# A spatio-directional model for extreme waves in the Gulf of Mexico

**Philip Jonathan**
Shell Technology Centre Thornton, Chester, UK.
philip.jonathan@shell.com

**Kevin Ewans**
Shell International Exploration and Production, Rijswijk, NL.
kevin.ewans@shell.com

10 September 2009

# Motivation

- Marine structures must be safe, so estimation of extreme environments is important.
- Gap to fill between regulatory requirements, engineering practice and latest statistical approaches with respect to modelling of covariate effects (e.g. space, time, depth, direction) and dependent extremes.
- Statistics literature provides framework for consistent and rational estimation.

# Issues with oceanographic extreme value analysis

- Extreme value analysis is difficult.
  - Modelling the most *unusual* events in the sample.
  - The extremes of the sample are highly influential in model estimation.
  - Extrapolating beyond the domain of the sample.
  - Theory is asymptotic but the sample may not be.
- Extremes vary systematically with a number of covariates (including storm direction, season and location).
- Extremes at neighbouring locations are dependent. Large values at one location are more likely given large values at one or more of its neighbours.
- Extremes are correlated in time.
- Reliable estimation of extreme events requires incorporation of covariate effects and dependence.

# Approach to modelling fitting and quantile estimation

- Peaks over threshold modelled using generalised Pareto (GP).
- GP model parameters vary smoothly in space, using natural thin plate spline (NTPS) form.
- Data standardised (or *whitened*) w.r.t. storm direction to accommodate covariate variation.
- Arrival rate of threshold exceedences characterised using Poisson model.
- Poisson rate varies smoothly with direction, using Fourier form.
- Maximise likelihood, penalised by parameter roughness. Diagnostics for model fit. Cross-validation for optimal roughness. Bootstrapping for parameter uncertainty point-wise.
- Simulate to characterise extreme quantiles (e.g. $H_{S100}$).
- Slick algorithm for maximum likelihood GP fitting with NTPS using reparameterised GP.

# Our work

Method development driven by application requirements. Our recent contributions include:

- Combining dependent samples of extremes (Jonathan and Ewans 2007b).
- Covariate effects on extreme quantile estimates (Jonathan et al. 2008).
- Directional extremes (Jonathan and Ewans 2007a, Ewans and Jonathan 2008).
- Seasonal extremes (Jonathan and Ewans 2008).
- Spatial modelling (Jonathan and Ewans 2009).
- Joint modelling of wave spectral parameters for extreme conditions (Jonathan et al. 2009).

# Basic references

Large body of statistical and engineering literature on extremes. Important method articles for current work include:

- Davison and Smith 1990 (maximum likelihood formation; reparameterised GP).
- Heffernan and Tawn 2004 (conditional joint extremes).
- Chavez-Demoulin and Davison 2005 (penalised likelihood for extremes; NHPP; spline covariate form in 1-D).
- Eastoe and Tawn 2009 (non-stationary extremes).
- Ramsay 2002 (finite element L-splines).

Reference books:

- Davison 2003 .
- Green and Silverman 1994 .

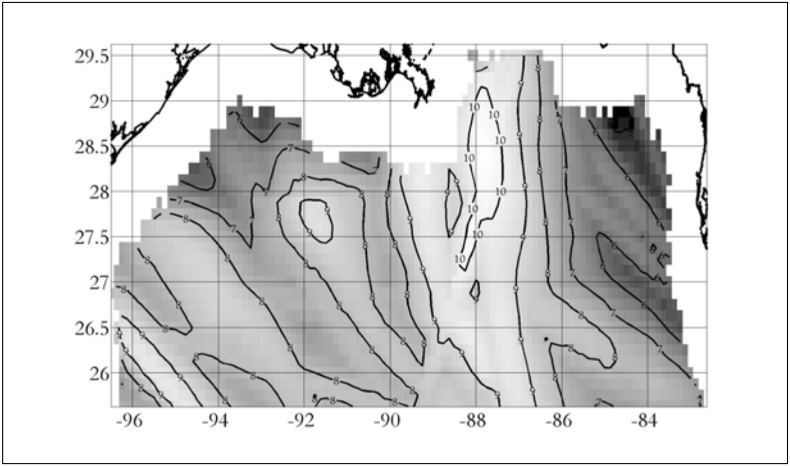
# Storm peak significant wave height data

