OMAE2013-10187

MODELLING COVARIATE EFFECTS IN EXTREMES OF STORM SEVERITY ON THE AUSTRALIAN NORTH WEST SHELF

David Randell[*]
Shell Research Ltd.,Chester, CH1 3SH, UK.
david.randell@shell.com

Yanyun Wu
Shell Research Ltd.,Chester, CH1 3SH, UK.

Philip Jonathan
Shell Research Ltd.,Chester, CH1 3SH, UK.

Kevin Ewans
Sarawak Shell Bhd., 50450 Kuala Lumpur, Malaysia.

ABSTRACT

Careful modelling of covariate effects is critical to reliable specification of 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 . 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 resampling procedure. The method is applied to estimate return values for a large spatial neighbourhood of locations off the North West Shelf of Australia, incorporating spatial and directional effects.

1 Introduction

The availability of comprehensive metocean data allows the effect of the heterogeneity 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 of 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 omni-Design criteria derived from a model that adequately incorporates covariate effects can be materially different from a model which ignores those effects(e.g. Jonathan et al. [2008]). Directional storm peaks H_{S100} derived from a directional model can be heavier tailed than that derived from a direction independent approach, indicating that large values of storm peak H_S 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 statistics literature regarding modelling of covariate effects in extreme value analysis; see, e.g., 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 straightforward 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 mul-

---

[*]Address all correspondence to this author.

tiple 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 extremes data prior to model estimation.

One of the first examinations of the 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 data base (Ward et al. [1978]) at a number of continental shelf locations from Mexico to Florida, and concluded that there was not 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 data base became available. Jonathan and Ewans [2011] used thin-plate splines to model the spatial characteristics of events in the Gulf of Mexico. The problem with the thin plate splines approach is that it is difficult to extend to included other covariate effects such as direction or season and therefore data need to be pre-processed prior to modelling.

Here we introduce a spatio directional model for extremes and apply it to data from the North West continental shelf of Western Australia. The model incorporates non-parametric model of extremes using P-splines formulation to characterise the smooth variation of extreme value parameters in space and directionally.

2 Data

The data used in the modelling covers 6156 hindcast storm events over a 9x9 grid of locations on the North West continental shelf of Western Australia during the period 1970-2007. The climate of the area is monsoonal, and displays two distinct seasons, "winter" from April to September and "summer" from October to March, with very rapid transition seasons, generally in April and September/October between the two main seasons. The winter "dry" is the result of a steady easterly air flow (North East - South East) originating from over the Australian mainland and travelling over the Timor Sea (known as the South East Trade

Winds). The summer "wet" is the result of the North West monsoon, a steady, moist predominantly West-South West and to a lesser extent North West wind. Tropical cyclones occur during these months and are clearly the most important for extreme metocean criteria. Tropical cyclones originate from South of the equator in the eastern Indian Ocean and in the Timor and Arafura Seas. The most severe cyclones most often occur in the months of December and March-April, when sea-surface temperatures are warmest. In the area under consideration, most of the storms are tropical lows or developing storms, but they can be very severe, such as tropical cyclones Thelma (1998) and Neville (1992). Most of the storms pass through this area head in a West or South West direction before turning southwards.
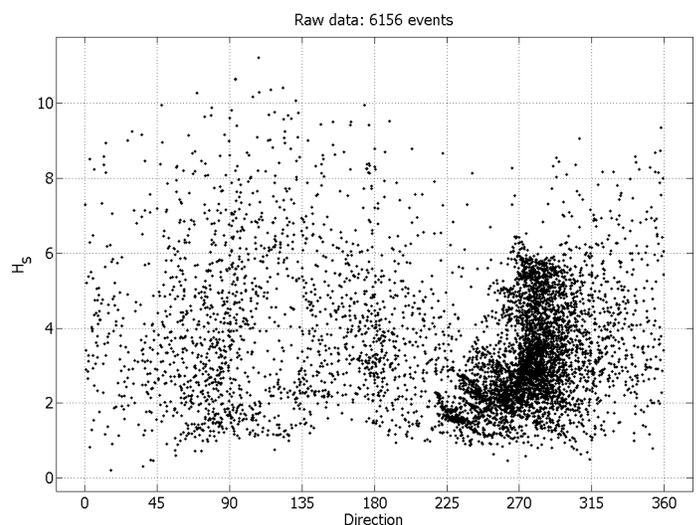
FIGURE 1. MARGINAL PLOT OF STORM PEAK H_S AGAINST DIRECTION. THERE IS CLEAR VARIABILITY WITH DIRECTION WITH THE MOST SEVERE STORMS SEEN FROM $90 - 130^O$. THERE ARE ALSO A LARGE NUMBER OF SMALL EVENTS FROM $250 - 290^O$.

In terms of the wave climate, which is the subject of this paper, the prevailing wave climate comprises contributions from Indian Ocean swell, winter easterly swell, westerly monsoonal swell, tropical cyclone swell, and locally generated wind-sea. The Indian Ocean Swell is a perennial feature typically, propagating from the South-West through North-West. The largest sea states are wind generated sea states associated with Tropical Cyclones.

