A HIERARCHICAL BAYESIAN SPATIODIRECTIONAL MODEL FOR WAVE HEIGHTS AND STRUCTURAL RESPONSE

Richard Gibson
BP, Chertsey Road, Sunbury on Thames, UK

12th International Workshop on Wave Hindcasting and Forecasting
November 2011

Abstract

The design of offshore structures requires estimates of extreme wave height and extreme response. It is often the case that a number of structures are spread across a region. Furthermore, a number of different datasets may be available from which extremes can be estimated. It is important that there is consistency in their design values and that covariate effects, such as directionality, are considered. In order to achieve this, a hierarchical Bayesian model is proposed that incorporates the spatial variation in the parameters of the directional extreme value distribution. This ensures the smooth variation in the parameters across the region whilst allowing for local differences. One of the principle advantages of this approach is that shorter period measurements can be combined with longer period hindcasts in a transparent manner. Furthermore, uncertainties in the extreme values can be estimated and incorporated in load factor calibration studies. The method has been applied to the reanalysis of a North Sea platform and the results are presented.

1 Introduction

Offshore structures must be designed for extreme environmental loads, which often comprises a combination of waves, winds and currents. Therefore, the joint distributions of these environmental parameters must be considered. There are a number of methods by which this can be achieved, but, as it is the extremes of the joint distributions that are of interest it can be challenging (Jonathan et al., 2009). One method is to consider the dependence between the waves, winds and currents implicitly by using the structural response as the parameter of interest (Tromans and Vanderschuren, 1995). However, in order to achieve this a concurrent data set of all the variables is required. In many regions of the world long period hindcast data sets of waves and winds exist, but for currents shorter measured data sets are often used. More generally, the question of how to combine different data sets, and types of data, in a logical manner often arises. Furthermore, in many cases platforms are distributed across a region and it is desirable that extreme values at different locations are consistent.

This paper proposed a method for achieving spatially consistent directional design values using different data sets, and applies it to the reanalysis of a structure in the central North Sea. The method is based upon a Bayesian hierarchical methodology. The prior distributions of the parameters of the directional extreme values distribution are described in terms of the latitude and longitude and then data at a point used to determine the posterior distribution. The paper begins by discussing previous work on Bayesian and spatiodirectional methods. It continues in section 3 by describing the method, and then in section 4 the results are presented. Finally, the conclusions are discussed in section 5 and further work suggested.

2 Background

Bayesian methodology, spatiodirectional models and extreme value analysis techniques have been widely used in environmental studies and engineering design. However the use of all three in combination is less common. For example, a review of the application of Bayesian methods in extreme value analysis is provided in Coles and Powell (1996). In particular, they consider the problem of estimating extreme wind speeds by defining the prior distribution from a analysis of a number of sites across a region. Wikle et al. (2001) use a hierarchical spatiotemporal model to derive high resolution estimates of the distribution of a wind field over a large area. They achieve this by combining wind data from satellites (scatterometer) with the results of weather center wind fields (NCEP). Wikle (2003) review the application of hierarchical models in environmental science. In particular, they focuss on spatiotemporal models and break the modeling down into three components: the data conditional on both the process and the parameters, the process conditional on the parameters, and the parameters themselves. More recently, Vanem et al. (2011) have used a hierarchical Bayesian approach in order to determine spatial and temporal variations in the mean significant wave height across a region of the North Atlantic.

In terms of extreme value analysis, Coles and Tawn (1996) use a Bayesian framework to calculate extreme rainfall and use this to provide design criteria that include the uncertainty. Coles and Tawn (2005) estimate extreme storm surges incorporating the seasonal variation in the parameters into the model. They also discuss a method by which the spatial variation in the parameters can be considered in order to provide spatially smooth estimates along a coastline. Finally, Jonathan and Ewans (2011) provide spatial estimates of the one hundred year return period significant wave height in the Gulf of Mexico by using a spatiodirectional model with the smoothness in the parameters of the extreme value distribution ensured by using a natural thin plate spline algorithm.

