Closure of discussion: Point-in-time and Extreme-Value Probability Simulation Technique for Engineering Design

R. Mínguez, Y. Guanche, F. J. Méndez

Environmental Hydraulics Institute "IH Cantabria", Universidad de Cantabria, Cantabria Campus Internacional, Spain

We would like to thank Dr. Solari and Prof. Losada for their comments and discussion [1] about our work [2]. Certainly, they raised very interesting points, which in our opinion, contribute to clarify the link between the point-in-time and the extreme value analysis for practitioners.

After careful reviewed of their comments, our conclusion is that they are right about the limitations of our proposed method to deal with the temporal dependence structure of the stochastic process if the following two conditions hold: i) the autocorrelation is positive, which unfortunately for us, occurs for most of the environmental variables in practice, and ii) annual maxima is used to deal with extremes. For this reason, we recognize those limitations and clarify those cases where the method is still valid and applicable.

On the other hand, we disagree about their concerns regarding the necessity of including the extremal index θ within the graphical representation trough equation:

$$T_r^{\mathrm{EV}} = \frac{n\theta}{1 - F^{\mathrm{EV}}(x)}, \tag{1}$$

where F^{EV} is the extreme-value probability distribution, and n is the number of hours within a year ($n = 8766$) in case of hourly values and annual maxima, because this way of including the extremal index distorts the extreme value analysis.

In the following sections, we will justify our arguments.

Email address: `roberto.minguez@unican.es` (R. Mínguez)

1. Discussion regarding the join graphical representation and the mixture distribution

As authors [1] claim, return periods T^{PT} and T_r^{EV}, in spite of having the same units, are not directly comparable except when data are independent. Thus, only T_r^{EV} truly correspond to return periods because they are obtained using any distribution for maxima, i.e. annual maxima, Pareto-Poisson, or the Peaks Over Threshold (POT) method, which ensure independency for the fitting process. For this reason, we called T^{PT} as "equivalent return period" instead of return period, and that is the reason why empirical points in the graphical representation do not coincide but for large values of x, i.e. $x \to \infty$. Note that using mass-stability, it can be proved that for large values of x then $T^{\text{PT}}(x) \approx T_r^{\text{EV}}(x)$. This behavior is observed in Figure 1 from [1], where points associated with the largest value of H_{m0}: i) the black dot (point-in-time), ii) red dot (annual maxima using (1b) in [1]) and iii) red square dot (peak over threshold using (1b) in [1]) are almost coincident.

The aim of the method proposed in [2] is to use the point-in-time and extremal fits without perturbing the extreme value analysis. We are aware that there is a portion of the distribution which is not appropriately reproduced, as shown in Figure 1 comparing the red and blue dashed lines, but we make sure that the upper tail associated with extremes is correct and consistent with state-of-the-art analysis of extremes, which is the most relevant issue from the engineering perspective. This is the reason why the method is still valid in case we are just interested on the marginal distribution of the variable under study, and thus it is applicable within first order reliability methods (FORM), as shown in [3].

Note that the question raised by [1] about including extremal index using (1) is very interesting because it allows explaining the differences between T_r^{EV} and T^{PT} (see the horizontal distance between T_r^{EV} and T^{PT} in Figure 1), and we fully agree with the authors about expression $T^{\text{PT}} \approx \theta T_r^{\text{EV}}$. Note that this expression allows moving the return period T_r^{EV} associated with independent data from right to left to make it coincident with T^{PT} (see Figure 1), which is precisely what authors [1] do in their example. However, we have two concerns about their proposal in [1]:

1. The resulting data set no longer correspond to return periods. Since state-of-the-art extreme value analysis is used to define F_{EV}, then $\dfrac{n}{1-F^{\text{EV}}(x)}$ is indeed the true return period expressed in hours instead

