Lectures on Modelling and Simulation; A selection from AMSE # 2017-N°2; pp 1–10

Selection at the AMSE Conference Calcutta/India, November 4-5, 2017

Statistical Aggregation of Wind Hazards Models

* Subrata Bera, **Dhanesh B. Nagrale, +U. K. Paul, ‡D. Datta

*Nuclear Safety Analysis Division (NSAD), Atomic Energy Regulatory Board (AERB) India, Anushaktinagar, Mumbai-400094, (sbera@aerb.gov.in)

**NSAD, AERB, India, Anushaktinagar, Mumbai-400094, (dbnagrale@aerb.gov.in)

+ NSAD, AERB, India, Anushaktinagar, Mumbai-400094, (ukp@aerb.gov.in)

‡Radiological Physics and Advisory Division, Bhabha Atomic Research Centre

India, Trombay, Mumbai-400085, (ddatta@barc.gov.in)

Abstract

Extreme value of the wind speed over the block of time period of one year in a given historical data set for a specific site are extracted for developing wind hazards model. In the hazard model development, year wise wind speed extrema are fitted with three parameters continuous generalized extreme value (GEV) distribution. Size of the sample space determines the variation of model parameters. Wind hazards model uncertainty has been assessed using the sampling based methodology. At a new site, there may not be available wind speed measurement station. In that case, extreme value analysis can be done using historical data provided by the stations near to the site. A methodology for statistical aggregation of multiple models is developed and demonstrated considering data from four measuring stations near to a site. In this statistical aggregation method, the statistical property of the GEV model has been preserved. The variation return period of wind speed due to individual models are compared with that of the statistically aggregated model.

Key words

Extreme value analysis, generalised extreme value distribution, empirical distribution, model uncertainty, statistical aggregation, external hazard, wind speed, wind load.

1. Introduction

Extreme value analysis[1] of collected historical data of meteorological, seismic events are important for designing the engineering structures, system and component robust enough to withstand external hazards such as wind load, flood level, earthquake etc. The probability and statistics [2, 3] are in extensive use in the design of engineering structures and safety assessment. Flood level, tsunami height, wind speed are beyond control of the human being. The behaviours of these variables are random in nature. Designing of a nuclear facility, safety important structure like cooling tower, stack, etc., requires prior knowledge about the variation of these variables more specific to a given site. Designer may want to know the extreme value such as minimum or maximum value of the hazard parameters to justify the design basis of the engineering structure. The design basis flood level, tsunami height, and wind load to the tall structure such as chimney and industrial stack are few examples where extreme value is important in design [4]. These extreme values do not follow the central limit theory (e.g. normal distribution) [5]. When the minimum value is of interest, the distribution is skewed toward lower values. Similarly, the distribution skewed towards higher value in case when the maximum value is the desired quantity. There are two approaches to estimate the extreme value: (1) block maxima approach [6] and (2) peak over threshold approach. There are many extreme value distribution functions such as Gumbel distribution, Weibull distribution and Generalised Extreme Value (GEV) distribution function, etc. [7, 8] used in block maxima approach. However, generalised Pareto distribution is used in peak over threshold approach [9, 10]. These distributions are utilised for estimation of extreme values. Most of the extreme value analyses are carried out using the generalised extreme value distribution[11].

In the present analytical study, wind speed is considered to be an extreme value variable. For the extreme value analysis, wind speed data for several years are required for the analysis. For a new site, such a detailed data based with measured meteorological parameter may not available. In that case, measurement in the nearest meteorological stations to the new site can be used. In the study reported here the year wise maximum wind speed data are fitted with generalised extreme value (GEV) distribution function. With known fitted parameters, the extreme value analysis model is generated to extrapolate the extreme value for the desired return period.

When a mathematical model for a physical system is developed, response of the model may not exactly replicate the actual system. The deviation of model response from actual system may be due to the error associated with input arguments or model parameter. (S. Bera et al. MS-17-58, 2017) has demonstrated the analytical input error propagation methodology for a double exponential function [12]. Estimation of model parameters through curve fitting leads to the uncertainty associated with the sample size. Low sample size data leads to increase the uncertainty in model parameters. In this study, estimation of model uncertainty has been emphasised. Probabilistic method based on random sampling from the desired distribution has been used to estimate the model uncertainty.