Figure 1 shows marginal plot of storm peak H_s against direction. There is clear variability with direction with the largest waves seen from $90 - 130^o$. There are also a large number of small events from $250 - 290^o$. This difference in the rate of occurrence of events can also been seen in figure 2 which shows
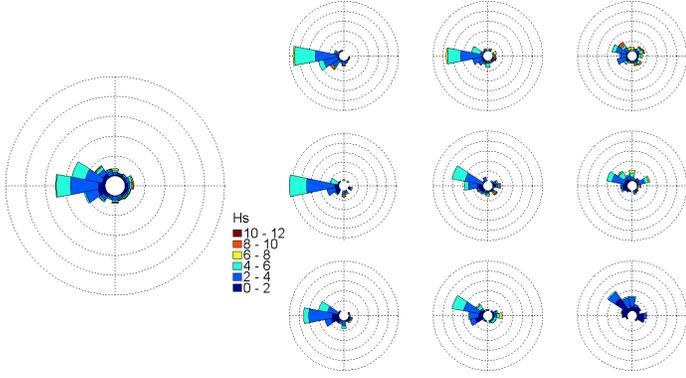
FIGURE 2. DIRECTIONAL ROSE HISTOGRAM PLOTS OF STORM PEAK H_S. THE COLOUR OF EACH BIN SHOWS PROPORTION OF DIFFERENT SIZED EVENTS WITHIN THAT BIN. LEFT-HAND PLOT SHOWS ROSE FOR ALL SITES POOLED. RIGHT-HAND PLOTS SHOW DIRECTION HISTOGRAMS FOR (FROM LEFT TO RIGHT FROM TOP) NW, N, NE, W, CENTRAL, E, SW, S AND SE LOCATIONS RESPECTIVELY.
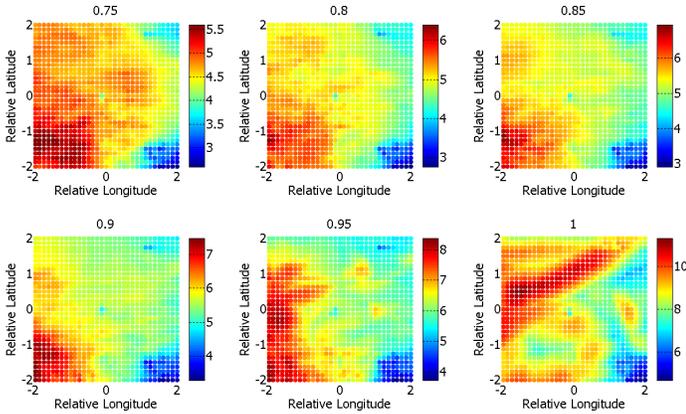
FIGURE 3. EMPIRICAL QUANTILES OF STORM PEAK H_S OVER SPATIAL DOMAIN OF DATA. CLEAR SPATIAL VARIABILITY, MOST SEVERE STORMS IN THE W TO SW. SMALLER EVENTS IN THE SE AND NE.

directional histograms of wave direction in different spatial locations. It can be seen that the vast majority of data comes from around $250 - 290^o$ however, these are associated with swell and not the big events, which are associated with intense local storms. Figure 3 shows empirical quantiles of storm peak H_s over spatial domain of data. There is clear spatial variability over spatial domain the biggest waves seen general in the West to South West and smaller events in the South East and North East.

3 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.

3.1 Model components

Following the work of Jonathan and Ewans [2008] & Jonathan and Ewans [2011], summarised in Jonathan and Ewans [2013], we model storm peak significant wave height, namely the largest value of significant wave height observed at each location during the period of a storm event. We assume that each storm event is observed at all locations within the neighbourhood under consideration. For a 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 storm peak directions $\{\dot{\theta}_i\}_{i=1}^{\dot{n}}$ (henceforth together referred to as covariates), we proceed using the peaks over threshold approach as follows.

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 ($\phi \overset{\triangle}{=} \phi(\theta, x, y)$) and is estimated using quantile regression. We retain 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 $\{\theta_i\}_{i=1}^{n}$ for further modelling.

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

Size of occurrence of threshold exceedance: We estimate the size of occurrence of threshold exceedance using a generalised Pareto (henceforth GP for brevity) 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 (see, e.g., Dixon et al. 1998, Jonathan and Ewans [2013]).

3.2 Parameter estimation

For quantile regression, we seek a smooth function ϕ of covariates corresponding to non-exceedance probability τ of storm peak H_S for any combination of θ, x, y. We estimate ϕ by minimising the quantile regression lack of fit criterion

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

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

Copyright © 2013 by ASME

roughness R_ϕ (with respect to all covariates), by minimising the 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.