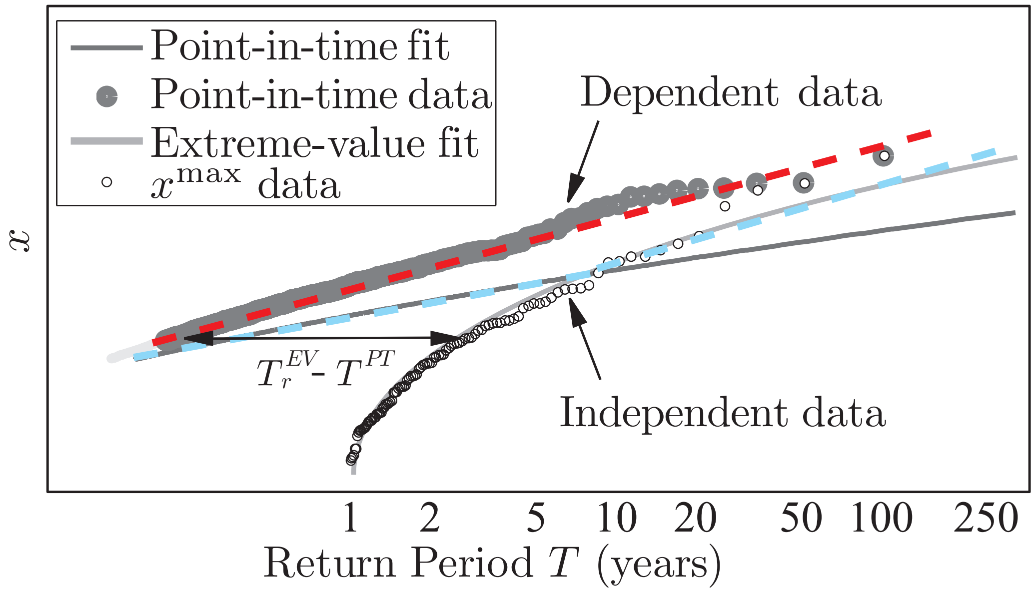

Figure 1: Graphical representation of the point-in-time and extreme mixture distribution proposed by [2].

of years. If we multiply this value by the extremal index, the “equivalent return period” associated with dependent data T^{PT} is obtained, which is not a return period.

2. It is also questionable to use a constant extremal index. As shown in Figure 1, the differences between T^{PT} and T_r^{EV} decrease when data increase. This non constant evolution of the extremal index is pointed out by the discussion of the paper [4] by Dr. Jonathan Tawn.

According to this reasoning, the blue dots in Figure 1 from [1] do not correspond to return periods associated with annual maxima. Accordingly, the fitted distribution to these maxima (blue line) does not provide return period estimates.

2. Discussion regarding the join graphical representation and the mixture distribution

Regarding comments about the autocorrelation structure, authors [1] are right about the limitation of our proposal to work appropriately for those cases where the temporal autocorrelation is positive and annual maxima are used for extreme value analysis.

We also agree with authors [1] that it is more appropriate and accurate to use transformation:

$$\begin{array}{rcll} \Phi(z) & = & F^{\mathrm{PT}}(x) & \text{if} \quad x \le x_{\mathrm{lim}} \text{ or } z \le z_{\mathrm{lim}} \\ p_{\mathrm{lim}}^{\mathrm{EV}} + \dfrac{\Phi(z) - p_{\mathrm{lim}}^{\mathrm{PT}}}{1 - p_{\mathrm{lim}}^{\mathrm{PT}}}(1 - p_{\mathrm{lim}}^{\mathrm{EV}}) & = & F^{\mathrm{EV}}(x) & \text{if} \quad x > x_{\mathrm{lim}} \text{ or } z > z_{\mathrm{lim}}, \end{array} \tag{2}$$

instead of

$$\Phi(z) = F^{\mathrm{PT}}(x). \tag{3}$$

Finally, there is a particular case in practice where the method proposed by [2] is still applicable even for positive autocorrelated stochastic processes. According to [5], univariate extreme value theory provides an asymptotic justification for the generalized Pareto distribution to be an appropriate model for the distribution of excesses over a suitably chosen high threshold u_{lim}. If F^{EV} corresponds to Pareto distribution, and x_{lim} is equal to the selected threshold u_{lim}, the mixture model proposed by [2] becomes:

$$\begin{array}{ll} F^{\mathrm{PT}}(x) & \text{if} \quad x \le u_{\mathrm{lim}} \\ p_{\mathrm{lim}}^{\mathrm{PT}} + (1 - p_{\mathrm{lim}}^{\mathrm{PT}}) F^{\mathrm{EV}}(x) & \text{if} \quad x > u_{\mathrm{lim}}, \end{array} \tag{4}$$

where $p_{\mathrm{lim}}^{\mathrm{EV}} = F^{\mathrm{EV}}(u_{\mathrm{lim}}) = 0$. In case F^{PT} corresponds to the empirical distribution function, this model is the same as the semiparametric model used by several authors [6, 7, 8] for multivariate extreme value analysis.

To show the performance of the method, we have run the simulation process using the same example of significant wave height at Bilbao given in [2], however, instead of using the POT model for annual maxima (light gray dashed line in both panels from Figure 2), we use Pareto distribution for exceedances over the threshold $u_{\mathrm{lim}} = 4.8$ (red dashed line in both panels from Figure 2)). Using transformation (2), as proposed by [1], we fit an ARMA$(3, 2)$ model to reproduce the temporal dependence structure of the stochastic process. We sample $n_s = 100 \times 24 \times 362.25$ hourly values using (4) method proposed in [2], and results from the simulation process are shown in the lower panel of Figure 2.