Statistical aggregation methodology developed by (S. Bera et al., MS-17-65, 2017)[13] for multiple stations has been used to obtain the average GEV model for return period estimation. In this case study, wind speed data collected over several years in nearest four stations as reported in (S. Bera et al., MS-17-65, 2017)[13] are considered for extreme value analysis.

2. Theoretical Methodology

2.1. Generalized Extreme Value Analysis model

The three parameter distribution function of the standard GEV is given in Equation (1).

$$F_{k,m,s}(x) = \begin{cases} exp\left(-\left(1+k\frac{x-m}{s}\right)^{-\frac{1}{k}}\right) & if\ k \neq 0 \\ exp\left(-\frac{x-m}{s}\right) & if\ k = 0 \end{cases} \tag{1}$$

where, 'k', 'm', 's' are known as all-important shape parameter, location parameter, scale parameter respectively. All-important shape parameter determines the nature of the tail distribution. The extreme value distribution in Equation (1) is generalised in the sense that parametric form subsumes three types of distributions which are known by other names according to the value of 'k'. If the data related to the year wise maximum value of wind speed is obtained for 'n' number of years, then distribution function F(x) will be calculated based on the empirical density estimation methodology.

2.2. Concept of Mean Return Period

Meteorological variables such as cyclonic wind speeds are non-deterministic values, as they randomly vary with time and hence, the probability of occurrence of a wind speed exceeding a given value has to be studied through the concept of mean return period. A characteristic wind speed can be defined as an order pair consisting of wind speed and associated the probability of occurrence. The mean return period of a characteristic wind speed, 'x', denoted as T(x) is defined as the mean time interval between two successive values which are greater than the characteristic wind speed, x. Hence, the probability of experiencing a characteristic maximum wind speed, 'X', greater than 'x', in any one year is equal to 1/T(x). In the mathematical notation, it can be written as given in Equation (2).

$$P_{k,m,s}(X > x) = \frac{1}{T_{k,m,s}(x)} \tag{2}$$

By definition of cumulative distribution function, F(x), can be written as given in Equation (3).

$$F_{k,m,s}(x) = P_{k,m,s}(X \leq x) = 1 - P_{k,m,s}(X > x) \tag{3}$$

Then $T_{k,m,s}(x)$ can be written in terms of $F_{k,m,s}(x)$ as given in Equation (4).

$$T_{k,m,s}(x) = \frac{1}{1 - F_{k,m,s}(x)} \tag{4}$$

Once the data are fitted with the extreme value model, ordered pair of variable 'x' can be generated. Equation (4) will be used to obtain return period.

2.3. Empirical Statistics

The empirical statistics is based on non-parametric distribution. This is an approximation of a population density function that is derived from a sample and has no unknown parameters is the empirical density. This is basically an order statistics [14, 15]. In order statistics, the random variables $(i.e. X = \{x_1, x_2, \dots, x_n\})$ are arranged in the ascending order. Each event in the sample space is given equal probability (i.e. equal to $\frac{1}{n}$). The empirical distribution is formed based on cumulative probability assignment.

2.4. Estimation of Model Parameters

Mathematical model of a physical system or process can be represented in an equation or set of equations. In a mathematical model, there will be some parameters apart from the arguments and responses. These parameters are called model parameters. Model parameters are being estimated using the argument and response data fitting with the desired model. From the station specific measured data, empirical distribution is calculated using Equation (5). This distribution is used to develop the GEV model by fitting the Equation (1). Each GEV model will have specific set of {k, m, s} values. There are various methods to estimate the model parameters such as moment method, list square fitting method, maximum likelihood estimation method and Bayesian estimation method [9]. As the model parameters control the response of the model, the sensitivity of the model parameter on the model response has to be studied. For that purpose, it is required to know the mean value of model parameters along with its variation.

2.5. Estimation of the Model Uncertainty

The dispersion of the model parameter will depend on the sample data size. Small sample space may lead to the large uncertainty whereas large data reduces the model parameter uncertainty. The methodology for assessing the model uncertainty due to the model parameter variation of a GEV distribution has been described below.

The GEV model is three parameter (i.e., k, m, s) continuous distribution function. While fitting the Equation (1) with measured data, it is obtained the mean value of three parameters along with their standard deviation. Model uncertainty is the estimation error due to the variation of the model parameters (i.e., k, m, s) not for the input parameter such as wind speed. Probabilistic method has been used to estimate the model uncertainty. In this method, random sample has been taken uncertainty domain of three parameters with normal distribution. The uncertainty analysis methodology has been shown in the Fig. 1.