For Poisson modelling, we use penalised likelihood estimation. The rate ρ of threshold exceedance is estimated by minimising the roughness-penalised (negative log) likelihood

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

where R_ρ is parameter roughness with respect to all covariates, λ_ρ is again evaluated using cross-validation, and Poisson (negative log) likelihood is given by

$$\ell_\rho = -\sum_{i=1}^{n} \log \rho(\theta_i, x_i, y_i) + \int \rho(\theta, x, y) d\theta dx dy$$

The generalised Pareto model of size of threshold exceedance is estimated in a similar manner by minimising the roughness penalised (negative log) GP likelihood

$$\ell_{\xi,\sigma}^* = \ell_{\xi,\sigma} + \lambda_\xi R_\xi + \lambda_\sigma R_\sigma$$

where R_ξ and R_σ are parameter roughnesses with respect to all covariates, λ_ξ and λ_σ are evaluated using cross-validation, and GP (negative log) likelihood is given by

$$\ell_{\xi,\sigma} = \sum_{i=1}^{n} \log \sigma_i + \frac{1}{\xi_i} \log(1 + \frac{\xi_i}{\sigma_i}(z_i - \phi_i))$$

where $\phi_i = \phi(\theta_i, x_i, y_i)$, $\xi_i = \xi(\theta_i, x_i, y_i)$ and $\sigma_i = \sigma(\theta_i, x_i, y_i)$, and a similar expression is used when $\xi_i = 0$ (see Jonathan and Ewans 2013). In practice, we set $\lambda_\xi = \kappa \lambda_\sigma$ for prespecified constant κ, so that only one cross-validation loop is necessary. The value of κ is estimated by inspection of the relative smoothness of ξ and σ with respect to covariates.

3.3 Parameter smoothness

Physical considerations suggest that we should expect the model parameters ϕ, ρ, ξ and σ to vary smoothly with respect to covariates θ, x, y. For estimation, this can be achieved by expressing each parameter in terms of an appropriate basis for the domain D of covariates, where $D = D_\theta \times D_x \times D_y$. $D_\theta = [0, 360)$
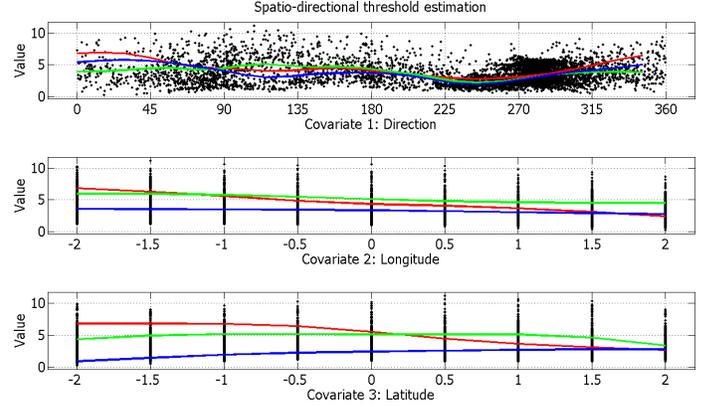
FIGURE 4. MARGINAL PLOTS OF STORM PEAK H_S AGAINST DIRECTION, RELATIVE LONGITUDE AND RELATIVE LATITUDE. A NON-EXCEEDANCE THRESHOLD PROBABILITY OF 0.5 WAS USED AND ϕ WAS ESTIMATED USING QUANTILE REGRESSION. THE TOP PLOT SHOWS THE 50% QUANTILE AS A FUNCTION OF DIRECTION FOR THE SW (RED), CENTRE (GREEN) AND NE (BLUE) RESPECTIVLY. THE MIDDLE PLOT SHOWS THE 50% QUANTILE AS A FUNCTION OF LONGITUDE FOR 0° AT THE MINIMUM LATITUDE (RED), FOR 120° FOR THE CENTRAL LATITUDE (GREEN) AND 270° FOR THE MAXIMUM LATITUDE (BLUE). THE BOTTOM PLOT SHOWS THE 50% QUANTILE AS A FUNCTION OF LATITUDE FOR 0° AT THE MINIMUM LONGITUDE (RED), FOR 120° FOR THE CENTRAL LONGITUDE (GREEN) AND 270° FOR THE MAXIMUM LONGITUDE (BLUE).

is the (marginal) domain of storm peak directions, and D_x, D_y are the domains of x- and y-values (e.g. longitudes and latitudes) under consideration. For each marginal domain, we calculate a B-spline basis matrix for an index set of $m(< n)$ combinations of covariate values (potentially we could calculate the basis matrix for each of the n observations, but usually avoid the case $m = n$ for computation efficiency). Specifically, for D_θ, we calculate basis matrix B_θ $(m \times p_\theta)$ such that the value of any function η at each point in the index set can be expressed as $\eta = B_\theta \beta$ for some vector β $(p_\theta \times 1)$ of basis coefficients.