- Significant wave height $H_S$ values from GOMOS Gulf of Mexico (GoM) hindcast study (Oceanweather, 2005), for September 1900 to September 2005 inclusive, at 30-minute intervals.
- $>$2500 locations on rectangular lattice with spacing of 0.125°.
- 315 storm periods identified (for whole GoM). For each grid point per storm, isolated storm peak significant wave height, $H_S^{sp}$, corresponding wave direction, $\theta$ and location.
- Coastal regions ignored.

Health warning:

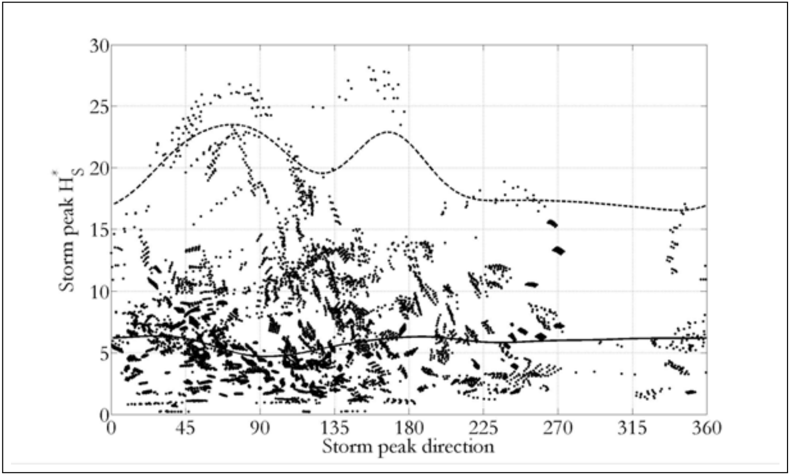
- Data are from a hindcast: simulator of meteorological - oceanographic physics, calibrated to observations of GoM hurricanes.
- Characteristics of sample change in time.
- Some graphs re-scaled for confidentiality.

# Observed maxima




- MATLAB contouring software
- Hurricane alleys (Chouinard et al. 1997)

# Variation with direction




$H_S^{sp*}$ with direction for a typical location (note re-scaled ordinate)

# Overview of modelling components

- Basics of generalised Pareto modelling.
- Penalised likelihood with Fourier covariate.
- Non-homogeneous Poisson process and Poisson arrivals with Fourier rate.
- Directional standardisation or whitening.
- GP modelling with univariate spline form.
- GP modelling with bivariate spline form.

# Generalised Pareto basics

$$\begin{aligned} P(X > x | X > u) &= \left(1 + \frac{\gamma}{\sigma}(x-u)\right)_+^{-\frac{1}{\gamma}}, \quad \gamma \neq 0 \\ &= \left(1 - \frac{y}{\sigma\alpha}\right)_+^{\alpha}, \quad \alpha = -\frac{1}{\gamma}, y = x - u \end{aligned}$$

Let $\alpha \uparrow \infty$, we get $e^{-\frac{y}{\sigma}}$. If $\gamma < 0$, then finite upper limit $u - \frac{\sigma}{\gamma}$.

$$P(X > x) = P(X > x | X > u) P(X > u)$$

Maximum likelihood estimates $\hat{\gamma}$ and $\hat{\sigma}$ are asymptotically correlated. We can reparameterise to $(\gamma, \nu = \sigma(1+\gamma))$ which are asymptotically independent. This facilitates a slick algorithm for bivariate spline GP models, and stabilises parameter estimation.