3 Method

Point estimates of extreme wave height and structural response (base shear) have been calculated using the method of Tromans and Vanderschuren (1995). The procedure is outlined below for the latter, as the similar, but simpler procedure is followed for wave heights.

1. Events associated with peak structural response (storms) are identified within a time series of metocean parameters (wave heights, wave periods, current speeds, wind speeds and their associated directions). Where the response is calculated using a global load model

$$\begin{aligned} X &= A_1u^2 + A_2uaT\Phi\cos\theta \\ &+ A_3\Phi^2a^2 + A_4u\Phi a^2\cos\theta/T \\ &+ A_5\Phi^2a^3/T^2 + A_6\Phi^2a^2T^2 \\ &+ A_7W^2\,, \end{aligned} \tag{1}$$

   where a is the wave amplitude, u the depth averaged current speed, W the wind speed, T the associated wave period, Φ the wave kinematics factor, θ the angle between the wave and the current, and A_{1-7} coefficients to be determined for different structures.

2. The distribution of the response associated with each event, $P(X|s)$, is calculated

$$P(X|s) = \prod_i P(X|i)\,, \tag{2}$$

where $P(X|i)$ is the distribution within an interval and the product is over all intervals i in the event. The most probable maximum response, X_{mp}, associated with each event is calculated as $P(X_{mp}|s) = 1/e$.

3. The distribution of most probable maximum response is estimated by fitting a Generalised Pareto distribution to peaks over a threshold.

$$P(X_{mp}) \quad = \quad 1 - (1 + \epsilon(X_{mp} - \gamma)/\sigma)^{(-1/\epsilon)} \,, \tag{3}$$

   where σ is the scale parameter, ϵ the shape, and γ the location. This is applied on the basis that provided the parent distribution is within the domain of attraction of an extreme value distribution, the Generalised Pareto distribution asymptotically models the tail (Pickands, 1975). Although, arguably the most probable maximum response is not a peak as such, but a derived parameter. The distribution $P(X_{mp})$ reflects the long term distribution of the environmental forcing.

4. The distribution of the response,

$$P(X) \quad = \quad \int P(X|X_{mp}) p(X_{mp}) dX_{mp} \,, \tag{4}$$

   incorporating both the long-term, $p(X_{mp})$, and the short-term variability, $P(X|X_{mp})$, is calculated and the return period value of interest determined. Where the asymptotic result (distribution of maxima in a narrow banded process)

$$P(X|X_{mp}) \quad \approx \quad \exp[-\exp(-\ln N((X/X_{mp})^{\beta} - 1))] \,, \tag{5}$$

   can be used if the short-term variation is described by a Weibull distribution (shape parameter β), where N is the number of extremes in an event and $\ln N \approx 8$.

The directional distribution of the response is included by allowing the shape and scale parameters of the extreme value distribution to vary as a function of the direction at the peak of the storm,

$$\begin{aligned} \sigma(\theta) &= a_0 + \sum_{i=1}^{N} (a_i \cos(i\theta) + b_i \sin(i\theta)) \\ \epsilon(\theta) &= c_0 + \sum_{i=1}^{N} (c_i \cos(i\theta) + d_i \sin(i\theta)) \,, \end{aligned} \tag{6}$$

as suggested by Robinson and Tawn (1997), and where the number of Fourier components N is typically small. This has previously been applied to wave heights by Jonathan and Ewans (2007) and the importance of including covariate effects, such as direction, demonstrated in Jonathan et al. (2008).

The spatial variation in extreme wave height and jacket response across the central North Sea has been defined using a hierarchical Bayesian approach. This uses prior belief about the extreme value distributions at a particular location. The general assumption that is inherent is that the parameters vary somewhat smoothly across the region. There are two principle advantages to this approach: there is less variation in estimates across the region (particularly when covariate effects such as directionality are considered); shorter data sets (such as from measurements) can be incorporated naturally.

