Modeling Spatial Extreme Events using Markov Random Field Priors

Hang Yu*, Zheng Choo†, Justin Dauwels*, Philip Jonathan‡, and Qiao Zhou*
*School of Electrical and Electronics Engineering, †School of Physical and Mathematical Sciences
Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798
‡Shell Technology Centre Thornton, P.O. Box 1, Chester, UK

***Abstract*—A novel spatial model for extreme events is proposed. The model may for instance be used to describe the occurrence of catastrophic events such as earthquakes, floods, or hurricanes in certain regions; it may therefore be relevant for, e.g., weather forecasting, urban planning, and environmental assessment. The model is derived from the following ideas: The above-threshold values at each location are assumed to follow a generalized Pareto (GP) distribution. The GP parameters are coupled across space through Markov random fields, in particular, thin-membrane models. The latter are inferred through an empirical Bayes approach. Numerical results are presented for synthetic and real data (related to hurricanes in the Gulf of Mexico).**

I. Introduction

Statistical models can help to assess the likelihood of extreme events [1], such as earthquakes, floods, or hurricanes, and the dependency among the events (see, e.g., [1], [2], [3]). The models may serve as quantitative tools to assess the risks associated with certain infrastructures and facilities exposed to extreme conditions.

Extreme value theory provides a solid basis for analyzing extreme events [1]. The Pickands-Balkema-de Haan (PBdH) theorem, often called the second theorem in extreme value theory [4], states that for a large class of unknown underlying distribution functions F of a random variable X , the conditional excess distribution F_u, for a large threshold u, is well approximated by the generalized Pareto (GP) distribution.

The characteristics (e.g., GP parameters) of extreme events often vary systematically with a number of covariates. For example, the characteristics of extreme waves in hurricane dominated areas vary systematically with location, time, and storm direction [5], [6]. Those covariates need to be incorporated in extreme value models. In this paper, we propose to use graphical models to capture spatial covariate effects.

So far, spatial extreme models have only received limited attention. It is common practice to locally fit the GP parameters, without exploiting the spatial dependency, leading to inaccurate characterization of the extreme events. In the following, we briefly review the literature on spatial extreme models. In [2] a procedure is proposed to compute the pairwise spatial dependence of extreme events, i.e., the probability of threshold exceedance at one site conditioned on exceedance at one other site [3]. Alternatively, Naveau et al. [7] quantify pairwise spatial dependence through the concept of variogram. Both studies are limited to pairwise dependency. Recently, Northrop et al. [6] proposed a parametric model where the GP parameters depend on the location. The thresholds are determined through quantile regression, using Legendre polynomials. Our proposed model is similar in spirit, but is based on graphical models instead of polynomial regression.

In the model proposed here, the threshold exceedance at each location is modeled by the GP distribution. Thin-membrane models characterize the variation of the GP parameters across space. We then follow an empirical Bayes approach, inferring the smoothness parameters of the thin-membrane models using expectation maximization (EM).

Numerical results for synthetic and real data (related to hurricanes in the Gulf of Mexico) show that the proposed spatial extreme model can indeed capture spatial variations in the characteristics of extreme events. By inferring the smoothness parameters of the thin-membrane model, the smoothness is adjusted automatically in a suitable manner. In some numerical examples, the MRF-GP model yields location-independent GP shape and scale parameters, while the GP threshold is location dependent. Spatial variations in the characteristics of extreme events are then fully captured by only varying the threshold in space, resulting in a model with significantly fewer parameters.

The paper is organized as follows. In Section II, we briefly review the GP distribution and thin-membrane model, which are the main components of the proposed model. The model itself is explained in detail in Section III. Numerical results on synthetic and real datasets are briefly presented in Section IV. We offer concluding remarks in Section V.

II. Preliminaries

In this section, we give a short description of the generalized Pareto (GP) distribution and the thin-membrane model.