# Single Fourier covariate

Given $\{X_i\}_{i=1}^n$, $\{\theta_i\}_{i=1}^n$, distribution of storm peaks above variable threshold $u(\theta)$ assumed GP with cdf $F_{X_i|\theta_i,u}$:

$$\begin{aligned} F_{X_i|\theta_i,u}(x) &= P\left(X_i \le x|\theta_i, u(\theta_i)\right) \\ &= 1 - \left(1 + \frac{\gamma(\theta_i)}{\sigma(\theta_i)}\left(x - u(\theta_i)\right)\right)_+^{-\frac{1}{\gamma(\theta_i)}} \end{aligned}$$

$\gamma$ and $\sigma$ vary smoothly with direction, assumed to follow Fourier form:

$$\sum_{k=0}^{p}\sum_{b=1}^{2} A_{abk}\, t_b(k\theta)$$

where $t_1 = cos$ and $t_2 = sin$, $u$ estimated as a moving quantile with respect to $\theta$. Possible to fix one or more of $\gamma$, $\sigma$ and $u$ if judged appropriate.

# Single Fourier covariate: penalised likelihood

Penalised negative log likelihood is $l^*$:

$$l^* = \sum_{i=1}^{n} l_i + \lambda \left( R_\gamma + \frac{1}{w} R_\sigma \right)$$

Unpenalised negative log likelihood is:

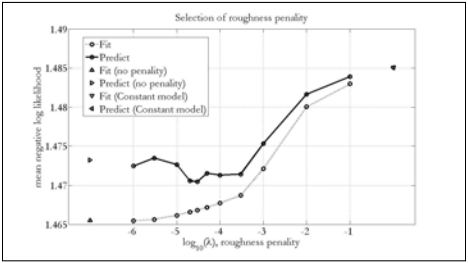
$$l_i = \log \sigma(\theta_i) + \left( \frac{1}{\gamma(\theta_i)} + 1 \right) \log \left( 1 + \frac{\gamma(\theta_i)}{\sigma(\theta_i)} (X_i - u(\theta_i)) \right)_+$$

Roughness of $\gamma$ is given by:

$$R_\gamma = \int_0^{2\pi} \left( \frac{\partial^2 \gamma}{\partial \theta^2} \right)^2 d\theta = \sum_{k=1}^{p} \pi k^4 \left( \sum_{b=1}^{2} A_{1bk}^2 \right)$$

Analogous expression for roughness of $\sigma$

# Single Fourier covariate: cross-validation and bootstrap



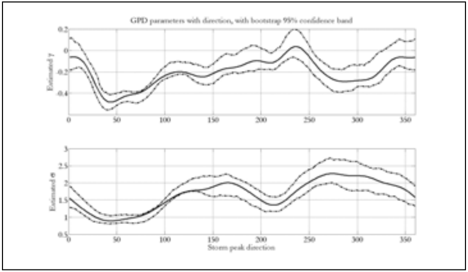




Illustration for directional covariate in Northern North Sea.

# Non-homogeneous Poisson process (NHPP) model

The negative log-likelihood written:

$$l(\rho, \gamma, \sigma) = l_N(\mu) + l_W(\gamma, \sigma)$$

where $l_N$ is the (negative) log-density of the total number of exceedances (with rate argument $\rho$), and $l_W$ is the (negative)log-conditional-density of exceedances given a known total number $N$). Inferences on $\rho$ made separately from those on $\gamma$ and $\sigma$.
The Poisson process log-likelihood, for arrivals at times $\{t_i\}_{i=1}^{n}$ in period $P_0$ is:

$$l_N(\rho) = -\left(\sum_{i=1}^{n} \log \rho(t_i) - \int_{P_0} \rho(t)dt\right)$$

# Non-homogeneous Poisson process (NHPP) model

Or approximately (Chavez-Demoulin and Davison 2005):

$$\hat{l}_N(\rho) = -\left(\sum_{j=1}^{m} c_j \log \rho(j\delta) - \delta \sum_{j=1}^{m} \rho(j\delta)\right)$$

where $\{c_j\}_{j=1}^{m}$ is the number of occurrences in each of the $m$ sub-intervals. We estimate storm occurrence rate adopting a Fourier form for Poisson intensity $\rho$, penalising its roughness $R_\rho$:

$$\hat{l}_N^*(\rho) = \hat{l}_N(\rho) + \kappa R_\rho$$

$R_\rho$ has form analogous to that of $R_\gamma$ or $R_\sigma$. Again use cross-validation to select $\kappa$ and (block) bootstrapping to quantify uncertainty.

