

Fast Computation of Large Scale Marginal Extremes with Multi-Dimensional Covariates

Laks Raghupathi^{a,*}, David Randell^b, Kevin Ewans^c, Philip Jonathan^b

^aShell India Markets Pvt. Ltd., Bangalore, 560048, India.

^bShell Global Solutions (UK), Manchester M22 0RR, United Kingdom.

^cSarawak Shell Bhd., 50450 Kuala Lumpur, Malaysia.

Abstract

Safe and reliable design and operation of fixed and floating marine structures often located in remote and hostile environments is challenging. Rigorous extreme value analysis of meteorological and oceanographic data can greatly aid the design of such structures. Extreme value analysis is typically undertaken for single spatial locations or for small neighbourhoods; moreover, non-stationary effects of covariates on extreme values are typically accommodated in an ad-hoc manner. The objective of the work summarised here is to improve design practice by estimating environmental design conditions (such as return values for extreme waves, winds and currents) for a whole ocean basin, including additional covariate effects (such as storm direction) as necessary, in a consistent manner. Whole-basin non-stationary extreme value modelling is computationally complex, requiring inter-alia the estimation of tail functions, the parameters of which vary with respect to multi-dimensional covariates characterised by us using tensor products of penalised B-splines. We outline two technical contributions which make whole-basin non-stationary analysis feasible. Firstly, we adopt generalised linear array methods to reduce the computational burden of matrix manipulations. Secondly, using high-performance computing, we develop a parallel implementation of maximum likelihood estimation for the generalised Pareto distribution. Together, these innovations allow estimation of rigorous whole-basin extreme value models in reasonable time. We evaluate the new approach in application to marginal extreme value modelling of storm peak significant wave heights in two ocean basins, accommodating spatial and directional covariate effects.

Keywords: extremes, splines, non-stationary effects, fast-computation, high-performance computation

1. Introduction

1.1. Background

Safe exploration and production of oil and gas from offshore locations requires the design and construction of marine structures able to withstand severe environmental conditions, more extreme than the worst foreseeable during the required structural lifetime, associated with an annual probability of failure typically less than 1 in 10,000. Robust modelling of extreme environmental conditions is therefore critical for safe day-to-day operation and long-term structural reliability. This requires the specification of return values for oceanographic phenomena such as waves, winds, currents and associated variables. The metocean engineer faces the challenge of analysing huge samples of measured or hind-cast meteorological and oceanographic data exhibiting complex dependencies, including non-stationarity with respect to multiple covariates and spatial dependence of extremes, in order to estimate both marginal and joint return values for a large number of wave-, wind- and current-related variables.

Extreme value analysis is motivated by asymptotic arguments. In the metocean context, only the largest values in the sample (for example, observations exceeding some threshold) should be used for extreme value modelling, but the sample size must also be sufficient for good empirical modelling. Therefore, the amount of data available for analysis at a specific location is often limited. Reducing the threshold value runs the risk of invalidating asymptotic arguments underlying the modelling strategy. Increasing the threshold value reduces sample size still further. One pragmatic solution to the sample size issue is to aggregate observations from a neighbourhood of spatial locations and analyse the pooled sample. However, neighbouring locations in general have different extreme value characteristic, and observations from neighbouring locations are typically not statistically independent. Naive analysis of spatial extremes over spatial neighbourhoods can lead therefore to erroneous inferences. Fortunately, recent developments in statistical modelling of extreme ocean environments offer solutions other than pooling; a recent review is presented in [JE13]. [JER13] and [JRWE14] present recent progress in the spline-based modelling of covariates effect. Here, we further develop these methods for basin-wide marginal modelling of extremes, incorporating non-stationarity due to direction, longitude and

*Corresponding author. Email: l.raghupathi@shell.com

latitude, and adjusting for spatial dependence between locations in all uncertainty analyses. An outline of the modelling procedure is given at the start of §2.

At present, preliminary design conditions for a specific location are typically estimated by the metocean engineer based on data for (the neighbourhood of) that location. This process is typically time-consuming and technically challenging. Moreover, industry design procedures are sufficiently vague that there could be inconsistency between estimated design conditions for the same location made by different metocean engineers, using different data sources, modelling assumptions and procedures. With a basin-wide extreme value model in place, spatio-directional return values for all locations in the ocean basin of interest can be estimated. The estimates obtained provide outline directional design conditions for locations of interest in the basin based on rigorous, consistent statistical analysis, allowing fair comparison of environmental risk for those locations. The estimates might further be pre-computed, stored in a database and accessed via a graphical user interface to a device of choice querying the database. In practice, it may further be necessary to calibrate hindcast data to measurements at specific locations.

The paper is organised as follows. §2 presents the model formulation for extreme values subject to spatio-directional non-stationarity, outlining computational challenges to its implementation basin-wide. §3 describes generalised linear array methods and reports their performance by incorporation within the extreme value model. §4 describes parallel adaption of the generalised Pareto algorithm and quantifies computational improvements achieved. §5 describes the practical application of the above methodology to estimation of basin-wide return values for ocean storm severity in two ocean basins. Discussion, conclusions and opportunities for further development are summarised in §6.

2. A spatio-directional marginal extreme value model

The objective is the estimation of *marginal* return values for storm severity (quantified using significant wave height) for locations within a spatial neighbourhood, accounting for spatial and storm directional variability of extremal characteristics.

2.1. Model Components

Following the work of [JE08] and [JE11], summarised in [JRWE14], 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. At a given location, storm peak events are reasonably assumed to be statistically independent given covariates since they correspond to occurrences of independent atmospheric pressure fields. We assume that each storm event is observed at all locations within the neighbourhood under consideration. For a sample $\{z_i\}_{i=1}^n$ of n storm peak significant wave heights (henceforth H_S) observed at locations $\{x_i, y_i\}_{i=1}^n$ with dominant wave directions $\{\theta_i\}_{i=1}^n$ at storm peak H_S (henceforth “storm directions”), we proceed using the peaks-over-threshold approach as follows.

We first estimate a *threshold* function ϕ above which observations 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. 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. We next estimate the *rate* of occurrence ρ of threshold exceedance using a Poisson process model with Poisson rate $\rho \triangleq \rho(\theta, x, y)$. Finally we estimate the *size* of occurrence of threshold exceedance using a generalised Pareto (henceforth GP) model; the GP probability density function is given in the Appendix. The GP shape and scale parameters ξ and σ are also assumed to vary smoothly as functions of covariates, with ξ real and $\sigma > 0$. Positivity of GP scale is ensured throughout in the optimisation scheme. The GP shape parameter is unrestricted in the full optimisation, but limited to the interval $(-0.5, +0.2)$ in the estimation of the GP starting solution.

This approach to extreme value modelling follows that of [CDD05] and is equivalent to direct estimation of a non-homogeneous Poisson point process model ([DTV98], [JE13]). Appendix A provides an outline of the corresponding penalised likelihood optimisations required to estimate model parameters. We emphasise that, in common with [CDD05] and [JE13], we perform marginal non-stationary extreme value analysis across a grid of (dependent) spatial locations, accounting (per location) for directional and spatial variability in extremal characteristics. We further account for the effects of (extremal) spatial dependence between locations on inferences using a block bootstrapping scheme.

2.2. Parameterising covariates

Applying the model to whole-basin applications is methodologically straightforward, but computationally challenging. Physical considerations suggest we should consider parameters ϕ, ρ, ξ and σ to be smooth functions of covariates θ, x, y [RYJE13]. For estimation, this can be achieved by expressing the value of each parameter in terms of a linear combination of appropriate basis for the domain D of covariates, where $D = D_\theta \times D_x \times D_y$. Here $D_\theta = [0, 360)$ 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 covariate (and marginal domain) in turn, we first calculate a B-spline basis matrix for an index set of size m ($m \ll n$) of covariate values; potentially we could calculate the basis matrix for each of the n observations, but usually avoid this for computational and statistical efficiency. For instance in the case of D_θ , we calculate an $m_\theta \times p_\theta$

basis matrix B_θ such that the value of any function at each of the m_θ points in the index set for storm direction can be expressed as linear combination $B_\theta\beta_\theta$ for some $p_\theta \times 1$ vector β_θ of basis coefficients. Note that periodic marginal bases can be specified if appropriate (e.g. for D_θ).

Then we define a basis matrix for the three-dimensional domain D using tensor products of marginal basis matrices. Thus

$$B = B_y \otimes B_x \otimes B_\theta \quad (1)$$

provides an $m \times p$ basis matrix (where $m = m_\theta m_x m_y$, and $p = p_\theta p_x p_y$) for modelling each of ϕ, ρ, ξ and σ on the corresponding ‘‘spatio-directional’’ index set of size m . Any of ϕ, ρ, ξ and σ (η , say, for brevity) can then be expressed in the form $\eta = B\beta$ for some $p \times 1$ vector β of basis coefficients. Model estimation therefore reduces to estimating appropriate sets of basis coefficients for each of ϕ, ρ, ξ and σ . The value of any marginal p_i (i.e. p_θ, p_x , or p_y) is equal to the number q_i of spline knots specified for periodic domains (e.g. D_θ), and to $q_i + d_i$ for aperiodic domains (e.g. D_x and D_y), where d_i is the order of the B-spline function specified (always 3 in this work, so that spline functions are twice differentiable).

