Ocean Engineering 88 (2014) 520–532

Contents lists available at ScienceDirect

Ocean Engineering

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

Return level estimation from non-stationary spatial data exhibiting multidimensional covariate effects

Philip Jonathan [a,*], David Randell [a], Yanyun Wu [a], Kevin Ewans [b]

[a] Shell Research Ltd., Manchester M22 0RR, UK
[b] Sarawak Shell Bhd., 50450 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 4 October 2013
Accepted 2 July 2014
Available online 31 July 2014

Keywords:
Offshore design
Extreme
Return value
Covariate
B-spline
Storm severity

ABSTRACT

Careful modelling of non-stationarity is critical to reliable specification of marine and coastal design criteria. We present a spline based methodology to incorporate spatial, directional, temporal and other covariate effects in extreme value models for environmental variables such as storm severity. For storm peak significant wave height events, the approach uses quantile regression to estimate a suitable extremal threshold, a Poisson process model for the rate of occurrence of threshold exceedances, and a generalised Pareto model for size of threshold exceedances. Multidimensional covariate effects are incorporated at each stage using penalised (tensor products of) B-splines to give smooth model parameter variation as a function of multiple covariates. Optimal smoothing penalties are selected using cross-validation, and model uncertainty is quantified using a bootstrap re-sampling procedure. The method is applied to estimate return values for large spatial neighbourhoods of locations, incorporating spatial and directional effects. Extensions to joint modelling of multivariate extremes, incorporating extremal spatial dependence (using max-stable processes) or more general extremal dependence (using the conditional extremes approach) are outlined.

1. Introduction

Availability of comprehensive met-ocean data allows the effect of heterogeneity (or non-stationarity) of extremes with respect to direction, season and location to be accommodated in estimation of design criteria. Jonathan and Ewans (2013) review statistical modelling of extremes for marine design.

Capturing covariate effects in extreme sea states is important when developing design criteria. In previous work (e.g. Jonathan and Ewans, 2007a; Ewans and Jonathan, 2008) it has been shown that omnidirectional design criteria derived from a model that adequately incorporates directional covariate effects can be materially different from a model which ignores those effects (e.g. Jonathan et al., 2008). Directional return values derived from a directional model can be heavier tailed than that derived from a direction-independent approach, indicating that large values of extreme events are more likely than we might anticipate were we to base our beliefs on estimates which ignore directionality. Similar effects have been demonstrated for seasonal covariates (e.g. Anderson et al., 2001; Jonathan et al., 2008).

There is a large body of statistical literature regarding modelling of covariate effects in extreme value analysis; for example, Davison and Smith (1990) or Robinson and Tawn (1997). The case for adopting an extreme value model incorporating covariate effects is clear, unless it can be demonstrated statistically that a model ignoring covariate effects is no less appropriate. Chavez-Demoulin and Davison (2005) and Coles (2001) provide straight-forward descriptions of a non-homogeneous Poisson model in which occurrence rates and extremal properties are modelled as functions of covariates. Scotto and Guedes-Soares (2000) describe modelling using non-linear thresholds. A Bayesian approach is adopted (Coles and Powell, 1996) using data from multiple locations, and by Scotto and Guedes-Soares (2007). Spatial models for extremes (Coles and Casson, 1998; Casson and Coles, 1999) have also been used, as have models (Coles and Tawn, 1996, 2005) for estimation of predictive distributions, which incorporate uncertainties in model parameters. Ledford and Tawn (1997) and Heffernan and Tawn (2004) discuss the modelling of dependent joint extremes. Chavez-Demoulin and Davison (2005) also describe the application of a block bootstrap approach to estimate parameter uncertainty and the precision of extreme quantile estimates, applicable when dependent data from neighbouring locations are used. Jonathan and Ewans (2007b) use block bootstrapping to evaluate uncertainties associated with extremes in storm peak significant wave heights in the Gulf of Mexico. Guedes-Soares and Scotto (2001) discuss the estimation of quantile uncertainty. Eastoe (2007) and Eastoe and Tawn (2009) illustrate an approach to removing covariate effects from a sample of extremes prior to model estimation.

* Corresponding author.
E-mail address: philip.jonathan@shell.com (P. Jonathan).

http://dx.doi.org/10.1016/j.oceaneng.2014.07.007
0029-8018/

One of the first examinations of spatial characteristics of extreme wave heights was reported by Haring and Heideman (1978) for the Gulf of Mexico. They performed extremal analysis of the ODGP hurricane hindcast database (Ward et al., 1978) at a number of continental shelf locations from Mexico to Florida, and concluded that there was no practical difference between the sites, but they did observe a gradual reduction in extreme wave heights with decreasing water depth. Chouinard et al. (1997) took the opportunity to re-examine the spatial behaviour of extremes in the Gulf of Mexico, when the GUMSHOE hindcast database became available. Jonathan and Ewans (2011b) used thin-plate splines to model the spatial characteristics of events in the Gulf of Mexico. Extending the thin-plate spline formulism to include other (possibly periodic) covariates is difficult; instead, the sample is typically pre-processed to remove the influence of all covariates other than the (2-D) spatial, prior to model estimation using thin-plate splines. Models estimated in this way suffer from the fact that interactions between various modelling steps (and the parameters estimated therein) cannot be easily quantified.

