Ocean Engineering 119 (2016) 30–46

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

Statistics of extreme ocean environments: Non-stationary inference for directionality and other covariate effects

Matthew Jones [a], David Randell [b], Kevin Ewans [c], Philip Jonathan [b,*]

[a] Department of Mathematical Sciences, Durham University, Durham DH1 3LE, United Kingdom
[b] Shell Projects & Technology, Manchester M22 0RR, United Kingdom
[c] Sarawak Shell Bhd., 50450 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 15 December 2015
Received in revised form
16 February 2016
Accepted 10 April 2016
Available online 29 April 2016

Keywords:
Extreme
Covariate
Non-stationary
Smoothing
Non-parametric
Spline
Gaussian process
mMALA
Kullback–Leibler

ABSTRACT

Numerous approaches are proposed in the literature for non-stationarity marginal extreme value inference, including different model parameterisations with respect to covariate, and different inference schemes. The objective of this paper is to compare some of these procedures critically. We generate sample realisations from generalised Pareto distributions, the parameters of which are smooth functions of a single smooth periodic covariate, specified to reflect the characteristics of actual samples from the tail of the distribution of significant wave height with direction, considered in the literature in the recent past. We estimate extreme values models (a) using Constant, Fourier, B-spline and Gaussian Process parameterisations for the functional forms of generalised Pareto shape and (adjusted) scale with respect to covariate and (b) maximum likelihood and Bayesian inference procedures. We evaluate the relative quality of inferences by estimating return value distributions for the response corresponding to a time period of 10 × the (assumed) period of the original sample, and compare estimated return values distributions with the truth using Kullback–Leibler, Cramer–von Mises and Kolmogorov–Smirnov statistics. We find that Spline and Gaussian Process parameterisations, estimated by Markov chain Monte Carlo inference using the mMALA algorithm, perform equally well in terms of quality of inference and computational efficiency, and generally perform better than alternatives in those respects.

1. Introduction

Accurate estimates of the likely extreme environmental loading on an offshore facility are vital to enable a design that ensures the facility is both structurally reliable and economic. This involves estimating the extreme value behaviour of meteorological and oceanographic (metocean) variables that quantify the various environmental loading quantities, primarily winds, wave, and currents. Examples of such parameters are significant wave height, mean wind speed and mean current speed. These characterise the environment for a given short period of time within which the environment is assumed to be stationary.

The long-term variability of these parameters is however non-stationary, in particular with respect to time, space and direction. From a temporal point of view metocean parameters generally have a strong seasonal variation, with an annual periodicity, and longer term variations due to decadal or semi-decadal climate variations. At any given location, the variability of a particular parameter is also dependent on the direction; for example, wind forcing is typically stronger from some directions than others, and fetch and water depth effects can strongly influence the resulting magnitude of the waves. Clearly these effects will vary with location: a more exposed location will be associated with longer fetches, resulting in a more extreme wave climate.

When estimating the long-term variability of parameters, such as significant wave height, the non-stationary effects associated with e.g. direction and season can be incorporated by treating direction and season as covariates. The common practice is to perform extreme value analysis of hindcast data sets, which include many years of metocean parameters, along with their associated covariates. Such data sets have all the information needed for input to covariate analysis.

From a design perspective, the metocean engineer is often required to specify return values for directional sectors such as octants centred on the cardinal and semi-cardinal directions. These directional return value estimates must be consistent with the estimated omnidirectional return value. In a similar manner, return values may be required corresponding to particular seasons or months of the year, consistent with an all-year return value.

* Corresponding author.
E-mail address: philip.jonathan@shell.com (P. Jonathan).

http://dx.doi.org/10.1016/j.oceaneng.2016.04.010
0029-8018/

Clearly, therefore, efficient and reliable inference for non-stationary extremes is of considerable practical interest, requiring estimation of (a) the rate and (b) the size of rare events. This work addresses the latter of these objectives.

A non-stationary extreme value model is generally superior to the alternative "partitioning" method sometimes used within the ocean engineering community. In the partitioning method, the sample is partitioned into subsets corresponding to approximately constant values of covariate(s); independent extreme value analysis is then performed on each subset. For example, in the current work we might choose to partition the sample into directional octants, and then estimate (8 independent stationary) extreme value models for each of the octants. There are two main reasons for favouring a non-stationarity model over the partitioning method. Firstly, the partitioning approach incurs a loss in statistical efficiency of estimation, since parameter estimates for subsets with similar covariate values are estimated independently of one another, even though physical insight would require parameter estimates to be similar. This problem worsens as the number of covariates and covariate subsets increases, and the sample size per subset decreases as a result. In the non-stationary model, we require that parameter estimates corresponding to similar values of covariates be similar, and optimise the degree of similarity during inference. For this reason, parameter uncertainty from the non-stationary model is generally smaller than from the partitioning approach. Secondly, the partitioning approach assumes that, within each subset, the sub-sample for extreme value modelling is homogeneous with respect to covariates. In general it is difficult to estimate what effect this assumption might have on parameter and return value estimates (especially when large intervals of values of covariates are combined into a subset). In the non-stationary model, we avoid the need to make this assumption.

