Covariate-Varying Threshold Selection Method in Non-Stationary Generalized Pareto Model

Afif Shihabuddin[1], Norhaslinda Ali*[2], and Mohd Bakri Adam[3]

[1,2,3]*Institute for Mathematical Research, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.*
[2,3]*Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

**Corresponding author: norhaslinda@upm.edu.my*

Non-stationary data usually exist in real life and influenced by covariates. The non-stationary extremes are usually modelled by setting a constant high threshold, u, where the threshold exceedances are modelled by Generalized Pareto distribution (GP). Covariates model is incorporated to the GP parameters to account for non-stationarity. However, the threshold, u, may be high enough for GP approximation on certain covariates but not on others, which in this case may violate the asymptotic basis of the GP model. In this paper, a covariate-varying threshold selection method based on regression tree is suggested and applied on simulated non-stationary data sets. The regression tree will be used to partition data sets into stationary groups with similar covariate condition. Thus, a constant high threshold can be fixed within each group. The tree-based threshold exceedances can then be modelled by stationary GP which is a simpler model compared to the non-stationary GP. Simulation study is done to demonstrate and assess the performance of this method compared to the conventional method. The results show that the proposed method is a reasonable complement to the conventional method.

I. Introduction

Extreme value techniques are widely used in many disciplines such as ocean wave modelling (Northrop and Jonathan, 2011), business cycle (de Carvalho et al., 2012), assessment of meteorological change (Wang et al., 2016) and also electricity demand (Sigauke and Bere, 2017). The objective of an extreme value analysis is to quantify the stochastic behaviour of a process at unusually large or small levels. Besides, extreme value analysis also provides a framework for forecasting via return levels.

There are two ways to define extremes in extreme value framework. The first one is block maxima where observations are partitioned or blocked into regular intervals and the maximum observations within each block (block maxima) are modelled by Generalized Extreme Value model (GEV). The next method is by setting a constant high threshold on the data sets and the exceedances are modelled by Generalized Pareto model (GP). The focus of this paper is on the second method.

Let y be the variable of interest and u is the fixed high threshold, Pickands III et al. (1975) shows that for values of y greater than u, given that u is sufficiently high, the exceedances of threshold u, can be modelled using GP model where the cumulative distribution function is

$$H(y) = 1 - \left[1 + \frac{\xi(y-u)}{\sigma}\right]^{-1/\xi},$$

where $y > u$ and $1 + \xi(-u)/\sigma > 0$. Scale and shape parameter are denoted by σ and ξ respectively.

Many methods on threshold selection have been proposed for the case of non-stationary extremes. Some of them uses goodness of fit tests to select the optimum threshold (Bader et al., 2018, Thompson et al., 2009). Thompson et al. (2009) use normality test on the difference of some estimated parameter obtained from two consequence thresholds. Starting from the lowest threshold, the goodness of fit test will be applied. The first threshold which failed to reject the null hypothesis of the test is selected as the optimum threshold. This is similar to Bader et al. (2018), except for a rule called ForwardStop rule which is applied to control the False Discovery Rate. Yang et al. (2018) proposed to use the plot of characteristic value against their corresponding thresholds as a threshold selection method. The lowest threshold value where the characteristic value is stable will be chosen as the optimum threshold.

In real world data sets, non-stationarity is apparent due to seasonal trends and covariates effect, (Northrop and Jonathan, 2011). The main issue in modelling non-stationary extremes using GP is threshold selection. Various methods have been proposed to tackle this issue. The most frequently used method in the literature is by incorporating covariate model in the parameter of GP to model the exceedances of a constant high threshold (Coles et al., 2001). However, the covariate model is usually incorporated in scale parameter, not in shape parameter. This is because the shape parameter is difficult to estimate precisely (Coles et al., 2001). Covariate model chosen to be incorporated in the scale parameter usually constructed by referring to the underlying trends that affect the data sets. However, it is difficult to determine the exact model for the covariates which affect the process. Besides, by setting a constant high threshold on a non-stationary extremes with covariates effect will violate the basis of the GP distribution approximation. At some covariate values, the high threshold might be high enough for GP approximation, however, at another different covariate values, the same high threshold would not be high enough for GP approximation.

To overcome this problem, a threshold that varies according to the covariate values is proposed. The covariate-varying threshold selection method is based on regression tree. The rest of this paper is organized as follows. In Section 2 the tree-based threshold selection method will be explained. In Section 3 the details of the simulation study on this method are described. In section 4, the results of the simulation study are presented. Finally, in Section 5 some concluding comments are discussed.

II. Methodology

Regression tree is a method to partition a data set recursively thus permitting a simple prediction model to be fitted within each cluster, represented by terminal node or leaf (Loh, 2011). The tree consists of parent node, internal node and terminal node. To determine which cluster an observation belongs to, firstly, all observations are placed at the root (parent node) of the tree. The observations at the root are split by answering binary questions which are related to the covariates affecting the observations, where observations with 'yes' answer will be placed at the left leaf (internal node) and observations with 'no' answer will be placed at the right leaf (internal node) (Shihabuddin et al., 2018). An example of the question asked is "Is $t < 228$?" where t is the covariate time of the observations.