The Bayesian approach follows from Bayes' Theorem

$$p(\theta|x) = \frac{p(\theta)p(x|\theta)}{\int_\theta p(\theta)p(x|\theta)d\theta}\,, \tag{7}$$

where $p(\theta)$ is the prior distribution (reflecting prior knowledge about the parameters), $p(x|\theta)$ is defined by the data at a point (in this case the likelihood) and $p(\theta|x)$ is the resulting posterior distribution. Hence, the posterior distribution is a function of both the data at a point and the prior distribution. The more data (the longer the data set), the less that the prior distribution influences the result. Although in the application to extreme value analysis Coles and Powell (1996) argue that as a longer data set becomes available the threshold will often increase and the number of independent peak events may not increase at all. Furthermore, it is obviously possible to construct prior distributions that are inherently more or less informative; the preference here is for the latter.

The entire distribution can be used to include parameter uncertainty in the estimates, through

$$p(X|x) = \int_\theta p(X|\theta)p(\theta|x)d\theta\,, \tag{8}$$

however, this study is concerned with estimating design values, and hence, only the most probable value (the mode of the posterior distribution) is of interest. This has the computational advantage of not requiring the denominator in equation 7 to be evaluated. However, the methodology could also be applied to load factor studies where equation 8 must be used.

4 Results

In order to determine the prior distribution, data at 16 Nextra (Oceanweather) grid points roughly covering an area of the central North Sea spanning $56° - 62°N$ and $0° - 4°E$, has been analysed. The spatial variation in the estimates of the parameters of a Generalised Pareto distribution have been determined using the following function of latitude (X) and longitude (Y)

$$\begin{aligned} \upsilon(X,Y) &= \upsilon_a + \upsilon_b X + \upsilon_c Y + \upsilon_d XY + \upsilon_e X^2 + \upsilon_f Y^2 + \varepsilon_\upsilon \\ \epsilon(X,Y) &= \epsilon_a + \epsilon_b X + \epsilon_c Y + \epsilon_d XY + \epsilon_e X^2 + \epsilon_f Y^2 + \varepsilon_\epsilon\,, \end{aligned} \tag{9}$$

where the transformation $\upsilon = \sigma(1+\epsilon)$ has been applied so that estimates σ and ϵ are uncorrelated, and ε is a truncated Normal distribution with zero mean and a variance to be determined.

An example of the applicability of the parametric form described in equation 9 is shown in figures 1 and 2, the first of which is the variation in ϵ for the distribution of H_{mp} over the region, and the second the results of the fit. The result for the scale parameter is similar. The residual is shown in figure 3, and indicates that a Normal distribution is plausible.

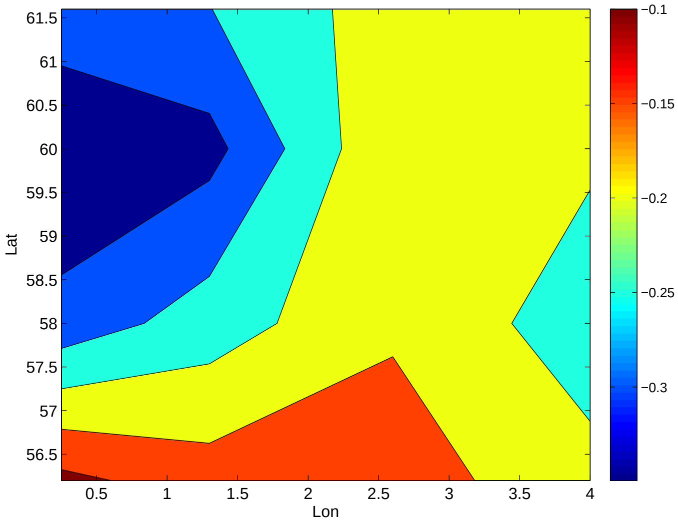

Figure 1: The variation in the mean of the Generalised Pareto shape parameter, ϵ, for Hmp over the region.