A. Generalized Pareto Distribution

We consider a random variable X with unknown distribution F. We are interested in inferring the conditional excess distribution function F_u. If the selected threshold u is high enough, for a large class of distributions F, the conditional excess distribution function F_u converges to a generalized Pareto (GP) distribution [4]:

$$F(x;u,\sigma,\gamma) = \begin{cases} 1-\left\{1+\frac{\gamma}{\sigma}(x-u)\right\}^{-\frac{1}{\gamma}}, & \gamma \neq 0 \\ 1-\exp\left(-\frac{x-u}{\sigma}\right), & \gamma = 0, \end{cases} \quad (1)$$

for $x \geq u$ and $1+\gamma/\sigma(x-u) \geq 0$, where $\gamma \in \mathbb{R}$ is the shape parameter and $\sigma > 0$ is the scale parameter.

If the random variable X has a GP distribution for a fixed threshold u, the conditional distribution of $X-t$, given $X \geq t$, corresponding to a higher threshold $u+t$, also has a GP distribution. The shape parameter γ_t of the conditional distribution remains unchanged, i.e., $\gamma_t = \gamma$, while the scale parameter σ_t is a linear function of the threshold, i.e., $\sigma_t = \sigma + \gamma t$.

B. Thin-Membrane Model

A Markov random field (MRF) or a graphical model is a collection of random variables indexed by the vertices of an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Each node $i \in \mathcal{V}$ is associated with a random variable X_i. An edge (i, j) is absent if the corresponding two variables X_i and X_j are conditional independent: $P(X_i, X_j | X_{\mathcal{V}|i,j}) = P(X_i|X_{\mathcal{V}|i,j})P(X_j|X_{\mathcal{V}|i,j})$, where $\mathcal{V}|i, j$ denotes all the variables except X_i and X_j. In particular, for Gaussian distributed X, the graph $\mathcal{G}$ is characterized by the inverse of the covariance matrix (precision matrix) K, i.e., $K(i, j) \neq 0$ if and only if the edge $(i, j) \in \mathcal{E}$ [8].

The thin-membrane model is a Gaussian MRF that is commonly used as smoothness prior. Such model tries to minimize the difference between neighbors, and its probability density function (pdf) can be written as:

$$\begin{aligned} P(X) &\propto \exp\{-\alpha \sum_{i \in V} \sum_{j \in \mathcal{N}(i)} (X_i - X_j)^2\} \\ &\propto \exp(-\alpha\, X^T K_p X), \end{aligned} \tag{2}$$

where $\mathcal{N}(i)$ denotes the neighboring nodes of node i, and α is the smoothness parameter. The matrix K_p is the adjacency matrix: its diagonal elements $[K_p]_{i,i}$ are equal to the number of neighbors of site i and its off-diagonal elements $[K_p]_{i,j}$ equal -1 if the sites i and j are adjacent and 0 otherwise.

III. Spatial Extreme Model

In this section, we introduce our novel spatial extreme model (denoted as MRF-GP model), which is based on the following two assumptions:

1) The threshold exceedance at each site follows a GP distribution (1) with threshold u_i, shape parameter γ_i, and scale parameter σ_i.
2) Spatial dependence is captured by the spatial-dependent parameters $u = (u_1, \cdots, u_p)$, $\gamma = (\gamma_1, \cdots, \gamma_p)$, $\sigma = (\sigma_1, \cdots, \sigma_p)$, where p is the number of sites. Specifically, the three parameter vectors u, γ, and σ each have a thin-membrane model as prior. Conditioned on the spatially dependent GP parameters, the extreme values at different sites are mutually independent.

Fig. 1 shows the factor graph of our model [9]. For simplicity, we depict the GP pdf of only one site on the grid (node indicated by "GP"). The three rectangular lattices on the right hand side represent the thin-membrane models; the nodes α_z represent the factors $\exp\{-\alpha_z(z_i - z_j)^2\}$, where z stands for either u, γ or σ.