The roughness R of any function η defined on the support of the spline basis can be easily evaluated on the index set (at which $\eta = B\beta$). For a one-dimensional (e.g. directional) spline basis, following the approach of [EM10], writing the vector of differences of consecutive values of β as $\Delta\beta$, and vectors of second and higher order differences using $\Delta^k\beta = \Delta(\Delta^{k-1}\beta)$, $k = 2, 3, \dots$, the roughness R of η is given by

$$R = \beta' P \beta \quad (2)$$

where $P = (\Delta^k)'(\Delta^k)$ for differences of order k (with appropriate modifications to preserve periodicity as necessary). For a spatio-directional spline basis, the penalty matrix P can be similarly defined using

$$P = P_y \otimes P_x \otimes P_\theta \quad (3)$$

in the obvious notation. We use $k = 1$ throughout this work. With this choice of k , heavy roughness penalisation results in stationarity of parameters with respect to periodic and aperiodic covariates. To avoid potential over-fitting, parameter estimation requires that the (negative log) likelihood function is penalised (see the appendix) using a linear combination of parameter roughnesses with roughness coefficients λ_η (for each model parameter η). The choice of optimal penalty coefficients is determined by block cross-validation, further adding to computational complexity. Since each storm is observed as a spatially-dependent event at all spatial locations, we define a storm ‘‘block’’ to be the set of occurrences of a particular storm at all locations for the purposes of both cross-validation (in estimation of model smoothness) and bootstrapping (for uncertainty quantification). For the spatio-directional problem, it is also possible to vary the relative size of basis penalisation with respect to direction, longitude and latitude. In this work, the same penalty value was adopted for both spatial co-ordinates, and the relative value of the directional to spatial penalty was decided using a series of experiments with different penalty choices.

2.3. Managing computational burden

We manage the considerable computational challenges of basin-wide extreme value modelling using the following techniques:

- Generalised linear array model (GLAM) framework ([CDE06]) for fast computation of matrix inner products and matrix vector products (described in §3)
- Parallel adaption of the generalised Pareto optimisation loop to exploit modern parallel computing environment (described in §4)

We outline these developments and the potential computational gains in the next two sections.

3. Generalised linear array model

As an illustration of the complexity of explicit matrix operations necessary for spatio-directional inference, consider the back-fitting iteration ([Dav03], referred to in equations A.32 and A.33 of the appendix) for estimation of generalised Pareto (and Poisson) spline basis coefficients and parameter estimates. Estimation involves repeated solution (for β) of equations of the form

$$(B'H_0B + \lambda P)\beta = B'H_0u_0 \quad (4)$$

where B is a tensor product of marginal spline bases $B_d \otimes \dots \otimes B_1$, and B_i is the $m_i \times p_i$ basis matrix for the i th marginal covariate (θ, x and y respectively for $i=1,2,3$ in the spatio-directional problem). Computation of the inner product $B'_{p \times m} H_{0m \times m} B_{m \times p}$ results in a $p \times p$ matrix for inversion. Since non-sparse weight matrix H_0 and working variable u_0 are functions of the starting value β_0 of β at the current iteration, the inversion must be repeated a number of times (especially in the case of generalised Pareto, as will be illustrated below) before convergence. This computational challenge

increases with increasing number $m (= \prod m_i)$ of covariate bins and total number $p (= \prod p_i)$ of spline coefficients. For a basin-wide spatio-directional model on a 1000 x 1000 spatial grid compared to a directional model for a single location, the value of m would increase by a factor of one million, and the value of p by a factor of 10,000. Block cross-validatory choice of λ , and bootstrap uncertainty analysis, further increases the number of inversions required. Explicit computation of inner products and their manipulation would be very demanding; but fortunately, due to generalised linear array model (GLAM) methodology, these computations are not necessary. Estimation requires computation of linear forms $B\beta$ and $B'H_0u_0$, and inner products of the form $B'H_0B$. These and similar matrix operations on d -dimensions are possible *without* explicitly evaluating full linear forms and inner products, dramatically reducing the number of arithmetic operations and memory requirements.

Let us illustrate this with a specific 2D case. The linear form $B\beta$ can be written

$$(B_2 \otimes B_1) \underset{p_1 p_2 \times 1}{\beta} \equiv B_1 \underset{p_1 \times p_2}{\mathcal{M}(\beta)} B_2' \quad (5)$$

where the $p_1 \times p_2$ matrix $\mathcal{M}(\beta)$ is the $p_1 p_2 \times 1$ vector β arranged in matrix form, and “ \equiv ” denotes this equivalence. Computationally, the right hand side of (5) provides a major advantage: it avoids explicit computation of B . The number of multiplications to evaluate it is very much smaller than that needed to evaluate the left-hand side: if B_1 and B_2 are both $10^3 \times 10^2$ then the left-hand side requires 10^{10} multiplications whereas the right-hand side requires just over 10^8 multiplications. In general, for $m_1 = m_2 = m$ and $p_1 = p_2 = p$, evaluation of the left hand expression involves $m^2 p^2$ multiplications, compared with $(m + p)mp$ for the right hand expression. Similarly, the computational complexity of the inner product operation is greatly reduced when the operation is expressed in terms of the row tensor operator $G(B)$, namely the tensor product of the i^{th} row of B with itself

$$G(B) = (B \otimes I') * (I' \otimes B). \quad (6)$$

Here, 1 is a column vector of appropriate length, each element of which is unity. The inner product $B'H_0B$ becomes

$$(B_2 \otimes B_1)' \underset{m_1 m_2 \times m_1 m_2}{H_0} (B_2 \otimes B_1) \equiv G(B_1)' \underset{m_1^2 \times m_2^2}{\mathcal{M}(H_0)} G(B_2) \quad (7)$$

where $m_1^2 \times m_2^2$ matrix $\mathcal{M}(H_0)$ is a rearranged form of the $m_1 m_2 \times m_1 m_2$ matrix H_0 . Again, the right hand side provides marked improvement in computational efficiency.

3.1. Improved computational efficiency

We now present the results of applying the GLAM approach for basin-wide extreme value analysis. We compare execution times and peak memory usage for an explicit non-GLAM (“baseline”) analysis with a GLAM implementation as a function of the number of covariate bins m , for relatively small problems only. We estimate the spatio-directional model for a fixed number of spatial and directional knots, (specifically for fixed $p = p_x p_y p_\theta = 13 \times 13 \times 32 = 5408$), for successively larger spatial domains, resulting in increasing sample sizes and number of covariate bins. For example, a model for 32 directional bins on a 10×10 spatial grid corresponds to 3200 covariate bins, and for 32 directional bins on a 20×20 spatial grid 12,800 bins. All experiments were run on an 8 core Intel Xeon 2.5GHz with 35 GB RAM. Results are illustrated in Figure 1. Identical inferences were obtained from both the baseline and GLAM analysis, but both run time and peak memory usage are considerably lower for the GLAM implementation. For larger problems, we are unable to execute the baseline analysis.

4. Parallel generalised Pareto algorithm

Here we present a parallel adaptation of maximum likelihood estimation for the non-stationary generalised Pareto likelihood, exploiting a multi-processor, multi-node distributed computing architecture. Using simulation we quantify the reduction in execution time achieved compared with sequential execution, and propose a strategy to optimise architecture given problem characteristics. Finally we incorporate bootstrap uncertainty analysis, exploiting the natural massive parallelism of the bootstrapping procedure.

4.1. Parallel Adaption of GP

The computational efficiency of generalised Pareto (GP) optimisation, using the back-fitting algorithm, has already been markedly improved using GLAM methods, as described in §3. However, in a trial modelling study involving 7 analyses employing between 3,000 and 384,000 covariate bins, estimation of GP parameters accounted for between 65% and 90% of overall computation time, making it an obvious candidate for further improvement in computational efficiency.

Algorithm 1 illustrates the naive approach to estimating GP shape and scale parameters, each of which is modelled as a linear combination of spline basis functions. In order to identify optimal spline roughness penalty coefficients, cross-validation ([Sto74]) is used; this involves repeated optimisation for n_{CV} pre-specified subsets of the original sample and

