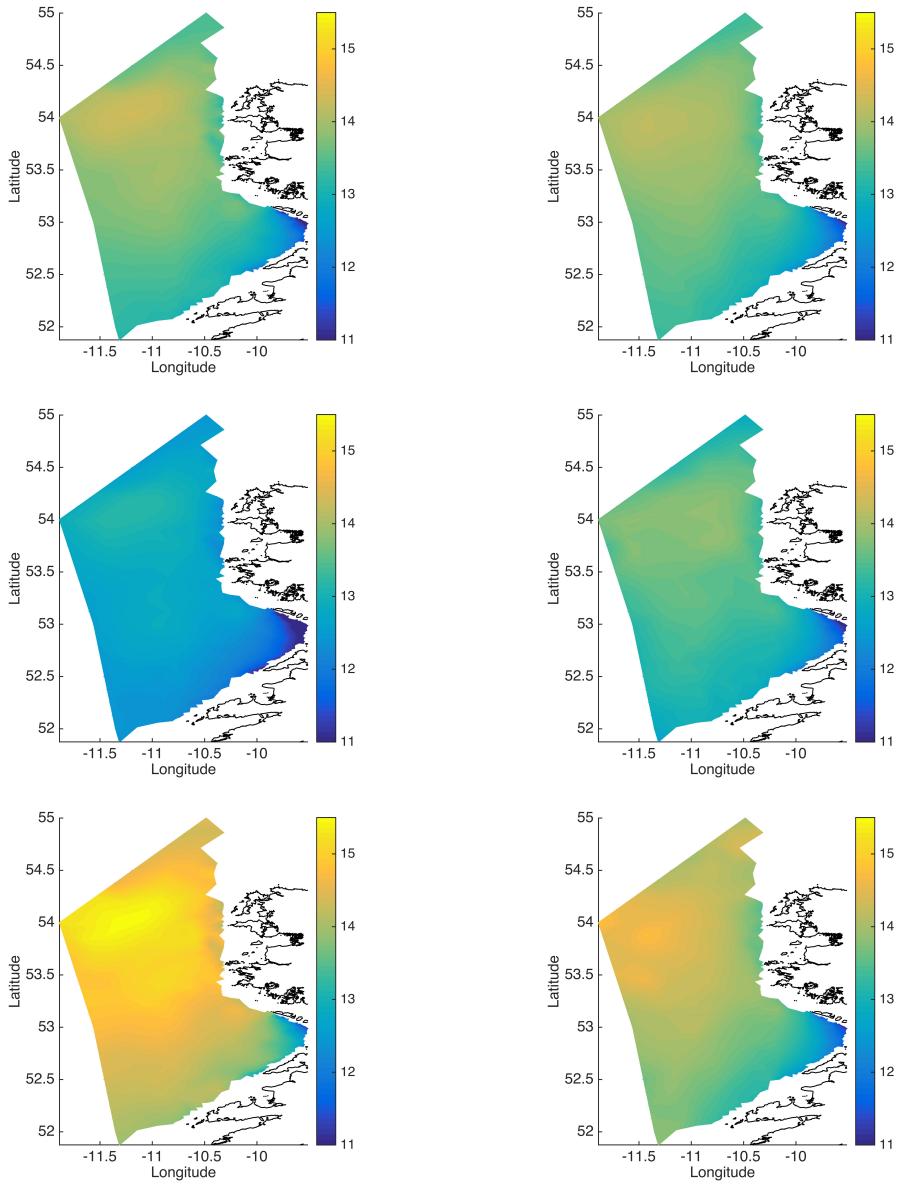


Figure 9: Maximum likelihood (left) and Bayesian (right) estimates of the 20-year return level of H_s. The estimates are shown (top), along with confidence interval bounds given by the 2.5th (middle) and 97.5th (bottom) percentiles.