# Form of $\rho$

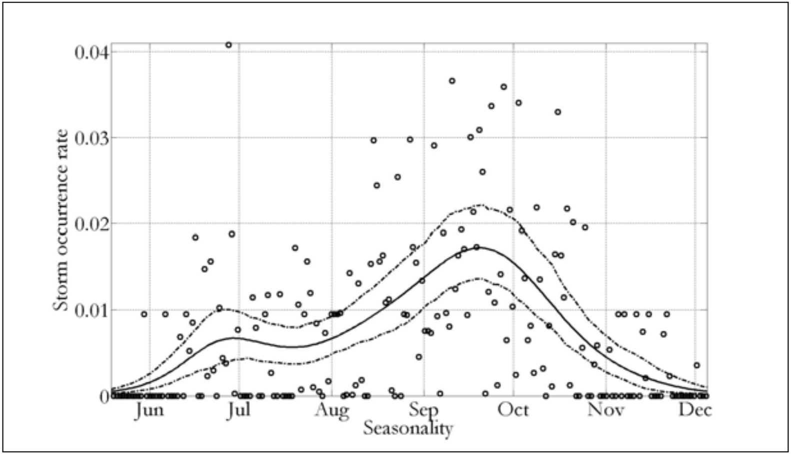



Illustration for seasonal covariate in Gulf of Mexico.

# A spatio-directional model : Directional standardisation + spatial model

Directional standardisation (see, e.g. Eastoe and Tawn 2009) might take the general form:

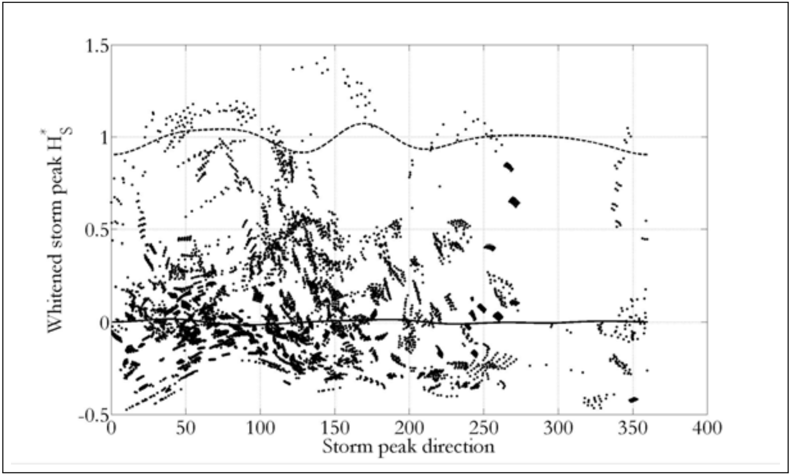
$$\frac{X_{ij}^{\beta(\theta_{ij})} - 1}{\beta(\theta_{ij})} = \mu(\theta_{ij}) + \eta(\theta_{ij}) W_{ij}$$

for storm $i$ at location $j$, where $\beta$, $\mu$ and $\eta$ are smooth functions of direction. Here we assume the simplified form:

$$W_{ij} = \frac{X_{ij} - \mu(\theta_{ij})}{\eta(\theta_{ij})}$$

- Standardisation removes directional *colour* from data and *whitens* it, and can be adopted for multiple covariates. Procedure is rather ad-hoc. We used local (wrt direction) median for $\mu$ and a local estimate of the difference between the 99%$^{\text{ile}}$ and the median for $\eta$.
- The standardised data are then modelled using NHPP with spatial covariate as explained below.

# Directionally standardised data



# GP model with univariate natural cubic spline form

Natural cubic spline (NCS):

- Sequence of cubic polynomial pieces on an interval joined together to form a continuous function,
- Continuous first and second derivatives,
- Zero second and third derivatives at ends of the interval.

$$f(r) \quad = \quad a_1 + a_2 r + \sum_{i=1}^{n} \delta_i (r - r_i)^3 \qquad s.t. \qquad \sum_{i=1}^{n} \delta_i = \sum_{i=1}^{n} \delta_i r_i = 0$$

Penalised (n.l.) likelihood $l^*$ for $\{x_i\}_{i=1}^n$ at *distinct* $\{r_i\}_{i=1}^n$ :

$$l^* \quad = \quad \sum_{i=1}^{n} l_i^*(\lambda_\gamma, \lambda_\nu) = \sum_{i=1}^{n} l_i(r_i) + \frac{\lambda_\gamma}{2} \int \gamma''^2(r) dr + \frac{\lambda_\nu}{2} \int \nu''^2(r) dr$$

- $l_i(r_i)$ is GP likelihood,
- $\{\gamma_i\}_{i=1}^n = \underline{\gamma}$ and $\{\nu_i\}_{i=1}^n = \underline{\nu}$ are spline coefficients to be estimated.