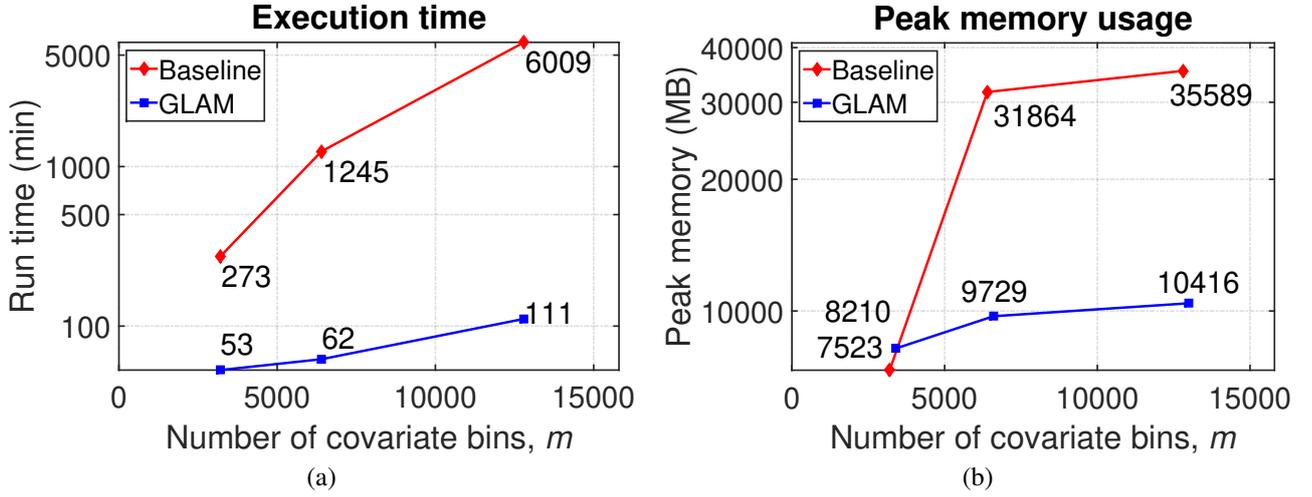


Figure 1: (a) Execution time for estimation with explicit (non-GLAM) “baseline” (red) and GLAM (blue) computation as a function of number of covariate combinations m on log scale; (b) Analogous comparison of peak memory usage.

```

input : Input data  $\text{dat}$ , number of penalties  $n_\lambda$ , number of cross-validations  $n_{CV}$ 
output: Optimal value of  $\lambda$  parameter

//loop over penalty coefficients;
for  $i_\lambda \leftarrow 1$  to  $n_\lambda$  do
  //loop over cross-validations;
  for  $i_{CV} \leftarrow 1$  to  $n_{CV}$  do
    //Divide the data into fit and test;
     $\text{datfit} \leftarrow \text{SplitDat}(\text{dat} = i_{CV})$ ;
     $\text{dattst} \leftarrow \text{SplitDat}(\text{dat} \neq i_{CV})$ ;
     $\sigma, \xi \leftarrow \text{GPFit}(\text{datfit}, i_\lambda)$ ;
     $\text{LOF}_{i_{CV}} \leftarrow \text{ComputeLOF}(\text{dattst}, \sigma, \xi)$ ;
  end
   $\text{LOF}_{i_\lambda} \leftarrow \sum_1^{n_{CV}}(\text{LOF}_{i_{CV}})$ ;
end
 $\text{LOF} \leftarrow \min(\text{LOF}_{i_\lambda})$ ;
 $\lambda \leftarrow \text{ComputeOptLmb}(\text{LOF})$ ;

```

Algorithm 1: Serial algorithm for GP fitting with nested loops over roughness penalty and cross-validations.

```

input : Input data  $\text{dat}$ , number of penalties  $n_\lambda$ , number of cross-validations  $n_{CV}$ 
output: Optimal value of  $\lambda$  parameter

//loop over both penalty coefficients and cross-validation sets;
for  $i_{\text{Loop}} \leftarrow 1$  to  $n_\lambda \times n_{CV}$  do
  //Obtain penalty and cross-validation set membership;
   $i_\lambda, i_{CV} = \text{LookUpTable}(i_{\text{Loop}})$ ;
  //Divide the data into fit and test;
   $\text{datfit} \leftarrow \text{SplitDat}(\text{dat} = i_{CV})$ ;
   $\text{dattst} \leftarrow \text{SplitDat}(\text{dat} \neq i_{CV})$ ;
   $\sigma, \xi \leftarrow \text{GPFit}(\text{datfit}, i_\lambda)$ ;
   $\text{LOF}_{i_{CV}, i_\lambda} \leftarrow \text{ComputeLOF}(\text{dattst}, \sigma, \xi)$ ;
end
for  $i_\lambda \leftarrow 1$  to  $n_\lambda$  do
   $\text{LOF}_{i_\lambda} \leftarrow \sum_1^{n_{CV}}(\text{LOF}_{i_{CV}, i_\lambda})$ ;
end
 $\text{LOF} \leftarrow \min(\text{LOF}_{i_\lambda})$ ;
 $\lambda \leftarrow \text{ComputeOptLmb}(\text{LOF})$ ;

```

Algorithm 2: Parallel adaption for GP fitting with single loop.

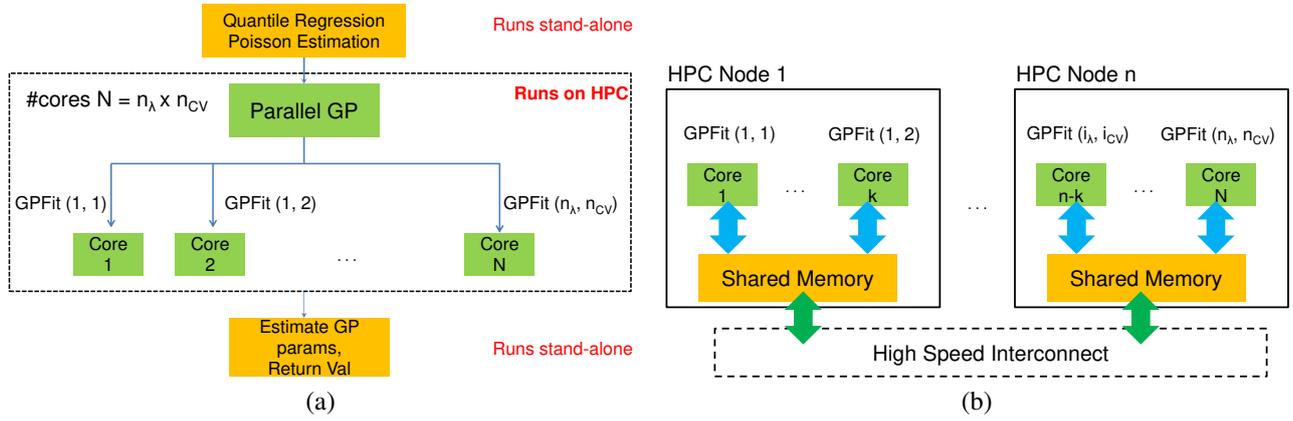


Figure 2: (a) Overall modelling work-flow showing parallel GP tasks. (b) Deployment of the parallel adaptation of the generalised Pareto algorithm in a distributed memory high-performance computing environment.

n_λ pre-specified values for spline roughness penalty. The optimal penalty is that which minimises the predictive lack of fit (LOF) of the GP model. There is clear scope to *unroll* this double loop over n_{CV} and n_λ , allocating tasks across multiple computing units.

Algorithm 2 illustrates the proposed parallel GP optimisation of effectively independent tasks over multiple computing cores. The double loop of algorithm 1 has been unrolled into a single loop using a pre-computed look-up table of combinations of roughness penalty and boolean memberships (for partitioning the sample into training and test subsets) for each cross-validation. This is a classic case of *task parallelism* where essentially the same computational task is performed in parallel on slightly different sets of data. The algorithm also illustrates the merging of results from different tasks so that overall predictive LOF can be evaluated.

Figure 2 (a) illustrates how the parallel GP algorithm described above is integrated into the modelling work-flow. Analysis steps prior to GP modelling (i.e., quantile regression for threshold estimation, Poisson modelling for estimation of rate of threshold exceedances, and a GP starting solution) are performed serially using improved GLAM algorithms described in §3. Subsequently, the parallel GP algorithm is executed in a high-performance computing environment and resulting output consolidated to estimate model parameters. Figure 2 (b) illustrates the distributed memory computing system. The parallel algorithm has $n_\lambda \times n_{CV}$ threads to compute, achieved by assigning up to $n_\lambda \times n_{CV}$ computing cores spread over many nodes. In the illustration shown we assume n independent nodes each with k cores, totalling $N (= k \times n)$ parallel cores. If we set $n_\lambda \times n_{CV} = N$, each thread can be assigned a unique core. If $N < n_\lambda \times n_{CV}$, fewer operations are performed in parallel; the remainder are scheduled after tasks from earlier batches are complete. In the next section, we discuss metrics to determine the optimal number of cores for efficient and rapid analysis.

4.2. Reduction in execution time

The parallel GP algorithm provides considerable reduction in execution time over serial. Figure 3 (a) shows the average run time (over three runs) of the parallel GP algorithm, as a function of the total number of cores N , for a trial analysis with $n_\lambda = 13$ and $n_{CV} = 10$. Serial computations previously taking days to execute are now complete in a few hours. Reductions of close to two orders of magnitude in execution times are achieved until $N \approx 75$, beyond which run time improvement diminishes. Figure 3 (b) compares execution time t_N for an N -core architecture against execution time t_1 using a single core. Defining ‘‘Speedup Factor’’ as

$$\text{Speedup Factor} = \frac{t_1}{t_N} \quad (8)$$

we can compare the realised speedup factor with the idealised speedup factor ($= N$ for an N -core architecture). Realised and idealised speedup factors diverge for $N > 75$, indicating limitations to the scalability of the parallel implementation due to communication, memory, thread wait and other overheads incurred. Quantifying these overheads is essential for judicious choice of architecture and estimation of execution times. Speedup characteristics are of course dependent on the architecture of computing resources available.

Figure 4 (a) illustrates that execution times for individual parallel tasks vary quite markedly, due to different values of penalty coefficient and cross-validation sample partition. Since the full analysis is complete only when all individual tasks are complete, this suggests different approaches to resource utilisation. We might *fully subscribe* as many cores as the number of parallel tasks ($N = n_\lambda \times n_{CV}$). Alternatively we might *under subscribe* cores, such that $N < n_\lambda \times n_{CV}$. When we fully subscribe (as shown in Figure 4 (a)), different tasks take different times. Once a core completes a relatively quick task it remains idle until it receives a communication that all tasks are complete; only then are resources released for other tasks. This leads to obvious inefficiencies and diminishing returns as more cores are added.