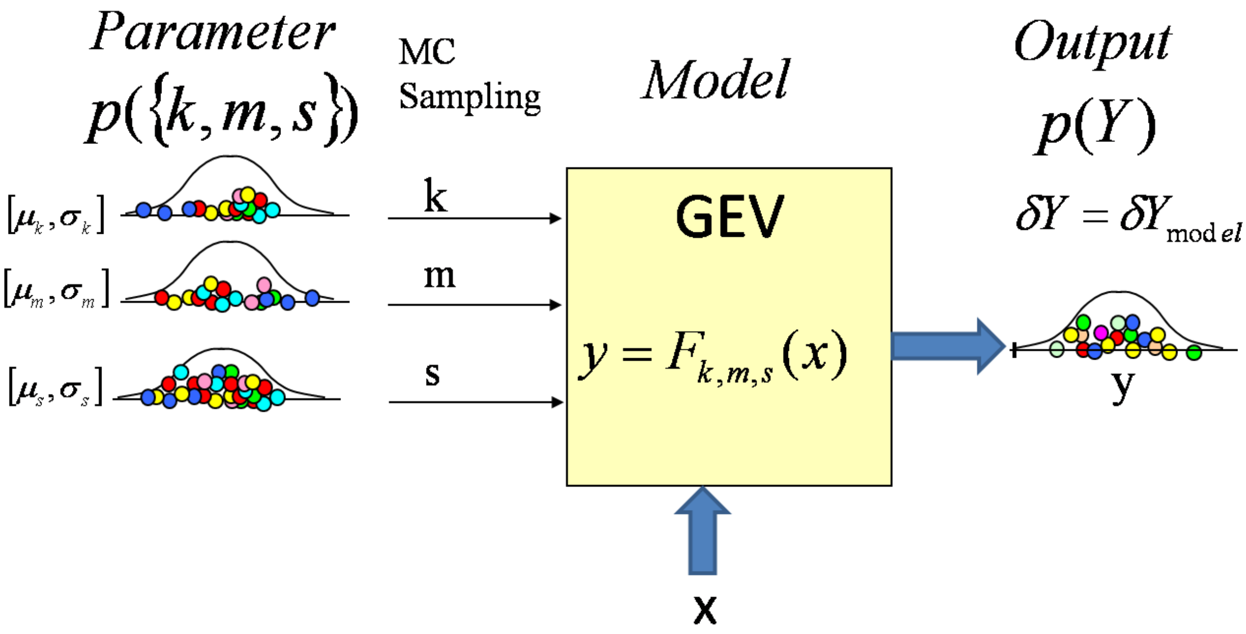

Fig.1. Model uncertainty analysis methodology

2.6. Sampling Methodology

In general sampling is a process to draw a sample from a statistical event space. The event space is described in the form of continuous statistical model or distributions such as uniform, normal, exponential, log-normal, Weibull, etc. In general, computer generates uniform random number in the range [0, 1]. These random numbers are used to generate random sample from the desired model or distribution by one-to-one mapping of cumulative distribution function (CDF) of desired distribution and uniform distribution. This methodology has been graphically represented in the Fig. 2. In the Fig. 2, bottom left side plot represent the uniform distribution in range [0, 1]. Top left plot shows the CDF of the bottom uniform distribution. Bottom right side plot of Fig. 2

represents the desired distribution and corresponding CDF is shown in the top right plot. As both the CDF is bounded by range [0, 1], these can be mapped one-to-one. For the continuous distribution model this method is known as inverse sampling method. The methodology for generating a sample from the computer generated random number (i.e., ξ) has been shown in the dotted arrow form in the Fig. 2.