Note that periodic marginal bases can be specified if appropriate (e.g. for D_θ). Moreover, we can define a basis matrix for the three dimensional domain D using Kronecker products of the marginal basis matrices. Thus

$$B = B_\theta \otimes B_x \otimes B_y$$

provides a $(m \times p)$ basis matrix (where $p = p_\theta p_x p_y$) for modelling each of ϕ, ρ, ξ and σ, any of which can be expressed in the form $B\beta$ for some $(p \times 1)$ vector of basis coefficients. Model estimation therefore reduces to estimating appropriate sets of basis coefficients for each of ϕ, ρ, ξ and σ.
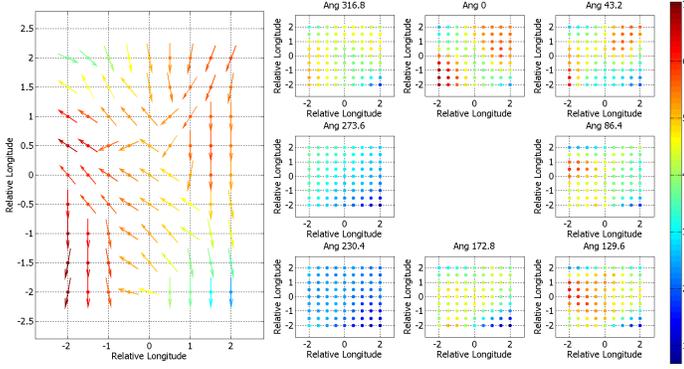
FIGURE 5. SPATIO-DIRECTIONAL PLOT FOR QUANTILE EX-CEEDANCE THRESHOLD, ϕ. THE 8 RIGHT-HAND PLOTS SHOW THE 50% THRESHOLD VALUES OF ϕ FOR EACH LOCATION FOR 8 DIRECTIONS. THE LEFT-HAND PLOT SHOWS THE DI-RECTION FROM WHICH ϕ IS LARGEST.

The roughness R of any function can be easily evaluated on the index set (at which $\eta = B\beta$). Following the approach of Eilers and Marx (e.g. Eilers and Marx 2010, Eilers and Marx [1996]), we define roughness using

$$R = \beta'P\beta$$

where P can be easily evaluated for the marginal and three dimensional domains. The form of P is motivated by taking differences of neighbouring values of β, thereby penalising lack of local smoothness. The values of p_θ, p_x, p_y are functions of the number of spline knots for each marginal domain, and also depend on whether spline bases are specified as periodic (e.g D_θ) or not (e.g D_x and D_y).

3.4 Algorithms

Quantile regression estimation is performed by direct minimisation of the criterion ℓ_ϕ^* from a good starting solution. The starting solution is estimated by fitting a smoothing spline to estimates of the spatio-directional quantile with non-exceedance probability τ at each of the m covariate combinations in the index set. Poisson and generalised Pareto estimation was achieved using iterative back-fitting (see, e.g., Davison 2003). Good starting solutions were found to be essential for GP minimisation in particular. These were achieved by estimating local GP models at each of the m members of the index set (or combinations of neighbours thereof to increase sample size), then fitting smoothing spline models for each of GP shape ξ and scale σ.

3.5 Return values

The return value z_T of storm peak significant wave height corresponding to some return period T, expressed in years, can
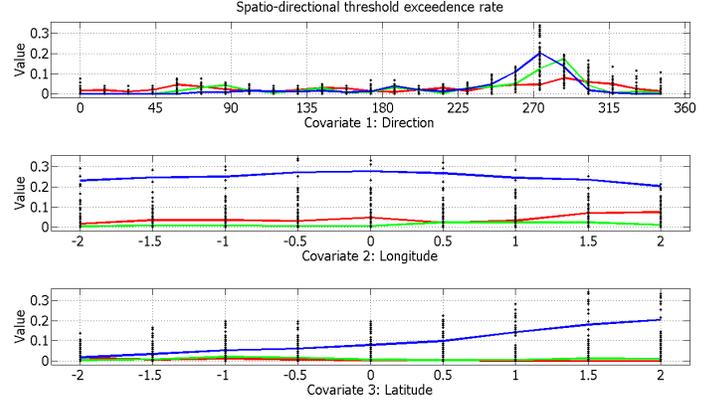
FIGURE 6. MARGINAL PLOTS OF RATE OF OCCURRENCE OF THRESHOLD EXCEEDANCE ρ AGAINST DIRECTION, REL-ATIVE LONGITUDE AND RELATIVE LATITUDE. THE TOP PLOT SHOWS THE RATE OF OCCURRENCE AS A FUNCTION OF DIRECTION FOR THE SW (RED), CENTRE (GREEN) AND NE (BLUE) RESPECTIVELY. THE MIDDLE PLOT SHOWS THE RATE OF OCCURRENCE AS A FUNCTION OF LONGITUDE FOR 0° AT THE MINIMUM LATITUDE (RED), FOR 120° FOR THE CENTRAL LATITUDE (GREEN) AND 270° FOR THE MAXIMUM LATITUDE (BLUE). THE BOTTOM PLOT SHOWS THE RATE OF OCCUR-RENCE AS A FUNCTION OF LATITUDE FOR 0° AT THE MINI-MUM LONGITUDE (RED), FOR 120° FOR THE CENTRAL LON-GITUDE (GREEN) AND 270° FOR THE MAXIMUM LONGITUDE (BLUE).

be evaluated in terms of estimates for model parameters ϕ, ρ, ξ and σ. For any choice of covariates θ, x, y, the return value is given by

$$z_T = \phi - \frac{\sigma}{\xi}(1 + \frac{1}{\rho}(\log(1 - \frac{1}{T}))^{-\xi})$$

where all of ϕ, ρ, ξ and σ are understood to be functions of θ, x, y, and ρ is expressed as an annual rate of threshold exceedance. Thus, z_{100} corresponds to the 100–year return value, often denoted by H_{S100}.