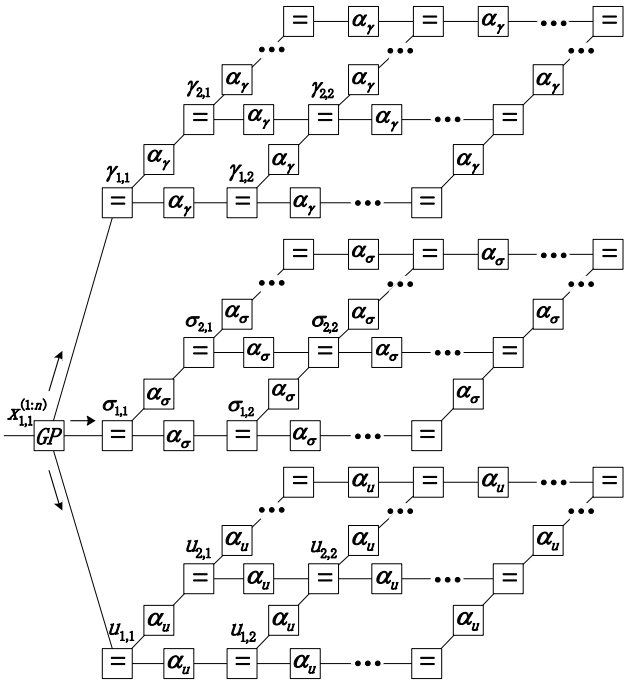


Fig. 1. Factor graph of MRF-GP model. The messages leaving the GP nodes towards the three thin-membrane subgraphs (indicated by the arrows) are approximated as Gaussian distributions [9].

Suppose that we have n samples $x_i^{(j)}$ at each of the p locations, where $i = 1, \cdots, p$ and $j = 1, \cdots, n$. Our objective is to infer the three parameters u, γ, and σ.

Let $y = (y_1, y_2, \ldots, y_p)$ denote the "observed" value of $z = (z_1, z_2, \ldots, z_p)$, where z is either u, γ or σ. In our approach, the "observed" y_i are computed at each site i from the n observed samples at that site. We model the observations as $y = z + b$, where $b \sim N(0, R_z)$ is a zero-mean Gaussian white noise with a diagonal covariance matrix R_z.

Since we assume that the prior distribution of z is a thin-membrane model (cf. (2)), the posterior distribution is given by:

$$\begin{aligned} P(z|y) &\propto \exp(-\alpha_z z^T K_p z) \exp\{-\frac{1}{2}(y-z)^T R_z^{-1}(y-z)\} \\ &\propto \exp\{-\frac{1}{2} z^T(\alpha_z K_p + R_z^{-1})z + z^T R_z^{-1} y\}. \end{aligned} \tag{3}$$

The maximum a posteriori estimate of z is then given by:

$$\hat{z} = \operatorname{argmax} P(z|y) = (\alpha_z K_p + R_z^{-1})^{-1} R_z^{-1} y. \tag{4}$$

In the following, we discuss how the "observed" value y, the noise covariance matrix R_z, and the smoothing parameter α_z are computed for each of the three parameters u, γ, and σ.

A. Local Observations of z

The thresholds $\hat{u}_i$ are selected so that the probability of threshold exceedance is identical (e.g., 5%) at all locations. The resulting thresholds are considered as local observations.

The local observed value of γ and σ is the vector of local maximum likelihood (ML) estimates ($\hat{\gamma_i}^{\mathrm{ML}}$, $\hat{\sigma_i}^{\mathrm{ML}}$) at each site, which can be obtained by numerically solving the following two coupled equations [11]:

$$\sum_{j=1}^{n} \frac{x_i^{(j)} - u_i}{\hat{\sigma_i}^{\mathrm{ML}} - \hat{\gamma_i}^{\mathrm{ML}}(x_i^{(j)} - u_i)} = \frac{n}{1 - \hat{\gamma_i}^{\mathrm{ML}}} \tag{5}$$

$$\sum_{j=1}^{n} \ln\left[1 - \frac{\hat{\gamma_i}^{\mathrm{ML}}(x_i^{(j)} - u_i)}{\hat{\sigma_i}^{\mathrm{ML}}}\right] = -n\hat{\gamma_i}^{\mathrm{ML}}. \tag{6}$$