Note that the hourly sample and their corresponding annual maxima present good agreement with respect to Pareto and the annual maxima distributions, respectively, with all sample points within the 95% confidence

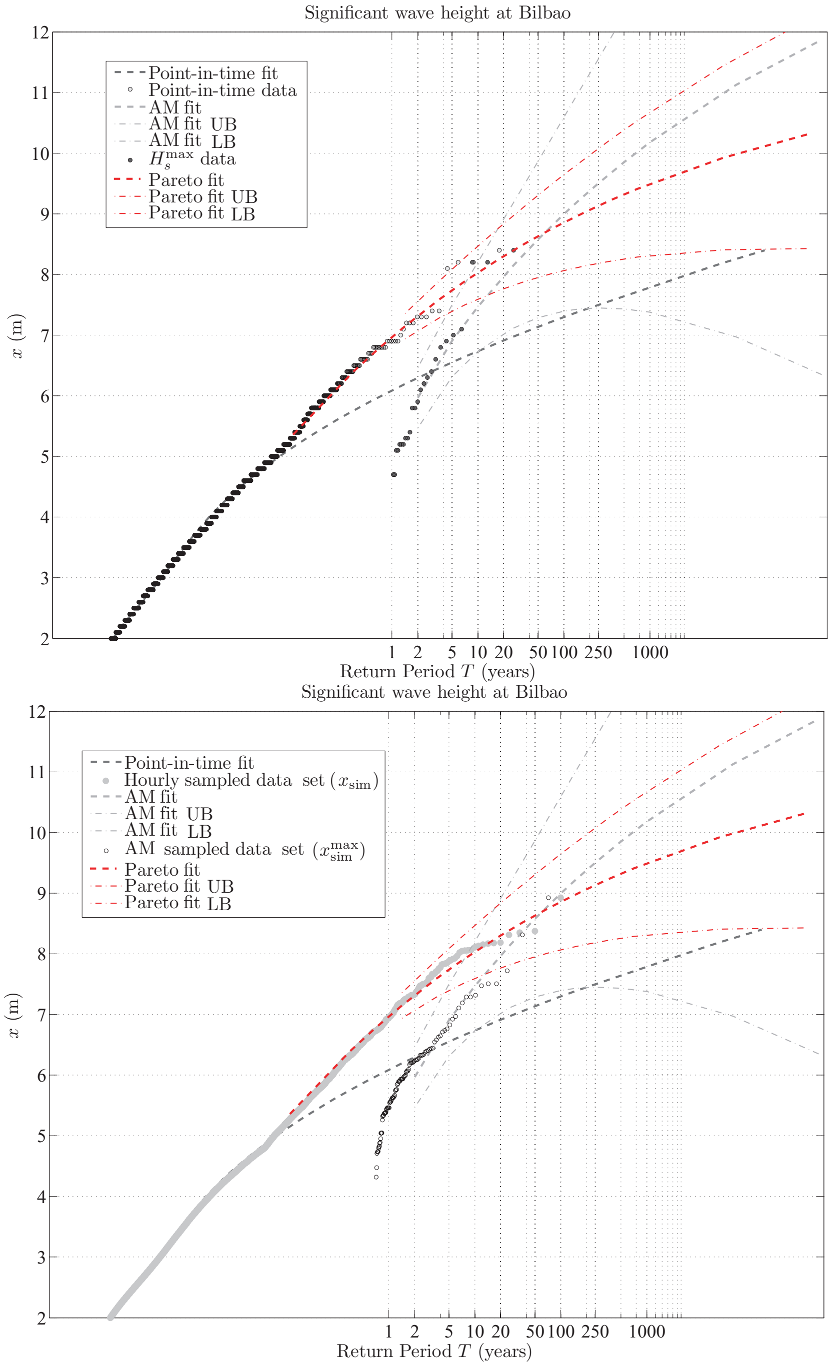

Figure 2: Graphical representation of the point-in-time and extreme fitting distributions for the significant wave height at Bilbao buoy location, and simulation process considering the positive autocorrelation structure.

bands. In addition the temporal dependence structure of the stochastic process is appropriately reproduced.

[1] S. Solari, M. A. Losada, Comments on point-in-time and extreme-value probability simulation technique for engineering design", Structural Safety ? (2013) ?–? doi:10.1016/j.strusafe.?

[2] R. Mínguez, Y. Guanche, F. J. Méndez, Point-in-time and extreme-value probability simulation technique for engineering design, Structural Safety 41 (2013) 29–36. doi:10.1016/j.strusafe.2012.10.002.

[3] R. Mínguez, Y. Guanche, F. F. Jaime, F. J. Méndez, A. Tomás, Filling the gap between point-in-time and extreme value distributions, in: Proceedings of the 11th International Conference on Structural Safety & Reliability (ICOSSAR 2013), New York, NY, 2013.

[4] A. C. Davidson, R. L. Smith, Models for exceedances over high thresholds, Journal of the Royal Statistical Society. Series B (Methodological) 52 (3) (1990) 393–442.

[5] J. Pickands, Statistical inference using extreme order statistics, Annals of Statistics 3 (1975) 119–131.

[6] S. G. Coles, J. A. Tawn, Modelling extreme multivariate events, J. R. Statist. Soc. B 53 (1991) 377–392.

[7] S. G. Coles, J. A. Tawn, Statistical methods for multivariate extremes: an application to structural design (with discussion), Appl. Statist. 43 (1994) 1–48.

[8] J. E. Heffernan, J. A. Tawn, A conditional approach for multivariate extreme values (with discussion), Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66 (3) (2004) 497–546. doi:10.1111/j.1467-9868.2004.02050.x.