Soheil Saeed Far
Universiti Teknologi Malaysia
+60 (17) 826 8048
soheilsaeedfar@gmail.com

Senior Scientist Andreas Sterl
Topic Editor,
Email: andreas.sterl@knmi.nl

August 21, 2016

Dear Dr. Sterl,

I am writing to inform you that we provided our point-by-point responses to the comments raised by the reviewer #2 dated August 15, 2016 for the following manuscript:

Journal: OS
Title: Evaluation of Peaks-Over-Threshold Method
Author(s): S. Saeed Far and A. K. Abd. Wahab
MS No.: os-2016-47
MS Type: Research article

C: Comment & **R:** Response

1. **C: 1.1. The authors claim that the GPD and the POT are two different models.**

   R: We do not claim that the GPD and POT (Goda) are different models. They are different models. The POT model was introduced by Goda (1988) and then was developed in a textbook (Goda, 2000). The new edition of the book was published in 2010 (Goda, 2010). This author was referred 23^{rd} times throughout our manuscript and the first 7 references in the list of references have been allocated to the Goda's publications. For example in page 6, line 17, title 2.2, Peaks-Over-Threshold Model introduced as Goda method (this is where the methodology of POT was described), or in the line of keywords (Pg. 1, line 12), the POT model accompanied by the name of Goda in parentheses. Therefore, the manuscript has left no any gap of explanation or any excuse to misunderstanding between the POT (Goda) model and other models.

   C: 1.2. The GPD plays a crucial role in extreme value theory as the distribution of the sample of excesses above a sufficiently high threshold, method known as the peaks-over-threshold (POT) -see, for instance, Davison and Smith (1990), Pickands (1975), Embrechts et al (1997).

   R: The publications of Pickands, Davison and Smith have been cited in our manuscript, and we used their ideas. We were aware about the similarity of the names (POT), that is

why many times the name of Goda were noted beside the POT model.

C: 1.3. The POT and the GPD are thus closely connected.

R: The POT (Goda) and GPD models use excesses data over a sufficiently high threshold value. GPD is an asymptotic model introduced by Picknads (1975), apart from this, the POT model for the first time introduced by Goda (1988). [All these information have been discussed and noted throughout the manuscript.]

C: 1.4. I do not understand the meaning of "to select the best fitting distribution for the dataset" (page 1, line 23).

Based on definitions, in the extreme data analysis, many theoretical distribution functions are employed for fitting to samples. In theoretical statistics, a data of extremes refers to the maximum or minimum among a sample of independent data. When extreme analysis is applied for a sample of such extreme data, it is known that three types of theoretical functions should fit such samples, depending on the population distribution of initial data. However, the data of extreme wave heights collected by POT are different from the extreme data of theoretical statistics (more details, Goda [Pg. 377](2000) Chapter 11, Statistical Analysis of Extreme Waves).

C: 1.5. The issue is to fit a GPD distribution to the excess (or exceedance) data. Obviously, the Generalized Extreme Value (GEV) or the Gumbel, Fréchet and Weibull are candidate distributions if the Annual Maxima Method is considered. For using this method the data has to exhibit a temporal structure (and obviously there is no threshold involved).

R: In the manuscript, we have not employed the GEV model. The reason of using the three **FT-I**, **FT-II** and **Weibull** distribution functions is to conduct the POT model, and figure out the best fitting distribution function for the dataset.

C: 1.6. (Some references of using the POT in waves context - Teena et al (2012), Thevasiyani and Perera (2014) and Cañellas (2007))

R: There are also many references of using the POT (Goda) model such as Li et al. (2012), Goda (2010) and Goda (2004).

C: 2.1. The authors, in page 6, present the hybrid method that starts with the choice of the threshold through the mean excess (ME) plot. In applications, the choice of the threshold just by looking at the ME plot is generally very difficult if not impossible (in some cases). Due to the practical difficulties in choosing the threshold, and considering the importance of this step in the method, some complementary approaches could have been mentioned (see, for instance, Beirlant et al (2004)).