From a message-passing perspective, the local estimates $\hat{\sigma_i}^{\mathrm{ML}}$ and $\hat{\gamma_i}^{\mathrm{ML}}$ are chosen as the means of the Gaussian messages leaving the GP node along the σ and γ edges respectively, towards the thin-membrane subgraphs associated with σ and γ respectively [9].

B. Covariance Matrices R_z

We use the bootstrap approach (as in [12]) to infer the noise covariance matrices R_u, R_γ and R_σ, and also the 95% confidence interval of the estimates of γ and σ as follows:

1) We generate m sample sets $S_1, \cdots, S_m$, each with size $n \times p$, by resampling at random with replacement from the original n observations (at each of the p locations).
2) The thresholds $u^k = (u_1^k, \cdots, u_p^k)$ are estimated by fixing the same quantile value (e.g., 5%) for each of the m subsets S_k, where $k = 1, \cdots, m$. With that choice of thresholds u^k, the parameters $\gamma^k = (\gamma_1^k, \cdots, \gamma_p^k)$, and $\sigma^k = (\sigma_1^k, \cdots, \sigma_p^k)$ are estimated using ML method (cf. (5)(6)) for each S_k.
3) The variance of $u_i^k (k = 1, \cdots, m)$ at site i is our estimate of $[R_u]_{i,i}$, with i = 1, ..., p. Similarly, we obtain estimates of the diagonal covariance matrices R_γ and R_σ.
4) The 95% confidence interval for γ and σ is estimated as the values corresponding to the 2.5% and 97.5% quantiles of γ_k and σ_k.

From a message-passing perspective, the diagonal elements of R_u, R_γ and R_σ are the covariances of the Gaussian messages leaving the GP node towards the thin-membrane subgraphs [9].

C. Smoothing Parameters α

The smoothness parameters α_z are hyperparameters in the overall model. Through EM we obtain point estimates of those hyperparameters, whereas we infer the posterior distributions of z. Such procedure corresponds to an empirical Bayes approach [10].

In the E-step, we compute [13]:

$$\begin{aligned} Q(\alpha_z, \hat{\alpha}_z^{(k-1)}) &= E_{Z|y,\hat{\alpha}_z^{(k-1)}}[\log P(y, Z|\alpha_z)] \\ &= -\frac{1}{2}\alpha_z\{\mathrm{trace}[K_p(\hat{\alpha}_z^{(k-1)}K_p + R_z^{-1})^{-1}] \\ &+ (\hat{z}^{(k-1)})^T K_p\, \hat{z}^{(k-1)}\} + \frac{1}{2}\log\det(\alpha_z K_p), \end{aligned} \tag{7}$$

where $\hat{z}^{(k)}$ is computed as in (4) with α_z is replaced by $\alpha_z^{(k)}$.

In the M-step, we select the value $\hat{\alpha}_z^{(k)}$ of α_z that maximizes $Q(\alpha_z, \hat{\alpha}_z^{(k-1)})$. A closed form expression of $\hat{\alpha}_z^{(k)}$ exists [13]:

$$\hat{\alpha}_z^{(k)} = \frac{p}{\mathrm{trace}[K_p(\hat{\alpha}_z^{(k-1)}K_p + R_z^{-1})^{-1}] + (\hat{z}^{(k-1)})^T K_p\, \hat{z}^{(k-1)}}, \tag{8}$$

where p is the number of sites. We iterate the E-step and M-step till convergence, yielding a local extremum of the marginal posterior of α_z.