The split is chosen such that it maximizes the reduction of the impurity level within the tree which is measured using sum of squared errors. The sum of squared errors for a tree T is

$$S = \sum_{c \in leaves(T)} \sum_{i \in c} (y_i - m_c)^2$$

where $m_c = \frac{1}{n_c} \sum_{i \in c} y_i$, is the mean of observations within leaf c. In other words, the impurity level is the deviation of each observations from their corresponding cluster mean. This step is repeated recursively to create new leaves(internal nodes). Each split will reduce the impurity level within the tree. However, the amount of the impurity level reduction for each split will be smaller than in the previous split. Nodes at the final level are called terminal nodes.

An important issue within regression tree is the stopping criterion, where the tree is stopped from growing. If the stopping criterion is too low, the tree will be overgrown whereby there will be a small number of observations in each cluster. However, too big a stopping criterion will produce clusters which are not homogeneous. The usual practice is by setting a bound, δ such that a split must reduce the sum squared errors of the tree not less than δ. The δ value is inversely related to the size of the tree or number of clusters produced by the tree. The smaller the δ value, the bigger the tree produced. In this study, the stopping criterion used is based on stationarity of the terminal nodes or clusters.

Let $x_1, \ldots, x_n$ be a non-stationary sequence of observations with $t_1, \ldots, t_n$ and $y_1, \ldots, y_n$ as covariates. Regression tree is used to partition the x_i sequence into m homogeneous stationary clusters. Since the purpose of using the regression tree in this study is to obtain stationary observations within each clusters, the stopping criterion for the regression tree is set such that if all clusters are stationary, the recursive partitioning will be stopped. However, for all data sets, the objective that all clusters are stationary cannot be achieved. Hence, the percentage of stationarity are selected arbitrarily at 90%. Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test is used in this study to test the stationarity of the observations within each cluster.

If all observations within each clusters are approximately stationary, a constant high threshold can be set within each cluster. The threshold is set in term of percentile. The percentile is kept similar for all clusters so that the rate of exceedances remain constant throughout the data set. By using percentile as threshold value, each cluster will have unique threshold value. In other words, each observation will have their own threshold value. These threshold values are arranged according to the index of observations to create varying threshold. In this study, the percentile value chosen to be threshold within each cluster is 95th percentile (Eastoe and Tawn, 2009).

III. Simulation Study

The tree-based threshold selection method will be demonstrated on simulated data sets. The data sets which are consist of GEV random variable, x are simulated using inverse sampling method. The simulated GEV random variables are as follows

$$x = \mu + \frac{\sigma((-\log F(x))^{-\xi} - 1)}{\xi}.$$

Let $X_1, \ldots, X_n$ is distributed by $\text{GEV}(\mu, \sigma, \xi)$, $M_n = \max(X_1, \ldots, X_n)$ will follow $\text{GEV}(\mu^*, \sigma^*, \xi)$ where

$$\mu^* = \mu + \frac{\sigma(n^{\xi} - 1)}{\xi} \text{ and } \sigma^* = n^{\xi}\sigma.$$

According to Coles et al. (2001), if block maxima of a dataset can be modelled with $\text{GEV}(\mu, \sigma, \xi)$, the exceedances of a high enough threshold, u set on the dataset can be modelled with $\text{GP}(\tilde{\sigma}, \xi)$. In our case,

$$\begin{aligned}\tilde{\sigma} &= \sigma^* + \xi(u - \mu^*) \\ &= n^{\xi}\sigma + \xi\left[u - \left(\mu + \frac{\sigma(n^{\xi} - 1)}{\xi}\right)\right]\end{aligned}$$

To induce non-stationarity in the data sets, covariates model is incorporated in the GEV location parameter, μ. Two covariate models are considered which are:

1. $\mu = \mu_0 + \mu_1\left(\frac{t}{n+1}\right) + \mu_2 y$ for linear trend.
2. $\mu = \mu_0 + \mu_1 \cos(\frac{2\pi t}{n}) - \mu_2 \sin(\frac{2\pi t}{n}) + \mu_3 y$ for cyclic trend.

These models are based on covariate models proposed by Eastoe and Tawn (2009). Here, t and n represent time covariate and number of observations respectively. Another covariate y is generated from standard normal distribution. Time covariate, t is included to create trends in the data sets, while the covariate y represents a random variable which might affect the variable x.