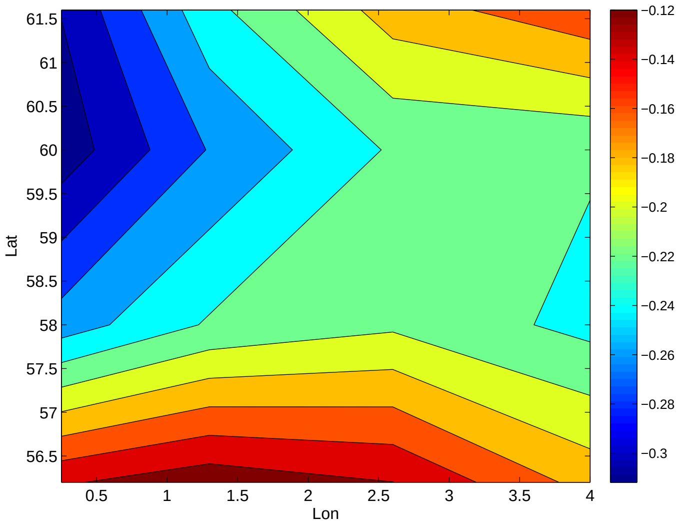

Figure 2: The fit to the mean of the Generalised Pareto shape parameter, ϵ, for Hmp over the region.

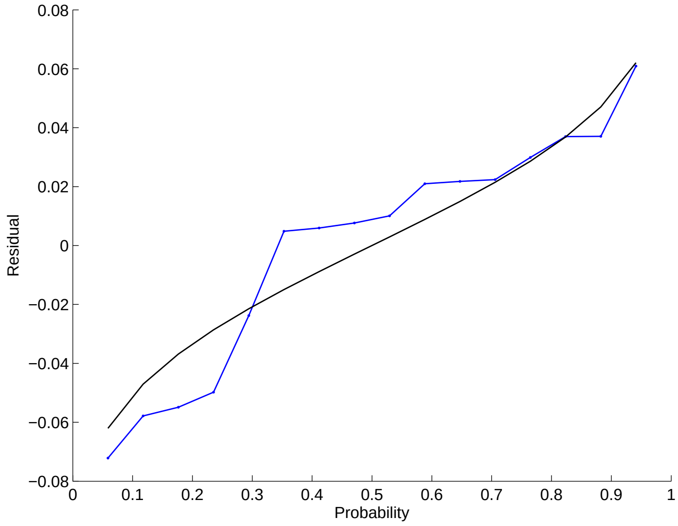

Figure 3: The residual of the fit to the Generalised Pareto shape parameter, ϵ, for Hmp over the region. The solid black line is the fit using a Normal distribution.

Point estimates (normalised) of the one hundred year return period most probable maximum wave height across the central North Sea are shown in figure 4 using all of the Nextra grid points within the region. Estimates of the one hundred year return period most probable maximum wave height (again normalised) calculated from the mode of the posterior distribution are shown in figure 5. The variation in Bayesian estimate of the wave height is smoother than that presented in figure 4, however, much of the spatial variation has been retained. This is to be expected as the data set used for wave height is 21 years of the continuous Nextra hindcast (rather than shorter measured data sets), and hence, the data tends to dominate the prior distribution.