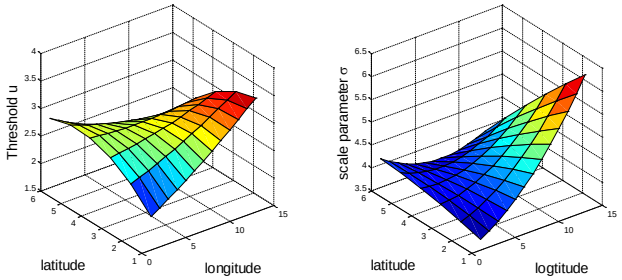
IV. Results

In this section, we apply the MRF-GP model to synthetic and real data. We compare it to a locally fit model, where the parameters $u = (u_1, \cdots, u_p)$, $\gamma = (\gamma_1, \cdots, \gamma_p)$, $\sigma = (\sigma_1, \cdots, \sigma_p)$ are all locally fit through ML estimation (cf. (5)(6)), without taking spatial priors into account. We compare the MRF-GP and the locally fit model based on three criteria:

1) We verify whether the shape parameter γ_i is independent of the threshold u_i, and the scale parameter σ_i depends linearly on u_i (cf. Section II-A).
2) We investigate how the shape and scale parameters depend on the threshold smoothness parameter α_v. Earlier studies suggest that a properly selected threshold surface (sometimes combined with scale surface) is sufficient to capture the spatial variation, and as a consequence, the shape surface may be flat [6].
3) We compute the 95% confidence interval of all estimates by bootstrapping.

A. Synthetic Data

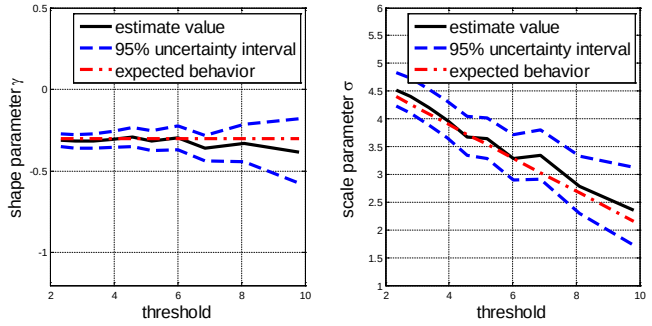
Here we present results for two case studies with synthetic data. Samples are drawn from GP marginals with location-dependent parameters. In both cases, the threshold surface is a quadratic Legendre polynomial, as shown in Fig. 2(a), whereas the shape and scale parameters are chosen differently in each case.



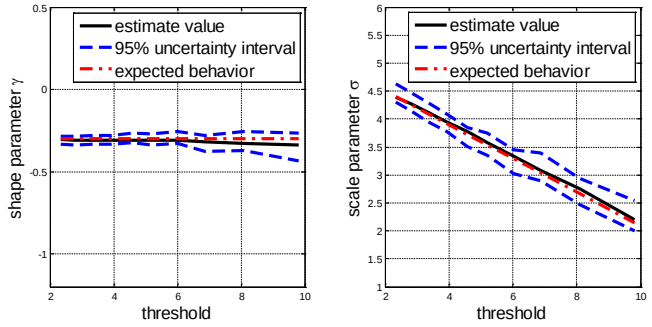
(a) (b)

Fig. 2. True GP parameters for synthetic data: (a) Threshold surface for case study 1 and 2; (b) Scale parameter surface for case study 2.

