The Importance of Spatial Dependency when Pooling Data to Estimate Return Values

Ross Towe[1] Emma Eastoe[1] Jonathan Tawn[1] Nicolas Fournier[2] Philip Jonathan[3]

[1]Maths & Stats, Lancaster University; [2]Shell International Exploration & Production, Netherlands; [3]Shell Technology Centre Thornton, Chester

Introduction

- Extreme metocean conditions are vital for constructing and deconstructing any offshore structures.
- Uncertainty in estimates has to be acknowledged when projections are made.
- The issue of spatial dependency has to be considered when calculating return values.

Motivation

Elisabeth Mannshardt-Shamseldin et al (2010), *The Annals of Applied Statistics* :

- *Found that averages of individual station return values should be larger than of station averages return values.*
- **How does this pooling affect the estimates ?**
- **Does spatial dependence affect the estimates ?**

Methodology

Method 1

- Observations are averaged over each time point then an extreme value distribution is fitted.

Method 2

- Marginal extreme value distributions are fitted with the return values averaged over the entire grid.

Accounting for spatial dependency (Coles and Tawn, 1996) :

Method 3

- Method 1 adjusted by the Areal coefficient.

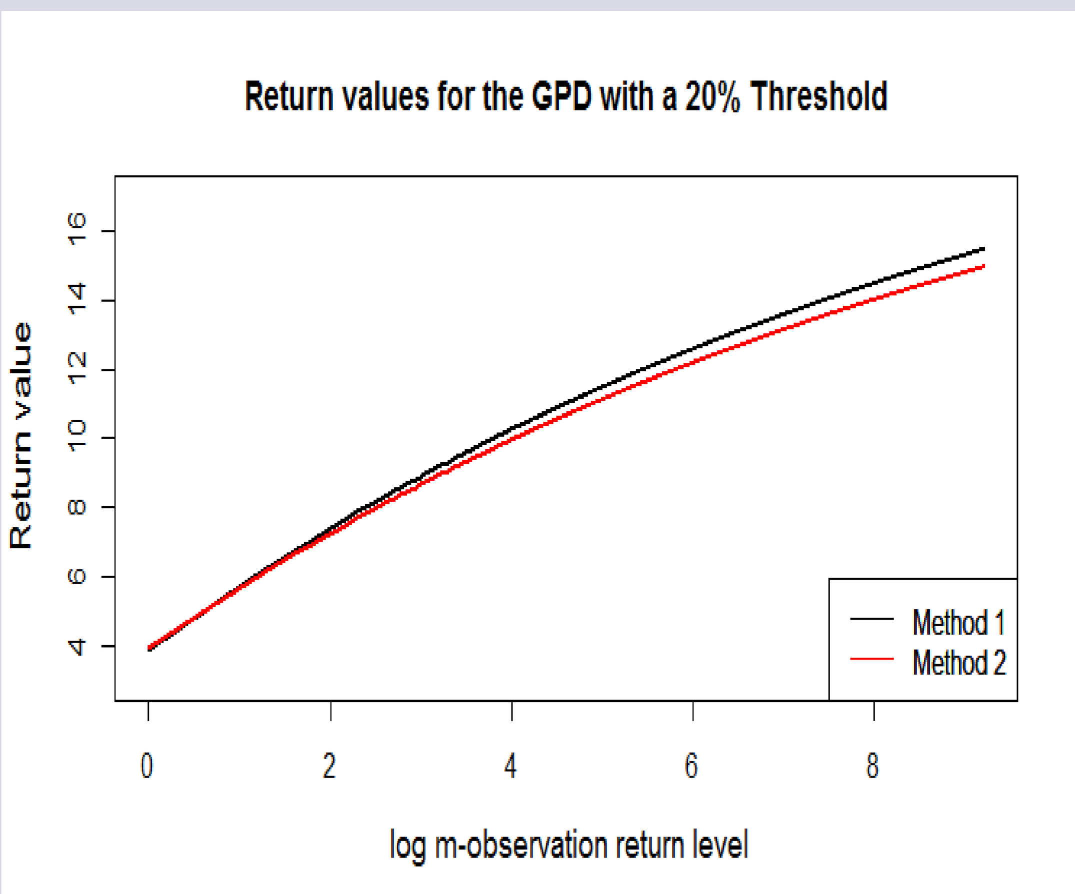
Method 4

- Method 2 adjusted by the Areal coefficient.