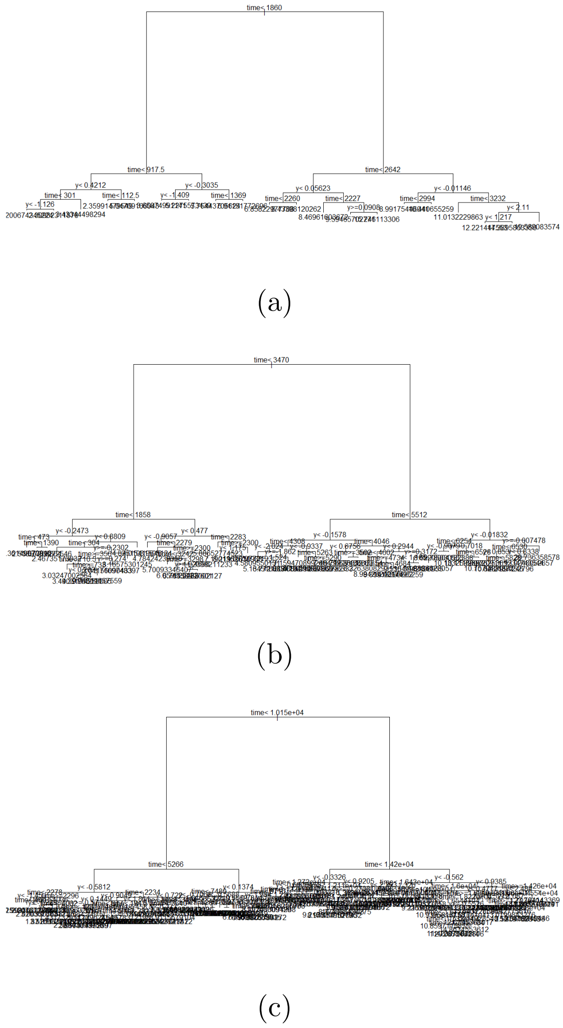
In this study, 36 different data sets with different number of observations and different parameter values are simulated. Three number of observations are selected arbitrarily, which are 3653, 7305 and 18263. These number of observations correspond to daily observations for 10 years, 20 years and 50 years respectively. The parameter μ_1(for linear and cyclic covariate model) and μ_2(for cyclic covariate model only) contribute more in making trend apparent in the data set because they are related to the covariate time. Hence, we vary the value of both parameters by setting them arbitrarily at 2.5, 5 and 10. The parameter μ_2(for linear covariate model), μ_3(for cyclic covariate model) and σ are kept constant value equal to 1. Besides, we also vary the shape parameter ξ at -0.4 and 0.4.

IV. Results and Discussion

The regression tree procedure explained before is applied on the simulated data sets. Number of clusters obtained and the reduction of SSE after the final split, δ are recorded for every procedure. Table 1 shows results for data sets with positive shape parameter; $\xi = 0.4$ while Table 2 shows results for data sets with negative shape parameter; $\xi = -0.4$.

From both tables, the results show that, as the number of observations within a data set increase, the partitioning needed to produce stationary clusters also increase. Besides, as the value of parameter related to the time covariate increase, the number of clusters produced also increase. This is related to the influence of the time covariate on the observations which increases when the value of the parameter get bigger. Data sets with negative shape parameter need more partitioning to obtain stationary clusters because the linear and cyclic trends are more apparent within these data sets.

For the sake of brevity, only one parameter value (3 data sets) for each combination of covariate model and shape parameter value are chosen to demonstrate the method. The data sets are chosen based on which parameter value produce largest range of number of clusters between different number of observations. The data sets chosen are highlighted in Table 1 and Table 2. The resulting regression tree are shown in Figure 1 and Figure 2 for data sets with positive shape parameter and Figure 3 and Figure 4 for data sets with negative shape parameter.



(a)

(b)

(c)

Figure 1: Regression tree for $\xi = 0.4$ and linear trend with (a) $n = 3653$, (b) $n = 7305$, (c) $n = 18263$

From Figure 1 until Figure 4, it can be observed that the size of the tree increases as

Table 1: Number of clusters obtained for regression tree in parentheses and the reduction of SSE after the final split, δ for data sets with positive shape parameter, $\xi = 0.4$

covariate model	parameter value	n=3653	n=7305	n=18263
linear	$\mu_1 = 2.5$	0.001895 (30)	0.001274 (40)	0.000719 (50)
	$\mu_1 = 5$	0.001357 (40)	0.000834 (42)	0.000464 (72)
	$\mu_1 = 10$	0.002119 (20)	0.000602 (52)	0.000023 (110)
cyclic	$\mu_1, \mu_2 = 2.5$	0.001106 (40)	0.000726 (50)	0.000362 (80)
	$\mu_1, \mu_2 = 5$	0.000654 (49)	0.000308 (73)	0.000109 (149)
	$\mu_1, \mu_2 = 10$	0.000101 (91)	0.000048 (162)	0.000035 (180)

Table 2: Number of clusters obtained for regression tree in parentheses and the reduction of SSE after the final split, δ for data sets with negative shape parameter, $\xi = -0.4$

covariate model	parameter value	n=3653	n=7305	n=18263
linear	$\mu_1 = 2.5$	0.001167 (30)	0.000433 (59)	0.000210 (86)
	$\mu_1 = 5$	0.000644 (52)	0.000318 (70)	0.000129 (114)
	$\mu_1 = 10$	0.000270 (66)	0.000105 (110)	0.000042 (181)
cyclic	$\mu_1, \mu_2 = 2.5$	0.000479 (57)	0.000152 (109)	0.000059 (181)
	$\mu_1, \mu_2 = 5$	0.000081 (109)	0.000058 (123)	0.000018 (239)
	$\mu_1, \mu_2 = 10$	0.000026 (124)	0.000013 (191)	0.000004 (377)

the number of observations in the data set increases. In addition to that, data sets with negative shape parameter tend to have bigger tree compared to data sets with positive shape parameter. Moreover, data sets with cyclic covariate model also have bigger tree than their counterparts with linear covariate model. This explains that data sets with more apparent trends need larger trees and produce more clusters so that each clusters are stationary.

The next step is to fix a constant high threshold within each stationary clusters obtained. The constant high threshold are chosen based on previous literature where the threshold is set at the nth percentile of the observations. In this study, 95th percentile is chosen as the threshold within the stationary clusters. Tree-based threshold can be obtained by arranging the threshold within clusters according to observations index. The resulting tree based threshold for all data sets are shown in Figure 5 until Figure 8. All the obtained tree-based thresholds follow the trend of the data. In other word, the obtained threshold vary according to the covariates affecting the observations.