1) Case Study 1: The shape and scale parameters γ and σ respectively are chosen to be constant, and equal to -0.3 and 4.4 respectively. We generate 1250 samples from the GP distributions at each site. From Fig. 3, we can see that the estimates resulting from the MRF-GP model follow

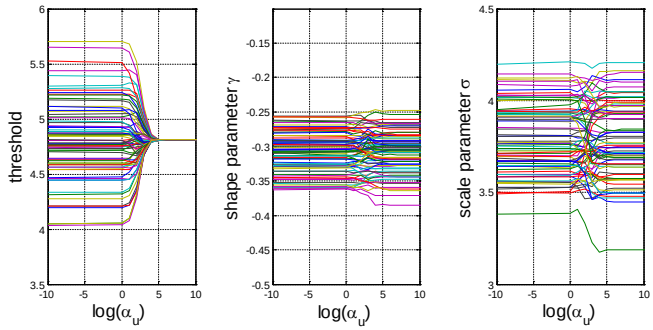


(a)

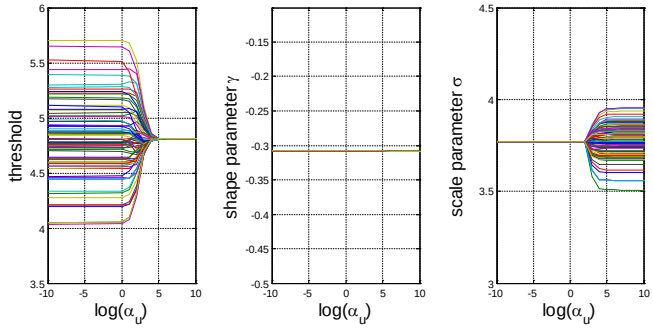


(b)

Fig. 3. Estimates of shape and scale parameters γ and σ as a function of threshold u, at one of the sites (Case study 1); the results at the other sites are similar. (a) Results for local fitting; (b) MRF-GP.



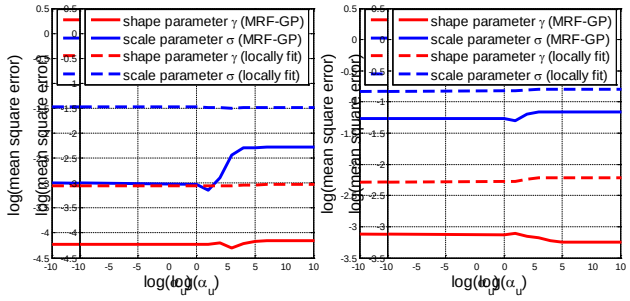
(a)



(b)

Fig. 4. Estimates of the thresholds u, and shape and scale parameters γ and σ as a function of threshold smoothness parameter α_u (Case study 1). The thresholds were chosen at each site to retain the 60% quantile. (a) Results for local fitting; (b) MRF-GP.

the predicted dependency on the threshold more closely: the shape parameter γ is nearly independent of the threshold u, whereas the scale parameter σ scales linearly with u (cf. Section II-A). Clearly, the local ML estimates fluctuate more, and hence are less reliable. Moreover, the MRF-GP estimates have narrower confidence intervals compared to the local ML estimates. Interestingly, Fig. 4 shows that for a large range of the threshold smoothing parameter α_u, the estimates of γ and σ do not depend on location, which is also the case for the true parameter values. On the other hand, the local ML estimates are significantly different at each site. Fig. 5(a) shows the mean square error (MSE) of the local and MRF-GP estimates, as a function of α_u: the MSE of the MRF-GP estimates is more than a order of magnitude smaller than the MSE of the local estimates.



(a) (b)

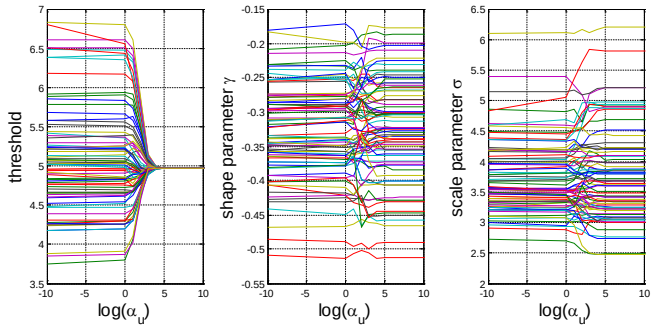
Fig. 5. Mean square error for ML and MRF-GP estimates of shape and scale parameters γ and σ as a function of threshold smoothness parameter α_u. The thresholds were chosen at each site to retain the 60% quantile. (a) Case study 1; (b) Case study 2.