Numerous articles have reported the essential features of extreme value analysis (e.g. Davison and Smith, 1990) and the importance of considering different aspects of covariate effects (e.g. Northrop et al., 2016). Carter and Challenor (1981) consider estimation of annual maxima from monthly data, when the distribution functions of monthly extremes are known. Coles and Walshaw (1994) describe directional modelling of extreme wind speeds using a Fourier parameterisation. Scotto and Guedes-Soares (2000) model the long-term time series of significant wave height with non-linear threshold models. Anderson et al. (2001) report that estimates for 100-year significant wave height from an extreme value model ignoring seasonality are considerably smaller than those obtained using a number of different seasonal extreme value models. Chavez-Demoulin and Embrechts (2006) describe smooth extreme value models in finance and insurance. Chavez-Demoulin and Davison (2005) provide a straight-forward description of a nonhomogeneous Poisson model in which occurrence rates and extreme value properties are modelled as functions of covariates. Cooley et al. (2006) use a Bayesian hierarchical model to characterise extremes of lichen growth. Renard et al. (2006) consider identification of changes in peaks over threshold using Bayesian inference. Fawcett and Walshaw (2006) use a hierarchical model to identify location and seasonal effects in marginal densities of hourly maxima for wind speed. Mendez et al. (2008) consider seasonal non-stationarity in extremes of NOAA buoy records. Randell et al. (2015a) discuss estimation for return values for significant wave height in the South China Sea using a directional-seasonal extreme value model. Randell et al. (2014) explore the directional characteristics of hindcast storm peak significant wave height with direction for locations in the Gulf of Mexico, North-West Shelf of Australia, Northern North Sea, Southern North Sea, South Atlantic Ocean, Alaska, South China Sea and West Africa. Fig. 1 illustrates the essential features of samples such as these. The rate and magnitude of occurrences of storm events vary considerably between locations, and with direction at each location. There are directional sectors with effectively no occurrences, there is evidence of rapid changes in characteristics with direction and of local stationarity with direction. Any realistic model for such samples needs to be non-stationary with respect to direction.

The objective of this paper is to evaluate critically different procedures for estimating non-stationary extreme value models. We quantify the extent to which extreme value analysis of samples of peaks over threshold exhibiting clear non-stationarity with respect to covariates, such as those in Fig. 1 or simulation case studies in Section 3 below, is influenced by a particular choice of model parameterisation or inference method. The 6 simulation case studies introduced in Section 3 are constructed to reflect the general features of the samples in Fig. 1, with the advantage that the statistical characteristics of the case studies are known exactly, allowing objective evaluation and comparison of competing methods of model parameterisation and inference. Our aim is that the results of this study are generally informative about any application of non-stationary extreme value analysis. We generate sample realisations from generalised Pareto distributions, the parameters of which are smooth functions of a single smooth periodic covariate. Then we estimate extreme value models (a) using Constant, Fourier, B-spline and Gaussian Process parameterisations for the functional forms of generalised Pareto parameters with respect to covariate and (b) maximum likelihood and Bayesian inference procedures. We evaluate the relative quality of inferences by estimating return value distributions for the response corresponding to a time period of $10\times$ the (assumed) period of the original sample, and compare estimated return values distributions with the truth using Kullback–Leibler (e.g. Perez-Cruz, 2008), Cramer–von Mises (e.g. Anderson, 1962) and Kolmogorov–Smirnov statistics. We cannot hope to compare all possible parameterisations, but choose four parameterisations useful in our experience. Similarly, there are many competing approaches for maximum likelihood and Bayesian inference, and general interest in understanding their relative characteristics. For example, Smith and Naylor (1987) compare maximum likelihood and Bayesian inference for the three-parameter Weibull distribution. In this work, we choose to compare frequentist penalised likelihood maximisation (see Section 2.3) with two Markov chain Monte Carlo (MCMC) methods of different complexities. Non-stationary model estimation is a growing field. There is a huge literature on still further possibilities for parametric (e.g. Chebyshev, Legendre and other polynomial forms) and non-parametric (e.g. Gauss–Markov random fields and radial basis functions) model parameterisations with respect to covariates. Moreover, in extreme value analysis, pre-processing of a response to near stationarity (e.g. using a Box–Cox transformation) is preferred.

The outline of the paper is as follows. Section 2 outlines the different model parameterisations and inference schemes under consideration. Section 3 describes underlying model forms used to generate samples for inference, outlines the procedure for estimation of return value distributions and their comparison, and presents results of those comparisons. Section 4 provides discussion and conclusions.

2. Estimating non-stationary extremes

Consider a random variable Y representing an environmental variable of interest such as significant wave height. The characteristics of Y are dependent on covariates such as (wave) direction, season, location and fetch. In this work we assume that a single periodic covariate θ (typically direction, or season) is sufficient to characterise the non-stationarity of Y. That is, we assume