Ocean Modelling 107 (2016) 1–13

Contents lists available at ScienceDirect

Ocean Modelling

journal homepage: www.elsevier.com/locate/ocemod

Virtual Special Issue
COWCLIP

Spatial Bayesian hierarchical modelling of extreme sea states

Colm Clancy[a,b], John O'Sullivan[a,b,*], Conor Sweeney[a,b], Frédéric Dias[a,b,c], Andrew C. Parnell[a,b]

[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

ARTICLE INFO

Article history:
Received 22 February 2016
Revised 21 September 2016
Accepted 28 September 2016
Available online 29 September 2016

Keywords:
Bayesian hierarchical modelling
Spatial modelling
Extreme value analysis
Ocean waves
Significant wave height

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 99.8th 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 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.

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 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). The background theory for this extreme value analysis is outlined in Section 2 below.

Models of 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 given the data (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 3. 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 implementation details, is given in Section 4. The results are presented in Section 5 with a discussion and conclusions in Section 6.

* Corresponding author.
E-mail address: john.o-sullivan.2@ucdconnect.ie (J. O'Sullivan).

http://dx.doi.org/10.1016/j.ocemod.2016.09.015
1463-5003/

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 (Pickands, 1975). We assume that our sequence of independent random variables, $Z_1, Z_2, \ldots,$ satisfies the extremal types theorem described above. For large enough threshold 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 $\{y : y > 0 \text{ and } (1 + \xi y/\sigma) > 0\}$. 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); Vinoth and Young (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 (1), 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

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

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 (3). Details of these are given in Coles (2001), along with a discussion of other methods for fitting and analysing uncertainty, such as the profile likelihood method.

A further 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 (4). 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 explic-