2) Case Study 2: In the second case study, we extend the scale parameter surface to be quadratic, as shown in Fig. 2(b), however, the shape parameter γ remains the value -0.3 as in the previous case study. We wish to verify whether the MRF-GP model can capture the additional spatial dependence. The results are qualitatively similar to the ones of the first case study. The only difference is that the estimates of σ are no longer independent of location (Fig. 6), which is not surprising since the true parameters σ follow a quadratic surface. In other words, by inferring α_σ and α_γ, the spatial smoothness of σ and γ can automatically and appropriately be adjusted. The MRF-GP estimates are more accurate than the local ML estimates, as shown in Fig 5(b), although the improvement is less pronounced than in Case Study 1 (cf. Fig 5(a)).

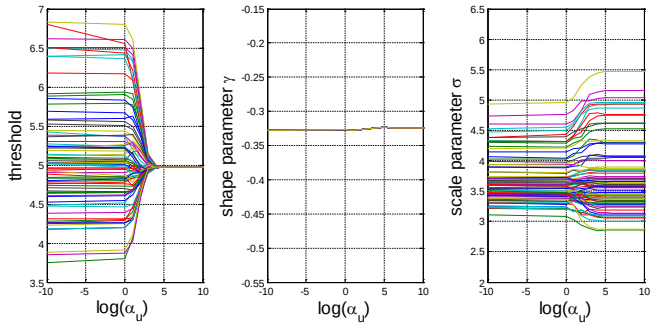
B. Real Data

We consider the GOMOS (Gulf of Mexico Oceanographic Study) data [14], which consists of 315 peak wave height values corresponding to hurricane events in the Gulf of Mexico. There are 78 sites arranged on a 6×13 rectangular lattice with spacing of 0.125° (approximately 14km).

The MRF-GP model yields more accurate GP parameter estimates than local fitting: The MRF-GP estimates of γ and σ again follow the predicted dependency on the threshold u more closely than the local estimates, and the 95% confidence intervals are narrower (not shown here). However, also for

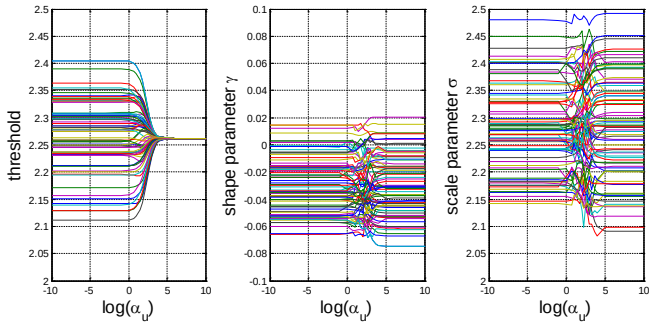


(a)

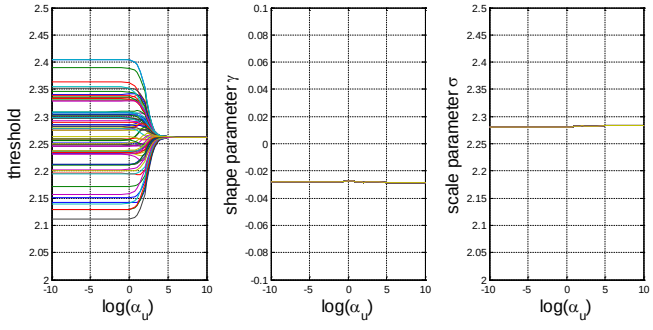


(b)

Fig. 6. Estimates of the thresholds u, and shape and scale parameters γ and σ as a function of threshold smoothness parameter α_u (Case study 2). The thresholds were chosen at each site to retain the 60% quantile. (a) Results for local fitting; (b) MRF-GP.