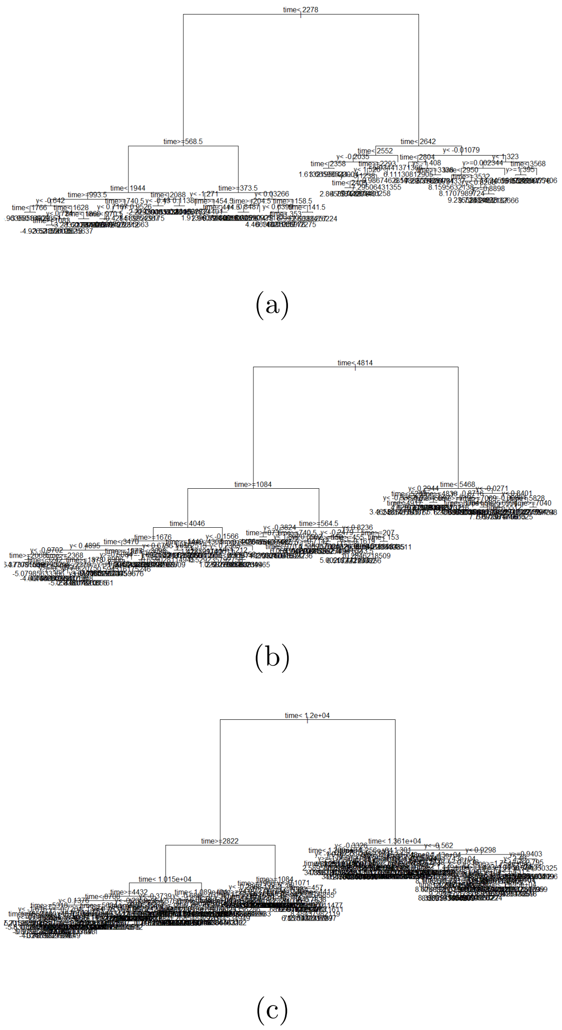
The exceedances of the tree-based thresholds are then modelled by stationary and non-stationary GP model. For the non-stationary GP model, covariate models incorporated in the GP scale parameter are given by:

$$\sigma = \exp\left\{\sigma_0 + \sigma_1\left(\frac{t}{n+1}\right) + \sigma_2 y\right\} \qquad (1)$$

$$\sigma = \exp\Big\{\sigma_0 + \sigma_1 \cos\left(\frac{2\pi t}{n}\right) - \sigma_2 \sin\left(\frac{2\pi t}{n}\right) + \sigma_3 y\Big\} \qquad (2)$$

where Equation 1 is for data set with linear trend and Equation 2 is for data set with cyclic trend.

The estimated parameter values are shown in Table 3 for $\xi = 0.4$ and in Table 4 for $\xi = -0.4$. Based on the results, most of the estimated shape parameter are close to the shape parameter value used for simulation. According to Coles et al. (2001), GP model fitted to

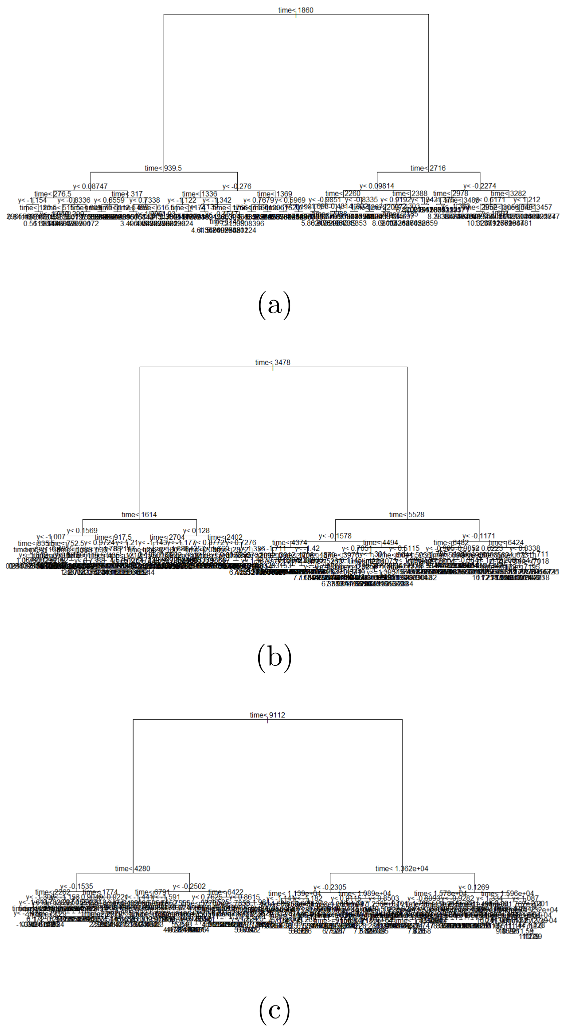


(a)

(b)

(c)

Figure 2: Regression tree for $\xi = 0.4$ and cyclic trend with (a) $n = 3653$, (b) $n = 7305$, (c) $n = 18263$



(a)

(b)

(c)

Figure 3: Regression tree for $\xi = -0.4$ and linear trend with (a) $n = 3653$, (b) $n = 7305$, (c) $n = 18263$

exceedances of high threshold, where the parent distribution is GEV, will have shape parameter same to the corresponding GEV shape parameter. Hence, it can be concluded that the tree-base threshold selection method does not violate the theoretical basis of GP approximation.