# GP model with univariate natural cubic spline form

Quadratic form for parameter roughness:

$$\int \gamma''^2(r)dr = \underline{\gamma}'\underline{K}\underline{\gamma}$$
$$\int \nu''^2(r)dr = \underline{\nu}'\underline{K}\underline{\nu}$$

- $\underline{K}$ is symmetric and easily computed.

Score equations to minimise $l^*$ :

$$\frac{\partial l}{\partial \gamma_i} - \lambda_\gamma \underline{K}\underline{\gamma} = 0$$
$$\frac{\partial l}{\partial \nu_i} - \lambda_\nu \underline{K}\underline{\nu} = 0$$

- Back-fitting based on Taylor expansion, similar to Newton-Raphson,
- Complexity reduced by adopting $(\gamma, \nu)$ parameterisation of GP, decoupling the system into separate schemes for $\underline{\gamma}$ and $\underline{\nu}$,
- Incidence matrix if multiple events at one or more locations.

# GP model with bivariate natural thin plate spline

Natural thin plate spline (NTPS):

- Function $f(\underline{r})$ of $\underline{r} = (r_{(1)}, r_{(2)}) \in \mathbb{R}^2$.

$$f(\underline{r}) = a_0 + a_1 r_{(1)} + a_2 r_{(2)} + \sum_{i=1}^{n} \delta_i \zeta(||\underline{r} - \underline{r}_i||) \quad s.t. \quad \sum_{i=1}^{n} \delta_i = \sum_{i=1}^{n} \delta_i \underline{r}_i = 0$$

Kernel:

$$\zeta(z) = \frac{1}{16\pi} z^2 \ln(z^2)$$

Roughness:

$$\begin{aligned} R(f) &= \int_{\mathbb{R}^2} \int \left( \frac{\partial^2 f}{\partial r_{(1)}^2} + \frac{\partial^2 f}{\partial r_{(1)} \partial r_{(2)}} + \frac{\partial^2 f}{\partial r_{(2)}^2} \right) dr_{(1)} dr_{(2)} = \underline{\delta}' \underline{E} \underline{\delta} \quad \textit{quadratic} \\ E_{ik} &= \zeta(||r_i - r_k||) \end{aligned}$$

- Note similarity of NTPS in 2-D and NCS in 1-D.

# GP model with bivariate natural thin plate spline

Roughness-penalised likelihood $l^*$ :

$$l^* = \sum_{i=1}^{n} l_i + \frac{\lambda_\gamma}{2} R_\gamma + \frac{\lambda_\nu}{2} R_\nu$$

- Minimising $l^*$ with respect to the four sets of parameters $\underline{a}_\gamma$, $\underline{d}_\gamma$, $\underline{a}_\nu$ and $\underline{d}_\nu$ using back-fitting.

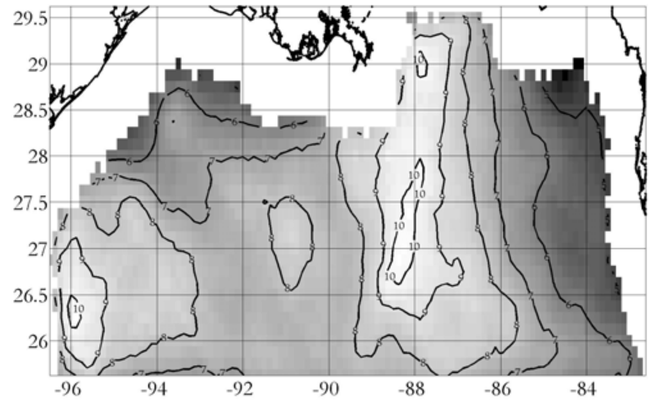
Issues:

- Integration over whole plane not domain of data.
- Threshold selection.
- NTPS is rotation-invariant, but $\zeta$ is not scale-invariant.