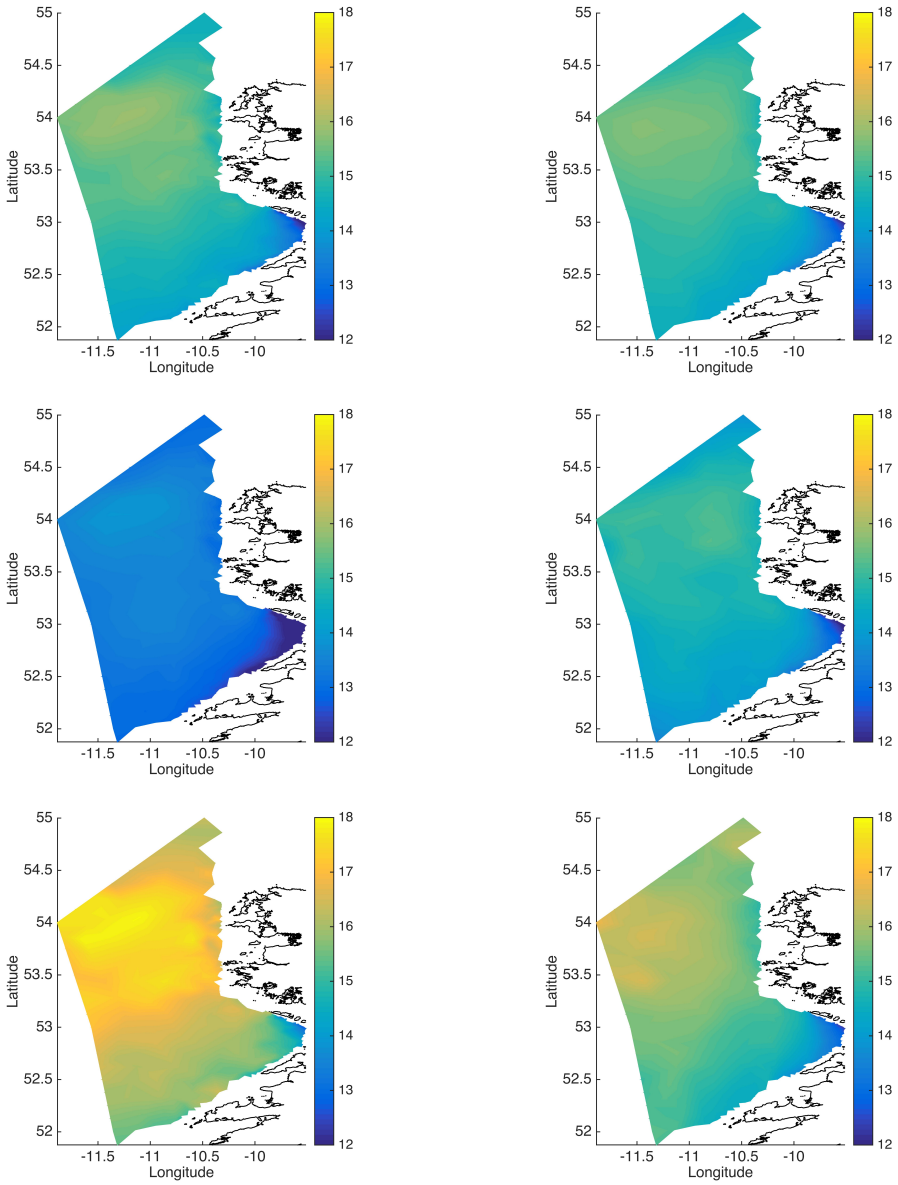


Figure 10: Maximum likelihood (left) and Bayesian (right) estimates of the 100-year return level of H_s. The estimates are shown (top), along with confidence interval bounds given by the 2.5th (middle) and 97.5th (bottom) percentiles.

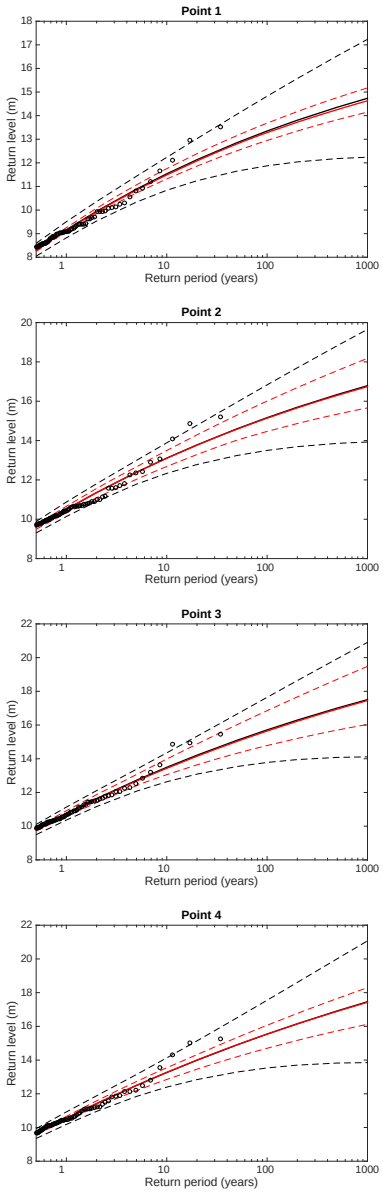


Figure 11: Model return level plots with empirical estimates (circles) for the four points (top to bottom) from Table **??**. Continuous curves are the return level estimates for the maximum likelihood (black) and Bayesian (red) models. The dashed curves are the corresponding lower and upper bounds of the 95% confidence interval. Note the logarithmic scale used for the return period.

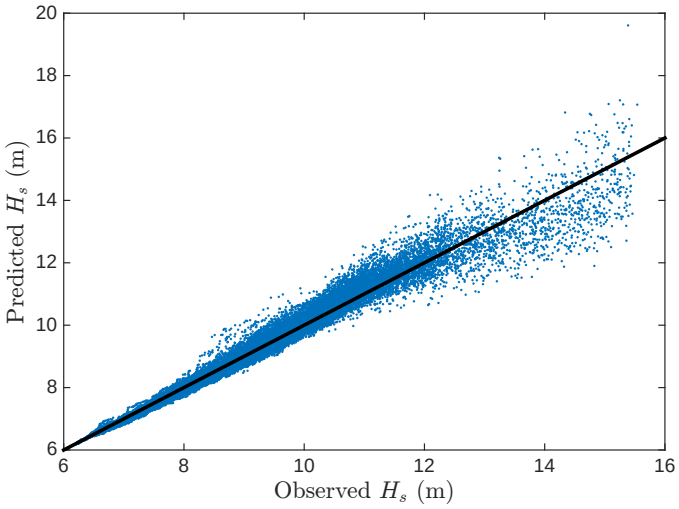


Figure 12: Validating the Bayesian spatial model by comparing observed exceedances of H_s with those predicted by the model. Further details may be found in the text.