4 Application

We now fit the spatio-directional spline model for a 9 x 9 spatial grid of locations on the North West continental shelf of Australia. A basis matrix B of $25 \times 13 \times 13$ knot locations (direction, longitude and latitude) for 6156 observations of storm peak H_S.

4.1 Threshold

We first estimate a threshold function ϕ using quantile regression spline above which observations $\dot{z}$ of storm peak H_S are
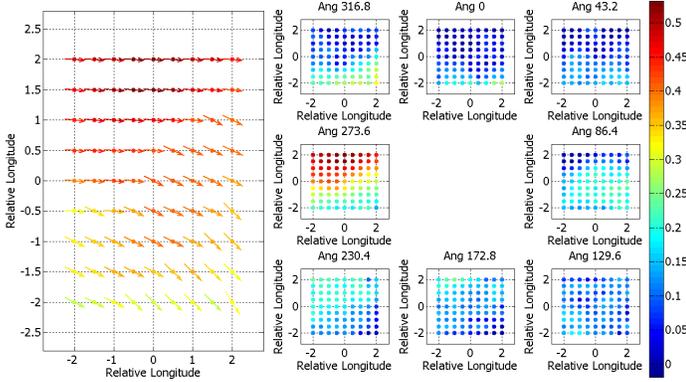
FIGURE 7. SPATIO-DIRECTIONAL PLOT FOR POISSON RATE OF OCCURRENCE ρ. THE 8 RIGHT-HAND PLOTS SHOW THE RATE OF OCCURRENCE ρ FOR EACH LOCATION FOR 8 DIRECTIONS. THE LEFT-HAND PLOT SHOWS THE DIRECTION FROM WHICH THE RATE OF OCCURRENCE ρ IS LARGEST.

FIGURE 8. MARGINAL PLOTS OF GP SHAPE ξ (LEFT-HAND PLOTS) AND GP SCALE, σ (RIGHT-HAND PLOTS) AGAINST DIRECTION, RELATIVE LONGITUDE AND RELATIVE LATITUDE. THE TOP PLOTS SHOW THE SHAPE AND SCALE AS A FUNCTION OF DIRECTION FOR THE SW (RED), CENTRE (GREEN) AND NE (BLUE) RESPECTIVELY. THE MIDDLE PLOTS SHOW THE RATE OF SHAPE AND SCALE AS A FUNCTION OF LONGITUDE FOR 0° AT THE MINIMUM LATITUDE (RED), FOR 120° FROM THE CENTRAL LATITUDE (GREEN) AND 270° FOR THE MAXIMUM LATITUDE (BLUE). THE BOTTOM PLOTS SHOW THE SHAPE AND SCALE AS A FUNCTION OF LATITUDE FOR 0° AT THE MINIMUM LONGITUDE (RED), FOR 120° FROM THE CENTRAL LONGITUDE (GREEN) AND 270° FOR THE MAXIMUM LONGITUDE (BLUE).

assumed to be extreme. A non-exceedance threshold probability of 0.5 was used to estimate ϕ. Figure 4 shows marginal plots of ϕ against direction, relative longitude and relative latitude. The highest thresholds can be seen from storms from the North and lowest threshold from storms from the West to South West which is consistent with the raw data. Figure 5 shows a spatio-directional plot for exceedance threshold, ϕ. The 8 right-hand plots show the 50% threshold values of ϕ for each location for 8 directions. Spatially the smallest thresholds occur in the South Eastern locations which is nearest land whereas larger thresholds are seen in the North and East further in more open ocean. The left-hand plot shows the direction from which ϕ is largest. The largest thresholds predominately come from either the North of South East.

4.2 Rate of occurrence of threshold exceedance

We next estimate the rate of occurrence ρ of threshold exceedance using a Poisson process model with Poisson rate $\rho(\stackrel{\triangle}{=} \rho(\theta, x, y))$. Figure 6 shows marginal plots of rate of occurrence threshold exceedances ρ against direction, relative longitude and relative latitude. The rate of occurrence is relatively similar for all longitudes and latitudes; however the rate of occurrence for events from the West (250-290°) is much higher. This is consistent with the directional histograms seen in figure 3. Similar effects can be seen in figure 7, a spatio-directional plot for rate of occurrence ρ.

4.3 Size of occurrence of threshold exceedance:

We estimate the size of occurrence of threshold exceedance using the GP model. Figure 8 shows marginal plots of GP shape ξ (left-hand plots) and GP scale, σ (right-hand plots) against

direction, relative longitude and relative latitude. The GP shape parameter is largest for events from between $225 - 280^0$ whereas the scale is higher for events from around $50 - 130^o$. Trends in longitude and latitude can also be seen. Figure 9 shows a spatio-directional plot for generalised Pareto shape of occurrence ξ. Generally the largest shape ξ is seen from the South West. Figure 10 shows a spatio-directional plot the generalised Pareto scale σ. The scale is largest for storms from the East.