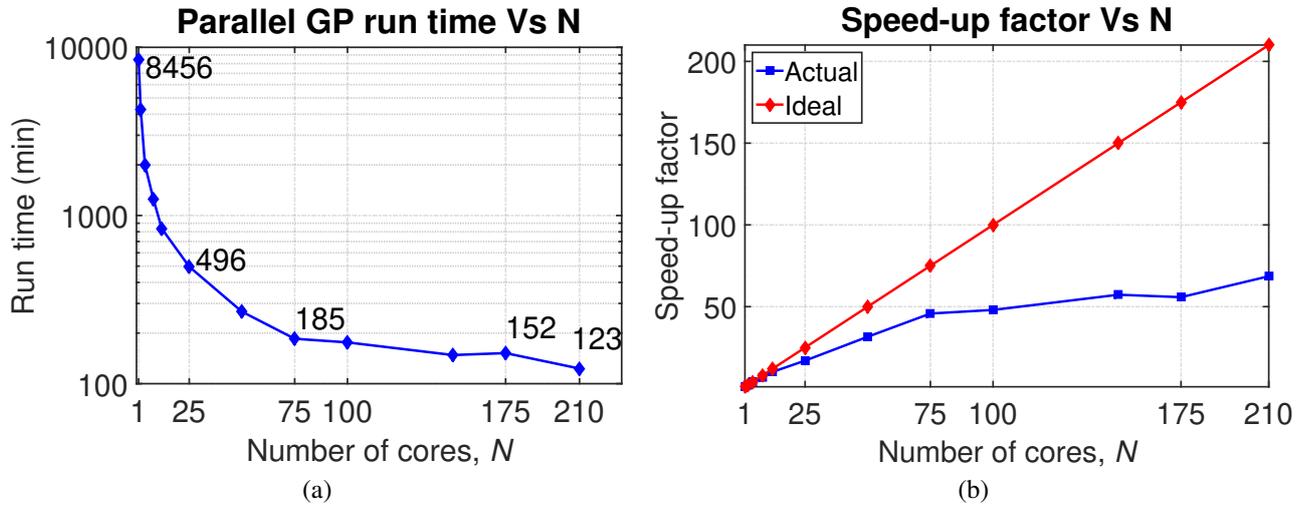


Figure 3: (a) Execution time of parallel GP algorithm as a function of the number of cores N assigned demonstrating close to two orders of magnitude reduction in execution time for $N < 75$. (b) Comparison of realised and idealised speedup factors which diverge with $N > 75$.

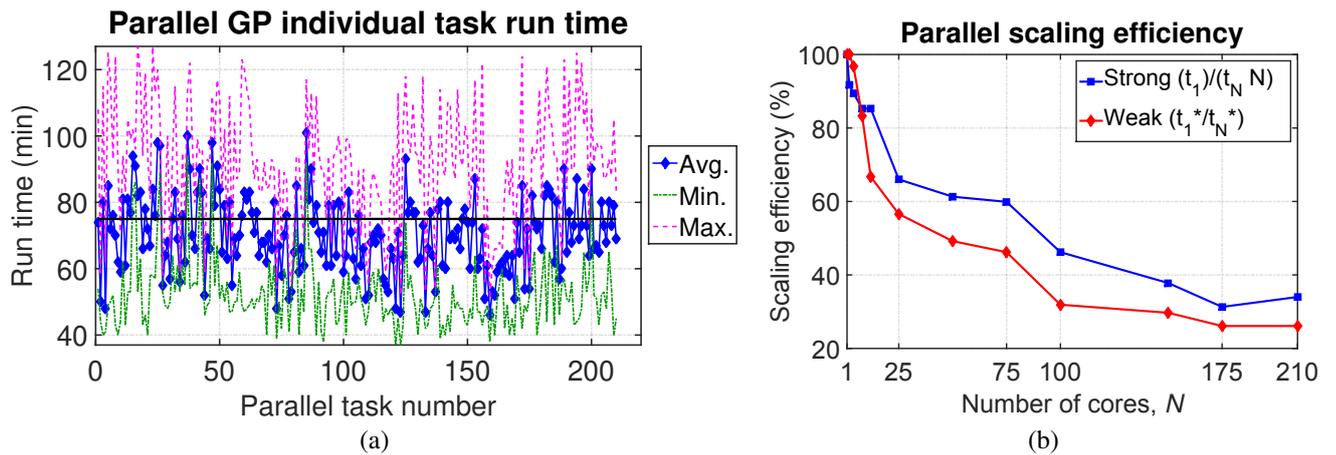


Figure 4: (a) Execution times for individual GP tasks for fully subscribed case ($N = n_l \times n_{CV}$) showing variations in run times of parallel tasks. The horizontal black line indicates overall mean execution time. (b) Strong and weak scaling efficiencies of the parallel GP algorithm to aid balancing of computing time with optimal resource usage.

4.3. Performance Test using Scalability Measures

We introduce two measures of scalability ([Kam13]) to determine the optimal number of cores empirically. For a constant problem size, we increase the number of cores and measure execution time and define a metric known as strong scaling efficiency:

$$E_{\text{STRONG}} = \frac{t_1}{t_N \times N} \times 100\% \quad (9)$$

where t_1 and t_N are defined as earlier. Using this measure, we determine a ‘‘sweet-spot’’ facilitating computations within reasonable time whilst also using allocated computing resources efficiently. A second measure of weak scaling efficiency is estimated by increasing problem size as a function of number of cores so that each core receives the same quantum of task:

$$E_{\text{WEAK}} = \frac{t_1^*}{t_N^*} \times 100\% \quad (10)$$

For example, we could use one core to fit one GP (say t_1^* for a given λ , CV), then two cores to fit two GP fits (t_2^*) and so on. Figure 4 (b) shows that both strong and weak scaling efficiencies drop below 50% with increasing N , corroborating inferences from Figure 4 (a). Setting expected efficiencies to around 60% and number of cores $N = 75$ appears a reasonable compromise.

The high performance computing architecture adopted here is a generic distributed memory computing system with multiple nodes. Each node has multiple CPU cores and shared memory linked via a high speed network. Additional enhancements to parallel task scheduling methodologies are largely unimportant in the current context of independent tasks not requiring inter-processor communication for data transfer. We might exploit relatively quick shared memory by scheduling certain tasks within the same computing node. To minimise wait cycles, we might also communicate task completion at a node and allocate pending tasks adaptively. A-priori analysis of task complexity might enable even greater efficiencies; particularly problematic individual GP optimisations might themselves then be partitioned into multiple tasks.

4.4. Uncertainty Analysis using Massive Parallelism

It is critical to quantify the uncertainty with which extreme value models are estimated. Re-sampling techniques such as bootstrapping can be used to estimate the uncertainty of model parameters and estimates of return values and other structure variables. [CDD05] 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. [JE07] use block bootstrapping to evaluate uncertainties associated with extremes in storm peak significant wave heights in the Gulf of Mexico.

In the current context, bootstrapping involves resampling the original sample with replacement to create a ‘‘bootstrap resample’’. The whole modelling procedure - including parameter and return value estimation - is then executed for the resampled data. By repeating this scheme for a large number of bootstrap resamples, we can estimate the size of sampling uncertainty on parameter and return value estimates. Model fitting for each bootstrap resample is independent of fitting for all others. Computationally, this problem is seen to be *embarrassingly parallel* ([Fos95]). In theory, given sufficient resources, all bootstrapping threads might be run simultaneously in parallel, potentially with each task utilising multiple cores using the parallel techniques described in §4. However in practice, there is a limitation on the number of computing resources available at a given time. Hence the task scheduler may impose restrictions on the number of cores that can be allocated to a given task. In such a scenario, a ‘‘job array’’ approach may be invoked to schedule only one or few bootstrapping iterations at a given time. The job array can be configured in such a manner as to automatically schedule the next set of bootstrapping iterations once the current tasks are done.

In §5.3, we discuss the estimation of parameter uncertainty for the Gulf of Mexico application in more detail.

5. Applications

This section illustrates the methodology in application to estimation of return values for storm severity, corresponding to long return periods, in the Gulf of Mexico (GoM). We also summarise a further similar application to the North West Shelf of Australia.

5.1. GoM: Background

The Gulf of Mexico is a major energy producing region accounting for 25% of oil and 14% of the natural gas production of the U.S. [GOM14a]. The Gulf is vast with a total area of 1,587,000 km². Tropical cyclones (or hurricanes) are intense revolving depressions that can generate extreme storm conditions. They can impact almost any part of the US Gulf coast, typically between June and November. In deep water, H_s exceeding 16 m is possible, with peak seas radiating outwards from the storm centre. Groups of large waves travelling in different directions close to the centre of the storm may combine to give exceptionally high individual wave heights in excess of 30 m ([API07],[ISO05]). Specification of realistic metocean criteria is critical for safe and efficient operation of offshore structures. Numerical

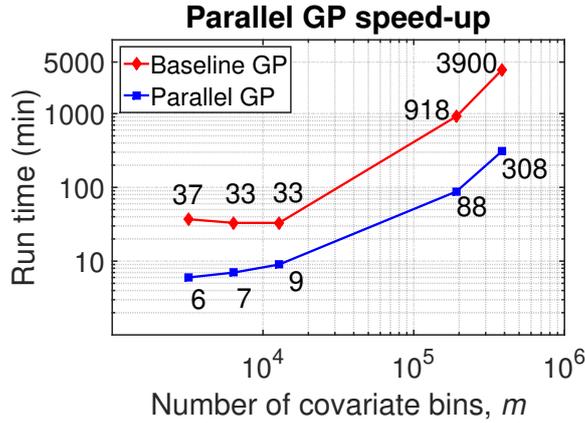


Figure 5: Comparative run-time performance (log-log scale) for GP fitting using HPC with different numbers of covariate bins.

hindcast studies of extreme and operational conditions include the Gulf of Mexico Oceanographic and Meteorological Study (GOMOS) [GOM14b]. GOMOS08 includes a 29-year continuous wind and wave hindcast for the period 1980-2008, 379 tropical events (hurricanes) for the period 1900-2008, and 68 extra-tropical events (winter storms) for the period 1950-2008. A 3rd generation wave model with 1/16th degree fine grid (7km) was used to model extreme storms. A total of 170 hurricane events at 9177 spatial locations were considered in this work.