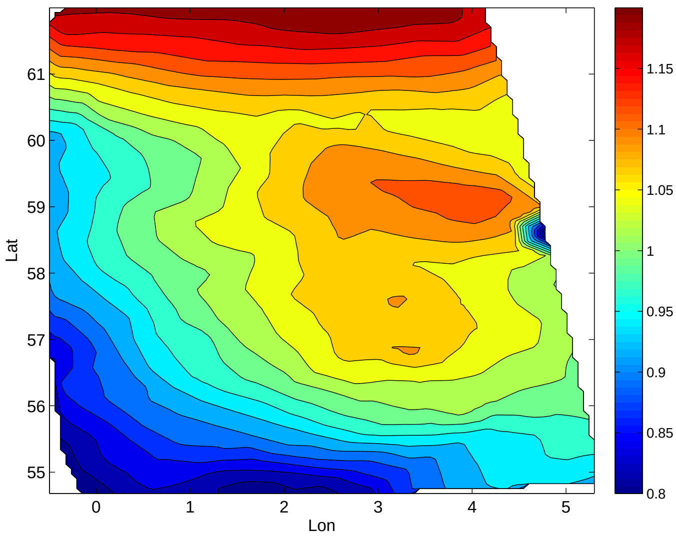

Figure 4: One hundred year return period estimate of the most probable maximum wave height H_{mp} (normalised) across the central North Sea from point estimates.

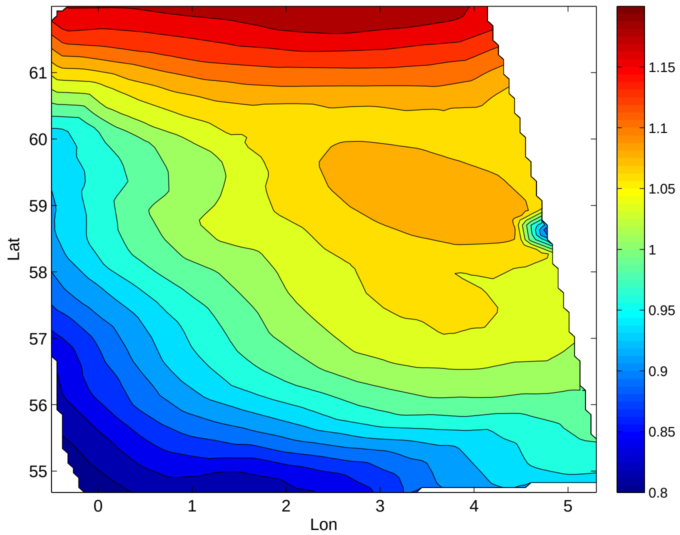

Figure 5: One hundred year return period estimate of the most probable maximum wave height H_{mp} (normalised) across the central North Sea from the mode of the posterior distribution.

The method has been applied to the reanalysis of a platform in the central North Sea. The omnidirectional fit to the distribution of most probable maximum wave height is shown in figure 6. In this case the posterior distribution lies between the prior and the point estimate, but the differences are small.

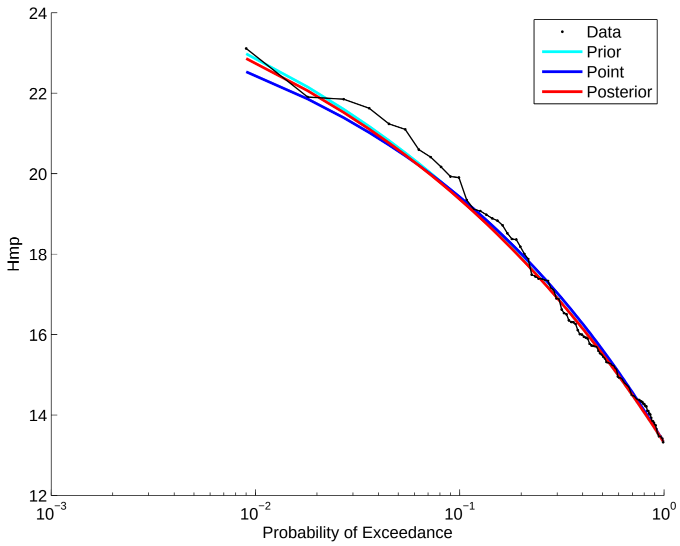

Figure 6: The distribution of most probable maximum wave height, Hmp, using 21 years of data from one location in the central North Sea.