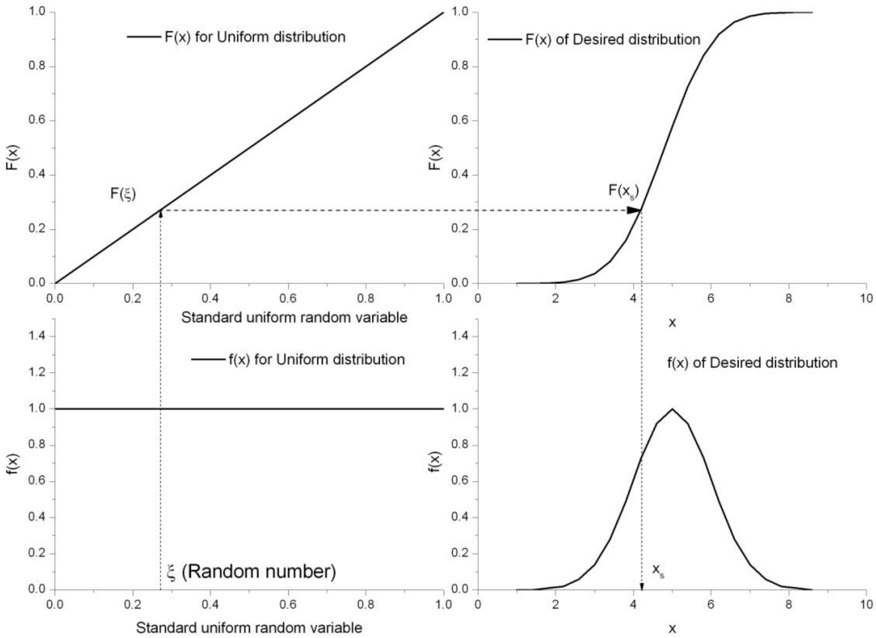

Fig. 2. Sampling methodology

2.7. Statistical Aggregation Methodology

Four stations generate the four GEV models with different set of three parameters (i.e. k, m, s). Average model from these three models has been generated weighted average of quantile data of each model. Quantile data is mathematically represented as inverse of the Equation (1). Mathematically the quantile information is represented in Equation (5).

$$x = \left(m - \frac{s}{k}\right) + \frac{s}{k}(-\ln F)^{-k} \quad (5)$$

For a given 'F', four values of 'x' can be generated for four models. If the weight given for each model is 'w', then the average value of 'x' will be as given in Equation (6).

$$\bar{x} = \frac{w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4}{w_1 + w_2 + w_3 + w_4} \quad (6)$$

Weights can be decided based on various attributes such as the distance of the measuring station from the site, no of data points available and reliability of the data, etc. Weights can also be generated based on expert elicitation method/process. The average value of ‘x’ will be generated for different value of ‘F’. These data can be used to regenerate the average model of four models. The average GEV model will have different set of three parameter data compared to the station wise GEV models. The generated average model preserved the statistical property of the GEV.

3. Results and Discussions

3.1 Statistical Aggregation of GEV models

Maximum wind speed data given in km/h unit in a block of time period of one year for four stations have been plotted in the Fig. 3. It is noted that the lowest and highest 50th percentile value is found in the measured data in station#1 and station#3 respectively. For the range of wind speed from 10 km/h to 80 km/h, it is found that station#1 and station#3 have enveloping distribution function. Each data have been fitted with GEV distribution function as given in Equation (1) in section 2.1 to obtain the shape parameter, location parameter and scale parameter.