Comparison between fitted stationary GP model and fitted non-stationary GP model are done using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The AIC and BIC values are shown in Table 5 for $\xi = 0.4$ and Table 6 for $\xi = -0.4$. The results show that most AIC and BIC values for stationary model are smaller than for non-stationary model. Hence, modelling the exceedances of tree-based threshold by stationary GP model leads to less information loss compared to modelling by non-stationary GP model. This also concludes that regression tree method reduce the non-stationarity within the data set which leads to the suitability of stationary GP model for the tree-based threshold exceedances.

The goodness of fit of the stationary model fitted to the exceedances of the tree-based threshold is checked using two goodness of fit

Table 3: The estimated parameters of stationary GP model (S) and non-stationary GP model (NS) fitted to the exceedances of tree-based threshold for $\xi = 0.4$

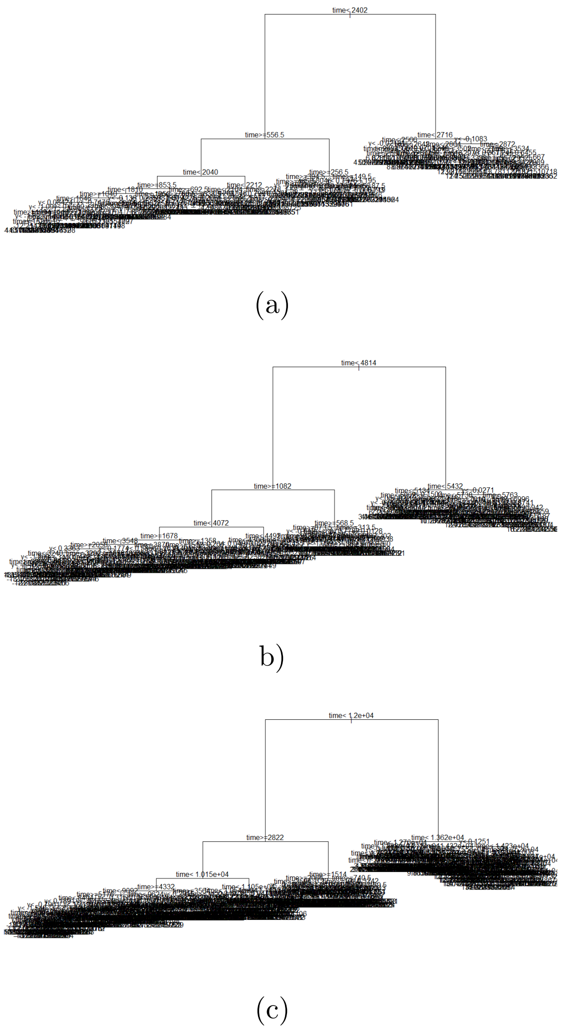
covariate model	n	model	σ_0	σ_1	σ_2	σ_3	ξ
linear	3653	S	2.773400	-	-	-	0.338944
		NS	1.140048	-0.254902	-0.038319	-	0.343374
	7305	S	3.173481	-	-	-	0.148101
		NS	1.175697	-0.035041	0.022537	-	0.144478
	18263	S	2.951356	-	-	-	0.239009
		NS	1.001323	0.165164	-0.079161	-	0.232361
cyclic	3653	S	2.126440	-	-	-	0.318758
		NS	0.762627	0.137990	-0.104260	-0.113842	0.300729
	7305	S	2.667115	-	-	-	0.230123
		NS	0.983395	-0.014219	0.117746	0.011770	0.226723
	18263	S	2.717581	-	-	-	0.274062
		NS	0.999397	0.007978	0.008781	-0.008925	0.273979

Table 4: The estimated parameters of stationary GP model (S) and non-stationary GP model (NS) fitted to the exceedances of tree-based threshold for $\xi = -0.4$

covariate model	n	model	σ_0	σ_1	σ_2	σ_3	ξ
linear	3653	S	0.341835	-	-	-	-0.289549
		NS	-1.148147	0.168006	0.038186	-	-0.311566
	7305	S	0.291734	-	-	-	-0.156084
		NS	-1.331015	0.282488	-0.006506	-	-0.203953
	18263	S	0.248132	-	-	-	-0.083597
		NS	-1.304679	-0.100424	0.065049	-	-0.130211
cyclic	3653	S	0.335565	-	-	-	-0.138367
		NS	-1.131935	-0.053628	-0.114716	0.173038	-0.171678
	7305	S	0.266744	-	-	-	0.017345
		NS	-1.304569	0.081954	0.071466	0.127846	-0.040992
	18263	S	0.243336	-	-	-	-0.000119
		NS	-1.403703	0.006401	0.030406	0.080007	-0.023372

tests which are Anderson-Darling (AD) test and Cramer von Mises (CVM) test. The p-values of the tests are shown in Table 7 for $\xi = 0.4$ and Table 8 for $\xi = -0.4$. According to the results, most of the data sets simulated with positive shape parameter, $\xi = 0.4$, shows that the p-values for CVM test is more than 0.05. However, for data sets with negative shape parameter, $\xi = -0.4$, only a few of the simulated data sets give a good fit for stationary GP model. Nonetheless, it would not be a problem, since extreme data sets with negative shape parameter are rarely encountered in real world. Hence, we can conclude that the stationary GP model is suitable in modelling the tree-based threshold exceedances.