Whilst there is some advantage in using a Bayesian approach for omnidirectional design values in cases where long data sets are available, the method is more useful for determining conditional distributions using shorter data sets. An example, is the directional distribution of base shear on a jacket platform. The distributions have been derived using the combined Nextra wave hindcast and DHI current model, which spans 9 years. For this example, the directional distribution of the parameters is described by a single Fourier component

$$\begin{aligned}\sigma(\theta) &= a_0 + a_1 \cos\theta + b_1 \sin\theta \\ \epsilon(\theta) &= c_0 + c_1 \cos\theta + d_1 \sin\theta\,,\end{aligned} \quad (10)$$

indicating that the spatial variation in six parameters must be estimated. In practice the likelihood ratio test can be used to determine the number of Fourier components required (Jonathan and Ewans, 2007). The spatial model that has been applied is the same as that used for omnidirectional estimates described in equation 9. This results in forty two parameters that must be determined (seven for each of a_0, a_1 etc: six for the spatial variation in the mean and one for the variance).

The results at the platform of interest are shown in figure 7 where the point estimate appears to be dominated by two large events, one from the North and one from the West, resulting in a very directional distribution. In contrast, the prior distribution, which has been derived using data from across the region, is much less directional. The resulting posterior distribution is somewhere between the two.

The maximum response for each direction is shown in table 1 and has been calculated by applying equations 4 and 5. The application is the reanalysis of an existing platform, and hence, the optimal directional distribution of loading that ensures the omnidirectional reliability has been determined by considering the directional distribution of the structure's utilisation (Forristall, 2004; Jonathan et al., 2008). Equation 1 has been inverted to provided a set of environmental conditions that correspond to the one hundred year response. One set of such conditions is presented in table 2. In practice, a number of such sets of conditions would be applied.

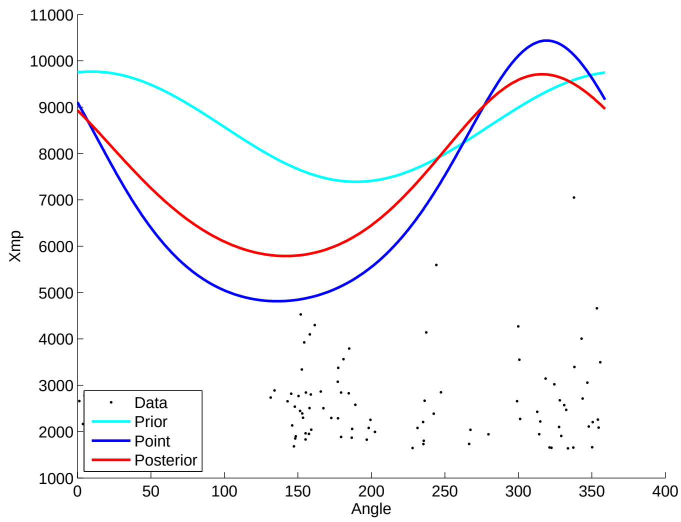

Figure 7: The one hundred year return period most probable maximum response X_{mp} (m) with direction.

	Return Period (yrs)			
Direction	1	10	100	10000
Omni	4171	6659	9248	14885
N	4586	7299	9965	15479
NE	4447	6770	8909	13197
E	4191	6168	7946	11512
SE	3957	5764	7404	10707
S	3856	5713	7457	10995
SW	3955	6105	8258	12781
W	4223	6816	9559	15638
NW	4497	7356	10360	16959

Table 1: Directional return period estimates of one hundred year return period maximum base shear (the units are arbitrary as equation 1 is inverted to determine design conditions).

Direction	H (m)	T (s)	$\bar{C}$ (m/s)
Omni	26.7	14.9	0.54
N	29.2	14.4	0.32
NE	27.6	14.0	0.33
E	26.2	13.7	0.31
SE	25.4	13.5	0.30
S	25.4	13.5	0.31
SW	26.6	13.8	0.33
W	28.7	14.3	0.32
NW	29.9	14.6	0.30

Table 2: One hundred year return period metocean criteria for wave dominated events with shorter periods, and associated currents: wave height H, wave period T and depth averaged current speed $\overline{C}$.