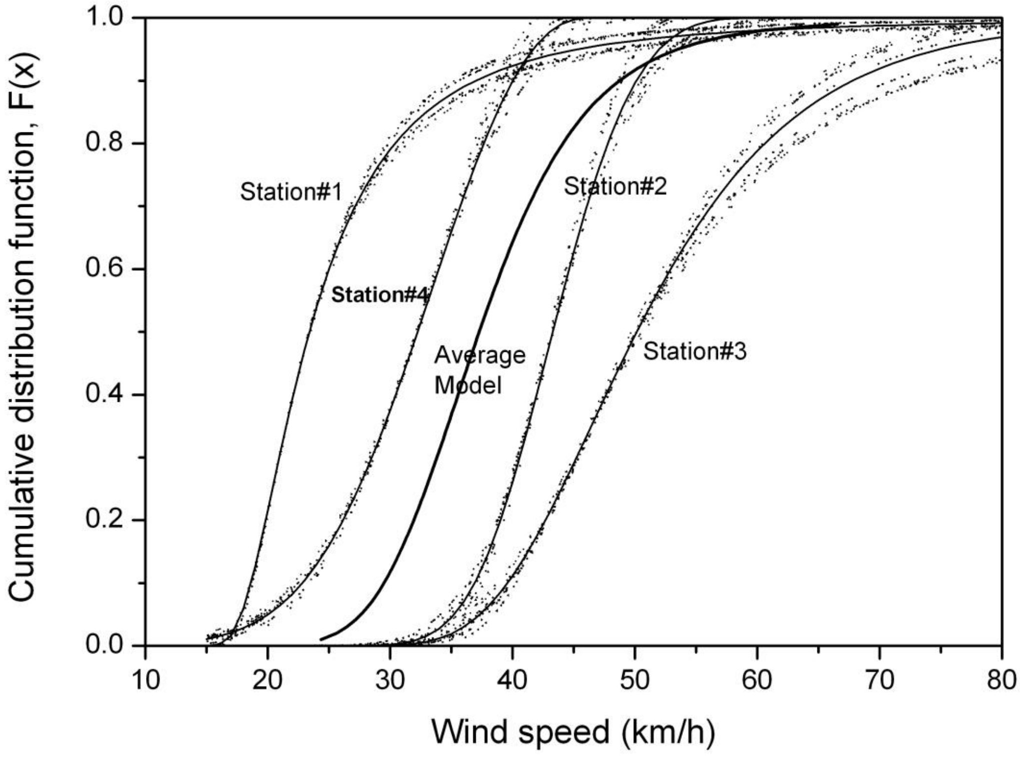

Fig. 3. GEV models for four stations

The details of the curve fitting has been reported in (S. Bera et al., MS-17-65, 2017)[14]. Model uncertainty for four GEV models has been included in Fig. 3.

For the statistical aggregation needs weights of individual GEV models. In this statistical aggregation study, equal weights are given for the each GEV model. The estimated average GEV model is shown as average model in Fig. 3. It is noted that the average GEV distribution follows in between the four individual GEV models. Again this average GEV model data has been fitted with GEV distribution function to obtain the three parameters. The estimated three parameters i.e. k, m, s are -0.04216, 34.9559, 6.34007 respectively.

3.2 Impact on Return Period

The variation of return period with wind speed has been estimated based on the probability of exceedance and shown in Figure 4.