5.2. GoM: Computational burden

Conventionally, metocean criteria are estimated using a site-specific approach due to lack of statistical and computational resources. Storm peak values from a small neighbourhood of locations centred around a particular reference location of interest are aggregated to create a sample for extreme value analysis, typically ignoring non-stationarity and spatial dependence. With the computational advances described in §3 and 4, we now have the capability to estimate marginal design conditions for the whole Gulf simultaneously. Physical considerations suggest that setting $p = p_x p_y p_\theta = 18 \times 18 \times 32$ provides good spatial and directional resolution. Using the GOMOS hindcast described above, we develop a marginal model for storm peak significant wave height for all spatial locations (at 1/16 degree resolution) and 32 directional sectors. This requires estimation on the index set of $m = m_x m_y m_\theta = 200 \times 60 \times 32 = 384,000$ covariate combinations for $p = 10$, 368 spline coefficients for each of GP shape and scale (and similar numbers for quantile threshold and Poisson rate). We might compare this with a typical problem for the neighbourhood of a single location of size $m = 10 \times 10 \times 32 = 3200$ covariate combinations.

Figure 5 shows comparative run-times for baseline estimation (GLAM version using 8 cores) and parallel approach (HPC version using 130 cores). Extreme value estimation for the whole of GOMOS can be completed (without uncertainty analysis) within a few hours.

5.3. GoM: Parameter estimation

Following a sensitivity study into the stability of GP shape parameter and median 100-year return value with threshold non-exceedance probability, a threshold non-exceedance probability of 0.5 was deemed appropriate. Figure 6 shows parameter estimates as functions of spatio-directional covariates, spatially for directional octants and for all directions (or “omni-directionally”). The threshold estimate is relatively stable with respect to both space and direction at around 2.5m. The rate of occurrence of threshold exceedances is highest for hurricanes from the East and South-East as expected from physical considerations. Omni-directionally, locations in the eastern Gulf exhibit both larger GP shape and scale, suggesting that extreme values in this region will be larger than elsewhere - again, consistent with historical evidence. We note that values of GP shape range from -0.1 to 0.06 and GP scale from 1.05 to 4.12. Note that the units for parameters are as follows: threshold (metres), rate of threshold exceedance (occurrences per degree per location per annum), GP shape (dimensionless) and GP scale (metres). Note that maximum likelihood estimators for GP shape and scale are asymptotically negatively dependent. It is common in applications to observe negative dependence between estimates of these parameters (see, e.g., [SM12]).

Figure 7 illustrates uncertainties in quantile regression, Poisson rate of threshold exceedance, GP shape and scale parameter estimates for the Gulf of Mexico estimated from 250 bootstrap resamples in terms of 2.5, 50 and 97.5 percentiles of bootstrap distributions on direction, longitude and latitude. Non-overlap of uncertainty estimates indicates (e.g. for Poisson rate) material difference between extreme value models for different combination of covariates. The relatively wide uncertainty bands for GP shape indicate difficulty in estimating the GP parameters for some combinations of covariates.

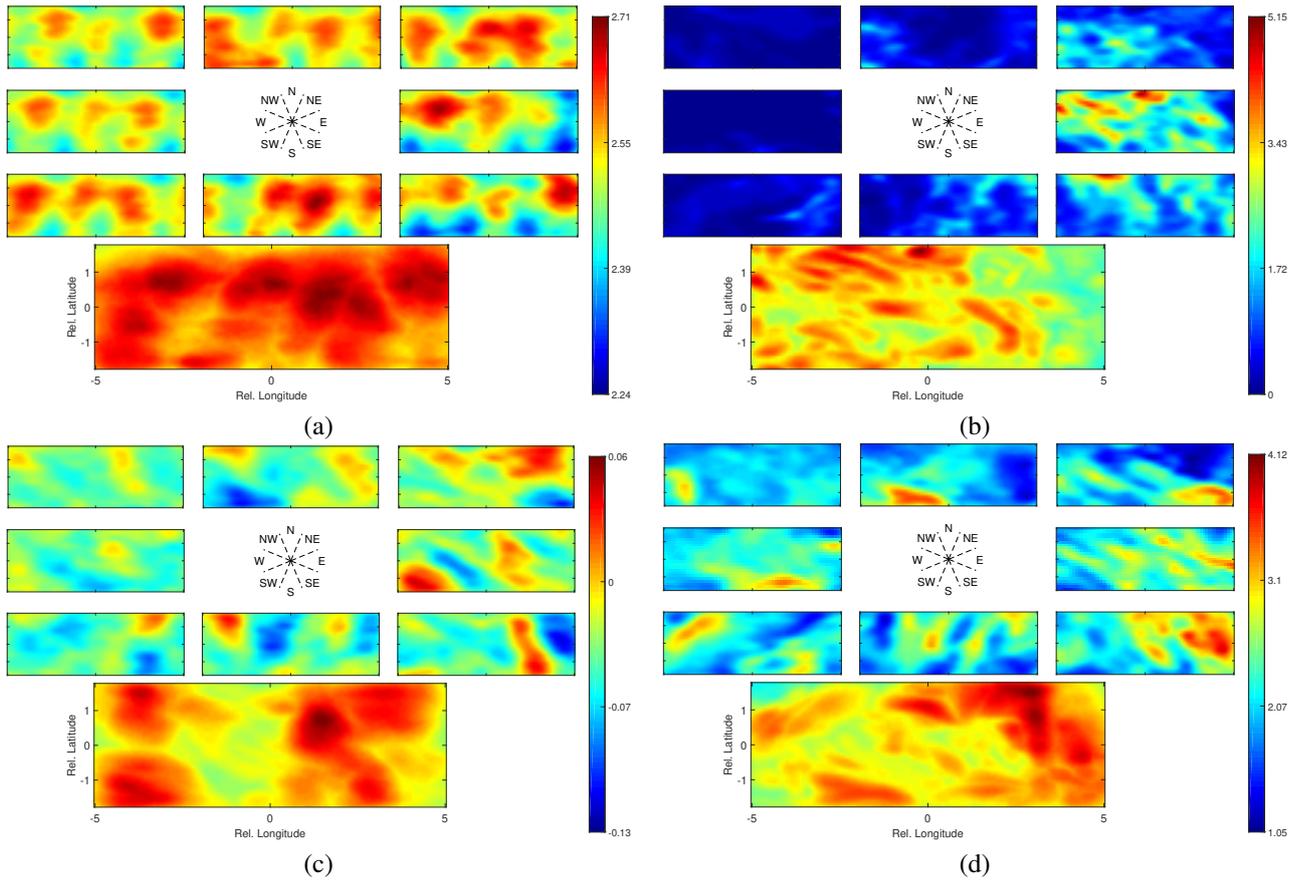


Figure 6: Spatio-directional plot of parameter estimates for the Gulf of Mexico: (a) threshold (metres), (b) rate of threshold exceedance (occurrences per degree per spatial location per annum), (c) GP shape (dimensionless) and (d) scale (metres). For each parameter, the lower plot shows median omnidirectional parameter estimates for each of 9177 locations. The 8 upper plots show median directional sector parameter estimates per location for 8 directional octants centred (from left to right, top to bottom) on storms from NW, N, and NE; W and E; SW, S and SE respectively.