The utility of the approach can more generally be demonstrated by determining the bias and coefficient of variation (COV) in estimates from artificially truncated data sets. Point and Bayesian estimates of the shape, scale and one hundred year return period values have been determined by bootstrapping data from one location in the North Sea. Ten thousand random samples (with replacement) have been analysed for different lengths of a data set and the bias and coefficient of variation of the estimated parameters calculated. These can be expressed as follows

$$Bias = \frac{\mu(\hat{X}) - X}{X}, \tag{11}$$

where $\mu(\hat{X})$ is the mean of the estimates of a parameter of interest and X the mode of the posterior distribution estimated using the full data set, and

$$COV = \frac{\sigma(\hat{X})}{\mu(\hat{X})}, \tag{12}$$

where here $\sigma(\hat{X})$ is the standard deviation of the estimate of the parameter of interest.

Table 3 indicates that, not surprisingly, shorter data sets have a greater bias and that as more data is available the point and Bayesian estimates converge. More significantly, table 4 demonstrates that the Bayesian estimate has a very low COV even when using very short data sets (the estimator is more efficient) and that the COV increases slightly as more data is available. Of course, this is to be expected, as when only a short data set is available the prior dominates. This provides a logical method for combining shorter measured data sets with longer period spatial hindcast data and achieving stable estimates of extreme values. This approach is demonstrated in figure 8 (which should be compared to figure 6), where a very short truncated data set is analysed. Of course, none of this considers the uncertainty in the spatial model itself, which could be included by using an extra hierarchy of parameters that describe the distributions of $\upsilon_{a,b,c,d,e,f}$ and $\epsilon_{a,b,c,d,e,f}$, but that is beyond the scope of this study.

	Point Estimate			Bayesian Estimate		
Duration (yrs)	σ	ϵ	Hmp	σ	ϵ	Hmp
2.5	0.148	2.318	-0.226	-0.015	-0.016	-0.005
5	-0.140	0.503	-0.007	-0.029	-0.012	0.036
7.5	-0.150	-0.269	0.063	-0.017	-0.017	0.039
10	-0.131	-0.068	0.008	-0.030	0.009	0.025
12.5	0.024	0.259	-0.028	-0.012	0.014	0.010
15	0.000	0.223	-0.034	-0.019	0.025	0.001
17.5	0.057	0.335	-0.046	-0.010	0.033	-0.007
20	0.053	0.194	-0.023	0.000	0.000	0.000

Table 3: The bias associated with point and bayesian estimates of the scale σ, shape ϵ and one hundred year estimate of the most probable maximum wave height Hmp for different lengths of data set.

	Point Estimate			Bayesian Estimate		
Duration (yrs)	σ	ϵ	Hmp	σ	ϵ	Hmp
2.5	0.582	0.722	0.129	0.009	0.008	0.005
5	0.488	1.376	0.187	0.020	0.029	0.015
7.5	0.309	1.036	0.113	0.023	0.028	0.015
10	0.201	0.579	0.080	0.025	0.032	0.017
12.5	0.163	0.359	0.053	0.026	0.038	0.019
15	0.147	0.341	0.050	0.026	0.041	0.019
17.5	0.129	0.282	0.045	0.026	0.044	0.020
20	0.141	0.299	0.042	0.027	0.041	0.019

Table 4: The coefficient of variation (COV) associated with point and bayesian estimates of the scale σ, shape ϵ and one hundred year estimate of the most probable maximum wave height Hmp for different lengths of data set.