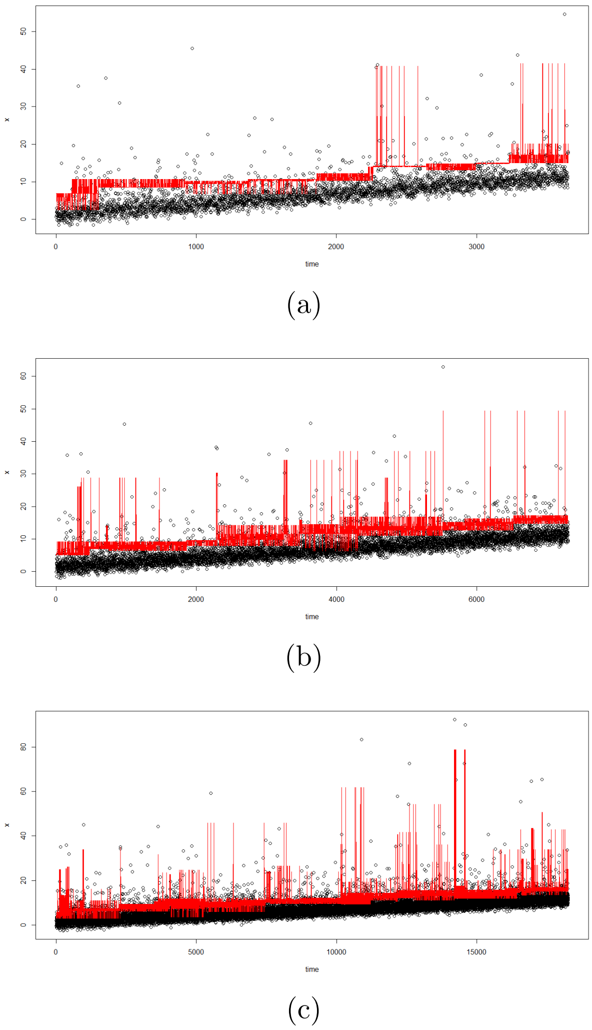
The proposed method which is the tree-based threshold selection also has been compared with the standard method which the exceedances of a constant high threshold are modelled by nonstationary GP model with covariate model incorporated in the parameter. In



(a)

b)

(c)



(a)

(b)

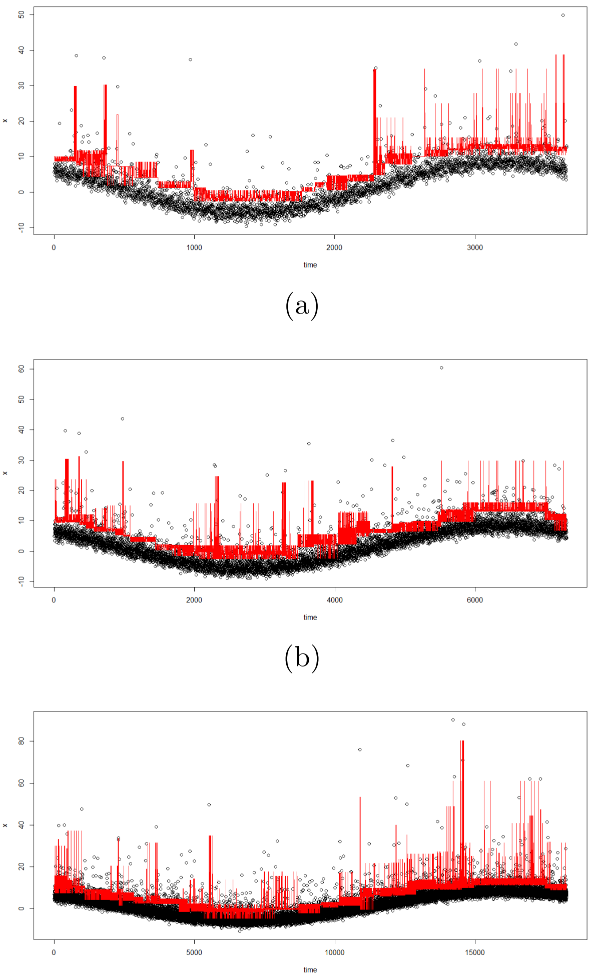
(c)

Figure 4: Regression tree for $\xi = -0.4$ and cyclic trend with (a) $n = 3653$, (b) $n = 7305$, (c) $n = 18263$

Figure 5: Tree-based threshold for $\xi = 0.4$ and linear trend with (a) $n = 3653$, (b) $n = 7305$, (c) $n = 18263$

this study, for each data set, a constant high threshold is fixed at 95th percentile of the data. The covariate models incorporated in the GP scale parameter are same as Equation 1 and Equation 2. The comparison between both methods are done by comparing the AIC and BIC values which determine the information loss due to the model estimating. The results are shown in Table 9 for $\xi = 0.4$ and in Table 10 for $\xi = -0.4$.

Based on the values of the AIC and BIC, it can be observed that there are not much difference between tree-based threshold method and standard method. In fact, for data sets with negative shape parameter, $\xi = -0.4$, the values of AIC and BIC for tree-based threshold method are smaller compared to the standard method, hence, it ca be concluded that the tree-based threshold selection method is better in terms of modelling due to less information loss.