Shell hindcast data set

- 1147 significant storm peaks.
- Recorded over a 10 by 10 grid.
- Empirically temporally declustered by Shell's standard procedure.
- Univariate GPD fits.
- 20% threshold consistent with previous studies using the same data set (Jonathan, Ewans and Forristall, 2008).

Return value plot



Comments :

- Issues in determining suitable confidence intervals.
- Need to adjust for using pointwise data to model an aggregated process.

Simulation study

Data simulated from a Smith max-stable model (Smith, 1990) :

$$Y_i = \max_i\{\xi_i f(s_i, t)\}$$

- ξ_i represents the magnitude of a storm
- $f(s_i,t)=f_0(s-t) \sim MVN(s,\Sigma)$

Key characteristics of the data set :

- A 4 by 4 regular grid over a unit square.
- Standard unit Gumbel ($\mu = 0$, $\sigma = 1$, $\xi = 0$) margins were used.
- Marginal GEV fits.
- Data sets were simulated with a number of different covariance structures.
- 100 realisations of the max-stable process.

Covariance structure

Name	σ_{11}	σ_{22}	σ_{12}
A	1	1	0.5
B	1	1	-0.5
C	2	1	0
D	2	1	1

True Return values

- *100 year return value = log(100) = 4.61.*
- *1000 year return value = log(1000) = 6.91.*