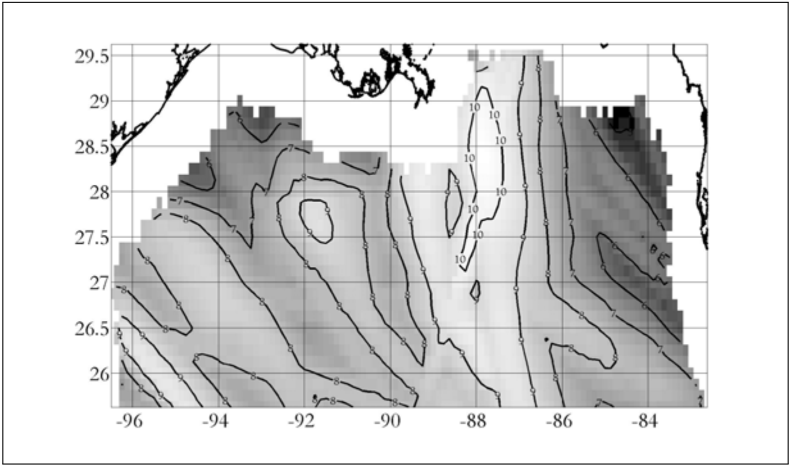
# Modelling procedure

1. At each location $j$, characterise variation of $\{X_i\}_{i=1}^n$ w.r.t. direction using standardisation. Whitened data $\{W_{ij}\}_{i=1,j=1}^{n,p}$ exhibit little directional variability in local *location* (e.g. the median value) and *spread* (e.g. a chosen inter-quantile range).
2. Select an appropriate threshold $u_j$ (typically a fixed quantile of the data per location) above which $\{W_{ij}\}_{i=1}^n$ exhibit a GP tail.
3. Use whitened data $\{W_{ij}\}_{i=1}^n$ to estimate the rate of occurrence $\rho_j(\theta)$ of exceedences of $u_j$, as a function of storm peak direction $\theta$, using a Poisson model.
4. For all whitened data at all locations, fit spatial GP model (with bivariate NTPS) to threshold exceedences.
5. Simulate from the fitted model to estimate extreme quantiles.

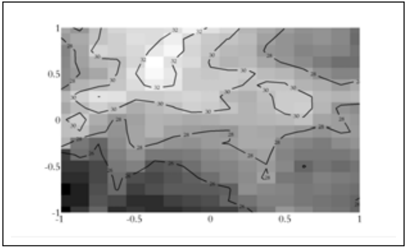
# Gulf-wide estimate for $H_{S100}$



# Observed maxima

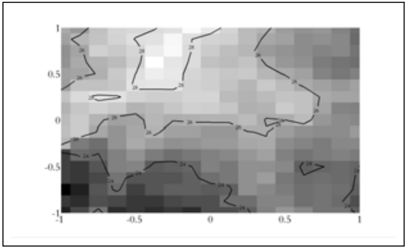


# $H_{S100}$ for NTPS model on 17 x 17 grid of locations

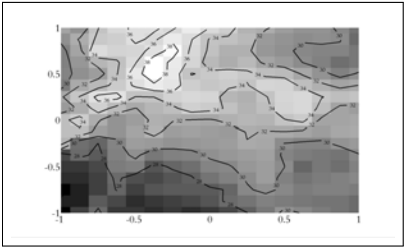


Median $H^*_{S100}$

Values of $H_S^{sp}$ have been re-scaled (to $H_S^*$) for reasons of confidentiality.