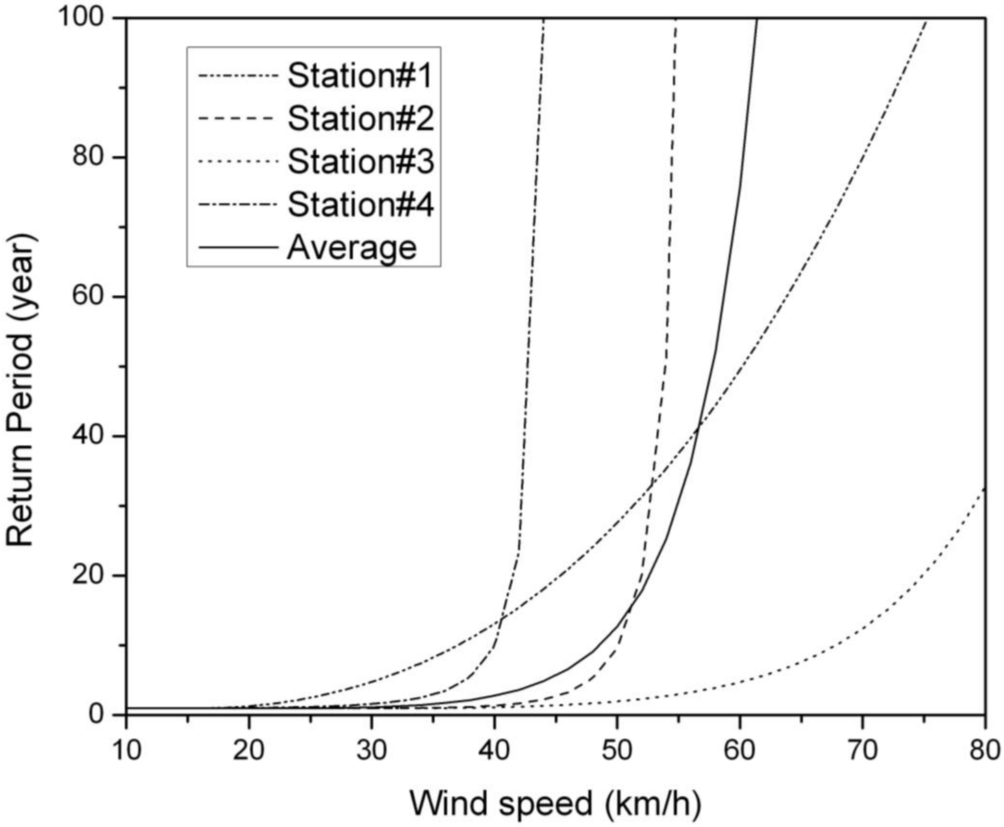

Figure 4. Variation of return period

It is found that for low wind speed, the variation of return period is small among all the models. However, the variation of the return period for high wind speed is found to be significantly different among models. As the stations are not actual representation of the site, average model can be used to predict return period.

4. Conclusion

Historical wind speed data collected from four measuring stations near to a new site are used to demonstrate the statistical aggregation of wind hazards models. Multiple models are developed for each measuring stations. Model uncertainty due to the low size of the sample size has been assessed using Monte Carlo based sampling techniques. The average GEV model is developed with model parameters i.e., 'k', 'm', 's' equal to -0.04216, 34.9559, 6.34007 respectively. It is found that for low wind speed, the variation of return period is small among all the models. However, the variation of the return period for high wind speed is found to be significantly different among models. As the stations are not actual representation of the site, average model can be used to predict return period and in turn to be used in designing safety important engineering structure, system and component till actual data is made available.

References

1. Laurens de Hann, Ana Ferreira, "Extreme value theory an introduction", Springer,(2006)Galambos, J. and Kotz, S., "Characterizations of Probability Distributions", 1978. Berlin: Springer.
2. Galambos, J. and Kotz, S., "Characterizations of Probability Distributions", 1978. Berlin: Springer.
3. Cox, D. R. and Hinkley, D. V., "Theoretical Statistics", 1974. London: Chapman and Hall.
4. AERB Safety Guide, "Extreme value of mateorological parameters", AERB/NF/SG/S-3,(2008)
5. R. D. Reiss, M. Thomas, "Statistical Analysis of extreme values with applications to insurance, Finance, hydrology and other fields", Birkhauser Verlag (2007)
6. Fisher, R. A. and Tippett, L. H. C., "Limiting forms of the frequency distribution of the largest or smallest member of a sample", Proceedings of the Cambridge Philosophical Society, vol 24, 1928, pp. 180–290.
7. Samuel Kotz, Saralees Nadarajah, "Extreme Value distributions theory and applications", Imperial College Press (2000)
8. Stuart Coles, "An introduction to statistical modeling of Extreme values", Springer (2001).
9. Castillo, E. and Hadi, A. S., "Fitting the generalized Pareto distribution to data", Journal of the American Statistical Association, vol 92, 1997, pp. 1609–1620.

10. Davison, A. C. and Smith, R. L., "Models for exceedances over high thresholds (with discussion)", Journal of the Royal Statistical Society, B, vol 52, 1990, pp. 393–442.
11. Hosking, J. R. M., Wallis, J. R. and Wood, E. F., "Estimation of the generalized extreme value distribution by the method of probability-weighted moments", Technometrics, vol 27, 1985, pp. 251–261.
12. Subrata Bera, U. K. Paul, D. Datta, A. J. Gaikwad, "Uncertainty in fission product transient release under accident condition", Proceedings of International Conference on Modelling and Simulation (MS-17), Association for the Advancement of Modelling and simulation Techniques in Enterprises(A. M. S. E), HHI, Kolakta, India, November 4-5, 2017.
13. Subrata Bera, Dhanesh B. Nagrale, U. K. Paul, D. Data, A. J. Gaikwad, "Statistical aggregation of extreme value analysis models", Proceedings of International Conference on Modelling and Simulation (MS-17), Association for the Advancement of Modelling and simulation Techniques in Enterprises(A. M. S. E), HHI, Kolakta, India, November 4-5, 2017.
14. Finner, H. and Roters, M., "On the limit behavior of the joint distribution function of order-statistics", Annals of the Institute of Statistical Mathematics, vol 46, 1994, pp. 343–349.
15. Galambos, J., "A statistical test for extreme value distributions", In Nonparametric Statistical Inference, 1982, pp. 221–230, editors B. V. Gnedenko, M. L. Puri and I. Vincze. Amsterdam: North-Holland.