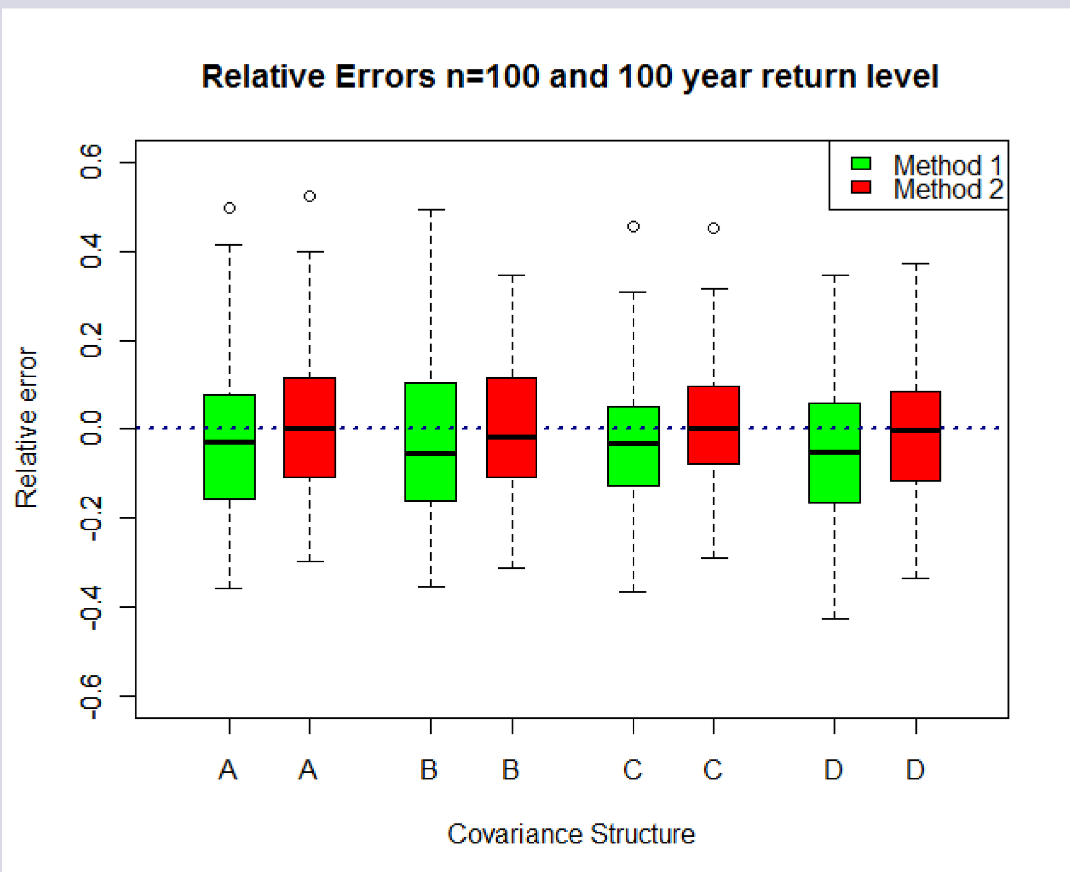
Simulation results

The average relative errors from the true 100 and 1000 year return values.

Sample size n=100 and n=10000

Mean *n=100*

Covariance	100yr		1000yr	
	1	2	1	2
A	-0.05	-0.01	0.01	0.01
B	-0.05	-0.01	0.01	0.01
C	-0.03	-0.01	0.01	0.01
D	-0.05	-0.01	-0.01	0.01



Mean *n=10000*

Covariance	100yr		1000yr	
	1	2	1	2
A	-0.04	-0.01	-0.01	-0.01
B	-0.04	-0.01	-0.01	-0.01
C	-0.02	-0.01	-0.01	-0.01
D	-0.05	-0.01	-0.02	-0.01

Comments

- Method 1 underestimates the return value more than method 2.
- Method 1 has a larger range than method 2.

Adjusting for spatial dependency

Areal Coefficient θ_V (Coles and Tawn, 1996) :

- Working again in Gumbel margins.
- Assumption of homogeneity in both scale and shape.

θ_V defined over the region $V = [a_1, b_1] \times [a_2, b_2]$

$$\theta_V = exp\left[-\frac{1}{24(1-\rho^2)}\left\{\left(\frac{b_1-a_1}{\sigma_x}\right)^2+\left(\frac{b_2-a_2}{\sigma_y}\right)^2\right\}\right]$$

- Use θ_V to adjust the return values to take account of the spatial dependency.

Spatial dependence results

Average relative errors again with 100 realisations of the max-stable process.

Mean *n=100*

Covariance	100yr		1000yr	
	3	4	3	4
A	-0.07	-0.03	-0.01	0.01
B	-0.08	-0.03	-0.03	-0.01
C	-0.08	-0.06	-0.05	-0.06
D	-0.05	-0.01	0.02	0.03

Mean *n=10000*

Covariance	100yr		1000yr	
	3	4	3	4
A	-0.07	-0.03	-0.03	-0.02
B	-0.07	-0.02	-0.03	-0.02
C	-0.07	-0.03	-0.04	-0.02
D	-0.03	-0.02	-0.01	-0.01

Comments

- Lower return values than method 1 and 2 as the return values have been adjusted for spatial dependence.
- The process was repeated for the median but no significant difference in estimates was found.
- However, the pooling method has an effect.
- Sample size has an effect though not majorly, this is good for Extreme value analysis.

Further Work

- Account for the bias that may arise from pooling.
- Incorporate the dependency between return values into the estimates.
- Minimise and quantify any uncertainty.
- Formulate confidence intervals for the return values.
- See if other max-stable models produce similar results.

References

- Coles, S.G and Tawn, J.A. (1996). *Journal of the Royal Statistical Society. Series B*, 58.2 :329–347. Modelling Extremes of the Areal Rainfall Process.
- Jonathan, P, Ewans, K and Forristall, G. (2008). *Ocean Engineering*, 11-12 :1211–1225. Statistical estimation of extreme ocean environments : The requirement for modelling directionality and other covariate effects.
- Mannshardt-Shamseldin, E. et al. (2010). *Ann. Appl. Stat*, 4.1 :484–502. Downscaling extremes : A comparison of extreme value distributions in point-source and gridded precipitation data.
- Smith, R. (1990). *Unpublished Manuscript* Max-Stable Processes and Spatial Extremes.

http://www.maths.lancs.ac.uk/department/info/people/person/towe r.towe@lancaster.ac.uk