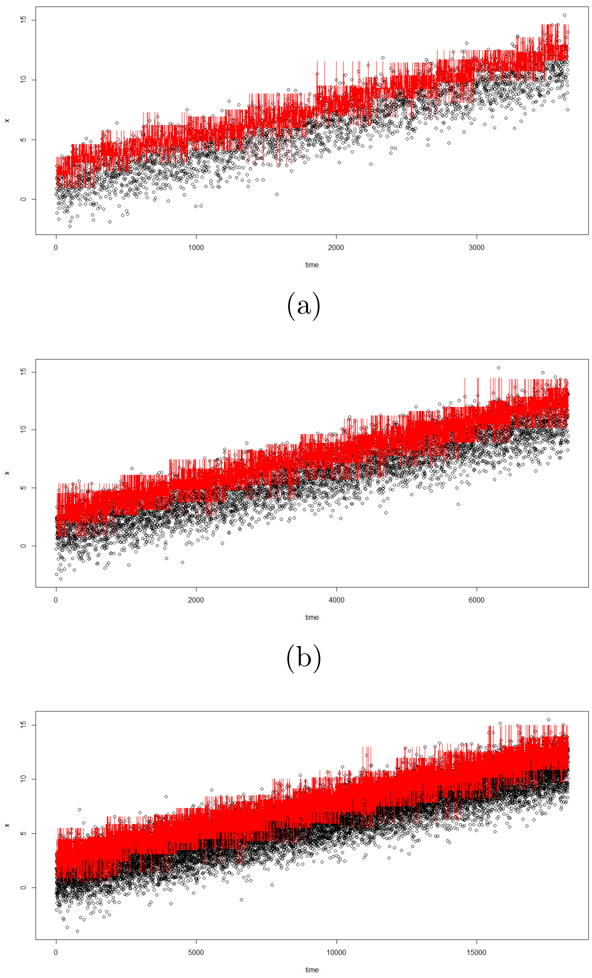


(a)

(b)

(c)

Figure 6: Tree-based threshold for $\xi = 0.4$ and cyclic trend with (a) $n = 3653$, (b) $n = 7305$, (c) $n = 18263$

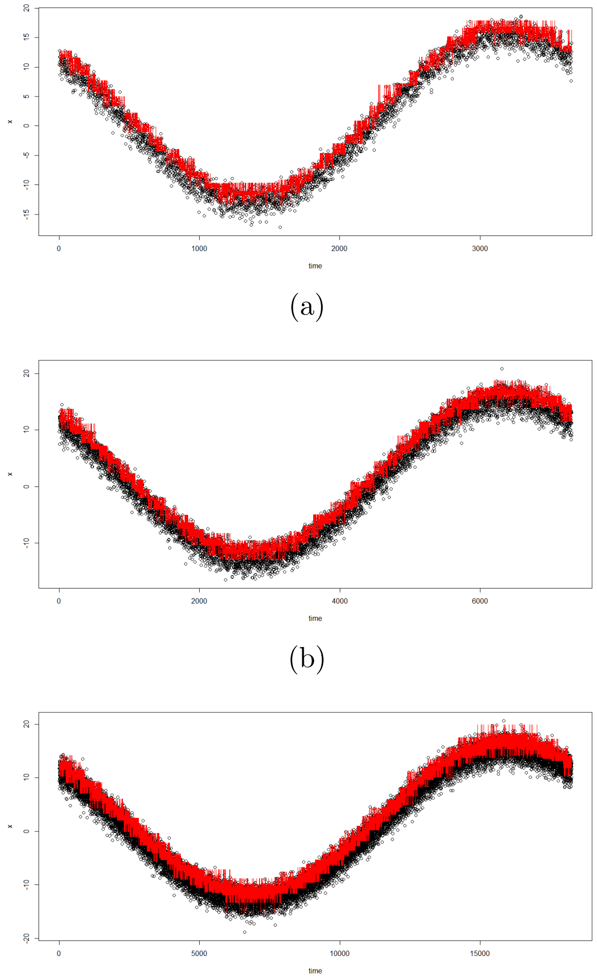


(a)

(b)

(c)

Figure 7: Tree-based threshold for $\xi = -0.4$ and linear trend with (a) $n = 3653$, (b) $n = 7305$, (c) $n = 18263$



(a)

(b)

(c)

Figure 8: Tree-based threshold for $\xi = -0.4$ and cyclic trend with (a) $n = 3653$, (b) $n = 7305$, (c) $n = 18263$

Table 5: The AIC and BIC values of stationary GP model (S) and non-stationary GP model (NS) for $\xi = 0.4$

covariate model	n	model	AIC	BIC
linear	3653	S	909.778	916.293
		NS	912.913	925.943
	7305	S	1790.925	1798.847
		NS	1794.751	1810.595
	18263	S	4497.751	4507.502
		NS	4496.108	4515.610
cyclic	3653	S	849.858	856.495
		NS	852.049	868.640
	7305	S	1750.734	1758.692
		NS	1754.786	1774.680
	18263	S	4460.792	4470.567
		NS	4466.708	4491.146