4.4 Return value z_{100} estimation

100–year return values z_{100} are then estimated. Figure 12 shows marginal plots of 100 year return level, z_{100} of storm peak H_S against direction, relative longitude and relative latitude. Return values are generally higher further East and further South. Lower return values can be seen for storms from around 260^0 the highest return values are seen for events from the North to East. Figure 11 spatio-directional plot of 100–year return level z_{100} for storm peak H_S. Generally lower return values are seen in the South East corner nearest land. The highest return value in most locations are generally from the North to North East. Return values are lower for storms from the West.

6
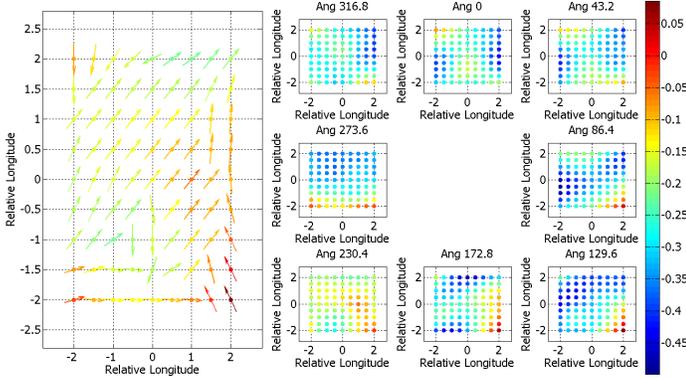
FIGURE 9. SPATIO-DIRECTIONAL PLOT FOR GENERALISED PARETO SHAPE OF OCCURRENCE ξ. THE 8 RIGHT-HAND PLOTS SHOW THE GENERALISED PARETO SHAPE ρ FOR EACH LOCATION FOR 8 DIRECTIONS. THE LEFT-HAND PLOT SHOWS THE DIRECTION FROM WHICH THE GENERALISED PARETO SHAPE ξ IS LARGEST.

FIGURE 11. SPATIO-DIRECTIONAL PLOT OF 100–YEAR RETURN LEVEL Z_{100} FOR STORM PEAK H_S. THE 8 RIGHT-HAND PLOTS SHOW THE 100–YEAR RETURN LEVEL Z_{100} FOR EACH LOCATION FOR 8 DIRECTIONS. THE LEFT-HAND PLOT SHOWS THE DIRECTION FROM WHICH THE 100–YEAR RETURN LEVEL Z_{100} IS LARGEST.
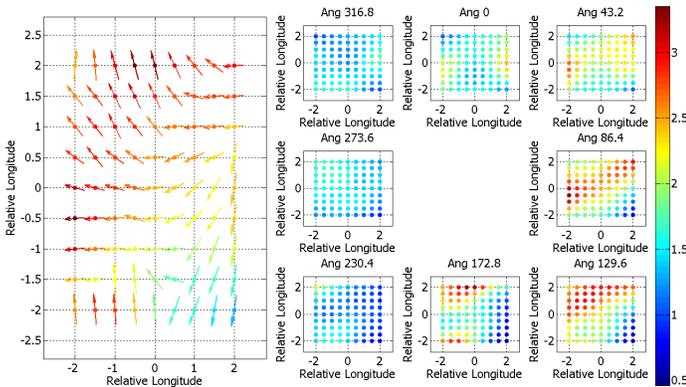
FIGURE 10. SPATIO-DIRECTIONAL PLOT THE GENERALISED PARETO SCALE σ. THE 8 RIGHT-HAND PLOTS SHOW THE GENERALISED PARETO SCALE σ FOR EACH LOCATION FOR 8 DIRECTIONS. THE LEFT-HAND PLOT SHOWS THE DIRECTION FROM WHICH THE GENERALISED PARETO SCALE σ IS LARGEST.

5 Discussion

In the paper, we introduce a marginal spatio directional model for extreme storm peak significant wave height, applied to estimation of spatio-directional design values for a neighbourhood of locations off the North West shelf of Australia. The model uses the peaks over threshold approach, incorporating estimation of an extreme value threshold and the rate and size of threshold exceedance. Model parameters are smooth spatio-directional functions. Cross-validation is used to estimate appropriate parameter smoothness in each case (results shown in appendix 6). Re-sampling techniques such as bootstrapping can be used to es-

timate the uncertainty of model parameters. The model yields parameter estimates and design values which are consistent with physical intuition and previous estimates.

The main advantage of the approach is that marginal spatial and directional variation of extreme value characteristics are incorporated in a rational manner eliminating the need for ad-hoc procedures such as site pooling. In isolating storm peak events, we also estimate the directional dissipation (see, e.g. Jonathan and Ewans 2007a) of storms across locations. This allows us also to estimate design criteria for arbitrary directional sectors for a given location together with the omni-directional estimate, in a consistent manner.

We are currently extending the model to incorporate multivariate spatial dependence, using composite likelihood methods, so that joint characteristics of extremes of storm peak significant wave height across multiple locations can also be estimated and studied.