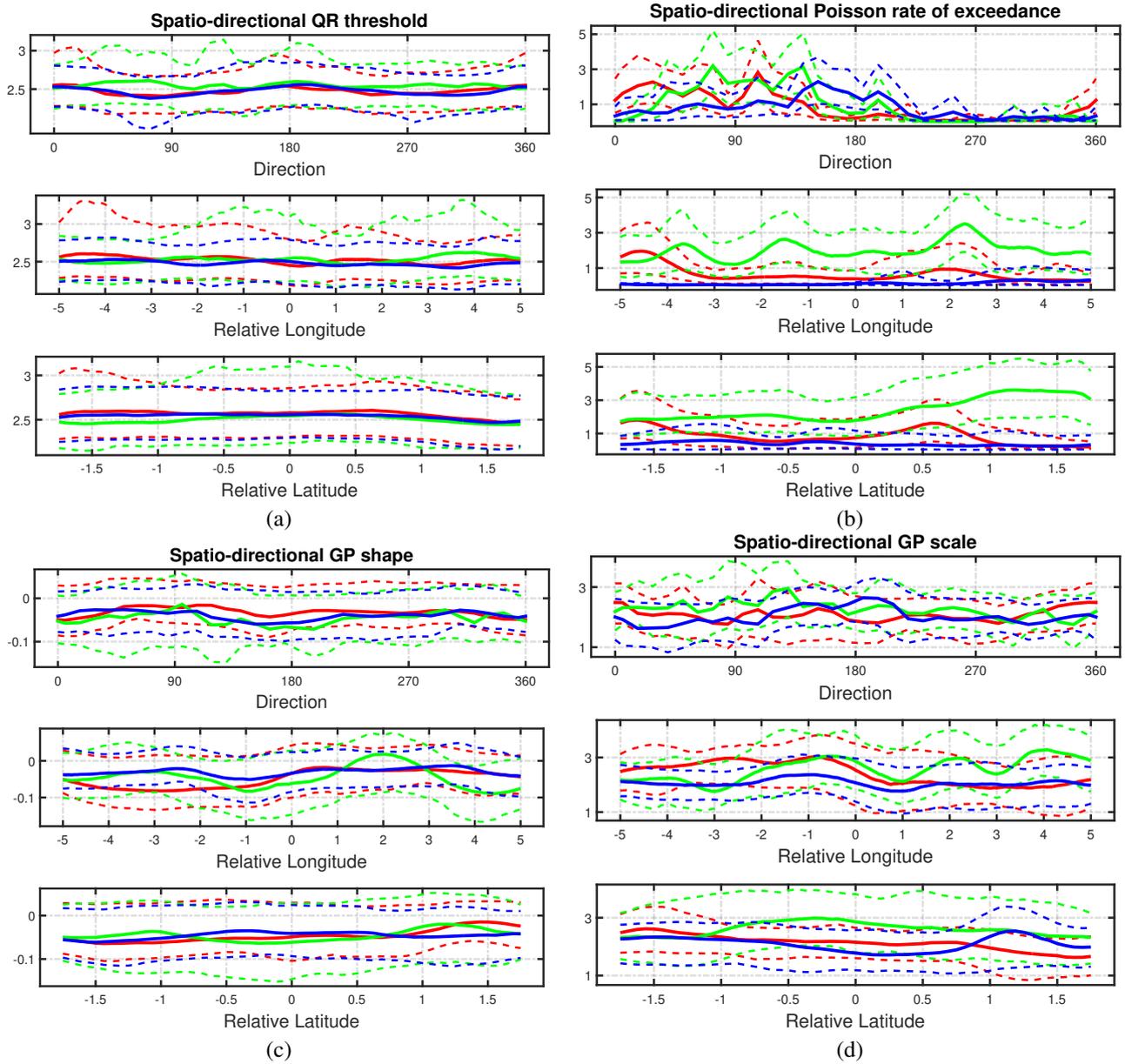


Figure 7: Model parameter uncertainty analysis for Gulf of Mexico, from 250 bootstrap iterations illustrated using the 2.5 (lower dashed), 50 (centre solid) and 97.5 percentile (top dashed) of the bootstrap distribution of parameter estimates. The panels show estimates for (a) quantile regression threshold (metres), (b) Poisson rate (occurrences per degree per location per annum), (c) GP shape (dimensionless) and (d) GP scale (metres). In each panel, the upper plot gives parameter estimates as a function of direction for typical SW (red), centre (green) and NE (blue) locations respectively; the middle plot as a function of longitude for storms from 0° and minimum latitude (red), from 120° and central latitude (green) and 240° and maximum latitude (blue); the bottom plot as a function of latitude for storms from 0° and minimum longitude (red), from 120° and central longitude (green) and from 270° and maximum longitude (blue).

5.4. GoM: Return Value Estimation

Return value distributions can be estimated by the simulation procedure described in [JRWE14]. Briefly, for each location in turn, the distribution of the 100-year return period is estimated omnidirectionally and for 8 directional sectors of equal size centred on cardinal and semi-cardinal directions, using all of the estimated extreme value models (for the actual sample and its bootstrap resamples). Figure 8 (a) shows median 100-year marginal return values for the entire Gulf, incorporating directional dissipation (as described in [JRWE14]). Omnidirectional estimates (lower panel) of approximately 15m are consistent with expectation. Directionally, the most severe storms occur in the eastern and south-eastern sectors, consistent with physical understanding. Figures (b) and (c) show estimates for the 2.5 and 97.5 percentiles of the return value distribution respectively. Largest return values, and largest return value uncertainties, occur in the “hurricane alley” region. Using model output such as that in the figure, the metocean specialist can access extreme value estimates for any spatial location and directional octant of choice, generated by a thorough consistent approach for the whole ocean basin.

5.5. GoM: Model diagnostics

Model diagnostics are essential to demonstrate adequate model fit. Of primary concern is that the estimated extreme value model generates spatio-directional distributions consistent with observed storm peak data (see [FRW⁺15], [JRWE14]). Figure 9(a) provides a spatial comparison of cumulative distribution functions for return values for the period of the original sample, estimated using the original sample (red) and using 1000 realisations under the model (black) for a 3×3 grid of locations in the eastern Gulf, corresponding approximately to the so-called “hurricane alley”, where interest in return value estimation is greatest. Figure 9(b) gives the corresponding directional comparison. There is good agreement. Wider uncertainty bands in western directional sectors are due to lower rates of occurrence of hurricanes.

5.6. North West Shelf application

We briefly describe a second application to estimation of return values of storm peak significant wave height for storms on the North West Shelf (NWS) of Australia. The regional climate is monsoonal, displaying two distinct seasons, “winter” from April to September and “summer” from October to March, with very rapid transitions between these, generally in April and September to October. Tropical cyclones during summer months are most influential for extreme metocean criteria, producing severe sea conditions from 4.0 m and often exceeding 8.0 m significant wave height. Typically, 5 named tropical storms per year occur in Western Australia [BOM14]. The North Australia Metocean Study (NAMOS) provides hindcast data for tropical cyclones [NAM14]. The application sample considered here corresponds to approximately 1.3×10^9 values of storm peak H_S and dominant wave direction (at storm peak H_S) for 76 hindcasts at 4095 locations on a 65×63 regular grid over the North West Shelf (NWS) of Western Australia for the period 1970-2007, corresponding to a spatial domain of approximately 4° longitude by 4° latitude. Physical considerations again suggest that setting the number of spline coefficients at $p = p_x p_y p_\theta = 18 \times 18 \times 32$ is reasonable. Based on stability analysis of GP shape and 100-year return value with threshold non-exceedance probability, we chose a non-exceedance threshold probability of 0.5 for extreme value analysis. The extreme value threshold is found to be approximately constant in space and direction at around 3.5m. However, the rate of threshold exceedance shows obvious spatial and directional variation, largest in the north-west for storms emanating from the west. Estimates for extreme value shape parameter are lower for NWS compared with GoM, suggesting that the distribution of storm peak H_S is more heavy-tailed in GoM. Return value distributions for the 100-year event, estimated by simulation, indicate omnidirectional return values from 8m to 9m, with highest values in the east and north-east. The largest return values occur for cyclones from the west, and lowest for storms from the south-west.

6. Discussion

Non-stationary marginal extreme value modelling for complete ocean basins is computationally complex, necessitating the estimation of millions of parameters, and large simulations under fitted models for return value estimation and uncertainty analysis. In this paper, we adopt two computational strategies to make this analysis feasible and orders of magnitude more efficient than naive approaches. Using generalised linear array methods, we reduce the computational burden of matrix manipulations of tensor products of spline bases. Using high-performance computing, we develop a parallel implementation of maximum likelihood estimation for the generalised Pareto distribution. We apply the methodology to marginal spatio-directional extreme value modelling of storm peak significant wave heights for the Gulf of Mexico and North West Shelf of Australia. Using the methodology, metocean design conditions can be estimated for large spatial grids of locations, incorporating the effects of non-stationarity, adjusting for spatial dependence and uncertainty analysis in a sound and consistent fashion. In this sense, the methodology is a considerable improvement on current practice. For preliminary design at least, the methodology offers the possibility that previously-computed extreme value parameter and return value estimates might be browsed interactively via a suitable user interface throughout an ocean basin, providing immediate design criteria and removing the need for the metocean engineer to perform extreme value analysis at all.