Characterising the joint structure of extremes for different environmental variables is also important for improved understanding of those environments. Yet many applications of multivariate extreme value analysis adopt models that assume a particular form of extremal dependence between variables without justification, or restrict attention to regions in which all variables are extreme. The conditional extremes model of Heffernan and Tawn (2004) provides one approach avoiding these particular restrictions. Extremal dependence characteristics of environmental variables also typically vary with covariates. Reliable descriptions of extreme environments should also therefore characterise any non-stationarity. Jonathan et al. (2013) extend the conditional extremes model of Heffernan and Tawn to include covariate effects, using Fourier representations of model parameters for single periodic covariates.

The last decade has seen the emergence of useable statistical models for spatial extremes based on max-stable processes, at least in academia. The application of max-stable processes is complicated due to unavailability of the full multivariate density function. Padoan et al. (2010) develop inferentially practical, likelihood-based methods for fitting max-stable processes derived from a composite likelihood approach. The procedure is sufficiently reliable and versatile to permit the simultaneous modelling of marginal and dependence parameters in the spatial context at a moderate computational cost. Davison and Gholamrezaee (2012) describe an approach to flexible modelling for maxima observed at sites in a spatial domain, based on fitting of max-stable processes derived from underlying Gaussian random process models. Generalised extreme value (GEV) margins as assumed throughout the spatial domain, and models incorporate standard geo-statistical correlation functions. Estimation and fitting are performed through composite likelihood inference applied to observations from pairs of sites. Davison et al. (2012) also provide a good introduction and review. Erhardt and Smith (2011) use approximate Bayesian computation to circumvent the need for a joint likelihood function by instead relying on simulations from the (unavailable) likelihood avoiding the need to construct composite likelihoods at higher computational cost.

In this work, we apply a marginal model for spatio-directional extremes to a sample of data for storm severity on the north west continental shelf of Western Australia. The model (developed in Section 2) adopts a penalised B-spline formulation to characterise smooth variation of extreme value parameters spatially and directionally. The North West Shelf application is then presented in Section 3. In Section 4, we discuss model extension to incorporate appropriate spatial extremal dependence, and also outline a non-stationary extension of the conditional extremes model of Heffernan and Tawn (2004).

2. Model

The objective is to estimate design criteria for individual locations within a spatial neighbourhood, accounting for spatial and storm directional variability of extremal characteristics.

2.1. Model components

Following the work of Jonathan and Ewans (2008, 2011b), summarised in Jonathan and Ewans (2013), we model storm peak significant wave height H_S^{sp}, namely the largest value of significant wave height H_S observed at each location during the period of a storm event. We assume that each of $\dot{n}_S$ independent storm peak events is observed at all of n_L locations within the neighbourhood under consideration. We therefore start with a total of $\dot{n} = \dot{n}_S \times n_L$ observations of H_S^{sp}. We refer to these as the sample $\{\dot{z}_i\}_{i=1}^{\dot{n}}$ of $\dot{n}$ storm peak significant wave heights observed at locations $\{\dot{x}_i, \dot{y}_i\}_{i=1}^{\dot{n}}$ with dominant wave directions $\{\dot{\theta}_i\}_{i=1}^{\dot{n}}$ (corresponding to the time of occurrence of H_S^{sp}, henceforth "storm directions"). We then proceed using the peaks over threshold approach as follows.

Extreme value threshold: We first estimate a threshold function ϕ above which observations $\dot{z}$ are assumed to be extreme. The threshold varies smoothly as a function of covariates ($\phi \triangleq \phi(\theta, x, y)$) and is estimated using quantile regression. As a result of thresholding, the same number of observations of smaller storm peak events is eliminated from all locations by construction. The number of storms peak events remaining per location reduces from $\dot{n}_S$ to n_S, so that the total number of observations (of threshold exceedances) for extreme value modelling is $n = n_S \times n_L$. We refer to these as the set of n threshold exceedances $\{z_i\}_{i=1}^{n}$ observed at locations $\{x_i, y_i\}_{i=1}^{n}$ with storm peak directions $\{\theta_i\}_{i=1}^{n}$.

Rate of threshold exceedance: We next estimate the rate of occurrence ρ of threshold exceedance using a Poisson process model with Poisson rate $\rho(\triangleq \rho(\theta, x, y))$.

Size of threshold exceedance: We estimate the size of occurrence of threshold exceedance using a generalised Pareto (henceforth GP) model. The GP shape and scale parameters ξ and σ are also assumed to vary smoothly as functions of covariates.

This approach to extreme value modelling follows that of Chavez-Demoulin and Davison (2005) and is equivalent to direct estimation of a non-homogeneous Poisson point process model (e.g., Dixon et al., 1998; Jonathan and Ewans, 2013).

2.2. Parameter estimation

Extreme value threshold: For quantile regression, we seek a smooth function ϕ of covariates corresponding to non-exceedance probability τ of H_S^{sp} given any combination of θ, x, y. We choose to estimate ϕ by minimising the quantile regression lack of fit criterion

$$\ell_\phi = \left\{ \tau \sum_{i, r_i \ge 0}^{n} |r_i| + (1-\tau) \sum_{i, r_i < 0}^{n} |r_i| \right\}$$

for residuals $r_i = z_i - \phi(\theta_i, x_i, y_i; \tau)$. We regulate the smoothness of the quantile function by penalising lack of fit for parameter roughness R_ϕ (with respect to all covariates), by minimising the revised penalised criterion

$$\ell_\phi^* = \ell_\phi + \lambda_\phi R_\phi$$

where the value of roughness coefficient λ_ϕ is selected using cross-validation to provide good predictive performance.

Rate of threshold exceedance: For Poisson modelling, we use penalised likelihood estimation. The rate ρ of threshold