(a)



(b)

Fig. 7. Estimates of the thresholds u, and shape and scale parameters γ and σ as a function of threshold smoothness parameter α_u (GOMOS data). (a) Results for local fitting; (b) MRF-GP.

the MRF-GP, the parameter γ significantly depends on the threshold u, which is also an issue for the approach of [6]. Consequently, selecting a threshold quantile is a delicate issue for the data set at hand. Interestingly, Fig. 7 shows that the MRF-GP estimates of the parameters γ and σ are identical for each site. The variation of the GP marginals across space can be captured effectively by a spatial-dependent threshold u only. Consequently, the MRF-GP model has vastly fewer parameter than for local fitting (where all parameters are location-dependent), but achieves about the same loglikelihood (for a wide range of quantiles). Therefore, the MRF-GP model is preferred over local fitting for the GOMOS data.

V. Conclusion

We introduced a novel spatial extreme model: the marginal excess probabilities are assumed to be GP distributions, and thin-membrane models serve as priors for the location-dependent GP parameters. The smoothness of the GP parameters across space can automatically be inferred from the data. In some cases, certain GP parameters may be become location-independent.

Acknowledgment

Zheng Choo wishes to acknowledge the funding support for this project from Nanyang Technological University under the Undergraduate Research Experience on CAmpus (URECA) programme.

References

[1] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling extremal events for insurance and finance,* Berlin: Spring Verlag, 1997.

[2] J. E. Heffernan and J. A. Tawn, "A conditional approach for multivariate extreme value", *J.R. Statist. Soc.* B 66, pp.1–34, 2004.

[3] C. Keef, C. Svensson, and J. A. Tawn, "Spatial dependence in extreme river flows and precipitation for Great Britain," *Journal of Hydrology*, vol. 378, pp. 240–252, 2009.

[4] J. Pickands, "Statistical inference using extreme order statistics," *Annals of Statistics*, vol. 3, pp. 119–131, 1975.

[5] P. Jonathan, K. Ewans, and G. Forristall, "Statistical estimation of extreme ocean environments: the requirement for modelling directionality and other covariate effects," *Ocean Engineering*, vol. 35, pp. 1211–1225, 2008.

[6] P. J. Northrop and P. Jonathan, "Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights," *Environmetrics*, vol. 22, pp. 799–809, 2011.

[7] P. Naveau, A. Guillou, D. Cooley, and J. Diebolt, "Modelling pairwise dependence of maxima in space," *Biometrika*, vol. 96(1), pp. 1-17, 2009.

[8] J. M. F. Moura and N. Balram, "Recursive Structure of Noncausal Gauss Markov Random Fields," *IEEE Transactions on Information Theory*, IT-38(2):334-354, March 1992.

[9] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, Li Ping, and F. Kschischang, "The factor graph approach to model-based signal processing," *Proceedings of the IEEE* 95(6), pp. 1295–1322, 2007.

[10] G. Casella, "An Introduction to Empirical Bayes Data Analysis," *American Statistician* 39(2), pp. 83-87, 1985.

[11] J. R. M. Hosking and J. R. Wallis, "Parameter and Quantile Estimation for the Generalized Pareto Distribution," Technometrics, Vol. 29, No. 3, 1987.

[12] P. Jonathan and K. Ewans, "Uncertainties in extreme wave height estimates for hurricane-dominated regions," *Journal of offshore mechanics and arctic engineering*, vol. 129(4), pp. 300-305, 2007.

[13] M. J. Choi, V. Chandrasekaran, D. M. Malioutov, J. K. Johnson, and A. S. Willsky, "Multiscale stochastic modeling for tractable inference and data assimilation", Comput. Methods Appl. Mech. Engrg. Vol. 197, pp. 3492-3515, 2008.

[14] Oceanweather Inc. *GOMOS - USA Gulf of Mexico Oceanographic Study, Northen Gulf of Mexico Archive*, 2005.