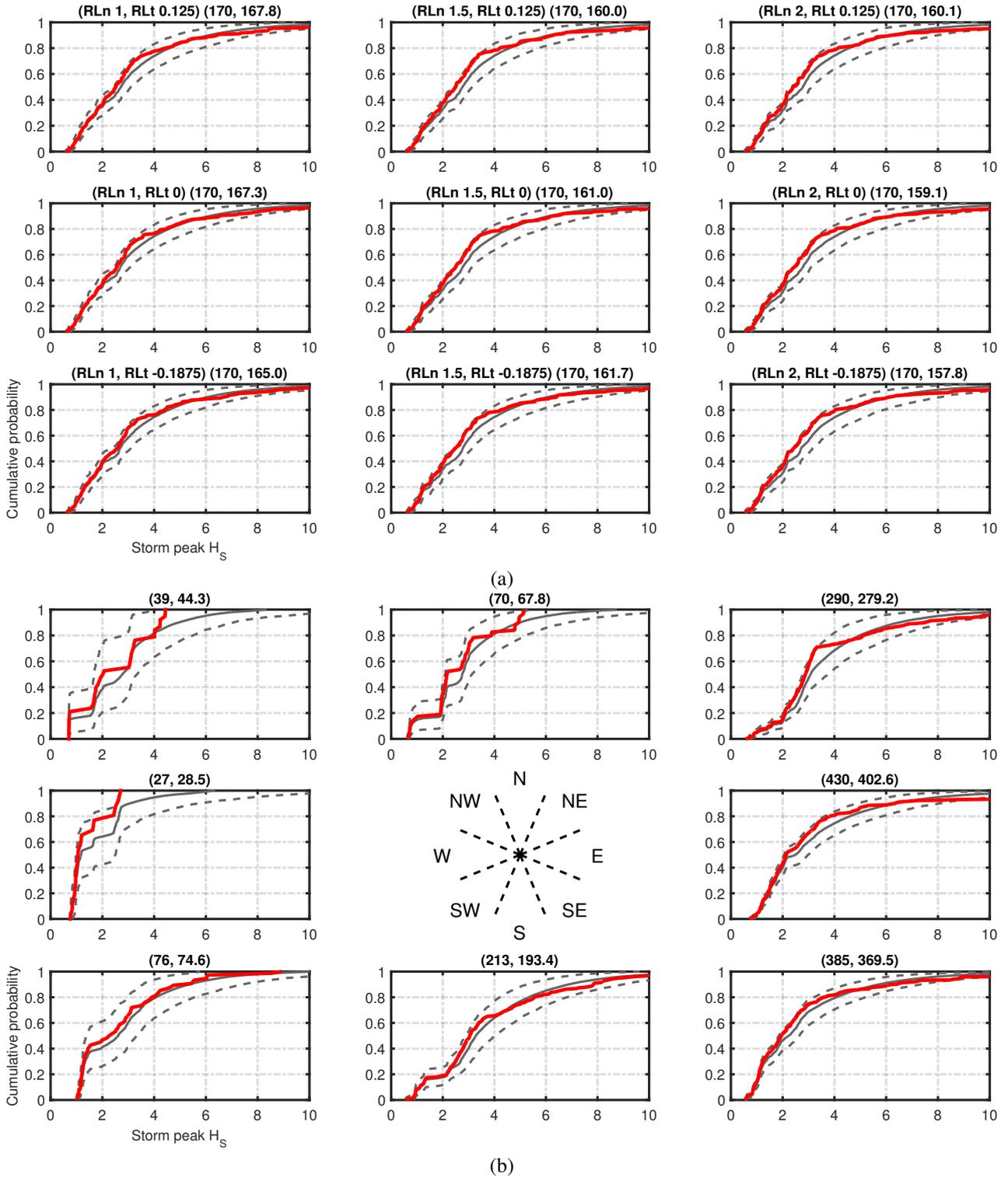


Figure 9: Validation of spatio-directional model by (a) spatial and (b) directional comparison of cumulative distribution functions (cdfs) for return value distributions of the original sample (red) with those from 1000 realisations under the model corresponding to the same time period as the original sample, shown in terms of the median (solid grey) and 95% uncertainty band (dashed grey). Figure (a) corresponds to omni-directional comparisons for a 3×3 grid of locations in the spatial domain, the relative longitudes and latitudes of which are given in the panel titles. The second pair of numbers in the panel title gives the number of observed and (average) simulated observations. Figure (b) gives a comparison of distributions for directional octants (for the same 3×3 grid of locations) centred (from left to right, top to bottom) on NW, N, and NE; W and E; SW, S and SE respectively. Panel titles give the number of observed and (average) simulated observations. Return values are measured in metres.

It is important to consider the extent to which the capability to estimate large-scale non-stationary extreme value models rapidly, demonstrated here, is practically useful. From an engineering perspective, the consistency offered by application of a single basin-wide analysis is very attractive compared to ad-hoc analyses (often stationary) for single locations. However, it is also clear that a high spatial and directional resolution exists beyond which further increases in resolution are of little value. For example, directionally, the characteristics of storms are approximately stationary within a $20^\circ - 30^\circ$ sector in general; this suggests that, for extreme value analysis of storm severities, a directional resolution of approximately 10° is probably sufficient in general. Similarly, spatial characteristics of storm severity are approximately stationary within a 100km neighbourhood in general, suggesting a limiting practical spatial resolution. Moreover, the characteristics of the sample from which inferences are to be made also obviously suggests limits to our ability to resolve non-stationary features. An important advantage of a non-parametric smoothing approach is the avoidance of the need for model comparison as a means of model selection. Penalised likelihood estimation provides optimally smooth parameter functions for a given sample. In this particular application, both the generalised Pareto shape and scale parameters are indeed near constant, but the rate of occurrence of threshold exceedances is clearly non-stationary. In other applications this may not be the case. The non-parametric penalised likelihood approach is equally and generally applicable.

There are considerable opportunities for application and further development of this approach. The current work addresses extreme value modelling of one variate (e.g. storm peak H_S) conditional on multidimensional covariates (e.g. space-direction). Other recent work (e.g. [JER13]) has outlined how the method can be extended to the conditional modelling of one variate (e.g. storm peak current speed) given extreme values of another (e.g. storm peak H_S) and one or more covariates (e.g. current direction, storm direction and season). Further, spatial dependence models (e.g. [DPR12]) seek to characterise extremal spatial dependence across multiple locations as well as marginal non-stationarity. Such models are central to good metocean design; they are computationally even more complex than the current model, and would benefit considerably from the computational improvements developed here.

Acknowledgments

We would like to acknowledge the support and encouragement of Bertwim van Beest and Vianney Koelman from the Computational Technology Platform, metocean guidance from Graham Feld and Michael Vogel, and useful discussions with Shell colleagues in Bangalore and Manchester, and with colleagues at Lancaster University.

References

- [API07] API. American Petroleum Institute - Interim Guidance on Hurricane Conditions in the Gulf of Mexico., *Bulletin 2INT-MET*, 2007.
- [BOM14] BOM. Tropical Cyclones in Western Australia, Bureau of Meteorology. <http://www.bom.gov.au/cyclone/index.shtml>, 2014. Accessed: 2014-06-17.
- [CDD05] V. Chavez-Demoulin and A.C. Davison. Generalized Additive Modelling of Sample Extremes. *J. R. Stat. Soc. C: Applied Statistics*, 54:207, 2005.
- [CDE06] I.D. Currie, M. Durban, and P.H.C. Eilers. Generalized Linear Array Models with Applications to Multidimensional Smoothing. *J. R. Stat. Soc. B*, 2006.
- [Dav03] A. C. Davison. *Statistical Models*. Cambridge University Press, 2003.
- [DPR12] A. C. Davison, S. A. Padoan, and M. Ribatet. Statistical Modelling of Spatial Extremes. *Statistical Science*, 27:161–186, 2012.
- [DTV98] J. M. Dixon, J. A. Tawn, and J. M. Vassie. Spatial Modelling of Extreme Sea-levels. *Environmetrics*, 9:283–301, 1998.
- [EM10] P.H.C. Eilers and B.D. Marx. Splines, Knots and Penalties. *Wiley Interscience Reviews: Computational Statistics*, 2:637–653, 2010.
- [Fos95] I. Foster. *Designing and Building Parallel Programs*. Addison-Wesley, 1995.
- [FRW⁺15] G. Feld, D. Randell, Y. Wu, K. Ewans, and P. Jonathan. Estimation of storm peak and intra-storm directional-seasonal design conditions in the North Sea. *J. Offshore. Arct. Eng.*, 137:021102:1–15, 2015.
- [GOM14a] GOM. U.S. Offshore Oil and Gas. http://en.wikipedia.org/wiki/Offshore_oil_and_gas_in_the_US_Gulf_of_Mexico, 2014. Accessed: 2014-12-15.
- [GOM14b] GOMOS08. Gulf of Mexico Oceanographic Study, Oceanweather, Inc. <http://www.oceanweather.com/metocean/gomos/index.html>, 2014. Accessed: 2014-12-15.
- [ISO05] ISO. International Standards Organization: Petroleum and Natural Gas Industries - Specific Requirements for Offshore Structures. *ISO 19901-1*, 2005.
- [JE07] P. Jonathan and K.C. Ewans. Uncertainties in Extreme Wave Height Estimates for Hurricane Dominated Regions. *J. Offshore Mech. Arc. Eng.*, 129:300–305, 2007.
- [JE08] P. Jonathan and K. C. Ewans. On Modelling Seasonality of Extreme Waves. In *Proc. 27th Intl. Conf. Offshore Mech. Arc. Eng., 4-8 June, Estoril, Portugal*, 2008.
- [JE11] P. Jonathan and K. C. Ewans. A Spatiodectional Model for Extreme Waves in the Gulf of Mexico. *ASME J. Offshore Mech. Arct. Eng.*, 133:011601, 2011.
- [JE13] P. Jonathan and K. C. Ewans. Statistical Modelling of Extreme Ocean Environments with Implications for Marine Design : A Review. *Ocean Eng.*, 62:91–109, 2013.
- [JER13] P. Jonathan, K.C. Ewans, and D. Randell. Joint Modelling of Extreme Ocean Environments Incorporating Covariate Effects. *Coastal Eng.*, 79:22–31, 2013.
- [JRWE14] P. Jonathan, D. Randell, Y. Wu, and K. Ewans. Return Level Estimation from Non-stationary Spatial Data Exhibiting Multidimensional Covariate Effects. *Ocean Eng.*, 88:520–532, 2014.
- [Kam13] Alan Kaminsky. *Big CPU, Big Data - Solving the World's Toughest Computational Problems with Parallel Computing*. Creative Commons, 2013.