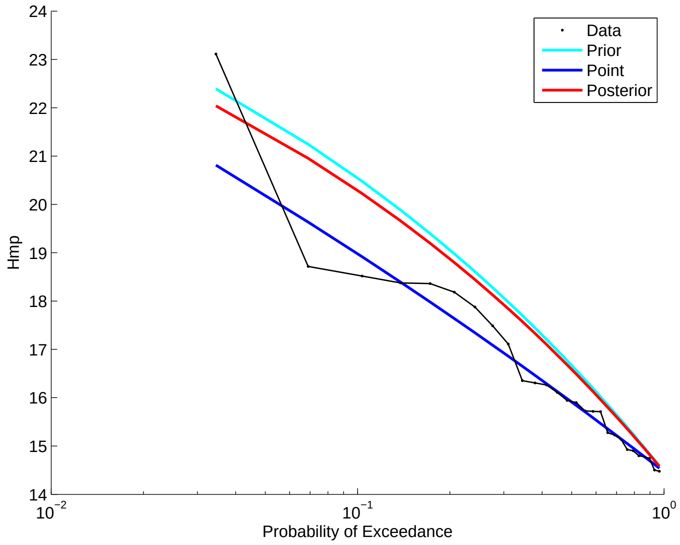

Figure 8: The distribution of most probable maximum wave height, Hmp, using five years of data from one location in the central North Sea.

5 Conclusions and Further Work

A Bayesian spatiodirectional model for wave heights and structural response has been proposed where the prior distributions of the parameters of the extreme value distribution are a function of location. The method has been applied to the central North Sea and estimates of the one hundred year return period wave height shown to be smoother than point estimates, but to still retain local features. The method has also been applied in order to determine directional values of extreme structural response for a North Sea platform and a highly directional distribution of loading moderated by data from other locations.

The methodology could be extended by considering the distribution of the coefficients that describe the spatial distribution of the parameters of the extreme value distribution (by increasing the number of levels to the model). It could also be applied in load factor studies in order to define safety factors that consider the spatial variation in uncertainty and variability across a region. Furthermore, whilst in this study it has been proposed that hindcast data could be used to determine the prior distribution. Alternatively, this suggestions could also be used in reverse in order to provide a spatially varying correction to a hindcast: the prior distribution would be defined using measurements and then hindcast data used at a point of interest in order to determine the posterior distribution.

References

S.G. Coles and E.A. Powell. Bayesian methods in extreme value modelling: A review and new developments. *International Statistical Review*, 64, 1996.

S.G. Coles and J.A. Tawn. A bayesian analysis of extreme rainfall data. *Applied Statistics*, 45, 1996.

S.G. Coles and J.A. Tawn. Bayesian modelling of extreme surges on the uk east coast. *Philosophical Transactions of the Royal Society A*, 363, 2005.

G.Z. Forristall. On the use of directional wave criteria. *Journal of Waterway, Port, Coastal and Ocean Engineering*, 130:272–275, 2004.

P. Jonathan and K. Ewans. The effect of directionality in extreme wave design criteria. *Ocean Engineering*, 34:1977–1994, 2007.

P. Jonathan and K. Ewans. A spatiodirectional model for extreme waves in the gulf of mexico. *Journal of OMAE*, 133, 2011.

P. Jonathan, K. Ewans, and G. Forristall. Statistical estimation of extreme ocean environments: The requirement for modelling directionality and other covariate effects. *Ocean Engineering*, 35:1211–1225, 2008.

P. Jonathan, J. Flynn, and K. Ewans. Joint modeling of wave spectral parameters for extreme sea states. In *11th International Workshop on Wave Hindcasting and Forecasting*, 2009.

Oceanweather. Nextra hindcast.

J. Pickands. Statistical inference using order statistics. *Annals of Statistics*, 1975.

M.E. Robinson and J.A. Tawn. Statistics of extreme sea currents. *Applied Statistics*, 46:183–205, 1997.

P.S. Tromans and L. Vanderschuren. Response based design conditions in the North Sea: application of a new method. In *OTC 7683, Offshore Technology Conference, Houston*, 1995.

E. Vanem, A.B. Huseby, and B. Natvig. A bayesian-hierarchical space-time model for significant wave height data. In *OMAE2011-49716*, 2011.

C.K. Wikle. Hierarchical models in environmental science. *International Statistical Review*, 2003.

C.K. Wikle, R.F. Milliff, D. Nychka, and L.M. Berliner. Spatiotemporal hierarchical bayesian modeling: tropical ocean surface winds. *Journal of the American Statistical Association*, 96, 2001.