6 Acknowledgements

REFERENCES

C.W. Anderson, D.J.T. Carter, and P.D. Cotton. *Wave Climate Variability and Impact on Offshore Design Extremes*. Report commissioned from the University of Sheffield and Satellite Observing Systems for Shell International, 2001.

E. Casson and S. G. Coles. Spatial regression models for extremes. *Extremes*, 1:449–468, 1999.

7   Copyright © 2013 by ASME
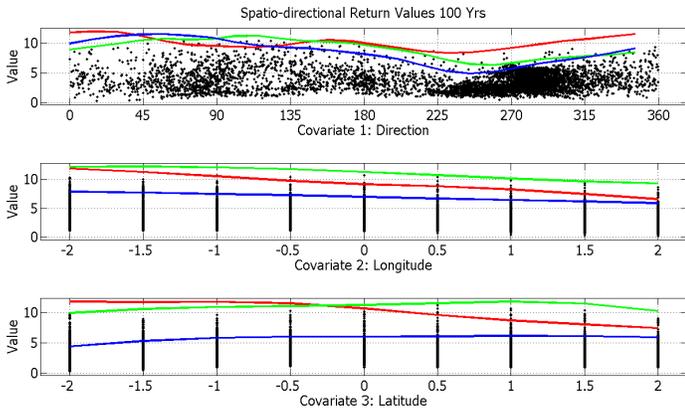
FIGURE 12. MARGINAL PLOTS OF 100–YEAR RETURN LEVEL Z_{100} OF STORM PEAK H_S AGAINST DIRECTION, REL-ATIVE LONGITUDE AND RELATIVE LATITUDE. THE TOP PLOT SHOWS THE 100–YEAR RETURN VALUE Z_{100} AS A FUNCTION OF DIRECTION FOR THE SW (RED), CENTRE (GREEN) AND NE (BLUE) RESPECTIVELY. THE MIDDLE PLOT SHOW THE 100–YEAR RETURN VALUE Z_{100} AS A FUNCTION OF LONGITUDE FOR 0° AT THE MINIMUM LATITUDE (RED), FOR 120° FROM THE CENTRAL LATITUDE (GREEN) AND 270° FOR THE MAXI-MUM LATITUDE (BLUE). THE BOTTOM PLOT SHOWS THE 100–YEAR RETURN VALUE Z_{100} AS A FUNCTION OF LATITUDE FOR 0° AT THE MINIMUM LONGITUDE (RED), FOR 120° FROM THE CENTRAL LONGITUDE (GREEN) AND 270° FOR THE MAXI-MUM LONGITUDE (BLUE).

V. Chavez-Demoulin and A.C. Davison. Generalized additive modelling of sample extremes. *J. Roy. Statist. Soc. Series C: Applied Statistics*, 54:207, 2005.

L. E. Chouinard, C. Liu, and C. K. Cooper. Model for severity of hurricanes in Gulf of Mexico. *J. Wtrwy., Harb., and Coast. Engrg.*, May/June:120–129, 1997.

S. Coles. *An introduction to statistical modelling of extreme values*. Springer, 2001.

S. G. Coles and E. Casson. Extreme value modelling of hurricane wind speeds. *Structural Safety*, 20:283–296, 1998.

S. G. Coles and E. A. Powell. Bayesian methods in extreme value modelling: a review and new developments. *International Statistics Review*, 64:119–136, 1996.

S. G. Coles and J. A. Tawn. A Bayesian analysis of extreme rainfall data. *Applied Statistics*, 45:463–478, 1996.

S. G. Coles and J. A. Tawn. Bayesian modelling of extreme sea surges on the uk east coast. *Phil. Trans. R. Soc. A*, 363:1387–1406, 2005.

A. C. Davison. *Statistical models*. Cambridge University Press, 2003.

A.C. Davison and R. L. Smith. Models for exceedances over high thresholds. *J. R. Statist. Soc. B*, 52:393, 1990.

J. M. Dixon, J. A. Tawn, and J. M. Vassie. Spatial modelling of extreme sea-levels. *Environmetrics*, 9:283–301, 1998.

E.F. Eastoe. *Statistical models for dependent and non-stationary extreme events*. Ph.D. Thesis, University of Lancaster, U.K., 2007.

E.F. Eastoe and J.A. Tawn. Modelling non-stationary extremes with application to surface level ozone. *Appl. Statist.*, 58:22–45, 2009.

P. H. C. Eilers and B. D. Marx. Flexible smoothing with b-splines and penalties. *Statistical Science*, 11(2):89–121, 1996.

P H C Eilers and B D Marx. Splines, knots and penalties. *Wiley Interscience Reviews: Computational Statistics*, 2:637–653, 2010.

K. C. Ewans and P. Jonathan. The effect of directionality on Northern North Sea extreme wave design criteria. *J. Offshore Mechanics Arctic Engineering*, 130:10, 2008.

C. Guedes-Soares and M. Scotto. Modelling uncertainty in long-term predictions of significant wave height. *Ocean Eng.*, 28: 329, 2001.

R. R. Haring and J. C. Heideman. Gulf of Mexico rare wave return periods. *Offshore Technology Conference, OTC3229*, 1978.