R: In the manuscript, the Hybrid method was proposed to determine true threshold value for the POT model. The method consists of two parts: the first part is using the Mean

Residual Life (MRL) plot, and the second part, which is described as the complementary method can be found in page 7 line 9 to 17, and step 10 in page 10. Therefore, the Hybrid method consists of **exploratory** and **complementary** methods.
The discussion of credibility of extracting true results via the MRL plot can be found in several literatures for instance Coles (2001) discussed about the interpretation of the graph, and noted that, "The interpretation of a mean residual life plot is not always simple in practice." To my knowledge, the use of MRL and extracting true result from the plot is intuitively understandable if the definition of deviation from mean in the population has already been understood.

C: 2.2. In page 7-line 11, the authors say that Weibull distributions were fitted to the wave data. Later on, the same is done for the Gumbel and for the Fréchet distributions. Due to the fact that ". . . the exceeded data over a certain threshold are employed in the analysis" (page 9, line 10) then the GPD should have been used instead because it is the proper distribution for modelling the excesses above the threshold, as was stated in my previous comment.

R: That is the explanation of step 8, which describes using the excess data to determine the plotting position for the POT model. The methodology of POT initiates from page 6, section 2.2 under the title of, "Peaks-Over-Threshold Model (Goda Method)" to page 11, section 3, entitled, "Results and Discussions".

C: The data lacks a convenient explanation. The authors seem to consider wave heights recorded at time intervals of at least three days apart (page 2, lines 28-29). Then, the authors state that they do linear interpolation "to fill in the large gaps" (line 29, page 2). I wonder why that is needed. The exceedances are supposed to be the focus. Additionally, it would have been nice to see the plot of the data which was analysed.

R: The process of choosing data in the time intervals of three days were done on the data to secure the required independence of the data. It has been done after data collection and prior to the determination of threshold value. After the process of data collection, usually some gaps of missing data for several days or more happen due to human error or instrument malfunction. Those gaps were filled by linear interpolation. This process is done before exerting the time intervals (three days) on the data.
Based on the regulation of using the data, we are not allowed to distribute or publish the data. However, the complete data analysis have been done, and this is not the first evaluation on this data. If any specific graph is required, which its distribution does not violate the regulations, please let me know, then I will gladly provide the arrangements to get permission to send it for you.

2. **Technical correction**
   C: 3. Expression (4) is not consistent with the parameterization of the GPD

the authors indicated in (1) – see Coles (2001).

R: Yes, it happened during the process of typesetting. The correct expression is,

$$x_m = u - \frac{\sigma}{\xi}\,[(m \cdot\ \zeta_u)^{-\xi} - 1] \tag{4}$$

And, then equation 7,

$$x_m = u - \frac{\sigma}{\xi}\,[(ARI.\ \frac{N}{k})^{-\xi} - 1] \qquad for\ (\xi \neq 0) \tag{7}$$

We appreciate your time taken to review the responses. Lot of thanks for your consideration and care.

Yours Sincerely,

Soheil Saeed Far

Goda, Y. (2010). Random seas and design of maritime structures. World scientific.

Goda, Y. (2000). Random seas and design of maritime structures. World scientific.

Goda, Y. (1988). On the methodology of selecting design wave height. Coastal Engineering Proceedings, 1(21).

Coles S. (2001). An introduction to statistical modeling of extreme values. London: Springer Series in Statistics;

Li, F., Bicknell, C., Lowry, R., & Li, Y. (2012). A comparison of extreme wave analysis methods with 1994-2010 offshore Perth dataset. Coastal Engineering, 69, 1-11.

Pickands J. (1975). Statistical inference using extreme order statistics. the Annals of Statistics. 1:119-31.

Goda, Y. (2004). Spread parameter of extreme wave height distribution for performance-based design of maritime structures. Journal of waterway, port, coastal, and ocean engineering, 130(1), 29-38. URL: http://ascelibrary.org/doi/abs/10.1061/(ASCE)0733-950X(2004)130:1(29)