- [NAM14] NAMOS. North Australia Metocean Study, Oceanweather, Inc. <http://www.oceanweather.com/metocean/namos/index.html>, 2014. Accessed: 2014-12-15.
- [RYJE13] D. Randell, Y. Yu, P. Jonathan, and K. C. Ewans. Modelling Covariate Effects in Extremes of Storm Severity on the Australian North West Shelf. In *Proc. 32nd Intl. Conf. Offshore Mech. Arc. Eng., Nantes, France*, 2013.
- [SM12] C. Scarrott and A. MacDonald. A review of extreme value threshold estimation and uncertainty quantification. *Revstat*, 10:33–60, 2012.
- [Sto74] M. Stone. Cross-validators Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc.*, 36:111–147, 1974.

A. Formulation of Spline Model

We represent model parameter η on the index set of m covariate bins as a linear combination of smoothly varying basis functions in matrix B for the covariate domain and spline coefficient vector γ , such that

$$\eta = B\gamma. \quad (\text{A.11})$$

A1. Quantile Regression

A smoothing spline through empirical quantiles per covariate bin is used as a starting solution for quantile regression whose penalised lack of fit criterion is

$$\ell_\phi^* = \ell_\phi + \lambda_\phi R_\phi \quad (\text{A.12})$$

where R_ϕ is the roughness parameter and λ_ϕ a roughness penalty. Lack of fit for a given quantile level $\tau \in [0, 1]$ ($\tau = 0.5$ implies 50% quantile or median level) is given by

$$\ell_\phi(\tau) = \tau \sum_{r_i \geq 0} |r_i| + (1 - \tau) \sum_{r_i < 0} |r_i| \quad (\text{A.13})$$

where r_i is the residual expressed in basis function terms as

$$r_i = y_i - (B\gamma)_i. \quad (\text{A.14})$$

To estimate the quantile function we solve

$$\min_{\gamma} \left\{ \tau \sum_{r_i \geq 0} |y - B'\gamma| + (1 - \tau) \sum_{r_i < 0} |y - B'\gamma| + \lambda\gamma'P\gamma \right\} \quad (\text{A.15})$$

using unconstrained optimization for a range of λ , using cross-validation to estimate the optimal λ^* and the corresponding γ^* . The quantile threshold on the index set is then computed by

$$\phi = B\gamma^*. \quad (\text{A.16})$$

A2. Poisson Process

The number of threshold exceedances on the index set of covariate combinations is characterised by the approximate Poisson (negative log) likelihood ([CDD05])

$$\ell_\zeta = \sum_j \left(\rho_j \Delta - c_j \log \rho_j \right) \quad (\text{A.17})$$

where c_j is the number of occurrences in covariate bin j and ρ_j the corresponding rate of occurrence to be estimated. Δ is the volume element of the covariate bin. Putting $r = B\zeta$ and $r_j = \sqrt{\rho_j}$ to impose non-negativity of rate, the penalised (negative log) likelihood is

$$\ell_\zeta^* = \sum_j \left(r_j^2 \Delta - c_j \log r_j^2 \right) + \lambda_\zeta \zeta' P \zeta. \quad (\text{A.18})$$

Applying the Newton-Raphson formulation as described in §A3, we estimate the Poisson parameter as a rearrangement of

$$2B'(B\hat{\zeta}_0\Delta - \frac{c}{\hat{r}_0}) + \lambda P \hat{\zeta}_0 + (2B'(\Delta + \frac{c}{\hat{r}_0^2})B + \lambda P)(\hat{\zeta} - \hat{\zeta}_0) = 0 \quad (\text{A.19})$$

yielding the iterative step for $\hat{\zeta}$ given current value $\hat{r}_0 = B\hat{\zeta}_0$

$$\hat{\zeta} = 2 \left[B' \text{diag} \left(\Delta + \frac{c_j}{\hat{r}_{0j}^2} \right) + \lambda_\zeta P \right]^{-1} B' \text{vec} \left(\frac{c_j}{\hat{r}_{0j}} \right). \quad (\text{A.20})$$

A3. Generalised Pareto Model

The generalised Pareto (GP) distribution of threshold exceedance y of high threshold ϕ is

$$P(Y > y | Y > \phi) = \left(1 + \frac{\xi}{\sigma} (y - \phi) \right)^{-\frac{1}{\xi}} \quad (\text{A.21})$$

for real shape ξ and positive scale σ , such that $1 + \frac{\xi}{\sigma} (y - \phi) > 0$. The penalised GP likelihood function is

$$\ell_{\xi, \sigma}^* = \ell_{\xi, \sigma} + \lambda_\xi R_\xi + \lambda_\sigma R_\sigma \quad (\text{A.22})$$

with negative log GP likelihood

$$\ell_{\xi, \sigma} = \sum_i \log \sigma_i + \left(\frac{1}{\xi} + 1 \right) \log \left(1 + \frac{\xi}{\sigma_i} (y_i - \phi_i) \right). \quad (\text{A.23})$$

GP parameters are expressed as

$$\begin{aligned} \xi &= B\alpha, \text{ and} \\ \nu &= B\beta \end{aligned} \quad (\text{A.24})$$

where $\nu = \sigma(1 + \xi)$ is such that ξ and ν provide a pair of asymptotically uncorrelated (and independently-distributed) parameter estimates, greatly simplifying the Newton-Raphson computation (see, e.g., [JRWE14]). Equation (A.22) can be written

$$\ell_{\alpha, \beta}^* = \ell_{\alpha, \beta} + \lambda_\alpha \alpha' P \alpha + \lambda_\beta \beta' P \beta. \quad (\text{A.25})$$

The solution is obtained with a back-fitting algorithm by applying Newton-Raphson to likelihood scores $\frac{\partial \ell}{\partial \beta_a}$ for parameter β_a ($a = 1, 2, \dots, p$)

$$\frac{\partial \ell}{\partial \beta_a} \Big|_{\hat{\beta}} = \frac{\partial \ell}{\partial \beta_a} \Big|_{\beta_0} + \sum_b \frac{\partial^2 \ell}{\partial \beta_a \partial \beta_b} \Big|_{\beta_0} (\hat{\beta}_b - \beta_{0,b}) = 0 \quad (\text{A.26})$$

Computing components for the Newton-Raphson formulation (A.26)

$$\begin{aligned} u_j^* &= \frac{\partial \ell^*}{\partial \alpha_j} \\ &= \frac{\partial}{\partial \alpha_j} \left(\ell_{\alpha, \beta} + \lambda_\alpha \alpha' P \alpha + \lambda_\beta \beta' P \beta \right) \\ &= \frac{\partial \ell}{\partial \alpha_j} + \lambda_\alpha P \alpha \\ &= \sum_k \frac{\partial \ell}{\partial \xi_k} \frac{\partial \xi_k}{\partial \alpha_j} + \lambda_\alpha P \alpha \end{aligned} \quad (\text{A.27})$$

Using $\frac{\partial \xi_k}{\partial \alpha_j} = B$ from (A.24) and $\frac{\partial \ell}{\partial \xi_k} = u_\xi$,

$$u_j^* = B' u_\xi + \lambda_\alpha P \alpha \quad (\text{A.28})$$

Further

$$\begin{aligned} \frac{\partial u_j^*}{\partial \alpha_l} &= \frac{\partial^2 \ell^*}{\partial \alpha_j \partial \alpha_l} \\ &= \frac{\partial}{\partial \alpha_l} \left(B' u_\xi + \lambda_\alpha P \alpha \right) \\ &= B' \frac{\partial}{\partial \alpha_l} \left(\frac{\partial \ell}{\partial \xi_k} \right) + \lambda_\alpha P \\ &= B' \frac{\partial}{\partial \xi_l} \left(\frac{\partial \ell}{\partial \xi_k} \right) \cdot \frac{\partial \xi_k}{\partial \alpha_j} + \lambda_\alpha P. \end{aligned} \quad (\text{A.29})$$

With $H_\xi = -\frac{\partial^2 \ell}{\partial \xi_j \partial \xi_l}$ and $\frac{\partial \xi_k}{\partial \alpha_j} = B$,

$$\frac{\partial^2 \ell^*}{\partial \alpha_j \partial \alpha_l} = -B' H_\xi B + \lambda_\alpha P. \quad (\text{A.30})$$

Substituting the values from (A.28) and (A.30) into (A.26)

$$\left(B' H_{\hat{\xi}_0} B - \lambda_\alpha P \right) (\hat{\alpha} - \hat{\alpha}_{\xi_0}) = B' u_{\hat{\xi}_0} + \lambda_\alpha P \hat{\alpha}_0. \quad (\text{A.31})$$

Rearranging and cancelling the $\lambda_\alpha P \alpha_0$ term, yields the iterative algorithm for estimating $\hat{\alpha}$ from the current $\hat{\alpha}_0$

$$\hat{\alpha} = \left(B' H_{\hat{\xi}_0} B - \lambda_\alpha P \right)^{-1} \left(B' u_{\hat{\xi}_0} + B' H_{\hat{\xi}_0} B \hat{\alpha}_0 \right). \quad (\text{A.32})$$

Similarly the iterative step for estimating $\hat{\beta}$ from $\hat{\beta}_0$ is

$$\hat{\beta} = \left(B' H_{\hat{\gamma}_0} B - \lambda_\beta P \right)^{-1} \left(B' u_{\hat{\gamma}_0} + B' H_{\hat{\gamma}_0} B \hat{\beta}_0 \right). \quad (\text{A.33})$$