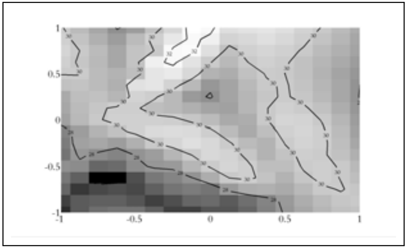


25%ile $H^*_{S100}$

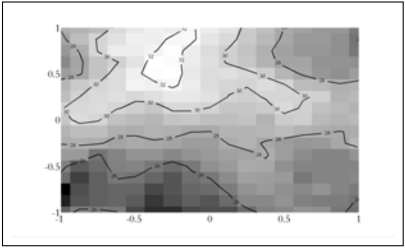


75%ile $H^*_{S100}$

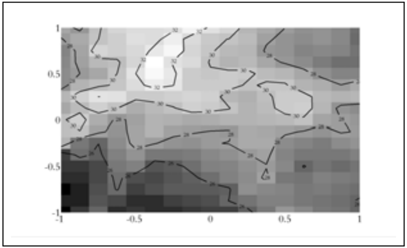
# Comparison of $H_{S100}$ for 17 x 17 grid of locations



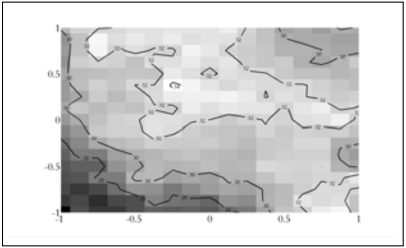
Observed maxima



Median $H^*_{S100}$, independent fits, whitened data



Median $H^*_{S100}$, NTPS, whitened data



Median $H^*_{S100}$, NTPS, original data

# Main findings

Pros:

- Rational, consistent approach.
- Accommodation of (multiple) covariate effects.
- Accommodation of spatial variation.
- Estimating spatial model is computationally faster than independent estimation over all locations.

Cons:

- Details of whitening step rather arbitrary, and hard to justify theoretically.
- Interpretation of GP fit to whitened data less intuitive.
- Sensitivity to more arbitrary choices (e.g. extreme value threshold, whitening parameters).

Other:

- Allowing threshold to vary w.r.t. covariates captures a considerable amount of the covariate effect.
- Solutions become quite large (simulations of $> 2500$ variates) and difficult to characterise concisely.

# Specific enhancements

- Incorporating uncertainties from model and threshold (mis-) specification in extreme quantile estimation.
- Develop improved rationale for parameter choices in whitening step.
- Consider variants of bivariate spline forms, in particular finite element L-splines (solution structure similar to NTPS but accommodates holes and concave regions in boundaries)

# General directions

- Realistic estimation of model uncertainty.
- Joint spatial modelling:
  - using conditional approach of Heffernan and Tawn [2004]
  - by modified likelihood to accommodate spatial dependence (e.g. Davison and Gholamrezaee 2009).
- Incorporation of multiple covariate effects (e.g. to accommodate direction, depth, temporal cycles and trends)
- Joint modelling of multiple variables (wind, waves, current, e.g. Heffernan and Tawn 2004), compare inferences with *response-based* approaches.
- Improved modelling of storm dissipation effects.
- Applications to controlled environments (e.g. wave basin experiments, where the physics is better understood and experiments repeatable).
- Promote use of appropriate statistical methods with regulators and influence marine design practice. Bridge academia and industry.

Thanks for listening.

philip.jonathan@shell.com

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.

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

A. C. Davison and M. M. Gholamrezaee. Geostatistics of extremes. Preprint. EPFL, stat.epfl.ch, 2009.

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

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

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.

P. J. Green and B.W. Silverman. *Nonparametric regression and generalised linear models: A roughness penalty approach*. Chapman and Hall, London, UK, 1994.

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. C. Ewans. The effect of directionality on extreme wave design criteria. *Ocean Engineering*, 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. A spatio-directional model for extreme waves in the Gulf of Mexico. *J. Offshore Mechanics Arctic Engineering*, 2009. Submitted February 2009.

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 Engineering*, 35: 1211–1225, 2008.

P. Jonathan, J. Flynn, and K. C. Ewans. Joint modelling of wave spectral parameters for extreme sea states. In *Proc. 11th International*

*Workshop on Wave Hindcasting and Forecasting, Halifax, Nova Scotia, Canada*, 2009.

T. Ramsay. Spline smoothing over difficult regions. *J. R. Statist. Soc. B*, 64:307–319, 2002.