J. E. Heffernan and J. A. Tawn. A conditional approach for multivariate extreme values. *J. R. Statist. Soc. B*, 66:497, 2004.

P. Jonathan and K. Ewans. A spatiodirectional model for extreme waves in the Gulf of Mexico. *ASME J. Offshore Mech. Arct. Eng.*, 133:011601, 2011.

P. Jonathan and K. C. Ewans. The effect of directionality on extreme wave design criteria. *Ocean Eng.*, 34:1977–1994, 2007a.

P. Jonathan and K. C. Ewans. Uncertainties in extreme wave height estimates for hurricane dominated regions. *J. Offshore Mechanics Arctic Engineering*, 129:300–305, 2007b.

P. Jonathan and K. C. Ewans. On modelling seasonality of extreme waves. In *Proc. 27th International Conf. on Offshore Mechanics and Arctic Engineering, 4-8 June, Estoril, Portugal*, 2008.

P. Jonathan and K. C. Ewans. Statistical modelling of extreme ocean environments for marine design: a review. *Ocean Eng. (accepted January 2013, draft at www.lancs.ac.uk/~jonathan)*, 2013.

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

A. W. Ledford and J. A. Tawn. Modelling dependence within joint tail regions. *J. R. Statist. Soc. B*, 59:475–499, 1997.

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

M.G. Scotto and C. Guedes-Soares. Modelling the long-term time series of significant wave height with non-linear threshold models. *Coastal Eng.*, 40:313, 2000.

M.G. Scotto and C. Guedes-Soares. Bayesian inference for long-

term prediction of significant wave height. *Coastal Eng.*, 54: 393, 2007.

E. G. Ward, L. E. Borgman, and V. J. Cardone. Statistics of hurricane waves in the Gulf of Mexico. *Offshore Technology Conference*, 1978.

Appendix

This appendix illustrates results of cross-validation to select roughness coefficients in roughness-penalised quantile regression for threshold estimation, and estimation of rate and size of occurrence of threshold exceedance using roughness-penalised maximum likelihood. We select values of roughness coefficients λ_ϕ (quantile regression), λ_ρ (Poisson rate) and λ_ξ ($= \kappa\lambda_\sigma$), λ_σ (for generalised Pareto exceedance size) to maximise the predictive performance of the respective models, or minimise the predictive lack of fit (LOF). The latter is quantified using the unpenalised fit criterion (ℓ) for the relevant model within the cross-validation procedure. Plots of predictive lack of fit as a function of roughness coefficient are given below in figures 13, 14 and 15.

FIGURE 14. PREDICTIVE LACK OF FIT (LOF) AS A FUNCTION OF ROUGHNESS COEFFICIENT (λ_ρ) FOR POISSON EXCEEDANCE RATE MODELLING.
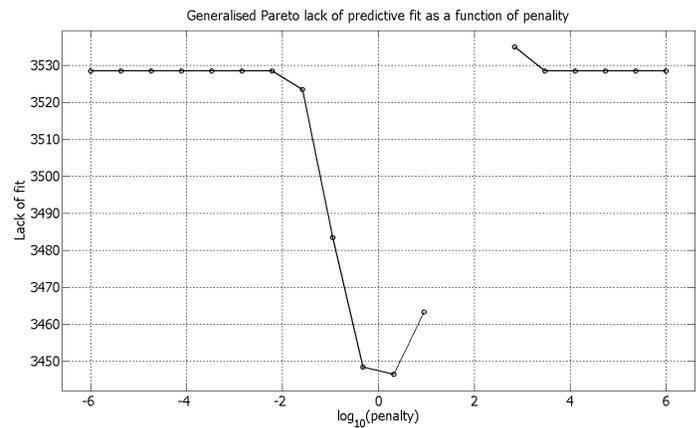
FIGURE 15. PREDICTIVE LACK OF FIT (LOF) AS A FUNCTION OF ROUGHNESS COEFFICIENT ($\lambda_\xi = \kappa\lambda_\sigma$ FOR PRE-SPECIFIED CONSTANT κ) FOR GENERALISED PARETO EXCEEDANCE SIZE MODELLING. MISSING VALUES IN THE PLOT INDICATE THAT FOR SOME VALUES OF ROUGHNESS COEFFICIENT, THE ESTIMATED GP MODEL (ESTIMATED ON A TRAINING SUBSET OF THE ORIGINAL SAMPLE) IS INCONSISTENT WITH THE REMAINDER OF THE SAMPLE, IN THAT SOME INDIVIDUALS ARE PREDICTED TO LAY BEYOND THE UPPER END POINT $\phi - \sigma/\xi$ OF THE CORRESPONDING GENERALISED PARETO DISTRIBUTION, FOR WHICH THE LOG LIKELIHOOD IS UNDEFINED. THE CORRESPONDING VALUES OF ROUGHNESS COEFFICIENT ARE THEREFORE REJECTED.

FIGURE 13. PREDICTIVE LACK OF FIT (LOF) AS A FUNCTION OF ROUGHNESS COEFFICIENT (λ_ϕ) FOR QUANTILE REGRESSION.