Table 6: The AIC and BIC values of stationary GP model (S) and non-stationary GP model (NS) for $\xi = -0.4$

covariate model	n	model	AIC	BIC
linear	3653	S	-147.016	-140.341
		NS	-145.123	-131.773
	7305	S	-314.145	-306.112
		NS	-313.859	-297.794
	18263	S	-945.291	-935.488
		NS	-951.379	-931.772
cyclic	3653	S	-102.817	-95.923
		NS	-110.466	-93.232
	7305	S	-268.262	-260.053
		NS	-276.264	-255.740
	18263	S	-879.217	-869.270
		NS	-883.743	-858.875

Table 7: P-values for Anderson-Darling (AD) and Cramer von Mises (CVM) goodness of fit tests for stationary GP model fitted to exceedances of tree-based threshold with $\xi = 0.4$

covariate model	n	AD test p-value	CVM test p-value
linear	3653	0.6795	0.7937
	7305	0.1232	0.3285
	18263	0.0095	0.1220
cyclic	3653	0.6643	0.7484
	7305	0.1007	0.2412
	18263	0.0023	0.0337

Table 8: P-values for Anderson-Darling (AD) and Cramer von Mises (CVM) goodness of fit tests for stationary GP model fitted to exceedances of tree-based threshold with $\xi = -0.4$

covariate model	n	AD test p-value	CVM test p-value
linear	3653	0.0261	0.0943
	7305	0.0043	0.0286
	18263	0.0066	0.0597
cyclic	3653	0.0063	0.0561
	7305	0.2095	0.3843
	18263	0.0010	0.0295

Table 9: The AIC and BIC values of tree-based threshold selection method (TB) and standard method (S) for $\xi = 0.4$

covariate model	n	model	AIC	BIC
linear	3653	TB	909.778	916.293
		S	885.880	898.718
	7305	TB	1790.925	1798.847
		S	1740.376	1755.986
	18263	TB	4497.751	4507.502
		S	4408.343	4427.15
cyclic	3653	TB	849.858	856.495
		S	844.380	860.427
	7305	TB	1750.734	1758.692
		S	1680.999	1700.512
	18263	TB	4460.792	4470.567
		S	4298.972	4323.061

Table 10: The AIC and BIC values of tree-based threshold selection method (TB) and standard method (S) for $\xi = -0.4$

covariate model	n	model	AIC	BIC
linear	3653	S	-147.016	-140.341
		NS	209.707	222.545
	7305	S	-314.145	-306.112
		NS	380.146	395.756
	18263	S	-945.291	-935.488
		NS	1030.396	1049.667
cyclic	3653	S	-102.817	-95.923
		NS	236.056	252.104
	7305	S	-268.262	-260.053
		NS	445.062	464.575
	18263	S	-879.217	-869.270
		NS	1110.373	1134.462

V. Conclusion

Non-stationary extremes modelling usually use constant high threshold to determine the extremes. However, with non-stationary data which is affected by covariates, a constant threshold might be an issue which possibly violate the GP model assumption. In this paper, a covariate-varying tree-based threshold has been proposed. The exceedances of the tree-based threshold are modelled by both stationary and non-stationary GP model. Results from AIC and BIC values show that stationary GP model fit the exceedances better. Hence, by applying the aforementioned method we have a simpler model to fit the data. Comparison has been made between the tree-based method and the usual constant high threshold method. The AIC and BIC values show that the tree-based threshold selection method is comparable to the standard method in terms of information loss.

With regards to future study, it is suggested that a method to select threshold within the stationary clusters in regression tree instead of using the 95th percentile value. Smoothing techniques can also can be applied on the tree-based threshold to reduce the roughness of the threshold.

References

[1] Brian Bader, Jun Yan, Xuebin Zhang, et al. Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *The Annals of Applied Statistics*, 12(1):310–329, 2018.

[2] Stuart Coles, Joanna Bawa, Lesley Trenner, and Pat Dorazio. *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.

[3] MB de Carvalho, K Feridum Turkman, António Rua, et al. Nonstationary extremes and the us business cycle, 2012.

[4] Emma F Eastoe and Jonathan A Tawn. Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(1):25–45, 2009.

[5] Wei-Yin Loh. Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1):14–23, 2011.

[6] Paul J Northrop and Philip Jonathan. Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*, 22(7):799–809, 2011.

[7] James Pickands III et al. Statistical inference using extreme order statistics. *the Annals of Statistics*, 3(1):119–131, 1975.

[8] Afif Shihabuddin, Norhaslinda Ali, and Mohd Bakri Adam. Simulation study on tree-based threshold in generalized pareto model with the presence of covariate. In *AIP Conference Proceedings*, volume 1974, page 040003. AIP Publishing, 2018.

[9] Caston Sigauke and Alphonce Bere. Modelling non-stationary time series using a peaks over threshold distribution with time varying covariates and threshold: An application to peak electricity demand. *Energy*, 119:152–166, 2017.

[10] Paul Thompson, Yuzhi Cai, Dominic Reeve, and Julian Stander. Automated threshold selection methods for extreme wave analysis. *Coastal Engineering*, 56 (10):1013–1021, 2009.

[11] Weiwen Wang, Wen Zhou, Edward Yan Yung Ng, and Yong Xu. Urban heat islands in hong kong: statistical modeling and trend detection. *Natural Hazards*, 83 (2):885–907, 2016.

[12] Xia Yang, Jing Zhang, and Wei-Xin Ren. Threshold selection for extreme value estimation of vehicle load effect on bridges. *International Journal of Distributed Sensor Networks*, 14(2):1550147718757698, 2018.